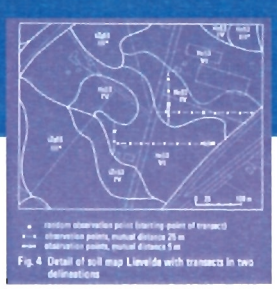
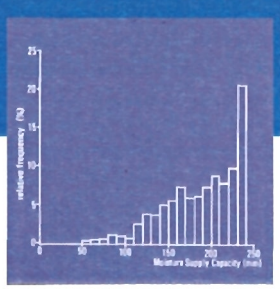
ALTERRA

Sampling for spatial inventory and monitoring of natural resources

J.J. de Gruijter





Alterra-rapport 070, ISSN 1566-7197

WAGENINGENUR

Sampling for spatial inventory and monitoring of natural resources

Sampling for spatial inventory and monitoring of natural resources

J.J. de Gruijter

Alterra-rapport 070

Alterra, Green World Research, Wageningen, 2000

ABSTRACT

J.J. de Gruijter, 2000. *Sampling for spatial inventory and monitoring of natural resources.* Wageningen, Alterra, Green World Research. Alterra-rapport 070. 90 pp 16 fig.; 91 ref.

A broad review is presented of statistical methodology for spatial inventory and monitoring of natural resources. The report deals with sampling design and analysis of sample data, and is intended to help researchers with developing good sampling schemes and monitoring systems. Emphasis is on the methodology for point sampling in space, in time or in space-time. Some basic knowledge of statistics is presupposed, but statistical theory is only dealt with as far as necessary for understanding of the methods and sound applications. Technical aspects of taking samples and measurements in the field are outside the scope of this report.

Keywords: ground water, inventory, monitoring, natural resources, Sampling, soil vegetation, statistics, survey,

ISSN 1566-7197

This report can be ordered by paying 40,00 Dutch guilders into bank account number 36 70 54 612 in the name of Alterra, Wageningen, the Netherlands, with reference to Rapport 070. This amount is inclusive of VAT and postage.

P.O. Box 47, NL-6700 AA Wageningen (The Netherlands).
Phone: +31 317 474700; fax: +31 317 419000; e-mail: postkamer@alterra.wag-ur.nl

Alterra is the amalgamation of the Institute for Forestry and Nature Research (IBN) and the Winand Staring Centre for Integrated Land, Soil and Water Research (SC). The merger took place on 1 January 2000.

Project 328-10161 [Alterra-rapport 070/IS/06-2000]

Contents

Sampling for spatial inventory and monitoring of natural resources

J.J. de Gruijter

June 2000

Preface

This is a broad review of statistical methods for spatial inventory and monitoring of natural resources. The report deals with sampling design and analysis of sample data, and is intended to help researchers in developing good sampling schemes and monitoring systems. The emphasis is on the methodology for point sampling in space, in time or in space-time. Although the users of this report are supposed to have some basic knowledge of statistics, statistical theory is only dealt with as far as necessary for understanding of the methods and sound applications. Technical and logistical aspects of taking samples and measurements in the field are outside the scope of this report.

This report is based on two sources of knowledge. First, my own expertise gained from statistical consultancy on sampling of natural resources in research projects of the Alterra institute and its ancestors. This relates mostly to soil, ground water and vegetation inventory. Second, a recent literature study on statistical methodology of monitoring of natural resources.

I thank Hans Bronswijk of R.I.V.M. for a fruitful discussion in the initial stage of this work, and my colleagues Mirjam Hack-ten Broeke, Martin Knotters, and Peter Finke for their useful comments on a draft version.

This report is the result of Project 81360 'Monitoring methodiek', in Research Program 328 'Geodata Groene Ruimte' of the Netherlands Ministry of Agriculture, Nature Management and Fisheries.

Chapter 1

Introduction

The aim of this report is to present and explain to researchers the statistical knowledge and methodology of sampling and associated data analysis that is useful for spatial inventory and monitoring of natural resources. Therefore, I intend to omit all theory not essential for applications or for basic understanding. Where possible I refer to the sampling and monitoring literature for specific topics. However, in one way this presentation is broader than standard statistical texts, where I pay much attention to how statistical methodology can be employed and embedded in real-life spatial inventory and monitoring projects. Thus I discuss in detail how efficient sampling schemes (Chapter 2) and monitoring systems (Chapter 9) can be designed in view of the aims and constraints of the project. The scope of this report is limited to natural resources that are part of the biosphere, hydrosphere or pedosphere. Geologic inventories as well as meteorologic monitoring are excluded as these are highly specialized topics. The spatial scale varies from a single agricultural field, as in precision agriculture, to continental, as in monitoring water quality of large rivers. The temporal scale varies from about one day, as may be the case in inventories, to many decades, as in long-term monitoring, e.g. of sea-water levels. The ultimate reasons for inventory and monitoring as discussed here are human health, rational production (food, fiber, timber) or environmental concern.

From a statistical point of view, the core of spatial inventory and monitoring is first of all *sampling*. For spatial inventory, sampling is done in space at one time. For monitoring it is done in time, either at a single location (temporal sampling), or at multiple locations (space-time sampling). Sampling is therefore the main focus of this report. The natural resources about which inventory or monitoring in a given application intends to give information are refered to as the 'universe of interest', or briefly the 'universe'. Universes in this context are biotic or a-biotic systems which generally vary in space and change with time. Some universes may, for the purpose of sampling, be considered as a physical continuum, e.g. the soil in a region at some time, the water of a river

passing a point during a period, the crop in an agricultural field at some time[1], or the natural vegetation in an area during a period[2]. The dimensionality of continuous universes can be 1, 2 or 3 for space, 1 for time, or combinations thereof. Other universes are populations of discrete, natural entities, such as the trees in a forest.

I use the term 'sampling' in the usual broad sense of selecting parts from a universe with the purpose of taking measurements from them. The selected parts may be measured *in situ*, they need not necessarily be taken out physically from the universe. The collection of selected parts is refered to as the 'sample'. To avoid confusion, a single selected part is called a 'sample element'. As a reasonable simplification, sample elements from continuous universes may be considered as points, lines, planes or (hyper-)volumes. This report focusses on point sampling because that is most relevant for practical applications.

An important sampling issue dealt with in this report is the difference and the choice between the design-based and the model-based approach to sampling (Chapter 3), as much confusion around this issue still exists. Because the choice between these two sampling approaches has major consequences, I use the distinction between them for a primary sub-division of the sampling strategies. More text is devoted to design-based strategies than to model-based strategies, not because they are more important but because there are more of them, and reference to existing literature is often problematic. In the general, statistical sampling literature design-based strategies are mostly presented in a non-spatial finite population framework, and 'translation' into the space-time context is in order.

The structure of this report is as follows. Part I deals with general issues of sampling, relevant to inventory and monitoring:

- Chapter 2: principles, practical, scientific and statistical issues in designing sampling schemes;
- Chapter 3: differences and choice between the design-based and the model-based approach to sampling;
- Chapter 4: possibilities and limitations of composite or bulk sampling.

Part II treats spatial inventory, i.e. spatial sampling at one time:

- Chapter 5 and 6: design-based and model-based strategies for sampling in space.

Part III treats monitoring, i.e. sampling in time or in space-time:

- Chapter 8: central issues in monitoring;

[1] A 'continuum view' at crop is appropriate if interest lies in crop properties per areal unit of the field. However, if interest lies in properties per plant, then the universe is to be considered as a discrete population.

[2] A monitoring period may or may not have a pre-planned end.

- Chapter 9: designing a monitoring system;
- Chapter 10: design-based and model-based strategies for sampling in time;
- Capter 11: design-based and model-based strategies for sampling in space-time.

Part I

General issues in sampling

Chapter 2

Designing a sampling scheme

2.1 Towards good design of sampling schemes

Inventory and monitoring of natural resources is often performed through the following activities:

- Planning of field activities: given the purpose of the project, the budget and possible logistic constraints, it is decided how many, where, how and when samples and/or field measurements are to be taken;
- Field activities: taking samples and/or field measurements;
- Laboratory work: sample preparation and analyses;
- Data recording;
- Data processing;
- Reporting.

Roughly speaking, these activities can be thought of as the consecutive stages of an inventory project. Obviously, in the case of monitoring field and laboratory work, as well as data recording and processing are done in some cyclic or continuous fashion. The activities mentioned above may overlap in time, for instance, data recording and field work are often done simultaneously. Also, there can be switching back and forth between activities. For instance, if during data processing some deficiency is discovered, it may be needed to do additional field work. Laboratory work is optional, as in some cases the measurements are taken in the field.

The main purpose of this section is to argue that, although the above sequence of activities may seem logical and sufficient, it does not constitute good

sampling practice. This is because an essential element is missing at the beginning: the element of planning the whole chain of activities, including the statistical procedure of data processing. Careful planning of the entire project is a pre-requisite of good sampling practice and should precede any other activity. Researchers usually put enough effort and ingenuity in deciding how, where and when to take samples and measurements. That is not the problem. Very often, however, the ideas about how to analyze the data remain rather vague until the data are there and crisp decisions must be made about what to do with them. In that case, more likely than not, data analysis and data acquisition will not be properly tuned to each other. Due to this mismatch, the potential qualities that a data acquisition plan might have are not fully exploited, and sub-optimal results are obtained. One example is where a stratified random sample has been taken, but this sample is analyzed as if it were a simple random sample. Another example is where the data are to be analyzed by some form of kriging, but it appears that the variogram needed for this can not be reliably estimated from the data. Finally, a situation often encountered is where the conclusions to be drawn from the sample data can only be based on questionable assumptions because the sample was not properly randomized. These examples will become more clear in the next sections.

In conclusion, it is recommended that not only the data acquisition is planned but the entire project, with special attention for the tuning of data acquisition with data processing and vice versa. Proper planning of the entire project will most likely pay itself back by increased efficacy and efficiency. I wish to emphasize this principle by referring to this complete plan as the *sampling scheme*. This broad concept of sampling scheme covers much more than just how, where and when to sample and measure. The sampling scheme captures all the decisions and information pertinent to data acquisition, data recording and data processing :

a. Objective of the inventory or monitoring: universe of interest defined in space and/or time, target variables, target parameters;

b. Objective function or quality criterion used to quantify the suitability of candidate sampling schemes;

c. Constraints: financial, logistic, operational;

d. Prior or ancillary information: existing sample data, maps, GIS files or models used in the design process;

e. Method of taking samples: dimensions of sample elements and sampling devices;

f. Methods of determination: field measurements and/or laboratory analyses;

g. Sampling design: sample size or frequency and how the sample points in space and/or time are to be selected;

h. The actually selected sample points in space and/or time;

i. Protocols on data recording and field work;

j. Method(s) of statistical analysis;

k. Prediction of operational costs and accuracy of results.

If we look at designing a sampling scheme as a case of problem solving, then the items *a* through *d* above form the information from which we search for a solution, items *e* through *j* together constitute the selected solution, and the last item is an *ex-ante* evaluation of that solution.

Apart from tuning data acquisition to data processing and vice versa, there is a more general reason why the project should be planned as a whole rather than to optimize parts of it in isolation from each other: the consequences of a decision about a single issue, in terms of quality and costs, depend on the decisions taken on other issues. A simple example is where two methods of determination are available for the target variable: an inexpensive but inaccurate method and an expensive but accurate method. The choice between either has an effect on both the costs and the accuracy of the final result, and these effects depend on the sample size. Given a fixed budget, choosing the inexpensive method implies that a larger sample size can be afforded. Whether or not this leads to a better result depends on various factors. How to design a sampling scheme is dealt with in the next sections.

2.2 Guiding principle in designing sampling schemes

A safe way to a good sampling scheme is this principle: '*Start at the end, then reason backwards*'. This means that one should first determine precisely what information is needed. Only when the information need is defined it becomes useful to search for a sampling scheme that leads to that result in an efficient way. The reason for this is that different types of results generally ask for different sampling schemes. Although this is an extremely important fact in sampling, it does not seem to be always clearly realized.

For instance, if the spatial mean of a region must be estimated, other, less expensive sampling schemes are needed than for local estimation at points, as for mapping. Another example is that data needs for generating hypotheses are totally different from those of testing hypotheses. The same is true for estimation of model parameters, for instance of variograms, compared with model validation.

Types of results can be divided into three broad groups. Firstly, the purpose of sampling may be estimation of the *frequency distribution* of a variable, or one or more parameters of that distribution. Examples are 'location' parameters such as the mean, quantiles (e.g. the median) and the mode, or 'dispersion' parameters such as the standard deviation, the range and tolerance intervals. These results are related to the universe as a whole; they have no geographical or time coordinates. Secondly, the purpose may be some kind of description of the *spatial and/or temporal distribution* of the variable within the universe.

Examples are: prediction of values at points within the universe, estimation of means within parts of the universe, or construction of contour maps. As opposed to the first group, these results contain geographical and/or temporal coordinates. Thirdly, there is a miscellaneous group of *special purposes* such as estimation of model parameters, model validation, generating hypotheses and multivariate statistics, including classification.

In principle, different types of results ask for different sampling schemes, because a given system may not yield the type of result that is required, or if it does, it may do so in an inefficient way. In conclusion, a good way of designing a sampling scheme is by reasoning backward through the following steps:

1. Decide precisely what type of result is needed, for instance, a map of a given variable, at a given scale and with a given accuracy. Or testing of a given hypothesis, at a given significance level and with a given statistical power.
2. Determine what kind of data analysis leads to that result.
3. Determine what the data needs are for this analysis.
4. Search for a sampling scheme to get those data in the most efficient way.

To aid the search for a good sampling scheme some practical, scientific as well as statistical issues are discussed in the following sections.

2.3 Practical issues

Avoid undue complexity

Researchers often know much about the processes that have generated the spatial pattern of soil or groundwater properties in the universe of interest. They may be tempted to express all this knowledge in detail in the form of a highly complex sampling design. Albeit understandable, this attitude entails two risks which are easily underestimated. Firstly, due to unforeseen operational difficulties during field work, it may prove impossible to carry out the design in all its complexity. The field work must then be adjourned until the design is re-adjusted. This may be time consuming and is likely to cause unwanted delay. Secondly, the complexities are introduced to increase the efficiency, but they may make the statistical analysis much more intricate and time consuming than expected. It is therefore usually wise to avoid highly complex sampling designs, because the theoretical gain in efficiency compared with simpler solutions is easily overridden by practical difficulties.

Allow for unexpected delay in field work

Even if you are familiar with the circumstances in the terrain, there can be factors beyond your control that prevent the field work to be completed within the

available time. Clearly, unfinished field work may seriously harm the statistical potential of the design. It is therefore prudent to allocate spare time in the system for contretemps, say 20 % of the total time for field work, and to include a number of optional sample points to be visited as far as spare time allows.

Include a test phase if necessary

If you are uncertain about the logistics of the field work or the spatial or temporal variability, a preliminary test phase is always worth the extra effort. The information that you get from even a small sample collected in advance will enable you to optimize the main sample more precisely and it will reduce the risk that the project will not meet its goal at all. In the final statistical analysis you can combine the test sample data with the main sample data, so the additional effort is limited to extra travel time and statistical analysis.

Evaluate the system beforehand

It is good practice to quantitatively predict the cost of operation of the system and the accuracy of the result, prior to the field work. Predicting cost and accuracy can be done in sophisticated ways, using mathematical models (Domburg e.a., 1994), or more globally, using experiences from similar projects, rules-of-thumb and approximations. A test phase will of course improve the prediction of cost and accuracy.

Explicit evaluation *ex ante* in terms of cost and accuracy is not only a final check of whether the system can be trusted to lead to the goal, it also enables comparison with evaluation *ex post*, i.e. after the project is finished. If this reveals significant discrepancies, the causes should be analyzed. This may provide a ground for better planning of future projects.

2.4 Scientific issues

Protocol for field work

Rules for field work will usually concern the physical act of taking samples and/or measurements in the field, but they should also tell what to do if a sample point is inaccessible or if it falls outside the universe. An example of the latter in soil sampling is where, on inspection in the field, it turns out that at the given point there is no 'soil' according to a given definition.

A poor protocol may seriously affect the quality of the results. Obvious requirements for a protocol are that it is complete, unambiguous, practically feasible and scientifically sound. The scientific aspect plays a role, for instance, when a rule says that an inaccessible sampling point is to be shifted to a nearby location in a certain way. In principle this leads to over-representation of boundary zones and, depending on the kind of design and the statistical analysis, this may result in biased estimates.

Protocol for data recording

As for field work, there should be sound rules for data recording. These rules should not only cover regular recording but also prescribe different codes for when a sampling point falls outside the universe, for when it is inaccessible, for when a variable cannot be measured because it is too large or too small ('censoring' in the statistical sense), and for when a variable cannot be measured for other reasons.

2.5 Statistical issues

Prior information on spatial and temporal variability

All prior information about the variability in the universe should be employed in the search for an efficient sampling design. Examples of prior information are satellite images, aerial photographs, thematic maps (e.g. groundwater, soil or vegetation maps) and theory about the spatial and/or temporal patterns of variation in the universe. This theory may be available in a verbal, qualitative form, or in a quantitative form of a mathematical model.

There are many ways in which prior information can be exploited in sampling schemes. Two modes can be distinghuished. The first mode is to use the prior information in the sampling design, i.e. in the data acquisition stage. The second mode is to use it in the statistical analysis of the sample data, i.e. in the data processing stage. In the following I give examples of each mode.

An example of the first mode is when images, photographs or maps are used for stratification of the universe. In that case the universe is split into a number of relatively homogeneous sub-universes (called 'strata'), which are then sampled independently from each other (Section 5.4). Another example is when genetic theory enables intelligent guesses about the spatial and/or temporal correlation. For instance, in the case of a universe consisting of the soil in a given area, eolian deposition of parent material in that area may be known to have resulted in little short-range variation of texture. Then, if the target variable is closely related to texture, it will be important for efficiency to avoid sampling at points close to each other. Finally, an example of the first mode is when a variogram is used to optimize the sample density or sample frequency.

An example of the second mode is when a variogram is used to optimize the weights of sample data for linear prediction as in Ordinary Kriging. Another example is the use of ancillary data in regression estimators (Section 5.8.3).

If prior information on the variability is captured in the form of variograms or correlograms, these functions can be used to predict the sampling variance for a given design (Section 5.9 and Chapter 6). If in addition a model for the costs is available then it is possible to optimize the sampling design in a fully quantitative way (Domburg e.a., 1997).

Modes of sample point selection

Three possible modes of sample point selection can be distinguished: *convenience* sampling, *purposive* sampling and *probability* sampling. The concept of convenience sampling is self-explanatory. An obvious example is when sampling is limited to road sides or other easily accessible spots. The advantage of this mode is that it saves time and cost. The disadvantage is that the statistical properties are inferior compared to the other modes. For instance, estimates from a convenience sample have to be considered as biased unless one is willing to accept specific assumptions about the sampling process and the spatial and temporal variation. These assumptions are often debatable, and this may or may not be acceptable, depending on the context of the project.

Purposive sampling tries to select the sample points such that a given purpose is served best. An example is the 'free survey' method of mapping soil classes, whereby the surveyor locates the sample points where they are expected to be most informative with respect to soil class delineation. In this example the points are selected in a subjective manner, using experience, visible landscape features and pedogenetic hypotheses. However, purposive sampling may also proceed by formally optimizing an objective function related to the purpose. For instance, if the purpose is to map a spatial distribution by kriging and if geographical boundary effects are disregarded, then it can be shown that the prediction error is minimized by a hexagonal grid of sample points, under assumptions of stationarity and isotropy (McBratney e.a., 1981). If boundary effects cannot be neglected, or if point data are available prior to sampling, then the grid that minimizes the prediction error will be irregular, and this can be found by simulated annealing (van Groenigen & Stein, 1998). Both methods are discussed in Chapter 6.

Probability sampling, unlike the other modes, selects sample points at random locations. Therefore the probabilities of selecting the points are known, and these probabilities provide the basis for statistical analysis of the data. As explained in Chapter 5, there are many techniques for random selection of sampling points. Collectively, this approach to sampling is referred to as the design-based approach, as opposed to the model-based approach, where the sample points are fixed instead of random and statistical analysis is based on a model of the variation in the universe. The choice between these two approaches is an important statistical issue, which is separately dealt with in Chapter 3.

Sources of error

It is important to realize that the accuracy of the final result is not only determined by sampling error, i.e. the error due to the fact that sampling is limited to a finite number of points. Other sources of error are: sample treatment, measurement and 'non-response': a term used in the general statistical literature to indicate the situation where for some reason no data can be obtained from a sample element. In groundwater, soil and vegetation sampling this occurs when a point in the field cannot be visited or when measurement is impossible for

other reasons.

Although any reduction of the sampling error will lead to a smaller total error, there is little point in putting all effort in further reduction of the sampling error if another source of error still has a higher order of magnitude. Therefore, in devising a sampling scheme, the relative importance of all error sources should be taken into consideration.

Chapter 3

Design-based versus model-based approach to sampling

There are two fundamentally different approaches to sampling: the *design-based* approach, followed in classical survey sampling, and the *model-based* approach, followed in geostatistics. Differences and relationships between these two approaches are extensively treated in Särndal e.a., (1992). De Gruijter & ter Braak (1990) discuss the issue in the spatial context, but the distinction holds and is equally relevant for sampling in time and in space-time. The difference between the two approaches is illustrated in Fig. 3.1 with a simple example, taken from Brus & de Gruijter (1997): a square area is sampled at 25 points and a 0/1 variable measured to estimate the fraction of the area with value 1. Fig. 3.1A shows a spatial distribution of the 0/1 variable and a configuration of 25 sample points. Averaging the observed values at these points yields an estimate of the fraction.

Now both approaches quantify the uncertainty of such an estimate by considering what would happen if sampling were repeated many times in a hypothetical experiment. Obviously, if in this experiment neither the pattern of values nor the locations of the sample points were changed there would be no variation, so one or the other has to be varied. The two approaches differ in which of the two is varied. The design-based approach evaluates the uncertainty by repeated sampling with different sets of sample points, while considering the pattern of values in the area as unknown but fixed. The sets of sample locations are generated according to a chosen random sampling design. The *row* of figures (A, B and C) represents three possible outcomes. As opposed to this, the model-based approach evaluates the uncertainty by repeated sampling with a fixed set of sample points, while varying the pattern of values in the area according to a chosen random model of the spatial variation. For this approach the *column* of figures (A, D and E) represents three possible outcomes. (Note that the target

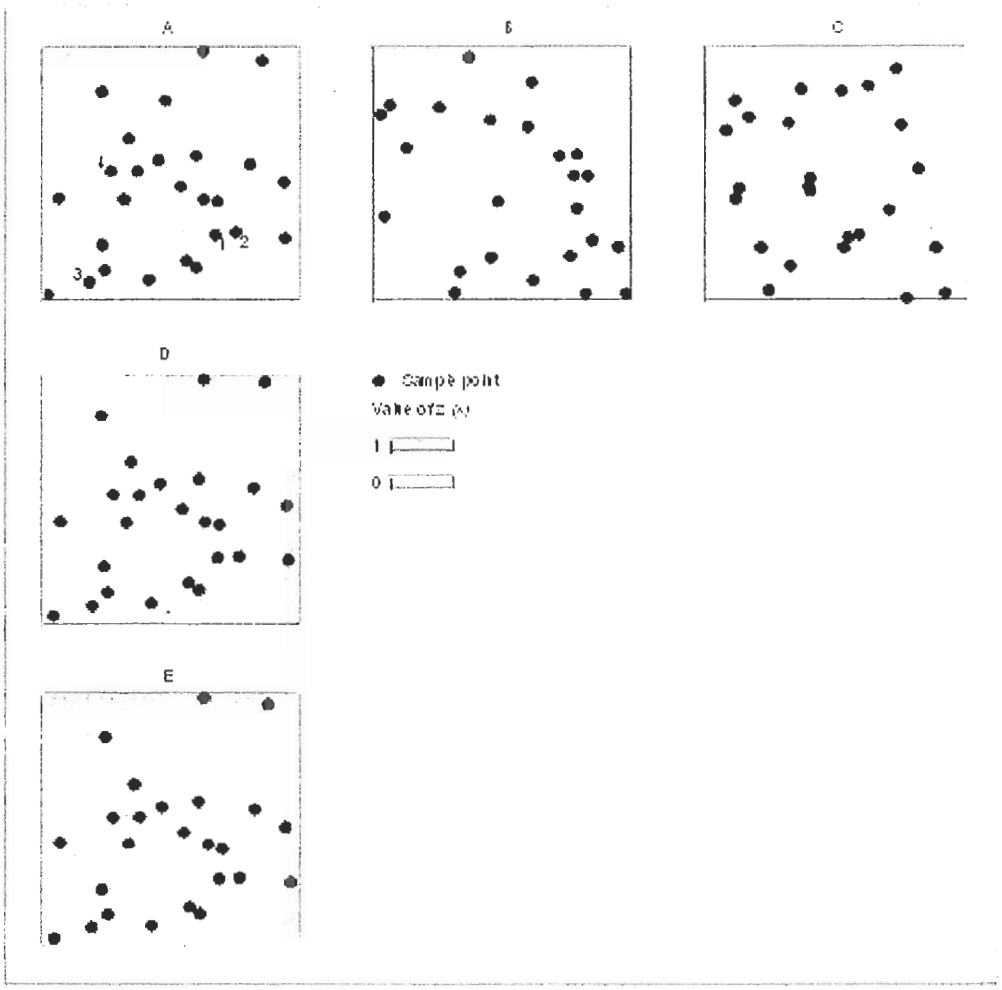


Figure 3.1: Repeated sampling in the design-based approach (A, B, C) and in the model-based approach (A, D, E). In the design-based approach the population is fixed and the sampling points are random. In the model-based approach the sampling points are fixed and the population is random. (From Brus & de Gruijter, 1997)

quantity in this approach is no longer constant: it varies among realizations from the model. The standard statistical terminology therefore speaks of prediction instead of estimation.)

The experiment can remain truly hypothetical in most instances because probability calculus enables to determine what happens on average over all possible realizations. In more intricate situations this is infeasible, however, and repeated sampling has to be simulated numerically, either varying the sample points or the pattern of values, or both.

The fact that the two approaches use a different source of randomness has several important practical as well as theoretical consequences. The main consequence is that the statistical inference from the sample data is entirely different. In the design-based approach estimation, testing and prediction are based on the selection probabilities as determined by the random design. This means that in calculating weighted averages, the data are assigned weights determined by their selection probabilities, not by their co-ordinates in space and/or time. In the model-based approach, inference is based on a stochastic model of the variation in the universe. Here the weights of the data are determined by covariances, given by the model as a function of co-ordinates in space and/or time.

Before deciding on the details of a sampling design, a choice between the design-based and the model-based approach should be made. It goes beyond the scope of this book to discuss this issue in detail, only an outline is given. An extensive discussion is presented in Brus & de Gruijter (1997). The 'ideal' circumstances for application of the design-based approach are as follows.

i. The required result is an estimate of the frequency distribution of the target variable in the universe as a whole, or a parameter of this distribution, such as the mean, the standard deviation or a quantile.

ii. A minimum sample size of, say, 5 or 10 points can be afforded, depending on the variation in the universe, to have at least a rough idea of the sampling error.

iii. It is practically feasible to sample at randomly selected locations and/or times.

iv. It is important to obtain an estimate that is unbiased in the sense that, averaged over all possible samples of the applied design, the estimate equals the true value of the target parameter.

v. It is important to obtain an objective assessment of the uncertainty of the estimate.

Around this 'ideal' there is a range of circumstances in which the design-based approach is still preferable to the model-based approach.

The 'ideal' circumstances for application of the model-based approach are as follows.

i. The required result is prediction of values at individual points, as with forecasting, or distribution of values over the entire universe, as with mapping.

ii. At least a medium sample size can be afforded, depending on the spatial/temporal variation. The model usually implies stationarity assumptions and a variogram. Accurate estimation of the variogram requires a sufficiently large numebr of observations. Following Webster & Oliver (1992) this number should be at least 100 to 150.

iii. A reliable model of the variation is available.

iv. High auto-correlations exist in the universe.

As before, around this 'ideal' there is a range of circumstances in which the model-based approach is still preferable to the design-based approach. A typical intermediate situation is where averages are required for a number of sub-universes or 'blocks', in which only sparse sampling can be done. Brus and De Gruijter (1997) explore this in a case study.

Chapter 4

Composite sampling

Composite sampling is the technique of putting the material of individual samples together and to mix and analyse this. As only the composite samples are analysed, the number of analyses is strongly reduced. The technique is often used in soil sampling because of its great advantage in saving laboratory costs. A vast amount of literature exists on this subject, both theoretical and applied, but a well-founded and generally applicable methodology of composite sampling does not seem to be available. Therefore, I merely give some indications on when and how to apply composite sampling, and I do this by discussing the assumptions that underly compositing.

The basic assumption in its most general form is that analysing a composite gives the same result as *analysing the individual samples* used to form the composite. Two special cases are mentioned here. The first case is where interest lies in the presence or absence of a qualitative variable, for instance, a species of soil microbe or a chemical substance. If the method used to determine presence/absence has a detection limit that is low enough, then a composite sample could be analysed instead of individual samples separately. This case is often referred to as group screening or group testing.

The second special case, more relevant to inventory and monitoring of natural resources, is where interest lies in the average value of a quantitative variable, for instance, phosphate content in the topsoil. Here the assumption is that analysing a composite gives the same result as *averageing the values measured on individual samples.* In other words: arithmatic averageing can be replaced by physical averageing. Of course, pre-assumptions are that averageing is meaningful, and that it is needed given the purpose of the project. I discuss these assumptions briefly.

Averaging of values is meaningful

This requires that the target *variable* is a quantitative property, which precludes composite sampling when the target variable is measured on a nominal or an ordinal scale.

Averaging of values is needed

Taking a non-composite sampling scheme as point of departure, this pre-assumption implies that, without compositing, the estimator of the target *quantity* would be a function of one or more arithmatic means of individual sample values. The simplest example of such estimators is the unweighted sample mean, as for instance used with Simple Random Sampling (Section 5.3), Systematic Sampling (Section 5.7), Systematic Unaligned Sampling or Markov Chain Sampling (Section 5.8.2). In these cases all individual samples could in principle be put together into one composite. Other examples, involving multiple arithmatic means, are the estimators used with Stratified Sampling (Section 5.4), Two-Stage Sampling (Section 5.5) and Cluster Sampling (5.6).In these cases all individual samples belonging to the same stratum, primary sampling unit or cluster could in principle be put together.

This requirement precludes compositing when the purpose is to estimate, for instance, a measure of dispersion (e.g. standard deviation or range), an extreme value, a quantile, or values at unsampled points via kriging. In these cases the estimators are not arithmatic means of sample values.

Arithmatic averageing can be replaced by physical averageing

In order to make this basic assumption valid, three requirements must be met. Firstly, the target variable must be directly measured in the samples, or be defined as a linear transformation of one or more measured variables. Otherwise, if the target variable is a non-linear transformation of one or more measured variables, the transformation of the mean value(s) from a composite sample is not equal to the mean of the transformed values from individual samples. Neglecting this fact can lead to an unacceptable systematic error. Examples of a target variable defined as a non-linear transformation are: the indicator variable indicating whether or not the phosphate content in the topsoil exceeds a given threshold, available soil moisture content calculated with a non-linear model from inputs measured at sample points, and pH as a logarithmic transformation of hydrogen activity.

Secondly, after putting individual samples together and mixing, no physical, chemical or biological interactions between the increments should take place that influences the value of the target variable. This precludes, for instance, compositing when the target variable depends on pH and some samples contain calcium carbonate while others don't. As above, compositing in such cases could easily lead to an unacceptable systematic error.

Thirdly, compositing reduces laboratory costs, but it introduces two inter-related sources of error: error by imperfect mixing of the composites and error by sub-sampling the mixed composite. Also, random measurement errors will cancel out less well in the case of composite sampling than with non-composite sampling, because fewer measured values are averaged. The additional error due to compositing should not enlarge the total error too much, and this puts a limit to the number of individual samples that can bulked. The increase of the contribution of measurement error to the total error could be counter-acted by taking multiple measurements from each composite while still preserving a cost advantage. Also, if mixing and sub-sampling are important error sources, one

could make a number of smaller composites from random subsets of individual samples, instead of one large composite.

Some influential theoretical publications on composite sampling are Duncan (1962), Brown and Fisher (1972), Rohde (1976) and Elder, Thompson and Myers (1980). Boswell et al. (1996) provide an annotated bibliography. Papers on composite soil sampling are e.g.: Baker et al. (1981), Brus et al. (1999), Cameron et al. (1971), Carter and Lowe (1986), Courtin et al. (1983), Onate (1953), Ruark et al. (1982), Reed and Rigney (1947), Webster and Burgess (1984) and Williams et al. (1989).

Part II

Inventory: spatial sampling at one time

Chapter 5

Design-based sampling in 2-dimensional space

5.1 Introduction

This chapter discusses how design-based sampling strategies work and how they can be applied in monitoring projects. The aim is to help understanding the basic principles at an intuitive level, the text is not meant as an exposé of sampling theory. A somewhat practically oriented textbook on design-based sampling strategies is Cochran (1977), from which most of the material presented here is borrowed or derived. A comprehensive textbook on sampling theory is Särndal e.a. (1992).

The general pattern in the development of sampling strategies is to take the simplest strategy (Simple Random Sampling, Section 5.3) as a starting point, with complete random selection of all sample points. Then restrictions on randomization are looked for, such that this would reduce the sampling variance or the cost of operation, or both. Different types of restrictions can be distinguished, each giving rise to a different type of sampling design.

Before discussing the basic designs, the statistical concept of *sampling design* itself need to be defined more precisely. In the spatial context it is defined as a function that assigns a probability of selection to any set of points in the area. For instance, the sampling design for Simple Random Sampling with sample size 25 assigns equal selection probabilities to every possible *set* of 25 points in the area and zero probability to any other set. (Note that a design assigns probabilities to sets of points, not to individual points.) A *sampling strategy* is defined as a combination (p, t) of a sampling design (p) and an estimator (t) for a given target parameter (T), such as the mean of the area. Statistical quality measures, like bias and variance, can only be defined and evaluated for these combinations, not for a design or an estimator on its own.

5.2 Scope of design-based strategies

A typical application of design-based strategies is to estimate the areal mean of a directly measured quantitative variable. However, the scope of these strategies is much wider than this. Extensions are possible in three directions: derived variables, other parameters, and smaller (sub-)areas.

Firstly, the target variable need neither be quantitative, nor directly measured. If the target variable is measured at a nominal or ordinal scale, then the sample data consist of class labels, and these can be analyzed statistically by first transforming them into 0/1 indicator variables. The presence or absence of a given class is thereby re-coded as 1 and 0, respectively. Of course, if there are k mutually exclusive classes, only $k-1$ indicator variables are needed. The mean of an indicator variable can be interpreted as the fraction of the area in which the class occurs.

Transformation into indicator variables can also be applied to quantitative variables in order to estimate the areal fraction in which the variable exceeds a given threshold. This technique can be extended to estimate the entire Spatial Cumulative Distribution Function (SCDF) of a quantitative variable. In that case areal fractions are estimated for a series of threshold values.

Apart from the simple 0/1 transformations, the target variable may be the output of a more or less complicated model for which the input data is collected at the sample points. Another important case of indirect determination is in validation studies, where the target variable represents an error, i.e. the difference between a measured value and a value predicted by a process model or a spatial distribution model, such as a thematic map. A common example is the error resulting from a classification algorithm applied to remotely sensed images. The errors determined at the sample points can be used to estimate their spatial mean (which equals the bias), the mean absolute error, the mean squared error, or the entire SCDF of the errors.

Secondly, the target parameter need not be the spatial mean. For instance, it may also be a quantile, such as the median, the spatial variance, a tolerance interval, or a parameter of a model relating one or more predictor variables with a variable of interest. See e.g. Krishnaiah & Rao (1988) and Patil & Rao (1994) for design-based statistical inference on these and other target parameters.

Thirdly, the region for which estimation or testing of hypotheses is demanded need not be the entire area sampled; interest may also be in one or more sub-areas. There are two different methods of estimation and testing in sub-areas. The first is to sample the sub-areas independently from each other, in which case they act as 'strata' in a stratified sampling design (Section 5.4). In the second method the sampling design is independent from any division into sub-areas. Estimation in sub-areas is then only based on sorting the sample data afterwards according to the sub-areas in which the sample points happen to fall. In this case the sub-areas are referred to as 'domains of interest', or briefly 'domains'.

The following sections describe each basic strategy by discussing the type of random selection restriction, a technique for selecting samples according to the

design, a simple example, the inference from sample data, the determination of sample sizes and advantages and disadvantages. I repeat from Chapter 3 that design-based statistical inferences such as given below are valid, regardless of the structure of the spatial variation, because they do not make any assumption about this structure.

5.3 Simple Random Sampling (SRS)

Restriction on random selection

No restrictions on random selection other than a previously chosen fixed sample size. All sample points are selected with equal probability and independently from each other.

Selection technique

An algorithm for SRS with sample size n, applicable to irregularly shaped areas, is as follows.

(a) Determine the minimum and maximum X and Y co-ordinates of the area: $X_{\min}$, $X_{\max}$, $Y_{\min}$ and $Y_{\max}$.

(b) Generate independently from each other two (pseudo-)random co-ordinates, X_{ran} and Y_{ran}, from the uniform distribution on the interval ($X_{\min}$, $X_{\max}$) and ($Y_{\min}$ and $Y_{\max}$), respectively.

(c) Determine with a point-in-polygon routine whether the point (X_{ran}, Y_{ran}) falls in the area. Accept the point if it does; skip the point if it does not.

(d) Repeat step (b) and (c) until n points are selected.

Example

Fig. 3.1A, B and C show three realizations of SRS with 25 points; Fig. 5.1 is an example with $n = 16$. Notice the irregularity, the clustering and the empty spaces in the configurations, which is typical of SRS.

Statistical inference

The spatial mean of the area, $\overline{Y}$, for a quantitative variable y is estimated by:

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

with n = sample size, and y_i = value of sample point i.

The strategy (SRS, $\overline{y}$) is 'p-unbiased'; this is a quality criterion defined as: $E_p[\overline{y}] = \overline{Y}$, where $E_p[.]$ denotes the statistical expectation over all possible sample realizations from a design p (in this case SRS). This means that if we

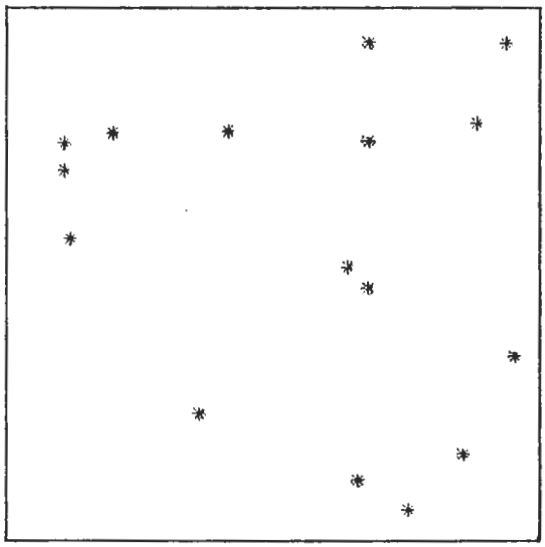


Figure 5.1: Notional example of Simple Random Sampling

would repeat sampling, measuring and calculating $\overline{y}$ in the same way again and again, we would find on average the true value $\overline{Y}$. If measurement errors are present, then the unbiasedness still holds if the errors are purely random, i.e. zero on average.

The variance of the variable y between points in the area is estimated by:

$$v(y) = \frac{1}{(n-1)} \sum_{i=1}^{n} (y_i - \overline{y})^2$$

and the standard deviation by $s(y) = \sqrt{v(y)}$. The sampling variance of the estimated mean $\overline{y}$ is estimated by:

$$v(\overline{y}) = \frac{v(y)}{n}$$

and its standard error by $s(\overline{y}) = \sqrt{v(\overline{y})}$

If the data contain random measurement errors, then their contribution to the total estimation error is automatically included in the estimates $v(\overline{y})$ and $s(\overline{y})$.

The two-sided 100(1-α)% confidence interval for $\overline{Y}$ is given by:

$$\overline{y} \pm t_{1-\alpha/2} \cdot s(\overline{y}) \tag{5.1}$$

where $t_{1-\alpha/2}$ is the $(1 - \frac{\alpha}{2})$ quantile of the Student distribution with $(n-1)$ degrees of freedom. This confidence interval is based on the assumption that y, and as a consequence $\overline{y}$, is normally distributed. If the distribution deviates

clearly from normality, the data should be first transformed to normality, for instance by taking the logarithm. The interval boundaries thus found are then back-transformed to the original scale. Transformation is not necessary if n is large, because then $\overline{y}$ is approximately normally distributed according to the Central Limit Theorem.

The above formulas for estimating means can also be used for areal *fractions*. The fraction of the area where a qualitative variable q has a given value, for instance 'very suitable', can be estimated by first generating a 0/1 indicator variable from the sample data, with value 1 if q= 'very suitable', and 0 otherwise. Then the above equations are simply applied to this indicator variable. The only exception is the calculation of confidence intervals because the indicator variable is clearly not normally distributed. The sample fraction has a Binomial distribution, and with small samples ($n < 20$) this distribution should be used to construct confidence intervals. With larger samples the distribution is close enough to normality and formula (5.1) will be accurate enough for most practical applications.

The above formulas can also be used for estimation in a domain (Section 5.2), if it contains sample points. A domain may or may not have a known geographical delineation. An example of the latter is where the spatial mean of a target variable within a given soil type is to be estimated, and no map of the soil types at an appropriate scale is available. This mean can be estimated if, in addition to the target variable, the soil type is recorded at the sample points.

The mean of a quantitative variable y in domain j, $\overline{Y}_j$, is simply estimated by averaging over the sample points that fall in this domain:

$$\widehat{\overline{Y}}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} y_{jk},$$

where n_j = number of sample points in domain j, and y_{jk} = value at point k in domain j.

Variances, standard deviations and confidence intervals are calculated in the same way as for the area. The same applies to estimation of fractions and SCDF's in domains.

Sample size

The sample size needed to estimate a mean such that, with a chosen large probability $1 - \alpha$, the relative error $\left|\frac{\overline{y}-\overline{Y}}{\overline{Y}}\right|$ is smaller than a chosen limit r, can be calculated by:

$$n = \left(\frac{u_{1-\alpha/2} \cdot S}{r\overline{Y}}\right)^2,$$

where $u_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution, and S is a prior estimate of the standard deviation of y in the area.

In this formula $\frac{S}{\overline{Y}}$ is the coefficient of variation of y in the area. Of course, this parameter is not known exactly beforehand. Instead, a prior estimate is

substituted, which can be obtained from a pilot or previous sampling in the same area, from sampling in a similar area, or from general knowledge of the spatial variation.

If instead of the relative error we wish the absolute error $|\overline{y} - \overline{Y}|$ to be smaller than a chosen limit d, we need sample size:

$$n = \frac{u_{1-\alpha/2}^2 \cdot S^2}{d^2}.$$

The sample size needed to estimate a fraction can be calculated in the same way as with a quantitative variable. In that case the prior estimate of the standard deviation of the corresponding indicator variable is derived from a prior estimate of the fraction P by: $S = \sqrt{P(1-P)}$.

Advantage

The simplicity of this type of design enables relatively simple and straightforward statistical analyses of the sample data, also with non-standard estimation and testing problems.

Disadvantages

(*i*) The sampling variance is usually larger than with most other types of design at the same cost, and (*ii*) because large empty spaces can occur between the sampling points, estimation in domains may be impossible.

5.4 Stratified Sampling (StS)

Restriction on random selection

The area is divided in sub-areas, called 'strata', in each of which SRS is applied with sample sizes chosen beforehand.

Selection technique

The algorithm for SRS is applied to each stratum separately.

Example

Fig. 5.2 shows an example with 16 square strata and 1 point in each stratum. Notice the more even spreading compared with SRS in Fig. 5.1.

Statistical inference

Means, areal fractions and SCDF's (after 0/1 transformation) of the area are estimated by:

$$\overline{y}_{\mathrm{St}} = \frac{1}{A} \sum_{h=1}^{L} A_h \cdot \overline{y}_h,$$

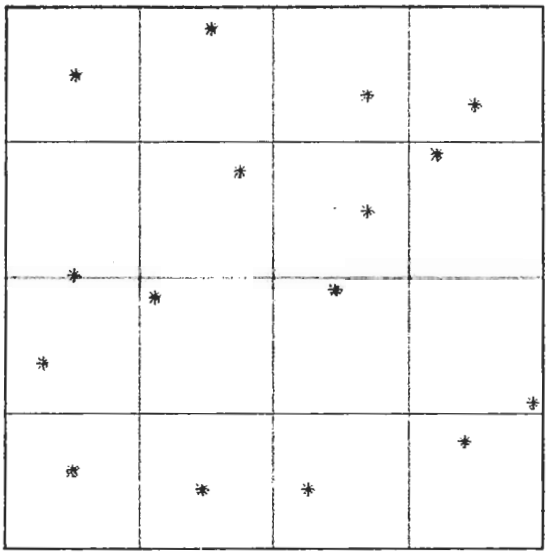


Figure 5.2: Notional example of Stratified Sampling

with L = number of strata; A_h = area of stratum h; A = total area, and $\overline{y}_h$ = sample mean of stratum h.

The strategy (StS, $\overline{y}_{\mathrm{St}}$) is p-unbiased. Provided all sample sizes are > 1, the variance of $\overline{y}_{\mathrm{St}}$ can be estimated by:

$$v(\overline{y}_{\mathrm{St}}) = \frac{1}{A^2} \sum_{h=1}^{L} A_h^2 \cdot v(\overline{y}_h),$$

where $v(\overline{y}_h)$ is the estimated variance of $\overline{y}_h$:

$$v(\overline{y}_h) = \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (y_{hi} - \overline{y}_h)^2,$$

with n_h = sample size in hth stratum.

The standard deviation is estimated by $s(\overline{y}_h) = \sqrt{v(\overline{y}_h)}$. Confidence intervals are calculated in the same way as with SRS, see Eq. 5.1.

The method of estimating means, fractions or SCDF's (after 0/1 transformation) in a *domain* depends on whether the areas of the domain within the strata are known. If they are, then the mean of the jth domain, $\overline{Y}_j$, is estimated by

$$\widehat{\overline{Y}}_j = \frac{1}{A_j} \sum_h A_{hj} \cdot \overline{y}_{hj} \tag{5.2}$$

with A_{hj} = area of domain j within stratum h; A_j = total area of domain j, and $\overline{y}_{hj}$ = sample mean of domain j within stratum h. The variance of $\widehat{\overline{Y}}_j$ is estimated by

$$v(\widehat{\overline{Y}}_j) = \frac{1}{A_j^2} \sum_h A_{hj}^2 \cdot v(\overline{y}_{hj}),$$

where

$$v(\overline{y}_{hj}) = \frac{1}{n_{hj}(n_{hj}-1)} \sum_{i=1}^{n_{hj}} (y_{hij} - \overline{y}_{hj})^2,$$

with n_{hj} = number of sample points falling in domain j within stratum h.

If the areas of the domain within the strata are not known, they have to be estimated from the sample. Unbiased estimates to be substituted in Eq. 5.2 are:

$$\widehat{A}_{hj} = A_h \cdot \frac{n_{hj}}{n_h} \quad \text{and} \quad \widehat{A}_j = \sum_h \widehat{A}_{hj}.$$

The variance is now larger, because of the error in the estimated areas. It is estimated by

$$v(\widehat{\overline{Y}}_j) = \frac{1}{\widehat{A}_j^2} \sum_h \frac{A_h^2}{n_h(n_h-1)} \left[\sum_i (y_{hij} - \overline{y}_{hj})^2 + n_h \left(1 - \frac{n_{hj}}{n_h}\right) \left(\overline{y}_{hj} - \widehat{\overline{Y}}_j\right)^2 \right].$$

Sample sizes

The sample sizes in the strata may be chosen to minimize the variance $V(\overline{y}_{\mathrm{st}})$ for a given maximum allowable cost, or to minimize the cost for a given maximum allowable variance. A simple linear cost function is:

$$C = c_o + \sum c_h n_h,$$

with c_o = overhead cost, and c_h = cost per sample point in stratum h.

If we adopt this function, the optimal *ratios* of the sample sizes to the total sample size n are:

$$\frac{n_h}{n} = \frac{A_h S_h / \sqrt{c_h}}{\sum \left(A_h S_h / \sqrt{c_h}\right)},$$

where the S_h are prior estimates of the standard deviations in the strata. This formula implies that a stratum gets a larger sample, if it is larger or more variable or less expensive to sample.

The total sample size affordable for a fixed cost C, given that optimal allocation to the strata is applied, is:

$$n = \frac{(C - c_o) \sum \left(A_h S_h / \sqrt{c_h}\right)}{\sum A_h S_h \sqrt{c_h}}.$$

The total sample size needed to keep the variance below a maximum value V_m, again presuming that optimal allocation to the strata is applied, is:

$$n = \frac{1}{V_m} \cdot \sum \left(W_h S_h \sqrt{c_h}\right) \cdot \sum W_h S_h / \sqrt{c_h},$$

where $W_h = A_h/A$. If the cost per point is equal for the strata, this reduces to:

$$n = \frac{1}{V_m} \cdot \left(\sum W_h S_h\right)^2 .$$

If, instead of V_m, an absolute error d has been specified with an allowed probability of exceeding α, then V_m can be derived from d and α, according to $V_m = d/u_{1-\alpha/2}$, where $u_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

When estimating areal fractions rather than means of quantitative variables, the above formulas for sample sizes can still be applied if S_h is replaced by $\sqrt{P_h(1-P_h)}$, where P_h is a prior estimate of the fraction in stratum h.

Advantages

There are two possible reasons for stratification. The first is that the efficiency as compared with SRS may be increased, i.e. smaller sampling variance at the same cost, or lower cost with the same variance. In this case the stratification is chosen such that the expected gain in efficiency is maximized. In practice this can be achieved by forming strata that are as homogeneous as possible. Also, if the cost per sample point varies strongly within the area, for instance with distance from roads, it is efficient to stratify accordingly and to sample the 'inexpensive' strata more densely. Another reason for stratification may be that separate estimates for given sub-areas are needed. If the strata coincide with these sub-areas of interest then, as opposed to SRS, one has control over the accuracy of the estimates by allocating sufficient sample sizes to the strata.

Disadvantage

With inappropriate stratification or sub-optimal allocation of sample sizes, there could be loss rather than gain in efficiency. This can occur if the stratum means differ little or if the sample sizes are strongly disproportional to the surface areas of the strata. If, for instance, one has many small strata with unequal area and a small sample in each, then these sample sizes are bound to be strongly disproportional because they must be integer numbers.

5.5 Two-stage Sampling (TsS)

Restriction on random selection

As with StS, the area is divided into a number of sub-areas. Sampling is then restricted to a number of randomly selected sub-areas, in this case called primary units. Note the difference with StS where all sub-areas (strata) are sampled. In large scale surveys this principle is often generalized to multistage sampling. (Three-stage crop sampling, for instance, could use sub-areas from RS images as primary units, fields as secondary units, and sample plots as tertiary units.)

Selection technique

A version is described by which the primary units (PU's) are selected with replacement and with probabilities proportional to their area. An algorithm to make n such selections from all N PU's in the area is as follows:

(a) Determine the areas of all PU's, $A_1, ..., A_N$, and their cumulative sums, $S_1, ..., S_N$, with $S_k = \sum_{i=1}^{k} A_i$.

(b) Generate a (pseudo-)random number x from the uniform distribution on the interval $(0, S_N)$.

(c) Select the PU of which the corresponding S_k is the first in the series that exceeds x.

(d) Repeat step (b) and (c) until n PU's are selected.

An alternative, sometimes more efficient algorithm works with a geographical representation of the area and its PU's:

(a) Select a random point in the area as in SRS.

(b) Determine with a point-in-polygon routine in which PU the point falls, and select this PU.

(c) Repeat step (b) and (c) until n selections have been made.

In the second stage, a pre-determined number of sample points, m_i, is selected within each of the PU's selected in the first stage. This is done in the same way as with SRS. If the geographical algorithm is applied, the random points used to select the PU's may also be used as sample points. If a PU has been selected more than once, an independent sample of points must be selected for each time the PU was selected.

Example

Fig. 5.3 shows four square PU's selected in the first stage, and four points in each in the second stage. Notice the stronger spatial clustering compared with SRS in Fig. 5.1. This is just a simple, notional example. It should be noted, however, that the PU's may be defined in any way that seems appropriate, and that the number of sample points may vary among units.

Statistical inference

Means, areal fractions and SCDF's (after 0/1 transformation) of the area are estimated by the remarkably simple estimator:

$$\overline{y}_{\mathrm{Ts}} = \frac{1}{n} \sum_{i=1}^{n} \overline{y}_i \tag{5.3}$$

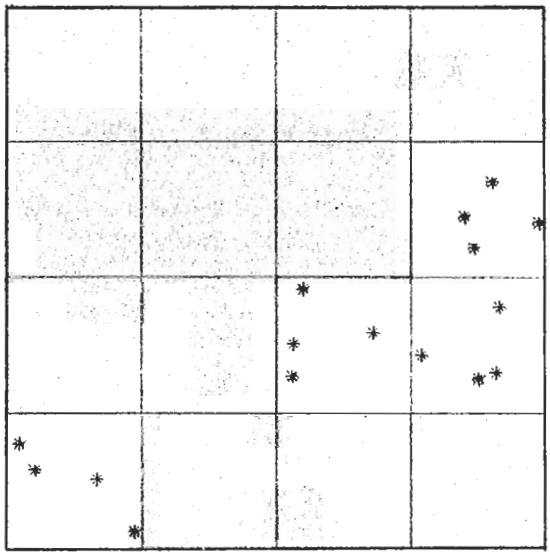


Figure 5.3: Notional example of Two-stage Sampling

with n = number of PU selections, and $\overline{y}_i$ = sample mean of the PU from selection i.

The strategy (TsS, $\overline{y}_{\mathrm{Ts}}$) is p-unbiased. The variance is simply estimated by:

$$v(\overline{y}_{\mathrm{Ts}}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\overline{y}_i - \overline{y}_{\mathrm{Ts}})^2.$$

Notice that neither the areas of the PU's, A_i, nor the secondary sample sizes m_i occur in these formulas. This simplicity is due to the fact that the PU's are selected with replacement and probabilities proportional to size. The effect of the secondary sample sizes on the variance is implicitly accounted for. (To understand this, consider that the larger m_i is, the less variable $\overline{y}_i$, and the smaller its contribution to the variance.)

The standard deviation is estimated by $s(\overline{y}_{\mathrm{Ts}}) = \sqrt{v(\overline{y}_{\mathrm{Ts}})}$. Confidence intervals are calculated in the same way as with SRS, see Eq. 5.1.

The method of estimating means, areal fractions and SCDF's in domains depends on whether the area of the domain, A_j, is known or not. If it is known, then the mean of the jth domain, $\overline{Y}_j$, is estimated by:

$$\widehat{\overline{Y}}_j = \frac{\widehat{Y}_j}{A_j} \tag{5.4}$$

where $\widehat{Y}_j$ is an estimate of the total (spatial integral) of variable y over domain j. To estimate this total, we first define a new variable y', which equals y everywhere in the domain, but is zero elsewhere. The total of y over domain

j equals the total of y' over the area, and this is estimated as A times the estimated mean of y', following Eq. 5.3:

$$\widehat{Y}_j = A \cdot \overline{y}'_{\mathrm{Ts}} = \frac{A}{n} \sum_{i=1}^{n} \overline{y}'_i,$$

where $\overline{y}'_i$ is the sample mean of the transformed variable y' from PU selection i. The variance of the domain mean is estimated by:

$$v(\widehat{\overline{Y}}_j) = \left(\frac{A}{A_j}\right)^2 \cdot \frac{1}{n(n-1)} \sum_{i=1}^{n} (\overline{y}'_i - \overline{y}'_{\mathrm{Ts}})^2.$$

If the area of the domain is not known, it has to be estimated from the sample. An unbiased estimate to be substituted for A_j in Eq. 5.4 is:

$$\widehat{A}_j = \frac{A}{n} \sum_{i=1}^{n} \frac{m_{ij}}{m_i},$$

with m_{ij} = number of points in PU selection i and domain j. Hence, the ratio estimator:

$$\widehat{\overline{Y}}_{\mathrm{R}j} = \frac{\widehat{Y}_j}{\widehat{A}_j} = \frac{\sum_{i=1}^{n} \overline{y}'_i}{\sum_{i=1}^{n} m_{ij}/m_i},$$

with estimated variance:

$$v(\widehat{\overline{Y}}_{\mathrm{R}j}) = \left(\frac{A}{\widehat{A}_j}\right)^2 \cdot \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\overline{y}'_i - \widehat{\overline{Y}}_{\mathrm{R}j} \frac{m_{ij}}{m_i}\right)^2.$$

Sample sizes

The primary and secondary samples sizes n and m_i can be optimally determined via dynamic programming, given a budget or variance requirement, any cost function and prior estimates of the within- and between-unit variances; see Domburg e.a. (1997). A simple approximation is by taking the m_i constant, say $m_i = m$. This is reasonable if the PU's have roughly the same area and internal variability. The variance of the mean is now

$$v(\overline{y}_{\mathrm{Ts}}) = \frac{1}{n}\left(S_{\mathrm{B}}^2 + \frac{1}{m} S_{\mathrm{W}}^2\right) \tag{5.5}$$

where S_{B}^2 and S_{W}^2 are the between-unit and the pooled within-unit variance, respectively. Given the linear cost function $C = c_1 n + c_2 nm$, the sample sizes minimizing the variance under the constraint that the cost does not exceed a budget C_m, can be found using the Lagrange multiplier method:

$$n = \frac{S_{\mathrm{B}} C_m}{S_{\mathrm{W}} \sqrt{c_1 c_2} + S_{\mathrm{B}} c_1}$$

and

$$m = \frac{S_W}{S_B}\sqrt{\frac{c_1}{c_2}}.$$

Conversely, minimizing the cost under the constraint that the variance does not exceed a maximum V_m:

$$n = \frac{1}{V_m}\left(S_W S_B\sqrt{\frac{c_2}{c_1}} + S_B^2\right)$$

and m as above.

If, instead of V_m, an absolute error d has been specified with an allowed probability of exceeding α, then V_m can be derived from d and α, according to $V_m = d/u_{1-\alpha/2}$, where $u_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

When estimating areal fractions rather than means of quantitative variables, the above formulas for sample sizes can still be applied if S_B is interpreted as a prior estimate of the standard deviation between the fractions in the units P_i, and S_W is replaced by a prior estimate of the square root of the average of $P_i(1-P_i)$ over the units.

Advantage

The spatial clustering of sample points created by TsS has the operational advantage of reducing the travel time between points in the field. Of course, the importance of this depends on the scale and the accessibility of the terrain. The advantage may be amplified by defining the PU's such that they reflect dominating accessibility features like roads and land ownerships.

Disadvantage

The spatial clustering generally leads to lower precision, given the sample size. However, the rationale is that due to the operational advantage a larger sample size can be afforded for the same budget, so that the initial loss of precision is outweighed.

5.6 Cluster Sampling (ClS)

Restriction on random selection

Pre-defined sets of points are selected, instead of individual points as in SRS, StS and TsS. These sets are referred to as 'clusters'.

Selection technique

In principle the number of clusters in the area is infinite, so it is impossible to create all clusters beforehand and to sample from this collection. However, only clusters which are selected need to be created, and selection of a cluster can take place via selection of one of its points. Hence the following algorithm:

(a) Select a random point in the area as in SRS; use this point as a 'starting point'.

(b) Find the other points of the cluster to which the starting point belongs, by applying predetermined geometric rules corresponding with the chosen cluster definition.

(c) Repeat step (a) and (b) until n clusters have been selected.

A condition for this algorithm to be valid is that the geometric rules are such that always the same cluster is created regardless of which of its points is used as starting point. A well-known technique satisfying this condition is random transect sampling with equidistant sample points on straight lines with a fixed direction. Given this direction, the random starting point determines the line of the transect. The other sample points are found by taking a pre-chosen distance in both directions from the starting point, until the line crosses the boundary of the area. Clusters thus formed will generally consist of a variable number of points, and the probability of selecting a cluster is proportional to the number of points in it. (This is taken into account in the statistical inference.)

Example

Fig. 5.4 shows four transects, each with four equidistant points. To limit the length of the transects, the area has first been dissected with internal boundaries perpendicular to the transects. Notice the spatial clustering and the regularity compared with SRS, StS and TsS (Fig. 5.1, 5.2 and 5.3). This is just a simple, notional example. It should be noted, however, that the clusters may be defined in any way that seems appropriate.

Statistical inference

For this type of design the same formulas are used as for TsS, clusters taking the role of primary sampling units. For clarity the inference is presented again, together with the 'cluster interpretation' of the quantities.

Means, areal fractions and SCDF's (after 0/1 transformation) of the area are estimated by the estimator:

$$\overline{y}_{\mathrm{Cl}} = \frac{1}{n} \sum_{i=1}^{n} \overline{y}_i \tag{5.6}$$

with n = number of clusters, and $\overline{y}_i$ = mean of cluster i.

The strategy (ClS, $\overline{y}_{\mathrm{Cl}}$) is p-unbiased. The variance is estimated by:

$$v(\overline{y}_{\mathrm{Cl}}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\overline{y}_i - \overline{y}_{\mathrm{Cl}})^2 .$$

Notice that the size of the clusters (number of points) don't occur in these formulas. This simplicity is due to the fact that the clusters are selected with

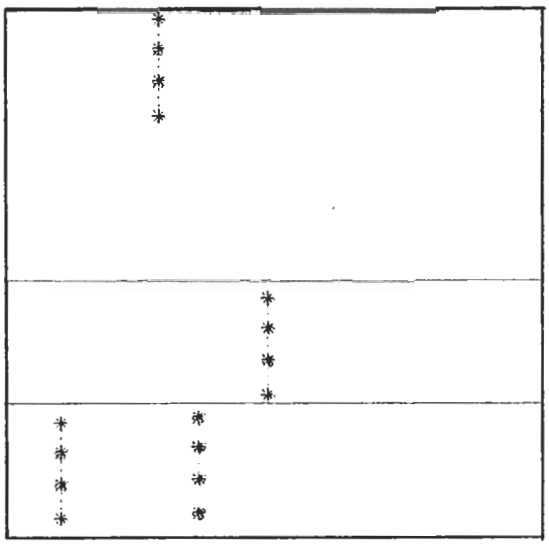


Figure 5.4: Notional example of Cluster Sampling

probabilities proportional to size. The effect of the cluster size on the variance is implicitly accounted for. (To understand this, consider that the larger the clusters are, the smaller the variance among their means must be.)

The standard deviation is estimated by $s(\overline{y}_{\mathrm{Cl}}) = \sqrt{v(\overline{y}_{\mathrm{Cl}})}$. Confidence intervals are calculated in the same way as with SRS, see Eq. 5.1.

The method of estimating means, areal fractions and SCDF's in domains depends on whether the area of the domain, A_j, is known or not. If it is known, then the mean of the jth domain, $\overline{Y}_j$, is estimated by:

$$\widehat{\overline{Y}}_j = \frac{\widehat{Y}_j}{A_j} \tag{5.7}$$

where $\widehat{Y}_j$ is an estimate of the total (spatial integral) of variable y over domain j. To estimate this total, we first define a new variable y', which equals y everywhere in the domain, but is zero elsewhere. The total of y over domain j equals the total of y' over the area, and this is estimated as A times the estimated mean of y', following Eq. 5.6:

$$\widehat{Y}_j = A \cdot \overline{y}'_{\mathrm{Cl}} = \frac{A}{n} \sum_{i=1}^{n} \overline{y}'_i,$$

where $\overline{y}'_i$ is the mean of the transformed variable y' in cluster i. The variance of the domain mean is estimated by:

$$v(\widehat{Y}_j) = \left(\frac{A}{A_j}\right)^2 \cdot \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\overline{y}'_i - \overline{y}'_{\mathrm{Cl}}\right)^2. \tag{5.8}$$

If the area of the domain is not known, it has to be estimated from the sample. An unbiased estimate to be substituted for A_j in Eq. 5.7 is:

$$\widehat{A}_j = \frac{A}{n} \sum_{i=1}^{n} \frac{m_{ij}}{m_i},$$

with m_{ij} = number of points in cluster i and domain j. Hence, the ratio estimator:

$$\widehat{\overline{Y}}_{\mathrm{R}j} = \frac{\widehat{Y}_j}{\widehat{A}_j} = \frac{\sum_{i=1}^{n} \overline{y}'_i}{\sum_{i=1}^{n} m_{ij}/m_i},$$

with estimated variance:

$$v(\widehat{\overline{Y}}_{\mathrm{R}j}) = \left(\frac{A}{\widehat{A}_j}\right)^2 \cdot \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\overline{y}'_i - \widehat{\overline{Y}}_{\mathrm{R}j} \frac{m_{ij}}{m_i}\right)^2.$$

Sample size

The number of clusters needed to keep the variance of the estimated mean below a given maximum V_m is given by $n = \frac{S^2_{\mathrm{B}}}{V_m}$, where S^2_{B} is a prior estimate of the variance between cluster means. Clearly, this variance depends on the number of points in the clusters and their spatial configuration. If prior information on the spatial variability is available in the form of a variogram, the method described in Section 5.9 can be used to estimate S^2_{B} for a given cluster definition.

If, instead of V_m, an absolute error d has been specified with an allowed probability of exceeding α, then V_m can be derived from d and α, according to $V_m = d/u_{1-\alpha/2}$, where $u_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

When estimating areal fractions rather than means of quantitative variables, the above formula for n can still be applied if S^2_{B} is interpreted as a prior estimate of the variance between cluster fractions.

Advantages

Like in TsS, the spatial clustering of sample points has the operational advantage of reducing the travel time between points in the field. In addition, the regularity may reduce the time needed to locate consecutive points in the cluster. Of course, the importance of these advantages depend on the scale, the accessibility of the terrain and the navigation technique used.

Disadvantages

As with TsS, the spatial clustering generally leads to lower precision, given the sample size. Again, the rationale is that due to the operational advantages a larger sample size can be afforded for the same budget, so that the initial loss

of precision is outweighed. If the spatial variation has a dominant direction, the precision can be optimized by taking transects in the direction of the greatest change.

Another disadvantage is that the sample size, i.e. the total number of points in the clusters which happen to be selected, is generally random. This may be undesirable for budgetary or logistic reasons. The variation in sample size can be reduced by defining clusters of roughly equal size.

5.7 Systematic Sampling (SyS)

Restriction on random selection

As with ClS, random selection is applied to pre-defined *sets* of points, instead of individual points as in SRS, StS and TsS. The difference with ClS is that only one cluster is selected. In this sense, SyS is a special case of ClS. (Note that the term 'cluster' as used here does not refer to geographical compactness, but to the fact that if one point of a cluster is included in the sample, all other points are included too.)

Selection technique

The selection algorithm for ClS is used with $n = 1$.

Example

Fig. 5.5 shows a random square grid. Notice the more even spatial spreading and the greater regularity compared with all other types of designs (Fig. 5.1 – 5.4).

Statistical inference

Means, areal fractions and SCDF's (after 0/1 transformation) of the area are simply estimated by the sample mean $\overline{y}$, as with SRS. The strategy (SyS, $\overline{y}$) is p-unbiased. This condition holds only if the grid is randomly selected, as is prescribed by the selection technique given above. With 'centered grid sampling', on the other hand, the grid is purposively placed around the center of the area, so that the boundary zones are avoided. This is a typical model-based strategy (see Chapter 6), which is p-biased.

Unfortunately, no unbiased variance estimators exist for this type of design. Many variance estimators have been proposed in the literature; all are based on assumptions about the spatial variation. A well-known procedure is Yates's method of balanced differences (Yates, 1981). An overview of variance estimation is given by Cochran (1977). A simple, often applied procedure is to calculate the variance as if the sample was obtained by SRS. If there is no pseudo-cyclic variation this over-estimates the variance, so in that case the accuracy assessment will be on the safe side.

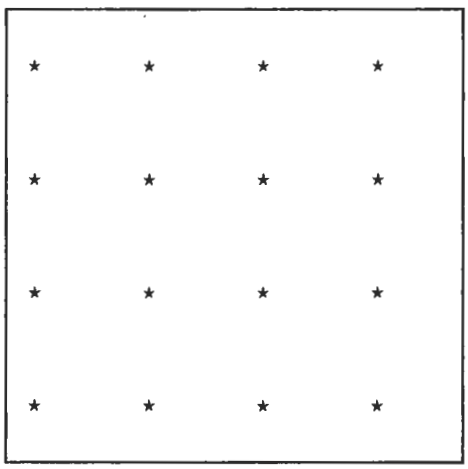


Figure 5.5: Notional example of Systematic Sampling

Means, areal fractions and SCDF's (after 0/1 transformation) in a domain are simply estimated by the sample mean in this domain:

$$\overline{y}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} y_{ij},$$

where m_j is the number of grid points falling in domain j.

Sample size

As indicated above, the sample size is random in general. The average sample size is determined by the choice of the grid size. A rough approach to this choice is to determine the sample size in the same way as for SRS (Section 5.3) and to reduce this with a empirical factor (for instance 2) to account for better precision of SyS relative to SRS. The average required grid size for a square grid is then $\sqrt{A/m}$. However, if an estimated variogram is available, it is more accurate to apply the method described in Section 5.9.

Advantages

Because only one cluster is selected, the clusters should be pre-defined such that each of them covers the area as good as possible. This is achieved with clusters in the form of regular grids: square, triangular or hexagonal. The statistical precision can thus be maximized through the definition of the grid. In addition, SyS has the same operational advantage as ClS: the regularity of the grid may reduce the time needed to locate consecutive points in the field. Again, the importance of this depends on the scale, the accessibility of the terrain and the navigation technique used.

Disadvantages

Because this type of design does not produce any random repetition, no unbiased estimate of the sampling variance is available. If the spatial variation in the area is pseudo-cyclic, the variance may be severely underestimated, thus making a false impression of accuracy. An operational disadvantage may be that the total travel distance between sample points is relatively long, due to the even spreading of the points. Finally, SyS has the same disadvantage as ClS: the sample size (number of grid points that happen to fall inside the area) is generally random, which may be undesirable for budgetary or logistic reasons. The possible variation in sample size will often be larger than with ClS, and it will be more difficult to reduce this variation.

5.8 Advanced design-based strategies

Apart from the basic strategies outlined in the previous sections, a large number of more advanced strategies have been developed. This section outlines some of the major possibilities.

5.8.1 Compound strategies

The basic strategies of the previous sections can be combined in many ways to form compound strategies. One example is given in Fig. 5.6, where TsS has been applied, however with SyS in both stages instead of SRS. In this case a square grid of 2×2 PU's was selected, and then a square grid of 2×2 points in each of the selected PU's. Notice that the total between-point distance is reduced as compared with SyS in Fig.5.5, that the risk of interference with possible cyclic variation has practically vanished, and that the operational advantage of regularity in the configuration still largely exists.

Fig. 5.7 shows another example of a compound strategy: Stratified Cluster Sampling with four strata and two clusters in each stratum. The clusters are perpendicular transects, each with two points at a fixed distance. Notice that, due to the stratification, a more even spreading is obtained as compared with ClS in Fig. 5.4, while the operational advantage of regularity still exists. See de

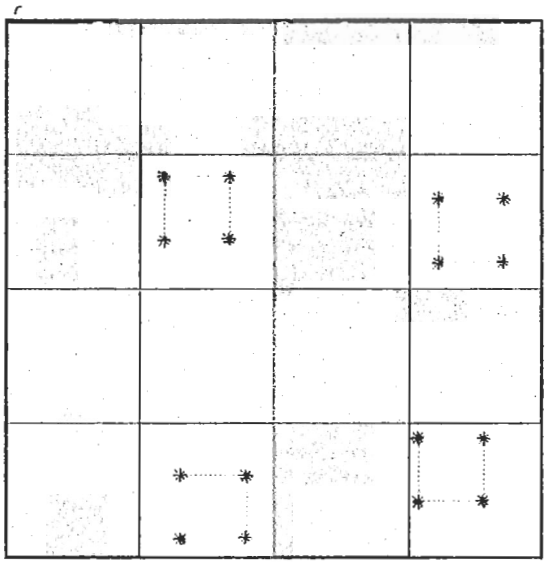


Figure 5.6: Notional example of Two-stage Sampling with Systematic Sampling in both stages

Gruijter & Marsman (1985) for an account of perpendicular random transect sampling and an application in quality assessment of soil maps.

As alluded in the examples above, the reason for combining two or more basic strategies is always an enhancement of advantages or mitigation of disadvantages of the basic strategies. As a final example, consider the situation in which the high precision and the operational advantage of regularity in SyS is wanted, however, it is desirable that the precision can be quantified from the data, without recourse to assumptions about the spatial variability. A possible solution is to adapt the Two-stage/Systematic compound strategy of Fig. 5.6. In order to enable model-free variance estimation, the PU's could be selected at random instead of systematically, while maintaining grid sampling in the second stage. In that case, the variance can be estimated in the same way as with basic TsS.

In devising a compound strategy, very often there are good reasons to stratify the area first, and then to decide which designs will be applied in the strata. It is not necessary to have the same type of design in each stratum. As long as the stratum means and their variances are estimated without bias, these estimates can be combined into unbiased overall mean and variance estimates using the formulas given in Section 5.4.

If a variogram for the area is available, the variance of a compound strategy can be predicted prior to sampling, using the Monte-Carlo simulation technique presented in Section 5.9. In the case of stratification this technique can be applied to each stratum separately, using different variograms if necessary.

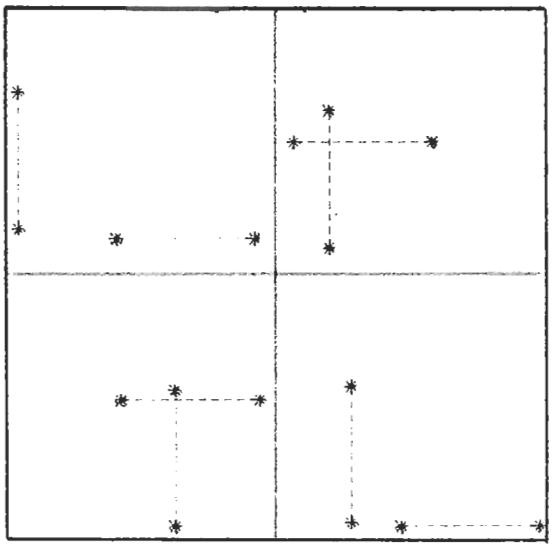


Figure 5.7: Notional example of Stratified Cluster Sampling

5.8.2 Spatial systematic strategies

Most strategies discussed so far are spatial in the sense that primary sampling units and clusters are defined on the basis of geographical co-ordinates. Also strata are usually defined that way. Given these definitions, however, the random selection restrictions do not refer to the co-ordinates of sample points. A category of more inherently spatial strategies exists of which the random selection restrictions make explicitly use of X and Y co-ordinates or distances in geographical space. Two examples are given.

Fig. 5.8 shows a 'systematic unaligned' sample. This technique was proposed by Quenouille (1949). The area is first divided into square strata and one point is selected in each stratum, however, not independently. A random X co-ordinate is generated for each row of strata, and a random Y co-ordinate for each column. The sample point in a stratum is then found by combining the co-ordinates of its row and column. Notice in Fig. 5.8 the irregular, but still fairly even spread of the points.

Fig. 5.9 shows a 'Markov chain' sample, a technique discussed by Breidt (1995). Again, notice the irregular but fairly even spread of the points. The underlying principle is that the differences between the co-ordinates of consecutive points are not fixed, as with systematic unaligned samples, but stochastic. These differences have a variance which is determined through a parameter ϕ, chosen by the user. Thus Markov Chain designs form a class in which one-per-stratum StR and systematic unaligned designs are special cases, with $\phi = 0$ and $\phi = 1$, respectively. The example in Fig. 5.9 was generated with $\phi = 0.75$.

As illustrated by the examples, the purpose of this type of strategies is to

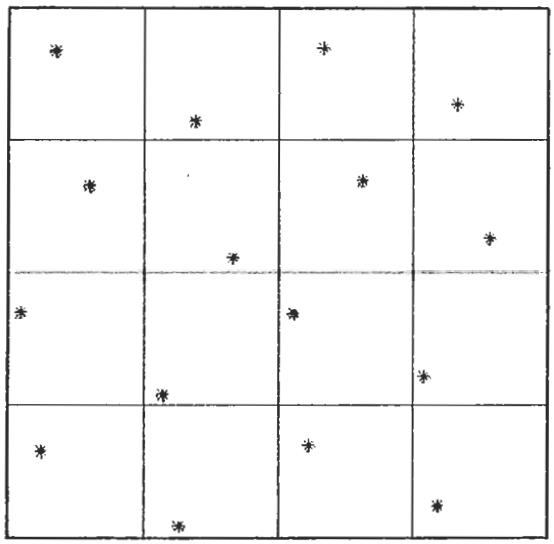


Figure 5.8: Notional example of Systematic Unaligned Sampling

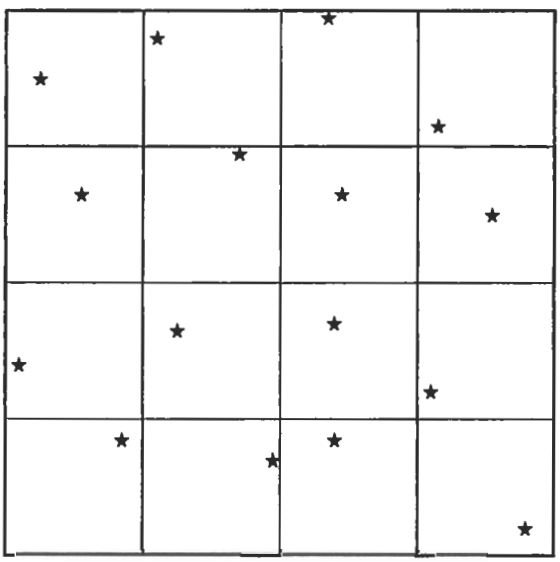


Figure 5.9: Notional example of Markov Chain Sampling

allow enough randomness to avoid the risk of interference with periodic variations and linear artifacts like roads, ditches, cables and pipelines, while still maintaining as much as possible an even spread of the points over the area.

5.8.3 Regression estimators

Suppose that an ancillary variable x is available which is roughly linearly related to the target variable y and known everywhere in the area, for instance from remote sensing or a digital terrain model. Then this information can be exploited by using a 'regression estimator'. For a simple random sample this is

$$\overline{y}_{\mathrm{Lr}} = \overline{y} + b(\overline{X} - \overline{x}),$$

where:

$\overline{y}$: sample mean of target variable;
$\overline{x}$: sample mean of ancillary variable, measured at the same points as y;
$\overline{X}$: areal mean of ancillary variable;
b : least squares estimate of the regression coefficient:

$$b = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

For large samples (say $n > 50$) the variance can be estimated by (Cochran, 1977):

$$v(\overline{y}_{\mathrm{Lr}}) = \frac{1}{n(n-2)} \sum_{i=1}^{n} \left[(y_i - \overline{y}) - b(x_i - \overline{x})\right]^2 .$$

If the ancillary variable is not known everywhere in the area, but can be measured cheaply in a large sample, then the relationship can be used by measuring y only on a random sub-sample, and again applying a regression estimator. This technique is known in the sampling literature as 'double sampling' or 'two-phase sampling'. Instead of the areal mean, $\overline{X}$, we now have the large sample mean $\overline{x}'$, so that

$$\overline{y}_{\mathrm{Lr}} = \overline{y} + b(\overline{x}' - \overline{x}),$$

with estimated variance (Chochran, 1977):

$$v(\overline{y}_{\mathrm{Lr}}) = s^2_{y.x}\left[\frac{1}{n} + \frac{(\overline{x}' - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right] + \frac{s^2_y - s^2_{y.x}}{n'},$$

where $s^2_{y.x}$ is the estimated residual variance:

$$s^2_{y.x} = \frac{1}{n(n-2)}\left[\sum_{i=1}^{n}(y_i - \overline{y})^2 - b^2(x_i - \overline{x})^2\right].$$

The regression estimators given above have been generalized to stratified sampling and to the case with more than one ancillary variable. They have a great potential for natural resource inventory, but their application in practice seems underdeveloped. Brus (2000) discussed in detail the use of regression models in design-based estimation of spatial means of soil properties.

5.9 Model-based prediction of design-based sampling variances

If prior information on the spatial variability is available in the form of a variogram, the following method can be used to predict the sampling variance of any design-based strategy. The core of the method is the general equation for predicting the variance of a design-based estimated mean from a variogram (Domburg e.a. 1994):

$$E_\xi[V_p(\widehat{\overline{Y}})] = \overline{\gamma}_A - E_p[\lambda' \cdot \Gamma_s \cdot \lambda] \tag{5.9}$$

where:

$E_\xi[.]$: statistical expectation over realizations from the model ξ underlying the chosen variogram;

$E_p[.]$: statistical expectation over realizations from the design p;

$V_p(.)$: variance over realizations from the design p (the usual sampling variance in the design-based approach);

$\overline{\gamma}_A$: mean semi-variance between two random points in the area;

λ : the vector of design-based weights of the points of a sample selected according to design p (For instance, if one cluster of 3 points and one of 2 points were selected, the weights in calculating the mean would be (cf. Eq. 5.6): 1/6, 1/6, 1/6, 1/4, 1/4.);

Γ_s : matrix of semi-variances between the points of a sample selected according to design p.

The first term, $\overline{\gamma}_A$, is calculated by numerical integration or by Monte-Carlo simulation, repeatedly selecting a pair of random points, calculating its semi-variance, and averaging. The second term can also be evaluated by Monte-Carlo simulation, repeatedly selecting a sample according to design p, calculating its mean semi-variance $\lambda' \cdot \Gamma_s \cdot \lambda$, and averaging. This generic procedure is computationally demanding but it is the only option for compound and spatial systematic strategies (Section 5.8). For the basic strategies, however, much more efficient algorithms are possible, making use of the structure of the design types. The following special prediction equations can namely be derived from the general Equation (5.9).

Simple Random Sampling

In the case of SRS, Equation (5.9) simplifies to:

$$E_\xi[V_p(\widehat{\overline{Y}})] = \frac{1}{n}\overline{\gamma}_A.$$

Stratified Sampling

For StS, Equation (5.9) becomes:

$$E_\xi[V_p(\widehat{\overline{Y}})] = \sum_{h=1}^{L} \frac{1}{n_h} \overline{\gamma}_{Ah},$$

where $\overline{\gamma}_{Ah}$ is the mean semi-variance between two random points in stratum h. Different variograms can be used for the strata.

Two-stage Sampling

For TsS and m_i constant the sampling variance is given by Equation (5.5). The variance components in this equation are the between-unit and the pooled within-unit variance, S^2_{B} and S^2_{W}. These components can be predicted from the two terms in Equation (5.9). The first term predicts the total variance, $S^2_{\mathrm{T}} = S^2_{\mathrm{B}} + S^2_{\mathrm{W}}$, while the second term predicts $S^2_{\mathrm{W}}/2$ if we take $n = 1$ and $m = 2$. In other words, the second term is calculated by repeatedly selecting one unit and two random points in it. The result is the mean semi-variance between pairs of random points within units, denoted by $\overline{\gamma}_U$. The sampling variance is then predicted by:

$$E_\xi[V_p(\widehat{\overline{Y}})] = \frac{1}{n}\left(\overline{\gamma}_A - \frac{m-1}{m} \cdot \overline{\gamma}_U\right).$$

Cluster Sampling

The sampling variance with ClS equals the between-cluster variance, S^2_{B}, divided by the number of clusters, n. To predict S^2_{B} for a given cluster definition, we apply Equaton (5.9) to ClS with $n = 1$. In other words, the second term is calculated by repeatedly selecting only one cluster. Within each cluster the points have equal weight $(1/m_i)$, so that $\lambda' \cdot \Gamma_s \cdot \lambda$ simplifies to the unweighted mean:

$$\lambda' \cdot \Gamma_s \cdot \lambda = \frac{1}{m_i^2} \sum_{k=1}^{m_i} \sum_{l=1}^{m_i} \gamma_{kl} = \frac{2}{m_i^2} \sum_{k=1}^{m_i-1} \sum_{l=k+1}^{m_i} \gamma_{kl},$$

because Γ_s is symmetric with zero diagonal. The result is the mean semi-variance between pairs of points within clusters, denoted by $\overline{\gamma}_C$. The sampling variance is then predicted by:

$$E_\xi[V_p(\widehat{\overline{Y}})] = \frac{1}{n}\left(\overline{\gamma}_A - \overline{\gamma}_C\right).$$

Of course, in the special case that all clusters have the same size and shape, $\lambda' \cdot \Gamma_s \cdot \lambda$ needs to be calculated only once.

Systematic Sampling

As SyS is ClS with $n = 1$, the sampling variance can be predicted by:

$$E_\xi[V_p(\widehat{\overline{Y}})] = \overline{\gamma}_A - \overline{\gamma}_C.$$

Again, in the special case that all clusters have the same size and shape, $\lambda' \cdot \Gamma_s \cdot \lambda$ needs to be calculated only once.

Chapter 6

Model-based sampling in 2-dimensional space

In the model-based approach the emphasis is on identifying suitable stochastic models of the spatial variation, which are then primarily used for prediction, given the sample data. This subject is treated in many textbooks on geostatistics. The models can also be used to find efficient sampling designs, but the main focus is on model building and inference, not on sampling design. This is natural, because the approach was developed to cope with prediction problems in the mining industry, where the data had already been collected via convenience or purposive sampling (Section 2.5). Nevertheless, stochastic models of the spatial variation have been successfully used in optimizing spatial sampling configurations for model-based strategies. Three different forms can be distinguished.

Firstly, if no prior point data from the area are available, the model can be used to determine the optimal sampling grid for point kriging or block kriging, given an accuracy requirement. It has been shown (Matérn, 1986) that if the spatial variation is second order stationary and isotropic, then equilateral triangular grids usually render the most accurate predictions, closely followed by square grids. In case of anisotropy the grid should be stretched in the direction with the smallest variability. McBratney e.a. (1981) presented a method to determine the optimal grid spacing for point kriging, given a variogram; a program and examples can be found in McBratney & Webster (1981). A similar method to determine the optimal grid spacing for block kriging is given by McBratney & Webster (1983). These methods are intended for large areas with a compact shape, so that boundary effects can be disregarded.

Secondly, if point data from the area pre-exist, the model can be used to find good locations for additional sampling. To that end a contour map of the kriging variance is made; additional sampling is then projected preferably in regions with high variance as this provides the largest reduction of uncertainty. This technique is practical and has found wide-spread application. It is only

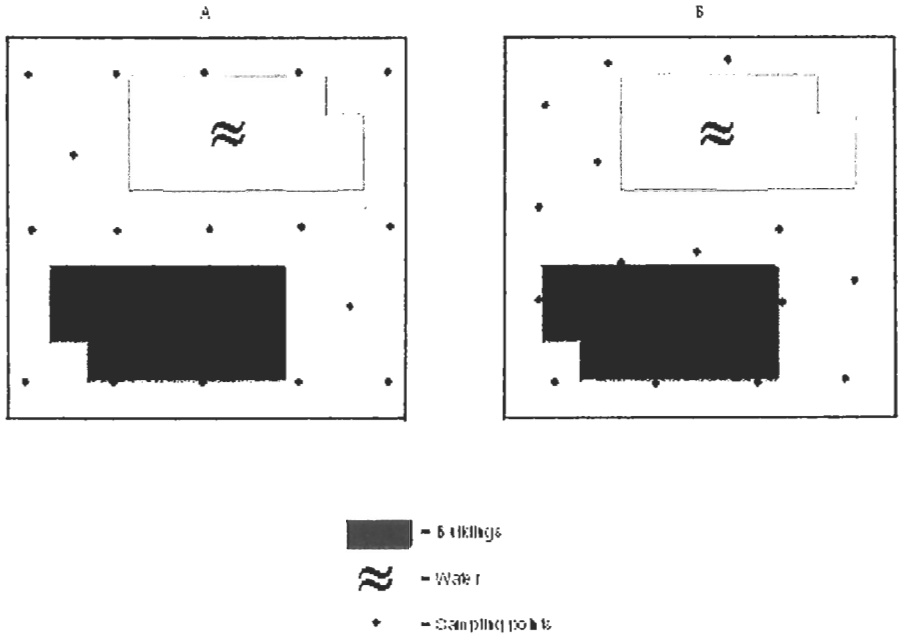


Figure 6.1: Examples of two model-based strategies: (A) centred equilateral triangular grid, and (b) configuration optimized by spatial simulated annealing with the Minimization of the Mean of Shortest Distances criterion. (From van Groenigen & Stein, 1998)

approximative, however, in the sense that it does not lead to an exact optimal configuration of sampling points.

Thirdly, if the area is small or irregularly shaped, then boundary effects cannot be disregarded and computationally more intensive methods are needed. Van Groeningen & Stein (1998) present such a method, using spatial simulated annealing. Fig. 6.1 shows an example of a point configuration optimized by their method.

The area contains two kinds of inclusions which cannot be sampled: a building in the South and water in the North. In this example, the soil under the building is part of the research area, say for soil sanitation, while the water is not. The optimized configuration shows that sample points are attracted by the 'research inclusion', but repelled by the 'non-research inclusion'. For comparison an equilateral triangular grid is shown, with the points removed that cannot be sampled. Using this method it is very easy to account for pre-existing data points; at the start they are simply added to the new points and their locations are kept fixed during the optimization process. The method then renders an optimized configuration, as opposed to the approximative method described above. Another advantage of this method is that it is versatile, because different

quality criteria for optimization can be build in easily.

The scope of model-based strategies is wider than that of design-based strategies. Firstly, the data requirements are more relaxed. Data from convenience, purposive as well as probability sampling can be used for model-based inference, while design-based inference requires probability sampling. Secondly, model-based inference can be directed towards a wider class of target quantities, including local functions and functions defined by geographic neighborhood operations. An example of the latter is the total surface area of land patches consisting of a minimum number of adjacent pixels classified as suitable for a given land use. A local function which can only be predicted by a model-based strategy is, for instance, the spatial mean of a small domain (or 'block') with no sample points in it.

The price paid by the model-based approach for its larger versatility is full dependency on a stochastic model of which the validity is more or less arguable. If the alternative of the design-based approach is not applicable, this dependency just has to be accepted. However, where the scope of the two approaches overlap, one has a choice as discussed in Chapter 3.

Chapter 7

Sampling in 1- or 3-dimensional space

The previous chapters are focussed on sampling in 2D space, but in practice very often other dimensions are involved: 1D or 3D space, time or space-time. In general, the methodology presented for 2D space can be easily transferred or adapted to these other dimensions as is outlined in this and the next chapters.

One-dimensional spatial universes can have a horizontal or a vertical orientation. Horizontal 1D universes are, for instance, projected trajectories of roads or pipelines. The methodology presented for 2D is directly transferable to this situation. Sampling in vertical 1D space, i.e. sampling at depth, is practically always done at more than one location, hence it is part of sampling in 3D space.

The universe of interest is very often embedded in 3D space. Sample points would then have three coordinates (X, Y and Z), and theoretically all three could be determined independently of each other, similarly to X and Y in 2D sampling. That would typically lead to sampling at a single variable depth at each location. However, this is hardly ever done in practice. There are two main reasons to treat the Z coordinate differently, and to decompose the 3D sampling problem into a 2D (horizontal) problem and a 1D (vertical) problem.

The first reason is when the target variable is defined as a function of soil properties at various depths, as is usually the case in the context of, for instance, plant growth and leaching. It is then logical to sample at these depths at each location. The second reason is when the target *variable* is defined at points in 3D space, e.g. the concentration of a contaminant, and the target *quantity* is defined over 3D space, e.g. the 3D spatial mean. In that case, although not a necessity, it is usually efficient to take samples at various depths at the same location. The sample is designed and analysed as a two-stage sample, with locations as primary units and depths as secondary units (see Section 5.5).

The methodology of sampling at depth is in principle the same as that for 2D space, however, cluster and two-stage sampling wil usually inefficient because their operational advantages in 2D space do not hold for sampling at depth. The

two dominant techniques in practice are purposive sampling at fixed depths and stratified systematic sampling, with soil horizons as strata and compositing of samples from the same horizon.

Part III

Monitoring: sampling in time or in space-time

Chapter 8

Central issues in monitoring

8.1 Purposes of monitoring

According to Webster's dictionary, monitoring is: "to watch, observe or check for special purposes". The good thing of this definition is that it links monitoring explicitely with "special purposes", to the extent that without a special purpose we should not even speak of monitoring. Nevertheless, two essential elements are missing in this definition. Firstly, that monitoring is repeated and continued for a shorter or longer period of time. Secondly, the observation is being done in a more or less systematic way. Generally speaking, monitoring of natural resources should provide the information that is necessary for taking proper decisions on natural resources management. With regard to soil, a traditional kind of monitoring is for the nutrient status of agricultural fields as a whole, while recently systems are being developed to monitor variations within fields, to support precision agriculture. Also, during the last decades systems have been set up to monitor soil quality and soil pollution at a regional or national scale. Monitoring in hydrology shows a large variety of aims and scales. As for soil, monitoring of open water as well as ground water may be directed to quality and pollution, or otherwise to quantity, with water level as an important aspect. Monitoring in ecology has still a wider scope than in soil science and hydrology; important objectives are evaluation of effects of environmental changes on species abundances and occurrence of vegetation types.

With a view on designing monitoring systems, it is useful to distinguish three categories of monitoring according to its purpose (Dixon & Chiswell, 1996; Loaiciga e.a., 1992):

- ambient or status monitoring for quantitative description of the universe as it changes with time;
- trend monitoring to test whether temporal trends are present in the universe;

- regulatory or compliance monitoring to test whether the universe satisfies regulatory conditions.

With status monitoring the sampling scheme should allow efficient estimation of decriptive parameters repeatedly; with trend and regulatory monitoring an important issue in system design is the validity and power of the relevant statistical tests.

8.2 Monitoring: sampling under changing circumstances

In view of the design of sampling schemes, an extremely important difference between inventory and monitoring is that inventory takes place within a relatively short period of time, during which neither the universe is supposed to change in any relevant way, nor the operational, organisational and budgetary conditions will generally alter. With monitoring, on the other hand, not only the universe may undergo large, unexpected changes, but especially in long-term monitoring the conditions may also alter in a way that makes adaption of the sampling scheme inevitable or desirable. There may be a change in the objectives or their priorities, or new knowledge or better models or techniques may become available. An important situation that will always change during monitoring is, the amount of available data. Although this is clear enough, it is less obvious how to exploit the increasing information on the universe for re-design or fine-tuning of the sampling scheme. The fact that monitoring, especially long-term monitoring, is generally bound to face changing circumstances calls for flexibility of the sampling scheme. The so-called dynamic and rotational systems are more powerful in this respect; see Chapter 11.

Chapter 9

Designing a monitoring system

I use the term 'monitoring system' here as a synonym of 'sampling scheme for monitoring', i.e. a scheme for sampling in time or in space-time. A term often used in this connection is 'monitoring network', which has been defined by Loaiciga e.a. (1992) as a fixed of set sampling locations and a sampling frequency. The 'system' concept is much more general than the 'network' concept, for two reasons. Firstly, a monitoring system is a sampling scheme, so it includes not only a sampling design in space-time but, for instance, also the measurement technique and the method of data analysis (see Section 2.1). Secondly, the set of sampling locations of a monitoring network does not change with time and the same sampling frequency is applied at each location: it represents a so-called *static sampling design*. A monitoring system, on the other hand, may use a *dynamic sampling design* with sampling locations that change with time, or a *rotation sampling design* in which part of the locations change with time. As dynamic and rotational systems have the highly desirable property of flexibility (Section 8.2), it is important not to exclude these in advance by using a too restrictive concept of sampling design.

Monitoring system design is decision taking under uncertainty and often has to deal with changing conditions, high complexity and competing objectives. Apart from flexiblity, this asks for:

1. *robustness* of the system, so that if the most questionable assumptions made in designing the system appear to be incorrect, it still produces useful information.

2. an *iterative approach*, to anticipate the situation that more data and better insight is available for fine-tuning and re-designing the system.

In Section 2.2 I proposed as a guiding principle in design of sampling schemes: "start at the end, then reason backwards". In other words: start the design

process with a precise as possible specification of the objective, by analysing the information need. I believe that this is a good principle in sampling for inventory as well as for monitoring. The objective is the main factor that determines the spatial and temporal scale, the target variables and parameters, the level of detail, and the objective function.

The objective function that you choose to optimize in designing a system should represent the final aim as accurately as possible, but a problem is that there is often more than one objective, and the objectives may change. Anyhow, it is good to realize that in principle there is a choice between two main kinds of objective function: surrogate objective functions (statistical quantities like error-variance or detection probability) and ultimate objective functions, which estimate beforehand the value of the monitoring information in achieving the ultimate goals. In principle, if an ultimate objective function can be defined realistically, that would be better than a surrogate function, because that makes it more likely that the system will be doing what it is supposed to do.

Given an objective function, the technique to optimize it will generally be some form of Monte-Carlo simulation with a stochastic model of the spatio-temporal variation in the universe. This generates equi-probable random realizations of the universe, and for each realization the response from a candidate monitoring system is evaluated. The value of the objective function for the candidate system is then calculated over these responses, and an iterative search algorithm repeats this process for other candidates in order to find the best system.

Some pitfalls in designing monitoring systems are worth mentioning. The first pitfall is that one does not fully acknowledge the dynamics of the problem, by not realizing or under-estimating how fast some circumstances may change while the system is used. Secondly, there may be an inclination to under-estimate the temporal variation, hence to choose a design that covers the time domain insufficiently, because it is usually easier to get a realistic prior estimate of the spatial variation than of the temporal variation. Thirdly, it may be tempting to adopt a cheap-to-measure target variable at the cost, however, of large bias in the final results. Suppose, for instance, that the objective is to estimate the total emission of a pollutant from the soil to the ground water in a given area during a given period. One possible strategy would then be to measure the concentration of the pollutant in the soil moisture at the sampling points, to estimate the mean concentration from these data, and to multiply the mean concentration with the total ground water recharge taken from a water balance for the area. The advantage of this strategy is that only concentrations need to be measured. However, the disadvantage is that the estimate of the total emission is possibly seriously biased. The cause of this bias is that the strategy assumes implicitly that concentration and recharge are independent variables, whereas in reality this will not be true; for instance, there may be a tendency for high concentrations at times and at places with low recharge to the ground water. A solution is to measure not only the concentration at the sample points but also the flux to the ground water, and to take the product of these two as the target variable.

The last pitfall to mention here is the use of a wrong formula for sample size or sampling frequency. To explain this, consider the variance of the estimated mean of some target variable x over a universe. Suppose we have n observations on x, with a deterministic mean μ plus a random component ε with variance σ^2:

$$x_i = \mu + \varepsilon_i \qquad (i = 1...n)$$

If we take the unweighted sample mean as estimator of μ:

$$\widehat{\mu} = \frac{1}{n} \sum x_i$$

and if the observations are independent, then the variance of the mean is given by:

$$\mathrm{Var}(\widehat{\mu}) = \frac{\sigma^2}{n} \tag{9.1}$$

However, if the observations are not independent, then it was realized long ago (Bayley & Hammerlsley, 1946), that this formula needs adjustment by taking account of the covariances between the observations:

$$\mathrm{Var}(\widehat{\mu}) = \frac{1}{n^2} \left\{ \sum \sigma^2 + 2 \sum \sum \mathrm{Cov}(x_i, x_j) \right\} = \frac{\sigma^2}{n} \{1 + (n-1)\overline{\rho}\} \tag{9.2}$$

where $\overline{\rho}$ denotes the average correlation between the observations. So, an equivalent sample size was defined, equal to the nominal sample size divided by the correction factor in Eq. 9.2:

$$n_{\mathrm{eq}} = \frac{n}{\{1 + (n-1)\overline{\rho}\}} \tag{9.3}$$

This formula for equivalent sample size has become rather popular and is applied in time serie analysis (Lettenmayer, 1976; Matalas & Langbein, 1962; Zhou, 1996) as well as in spatial statistics, for instance in Gilbert's book on ecological monitoring (Gilbert, 1987). The formula seems entirely correct, but if we look at what happens with the variance of the mean when we increase the sample size, some odd behaviour can be noticed. Take as a simple example an equidistant time series with the exponential autocorrelation function $\rho(T) = e^{-3T}$ (see Fig. 9.1).

Futhermore, we take both σ^2 and the monitoring period equal to 1, and increase the sample size by increasing the sampling frequency. Using Equation (9.1) and (9.2), respectively, for independent and dependent observations we obtain the variance of the estimated mean ($\widehat{\mu}$) as a function of sample size, depicted in Fig. 9.2.

In Fig. 9.2 we see that with independent observations the variance continuously decreases with increasing sample size, however with dependent observations the variance first drops, but not lower than a certain level, and after that it stays nearly constant, in fact even increases somewhat. In other words,

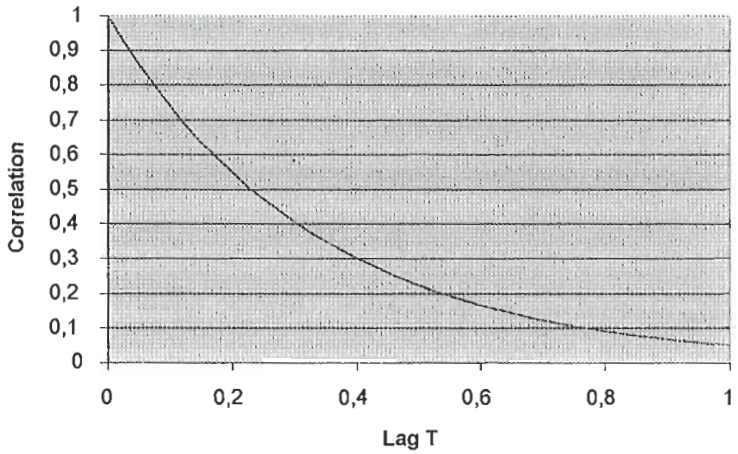


Figure 9.1: Autocorrelation function used to calculate the variance of the estimated mean; see text.

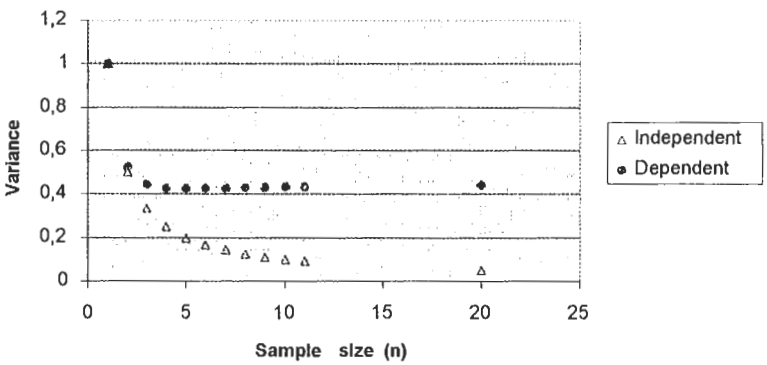


Figure 9.2: Variance of the estimated mean as a function of sample size, for independent and dependent observations.

according to Equation (9.2) we cannot reach a precision beyond a certain level no matter how many samples we take, which counters the intuition that the larger the sample, the more we know about the universe. The reason for this is not that Equation (9.2) is incorrect, but that we were estimating the wrong mean. Remember that μ is just a model parameter. The solution is that we don't take the model mean as the target quantity to estimate, but the average of x over the universe of interest:

$$\overline{x} = \frac{1}{\|\mathbf{U}\|} \int_{\mathbf{U}} x \mathrm{d}u$$

where the integration is over the universe of interest $\mathbf{U}$, with size $\|\mathbf{U}\|$.

As explained in Chapter 3 there are two different approaches to estimate $\overline{x}$: the model-based appoach, which leads to some form of block-kriging or a time series equivalent of that, and the design-based approach, using probability sampling. The methods derived from these two approaches are dealt with separately in Chapter 5 and 6 for spatial universes, in Section 10.1 and 10.2 for temporal universes, and in Section 11.1 and 11.2 for space-time universes.

In summary, when designing a monitoring system it seems useful to:

- optimize an ultimate objective function whenever possible;
- consider the use of a rotational system because of its inherent flexibility (Section 11.1);
- be careful with the use of stratifications, because changes in the universe may render the original stratification obsolete or hard to maintain.

Chapter 10

Sampling in time

Sampling in time is done, for instance, to monitor the quality of surface water or groundwater at a single critical location. It is similar to sampling in space in the sense that, although the practical aspects may differ, the same principles, theory and problems of choice play a role. Therefore, much of what has been said about spatial sampling in Part II is applicable to temporal sampling as well. In particular, the distinction and the choice between the design-based and the model-based approach is again of paramount importance and is taken here too as the main sub-division of the methodology. See Chapter 3 for a general discussion of how to choose between these two appoaches. In the special case of sampling in time it should be added that cyclic variations seem to be more common in time than in space. If this is true, then for sampling in time more caution is needed with model-based strategies involving systematic sampling, i.e. at constant time-intervals, because of a greater risk that the sampling interval interferes with some cyclic pattern of variation (see also Section 5.7). On the other hand, taking samples at constant intervals is often more convenient. Of course this advantage vanishes when a programmable automatic sampling device can be installed. With sampling in time too, design-based strategies have the advantage of greater simplicity and more robustness in the sense that the statistical inference from the sample data does not rely on the validity of a time series model. The scope of design-based strategies, however, is limited to estimation of parameters related to the universe or to sub-universes as a whole, e.g. means, totals, quantiles or distribution functions (see also Section 5.2).

Notwithstanding the similarities, temporal sampling may differ in one principal aspect from spatial sampling: while spatial sampling always takes place in a bounded universe, the temporal universe to be monitored may have an end-point that is still undetermined when the monitoring system is designed. This has several consequences, as discussed in the following sections. One practical consequence is that, instead of the total sample size, the average sample size per unit of time or the sampling frequency becomes the major parameter of the monitoring system.

10.1 Design-based sampling in time

Although random sampling seems to be less usual in time than in space, the same methods as discussed in Chapter 5 for spatial sampling can be used for temporal sampling. Clearly, where in 2D spatial sampling populations, domains, strata and primary sampling units are all areas, in temporal sampling they are periods of time. The advantages and disadvantages indicated for the various spatial sampling strategies in Chapter 5 hold, *mutatis mutandis*, for sampling in time. If the end of the monitoring period is pre-determined, the selection techniques presented in Chapter 5 for the 2D spatial context only need obvious adaptations to the 1D temporal context. Long-term monitoring projects often have no pre-determined end, but budgets tend to be allocated annually. In that case it is practical to take the budgetary years as strata, and to determine the sample size for each successive year from the available budget.

There is one exception to the rule that the 2D spatial sampling designs of Chapter 5 are applicable in time: the systematic unaligned type of design in Section 5.8.2 needs 2 dimensions. In place of that, the Markov Chain design discussed in Section 5.8.2 is well suited to achieve a fairly even spread of sampling points over time, while still avoiding the risk of interference with (pseudo-)cyclic variations.

One important purpose in temporal sampling is not covered by the design-based methods presented for spatial sampling in Chapter 5: estimation or testing of a step trend. If interest lies in possible effects of a sudden natural change or certain human activities that start at a given point in time, then a relevant quantity to estimate may be the difference between the temporal means before and after the change:

$$D = \overline{x}_{\mathrm{a}} - \overline{x}_{\mathrm{b}} = \frac{1}{t_{\mathrm{e}} - t_{\mathrm{c}}} \int_{t_{\mathrm{c}}}^{t_{\mathrm{e}}} x \mathrm{d}t - \frac{1}{t_{\mathrm{c}} - t_{\mathrm{b}}} \int_{t_{\mathrm{b}}}^{t_{\mathrm{c}}} x \mathrm{d}t$$

where $\overline{x}_{\mathrm{a}}$ and $\overline{x}_{\mathrm{b}}$ are the temporal means after and before the change, respectively, t_{b} and t_{e} are the beginning and end time of the monitoring, and t_{c} is the time at which the change happens. This effect is simply estimated by:

$$\widehat{D} = \widehat{\overline{x}}_{\mathrm{a}} - \widehat{\overline{x}}_{\mathrm{b}}$$

where $\widehat{\overline{x}}_{\mathrm{a}}$ and $\widehat{\overline{x}}_{\mathrm{b}}$ are estimators of the temporal means, depending on the applied type of sampling design (see Chapter 5). If the samples taken before and after the change are taken independently from each other, then the variance of $\widehat{D}$ equals:

$$V(\widehat{D}) = V(\widehat{\overline{x}}_{\mathrm{a}}) + V(\widehat{\overline{x}}_{\mathrm{b}})$$

where $V(\widehat{\overline{x}}_{\mathrm{a}})$ and $V(\widehat{\overline{x}}_{\mathrm{b}})$ are the true sampling variances of the estimated means. An estimate $v(\widehat{D})$ of $V(\widehat{D})$ can simply be obtained by inserting the estimates of $V(\widehat{\overline{x}}_{\mathrm{a}})$ and $V(\widehat{\overline{x}}_{\mathrm{b}})$, as given in Chapter 5 for the various designs:

$$v(\widehat{D}) = v(\widehat{\bar{x}}_{\mathrm{a}}) + v(\widehat{\bar{x}}_{\mathrm{b}})$$

A two-sided $100(1-\alpha)$ % confidence interval for D is given by:

$$\widehat{D} \pm t_{1-\alpha/2} \cdot \sqrt{v(\widehat{D})} \tag{10.1}$$

where $t_{1-\alpha/2}$ is the $(1-\frac{\alpha}{2})$ quantile of the Student distribution with $(\eta_{\mathrm{a}}+\eta_{\mathrm{b}})$ degrees of freedom; η_{a} and η_{b} denoting the degrees of freedom on which the estimates $v(\widehat{\bar{x}}_{\mathrm{a}})$ and $v(\widehat{\bar{x}}_{\mathrm{b}})$ are based. The null-hypothesis of no effect ($D = 0$) can be tested against the alternative $D \neq 0$ with the two-sided two-sample t-test. The null-hypothesis is rejected if the confidence interval of formula (10.1) does not contain zero.

10.2 Model-based sampling in time

The simplest and most usual model-based method of sampling in time is to sample at equidistant points in time, and to use a time series model for statistical inference from the sample data. Time series analysis is a broad subject on its own, and a vast literature exists on its methodology. A practical textbook is Box & Jenkins (1976); see Hipel &McLeod (1994) for applications in natural resources. These books discuss in detail how models of the temporal variation may be selected and fitted to the data, and how these models can be used to test, estimate and forecast quantities of interest. We repeat the warning in Chapter 9 that some formulas for sample size imply that the *model* mean is to be estimated, and therefore render sample sizes that are larger than needed for estimating the more relevant *temporal* mean, the average of the target variable over the monitoring period.

The tendency to sample equidistantly in time is obviously caused by operational advantages but probably enforced by the fact that equidistant series can be analysed by methods which are mathematically relatively simple. However, taking spatial sampling as an analogy, it can be conjectured that equidistant sampling in time is not always the best option even in the model-based approach. As explained in Section 5.4, if in spatial sampling sub-areas can be outlined beforehand that are expectedly more variable than others, then it is efficient to stratify accordingly and to sample more densely in the more variable strata. Similarly, if the temporal variation varies with time then it should be efficient to sample more densely in periods with larger variability. Another example can be borrowed from Chapter 6 on model-based sampling in space, where it was demonstrated that when optimizing a configuration of sample points for an area in which certain parts are not accessible for sampling, the sample points are attracted or repelled by the boundaries of the inaccessible parts, depending on whether they belong to the target area or not. Similarly, if sampling is impossible in one or more sub-periods then a sampling design adapted in the same way should be more efficient than an equidistant one.

Chapter 11

Sampling in space-time

The general remarks made in Chapter 10 about sampling in time also apply to sampling in space-time. In addition, two specific concepts need to be mentioned here: *static* and *dynamic* monitoring systems. Static monitoring systems are systems by which samples are taken each time at the same locations. With dynamic systems a new set of sampling locations is selected at each sampling time. The difference between static and dynamic systems is illustrated by Fig. 11.1 and Fig. 11.2.

The choice between a static and a dynamic system should be guided by operational as well as statistical considerations. Obviously, a static system has an operational advantage if the costs of repeated sampling at the same location are lower than for sampling at different locations with the same total sample size. Common reasons for this are when locating sample points in the field is easier because they can be marked, or when sampling, measuring or recording equipment is installed at fixed points in the field on a semi-permanent basis. A statistical advantage of static systems is that estimation of temporal trends will often be more efficient than other systems. A statistical disadvantage is that while monitoring goes on, only the information on temporal variability increases, not that on spatial variability.

The main advantage of dynamic systems is that they are much more flexible than static ones. This is because at each sampling time the system can be adapted to altered circumstances with respect to the spatial or temporal variablity existing in the universe, the accumulating amount of information on both these variabilities, or the information needs. The importance of flexibility in monitoring systems has been discussed in Section 8.2.

Rotational systems are a compromise between static and dynamic systems in that at each sampling time a fraction of the locations is rotated out of the sample and replaced by new ones (see Fig. 11.3). Advantages compared with static systems are greater flexibility and better spatial coverage. If there is a fair amount of temporal correlation between sampling times, advantages compared with dynamic systems are their higher efficiencies in estimating temporal trends as well as status; see Section 11.1.

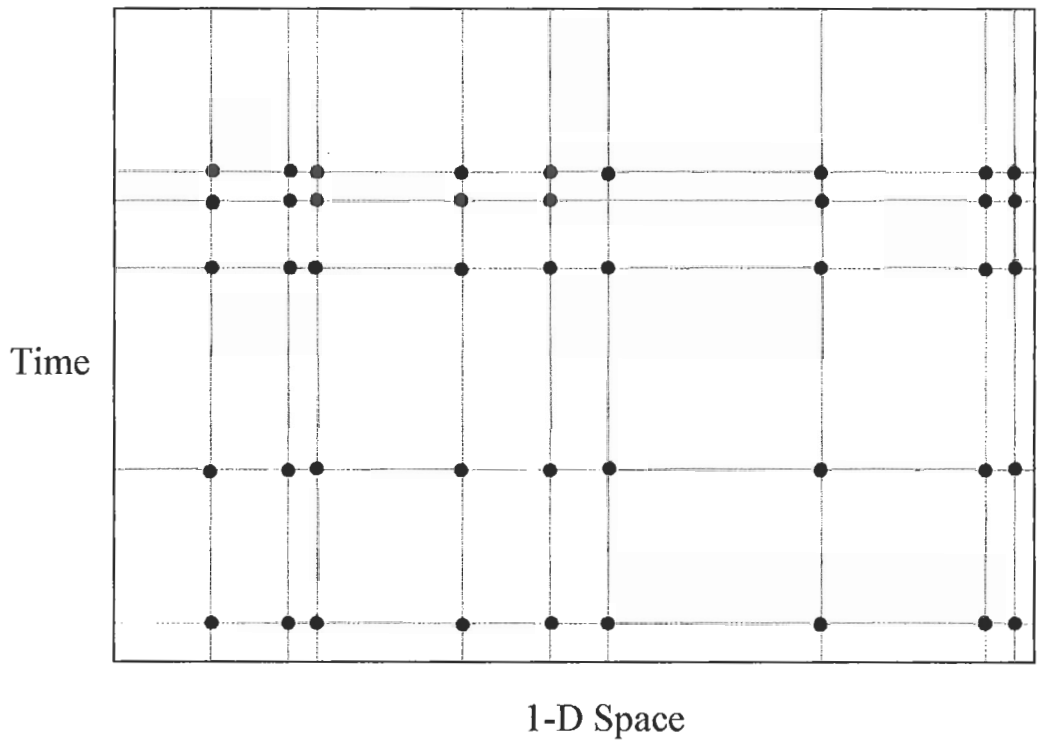


Figure 11.1: Notional example of a static system, with Simple Random Sampling in both space and time.

11.1 Design-based sampling in space-time

11.1.1 Static systems

A static system can be viewed as a combination of a spatial sampling design and a temporal sampling design, such that at each sampling time all points of the spatial design are sampled and vice versa (see Fig. 11.1). The inference in static systems will depend primarily on these two constituting designs. For both space and time a choice has to be made between the design-based and the model-based approach, so there are four possible combinations: design-based in space and time, design-based in space plus model-based in time, model-based in space plus design-based in time, and model-based in space and time. Only the first combination, leading to purely design-based systems, is dealt with in this

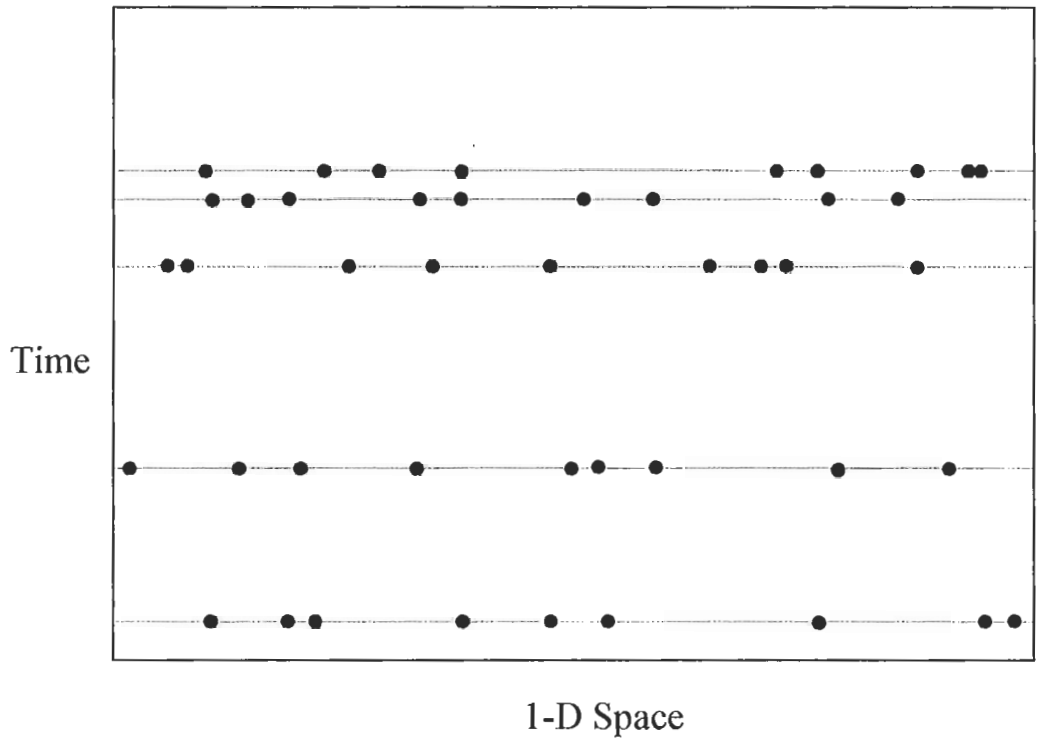


Figure 11.2: Notional example of a dynamic system, with Simple Random Sampling in both space and time.

Chapter; the other combinations use a model of the spatial and/or temporal variation for inference and are dealt with in Section 11.2.

The set of sampling locations as employed in design-based static systems can be selected by the same designs as described in Chapter 5 on design-based sampling in space, while the set of sampling times can be selected by the methods discussed in Section 10.1 on design-based sampling in time. Inference on (a parameter of) the Spatial Cumulative Distribution Function at any given sampling time can be done by applying the appropriate method from Chapter 5 on the data collected at that time.

Inference on a space-time mean is done in two steps. First, for each sampling location the temporal mean is estimated from the data at that location, using the method associated with the temporal design. Then the space-time mean and its standard error is estimated using these temporal means as 'observations', using

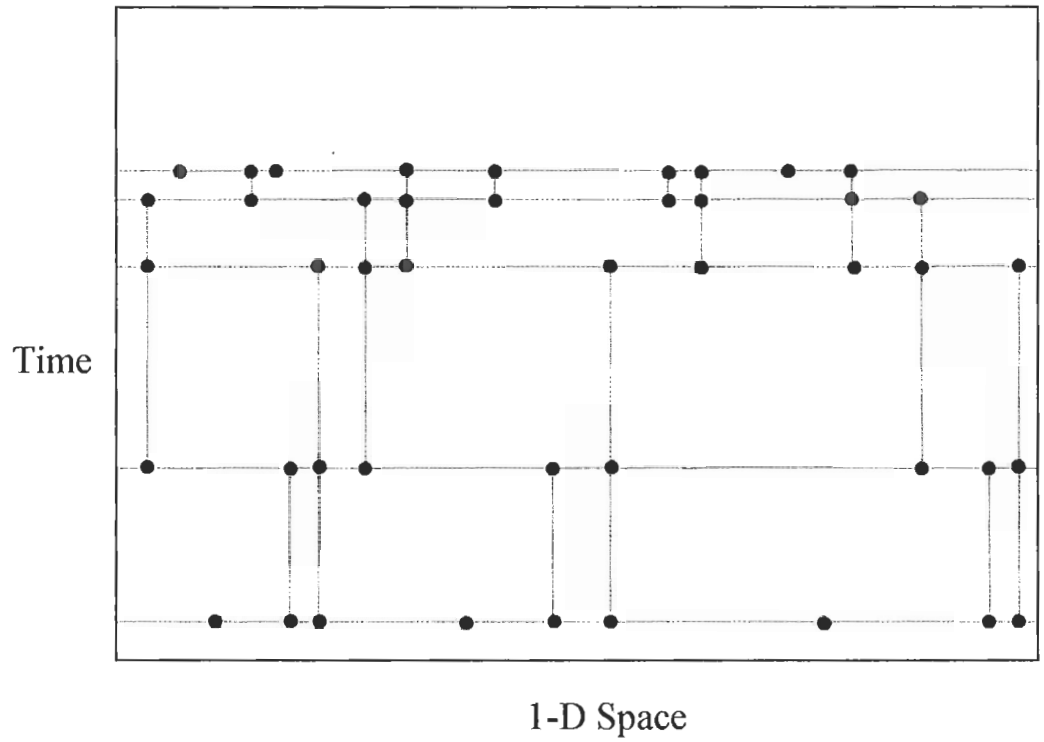


Figure 11.3: Notional example of a rotational system, with Simple Random Sampling in both space and time.

the method associated with the spatial design. This standard error accounts automatically for errors due to sampling in space and sampling in time. The same two-step procedure can be followed for estimating space-time totals.

If the difference between the spatial means at two different times, $D = \overline{Y}(t_2) - \overline{Y}(t_1)$, must be estimated or tested, then a possible temporal correlation between the estimated means $\widehat{\overline{Y}}(t_1)$ and $\widehat{\overline{Y}}(t_2)$ should be taken into account. A simple, implicit way of doing this is similar to the two-step procedure indicated before, and is based on the fact that the difference between two spatial means is equal to the spatial mean of the differences. So, first calculate the difference $d_i = y_i(t_2) - y_i(t_1)$ at each sample point i, then apply the appropriate method of inference from Chapter 5 to those differences. In the case of classical testing, this procedure leads to the common t-test for paired observations. The same procedure can be followed for inference on the spatial mean of any temporal

trend parameter, such as the difference between the temporal means before and after some event (step trend), or the average change per unit of time (linear trend).

11.1.2 Dynamic systems

With dynamic systems, at each sampling time one is free to choose a spatial sampling design from Chapter 5 that seems most appropriate given the circumstances at that time. For simplicity, Fig. 11.2 shows the same spatial design at every sampling time (SRS with $n = 9$), but in applications one may adapt the sample size, possible stratification, clusters and/or primary sampling units, and even the very type of design. Because the samples taken at different times are mutually independent, the estimated means $\widehat{\overline{Y}}(t_1)$ and $\widehat{\overline{Y}}(t_2)$ are so too. Hence the difference D is estimated by

$$\widehat{D} = \widehat{\overline{Y}}(t_2) - \widehat{\overline{Y}}(t_1)$$

and its standard error is simply estimated from the variances:

$$s(\widehat{D}) = \sqrt{v\left(\widehat{\overline{Y}}(t_2)\right) + v\left(\widehat{\overline{Y}}(t_1)\right)}$$

In the case of classical testing, this procedure leads to the common two-sample t-test.

More generally, the two-step procedure for static systems (estimating first temporally, then spatially) is now reversed into estimating first spatially, then temporally. For instance, inference on a space-time mean proceeds by first estimating the spatial mean at each sampling time (using the method associated with the spatial design at that time), and then estimating the space-time mean from these means as 'observations' (using the method associated with the temporal design). Inference on totals and trend parameters is similar.

Dynamic systems can be considered as a special case of two-stage sampling in space-time, using spatial sections of the universe at given times as primary sampling units, and sampling locations as secondary units. Therefore, the methods of inference for two-stage sampling in space, given in Section 5.5, can be applied. In the extreme case when one location is selected at each sampling time, this is equivalent to Simple Random Sampling in space-time. (Note that the designs in the examples of Fig. 11.1, 11.2 and 11.3, although with SRS in both space and time, are not equivalent with SRS in space-time.)

Of course one may reverse the order of space and time in the two stages, by using temporal sections of the universe (i.e. periods of time) at given locations as primary sampling units, and sampling times as secondary units. Now the set of sampling locations remains fixed through time, as with static systems, which brings the same operational advantages. The difference with static systems is that sampling at the various locations is not synchronized, and that the temporal design may be adapted to local circumstances. This kind of dynamic system is

attractive when considerable spatial variation between time series is known to exist, and the mentioned operational advantages are real. The inference is as for static systems.

11.1.3 Rotational systems

Rotation sampling or 'sampling with partial replacement' represents a compromise between static and dynamic systems. The rationale is to avoid on the one hand the unbalancedness of static systems that accumulate more data only in time. On the other hand, the relative inefficiency of dynamic systems for estimating temporal trends is partially avoided because repeated measurements are made at the same locations.

The principle of rotation sampling is to divide the locations of an initial spatial sample into different rotation groups, and to replace each time one group by a new set of locations (see Fig. 11.3). The spatial mean at any time t_0, $\overline{Y}_0$, is estimated by the composite estimator (Rao and Graham, 1964; Salski, 1990):

$$\hat{\overline{Y}}'_0 = Q(\hat{\overline{Y}}'_{-1} + \hat{\overline{Y}}_{0,-1} - \hat{\overline{Y}}_{-1,0}) + (1-Q)\hat{\overline{Y}}_0$$

where Q is a weighing constant ($0 < Q < 1$), $\hat{\overline{Y}}_0$ is the estimator for t_0 based on the entire sample at t_0, $\hat{\overline{Y}}_{0,-1}$ is the estimator for t_0 based on only those samples common to t_0 and the previous time t_{-1}, $\hat{\overline{Y}}_{-1,0}$ is the estimator for t_{-1} based on only those samples common to t_0 and t_{-1}, and $\hat{\overline{Y}}'_{-1}$ is the composite estimator for t_{-1}. The difference D is estimated by

$$\hat{D} = \hat{\overline{Y}}'_0 - \hat{\overline{Y}}'_{-1}$$

Rao and Graham (1964), in an extensive study assuming a finite population, calculated optimum values for Q and gains in efficiency over the simple estimators of level ($\overline{Y}_0$) and change (D). It appeared that when the temporal correlation is high moderate gains (15 - 55 %) for level and large gains (100 - 800 %) for change are achieved. Keeping the sample size constant, the gains for level decrease with increasing number of repeated measurements, while the gains for change increase.

Many different strategies of rotation sampling have been developed, including improved estimation procedures. In some strategies a set of locations would be re-introduced into the sample after having been rotated out for some time. See Binder & Hidiroglou (1988) for a review on rotation sampling.

11.2 Model-based sampling in space-time

Broadly speaking four approaches can be followed for model-based sampling in space-time, using:

- a geostatistical model of variation in space-time (an extension of spatial geostatistics with the time dimension);
- a multivariate time-series model (a generalisation of univariate time-series, the respons being the time-dependent vector of observations at the various sampling locations);
- a regionalized time-series model (a spatial geostatistical model with the parameters of time-series models fitted at the various sampling locations as regionalized variables);
- a space-time Kalman filter (a data assimilation technique with optimality properties, especially useful for short time-series).

All four approaches involve complicated and highly specialized statistical techniques. Therefore a presentation of these methods falls outside the scope of this report, and the reader is referred to the literature: Kyriakidis & Journel (1999) for the geostatistical approach, Box & Jenkins (1976) for time-series analysis, and Anderson & Moore (1979) and Binder & Hidiroglou (1988) for the Kalman filter approach.

The geostatistical approach is generally considered to be less promising than the regionalized time-series and the Kalman filter approach, because it is often difficult to construct a realistic model of the variation in space-time. The same applies to the multivariate time-series approach. Although much depends on what data are available, it seems that in many practical situations the regionalized time-series approach and the Kalman filter approach are the most suitable candidates for model-based inference from sample data.

11.3 Final remarks

Clearly, several important issues received little or no attention in this report. This is partly due to the fact that I could not find much practical methodology to tackle these problems. Therefore I suggest priority for further research on the following themes:

- combining data from probability samples with data from purposive or convenience samples;
- suitable techniques of cartographic presention of monitoring results on *multiple* target variables;
- specialized sampling strategies for locating 'hot-spots';
- efficient design-based rotational monitoring systems;
- regionalized time-series modeling and Kalman filters, using available prior information.

Bibliography

[1] Anderson, B.D.O. and J.B. Moore. 1979. Optimal Filtering. Prentice-Hall, Englewood Cliffs, NJ.

[2] Andricevic, R. 1996. Evaluation of sampling in the subsurface. Water Resources Research 32:863-874.

[3] Angulo, M. and W.H. Tang. 1999. Optimal ground-water detection monitoring system design under uncertainty. J. of Geotechnical and Geoenvironmental engineering 125:510-517.

[4] Baker, A.S., S. Kuo, and Y.M. Chae. 1981. Comparison of arithmatic soil pH values with the pH values of composite samples. Soil Sci. Soc. Amer. J. 45:828-830.

[5] Bayley, G.V. and J.M. Hammersley. 1946. The effective number of independent observations in an autocorrelated time series. J. Roy. Statist. Soc. 8(1b):184-197.

[6] Binder, D.A. and M.A. Hidiroglou. 1988. Sampling in time. In: P.R. Krishnaiah and C.R. Rao (eds.) Handbook of Statistics, Vol. 6. Elsevier Science Publishers, 187-211.

[7] Boswell, M.T., S.D. Gore, G. Lovison, and G.P. Patil. 1996. Annotated bibliography of composite sampling, Part A. Environmental and Ecological Statistics 3:1-49.

[8] Box, G.E.P. and G.M. Jenkins. 1976. Time series analysis, forecasting and control. Revised Edition. Holden-Day, San Francisco.

[9] Breidt, F.J. 1995. Markov chain designs for one-per-stratum spatial sampling. Survey methodology 21:63-70.

[10] Brown, G.H. and N.I. Fisher. 1972. Subsampling a mixture of sampled material. Technometrics 14:663-668.

[11] Brus, D.J. 2000. Using regression models in design-based estimation of spatial means of soil properties. Eur. J. Soil Sci. 51:159-172.

[12] Brus, D.J. and J.J. de Gruijter. 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma 80:1-59.

[13] Brus, D.J., L.E.E.M. Spätjens, and J.J. de Gruijter. 1999. A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. Geoderma 89:129-148.

[14] Cameron, D.R., M. Nyborg, J.A. Toogood, and D.H. Laverty. 1971. Accuracy of field sampling for soil tests. Can. J. Soil Sci. 51:165-175.

[15] Candela, L., R.A. Olea, and E. Custodio. 1988. Lognormal kriging for the asessment of reliability in groundwater quality control observation networks. J. Hydrol. 103:67-84.

[16] Carter, R.E. and L.E. Lowe. 1986. Lateral variability of forest floor properties under second-growth Douglas fir stands and the usefulness of composite sampling techniques. Can. J. For. Res. 16:1128-1132.

[17] Chen, M. and S. Cheng. 1989. Method for the optimum sampling frequency selection in a regular water quality monitoring system. Huanjing Kexue 10:58-63.

[18] Close, M.E. 1989. Effect of serial correlation on ground-water quality sampling frequency. Water Resour. Bull. 25: 507-515.

[19] Cochran, W.G. 1977. Sampling techniques (3rd ed.). Wiley, New York.

[20] Conquest, L.L. 1993. Statistical approaches to environmental monitoring: did we teach the wrong things? Environmental Monitoring and Assessment 26:107-124.

[21] Courtin, P., M.C. Feller, and K. Klinka. 1983. Lateral variability in some properties of disturbed forest soils in southwestern British Columbia. Can. J. Soil Sci. 63:529-539.

[22] de Gruijter, J.J. 1999. Spatial sampling schemes for remote sensing. In: Spatial Statistics for Remote Sensing (eds. A. Stein and F.D. van der Meer), Kluwer Acad. Publ., Dordrecht.

[23] de Gruijter, J.J. and B.A. Marsman. 1985. Transect sampling for reliable information on mapping units. In: Proceedings of the SSSA-ISSS Workshop on Spatial Variability, Las Vegas, U.S.A., Nov. 30 - Dec. 1, 1984 (ed. J. Bouma), Pudoc, Wageningen.

[24] de Gruijter, J.J. and C.J.F. ter Braak. 1990. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. Mathematical Geology 22:407-415.

[25] Dixon, W., and B. Chiswell. 1996. Review of aquatic monitoring program design. Water Research 30:1935-1948.

[26] Domburg, P., J.J. de Gruijter, and D.J. Brus. 1994. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. Geoderma 62:151-164.

[27] Domburg, P., J.J. de Gruijter, and P. van Beek. 1997. Designing efficient soil survey schemes with a knowledge-based system using dynamic programming. Geoderma 75:183-201.

[28] Duncan, A.J. 1962. Bulk sampling. Problems and lines of attack. Technometrics 4:319-343.

[29] Dutta, D., A. Das Gupta, and V. Ramnarong. 1998. Design and optimization of a ground water monitoring system using GIS and multicriteria decision analysis. Ground Water Monitoring and Remediation 18:139-147.

[30] Elder, R.S., W.O. Thompson, and R.H. Myers. 1980. Properties of composite sampling procedures. Technometrics 22:179-186.

[31] Ellis, J.C. 1989. Handbook on the Design and Interpretation of Monitoring Programmes. Publication NS 29, Water Research Centre, Medmenham, England.

[32] Ellis, J.C. and R.F.,Lacey. 1980. Sampling: defining the task and planning the scheme. Wat. Poll. Control 79:452-467.

[33] Gilbert, R.O.1987. Statistical Methods for Environment Pollution Monitoring. Van Nostrand Reinhold Company, New York.

[34] Helsel, D.R. and R.J. Gilliom.1986. Estimation of distributional parameters for censored trace level water quality data: 2. Verification and applications. Water Resour. Res. 22:147-155.

[35] Hipel, K.W., W.C. Lennox, T.E. Unny, and A.I. McLeod. 1975. Intervention analysis in water resources. Water Resour. Res. 11:855-861.

[36] Hipel, K.W. and A.I. McLeod. 1994. Time series modelling of water resources and environmental systems. Elsevier, New York.

[37] Hipel, K.W., A.I. McLeod, and W.C. Lennox. 1977. Advances in Box-Jenkins modeling I. Model construction. Water Resour. Res. 13:567-575.

[38] Hirsch, R.M., R.B. Alexander, and R.A. Smith. 1991. Selection of methods for the detection and estimation of trends in water quality. Water Resour. Res. 27:803-813.

[39] Hirsch, R.M. and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resour. Res. 20:727-732.

[40] Hirsch, R.M., J.R. Slack, and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. Water Resour. Res. 18:107-121.

[41] Hudak, P.F., H.A. Loaiciga, and M.A. Marino. 1995. Regional-scale ground water quality monitoring via integer programming. J. Hydrol. 164:153-170.

[42] Jankovic, I. and R. Andricevic. 1996. Spatial and temporal analysis of groundwater recharge with application to sampling design. Stochastic Hydrology and Hydraulics 10:39-63.

[43] Knopman, D.S. and C.I. Voss. 1989. Multiobjective sampling design for parameter estimation and model discrimination in groundwater solute transport. Water Resour. Res. 25:2245-2258.

[44] Krishnaiah, P.R. and C.R. Rao (Eds.). 1988. Sampling. Handbook of Statistics, Vol. 6. North-Holland, Amsterdam.

[45] Kyriakidis, P.C. and A.G. Journel. 1999. Geostatistical space-time models: a review. Mathematical Geology 31:651-684.

[46] Lettenmaier, D.P. 1976. Detection of trends in water quality data with dependent observations. Water Resour. Res. 12:1037-1046.

[47] Lettenmaier, D.P. 1978. Design considerations for ambient stream quality monitoring. Water Resour. Bull. 14:884-902.

[48] Loaiciga, H.A., R.J. Charbeneau, L.G. Everett, G.E. Fogg, B.F. Hobbs, and S. Rouhani. 1992. Review of ground-water quality monitoring network design. J. of Hydraulic Engineering, 118:11-37.

[49] Loftis, J.C., G.B. McBride, and J.C. Ellis. 1991. Considerations of scale in water quality monitoring and data analysis. Water Resour. Res. 27:255-264.

[50] Loftis, J.C., R.D. Ward, R.D. Phillips, and C.H. Taylor. 1989. An evaluation of trend detection techniques for use in water quality monitoring programs. U.S. Environmental Protection Agency, EPA/600/3-89/037.

[51] Loftis, J.C., R.D. Ward, and G.M. Smillie. 1983. Statistical models for water quality regulation. J. Wat. Poll. Control Fed. 55:1089-1104.

[52] MacKenzie, M.C., R.N. Palmer, and S.P. Millard. 1987. Analysis of statistical monitoring designs. J. Wat. Resour. Plan. Manage. 113:599-615.

[53] Maher, W.A., P.W. Cullen, and R.H. Norris. 1994. Framework for designing sampling programs. Environ. Mon. Assess. 30:139-162.

[54] Mar, B.W., W.S. Mitter, R.N. Palmer, and R.A. Carpenter. 1987. Cost-effective data acquisition. Guidelines for surveying and monitoring watersheds. East-West Environment and Policy Institute. Workshop Report No 1.

[55] Matelas, N.C. and W.B. Langbein. 1962. Information content of the mean. J. of Geophysical Research 67:3441-3448.

[56] Matérn, B. 1986. Spatial Variation (2nd Edition). Springer-Verlag, New York.

[57] McBratney, A.B. and R. Webster. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables: II. Program and examples. Comput. and Geosciences 7:335-365.

[58] McBratney, A.B. and R. Webster. 1983. How many observations are needed for regional estimation of soil properties? Soil Science 135:177-183.

[59] McBratney, A.B., R. Webster, and T.M. Burgess. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables-I Theory and method. Computers and Geosciences 7:331-334.

[60] Mol, G., S.P. Vriend, and P.F.M. van Gaans. 1998. Future trends, detectable by soil monitoring networks? J. of Geochemical Exploration 62:61-66.

[61] Montgomery, R.H. and K.H. Reckhow. 1984. Techniques for detecting trends in lake water quality. Water Resources Bulletin 20:43-52.

[62] Nelson, J.D. and R.C. Ward. 1981. Statistical considerations and sampling techniques for groundwater quality monitoring. Ground Water 19:617-625.

[63] Onate, B.T. 1953. Some statistical aspects of the use of composites in soil sampling. The Philippine Agriculturist 241-257.

[64] Overton, J.McC., T.C. Young, and W.S. Overton. 1993. Using 'found'data to augment a probability sample: procedure and case study. Environmental Monitoring and Assessment 26:65-83.

[65] Patil, G.P. and C.R. Rao (Eds.). 1994. Environmetal statistics. Handbook of Statistics, Vol. 12. North-Holland, Amsterdam.

[66] Quenouille, M.H. 1949. Problems in plane sampling. Ann. Math. Stat. 20:355-375.

[67] Radford, P.J. and J.West. 1986. Models to minimize monitoring. Wat. Res. 20:1059-1066.

[68] Rao, J.N.K. and J.E. Graham. 1964. Rotation designs for sampling on repeated occasions. JASA 59:492-509.

[69] Reed, J.F. and J.A. Rigney. 1947. Soil sampling from fields of uniform and nonuniform appearance and soil type. J. Amer. Soc. Agron. 39:26-40.

[70] Reinelt, L.E., R.R. Horner, and B.W. Mar. 1988. Nonpoint source pollution monitoring program design. J. Wat. Res. Plan. Manage. ASCE 114:335-352.

[71] Rohde, C.A. 1976. Composite sampling. Biometrics 32:273-282.

[72] Ruark, G.A., D.L. Mader, and T.A. Tattar. 1982. A composite sampling technique to assess urban soils under roadside trees. J. Arboriculture 8:96-99.

[73] Särndal, C.E., B. Swensson, and J. Wretman. 1992. Model assisted survey sampling. Springer-Verlag, New York.

[74] Scott, A.J. and T.M.F. Smith. 1974. Analysis of repeated surveys using time series methods. JASA 69:674-678.

[75] Serón Arbeloa, F.J., C. Pérez Caseiras, L.J. Nogué Lahuerta, and P.M. Latorre Andrés. 1993. Air quality monitoring for multple pollutants: optimization of a network around a hypothetical potash plant and two thermal power stations in open countryside. Environmental Monitoring and Assessment 27:107-134.

[76] Sharp, W.E.1971. A topographically optimum water sampling plan for rivers and streams. Water Resour. Res. 7:1641-1646.

[77] Skalski, J.R. 1990. A design for long-term status and trends monitoring. J. of Environmental Management 30:139-144.

[78] Stout B.B. 1993. The good, the bad and the ugly of monitoring programs: defining questions and establishing objectives. Environmental Monitoring and Assessment 26:91-98.

[79] Taylor, C.H. and J.C. Loftis. 1989. Testing for trend in lake and groundwater quality time series. Water Resour. Bull. 25:715-726.

[80] Todd, D.K., R.M. Timlin, K.D. Schmidt, and L.G. Everett. 1976. Monitoring ground water quality: Monitoring methodology. United States Environmental Protection Agency, Las Vegas, Nev.

[81] van Belle, G. and J.P. Hughes. 1984. Nonparametric tests for trend in water quality. Water Resour. Res. 20:127-136.

[82] van de Wetering, B.G.M. and S. Groot. 1986. Water quality monitoring in the state-managed waters of the Netherlands. Wat. Res. 20: 1045-1050.

[83] van Groenigen, J.W. and A. Stein. 1998. Spatial Simulated Annealing for constrained optimization of spatial sampling schemes. Journal of Environmental Quality 27:1078-1086.

[84] Ward, R.C., J.C. Loftis, and G.B. McBride (Eds). 1990. Design of Water Quality Monitoring Systems. Van Nostrand Reinhold, New York.

[85] Ward, R.C., J.C. Loftis, K.S. Nielsen, and R.D. Anderson. 1979. Statistical evaluation of sampling frequencies in monitoring networks. J. Wat. Poll. Control Fed. 51:2292-2300.

[86] Webster, R. and T.M. Burgess. 1984. Sampling and bulking strategies for estimating soil properties in small regions. J. of Soil Sci. 35:127-140.

[87] Webster, R. and M.A. Oliver. 1992. Sample adequately to estimate variograms of soil properties. J. of Soil Sci. 43:177-192.

[88] Williams, L.R., R.W. Leggett, M.L. Espegren, and C.A. Little. 1989. Optimization of sampling for the determination of mean radium-226 concentration in surface soil. Environ. Monitoring and Assessment 12:83-96.

[89] Woldt, W. and I. Bogardi. 1992. Ground water monitoring network design using multiple criteria decision making and geostatistics. Water Resour. Bull. 28:45-62.

[90] Yates, F. 1981. Sampling methods for censuses and surveys. Griffin, London, 4th ed.

[91] Zhou, Y. 1996. Sampling frequency for monitoring the actual state of groundwater systems. J. of Hydrology 180:301-318.