SILVA FENNICA VOL. 13, 1979, *N:o* 4: 368—384

OUTLINE FOR A DATA PROCESSING SYSTEM IN FOREST MENSURATION

PEKKA KILKKI

SELOSTE:

EHDOTUS METSÄNMITTAUSTULOSTEN LASKENTAMENETELMÄKSI

Saapunut toimitukselle 1979-09-12

This paper presents the principles of a unified data processing system suitable for derivation of the most variables of interest in forest mensuration. The precedence (succedence) relations between the tree and forest stand variables are analyzed and a blockwise simultaneous recursive multi-equation model is suggested to describe these relations. Regression analysis is used in the estimation of the model parameters and Taylor's series and Monte Carlo simulation are available in the derivation of the unbiased results.

1. INTRODUCTION

The purpose of data processing in forest mensuration is to transform the original measurements into a form more applicable to forestry decision making. For instance, tree diameter and height measurements are transformed into volumes and single tree parameters into growing stock parameters. Before the introduction of electronic computers, the data processing of forest measurements was necessarily very simple. Quite often the measurements were designed in such a way that only very few calculations were needed. In the visual estimation of stands, for example, all the growing stock characteristics were estimated directly in the forest. The data processing only consisted of the multiplication of the estimated per hectare values by the areas of the stands, and the summation of the results.

In the first phase of the introduction of computers, the traditional calculations were simply transferred to the computers. However, it was soon discovered that data processing capacity of the computers not only facilitated traditional calculations but made new calculation procedures feasible. These new computational procedures have made it possible to obtain more information from the traditionally measured forest data, and computational difficulties no longer constrain the development of forest measurement techniques.

The origin of this paper lies in the belief that the potentials of the computer are still not fully exploited in the data processing of forest measurements. Especially, the exploitation of data measured earlier is often insufficient. Further considering the fast development of the computers, it is not unrealistic to develop methods which computationally are today still too time consuming.

The first task in practical forest mensuration work is to analyze what kind of information is needed. Thereafter, the measurement and data processing system has to be chosen which provides the decision maker with this information at least cost. In this paper the main interest is focussed on the data processing part of the information system. However, the study does not limit itself to producing some particular information but tries to present principles of a unified data processing system suitable for most if not all tasks in forest mensuration.

The unification of the data processing system tries to serve several purposes: (1) To avoid repetative work in the derivation of results from different measurements by using common parts of the data processing system whenever possible. (2) To use the existing information as efficiently as possible whether it is in the form of measured data or mathematical models. (3) To avoid internally contradictory results due to the separate estimation of different tree and growing stock parameters.

The scope of this paper is so wide that a complete solution cannot be given. The main purpose is to introduce the underlying theory and present some examples of the possible practical applications.

The manuscript has been read by prof. SIMO POSO. His criticism has been valuable and is acknowledged.

2. THEORETICAL BACKGROUND

21. Systems analysis

LANGEFORS (1973, p. 35) defines a system as a collection of objects, called parts, which are correlated in some way. According to this definition almost anything can be a system. However, the concept »system» is mainly reserved to describe the whole of a system and the relations between the parts of a system, but not the properties of the parts of the system.

Systems can be simply described diagrammatically. The nodes represent the objects of the system and the relations between these objects are described by arrows, Fig. 2.1.

Fig. 2.1. shows that certain nodes of the graph are precedents of other nodes. This relation is called a precendence relation. Respectively, some nodes are succedents of others. This relation is called the succedence relation. The precedence (succedence)

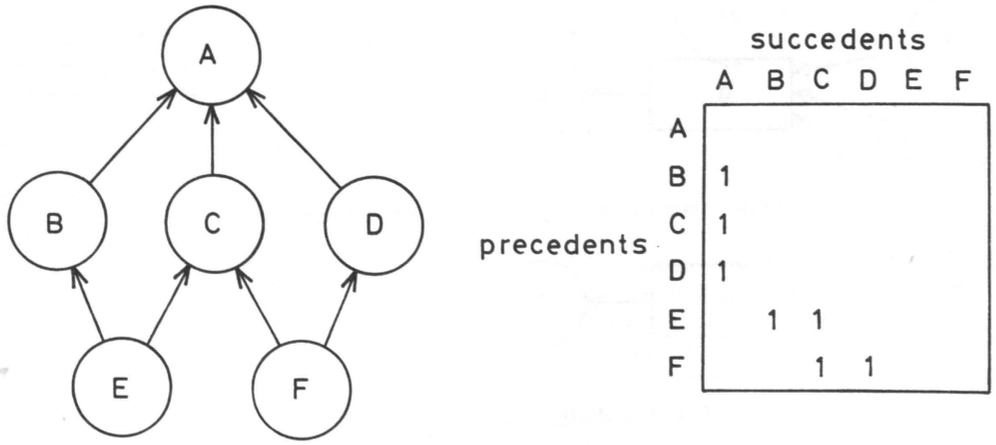


Fig. 2.1. Diagrammatic and matrix representations of a system.

relation can be described also by a precedence matrix.

Precedence matrices are useful in the description of a large system, because they make possible the algebraic analysis of the system. Thus, for instance, the precedents of the precedents can be determined by multiplying the matrix by itself.

An information system is a system of information sets needed for decision and signalling in a larger system called the object system (LANGEFORS 1973, p. 195). The object system can be an organization or enterprise, for instance. Data and algorithms are the central concepts in the description of information systems. The value of a variable, a vector or an observation matrix are examples of data. Also the precedence matrices belong to the data. The algorithms operate on the data to produce new sets of data. Thus, for example, tree volume is calculated from tree measurements by such algorithms as volume functions.

An example of an information system and the corresponding precedence matrix are given in Fig. 2.2.

The relevance of any information in the information system can be determined by precedence analysis. The precedence analysis also helps to find the most effective methods to derive certain information.

22. Models

221. Classification of the models

It is possible to define model as anything that represents something else (KLEIJNEN 1974, p. 3). In the present paper only abstract models consisting of mathematical symbols are considered. These models are used as algorithms of the information system to produce new data from existing data. The models have to be in an operational form, i.e. they have to be applicable to practical calculations.

Models are employed for two reasons: (1) to save work in the measurement of certain variables and (2) to estimate variables whose measurement is impossible. The first case is typical in ordinary forest mensurational problems. There are no theoretical difficulties in the measurement of the upper stem diameters, for instance, but the costs involved are so high that measurements are frequently replaced by models which yield estimates for the upper diameters. There are also examples of the second case in forest measuration. The direct measurements of the volumes of the different parts of the tree are impossible without killing the tree. Moreover, all variables which describe some future events must be predicted by models.

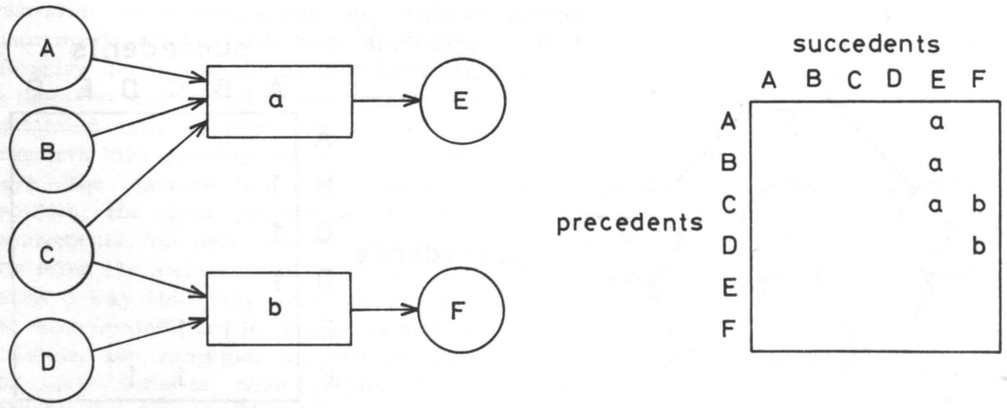
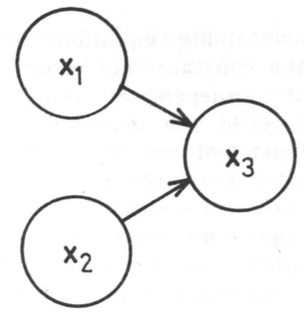


Fig. 2.2. Example of an information system.

Single-equation model:

$$x_3 = a + bx_1 + cx_2 + \varepsilon$$

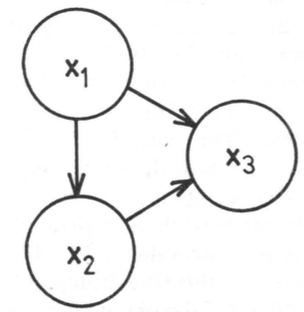


Multi-equation recursive model:

$$x_1 = a_{10} + \varepsilon_1$$

$$x_2 = a_{20} + a_{21}x_1 + \varepsilon_2$$

$$x_3 = a_{30} + a_{31}x_1 + a_{32}x_2 + \varepsilon_3$$



Simultaneous equation model:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \varepsilon_1 = 0$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2 = 0$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \varepsilon_3 = 0$$

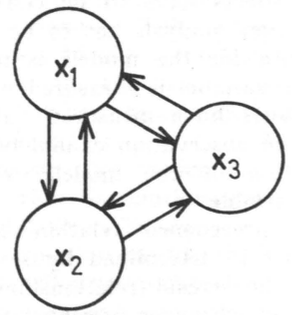


Fig. 2.3. Examples of different models.

RIIHINEN (1963, p. 12) has divided the mathematical models used in econometrics into the following categories:

1. Single-equation models
2. Multi-equation models
   21. Recursive models
   22. Simultaneous models

Examples of these models and the corresponding graphs which illustrate the precedence relations between the variables in different model categories are given in Fig. 2.3.

In the single-equation model, one and only one variable is the succedent of the other variables of the model. The preceding variables are called independent variables and the succeding variable the dependent variable. The error term (ε) can also be treated as an independent variable. One special class of single equation models is

that containing equations which include only the constant and error terms but no explicit independent variables. These models yield the marginal distribution of the dependent variable. Then, the estimation of the error term distribution has even stronger emphasis.

In the multi-equation recursive model one variable may be the dependent variable in one equation and an independent variable in another equation. However, the precedence relations are always in one direction, i.e. the same variable cannot both precede and succede another variable.

In the simultaneous equation model a variable can be both a precedent and succedent to another variable. Such variables are called intervening or endogenous variables. The variables with no precedents in the simultaneous equation model are called exogenous variables. Thus, the endogenous variables represent a collective dependent variable and the exogenous variables a collective independent variable.

In order to decide which type of model is the most suitable for a certain application, a thorough precedence analysis is needed. In this analysis both theoretical and practical aspects have to be considered. The precedence analysis has to be made before the data for the models is measured. A certain variable is measured only if all its precedents have measured values. Otherwise, the observation cannot be used in the derivation of the models which predict this variable.

The precedence relation between the variables is determined by several factors. First, the causal relationships should be honoured whenever possible. Secondly, the mensurational aspects have to be taken into account. Thirdly, the simplicity of the models is an important factor. Thus, for example, recursive models may be preferred even though the causal relationships would suggest simultaneuos models.

After the precedence analysis has been completed, it is still necessary to analyze the functional form of the models. Sometimes it is possible to determine the functional form a priori, before the data are measured. Then, the functional form is based upon an existing theory. More often, however, the exact functional form can only be determined after the actual data are available.

Whenever transformations are applied to the dependent variables, the effect is the same as if a new equation were introduced into a multi-equation model (cf. HOLM 1977). Thus, for instance, any single-equation model becomes a two-equation recursive model after the transformation of the dependent variable. If a logarithmic transformation is applied to the dependent variable (y), the derivation of the value of the dependent variable requires the following two-equation recursive model:

$$\log (y) = f(x)$$
$$y = e^{\log(y)}$$

where x represents the set of independent variables.

Another way to classify mathematical models is to see whether the dependent variable is quantitative or qualitative. The models with quantitative dependent variables have until now dominated in forest mensuration. Recent development of computers and estimation algorithms have also made the models with qualitative dependent variables accessible (e.g. BISHOP et al. 1975, NERLOVE and PRESS 1973).

222. Estimation of the parameters

First, it is assumed that the measurements upon which the models are based are correct. Then, ordinary least squares (OLS) estimation of the single-equation model parameters yields unbiased estimators for quantitative variables. This assumes proper precedence analysis, proper functional form of the model, and homogenous error variance of the dependent variable. The residual error variance is then due to the missing independent variables. The models consequently yield unbiased results when applied to the same parent population from which the data for the model has been measured.

If there are measurement errors in the variables, the estimation of the model parameters may become more complicated. If the errors occur in the dependent variables at random, the OLS method still gives unbiased models. In the multi-equation simultaneous models, however, errors in any variables are dangerous since a dependent variable in one equation may also be an independent variable in another equation.

If there are errors in the independent variables, the OLS method is no longer valid in the derivation of the models which exhibit the true relationships between the dependent and independent variables. For single-equation models, this distortion may not lead to biases if there are equal errors in the independent variables while the models are applied. However, the derivation of models from data in which the independent variables do not have exactly known values should be avoided.

Fortunately, forest measurements can be made very accurately compared to many other areas of measurement. In measuring trees almost an arbitrarily high degree of accuracy can be attained. Therefore, no more attention will be paid in this paper to the possible errors in the model data. The problem of errors in the measured values should not, however, be overlooked. In econometrics, for instance, a whole branch of statistics has been developed to deal with this problem.

In recursive models it might appear attractive to employ the estimated values of the dependent variables as values of the independent variables in the succeeding equations. The model would then give unbiased estimates when directly applied to a situation where the values of the respective variables are derived by the model. However, the model may give biased results wherever the independent variables assume values with a different error distribution than that of the model. This happens, for instance, if the independent variables assume exact measured values. In order to have consistent models, exact values of the independent variables should be also used when the parameters of the recursive models are calculated. In the following section it will be shown how the possible biases in the results of the unbiased multi-equation models can be avoided.

The models are not usually exact, but there is an error term. Therefore, due attention must be paid to the estimation of the parameters of the error term. Usually, the distribution of the errors are assumed to follow a normal distribution, with zero expected value and with certain variance. If the normality assumption does not hold, the description of the error distribution usually requires more parameters, higher order central moments, for instance.

If the assumation that the residual errors of the dependent variable are homogenous does not hold, and this may be especially true when the dependent variable is qualitative, OLS estimation no longer gives unbiased models. Then, more complicated methods, such as weighted linear regression analysis, nonlinear regression analysis, etc. must be employed.

223. Use of the models

2231. Analytical methods

The main purpose of the mathematical models in forest mensuration is not to verify hypotheses of the relations between different variables but to provide means for estimating distributions for those variables whose values cannot be measured directly. The distributions of the variables can be condensed into parameters such as sums, means, and variances. This condensation is often necessary to aid comprehension of the distribution. On the other hand, it is possible to later derive the real distributions from these parameters.

It is now shown how to derive unbiased estimates for the distribution parameters of the quantitative variables.

With single-equation models the calculation of the distribution parameters of the quantitative dependent variable may be quite straight-forward. If the model is as follows:

$$y = f(x_1, \ldots, x_k) + \varepsilon_y \qquad (2.1)$$

where y = dependent variable
x_i = independent variable i
ε_y = error term of the equation

and the values of the independent variables are known exactly, the expected value of the dependent variable can be calculated from the following formula:

$$E(y) = f(\hat{x}_1, \ldots, \hat{x}_k) + E(\varepsilon_y) \quad (2.2)$$

where $\hat{x}_i$ = true value of x_i

If the model is unbiased the expected value of the error term, or the first central moment (m_1) equals zero. The higher central moments can be calculated easily:

$$m_2 = E(\varepsilon_y^2) \quad (2.3)$$

$$m_3 = E(\varepsilon_y^3) \quad (2.4)$$

etc.

If the central moments have to be estimated from a sample, then the third and higher order central moments have to be replaced by Fisher's (1928) k-statistics.

The first ordinary moment (M_1) equals the expected value of the dependent variable, and the higher ordinary moments can be derived from the central moments:

$$M_2 = M_1^2 + m_2 \quad (2.5)$$

$$M_3 = 3M_2M_1 - 2M_1^2 + m_3 \quad (2.6)$$

$$M_4 = 4M_3M_1 - 6M_2M_1^2 + 3M_1^4 + m_4 \quad (2.7)$$

etc.

If the values of the independent variables are not known exactly, or the error term of the model is correlated with the independent variables, the use of model (2.1) in the derivation of the distribution parameters of the dependent variable becomes more complicated. If the equation is continuous with respect to all its derivatives, the expected value of the dependent variable can be derived by the application of the Taylor's series:

$$E(y) = f(\hat{x}_1, \ldots, \hat{x}_k) + \sum_{r_1=1}^{R_1} \cdots \sum_{r_k=1}^{R_k} \left(\prod_{i=1}^{k} \frac{1}{r_i!} \right) \frac{\delta^{\sum r_i} y}{\delta x_1^{r_1} \ldots \delta x_k^{r_k}} m_{r_1 \ldots r_k} + E(\varepsilon_y) \quad (2.8)$$

where R_i = degree of the equation with respect to x_i

$m_{r_1 \ldots r_k} = r_1 \ldots r_k{}^{th}$ central moment of vector x

The error term of the equation (ε_y) is separated from the independent variables, but if it covariates with them it must be handled as another independent variable. Independent variables may also be qualitative, but then they must assume exact values.

The calculation of the higher order ordinary moments requires equation (2.1) to be first raised to the respective power and the new equation introduced to formula (2.8). The central moments are obtained by solving equations (2.5, . . . , 2.7, etc).

Formula (2.8) becomes computationally easier if some assumptions concerning the distribution of the deviation of x_i:s from their expected values can be made. If the deviations around the expected values of x_i:s are symmetrical, only the central moments with even number as a sum of r_i:s need to be taken into account. Furthermore, if it can be assumed that there exists no covariation between the deviations of the x_i:s, only the even values of r_i:s need to be taken into account.

Some examples of the use of formula (2.8) are now given. The first of the two formulas indicates the original equation and the second formula gives the expected value of the dependent variable or the first ordinary moment of its conditional distribution as a function of the expected value and central moments of the independent variable.

$$y = x + \varepsilon_y \quad (2.9\,a)$$

$$E(y) = \hat{x} + m_1 + E(\varepsilon_y) \quad (2.9\,b)$$

$$y = x^2 + \varepsilon_y \quad (2.10\,a)$$

$$E(y) = \hat{x}^2 + 2\hat{x}m_1 + m_2 + E(\varepsilon_y) \quad (2.10\,b)$$

$$y = x^3 + \varepsilon_y \quad (2.11\,a)$$

$$E(y) = \hat{x}^3 + 3\hat{x}^2 m_1 + 3\hat{x}m_2 + m_3 + E(\varepsilon_y) \quad (2.11\,b)$$

$$y = x^4 + \varepsilon_y \quad (2.12\,a)$$

$$E(y) = \hat{x}^4 + 4\hat{x}^3 m_1 + 6\hat{x}^2 m_2 + 4\hat{x}m_3 + m_4 + E(\varepsilon_y) \quad (2.12\,b)$$

$$y = x^{1/2} + \varepsilon_y \quad (2.13\,a)$$

$$E(y) = \hat{x}^{1/2} + 1/2\hat{x}^{-1/2} m_1 - 1/8\hat{x}^{-\frac{3}{2}} m_2 + \frac{1}{48}\hat{x}^{-\frac{5}{2}} m_3 - \ldots + E(\varepsilon_y) \quad (2.14\,b)$$

$$y = e^x + \varepsilon_y \quad (2.15\,a)$$

$$E(y) = e^{\hat{x}} + e^{\hat{x}} m_1 + 1/2e^{\hat{x}} m_2 + 1/6e^{\hat{x}} m_3 + \ldots + E(\varepsilon_y) \quad (2.15\,b)$$

The term with m_1 gives the effect of bias, the term with m_2 gives the effect of variance, etc. If it can be assumed that the original models are unbiased, term $E(\varepsilon_y)$ cancels out from the expected value formulas.

The derivation of unbiased expected values of the dependent variable after transformations is identical to the previous examples with only one independent variable. If a logarithmic transformation, for example, is applied to the dependent variable before regression analysis, the logarithmic value of the dependent variable yielded by the regression equation can be transformed back to its antilogarithm by using formula (2.15 b). Thus, formula (2.15 b) is a generalization of the well-known correction formula for the case of normally distibuted logarithmic residuals (see MEYER 1941).

If it can be assumed that the residuals of the dependent variable of the regression equation are normally distributed after a nonlinear transformation, only the even order central moments have to be taken into account in the derivation of the unbiased expected values for the untransformed dependent variable. The higher even order central moments of the normal distribution can be calculated from the following formula (cf. FRASER 1976, p. 227).

$$m_{2n} = (2n-1)\,(2n-3)\cdots 3\cdot 1\cdot m_2^n \quad n = 2, 3, 4, \ldots \quad (2.16)$$

Next, two examples involving two independent variables (x and z) are given.

$$y = xz + \varepsilon_y \quad (2.17\,a)$$

$$E(y) = x\hat{z} + m_{1_x} + m_{1_z} + m_{1_x 1_z} + E(\varepsilon_y) \quad (2.17\,b)$$

$$y = x^4 z^3 + \varepsilon_y \quad (2.18\,a)$$

$$E(y) = \hat{x}^4\hat{z}^3 + 4\hat{x}^3\hat{z}^3 m_{1_x} + 6\hat{x}^2\hat{z}^3 m_{2_x} + 4\hat{x}\hat{z}^3 m_{3_x} + \hat{z}^3 m_{4_x} + 3\hat{x}^4\hat{z}^2 m_{1_z} + 12\hat{x}^3\hat{z}^2 m_{1_x 1_z} + 18\hat{x}^2\hat{z}^2 m_{2_x 1_z} + 12\hat{x}\hat{z}^2 m_{3_x 1_z} + 3\hat{z}^2 m_{4_x 1_z} + 3\hat{x}^4\hat{z} m_{2_z} + 12\hat{x}^3\hat{z} m_{1_x 2_z} + 18\hat{x}^2\hat{z} m_{2_x 2_z} + 12\hat{x}\hat{z} m_{3_x 2_z} + 3\hat{z} m_{4_x 2_z} + \hat{x}^4 m_{3_z} + 4\hat{x}^3 m_{1_x 3_z} + 6\hat{x}^2 m_{2_x 3_z} + 4\hat{x} m_{3_x 3_z} + m_{4_x 3_z} + E(\varepsilon_y) \quad (2.18\,b)$$

The higher order central moments of the dependent variable may be of interest when single-equation models are employed. The information they yield is indispensable when using multi-equation models. In the multi-equation model the dependent variable may be an independent variable in a succeeding equation. In this case, it is necessary to know the parameters or the moments which describe its variation and covariation with the other independent variables.

When multi-equation recursive models are used to derive parameters which describe the conditional distributions of the dependent variables, the calculation proceeds in accordance with the precedence relations. At each stage the necessary steps have to be taken to calculate all the moments required in the succeeding stages. The calculations may become quite overwhelming with large and complicated models. A judgement has then to be made up which order it pays to compute the moments. Quite frequently only the means, variances, and covariances are needed.

The derivation of results from simultaneous equation models is more difficult than from single-equation models or multi-equation recursive models. The calculations are still relatively easy if the simultaneous equations are linear. Then, linear matrix operations can be employed to derive the ordinary

moments of the conditional distributions of the endogenous variables.

The linear simultaneous equation model can be written by using the matrix notations:

$$Ay + Bx + \varepsilon_y = 0 \qquad (2.19)$$

where A = coefficient matrix of the endogenous variables
y = endogenous variable vector
B = coefficient matrix of the exogenous variables
x = exogenous variable vactor
ε_y = error vector of the equations

By multiplying the matrix equation (2.19) by the inverse matrix A^{-1} we get:

$$y = -A^{-1}Bx - A^{-1}\varepsilon_y \qquad (2.20)$$

If the equations are unbiased and the values of the exogenous variables are known exactly, the expected values of the endogenous variables can be calculated from the following formula:

$$E(y) = -A^{-1}B\hat{x} \qquad (2.21)$$

If there are nonlinear transformations in the exogenous variables and the values of the exogenous variables are not exactly known, the exogenous parts of the original equations have to be replaced by equations corresponding to formula (2.8). The resulting vector is then inserted into formula (2.21) to replace vector $B\hat{x}$.

The expected values of the variances (V_i) of the endogenous variables can be calculated from the following formula:

$$E(V_i) = \sum_{k=1}^{n} a_{ik}^2 m_{2_k} + 2\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} a_{ik}a_{il}m_{1_k1_l} \qquad (2.22)$$

where
a_{ij} = ij^{th} element of matrix $-A^{-1}$
m_{2_k} = error variance of equation k
$m_{1_k1_l}$ = covariance of the errors of equations k and l
n = number of equations (= number of endogenous variables)

The expected values of the covariances (cov_{ij}) between the endogenous variables can be calculated by the following generalized formula:

$$E(cov_{ij}) = \sum_{k=1}^{n}\sum_{l=1}^{n} a_{ki}a_{jl}m_{1_k1_l} \qquad \begin{matrix} i = 1, \ldots, n \\ j = 1, \ldots, n \end{matrix} \qquad (2.23)$$

The third central moments (m_{3_i}') can be derived from the following formula:

$$E(m_{3_i}') = \sum_{k=1}^{n} a_{ik}^3 m_{3_k} + 3\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} a_{ik}^2 a_{il} m_{2_k1_l} + 3\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} a_{ik}a_{il}^2 m_{1_k2_l} + 6\sum_{k=1}^{n-2}\sum_{l=k+1}^{n-1}\sum_{p=l+1}^{n} a_{ik}a_{il}a_{ip}m_{1_j1_k1_l} \qquad i = 1, \ldots, n \qquad (2.24)$$

The symbols are the same as used in formulas (2.8) and (2.22). It should be noticed that if the number of endogenous variables is less than three, the last term disappears. Similarly, formulas for any higher order central moments can be derived.

If the equations in the simultaneous model are nonlinear with respect to the endogenous variables, the solution of the system of equations becomes more complicated. In fact, no analytical solution usually exists. However, several iterative techniques and computer algorithms are available to solve the problem (see e.g. Brown and Dennis 1971; Imsl 1977). The Jacobian matrix contains the first derivatives of the equations with respect to all endogenous variables:

$$J = \begin{pmatrix} \frac{\delta y_1}{\delta y_1}, \ldots, \frac{\delta y_1}{\delta y_n} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\delta y_n}{\delta y_1}, \ldots, \frac{\delta y_n}{\delta y_n} \end{pmatrix} \qquad (2.25)$$

The Jacobian matrix of the system of simultaneous linear equations equals the coefficient matrix A (cf. formula 2.19).

The estimates of the second and higher order central moments of the distributions of the endogenous variables can be calculated by using the inverse of the Jacobian matrix (2.25) in the respective formulas (e.g. 2.22, 2.23, and 2.24). The information provided by these central moments is necessary if the system of equations is nonlinear (cf. Kilkki and Varmola 1979). The equations then have to be expressed in the form of formula (2.8). Since the Jacobian matrix may change with the change of the values of the endogenous variables, the central moments should, in theory, be calculated anew with each round of iteration. Fortunately, the central moments usually change rather slowly, and they need to be updated only after k iterations.

If the dependent variables are qualitative, the mathematical model yields the probabilities that the dependent variables assume certain values. Thus, for instance, a logistic model used by Walker and Duncan (1967) yields the probability of the presence or absence of disease. If the estimated qualitative variables are used as independent variables for the succeeding model, either the population has to be divided into subpopulations according to the values of these variables or the Monte Carlo method described in the following section has to be employed.

2232. Monte Carlo method

Even though the basic requirements, the continuity of the functions and the exact values of the qualitative independent variables, have been fulfilled, derivation of the population parameters by the methods described in the previous section may prove to be quite laborious for large multiequation models. To be sure that no important moments are missing, an overwhelminly large number of them should be taken into account. Moreover, there are situations in which the population cannot be described by a few parameters such as means and variances. Quite frequently, it would be more practical to handle the population elements instead of the parameters of the population in further analysis. It is also possible that the methods described in section 2231 may require an excessive amount of computer time, even though this point will lose its importance given the prevalent trend in computational costs.

An alternative method is therefore required to derive results from mathematical models. Such is the Monte Carlo method (see e.g. Hillier and Lieberman 1967). In the Monte Carlo method, the theoretical distribution described by a mathematical model is decomposed into a number of population elements by taking a sample of the distribution. Since the sampling is done by computer, all variables of the sample elements assume exact »measured» values, and the size of the sample may also be relatively large compared to the samples taken in real conditions.

The mathematical model describes the population as a n-dimensional distribution, where dimension n expresses the number of variables in each population element. The distribution of each variable is described as a conditional distribution of its precedents. The Monte Carlo sampling progresses in accordance with the precedence relations. After the precedents of a certain variable are known, the value of this variable is sampled from its conditional distribution. The conditional distributions can be determined easily because the precedents assume exact values. In a typical case the conditional distribution equals the distribution of the residuals of a regression model.

In its simplest form the Monte Carlo sampling is unrestricted random sampling. The value of the variable is taken randomly from its conditional distribution as determined by its precedents. The conditional probability density function determines the probability that the variable gets a certain value.

As in the sampling of real populations (see e.g. Nyyssönen et al. 1967), the unrestricted random sampling is not usually the most efficient sampling technique in the Monte Carlo method. A number of variance reduction techniques in the Monte Carlo method have been developed (Kleijnen 1974). Among these techniques are stratified sampling, selective sampling or

fixed sequence techniques, control variates or regression sampling, importance sampling, antithetic variates, and common random numbers. The most interesting of these techniques in our case are stratified sampling and antithetic variates.

The antithetic variate technique tries to create negative correlations between the observations. One observation is generated from a random number (r) taken from interval 0 ... 1, and its antithetic pair from random number (1-r). If the values of the variable are normally distributed around its expected value, the two observations are located at equal distances from the expected value. The same is true for any symmetric distribution.

For multi-dimensional distributions two sequences of random numbers $r_1, \ldots, r_n$ and $(1\text{-}r_1), \ldots, (1\text{-}r_n)$ have to be developed. The error variance of the mean of any variable is calculated as the variance of the means of the antithetic pairs.

The use of the antithetic variates may not be useful with qualitative variables or if relations between the variables are non-monotonic (see KLEIJNEN 1974, p. 189). Since there is no harm of using the antithetic variates technique even with qualitative variables and since most of the relations between the quantitative variables are monotonic in forest mensuration (e.g. h = f(d)), there is little danger of increased variance if antithetic variates are employed for all randomly sampled variables.

One major advantage of the Monte Carlo method is the fact that the method itself takes care of possible biases arising from nonlinear transformations of the dependent variables. Thus, for example, no correction to the constant term of the regression model due to logarithmic transformation (see e.g. MEYER 1941) is required. Since the conditional distribution of the transformed values often follows the normal distribution, the description of the distribution is also easy.

In simultaneous equation models, the sample must be taken from a multi-dimensional conditional distribution of the endogenous variables. If the endogenous variables are continuous and if their conditional distributions with respect to the exogenous variables are normal, the multivariate normal distribution can be employed for sampling (see e.g. WILKS 1962, p. 164).

After the sample from the distribution is taken, the frequencies of the sample elements are summed. The total frequency of the population is divided by this sum and the result is used as a multiplier to calculate the true frequencies of the sample elements. The population parameters are calculated by standard algorithms from the artificial observation matrix derived by the Monte Carlo method.

3. MODEL OF THE FOREST

Forest stand and/or tree populations are the objects of forest mensuration. In covering the whole area inside the boundaries of a forest area, compartments which do not belong to forest land are also frequently included in the stand population.

The elements of the stand population, the stands, are characterized by a number of variables describing the location, site, and growing stock. The tree population elements, the trees, may possess all these stand variables besides such individual variables as tree species, quality of the tree, various diameters, height, etc. The frequency attached to each stand indicates the area of the stand. The frequency of the tree population element usually indicates the number of trees, but it may also indicate the basal area or volume or some other quantitative measure.

Sometimes it is feasible to measure all interesting variables of all elements of either stand or tree populations. More often, however, the measurements have to be limited to only a small fraction of the whole population. Even from that sample, only part of the interesting variables can be measured. Therefore, we need an information system which reconstructs from these partial measurements, a model of the orginal population. The first task is to study the precedence relations of the

variables of the stand and tree population elements. Since all stand characteristics are also tree characteristics, discussion can be limited only to the precedence relations between the tree characteristics.

Some relations between the tree characteristics are clearly causal. For example, the environment and the age of the tree, as well as the past development of the tree affect the growth of the tree. Quite frequently, however, the relations between the tree characteristics are noncausal. For example, it is not possible to say that dbh precedes the height or vice versa. Thus, there are features which favour both recursive and simultaneous models in the description of the relations between the tree characteristics.

Unfortunately, it is not possible to base the models only upon causal relations. It must be remembered that to estimate the parameters of the model, the values of all independent and dependent variables have to be measured (see p. 372). If a multi-equation simultaneous model for instance, is chosen, all endogenous (as well as exogenous) variables have to be measured from the sample trees, otherwise the parameters of the model cannot be estimated. Thus, a model supported by the causal precedence analysis may not be chosen, but measurement costs must also be taken into account. Then dbh, for example, precedes height in field measurements.

Finally, model simplicity also forms a relevant basis for making choices between models. One might argue that single-equation models offer the simpliest model structure. Given unbiased models and correct measurements of the independent variables, the single-equation models do give unbiased estimates. Thus, the volume of all tallied trees might be estimated directly by an equation in which only the variables measured from all trees, tree species and dbh, for instance, would be the independent variables.

The use of the single-equation models has its drawbacks. In some cases equations become quite awkward and complicated since they are aggregates of several interdependencies. Moreover, the timber assortment models, for instance, are not continuous with respect to the breast height diameter. The number of sample trees may also become too small for some models since only those trees with sufficient measurements can be employed. The single-equation volume increment models, for instance, must incorporate in themselves the height and taper of the tree. However, sample trees with only height and taper measurements but without increment measurements cannot be utilized at all.

The variables predicted by single-equation models may also contradict with each other unless strictly linear relations prevail between the variables. For instance, if the height, taper class ($dbh\text{-}d_6$), and volume of the tree are predicted by their own equations in which tree species and dbh are independent variables, the predicted values do not represent any real trees.

As a flexible approach, a blockwise simultaneous recursive model (BSRM) is suggested in this paper. This model is a combination of simultaneous models which are arranged in a recursive order. The simultaneous parts of the model take advantage of the causal nature of the relations between the stand and tree characteristics. The overall recursive organization allows the use of cost saving multi-phase sampling techniques in which the most difficult measurements are restricted to a small number of sample trees and ultimately may be totally avoided in favour of the existing models.

The more numerous the blocks are the closer the model becomes to a pure recursive multi-equation model. In the extreme, there is only one equation per block and the model is completely recursive. Furthermore, since all dependent variables of the preceding equations do not have to be employed as independent variables in succeeding equations, a set of single equations is an extreme case of a multi-equation recursive model. Thus, the blockwise simultaneous recursive model, in fact, covers all the models and combinations described in section 221.

It would require a thorough cost-benefit analysis to decide the optimal division of the blocks and the optimal sample size in each block. The age of the tree, for instance, is an important precedent to a number of tree characteristics. However, its measure-

ment by boring is costly and, consequently, the age cannot in most cases precede the height, not to mention dbh.

In the following, a provisional block arrangement is suggested. It results from a priori known precedence relations between the variables. For the time being, only the tree characteristics related to the stem are included.

The first block of the simultaneous equations is reserved for the stand characteristics. These are obtained mainly by ocular estimation. In an extreme case, all stand characteristics are estimated directly and consequently there are only exogenous variables in the simultaneous model of the first block. The employment of remote sensing and two-phase sampling represents another extreme. Then, the stand characteristics estimated from the aerial photographs, for example, are exogenous variables and the variables measured on ground are endogenous variables.

Some relationships between the stand characteristics are deterministic and cause no problems. Most of the relationships are stochastic and since a number of the endogenous variables are measured either by nominal or order scale, ordinary regression analysis cannot be employed. Because of the computational difficulties involved in the estimation of the models with several qualitative polytomous variables (Nerlove and Press 1973) it may be useful to break the large simultaneous models into smaller parts. Also the qualitative variables should be avoided whenever practicable. Thus, a continuous variable should not be broken into discrete classes. As a result of the first block, a multidimensional probability distribution of the endogenous variables emerges.

In the second block, the dbh distribution is estimated. As with any variables, the breast height diameters can be measured directly; they are then exogenous variables of the model and the solution is trivial. If the number of dbh measurements is too small to estimate the dbh distribution conditioned by the stand characteristisc as a discrete distribution, some theoretical distribution has to be employed. Beta-distribution is the most commonly used distribution for breast height diameters (cf. Loetsch *et al.* 1973). Beta-distribution needs four parameters which are often strongly intercorrelated. Thus, the derivation of the parameter estimates suggests the use of a simultaneous model. The endogenous variables of the model are the four parameters and the exogenous variables are the stand characteristics; mean diameter being the most useful one. It is also advisable to try certain transformations of the dbh in order to make the distribution closer to the normal distribution. The estimation of the parameters is then more accurate. One simple transformation is to square the diameters.

In the third block, the conditional distributions of height, the upper diameter (d_6), and bark thickness at breast height are calculated. Such transformations should be applied to the endogenous variables that make their conditional distribution multinormal. The most important exogenous variable is certainly the breast height diameter, but all stand characteristics are also potential exogenous variables.

In the fourth block, the endogenous variables include i_d, i_h and the age of the tree. These variables should also be transformed to produce multinormal conditional distributions.

In the fifth block, the endogenous variables might be relative-height diameters above bark (cf. Kilkki *et al.* 1978), and the relative-height diameters under bark.

The diameter increments at various relative heights can be obtained by employing the Jacobian matrices of the models of the fifth and sixth blocks. However, more accurate results can be derived by estimating the diameter increments separately in the seventh block. Then, the influence of the environment can be taken into account more accurately.

The eighth block consists of an interpolation formula which connects the relative-height diameters and diameter increments to form continuous taper and diameter increment curves.

In the ninth block, models to divide the stem into assortments are employed.

In the tenth block the volumes of different parts of the stem are calculated by general integration formulas.

The blockwise simultaneous recursive model yields, at each stage, the conditional distributions of the endogenous variables with respect to the exogenous variables. To activate the model, the values of the exogenous variables in any of the blocks are required. Thereafter, the conditional distributions of the succeeding variables can be estimated. Of course, the reliability of the results is improved whenever measured values can be applied. The employment of the measured values must be blockwisely in the recursive order.

4. APPLICATION OF THE METHOD

The application of the blockwise simultaneous recursive model (BSRM) can be divided into two parts. In the first part, the measurements are used to adjust the parameters of the model to the actual population. In the second part, the model is activated by inserting into it the values of the exogenous variables of the first relevant block. The results are then calculated.

Theoretically, the BSRM as a whole could be derived from the actual measurements in a certain forest mensuration job. In practice, however, the derivation of all parts of the model is in most cases impracticable. First, the task requires too many measurements, and secondly, people working with practical forest mensuration problems seldom possess sufficient skills to derive the whole model. Consequently, an existing model is a necessary prerequisite of the application of the BSRM. Otherwise, more traditional methods should be employed.

Given an existing BSRM, it can be activated from any block by inserting into the model the exogenous variables of the block. Thereafter, it is possible to calculate the parameters which describe the conditional distributions of all succeeding variables. However, if both the endogenous and exogenous variables of a certain block are known for some observations, these measurements can be utilized in adjusting the basic model to the population under surveillance.

There are three main ways to utilize the measured observations of both the endogenous and exogenous variables: (1) the measured values of the endogenous variables are employed as such to form a discrete distribution, (2) completely new equations are derived from the measured values, and (3) original functional forms are used, but one or more parameters of the equations are estimated from the measured values. Dbh-distribution of the sample trees is a traditional example of the first case. Height curves are often based solely upon the sample tree measurements and represent the second case. In the third case, frequently, only the constant term of the equation is changed.

To optimize the number and quality of the measurements, the principles of sequential sampling should be utilized. The number of observations and their deviation from the expected values derived by the existing BSRM determine whether it pays to make any changes in the existing model.

A computer system has to be built to handle the calculation of the results. This system may employ both the analytical methods and Monte Carlo method. The analytical methods are practical for the derivation of the results from some parts of the BSRM. As a general purpose algorithm, however, only the Monte Carlo method is serviceable.

The practicality of the Monte Carlo method greatly depends upon the number of observations required for a given accuracy. The most promising variance reduction techniques, in our case, seem to be a combination of stratified and systematic sampling, and the use of antithetic variates.

The discrete variables without any precedents form a natural basis for stratification. Each tree species, for example, may form one stratum. Continuous variables can also be used for stratification. If the class intervals are equally large, and the value of the variable is taken from the middle of the stratum, the sampling is equivalent to systematic sampling. The middle values of the dbh-classes are an

example of this kind of sampling. However, the merits of systematic sampling are evident only if the sample is large enough and sufficiently covers the whole range of the variable values. With small samples, the tails of the distribution in particular may be left without representatives if systematic sampling is employed.

Both stratification and systematic sampling have to be restricted to the first variables in accordance with the precedence relations. Otherwise, the number of observations becomes too large, or the succeeding variables will have too few values in the densest part of the distribution. When stratified and/or systematic sampling are applied, the frequencies of the observations may no longer be equal. The frequency of each element has to be expressed explicitely.

Random sampling with the use of antithetic variates is applied to the bulk of the variables.

The observation matrix produced by the Monte Carlo method is directly applicable to the calculation of any population parameters. In these calculations standard algorithms are used.

5. DISCUSSION

The data processing system for forest measurements described in this paper is still mainly a paper tiger. It is also possible that some of the ideas embodied in it will change in the ultimate test of practical data processing. Especially, the optimal division of the blockwise simultaneous recursive model (BSRM) into blocks may be different from the one given in chapter 3. Of course, it is not even necessary to build a complete BSRM to apply the ideas presented in this paper, but these ideas can be utilized in solving any partial data processing problems in forest mensuration.

Even though no complete BSRM can be presented to prove to applicability of the approach, a number of partial models have been developed in recent years. Eventually, at least some of these models can be incorporated into a BSRM. The models needed in the second block are under preparation. PÄIVINEN (1978) has published taper class (dbh-d_6) and breast height bark thickness equations for pine, spruce, and birch. Furthermore, unpublished equations for height, i_d, i_h, and age of the tree have been developed. These models, however, are purely recursive.

The greatest advance has been made in the development of the simultaneous equation models of the fifth block. Systems of both linear and nonlinear equations have been calculated to estimate the diameters at 11 and 10 relative heights of the tree (KILKKI *et al.* 1978; KILKKI and VARMOLA 1979). In these studies interpolation formulas for the seventh block have also been developed.

At present, a working system resembling the BSRM, has been built to calculate the main results of the national forest inventory of Finland. The system is based upon a multi-equation recursive model. The system is basically capable of inclusion in a larger system which would produce results from any forest measurements. However, at present it lacks the parts needed to estimate stand characteristics and dbh-distributions, as well as those parts needed to estimate diameters, bark thicknesses, and diameter increments at arbitrary heights of the stem. Stand characteristics and breast height diameters are measured exogenous variables. Traditional models based upon two fixed height diameters (dbh and d_6), height, i_d, i_h, and bark thickness at breast height are employed in the estimation of the volumes of different parts of stem.

The numerical results of the inventory are derived mainly by the Monte Carlo method. A combination of stratified and sytematic sampling, and the use of antithetic variates is employed.

In future, other parts of the tree in addition to the stem should be included in the model. There is a correlation between the form of the stem and the variables describing the crown. Similar relations may be found between the rootsystem and the parts of the tree above ground. Growth is also related to the crown and root characteristics. Consequently, a new evaluation of the block division may be necessary.

Another extension of the data processing system is the simulation of the future development of the growing stock. This question also belongs to our study program. The set of the trees derived by the Monte Carlo method serves as a starting point for the simulation of growth.

BSRM is not restricted only to the derivation of results from forest measurements. It is also a system to test data. When the conditional distributions of the variables are known, it is easy to find out exceptional values of the variables.

Most of the discussion has so far concerned the practical data processing problems encountered in forest mensuration. The methodology described in this paper is also applicable in the problems of growth and yield studies, in which the basic relations between the environment and the development of the trees are studied. For growth and yield studies, most probably, very large simultaneous equation models will be needed, and the application of differential equations will also be necessary.

REFERENCES

BROWN, K. M. & DENNIS, J. E., Jr. 1971. On the second order convergence of Brown's derivate-free method for solving simultaneous nonlinear equations. Yale University Department of Computer Science. Technical Report: pp 71—77.

FRASER, D. A. S. 1976. Probability and statistics. Theory and applications. Duxbury Press. North Scituate, Massachusetts, 623 p.

HILLIER, F. S. & LIEBERMAN, G. J. 1967. Introduction to operations research. Holden-Day, Inc. San Francisco, Cambridge, London, Amsterdam. 639 p.

HOLM, S. 1977. Transformationer av en eller flera beroende variabler i regressionsanalys. Hugin rapport nr 7, 21 p.

IMSL Library 2, Reference Manual, Edition 6. 1977. International Mathematical and Statistical Libraries, Inc. Houston, Texas.

KILKKI, P., SARAMÄKI, M. & VARMOLA, M. 1978. A simultaneous equation model to determine taper curve. Silva Fenn. 12(2): 120—125.

— » — & VARMOLA, M. 1979. A nonlinear simultaneous equation model to determine taper curve. Silva Fenn. 13(4): 293—303.

KLEIJNEN, J. P. C. 1974. Statistical techniques in simulation. Marcel Dekker, Inc. New York, 775 p.

LANGEFORS, B. 1973. Theoretical analysis of information systems. Fourth edition. Auerbach Publishers Inc. Philadelphia, Pa., 489 p.

LOETSCH, F., ZÖHRER, F. & HALLER, K. E. 1973. Forest Inventory II. BLV Verlagsgesellschaft, München, 469 p.

MEYER, H. A. 1941. A correction for a systematic error occuring in the application of the logarithmic volume equation. The Pennsylvania State Forest School, Res. paper No. 7, 3 p.

NERLOVE, M. & PRESS, S. J. 1973. Univeriate and multivariate log-linear and logistic models. Rand. Santa Monica, Ca., 134 p.

NYYSSÖNEN, A., KILKKI, P. & MIKKOLA, E. 1967. On the precision of some methods of forest inventory. Acta For. Fenn. 81.4: 1—60 p.

PÄIVINEN, R. 1978. Kapenemis- ja kuorimallit männylle, kuuselle ja koivulle. Summary: Taper and bark thickness models for pine, spruce and birch. Folia For. 353: 1—31.

RIIHINEN, P. 1963. Sales of newsprint in Finland 1949—1959. Models for short term forecasting. Acta For. Fenn. 74.7: 1—82.

WALKER, S. H. & DUNCAN, D. B. 1967. Estimation of the probability of an event as a function of several independent variables. Biometria 54: 167—179.

WILKS, S. S. 1962. Mathematical Statistics. John Wiley & Sons, Inc. New York—London, 644 p.

SELOSTE:

EHDOTUS METSÄNMITTAUSTULOSTEN LASKENTAMENETELMÄKSI

Tutkimuksessa esitellään metsänmittaustulosten laskentamenetelmä, joka kattaa valtaosan puun ja metsikön mittauksessa esiintyvistä laskentatilanteista. Ensiksi tutkitaan puuta ja metsikköä kuvaavien muuttujien edeltäjä-seuraajarelaatiot. Näitä relaatioita kuvaamaan laaditaan moniyhtälö-

malli (BSRM), joka koostuu rekursiiviseksi malliketjuksi järjestetystä joukosta simultaanisia yhtälöryhmiä. Mallin parametrit saadaan pääosin regressioanalyysillä. Tulosten laskenta mallilla edellyttää joko Taylorin sarjakehitelmän tai Monte Carlo simulointiin perustuvan korjausmenetelmän käyttöä harhojen poistamiseksi.

Laskentamenetelmän etuina ovat: (1) samoja matemaattisia malleja ja tietokoneohjelmiston osia voidaan käyttää lukuisten tunnusten laskentaan, (2) olemassaolevaa tietoa voidaan käyttää tehokkaasti hyväksi ja (3) lasketut puutunnukset ovat keskenään ristiriidattomia.

UUSI OPAS TUTKIMUSRAPORTIN LAATIJOILLE

Sitten professori V. T. Aaltosen vuonna 1945 julkaiseman teoksen »Tieteellinen tutkimustyö» metsäntutkijain käytettävissä ei ole ollut hyvää suomenkielistä opasta tutkimusraportin laadinnasta. Tästä syystä viime keväänä vt. professori *Matti Kärkkäisen* julkaisema käsikirja »*Tutkimusraportin laadinnan perusteita*»[1]) on tervetullut kirjoitus- ja painatustekninen opas kenelle tahansa julkaisun kirjoittajalle. Teoksessa on käsitelty kaikkia tärkeimpiä tutkijan eteen tulevia julkaisuongelmia kirjainlajeista ja piirrosmalleista tekijänoikeuskysymyksiin.

Uuden tiedon tuottaminen on monesti vaikea tehtävä ja vieläpä tutkimusraportin laadinta saattaa muodostua kompastukseksi. Puheena oleva teos on tarkoitettu ensisijassa käytännön oppaaksi aloitteleville tutkijoille ja opinnäytetöiden laatijoille, mutta varttuneempikin kirjoittaja löytää siitä runsaasti hyödyllisiä ohjeita mukaanlukien käsikirjoituksen tarkastuksen. Oppaan pääpaino on metsäntutkimuksessa, mutta monet siinä esitetyt näkökohdat ovat siinä määrin yleisiä, että ne sopivat esimerkiksi muidenkin alojen tutkijoille.

Opaskirjan ilmestymistä on tervehdittävä ilolla erityisesti siksi, että tutkimusraportin laadinnasta ei ole tutkijoiksi aikoville järjestetty varsinaista opetusta opinnäytetöistä saatavaa arviointia lukuunottamatta. Oppiminen on siten monesti tapahtunut erehtymisen ja virheiden kautta. Suomen Metsätieteellisen Seuran julkaisusarjojen entisenä toimittajana Kärkkäinen on seurannut läheltä julkaisemiseen liittyviä ongelmia, joten on varsin luonnollista, että hän on nähnyt välttämättömäksi auttaa kirjoittajia laajalla kokemuksellaan.

Kirjassa annetaan aluksi hyödyllisiä teknisiä ohjeita kirjoitusvaikeuksien voittamiseksi ja kirjoittamisen tuloksellisuuden parantamiseksi. Nämä vaikeudet lienevät tuttuja monelle aloittelevalle tutkijalle. Tutkimusraportin sisällölle ja muodolle asetettavat vaatimukset käydään yksityiskohtaisesti läpi. Tämän jälkeen Kärkkäinen esittelee havainnollisesti kirjallisuuden luokitusjärjestelmiä, joiden tunteminen on edellytys kulloinkin tarpeellisen tietoaineksen löytämiseksi. Tässä yhteydessä on erityisesti aiheellista mainita uudet vuoden 1978 alussa käyttöön otetut FA ja FPA järjestelmät. Muutoinkin kirjasta löytyy runsaasti hyviä ohjeita, jotka helpottavat tutkijaa löytämään kirjallisuutta tutkimusongelman tarkastelua varten.

Erityisen hyödyllisinä kokenee moni kirjoittaja taulukoiden ja piirrosten rakennetta ja havainnollisuutta koskevan yksityiskohtaisen tarkastelun. Samalla selvitetään piirrosten laadinnassa kirjapainon kannalta huomioon otettavia seikkoja sekä taulukoiden ja kuvien vaikutusta taittoon.

Ryhtyessään oppaan kirjoittamiseen Kärkkäinen on tarttunut vaikeaan tehtävään, vaikka ohjeiden antaminen on monesti helpompaa kuin niiden noudattaminen. Opas on kuitenkin pääosaltaan helppolukuinen ja sanomaltaan selkeä. Jossain määrin yllättävältä tuntuu kuitenkin se, että eräät yksittäiset esimerkit on jätetty melko triviaaleiksi muutoin perusteelliseen tarkastelutapaan nähden.

Tämä käsikirja olisi hintansa arvoinen jo pelkästään siihen koottujen liitteiden, kuten UNESCOn antamien julkaistavaksi tarkoitettujen tieteellisten kirjoitusten laatimisohjeiden, metsätieteellisten sarjojen lyhenteiden sekä mittayksiköitä koskevan standardin ansiosta. Lisäksi teokseen liitetty kirjapaino- ja kirjastoalan keskeisin terminologia helpottaa asiointia kirjapainon kanssa.

Eino Mälkönen

[1]) Matti Kärkkäinen: Tutkimusraportin laadinnan perusteita. Helsingin Yliopiston monistuspalvelu. Helsinki 1979.