ARTIKLER 144

LAND USE AND ENVIRONMENTAL STATISTICS OBTAINED BY POINT SAMPLING

AREAL- OG MILJØSTATISTIKK UTARBEIDET VED HJELP AV PUNKTUTVALG

BY
HANS VIGGO SÆBØ

STATISTISK SENTRALBYRÅ

CENTRAL BUREAU OF STATISTICS OF NORWAY

ARTIKLER FRA STATISTISK SENTRALBYRÅ NR. 144

LAND USE AND ENVIRONMENTAL STATISTICS OBTAINED BY POINT SAMPLING

AREAL- OG MILJØSTATISTIKK UTARBEIDET VED HJELP AV PUNKTUTVALG

BY
HANS VIGGO SÆBØ

STATISTISK SENTRALBYRÅ
OSLO–KONGSVINGER 1983

ISBN 82-537-1966-3
ISSN 0085-431x

PREFACE

This article is written as an invited paper on natural resource accounting to the 44th session of the International Statistical Institute in Madrid in September 1983. It gives a review of methods used for estimating and evaluating natural resources, and a description of land use and environmental statistics obtained by point sampling in Norway. In addition, the paper contains detailed considerations on the statistical properties of point sampling, especially on the precision of systematic point sampling.

Central Bureau of Statistics, Oslo, 9 August 1983

Arne Øien

FORORD

Denne artikkelen er skrevet som et invitert bidrag om naturressursregnskap til den 44. sesjonen til Det internasjonale statistikkinstituttet i Madrid i september 1983. Den gir en oversikt over metoder som brukes til å beregne naturressurser, og en beskrivelse av areal- og miljøstatistikk utarbeidet ved hjelp av punktutvalgsmetoder. Artikkelen inneholder også detaljerte vurderinger av de statistiske egenskapene til slike metoder. Presisjonen ved systematiske punktutvalg er vurdert spesielt.

Statistisk Sentralbyrå, Oslo, 9. august 1983

Arne Øien

CONTENTS

LAND USE AND ENVIRONMENTAL STATISTICS OBTAINED BY POINT SAMPLING

by Hans Viggo SÆBØ,
Central Bureau of Statistics, Oslo, Norway

1. INTRODUCTION

1. A major purpose of the Norwegian resource accounts is to provide data as a basis for resource management. The accounts are based partly on existing data, but it has also been necessary to work out improved and new statistics on natural resources. This especially concerns data on land use. In this area there has been a need to establish a unified data base comprising all land use classes as well as the whole country. Besides, such a data base should give information on potential land use and planned land use.

2. The resource accounts have been established for energy, some minerals, fish, forest and land. In addition, the work on developing environmental accounts has been given priority. So far this work is on a preliminary stage. To survey environmental status and changes in water, air and soil several problems have to be solved. One is the problem of generalization: How should the environmental quality of a geographical area as a whole be estimated and expressed? In order to work out generalized environmental statistics there is a need for new data as well as new methods for evaluating existing data.

3. The preceding paragraphs underline the link between resource accounting and use of statistical methods in collecting and integrating resource data. This paper gives a survey of data obtained by the method of point sampling, which has turned out to be successful especially in land use statistics. Examples will be given both from land use statistics and from applications in environmental statistics. Besides, the method itself and its statistical features will be considered in some detail.

4. In fact, sampling methods yield the basis for a large part of the national statistics. Examples are statistics on social conditions being based on inquires to persons or households and price and production statistics collected from a sample of firms. By point sampling one obtains information from points in space, like data on land use. In this sense the method is analogous to sampling people in surveys presenting information on social and economic conditions.

5. The paper opens with a brief review of methods currently used for resource assessments and evaluations. The point sampling method is then described, and examples of data obtained by this method are presented. The last part of the paper concerns the statistical properties of point sampling, and discusses the precision of systematic point sampling. Although the discussion is based on sampling to provide land use data, the methodological aspects may concern the collection of data on other natural resources mentioned in the paper as well: Petroleum, minerals, fish, forest and water.

2. EVALUATION METHODS FOR NATURAL RESOURCES

6. The natural resources provide materials and energy for the production processes as well as the environment for humans, animals and plants. For accounting purposes it is convenient to distinguish between the resources which go directly into the production process (material resources) and the resources which more indirectly constitute a basis for production and life (land and environmental resources). Material resources comprise both renewable, conditionally renewable and non-renewable resources, whereas the environmental resources are mainly considered to be conditionally renewable.

7. When it comes to evaluating natural resources, the most important aspect concerning material resources is to estimate the quantities which have a concentration which is technically and economically exploitable, that is the reserves. Land and environmental resources are characterized by the actual use and quality of the resource.

8. Figure 1 shows the most common procedures traditionally used for mapping or estimating the amount or quality of natural resources.

Resource		
Economic classification	Physical classification	Procedures or surveys
Material resources	<u>Mineral resources</u>:	
	Minerals and hydro carbons	-Seismic registrations -Drilling
	<u>Biological resources</u>	
	Fish	-Catch per unit of effort -Counting samples -Sonar counts -Marked members
	Forest	-Forest assessments -Forest censuses and surveys
Environmental resources	Land	-Mapping based on air photos -Map derived registers -Registers based on administrative routines -Agricultural and forest censuses and surveys
	Water Air Soil	-Measurement stations in connection with monitoring programs

FIG. 1
Procedures for evaluation natural resources

9. With the exceptions of mapping (which does not provide statistics directly), censuses and administrative registers, all the methods are sampling methods. Sampling errors therefore contribute to the uncertainty of most resource estimates. However, such errors may at least theoretically be quantified and ex-

pressed by statistical measures like standard deviations or confidence intervals. Our objective is to publish such measures of uncertainty in connection with the resource accounts. Figure 2 shows a typical probability curve for a mineral reserve. The figure shows that the reserve is 1,5 million tons (mean value) with a 90 per cent confidence interval from 0,5 to 3 million tons. This interval covers the true value with a probability of 90 per cent. The curve characterizes the quality of the reserve estimate, i.e. how much we know about the reserve. Well known reserves will have peaked curves and narrow confidence intervals, whereas less known reserves will have flat curves and wide confidence intervals. Knowledge of the accuracy of resource estimates is for example important to resource managers responsible for making investment decisions. In the following paragraphs the methods in figure 1 are briefly described and considered with respect to the estimates on which the Norwegian resource accounts are being based.

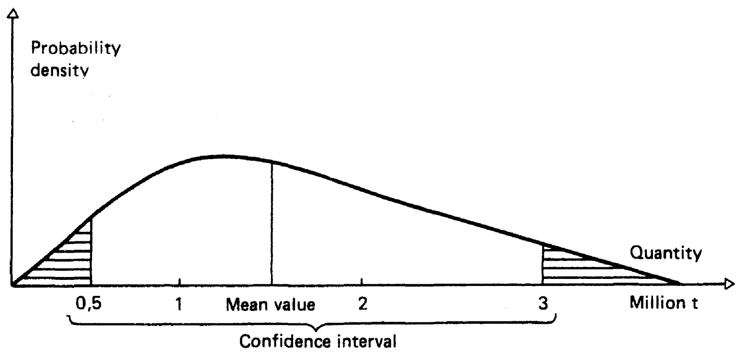


FIG. 2
The probability density of a reserve estimate

10. Minerals and petroleum cannot be proved without measurements obtained by drilling. Such measurements are expensive, and the number of bore holes therefore are minimized. However, knowledge of geological structures may give indications of the possibilities of finding the resources. As for petroleum, such knowledge is often provided by seismic surveys which are less expensive than exploration drilling. Reserve estimates are therefore based on a combination of geological information and a sample of measurements. In fact, drilling for petroleum or minerals is an application of point sampling frequently being used. In order to minimize the number of bore holes one has to know how the petroleum well or ore body varies in space. Considering correlations in spatial structure is an important aspect of point sampling.

11. As an example of resource stocks which have been published with an additional measure of uncertainty, one can mention the known Norwegian metal reserves. On the other hand we have not succeeded in obtaining uncertainty estimates concerning the Norwegian petroleum reserves. However, the Norwegian Petroleum Directorate, which is responsible for the official reserve estimates, has planned to work out such measures.

12. Several methods are used for estimating fish stocks. The estimates in the fish accounts are prepared by the Norwegian Institute of Marine Research, but put together by the Central Bureau of Statistics.

13. Stock estimates of cod, saith and haddock are based on the catch per unit of effort. One can reasonably assume that each unit of a fishing operation, say one day's fishing by a trawler, captures a constant but unknown fraction of the stock being fished. Assumptions on mortality and information from several years make it possible to estimate the stock size. The number of fish in each year class is reestimated every year (VPA-analysis). The stock is thus revaluated several times, the latest estimate being considered as the "best".

14. Until now, the estimates of fish stocks have been published without any measure of uncertainty. However, the yearly revaluation may give some ideas of the precision of these estimates. Tables and figures showing these revaluations therefore reflect the precision level.

15. The stocks of capelin is estimated on the basis of sonar registrations combined with trawl samples. The sonar studies are being undertaken from ships following line transects in the sea. The correlations in space are important aspects of this method.

16. The last method which should be mentioned concerning estimating fish stocks, is the method of marked members. This method is used for herring and mackerel. When a sample of fish, m , is marked and released into a stock with a number N and another catch, C , is made, the estimated number $\hat{N}$ is

$$\hat{N} = C \cdot \frac{m}{n} ,$$

where n is the number of marked fish recaptured. The method is a standard sampling method, based on the assumption that the marked fish are taken at random from catches and then distributed at random among the unmarked population.

17. Forest assessments have been carried out in Norway since 1919. Most of the productive forest land has been covered by the National Forest Survey in the latest survey 1964-1976. Information on land use and forestry data have been collected by fieldwork on sample sites identified on topographical maps. In fact, this method represents an application of point sampling. The sampling design being used in Norway and several other countries consists of placing 20 sample sites along the contour of a square-km. Thus the distance between neighbouring points is 200 meters. The procedure is repeated each third km. The reason for concentrating the sample sites in this way is to minimize the fieldwork efforts. From a theoretical point of view, a sampling scheme with more scattered points or sites would give better precision. Provided strong correlations in space, a site placed near to another will not represent much new information.

18. Another data base comprising information on forest should be mentioned. This is the forestry census being undertaken each 10 years. Besides, a sample of forest owners provides information on logging to the Central Bureau of Statistics each year. The last Norwegian forestry census was carried out in 1979 together with the agricultural census. Data from forest assessments and

these surveys are limited to agricultural and forest districts, and land use categories like built-up land and barren land are not covered.

19. The most extensive and comprehensive land use information is to be found on maps and air photographs. The whole of Norway is covered by topographical maps at the scale of 1: 50 000. Total land and surface water area has been measured by the Norwegian Geographical Survey on the basis of these maps.

20. Economic maps at the scale of 1: 5 000 and 1: 10 000 are made for about half of the Norwegian land area which totals 323 900 km^2. These economic maps contain property limits as well as a detailed mapping of the area by land cover and land capability classes, mainly from the point of view of agriculture and forestry. The air photographs on which these maps are based have been obtained at different points in time over the last 20 years. The Land Register is based on digitized information from these maps, and may thus also provide some land use statistics. So far, however, this register has only been completed for about 100 of the more than 450 Norwegian municipalities.

21. The land accounts have to provide comprehensive land use statistics. In order to obtain the data required with reasonable efforts, we have employed point sampling. Land use and other related information have been classified or measured at points identified on the maps or air photos. In this way two point registers or data bases have been established.

i) A national register comprising about 6 000 points to provide national land use statistics.

ii) An urban register comprising about 135 000 sample points, which makes it possible to provide separate land use statistics for urban settlements exceeding 1 000 inhabitants.

The methods used and statistics provided in these projects are considered further in the next section of this paper.

22. As mentioned in the beginning of the paper, a major problem regarding environmental statistics is the problem of generalizing information gathered at fixed points or stations. In Norway such measurements are carried out as a part of monitoring programs concerning water and air quality. So far statistics from these programs have been presented as data concerning the specific measurement sites in rivers, lakes and towns (air quality). In the next section an example will be given of how such measurements may be adapted to a statistically sound sampling framework.

3. LAND USE AND ENVIRONMENTAL STATISTICS OBTAINED BY POINT SAMPLING IN NORWAY

23. We have seen that sampling procedures play an important part in resource data evaluation. Spatial sampling like point sampling has been applied in geographical research for several years, see for instance Berry and Marble (1968), Haggett, Cliff and Frey (1977) and Dixon and Leach (1978). Matern (1960) has given a detailed evaluation of point sampling methods used in connection with forest assessments. As for land use statistics, larger point sampling projects have been or are being carried out in the United States, Canada, England, France, Sweden, Switzerland and Norway.

24. Point sampling consists of sampling points in an area which is to be surveyed, and registering characteristics like land use, altitude etc. in the points (in practice a small area surrounding the points). The sampling and registration

are most often done on maps or air photograhps. A complete inventory is likely to be time-consuming and economically infeasible, sampling procedures therefore are preferable to permit rapid collection of data. Samples referring to different points in time may in a simple way provide a basis for estimating changes. Another advantage of point sampling is that the method provides possibilities for combining different types of data, as a basis for both statistical tables and analyses.

National land use

25. Figure 3 shows the national point grid in the southern part of Norway. Above the forest line, the distance between sample points is 12 km. Below this line the grid spacing is 6 km. Sources of information are maps in different scales and coverage, and to some extent, air photographs.

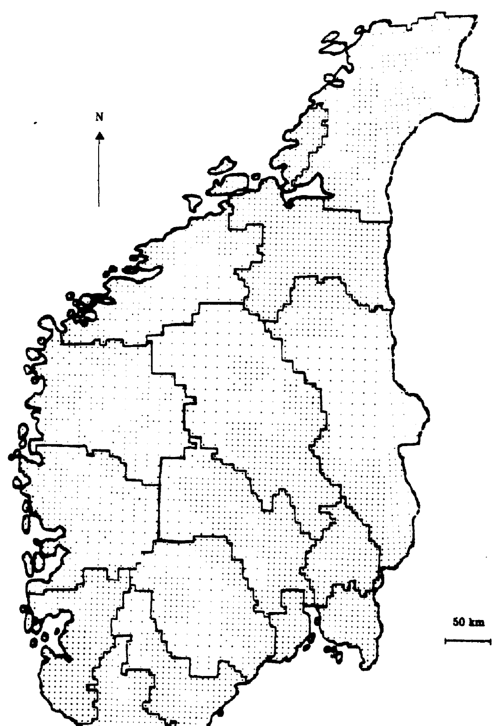


FIG.3
Point sample in the southern part of Norway

26. Various institutions working with vegetation mapping, geological mapping etc. have cooperated in terms of producing data for various areas for which maps have not yet been produced. Thus we will have a complete set of data for almost all sample points, for which ordinary mapping would have been too expensive. The following examples illustrate the kind of statistics which can be produced by means of such point sampling.

27. Table I shows actual land use and cultivation land in Norway. Cultivation land is land which is not fully cultivated but suited for cultivation. The table shows that the amount of cultivation land in Norway is of the same magnitude as the amount of land being cultivated today. About half of this land is covered by productive forest today. Such estimates of cultivation land have not been presented before in Norway.

Table I

Land use and cultivation land in Norway. Km^2

	Total	Cultivation land
Total	323 900	8 900
Built-up land	3 700	-
Agricultural land	11 500	400
Fully cultivated land	8 600	.
Surface cultivated and permanent grassland	2 900	400
Forest land	119 200	5 500
Productive forest	72 500	4 500
Other forest	46 700	1 000
Bogs, wet land	20 300	2 400
Open land with soil cover	99 100	600
Barren land	52 500	-
Fresh water	17 600	-

28. Another example will describe a way in which point sampling may provide data on wilderness. Wilderness is associated with the distance to such physical encroahments as roads, railways and regulated lakes and rivers. Such distances have been registered in the point sampling project, and we can for example classify the land in three different zones:

- land located less than 1 km from a road, a railway or a regulated water course
- land located between 1 km and 5 km from an encroachment
- land located more than 5 km from the nearest encroachment. This land could be defined as "wilderness".

According to this definition, 23 per cent of Norway is described as wilderness. Most of it is open (mostly unproductive) and barren land. The regional differences are substantial, the amount of wilderness in northern Norway being 44 per cent. In the national parks the corresponding figure is 80 per cent.

Land use in urban settlements

29. In total, approximately 250 urban areas, that is all urban areas with more than 1 000 inhabitants, have been classified for the years 1955, 1965 and 1975 (or years as close as possible if air photos for the years in question do

not exist). We have used different regular grids, varying from 100 to 300 m spacing. The classification at all three points of time has been carried out simultaneously in the same points. This reduces classification errors and gives reliable information on changes in land use.

30. The classification is done for three separate levels, the area level (residential areas, industrial areas etc.), field (or parcel) level, and the lowest level, which we have called physical structure. Each sample point is thus classified according to the physical surface or structure, to the field, and to the area class surrounding the sample point. This procedure makes it possible to combine data on different dimensions and geographical levels of land use. Figure 4 illustrates the differences between different levels of classification.

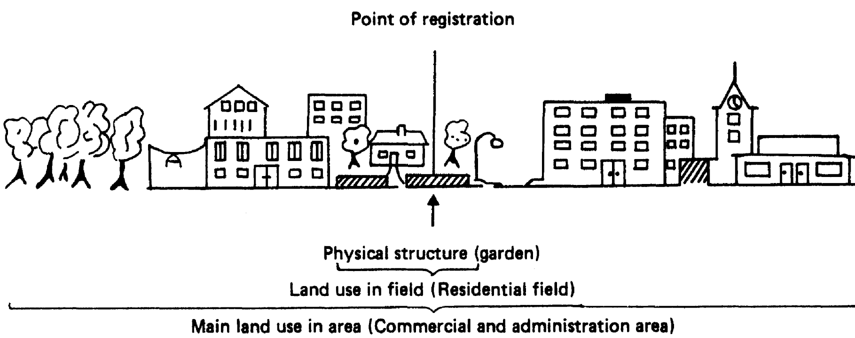
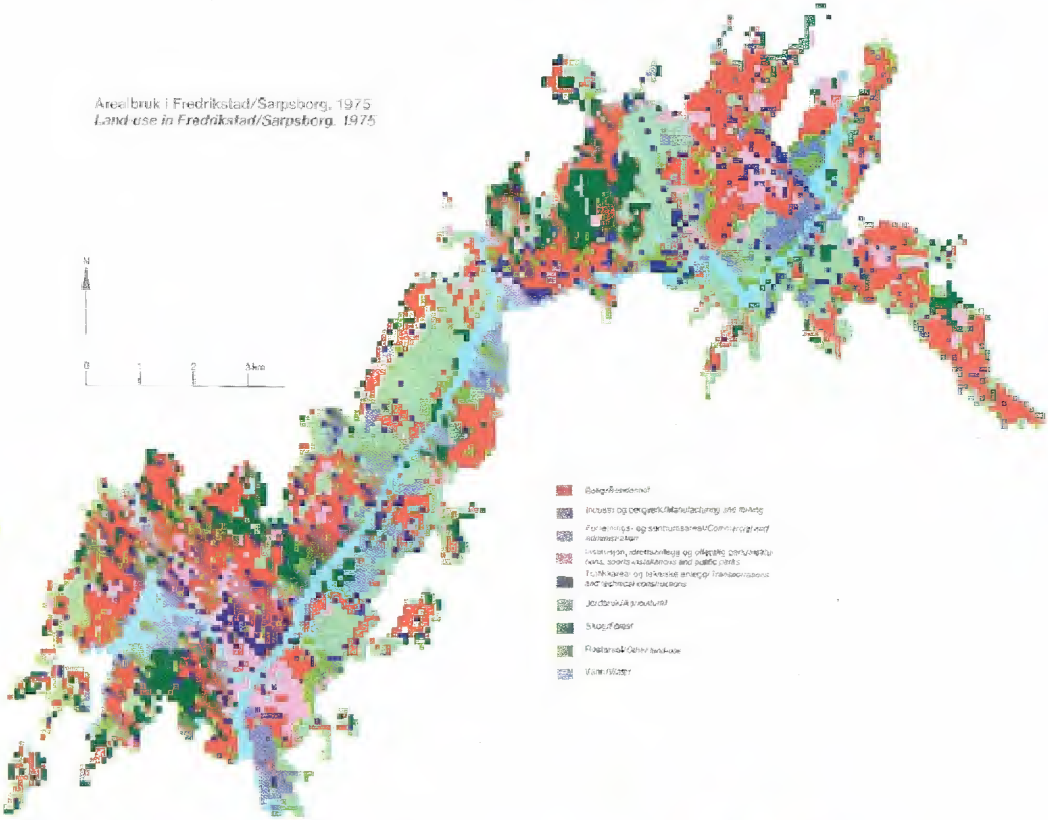


FIG.4
Classification of land use at different levels

31. The method enables us to calculate building densities (defined as the part of built-up land covered by buildings). About 15 per cent of built-up land in Norway is covered with buildings. This density varies between 11 and 22 per cent when the different urban settlements are considered. The amount of built-up land being used for garden purposes is 35 per cent on average.

32. Table II shows the transitions from the different land use categories in 1955 to built-up land in 1965 and the corresponding transitions during the period 1965-1975. The table indicates that the Norwegian land use policy which seeks to conserve agricultural land, has had limited success, although a smaller part of the development has taken place on agricultural land in the last period compared to the first. To some extent residential development has been forced to take place on forest land as opposed to agricultural land. Other built-up categories, however, have the same land "consumption pattern" in the last as in the first period, using mainly agricultural land and other non-built-up land. A considerable part of the land being classified as other non-built-up land consists of (former) agricultural land lying fallow.

Arealbruk i Fredrikstad/Sarpsborg, 1975
Land-use in Fredrikstad/Sarpsborg, 1975



Potensielle fortettingsarealer i boligstrøk med småhusbebyggelse. Fredrikstad/Sarpsborg, 1975
Undeveloped land within residential areas. Fredrikstad/Sarpsborg, 1975

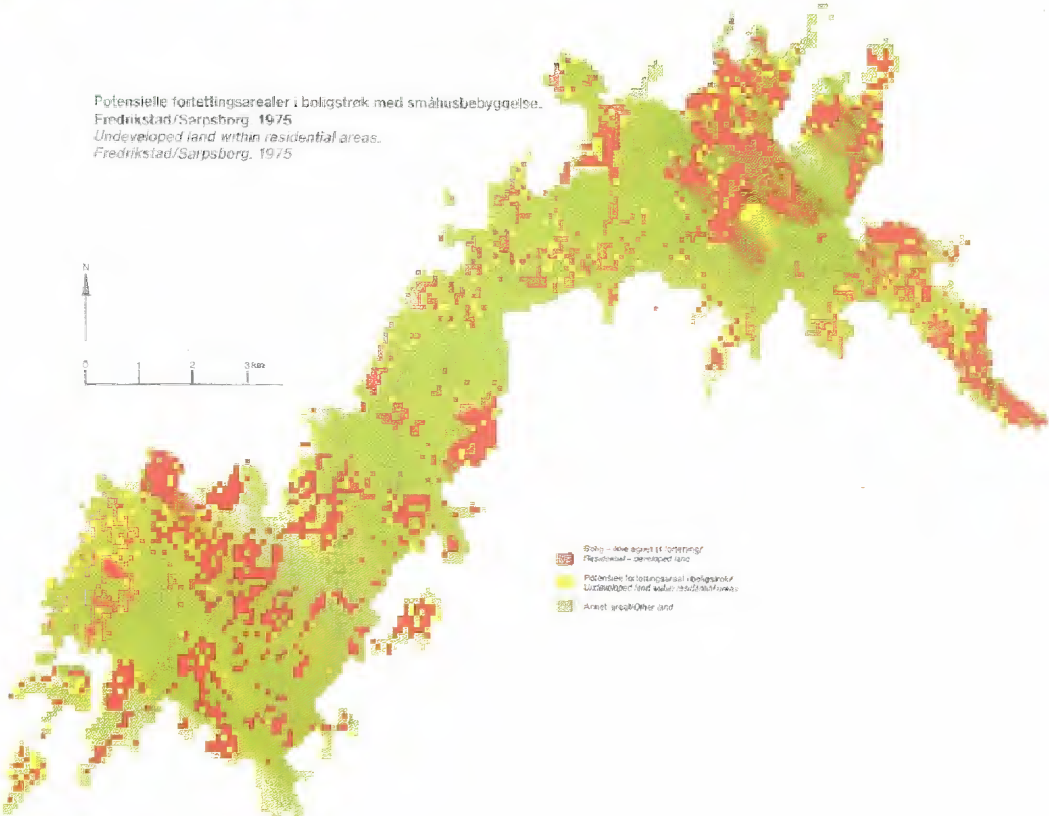


Table II

Urban development (changes from unbuilt to built-up land) 1955-1965 and 1965-1975

From	To built-up land				
	Total		Residential	Manufacturing, commercial etc.	Transportation and technical constructions
	Km^2	Per cent			
1955-1965					
Total	173,2	100	100	100	100
Agricultural land	62,9	36	38	35	33
Forest land	57,9	33	38	24	30
Other unbuilt land	52,4	31	25	41	37
1965-1975					
Total	215,3	100	100	100	100
Agricultural land	68,4	32	31	34	30
Forest land	84,6	39	45	26	39
Other unbuilt land	62,2	29	23	40	31

33. Land within built-up areas where it is possible to increase the density of residential buildings has also been registered. In the largest Norwegian urban settlements this land corresponds to more than 20 per cent of the land occupied by small houses. This concerns both the possibility of dividing large residential properties and of developing vacant sites in residential areas.

34. Maps will be produced to present supplementary information on large urban areas. The maps are not accurate, being based on point sampling data. They will, nevertheless, give an impression of land use patterns.

35. The production of such land use statistics should be repeated every 5 or 10 years. We will therefore produce new statistics for 1980. The registrations are now being undertaken by the Norwegian Geographical Survey, making the Central Bureau of Statistics responsible only for controlling and publishing the data. The collection of data from Oslo, which is the largest urban area in Norway, is just finished. Land use data for 1980 can be linked to the data from the last Norwegian population census, to constitute a basis for analyses.

Point sampling and remote sensing

36. Air photos and maps covering rural and mountainous areas are not updated regularily. Point sampling, therefore, cannot provide statistics on current land use changes in these areas. One possibility for estimating changes in built-up land, agricultural land and forest is to utilize remote sensing. The Norwegian Central Bureau of Statistics therefore is evaluating the simulations being carried out by plane to prepare data collection by means of the French satellite SPOT. This satellite will be put in service from 1984/85. We are

particularly interested in testing statistically the quality of remote sensing data. The quality will be analyzed by help of a comparison of land use classification carried out in points on air photos and on simulated satellite registrations.

Point sampling in environmental statistics

37. In this paper we will present a method for producing generalized statistics on water quality on the basis of measurements carried out at specific points (measurements stations). The example concerns the Laagen river system in the central part of Norway, covering a total catchment of 11 500 km^2.

38. The basis for a statistical sampling design is a river basin archive system, which has been developed by the Central Bureau of Statistics. This register has now been transferred to the Norwegian Water and Electricity Board. It comprises all river reaches and lakes in Norway. To each river reach there are attached data on length, rate of flow etc.

39. In the sampling design river reaches are weigted, not only by their length but also by the flow rate. Thus, each point in a river is given a sampling probability proportional to the rate of flow. The example presented here is based on 20 points placed systematically in the Laagen river system as shown in figure 5. Since the rate of flow increases down the river, the points are placed denser here than in the upper part of the river system.

40. The water quality in a river is correlated due to the fact that measurements carried out in two neighbouring points are likely to give very similar results. The measurement sites therefore are placed systematically. Estimates of "mean water quality" in the river will have less variance by this method than by a random procedure (see next chapter).

41. In addition to providing measures of mean water quality, this method makes it possible to estimate the distribution of water by quality. Figure 6 shows such a distribution by acidity (categories of pH) in the Laagen river system except for the lake Mjösa. The figures are mainly referring to the summer of 1970, and show that acidity is not a critical problem in this part of Norway.

42. It should be emphasized that the example presented here is just meant to give an illustration of how statistical methods may be implemented in evaluating environmental data. "Mean water quality" is not necessarily a valid measure. The number of people being affected by water of certain quality categories usually is more important to resource managers. It is our object to account for such aspects in a more appropriate manner in the further development of environmental statistics in Norway.

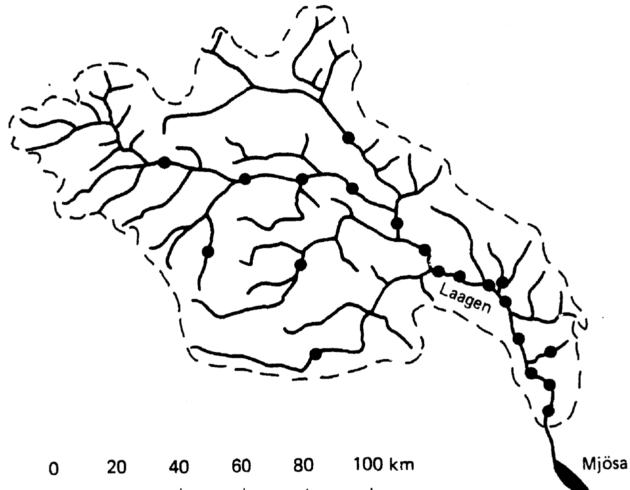


FIG.5
The Laagen river system with sampled measurement points. An example

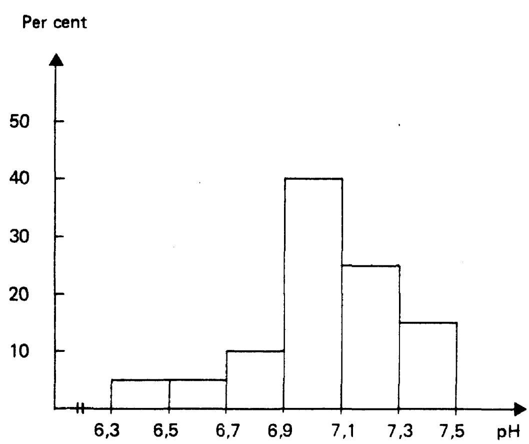


FIG. 6
Water by acidity in the Laagen river system

4. ON THE PRECISION OF POINT SAMPLING

43. The precision of sampling depends on the number of sample units, here points, and the sampling procedure or design. One issue turns out to be important concerning sampling in space in particular, namely the correlation between adjacent points. This implies that the precision of systematic sampling usually is better than the precision of a more random procedure. Unfortunately, it is difficult to take this correlation into account in a simple and exact way when making and comparing sampling designs.

44. According to sampling theory as described in a standard textbook like Cochran (1977), measures expressing the accuracy of point sampling when using a random sampling procedure (simple random sampling and stratified random sampling) may be deduced without considering correlations in space. Usually the problem is to estimate the share P of the total area A that belongs to a specific land use category. The estimator applied is

$$\hat{P} = \frac{1}{n} \sum_{j=1}^{n} I_j , \tag{1}$$

where n is the number of points, and I_j is an indicator which equals 1 if point j belongs to the class we want to estimate and 0 otherwise. In simple random sampling the variance of this estimator is given by

$$\text{Var } \hat{P} = \frac{P(1-P)}{n} . \tag{2}$$

This gives the standard deviation

$$S = \sqrt{\frac{P(1-P)}{n}} , \tag{3}$$

and a 95 per cent confidence interval may be expressed as

$$[\hat{P}-2S,\ \hat{P}+2S] .$$

This interval includes the correct value of P with a probability of 95 per cent. Formulae concerning stratified random sampling and estimation of land use changes by applying fixed points have been given by Garnåsjordet and Sæbø (1980). In practice, however, square point grids have been and are being used, and this paper, accordingly, will focus on variance estimation in systematic sampling.

45. A systematic sample as a whole only represents one realization of a random procedure. First of all, therefore, it should be recognized that the only way of estimating the exact value of the sampling variance when using a systematic sample, is to repeat the sampling procedure many times with a random location of the first point. This method is extremely laborious and, of course, not applicable in practice, except for testing or controlling approximation methods. So far, we have not used variance estimation methods which require repeated sampling or information on the complete land use structure (also between the grid points) in Norway. Our approach, which is outlined in the next paragraphs, has been based on studies of variations or correlations of land use categories in space.

The model of regionalized variables

46. Let $\underline{x}$ be a point in space (vector), and $z(\underline{x})$ the value of a function which we are interested in at point $\underline{x}$ (for instance the temperature). Such a function is called a regionalized variable, and can be regarded as a realization of a certain random function or stochastic process $Z(\underline{x})$. This model is convenient when taking correlations in space into account. The theory of regionalized variables has been developed by Matheron (1971) in particular.

47. As for regionalized variables representing natural phenomena, there exists only one observable realization of the corresponding random function. We are facing the problem of finding the statistical characteristics such as the expected value and the variance of this function. This is possible if we assume stationarity, which means that the expected value and the spatial covariance of the random function $Z(\underline{x})$ is the same all over the field of interest. If the natural phenomenon to be considered cannot be described realistically by the above assumptions, weaker assumptions may be established.

48. Let the expected value be:

$$E[Z(\underline{x})] = p, \tag{4}$$

and the covariance:

$$Cov[Z(\underline{x}),Z(\underline{x}+\underline{h})] = E\{[Z(\underline{x}) - p][Z(\underline{x}+\underline{h}) - p]\} = K(\underline{x},\underline{x}+\underline{h}) = K(\underline{h}) \tag{5}$$

where $\underline{h}$ is a vector in R^n. Hence the variance of the random function $Z(\underline{x})$ is:

$$Var[Z(\underline{x})] = K(\underline{x},\underline{x}) = K(0). \tag{6}$$

49. In the case of land use, the random function may be regarded as a binomial variable in each point of the plane:

$$Z(\underline{x}) = \begin{cases} 1 \text{ if the point } \underline{x} \text{ belongs to the class considered,} \\ 0 \text{ otherwise.} \end{cases}$$

50. The expected value of $Z(\underline{x})$ is the probability that the point belongs to the land use class considered. We have

$$E[Z(\underline{x})] = Pr[Z(\underline{x})=1] = p, \tag{7}$$

$$Var[Z(\underline{x})] = p(1-p). \tag{8}$$

51. The covariance may be expressed by p and the conditional probability $p(\underline{h})$ for $(\underline{x}+\underline{h})$ to be in the class, given that $\underline{x}$ is in the class, as shown by Garnåsjordet and Sæbø (1980) is:

$$K(\underline{h}) = p(p(\underline{h})-p). \tag{9}$$

52. The average of $Z(\underline{x})$ in an area A is given by

$$P = \frac{1}{A} \int_A Z(\underline{x}) d\underline{x}. \tag{10}$$

This average is a random variable, the only observable realization of which is the share of the area A that belongs to the land use category considered.

The estimation variance

53. The estimator being used when point sampling land use classes can be expressed as

$$\hat{P} = \frac{1}{n} \sum_j Z(\underline{x}_j), \tag{11}$$

where $\underline{x}_j$ denotes sampling point j and n the total number of points.

54. The expectation of $\hat{P}$ and P is following (7)

$$E\hat{P} = EP = p, \tag{12}$$

thus the estimator $\hat{P}$ is without bias. The variance to be estimated is

$$\mathrm{Var}(\hat{P}-P) = \mathrm{Var}\ \hat{P} + \mathrm{Var}\ P - 2\ \mathrm{Cov}\ (\hat{P},P). \tag{13}$$

55. A straightforward procedure inserting (10) and (11) into (13) gives

$$\mathrm{Var}(\hat{P}-P) = \frac{1}{n^2} \sum_i \sum_j K(\underline{x}_i-\underline{x}_j) + \frac{1}{A^2} \int_A d\underline{y} \int_A K(\underline{y}-\underline{x})d\underline{x} - \frac{2}{nA} \sum_i \int_A K(\underline{x}_i-\underline{x})d\underline{x}. \tag{14}$$

56. If we assume that $K(\underline{h}) = K(|\underline{h}|) = K(h)$, i.e. that the covariance is isotropic, the distance where the covariance equals 0 is called the range of the actual land use class. For this distance the conditional probability of $(\underline{x}+\underline{h})$ being in the class provided that $\underline{x}$ is in the class equals the unconditional probability p, and one point does not provide information about other points beyond this distance. If the covariances are not isotropic, ranges which depend on the directions in space can be defined analogously.

57. To simplify the variance expression (14), we now assume that there are no border effects, which means that $\int_A K(\underline{x}_i-\underline{x})d\underline{x}$ is independent of $\underline{x}_i$. This corresponds to the assumption that the area A is unlimited or large compared to the range of the phenomenon being studied.

58. The variance can now be written as:

$$\mathrm{Var}(\hat{P}-P) = \frac{1}{n} \sum_j K(\underline{x}_i-\underline{x}_j) - \frac{1}{A} \int_A K(\underline{x}_i-\underline{x})d\underline{x}. \tag{15}$$

The variance turns out to be the difference between an approximate value and the exact value of the integral $\frac{1}{A} \int_A K(\underline{x}_i-\underline{x})d\underline{x}$.

Hence it becomes smaller as:

- the grid spacing decreases
- the function K, and therefore the regionalized variable itself, becomes more regular.

The integral is proportional to the volume below the covariance function, as shown in figure 7.

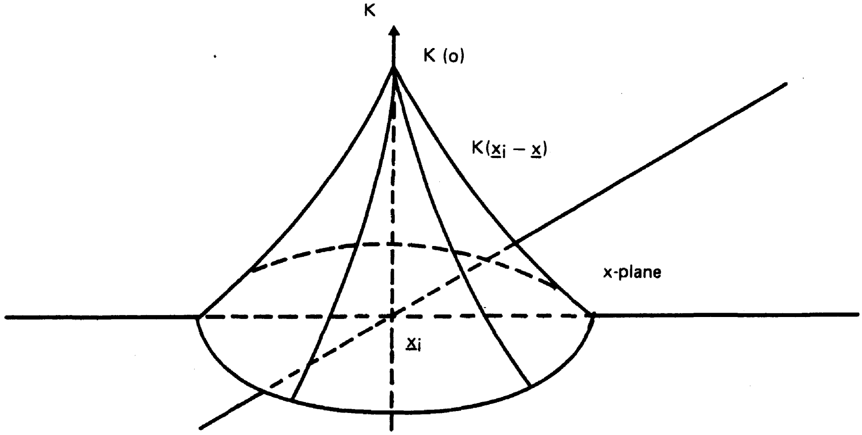


FIG. 7

The covariance cone about a random point $\underline{x}_i$

59. Unfortunately, the integral in (15) cannot be estimated by the point observations alone. To estimate the variance in systematic point sampling, we have to use a model in addition to point observations. The basis for the model has to be estimations of the covariances between points in our grid: $K(\underline{x}_i-\underline{x}_j)$. Such estimations have been carried out for a large Norwegian urban settlement. The covariances have been estimated assuming isotropy. Results from this study are presented in the last part of the paper. We are later going to carry out corresponding estimations for other urban areas as well as for a rural district.

Approximation methods

60. Several approximation methods have been proposed to simplify estimation of sampling variances when using systematic point procedures. Matheron (1971) presents a formula which can be useful in cases where the grid spacing is small compared to the range of the phenomenon being studied. The formula is based on an expansion of the covariance near the origin. The shape of the covariance curve is especially important here. The application of the formula in the two-dimensional case requires a count of pairs of vertically and horizontally adjacent points of which one point is classified in the class where we want to estimate the variance, and the other is not. The members of such pairs are called $2n_1$ and $2n_2$ respectively, n_2 being the smallest of them (the indices 1 and 2 refer to the main directions of the grid). In our notation the formula can be written as:

$$\mathrm{Var}(\hat{P}-P) \approx \frac{1}{n^2}\left[\frac{1}{6}n_2 + 0{,}061\,\frac{n_1^2}{n_2}\right], \qquad n_2 \leq n_1. \qquad (16)$$

61. One possibility to approximate the sampling variance in systematic sampling is to regard the sample as a stratified random sample with two or even more points per stratum. The area considered is divided into regular strata and the usual variance formula concerning proportional stratification is then applied (see for instance Cochran (1977) or Garnåsjordet and Sæbø (1980)). Milne (1959) has done an empirical study that suggests that this method may give a good idea of the true sampling variance. In general, the method will probably overestimate the sampling variance, though not to the same extent as will be the case if formula (2) concerning a complete random sample is used. The method has been applied by the National Forest Survey of Norway.

62. Several authors like Zöhrer (1978) have expressed the standard deviation when using systematic point sampling in formulae similar to (3) concerning simple random sampling, but with a higher exponent than 0,5 attached to n (for example 0,75). In addition, these formulae usually contain a factor expressing the shape of the land use class to be estimated. Low factors apply to regular shapes like circles, whereas high factors apply to irregular shapes resulting in larger standard deviations. Other kinds of formulae have also been considered, see for instance Emmott (1981). All these formulae, however, should be fitted to the actual land use structures or at least tested before they can be used as approximation formulae in new areas. In Norway, we have not considered the use of such formulae.

63. Matern (1960) considers several quadratic forms, that is quadrats of linear combinations of the observed values in the sample points, as estimates for sampling variance when using systematic point samples. One such form in particular, should be mentioned since it also has been applied by Kölbl and Trachsler (1981) for estimating variances when obtaining land use statistics by point sampling in Switzerland. This is the formula of "cross-differences", being based on observed differences between point values in quadrats consisting of 4 neighbouring points.

64. Let x_{ij} denote a point in the grid with coordinates (i,j). The formula of cross differences is based on the quadratic form

$$T_{ij} = \frac{1}{4}[Z(x_{ij}) - Z(x_{i,j+1}) - Z(x_{i+1,j}) + Z(x_{i+1,j+1})]^2. \quad (17)$$

The variance estimator is given by

$$V = \frac{1}{n^2} \sum_i \sum_j T_{ij}. \quad (18)$$

65. Assuming isotropic covariance and point distance d in a square grid, a straightforward calculation of the expected value of V, provides

$$EV = \frac{1}{n}[K(0) - 2K(d) + K(d\sqrt{2})]. \quad (19)$$

66. It may be shown that this expression is a reasonable approximation for the variance formula (15) if the land use category considered has a range of the same magnitude as the smallest grid distances.

67. Matern (1960) has shown that the cross-difference formula overestimates the true variance in cases with a "smooth" covariance function. This especially concerns variables with long ranges as compared to the point distance. Later it will be shown how this approximation formula tends to overestimate the variance in an example from a Norwegian urban settlement.

Empirical studies

68. As pointed out, application of the variance formula (15) requires a model for the covariance function. To estimate this function we make the assumption of stationarity, and usually we will need to assume isotropy as well to obtain a sufficiently compact description of the covariance. If there are a priori grounds for assuming that one direction differs in character from the others, we might assume geometric anisotropy and take $K(\underline{x})$ as a function of, say, $\sqrt{x_1^2 + ax_2^2}$ where a is an additional parameter. Such anisotropies may imply increasing estimation variance when using a systematic point sample. In cases where regularities in the terrain coincide with the grid distances, the variance may even be larger than in simple random sampling. However, we have no reason to believe that such structures have affected the precision of point sampling land use in Norway in general. Milne (1959) has discussed the problem of systematic sampling and regularities in space or in populations in environmental studies, concluding that the risk of periodic variation defeating the systematic sampling method is small.

69. Convenient classes of isotropic covariances are rather few. The functions or combinations of the functions e^{-ah} and e^{-ah^2} have often been proposed. Ripley (1981) mentions some other functions which have been considered too, but remarks that very little work has been done on fitting the parameters of any of these models. We have tried several models for expressing correlation. The models have been fitted to estimated covariances between land use classes in our point grid in a large Norwegian urban settlement, Fredrikstad/Sarpsborg. The estimation of covariances is considered by Omre (1980).

70. The (isotropic) correlation coefficient R(h) may be expressed as

$$R(h) = \frac{K(h)}{K(0)} = \frac{K(h)}{p(1-p)} \qquad (20)$$

if binominal variables like land use categories are considered.

Inserting into (15) gives

$$\mathrm{Var}(\hat{P}-P) = \frac{1}{n}\, p(1-p)\Big[\sum_j R(|\underline{x}_j|) - \frac{2\pi}{d^2}\int_0^\infty h\,R(h)\,dh\Big]. \qquad (21)$$

Here $\underline{x}_i$ is set equal to 0 and $A=nd^2$, where the smallest point distance is d. The total area A is assumed to be large compared to the range of the land use classes, so that no border effects occur. This formula has been used to estimate variances based on fitted correlation coefficients (correlogram). The expression within the parenthesis equals the ratio between the variance when using a square point grid and $\frac{1}{n}\,p(1-p)$, which corresponds to the variance in simple random point sampling.

71. The correlogram should be 0 or nearly 0 for a "reasonable" distance corresponding to the range of the actual land use class. Due to the limitations of the urban area, the empirical correlogram will cross the 0-line and become negative for some distances. We have used positive functions as if the area was unlimited. The functions should differ very little from 0 at a distance corresponding to the "empirical range". This might cause a problem when choosing convenient functions.

72. Several functions have been tried, the "best" and most convenient ones have turned out to be:

Model 1:

$$R_1(h) = \frac{a_1}{a_1+h} \exp(-b_1 h^2), \qquad (22)$$

model 2:

$$R_2(h) = \frac{a_2}{a_2+h} \exp(b_2 h - c_2 h^2), \qquad (23)$$

where R(h) denotes the correlation as a function of the distance h in the plane. The models have been estimated in a log-linear form.

73. The results of fitting model 1 are shown in table III. This model gives good correlograms for most land use classes, except for transportation land and forest, where $\hat{b}<0$, which means that $R \to \infty$ when h increases. Model 2 gives valid correlograms for these classes.

Table III

Correlation in space for land use categories in Fredrikstad/Sarpsborg

Correlation function $R = \frac{a}{a+h} \exp(-bh^2)$

Land use category	Estimated constants		Goodness of fit	Note
	$\hat{a}$	$\hat{b}$	RSQ	
	m	$10^{-6} m^{-2}$		
Residential land	65,6	2,69	0,984	
Manufacturing	89,6	1,22	0,996	
Commercial and city land	20,3	0,31	0,920	
Institutions etc.	45,7	1,18	0,876	
Transportation etc.	10,4	[< 0]	0,586	Not valid
Agricultural land	94,7	0,20	0,983	
Forest	61,6	[< 0]	0,943	Not valid
Other land use	19,9	6,36	0,925	
Water (river)	82,6	2,18	0,883	

74. Figure 8 shows the fitted correlogram for some land use categories in the urban area: Residential, manufacturing and transportation land. The practical range could be defined as the distance where the correlogram function equals

0.05. In the example this distance varies as follows:

Practical range	- residential land	500 m
	- manufacturing	800 m
	- transportation etc.	150 m

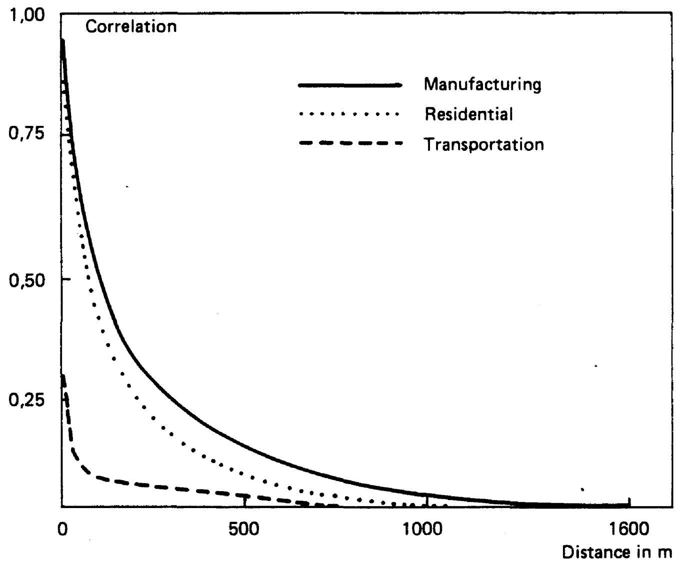


FIG. 8
Correlation for land use categories in Sarpsborg/Fredrikstad

75. Expression (21) has been applied directly when estimating sampling variances. The most important terms of $\Sigma\, R(\underline{x}_j)$ are

$$\sum_j R(\underline{x}_j) = 1 + 4R(100) + 4R(100\sqrt{2}) + 4R(200) + \ldots \qquad (24)$$

The summation and integration is undertaken up to 1600 meters. The robustness of the calculations has been checked by varying this distance.

76. Table IV shows the estimated variances and standard deviations when using a 100 m-grid compared to the variances and standard deviations to be considered in simple random point sampling. The estimated functions are assumed to be valid. The variance reduction is considerable, more than 50 per cent for most classes. The reduction is largest for agriculture and manufacturing land, two land use classes consisting of larger, homogeneous units in the actual urban area. There is almost no variance reduction for transportation land, which has a flat correlogram from a distance of 10 meters.

Table IV

Variance and standard deviation in land use sampling in Fredrikstad/Sarpsborg. Values for 100 x 100 m point grid relative to those obtained by simple random sampling

Land use category	Model used for correlogram	Relative variance	Relative standard deviation
Residential	1	0.29	0.54
	2	0.34	0.58
Manufacturing	1	0.23	0.48
	2	0.25	0.50
Commercial and city land	1	0.58	0.76
Institutions etc.	1	0.38	0.62
Transportation etc.	2	0.87	0.93
Agricultural land	1	0.22	0.47
	2	0.22	0.47
Forest	2	0.39	0.63
Other land use	1	0.59	0.77
	2	0.50	0.71
Water (river)	1	0.25	0.50

77. The estimated land use categories in Fredrikstad/Sarpsborg together with standard deviations and confidence intervals are given in table V, assuming correlation function 1 (function 2 for transportation land and forest).

Table V

Land use in Fredrikstad/Sarpsborg. Standard deviations and 95-per cent confidence intervals

Land use category	Estimated size	P	Standard deviation S	Relative standard deviation 100 S/P	Confidence interval
	Number of points[1]	Per cent			
Total	6 896	100,0	-		
Residential land	2 028	29,4	0,3	1,0	28,8-30,0
Manufacturing	571	8,3	0,2	1,9	8,0- 8,6
Commercial and city land	115	1,7	0,1	7,0	1,5- 1,9
Institutions etc.	345	5,0	0,2	3,3	4,7- 5,3
Transportation etc.	505	7,3	0,3	4,0	6,7- 7,9
Agricultural land	1 313	19,0	0,2	1,2	18,6-19,4
Forest	712	10,3	0,2	2,2	9,8-10,8
Other land use	824	11,9	0,3	2,5	11,3-12,5
Water	483	7,0	0,2	2,2	6,7- 7,3

1) 1 point corresponds to 1 hectare

78. The variances when using grids with different spacing have been estimated on the basis of the correlation functions described. Some of the results are shown in figure 9. The relative variance of grid sampling compared to simple random sampling approaches 1 when the point distance increases and the number of points decreases.

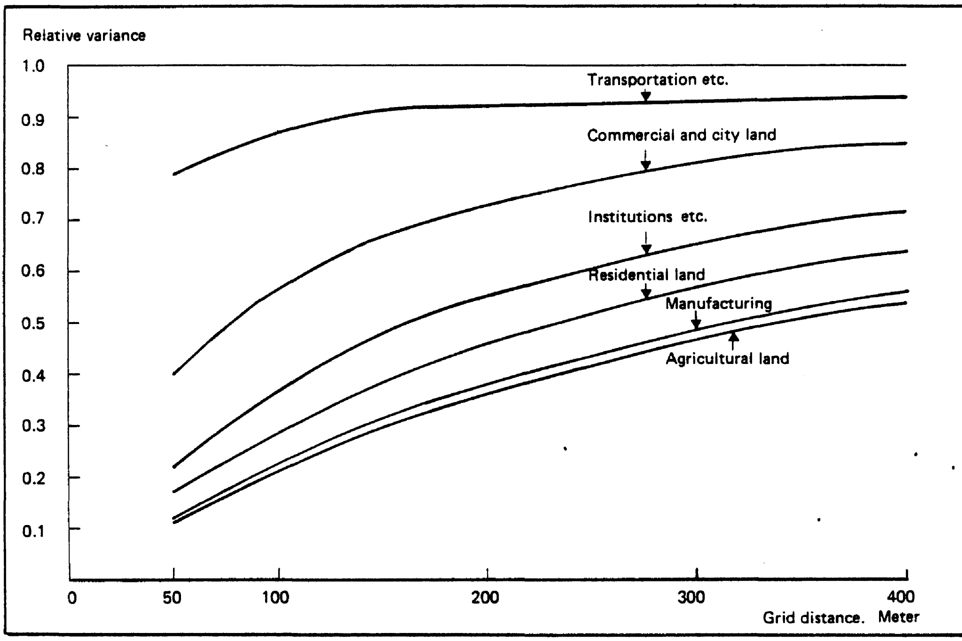


FIG. 9
The variance when using a square point grid relative to the variance of simple random sampling

79. The relative variance considered approaches zero as the number of points increases, but measurement errors will gradually dominate the sampling errors, and the benefit of increasing the point number will be limited. Experiments done in Norway suggest that measurement errors do not contribute considerably to the total error when using a 100 m-grid, see Engebretsen and Sæbø (1979).

80. The approximation methods mentioned in paragraph 60 and 64 have also been applied in the Fredrikstad/Sarpsborg study. Matherons formula (16) results in lower variances and standard deviations than those estimated by means of the fitted correlograms. The deviations are smallest for the land use categories with long ranges like agricultural land. This is what should be expected since the approximation formula is based on the assumption that the smallest grid distance is small compared to the extent of the phenomenon to be conside-red. In our case for example, this approximation formula yields a variance which is about 40 per cent lower than the variance based on the fitted correlo-gram for agricultural land, and 75 per cent lower for transportation. Hence, the point distance is not small enough for this approximation formula to be applied successfully.

81. The expectation of the "cross-difference" variance formula has been given in (19). Expressed by the correlation coefficient R this expectation may be written as:

$$EV = \frac{p(1-p)}{n}[R(0) - 2R(d) + R(d\sqrt{2})] \ . \qquad (25)$$

This expression has been applied directly assuming the fitted correlograms for the land use classes in Sarpsborg/Fredrikstad. The resulting variances are compared to the variances from table IV in table VI. The "cross-difference" formula overestimates the variances, provided that our correlogram functions are valid, as is the case when using simple exponential correlograms, which has been done by Matern (1960). The variances and standard deviations, however, are moderately overestimated for classes with ranges of about the same magnitude as the grid distance. Nevertheless, this approximation formula seems to be much better than the formula concerning simple random sampling when using a systematic point grid.

Table VI

Expected variance for land use classes according to cross-differences, and exact calculations. Values for 100 x 100 m point grid relative to simple random sampling

Land use category	Model for correlogram	Relative variance: Expectation of cross-difference formula	Relative variance: Exact value
Residential land	1	0,53	0,29
Manufacturing	1	0,44	0,23
Commercial and city land	1	0,78	0,58
Institutions etc.	1	0,62	0,38
Transportation etc.	2	0,96	0,87
Agricultural land	1	0,43	0,22
Forest	2	0,66	0,39
Other land use	1	0,80	0,59
Water (river)	1	0,47	0,25

LAND USE AND ENVIRONMENTAL STATISTICS OBTAINED BY POINT SAMPLING

SUMMARY

The Norwegian resource accounts are based partly on existing data, but it has also been necessary to develop improved and new statistics on natural resources. This especially concerns data on land use, where the method of point sampling has turned out to be successful. The paper gives a survey of data obtained by this method.

Sampling methods provide the basis for a large part of the national statistics. By point sampling one obtains information from points in space, such as data on land use. The paper opens with a brief review of methods currently used for resource assessments and evaluations. Examples of land use and environmental statistics obtained by point sampling are then presented. Two point registers have been established in Norway:

i) A national register comprising about 6 000 points to provide national land use statistics.

ii) An urban register comprising about 135 000 points, which makes it possible to provide separate land use statistics for urban settlements.

In the first project, information from different kinds of maps and air photos has been collected: Land use, altitude, slope, vegetation, geology, distance to nearest road etc. These data provide a complete statistics on land use and related information covering the whole country.

The point sampling in urban settlements has been carried out on the basis of air photos referring to three periods in time: 1955, 1965 and 1975, thus providing statistics on land use changes as well.

This year, priority has been given to the development of environmental accounts as an integrated part of the Norwegian resource accounts. Such accounts will provide systematic statistics on emissions together with surveys reflecting environmental quality. In order to provide statistics on water resources, point sampling in water courses is being evaluated. The paper contains an example of how sampling methods may be used to establish a statistically sound framework for statistics on water quality.

The precision of sampling depends on the number of sampling units (points) and the procedure or design of sampling. Simple formulae express the accuracy of the method if the sampling points are located at random. However, the points are most often located regularily in a grid system. This is practical, and it may be shown that the precision usually is better than the precision of simple random sampling.

The last part of the paper is devoted to an evaluation of systematic point sampling in the light of statistical theory. An example from a large Norwegian urban settlement indicates that the precision of systematic point sampling may be considerably better than the precision of simple random sampling when estimating land use categories. Standard deviations and confidence intervals when for example estimating the amount of agricultural or manufacturing land by means of a 100 m point grid, may be less than one half of the corresponding uncertainty measures of simple random sampling.

BIBLIOGRAPHY

Berry, B.J.L. and Marble, D.F. (1968): "Spatial Analyses". Prentice-Hall, New Jersey.

Cochran, W.C. (1977): "Sampling Techniques". John Wiley & Sons, New York.

Dixon, C. and Leach, B (1978): "Sampling Methods for Geographical Research". Concepts and Techniques in Modern Geography no. 17, University of East Anglia, Norwich.

Emmott, C. (1981): "Computer aided Photo Interpretation for Land Use". Survey and Mapping 1981, Paper No. F5.

Engebretsen, Ø. and Sæbø, H.V. (1979): "En vurdering av kvaliteten ved areal-klassifisering på flybilder". Metodehefte 24 from the Central Bureau of Statistics (IO 79/6). Oslo.

Garnåsjordet, P.A.; Lone, Ø. and Sæbø. H.V. (1980): "Two notes on land use statistics". Report from the Central Bureau of Statistics (80/31).

Haggett, P.; Cliff, A.D. and Frey, A. (1977): "Locational Methods". Arnold, London.

Kölbl, O.; Mahrer, F; Meyer, B. and Trachsler, H. (1981): "Stichprobenweise Auswertung von Luftaufnahmen für die Erneuerung der Eidsgenössischen Arealstatistik". Arbeitsdokumente für die schweizerische Statistik/Heft 5, Bern 1981.

Matern, B. (1960): "Spatial variation: Stochastic Models and their Application to some Problems in Forest Surveys and other Sampling Investigations". Meddelanden från Staten Skogforskningsinstitut, 49, 1-144.

Matheron, G. (1971): "The Theory of Regionalized Variables and its Applications", Ecole Nationale Superieure des Mines de Paris.

Milne, A. (1959): "The Centric Systematic Area-Sample treated as a Random Sample". Biometrics, June 1959.

Omre, H. (1980): "Romlig avhengighet i arealbruken i tettsteder - vurdering av ulike estimatorer". The Norwegian Computing Center. Oslo.

Ripley, B.D. (1981): "Spatial Statistics". John Wiley & Sons, New York.

Zöhrer, F. (1978): "On the Precision of Dot Grid Estimates". Resource Inventory Notes, USDII Bureau of Land Management, Denver.

Utkommet i serien Artikler fra Statistisk Sentralbyrå (ART) - ISSN 0085-431X

Issued in the series Articles from the Central Bureau of Statistics (ART)

Nr. 134 Lars Østby: Norwegian Fertility Survey 1977 A Summary of Findings Fruktbarhetsundersøkelsen 1977 Et sammendrag av resultatene 1982 22 s. kr 10,00 ISBN 82-537-1710-5

" 135 Bjørg Moen og Per Sevaldson: Fødselstallene i Norge 1950 - 1975 Endringsfaktorer Births in Norway 1950 - 1975 Components of Change 1982 81 s. kr 15,00 ISBN 82-537-1716-4

" 136 Petter Frenger, Eilev S. Jansen og Morten Reymert: MODEX - En modell for verdenshandelen og norsk eksport av bearbeidde industrivarer A Model of World Trade and the Norwegian Export of Manufactured Goods 1982 35 s. kr 10,00 ISBN 82-537-1717-2

" 137 Petter Jakob Bjerve: Three Decades of the Conference of European Statisticians: Past Achievements and Perspectives for the Future Den europeiske sjefsstatistikarkonferansen gjennom tre ti-år: Mål som vart nådd og perspektiv for framtida 1982 22 s. kr 10,00 ISBN 82-537-0000-0

" 138 Arne Rideng: Befolkningsutviklingen i Norge fram til år 2025 The Changes of Population in Norway to 2025 1982 83 s. kr 15,00 ISBN 82-537-1852-7

" 139 Nils Terje Furunes and Svein Lasse Røgeberg: Compilation of Input-Output Tables in Norway Beregning og sammenstilling av kryssløpstabeller i Norge 1983 37 s. kr 12,00 ISBN 82-537-1885-3

" 140 Turid Noack og Lars Østby: Barnløshet - savn eller ønske? Om ufrivillig og frivillig barnløshet Childless or Childfree? About Infecundity and Intentional Childlessness 1983 50 s. kr 12,00 ISBN 82-537-1894-2

" 141 Olav Bjerkholt: Experiences in Using Input-Output Techniques for Price Calculations Bruk av kryssløpsmodeller for prisberegninger 1983 22 s. kr 12,00 ISBN 82-537-1926-4

" 142 Finn R. Førsund and Eilev S. Jansen: The Interplay between Sectoral Models Based on Micro Data and Models for the National Economy Samspillet mellom sektormodeller basert på mikrodata og modeller for økonomien som helhet 1983 22 s. kr 12,00 ISBN 82-537-1934-5

" 143 Svein Longva and Øystein Olsen: Price Sensitivity of Energy Demand in Norwegian Industries Prisfølsomheten i energietterspørselen i norske næringssektorer 1983 25 s. kr 12,00 ISBN 82-537-1960-4

" 144 Hans Viggo Sæbø: Land Use and Environmental Statistics Obtained by Point Sampling Areal- og miljøstatistikk utarbeidet ved hjelp av punktutvalg 1983 31 s. kr 12,00 ISBN 82-537-1966-3

Fullstendig oversikt over tidligere nummer av serien Artikler finnes i nr. 140.

Publikasjonen utgis i kommisjon hos
H. Aschehoug & Co. og Universitetsforlaget, Oslo,
og er til salgs hos alle bokhandlere
Pris kr 12,00

Omslag trykt hos Grøndahl & Søn Trykkeri, Oslo

ISBN 82-537-1966-3
ISSN 0085-431x