WATER RESOURCES RESEARCH, VOL. 44, W11410, doi:10.1029/2007WR006123, 2008

Sampling design for compliance monitoring of surface water quality: A case study in a Polder area

D. J. Brus and **M. Knotters**
Alterra, Soil Science Centre, Wageningen University and Research Centre,
Wageningen, Netherlands

Abstract

[1] International agreements such as the EU Water Framework Directive (WFD) ask for efficient sampling methods for monitoring natural resources. In this paper a general methodology for designing efficient, statistically sound monitoring schemes is described. An important decision is the choice between a design-based and a model-based method, implying the choice between probability (random) sampling and purposive sampling. For mapping purposes, model-based methods are more appropriate, whereas to obtain valid results for the universe as a whole, such as in testing water quality standards against legal standards, we generally prefer a design-based method. Four basic sampling patterns in space-time universe are described: static, synchronous, static-synchronous, and rotational. A case study is carried out for monitoring the quality of surface water at two farms in western Netherlands, wherein a synchronous sampling design is applied, with stratified simple random sampling in both space and time. To reduce laboratory costs the aliquots taken at the locations of a given sampling round are bulked to form a composite. To test the spatiotemporal mean N-total concentration during the summer half-year against the MAR standard with a power of 80% at a concentration 15% below the MAR standard and with a confidence of 95%, six to nine sampling rounds are needed with 50 to 75 locations per sampling round. For P-total the required number of sampling rounds differs strongly between the two farms, but is for both farms much larger than for N-total.

Received 17 April 2007; revised 8 February 2008; accepted 30 July 2008; published 8 November 2008.

Keywords: design-based sampling, Water Framework Directive, statistical testing.

Index Terms:
1848 Hydrology: Monitoring networks; 1871 Hydrology: Surface water quality; 1873 Hydrology: Uncertainty assessment (3275).

1. Introduction

[2] The implementation of international agreements such as the Kyoto Protocol [*United Nations*, 1992] and, at the European level, the Habitats Directive [*Council of the European Communities*, 1992], the Water Framework Directive [*Council of the European Communities*, 2000; *Blöch*, 2001], and the Soil Thematic Strategy [*Commission of the European Communities*, 2002] ask for efficient sampling methods for monitoring natural resources such as biotic populations, groundwater, surface waters and soil. The monitoring aspects of the Water Framework Directive have been worked out in a guidance manual [*Water Framework Directive Common Implementation Strategy Working Group 2.7 Monitoring*, 2003]. Monitoring can be defined as collecting information on an object through repeated or continued observation in order to determine possible changes in the object. The monitoring object may or may not involve a spatial extent. If it does, then observations can be collected via sampling in space-time, the subject of this paper. An example of monitoring an object without spatial extent is to sample repeatedly over time at a particular location of a river where the water quality or level is measured.

[3] *de Gruijter et al.* [2006]

present several basic principles and major decisions for designing monitoring schemes. In this paper we will briefly summarize some of these principles and decisions in developing efficient strategies for natural resource monitoring. In the second part of this paper we will illustrate the proposed methodology with a case study in the field of water resources management: the development of a strategy to monitor the quality of the surface water at dairy farms in the Dutch peat district. The purpose is to test at a farm level whether water quality complies with the standards of the European Water Directive Framework or not, and to conclude whether farm management required intervention in order to obey to these standards. We will explain why we preferred a full design-based approach, involving probability sampling both in space and time, for the purpose of testing water quality standards against legal standards. To our knowledge the application of a full design based approach in testing surface water quality against legal standards, as we illustrate in this paper, is new. The case study demonstrates how decisions on the sampling design for monitoring can be based on statistical inference, and how new information on the variation of the target variable collected in a first monitoring project can be used to update the sampling design.

2. Designing a Monitoring Scheme

[4] When designing a monitoring scheme many decisions must be taken, the most obvious of which concern the boundaries of the area and the period of time to be monitored (the boundaries of the so called universe), the number of sampling locations, the sampling frequency, and where and when to take samples [*Harmancioglu et al.*, 1999]. However, as will be shown, there is much more to decide on. All these decisions can only be taken after a thorough analysis of the objective of the monitoring project at hand, and a specification of the constraints. Also, for some decisions prior information is needed. For instance, to determine the required number of sampling rounds and number of locations per sampling round, prior estimates of components of the variance of the target variable in the universe to be monitored are required. A complete monitoring scheme specifies not only the proposed solution to the monitoring problem at hand, i.e., the sampling plan, but also the objective, the constraints, and the prior information on which the solution is based, together with a description of how the sample data should be analyzed statistically [*de Gruijter et al.*, 2006].

[5] Early in the process important decisions need to be taken which have dominant effects on both costs and quality, and on which most other decisions depend. These major design decisions include the choice between a design-based or a model-based monitoring strategy, the choice of sample support (volume of water samples), whether and how to use composite sampling, and the choice of an observation method. For further details, we refer to [*de Gruijter et al.*, 2006]. Hereafter we will elaborate on the choice between a design-based and a model-based method, which implies choosing between probability sampling and purposive sampling. In a next subsection we will give some more details on the choice of the sampling pattern types in space-time, in space and in time.

[6] *de Gruijter et al.* [2006]
give several design principles, one of which is of special importance for monitoring, namely “anticipate changing conditions during monitoring”. While a survey takes place within a relatively short period of time, the monitoring period can be so long that not only the universe itself changes (being the motivation for repeating the survey), but the objectives of the monitoring project and/or the constraints posed on the monitoring scheme may also change over time. For instance, during the monitoring period interest may shift to other target variables or other subregions or subperiods, and the budget may have changed considerably. One condition that is always changing during the monitoring is the amount of prior data as a result of obtaining more and more information on the variation in space and/or time in a sequential manner. Any of these changes may provide a basis for fine tuning or for a thorough redesign of the monitoring scheme. To accommodate these changes *Overton and Stehman* [1996] recommend a simple design structure.

2.1. Design-Based or Model-Based Method

[7] The choice between a design-based method involving probability sampling and design-based

inference, or a model-based method involving purposive sampling and model-based inference, is one of the major design decisions one must make in designing any monitoring scheme[*Brus and de Gruijter*, 1997]. In a design-based method sampling units (locations and/or instants of time) are selected by probability sampling according to some well-defined sampling design. This sampling design determines the probabilities (for infinite populations probability densities) that a sampling unit is included in the sample, as well as the inclusion probabilities for pairs of sampling units. All sampling units in the universe must have a positive probability of being selected. The inclusion probabilities are used in the statistical inference (e.g., estimating the mean and the sampling variance of the estimated mean), so the inference is based on the sampling design.

[8] In a model-based method there are no requirements on the method used for selecting the sampling units. Typically the sampling units are not selected by probability sampling but purposively, such that the prediction error variance is minimized. Commonly used spatial sampling patterns for model-based methods are regular grids and spatial coverage samples. A model for spatial and or temporal variation, including a random error term, is used for predicting the target quantity (e.g., the values at points in the space-time universe, the spatiotemporal mean or the temporal trend), and for estimating the prediction error variance of this quantity. The quality of these predictions depends on the quality of the model. In design-based methods no model of variation in space and or time is used. This makes that design-based methods have better validity properties, i.e the quality of the result is independent of the quality of model assumptions.

[9] The appropriateness of these two approaches is partly determined by the objective of monitoring, more specific the number of domains. At one extreme there is only one domain being the entire universe, at the other extreme there is an infinite number of domains being all points (s,t) in a space-time universe, all locations (s) in a spatial universe or all instants of time (t) in a temporal universe. For mapping the current values (values at latest sampling time) or the temporal trend at locations, model-based methods are the best option, whereas for estimating the spatiotemporal mean, the current spatial mean, or the spatial mean temporal trend both approaches are appropriate in principle. In the latter case the choice between the two approaches should be guided by the relative importance of validity and efficiency. *Cooper* [2004]
discusses the choice between design-based and model-based inference in the context of estimating water quality in streams.

[10] In compliance or regulatory monitoring, where for example one tests whether concentrations of pollutants exceed regulatory standards or not, the validity of the result (conclusion of hypothesis testing) is of primary interest, it may be argued that design-based methods are more appropriate.

2.2. Types of Sampling Pattern

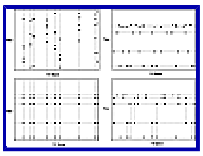


Figure 1. Four basic types of sampling pattern for monitoring: static (top left), synchronous (top right), static-synchronous (bottom left), and rotational (bottom right).

[11] The efficiency (precision and costs) of a sampling pattern for monitoring is partly determined by the distribution of the sampling units in the space-time universe. A relevant concern then is whether the same locations are observed at all sampling times, or whether this restriction is relaxed so that all or part of the sampling locations are replaced by new ones. On the basis of this aspect four basic types of sampling pattern are distinguished: static, synchronous, static-synchronous and rotational patterns (Figure 1). In static sampling all sampling takes place at a fixed set of locations. Sampling at the various locations may or may not follow the same pattern in time. In synchronous sampling, also referred to as repeated or dynamic sampling, a different set of sampling locations is selected for each sampling time; that is, the sampling locations are not revisited. The spatial patterns used at different times may or may

not be the same. If they are the same, then they do not coincide spatially, because otherwise the pattern would be static-synchronous. When static sampling and synchronous sampling are combined with each other, we speak of static-synchronous sampling, also referred to as a pure panel. Rotational sampling is a compromise between static sampling and synchronous sampling, in the sense that the locations of the previous sampling time are partially replaced by new ones. The choice of a pattern type for monitoring should be guided by statistical as well as operational considerations. For instance, if the aim is to estimate the (spatial) mean change of the target variable from the previous round to the next sampling round, in general a static-synchronous pattern is the best choice because it gives most precise estimates. On the other hand, for flexibility reasons, we might prefer a synchronous pattern.

[12] A more complete design of a sampling pattern for monitoring specifies spatial as well as temporal patterns describing the distribution of the sampling units in both space and time, respectively. This leads for instance to the following descriptions: (1) synchronous sampling with random grid sampling in space and stratified simple random sampling in time; and (2) static-synchronous sampling with centered grid sampling in space and systematic sampling in time.

3. Case Study: Compliance Monitoring of Quality of Surface Water

[13] The process of designing efficient sampling patterns will now be illustrated with a real world case study on water resource monitoring. Article 8 of the European Water Framework Directive [*Council of the European Communities*, 2000] states that "Member states shall ensure the establishment of programmes for the monitoring of water status in order to establish a coherent and comprehensive overview of water status within each river basin district". For surface waters such programmes shall cover amongst others the ecological and chemical status. In this study a monitoring scheme is designed for the surface waters (ditches) in a lowland peat area with dairy farms in the Netherlands. The aim of the monitoring scheme is to test whether the quality of the surface waters complies with the standard, so it is an example of compliance monitoring, or in terms of the European WFD: operational monitoring. Annex 5, section 1.3 of the WFD states that "the level of confidence and precision of the results provided by the monitoring programme should be given in the plan" [*Council of the European Communities*, 2000], thus demanding a statistical approach to the problem.

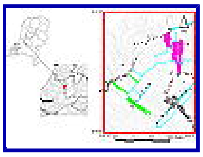


Figure 2. Location of the farms Spruit (bottom left) and Zegveld (top right) for which a monitoring scheme was designed.

[14] In the Netherlands one of the questions is whether measures must be taken so that the quality of the surface water in agricultural areas complies with the WFD standards. The WFD standards are still qualitative, and therefore we used the Dutch Maximum Allowable Risk (MAR) standards. Two dairy farms, Spruit and Zegveld, both situated in the lowland peat area in the west of the Netherlands (Figure 2), were selected, and for both farms a monitoring scheme was designed in accordance with the guidelines of the WFD [*Council of the European Communities*, 2000] and the Guidance on Monitoring for the Water Framework Directive [*Water Framework Directive Common Implementation Strategy Working Group 2.7 Monitoring*, 2003].

3.1. Aim

[15] First the aim of the monitoring was further detailed (see Table 1):

[16] 1. In time the universe consists of a single period from 1 April to 30 September (year unspecified); in geographical space the universe is defined as the surface water in the ditches that are bordered on one or two sides by the fields of Spruit or Zegveld.

[17] 2. Domains: Spruit and Zegveld; that is, a result is required for these two farms separately.

[18] 3. Target variables: concentrations of N-total and P-total in the surface water. For these two variables, maximum allowable risk (MAR) standards are available in the Netherlands.

[19] 4. Target parameter: the spatiotemporal mean

[20] 5. Type of result: the study seeks an answer (yes or no) to the question "does the spatiotemporal mean of the N-total and P-total concentrations in the surface water during the summer half-year in 2006 at Spruit (Zegveld) comply with the MAR standards (2.2 mg/l and 0.15 mg/l respectively)?" The decision is taken by statistical testing of the null hypothesis (H_0): $c \geq c_{\mathrm{MAR}}$ against the alternative hypothesis, (H_1): $c < c_{\mathrm{MAR}}$, where c is the actual spatiotemporal mean N-total (P-total) concentration, and c_{MAR} is the MAR standard for N-total (P-total). The reason for choosing $c \geq c_{\mathrm{MAR}}$ as the "benefit of doubt" hypothesis H_0 is that in large parts of the Netherlands N-total and P-total concentrations above the MAR standard occur. To improve the quality of surface waters, the application of N and P is regulated. Farmers who want to be exempted from these general regulations must show that there is strong evidence that the concentrations are already below the MAR standard.

[21] 6. Quality measure: two types of error can be made in statistical testing. We may incorrectly conclude that the concentration complies with the standard (type I error), or incorrectly conclude that the concentration does not comply with the standard (type II error). The probabilities of these two errors are taken as a quality measure.

3.2. Constraints

[22] The next step is to specify the constraints. First a requirement was specified for the quality of the result. The probability of a type I error (incorrect conclusion $c \leq c_{\mathrm{MAR}}$) is set to a maximum of 5%, and the probability of the type II error (incorrect conclusion $c \geq c_{\mathrm{MAR}}$) is set to a maximum of 20% at a concentration of the standard minus 15% of the standard. The latter implies a power of 80% at an N-total concentration of 2.2 - (0.15 * 2.2) = 1.87 mg/l and a P-total concentration of 0.15 - (0.15 * 0.15) = 0.125 mg/l. The yearly costs of monitoring should not exceed euros 1,000. = per farm.

3.3. Prior Information

[23] In 2004 at Spruit's farm water samples were collected in nine sampling rounds between 23 April and 28 September, at five locations divided over four ditches. The data were treated as a simple random sample. Table 1 shows the estimated spatial variance of observations at locations within a sampling round ($\tilde{V}_S$), and the estimated temporal variance of the spatial means ($\tilde{V}_T$).

3.4. Sampling Plan

[24] We preferred a design-based approach for two reasons: (1) design-based methods are well suited for global quantities such as the spatial-temporal mean; and (2) design-based methods have in general better validity properties, which is of great importance for compliance monitoring.

[25] We chose a synchronous pattern type mainly because of its flexibility; that is, the sampling design (number of locations and even the type of sampling design) can easily be adapted during the monitoring period. Also, estimation of the sampling variance, needed in statistical hypothesis testing, is simple and straightforward compared to a static-synchronous pattern. In space the locations are selected by stratified simple random sampling (STSI), with ditches as strata. In time also STSI is selected as a design type, with periods of 2 months as strata (which implies that we have three temporal strata).

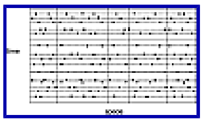


Figure 3. Schematic presentation of the synchronous sampling pattern, with stratified simple random sampling in space and with stratified simple random sampling in time.

[26] Figure 3 is a schematic presentation of the sampling plan.

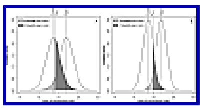


Figure 4. (a) Testing of H_0 hypothesis: spatiotemporal mean N-total concentration $\geq c_{\mathrm{MAR}}$, for number of sampling rounds (n) = 4 and number of sampling locations per sampling round (m) = 10. The maximum probability of type I error (α) was set to 5%. For this sample size the probability of type II error (β) was 42% (power: 58%). By increasing the sample size, the sampling variance of the estimated mean decreases, and consequently, c_{crit} shifts toward c_{MAR} (for given α), and β decreases. (b) For n = 4 and m = 32, the power reaches the required level of 80%.

[27] The prior information was used to determine the required number of sampling locations and sampling times given the quality requirements. We do not have prior information on the spatial variance within spatial strata nor on the temporal variance of spatial means within and between temporal strata, and therefore we used the prior information of Table 1 to obtain prior estimates of the required sample sizes for a synchronous pattern with simple random sampling (SI) in space and SI in time. The temporal sections (horizontal lines in Figure 3) can be considered as primary sampling units in two-stage sampling, and the locations as secondary units. Therefore, the prior estimate of the sampling variance of the spatiotemporal mean $\bar{z}$ equals

$$\tilde{V}(\bar{z}) = \frac{1}{n}\{\tilde{V}_T + \frac{1}{m}\tilde{V}_S\}, \quad (1)$$

where n is the number of sampling times (sampling rounds), m is the number of selected sampling locations per sampling round, $\tilde{V}_S$ is the prior estimate of the spatial variance of the observations at locations within sampling rounds (variance within primary units), and $\tilde{V}_T$ is the prior estimate of the temporal variance of the spatial means (variance between primary units). This variance was calculated for many combinations of n and m. Given the prior estimate of the variance of the spatiotemporal mean, the power of the test can be calculated with

$$1 - \beta = \Phi\left(c_{\mathrm{crit}}; c_{\mathrm{power}}; \tilde{V}(\bar{z})\right), \quad (2)$$

where 1-β is the power (β is the probability of a type II error), c_{crit} is the critical value of the mean concentration beyond which H_0 is rejected, c_{power} is the concentration for which the power is estimated ($0.85c_{\mathrm{MAR}}$), and Φ is the cumulative normal distribution. c_{crit} is given by

$$c_{\mathrm{crit}} = \Phi^{-1}\left(\alpha; c_{\mathrm{MAR}}; \tilde{V}(\bar{z})\right) \quad (3)$$

with α being the maximum tolerable probability of a type I error, here $\alpha = 0.05$. Figure 4 illustrates the calculation of the power at $c = 0.85\ c_{\mathrm{MAR}}$, for $\alpha = 0.05$.

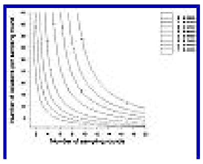
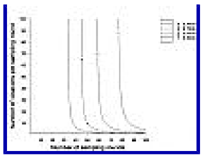


Figure 5. Power based on prior data of 2004 for N-total (top) and P-total (bottom) at a concentration of 1.87 mg/l (N-total) and 0.125 mg/l (P-total). Action level (MAR standard) is 2.2 mg/l for N-total, and 0.15 mg/l for P-total. Confidence level 95%.



[28] The result is shown in Figure 5. Table 2 shows combinations of *n* and *m* leading to the required power. Table 2
shows that for P-total the required number of sampling rounds is much larger than for N-total. For P-total the temporal variance of spatial means is only slightly smaller than the spatial variance of observations at locations, whereas for N-total the temporal variance is approximately 5% of the spatial variance (Table 1).

[29] To reduce laboratory costs all water samples (aliquots) taken during a sampling round are bulked to form composite aliquots (composites). To obtain unbiased estimates of the spatial mean, the volume of water of a given spatial stratum in the composite is proportional to the total volume of water in that spatial stratum.

[30] The selection of the ultimate combination of *n* and *m* from all combinations that comply with the requirement on the power (Table 2) should be based on the costs. As the travel time to the study area is a considerable portion of the total time for fieldwork, we decided to take for *m* the number of locations that could be sampled in 1 day. Given this choice of *m*, the required number of sampling rounds was determined.

[31] The monitoring strategy was tested in the period 1 June to 31 September of 2006. The two farms have been sampled in five sampling rounds, two in the temporal stratum June-July, and three in the temporal stratum August-September. Note that the first temporal stratum April-May was not sampled. At Spruit 50 locations were sampled per sampling round, at Zegveld 75 locations. Locations were selected by STSI, using ditches as strata. All samples of a given sampling round were bulked.

3.5. Statistical Inference

[32] After the data have been obtained the spatiotemporal mean was estimated by

$$\hat{\bar{z}} = \sum_{h=1}^{L} w_h \hat{\bar{z}}_h \qquad (4)$$

where L is the number of temporal strata, w_h is the weight of temporal stratum h being equal to the number of days of stratum h
divided by number of days of whole monitoring period (relative duration), and $z(\hat{\bar{h}})$ is the estimated spatiotemporal mean concentration for temporal stratum h, that can be estimated on his turn by

$$\hat{\bar{z}}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} z_{hi}, \qquad (5)$$

where n_h is the number of sampling rounds in temporal stratum h, and z_{hi} is the concentration of composite i in temporal stratum h used as an estimate of the spatial mean for sampling round i. The sampling variance of the estimator equation (4) was estimated by

$$\hat{V}\left(\hat{\bar{z}}\right) = \sum_{h=1}^{L} w_h^2 \frac{\hat{V}_{T,h}}{n_h}, \qquad (6)$$

where $\hat{V}_{T,h}$
is the estimated temporal variance of the estimated spatial means (composite means) for stratum h,

$$\hat{V}_{T,h} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(z_{hi} - \hat{\bar{z}}_h \right)^2 \qquad (7)$$

where $\hat{\bar{z}}_h$ is the unweighted mean of the composites in temporal stratum h. Note that in equation (6) the spatial variance of observations at locations within sampling rounds (variance within primary units) of equation (1) does not show up. The reason is that $\hat{V}_{T,h}$ is the estimated temporal variance of the estimated spatial means, whereas V_T is the temporal variance of the true spatial means. The error in the estimated spatial means due to the spatial variance within sampling rounds is automatically incorporated in $\hat{V}_{T,h}$.

[33] Because of the bulking of the water aliquots taken during a sampling round, and the small number of sampling rounds, testing with a normal distribution might not be very realistic. We used the Satterthwaite approximation [*Satterthwaite*, 1946; *Cochran*, 1977], saying that the standardized estimated spatiotemporal mean has Student *t* distribution:

$$\frac{\hat{\bar{z}} - c_{\mathrm{MAR}}}{\sqrt{\hat{V}\left(\hat{\bar{z}}\right)}} \sim t_{df} \qquad (8)$$

with the degrees of freedom equal to

$$df \approx \frac{\left(\sum_{h=1}^{L} w_h^2 \hat{V}_{T,h}^2\right)^2}{\sum_{h=1}^{L} w_h^4 \left(\hat{V}_{T,h}^2\right)^2 \frac{1}{n_h - 1}}. \qquad (9)$$

[34] Table 3 shows that for both target variables and at both farms the estimated spatiotemporal means were larger than the MAR standards. This explains the large *p* values, from which we can conclude that we certainly cannot reject the null hypothesis $c \geq c_{\mathrm{MAR}}$. If we would have chosen as a null hypothesis $c \leq c_{\mathrm{MAR}}$, then this hypothesis would have been rejected ($\alpha = 0.10$) for N-total at both farms, and for P-total at Spruit.

3.6. Updating the Required Number of Sampling Rounds

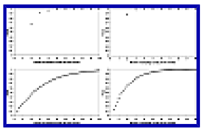


Figure 6. Updated power for N-total (top) and P-total (bottom) for Spruit (left) and Zegveld (right) at a concentration of 1.87 mg/l (N-total) and 0.125 mg/l (P-total). Action level (MAR standard) is 2.2 mg/l for N-total, and 0.15 mg/l for P-total. Confidence level 95%.

[35] We used the monitoring data to update the prior estimate of the temporal variance of spatial means. Additional prior information on the spatial variance within sampling rounds is not available, since all water samples (aliquots) taken during a sampling round were bulked to form composite aliquots in order to reduce laboratory costs. Therefore, we were not able to update both the required number of sampling rounds and the number of locations per sampling round. Table 4 shows the temporal variance of estimated spatial means within strata as estimated from the monitoring data. The estimated variances for the two temporal strata have been pooled, using the numbers of sampling times as weights. For N-total the temporal variance of spatial means at Zegveld is close to the prior estimate of Table 1, but for Spruit it is approximately twice as large. For P-total the temporal variance at Spruit is close to the prior estimate, but for Zegveld it is considerably smaller. We used the unweighted average of the pooled temporal variances as estimated from the monitoring data and the prior estimates of this variance of Table 4 as the updated estimates of the temporal variance within the temporal strata. Given the number of sampling locations per sampling round (50 for Spruit and 75 for Zegveld) and the allocation of these numbers to the spatial strata, we used this updated temporal variance to update the number of sampling

rounds n_h to achieve the required power of the test (equation (6)). To compute the power, we used the Student *t* distribution. Figure 6 shows the results. For N-total the required number of sampling rounds per 2 months is 3 (Spruit) and 2 (Zegveld), leading to a total of 9 and 6 sampling rounds for the summer half-year, which is slightly more than the prior estimate of 4 sampling rounds (Table 2). For P-total the required number of sampling rounds equals 25 per 2 months for Spruit and 15 for Zegveld, leading to a total of 75 and 45 sampling rounds. For Spruit this is of the same order of magnitude as the prior estimate of Table 2 (77), but for Zegveld it is much less. Uncertainty about the variation in time and space of P-total therefore remains, and consequently at this stage we cannot be conclusive about the required number of sampling rounds and number of locations per sampling round for P-total. As monitoring continues, the information on the variation of N-total and P-total increases, so that we gradually become more certain about the required sample size. More than ten to twenty sampling rounds per year are unfeasible. If so many sampling rounds are required indeed to achieve the quality constraints, then we should think of an alternative sampling strategy, for instance sampling with automatic samplers at fixed locations (static or static-synchronous pattern, see Figure 1).

4. Conclusions

[36] A sampling plan for monitoring comprises much more than the number of sampling locations, the sampling frequency, and where and when to take samples. Many other decisions must be taken, for instance on the statistical approach (design-based or model-based), sample support, the observation method, whether or not to form composite samples et cetera. All these decisions can only be taken after a thorough analysis of the objective of monitoring and the constraints such as budgetary constraints and quality requirements. For regulatory or compliance monitoring in which a global quantity is to be tested against a (legalized) standard, we generally prefer a design-based method because in estimating the quantity from the monitoring data no assumptions on the variation in the space-time universe need to be made, leading to a valid result.

[37] For compliance monitoring (or operational monitoring in the context of the WFD) of the quality of surface waters by means of statistical testing of the spatiotemporal mean concentration of N-total and P-total during the summer half-year against WFD standards, synchronous sampling with stratified simple random sampling in space and stratified simple random sampling in time, works adequately. To reduce laboratory costs we propose to bulk the aliquots taken at the locations of a given sampling round, i.e., composite sampling. To achieve a power of 80% at a concentration 15% below the WFD standard and a confidence of 95%, six to nine sampling rounds are needed with 50 to 75 locations per sampling round. For P-total the required number of sampling rounds differs strongly between the two farms, but is for both farms much larger than for N-total.

[38] This paper illustrates the usefulness of design-based methods for compliance monitoring of surface water quality in stagnant waters. We think that a design-based approach may also be worthwhile for compliance monitoring of streams. However, strong dynamics of the target variable to be monitored, may have consequences for the sampling design to be applied.

Acknowledgments

[39] We are grateful to Bram de Vos and Jan van Kleef for providing us with prior information on spatial and temporal variances in the case study and to Igor Staritsky for the GIS work.

Citation: Brus, D. J., and M. Knotters (2008),
Sampling design for compliance monitoring of surface water quality: A case study in a Polder area, *Water Resour. Res.*, *44*, W11410, doi:10.1029/2007WR006123.