

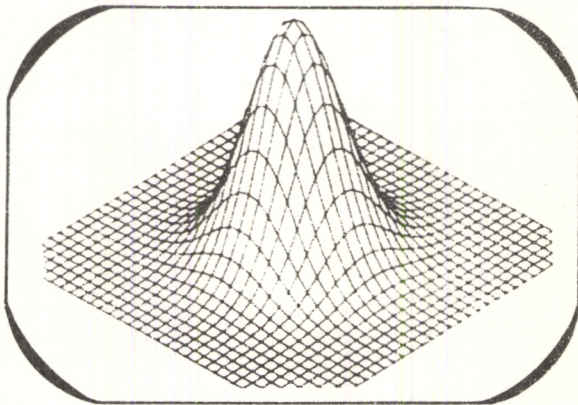
METSÄNTUTKIMUSLAITOKSEN  
TIEDONANTOJA 194

Matemaattinen osasto



# SPATIAALISEN TILASTOTIETEEN SOVELTAMINEN METSÄNTUTKIMUKSEEN

TOIM. Heinonen, J.,  
Penttinen, A.,  
Salminen, S.,  
Tomppo, E.



Helsinki 1985





Metsäntutkimuslaitoksen tiedonantoja 194

SPATIAALISEN TILASTOTIETEEN SOVELTAMINEN METSÄN-  
TUTKIMUKSEEN

SEMINAARIRAPORTTI

4.2.1985

KEILANIEMI, ESPOO

TOIM. HEINONEN, J.

PENTTINEN, A.

SALMINEN, S.

TOMPPU, E.



ISBN 951-40-0911-8  
ISSN 0358-4283

Valtion painatuskeskus 1985



## ALKUSANAT

Metsänarviointi on pioneerin asemassa niiden tutkimusalojen joukossa, joilla on systemaattisesti otettu huomioon tilassa tapahtuvaa vaihtelua (spatiaalista vaihtelua). Ongelmana on ollut paitsi kehittää tehokas otantamenetelmä myös virhearvion laatiminen systemaattisen otannan tapauksessa. Ruotsalaisen professorin Bertil Matérnin työt 40-luvun loppupuolelta alkaen ovat merkitykseltään keskeisiä spatiaalisen vaihtelun analysoimiseksi stokastisia malleja käyttäen. Tutkimusala on laajentunut 70-luvun aikana, jolloin myös spatiaalisen tilastotieteen käsite on otettu käyttöön. Kehitykseen ovat vaikuttaneet sekä haastavat käytännön ongelmat useilta eri sovellusaloilta että tietokoneiden kehitys. Spatiaalinen tilastotiede on ala, jolla tietojenkäsittely on keskeistä aineistojen analyysissä ja tulosten graafisessa esittämisessä.

Mikä sitten on ominaista spatiaaliselle tilastotieteelle? Kyseessä on tilassa tapahtuvan vaihtelun tilastollinen analyysi. Matemaattisen tilastotieteen termein tämä merkitsee riippuvien havaintojen analyysia tilanteessa, jossa riippuvuusrakenne ottaa havaintojen tilakoordinaatit huomioon. Esimerkkinä on satelliittikuva, jonka kuvanalkioista, "pikseleistä" tiedämme, mitkä ovat naapureita keskenään ja oletamme, että havaitut harmaansävyt vierekkäisissä pikseleissä ovat voimakkaasti toisistaan riippuvia. Mallinrakennuksessa tämä riippuvuusrakenne on otettava huomioon.

Spatiaaliset aineistot ovat jaettavissa luonnollisella tavalla kahteen luokkaan: ajan suhteen puhtaasti poikkileikkausaineistoihin ja ajasta riippuviin spatiaalisiin aikasarjoihin. Tyyppiesimerkki edellisestä on taimikkoa kuvaava

kartta ja jälkimmäisestä taas peräkkäisinä vuosina tehdyt kartat, joista ilmenevät taimikon tilassa tapahtuvat muutokset. Mallinrakennuksen lähestymistapoja on useita. Ongelmaan liittyvää matemaattista koneistoa esiintyy kirjallisuudessa mm. seuraavien otsikoiden alla: kuvananalyysi, matemaattinen morfologia, stokastinen geometria, stereologia ja matemaattinen ekologia. Tätä kirjavuutta voidaan pitää yhtenä vaikeutena.

Metsäntutkimus tuo esille useita spatiaalisen vaihtelun ongelmia. Edellä mainitun klassisen inventointiprobleeman lisäksi spatiaaliset tekijät esiintyvät jo mm. taimikon kehityksen malleissa ja metsän harvennismalleissa. Myös metsäntutkimuksen koesuunnitteluun voidaan spatiaalisen vaihtelun ajatus kytkeä. Temporaaalisia (ajasta riippuvia) ongelmia ovat mm. metsän kasvu ja moninaiset metsätuhoihin liittyvät tartuntamekanismit.

Satelliittikuviin perustuvat metsien tilaa kuvaavat analyysit ovat olleet Suomessa esillä jo 60-luvun lopulla - kansainvälisestäkin ottaen hyvin varhain - ja niiden käyttöön metsäntutkimuksessa liittyy paljon odotuksia. Toisaalta kaukokartoituksella hankitun aineiston ekonomiseen käyttöön liittyy avoimia kysymyksiä, joista osa kuuluu spatiaalisen tilastotieteen alaan. Tämä on haaste sekä tilastotieteilijöille että metsäntutkijoille, josta haasteesta ei selvitä ilman tehokasta yhteistyötä. Se on osa laajenevaa tietotekniikan hyödyntämistä.

Biologiassa on tunnettua, että diversiteetti on usein suurinta vaihettumisvyöhykkeissä kuten viljelysmaan ja metsämaan kohdatessa. Samoin on uusien ideoiden laita: innovatiivisuus voi olla vilkkainta tieteenalojen, tässä tapauksessa metsäntutkimuksen, matematiikan ja tilastotieteen kohdatessa. Tämä raportti on dokumentti yrityksestä tuoda



tilastotieteilijöitä ja metsäntutkijoita yhteiseen seminaariin, jonka tavoitteina on ollut tehdä tunnetuksi metsäntutkimuksen spatiaalisia ongelmia sekä metsäntutkimuksessa ja tilastotieteessä tehtyä alan tutkimusta. Uskomme esitelmien nopean julkaisemisen palvelevan myös näitä tavoitteita.

Seminaarin esitelmät kattavat melko laajasti spatiaalisen tilastotieteen kenttää: ajassa riippuvat spatiaaliset prosessit (Kuulasmaa), spatiaalisen kilpailun mallit (Penttinen), kokeensuunnittelu (Lahtinen), kuvananalyysi mukaanlukien kaukokartoituksen menetelmät (Arkimaa, Oja, Suomela, Tomppo), metsäninventointi (Ranneby, Salminen) ja spesifejä ongelmia (Niinimaa, Saarenmaa). Raportti sisältää myös kaksi innovatiivista keskustelupuheenvuoroa (Hari, Poso).

Seminaarin järjestäjinä ja raportin toimittajina olemme kiitollisia Suomen Akatemian maatalous-metsätieteelliselle toimikunnalle taloudellisesta tuesta sekä professori Risto Seppälälle seminaarin puheenjohtajana toimimisesta. Erityisesti lausumme kiitoksemme esitelmänpitäjille talkoovoimin tehdystä työstä.

Helsingissä 6.6.1985

Jaakko Heinonen  
Antti Penttinen  
Sakari Salminen  
Erkki Tomppo

ESITELMÄT	sivu
RANNEBY, B.: Spatial statistics in forest inventories	1
SALMINEN, S.: Metsien linjainventointien tilastotieteellinen tausta	17
PENTTINEN, A.: Spatiaalisista kilpailumalleista	27
LAHTINEN, M.: Lähimmän naapurin korrelaation huomioonottaminen koesuunnittelussa	38
NIINIMAA, A.: Eliminating edge effects in distance sampling	48
KUULASMAA, K.: Epidemian spatiaaliset tartuntamallit	57
SAARENMAA, H.: Tilajärjestys kaarnakuoriaisten kilpailun selittäjänä	70
SUOMELA, P.: Kuva-analyysi ranskalaiseen tapaan	80
OJA, E.: Tekstuurin luokittelu	89
TOMPPO, E.: Satelliittikuvien tulkintamenetelmiä	91
ARKIMAA, H.: Satelliittikuva-aineiston ja geofysikaalisen aineiston numeerisesta luokituksesta	116
PUHEENVUOROT	
HARI, P.: Metsäluonnon muuttuminen ja spatiaalinen tilastotiede	121
POSO, S.: Spatiaaliset menetelmät, uusi keino metsätalouden suunnitteluun	123



## SPATIAL STATISTICS IN FOREST INVENTORIES

Bo Ranneby

### 1. Introduction

Statistical methods and principles are an unvaluable tool in forestry. The following three methods are used very frequently.

1. Sampling
2. Design of experiments
3. Regression analysis.

To evaluate, in a proper way, the result from a statistical analysis some assumptions have to be fulfilled. In the beginning these assumptions were rather restrictive. Fortunately a lot of (but not all) statistical methods are robust - they give satisfactory results also when some assumptions are violated. As the time passed statistical methods based on more realistic and sophisticated assumptions were developed. The use of spatial statistics and spatial models is one example.

In forestry it is very important to be aware of the spatial influence on the studied phenomenon. Spatial models can be used in connection with all three methods mentioned above.

I shall concentrate myself on forest inventories which is an application of sampling. I shall also say some words about regression analysis but omit the topic about the theory for designing experiments.

## 2. Regression analysis

Mathematical models are often used in forestry. Suitable models are assigned and then certain unknown parameters are estimated. Very often the method of regression analysis is used to determine the parameters. Normally the value of  $R$  (the multiple correlation coefficient) is used to choose between different models. However, this method gives only a fair comparison when we have the same dependent variable in all models.

Because of that other criteria have been given to choose between different models, see for example Akaike (1974, 1977), Schwarz (1978), Stone (1974, 1977) and Geisser (1975). None of these criteria gives us any possibility to decide whether the chosen model gives a satisfactory description. A method giving us this possibility has been suggested in Ranney (1984). Most of the methods just mentioned are more or less related to the maximum likelihood method.

A method which, at first sight, seems very attractive is cross-validation. Let us assume that we want to determine a volume function according to some model. We then make a first computation using all observations except the first one. By aid of the function thus obtained and the values of the independent variables for the first stem we estimate a value  $\tilde{v}_1$  of the volume of the first stem. In the next step we omit stem number 2 and proceed similarly to form an estimate  $\tilde{v}_2$  of the volume of the second stem. We proceed

like this, using all observations except one. Finally, we compute the sum of squared deviations

$$(v_1 - \tilde{v}_1)^2 + (v_2 - \tilde{v}_2)^2 + \dots + (v_n - \tilde{v}_n)^2 \quad (2.1)$$

and use this as a basis for comparison. My opinion is that in this form the cross-validation method gives too much weight to large stems. The small stems are more or less neglected. However, the same technique can be used with another criterion (giving more weight to small stems) than (2.1) for the comparison.

All methods have certain advantages and disadvantages and they must be supplemented by residual studies. Moreover, it is very important that we try to use all our prior knowledge when the models are formulated. The prior knowledge shall be used to give a realistic description of the deterministic part of the model, but also to give a good description of the residual term. For instance, the data behind a volume function can often be "clustered" because some stems belong to the same stand or the same plot. Taking this spatial dependence into account leads to a more complicated procedure but gives also a better model and better possibilities to give correct statements about the reliability of the model.

### 3. Forest inventories

We have different types of forest inventories. I shall mainly discuss traditional inventories - inventories where we want to estimate volume, growth, depletion etc. - but I shall also make some comments about more specific inventories as estimation of road lengths and of animal populations.

### 3.1. Traditional forest inventories

It was within forest surveying that sampling first began to be applied. The method used in the beginning was strip surveys. Quite soon it was recognized that almost the same precision was achieved with a plot survey where the surveyed area was much smaller. The reason for this is the spatial dependence between neighbouring plots. Plots close to each other tend to be similar - they are positively correlated.

To construct a good sampling design we must be aware of the spatial dependence between plots. An efficient way of doing this is to study sampling and estimation procedures in mathematical model forests. By varying the fundamental properties of the model, one can find how the performance of different methods is affected by these fundamental properties. We have the possibility to compare a lot of different sampling designs. If we base our decision on empirical studies the number of possible designs in the comparison is much smaller. Moreover, the estimated error variances from empirical data may be rather imprecise. This effect especially the studies of systematic sampling (see Matérn & Ranneby (1983)). However, in order to obtain real useful results from mathematical studies, the results must hold for a large class of models (or the model must be very realistic). Unfortunately, this is not always the case as the following example illustrates.

Example: Suppose that we want to have an estimator of the number of plants per area unit of land. Assuming that the plants have been located according to a spatial Poisson process (we have a "Poisson forest") the expression  $0.25/\bar{x}^2$ , where  $\bar{x}$  denotes the average distance from  $n$  random points to their respective nearest plant, is a reasonable



estimator of the plant density. Although this estimator behaves well in the "Poisson forest", it can be seriously biased in patterns generated by other mechanisms. The estimator  $0.25/\bar{x}^2$  is too model-dependent to be useful.

Suppose that we want to make a statistical design for a forest inventory covering a large area. To do this properly we need a mathematical model describing the spatial variation.

As a first choice we assume that our spatial process is locally stationary and isotropic. The process is isotropic if the correlation is the same in all directions.

For a very large class of isotropic correlation functions the best sample design is a systematic equilateral triangular lattice. However a quadratic lattice performs almost equally well.

In practice, we also have to take the surveying costs into consideration and, when it is expensive to reach the plots, it might be meaningful to gather them into clusters. The size and shape of the cluster varies according to the circumstances. In the Swedish National Forest Survey (NFS) the clusters are squares and the sample plots are located systematically along the perimeters of the squares. These squares are called tracts and the size of the tract corresponds to one day's work.

A good approximation of the correlation functions of variables studied in the forest, at least for distances larger than 100 metres, can be obtained by linear combinations of exponential correlation functions; see Matérn (1947, 1960), Ohlsson (1975) and Ranney (1981a). We shall therefore consider exponential correlation

functions. (From these results we can easily derive the corresponding results for linear combinations of exponential correlation functions.)

Table 1. The variance of the mean of the plot observations of a tract with perimeter 4 kilometres and 4m sample plots. Covariance function  $C(v) = \exp(-hv)$ . (The symbol  $v$  denotes distance and is measured in kilometres, while  $h$  is a parameter.)

m \ h	0.25	1	4	8
2	0.8293	0.5042	0.1727	0.1301
3	0.8326	0.5080	0.1558	0.0971
4	0.8338	0.5096	0.1502	0.0842

We see from the table that when there is high correlation between the observations (low value of  $h$  in the table), the variance increases with increasing number of sample plots. This is quite unsatisfactory. To avoid this disadvantage we can use a weighted mean  $Z_a = \sum a_i Z(x_i)$ , where  $\sum a_i = 1$ . However, with high correlation the increase of information by new observations is small and consequently the optimal weighted mean will not be much better than the mean of equally weighted observations. Instead of using weighted means it is better to locate the sample plots so that all of them contain real information.

Of course we have to consider the effect of different plot sizes and the number of sample trees when the size and shape of the tracts is to be determined. All of the tracts we shall compare constitute one day's work. Thus if the

plot size is increased, we have to decrease the side length and/or the number of plots on the tract.

By analyzing the results from old inventories we found that the extra information obtained by measuring circular plots having a radius of 10 metres instead of 5 metres was rather small. For example to estimate volume/ha in Västerbotten in Sweden 135 "5-metres" plots will give the same precision as 100 "10-metres" plots.

The optimum plot size depends on the variation between plots compared to the variation within plots and the costs for measuring plots.

In the sixth NFS of Sweden which started in 1983 both temporary and permanent plots are used. The radius of the temporary plots is (essentially) 7 metres, while the permanent plots have a radius of 10 metres. Further details may be found in Ranney (1981b).

Studies of the type just described will give us a certain tract form. But an efficient location of the tracts is also necessary. If we still assume that the correlation function is isotropic the best precision is obtained if the tracts are located in an equilateral triangular lattice, but a quadratic lattice performs almost equally well.

However one often has reason to assume that the correlation is higher in some direction and then it is necessary to modify the assumption about an isotropic correlation function. Starting with an isotropic correlation function we can easily define a process which is such that the correlation is highest in a specific direction. We obtain processes having an elliptic correlation function.

Since a quadratic lattice is "almost optimal" for processes having an isotropic correlation function it follows from the construction of the "elliptic process" that a rectangular lattice with the correct ratio between the sidelengths and the correct orientation is "almost optimal".

In situations when we feel unsure about how much stronger the correlation is in a certain direction we may still use a quadratic lattice. However, if the quadratic lattice is not oriented in a proper way the difference to the "almost optimal" lattice may be large. If we locate the quadratic lattice so it has an angle of  $45^{\circ}$  with the "correlation-direction" a much better performance is obtained. Consequently, in situations when we have some vague idea that the correlation is higher in a certain direction a quadratic lattice of this type gives us a robust design. For further details, see Cruse et al. (1985).

So far we have only discussed inventory designs when no supplementary information is available. However, a design which is superior without auxiliary information is also superior with this information. For example, since a systematic (quadratic lattice) sample gives higher precision than a simple random sample this systematic sample gives also a better precision when auxiliary information (from aerial photos, say) is used.

A situation related to that with supplementary information is sampling on successive occasions. Using both temporary and permanent plots it is often possible to obtain better estimates. The optimal proportion of permanent plots depends on which quantity we want to estimate. If we are interested in changes of the volume from occasion 1 to occasion 2, all plots should be permanent. Contrarily, if



we want to estimate the mean of the volumes at occasion 1 and occasion 2, all plots should be temporary. If we want to estimate current volume it is optimal to use both temporary and permanent plots. The optimum percent of permanent plots and the gain in precision (compared with temporary plots only) depend on the time-correlation, which is closely related to the length of the rotation period. The gain is of some importance only if the correlation is rather high.

When estimating the amount of fellings the advantage with permanent plots is even more significant. In this situation the permanent plots give unbiased estimates with higher precision. It is doubtful if it is possible to obtain unbiased estimates from temporary plots only. To obtain a really high efficiency the time until the remeasurement of the permanent plots must not be too short.

An often reported drawback with sampling with partial replacement (SPR) is the complexity of the estimators and that we cannot intuitively understand the estimators. I have read in an article, that "SPR is a statistician's paradise. Even when dealing with the most simple cases the estimation theory is cumbersome". This is a gross exaggeration as the following example illustrates.

Example: Let us consider a situation where we have the following observations:

T(1)	temporary plots measured in year 1	
P(1)	permanent	-"-
T(2)	temporary plots measured in year 2	
P(2)	permanent	-"-

- T(3) temporary plots measured in year 3
- P(3) remeasurement of the permanent plots from year 1
- T(4) temporary plots measured in year 4
- P(4) remeasurement of the permanent plots from year 2

Thus we have 6 different samples.

Suppose that we want to estimate current volume/ha at year 4. We construct the following three estimates:

- I: T(4)
- II: P(4)
- III:  $T(2) + (P(4) - P(2))$

The final estimate is obtained by forming a weighted mean of these estimates. The weights are chosen so that the variance is minimized.

The optimal weights are functions of the variances of the observations from the temporary and permanent plots and the correlation between the successive measurements on the permanent plots. In practice we either have to estimate these parameters or to use fixed values - determined from earlier experiences - of the weights. This is possible since the variance of the estimators of current volume (as a function of the weights) is rather flat in a neighbourhood of the optimal values. Thus small (or moderate) deviations from the optimal values only increase the variance by very little.

### 3.2. Estimation of total length of forest roads.

The problem is to estimate the total length of forest roads of different categories. The method described below, which

is a variant of Buffon's needle problem, has been advised by Bertil Matérn and used in the Swedish NFS.

Consider an area with a network of roads with total length  $L$ . If a system of parallel equidistant lines (at distance  $b$  apart) is placed according to a randomly chosen direction, the expected number of intersections with the road net is  $2L/(\pi b)$ . Let now  $X$  be the number of intersections between the line system and the network of roads and define

$$Y = \frac{\pi}{2} bX$$

The random variable  $Y$  has the expectation  $L$ , i.e.  $Y$  is an unbiased estimator of the total length of roads.

The unbiasedness of  $Y$  holds also in the case of two line systems at right angles or if we have fragments of the two systems. The tract system used in the Swedish NFS is an example where we have fragments of two line systems. However, we have to redefine  $b$  as  $1/(\text{total length of the line system per unit area})$ .

The unbiasedness of  $Y$  holds if the direction of the line system is determined at random, i.e. the average value of  $Y$  for a very great number of systems with different directions equals  $L$ . However, for an individual system with lines in only one direction  $Y$  may considerably deviate from  $L$  even if the lines of the system are spaced very closely. This situation may occur if the road net is located in such a way that most of the roads run in the same direction. The following example illustrates that the bias is considerably reduced if we have two systems at right angles (or fragments of two systems of this kind).

Example: In the example the length of railways on the isle of Öland has been estimated. The following results were obtained:

Length according to curvemeter (km)	Estimated length in km		
	Line system in		Both systems
	W-E	N-S	
146	192	90	140

If only one line system is used the method may produce rather biased estimates. Using two line systems at right angles will not give any essential bias and is thus the method to recommend.

For further details about this application I refer to von Segebaden (1964). Other applications concern the length of roots, Newman (1966), and the amount of logging residue, Warren and Olsen (1964). The principle can also be used to estimate the length of rivers, forest borders etc.

### 3.3. Estimation of the size of wildlife populations.

A lot of methods have been developed to estimate the size of wildlife populations. The relevance of the different methods depends on certain assumptions which sometimes are rather special and unrealistic. I shall briefly describe a few of the methods and refer the interested readers to Davis (1963) or Seber (1973).

The change in ratio method is based on the change in ratio due to removal of some individuals. The method is best described by an example from Davis (1963). The number of deer tracks per unit area ( $T_1$ ) is counted, a known number



of deers ( $n$ ) is removed and then the tracks ( $T_2$ ) is counted again. The change in number of tracks per unit area ( $T_1 - T_2$ ) is due to  $n$  deers and hence:

$$\frac{T_1 - T_2}{n} = \frac{T_1}{N_1} = \frac{T_2}{N_2},$$

where  $N_1$  is the size of the population before removal, and  $N_2 = (N_1 - n)$  is the size after removal of  $n$  animals.

The formula above can be applied to any change in a ratio, e.g. the ratio of males to females or marked to unmarked.

The latter example is called the method of capture and recapture. In an investigational area  $M$  animals are captured, marked in some manner, and released. On a later occasion  $n$  animals are captured,  $m$  of which are found to be marked. Then the size of the population can be estimated from the following formula:

$$\frac{N}{M} = \frac{n}{m}; \text{ i.e. } \hat{N} = \frac{Mn}{m}.$$

The catch-effort method is based on the belief that changes in the number of animals taken by successive units of efforts (e.g. man-hours of hunting, nets fished per day) reflect changes in the size of the animal population. Let  $c_1, c_2, \dots$  be the catches made at successive occasions and  $e_1, e_2, \dots$  the used efforts. At each occasions we calculate the catch per effort unit (denoted by  $d_i = c_i/e_i$ ) and the accumulated catch  $C_i(a) = c_1 + c_2 + \dots + c_{i-1}$ . Under certain assumptions we have  $d_i = kN_i$ , where  $N_i$  equals the size of the population at occasion  $i$  and  $k$  is the average catch per effort unit.

As  $N_i = N_1 - C_i(a)$  we get  $d_i = kN_1 - kC_i(a)$ . Plotting the

points  $(C_i(a), d_i)$  in a coordinate system and fitting a straight line we can estimate  $k$  and  $N_1$ .

The size of an animal population can also be estimated indirectly. The faecal pellet-group count is a widespread technique. The number of pellet groups within an area is estimated by means of sample plots. If the daily defecation rate and the period of deposition are known, the size of the population can be estimated.

I conclude with a warning. As I have mentioned all these methods are based on special assumptions and before using them I recommend you to consult the book by Seber.

#### References

- Akaike, H (1974). A new look at the statistical identification model. IEEE Trans. Auto Control 19, 716-723.
- Akaike, H (1977). On entropy maximization principle. In Krishnaiah, P.R. (ed.): Applications of statistics, 27-41.
- Cruse et al. (1985). Designing a new national forest survey for Sweden. To appear in Studia Forestalia Suecica.
- Davis, D E (1963). Estimating the numbers of game populations. In Mosby, H S (ed.): Wildlife Investigational Techniques, 2nd ed. 89-118. The Wildlife Society, Washington.
- Geisser, S (1975). The predictive sample reuse method with applications. Jour. Am. Stat. Ass. 70, 320-328.

- Matérn, B (1947). Methods of estimating the accuracy of line- and sample plot surveys. (In Swedish, English summary.) Reports of the Forest Research Institute of Sweden 36(1), 138 pp.
- Matérn, B (1960). Spatial variation - stochastic models and their application to some problems in forest survey and other sampling investigations. Reports of the Forest Research Institute of Sweden 49(5), 144 pp.
- Matérn, B & Ranneby, B (1983). Variational structure in forest. Implications for sampling. In Forest inventory for improved management, University of Helsinki, Dept. of forest mensuration and management, Research notes No. 17, 1-10.
- Newman, E I (1966). A method of estimating the total length of root in a sample. J. Appl. Ecology 3, 139-145.
- Ohlsson, S (1975). Riksskogstaxeringen. Ändamål, utformning och användningsområden. Sw. Univ. of Agric. Sciences, Dept. of Forest Survey, Report No. 16.
- Ranneby, B (1981a). Den topografiska variationen inom olika områden. En redovisning av skattade korrelationsfunktioner. Sw. Univ. of Agric. Sciences, Dept. of Forest Survey. Project Nytax 83, No. 5.
- Ranneby, B (1981b). Provytestorlekens betydelse vid skogsinventeringar. Sw. Univ of Agric. Sciences, Dept. of Forest Survey. Project Nytax 83. No 4.
- Ranneby, B (1984). The maximum spacing method. An estimation method related to the maximum likelihood

- method. Scand. J. Statist. 11, 93-112.
- Schwarz, G (1978). Estimating the dimension of a model. Ann. Statist. 6, 461-464.
- Seber, G A F (1973). The estimation of animal abundance and related parameters. Griffin, London.
- Segebaden, G von (1964). Studies of cross-country transport distances and road net extension. Studia Forestalia Suecica, 18.
- Stone, M (1974). Cross-validators choice and assessment of statistical predictions (with discussion). J. R. Statist. Soc. B, 36, 111-147.
- Stone, M (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. J. R. Statist. Soc. B, 39, 44-47.
- Warren, W G & Olsen, P F (1964). A line intersect technique for assessing logging waste. Forest Science 10, 267-276.

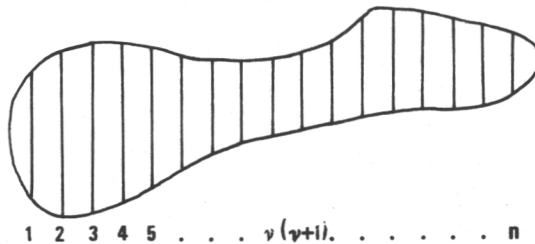
## METSIEN LINJAINVENTOINTIEN TILASTOTIETEELLINEN TAUSTA

Sakari Salminen

Pohjoismaisia valtakunnan metsien inventointeja edeltäneet koeinventoinnit ajoittuvat 1910-luvulle ja koko valtakunnan aluetta kattavien inventointien aloitus 1920-luvun ensipuoliskolle. Tilastollisena menetelmänä inventoinnit olivat alun alkaen systemaattista otantaa. Metsänarvioinnin tyyppisille tehtäville on ominaista, että systemaattisesta otoksesta lasketut arviot ovat yleensä täsmällisempiä kuin vastaavan kokoisesta rajoittamattomasta satunnaisotoksesta lasketut arviot. Harhattomien keskivirhe-estimaattien laskenta suoraan systemaattisesta otoksesta on kuitenkin toistaiseksi ratkaisematon ongelma. Keskivirheiden laskentaan onkin koko inventointien ajan kehitetty teoreettiselta taustaltaan tarkentuvia menetelmiä. Menetelmien lyhyt esittely on johdanto B. Rannebyn nykytilanteen kuvaukselle.

### 1. Koeinventoinnit

Koeinventoinnit olivat täydellisiä linjainventointeja, joissa sekä puusto- että pinta-alatiedot kerättiin tasavälisiltä yhdensuuntaisilta linjoilta. Värmlannin koeinventoinnissa linjat jaettiin ryhmiin antamalla linjojen 1, 11, 21, ... muodostaa ryhmän 1, linjojen 2, 12, 22, ... ryhmän 2 jne (kuva 1) (Lindeberg, 1924). Tarkastellaan jonkin inventointialueella  $Q$  keskimäärin vallitsevan



Kuva 1. Arviointilinjasto alueella  $Q$ .



tilanteen - esimerkiksi metsämaan osuuden tai keskitilavuuden - estimointia. Kun linjaryhmään kuuluvien linjojen  $\mu$  pituudelle käytetään symbolia  $L_\mu$  ja koko linjaston pituudelle symbolia  $L$ , saadaan aluearvio kaavasta

$$(1) \quad \hat{f}(Q) = \sum_{\mu=1}^{10} \frac{L_\mu}{L} f(\mu)$$

Keskivirhe-estimaatti arviolle  $\hat{f}(Q)$  lasketaan kaavasta

$$\epsilon\{\hat{f}(Q)\} = \sqrt{\sum_{\mu=1}^{10} \left(\frac{L_\mu}{L}\right)^2 (f(\mu) - \hat{f}(Q))^2}$$

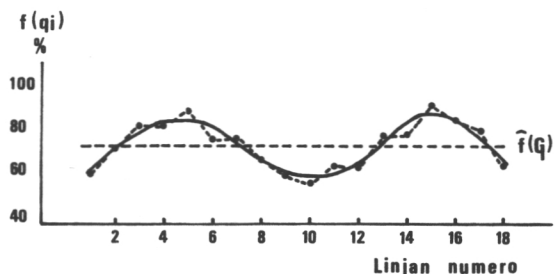
Menetelmä johtaa Lindebergin mukaan harhattomaan lopputulokseen seuraavin oletuksin

- linjaryhmittäiset havainnot ovat toisistaan riippumattomia
- linjaryhmän  $\mu$  kattaman alueen keskivirhe on kääntäen verrannollinen linjapituuden  $L_\mu$  neliöjuureen.

Kritiikin kärki kohdistui jälkimmäiseen oletukseen. Se on liian spesifinen yleisesti sovellutuskelpoiselle menetelmälle.

Suomessa Saha- ja Kuhmalahden arvioinnissa piirrettiin linjanumeron funktiona tasoitusviiva kuvaamaan tarkasteltavan muuttujan odotusarvon vaihtelua alueen eri osissa (kuva 2)

Kuva 2. Subjektiiivisesti piirretty tasoitusviiva linjanumeron funktiona.  $\hat{f}(Q)$  on arvio koko alueelle  $Q$ .



(Ilvessalo, 1923). Tasoitusviivan käyttö lienee ollut perintöä Cajanuksen ajatuksista. Keskivirhearvion neliöpoikkeamat laskettiin yksittäisten linjahavaintojen erona tasoitusviivasta. Lindebergin kritiikki kohdistui tasoitusviivan subjektiiviseen piirtämiseen. Hänen mukaansa tasoitusviiva saattoi yhtä hyvin olla piirtäjänsä funktio kuin todellisen vaihtelun kuvaaja; piirroksen onnistuminen vaati tekijältään vankkaa ammattitaitoa ja alueen metsävarojen jakautumisen tarkkaa tuntemusta.

## 2. Viereisten linjojen vertailu

Tasoitusviivan perusajatus oli oikea; pyrittiin käyttämään hyväksi systemaattisen otoksen tarjoama informaatio muuttujan arvon vaihtelusta inventointialueen sisällä. Ratkaisun subjektiivisuus vain teki siitä epätydyttävän. Käyttökelpoinen yleissovellutus edellyttää, että keskivirhe-estimaatti on laskettavissa kaavasta, jossa muuttujan arvot ovat mukana yksinomaan otoshavaintoina. Lindeberg (1924) päätyi menetelmässään ehdottamaan viereisten linjahavaintojen eron neliöintiä. Kaavan (1) symbolein, kun alaindeksi  $\mu$  yksilöi linjar ryhmän sijaan yksityisen linjan, jonka linjapituus on  $l_{\mu}$ , on Lindebergin keskivirhekaava yleisimmässä muodossaan

$$(2) \quad \epsilon\{f(Q)\} = \sqrt{\frac{n}{2(n-1)} \sum_{\mu=1}^{n-1} (l_{\mu} - l_{\mu+1})^2 (f(\mu) - f(\mu+1))^2}$$

Linjojen kokonaislukumäärä on  $n$  (kuva 1).

Läheisten otososien havaintojen erotuksilla on Lindebergin ratkaisusta lähtien ollut keskeinen asema inventointien keskivirheitä estimoitaessa. Suoraan Lindebergin ajatuksia käyttäen laskettiin keskivirheet valtakunnan metsien 1. - 3. in-

ventoinnissa v. 1921-24, 1936-38 ja 1951-53 (Ilvessalo, 1927, 1942 ja 1956). Vuosien 1936-38 inventoinnissa siirryttiin puhtaasta linjainventoinnista yhdistettyyn linja- ja koeala-arviointiin. Mitattuja puustotietoja ei enää kerätty koko linjalta, vaan ainoastaan linjalle määräväleihin sijoitetuilta koealoilta.

Kaavan (2) sovelluskelpoisuus heikkenee, jos linjoja on vähän tai viereiset linjat ovat pituuksiltaan aivan eri suuruusluokkaa. Molempien häirtatekijöiden vaikutusta voidaan vähentää, jos inventointialue jaetaan sopivasti osa-alueisiin ja kaavaa käytetään osa-alueiden keskivirheiden estimointiin. Lindeberg oli tietoinen kaavansa heikkouksista mainituissa olosuhteissa ja ehdotti osa-alueiden käyttöä.

### 3. Linjainventointi stokastisena prosessina

Matérn (1947) tarkasteli linjainventointia stokastisena prosessina (Salminen, 1983). Tämä työ on ollut pohjana keskivirhe-estimoinnin kehitystyölle viimeiset 40 vuotta. Käytännön sovellutuksiin johtaneissa ratkaisuissa on inventointiprosessi oletettu stationääriseksi ja isotrooppiseksi. Eritellyinä ehtoina tämä merkitsee, että

- pisteessä  $u$  on satunnaismuuttujalla  $f(u)$  vakiot  $\mu$  ja  $\sigma^2$  odotusarvona ja varianssina
- satunnaismuuttujien  $f(u)$  ja  $f(v)$  välinen korrelaatio  $\varphi(t)$  riippuu vain pisteiden  $u$  ja  $v$  välisestä etäisyydestä  $t$ , mutta ei niitä yhdistävän vektorin suunnasta.

Sekä prosessin teoreettisessa tarkastelussa että keskivirheiden estimoinnissa käytetään apuna neliömuotoja. Yleis- muodossaan neliömuoto  $T$  voidaan esittää kaavana

$$T = \sum_i \sum_j c_{ij} f(q_i) f(q_j), \text{ jossa}$$

kertoimet täyttävät ehdot

$$c_{ij} = c_{ji}, \sum_i \sum_j c_{ij} = 0 \text{ ja } \sum_i c_{ii} = 1$$

ja  $f(q_i)$  on satunnaismuuttujan arvo alueella  $q_i$ . "Alue" voi olla yhtenäinen tai erillisten koealojen muodostama todellinen alue, linjasto, linja, linjanosa tai pisteistö.

Keskivirheiden estimoinnissa Matérn päätyy ratkaisuun, jossa alueet  $q_i$  ovat tasapituisia lähekkäisiä linjanosia. Eri neliömuotojen käyttökelpoisuuden vertailumalliksi johdetaan neliömuoto  $T_0$ .

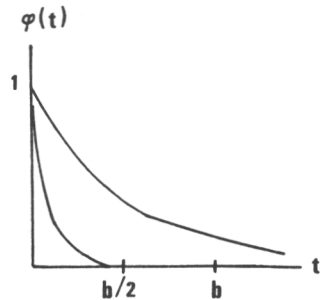
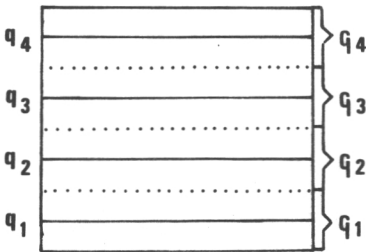
$$T_0 = l_i * \{f(q_i) - f(Q_i)\}^2,$$

jossa  $l_i$  merkitsee linjan  $q_i$  pituutta ja  $Q_i$  linjan  $q_i$  halkomaa suorakaidetta (kuva 3). Odotusarvo

$$(3) \quad E(T_0) = l_i * E\{f(q_i) - f(Q_i)\}^2$$

kirjoittuu muotoon

$$(3a) \quad E(T_0) = \int_0^\infty \phi(t) a_{T_0}(t) dt.$$



Kuva 3. Alueella  $Q = \sum_i Q_i$  tasavälein sijaitsevat arviointilinjat  $q_i$ .

Kuva 4. Pari esimerkkiä jatkuvista, aidosti laskevista korrelaatiofunktioista.

Kaavan (3a) integroitavan tulossa  $\varphi(t)$  on isotrooppisen prosessin korrelaatiofunktio ja  $a_{T_0}(t)$  neliömuodon  $T_0$  etäisyysfunktio. Korrelaatiofunktio on jatkuva, laskeva ja konvekssi sekä täyttää ehdon  $\Phi(0) = 1$  (kuva 4). Etäisyysfunktio

$$a_{T_0}(t) = l_i \{a_{q_i q_i}(t) - 2a_{Q_i Q_i}(t) + a_{Q_i Q_i}(t)\}$$

on tiheysfunktioiden lineaarinen yhdistely, missä  $a_{q_i q_i}(t)$  on linjalta  $q_i$  satunnaisesti valitun kahden pisteen etäisyyden  $t$  tiheysfunktio, tiheysfunktion  $a_{Q_i Q_i}(t)$  tapauksessa pisteet on valittu suorakaiteelta  $Q_i$  ja funktion  $a_{q_i q_i}(t)$  tapauksessa toinen piste linjalta  $q_i$  ja toinen suorakaiteelta  $Q_i$ . Yleisesti neliömuodon  $T$  etäisyysfunktio on muotoa

$$a_T(t) = \sum_{ij} c_{ij} a_{ij}(t).$$

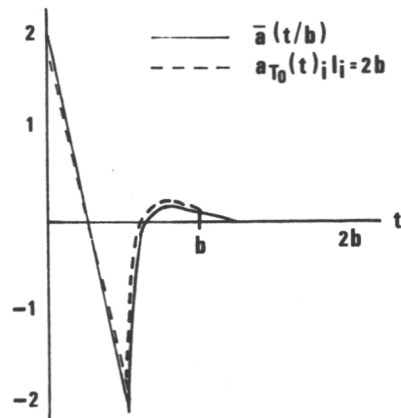
Linjapituudesta riippumattomaan etäisyysfunktioon päädytään antamalla linjojen väliselle etäisyydelle merkki  $b$  ja siirtymällä raja-arvoon

$$\lim_{l_i \rightarrow \infty} a_{T_0}(t) = \bar{a}(t/b).$$

Kuvan 5 mukaan

$$\bar{a}(t/b) \approx a_{T_0}(t)$$

Kuva 5. Etäisyysfunktiot  $\bar{a}(t/b)$  ja  $a_{T_0}(t)$ .





jo niin lyhyellä linjan pituudella kuin  $l_i = 2b$ . Koska  $\Phi(t)$  on laskeva funktio, muodostaa raja-arvo

$$\lim_{l_i \rightarrow \infty} E(T_0) = \epsilon_b^2$$

käyttökelpoisen konservatiivisen estimaatin odotusarvolle  $E(T_0)$ .

Olkoon  $Q$  mielivaltainen inventointialue,  $q$  alueen kattava systemaattinen linjasto sekä  $L$  linjojen yhteispituus. Todistus sivuuttaen todetaan, että

$$E\{f(q) - f(Q)\}^2 \approx \epsilon_b^2/L.$$

Käytettäessä linjanosia keskivirheiden estimointiin on yleisimmin sovellettu neliömuotoa

$$T_1 = c * \{f(q_{ij}) - f(q_{i(j+1)})\}^2,$$

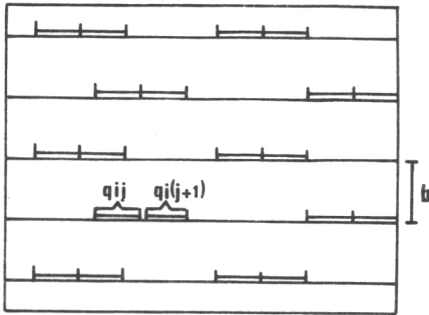
jossa linjanosaan pituus on  $c$  ja verrattavat linjanosat sijaitsevat välittömästi peräkkäin samalla linjalla  $i$  (kuva 6).  $T_1$  on kasvava  $c:n$  funktio. Yhtälöstä

$$(4) \quad \left(\frac{d a_{T_1}(t)}{dt}\right)_{t=0} = \left(\frac{d \bar{a}(t/b)}{dt}\right)_{t=0}$$

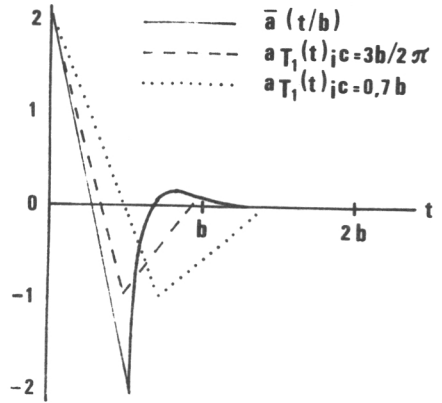
on määrättävissä minimipituus  $c_{T_1}$ , jota lyhyemmän linjanosan käyttö aiheuttaa todennäköisesti keskivirheen systemaattisen aliarvion. Yhtälöllä (4) on ratkaisuna

$$c_{T_1} = 3b/2\pi.$$

Kuvassa 7 on vertailtu  $c:n$  arvoilla  $3b/2\pi$  ja  $0,7b$  etäisyysfunktioita  $a_{T_1}(t)$  raja-arvofunktioon  $\bar{a}(t/b)$ . Teorian mukai-



Kuva 6. Esimerkki systemaattisesta linjanosien valinnasta tasavälisessä linjastossa. Linjanosan  $q_{ij}$  pituus =  $c$  ja linjojen  $i_j$  välinen etäisyys =  $b$ .



Kuva 7. Etäisyysfunktio  $\bar{a}(t/b)$  ja kaksi esimerkkiä funktiosta  $a_{T_1}(t)$ .

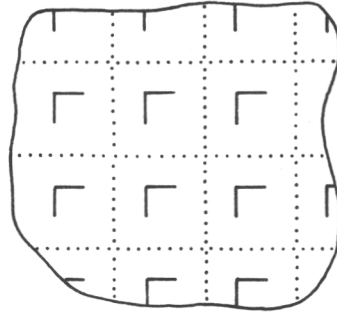
sesti mainituilla linjanosapituuksilla neliömuodosta  $T_1$  saadaan keskimäärin liian suuria keskivirhearvioita. Ne ovat kuitenkin tyydyttäviä täsmällisyyden mittoja tapauksissa, joissa korrelaatiofunktio laskee suhteellisen nopeasti.

Suomen valtakunnan metsien inventoinneissa linjanosien vertailua ei keskivirheiden estimoinnissa koskaan käytetty. Ajallisesti vuosien 1951-53 inventointiin menetelmä olisi ollut käytettävissä, mutta meillä pitäydyttiin vanhasta tottumuksesta yhä Lindebergin kaavoihin (Ilvessalo, 1956).

#### 4. Inventoinnit tänään

Varsinaista linjainventointia ei enää ole. Yhtenäiset linjat on katkottu neliön tai puolinelion piirin muodostamiksi lohkoiksi (kuva 8). Mittauksia tai havaintoja ei enää tehdä koko linjalta, vaan ainoastaan koeloilta. Osittain lohkot

Kuva 8. Kaavio nykyisin käytössä olevasta arviointilohkojen sijoittelusta.



ovat pysyviä, lohkolinjalla sijaitsevia koealoja mitataan uudelleen ja uudelleen. Keskvirheiden estimoinnissa vertailtavat alueet ovat viereisiä lohkoja.

Runsas pohdintojen aihe on lohkon optimimuoto ja -koko sekä koealojen keskinäinen etäisyys lohkolla (Ranneby 1979). Korrelaatio- ja etäisyysfunktiot ovat käyttökelpoinen apu näiden ongelmien käsittelyyn.

#### Lähdeluettelo

- Ilvessalo, Y. (1923): Tutkimuksia yksityismetsien tilasta Hämeen läänin keskiosissa. Sahalahden ja Kuhmalahden pitäjäien metsät. Acta For.Fenn. 26(2), 1-137.
- (1927): Suomen metsät. Tulokset vuosina 1921-1924 suoritetusta valtakunnan metsien arvioimisesta. Commun.Inst. For.Fenn. 11, 1-421 sekä taulukkosivut 1-192.
  - (1942): Suomen metsävarat ja metsien tila. II valtakunnan metsien inventointi. Commun.Inst.For.Fenn. 30, 1-446.
  - (1956): Suomen metsät vuosista 1921-1924 vuosiin 1951-1953. Kolmeen valtakunnan metsien inventointiin perustuva tutkimus. Commun.Inst.For.Fenn. 47(1), 1-227.
- Lindeberg, J.W. (1924): Über die Berechnung des Mittelfehlers des Resultates einer Linientaxierung. Acta For.Fenn. 25 (5), 1-22.

- Matérn, B. (1947): Metoder att uppskatta noggrannheten vid linje- och provytetaxering. Medd. från Statens Skogsforskn. Inst. 36(1), 1-138.
- Ranneby, B. (1979): Model studies of tract sizes in forest survey. Forest Resource Inventories. Workshop Proceeding. Colorado State University; Fort Collins. July 23-26, 1979. Vol. I, 289-297.
- Salminen, S. (1983): Nykyisen valtakunnan metsien inventoinnin tilastomatemaattinen perusta. Suomen Akatemian seminaari Hyytiälässä 21-22.11.1983. Uudet menetelmät metsäinventoinnissa. Seminaariesitelmät. 8 s.

## SPATIAALISISTA KILPAILUMALLEISTA

Antti Penttinen

Kilpailulla on keskeinen asema ekologisissa ilmiöissä. Intuitiivisesti näyttää ilmeiseltä, mitä kilpailulla tällöin tarkoitetaan. Kuitenkin tämän käsitteen eksakti määrittelyminen ei ole ongelmattonta. Ensinnäkin, tarkoitammeko kilpailulla esimerkiksi populaatioiden välistä kilpailua vai yksilöiden välistä kilpailua? Toisaalta, asetammeko määritelmän verbaalisti vai matemaattisen mallin avulla? Kolmanneksi, missä avaruudessa kuvauksen teemme (ks. Cormack, 1979)? Rajoitun tässä yhteydessä käsittelemään yksilöiden välistä kilpailua kaksiulotteisessa tila-avaruudessa ja valitsen lähtökohdakseni matemaattisen kuvauksen siitä, miten kilpailu vaikuttaa tilajärjestykseen.

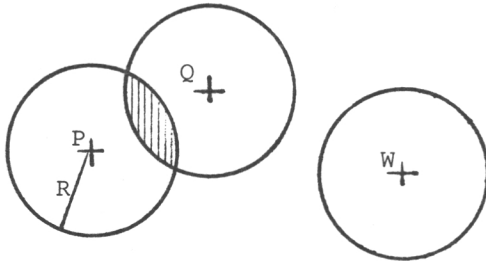
### 1. Kilpailu ja tilajärjestys

Oletamme, että havaintonamme on puiden tai vastaavien objektien tilajärjestys, so. objektien suhteelliset sijainnit toisiinsa nähden ja oletamme, että puut vuorovaikuttavat (kilpailevat) keskenään. Varhaisimpia tilajärjestyksen malleja, jotka ottavat kilpailun huomioon, ovat Matérnin kovan ytimen mallit (Matérn, 1960). Pyrimme kuitenkin yleisempään kilpailun kuvaamiseen.

Kilpailussa on kyse vuorovaikuttamisesta. Fysiikassa vuorovaikutusta on kuvattu matemaattisten mallien, ns. Gibbs'in tilojen avulla. Nämä alunperin termodynamiikassa kehitetyt mallit ovat tulleet tilastotieteeseen 70-luvun loppupuolella ja tarjoavat monipuolisen työkalun myös biologisessa ym-

ympäristössä. Tarkastelemme tätä malliperhettä esimerkin avulla.

Oletamme, että puu käyttää ravinteita R-säteisen ympyrän alalta. Mikäli muita tekijöitä ei oteta huomioon, niin kaksi puuta kilpailevat keskenään ravinteista, jos niiden välinen etäisyys on enentään  $2R$ . Päädymme yksinkertaiseen kvalitatiiviseen kuvaukseen (Kuvio 1).



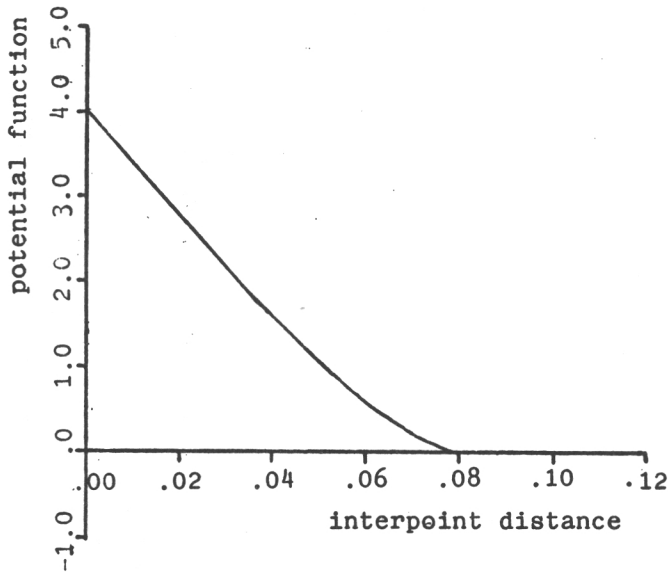
Kuvio 1. Esimerkki spatiaalisesta kilpailusta, jossa puu hyödyntää maaperää R-säteisen ympyrän alalta: P ja Q kilpailevat, kun taas P ja W sekä Q ja W eivät kilpaile.

Pyrkimyksenämme on kuitenkin kilpailun kvantitatiivinen kuvaus. Merkitään  $\psi(P,Q)$ :lla puiden P ja Q välisestä vuorovaikutuksesta johtuvaa potentiaalia. Tässä tapauksessa on luonnollista olettaa, että  $\psi(P,Q) = \psi(|P-Q|)$ . Yksinkertainen tapa konstruoida potentiaalifunktio esimerkkinme tapauksessa on olettaa, että  $\psi(|P-Q|)$  on verrannollinen P:n ja Q:n yhteisesti hyödyntämän alueen pinta-alaan (ks. Kuvio 1), jolloin potentiaalifunktio  $\psi$  on



$$(1) \quad \psi(d) = \begin{cases} -\theta \left\{ 2R^2 \arccos\left(\frac{d}{2R}\right) - d \sqrt{R^2 - \left(\frac{d}{2}\right)^2} \right\}, & 0 \leq d \leq 2R \\ 0, & d > 2R, \end{cases}$$

missä  $d = |P-Q|$  ja  $\theta$  on (skaalaus)parametri. Kuviossa 2 on tämän potentiaalifunktion kuvaaja.



Kuvio 2. Potentiaalifunktion (1) kuvaaja, kun  $R = 0.04$  ja  $\theta = -800$ .

Siirrymme nyt kuvaamaan koko metsän tilajärjestystä, jossa metsässä vallitsee esimerkkinne mukainen kilpailu. Oletetaan, että puiden lukumäärä metsäalueessa  $E$  on  $N$  ja alueen pinta-ala on  $|E|$ . Merkitään puiden sijaintien koordinaatteja vektorilla  $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_N)$ . Tällöin systeemin kilpailusta johtuva kokonaisenergia on

$$(2) \quad V(\tilde{X}; \psi) = \sum_{i < j} \psi(|\tilde{X}_i - \tilde{X}_j|)$$

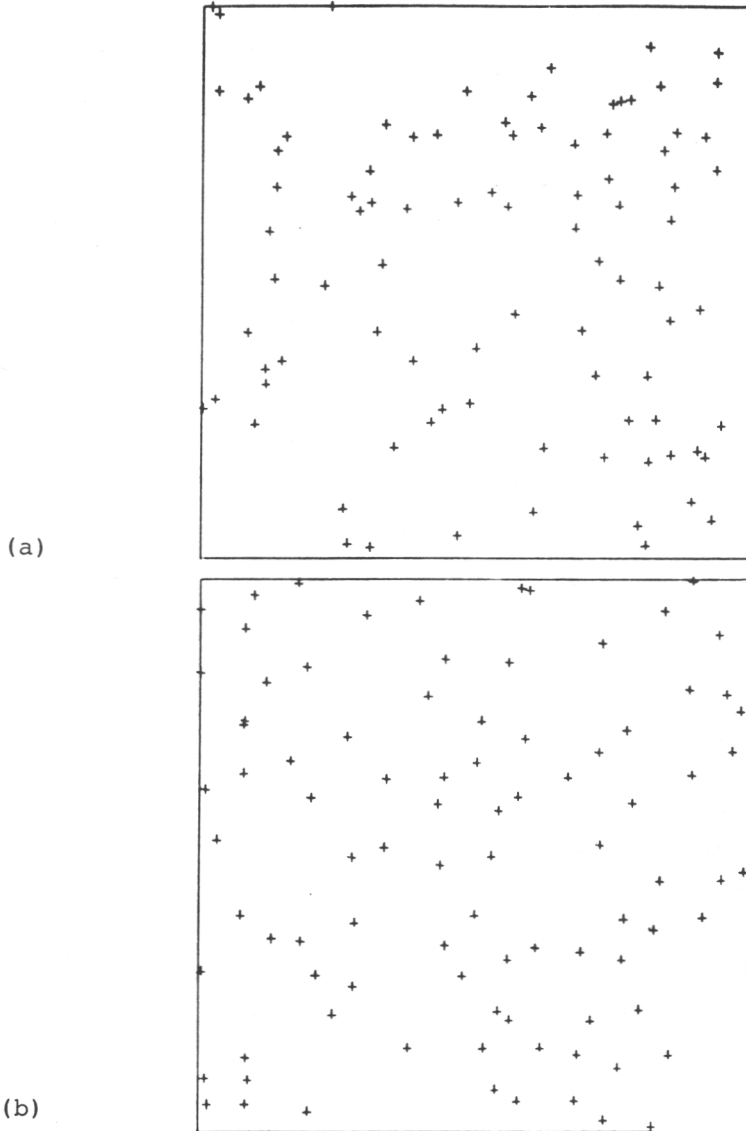
ja systeemin tilaa kuvataan tiheysfunktiolla

$$(3) \quad p(\tilde{X}; \psi) = c(\psi, E, N)^{-1} \exp\left\{-\sum_{i < j} \psi(|X_i - X_j|)\right\} |E|^{-N}.$$

Tätä jakaumaa kutsutaan Gibbs'in tilaksi ja sen määrittelemää pisteprosessia Gibbs'in prosessiksi.

Gibbs'in prosessilla on seuraava tulkinnan kannalta tärkeä ominaisuus: Tilajärjestyksen ehdollinen todennäköisyyslaki koealassa  $A \subset E$  ehdolla, että puiden tilajärjestys  $A$ :n ulkopuolella tunnetaan, riippuu tästä ehdollistamisesta ainoastaan  $A$ :n ympäristössä olevien puiden tilajärjestyksestä. Esimerkkinne tapauksessa ko. ympäristö on  $A$ :n  $2R$ -säteinen ympäristö. Tätä ominaisuutta sanotaan Markovin ominaisuudeksi. Varsin heikon lisäoletuksen vallitessa myös käänteinen on voimassa: Ehdollinen riippumattomuus johtaa tyyppiä (3) olevaan multiplikatiiviseen todennäköisyystiheuteen. Tämän vuoksi (3):a sanotaan myös Markovin malliksi ja vastaavaa prosessia Markovin pisteprosessiksi. Toinen tärkeä ominaisuus on se, että Gibbs'in prosessi voidaan esittää erään ajasta riippuvan prosessin tasapainojakaumana, mikä mahdollistaa realisaatioiden simuloinnin ja siten mallin tilastollisen analyysin. Kuviossa 3b olevan kilpailumallin realisaatio on generoitu juuri tällä periaatteella.

Mistä esimerkissämme on oikeastaan kyse? Itse ilmiö on lokaali kilpailu mikrotasolla, jonka oletetaan vaikuttavan makrotasolla havaittavaan tilajärjestykseen. Kilpailumallimme on konstruoitu lähtien parittaisista vuorovaikutuksista mikrotasolla ja lopputuloksena on makrotasoa vastaava tilastollinen malli. Tämä ajatus on selvimmillään nähtävissä termodynamiikassa, jossa kaasun havaittavia ominaisuuksia kuten tilayhtälöä selitetään molekyylien välisillä vuorovaikutuksilla (törmäyksillä). Mallin tehtävänä on välittää itse ilmiö havaintotasolle.



Kuvio 3. Tyypillinen realisaatio kahdesta tilajärjestyskuvioista, jossa kummassakin on 30 "puuta". (a)  $\psi \equiv 0$  (ei kilpailua ja (b)  $\psi$  on kaavan (1) mukainen ( $\theta = -800$ ,  $R = 0.04$ ), (Penttinen, 1984; s. 22). Kuviot ovat yksikköneliössä.

## 2. Markovin mallin tilastollisen analyysin mahdollisuuksia

Siirrymme hetkeksi ajatusleikkiin. Kuvittelemme olevamme biologeja tai metsämiehiä, joiden tehtävänä on analysoida taimien välistä kilpailua. Meille on kerrottu Markovin malleista. Pidämme tätä malliperhettä realistisena kuvauksena ilmiölle, ainakin, jos populaatio on tasapainossa. Edelleen pidämme näitä malleja helppotajuisina ja helposti tulkittavina. Miksi emme käyttäisi niitä?

Vaihdamme nyt roolia ja kuvittelemme olevamme tilastotieteilijöitä, joiden tehtävänä on suorittaa mallin sovittamiseen tarvittavat laskut. Voin hyvin kuvailla tunteitamme: kerrassaan ikävä malliperhe. Syynä on se, että rutiininomainen tilastollinen analyysi ei tule kysymykseen. Esimerkiksi tilastotieteilijän jokapäiväistä työkalua, uskottavuusfunktioita, ei voida esittää analyttisesti. Tuoreessa artikkelissaan Diggle ja Gratton (1984) kutsuvat tämäntapaista mallia implisiittiseksi tilastolliseksi malliksi vastakohtana eksplisiittiselle tilastolliselle mallille, joihin olemme tottuneet. Implisiittinen tilastollinen malli on sellainen, jota on vaikea käsitellä analyttisesti, mutta jonka konstruktio sallii esimerkiksi tietokonesimulointien hyödyntämisen. Markovin tilajärjestysmalli on tyypillinen esimerkki. Nykyinen teoreettinen tietämyksemme mahdollistaa kuitenkin tämän malliperheen soveltamisen tietokonetta hyväksi käyttäen.

Tilajärjestyksen analyysi perustetaan kenttähavainnoilla (ts. harvalla otannalla) saatuun aineistoon, jossa mitataan etäisyyksiä tai vaihtoehtoisesti lukumääriä (tai molempia). Toinen, informatiivisempi aineiston muoto on täydellinen kartta puiden sijainneista. Tarkastelen seuraavassa karttamuotoisen aineiston estimoinnin ja testauksen mahdollisuuksia.

## 2.1. Potentiaalifunktion estimointi

Oletetaan, että potentiaalifunktiolle on valittu parametri-  
nen esitys, ts.  $\psi(\cdot) = \psi(\cdot; \theta)$ , missä  $\theta$  on parametriavaruuden  
piste. Potentiaalifunktion estimointi tarkoittaa parametrin  
arvon määrittämistä aineiston perusteella. Useita vaihtoehtoisia  
estimointimenetelmiä on esitetty käytettäväksi  
Markovin malleille. Näistä mainittakoon approksimatiivinen  
suurimman uskottavuuden menetelmä (Ogata ja Tanemura, 1981;  
1983a; 1983b; Penttinen, 1984), pseudo-uskottavuusmenetelmä  
(Besag, 1978) ja pns-tyyppinen Monte Carlo -menetelmä (Diggle,  
1979; Diggle ja Gratton, 1984). Tätä estimointimenetelmien  
joukkoa on äskettäin täydennetty kaksivaiheisella tekniikalla,  
jossa hyödynnetään ns. Palmin todennäköisyyksiä  
(Takacs, 1983a; 1983b; Fiksel, 1984). Jos mallin käyttö on  
eksploratiivista, parametrin arvo on mahdollista "arvata"  
joissain yksinkertaisissa tapauksissa (Ripley, 1981). Ver-  
tailevia tutkimuksia em. menetelmien suoriutumuksesta ei ole  
käytössämme.

## 2.2. Yhteensopivuuden testaus

Markovin tilajärjestysmallin teoreettiset vaikeudet näkyvät  
myös yhteensopivuuden testauksessa. Konventionaaliset mene-  
telmät eivät yleisesti tule kysymykseen. Poikkeuksena on  
malli  $\psi \equiv 0$  (spatiaalinen Poissonprosessi). Tämä malli vas-  
taa tilannetta, jossa yksilöiden välillä ei ole vuorovaiku-  
tuksia. Tämän Poisson-hypoteesin testaamiseen on olemassa  
asymptoottisia testejä (vrt. Ripley, 1981: s. 160). Pää-  
sääntöisesti Markovin malleille yhteensopivuuden testaus  
joudutaan tekemään Monte Carlo -testinä (ks. esim. Diggle,  
1983: s. 76). Pääpiirteissään Monte Carlo -testi on seuraava:  
Oletetaan, että estimoinnin tuloksena on saatu potenti-

aalifunktio  $\hat{\psi}$ . Simuloidaan tätä potentiaalifunktiota vastaavaa tilajärjestysmallia ja simulointituloksia verrataan havaittuun tilajärjestyskuviioon käyttäen tiivistelmiä, joita ovat mm. Ripley'n K-funktio, tyhjän tilan tunnusluku ja lähimmän naapurin etäisyyden jakauma. Mikäli aineisto on riittävän suuri, on suositeltavaa puolittaa se ja käyttää toista osaa estimointiin ja toista testaukseen. Tällöin vältytään testin konservatiivisuudelta. Testiä tehtäessä Markovin mallin simulointi tapahtuu iteratiivisesti kuviota lokaalisti muuttaen (Ripley, 1979). Simulointialgoritmin suoritusnopeus riippuu mallista. Se saattaa vielä tällä hetkellä olla liian suuritöinen mikrotietokoneella suoritettavaksi.

### 3. Yleistyksiä

Edellä on kilpailua kuvattaessa otettu huomioon vain parittaiset vuorovaikutukset. Lisäksi on oletettu, että sekä puut että metsäalue ovat homogeenisia. Näistä yksinkertaistuksista voidaan jossain määrin luopua. Yleistyksestä joudutaan maksamaan oma hintansa mallin muuttuessa komplisoidummaksi. Matemaattiseen malliin ei liittynyt ongelmia, mutta soveltamiseen kylläkin. Esimerkiksi mallin identifiointi voi osoittautua mahdottomaksi, mikäli vahvaa sisältötietoa ei ole käytössä. Luonnehdin seuraavassa kahta yleistysmahdollisuutta.

Kolmen ja useamman puun yhdysvaikutus on mahdollista ottaa malliin. Esimerkiksi kolmen puun vuorovaikutusten tapauksessa tilajärjestyskuviion kokonaispotentiaali on

$$V(\tilde{X}; \psi_2, \psi_3) = \sum_{i < j} \psi_2(\tilde{x}_i, \tilde{x}_j) + \sum_{i < j < k} \psi_3(\tilde{x}_i, \tilde{x}_j, \tilde{x}_k) .$$



Kolmen puun vuorovaikutusta tarvitaan esimerkiksi testattaessa, muodostavatko puut "linjoja". Kolmannen asteen vuorovaikutusten lisääminen malliin kasvattaa parametrien määrää.

Jos puut ovat epähomogeenisia ja erilaisuus vaikuttaa puiden väliseen kilpailuun, niin Markovin malli on modifioitavissa tähänkin tapaukseen. Jos esimerkiksi metsä muodostuu kahdesta puulajista, vuorovaikutusten kuvaamiseen tarvitaan kolme funktiota:  $\psi^{(1)}$  lajin 1 sisäistä kilpailua varten,  $\psi^{(2)}$  lajin 2 sisäistä kilpailua varten sekä  $\psi^{(12)}$  kuvaamaan lajien välistä vuorovaikutusta. Malli (3) pätee edelleen, mutta  $\psi$  riippuu siitä, mitä lajia kilpaileva puupari  $X_i$  ja  $X_j$  ovat. Tällaista yleistystä sanotaan merkatuksi Markovin tilajärjestysmalliksi.

Edellä on Markovin mallien tilastollista analyysia kuvattu siinä tapauksessa, että aineisto on karttamuotoinen. Usein metsätieteissä aineisto on mahdollista saada vain otannalla. Yksi työstettävänä olevista ongelmista on juuri kenttähavaintoihin perustuvien analyysimenetelmien kehittäminen.

#### 4. Mallin käytöstä

Tilajärjestystä analysoiva tutkija voi asettaa tavoitteeseen eksploratiivisen analyysin, jossa mallin merkitystä ei korosteta. Potentiaalifunktiota voidaan käyttää helposti tulkittavana tunnuslukuna täydentämään muita deskriptiivisiä analyysimenetelmiä. Jos sen sijaan tavoite on konfirmatorinen, ts. ennalta asetettua vuorovaikutushypoteesia halutaan testata, niin Markovin malli on luonnollisia hypoteesikandidaatteja konstruktioperiaatteensa perusteella. Lisäksi Markovin malli yleistyksineen voi tulla kysymykseen tilajärjestysmallien simulointitutkimuksissa. Esimerkkinä

mainittakoon ns. relaskooppiotannan (Bitterlich -otannan) suorituskyvyn tutkiminen merkatun Markovin tilajärjestysmallin avulla, missä "merkinä" on puun paksuus - oleellinen tekijä relaskooppiotannassa.

#### Lähdeluettelo

Besag, J. (1978): Some methods of statistical analysis for spatial data. Bull. Int. Statist. Inst. 47(2), 77-92.

Cormack, R.M. (1979): Spatial aspects of competition between individuals. Kokoomateoksessa "Spatial and temporal analysis in ecology", toim. Cormack, R.M., Ord, J.K., 151-212. International Co-operative Publishing House, Fairland, Maryland.

Diggle, P.J. (1979): On parameter estimation and goodness-of-fit testing for spatial point patterns. Biometrics 35, 87-101.

Diggle, P.J. (1983): Statistical analysis of spatial point patterns. Academic Press, London.

Diggle, P.J., Gratton, R.J. (1984): Monte Carlo methods of inference for implicit statistical models. J.R. Statist. Soc. B 46, 193-227 (with discussion).

Fiksel, T. (1984): Estimation of parametrised pair potentials of marked and nonmarked Gibbsian point process. Elektron. Informationsverarb. Kybernet. 20 .

Matérn, B. (1960): Spatial variation. Meddelanden från Statens Skogsforskningsinstitut 49,5, 1-144. Stockholm.

Ogata, Y, Tanemura, M. (1981): Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. Ann. Inst. Statist. Math. B 33, 315-338.

Ogata, Y, Tanemura, M. (1983a): Likelihood analysis of spatial point patterns. J.R. Statist. Soc. B.

Ogata, Y., Tanemura, M. (1983b): Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method. Research Memorandum No. 260, The Institute of Statistical Mathematics, Tokyo.

Penttinen, A. (1984): Modelling interactions in spatial point patterns: parameter estimation by the maximum likelihood method. Jyväskylä studies in computer science, economics and statistics, n:o 7, Jyväskylä (Diss.).

Ripley, B.D. (1979): Simulating spatial patterns. Dependent samples from a multivariate density. J.R. Statist. Soc. C (Applied Statistics) 28, 109-112.

Ripley, B.D. (1981): Spatial statistics. Wiley, New York-Chichester-Brisbane-Toronto.

Takacs, R. (1983a): Estimator for the pair-potential of a Gibbsian point process. Institutsbericht Nr. 238, Institut für Matematik, Johannes Kepler Universität, Linz.

Takacs, R. (1983b): Schätzen des Paarpotentials eines Gibbsprozesses (Fallstudie). Institutsbericht Nr. 248, Institut für Matematik, Johannes Kepler Universität, Linz.

LÄHIMMÄN NAAPURIN KORRELAATION HUOMIOONOTTAMINEN  
KOESUUNNITTELUSSA

Matti Lahtinen

Käsittelen ns. Papadakisin menetelmää, joka on maanviljelykokeissa käytetty keino maan korrelaation huomioonottamiseksi tulosten analysoimisessa.

1. Papadakisin menetelmä lohkokokeen täydentäjänä

Tarkastellaan seuraavaa ongelmaa: halutaan kokeellisesti selvittää lannoitteiden tehoja. Asiaankuulumattomien ympäristötekijöiden vaikutuksen eliminoimiseksi koealueeksi valitaan mahdollisimman tasalaatuinen ja siksi yhtenäinen alue. Se jaetaan sitten ruutuihin ja ruudut käsitellään eri lannoitteilla. Vertaamalla eri tavalla käsiteltyjen ruutujen satojen keskiarvoja päästään tulokseen.

Tasalaatuisen maan valinta ei ole kuitenkaan aina mahdollista esim. pinnanmuodostuksen vaihtelun vuoksi. Siksi koealue jaetaan yleensä vielä lohkoihin, joiden sisällä em. tasaisuus toteutuu. Lisäksi lannoitteet jaetaan satunnaisesti kunkin lohkon sisällä. Kuviossa 1 on esimerkki. Yksinkertaisin tähän perustuva malli on

$$Y_{ij} = \mu + \mu_i + \beta_j + \varepsilon_{ij}.$$

Ts. sadon katsotaan koostuvan koko alueen yleistasosta  $\mu$ , lannoitteen vaikutuksesta  $\mu_i$  ja lohkon vaikutuksesta  $\beta_j$ ,

sekä satunnaisesta vaihtelusta  $\varepsilon_{ij}$ , jotka ovat kaikki riippumattomia ja voidaan laskea yhteen (siis ei ole yhdysvaihteluksia). Tämä on ns. lohkokoe.

Mutta tulokset ovat luotettavia vain kun malli on oikea.

Milloin näin on? Se nähdään jäännöksistä  $\varepsilon_{ij}$ . Niiden pitää olla satunnaisia ja riippumattomia. Lohkoihin jako on saatanut olla virheellinen, mutta paraskaan valinta ei voi poistaa kaikkea systemaattista. Muuten: lohkokokeessa käsittelyjen satunnaistaminen lohkojen sisään sekoittaa jäännöksiä niin, että on vaikea nähdä pelkällä silmällä ovatko ne riippumattomia.

Onhan ymmärrettävää, että hyvän maan vieressä on todennäköisesti hyvää ja huonon huonoa. Tätä malli ei ota tarkasti huomioon, sillä siinä oletetaan, että lohkot ja niiden tasot  $\beta_j$  kertovat kaiken maan laadun vaihtelusta. Jäännökset voisivat olla kuvion 2 kaltaisia (kuvion 1 aineistosta).

Voidaanko mallia parantaa? Yksinkertainen ja intuitiivisesti luonnollinen tapa selittää näitä jäännöksiä, joita sanon ruutujen tuottavuuksiksi, on tehdä se naapuriruutujen tuottavuuksien avulla. Ruudun naapurien tuottavuuksien keskiarvon voi ajatella antavan jollakin tarkkuudella itse ruudun tuottavuuden ja ylimenevän osan olevan satunnainen jäännös. Siis nämä korjatut jäännökset esimerkin tapauksessa olisivat kuvion 3 mukaisia.

Naapureiksi on luonnollisinta valita lohkonaapurit, jos lohkoissa on tasoeroja, sillä tuottavuudet eivät varmaankaan korreloi tasoerojen yli. Esimerkissä naapureita olisi siis kaksi, mutta neljäkään ei ole poissuljettu (myös rivinaapurit mukaan). Korjatut jäännökset lasketaan vähentämällä tuottavuudesta samassa lohkoissa kummallakin puolella olevien tuottavuuksien keskiarvo, paitsi lohkojen päissä, joissa vähennetään ainoa lohkonaapuri. Kuten kuviosta 3 näky,

jäännökset ovat nyt ainakin itseisarvoltaan pienempiä.  
Kyseessä on lohkomallin laajennus

$$\begin{aligned} Y_{ij} &= \mu + \mu_i + \beta_j + \epsilon_{ij} \\ \epsilon_{ij} &= \frac{1}{2}(\epsilon_{i-1,j} + \epsilon_{i+1,j}) + \delta_{ij} \quad (i=2,3,\dots,12) \\ \epsilon_{1j} &= \epsilon_{2j} + \delta_{1j} \\ \epsilon_{13,j} &= \epsilon_{12,j} + \delta_{13,j} \\ \delta_{ij} & \text{ satunnaisia ja riippumattomia} \end{aligned}$$

Mutta miten parannus kuvastuu käsittelykeskiarvoissa, sillä niidenhän tarkentamiseen pyritään?  
Korjataan ensin satoja kuten äsken jäännöksiä, vähentämällä siis naapurien tuottavuuksien keskiarvo. Tämähän jättää satoon virheeksi vain äsken mainitun korjatun jäännöksen  $\delta_{ij}$  ja lohko vaikutuksen, joka eliminoidaan seuraavaksi, jos se yleensä halutaan pitää mallissa mukana. Sitten korjatuista sadoista lasketaan korjattu käsittelykeskiarvo. Nyt tulokset ovat entistä tarkempia (tietenkin edellyttäen, että uusi malli on oikea). Maaperän vaikutuksesta aiheutunut systemaattinen virhe on tasoitettu kaikkiin ruutuihin.

Tämän jälkeen kaikki vaiheet voidaan toistaa lähtemällä juuri tarkennetuista satoestimaateista. Jos jäännökset jo ovat satunnaisia ja riippumattomia, ei saada parannusta käsittelykeskiarvoihin. Muuten jatketaan kunnes tähän päästään tai jäännösten perusteella halutaan keskeyttää (malli on jo tehnyt tehtävänsä aineiston selittäjänä). Kierros kierrokselta satoihin vaikuttavat yhä kauempana sijaitsevat ruudut, ja malli monimutkaistuu.

## 2. Yleistä menetelmästä ja sen käytöstä

Yllä kuvatun ajatuksen viereisten ruutujen korrelaatioiden poistamiseksi esitti kreikkalainen Papadakis v. 1937. Seuraavana vuonna Bartlett kommentoi sitä, mutta sitten asia jäi unohduksiin 40 vuodeksi. Syynä oli Toisen Maailmansodan jälkeinen varianssianalyysin kehittyminen ja sen tarjoamat monet mahdollisuudet. Alunperin Papadakis sisällytti menetelmäänsä vain yhden korjauksen, mutta myöhemmin hän lisäsi iteroinnit huomattuaan, että se johtaa tarkentuviin tuloksiin. Hän kehitti myös keinon virheen arvioimiseksi.

Uudelleen menetelmä herätti laajaa kiinnostusta Bartlettin toimesta 1978, ja sen jälkeen sitä on ryhdytty tutkimaan ja käyttämään toden teolla. Pitkän aikaa teoreettine perustelu puuttui, mutta nyt spatiaalisten stokastisten prosessien teoria tarjoaa sellaisen.

Laskennallisesti toteutus on helppoa, eikä ohjelmoiminen tietokoneelle ole vaikeaa. Esitän algoritmin vielä kerran täydellisemmin (ilman lohkoja):

1. Lasketaan käsittelykeskiarvot.
2. Lasketaan tuottavuudet = havaittu sato - vastaava käsittelykeskiarvo. Jos tuottavuus eroaa liian paljon kummankin naapurin tuottavuudesta suuntaan tai toiseen (esim. 70% yleiskeskisarvosta), ko. ruutu jätetään huomiotta tällä iteraatiokierroksella (seur. 3.- ja 4.-vaihe). Jos toisaalta tuottavuudessa on epäilyttävän suuri hyppäys naapuriruutuun nähden, ko. ruutujen välillä ei katsota olevan naapuruus-



suhdetta vaiheessa 3. Ensimmäisellä kierroksella huolehditaan varovaisuuden vuoksi siitä, että jokaiselle ruudulle jää naapuri. Siis poistetaan vain huonompi naapurisuus, ei molempia.

3. Sadosta vähennetään naapurikoealojen tuottavuuksien keskiarvo. Naapureita on siis yksi tai kaksi kappaletta.
4. Lasketaan korjauksen saaneista sadoista käsittelykeskiarvot. Jos satoestimaatti eroaa liikaa edellisen kierroksen estimaatista, sitä ei oteta mukaan.
5. Toistetaan 2.-4. -vaiheita kunnes käsittelykeskiarvoissa ei enää tapahdu muutoksia, tai tulokset alkavat huonontua.
6. Lopuksi tarkennus: kovarianssianalyysi koealan ja sen naapurien tuottavuuksien keskiarvon välillä. Sadoista vähennetään naapurien tuottavuuksien keskiarvo kovariaatin kertoimella kerrottuna. Sitten niistä lasketaan lopulliset käsittelyestimaatit.

Papadakis siis suosittaa huolellista poikkeavien havaintojen käsittelyä. Kun se tehdään oikein, hänen mukaansa 1, 2 tai 3 iteraatiokierrosta riittää. Itse asiassa algoritmin 3. vaiheessa tehdään likimääräisiä kovarianssianalyysyjä, ja viimeisen vaiheen kovarianssianalyysi on niiden tarkennus.

Voidaan osoittaa, että iteroinnit sisältävä menetelmä vastaa seuraavankaltaista mallia:

$$y_{ij} = \mu + \mu_i + \beta_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} = \frac{\alpha}{2} (\varepsilon_{i-1,j} + \varepsilon_{i+1,j}) + \delta_{ij} \quad (i=2,3,\dots,12)$$

$$\varepsilon_{1j} = \alpha \varepsilon_{2j} + \delta_{1j}$$

$$\varepsilon_{13,j} = \alpha \varepsilon_{12,j} + \delta_{13,j}$$

Tästä ei voida suoraan estimoida käsittelyvaikutuksia  $\mu_i$ , koska ei tunneta jäännösten  $\varepsilon_{ij}$  korrelaatorakennetta, ja on siis toimittava iteratiivisesti.

Naapureitten puuttuminen, joka on väistämätöntä ainakin koealueen reunoilla, voi kuitenkin vaikuttaa arvaamattomasti tuloksiin. Yhdessä ulottuvuudessa, jolloin naapureita on yksi tai kaksi, tällä ei ole suhteellisesti niin suurta merkitystä kuin kahdessa ulottuvuudessa.

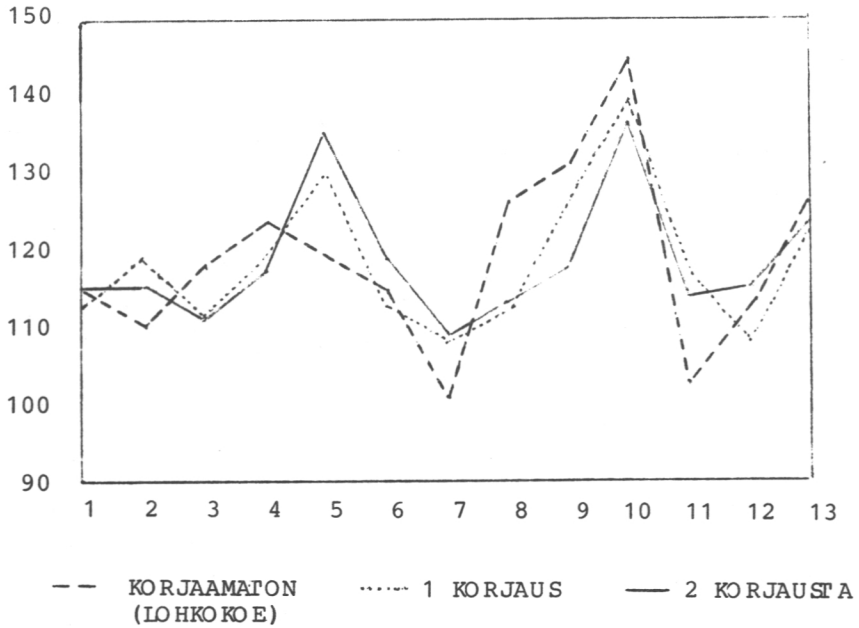
Ilmeisesti ruutujen koko ja muoto vaikuttavat niiden keskinäisiin korrelaatioihin. Näitä kysymyksiä on tutkinut ainakin Matérn. Niinpä algoritmin viimeisen vaiheen kovarianssianalyysi voidaan toteuttaa kaksisuuntaisena - silloin kun ruudut poikkeavat huomattavasti neliöstä tai on muita syitä odottaa, että korrelaatio on eri suuruista eri suunnissa.

### 3. Esimerkkiaineiston analysointia

Esimerkkiaineisto (kuvio 1) on kokeesta, jolla pyrittiin vertaamaan eri menetelmiä. Siksi sen käsittelyt ovat kuvitteellisia. On odotettavissa, että kaikki 13 antavat saman tuloksen.

Kuvasta 1 ilmenevät käsittelykeskiarvojen muutokset.

SATOKESKIARVO



Kuva 1. Käsittelykeskiarvot algoritmin eri vaiheissa.

Korjaamattomien käsittelykeskiarvojen neliösumma (keskiarvosta) on 1701, yhden korjauksen jälkeen se on 1033 ja kahden jälkeen 881. Siis Papadakis'in menetelmä vähentää

vaihtelua puoleen lohkokokeeseen nähden, ja siis puolet siitä voidaan selittää naapuriruutujen korrelaatiosta johdettavaksi.

Varianssitaulut ovat seuraavassa taulukossa.

TAVALLINEN LOHKOKOE

	D.F.	SS	MS
KÄSITTELYT	12	8416	701
LOHKOT	4	38374	9593
VIRHE	48	53759	1120
YHTEENSÄ	64	100549	

PAPADAKISIN MENETELMÄ (2 ITERAATIOA)

	D.F.	SS	MS
KÄSITTELYT	12	5113	426
VIRHE	46	24823	540
YHTEENSÄ	58	29936	

Taulukko 1. Varianssitaulut.

Vapausasteiden lukumäärä väheni 64:stä 58:ään, koska epänormaaleja ruutuja poistettiin.

Lopuksi verrataan satojen keskineliövirhettä, kun lohkoja ei oteta huomioon (korjaukset johtivat lohkoerojen häviämiseen). Tulos on:

Lohkokoe:  $(8523+53759)/(12+48)=1036$   
Papadakisin menetelmä:  $(5113+24823)/(12+46)=516$ .

Taas nähdään, kuten käsittelyille edellä, että noin puolet jäännösneliövirheestä voidaan selittää korreloinnin avulla.

Kaikkiaan Papadakisin menetelmä on usein hyvä keino tarkentaa muilla menetelmillä saatuja tuloksia, ja varsinkin silloin kun ne pettävät. Esimerkiksi kun lohkokoe epäonnistuu, lohkoista voidaan luopua ja sen sijaan eliminoida kasvupaikakatekijöitä Papadakisin iteroinneilla. Tämä on menetelmä, jota kannattaa kokeilla.

Lähteitä:

Discussion of Professor Bartlett's paper (1978). Nearest neighbour models in the analysis of field experiments. J. R. Statist. Soc. B, 40, 158-174.

Matérn, B. (1971). Stokastiska modeller för variation i planet (Den tredje nordiska konferensen i matematisk statistik). Svenska statistikersamfundet, Stockholm.

Papadakis, J. (1970). Agricultural Research: Principles, Methodology, Suggestions, 10-17, Buenos Aires.

Papadakis, J. (1984). Advances in the analysis of field experiments. Communications in the 31 May session of the Academy of Athens.

Ripley, B. D. (1981). Spatial Statistics, 94-97, John Wiley & Sons.

	L1	L2	L3	L4	L5					
1	13	7	2	8	6	83	88	88	63	132
2	2	6	8	10	11	85	113	110	102	113
3	11	2	4	5	1	72	116	121	132	129
4	9	8	3	12	2	66	63	77	105	110
5	4	10	7	11	12	74	105	94	99	138
6	12	9	6	7	7	85	91	116	74	138
7	6	1	11	4	10	99	91	143	182	154
8	1	5	12	13	5	83	105	158	179	160
9	3	4	13	2	13	91	66	94	151	165
10	5	11	1	9	9	105	88	151	138	201
11	7	12	5	1	3	110	74	91	121	182
12	8	13	9	3	8	140	113	160	132	253
13	10	3	10	6	4	140	107	223	113	179

Kuvio 1. Esimerkkiaineisto: 13 käsittelyn sijoittuminen 5 lohkoon ja vastaavat sadot.

Lohko	1	-44	-25	-31	-65	...
	2	-13	-2	+6	-63	...
	3	-22	-16	-3	-41	...
	4	-63	-43	+13	-7	...
	5	+17	+10	+14	+0	...

Kuvio 2. Jäännökset lohkokoeanalyysistä.

Lohko	1	-19	-12.5	+14	...
	2	-11	-1.5	+38.5	...
	3	-6	-3.5	+25.5	...
	4	-20	-18	+38	...
	5	+7	-5.5	+9	...

Kuvio 3. Korjatut jäännökset.

## ELIMINATING EDGE EFFECTS IN DISTANCE SAMPLING

Ahti Niinimaa

Abstract Let  $Y, X_1, \dots, X_n$  be points located uniformly and independently in a bounded  $R^2$  region  $S$ . An algorithm is proposed for calculating probabilities of type  $P(\min_{1 \leq i \leq n} |Y - X_i| \leq t)$ .

Key words: distance methods, edge effect

### 1. Introduction

Let  $Y, X_1, \dots, X_n$  be points distributed uniformly and independently in a  $R^2$  region  $S$ . We use notation  $\gamma_S^{(n)}(t)$  for the distribution function of the distance from an arbitrary point in  $S$  to the nearest neighbor of it, thus

$$\gamma_S^{(n)}(t) = P(\min_{1 \leq i \leq n} |Y - X_i| \leq t).$$

The distribution function depends strongly on the shape and the size of the region and the calculation of the exact value of it is troublesome even for a simple compact region (Kendall & Moran 1963). With our algorithm it is possible to get good approximations for these probabilities also in the case of highly irregular regions.

### 2. Algorithm

Let  $S$  be a bounded region not necessarily compact or connected. If  $X_1, \dots, X_n$  are points located uniformly and

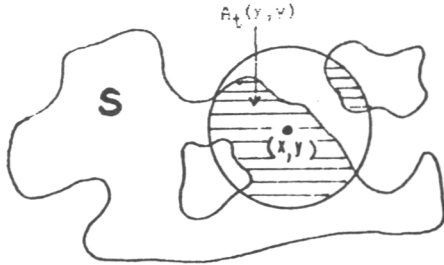


independently in  $S$  and if  $(x,y)$  is any fixed point of  $S$ , then

$$P(|Y - X_i| > t, i=1, \dots, n | Y = (x,y)) = \left(1 - \frac{A_t(x,y)}{\mu(S)}\right)^n,$$

where  $\mu(S)$  is the area of region  $S$  and  $A_t(x,y)$  is the common area of region  $S$  and the circle with radius  $t$  and centre  $(x,y)$  (fig. 1).

figure 1



Integrating over  $S$  we obtain

$$1 - \gamma_S^{(n)}(t) = \frac{1}{\mu(S)} \int \int_S \left(1 - \frac{A_t(x,y)}{\mu(S)}\right)^n dx dy.$$

If  $[a,b] \times [c,d]$  is any rectangle such that  $S$  is contained in it, we can write

$$(1) \quad 1 - \gamma_S^{(n)}(t) = \frac{1}{\mu(S)} \int_a^b \int_c^d \left(1 - \frac{A_t(x,y)}{\mu(S)}\right)^n I(x,y) dx dy,$$

where

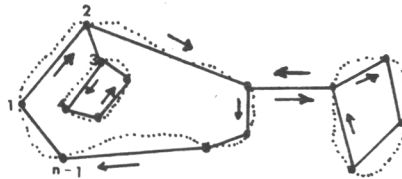
$$I(x,y) = \begin{cases} 1 & , \text{ if } (x,y) \in S \\ 0 & , \text{ otherwise.} \end{cases}$$

Using this formula we can calculate numerically the value of  $\gamma^{(n)}(t)$ .

### Approximation of region S

We approximate the boundary of the region S by line segments connecting points  $P_0, \dots, P_n$  ( $P_0 = P_n$ ). We choose the points P so that we travel the boundary of S in negative (clockwise) direction. The region is always kept to the left. If there is holes in S we enter and return from holes using the same path (fig. 2).

figure 2



### Area of S

The following formula gives exact area to the region defined by broken line  $P_0P_1 \dots P_n$  ( $P_i = (x_i, y_i)$ ) and thus gives good approximation to the area of S

$$\mu(S) \approx \sum_{i=0}^{n-1} (x_{i+1} - x_i)(y_{i+1} + y_i)/2 .$$

### Indicator function I(x,y)

Let P be an interior point of S. If we travel along the line segment starting from point P, we must cross the boundary of S odd number of times and 0,2,... times if the point P lies outside of S (fig. 3). Using this fact we obtain a simple algorithm for the indicator function  $I(x,y)$ .

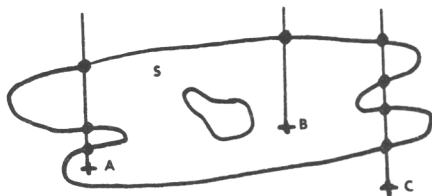


Figure 3. Let  $N^+(P)$  be the number of times vertical line segment starting from  $P$  crosses the boundary of  $S$  above the point  $P$ .  $N^+(A) = 3$ ,  $N^+(B) = 1$  and  $N^+(C) = 4$  and thus  $A$  and  $B$  are inside and  $C$  outside of  $S$ .

Program listing (FORTRAN)

```

FUNCTION  INSIDE(X,Y,XP,YF,MMM)
C  INSIDE = 0 / 1 , IF THE POINT (X,Y) LIES OUTSIDE/INSIDE
C  OF THE REGION DEFINED BY THE POINTS
C  (XP(I),YP(I)),  I=1,...,MMM
REAL XP(MMM),YP(MMM)
INSIDE=0
DO 3 I=1,MMM
X1=XP(I)
X2=XP(I+1)
IF((X2-X)*(X-X1).LT.0)GO TO 3
Y1=YP(I)
Y2=YP(I+1)
T=(Y-Y1)*(X1-X2)-(Y1-Y2)*(X-X1)
IF(T.EQ.0)THEN
INSIDE=0
RETURN
END IF
IF(X1-X2.LT.0)THEN
IF(T.GT.0)INSIDE=INSIDE+1
GO TO 3
END IF
IF(T.LT.0)INSIDE=INSIDE-1
CONTINUE
RETURN
END
3

```

Area  $A_t(x,y)$

The common area of circle  $C = \{(u,v) | (x-u)^2 + (y-v)^2 \leq t^2\}$  and the region defined by broken line  $P_0P_1 \dots P_n$  is calculated as a sum

$$A_t(x,y) = \sum_{i=0}^{n-1} \pm V_i,$$

where  $V_i$  is the common area of circle  $C$  and the triangle with corners  $(x,y)$ ,  $P_i$ ,  $P_{i+1}$  (fig. 4).

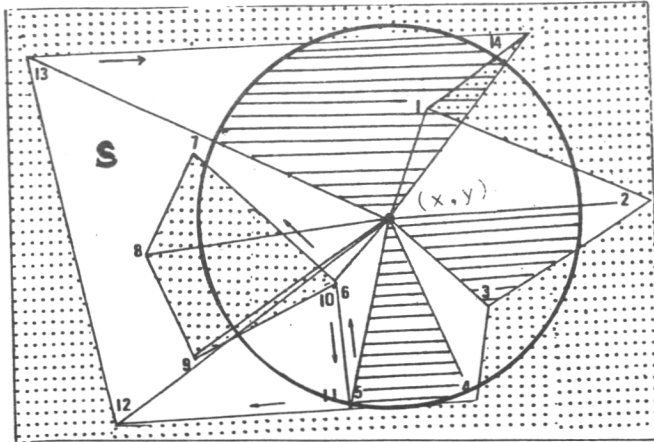


Figure 4.  $A_t(x,y) = V_1 + V_2 + V_3 + V_4 + V_5 + V_6 + V_{11} + V_{12} + V_{13} - V_7 - V_8 - V_9 - V_{10} - V_{14}$   
(shaded  $V_2$ ,  $V_4$  and  $V_{13}$ ).

Program text (FORTRAN)

parameters:

- N            Number of corner points of region S
- XP        } x- and y-coordinates of corner points
- YP        }        defining the region S
- } XP(N+1)=XP(1) and YP(N+1)=YP(1)
- X0,Y0      The center of circle
- R            Real array. Before entry R must contain the
- } t-values where the areas  $A_t(x,y)$  are computed
- TUL        Real array. On exit TUL contains the IMAX different
- } intersection areas

```
SUBROUTINE ATXX(X0,Y0,R,IMAX,XP,YP,N,TUL)
REAL R(IMAX),TUL(IMAX),XP(N),YP(N)
EXTERNAL INSIDE
IF(INSIDE(X0,Y0).EQ.0)RETURN
DO 100 L=1,N-1
  X1=XP(L)
  X2=XP(L+1)
  Y1=YP(L)
  Y2=YP(L+1)
  IK=-1
  TEST=(Y2-Y0)*(X1-X0)-(Y1-Y0)*(X2-X0)
  IF(TEST.EQ.0)GO TO 100
  IF(TEST.LT.0)IK=1
  A=(X1-X2)**2+(Y1-Y2)**2
  DX0=X1-X0
  DY0=Y1-Y0
  E1=SQRT(DX0*DX0+DY0*DY0)
  E2=SQRT((X2-X0)**2+(Y2-Y0)**2)
  EMIN=AMIN1(E1,E2)
  EMAX=E1+E2-EMIN
  APB=DX0*(X2-X0)+DY0*(Y2-Y0)
  E=SQRT(E1*E1-(APB-E1*E1)**2/A)
  FII=ACOS(APB/(E1*E2))
  DO 1 I=1,IMAX
    RRR=R(I)
    IF(RRR.GT.E)GO TO 77
    TUL(I)=TUL(I)+IK*.5*FII*RRR*RRR
    KA=I
    IF(KA.GE.IMAX)GO TO 100
    DO 2 I=IMAX,KA,-1
      IF(R(I).LT.EMAX)GO TO 88
      TUL(I)=TUL(I)+IK*.5*E1*E2*SIN(FII)
      KY=I
      IF(KY.LT.KA)GO TO 100
      W1=ACOS(E/EMAX)
      W2=ACOS(E/EMIN)
      IT=1
      IF(A.GT.EMAX*EMAX-E*E)IT=-1
      DO 3 I=KA,KY
        RRR=R(I)
        FIIR=ACOS(E/RRR)
        SINI=SIN(2*FIIR)
        D1=W1-FIIR
        PLUM=.5*(W2-FIIR)*RRR*RRR
        IF(IT.EQ.1)PLUM=-PLUM-.5*RRR*RRR*SINI
        IF(EMIN.LE.RRR)PLUM=-0.5*EMIN*RRR*SIN(FIIR+IT*W2)
    TUL(I)=TUL(I)+IK*.5*RRR*RRR*(D1+SINI)+IK*PLUM
  CONTINUE
  RETURN
END
```

Numerical integration

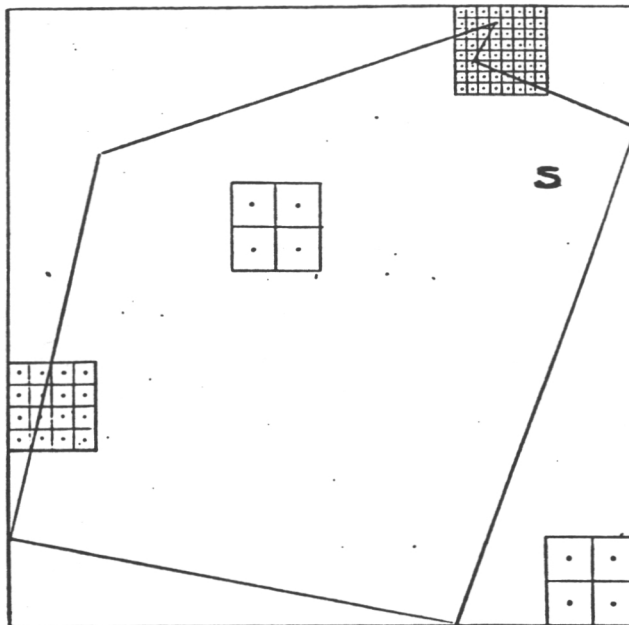
We approximate the integral (1) by dividing the rectangular  $[a,b] \times [c,d]$  into smaller rectangulars and summing over these

$$1 - \gamma_S^{(n)}(t) \approx \frac{1}{\mu(S)} \sum \left( 1 - \frac{A_t(x_1, y_1)}{\mu(S)} \right)^n I(x_i, y_i) \Delta x_i \Delta y_i .$$

The function  $A_t(x, y)$  is most poorly behaved close to the boundary of region  $S$  and therefore we use more function evaluations there. Following simple method is used:

1. Rectangular  $[a,b] \times [c,d]$  is first divided into  $k \times k$  equal parts.
2. The length of line  $P_0 P_1 \dots P_n$  is measured inside each cell and, using this length to determine the fineness of division, each subregion is split into parts (fig. 5).

figure 5



Time and accuracy

The values of  $\gamma_S^{(1)} t'$  was calculated using three different accuracy for the square  $[0, ] \times [0, ]$  and circle with radius 1.

square

t	$\gamma_S^{(1)}(t)$	numerical algorithm					
		exact value(*)	ACC=1	error%	ACC=2	error%	ACC=3
.0020	.00000387	.00000388	.09	.00000388	.09	.00000388	.09
.0079	.00005182	.00006206	.38	.00006206	.38	.00006203	.35
.0177	.00031149	.00031412	.84	.00031359	.67	.00031207	.18
.0315	.00097795	.00098921	1.15	.00097991	.20	.00097879	.08
.0492	.00236717	.00237910	.50	.00236990	.11	.00236805	.04
.0709	.00485709	.00486435	.15	.00486057	.07	.00485842	.03
.0965	.00888610	.00891440	.32	.00888415	.09	.00888805	.02
.1260	.01494002	.01495895	.13	.01494998	.07	.01494372	.02
.1595	.02353535	.02357710	.18	.02354565	.04	.02353905	.02
.1969	.03520280	.03523000	.08	.03521657	.04	.03520885	.02
.2383	.05046754	.05051969	.10	.05048599	.04	.05047309	.01
.2836	.06982877	.06990172	.10	.06987356	.06	.06985672	.04
.3328	.09373805	.09400916	.29	.09380131	.07	.09377650	.04
.3860	.12257688	.12273274	.13	.12271664	.11	.12262217	.04
CPU (s)		25		49		118	

circle The boundary of circle was approximated with broken line  $P_i = (\cos(m - i)\pi/m, \sin(m - i)\pi/m)$ ,  $i = 1, 2, \dots, m$ ,  $m = 30$  or  $60$ .

t	$\gamma_S^{(1)}(t)$	numerical algorithm						
		exact (*)	m=30		m=60		m=60	
			value	ACC=0	error%	ACC=1	error%	ACC=2
.0020	.0000039	.0000039	.12	.0000039	.12	.0000039	.12	
.0079	.0000518	.0000621	.47	.0000620	.34	.0000518	.13	
.0177	.0003108	.0003119	.35	.0003117	.27	.0003115	.20	
.0315	.0009742	.0009790	.49	.0009777	.35	.0009771	.29	
.0492	.0023529	.0023659	.55	.0023639	.47	.0023579	.21	
.0709	.0048145	.0048474	.68	.0048412	.56	.0048177	.07	
.0965	.0087787	.0088434	.74	.0088026	.27	.0087838	.06	
.1260	.0147005	.0147603	.41	.0147165	.11	.0147103	.07	
.1595	.0230497	.0231207	.31	.0230786	.13	.0230589	.04	
.1969	.0342888	.0343450	.16	.0343604	.21	.0342983	.06	
.2383	.0488482	.0489753	.26	.0489105	.13	.0488654	.04	
.2836	.0670991	.0672825	.27	.0671563	.09	.0671202	.03	
.3328	.0893253	.0894483	.14	.0894301	.12	.0893774	.04	
.3860	.1156929	.1158254	.11	.1137606	.06	.1157202	.02	
CPU (s)		30.9		146.5		335.9		

(\*) Kendall & Moran : Geometrical Probability

REFERENCES

Kendall, M.G. and Moran, P.A.P. (1963): Geometrical Probability. Griffin, London.



## EPIDEMIAN SPATIAALISET TARTUNTAMALLIT

Kari Kuulasmaa

Ekologiassa erilaiset leviämistilanteet ovat yleisiä. Tässä kirjoituksessa luokittelemme yksinkertaisia leviämismekanismia ja esittelemme matemaattisten mallien käyttömahdollisuuksia ja käytössä huomioitavia asioita leviämiseen liittyvien kysymysten selvittämiseksi.

### 1. Metsäntutkimukseen liittyviä spatiaalisia tartuntailmiöitä

Tarkastelemme tilannetta, jossa ilmiö leviää yksilöistä toisiin jonkinlaisen tartuntamekanismin mukaisesti. Tyypillinen tällainen tilanne on tarttuva tauti. Vaikka käyttämämme sana "epidemia" liittyy nimenomaan ihmisten sairauksiin, nyt käsiteltävät ilmiöt ja yksilöt voidaan ymmärtää paljon yleisemmin. Metsäntutkimukseen liittyviä tilanteita ovat esimerkiksi seuraavat:

- tarttuva sairaus puustossa. Tällöin puut ovat yksilöitä, joissa ilmiö (sairaus) leviää ja tartuntamekanismin määrittelee taudin aiheuttajan, esimerkiksi sienien, itiöiden tai hyönteisten, siirtymistä puusta toiseen.
- sienitauti puussa. Yksilöitä ovat puun solut ja tartuntamekanismin määrittelee taudinaiheuttajan lisääntymistä ja siirtymistä solusta toiseen.

- puuston leviäminen maastossa. Yksilöitä, joiden välillä leviäminen tapahtuu voivat olla maaston alueet, esimerkiksi karttaruutua vastaavat ne-  
liöt, ja "tauti" on itse puusto, joka leviää sie-  
menten välityksellä.
- myyriä leviäminen. Tämä tilanne on edellisen kal-  
tainen.
- metsäpalo. Tässä yksilöitä ovat joko puut (tai  
pensaat) tai maaston alueet, ja taudin aiheuttaja  
on tuli.

Kaikille edellä mainituista tilanteista on tyypillistä, et-  
tä taudin kantajilla, eli yksilöillä joiden keskuudessa  
tauti leviää, on kiinteät sijainnit tilassa. Juuri sijain-  
tien tärkeys tekee leviämisen spatiaalisen. Eläin-  
ten, kuten ihmisten tai kalojen epidemioita tutkittaessa  
sijainteja ei useinkaan oteta huomioon, vaan ajatellaan,  
että kuka tahansa voi tartuttaa minkä tahansa toisen yksi-  
lön. Tällöin sijainneilla ei ole merkitystä, ja puhutaan  
epäspatiaalisista tartuntailmiöistä. Käsittelemme nyt  
spatiaalista leviämistä.

Tila, jossa leviäminen tapahtuu on maantieteellisissä yh-  
teyksissä kaksiulotteinen. Solukon sairaudet leviävät kol-  
messa ulottuvuudessa. Esimerkki yksiulotteisesta leviämi-  
sestä on kasvin leviäminen rannalla tai radan varressa.

Keskitymme tässä tilanteisiin, joissa taudin kantajilla  
voidaan ajatella olevan kiinteät sijainnit, kuten edelli-  
sissä esimerkeissä. Ajalla on keskeinen merkitys leviämi-  
stä tarkasteltaessa.

## 2. Mikä spatiaalisessa leviämisessä kiinnostaa?

Tautien leviämissovellutuksissa tärkein kysymys on yleensä: miten leviäminen voidaan pysäyttää? Jos epidemiaa ei saada pysäytetyksi, miten nopeasti tauti leviää ja miten suuri osuus mahdollisista taudin kantajista sairastuu. Joillekin epidemioille on tyypillistä, että ne palaavat takaisin jonkin ajan kuluttua. Halutaan tietää milloin on seuraava epidemia ja kuinka säännöllisin väliajoin epidemia uusiutuu.

Yksi tapa lähestyä näitä kysymyksiä on matemaattinen mallittaminen: pyritään pelkistämään leviämismekanismin oleelliset osat matemaattiseen malliin, tutkitaan millaisen vastauksen mallin käyttäytyminen antaa kysymyksiin ja yleistetään mallin vastaukset alkuperäiseen tilanteeseen.

Leviämismekanismilla tarkoitamme tapaa, jolla leviäminen tapahtuu paikallistasolla: miten yksilöt tartuttavat toisiaan, kuinka kauan sairast elävät ja niin edelleen. Leviämistapaa koko populaation kannalta katsottuna sanomme leviämisen kokonaiskäyttäytymiseksi: miten laajalle ja miten nopeasti leviäminen tapahtuu, tai jääkö tauti pysyväksi.

Leviämismekanismiin perustuvilla malleilla voidaan tutkia leviämistilanteita, joista ei ole aikaisempaa kokemusta. On hyödyllistä, jos paikallinen leviämismekanismi tunnetaan hyvin. Muutoin on syytä tutkia erikseen erilaisia potentiaalisia leviämismekanismeja vastaavia malleja ja kiinnittää erityistä huomiota näiden kokonaiskäyttäytymisen eroihin. Jos esittämämme kysymyksen suhteen kokonaiskäyttäytymisissä ei ole suuria eroja, potentiaalisten leviämismekanismien erot tuskin ovat merkittäviä tarkastelemamme käyttäytymispiirteiden kannalta. Jos sen sijaan mallit käyttäytyvät hyvin eri tavoin, tutkittavan ilmiön leviämismekanismia on syytä selvittää perusteellisemmin ennen kuin mallien antamia tuloksia voidaan käyttää hyväksi.

Matemaattisia malleja voidaan käyttää myös päinvastaisissa ongelmissa: tutkittavan ilmiön kokonaiskäyttäytyminen tunnetaan, mutta paikallinen leviämismekanismi on epäselvä. Tällöin katsotaan millaisiin leviämismekanismeihin perustuvat mallit johtavat tutkittavassa ilmiössä havaittuun kokonaiskäyttäytymiseen. (Tällaisesta metsäntutkimukseen liittyvästä sovellutuksesta Rayner et. al., 1981, antavat mielenkiintoisen esimerkin.)

### 3. Mallien luokittelua

Tarkastelemme nyt yksityiskohtaisemmin eri tyyppisiä leviämistilanteita mallien pohjalta. Sama luokitus on paljolti voimassa myös epäspatiaalisilla malleilla, joista yhteenve-toja ovat kirjoittaneet mm. Bailey (1975) ja Bartholomew (1982). Näissä kirjoissa käsitellään lyhyesti myös spatiaa-

lisiä malleja. Spatiaalisia malleja ekologisten ja epidemiologisten sovellutusten yhteydessä ovat tarkastelleet mm. Mollison (1977), Kuulasmaa (1982) sekä Mollison ja Kuulasmaa (1985).

Yksinkertaisuuden vuoksi oletamme, että yksilöt sijaitsevat tason kokonaislukukoordinaattisissa pisteissä. Lisäksi oletamme, että tartuntamekanismi ei riipu tartuttajasta tai sen sijainnista.

Yksinkertaisissa epidemiamalleissa on kahdenlaisia yksilöitä: terveitä ja sairaita. Leviäminen alkaa yhdestä tai muutamasta sairaasta. Kaikki muut ovat terveitä. Tauti tarttuu sairaista terveisiin. Kerran sairastuttuaan yksilö on sairas ikuisesti ja levittää tautia edelleen. Ennen pitkää jokainen yksilö on sairas.

Tutkituimpia yksinkertaisista epidemiamalleista ovat "first passage percolation"-mallit (katso esim. Welsh, 1977). Niissä aika, jonka kuluttua sairastumistaan yksilö tartuttaa tietyn neljästä naapuristaan, noudattaa sovittua todennäköisyysjakaumaa, ja eri naapurit tartutetaan toisistaan riippumatta saman jakauman mukaisesti. Richardson (1973) on tutkinut monenlaisten yksinkertaisten epidemioiden käyttäytymistä. Osa hänen malleistaan on jatkuva-aikaisia ja osa diskreettiaikaisia.

Edellä mainituissa yksinkertaisissa epidemioissa sairas yksilö voi tartuttaa vain lähimpiä naapureitaan. Mollison (1972) toi mukaan tartuntajakauman käsitteen, jonka avulla voidaan helposti määritellä malleja, joissa mielivaltaisen pitkät tartuntaetäisyydet ovat mahdollisia.

Leviämiseen liittyvät kysymykset, joita yksinkertaisilla epidemiamalleilla voidaan selvittää, ovat leviämisnopeus ja muoto. Hyvin yleisesti voidaan osoittaa, että yksinkertaisessa epidemiassa sairaan alueen muoto lähenee asymptoottisesti (jonkin normin mukaista) palloa, jonka säde kasvaa lineaarisesti. On kuitenkin myös tilanteita, joissa taudin eturintaman nopeudella ei ole äärellistä raja-arvoa. (Mollison, 1978).

Yleisissä epidemioissa on se ero yksinkertaisiin epidemioihin, että sairaut eivät ole sairaita ikuisesti vaan ne poistuvat systeemistä sairastettuaan aikansa. Poistumisen syy voi olla esimerkiksi tartuttajan kuolema. Edellä mainittujen yksinkertaisten epidemioiden pohjalta voidaan helposti määritellä yleisiä epidemiamalleja lisäämällä tartuttajien poistamismekanismit. Lisäksi "bond percolation"- mallit, joita on paljon tutkittu (ks. esim. Welsh, 1977), ovat diskreetti-aikaisia yleisiä epidemioita. Niissä yksilö, joka sairastuu hetkellä  $t$ , tartuttaa hetkellä  $t+1$  naapureitaan jonkin satunnaismekanismin mukaisesti ja kuolee välittömästi sen jälkeen.

Yleisille epidemioille on tyypillistä kynnyskäyttäytyminen: Jos sairaut tartuttavat elinaikanaan liian vähän terveitä yksilöitä, sairaut yksilöt loppuvat ja tauti häviää. Jos sitä vastoin sairaut tartuttavat paljon terveitä, tauti leviää rajatta. Yleensä on olemassa niin sanottu kriittinen arvo jollakin parametrilla, kuten sairaan keskimääräisellä potentiaalisten tartuntojen määrällä, siten, että jos parametri on kriittistä arvoa pienempi, tauti häviää todennäköisyydellä yksi, kun taas kriittisen arvon yläpuo-

lellä on positiivinen todennäköisyys sille, että tauti leviää rajatta. Epidemian sanotaan olevan kynnyksen ala- tai yläpuolella sen mukaan, onko parametrin todellinen arvo kriittistä arvoa pienempi vai suurempi. Kriittisen parametrin olemassaolo ei aina ole itsestään selvä. On yleisiä epidemioita, tyypillisesti yksiulotteisia, joissa taudin häviäminen on varma kaikilla parametrien arvoilla (Kelly, 1977). Kuviossa 1 on esimerkki jatkuva-aikaisen yleisen epidemian leviämisestä.



Kuvio 1. Simulointiesimerkki lähellä kynnystä olevan jatkuva-aikaisen yleisen epidemian leviämisestä. Tumma piste keskellä ilmaisee leviämisen alkamiskohdan.

Vertaamalla yleisiä epidemioita niin sanottuihin haarautumisprosesseihin voidaan heti todeta yleisen epidemian olevan kynnyksellä, jos sairaut tartuttavat keskimäärin yhden terveen. Käytännössä kriittisen parametrin arvon löytäminen on kuitenkin paljon vaikeampaa, koska sairaut yrittävät tartuttaa myös jo sairaita ja poistettuja yksilöitä, jolloin osa tartuntayrityksistä jää tehottomiksi, ja tehokkaiden tartuntojen keskimäärää ei pystytä laskemaan. Muutamaa poikkeusta lukuunottamatta kriittinen arvo on esitöitä tietokonesimulointien avulla.

Haitallisten tautien yhteydessä olosuhteita pyritään muuttamaan siten, että epidemia saadaan kynnyksen alapuolelle. Toisaalta toivotuissa leviämistilanteissa on pyrittävä varmistamaan, että ollaan kynnyksen yläpuolella. Tämä ei kuitenkaan aina takaa, että leviäminen jatkuu. Yleensä riittää, että lisäksi varmistetaan hyvä alku leviämiselle. Yleisille epidemioille on nimittäin tyypillistä, että mikäli tauti häviää, se yleensä häviää hyvin alkuvaiheissaan.

Sekä yksinkertaisille että yleisille epidemioille on ominaista, että tauti ei ole samalla alueella pysyvä. Malleissa, joilla tutkitaan taudin palautuvuutta tai pysyvyyttä, on taattava uusien terveiden yksilöiden saanti poistettujen tilalle. Tällaisia palautuvia tai pusuviä epidemioita voidaan määrittellä esimerkiksi yleisten epidemioiden pohjalta siten, että lisätään mekanismi, jolla, joko itsensä tai terveiden yksilöiden vaikutuksesta, poistettuja yksilöitä korvataan uusilla terveillä yksilöillä. Yksinker-



taisia palautuvia epidemioita on tutkittu paljon lähinnä statistiseen mekaniikkaan liittyvien ongelmien yhteydessä nimellä "interacting particle systems" (ks. esimerkiksi Griffeth, 1979).

Palautuvilla ja pysyvillä epidemioilla mielenkiinto kohdistuu kynnyškäyttäytymisen lisäksi myös palautumisjaksoon ja prevalenssiin. Epäspatiaalisille palautuville epidemioille on usein tyypillistä, että epidemia palautuu voimakkaana säännöllisin välein. Spatiaalisissa malleissa sitävastoin tauti saatta levitä vaeltavina sairaina läiskinä, jotka alati hajoavat useksi läiskiksi, ja joista osa häviää. Taudin prevalenssilla tarkoitetaan osuutta populaatiosta, joka kulloinkin on sairaana. Vaikka pysyvän taudin prevalenssi saatta paikallisesti vaihdella hyvinkin voimakkaasti, laajoja alueita kokonaisuutena katsottuna sairaiden osuus pysyy likimain vakiona. On yllättävää, että hyvinkin pieni prevalenssi voi pitää taudin pysyvänä (Mollison ja Kuulasmaa, 1985).

#### 4. Mallien soveltamisesta

Matemaattisia malleja rakennettaessa tai sovellettaessa on aina aluksi tarpeen selvittää, mitä todella ollaan tutkimaan ja mihin kysymyksiin halutaan vastaus. Yleisellä tasolla nämä kysymykset on asetettava ottaen huomioon pelkääntään sovellutus. Yksityiskohdat sitävastoin voidaan täsmentää vasta, kun ongelmaa ajatellaan myös mallin kannalta. Tämä johtuu siitä, että malleihin liittyy aina rajoituksia, jotka joudutaan sovittamaan tutkittavaan tilanteeseen.

Mallien rajoitukset liittyvät niiden oletuksiin eli määrittelyn yksityiskohtiin. Mitä vähemmän tehdään oletuksia siitä yksinkertaisempi malli saadaan. Mitä yksinkertaisempi malli valitaan sitä enemmän malli pelkistää sovellutustilannetta, ja on vaarana, että ongelman kannalta tärkeitä todellisen leviämismekanismien piirteitä jää ottamatta huomioon. Toisaalta paljon yksityiskohtia sisältävän mallin rakentaminen ja tutkiminen on vaikeaa. Tällöin mallin matemaattinen tutkiminen käy helposti hyvin ongelmalliseksi. Suurempi vaara on kuitenkin siinä, että monimutkaisiin malleihin sisältyy helposti implisiittisiä oletuksia, joita ei tultu ajatelleeksi mallia valittaessa, mutta jotka saattavat vaikuttaa ratkaisevasti mallin kokonaiskäyttäytymiseen.

Mallia valittaessa joudutaan siis aina tekemään kompromisseja mallin yksinkertaisuuden ja yksityiskohtaisuuden välillä. Tärkeintä on, että mallia sovellettaessa tunnetaan mahdollisimman tarkoin siihen liittyvät oletukset. Spatiaalisten mallien yhteydessä tämä käytännössä merkitsee sitä, että lähtökohdaksi otetaan mahdollisimman yksinkertainen malli, joka kuitenkin sisältää kaikki välttämättömiksi katsotut oletukset. Asetettujen oletusten tärkeyttä voidaan selvittää tutkimalla mallien kokonaiskäyttäytymistä samalla kun oletuksia muutellaan. On huomattava, että jonkin kokonaiskäyttäytymisen piirteen suhteen tärkeät oletukset saattavat olla merkityksettömiä toisen piirteen suhteen. Viimeaikoina on mallien tutkimisessa kiinnitetty entistä enemmän huomiota eri oletusten merkitykseen (ks. esim. Molison, 1984).

Luvussa 3 esitellyistä malleista yksinkertaiset epidemiat ovat yksinkertaisimpia. Niistä ei kuitenkaan ole apua selvittäessä leviämislämmön kynnyksikäyttäytymistä tai pysyvyyttä. Näiden ongelmien selvittäminen vaatii vähintään yleisten tai palautuvien epidemiamallien käyttöä. Yksi luvun 3 oletuksista oli yksilöiden sijaintien säännöllisyys. Tämä oletus on kuitenkin paljon lievempi kuin miltä ensimmäisellä silmäyksellä näyttäisi: yksilöiden sijaintien säännöllisyyden ja populaatiotiheyden homogeenisuuden merkitystä voidaan huomattavasti vähentää laajentamalla mahdollisia tartahtaetäisyyskäsiä.

Jo hyvin yksinkertaisiin spatiaalisiin leviämismalleihin liittyy kysymyksiä, joita ei ole toistaiseksi pystytty ratkaisemaan teoreettisesti. Tällöin on turvauduttava tietokoneella suoritettaviin simuloointeihin. Simulointien antamat ratkaisut ovat yleensä oikeita. Simulointeja suoritettaessa on kuitenkin oltava varuillaan kahden seikan suhteen: Aluetta, jossa leviäminen tapahtuu, joudutaan yleensä rajoittamaan, jotta simuloinnin vaatima tietokoneen muistitila tai simuloinnin vaatima aika ei kasvaisi liian suureksi. Tämä alueen rajoitus saattaa vaikuttaa oleellisesti simuloinnin tuloksiin. Toisaalta, spatiaalisten mallien yhteydessä, generoitujen satunnaislukujen säännöllisyys saattaa vaikuttaa tuloksiin.

Huolimatta edellä mainituista ongelmista ja vaaroista uskon, että spatiaalisilla malleilla on jo nykyisen tietämyksen pohjalta paljon annettavaa myös metsäntutkimukselle. Ongelmat liittyvät ensisijaisesti kvantitatiivisten päätelmien tarkkuuteen. Spatiaalisissa tilanteissa leviämismekanismien ja kokonaiskäyttäytymisen välisen yhteyden intuitiivinen päättely on yleensä hyvin vaikeaa ja johtaa valitettavan usein harhaan. Tällöin jo mallien antamat kvalitatiiviset tulokset ovat arvokkaita.

#### Lähdeluettelo

Bailey, N.T.J (1975): The Mathematical Theory of Infectious Diseases, 2nd ed. Griffin, London.

Bartholomew, D.J. (1982): Stochastic Models for Social Processes, 3rd ed. Wiley, New York.

Griffeath, D. (1979): Additive and cancellative interacting particle systems. Lecture Notes in Mathematics 724. Springer, Berlin.

Kelly, F.P. (1977): Keskusteluosassa julkaisuun Mollison (1977).

Kuulasmaa, K. (1982): Threshold and Endemic Behaviour of Spatial Contact Processes. PhD thesis, Heriot-Watt University, Edinburgh, Scotland. Monistettu sarjassa Mathematics, University of Oulu, No. 1/83.

Mollison, D. (1972): The rate of spatial propagation of simple epidemics. Proc. 6th Berkley Symp. on Math. Statist. and Prob. 3, 579-614.

Mollison, D. (1977): Spatial contact models for ecological and epidemic spread. J. R. Statist. Soc. B 39, 283-326.

Mollison, D. (1978): Markovian contact processes. Adv. Appl. Prob. 10, 85-108.

Mollison, D. (1984): Simplifying simple epidemic models. Nature 310, 224-225.

Mollison, D. ja Kuulasmaa, K. (1985): Spatial epidemic models: theory and simulations. Kokoomateoksessa "The Population Dynamics of Wildlife Rabies", toim. Bacon, P.J.. Academic Press, London. Ilmestyy lähiaikoina.

Rayner, A.D.M., Bevercombe, T.C., Brown, T.C. ja Robinson, A. (1981): Fungal growth in a lattice: a tentative explanation for the shape of diamond-cancers in syncamore. New Phytol. 87, 383-393.

Richardson, D. (1973): Random growth in a tessellation. Proc. Camb. Phil. Soc. 74, 515-528.

Welsh, D.J.A. (1977): Percolation and related topics. Science Progress 64, 65-83.

## TILAJÄRJESTYS KAARNAKUORIAISTEN KILPAILUN SELITTÄJÄNÄ

Hannu Saarenmaa

### 1. Johdanto

Kaarnakuoriaisemot saapuvat keväällä puun tuoksujen ja feromonien houkuttelemina tuoreeseen puuhun tai puutavaraan, porautuvat kaarnan alle ja konstruktoivat sinne mutkikkaat käytävänsä (ks. esim. Saalas 1949), joiden varsille munat munitaan. Jotta puun pihka ei tappaisi kaarnakuoriaisia, näiden on iskeydyttävä puuhun joukkovoimalla, mistä taas seuraa jatkossa ylitiheyttä. Niinpä nilasta, joka on toukkien ravintoa, käydään kilpailua saman emokäytävän toukkien kesken, lajin sisällä eri emokäytävien välillä sekä eri kaarnakuoriaislajien välillä. Myös pedot, loiset, taudit, pihka, ym. ottavat veronsa, mutta useimmiten kilpailu on kaarnakuoriaisten suurin kuolleisuustekijä. Pystynävertäjällä (*Tomicus piniperda*) yleensä 5 - 20 % alunperin munituista munista selviää aikuisiksi asti (Saarenmaa 1985); mitä suurempi emojen iskeytymistiheys, sitä alhaisempi eloonjääminen. Vaihtelu lisääntymistuloksissa on kuitenkin valtavaa; esim. pystynävertäjällä tuotos voi vaihdella 1000 - 6000 /m<sup>2</sup> kuoripinta-alaa samassa tiheydessä kontrolloiduissa laboratorio-oloissa (Eidmann & Nuorteva 1968).

Koska kaarnakuoriaiset elävät lisääntymiskautenaan puun ja kuoren välisessä nilakerroksessa kaksiulotteisessa jatkuvassa tilassa ja kilpailevat siellä ravinnosta, on ilmeistä, että tilajärjestyksellä on merkitystä niiden kuolleisuudelle. Tätä selvitettiin kahdella tavalla: (Luku 3) askeltavalla regressiolla, jossa selittävinä muuttujina olivat emojen tiheys ja erilaiset tilajärjestysindeksit (Saarenmaa 1983), sekä (Luku 4) simuloimalla toukkien liikkumis- ja syöntiprosessia kaarnan alla (Jong & Saarenmaa 1985).

## 2 Aineisto

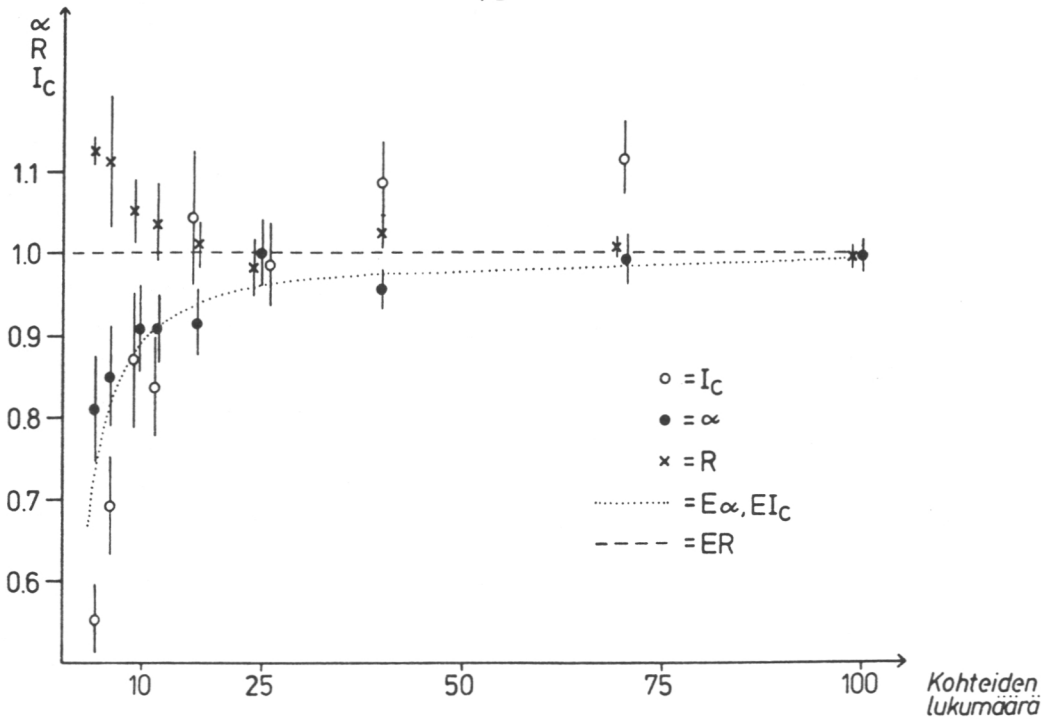
Tutkimuksen kohteena oli pystynävertäjä, joka on männyn tärkein kaarnakuoriaistuholainen. Pystynävertäjien annettiin vapaasti iskeytyä 83:een noin 40 cm pituiseen koepölkkyyn laboratoriossa ja maastossa. Kun kaikki jälkeläiset olivat kuoriutuneet, pölkyistä kartoitettiin emokäytävät numeerisesti mittanauhalla. Mittauksen vuoksi pöllisiin vedettiin keinotekoinen nollaviiva, mutta kaikki koordinaatit projisoitiin toiseen kertaan tämän yli jatkuvan pinnan aikaansaamiseksi.

## 3. Tilajärjestysindeksit selittäjinä

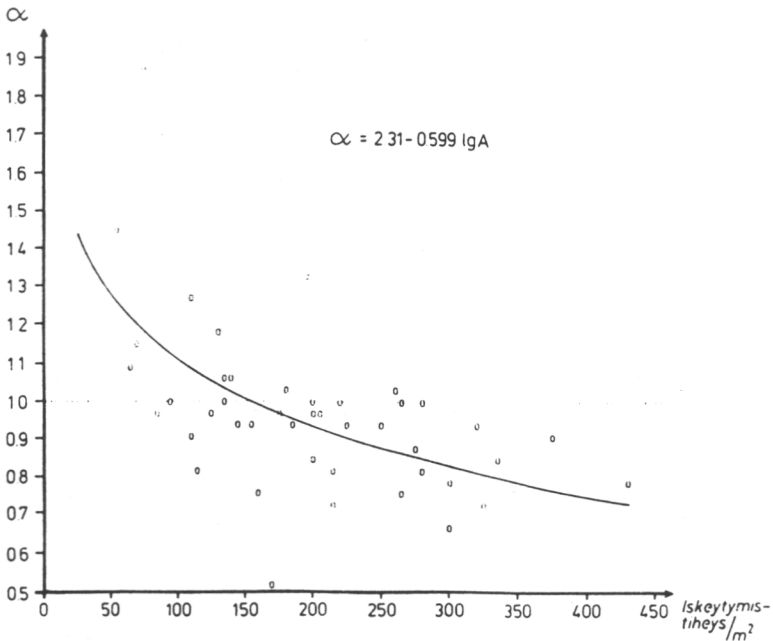
Pöllien kartoitus mahdollisti minkä tahansa tilajärjestysindeksin käyttökelpoisuuden testauksen. Tutkittavaksi valittiin (1) nearest-neighbor R (Clark & Evans 1954), (2) point-to-individual alfa (Pielou 1959, 1977) sekä (3) nollaruutudiagrammaan perustuva Ic (Cox 1971, Pohtila 1980).

Jotta reunaefekti ei aiheuttaisi systemaattista virhettä tuloksiin, pöllin päähän rajattiin reunus, jolta mittauksia ei aloitettu. Simuloinnin tuloksena tämän reunuksen leveydeksi saatiin  $0.75 \cdot$  yksilöiden keskimääräinen välimatka satunnaisessa tilajärjestyksessä (Saarenmaa 1983 kuva 4).

Jotta alhainen iskeytymistiheys (= yksilöiden lukumäärä) ei aiheuttaisi systemaattista virhettä indekseihin, sen vaikutusta simuloitiin. Osoittautui, että R antoi harhattomia tuloksia vasta kun pöllissä oli yli 12 pistettä, alfa oli harhaton jo 6 pisteessä, mutta Ic:tä ei saatu lainkaan harhattomaksi käytetyillä menetelmillä (kuva 1, Saarenmaa 1983 kuva 5). Tämän vuoksi jatkossa käytettiin vain alfaa, jonka oletusarvo skaalattiin tiheydestä riippumattomaksi  $(n-1)/n$ :stä 1:een regressioanalyysija varten.



Kuva 1.  $R$ :n,  $\alpha$  ja  $I_C$ :n keskiarvot ja yksikertaiset keskivirheet satunnaislukugeneraattorilla tehdyissä tilajärjestyksissä kohteiden lukumäärän funktiona.



Kuva 2. Emokäytävien alkukohtien skaalatun  $\alpha$  riippuvuus iskeytymistiheydestä. Alhaisessa tiheydessä ryhmittäinen ja suuressa säännöllinen.



Pystynävertäjän emokäytävien alkukohtien (sisäänmenoreikien) tilajärjestys oli ryhmittäinen alhaisessa tiheydessä, satunnainen keskitiheydessä ja lievästi säännöllinen suuressa tiheydessä (kuva 2, Saarenmaa 1983 kuva 7). Tämä tulkittiin siten, että ensimmäisenä saapuvat emot valtaavat pöllistä lämpötilaltaan parhaat paikat, mutta kun tiheys kasvaa, yhä epäsuotuisampia kohtia täytyy valita (yläpuolella lämpötila voi auringonpaisteessa nousta 40 °C asti ja alapuolella taas keväällä on vain noin 10 °C). Samalla kun tilajärjestys muuttuu säännölliseksi, iskeytyminen päättyi: yli 400 /m<sup>2</sup> tiheyksiä ei luonnossa juuri tapaa. Tämä on osoitus siitä, että emot pitävät tietyn minimivälimatkan iskeytymisten välillä.

Pelkkä iskeytymistiheys selitti 30-43 % jälkeläistuotoksen varianssista. Kun alfa lisättiin malleihin, selitysaste ei noussut tilastollisesti merkitsevästi (Saarenmaa 1983 taulukko 10). Tämän tulkittiin johtuvan pääasiassa siitä, että varsinainen kilpailu tapahtuu laajalla alalla nilassa eikä iskeytymiskohdassa. Tämän vuoksi emojen iskeytymiskohtien tilajärjestyksellä ei ole suurta merkitystä kilpailulle ja jälkeläistuotokselle. Jos emot voisivat säädellä myös käytävien kulkua kaarnan alla, vaikutus voisi olla selvempi, mutta luonnossa kaksi toisistaan kaukanakin alunperin ollutta käytävää voi suorastaan törmätä toisiinsa kaarnan alla. Edelleen, toukkakäytävien kulku toisiinsa nähden on ratkaisevampi tekijä kuin emokäytävien. Tämän vuoksi tilastollinen tilajärjestyksen tutkiminen todettiin tässä tapauksessa hedelmättömäksi.

#### 4. Mekanistinen simulointimalli kilpailulle

Mallin perusajatus on esittää nila kaksiulotteisena taulukkona, jonka jokainen elementti edustaa 2\*2 mm nilapinta-alaa. Toukat liikkuvat elementistä toiseen kuluttaen niitä siten, että toinen toukka, joka saapuu ko. elementtiin, ei saa siitä ravintoa. Kuitenkin toukat pysyvät käyttämättö-

missä elementeissä niin kauan kuin mahdollista. Mallin, joka tavallaan muistuttaa shakkia tai videopeliä, oletukset ja rakenne ovat seuraavat (Jong & Saarenmaa 1985):

- 1) Aika on vakio kaikille ja jokaisella aika-askelella jokainen toukka joko a) liikkuu ja kasvaa vakiomäärän, b) kuoriutuu tai c) kuolee.
- 2) Emokäytävien tilajärjestys on annettu.
- 3) Munatiheys 0.332/mm on empiirinen vakio emokäytävän varrella, kuten myös emokäytävän munattoman alkuosan pituus 7.7 mm.
- 4) Luonnollisen ikärakenteen jäljittämiseksi toukat kuoriutuvat munista 0.35 aika-askelen välein.
- 5) Toukat liikkuvat preferenssillä 0.95 pitkittäin puun syiden suuntaan.
- 6) Ennen kuoriutumista kaivettavien toukkakäytävien pituus noudattaa empiiristä beta-jakaumaa vaihdellen 52 - 150 mm, keskimäärin 94 mm.
- 7) Toukat "kuuntelevat" toisiaan tietyn säteen sisällä ja suuntautuvat pois toisistaan kilpailun välttämiseksi. Säteeksi iteroitiin 8 mm.
- 8) Jos toukka ajautuu umpikujaan, missä se on kulutetun nilan ympäröimä joka puolella, sillä on kaksi mahdollisuutta: jos se on saavuttanut minimikäytävän-pituuden, joksi iteroitiin 46 mm, se kuoriutuu vaikka se ei olisi vielä saavuttanut empiirisestä beta-jakaumasta generoitua satunnaista kuoritumispainoaan, mutta...
- 9) jos toukka ei kuitenkaan ole vielä tehnyt minimikäytävää, se ryhtyy liikkumaan jo kulutetuissa elementeissä. Jos se ei selviä kuluttamattomille elementeille ennenkuin sen ravintovarot ehtyvät, se kuolee. Ravintovaroiksi iteroitiin 10 mm liikkumismatkaa kulutetuilla elementeillä vastaava käytävän pituus.

Time 0  
Mortal-0.0  
ity-%

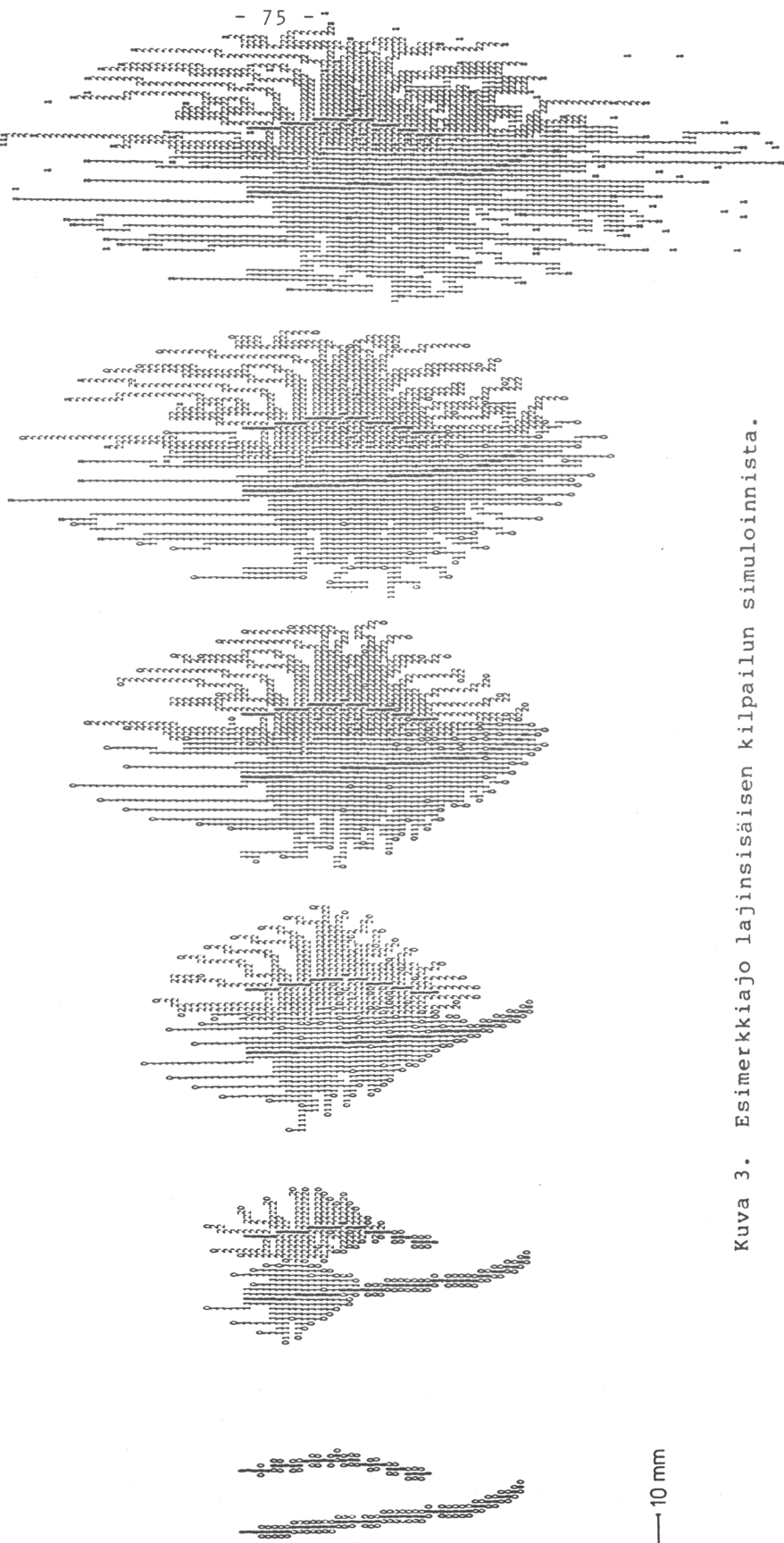
10  
0.0

20  
0.8

30  
25.8

40  
45.8

71  
50.0



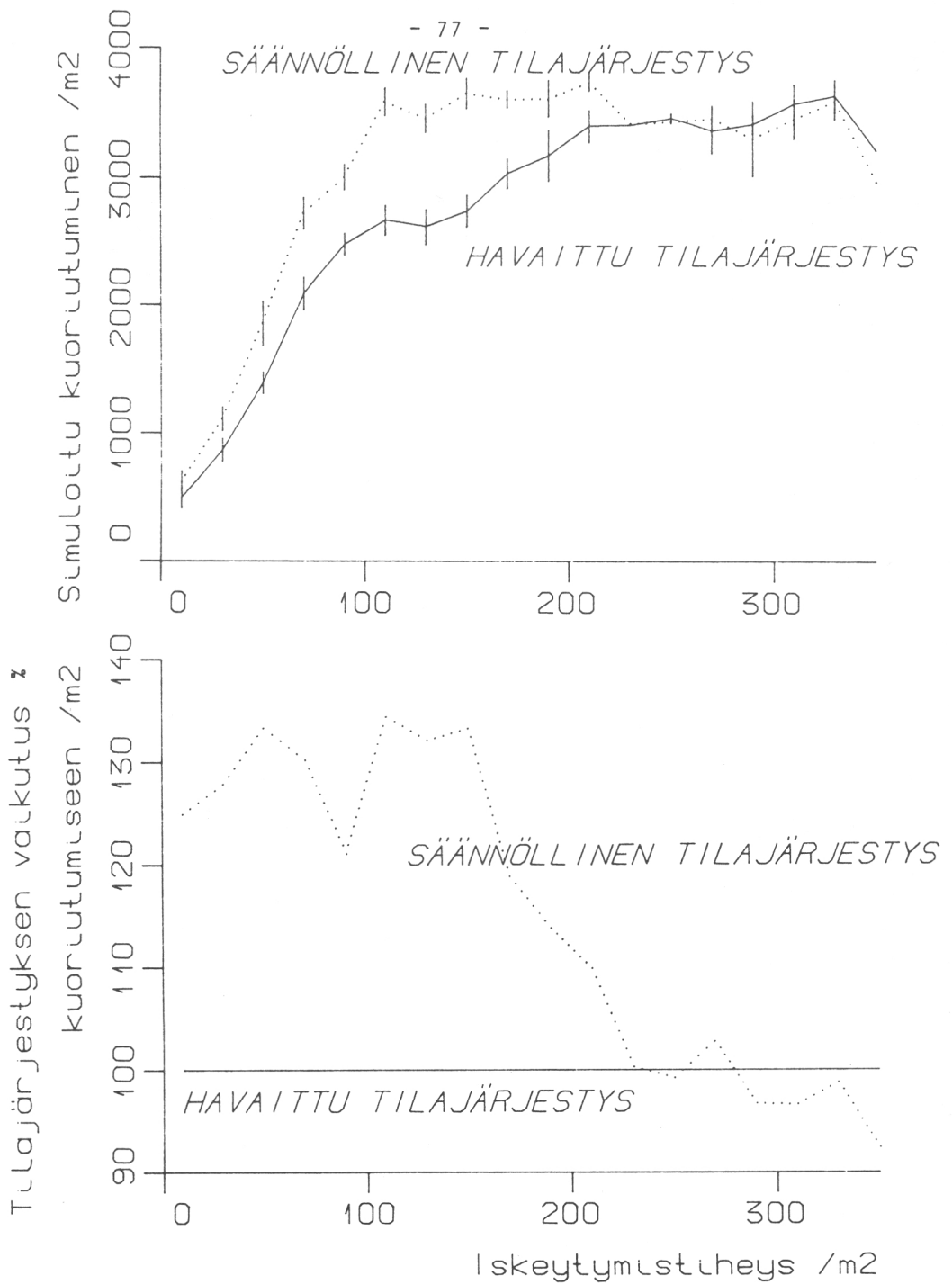
Kuva 3. Esimerkkiajo lajinsisäisen kilpailun simuloinnista.

— 10 mm

Kuva 3 näyttää esimerkkiajon eräästä pöllistä, jossa kahden emokäytävän toukat kilpailevat keskenään. Kun pöllissä voi olla kaksi tuhatta toukkaa, jotka kulkevat keskimäärin noin kaksikymmentä askelta tutkien ympäristöönsä suuressa taulukossa, on ymmärrettävää, että mallin käyttö ja edelleen kehittäminen vaatisi supertietokonetta. Nykyisin käytettävissä olevilla virtuaalimuistia käyttävillä superminikoneilla CPU-ajan käyttö on valtavaa.

Herkkyyksianalyysi paljasti, että malli on hyvin robusti lukuunottamatta muutoksia kolmessa parametrissa, jotka olivat minimikäytävänpituus, ravintovarot kulutetulla alueella ja "kuuntelu". Malli ei toiminut lainkaan, jos toukat eivät kuuntele ympäristöönsä. Niinpä pääteltiin, että nämä ovat oleellisia mekanismeja kaarnakuoriaisten lajinsisäisessä kilpailussa. Malli selittää noin 58 % jälkeläistuotoksen varianssista ja on siis selitysvoimaltaan selvästi regressiomalleja parempi. Tällainen biofyysinen mekanistinen malli on myös ekstrapoloitavissa (ks. Saarenmaa 1983 kuva 18) niin kauan kuin perusprosessien voidaan olettaa olevan voimassa päinvastoin kuin regressiomallit, joilta biologinen pohja usein puuttuu.

Malli mahdollistaa myös kokeilut erilaisilla tilajärjestyksillä. Näin verrattiin jälkeläistuotoksia havaitussa tilajärjestyksessä (joka muuttui siis tiheyden myötä kuten kuva 2 osoittaa) sekä sellaisessa säännöllisessä tilajärjestyksessä, jossa emokäytävien minimietäisyydet oli iteratiivisesti maksimoitu. Tuloksena oli (kuva 4), että säännöllisessä tilajärjestyksessä jälkeläistuotos /m<sup>2</sup> kaarnapinta-alaa on tilastollisesti merkitsevästi suurempi noin tiheyteen 200 emokäytävää /m<sup>2</sup>, mutta tämän jälkeen ero katoaa. Tämä johtuu siitä, että myös havaittu tilajärjestys alkaa tällöin olla säännöllinen. Ero on suurimmillaan noin 30 %. Jos vastaavat kuvat tehdään lisääntymiskertoimelle, tulos on samankaltainen. Säännöllisellä tilajärjestyksellä näyttäisi siis olevan jonkin verran merkitystä lajin lisääntymiselle. Kun tilajärjestys ei kuitenkaan ole säännöllinen jo



Kuva 4. Ylempänä simuloidut jälkeläistuotokset /m<sup>2</sup> kaarnapinta-alaa emojen iskeytymistiheyden funktiona havaitussa ja maksimaalisessa säännöllisessä tilajärjestyksessä. Yksikertaiset keskivirheet esitetty pystyviivoin. Alempi kuva: sama suhteellisena.

alhaisissa tiheyksissä, mistä näyttää koituvan lajille menetyksiä, herää kysymys, onko tilajärjestys sittenkään kovin tärkeää luonnon oloissa? Ilmeisesti iskeytyvillä pystynävertäjäemoilla on muutakin tekemistä kuin huolehtia tilajärjestyksestä. Suurissa tiheyksissä tavattava säännöllien tilajärjestys onkin paremmin tulkittavissa sellaisen mekanismin sivutuotteeksi, jonka tärkein tehtävä on katkaista iskeytyminen puun jo tultua täyteen, eikä sillä ilmeisesti pyritäkään maksimaaliseen jälkeläistuotokseen.

## 5. Viitteet

- Clark, P.J. & Evans, F.C. (1954): Distance to nearest neighbor as a measure of spatial relationship in populations. *Ecology* 35, 445-453.
- Cox, F. (1971): Dichtebestimmung und Strukturanalyse von Pflanzenpopulationen mit Hilfe von Abstandsmessungen. *Mitt. Bundesforschungsanstalt Forst- und Holzw.* 87, 1-182.
- Eidmann, H.H. & Nuorteva, M. (1968): Der Einfluss der Siedlungsdichte und anderer Faktoren auf die Anzahl der Nachkommen von Blastophagus piniperda L. (Col., Scolytidae). *Ann. Entomol. Fenn.* 34, 135-148.
- Jong, M.C.M. de & Saarenmaa, H. (1985): A mechanistic simulation model for the movement and competition of bark beetle larvae. *Ecol. Model.* (In print).
- Pielou, E.C. (1959): The use of point-to-plant distances in the study of the pattern of plant populations. *J. Ecol.* 47, 607-613.
- (1977): *Mathematical ecology.* 385 pp. John Wiley & Sons, New York.
- Pohtila, E. (1980): Havaintoja taimikoiden ja nuorten metsien tilajärjestyksen kehityksestä Lapissa. *Commun. Inst. For. Fenn.* 98, 1-35.
- Saalas, U. (1919): *Suomen metsähyönteiset.* 719 pp. WSOY, Porvoo.
- Saarenmaa, H. 1983. Modeling the spatial pattern and

intraspecific competition in Tomicus piniperda  
(Coleoptera, Scolytidae). Commun. Inst. For. Fenn.  
118, 1-40.

- (1985): Within-tree population dynamics models for  
integrated management of Tomicus piniperda  
(Coleoptera, Scolytidae). Commun. Inst. For. Fenn.  
(In print).

## KUVA-ANALYYSI RANSKALAISEEN TAPAAN

Pentti Suomela

Ranskalaisuus viittaa tässä Fontainebleaussa sijaitsevaan matemaattisen morfologian tutkimuskeskukseen, joka toimii Pariisin vuoriteollisen korkeakoulun yhteydessä (Centre de Morphologie Mathématique, École des Mines de Paris). Sen primus motor on Georges Matheron, jonka väitöskirja spatioaalisten muuttujien estimoinnista ilmestyi 1965. Hänen luomansa koulun työkenttä on laajentunut maaperä- ja mineraalinäytteiden kovarianssianalyysistä kuva-analyysin kokosaralle. Koulun tavoitteena on matemaattisesti hyvin perustellun kuva-analyysin metodiikan ja siihen liittyvän automatiikan ('texture analysers') kehittäminen. Tuore esitys Fontainebleaun koulun saavutuksista on Jean Serran kirja *Image analysis and mathematical morphology* (1982).

### 1. Muunnosten tuoma informaatio

Tässä esityksessä kuva tarkoittaa valmiiksi rasteroitua kuvaa, jossa on vain kaksi harmaan sävyä, musta ja valkoinen. Kuva koostuu mustasta kuviosta  $X$  ja sen valkoisesta taustasta  $X^C$ . (Tämä on esityksen lyhyden sanelema rajoitus, jota Matheron ja kumppanit eivät suinkaan tee.)

Tunnusomaista ranskalaiselle kuva-analyysille on kuvien (kuvioiden) muuntaminen ennen mittausten suorittamista: kuvion käyttäytyminen erilaisissa muunnoksissa kertoo siitä asioita, joita voi olla suoraan vaikea tavoittaa. Muunnoksilla pyritään epäolennaisuuksien karsimiseen kuvioista ja mittausten palauttamiseen sellaisiin helppoihin toimituksiin



kuin pikselien tai komponenttien lukumäärän laskeminen. Lopussa esitettävissä esimerkeissä kaikki mittaukset palautuvat muunnetun kuvion pikselien lukumäärän laskemiseen. Analyysin etenemisjärjestys on silloin seuraava:

kuvio  $X \rightarrow$  muunnos  $T(X) \rightarrow$  mittaluku  $\#T(X) \rightarrow$  tunnusluku  $m(X)$ ,

missä  $m(X)$  on haettu alkuperäisen kuvion tunnusluku.

## 2. Omaleimaisuutta analyysin kaikissa vaiheissa

Ranskalaiset ovat tuoneet oman panoksensa kuva-analyysin kaikkiin vaiheisiin, kuvioiden mallintamisesta niiden morfologisten tunnuslukujen estimointiin asti.

Kuvioiden mallintamiseksi Matheron on kehittänyt satunnaisjoukkojen teoriaa ja stokastista geometriaa (ks. Matheron, 1975), jossa topologia on valittu morfologian tarpeita silmällä pitäen ('hit or miss topology'). Tyypillisesti topologinen ongelma on kuvion reunan määrittäminen, kun taas todennäköisyyslaskenta tapahtuu topologian virittämissä Borel-joukkojen perheessä, missä Matheronin lähestymistavan omaperäisyys häipyy (ks. Ripley, 1976). Kuvioiden mallintamisen saralla näyttää DDR:n pisteprosessikoulu, jolla on tukikohta myös Freibergin vuoriakatemiassa (!), etenevän ranskalaisia voimallisemmin.

Kuvioiden muunnosten teoriaa Fontainebleaun koulu on kehittänyt matemaattisen morfologian nimellä (ks. Serra, 1982). Ns. Serran kalkyyllissä kaikki morfologiset muunnokset palautetaan rakenteellisten alkioiden ('structuring elements') avulla suoritettaviin eroosioihin ja dilataatioihin (ks. al.).

Muunnosten valintaan ranskalaiset ovat esittäneet erinäisiä mittauksilta toivottuihin ominaisuuksiin perustuvia kriteerejä, joista Serra (1982) käsittelee mm. seuraavia:

- kovarianssi,
- suuruuskriteerit (granulometriat),
- muotokriteerit (erit. konveksisuus-kriteerit),
- yhtenäisyyskriteerit.

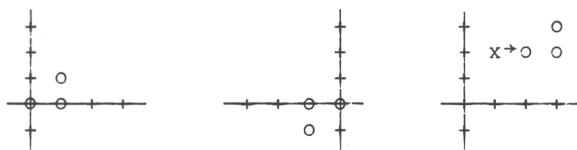
Ellei kyse ole diskreetin kuvan analyysistä, huomioon on otettava myös digitalisoituvuus-kriteerit (estimoinnin tarkentuvuus rasterointia hienonnettaessa). Muunnosten valinnassa on ensin kyse rakenteellisen alkion valinnasta ja sitten sopivan eroosioiden ja dilataatioiden yhdistelmän löytämisestä. Vaikka kriteerit sanelevatkin muunnosten valinnan varsin pitkälle, käytännön kuva-analyysiin tarvitaan hyviä morfologisten muunnosten kirjastoja. Niitä on laadittu vasta joihinkin erityisiin tarkoituksiin.

Alkuperäisen kuvion morfologiset tunnusluvut lasketaan sopivasti muunnetun kuvion mittaluvuista. Ihanteena on muunnosten sarja, joka antaa mitta- ja tunnusluvut helposti automatisoitavilla algoritmeilla. Tässä rakenteellisen alkion käyttö on ilmeisen hyvä oivallus. Globaalien mittaus-ten ohella voidaan käyttää stereologisia ja tilastollisia menetelmiä.

Fontainebleaun koulun kehittälemässä kuva-analyysin metodiikassa on vielä aukkoja mm. tilastollisten menetelmien perusteluissa ja kuvioiden rakenteiden (tekstuuri) mallintamisessa ja etenkin mallien hyödyntämisessä. DDR:n pisteprosessikoulu on paikkaamassa näitä aukkoja (ks. Stoyan und Mecke, 1982).

### 3. Alkeismuunnokset

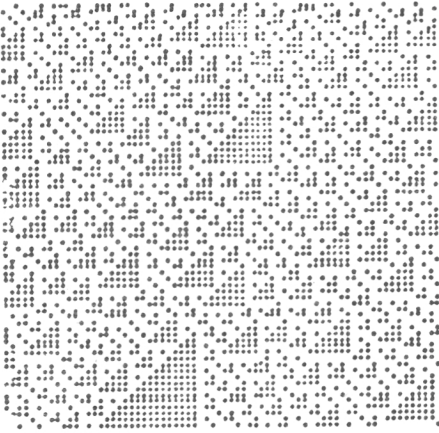
Tarkastellaan neliöhilan  $Z^2$  pisteistä koostuvia kuvioita. Rakenteelliseksi alkiksi käy mikä tahansa  $Z^2$ :n osajoukko  $B$ , joka sisältää origon. Merkitään  $\check{B}$ :llä  $B$ :n peilikuvaa origon suhteen ja  $B_x$ :llä joukkoa, joka saadaan  $B$ :stä  $Z^2$ :n vektorin  $x$  suuruisella siirrolla (kuva 1).



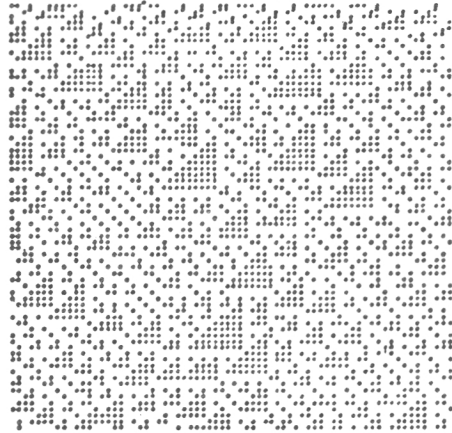
Kuva 1. Rakenteellinen alkio  $B$  ja sen muunnokset  $\check{B}$  ja  $B_x$ .

Eroosio alkiolla  $B$  jättää kuvioista  $X$  jäljelle ne pisteet  $x$ , joissa  $B_x$  sisältyy kokonaan  $X$ :ään; saatua kuviota merkitään  $X \ominus \check{B}$ . Dilataatio alkiolla  $B$  lisää kuvioon  $X$  ne pisteet  $x$ , joissa  $B_x$  leikkaa  $X$ :ää; saatua kuviota merkitään  $X \oplus \check{B}$ . Peräkkäinen eroosio  $B$ :llä ja dilataatio  $\check{B}$ :lla muuntaa kuvion  $X$  suppeammaksi kuvioksi  $X_B = (X \ominus \check{B}) \oplus B$ , joka saadaan kokonaan  $X$ :ään sisältyvien  $B_x$  yhdisteenä; tätä operaatiota sanotaan  $B$ :llä avaamiseksi. Peräkkäinen dilataatio  $B$ :llä ja eroosio  $\check{B}$ :lla muuntaa kuvion  $X$  suuremmaksi kuvioksi  $X^B = (X \oplus \check{B}) \ominus B$ , joka saadaan joukkojen  $B_x$  yhdisteenä, kun  $x$  käy läpi kaikki  $X$ :n pisteet; tätä operaatiota sanotaan  $B$ :llä sulkemiseksi. Kuten sanottua, Serran kalkyyllissä kaikki morfologiset muunnokset saadaan eroosioiden ja dilataatioiden yhdistelminä niin kuin avaaminen ja sulkeminenkin.

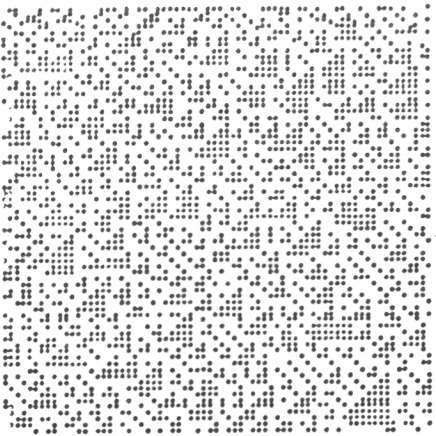
Kun rakenteellisena alkiona on kuvan 1 mukainen pienin kolmio  $\therefore$ , mainitut neljä muunnosta vaikuttavat kuvioon 2-a seuraavasti: eroosio poistaa erilliset pisteet ja kolmioiden oikeat reunat; dilataatio ja sulkeminen täyttävät kuvan laitoja lukuunottamatta; avaaminen poistaa vain erilliset



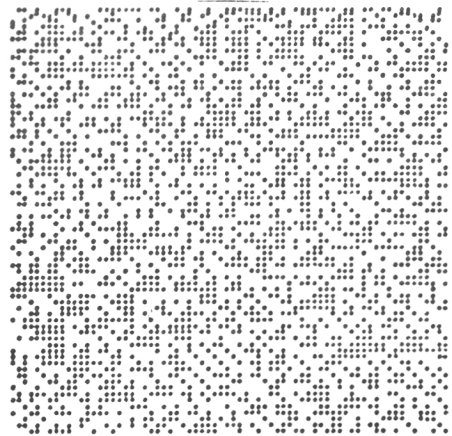
2-a



2-b



2-c



2-d

Kuva 2. Eräs Markov-aihe puhtaana ja hämärrettynä.

pisteet.

#### 4. Muunnosten käyttö kuva-analyysissä

Jonkinlaisen aavistuksen muunnostekniikan käyttökelpoisuudesta antaa ehkä kuvan 2 kuvioiden analyysi ranskalaiseen tapaan. Ne ovat simuloituja Markov-kuvioita, joiden rakenne on tiedossani (ks. Suomela, 1979).

Kuvan 2-a granulometria (kolmioiden suuruusjakauman määrittäminen) voidaan perustaa kuvion avaamiseen erikokoisilla kolmioilla

.    .:    .::    .:::

Merkitään k-kantaisella kolmiolla avaamisen tuottamaa kuviota  $X_k$ :lla. Yhdellä pisteellä avaaminen antaa tietysti kuvion  $X = X_1$  itsensä. Kolmen pisteen kolmiolla avaaminen poistaa kuviosta erilliset pisteet, joten niiden lukumäärä saadaan erotuksena

$$n(1) = \#X_1 - \#X_2 .$$

Kuuden pisteen kolmiolla avaaminen poistaa kuviosta  $X_2$  2-kantaiset kolmiot, joten niiden lukumäärä saadaan las-  
kulla

$$n(2) = \frac{1}{3} (\#X_2 - \#X_3) .$$

Näin edeten saadaan peräkkäin määrättyksi lukumäärät

$$n(k) = \frac{2}{k(k+1)} (\#X_k - \#X_{k+1}), \quad k = 1, 2, \dots .$$

Kolmioiden teoreettinen suuruusjakauma kuvassa 2-a on geometrinen parametrilla  $1/2$ , niin että suhteen  $n(k):n(k+1)$  odotusarvo on  $2:1$  kaikilla  $k$ . Nopeampi estimointikeino lukumäärien  $n(k)$  suhteille saadaan siitä, että pistejuoksujen pituuksilla on niin riveillä kuin sarakkeilla sama teoreettinen jakauma kuin kolmioiden suuruuksilla.

Entä mikä on komponenttien suuruusjakauma kuvassa 2-b, jossa kolmioaihe ei ole aivan puhdas? Periaatteessa senkin arviointiin voitaisiin käyttää äskeistä seulontamenettelyä, mutta komponenttien vaihtelevat muodot tekevät jäykällä rakenteellisella alkiolla seulonnan epätarkaksi. Tällaisissa tapauksissa Serra kehottaa ensin siistaamaan kuviot avaamalla tai sulkemalla ne siten, että komponentit tulevat selväpiirteisemmiksi. Siistauksen hintana saattaa kuitenkin kadota sellaisia detaljeja, joissa on pikkupiirteisen kuvioidinnin koko salaisuus.

Hilakuvioiden kovarianssianalyysi voidaan perustaa eroosioille pistepareilla  $B = \{0, h\}$ , missä  $h$  on hilan vektori. Joukon  $X \in \check{B}$  mahtavuus on sinänsä eräs globaali kovarianssitunnus. Jos kyse on stationaarisesta satunnaiskuviosta (niin kuin kuvissa 2-a-d), joukon  $X \in \check{B}$  suhteellinen koko ikkunaan verrattuna käy lokaalin todennäköisyyden  $C(h) = P\{x \in X, x+h \in X\}$  estimaattoriksi. Vastaavasti kuvion  $X$  osuus koko kuvasta käy todennäköisyyden  $p = P\{x \in X\}$  estimaattoriksi. Jos  $I_X$  merkitsee joukon  $X$  indikaattoria, näiden avulla saadaan muuttujien  $I_X(x)$  ja  $I_X(x+h)$  kovarianssille estimaattori

$$\text{Cov}(h) = C(h) - p^2 .$$

Kuvan 2 kuvioiden eroista parittaiset kovarianssit eivät kerro mitään, sillä jokaisessa kuvassa kaikki muuttujat

$I_X(x)$  ja  $I_X(x+h)$ ,  $h \neq 0$ , ovat (mallin mukaan) korreloimattomia. Niissä  $p:n$  odotusarvo on  $1/2$  ja  $C(h):n$  odotusarvo  $1/4$  kaikilla  $h \neq 0$ . Kuvien rakenteen tyhjentävään kuvaukseen tarvitaan kolmen tai useamman muuttujan yhteisjakautumia. Niiden Markov-ominaisuuden paljastaminen käy luontevimmin lokaaleja vuorovaikutuksia tutkimalla (vrt. Antti Penttisen puheenvuoroon tässä seminaarissa), joskin sekin voidaan tehdä ranskalaiseen tapaan.

Kuvasta 2 kerrottakoon sen verran, että sen matemaattisessa mallissa on vain yksi vapaa parametri, nimittäin ehdollinen todennäköisyys että pienimpään kolmioon  $\cdot$ : tulee kaksi kuvion pistettä, kun pisteiden määrä kyljellä  $\cdot$  tunnetaan; kuvien vasen ja yläreuna ovat täysin umpimähkäisiä. Kuvassa 2-a parametrin arvo on 0, muissa kuvissa sitä on suurennettu niin, että kuvien entropia kasvaa neljänneksen askelin 0:sta  $3/4$ :aan (yksikkönä täysin umpimähkäisen kuvan entropia).

Ehkä nämäkin esimerkit riittävät kertomaan sen, että morfologisten muunnosten käyttö kuva-analyysissä vaatii kekseliäisyyttä, joka ei aina maksa vaivaa. Parhaimmillaan se johtaa helposti automatisoitaviin algoritmeihin, mutta monissa tilanteissa rakenteellisen alkion jäykkyys johtaa menetyksiin, joita muunnosten kautta saavutettu informaatio ei riitä korvaamaan.

Mielenkiintoisen tehtäväkentän muodostaa Fontainebleaun koulun kehittämän morfologian kytkeminen kuvioiden Markov-malleihin. Voitaisiinko kuvioiden lokaalien vuorovaikutusten tuntemusta käyttää hyväksi morfologisten tunnuslukujen estimoinnissa ranskalaiseen tapaan?

Lähdeluettelo

Matheron, G. (1965): Les variables régionalisées et leur estimation. Masson, Paris.

Matheron, G. (1975): Random sets and integral geometry. Wiley, New York etc.

Ripley, B.D. (1976): Locally finite random sets: foundations for point process theory. Ann. Probability 4, 983-994.

Serra, J. (1982): Image analysis and mathematical morphology. Academic Press, London.

Stoyan, D. und J. Mecke (1982): Stochastische Geometrie. Akademie-Verlag, Berlin.

Suomela, P. (1979): Markov-aiheet ja -kuviot. Reports on computer science, economics and statistics 29/1979. University of Jyväskylä, Jyväskylä.



## TEKSTUURIN LUOKITTELU

Erkki Oja

Maaston luokittelussa satelliittikuvien perusteella käytetään nykyään lisääntyvässä määrin digitaalisen kuva-analyysin menetelmiä. Nykyisillä mittalaitteilla, etenkin uusimmilla monikanavakeilaimilla, saadaan nopeasti ja suhteellisen halvalla suuret määrät aineistoa, mutta tulosten hyödyntäminen on paljon hitaampaa ja kalliimpaa. Olisi tärkeää kehittää tietokonepohjaisia tulkintamenetelmiä, joiden kustannukset olisivat kohtuulliset.

Useat tulkintamenetelmät perustuvat hahmontunnistusteorian hyödyntämiseen. Tyypillisesti pyritään yksittäisiä kuvapisteitä luokittelemaan tai ryhmittelemään joidenkin tilastollisten kriteerien avulla. Pelkkien pisteiden tarkastelu ei kuitenkaan ole kuvatulkinnessa riittävää, vaan tarvitaan myös muita piirteitä, jotka tyypillisesti riippuvat useampien pisteiden muodostamista ympäristöistä. Erityisesti ilmakuvissa suuri osa informaatiosta sisältyy ns. tekstuuriin eli pintakuviointiin. Tekstuurien luokittelu on eräs yleisen kuva-analyysin keskeisiä probleemoja, jolla kaukokartoituksen lisäksi on sovelluksia teollisuusautomaatiossa ja lääketieteellisten kuvien käsittelyssä.

Huolimatta tekstuuripiirteiden ilmeisestä hyödyllisyydestä ilmakuvien analyysissa, ei muodollista ja hyväksyttävää määritelmää tekstuurille ole olemassa. Tyypillisesti tekstuurin ajatellaan koostuvan tietyistä alkeisosista ja niiden spatiaalisesta järjestyksestä, joka usein on karkeasti jaksollista. Tekstuurianalyysimenetelmät voidaan jakaa strukturaalisiin menetelmiin ja tilastollisiin menetelmiin.

Alan standardiviitteessä (Haralick, 1979) mainitaan kaikkiaan kahdeksan erilaista tilastollista lähestymistapaa tekstuurin luokitteluun. Nämä ovat autokorrelaatiofunktion käyttö, optiset muunnokset, digitaaliset muunnokset kuten DFT, kuvioinnin

reunaisuuden mitat, ns. strukturaalisen alkion käyttö, harmaatason yhteisjakauhistogrammit, harmaatason juoksupituudet ja kaksiulotteiset autoregressiiviset mallit. Osa menetelmistä on siis lähellä aikasarja-analyysiä, osa on varsin heuristisia ja lähinnä kokeellisesti testattavia.

Eräs lupaava uusi menetelmä tekstuurin luokitteluun on soveltaa Fourier-amplitudispektreihin ns. aliavaruusluokittelijaa (Oja, 1983). Tällä luokittelijalla on tehty joukko kokeita käyttäen standarditekstuureja (Parkkinen ja Oja, 1985). Vertailumenetelmänä on ollut osa Haralickin tekstuuripiirteistä ja joukko standardiluokittelijoita (ks. esim. Rosenfeld ja Kak, 1982). Jatkossa samaa menetelmää on tarkoitus soveltaa käytännön satelliittikuviin niissä olevien tekstuuriin luokitteluun ja ryhmittelyyn.

#### Lähdeluettelo

Haralick, R.M. (1979): Statistical and structural approaches to texture. Proc. IEEE 67(5), 786-804.

Oja, E. (1983): "Subspace methods of pattern recognition". Research Studies Press, Letchworth, ja John Wiley, New York.

Parkkinen, J. ja Oja, E. (1985): Texture classification by the subspace method. Ilmestyy konferenssikirjassa Proc. 4th Scand. Conf. on Image Analysis, Trondheim, June 18-20, 1985.

Rosenfeld, A. ja Kak, A.C. (1982): "Digital picture processing". Academic Press, New York.

## SATELLIITTIKUVIEN TULKINTAMENETELMIÄ

Erkki Tomppo

### 1. Johdanto

Satelliittikuvien käyttö metsäntutkimuksessa alkoi Suomessa 1960-luvun lopulla Metsäntutkimuslaitoksessa. Ensimmäiset kuvat olivat erotuskyvyltään varsin vaatimattomia, kuvan-alkion sivun pituus (eli n.s. spatiaalinen erotuskyky) vastasi maastossa 5 km:iä. Kuitenkin näiden kuvien antamat tulokset puuston kokonaistilavuuden estimoinnissa olivat jo lupaavia. (Korrelaatio maastossa mitatun puuston tilavuuden ja kuvan tummuusarvon välillä oli 0,84.) 1970-luvun puolesta välistä alkaen metsätaloudessa on käytetty yksinomaan numeerista kuvatulointia. Keskeinen tutkimuksen kohde on edelleen ollut puuston kokonaistilavuus ja tärkein kuvamateriaali vuoden 1983 loppuun saakka Landsat MSS-kuvat, joiden spatiaalinen erotuskyky on n. 80 metriä.

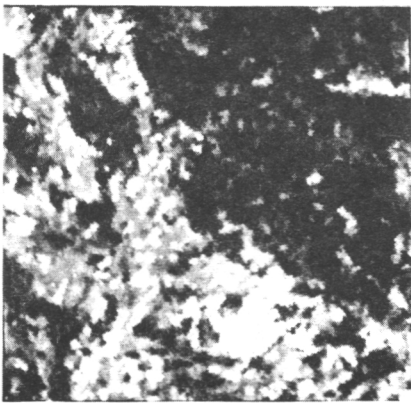
Tilavuuden arvioinnissa tulokset ovat suurehkolla alueella olleet tyydyttäviä: alueen ollessa vähintään 1000 ha ero maastossa mitattuun tilavuuteen on jäänyt alle 10 %:iin. Lisäksi joissakin erikoisprobleemoissa on päästy käytännössä toimiviin ratkaisuihin.

Tähän asti saatavilla olevien satelliittikuvien spatiaalinen erotuskyky on riittänyt vain joihinkin metsälökohtaisiin sovellutuksiin. Vuodesta 1984 alkaen tilanne on olennaisesti parantunut, kun Landsat TM-kuvien sekä spatiaalinen että spektrinen erotuskyky ovat olennaisesti parempia kuin aikaisempien kuvien (kuvaelementin koko maastossa 30x30 m<sup>2</sup>).

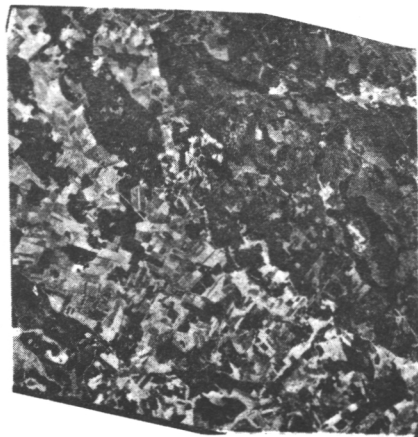
Uusien kuvien täysipainoinen hyödyntäminen edellyttää myös kuvan tulkintamenetelmien kehittymistä. On jopa mahdollista, että voimakkaasti vaihtelevilla alueilla erotuskyvyn lisääntyminen lisää kuvan sekavuutta ja johtaa huonompaan tulkintatulokseen kuin pienemmän erotuskyvyn kuvat, jos käytetään perinteisiä menetelmiä, (vrt. kuvat 1a ja 1b).

Tulkintamenetelmiä kehittämällä on nykyiset kuvat mahdollista saada palvelemaan esimerkiksi tilakohtaista metsätalouden suunnittelua sekä valtakunnan metsien inventointia.

Seuraavassa esitetään yhteenveto satelliittikuvien tulkinnaissa käytetyistä menetelmistä (jotka edelleenkin tulevat olemaan osa tulkintaa) sekä menetelmistä, joita mahdollisesti jatkossa tullaan käyttämään.



a



b

Kuva 1. Oikaistu Landsat MSS-kuva (erotuskyky 80 m) (a), simuloitu Spot-kuva (erotuskyky 10 m) (b) etelä-suomalaiselta metsä- ja viljelysalueelta, mittakaava 1:150000 (lähde Parmes (1984)).

## 2. Satelliittikuvan matemaattinen malli

Kaukokartoituksessa käytettävät ilmaisimet mittaavat kohteen heijastamaa tai emittoitamaa sähkömagneettista säteilyä. Säteily mitataan painotettuna keskiarvona tietyn kokoisilta maastonosilta eli maastonalkioilta, joita kuvalla vastaavat kuvanalkiot eli pikselit. Koska säteily mitataan samanaikaisesti useilla spektrin alueilla, kuvanalkio muodostuu useasta säteilyarvotiedosta (jotka normaalisti skaalataan välille 0-255). Seuraavassa kutsutaan näitä lukuja **sävyarvoiksi**.

Merkitään kuvanalkion (maantieteellisen) keskipisteen koordinaatteja  $x = (x_1, x_2)$ . (Samaistamme seuraavassa merkinnällisesti maastonalkion ja kuvanalkion, joita molempia merkitsemme keskipisteellä  $x$ .) Jokaiseen  $x$ :ään liittyy siis  $m$ -ulotteinen vektori

$$(1) \quad Z(x) = (Z^1(x), Z^2(x), \dots, Z^m(x)), \quad x \in X,$$

missä  $Z^i(x)$  on maastonalkion emittoima tai heijastama mitattu energian määrä aallonpituuskanavalla  $i$ .

Tämän mittaustiedon matemaattisena mallina voidaan käyttää tasossa määritellyn stokastisen prosessin integraalia

$$(2) \quad Z(x) = \int_{\eta} (S(x+t) + \varepsilon_1(x+t)) dt,$$

missä

$S(x)$  =  $m$ -ulotteinen tason stokastinen prosessi, n.s. signaali maantieteellisessä pisteessä  $x$  (eli kohteen lähettämä todellinen säteilyn määrä)

$\varepsilon_1(x)$  = m-ulotteinen tason stokastinen prosessi. n.s. kohina pisteessä  $x$ , joka kuvaa mm. ilmakehän ja mittauslaitteiden aiheuttamaa virhettä.  
 $dt$  = tason Lebesquen mitan (pinta-alan) differentiaali  
 $\eta$  = origokeskinen yksikköneliö, sivu maastonalkion mukaan suunnattu

(Ks. Switzer 1983)

Kuvatulkinnassa  $x$ :ään liitetään muuttuja  $I(x)$  (esim. metsätyyppi, puuston tilavuusluokka), joka voi saada äärellisen määrän arvoja  $1, \dots, K$ . Samoin signaalin  $S$  arvot luokitellaan ja oletetaan että  $S$  on funktio  $\mu$   $I$ :stä.

$$(3) \quad S(x) = \mu(I(x))$$

Tulkinnassa oletetaan usein vielä, että  $I(x)$  on vakio yhden kuvanalkion alueella, jolloin (1) saa muodon

$$(4) \quad Z(x) = \mu(I(x)) + \int_{\eta_1} \varepsilon_1(x+t) dt = \mu(I(x)) + \underset{\text{merk}}{\varepsilon(x)}$$

Tulkinnan tavoite on päätellä  $I(x)$  vektorin  $Z(x)$  avulla kullekin kuvanalkiolle. Perinteisissä tulkintamenetelmissä kukin kuvanalkio on liitetty johonkin luokkaa  $k$  oman sävyarvonsa perusteella. Kuvan muita alkioita ei ole käytetty hyväksi. Menetelmät olettavat siis, että  $\varepsilon(x)$ :t ovat toisistaan riippumattomia. Lisäksi yleensä on oletettu, että  $\varepsilon$  on multinormaalijakautunut. Muuttujaa  $I(x)$  on pidetty joko kiinteänä funktiona tai multinomijakautuneena satunnaismuuttujana, missä tapauksessa  $I(x)$ :iä on pidetty toisistaan riippumattomina.

### 3. Yhden kuvanalkion sävyarvoon perustuvat luokitusmenetelmät

#### 3.1 Ohjattu tulkinta

Ohjatussa tulkinnassa parin  $(Z(x), I(x))$  arvot tunnetaan joillakin  $X$ :n arvoilla (n.s. tukialueelta). Muuttujan  $I$  luokat muodostetaan yleensä näiden kuvanalkioiden perusteella. Luokitus sääntö perustuu tukialueen sävyarvo- ja maastotietoihin.

a) Lähimmän luokkakeskuksen menetelmä. Yksinkertaisimmassa luokitusmenetelmässä lasketaan tukialueelta kullekin  $I$ :n arvolla  $i$  siihen liittyvien sävyarvovektoreiden keskiarvo  $\bar{Z}_i$ . Luokitettava kuvanalkio  $x$  liitetään luokkaan, jonka sävyarvojen keskiarvoa lähinnä  $x$ :n sävyarvo on eli

$$(5) \quad I(x) = i \iff \|Z(x) - \bar{Z}_i\| \leq \|Z(x) - \bar{Z}_j\|, \quad j \neq i.$$

b) Laatikkoluokituksessa (parallelepiped classification) kuhunkin luokkaan liitetään kanavittaiset sävyarvojen ylä- ja alarajat. Kaikki kuvanalkiot, joiden sävyarvot mahtuvat näiden rajojen väliin liitetään k.o. luokkaan. Jos kuvanalkion sävyarvo ei mahdu minkään luokan rajojen sisälle, se jätetään luokittamatta. Luokkien päällekkäin menon estämiseksi luokkarajoja voidaan porrastaa.

c) Suurimman uskottavuuden (maximum likelihood) luokitusmenetelmässä oletetaan, että kunkin luokan  $i$ ,  $i = 1, \dots, K$  sävyarvovektorilla on joku todennäköisyysjakauma  $p(Z|i)$ . Jakauma voidaan periaatteessa

estimoida tukialueiden avulla. Kuvanalkio liitetään siihen luokkaan, johon se uskottavammin kuuluu eli

$$(6) \quad I(x) = i \Leftrightarrow p(Z(x)|i) \geq p(Z(x)|j), \quad j \neq i.$$

Käytössä olevat luokitusalgoritmit olettavat yleensä  $p(Z|i)$ :n multinormaaliseksi, jolloin puhutaan myös Gaussin suurimman uskottavuuden menetelmästä. Kunkin luokan  $i$  jakauman estimoimiseksi riittää estimoida tällöin luokan keskiarvovektori  $\mu_i$  ja kanavien sävyarvojen kovarianssimatriisi  $\text{cov}(Z_i)$ .

d) Bayesiläisessä suurimman uskottavuuden menetelmässä kullekin luokalle  $i$  annetaan a priori todennäköisyydet  $\pi_i$ . Muuttujaa  $I(x)$  piidetään siis multinomijakautuneena satunnaismuuttajana eli

$$(7) \quad P(I(x)=i) = \pi_i, \quad i=1, \dots, K, \quad x \in X.$$

Todennäköisyydet  $\pi_i$  on pääteltävä jonkin ennakkoinformaation avulla. Luokitussääntö on

$$(8) \quad I(x) = i \Leftrightarrow p(Z(x)|i) \pi_i \geq p(Z(x)|j) \pi_j, \quad j \neq i.$$

Ehdollisen todennäköisyyden määritelmästä seuraa, että luokitussääntö (8) on yhtäpitävä säännön (8') kanssa

$$(8') \quad I(x) = i \Leftrightarrow p(i|Z(x)) \geq p(j|Z(x)), \quad j \neq i.$$

Käytännössä tässäkin oletetaan, kuten edellä, että  $p(Z|i)$  on multinormaalisti jakautunut. Jos valitaan



kaikkien luokkien a'priori todennäköisyydet samoiksi, saadaan suurimman uskottavuuden luokitusmenetelmä.

e) Luokitus päätöksessä voidaan edelleen ottaa huomioon mahdollisista virheluokituksista aiheutuvat tappiot

$$(9) \quad \lambda(i|j), \quad i, j = 1, \dots, K.$$

Tässä  $\lambda(i|j)$  on tappio, joka aiheutuu, kun todellisuudessa luokkaan  $j$  kuuluva havainto luokitetaan luokkaan  $i$ . Bayes-optimaalisessa luokituksessa minimoidaan keskimääräinen tappio. Kuvanalkio  $x$  liitetään luokkaan  $i$ , jolle

$$(10) \quad L_{Z(x)}(i) = \sum_{j=1}^K \lambda(i|j) p(j|Z(x)) \quad \text{on}$$

pienin. Yhtälö (10) voidaan kirjoittaa myös

$$(10') \quad L_{Z(x)}(i) = \sum_{j=1}^K \lambda(i|j) p(Z(x)|j) \frac{\pi_j}{p(Z(x))}.$$

Valitsemalla tappiofunktiksi

$$(11) \quad \lambda(i|j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}$$

saadaan Bayesiläinen suurimman uskottavuuden luokitusmenetelmä.

f) Lineaarinen diskriminanttianalyysi (käytetään seuraavassa lyhennystä LDA) on paljon käytetty menetelmä kuvatulkinnessa. Tukialueiden tietojen avulla lasketaan luokkien väliset erottelufunktiot  $Y = \sum a_i z^i$ , jotka jakavat sävyarvoavaruuden hypertasojen avulla pistevieraisiin sävyarvoluokkiin.

Luokitussääntö on yhtäpitävää sen kanssa, että säteilyarvon  $Z(x)$  omaava pikseli liitetään luokkaan  $i$ , jonka keskukseen  $Z(x)$ :stä laskettu Mahalanobis-etäisyys on pienin eli

$$(12) \quad I(x) = i \Leftrightarrow D^2(Z(x), \bar{Z}_i) \leq D^2(Z(x), \bar{Z}_j), \quad j \neq i,$$

missä  $D^2(Z, \mu) = (Z - \mu)' S^{-1}(Z - \mu)$  ja  $S$  on sävyarvojen (pooled) kovarianssimatriisi.

Lineaarinen diskriminanttianalyysi on likimain Bayes-optimaalinen, kun  $\varepsilon(x)$  on multinormaalinen ja sama kaikilla  $x$  sekä a priori todennäköisyydet ovat samat kaikilla luokilla ja maksimoidaan oikein luokiteltujen kuvanalkioiden kokonaismäärä (ks. esim. Switzer 1980). Myös LDA:ssa voidaan antaa luokille erisuuret a priori todennäköisyydet  $\pi_i$ . Luokitussääntö on tällöin

$$(13) \quad I(x) = i \Leftrightarrow$$

$$D^2(Z(x), \bar{Z}_i) + 2 |\ln \pi_i| \leq D^2(Z(x), \bar{Z}_j) + 2 |\ln \pi_j|,$$

$$j \neq i.$$

### 3.2 Piirteiden valinta

Satelliittikuvan eri kanavien sävyarvot  $Z_i$  ovat yleensä korreloituneita, joten jättämällä osa kanavista pois tulkinnessa informaation menetys on pieni. Sen sijaan havaintoaineiston määrä voi pienentyä olennaisesti. Lisäksi käytämällä alkuperäisten kanavien sijasta niiden funktioita sävyarvoluokat saattavat erottua paremmin toisistaan. Piirteiden valinnan tarkoitus on selvittää mitä kanavia tai kanavien funktioita tulkinnessa käytetään. Paljon käytetty menetelmä on pääkomponenttianalyysi (eli Karhunen-Loève kehittämä). Tässä kanavien sävyarvoavaruudesta valitaan aliavaruus, joka sisältää suurimman osan sävyarvovaihtelusta.

Toinen mahdollisuus on käyttää erilaisia luokkien välisiä etäisyysmittoja piirteiden valinnassa. Piirteet valitaan siten, että luokkien väliset etäisyydet tulevat mahdollisimman suuriksi. Eniten käytetty lienee ns. Bhattacharyya-etäisyys eli B-etäisyys, joko minimoi virheluokitusten lukumäärän odotusarvon (ks. Fukunaga 1972).

Piirteinä voidaan käyttää lisäksi luvussa 4 lueteltavia tekstuurimittoja.

### 3.3 Ohjaamaton kuvatulkinna

Ohjaamattomassa tulkinnessa ei sävyarvoluokkia muodosteta maastotukialueiden perusteella. Tulkinna ensimmäisessä vaiheessa valitaan piirrevektorit tutkittavalle alueelle. Nämä ryhmitellään spektrisesti homogeenisiin luokkiin ryhmittelyanalyysia käyttäen. Kun luokat on muodostettu, niiden sisältö tulkitaan joko ennakkotietoa tai maastotukialueita käyttäen. Menetelmää käytetään usein ohjatun tulkinna ohella, esim. tukialueiden valinnassa.

#### 4. Tulkintamenetelmät, joissa käytetään hyväksi kuvanalkion ympäristöä

Luvussa 3 esitetyt menetelmät käyttävät luokituksessa vain osittain hyväksi kovalta saatavaa informaatiota, nimittäin yhden kuvanalkion sävyarvoa kerrallaan. Menetelmät eivät ota huomioon kohteen spatiaalista jatkuvuutta, t.s. sävyarvot luokitellaan ikään kuin ne olisivat toisistaan riippumattomia. Menetelmät saattavat antaa kohtalaisen tuloksen spektraalisesti selvästi erottuvilla alueilla, mutta voimakkaasti vaihtelevilla alueilla ja kuvioiden reunoilla virheluokitusten osuus on suuri. Samoin viivamaiset (tiet y.m.) ja pienet kohteet tulevat huonosti tulkituiksi.

Luokitustulosta voidaan parantaa tarkastelemalla yhden kuvanalkion sävyarvon sijasta samanaikaisesti useamman kuvanalkion sävyarvoa ja sävyjen sijaintia toisiinsa nähden. (Visuaalinen kuvatulkinta toimii tällä tavalla.) Myös kuvien parantunut spatiaalinen erotuskyky lisää tarvetta käyttää kuvanalkion ympäristöä tulkinnassa (vrt. kuva la ja lb).

Kun kuvanalkio luokitellaan ympäristöään käyttäen puhutaan sävyn lisäksi käsitteistä primitiivi eli tekstuurinalkio ja tekstuuri (Haralick et al. 1973, Haralick 1978 ja Laporte 1983). Vaikka näitä termejä on jo melko pitkään käytetty, täsmällistä määritelmää niille ei vielä ole esitetty. Näitä voidaan kuitenkin luonnehtia seuraavasti.

- 1) Sävy kuvaa kohteen emittoimaa tai heijastamaa energian määrää. Sävyn antama informaatio kohteesta riippuu aina kohteen oman sävyn lisäksi ympäristön sävystä (Irons & Petersen 1981).

- 2) Primitiivi eli tekstuurinalkio on yhden tai useamman kuvanalkion muodostama yhtenäinen kuvan osajoukko, jota karakterisoivat alkion dimensiot sävy, koko ja muoto (Laporte 1983).

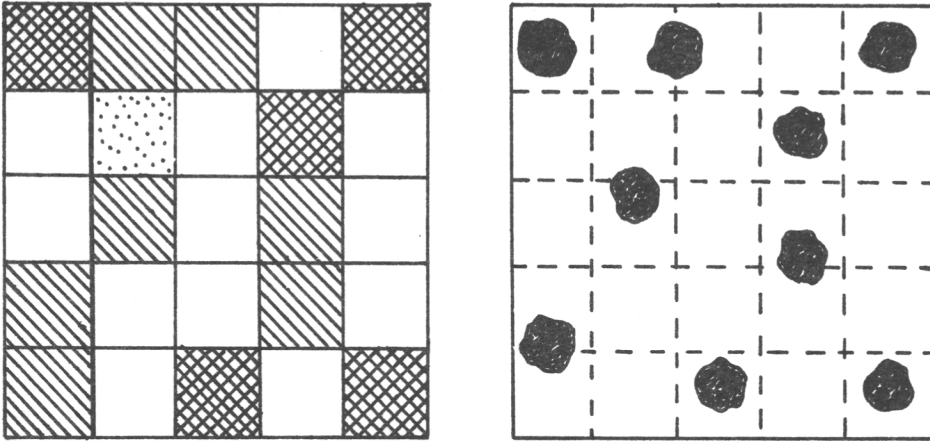
Esimerkkeinä primitiiveistä voidaan mainita

- a) kuvanalkio ja sen sävy
- b) samansävyisten kuvanalkioiden yhtenäinen joukko
- c) jonkin lokaalin operaation (esim. hajonnan) arvojen määräämä kuvanalkioiden joukko

- 3) Tekstuuri voidaan luonnehtia tekstuurinalkioiden ominaisuuksien ja alkioiden spatiaalisen riippuvuuden eli vuorovaikutuksen avulla (siis kuinka primitiivit sijaitsevat toisiinsa nähden).

Alkioiden sijainti toisiinsa nähden voi olla satunnaista tai siinä voi esiintyä pareittaista tai useamman kertaluvun vuorovaikutusta. Vuorovaikutus voi olla rakenteellista (structural) eli alkioiden spatiaalinen järjestäytyminen on lähes säännöllisesti kertaautuvaa tai se voi olla stokastista.

Esimerkiksi oletetaan, että tulkittava kohde kuvaa siemenpuuasennossa olevaa männikköä. Jos kuvanalkion koko on  $80 \times 80 \text{ m}^2$ , jokaista alkiota vastaava maastoelementti sisältää likimain yhtä monta puuta ja yhtä paljon puutonta aluetta, joten kuvanalkioiden sävyt ovat likimain samat. Jos taas kuvanalkion koko on  $10 \times 10 \text{ m}^2$ , osa maastoelementeistä on puuttomia ja osa sisältää yhden puun tai osan puusta (kuva 2). Seurauksena on, että tämän metsikkökuvion luokitus on tehtävä tekstuurin eikä sävyn perusteella.



Kuva 2. Esimerkki sävyjen vaihtelusta kuvalla, joka esittää siemenpuuasennossa olevaa männikköä.

Esimerkkinä kohteesta, jossa sävyjen järjestäytyminen on lähes säännöllisesti kertautuvaa ovat kuvioiden rajat ja viivamaiset kohteet.

Tekstuurin analysoimiseen on käytetty erilaisia tunnuslukuja eli tekstuurimittoja ja jonkin verran koko kuvan sävyarvojen mallittamista tekstuurimalleilla. Tekstuurimittoja ja mallien parametreja voidaan käyttää joko sellaisenaan luokitteluun (esim. lisäpiirteinä kanavien ohella tai sijalla) tai sitten niiden avulla kuva voidaan segmentoida tekstuuriltaan homogeenisiin osiin.

Tekstuurianalyysi on perinteisesti jaettu tilastolliseen ja rakenteelliseen analyysiin (Haralick 1978, Queil 1984). Ensimmäisessä käytetään erilaisia tilastollisia tunnuslukuja ja malleja tekstuurin luokittamiseen. Rakenteellinen

lähestymistapa mallittaa primitiivien geometriaa ja järjestäytymistä. (Puhdas rakenteellinen lähestymistapa perustuu näkemykseen, että primitiivien järjestäytyminen on lähes jaksollisesti toistuvaa (Haralick 1978).) Mauer ja Schärf (1982) luonnehtivat tilastollisen ja rakenteellisen analyysin eroja seuraavasti. Tilastollisella tekstuurianalyysillä

- havaitaan (multispektraalianalyysia paremmin) kohteet, joiden piirteet ovat teksturaalisia
- myös spektrisesti homogeeniset kohteet havaitaan
- huono erotuskyky kohteen reunoilla
- viivamaisia ja pistemäisiä kohteita ei havaita

Rakenteellisella tekstuurianalyysillä

- havaitaan myös viivamaiset ja pistemäiset kohteet
- saavutetaan kohteen reunoilla myös hyvä erotuskyky
- on vaikea havaita spatiaalista vuorovaikutusta, joka on luonteeltaan stokastista.

Ahuja ja Schachterin (1983) mielestä tekstuurianalyysin perustana pitäisi aina olla tekstuurimalli. Heidän mielestään parempi kuin jako tilastolliseen ja rakenteelliseen analyysiin olisi jako kuvanalkioon perustuviin ja alueisiin perustuviin malleihin. Kuitenkin monet tekstuurimitat (esim. co-occurrence matriisi), jotka eivät perustu mihinkään malliin toimivat käytännössä hyvin.

Seuraavassa tekstuurianalyysi jaetaan ei-malleihin perustuvaan ja malleihin perustuvaan analyysiin. Edellisessä käytetään n.s. informaaleja tekstuurimittoja.

#### 4.1 Informaalit tekstuurimitat

Haralick (1978) esittää yhteenvedon tekstuurin luokitteluun käytetyistä mitoista, jotka hän jakaa tilastollisiin ja rakenteellis-tilastollisiin mittoihin.

Tärkeimmät näistä voidaan jakaa seruaaviin ryhmiin.

- 1) Spatiaalinen frekvenssianalyysi, joka sisältää
  - 1a) sävyjen autokorrelaation (mittaa sävyjen jaksollisuutta ja riippuvuutta)
  - 1b) digitaaliset transformaatiomenetelmät
- 2) Reunaviivatiheyttä mittaavat tunnusluvut, mihin kuuluu
  - 2a) erilaiset gradienttimenetelmät
  - 2b) eräitä suodatusmenetelmiä

Esimerkkinä a)-kohdasta on kokonaisvaihtelu, joka kuvalla

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \text{ on}$$

$$(14) \quad TV = |a-b| + |b-c| + |d-e| + |e-f| + |g-h| + |h-i| \\ + |a-d| + |d-g| + |b-e| + |e-h| + |c-f| + |f-i| .$$

- 3) Sävyjen yhteisesiintymiseen perustuvat mitat, esim.
  - 3a) co-occurrence todennäköisyyksistä johdetut tunnusluvut (esim. co-occurrence-matriisiin perustuvat tunnusluvut)



- 3b) sävyn "pituuteen" (run length) perustuvat tunnusluvut.

Run length primitiivi on maksimaalinen saman sävyn muodostama kuvanalkioiden joukko, jota karakterisoi sen sävy, pituus ja suunta.

- 4) lokaaleihin ääriarvoihin perustuvat tunnusluvut

Lisäksi voidaan mainita

- 5) Hajontamittoihin perustuvat tunnusluvut esimerkkinä variogrammi, joka mittaa kuvanalkioiden sävyjen hajontaa alkioiden etäisyyden  $d$  ja suunnan  $\theta$  funktiona eli

$$(15) \quad v(d, \theta) = E (Z(x) - Z(y))^2 ,$$

missä  $Z(x)$  ja  $Z(y)$  ovat kuvanalkioiden  $x$  ja  $y$  sävyt.

- 6) Kuvanalkiota ympäröivien kuvanalkioiden sävyjen järjestetty joukko. Joissakin tapauksissa tällainen piirrevektori antaa vähintään yhtä hyvän luokitustuloksen kuin muut lokaalit operaatiot (ks. Ahuja et. al. 1979).

Tekstuurimittojen yksityiskohtaisemmasta esittelystä ks. Oja (1985).

#### 4.2 Mallien käyttö kuvatulkinnassa

Stokastisia malleja voidaan käyttää tekstuurin analyysissä ainakin kahdella tavalla.

1) Käytetään mallin antamia sävyarvoja, 2) käytetään mallin parametrejä lisäpiirteinä luokituksessa. Riippuen siitä onko primitiivinä kuvanalkio voi joku kuvanalkioiden joukko mallit jaetaan kuvanalkioon ja alueisiin perustuviin malleihin.

#### 4.2.1 Kuvanalkioon perustuvat mallit

Ahuja ja Schachter (1983) jakaa kuvanalkioon perustuvat mallit yksiulotteisiin aikasarjamalleihin ja satunnaiskenttämalleihin. Molemmat mallittavat kuvanalkion sävyä ja sävyjen sijaintia. Yksiulotteisessa aikasarjamallissa kuvanalkion sävy esitetään edeltävien alkioiden sävyjen avulla. Esimerkkinä ovat autoregressiivinen (AR) malli ja liukuvan keskiarvon (MA) malli tai yhdistetty (ARMA) malli.

Satunnaiskenttämalleissa kuvanalkion sävyarvo lausutaan ympäröivien alkioiden (lokaalit mallit) tai kuvan kaikkien muiden alkioiden avulla (globaalit mallit). Kuvan kaksikulotteisuudesta johtuen nämä soveltuvat aikasarjamalleja paremmin kuva-analyysiin.

Esimerkkinä globaalista mallista voidaan pitää ns. entistämisyyhtälöä

$$(16) \quad Z = BS + E,$$

missä  $Z$  on havaittu kuva (pinomuotoon järjestetty),  $S$  ideaalinen kuva, jota on sumennettu matriisilla  $B$  sekä  $E$  kohina vektori.

Esimerkkinä lokaalista mallista on kaksikulotteinen aikasarjamalli

$$(17) \quad z_{st} = \sum_{p=-M}^M \sum_{q=-N}^N b_{pq} z_{s-p, t-q} + e_{st}$$

missä  $b_{pq}$  ovat mallin kertoimia ja  $e_{st}$  virhetermi. (Tällaisten mallien estimointi voidaan nykyisin suorittaa valmiilla kirjasto-ohjelmilla (mm. SPIDER). Ohjelma etsii optimaaliset M:n ja N:n arvot, laskee kertoimet sekä estimoidun kuvan.)

Tarkastellaan seuraavassa muutamaa esimerkkiä mallin käytössä kuvatulkinnassa.

**Esimerkki 1** Mallin parametrin käyttö teksturaalisena piirteenä.

Tekstuurin epätasaisuus (roughness) voidaan määritellä kuvan reunaviivojen kokonaispituuden odotusarvona. Jos oletetaan, että kuva on Gaussin satunnaiskentän generoima, saadaan tälle tunnukselle lauseke  $\sqrt{1/2\pi\rho''(0)}$ , missä  $\rho''$  on prosessin korrelaatiofunktion toinen derivaatta (ks. Ahuja ja Schachter 1983).

**Esimerkki 2** Lokaalin mallin käyttö tulkinnassa. Switzer et al. (1981) ja Switzer (1983).

Tämä menetelmä on itseasiassa parannettu versio Bayesiläisestä suurimman uskottavuuden menetelmästä. Siinä spatiaalinen jatkuvuus otetaan huomioon antamalla a priori todennäköisyyksien  $\pi$  riippua kuvanalkion paikasta  $x$ .  $I(x)$ :ää pidetään siis spatiaalisena prosessina. Antamalla a priori todennäköisyyksien riippua  $x$ :stä voidaan kuvatulkinnassa ottaa huomioon myös mahdollinen lisäinformaatio kuten digitaalinen maaston korkeusmalli tai muu digitoitu maastotieto.

Todennäköisyyksien  $\pi(x)$  estimointi tapahtuu seuraavasti:

1) Suoritetaan ohjattu tulkinta kuvanalkioittain esimerkiksi suurimman uskottavuuden menetelmää tai LDA:ta käyttäen.

2) Lasketaan tukialueilta "sekoitusmatriisin"  $f_{KxK}$  estimaatti, missä

$f_{ij}$  = todennäköisyys, että luokkaan  $i$  kuuluva kuvanalkio liitetään luokkaan  $j$ ,  $\sum_j f_{ij} = 1$ .

3) Lasketaan jokaisen kuvanalkion kiinteän kokoisessa ympäristössä (esim.  $5 \times 5$ ) luokkaan  $i$  luokiteltujen alkioiden osuus  $p_i(x)$ . Merkitään

$$p(x) = (p_1(x), \dots, p_K(x)).$$

4) Jos oletetaan sekoitusmatriisi koko kuvan alueella vakioksi nämä lokaalisti estimoidut frekvenssit riippuvat a priori todennäköisyyksistä seuraavasti (seuraa kokonaistodennäköisyyden kaavasta)

$$(18) \quad E[p(x)] = \pi(x)E[\hat{f}].$$

Tästä saadaan  $\pi(x)$ :lle estimaattoriksi (mikäli  $f$  on ei-singulaarinen)

$$(19) \quad \hat{\pi}(x) = f^{-1}p(x).$$

5) Estimaattoria  $\hat{\pi}(x)$  käyttäen suoritetaan tämän jälkeen Bayesiläinen suurimman uskottavuuden luokitus.

Tällä menetelmällä Switzer (1983) on saanut virheluokitusten määrän pienenemään 39 %:sta 19 %:iin.

Suorittamalla vaiheet 2)-5) useaan kertaan lienee mahdollista edelleen pienentää virheluokitusten määrää.

### Esimerkki 3

Kolmantena esimerkkinä tarkastellaan Besagin (1983) esittämää spatiaalista diskriminanttianalyysiä, joka perustuu koko kuvan mallittamiseen. Oletetaan, että

- 1) kuvalla on määritelty symmetrinen naapuruusrelaatio, merkitään  $x \sim y$ . (Kuvanalkion naapureita voivat olla jälleen esimerkiksi kaikki sitä ympäröivän  $k \times k$ -neliön alkiot.) Merkitään

$$n_i = \#\{x | I(x) = i, x \in X\}$$

$$n_{ij} = \#\{(x, y) | x \sim y, I(x) = i, I(y) = j\}$$

(# tarkoittaa lukumäärää).

- 2)  $\underline{I}$ :llä on todennäköisyysjakauma, esimerkiksi

$$f_{\underline{I}}(i) = \frac{\exp\left(\sum_{i=1}^K \alpha_i n_i - \sum_{i,j=1}^K \beta_{ij} n_{ij}\right)}{C(\alpha, \beta)},$$

$$\alpha = (\alpha_i, i=1, \dots, K) \text{ ja } \beta = (\beta_{ij}, i, j = 1, \dots, K)$$

- 3)  $Z(x)$  ja  $Z(y)$  ovat ehdollisesti riippumattomia ehdolla, että  $\underline{I}$  tunnetaan
- 4)  $Z(x)$  on multinormaalisti jakautunut parametrein  $\mu, V$ .

Tavoitteena on ennustaa  $\underline{I}$   $Z$ :sta lähtien. Tämä tehdään maksimoimalla  $I$ :n todennäköisyys annetulla  $Z$ :n arvolla eli

$$(20) \quad \max_{\underline{I}} f_{\underline{I}|\underline{Z}}(\underline{i}|\underline{z}) = \max_{\underline{I}} \frac{f_{\underline{Z}|\underline{I}}(\underline{z}|\underline{i}) f_{\underline{I}}(\underline{i})}{f_{\underline{Z}}(\underline{z})} .$$

Tämä on yhtäpitävää sen kanssa, että maksimoidaan  $\underline{I}$ :n suhteen

$$(21) \quad f_{\underline{Z}|\underline{I}}(\underline{z}|\underline{i}) \cdot f_{\underline{I}}(\underline{i}) .$$

Käytännössä lausekkeen (21) maksimi löydetään seuraavasti:

- 1) Tukialueilta lasketaan  $\mu$ :n ja  $V$ :n estimaatit  $\hat{\mu}$ ,  $\hat{V}$ .
  - 2) Sovelletaan  $Z$ :aan tavallista LDA:ta,olloin saadaan estimaatit
- $$(22) \quad \hat{\underline{I}}^1 = (\hat{n}_i, i = 1, \dots, K; \hat{n}_{ij}, i, j = 1, \dots, K)$$
- ja  $\hat{\alpha}^1$  ja  $\hat{\beta}^1$ .
- 3) Määritellään kuvanalkion  $x$  naapurit (esim.  $X$ -keskinen  $k \times k$  ikkuna).
  - 4) Jokaisella  $x$  minimoidaan lauseke (21) tässä  $x$ :n ympäristössä käyttäen jo saatuja estimaatteja.
  - 5) Tällöin saadaan uudet estimaatit  $\hat{\underline{I}}^2$ ,  $\hat{\alpha}^2$  ja  $\hat{\beta}^2$ .

Vaiheet 4) ja 5) voidaan toistaa useita kertoja jolloin saadaan parempi estimaatti.

Keinotekoisella aineistolla Besag on tällä tavoin saanut pienennettyä LDA:n virheluokitusten määrää 40 %:sta 11 %:iin ja vastaavasti 10 %:sta 2,5 %:iin. Valitsemalla edellä  $\beta_{ij} = 0$  saadaan tavallinen LDA.

#### 4.2.2 Alueisiin perustuvat mallit

Alueisiin perustuvissa malleissa kuvan ajatellaan muodostuvan osahahmoista, jotka ovat sijoittuneet tasoon tietyn säännön mukaan. Mallin määrittelyssä primitiiveinä käytetään kuvanalkion asemasta alueita (usean kuvanalkion joukkoja). Samoin tulkinta tapahtuu alueittain. Mallien avulla voidaan kuvata primitiivien kokoa, muotoa, sijaintia ja sävyä.

Esimerkkeinä alueisiin perustuvista malleista ovat peite (coverage) mallit ja mosaiikkimallit.

Peitemallin realisaation voidaan ajatella syntyvän siten, että geometrisiä kuvioita sijoitetaan satunnaisesti tasoon ja annetaan kuvioille väri jonkun säännön mukaan. Ahuja ja Schachter (1983) esittävät näiden mallien parametreihin perustuvia tilastollisia tekstuuriimittoja.

Satunnaismosaiikkimalleissa taso jaetaan konvekseihin osajoukkoihin, mosaiikkeihin. Todennäköisyys, että kaksi pistettä kuuluvat samaan mosaiikkiin riippuu vain pisteiden etäisyydestä. Kullekin mosaiikille annetaan väri  $c_i$ ,  $i=1, \dots, m$  toisista riippumatta kiinteillä todennäköisyyksillä  $P_i$ ,  $i=1, \dots, m$ . Mosaiikit voivat olla esimerkiksi säännöllisiin kuvioihin (kolmio, suorakaide jne.) perustuvia. Poisson-viiva mosaiikki syntyy siten, että tasoon sijoitetaan satunnaisesti toisiaan leikkaavia suoria ja annetaan syntyneille konvekseille soluille väri em. ta-

valla. (Tällaisella mosaiikilla on mm. seuraava mielenkiintoinen ominaisuus. Kun tasoon sijoitetaan tasavälein pisteitä, pisteiden värien muodostama jono on m-tilainen Markovin ketju.)

Myös satunnaismosaiikkimallin muodostamalle tekstuurille voidaan johtaa tekstuuriominaisuuksien eksakteja lausekkeita. Näistä voidaan puolestaan johtaa tunnuksia joillekin tekstuuria kuvaaville termeille. Esimerkiksi korrelaatiofunktion ja variogrammin ensimmäisten derivaattojen avulla voidaan kuvata tekstuurin karkeutta (coarseness) ja laikkuisuutta (patchiness) (ks. Matern 1960 sekä Ahuja ja Schachter 1983). Karkeus mittaa sitä kuinka suurista osista tekstuuri on muodostunut. Laikkuisuus taas mittaa kuvioiden kokoa, tiheyttä ja kontrastia.

Oletetaan, että pisteitä sijoitetaan tasoon Poisson prosessin mukaan intensiteettinä  $T$  ja pisteiden kautta asetetaan suora, jonka suunta valitaan satunnaisesti. Tällöin Poisson-viiva mosaiikille karkeus on

$$(23) \quad -\rho'(0) = \frac{2T}{\pi} \text{ ja}$$

laikkuisuus on kahden värin tapauksessa (värien todennäköisyydet  $p_0$  ja  $p_1 = 1-p_0$ )

$$(24) \quad -\nu'(0) = \frac{T}{\pi} p_0 p_1.$$

Alueisiin perustuvia malleja voitaneen pitää sopivina juuri metsän spatiaalisen vaihtelun ja metsikkökuvioiden mallina. Näiden huonona puolena on kuitenkin se, että ne johtavat helposti melko raskaaseen matemaattiseen kalustoon.



Luonnossa esiintyvä vaihtelu on kuitenkin usein monimutkaista. Sen totuuden mukaisessa kuvaamisessa ei voitane välttää monimutkaisia malleja.

## 5. Yhteenveto

Edellä on lueteltu muutamia esimerkkejä menetelmistä, joita voidaan käyttää satelliittikuvien tulkinnassa. Luettelo ei ole kaiken kattava. Siitä puuttuu kokonaisia menetelmäryhmiä, esim. syntaktiset menetelmät ja kohteen yhteyteen perustuvat menetelmät. Samoin on rajoitettu vain kuvan tulkintaan, mukaan ei ole otettu kuvan esi- eikä jälkikäsittelemenetelmiä (segmentointimenetelmät mukaanlukien). Samoin tilastotieteellisesti mielenkiintoista virheluokitusten määrän arviointia ei ole tarkasteltu.

Yhteenvetona voidaan todeta, että käytettävä tulkintamenetelmä riippuu kuvan laadusta, kohteesta sekä siitä mitä tietoja kovalta halutaan. Edelleen eri menetelmiä ei ole pidettävä pelkästään toistensa kilpailijoina vaan myös täydentäjinä. Itse tulkinta voidaan tehdä hierarkisesti käyttäen eri vaiheessa eri menetelmiä. Tulkinta voidaan aloittaa esimerkiksi selvien kohteiden multispektraalitulkinalla ja edetä kohti vaikeampia kohteita tekstuurimitto- ja ja -malleja käyttäen.

Lähdeluettelo

Ahuja, N., Haralick, B.M. ja Rosenfeld, A. (1979): Neighbor grey level as features in pixel classification. Pattern recognition, Vol. 12, 251-260.

Ahuja, N. ja Schacter, B.J. (1983): Pattern Models, Wiley, New York.

Besag, J. (1983): Spatial Discriminant analysis. Bull. Int. Statist. Inst., Vol. 50.3, 422-425.

Fukunaga, K. (1972): Introduction to statistical Pattern Recognition, Academic Press, New York.

Haralick, R.M. (1978): Statistical and structural approaches to texture, Proc. 4th Int. Joint Conf. on Pattern Recognition, 45-69.

Haralick, B.M., Shanmugam, K. ja Dinstein, I. (1973): Textural features for image classification, IEEE Trans. Systems, Man, Cybernet. SMC-3, 610-621.

Irons, R.I. ja Petersen, G.W. (1981): Texture transforms of Remote Sensing Data. Remote sensing of environment 11, 359-370.

Laporte, J. (1983): Les directions de recherche dans l'analyse des textures. Käsikirjoitus 1983.

Matern, B. (1960): Spatial variationen. Meddelanden från statens skogsforskningsinstitut, Tukholma, 1-144.

Mauer, E. ja Schärf, R. (1982): Picture classification and segmentation by feature combination in multispectral data. Actes du Symposium International de la Commission VII de la Societe Internationale de Photogrammetrie et Teledetection, Toulouse, 1982, 35-43.

Oja, E. (1985): Tekstuuriluokittelu, Metsäntutkimuslaitoksen tiedonantoja. (Ilmestymässä.)

Parmes, E. (1984): Segmentation as an initial step in land use interpretation from Spot imagery. Käsikirjoitus.

Queil, R. (1984): Trends and developments in the classification of multispectral data. International Society for Photogrammetry and Remote Sensing XVth Congress, Rio de Janeiro, 1984, 440-451.

Swain, P.H. ja Davis, Sh.M. (1978): Remote Sensing: The Quantitative Approach, Mc Graw-Hill, New York.

Switzer, P. (1980): Extension of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. Mathematical Geology, 12, 367-376.

Switzer, P. Kowalick, W.S. ja Lyon, R.J.P. (1982): A Prior probability method for smoothing discriminant analysis classification maps. Mathematical Geology, 14, 433-444.

Switzer, P. (1983): Some spatial statistics for the interpretation of satellite data (with discussion). Bull. Int. Statist. Inst., Vol. 50.2, 962-972.

## SATELLIITTIKUVA-AINEISTON JA GEOFYSIKAALISEN AINEISTON NUMERISESTA LUOKITUKSESTA

Hilkka Arkimaa

### JOHDANTO

Satelliittikuvia on Suomessa käytetty jo kuusikymmenluvun lopulta lähtien kallioperän rakenteiden ja kivilajien kartoituksen apuna. Kallioperä ja sitä lähes kauttaaltaan peittävä irtomaakerros ovat kuitenkin vasta toissijaisia vaikuttajia maanpinnasta heijastuneen sähkömagneettisen säteilyn intensiteetin vaihtelussa. Ensisijaisena vaikuttajana on kasvillisuuspeite. Maa- ja kallioperän koostumus saattaa aiheuttaa kasvillisuuteen vaihtelua, joka antaa siten välillistä tietoa kasvualustan geologisista yksiköistä.

Numeerisessa muodossa olevasta säteilyaineistosta on tilastollisilla luokittelumenetelmillä mahdollista erotella geologisia yksiköitä luonnehtivia intensiteettitasoja ja tuottaa näin kivilajikartta

### OHJATTU LUOKITTELU SOKLIN ALUEELLA

Itä-Lapissa sijaitsevan Soklin karbonatiittimassiivin alueella eri kivilajiyksiköitä luonnehtii varsin pitkälle niille ominaiset kasviyhdykunnat. Niinpä siellä kokeiltiin kivilajien kartoittamista digitaalisesta Landsat-aineistosta ohjattua luokittelumenetelmää käyttäen.

### Alueen geologiasta

Soklin alueen geologia koostuu lähes pyöreästä piippumaisesta noin 5 kilometrin läpimittaisesta karbonatiittintrusiivista, tätä ympäröivästä feniittikehästä ja kaikkein uloimpana gneissigraniitista (Kuva 1).

Karbonaattiin sisältämä fosfori yhdessä kalsiitin kanssa muodostaa kasvillisuuteen rehevöittävästi vaikuttavan ravinnetekijän. Karbonaattiinmassiivin alue on lähes kauttaaltaan paksuhkojen irtomaa-ainesten peittämä. Topografia alueella on muodostunut kivilajiyksiköiden mukaisesti siten, että karbonaattiipiippu muodostaa allaslaakson ja feniitti tätä ympäröivän kukkulaisen renkaan.

#### Säteilyaineiston luokittelu

Landsat-aineiston soveltuvuutta kivilajien kartoittamiseen alueella kuvastaa massiivin halki kulkevalta profiililta piirretty säteilyn intensiteettijakauma eri kanavilla ja vastaava kallioperägeologinen profiili (kuva 2). Pitenevän aallonpituuden mukana geologisten yksiköiden kuvastuminen tyypillisinä säteilyn intensiteettitasoina lisääntyy. Numeerisessa muodossa oleva säteilyaineisto luokiteltiin maximumlikehood-periaatetta käyttävän ohjatun luokittelumenetelmän avulla. Ennen varsinaista luokittelua määritettiin luokkarajat maastossa tunnettujen pienialaisten ns. tukialueiden keskiarvovektoreiden ja kovarianssimatriisien avulla. Ohjattu luokittelu suoritettiin valittujen tukialueiden tilastolisten tunnuslukujen pohjalta koko tutkittavalle alueelle käyttäen neljän aallonpituuskaistan mittausaineistoja muutujina (kuva 3).

Laaja-alaiset geologiset yksiköt, kuten gneissigraniitti, feniitti ja karbonaatti kartoituivat tunnetuilta osin hyvin. Poikkeuksen muodosti karuinta metsätyyppiä karbonaattiin päällä edustava harjujakso, joka luokiteltiin ympäristön kivilajeihin. Karbonaattiinluokkien esiintyminen paikoin massiivin rajojen ulkopuolella selittyy sillä, että ravinteikkaasta kallioperästä moreeniin joutunut aines on kulkenut jokiuomia pitkin massiivilta ulos, jonne syntyy rehevän kasvillisuuden alueita.

## GEOFYSIKAALISESTA MITTAUSAINEISTOSTA

Satelliittikuvien käsittelyyn alunperin suunniteltuja numeerisia tulkintamenetelmiä on sovellettu myös geofysikaalisten mittausaineistojen käsittelyyn. Koska geofysikaalinen mittausaineisto antaa suurempaa tietoa itse kallioperästä, sen fysiikkalisista ominaisuuksista, tuo niiden käyttö muuttujina tulkintamenetelmissä paremmat mahdollisuudet kallioperän kivilajien ja rakenteiden sekä malmikriittisten muodostumien kar-toitukseen.

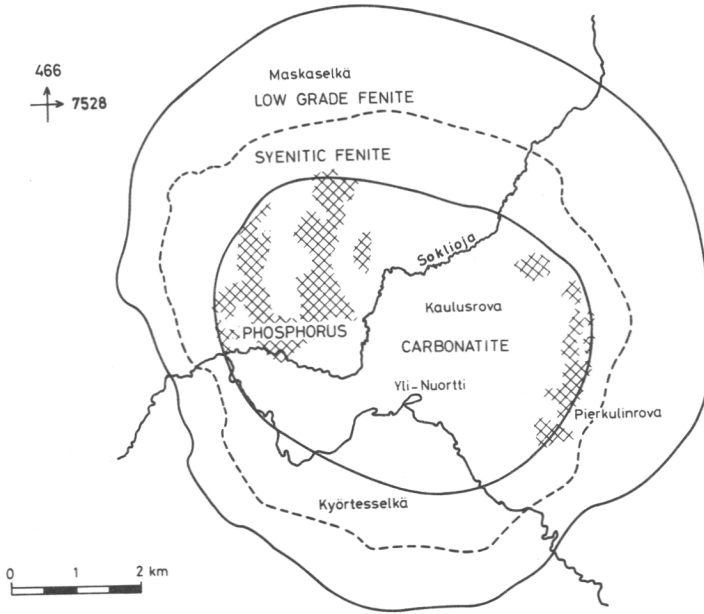
## KIRJALLISUUS

Aarnisalo, J. et al., 1982. On the integrated use of Landsat, geophysical and other data in exploration in the Baltic Shield, Finland. The Photogrammetric Journal of Finland 9, 1:48-64.

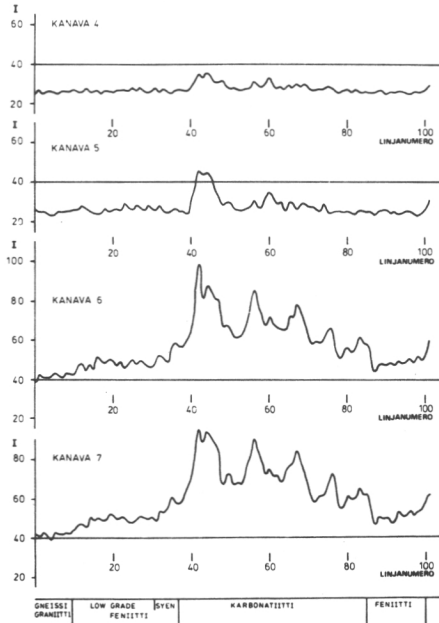
Arkimaa, H., 1982. Some examples of multichannel analyses of Landsat and geophysical data. The Photogrammetric Journal of Finland 9, 1:38-47.

Talvitie, J., 1979. Remote Sensing and geobotanical prospecting in Finland. Bull. Geol. Soc. Finland 51, 63-73.

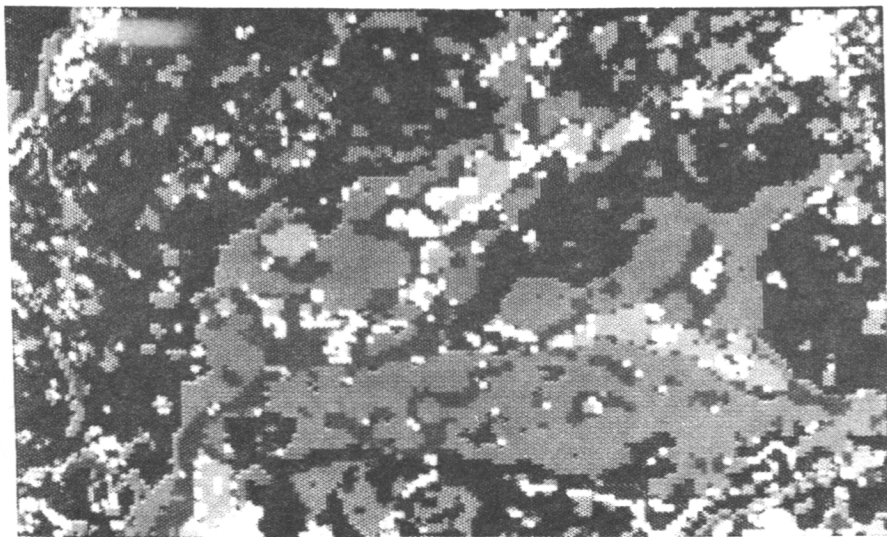
Suomi avaruudesta (toim. Mikko Punkari), 1984.  
Tähtitieteellinen yhdistys Ursa.



Kuva 1. Soklin alueen geologinen kartta.



Kuva 2. Intensiteetin jakauma eri aallonpituuskanavilla ja vastaava kallioperägeologinen profiili tutkimusalueen poikki.



Kuva 3. Landsat-aineiston ohjattu luokittelu.



## METSÄLUONNON MUUTTUMINEN JA SPATIAALINEN TILASTOTIEDE

Pertti Hari

Ihmiskunnan teollinen toiminta on kasvanut eksponentiaalisesti yli sadan vuoden ajan. Samaa tahtia on lisääntynyt myös ilmakehään päästettyjen aineiden määrä. Päästöjen takia ovat ilmakehän ominaisuudet alkaneet muuttua. Tämä vuorostaan generoi muutoksen virtaamiin ilmakehästä metsiin. Näin metsien kasvun ja terveyden kannalta olennaiset ympäristötekijät ovat alkaneet muuttua.

Maastosta esiin työntyvät kohteet suodattavat tehokkaasti ilmaa. Pinnan rakenne vaikuttaa lisäksi ilman virtaukseen, mikä lisää paikallista vaihtelua. Näin maaston muoto ja metsän rakenne vaikuttavat voimakkaasti kunkin paikan laskeuman määrään.

Maaperän kemialliset ominaisuudet, pinnan muoto sekä veden virtaus maassa aiheuttavat vaihtelua kunkin paikan laskeuman sietoon. Siis myös laskeuman vaikutuksissa on selviä paikallisia piirteitä.

Spatiaalisen tilastotieteen kannalta on olennaista, että metsien muuttumisen analysoimiseksi on käytettävissä hyvin monista lähteistä peräisin olevaa spatiaalista tietoa. Maaston muoto määritetään kartoittamisen yhteydessä. Puuston ominaisuudet voidaan määrittää joko inventoimalla tai kaukokartoitusmenetelmillä. Maaperän laskeuman kestokykyä voidaan arvioida geologisten ominaisuuksien perusteella. Metsien vaurioitumista voidaan seurata satelliittikuvien ja inventointien avulla.

Muuttuvan ympäristön vaikutuksilla on voimakkaita spatiaalisia piirteitä. Lisäksi on käytettävissä runsaasti spatiaalista tietoa muutoksen analysoimiseksi. Muuttuvan ympäristön metsävaikutusten analysoinnissa on spatiaalisen tilastotieteen suurin ja selvin tehtävä metsätieteissä.

SPATIAALISET MENETELMÄT, UUSI KEINO METSÄTALouden  
SUUNNITTELUUN

Simo Poso

Ihmisen toiminta ja suhde ympäristöön on niin moninainen, että se jää suurelta osalta ymmärtämis- ja hallintakykyimme ulkopuolelle. Kuitenkin esimerkiksi metsäammattimiehillä on vastuu ja myös tarve hallita toimintansa kohdealueita. Hallintaa varten asioita joudutaan yksinkertaistamaan mm. käsitteiden määrittelyin ja erilaisin luokituksin. Yksinkertaistukset merkitsevät kuitenkin, että samaan ryhmään joudutaan panemaan käsittely- tai suunnitteluyksiköitä, jotka saattavat poiketa toisistaan huomattavastikin. Hallintatekniikka taas määrää, että samaan ryhmään luokitettut yksiköt saavat samanlaisen kohtelun.

Metsien käytössä ja sen suunnittelussa on vallitsevana ollut tarkastelutapa, jota voidaan kutsua staattiseksi. Se on perustunut ajatukseen, että on olemassa jokin tietty menettelytapa, jota noudattaen metsätalous saadaan ikiajoiksi kuntoon. Selvimpiä esimerkkejä ovat hakkuulaskelmien kaava-menetelmät, joita kehitettiin 1800-luvulla ja joita on sovellettu sellaisinaan hakkuusuunnitteina. Staattiselle tarkastelutavalle on katsottu saatavan oikeutusta metsätalouden epätavallisesta pitkäjänteisyydestä.

Elinkeinoelämän monipuolistuessa staattisen tarkastelutavan arvo on vähentynyt dynaamisemman otteen tieltä. Metsätalouden harjoittajat ovat erilaisia, metsien omistajat vaihtuvat ja heidän tarpeensa ja tavoitteensa muuttuvat. Myös metsissä on luontaista vaihtelua, jota voidaan käsittelyin joko lisätä tai vähentää. Metsät tarjoavat erilaisia mahdolli-

suuksia taloudelliselle ja muulle toiminnalle.

Moninaisuuden ja dynaamisuuden tunnustaminen aikaisempaa tärkeämmiksi tekijöiksi metsän käytön suunnittelussa merkitsee, että suunnitelmien kehittelyä on tarkasteltava uudesta näkökulmasta. Tekniset mahdollisuudet tähän ovat lisääntyneet erityisesti atk:n ja satelliittikuvien kehittymisen johdosta. Mahdollisuuksien täysimääräinen hyväksikäyttö metsän inventoinnissa ja metsätalouden suunnittelussa vaatii uudenlaista tilastotieteellistä lähestymistapaa. Tälläisen tarjonnan spatiaaliseksi tilastotieteeksi kutsuttu tieteenala. Se voi käsitellä moniulotteista avaruutta, jossa tieto on kuvattavissa esimerkiksi paikan ja ajan suhteen. Tieto-osasten välisillä riippuvuuksilla on usein tärkeä merkitys. Esimerkkinä spatiaalisesta tarkastelusta on metsikön (puiden) tilajärjestyksen kuvaus matemaattis-tilastotieteellisesti siten, että metsikön ominaisuuksista saadaan riittävän yksityiskohtaiset tiedot erilaisiin tarpeisiin kuten inventoinnin, hakkuun ja kuljetuksen suunnitteluun.

Metsän tilajärjestystä kuvaavien dynaamisten mallien avulla voitaneen kehitellä menetelmiä, joilla haitallisia yksinkertaistuksia ja subjektiivisten luokitusten tarvetta on mahdollista vähentää. Samalla lisätään inventointimenetelmien joustavuutta ja tehokkuutta sekä saadaan kasvu- ja tuotomallit entistä yksityiskohtaisemmiksi ja toimenpiteiden vaikuttavuusanalyysiin paremmin soveltuviksi.

Hyvän mahdollisuuden kehittää ja soveltaa spatiaalisia menetelmiä tarjoavat satelliittikuvat. Satelliittitieto saadaan maastoerotuskykyä vastaavista ruuduista usealla aallonpituuskanavalla. Esimerkiksi Landsat 5-kuvasta saadaan 30 m x 30 m ruuduista säteilyarvot kuudella kanavalla.

Tämä merkitsee, että hehtaaria kohden on käytettävissä 67 apumuuttujaluoteista tietoa. Hyvien sääolosuhteiden vallitessa tällaista tietoa saadaan useita kertoja vuodessa. Spatiaalisten menetelmien keinoilla voidaan pyrkiä käyttämään hyväksi säteilyarvojen välisiä riippuvuuksia, jotka koskevat lähekkäisiä ruutuja ja eri kuvausajankohtia sekä sitomaan säteilyarvot lopulta maastossa kontrolloitaviksi tuloksiksi.

Voidaan olettaa, että spatiaalisin menetelmin päästään entistä yksityiskohtaisempiin metsää kuvaaviin malleihin, joilla on huomattava merkitys inventointien ja koko metsätalouden suunnittelussa. Kuitenkaan kuva siitä, millaisiksi menetelmät tulevat muodostumaan ja mihin ne lopulta soveltuvat, ei ole selvä. Näkymät ovat joka tapauksessa niin lupaavat, että mahdollisuuksien kartoittamista ja menetelmien kehittelyä voidaan pitää tärkeänä ja kiireellisenä tutkimustehtävänä. Selvää on, että tehtävästä suoriutuminen edellyttää metsä- ja tilastotieteiden edustajien yhteistyötä.

KIRJOITTAJAT

Arkimaa, Hilikka, FK, geofyysikko  
Geologian tutkimuskeskus, c/o VTT, Maankäytön laboratorio,  
Itätuulentie 2 02100 Espoo

Hari, Pertti, MMT, dosentti  
Suomen Akatemia  
c/o Helsingin yliopisto, Metsänhoitotieteen laitos  
Unioninkatu 40 B 00100 Helsinki

Kuulasmaa, Kari, Ph.D., biometrikko  
Kansanterveyslaitos, Epidemiologian osasto  
Hämeentie 155 00560 Helsinki

Lahtinen, Matti, FK, vs. assistentti  
Jyväskylän yliopisto, Tilastotieteen laitos  
Seminaarinkatu 15 40100 Jyväskylä

Niinimaa, Ahti, FL, vs. apul, prof.  
Oulun yliopisto, Sovelletun matematiikan ja tilastotieteen  
laitos  
Linnanmaa 90570 Oulu

Oja, Erkki, TkT, apul. prof.  
Kuopion yliopisto, Sovelletun matematiikan laitos  
PL 6 70211 Kuopio

Penttinen, Antti, FT, assistentti  
Jyväskylän yliopisto, Tilastotieteen laitos  
Seminaarinkatu 15 40100 Jyväskylä

Poso, Simo, MMT, vs. prof.  
Helsingin yliopisto, Metsänarvioimistieteen laitos  
Unioninkatu 40 B 00100 Helsinki 10

Ranneby, Bo, Ph.D., dosentti  
Swedish University of Agricultural Sciences, Section of  
Forest Biometry  
S-90183 Umeå, Sweden

Saarenmaa, Hannu, MML, tutkija  
Metsäntutkimuslaitos, Rovaniemin tutkimusasema  
Eteläranta 55 96300 Rovaniemi

Salminen, Sakari, MH, vanhempi tutkija  
Metsäntutkimuslaitos  
Unioninkatu 40 A 00170 Helsinki

Suomela, Pentti, FT, dosentti  
Jyväskylän yliopisto, Matematiikan laitos  
Seminaarinkatu 15 40100 Jyväskylä

Tomppo, Erkki, FL, tutkija  
VTT, Maankäytön laboratorio  
Itätuulentie 2 02100 Espoo

OSANOTTAJALUETTELO

N i m i

T y ö p a i k k a

Arkimaa, Hilikka	Geologian tutkimuskeskus
Arovaara, Hannu	HY, metsänhoitotieteen laitos
Ferm, Ari	Metsäntutkimuslaitos, Kannus
Hari, Pertti	HY, metsänhoitotieteen laitos
Heikkilä, Elina	HY, metsänhoitotieteen laitos
Heinonen, Jaakko	Metsäntutkimuslaitos, Joensuu
Hiltunen, Jaana	Metsäntutkimuslaitos, MSE
Hovi, Arto	Joensuun yliopisto
Holmberg, Maria	HY, metsänhoitotieteen laitos
Häkkinen, Risto	Metsäntutkimuslaitos, MAT
Häme, Tuomas	VTT, maankäytön laboratorio
Hänninen, Harri	Metsäntutkimuslaitos, EKK
Ihalainen, Antti	Metsäntutkimuslaitos, ARI
Ilvesniemi, Hannu	HY, metsänhoitotieteen laitos
Isomäki, Antti	Metsäntutkimuslaitos, ARP
Ikäheimo, Erkki	Metsäntutkimuslaitos, Vantaa
Korhonen, Kirsi-Marja	Metsäntutkimuslaitos, MSE
Koski, Timo	Mat. inst. vid Åbo Akademi
Kotisaari, Ahti	HY, metsänhoitotieteen laitos
Kulmala, Markku	HY, Fysiikan laitos
Kurimo, Heikki	SA, Joensuun yliopisto
Kuulasmaa, Kari	Kansanterveyslaitos
Lahti, Tapani	HY, kasvimuseo
Lahtinen, Matti	Jyväskylä yliopisto
Laininen, Pertti	TKK, Otaniemi
Lappi, Juha	Metsäntutkimuslaitos, Suonenjoki
Latja, Ari	Joensuun yliopisto
Linnilä, Kimmo	Metsäntutkimuslaitos, MHO
Luoma, Martti	Vaasan korkeakoulu
Luoma, Pirkko	Metsäntutkimuslaitos, ARP
Mauranen, Kari	Jyväskylän yliopisto
Niemi, Hannu	HY, tilastotieteen laitos
Niemistö, Pentti	Metsäntutkimuslaitos, ARP
Niinimaa, Ahti	Oulun yliopisto
Nöjd, Pekka	Metsäntutkimuslaitos, ARP
Oja, Erkki	Kuopion yliopisto



Ojansuu, Risto	Metsäntutkimuslaitos, ARP
Oker-Blom, Pauline	HY, metsänhoitotieteen laitos
Pekkonen, Timo	Metsäntutkimuslaitos, MAT
Penttinen, Antti	Jyväskylän yliopisto
Poso, Simo	HY, metsänarvioimistieteen laitos
Pukkala, Timo	Joensuun yliopisto
Ranneby, Bo	Sveriges lantbruksuniversitet
Repo, Seppo	Metsäntutkimuslaitos, EKK
Saarenmaa, Hannu	Metsäntutkimuslaitos, Rovaniemi
Saksa, Timo	Metsäntutkimuslaitos, Suonenjoki
Salmi, Veli-Pekka	Metsäntutkimuslaitos, MSE
Salminen, Sakari	Metsäntutkimuslaitos, ARI
Seppälä, Risto	Metsäntutkimuslaitos, MAT
Siitonen, Markku	Metsäntutkimuslaitos, ARI
Similä, Markku	HY, metsänarvioimistieteen laitos
Suomela, Pentti	Jyväskylän yliopisto
Teivainen, Terttu	Metsäntutkimuslaitos, MSE
Timonen, Mauri	Metsäntutkimuslaitos, Rovaniemi
Tomppo, Erkki	VTT, maankäytön laboratorio
Viitala, Risto	Metsäntutkimuslaitos, MHO



METSÄNTUTKIMUSLAITOS

Matemaattinen osasto

Osoite: PL 37, 00381 HELSINKI (Kornetintie 8)

Unioninkatu 40 A, 00170 HELSINKI

Puhelin: (90) 556 276

(90) 661 401

Seppälä, Risto, professori

Löfgrén, Raija, tutkimussihteeri

Menetelmät

Häkkinen, Risto, matemaatikko

Heinonen, Jaakko, tutkija (Joensuun tutkimusasema)

Pekkonen, Timo, tutkija

Sievänen, Risto, tutkija

Atk

Pöntinen, Jukka, atk-päällikkö

Herrala-Ylinen, Helena, tutkija

Kaila, Erkki, tutkija (Rovaniemen tutkimusasema)

Kinnunen, Hilikka, tutkija (Rovaniemen tutkimusasema)

Snellman, Carl-Gustaf, tutkija

Taipale, Markku, tutkija (Rovaniemen tutkimusasema)

Viitanen, Erkki, tutkija

Granlund, Hilikka, operaattori

Kuirinlahti, Eila, tutkimusapulainen (Rovaniemen tutkimusasema)

Soimula, Maire, operaattori

Metsätilasto

Uusitalo, Matti, tutkija

Aarne, Martti, tutkija

Lehto, Kari, tutkija

Mäki, Elina, tutkija

Leppäkumpu, Tuula, toimistos sihteeri

Nurmenniemi, Paula, toimistos sihteeri

Metsäverotus

Rauskala, Raimo, vanhempi tutkija

Kakkuri, Eero, tutkija

Kulju, Irma, kanslisti

Mäkinen, Kaija, tutkimusmestari

Sivulliset tutkijat

Hari, Pertti

Heikinheimo, Lauri

Rytkönen, Antti



Matemaattisella osastolla ilmestyneet Metsäntutkimuslaitoksen tiedonantoja -sarjan viimeisimmät julkaisut:

- nro 83 Eero Kakkuri, Raimo Rauskala & Erkki Viitanen. Kantohintatason määrittäminen. 20 s. 1983.
- nro 86 Timo Pekkonen. Leimikon puuston tilavuuden arviointi regressioennustinta käyttäen. 63 s. 1983.
- nro 87 Raimo Rauskala ja Carl-Gustaf Snellman. Kunnittaiset kantohinnat ja puukuutiometrin bruttoarvot hakkuuvuosina 1980/81 ja 1981/82. 11 s. 1983.
- nro 88 Martti Aarne & Matti Uusitalo. Ylsityisluontoisten metsien raakapuun kanto- ja hankintahinnat hakkuuvuoden 1982/83 alkupuoliskolla. 3 s. 1983.
- nro 92 Martti Aarne. Markkinahakkuut 1.7. - 31.12.1982 ja kalenterivuonna 1982 piirimetsälautakuntien alueittain. 14 s. 1983.
- nro 93 Eero Kakkuri. Ilomantsin luonnonsuojelualueiden taloudellinen merkitys puuntuotannolle ja matkailulle. 23 s. 1983.
- nro 128 Raimo Rauskala. Kunnittaiset kantohinnat ja puukuutiometrin bruttoarvot hakkuuvuonna 1982/83. 38 s. 1984.
- nro 149 Pertti Hari, Kullervo Kuusela, Pentti K. Räsänen, Risto Seppälä. Metsäntutkimukseen liittyvistä kehityssuunnista. 38 s. 1984.
- nro 152 Eero Kakkuri. Yksityismetsänomistajien puun kasvatuksen kulut vuosina 1981 ja 1982. 17 s. 1984.
- nro 157 Erkki Kaila ja Markku Taipale. Tutka-tiedonhallintaohjelmisto Tietokannan muodostus ja käyttö. 113 s. 1984.
- nro 176 Raimo Rauskala. Forest taxation and roundwood supply in Finland. 12 s. 1985.
- nro 183 Staffan Ringbom. Virkesproduktionens totala lönsamhet och dess mätning. 32 s. 1985.
- nro 191 Raimo Rauskala. Kunnittaiset kantohinnat ja puukuutiometrin bruttoarvot hakkuuvuonna 1983/84. 44 s. 1985.