IIASA
International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-98-106/December

Identification of Biodiversity and Other Forest Attributes for Sustainable Forest Management: Siberian Forest Case Study

Szymon Wilk (szymon.wilk@cs.put.poznan.pl)
Matti Flinkman (matti.flinkman@stat.umu.se)
Wojtek Michalowski (wojtek@business.carleton.ca)
Sten Nilsson (nilsson@iiasa.ac.at)
Roman Slowinski (slowinsk@sol.put.poznan.pl)
Robert Susmaga (robert.susmaga@cs.put.poznan.pl)

Approved by
Gordon J. MacDonald (macdon@iiasa.ac.at)
Director, *IIASA*

Contents

Abstract

This paper attempts to identify characteristics for biodiversity and other (forest) ecosystem conditions that are considered essential for a description of ecosystem functioning and development of sustainable forest management practices in the Siberian forests. This is accomplished through an analysis of net primary production of phytomass (NPP) which acts as a proxy for ecosystem functioning. Rough Sets (RS) analysis is applied to study the Siberian ecoregions classified into compact and cohesive NPP performance classes. Through a heuristic procedure, a reduced set of attributes is generated for a NPP classification problem. In order to interpret relationships between various forest characteristics, so-called *interesting rules* are generated on a basis of reduced problem description. These *interesting rules* provide means to draw conclusions in the form of knowledge statements about functioning of the Siberian forests.

Identification of Biodiversity and Other Forest Attributes for Sustainable Forest Management: Siberian Forest Case Study

Szymon Wilk, Matti Flinkman, Wojtek Michalowski, Sten Nilsson, Roman Slowinski and Robert Susmaga

1. Problem Statement

In the context of developing sustainable forest management practices in the boreal forest zone a key issues are:

a) to identify and evaluate in a holistic manner the current and desirable state of ecosystem functioning essential for (human) life;

b) to study the impact of alternative forest management regimes on the ecosystem functioning.

The present paper deals primarily with holistic analysis of ecosystem functioning in the Russian Siberia using comprehensive database maintained at the International Institute for Applied Systems Analysis (IIASA). Knowledge established through such an analysis will in turn form a platform for further work on the development of sustainable management practices.

The ongoing debate on biological diversity constitutes a convenient starting point for the study of ecosystem functioning. The concept of biological diversity (biodiversity) of the natural resources has become a worldwide concern as exemplified by the discussion at the 1992 UN Conference on Environment and Development (UNCED, 1992). Since that conference "biodiversity" has become one of the main topics on the agenda of a number of international institutions and non governmental organizations dealing with environmental and forestry issues. However, despite that interest, the concept of "biodiversity" is still in its infancy in both forestry and environmental debates. Consequently, there exists neither a widely accepted definition of "biodiversity" nor any agreement regarding its context (Duinker *et al.*, 1996).

In general, there are two approaches to the studies of biodiversity. The first approach, originating from the Convention on Biological Diversity (UNCED, 1992), is focused on the "diversity of populations" at various levels, including genetic diversity, species diversity and ecosystem diversity. A related issue of interest concerns the description of the “degree” of biodiversity within these levels. There is, however, ambiguity regarding specific ecosystem boundaries and the use of appropriate scale to address biodiversity descriptions and analysis (Noss, 1990). In this context, the “landscape approach” has been proposed as addressing some of the above concerns, but the question remains of what kind of explanatory framework should be used, especially when considering biodiversity in relation to other ecosystem functions.

In the second approach which originated from the Statement of Principles on Sustainable Forest Management (UNCED, 1992) the concept of biodiversity as described above is used in conjunction with the development of sustainable and conservation-based forest management principles. This approach advocates taking a broader perspective, including ecological, economic, socio-cultural and related socio-economic aspects of forestry. The following theme areas of analysis are considered within this approach:

- Global carbon cycles,
- Health and vitality,
- Wood and non-wood productive functions,
- Biological diversity,
- Protective functions as regard soils and waters,
- Socio-economic functions and conditions.

Specific criteria and indicators are being developed for each of the theme areas. This seemingly holistic review of aspects related to different functions of forest ecosystem has a principal shortcoming, namely that the theme areas, and performance indicators, are treated in isolation instead of in a holistic manner (Nilsson, 1997a).

Drawing partly on the second approach it is emphasized in this paper that the theme areas and the inter-linkages between them should be considered simultaneously in the analysis of various functions of an ecosystem. Thereby, *ecosystem functioning* ought to be used as a core concept, implying that the appropriate and desirable functioning of all theme areas is paramount for the support of ecosystem services[1] and for overall understanding of the consequences of natural or human made changes within a specific theme. Such view should provide a comprehensive framework for incorporating the theme areas and the linkages between them into a study of (bio) diversity.

Accordingly, a general framework for identification of (bio) diversity and other forest attributes in this Siberian forest case study is built on the premises that a possible impact of descriptive attributes identified within different theme areas should be scrutinized within a core concept of ecosystem functioning. The explanatory attributes to be chosen among abiotic, biotic factors and factors indicating human impact should thus describe a

[1] Delivery of ecosystem services involves (Cairns, 1997): (1) Capture of solar energy and conversion into biomass that is used for food, building materials and fuels, (2) Breakdown of organic wastes and storage of heavy metals, (3) Maintenance of gas balance in the atmosphere that supports human life: absorption and storage of carbon dioxide and release of oxygen for breathable air, (4) Regeneration of nutrients in form essential to plant growth, e.g. nitrogen fixation and movement of those nutrients.

pallet of structures as well as interactions between land-uses, vegetation types, forest density, site-class, age, and different aspects of human activities.

The data component of this study is described in section 2. The data set contains information on a number of attributes recorded at ecoregion level. Due to a significant number of possible attributes to be considered, one stands in front of a complex decision problem while selecting appropriate ones. Following the idea of a holistic approach, it is necessary to consider different cross-classifications, reflecting different roles of attributes describing various conditions. This task will be accomplished through an amalgamation of the rough sets (RS) analysis with a heuristic evaluation of possible sets of attributes that guarantee similar descriptive accuracy. The principles for RS analysis are given in section 3 and in Appendix 2. An implementation of the RS methodology on the data set derived from the Siberian forest database is also described in section 3. Section 4 presents discussion of the results.

2. Siberian Forest Database

The Siberian forest database contains information pertaining to the cornerstone areas of the Sustainable Boreal Forest Resources Project at IIASA (Nilsson, 1997b). Nearly 5000 attributes describing abiotic, biotic and human induced conditions are included in the database. The spatial coverage of the gathered information is aggregated at different levels. The highest level covers the whole of Siberia. Sub-levels are for 65 administrative regions, 65 ecological regions (ecoregions), 360 landscapes and 2500 forestry enterprises. All database items can be related to some spatial aggregation level that allows spatial descriptions of abiotic, biotic and anthropogenic conditions.

As the analysis should allow for description of the linkages between theme areas and different diversity aspects for ecoregions, it seems appropriate to use distributions of various attributes instead of single mean values, as they are considered to provide better descriptive information. Therefore, for a purpose of this study a smaller data set was derived from Siberian forest database (see Appendix 1). This data set contains a sample of original abiotic and biotic attributes and attributes for human induced conditions. Also, for each ecoregion a number of modified attributes, so-called CODE-descriptors, describing the structure of certain distributions (such as for example, age distribution of forested area) have been developed. In creating the CODE-descriptor, the original data on distributions has been categorized into few (4-7) share classes giving opportunity to create a number of distribution “profiles”. Moreover, for illustration and assessment of diversity aspects so-called SHDI-descriptors were also included. The calculation of SHDI-descriptors is based on Shannon diversity index formula (Shannon and Weaver, 1962). The SHDI-descriptor illustrates the degree of diversity of the attribute being considered. The actual distribution of values for an attribute with few dominating classes generates low diversity value of the SHDI-descriptor, while an evenly distributed share is coded as a high value.

3. The Net Phytomass Production Classification Problem

The goal of the current Siberian forest case study is to identify characteristics for (bio) diversity and other (forest) ecosystem conditions important for ecosystem functioning. It is not a straightforward exercise to create a cohesive description of each ecoregion while analyzing a magnitude of factors related to ecosystem functioning. This is so because there is a significant number of attributes that might be considered as the candidates for such a description. Methodological considerations associated with the analysis of a data set are presented in section 3.1. In order to create a compact and cohesive description of ecosystem functioning, we apply the RS analysis to identify a smaller subset of the attributes that need to be evaluated. Following the RS principles we focus on the identification of specific subsets of the attributes and in section 3.2 we describe a heuristic procedure which we used to generate a “good” subset (so-called “good” reduct) for a given classification. Identification of a “good” reduct allows achieving a significant reduction of the number of attributes that are to be considered further.

An important aspect of any policy analysis is related to an explanation of the relationships between problem components. One of the best vehicles for conveying such an information is provided by the decision rules being logical statements of the type “*if…, then…*”. We use them in our study, and we generate so-called *interesting* rules for each of the “good” reducts (the general principles for generating rules are described in Appendix 2 section 3). The *interesting* rules provide helpful explanation of the role of attributes and the significance of their specific values, and allow drawing conclusions in terms of knowledge statements.

The idea of the study regarding the assessment of diversity aspects in conjunctions with other important functions of forest ecosystem is assumed to be captured by the net primary production of phytomass (NPP) of an ecoregion[2]. With that on mind, the ecoregions were classified into performance categories following their NPP, and the RS analysis (Pawlak, 1991; Slowinski, 1992) was applied to draw the conclusions from such a classification. This part of the study is reported in Section 3.3.

3.1. Methodological Considerations

The methodology used to analyze the relationships among the attributes describing ecosystem functioning of the Siberian forests is based on the RS theory. This theory was first proposed by Pawlak, (1991) to study classification problems. From the point of view of usefulness of results obtained by the basic RS analysis, it is advised to consider symbolic (qualitative) data rather than continuous-valued (quantitative) information. In the later case, domains of continuous-valued attributes should be discretized (categorized) prior to the analysis (see Appendix 2 section 4). The data set being considered consists of *objects* (also *examples*, *cases*) which here represent Siberian

---

[2] The net primary production of phytomass (NPP) is an estimated measure of an ecoregion’s total production potential of phytomass in t/ha/year according to Bazilevich (1993). The NPP measure comprises all land uses, including agriculture land, within an ecoregion.

ecoregions described by discrete values of the *attributes*, representing characteristics of these ecoregions. The set of attributes is usually divided into two disjoint subsets, called *condition* and *decision* attributes. The important distinction between these two sets is that the condition attributes express some descriptive information about the ecoregions (objects), while the decision attributes express some decisions or classifications made about the ecoregions. The set of ecoregions (objects) described by attributes and represented in a table format is called *a decision table*. This table is further analyzed from a point of view of reduction of a number of condition attributes while maintaining at the same time a quality of approximation of the original set[3].

It is important to stress that the NPP classification problem considered in this paper could also be analyzed using statistical methods based on discriminant analysis. The main goal of discriminant analysis is to create functions, which can be later used to assign a given ecoregion to one of the predefined classes depending on the scores of these functions associated with the classes. Moreover, discriminant analysis can also be used to reduce the number of attributes (features) and to select the most important ones.

Most of the discriminant analysis methods are applicable only for continuous-valued attributes. Only a few (for example, so-called "location model approach" (Krzanowski, 1980) can deal with the mixture of continuous-valued and symbolic attributes.

Considering the properties of the Siberian data set and the comprehensibility of the generated output, the RS theory has several advantages over discriminant analysis (Stefanowski, 1992):

a) discriminant analysis methods are very demanding regarding the quality of the input data - normal distribution of continuous attributes is assumed and considered classes should contain comparable number of objects (both of these requirements are not satisfied for the Siberian forest data), while such a requirement needs not to be satisfied for the RS.

b) in the location model approach all qualitative attributes have to be transformed into the binary ones; when qualitative attributes have large number of values (this is the case of all CODE-descriptors attributes), the resulting number of attributes increases rapidly; in practical applications it is suggested that the number of binary attributes does not exceed six (Krzanowski, 1983) while in the analyzed data set there are 22 qualitative attributes; the RS methodology does not neither require any transformation of qualitative attributes nor limit their number.

c) the methods of discriminant analysis generate the final result in a form of discriminant functions, which are aggregating the input information in a non transparent way; methods based on the RS theory produce decision rules that are much more readable than the discriminant functions and can be easily interpreted by a prospective user.

The above arguments support our choice of the RS for the analysis of the Siberian forest database and are further strengthen by the results of this analysis.

---

[3] Basic notions of the RS theory are described in greater detail in Appendix 2.

3.2. Procedure to Identify a "Good" Reduct

Applications of the RS analysis to the selected subsets of the Siberian forest database resulted in generation of empty cores and large numbers of reducts. Therefore, in order to identify some "good" reduct as a final result of the analysis, one needs to resort to a heuristic procedure based on the notion of β-core and β-reducts. The procedure used in this study is given below.

1. Generate all existing reducts.
2. For every conditional attribute calculate its relative frequency of occurrence in the reducts.
3. Establish the threshold $\beta \in [0, 1]$ (this may be done, for example, after analyzing the histogram of relative frequencies), and establish the β-core, i.e. select those attributes whose frequencies (calculated in 2) are not lower than β.
4. Find all β-reducts, i.e. reducts that include the β-core (if such reduct does not exist, modify the β-core by dropping the attribute with smallest relative frequency and repeat this step).
5. Using the β-reducts identified in 4, find those which have the smallest cardinality (i.e. the smallest number of attributes). If there is only one such reduct, then it is the "good" one. Otherwise go to the next step.
6. For each reduct identified in 5 test its ability of constructing an accurate classifier representing the data set in terms of decision rules[4]. Identify a reduct with the best result of that test and select it as the "good" one.

The above procedure strictly follows the purpose of our study, which is to find a smallest subset of the attributes that guarantees no drop of the discrimination degree in the Siberian forest data set. Moreover, characterization of the attributes through the frequencies of occurrence in the reducts is motivated by the search for attributes that resemble as close as possible the indispensable attributes belonging to the classical RS core (which has $\beta = 1$).

3.3. Ecosystem Functioning

The working hypothesis for this study is that the classification of ecoregions into different classes according to the NPP in t/ha/year reflects various uses of land and biogeophysical conditions (Shvidenko, *et al.*, 1997). Thereby, such classification will capture a number of factors assumed to be associated with different levels of ecosystem

---

[4] This test is called cross-validation (CV) test (See Appendix 2 section 3). Presence of "noise" in data suggests giving priority to a self-test while selecting the attributes for further analysis. At the same time, due to a big variance associated with the CV test results, cognitive validity of this particular test should be downplayed and used only as a last resort. Nevertheless, discrimination among β-reducts with the smallest cardinality (step 5 of the procedure) using the results of CV test encourages to consider β-reducts with best predictive powers.

functioning. The set of condition attributes used in this classification problem consists of[5]: MOUNTAIN, PERMAFROST, AV_AIR_TEM, AV_SOIL_TE, AV_MAX_SOI, AV_MIN_SOI, TOT_PRECIP, WIND, SUM_T10, SUM_T5, SUM_PREC10, SUM_PREC5, DURATION_1, DURATION_5, SNOW_COVER, Vext-SHDI, FA/Area, FF-CODE, FF-SHDI, BON-CODE, BON-SHDI, DENS-CODE, DENS-SHDI, AgAr-Code, AgAr-SHDI, AgVo-Code, AgVo-SHDI, POP/sqkm, Autow/sqkm, Railw/sqkm and Riverw/sqkm.

3.3.1 The Resulting "Good" Reduct

Each ecoregion was assigned into one of three NPP classes (L, M and H, denoting the low, medium and high classes of the NPP, respectively), according to ecoregion's potential phytomass production capacity.

Through the RS analysis an original set of 31 attributes was significantly reduced, and the following "good" reduct was identified:

- Relief conditions (MOUNTAIN)
- Snow cover conditions (SNOW_COVER)
- Share of forested area of total ecoregion area (FA/Area),
- Forest Fund profile consisting of Forest land, Non-forest land and Lease (FF-Code)
- Age profile of growing stock consisting of 5 age class categories (AgVo-Code)
- Density of railway network (Railw/sqkm)

The AgVo-Code, FA/Area and FF-Code are all forest related attributes, while MOUNTAIN and Snow_Cover can be regarded as describing biogeophysical conditions, while the Railw/sqkm can be considered as an indicator of ecoregion's development. Reduction of the original set of attributes to 6 most relevant ones, constitutes a significant improvement over other studies where such information was not available.

3.3.2 Generation of *Interesting* Rules

General knowledge statements were built using the *interesting* rules only (see Appendix 2 section 3). Table 1 presents the *interesting* rules induced from the Siberian forest database which description was reduced to contain only relevant attributes from the "good" reduct.

Each row in Table 1 represents one decision rule. Condition part of the rule is a conjunction of elementary conditions on those attributes for which values are specified (elementary condition has syntax *attribute* = *value*) and decision part reflects assignment of an ecoregion to the specified NPP class. For example, rule 7 should be read as:

if AgVo-Code equals to ABDBC and MOUNTAIN equals to 1, then NPP class is M.

---

[5] See Appendix 1 for a full list of the attribute and for an explanation of their abbreviations.

An interpretation of the same rule in a subject domain maybe following: **if** the distribution of growing stock into age classes is such that the age class “youngest forest” accounts for somewhere between 0-5%, of the growing stock; the “young forest” accounts for somewhere between 5-20%; the “middle aged forest” accounts for somewhere between 40-60%; the “immature forest” accounts for somewhere between 5-20%; and the “mature and overmature forest” accounts for somewhere between 20-40% of the growing stock and at the same time the relief conditions are mountainous, **then** the NPP class is medium.

The profile of forest fund (FF-Code) appears to be the most frequent attribute present in condition part of the *interesting* rules. Especially, this is true for a high NPP class where it appears in condition part of all decision rules. Furthermore, it also appears in combination with a number of other attributes in most of the rules for two other NPP classes.

Table 1. *Interesting* rules for the NPP classification problem[6]

Rule no.	NPP class	Elementary conditions						Relative rule strength
		AgVo-Code	FA/Area	FF-CODE	MOUNTAIN	Railw/ sqkm	SNOW COVER	
1	L	AABAF						12%
4	L		0				1	56%
5	L			ECA	1		1	12%
6	L			ECA			1	12%
7	M	ABDBC			1			10%
8	M	AABBE			2			16%
9	M	ABDBC	1					10%
10	M		1	ECA		1		13%
11	M		1	GAA		0		13%
12	M	AACBD	1			1		10%
13	M		1	FBA	2		0	10%
14	M		1		2	0	0	16%
15	M			FBA	2	0	0	10%
16	H		0	FBA	2			19%
17	H			FBA	2	1		19%
18	H		0	FBA		1		25%

Column “relative rule strength” gives percentage of the ecoregions “covered” by a given rule (i.e. those which are classified into appropriate class by this rule). While generating the *interesting* rules we used a threshold of 10% (i.e. only those rules are interesting which “cover” at least 10% of the cases (ecoregions)).

---

[6] Values 0 and 1 in columns FA/Area and Railw/sqkm indicate either first or second interval generated by *Recursive Minimal Entropy Partitioning* discretization method (Fayyad and Irani, 1993) applied for these two attributes. All other attributes were discretized according to the intervals defined by an expert.

3.3.3 Extracting Knowledge from the Rules

The analysis of the NPP classification problem suggests that a typical feature for ecoregions classified into high (H) NPP class is a low share of the landmass covered by forested area (FA/Area = 0). The existing forest fund within these ecoregions consists of mainly forest land and, to a lesser extent, non-forest land (FF-Code = FBA). Ecoregions seem to be well developed (Railw/sqkm = 1, MOUNTAIN = 2) and the climate conditions appear also to be relatively favorable for a high NPP due to their Southern location (mainly in West and South-West Siberia).

Ecoregions classified into the low (L) NPP class are characterized by mountainous and harsh climatic conditions, and consequently are difficult to access. Also, the share of non-forest land of the forest fund appears to be relatively high in these regions (FF-Code = ECA). Thus, the production of phytomass is based, to a large extent, on growing potential outside of the forests which is also confirmed by the low share of forested area identified for those ecoregions (FA/Area = 0). Otherwise, in cases where the conditions described above are not applicable, the low NPP can be associated with uneven distribution of the growing stock into different age classes (AgVo-Code = AABAF). A share of growing stock in the mature and overmature age class is clearly dominant over other age classes, i.e. the forested area of such an ecoregion is approaching the “climax” stage of its development cycle.

The ecoregions classified into medium (M) NPP class represent “forested” regions because the forest cover of the total ecoregion area is clearly predominant (FA/Area = 1). This is also supported by the forest fund consisting, to a large extent, of forest land (FF-Code = ECA or GAA). In addition, the age class distribution of growing stock within forested area (AgVo-Code = ABDBC or AABBE) obviously indicates a certain degree of utilization of the forest resources or possible “natural management” through disturbances like fires and insect attacks which have brought down the volume of the “old growth”. So, from the point of view of the NPP the current distribution of growing stock is regarded as more desirable than, for example for the low class.

In conclusion, the ecosystem functioning or the delivery of ecosystem services in the ecoregions belonging to low and high NPP class is to a large extent dependent on other than forests (forested areas) life forms. At the same time, the ecoregions classified into medium NPP classes are predominantly characterized by forests that appear to play a crucial role in supplying ecosystem services.

Future analysis should identify and clarify the actual stage of these services, for example actual NPP of the entire ecoregion or its forests, and also possible future actions in order to improve the performance of the ecosystems in terms of desirable supply of the services. From the point of view of forestry and forest management practices the interest should be focused first on ecoregions belonging to medium class NPP. The findings of our study so far confirm the importance of forests for ecosystem functioning in these ecoregions, which in turn implies considerable potential for the implementation of desirable forest management policies.

4. Discussion

We evaluated the ecosystem functioning from the point of view of the NPP, what required incorporation of several descriptive aspects of commonly considered factors essential for sustainable functioning of ecosystems. Such a problem area should be studied in a holistic manner and with understanding that ecosystems constitute complex structures.

Analysis of complex situations, characterized by many decision attributes of different character and different level of detail calls for methodology that allows simplifying a problem in terms of its descriptive requirements. In a case of the Siberian forest database, the RS methodology enhanced with the procedure for identification of “good” reducts, enabled to develop reduced and compact description for the classification problem. Creation of such a compact description has advantages from a data mining perspective, as it requires less information to be collected and accessed, and it also facilitates analysis of data dependencies. Generation of the *interesting* rules demonstrated that it is possible to identify certain commonalties for the ecoregions belonging to the same class. We attempted to translate these commonalties into general knowledge statements. A promising aspect which emerged while creating these statements is that the regularities discovered in the Siberian forests are in line with forces shaping ecosystem functions in other boreal regions outside Siberia.

One of the principal issues related to studying forest ecosystems is a necessity to consider several aspects of the problem, as exemplified by the appropriate theme areas. In order to address this issue, one needs to evaluate a set of diversified attributes. In the study we managed to accomplish this and, moreover, to identify the relations between specific attributes coming from different theme areas. Our findings should help in further forest studies to focus on those specific aspects of theme areas that are deemed as important and thus create a basis for the interpretation of the results for sustainable forest management policies.

In future research, some recent extensions of the RS methodology could be used for a more detailed Siberian forest data set. In particular, the extensions concerning the attributes with preference ordered domains and the approximation of classes by dominance relations instead of the classic indiscernibility relation (see Greco *et al.*, 1999) should prove useful for this kind of analysis.

Acknowledgement

Slowinski, Susmaga and Wilk wish to acknowledge financial support from the State Committee for Scientific Research, KBN, and from CRIT 2 Esprit Project. Research conducted by Michalowski was supported by the Decision Analysis and Support (DAS) Project of IIASA while he was a senior research scholar at the Institute, and partially supplemented by a research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Bazilevich, N.I. (1993). “Bilogical Productivity of Ecosystems of Northern Eurasia”. Nauka, Moscow. In Russian.

Cairns, J. Jr. (1997). Protecting the delivery of Ecosystem Services, *Ecosystem Health*. Vol.3 No.3, September 1997.

Dougherty, J., Kohavi, R. and Sahami, M. (1995). “Supervised and Unsupervised Discretizations of Continuous Features”. In: *Proceedings of the 12th International Conference on Machine Learning*, Morgan Kaufmann Publishers, New York, pp. 194-202.

Duinker, P. (*ed.*) Carlsson, M. Gluck, M. Plinte R. and I. Venevskaia, I. (1996). “Biodiversity of Siberian Forests: Concepts, Preliminary Analysis, and Proposed Research Directions.” WP-96-79. International Institute for Applied Systems Analysis. Laxenburg.

Fayyad, U.M. and Irani, K.B. (1993). “Multi-interval discretization of continuous-valued attributes for classification learning”. In: *Proceedings of the 13th International Conference on Artificial Intelligence*, Morgan Kaufmann, pp. 1022-1027.

Greco, S., Matarazzo, B. and Slowinski, R. (1999). “The Use of Rough Sets and Fuzzy Sets in MCDM”. Chapter 14 in: Gal, T., Hanne, T. and Stewart, T. (eds.), *Advances in Multiple-Criteria Decision Making*, Kluwer Academic Publishers, Boston, pp. 14.1-14.59.

Grzymala-Busse, J.W. (1992). „LERS – “A system for learning from examples based on rough sets”. In: Slowinski, R. (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*. Kluwer Academic Publishers, Dordrecht, pp. 3-18.

Krzanowski, W.J. (1980) “Mixtures of Continuous and Categorical Variables in Discriminant Analysis”. *Biometrics* No. 36, pp. 493-499.

Krzanowski, W.J. (1983) “Stepwise Location Model Choice in Mixed-variable Discrimination”. *Applied Statistics* No. 32, pp. 260-266.

Mienko, R., Stefanowski, J., K. Tuomi, D. Vanderpooten. (1996). Discovery-oriented Induction of Decision Rules. *Cahier du LAMSADE*, No. 141.

Nguyen, H.S. (1998). “Discretization problems for rough set methods.” In: Polkowski, L. and Skowron, A. (eds.), *Lecture Notes in Artificial Intelligence 1414. Rough Sets and Current Trends in Computing. (Proceedings of RSCTC’98)*, Springer Verlag, Berlin-Heidelberg, pp. 545-552.

Nilsson, S. (1997a). Challenges for the Boreal Forest Zone and IBFRA. A Key Note Paper for IBFRA Conference.

Nilsson, S. (ed.) (1997b). Dialogue on Sustainable Development of the Russian Forest Sector - Volume II, IIASA Interim Report (IR-97-010), Laxenburg, Austria: International Institute for Applied Systems Analysis.

Noss, R.F. (1990). “Indicators for Monitoring Biodiversity. A Hierarchical Approach”. *Conservation Biology* No. 4, pp. 355-363.

Pawlak Z. (1991). *Rough Sets. Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, Dordrecht.

Predki B., Slowinski R., Stefanowski J., Susmaga, R., Wilk, Sz. (1998). “ROSE – software implementation of the rough set theory”. In: Polkowski, L. and Skowron, A. (eds.), *Lecture Notes in Artificial Intelligence 1414. Rough Sets and Current Trends in Computing. (Proceedings of RSCTC’98),* Springer Verlag, Berlin-Heidelberg, pp. 605-608.

Shannon, C.E and Weaver, W. (1962). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

Shvidenko, A., S. Nilsson, Roshov, V. (1997). Possibilities for Increased Carbon Sequestration through the Implementation of Rational Forest Management in Russia. *Water, Air and Soil Pollution* No. 94, pp. 137-162.

Slowinski, R. (ed.) (1992). *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*. Kluwer Academic Publishers, Dordrecht.

Slowinski, R. and Stefanowski, J. (1994). “Rough Set Reasoning about Uncertain Data”. *Fundamenta Informaticae*, Vol. 27, No 2/3, pp. 229- 244.

Stefanowski, J. (1992) “Rough Sets Theory and Discriminant Methods as Tools for Analysis of Information Systems. A Comparative Study”. *Foundations of Computing and Decision Sciences* No. 2, pp. 81-98.

Susmaga, R. (1997). Analysing Discretizations of Continuous Attributes Given a Monotonic Discrimination Function. *Intelligent Data Analysis 1*, no. 3.

UNCED, (1992). *The Global Partnership for Environment and Development*. United Nations, Geneva.

Appendix 1

List of attributes used in the study

Attribute Name	Description
PhyProClass	Net Primary Production classes of Phytomass
MOUNTAIN	Relief conditions: mountain, plain or far east mountain
PERMAFROST	Permafrost: year round, seasonally or no frozen ground
AV_AIR_TEM	Average air temperature
AV_SOIL_TE	Average soil surface temperature
AV_MAX_SOI	Average Max soil surface temperature
AV_MIN_SOI	Average Min soil surface temperature
TOT_PRECIP	Average total precipitation
WIND	Average wind speed
SUM_T10	Total number of days during the growing season with average temperature above 10^{o}C
SUM_T5	Total number of days during the growing season with average temperature above 5^{o}C
SUM_PREC10	Total precipitation during the growing season for days with average temperature above 10^{o}C
SUM_PREC5	Total precipitation during the growing season for days with average temperature above 5^{o}C
DURATION_1	Duration of vegetation period where average temperature is above 10^{o}C
DURATION_5	Duration of vegetation period where average temperature is above 5^{o}C
SNOW_COVER	Duration of snow cover
Vext-SHDI	Shannon diversity index for vegetation types
FA/Area	Forested area of total ecoregion area in %
FF-CODE	Forest fund profile distributed by forest land, non-forest land, and lease
FF-SHDI	Shannon diversity index for forest fund profile
New NFL-Code	Non-forest land profile distributed by arable land, hayfield, pasture; water, sand, glacier, other; garden, houses; roads; and bogs
New NFL-SHDI	Shannon diversity index for non-forest land profile
FL-CODE	Forest land profile distributed by unforested land and forested area
FL-SHDI	Shannon diversity index for forest land profile
UNF-CODE	Unforested land profile distributed by sparse woodlands, burnt areas, clearcuts, and glades
UNF-SHDI	Shannon diversity index for unforested land profile

Attribute Name	Description
FA-CODE	Forested area profile distributed by virgin, natural, and antropogenic forests
FA-SHDI	Shannon diversity index for forested area profile
BON-CODE	Site class profile for all age classes
BON-SHDI	Shannon diversity index for site class profile of all age classes
BOYU-CODE	Site class profile for young stands
BOYU-SHDI	Shannon diversity index for site class profile of young stands
BOMI-CODE	Site class profile for middle-aged stands
BOMI-SHDI	Shannon diversity index for site class profile of middle-aged stands
BOIM-CODE	Site class profile for immature stands
BOIM-SHDI	Shannon diversity index for site class profile of immature stands
BOOV-CODE	Site class profile for mature and overmature stands
BOOV-SHDI	Shannon diversity index for site class profile of mature and overmature stands
DENS-CODE	Density class profile for all age classes
DENS-SHDI	Shannon diversity index for density class profile of all age classes
DEYO-CODE	Density class profile for young stands
DEYO-SHDI	Shannon diversity index for density class profile of young stands
DEMI-CODE	Density class profile for middle-aged stands
DEMI-SHDI	Shannon diversity index for density class profile of middle-aged stands
DEIM-CODE	Density class profile for immature stands
DEIM-SHDI	Shannon diversity index for density class profile of immature stands
DEMA-CODE	Density class profile for mature and overmature stands
DEMA-SHDI	Shannon diversity index for density class profile of mature and overmature stands
AgAr-CODE	Age class profile of total forested area
AgAr-SHDI	Shannon diversity index for age class profile of total forested area
AgArExp-Code	Age class profile of exploitable forested area
AgArExp-SHDI	Shannon diversity index for age class profile of exploitable forested area
AgArNoEx-Code	Age class profile of non-exploitable forested area
AgArNoEx-SHDI	Shannon diversity index for age class profile of non-exploitable forested area

Attribute Name	Description
AgVo-Code	Age class profile of growing stock within total forested area
AgVo-SHDI	Shannon diversity index for age class profile of growing stock within total forested area
AgVoExp-code	Age class profile of growing stock within exploitable forested area
AgVoExp-SHDI	Shannon diversity index for age class profile of growing stock within exploitable forested area
AgVoNoEx-Code	Age class profile of growing stock within non-exploitable forested area
AgVoNoEx-SHDI	Shannon diversity index for age class profile of growing stock within non-exploitable forested area
POP/sqkm	Population density per square kilometer
Autow/sqkm	Road density per square kilometer
Railw/sqkm	Railways density per square kilometer
Riverw/sqkm	Waterway density per square kilometer

Appendix 2

2.1. Basic notions of the Rough Sets theory

The fundamental notion of the RS theory when it is applied to analyze a given classification is the *indiscernibility relation* among the objects. Indiscernibility applies to objects that are indiscernible from one another when only a given set of attributes is taken into account. The relation exemplifies the fact that the values of attributes are the sole source of knowledge about the objects. Indiscernibility is an equivalence relation, so it defines a *partition* (also called *classification*) of objects into disjoint subsets, called *elementary sets*. The main concern of the RS theory is to examine different partitions of objects induced by sets of condition and decision attributes, and the relationship between these partitions.

Two particular partitions of the objects are most frequently studied. One of them is the partition induced by the set of all the decision attributes. The elementary sets of this partition are called *classes* – sets of objects described by the same value of a decision attribute. The other partition of interest is induced by the set of all condition attributes. The elementary sets of this partition, called *atoms,* contain objects that are indiscernible from one another with regard to all condition attributes. The name “atom” is intended to stress that it represents the smallest and unsplittable granule of knowledge that can be used to approximate (build) another knowledge, namely this regarding partition of objects into classes. The definitions of atoms and classes are followed by the next step of the RS analysis, in which the different partitions are matched and analyzed.

Any subset of objects (called a *concept*) is *definable* by a set of attributes if this concept can be represented as a union of the elementary sets generated by this set of attributes. If this is not possible, the RS theory introduces the notion of a concept approximation, which consists of the *lower approximation* and the *upper approximation*. The lower approximation of a concept is the union of all elementary sets that are included in this concept, while the upper approximation is the union of all elementary sets that have non-empty intersection with the concept. Thus, the lower approximation is always a subset of the concept, while the upper approximation is a superset of the concept. Concepts for which the lower and the upper approximations are equal are called *crisp* sets; otherwise they are referred to as *rough* sets. Every rough set is characterized by a non-empty *boundary region,* which is defined as the difference between its upper and lower approximation.

Two particular partitions of objects, namely the partition defined by the decision attributes (consisting of classes) and the partition defined by the set of all condition attributes (consisting of atoms), are of special interest. If each class is definable by the set of all condition attributes then the values of the condition attributes provide enough information to distinguish between objects belonging to different classes. Otherwise the non-definable classes are represented in form of approximations. The situation is referred to as *data inconsistency*.

To control the level of inconsistency in the data, the RS theory introduces a special measure called the *quality of approximation,* which is defined as the ratio of all objects belonging to lower approximations of all classes to all objects in the decision table. Maximum value of this measure, equal to *1.0*, indicates that all the classes may be fully distinguished from one another using the information supplied by the condition attributes.

It may be interesting to explore if there are some proper subsets of condition attributes which are sufficient to generate the same quality of approximation as a whole set. This leads directly to the idea of attributes' reduction.

2.2. Reducts and Their Computation

A *reduct* is defined as a subset of attributes that is minimal with regard to inclusion and that ensures the same quality of approximation as the whole set of attributes. In general, it is possible that there is more than one reduct for a given decision table. In that case, the set called the *core* of attributes is defined as the intersection of all reducts. In other words, the core consists of those common attributes that belong to all reducts. As far as data consistency is concerned the core is the set of most relevant and indispensable attributes in the table – removal of any of the core attributes from the decision table leads to the increase of data inconsistency, which is manifested by a drop of the quality of approximation.

Generating the core is easy because it does not involve finding all the reducts and producing their intersection. A much more convenient way is to remove one by one each of the attributes and to check the quality of approximation: if the quality drops then the given attribute should be included in the core.

The process of generating reducts, on the other hand, is computationally complex (NP-complete). As a result, apart from exact algorithms, designed for generating all reducts from a decision table, there exist approximate algorithms, designed for generating single reduct or population of reducts, with the aim of decreasing computing time. The main disadvantage of the approximate algorithms is that it is not possible to maintain that reducts generated in such a way are indeed minimal.

In many practical situations the difficulties associated with core and the reducts are:

- a number of reducts is usually very large (frequently too big to be effectively analysed),
- the core is often empty.

This indicates that the regularities in the data are not clear enough to be captured in a form of the classical RS concepts like core or reducts. It does not mean, however, that such regularities do not exist.

In an attempt to catch and express those regularities a notion of β-core is introduced, which is a natural generalization of the classical RS core. Assuming that β be a real number from [0,1], the β-core is the set of all attributes whose relative frequency of occurrence in all reducts is not lower than β. This definition ensures that the β-core is equivalent to the core in the classical RS sense.

It is important to stress that unlike the classical RS core, the β-core must be generated by computing all reducts and calculating the attributes' frequencies. This may be quite difficult, especially when the number of reducts is very large. The β-core may be useful, however, in exposing some interesting regularities in the decision table. Additionally, the β-core may prove helpful in handling the large number of reducts: the reducts that do not include the β-core are discarded and only a small set of reducts remains to be analysed. Following the terminology, these remaining reducts may be referred to as β-reducts.

2.3. Decision Rules

A decision rule is a logical statement defined as "***if*** *some conditions are met,* ***then*** *some decisions are recommended*", where the condition part is a conjunction of elementary conditions (i.e. elementary tests on attribute values), and the decision part is a disjunction of recommended decisions (i.e. assignments to classes). A rule is said to cover an object if all conditions in the condition part are matched by the attribute values of an object.

Decision rules are generated by the way of induction. During this process, two sets of objects are considered: a set of positive objects and a set of negative ones. For the decision rule being induced, positive objects covered by the rule are *supporting* it, and negative objects covered by this rule are *contradicting* it. The ratio of the number of covered positive objects to the number of all objects covered by the rule is called *discrimination level*.

The set of negative objects is always defined as a complement of the set of positive objects. The set of positive objects may be defined in one of two ways:

- in consequence of the RS approach, as either lower approximation or the boundary of a given class; in this case, *exact* and *approximate* rules are induced, respectively; discrimination level of induced rules is equal to 1 (Grzymala-Busse, 1992; Predki *et al.*, 1998);
- directly, as a given class; in this case discrimination level of induced rules is less than or equal to 1.

It is possible to generate the rules using the following induction strategies:

1. Induction of all possible rules. This approach provides the deepest insight into the analyzed data set (all existing relationships between attribute values and definition of classes are shown), but may be computationally inefficient even for small data sets.
2. Induction of a minimal set of rules (so-called *minimal covering*). This approach provides a minimal number of rules that cover all objects from the analyzed data set.
3. Induction of rules satisfying some user requirements (so-called *interesting rules* or *satisfactory description*; Mienko *et al.*, 1996). This approach provides a set of rules that represents some information patterns and regularities in the analyzed data set, and as such can be helpful in understanding and explaining relationships between attribute values and definition of classes.

For the *interesting* rules, user requirements are defined in terms of:

- minimal strength of a rule (either absolute, as a number of positive objects covered by the rule; or relative, as a ratio of a number of positive objects covered by the rule, to a number of all objects in the class); rules that are weaker than a given threshold are not induced;
- maximal length of a rule (defined through a number of elementary conditions in the condition part of the rule); rules longer than a specified threshold are not induced;
- minimal discrimination level of a rule; rules with discrimination level smaller than a given threshold are not induced.

The induction of *interesting* rules (discovery-oriented induction) is based on a breadth-first exploration of the rule space restricted through the thresholds defined above.

The induction process of *interesting* rules starts with the shortest rules (length equal to 1) and the rule length increases in next steps. In each step all rules are evaluated against the threshold values of length and strength. Rules that are too long or too weak are discarded. Then the level of discrimination of remaining rules is evaluated and all rules with an acceptable value of this measure are stored. Rules with unacceptable value of discrimination level are further specialized by adding new elementary conditions. The process stops when there are no more rules to consider.

It must be stressed that there is no claim as to the *interesting* rules constituting a complete description of the classification in terms of the condition attributes. The interesting rules represent only a part, although a well-founded one, of domain knowledge. The reasons for that are the following:

- the *interesting* rules do not cover all the objects from the decision table,
- the decision table need not contain a representative sample of objects,

A set of decision rules describing a given database of objects can be seen as a classifier. In order to asses its ability of accurate reclassification of objects (examples) one of cross-validation tests might be performed. For a medium size database (up to 100 objects) it is recommended to use, so-called, “leaving-one-out” test. Such test is a loop with number of iterations equal to number of objects in the database. In each iteration one object is considered as a test object, and the remaining objects are used for inducing decision rules. Generated rules are in turn used to classify the test object. Each object from the database is classified exactly once.

2.4. Discretization of Continuous Attributes

From the practical point of view, the indiscernibility relation may be applied only if the values of the attributes are symbolic (qualitative, discrete), as even very small differences in continuous values affect considerably the definition of atoms. To prevent this from happening, the continuous attributes should be discretized. As a result of discretizing, precision of the original data is decreased (in the sense that the original values of the attributes cannot be reconstructed on the sole basis of the discrete values), but its generality is increased.

It should be also stressed that discretization of continuous values is deeply embedded in human reasoning. For example, a decision maker often groups actual values together and considers discretized values such as “low”, “medium” or “high”, *etc*.

Most typical discretization information (called a *hard discretization*) consists of a finite set of numbered subintervals defined over the range of continuous attribute values. This type of discretization is also referred to as *norms*, because the subintervals are frequently defined following some norms in the subject domain. The subintervals are used to discretize the continuous values by substituting an original value with a number of an interval to which it belongs. A more advanced form of discretization involves subintervals represented as fuzzy numbers with overlapping bounds. This fuzzy form of discretization requires different, usually more advanced techniques for processing the discretized decision tables (Slowinski and Stefanowski, 1994).

When a domain expert following his/her judgment specifies the subintervals for the discretization, then they are called *expert discretizations*. On the other hand, when they are defined automatically, then they are called *automatic discretizations* (for a review of automatic discretization procedures see Dougherty *et al*., 1995; Fayyad and Irani, 1993; Nguyen, 1998; Susmaga, 1997).