ELSEVIER

Journal of Hydrology 200 (1997) 364–386

Journal of Hydrology

Mapping groundwater quality in the Netherlands

Edzer J. Pebesma[1,*], J.W. de Kwaadsteniet

National Institute of Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, The Netherlands

Received 18 December 1995; revised 14 February 1997; accepted 14 February 1997

Abstract

Maps of 25 groundwater quality variables were obtained by estimating 4 km × 4 km block median concentrations. Estimates were presented as approximate 95% confidence intervals related to four concentration levels mostly obtained from critical levels for human consumption. These maps were based on measurements from 425 monitoring sites of national and provincial groundwater quality monitoring networks. The estimation procedure was based on a stratification by soil type and land use. Within each soil–land use category, measurements were interpolated. Spatial dependence between measurements and regional differences in mean level were taken into account. Stratification turned out to be essential: no or partial stratification (using either soil type or land use) results in essentially different maps. The effect of monitoring network density was studied by leaving out the 173 monitoring sites of the provincial monitoring networks. Important changes in resulting maps were assigned to loss of information on short-distance variation, as well as loss of location-specific information. For 12 variables, maps of changes in groundwater quality were made by spatial interpolation of short-term predictions calculated for each well screen from time series of yearly measurements over 5–7 years, using a simple regression model for variation over time and taking location-specific time-prediction uncertainties into account.

From a policy point of view, the resulting maps can be used either for quantifying diffuse groundwater contamination and location-specific background concentrations (in order to assist local contamination assessment) or for input and validation of policy supporting regional or national groundwater quality models. The maps can be considered as a translation of point information obtained from the monitoring networks into information on spatial units, the size of which is used in regional groundwater models. The maps enable location-specific network optimization. In general, the maps give little reason for reducing the monitoring network density (wide confidence intervals). © 1997 Elsevier Science B.V.

Keywords: Groundwater quality; Monitoring; Mapping; Statistics; Geostatistics; Optimization; Stratification; Network density

* Corresponding author.

[1]Present address: Landscape and Environmental Research Group, Faculty of Environmental Sciences, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 W Amsterdam, The Netherlands.

PII S0022-1694(97)00027-9

1. Introduction

In the first half of the 1980s, the national groundwater quality monitoring network (NGM) was installed for the monitoring of groundwater quality and changes in groundwater quality in the Netherlands (Van Duijvenbooden et al., 1985). For this purpose 370 monitoring sites, spread fairly evenly across the country, are sampled yearly at two depths (from well screens at about 10 m and 25 m below the surface). Samples were analysed for at first 19, and later 25, groundwater quality variables.

More recently (during the first half of the 1990s), the 12 provinces of the Netherlands started to install similar monitoring sites (the provincial groundwater quality monitoring networks, the PGMs), in order to study groundwater quality at a denser spatial resolution. Currently, the PGMs have as many monitoring sites as the original NGM.

For all variables measured (Table 1) maps were made, using measurements from both monitoring networks. Uncertainties in predicted values were quantified and presented on the maps. Full results were presented in separate reports (Pebesma and De Kwaadsteniet, 1994, 1995). Here we illustrate the methods used with results on one groundwater quality variable, namely potassium concentration.

From geohydrology we know that many factors influence groundwater quality, like soil type, land use and geohydrological situation (e.g. depth of groundwater table, infiltration or seepage, marine influence, precipitation). In this study, soil type and land use can be distinguished from the other factors since the data on Dutch soil type and land use are available as digitized maps (Steur et al., 1985; Thunnissen et al., 1992). Hence, we can use them to help make groundwater quality maps, and we stratified by soil type and land use prior to the extrapolation. The effects of partial or no stratification were also investigated.

For each soil–land use category and for each groundwater quality variable, attention was paid to the level of, the variation in and the spatial dependence in the values measured (in 1991) or the values predicted (for 1980 and 2000) by using kriging interpolation (Journel and Huijbregts, 1978) within categories. Non-stationary, location-specific uncertainties in time prediction were taken into account by a simple modification of the spatial interpolation procedure for the time predictions (Delhomme, 1978).

Since a groundwater quality monitoring network that has a national or provincial extent can by no means provide accurate estimates of groundwater quality variables for specific unmeasured locations, a coarser spatial resolution for the groundwater quality maps was adopted, and block mean values of 4 km × 4 km square elements were estimated for a 2 km × 2 km grid covering the whole country, using block kriging (Journel and Huijbregts, 1978). The size of the square elements was chosen with the spatial differentiation of soil type and land use in the Netherlands in mind. The resulting maps of block mean values of groundwater quality (or, more precisely, block median values) can be regarded as a

Table 1
The 25 groundwater quality variables measured in the Dutch monitoring networks (NGM and PGMs) in 1991

Al, As, Ba, Ca, Cd, Cl, Cr, Cu, DOC, EC, Fe, pH, HCO_3, K, Mg, Mn, Na, NH_4–N, Ni, NO_3–N, Pb, P–tot, SO_4, Sr, Zn

translation of point information obtained from the monitoring networks into information on spatial units, the size of which is used in regional groundwater models.

The effect of monitoring network density was quantified by comparing maps based on the full dataset (NGM + PGMs) with maps based on the information originally available from the NGM.

Sections 2–4 describe the stratification, data selection and spatial interpolation. Section 5 explains the groundwater quality maps. Effects of stratification and monitoring density

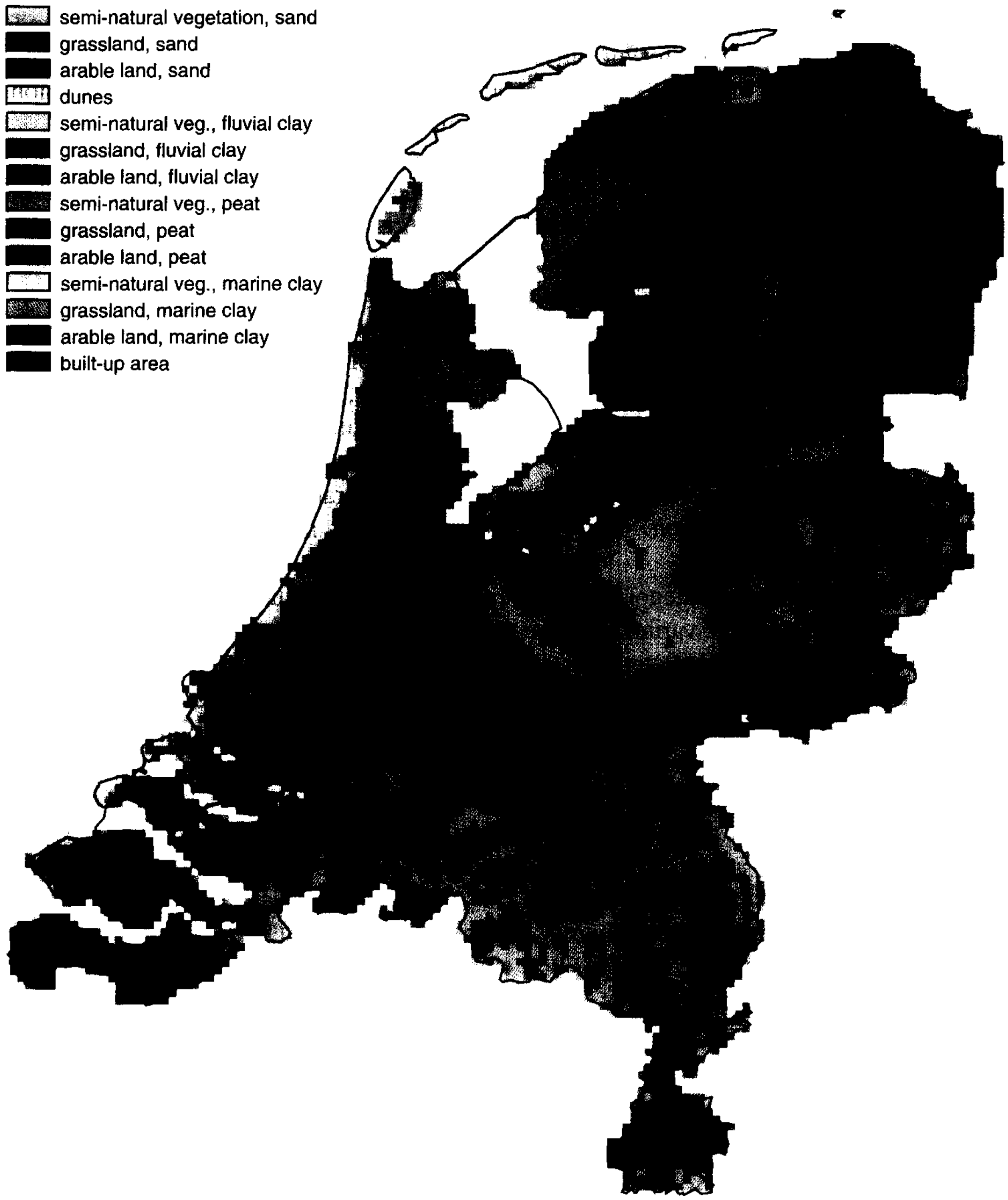

Fig. 1. Soil type and land use map. Dominant soil–land use categories for 2 km × 2 km cells.

are presented and discussed in Sections 6 and 7. Section 8 deals with the mapping of temporal changes and Section 9 concludes with a discussion.

2. An aggregated map comprising soil type and land use

The available maps regarding soil type (Steur et al., 1985) and land use (Thunnissen et al., 1992) are so detailed that hundreds of combinations of classes can be made. For the purpose of this study such detail does not make sense: many categories would represent only a small fraction of the total area and would consequently contain too few (or no) measurements. Therefore, and analogously with considerations on the design of the monitoring network (Van Duijvenbooden et al., 1985), we aggregated classes to the 13 categories—apart from built-up areas—shown in Table 2. Here, semi-natural vegetation comprises forests, heaths and nature reserves. The category 'dunes' is, strictly speaking, not a soil–land use category, but was obtained from a map of ecological districts (Klijn, 1988).

For each element on a 2 km × 2 km grid the dominant soil–land use category was determined from an overlay of both aggregated maps. The result is shown in Fig. 1. The soil type and land use information for measurements was derived from this map.

3. Data selection

For the generation of the groundwater quality maps a number of measurements were left out because they seemed unsuitable for this purpose. For instance, monitoring sites from

Table 2

Basic information: soil–land use categories, area covered by a category, number of selected monitoring sites from the national groundwater quality monitoring network for each category, and number of selected monitoring sites from the provincial groundwater quality monitoring networks for each category. The spatial distribution of this information is shown in Figs. 1 and 2

Category	Area	Number of monitoring sites		
	(%)	NGM	PGMs	Total
semi-natural vegetation on sand	10	36	26	62
grassland on sand	25	79	45	124
arable land on sand	7.1	29	15	44
dunes	2.8	11	4	15
semi-natural vegetation on fluvial clay	0.2	0	2	2
grassland on fluvial clay	8	25	26	51
arable land on fluvial clay	0.5	1	0	1
semi-natural vegetation on peat	0.3	1	4	5
grassland on peat	8.6	24	21	45
arable land on peat	0.2	2	1	3
semi-natural vegetation on marine clay	1	2	2	4
grassland on marine clay	9	16	10	26
arable land on marine clay	15.5	26	17	43
built-up area	11.8			
TOTAL	100	252	173	425

the category 'built-up area' were left out because the soil map (Steur et al., 1985) does not provide information about soil type in built-up areas. Extrapolation from measurement information, taking soil type into account, is therefore impossible in built-up areas with the information available, and no estimates for this class were made. Monitoring sites that could not be matched to a soil–land use category (Fig. 1) because they were influenced directly by water from rivers or the sea were also left out. Fig. 2 shows the locations of the

Fig. 2. Monitoring network locations. Sites of the national groundwater quality monitoring network (●) and the provincial groundwater quality monitoring networks (+) that contributed to the groundwater quality maps.

425 monitoring network sites that contributed to the groundwater quality maps. Their distribution over the soil–land use categories is summarized in Table 2.

The groundwater quality maps for 1991 were based on the values of 25 groundwater quality variables (Table 1) in the selection of monitoring sites described above, with monitoring screens all between 5 and 17 m below ground surface. The maps of changes in time of groundwater quality were based on the values of the twelve groundwater quality variables of the NGM for which long-term series were available (Table 3). Detection limit values were assigned to measurements below the detection limit (using the highest value when different detection limits occurred—in exceptional cases, when only a tiny fraction of the measurements was determined with a considerably higher detection limit, measurements below such a high detection limit were treated as missing values).

Visual examination of the time series plots for each variable and for each monitoring screen revealed that for many variables the first years of monitoring (1985 and 1986) did not fit the general pattern of the series for a substantial fraction of screens, leading to doubts about the quality of these measurements (quality of sampling, sample treatment or analysis). We opted for uniform treatment, and for all variables and monitoring screens only measurements from 1987 and later were used. Usually, series up to 1992 or 1993 were available. Visual examination of the time series plots revealed, furthermore, that for some samples a considerable part (more than three) of the variables did not fit in the series, which was interpreted as an indication of erroneous sampling or sampling treatment. For this reason, three samples were left out of the maps.

In addition, some doubtful measurements were left out. In the quality control we used the measurements from the NGM as a reference set. The measurements from the NGM themselves were checked only marginally: a measurement from the NGM set was left out from the mapping only if (i) the measurement differed by more than a factor of 10 from all other measurements of that variable in the same monitoring screen in the period 1987–1992, and (ii) the factor of 10 difference did not arise from a change in the detection limit. From the PGM wells one or more measurements of a variable were left out when they could be marked as an outlier (as a result of sampling or analysis errors, or local pollution). The presence of outliers was suspected when addition of the PGM set to the NGM set resulted in a substantial increase of the within-category variation (the long-distance value of the variogram concerned, Section 4). A suspected value was classified as an outlier when it did not belong to the interval $[\bar{z} - 3s, \bar{z} + 3s]$, where $\bar{z}$ and s are the within-category mean value and standard deviation of the log-transformed measurements in the NGM set that were not equal to the detection limit. Less than 0.2% of the measurements were marked as outliers.

After applying all previously described selections, only the measurements from monitoring screens with at least five measurements in the period considered were used for mapping temporal changes in groundwater quality.

Table 3
The 12 groundwater quality variables measured yearly in the national monitoring network over the period 1987–1993

Ca, Cl, EC, pH, HCO_3, K, Mg, Na, NH_4–N, NO_3–N, P–tot, SO_4

Very few monitoring network sites had two screens at the depth interval considered. For these sites, the mean value of the log-transformed values was used as one observation for the spatial interpolation. The values from separate screens were used individually for the variogram, thus contributing to the short-distance variation, the start of the variogram (assuming that variation of groundwater quality variables at a vertical distance of 10 m is similar to the spatial variation at a horizontal distance of a few kilometres). Time predictions and time prediction uncertainties (estimation variances) of different screens in the same well were also simply averaged on the log scale, assuming independence.

4. Spatial interpolation

Maps of groundwater quality variables were obtained by means of ordinary block-kriging (Journel and Huijbregts, 1978) of log-transformed measured concentrations within each of the soil–land use categories of Table 2, using the measurements available. For categories with only very few measurements a modified procedure was used, as described at the end of this section. A small modification of the interpolation procedure used for mapping temporal changes in groundwater quality is described in Section 8.

For the spatial interpolation of the measurements a model concept was adopted which considered log-transformed measurements of a groundwater quality variable $z(x)$ as a sample from a realization of the random field $Z(x)$ with the following properties:

$$Z(x)=m+e(x) \tag{1}$$

with

$$m \text{ constant} \tag{1a}$$

$$\mathrm{E}(e(x))=0, \quad \forall x \in \boldsymbol{R} \tag{1b}$$

and

$$\tfrac{1}{2}\mathrm{E}\left[(Z(x_1)-Z(x_2))^2\right]=\gamma(\|x_1-x_2\|), \quad \forall x_1,\, x_2 \in \boldsymbol{R}. \tag{1c}$$

That is, within an area of size $\boldsymbol{R}$ inside a soil–land use category we used an additive model (on log scale) with (i) a constant expected value and (ii) a spatial dependence between measurements that is a function only of the distance h between the measurement locations. Additionally we assumed that in a soil–land use category for a groundwater quality variable only one function $\gamma(h)$ is needed (independent of the location of the area of size $\boldsymbol{R}$).

Depending on the the sample variogram (or experimental variogram), the form of the theoretical variogram $\gamma(h)$ was chosen from a nugget model, a spherical model, or a mixture of these two models (Journel and Huijbregts, 1978). An ad hoc method for fitting the variogram model to the sample variogram was used: the short-distance variance (the nugget) and the maximum correlation distance (the range of a spherical model) were chosen, whereas the long distance variance (the sill) was fitted (using weighted least-squares fit with weights equal to the number of pairs used for each sample variogram estimate). Fig. 3 shows the variograms for potassium. The area size $\boldsymbol{R}$ used was defined (in

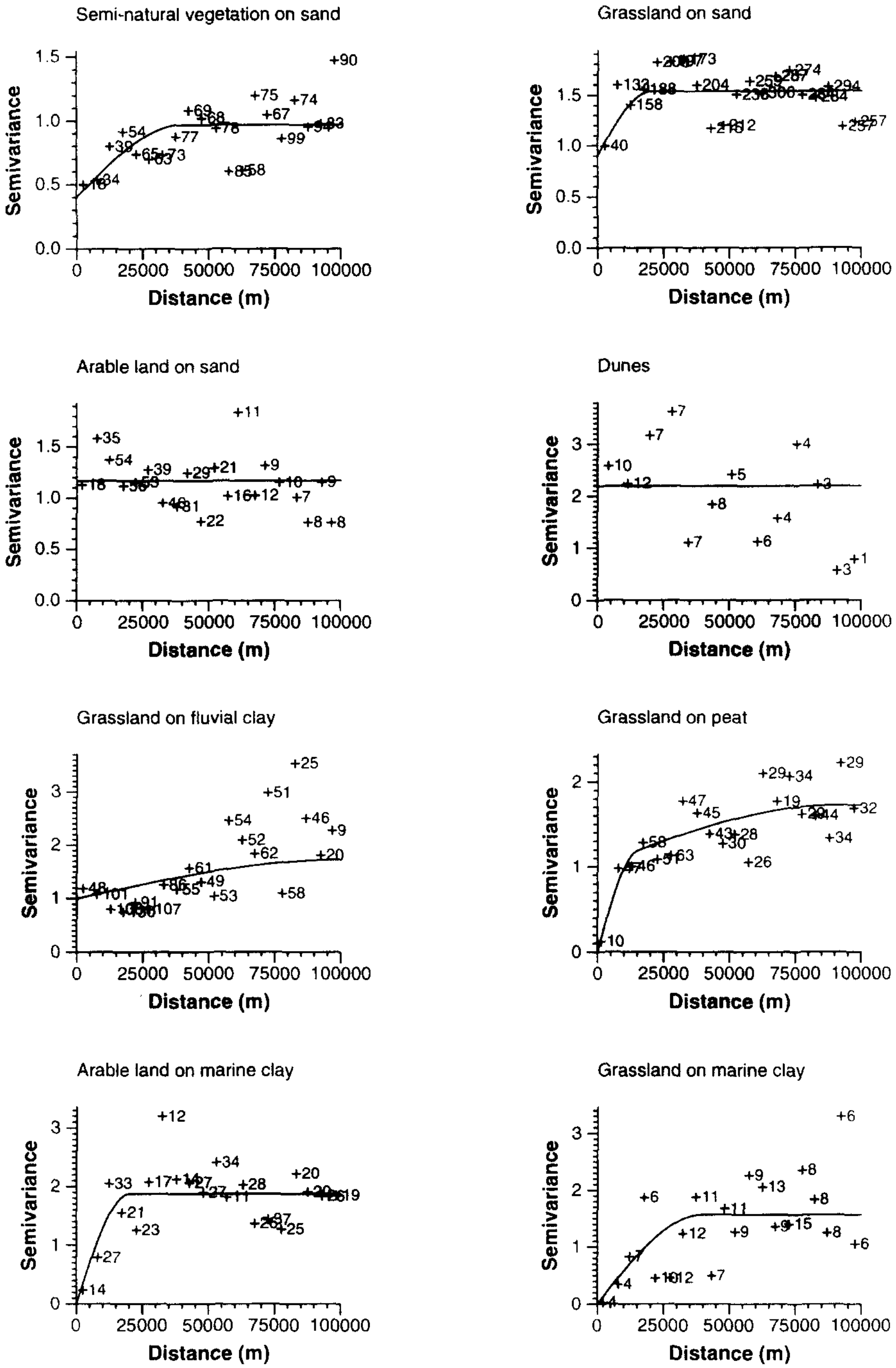

Fig. 3. Variograms of ^{e}log (K), after stratification by soil type and land use. A number reflects the number of observation pairs used for an estimate (+) of a variogram point.

order of preference) by (i) the 20 closest observations when more than 20 observations were available within a distance of 100 km, (ii) 100 km when 10–20 observations were available within this distance, (iii) the 10 nearest observations.

The estimates given in the maps concern mean values on log scale of 4 km × 4 km blocks. Estimates are presented as approximate 95% confidence intervals, calculated by

$$[\hat{z}(x_0) - 2\sigma_k(x_0),\ \hat{z}(x_0) + 2\sigma_k(x_0)] \tag{2}$$

with $\hat{z}(x_0)$ the point estimate (kriging estimate) on log scale, and $\sigma_k(x_0)$ the estimated standard deviation (kriging standard deviation) of the estimate (Journel and Huijbregts,

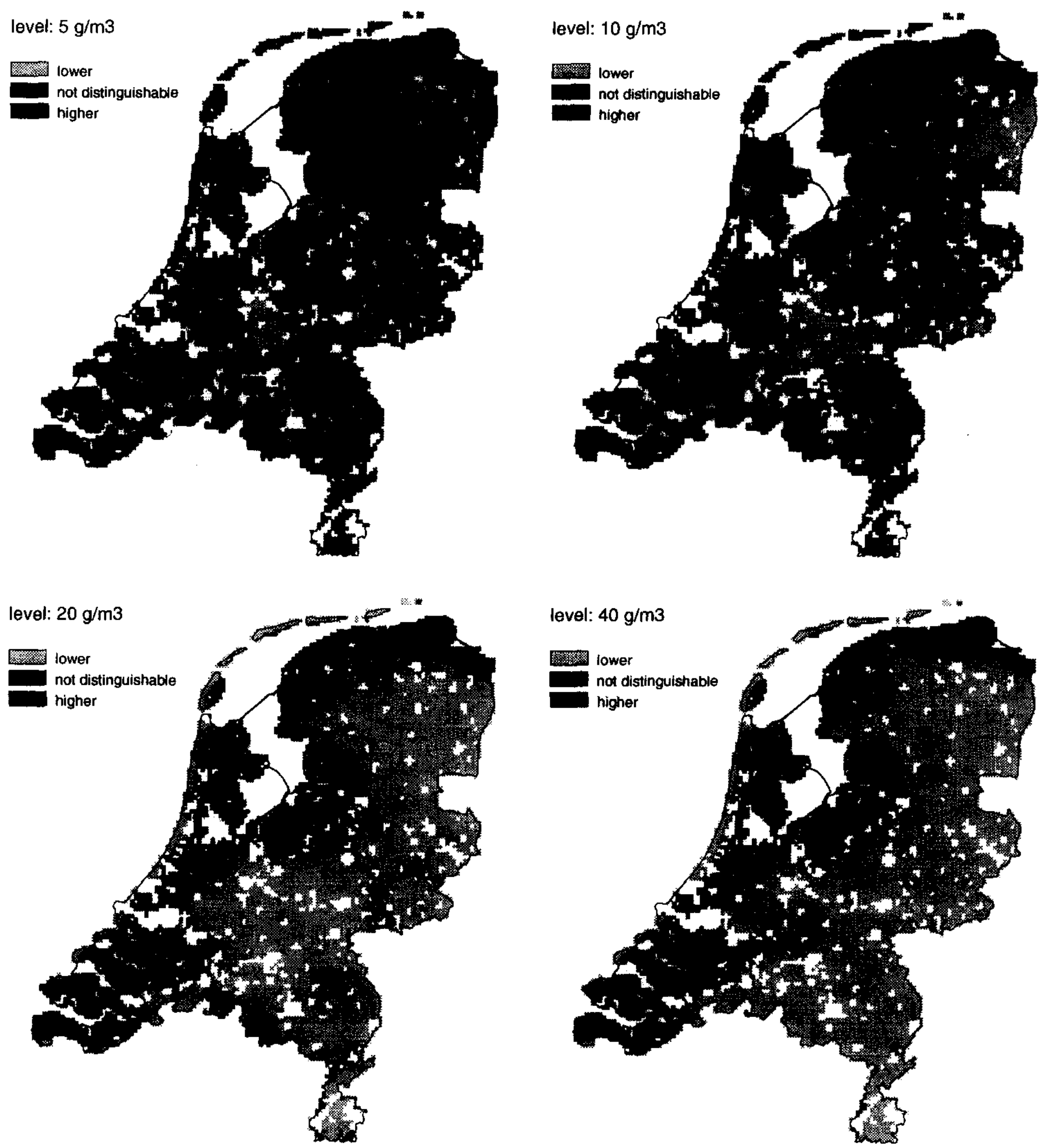

Fig. 4. Map of potassium in the groundwater. 95% confidence intervals for 4 km × 4 km block median values related to four concentration levels.

1978). Back transformation of this interval to the original scale (taking the exponent of both sides) yields an approximate 95% confidence interval for the block median value.

All 4 km × 4 km blocks considered are centred at the cells of the 2 km × 2 km grid of Fig. 1 (assuming that soil–land use dominance of 2 km × 2 km cells still holds for the 4 km × 4 km blocks). In this way two adjacent 4 km × 4 km blocks have a 50% overlap, resulting in the finer spatial resolution of 2 km × 2 km cells, while the limited estimation variance of 4 km × 4 km block mean values remains relevant. The results have been presented in maps (e.g., see Figs. 4 and 5). Only the 2 km × 2 km centres of blocks for which estimates were made are shown on the maps.

For the small soil–land use categories with less than ten observations of a variable the estimation procedure was modified, because too little information is available in these categories for modelling the variogram and for estimating block mean values in the way described above. For spatial interpolation in such a small category, for a specific variable (i) the variogram of the largest category with the same soil type was used (implying that soil type is considered to be the major structure-determining factor), and (ii) the observations of the category from which the variogram was taken were added to the set of observations of the small category, but only when the 80-percentile range of the observations in the main category covered the observations of the small category; this implied that the values of both categories should be similar: in any other case no estimates were made for the small category. The small categories cover 2.2% of the maps.

5. The groundwater quality maps

In the groundwater quality maps presented, the 95% confidence intervals for 4 km × 4 km block median values are related to four reference levels obtained from multiples of target levels or critical levels for human consumption. These target levels or critical levels, however, are defined in relation to individual samples. If a block median value exceeds a critical level, more than half of the individual samples in the block exceed this level. If, however, a block median value does not exceed a critical level, many individual samples within the block may still exceed it.

The maps are shown in two forms. In the base maps (the black-and-white maps, e.g. potassium concentration in Fig. 4) the estimates are related to the four reference levels considered in four separate sub-maps. Each of the sub-maps shows the position of the confidence intervals relative to one reference level. The interval can be completely below the level (*lower*), completely above the level (*higher*) or it can straddle the level (in which case estimated value and reference level are not distinguishable on the basis of the information available). In the coloured maps (e.g. potassium concentration in Fig. 5) four black-and-white sub-maps are merged into a single map (holding exactly the same information).

In general, the black-and-white maps are easier to read, but the coloured maps immediately show the precision of the estimates: cells with confidence intervals falling completely in one legend class have a solid colour, whereas green–red cells are cells for which no information is available at the resolution level considered. The doubling of the reference levels in the maps reflects the choice for equidistant levels on log scale.

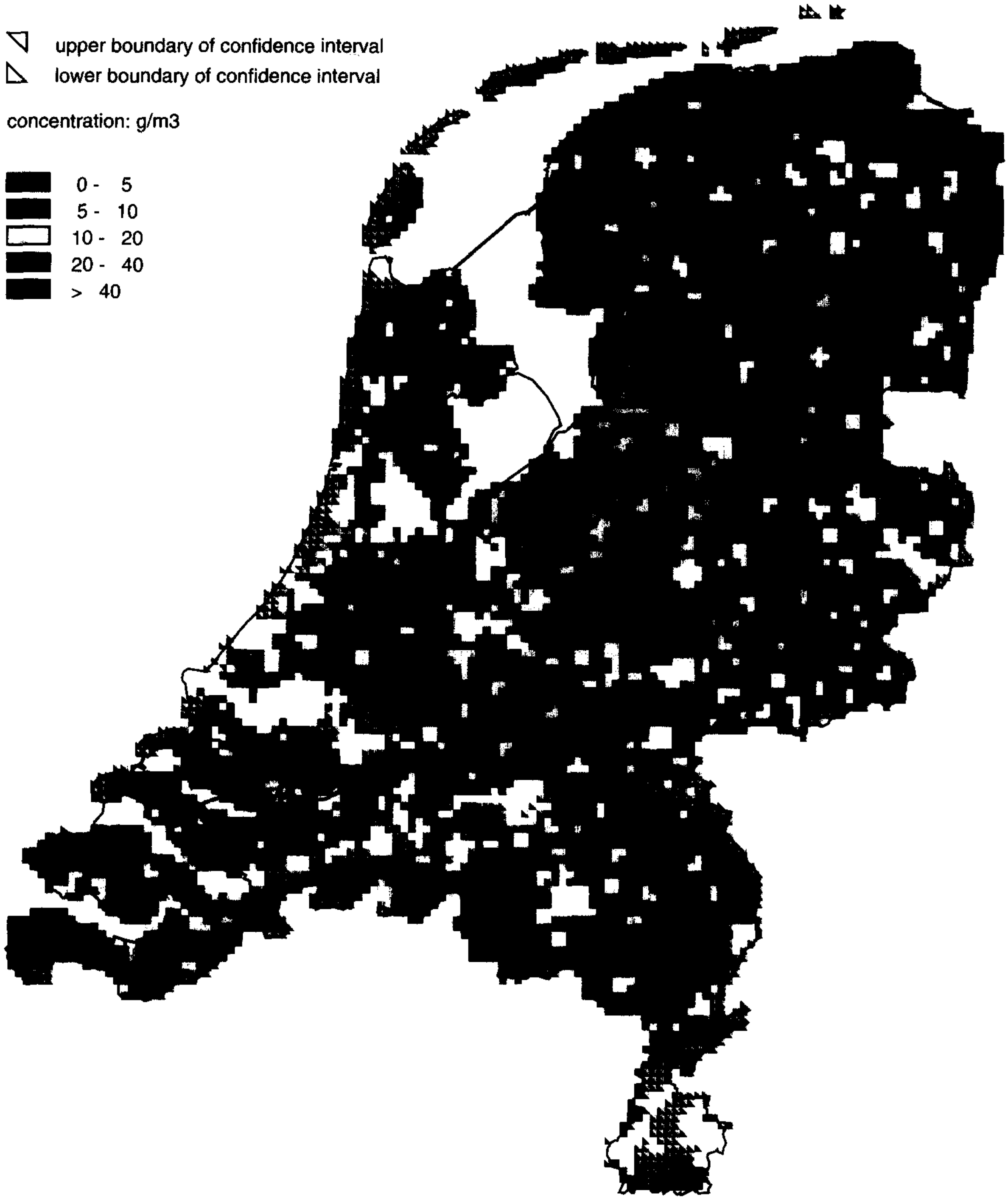

Fig. 5. Map of potassium in the groundwater. 95% confidence intervals for 4 km × 4 km block median values related to four concentration levels (alternative display of Fig. 4).

6. The effects of stratification

Groundwater quality maps can also be obtained while partially or completely ignoring the geohydrological knowledge that resulted in stratification by soil type and land use. For nine groundwater quality variables (Table 4) the effects of stratification were investigated by making groundwater quality maps with partial stratification and without stratification.

Table 4

The 9 groundwater quality variables involved in assessing the effects of stratification and network density on mapping

Al, Ca, Cd, Cu, K, NO_3–N, Pb, P–tot, Zn

These maps are based on the same data as the main maps, using the same model (1) but now (a) for one category only (containing all 13 categories of Fig. 1), (b) for land use categories only, or (c) for soil type categories only. For potassium, the results are summarized in Table 5. For one concentration level the effects of no stratification are shown in Fig. 6a, the effects of stratification by land use only—ignoring soil type—are shown in Fig. 6b.

In general, it can be said that:

(i) In this study the effects of omitting or partially applying stratification are large, even in the current situation where usually very wide confidence intervals are related to four reference levels. For almost all groundwater quality variables considered, partial stratification results in a 10–25% change in at least one of the four sub-maps (Pebesma and De Kwaadsteniet, 1994).
(ii) The simpler the approach, the less the spatial differentiation.
(iii) The large differences between resulting maps demonstrate that it is essential to incorporate geohydrological knowledge into a mapping procedure such as this (De Kwaadsteniet, 1990).
(iv) The full stratification used for the base maps represents the safest approach, compared to the simpler approaches, because all simpler models are a special case of the fully stratified model. Only when one has good reason to believe that for some groundwater quality variable stratification should have been omitted (partially), may one prefer the corresponding map.

7. The effects of monitoring network density

The effects of the density of the monitoring network on the groundwater quality maps were studied by comparing groundwater quality maps based on the limited information

Table 5

The effects of stratification for potassium. Number of cells for which the value of a 95% confidence interval for a 4 km × 4 km block median concentration related to a reference level (possible values: higher, lower and not distinguishable) changed as a result of omitting stratification by (a) land use, (b) soil type, and (c) both (total number of cells: 8197). Changed cells for the 10 g m^{-3} level of columns (b) and (c) are shown in Fig. 6

Level	Number of changed cells as a result of omitting		
(g m^{-3})	Land use	Soil type	Both
5	1426	1147	1866
10	1926	1936	2247
20	817	1052	1394
40	392	907	1170

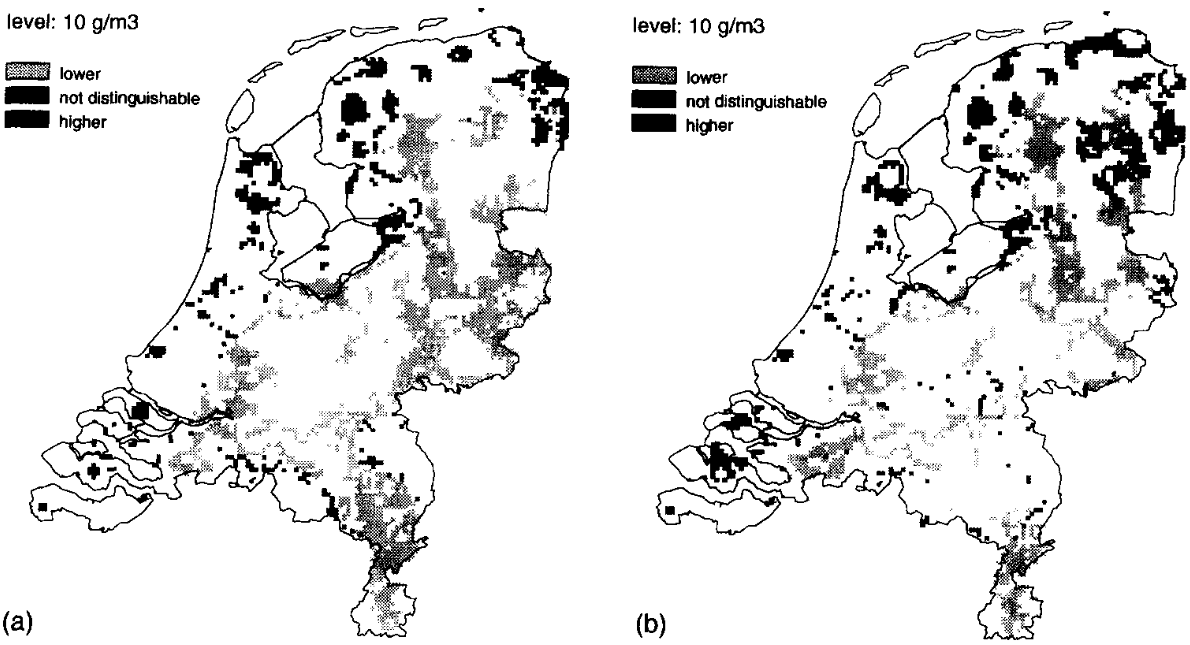

Fig. 6. The effects of stratification for potassium. 95% confidence intervals for 4 km × 4 km block median values when (a) soil and land use information is ignored or (b) soil information is ignored, related to a single concentration level. Only results that differ from those of Fig. 4 are shown.

from the observations of the NGM with the maps based on the full set of observations (NGM + PGMs). This was done for nine groundwater quality variables (Table 4). For potassium, the results are summarized in Table 6. For one concentration level, the results are presented in Fig. 7a. Changes can be assigned to (i) loss of information on the mean level per soil–land use category, and/or location-specific information (loss of first-order information), or (ii) loss of spatial dependence (variogram) information (loss of second-order information). Changes that result from either one of these two specific components are shown in Fig. 7c and Fig. 7b respectively.

Table 6

The effects of monitoring network density for potassium. Number of cells for which the value of a 95% confidence interval for a 4 km × 4 km block median concentration related to a reference level (possible values: higher, lower and not distinguishable) changed as a result of omitting the information from the provincial groundwater quality monitoring networks (PGMs) (total number of cells: 8197). Changes are a result of (a) loss of information on variograms (second-order information) and/or (b) loss of information on the mean level and location-specific information (first-order information). Numbers of changes from each of these components are also given. For the 10 g m^{-3} level the changed cells are shown in Fig. 7

Level	Number of changed cells		
(g m^{-3})	Total	As a result of (a)	As a result of (b)
5	1584	1765	494
10	3153	2728	738
20	1165	1045	461
40	750	1067	174

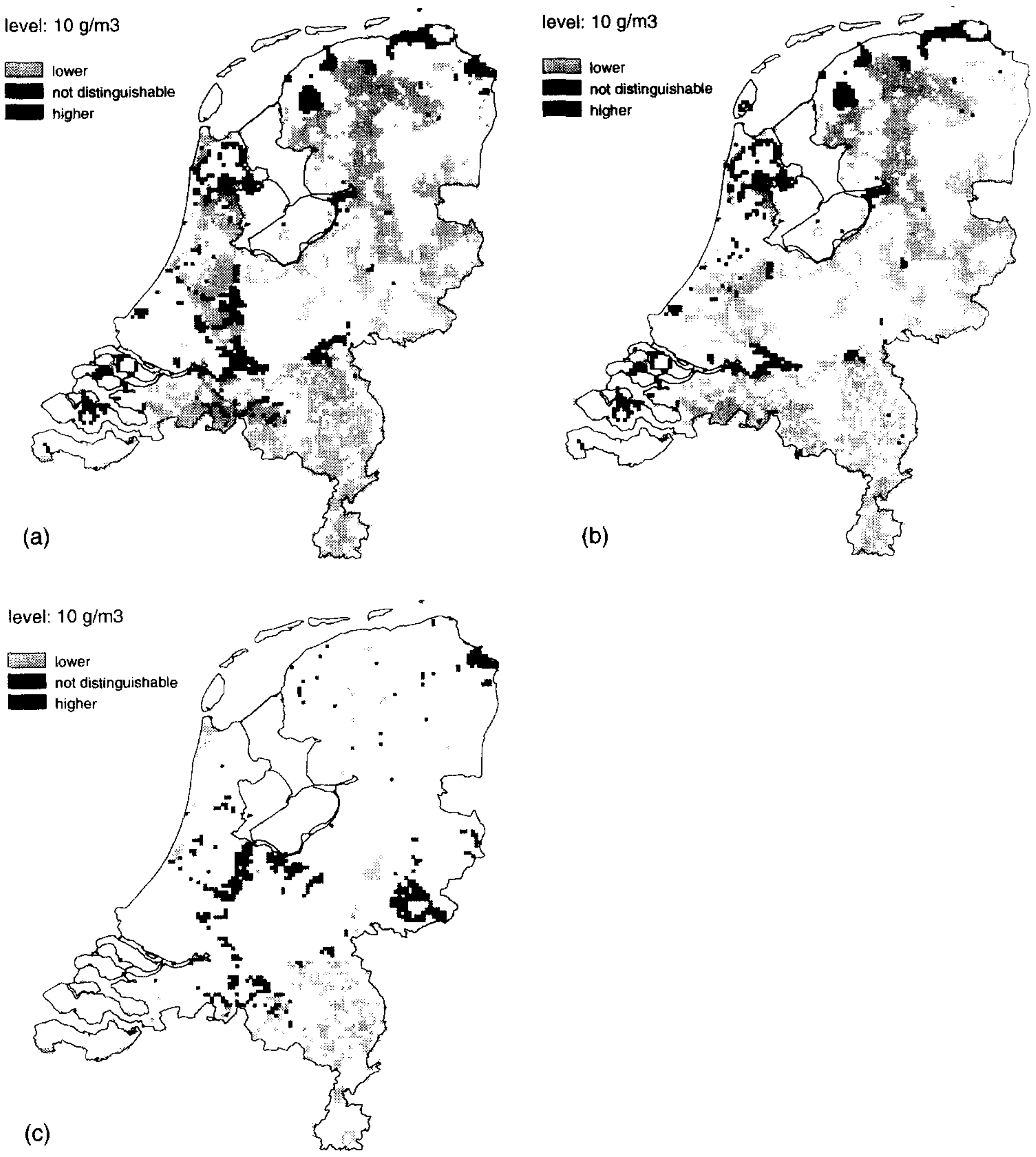

Fig. 7. The effects of network density for potassium. 95% confidence intervals for 4 km × 4 km block median values related to one concentration level. Maps are based on (a) measurements from the national groundwater quality monitoring network only (provincial monitoring network information is left out), (b) first-order

It is evident that the extension of the NGM with the PGMs has resulted in important improvements of the groundwater quality maps (Fig. 7a, 'total' column of Table 6). For a large part these improvements result from the increase of spatial dependence information (Fig. 7b and column (a) of Table 6), especially from information on short-distance variation (see also Fig. 2). However, the effects of increased first-order information also lead to important improvements of the groundwater quality maps (Fig. 7c and column (b) of Table 6).

8. Changes in groundwater quality

Since starting level and time gradient are of combined interest in the assessment of changes in groundwater quality, we chose to examine temporal changes in terms of short-term predictions. In particular, we are interested in first indications of meaningful, steady (i.e. monotonic), long-term components in the changes in time. To that end, for each separate monitoring well screen, the behaviour of a groundwater quality variable on the log scale in time $y(t)$ was represented by a first-order linear model with time, having independent, identically distributed errors

$$Y(t)=\alpha_0+\alpha_1 t+e(t),\quad \mathrm{E}(e(t))=0,\quad \mathrm{Cov}(e(t))=\sigma^2 I \tag{3}$$

which is a special case of the linear model $Y(t)=F\beta+e(t)$ when $f(t_i)=(1, t_i)$, $i = 1...n$, is the ith row of F, and $\beta = (\alpha_0, \alpha_1)'$.

Using standard regression analysis (Draper and Smith, 1981), the ordinary least-squares estimate of β is $\hat{\beta} = (F'F)^{-1}F'y$, with $y=(y(t_1), ..., y(t_n))'$ the log-transformed measurements. Given $\hat{\beta}$, the estimate for the mean value of a short-term prediction at extrapolation time t_{new} is

$$\hat{y}(t_{\mathrm{new}})=f(t_{\mathrm{new}})\hat{\beta}=\hat{\alpha}_0+\hat{\alpha}_1 t_{\mathrm{new}} \tag{4}$$

having estimation variance

$$\sigma^2(t_{\mathrm{new}})=f(t_{\mathrm{new}})(F'F)^{-1}f(t_{\mathrm{new}})'\sigma^2. \tag{5a}$$

$\sigma^2(t_{\mathrm{new}})$ is estimated by $s^2(t_{\mathrm{new}})$ when

$$s^2=y'\left(I-F(F'F)^{-1}F'\right)y/(n-2) \tag{5b}$$

is substituted for σ^2 in Eq. (5a).

Using the log transformation not only guarantees predictions that are strictly positive on the original scale; it also yields predictions and prediction inaccuracies at the monitoring network sites on the most obvious scale for spatial interpolation.

For creating maps of changes in groundwater quality, the short-term predictions from Eq. (4) were taken as 'new observations' at the monitoring network sites. Location-specific 'observation inaccuracy' was derived from Eq. (5a) and Eq. (5b). More specifically, for two moments at equal times from the centre of the observation time series 1987–1993 (at t = 1980.5 and t = 2000.5), a set of 'new observations' was computed. For each pair of 'new observations' at a specific monitoring site, this equitemporal choice of extrapolation times results—on the log scale, and by absence of a trend on the original scale as well—in equal 'observation inaccuracies' (Fig. 8). In choosing a twenty-year extrapolation span we aimed at visualizing trends despite observation inaccuracies.

Quantifying the observation inaccuracy with Eq. (5a) implies that the 'new observations' relate only to the situation *around* 1980 and *around* 2000: the short-term components of temporal variation, the year-to-year deviations from the regression line, are filtered out, within the context of the current model structure.

The maps were based on a concept where the variable we want to map—the spatial variable $z(x)$ at an extrapolation time—is not observed at an observation location x_i, but where only approximate observations $\hat{y}(x_i)$ are available, which are subject to an unknown

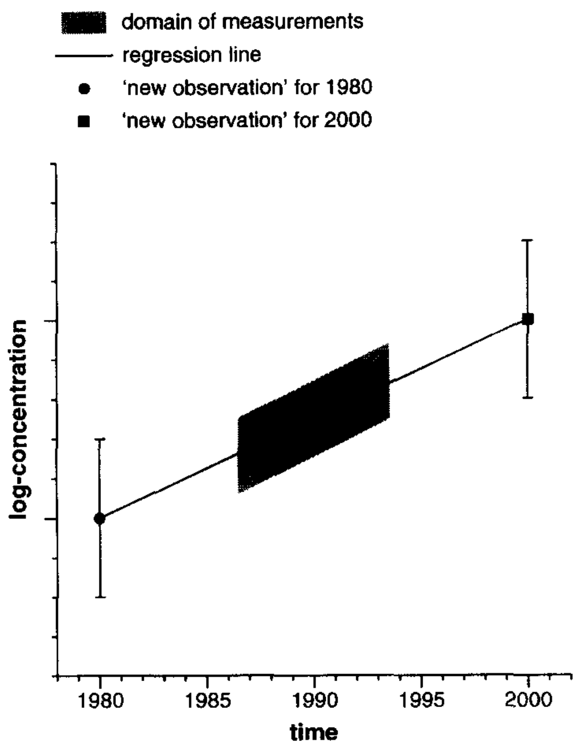

Fig. 8. Schematic representation of 'new observations' and their inaccuracy (as error bars) at a single monitoring well screen (no actual measurements drawn).

observation error. We can model such approximate observations as the sum of a stationary variable $Z(x_i)$ (as in Eq. (1)) that represents the variable of interest $z(x_i)$, and a measurement error $\epsilon(x_i)$:

$$\hat{Y}(x_i)=Z(x_i)+\epsilon(x_i). \tag{6a}$$

For the spatial dependence of the variable $Z(x_i)$ we used the variograms obtained for 1991 (e.g. Fig. 3 for potassium). For the observation error $\epsilon(x_i)$ we assumed that it has zero mean, that it is independent from other $\epsilon(x_j), j \neq i$, and $Z(x)$,

$$\mathrm{E}(\epsilon(x_i))=0,\ \ \mathrm{Cov}\big(\epsilon(x_i),\ \epsilon(x_j)\big)=0,\ \ \mathrm{Cov}(\epsilon(x_i),\ Z(x))=0,\ \ \forall i,j \neq i,x \tag{6b}$$

and that it has a location-specific, known variance $\sigma_\epsilon^2(x_i)$ as was estimated from Eq. (5a) and Eq. (5b).

This implies that beyond the standard kriging assumptions of Eq. (1) the maps of changes in groundwater quality were based on the assumption that at a monitoring site we do not know the exact value of $z(x_i)$, but that (for convenience's sake, in the case of a normally distributed observation error on the log scale), independently of everything else, 95% confidence intervals for $z(x_i)$ are given by

$$[\hat{y}(x_i)-2\sigma_\epsilon(x_i),\ \hat{y}(x_i)+2\sigma_\epsilon(x_i)],\ \ \forall i.$$

The adaptation of the standard kriging procedure to handle observations that are subject to an observation error that satisfies Eq. (6a) and Eq. (6b) was given by Delhomme (1978) and Pebesma (1996). In comparison with the situation where the observation error would have been ignored, applying this adaptation represents a more careful approach that results in wider confidence intervals, and an approach in which less weight is assigned to observations with relatively large inaccuracies.

For each variable, changes in groundwater quality were revealed by showing maps for 1980 and 2000 side by side, and by only showing their mutual differences—where the maps were identical, no values were shown. Fig. 9 shows the maps of indications of changes in potassium concentration for two concentration levels.

The maps reflect the development of groundwater quality over a 20-year span, within the context of the current model structure. The differences shown arise directly from the

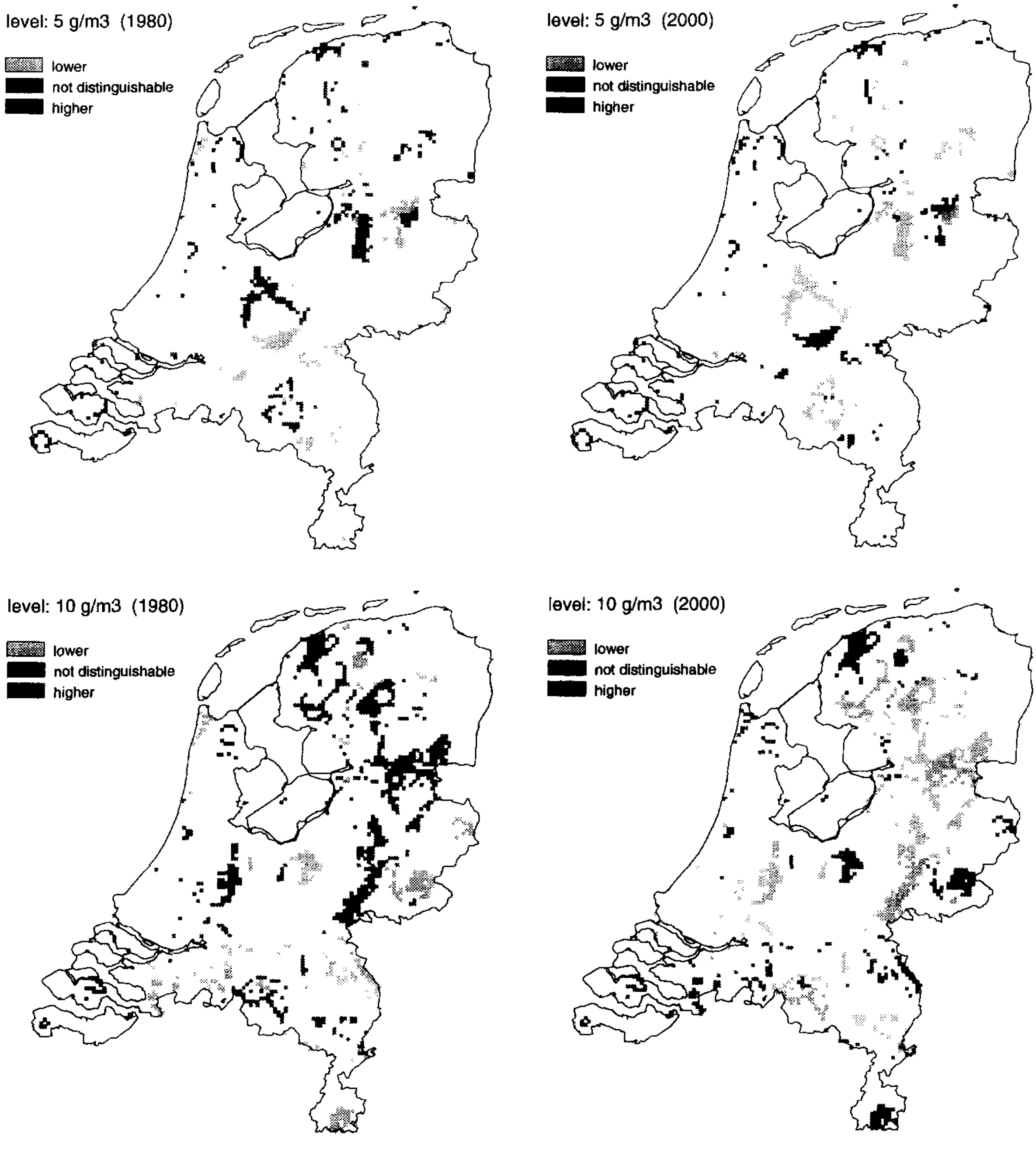

Fig. 9. Maps of potassium in the groundwater around 1980 (left) and around 2000 (right). 95% confidence intervals for 4 km × 4 km block median values related to two concentration levels. Only cells for which the 1980 and 2000 maps differ are shown.

slope of all or part of the regression lines, as fitted to the measured time series: if the slopes were zero the maps would have been identical.

The indicated trend direction in a cell can be derived from the direction of shift in legend class from one extrapolation time to the other. In general, for instance in the case of a single legend class shift, the trend magnitude cannot be derived from this shift. It can be stated, however, that all trend indications shown are of vital importance in the sense that they concern a shift in estimates of 4 km × 4 km block median concentration that results in a different map for at least one of the reference levels.

The maps (Fig. 9; Pebesma and De Kwaadsteniet, 1995) show a wealth of trend indications. Primarily we can conclude that a trend approach with such a high spatial differentiation is attainable and advisable—advisable because (i) in that way cells with a pronounced trend can be distinguished from a neighbourhood with less pronounced trends; (ii) trends in adjacent or nearby cells (even within the same soil–land use category) can have opposite directions.

9. Discussion and conclusions

9.1. Groundwater quality in 1991

In this study, estimates of 4 km × 4 km block median values were presented as approximate 95% confidence intervals: a compact way of combining the presentation of expected values (point estimates) and uncertainty (estimation variances) that prevents misinterpretation of point estimates. If there is an area where confidence intervals are wide and block median values cannot be distinguished from an important critical level, one could decide to place an additional monitoring site in or close to that area. On the other hand, where such a distinction can easily be made, one may decide that fewer measurements would be sufficient. For considerations of this kind, the critical levels should apply to block median values and not to individual measurements. In this way, optimization of a monitoring network can be directed efficiently and location-specifically to relevant aspects of groundwater quality maps. In general, the maps give little reason for reducing the monitoring network density (they have wide confidence intervals).

From a policy point of view, the maps can be used directly either for quantifying diffuse groundwater contamination (because the measurements used do not represent strictly local contaminations) or for obtaining location-specific background concentrations, e.g. for assisting local contamination assessment. More indirectly, the maps can be used as an input for, and for validation of policy supporting, regional or national groundwater quality models. The maps can be considered as a translation of the information available on the measurement scale into information on a scale used in regional groundwater models.

The spatial resolution of the groundwater quality maps (half-overlapping 4 km × 4 km blocks) can be seen as a maximum, given the available monitoring network density and the observed variation in groundwater quality. For purposes where the properties of larger spatial units are sufficient, using the current approach but with simply larger blocks would be unsatisfactory in the Netherlands: because of the spatial variation of soil type and land use in the Netherlands, the meaning of block-dominant soil–land use as a basis for stratification would lose its relevance. When estimates of groundwater quality variables are needed at a lower spatial resolution, it may be better to combine the available measurements within a soil–land use category, treat them as a (random) sample, and use classical statistics (design-based inference). This may also permit the estimation of higher quantiles (the tail of the distribution of values occurring in a soil–land use category) or the estimation of the fraction of individual values exceeding a critical level. Provided that the latter approach also includes temporal aspects of groundwater quality, it can be considered as a second, alternative approach to monitoring network evaluation (and optimization).

Stratification by land use and soil type preceding spatial interpolation turned out to be essential: the necessity of using process knowledge because of the non-uniqueness of a strictly geostatistical approach (De Kwaadsteniet, 1990) is not just a theoretical point, but results in maps that differ substantially from maps obtained when no or partial stratification is applied. Differences can be viewed as the translation of the way in which, with respect to the groundwater quality variables, strata differ in level, in order of magnitude of variation, and in spatial dependence between measurements. For the last two points, see also Fig. 3.

More generally, it can be said that the current approach closely follows a classic statistical approach: we start with a stratification by soil type and land use. It improves on a purely classic statistical approach, however, by its combined use of location-specific information and spatial dependence through stratified kriging. The effects of monitoring network density clearly demonstrate the added value of using location-specific information and spatial dependence: the effects of short-distance variation, especially, but also the effects of local information on the groundwater quality maps are substantial.

The stratification used for this study is not ideal. Soil type information was obtained from a national soil survey (Steur et al., 1985) that describes the top metre of the soil, whereas the soil characteristics from soil surface to the depth of the monitoring screens would have been more relevant. Land use information (Thunnissen et al., 1992) describes the (estimated) situation in 1986–87, whereas land use t years ago (where t depends on local groundwater flow) would yield more adequate information. The suboptimal information availability results in larger within-category variation (wider confidence intervals) than the variation that would be obtained with the ideal soil and land use information. Also, errors occurring in both maps, clustering of categories and using dominant soil–land use result in an increase of within-category variation. Finally, an important part of the within-category variation may stem from the absence of maps of relevant geohydrological attributes. When such maps (e.g. on infiltration versus seepage) became available, it might be worthwhile to include them in the stratification. This would result in fewer measurements per category, but with smaller within-category variation more accurate estimates might nevertheless be obtained.

Although geostatistical models are often used for the estimation (mapping) of groundwater flow model parameters and groundwater head (e.g. Hoeksema and Kitanidis, 1984), few reports of the direct application of geostatistical methods to the mapping of groundwater quality variables have been found. This may be so because of the large variations often found in the sampling material (e.g. Bjerg and Christensen, 1992). This, however, does not seem a structural objection to the current study, and presenting results as 95%

confidence intervals prevents, furthermore, the ignorance of the estimation uncertainty of the results. Rubin (1991) claims that "kriging techniques ... have rarely been applied in the context of solute concentrations" because "neither the concentration expected value nor its spatial moments are readily available. It is particularly difficult to infer the covariances of the concentration since the concentration field is time-dependent and nonstationary. In order to empirically derive the covariance of a covariance field a huge data base is needed which translates practically to a situation where an estimation problem is nonexistent." The present study can be seen as a counter-example to this claim.

9.2. Temporal changes in groundwater quality

Initially, the intuitive and, except for large factorial trends, most obvious simple approach for modelling steady trends would be to use a first-order linear model with time on the natural (i.e. measurement) scale. However, when we consider 10-year extrapolations, maps from this approach would differ significantly from maps using first-order linear trends on the log scale (this study) only when factorial trends of at least a factor of 2 occur and prediction inaccuracies are ignored. For large factorial trends the log scale is already preferred on intuitive grounds. Furthermore, the log procedure is found to be more robust (Hirsch et al., 1991). For less pronounced factorial trends the differences between both approaches fade away when prediction inaccuracies are taken into consideration. The quantitative incorporation of location-specific prediction inaccuracies is an essential aspect of the current study. As the working scale, the log scale permits this.

In this paper, studying changes in groundwater quality is primarily aimed at obtaining first statistical indications of steady time trends in block median values (steady trends in a first-order statistic). From a methodological point of view, it is then a correct first approach to work with time-invariant second-order statistics (spatial variation, spatial dependence in variation, variogram) within soil–land use categories. We decided to do so on the basis of the following arguments:

(i) Statistical distributional properties of variograms are complex, and a statistical comparison of variograms is a subject the complexity of which is beyond the context of this study.

(ii) Using a different variogram for each extrapolation time would comprise a modelling detail that was not statistically shown to be necessary (and probably cannot be shown to be necessary, given the number of observations per soil–land use category (Table 2)). In addition, it would represent a modelling approach in which maps no longer exclusively reflect the primary goal: differences in maps would in that case reflect steady time trends as well as the newly introduced time differences in second-order statistics.

For the variograms of $Z(x)$, in Section 8 we used the variograms fitted to the 1991 measurements (Fig. 3). The primary reason for this is that, for studying spatial variation of a groundwater quality variable within a soil–land use category, for 1991 we had a much larger set of observation points than was available on an extrapolation time: for the 1991 study we could use about 170 observation points of provincial monitoring networks in addition to the national groundwater quality monitoring network. Distinguishing the 1991 situation from the 'around 1991 situation' with respect to the variogram specification was—with all other simplifications in mind—considered to be overdone.

Taking the estimated 'observation inaccuracy' as a deterministic quantity when, as in this study, a 'new observation' (a short-term prediction) in a monitoring well screen is based on a time series of 5–7 measurements, implies that in this first approach for the determination of the 'observation inaccuracy' the estimated variance $s^2(t_{new})$, based on three degrees of freedom when five observations are available, was not distinguished from the true variance of Eq. (5a). This results in a serious underestimation of 'observation inaccuracy', and underlines the caution that should be taken with respect to the interpretation of the maps. The reason why we used 'new observations' based on time series of five measurements, despite the shortcomings of the methods used, is the observation from practice that the influence of such series on the maps cannot be ignored.

More generally, it can be stated that in trend studies the inaccuracies of trend observations deserve explicit, quantified attention. In order to be able to present well-founded estimates of changes in groundwater quality over a reasonable time span, the—now yearly—measurement frequency should not be reduced.

It should be noted that the maps do not pretend to show the actual changes that occur over the 20-year time span considered. The changes shown only involve extrapolations from observed changes in the measurement series 1987–1993, within the context of the current model structure.

Serious modifications of the maps in this study may be expected when, after a few years, measurement series of the provincial groundwater quality monitoring networks become available and are incorporated in the mapping.

9.3. Mathematical aspects

Mathematical aspects were kept simple, primarily because we aimed at (a) an optimal accessibility of the various methodological items mentioned, and (b) the avoidance of more complex assumptions than necessary, given the objective of this study and the information available. The simplicity refers to, e.g., models (1), (3) and (6), the choice and fit of the variogram models, the choice of $\boldsymbol{R}$, the way approximate confidence intervals for block median values were calculated, and the outlier procedure. Some general remarks on the mathematical aspects are due:

(i) Kriging, as applied in this study (using model (1) for each variable for each soil–land use category), can be considered as a first step beyond and an improvement of the classic statistical approach (for each variable for each category) in its four basic forms using spatial dependence or independence, and one mean per category or local means of regions within a category, respectively. The kriging approach used improves on these classic statistical approaches by completing the specific realization of the spatial variable considered, instead of paying attention to the ensemble of possible realizations (or a new realization) under the classic model assumptions.

(ii) The (approximate) 95% confidence intervals presented in this study are consistent with interpolations and short-term predictions on a log scale: an obvious choice of scale given the character of the measurements at hand. For a first quantification of uncertainties (the calculation of confidence intervals using Eq. (2)) the estimation error of block mean values on the log scale is assumed to be normally distributed, and uncertainty of the (kriging) standard deviation (resulting from the estimation of variograms and 'observation inaccuracies') is ignored. This implies that only a rough indication of reliability is obtained, and that the widths of the confidence intervals should be considered as lower bounds for more carefully obtained confidence intervals.

(iii) For the interpretation of the intervals of Eq. (2) as 95% confidence intervals for block median concentrations, it is necessary that on the log scale the block mean and block median coincide. This is true, for instance, when, for each block on the log scale, individual samples within the block are distributed symmetrically. When this assumption is too restrictive, an alternative and, with respect to within-block distribution, non-parametric interpretation of Eq. (2) is that of an approximate 95% confidence interval for the block geometric mean.

(iv) In the current context (block kriging on a log scale) the central measure that is estimated is the block median, not the block mean. When the block mean is needed (e.g. for comparing the maps with groundwater quality model results), first indications of block means can be obtained from the maps of block medians by using an approximate relation between block mean $m_\Box$ and block median $M_\Box$:

$$m_\Box = M_\Box \exp(s_\Box^2/2) \tag{7}$$

where $s_\Box^2$ is the short-distance variance, the short-distance value of the variogram for the relevant variable and category. Eq. (7) is derived from the well-known relation between mean and median of a lognormal distribution (Aitchison and Brown, 1957). First indications of 95% confidence intervals for 4 km × 4 km block mean values are obtained when the legend class boundaries of a map are multiplied by the factor of $M_\Box$ from Eq. (7). For a groundwater quality variable, this will result in different legend class boundaries for different soil–land use categories.

References

Aitchison, J., Brown, J.A.C., 1957. The Lognormal Distribution. Cambridge University Press, Cambridge.

Bjerg, P.L., Christensen, T.H., 1992. Spatial and temporal small-scale variation in groundwater quality of a shallow sandy aquifer. J. Hydrol., 131: 133–149.

De Kwaadsteniet, J.W., 1990. On some fundamental weak spots of kriging technique and their consequences. J. Hydrol., 114: 277–284.

Delhomme, J.P., 1978. Kriging in the hydrosciences. Adv. Water Resour., 1: 251–266.

Draper, N., Smith, H., 1981. Applied Regression Analysis, 2nd edition. Wiley, New York.

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

Hoeksema, R.J., Kitanidis, P.K., 1984. An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. Water Resour. Res., 20: 1003–1020.

Journel, A.G., Huijbregts, Ch.J., 1978. Mining Geostatistics. Academic Press, London.

Klijn, F., 1988. Milieubeheergebieden. Deel A: Indeling van Nederland in ecoregio's en ecodistricten (Environment management areas. Part A: subdivision of the Netherlands into ecological regions and ecological districts). Centrum voor Milieukunde, Rijksuniversitelt Leiden, Leiden. Mededelingen No. 37 (in Dutch).

Pebesma, E.J., 1996. Mapping groundwater quality in the Netherlands. PhD thesis, Utrecht University, Utrecht. Netherlands Geographical Studies No. 199.

Pebesma, E.J., De Kwaadsteniet, J.W., 1994. Een landsdekkend beeld van de Nederlandse grondwaterkwaliteit op 5 tot 17 meter diepte in 1991 (Maps of groundwater quality in the Netherlands at 5-17 m depth in 1991). National Institute of Public Health and the Environment, Bilthoven. Report No. 714810014 (in Dutch).

Pebesma, E.J., De Kwaadsteniet, J.W., 1995. Een landsdekkend beeld van veranderingen in de Nederlandse grondwaterkwaliteit op 5 tot 17 meter diepte (Maps of temporal changes in groundwater quality in the Netherlands at 5-17 m depth). National Institute of Public Health and the Environment, Bilthoven. Report No. 714810015 (in Dutch).

Rubin, Y., 1991. The spatial and temporal moments of tracer concentration in disordered porous media. Water Resour. Res., 27: 2845-2854.

Steur, G.G.L., De Vries, F., Van Wallenburg, C., 1985. Bodemkaart van Nederland 1:250000 (Soil map of the Netherlands 1:250000). Stichting voor Bodemkartering, Wageningen (in Dutch).

Thunnissen, H., Olthof, R., Getz, P., Vels, L., 1992. Grondgebruiksdatabank van Nederland vervaardigd met behulp van Landsat Thematic Mapper opnamen (Land use database of the Netherlands constructed with the aid of Landsat Thematic Mapper images). DLO-Staring Centrum, Wageningen. Report No. 168 (in Dutch).

Van Duijvenbooden, W., Taat, J., Gast, L.F.L., 1985. Landelijk meetnet grondwaterkwaliteit: eindrapport van de inrichtingsfase (The national groundwater quality monitoring network: final report of the installation phase). National Institute of Public Health and the Environment, Bilthoven. Report No. 840382001 (in Dutch).