Estimating space-time mean concentrations of nutrients in surface waters of variable depth

M. Knotters[1] and D. J. Brus[1]

Received 2 July 2009; revised 25 February 2010; accepted 5 March 2010; published 3 August 2010.

[1] A monitoring scheme has been designed to test whether the space-time mean concentration total Nitrogen (N-total) in the surface water in the Northern Frisian Woodlands (NFW, The Netherlands) complies with standards of the European Water Framework directive. Since in statistical testing for compliance monitoring valid estimators for the mean and its variance are important, a design-based method is preferred above a model-based method. In the NFW-area the surface water depth varies in both space and time and can periodically equal zero, due to variation in precipitation and evapotranspiration. To account for this, space-time mean concentrations are estimated by the ratio of the estimated total mass of nutrient and the estimated total volume of water. The method is applied in the period from 1 April to 30 September 2008 to four hydrologically different subareas. Besides, the aim was to use the information on the spatial and temporal variance of N-total concentrations to optimize the numbers of sampling rounds and sampling locations per sampling round in future monitoring campaigns, given budgetary constraints. A bootstrap procedure was applied to account for uncertainty about the temporal and spatial variances in estimating the optimal number of sampling rounds and sampling locations. For two subareas the accuracy of the estimated space-time means can be improved by sampling more frequently at less locations (compared to the design applied in 2008), whereas for one subarea sampling less frequently at more locations increases the precision. For one subarea the sample data were rather inconclusive about the optimal sample sizes.

Citation: Knotters, M., and D. J. Brus (2010), Estimating space-time mean concentrations of nutrients in surface waters of variable depth, *Water Resour. Res.*, *46*, W08502, doi:10.1029/2009WR008350.

1. Introduction

[2] Although Dutch farmers generally endorse the goals of manure policy, they increasingly question the measures dictated by law to achieve these goals. As an alternative to obeying rules on the way of manure application, a farmers cooperative in the Northern Frisian Woodlands (NFW) in the Netherlands was given the responsibility for reaching the environmental goals in their region, by applying innovative management systems for sustainable forms of agriculture [*Bouma et al.*, 2008, pp. 216–217]. A monitoring scheme had to be designed to test whether the environmental goals are reached or not. One of these goals is that the surface water quality complies with the standards of the European Water Framework directive [*Council of the European Communities*, 2000]. To this end the space-time mean concentrations of nutrients in surface water were tested against standards. In statistical testing it is important to obtain valid estimates of the mean (total) and its error variance. Validity means that the outcome of the test does not depend on the quality of model assumptions. Therefore, a design-based method is preferred above a model-based method, because in a design-based method no assumptions on the variation of the concentrations in space and in time are made.

[3] *Dobbie et al.* [2008] reviewed model-based and design-based strategies for monitoring stream networks and discussed the differences between both approaches. In a *design-based approach* sampling units (locations and/or instants of time) are selected by probability sampling. The 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 inference (e.g. estimating the mean and the sampling variance of the estimated mean) is based on the inclusion probabilities and thus on the sampling design. In a *model-based approach* there are no requirements on the method used for selecting the sampling units. The sampling units can be selected purposively, such that the prediction error variance is minimized. Regular grids and spatial coverage samples are commonly used spatial sampling patterns for model-based methods. A model for spatial and or temporal variation, including a random error term, is used for predicting the target quantity (e.g. point values, a spatio-temporal mean or a temporal trend), and for estimating the

[1]Soil Science Centre, Wageningen University and Research Centre, Wageningen, Netherlands.

0043-1397/10/2009WR008350

prediction error variance of this quantity. The quality of these predictions depends on the quality of this model. *Brus and de Gruijter* [1997] discussed the advantages and disadvantages of the design-based and the model-based approach, and how to choose between them. In contrast to model-based methods, in design-based methods no model of variation in space and or time is used. Therefore a design-based approach has better validity properties, i.e the quality of the result is independent of the quality of model assumptions. *Dobbie et al.* [2008] concluded that most of the model-based approaches in the literature they considered were applied in two-dimensional domains, and only a few in linear systems such as streams. Besides, they observed that some recent developments in probability-based designs were motivated by applications in monitoring large-scale natural resource systems such as stream networks. Furthermore they concluded that "probability-based design approaches appear to be better suited both theoretically and practically for the sparse sampling of stream networks", and mention the generalized random-tessellation stratified (GRTS) sampling design [*Stevens and Olsen*, 2003, 2004] as a promising approach for sampling large-scale stream networks. *Southerland et al.* [2009] preferred probability-based sampling above "ad hoc" (targeted) sampling of stream segments in assessment of biological water quality standards in Maryland, because conditions at a watershed scale cannot be reliably inferred from "ad hoc" sampling. *Peterson et al.* [1998] described the advantages of probability sampling over non-random sampling in assessment of lake trophic states in the Northeastern United States. An advantage is in the fact that inference is objectively based on the probability sampling design. The probability-based design permits sample findings to be inferred to the entire lake population. Assessment programs based on probability sampling in space as described by (amongst others) *Southerland et al.* [2009] and *Peterson et al.* [1998], followed from the US Clean Water Act [*U.S. Environmental Protection Agency*, 2002], which has similar aims as the European Water Framework Directive.

[4] *Brus and Knotters* [2008] presented a design-based method to test space-time mean concentrations of nutrients in the surface water of a polder area. In this fully design-based method both sampling rounds and sampling locations are selected by probability sampling. Random selection of sampling rounds is relatively new. Random selection of sampling times for monitoring the quality of surface water has been applied before by *Thomas* [1985] and *Thomas and Lewis* [1995].

[5] For the polder area in the study by *Brus and Knotters* [2008] the water depth was assumed to be constant in space and in time, which was realistic because a constant water level was maintained during the monitoring period. In the NFW-area the surface water depth varies in both space and time and can periodically equal zero, due to variation in precipitation and evapotranspiration. Therefore, the estimation method presented by *Brus and Knotters* [2008] cannot be applied straightforwardly in this case and must be adapted to account for this variable depth. We will show that in this case space-time mean concentrations can be estimated by the estimator for the ratio of the total mass of nutrient and the total volume of water [*Cochran*, 1977, p. 30–34].

[6] The aim of this study is to estimate space-time mean concentrations of total Nitrogen (N-total) in surface water with a depth varying in space and time. The method is applied to four hydrologically different subareas of the NFW-area. Besides, the aim was to use the information on the spatial and temporal variance of N-total concentrations to optimize the numbers of sampling rounds and sampling locations per sampling round in future monitoring campaigns, with the costs of the executed campaign as budgetary constraint.

2. Sampling Procedure

[7] Figure 1 shows the study area. Four subareas of the NFW-area have been selected to develop a monitoring procedure which can be applied to larger areas in a later stage. Subarea 1 (approximately 22.4 hectares of surface water) is situated in a polder with peat soils and relatively constant water levels. Subarea 2 (1.5 hectares) is located in a slightly undulating area with cover sands overlaying boulder clay, and gravitational discharge. Subarea 3 (4.2 hectares) covers a gradient from sandy to peaty soils, with gravitational discharge. Subarea 4 (9.8 hectares) is in a polder with clay soils. The aim was to estimate the space-time mean concentrations for all four subareas, and not merely the space-time mean concentration for the area as a whole.

[8] The Dutch Maximum Allowable Risk (MAR) standards are defined for the space-time mean concentration of N-total in surface water up to 50 cm depth, during the summer period from April 1 to September 30. For this period in 2008 we applied a synchronous sampling pattern [*de Gruijter et al.*, 2006; *Brus and Knotters*, 2008], with stratified simple random sampling (STSI) in time and simple random sampling (SI) in space. The spatial SI sample of a given sampling round was selected *independently* from the SI samples of the other rounds. This is referred to as independent synchronous sampling. As shown by *Brus and Knotters* [2008] this space-time sample can be analyzed as a two-stage sample, with sampling rounds as primary sampling units (psu's), and sampling locations as secondary sampling units (ssu's).

[9] Six sampling rounds were selected by STSI, with temporal strata of two months length and two sampling rounds per stratum, so that temporal coverage is improved as compared to simple random sampling. The selected datums are April 2, April 23, June 6, July 24, September 25 and September 26. Next, for each sampling round sampling points were selected by SI. The numbers of sampling locations per sampling round were 18 points in subarea 1, 6 points in subarea 2, 6 points in subarea 3 and 10 points in subarea 4. These numbers are approximately proportional to the area of surface water in the subareas. Figure 1 shows the positions of the sampling points for the first sampling round at April 2. At each sampling event (combination of sampling location and sampling time) an aliquot of the surface water was taken over the full depth, up to a maximum of 50 cm. If a randomly selected location happened to fall in a dry ditch, then an extra location was randomly selected, until the number of aliquots of surface water per sampling round was 18, 6, 6, and 10 for subareas 1, 2, 3 and 4, respectively. Water depths at the sampling

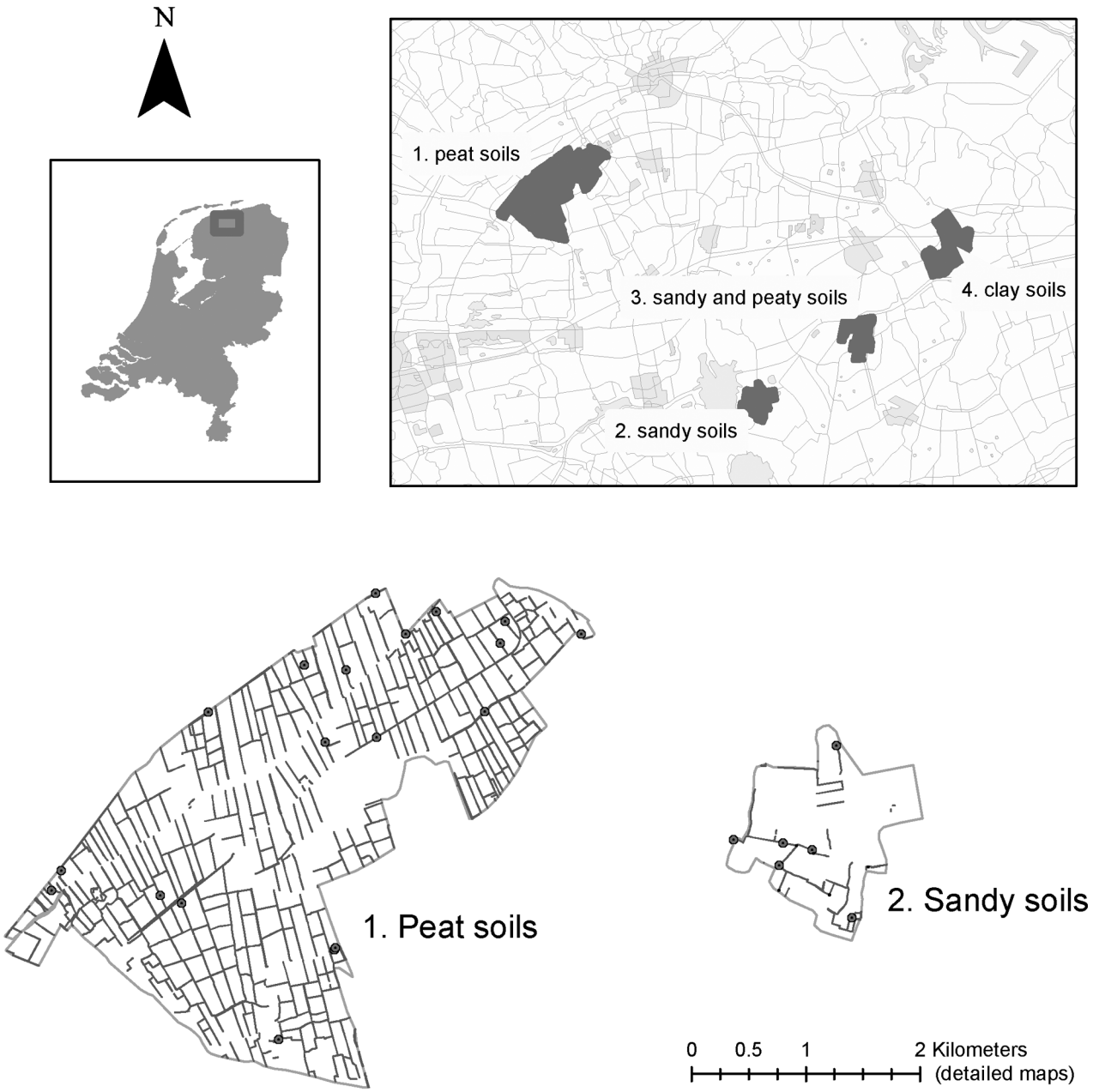

Figure 1. Study area. The dots indicate the sampling locations during the first sampling round, 2 April 2008.

locations were measured when the aliquots were taken. These water depths are used in the estimation procedure for volume weighting.

3. Estimation Procedure

3.1. Target Parameter

[10] The space-time mean concentration of N-total is defined as:

$$\bar{\bar{y}} = \frac{\int_{t\in\mathcal{T}} \int_{\mathbf{s}\in\mathcal{S}} y(\mathbf{s},t)\cdot l(\mathbf{s},t)\mathrm{d}\mathbf{s}\mathrm{d}t}{\int_{t\in\mathcal{T}} \int_{\mathbf{s}\in\mathcal{S}} l(\mathbf{s},t)\mathrm{d}\mathbf{s}\mathrm{d}t}, \quad (1)$$

with $y(\mathbf{s},t)$ the concentration of N-total at location $\mathbf{s}$ and time t (averaged over the water depth), $l(\mathbf{s},t)$ the water depth at location $\mathbf{s}$ and time t, $\mathcal{S}$ the (two-dimensional) area, and $\mathcal{T}$ the monitoring period. In this definition it is assumed that the area $\mathcal{S}$ is constant over time. This implies that the water volume only varies with water depth, which is a reasonable assumption given the dimensions of the surface waters (mainly ditches) in the study area. We use two bars ($\bar{\bar{y}}$) to indicate that the target quantity is an average over both space *and* time. Note that $\int_{\mathbf{s}\in\mathcal{S}} l(\mathbf{s},t)$ equals the volume of water at time t. This target parameter is a volume-weighted average concentration.

3.2. Estimation of a Ratio

3.2.1. SI in Time and SI in Space

[11] First, we describe the estimation procedure for SI in time and SI in space. In section 3.2.2 we extend the estimation procedure to STSI in time and SI in space, which was applied in this study. Note that the denominator in equation (1), the volume of water at time t integrated over the monitoring period, is unknown and must be estimated from the sample. This implies that the space-time mean must be estimated by estimating the numerator and the denominator of the ratio of equation (1). For SI sampling in time and SI sampling in space this estimator equals:

$$\hat{\bar{\bar{y}}}_{\mathrm{SI,SI}} = \frac{\frac{T}{n}\sum_{i=1}^{n}\frac{A}{m_i}\sum_{j=1}^{m_i} y_{ij}\cdot l_{ij}}{\frac{T}{n}\sum_{i=1}^{n}\frac{A}{m_i}\sum_{j=1}^{m_i} l_{ij}} \tag{2}$$

with T the length of the monitoring period, n the number of selected sampling times (psu's), A the (surface) area of the study area, m_i the number of selected sampling locations (ssu's) in the ith psu, y_{ij} the concentration at the ith selected psu and the jth selected ssu, and l_{ij} the water depth at that time and location. With $\hat{t}_i(y) = \frac{A}{m_i}\sum_{j=1}^{m_i} y_{ij}\cdot l_{ij}$ the estimated total mass of N-total in the ith psu and $\hat{V}_i = \frac{A}{m_i}\sum_{j=1}^{m_i} l_{ij}$ the estimated water volume in that psu, equation (2) reduces to

$$\hat{\bar{\bar{y}}}_{\mathrm{SI,SI}} = \frac{\sum_{i=1}^{n}\hat{t}_i(y)}{\sum_{i=1}^{n}\hat{V}_i} \tag{3}$$

[12] The variance of $\hat{\bar{\bar{y}}}_{\mathrm{SI,SI}}$ can be estimated by [*Cochran*, 1977, p. 32, equation (2.45)]

$$\begin{aligned}\mathrm{var}\left(\hat{\bar{\bar{y}}}_{\mathrm{SI,SI}}\right) &= \frac{1}{n\left(\hat{\bar{V}}\right)^2}\frac{\sum_{i=1}^{n}\left[\hat{t}_i(y)-\hat{\bar{\bar{y}}}_{\mathrm{SI,SI}}\hat{V}_i\right]^2}{n-1}\\ &= \frac{1}{n\left(\hat{\bar{V}}\right)^2}\frac{\sum_{i=1}^{n}[\hat{t}_i(e)]^2}{n-1}\\ &= \frac{s^2[\hat{t}(e)]}{n\left(\hat{\bar{V}}\right)^2},\end{aligned} \tag{4}$$

with $\hat{t}_i(e)$ the estimated residual in total mass of N-total (estimated residual mass) for the ith psu, $s^2[\hat{t}(e)]$ the estimated temporal variance of these estimated total residuals per psu, and $\hat{\bar{V}} = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i$ the estimated water volume averaged over the psu's.

3.2.2. STSI in Time and SI in Space

[13] The space-time universe was stratified along the time-axis, to improve temporal coverage. For STSI sampling of times (psu's) and SI sampling of locations (ssu's), the space-time mean can be estimated by

$$\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}} = \frac{\sum_{h=1}^{\ell}\frac{T_h}{n_h}\sum_{i=1}^{n_h}\frac{A}{m_{hi}}\sum_{j=1}^{m_{hi}} y_{hij}\cdot l_{hij}}{\sum_{h=1}^{\ell}\frac{T_h}{n_h}\sum_{i=1}^{n_h}\frac{A}{m_{hi}}\sum_{j=1}^{m_{hi}} l_{hij}}. \tag{5}$$

In this combined ratio estimator the numerator and denominator are estimates of totals for the entire universe.

[14] The three time-strata are equally sized (61 days each). In each stratum two psu's were selected. The resulting sampling design is self-weighting, which implies that equation (2) can be used as an estimator of the space-time mean concentration, i.e., in this particular case $\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}} = \hat{\bar{\bar{y}}}_{\mathrm{SI,SI}}$.

[15] To estimate the variance of $\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}}$ we first calculate the residuals in mass of N-total per psu, analogous to [*Cochran*, 1977, p. 32]:

$$\hat{t}_{hi}(e) = \hat{t}_{hi}(y) - \hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}}\cdot\hat{V}_{hi}, \tag{6}$$

with $\hat{t}_{hi}(y) = \frac{A}{m_h}\sum_{j=1}^{m_h}(y_{hij}\cdot l_{hij})$ the estimated mass of N-total for the ith psu in stratum h, and $\hat{V}_{hi} = \frac{A}{m_h}\sum_{j=1}^{m_h} l_{hij}$ the estimated water volume of that psu. The variance of $\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}}$ can then be estimated by

$$\mathrm{var}\left(\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}}\right) = \frac{\sum_{h=1}^{\ell} w_h^2\cdot\mathrm{var}_h\left[\hat{\bar{t}}(e)\right]}{\left(\hat{\bar{V}}\right)^2}, \tag{7}$$

with

$$\mathrm{var}_h\left[\hat{\bar{t}}(e)\right] = \frac{1}{n_h(n_h-1)}\sum_{i=1}^{n_h}[\hat{t}_{hi}(e)]^2 = \frac{s_h^2[\hat{t}(e)]}{n_h}, \tag{8}$$

and

$$\hat{\bar{V}} = \sum_{h=1}^{\ell} w_h\cdot\hat{\bar{V}}_h. \tag{9}$$

3.3. Optimization of Numbers of Sampling Rounds and of Sampling Locations per Sampling Round

[16] In equation (8) $s_h^2[\hat{t}(e)]$ is the estimated temporal variance of the estimated residual mass of N-total per primary unit. This temporal variance automatically includes the uncertainty in the estimated residual amount due to spatial variation of the residuals per sampling location. This explains that the number of sampling locations per sampling round m_h does not show up in equation (8). For optimization of the number of sampling rounds and the number of sampling locations per sampling round, we decompose $s_h^2[\hat{t}(e)]$ in equation (8) in a temporal variance and a spatial variance component as follows:

$$s_h^2[\hat{t}(e)] = s_{Th}^2[t(e)] + A^2\frac{s_{Sh}^2(e)}{m_h} \tag{10}$$

with $s_{Th}^2[t(e)]$ the temporal variance within stratum h of the (errorless) total residuals per primary unit, and $s_{Sh}^2(e)$ the spatial variance within stratum h of the residuals at point locations. Inserting equation (10), and noting that for proportional allocation to temporal strata of equal length $n_h = \frac{n}{\ell}$ and $w_h = \frac{1}{\ell}$, equation (7) can be rewritten as:

$$\mathrm{var}\left(\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}}\right) = \frac{\sum_{h=1}^{\ell} s_{Th}^2[t(e)]}{\ell\left(\hat{\bar{V}}\right)^2}\frac{1}{n} + \frac{A^2\sum_{h=1}^{\ell} s_{Sh}^2(e)}{\ell\left(\hat{\bar{V}}\right)^2}\frac{1}{nm}. \tag{11}$$

As we have rather sparse information on the temporal variance and spatial variance, we assumed that the temporal and spatial variances within strata are equal for the three temporal strata, so that the observations of all three strata can be

Table 1. Summary of the Collected Data in Four Subareas of the Northern Frisian Woodlands in the Period April 1, 2008 to September 30, 2008[a]

		Subarea									
		1					2				
			l		*y*			*l*		*y*	
Temporal Stratum	Datum (Day-Month)	m_{hi}	Min.	Max.	Min.	Max.	m_{hi}	Min.	Max.	Min.	Max.
1	2-4	18	20	50	1.1	4.3	6	10	40	1.5	7.8
1	23-4	18	5	50	0.73	4.5	6	5	20	1.8	5.0
2	11-6	18	5	50	1.3	5.7	6	5	20	2.2	5.3
2	24-7	18	10	50	0.63	2.7	10	0	50	0.	3.6
3	25-9	19	0	50	0.	9.7	12	0	50	0.	23.9
3	26-9	19	0	50	0.	6.9	7	0	20	0.	3.3
		Subarea									
		3					4				
			l		*y*			*l*		*y*	
Temporal Stratum	Datum (Day-Month)	m_{hi}	Min.	Max.	Min.	Max.	m_{hi}	Min.	Max.	Min.	Max.
1	2-4	6	10	35	1.2	4.0	10	10	25	1.0	2.9
1	23-4	8	0	20	0.	7.4	10	5	50	1.3	4.5
2	11-6	6	5	35	0.69	20.0	10	20	35	1.6	5.8
2	24-7	9	0	50	0.	3.0	10	15	30	1.0	7.0
3	25-9	8	0	50	0.	51.1	10	5	20	1.1	5.5
3	26-9	7	0	50	0.	4.2	10	5	30	1.3	5.0

[a]See Figure 1. m_{hi} is the number of sampling locations per observation round. *l* is surface water depth in cm, up to 50 cm. *y* is concentration N-total in $mg \cdot l^{-1}$.

used in estimating $s^2_{T_h}(\bar{e})$ and $s^2_{S_h}(e)$. In this case equation (11) reduces to

$$\mathrm{var}\left(\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}}\right) = \frac{s^2_{T_h}[t(e)]}{\hat{\bar{V}}^2}\frac{1}{n} + \frac{A^2 \cdot s^2_{S_h}(e)}{\hat{\bar{V}}^2}\frac{1}{nm}. \quad (12)$$

[17] We estimated the multivariate sampling distribution of $s^2_{T_h}[t(e)]$, $s^2_{S_h}(e)$ and $\hat{\bar{V}}^2$ by a bootstrap procedure [*Efron and Tibshirani*, 2006], see Appendix A. The estimated multivariate sampling distribution is used to account for the uncertainty about the temporal and variances $s^2_{T_h}[t(e)]$ and $s^2_{S_h}(e)$ and the volume $\bar{V}$ in optimizing the number of sampling rounds and sampling locations.

[18] We now postulate a simple linear cost model:

$$C = c_1 n + c_2 nm, \quad (13)$$

with c_1 the costs per sampling round and c_2 the costs per sampling location. The sample sizes minimizing the variance under the constraint that the total costs does not exceed a budget $C_{\max}$ can be found by the Lagrange multiplier method:

$$n = \frac{C_{\max} S_1}{S_2\sqrt{c_1 c_2} + S_1 c_1} \quad (14)$$

and

$$m = \frac{S_2}{S_1}\sqrt{\frac{c_1}{c_2}} \quad (15)$$

with

$$S_1 = \sqrt{\frac{s^2_{T_h}[t(e)]}{\hat{\bar{V}}^2}} \quad (16)$$

and

$$S_2 = \sqrt{\frac{A^2 s^2_{S_h}(e)}{\hat{\bar{V}}^2}}, \quad (17)$$

The numbers obtained with equations (14) and (15) are reals. These reals were rounded to the nearest integers, and we searched in the neighborhood of these integers for the optimal sample sizes, i.e. the number of sampling rounds n and the number of sampling locations m for which the variance $\mathrm{var}(\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}})$ is minimal and the costs C are smaller than $C_{\max}$. The bootstrap procedure thus results into 10,000 estimates of the optimal values for n and m.

4. Results

4.1. Descriptive Statistics

[19] Table 1 summarizes the collected sample data. In the subareas 1, 2 and 3 ditches frequently dried up. In subareas 2 and 3 the concentrations N-total showed relative large variations. After removing the outlying value of 51.1 $mg \cdot l^{-1}$ in subarea 3 from the data set, the maximum concentration for September 25 was 8.4 $mg \cdot l^{-1}$. Note that the number of sampling locations per sampling round, m_{hi}, varied,

Table 2. Estimated Space-Time Mean Concentrations of Total Nitrate in Surface Water in Four Subareas of the Northern Frisian Woodlands in the Period April 1, 2008 to September 30, 2008[a]

Subarea	$\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}}$ ($mg \cdot l^{-1}$)	$\mathrm{var}(\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}})$ $mg^2 \cdot l^{-2}$
1 (peat soils)	2.45	0.017
2 (sandy soils)	3.44	0.493
3 (sandy and peaty soils)	2.23	0.066
4 (clay soils)	2.31	0.030

[a]See Figure 1.

Table 3. Optimized Numbers of Sampling Rounds (n) and Sampling Locations per Sampling Round (m) in the Period April 1 to September 30[a]

	Based on Sample		Based on Bootstrap Procedure		
Subarea	n	m	n	m	Relative Frequency
1 (peat soils)	16	5	3	38	0.14
			4	28	0.01
			5	22	0.01
			6	18	0.01
			7	15	0.02
			8	12	0.01
			9	11	0.03
			10	9	0.02
			11	8	0.03
			12	7	0.03
			14	6	0.07
			16	5	0.11
			18	4	0.15
			21	3	0.24
			26	2	0.13
			29	1	0.00
2 (sandy soils)	5	7	3	14	0.18
			4	10	0.06
			5	7	0.04
			6	6	0.15
			7	4	0.18
			9	3	0.27
			11	2	0.11
			14	1	0.02
3 (sandy and peaty soils)	3	14	3	14	0.31
			4	10	0.05
			5	7	0.03
			6	6	0.15
			7	4	0.16
			9	3	0.22
			11	2	0.06
			14	1	0.00
4 (clay soils)	7	8	3	22	0.23
			4	14	0.02
			5	12	0.02
			6	10	0.04
			7	8	0.08
			8	6	0.09
			9	5	0.09
			11	4	0.22
			13	3	0.15
			16	2	0.05
			20	1	0.00

[a]The optimized sample sizes are estimated from the results obtained in 2008. The uncertainty about the input parameters, such as the spatial and temporal variance, see equation (12), is quantified by the bootstrap procedure. The last column shows the relative frequency for which the sample sizes n and m were optimal, as computed from 10,000 bootstrap samples.

because extra locations were selected if locations were situated in dry ditches (see section 2).

4.2. Space-Time Mean Concentrations of N-Total

[20] Table 2 presents the estimated space-time mean concentrations of N-total and their standard errors. In subarea 3 an outlying value of 51.1 $mg \cdot l^{-1}$ was observed at September 25, 2008. This outlier was omitted in the calculations. If this outlying observation would be included, the space-time mean concentration for subarea 3 is estimated at 3.77 $mg \cdot l^{-1}$, with a variance of 3.416 $mg^2 \cdot l^{-2}$.

[21] The estimated space-time mean concentrations exceed the legal MAR standard for N-total of 2.2 $mg \cdot l^{-1}$ in all four subareas. It should be noted, however, that the MAR standard is defined for stagnant open waters, like lakes, whereas the surface water in the study area consists of a system of small ditches draining land that is mainly used for dairy farming.

4.3. Optimal Number of Sampling Rounds and Sampling Locations

[22] We optimized the total number of sampling rounds $n = \sum_{h=1}^{\ell} n_h$ and the number of sampling locations per sampling round m for the STSI-SI design, for a budget that equals the costs of the monitoring project executed in 2008. These costs were computed with equation (13). For c_2 we took the costs of one laboratory analysis of N-total concentration, being € 158.50 per aliquot. The costs of a sampling round (costs of analyses not included), c_1 in equation (13) equaled € 430. This resulted into a costs of € 19,698 for subarea 1, € 8,286 for subareas 2 and 3, and € 12,090 for subarea 4. Table 3 gives the optimized numbers of sampling rounds and sampling locations per sampling round for the four subareas. The optimized numbers are calculated on the basis of the monitoring data collected in 2008 directly, as well as on the basis of the multivariate sampling distribution of $s^2_{T_h}[t(e)]$, $s^2_{S_h}(e)$ and $\hat{\overline{V}}^2$ as estimated from the monitoring data collected in 2008, see section 3.3.

[23] For subarea 1 (peat soils) the optimal total number of sampling rounds, n_{opt}, and optimal number of sampling locations per sampling round, m_{opt}, equaled 16 and 5, respectively, following from the monitoring data collected in 2008. The bootstrap shows that we are rather uncertain about the optimal numbers of sampling rounds and locations per round. Nevertheless, the bootstrap results seem to confirm that the precision of the estimated space-time mean can be potentially increased by sampling more frequently and less densely: the solutions with 16 sampling rounds or more have a cumulative probability of 63%. Note that for 24% of the bootstrap samples the optimal sample size was 21 rounds of 3 locations per round, which is somewhat different from the optimal solution based on the estimated spatial and temporal variance. This difference can be explained by the nonlinear relation between the sample sizes n and m and the spatial and temporal variance. Finally, note that for 14% of the bootstrap sample the optimal sample size was three rounds of 38 locations per round, which is pretty large. This underpins the need for more precise information on the spatial and temporal variance in order to determine the optimal sample size.

[24] Note that for proportional allocation of the sampling rounds to the three temporal strata of equal length the total number of sampling rounds must be divisible by 3. So, for subarea 1 we conclude that the accuracy can be increased by sampling more often at less locations: 18 or 21 sampling rounds with respectively 4 or 3 locations per round is close to optimal.

[25] For subarea 2 (sandy soils) the bootstrap results indicate that given the maximum budget the precision can be improved by increasing the number of sampling rounds from 6 to 9, and reducing the number of sampling locations per round from 6 to 3. For subarea 3 both the results based on the estimated spatial and the temporal variance, and based on the bootstrap procedure indicate that given the maximum budget the accuracy can be improved by reducing the number of sampling rounds to 3, with 14 locations per

sampling round. For subarea 4 the optimum number of sampling rounds was 7 with 8 locations per round. However, the bootstrap indicated two optima: 3 sampling rounds with 22 locations per round and 11 sampling rounds with 4 locations per round, showing that we cannot draw definite conclusions about the optimal sample size.

5. Discussion and Conclusions

[26] This paper presents a fully design-based method to estimate space-time means, in which both sampling rounds and sampling locations are selected by probability sampling. A design-based approach is preferable above a model-based approach if valid estimates of the mean (total) and its accuracy are required, as in testing against standards in compliance monitoring. In that case it is important that the quality of the test results does not depend on the quality of model assumptions [*de Gruijter et al.*, 2006].

[27] We estimated the space-time mean concentration of N-total in surface waters with variable depth by estimating the ratio of the estimated total mass of nutrient and the estimated total volume of water. It is important to note that in this procedure the sampling units selected in dry ditches are not removed from the data set. The water depth l_{hij} at these locations equals 0, and so is the amount of N-total at that location, $y_{hij} \cdot l_{hij}$.

[28] In this way an unbiased estimate of the space-time mean concentration in surface waters with varying depths is obtained. The method applied in this study is a valuable extension to the method described by *Brus and Knotters* [2008], because the surface water depth may vary in many areas.

[29] The evaluation of the number of sampling rounds and sampling locations indicated that for subareas 1 and 2 the precision of the estimated space-time means can possibly be increased by sampling more frequently at less locations. This is particularly true for subarea 1, the peat area. For subarea 3 the accuracy can be improved by sampling less frequently at more locations. The bootstrap shows that we are rather uncertain about the optimal numbers of sampling rounds and locations per round. This is especially true for subarea 4.

[30] We emphasize that the number of sampling rounds and sampling locations is evaluated with respect to accurate design-based estimation of space-time mean concentrations, since testing these space-time means against legal standards is the objective of the compliance monitoring described in this study. If, for example, tracing locations and periods at which critical conditions occur were the objective, a model-based approach with purposively selected samples would have been recommendable. In that case the number of sampling rounds and sampling locations, as well as the temporal and spatial co-ordinates, are optimized aiming for maximizing the probability of tracing critical conditions.

[31] As an alternative to stratified simple random sampling of sampling rounds with temporal strata of equal length, sampling rounds can be selected by systematic random sampling (SY), i.e., at constant interval. For SY temporal coverage is optimal, and it can be expected that this improved temporal coverage compared to STSI leads to a gain in precision of the estimated space-time mean. A drawback of SY is that no unbiased estimator of the sampling variance exists. Estimating the sampling variance as if the sample was selected by STSI will in general (slightly) overestimate the sampling variance.

[32] We focused on four relatively small areas. If the aim is statistical inference on the water quality in the streams of a large area, random selection of stream segments can be done for instance by Generalized Random Tessellation Sampling [*Stevens and Olsen*, 2003, 2004]. *Brus et al.* [2002] proposed to select sampling locations in a stream network by stratified three-stage sampling, using stream segments as primary sampling units, transects perpendicular to the axis of the watercourses as secondary units, and point-locations as tertiary sampling units. They illustrated how the sample sizes can be optimized given a cost model and prior knowledge on variance components.

Appendix A: Bootstrap Procedure

[33] The bootstrap procedure to estimate the multivariate sampling distribution of $s^2_{\mathcal{T}h}[t(e)]$, $s^2_{\mathcal{S}h}(e)$ and $\hat{\overline{V}}^2$ is as follows:

[34] 1. Select by simple random sampling with replacement (SIR) 6 psu's out of the sample of 6 psu's. Temporal stratification has not been taken into account, because of the small number of primary units within the strata ($n_h = 2$).

[35] 2. Select from each selected psu of the previous step by SIR m_{hi} ssu's (sampling locations), with m_{hi} the number of ssu's that have been selected from the ith psu in stratum h.

[36] 3. From the sample resulting from steps 1 and 2 compute $\hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}}$ (equation (3)) and $\hat{\overline{V}}^2$ (equation (9)).

[37] 4. Compute residuals per ssu (sampling location) by

$$e_{hij} = y_{hij} \cdot l_{hij} - \hat{\bar{\bar{y}}}_{\mathrm{STSI,SI}} \cdot l_{hij}. \tag{A1}$$

[38] 5. Estimate the spatial variances of residuals per ssu within the psu's by

$$s^2_{hi}(e) = \frac{1}{m_{hi}-1}\sum_{j=1}^{m_{hi}}\left(e_{hij} - \hat{\bar{e}}_{hi}\right)^2, \tag{A2}$$

and further estimate the sampling variance of the estimated total residual per psu by

$$\mathrm{var}[\hat{t}_{hi}(e)] = A^2\frac{s^2_{hi}(e)}{m_{hi}}. \tag{A3}$$

[39] 6. Estimate the temporal variance within strata of estimated totals per psu by

$$s^2_h[\hat{t}(e)] = \frac{1}{n_h-1}\sum_{i=1}^{n_h}\left[\hat{t}_{hi}(e) - \hat{\bar{t}}_h(e)\right]^2, \tag{A4}$$

and further pool the estimated sampling variances of estimated totals per psu:

$$\mathrm{var}[\hat{t}_h(e)] = \frac{\sum_{i=1}^{n_h}(m_{hi}-1)\mathrm{var}[\hat{t}_{hi}(e)]}{\sum_{i=1}^{n_h}(m_{hi}-1)}, \tag{A5}$$

and finally, estimate the temporal variance within strata of the *errorless* totals per psu by

$$s_h^2[t(e)] = s_h^2[\hat{t}(e)] - \mathrm{var}[\hat{t}_h(e)]. \quad \text{(A6)}$$

[40] 7. Finally, compute the pooled spatial variance within sampling rounds as

$$s_{\mathcal{S}h}^2(e) = \frac{\sum_{h=1}^{\ell}\sum_{i=1}^{n_h}(m_{hi}-1)s_{hi}^2(e)}{\sum_{h=1}^{\ell}\sum_{i=1}^{n_h}(m_{hi}-1)}, \quad \text{(A7)}$$

and estimate the temporal variance of errorless totals per psu by

$$s_{\mathcal{T}h}^2[t(e)] = \frac{1}{\ell}\sum_{h=1}^{\ell} s_h^2[t(e)]. \quad \text{(A8)}$$

The procedure described above is repeated 10,000 times, resulting in 10,000 independent estimates of $s_{\mathcal{S}h}^2(e)$, $s_{\mathcal{T}h}^2[t(e)]$ and $\hat{\overline{V}}^2$.

[41] **Acknowledgments.** This research is part of the strategic research program "Sustainable spatial development of ecosystems, landscapes, seas and regions" and is funded by the Dutch Ministry of Agriculture, Nature Conservation and Food Quality under BAS KB-01-001-003-ALT. We thank Wetterskip Fryslân (District Water Board) for collecting the data.

References

Bouma, J., J. de Vos, M. Sonneveld, G. Heuvelink, and J. Stoorvogel (2008), The role of scientists in multiscale land use analysis: Lessons learned from Dutch communities of practice, *Adv. Agron.*, *97*, 175–237.

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.

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.

Brus, D. J., M. J. W. Jansen, and J. J. de Gruijter (2002), Optimizing two- and three-stage designs for spatial inventories of natural resources by simulated annealing, *Environ. Ecol. Stat.*, *9*, 71–88.

Cochran, W. G. (1977), *Sampling Techniques*, John Wiley, New York.

Council of the European Communities (2000), Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, *Off. J. Eur. Commun.*, *L327*, 1–72.

de Gruijter, J. J., D. J. Brus, M. F. P. Bierkens, and M. Knotters (2006), *Sampling for Natural Resource Monitoring*, Springer, Berlin.

Dobbie, M. J., B. L. Henderson, and D. L. Stevens Jr. (2008), Sparse sampling: Spatial design for monitoring stream networks, *Stat. Surv.*, *2*, 113–153.

Efron, B., and R. Tibshirani (2006), *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Peterson, S. A., D. P. Larsen, S. G. Paulsen, and N. S. Urquhart (1998), Environmental auditing. Regional lake trophic patterns in the northeastern United States: Three approaches, *Environ. Manage.*, *22*, 789–801.

Southerland, M., J. Vølstad, E. Weber, R. Klauda, C. Poukish, and M. Rowe (2009), Application of the probability-based Maryland Biological Stream Survey to the state's asssessment of water quality standards, *Environ. Monit. Assess.*, *150*, 65–73.

Stevens, D. L., and A. R. Olsen (2003), Variance estimation for spatially balanced samples of environmental resources, *Environmetrics*, *14*, 593–610.

Stevens, D. L., and A. R. Olsen (2004), Spatially balanced sampling of natural resources, *J. Am. Stat. Assoc.*, *99*, 262–278.

Thomas, R. B. (1985), Estimating total suspended sediment yield with probability sampling, *Water Resour. Res.*, *21*, 1381–1388.

Thomas, R. B., and J. Lewis (1995), An evaluation of flow-stratified sampling for estimating suspended sediment loads, *J. Hydrol. 170*, 27–45.

U.S. Environmental Protection Agency (2002), Summary of biological assessment programs and biocriteria development for states, tribes, territories, and interstate commissions: Streams and wadeable rivers, *Rep. EPA-822-R-02-048*, Washington, D. C.

D. J. Brus and M. Knotters, Soil Science Centre, Wageningen University and Research Centre, PO Box 47, NL-6700 AA Wageningen, Netherlands. (martin.knotters@wur.nl)