RECONCILIATION OF FOREST PRODUCT TRADE DATA

Gábor Kornai

August 14, 1985
WP-85-48

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

FOREWORD

The objective of the Forest Sector Project at IIASA is to study long-term development alternatives for the forest sector on a global basis. The emphasis in the Project is on issues of major relevance to industrial and governmental policy makers in different regions of the world who are responsible for forest policy, forest industrial strategy, and related trade policies.

The key elements of structural change in the forest industry are related to a variety of issues concerning demand, supply, and international trade in wood products. Such issues include the growth of the global economy and population, development of new wood products and of substitute for wood products, future supply of roundwood and alternative fiber sources, development of new technologies for forestry and industry, pollution regulations, cost competitiveness, tariffs and non-tariff trade barriers, etc. The aim of the Project is to analyze the consequence of future expectations and assumptions concerning such substantive issues.

The research program of the Project includes an aggregated analysis of long-term development of international trade in wood products, and thereby analysis of the development of wood resources, forest industrial production and demand in different world regions. The analysis is carried out by means of a model of the global sector. The purpose of this article is to describe the preliminary model of the Soviet forest sector in connection with a global model. Some historical data, statistical forecasts as well as simulation runs with the model for demonstration purposes are presented.

The work reported here was done in close collaboration between the Forest Sector Project of the International Institute for Applied Systems Analysis (IIASA) and the Food and Agriculture Organization of the United Nations (FAO). The reasons for this joint effort were twofold. As far as IIASA is concerned, the incentive was our need for a consistent and full-scale historical data set on bilateral trade flows to be used in the global model. The FAO had several interests for participating in this analysis. First, to gain a fast computerized access to the very large UN data bank on trade in forest products. Then to carry out a series of comparisons between the UN and the FAO data, and, finally, to generate a set of standardized trade matrices. It is expected that the results will provide the basis for further validation of FAO data.

Both the aims of IIASA and of FAO require an algorithm for "reconciling" the data of different origins with each other. This algorithm can then be used to calculate a set of trade matrices that may be closer to reality. The importance of such matrices is evident: past trade and other economic analyses having been based on any one of the various original trade matrices led to conclusions of a somewhat diverse and conflicting nature, reflecting sometimes the specific features – and errors – of the data set. The reconciliation method reported here produces historically consistent trade matrices. Even though some rules of the algorithm may be questioned, it is believed that the simplicity of our procedure makes it easier to use than many of the other known and more sophisticated data reconciliation techniques.

Markku Kallio
Leader
Forest Sector Project

Gábor Kornai
Forest Sector Project

ACKNOWLEDGMENTS

I am deeply indebted for the generous and extensive help I have received from the FAO Forestry Division, Economic and Statistical Analysis Unit in Rome, and, I owe a very special debt of gratitude to Philip A. Wardle, head of the Unit, for his encouraging assistance, advice and push in completing this exercise. The valuable contribution of Serge Medow (IIASA Computer Services) and Felice Padovani (FAO) in ironing out some of the not so evident computational problems in the early phases of the work, and the suggestions I have received from Daniel Chappelle (Michigan State University), Michael Martin (FAO) and Jack Weeks (UNIDO) are also acknowledged here. However, I am alone responsible for the errors.

CONTENTS

RECONCILIATION OF FOREST PRODUCT TRADE DATA

Gábor Kornai

1. THE DISCREPANCIES IN THE DATA

In principle, a trade flow may exist between any pair of countries for each commodity in each year. Let, for a given commodity:

$e_{i,j,k,t}$ represent a bilateral trade flow from country i to j for commodity k in year t;

$x_{i,j,k,t}$ represent exports from country i to j;

$m_{i,j,k,t}$ represent imports to country j from i.

In theory,

$$e_{i,j,k,t} = x_{i,j,k,t} = m_{i,j,k,t} \quad ,$$

since they represent the very same commodity flow. In fact, however, $x_{i,j,k,t}$ and $m_{i,j,k,t}$, as determined from trade flow statistics, are seldom equal. The differences in fact can be huge, and as the literature of the problem indicates there is no sign of improvement yet. Morgenstern observes that the problem was already noticed in 1885 by C.F. Ferraris (Morgenstern 1963:139), and Párniczky refers to a 1920 study by Zuckermann (Párniczky 1984).

The matrices $\{X\} = x_{i,j,k,t}$ and $\{M\} = m_{i,j,k,t}$ of observable trade flows are rather sparse. The "deeper," the more specific the definition of a commodity is, the least dense the trade matrix will be: most countries do not trade with each other. There are several economic, geographical, historical and political reasons for this, e.g., high tariffs are imposed, transportation costs are high because the distance is too big between a pair of countries, there are no traditional links between them, embargos prohibit the trade, etc. (Even in the case of our highly aggregated GTM with 18 model regions and 16 commodities defined, only about 30% of all possible trade routes is actually used. The number of all possible trade routes in the GTM is 18 × 17 regions × 16 commodities = 4896, but only 1465 flows were observed in 1980.)

In international trade statistics commodities are usually defined by the Standard International Trade Classification (SITC, see UNITED NATIONS 1975) at various "depths" of breakdown. The level of aggregation is generally referred to by the number of digits in the SITC-code of a commodity. The most aggregated level has a 1-digit code, such as SITC 6 for "Manufactured Goods Classified Chiefly by Material," while a 4-digit code refers to an already rather homogeneous product, e.g., SITC 641.1 stands for "Newsprint." The system of SITC commodity codes has so far undergone two major revisions, first in 1960 and again in 1975. To avoid confusion, we will refer to the latest version, the so-called SITC Revision 2 classifications.

There are two major more-or-less independent sources of bilateral trade data on forest products. The first is the "Direction of Trade" tables published by the Food and Agriculture Organization of the UN, the FAO (See FAO 1983). We will refer to this source henceforth as the "FAO" data. The second set of bilateral trade observations is also available on a computer tape from the UN Statistical Office (see UNITED NATIONS 1981), referred to as "UN" data below. Besides the above two sources, the Organization for Economic Cooperation and Development, (OECD), the International Monetary Fund (IMF), and the UN Commission for Trade and Development (UNCTAD), among others, also publish "Direction of Trade" matrices, originally subsets of the UN data. The analysis in this paper, however, is limited to reconciliation between the FAO and UN data only.

Both data sets include so called "partner country" reports. In principle, all trading should be reported both by the exporter and the importer country identically:

the exporter country i reports $x_{i,j,k,t}$ for all $j = 1, \cdots, n$,

the importer country j reports $m_{i,j,k,t}$ for all $i = 1, \cdots, n$,

for all commodities traded each year (or month). However, the principle that both reports should be equal, that is, $x_{i,j,k,t} = m_{i,j,k,t}$, so far could not be fully implemented. (For empirical evidence see Durst et al. 1985, Luey 1971, Naya 1969, Párniczky 1980 and 1984, Yeats 1978.) To illustrate this statement we selected some bilateral trade flows from the group of highly developed industrial nations which are believed to have reasonably good foreign trade statistics (see Table 1).

In the case of the UN data a series of simple computations has helped us to find some characteristics of the partner country reports. A brief summary by some selected products is given in Table 2. There are several studies available analyzing systematically the statistical magnitude of the discrepancies. (See Durst et al. 1985, Naya 1969, and Yeats 1978.) Our intention here was only to indicate the problem.

What are the lessons from the "observations" in Tables 1 and 2?

(a) Neither of the statistical sources can be considered as really superior to the other. It seems to be necessary to have an "independent" set of information to make quality judgements.

(b) Even the reports of the most developed countries are suspicious. Discrepancies are most likely to be bigger in the case of developing countries. This is clearly indicated in Table 2, where reports on Newsprint (a commodity traded mostly among the developed countries) coincide more closely than those on mechanical wood products.

(c) There is no indication that quantity reports are less reliable than value data, although volumes may be measured in different units, even if they bear the same name (e.g., a cubic meter can refer to a stacked or a solid

Table 1: Discrepancies in the direction of trade data (illustration).

Route 1. Reported for	As reported by Canada	USA	Converted to metric tons by using	
FAO (metric tons)	5635600	5734000	–	
UN (volume, mt)	5660376	5733723	–	
UN (value,1000 US$)	5787770	6139339	FAO Unit value	
FAO/UN (volume)	0.9956	1.0000	Route 1: Exports of Canada to USA, 1970. Newsprint (SITC 641.1)	
FAO/UN (value)	0.9737	0.9340		
UN (vol)/UN (val)	0.9780	0.9339		

Route 2. Reported for	As reported by USA	Japan	Converted to cubic meters by using	USA/Japan
FAO (cubic meters)	1195000	na.	–	–
UN (volume, CuM)	670176	1002494	FAO	Conv.factor
UN (value,1000 US$)	1206014	1422685	FAO	Unit value
FAO/UN (volume)	1.7831	–	Route 2: Exports of USA to Japan, 1981. Coniferous Sawnwood (SITC 248.2)	
FAO/UN (value)	0.9909	–		
UN (vol)/UN (val)	0.5557	0.7046		

Route 3. Reported for	As reported by France	FRG	Converted to cubic meters by using	France/FRG
FAO (cubic meters)	135000	na.	–	–
UN (volume, CuM)	130940	167105	FAO Conv.factor	0.7836
UN (value,1000 US$)	171452	178964	FAO Unit value	0.9580
FAO/UN (volume)	1.0310	–	Route 3: Exports of France to FRG, 1981. Non-coniferous Sawlogs and Veneer Logs (SITC 247.2)	
FAO/UN (value)	0.7874	–		
UN (vol)/UN (val)	0.7637	0.9337		

Route 4. Reported for	As reported by USA	Canada	Converted to cubic meters by using	USA/Canada
FAO (cubic meters)	401000	na.	–	–
UN (volume, CuM)	365006	597900	FAO Conv.factor	0.6105
UN (value,1000 US$)	310710	427935	FAO Unit value	0.7261
FAO/UN (volume)	1.0986	–	Route 4: Exports of USA to Canada, 1981. Non-coniferous Sawnwood (SITC 248.3)	
FAO/UN (value)	1.2906	–		
UN (vol)/UN (val)	1.1747	1.3972		

Table 2: Some Features of the UN Trade Flow Reports by Selected Commodities.

Product	(SITC)	Year	Unit (1000)	Total Number of reports	Double reports (see Notes)	World Trade by (M) importers	(X) exporters	M/X
Logs, (C)	(247.1)	1981	CUM	565	110	20546.8	14437.9	1.42
Logs, (NC)	(247.2)	1981	CUM	1091	213	30456.0	16363.0	1.86
Sawnwood, (C)	(248.2)	1981	CUM	1555	368	43420.9	15112.3	2.87
Sawnwood, (NC)	(248.3)	1981	CUM	2045	430	10103.1	4831.4	2.09
Pulp	(251.0)	1981	MT	756	350	16243.1	17545.7	0.93
Printing & Wr.	(641.2)	1981	MT	3026	924	6325.6	7208.5	0.88
Newsprint	(641.1)	1970	MT	1173	320	10016.4	10245.0	0.98
Newsprint	(641.1)	1975	MT	1168	315	8710.9	8839.4	0.99
Newsprint	(641.1)	1980	MT	1167	324	11915.1	11542.6	1.03
Newsprint	(641.1)	1981	MT	1146	297	11586.6	12154.2	0.95
Paper N.E.S.	(ex641)	1981	MT	2798	1268	14933.4	13674.1	1.09

Notes. In column "Double reports" we give the number of trade flows reported by both the exporter and the importer country. (C) stands for Coniferous, (NC) for Non-coniferous commodities. Paper N.E.S. is Paper-Not-Elsewhere-Specified (total Paper-Newsprint-Printing and Writing Paper). Units were re-converted into the indicated measurement (Cubic Meters or Metric Tons) using the FAO standard conversion factors (See FAO 1983:61).

wood measure). As the above examples indicate, any re-conversion of different units may be questionable. However, there is no evidence that the unit values used for conversion into quantities are any better.

Without trying to mention them in the order of their likely magnitude, the basic well-known reasons for the frequent discrepancies are compiled below using the following works: Bhagwati 1967 and 1981, DeWulf 1981, Durst et al. 1985, Luey 1971, Morgenstern 1963, Naya 1969, Párniczky 1980 and 1984, Richter 1970 and 1971, Simkin 1970, and Yeats 1978.

(1) Statistics are – in general – generated by the customs authorities. Simple unintentional counting and recording mistakes made by customs officers and reporting officials may result in errors. However, some countries collect data directly from trading enterprises, who, in principle, "should know it better." But they may also make mistakes.

(2) There is a certain shipping and recording time lag due to distance and administrative procedures. Thus, "a specific period of time has a different coverage in terms of exports and imports." (C.f. Párniczky 1980:44)

(3) Origins and/or destinations of shipments may be identified incorrectly or differently by the trading partners. The "middleman" transit trade, the so called "free-trade zones" and the recordings based on the "country of production" principle instead of the preferred "country of consignment" concept are the widest-spread sources of these errors.

(4) Diverse definition of commodities by the trading partners may also lead to discrepancies. We should mention here the effect of the various physical conversion factors in use (c.f. Darr 1984).

(5) In a good number of cases, trade remains simply unrecorded. Besides smuggling and deliberate falsification by one or both partners (c.f. Bhagwati 1967 and 1981), there are several countries making no effort to report their trade at all. Besides secrecy (and partly for it), many countries use special systems of statistical data compilation which are hard – if not impossible – to correlate with the internationally standard classification systems of trade. (Among these countries – considering the role they play in the trade of forest products – I think, the most important non-reporters are the socialist countries: there is basically no information available on their trade with each other, while some bilateral trade statistics can be obtained only through the reports of their trading partners. C.f. UNITED NATIONS 1981:1–8.)

(6) There is an intentional (official) under- and over-invoicing of trade practiced in several countries. The basic reasons for such are to receive and/or provide advantages (e.g., subsidies, price differentials, duties, tax deductions etc.) to domestic buyers and/or suppliers of foreign trade, and, to loosen or tighten quotas.

(7) Trade flows measured in value terms will in most cases differ because of the so-called "free on board" (or f.o.b.) valuation of the exporter parties versus the "cost, insurance, freight" (c.i.f.) recordings of the importers. However, this traditional way of valuation may not be implemented by several countries which use a different set of rules.

(8) Value reports may also differ due to the exchange rates used to convert local currency values into US dollars. These "official" rates of exchange may (and sometimes do) significantly differ from those actually used in the transaction, partly due to the time lag mentioned before, partly because of the various rules "official" rates of exchange are derived in several countries.

Reasons (1) through (6) affect both the quantity and value data, while the consequences of (7) and (8) appear only in the value figures. The existence of the (mismatching) partner country reports makes it possible – and necessary – to look for the "real" trade figures.

2. SOME ESTABLISHED METHODS AND EXPERIMENTS

It is nearly impossible to decompose the data to isolate the influence of each and every problem mentioned above. Thus, there is no overall method available for a consequent reconciliation of trade reports. There is no doubt that bilateral trade data should be reconciled somehow before one utilizes it for any serious analysis or research. However, the actual method chosen for the reconciliation should depend on the purpose for which the data are intended.

For an in-depth research with a single commodity, a mainly manual reconciliation of the historical data would be most advantageous. One may try to identify and quantify the effects causing the most striking discrepancies, then check and verify each individual trade flow, utilizing many pieces of microeconomic information ranging from monthly trade data to transportation cost figures, etc. Apart from some well known guiding principles, there cannot be a general algorithm designed for this type of work.

A study on general patterns of bilateral trade, however, may require masses of trade flow data to be processed without much attention paid to any single flow. There are several such "automatic" data reconciliation methods available, each utilizing some statistical or mathematical properties the data under inspection are assumed to have.

According to the rather loose classification above, the scale of reconciliation methods available ranges from fully "manual" to fully "automatic" procedures. Before going into details on the method we applied, a brief overview of some previous reconciliation practices might be helpful to identify our basic points of departure. We have no intention, however, of reviewing those algorithms. Our only question is whether they are in any respect superior to our way of doing the job. (I am concerned here only with reconciliation of data for the historical, or observation period; thus no conclusions will be drawn on the forecasting capabilities of the methods to be mentioned below.)

The best known "automatic" procedure is the so called "RAS" – or biproportional method of Stone (see Lecomber 1975:1–25). RAS was devised as an operational technique to solve the problem of updating input-output matrices. These matrices are very similar to trade flow tables as there are industries (rows) providing input for user industries (columns): analogous to exporters and importers of a commodity. (See Lecomber 1975:8–9 for references on RAS applications with trade flow data.) Without going into details on the mathematical properties of this method (and the closely related other "limited information" problems), the essence of the procedure is the following. Given the total exports and imports by country, and having obtained a first estimate on the nonzero trade flows (thus defining the cells for the actual trade routes), the algorithm finds a value for each observed trade route, so that both export and import totals are satisfied by country. However, there is no evidence that the values allocated to the individual "trade flow" cells of the matrix will be in the range of the observed figures: the concern of the method is to satisfy the biproportionality conditions. Recent versions of RAS allow for constraining some of the "most important" individual matrix cells, thus fixing their range to be "close to reality." But then we can hardly call RAS an "automatic" procedure any more.

My deepest concern with RAS and the other closely related limited-information procedures used for historical data reconciliations is that the problem of trade matrices we face here is not at all a usual "minimal information" type. We have a great number of matrix elements, for which there are more that enough, sometimes even quadruple – basically mismatching – observations.

Such problems might be better fit with the now available mathematical programming solutions (see Harrigan 1983). These methods allow for the imposition of more and more actual economic observations, constraints, and other variables; they are capable of using more general functions than RAS, etc. However, the application of these methods are still rather complicated, and often are not yet fully tested for possible economic consequences of the main underlining mathematical assumptions. For example, the "biproportionality" condition may be true as a purely mathematical assumption, but the real world does not back it up: there is some evidence available that export and import totals as were reported by the countries are not the same reliable and consistent economic observations (see Párniczky 1980. In fact, world trade – even in quantity terms – never adds up to the same number for exports as for imports, thus reconciliation results obtained assuming and using biproportionality conditions are questionable.

Partly for the reasons stated above, partly for the sake of simplicity we prefer more easy to follow, though much more naive methods. Such an algorithm was used to reconcile the trade matrices of six major agricultural products (see Párniczky 1984). This semi-automatic procedure consist of the following two steps:

1. An automatic (computerized) assessment of trade flow data was undertaken for the selected commodity from the UN data set. Trade flows were selected and ranked by their volume until either a 90 percent coverage

of world trade, or a 30 country limit was reached. Then a temporary set of trade was computed according to simple rules (C.f. Párniczky 1984:6):

(i) If there are no data reported: the partner country's figure is accepted;

(ii) If both the reporter's and the partner's figure is available, and the ratio of the two is between 0.8 and 1.2: the importer's declaration is accepted. (Párniczky's arguments for preferring import figures over exports are reasonable: it is easier to establish the origins of imports than to foresee the real export destinations at the time of declaration, and, customs administration pays generally much more attention to the inspection of imports than exports. However, the "ratio rule" applied here may be questionable.)

(iii) If both the reporter's and the partner's figures are available, and the ratio of the two is outside the given range, a question mark is printed to call the attention of the next step's expert to the mismatch.

2. A manual operation finalizes the reconciliation, with an international trade expert making individual decisions on each trade flow figure according to the following rules:

(i) Maintain country total export and import figures, if possible;

(ii) Establish the reasons for large discrepancies;

(iii) Detect and eliminate middleman trade flows and reallocate them to real origins;

(iv) Check consistency of world totals;

(v) Check residual data arising from the differences of country total exports and imports with respect to the reconciled flows.

With the help of the above very simple algorithm, FAO could prepare three matrices for each of the six commodities they investigated: an export, an import and the "reconciled" matrix of approximately 20 exporting by 30 importing country size. However, I think that Párniczky's procedure is far too labor-intensive to be fruitfully used with a long list of years and commodities. Moreover, in the case of the forest products, if applied, a high-percentage cut-off point may jeopardize the full coverage principle of our experiment.

3. THE SOURCE DATA

Before going into detail about our method of reconciliation, we should first describe the most important features of the data we intend to use. Some of the consequences we arrive at should be taken care of during the reconciliation.

The United Nations collects trade data according to the SITC codes. Bilateral trade figures are assembled as direct copies from the national trade yearbooks of the member countries of the UN. These data, previously published only in a considerably summarized printed form (see UN 1981) or in detail in the form of microfiche, are now available in computer readable form back to 1961. This computer tape database contains all available reports on bilateral trade, including

$$x_{i,j,k,t},\ m_{i,j,k,t}, \text{and } x_{(i).,j,k,t} \ .$$

(The latter are the re-exports of unknown original supplier, denoted with ".").

Each edition of the FAO Yearbook of Forest Products (see FAO 1983) has a summary trade matrix for major products for the latest two years. The matrix for the second latest year in each edition includes revision of the data included in the previous edition. These data were recently made available to us in computer readable form. FAO publishes export matrix data $\{x\}$ for the years 1966–1983, while the import matrices $\{m\}$ are available only for years 1966–1976.

As a consequence, the reconciling algorithm may take advantage of having as much as four corresponding reports (the $x_{i,j,k,t}$ and $m_{i,j,k,t}$ of the UN, and the same entries of the FAO), but should also work for years when only two reports can be obtained (from the UN).

The UN data covers each country that has appeared in a trade report either as a reporter or as a partner. The FAO, however, filters the information. With limited staff available to compile forest products trade statistics in FAO, first dependence is placed on national returns to the FAO questionnaires. Trade data are completed by reference where necessary to national Trade Yearbooks. The UN microfiche, which usually arrives there later, is used when no report is received in FAO. In the event that no national data are available, information is extracted from reports of major traders which may show the country as a trading partner. It will be appreciated that whether using national yearbooks or microfiche these methods of searching for missing data are extremely laborious. Thus, the trade matrix published by FAO currently includes only countries where total imports or exports exceed one percent of the total volume of world trade of a given product. Less important residual traders are lumped together into several categories of "OTHERS". Thus, the actual size of an FAO trade matrix is about 20 by 20 countries. This means that the FAO data can be used only as far as it covers the world geographically. However, the economic coverage of the world by FAO is not as poor as the geographical – representing 80–90 percent of total trade – since international trade in forest products is rather concentrated (c.f. Francescon et al. 1983).

From a technical point of view, the UN and the FAO use different systems of country codes. This not only makes cross section comparisons more complicated, but in some cases, due to political changes, intertemporal comparisons may become harder (e.g. Vietnam, Tanzania, etc.). These definitional problems should be also solved before any reconciliation.

Similarly to the large number of countries appearing in the UN reports, the UN covers all commodities traded, that is, several thousand goods classified in one to five digits of SITC. FAO, however, focuses only on 10–12 major forest products defined according to the role they play in the forest economy. Thus, some of the FAO-commodities cannot be identified directly with any one of the SITC categories, however, most of them may be composed as an aggregate of several SITC commodities, all in a three to five digits definition. These – sometimes rather loose – SITC categories appear in the Yearbooks (see FAO 1983) specified in Revision 2 of the SITC. Lacking detailed FAO data, the reconciling algorithm should work only using UN data, and, should be capable of aggregating several primary commodities into the broader categories of the FAO. From the technical aspect, the UN is still obliged to use the SITC Revision 1 commodity classifications for the countries reporting in that system only (c.f. UNITED NATIONS 1983). These codings are not always directly convertible or aggregatable into the Revision 2 categories referred by the FAO. (For example, 241.1 of SITC Rev. 1 – Fuelwood and wood waste – is now broken into 245.01 Fuelwood, and 246.03 Wood waste. For the correlation between the two Revisions, see FAO 1982.)

The UN provides the trade flow data basically as a US dollar value converted from national currencies using exchange rates mostly furnished by the IMF. When the homogeneity of the commodity allows, some quantity (volume) figures are also shown. However, there is no guarantee that all countries report in the same unit of measurement which is suggested by the UN (metric tons). The FAO does not publish value data for the trade flows, but the volumes reported there are all measured in consistent physical units suitable to the commodities (cubic meters for mechanical wood products and metric tons for pulp and paper products). Thus, any reconciling procedure making use of both value and volume observations should incorporate options for re-converting value data into volumes (e.g. using unit values), or volume data into other quantity measures (e.g. weight to volume).

The primary sources of UN data are official government reports generally derived from customs records. The UN makes only the technical updating of its trade data files (currency conversions, aggregation of monthly and quarterly information, summation for higher level commodity classes, etc.), but does not carry out any further checking or revision on the source data: reports are accepted irrespective of their quality. The FAO, however, continuously checks and double-checks the sources, verifies, revises and even estimates the data when it is missing from their original questionnaires supplied by national forestry authorities. The experts of FAO make use of the other available information (on production, consumption, processing technologies, forest resources, etc.) before publishing their trade matrices, which are then even frequently updated for historical years. These important features of the two data sources are summerized in Table 3.

It seems rather obvious that for a complete trade flow assessment the UN data may be considered somewhat superior to that of FAO since it covers more products and years, and provides both values and volumes. However, there are a number of problems with the quality of the UN data. The discrepancies between individual trade reports then are especially amplified in comparison with FAO trade flow statistics.

Table 3: Some features of UNSO and FAO trade flow statistics.

	UNSO	FAO
Years covered:	1961–1983 (23) Importers' reports:	Exporters' reports: 1966-1983 (18) 1966–1975 (10)
Countries covered:	All reporters.	"Major" reporters (1966–1977). Reporters with trade ≥ 1 % of world exports (1978–1983).
Products covered:	All products traded.	Some important forest products (3 to 5 digit SITC definition).
Product definition:	SITC Revisions 1 and 2. (as reported)	FAO-definition, mostly related to SITC Revision 2.
Measurement used:	Value in US$. Volume in metric tons (but some countries may use other units).	No values available. Volume in cubic meters or in metric tons (all countries use the same unit for a given product).
Primary source of data:	Not verified or checked government reports.	Questionnaires to forestry authorities revised by FAO experts and/or estimates based on other information.

4. OUR RECONCILIATION ALGORITHM

The trade matrix data reconciling procedure I will describe here is also of an "automatic" type. It is rather "naive", thus somewhat more transparent than the ones mentioned before. I consider these features as advantages over other procedures which may work in the shadow of some non-evident mathematical assumptions, or which require substantial manual intervention.

The procedure computes a single "final" reconciled matrix, which contains the full direction of trade information (i.e., I obtain the

$$\{e\} = e_{i,j,k,t}$$

figures which are both export and an import matrices at the same time). The dimensions of a matrix $\{e\}$ are limited only by the number of countries appearing in all of the trade reports for a given year and commodity, thus providing a full coverage of world trade. The trade flows are expressed in volume terms only to ensure full compatibility with the FAO trade matrices (if they exist). The only overall criterion I apply is that the "final" matrix $\{e\}$ should be fully consistent with the country-by-country export totals:

$$\sum_{j=1}^{n} e_{i,j,k,t} = X_{i,k,t} \quad , \qquad \text{for } i=i' \ , \ k=k' \ , \ t=t' \ .$$

That is, I make use of a single-sided proportionality only. The export total figures $\{X\} = X_{i,k,t}$ published annually by FAO (see FAO 1983) are considered to be the most reliable ones among the FAO trade statistics, since these data are continuously revised and updated so that each edition of the Yearbook includes data as revised up to the date of publication. (Thus, the general "rule" that more attention is paid to imports than to exports is not always true.) In fact, the export totals by country comprise "exogenous" additional information upon which I depend heavily when computing the final matrices.

I work on a rather detailed commodity breakdown; namely, on commodities defined with four to five digit SITC codes. In spite of the fact that the more specific the commodity definition is, the bigger the discrepancies in the data will be (see Yeats 1978:355–358), I think that forest products are homogeneous enough to justify this depth of specification.

My algorithm consists of the following steps:

(1) Technical preparations (following data input):

   (a) Re-coding of UN source data specifications (country and commodity codes) to their corresponding FAO equivalents. Varying quantity measurement units are converted into the common unit defined by the $X_{i,k,t}$ set using the standard conversion factors of the FAO (see FAO 1983:57–59). Due to the reasons mentioned in Sections 1 and 2 of this paper, this is not always a trivial task. However, with the help of some trial runs, we could identify most of the misleading specifications and conversions.

   (b) Filtering the FAO source data $\{x\}$, $\{m\}$, and $\{X\}$ of all – otherwise redundant – subtotals (e.g., entries like "Developing countries", "Africa", "Others", etc).

   (c) Dropping of all UN reports on re-exports (transit- or middleman trade) partly to avoid double counting, since they may already be recorded correctly by the final importer, partly because the original sources of these flows cannot be identified from the reports by any automatic means.

(2) Compiling the UN trade matrix. A union of exporters' and importers' reports is provided, so that both $x_{i,j,k,t}$ and $m_{i,j,k,t}$ appear in the same i,j,k,t cell. A full list of discrepancies (differences and ratios) is printed, provided there are double reports available.

(3) Unifying the UN data. Each i,j,k,t cell of the matrix will be substituted to contain the larger of x or m:

$$MAXUN_{i,j,k,t} = \max(x_{i,j,k,t}, m_{i,j,k,t})$$

(4) Compiling the FAO trade matrix, similarly to step (2), provided that there are FAO data available for the commodity. The procedure can work without any FAO data. If possible, a full list of discrepancies is printed, where there are double reports available.

(5) Unifying the FAO data, if available. Similarly to step (3), each i,j,k,t cell of the matrix will be substituted to contain the larger of x or m:

$$MAXFAO_{i,j,k,t} = \max(x_{i,j,k,t}, m_{i,j,k,t}).$$

(For years beyond 1976, where there is only an export matrix available from FAO, $MAXFAO_{i,j,k,t} = x_{i,j,k,t}$.)

(6) Unifying the UN and the FAO data, so that both $MAXUN_{i,j,k,t}$ and $MAXFAO_{i,j,k,t}$ appear in the same i,j,k,t cell. A full list of discrepancies is printed where there are double entries in a cell. Then, each [i,j,k,t] cell of the pre-reconciliation matrix $\{e'\}$ will be replaced by the bigger of $MAXUN$ or $MAXFAO$:

$$e'_{i,j,k,t} = \max(MAXUN_{i,j,k,t}, MAXFAO_{i,j,k,t})$$

(7) Computing of pre-reconciliation export totals $\{X'\}$ by country:

$$X'_{i,k,t} = \sum_{j=1}^{n} e'_{i,j,k,t} \ .$$

(8) Merging "real" and computed matrices $\{X\}$ and $\{X'\}$ with each other so that both $X_{i,k,t}$ and $X'_{i,k,t}$ appear in the same cell i,k,t. A full list of discrepancies $D_{i,k,t}$ and trade-flow multipliers

$$D_{i,k,t} = X_{i,k,t} - X'_{i,k,t}$$
$$R_{i,k,t} = X_{i,k,t} / X'_{i,k,t}$$

are computed and printed.

(9) Reconciliation by exporter i:

if $D_{i,k,t} > 0$, then a new "flow" is introduced:

$$e_{i,UNSPECIFIED,k,t} = D_{i,k,t} \ ,$$

since we do not now the real destination.

if $D_{i,k,t} = 0$, then no action is needed:

$$e_{i,j,k,t} = e_{i,j,k,t} \text{ for each } j=1,\cdots,n \ .$$

if $D_{i,k,t} < 0$, then we "compress" each flow:

$$e_{i,j,k,t} = e_{i,j,k,t} * R_{i,k,t} \ .$$

The flows $e_{i,j,k,t}$ obtained in this way will satisfy the adding-up criterion

$$\sum_{j=1}^{n} e_{i,j,k,t} = X_{i,k,t}, \quad \text{for } i=i',\ k=k',\ t=t' \ .$$

The procedure of the above 9 steps will definitely improve the coverage of the trade matrix. However, discrepancy reasons (1) through (8) may all still have their effects. The rule of selecting the biggest flow appearing in the same i,j,k,t cell of the matrix can be justified only by achieving useful results. The only real reason in favor of this rule is, that this way we will not overlook those frequent cases when there is only a single report available for a given flow. This situation arises when – say – since the flow is a minor one, FAO will not mention it, while only one of the parties reported it for the UN. In the light of the several tests we made, I think this "rule" is just acceptable.

5. TEST RESULTS

Due to the enormous quantity of data to work with, the method described above can only be used effectively on large "number-crunching" computers with fast input–output facilities. (The theoretical size of the database for a reconciliation exercise, assuming 150 countries, and two data sources with both exporters' and importers' reports is 150 × 149 × 4 = 89400 pieces of data for a single year and commodity. Our actual database consists only some 2 million observations.) Even if the simplicity and compactness of the computer programs required would allow to run the procedure on a Commodore 64, we need huge disk storage and a tape-device usually not available for the PC's.

The computer algorithm was written both in FORTRAN77 and "SAS" codes. The former was tested on a VAX 11/780, the latter on an IBM 370. Due to these two completely different computing environments, no really comparable run times or CPU statistics can be given here. Once working, a singe reconciliation job takes approximately 30 CPU seconds (including data input). The overwhelming majority of the computer resources is used for such trivial manipulations as reading, filtering, sorting and merging the data. The actual reconciliation – once the full matrix of all available (sometimes multiple) entries was set up – takes less then 3 seconds of the central processor, due to the simplicity of transformations needed. The size of the non-optimized FORTRAN77 code compiled for the VAX is less then 300 kilobytes, the SAS interpreter used 960 kilobytes on the IBM.

We have run several related data reconciliation problems, partly to test our algorithm, partly to obtain some simple measures of discrepancies. Altogether, these 19 exercises included runs with two options related to the type of the UN data used, where

- option "Q" is based on volume figures re-converted into a unified measurement unit on the basis of the standard conversion factors of FAO,
- option "V" is based on value figures re-converted into a volume unit using unit values derived from export totals $\{X\}$ by country in value over volume.

Six runs were based entirely on the UN figures without using FAO direction of trade data (option "UN"). The combinations of these options were selected in order to show the sensitivity of results to the various options. Some basic parameters and statistics of the test runs are summarized in Table 4.

As shown in columns (a) and (b) of Table 4, the geographic coverage of the world by the reconciled trade matrix was in fact doubled with respect to the country total export data originally appearing in the FAO Yearbooks (FAO 1983). However, as the column (c) of the share of this "newly" identified trade indicates, the

Table 4: Summary of reconciliation tests.

Commodity	SITC	YEAR	Option	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Logs (C)	247.1	1981	Q (UN)	82	44	0.73	108.2	455	110	–
Logs (C)	247.1	1981	V (UN)	85	47	2.15	147.0	499	116	–
Logs (NC)	247.2	1981	Q	126	60	0.93	110.6	878	213	45
Logs (NC)	247.2	1981	V	130	64	0.74	148.2	938	222	45
Sawnwood (C)	248.2	1981	Q	122	64	0.22	97.8	1187	368	100
Sawnwood (C)	248.2	1981	V	128	70	0.25	112.5	1368	475	101
Sawnwood (NC)	248.3	1981	Q	145	61	2.00	120.4	1615	430	139
Sawnwood (NC)	248.3	1981	V	150	66	2.67	152.3	1927	650	145
Pulp	ex251.0	1981	Q	64	29	0.25	102.9	756	350	130
Pulp	ex251.0	1981	V	65	30	0.25	107.1	780	367	130
Pulp	ex251.0	1981	Q (UN)	64	29	0.25	95.4	756	350	–
Pulp	ex251.0	1981	V (UN)	65	30	0.25	100.4	780	367	–
Newsprint	641.1	1981	Q	63	28	0.06	101.7	849	297	63
Newsprint	641.1	1980	Q	72	36	0.09	104.3	843	324	64
Newsprint	641.1	1975	Q	65	28	0.09	109.7	853	315	139
Newsprint	641.1	1970	Q	59	27	0.04	101.8	853	320	200
Newsprint	641.1	1970	V	64	32	0.05	110.5	926	333	200
Printing & Wr.	641.2	1981	Q (UN)	96	47	0.08	107.3	2102	924	–
Paper NES	ex641.0	1981	Q (UN)	127	65	0.20	101.9	2798	1268	–

Notes.

(a) The number of countries appearing in the reports; also refers to the dimensions of the reconciled matrix.

(b) The number of countries appearing here, but not shown in the FAO country total export tables (published in FAO 1983). The only source of data on these countries is the UN trade data set.

(c) The percentage share in world trade of these – previously excluded – countries.

(d) The percentage ratio between the world total of the pre-reconciliation $\{e'\}$ data set and the world total exports published in the FAO Yearbook (FAO 1983).

(e) The total number of trade flows reported by both sources.

(f) The number of conflicting reports within the UN data set.

(g) The number of conflicting reports between the UN and FAO data sets.

Pulp (ex251.0) excludes "Other fiber pulp (SITC 251.6)". Paper Not Elsewhere Specified (ex641.0) excludes Newsprint (641.1) and Printing and Writing Paper (641.2). See text for further explanations.

supplementation to the FAO global figures was in fact very minor; even the largest differences are within a 3% range of the original world totals. There was virtually no addition to pulp and paper commodities, and Coniferous Sawnwood, while the largest discrepancy appeared for Nonconiferous Sawnwood.

Column (d) of Table 4 shows the percentage magnitude of disagreement between the FAO Yearbook world total exports and the data assessed according to the "bigger" rule. The discrepancies obtained in runs based on value data (option "V") are significantly higher than those resulting from volume data. The reasons for this upward bias are rather obvious: since the c.i.f. trade figures of the importers tend to be higher than the corresponding f.o.b. reports of the exporters, in most of the conflicting cases the importers' reports will be selected by the algorithm. Then, these data will be re-converted to quantities using export unit values obtained from the FAO value and volume data on export totals. As column (e) on the total number of trade flows indicates, the gain of having 5–15% more reports when using value data may not pay for the loss in accuracy. (A better solution would be to work on quantity data as long as it is available, while for flows reported only in value terms use the corresponding – export or import – unit value for conversions. However, this rule is not yet implemented into the tested algorithm.)

In the case of mechanical wood products (logs and sawnwood) about 70% of all reports in the UN data set had to be reconverted from metric tons to cubic meters. Since pulp and paper are measured in metric tons in both data sets, there was virtually no need for reconversions. The differences in the magnitude of disagreement – i.e., figures in column (d) of Table 4 for tests with option quantity ("Q") – between mechanical wood products and paper and pulp do not question the relevance of the standard FAO measurement conversion factors I used (see FAO 1983:59).

Most of the actual reconciliation is done within the UN data, since the number of FAO reports is very low (5 to 18% of all observable flows). Column (f) of the above table shows the number of conflicting reports within the UN data set. The relatively low figures – 25–45% of all reports – justify the "bigger" rule; 55 to 75% of all trading always remain unrecorded – or at least unreported – by one of the parties.

The only "dynamic" set of tests were made for Newsprint: in terms of geographic coverage, the data sets used do not reveal any improvement over time: in 1970 46% of the countries was not mentioned in the FAO Yearbook, in 1981 this ratio was 44%; the number of flows observed by the UN was virtually constant – see column (e) –, while the FAO has dramatically cut the scope of observations since 1977 – see column (f).

For commodity Pulp we run parallel tests with and without using the FAO direction of trade data. The tests with option "UN" reveal a smaller magnitude of difference than their counterparts utilizing FAO information. The reason I assume is that there should be such trade flows mentioned in the relatively small FAO data set which were not reported for the UN (For example, trade among the socialist countries). This underlines the importance of utilizing both data sources together.

The most important conclusion, I think, is that both data sources (i.e. the FAO Yearbooks and the UN international trade statistics) are telling basically the very same story about world trade in forest products as a whole: the correspondence of the reconciled and the original world totals is virtually one to one – see column (c). This fact indicates that our rule on selecting the "biggest" entry does not cause real problems – at least in the cases tested.

It is of no doubt that this reconciling algorithm should be further tested. I think, that the real advantages and shortcomings of the present procedure can only be identified in comparison to results of other trade matrix reconciling exercises. Since the software for the reconciliation method of Párniczky is also available at FAO, there will be several comparative runs made in the near future. These tests will give a better indication on the magnitude of bias imposed by our simple rule of selecting the bigger entries.

The present software – besides its capability of generating standardized trade flow matrices incorporating the vast number of observations by the UN – may also serve as a tool for easy access to the UN trade data base in forest products. Thus it may directly help the trade data validation procedures of the FAO.

It may also be useful to apply the suggested trade matrix reconciliation procedure as a preparatory step to any other analyses on bilateral trade.

REFERENCES

Bhagwati, J.N. 1967. Fiscal Policies, the Faking of Foreign Trade Declarations, and the Balance of Payments. *Oxford Bulletin of Economics and Statistics*, 27(1):61–77.

Bhagwati, J.N. 1981. Alternative Theories of Illegal Trade: Economic Consequences and Statistical Detection. *Weltwirtschaftliches Archiv*, 117(3):409–427.

Darr, D.R. 1984. Conversion Factors Can Affect Forest Products Trade Policies. *Journal of Forestry*, August:489–491.

DeWulf, L. 1981. Customs Valuation and the Faking of Invoices. *Economia Internazionale*, 34(1):13-33

Durst, P.B., C.D. Ingram, and J.G. Laarman. 1985. Statistics on Forest Products Trade: Are They Believable? Paper presented at the II. CINTAFOR Conference on the International Trade in Forest Products, Seattle, March 1985. (Manuscript)

FAO. 1982. Classification and definition of forest products. FAO Forestry Paper no. 32: Rome: Food and Agriculture Organization of the United Nations.

FAO. 1983. Yearbook of Forest Products. Several issues. Rome: Food and Agriculture Organization of the United Nations.

Francescon, A., A. Nagy, and G. Kornai. 1983. An Historical Analysis of International Trade in Forest Products. WP–83–80. Laxenburg, Austria: International Institute for Applied Systems Analysis,

Harrigan, J.F. 1983. A Quadratic Programming Approach to Data Reconciliation: Contrasts with RAS. In Grassini, M. and A. Smyshlyaev (eds), *Input-Output Modeling*. CP–83–S2. Laxenburg, Austria: International Institute for

Applied Systems Analysis.

Lecomber, J.R.C. 1975. A Critique of Methods of Adjusting, Updating and Projecting Matrices. In Allen, R.I.G. and W.F. Gossling (eds), *Estimating and Projecting Input-Output Coefficients*. London: Input-Output Publishing Company.

Luey, P. 1971. On Naya and Morgan's Criterion of Accuracy in Trade Recording. *Journal of the American Statistical Association*, 66(336):763–765.

Kallio, M., D. Dykstra, and C. Binkley. 1986. *The Global Forest Sector: An Analytical Perspective*. London: Wiley. (in preparation)

Morgenstern, O. 1963. On the Accuracy of Economic Observations. 2nd ed., Princeton: Princeton University Press.

Naya, S. 1969. The Accuracy of International Trade Data: the Case of Southeast Asian Countries. *Journal of the American Statistical Association*, 64(326):452–467.

Párniczky, G. 1980. On the Inconsistency of World Trade Statistics. *International Statistical Papers*, 48:43–48.

Párniczky, G. 1984. The Reconciliation of Agricultural Trade Flows. Report on an experimental trade matrix project. (W/R1984/E/1.85/1/300). Rome: Food and Agriculture Organization of the United Nations.

Richter, H.V. 1970. Problems of Assessing Unrecorded Trade. *Indonesian Economic Studies*, VI.1. pp. 45–60

Richter, H.V. 1971. Exports: Statistical Problems and Country Patterns. *Bulletin of Indonesian Economic Studies*, 7(3):124–138.

Simkin, C.G.F. 1970. Indonesia's Unrecorded Trade. *Bulletin of Indonesian Economic Studies*, VI.1. pp. 17–44.

UNITED NATIONS. 1975. Standard International Trade Classification Revision 2. Statistical Papers Series M No. 34/Rev.2. New York: United Nations.

UNITED NATIONS. 1981. Yearbook of International Trade Statistics. Trade by Commodity. ST/ESA/STAT/SER.G/30. New York: United Nations.

Yeats, A.J. 1978. On the Accuracy of partner Country Trade Statistics. *Oxford Bulletin of Economics and Statistics*, 40(4):341–361.