3

Basic Design Principles

The aim of this chapter is to offer help in designing schemes for survey or monitoring. To this end, we present in the following sections seven principles that we consider essential for good design. These principles are:

1. Develop a complete scheme (Sect. 3.1).
2. Structure the design process (Sect. 3.2).
3. Pay ample attention to practical issues (Sect. 3.3).
4. Employ prior information on variation (Sect. 3.4).
5. Balance the various sources of error (Sect. 3.5).
6. Anticipate changing conditions during monitoring (Sect. 3.6).
7. Calculate the sample size appropriately (Sect. 3.7).

3.1 Develop a Complete Scheme

Survey and monitoring of natural resources often involves the following activities.

- Planning field activities: given the purpose of the project, the budget and possible logistical constraints, it is decided how many samples and/or field measurements are to be taken, as well as where, how and when.
- 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 a survey project. In the case of monitoring, field and laboratory work, as well as data recording and processing are obviously 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, the process may involve switching back and forth between activities. For instance, if some deficiency is discovered during data processing, additional field work may be needed. Laboratory work is optional, as measurements may be taken in the field.

The main purpose of this section is to argue that, although the above sequence of activities may seem reasonable, it does not make good sampling practice. The reason is that an essential element is missing at the beginning of the sequence: the element of planning the whole chain of activities, including the statistical procedure to be used in data processing. Careful planning of the entire project is a prerequisite of good sampling practice and should precede any other activity. Whereas researchers usually invest enough effort and ingenuity in deciding how, where and when to take samples and measurements, their ideas about how to analyze the data very often 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 attuned 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 is found that the variogram needed for this cannot be reliably estimated from the data. Finally, a common situation 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, like most legacy soil survey data. These examples will be discussed in greater detail in the next sections.

Apart from attuning 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 by optimizing parts of it in isolation: the consequences of a decision about a particular issue, in terms of quality and costs, depend on the decisions taken on other issues. A simple example is where two assessment methods are available for the target variable: an inexpensive but inaccurate method and an expensive but accurate method. The choice between the two affects 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 used, which may or may not lead to a better result.

In summary, we recommend planning not only the data acquisition but the entire project, paying special attention to the agreement between data acquisition and data processing. Proper planning of the entire project will most likely pay itself back by increased efficacy and efficiency. We want to emphasize this principle by referring to the entire plan as the *scheme*. Our broad concept of scheme covers much more than just how, where and when to sample and measure. A scheme captures explicitly all the decisions and information pertinent to data acquisition, data recording and data processing. It consists of the following items.

1. Detailed analysis and specification of the *objective*.
   a) *Target universe*: a precise definition of the universe of interest, with boundaries in space and/or time, and possibly a specification of exclusions. (Note that, for various reasons, the target universe may differ from the actually *sampled universe*.) In case of ecological populations, a decision on how the universe will be treated, as continuous or as discrete (Sect. 4.2).
   b) *Domain(s)* of interest: a specification of the part(s) of the universe for which a separate result is required. At one extreme (in terms of extent), this is the entire universe, at the other extreme it is a point or a set of points in the universe (e.g., grid nodes used to prepare a contour map). In between these extremes, there may be a number of domains with smaller or larger extents in space and/or time. Examples include sub-areas within a spatial universe, or spatial cross-sections through a space–time universe (i.e., the space at given points in time).
   c) *Target variable(s)*: a precise definition of the variable(s) to be determined for each of the sampling units. (Note that target variables are generally not identical with actually *measured variables*, because of measurement errors or transformations of measured variables prior to statistical inference.)
   d) *Target parameter*: the type of statistic for which a result is needed, given the target variable(s) and the domain(s). Examples include total, mean, fraction, median, standard deviation or trend parameter.
   e) *Target quantity*: the combination of a domain, target variable and target parameter is referred to in this book as a target quantity. An example is the mean (parameter) phosphate content in the topsoil (target variable) of the agricultural soils in the Netherlands (domain). A target quantity that is related to the entire universe is referred to as a 'global quantity'; in all other cases it is referred to as a 'local quantity'.
   f) *Type of result*: quantitative or qualitative. If a quantitative result is desired, then the mode of inference has to be *estimation* or *prediction*. If a qualitative result is needed, e.g., an answer to the question whether or not the target quantity exceeds a given level, then the mode of inference has to be *testing*, *classification* or *detection* (Sect. 2.2).
2. *Quality measure*: the quantity used to express numerically the statistical quality of the survey or monitoring result (Sect. 5.1). Examples include the half-width of a 95% confidence interval in estimation, the error variance in prediction, the power in hypothesis testing, the error rate in classification.
3. *Constraints*: the allocated budget and/or minimum quality of the result, fieldwork (optional), transport (optional) and laboratory capacity (optional).

4. *Prior information* (Sect. 3.4).
   a) *Sampling frame*: the list, file or map identifying the sampling units from which the sample is selected.
   b) *Miscellaneous information*: general knowledge, experience and information from comparable projects, existing sample data, maps or GIS files.
   c) *Model of the variation* of the target variable within the universe (eventually needed if model-based inference is chosen; see item 8). Examples include a variogram adopted from a comparable case or estimated from a preliminary sampling round (Chap. 9).
5. *Sample support*, in the case of a continuous universe (Sect. 4.2); physical sampling devices for taking aliquots (optional).
6. *Assessment method*: field and/or laboratory measurement procedures (reference to existing protocols where possible); method of data pre-processing to calculate the target variable from measured variables (optional).
7. Whether and how to use *composite sampling*, i.e., bulking aliquots (Sect. 4.3).
8. Choice between *design-based* and *model-based inference* from sample data (Sect. 4.1).
9. For design-based inference: choice of random sampling *design type* and *attributes* of the chosen design type (Sect. 5).
10. For model-based inference: choice of *sampling pattern type* and *optimization* algorithm and restrictions (optional), (Sect. 5).
11. Identification of the actually selected *sample*: a list of sampling unit labels, a map with sampling locations, a table of sampling times or coordinates of sampling events in space–time.
12. *Protocols* on data recording and field work (Sect. 3.3).
13. Method to be used for *statistical inference*.
14. *Prediction* of operational *costs* and *quality* of results: ex-ante evaluation (Sect. 3.3).

The scheme item 'target parameter' deserves more attention before we continue with examples of a scheme. Parameters may be related to either a frequency distribution or a probability distribution. Frequency distributions are integrals over space and/or time, which of course applies only to non-point domains, i.e., domains with an extent. Parameters related to frequency distributions are the total, mean, mode, standard deviation, percentiles (e.g., the median) and fractions. (A fraction is the relative size of that part of the domain where a given state is present, such as exceedance of a threshold value). Examples include an areal fraction, a temporal mean and a spatio-temporal total. We refer to such parameters as 'frequential parameters'. Probability distributions, on the other hand, are integrals of probabilities defined by a stochastic model. Parameters usually related to probability distributions are the expected value, mode, standard deviation, percentiles (e.g., the median) and probabilities of exceedance. Such parameters are referred to as 'probabilistic parameters'. They may be defined for the target variable at a given point of the universe but also for frequential parameters, which unfortunately complicates matters. For instance, the chosen parameter may be the probability that the spatio–temporal total of the target variable over a given domain exceeds a given threshold value.

To illustrate our concept of a survey or monitoring scheme, we give two hypothetical examples. In order to avoid lengthy descriptions and details not required for a correct understanding, we present the examples in a concise form. In a real project, they should of course include the full details.

Example I, 2D survey

1. *Objective.*
   a) *Target universe*: the topsoil of all arable fields of a specified class in region R at time T, a continuous universe.
   b) *Domain*: the entire region.
   c) *Target variable*: a variable indicating whether or not the (true) mean cadmium concentration in the topsoil at a location in the region exceeds critical concentration level C (0 means 'no'; 1 means 'yes'). This 'indicator variable' is needed because the corresponding areal fraction is to be used as the target quantity (see item 1e).
   d) *Target parameter*: the spatial mean.
   e) *Target quantity*: the areal fraction of the arable fields in region R where at time T the (true) cadmium concentration in the topsoil exceeds level C. (Note: sampling is done over a relatively short period, during which the concentrations are assumed to be constant.)
   f) *Type of result*: qualitative result, accepting or rejecting the null-hypothesis that the areal fraction is below a given critical level F, by testing at the 95% confidence level.
2. *Quality measure*: the power of the test (the probability of rightly rejecting the null-hypothesis), when the actual areal fraction exceeds the critical level F by a given amount.
3. *Constraints*: the power must not be less than a given value (a quality requirement). This implies that the scheme should aim to minimize the costs while satisfying the quality requirement.
4. *Prior information*:
   a) *Sampling frame*: a GIS file containing the boundaries of the arable fields in region R. (Note that this may not be an error-free representation of the target universe.)
   b) *Miscellaneous information*: a map showing expected cadmium concentrations, compiled from local knowledge of pollution sources.
   c) *Model of the spatial variation*: not available.
5. *Sample support*: standard auger core from the topsoil.
6. *Assessment method*: a specified laboratory method for concentration (the measured variable), followed by 0/1 transformation of the cadmium con-

centrations, with 0 for no exceedance and 1 for exceedance of critical level C.

7. *Composite sampling?* No Composite sampling. (Note: bulking aliquots would in this case lead to biased estimates of the target variable; see Sect. 4.3.)
8. *Design-based or model-based inference?* Design-based inference.
9. *Design type* and *design attributes*: Stratified Simple Random Sampling (design type), with specified strata (attribute 1) derived from the map of expected concentrations, and specified sample sizes in the strata (attribute 2) chosen to ensure that costs are minimized while the quality requirement is satisfied.
10. (For model-based inference: not applicable)
11. *Sample*: reference to a map indicating the sampling locations, selected according to the chosen design type and design attributes.
12. *Protocols*: reference to documents.
13. *Method of inference*: standard one-sided test of the null-hypothesis, with the assumption that the estimated areal fraction is approximately normally distributed (the sample size should be large enough).
14. *Ex-ante evaluation*: reference to a document.

Example II, monitoring in 2D space and time

1. *Objective.*
   a) *Target universe*: the water passing through a cross-section S of a river during year Y, a continuous universe.
   b) *Domains*: the water passing through the cross-section during the summer and during the winter.
   c) *Target variable*: the (true) amount of salt passing through $1\,\mathrm{m}^2$ of the cross-section during 1 min.
   d) *Target parameter*: spatio–temporal total.
   e) *Target quantities*: the total salt loads passing through cross-section S during the summer and the winter of year Y.
   f) *Type of result*: quantitative, as 95% prediction intervals.
2. *Quality measure*: half-width of the widest of the two prediction intervals.
3. *Constraints*: a limited budget is available, which implies that the scheme should aim at minimizing the half-width of the widest prediction interval while keeping the costs within budget; due to shipping traffic on the river, sampling is confined to a specified acceptable zone of the cross-section.
4. *Prior information*:
   a) *Sampling frame*: a map of cross-section S, conceptually combined with the continuous time scale of a year with 365 days.
   b) *Miscellaneous information*: sample data from the same river but for previous years.
   c) *Model of the variation*: a space–time variogram of the salt flux was developed from the available sample data.

5. *Sample support*: implicitly defined by the assessment method.
6. *Assessment method*: a specified sensing method to measure salt concentration and a specified sensing method to measure water flux (two measured variables), followed by multiplication of the salt concentrations with the water fluxes to calculate the target variable.
7. *Composite sampling?* No composite sampling (measurement in situ; no aliquots are taken).
8. *Design-based or model-based inference?* Model-based inference.
9. (For design-based inference: not applicable)
10. *Sampling pattern type and optimization*: a number of fixed sampling locations where measurements are taken simultaneously at equidistant sampling times (sampling pattern type: regular grid in space–time); optimization through evaluation of costs and quality measure for all combinations of eligible sampling locations (e.g., from a $1 \times 1\,\mathrm{m}^2$ grid on the acceptable zone of the cross-section) and eligible sampling time intervals (e.g., in a series of one day, one week, two weeks, four weeks, 13 weeks).
11. *Sample*: a map indicating the sampling locations in the cross-section and a table of sampling times.
12. *Protocols*: reference to documents.
13. *Method of inference*: calculation of the 95% prediction interval for the salt loads from space–time block-kriging predictions and associated kriging variances.
14. *Ex-ante evaluation*: reference to a document.

3.2 Structure the Design Process

If designing a scheme is regarded as an instance of problem solving, items 1–4 can be seen as the information which is used to find a solution: the 'design information'. From this strict perspective, items 5–13 together constitute the selected solution, and the final item (14) is an ex-ante evaluation of that solution. From a broader perspective, the design information, especially the items 'objective', 'quality measure' and 'constraints' will already be the result of a 'translation' of an initial, more general description of the aim of the project. This translation typically settles various details that were left undecided until then. This can usually be done in different ways, the alternatives having a potential effect on the costs and quality of the results. Therefore, we consider it to be a fundamental part of the design process as a whole. Clearly, the translation requires close interaction with and agreement from the stakeholders. It is probably not unusual that, at some point in the design process, parts of the translation will have to be reconsidered and repeated.

A safe way to obtain a good scheme is based on the following 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 has been defined it does become useful to search for a scheme that satisfies this need in an efficient way. The reason for this is that different information needs generally ask for different schemes. Although this is one of the most important facts in sampling, it does not seem to be always clearly acknowledged. We shall therefore discuss this in more detail.

Information needs in the context of survey or monitoring can be divided into two broad groups. In the first group, the purpose may be to estimate a global quantity, i.e., a parameter of the *cumulative distribution function* of the target variable over the entire universe. Examples are 'location'[1] parameters such as the mean, quantiles (e.g., the median) and the mode, and 'dispersion' parameters such as the standard deviation, range and tolerance intervals.

In the second group, the purpose may be some kind of description of the *spatial and/or temporal distribution* of the target variable within the universe. Examples are prediction of values at specific points within the universe, estimation of means within parts of the universe and construction of contour maps.

In general, different types of results ask for different sampling designs, because a given scheme may not yield the type of result that is required, or it may do so in an inefficient way. For instance, estimating the spatial mean of a region (a global target quantity) requires other, less expensive schemes than the prediction of the values at grid nodes (local quantities), as is done for mapping. In conclusion, a good way of designing a scheme is by reasoning backwards through the following steps:

1. Decide precisely what information is needed. Examples include a map of a given variable, at a given scale and with a given accuracy, or testing a given hypothesis, at a given confidence level and with a given statistical power.
2. Identify the constraints that apply to the production of the required information.
3. Identify what useful information is already available.
4. Determine what kind of data analysis leads to the required type of result.
5. Identify the data needs for this analysis and search for a strategy to obtain these data in the most efficient way.

This sequence of steps implies that, prior to making any design decision, one should first collect all design information. The reason for this is that the design decisions are otherwise likely to be premature and need to be reconsidered once the design information has been made more complete or explicit.

After all design information has been collected, the question remains how to organize the rest of the design process. More specifically, in what order should the various other items of the scheme be decided? It would be unfeasible to provide a detailed design protocol that is suitable for all possible circumstances. Nevertheless, the following global guidelines seem to be useful:

[1] In this context 'location' does not refer to position in geographical space, but to position on the measurement scale.

1. Make a provisional decision on the assessment method (scheme item 6).
2. Choose a quality measure (scheme item 2, see Sect. 5.1).
3. Make a rough estimate of the sample size affordable at a given budget or needed to meet a given quality requirement.
4. Make provisional decisions on the following major design issues:
   a) Choice between design-based and model-based inference (scheme item 8; see Sect. 4.1).
   b) Choice of sample support (scheme item 5; see Sect. 4.2).
   c) Choice of whether and how to bulk aliquots (scheme item 7; see Sect. 4.3).
5. In the case of design-based inference: search for an optimal random sampling design (scheme item 9; Sect. 5.2.1).
6. In the case of model-based inference: choose a sampling pattern type and optimization technique (scheme item 10), such as simulated annealing and genetic algorithms (Appendix A), and optimize the sample.
7. Make a prediction of operational costs and quality of results (scheme item 14). If the expected costs are too high or the quality too low, then revise one or more of the provisional design decisions. Otherwise, the decisions are final.
8. Draw a random sample according to the chosen sampling design, or optimize a sample (scheme item 11).
9. Work out the method of statistical inference (scheme item 13).
10. Develop protocols on field work and data recording (scheme item 12).

The reasons for making the major design decisions (mode of inference, sample support and aliquot bulking) early in the process is that these decisions tend to have dominant effects on both costs and quality, and that most other decisions depend on them.

The scheme above assumes a single target quantity. In practice, especially in monitoring situations, one has multiple target quantities, which makes optimization a more complex problem, further discussed in Sect. 5.2.

Design processes are seldom linear and one-way. There are often good reasons to loop back to earlier provisional design decisions, or even to the design information, e.g., to switch to a less demanding aim, to relax a constraint, or to search for other prior information. Our advice is to keep track of the process to prevent it from becoming haphazard or chaotic, and also to enable reporting of the reasons for the various choices that are finally made.

3.3 Pay Sufficient Attention to Practical Issues

Designing a scheme for survey or monitoring is not just a matter of statistical methodology. On the contrary, if the practical issues are disregarded, there is a high risk that the project will be unsuccessful. Therefore, without pretending to be exhaustive, we discuss what would seem to be the main practical issues.

Avoid Undue Complexity

Researchers often know a great deal about the physical processes that generate spatial patterns or time series of 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. Although understandable, this attitude entails three risks which are easily underestimated.

First, 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 has been adjusted. This may be time-consuming and is likely to cause undesirable delay.

Second, 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 outweighed by the practical difficulties. Also, multiple target variables may be of interest, and one may face the problem that an efficient design for one target variable can be inefficient for another.

Third, complex sampling designs can be efficient for one target variable, but inefficient for another variable. For instance, stratification of the target area may lead to increased precision for a target variable related with the stratification variable, but for target variables that are not related, there may be no gain in precision or even a loss of precision. Therefore, for surveys and monitoring with multiple target variables we recommend keeping the sampling design as simple as possible, and using instead the ancillary information at the estimation stage, for example by using the post-stratification estimator (Sect. 7.2.11).

Allow for Unexpected Delay in Field Work

Even if one is familiar with the circumstances in the terrain, there may be factors beyond one's control that prevent the field work from being completed within the available time. Clearly, unfinished field work may seriously harm the statistical potential of the design. It is therefore prudent to allow some extra time in the scheme for contretemps, say 20 % of the total time for field work, and to include a number of optional sampling locations to be visited as the extra time allows.

Include a Test Phase if Necessary

If there is significant uncertainty about the logistics of the field work or the spatial or temporal variability, a preliminary test phase is always worth the extra effort. The data from even a small sample, collected prior to the main sample, enables the latter to be optimized more precisely and reduces the risk that the project will not meet its goal at all. In the final statistical analysis, the data from the test phase can often be combined with those for the main sample, so that the additional effort is limited to extra travel time and statistical analysis.

Evaluate the Scheme Beforehand

It is good practice to quantitatively predict the operational costs of the scheme, and the accuracy of the result, prior to starting the field work. Predicting cost and accuracy can be done in sophisticated ways, using mathematical models (Domburg et al., 1994), or more generally, using experience from similar projects, rules-of-thumb and approximations. A test phase will of course improve the prediction of costs and accuracy.

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

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 indicate what should be done if a sampling location 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 location there is no 'soil' according to an intended 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 location 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, may result in biased estimates.

Protocol for Data Recording

Just 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 location falls outside the universe, when it is inaccessible, when a variable cannot be measured because its value is too large or too small ('censoring' in the statistical sense), and when a variable cannot be measured for other reasons.

3.4 Employ Prior Information on Variation

Any prior information about the variation in the universe should be utilized as good as possible 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 theories about the spatial and/or temporal patterns of variation in the universe. Such theories may be available in a verbal, qualitative form, or in the quantitative form of a mathematical model.

There are many ways in which prior information can be exploited in schemes. Two modes can be distinguished. 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 we give examples of each mode.

An example of the first mode is when images, photographs or maps are used to stratify the universe. In this case, the universe is split into a number of relatively homogeneous sub-universes (called 'strata'), which are then sampled independently (Sect. 7.2.4). Another example is when genetic theory enables intelligent guesses about spatial and/or temporal correlations. For instance, in the case of a universe consisting of the soil in a given area, aeolian deposition of parent material in that area may be known to have resulted in little short-range variation of texture. If the target variable is closely related to texture, it will be then important for the sake of efficiency to avoid sampling at locations in close proximity. A final example of the first mode is when a variogram (Chap. 9) is used to optimize the sampling density or sampling frequency.

An example of the second mode is when prior point data are used to design a space-filling sample, and prior and new data are used in spatial interpolation. Another example is the use of ancillary data in regression estimators (Sect. 7.2.11). Brus and de Gruijter (2003) present a method for using prior point data from non-probability sampling in design-based estimation of spatial means.

If prior information on the variation is captured in the form of variograms, these functions can be used to predict the sampling variance for a given design (Sect. 7.2.15). If in addition a model of the costs is available, then it is possible to optimize the sampling design in a fully quantitative manner (Domburg et al., 1997).

3.5 Balance the Sources of Error

It is important to realize that the accuracy of the final result is not only determined by the sampling error, i.e., the error due to the fact that sampling is limited to a finite number of units. Examples of other sources of error are sample treatment, observation, model of the variation, 'censoring' and 'non-response'. Censoring means that no quantitative measurement is possible on a particular sampling unit, because its true value falls outside the measurable range of the measuring device used for measuring. An example would be a situation in which the water table is to be measured in an auger hole with a depth of 1.5 m depth and no groundwater is observed in the hole. This indicates that the level is deeper than 1.5 m ('right censoring'), and a quantitative assessment can only be produced by guessing or by statistical estimation based on an assumed distribution function. (See Knotters et al. (1995) for an example of the latter.) Another example is where the true value of a concentration is below the detection limit of a chemical analysis ('left censoring').

Non-response is a term used in the general statistical literature to indicate the situation where for some reason no data can be obtained from a sampling unit. In groundwater, soil and vegetation sampling this occurs when a location in the field cannot be visited or when measurement is impossible for a different reason than censoring, e.g., loss of an aliquot.

When the inference from the sample data is based on a model of the spatial variation, this model will generally be a source of error, because the underlying assumptions deviate from reality (see Sect. 4.1).

In many cases the target variable can not be observed without error. Examples are measurements of chemical and physical properties on soil- and water-aliquots. Also, in surveys of elusive populations of plants or animals the observer is typically unable to detect every individual in the neighbourhood of the sampling location or line-transect.

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 groundwater in a given area during a given period. One possible strategy would be to measure the concentration of the pollutant in the soil moisture at the sampling locations, to estimate the mean concentration from these data, and to multiply the mean concentration with the total groundwater 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 groundwater. A solution is measuring not only the concentration at the sampling locations but also the flux to the groundwater, and to take the product of these two as the target variable.

Although any reduction of the sampling error will lead to a smaller total error, there is little point in investing all efforts in further reduction of the sampling error if another source of error has a higher order of magnitude. Therefore, in devising a scheme, the relative importance of all error sources should be taken into consideration. For instance, if the spatial variation within a sampling unit (plot) is small compared to that within the domain, it does not pay to take many aliquots in the selected plots to estimate the means of the plots. The optimal number of aliquots in a plot also depends on the time needed for taking the aliquots.

See Gy (1979) for a comprehensive theory of error sources in sampling, especially sampling of particulate materials.

3.6 Anticipate Changing Conditions During Monitoring

An extremely important difference between survey and monitoring with respect to scheme design is that survey 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 alter. With monitoring, on the other hand, not only the universe may undergo large, unexpected changes, but especially in long-term monitoring the conditions will often alter in a way that makes adaptation of the scheme inevitable or at least desirable. Budgets may vary from one year to another, and operational constraints that were originally present may be relaxed, or new unforeseen ones may come into force.

It is also quite common that the focus is shifted, or that new objectives are defined, e.g., through the introduction of new target variables or domains of interest. Better measurement techniques may become available and, last but not least, spatial variation patterns often change in time. For instance, the spatial variation within strata, as originally defined, may increase to a level that makes stratified sampling on the basis of these strata no longer efficient.

One important condition that will always change during monitoring is the amount of available data: more and more data about the universe will become available through monitoring itself. Thus, the available knowledge about the variation in space and/or time will accumulate to a higher level than the prior information that was used to design the scheme. This in itself may be a good reason for fine-tuning or redesigning the scheme.

All changes mentioned above may call for specific adaptations of the scheme, but the point is that some schemes do not lend themselves well to particular adaptations. For instance, suppose that the target area has been divided into small strata, with only one (permanent) sampling location in each, that the statistical inference will be design-based, and that after some years the total number of locations must be reduced due to budget cuts. One then has to choose between (a) leaving some strata unsampled, which leads to biased results, or (b) switching to a new stratification with fewer strata and newly selected sampling locations within them. This may lead to extra costs, and to less precise estimates of change over time.

Clearly, both options are undesirable. The chosen type of sampling design does not survive any reduction of the sample size, i.e., it has no flexibility in this respect. This trap might be avoided, for instance, by using a non-stratified type of design or by defining fewer strata, allowing for more sampling locations in each. Such a choice may yield less precise results at the original budget, but the expected loss of initial precision may be less important than greater adaptability to changing budgets. As for surveys with multiple target variables (see Sect.3.3), we recommend strongly to avoid complex sampling designs for selecting the sampling locations of a monitoring scheme. See Overton and Stehman (1996) for a discussion of desirable design characteristics for long-term monitoring.

More generally, it would be unwise to limit the ex-ante evaluation of long-term monitoring schemes to cost and quality based on the initial design information. Different 'what–if' scenarios in terms of changes in conditions and possible adaptations to such changes should be worked out before final decisions are made.

The fact that monitoring, especially long-term monitoring, is bound to face changing conditions calls for flexibility of the scheme. This flexibility is largely determined by the installation costs of the monitoring locations. When these costs constitute a considerable part of the total costs of monitoring, one will be reluctant to move at the next sampling round to other locations, leading to a *static* or a *static-synchronous* sampling pattern, see Sects. 14.1 and 14.2 for a discussion of these pattern types.

3.7 Calculate the Sample Size Appropriately

It is perfectly obvious that in scheme design a correct formula for the sample size must be used, so why should we stress this by presenting it as a design principle? The reason is that we have repeatedly encountered cases in the literature where an incorrect formula was used, sometimes leading to a much larger sample than necessary and a waste of time and money. We discuss three different kinds of mistake in calculating the sample size.

Design Effect Disregarded

A mistake sometimes made in design-based sampling is to use a sample size formula intended for Simple Random Sampling (Eq. 7.8 or 7.9), when a different sampling design will be used, such as Stratified Simple Random Sampling (Sect. 7.2.4). By doing this, the effect of the chosen sampling design on the sampling variance, compared with Simple Random Sampling, is disregarded and as a result the calculated sample size may be either too large or too small. For instance, if stratification is applied, the design effect is usually a reduction of the sampling variance. Disregarding this effect by using a sample size formula for Simple Random Sampling would then lead to a sample that is larger than necessary.

Sample size calculation specifically for designs other than Simple Random Sampling may be problematic in practice, because it needs prior information that is difficult to obtain. In that case one can deliberately choose to calculate the sample size as if Simple Random Sampling would be applied. If one expects a positive design effect (reducing the sampling variance compared with Simple Random Sampling), one can either accept the calculated sample size as conservative estimate (a number on the safe side), or one can correct it with a reduction factor equal to a prior estimate of the design effect based on experience in comparable projects. If a negative design effect is to be expected, for instance when Cluster Random Sampling or Two-Stage Random Sampling is adopted for operational reasons, then it is especially important to correct the calculated sample size with a prior estimate of the design effect, in order to avoid undersized sampling.

Autocorrelation Disregarded

When sample data are collected not by random but by purposive selection, they should be analyzed by model-based inference, such as block-kriging for prediction of the spatial mean. From a model-based point of view the observations will usually be auto-correlated, which makes the prediction error variance smaller than when no autocorrelation exists. In that sense, kriging takes advantage of auto-correlation. However, if one calculates the sample size from the assumption that there is no autocorrelation (in other words: assuming a pure nugget variogram), while in reality there is, then this advantage is not accounted for. The result is an oversized sample.

Estimation of Model Mean Instead of Spatial or Temporal Mean

A pitfall also worth mentioning here is using a formula that is appropriate for estimating a model mean but not for a spatial or temporal mean. To explain this, consider the variance of the predicted mean of some target variable Z over a universe. Suppose we have n observations on Z, where z satisfies a model with mean μ plus a random component ϵ with variance σ^2:

$$Z_i = \mu + \epsilon_i \qquad\qquad (i = 1, \ldots, n) \tag{3.1}$$

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

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Z_i \ , \tag{3.2}$$

then, if the observations are independent, the variance of this estimator is given by the well-known formula:

$$V_{\text{ind}}(\hat{\mu}) = \frac{\sigma^2}{n} \ . \tag{3.3}$$

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

$$V_{\mathrm{dep}}(\hat{\mu}) = \frac{1}{n^2}\left\{\sum_{i=1}^{n}\sigma^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{n}C(z_i, z_j)\right\} = \frac{\sigma^2}{n}\left\{1 + (n-1)\bar{\rho}\right\} \ , \quad (3.4)$$

where $\bar{\rho}$ 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 (3.4):

$$n_{\mathrm{eq}} = \frac{n}{\{1 + (n-1)\bar{\rho}\}} \ . \quad (3.5)$$

This formula for equivalent sample size has become rather popular and is applied in time-series analysis (Lettenmaier, 1976; Matalas and Langbein, 1962; Zhou, 1996) as well as in spatial statistics, for instance in Gilbert's book on ecological monitoring (Gilbert, 1987). The formula is entirely correct, but if one looks at what happens with the variance of the mean when the sample size is increased, an odd behaviour can be noticed. Take as a simple example an equidistant time series with the exponential autocorrelation function $\rho(t) = \mathrm{e}^{-3t}$ (see Fig. 3.1). Furthermore, take both σ^2 and the monitoring period equal to 1, and increase the sample size by increasing the sampling frequency.

Using (3.3) and (3.4), respectively, for independent and dependent observations we obtain the variance of the estimated mean ($\hat{\mu}$) as a function of sample size, depicted in Fig. 3.2. This figure shows that, with independent observations, the variance decreases continuously 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 other words, according to (3.4) one cannot reach a precision beyond a certain level no matter how many observations one takes. This counters the intuition that the larger the sample, the more one knows about the universe. The reason for this is not that (3.4) is incorrect, but that it is intended for estimating the model mean, not the integral mean, i.e., the average of z over the universe of interest:

$$\bar{z} = \frac{1}{|\mathcal{U}|}\int_{u\in\mathcal{U}} z \,\mathrm{d}u \ . \quad (3.6)$$

If the integral mean (spatial, temporal or spatio-temporal) is to be estimated or predicted, then (3.5) is not applicable and a different formula should be applied, which depends on how the sample will be taken. Following the design-based approach, with some form of random sampling to estimate the integral mean, the correct sample size formula depends on the chosen type of sampling design (see Sect. 7.2). For instance, if Simple Random Sampling is chosen as type of design, then (7.8) or (7.9) should be applied.

Using the model-based approach, the variance of the prediction error of the Best Linear Unbiased Predictor of the integral mean is given by (2.16), which can be used to calculate the required sample size via some optimization routine (see Sect. 7.3 and Appendix A).

It follows from the above example that it is important to choose the target quantity carefully. Just 'the mean' is not enough; the kind of mean is what counts for the sample size. We expect that for surveys the integral mean rather than the model mean would usually be relevant, because it reflects directly the actual state of the universe. The same applies for status and compliance monitoring. For effect and trend monitoring, on the other hand, the model mean may be more relevant.

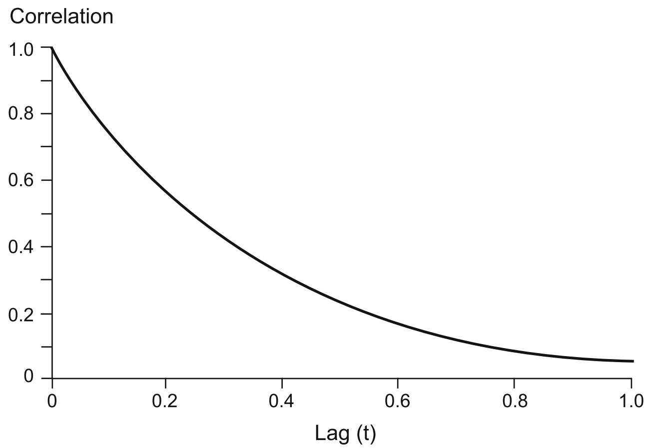

Fig. 3.1. Autocorrelation function used to calculate the variance of the estimated mean in Fig. 3.2

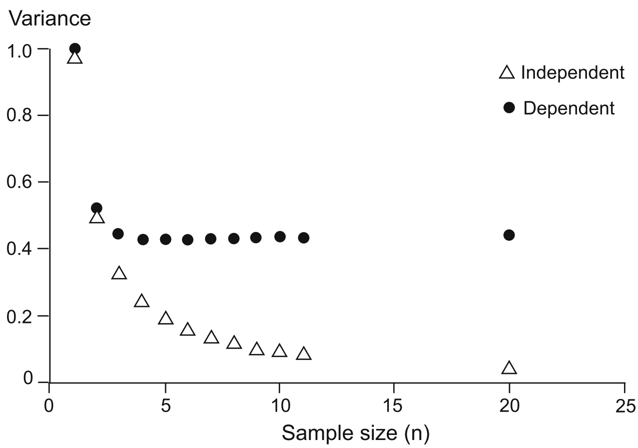

Fig. 3.2. Variance of the estimated mean as a function of sample size, for independent and dependent observations (see text)