Sampling of aquatic sediments. Design of a decision-support system and a case study

R Wehrens, P van Hoof, L Buydens and G Kateman

Laboratory for Analytical Chemistry, Catholic University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen (Netherlands)

M Vossen, W H Mulder and T Bakker

RIZA, P O Box 17, 8200 AA Lelystad (Netherlands)

(Received 1st June 1992, revised manuscript received 14th September 1992)

Abstract

The structure of a knowledge-based decision-support system for the sampling of aquatic sediments in lakes is discussed. The system is in the implementation stage and will advise on a sampling strategy that reveals as much as possible of the pollution level of the underwater soil, based on geomorphic information and analytical results for previous samples (if available). Constraints such as the desired precision and maximum sampling costs are taken into account. A case study on Ketelmeer lake (Netherlands) illustrates the techniques and heuristics that will be used in the decision-support system.

Keywords: Decision-support system, Sampling, Sediments, Waters

In each water system, there is a close relationship between the quality of (the suspended matter in) the surface water and the quality of the surface layers in the sediments underneath. As a consequence of water pollution in the past, the sediments in many industrialized countries are polluted in many places. Because of the close relationship between the quality of sediments and the quality of surface water, contaminated sediments can form a long-duration source of diffuse environmental pollution even when water pollution has been drastically reduced.

The potential risks for the environment have been the motivation in The Netherlands for the start of a water-bed cleaning programme in order to reduce the risks where needed. The programme is intended to reach a sediment quality

such that there are only negligible risks to the functioning of balanced aquatic ecosystems. In addition to the direct risks of contaminated sediments there are also problems of coping with contaminated dredging spoil. To describe the potential risks of water-bed pollution and in order to give practical directives for coping with contaminated dredging spoil, five categories of pollution have been defined in The Netherlands [1] (see Table 1).

In order to obtain an overview of the pollution of water-beds, a research programme throughout The Netherlands has been started. The aim is to categorize all water-beds in one of the above five categories. The category in which a sediment is classified depends on the number of compounds exceeding the standards (quality limits) that are defined for compounds most commonly present. Polluting compounds that are taken into account are heavy metals and organic micropollutants such as polycyclic aromatic hydrocarbons (PAHs),

Correspondence to: R Wehrens, Laboratory of Analytical Chemistry, Catholic University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen (Netherlands)

polychlorinated biphenyls (PCBs) and pesticides. Consequences such as water-bed cleaning or storage of contaminated dredging spoil in storage depots depend on the classification of the sediment or spoil.

One of the first problems in starting the research programme is the question of where to take the samples. Whereas a lot of work has been done on the sampling of surface waters, sampling of aquatic sediments has been neglected. Further, as the sampling is performed by different water authorities such as the water boards, it is of great importance that one uniform sampling strategy is followed in the different water districts, otherwise the results would not always be intercomparable.

This sampling problem has given rise to start the development of a knowledge-based decision-support system, BIAS (a Dutch acronym for sampling of aquatic sediments). In the last few years, more and more knowledge-based systems have been used in environmental applications [2]. Some systems for sampling strategies have also been developed [3], although they concerned sampling strategies for dry soil, whereas BIAS is meant to provide consistent sampling strategies for waterbeds. A first restriction was made to strategies for great lakes. The heuristics and statistical operations used in the expert system are validated using a number of test cases. From these, missing

knowledge may be extracted and existing knowledge may be refined.

Geomorphic information on the water system to be investigated and analytical results for previous measurements, if available, are included in the advice-forming model. General heuristics and statistically calculated parameters are then used to provide the optimum sampling strategy. Constraints such as the costs of sampling and the desired reliability should also be taken into account. The ideal sampling scheme of taking as many samples as possible is not, of course, feasible in practice.

In this paper, the structure of the knowledge-based system is discussed and an outline of the techniques used in the analysis of previous sampling data are given. Results of a case study on the Ketelmeer, a lake in The Netherlands, are discussed. Heuristics and plausible generalizations derived from this case study with respect to sampling strategies are presented. Finally, conclusions and directions for further research are given.

SETTING UP SAMPLING STRATEGIES WITH BIAS

Sampling strategies are used to minimize the amount of samples while still guaranteeing a specified accuracy [4]. To be able to choose the

TABLE 1

Classification of sediments according to level of pollution

Class	Standard	Consequences for water-bed cleaning policy	Consequences policy on dredging spoils
4	Warning value	Research into necessity for cleaning urgent because of risk to public health and environment	Processing under controlled conditions which become stricter the more the quality of the dredging spoils exceeds the test value
3			
2	Test value	Research into need for cleaning not urgent	Use and dispersal in the water possible under certain conditions
1	Quality objective 2000		
0	Target values	No cleaning	Use and dispersal in the environment possible. Waterbed quality may not deteriorate
0			No restrictions on use and dispersal in environment

most profitable trade-off value in the case of sampling of sediments, one must have a good idea of the geomorphic properties of the water system and the relationships between them and diffusion patterns of the sediments. For the compounds that are of interest here, different distribution patterns may be encountered because of different adsorption on sediment material. Knowledge about the sources of pollution is also very important in this respect. Analytical results for previous samplings will also be very helpful in devising a relevant sampling strategy, as they can provide a model for the spatial correlation in the soil. A knowledge-based system that advises on a sampling strategy for aquatic sediments should be able to combine these different types of knowledge and should contain both statistical and heuristic knowledge. In BIAS, this is achieved by a division into modules each with a distinct task. The structure of BIAS is given in Fig. 1.

Module 1 Input and data validation

In Module 1 the characteristics of the water system are given by the user. These include, amongst others, shape and depth of the water system, soil types, direction and strength of the

current, prevailing direction and strength of the wind, sources of pollution, inlet(s) and outlet(s) of the water system, shipping routes, harbours and special items such as dredging areas. If available, data on previous samplings can be read from a database. These data are validated first to rule out typing and other errors. If all data are validated, control is passed to Module 2.

Module 2 Definition of homogeneous subareas

In Module 2 the water system is divided into more or less homogeneous subareas. These areas then are treated separately. The rationale behind this subdivision is that for each subarea a separate sampling strategy may be feasible, in calm, deep waters where little resuspension takes place only small deviations from the area mean will be expected, and in many cases only a few samples will suffice. In shallow waters with a strong current, many more samples will have to be analysed to obtain a good overview of the local situation.

The division into subareas is based on a set of parameters that influence the pollution level of an underwater soil. They can be divided into the following:

- climatological conditions—wind, prevailing direction and strength,
- hydrological conditions—depth, and direction and speed of the current,
- geomorphological conditions—soil type,
- human activities—shipping, harbours, dredging, sources of pollution.

If results of previous samplings are available, a cluster analysis is done to see whether the clustering of the samples fits the predefined areas. If agreement is found, this means an extra affirmation that the division into subareas was done on relevant grounds. If no agreement is found, the data should be examined more carefully. In some instances, missing knowledge can be detected in this way, e.g., a sample location may be influenced by a parameter that has been overlooked in the input phase. Outliers may also be detected in this way.

Module 3 Analysis of subareas

If results of previous samplings are available, they are analysed here to estimate pollution lev-

Fig. 1 The structure of BIAS

els at unsampled locations. The assumption is that locations that are spatially close to each other tend to have similar soil characteristics. Two techniques can be used here: a two-dimensional interpolation technique often used in mining, called kriging [5,6], and a one-dimensional technique, stemming from time series analysis and using autocorrelation [7,8]. Because all spatial correlations can be taken into account in the two-dimensional plane, kriging will give better predictions. Autocorrelation methods, on the other hand, do not require as many data as kriging techniques and are therefore applicable in (frequently occurring) cases where relatively few data are available.

Both methods require a set of conditions to be fulfilled to produce useful predictions. First, the variance of the variable to be estimated depends only on the lag vector between the two locations. Often this condition is replaced by an even stronger one in which only the length of the lag is taken into account and the direction is neglected. This, however, will yield incorrect results in many instances as different correlations will exist in different directions. Second, the variable is stationary, i.e., no drift is present. Effectively, this means that the a priori expectation for all locations in the area is equal. In practice, additional knowledge about the area will often make this assumption false. The subareas obtained in the previous module are meant to satisfy these conditions. In general, it is very difficult to determine whether this is the case. The most popular method to validate the method is cross-validation, where values at sampled locations are predicted using the other locations. However, good predictions still are no guarantee that the conditions are satisfied.

Kriging and related methods Kriging can be seen as an optimum unbiased interpolation method, based on known values in sampled locations, values at other locations are predicted. A measure indicating the amount of correlation between two locations separated τ times a lag distance h apart is the semivariance γ_τ . The semivariance for a lag τ is given by

$$2\gamma_\tau = E\{[z(x) - z(x + \tau h)]^2\} \quad (1)$$

Fig. 2. Idealized example of a semivariogram. In this instance a linear model is chosen to represent the data (see text).

and can be estimated by the equation

$$\hat{\gamma}_\tau = \frac{1}{2n(\tau)} \sum_{i=1}^{n(\tau)} [z(x_i) - z(x_i + \tau h)]^2 \quad (2)$$

where $z(x_i)$ and $z(x_i + \tau h)$ are the values of the variable of interest, x , at location i and a location τ lags away, respectively (in kriging and related topics, variables are termed regionalized, indicating that the variables are a non-random sample from one realization of a random function, this is different from the usual idea of multiple realizations of a random variable but for the present purposes, this difference can be ignored). The number of pairs $[z(x_i), z(x_i + \tau h)]$ is given by $n(\tau)$. The semivariogram, in which semivariances are plotted against the lag number (see Fig. 2), is modelled by fitting a simple function through the data points. Linear, exponential and spherical models are often used. At lag zero, the semivariance is 0. In many instances a discontinuity is present called the nugget effect. This indicates that there is significant variance over distances smaller than the lag distance. At large distances, the semivariance is equal to the variable variance and reaches a ceiling, called the sill. These parameters, and the range where the semivariance increases with the lag, are indicated in Fig. 2.

The value of a variable x at an unsampled location is derived from a linear combination of nearby points. Kriging attempts to minimize the

prediction error This can be achieved by solving the following matrix equation

$$\begin{pmatrix} \gamma(h_{1,1}) & \gamma(h_{1,n}) & 1 \\ \gamma(h_{1,n}) & \gamma(h_{n,n}) & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} W_1 \\ W_n \\ \lambda \end{pmatrix} = \begin{pmatrix} \gamma(h_{1,p}) \\ \gamma(h_{n,p}) \\ 1 \end{pmatrix} \quad (3)$$

or the equivalent form

$$\mathbf{A} \mathbf{w} = \mathbf{b} \quad (4)$$

where $h_{i,j}$ is the distance between the sampled locations i and j , $\gamma(h_{i,j})$ is the semivariance between locations i and j , $\gamma(h_{i,p})$ is the semivariance between the sampled location i and the unsampled location p , obtained from the semivariogram, W_i is the weight given to location i in the prediction of location p and λ is the La-

grange multiplier The solution for the unknown vector $\mathbf{w}$ is simple

$$\mathbf{w} = \mathbf{A}^{-1} \mathbf{b} \quad (5)$$

This gives for an estimate on location p the value

$$\hat{x}_p = \mathbf{w} \mathbf{x} \quad (6)$$

with prediction error variance

$$s_e^2 = \mathbf{w} \mathbf{b} \quad (7)$$

The weights obtained in this way are optimum in the sense that the predictor is unbiased and has a minimum prediction error variance. The predictor is also exact, which means that the value at a sampled location is predicted exactly and with zero error. Contour maps may be drawn that provide a global view of the situation.

Several other types of kriging exist. In universal kriging [9,10], a polynomial trend is permitted. In co-kriging [11,12], the value of the variable of interest in an unsampled location is determined not only by interpolation from sampled locations, but also by measurements of correlated variables. Disjunctive kriging [13–15] is a variant in which

Fig 3 Example of an autocorrelogram

the conditional probability that a soil value exceeds a predefined threshold can be calculated and plotted

Autocorrelation methods Autocorrelation techniques stem from time series analysis. They are based on the assumption that the value of a measurement can be related to the value of previous measurements. Time can also be replaced with distance, and then the assumption is that the value of a variable at a specific location is related to values at nearby locations. The methods are inherently one-dimensional.

In autocorrelation techniques, autocovariance is used rather than semivariance to express the similarity between two measurements τ times a lag distance h apart

$$\begin{aligned} \text{cov}(x, x + \tau h) \\ = E\{[z(x) - \mu][z(x + \tau h) - \mu]\} \end{aligned} \quad (8)$$

where μ is the process mean value. If only a few measurements are available, the autocovariance can be estimated by

$$\begin{aligned} \widehat{\text{cov}}(x, x + \tau h) = \frac{1}{n - \tau - 1} \sum_{i=1}^{n-\tau} [z(x_i) - \mu] \\ \times [z(x_i + \tau h) - \mu] \end{aligned} \quad (9)$$

where n is the total number of measurements. The (dimensionless) autocorrelation then is defined by

$$\Phi_{x,x}(\tau) = \text{cov}(x, x + \tau h) / s_x^2 \quad (10)$$

where s_x^2 denotes the variance of x . In an autocorrelogram the autocorrelation is plotted against τ . In many instances, linear first-order models are found to be satisfactory. They give an exponentially decreasing curve such as that depicted in

Fig. 4 Relationship between autocovariance $\text{cov}(x, x + \tau)$, semivariance γ_τ and process variance σ_0^2

Fig 3 The variance of such an autocorrelation function can be estimated by

$$\hat{s}^2[\Phi_{x,x}(\tau)] = \frac{1}{N-\tau} \left[\frac{(1 + e^{-2/T_x})(1 - e^{-2\tau/T_x})}{1 - e^{-2/T_x}} - 2\tau e^{-2\tau/T_x} \right] \quad (11)$$

where T_x and τ are given in lag units. From this function, the upper and lower confidence limits for the autocorrelation function can be calculated. These are depicted in Fig 3 as dotted lines.

The distance at which the value of the autocorrelation has decreased with a factor of e^{-1} is called the correlation distance T_x . If the autocorrelation at this distance differs significantly from zero, the correlation distance can be seen as a measure of the maximum distance at which a parameter still shows significant autocorrelation (if autocorrelation is used in time series analysis, the parameter T_x is called the time constant). A large correlation distance indicates that the value of a certain parameter does not change much with distance. Whereas in many instances it is not possible to calculate an upper confidence limit for T_x because the upper border diverges from the exponential curve, it is possible to calculate a minimum value $T_{x,min}$. This is the value where the lower confidence limit has decreased with a factor of e^{-1} .

In practice, the correlation distance may be calculated in a direction parallel as well as perpendicular to the current. If the two are equal, the system is isotropic, in most instances, however, anisotropic systems will be found.

The autocorrelation equations are very closely related to the semivariance equations. If the variable is stationary (i.e., the expected value of the variance is equal in the whole area), the semivariance for a distance h is equal to the difference between the variance σ_0^2 and the spatial autocovariance σ_n^2 for the same distance (see Fig 4). If the variable is also standardized to have a mean of zero and a variance of 1.0, the semivariogram is a mirror image of the autocorrelation function.

Comparison of autocorrelation methods and kriging. From the above, some conclusions can be

drawn. Both methods require the same assumptions to be valid, with the added condition in autocorrelation techniques that the process variance, s_x^2 , exists. Kriging is the preferred method if many data are available and a reliable semivariogram can be set up. The interpolations take full advantage of the information in the semivariogram, and prediction variances can be plotted. Also sampling strategies can readily be obtained (see below). In cases where only a few data points are sampled, however, often no semivariogram can be calculated. It has been estimated that at least 100 data points, depending on the size and structure of the sampling area, are necessary in order to perform an accurate geostatistical analysis. This value may be different with aquatic sediments, and future research should clarify this point. Until then, the user will have to decide whether the number of data points is sufficient to provide a semivariogram or not. In the latter instance, autocorrelation techniques may provide a more reliable estimate. The upper and lower limits on the autocorrelogram provide useful information on the representativity and reliability of the samples. Although autocorrelation techniques are in most instances inappropriate for stratified objects such as soil and rock [16], they can be used here because of the dynamic nature of the sedimentation process. Another advantage of autocorrelation techniques is that typical values for correlation distances in different water systems can be tabulated and used as a first estimate in case the development of a sampling strategy has to start from scratch.

Module 4 Calculation of a sampling strategy

Several sampling strategies are possible, depending on the aim of the investigation: a strategy to obtain a global view of concentrations of (polluting) compounds with a predetermined reliability, a strategy to obtain an idea of concentrations of (polluting) compounds, as reliable as possible, given the allowed costs, a strategy that does not focus on exact concentrations of (polluting) compounds but more on classifications of soil in pollution classes with a given reliability, a strategy used for monitoring. Each strategy has its own requirements. For both autocorrelation

techniques and kriging and related techniques, such sampling strategies may be set up

Sampling strategies and kriging If kriging techniques are applied to develop sampling strategies, several situations may occur [17]. In the simplest case the semivariogram is known. It is then possible to calculate the prediction variances of several sampling set-ups irrespective of the values that will be found. A triangular grid has been shown to yield optimum results, but a cubic grid is in most instances almost as good and easier to achieve. If only additional samples are needed, their location can be derived from the kriged contour map of the prediction variances [18]. If, for instance, only one location has to be added to a set of sampled locations, this can best be done at the place where the prediction variance is largest. In this way, the overall standard deviation is minimized.

If no variogram has been determined previously, a sampling strategy should aim to optimize the sample locations with respect to variogram estimation [19,20]. For this, the sampling locations should be chosen in such a way that all lags contain an approximately equal number of pairs. In general, a nested sampling design is chosen. Whereas many samples are needed to obtain a reliable variogram, for a reliable interpolation given a good variogram many fewer samples will suffice.

In the frequently occurring situations where different variables are correlated, co-kriging methods may be used [11]. In such a case the costs of analysis of different compounds may be taken into account, and cheap analyses may be performed much more often than expensive ones. In this way, it is possible to obtain reliable estimates at much lower cost. Finally, disjunctive kriging may be used to determine what pollution classes are present in the water system, and if the uncertainty of the presence of a certain class is too high additional samples may be taken.

Sampling strategies and autocorrelation techniques Autocorrelation techniques have been used earlier to predict an optimum sampling frequency for the surveillance of surface water quality [21], and also other applications have been reported [22]. Previously, equations have been

derived to calculate the prediction variance, σ_{est}^2 , as a function of the number of samplings n , the process variance σ_0^2 and the size and spacing of the samples [8]. If the sample size is very small compared with the size of the area, the following equation holds:

$$\frac{\sigma_{\text{est}}^2}{\sigma_0^2} = \frac{1}{n} \left[1 + \frac{2e^{-a}}{1 - e^{-a}} - \frac{2e^{-a}(1 - e^{-p})}{n(1 - e^{-a})^2} \right] - \frac{2}{np} \left(2n - \frac{1 - e^{-p}}{1 - e^{-a}} + \frac{1 - e^{-p}}{1 - e^{-a}} \right) + \frac{2}{p^2} (p - 1 + e^{-p}) \quad (12)$$

where a is the distance between two adjacent samples and p is the length of the sampled area, both divided by the correlation distance. This equation has to be solved iteratively to obtain the distance between adjacent samples in the area in which the correlation distance holds. Autocorrelations obtained in different directions must be used to determine the distances between samples in the grid. Again, if the autocorrelogram has been determined with sufficient reliability, the number of samples needed for a reliable prediction is much smaller than the number of samples needed to set up a good autocorrelogram. In this respect the same situation as in the kriging case is encountered.

Constraints on sampling strategies Any proposed sampling scheme should be reasonable with respect to constraints such as the number of analyses and the total costs associated with sampling. If the sampling strategy proposed to obtain a required reliability cannot be performed with the means available, a new scheme will have to be set up. Either the reliability will be lower, if the cost constraint is satisfied, or the sampling operation will be more expensive. In both instances, the system may be consulted again with different input values for the constraints. The use of correlation between different compounds may be a very important tool to diminish the costs of sampling either using a few compounds as guiding parameters or using co-kriging or related methods will decrease the number of samplings of the associated costs.

Fig 5 The Ketelmeer

THE KETELMEER

The Ketelmeer is a lake 12 km long and 4 km wide, in central Netherlands (Fig. 5). It is shallow, most areas being 2–3 m deep. In the very shallow (< 1 m) part in the east of the lake, the sediment consists of sand, in the other parts the sand is covered with more loam-like sediment. The river IJssel debouches into the south-east corner of the lake and water flows out from the north-east side. There is some shipping and two small harbours can be found on the north and south coasts. The prevailing direction of the wind is south-west. In the following sections, the Ketelmeer will be treated as an example of how a sampling strategy is set up in BIAS. The sections follow the general structure of BIAS: input and data validation, definition of homogeneous subareas, analysis of subareas and calculation of the required sampling strategy.

Validation of previous sampling data of the Ketelmeer

As the Ketelmeer has long been polluted, a relatively large amount of sampling has already taken place [23]. A data set of sampled locations was selected on the following criteria. All locations should be sampled to obtain at least 80% of the top layer. The top 10–20 cm are easily disrupted, even in calm and sheltered areas. Moreover, the degree of pollution of sediment has changed considerably with time, and comparable periods of time should be represented in the samples. Care was taken not to include locations where more than just the top layer was sampled, because the underlying sand will be much cleaner and the samples would not be intercomparable. The depth at the sampled locations should not have changed much since the sample was taken. If significant suspension or erosion has taken place since then, the sample is not representative for the current situation. Local human activities should be taken into account. If, e.g., a harbour is present near one of the sampled locations, or dredging has taken place, again these samples are not representative for the area. Further, most human activities constitute a direct source of pollution and should be considered if samples

Fig. 6 Schematic drawing of the Ketelmeer. The numbers on the axes are the distances in kilometres to a reference point in Paris, France. Inlet and outlet are indicated by arrows. Harbours are indicated by H. Each point indicates a sampling location. Subareas are drawn and indicated with roman numerals.

have to be taken in an area where much activity takes place.

Definition of homogeneous subareas in the Ketelmeer

At all sampled locations included in the analysis, the same soil type should be present, because different sediment compositions will show different degrees of pollution. Because of the availability of the data, heavy loam [12–25% clay (particles smaller than $2\mu\text{m}$)] was selected. This caused the analysis to concentrate on the central and western part of the lake. In the sandy part in the east, not enough samples were taken for a reliable analysis. Overall, 23 of the locations that satisfied the conditions in module 1 were in the heavy loam area. These are drawn in Fig. 6. The heavy loam part was subdivided into a part sheltered from the wind along the southern dike (area I in Fig. 6 containing seven sampling locations), a central part, exposed to the wind (area II, containing ten locations), a shipping route (the narrow area above area I) and a region containing sand-pits in the western part (left of area II).

A cluster analysis using Euclidean distance confirmed the above subdivision. Further, an area along the northern dike (north of area II) was defined because cluster analysis showed that the samples in this area were not correlated with the other samples in area II, probably because the distances between them were too large. Two sampling locations fell in the shipping route where,

because of dredging, no reliable estimates could be made. One location was very near a harbour and was also excluded. Some locations were considered to be outliers, such as the middle sampling location in area I. Eventually, areas I and II, containing seventeen sampling locations, were used in the following analysis, and sampling strategies for these areas were set up. If sampling strategies have to be set up in areas where no samples have been taken, it must be done on heuristic grounds.

Analysis of subareas of the Ketelmeer

As the Ketelmeer is essentially oblong-shaped, the current plays an important role in the sedimentation process. As the data were scarce, no relevant models could be fitted through the semi-variograms and kriging techniques could not be applied. Therefore, autocorrelation techniques were used. For each of the subareas I and II in Fig. 6 the autocorrelograms for all compounds to be analysed for are calculated. Because over 90% of the pollution comes from the river IJssel, autocorrelograms are drawn using one dimension, the distance to the mouth of the river. This dimension is more or less equal to the direction of the current. Perpendicular to the chosen direction, too few samples were taken to obtain an autocorrelogram.

Fig. 8 Autocorrelogram for fluoranthene in the central area II of the Ketelmeer

The autocorrelograms for fluoranthene in the two areas are depicted in Figs. 7 and 8. Correlation distances are calculated by fitting exponential curves to the data points. Correlation distances are given in Table 2. The values indicated are class means, i.e., the value for the heavy metals is the mean of the value found for Cu, Cd, Cr, Hg, Ni, Pb and Zn. As can be seen, the correlation distance in area I is much larger than that found in area II. This means that in area I the amounts of fluoranthene in the soil change more slowly on going from one place to another than in area II. Also, the minimum correlation distance in area I is much larger than that in area II. This is what would be expected, as the influence of the wind in the sheltered part is much less than the influence in the central part, where currents can easily be induced.

The 95% confidence limits are shown. It is clear that the resulting autocorrelograms are not very reliable. The reason for this is that the correlation distances found are smaller than or equal to the lag distance. In some instances in area II, no correlation distance could be calculated at all. The only way to improve the precision of the functions is to use more sampled locations to set up the autocorrelation functions. At present, there is no reason to doubt the valid-

Fig. 7 Autocorrelogram for fluoranthene in the sheltered area I of the Ketelmeer

TABLE 2

Correlation distances in the Ketelmeer^a

Substance class	T_x (area I)	$T_{x \min}$ (area I)	T_x (area II)	$T_{x \min}$ (area II)
Heavy metals	530	250	390	140
PCBs	460	210	250	110
PAHs	800	320	180	80

^a Distances are given in metres. Areas I and II refer to Fig. 6. The lag distance in area I is 1000 m and in area II 500 m. For three compounds in the PAH class, two compounds in the PCB class and one heavy metal (Cd), no autocorrelation function could be set up in area II because the autocorrelation at the first lag was already less than zero. The figures given are based on the other compounds in the classes.

ity of the first-order linear model, as all data points lie well within the confidence limits.

Calculation of a sampling strategy for the Ketelmeer

Although far from being optimum, the above results reveal a lot of information that can be used to set up a sampling scheme. Because the correlation distances that have been found are in the same order of magnitude as the sampling distances, the resulting autocorrelograms are not very reliable. Therefore, additional sampling should in the first place aim to improve the reliability of the autocorrelograms, which means that extra samples have to be taken at smaller distances from each other. At least three lags should lie within the correlation distance. If one considers the values for the heavy metals, which are determined with the greatest accuracy, this criterion would yield a distance between adjacent samples of 200–250 m for area I and 100–150 m for area II. As these values are estimated from the correlation distances in the direction of the current, they are valid in that direction only. As a first approximation, the same values may be used in the perpendicular direction. It is not necessary to sample the complete (sub)area with the same frequency as soon as there are enough data points to set up a reliable autocorrelogram, the sampling frequency may be decreased to obtain the required precision (cf., Eqn. 12).

For example, if it is assumed that the above correlation distances are correct and if it is wanted

that the reliability, expressed as the estimation variance divided by the process variance, should have a value of 0.1, then it can be calculated that for area I (length 7 km, mean correlation distance 600 m) one row of five samples is enough. Similarly, for area II (length 6 km, average width 2 km, mean correlation distance 260 m) a grid with spacings of 1000 m in the direction of the current and 500 m in a perpendicular direction suffices. Here, the correlation distance found in the direction of the current is also used in the perpendicular direction (Eqn. 12 also contains the length of the area sampled, and therefore the spacings in the two perpendicular directions are not equal). It is clear that sampling where a reliable autocorrelogram is known is much cheaper than sampling to obtain the autocorrelogram, for the latter objective, in this instance more than seven times as many samples are needed in both areas.

Extrapolations to other parts of the lake will have to be made, based on heuristics. This will provide a complete sampling scheme for the lake. If the costs for this sampling scheme violate the constraints concerning budgets, a lower reliability must be specified, or some subareas will have to be treated separately, with a smaller sampling frequency.

General results obtained in the Ketelmeer case study

For the Ketelmeer, it has been shown that autocorrelation methods can be useful in devising sampling strategies for aquatic sediments. Despite the small number of data points, useful information has been obtained to guide further sampling in the area. The correlation distance determined from previous samplings or geomorphic information can be used to fulfil two objectives simultaneously: to devise a sampling scheme that permits a more precise estimation of the correlation distance, and a sampling scheme that yields a subsequent interpolation to unsampled locations of a specified accuracy. For the first goal, two perpendicular rows of samples may be taken over a distance of four times the correlation distance, with a spacing of one quarter of the correlation distance. This yields a total of sixteen samples. In smaller subareas (such as area I in

the Ketelmeer), fewer samples or one row of samples may suffice. The second goal is achieved by calculating the distances between samples in a rectangular grid that will provide an interpolation with a desired accuracy. In general, the number of samples needed to satisfy the second goal is much smaller than that required to satisfy the first.

The results of the Ketelmeer study, notably the correlation distances and sampling schemes derived from them, may also be extrapolated to other water systems, where depth, current and other parameters are comparable. What parameters exactly are of influence is still a subject of investigation. In cases where different conditions are met, results will be different. For instance, the influence of the wind that was clearly present in the Ketelmeer will be much smaller with deep (> 5 m) waters. There, sedimentation will proceed much more calmly, and larger correlation distances will be expected there. In the Ketelmeer, it proved to be necessary to analyse only those samples which constituted a large portion of the top layer. Correlation values obtained from locations where only small portions of the top layer were sampled showed large scatterings because of continuous whirlings of the top sediment. In deeper waters, this will not be so much of a problem. In the Ketelmeer, good correlations were found between the concentrations of the different compounds of interest. It may be expected that this is generally the case when all pollution stems from the same source. In such a situation, some of the compounds may serve as guiding parameters for others that are perhaps difficult to measure or for which expensive methods are necessary.

Conclusion

The importance of consistent and reliable sampling strategies for underwater soils cannot be overestimated. The amount of money that can be saved and the amount of extra information that can be obtained using sensible sampling schemes are considerable. In this paper, an approach has been presented that aids the water managers of the separate districts in The Netherlands in setting up such strategies. A case study, the

Ketelmeer, has been used to test the validity of the approach and to test the knowledge for a practical situation. The knowledge that is acquired in this way will be implemented in a knowledge-based system, BIAS. More test cases will follow, thus enabling the system to be tested on a wide variety of lakes. This is necessary in order to be able to set up heuristics that can help in the advice-forming module when no previous sampling data are available. Although some heuristics have already been established, the list is far from being complete yet. In the near future, new test cases will be tackled and extra samplings on test cases already analysed will be performed.

The authors thank H.J. Winkels and H.C. Klavers for valuable comments and discussions. This research was carried out with financial support from the Dutch Ministry of Transport and Public Works.

REFERENCES

- 1 Water in The Netherlands: a Time for Action, Ministry of Transport and Public Works, The Hague, 1991.
- 2 J.M. Hushon (Ed.), *Expert Systems for Environmental Applications* (ACS Symposium Series, No. 431), American Chemical Society, Washington, DC, 1990.
- 3 R.A. Olivero and D.W. Bottrell, in J.M. Hushon (Ed.), *Expert Systems for Environmental Applications* (ACS Symposium Series, No. 431), American Chemical Society, Washington, DC, 1990, pp. 69–81.
- 4 G. Kateman and F.W. Pijpers, *Quality Control in Analytical Chemistry*, Wiley-Interscience, New York, 1981.
- 5 A.G. Journel and Ch. Huybrechts, *Mining Geostatistics*, Academic Press, London, 1978.
- 6 D.E. Myers, *Chemometr. Intell. Lab. Syst.*, 11 (1991) 209.
- 7 G.E.P. Box and G.M. Jenkins, *Time Series Analysis*, Holden-Day, San Francisco, 1976.
- 8 P.J.W.M. Muskens and G. Kateman, *Anal. Chim. Acta*, 103 (1978) 1.
- 9 A. Stein and L.C.A. Corsten, *Biometrics*, 47 (1991) 575.
- 10 R. Webster and T.M. Burgess, *J. Soil Sci.*, 31 (1980) 505.
- 11 A.B. McBratney and R. Webster, *J. Soil Sci.*, 34 (1983) 137.
- 12 D.E. Myers, *Math. Geol.*, 14 (1982) 249.
- 13 R. Webster and M.A. Oliver, *J. Soil Sci.*, 40 (1989) 497.
- 14 S.R. Yates, A.W. Warrick and D.E. Myers, *Water Resour. Res.*, 22 (1986) 615.
- 15 S.R. Yates, A.W. Warrick and D.E. Myers, *Water Resour. Res.*, 22 (1986) 623.

- 16 G Kateman, *Top. Curr. Chem.*, 141 (1987) 43
- 17 T M Burgess, R Webster and A B McBratney, *J. Soil Sci.*, 31 (1981) 643
- 18 H J Di, B B Trangmar and R A Kemp, *Soil Sci. Soc. Am. J.*, 53 (1989) 1163
- 19 D Russo and W A. Jury, *Soil Sci. Soc. Am. J.*, 52 (1988) 1228
- 20 A W Warrick and D E Myers, *Water Resour. Res.*, 23 (1987) 496
- 21 P J W M Muskens, *Anal. Chim. Acta*, 103 (1978) 445
- 22 J Einax, B Machelett, S Geiss and K Danzer, *Fresenius' Z. Anal. Chem.*, 342 (1992) 267
- 23 H J Winkels and A van Diem, *The Makeup and Quality of the Ketelmeer Soil* Technical Report, Ministry of Transport and Public Work, The Hague, 1991, Flevobbericht 325 (in Dutch)