



FUNCTIONAL CLASSIFICATION OF SPACE

**Aspects of site suitability assessment in
a decision support environment**

René F. Reitsma
International Institute for Applied Systems Analysis
Laxenburg, Austria

RR-90-2
April 1990

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

International Standard Book Number 3-7045-0100-X

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Copyright ©1990
International Institute for Applied Systems Analysis

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system, without permission in writing from the publisher.

Cover design by Martin Schobel
Printed by Novographic, Vienna, Austria

Foreword

The research described in this report resulted from a series of international collaborations, beginning with the Faculty of Policy Sciences of the University of Nijmegen, the Netherlands, and IIASA, within the framework of IIASA's Young Summer Scientists Program (YSSP).

What started as an innocent YSSP assignment was turned into one of the building blocks of a project on *Expert Systems for Integrated Development: A Case Study of Shanxi Province*, a collaboration between IIASA's Advanced Computer Applications (ACA) group and the State Science and Technology Commission of the People's Republic of China.

REPLACE (the remainder of this report explains this acronym and its deeper eschatological meaning) is one of the true 'expert systems' components of the Shanxi Province prototype; not only was it implemented in Prolog, which made it attractive to the ACA project and our Chinese colleagues, it also represents, with a considerable amount of hindsight, an extension to the decision support philosophy of 'satisficing', cultivated at IIASA's Systems and Decision Sciences' Methods of Decision Analysis (MDA) project.

Combining a soft optimization paradigm with Prolog's power to use numerical, as well as symbolic, descriptors in describing a location's properties as well as an activity's requirements, and matching them, using logic rather than number crunching, resulted in a system of considerable appeal and great promise. Adding the graphical interaction and display features that are the hallmark of ACA's developments resulted in a package that impressively illustrates the underlying concepts and ideas in computational geography and locational analysis.

As proof of the system's more applied powers, it successfully matched **Drs.** René Reitsma with a proper **Dr.**

Kurt Fedra
Advanced Computer Applications

Preface and Acknowledgments

The research reported on here was conducted at both the Department of Geography of the Faculty of Policy Sciences, the University of Nijmegen, the Netherlands, and the Advanced Computer Applications (ACA) project at IIASA. Most of the work presented in the first four chapters was carried out at Nijmegen. The result of this, a prototype system for developing and computing functional classifications, was taken to IIASA where it was enhanced and extended into the REPLACE (RELational Plant Location and ACquisition Enquiry) system. REPLACE itself was again integrated into a larger decision support system developed for assistance in planning the reorganization of the regional economy of the province of Shanxi, the People's Republic of China. The work presented here therefore contains elements of both a theoretical and an application-oriented nature. The work done at IIASA, the application, provided excellent opportunities for critically evaluating the theoretical work. The writing of the manuscript implied reconsidering the theory and this again created ample opportunities to reflect on the application.

In order to understand how the REPLACE applications finally turned out, it might be useful to explain a little more of the Nijmegen-IIASA-Shanxi connection and of how REPLACE was empirically embedded in the Shanxi Province Decision Support System.

At the time I contacted IIASA, the ACA research group, in collaboration with the State Science and Technology Commission of China (SSTCC), was already conducting the Shanxi Province case study. The research problem was the one outlined in Chapter 5—model-based decision support for integrated planning. The research was carried out by IIASA scholars and Chinese scholars invited to work at IIASA. It is worth mentioning that the

main division of tasks was that IIASA would provide most of the models and the interfacing (model-model, model-data base, and system-user), whereas the Chinese were responsible for the empirical content of the system, i.e., the content of the data bases and the empirical case studies relevant to the Shanxi problem.

In its core-form—the model as presented in Chapter 4—REPLACE was considered a potentially useful contribution to the Shanxi Province DSS. The more so, since no site suitability locational planning model was available at that moment. It was therefore decided to port the model over to the Shanxi DSS, augment and enhance it with all the frills and furbelows mentioned in Chapter 5, and implement some test cases which were considered relevant in the Shanxi Province context.

Activity and locational data, as well as the initial rule bases, were delivered by the Chinese scholars working at IIASA during the development of the system. Decisions had to be made on how to proceed in the development of a few prototype rule bases. It was decided to concentrate on only a few activities the locational aspects of which were considered particularly interesting for the province of Shanxi: the chemical industry and the aluminum production and processing industries. The examples in Chapter 6 are therefore taken from these three activities.

The dedicated GIS presented in Section 5.7.9 was developed together with Brian Makare (currently at the University of Colorado at Boulder, USA). Much of the work on the implementation of the optimization process presented in Section 6.8 was carried out by Maurits van der Vlugt (currently at the Physical Planning Agency (RPD), The Hague, the Netherlands), and Steven Markstrom (currently at IIASA).

The Manuscript Committee consisted of Prof. Dr. E. Wever, Prof. Dr. H.J.P. Timmermans and Dr. K. Strzepek. The Promoter was Prof. Kouwe and the Co-promoters were Dr. K. Fedra and Dr. T. van der Smagt.

Trying to think of all the people who contributed to the research presented in this report, however, makes one feel uncanny about all those little things which, in case they would not have happened, or in case they would have occurred in just a slightly different version, would have seriously jeopardized the necessity of these very acknowledgements. It reminds one of the ‘butterfly-effect’; that if somewhere, sometime in the history of this planet, a butterfly had decided to ‘take-off’ from the flower it had sat on just a few seconds earlier than it actually did, it would have triggered a chain reaction changing precisely one or two of those things that kept this project going during its various phases. But as Gleick (1988) and many of the scientists

whose work he presents explains, many chaotic processes contain structure in the form of so-called ‘strange attractors’, sequences of system’s states which once mapped into phase-space show the underlying stable dynamics of the system’s evolution. I suppose that I have been very fortunate to have encountered so many people featuring as ‘attractors’, stabilizing the otherwise chaotic process of dissertation research.

I would like to thank Dr. Ton van der Smagt, Dr. Paul Hendriks, Larry Lucardie, Prof. Kouwe and Dr. Fedra for a lot of thinking; Emile Gemmeke and Shetang Yang for getting me started; Lothar Winkelbauer for keeping me going (and for some interesting interludes); Dr. Fedra and Prof. Z. Wang for giving me a chance; Elisabeth Weigkricht for helping with the colors and a good game of tennis; Anna Korula John for doing the editing, for supplying me with caloric input and who, together with Barbara Hauser, administered good spirits; Yongtai Liu, ‘Madam’ Wang and H. Xu for providing lots of data and for helping me understand some of Shanxi Province; Brian Makare for helping me to learn C, for many valuable ideas, and lots of fun; and finally, all those people at Nijmegen, IIASA, or anywhere else, who contributed to this project and who made the past five years so rewarding.

René F. Reitsma
Nijmegen/Laxenburg
April, 1990

Contents

Foreword	iii
Preface and Acknowledgments	v
Introduction	1
1 A Relational Approach to Spatial Choice Modeling	9
1.1 Introduction: spatial choice and decision making	9
1.2 Constraints versus preference	11
1.2.1 Utility and spatial choice	13
1.2.1.1 Attributes and their categories	14
1.2.1.2 Problems with preference	17
1.2.2 An alternative based on functionality	19
1.2.2.1 Taxonomic versus functional classification	20
1.3 A two-stage model of spatial decision making	22
1.4 Functional concepts and relational matching	25
1.4.1 INUS conditionality and relational definitions	27
1.4.2 Representing relational definitions	28
1.5 Choice behavior and decision making	30
1.6 Research questions and objectives	31
2 Site Suitability in Location Theory	35
2.1 Location theory and spatial planning	36
2.1.1 Behavioral location theory	38
2.1.2 Functional location theory	39
2.1.3 Functional location theory: implications for model building	42
2.2 The concept of production milieu: definition and measurement	43
2.2.1 The R.E.B. approach toward production milieu	44

2.2.2	The production milieu matrix method	47
2.2.3	R.E.B. methodology and production milieu matrix: conclusions	51
2.3	The concept of regional indigenous potential	51
2.4	Regional potential and production milieu: a closer look . . .	54
2.4.1	Regional potential and production milieu: some conclusions	56
2.5	Production milieu as a choice set	57
2.5.1	Rawstron's three restriction principles	58
2.5.2	Feasibility studies: the matching approach to site suitability	61
2.5.2.1	The "Standortfaktoren" catalog of Schilling	61
2.5.2.2	Evaluating the matching approach: advantages	62
2.5.2.3	Evaluating the matching approach: problems	69
2.5.3	Intermezzo: automating the matching process	75
2.5.4	Computer-based matching: Cullen's proposal	77
2.5.5	Studying individual firms	78
2.6	Conclusion: toward a relational matching model	80
3	Relational Inference Trees	83
3.1	Introduction	83
3.2	Relational definitions as inference trees	84
3.2.1	Formal properties of a relational inference tree	87
3.2.2	Dimensions	88
3.2.3	Internally related variables	92
3.2.4	Internally related dimensions	93
3.2.5	Dimensions: additivity and the partitioning problem	95
3.3	Relational inference trees versus decision plan nets	97
3.4	Model content: formal evaluation procedures	99
3.4.1	Completeness and consistency	101
3.4.2	Optimization	103
3.4.2.1	Optimizing with information entropy: ID3	104
3.4.2.2	Pitfalls: inconsistency and incompleteness	107
3.4.2.3	Single-action rule sets	108
3.4.2.4	ID3: conclusion	109
3.5	Actor attributes and requirements in the inference tree: some critical remarks	111
3.5.1	The redundancy problem	111
3.5.2	The modeling problem	112

3.6	Modeling demands generated by objects	114
3.7	Two-way matching	117
3.8	Model content: empirical procedures	120
3.8.1	Knowing the objects	120
3.8.2	Knowing the actors	121
3.8.3	The problem of generalization	123
3.9	Knowing the inference tree: techniques of knowledge acquisition	123
3.9.1	Inducing inference trees: efficiency versus understandability	127
3.9.2	Validating the model	128
3.10	Multiple matching solutions	130
3.11	Multi-locational actors	132
3.12	Conclusions	134
4	Implementation	135
4.1	Structure of the matching system	135
4.2	Frames for representing facts and rules	139
4.3	Find-deduce-inherit-ask inference	143
4.4	A Prolog implementation	145
4.4.1	Frames and inference in Prolog	146
4.4.2	Inference trees and matching rules in Prolog	150
4.4.3	The matching program	152
4.4.4	Explanation	155
4.5	Implementing generalization	156
4.6	Conclusion	158
5	Decision Support Systems and the REPLACE System	161
5.1	Introduction: model use and decision support	162
5.2	Decision Support Systems: an overview	164
5.2.1	Why: ill-structured problems	165
5.2.1.1	Well-defined, ill-structured, and functional equivalence	166
5.2.2	What: constraints and pre-defined goal states	168
5.2.3	How: data bases, models, user-interaction	169
5.3	Intermezzo: expert systems (ESs), DSSs, and spatial planning	171
5.4	Decision support in planning a regional economy	173
5.5	Model integration	178
5.6	The Shanxi Province DSS	179

5.6.1	Shanxi Province: geography and development	179
5.6.2	Shanxi Province: the problem	184
5.6.3	Overview of the DSS	188
5.6.4	REPLACE: RElational Plant Location and ACquisition Enquiry	192
5.7	REPLACE: functions, implementation and 'looks'	192
5.7.1	Main menu	197
5.7.2	'Run the matching model'	199
5.7.2.1	'Explain matching results'	200
5.7.3	'Select the activity'	202
5.7.4	'Select set of counties'	204
5.7.5	'Select the dimensions'	205
5.7.6	'Edit current activity'	205
5.7.6.1	Dynamic editing of activity attributes	206
5.7.6.2	Select/deselect 'don't care' values	209
5.7.7	'Edit county data base'	210
5.7.8	'Set decision attributes'	211
5.7.9	REPLACE data-analysis module: a dedicated GIS	213
5.7.9.1	'Reclassify the attribute'	215
5.7.9.2	'Compute Bi-variate statistics'	217
5.8	Conclusion	219
6	Examples and Applications	221
6.1	Introduction: REPLACE in the Shanxi Province case study	221
6.2	Data problems	223
6.3	Generating production milieu maps	230
6.3.1	Pattern interpretation and inductive knowledge acquisition	232
6.3.2	Iterative model building: aluminum industry	234
6.4	Matching with demands by locations	238
6.5	Modeling agglomeration economies by means of linkages	241
6.6	Complex matching and 'don't care' values: an example	244
6.7	Implementing a two-step choice process: PDA, PDAS and REPLACE	247
6.8	Implementing a two-step choice process: REPLACE-DISCRET	249
6.8.1	Exporting attributes to the evaluation model	250
6.8.1.1	A measure of the robustness of matching solutions	252

<i>Contents</i>	xiii
6.9 Conclusion: Relational complexity and relational model building	253
7 Conclusions and Discussion	255
7.1 Relational modeling in a two-step process of choice modeling	255
7.2 Relational modeling by means of inference trees	257
7.3 Matching with dimensions	258
7.4 Relational modeling: explanation versus decision support . .	259
7.5 Building real-world applications: accepting complexity	261
References	265
Appendix 1: Some remarks on INUS conditionality	287
Appendix 2: Translations of the Schilling Catalog (Table 2.1)	291
Appendix 3: The Prolog programming language	295
Appendix 4: Hardware and software	301

Introduction

Bavaria is a great place for locating your business; that is what the advertisement shown in Figure 0.1 claims to engrave into the reader's long-term memory ((LTM), Smith *et al.*, 1982, 1984). Some day, if the reader is in a position to take a decision as to where a new branch plant is to be established, the advertisers hope that Bavaria will be recalled immediately as a suitable location.

The advertisement is interesting for several reasons. First of all, it not only informs us that Bavaria offers a very fertile soil for entrepreneurship, it also explains why this is so. Bavaria is characterized by a stable government, it is the largest state in the Federal Republic of Germany, and ample industrial sites are available. It is furthermore claimed to be Europe's continental semi-conductor center, it is located in the heart of one of Europe's highest concentrations of user industries of advanced technology products, it is a land of traditional values, and so forth.

A second aspect that makes the advertisement interesting is that it constitutes a nice example of what is sometimes called 'taxonomic classification' (van der Smagt, 1988), a means of classification by which objects get assigned to classes on the basis of empirical similarity. If objects are similar with respect to their empirical properties they are assigned to identical classes. The advertisement seems to be based on the idea that areas can be good or bad for establishing a business, or that areas can have a value assigned on a 'business locational suitability index' scale ranging from bad to good. Regions that score high on the favorable characteristics are then assigned to the 'attractive', the 'fertile' group. Bavaria: fertile soil for high-performance companies, therefore.

'Fertile soil' here, of course refers to the suitability for locating a company at a specific site. The metaphor is well chosen, because if one tries to come up with a definition of either soil fertility or site suitability, it has to be recognized that what is really being evaluated is a functional relationship

Love at first site.

Obviously something has been attracting technically advanced firms from around the world to Bavaria, making it Europe's premier location for high-performance companies. Consider this combination of typical strengths:

1. Bavaria has a traditionally stable government committed to the principle of help for self-help. Solid support for education and vocational training, research and technology transfer, a consistent strengthening of the state's infrastructure, and a variety of investment incentives.

2. Germany's largest federal state, covering 28% of the nation's land area, Bavaria offers sites for industrial settlement practically throughout the state. A company locating in Bavaria can tailor its facilities from available space amounting to more than 27,000 acres.

3. Bavaria, Continental Europe's semiconductor center, is a technically advanced economy employing, for example, one-fourth of all Germans active in electronics and electro-technology—in industries such as components manufacture, telecommunications, dataprocessing, entertainment electronics, and electromedicine.

4. Bavaria is at the heart of one of Europe's highest concentrations of user industries of advanced technology products, with electrical and mechanical engineering, automakers, and the German aviation and aerospace industries—about 50% of which are in Bavaria—heading the list.

5. Bavaria, where more people are employed in R&D in the industrial sector than in any other state, is a leading research center. It is the home of numerous universities and technical colleges as well as of the world-famous Max-Planck Institute, Fraunhofer-Gesellschaft and new research facilities for microelectronics and biotechnology.

6. Bavaria is a land of traditional values. Entrepreneurial drive and a strong work ethic are the bedrock of growing prosperity in Bavaria where government, the business and labor community, as well as the academic world have linked their resources to create an ideal environment for high-performance companies.

Not bad for a state famous for its charming lifestyle. If your company is considering expansion in Europe, take a close look at Bavaria. It'll be love at first site.



Bavarian State Ministry of Economics
D-8000 Munich 22, P.O. B.
Federal Republic of Germany

Please send information about location opportunities in Bavaria.

Name _____
Company _____
Street _____
City _____

Bavaria. Fertile soil for high-performance companies.

SAD

Figure 0.1: Advertising a location for entrepreneurship (Source: Scientific American, October 1986; p. 5)

between a piece of land, a region, or more generally, 'space', and the specific objective to be realized by the use of this space. Soil is fertile if crops can be grown on it; locations offer suitable sites if activities can be located and run on them.

What is at issue therefore are end-means relationships. There are objectives to be met e.g., growing crops, the establishment of residential environments, or operating a business; the allocation of space must help realize these objectives. Whether the allocation of a specific space is the only means of realizing an objective, one of several possibilities, or just one aspect out of a whole range of means, depends on the nature of the objective. If the objective is the establishment of a company's branch plant with its associated production scheme, a specific location only fulfills a role in a much more complex system of end-means relationships.

Classification based on end-means relationships can be denoted 'functional classification'. Objects (areas) are assigned to distinct classes if they can fulfill specific functions associated with these classes. Unlike taxonomic classifications which are based on empirical similarity, functional classifications are based on similarity in end-means relationships.

Functional classifications are quite familiar, although one does not always realize that such classifications are being used. Sometimes they evolve rather naturally and their functional character is easily recognized. Adewolde-Osunade (1988), for instance, studied soil classification schemes used among small farmers in Southwestern Nigeria. It appears that these people categorize their soils in clear functional classes. Categories such as 'Yanrin', 'Bole', or 'Alaadun' soils have very specific utilities for growing crops; they represent crop-specific fertility classes. Or as Adewolde-Osunade (1988; p. 200) puts it, "The approach to soil suitability by the small farmers is a pragmatic one. Thus, it is an example of an empirical classification system, in which the properties that appear significant for crop growth are those used for classification."

Of course, each of the soil types recognized by the farmers can be characterized by a set of empirical properties (Table 0.1). But what is important here is that it is not the empirical properties which determine the soil classification. Instead, it is the functional relationship between sets of empirical soil properties and a specific type of soil use that determines the classification. In other words, although similar soils will be assigned to the same class, the criteria determining whether or not soils are to be considered similar depend on the objective; the use it will be put to. Unlike the above example, the functional nature of many other classifications used is much harder to

Table 0.1: Soil classification by small farmers in Nigeria (Source: Adewolde Osunade, 1988)

Classes	Soil types
<i>Wokuta</i> (stony soils)	<i>Taraa</i> (gravelly soil) <i>Yangi</i> (lateritic soil) <i>Wokuta</i> (stony soil) <i>Wokuta pupa</i> (reddish brown stony soil) <i>Wokuta dudu</i> (dark stony soil)
<i>Yanrin</i> (sandy soils)	<i>Yanrin</i> (coarse sandy soil) <i>Yanrin pupa</i> (reddish brown sandy clay soil) <i>Yanrin funfun</i> (light sandy soil) <i>Yanrin dudu</i> (dark sandy clay soil)
<i>Bole</i> (clay soils)	<i>Bole</i> (clay soil) <i>Bole pupa</i> (reddish brown clay soil) <i>Bole dudu</i> (dark clay loamy soil) <i>Bole funfun</i> (light clay sandy soil) <i>Bole olokuta</i> (stony clay soil) <i>Bole alaadun</i> (loamy clay soil)
<i>Alaadun</i> (loamy soils)	<i>Alaadun</i> (loamy soil) <i>Alaadun pupa</i> (reddish brown loamy soil) <i>Alaadun dudu</i> (dark loamy soil)

recognize. Hendriks (1986; pp. 3-4), for instance, discusses the example of the concept of a 'house'. At first sight, there seems to be no problem at all. One can wander through a neighborhood and count houses. Or can one? What for instance would have to be done if one is asked to look for a house suitable for a family with a specific income and family composition? It is possible that the number of 'houses' in that same neighborhood can be significantly reduced, simply because most houses are either too small, too large, or too expensive. Apparently, the same neighborhood with the same buildings in it can have different numbers of houses, depending on who the classification is meant for.

This intrinsically functional character of a concept such as 'house' or 'dwelling' is elegantly represented in a conceptual model of urban residential mobility by Brown and Moore (1970, 1971). The concept of 'stress' occupies a central position in this model: Stress is supposed to be a measure

of the discrepancy between what the subjects, whose migration behavior is studied, require of their home, and what their current house and its environment provides. Stress, therefore, is the interactive result of confronting needs with provisions, and characterizes the utility of the place of residence. This implies a rather functional approach to utility and the concept of house or dwelling. Subjects need things, they want their house to provide certain functions, and they try to realize them by finding the house that provides those functions. If such an alternative cannot be found, four options remain. The first three concern the aspiration level; adjust (lower) the aspiration level, maintain the aspiration level and get terribly frustrated, or lower the aspiration level temporarily and change the personal or household characteristics in a direction that enables a future resurgence of the original aspiration level with a better chance of realizing it. A fourth, and for functional classification most interesting alternative, concerns the adaptation of the residential environment itself in order to meet the aspiration level. In many instances such alternative lines of action, each of which may lead to a significant decrease or even entire resolution of the stress, do indeed exist. Such alternatives may be denoted as 'functionally equivalent'. Each of the alternatives, although empirically entirely dissimilar, can be considered a possible means in realizing a specific goal. Therefore, although such alternatives will hardly be put into one class if the classification procedure is taxonomic, they will be put in one and the same class if the classification procedure is functional. The Brown-Moore model is conceptually attractive. Functionality as the critical variable in decision making is at its heart, and it is this functionality that is often recognized in modern theories of decision making. In Chapter 2, functional aspects in economic geographical location theory are discussed. It is argued that modern versions of this body of theory take an explicit functional position concerning concepts such as production milieu or site suitability.

The problem

Unfortunately, however, much of that functionality is often lost during the actual process of model building and the development of measurement procedures. This certainly seems to hold for location theory, and some examples of this will be presented. Traditionally, geographical models are of a statistical or mathematical nature. These models, however, do not seem well suited to capture the notion of functionality. There are several reasons for this. First, traditional mathematical models describe the dynamics of, and

associations between, patterns and processes regarding the mechanisms that generate them as more or less given facts. What is described are the dynamics of these patterns and processes as generated by aggregates of individual behavior. Functionality, on the other hand, implies great variability in interests and objectives among individual subjects and social groups, something which is hard to capture in, for instance, a spatial interaction model.

A problem with traditional statistical modeling is that it is based on empirical similarity. Objects and locations are characterized by their similarity in empirical properties. The soil classification example, however, shows that empirical similarity is basically irrelevant where functionality is concerned. What is important is similarity between the actor's requirements and the object's properties, since it is by this interaction relationship that functionality is determined. Moreover, statistical and mathematical models can be considered members of a broader class of models, often denoted as 'equation models'. Functionality, however, is something that is often hard to model by means of an equation. The reason being that modeling functionality requires that there is a possibility to incorporate semantics in the model. This implies that the interpretation of variables can be different in different situations; something that is hard, if not impossible, to incorporate into an equation.

It is perhaps for these reasons that the Brown–Moore model proved to be so hard to implement and apply in a way that was consistent with its conceptual, theoretical content. At the time it was 'invented', the spatial analysis paradigm with its emphasis on aggregate patterns and statistical relationships dominated the scene. Its tools, however, were not really suitable for tackling the problem (Reitsma and Vergoossen, 1988).

How does all this relate to the 'fertility' of Bavaria's soil? The fact that geographers and economists have been working on location theory for such a long time suggests that for the 'economic fertility' of regions, no straightforward functional classification(s) such as the one maintained by the Nigerian farmers was available. This is not surprising. Different economic activities can have very different objectives. And even if they have similar objectives, the means for realizing them do not have to have much in common. Similarly, one specific activity often has a choice between various objectives and alternative means for realizing them. As Massey (1979b) has pointed out, a particular stretch of geographical space can represent many different opportunities as well as impediments to a large variety of economic agents. The number of ways in which this space can function as a location is far too large to be contained in a simple classification scheme. The feature economic fertility and agricultural fertility have in common, however, is that the 'fer-

tility' of the 'soil' cannot be assessed by means of a set of empirical properties such as the ones shown in Figure 0.1 and Table 0.1. Fertility, production milieu, accessibility, service level, and site suitability: the meaning of all these concepts contains a clear functional aspect, to be revealed and modeled by means of a functional classification based on functional modeling.

Nevertheless, although various bodies of modern theory on spatial behavior do indeed recognize the relevance of functionality as a key issue, most methods and techniques currently available for the development of explanatory models of such behavior must be considered incapable of adequately representing functionality. The main objective of this study, therefore, is to find ways of modeling functionality and developing functional concepts and classifications, for use in geographical research applications. An attempt is made to formulate an alternative to mathematical and statistical models which suits the functional characteristics of (geographical) concepts better and which can be used in empirical applications and decision support. As is explained in Chapter 5, decision support presupposes a typically functional point of view.

What then are the problems which make functionality hard to model? This is discussed in Chapter 1. It is argued that for modeling functionality two important problems need to be resolved. The first concerns the origin and reconstruction of the functions actors require from objects or, when restricted to geographical applications, from space. When functionality is the central issue, which functions are important, and which objects can fulfill which functions? In this chapter a general method for representing functional concepts, based on earlier research (van der Smagt, 1985; Hendriks, 1986) is introduced. The second problem concerns the way these methodological guidelines can be integrated into an actual modeling procedure; something dealt with in Chapter 3.

Chapter 2 is meant to show that many of the more general modeling problems discussed in Chapter 1 can be recognized in the development and applications of economic geographical industrial location theory. Several attempts at modeling the concepts of production milieu, site suitability, and regional potential will be discussed and critically evaluated. The conclusion of the chapter is that although functionality is at the heart of modern location theory, modeling attempts have been unsuccessful in capturing this functional character.

A proposal for an alternative is then formulated in Chapter 3. A modeling technique based on the argumentation of the first two chapters is presented and evaluated. This technique of 'relational matching' is extensively

discussed and evaluated. Many aspects of the functional classification of space that were not discussed, or only briefly mentioned, in the first two chapters are elaborated on in this chapter.

Chapter 4 presents an implementation of this modeling technique on a computer. The techniques for implementing a 'relational matching model' are discussed and presented, and various implementational issues are dealt with. Special attention is given to the Prolog programming language and expert system technology because of their usefulness in implementing relational matching models and automated matching systems.

In Chapter 5 the discussion turns to the issue of decision support, which was already touched upon in Chapter 1. The chapter opens with a brief discussion on the nature and characteristics of decision support. It is argued that many spatial planning problems can be formulated in a form which is suited for decision support system implementations. As an example, a decision support system developed at IIASA, for the re-organization of the regional economy in the province of Shanxi, PRC, is presented. This system contains an enhanced version of the relational matching model presented in Chapter 4 as one of its modules; the REPLACE system is presented and its functions are discussed.

Chapter 6 contains a few generic examples from the empirical application introduced in Chapter 5. Some of the problems associated with applying the REPLACE system to a case study of Shanxi Province are mentioned. The examples address some well-known issues in location theory such as the modeling of linkages, and the interpretation of site suitability patterns. Other examples show some initial attempts at arriving at a structured method for determining the empirical content of a relational matching model, and a possibility to link the site suitability model with various optimization models.

Chapter 7 briefly reconsiders matters as discussed in the previous chapters. An overall evaluation of the research project reported on here is presented.

The research draws on various disciplines: social science methodology in Chapters 1 and 3, economic geography in Chapters 2 and 6, and computer science, logic, and applied artificial intelligence (AI) in Chapters 3, 4 and 5. One of the results of this study is that modeling and applying functionality in empirical research does indeed require a somewhat 'integrated' approach. The theory is methodological, the application is geographical, and computer science provides the means with which the measurement tools can be developed. However, the issues have been kept separate where possible.

Chapter 1

A Relational Approach to Spatial Choice Modeling

ABSTRACT

The discussion concentrates on the methods and techniques in theoretical geography that were developed for modeling spatial choice and spatial decision behavior. Particular attention is paid to the role of utility and preferences in some of these models and the associated problems. As an alternative to these approaches, a recently developed, reconstructive, relational alternative for modeling functionality is presented. The objective of the study is the development of a set of model building and measurement tools for implementing this methodological alternative. From this more general objective, a set of more precise research questions and tasks is formulated.

Keywords: spatial choice, preference modeling, utility, constraints, inus-conditions, internal relations, functional equivalence, goal rationality, relational definition

1.1 Introduction: spatial choice and decision making

Patterns and processes studied by social geographers result from the behavior of many different individual actors and groups of actors. People reside and

migrate, they work and indulge in recreation. Companies locate and relocate plants, distribute goods and information, attract workers in the morning and disperse them again at five. The spatial behavior of individual actors and groups of actors is studied in order to understand better the mechanisms by which these aggregate patterns and processes are generated. These individuals are not mere creatures without volition 'behaving' as would particles, mediating the forces exerted on them. Instead, they must be considered subjects with interests and objectives that process information about the world and use it to generate their actions by means of a decision. As a result, (spatial) decision making and (spatial) choice became important research issues in geography. The numerous articles, papers, and reviews published over the last ten years or so, show that its importance in geography has rapidly increased over a relatively short period of time and is now firmly established (see e.g., Golledge and Rushton, 1976; Cox and Golledge, 1981; Golledge and Rayner, 1982; Timmermans, 1984; Golledge and Timmermans, 1988; Smith *et al.*, 1984). Although the extensive literature might give the impression of a 'mature' field of research, many issues are still strongly debated. In fact, subjects such as the nature of the decision-making process, the relations decision makers have with their environment, the role of preferences versus constraints, the rationality of decision making, the way information is processed by a decision maker, as well as how spatial choice behavior and decision making can or must be modeled, are still the subject of ongoing discussion and debate, and several different approaches to modeling spatial choice behavior exist.

One way of reducing this large variety in the approaches and methods into a smaller set of more generic types, is to classify them according to the conceptual and methodological views and opinions out of which they are constructed.

One way of looking at decision making is to regard the behavior of actors as given, as a deterministic response to certain stimuli. This is the approach mentioned above in which actors are regarded as the elements in the system mediating its forces in a well-described manner. It is the approach which in psychology is known as the 'behavioristic' approach (Gould and Kolb, 1964; p. 54; Kuper and Kuper, 1985; p. 65). In this approach actors are treated as black boxes, accepting stimuli from the environment and reacting to those stimuli by a specific response. Mediating or intervening cognitive processes are not modeled. Specific stimuli are associated with specific responses. In geography, this idea of a basically non-decisive, either deterministic or probabilistic, though fully informed, reactive actor

is implicitly assumed in (neo)classical location theory or spatial interaction modeling. However, although this kind of approach might generate interesting associations between spatial patterns and may be used for predictive purposes, its theoretical content is generally considered insufficient and unrealistic. Not many representatives of the ideal-typical actors that these approaches are based on can be found in the real world.

Once it was realized that in order to really explain spatial patterns and processes one had to concentrate on how the behavior causing these patterns itself was generated, attention turned toward spatial behavior in general, and the cognitive processes underlying this behavior in particular.

1.2 Constraints versus preference

It would require considerable study to reconstruct how these somewhat more behavior-oriented views on the explanation of spatial patterns and processes extended throughout the various fields in human geography. It is nevertheless important to note that the overall result is known as 'behavioral geography'. However, Johnston (1981; pp. 19–20) describes behavioral geography as, "An approach to human geography which draws on the assumptions, methods and concepts of behaviorism to identify the cognitive processes through which individuals codify, respond to and react upon their surrounding environments". The interpretation of 'behaviorism' by Johnston is very different from what it normally stands for in psychology. Unlike behavioral psychology, behavioral geography aims at a white-box reconstruction of the actor, coupling stimuli by means of cognitive processes to responses. However, Johnston's description coincides well with what is generally known as the 'behavioral approach' in geography. This is the interpretation which is followed here. As usual, after the initial discussions and theoretical discourse on the usefulness and significance of behavior-oriented geography, two types of developments occurred. There was a strong tendency to introduce the so-called behavioral approaches in the thematic fields such as economic geography (Pred, 1967; Townroe, 1969). The goal here was of course to use the new insights and opinions to arrive at better explanations, better understanding of the issues one was studying. The other development was a specialized reaction in theoretical geography with the objective of the study of spatial choice behavior proper, with emphasis on the decision process itself rather than on the explanation of spatial outcomes. Here it suffices to say that in both fields, thematic geography and theoretical geography, this

behavioral approach led to the application of many alternative methods and techniques of analysis and measurement. To a greater extent than was earlier the case, attention was paid to the decision processes leading to specific types of locational behavior.

One of the early versions of the spatial choice modeling endeavors in theoretical geography attributed spatial choice behavior mainly to actor-specific preferences. In this 'revealed preference' approach behavior is explained by regarding it as a means to satisfy the actor's preferences. As a consequence, various methods and techniques for 'revealing' these preference structures directly from spatial choice behavior were developed (for an overview refer to e.g., Rushton, 1969; Timmermans and Rushton, 1979).

The objection against this approach was that the assumption that behavior is generated by preferences can only be valid in a situation of complete freedom of choice. However, freedom of choice is rare and unevenly distributed among different groups of subjects or interest groups. As a result, behavior cannot be regarded the mere expression of preferences. Instead, it is the more or less complicated result of an interaction of preferences, expressed as objectives and interests, and the opportunities and constraints within the limits of which these objectives can be realized. (Pirie, 1976; Sheppard, 1979; Thrift, 1981; Desbarats, 1983; Hendriks, 1986). Or, as Short (1977; p. 442) remarks in the context of residential choice modeling: "...behavioral aspects of residential mobility are more realistically explained as a form of adaptive behavior to the system of housing supply and allocation, which is of course, dependent on the structure of the wider society".

Another approach, often denoted as the 'expressed preference' approach, tries to bypass the above mentioned problem by explicitly separating the preference structure of subjects from their overall behavior (e.g., Schuler, 1979; Lieber, 1979; Louviere, 1981; Hendriks, 1983; Timmermans, 1984, 1986; Timmermans *et al.*, 1984; van der Heijden, 1986). The preference structure is measured under laboratory conditions representing a constraint-free choice environment. The results are thus presumed to represent 'clean and pure' preferences. The idea is elegant and can be boldly summarized as follows:

- 0) behavior is preference plus constraints;
- 1) model and measure the preference structure;
- 2) predict the behavior on the basis of preference;
- 3) compare the behavior as it is observed outside the laboratory with the predicted behavior;

- 4) if the prediction turns out all right, the preference structure can be regarded a good predictor of behavior.

It should be noted that such an approach, which concentrates entirely on preferences, does not in itself, and automatically, deny the existence and importance of constraints. On the contrary, constraints do get incorporated into the model (step 0). The underlying assumption which allows this is, of course, that preferences and constraints can be independently described and modeled. In the expressed preference approach, these constraints are typically modeled as a kind of random variable. By using various functions representing different possibilities of how the constraint variable is distributed [logit, probit, dogit (Timmermans, 1984; Timmermans and van der Heijden, 1984)], the constraint component in the expressed preference model is basically regarded as ‘noise’.

In expressed preference modeling the preference structure can be uncovered in two ways; compositional and decompositional. In the compositional case, various aspects or ‘dimensions’ of choice alternatives are evaluated separately and then combined into an overall score for the choice alternative by some sort of combination rule. In the decompositional approach, complete choice alternatives are evaluated by respondents. By carefully designing these alternatives so that they represent combinations of attributes and scores, and then analyzing the patterns of the evaluations by the respondents by using a factorial design, the overall evaluation or ‘utility’ scores are decomposed into their constituent aspects, and the combination rule is derived (refer to Timmermans (1984) for a review of the different options and techniques for developing these so-called ‘(decompositional) multi-attribute preference models’).

1.2.1 Utility and spatial choice

An important concept in this kind of spatial choice modeling is that of the ‘utility’ or attractiveness of choice alternatives (Samuelson, 1947) as a combinatorial function of the attributes of the choice alternatives (Luce and Tukey, 1964; Anderson, 1974; Louviere *et al.*, 1980). Hendriks (1986; p. 62) characterizes the utility concept as “...a virtual (latent) variable which performs the function of an intermediary between the objective reality of the choice alternatives and spatial behavior” [my translation]. Actors make decisions as to the selection of a choice alternative on the basis of the (expected) utility of the alternatives. In the expressed preference approach,

choice alternatives are rated and ranked on the basis of their utilities which are considered to represent the actor's preference values.

Van der Smagt (1985) and Hendriks (1986) criticize the way utility is defined and measured in the expressed preference approach. Two objections to the way utility is coupled to spatial behavior via expressed preference models seem to dominate their critique. First, they argue that the methodology of preference modeling, in particular in the form of a multi-attribute preference model, suffers from problems which seriously jeopardize the validity of the model results (Hendriks and van der Smagt, 1988). The second objection is that modeling the concept of utility with preferences tends to conceal the real mechanisms and processes underlying choice behavior. Instead, they argue that utility is a concept which links objectives with the means by which they can be realized; therefore it is end-means relationships that the modeling should concentrate on. Each of these objections are discussed below in somewhat more detail.

1.2.1.1 Attributes and their categories

The central component of a (multi-attribute) preference model is formed by the so-called 'combination rule'. As mentioned above, in compositional and decompositional preference models the overall utility of a choice alternative is the result of combining several so-called 'part-worth utilities', generated by various aspects or dimensions of the choice alternative. The combination rule is a mathematical function expressing the way in which these part-worth utilities are combined in the overall utility score. Many combination rules can be specified, each of them representing different assumptions as to how an actor unifies the different aspects of a choice alternative into an overall utility assessment (Timmermans, 1984,1987; van Dinteren and Reitsma, 1985).

An important part of the critique, as formulated by van der Smagt and Hendriks (1988), concentrates on the nature of the process represented by a mathematical combination rule. For example, application of a weighted additive combination rule of the form

$$U_i = \sum_{j=1}^k (W_j * X_{ij}) \quad (1.1)$$

where U_i : utility of alternative i ,
 W_j : weight of dimension j ,

X_{ij} : score of alternative i on dimension j ,
 k : number of dimensions,

assumes a specific and constant way of how part-worth utilities ($W_j * X_{ij}$) are arrived at, and how they are combined into an overall utility score. An important aspect here concerns the way the dimensions are split up in categories. This kind of model deals with variables of which the categories are fixed *a priori*. Categorizations may vary between individual decision makers, but within the context of one individual decision maker and one combination rule, categorizations are fixed. As van der Smagt (1985; pp. 113–115) and van der Smagt and Lucardie (1990) have pointed out, however, there is no good reason for *a priori* assuming that the attributes of choice alternatives can be independently categorized. The point is illustrated by means of an example of how a specific categorization of an attribute becomes necessary as a result of a score assigned to another variable.

	Elevator	Floors
Able	Y	1,2,3,4
	N	1,2,3,4
Disabled	Y	1,2,3,4
	N	1, Others

Figure 1.1: Categorization of the floor-attribute for different combinations of actor and object attributes (Source: van der Smagt, 1985)

The example (Figure 1.1) concerns a hypothetical residential (dwelling) choice problem. The choice objects are apartments which are described by only two attributes and their categories: floor level (one to four) and the availability or lack of an elevator. Two types of actors are assumed: an actor who can use both the stairs and the elevator, and an actor who, for whatever reason (e.g., physical handicap), is entirely dependent on the availability of an elevator to reach levels other than the ground floor. The apartments have to be described in such a way that the description contains those attributes and attribute levels that can be considered relevant in determining the qualities that would make it an acceptable choice. In this case the characteristics of the handicapped actor will force a recategorization of the ‘floor’ variable.

The reason for this is that if an apartment building does not contain an elevator, only two categories of floor level can be meaningfully designated: ground floor and other floors. However, in case the apartment building does offer access to an elevator, all floors become possible alternatives and thus valid categories of the variable 'floor'. For the actor who is not limited by the absence of an elevator, there is no reason to use different categorizations of 'floor' in connection with the availability or absence of an elevator; he can use the stairs. Since the disabled actor cannot walk the stairs, the attributes 'floor' and 'availability elevator' are said to be internally related.

Note that this categorization problem is not caused by the additivity of the combination rule. Of course, additivity can be an erroneous assumption, but even if the different dimensions could substitute for each other, the internal relationship causing the necessity to use different categorizations for different type of actors is retained.

The possibilities of internal relationships between variables are not limited to the inter-dependence of categorizations of variables only. Not only can the categorization of a variable be determined by another variable, but the same also holds for the significance of complete attributes. Especially in compensatory structures, certain attributes only become relevant if certain other attributes are assigned particular values. For a production plant, for instance, accessibility to a regional high-tension line may only become relevant in case local electricity supply is unavailable. This implies that whether or not a variable should be contained in the assessment of utility at all can be the consequence of scores given to other variables.

In short, even though individual variation in how the utility of a choice alternative is generated can be accounted for by means of different algebraic combination rules, these rules cannot comply with internally related variables, simply because they assume *a priori* fixed categorizations of the attributes as well as *a priori* relevance of these attributes. This incapacity to integrate internally related attributes not only causes problems for expressed preference models but for all models that consist of an equation, or a system of equations (van der Smagt, 1985; Hendriks, 1986). They either tend to neglect the large individual variation in how attributes must be categorized or, even if applied on an individual level, fail to represent internally related variables.

In the remainder of this chapter and in the next two chapters a possible solution for representing the interrelations between variable categorizations is put forward. It suffices to say here that taking internally related variables into account requires some form of algorithmic, or logical approach rather

than an equation-oriented one. The elevator–floor example makes clear that internally related variables introduce an element of dependency in the categorization of the attributes. Attributes are to be categorized in one way **IF** a specific situation occurs. If not (**ELSE**), a different categorization should be used.

1.2.1.2 Problems with preference

Although these objections do cause a problem for multi-attribute preference modeling, they do not directly address the validity of the basic principle underlying expressed preference choice modeling, namely the assumption that choice behavior is generated by preferences, plus constraints. This assumption can be combined with the utility-theory principle that choice alternatives can be ranked on the basis of their utility which, in case of multi-attribute preference modeling, is a measure of preference. Together with the assumptions concerning the distribution of the constraint variable, this yields that the alternative with the highest preference value will be selected if it were not for some constraints and external factors that disturb this relationship and turn it into a probabilistic rather than a deterministic one.

Three issues concerning these assumptions appear to be important. The first relates to the nature of the constraints. Modeling them as ‘noise’ may be an undesirable approach, as it tends to conceal many aspects of spatial choice and spatial decision making which cannot be considered ‘noise’ at all, but which, instead, represent very distinct constraint-related mechanisms. This is not to say that in the expressed preference approach constraints are not important. They are considered important, and they are, in a sense, modeled as well. The problem, however, is that they are considered to have their own, independent contribution to generating spatial choice, as is the case with preferences. It is assumed that the variation in behavioral constraints is rather large and unsystematic, and can therefore be regarded as statistical error.

Two other issues which are important in a discussion on the characteristics and usefulness of preference modeling concern the validity of the preference measurements, and the presumed independent status of preferences and constraints. The first issue can be stated as: what exactly is measured in many preference assessments? Preference measurements are typically based on rating and scaling exercises. Sometimes in a crude manner, sometimes very delicately by means of tools such as pair-wise comparisons combined

with multi-dimensional scaling or the repertory grid method; Timmermans *et al.* (1984) and Hendriks (1985) provide an overview of these techniques.

Pawson (1982) argues that ranking and scaling assignments are tasks which respondents are easily prepared to perform. Most of the time the job is fairly easy, so that respondents have no trouble being 'good' respondents. Moreover, since they are requested to rank or scale, they rank or scale, regardless of whether what has to be ranked makes much sense or whether the items to be ranked apply to that respondent at all. Even though a respondent has no car, for example, there is a fair chance that he will be able to assess the quality of the parking facilities of a supermarket. He may even tell the investigator that he thinks it is very important that supermarkets have good parking facilities.

With respect to a multi-dimensional scaling analysis of occupations Pawson (1982; p. 55) puts it like this: "We do not naturally and automatically see occupations as points in multidimensional spaces, though we can do so at will, especially if we are encouraged". Application of this kind of measurement may very well lead to what Blinkert (1978) calls 'Methodische Realitätskonstruktionen', or as Pawson puts it, 'intrusion of the method into the results'; "The technique will always produce results, and so their pre-suppositions, if they ever come under question, are apparently exonerated" (Pawson, 1982; p. 54).

Although part of this problem can perhaps be solved by using a different type of measurement, there is a second problem which is much more serious. This problem is caused by the 'fuzziness' of a concept like 'preference'. What are preference, importance, satisfaction, and so forth, as measured in these procedures? Even if the problems mentioned by Pawson can be overcome, what does one actually measure? It is to be expected that what one gets when asking actors (people, firms, households, etc.) about their preferences, is some sort of composite evaluation index, based on sometimes very complex structures in which objectives and possible means for realizing them play a major role.

It is well known that it takes an effort to separate 'real needs' and 'ultimate desire' from 'reasonable demands' and 'realistic wishes' (Ispen, 1978; Sheppard, 1979). How much of the expressed preferences is 'ulterior aspiration' and how much of it is induced by constraints limiting the possibilities for realizing these aspirations? And even more importantly, how many of these constraints are really random, and how many can be modeled as relationships between objectives, actor's characteristics, and the properties of choice alternatives? A part of these relationships is known by the actors

themselves. They may well know that, because of their income level, they will never be able to obtain an expensive house. Therefore, they may adjust their preference and inform the investigator that they prefer a modest dwelling over an expensive one. What is significant here is not that they want a modest kind of house. What is important is that they want a modest house because they cannot afford an expensive one, something which is not expressed in the preference rating; yet the actor may be totally unaware of another part of the relationship between objective and means. One reason that people and organizations pay consultants for advice is that they do not know their own situation well enough to determine what they want. Nevertheless, the relations still exist, and they might or might not be expressed in the preference ratings.

Different actors have different objectives. Depending on their objectives, their capabilities, and the available means, they develop strategies for realizing their goals. Modeling choice behavior should concentrate on these strategies since it is these which form the link between goals and means, and thus the blueprints for the resultant behavior.

1.2.2 An alternative based on functionality

An approach based on end-means relationships, however, implies a conception of spatial decision making which is different from the one applied in spatial choice modeling based on preferences. The main difference lies in the type of rationality a decision maker can be expected to apply. Unlike a utility-theoretical rationality which is characteristic of the preference approach, analyzing spatial choice on the basis of end-means relationships implies goal-rationality. What is important is not to find the object with the highest preference value, but the object or set of objects that can satisfy a pre-defined goal or objective. This implies the notion of functionality. Objects, be they buildings, people, plans, or areas, can under certain conditions, fulfill a specific function for an actor. Here it is important to recognize that an actor may or may not have various alternative ways or 'strategies', to realize a specific objective. Therefore, many empirically different 'objects' can fulfill one specific function. Moreover, as *The Hitch Hiker's Guide to the Galaxy* (Ursa Minor, quoted by Douglas Adams, 1979; pp. 24–25) points out, one and the same object, therefore one and the same set of empirical properties, can serve many different functions:

“A towel ... is about the most massively useful thing an interstellar hitch hiker can have. Partly it has great practical value—you

can wrap it around you for warmth as you bound across the cold moons of Jaglan Beta; you can lie on it on the brilliant marble-sanded beaches of Santraginus V, inhaling the heady sea vapors; you can sleep under it beneath the stars which shine so redly on the desert world of Kakrafoon; use it to sail a mini raft down the slow heavy river Moth; wet it for use in hand-to-hand combat; wrap it round your head to ward off noxious fumes or avoid the gaze of the Ravenous Bugblatter Beast of Traal (a mindboggingly stupid animal, it assumes that if you can't see it, it can't see you—daft as a brush, but very very ravenous); you can wave your towel in emergencies as a distress signal, and of course dry yourself off with it if it still seems to be clean enough.”

If objectives change, the attributes objects must have in order to be able to satisfy this objective may or may not change. Likewise, if the objective remains the same, a change in the characteristics of the actor may generate different requirements with regard to the objects. For instance, if the objective is something like ‘habitable’, requirements concerning houses will include attributes such as the available spatial arrangement of the rooms and specific plumbing facilities. If, however, the family situation of the actor changes, it is very well possible that the requirements concerning living space change, whereas those associated with the plumbing do not. In other words, whether or not an object can be considered ‘functionally adequate’, depends on how both the characteristics of the actor and the properties of the object match. Objects contained in the set of objects that can, in principle, fulfill a specific function for a specific actor are denoted ‘functionally equivalent’.

Note that what is of central importance here is not that actors have objectives that structure their behavior. The difference between a functional approach based on end-means relationships and one based on preference is that in a functional approach the end-means rationality is explicitly modeled, whereas in a preference approach the objectives are reconstructed from the preference measurements.

1.2.2.1 Taxonomic versus functional classification

The difference between choice modeling based on preference and modeling based on functional equivalence implies a difference in views on the classification of choice alternatives. Van der Smagt (1988) discusses two

different approaches to classification and concept definition; the taxonomic approach and the functional approach. The taxonomic approach seems to be the one underlying expressed preference modeling, whereas the functional approach is advocated here.

In taxonomic classification, the definition of a concept consists of a set of observable object properties that in conjunction characterize an object as an instance of the class. Modern varieties of this method of classifying objects stress the flexibility or ‘fuzziness’ that has to be taken into account in such classifications. The properties cannot always be described unambiguously, and combinations of attribute-value scores are not strictly necessary and sufficient, but should instead be considered ‘characteristic’. Membership in a category is gradual and increases with the number of properties the object has in common with some sort of archetype (see Zadeh *et al.*, 1975; Kickert, 1978; Gupta *et al.*, 1979; Negoita, 1985; and Smithson, 1987, for overviews and possible applications of fuzzy set theory). Such an approach can thus be considered inductive–probabilistic. There is a central archetype and actual objects are classified according to their degree of empirical similarity with that archetype. The more the similarity, the higher the likelihood that the object is an instance of the concept.

Functional classification, on the other hand, is based on functional equivalence. Therefore, empirical similarity is irrelevant for functional classification. Instead, it is the similarity between the properties of an object on the one hand, and the means to satisfy an objective on the other that determines class membership. Since many of the relationships between end and means are of a rather structural nature, in the sense that they limit the freedom of choice, functional classification mainly considers the constraint side of a choice problem. Therefore, modeling based on functional equivalence implies a much more reconstructive, deductive method. It implies furthermore, a different perspective on the significance of preference versus constraints, where their supposed contribution to the explanation of choice behavior is concerned. Where modeling based on preference concentrates on individual preference structures with constraints as an additional ‘noisy’ factor, a reconstructive, relation-oriented approach puts emphasis on modeling the systematics in the objectives, means, and constraints, with the individual variation in ‘pure’ preference as noise.

Van der Smagt (1988; pp. 45–46) also mentions the difficulties associated with constructing functional classification procedures. First of all, the fact that one and the same objective can be fulfilled in many different ways, whereas one and the same object or means can fulfill many different func-

tions, creates the problem of many definitions being possible. Uncovering the relevant end-means relationships from such a complex structure can be a cumbersome endeavor. A second problem is the analysis of objectives. Objectives often are not well-formulated because they are not well-known in the first place. Yet another problem is that no analysis will ever be able to cover all possible means for a given objective. Theoretically, the set of means to satisfy a specific objective is infinite. One therefore needs methods to reduce the set of means to a set that is both plausible and manageable. Finally, there is the problem of how to model functional equivalence.

1.3 A two-stage model of spatial decision making

Explanatory theories of spatial behavior should provide models of how behavioral 'variables' such as attitudes and preferences interact with structural ones such as the constraints limiting spatial behavior and spatial choice. In the expressed preference approach this interaction is considered minimal. Both preferences and constraints have their own, independent contribution to choice behavior. Here, however, it is argued that from a functional point of view, this is not an attractive approach. I would therefore like to suggest an alternative, namely, two-stage modeling of spatial choice behavior. In a first, reconstructive stage, an attempt is made to model those relationships and constraints which, in interaction with specific objectives, determine a feasible set of functionally equivalent choice alternatives.

Specific alternatives can fulfill specific functions for actors. Determination of which functions an alternative can fulfill or whether or not an alternative can, in principle, fulfill such a function, depends on what both the actor and the alternative require, and what both possess in terms of properties. Part of these requirements may belong to the preference structure or the objectives of the actor. Others though, and it is interesting to investigate how different choice processes contain different amounts of preferential freedom, are highly structured in the sense that large parts of the requirements are functionally dependent on the characteristics of the actor. The fact that the disabled actor cannot use the stairs, makes it imperative for him to have either an apartment on any floor in case an elevator is available, or an apartment on the ground floor in the absence of one. Of course, this assumes that dragging oneself up the stairs by the banisters is not a feasible alternative. But even in the unlikely event that the actor does consider it an

opportunity, it can still be modeled as a relation between an objective and the means to satisfy it.

The result of this first stage of reconstruction, is a so-called choice set of functionally equivalent alternatives. The magnitude of this set, relative to the total set of objects or locations the analysis starts with, depends on two things: the number and nature of objectives, and the properties of the objects that must satisfy them. The more objectives are added, the more requirements must be satisfied. How many more, however, depends on the characteristics of the relations associating the objectives with the means to satisfy them. Similar objectives may imply very different sets of requirements depending on the characteristics of the actor. A family with three children may well have requirements for a house which differ from those of a retired couple, even though the general objective, residing in an affordable house, is similar. Even so, many different types of houses may function as a dwelling for the family because often requirements can be fulfilled in various ways. The more objectives are added, however, and the more a choice situation is constrained by limiting factors, the smaller the set of functionally equivalent alternatives will be, the empty set being the smallest possible. The amount of reduction in the original set of choice objects brought about by the reconstruction of the feasible set can then be interpreted as a measure of freedom of choice.

The model covering this first reconstructive stage will be denoted a 'relational model', because it contains a set of declarations on how to reconstruct a set of functionally equivalent choice alternatives on the basis of the existing or possible relations between an actor's objectives and characteristics on the one hand, and the resultant requirements expressed as sets of object properties on the other. Object and actor are said to be 'relationally' linked by the model. The actual classification then, consists of a 'matching operation' by which all objects in an initial set are tested on the set of requirements represented in the relational model. The measurement here is on a nominal, dichotomous level. Alternatives belong to the feasible set or they do not. They can be considered either functionally equivalent or not. No rating or ranking can happen here; the modeling of the first stage implies selection, not ranking.

However, although the alternatives in the reconstructed set can be considered functionally equivalent, it may be that according to one or more additional criteria (*not* objectives), some alternative or set of alternatives is better—has a higher utility—than the others.

A second, optimization stage may be used to figure this out. Of course, in case none (or only one alternative) remain after conducting the first stage, this second modeling step may not be needed, but if the set of functionally equivalent alternatives contains more than one element, optimization may be used to select the best one.

Note that within the framework of a functional approach to choice modeling, such a second stage of optimization can only be applied to the set of functionally equivalent alternatives. Within this set there is a relative freedom of choice. Of course, if new objectives are added, the set may change, but in a given choice situation, the resulting choice set contains alternatives that may be compared and ranked on what may be called 'common dimensions'. These common dimensions cannot be the 'alternatives' empirical attributes contained in the relational model linking objectives with object properties. These have very specific meanings and can be subject to internal relations. Therefore they cannot be considered common dimensions. However, other common dimensions can be recognized and used as optimization criteria. This is the case, for instance, if the alternatives are ranked on the susceptibility to change in all or specific characteristics or objectives of the actor. This kind of criteria can be considered 'meta-criteria' of sorts; criteria that do not refer to any empirical property of the choice object, but instead refer to its position in the end-means structure. Another possibility is that although all objects could, in principle, fulfill a specific function for an actor, they could still be ranked on certain, very general dimensions such as costs, distance, time, and so forth. In Chapter 5, an example of how such an optimization could be integrated with a relational model of site suitability is discussed briefly.

It is important to note that matters of optimization are not considered 'objectives' here. If they were, they must be part of the first, reconstructive stage and should not be considered separately. Therefore, the objectives of the reconstructive stage must not contain criteria such as finding the 'cheapest', the 'closest', the 'most' or the 'least', and so forth. What distinguishes the objectives on which the reconstruction is based is that, unlike an optimization objective, they are based on functionality.

Emphasis in this report is on the first stage of choice modeling based on functional equivalence. This however, requires a method as well as a set of tools by means of which such a functional approach may be applied. This is the objective of this study—to develop a general method for reconstructing spatial choice processes from the point of view of end-means relationships,

and to construct means and tools by which such a method can be implemented and applied in empirical research.

1.4 Functional concepts and relational matching

Van der Smagt (1985) and Hendriks (1986) apply this notion of functionality in a theory on the meaning and definition of theoretical concepts. Examples of such concepts frequently used in geography are ‘service level’, ‘accessibility’, ‘residential utility’, ‘site suitability’, and ‘production milieu’. Concepts such as these typically fulfill a role in theories of spatial behavior. Service level and residential utility may be important concepts in an explanatory theory of residential choice behavior or migration, whereas accessibility and site suitability feature in, for example, industrial location theory. Many of the applications of these concepts are based on either a statistical or mathematical approach to concept definition, or on a purely behavioral-cognitive one. Neither of these methods, however, seems very appropriate for capturing and representing the functional character of the concepts, although functionality does play an important role in modern location theory. At the level on which concepts appear in theories, many of them are defined in an explicitly functional manner. Many attempts to model site suitability or its more aggregate companion ‘production milieu’, for instance, depart from a functional point of view. According to de Smidt (1975; p. 48), for example, production milieu can be defined as “the composite of external conditions, i.e., determinants external to the firm, which influence its siting and functioning.” [my translation]

Such a definition can be denoted a ‘nominal’ or functional definition. It designates the function of an object (here a location) which can be classified as an instance of the concept. Although a functional definition is an attractive point of departure, the next step would be of course converting it into an appropriate measurement procedure. Usually this happens by translating the definition directly into a set of empirical indicators in terms of which the object or location is to be measured. Scores on these indicators are then somehow combined into an overall score on the concept. Strictly speaking there is nothing wrong with that provided the empirical indicators as well as the way they are categorized, measured, and combined, represent the functional character of the theoretical concept. They therefore must represent goal-means relationships. In accordance with the rejection of empirical similarity as a basis for functional classification, Hendriks (1986), for instance,

points out that empirical, observable object attributes can never generate functionality just by themselves. It is only in the context of a specific actor with objectives and requirements that specific properties of objects can contribute to functionality. Again, the difference between the second and third floor can only have an effect on the residential function of an apartment if the actor is not disabled, or in case he is, if an elevator is available. Hendriks (1986; p. 225) puts it like this:

“Crucial in the real (relational) meaning of the concept are therefore not the properties of an object in their own right. Of vital importance is the relationship between the object and the context represented as a set of demands or conditions relating to the object. The real definition of any concept will therefore be a relational definition, as it is to be centered on this relationship between object and context.”

The result of this is that in a relational model, concepts such as site suitability, accessibility, service level, or even utility, are represented as the interactive result of requirements generated by an actor, and the properties characterizing the choice alternatives. In short, the difference between a functional and a relational definition of a concept is that whereas the functional definition only declares the classification to be based on the ability of objects to fulfill specific functions for an actor, the relational definition is a declaration of how this can be achieved.

Investigation of the choice object's properties and the rules linking these properties with the actor's objectives and characteristics are the two main aspects of a relational definition of concepts as recognized by Hendriks (1986; p. 53):

- *the relational aspect*: the assignment of an object to the set of functionally equivalent alternatives is the result of a matching of object properties and actor's objectives and characteristics;
- *the data aspect*: in order to establish whether the matching relation is satisfied, both the actor's objectives and characteristics, and the object's properties need to be known.

In the context of relationally defining concepts, classification therefore comes down to a matching procedure. From the characteristics and objectives of an actor, a number of requirements follow. The 'extension' or 'domain' of the concept then contains all those objects the properties of which match the requirements of the actor.

1.4.1 INUS conditionality and relational definitions

Van der Smagt (1985; p. 36) and Hendriks (1986; p. 116) suggest the modeling of the relationships representing sets of functionally equivalent alternatives by means of first-order logic. This proposal is based on first, a theory of ‘causal factors’ presented by Mackie (1965), offering an attractive conceptual framework for modeling necessary and sufficient conditions (requirements), and second, the notion that predicate logic might offer an appropriate language for representing a relational match (Hendriks, 1986, Chapter 4; Reitsma, 1986). Mackie argues that causal factors must be considered so-called ‘INUS conditions’:

“*A* is an INUS condition of a result *P* if and only if, for some *X* and for some *Y*, (*AX* or *Y*) is a necessary and sufficient condition of *P*, but *A* is not a sufficient condition of *P* and *X* is not a sufficient condition of *P*.” (Mackie, 1965; p. 246).

Denise (1984; p. 49) defines an INUS condition in less formal terms as “an insufficient but necessary part of a condition that is itself unnecessary but sufficient for a result” (refer to Appendix 1 for Denise’s amendment of the formal definition by Mackie).

For example, a smoldering cigarette in a forest might ‘cause’ a fire, but only if a large number of additional conditions, such as the availability of dry fuel, enough airflow to heat up the cigarette’s end, etc., is satisfied. Each of the components in the ‘smoldering-cigarette-plus-additional-conditions’ scenario is a necessary, though insufficient, condition of that particular set. The set itself is one out of many that can cause a fire. It is therefore a sufficient, although unnecessary, condition for a fire.

When applied to the analysis of how an object might fulfill a certain function, it means that no empirical property of an object can by itself be considered a necessary and sufficient condition of such a function. It is only in the context of an actor that object properties can be considered as contributing to such a function. For example, no characteristic of a location or region can, on its own account, generate a suitable environment for locating economic activities. Whether or not such a characteristic can contribute to site suitability depends on the spatial production requirement of the economic activity for which it must be assessed. Moreover, often these locational requirements cannot be expressed in terms of one or only a few characteristics, to be added, multiplied or in any other way combined, without taking internal relationships into account. In most cases, individual

locational properties can be combined in many different ways in order to be able to fulfill the function of a production site. Within each of these combinations, attributes may have to be categorized in different ways. Also, within each of these combinations, the individual locational properties represent necessary but insufficient conditions for a positive result on site suitability. The total combination, however, is one out of many that may generate such a result; a sufficient but unnecessary condition therefore.

1.4.2 Representing relational definitions

Employing the framework of INUS conditionality for modeling the relational definition of concepts implies that for representing a definition for a concept such as site suitability, a strictly logical structure, a disjunction of conjunctions, can be used:

$$R \text{ IF } ((L1 \text{ AND } L2) \text{ OR } (L1 \text{ AND } L3) \text{ OR } (L2 \text{ AND } L4))$$

where R = positive result on a relational matching,
and L1 to L4 = individual locational requirements.

This disjunction of conjunctions forms the basic structure of a relational definition based on functional equivalence. Each of the conjunctions represents a possibility for realizing a specific function (R). Within each of these disjunctive terms, however, all elements are necessary conditions. However, this kind of structure is still inconclusive for application in functional classification. At least two aspects must be added; the relations between sets of conditions and the characteristics of actors, and the possibility that individual locational requirements can be fulfilled in various ways.

Earlier it was argued that in order to understand choice behavior better, the constraint element must be modeled, and that this is only meaningful if relations between the nature of the constraints and the nature of the actor's characteristics are contained in the model. A model of a decision procedure of an individual is a nice thing, but it is much more interesting to be able to model this individual decision procedure as an instance, a special case of a more general model in which the individual is represented by a set of attributes, associated with sets of requirements. In the context of modeling site suitability, this implies that locational production requirements are associated with characteristics of the economic activity; either in the form of production properties or via objectives such as the estimated amount of pro-

duction, whether or not the activity wants to produce for an extra-regional market, and so forth.

These actor characteristics can be incorporated into the logical ground structure of a relational match:

```
R IF ((A1 AND A2 AND ((L1 AND L2) OR
                        (L1 AND L3) OR
                        (L2 AND L4)))
      OR
      (A1 AND A3 AND ((L1 AND L4) OR
                        L5)))
```

where R = positive result on locational match,
 A1 to A3 = actor characteristics,
 and L1 - L5 = locational requirements.

Ideally, the associations (combinations) of actor characteristics and locational requirements have a causal character. This means that the requirements are causally dependent on the characteristics. In the case of economic activities, these relationships are given by the locational requirements associated with characteristics of the production process. Note that what is denoted 'actor characteristics' here (A1 to A3), can be both empirical properties of actors as well as the actor's objectives. In the case of economic activities, both can lead to specific kinds of locational production requirements. Here they are both considered 'factors' or 'characteristics' generating locational requirements, therefore they are treated identically in the model.

The second extension of the ground structure concerns the possibility that locational production requirements may be fulfilled in many different ways. This implies that rather abstract production requirements must themselves be modeled as disjunctions of conjunctions. For example, if 'L1' in the logical structures presented above is supposed to represent something like 'accessibility', this accessibility may be achieved by, for instance, access to the rail network (L6), or the availability of a highway connecting the location with a large market abroad (L7). The logical representation of the relational match can then be rewritten as:

```
R IF (((A1 AND A2 AND (((L6 OR L7) AND L2) OR
                        ((L6 OR L7) AND L3) OR
                        (L2 AND L4)))
      OR
      (A1 AND A3 AND ((L6 OR L7) AND L5)))
```

$$(A1 \text{ AND } A3 \text{ AND } ((L6 \text{ OR } L7) \text{ AND } L4) \text{ OR } L5))$$

where R = positive result on locational match,
 A1 to A3 = actor characteristics,
 and L1 - L7 = locational requirements.

In Chapter 3 this process of specifying requirements on different levels of abstraction is discussed further. It will be argued that it is possible to build systems of hierarchical nested relational structures, each of which can be modeled separately, and can be considered a 'dimension' of the total matching problem.

1.5 Choice behavior and decision making

Reconstruction of functionally equivalent choice sets by means of a relational matching model represents a typical normative–deductive approach, rather than a cognitive–inductive one as applied in preference modeling. It is the researcher who determines the relationships between the actor's objectives or characteristics and the associated requirements. Of course this may imply the use of, among other sources of information, the views and opinions supplied by the actor. However, it is not the actor's cognitive structure underlying his behavior and decisions that is to be reconstructed, but the mechanisms leading to an 'objective' choice set, given a specific description of the actor in terms of goals and/or attributes. 'Objective' here means that the only way in which the actor can change the set is by changing his goal or characteristics, not his views, opinions and attitudes concerning the set.

This emphasis on the more structural relationships governing choice behavior not only implies that modeling functional equivalence is much more concerned with the constraint aspect of choice behavior than with the aspect of preference, it also offers opportunities for supporting spatial decision making.

One of the issues discussed in Chapter 5 is that decision support implies a normative–deductive approach. Not what and why actors perform specific actions is modeled, but what actors could or perhaps even should do given an initial situation. A two-stage modeling process, as suggested here, could offer a useful framework for such a decision support orientation. In the first stage of relational reconstruction, a feasible set of alternatives is deduced on

the basis of a set of choice objects, an actor with objectives and characteristics, and a relational model declaring the relationships between these two. Next, on the basis of common criteria, the remaining alternatives can be ranked and evaluated. Especially if the alternatives imply concrete lines of action to be followed by a decision maker or when the evaluation functions become complicated, the exercise acquires the character of an investigation of alternative strategies and their consequences. One can start looking at, for instance, a spatial economy from many different points of view, each of them formulated as a combination of a specific type of actor and a specific set of evaluation functions. As is discussed in Chapter 5, this is what decision support is about, and such an approach offers opportunities to support spatial decision making and spatial policy formulation.

1.6 Research questions and objectives

Developing a relational model implies the reconstruction of the relationships between actor characteristics and objectives, and the resulting requirements the choice alternative has to fulfill. Since functional equivalence can be defined in a logical manner and does not require empirical similarity, this reconstruction is based on logical rather than data processing (Sutherland, 1988; van der Smagt and Lucardie, 1990). The result of such a reconstruction is a relational definition, represented as a logical structure; a disjunction of conjunctions. The actual matching then consists of processing the relational definition, i.e., finding the requirements for choice alternatives given a set of actor characteristics, and comparing them with the properties of the individual members of a set of possible choice alternatives.

The development of such a model in terms of its formal characteristics, its implementation on a computer, its application to the problem of defining and classifying sites as to their suitability for economic activities, and its integration into a system that can be used for both model development and decision support, are the main problems covered by the research reported on here. The general research objective therefore was to implement these ideas of functional equivalence, internal relationships and relational matching as a modeling and classification system that can be used in empirical applications, both in the fields of theory construction, and model development and decision support.

This general objective can be reformulated as a set of more precise research questions, which are further discussed and elaborated on in the following chapters.

1) What is the role of concepts such as site suitability and production milieu in industrial location theory: What is their explanatory power, how have they been defined and modeled, and what do these definitions look like from a relational point of view?

This question, dealt with in Chapter 2, illustrates what has been discussed here, but this time in the context of economic geography and industrial location theory. As such, it does not suggest any solutions to the general problem just mentioned. Instead, it restates the problem in the context of industrial location behavior and emphasizes the need for an applicable, relational alternative.

2) How can the logical representation of a relational match presented above be incorporated into a model structure that covers both the declarative and procedural aspects of computer-based relational matching?

The logical structures presented above only represent the basic, declarative structure of a relational definition. Empirical applications, however, require a much more versatile kind of model structure in order to handle many different kinds of matching situations. What should the formalism for a relational match look like if it must be processed by a computer? Can a general model structure be developed that supports the division of a complex matching problem into a set of smaller, dimensional problems? Can one come up with a model structure in which one or more forms of generalized abstraction can be represented? These problems are dealt with in Chapter 3. Note that the discussion of these questions barely touches upon the problem of knowledge acquisition, the problem of how to collect and develop a body of empirical knowledge that can indeed be represented in a relational model. Although very important, the development of a knowledge acquisition technique for relational modeling goes beyond the scope of this study. However, a number of formal criteria and procedures that are important for knowledge acquisition are discussed, simply because they are important elements in a discussion of what a relational model is.

3) How can all this be implemented on a computer?

Building and processing a relational matching model and a functional classification procedure requires some special kind of computer programming and computational techniques. Processing a relational matching model contains elements of rule-based computation, knowledge representation, data base searching and traditional computing and calculation. Although a proposal for a general model outline is presented in Chapter 3, the implementation issues are dealt with in Chapter 4.

4) How can the resultant model and automated relational matching procedure be incorporated in a user-friendly, relatively smart system, which provides decision support in problem situations?

Even though an automated relational matching system can be developed, application in a decision support environment generates its own requirements. In a decision support situation, for instance, it is no longer sufficient if the researcher knows how to handle his program, read the output, and understand what happened during a model run. Instead, input and output must be easily understandable and quickly modifiable by the user. Suitable sites must be mapped, and the reasons why areas did not match the activity's requirements must be explained. And what if a relational matching model forms just one out of many models in a decision support system, each of which emphasizes a different aspect of the decision problem? Can they be coupled and can they somehow interact with each other? The problems of computer-based decision support are discussed in Chapter 5, together with a complete version of the relational matching model in a real-life decision support system.

5) What kind of empirical problems can be tackled with such a system?

With the system as it is at the end of Chapter 5, I return to the discussion on site suitability and production milieu in Chapter 2, and present some examples of applications of relational site suitability measurement in Shanxi Province, the People's Republic of China. As explained in Chapter 6, the examples are of a distinct generic character. They do not claim to represent accurate measurements of the site suitability of the counties in the province for specific activities, nor were they intended to do so. Instead, they show the possibilities for much more serious and accurate measurements, as well as modeling attempts for some of the questions and situations that appear frequently in industrial location theory and therefore in the explanation of locational choice behavior of industries.

Chapter 2

Site Suitability in Location Theory

ABSTRACT

In this chapter the need for a relational approach to modeling spatial decision making is illustrated in the context of industrial location theory and site suitability assessment. The starting point is the usefulness of location theory for spatial planning in general and regional policy in particular. Various approaches, such as the (neo)classical approach and the behavioral approach in location theory are discussed briefly. Attention then turns to a modern, more functional approach to location theory. It is argued that some of the procedures by which it has been applied in modeling are inappropriate and hence reduce the applications to mere behavioral ones. The argument is illustrated by several examples from research conducted on the concepts of 'production milieu', 'regional production potential', and 'site suitability'. It is argued that application of the functional approach in location theory requires a relational approach to modeling. Feasibility and matching studies conducted in the sixties and seventies are considered a starting point for such modeling. Some of these studies are critically evaluated from a relational point of view.

Keywords: location theory, production milieu, regional production potential, functional versus relational definitions, feasibility studies, choice-set, matching approach

2.1 Location theory and spatial planning

Spatial planning concerns the monitoring and guidance of the spatial arrangement of activities toward a predetermined, 'better' distribution of these activities. This evaluation aspect renders spatial planning a goal-driven activity of a very distinct instrumental character. The fact that it is goal-driven makes it an activity that seeks for instruments with which a change from a less favorable toward a more favorable configuration of space can be brought about. An activity, therefore, that is characterized by a clear objective-means relationship.

Spatial planning is not an activity pertaining to governments and/or lower-level administrative authorities only. It is just as much a part of the decision-making behavior of individual households and firms as it is of administrative bodies. In case of the former, spatial planning is 'part of life' and constitutes a means to promote one's individual interests; in the case of the latter, it is meant to influence the behavior of these individual decision makers in such a way that the aggregate result of this behavior is similar to the pre-defined goal-state.

With regard to spatial planning by a public authority, Cullen (1986; p. 242) states that "...planning analysis is required to explore the impacts of complex structural relationships upon social processes, the 'how' question must be answered in ways which avoid the unfortunate side effects of the current strategies." This statement could be rephrased in a somewhat more general form as: the better an actor is equipped with knowledge about the system, the better the chances are that something can indeed be changed in the desired direction. Although there is much more to planning than mere knowledge of the 'object system', knowledge of that system may reduce the uncertainty as to what the consequences of a certain line of action will be. Or as Kutter puts it, "Every type of planning is essentially interested in an explanation in terms of causal relations." (Kutter, 1973; p. 74 [my translation]). Of course, it is assumed here that the system can be altered in a desired direction in the first place. In the worst case, accumulating knowledge about a system will result in the recognition that the system cannot be changed at all. But in that case, one at least realizes that it cannot be changed and one can try to find something better to do.

This perspective on 'the rationality of planning and policy formulation'; the perspective of scientific or technical rationality, is certainly a limited one. In modern theories on policy formulation and the rationality of policy-making, different, partly incompatible aspects of the rationality of planning

such as political, economic, and technical or scientific aspects are recognized (e.g., Snellen, 1987). These studies also show that the relationship between 'knowing more' and better policy making is a conditional one, a relationship which can only be realized under specific conditions.

Location theory tries to explain the spatial patterns and processes to be associated with the behavior of economic activities. As such, it can form an important source of knowledge to be applied in spatial planning, both on the level of individual actors and on the level of a public authority. On the level of individual decision makers such as firms, an adequate theory of locational decision making can provide means and instruments for supporting decisions about where to locate the activity. If the theory as well as the model derived from that theory represent the interactions between the activity and the environment appropriately, it can be used as a guideline for the decisions to be taken by the decision makers. This resembles what was characterized as the application and decision support orientation in modeling functionality in Section 1.5. Modeling end-means relationships serves two goals: theory formation and theory-induced decision support.

Of course, spatial planning by economic activities does not stand on its own, i.e., it is not an objective in itself. Instead, it can be considered one particular aspect of the decision-making behavior and strategy determination by firms. Often, actions with a spatial dimension are consequences or 'derivatives' of other, more complex or higher-order strategies. Locational decision making in industry, for instance, is part of a much more complex process of determining an economic production strategy.

For a public authority, location theory might be even more important. The spatial arrangement of economic activities is strongly associated with other variables that have a clear societal concern. Therefore, being able to assess the inter-relationships between the interests and characteristics of economic activities and their environment on the one hand, and the resultant shifts in geographical space as a consequence of specific changes in either these interests, or the environment on the other, can be an important instrument in developing strategies for governmental spatial planning. Regional policy, for instance, has a clear interest in an adequate location theory. Regional policy constitutes a form of governmental spatial planning which concentrates on regional imbalances in economic development, i.e., the uneven distribution of capital, economic activity, and employment over various regions. Generally, regional policy seeks to reduce this imbalance by means of legislation (e.g., location permits) or incentives (tax reduction, investment premiums etc.). Regional policy thus aims to change an existing and undesir-

able situation into a more desirable one. Most important for regional policy, therefore, are possible directions for action and strategic planning. This is a field where the insights derived from location theory could be applied in order to assess the consequences of changes in the environment due to the locational behavior of economic activities.

2.1.1 Behavioral location theory

During its evolution, location theory went through various stages reflecting the more general methodological trends throughout the field of geography (Townroe, 1968; Smith, 1971; Lloyd and Dicken, 1977; Conkling and Yeates, 1976; Carr, 1983; Chapman and Walker, 1987). For a long period of time classical least-cost theory, carried out within the framework of uniform space, profit maximization, and a basically non-decisive, deterministic though fully informed actor, dominated the field. Many of these assumptions were later on considered too simple, unrealistic, and implausible. As an alternative, a more behavioral approach was advocated (Pred, 1967; Krumme, 1969; Townroe, 1969, 1972). According to this approach, the firm or economic activity in general is considered a decision-making actor, organized in a specific manner. Locational decision making is then regarded as the outcome of the confrontation of an organization which has to make choices as to where, when and how to arrange its future activities, within an uncertain environment. This paradigmatic shift in theory and research gave way to a wide range of new topics to concentrate on, which in turn required alternative methodologies, and new techniques of measurements and data-analysis. As a result, a large body of location theoretical research was conducted on subjects like the corporate character of industrial locational decision making (e.g., Krumme, 1969; Chapman, 1974; Hamilton, 1974), organizational versus spatial structure of enterprises (e.g., Erickson, 1972; Watts, 1974, 1978; Pred, 1977), external control (e.g., Erickson, 1974; Firn, 1975; Dicken, 1976; Holland, 1976; Smith, 1979), and regional preferences and attitudes (e.g., Green, 1981; Pellenbarg, 1985; Timmermans, 1986; van Dinteren and Reitsma, 1985).

Generally speaking, the behavioral approach to the explanation of spatial choice and spatial decision making seems to have evolved along a line which consists of a series of 'paradigmatic' stages. After an initial stage in which a fairly straightforward stimulus-response paradigm was followed, one gradually shifted to a more cognitive, psychological approach. It was realized that decision makers not only 'behave' according to certain regularities

and in a stimulus–response framework, but also perceive their own situation on the basis of which they make decisions. Approaches such as preference-modeling and the behavioral studies in geography mentioned above, fit into this approach. However, this purely cognitive approach was again criticized because of its overestimation of the role of cognitive elements such as attitudes and preferences. When applied to the theory of locational decision making, for instance, it was argued that although attitudes and preferences may be important in the explanation of behavior, the individual decision making of firms must be considered within the context of much larger, supra-individual structures governing and directing it (Carr, 1983; Massey, 1975a). An approach, therefore in which behavior-limiting constraints are explicitly included; Carr (1983; p. 396) puts it like this:

“Therefore, it is all very well advocating the behavior of enterprises as the causative element in industrial change, but this does not recognize the permissive or limiting influences of structure and changes in structure independent of behavior.”

Carr refers to a study by Steed (1971) which “appeared to involve a wider range of processes than those recognized in industrial location theory as it stood” (Carr, 1983; p. 395). In this study various external ‘structural’ factors such as changes in ownership and centralization processes throughout an entire industry were investigated. Such processes operate on a much wider scale than that of the individual firm, and partly determine the form and character of the internal decision-making processes. As such they should be taken into account or “merged” with behavioral variables in order to provide better explanations.

2.1.2 Functional location theory

It is interesting to note that Carr refers to earlier work by Massey (1974a, 1974b, 1975a, 1975b, 1979a) who presented a number of critical evaluations of the behavioral approach to location theory. Concerning Massey’s work he states that “...this has not resulted in a concerted movement to define the faults and inadequacies of the prevailing behavioral orthodoxy.” (p. 386). Unfortunately, Carr does not provide the reader with any kind of supporting argumentation for this claim. It is nevertheless interesting to compare Massey’s ideas concerning the behavioral approach in economic geography with those of Carr. Massey does point out what she thinks is wrong with a purely behavioral approach and where much of “behavioral theorizing is

merely reproducing the faults of classical location theory” (Massey, 1975a; p. 85). She makes it clear that:

“What is omitted is the fact that behavior is itself produced—in the case of the firm primarily through the structural interrelationships of the economic system. It is at that level that a theoretical explanation can be constructed which produces the concrete variations in behavior that appear in reality The attempt therefore should be to develop an approach to location which both relates spatial behavior to the development of the economic system, and does not set up ahistorical ideal types, but sees the behavior of groups of economic activities as largely a function of their structural relationships with each other and with the economic system as a whole.”

On the opinion that attitudes should be the main issue to concentrate on in studying decision making behavior (Svart, 1974) notes:

“Forms of behavior can never be taken, theoretically, as given; they are always ‘produced’, that is they are the outcome of the structured context in which they occur.... The argument is a general conceptual one: it should be applied equally to mental processes.”

and

“Instead of delving deeper into the individual’s psyche in a search for something which is immutable, geographical research should be investigating the systemic structure which produces those resultant forms of behavior.” (Massey, 1975b; p. 202)

What Massey seems to advocate here is a kind of location theory in which the relations between the individual decision maker and the framework of supra-individual processes in which this decision maker ‘behaves’, must be integrated into location theory. Basically, this seems compatible with the opinion held by Carr.

It seems that all this still fits into the framework of a more constraint-oriented approach to spatial choice, mentioned above. Massey, however, goes beyond this position and introduces yet another perspective on spatial decision making, in an article on the meaning of the concept of ‘regional inequality’ (Massey, 1979b). In this article she states that regional inequality is to be considered as:

“...inequality in the degree of attractiveness to, and suitability for, economic activity. At any point in time, in other words, there is an uneven geographical distribution of the conditions necessary for profitable, and competitive, production.” (Massey, 1979b; p. 234)

Massey then goes on to point out that different forms of economic activities respond differently to the same kind of regional inequality:

“This manner of response to geographical unevenness will vary both between sectors and, for any given sector, with changing conditions of production.”

According to Massey, therefore, regional problems are not simply ‘regional’. They are the result of how economic activities respond to uneven distributions of production factors, and as such, manifest themselves in regional inequalities in e.g., employment and economic growth.

This notion has some important implications, not in the least for location theory and the development of location theoretical models. First, the point Massey makes here does indeed go beyond the earlier criticism of the behavioral approach in location theory. Not only must macro-structures (available technology, labor division, spatial patterns of production, etc.), be included in theories of locational decision making, but it is the functional relationships between an economic activity and its (regional) environment that location theory should concentrate on. Moreover, the idea of a functional description of spatial distributions, spatial (in)equality and locational suitability, seems compatible with a relational point of view. Massey’s conception of regional inequality implies that modeling locational utility and site suitability, concepts that refer to the functionality of locations for locating their activities, cannot be based on empirical similarities only. It is the interaction between these empirical characteristics and the strategies, interests and properties of the activity that determine the function the region can fulfill for the activity. The same empirical distribution of spatial variables should be considered an uneven and variable distribution for activities for which the production requirements maintain different relationships with these variables.

This more functional orientation might be considered a third paradigmatic shift toward spatial choice in general, and locational decision making in particular. A functional approach not only recognizes the importance of structural, supra-individual constraints on behavior, it also implies a rather

actor-specific line of thought, with special attention paid to the interactions between actor and environment. Earlier it was argued that such an interaction-oriented perspective implies important consequences for how explanatory concepts should be defined and measured.

2.1.3 Functional location theory: implications for model building

The shift from a mere behavioral to a more constraint-oriented, and more functional location theory, was not accompanied by many alternative approaches in model building. Many of the methods and techniques used in the behavioral tradition were retained. And although this does not detract anything from the plausibility of functional location theory and its potential for regional policy, functional location theory can only contribute to the success of regional policy if it is accompanied by model-building procedures which are able to incorporate its most important characteristics. Model-building procedures therefore, that can handle the interactions of a firm's spatial production requirements on the one hand, and the locational characteristics, the uneven distributions of production factors, on the other.

To illustrate the problems associated with a functional approach to location theory which is not accompanied by functional (relational) modeling, I would like to discuss some approaches to modeling site suitability, which depart from a functional point of view, but which do not incorporate this functionality in the modeling procedure. Instead, the modeling takes a typically behavioral form; measurements are based on similarities in empirical properties, and the significance or importance of these properties is measured as expressed preference or attitudes. The discussion concentrates around two concepts which have been of some importance in the economic geographical analysis of site suitability: 'production milieu' and 'regional (indigenous) potential'. Both concepts are products of a more functionally-oriented approach in location theory. Although it cannot be claimed that the applications discussed here aimed to apply the ideas as formulated by Massey, the concept definitions they depart from are of a typical functional character.

The main point these examples are meant to illustrate is that a functional approach in location theory requires that end-means relationships are integrated into the model. In the previous chapter it was argued that this is only possible if functional concept definitions on the theoretical level are replaced by relational ones associating objectives with means. The examples

will show that translating functional definitions directly into a measurement procedure, and thus omitting the relational definition of the concept, will prevent this integration of functional dependencies, and therefore the integration of the merits of the theory into the models.

2.2 The concept of production milieu: definition and measurement

Site suitability studies were conducted throughout the history of economic geography. The concept was defined and measured in accordance with the predominant methodological paradigm. In (neo)classical location theory, site suitability is implicit. Classical and neo-classical location theory are typically normative and site suitability is simply equivalent to the calculated zones and areas of production (e.g., Lösch, 1954; Greenhut, 1956; Isard, 1956). A normative approach to site suitability, however, is not something that is limited to classical location theory. In an overview of methods for defining and assessing site suitability for various forms of land-use, Hopkins (1977) discusses various techniques and methods for site suitability assessment. Many of them are of a typically normative character. As discussed in Chapter 1, a relational reconstruction of concepts such as site suitability can be considered normative as well. In more behavioral approaches, a much more inductive kind of analysis is followed, based on empirical similarities and statistical associations. Particular attention is given to how entrepreneurs evaluate the regional or local characteristics of places.

In a review of the literature on 'regional economic potential' Roelofs and Wever (1985) mention a number of approaches for the definition and measurement of what is sometimes called 'regional production milieu'. Similarly, Hendriks *et al.* (1984) present a number of evaluating reviews on the definition and measurement of this concept. Production milieu has been the object of many Dutch studies, undertaken primarily with the objective of generating a set of tools for use in the acquisition of economic activities, both at the national and regional level. Vonk and Willems (1972; p. 15) define production milieu as:

“...the composite of locational circumstances external to the firm or establishments, which directly or indirectly influences the locational and economic activities at that location or in that region.”
[my translation]

According to de Smidt (1975; p. 48) production milieu can be defined as:

“...the composite of external conditions i.e., determinants external to the firm, which influence its siting and functioning.” [my translation]

These definitions are of a distinct functional nature. They refer to the siting function a location can or cannot fulfill for an economic activity. This in itself does not generate any problem, if, and only if, one realizes that such a definition needs a relational specification before it can be meaningfully applied in measurement. In the form of a functional definition it only represents a set of possible (relational) definitions, each of them relating the ‘external circumstances’ to the spatial production requirements of a specific kind of economic activity. However, attempts at making a functional definition of production milieu operational by means of a direct translation of the concept into sets of empirical indicators are certain to yield questionable results. This can be illustrated by examining two empirical applications of the concept: production milieu measured by means of the R.E.B. methodology and by means of a production milieu matrix. Both methods have been critically evaluated by van Blokland and Roelofs (1984). Much of what follows in the next two paragraphs is based on their investigations.

2.2.1 The R.E.B. approach toward production milieu

The R.E.B. (Regionaal Economische Beleidsadvisering—regional economic policy advice) investigations were developed by collaboration between Chambers of Commerce (Samenwerkende Kamers van Koophandel) in the Netherlands (R.E.B. coördinatie, 1979; R.E.B. coördinatie, 1980). The objective of the research was to gain insight into the economic development of a region by a careful monitoring of its economic activities and the regional characteristics that can be considered relevant for those activities. Part of the investigations dealt with an assessment of the quality of the regional production milieu. The definition of production milieu applied is very similar to the ones mentioned above, “The composite of regional circumstances external to the individual firm which are important for the existence and operation of that firm in a region” (R.E.B. coördinatie, 1979 [my translation]). Two approaches: an objective and subjective one, are followed. The former consists of an as-complete-as possible description of those location factors which are considered important indicators of the quality of the regional production

milieu. This is achieved by the construction of a list of 11 categories of 'objective' location factors, each of which is subdivided into a set of empirical indicators of which the measurement is considered unproblematic. Comparison of various regions is then a means to say something about the differences in the quality of the regional production milieu, while comparison of the configurations of indicators for one region over various time periods should give an indication of the intra-regional production milieu development.

Van Blokland and Roelofs mention a number of objections to this objective part of the R.E.B. method for measuring production milieu. They stress the problem of the determination of the relevant production factors. The main objection then is that empirical characteristics of a region can only be considered indicators for production milieu, hence location factors, if the functionality of these characteristics for the site suitability for a specific activity can be made plausible. An inventory of spatial variables can be very useful for constructing a locational data base by means of which an area is empirically described. But such an unrelated collection of factors and their regional scores does not by itself yield any information about production milieu. Actually, it can be considered an extended version of the advertisement shown in the introduction, presenting the production milieu of Bavaria as a set of regional properties. Application of functional location theory requires that the relations between locational characteristics and the properties and production requirements of economic activities are made explicit. Enumeration of a range of empirical properties that might, under favorable conditions, act as spatial production factors, does not suffice to achieve this. Put in terms of relational modeling; not a single locational property is by itself capable of generating site suitability. There is nothing inherent in a locational property that makes it into a necessary condition under which the location can serve as a site for production. It is only in an interactive co-production relation of location and activity that locational properties can be considered spatial production factors.

Concerning the 'unproblematic' character of the measurement of the indicators, Roelofs and van Blokland mention the problem that some of the indicators such as the variable 'availability of skilled labor' are rather abstract. Clearly, what 'skilled' is will depend on the type of labor required by the activity. Therefore, the availability of skilled labor can only be defined by referring to the activity the measurement is meant for. Any direct translation of this theoretical variable into one or more empirical indicators, without reference to the functionality of the indicator for specific activities, must be rejected.

In the subjective part of the R.E.B. method, one centers on the evaluation by the entrepreneurs of the factors/indicators considered in the objective phase, both in terms of the importance those factors have for their firm, and in terms of their satisfaction with the regional situation regarding those factors. These evaluations are collected by means of interviews and surveys consisting of closed questions with fixed evaluation categories. The satisfaction scores are then amalgamated by counting the number of favorable/unfavorable responses and sometimes by computing averages. Several objections can be raised. Van Blokland and Roelofs object that an evaluation of locational properties can only shed light on the evaluation of production milieu if the relations between various factors are made explicit in terms of the spatial production requirements of the firm. For example, an evaluation of available railway connections in terms of satisfaction is not independent of the function these connections have for the activity for which they are evaluated. If a location does not possess enough coal to feed a coal-based chemical plant with, it can still serve as a site for such a plant if the coal can be transported from elsewhere, for instance, by means of rail. This, however, implies that rail access only becomes a location factor if local coal resources are lacking. Similarly, if coal is not the main raw material, then the availability of both local coal and rail access for transporting become irrelevant. A satisfaction score devoid of any reasons for how it came about does not contain much information about production milieu.

Additional objections can be raised against the applied scoring system and the way scores are combined to form an overall measure of the quality of the production milieu. The lack of a response category 'unimportant' or 'neutral' in certain circumstances, forces a respondent to evaluate factors as either negative or positive. Furthermore, what would be the implications in case a factor would indeed be listed as unimportant? In case no information is provided about why the factor is considered unimportant, or about which characteristics of the firm render it unimportant, not very much is gained. And since relevance or irrelevance cannot always be assessed independently of the availability of other factors, they can be considered not very meaningful.

These and related problems become worse as a result of the procedure that is applied for aggregating the results from the evaluation. This is done by computing the percentage of firms that consider a factor (un)important, or by computing an average score on the importance ratings. The resulting numbers are then interpreted as a measure of the quality of the various production factors. It will be clear that this 'averaging' of importance ratings

is something that should be avoided. Different firms can have very different reasons for considering production factors (un)important. These reasons are associated with the characteristics of the production and should instead be the nexus of the model rather than be ‘averaged away’ in an aggregation procedure.

Another problem associated with ‘importance’ measurements concerns a relation between the degree to which a production factor causes the entrepreneur problems, and the likelihood that it is recognized as important. Maybe production factors can be very important, but because of the fact that they are satisfied by the current location, they are never realized. Only on occasions where such factors are not satisfied anymore do they suddenly become manifest, and are only then recognized as ‘important’.

With regard to R.E.B. methodology, it can be concluded that the conception of production milieu it departs from, at least in its functional form, indicates a functional approach. When modeled by means of R.E.B. methodology, however, this functional approach is lost in the applied measurement procedures. The attempt to directly ‘translate’ a functional definition of a theoretical concept like ‘production milieu’ into sets of empirical indicators and value judgments, necessarily fails to incorporate the conditional nature of factors. A functional definition pertains to the function an object has to fulfill, not to how such a function can be fulfilled. Treating empirical characteristics of the locations as indicators of these functions without making explicit the conditions under which this can actually occur, renders the measurement invalid.

2.2.2 The production milieu matrix method

Another method for assessing the quality of production milieu is measurement by means of a production milieu matrix (New Town Guide, 1978; van Oudheusden *et al.*, 1981; Stijnenbosch *et al.*, 1983). Again, one departs from a typically functional point of view. Not all locations are suitable for locating every different kind of economic activity. “Attractive production milieu, therefore, means that at a location there is a favorable combination of conditions in order to attract and develop specific types of production.” (Stijnenbosch *et al.*, 1983; p. 26 [my translation]).

The method starts with the determination of a number of rather abstract, theoretical location factors representing the different aspects of production milieu, e.g., labor market, physical space, local authority policy, service structure, etc. These are then broken up into a fairly large number

of empirical indicators, the measurement of which can be considered rather unproblematic (percentage unemployment, percentage commuters, ground prices, distance to airport, number and types of schools, etc.). The first step then consists of an inventory of the locations included in the measurement and their scores on these indicators. In order to make scores on different indicators comparable, they are expressed on a three-point scale, the actual scores of which are determined by the magnitude of the deviations of the location specific scores from the mean score.

In a second step, recently relocated firms are requested to express their opinion about the indicators. First, the relative importance of the indicators is measured (five-point scales). Next, for each of the main factors mentioned earlier, importance assessments are collected by asking the respondent to rank the factors in order of importance. These ordinal scale values are then converted to numerical scores, which are then aggregated by taking the average scores of the firms contained in various industrial sectors.

In a third step, the results from the previous stages are confronted with each other in a matrix. Each of the empirical scores of an indicator for a specific factor is multiplied by the importance score of that factor. In order to correct for the problem that not every factor is represented by the same number of indicators, the resultant scores are weighted relative to a standard of five indicators per factor. This procedure is conducted for each industrial sector separately, because it was realized that different types of industry evaluate production factors differently. The resultant weighting factors are then multiplied by the ordinal scores on the indicators for each of the locations included in the measurement. Summation over the indicator scores then results in a measure of the quality of the industry-specific production milieu at that specific location or area.

Compared with R.E.B. methodology, measuring production milieu by means of this production milieu matrix technique comprises some significant improvements. First of all, the quality of a production milieu is not measured by means of querying entrepreneurs on what they think of it, but by means of an analytic procedure by which empirical measurements and importance assessments are used to arrive at a score for the quality of the production milieu. Apart from the fact that this gives the method a formal, much more model-like basis, it also enhances the clarity of the measurement and makes results easier to interpret and compare. Another important advantage when compared with R.E.B. methodology involves the industry- or production-specific measurements of production milieu. Therefore, the production milieu matrix method constitutes a valuable attempt to integrate

this activity-specific aspect of functional location theory into the modeling procedure.

In spite of these improvements, however, the procedure contains a number of problems of a more conceptual nature, which seriously jeopardize the validity of its results. Van Oudheusden *et al.* (1982) themselves observe a weak correlation between the evaluations of the abstract production factors by the entrepreneurs and those of the 'objective' indicators. They suggest that their (the researcher's) choice of indicators might not correspond to what the entrepreneur thinks about when he performs the scoring on the factors. This is indeed very likely. As in the case of R.E.B. methodology, the indicators of theoretical production factors are expressed solely in terms of limited sets of empirical properties of the various locations. Whether an empirical property of the location can act as a location factor depends on the spatial production requirements of the firm. Note that this objection is not inconsistent with the advantage of using the formal evaluation procedure mentioned earlier. The use of a formal procedure itself can be considered an advantage. Whether or not the information that is used by the procedure is adequate is another matter.

Another important problem concerns the way importance weights are considered, and the role they play in the aggregation of the individual results. Concerning the importance assessments themselves, the same objections as raised against the R.E.B. methods apply here. What should be modeled is the relationships between activity or production-specific properties, and the importance of locational characteristics. This importance is dependent on production requirements and possibly on other locational characteristics. Importance of locational properties can only be assessed in the context of a conjunctive/disjunctive set as discussed in Chapter 1. Importance weights devoid of such context are rather pointless.

One might want to argue, however, that within the context of one specific actor, importance weights are meaningful and can be used to assess the quality of the production milieu. Although this is true, this would constitute a clear example of what can be characterized as 'context-specific measurement' (van der Smagt, 1985; pp. 53–57; Hendriks, 1986; pp. 32–35). This occurs when the context (variable requirements for various types of activities) is included in the model, but only in the part which considers the measurement. Not, however, in the structural part of the model. In context-specific measurement, different scoring systems and different variables are used to measure the same theoretical variable for different types of actors. The next step is then the inclusion of this theoretical variable in a structural model

which may contain other theoretical variables as well. According to van der Smagt (1985; pp. 53–57) and Hendriks (1986; pp. 32–35), however, context-specific measurement is not unproblematic. Although variation in context can be taken into account, within one specific context, object attributes are considered intrinsic characteristics or indicators of the theoretical concept. The problem then is that although context-specific measurement may yield valid results for that one, specific context, if something in either the characteristics of the actor or the context changes, the results may not be valid anymore, and a new measurement is needed.

A similar objection can be raised against how sector-specific results are aggregated. Aggregation should be carried out on the basis of functional generalization, not on the basis of an *a priori* fixed typology of actor types. It is very likely that various locational characteristics play similar roles for various firms across different sectors, while they play different roles for firms within the same sector.

Yet another problem connected with the way the importance weights are derived and interpreted has to do with the manner in which these (weighted) indicators are combined to form an overall score on production milieu. The scores are combined additively. This assumes an independent contribution of individual production factors to the value of production milieu; a fully compensatory structure, therefore. Both assumptions are unrealistic. It is very well possible that the significance of a production factor for an activity is dependent on the availability of another production factor. This would, for instance, occur if the availability of a railway connection becomes important only in case adequate road connections are lacking. Similarly, in case the electricity production in a certain area is evaluated as being insufficient, access to a national high-tension network can become important. The importance of a production factor is therefore not always independently assessable. Concerning the implicit assumption of a compensatory structure, an objection could be that it is likely that the lack of certain spatial production requirements cannot be compensated for by other factors, or that only specific factors can compensate for the lack of other factors. In the production milieu matrix method, however, low scores on one indicator can always be compensated by high scores on another indicator.

2.2.3 R.E.B. methodology and production milieu matrix: conclusions

It can be asserted that the major objections to the production milieu matrix method are very similar to the ones made against R.E.B. methodology. Of course there are some technical problems such as multiplication of ordinal scale values, the applied weighting schemes, etc. Other problems concern the applied techniques for measuring importance weights and attitude values (refer to Pawson's (1982) critical evaluation of the use of scaling and ranking techniques mentioned in Section 1.2.1.2). More important, however, are problems of a more conceptual nature.

Functional location theory emphasizes the importance of the relations among spatial characteristics on the one hand, and the nature and objectives of the economic actors that act upon that space, on the other. Production milieu matrix and R.E.B. methodology, however, model production milieu by means of an *a priori* fixed set of independent and compensatory indicators. This inability to proceed beyond a mere behavioral approach toward locational decision making is inherent in the direct translation of functional definitions into empirical terms. Functional definitions are important: they preclude confusion by ascertaining the function of an object as an instance of the concept. A functional concept definition, however, is only the starting point in the process of modeling a theoretical concept. The ways in which a function represented in a functional definition can be fulfilled must be incorporated in the model of the concept. R.E.B. methodology and production milieu matrix method, therefore, seem to put the cart before the horse when departing from a functional definition whilst applying model-building techniques based on empirical similarity.

2.3 The concept of regional indigenous potential

Another concept frequently used in location theoretical research is that of regional (indigenous) potential. It has been defined in several ways (Strassert, 1984; Roelofs and Wever, 1985). Roelofs and Wever define regional potential as "the configuration of regional and/or firm characteristics contributing to the region's comparative advantage over other regions concerning the generation of a specific type of economic growth." [p. 21; my translation]

It is obvious that defined like this, regional potential and production milieu have a lot in common. Regional economic activity and economic growth are often strongly associated with the presence of business establishments

(for an elaborate analysis of the relations between the spatial distribution of economic growth and that of the spatial distribution of business corporations refer to Holland (1976) and Pred (1977)). Therefore the type, as well as the amount, of possible economic growth that can be established in a region is dependent on the production milieu of that region. The main difference between the two concepts stems mainly from the fact that regional potential refers to the potential, the capability of a region to generate economic activity, whereas production milieu concentrates much more on site suitability proper.

In their review of the literature on regional potential, Roelofs and Wever mention a large variety in the approaches to the analysis and measurement of regional potential, ranging from classical ones such as growth pole theory and cumulative causation theory, to attempts to assess regional potential solely by means of production milieu. The former are then categorized as belonging to what is called a 'direct' approach, whereas the latter is an example of an 'indirect' approach. Direct approaches aim to formulate a model for regional potential, as well as procedures to measure it. Indirect methods try to evaluate regional potential by means of modeling and measuring concepts that are expected to be strongly correlated with regional potential.

Measuring regional potential by means of the marginal capital return rates or marginal factor productivities (van de Vooren, 1980; van de Vooren and Wagenaar, 1984) is an example of a direct approach based on (neo)classical economic theory. In this type of analysis one concentrates on investment return rates and employment growth. Those regions showing the highest productivity of a set of production factors—the highest regional potential—are the most promising regions for locating new investments. Computing this potential for different business sectors separately highlights those sectors in a region that constitute the most promising targets for policy incentives.

Van de Vooren and Wagenaar (p. 802) themselves state that traditionally regional policy is directed toward equity, i.e., aimed at reducing regional differences in wealth and economic development. Consequently, a policy based on equity will not automatically result in a distribution of production factors that yields maximum results on a national scale. This constitutes the problem of inter-regional equity versus aggregate efficiency. The approach by van de Vooren and Wagenaar fits into an efficiency framework. In itself this seems reasonable. One of the reasons the concept of regional potential was 'invented' in the first place, was that in order to achieve the maximum likelihood that regional policy incentives are successful, they should be directed toward those elements that comprise the 'strong' aspects of the region. Many

regional incentives in the sixties and early seventies, at least in the Netherlands, were considered too general. Instead, it was suggested that focus be on those sectors of the regional economy that had certain advantages when compared with the same elements in other regions. It would, for instance, then be possible to select certain areas for regional policy incentives, thereby focusing actual incentives on those sectors that show the highest potential for that region: clearly an efficiency-based strategy.

There are, however, some disadvantages connected with this method of assessing regional potential. Roelofs and Wever mention a number of objections to the validity of the assumptions implicit in this neoclassical approach (unconstrained mobility of production factors, only labor and capital are included in the measurements, etc.). These assumptions render the measurements questionable indicators of promising targets for regional policy.

From a purely methodological point of view, the sector-specific measurements reported in van de Vooren and Wagenaar (1984; p. 803) constitute, at best, a context-specific measurement. For different types of actors separate measurements are conducted. This is because it is realized that for each of these classes the context—here the distribution of production factors—must be evaluated differently. The sector-specific marginal capital return rate is a variable which can be associated with a region and used in a model of regional potential. This implies abstraction from the processes generating this return rate. Should this be unproblematic in this one context, there is nothing in the model that safeguards the validity of a similar measurement in case changes in either the context or the requirements of the actor occur. In case context-specific measurement is nevertheless preferred by the researcher, he should proceed by using a proper procedure for selecting contexts and actor-types. It is simply insufficient to distinguish between different types on a purely *a priori* basis. According to the table of measurement results (p. 803), van de Vooren and Wagenaar distinguish different types of actors by adopting the Dutch standard enterprise classification scheme (S.B.I). It is very unlikely, though, that this classification corresponds accurately with a classification developed on the basis of differences in spatial production requirements.

Other problems are associated with the explanatory part of the approach. Van de Vooren (1980) 'explains' the differences in marginal capital return rates by means of variance in regional characteristics, applying a multiple regression model. Following a relational point of view, however, statistical associations thus derived cannot be interpreted causally and are therefore unfit for explanation. Especially in the case of an instrumental problem such as

regional policy, knowledge about causal relationships between regional characteristics and production-specific properties is of vital importance. Again, the results from a statistical estimation apply to average patterns of association. But in case of a population of subjects that can be expected to show profound differentiation in how certain elements in a spatial environment can perform certain functions, statistical associations are no good. A statistical model does not permit re-categorization of variables as a function of the value other variables take, and it does not provide the means with which to handle non-compensatory relations.

Finally, Wever and Roelofs also mention the neoclassical assumption of profit maximization as an objection to be raised against the approach by van de Vooren and Wagenaar. This objection is a typical behavioral one pertaining to the characteristics of the decision behavior of the economic activity. However, in a reconstructive-normative approach such as neoclassical location theory, or a relational one, this assumption is perfectly acceptable. Of course, differences in how this maximization must be defined and modeled can exist, however, since normative reconstruction does not aim to model the decision behavior itself, but is directed at the structural situation in which this decision is to be set, the question whether or not decision makers are optimizers is not really relevant.

2.4 Regional potential and production milieu: a closer look

It is interesting to take a closer look at the definition of the concepts of regional potential and production milieu, as well as at their mutual relations. Both concepts are used and dealt with in the literature, but it appears difficult to make a clear distinction between them. Roelofs and Wever, for example, review both R.E.B. methodology and the production milieu matrix method as instances of the indirect approach in assessing regional potential. A similar type of confusion is exhibited in a study by Meyer-Krahmer (1985) on regional innovation potential. In this study Meyer-Krahmer investigates factors and processes influencing the innovation behavior of firms with regard to their location. Of course, aggregated to the regional level these innovations together constitute some sort of regional score on innovation performance. The study itself, however, is about *production milieu*. Statements from the analysis illustrate this (p. 531):

“According to their own indications, their geographic locations and those of the know-how vendors almost without exception do not play a role.”

and

“Unlike firms with pronounced outward orientedness, these firms regard their geographic location as an important factor, especially with respect to polytechnic schools, universities, and research establishments.”

As part of an economic analysis of the concept of regional potential, Günter Strassert (1984; p. 25) has a serious complaint regarding the value of the concept: “Is a theoretical advance taking shape here? I think: No.” [my translation]. He arrives at this verdict after examining a series of (German) definitions and applications of the concept of regional (indigenous) potential (Giersch, 1963; Biehl *et al.*, 1974; Thoss, 1977, 1983; Spehl *et al.*, 1981). Rearranging thoughts and ideas concerning production milieu, regional potential and underlying location theoretical issues, Strassert makes a distinction between “resource potential” (Das Potential einer Resource) and “capacity” (Kapazität). Resource potential is then defined as “the opportunities to technically apply them, i.e., one is able to point out for which types of utilization a resource is appropriate.” [p. 20; my translation]. Capacity, on the other hand, pertains to the maximum turnout a resource can generate once it is actually exploited (p. 21). Assessing the capacity, however, presupposes a decision concerning the allocation of resource potential. This decision, in turn, demands answers on questions about the type of utilization to be applied, and the technology by which that application is implemented. Strassert (p. 26) then argues that regional potential pertains to capacity rather than resource potential:

“This makes clear that ‘regional potential’ indicates production capacity, albeit only the effective or available capacity, which one can achieve by eliminating the existing impediments.” [my translation]

Strassert therefore concludes that the concept of regional potential has no additional value when compared with the more traditional capacity concept.

When put in terms of the preceding discussion concerning attempts at modeling regional potential and production milieu, it can be argued

that Strassert's resource potential can be associated with production milieu, whereas regional potential refers to the attainable economic growth, once decisions on the allocation and utilization of the prevailing production milieu have been taken.

Production milieu pertains to the possible ways of utilizing the region's resources by activities, the potential then refers to the (estimated) results of actually utilizing some or all of these opportunities. Regional potential as such is always a derived measure. It is a measure of the possible effects in terms of economic growth, given that certain economic activities are maintained or initiated. But whether the latter is feasible or not is determined by the quality of the production milieu which refers to the individual relations between an activity and its regional environment. Regional potential refers to the overall results, calculated on a regional scale. This, however, makes production milieu the principle concept to concentrate on. The more so, since increase or change of the regional potential by means of, for instance, regional policy incentives, always has to be achieved by changes in the region's production milieu. Altering the properties of the region can have consequences on its suitability for locating certain types of economic activities.

2.4.1 Regional potential and production milieu: some conclusions

The discussion so far concentrated on the relation between production milieu and regional potential on the level of functional definitions. This applies to attempts to model production milieu by means of either R.E.B. methodology or production milieu matrix, as well as to the views concerning regional potential by Strassert. His description of resource potential is very similar to the functional definitions of production milieu mentioned earlier. Central to the critique on these attempts at modeling is the notion that functional definitions cannot serve as a model for the concept. Functional definitions serve an important, but limited purpose. They designate the function an object has to perform for an actor, but they do not contain information on the (necessary and sufficient) conditions that need to be satisfied in order to realize this function. They can, furthermore, be regarded as statements that say that in order to operationalize such a definition, the functional interactions between actor and environment must be explicitly declared. Translating a functional definition directly into a set of empirical indicators and a formal procedure to combine them into a score, however, fails to incorporate the

relational nature of concepts and the views of functional location theory. As a consequence, the validity of the results is highly questionable.

What is required, therefore, is an alternative framework of analysis and model building that makes it possible to model production milieu in accordance with the postulates of functional location theory and that incorporates this relation between regional potential and production milieu. So far, however, no real proposal for such an alternative has been put forward. In the remainder of this chapter, therefore, an alternative approach is discussed. The approach, which is based on matching production requirements with locational properties goes in a direction which contains some valuable elements that fit a relational modeling approach.

2.5 Production milieu as a choice set

As with spatial decision making in general, locational decision making can be modeled as a two-stage procedure. First, a choice-set is reconstructed. This choice-set constitutes a set of choice alternatives which are considered functionally equivalent, i.e., all the elements of the choice-set satisfy the necessary conditions for fulfilling a certain function for a certain actor. Modeling the actual choice, then, belongs to the next stage which can include preferences as well as optimization procedures to determine which choice alternative should or will be selected. Of course, in case the modeling needs to generate advice about which alternative is to be preferred, the emphasis will be on optimization. If, on the other hand, the model is supposed to represent the actual behavior of subjects, then preferences and attitudes might be emphasized.

Regarding production milieu as an activity-specific choice set is something which, as an idea, is not new. It is, for example, present in "The Three Principles of Industrial Location" formulated by Rawstron in 1958. It was furthermore applied in some 'feasibility studies', carried out mainly at the end of the sixties and the early seventies (e.g., Schilling, 1968; Pellenbarg *et al.*, 1974). Indications of the relevance of matching can also be found in site suitability studies for locating energy generation facilities (Hobbs, 1980; Church and Bell, 1981; Calzonetti and Eckert, 1981) but just as well in a book on "the art of arranging buildings and other structures on the land in harmony with each other" (Lynch, 1962; preface). And although it has hardly ever been formalized in an actual modeling procedure, a similar notion can also be found in many of the introductory sections of modern studies on

locational decision making (e.g., de Pater, 1980; Mason, 1985; Perry, 1985; Wilder, 1985).

Every single one of these studies is characterized by its own approach, but although the choice-set idea is formulated in many different ways and sometimes pretty much concealed or hidden in a footnote (Church and Bell, 1981) there is this common notion about a set of locations from which, in principle, a choice can be made.

2.5.1 Rawstron's three restriction principles

In his book on industrial location David M. Smith (1971) touches on the ideas of Rawstron (1958) who put forward a conceptual model of locational choice by economic activities. Rawstron represents locational choice as a process of narrowing down the set of feasible alternatives as a result of a number of 'restrictions' or spatial production requirements set by a firm. Three principle restrictions can be recognized: physical, economic, and technical restrictions. Physical restrictions obtain if some kind of natural resource is to be produced. They determine where the production cannot be located; where one cannot mine or dig. The economic restrictions presume knowledge of the cost structure of the firm. Labor costs, real estate and utility costs, marketing etc., are considered expenditures that reduce the number of viable alternatives. Transport costs are not considered. The reason for this is that Rawstron treats transport costs as a spatial variable to be expressed in the cost structure of the firm as a function of its location. One could say that the economic restrictions image a spatial break-even analysis. The third type of restriction, the technical ones, pertain to locational economies in terms of the needs for technological innovation. According to Smith, they can often be analyzed in more or less the same way as the economic restrictions.

Figure 2.1 represents Rawstron's idea of the effect of location on the cost structure of three imaginary firms. The shaded portions represent locational costs. All three firms are assumed to have identical needs for labor, materials, land, marketing, and capital. But the costs for these production factors vary according to the location. Figure 2.1a shows that labor is more expensive in B than it is in A or C. Materials, however, are the most expensive in C. Figure 2.1b shows total locational costs. It is evident that production at A is cheapest whereas production at C is the most expensive.

Underlying the idea of these restrictive principles is Rawstron's notion of locational choice as a process of elimination of unsuitable locations. The result of that process can then be seen as a set of feasible alternatives; a

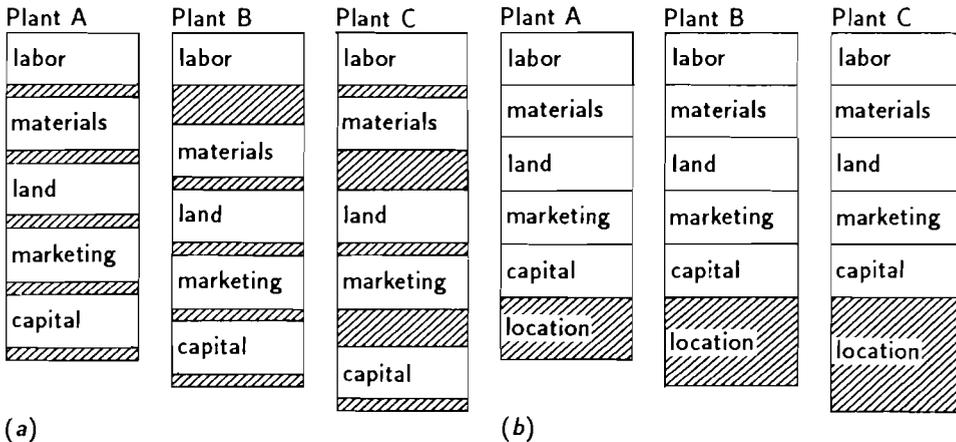


Figure 2.1: The effect of location on the cost structure of three hypothetical plants (Source: Smith, 1971)

choice-set therefore. Not only as a model of locational decision making, but also as a method for reconstructing locational choice, Smith judges Rawstron's approach as very valuable:

“It can help to sort out the importance of different causal factors, and the restrictions they impose. And in some instances, where cost data are unavailable or imprecise, this approach may be as near as one can get to a sensible evaluation of the effect of different variables.” (p. 105)

Rawstron empirically applied his ideas in studies on the location of electricity-generating facilities in Trent Valley in the Midlands, England (Rawstron, 1966).

This idea of narrowing down an initial set of alternatives to a limited set of feasible alternatives is rather attractive, primarily because it offers a way of introducing activity-specific locational requirements, and because it introduces a non-compensatory element. Specific sites cannot be considered suitable because they miss specific characteristics which are required by the activity. However, from the way Rawstron illustrates the effects of, in particular, the economic restrictions, it seems that filtering and optimization are not kept separate. The way Figure 2.1 illustrates the results of economic restrictions seems to imply an optimization procedure based on ranking total costs rather than a filtering process based on restriction and selection.

In comparison to the two-stage approach advocated here, the filtering down belongs to the first stage, the reconstruction of the choice-set of functionally equivalent alternatives. Any kind of optimization, however, would belong to the second stage of modeling. As explained in Section 1.3, these stages should be kept apart because of the different types of modeling required for each of them. Therefore, some serious doubts can be raised as to how the idea of restrictions and a limited choice-set, as suggested by Rawstron, are to be applied in a modeling procedure.

A serious problem associated with Rawstron's approach is the lack of any kind of representational scheme by which restrictions can be associated with characteristics or combinations of characteristics of activities. It does not become clear how restrictions are generated or where they stem from. These relations, however, need to be included in the model.

A different, and much less promising, approach with regard to the filtering-down character of locational decision making is put forward by, e.g., McMillan (1965), Massey (1975a), and Wheeler (1981). Here, the filtering concerns sequential stages in the locational decision behavior of firms, each of them conducted on decreasing geographical scales. For instance, during a search for a site, certain activities may initially consider a specific region on a national level, then a city or town, and finally a location within that town. Different types of activities follow different scale-related choice strategies.

These studies are interesting because they show relations between types of locational choice processes and geographical scale. The problem of this approach, however, is that the emphasis is on the behavioral characteristics of the choice process rather than on the nature of the causal processes and mechanisms driving the choice process. Emphasis on scale tends to conceal real locational constraints, especially when traditional scales such as 'national' and 'regional' are applied. What is called 'winnowing-down' in these studies is therefore quite different from what is called 'filtering-down' here. The winnowing down in the scale-studies is typically behavior oriented. It describes the actual decision making as a winnowing-down process, different types of locational decisions being taken on different levels of scale. Filtering down, as proposed here, is the result of a theory for explaining such locational decision making: it is much more a reconstruction by the researcher than a description of the decision process as it is actually performed by the entrepreneur.

2.5.2 Feasibility studies: the matching approach to site suitability

In order to model site suitability relationally, it is necessary to define it in terms of sets of necessary and sufficient conditions. These conditions, the spatial production requirements of a firm, are a function of the characteristics of the production process. Depending on the type and quantity of raw materials that are needed, for example, specific forms of transportation and infrastructure will be required. Production processes, or in more general terms ‘technology’, also implies demands regarding the quality and quantity of employment. The same can be said of energy requirements. Smelting aluminum (ore) requires a type of energy facility that is different from that required for fertilizer production. Many plants that process raw materials, such as sand or clay (e.g., the production of porcelain out of kaolin) are bound to locations where the resource is available. Other industries have a much more urban-oriented location (e.g., printing industry) because the finished product is heavy and expensive to transport over great distances. Of course, the amount and types of spatial production requirements associated with specific production processes varies.

At the end of the sixties and the beginning of the seventies, a number of feasibility studies were conducted. These studies—e.g., the ones carried out by Schilling (1968) and Pellenbarg *et al.* (1974)—consisted of a careful matching of technology requirement profiles on the one hand, with a regional profile on the other. These feasibility studies deserve attention here because they contain some elements which are compatible with a relational matching approach; the idea of matching as a measurement procedure and the emphasis on non-compensatory requirement structures. A good example of this kind of feasibility study based on matching, the one by Schilling on the site suitability of locations in a province of Austria (Schilling, 1968), is used here to illustrate the merits and problems of this traditional matching approach.

2.5.2.1 The “Standortfaktoren” catalog of Schilling

In his “Standortfaktoren für die Industrieansiedlung (siting factors for industrial location) Helmut Schilling (1968) presents a method of site suitability assessment based on the idea of matching spatial requirement profiles and locational property profiles. The work consists of two parts: an introductory part in which the method is explained, and a catalog containing spatial

requirement profiles for a large number of different categories of industry. These can be confronted with the property profile of locations, the result of which constitutes an assessment of the suitability of that location for locating the activity associated with the requirement profile.

Table 2.1 shows an excerpt from the catalog (for a translation see Appendix 2). On the left, 12 production factors are listed. For most of them various indicators are presented. At the top of the table, a number of industry types can be found. The right of the table shows the combinations of indicators with industry types. The symbols represent either the 'importance' of that locational factor for an activity, or the degree to which this factor is used. Vertical arrays of symbols then represent the activity-specific requirement profiles. Schilling's idea was to 'match' these profiles with the property profiles of locations and thus be able to draw conclusions regarding the site suitability of that location for a given activity. Schilling (pp. 22–27) suggests three types of applications:

1. Assessment of the local production milieu (örtlichen Industriestand-
orteignung);
2. Establishing policy incentives on the basis of the differences between
requirement profiles and locational property profile;
3. Supporting locational decisions for an industrial plant.

Table 2.2 shows part of an application for a number of 'zones' and local authorities in the province of Lower Austria.

The symbols in the table represent the outcome of the matching of the activity-specific requirement profile and the property profile of the location (refer to Appendix 2 for a translation of the symbols). Vertical arrays of symbols contain the activity-specific matching results for various locations, horizontal arrays provide location-specific matchings on various activities. The table constitutes a mixture of the first and second type of application suggested by Schilling. They indicate whether or not a location is a suitable site, or whether it could be a suitable site if, in addition, certain conditions can be satisfied. Comparing the two types of profiles with each other highlights which requirements are not met by a location. This, again, is information that can be useful for regional policy making, because the procedure constitutes a technology-specific evaluation of a production milieu. Its results are material recommendations as to which bottlenecks and niches should be eliminated in order to satisfy the spatial production requirements of that specific technology.

Table 2.1: Industrial activities and their locational requirement profiles (Source: Schilling, 1968)

Standortfaktoren		01 Stein- und keramische Industrie									
		010100 Natursteinindustrie (40), 402	010200 Sand- und Kiesindustrie (401)	010300 Kalkindustrie (401); Gips- und Kreideindustrie (205, 401, 405)	010400 Rohlen- und Koolin-industrie (403)	010500 Ziegelindustrie (404)	010600 Zementindustrie (405)	010700 Betonsteinindustrie (406)	010800 Steinzeugindustrie; Schamotte- und Silika-industrie (407)	010900 Feinkeramische Industrie (408)	
Bewertung:											
absolut											
relativ											
Zusatz											
z. B.: M		sehr große Bedeutung									
M		große Bedeutung									
M		geringe Bedeutung									
		unbedeutend									
Sonstige Aussagen: z. B.: m		(ohne Bewertung)									
Al'gemein wichtig:		*									
01 Arbeitskräfte	011 Personalkosten	●	●	●	●	●	●	●	●	●	●
	012 Arbeiter	● ^A	● ^A	● ^A	● ^A	● ^A	● ^A	● ^A	● ^A	● ^A	● ^A
	013 Arbeiterinnen	□	□	□	□	□	□	□	□	□	□
	014 Angestellte	□	□	□	□	□	□	□	□	□	□
	015 Saisonschwankungen (sa), Heimarbeit (h), Schichtbetrieb (sb)	sa	sa	sa	sa	sa sb	sa sb	sa	sa sb	sb	sb
02 Grundstücke und Gebäude	021 Flächenbedarf	●	●	●	●	●	●	●	●	●	●
	022 Bauliche Investitionen, Nutzungskosten	●	●	●	●	●	●	●	●	●	●
	023 Erschließung	●	●	●	●	●	●	●	●	●	●
03 Maschinelle Anlagen		●	●	●	●	●	●	●	●	●	●
04 Finanzierung		*	*	*	*	*	*	*	*	*	*
05 Roh- und Hilfsstoffe	051 Rohstoffe	●	●	●	●	●	●	●	●	●	●
	052 Hilfsstoffe, fertig bezogene Teile	○	○	○	○	○	○	○	○	○	○
06 Energie	061 Kohle	○	○	●	○	●	○	○	○	○	○
	062 Heizöl	○	○	○	○	○	○	○	○	○	○
	063 Strom	○	○	○	○	○	○	○	○	○	○
	064 Gas	○	○	○	○	○	○	○	○	○	○
07 Wasserbedarf		○ ^M	○ ^M	○ ^M	○ ^M	○ ^M	○ ^M	○ ^M	○ ^M	○ ^M	○ ^M
08 Örtliche Kontakte	081 Lieferanten (L), Abnehmer (A), gleichartige Betriebe (G)	○ ^A	○ ^A	○ ^A	○ ^A	○ ^A	○ ^A	○ ^A	○ ^A	○ ^A	○ ^A
	082 Zentralörtliche Dienste	*	*	*	*	*	*	*	*	*	*
09 Verkehr	091 Personenverkehrs- und Nachrichtenverbindungen	*	*	*	*	*	*	*	*	*	*
	092 Güterverkehr	●	●	●	●	●	●	●	●	●	●
	0921 Verkehrsintensität	●	●	●	●	●	●	●	●	●	●
	0922 Straße	■	■	■	■	■	■	■	■	■	■
0923 Bahn	■ ^g	■ ^g	■ ^g	■ ^g	■ ^g	■ ^g	■ ^g	■ ^g	■ ^g	■ ^g	
0924 Wasserverkehr (w) Luftverkehr (l)	○	○	○	○	○	○	○	○	○	○	
10 Immissionen	101 Betrieb → Umgebung	○	○	○ ^v	○	○	○ ^v	○ ^v	○ ^v	○ ^v	○ ^v
	102 Umgebung → Betrieb	○	○	○	○	○	○	○	○	○	○
11 Produktionsrückstände	111 Abwasser	○ ^Q	○ ^Q	○ ^Q	○ ^Q	○ ^Q	○ ^Q	○ ^Q	○ ^Q	○ ^Q	○ ^Q
	112 Sonstige Rückstände	○ ^v	○ ^v	○ ^v	○ ^v	○ ^v	○ ^v	○ ^v	○ ^v	○ ^v	○ ^v
12 Absatz		●	●	●	●	●	●	●	●	●	●
Zusatz- angaben (Österr. 1962)	Zahl der Mitgliedsfirmen	113	38	35; 16	16	212	11	134	5; 18	33	
	Zahl der Beschäftigten	4731	782	1217; 427	727	7666	4295	4251	308; 1033	3276	
	Exportanteil am Umsatz in %	0,7	0	0,5; 7	31	0,1	2	0	25; 2	13	

Table 2.2: Local authorities in Lower Austria and their suitability for industrial activities (Source: Schilling, 1968)

Betriebszweige (lt. Betriebesystematik des Standortfaktoren-Kataloges)	Stein- u. keramische Industrie				Glasindustrie				Metall- ind.				Gießerei				Maschinen- u. Stahl- u. Eisenbauindustrie												
	1 Betonsteinindustrie	2 Stein- u. Schmelz- und Ziegelindustrie	3 Keramische Industrie	4 Schalsteinindustrie	5 Keramik- und Leichtbaustoffindustrie	6 Hohlglasverzeugung	7 Flachglasverzeugung	8 Allgem. Glasbearbeitung	9 Glaschmuck, optisches Glas	10 Metall- u. Feberze, Eisenbearbeitung	11 Schneid- u. Zähl-Industrie	12 Rohblechbearbeitung	13 Metallblechfabrikate	14 Eisengießerei	15 Metallgießerei	16 Stahlbau	17 Zentralheizungs- und Lüftungsbau	18 Schiffbau	19 Baupfostenbau	20 Kraftmaschinen	21 Turbinen	22 Kfz- u. Flugzeug-, Pumpen- und Druckluftmaschinen	23 Arbeitsmaschinen	24 Fördermittel u. Aufzüge	25 Korn-, Berg-, Rüttel- u. Aufbereitungsmaschinen	26 Landmaschinen	27 Wahlmaschinen u. Gummi-, Textil-, Leder-, chem., Holz- u. Holzleimbau	28 Wagen	29 Wahlgerate und Messergeräte
1 Amstetten	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
2 Berndorf (Zone) ⁴⁾	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
3 Bruck a. d. Leitha	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
4 Ebnofurth	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
5 Eberggening/ Grossteneusiedl	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
6 Ebreichsdorf/Waigelsdorf/ U.-Waltersdorf	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
7 Eggenburg	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
8 Ernetbrunn	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
9 Fischamend Dorf/Markt	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
10 Gänserndorf (Zone) ⁴⁾	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
11 Geras	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
12 Gloggnitz (Zone) ⁴⁾	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
13 Gmünd	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
14 Groß-Gerunze	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
15 Groß-Siegharts/ Dietmanns	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
16 Groß-Weikersdorf	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
17 Rainburg	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
18 Rainfeld	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
19 Raasdorf	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
20 Heidenreichstein	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
21 Herzogenburg (Zone) ⁴⁾	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
22 Rohrau	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	

Zur Industrieansiedlung	geeignet	mit Vorbehalt geeignet ²⁾
sofort	■	▲
nach Ausbau der Infrastruktur ³⁾	□	△

Bearbeitet im Österreichischen Institut für Raumplanung
 Entwurf: E. Schilling, R. Petrusenn
 Graphische Gestaltung: E. Heizer
 Anordnung: P. Waneck, K. Bauer
 Druck: H. Bergmanns, Wien VII

Erläuterungen:

1) Die Industriestandortseignung wurde unter Heranziehung der Standortfaktoren-Kataloge beurteilt, dessen Betriebesystematik und Bewertung der vorliegenden Tabelle zugrundeliegende betriebliche Standortfaktoren mit den oben Standortvoraussetzungen der 56 für eine Industrie geschiedenen n.d. Gemeinden und Zonen Bewertung Strombedarf, Gasbedarf, Wasserbedarf, örtliche Konzentration (Straßen-, Bahn-, Wasser- und Luftverkehrsmitteln (= Störfaktoren) und Abwasserbelastung "Arbeitskräfte" wurde nicht einbezogen.

2.5.2.2 Evaluating the matching approach: advantages

The approach followed by Schilling deserves special attention because it contains some very valuable elements. The idea of the definition of production milieu as a matching result and the extension of the concept as the set of locations satisfying the requirements in the definition, is attractive. It offers means for integrating some ideas from functional location theory in a much better and smoother way than techniques such as R.E.B. methodology and production milieu matrix. However, there are also a number of important problems connected with Schilling's approach. The most important ones have to do with the status and forms of the requirement profiles. The typology of generic production types presented appears to be the result of a standard operation of generalization by empirical similarities, rather than by functional requirements. Before going into these problems, however, the advantages of an analysis such as Schilling's must be mentioned.

1. *Activity-specific:*

Modeling site suitability as a matching problem implies that its definition and measurement should be actor-specific. Depending on the characteristics of the actor, different properties of the location become important. Each of the abstract production factors in Table 2.1 (labor, energy, infrastructure, etc.) is defined differently for different types of technologies. This implies that for different types of actors, site suitability is defined in a different manner. A matching approach such as Schilling's denies the assumption that there could be one score or quality judgement about the site suitability of a location. On the contrary; whether or not a site is suitable depends on its ability to satisfy the requirements associated with a particular type of actor. Great care should therefore be taken when interpreting a map such as shown in Figure 2.2, a map of Niederösterreich, derived by application of Schilling's technique (Österreichisches Institut für Raumplanung (Austrian Institute for Spatial Planning), 1965).

The area has been divided into a number of categories representing the overall site suitability of the locations, measured as the amount of requirements met for all activities included in the analysis. The dark shading represents large numbers of available location factors, the light shading, small numbers.

Aggregating activity-specific measurements to an overall measurement can be interesting if one wishes to investigate the range of activities for which a location offers a suitable site. This is how the map in Figure 2.2 should be understood. But great care should be taken when using it as a guide for the

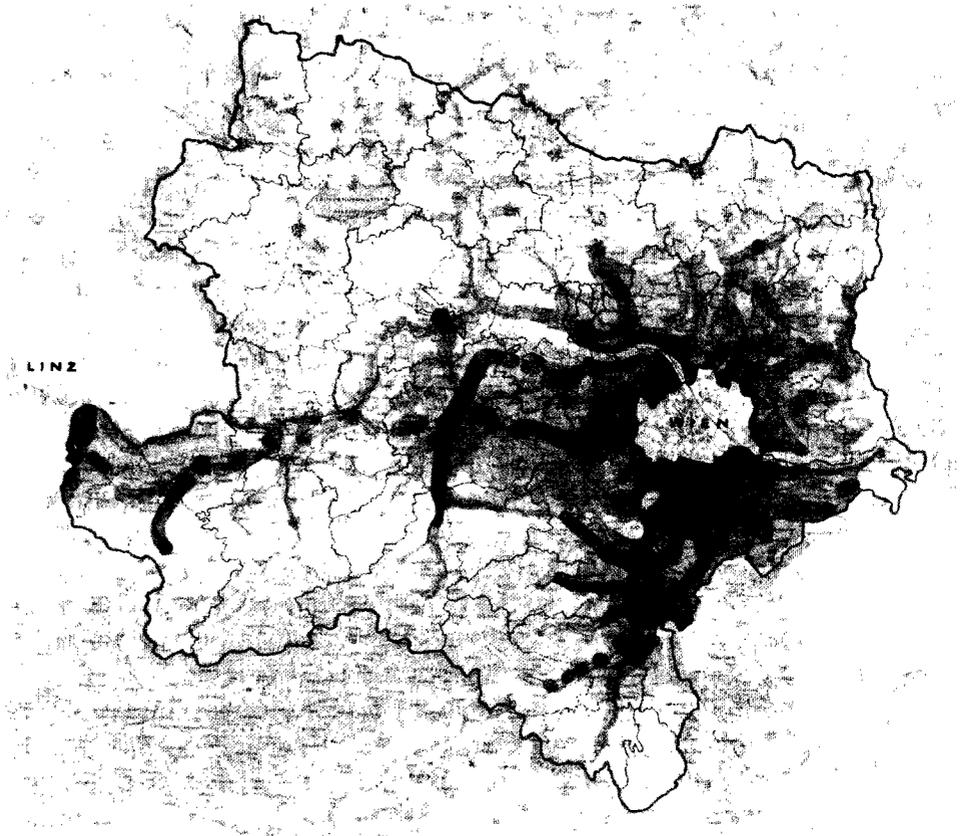


Figure 2.2: Distribution of the availability of locational factors as described in Schilling (1968) (Source: Österreichisches Institut für Raumplanung, 1965)

formulation of regional policy incentives. Serious doubts can be raised as to whether such a result offers any starting point for regional policy at all! Matching as executed by Schilling has its attractions because it departs from the view that site suitability measurements should be activity-specific. The activity-specific definitions and measurements offer suggestions for regional policy incentives, because they show why certain activities cannot establish a plant at certain locations. Aggregating over the activity types implies the loss of this crucial piece of information. What is alluring is a classification in terms of 'good' and 'bad' suitability of the site, regardless of the activity for

which it is measured. This kind of conclusion must be rejected. Aggregations such as those shown in Figure 2.2 can easily persuade one to interpret site suitability as a property of an area instead of the outcome of a (relational) match. The results of the aggregation should therefore be used with utmost care.

2. Separation of preferences and requirements:

The kind of matching advocated by Schilling corresponds well with a two-stage modeling of locational decision making. The matching conducted on the basis of requirements belongs to the first stage. Starting with an initial set of locations, matching for a set of requirements will result in a reduced set of alternatives, each of which can be considered a feasible site. This is consistent with the idea of production milieu as a choice-set from which, by means of optimization based on a set of common dimensions, a best or most attractive alternative may be selected. The requirement profiles themselves represent the activity-specific conditions under which a location can function as a site for that activity. This emphasis on the relation between activity-specific needs and locational properties is consistent with functional location theory.

3. Non-compensatory matching:

Unlike techniques such as R.E.B. and production milieu matrix which apply completely compensatory combination rules, the method by Schilling has a typically non-compensatory character. Locations must meet the requirements by the activity. In case they do not, they fail to pass the site suitability test and cannot be considered part of the production milieu of the activity. The (undesired) additive character that is inherent in virtually all algebraic models combining the contributions of individual indicators to the overall measure of site suitability is not present here.

4. Constraints and freedom of choice:

In addition to the mere methodological advantages connected with matching, there are also some advantages of a more theoretical nature involved. Matching provides a means of incorporating the constraints on locational choice into the model. Constraints on locational choice limit the freedom of choice. They more or less 'set the stage' for the actual choice process. Matching, therefore, provides a very distinct interpretation of freedom of choice. Being 'footloose' is often associated with a large degree of locational freedom. In terms of matching this implies that in case of 'footloose' industries the initial set of locations will hardly be reduced because of spatial

production requirements. The term 'footloose' can thus be interpreted in terms of the reduction of alternatives as a consequence of a specific configuration of production requirements. Given an initial set of locations, the degree to which an activity is footloose can be expressed as the inverse of the amount of reduction in this initial set as a consequence of the spatial production requirements associated with that activity. Matching thus provides a formal interpretation and a way of measuring the degree of being footloose. However, some caution is required. From Table 2.2, for example, it becomes clear that for the abrasives industry [Schleifmittelindustrie (4)] each of the locations listed is considered suitable. Its requirement profile in Table 2.1 does not contain many indicators that might limit the number of feasible locations. Although tempting, concluding that the abrasives industry is a footloose activity from the requirement profile alone would nevertheless be incorrect. The relative freedom of choice (a large choice-set) is a result of matching the requirement profile with the property profiles of the locations. The degree to which an activity is footloose can therefore only be meaningfully assessed when compared with another activity on a common, initial set of locations. This is illustrated in Table 2.3. A '+' denotes a positive score on site suitability, '-' a negative score. Because of the different nature of the requirements of A and B, they cannot be compared in terms of 'more' or 'less' requirements. If the locations represent the initial set, both activities can be located in only two locations. This means that relative to this initial set their degree of being footloose is identical. C is more footloose, however, since it can be located in three of the four initial locations. However, had we chosen another initial set of locations, it would have been possible that A or B would have turned out more footloose. This shows that an activity cannot be labeled 'footloose' by virtue of its own characteristics alone. Statements about the degree of being more or less footloose are only meaningful within the context of a set of actors and one, common set of locations.

5. Regional potential:

Matching also allows a somewhat alternative view of the concept of regional potential. Strassert argues that regional potential should be understood as a measure of capacity, to be estimated only after a (hypothetical) decision has been taken on how the resources, in terms of production milieu, ought to be utilized. An alternative interpretation that deviates somewhat from the idea of capacity is based on a set-theoretic operation conducted on various production milieux. Production milieu represents a set of locations that satisfy the site suitability problem for a given activity. Regional potential can then

Table 2.3: Production requirements, locational properties and relational matches

	Activities	A	B	C	
Locations	Attributes	1,2,3	7,8	1,3,9	total
1	1,2	-	-	-	0
2	1,2,3,9	+	-	+	2
3	1,4,5,6,7,8,9	-	+	+	2
4	1,2,3,7,8,9	+	+	+	3

be defined as the union of these sets, hence the set of all production milieux. This is the complete set of all combinations of activities and locations of which requirement profile and locational profile match. A set, therefore, of which each of the elements designates a positive outcome on site suitability, given a location and an activity. For Table 2.3 this means that the regional potential contains those combinations of activity and locations as elements that have a '+' value. Note that this is not the same as what is depicted in Figure 2.2. What is shown there is the number of positive results on site suitability per location (column 'total' in Table 2.3).

This definition of regional potential differs substantially from the interpretation by Strassert. Strassert considers capacity the main characteristic of regional potential. Regional potential as the set of production milieux, a set of sets, does not signify capacity, but denotes the total set of locational opportunities for a given set of activities. Of course, each of the opportunities that is an element of this set has a capacity associated with it.

6. *Diagnostics:*

Schilling's second suggestion for application of the catalog has to do with diagnostics. If comparison of requirement profiles with a locational property profile yields a location classified unsuitable, it is possible to diagnose the problem. Because of the clarity of the procedure and the non-compensatory character of the classification, it can easily be determined why a location cannot be considered a suitable site for an activity. This again makes it possible to formulate very precise and straightforward recommendations about how to improve the suitability of a given location for a given activity.

2.5.2.3 Evaluating the matching approach: problems

Schilling's matching approach contains some valuable elements. Looking at the site suitability problem from a matching perspective offers opportunities to include many features that non-matching approaches do not. The matching perspective, furthermore, establishes a very close relationship between a functional theory of site suitability and a procedure for modeling the concept. Schilling's catalog constitutes a model of site suitability, site suitability itself being defined as a match between actor-specific requirements and locational characteristics.

As a modeling procedure, however, Schilling's method also contains problematic elements just as much as there seem to be some missing elements:

1. *Compensatory versus non-compensatory:*

Schilling's model of activity-specific site suitability is purely non-compensatory. If a certain requirement is not satisfied, the location is eliminated. Although this may be fine on a rather abstract level of general production factors, Schilling's method does not provide any means to incorporate compensatory elements in the formulation of the actual locational requirements. However, requirements can be modeled as conjunctive and disjunctive structures, the disjunctions representing the various ways in which a requirement can be fulfilled (Section 1.4.2). Energy needs might be met in different ways. The same holds for transportation of raw materials, the acquisition of labor, etc. A matching model of site suitability should provide means to incorporate both compensatory and non-compensatory structures.

2. *Dubious locational requirements:*

Only some of the locational requirements present in the requirement profiles in Schilling's catalog are expressed in terms of empirical locational characteristics. Several requirements listed in the profile, however, are actor attributes rather than locational requirements. Looking at the production factors 'labor' and 'energy', for instance, it becomes clear that what is actually put into the requirement profile are the types and relative quantities of labor or energy needed by the activity. The energy requirements (factor 06), for instance, are defined as the percentages of energy needs to be covered by the various fuels. But although this is a requirement for the activity to continue to exist, it is hardly a locational requirement. That 50% of the employment is to be occupied by employees and the other 50% by laborers is interesting and relevant, but it does not specify the locational requirements that can be derived from this employment configuration. This partial

lack of a translation of production requirements into locational requirements raises some doubts concerning the measurements of the Lower Austrian application. How were the energy and labor demands, such as they are in the catalog, translated into locational properties? This translation process remains unexplained.

A somewhat similar objection can be raised against the way some of the other locational requirements are listed. Some of the indicators of production factors are represented in terms of ranking concepts such as 'important', 'very important', and 'not important'. But as was mentioned in the discussion on R.E.B. methodology and production milieu matrix method, these kinds of categorizations should be avoided since they do not contain much useful information. What should be represented instead is what this importance amounts to in terms of locational requirements. What should be contained in the model is first, how actor properties are associated with certain forms of energy, raw materials, labor, etc., and second, how these are again related to locational characteristics. Furthermore, the importance of locational properties as production factors is not only dependent on actor characteristics and objectives, but can also depend on the relevance of other locational properties. Locational characteristics become important, depending on the possibilities for establishing a positive outcome on site suitability.

3. Fixed activity types:

Schilling's catalog contains 158 different categories of industrial activity. These are generic types of industry. This means that one does not know, for example, how much they actually produce, which types of raw material they use, whether substitution can occur or not and, if yes, which materials can be substituted for each other, etc. What is presented are the relative amounts of certain production factors. Thus it is given that in radio and T.V. manufacturing 65–75% of the employment is covered by women laborers, 10% being professionally trained. It is furthermore given that labor costs make up 20% of the total production costs. Of course this is useful and even necessary information, but in order to carry out a real matching one has to know the estimated total production. If this kind of information is not available, one neither knows how much space is needed, nor how much capacity a high-tension line should have.

An additional disadvantage of these kinds of fixed and rigid profiles is that any new type or variant on one of the existing types, should be added to the catalog as a new and separate profile. Apart from the fact that this can make

things rather cumbersome to work with, it introduces a lot of redundancy because many industries have common characteristics and therefore common locational requirements. Ideally, these should only be represented once in a model of site suitability rather than being associated with each and every type of industry that displays them.

Underlying these problems with the activity types as presented by Schilling, however, is a much more serious methodological problem, namely that of the 'objective' status of the typology. Earlier (Section 2.1.2), Massey (1975a; p. 85) was cited, where she advised "not to set up ahistorical ideal types", but rather to consider the behavior of activities as a function of their relationships with each other and the economic system as a whole. Now, one interpretation of such a statement could be that what one should concentrate on is not empirical similarities between activities, but similarities in functional relationships. Schilling's typology is based on a standard categorization of industry and for each of the existing categories average values for locational requirements have been taken. This, however, assumes an objective functionality of locational properties with regard to the types in the typology. But as Massey argues, if something in the strategy or objectives of an activity changes, it cannot be considered an instance of the ideal type anymore. Ideal types must be associated with inductive modeling based on empirical similarity. As in the other inductive approaches discussed before, Schilling does not model the locational requirements as functions, as consequences of activity characteristics. They are associated with generic activity types that have these characteristics, but not with the characteristics themselves. This is dangerous and tends to conceal the true relations between locational requirements and activity properties. The fact that processing huge amounts of iron ore requires deep water in order to transport the ore, links smelting-furnaces with deep water. This is the kind of link that is present in Schilling's catalog. What should have been declared, however, is that deep water is needed because a bulky material such as iron ore needs to be transported by ship, and that iron smelting implies the use of bulky material. If transport technology and/or costs change, the relation between bulky material and deep water may change, and therefore the relation between iron smelting and deep water. The developments in transport technology are a factor external to iron smelting (contrary, to e.g., improvements in iron ore extraction technology). Explicitly linking iron ore smelting with deep water instead of linking it with the processing of bulky material and linking bulky material with deep water, conceals the relationships that explain why

large blasting-furnaces need access to deep water. It furthermore makes it difficult to adjust the model in case transport technology changes in such a way that the relations between bulky material and means of transport have to be revised. This last point illustrates once more the disadvantages of context-specific measurement. Although the fact that iron smelting may, at one moment, be correctly associated with the requirement of deep water, as soon as something changes, the association becomes faulty. This would require a new measurement and a new profile to be set up. In other words, from a relational approach based on the concept of functional equivalence, it follows that developing profiles based on the empirical similarity of activities *must* be wrong, because they have no real functional meaning. The point of context-specific measurement just mentioned illustrates this very clearly. Even though, at a specific moment, such profiles might correspond nicely with functional profiles, as soon as something changes in the functional relationships, the profiles are no good anymore. Therefore, profiles for matching may be used, but they have to be derived from functional profiles, stored in a relational model.

4. *No conditional dependency:*

A problem which was also mentioned in relation to the production milieu matrix is that every production factor is present in each requirement profile, even if it is completely irrelevant. Of course, it can get assigned an 'irrelevant' value. Although this does not directly influence the measurements, including irrelevant dimensions is something that is a consequence of ignoring the conditional dependence of production factors. It also introduces redundancy such as mentioned in connection with the point of the fixed-activity types.

5. *Actor versus object demands:*

In the modeling approach proposed by Schilling, it is not clearly recognized that it is not only activities that have requirements. Often, the locations themselves have requirements regarding the characteristics of the activity. If, because of legal regulations for example, a location can only be allocated to certain types of activity, then these regulations form another limitation on the possible result of the matching process. This notion of a two-way matching process is depicted in a scheme by Lucardie (1988a; p. 62 [my translation]), shown in Figure 2.3. Both object and actor have requirements, and a relational match can only come about if both are able to satisfy each other's requirements.

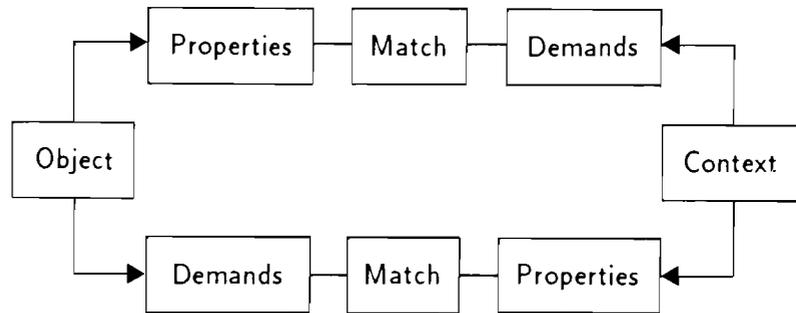


Figure 2.3: Matching as concurring profiles of requirements and actual characteristics (Source: Lucardie, 1988a)

Schilling's catalog contains a production factor 'emissions' (no. 10) which again refers to an assessment of the influences the activity has on the environment. But from the catalog it does not become clear what is to be done with that information. It does not include a possibility to limit the number of suitable sites because the activity does not comply with the requirements by the location.

Such a matching in two directions, however, was conducted by Pellenberg *et al.* (1974) in an application of the Schilling catalog for the Dutch town Medemblik. After Medemblik was scanned on its suitability for a number of activities, the activities were scanned for their appropriateness for Medemblik. A procedure for modeling and measuring site suitability by means of matching must contain these two types of requirements.

In conclusion one can say that the matching method suggested by Schilling avoids a lot of problems inherent in approaches like R.E.B. methodology and production milieu matrix. It explicitly recognizes the activity-dependent nature of site suitability, it contains the idea of suitable sites as elements in the choice set, and it introduces non-compensatory elements. As such, it has numerous advantages over the other methods discussed earlier. As a modeling procedure, however, it contains some serious weaknesses. The compensatory element is lost, activity types are treated rather rigidly, requirements are not quantified and not connected with activity characteristics, and actor attributes and locational requirements are sometimes confused.

2.5.3 Intermezzo: automating the matching process

In spite of the problems mentioned above, the matching conducted in the feasibility studies constitutes a promising technique for modeling and measuring site suitability and production milieu. But then why was it forgotten, or disappear for such a long time?

Most feasibility studies based on matching were conducted at the end of the sixties or the beginning of the seventies. Since then they seem to have more or less disappeared from the geographical stage. One of the few applications of a form of matching approach to measuring site suitability, concerns energy facility siting studies (Hobbs, 1980; Calzonetti and Eckert, 1981; Church and Bell, 1981). Hobbs (1980), for instance, mentions some advantages and disadvantages of a matching approach for determining suitable sites for energy facilities. In an attempt to assess the appropriateness of baseline screening techniques based on a standard additive weighted utility function *versus* optimization by means of mathematical programming for determination of site suitability for energy facility location, Church and Bell (1981) conduct a pre-screening of an initial set of alternatives to derive a feasible set. The screening is mentioned only in a footnote and turns out to be rather simple. They merely list a few simple requirements (availability water, availability railway, etc.) and test the alternatives on these without differentiating between the different types of power plants and without explaining how they were operationalized. Why nuclear power plants need access to a railway, for example, or how this access was measured, is not explained. Neither compensatory nor quantified information is used.

Matching approaches were no longer popular in location theory and its applications in economic geography, since the period around the beginning of the seventies. It is not clear whether this was because matching did not constitute a promising modeling technique. Or was it a mistake to assume that the definition and measurement of concepts from a matching point of view could contribute substantially to the theory of spatial choice, location theory and their applications? I think not. A matching approach offers ample opportunities for modeling spatial choice and spatial decision-making processes, simply because it is consistent with a two-stage reconstruction of spatial choice, and because it implies logical modeling. A relational model is a logical model, like a matching model is a logical model. Schilling's model suffers from various drawbacks. In the next chapter, however, it is shown that many of these can be solved by departing from a purely relational point of view. But if matching models are promising, then why did they vanish

so quickly after they showed up? A quick and obvious answer would be to say that the method contained too many flaws, did not work as a result, and therefore was not acceptable. However, the answer may be a little more complicated than that. Ever since Kuhn wrote his book on scientific revolutions (Kuhn, 1962), it was accepted that unsuccessful theories, models, and methods are not automatically and immediately replaced by other ones. There could be at least two reasons for not pursuing matching modeling any further: the methodological shift toward 'spatial analysis' with its emphasis on statistical and mathematical models, and the lack of opportunities to automate the matching process and to store huge amounts of information about activities and locations.

A matching procedure like the one proposed by Schilling is a very qualitative approach. No parameters are estimated, no algebraic functions specified or calibrated. During the period of 'spatial analysis' a matching analysis like this must have been considered a rather archaic approach, not exactly tallying with the predominating methodological paradigm. Sophisticated statistical techniques and mathematical modeling must have appeared a lot more promising than a rather laborious and painstaking endeavor implied by a technique such as Schilling's. Moreover, Schilling's approach did not result in a real model, or at least not in a model resembling an elegant mathematical equation or set of equations. Instead, it much more resembled a complex data matrix that could be the starting point rather than the result of the model-building process.

Very much related to this development, was the shift in economic geography from the normative, deterministic, classical location theory, toward a more inductively oriented, behavioral approach with its emphasis on uncertainty and probability. Matching, as it stood, had a very deterministic character and did not pay any attention to behavioral variables such as regional perception, incomplete knowledge about the environment, corporate decision making, and entrepreneurial attitudes. As such, it did not fit in very well.

Conducting a matching study, furthermore implies a large amount of work, most of which has to do with collecting and storing information about locational properties and activity requirement profiles, as well as with the matching of these two. A study, such as the one carried out for Lower Austria (Österreichisches Institut für Raumplanung, 1965) implies a tremendous amount of work which, at that time, had to be carried out without the assistance of computers. Moreover, if something changed in the requirement profile of an activity or in the property profile of the location, the entire match-

ing for that activity or location had to be done again, manually. Although computers were available by then, they were primarily used for mathematical and statistical calculations. Data bases and tools to represent requirement profiles or to carry out a matching simply did not exist. It is only very recently that these tools have become widely available. In Chapters 4 and 5 it is explained that the results from AI research, such as the development of tools for logical programming, knowledge representation, and inference engines, together with the availability of fast and powerful machines, made it possible to implement matching models and to automate their execution.

To a certain extent this implies that the matching studies were ahead of their time. Only after some time was it apparent that a purely behavioral approach toward locational decision making could not offer any real solution. It was recognized that constraints limiting the decision space have to be taken into account. This does not imply the re-appearance of uniform space, profit maximization and complete information, but it does imply the concept of activity-dependent choice sets. Further, matching contains numerous qualitative moments. It was a long time before the tools for working with this kind of qualitative information became available. I expect that in the near future many modern varieties of matching models will (re)appear.

2.5.4 Computer-based matching: Cullen's proposal

An interesting version of one of those automated matching approaches for the assessment of land-use suitability is the one proposed by Cullen (1986). The proposal is embedded in a more general article on the application of expert systems in planning analysis, but it shows an example of a modern, computerized version of the approach by Schilling. The proposal is interesting for two reasons; first, because it shows some of the possibilities of modern computer technology, and second, because it illustrates that automation by itself does not improve the quality of the modeling. Cullen (p. 244) mentions the following conditional matching statement:

```

IF      there are no geological problems
AND    there are no access problems
AND    all the basic infrastructure services are available
AND    EITHER      outline planning permission is granted
           OR      the site is in an enterprise zone
           OR      it is zoned for industrial use
THEN   it is technically suited for industrial use

```

Application of expert systems technology then implies that a computer program can handle this kind of rule and can investigate which locations, from an initial set of locations, is suitable for locating an industry. It will be clear that this kind of definition is very similar to the ones suggested by Schilling. Both approaches contain the concept of functional equivalence, but both suffer from the same problems. The rule presented by Cullen can be part of a matching model for site suitability, but only if elsewhere in the model 'access' and 'infrastructure services' are clearly defined. 'Access' and 'services' themselves should be defined in such a way that for different types of activities, different requirements may become important. What should be represented in the model is how various empirical properties of locations are to be associated with characteristics of activities where it concerns accessibility and service-level.

The two points I would like to make here are the following. First of all, mere automation of the matching process does not constitute an improvement of the model. The second point is that new technology, such as expert systems technology, does indeed offer many attractive opportunities in the sense that it provides the means to build alternative types of models that, by virtue of the automation, can easily be handled, formally checked, and automatically extended and modified.

2.5.5 Studying individual firms

Recently published studies on the individual production milieu of firms (Glasmeyer, 1988; Ellegard and Alvstam, 1987; Alvstam and Ellegard, 1989; Vaessen, 1989; de Smidt and Wever, 1990) seem partly to have evolved from a realization that a functional interpretation and approach to production milieu and site suitability must indeed relate the firm's individual characteristics and objectives with the economic geographical environment it is set in. Some of the studies were more general or had a slightly different objective than the development of better ways to study production milieu. The study by Vaessen (1989), for instance, was directed at an assessment of the relevance and existence of regional differences in production milieu in the Netherlands.

The argument underlying the studies on the behavior and location of individual firms is simple. Since a functional approach pinpoints the relevant locational choice mechanisms to actor-specific relations between objectives and environment, actor-specific investigations should lead to a better under-

standing of location decisions and the meaning and relevance of production milieu.

The general inclination in most of these studies seems to be that, first of all, production milieu as a concept should not be considered in too isolated a manner. This concurs with the assertion put forward at the beginning of this chapter that the locational decisions of firms must be regarded as only one aspect, one feature of a much broader and more complex structure of decision making associated with the planning and implementation of an entire production strategy. A second kind of general agreement among these studies is that at least in what can be denoted 'advanced economies' located within a fairly limited space, regional differences in production milieu seem to diminish at an increasing rate. And since this seems to be the case for many different types of firms and activities, the image of a modern, concentrated spatial economy as one rather uniform 'production field' instead of a configuration of many, individually different production milieux emerges. In terms of a choice-set approach to production milieu, this implies that in these economies, the sets of functionally equivalent choice alternatives become relatively large since the number of constraints decrease or the number of compensatory opportunities increase. In the following this difference between the initial set of locations and the subset of feasible locations is interpreted as a measure of footlooseness. Here it suffices to say that although these recent studies indicate that in some cases the reduction of the initial set of alternatives to a smaller set of functionally equivalent alternatives may be rather small, the relational approach of choice-set reduction itself is not at stake here.

It is clear that from a relational analysis point of view, individual firm location decision studies are preferred over aggregate approaches such as R.E.B. or the production milieu matrix method. However, purely individual studies of behavior alone must be considered insufficient for understanding, and perhaps even predicting, spatial patterns of, in this case, the location behavior of firms. A collection of individual cases is fine and more or less unproblematic, but things only become really interesting if the individual cases can be considered instances of more general, more structural types or mechanisms. It will be in the area of decision support that being able to recognize a specific case as an instance of a wider class or a configuration of instances from wider classes, becomes particularly important. If one only knew how the firm SAS (so-and-so) behaved, without recognizing that in the case of SAS a number of structural relationships hold, then how can one hope to say something meaningful about the situation for a case which is similar,

but not identical, to SAS. Or worse, without connecting SAS's individual requirements with some of its objectives and characteristics that give rise to these requirements, how can one ever decide that another firm is similar to, or different from, SAS? The ability to recognize a case as (dis)similar to SAS presupposes the availability of a set of criteria by means of which this (dis)similarity can be assessed. Absence of these criteria leaves one with a collection of individual cases devoid of any cross connections, regularity or whatever kind of systematic.

One can conclude that studies of individual firms are good and necessary, although they are only part of the story. For a meaningful, as well as applicable model, of for instance, production milieu, a model structure is required in which these individual cases can be represented as instances of a more general structure of relationships.

2.6 Conclusion: toward a relational matching model

Relational analysis can be part of two-stage modeling of locational decision making, the first stage consisting of a reconstruction of the choice set. This implies matching an initial set of alternatives against the set of requirements representing an activity. This matching can be considered the processing of a set of relational decision rules linking the characteristics of an activity with requirements, and comparing these with locational profiles.

A relational model is a logical structure, representing the compensatory and non-compensatory rules an object has to satisfy in order to become part of the extension of the concept represented by the model. Schilling's method provides a useful start-up for building relational matching rule-bases, since it contains a lot of actor attributes and higher-order dimensions that could be incorporated in a relational model. In other words, it provides some clues, some general aspects of locational decision making that might be used as a framework for building a relational model of site suitability. From a relational point of view, however, the method contains too many flaws and faulty elements. In the next chapter I therefore suggest an alternative approach that nevertheless implies the application of matching. Based on a relational framework, a formal representation of site suitability avoiding most of the problems associated with Schilling's approach is proposed. Such a model which takes the form of an 'inference-tree' constitutes the relational definition of a concept such as site suitability. It serves two functions. As

the relational definition of the concept it contains the rules specifying under what conditions a location can be regarded a suitable location for an activity. Moreover, it also provides an attractive framework for conducting the actual matching operations.

Chapter 3

Relational Inference Trees

ABSTRACT

A technique and representation scheme for a relational matching model are presented here. It is argued that a relational definition can be represented in a tree-like or decision-table format. This tree, which can take a nested or ‘dimensional’ form, can be processed by a program searching for a path through it. On the basis of this path a matching is carried out. The formal criteria and techniques for evaluating the model’s content are explained and discussed. The inference-trees method is then evaluated in the light of relational modeling and locational choice analysis. Although only mentioned briefly, some attention is paid to the process of knowledge acquisition and the role the matching model could play in that context.

Keywords: matching model, inference trees, decision rule, matching rule, nested trees, dimensions, tree optimization, ID3 (information entropy), multi-locational actors, two-way matching, knowledge acquisition

3.1 Introduction

Defining a concept relationally implies declaring how an object can fulfill a certain function, given a specific context. When dealing with the explanation of (spatial) choice behavior, this context consists of an actor such as an individual, a household, or a firm. The characteristics and interests of this actor have to be relationally linked with the empirical properties of

the object. This relational linking means that the object must be able to fulfill the requirements by the actor, whereas the actor has to be able to conform to the set of demands generated by the object. Confronting the requirement profiles with the property profiles is denoted 'matching'. The term relational matching stands for the entire operation of first, deducing requirements on the basis of the actor's characteristics, and next, conducting the actual matching. A relational match occurs in case the characteristics of the actor and the properties of the object coincide in such a way that the property and requirement profiles match each other. In terms of relational modeling this means that the object can, in principle, fulfill its role, as is meant by the functional definition describing it. Objects contained in the matching set are said to be functionally equivalent. In most cases there is a great variety of combinations of actor characteristics and object properties, giving rise to many different requirement profiles, of which the matching against a set of objects may generate as many different solutions. Different requirement profiles can be the result of either differences in the actual characteristics of an actor, or variation in strategies and objectives by the same actor. Apart from this variation in activity types, a relational model must also be able to represent the disjunctive variation in requirement profiles, given a specific activity. These profiles then take the form of disjunctive-conjunctive networks. Furthermore, it must be possible to represent internally related variables in the relational model. Variables are internally related in case the categorization of one variable is a function of the score of another variable (recall the disabled-floor-elevator example of Section 1.2.1.1). Inference trees are suggested here as a scheme, and structure, for modeling the relational definition of concepts. Inference trees offer means to represent the variation in activity types, the functional dependencies of requirement sets, and the disjunctive-conjunctive nature of these sets themselves (Reitsma, 1987, 1988b, 1990). Moreover, they provide an attractive scheme for the procedural part of relational matching: the confrontation of requirement and property profiles.

3.2 Relational definitions as inference trees

An inference tree is a tree-like representation of a decision table representing a logical decision and classification procedure (e.g., Quinlan, 1979; Verhelst, 1980; Arbab and Michie, 1985, 1987; Thompson and Thompson, 1986). Such a tree can be used as a model for many things, but interpreted in a specific

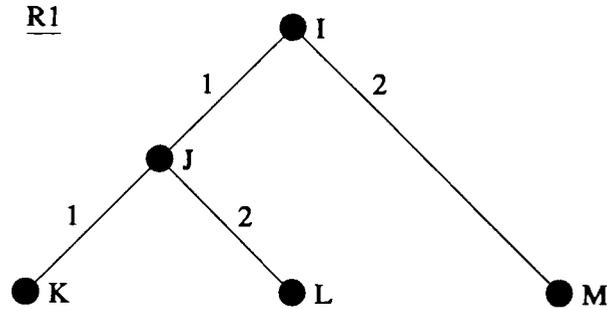


Figure 3.1: Relational inference tree

manner it can be regarded a model of a relational match. An inference tree for a hypothetical relational variable R1 (e.g., site suitability) is shown in Figure 3.1.

As with any symbolic tree, the tree in Figure 3.1 consists of two types of elements: nodes and links or arcs. Three kinds of nodes are present: a starting node or root {I}, terminal nodes or leaves {K,L,M}, and the intermediate nodes {J}. In order to make the tree designate a relational definition, let the root and intermediate nodes {I,J} represent actor characteristics, and let the terminal nodes {K,L,M} represent sets of requirements, each of which is associated with the set of actor characteristics represented by the path leading to it. Interpreted this way, a (relational) inference tree contains a number of routes representing possible ways of generating a relational match. Each of these routes denotes a decision rule.

The tree in Figure 3.1, therefore, contains three actor types, each of which can be represented as a set of actor characteristics: {I1,J1}, {I1,J2}, and {I2}. Each of these sets has a set of requirements attached: K for {I1,J1}, L for {I1,J2}, and M for {I2}. The set of decision rules represents the disjunctive part of a relational definition. It designates the different combinations of actor characteristics and object properties that can make the object fulfill a certain function for the associated actor type. Each of the individual decision rules, in turn, represents the conjunctive part of a relational match. In order to fulfill a function for an actor, all of the conditions that are part of the decision rule have to be satisfied. The requirement sets {K,L,M} themselves refer to conjunctive/disjunctive structures, also denoted matching rules. In case of site suitability, for instance, they represent the compensatory and non-compensatory parts of the locational requirement profiles of firms. It will be clear that since these matching rules constitute

purely logical networks, they too can be modeled in a tree-like manner. The nodes of those trees cannot, of course, be interpreted as actor characteristics. Instead, they simply refer to object properties. The routes in the tree refer to the different combinations of object properties that satisfy the actor's needs. Representing a relational definition in the form of an inference tree thus satisfies the need for a model that can incorporate both the inter- and intra-actor diversity in the ways relational matches can occur.

An important distinction between different ways of representing concepts, concerns the difference between declarative and procedural definitions (Simon, 1969; Sowa, 1984; Hendriks, 1986; pp. 153–160). A declarative representation of a concept refers to its 'meaning'. It does not refer to a method or procedure by means of which it can be measured, or by means of which an object satisfying the definition can be constructed. A declarative representation pertains to 'knowing *what*'. A procedural representation, on the other hand, is a representation in terms of a set of actions to be conducted in order to measure the concept represented, or in order to construct an object that satisfies it. A recipe for a cake constitutes a procedural representation of a cake. An exuberant poem describing its form, color, composition, and exquisite taste, however, forms a declarative representation. Similarly, a function derived by an ordinary least squares multiple regression procedure constitutes a declarative representation of the association between a dependent variable and a set of independent variables. The arithmetic procedures that must be applied in order to find the proper parameter values for the function, either in the form of a text or as a computer program, form the procedural representation. The use of a logical model such as an inference tree implies a close connection between declarative and procedural representation. The inference tree as a whole forms the relational definition of a concept. It represents the notion that a relational match means that an object can fulfill a function for an actor, if the actor is of a certain type and the object exhibits certain characteristics. But apart from this declarative aspect, it also implies that a relational match is the result of matching requirement profiles with property profiles. This means that in order to find out whether relational matches do exist, i.e., conducting a relational measurement operation, the inference tree has to be traversed. This implies finding a route through the tree, by comparing the actor characteristics declared in the nodes of the tree with the activity for which the matching is performed. After the tree is traversed and a terminal node is reached, each of the locations that is part of the initial set of alternatives has to be tested on the requirements associated with that terminal node. If they match, they

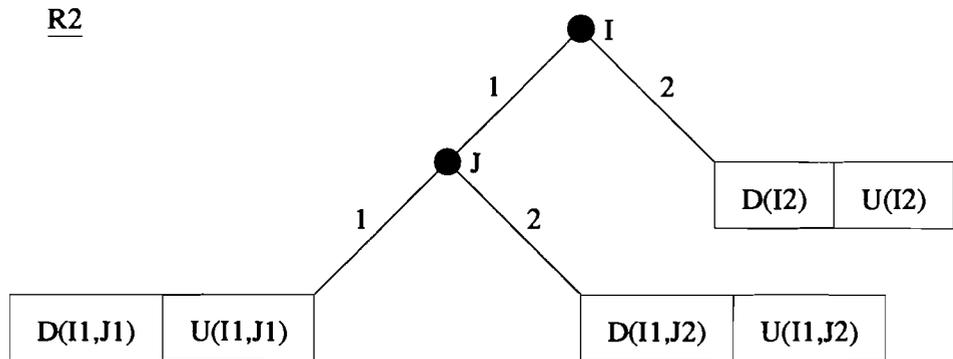
can be regarded a member of the choice set. The locations in this set are functionally equivalent.

3.2.1 Formal properties of a relational inference tree

It is now possible to interpret the inference tree in Figure 3.1 in terms of three features that together satisfy both the procedural and declarative aspects of a relational definition.

1. *The decision attribute:* When going through the tree, a decision about which path to follow next should be taken at any node but the terminal nodes. This decision has to be taken on the basis of the value of a decision attribute, here a property of an actor. During a relational matching operation an actor attribute is evaluated in each of the intermediate nodes of the inference tree. The outcome of this evaluation determines the path to be followed next. The total path leading to a terminal node represents a description of the actor.
2. *The set of demands:* A node in the tree can thus be associated with a certain actor attribute and a route through the tree, hence with a set of actor-attribute-value tuples. Once a terminal node is reached, all the necessary information on the demands an object has to fulfill has been collected, and the requirements for the object are known. This set of requirements constitutes the matching rule. Each of the matching rules is again represented as a disjunctive/conjunctive structure and is also modeled as a tree. Each of these combinations represented by a path through the tree and an associated matching rule forms a relational decision rule.
3. *The set of alternatives:* The third element of a relational match is the set of objects that match the demands associated with a certain actor-type and hence with a certain path through the inference tree. This matching set will be a subset of the initial set of objects, the ultimate cases being a subset that equals the initial set (all objects fulfill all demands), and the empty set (no objects fulfill the demands).

Figure 3.2 represents the inference tree of Figure 3.1, but now in terms of the three attributes described above. Of course there is an empty set of actor attributes associated with the root of the tree, whereas the set of alternatives is complete. After passing through the tree, the leaves of the



I,J: actor characteristics;
 D(I1,J1): requirements following from actor characteristics I1 and J1;
 U(I1,J1): the set of alternatives satisfying the requirements D(I1,J1).

Figure 3.2: Relational inference tree with actor characteristics, requirements, and set of alternatives

tree represent the sets of demands, the matching rule, and the objects that satisfy that rule. No further attributes are needed anymore.

3.2.2 Dimensions

It can be convenient to split up the relational definition of a concept—and therefore the relational inference tree representing this concept—into a set of other (sub)concepts or dimensions. Of course, logically there is no need to split up a relational concept into a set of dimensional sub-problems. The logical structure of a relational definition implies that we have in fact nothing but sets of disjunctions and conjunctions to be combined and added. No dimensional structure is required by the modeling procedure. Nevertheless, splitting up the problem in a set of dimensions can be very practical, especially when empirical application is concerned. People, even geographers, seem to order their knowledge in dimensional structures—big chunks of structured knowledge that are easy to deal with and which prevent the ‘user’ from accessing all the knowledge at once. This becomes apparent, for instance, when looking at the site suitability and production milieu modeling procedures discussed earlier. Both in the R.E.B. methodology and produc-

tion milieu matrix approach, but also in the matching approach by Schilling (1968), one departs from a set of general, abstract production factors such as 'labor market', 'accessibility', 'energy', 'service level', etc. These are then translated into sets of empirical indicators. This common starting point of a set of general and abstract production factors supports the hypothesis that researchers tend to think in dimensions; broad categories each of which covers a part of the total problem under consideration. It is important not to misunderstand this notion. For example, it does not after all imply that techniques such as factor analysis and principal component analysis, multi-dimensional scaling, and correspondence analysis are the obvious and appropriate tools with which to reveal the structure of the social world. Van der Smagt (1985) clearly shows that these kinds of techniques implicitly deny the relational nature of concepts. They might, perhaps, work on the level of the broad and general categories people deal with when trying to denote roughly what they are talking about. But this seems to be more than it actually signifies. Generally, one does not need intricate techniques to come up with a number of broad categories or dimensions describing a problem or concept. Moreover, although division into dimensions can be helpful in that it allows one to tackle a problem in a more or less modular manner, it does not add much to the deeper understanding of the problem itself. Division of an abstract concept into a set of other, abstract concepts, typically pertains to the functional definitions of those concepts. But functional definitions do not suffice as a model for the concept they denote, since they do not exemplify the conditions under which the object can fulfill the function expressed by the definition. Therefore, splitting up a complex problem into a set of smaller dimensional problems is necessarily limited to the functional level of concept definitions. It is, however, something which people seem to do quite easily. It can be expected that concurring with this dimensional structure in the modeling procedure offers some important advantages of a predominantly practical nature. It can facilitate the knowledge acquisition process, the process by means of which the empirical content of the inference trees has to be recovered. Building the inference trees can be a rather complex task. Being able to conduct that task in parts can make it a lot easier. If it is possible to split up a complex relational problem into less complex dimensional problems, then there is no reason to refrain from splitting up these dimensions into sub-dimensions again, and so on. The result will be a hierarchy of dimensions. An example of such a hierarchy is presented in Figure 3.3. The matching problem considers a relational site suitability assessment for a chemical industry. Of the five dimensions designated, 're-

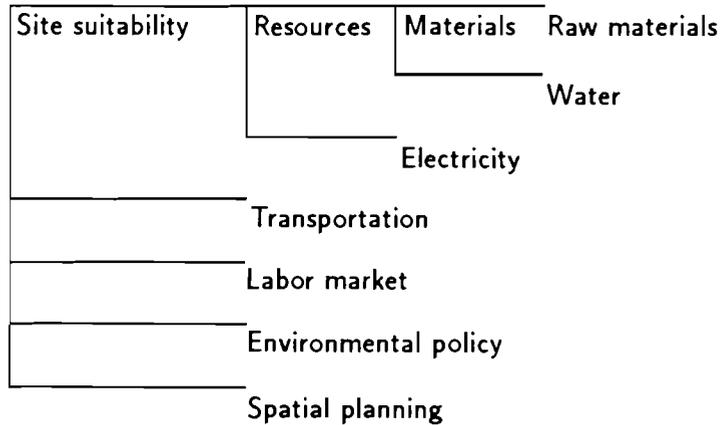


Figure 3.3: Hierarchical dimensional representation of a relational concept

sources' is split up in 'materials' and 'electricity'. The dimension 'materials' is again split up in the dimensions 'raw-materials' and 'water'. The result of this is a dimensional tree.

It must be recalled, however, that dimensions pertain to the functional level of a concept only. As such, each of the dimensional problems needs to be defined in terms of either sub-dimensions or, when no further sub-dimensions are needed, into a relational definition in the form of an inference tree such as the one in Figure 3.1.

Introduction of dimensions compels one to revise the formal structure of a relational definition a little. This is necessary because introducing dimensions implies an initial definition of a concept that is of a higher, more abstract level than that of the inference trees. The dimensional structure of a concept can therefore be regarded as a relational concept containing other relational concepts as part of its definition. This means that in order to solve the top matching problem (a conjunction) each of the nested problems (the dimensions or conjunctive terms) has to be solved first. The example in Figure 3.3 may illustrate this. In order to match for site suitability, the following conjunction has to be satisfied:

```

site suitability IF (
    resources                AND
    transportation           AND
    labor market             AND
    environmental policy     AND
  )
  
```

spatial planning
).

But since 'resources' is defined as a conjunction of 'materials' and 'electricity', and since 'materials' is again defined as a conjunction of 'raw materials' and 'water', the 'resources' condition has to be replaced by a nested conjunctive term:

```

site suitability IF (
    (
        (
            raw materials  AND
            water
        ) AND
        electricity
    ) AND
    transportation      AND
    labor market        AND
    environmental policy AND
    spatial planning
).
```

And since this constitutes a purely conjunctive structure, the nesting of the conditions (the sub-dimensions) can be eliminated:

```

site suitability IF (
    raw materials      AND
    water              AND
    electricity        AND
    transportation     AND
    labor market       AND
    environmental policy AND
    spatial planning
).
```

Each of the leaves of the dimensional tree—and therefore each of the conjunctive terms in the non-nested conjunction—refers to a relational inference tree. This makes the dimensional structure of a concept constitute a conjunction of dimension-specific inference trees.

3.2.3 Internally related variables

The possibility that variables can be internally related was one of the main reasons for developing relational methodology. The example of the (not)disabled person and the ways in which apartments in an apartment building can be categorized in terms of the availability of an elevator and the floor the apartment is on (Figure 1.1), clearly reveals the importance of being able to include internally related variables in the model. Note that this type of inter-relatedness does not pertain to the requirements or means only. Similar dependencies in categorizing variables may occur in the actor attributes. The way an actor attribute is categorized may depend on where this attribute is located in the inference tree, i.e., which other attribute-value combinations are contained in the path leading to it. Of course, internally related activity attributes are represented by different categorizations of these attributes in the inference tree. Different criteria, in other words, for following the next path from the attribute node. Requirement related internal relations, however, can also easily be incorporated. They can be represented by associating different requirements with different terminal nodes of an inference tree. The (not)disabled-elevator-floor example can be translated into the inference tree displayed in Figure 3.4.

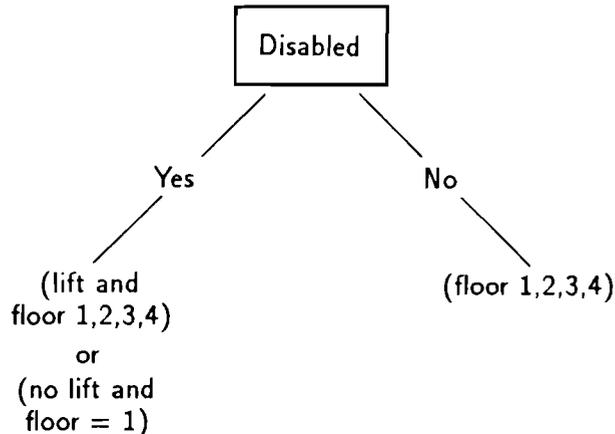


Figure 3.4: Categorization of the floor attribute by means of a relational inference tree

One could argue that internal relationships as present in the disabled-floor-elevator example, occur only rarely. Much more frequently, the variables contained in the different matching demands will be of an entirely

different nature, without a common set of categories or a common basis for categorization. For representation in a relational inference tree, this does not make a real difference. Basically, the different sets of demands represented by the terminal nodes of an inference tree are fully independent of each other. Neither the data base representing the trees nor the program processing them takes into consideration how the variables are categorized. It is the researcher who develops the trees who should take care in applying the appropriate classes and categories.

3.2.4 Internally related dimensions

Although in a rather different manner, dimensions too can be internally related. Once a node in an inference tree has been processed, this will have consequences for the way another tree can be processed. Dimensions are internally related if, once a specific route in an inference tree has been traversed, the number of opportunities to satisfy another, related tree decreases. The second inference tree is pruned; the search space for solving the matching problem decreases. To illustrate this, suppose that besides the inference tree for R1 (Figure 3.1) representing the first dimension of a relational concept, we have another inference tree for the second dimension R2 (Figure 3.5). Now, suppose that R1 gets K-instantiated, meaning that actor attribute scores I1 and J1 were set. Clearly, this must have implications for the way the tree for R2 can be instantiated since this tree contains the actor attribute I as well; since I1 is already known, the tree for R2 is reduced to the one in Figure 3.6.

Dimensions can be internally related, but they do not have to be and it can be interesting to inspect the model results for one dimension only. One might even want to investigate the model results of successive matchings for individual dimensions when compared with a matching for various dimensions simultaneously. This comparison of various one-dimensional matching solutions can become important when dimensions are considered to represent the activity's strategies or objectives. In that case, one might want to know whether or not various dimension-specific solutions are compatible. In other words, does a solution-set for one dimension or objective contain elements which are contrary to another dimension or objective? Or to which degree do solutions for different objectives correspond to each other? In case a matching exercise is conducted on more than one dimension, the total matching result is generated by a series of consecutive, possibly choice-set limiting, dimension-specific matchings. Each dimension represents an independent

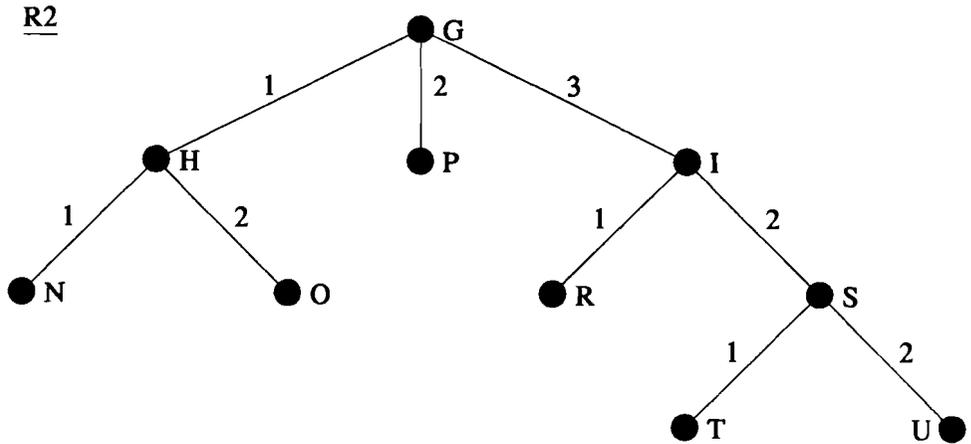


Figure 3.5: Relational inference tree for variable R2

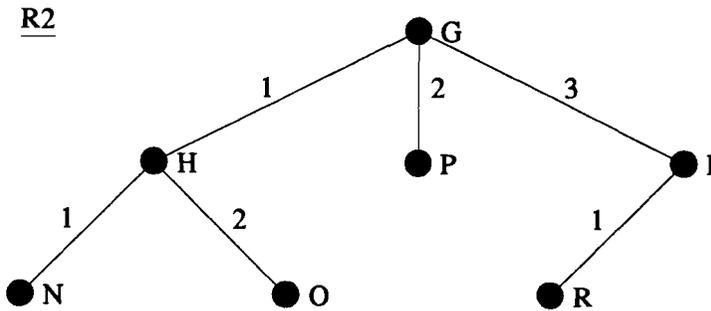


Figure 3.6: Relational inference tree for R2 given a k-instantiated tree for R1

matching sequence and has therefore its own, independent impact on the total matching result. From the total matching result, however, one cannot draw conclusions on the impact each of the individual dimensions had (unless, of course, the matching contained only one dimension). By comparing and inspecting the dimension-specific matching patterns, however, one can obtain an impression of the discriminatory power of each of the individual dimensions. By using these set-theoretic operations on dimension-specific matching results such as their unions, intersections, and differences, one can acquire a much more detailed overview of the discriminating power of individual dimensions or combinations of dimensions.

3.2.5 Dimensions: additivity and the partitioning problem

An objection one could raise against modeling a concept such as site suitability using a dimensional approach, is that it implies that additivity is brought back into the model. Each of the dimensions is a term in a conjunction and can be independently evaluated, thereby having its own, independent contribution to the value of the classification of the object. Two remarks can be made against such an objection. First of all, a dimensional structure in the form of a logical conjunction is not really additive because the non-compensatory character of the model is retained. Once a term (a dimension) in the conjunction fails to be satisfied, the entire conjunction fails. This is an important difference when compared to the traditional form of additivity as present in, for instance, a weighted additive combination rule. Further, it should be noted that additivity on the dimensional level does not imply additivity on the deeper level of the relational definition. Dimensions split up an abstract problem into a set of just as abstract sub-problems. Each of the dimensions must, in the end, be defined relationally by means of an inference tree. Only there does the real definition appear, and it is there that non-additivity and internal relations are added. Internal relations on the dimensional level only pertain to the configuration of the search space, i.e., the number of possibilities to traverse a tree. Internal relations between variables, however, are limited to the inference trees. And since inference trees form the leaves of the dimensional trees, this kind of internal relation is retained.

However, a more serious, and as yet unresolved, problem remains. Splitting up a complex problem into a set of smaller problems presupposes the possibility of attributing every aspect of the problem to one and only one sub-problem. It is like dividing an n -dimensional space into various n -dimensional regions such that every point in that space belongs to one and only one region. Designating the dimensional structure of a concept, therefore, presupposes that it is possible to 'partition' a concept such as site suitability into a fixed set of dimensions, each of which covers a distinct part of the problem. The total of the dimensions then covers the entire problem. This seems rather unrealistic. Certainly, the abstract production factors that are often mentioned when dealing with site suitability or production milieu show that certain different types of aspects can be recognized. But for some aspects of the problem it can be difficult to decide to which dimension they belong. Should the transport requirements for coal be part of the energy or the material dimension? And if coal is used for both fueling and

as a raw material, to which dimension(s) do the coal related requirements belong? Should 'coal' perhaps be a dimension on its own? Splitting up a large relational matching problem into a set of smaller dimensional ones thus generates the problem of choice of dimensions, and of how to decide what exactly they represent. These decisions are by nature *ad hoc* and more or less arbitrary and external to the model. This can cause confusion and can make various models difficult to compare. Davis *et al.* (1987; pp. 247–248) mention several possible solutions for situations in which fixed partitioning of the problem space is impossible. A first approach is that the problem is conceptualized at a number of levels and the problem is then completely solved at each level, going from the general to the more specific (Sacerdoti, 1974). Clearly, this is not of much use here as the problem of partitioning, as discussed here, pertains to partitioning at a very general level. Moreover, this approach is more like a statement of the problem than a solution to it. Conceptualization at different levels is what would be most attractive, but the question is just how can this be done? A second approach involves the principle of 'least commitment' (Stefik, 1981; Wilensky, 1983) which implies that decisions within a sub-problem are only taken when enough information is available. Yet another possibility for tackling the partitioning problem is that earlier decisions can be retracted when they later on turn out wrong or become unattractive. Implementation of such an approach, however, requires the availability of non-monotonic logic (McDermott and Doyle, 1980); it is therefore not widespread. The second and third strategy for solving the partitioning problem clearly represent a very distinct view of dimensions. They are typically decision-process oriented whereas the dimensions as suggested here are based much more on a conceptualization of a specific situation. The possibilities these alternative approaches offer for tackling the partitioning problem in a relational matching model were not investigated further. It was nevertheless deemed preferable to work with a dimensional structure because of its many practical advantages. It does not seem realistic to expect that many matching problems have their 'natural' dimensions attached. The best that can be achieved at this stage is careful documentation of what is represented by which dimension, and a transparent automated matching process that can explain its own behavior. Self-explanatory features will show which aspects are and which are not covered by certain dimensions.

3.3 Relational inference trees versus decision plan nets

It seems appropriate to pay some attention here to the relationships between the method of relational inference trees and an alternative, non-algebraic method for modeling choice behavior, namely, decision plan nets (Park *et al.*, 1981; Op 't Veld *et al.*, 1986, 1987; Timmermans and van der Heijden, 1987; Op 't Veld, 1988). A decision plan net constitutes a model of an actor's decision process represented in the form of a heuristic kind of network structure. With regard to the claim that a decision plan net models the choice process, it must be mentioned that it is assumed that specific stages or aspects of the decision-making process, such as the initial stages of information gathering and exploration of alternatives, are ignored. As with an approach based on matching, the ideas of modeling goal-oriented behavior by means of a heuristic network structure have been already explored. During the sixties, proposals for such an approach had been put forward (Miller *et al.*, 1960; Bettman, 1970). Due to the problems associated with preference modeling by means of algebraic models however, interest in this method has increased recently.

Figure 3.7 shows the structure of a decision plan net. The network contains two types of choice-alternative attributes; primary and secondary attributes. The primary ones are considered to be evaluated first by the decision maker. Only if the required values and combinations of values of these attributes cannot be satisfied are the secondary ones inspected. Primary alternatives are ordered from right to left corresponding to increasing order of importance.

In the context of a discussion on relational inference trees, the question as to how such trees relate to decision plan nets is an interesting one. I think that there are a few important differences, and would like to argue that relational inference trees offer more, and better, opportunities for modeling spatial choice based on functional equivalence. Van der Smagt and Lucardie (1990; p. 10) mention some objections to the method of decision plan nets. An important problem of decision plan nets concerns their dichotomous nature. The tree in Figure 3.7, for instance, is of a typically binary character. Attributes are only organized into dichotomous categories on which one can score 'yes' or 'no'. This implies that one either has to restrict oneself to dichotomous categorizations of variables, something which is rather unattractive, or one turns every category of a polytomous variable into a separate

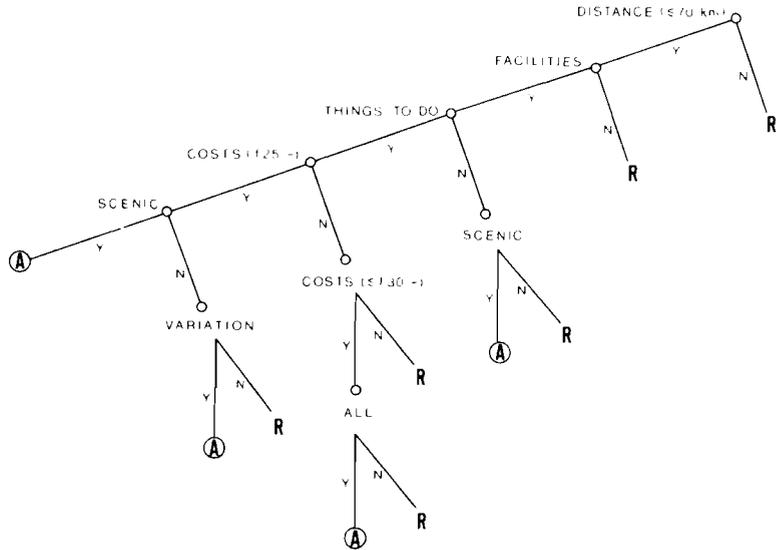


Figure 3.7: Example of a decision plan net (Source: Timmermans and van der Heijden, 1987)

dichotomous variable. The latter option, however, implies a very unfortunate approach and very large and complex decision plan nets. Another objection van der Smagt and Lucardie mention is the possible ‘spreading’ of attributes around the decision plan net. Later on in this chapter it is shown that a relational inference tree or decision table perspective offers ample opportunities for ‘optimizing’ the classification procedure into networks that are as sparse as possible. For decision plan nets, this kind of optimization is a lot more difficult and a lot harder to realize. A third, and by far the most serious, problem concerns the way several authors propose decision plan net applications (Op ’t Veld *et al.*, 1986, 1987; Op ’t Veld, 1988). In these proposals, decision plan nets become representations of very individual, very actor-specific decision processes. This, however, brings us right back to the problems associated with firm-specific location decision studies discussed in Section 2.5.5. There it was argued that a collection of individual cases (here a collection of a set of actor-specific decision plan nets) is basically unproblematic, but that such a collection must be integrated into a more structural, more general model of actor-specific functional equivalence. It was argued that only if individual cases can be recognized as instances of more general, more structural types or mechanisms, real explanatory information is added. If the individual requirements, as represented in the decision plan net, are

not connected to some of the actor's objectives and characteristics that give rise to these requirements, what could conceivably be the advantage? It was argued that the ability to recognize a case as (dis)similar to another one, presupposes the availability of a set of criteria by means of which this (dis)similarity can be assessed; absence of these criteria leaves one with a collection of individual cases devoid of any cross connections, regularity or whatever kind of systematic. If one disregards the problems caused by the binary character of decision plan nets for a moment, one might say that decision plan nets might be compared to the matching rules contained in the inference trees. They represent conjunctive–disjunctive sets of requirements. However, in a relational model, specific matching rules are the result of comparing an actor's objectives and characteristics with the ones declared in the other, relational part of the inference tree. In the proposed use of decision plan nets for modeling choice behavior, this linking of requirements to objectives and characteristics has yet to be executed.

3.4 Model content: formal evaluation procedures

Inference trees are models of decision procedures. Applied to the relational definition of concepts, they become a scheme and procedure, specifying how actor characteristics must be linked to sets of object properties in order to arrive at a relational match. Inference trees can also be regarded as formal representations of decision and classification procedures. On the basis of a set of criteria (combinations of actor properties) an action is conducted (matching with a particular matching rule). Each of the combinations of conditions and the associated matching rule denotes a decision rule. In building inference trees, various formal quality requirements have to be obeyed. And since an inference tree constitutes a purely logical structure, the formal criteria are of an entirely logical nature. Formal evaluation of a logical classification procedure in the form of an inference tree involves three criteria: consistency and completeness (Hofstadter, 1979; Sinth, 1985; Verhelst, 1980; Nguyen *et al.*, 1985; Finin, 1986), and optimization (Verhelst, 1980; Quinlan, 1979, 1983; Thompson and Thompson, 1986; Arbab and Michie, 1985, 1987). An inference tree such as the one in Figure 3.1 can be viewed in different ways. One possibility is the decision table perspective (Montalbano, 1974; Verhelst, 1980; Hendriks, 1986; Reilly *et al.*, 1987; Lucardie, 1988b). A decision table represents a set of actions, each of which is a function of a distinct set of conditions. These actions may represent single actions like a single-valued

Table 3.1: Inference tree for R1 in decision table format

R1			
I	1		2
J	1	2	-
Action	K	L	M

conclusion, or a reference to another decision problem, a set of actions to be taken, and so forth. In case of a relational inference tree, the actions represent the matching rules, the terminal nodes of the tree. As mentioned earlier, these represent decision procedures themselves, which can again be modeled as decision tables. Decision tables can be used for representing inference tree classification procedures. If we denote the matching rules K, L and M of Figure 3.1 the actions, then the conditions for these actions are formed by the paths leading to them, i.e., {I1,J1}, {I1,J2}, and {I2}. As was mentioned earlier, each of the matching rules themselves, being a conjunctive and disjunctive structure, can be considered a decision table too, albeit rather special ones. They contain only one action 'succeed', which is a function of different combinations of object characteristics. The decision table corresponding to the inference tree of Figure 3.1 is shown in Table 3.1.

Many classification procedures can be displayed in this fashion. Suppose, for example, the hypothetical classification scheme regarding the site suitability of a location for a given type of company is as in Table 3.2:

Table 3.2: Hypothetical inference tree

Volume	Coal	Cooling	Matching rule
Large	No	Water	B
Medium	Yes	Water	A
Medium	No	Air	B
Large	No	Air	B
Small	No	Air	A
Small	No	Water	A
Medium	No	Water	B
Small	Yes	Air	A
Medium	Yes	Air	A
Large	Yes	Air	B

If the matching rule to be applied is taken as the action variable, the following decision table can be constructed (Table 3.3):

Table 3.3: Decision table of the classification in Table 3.2

Volume	Small			Medium				Large		
Cooling	Water		Air	Water		Air		Water		Air
Coal	N	Y	N	Y	N	Y	N	N	Y	N
Matching rule	A	A	A	A	B	A	B	B	B	B
	0	1	2	3	4	5	6	7	8	9

3.4.1 Completeness and consistency

A decision table, and hence an inference tree, is logically correct if it is consistent and complete. Logical completeness is a special form of completeness. It differs, for instance, from empirical or effective completeness. A classification procedure is effectively complete if it contains all relevant and possible combinations of conditions and all possible actions. In case of the inference tree/decision table in Table 3.1 this would mean that the tree as it is presented is the actual tree for a generic type of actor. It is effectively or empirically complete if no other combinations of actor attributes and no other combinations of object characteristics can occur. All possible cases to be found in the world are covered. Another type of completeness is logical completeness. Verhelst (1980) and Nguyen *et al.* (1985) mention what this means in the context of a decision table or classification procedure. In a logically complete decision table not a single action is unreachable and each and every possible combination of conditions leads to at least one action. An unreachable action is an action that is present in the decision table, but for which no conditions are specified. In the context of an inference tree this means that the tree contains a matching rule but no path leading to it. This may seem a little odd, since a tree consists of branches that terminate in leaves. Hence, a leaf cannot exist by itself. It can nevertheless occur when classification procedures are specified for situations when the actions are known *a priori*, but for which the combinations of conditions still have to be specified. This is something that frequently occurs in text-based regulations, diagnostics and instruction manuals. Nguyen *et al.* (1985) denote this problem as the ‘missing rule’ problem. The second type of incompleteness, missing actions, is much easier to trace. A glance at Table 3.3 shows

that no actions are specified for the cases {large volume,coal,water cooling} and {small volume,coal,water cooling}. Of course, this was clear from the beginning since a trichotomous variable 'volume', and dichotomous variables 'coal' and 'cooling' give rise to $3 \times 2 \times 2 = 12$ possible combinations, whereas Table 3.2 only contains ten. But in many instances of classification procedures, especially when they are in the form of a more or less complex text, lacking actions only become evident when a decision table is constructed (Verhelst, 1980).

An adequate classification procedure is not only logically and empirically complete, it is also internally consistent. A classification procedure can be considered inconsistent if at least one combination of conditions points to more than one class; "two 'positions' belonging to different classes share the same description" (Quinlan, 1979; p. 170). Neither Table 3.2 nor Table 3.3 contain such an inconsistency, but as Verhelst (1980) illustrates, many regulations and instruction booklets do. A somewhat more subtle form of inconsistency that is important for inference trees is the problem of subsumed rules. This occurs if two sets of conditions point to the same action, but one of the sets is a subset of the other. The following two (hypothetical) matching rules, for example, show this type of inconsistency:

- | | | | |
|----|---------------------|------------------------|-----|
| 1) | site suitability IF | urban area | AND |
| | | railway available | AND |
| | | international airport | AND |
| | | subcontractors present | |
| | | | |
| 2) | site suitability IF | subcontractors present | AND |
| | | railway available | AND |
| | | urban area. | |

Clearly, according to the second rule the presence of an international airport is not required, whereas the first rule declares that it is. These two types of inconsistency are important in the context of inference trees. Presenting them in the form of a decision table can help to discover and eliminate them. Nguyen *et al.* (1985), mention two more forms of inconsistency: circular rules and redundant rules. Circular rules are rules that somehow refer to themselves, either directly or indirectly via one or more other rules. Redundant rules are rules that signify identical condition-action combinations.

3.4.2 Optimization

Although circularity and redundancy are not really relevant in securing the consistency of relational inference trees, redundancy constitutes an important issue in the matter of optimizing the trees. Concerning the adequacy of (automated) procedures for the classification of events or situations, Quinlan (1979; p. 170) remarks that, “Ideally the properties to characterize them should be adequate for the task, give rise to substantial compression of the data base, be readily computable, and lend themselves to the development of concise rules”. The degree to which a classification procedure can be considered optimal can be expressed in many ways. One of the possibilities is a measure of efficiency in terms of minimizing a mathematical function describing the average time needed for processing the classification (Verhelst, 1980; pp. 159, 160, 163). The processing of a classification scheme in the form of a non-optimal decision table can be enhanced by two types of operations: rule collapsing and changing the order of the conditions. Adjacent rules can be collapsed if they point to an identical action, and if this common action can be attributed to a common condition. Adjacent groups of rules can be collapsed if the groups show identical patterns of actions, and if these patterns can be attributed to a common condition. In Table 3.3, for example, rules seven, eight and nine have identical actions, to be attributed to the common condition of ‘large volume’. Once it is known that volume is large, the matching rule will always be B. Similarly, rules zero, one and two, each have an associated matching rule A, to be attributed to the common condition ‘small volume’. That collapsing an incomplete table implies some risk for the adequacy of the remaining collapsed table is explained below. Collapsing adjacent rules in Table 3.3 results in a reduced, smaller Table 3.4.

Table 3.4: Reduced version of Table 3.3 by collapsing

Volume	Small	Medium				Large
Cooling	–	Water		Air		–
Coal	–	Y	N	Y	N	–
Matching rule	A	A	B	A	B	B
	0	1	2	3	4	5

Table 3.4 is smaller than Table 3.3, not only because the number of rules has been reduced from ten to six, but also because for both rules zero and five, only one condition needs to be evaluated before a conclusion or action

Table 3.5: Optimal decision table

Volume	Small	Medium	Large	
Coal	–	Y	N	–
Matching rule	A	A	B	B
	0	1	2	3

is reached. Notwithstanding this significant reduction of the original table, Table 3.4 still contains some redundancy. Rules one and two, and three and four, show an identical action pattern (A,B), which can be attributed to ‘medium volume and using/not using coal’. This implies that the ‘cooling’ variable never needs to be tested. The resultant decision table is shown in Table 3.5. The table has now been reduced to only 4 rules with only a maximum of two conditions to be tested.

Now, removing a condition just like that is a rather inelegant way of reducing a decision table into a sparser one. This could have been avoided if the order of the conditions in the original table (Table 3.3) had been different. Suppose, for instance, that ‘coal’ and ‘cooling’ had been entered in reverse order. In that case, a simple collapse over adjacent rules and rule groups would have resulted in the final output as in Table 3.5.

3.4.2.1 Optimizing with information entropy: ID3

The optimization of decision tables can be automated by developing appropriate algorithms that do the job. Although various solutions have been suggested, the approaches based on the concept of information entropy seem particularly useful for the optimization of inference trees (Shwayder, 1971, 1974; Ganapathy and Rajaraman, 1974; Quinlan, 1979, 1983; Arbab and Michie, 1985, 1987; Thompson and Thompson, 1986; Hart 1986). Quinlan (1979, 1983) applied the concept in the development of the ID3 (Iterative Dichotomizer 3) algorithm. Quinlan and Thompson and Thompson (1986) advocate an inductive use of the technique. This implies that the technique can be used for constructing (sub)optimal decision trees out of a raw set of rules. ID3 removes the redundancy from the rules and delivers the associated (sub)optimal decision tree, representing the essential content of the raw set of rules (refer also to Hart (1986) for some illustrations of the technique). As stated, ID3 is based on the concept of information entropy. In the context of optimizing a classification procedure such as an inference tree, information

entropy constitutes a measure of the uncertainty of how a situation must be classified. A measure, in other words, of the uncertainty as to which class of a given set of classes an object, situation, or event belongs. In an inference tree we deal with actor characteristics. On the basis of specific combinations of actor attributes, a decision is taken as to which matching rule is to be applied. Information entropy is a measure for the uncertainty as to which matching rule has to be applied, given an initial actor, say a firm. The more is known about the actor, the less uncertainty exists as to which matching rule applies. This entropy or uncertainty as to which class a situation belongs can be calculated as:

$$H(C) = -\sum P(C_i) * \text{Log}_2(P(C_i)) \quad (I = 1, N) \quad (3.1)$$

where H(C): entropy of classification;
 N: number of classes;
 C_i: class i;
 P(C_i): probability that C_i is the correct class.

The inference tree in Figure 3.1 illustrates the meaning of information entropy. If the only available information about the actor is the fact that one of the matching rules {K,L,M} applies, each of them has a $1/N = 0.333$ probability of being the correct one. The associated entropy calculated as in (1) is 1.588. In instances where each of the classes has an equal probability of being the correct one, formula 1 can be rewritten as:

$$- N * (1/N * \text{Log}_2(1/N)) = -\text{Log}_2(1/N). \quad (3.2)$$

In a case like this, the entropy is a simple logarithmic function of the number of classes. Under *ceteris paribus* conditions, more classes implies more uncertainty, hence higher entropies. Adding information about the actor reduces the uncertainty. If, for instance, it is known that the actor scores '1' on I, the probability for matching rule M in Figure 3.1 becomes zero, whereas the probabilities for rules K and L become 0.5. The associated entropy becomes 1.0. The entropy decreases because the classification problem has been reduced to two classes, as a consequence of the information that the actor scores a '1' on property I.

The ID3 algorithm uses this measure of entropy to compute a (sub)optimal decision table. This means that it performs rule collapsing and condition order changing. From the point of view of information entropy, the condition that reduces the uncertainty of the classification most should be the first in

the decision table. The next attribute should be that which reduces entropy most for the remaining classification problem. What therefore has to be computed is the entropy of the (remaining) classification problem, given a current attribute to be tested. If for each attribute, each condition is tested, the attribute that can be associated with the smallest entropy will be taken as the next attribute in the decision table. The conditional entropy, i.e., the entropy given a partitioning attribute, is computed as:

$$H(C|A) = \sum P(A_j) * H(C|A_j) \quad (3.3)$$

where $H(C|A)$: entropy after partitioning on attribute A;
 $P(A_j)$: probability that attribute A has value j;
 $H(C|A_j)$: the entropy of the classification problem remaining after partitioning on attribute A.

The entropy of the remaining classification problem— $H(C|A_j)$ —can be calculated as:

$$H(C|A_j) = -\sum P(C_i|A_j) * \text{Log}_2(P(C_i|A_j)) \quad (3.4)$$

where $P(C_i|A_j)$: probability that the correct class is C_i , given a value j on attribute A.

Note that the conditional entropy $H(C|A)$ is computed as the average of the entropy for each value j of the attribute, weighted with the probability of that value. Application of the algorithm comes down to a recursive procedure in which after a partitioning attribute has been selected, each of the resultant classification problems for each of the values of the partitioning attribute becomes subject to a new application of the algorithm. Rule collapsing is implicit in this method. Once the entropy in one of the (sub)classification problems is reduced to zero, everything to be known is indeed known, and no more partitions are required.

To illustrate the operation of the algorithm, the examples in Table 3.2 can be interpreted as rules:

```

IF (
    volume = large AND
    coal = yes      AND
    cooling = air
)

```

```

    THEN      (matching rule = B)
and:
    IF   (
           volume = small AND
           coal = no      AND
           cooling = air
        )
    THEN      (matching rule = A)

```

After randomizing both the order of the rules and the order of the conditions inside the rules, and after submitting the complete set of rules to a (Prolog) version of ID3, the resultant set of rules is shown in Table 3.6. As can be seen, it is equivalent to Table 3.5.

Table 3.6: Classification as a set of rules and nested list, optimized by ID3

RULES
rule(b) :- volume(large). rule(b) :- volume(medium),coal(no). rule(a) :- volume(medium),coal(yes). rule(a) :- volume(small).
TREE
[[volume(large),rule(b)], [volume(medium),[[coal(no),rule(b)], [coal(yes),rule(a)]]], [volume(small),rule(a)]]

In addition to a new, optimal set of rules, Table 3.6 also shows the resultant inference tree in a nested Prolog list representation. A Prolog list is a set of elements, contained in rectangular brackets, with the elements separated by commas. A nested list is a list which contains one or more other lists as its elements. As shown in Table 3.6, a tree can easily be represented by a nested list.

3.4.2.2 Pitfalls: inconsistency and incompleteness

ID3 is a nice and very useful algorithm for removing redundancy from a complex inference tree. When applying the algorithm, however, one has to realize

that it does precisely that and nothing more. ID3 will attempt to optimize whatever tree it gets, regardless of its (in)consistency and (in)completeness. It can be shown, for instance, that out of an incomplete table, ID3 can produce a complete table that will pass any completeness test. The reason for this is that ID3, like optimization by hand, ‘assumes’ that the table submitted is already a complete table. Originally missing, but logically possible combinations of conditions are neglected as if they are irrelevant or impossible. Similarly, if specific cases are not declared ‘missing’ or ‘unknown’ in the original table, ID3 will generate logically complete versions from logically and empirically incomplete input tables.

Another problem that should be taken into consideration is that there may be a few cases that ID3 is unable to optimize. One of them is the following:

```
IF type(urban) THEN suitability(yes)
IF type(rural) THEN suitability(yes)
```

ID3 does not change this small rule set into the even smaller set:

```
suitability(yes),
```

meaning that `suitability(yes)` will always occur: an action without conditions, therefore. Now, from a purely logical and set-theoretic point of view this might not appear appropriate, but there are instances in which it is gratifying that ID3 leaves these kinds of rule sets unaltered. Take, for instance, the following set of consistent rules:

```
IF type(urban) THEN suitability(yes)
IF type(rural) THEN suitability(yes)
IF railway(yes) THEN suitability(yes)
```

In case ID3 did collapse the first two rules, the optimized result would be inconsistent. Whereas the first (collapsed) rule declares that site suitability is always ‘yes’, the second rule formulates a condition for suitability. This second rule is subsumed by the first one.

3.4.2.3 Single-action rule sets

As mentioned earlier, matching rules themselves can be modeled as decision tables with only one action ‘yes’ or ‘true’. They describe the conjunctive/disjunctive set of demands an object has to fulfill for a certain actor.

Table 3.7: Decision table of classification only containing rules leading to a matching rule A

Volume	Small			Medium	
	Water	Air		Water	Air
Coal	N	Y	N	Y	Y
Matching rule	A	A	A	A	A
	0	1	2	3	4

Table 3.8: Classification including 'unknown' cases

RULES
<pre>rule(unknown) :- volume(large). rule(a) :- volume(medium). rule(a) :- volume(small),cooling(water),coal(no). rule(unknown) :- volume(small),cooling(water),coal(yes). rule(a) :- volume(small),cooling(air).</pre>
TREE
<pre>[[volume(large),rule(unknown)], [volume(medium),rule(a)], [volume(small),[[cooling(water),[[coal(no),rule(a)], [coal(yes),rule(unknown)]]], [cooling(air),rule(a)]]].</pre>

As such, a matching rule only describes how the demands can be fulfilled; it only describes the demands for a positive outcome. If an object fails to meet one of them it is eliminated from the choice set. ID3 can optimize single-action rule sets just as well as it handles multiple-action rule sets. Table 3.7 contains that part of the decision table of Table 3.3 which has only 'A' matching rules as actions. Table 3.8 shows the results after processing by ID3 after the unknown cases are incorporated into the original rule set.

3.4.2.4 ID3: conclusion

It can be concluded that ID3 is an attractive and powerful tool for optimizing decision tables. Inference trees for relational modeling are closely connected to decision tables and rule-based classification procedures. Therefore, ID3 can be considered an adequate tool for helping to build them. Of course, ID3

is only an optimization procedure. It removes redundancy from a classification procedure by altering the order of conditions and by collapsing adjacent rules. Submitting incomplete and/or inconsistent inference trees to ID3 can therefore be dangerous. Certain types of inconsistency are 'detected' by ID3, because the algorithm terminates when, for instance, attributes are exhausted before the tree is complete (Hart, 1986; p. 117). Other inconsistencies, however, are invisible for ID3 and are therefore retained. Inconsistent rule sets will then be converted into inconsistent optimal versions. This is something that should be prevented by checking the consistency of the input inference tree. Logically incomplete input trees are more hazardous for the quality of the resultant rule base/inference tree. Whereas inconsistent trees might yield inconsistent results that can always be tested retrospectively, logically incomplete inference trees can be processed into logically complete optimal versions. This makes posterior completeness checking impossible. Input trees should therefore always be checked on completeness.

One can think of many methods and approaches for developing empirical inference trees when considering questions such as what completeness and consistency checking entail in the much broader context of building inference trees with a particular empirical content, and where these tests fit in and how they can be applied empirically. Of course, the examples presented before are rather simple and unlikely ever to occur. More complex and more realistic relational models in the form of inference trees can be built either by hand, or with the help of a computer. An inference tree is a strictly logical structure so it is possible to construct a computer program that reads in combinations of actor attributes and matching rules and then converts them into inference trees. The researcher then formulates the empirical content of the rules and trees, the computer treats them like chains of symbols with a specific logical structure. Such a partly automated process of knowledge acquisition would greatly facilitate consistency and completeness checking. Inconsistent rule sets would be discovered by the program and presented to the researcher who can then decide how the inconsistency is to be resolved, after which the program can check again. Incompleteness can be treated similarly. The only difference here is that although a program can identify incompleteness, it cannot make a decision concerning the relevance of the missing cases. Whether or not a case is 'legally' omitted is something the researcher has to decide. These cases can then be marked by the program and treated separately in further modifications of the tree.

3.5 Actor attributes and requirements in the inference tree: some critical remarks

Although the formalism of inference trees as presented here seems appropriate for representing the relational definition of concepts, it suffers from a problem which is strongly associated with the problem of optimizing decision tables, as discussed above. This problem stems from the fact that both the inference tree and the associated matching-rules are treated as separate trees, linked together by the terminal node or leaf of the inference tree. Each of the paths through the inference tree describes an actor type as a set of actor-attribute-value tuples. Individual matching rules, the disjunctions of conjunctive requirements, are associated with each of these paths. In other words, the set of requirements is always associated with an actor type, a complete set of actor-attribute-value tuples. This, however, causes two kinds of problems: a data redundancy problem and a modeling problem.

3.5.1 The redundancy problem

These problems are illustrated in Figure 3.1. Now suppose that actor-attribute-value tuple I1, independent of other actor-attribute-value combinations, causes a demand Dx. Clearly, in the current situation Dx is part of both the matching rules associated with nodes K and L. But this implies a kind of redundancy, because if Dx could be associated with just I1 rather than the tuple combinations of which I1 is a member, the demand only needs to be represented and stored once. This problem is identical with the situation of a decision table in which more than one action is associated with one combination of conditions, and where a subset of the actions in the action set is a function of a subset of the conditions associated with this action set. This problem is not solved by an algorithm like ID3, since ID3 only handles one-class-only classification problems. The reason that ID3 can be used for optimizing inference trees, is that the ‘actions’ of the inference tree are complete matching rules, i.e., single actions, matching-rule 1, 2, or n. In case one regards the matching rule as a set of actions however, the situation changes, and ID3 can no longer handle this kind of problem. The question then is whether or not such a kind of redundancy can be treated in decision tables altogether. In order to investigate this problem a little further, the inference tree in Figure 3.1, can be used again with three new matching rules K,L and M. The resulting decision table is shown in Table 3.9. The table

Table 3.9: Table with multiple actions; no dependency specified

R1			
I	1	2	
J	1	2	-
Action	K	L	M
	A	A	D
	B	C	

Table 3.10: Dependency specification by means of multiple tables

R1			R2		
I	1	2	J	1	2
Action	A	D	Action	B	C
	R2				

shows that the action A, which is a part of both the action sets K and L, is stored twice.

To solve this redundancy in the decision table, a sub-table may be used. This implies that instead of using one table in which the complete action sets are represented, there is a main table which contains as an action a reference to another table in which the actual action is specified. Table 3.10 shows the result for the decision table in Table 3.9. Now action A is exclusively associated with I1, while B and C are associated with the different values of J; the redundancy is removed.

3.5.2 The modeling problem

The second and more serious problem concerning the way actor attributes and requirements are associated involves the structure of the inference tree itself. Again, the problem can be illustrated by taking a look at Table 3.9. Both the action sets K and L contain an action A. This could be caused by the common condition I1. However, another possibility is that both A and B in rule K are generated simultaneously by the combination (I1,J1), whereas A and C in rule L are generated by (I1,J2). Although A would still be a common action of both K and L, the table does not contain any redundancy, and an operation such as in Table 3.10 could not be conducted.

Table 3.11: Table with multiple actions; no dependencies specified

R1				
H	1		2	
I	1	2	-	
J	1	2	-	-
Action	K	L	M	N
	A	A	D	E
	B	C		

Table 3.11 illustrates this point. Compared with Table 3.9, Table 3.11 shows a slightly more complicated decision procedure. Attribute I is no longer the first attribute to be tested, instead it is only examined after H1 is established.

In this case, matching-rules K and L not only have the condition I1 in common, but also H1. As a consequence, action A can be the result of the common conditions I1 or H1 or any of their combinations, or it is the result of the entire set of conditions. The crucial point here is that in order to be able to decide whether or not the table contains redundancy, external knowledge, i.e., knowledge that is not contained in the decision table itself, is required. What does this imply for the formalism of relational inference trees presented earlier? The decision tables in Table 3.9 and Table 3.11 are good representations of the way the inference tree with the actor attributes and the disjunctive sets of requirements are coupled. Because of the fact that entire requirement sets are coupled with complete paths through the inference tree, no information about which attributes or combinations of attributes generate which requirements is retained. This means that not only do various matching rules contain identical parts, but the external knowledge which specifies how these requirements became part of these sets in the first place is not retained. The functional relations between actor attributes and requirements remain concealed and once both the inference tree and the matching rules are constructed, they are lost altogether. All that remains is an actor type associated with a set of demands. This, however, can cause serious problems, for instance, in case the model needs to be modified. Since from the inference tree it cannot be deduced which attributes or combinations of attributes generate which requirements, it is difficult to infer the consequences of local changes in the model for the rest of the model. What, for instance, should happen if it is realized that I1 should lead to E. If it is

known that I1 was originally declared to lead to A, A can simply be replaced by E. But if the relation between I1 and A is lost altogether, how should the matching rules change after it gets established that I1 leads to E? For these reasons it is important, if not crucial, that the relations between actor attributes and requirements are retained in the model. In the current version of the relational matching model, these relations are not modeled explicitly; although the model content may still be acceptable and it is possible to work with it, it is still an undesirable situation. For future versions of the relational matching model it is therefore suggested that something like a pre-processor or model-building program be considered, to which the researcher submits the relations between actor attributes and requirements. This program then builds its procedural and optimal counterpart from these associations.

3.6 Modeling demands generated by objects

In the discussion thus far no attention has been paid to how demands generated by the objects should be dealt with. For an appropriate matching it might be necessary to take the requirements by both the actor and the object (location) into account. Economic activities have spatial production requirements, but locations, or better, people and organizations with authority over that location, in their turn can specify regulations stipulating which type of activities may be allowed and which not. So far the discussion of inference trees concentrated fully on the requirements by the actor, but what of the requirements of the objects? Two approaches can be followed: (1) taking the object requirements into account by conducting a second and independent matching procedure scanning the actors or (2) masking the object requirements into 'inverted' actor requirements. In the Medemblik-study by Pellenbarg *et al.* (1974), two matching rounds were conducted. First, the locations were matched against the Schilling-profiles of activities. Next, the remaining activities were matched against the locational regulations for the area. A similar approach is advocated by Lucardie (1988b). Within the context of such a cross-matching, inference trees can, at least in principle, be applied for both the actor's and the object's part of the matching; the actor's part has been discussed but not the object's. One way of achieving this is by associating requirements by locations with their characteristics. Inference trees can then be used for modeling them and two separate matching rounds constitute the actual matching process. A problem associated

with applying such an approach, however, is that it can be expected that on many occasions the requirements of the object cannot be attributed to a definite set of object characteristics. In the case of locations, for instance, many different regulations on locational permits issued by local authorities are not subject to general rules, applied to all locations. This inherently random variation in requirements by objects reduces the usefulness of inference trees in modeling them. Note that this is not an argument against conducting two matching rounds, it just implies some reluctance concerning the use of inference trees for modeling object requirements.

A second way of incorporating object requirements into the relational matching model is by representing them as inverted actor requirements. This was the approach followed by e.g., Hendriks (1986) and it is also applied in the case study discussed in Chapter 6. Modeling object requirements as inverted actor requirements can be done by formulating an actor requirement stating that the object that is matched must not require anything the actor cannot fulfill. The advantage of such an approach is that one can incorporate the object requirements directly into the actor's inference tree (e.g., as a dimension). Moreover, no second matching round is needed anymore. A disadvantage might be that it is (perhaps) a computationally more intensive solution since the checking of inverted requirements implies some fairly complex logical and search operations. A second disadvantage is that although it is rather elegant that inference trees can do the entire job, from a relational matching point of view a double matching as proposed by Lucardie (Figure 2.3) might be preferred.

An issue that is worth mentioning is the incorporation of object requirements altogether. Two, perhaps three, opinions could be put forward, two of which are rather theoretical, and one which is more practice oriented. One position could be that object requirements such as locational permits in the case of, for example, industrial location, should not be part of a relational model of site suitability at all. Site suitability pertains to the appropriateness of a location for establishing an economic activity. Locational permits have nothing to do with that. Of course, when it comes down to actually making decisions as to where to locate a plant it will become important, but in the reconstruction of the areas that are, in principle, suitable, locational permits should not be included. From the point of view of regional policy makers, this might not be such a bad option. In an article about the applicability of knowledge-based decision support systems for land-use planning, Davis and Grant (1987; p. 55) formulate the task of land-use planners as:

“...drawing up any land-use plan is to assemble information from diverse sources (for example, special interest groups, political parties, scientific predictions, personal experience) and make decisions that best satisfy the multiple goals advocated by these sources. To do this, they must recognize and promote socially desirable opportunities and avoid socially costly hazards, assess and resolve as far as possible any inherent conflicts between the different parties, be aware of the long-term and short-term impacts of decisions on the various parties, using whatever data are available.”

One could argue that in order to be able to take such an informed and well-considered decision about how to allocate space, it is important to know which space can be of interest for which activities. An overview of possible conflicting land-uses and their geographical distribution, if available, can certainly improve the political decision-making process. Not including object requirements, therefore, reveals the interests of different groups as well as is possible.

An alternative opinion can be that object requirements should be part of the relational definition of a concept such as site suitability. If certain regulation prevents an activity from locating somewhere, then this location cannot be regarded a feasible choice alternative; it is clearly excluded from the available alternatives. In order to be able to reconstruct choice behavior, therefore, object requirements should be included in the model.

Both positions seem reasonable and they differ only because the rationale for building the model and conducting the analysis is different. If one wants to figure out which locations are suited for a certain activity from the point of view of the activity, then object requirements are irrelevant. If, on the other hand, one is interested in reconstructing locational choice behavior, object requirements should be taken into account. So why not settle for a structure in which both positions can be incorporated? A structure, therefore, by which both kinds of analyses can be conducted; matching with and without object requirements involved. Such a structure can be accomplished by putting all the object requirements into one or more separate dimensions of a one-sided matching model. The user should then be allowed to specify which dimensions must be included in the analysis. If object requirements must be included, then the user can specify that the associated dimension(s) should be part of the highest inference tree. In case object requirements are not to be included, the dimension(s) can be omitted. The dimensional

tree in Figure 3.3 shows this solution of representing object requirements by specific dimensions. The dimensions ‘spatial planning’ and ‘environmental policy’ contain requirements of locations in terms of locational regulations, identifying which conditions have to be satisfied by an activity before it can be considered a possible candidate for locating its plant. The computer program monitoring the matching process enables the user to declare which dimensions should be included in the measurement. Excluding ‘spatial planning’ and ‘environmental policy’ implies excluding the requirements generated by the locations.

3.7 Two-way matching

Conceptually, the term ‘two-way matching’ is not as unambiguous as it might seem at first sight. What has been called two-way matching before, actually consists of two separate rounds of one-way matching. In the application by Pellenbarg *et al.* (1974) this is rather obvious, but in Lucardie’s proposal (1988b) as well as in the dimensional solution of Figure 3.3 it is more or less hidden. What I think that the term ‘two-way matching’ should be reserved for, is something entirely different, and a lot more difficult to realize. The notion of ‘true’ two-way matching becomes clear from two of the three types of applications Schilling (1968; pp. 22–26) proposed for his catalog:

1. Search for suitable sites, given a specific activity;
2. Search for ‘suitable’ activities, given a specific site.

Clearly, a match is the positive result of the confrontation and comparison of requirements by an actor and an object, with their characteristics and properties. As such, a relational match constitutes a symmetrical structure; objects and context on either side, and a relational model in the middle. Theoretically, this implies that if one of the sides of the structure is initialized, and if the relational model is known, it must be possible to deduce the result on the opposite side. Therefore, once a context is fixed, i.e., once a specific activity is chosen, the resultant set of functionally equivalent locations can be determined. By the same token, once an object is fixed, i.e., once a specific location is chosen, the set of activities for which this location can be regarded suitable can be deduced. The relational model as a (nested) inference tree and its processing by comparing requirements with properties, reducing the set of feasible alternatives if these do not match, fits one side of

the two-way matching process. Modeling and processing as discussed above, make it possible to reconstruct a set of feasible alternatives, given a specific activity. But it becomes a lot more difficult to go the other way around, i.e., reconstruct activities, given an initial location. The main reason for this is that matching in either direction requires that the set of alternatives to be matched is fixed in both number and characteristics. In the case of matchings such as Schilling's and the ones carried out on the basis of inference trees, the matching is conducted on a set of locations fixed *a priori*, each of which is described by a fixed set of attributes and fixed scores on those attributes. The activity side, however, consists of only one specific type of actor, also represented by a set of fixed attributes and scores. Of course, it is possible to modify scores for both locations and activity prior to the matching, but once the matching starts, the set of alternatives as well as the activity are fixed. Matching the other way around, looking for activities given a specific location, thus requires a fixed set of activities, described as combinations of fixed attributes with fixed scores. The problem, however, is that such a set is just not available. The activities are implicitly described in the inference trees. If necessary, such as prior to matching, it is possible to fix those characteristics that are important for the matching (see Section 5.7.6 for an explanation of how this can be done). What we do not have, however, is a fixed set of empirically well-described actors. As should be recalled from Chapter 1 and 2, the aim of the analysis is to develop a model procedure which connects actor characteristics with object properties on the level of generalized abstraction. Certain actor characteristics or combinations of characteristics lead to certain requirements concerning the objects. The inference trees, therefore, contain implicitly general actor types, described as paths representing combinations of attributes and scores, but no real, existing, empirical actors; only generic types. Of course, these types might be derived from collections of existing actors exhibiting common traits, but once the abstraction into types has been carried out, the individual information is lost. It can be argued that one could keep a collection of empirical actors, for example in a data base, and conduct two-way matching for only those actors contained in that data base. On many occasions, however, it will be the researcher who determines the actor types because they constitute the outcome of a theory of locational choice set generation and site suitability. As such they are of a hypothetical nature and do not necessarily have any empirical counterparts. Of course, if the theory is adequate, the hypothetical actor type(s) will cover a set of existing activities. But developing the relational model on the basis of theoretical

insights does not necessarily require a data base of real, existing activities. It can even be argued that keeping a data base of real, existing actors to be matched on locational characteristics is not very useful at all, especially from the point of view of location theory. If, as Schilling suggests, a model and matching procedure that generates a set of suitable activities, given an initial location, is required, then what policy makers will be interested in is not names of particular, existing companies, but generic type descriptions of activities they can aim their acquisition policy at.

But even if such a data base is available, and even if one would want to know for which particular companies some location would be interesting, then there still would be considerable problems for two-way matching. These problems stem from two sources: (1) inversion of actor requirements and the problem of continuous variables, and (2) the lack of clues as to which locational characteristics are important. Suppose that there is a data base containing real actors. Suppose, furthermore, that one of those actors needs an area of N square kilometers to produce on. What would this mean for a matching departing from a specific location with 'area' as the only relevant attribute? Suppose that it is known that the area of the location is M square kilometers. Evidently, the location would be suitable for the activity if N is smaller than M , but the model does not know this. The information must somehow be incorporated in the model that in order to know whether the location is suitable for the activity, how much area the activity needs must first be checked, and this data should be checked against the area of the location. But this constitutes a concealed way of matching from activity to location, rather than the other way around. The problem is that many of the requirements of locations are implicit. If the location has a certain area, then only activities that require less can be considered feasible activities. But whether activities need this amount of area, or more or less, is not always simple to determine. It is very likely that the amount of area needed by an actor is a function of its production characteristics. And it is very well possible that this function is discontinuous and technology-dependent. For a pre-defined, existing actor, this is not a problem; the needs are defined and known and do not need to be calculated anymore, but it has already been suggested that this is not really interesting. But it is much more difficult to reverse relationships: reversing would imply that all the possible combinations of characteristics that can lead to another area estimate must be tried, before it can be concluded that an activity does not match the location's characteristics. This is difficult, especially when continuous functions are considered. If, for instance, the area needed for a specific production

plant is a function of two variables, then in order to arrive at a description of a suitable activity, the matching mechanism has to solve the equation for the area, such that the resultant area is smaller than what the location offers. In case more than one of these kinds of functions are involved, the problem becomes a lot worse. An additional problem is that of selecting the proper locational characteristics to match against the activities. The only way of knowing which characteristics have to be included, is by examining the activity-specific matching rules for the locational characteristics present there. But then one might as well conduct a one-way matching, searching for locations, given an initial activity. It can therefore be concluded that a truly two-way matching model will be very hard to build. If no empirical activities are available, and it was argued that it was not very interesting for them to be available, the search space for matching activities becomes enormous, if not infinite. Inference trees must be traversed in reverse direction, giving rise to numerous logical and mathematical problems. Their solution requires smart equation-solving algorithms and fairly intelligent logical inference and, of course, a lot of extra work. For these reasons it was decided to conduct matching only in one direction: finding suitable locations, given an *a priori* fixed activity.

3.8 Model content: empirical procedures

Application of a relational matching system requires three kinds of empirical data: actor data, object data and inference trees. Of course, in order to be able to conduct matching one also needs an algorithm that can apply this information in a matching procedure, but such an algorithm is a computational procedure, not empirical data. The process of collecting these empirical data is called the process of 'knowledge acquisition'.

3.8.1 Knowing the objects

From the discussion on two-sided matching it can be inferred that one of the components needed for relational matching is a data base representing the objects in terms of attribute-value combinations. Such a data base can contain facts as well as rules. A fact is a plain attribute-value combination, a rule is a procedure for inferring attribute-value information from existent information. For instance, if the data base contains the facts that region X has a population P and an area Q, then there could be a rule declaring that the population density Z can be calculated as P/Q . An important issue

in designing the object data base considers the selection of the attributes describing the objects. During the matching, objects are compared with the requirements by actors. Therefore it is important that requirements and objects are described in the same language, i.e., identical attributes. In real, empirical applications this can sometimes be hard to accomplish. Reasons for this could be the following: developing inference trees generally happens in a sequence of steps or rounds, during each of which the resultant rule bases are modified and refined. This implies that only at the end of the process of model building the correct set of attributes is known. It is unlikely, however, that these attributes are exactly those that can be found in, for instance, public data bases and governmental statistics. Typically, standard statistical information differs from the information one needs for the matching application. Either because the objects do not exist in statistical data sources, or because the statistical information about the objects does not conform to what is required. The former will happen frequently when dealing with spatial objects such as regions. The spatial behavior of individuals, households, or firms only partly coincides with standard statistical areas. As a result, much of the information needed for the matching will have to be calculated from standard statistical area data. But even if the objects about which statistical information is available do coincide with the ones needed in the matching, it is unlikely that this information coincides with the information needed for the matching. Statistical bureaus do not reckon with the individual spatial requirements of activities, and their data will therefore only be of limited direct use for a relational matching model.

This problem can be tackled in various ways. One solution is that after the matching units (the objects) are determined, object properties are defined in terms of available information. An alternative approach is that one conducts the development of the inference trees independently, after which rules are specified that convert available information into the desired attribute-score combinations. In most cases a mixed strategy will do. In general, it is not such a bad idea to choose as matching locations those areas for which statistical information is available, even if these areas do not quite coincide with the areas that are relevant for the spatial choice behavior under consideration. As long as the relevant areas are aggregates of statistical areas (Figure 3.8(a)), it is not too difficult to construct a data base for these aggregates. However, in case the relevant areas do not form such clear aggregates (Figure 3.8(b)), sophisticated equipment and algorithms would be required to conduct the data-transformation.

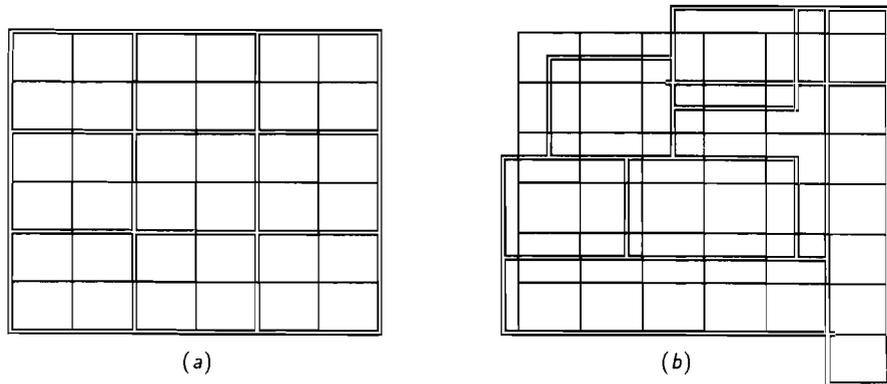


Figure 3.8: Matching (a) and non-matching (b) of administrative boundaries and locational alternatives

Such techniques and algorithms are available in several large geographical information systems (GIS) such as ARC-INFO or GRASS (Green *et al.* 1985; Burrough, 1986; Dueker, 1987; Tomlinson, 1987; Scholten and van der Vlugt, 1988). Hooking up the locational data base of the matching system with these types of GIS, so that it is possible to work with customized areas, is certainly worth pursuing. But as long as these transformations cannot be made in a fast and flexible manner, it seems wise to either accept as matching areas those for which statistical data are available or to create one's own by aggregating areas for which statistical data are available. Next, an attempt could be made to develop the matching rules while bearing in mind what information is available on the objects. This means that an attempt is made to express as many of the requirements as possible, in terms of the available information or in terms of attributes that can somehow be calculated from the information so made available.

3.8.2 Knowing the actors

Like objects and object characteristics, actor information is needed for the matching. When an inference tree is traversed, actor attributes are inspected and depending on the values the actor scores on those attributes, a route through the inference tree is chosen. As in the case of objects, many of these attributes and their scores can be stored in the form of facts. Information that can be derived from information that is available elsewhere in the data base, can be stored in the form of rules. Of course, one has to take care that

for each and every attribute that is part of an inference tree, either facts or rules are known. More information on an actor than is actually needed only slows down the process of information retrieval, less information than needed is harmful. Lacking actor knowledge generates the problem that no decision about how to go on through the inference tree can be taken; a deadlock therefore.

3.8.3 The problem of generalization

The various routes from the root of an inference tree to its terminal nodes (the matching rules), describe different activity types, each of them seen as a set of activity attribute-score combinations. Inference trees as sets of activities also introduces the idea that with each branching downward, instances of a higher-level class are specified. Coal-based chemical industries with high and low amounts of production are both instances of the higher class of chemical industry. Inference trees, therefore, give us an implicit way of carrying out generalizations within the relational model. The hierarchy contained in the inference tree makes it possible to conduct relational matchings for more general types of activities, for example, if one is not interested in a very specific case, but wants to explore the matching solutions for a more general category of actors. Another possibility is that information on more specific kinds of actors is missing. Being able to conduct a matching for the above generalization can then be an attractive alternative. One way of having generalizations contained in the model and thus in the measurement procedure, is by introducing 'don't care values'. Specification of such a value can be seen as replacing a node in the inference tree representing an actor attribute, with an instruction that every path beneath that node is a valid path. In effect, this means that the matching program carrying out the comparison of the object properties with the requirements expressed in the matching rule can access more than one matching rule to try to find a successful match. Each of the paths under the 'don't care' value namely, leads somehow to a terminal node and thus a matching rule. If all the paths under a 'don't care' value are possible paths, then all the matching rules they lead to are possible matching rules. Logically, this kind of generalization means the replacement of one or more 'AND' conditions in the model, by (inclusive) 'OR' conditions, thus increasing the possibilities of finding a successful match. Generalization as applied here, therefore, can be formulated as a set-theoretical operation, namely, that of set union. More formally: if U and V are sets of objects matching the requirements of

activities P and Q, and if both P and Q are the only possible instances of a higher class Z, then the set of objects successfully matching Z is equivalent to the union of U and V.

The possibility of dealing with generalization in a relational matching model also introduces a problem. Not only can more specific formulations of an actor or activity be seen as instances of a class that is less well expressed, but these classes themselves can be instances of yet higher classes. As long as this implies searching the inference tree upward, there is no real problem, because it is possible to conduct the set-theoretic operations mentioned above. But what if the root of the tree is reached and one arrives at the level of a generic type of actor, say 'chemical industry', which can be regarded a kind of industry. An activity such as industry requires space for its operation and therefore needs to be located somewhere, just as services and agriculture. Should these then again be considered instances of a yet more general class 'spatial activity'? The result of following such an approach would be one, huge classification tree, departing from some kind of root, first branching into large categories such as industry, agriculture, services, etc., then branching again into sub-classes such as chemical industry, printing industry, metallurgy etc. However, what should be contained in a relational model of site suitability are the causal relationships that explain the associated requirements. Following from this one could expect a completely different kind of inference tree. An inference tree that is still large, but one not containing any kind of standard typological classes, such as just mentioned. Instead, one that branches in a rather large number of production specific attributes such as the types of material that are used as a resource, the amounts produced, the type of labor needed, etc. Yet, the reason for introducing dimensions into the matching models was that these broad categories can be treated as more or less separate, although possibly interrelated, abstract variables. The generalization problem thus seems unavoidable and huge. How and where should different activities be divided into classes and sub-classes? What consequences does this have for the model structure? Do things become so complex that they cannot be understood anymore? Although this problem is indeed serious and needs further investigation in order to be solved in both a theoretically and practically adequate manner, it was decided to bypass the problem by allowing for an initial set of, perhaps arbitrary and traditional, but understandable and practically applicable, generic types of actors or activities. Within types such as chemical industry and aluminum processing industry, relational modeling by means of inference trees can be conducted. If, in practical applications, the classes

turn out to be too large, giving rise to substantial problems in the modeling process, new sub-types, again based on essentially arbitrary decisions by the researcher, can be constructed. I fully realize that, if possible, these kinds of operations have to be avoided. It is for these types of operations on concepts and classes that relational methodology tries to offer an alternative. On the other hand, it does encourage the further development of empirical applications, which in their turn, might provide instruction on how the modeling can be improved, where it is wrong, where it is good, and how to pursue further research.

3.9 Knowing the inference tree: techniques of knowledge acquisition

The third, and most important, knowledge element to be implemented in a relational matching system is the empirical content of the inference trees. Apart from the actual content of the inference trees, it is important to have an idea of how this knowledge is gained in the first place. This part of the knowledge acquisition is often denoted as the problem of knowledge elicitation (for a brief overview of the field, refer to, e.g., Davis, 1979; Boose and Bradshaw, 1987; Kahn *et al.*, 1985; de Greef and Breuker, 1985; Haley and Williams, 1986; Hart, 1986; Wielinga and Breuker, 1986; Diederich *et al.*, 1987; Eshelman *et al.*, 1987; Gale, 1987; Littman, 1987; Reitman Olson and Rueter, 1987; Rajamoney *et al.*, 1987; Swartout and Smoliar, 1987; Lundberg, 1989). Knowledge elicitation and the development of systems or tools for conducting knowledge elicitation are important issues in the development of expert systems. Expert systems purport to represent expert knowledge and at least a significant part of the expert's reasoning and inference abilities. As such it is important to try to find out how people have their knowledge arranged, how it is accessed, and how conclusions and premisses can be linked. The elicitation process has two aspects; a formal and an empirical one. As far as the former is relevant for the development of relational inference trees, it has been discussed when dealing with the decision table perspective on inference trees (Section 3.3). This formal aspect of knowledge elicitation is the least problematic of the two. However complicated rule bases and various forms of knowledge representation might be, in a formal analysis one can nearly always find a rigid method, structure, or scheme with which to analyze the existing knowledge base. It is perhaps for this reason that most of the work done on elicitation concerns this formal aspect. The problems

connected with the empirical aspects of knowledge elicitation are much more difficult to solve. Work done in this field concentrates on problems such as the development of general elicitation methodologies and techniques that can be applied in elicitation processes, regardless of their empirical content. How does this relate to the development of relational matching models? Is knowledge elicitation important, and if yes, is it possible to conceive of some sort of general technique of developing adequate relational matching models? A clear answer to this question is difficult to give. The literature on knowledge acquisition is massive, and tends to concentrate on a limited number of rather specific techniques and methods. The problems associated with the development of a knowledge acquisition methodology for relational modeling are complex and require a separate research effort. However, a few ideas on what a more structured knowledge acquisition for relational modeling could look like may be mentioned. First of all it is important to note that, to a certain degree, the relevance of the availability versus absence of such a technique depends on the objective for which a relational model is developed. In case the model is required to be reliable and 'unproblematic', for instance if it is supposed to be used as a prediction or a decision support tool, the availability of structured knowledge acquisition is important. The researcher might have to develop a model with the help of domain experts and has to come up with a model which is both logically and empirically acceptable. The availability of a knowledge acquisition method will certainly facilitate this process. In case the model has a much more hypothetical character, for instance if it constitutes an explanatory model of, say, production milieu, knowledge acquisition becomes much less important. Of course, logical and completeness checking are still important, but since the researcher in this case is his own knowledge engineer, guidance by a structured method becomes less important. As mentioned in Section 3.9.2 however, the latter case requires reliable test sets in order to be able to validate the model.

In general, two aspects of knowledge acquisition for the development of relational matching models seem important. The first is how to translate empirical knowledge into inference trees. The second considers the problem of how to accumulate knowledge in the first place. One might prefer to designate a third aspect, namely, testing accumulated knowledge for accurateness and reliability, but here it is considered to be part of the broader aspect of accumulating empirical knowledge.

3.9.1 Inducing inference trees: efficiency versus understandability

Up to now no really structured or automated method for translating empirical knowledge into an inference tree has been developed. Of course, techniques such as ID3 can be used not only for optimizing trees, but also for building trees out of an initial set of examples. Building rule bases out of sets of examples, a process denoted as ‘learning by example’ or ‘inductive learning’, constitutes one possibility for covering at least the initial stages in the development of a rule base of relational decision and matching rules. One could, for instance, think of a knowledge acquisition process that starts with an ‘example stage’ in which some initial sets of actor’s and object’s attribute-value combinations are joined to relational decision rules. After testing for completeness and consistency, an induction technique such as ID3 then processes these examples into a first-draft version of an inference tree. Shapiro and Niblett (1982) and Shapiro (1983) call this process ‘structured induction’. When an optimization technique such as ID3 is used in an inductive manner, i.e., for generating inference trees out of raw and unconnected examples, the question of efficiency versus understandability becomes important. Arbab and Michie (1987), for example, mention a fundamental trade-off between what they call the ‘understandability’ and ‘efficiency’ of rule sets generated by means of inductive optimization techniques. More efficient rule sets are not necessarily more understandable. They explain that ID3 tends to generate efficient but not very comprehensible trees. As a measure of understandability they suggest the ‘(non)linearity’ of the decision tree (Figure 3.9). Shapiro (1981) and Shapiro and Niblett (1982) mention a positive relation between the understandability and linearity of a decision tree. For applications in knowledge acquisition, Arbab and Michie suggest an optimization of decision trees, such that linearity is preferred over efficiency. If several trees with equal linearity can be found, the most efficient should be chosen. The optimization algorithm they suggest is called RG (Rule Generator).

Both ID3 and RG are important and should be considered tools for supporting the (initial) development of relational matching models. Whether RG must be preferred above ID3 or whether ID3 proves to be more satisfying, is something that must be explored in future applications.

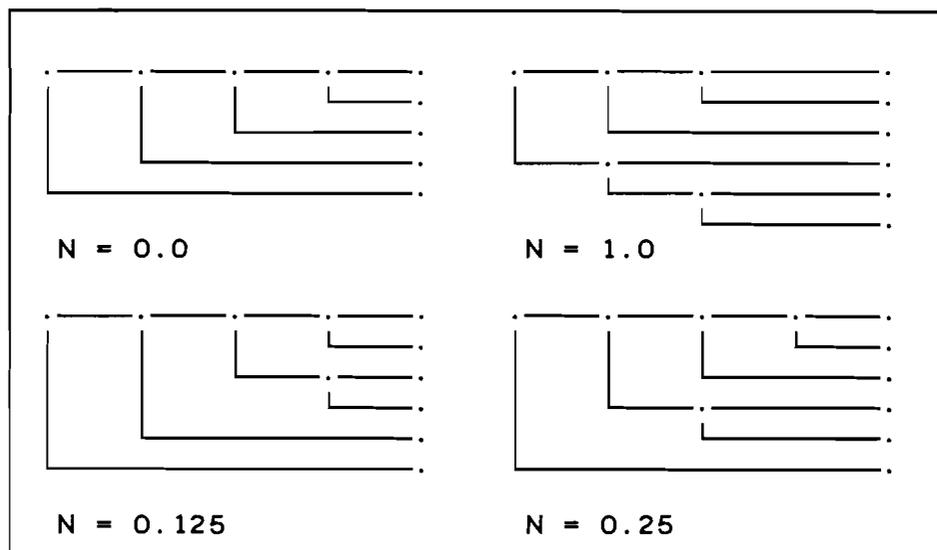


Figure 3.9: Non-linearity measures of four trees (Source: Arbab and Michie, 1987)

3.9.2 Validating the model

Apart from the evaluation of a relational matching model on a set of formal criteria, validating the empirical content of the model forms a second moment of model evaluation. A complete and consistent inference tree is no guarantee for an empirically adequate inference tree. The empirical test, therefore, has nothing to do with the logical quality of the model, but only with its empirical content. Running the model on a pre-defined set of alternatives, of which the correct measurements are already known, constitutes one possibility. Comparison of the model results with the actual scores thus yields an indication of the empirical quality of the model. The feasibility of such a procedure for validating the model therefore depends on the possibilities to generate an appropriate set of test alternatives. One type of candidate for a test set is formed by choices which have actually taken place. In case of locational behavior, for example, one could use the actual locations of firms as the members of the test set. However, using actual choices as members of the test set is only of limited value. The main problem is that a relational matching model does not predict behavior; it reconstructs constraints on behavior. Therefore, matches cannot really be compared with choices

actually made. The more so since the ideal, generic types of activity represented by the matching model, do not exist in reality. Only in case one has good reasons to assume that the model activity adequately represents the existing activity, the actual behavior of which is used as the test case, can this behavior serve as a test case for a relational matching model. Such good reasons, however, can only stem from a careful comparison of both activities; a piece of work that in itself requires extensive study and careful analysis. Characteristics of activities change and so do their objectives. As Massey (1979b) explains in her analysis of regional problems, the behavior of firms and enterprises is a continuous response to whatever the environment offers, all within a framework of evolving objectives. This renders past choices not automatically appropriate candidates for testing a model that tries to incorporate both environmental influences and the activity's objectives. Hendriks (1986; pp. 107–122) suggests process validation rather than result validation as an approach for constructing empirically adequate relational matching models. The idea is that the development of the model is seen as an (iterative) process that is validated on a continuous basis. Every time empirical knowledge is added to the model it is tested. Hendriks therefore suggests the use of 'key informants' or domain experts for assisting in this process of validation. Figure 3.10 displays such a process in its most general form.

An initial 'guess', for example a set of raw examples, is formally tested for completeness and consistency. If correct, it is submitted to an optimization procedure. Next, a matching on the basis of this model is conducted. The results are then evaluated by the domain expert(s), for instance employees of a specific firm that know about its locational decision making and thus its locational requirements. At this stage result validation could be useful, but only if the actual choice against which the model is tested is one that the domain experts know sufficiently well. The reason to stick with this one activity is that only the domain experts can interpret the results of confronting the model prediction with the choice actually taken. Of course, it may require a lot of ingenuity on the part of the researcher to extract this information from the domain experts. It will require the same degree of ingenuity to designate true reasons from rationalizations and to overcome the problem of cognitive dissonance (Festinger, 1962). Of course, experts can be wrong too and it might be advisable to consult more than one and from different disciplines, but it is they who can best judge the adequacy of the model predictions. By embedding such a process validation into an iterative procedure of knowledge acquisition, it should be possible to develop

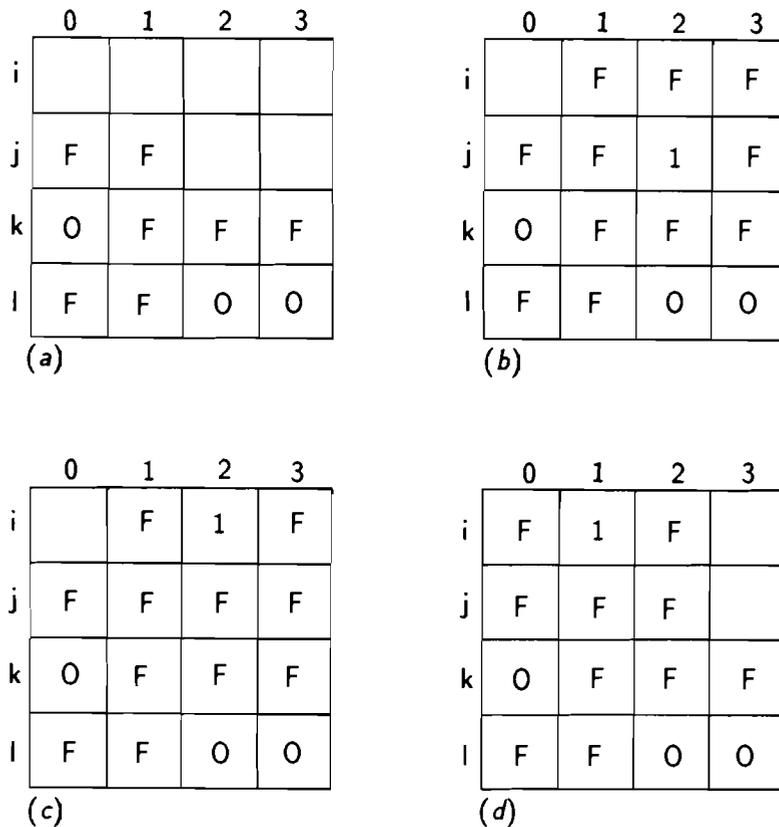


Figure 3.11: The problem of multiple matching solutions

This rule could be read as: A location X can be considered suitable for an activity if none of its neighboring locations has an activity. An example of such a rule could be the result of some kind of policy saying that, for instance, refuse incinerators should be spread more or less evenly over a region. Siting one of them will have consequences for where others may be put. This, however, constitutes a problem for the kind of relational matching suggested here, because it would imply that the order in which the alternative locations are tested becomes important. This is illustrated in Figure 3.11.

Figure 3.11(a) represents the initial situation before any matching has taken place. Locations k0, l2, and l3, already have an activity. Since the matching result of the neighboring locations is already known to be negative, they are excluded from the matching. Figure 3.11(b), Figure 3.11(c), and

Figure 3.11(d), each represent the consequences of a successful matching of a different alternative. Clearly, if alternative j_2 is matched first (b), a match for alternative i_2 becomes unattainable. If, on the other hand, i_2 is the alternative that is matched first (c), j_2 becomes unattainable. And if i_1 is matched first (d) both i_2 and j_2 become impossible to match. What then is the real meaning or the actual relevance of this problem of multiple matching solutions? Does it constitute a serious objection to relational matching as presented earlier in this chapter? I think not. The problem of multiple matching solutions stems from confusing relational reconstruction of a set of choice alternatives with questions about the dynamic, time-dependent character of spatial choice processes. In the relational matching approach outlined here, each potential choice alternative is matched individually, given a stable, unchanging context. Of course, it is more than likely that once certain choices are actually made, i.e., once one of the possible alternatives in the reconstructed set of functionally equivalent objects is chosen or implemented, this may have consequences for the possibilities of successive choices. This implies that if one wants to incorporate dynamics of this sort into the system, one would have to simulate changes in the environment, changes in the characteristics of objects and actors. These changes can then be induced either by means of arbitrarily modifying the properties of objects and/or actors, or they can be the result of simulated choices made on the basis of the outcome of a relational matching model. How this can be done is discussed in Chapter 5.

3.11 Multi-locational actors

Another, much more serious problem is associated with the possibility of multi-locational actors. Until now, actors were implicitly assumed to be units; single, decision-making entities. In many instances, this type of actor can be considered a proper representation of decision-making entities in real life: individuals, households, institutions, and many types of firms and economic activities. For the reconstruction of locational choice by economic activities and for the reconstruction of production milieu, however, one has to reckon with multi-locational activities, for instance in the form of either corporate firms, or as a set of individual activities, the location of which depends on the location of the others. Examples of the latter are locating a set of energy facilities (Church and Bell, 1981), or multi-locational problems such as locating schools or hospitals, or allocating telephone booths or

mailboxes to a new neighborhood. Multi-locational actors seem to be a serious problem for a relational matching model as outlined here. The question therefore is whether a relational matching model based on inference trees is suited to the reconstruction of feasible locational choice alternatives for multi-locational actors. It will be clear that although not identical, the problem has some resemblance to that of multiple matching solutions. Because of this resemblance, it is possible to find the same way out of the problem. The problem of multi-locational actors is partly resolved when regarding each locational member of this composite actor an individual activity, for which a suitable set of alternatives is reconstructed. If one or more other members of the composite actors force certain requirements upon this individual member, then they can be incorporated in its inference tree. Each of the members must be located, otherwise the multi-locational activity cannot be located. If one or more sub- or branch-activities do not find suitable locations, then the composition of the multi locational activity must be changed. Note that conducting a relational matching for each of the members of the multi-locational activity does not imply that each of these members is considered a decision-making actor. On the contrary, the actual decisions as to which line of action to follow and which alternative will be chosen is exogenous to the matching model. The matching model just reconstructs feasible alternatives on the basis of characteristics. Whether one or another alternative is actually chosen is an entirely different matter, just as it was in the case of multiple matching solutions.

This solution must be considered 'a way out of the problem' because it helps to retain the two-stage process of choice modeling advocated here (Section 1.3). However, in case of multi-locational actors, the second stage, the stage in which a choice set is analyzed by an optimization technique, will be a very important one. Multi-locational matching problems might tend to generate very many solutions because many different combinations of location allocations will satisfy the relational constraints. It is especially for this kind of solution that the second stage can turn out to be very useful. This stage will then address problems such as the selection of the solution which implies the least amount of average traveling between the various allocated locations, or the maximum average market or service area covered by the activities on the allocated locations. These are, of course, problems that are dealt with in spatial allocation modeling (Scott, 1971; Karlqvist *et al.*, 1975; Church and Bell, 1981).

3.12 Conclusions

Despite problems connected with multi-locational actors and multiple matching solutions, relational inference trees seem to provide ample means for representing the relational definition of concepts. Paths through the tree representing actor types lead to terminal nodes referring to a matching rule which is a disjunctive set of conjunctive terms containing the associated matching requirements. Apart from this declarative, definition-oriented aspect, relational inference trees are also attractive from a more procedural perspective. Relational matching can be viewed as finding a route through the tree upon which the associated matching rule is executed. An additional advantage of inference trees is that the formalism offers some means to partition a complex matching problem into a hierarchical structure of dimensions, each of which can be considered a matching problem of its own. Inference trees also offer ample opportunities for formal evaluation on criteria such as logical consistency and completeness. Apart from these advantages, however, the current representation and processing of the inference tree suffers from a number of problems reducing its potential. In order to be able to use inference trees more adequately, problems such as complete paths being associated with complete matching rules, and the redundancies contained in the matching rules, must be resolved. The next step, presented in the next chapter, is the construction of an automated system that can conduct matching on the basis of the type of inference trees presented here. Such a system is attractive because carrying out an actual matching on the basis of a complex relational definition and a possibly large set of choice alternatives—locations in case of the relational measurement of site suitability—implies a huge amount of dull, but precise, work. Moreover, the useful application of this kind of modeling in spatial planning and decision making requires the possibility of conducting many model runs and measurements, each of which may contain different actor and object properties. For this kind of application, automation of the matching is essential.

Chapter 4

Implementation

ABSTRACT

In this chapter the implementational aspects of a relational matching model and an automated matching procedure are discussed. The implementational requirements are outlined, and a proposal for a relational matching system based on expert system technology is formulated. The main techniques which have been applied are knowledge representation by means of frames, various forms of inference, and a structure for controlling the system's general behavior. Prolog is presented as an attractive language for the implementation of such a system.

Keywords: matching system, data bases, knowledge representation, facts, rules, inheritance, inference engine, frames, Prolog

4.1 Structure of the matching system

The availability of an automated system that is able to conduct relational matching on the basis of a relational matching model is important for several reasons. Given an initial set of possible choice alternatives and a specific actor type, carrying out a relational matching entails a lot of rather simple, dull work. Inference trees must be traversed, something which may require inferences and calculations about actor characteristics. Once a terminal node is reached, the total set of alternatives must be matched against the requirements associated with the matching rule. Again, this may require somewhat complicated inferences about object characteristics, but basically

the work comes down to a considerable amount of calculation and computing, with the need to drag many objects through the same test, and compare scores with other scores.

As mentioned in Chapter 2, for his application, Schilling (1968) must have excuted the work manually. The lack of automated procedures as one possible reason for the (hopefully) temporary disappearance of matching studies from geography has already been mentioned. A lot has changed since the sixties, however. Computers are widely available now, and are faster and smarter, just as the amount and variety of applications increased enormously, over the last decade or so.

The implementation of a system that can process relational matching models requires four kinds of functions which today's computers are able to provide: searching, knowledge representation, inference and calculation. In Chapter 5 it is suggested that in case a relational model is to be part of a larger and integrated decision support system, extensive graphics display and user-friendly interfacing are two additional functions that are required. Conducting relational matching implies calculation, because many actor, as well as object, characteristics must be computed out of other information with the help of rules. The same holds for the computation of requirements if they are defined as numerical functions of actor characteristics. The total amount of electricity needed, for example, can be a function of the production volume. Matching also implies a lot of searching. The activity data base and the relational rule base must be searched in order to infer a route through the inference tree. Once a matching rule is found, the object data base must be searched to find objects the characteristics of which satisfy the disjunctive matching rule. Searching must also be done on the evaluation of actor and object characteristics. Is a property about which information is required available? If not, which are the rules that could be used to infer the information?

Knowledge representation provides an answer as it refers to the broader area of how to represent different kinds of knowledge on a computer. The literature on this issue is extensive (e.g., Minsky, 1975; Levesque, 1984; Mylopoulos and Levesque, 1984; IEEE, 1986 (special issue); NIJCAI, 1987; Woods, 1988; etc.). Knowledge representation is an important area of research in artificial intelligence. It was realized that in order to make systems behave somewhat intelligently, it is not enough to make them fast and to equip them with inference capabilities. In order to display some intelligent behavior, systems need to be equipped with a significant amount of knowledge about 'the world'. Although developing an intelligent system is not

what automating relational matching is about, techniques of knowledge representation are very useful for the implementation of the relational model, as well as for constructing the actor and object data bases and the search techniques which must operate upon them. As yet another task to be implemented in a relational matching system, inference is closely linked with knowledge representation. Perhaps one can say that inference capabilities constitute the dynamic element in knowledge representation. One way of structuring knowledge is by dividing it into facts and rules. Inference is the activity of accessing that knowledge. The term 'inference' applies to logical procedures to generate information with. On a computer this can mean things like looking up a fact in a data base, but also the scanning of a data base for rules representing procedures for how a specific element of information can be derived from other information, as well as the process of actually deducing or computing that information.

Thus, calculation, searching, knowledge representation, and inference are the main forms of computing to be conducted by a relational matching system. Another way of looking at such a system, though, is by formulating its components. These can also be considered tasks to be conducted by the system, but they concern these tasks as seen by the user. When using the system it is not very important how these components are implemented, and by which computing method they are driven. On the contrary, from the user's point of view a relational matching system must contain a number of components, linked in such a way that conducting relational analysis is possible. To achieve this, five major components need to be incorporated into an automated relational matching system.

1. Representation of the objects:

Relational matching requires a data base containing information on objects. This information can be in the form of either facts or rules. For both the user and the matching program, it must be possible to access object-specific information. The matching program needs the information to compare objects with requirements. The user must be able to inspect and modify the characteristics of the objects. This enables the user to either simulate changes in the object characteristics or make them permanent in case of an updating or modification operation.

2. Representation of the activities:

As with the objects, activity properties have to be represented in a data base as well. Depending on the characteristics of an actor, certain properties of

the objects are important, others will become irrelevant. This relevance of object characteristics is expressed in the matching rules, which are again a function of a set of actor attribute-value combinations. Therefore, information on these actor-attribute combinations must be available to the matching system; either in the form of facts or as rules.

3. *Representation of the inference trees:*

Of course a very important component is the one containing the inference trees, the relational matching model. Somehow, these trees must be represented so that the matching program can traverse them and find the proper matching rules. Ideally, the inference trees should be in a format, or should be convertible to a format, that can be handled by a program for optimizing them (ID3, Section 3.4.2.1) and for checking their completeness and consistency. One way of achieving this could be that a separate system handles the building and developing of the inference trees, the result of which is then passed on to the relational matching system.

4. *The matching program:*

The matching program is that component of the system that deals with the actual processing of the inference trees, the identification of the matching rules, and the comparison of objects with the requirements associated with these matching rules. It is the engine driving the matching. It must check the activity for which a matching has to be conducted, the objects that must be contained in the matching, and the dimensional structure of the matching problem. Its result should be a subset of the initial set of objects, the members of which can be regarded functionally equivalent.

5. *Explanation:*

Although not strictly necessary, a utility that can explain the outcome of a matching process can enhance the usefulness of a matching system considerably. Of course it is interesting to know which sites can be regarded suitable for a certain activity, but it is perhaps just as interesting to be able to ask the system why another location cannot. What requirement or set of requirements was the location unable to fulfill? And what about some other objects? Did they fail to pass the matching for the same reason? Schilling's proposal for application mentioned in Chapter 2, contains a 'diagnosis' option. This means that if a location does not pass the test, it should be possible to know why it did not. Only if it is known which requirements were not satisfied, can proposals be made for changing the current situation. Clearly, this holds some important implications for regional policy

incentives. It is perfectly possible that in spite of certain regional incentives, specific locations still do not match the requirements generated by certain types of activity. By the same token, diagnosis of mismatches can show where possible incentives and changes in the existing situation might have effect.

Figure 4.1 shows a conceptual scheme of a relational matching system in which the basic components mentioned above are included.

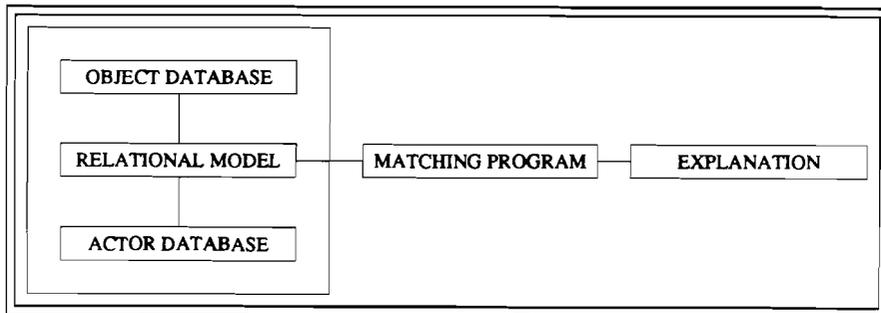


Figure 4.1: Conceptual scheme of a relational matching system

4.2 Frames for representing facts and rules

The static part of knowledge representation comprises the methods and techniques that can be used to represent facts and rules on a computer in such a way that they can be used by the dynamic part, the inference component. One approach which is particularly useful for representing knowledge about actors and objects in a relational matching system, is representation by means of ‘frames’ (Minsky, 1975; Forsyth, 1984; Savory, 1985; Cuadrado and Cuadrado, 1986; Reitsma, 1986; Lucas and van der Gaag, 1988). What is termed a ‘frame’ here, can be defined as a unit of knowledge consisting of a set of ‘slots’, each of which is used for representing a specific aspect of this unit of knowledge. One possibility for describing actors or objects as combinations of attributes and values by means of frames is the construction of a set of relations or ‘tuples’, each of which contains four entries or slots:

1. The name of the object described by the frame;
2. The name of the attribute the relation is about;

3. A 'facet' containing information on how the attribute is represented;
4. A value.

Note that within the context of representation by means of frames, it is irrelevant what the object described by the frame actually is. Both the actors and objects of a relational matching are 'objects' in the context of frame representation.

Figure 4.2 shows an example of a frame for the object (activity) 'limestone industry'. All relations in the frame contain the object-name. The reason for this is that it was chosen to consider frames as dynamic collections of relations to be constructed 'on the fly' rather than having them stored in large, static structures. The frame in Figure 4.2, shows two types of facets: 'value' and 'if needed'. A 'value' facet declares that what is contained in the value slot is actually a value, i.e., a numeric or symbolic value. The tuple therefore represents a fact. In contrast to a 'value' facet, an 'if needed' facet declares that the value slot contains a rule, or a reference to a rule, which can be used to infer the value of the attribute represented by the attribute name. The fourth tuple of Figure 4.2, for example, declares that in order to infer the value for 'typical employment' for limestone industry, a rule must be triggered. This rule might use the total number of workers and the percentages of seasonal, home, and shift work, to calculate the typical employment figures. The use of a rule can imply the use of other rules. Suppose, for instance, the rule for computing the typical employment shown in Figure 4.2, contains a reference to the total number of workers required. It is then possible that the calculation of the total number of workers required in its turn can only be done by using a rule. This might contain other references to, for instance, the annual production and the employment rate(s). The last tuple in Figure 4.2 shows an attribute 'ako' (a kind of), and can be used to define relations of inheritance. The ako-relation defined in Figure 4.2, for example, defines limestone industry as 'a kind of' industry. An appropriate frame driver, a computer program that can handle frames in order to represent knowledge, can use this information in case declared facts and rules about limestone industry are insufficient to deduce the desired information. Declaring limestone as an industry will then make the frame driver try to deduce the desired information out of the information associated with industry, rather than limestone industry.

The frame as displayed in Figure 4.2 is an example of how knowledge on objects (either actors or choice alternatives), can be represented. The

FRAME: limestone industry

OBJECT: limestone industry
ATTRIBUTE: annual production
FACET: value
VALUE: 2000

OBJECT: limestone industry
ATTRIBUTE: employment costs
FACET: if needed
VALUE: rule_employment_costs

OBJECT: limestone industry
ATTRIBUTE: number of workers(m/w)
FACET: if needed
VALUE: rule_number_of_workers(m/w)

OBJECT: limestone industry
ATTRIBUTE: typical employment
FACET: if needed
VALUE: rule_typical_employment

OBJECT: limestone industry
ATTRIBUTE: ako
FACET: value
VALUE: industry

.
. .
. .
etc.

Figure 4.2: Frame for the limestone industry

dynamic element of such a representation consists of a set of operations to be conducted on the tuples contained in those frames. In their application of a frame-based knowledge representation system, Cuadrado and Cuadrado (1986) define a number of operations to be carried out on the tuples. They distinguish five basic actions: retrieving information from tuples, storing information into tuples, removing information from tuples, appending an item to a list contained in a value slot, and replacing information in tuples by other information. Each of these actions has its associated facet (*if_needed*, *if_added*, *if_removed*, *if_appended*, and *if_replaced*). Although the proposal by Cuadrado and Cuadrado is generally attractive, it has some strange aspects. This holds in particular for the way they implement the system in the Prolog programming language, which is discussed further in Section 4.4.1. What is somewhat surprising at the conceptual level, however, is that for the special case of list values, only an 'append' action is defined. Clearly, introduction of a list as something to be contained in a value slot changes the concept of the tuples. If a value entry is allowed to hold multiple values by means of a list, then a lot of other things have to change too. First of all, each of the other actions to be performed on the tuples should have its list equivalent. Removing elements from a list or replacing them with others is just as important as appending new elements to a list. More serious, however, is that if both list and non-list value entries are allowed, either the frame driver needs a lot more intelligence to know where to put, and how to retrieve, specific knowledge from the tuples, or this knowledge has to rest with the user. It seems that Cuadrado and Cuadrado assume the latter, but I do not agree with them on this. A knowledge representation system should be able to hold and maintain knowledge. The user must be able to use that system without knowing how its knowledge is organized. This might of course require some special query language. But it should not be the case that the user has to know whether knowledge is stored as lists or non-lists in order to use the proper queries. Either one chooses lists as the formal representation of values, in which case single values form single element lists, or one chooses a non-list value format with one single value per tuple. The alternative is to allow both, and sharply increase the intelligence of the frame-driver program. The latter alternative, however, would imply the implementation of quite a lot of meta-knowledge, since the system would have to know how its knowledge is organized.

4.3 Find-deduce-inherit-ask inference

It will be clear that this kind of frame representation is closely related to aspects of inference. Extracting knowledge from a frame data base implies the application of inference techniques. Obviously, the facet slots refer to various inference techniques to be applied in extracting knowledge from such a frame data base. The example of Figure 4.2 shows three types of inference; 'find' (value facet), 'deduce' (if_needed facet) and 'inherit' (ako-attribute). Find inference is the basic form of inference. The other two are of a somewhat higher order, because they can be defined in terms of each other and in the end come down to find inferences. Find inference means that when information on an object is needed, for example an attribute-score combination, the program simply looks for that information in the value entry of the attribute-specific frame item for that object. Deduction is the type of inference that is applied if the desired information is to be generated by means of the application of a rule. Inherit inference refers to the derivation of information by inspecting the inheritance relations, which are then used to extract the required information from the frames of higher-order or 'parent' objects. Find, deduce, and inherit actions must be regarded as alternative means to infer information. As such they represent several possible attempts to perform a search operation on an item of information, to be carried out one after the other. If a find action remains unsuccessful, a deduce is attempted. If this still does not yield the required information, the hierarchy of classes is inspected. Combining deduce and inherit actions can turn out to be very efficient for representing knowledge. For instance, if the population density of region X is to be calculated, population density can be represented by a rule which declares that the density is the result of the population divided by the area. But where should this rule be declared? Not for each of the regions stored in the data base. Since the rule applies to each region, it is much more efficient to declare the population density rule for the class 'region', and declare each of the separate regions 'a kind of' region. A call for the population density of region X will then result in an inherit action that will access the class 'region', at which level the population density rule is found, which can then be applied to region X. Combining find, deduce, and inherit inference in one, general inference engine, yields a structure as displayed in Figure 4.3.

The deduce and inherit strategies will result in recursive calls to the inference engine. The application of a rule requires a new inference action for the required information referenced by the rule. By the same token,

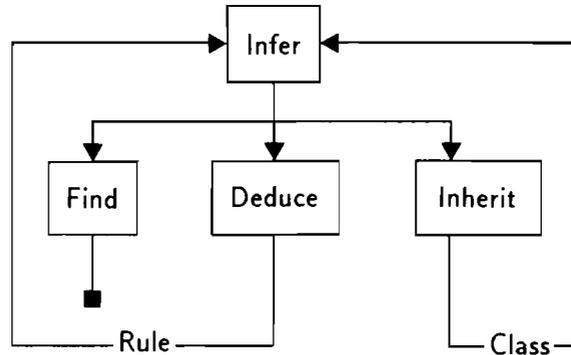


Figure 4.3: Inference engine with find-deduce-inherit inference

inference by inheritance implies inference actions on objects higher up in the hierarchy. From the scheme in Figure 4.3 it becomes clear that inference actions can only ‘bottom out’ by successful find actions. All other actions are, in the end, defined in terms of finds.

A fourth and rather brusque method of inference is that of simply ‘asking’ the user. In case neither find, nor deduce, nor inherit inference is able to generate the desired information, there is always the possibility that the user might know, or that the user might want to give an estimate or make the program simulate a certain value. The result of an ask action can then be stored in the appropriate frame item, so that the next time that this information is needed, it can be found without consulting the user. However, storing a value retrieved by means of ask inference can be dangerous, because it may threaten data quality.

Figure 4.4 displays the structure of a find-deduce-inherit-ask inference engine. Note that the ask-inference action has a separate status. It is not contained in the ‘trace’ alternatives (find, deduce, inherit). Instead, it is implemented as a separate action, not to be triggered by recursive calls from any of the other actions. The reason for this is that if the ask opportunity is treated as equal to the other three inference methods, the inference engine will exhibit rather silly kind of behavior. For instance, the failure of a find and deduce action for a value on the amount of limestone processed by some plant. Suppose furthermore that limestone industry is declared as an instance of the higher-class ‘industry’. If *ask* would be reachable via a recursive call by *inherit*, and if on the level of industry *find* and *deduce* attempts would again fail, then the system would not hesitate to ask the user how many tons of limestone is processed by industry. The conclusion

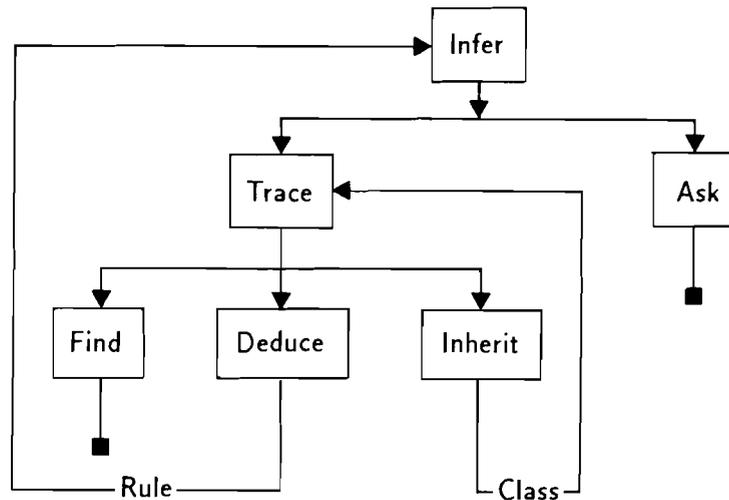


Figure 4.4: Inference engine with find-deduce-inherit-ask inference

is therefore that allowing ask actions to be triggered by recursive calls from inherit, generates either misleading or absurd questions at the wrong class level. Incorporating an ask opportunity as in Figure 4.4 avoids this problem. In this set-up, an ask action is only conducted if the complete set of find, deduce, and inherit attempts fails to come up with a solution. Therefore, the question asked of the user always pertains to the object for which the information was originally required.

4.4 A Prolog implementation

To implement all this on a computer an appropriate programming tool is required. The programming language Prolog is very suitable for the implementation of both a frame-based knowledge representation system and inference engine (for a brief introduction to the Prolog programming language, refer to Appendix 3). Prolog provides built-in inference and data base operations, a combination which is very useful for developing a relational matching system. Moreover, since Prolog is a logic programming language, it is very appropriate for implementing logical, conjunctive-disjunctive structures such as inference trees and matching rules. Prolog provides its own inference engine which is driven by matching, unification and backtracking. The find-deduce-inherit-ask inference engine described in Section 4.3, pro-

vides inference techniques that are of a somewhat higher level. They can, however, be implemented by means of Prolog's own inference strategy, i.e., they can be programmed in Prolog. Apart from knowledge representation and inference, Prolog is also well suited to representing a relational model (inference trees and matching rules), for implementing a matching program, and for implementing an explanation facility.

4.4.1 Frames and inference in Prolog

Cuadrado and Cuadrado (1986) provide a Prolog implementation that supports both frame representation and find-deduce-inherit inference in one program. The program provides the essential operations to be conducted on frames (storing, retrieving, removing, and replacing information). The retrieving is done by means of find-deduce-inherit inference.

```

limestone_industry(ako,value,industry).
limestone_industry('annual production',value,2000).
limestone_industry('employment costs',
                    if_needed,
                    rule_employment_costs).
limestone_industry('number of workers (m/w)',
                    if_needed,
                    rule_number_of_workers_(m/w)).
limestone_industry('typical employment',
                    if_needed,
                    rule_typical_employment).

```

Figure 4.5: A Cuadrado and Cuadrado frame in Prolog

The basic representation structure is simple. Every object-attribute-facet-value tuple is represented by a simple Prolog structure. Figure 4.5 contains a translation of the frame in Figure 4.2 into the implementation suggested by Cuadrado and Cuadrado. The clauses are of the form

```
object(attribute,facet,value).
```

Although the idea of using this kind of simple Prolog structures for implementing frame items is all right, it is a little surprising that Cuadrado and Cuadrado use the object as the functor, the predicate name of the Prolog clauses. Both from the perspective of predicate logic and Prolog

programming, this solution should be rejected. In predicate logic objects are represented either as constants or variables, but not as predicates. In a frame representation, the predicates should be reserved for representing the attributes. The objects should be represented by one of the predicate's arguments:

```
attribute(object,facet,value).
```

Further, from the perspective of Prolog programming it is unwise to use the object as the predicate. Prolog performs actions on its data base by accessing predicates. It can access clauses exclusively by reference to their functor, the predicate. To illustrate the inadequacy of the Cuadrado and Cuadrado implementation, take the following two Prolog clauses:

```
limestone_industry(ako,value,industry).  
brick_manufacturing(ako,value,industry).
```

If one wants Prolog to figure out the derivations for the following two statements:

1. Which is the parent frame of limestone_industry?
2. Which are the descendants of industry?

The first question could be submitted to Prolog like this:

```
limestone_industry(ako,value,Parent).
```

Prolog will reply with:

```
Parent = industry  
yes.
```

But then the second question: just on the basis of the two clauses for limestone_industry and brick_manufacturing, there is no way Prolog can ever find a derivative for this question. Put a little differently, it is impossible to submit a valid clause for answering the question. Or in terms of predicate logic, there is no well formed formula (wff), the derivative of which yields an answer to the question. The reason is of course that the answer to the question consists of predicates (limestone_industry and brick_industry) rather than arguments or clauses of which the predicates are already known.

Predicates are the keys to the statements or clauses in predicate calculus. If the predicate is unknown, there is no way of accessing a clause, something which seems very reasonable for a language based on predicates. Treating the objects of a relation such as ‘a kind of’ as predicates thus makes it impossible to bind those objects to a variable in a clause, simply because predicates cannot be variable at the moment a matching clause has to be found. Apart from this somewhat formal reason, there is also a more intuitive one for rejecting the solution suggested by Cuadrado and Cuadrado. What the clauses purport to say is that limestone industry and brick manufacturing are instances of the class industry. As such there is an ako-relation that associates two objects which each other; the parent class and the instance. Therefore, it seems reasonable to treat the relation as the predicate, and the objects as its arguments:

```
ako(limestone_industry,value,industry).
ako(brick_manufacturing,value,industry).
```

Question 1 can now be submitted to Prolog as:

```
ako(limestone_industry,value,Parent).
```

Question 2 can now be formulated as:

```
ako(Descendant,value,industry).
```

This second question will generate two answers:

```
Descendant = limestone_industry,
Descendant = brick_manufacturing
yes.
```

Although this modification of the Cuadrado and Cuadrado implementation offers many advantages, there is also a (modest) price to be paid. Using attributes as functors implies that the concept of a frame as a dynamically constructed set of tuples with identical object slots becomes somewhat obsolete. To illustrate this a Prolog clause could be written which gets the name of an object as its first argument, and which instantiates its second argument to a list of tuples containing the object in its object slot:

```

get_tuples(Object, Tuple_list) :-   condition_1,
                                   condition_2,
                                   .
                                   .
                                   condition_n.

```

Unfortunately, this predicate cannot be defined in Prolog when using the modified frame representation. The reason is that what has to be instantiated in the variable `Tuple_list` concerns attributes, which in the implementation suggested here, are functors. Unknown predicates preclude access to clauses; the `get_tuples` predicate cannot be defined. However, the advantages of an implementation with attributes as functors certainly outweigh the disadvantages of an implementation with objects as functor. Therefore, the inference engine in the Cuadrado and Cuadrado implementation was modified, and various other aspects were added. One of these modifications was that objects are represented in lists, rather than in atomic form, for instance:

```

predicate_a([object_1], value, v1).
predicate_a([object_2, object_3], value, v2).
predicate_b([object_1, object_2], value, v3).
predicate_b([object_3], value, v4).

```

The advantage of putting objects in lists is that only one clause is needed to store all objects that have identical scores on a certain attribute. Especially in the case of categorical variables this can lead to a significant reduction of the data base. Of course one has to equip the inference engine with extra knowledge so that it handles the lists correctly. Removing the item for `object_2` for `predicate_b`, for instance, implies that `object_2` is removed from the list in the associated clause for `predicate_b`. Likewise, if the value of `predicate_a` for `object_2` is changed into `v1`, `object_2` must be removed from the list in which `object_3` is declared to have a value `v_2`, and must be added to the list in which `object_1` is declared to have a value `v1`. Another enhancement of the Cuadrado and Cuadrado implementation was the addition of a value 'unknown' to be used in cases where values are really unknown. Especially in instances where the matching program needs some information, unknown scores can be very convenient (see Chapter 5 for a discussion of how unknown scores can be treated in the relational matching process). Introduction of 'unknown' scores in general provides an opportunity to distinguish the different ways in which a Prolog program generates the verdict 'false'. Prolog's

closed-world assumption guarantees that everything that cannot be derived is automatically considered false. Adding special scores such as ‘unknown’ offers the possibility to discriminate between the various ways clauses can fail to be derived. Apart from these and other extensions to the frame representation, the inference engine was extended with an ask-inference opportunity and multiple inheritance. This modified inference engine was equipped with sufficient intelligence to handle the ask-(multiple) inheritance interactions as mentioned above in an appropriate manner. Reasoning with exceptions was not implemented.

4.4.2 Inference trees and matching rules in Prolog

As was mentioned before (refer to Table 3.6 and Table 3.8), inference trees can be represented as nested lists. As it happens, the nested lists presented in Table 3.6 and Table 3.8 are not only some general form of nested list representation, they are also perfect Prolog. Together with a special inference tree processing program, they can be used in the matching process. The trees in Table 3.6 and Table 3.8 are one-dimensional. They can be considered inference trees belonging to the lowest level of dimensions in a matching problem. Like inference trees, the dimensional structure of a relational matching model takes the form of a tree. Again this can be represented in the format of (a set of) lists. Figure 4.6 contains both the dimensional tree of Figure 3.3 and its representation as a set of ‘linked’ Prolog lists.

Since matching rules consist of a conjunction of disjunctions they are easy to implement in Prolog. It was decided to choose a representation that is somewhat different from the frame representation discussed above. Figure 4.7 shows two examples of matching rules.

The second argument of the ‘rule’ predicate specifies the terminal node of an inference tree. A terminal node ‘dummy’ represents an inference tree without activity characteristics. This second argument is used by the matching program when it searches for matching rules to trigger. In order to prevent the triggering of matching rules that belong to dimensions other than the one that is currently matched, the rule predicates have a first argument specifying the dimension. The variables Y and X in the rule clauses represent the activity and the object that are currently matched. These variables are used to pass the activity and object to the matching rule, so that the `v_get` predicates can conduct the calls to the frame data bases in order to infer the desired information. The `v_get` predicates collect and accumulate the information needed for the final requirement test. Requirements

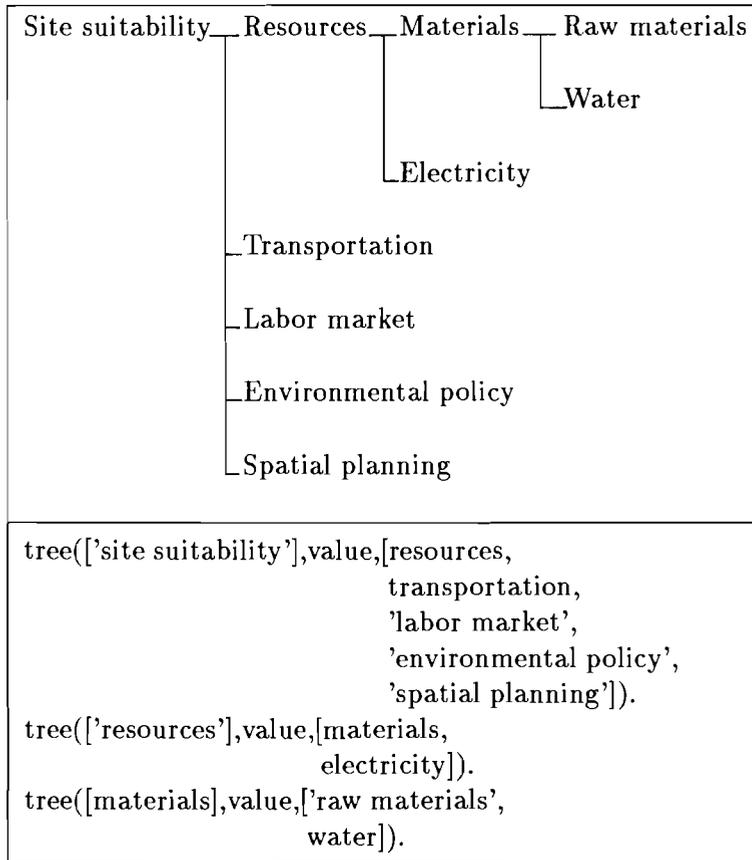


Figure 4.6: Dimensional structure in graphical form and as a set of Prolog clauses

are represented as lists. Multiple requirements form nested lists. In the prototype relational matching system, the implementation of matching rules as in Figure 4.7 was used. This implementation, however, is an ‘expensive’ one, because it implies a rather inefficient way of representing the disjunctive element of a matching rule. Each of the matching rules in Figure 4.7 only contains conjunctive elements. They represent one possibility for satisfying some type of production requirement for some kind of activity. But as discussed in Chapter 1, many requirements can be fulfilled in more than one way. Figure 4.8 shows an example of how this disjunctive element is represented in the prototype relational matching system.

```

rule('raw material','source orientation'(Y,yes,X),
  [[v_get('bauxite needed',Y,Baux),
    v_get('bauxite production',X,Prod),
    Baux > Need],
  [v_get('limestone reserve',X,Lime),
    v_get(Lime > 500.0)],
  [v_get('limestone % CaO',X,Cal),
    Cal > 52.0],
  [v_get('limestone % MgO',X,Mag),
    Mag =< 1.50]]).

rule(environmental planning,dummy(Y,-,X),
  [[v_get('water needed',Y,Need),
    v_get('water available',X,Avail),
    Avail > Need],
  [v_get('total water consumption',X,Wcons),
    v_get('agr. water consumption',X,Acons),
    Acons/Wcons < 0.80]]).

```

Figure 4.7: Examples of matching rules

Clearly, in case the first rule fails to be satisfied by the location, the second rule is tried. The problem here is that both rules have an identical first (conjunctive) part. The second part of the matching rules represents the disjunctive element. Treating both rules as separate possibilities to satisfy a matching requirement thus implies that the first part of the rule is executed twice. This is expensive and unnecessary. An alternative implementation is shown in Figure 4.9. This implementation contains explicit declarations of the disjunctive (or) and conjunctive (and) elements of the matching rule. Together with a matching program that understands this kind of structure, the implementation becomes a lot more efficient, and the speed of matching can increase significantly. Future versions of the current system may therefore contain this augmented representation of matching rules.

4.4.3 The matching program

Relational matching is the procedure that takes a relational model in the form of a dimensional inference tree with matching rules, traverses them, and

```

rule('spatial planning',dummy(-,X),
    [[v_get('total energy consumption',X,Cons),
      v_get('total area',X,Area),
      Cons/Area =< 195],
     [v_get('total retail value',X,Retail),
      v_get(population,X,Pop),
      Retail/Pop > 0.039]]).

rule('spatial planning',dummy(-,X),
    [[v_get('total energy consumption',X,Cons),
      v_get('total area',X,Area),
      Cons/Area =< 195],
     [v_get('gross agr./ind. product',X,Prod),
      v_get('gross county enterprise product',X,Gep),
      Prod/Gep > 0.33]]).

```

Figure 4.8: Prototype disjunctive matching rule

```

rule('spatial planning',dummy(-,X),
    and([[v_get('total energy consumption',X,Cons),
          v_get('total area',X,Area),
          Cons/Area =< 195],
         or([[v_get('total retail value',X,Retail),
              v_get(population,X,Pop),
              Retail/Pop > 0.039],
             [v_get('gross agr./ind. product',X,Prod),
              v_get('gross county enterprise product',X,Gep),
              Prod/Gep > 0.33]])]))).

```

Figure 4.9: Disjunctive matching rule with explicit conjunctive and disjunctive declarations

tests a set of objects on the requirements found in the matching rules. As with the object and actor data bases, the inference engine, and the relational model itself, this component of the relational matching system was written in Prolog.

The basic structure of such a program is displayed in Figure 4.10. It consists of three stages. First, all the necessary information for conducting a

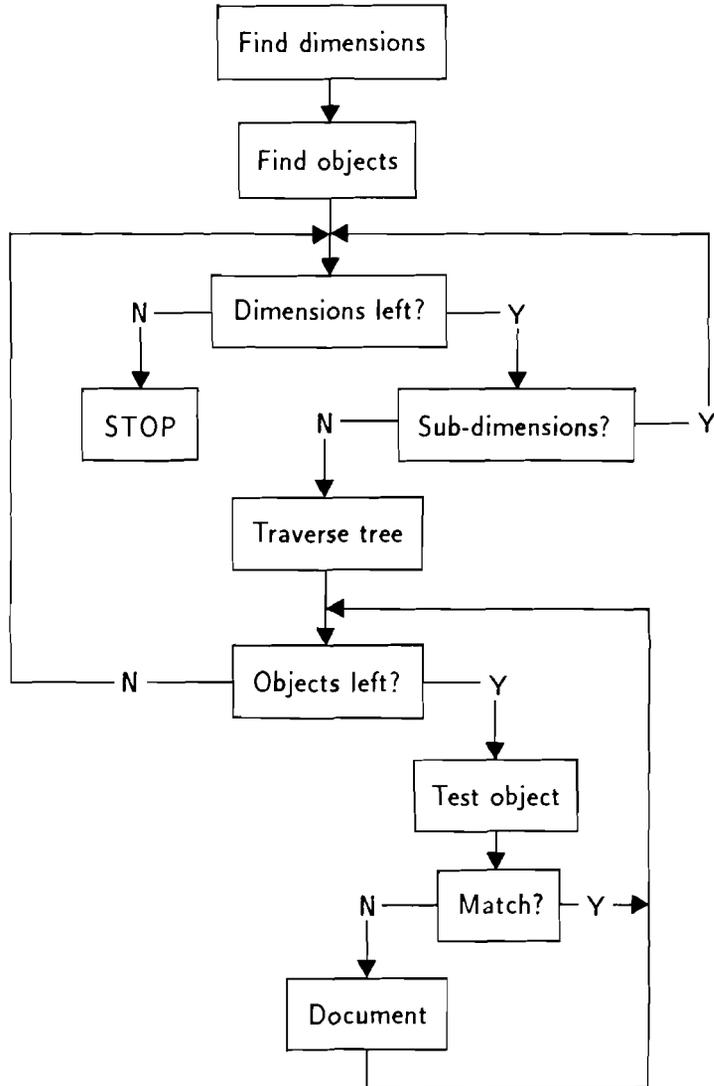


Figure 4.10: The matching program

matching is collected; the dimensions to be contained in the matching, and the set of objects that must be matched. Next, a nested looping structure is set up. Dimensions are sequentially processed, and within a dimension-specific cycle, objects are sequentially matched, after the associated inference tree is traversed and the matching rule(s) are known. In case dimensions are

nested, nested dimensions are processed before going on to the next dimension. If the only thing one would be interested in was the failure or success of an object in satisfying the requirements of an activity, it would have been much more efficient to put the dimensional loop inside the loop over the objects. This would have established that once an object fails to pass a test, it does not have to be checked on other requirements anymore. Once an object fails to pass a matching rule, its result on the total matching must be negative. From Figure 4.8, however, it can be inferred that since the object loop is contained inside the dimensional loop(s), every object is always matched on every dimension. Again, if only the failure or success of an object on the total matching problem is of interest, a lot of the testing and matching is unnecessary. But as was mentioned before, knowledge of why an object does not match can be very important. Schilling's 'diagnostics' application is based on the availability of data as to why certain alternatives do not pass the matching. Clearly, a proper explanation of why an object does not pass the matching should contain all the requirements that were not matched by the object rather than only the first one. Or put a little differently, the explanation of why an object fails to get through the matching successfully should not depend on the order in which dimensions are processed. In Section 4.4.1 it was mentioned that by adding special values such as 'unknown', programs can distinguish between different reasons for failure to infer certain bits of information. In the matching program, this is done by the 'v_get' predicates present in the matching rule. The v_get predicates keep track of the occurrence of unknown values.

4.4.4 Explanation

If a program needs to be able to explain what happened during its execution, and that is in some sense what the explanation of the matching results is about, then it needs to keep track of its own execution. When programming in Prolog, there are basically two ways of achieving such self-monitoring behavior. One can make the program store information on its behavior in the data base, and later on use that information for explanation purposes. The other possibility is to store information in lists and structures, that the program keeps carrying with it during execution. Upon finishing the execution this information can then be put into the data base or just handed over to another function (predicate) as one or more arguments. For an example of the latter, refer to Sterling and Shapiro (1986; pp. 313–314). Here, the former alternative was chosen for various reasons. First of all, it is easier to

program. Carrying information along while updating it, when necessary, is quite a complex task. Putting it in Prolog's data base and removing it from there again once it has been used for explanation purposes, is a lot easier to accomplish. A second reason was that if the set of objects becomes large—in the empirical application presented in Chapters 5 and 6 up to 107 locations are contained in a single multi-dimensional matching—the information to be carried along becomes rather unwieldy. This increases complexity and the program's appetite for (stack) memory, especially when this information is to be carried along in deep recursive procedures. Therefore a decision was made in favor of a simple and fast, albeit not entirely elegant solution of storing information in the data base during matching. This is done by the 'document' component in the matching program (Figure 4.10), and by the `v_get` predicates that 'make notes' in case information is found to be 'unknown'. In case an object fails to pass a requirement, information about the route followed through the dimensional structure as well as the fail information is stored in the data base, using the frame representation and frame driver. If the user requires an explanation on why a certain object was not considered a feasible choice alternative, a separate explanation program, again written in Prolog, is executed. The only parameter this program gets passed is the name of the object for which matching explanation is requested. The explanation program then searches Prolog's data base for the matching information of this specific object, grabs it, analyses it in a dimension-specific way, then searches for the associated matching rules, and finally parses them into a comprehensible format akin to English. The result can then be shown on a computer screen or stored in a file.

4.5 Implementing generalization

Although an integral part of the matching program, the implementation of the possibility to conduct relational matching with pre-specified generalizations needs separate attention. Although the basic idea of generalization is simple—the replacement of nodes in the inference tree representing an actor attribute by a 'don't care' value, such that all paths beneath that node become possible paths—implementation is somewhat troublesome because it implies the modification of the inference trees prior to, or during, the matching. One of the problems, for instance, is the kind of method or algorithm to be applied for finding the set of matching objects. In Section 3.8.3 it was mentioned that the solution set associated with a particular generalization

can be derived by means of the union of the solution sets associated with each of the instances of the abstraction. However, computation of the solution set by that method would require as many matchings as there are instances, after which the various sets need to be dragged through a series of union operations. A much easier and more efficient way of carrying out matching with generalization can be achieved by making use of the equivalence of set union and logical 'OR' declarations. Instead of conducting a matching for each of the instances and then computing the union of the associated solution sets, one could just as well try to match an object, first for the first instance, if this does not work for the second, if that does not work for the third and so on, until a match is found, or until all possibilities are exhausted. Repeating this procedure for all objects in the initial set will then generate the same result as the union-alternative, but the procedure is a lot more efficient. The more so, since it can be elegantly implemented by making use of Prolog's backtracking facilities. An example is shown in Figure 4.11.

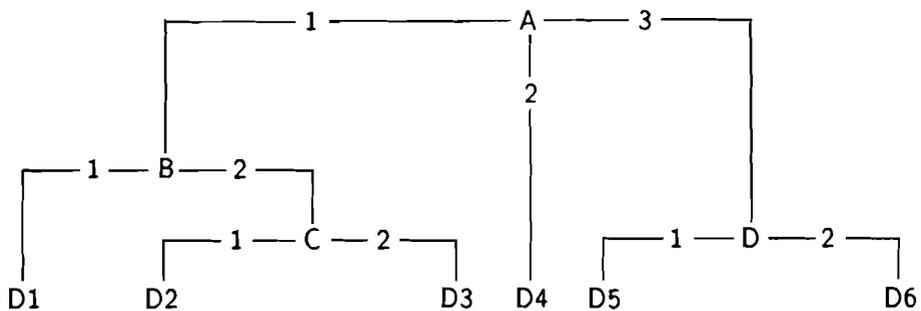


Figure 4.11: Traversing an inference tree containing 'don't care' values using backtracking

Suppose that the actor attribute B would be replaced by a 'don't care' value, and the resultant tree would be submitted to Prolog. Suppose further that the activity scores a 1 on attribute A. Prolog would then proceed to node B and first try path B1 leading to the matching rule D1. If the object would be found not to match the requirements of D1, and if Prolog would be allowed to backtrack, its next attempt would be the path B2. Its first possibility there would be path C1 with matching rule D2. If this still does not result in a match, Prolog would backtrack to node C and attempt path

C2 with rule D3. This way Prolog would try all possible paths beneath B in order to find a match, and this is exactly what one wants.

As can be inferred from what has been said about the implementation of inference trees before, things are a little more complicated from an implementational point of view. Although the tree itself is in Prolog, it is not an 'executable' Prolog program of its own. The processing of the trees has a meta-character in that the matching program is the program that interprets the trees and processes them. Therefore, the backtracking was integrated into the matching program as a special way of processing the inference tree. When the matching program traverses the tree, it checks at each node whether or not a 'don't care' has been declared for that node (refer to Section 5.7.6.2 for a description of how these 'don't care' values are set). If it has, the backtracking alternative is used. If not, normal non-backtracking traversing is applied. Application of the backtracking alternative generates some additional problems. These have to do with attributes of which the values are calculated on the basis of the attribute(s) for which 'don't care' values have been declared. For example, if the main source of energy is a node in the inference tree that can take the values 'coal' or 'natural gas', and if the derived variable is 'energy costs', then this derived variable will take a different value if the source of energy changes. But if the energy costs turn up in one or more matching rules, then they must be recalculated for each different path under the generalization node. To solve this problem, it was decided to 'simulate' values for the 'don't care' attributes, the values being those that are associated with the path followed. For the example in Figure 4.11 this would mean that for the path B1, a value 1 for B would be assumed, whereas for paths C1 and C2 a value 2 for B would be assumed. The matching program actually puts these values into the data base, so that the inference engine will 'think' that is the value for that specific attribute. For each new path, the value is changed, and after all the paths have been processed or after a match has been found, the original value is reinstalled.

4.6 Conclusion

The five components of the system discussed in this chapter together constitute a complete, though minimal, version of a relational matching system. Objects and actors can be represented with frames written in Prolog. The inference engine can manipulate the information in these frames. Inference trees and matching rules can also be written in Prolog, just like a matching

program that conducts the actual matching. Prolog offers ample opportunities for writing programs that monitor their own behavior, and as such the matching program can document special cases it encounters during a run. This information can then again be used by an explanation program, also written in Prolog. The main task of this program is to figure out what went wrong with the matching of certain objects, and to present this information in an understandable format to the user. When all the components shown in Figure 4.1 are implemented, the matching program integrates them into a system that can be used for relational matching. In principle, this is sufficient for conducting relational analysis, but it would not be sufficient for building meaningful empirical applications. Earlier it was suggested that the matching be made an integral part of the knowledge acquisition process. Execution of the matching model by a relational matching system gives results that must be tested against real situations or against the views of domain experts, preferably by means of using a reliable test set. In order to carry out such an iterative procedure efficiently, a lot more than just a core relational matching system is needed. Model results must be easy to interpret. Both activity characteristics and object properties must be easy to modify and update. Explanation must be something that the user can request from the system whenever required. If one wishes to apply relational analysis for spatial decision making, one needs the spatial perspective in the form of maps; single maps with model results, or multiple maps in which the model results can be compared with regional characteristics. These and other issues become very important if relational matching is to be used in empirical applications, both as a modeling tool, and as a consultation system. This brings us into the realm of decision support systems.

Chapter 5

Decision Support Systems and the REPLACE System

ABSTRACT

Decision Support Systems (DSSs) are discussed, with attention paid to what they are about, their function and character, as well as the role they could play in model-based development planning, such as regional planning. Most important in this respect is the issue of how various models can be integrated into one system such that they represent various perspectives on the empirical system that is modeled. The argument is illustrated by a case study of a DSS built to support the reorganization of a regional economy in the People's Republic of China. The relational site suitability matching system REPLACE (RELational Plant Location and ACquisition Enquiry) is part of that DSS and is an extension and application of the approach discussed in Chapters 3 and 4. Its characteristics, form and content, as well as its conceptual relations with other parts of the overall DSS are discussed in detail in the second half of the chapter. Some preliminary results of implementing a two-stage choice model are also presented.

Keywords: decision support, data bases, models, optimization, evaluation, user-interaction, model-based decision support, Shanxi Province, REPLACE

5.1 Introduction: model use and decision support

Many implementations of scientific models suffer from the problem that although the models themselves might provide useful and accurate views and abstractions of reality, they are difficult to apply in empirical research. Ample reasons can be thought of. Many models need large amounts of data input, some of which are difficult to collect, and some of which require so much domain knowledge, that the model can only be used by experts. In order to calculate the track a specific type of pollution will follow through a groundwater aquifer, for instance, the spatial distribution of the permeabilities of the bedrock need to be known in great detail. Calculation of the development of the water resources in a regional system over a series of time steps also requires the availability of large amounts of often very specific data. Specialist knowledge is also required when dominance factors for some forms of multi-objective, multi-criteria evaluation models have to be determined.

Similar problems surface with regard to the output of the models as well. Many spatial models, for example, generate output in the form of tables. It can, however, be rather difficult to find the proper associations between a complex table or set of tables and a spatial, map-like picture of the same output. Mapping model output, in its turn, generates further problems. If colors are to be used to represent a region's value on an attribute, for example, continuous variables must be classified in discrete categories. Of course default categorizations can be used, but a smart mapping tool must be able to use different, user-defined ones as well. As with what is true on the input side of the model, the output can contain information that, although it could in principle be understood by non-domain specialists, is highly technical and in a format that only the specialist, and sometimes only the model builder, can interpret. But even if the proper data can be acquired and even if the specialists who know how to provide the model with the correct parameter values and who can read the model's output are available, many practical problems remain. If a model needs a lot of data input and if this requires considerable work maybe even by the specialist, application of such a model becomes a cumbersome, and expensive, procedure.

These problems become worse when the same model needs to be run many times to solve a specific problem. This is the case, for instance, when the input data of the model represent several alternative options or lines

of action that can be chosen by a decision maker, or that may occur. The output then shows what the consequences of that initial situation or the decisions taken over a certain period of time will be. This situation is often denoted as 'what if' or 'scenario' analysis. Given a model describing a system, what happens if a specific initial situation is set, or, if given an initial situation, some of the process data are changed. For example, take a model describing what happens to the atmosphere over a specific region at time $t+x$ if somewhere at time t a certain load of pollutant is emitted. What if this is done in December? How different would this be in June? What if the pollution is not emitted at one particular moment, for example due to an accident, but is instead emitted in a continuous flow so that it gets evenly spread over a certain period of time? What do differences between December and June look like then? For this type of model use, the model may need to be run many times, each run representing different initial states of the system, while for each of the runs data must be modified and output interpreted. Similarly, for a relational site suitability matching model, one might want to ask many what-if questions such as what happens with the site suitability pattern if cities of over a million inhabitants experience a population rise of 10 per cent? Or what happens to the site suitability pattern if an activity decides to increase its production by 20 per cent? To find the answer to each of these questions, modifications must be made in either the locational or the activity data bases.

Problems can get even more complicated in case more than one model is included in the analysis. Integrating more models into one analysis does not only introduce the prerequisite demand that different models need to be able to 'understand' each other, but getting all the data right, putting it into the right places and formats, changing the output of one model into an input that the next model can handle, becomes a procedure which, if not automated, can seriously depreciate empirical applications, simply because there is a lot of work involved which is sensitive to mistakes. The consequences of this can be rather portentous. A lot of work spent on modeling remains unused and many multi-disciplinary approaches to problem solving are delayed, if not inhibited, because of the practical difficulties in applying it. Especially for problems of a typically multi-disciplinary nature, integration of several models constitutes a worthwhile endeavor. It is much like integrating several points of view, i.e., several scientific approaches, into one operational system that can handle the joint results of these individual analysis tools.

Clearly, all this requires that model input, model output, and model runs must be made easy and efficient. Not only because it is the only way

to conduct this type of integrated approach to problem solving, but also to let the non-specialist use the models for the purpose they were originally built. Another reason for making models more accessible is that it may well be the only way to make people want to use them. Data preparation and output reformatting and mapping can be tedious and as with any kind of monotonous work it is prone to generate errors. Especially when the format and data source/destination transformations can be done in a rigid manner, computers can do that kind of work a lot faster and more reliably.

Making models easier to use by taking away a lot of tedious work and many awkward tasks also has other advantages. Not only does it become possible to use a much larger variety of models and approaches, but it can be expected that the possibility to conduct a large number of model runs without too much effort to be invested in data preparation and output handling will enable the model builder to pay more attention to the development and improvement of the model itself. In Section 3.9.2 it was suggested that the empirical content of a relational matching model be determined by an iterative process in which model tests form a necessary component. Conducting model tests implies input preparation and output explanation. It also implies many model runs each of which can be more or less different from the previous ones in order to find when model results start to change, or where critical points occur. Clearly, such a procedure requires an efficient interfacing with the model. Relations between the content and character of a model and the possibilities of its application in situations of decision making are important issues in decision support systems research.

5.2 Decision Support Systems: an overview

“A decision support system, DSS for short, is a *computer-based information system* that *helps a manager* make decisions by providing him or her with all the *relevant data* in an easily *understandable form*. As the user of DSS, the manager formulates the problem by using an *interactive* and probably *menu-driven* front end. The system then accesses a data base to locate the necessary data, utilizes a repertoire of mathematical and/or statistical *models*, and finally produces the desired information at the user’s terminal. The user can explore several ‘*what if*’ scenarios in order to arrive at a decision.” (Mitra, 1986; p. vii [my italics]).

Although this (functional) definition of a DSS by Mittra might not be the best possible one, it serves well as a point of departure for discussing some more or less common traits of DSSs, as well as for a discussion on some features that may be used not only in a managerial or organizational environment, but also in a more applied, scientific context. The description by Mittra contains three kinds of elements the discussion can concentrate on: why (aid decision making), what (goal state and constraint orientation), and how (computer-based, interactive, data bases, models, 'what if' scenarios). Each of these aspects is discussed briefly below.

5.2.1 Why: ill-structured problems

Several authors point out why DSSs were developed (e.g., Bosman, 1983; Mittra, 1986; Sprague, 1986; Keen, 1986; Brennan and Elam, 1986; Turban and Watkins, 1986; Davis and Grant, 1987; Densham and Rushton, 1988). When studying the ends and objectives of DSSs, a difference in opinion emerges regarding the possible domain of applications of DSSs. A somewhat narrow view regards the objective of DSSs to be "to improve the performance of knowledge workers in organizations" (Sprague, 1986; p. 10) or "to help improve the effectiveness and productivity of managers and professionals." (Keen, 1986; p. 48). According to this view, DSSs seem to be limited to applications in an organizational context, and are aimed at improving the performance of the organization by improving the performance of its members. As a consequence, most if not all, of the examples and applications of DSSs these authors mention are taken from the business or corporate domain. The associated tasks and theoretical design of the DSS are taken from this domain as well. The position taken by Mittra or Densham and Rushton (1986), Fedra *et al.* (1986, 1987), or Brennan and Elam (1986), however, shows a much wider perspective for the use and application of DSSs. They concentrate on the more general character of the tasks for which DSSs are developed. Perhaps the most important aspect of the objective DSSs are developed for is that they are supposed to assist in solving ill-structured decision problems: "Decision Support Systems are computer based systems whose objective is to enable a decision maker to devise high-quality solutions to what are often only partially formulated problems." (Brennan and Elam, 1986; p. 130). Bosman (1983; p. 80) characterizes ill-structured problems as problems for which at least one of the following conditions is not met:

1. The set of action alternatives is finite and identifiable. (A problem for which the set of action alternatives is identifiable is called 'well-defined'.);
2. The solution is consistently derived from a model that shows a good correspondence;
3. The effectiveness or the efficiency of the action alternatives can be numerically evaluated.

Simon (1960; p. 6) rephrases these formal criteria into a conception of unstructured (unprogrammed) problems as those which cannot be solved with a "...cut and dried method for handling the problem, because it hasn't arisen before, or because its precise nature and structure are elusive or complex, or because it is so important that it deserves custom-tailored attention". Fedra *et al.* (1986; p. 169) describe what this amounts to in practical decision-making situations:

"...there is a class of (decision) problem situations that are not well understood by the group of people involved. Such problems cannot be properly solved by a single systems analysis effort or a highly structured computerized decision aid. They are neither unique—so that a one-shot effort would be justified given the problem is big enough—nor do they recur frequently enough in sufficient similarity to subject them to rigid mathematical treatment. Due to the mixture of uncertainty in the scientific aspects of the problem, and the subjective and judgmental elements in its socio-political aspects, there is no wholly objective way to find a best solution."

5.2.1.1 Well-defined, ill-structured, and functional equivalence

Perhaps the term 'ill-structured' is a little unfortunate. Of course it cannot be that decision-making situations for which a DSS is useful are really unstructured. On the contrary, many decision-making situations do contain structure in the sense that the problem involves a number of aspects that can indeed be modeled. What makes the problem difficult, or even impossible to solve in a straightforward manner, is that its analysis contains various qualitative and arbitrary moments which stem from two, conceptually different

aspects of the fact that a decision task may be not well-defined. First there is a set-theoretic problem, namely that the amount of theoretically possible ways in which a system can be changed is virtually infinite. Second, the system and its possible directions of development can be not well-defined in the sense that identical combinations of some of the system's parameters can have an entirely different meaning and significance for different groups of actors or different pre-defined objectives. This aspect concerns the non-commensurability of various criteria and objectives. Both aspects amount to the problem of a not well-defined search space for finding solutions to the problem. The question is first, how to describe or 'construct' this space such that it becomes well-defined, and next, how to limit and search this space. If a methodology is applied that is based on the assumption that a system can be described unequivocally in terms of its empirical properties, the problem of a not-well defined system remains. Within such a conception there does not seem to exist a set of criteria by which both the problem space and the set of possible decision-induced impacts on the system can be meaningfully described and limited. Or in terms of the two-stage choice modeling advocated in Chapter 1, alternatives (different, decision-induced states of the system) may be evaluated on a common set of dimensions, but only if they are all part of the set of functionally equivalent alternatives. Their membership of this set is determined by a matching of their empirical properties with a disjunctive set of conjunctive requirements, representing the necessary and sufficient conditions for satisfying a specific objective by a specific actor. Describing a system in terms of end-means relationships may help the decision problem to become less ill-defined. Functionality provides the criteria the lack of which creates problems when using an approach based on the assumption of intrinsic properties of objects. Modeling functionality means that the relevance of the alternative's properties is the result of how a specific objective can be satisfied.

How can this idea of defining sets of action alternatives by means of functional equivalence be incorporated or modeled into a DSS? To some extent this is a matter of scale. It depends on what is defined as an objective and what is considered the means. If the objective is to expand a region's economy, the number of ways in which this can be done is enormous. Here it is the complexity of the problem, as mentioned earlier by Fedra *et al.* (1986), that makes it difficult to model functional equivalence. One might as well try to model functional equivalents for the objective of 'living a happy life'! The objective is simply too abstract and represents such a vast choice set that it cannot be known nor modeled. Nevertheless, people seem to have

ideas on what plausible lines of action are, so why not leave the (implicit) construction of sets of functionally equivalent alternatives for such a complex problem as the planning of a regional economy to the decision makers themselves, rather than trying to model them into a system? The problem space can be considered a collection of nested sets of objectives and means at different levels of abstraction, the objectives of one level becoming the means of a higher level. Establishing a few power plants can be a means for the objective of expanding a regional economy. This, however, implies that they themselves can be the objectives in a site suitability model looking for functionally equivalent sites for the power plants. Whether or not the objective to establish power-plants was a means of satisfying an even 'higher' objective, is irrelevant to the site-suitability model. Similarly, whether or not the (simulated) decision to locate new power plants was an appropriate or perhaps even the 'best' option, is something that must be decided at the appropriate level. This can be done by either another model based on functional equivalence to decide whether it is 'appropriate' and an optimization model to decide which is the 'best', or by a decision maker who represents his own model and who uses the decision support system as a means to gain better insight on detailed, less complex problems for which models are available. At those lower levels these problems form objectives; on the decision maker's level they represent complex means. Of course both points, finding 'appropriate' and finding 'the best' solution, are related. Finding appropriate solutions is a matter of functional equivalence. The 'best' solution—by analytical and objective standards—is always a solution under a number of assumptions concerning the common criteria on the basis of which the optimization is carried out. Once the determination of these assumptions makes up a considerable part of the problem, the resolution of the problem requires that the variability of these assumptions becomes an integral part of it.

5.2.2 What: constraints and pre-defined goal states

Perhaps the first of the two key characteristics of DSSs concerns these more or less arbitrary moments of the decision-making process; not, however, the complexity of the problem. Although a simulation of what happens if somewhere in Central Europe a load of dangerous chemicals is spilled into the river Rhine may precipitate decisions on new permits of who may drain how much of what substances into the river next year, the simulation only becomes a part of a DSS if there is a possibility to include 'arbitrary' assumptions or constraints such as, for instance, what the maximum load of a certain chem-

ical in the Rhine may be. In other words, in a DSS the constraints within which a problem must be solved must be variable and available for setting by the decision maker. Both must represent aspects that are the subject of decision making in the real world.

The constraints constitute aspects that are important for investigating the possibilities for realizing an underlying, pre-defined goal state; a more or less composite objective or intention, such as increasing a region's employment possibilities while keeping pollution at a low level, or the location of a set of production units under the assumption of specific developments in the labor market. The goal state itself may be variable as well. In many instances there simply is no clear goal description. In those cases the DSS may assist the user in formulating one or several.

The second major characteristic of a DSS consists of objective functions that inform the user about the extent to which a pre-defined goal state can be realized within the limits of the constraints. This concerns the second stage of the modeling process suggested earlier. Again, it may very well be the case that the choice of the objective function itself can be subject to decision-making. Should the costs of a waste disposal site be calculated in money, or perhaps in the risks of spreading diseases? Are possible casualties more or less important than annual maintenance costs?

Summarizing, one could say that throughout the rest of the text the term 'DSS' is meant to refer to computer-based systems aimed at investigating the possibilities of realizing a pre-defined goal or objective, given a set of pre-defined constraints, each of which refers to aspects of the real world that are subject to decision making by the user. In short, multi-criteria, multi-objective complex problems. Note that in this context the term 'constraints' is reserved exclusively for those aspects of the problem that are, partly or wholly, subject to decision making. Constraints stemming from relationships and processes that are basically 'given' to the decision maker are included in the models contained in the DSS.

5.2.3 How: data bases, models, user-interaction

Most, if not all, of the overviews and introductions to DSSs acknowledge that computer-based decision support requires the implementation of data bases, models and a form of user interaction: data bases for representing and holding relevant information such as model input and output, and general information to be used by the decision maker, user interaction because without this nothing much would happen. But what of the model component?

Although there is general agreement that a proper DSS is model-based, there is less consensus about what these models actually are, or what they are supposed to represent. Again, differences of opinion evolve from the divergence of views on the use intended for a DSS. Generally, two conceptions of 'model' seem to predominate the DSS literature, each of them referring to a specific task the DSS should perform. The first use of the term 'model' is reserved for models in the traditional, empirical sense. Models, therefore, that represent part of the empirical world. They take simulated decisions as their input, and compute their consequences for a part of the world. Examples are simulation models for atmospheric pollution or groundwater or surface water pollution, economic input-output models, migration models, site suitability models, and so on. A second type of model that is often mentioned in connection with DSSs is a model representing the structure and decision-making processes in the organization itself. Figure 5.1 taken from Sprague (1986; p. 23), shows an example of this.

Clearly, the components of the 'models' sub-system refer to decision processes themselves. Sprague (1986; p. 22) states that "A very promising aspect of DSS is its ability to integrate data access and decision models. It does so by embedding the decision models in an information system which uses the data base as the integration and communication mechanism between models". The reason Sprague emphasizes decision models stems from the fact that he seems to adhere to the organizational/management interpretation of DSS. Strategy and tactics (Figure 5.1) are of course important issues in such an environment. Here, however, 'models' are supposed to refer to empirical models. The reason for this is that the main interest here is in possible applications of DSSs in spatial planning and spatial decision making on the level of spatial policy, e.g., regional policy. The decision aids to be used in the formulation of spatial policy measures, for example by a regional or national administration, are meant to provide as-objective-as-possible information on the consequences of these measures. Whether or not these consequences are acceptable and politically sound, realistic, or simply desirable, is a decision that may be very hard, or even impossible, to implement in a DSS. Brennan and Elam (1986) denote systems that contain explicit and accurate models of decision-making processes 'smart systems'. Applications in a societal context may require 'brilliant' ones.

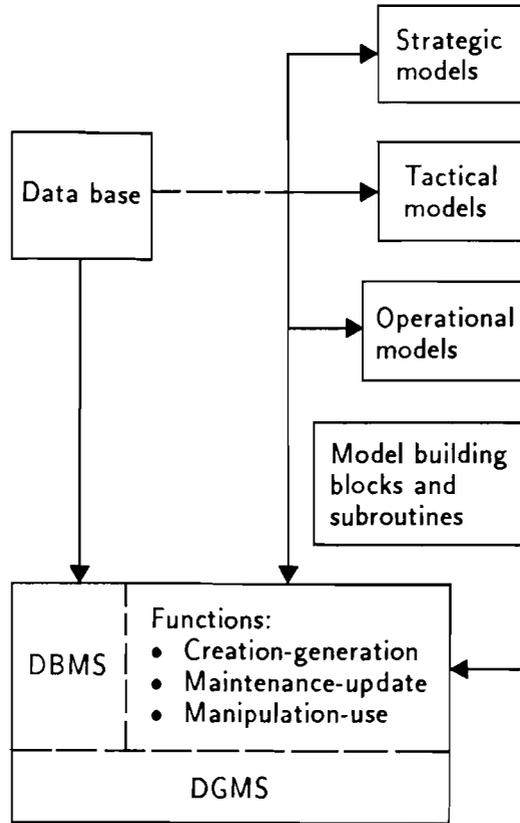


Figure 5.1: Conceptual architecture of a DSS (Source: Sprague, 1986)

5.3 Intermezzo: expert systems (ESs), DSSs, and spatial planning

The rapid increase in the popularity of ESs over the last, say five to seven years, made it inevitable that the subject would eventually turn up in geographical and planning literature. ESs can be considered a spin-off from artificial intelligence research. Once it was recognized that problem-solving capabilities depend to a large extent on the availability of knowledge, and to a smaller extent on inference capabilities, and once the basic techniques for combining these two into knowledge-based systems became widely available, a large, application-oriented market emerged. It became apparent that significant parts of knowledge about a small and specific domain held by human

experts could be modeled in computer programs combining factual knowledge with inference capabilities for tapping this knowledge. These programs could then solve domain-specific problems, just as an expert would. Experts moreover, tend to be expensive, take the knowledge with them when they leave, and sometimes make mistakes, whereas computer programs do not suffer from these 'disorders', so quite naturally, expert systems have become popular.

ESs 'invaded' geography and planning in two ways: as expert systems, i.e., as problem-solving, knowledge-based programs capable of carrying out specific kinds of tasks which are of interest to geographers and planners, and as a technology, offering opportunities for new types of modeling and analysis. Two types of application of the latter alternative, the use of knowledge-based systems as a new kind of modeling technique, can be singled out: the work by Smith (1983) and Smith *et al.* (1982, 1984). In these applications, ES technology is used for the development of computational process models of spatial choice behavior. Underlying the approach is the idea that specific types of knowledge-based systems form a model of the cognitive processes generating choice behavior. The structure of 'rule based' expert systems is a good example of such a cognitive structure. In its simplest form, such a system consists of a set of rules R, each of which is applicable in specific situations or states of the system S(t), and an inference engine which somehow matches this state against the states contained in the rules. If a match occurs, the rule is executed (A) after which the system enters a different state S(t+1):

```

IF    (RULE(IF  S(t)
              AND knowledge (R)
              AND inference (I)
              THEN action (A))
      AND S(t)
THEN  S(t+1).

```

This process continues until some desired or pre-defined goal state obtains. Smith (1983) and Smith *et al.* (1982, 1984) regard this type of structure as a model of the cognitive process underlying spatial choice. The system's states on which the rules are matched are denoted the actor's 'short-term memory' or STM, whereas the set of rules associating these states with the actions together with the inference engine are denoted the 'long-term memory' or LTM:

```
IF (RULE(IF STM(t)
          AND LTM
          THEN action)
    AND STM(t)
  THEN STM(t+1).
```

Translated into spatial choice terms, the *STM* is a model of the decision environment or decision situation, whereas the *LTM* represents such things as preferences and information processing or intelligence. In a series of articles the authors furthermore present inductive techniques and examples of how this *LTM* can be experimentally derived and modeled.

As explained in Chapter 1, however, relational modeling is based on a normative–deductive approach instead of one aimed at reconstructing the cognitive processes underlying spatial choice. Therefore, in relational modeling expert systems or knowledge-based systems are not used as models. Instead, only some of the technology applied in knowledge-based systems is used for implementing a relational matching model. Other examples of the use, or suggested use, of expert system technology for model building are the ones by Cullen (1986), Leary (1987), Timmermans and van der Heijden (1987), and Op 't Veld (1988). Tasks in which expert systems can be used as problem-solving programs, i.e., as expert systems proper, are limited mostly to rule-oriented problems. Typical applications are then the automation of classification in, for example, locational planning, for which the classification criteria are known. Examples of discussions on these kinds of ES applications can be found with Langendorf (1985), Davis *et al.* (1987), Ortolano and Perman (1987), and Fedra *et al.* (1987). From these, the positions taken by Langendorf and Fedra *et al.* offer a more integrative perspective. The role they suggest for knowledge-based systems is as a part of a larger DSS in which several of these systems can be contained. Knowledge-based systems simply represent knowledge-based models; special parts of the model base for which rule-based approaches are most appropriate. Similar suggestions for integrating ESs into DSSs are put forward by Turban and Watkins (1986), Coulson *et al.* (1987), and Winkelbauer and Reitsma (1988).

5.4 Decision support in planning a regional economy

‘The coordinated development of a region, and its industrial structure in particular, requires the simultaneous consideration

of numerous inter-relationships and impacts, e.g., resource requirements, environmental pollution, and socio-economic effects. Plans and policies for a rational and coordinated development need a large amount of background information from various domains such as economics, industrial and transportation engineering, and environmental sciences, in a readily available format, directly usable by the planner and decision maker. However, the vast amount of complex and largely technical information and the confounding multitude of possible consequences and actions taken on the one hand, and the complexity of the available scientific methodology for dealing with these problems on the other hand pose major obstacles to the effective use of technical information and scientific methodology by decision makers." (Fedra *et al.*, 1987; p. 1).

This quotation sums up what has been said about DSSs so far, and their possible application in regional policy: it addresses the situation in which the decision maker has the opportunity to (re)arrange a complete regional economy. Although many aspects of the regional economy in capitalist countries 'emerge' as the result of the decisions taken by many individual decision makers rather than as a result of an integrated planning effort, the point made in the statement by Fedra *et al.* also applies to many aspects of societies characterized by free economies. Regional and environmental policy, for instance, are areas for which a clear societal interest can be recognized. It is of course for this reason that administrations on different levels of authority try to maintain control over these segments of society. But even though the actual application mentioned by Fedra *et al.* may be confined to centralized economies, it constitutes an appropriate example for studying the general use and characteristics of DSSs in (governmental) spatial or environmental planning.

Clearly, a DSS for assisting the reorganization of a spatial economy conforms to the above-mentioned characteristics of DSSs in general. Fedra *et al.* mention the complexity of the problem, the multi-disciplinary character, the large variability in effects different measures may have, and the complexity of the information involved. Expressed in more general terminology: the problem is ill-structured or better, semi-structured. It contains a large number of decision-dependent, arbitrary moments (multi-objective), and alternatives must be evaluated on their contribution to a pre-defined goal state, which itself is subject to decision making (multi-criteria). Depending

on the chosen objectives, different aspects and questions become important. For example, if a decision to expand a coal-fired chemical industry is made, what products and production volumes should be implemented if a minimum of atmospheric pollution over densely populated areas is required, while attempting to achieve maximum efficiency in coal transportation and energy consumption? What demands would that cause for the labor market? How can the suitability for industrial production in a region that is inaccessible to certain types of industry be improved? And if it can be done, what does it cost and how would it change the overall regional input-output patterns? Formulating a regional economic policy also implies making decisions about the selection of sometimes contradictory or competing goals. What consequences does a specific production and investment plan have for the availability of water required for agricultural purposes? Is it possible to implement a production structure that minimizes a number of negative effects and maximizes the positive ones? Which are the positive and the negative ones in the first place, and how do things turn out regionally or for different groups of people? What should be the role of agriculture? If many more people will be absorbed by the industry, will this cause labor shortages in agriculture? If yes, would it be wise to decrease agricultural production, which might be possible because of the increasing revenues from industry?

It is clear that the range of possible questions is infinite; so are the possible combinations of objectives, policy measures, and their effects. The problem can be rephrased so that it at least sounds a lot nicer: '... the environmental planning process is an attempt to achieve greater rationality in the solving of problems and making of decisions on the protection, use, and management of the physical environment. Rationality here is a methodological rule which requires systematic and explicit integration of means and ends. It calls for the design and comparison of alternative means, in terms of the ends they are intended to serve.' (Davis *et al.*, 1987; p. 241). This abstraction must of course be associated with a (normative) model of environmental planning (Figure 5.2). Such a model provides a structured approach to tackling the complex problem mentioned above. Application of such a model guarantees neither its solution, nor the generation of a satisfying composite of decisions and their effects. But it does provide a framework, a method for tackling complex planning problems, as well as for a systematic search for feasible solutions.

The items in the 'problem-solving' box of Figure 5.2 can be associated with the basic characteristics of a DSS. Basic problems must be identified and defined, alternative strategies formulated. Different strategies must be

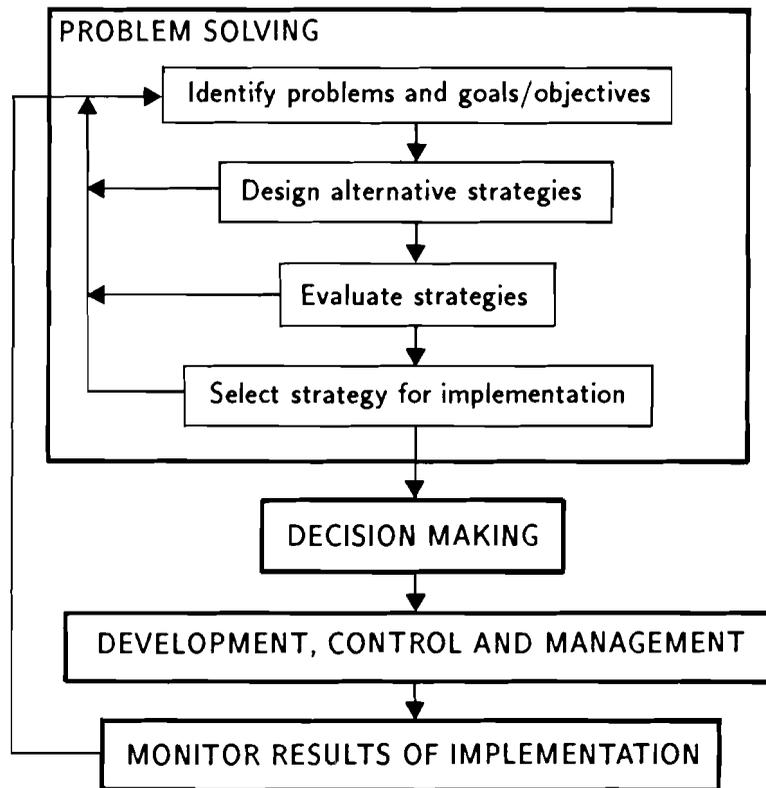


Figure 5.2: A conceptual model of planning (Source: Davis *et al.*, 1987)

separately evaluated, after which the best one available must be chosen. What is 'best' depends on the evaluation function(s), the criteria. This pertains to the multi-criteria character of the planning problem/DSS. The multi-objective aspect is covered by the determination of the problems and the strategies. Again, formulation and evaluation of strategies brings us back to the problem of modeling functional equivalence and the formulation of common dimensions of functionally equivalent alternatives. Strategies must be formulated according to the objectives they must satisfy and evaluation functions are chosen by the decision-maker and can be the object of political debate.

In Figure 5.3 the conceptual scheme for a DSS for planning a regional economy, or for 'integrated development' is shown.

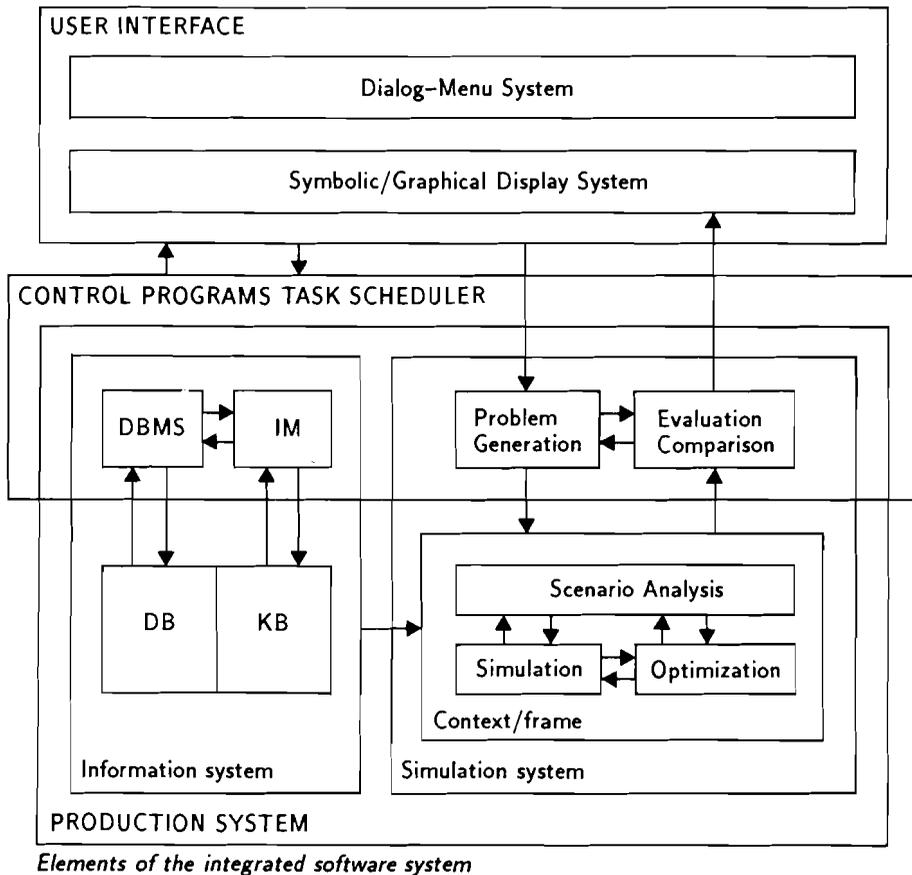


Figure 5.3: Conceptual architecture of an integrated DSS (Source: Fedra *et al.*, 1987)

The three basic elements mentioned earlier: data bases, models and user-interaction, are represented by the ‘information system’, the ‘simulation system’ and the ‘user interface’ components respectively. The simulation system contains the components for simulating decisions, their consequences for the environmental, economic, or social system that is modeled, and the optimization functions. The evaluation aspect, together with the problem definition, are contained in the ‘control programs task scheduler’.

5.5 Model integration

The simulation system in Figure 5.3, contains a set of models about the world, each of them representing a different perspective. A regional economy can be modeled in terms of input-output relations, but also in terms of transportation, migration, labor market conditions, water management, atmospheric or groundwater pollution, and so forth. Since each of these models offers a different perspective on the same empirical environment, while integrated planning requires that these aspects are somehow merged and blended into a final policy, each of them must be able to communicate with at least some of the other parts of the system. They need to have access to the data bases in order to find their input and deposit their output. Whether or not this inter-model communication is to be realized on the level of the models themselves is mostly a technical problem. In many cases where different processes are represented by different models, it will be sufficient that there is a common data base from which the models can both obtain their input data, and return results, that may then in turn be read from the data base by another model.

The data base thus represents the (simulated) regional economic system and its physical (geographical, environmental) configuration, modified by simulated decisions and models computing the consequences of these decisions. Consider, for example, a production plan, constructed on the basis of a dynamic input-output model (I/O; Zhang *et al.*, 1988) and an energy demand optimization model (MAED; Vallance and Weigkricht, 1988). The plan contains production volumes and product specifications as well as production technologies. Is this plan realistic? What, for instance, are the locational consequences? Where can the extra amount of, say, 500,000 units of fertilizer be produced, and where the aluminum? And what about the coal-based chemical industry? This locational problem may be tackled with a (relational) site suitability model which selects suitable sites for the specific types of economic activity, the results of which are then submitted (or read by) a model that optimizes the production distribution area (PDAS; Zebrowski *et al.*, 1988). The consequences of all this for atmospheric pollution may then be computed with a Gaussian dispersion model (ISC; Posch, 1988), whereas the effects on water management are taken into account by a water resources simulation model (MITSIM; Strzepek and Fedra, 1988). At each point during such a process, it might be interesting to evaluate the situation on the basis of a set of user-defined criteria and constraints (DISCRETE; Fedra and Zhao, 1988). If, for instance, the production plan implies

too much localized pollution of a specific nature, adjustments may be made. Of course they might increase the costs and lower the energy efficiency, but whether or not this is acceptable is a decision that only the decision maker can make. Each of these models might require data from other models, although some data can be obtained directly from the data base. Some models depend on each other, some just represent different perspectives of the regional economy.

5.6 The Shanxi Province DSS

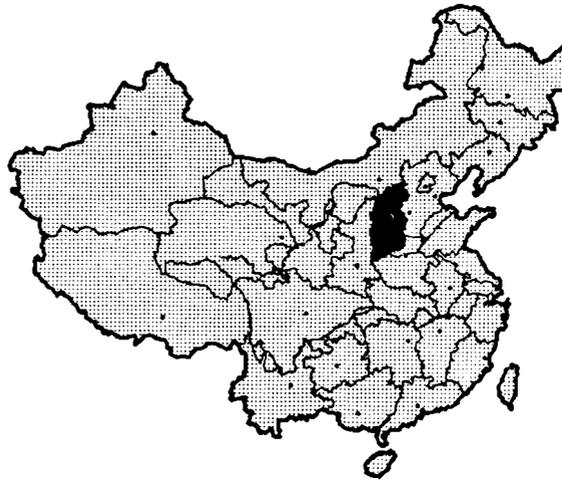
The example of a (partly) causal chain of decisions simulated by a multi-model simulation system such as just mentioned, is taken from a DSS developed at IIASA for developing the economy of Shanxi Province, the People's Republic of China. The overall problem situation addressed by the system is formulated as 'how to plan for integrated industrial development centered on a primary resource, namely coal, maximizing revenues from industrial production for a set of interdependent activities, subject to resource constraints and minimizing external (i.e., environmental) costs' (Fedra *et al*, 1987; p. 1). Before going into how this problem can be tackled by means of a hybrid approach to decision support, it is important to have some background information on Shanxi province, its main characteristics, and its major problems.

5.6.1 Shanxi Province: geography and development

A number of Shanxi's geographical characteristics and its recent achievements in economic development are outlined by Gao (1985), and its role in the interprovincial trade and development between 1957 and 1979 is discussed by Lyons (1987). The maps in Figure 5.4 show the position of the province in the People's Republic of China (a), and a few of its topographical characteristics (b). The maximum east-west distance through the province is about 384 kilometers, the distance from north to south approximately 682. The total area of land is 156,286 square kilometers, 25% of which is covered by arable land, and 10% by forest area.

Shanxi's physical geography causes groundwater and surface water systems to flow east, south and west. Because of great variations in elevation, temperature and precipitation, the water resources in the various areas exhibit strong variation as well. All the rivers in Shanxi flow outward. The drainage area of the Yellow river system with the Yellow river and the Fenhe river as its two largest contributors, covers 62% of the total area of the

(a)



(b)



Figure 5.4: Shanxi Province: location and some topographical features

province. The total annual volume of water resources in Shanxi is about 19 billion cubic meters, about 80% of which stems from river water and 20% from groundwater. Distribution of the total water supply and degree of water exploitation are displayed in Figure 5.5.

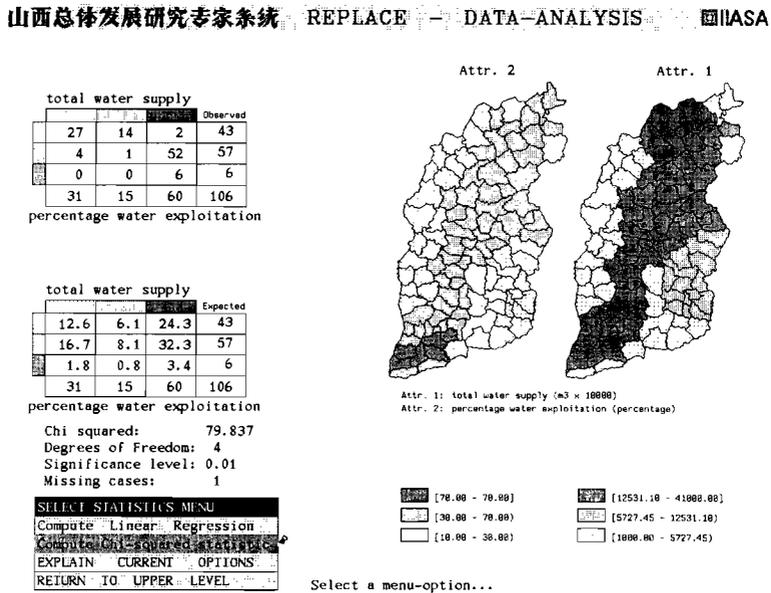


Figure 5.5: Shanxi Province: total water supply and water exploitation

Shanxi Province contains a total of 87 different types of mineral resources, the most important ones of which are coal, iron, copper, aluminum, molybdenum, titanium, lead, gold, silver, cobalt, and limestone. Shanxi has the largest deposits in China for 16 of these. By far the largest of these resources is coal, the fields spreading over 80 of the 107 counties in the province, covering 56,700 square kilometers, 36.5% of the total area. The prospected reserves are 900 billion tons, 203.5 billion tons of which are proven. The total of Shanxi's coal resources accounts for about one third of the total coal reserves in China. On a global scale this implies that Shanxi contains approximately 3.0–3.3% of the world's coal reserves. Moreover, Shanxi's coal is of a superior quality in most of its varieties, is contained in stable seams, and is located close to the surface.

Figure 5.6 shows both coal production and reserves, their respective spatial distribution and statistical relationship. Both variables were categorized

山西总体发展研究专家系统 REPLACE - DATA-ANALYSIS IASA

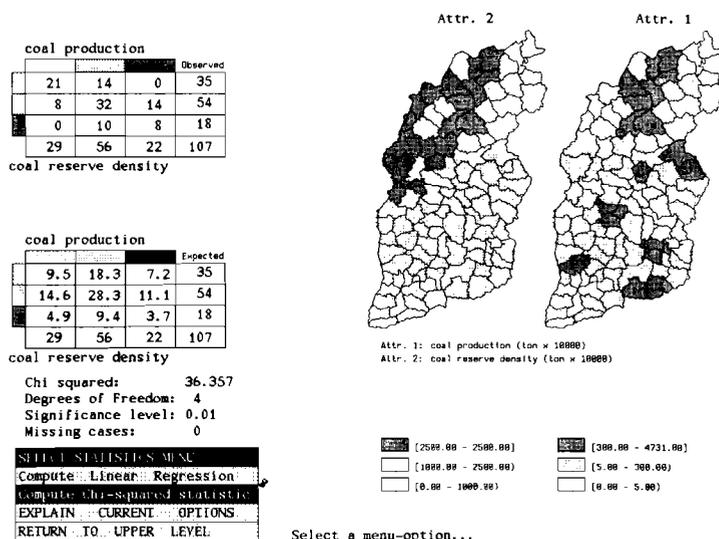


Figure 5.6: Shanxi Province: coal production and coal reserve density

because one of the variables (coal production) was only available in a classified format. The largest producer of coal for power generation in China is Datong county (38.6 billion tons reserves) in the north of the Province. Yangquan (19.2 billion tons reserves) is the nation's biggest producer of anthracite, and Xishan (10.7 billion tons reserves) is the fuel base of Shanxi's capital, Taiyuan city.

In 1983 Shanxi counted 25,723,053 inhabitants, generating an average population density of 165 inhabitants per square kilometer. The maps in Figure 5.7, however, show that there are considerable differences in population density, due to the large concentrations of people in large cities such as Taiyuan city (2,344,452), Datong city (1,000,062), Changzhi city (2,713,984), Yangquan city (1,048,443), and Jincheng city (1,820,330). The distribution of population density is the first indicator of Shanxi's main axis of economic development, running from north-east to south-west, with a second center of development in the south-east.

This axis of economic activity can be inferred from various kinds of data. Industrial production is shown in Figure 5.8, energy consumption in Figure 5.9. Another indicator is the relation between the total water supply and total water exploitation as shown in Figure 5.5).

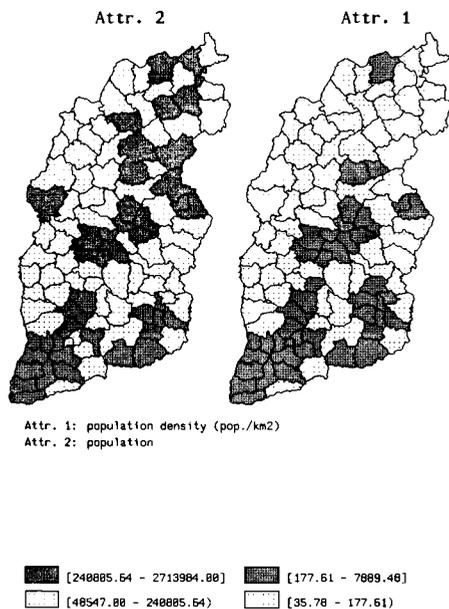


Figure 5.7: Shanxi Province: population and population density

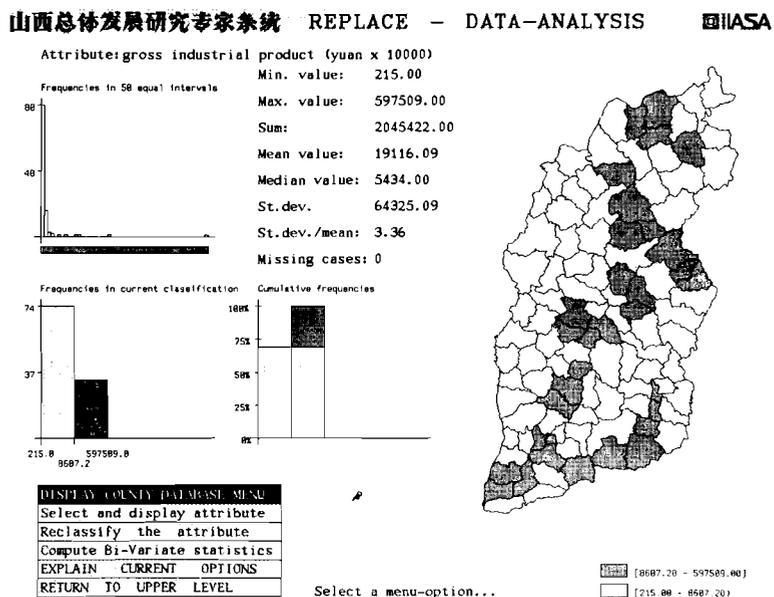


Figure 5.8: Shanxi Province: industrial production

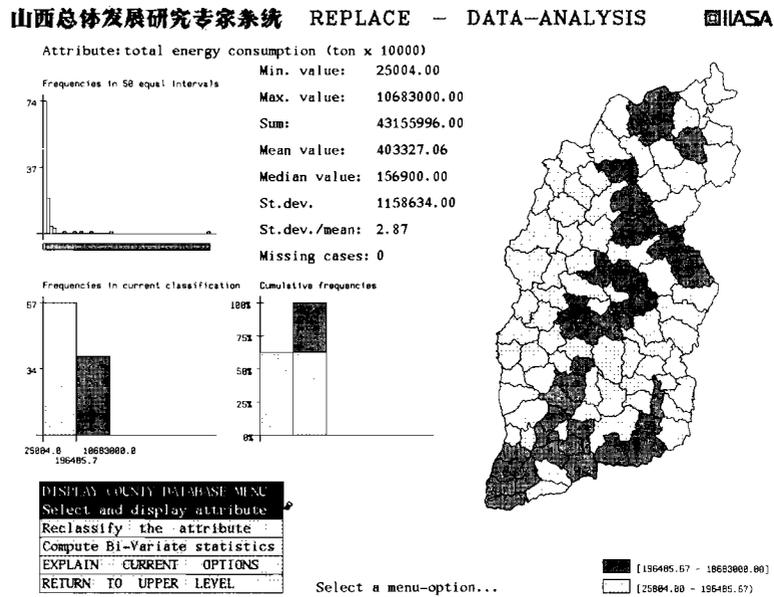


Figure 5.9: Shanxi Province: energy consumption

Agricultural activity, on the other hand, shows a pattern that is different from that of industrial activity. Figure 5.10, for instance, shows that although the wheat production areas are part of the central axis of development, corn production is mainly in the east of the province. Figure 5.11 shows the distributions for cattle and pig raising. Again, large parts of the eastern counties are characterized by large numbers.

A picture emerges of an industrialized, developing axis down the middle of the province, along the Fenhe river, where the cities and many of the coal reserves are situated, supplemented with a somewhat local center of development in the south-east, concentrated around a few large cities. The eastern part of the province has a typically agricultural role. The western parts seem to lack any significant economic activity, although a few of the counties might act as local growth poles because of the presence of mineral resources.

5.6.2 Shanxi Province: the problem

After the founding of the People's Republic of China in 1949, the province went through three stages of development:

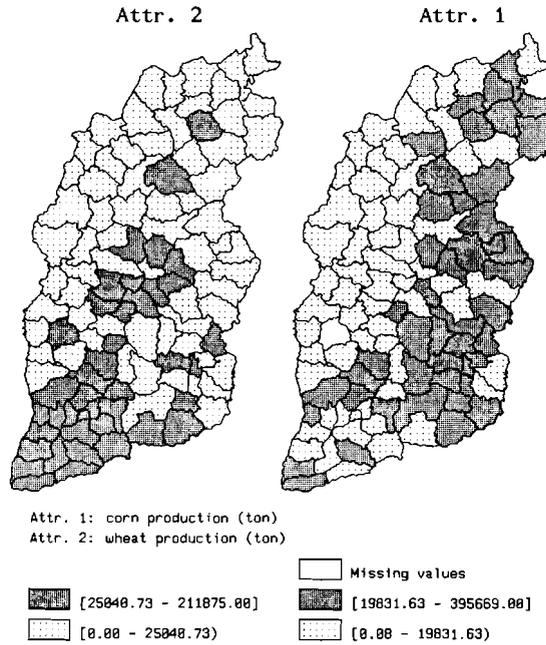


Figure 5.10: Shanxi Province: corn and wheat production

山西总体发展研究专家系统 REPLACE - DATA-ANALYSIS IIASA

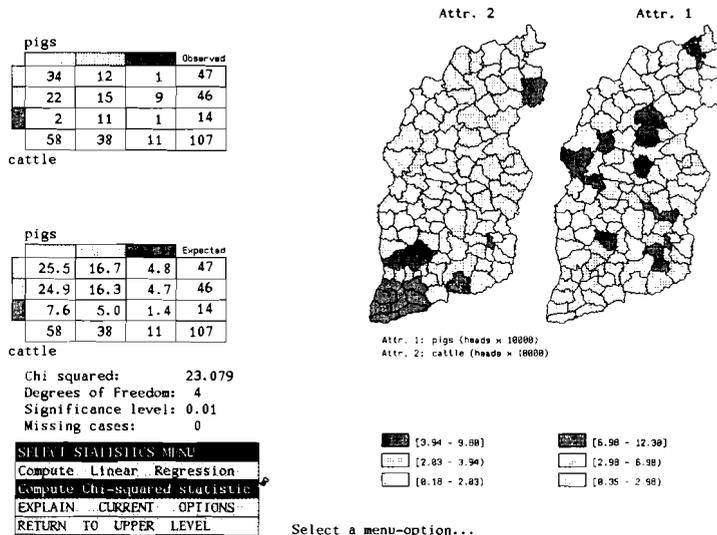


Figure 5.11: Shanxi Province: livestock

1. *Restoration and economic adjustment (1949–1965)*: During this period, “the establishment of the industrial technical basis had heightened the productive capacity to a new level, ascertained the leading position of industry in the whole economy and laid a solid foundation for setting up a comparatively coordinated developing system of national economy and a socialist industrial system.” (Gao, 1985; p. 15). Coal production was increased sharply. The same happened in the metallurgical, engineering, chemical and textile industries. City and town infrastructures were greatly improved.
2. *Third and fourth five-year plan period (1966–1975)*: The cultural revolution; period of vigorous political struggle and social turbulence. Investment and economic development slowed down considerably.
3. *Fifth and sixth five-year plan period (1976–1986)*: This period is characterized by the “strategic decision of shifting the stress of work to the socialist construction of modernization” (Gao, 1985; p. 19). Shanxi was planned as the center of a large geographical base for energy generation and the heavy and chemical industries. 16 Projects were set up in Shanxi, ranging from large coal-based industries to a 500,000 volt high-tension transmitting and transformation project.

By 1983 Shanxi contained 9,809 industrial enterprises, 3.42 times as many as in 1949. Their fixed assets amounted to 85 times that of 1949, and the industrial output had been multiplied by a factor of 68.8. An example: in 1983, an average 3.7 days of steel smelting equaled the largest annual production of steel before 1949. The figures for producing raw coal, electric power, and caustic soda were 14.2, 1.25, and 2.58 respectively. The large and medium-sized industrial projects were spread over 42 different counties.

Notwithstanding these achievements, the economic development of Shanxi is characterized by some important problems (Fedra *et al.*, 1987; p. 10):

1. The province’s economy is characterized by a low degree of efficiency. The industrial output values and revenues are about 36% beneath the national average, profits and taxes are 30% below, and labor productivity is 29% below. As a result, income and living standards of the population are beneath the national average. Development alternatives that better suit Shanxi’s character and situation, leading to steadier growth and increased welfare should be pursued.

2. Earlier, focus on the growth of industrial output, especially heavy industry, led to serious imbalances with regard to agriculture and light industry. Industry tends to be labor intensive with low profits and little flexibility. The introduction of more flexibility and a better balanced industrial production is of great importance.
3. Environmental problems are becoming serious. This currently pertains to pollution of surface water, groundwater, and air, but also to a looming water shortage. Future developments must therefore take ecological and water supply factors into account.

A number of development objectives have therefore been formulated:

1. By the year 2000, annual industrial and agricultural output should be quadrupled from the 1980 basis, implying an annual 7.5% growth rate over the next 10–15 years;
2. Substantially raise the living standards of the population, taking into account the needs of both production and consumption;
3. The province aims to achieve 270 million tons of annual coal output by the end of the century;
4. Simultaneous development of the economy, society, science, and technology, with attention paid to an ecological balanced development.

Solving the problems and achieving the objectives is subject to various constraints:

1. Capital constraints: the 1984 level of investment was about 40 billion yuan, with an estimated annual growth rate of 7.5%;
2. Water resources: currently, with about 50% of all water resources developed and a planned fourfold increase in production, problems of quantity, location, and distribution have begun to become serious;
3. The current transportation network has insufficient capacity for processing the total freight volume;
4. Serious environmental problems are anticipated;
5. There is a considerable shortage of skilled labor;

6. Technology and management are Shanxi's weak points.

These and other constraints must be taken into account in further planning the economic development of Shanxi Province. The leaders of Shanxi Province have a lot of decisions to make as to how to rearrange overall planning, implementation of policies, coordination and the use of economic means of regulation in such a way that the objectives can be achieved. Earlier, it was mentioned that this kind of problem is well suited for implementation in a DSS. A significant part of the relations and mechanisms of a spatial economy can be captured in models. Aims and constraints are subject to decision making and determine the fixed parameters of the models. The problem is a clear example of one for which there is no single, or only a few, ideal solutions. Many courses of action can be thought of, each of them having different consequences for different parts of the economy or regional society as a whole. Moreover, various interests and objectives may require different measures since they are characterized by different requirements. Some of these objectives may furthermore be mutually contradictory or may compete with each other for the same resources. Evaluation of the sets of functionally equivalent means can then be done on common dimensions such as 'which alternatives are the least competitive', 'which are the cheapest', or 'which are most sensitive to future changes in object characteristics'.

5.6.3 Overview of the DSS

Fedra *et al.* (1987) suggest a DSS that supports decision-making and strategic planning on a scientific basis. The DSS should provide direct assistance to top-level decision makers. "It must accentuate problem-mindedness over solution-mindedness in that it perceives as critical the need to explore the nature of the problem and to generate alternatives rather than purely to dwell on the choice among alternatives, as this last precludes, or at least inhibits, the exploration of novel avenues." (Fedra *et al.*, 1987; p. 13). This 'objective' goes well with the notion of a decision maker as his own model, assisted by a system that can help explore smaller, complex problems. The idea behind such a system is thus that it is the decision-maker who, by virtue of his knowledge and experience, is able to reduce the huge search space to a much smaller space of possible lines of action. This smaller space, however, is still large enough to contain many possible combinations of actions and various versions of actions. Furthermore, each of these lines of action might still have to be investigated on its actual feasibility. This makes the

objective-means relationships go one level deeper. The decision maker's means become the objective of a deeper model. What is attractive though, is that the provision of models and means to simulate changes in the system by a DSS, can also stimulate the decision maker to develop new means and strategies.

The approach suggested by Fedra *et al.* should lead to some scenarios of the overall structure and relationships characterizing certain types of policy measures. Definite values must be subject to the user's preference, since they provide the decision makers with the opportunity to introduce their own insights and knowledge. The result of this was a prototype DSS, the architecture of which was displayed in Figure 5.3. The overall system was designed as a 'hybrid' system, meaning that the system contains a number of different approaches to implementation. Due to the diverse nature of the information required, multiple knowledge representation paradigms are integrated. Some aspects, like the relational matching model and a symbolic macroeconomic simulator require knowledge-based, rule-oriented information processing and modeling, whereas a Gaussian dispersion model or a specific water management model use traditional numerical data-processing. The start-up screen of the Shanxi Province Prototype DSS (Figure 5.12) shows a number of aspects of the implementation and approach to DSSs developed at IIASA.

1. *Models and data bases:*

The system contains thirteen different models (for a list of contributors, refer to IIASA (1988; p. iii, iv), three user interactive data bases, and a prototype interactive problem definition module (Winkelbauer and Reitsma, 1988). Each of the models covers a different perspective on the province, and can be used to evaluate policy measures, changes in Shanxi's characteristics, or can be consulted for optimum solutions, given an initial situation. Currently, only a few models are truly coupled. In future releases, however, much more attention may be paid to integrating the models in the way discussed earlier. Models and data bases are separated, although they can be linked conceptually. The data bases (GEO, COMP, and a GIS-component of REPLACE), provide information on a large range of attributes. The information is both spatial and statistical (Figures 5.5–5.11). The REPLACE data base/GIS is completely linked to the relational matching model for site suitability (refer to Section 5.7.9 for a more detailed discussion).

2. *Graphic presentation:*

At the beginning of this chapter the problem of models which are difficult to

山西总体发展研究专家系统 ShanXi Regional Development DSS IIASA

Prototype Release Version 2.0 July 1988
All Rights Reserved.

This software system is developed under contract to
State Science and Technology Commission (SSTC)
Beijing, The People's Republic of China, by
the International Institute for Applied Systems Analysis
IIASA, A-2361 Laxenburg, Austria.

ShanXi Development DSS - MAIN MENU	
Interactive Problem Definition	KIM
Long-Term Macro-Economic Model	MACSIM
Economic Sector Interaction	MACEDIT
Input/Output Modeling and Analysis	I/O
Regional Investment Analysis	GLOBINVEST
Macro-Economic Regional Comparison	COMP
Site Suitability Analysis	REPLACE
Transportation Systems Analysis	TRANS
Industrial Structure Optimization	PDAS
Global Energy Demand Analysis	MAED-BI
Global Coal Mining Sector Analysis	COAL
Water Resources Planning Model	MTYSIM
Environmental Impact Assessment	ISC
Geographical & Regional Data Bases	GEO
Urban/Rural Development Game	CONFRES
Multi-Criteria Data Evaluation	DISCRETE
EXPLAIN ONE OF THE CURRENT MENU OPTIONS	
STOP AND EXIT DECISION SUPPORT SYSTEM	

select a menu item with the mouse (left button) ---->

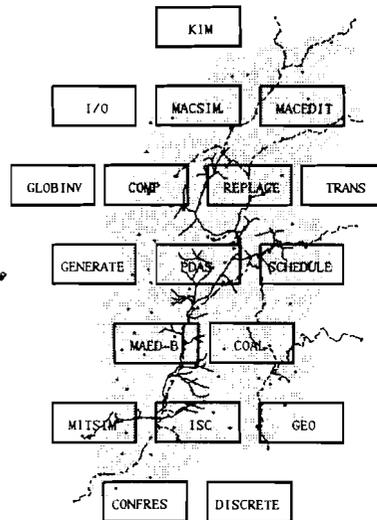


Figure 5.12: Start-up screen of the IIASA-ACA Shanxi Province DSS

access, use and understand, and the need for user-friendly interfaces, efficient and clear presentation of input and output of model data and information in general, was addressed. In the system presented here this is realized by a fully graphic implementation. Data are presented by icons, maps or a combination of these. Model results are directly projected onto maps or, if the output is non-spatial, in graphs, icons, or combinations of either of them with numbers. The system is completely mouse-driven, so that hardly any keyboard entry is required. Setting of parameters and constraints is done graphically and by means of the mouse. A nice example of replacing complicated input, both conceptually and technically, by a user-friendly, understandable procedure, is shown in Figure 5.13 and color Plate 1.

This screen is of the specific task to be conducted in the preparation of the data for the execution of an atmospheric pollution model (for an overview of this model refer to Posch (1988); for an application refer to Fedra *et al.* (1987)). The model needs considerable quantities of meteorological data including information about temperature, wind-speed and wind direction, atmospheric stability, inversion layers, etc. However, if the model is to be used by planners who must take decisions as to where to locate certain production plants that pollute the air, these data need to be replaced by

山西总体发展研究专家系统 Industrial and Urban Air Pollution IIASA

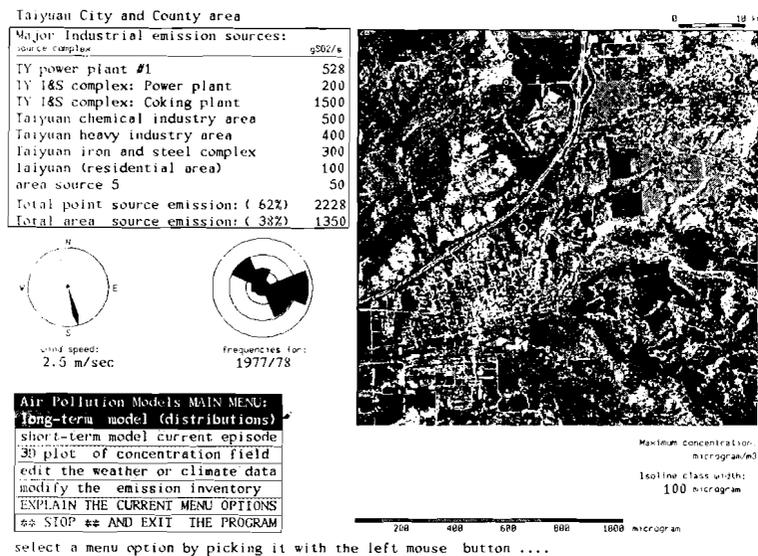


Figure 5.13: Setting up the data for the atmospheric pollution model

less specialized and easier-to-understand representations. Therefore, various combinations of these factors representing frequently occurring weather patterns, are presented as icons, which can be selected by the user of the model by means of the mouse. They form an alternative way of preparing the input for the model. The output of the model is shown in color Plate 1. Instead of complicated tables and graphs, the results are directly projected onto the map of the area, and various explanation boxes inform the user of what is on the screen and the available features.

3. Menu- and default-driven:

The system is menu- and default-driven. This means that the sequence of actions a user can ask the system to perform, is limited to the extent that certain sets of actions have to be conducted in a specific order, and that the feasibility of an action depends on the context of its use. Nested use of menus results in a tree-like procedure. From each of the menus the user can exit to the above level, or go deeper, by choosing another option. Every menu is accompanied by an explanation facility that tells the user what the options of the menu stand for. For all the models default values are set. This means that upon entry, there is always a valid input set-up on which a model can run. The user can of course alter the default settings by telling the system

(either by means of a menu or by pointing with the mouse device to the appropriate locations on the screen). This will then either provide the user with instructions of how the attributes and parameter(s) can be changed, or with a graphic editor for modifying attribute values. Most editors are 'smart' in the sense that they 'know' the limits values can take, either because of the values other variables have, or because of pre-defined boundaries set by the developer.

5.6.4 REPLACE: Relational Plant Location and ACquisition Enquiry

As is apparent from Figure 5.12, the relational site suitability matching system described in Chapters 3 and 4 constitutes a separate module in the overall Shanxi Province DSS under the name 'REPLACE' (for some data on applied software and hardware, refer to Appendix 4). The significance and possible usefulness of a site suitability evaluation model for the overall DSS does not need to be discussed here in great detail. From what has been discussed in Chapter 2 and from the examples presented earlier in this chapter, it will be clear that for the planning of regional development, production milieu and site suitability measurements can be important aspects. Planned production must be located somewhere. Questions as to where this is feasible need to be answered. Conducting several site suitability measurements for different types of activities may point out areas that might be prone to competing for allocation.

5.7 REPLACE: functions, implementation and 'looks'

The core model of REPLACE and its implementation were described in Chapter 4. But as with the other models in the Shanxi Province DSS, the implementation had to be enhanced by a whole series of extensions and interfaces before it could be considered compatible with the overall features of the system. For example, no selection/editing opportunities for either the locations or the activities had been incorporated. The same holds for the way the output of the model is displayed, the implementation of the explanation routines, the opportunities for introducing 'don't care' values into the model, the way the user can specify dimension-specific matchings, and so forth. Figure 5.14 shows the tasks REPLACE can perform. As with

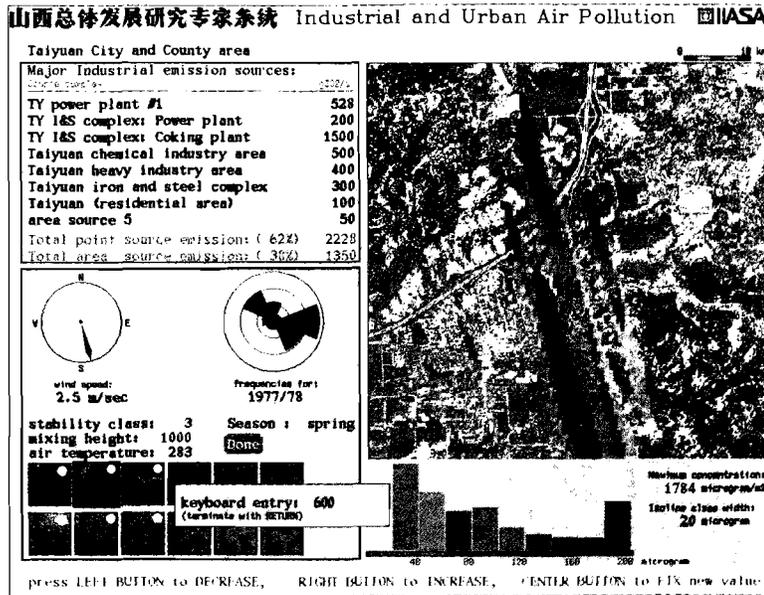


Plate 1: Output of the atmospheric pollution model

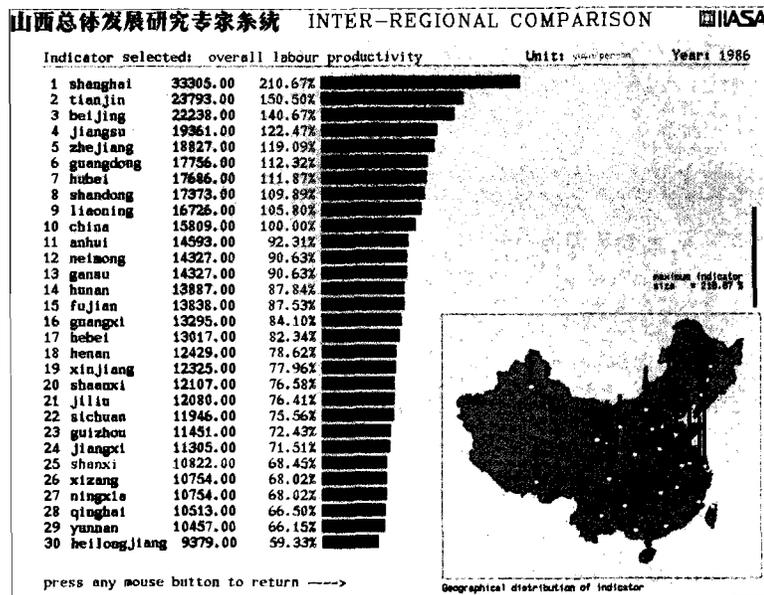


Plate 2: Inter-regional comparison of overall labor productivity

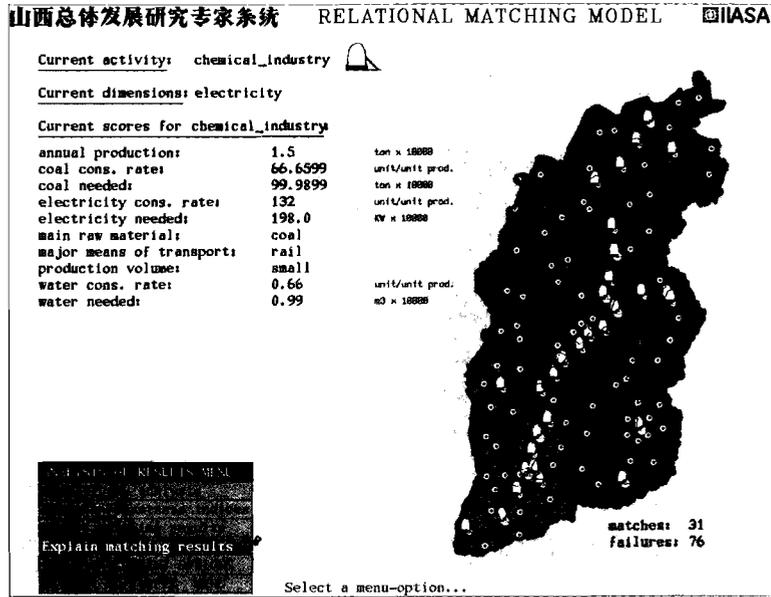


Plate 3: Showing the matching results

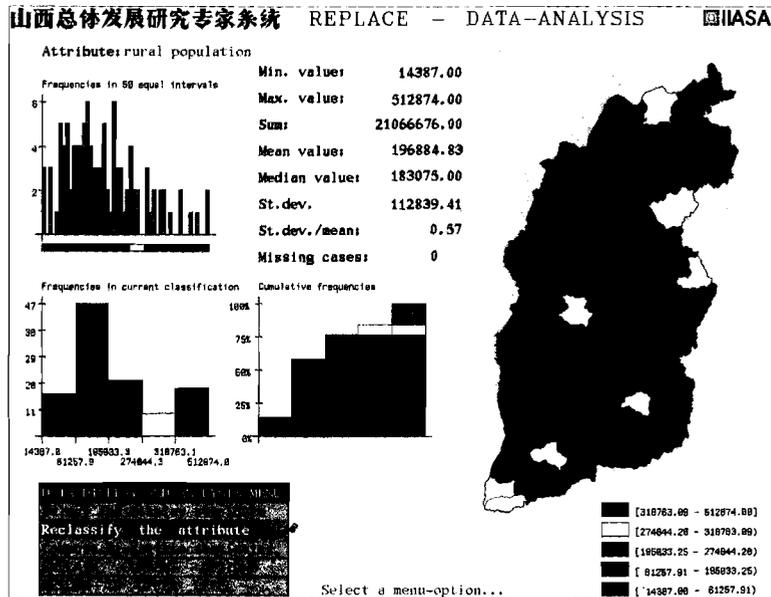


Plate 4: Display of a reclassified county attribute

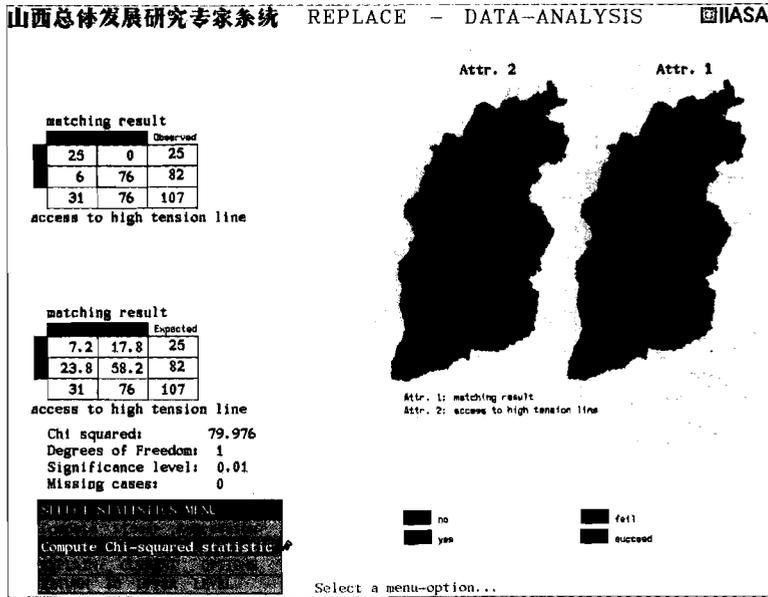


Plate 5: Matching result for electricity and access to high-tension lines

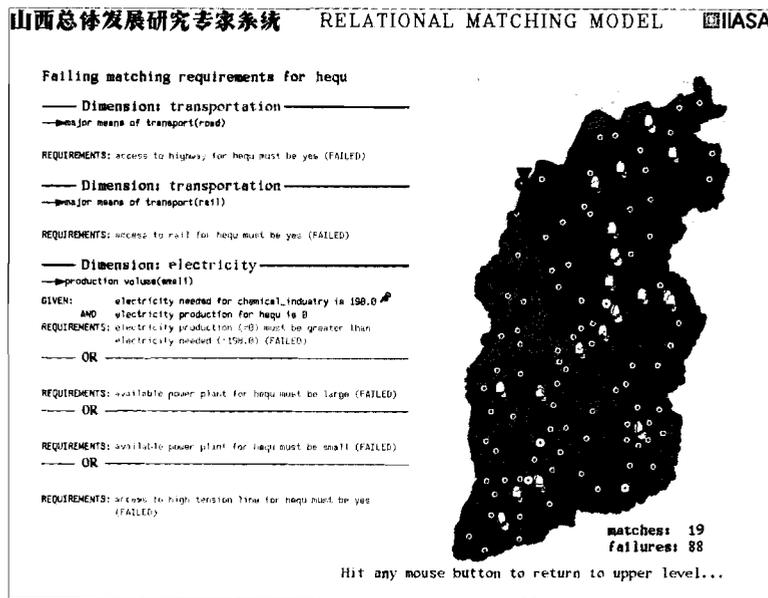


Plate 6: Explanation of the failure of a county to match

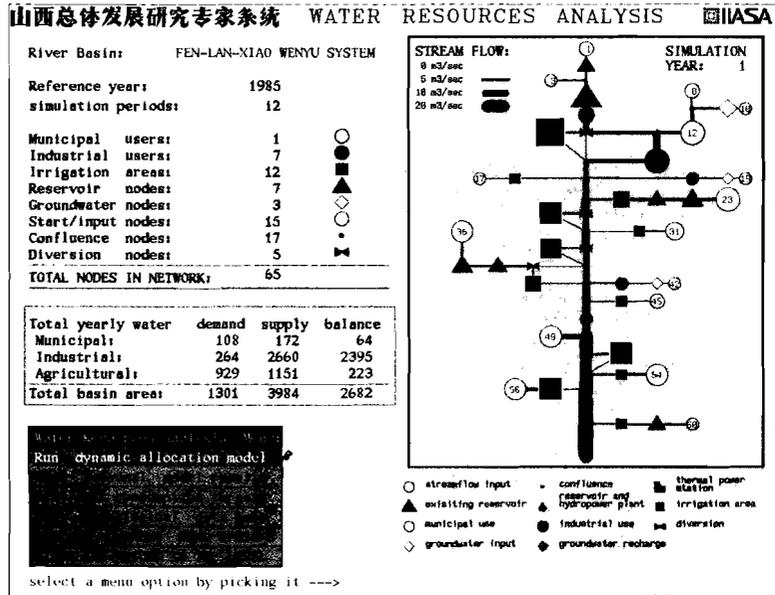


Plate 7: The MITSIM water management model

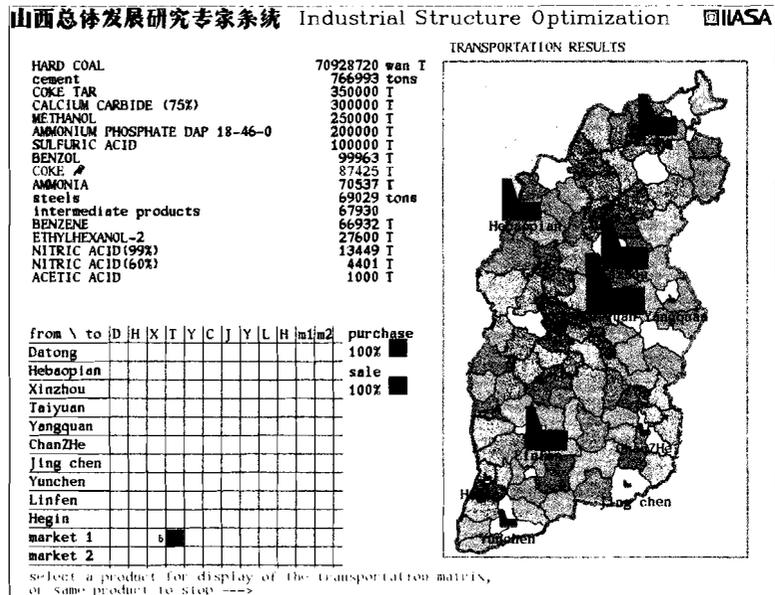


Plate 8: The PDAS industrial structure optimization model

the other modules in the DSS, REPLACE is completely menu-driven, the functions can be displayed as a tree of menus and menu options.

The remainder of this chapter consists of a sequence of paragraphs, each of them allotted to a specific function of the REPLACE system and its implementation. The discussion of these functions and options takes the form of a number of examples and illustrations from case studies, some of which are dealt with in more detail in Chapter 6. The illustrations are color screen dumps, transformed into black-and-white raster dumps. To obtain an idea of what things look like on a 19 inch, 1152 x 900 pixel color monitor, refer to the color plates in this report.

5.7.1 Main menu

The main menu constitutes the initial state of the REPLACE system from where major functions and actions can be executed. The associated screen (Figure 5.15) shows the main menu itself, as well as information on the current settings of a number of components of the matching process.

A map of the counties of Shanxi Province with their capitals appears on the screen. It was decided to use the 107 counties of the province as the total set of locations from which the activity-specific production milieu could be constructed. Arguments for taking the county as the object of measurement are given in Chapter 6. The triangles pointing to some of the capitals represent a limited set of counties that is currently considered to be the total set for which a matching must be conducted. In many instances a user might not be interested in a total matching for all 107 counties, but instead may only want to inspect the matching results for a limited set of counties or for just one county. The 'Select set of counties' option enables the user to modify this set of locations. The number of currently selected counties is displayed on the screen. Also displayed are the 'current activity', the 'current dimensions', three 'decision parameters' that inform the matching program of what is to be done in case information needed for the matching is missing and how a numerical mismatch must be interpreted, as well as a bit of general information on the REPLACE system. Activity, dimensionality, set of counties, and the value of the decision parameters, is information that is needed before a matching can be conducted. Any matching needs an initial setting for these four types of data. As is the case with the other models in the Shanxi Province DSS, REPLACE is default driven. This means that there is always a current activity, dimensionality, etc., so that the matching model can be run at any time, even upon entry

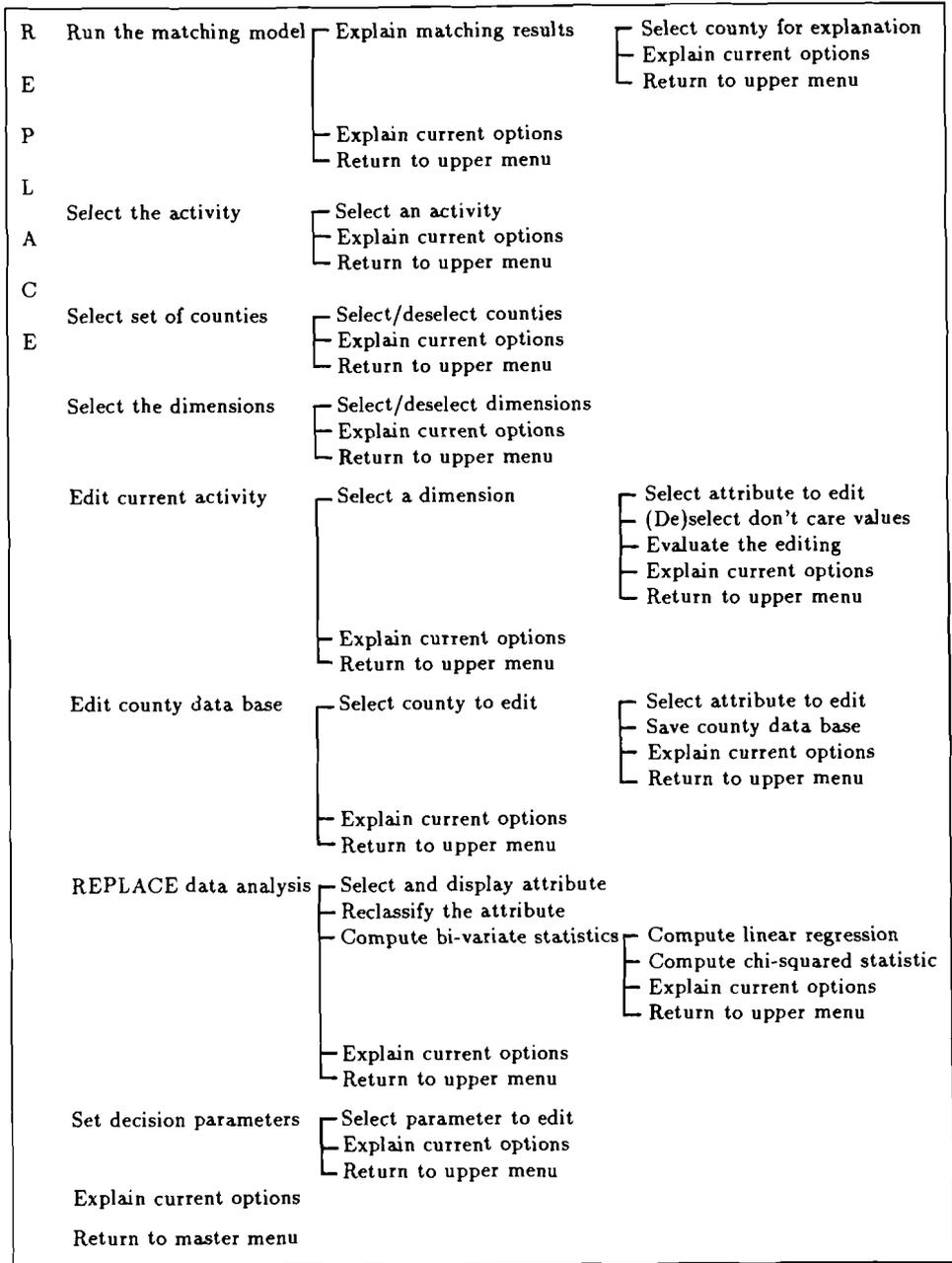


Figure 5.14: Menu structure of the REPLACE system

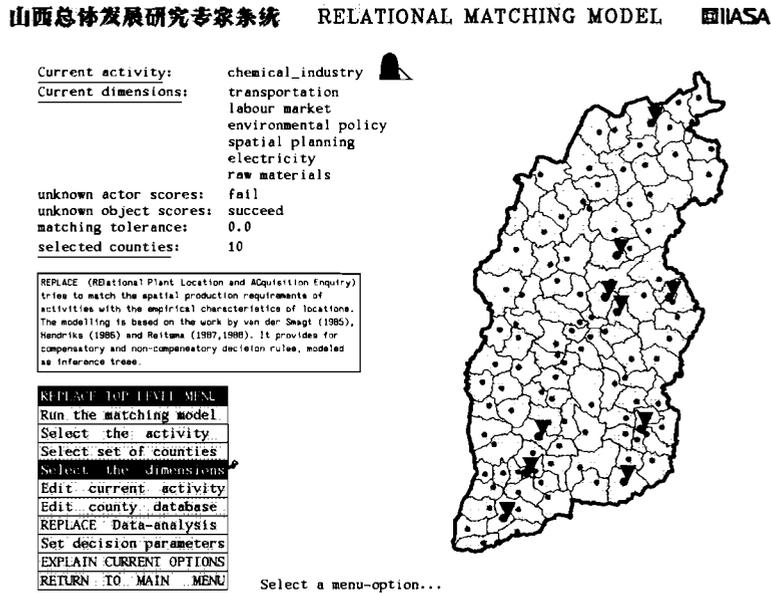


Figure 5.15: Start-up of the REPLACE system

into the system. Current settings can be changed by selecting the ‘Select the activity’, ‘Select the dimensions’, and the ‘Set decision parameters’ options, respectively. Invoking these and other actions from the main menu will result in a complete or partial change of the screen. The screen belonging to the new task is called up, together with a new menu, or some instructions on options are available to the user. In case the program is doing something on its own, such as running the matching model, the user is informed. After the task is completed, the program returns to the main menu and its associated screen. If the set of counties, activity, dimensionality, or decision parameters have been altered, the corresponding information on the main menu screen is changed. Selection of the ‘RETURN TO MAIN MENU’ option terminates the REPLACE session and the system returns to the Shanxi Province DSS master menu (Figure 5.12).

5.7.2 ‘Run the matching model’

Selection of this option invokes the site suitability matching process. The matching program collects the necessary information: current activity, initial set of counties, and dimensionality, then loads the activity-specific rule base

(inference tree(s)) into the Prolog data base, and conducts the matching as explained in Chapters 3 and 4.

The associated screen (color Plate 3) shows information on the activity that is currently matched for. As mentioned earlier, many of the activity's characteristics are derived from other, more basic characteristics. In color Plate 3, for instance, the coal requirement is a clear function of the annual production (basic data) and the coal consumption rate for that specific industry. Matching the counties against the activity is a simple sequential process. In Section 4.4.3 it was explained that it was chosen to conduct the matching on a county-specific basis, thus conducting a full matching for one county before continuing with the next one. One of the advantages of this approach is that the matching results can be mapped onto the map of Shanxi Province during the matching process. Counters in the lower right of the screen show the number of matches, failures, and counties yet to be matched, while each time a county has been matched and the match turns out successful, a little mapping routine places the activity symbol in the county on the map. If the initial set contained enough locations, and if matches are found, patterns such as the ones in color Plate 3 or Figure 5.16 may emerge. Finally, after the matching process is finished, a menu is put up that enables the user either to go into the explanation phase, or to return to the main menu.

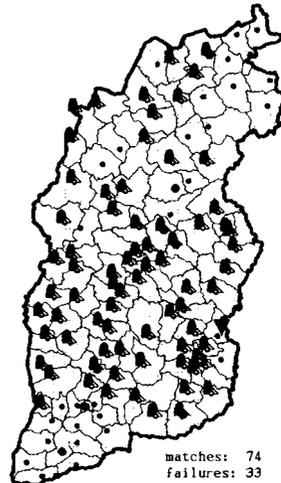
5.7.2.1 'Explain matching results'

If the user wants to have some explanation as to why specific counties did not match (failed), he can ask the system for an explanation. Selection of the 'Explain matching results', will first make the system request the user to select a county for explanation (Figure 5.16). This is done by leaving the map with matching results on the screen (note that the maps in color Plate 3 and Figure 5.16 are taken from different matchings), and putting the names of the counties that were contained in the matching (here all 107 of them) to the left of it. Names of counties that matched are displayed in blue, names of those which failed the matching requirements are in red. The user is instructed to select a county that failed the matching (red) for explanation by means of the mouse. Selection is shown by displaying the name of the county in yellow, and by putting up a green triangle on the corresponding location on the map. Evidently, a triangle can only point to a county that does not contain an activity symbol representing a successful matching. Selection routines such as these are made flexible by combining continued selection

山西总体发展研究专家系统 RELATIONAL MATCHING MODEL IASIA

Counties contained in the matching...

- | | | | |
|---------------|---------------|-------------|---------------|
| ance | jiangfan | pinlu | xiaoyi |
| baode | jiachang | puolan | xiaolan |
| changzhi | jiakou | qingou | xingfan |
| changzhi city | jiakou | qinshui | xinjiang |
| daxian | jinzhong city | qinshan | xinzhong city |
| daming | jingde | qinyuan | xixian |
| dating | jishan | qixian | xiyang |
| dalong city | jishan | qiao | yangcheng |
| dingxiang | kalan | ruicheng | yanggao |
| fangshan | lanxian | shanyin | yanggu |
| fansi | lancheng | shanchi | yangquancity |
| fansi | linchuan | shilou | yicheng |
| fanyang | linfeng city | shouyang | yingtan |
| fushan | lingzi | suokan | yonghe |
| gaoping | lingxi | taigu | yongji |
| guanglin | linxi | taiyuancity | youyu |
| gujiao | linyuan | tianzhen | yuanning |
| guwan | liang | luanlu | yuanyuan |
| hejin | liulin | wanrong | yuci city |
| hequ | loufan | wanshui | yuncheng city |
| heshun | lucheng | wenxi | yushu |
| hongtong | ningguo | wutai | yuxian |
| houmacity | pingguan | wuliang | zhangzi |
| huxian | pingding | wuzhai | zhongyang |
| huguan | pinglu | wangfen | zuoquan |
| hunyan | pingshun | wangling | zuyun |
| huokan | pingyao | wangyuan | |



Select a county which failed the matching requirements (red) for explanation by picking it at its name or by picking it from the map.

```

SELECT FOR EXPLANATION MENU
Select county for explanation
EXPLAIN CURRENT OPTIONS
RETURN TO UPPER LEVEL
    
```

Figure 5.16: Selecting a county for explanation

with termination by repeated menu option choice. For the user this means that once he has selected the 'Select county for explanation' option from the menu in Figure 5.16, he can keep selecting counties—each new selection will undo the previous selection—until the 'Select county for explanation' option (marked with a blue arrow to show that it is the option that is currently active), is selected again. This will then fix the selection, which is then handed over to the control program that passes it to the actual explanation routine. This continued selection with menu termination is the convention that was chosen for the whole of the Shanxi Province DSS. Upon return from the explanation, the user can select another county for explanation by choosing the select option again. Selection of the 'RETURN TO UPPER LEVEL' option makes the program return to the main menu.

Once the county for which explanation is requested is fixed, the list of names disappears from the screen and is replaced by the explanation results (see color Plate 6). In order to be able to fully understand output such as shown in color Plate 6 one has to consider the way the explanation is built up and displayed on the screen. The explanation routine consists of two parts: the first collects all the information needed in the explanation, the second orders it and puts it on the screen. The collecting of the information is

done by inspecting matching documentation from the Prolog data base that was left there by the matching program. In Section 4.4.4 it was explained that it was chosen to leave information in the data base while matching, rather than carrying it along during the matching. Also left in the data base is the route through the various dimension-specific inference trees. This information is used by the explanation routine and is removed upon return to the main menu. The matching information contains only items from the inference trees, and the locations in those trees where failures occur. No specific values on requirements and county characteristics are documented. It is for this reason that the gathering of the explanation information does not only involve the collecting of the information left in the data base, but also a second 'going-through-the-inference-trees' in order to find the relevant attribute values. Once all this is done, the information is handed over to the analyzing and the parsing programs. The analyzing routine (re)organizes the information so that it takes the format of decision rules. This implies the separation of a route through the tree (the entries behind the arrow symbols in color Plate 6, Figure 6.12 and Figure 6.13) from the matching rule, and within the matching rule, the separation of the 'given' part (the `v_get` clauses in Figure 4.7) from the actual requirements. The parsing routine then parses this information into a more natural language oriented format, after which everything is written to the screen. In case the explanation requires more than one page, pages are displayed sequentially. Note that the explanation in color Plate 6 contains multiple entries for the 'raw materials' dimension. This is caused by setting of 'don't care' values.

5.7.3 'Select the activity'

In Chapter 3 it was explained that it was preferred to work with separate, generic types of activities, rather than with one large classification tree of which the root would be something like 'spatial activity'. As an empirical case study for both the development of REPLACE and the content of the Shanxi Province DSS, several generic types of industries were selected. For these types (simple) rule bases were developed. But if various generic types of activities are available, the user must be able to specify the type the site suitability measurements must be carried out for. For reasons of clarity, full specification of an activity may require two steps: selection of a generic type, and the setting of specific attribute values for that type (see paragraph 5.7.6). Figure 5.17 shows the activity type selection screen. Only a few types

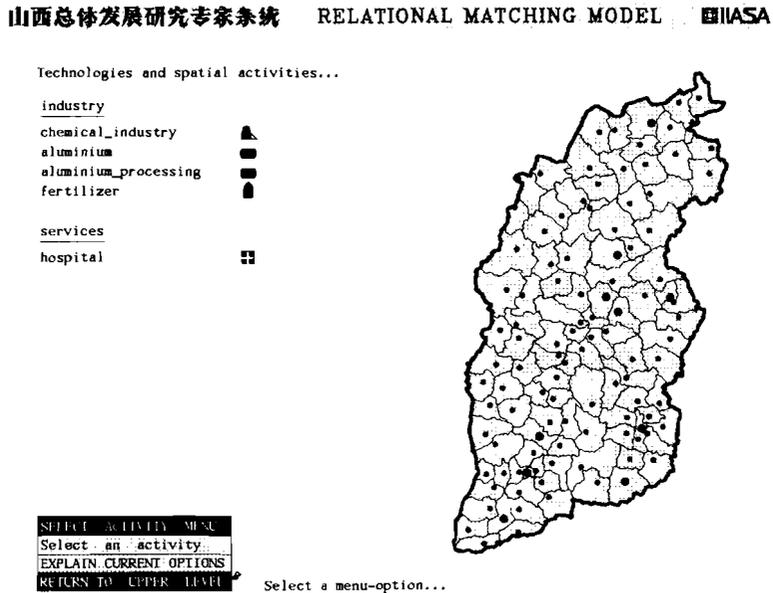


Figure 5.17: Selecting an activity for matching

are available, each of them having their own pictorial symbol (the symbols for aluminum and aluminum processing have different colors).

A problem that has to be dealt with in loading rule bases (inference trees) for different activity types is that the various rule bases must not get mixed up with each other. The best way of avoiding this, is to allow only one activity-specific rule base in memory at any one time. To achieve this in (Quintus) Prolog, different rule bases were stored in different files. A default activity rule base is loaded into the Prolog data base when REPLACE is started up. Special loading routines were therefore written that would monitor the loading and administer the predicates and their arguments when loaded from the file. The combination of file name and list of predicates and arguments enables removal of the entire rule base from the data base again if the user selects another activity type. After the rule base for the old activity type has been removed, the one for the new type is loaded using the same procedure. To avoid a lot of useless loading and unloading of rule bases, checks were incorporated for instances when the user selects a new activity which is the same as the current one.

5.7.4 'Select set of counties'

It is very well possible that the user is not interested in matching results for the entire province. Perhaps he is only interested in the results for a small set of counties (e.g., the large cities) or in just one particular county. For this reason it was decided to keep the initial set of locations as a variable so that the user can specify his own set of counties for which a matching must be carried out. Figure 5.18 shows the screen that is associated with the routine that takes care of this.

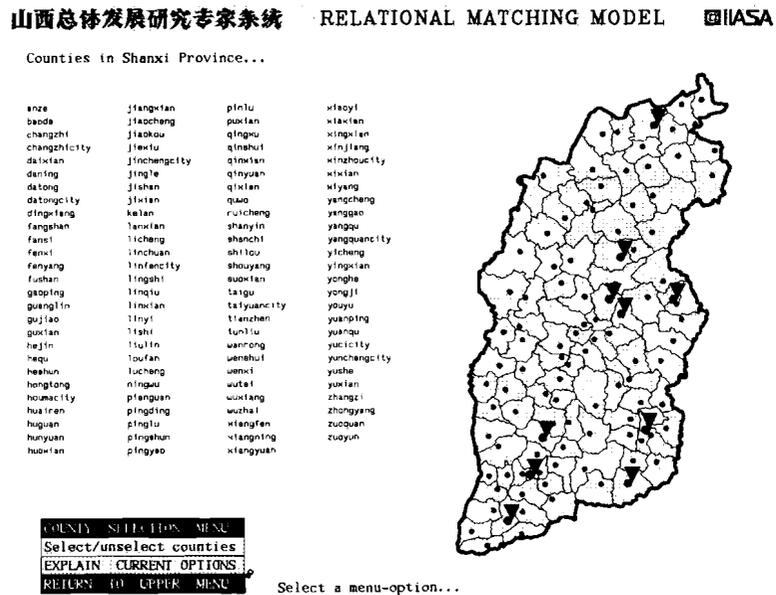


Figure 5.18: Selecting countries for matching

The counties on the map correspond with the alphabetically ordered list of names on the left. Counties can be selected by clicking a mouse button while pointing on the screen at either their names, or at the capitals on the map. The reason for allowing selection from the map is that it is conceivable that the user is interested in counties that have a special topographical character, such as border counties, counties from a specific part of the province, or counties representing large cities. Selected counties are displayed in red in the names list, and with a green triangle on the map. If a selected county is selected again, the selection is undone. The selection triangle is taken away and the name turns to blue again. Toggling between

select and de-select enables the user to correct selections made by mistake. The full set of 107 counties is the default.

5.7.5 ‘Select the dimensions’

In Chapter 3 a number of advantages of working with a dimensional structure of the matching problem were mentioned. Among them were the possibilities of dimension-specific matching, and the integration of object demands by putting them into separate dimensions. Prerequisite for this kind of flexibility is that the user must be able to make his own choice for the dimensionality of the matching problem, of course within the limits of the total dimensionality as defined in the model. Furthermore, in Section 3.6, it was suggested that demands generated by objects be represented in separate dimensions. A routine that enables the user to incorporate these dimensions into the matching problem or exclude them from it is therefore a requisite. Selection of the ‘Select the dimensions’ option from the main menu triggers a routine that enables the user to do this. Figure 3.3 shows a possible screen that appears. The left side of the screen shows the current activity, the dimensional tree and a menu. The content of the dimensional tree is dependent on the current activity. A different activity may be associated with a different dimensional tree. The dimensional tree is also hierarchical. If a dimension is selected then all its sub-dimensions are also selected. Selected dimensions are displayed in red, unselected ones in grey. As with selecting counties for matching, selection is done by clicking a mouse button while pointing at the dimension name. Selection and de-selection toggle. Full dimensionality is the default.

5.7.6 ‘Edit current activity’

Determining the characteristics of the activity for which a matching is conducted requires two steps; selection of a generic type (Section 5.7.3) and, if required, editing of this type of specific attributes. Activity attributes can be divided in three ways that are important for editing them. First, they can be split up in independent, basic attributes, and dependent, derived ones. Another distinction is the one that separates inferential variables from non-inferential ones. The inferential variables are the ones that serve as nodes in the inference trees. These inferential variables, in their turn, can be separated into attributes that can take ‘don’t care’ values, and attributes that cannot. Yet another distinction is the separation into categorical and

numerical attributes. The editor was constructed to handle both the modification of independent variables and the setting/cancellation of 'don't care' values. These dichotomies are independent, however, so we can treat them separately.

If one disregards the 'don't care' values for a moment, it will be clear that the editing of activity attributes only pertains to the independent ones. Dependent variables are functionally derived from either other dependent ones, independent ones, or a combination. Their dependent character renders their editing 'impossible', even if they are inferential attributes. Figure 5.19, for example contains the inferential attribute 'production volume'. It can take the values 'small' and 'large', but the proper category is determined by a hidden function on the basis of the 'annual production'. Suppose that the small-large boundary is 100 units of production. Suppose furthermore that the user specifies that the production volume should be large. Which annual production should go with that? Clearly, a production larger than 100 units, but how much larger? That this is not an arbitrary matter becomes clear if it is known that other attributes such as the amounts of coal and water that are needed are also partly computed on the basis of the annual production. It was therefore decided to restrict the editing of activity attributes to the independent attributes only. For these, a special-purpose editor was implemented, a screen of which is shown in Figure 5.19.

The screen contains the current activity, a dimension, a number of attributes, their current values and their possible values, and a menu. Current values are displayed in red, possible values in blue. Attributes are separated in categorical and numeric ones. Editing of the variables happens by clicking a mouse button while pointing at a categorical score for categorical attributes, or by dragging the value pointer (little triangles) along the associated numerical scale. Alternatively, the new values can be entered by means of the keyboard.

There is a relation between the dimension displayed on the screen, and the variables listed. Editing of the activity attributes occurs on a dimensional basis. Preceding the actual editing, therefore, the user is prompted to select a dimension for editing. This is done by means of a routine that is similar to the one described in Section 5.7.5. The only difference is that only the leaves of the dimensional tree can be selected. The reason for this is that only these basic dimensions are connected with inference trees, and it is in the inference trees that activity attributes are contained. These are then included in a technique for dynamically editing activity attributes.

山西总体发展研究专家系统 · RELATIONAL MATCHING MODEL IASAS

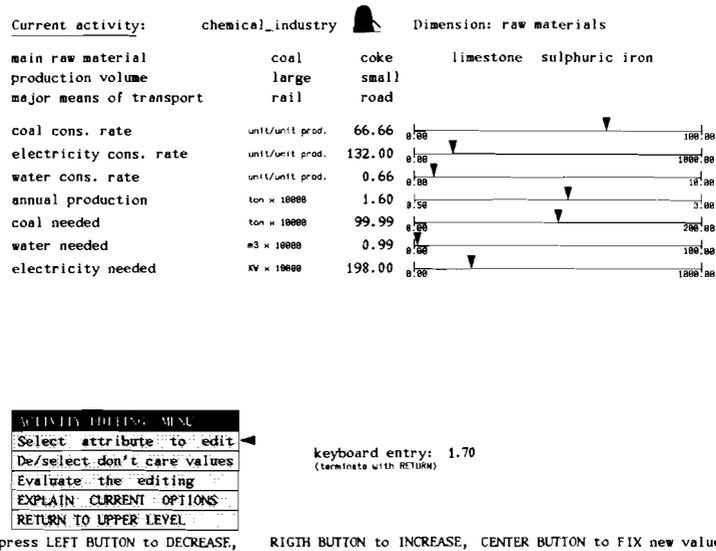


Figure 5.19: Editing the activity attributes

5.7.6.1 Dynamic editing of activity attributes

It was decided to implement the editing of the activity attributes in such a way that after the user modifies the values of the attributes, the system can show the consequences of these modifications for the inference process. A possible solution for this is a process in which the editor simulates the inference on the basis of the current activity and current values of the inferential activity attributes. In other words, it searches its route through the inference tree, guided by the current values of the inferential variables. In the start-up situation these are all the default values. The result of these default settings is therefore a default inference, since the total set of activity attribute values is logically equivalent to a specific route through the inference tree.

If the user tells the system he wants to change an inferential activity attribute, a change in its value may or may not lead to a different path in the inference process. How can that be shown? It was decided to provide the user with only a subset of the inferential variables. This subset contains only those variables that represent the appropriate route through the inference tree on the basis of the attribute values as they are at that particular moment. If the user chooses to change the value of one or more of the attributes in this subset, the 'Evaluate the editing' option gives control back to the inference

program, which will process the inference tree on the basis of the values of the attributes at that moment, including the change just made. In case this change affects the route taken through the network, it will be shown at the next editor action since by then the subset of attributes that can be edited will have changed. If, on the other hand, the subset did not change, the user knows that the modification of the attribute value just made was not significant.

Apart from changes in the set of attributes, this process will also update all the values of dependent attributes that need to be modified as a result of the editing of the independent attribute. It is worth noting that this approach is dynamic in the sense that modifications in one attribute can effect the 'editability' of other attributes. In terms of the relational meaning of the variables: setting certain variables to certain values can rearrange the inference track, thereby causing new categorizations of other variables. This dynamic element, however, implies specific requirements for the part of the program that simulates the inference and builds up the subset of variables that can be edited at the next editor action.

First of all, there must be an inference simulator, a routine that simulates the inference by finding its way through the inference tree according to the rule base on the one hand and the values of the actor attributes as they are set during the editing on the other. This, however, is not really problematic since this routine is very similar to the one that guides the actual matching and which has to be available anyhow. A more difficult problem concerns the building of the subset of changeable attributes, especially if case-dependent variables are involved. The technique applied here is that although for each editing action the complete activity attribute list is put on the screen, the editor determines which of them can be edited by comparing each of the attributes with each of the ones in a list built up during the simulation of the inference. As long as the variables in this list are independent variables, this is a straightforward procedure. However, in case of a dependent attribute, things become more awkward and a special procedure must then be followed. The problem to be solved here can be described as follows. The inference simulation is driven by the content of the inference tree, as well as by knowledge about current values of actor attributes. A dependent attribute, however, is nowhere in the data base represented by a fixed value, whereas the variables needed to compute that value are not necessarily present in the inference tree. The inference tree contains only inferential variables and from the example above it becomes clear that these can be dependent ones. This means that while the inferential attribute itself should never be edited

(it is dependent), the basic attributes of which its value is a function and which therefore should be changeable, cannot be reached by the inference simulator. To solve this problem a set of ‘actor_demon declarations’ was added to the actors data base. Each of these declarations is a slot of the form:

```
actor_demon(Inferential_attribute,value,Demon_attributes).
```

where Demon_attributes is a list of the actor attributes contained in the demon for Inferential_attribute. In case the inference simulator encounters an inferential attribute that can be recognized as a dependent attribute:

```
Attribute(Object,if_needed,Rule),
```

it marks this attribute as ‘un-editable’, looks in the corresponding actor_demon clause for the list of associated attributes, and appends this list held by the inference simulator, so that when this attribute is encountered by the editor, it will know that it can be edited.

There are, however, some problems associated with this technique. A rather serious disadvantage is that the order of the actor variables in the actor’s attribute list becomes relevant. It should not be possible that an attribute in the attribute list is processed as ‘un-editable’, while later on, because of this demon dependency, it is added to the simulator’s list of attributes. Another problem concerns the actor_demon declarations themselves. Basically, they are redundant and should, if possible, be avoided. Perhaps future versions will be rid of this problem.

5.7.6.2 Select/deselect ‘don’t care’ values

The ‘Edit current activity’ option does not only allow the user to assign new values to activity attributes, it also enables the setting or undoing of ‘don’t care’ values. The ‘editability’ of attributes when ‘don’t cares’ are considered is somewhat different from the editing process just described. ‘Don’t care’ values can only be assigned to inferential variables; whether they are independent or not does not matter (refer to the ‘production volume’ example above for clarification). However, not every inferential variable can get assigned a ‘don’t care’ score. Numeric attributes have the problem that even though they might, conceptually, get a ‘don’t care’ value assigned, their value might be needed to calculate dependent variables. If this is the case, then with what discrete value should a ‘don’t care’ be associated? For that

reason it was decided that no numerical inferential variables can get 'don't care' values assigned; they are reserved for categorical variables only. Furthermore, those categorical variables that are dependent ones, and of which the numeric variables they were derived from are needed in the matching process in order to derive yet other numeric variables, cannot get 'don't care' values assigned either. Allowing this would mean that the functional relation between the independent and dependent attributes is given up.

Once an attribute gets a 'don't care' value assigned (Figure 5.19) normal editing is inhibited until the 'don't care' is undone. It will be clear that with these kinds of data dependencies, determining the 'editability' of variables and determining whether or not variables can be subject to generalization, is a rather complex procedure, which required quite a lot of editor-Prolog control and matching program interaction.

An extra complication associated with 'don't care' values is that of how to handle them in the explanation of the matching results. Because a 'don't care' value implies that more than one route through the inference tree can be followed, failure to match the requirements means that all these alternative routes must also be part of the explanation. The example of an explanation screen in color Plate 6 shows how this is solved in REPLACE. As can be seen, explanation is dimension-specific, but since the attribute 'major means of transport' was assigned a 'don't care' value, various attempts at matching for the 'transportation' dimension were carried out. These are represented with the various entries for the 'transportation' dimension. The different scores associated with the 'major means of transport' attribute represent the alternative routes through the inference tree as a result of the 'don't care'.

5.7.7 'Edit county data base'

One of the possible applications of a relational matching model of site suitability is the exploration of possible consequences if characteristics of either the activity or the empirical environment (the locations) change. An automated system should therefore be equipped with possibilities to modify the data bases so that county data can be replaced or erased and modifications and updates can be implemented. For these reasons a county editor was developed (Figure 5.20).

The editor operates in much the same way as the activity editor, but its control is by far not as complex. Editing a county is, of course, county-specific. Therefore, after the user selects the 'Edit county data base' option,

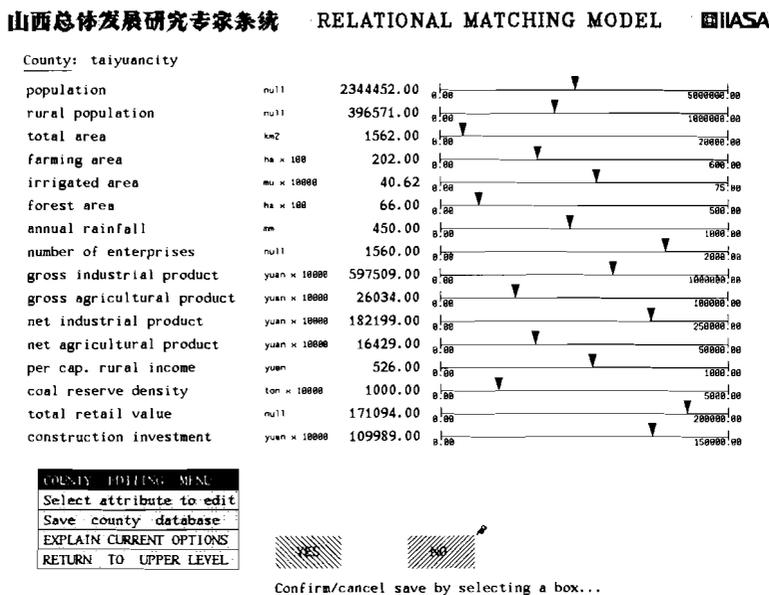


Figure 5.20: Editing the county data base

he is first prompted to select a county for editing. This is done by a routine that is similar to the one displayed in Figure 5.18. The only difference is that in this case only one county can be selected at a time.

Editing county attributes happens in the usual way: by pointing at the attributes and dragging along the number bars, by entering values by means of the keyboard, or by selecting the required category in case of nominal variables. The editor works in page mode, so that the user can step through the pages until the page with the wanted attribute is displayed. Since the user may want to make modifications permanent, a 'Save county data base' option is provided. After selection of this option, the user is asked for confirmation (Figure 5.20). If the saving is confirmed, a new version of the county data base is installed. The old version is retained as a back-up copy. Only one back-up copy is retained.

5.7.8 'Set decision attributes'

The user can specify some global parameters that are taken into account when the matching takes place. These parameters tell the matching system what to do in case specific situations occur. Three of these are defined in the current version of REPLACE: missing activity information, missing county

information, and numerical mismatch. Numerical mismatch cannot really be regarded a 'special' circumstance. Within the framework of relational measurement there is nothing special about it. If the requirements by either a county or activity are not met by the counterpart; activity or county, the matching fails. However, the sensitivity of a matching is something that deserves some attention. What, for instance ought to be done, if a firm requires the county to have a 1000 units of something, and only 989 are available? Strictly speaking, the county does not match the requirements by the firm and no relational match may occur. But it can be interesting to get an idea of the robustness of the resulting matching patterns by allowing for a margin or tolerance of numerical mismatch. The following rules for a percentage tolerance level T , were implemented.

$$\begin{aligned}
 a = b : & \quad \left[\frac{a+b}{a-b} \right] * 100 \leq T; \\
 a < b : & \quad \frac{b}{100} * T + b > a; \\
 a > b : & \quad \frac{b}{100} * T - b < a; \\
 a \leq b : & \quad a = b \text{ OR } a < b; \\
 a \geq b : & \quad a = b \text{ OR } a > b.
 \end{aligned}$$

The default value of this parameter (the 'matching tolerance' from Figure 5.15) is set to 0.0, the maximum value is 0.15. The current value of the parameter is checked by the matching program, whenever a numerical requirement has to be evaluated. Detection of this type of requirement is done by comparing the requirement operator with the list [$<$, \leq , $=$, \geq , $>$].

The other two global decision parameters tell the matching mechanism what to decide on in case either activity or county information is missing. This constitutes a problem, because it introduces indecisiveness into the model. If an activity requires 1000 units of something but the amount of units that a location can provide is unknown, what should the outcome be? By the same token, if the inference trees declare that in order to calculate how much an activity needs of something A must be multiplied by B, but B is unknown, what implications does that have for the matching? It was decided to give the user the opportunity to choose between two options: requirements in which unknown information is involved either fail or succeed.

The 'unknown actor scores' and 'unknown object scores' parameters (Figure 5.15) tell which of these alternatives should be used for missing activity

and county information respectively. If the user specifies the 'succeed' option, requirements involving missing information succeed, but only if additional parts of the requirement, if any, do not fail to be satisfied. As with the introduction of 'don't care' values, the possibility of influencing the matching process by these 'what-if-information-is-missing' parameters must be taken into account during the explanation of the matching result.

5.7.9 REPLACE data-analysis module: a dedicated GIS

The main menu contains one more option not yet dealt with: 'REPLACE data-analysis'. Selection of this option activates a more or less separate data-analysis module, connected to REPLACE in several ways, but basically independent. It constitutes a prototype of a dedicated GIS.

The term 'dedicated GIS' requires some explanation. The Shanxi Province DSS can be considered a 'dedicated' DSS. It is a DSS of which the models and their conceptual and implemented interactions, the representational formats (looks), the decision structures, etc., are attuned to the Shanxi Province economic developmental circumstances. Of course, it contains numerous elements that can be used and combined into other DSSs to serve other, similar, or perhaps quite different goals. But a number of features such as the choice of models, the mapping, the geographical aggregation levels, etc., are typical for Shanxi or perhaps for a Chinese Province. In that sense, the DSS is 'dedicated'. And just as the system as a whole is constructed for tackling a specific problem, many of its models and data bases are tailored in order to be compatible with the typical situation of Shanxi Province.

Geographical information systems provide ample opportunities for storage and retrieval of geographical data. Several large scale GISs have been developed and are available (e.g., ARC/INFO and GRASS), some of them equipped with impressive abilities to transform data to whatever kind of aggregations, to generate multiple overlays, three dimensional bird-perspective maps, and so on. These massive packages are good in what they were developed for: storage and retrieval of large quantities of spatial data in many different forms and combinations. As mentioned in the introduction to the Shanxi Province DSS, provision of information, even standard statistical information, is an important component of decision support. The more so, if this standard statistical information can be merged with model results into, for instance, a map-like format. But because of the 'dedicated' character of the Shanxi Province DSS, and also because of the user-friendly, menu/mouse

driven, graphical character of its implementation, large scale GISs can hardly be used for this purpose. The same thing that makes these large systems so good at what they are good at, i.e., their size and the multitude of possible operations offered, often make them difficult to handle. Of course, the relational nature of concepts applies to what is 'difficult' too, but it is well known that for non-specialists these systems can be rather hard to use and maintain. A system like the Shanxi Province DSS, however, is made for decision makers, not for GIS specialists.

The problem, therefore, is to find a way to be able to store and retrieve geographical information in such a way that it satisfies the needs for statistical and mapped information relevant to the problem, without having to deal with the drawbacks of working with large and unwieldy systems to get the job done. Dedicated GISs may offer a solution here. A dedicated GIS, i.e., a GIS of which the capabilities are attuned to the data, data structures, spatial aggregation levels, mappings, and predominant types of analyses that can be expected given a specific problem, can be constructed in two ways: either by decomposing an existing GIS and composing a new GIS from some of its parts complemented by new cross-connections and interfaces, or by developing it from scratch. In this case, the latter option was chosen, mainly because of the limited time available for implementation. For this small and dedicated GIS, several tasks were recognized as being of importance:

1. Uni-variate statistical data description;
2. (Re)classification of county attributes;
3. Bi-variate statistical analysis;
4. Matching results must be treated as variables;
5. Mapping on a county-specific basis.

Figure 5.21 shows the screen after the user selects the 'REPLACE data-analysis' option from the main menu. The screen shows the map of the province containing the categorized scores of the attribute displayed on top of the screen, some uni-variate statistics, three frequency histograms, explanation of the colors, and a menu containing further options. 'Rural population' is the default attribute the GIS is started up with. Other attributes can be selected by choosing the 'Select and display attribute' option from the menu. This will then invoke a screen that enables the user to select any of the attributes available in the system.

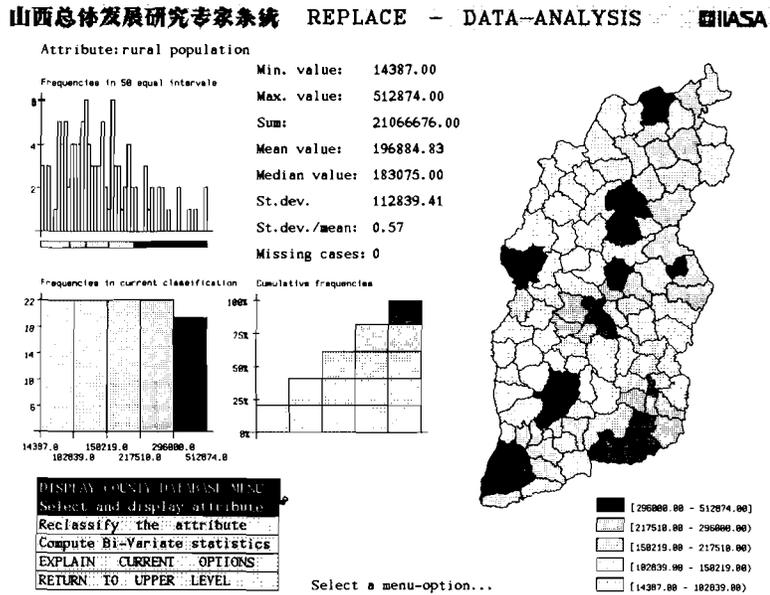


Figure 5.21: Display of a county attribute

After selection of an alternative attribute, the screen of Figure 5.21 reappears, but now showing information for the newly selected attribute. Three frequency tables are shown: equal interval frequencies, frequencies of the current classification, and cumulative frequencies of the current classification. Attributes are always categorized. Numerical variables are by default categorized in 5 equal count classes (identical numbers of observations per class). This is reflected in the structure of the two bottom frequency graphs in Figure 5.21. The top frequency graph displays numbers of observations in 50 equal intervals. The colored bar beneath it corresponds to the current classification. The values associated with the current class boundaries are shown in both the explanation beneath the map, and the classification frequency graph. In case variables are indeed numerical, uni-variate statistics are displayed. Nominal variables do not need categorization by the system. Neither do an equal interval frequency graph and uni-variate statistics apply. These are therefore not displayed. Missing values ('unknown' information) are displayed as white, uncolored counties on the map.

5.7.9.1 'Reclassify the attribute'

In order to enable the user to reclassify numerical attributes into a categorization that better suits his purposes, a special graphical reclassification routine was developed (Currently the GIS does not offer reclassification for nominal variables). Figure 5.22 shows a screen from a reclassification of the variable displayed in Figure 5.21.

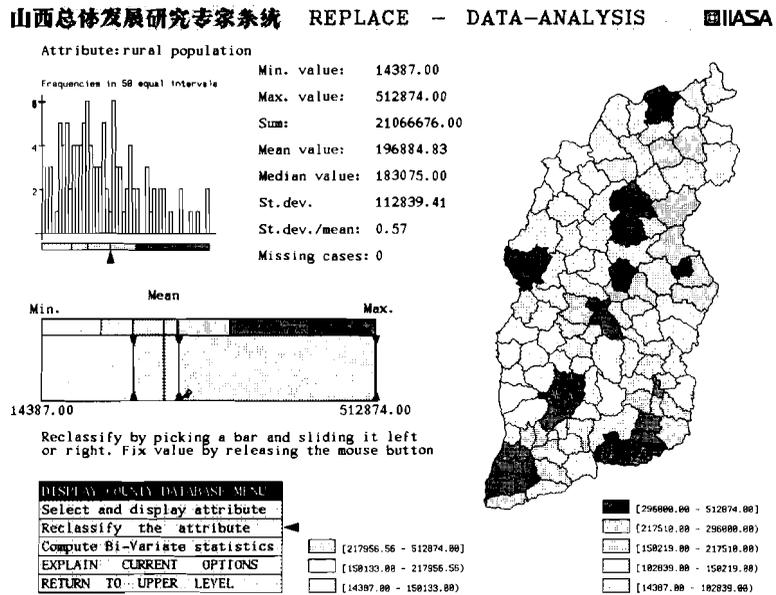


Figure 5.22: Reclassifying a county attribute

The map, explanation, and statistics are left unaltered. The remaining two frequency graphs have been replaced by a (re)classification box. Minimum, maximum, and mean value of the attribute are shown. The bar on top of the box shows the current classification boundaries; they correspond with the bar under the frequency graph. Inside the lower part of the box, variable class boundaries are displayed by vertical lines, ending in arrowheads. They correspond to the values displayed on the right of the menu. These lines can be dragged left or right in order to alter the classification boundaries. The arrow under the bar beneath the frequency graph indicates the current position of the dragged boundary. Reclassification is terminated by selecting the 'Reclassify the attribute' option again. Color Plate 4 shows

the result. This kind of reclassification operation enables the user to study the statistical and spatial distribution in a rather flexible manner.

5.7.9.2 'Compute Bi-variate statistics'

The GIS component of REPLACE offers the user the opportunity to calculate some bi-variate associations between variables. Currently only a simple, two-variable linear regression and a chi-square test for statistical independence of categorized variables are implemented. Although many more techniques for statistical analysis of the Shanxi Province data can be thought of, these two yield some interesting possibilities for the exploration of statistical relations, especially when combined with the spatial distributions of their variables. Naturally, regression can only be conducted for interval and ratio variables. If regression is requested for a problem containing a nominal or ordinal variable, the system will tell the user that he does not really want that. Figure 5.23 shows the screen after a linear regression has been conducted.

The screen contains a scatterplot of the points, the regression line, the parameters of the regression and significance test, as well as a map of the residuals categorized in five equal interval classes. The x- and y-axis of the scatterplot are linearly scaled with respect to the minimum and maximum values of both variables, the minimum values corresponding with the origin, the maxima with the endpoints of the axes. As illustrated by the plot in Figure 5.23, linear scaling of variables with a high skewness level results in scatterplots in which many points are concentrated in a very small area. This corrupts the visual relation between the regression line and the cloud of points. In the current version of REPLACE, however, no alternative scaling can be selected. In a future implementation, the user might be provided with an on-screen interactive utility by which he can 'jump' between alternative scalings (e.g., linear, logarithmic, square root, etc.). Likewise, future versions of the regression test must contain options for transforming skewed distributions toward bi-variate normal distributions, in order to get more meaningful significance estimates.

Whereas regression can be used for numeric attributes, a chi-squared test for statistical independence can be applied to a combination of categorized variables. And since, for mapping reasons, numeric variables are also categorized, they can be included in a chi-square test as well. Statistically there is of course no reason to conduct a chi-square test on two numeric

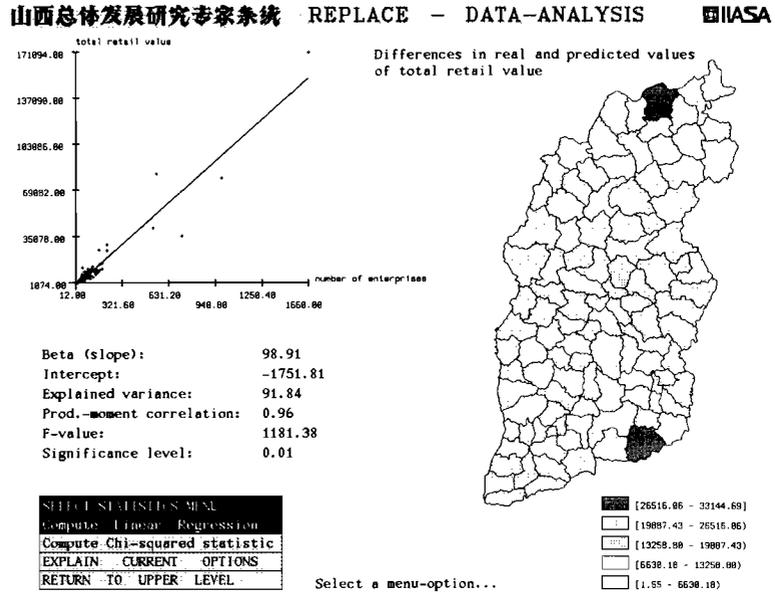


Figure 5.23: Display of the regression results

variables, but combinations of a numeric variable with a nominal one can be interesting. Such a combination is displayed in Figure 5.24.

The test contains the dichotomous variable 'access to high tension line', and the numeric variable 'total retail value', classified into two categories. The screen shows the spatial distributions of both variables, the observed and expected contingency tables, and the chi-square test parameters.

All in all, the GIS part can be considered an important functional augmentation to the REPLACE site suitability measurement system. Without a utility that can map statistical information and the spatial distribution of variables, it can be rather difficult to interpret matching results, or to compare matching results with existing distributions of attributes.

It was mentioned that the GIS regards the results of the matching model as a normal dichotomous (match/fail) county attribute. The possibility of thus linking into model results, opens up a whole range of interesting applications. If, for instance, the site suitability measurements from the relational matching model can be considered a normal variable, then this also applies to site suitability measurements stemming from any other model. Since the GIS offers the opportunities to compare distributions of variables with each other, both statistically and spatially, various models can be compared and

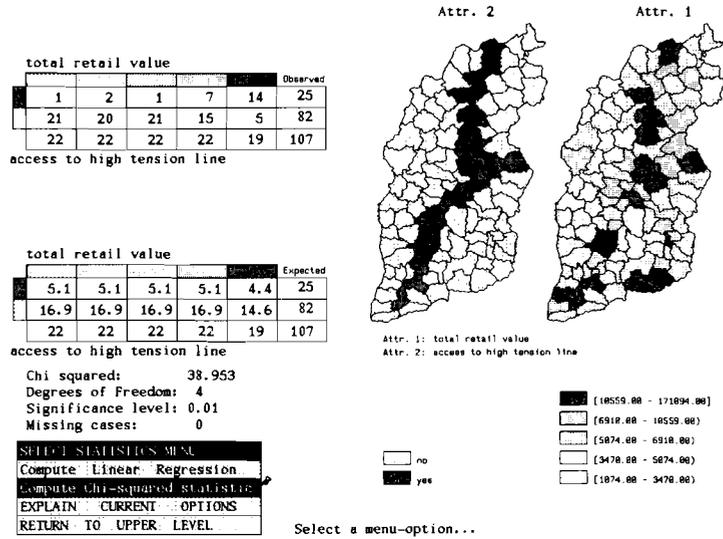


Figure 5.24: Display of the bivariate spatial distribution of a classified variable and chi-squared statistic

evaluated with respect to each other. An application that is mentioned in the next chapter is that in which the knowledge acquisition process is supported by the interaction between matching model, GIS, and researcher or domain specialist.

5.8 Conclusion

This chapter began with the assertion that many scientifically interesting and empirically applicable models are often hard to use, because they are somewhat inaccessible to the people who might want to use them. Especially in situations of complex decision making, formal representations of various parts of these situations might be contained in models that can be integrated by means of a DSS. Several characteristics of DSSs were discussed, and it was argued that they might offer a useful approach to the problem of decision making in spatial planning. As an example, the problem of the (re)organization of the regional economy of Shanxi Province was discussed, together with some aspects of a prototype DSS developed at IIASA. The

REPLACE system forms part of this large scale DSS. Several conceptual relations between the relational site suitability matching model and other models from the overall DSS were mentioned, and a rather detailed overview of the characteristics and implementation of REPLACE was presented. One last issue to be discussed here is the question whether a system such as REPLACE (or such as REPLACE was meant to be), can function as a DSS on its own. I think that a meaningful answer to this question implicitly refers to the relational nature of a concept such as 'DSS'. The definitions of DSSs at the beginning of this chapter were of a typical functional character. A computer-based system is a DSS if it can support decision making. Determining whether a specific system can actually function as a DSS, however, requires a relational definition in terms of the requirements of the actor (the decision maker) and the properties of the system. Clearly, there is nothing inherent in any kind of system that makes it into a DSS. Only if the decision-making situation and system match each other can a system function as a DSS. In more application-oriented terms: in situations where decisions have to be taken about where to locate a specific activity, or about where and which locational circumstances must change so that a location becomes suitable for specific types of activities, can a system such as REPLACE be considered a DSS. These situations frequently occur in establishing regional policy measures as well as in spatial decision making by spatial activities themselves. If, however, the decision-making situation is as in the case of Shanxi Province, REPLACE can be no more than part of a DSS; one model, one perspective on the regional system that can be combined and integrated with many others in order to arrive at a well-balanced decision.

Chapter 6

Examples and Applications

ABSTRACT

In this chapter some empirical work on the assessment of site suitability in Shanxi Province is presented. The examples discussed are of a generic character. Rather than an attempt at reconstructing some precise and accurate measurements of site suitability for some activities, the examples address more general problems such as how to gradually build up a relational model, how to interpret the map patterns generated by a relational site suitability model, and how to represent the concept of ‘linkages’ by means of a relational matching model. As such, they serve to evaluate some of the merits, as well as problems, associated with the method of relational matching modeling presented in the previous chapters.

Keywords: matching pattern, linkages, optimization, multi-criteria evaluation, robustness

6.1 Introduction: REPLACE in the Shanxi Province case study

In the previous chapter, the Shanxi Province case study was introduced. This case study was discussed from the perspective of the potential DSSs offer for model-based integrated development in general, and regional economic planning and spatial decision making in particular. As a separate

module in the much larger Shanxi Province DSS, the prototype implementation of the REPLACE site suitability matching system was discussed in detail. Emphasis was on the generic properties of the system: the functions it contains, how things are implemented, what they look like on the screen, and how the user interface was configured.

Ample attention was also given to what was termed 'problems of data-dependency' and their implementational consequences. This 'dependency' refers to the variation in the meaning and categorization of variables as well as the interpretation of seemingly unambiguous and straightforward empirical data as a function of where and how these data are used in the relational model. From the point of view of relational analysis, this qualitative variation in the meaning of data is one of the most important aspects of the methodology. Depending on the characteristics of actors and the functional relations they maintain with objects, the empirical properties of either of them must be interpreted differently. The categorization of variables is often the result of a score one or more other variables get assigned.

Another reason for the way in which these data dependencies turn up in the application of a relational matching model might be associated with the computational nature of such a model. Although mere conjecture, it does appear as though there is a rather strong positive relation between the intelligence of a model or system and its computational complexity. Notions such as 'don't care' values, 'unknown object scores' parameters, or 'inferential' variables, presume a certain level of semantic interpretation, something which appears to go hand in hand with increasing overhead and complexity in computational effort.

Models such as REPLACE's core matching model are characterized by a decreasing separation of model and implementation, representation and process, the declarative and procedural part. Models in which this distinction is more or less absent, for instance, specific types of process models the implementation of which is a model of the process, are sometimes referred to as 'computational models' (Coulclis, 1986a, 1986b; Smith, 1983, 1984). Prolog is a programming language that is very well suited for building these types of models. The nature of Prolog is such that there is no distinction between data and program, model and process. Of course there is a very distinct separation on the level of implementation, but on the level of analysis done by the model-cum-program, the distinction vanishes. These, and other, issues dealt with in previous chapters all pertain to the formal characteristics of the relational matching model contained in the REPLACE system. The empirical contributions, however, have as yet only been mentioned in

connection with the overall objectives and goals associated with the Shanxi Province DSS. They were only implicitly presented as rather disconnected pieces of information, hidden in examples and sample screens.

In this chapter some of these first empirical applications of the REPLACE system are discussed in somewhat more detail. They were closely connected with the goals as determined by the Chinese contributors and had rather significant implications for the provision of data, both in terms of activity data such as rule bases, and locational data. Some remarks and examples of the solutions that had to be found in order to overcome these problems are mentioned. The overview of the data aspects will give the user a healthy distrust regarding the empirical validity of the example applications but it also generates some discussion on the sensitivity of relational matching results to faulty input data.

Next, some rather generic examples of how a relational matching model could be utilized in empirical applications are discussed; generic, because they cannot be considered full-fledged and empirically sound. They are based on the Shanxi Province case study material and do show some of the strong and weak points of the relational matching approach as presented here. This chapter therefore has the character of a rather loosely organized set of examples and evaluations, set against the background of the Shanxi Province case study material. Some examples are considered self-explanatory or obvious. Others, however, serve to illustrate specific theoretical and methodological points; although most of them were dealt with in earlier chapters they are worth mentioning in a more application-oriented context.

6.2 Data problems

A problem which seriously depreciates the empirical validity of the suitability measurements presented here is the quality of the data the model is run on. Most of the available locational data was on the county level. Therefore, it was decided to take the county as the locational unit; the 'object' in the relational matching model. Such a rigid choice, however, has disadvantages. Taking the county as the unit of analysis, for instance, implies that each county is considered internally uniform, something which is certainly not true. Further, there is no good reason to expect that the county is the basic unit of choice when locational decision making is considered. But necessity knows no law, therefore for reasons of data availability it was decided to choose the county as the unit of analysis in the site suitability measurements.

Table 6.1: Basic county attributes

Attribute	Min	Max	Unit
Total population	0	5000000	
Total area	0	20000	square km
Farming area	0	600	ha × 100
Forest area	0	500	ha × 100
Annual rainfall	0	1000	mm
Number of enterprises	0	2000	
Gross ind. product	0	1000000	yuan × 10000
Gross agr. product	0	100000	yuan × 10000
Net ind. product	0	250000	yuan × 10000
Net agr. product	0	50000	yuan × 10000
Per cap. rural income	0	1000	yuan
Construction investment	0	150000	yuan × 10000
Coal production	0	5000	ton × 10000
Coal reserve density	0	5000	ton × 10000
Coke production	0	500	ton × 10000
Bauxite production	0	500	ton × 10000
Aluminum production	0	25	ton × 10000
Sulfur-iron reserve	0	2000	ton × 10000
Limestone reserve	0	30000	ton × 10000
Limestone % CaO	40	75	
Limestone % MgO	0	5	
Electricity production	0	500000	kwh × 10000
Total water supply	0	50000	cubic m × 10000
Per cap. res. water cons.	0	100	cubic m
Water exploitation %	0	100	
Steel production	0	2000000	ton
Iron production	0	1500000	ton
Cement production	0	1500000	ton
Fertilizer production	0	150000	ton
Wheat production	0	300000	ton
Corn production	0	500000	ton
Cereal production	0	200000	ton
Sorghum production	0	120000	ton
Tuber production	0	75000	ton
Students (gen. sec.)	0	150000	
Students (gen. mid.)	0	50000	
Students (prof. sec.)	0	5000	
Students (prof. mid.)	0	12000	
Pupils	0	500000	
Rural population	0	1000000	
Irrigated area	0	75	mu × 10000
Total retail value	0	200000	yuan × 10000
Pigs	0	10000	
Sheep	0	10000	
Cattle	0	10	heads × 10000
Workers in industry	0	100000	
Gross enterprise prod.	0	150000	yuan × 10000

Table 6.2: Provincial constants (scores are in units/units)

Provincial attribute	Score
Average ind. coal consumption coeff.	15.70
Coal export rate	0.65
Ind./res. coal consumption ratio	3.50
Average ind. water consumption coeff.	0.0834
Average agr. water consumption coeff.	306.0
Per capita res. rural water consumption	0.00073
Per capita pig water consumption	9.1
Per capita sheep water consumption	2.9
Per capita cattle water consumption	11.0
Average ind. energy consumption coeff.	17.4
Average agr. energy consumption coeff.	11.0

Table 6.1 shows the list of the 49 basic (non-derived) locational variables available for REPLACE. Table 6.2 shows some additional provincial constants. Most of these variables came from statistical sources, some of which only existed in the form of hand-written lists of variables and numbers, translated from Chinese. Many other variables, however, had to be inferred from contour maps such as the one displayed in Figure 6.1. It presents a map of the five major coal fields, superimposed on the map of the counties. The coal fields are divided into two classes of densities: 1000 and 2500 units of coal per square kilometer (1 unit = 10000 tons). Why the classes are discrete scores rather than intervals is unknown. This was all the data that was available.

This kind of data had to be used for the derivation of other information. In one of the applications, for example, an estimate of the total coal reserve per county was required. Given the data of Table 6.1, and the coal density map in Figure 6.1, the only possible way of inferring this attribute's value was first, to assume that the coal density is uniformly spread throughout the fields, and furthermore, that the coal reserve per county can be calculated by multiplying the coal deposit density by the county's area. But in order to do this, the contour map information had to be converted to county-specific information. It was in these situations where the need for a good GIS was felt most strongly. Given the map in Figure 6.1 and given a rule saying that, for instance, the reserve equals the density times the area covered by the coal field, a good GIS would be able to calculate the estimated reserves for

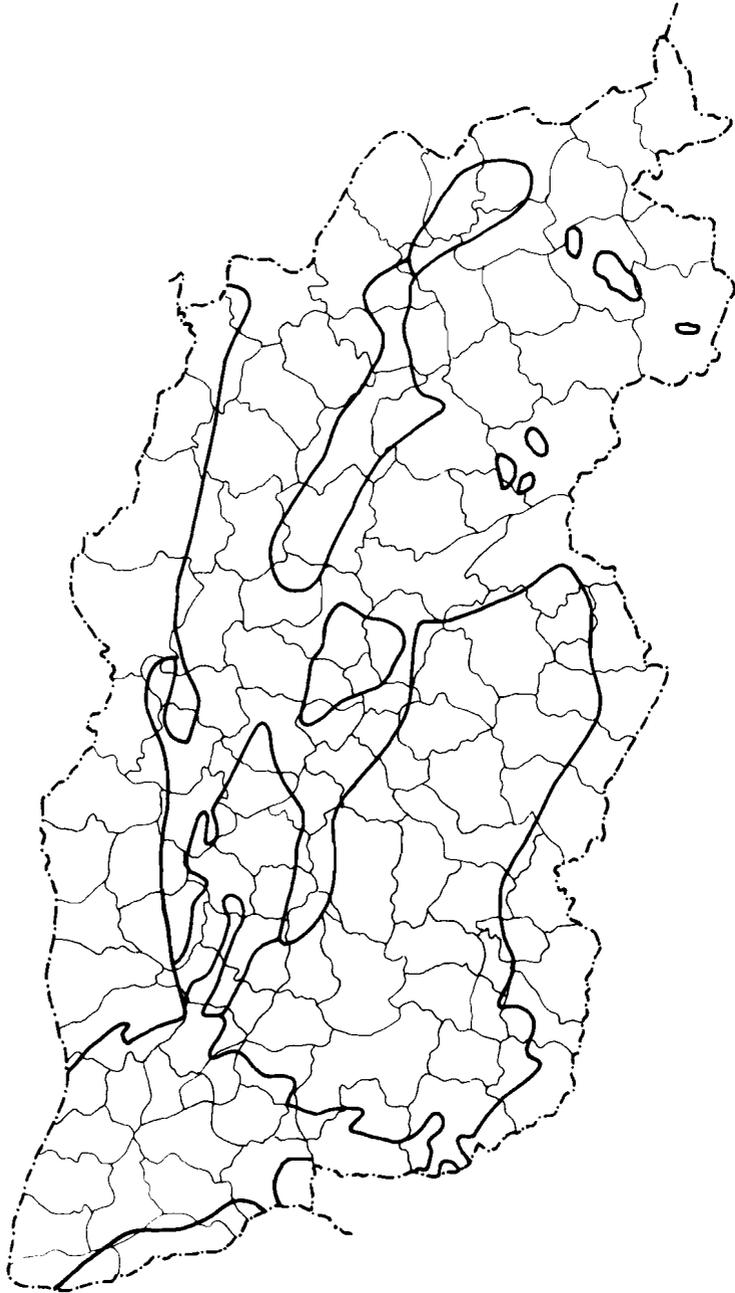


Figure 6.1: Distribution of the coal reserves

each of the counties. This case study, however, lacked such a system or a clear-cut methodology as developed by, for instance, Flowerdew and Green (1989) for inferring information across incompatible zonal systems. Instead, each county was assigned to a coal field if a significant part of its area was covered by it. 'Significance' had to be determined by interpretation. Formal raster-based procedures are of course a much better possibility, and this is the kind of task a traditional GIS can perform easily. To conduct them by hand, however, would not only have been extremely time-consuming, it would also have been very tedious and almost certain to generate mistakes.

Another example of how derived variables had to be computed is that of the calculation of the 'available water surplus' as displayed in Figure 6.2. Of course, for this particular variable much better estimates can be produced by a model such as MITSIM (Section 5.5), but at the time, these inter-model linkages had not been set up, so again procedures like the one in Figure 6.2 for computing the derived attribute's value had to be designed.

Looking at the list of variables in Figure 6.2, one may be a little surprised to discover entries such as the 'per capita water consumption per pig', and the total number of sheep. It is also surprising that Table 6.1 contains the rural population and the total population as basic variables, whereas the urban population must be derived by subtraction. Things might be biased because of the composition of the group of Chinese scientists that collected most of the data and the difficulties that were experienced in getting data over from the Province; perhaps it is just a misinterpretation, but statistics of this sort might be typical for a country such as China, or for Shanxi Province. It was interesting to learn that the average amount of water sheep drink over the year was known, whereas the total mileage of the railways was unknown, or could not be provided from public statistics. It appeared that many variables that only apply to the provincial level and that therefore could not be collected on a county-specific basis, were either missing or very roughly categorized (the example of the coal-field densities). Surprisingly, in spite of the importance of coal production, only a few county-specific coal production figures were known. What appears to be the case, is first, that the collection of information on attributes for which the county is not an adequate administrative unit (such as coal mining), is very limited. This again seems to indicate that collection and maintenance of statistics is something that is indeed done at the county level, or, given the often very detailed agricultural information, in communities within the county. That the rural population is a basic variable whereas the urban population is only

W_RES	:	Available water surplus (m3 * 1000000)
*W_PROD	:	Total water supply (m3 * 1000000)
W_CONS	:	Total water consumption (m3 * 1000000)
W_ICONS	:	Industrial water consumption (m3 * 1000000)
*W_AICONS	:	Average industrial water consumption (m3/10000 yuan)
W_ACONS	:	Agricultural water consumption (m3 * 1000000)
*W_AACONS	:	Average agricultural water consumption (m3/mu)
W_RUCONS	:	Residential urban water consumption (m3 * 000000)
*W_CRUCONS	:	Per capita residential urban water consumption (m3)
W_RRCONS	:	Residential rural water consumption (m3)
*W_CRRCONS	:	Per capita residential rural water consumption (m3)
W_ARCONS	:	Animal rural water consumption (m3 * 1000000)
*W_CPIG	:	Per capita water consumption of pigs (m3)
*W_CSHEEP	:	Per capita water consumption of sheep (m3)
*W_CCATTLE	:	Per capita water consumption of cattle (m3)
*IGO	:	Industrial gross output (yuan * 10000)
*A_PROD	:	Agricultural production (yuan 10000)
*I_AREA	:	Irrigated area (km2 * 1000)
*U_POP	:	Urban population
*R_POP	:	Rural population
*PIGS	:	Total number of pigs
*SHEEP	:	Total number of sheep
*CATTLE	:	Total number of cattle
$W_RES = W_PROD - W_CONS.$		
$W_CONS = W_ICONS + W_ACONS + W_RCONS + W_ARCONS.$		
$W_ICONS = IGO * W_AICONS.$		
$W_ACONS = I_AREA * W_AACONS.$		
$W_RCONS = W_RUCONS + W_RRCONS.$		
$W_RUCONS = U_POP * W_CRUCONS.$		
$W_RRCONS = R_POP * W_CRRCONS.$		
$W_ARCONS = (PIGS * W_CPIG) + (SHEEP * W_CPIG) +$ $(CATTLE * W_CCATTLE)$		

Figure 6.2: Computation of the county-specific water surplus; variables marked by an asterisk represent basic county data

a derivative, might be an indication of the predominantly agrarian and rural orientation, even in a partly industrialized region such as Shanxi.

The type of data that is collected is further dependent on the relevance the data have for the people collecting them. It can be assumed that if agriculture is the prime source of income in many counties, the agricultural statistics are carefully maintained, whereas information on, for instance, across (county) border coal mining is considered to be a provincial matter and will therefore not, or only defectively, be recorded.

A general point to note here therefore, is that the reliability of the data used in the examples presented in the remainder of the chapter is difficult to assess. Many variables involved in the matching rules had to be estimated in a rather primitive manner, either from basic variables, or from other estimates. The example of the 'coal reserve' and 'water surplus' computations can be considered fairly representative cases.

In a research note on the potential of the 'decision nets' (briefly discussed in Section 3.3) Timmermans and van der Heijden (1987; p. 302) mention the absence of an error theory from intrinsically deterministic approaches such as decision nets and relational matching models. This creates a problem, because without an error theory such as the one inherent in a statistical approach, there is some uncertainty as to what ought to be done with measurement errors. Although this objection only applies to non-systematic errors—statistical error theory does not help much in case of systematic errors—this is indeed a serious problem that needs more attention. Quick and dirty solutions such as the availability of global matching parameters by which one can introduce a kind of tolerance factor into the measurements, are certainly insufficient. They nevertheless seem to offer some opportunities, the future study of which might yield some interesting possibilities for incorporating uncertainty into the model. One might think here of the inclusion of (un)certainly assessments considering the significance of specific requirements or actor and object attributes, and matching tolerance parameters that take into account pre-defined assumptions concerning the reliability of the object data. Although very important, however, a rigorous treatment of the problem of data-errors goes well beyond the scope of this study. It was therefore decided to accept the data as they were. Of course this seriously depreciates the value of the measurements conducted, but since the prime interest of this study was on the structure of the model, this was accepted regretfully.

Quite another aspect for evaluating the version of the relational matching approach presented here concerns the sort of questions and the various

types of analysis for which such a model can be used. It will be clear that different types of use will require different features and will therefore entail different evaluation criteria. In the following paragraphs a few examples of the purposes for which a model or system such as REPLACE could be used are discussed.

6.3 Generating production milieu maps

Although REPLACE can be used as a dedicated GIS, its main objective is the assessment of site suitability for economic activities, primarily for industrial production. The routines and options described in the previous chapter offer various ways in which these measurements can be conducted. Within a given type of activity, for instance, it is possible to conduct measurements with different sets of locations, different dimensionality, different global matching parameters, etc.

An interesting aspect of relational modeling concerns the status or significance of spatial patterns of production milieu. Consider, for example, Figure 6.3. It presents a matching result for the transportation dimension for coal-based chemical industries, with rail as a major means of transport. The possibility to define a major means of transport stems from the fact that no transportation cost functions were known. The requirements contained in this dimension have to do with the transport of raw material to the location of production and the transport of finished products.

Data on both the network of railways and major roads were only available as maps, to be superimposed on the county map. Because of the lack of transportation cost functions, it was decided to compute two types of variables that could be used as indicators for the quality of the infrastructure of the county: 'direct' and 'indirect access' to the railway system and 'access' to the system of major roads. A county was designated to have direct access if it was part of the network. Indirect rail access was defined as a combination of a road connection to a neighboring county which has direct rail access. According to the map in Figure 6.3, the accessibility of most of the counties in the province is sufficient for the specified activity type, except for an area along the western border, and a few counties elsewhere in the province. Another, much clearer 'pattern' is the one in color Plate 5. It constitutes a chi-square test for the matching result on the electricity dimension and the location of the high-tension line network.

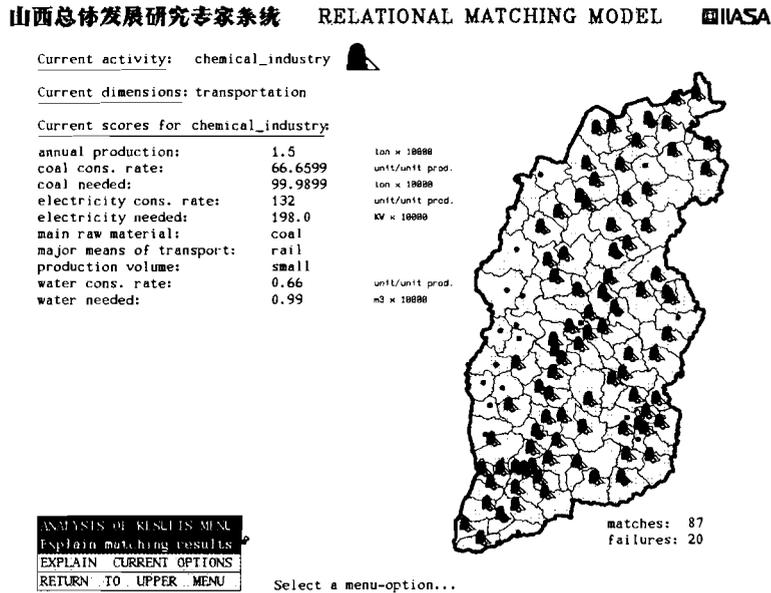


Figure 6.3: Matching result for transportation dimension

The meanings of the patterns and how they can be interpreted is of interest here, as the question may be raised whether they can be interpreted as spatial regularities in the distribution of production milieu. Of course they do represent spatial regularities in the distribution of the matching result. The more interesting question, however, is whether these patterns can also be considered to represent causal regularities in the generation of production milieu.

The answer to this second question should be a hesitant 'perhaps'. The reason for this is that it should be realized that the matching results are in fact nothing but the result of a set-theoretic, basically non-spatial operation. Of course, the matching itself contained spatial variables, but a resultant spatial pattern of matches does not necessarily indicate a causal spatial pattern as well. A matching model consists of conjunctive and disjunctive terms. Sometimes just a few (as in most of the illustrations considered here), but possibly very many. The disjunctive elements of a matching model offer the possibility of many different combinations of attributes and their values to be contained in the matching. For the resultant spatial pattern this means that identical patterns can be generated by more than one, and sometimes very many combinations of locational attributes and attribute values. A

proper interpretation of such a pattern must therefore include the matching process and all the locational attributes and value combinations involved.

Although this implies that one should hesitate to interpret the matching results as causally associated with the spatial distribution of locational variables, it does not mean that spatial regularities in matching can never be causally interpreted. From this argument, it can be inferred that if the number of locational attributes involved in the matching is small, the spatial pattern of matches must be causal, because the matching model itself is of a causal nature. In the matching of color Plate 5, for example, only a few locational attributes were included. A county was declared to match the electricity requirements if the electricity production in that county exceeded the amount needed by the activity, or if the county had access to a high-tension network. In this particular case, the spatial patterns could be causally interpreted. Comparison of the map of the matching results with the map of the access to high-tension lines, shows a clear relationship between these two variables, the only exceptions being a few counties that generate sufficient electricity without having to rely on the availability of a high-tension line connection. But even in this case, caution is required. It is very well possible that counties that do have access to the high-tension network, and hence match the requirement, also produce sufficient electricity by means of their own energy facilities. A complete explanation of this pattern, therefore, must include comparisons of the matching pattern with both the pattern for electricity generation and access to high-tension lines. It will be clear that for intricate matchings these kinds of investigations into the causal relationships hidden in the patterns becomes an extremely complicated matter.

6.3.1 Pattern interpretation and inductive knowledge acquisition

This issue of how to interpret patterns of matching and how to interpret possible associations between matching patterns and the spatial distribution of locational variables is not important just from an empirical or economic-geographical point of view. Recognition of pattern in data, be it spatial or statistical distributions, categorical associations, or deterministic rules, also plays a role in many inductive methods for generating knowledge in a specific domain.

In Section 3.4.2.1 classification optimization by means of information entropy was discussed. Applying such a technique as part of the knowledge

acquisition process was characterized as an 'inductive' approach to knowledge acquisition. The basic idea was that from a raw set of data the technique extracts the essential information and discards the redundancies. It was argued that this kind of inductive reasoning could be very helpful in the development of relational matching models, not only for reasons of efficiency, but also because the modeling process can only be improved and facilitated by showing the 'true' and essential content of a specific version of the model.

Another type of inductive approach one might want to suggest for developing relational matching models is based on the analysis of spatial patterns and pattern associations. The basic line of reasoning would be something like this: if the members in the matching set can be characterized by a specific combination of empirical attributes and values, and if this combination makes them differ from the locations not contained in the set, then this combination of attributes and values can be considered a sufficient description of the choice set.

Although the suggestion of a procedure for associating choice sets with a specific actor type does sound tempting, and although it might seem an almost natural way of separating the essential from the redundant, this kind of inductive generalization must be rejected. The reason is of course that it leads straight back to the empiricist, statistical, non-causal approach to definition and measurement, based on empirical similarity. What the procedure really describes is a kind of reverse cluster-analysis or multi-dimensional scaling analysis in which the clusters are already known; it is just the combinations of attributes that best describe the clustering that have to be found. It will be clear that accepting such an approach would be an implicit denial of the relational nature of concepts and their measurements. As stated several times earlier, there is nothing inherent in the empirical characteristics that determine objects (locations) as suitable to perform a specific function for a specific actor. Therefore, it simply cannot be valid to reconstruct choice sets on the basis of their empirical properties alone. Of course, for this specific combination of activity and set of locations, the technique would produce valid results. But there is no reason whatsoever to expect that if another object is added to the initial set of objects, or if a completely different initial set of objects is taken, the measurement will again produce valid results. The functional equivalence of objects in a choice set does in no way presuppose empirical similarity between these objects. It is the result of a matching of the requirements generated as a result of empirical characteristics of both actor and object, and the empirical properties of the other.

Procedures that reduce functional equivalence to empirical similarity must therefore be rejected.

6.3.2 Iterative model building: aluminum industry

As an alternative to the application of inductive techniques of knowledge acquisition based on empirical similarity, it may be useful to explore some of the potential REPLACE has for being employed in the iterative process of model building advocated in Chapter 3 (Figure 3.10). Since REPLACE displays model results in a map format and can be asked to explain the matching result, it may be worthwhile using the system as a means to test the content of, and modifications made on, a specific relational matching model. In this proposal for iterative model building it is suggested that one start with a simple initial model that can then be modified and extended during several rounds of evaluation. The evaluation itself involves the execution of the model by means of the matching system, and an assessment of the quality of its results by either a comparison with observed frequencies and patterns, or by a domain expert.

The difficulty of developing proper tests for the accuracy of the model results was also mentioned. Generally, past location decisions are not good candidates for such a set since they may have been taken on the basis of a very different situation. Locational decisions may have been taken on the basis of either different objectives or actor characteristics, or a different empirical environment. Changes in either of these three factors: actor characteristics, actor objectives and locational properties may change the outcome of a matching operation dramatically. Process validation, on the other hand, requires the availability of domain experts, and those were only marginally available during the work on the Shanxi application of REPLACE. An example involving aluminum production is discussed in order to give an idea of what such an iterative process of model development might look like. Since the Shanxi province collaborators showed an interest in investigating possible investments in the aluminum production and processing industries, some work was done on the construction of an initial rule base for both types of industry. One of them, aluminum production, is therefore taken as the example in an illustration of how REPLACE could be used in the process of iterative model development.

For practical reasons it is assumed that the existing set of aluminum production plants in Shanxi Province constitutes a proper test set for the model. Under this assumption, looking for cases in which the model finds

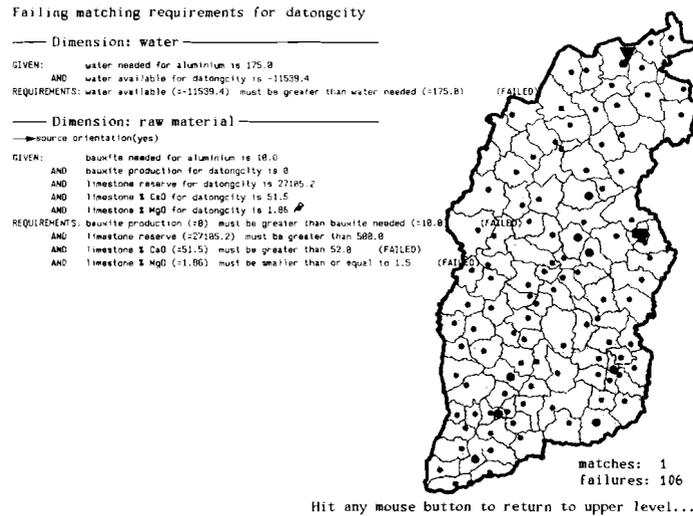


Figure 6.4: Full matching for source-oriented aluminum production

a negative matching result while the activity that is matched for is indeed located at that specific site, is a possible test procedure.

Figure 6.4 shows the results of a full matching (all locations included and full dimensionality) for source-oriented aluminum production, and some explanation on the failure of the county of Datong city to meet the requirements. Only one county (Yangquan city) matches the requirements. None of the actual production sites for aluminum (Ningwu, Taiyuan city, Xiangning and Yuncheng city) is contained in the set of suitable sites. Under the assumption that these are indeed suitable sites, the model thus fails to include them in the choice set. That Yangquan city is included although no production takes place there is of course no problem.

One way to try and find possible errors in the model is to review the results of dimension-specific matchings. This is a valid procedure because the total matching is the conjunctive result of a series of consecutive dimensional matchings. At the time the examples discussed here were worked out, however, dimension-specific sensitivity analysis could only be conducted by comparing various one dimensional matchings with each other and with various multi-dimensional results (refer to Section 3.2.4 for some ideas on how to implement more complex forms of inter-dimensional comparisons).

The matching for the dimension 'raw material', for instance, generates the same result as the matching for the total dimensionality. Therefore, 'raw material' must be one of the dimensions that is responsible for the errors. Inspection of the explanations given by the explanation facility revealed that the 'source orientation' characteristic that was set in the default description of the activity lead to a series of very strict locational requirements concerning the availability of the various raw materials required to make aluminum.

This is a good example of the difficulties associated with the development of accurate test procedures. Whether or not the requirements associated with a 'source orientation' can be considered erroneous or not, depends on the *characteristics of the activities* in the test set. It is very well possible that 'source orientation' is an objective of the decision makers that determine the decision-making behavior of the activity. In that case, again under the assumption that the test set is correct, the associated requirements must be wrong, because all the currently exploited sites are characterized unsuitable. If, on the other hand, the production plants do not have a source orientation, the wrong type of activity was matched. Rather than matching for a source-oriented production plant, a plant that is not source-oriented should have been matched.

An indication that this is indeed the case is Figure 6.5. Here it can be seen that none of the aluminum producing counties is a bauxite producer, although some of them have bauxite producers as their neighbors. The result of a matching for aluminum production without source orientation is displayed in Figure 6.6.

Switching to production without source orientation implies that a lack of local material resources can be compensated by the proper means of infrastructure, such as direct accessibility to the railway network. Compensation implies disjunctive elements and thus more possibilities to generate matches. The result is thus that many more locations can be considered suitable, albeit only from the raw material perspective.

The comparison with actual production sites turns out better now; three out of the four sites are in the set. Xiangning, however, lacks direct access to the railway network and is therefore not contained in the matching set.

The demands can further be released by assuming that the raw material can be transported by means of trucks over a relatively small distance before it is transported further by means of rail (again no cost functions for, in this case, transfer costs were incorporated). The demand of direct accessibility to the railway network is then released to indirect access, so that Xiangning

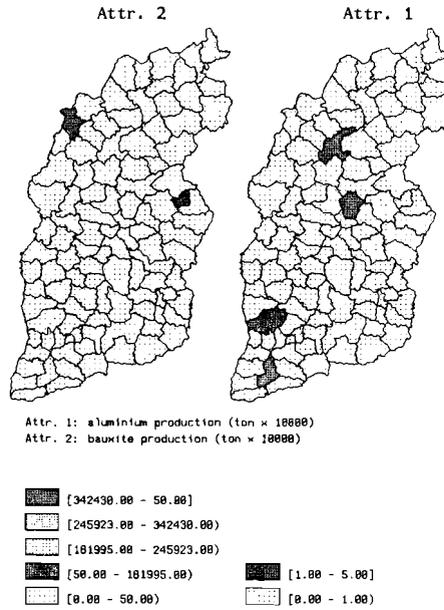


Figure 6.5: Aluminum production versus bauxite mining

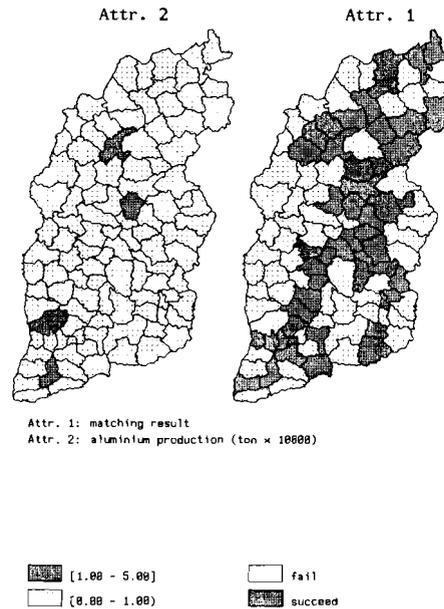


Figure 6.6: Full matching for aluminum production without source orientation

will become part of the set as well. The resultant map of the matchings is shown in Figure 6.7.

The next dimensions to be considered are those of coal and electricity. In a first version of the model these were treated as separate dimensions. Matching for coal revealed that Yuncheng city was not in the set because it did not have any coal. But after modifying the model so that coal and electricity became alternative, disjunctive possibilities for fulfilling the plant's energy needs, and after assuming that coal could be transported by rail as well, the results became better. Figure 6.8, for instance, shows the matching results for the dimensions 'raw material' and 'energy' together, and the actual production sites.

As will be obvious, however, there is a price to be paid for continuing relaxation of the spatial production requirements in the model. In the last solution presented here the number of matching sites becomes so large that virtually any possible site becomes part of the matching set. In the example considered only the conditions were relaxed, thereby replacing more and more conjunctive elements by disjunctive ones, thus providing the sites with even more opportunities to satisfy dimensional demands. Of course, in much more serious applications, modification rather than relaxation will be the most appropriate way of adjusting the rules of the rule base. This implies replacement of conjunctive terms with just different conjunctive terms, or the introduction of disjunctive terms that represent new conjunctive networks.

6.4 Matching with demands by locations

Another possibility of the REPLACE relational matching system which was interesting to explore in the Shanxi case study, was to see which spatial effects are generated by policy regulations concerning the location of specific kinds of activities. This topic relates to two questions: first, what is the spatial interpretation of these regulations or proposals for regulations, and second, how is the spatial outcome of these regulations related to the suitability patterns for the activities?

From a relational matching point of view, talking about restrictive policy regulations is talking about 'demands by objects'. The necessity of being able to include demands by locations into the relational matching model was discussed in detail in Section 3.6. It was argued that in order to be able to assess the possibilities for locating an activity from the perspective of the activity itself, locational demands are indeed important constraints on the

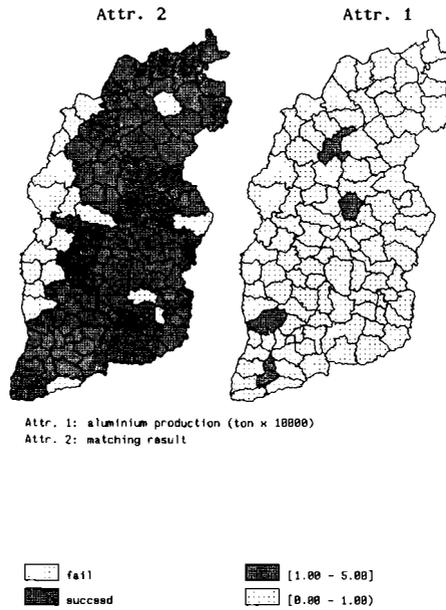


Figure 6.7: Matching with released transport requirements

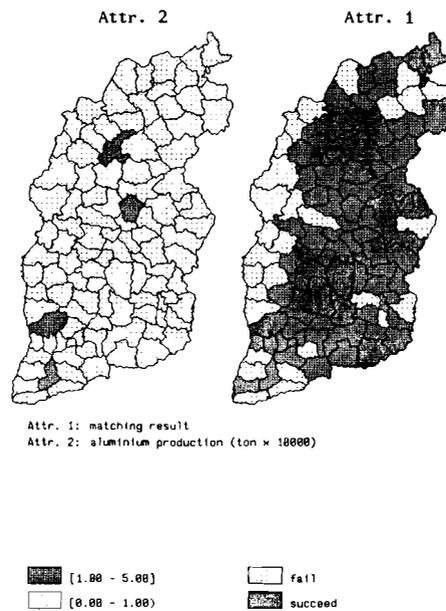


Figure 6.8: Matching results for raw materials and energy, and actual production sites

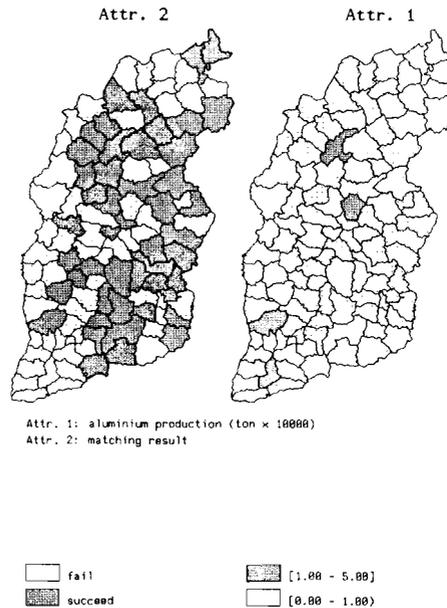


Figure 6.9: Matching after adding spatial policy and environmental regulations

production milieu. From the perspective of the governmental planner, on the other hand, it is useful to have an overview of which areas are basically suited for various activities. From these data, spatial conflict and competition maps can be constructed, which can again be objects for policy formulation.

Separate dimensions were mentioned as a way of incorporating object demands into the matching model. Combined with a routine by which the user can determine the dimensionality of the matching problem, this has the advantage that the user can decide whether or not he wants to include the object demands in the matching.

Figure 6.9 shows an example of adding a number of 'spatial policy' and 'environmental policy' regulations to the matching shown in Figure 6.8. As appears from a comparison of the maps, including these locational constraints in the model reduces the number of matches considerably.

The regulations contained in the 'environmental policy' dimension restrict the siting of new industries to zones that are not yet too heavily polluted. But since no data on pollution were available, this dimension did not have any effect on the matching at all (note that to prevent this unknown information from making all the locations fail the test on this dimension,

the global matching parameter for ‘unknown object scores’ had to be set to ‘succeed’). The reduction in matches is therefore completely the result of the regulations formulated in the ‘spatial policy’ dimension. The exclusion of both of the industrial production areas of Taiyuan city and Yuncheng city from the set shown in Figure 6.8 was the result of an excess energy consumption density in these regions. It must be stressed here once more that this kind of exclusion does not mean that these sites could not serve as production sites for the activities as such; it is just that they could not fulfill the function of a production site under the assumed policy regulations.

6.5 Modeling agglomeration economies by means of linkages

The study of agglomeration economies with its emphasis on spatial connections and linkages is a widely comprehended issue in industrial location theory (e.g., Conckling and Yeates, 1976; Lloyd and Dicken, 1977; Chapman and Walker, 1987). Many types of linkages (e.g., production, service and marketing linkages) and various types of agglomeration economies (e.g., localization economies and urbanization economies) have been specified and studied.

A rather functionally-oriented approach to linkages is presented by Lloyd and Dicken (1977). They define linkages as the functional relations between economic activities within a relatively restricted space:

“In the final analysis, of course, any firm is but one part of a complex chain of production held together by direct or indirect linkages between a series of firms. It is through such linkages that *external* economies are transmitted to the individual production unit through its network of interconnections with other elements in the system.” (Lloyd and Dicken, 1977; p. 288 [my italics]).

Of course there are many different types of linkages, but for location theory in general and relational modeling in particular, only those linkages imposing constraints on the possible choice of location are important. This brings us back to the restrictive principles by Rawstron discussed in Section 2.5.1., and via Rawstron, to the matching approach to locational choice set modeling.

From the perspective of the Shanxi Province case study, the relations between linkage networks and the site suitability patterns of specific types of

activities is something that might be worth studying. Strong local linkages tend to generate local 'growth poles' or concentrations of activities which mutually contribute to each other's development. In the context of the reorganization of a regional economy it can therefore be interesting to integrate the issue of linkages into the site suitability measurements.

From the perspective of relational modeling it is also interesting to have a somewhat closer look at linkages. Linkages concern relations between different activities, either different kinds of activities, or different activities of the same kind. A relational matching model as presented here, however, only provides a one-actor site suitability perspective on the geographical environment. A positive or negative outcome on site suitability is the result of the confrontation of the actor's requirements and the environment's properties. This implies that in the study of linkages by means of a relational matching model such as the one in REPLACE, linkage-requirements must be formulated as locational production requirements. Or put somewhat differently, in order to study the outcome of inter-industry relationships on the outcome of the site suitability pattern, these relationships or linkages must be translated into sets of locational properties. This then leads to formulations of linkages such as the following:

```
[
    v_get('annual production,Y,Prod),
    v_get('surrounding aluminum
production',X,Surround),
    Prod + Surround > 16 * Prod
]
```

So the more activity-oriented requirement that in order to process aluminum there have to be nearby producers of the resource, is here translated into a requirement in terms of the locational property 'surrounding aluminum production'.

To illustrate a possible application of the use of linkages, Figure 6.10 shows the matching results for the 'raw materials' dimension of aluminum processing. A linkage factor was included by formulating a requirement that sufficient aluminum for covering the needs of the production plant must be produced by the county and its neighbors. The result of such a requirement is that matching clusters of counties around aluminum producing counties are generated by the matching model. In the example in Figure 6.10, Taiyuan city and Yuncheng city drop out because they and their surrounding counties do not reach the required aluminum production threshold.

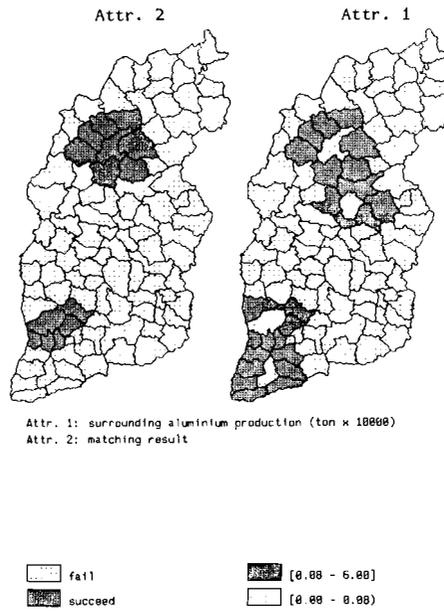


Figure 6.10: Matching for aluminum processing including linkages

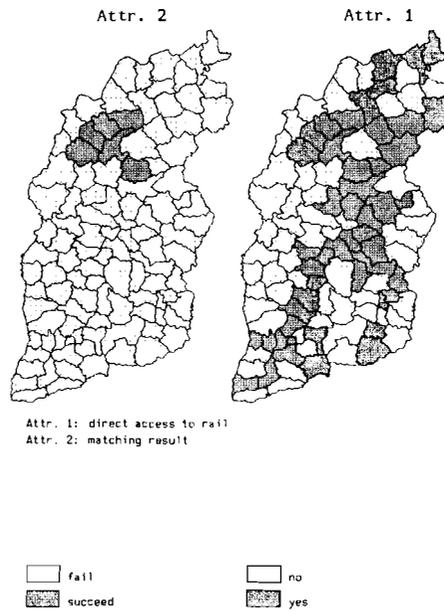


Figure 6.11: Matching for aluminum processing including linkages and compensating transport requirements

It is interesting to see how things change when transportation requirements are added. These specify that the surrounding aluminum production must not only be sufficient, but the producing counties must also be accessible from the processing location, for example by rail (Figure 6.11). Of course the specific requirements can again be made dependent on the specific characteristics of the aluminum processing plant.

Adding these requirements thus leads to an additional reduction in the set of counties. These are just simple illustrations of how linkages and spatial production dependencies between various types of production might be included into a relational matching model of production milieu. Although the examples are not very realistic, they do show that if interpreted and defined in a functional manner and therefore as constraints on the formation of the choice set, linkages can be incorporated in a relational model of site suitability.

6.6 Complex matching and ‘don’t care’ values: an example

Quite another example of what working with a system such as REPLACE may be like is an illustration of how complex, under certain circumstances, matching operations can become, and how ‘semantic’ aspects of the system become important. During the development of the REPLACE system it was decided to provide the user with a possibility to declare ‘don’t care’ values. This decision was based on the recognition that a user might want to conduct matchings on different levels of (generalized) abstraction. Various reasons for conducting matches on generalized types can be thought of. One of them might be that the information a user has on the relevant characteristics of an activity is less than presumed in the inference tree(s). Another might be that the user is interested in the differences between the matching results of various levels of generalization of the same generic activity type.

When experimenting with the DSS in general and REPLACE in particular, reasons for trying out abstractions cropped up frequently. Sometimes because it was interesting to see how much each of the differentiating actor characteristics was responsible for reducing the matching set of locations, sometimes because it was simply not yet known what the activity would really be like, and one was more interested in more general types than in specific instances of that type. ‘Don’t cares’ increase the disjunctive freedom of establishing matchings by providing the matching program with extra

paths in the inference trees. Undoing them removes these extra disjunctions. The result may be a complex process of matching, the more so since the setting of ‘don’t care’ values for inferential attributes does not always result in extra opportunities for finding a match. The dependency relations among attributes that are typical for a relational matching model may intervene with the changes made in the inference tree by the setting and undoing of ‘don’t care’ values.

To investigate some of these complexities, a number of hypothetical matching models were constructed. Figure 6.12 and Figure 6.13, for instance, show results of such a matching for a hypothetical ‘transport’ dimension with different combinations of pre-defined ‘don’t care’ values, for the dichotomous attributes ‘magnitude’ and ‘exporting’. How many extra possibilities for generating a match do these ‘don’t cares’ generate? According to the differences between the explanation in Figure 6.12 and Figure 6.13, that would depend; as is shown below. This can also be inferred from Figure 6.14 which shows the (hypothetical) inference tree for this example.

Figure 6.13 shows the matching results for the ‘transport’ dimension, with declared ‘don’t care’ values for the inferential variables ‘magnitude’ (large_scale/small_scale) and ‘exporting’ (yes/no). Given the inference tree in Figure 6.14, however, these ‘don’t cares’ do not give rise to four (two times two), but only to three alternative routes. The reason is that once the ‘magnitude(small scale)’ path is chosen, the attribute ‘exporting’ is not an inferential variable any longer. Declaring ‘exporting’ a ‘don’t care’ value therefore, does not generate additional matching opportunities once the ‘magnitude(small scale)’ path has been selected.

Figure 6.12, on the other hand, shows the matching result with only ‘exporting’ declared a ‘don’t care’ value. It will be clear that the difference of the number of matches between Figure 6.12 and Figure 6.13 must thus be attributed solely to the magnitude (small scale) path.

That this form of complexity cannot be solved by application of the technique of dynamic editing of inferential variables (refer to Section 5.7.6.1 for an explanation) becomes clear by taking another look at the inference tree in Figure 6.14. Dynamic editing precludes the setting of a ‘don’t care’ value for ‘exporting’ once the value for magnitude equals ‘small scale’, because in that case ‘exporting’ is not an inferential variable any more. In case the value for magnitude is ‘large scale’, however, ‘exporting’ is an inferential attribute and a ‘don’t care’ can be declared (Figure 6.12).

But what are the consequences for an eventual ‘don’t care’ value for ‘exporting’ if a ‘don’t care’ value is declared for ‘magnitude’? Clearly, this

山西总体发展研究专家系统 RELATIONAL MATCHING MODEL IIASA

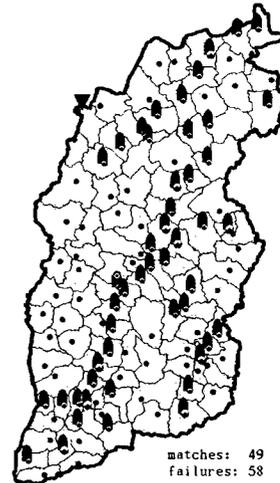
Failing matching requirements for hequ

Dimension: transport
magnitude(large scale) exporting(no)

REQUIREMENTS: direct access to rail for hequ must be yes (FAILED)

Dimension: transport
magnitude(large scale) exporting(yes)

REQUIREMENTS: direct access to rail for hequ must be yes (FAILED)
AND direct road abroad for hequ must be yes (FAILED)



matches: 49
failures: 58

Hit any mouse button to return to upper level...

Figure 6.12: Matching results with 'don't care' (A)

山西总体发展研究专家系统 RELATIONAL MATCHING MODEL IIASA

Failing matching requirements for hequ

Dimension: transport
magnitude(small scale)

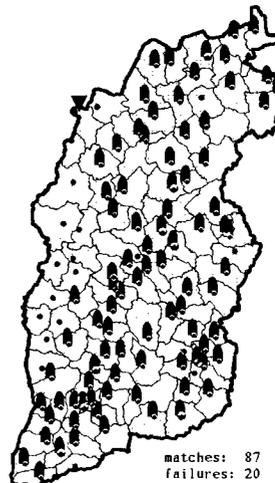
REQUIREMENTS: access to rail for hequ must be yes (FAILED)

Dimension: transport
magnitude(large scale) exporting(no)

REQUIREMENTS: direct access to rail for hequ must be yes (FAILED)

Dimension: transport
magnitude(large scale) exporting(yes)

REQUIREMENTS: direct access to rail for hequ must be yes (FAILED)
AND direct road abroad for hequ must be yes (FAILED)



matches: 87
failures: 20

Hit any mouse button to return to upper level...

Figure 6.13: Matching results with 'don't care' (B)

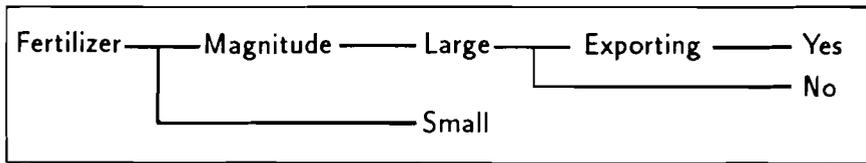


Figure 6.14: Hypothetical inference tree

implies that both the ‘small scale’ and the ‘large scale’ paths can be taken. But although for the ‘large scale’ alternative, a ‘don’t care’ for ‘exporting’ is valid, for the ‘small scale’ alternative it is not. Dynamic editing cannot solve this problem. Setting the ‘don’t care’ value for exporting while ‘magnitude’ has a ‘don’t care’ as well, however, does not harm the process either. The ‘exporting’ attribute is simply not used in case the ‘small scale’ path for ‘magnitude’ is followed.

Nevertheless, the resultant matching process may become complex to interpret, even though the path followed through the inference tree is displayed in the explanation. The conclusion could be that generalized abstraction by ‘don’t care’ values might, under certain conditions, lead to complex matching processes. This might be a disadvantage because things become more difficult to interpret and evaluate. On the other hand, it does show that the relational matching model as presented here can handle complex generalization and abstraction. It also shows that relational modeling does not necessarily result in a modeling-per-individual. Generalization can be incorporated in a relational model, provided that the generalizations conform to the rules of relational definition and modeling as well.

6.7 Implementing a two-step choice process: PDA, PDAS and REPLACE

Certainly at least as important as the availability of a stand-alone site suitability assessment model are the possibilities for integrating such a model with other models in the system. Currently, linkages with three or four other models are conceivable. First, there is the MITSIM water management model (Strzepek and Fedra, 1988), and the ISC atmospheric pollution model (Posch, 1988). Relations in either direction between these two models and REPLACE are possible. The most obvious way to go about it would be to provide direct linkages of the kind where both MITSIM and ISC provide

REPLACE's data bases with data about site-specific water availability or pollution levels. Both kinds of data can be important in the site suitability matching process. Water, because it can be an important means of production, atmospheric pollution, because of its importance for environmental policy; it may be a determining factor in whether or not location permits are granted. Interrelationships can also take the opposite direction. The results of REPLACE may be used for suggesting a specific choice of location, which again may have consequences for water management and atmospheric pollution. These kinds of interrelationships are of the type where various models operate on a common data base, a type of interaction that was already mentioned in Section 5.5.

The other type of interaction mentioned there was the one in which various models form a kind of analysis-chain, a sort of step-wise method for finding a solution to a specific problem. Within the Shanxi Province DSS, such an interaction between the PDA, REPLACE and PDAS models appears to be quite promising because it would constitute a good example of the two-stage spatial choice modeling suggested in Section 1.3.

The PDA (Production Distribution Area) and PDAS (Production Distribution Area Spatial) concepts were developed by a group of Polish researchers, as an integral part of an Industrial Development Strategy (IDS). The idea of IDS is to change the structure of production (product range, production volumes, etc.) over time and space by means of investments, in such a way that an optimal structure emerges. The PDA model (Dobrowolski *et al.*, 1984) generates such an optimal industrial structure, given an initial situation, a pre-defined optimal goal state, and a set of constraints, the optimal being defined as, for instance, maximum energy and material resource efficiency. The properties of the process require it to be treated dynamically over a time span of 10 to 15 years. The PDA model generates a plausible target structure which can be reached starting from the initial situation.

The PDA model is essentially non-spatial. It constitutes an optimization of the structure of an industrial complex on the basis of energy and material resources, but it does not take into account the allocation of space associated with such an industrial structure. For this reason the PDAS approach was developed; a linear programming optimization model which takes spatial allocation into account by means of integrating transportation costs and local constraints (Zebrowski *et al.*, 1988). The result of the model is an optimal spatial distribution of the PDA optimal industrial production structure (color Plate 8).

In the concept of spatial choice in general, and locational decision making in particular, a model such as PDAS typically belongs to the second stage of spatial choice modeling. Optimization, selecting the best site(s), requires a set of suitable, functionally equivalent sites. Different technologies and production characteristics require different types of sites. This reconstruction of a set of production-specific feasible sites belongs to the first stage of spatial choice modeling, the reconstruction of the choice set. A relational site suitability matching system such as REPLACE can be used to do this. The results of this modeling step are then forwarded to an optimization model such as PDAS, which then computes an optimal case within the limits of the choice set determined by REPLACE. Needless to say, the set of technologies and production volumes for which REPLACE would have to reconstruct production milieus can be generated by the PDA model. The PDA → REPLACE → PDAS chain thus forms another example of how REPLACE could be linked with various models in the Shanxi Province DSS.

6.8 Implementing a two-step choice process: REPLACE-DISCRET

Yet another way of implementing a version of a two-stage decision model is the use of REPLACE in combination with a multi-criteria evaluation technique for the second stage. In the context of the Shanxi Province DSS this second modeling step cannot be ignored. It is simply insufficient to reconstruct choice sets that consist of more than say, three or four alternatives. Decision makers must make discrete choices, and would therefore be helped with one or more options for optimizing a discrete choice from the resultant choice set. Using REPLACE as a kind of pre-processor for PDAS is one possibility for achieving this. The problem, however, is that the use of PDAS implies the use of a set of very specific optimization criteria which have to do with the optimal use of energy, material resources and transport costs. A set, therefore, which does not necessarily represent the criteria of the decision maker.

Another, and perhaps more flexible, way of determining a set of optimization criteria concerns the use of a general multi-criteria evaluation model. The idea is then that REPLACE can be used to reconstruct a reduced set of functionally equivalent, suitable alternatives, whereas the remaining freedom of choice in this set can be used to search for an optimal solution, given a pre-defined set of evaluation criteria.

Generally such a model contains a set of criteria on which the alternatives score, as well as combination rules specifying how the individual scores must be combined into an overall score. The additive combination rule in Section 1.2.1.1 is one example of such a model. Many different ways of combining criteria scores into an overall score exist. Similarly there are many ways of ordering alternatives without first calculating alternative-specific overall scores. Here one can think of the application of 'dominance principles' as well as of many different kinds of methods for handling alternative ordering with discrete data (for an overview of the various methods refer to, for example, van Delft and Nijkamp, 1977; Wierzbicki, 1979,1980; Nijkamp and Spronk, 1981; Grauer *et al.*, 1982). It is important here that, to a certain extent, these models enable the user to define his own criteria and combination functions, thus generating more freedom and flexibility in defining the criteria to be used in the optimization stage. It was therefore decided to develop a prototype combination of REPLACE with the DISCRET model, a member of the DIDASS family of multi-criteria evaluation models developed at IIASA (Grauer *et al.*, 1984).

6.8.1 Exporting attributes to the evaluation model

In order to establish a meaningful connection between REPLACE and a multi-criteria evaluation model, both the user and the evaluation model must be provided with a set of meaningful evaluation criteria. In Section 1.3 it was mentioned that for the optimization in the second stage of the choice modeling, inter-alternative compatibility of the variables is required. A model such as a weighted additive combination rule, for instance, pre-supposes such inter-alternative compatibility. In other words, optimization of the choice set requires common dimensions, attributes on which each of the objects can have a score which can only vary in a quantitative manner, not qualitatively.

What does this imply for the information on alternatives in the choice set that can be used in the multi-criteria evaluation? It simply means that no information that was used during the matching can again be used for the optimization. In the matching model, locational variables are only used within very specific contexts; specific combinations of actor attributes leading to specific sets of requirements. Using the same information, the same variables again in the optimization, would imply a sudden abstraction from these specific contexts, something for which no good reason seems to exist. Note that this applies not just to those variables which were actively

engaged in the matching. In case a requirement can be fulfilled in more than one way, and the first attempt is successful, the variables associated with the other possibilities are not used at all. Nevertheless, they cannot be used for optimization any more, simply because the matching model declares them as having a different meaning in different contexts.

However, if information contained in the matching model cannot be used, then what kind of information can be used for optimization? Three sources of information come to mind. First of all, there is the information which is empirical, but which is neither mentioned nor used in the matching model. A second type of information which may be used for optimization concerns what could be called 'excess or surplus information'. The idea of not using matching information, can be rephrased by saying that all information which is not declared or used in the matching is valid information for optimizing. An interesting implication of this formulation is that if a specific requirement is formulated as say, 'the available amount of electric energy must be higher than 1000 units', surplus (or deficit) in electricity (the total amount available minus the required amount), can be considered unused information and can therefore be considered valid information for the optimization. But although this sounds tempting, there is a snake in the grass, which reduces the possible contribution of this kind of information for optimization considerably. Since the requirements such as electricity demands are contained in disjunctive matching rules, it is very well possible that for some alternatives the electricity surplus/deficit is used as a matching requirement, whereas for another object it is not. In such a case, 'electricity surplus' cannot be considered a common dimension of the two alternatives, unless it can be shown that for the alternative that was matched on another matching rule, the rule containing the electricity problem also leads to a match. Needless to say, the likelihood of all alternatives in the choice set to match this particular rule is rather small. Moreover, implementation of the use of this second type of information will be a tough job. It was therefore decided to discard 'surplus/deficit information' for the moment. A third and most interesting type of information for optimization consists of a set of newly derived or 'meta-variables', on which the various elements of the choice set can be compared. Here one can think of several possibilities. One of them is the use of variables from other models in the Shanxi Province DSS; e.g., transport costs, pollution values, energy efficiency measures, and so on. Another type of variable relates to the matching process itself. Here one can think of, for instance, stability measures expressing the robustness of the individual matching solutions. For each alternative in the choice set such a measure

can be computed. It can also be considered a common dimension and can thus be used for optimization.

6.8.1.1 A measure of the robustness of matching solutions

The robustness of a match consists of two elements; surplus value, and the number of 'fallbacks'. Surplus value occurs when an alternative satisfies a numeric requirement. The 'ease' by which it passes the test, the magnitude of the surplus therefore, contributes to the robustness or stability of the match. The larger the surplus, the higher the robustness. Surplus, however, only applies to numeric requirements. In order to deal with qualitative, categorical requirements, a second element of robustness can be introduced; the number of 'fallbacks'. The matching rules associated with the relational model usually are (partially) disjunct. This means that a matching rule can often be satisfied in various ways. Although only one successful one is sufficient for satisfying the entire rule, one can consider additional ways which would also have led to success, i.e., backing up the success of the alternative, thus increasing the robustness of the match.

Starting with the latter element, and forgetting about the surplus for the moment, the number of successful disjunctive terms in the matching rule can be considered an index of robustness based on fallbacks. Computation of this index happens by traversing the matching rule in a bottom-up manner. Each successful disjunctive term has a value 1.0. Disjunctions are added. In case a conjunction is encountered, the lowest value is taken. This represents the idea that a match is as robust as its least robust conjunctive term.

Next, the 'fallback' index must be combined with robustness information derived from surpluses. Here it was chosen to use surplus value as a weight for each of the successful disjunctive terms in the matching rule. As a measure for these surplus weights, one can take the ratio of the locational value and the required value in case the rule contains a 'greater than' operator, and the reversed ratio in case of a 'smaller than' operator. In order to correct for the effect that very large differences increase the robustness in a disproportionate manner, the weights had to be computed in such a way that they follow something like a 'decreasing marginal utility' rule; the growth of the weight decreases if the surplus increases. Of course various weighting functions that conform to such a rule can be thought of (logarithmic, inverse power, etc.). However, since it was realized that this first version of a robustness index was a rather crude one anyway, it was decided to go the easy way, and a simple maximum weight-value was set.

Applying this technique to each of the dimensional sub-problems, as many robustness measures as dimensions are computed. Dimensions, however, are combined in a conjunctive manner, thus making the general robustness of the match the minimum of each of the dimension-specific robustnesses.

Of course, the method described above is not without problems and limitations. An important issue that needs further study is that the rule base must be 'realistic'. By this it is meant that if the rule base includes rules which are always evaluated 'true' or 'false' the robustness measure becomes perturbed. Suppose, for instance, a dimensional rule of the form: '(A and (B or C))'. In case A is a requirement which is always satisfied, the overall robustness is determined by the top-level conjunction, whereas in fact it is the disjunction 'B or C' that determines the actual robustness.

Another, more serious, problem associated with this method of robustness determination is that it requires optimal rule bases. In Section 3.5.2 it was mentioned that one of the disadvantages of the current representation of inference trees is that only complete requirement sets are associated with completely described actor types. This implies that the inference tree can contain rules such as '((A and B) or (A and C))'. Clearly this is a non-optimal rule which can be reformulated more efficiently as '(A and (B or C))'. If the robustness is computed on the non-optimal rule, a higher value will be the result, since a disjunction rather than a conjunction is at the top of the tree. In order to bypass this problem without, at this stage, having to redesign the complete structure of the inference trees, a kind of pre-processor was developed. This program takes in a non-optimized matching rule, optimizes it, and computes the associated robustness. This is of course a 'quick and dirty' solution. Rather than using this kind of palliative solution, the structure of the inference tree should be revised so that this problem does not occur any more. But again, deadlines and other organizational issues made this impossible.

6.9 Conclusion: Relational complexity and relational model building

On reading through the examples presented here, one easily gets the impression that they are not very realistic: they are not. Although they were presented as 'generic' illustrations of how several aspects of (integrated) location theory can be represented in a relational matching framework, it must

be said that these examples come from the initial rule bases of REPLACE as it was implemented in the Shanxi Province DSS. Very quickly after the first and very simple versions of inference trees for 'real' production activities were constructed we were confronted with a sometimes overwhelming complexity and intricacy, making the examples appear very naive and simplistic. However, what was attempted was to show by means of the examples that various aspects of industrial locational behavior and industrial location theory can in principle be treated in a relational manner. The examples represent mere starting points; initial attempts at building relational rule bases, based more on commonsense and methodological thinking than on solid economic-geographical investigation. In order to become really applicable in actual decision support, they would need careful inspection and evaluation, but most of all they need to be significantly extended and improved.

Chapter 7

Conclusions and Discussion

ABSTRACT

This chapter contains an evaluation of some of the issues dealt with in the previous chapters. An attempt has been made to formulate some general conclusions from the material and results discussed earlier and to formulate the potential for future developments.

7.1 Relational modeling in a two-step process of choice modeling

The objective of relational modeling is to reconstruct sets of functionally equivalent choice alternatives. In the context of empirical research, one can think of two types of applications; those aimed at a better understanding of spatial choice behavior, and those meant to support decision making. Although the two-stage process advocated here might perhaps be a useful one for the former type of applications, it certainly seems a useful approach for the latter.

As pointed out in Chapter 1 and as mentioned by van der Smagt and Lucardie (1990; p. 7), relational reconstruction of functionally equivalent alternatives considers a process of logical selection rather than ranking or optimizing. Whether or not a second, optimizing modeling step is necessary is something which is certainly open to debate. According to van der Smagt and Lucardie, reconstruction by selection can do the entire job; no ranking or optimization in a second step is necessary. To a certain extent this is indeed

true. If one keeps adding requirements associated with increasing numbers of dimensions or objectives, the choice set will be reduced to a set containing only one or no alternatives. Hence, no optimization is necessary any more. In other words, proper relational reconstruction will always reduce the choice set to such an amount that the final choice of a discrete alternative will be a more or less trivial matter. Nevertheless I doubt that such a procedure is really feasible, especially when considering empirical applications and decision support. The basic problem here is that there seems to be no way of knowing beforehand which, and how many, objectives must be added in order to reduce the choice set to just one alternative. Similarly, in case the matching yields a set of alternatives larger than one, the system can hardly demand the adding of objectives, just because a set of only one alternative is required. In Section 1.2.1.2 it was suggested that one of the reasons that people and organizations pay consultants for advice is that they do not know their own situation well enough to determine what they want. In the context of decision support this implies that a decision support system can be used not only to figure out specific aspects of a particular decision, but also for determining objectives. This again implies that one cannot expect a user to be able to determine a very specific and unequivocally defined set of objectives. However, what he might be able to do is to define different objectives and perhaps even different sets of objectives, each of which will generate different sets of choice sets. Sometimes such a choice set will be very small, sometimes it will be large. Large choice sets imply a large freedom of choice, or in case of site suitability or production milieu, a high degree of footlooseness as defined in Section 2.5.2.2. The use of abstractions by means of 'don't care' values also implies a potential increase in the number of alternatives in the choice set. One of the reasons for including such a thing as abstraction in the system was that users are not always able to define their objectives in terms of what the system offers them, or are simply indifferent to specific distinctions between actor-types. All this means that we might end up with choice sets that contain so many elements that a second, discrete optimization can be very helpful. The examples of linking REPLACE with PDAS or DISCRET presented earlier furthermore indicate that by providing such a second modeling stage, merits or 'knowledge' from other fields of research can be combined and utilized within a decision support context.

7.2 Relational modeling by means of inference trees

In Chapters 1 and 2 it was explained that relational modeling not only departs from a perspective on spatial choice that is different from a mere behavioral or cognitive one, it also implies a different kind of modeling. In Chapter 2 it was argued that although the thinking about concepts such as site suitability and production milieu has evolved in a more functional direction, many 'translations' of this thinking into modeling turned out to be unfortunate, simply because theory and modeling did not really 'match'. It was argued that instead of (statistical) modeling based on empirical similarity, a functional approach requires a more heuristic or algorithmic modeling based on functional equivalence.

In Chapter 3 relational inference trees were suggested as a means for establishing such modeling. Inference trees contain routes representing actor-types as actor-attribute-value combinations, and matching-rules associated with each of these routes. In Chapters 3 and 4 it was furthermore argued that these inference trees offer ample opportunities to represent the various aspects of relational modeling such as the re-categorization of variables as a result of a score on one or more other variables, or the conditional relevance of entire variables. Together with a logical programming language such as Prolog, they also provide an attractive representational scheme, not only from a declarative, but also from a procedural point of view. Another attractive aspect of inference trees is that they have a lot in common with a classification scheme known as a decision table. It was shown that such a scheme offers ample opportunities for checking formal adequacy and for inferring its 'essential' content.

The specific form of the inference trees presented here, however, is not entirely unproblematic. Especially on a conceptual level, two problems predominate; a generalization problem and a modeling problem. The generalization problem concerns the question of how to make decisions on the construction of various actor-types. The problem, in other words, of how to build a formal structure representing functional equivalence, without stepping back into the trap of fixed typologies based on empirical similarity. In this project it was decided not to worry too much and to comply with some of the major types of activities recognized as useful and important ones for the Shanxi Province case study. However, the problem is definitely a serious one and needs further investigation.

The same applies to another unfortunate situation, for in the current version of the inference tree it is impossible to associate requirements with only a subset of the actor attributes contained in a specific route through the tree. In Section 3.5.2 this problem was referred to as the 'modeling problem'. This causes serious problems for an iterative process of model development, for modifying and revising an already existing model, and for computing the robustness of matches. Solving this problem is of great importance for successful application of relational modeling by means of inference trees. Although I expect that the problem can be solved within the framework of inference trees, a somewhat different representation will nevertheless be necessary.

7.3 Matching with dimensions

A matching problem can often be partitioned into a set of so-called 'dimensions'; more or less 'independent' aspects of the matching problem which seem to coincide nicely with the 'natural', although abstract, aspects of, in this case, production milieu. It was argued that this independent status of dimensions is not in contradiction to a relational approach, since this independence only applies to the abstract level of dimensions. As with the relational definition of concepts, dimensions are modeled by means of relational inference trees, or consist of conjunctions of sub-dimensions. Dimensions are 'internally related' in the sense that processing of a dimension can limit the search space for other dimensions. Internal relations such that categorizations vary with different scores on certain variables occurs only inside the inference trees and matching rules referred to by dimensions.

Introduction of a dimensional structure in a relational model, however, induces a somewhat perturbing problem. While working with a dimensional model and while conducting various tests and applications, it turned out that dimensional structure was something that indeed concurred nicely with many questions users would like to ask a site suitability matching system. Of course, since the system was set up in a dimensional fashion, it is difficult to guess what people would have thought of a system that would not work with dimensions. Nevertheless, during many sessions with both visitors and IIASA scientists, it turned out that dimensions are attractive means, not only to explain relational analysis with, but also for studying a problem such as site suitability. Such a problem is so complicated that partitioning it

into a set of broad, rather vague and imprecise aspects is an almost natural thing to do.

However, as was discussed in Section 3.2.5, working with dimensions forces one to take decisions on how to allocate specific aspects of the problem to specific dimensions. Therefore, although partitioning a complex problem into dimensions is easily done at the abstract level of concepts that need further relational definition, defining them relationally brings one right back to the question of how best to partition.

In conclusion, one can say that working in dimensions is attractive. Therefore, I would like to argue for further investigations concerning the question of how relational models can be kept dimensional, while overcoming the problem of how to allocate specific issues or sub-dimensions to the higher-order dimensions.

7.4 Relational modeling: explanation versus decision support

Chapters 3, 4, and 5 discussed the development of a set of tools by means of which relational analysis can be carried out. These tools were put together in the REPLACE system by means of application of a specific method for building integrated, model-based software; a method advocated by IIASA-ACA. In Chapter 6 some general possibilities for using the system were presented.

In Chapter 1 it was suggested that relational analysis could serve two, more general types of goals; the development of explanatory models of spatial decision making and choice behavior, and the use of relational models in decision support. To what extent can REPLACE contribute to each of these fields? This is of course a very important, and even necessary, question to raise. However, it is also a tricky one. First of all, when looking back at one's own product and onto all the work associated with it, it becomes clear that although the question is easily stated, an intelligible but concise answer is very difficult to give. As was pointed out above and in earlier chapters, the system and core-model can only be considered to be a 'try-out'. It constitutes some first attempts at building a formal model structure into which relational definitions of concepts can be fitted and at developing a prototype system which can handle both this formal structure and a large series of requests submitted by a user. A system such as REPLACE does seem to

offer opportunities for developing explanatory models of choice behavior and to be engaged in decision support.

To start with the latter, the system presented in Chapter 5 clearly shows that if the relational models contained in REPLACE are adequate, the system provides many opportunities to 'play' with site suitability. The various options offered to the user make it possible to carry out many matchings in a relatively short period of time. Each of these matchings is the result of a different setup representing different actor objectives or properties and/or a different environment. In combination with the explanation routines and the GIS part, the user might get an inclination of 'how things look' for some of the options he has in mind. By means of, for instance, the use of models such as PDAS or DISCRET, the user can obtain some idea of how choice sets are structured internally after optimization.

A good reason for building DSSs is that problems, even when structured, can remain complex and even ill-defined. However, in these cases the best model of the actual decision-making behavior is formed by the decision maker himself. If things can be worked out in such a way that a DSS can show him what the structural aspects of the problem imply in the context of the complexity of the problem, quite a bit would have been achieved. Regarding the potential for developing explanatory models of spatial choice behavior, however, things are somewhat different. Decision support requires more or less 'true' or 'unproblematic' model content. In case the objective is the development of a model with a high(er) degree of explanatory power, however, the models formulated obtain a much more hypothetical character. Validation of the hypothesis requires empirical testing. However, it has been noted that developing adequate test sets for testing reconstructed sets of functionally equivalent alternatives is difficult. As in the development of knowledge-based systems, in general it seems that validation must be conducted during the model building, i.e., during the various rounds of the knowledge-acquisition process. I think that apart from resolving the problems mentioned in some of the earlier conclusions, working out a far more structured methodology for acquiring the empirical content of relational matching models deserves high priority, and needs serious investigation in the near future. Only if these conditions are satisfied can a relational approach to spatial choice modeling increase explanatory power.

7.5 Building real-world applications: accepting complexity

At the end of Chapter 6, a few remarks concerning the simplicity of the example applications dealt with in that chapter were made. It was stated that once confronted with real world cases, relational modeling becomes a very complex exercise.

As argued in the previous section the difficulties in developing meaningful empirical applications of relational site suitability assessment do not stem from fundamental weaknesses in the relational methodology as such. Problems originate from two other sources: deficiencies in the modeling procedure, and the way the complexity of reality becomes manifest in a relational approach.

To start with the latter, a relational approach in modeling choice behavior in general, and production milieu in particular, confronts the researcher with the full complexity of the empirical world in a much stronger way than any statistical or mathematical approach could ever do.

Whereas, for example, many inductive approaches mold and transform the definition and classification problem into a rather simplistic statistical representation based on empirical similarities, the relational method simply forces the researcher to face the complexity of the empirical problem at hand. The researcher cannot get away with inappropriate abstractions and generalizations, because the relational approach compels him to incorporate explicitly that complexity into the model. It is indeed striking how clear the complexity of meaningful site suitability assessments becomes when tackling the problem in a relational manner. Even when inference trees start branching quite a few times, the resultant decision and matching rules often look simplistic and unrealistic. The reason is that formulating the meaning of a concept such as site suitability in a branching network of inference trees makes the huge amount of simplifying assumptions that are implicit and often deeply hidden in, for instance, a statistical approach, so clearly apparent. It seems that the more complex and difficult the problem at hand is to model in a relational manner, the higher the degree of accuracy and adequacy the resulting model will have. This, however, might actually have more implications than it seems to have at first sight. Relational analysis as a functional approach toward concept definition and classification was developed as a reaction against those procedures which explicitly or implicitly treat problems of classification and generalized abstraction as a search for

empirical similarities. These similarities can be found in a rather straightforward manner by the application of statistical techniques. But empirical similarity as such is not very informative for the explanation of, say, choice behavior and decision making. Instead, it is the functional relation between empirical characteristics of alternatives and the properties of actors that determine which empirical characteristics and combinations of characteristics are important. The similarity only becomes important in assessing the extension of the concept; the set of objects that can be considered functionally equivalent under the given actor and goal description.

This issue of how valuable different approaches in modeling are for understanding somewhat better what goes on in the world around us also has some very interesting relations to the question of the predictive power of models. One can argue that heuristic and algorithmic modeling are inherently deterministic. As a result, there is no room for uncertainty in the meaning of probability. Therefore predictions on the basis of such models are good or bad, and it is likely that they are more often bad than good. However, how attractive is the alternative? One could say that the probability distributions associated with statistical models contain an assessment of the uncertainty of the adequacy of the structural relationships contained in these models. This implies that the predictions by statistical models are always 'good to a certain extent'. Of course, since averaging is at the heart of every statistical model, the assertion that the *a priori* probability that a specific statistical model will, on average, be a better predictor than a deterministic one, is reasonable. The choice between a relational or a statistical approach is therefore determined by what we want to have predicted. If it is patterns and processes that are of a typically aggregate character, for instance inter-regional migration streams, it may well be that statistical associations based on empirical similarity offer attractive opportunities. If, however, there is interest in what a single empirical distribution of regional variables will imply for various rather different types of economic activities, and if one also wants to influence the allocation of that regional space so that it conforms better to a pre-defined goal state, then an actor-specific, functionally-oriented model such as a relational matching model might be an attractive tool.

Recently, interesting new insights concerning the problem of deterministic relations versus statistical behavior of systems have been put forward in the theory of 'complex dynamical systems', also known as chaos theory (see Gleick (1988) or Crutchfield *et al.* (1986) for excellent introductory overviews of the field). One of the most important issues this theory has

to say something about is that even very simple, deterministic systems may exhibit relative chaotic behavior. Relative, because within this chaos very specific forms of regularity (attractors) can be found. This regularity itself is statistical rather than deterministic. However, the attractors can represent specific causal characteristics of the system that may help find possibilities to change the system's behavior in a desired direction. The insights in this new body of theory, at least on the conceptual level, offer a perspective on systems dynamics, into which both determinism and probability, both non-compensatory rigidity and statistical averaging can be fit together, to generate a new and fresh view on prediction.

Nevertheless, the introduction of a functional approach toward concept definition constitutes an attempt at achieving more valid models of choice behavior, at the expense of much higher complexity. The question, therefore, is whether one wants to pay that price. I think one has to. On account of the complexity of the relational nature of concepts, the associated model building will be difficult and time consuming. But these difficulties unmistakably indicate the complexity of choice and decision behavior, the intricacy of goal-driven behavior constrained by the availability of limited means. If it is these that are to be modeled as realistically as possible, I think that one simply has no choice. One either sticks with simple but, as I see it, invalid views on the nature of concepts and their measurement, and thus with models the validity of which is clearly in question, or one accepts the complexity of choice behavior as the starting point, and tries to model it accordingly, even if that requires a considerable increase in work. In return, it offers a perspective on better models to be applied in (spatial) decision making.

References

- Adams, D., 1979, *The Hitch Hiker's Guide to the Galaxy*, Pan Books, London.
- Adewolde Osunade, M.A., 1988, Soil Suitability Classification by Small Farmers, *Professional Geographer* **40**: 194–201.
- Alvstam, C.G., Ellegard, K., 1989, Volvo, the Organization of Work, a Determinant of the Future Location of Manufacturing Enterprises, in: Smidt, M. de, Wever, E., eds., *The Corporate Firm in a Changing World Economy, Case Studies in the Geography of Enterprise*, Routledge, London.
- Anderson, N.H., 1974, Information Integration Theory: A Brief Survey, in: Krantz, D.H., Atkinson, R.C., Luce, R.C., Suppes, P., eds., *Contemporary Developments in Mathematical Psychology 2*: 236–305, W.H. Freeman, San Francisco.
- Arbab, B., Michie, D., 1985, Generating Rules from Examples in Prolog, in: Haynes, J.E., Michie, D., Richards, J., *Machine Intelligence 11*, Horwood, Chicester, Halsted, New York.
- Arbab, B., Michie, D., 1987, Generating Rules from Examples, in: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, August 18-23, Los Angeles, CA, pp. 631–633.
- Bettman, J.R., 1970, Information Processing Models of Consumer Behaviour, *Journal of Marketing Research* **7**: 370–376.
- Biehl, D., Hussman, E., Schnyder, S., 1974, Bestimmungsgründe des Regionalen Entwicklungspotentials–Infrastruktur, Wirtschaftsstruktur und Agglomeration, *Die Weltwirtschaft*, Heft 1.
- Blinkert, B., 1978, Methodische Realitätskonstruktionen oder soziale Tatbestände?, *Soziale Welt* **29**: 358–369.
- Boose, J.H., Bradshaw, J.M., 1987, Expertise Transfer and Complex Problems. Using AQUINAS as a Knowledge Acquisition Workbench for Knowledge Based Systems, *International Journal of Man–Machine Studies* **26**: 3–28.
- Bosman, A., 1983, Decision Support Systems. Problem Processing and Coordination, in: Sol, H.G., ed., *Processes and Tools for Decision Support*, North Holland, pp. 79–92.

- Bouma, L.G., 1987, Formele Achtergronden van Prolog, *Informatie* 29: 316–327.
- Brand, M., ed., 1976, *The Nature of Causation*, University of Illinois Press, Urbana, Chicago, IL.
- Bratko, I., 1986, *Prolog Programming for Artificial Intelligence*, Addison Wesley, Workingham, England.
- Brennan, J.J., Elam, J., 1986, Enhanced Capabilities for Model-based Decision Support Systems, in: Sprague, R., Watson, H.J., eds., 1986, *Decision Support Systems. Putting Theory into Praxis*, Prentice Hall, Englewood Cliffs, NJ., pp. 130–137.
- Brown, L.A., Moore, E.G., 1970, The Intra-Urban Migration Process, a Perspective, *Geografiska Annaler* 52b: 1–13.
- Brown, L.A., Moore, E.G., 1971, The Intra-Urban Migration Process, a Perspective, in: Bourne, L.S., ed., *The Internal Structure of the City*, Oxford University Press, New York, pp. 200–209.
- Burrough, P., 1986, *Principles of Geographical Information Systems for Land Resource Assessment*, Clarendon, Oxford.
- Calzonetti, F.J., Eckert, M.S., 1981, Finding a Place for Energy, Siting Coal Conversion Facilities, *Resource Publication in Geography*, Association of American Geographers, Washington, DC.
- Carr, M., 1983, A Contribution to the Review and Critique of Behavioral Industrial Location Theory, *Progress in Human Geography* 7: 386–401.
- Chapman, K., 1974, Corporate Systems in the United Kingdom Petrochemical Industry, *Annals of the Association of American Geographers* 64: 126–137.
- Chapman, K., Walker, D., 1987, *Industrial Location, Principles and Policies*, Blackwell, Oxford.
- Church, R.L., Bell, T.L., 1981, A Comparison of Two Baseline Screening Approaches to Regional Energy Facility Siting, *Geojournal* supplementary issue nr. 3: 17–36.
- Clocksins, W.F., Mellish, C.S., 1984, *Programming in Prolog*, Second Edition, Springer Verlag, New York.
- Conckling, E.C., Yeates, M., 1976, *Man's Economic Environment*, McGraw-Hill, New York.

- Couclelis, H.M., 1986a, Artificial Intelligence in Geography, Conjectures on the Shape of Things to Come, *Professional Geographer* **38**: 1–10.
- Couclelis, H.M., 1986b, A Theoretical Framework for Alternative Models of Spatial Decision and Behavior, *Annals of the Association of American Geographers* **76**: 95–113.
- Coulson, R.N., Joseph Folse, L., Loh, D.K., 1987, Artificial Intelligence and Natural Resource Management, *Science* **237**: 262–267.
- Cox, K.R., Golledge, R.G., eds., 1981, *Behavioural Problems in Geography Revisited*, Methuen, New York.
- Crutchfield, J.P., Farmer, J.D., Packard, N.H., Shaw, J.D., 1986, Chaos, *Scientific American* **255**: 38–49.
- Cuadrado, J.L., Cuadrado, C.Y., 1986, A.I. in Computer Vision, Framing Doors and Windows, *Byte*, January, pp. 237–258.
- Cullen, J., 1986, Expert Systems in Planning Analysis, *Town Planning Review* **57**: 239–251.
- Davis, R., 1979, Interactive Transfer of Expertise, Acquisition of New Inference Rules, *Artificial Intelligence* **12**: 121–157.
- Davis, J.R., Grant, I.W., 1987, ADAPT: A Knowledge-based Decision Support System for Producing Zoning Schemes, *Environment and Planning B* **14**: 53–66.
- Davis, J.R., Compagnoni, P.T., Nanninga, P.M., 1987, Roles for Knowledge-based Systems in Environmental Planning, *Environment and Planning B* **14**: 239–254.
- de Greef, P., Breuker, J., 1985, A Case Study in Structured Knowledge Acquisition, in: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, August 18–23, Los Angeles, CA, pp. 581–84.
- Denise, T.C., 1984, On the Nature of INUS Conditionality, *Analysis* **44**: 49–52.
- Densham, P., Rushton, G., 1988, Decision Support Systems for Locational Planning, in: Golledge, R.G., Timmermans, H.J.P., eds., *Behavioral Modelling in Geography and Planning*, Croom Helm, New York, pp. 56–90.
- de Pater, B.C., 1980, Locatietheorie, Locatiekeuze en Ruimtelijke Relaties, *Stedebouw en Volkshuisvesting*: 569–579.

- Desbarats, J., 1983, Spatial Choice and Constraints on Behaviour, *Annals of the Association of American Geographers* **73**: 340–357.
- de Smidt, M., 1975, *Bedrijfsstructuur en Arbeidsmarkt in een Ruimtelijk Kader*, Dissertation, Department of Geography, State University of Utrecht, The Netherlands.
- de Smidt, M., Wever, E., eds., 1990, *The Corporate Firm in a Changing World Economy, Case Studies in the Geography of Enterprise*, Routledge, London.
- Dicken, P., 1976, The Multi-plant Business Enterprise and Geographical Space: Some Issues on the Study of External Control and Regional Development, *Regional Studies* **10**: 401–442.
- Diederich, J., Rushman, J., May, M., 1987, KRITON, A Knowledge Acquisition Tool for Expert Systems, *International Journal of Man–Machine Studies* **26**: 29–40.
- Dobrowolski, G., Kopytowski, J., Woytania, J., Zebrowski, M., 1984, *Alternative Routes from Fossil Resources to Chemical Feedstocks*, IIASA RR-84-19, Laxenburg, Austria.
- Dueker, K.J., 1987, Geographic Information Systems and Computer Aided Mapping, *Journal of the American Planning Association* **53**: 383–390.
- Ellegard, K., Alvstam, C.G., 1987, People-Production-International Division of Labour, in: Hamilton, F.E.I., ed., *Industrial Change in Advanced Economies*, Croom Helm, London, pp. 23–44.
- Erickson, R.A., 1972, The ‘Lead Firm’ Concept, *Tijdschrift voor Economische en Sociale Geografie* **63**: 426–437.
- Erickson, R.A., 1974, The Regional Impact of Growth Firms: The Case of Boeing, 1963–68, *Land Economics* **50**: 127–136.
- Eshelman, L., Ehrt, D., McDermott, J., Tan, W., 1987, MOLE: a Tenacious Knowledge Acquisition Tool, *International Journal of Man–Machine Studies* **26**: 41–54.
- Fedra, K., Weigkricht, E., Winkelbauer, L., 1986, A Hybrid Approach to Information and Decision Support Systems: Hazardous Substances and Industrial Risk Management, in: *IFAC Proceedings of the Conference on Economy and Artificial Intelligence*, September, Aix-en-Provence, France, pp. 169–175.

- Fedra, K., Li, Z., Wang, Z., Zhao, C., 1987, *Expert Systems for Integrated Development: a Case Study of Shanxi Province, The People's Republic of China*, IIASA SR-87-1, Laxenburg, Austria.
- Fedra, K., Zhao, C., 1988, Multi-criteria Data Evaluation Based on DI-DASS/DISCRET, in: *IIASA, Expert Systems for Integrated Development, a Case Study of Shanxi Province, People's Republic of China: Final Report, Volume 1: General Systems Documentation*, Laxenburg, Austria, pp. 209–218, limited distribution.
- Festinger, L., 1962, *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford, CA.
- Finin, T.W., 1986, Interactive Classification: a Technique for Acquiring and Maintaining Knowledge, *Proceedings of the IEEE* **74**: 1414–1419.
- Firn, J.R., 1975, External Control and Regional Development: The Case of Scotland, *Environment and Planning A* **7**: 393–414.
- Flowerdew, R., Green, M., 1989, *Inference Between Incompatible Zonal Systems Using the EM Algorithm*, Paper presented at the Sixth Colloquium on Quantitative and Theoretical Geography, September, 5–9, Paris.
- Forsyth, R., ed., 1984, *Expert Systems, Principles and Case Studies*, Chapman and Hall, London.
- Gale, W.A., 1987, Knowledge Based Knowledge Acquisition for a Statistical Consulting System, *International Journal of Man-Machine Studies* **26**: 55–64.
- Gallaire, H., Minker, J., Nicolas, J., 1984, Logic and Databases, a Deductive Approach, *Computing Surveys* **16**: 153–185.
- Ganapathy, S., Rajamaran, V., 1973, Information Theory Applied to the Conversion of Decision Tables to Computer Programs, *Communications of the ACM* **16**: 532–539.
- Gao, H., 1985, A Brief Introduction to the Achievements of Shanxi Capital Construction, Foreign Affairs Office of the Peoples Government of Shanxi Province, *Shanxi in Progress*, Taiyuan City, Shanxi, PRC.
- Giersch, H., 1963, Das Ökonomische Grundproblem der Regionalpolitik, in: Jürgensen, H., ed., *Gestaltungsprobleme der Weltwirtschaft*, Festschrift für Andreas Predöhl.

- Glasmeier, A., 1988, Factors Governing the Development of High Tech Industry Agglomerations, a Tale of Three Cities, *Regional Studies* **22**: 287–301.
- Gleick, J., 1988, *Chaos. Making a New Science*, Penguin Books, New York.
- Golledge, R.G., Rushton, G., eds., 1976, *Spatial Choice and Spatial Behavior: Geographic Essays on the Analysis of Preferences and Perceptions*, Ohio State University Press, Columbus.
- Golledge, R.G., Rayner, J.N., eds., 1982, *Proximity and Preference, Problems in the Multi-dimensional Analysis of Large Data Sets*, University of Minnesota Press, Minneapolis.
- Golledge, R.G., Timmermans, H.J.P., eds., 1988, *Behavioural Modelling in Geography and Planning*, Croom Helm, New York.
- Gould, J, Kolb, W.L., eds., 1964, *A Dictionary of the Social Sciences*, The Free Press of Glencoe, New York.
- Grauer, M., Lewandowski, A., Wierzbicki, A., 1982, *Methodology and Use of Multiple Objective Decision Making in Chemical Engineering*, IIASA WP-82-37, Laxenburg, Austria.
- Grauer, M., Thompson, M., Wierzbicki, A., 1984, Plural Rationality and Interactive Decision Processes, *Proceedings of a IIASA Summer Study on Plural Rationality and Interactive Decision Processes*, August 16–26, Sopron, Hungary, Springer Verlag.
- Green, M.B., 1981, Regional Preferences for Interlocking Directorates among the Largest American Corporations, *Environment and Planning A* **13**: 829–839.
- Green, N.P., Finch, S., Wiggins, J., 1985, The ‘State of the Art’ in Geographical Information Systems, *Area* **17**: 295–301.
- Greenhut, M.L., 1956, *Plant Location in Theory and in Practice*, University of North Carolina Press, Chapel Hill, NC.
- Gupta, M.M., Saridis, G.N., Gaines, B.R., eds., 1979, *Fuzzy Automata and Decision Processes*, North Holland, New York.
- Hamilton, F.E.I., 1974, A View of Spatial Behavior, Industrial Organization and Decision-making, in: Hamilton, F.E.I., ed., *Spatial Perspectives on Industrial Organization and Decision-making*, John Wiley, Chichester.
- Hart, A., 1986, Knowledge Acquisition for Expert Systems, *New Technology Modular Series 8605*, Kogan Page, London.

- Haley, P., Williams, Ch., 1986, Expert System Development Requires Knowledge Engineering, *Computer Design*, February, pp. 83–88.
- Hendriks, P.H.J., 1983, Unidimensionele Conjunctanalyse en het Voorspellen van Geaggregeerd Ruimtelijk Koopgedrag, *Planning, Methodiek en Toepassing* 21: 11–20.
- Hendriks, P.H.J., 1986, De Relationale Definitie van Begrippen: Een Relationeel Realistische Visie op het Operationaliseren en Representeren van Begrippen, *Netherlands Geographical Studies* 24.
- Hendriks, P. et al., 1984, *Regionaal Beleid en het Produktiemilieu van Bedrijven*, Research publication nr.11, Research Group T.M.T., Dept. of Geography, University of Nijmegen, The Netherlands.
- Hendriks, P.H.J., Smagt, A.G.M. van der, 1988, Definition of Choice Sets and Choice Set Attributes, Some Problems Inherent in Decompositional Modelling, in: Golledge, R.G., Timmermans, H.J.P., eds., *Behavioural Modelling in Geography and Planning*, Croom Helm, New York, pp. 27–37.
- Hobbs, B.F., 1980, Multiobjective Power Plant Siting Methods, *Journal of the Energy Division* 106: 187–200.
- Hobbs, B.F., 1984, Regional Energy Facility Models for Power System Planning and Policy Analysis, in: Lev, B., et al., *Analytic Techniques for Energy Planning*, Elsevier Science Publishers B.V., Amsterdam, pp. 53–66.
- Hofstadter, D.R., 1979, *Gödel, Escher, Bach: An Eternal Golden Braid*, Harvester Press, Sussex.
- Holland, S., 1976, *The Regional Problem*, Macmillan, London.
- Hoogkamer, J.J., Scholte, R.H., 1984, Regionale Verschillen in de Kwaliteit van het Produktiemilieu voor Bedrijven, *Economisch Statistische Berichten*: 560–564.
- Hopkins, L.D., 1977, Methods for Generating Land Suitability Maps: A Comparative Evaluation, *Journal of the American Institute of Planning* 43: 386–400.
- IEEE, 1986, *Proceedings of the Institute of Electrical and Electronic Engineers Inc.*, New York, Special Issue on Knowledge Representation, 74, October.

- IIASA, 1988, *Expert Systems for Integrated Development. A Case Study of Shanxi Province, People's Republic of China: Final Report, Volume 1: General Systems Documentation*, Laxenburg, Austria, limited distribution.
- Isard, W., 1956, *Location and Space Economy*, Massachusetts Institute of Technology Press, Cambridge, MA.
- Ispen, D., 1978, Das Konstrukt Zufriedenheit, *Soziale Welt* **29**: 44–53.
- Jackson, F., 1982, Review on Brand, M., 1976, The Nature of Causation, *Journal of Symbolic Logic* **47**: 470–473.
- Johnston, R.J., ed., 1981, *The Dictionary of Human Geography*, Blackwell Reference, Oxford.
- Kahn, G., Nowlan, S., McDermott, J., 1985, MORE: an Intelligent Knowledge Acquisition Tool, in: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, August 18–23, Los Angeles, CA, pp. 581–84.
- Karlqvist, A., Lundqvist, L., Snickars, F., eds., 1975, *Dynamic Allocation of Urban Space*, Saxon House, Farnborough, England.
- Keen, P.G.W., 1986, Value Analysis: Justifying Decision Support Systems, in: Sprague, R., Watson, H.J., eds., *Decision Support Systems. Putting Theory into Praxis*, Prentice Hall, Englewood Cliffs, NJ, pp. 48–64.
- Kickert, W.J.M., 1978, *Fuzzy Theories on Decision Making, a Critical Review*, Martinus Nijhoff Social Science Division, Leiden, The Netherlands.
- Krumme, G., 1969, Towards a Geography of Enterprise, *Economic Geography* **45**: 30–40.
- Kuhn, T.S., 1962, The Structure of Scientific Revolutions, in: Neurath, O., Carnap, R., Morris, Ch., eds., *The International Encyclopedia of Unified Science*, University of Chicago Press, Chicago, IL.
- Kuper, A., Kuper J., eds., 1985, *The Social Science Encyclopedia*, Routledge and Kegan Paul, London.
- Kutter, E., 1973, Aktionsbereiche des Stadtbewohners, *Archive der Kommunalwissenschaften* **12**: 69–85.
- Langendorf, R., 1985, Computers and Decision Making, *Journal of the American Planning Association* **51**: 422–433.

- Leary, M., 1987, Expert Systems in Planning, Two Responses to Issues Raised by Cullen, *Town Planning Review* **58**: 331-342.
- Levesque, H.J., 1984, Foundation of a Functional Approach to Knowledge Representation, *Artificial Intelligence* **23**: 155-212.
- Lieber, S., 1979, An Experimental Approach for the Migration Decision Process, *Tijdschrift voor Economische en Sociale Geografie* **70**: 75-85.
- Littman, D.C., 1987, Modelling Human Expertise in Knowledge Engineering, some Preliminary Observations, *International Journal of Man-Machine Studies* **26**: 81-92.
- Lloyd, P.E., Dicken, P., 1977, *Location in Space: A Theoretical Approach to Economic Geography*, second edition, Harper and Row Publishers, London.
- Lösch, A., 1954, *The Economics of Location*, Yale University Press, New Haven, CT.
- Louviere, J.J., Wilson, E.H., Piccolo, M., 1980, Applications of Psychological Measurement and Modelling to Behavioural Travel-Demand Analysis, in: Henscher, D.A., Stopher, P.R., eds., *Behavioural Travel Modelling*, Croom Helm, London, pp. 713-738.
- Louviere, J.J., 1981, A Conceptual and Analytical Framework for Understanding Spatial and Travel Choice, *Economic Geography* **57**: 304-314.
- Lucardie, L., 1988a, Relationale Begripsanalyse en Beslissingstabellen: Aanzetten tot Generalisatie, in: Smagt, A.G.M. van der, Hendriks, P.H.J., eds., *Methoden op een Keerpunt*, *Netherlands Geographical Studies* **77**: 59-81.
- Lucardie, L., 1988b, *Beslissingstabellen en Kennissystemen*, Dept. of Geography, Faculty of Policy Sciences, University of Nijmegen, The Netherlands.
- Lucas, P.J.F., Gaag, L.C. van der, 1986, Prolog en Expertsystemen, *Informatie* **28**: 16-25.
- Lucas, P.J.F., Gaag, L.C. van der, 1988, *Principes van Expertsystemen*, Academic Press, Schoonhoven, The Netherlands.
- Luce, R.D., Tukey, J.W., 1964, Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement, *Journal of Mathematical Psychology* **1**: 1-7.

- Lundberg, C.G., 1989, Knowledge Acquisition and Expertise Evaluation, *Professional Geographer* 41: 272–282.
- Lynch, K., 1962, *Site Planning*, Massachusetts Institute of Technology Press, Cambridge, MA.
- Lyons, T.P., 1987, Interprovincial Trade and Development in China, 1957–1979, *Economic Development and Cultural Change* 35: 223–256.
- Mackie, J.L., 1965, Causes and Conditions, *American Philosophical Quarterly* 2: 245–264.
- Mason, C.M., 1985, The Geography of ‘Successful’ Small Firms in the U.K., *Environment and Planning A* 17: 1499–1513.
- Massey, D., 1974a, Towards a Critique of Industrial Location Theory, *Centre for Environmental Studies Research Paper* 5: 1–27.
- Massey, D., 1974b, Is the Behavioural Approach Really an Alternative? in: Massey, D., Morrison, W.I., eds., *Industrial Location: Alternative Frameworks*, *Centre for Environmental Studies Research Paper* 15: 79–86.
- Massey, D., 1975a, Approaches to Industrial Location Theory: A Possible Spatial Framework, in: Cripps, E.L., ed., *Regional Science – New Concepts and Old Problems*, Pion, London, pp. 84–108.
- Massey, D., 1975b, Behavioural Research, *Area* 7: 201–203.
- Massey, D., 1979a, A Critical Evaluation of Industrial Location Theory, in: Hamilton, F.E.I., Linge, G.D.R., eds., *Spatial Analysis, Industry and the Industrial Environment, Volume 1: Industrial Systems*, John Wiley, Chichester, pp. 57–72.
- Massey, D., 1979b, In What Sense a Regional Problem? *Regional Studies* 1: 233–243.
- McDermott, D., Doyle, J., 1980, Non-Monotonic Logic 1, *Artificial Intelligence* 13: 14–72.
- McMillan, T.E., 1965, Why Manufacturers Choose Plant Locations vs. Determinants of Plant Locations, *Land Economics* 41: 239–246.
- Meyer-Kramer, F., 1985, Innovation Behaviour and Regional Indigenous Potential, *Regional Studies* 19: 523–534.
- Miller, G.A., Galanter, E., Pribram, K.H., 1960, *Plans and Structure of Behaviour*, Holt, Rinehart and Winston Inc., New York.

- Minsky, M., 1975, A Framework for Representing Knowledge, in: Winston, P.H., ed., *The Psychology of Computer Vision*, McGraw Hill, New York, pp. 211–280.
- Mittra, S.S., 1986, Decision Support Systems, *Tools and Techniques*, John Wiley & Sons, New York.
- Montalbano, M., 1974, *Decision Tables*, Scientific Research Associates Inc., Chicago.
- Mylopoulos, J., Levesque, H.J., 1984, An Overview of Knowledge Representation, in: Brodie, M.L., Mylopoulos, J., Schmidt, J.W., eds., *On Conceptual Modelling, Perspectives from Artificial Intelligence, Databases and Programming Languages*, Springer Verlag, New York, pp. 3–17.
- Negoita, C.V., 1985, *Expert Systems and Fuzzy Systems*, Benjamin/Cummings Publ. Co, Menlo Park, CA.
- New Town Guide, 1978, *How to Choose the Best New Town for your Business*: 10–14.
- Nguyen, T.A., Perkins, W.A., Laffey, T.J., Pecora, D., 1985, Checking an Expert System's Knowledge Base for Consistency and Completeness, in: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, August 18–23, Los Angeles, CA, pp. 375–378.
- NIJCAI, 1987, *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 1987, August 18–23, Los Angeles, CA.
- Nijkamp, P., Spronk, J., eds., 1981, Multiple Criteria Analysis, *Operational Methods*, Aldershot, Gower.
- Op 't Veld, D., 1988, Een Beslissingsondersteunend Model voor Woningmarktscenario's, in: Smagt, A.G.M. van der, Hendriks, P.H.J., eds., *Methoden op een Keerpunt*, *Netherlands Geographical Studies* 77: 128–142.
- Op 't Veld, D., Bijlsma, E., Starmans, J., 1986, *Individuele Gedragsmodellen: een Verkenning van Kennissystemen en Decision Plan Nets*, Planning Study-Center TNO, Delft, The Netherlands.
- Op 't Veld, D., Bijlsma, E., Starmans, J., Timmermans, H., 1987, Kennissystemen in de Ruimtelijke Planning, *Planning, Methodiek en Toepassing* 29: 2–10.
- Ortolano, L., Perman, C.D., 1987, A Planner's Introduction to Expert Systems, *Journal of the American Planning Association* 53: 98–103.

- Österreichisches Institut für Raumplanung, 1965, Industrie-Entwicklungsprogramm Niederösterreich: Raumordnungsgutachten, *Veröffentlichungen des Österreichisches Institut für Raumplanung* nr.23, Vienna.
- Oudheusden, J., et al., 1982, *Bedrijfsverplaatsing en Produktiemilieu*, Dept. of Geography, State University of Utrecht, The Netherlands.
- Park, C.W., Hughes, W., Thurskal, V., Friedman, R., 1981, Consumer's Decision Plans and Subsequent Choice Behaviour, *Journal of Marketing* 45: 33-47.
- Pawson, R., 1982, Desperate Measures, *The British Journal of Sociology* 33: 11-25.
- Pellenbarg, P.H., 1985, *Bedrijfsrelokatie en Ruimtelijke Kognitie*, Dissertation, Dept. of Geography, State University of Groningen, The Netherlands.
- Pellenbarg, P., Schuurmans, F., Wouters, J., 1974, *De Ontwikkelmogelijkheden van Medemblik: Proeve van een Feasibility Study*, Department of Geography, State University of Groningen, The Netherlands.
- Perry, M., 1985, New Small Factories in the Public Sector, *Area* 17: 185-191.
- Pirie, G.H., 1976, Thoughts on Revealed Preference and Spatial Behaviour, *Environment and Planning A* 8: 947-955.
- Posch, M., 1988, Gaussian Dispersion Models: Application to the Shanxi Province Case Study: in: *Expert Systems for Integrated Development, a Case Study of Shanxi Province, People's Republic of China: Final Report, Volume 1: General Systems Documentation*, IIASA, Laxenburg, Austria, pp. 179-196, limited distribution.
- Pred, A., 1967, Behaviour and Location, Lund Studies in Geography, Series B, *Human Geography* 27 and 28, Gleerup, Lund.
- Pred, A., 1977, *City Systems in Advanced Economies: Past Growth, Present Processes, and Future Development Options*, Hutchinson, London.
- Quinlan, J.R., 1979, Discovering Rules by Induction from Large Collections of Examples, in: Michie, D., ed., *Expert Systems in the Micro Electronic Age*, Edinburgh University Press, Edinburgh, pp. 168-201.
- Quinlan, J.R., 1983, Learning Efficient Classification Procedures and their Application to Chess End Games, in: Michalski, R.S., Carbonell, J.G., Mitchell, T.M., eds., *Machine Learning, an Artificial Intelligence Approach*, Tioga Publ. Comp., Palo Alto, CA, pp. 463-482.

- Rajamoney, S., Jong, G. de, Faltings, B., 1987, Towards a Model of Conceptual Knowledge Acquisition Through Directed Experimentation, in: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, August 18–23, Los Angeles, CA, pp. 688–690.
- Rawstron, E.M., 1958, Three Principles of Industrial Location, *Transactions and Papers of the Institute of British Geographers* 25: 132–142.
- Rawstron, E.M., 1966, Electric Power Generation, in: Edwards, K.C., ed., *Nottingham and its Region, Prepared for the meeting of the British Association of the Advancement of Science*, Nottingham.
- R.E.B. Coördinatie, 1979, *Regionaal Produktiemilieu Onderzoek*, Utrecht, The Netherlands.
- R.E.B. Coördinatie, 1980, *Een Systeembenadering van de Regionale Economie*, Utrecht, The Netherlands.
- Reilly, K.D., Salah, A., Yang, C., 1987, A Logic Programming Perspective on Decision Table Theory and Practice, *Data and Knowledge Engineering* 2: 191–212.
- Reitman Olson, J., Rueter, H.H., 1987, Extracting Knowledge from Experts, Methods for Knowledge Acquisition, *Expert Systems* 4: 152–168.
- Reitsma, R.F., 1986, *Reconstructing Choice Sets Using Prolog*, Paper presented at the I.G.U. conference on mathematical modelling, August, Madrid.
- Reitsma, R.F., 1987, Inferentiebomen: Een Techniek voor het Modelleren van Relationele Beslisregels, *SWIM* 2: 2–8.
- Reitsma, R.F., 1988a, *REPLACE: A Computational Approach to Spatial Choice Modelling*, Paper presented at the Annual Conference of the Institute of British Geographers, January, Loughborough, England.
- Reitsma, R.F., 1988b, Relationeel Modelleren van het Produktiemilieu van Typen Bedrijvigheid, in: Smagt, A.G.M. van der, Hendriks, P.H.J., eds., *Methoden op een Keerpunt, Netherlands Geographical Studies*, nr. 77, pp. 82–105.
- Reitsma, R.F., 1990, Inference Trees: A Method for Modelling Relational Decision Rules, in: Pfersmann, O., ed., *Proceedings of the Conference on Intelligence and Society*, March 23–31, 1988, Vienna, Austria, forthcoming.

- Reitsma, R.F., Vergoossen D., 1988, A Causal Typology of Migration: the Role of Commuting, *Regional Studies* **22**: 331–340.
- Roelofs, B., Wever, E., 1985, *Regio en Economische Potentie: een Literatuurverkenning*, Studierapport nr.28, Physical Planning Agency, RPD, The Hague, The Netherlands.
- Rushton, G., 1969, Analysis of Spatial Behavior by Revealed Space Preference, *Annals of the Association of American Geographers* **59**: 391–400.
- Sacerdoti, E.D., 1974, Planning in a Hierarchy of Abstraction Spaces, *Artificial Intelligence* **5**: 115–135.
- Samuelson, P.A., 1947, *Foundations of Economic Analysis*, Harvard University Press, Cambridge, MA.
- Savory, E.S., ed., 1985, *Künstliche Intelligenz und Expertensysteme*, R. Oldenburg Verlag GmbH, München, FRG.
- Schilling, H., 1968, *Standortfaktoren für die Industrieansiedlung, ein Katalog für die Regionale und Kommunale Entwicklungspolitik sowie die Standortwahl von Unternehmen*, Österreichisches Institut für Raumplanung, Veröffentlichungen nr.27, Kohlhammer GmbH, Stuttgart, FRG.
- Scholten, H.J., Vlugt, M. van der, 1988, *Applications of Geographical Information Systems in Europe*, Paper presented at the UNISYS Management Seminar on Geographic Information Systems, Pacific/Asia Division, October 25–26, Sydney.
- Schotel, H., 1987, *Programmeren in Prolog*, Coutinho, Muiderberg, The Netherlands.
- Schuler, H., 1979, A Disaggregate Store-Choice Model of Spatial Decision Making, *Professional Geographer* **31**: 146–156.
- Scott, A.J., 1971, *An Introduction to Spatial Allocation Analysis*, Association of American Geographers, Washington D.C.
- Shapiro, A., 1983, *The Role of Structured Induction in Expert Systems*, Dissertation, Machine Intelligence Research Unit, University of Edinburgh, United Kingdom.
- Shapiro, A., Niblett, T., 1982, Automatic Induction of Classification Rules for a Chess Endgame, in: Clarke, M.R.B., ed., *Advances in Computer Chess* **3**: 73–92.
- Shapiro, E.Y., 1981, *Inductive Inference of Theories from Facts*, Dept. of Computer Science, Yale University.

- Sheppard, E., 1979, Notes on Spatial Interaction, *Professional Geographer* **31**: 8–15.
- Short, J.R., 1977, The Intra-Urban Migration Process, Comments and Empirical Findings, *Tijdschrift voor Economische en Sociale Geografie* **68**: 362–370.
- Shwayder, K., 1971, Conversion of Limited-Entry Decision Tables to Computer Programs. A Proposed Modification of Pollack's Algorithm, *Communications of the ACM* **14**: 69–73.
- Shwayder, K., 1974, Extending the Information Theory Approach to Converting Limited-Entry Decision Tables to Computer Programs, *Communications of the ACM* **17**: 532–537.
- Simon, H.A., 1960, *The New Science of Management Decisions*, Harper and Row Publishers, New York.
- Simon, H.A., 1969, *The Sciences of the Artificial*, Massachusetts Institute of Technology Press, Cambridge, MA.
- Sinth, M., 1985, *Kunstmatige Intelligentie*, Syllabus, Dept of Computer Science, University of Nijmegen, The Netherlands.
- Smith, D.M., 1971, *Industrial Location, an Economic Geographical Analysis*, John Wiley & Sons Inc., New York.
- Smith, I.J., 1979, The Effect of External Takeovers on Manufacturing Employment Change in the Northern Region Between 1963 and 1973, *Regional Studies* **13**: 421–437.
- Smith, T.R., 1983, Computational Process Models of Individual Decision Making, in: Crosby, R.W., ed., *Cities as Non-linear Decision Systems*, American Association for the Advancement of Science, Washington D.C., pp. 175–210.
- Smith, T.R., 1984, Artificial Intelligence and the Applicability to Geographical Problem Solving, *Professional Geographer* **36**: 147–158.
- Smith, T.R., Pellegrino, J.W., Golledge, R.G., 1982, Computational Process Modeling of Spatial Cognition and Behavior, *Geographical Analysis* **14**: 305–325.
- Smith, T.R., Clark, W.A., Cotton, J.W., 1984, Deriving and Testing Production System Models of Sequential Decision Making Behavior, *Geographical Analysis* **16**: 191–222.

- Smithson, M., 1987, *Fuzzy Set Analysis for Behavioral and Social Sciences*, Springer Verlag, New York.
- Snellen, I.T.M., 1987, *Boeiend en Geboeid, Ambivalenties en Ambities in de Bestuurskunde*, Rede uitgesproken bij de aanvaarding van het ambt van gewoon hoogleraar in de bestuurskunde aan de Katholieke Universiteit Brabant, Samson & Tjeenk Willink, Alphen a/d Rijn, The Netherlands.
- Sowa, J.F., 1984, *Conceptual Structures, Information Processing in Mind and Machine*, Addison-Wesley, Reading, MA.
- Spehl, H. et al., 1981, *Wirtschaftspolitik und regionale Entwicklungsplanung in strukturschwachen Regionen. Erfolgskontrolle und alternative Entwicklungskonzeptionen*, Schriftenreihe, Band 4, Gesellschaft für Regionale Strukturentwicklung, Bonn.
- Sprague, R., 1986, A Framework for the Development of Decision Support Systems, in: Sprague, R., Watson, H.J., eds., *Decision Support Systems. Putting Theory into Praxis*, Prentice Hall, Englewood Cliffs, NJ, pp. 7-32.
- Steed, G.P.F., 1971, Internal Organization, Firm Integration and Locational Change: the Northern Ireland Linen Complex, 1954-64, *Economic Geography* 47: 371-383.
- Stefik, M., 1981, Planning with Constraints, MOLGEN: part 1, *Artificial Intelligence* 16: 111-139.
- Sterling, L., Shapiro, E., 1986, *The Art of Prolog, Advanced Programming Techniques*, Massachusetts Institute of Technology Press, Cambridge, MA.
- Stijnenbosch, M.H., Witlox, A.H.T., Oudheusden, J.C. van, 1983, Een Methode voor de Vergelijking van Productiemilieus in Gemeenten, *Planning, Methodiek en Toepassing* 20: 26-33.
- Strassert, G., 1984, Regionales Entwicklungspotential: ein Versuch der Ent-rätselung eines Schlagwortes, *Raumforschung und Raumordnung* 42: 19-26.
- Strzepak, K.M., Fedra, K., 1988, Water Resources Simulation: MITSIM, in: *Expert Systems for Integrated Development, a Case Study of Shanxi Province, People's Republic of China: Final Report, Volume 1: General Systems Documentation*, IIASA, Laxenburg, Austria, pp. 175-178, limited distribution.

- Sutherland, J.W., 1988, Prospects for Logical, Versus Data, Processing Facilities as Strategic Decision Aids, *IEEE Transactions on Systems, Man, and Cybernetics* 18: 885-898.
- Svart, L.M., 1974, On the Priority of Behaviour in Behavioural Research: a Dissenting View, *Area* 6: 301-305.
- Swartout, W.R., Smoliar, S.W., 1987, On Making Expert Systems More Like Experts, *Expert Systems* 4: 196-207.
- Thompson, B., Thompson, W., 1986, Finding Rules in DATA, an Algorithm for Extracting Knowledge from Data, *Byte*, November, pp. 149-158.
- Thoss, R., 1977, Die Verteilung des Entwicklungspotentials als Problem der Regionalpolitik in der Bundesrepublik Deutschland, in: Bühr, W., Friedrich, P., ed., *Konkurrenz zwischen kleinen Regionen*, Schriften zur Öffentlichen Verwaltung und Öffentlichen Wirtschaft, Band 23, Baden-Baden.
- Thoss, R., 1983, Qualitatives Wachstum in der Raumordnungsregionen der Bundesrepublik Deutschland, in: *Gleichwertige Lebensbedingungen durch eine Raumordnungspolitik des mittleren Weges. Indikatoren, Potentiale, Instrumente*, Forschungs- und Sitzungsberichte der Akademie für Raumforschung und Landesplanung, Band 140, pp. 1-23, Hannover.
- Thrift, N., 1981, Behavioural Geography, in: Wrigley, N., Bennet, R.J., eds., *Quantitative Geography*, a British Review, Routledge & Kegan Paul, London, pp. 352-365.
- Tijdink, N., 1981, *Knelpuntenanalyse ERBO 1981*, VVK, Woerden, The Netherlands.
- Timmermans, H.J.P., 1984, Decompositional Multi-attribute Preference Models in Spatial Choice Analysis, a Review of some Recent Developments, *Progress in Human Geography* 8: 189-221.
- Timmermans, H.J.P., 1986, Locational Choice Behaviour of Entrepreneurs, an Experimental Analysis, *Urban Studies* 23: 231-240.
- Timmermans, H.J.P., 1987, Individuele Beslismodellen en Vervoerswijzekeuze 1, Compensatorische en Non-Compensatorische Beslisregels, *Verkeerskunde* 38: 350-353.
- Timmermans, H.J.P., Rushton, G., 1979, Revealed Spatial Preference Theory, A Rejoinder, *Tijdschrift voor Economische en Sociale Geografie* 70: 309-312.

- Timmermans, H.J.P., Heijden, R. van der, Westerveld, H., 1984, Decision Making between Multi-attribute Choice Alternatives: a Model of Spatial Shopping Behaviour Using Conjoint Measurements, *Environment and Planning A* **16**: 377–387.
- Timmermans, H.J.P., Heijden, R. van der, 1984, The Predictive Ability of Alternative Decision Rules in Decompositional Multi-attribute Preference Modelling, *Sistemi Urbani* **1**: 89–101.
- Timmermans, H.J.P., Heijden, R. van der, 1987, Uncovering Spatial Decision Making Processes: a Decision Net Approach Applied to Recreational Choice Behaviors, *Tijdschrift voor Economische en Sociale Geografie* **78**: 297–304.
- Tomlinson, R.F., 1987, Current and Potential Uses of Geographical Information Systems, the North American Experience, *International Journal of Geographical Information Systems* **1**: 203–218.
- Townroe, P.M., 1968, *Industrial Location and Regional Economic Policy*, a Selected Bibliography, Centre for Urban and Regional Studies, University of Birmingham.
- Townroe, P.M., 1969, Locational Choice and the Industrial Firm, *Regional Studies* **3**: 15–24.
- Townroe, P.M., 1972, Some Behavioural Considerations in the Industrial Location Decision, *Regional Studies* **6**: 261–272.
- Townroe, P.M., 1973, Industrial Location Search Behaviour and Regional Planning, in: Rees, J., Newby, P., eds., Behavioural Perspectives in Geograph, *Middlesex Polytechnic Monographs in Geography* **1**: 44–58.
- Turban, E., Watkins, P.R., 1986, Integrating Expert Systems and Decision Support Systems, in: Sprague, R., Watson, H.J., eds., *Decision Support Systems. Putting Theory into Praxis*, Prentice Hall, Englewood Cliffs, NJ, pp. 138–152.
- Vaessen, P., 1989, *Bedrijf, Regio en Succes, een Micro-onderzoek*, Dept. of Social and Economic Geography, Faculty of Policy Science, University of Nijmegen, The Netherlands.
- Vallance, B., Weigkricht, E., 1988, *Interactive Energy Demand Analysis: A Case Study of Shanxi Province, PRC*, IIASA WP-88-97, Laxenburg, Austria.

- van Blokland, F., Roelofs, B., 1984, Het Begrip Produktiemilieu, Definitie en Meting, in: Hendriks, P. *et al.*, *Regionaal Beleid en het Produktiemilieu van Bedrijven*, Research publication nr.11, research group T.M.T., Dept. of Geography, University of Nijmegen, The Netherlands, pp. 9–65.
- van Delft, A., Nijkamp, P., eds., 1977, Multi Criteria Analysis and Regional Decision Making, *Studies in Applied Regional Science* 8, Nijhoff, Delft, The Netherlands.
- van der Gaag, L.C., Lucas, P.J.F., 1988, Prolog: Specificatie=Implementatie, *Informatie* 27: 766–774.
- van der Heijden, R.E.C.M., 1986, *A Decision Support System for the Planning of Retail Facilities, Theory, Methodology, and Applications*, Dissertation, Technical University of Eindhoven, The Netherlands.
- van der Smagt, A.G.M., 1985, *Definiëren en Relateren in Sociaal Wetenschappelijk Onderzoek*, Dissertation, Dept. of Geography, University of Nijmegen, The Netherlands.
- van der Smagt, A.G.M., 1988, Het Semantische Datamodel als Grondslag voor Informatie-systemen, in: Smagt, A.G.M. van der, Hendriks, P.H.J., eds., *Methoden op een Keerpunt, Netherlands Geographical Studies* 77: 38–58.
- van der Smagt, A.G.M., Lucardie, L., 1990, *Decision Making Under Not Well-Defined Conditions: From Data Processing to Logical Modelling*, forthcoming in *Tijdschrift voor Economische en Sociale Geografie*.
- van de Vooren, F.W.C.J., 1980, *Regionaal Ontwikkelingspotentieel, een Wezenlijk Aspect voor het Regionaal Beleid*, Economisch Technologisch Instituut Overijssel, Zwolle, The Netherlands.
- van de Vooren, F.W.C.J., Wagenaar, S., 1984, Het Regionale Ontwikkelingspotentieel, *Economisch Statistische Berichten*: 800–803.
- van Dinteren, J., Reitsma, R., 1985, Keuzemodellen in het Onderzoek naar de Vestigingsplaats van Kantoorondernemers, in: Dinteren, J.H.J., Hart, H.W. ter, eds., *Geografie en Kantoren 1985, Netherlands Geographical Studies* 21: 83–97.
- Verhelst, M., 1980, *De Praktijk van Beslissingstabellen*, Kluwer, Deventer.
- Vonk, F., Willems, J., 1972, *Strukturonderzoek Oude IJsselstreek*, Vereniging van Nederlandse Gemeenten, The Hague, The Netherlands.

- Watts, H.D., 1974, Spatial Rationalization in Multi-plant Enterprises, *Geoforum* 17: 69–76.
- Watts, H.D., 1978, Interorganizational Relations and the Location of Industry, *Regional Studies* 12: 215–225.
- Wheeler, J.O., 1981, Effects of Geographical Scale on Location Decisions in Manufacturing: The Atlanta Example, *Economic Geography* 57: 134–145.
- Wielinga, B., Breuker, J., 1986, *Model Driven Knowledge Acquisition for Expert Systems*, Dept. of Social Science Informatics, University of Amsterdam, The Netherlands.
- Wierzbicki, A., 1979, *A Methodological Guide to Multi-Objective Optimization*, IIASA WP-79-122, Laxenburg, Austria.
- Wierzbicki, A., 1980, *A Mathematical Basis for Satisficing Decision Making*, IIASA WP-80-90, Laxenburg, Austria.
- Wilder, M., 1985, Site and Situation Determinants of Land-use Change: an Empirical Example, *Economic Geography* 61: 332–344.
- Wilensky, R., 1983, *Planning and Understanding, a Computational Approach to Understanding*, Addison Wesley, Reading, MA.
- Winkelbauer, L., Reitsma, R.F., 1988, Expert Systems for Integrated Development: a Hybrid Approach to Decision Support, in: Cheng, W., ed., *Proceedings of the International Conference on Systems, Science and Engineering, ICSSE*, July 25–28, Beijing, PRC, International Academic Publishers, pp. 909–912.
- Woods, W.A., 1988, Knowledge Representation, in: Bartee, T.C., ed., *Expert Systems and Artificial Intelligence, Applications and Management*, Howard W Sams & Company, Indianapolis, IN, pp. 147–176.
- Zadeh, L.A. et al., eds., 1975, *Fuzzy Sets and their Applications to Cognitive and Decision Processes*, Academic Press, New York.
- Zebrowski, M., Dobrowolski, G., Rys, T., et al., 1988, Industrial Structure Optimization: the PDAS Model, in: *Expert Systems for Integrated Development, a Case Study of Shanxi Province, People's Republic of China: Final Report, Volume 1: General Systems Documentation*, IIASA, Laxenburg, Austria, pp. 87–134, limited distribution.
- Zhang, R., Yang, S., Zhou, D., 1988, The Input-Output Model System, in: *Expert Systems for Integrated Development, a Case Study of Shanxi*

Province, People's Republic of China: Final Report, Volume 1: General Systems Documentation, IIASA, Laxenburg, Austria, pp. 31–56, limited distribution.

Appendix 1:

Some remarks on INUS conditionality

In an article in *Analysis* (1984; pp. 49–52), Denise discusses some of the problems associated with the definition of INUS conditionality by Mackie. Mackie (1965; p. 246) defines an INUS condition as:

A is an INUS condition of a result *P* if and only if, for some *X* and for some *Y*, $((A \text{ AND } X) \text{ OR } Y)$ is a necessary and sufficient condition of *P*, but *A* is not a sufficient condition of *P* and *X* is not a sufficient condition of *P*.

Intuitively this definition is vulnerable for (possibly) irrelevant or uninteresting conditions, for one can define any condition (or its negation) as an INUS condition. Denise cites Jackson's review (Jackson, 1982) of some articles on causation theory from a book by Brand (1976), who shows that with a bit of Boolean trickery, this can indeed be proven. Jackson's proof goes like this:

“Let ‘*A*’ be, say, ‘My office is in the Hall of Languages’, ‘*P*’ be ‘No two snowflakes have identical patterns’, and ‘*S1*’ and ‘*S2*’ be marks for distinct sufficient conditions of *P* such that $(S1 \text{ OR } S2)$ is necessary and sufficient for *P*. Now let $X = (S1 \text{ OR } \text{not}(A))$ and $Y = (S2 \text{ OR } (\text{not}(A) \text{ AND } S1))$. Then under these assumptions, “ $(A \text{ AND } X)$ is a sufficient condition for *P*, since if it obtains, then *S1* does [if $(A \text{ AND } (S1 \text{ OR } \text{not}(A)))$, i.e., $((A \text{ AND } S1) \text{ OR } (A \text{ AND } \text{not}(A)))$, i.e., $(A \text{ AND } S1)$, then *S1*]. But *A* is not a sufficient condition for *P*, and neither is *X*. Moreover, $((A \text{ and } X) \text{ OR } Y)$ is a necessary and sufficient condition for *P*, since it obtains if and only if $(S1 \text{ OR } S2)$ does [$((A \text{ and } S1) \text{ OR } (S2 \text{ OR } (\text{not}(A) \text{ AND } S1)))$, i.e., $(S1 \text{ OR } S2)$ therefore $(S1 \text{ or } S2) = P$]; and so *A* satisfies Mackie's definition” (Jackson, 1982; p. 492).

Denise (p. 50) proposes to augment Mackie's definition somewhat in order to make it immune for operations like Jackson's:

“*A* is an INUS condition of a result *P* if and only if, for some *X* and for some *Y*, $((A \text{ AND } X) \text{ OR } Y)$ is a necessary and sufficient condition of *P* and *A* is a necessary condition of $(P \text{ AND } X)$, but *A* is not a sufficient condition of *P* and *X* is not a sufficient condition of *P*.”

Considering the same assumptions as in the Jackson example, A is not a necessary condition of $(P \text{ AND } X)$, i.e., $(P \text{ AND } X)$ can obtain, even if A fails to, since if $(P \text{ AND } \text{not}(A))$, then $(P \text{ AND } X)$ does, since $(P \text{ AND } (S1 \text{ OR } \text{not}(A)))$ does. This renders the individuality of snowflakes no longer dependent on the location of Denise's office location!

Denise also discusses the use of this augmentation of Mackie's definition for preventing redundant factors to be denoted causal factors. He provides the following example of a table containing truth values for the conditions A, B , and C , and the result P :

	A	B	C	P	
(1)	t	t	t	f	
(2)	t	t	f	t	S1
(3)	t	f	t	t	S2
(4)	t	f	f	f	
(5)	f	t	t	f	
(6)	f	t	f	t	S3
(7)	f	f	t	f	
(8)	f	f	f	f	

The combinations of truth conditions for A, B , and C in $S1, S2$, and $S3$ are sufficient although unnecessary conditions of P , while the truth conditions of A, B , and C are necessary but insufficient conditions of P . According to Mackie, the truth conditions of A, B , and C are therefore INUS conditions of P . However, $S1$ represents $(A \text{ AND } B \text{ AND } \text{not}(C))$, while $S3$ represents $(\text{not}(A) \text{ AND } B \text{ AND } \text{not}(C))$. This implies, that what causes P is $(B \text{ AND } \text{not}(C))$, regardless of the truth value of A . Denise's extension of Mackie's definition of INUS conditionality prevents A and $\text{not}(A)$ to be considered INUS conditions for P , because under the amended definition, only each conjunct in the disjunction $((B \text{ AND } \text{not}(C)) \text{ OR } (A \text{ AND } \text{not}(B) \text{ AND } C))$ is considered an INUS condition.

It is interesting to note the connection between this problem of redundancy in standard INUS conditionality and the optimization of decision tables as discussed in Chapter 3. Essentially, the truth table presented above is a decision table with conditions A, B , and C , and their negations, and actions

P and $\text{not}(P)$. Optimizing the table with the help of an algorithm like ID3 (refer to Section 3.4.2.1 for an overview), will eliminate the above mentioned redundancy, by removing A as a condition from the decision table.

Appendix 2: Translations of the Schilling Catalog (Table 2.1)

01 Arbeitskräfte (employees)	011	Personalkosten (labor costs)	
	012	Arbeiter (male workers)	Q qualifiziert (qualified)
	013	Arbeiterinnen (fem. workers)	A angelernt (semi-skilled)
	014	Angestellte (employees)	R Routinepersonal (unskilled)
	015	Saisonschwankungen (sa), (seasonal fluctuations) Heimarbeit (h), (homework) Schichtbetrieb (sb) (shift work)	
02 Grundstücke und Gebäude (building plots and buildings)	021	Flächenbedarf (area need)	m ausschließlich Menge (quantity only)
	022	Bauliche Investitionen, Nutzungskosten (Construction investment, operating costs)	
	023	Erschließung (access to public utilities)	
03 Maschinelle Anlagen (machinery)			
04 Finanzierung (financing)			
05 Roh- und Hilfsstoffe (raw materials and auxiliary materials)	051	Rohstoffe (raw materials)	
	052	Hilfsstoffe, fertig bezogene Teile (auxiliary materials, ready-made parts)	
06 Energie (energy)	061	Kohle (coal)	
	062	Heizöl (oil)	
	063	Strom (electricity)	
	064	Gas (natural gas)	

07	Wasserbedarf (water demand)	M Menge (quantity)	Q Qualität (quality)	
08	Örtliche Kontakte (local contacts)	081	Lieferanten (L), (suppliers) Abnehmer (A), (clients), gleichartige Betriebe (G) (similar companies)	
		082	Zentralörtliche Dienste (centralized services)	p persönliche Dienste (personal services)
09	Verkehr (transportation)	091	Personenverkehrs- und Nachrichtenverbindungen (personnel transport and communications)	
		092	Güterverkehr (commodity transport)	v Transportvolumen (transport volume)
			0921 Verkehrsintensität (transport intensity)	
			0922 Straße (streets and roads)	
			0923 Bahn (railroads)	g Gleisanschluß (rail access)
			0924 Wasserverkehr (w) (water transport) Luftverkehr (l) (air transport)	
10	Immissionen (Emissions)	101	Betrieb→Umgebung (plant→environment)	e Erschütterung (disturbance) g Gefährlichkeit (danger)
		102	Umgebung→Betrieb (environment →plant)	l Lärm (noise) v Luftverunreinigung (air pollution)
11	Produktionsrück- stände (waste)	111	Abwasser M Menge (waste water) (quantity)	Q Qualität (quality)
		112	Sonstige Rückstände (other wastes)	b beseitigbar (removable) v verwertbar ((re)usable)
12	Absatz (volume of sales)			

01 Stone and Ceramics Industry	
010100	Natursteinindustrie (natural stone industry)
010200	Sand- und Kiesindustrie (sand and gravel industry)
010300	Kalkindustrie; Gips- und Kreideindustrie (limestone industry; gypsum and chalk industry)
010400	Rohton- und Kaolinindustrie (clay and kaolin industry)
010500	Ziegelindustrie (brick manufacturing industry)
010600	Zementindustrie (cement industry)
010700	Betonsteinindustrie (concrete industry)
010800	Steinzeugindustrie; Schamotte- und Silikaindustrie (building-stone industry; chamotte and silica industry)
010900	Feinkeramische Industrie (fine ceramics industry)

Standortfaktoren (location factors)		
Bewertung: (scoring)		
absolut (absolute)	relativ (relative)	Zusatz (additional)
●	■	z.B.: <u>M</u> sehr große Bedeutung (eg.) (very important)
◐	◑	<u>M</u> große Bedeutung (important)
◒	◓	M geringe Bedeutung (low importance)
○	□	unbedeutend (unimportant)
Sonstige Aussagen: (other qualifications)		z.B.: m (ohne Bewertung) (no scoring)
Allgemein wichtig: (generally important)		*

Translation of symbol explanation in Table 2.2.

For location (zur Industrieansiedlung)	Suitable (geeignet)	Suitable with restrictions (mit Vorbehalt geeignet)
Immediately (sofort)	■	▲
After extension of the infrastructure (nach Ausbau der Infrastruktur)	□	△

Appendix 3:

The Prolog programming language

Prolog is a (first-order) predicate calculus-oriented programming language. This leaves room for at least two ways of viewing Prolog: as a first-order predicate language (e.g., Gallaire *et al.*, 1984; van der Gaag and Lucas, 1988; Bouma, 1987), and as a programming language (e.g., Sterling and Shapiro, 1986; Clocksin and Mellish, 1984; Bratko, 1986; Schotel, 1987).

Clocksin and Mellish (1984), Bouma (1987), Lucas and van der Gaag (1986, 1988), and many others, provide overviews of the relations between first-order predicate logic (FOPC) and Prolog. Bouma argues that Prolog can be seen as a very useful attempt at implementing first-order predicate logic.

A Skolem normal form first-order formula is as follows:

$$\text{ALL}x_1, \dots, \text{ALL}x_n(L_{1,1} \text{ OR } L_{1,2}) \text{ AND } (L_{2,1} \text{ OR } L_{2,2})$$

All variables are universally quantified and $L_{i,j}$ is either an atomic formula or its negation. In other words: existential quantifiers are eliminated, and universal quantifiers are in front of a purely disjunctive and conjunctive formula. A clause is a conjunctive term of the Skolem function ($L_{1,1} \text{ OR } L_{1,2}$). This clause can be alternatively written as:

$$A_1 \text{ AND } A_2 \rightarrow B_1 \text{ OR } B_2$$

with A_i as the negations of the atoms under L_j , and B_k as the atoms (universal quantification assumed). A Horn clause is a Skolem function containing only one B . A logical program is a set of these Horn clauses:

$$\begin{aligned} B_1 &\leftarrow A_{1,1} \text{ AND } A_{1,2} \text{ AND } \dots \text{ AND } A_{1,n_1} \\ B_2 &\leftarrow A_{2,1} \text{ AND } A_{2,2} \text{ AND } \dots \text{ AND } A_{2,n_2} \\ &\cdot \\ &\cdot \\ &\cdot \\ B_k &\leftarrow A_{k,1} \text{ AND } A_{k,2} \text{ AND } \dots \text{ AND } A_{k,n_k} \end{aligned}$$

Definite predicate logic is a kind of part-wise predicate logic in which the indecision in standard FOPC is avoided by the closed-world assumption. This assumption implies that there is no possible domain interpretation except from the ones implicitly given by the set of all possible terms to be

formed from the constants and functions of a logical program P. In other words, all formulae that can be derived on the basis of a given set of facts and Horn clauses are considered true. Put differently, the negation of all possible wff's that are not present, is assumed. The universe is given. All decisions as to the derivation of formulae is relative to this universe. All models of the formula are therefore implicitly and *a priori* known.

A Prolog program is a logic program. Its execution consists of the loading of a data base of facts and Horn clauses, and a request for the derivation of any kind of Horn clause. A derivation may be found while some or all of the variables occurring in the clauses get instantiated to values (interpreted).

Prolog is therefore both a language and a derivation mechanism. The syntactics provide means to build a data base of Horn clauses, while Prolog's theorem prover can be asked to see if a new goal clause can be derived from the already available ones. If clauses can be derived, Prolog answers with 'yes', otherwise Prolog answers with 'no'. The procedure by which Prolog tries to prove (or tries to deny the negation of) a clause can be viewed in different ways. The FOPC point of view on that method is that of resolution (Sinth, 1985), a technique for proving the inconsistency of FOPC statements.

Programming in Prolog

Just as decision making relates to logic, so are programming problems of a logical nature a nice consequence of a language like Prolog. Programming in Prolog can be seen as declaring things as true, and asking about other things whether they are consistent with the declared part. Such a declaration of true statements can look something like this (Note that the ':' sign means \leftarrow . The comma separates the arguments in predicates, and means 'AND' when it separates predicates).

- (0) `building(empire_state).`
- (1) `building(my_house).`
- (2) `building(world_trade_center).`
- (3) `height(empire_state,381).`
- (4) `height(world_trade_center,411).`
- (5) `height(my_house,10).`
- (6) `geographer(humboldt).`
- (7) `geographer(hettner).`

- (8) `likes(X,phone_booth) :- geographer(X).`

```

(9) likes(X,Y) :- geographer(X),
                  high_building(Y).
(10) likes(hettner,Y) :- building(Y).

(11) high_building(X) :- building(X),
                        height(X,H),
                        H > 100.

```

All but the 'high_building' and 'likes' clauses are facts (not premises). These facts consist of a functor (predicates such as 'likes', 'high_building', 'building', and 'height'), and arguments (such as 'geographer', 'empire_state', and '10'). The conditional statement clause on high buildings consists of a head, separated from the body by the 'reversed implication sign (:-)'. The implication sign, however, is just a special kind of predicate. The clause could therefore also be written as:

```

' :- ' (high_building(X), building(X), height(X,H), H > 100) .

```

Capitals stand for variables. All statements are considered universally quantified.

If this small data base is read into the Prolog system, one can start asking Prolog questions such as whether my_house is liked by humboldt:

```

:- likes(humboldt, my_house) .

```

Prolog will search its data base for possibilities to prove this. It will search through the data base from top to bottom. It tries to find a clause that matches the goal clause. It finds clause (9) as the matching clause and unifies the variable X with the constant 'humboldt' and variable Y with the constant my_house. Note that the goal clause does not match with clause (8) because of the incompatibility of the constants 'phone_booth' and 'my_house'.

But clause (9) is a conditional clause. It can only be considered true if it can be proved that geographer(X) and that high_building(Y), hence that geographer(humboldt), and high_building(my_house). That humboldt is a geographer can be proved because the data base contains a fact (6) declaring this. But that my_house is a high_building is something that can only be inferred from rule (11). The new goal clause 'high_building(my_house)', thus matches with clause (11), unifying variable X with 'my_house'. Again this is a conditional clause. By (1) it is proved that building(my_house). H gets

instantiated to '10' by means of (5), but the test 'H > 100' fails, because 10 is not larger than 100. The result of all this is that the attempt to prove that humboldt likes my_house fails, and Prolog returns with the answer 'no'.

Had one asked whether hettner likes my_house, the answer would have been 'yes'. Again Prolog would have tried the proof as for humboldt, something which again would have failed. But on backtracking from the nested sub-proofs, it would find another possibility for proving that hettner likes something, namely rule (10). This time it needs only to be proved that my_house is a building, something which works because of (1). Clearly, we can avoid the useless attempt to prove that hettner likes my house by means of (9), by changing the order of clauses (9) and (10). This does not hurt the logic in the data base because of the constant 'hettner' in clause (9).

That the order of clauses is important shows that Prolog is not a 'pure' logical language. It does contain a number of procedural elements, something which may pose a problem for certain kinds of applications.

The only thing Prolog does is to try to prove a clause. However, when using variables rather than constants in the main goal clause, the proof of the clause might imply the instantiation of those variables by constants (either atoms, or a list, or another clause). For example, given the above mentioned data base, the request for proof

```
likes(X,Y).
```

will generate the following proofs:

```
X = humboldt
Y = phone_booth
```

```
X = hettner
Y = phone_booth
```

```
X = humboldt
Y = empire_state
```

```
X = humboldt
Y = world_trade_center
```

```
X = hettner
Y = empire_state
```

```
X = hettner
Y = world_trade_center
```

```
X = hettner
Y = empire state
```

```
X = hettner
Y = my_house
```

```
X = hettner
Y = world_trade_center
```

Apparently, backtracking caused Prolog to find multiple proofs for both the facts that hettner likes the empire state building and that hettner likes the world trade center. The reason is of course that these can be proven by either using rule (9) or rule (10). Prolog does provide means to avoid multiple proofs of clauses, a topic extensively covered by many standard Prolog textbooks.

In Prolog one can easily work with lists. A list is a sequence of elements. A Prolog list consists of a head (the first element of a list), and a tail (the rest of the list). The tail by itself is again a list that has a head and a tail, so that one can perform recursive actions on lists. A list consisting of head H and tail T is written as [H|T]. The only list that has neither a head nor tail is the empty list ([]). A list with only one element has the empty list as its tail: [X] = [X|[]]. One can use these characteristics of a list to define in Prolog what a list actually is:

```
is_list([]).
is_list([_|_]).
```

The first clause declares the empty list as a list. The second clause declares anything that matches the form of a list, i.e., something with a head and tail, as a list. The underscore signs stand for anonymous variables.

An example of combining list characteristics with a recursive definition is a Prolog function for list membership 'member(Member,List)' saying that Member is a member of list List.

```
member(H, [H|_]).
member(X, [_|T]) :- member(X,T).
```

The first clause declares that something is a member of a list if it is the head of the list. The second clause covers all other cases by saying that X is a member of a list if it is a member of its tail.

Prolog implementations also provide so-called built-in predicates for analyzing and building Prolog clauses. A program can assert new clauses to the data base or retract existing ones, thus modifying the data base, something which can influence the execution of the program itself. This again shows that in Prolog there is no real difference between program and data. Prolog just tries to prove the things it is asked. If this proof requires modification of its set of clauses, then this is done. In this way one can write self-modifying programs, and programs that 'know' about their own execution.

Appendix 4: Hardware and software

The Shanxi Province DSS was implemented at IIASA-ACA between July, 1986-June, 1988, on a SUN 3/260 C (color) workstation with graphics coprocessor and 16 megabytes of main memory, running the Sun 3.4 operating system. Resolution on the 19-inch color monitor is 1152 x 900 pixels. The various models were implemented in a variety of programming languages such as C, FORTRAN, Lisp and Prolog.

REPLACE was written in Quintus Prolog 2.0 and C. Prolog was used to implement the core model (the inference tree(s)), the data bases, the inference engine, and some of the additional functions such as the explanation routines and the dimensionality selector. Typical procedural components such as the selection of counties and activities, and the display of the spatial and statistical distribution of variables, were implemented in C. Because both the data bases and the system control were in Prolog, C routines are called from Prolog by means of the Quintus Prolog Foreign Function Interface.

That the data bases of REPLACE are in Prolog also implies that in some cases routines show a high degree of interaction between Prolog and C. Examples of these are the various editors, used to modify the county or activities data bases. The technique used in REPLACE is that Prolog calls the C routines with specific parameter values, so that the C routines know what actions to perform. C, in its turn, returns to Prolog with specific values, so that Prolog knows what happened, e.g., which menu option a user selected. Prolog then processes this information, for example by updating its data base, and then calls the C routine again. To exit an editor, for instance, C returns a flag to Prolog indicating that the user wants the editing to be terminated. Prolog interprets the flag, and calls the editor with a 'destruct' flag, instructing the editor to self-destruct. The editor terminates itself and returns to Prolog with a flag saying that it has terminated, upon which Prolog can return to the level from where the editor was called in the first place, and continue its 'normal' process.

The graphics of REPLACE, like the graphics of the other modules in the Shanxi Province DSS, were implemented in Sun Core and Sun Pixrect, and called from C programs, or via the QP foreign function interface, directly from Prolog.

