

MAPPING GROUNDWATER QUALITY IN THE NETHERLANDS

CIP-GEGEVENS, KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Pebesma, Edzer Jan

Mapping Groundwater Quality in the Netherlands / Edzer Jan

Pebesma. - Utrecht: Faculteit Ruimtelijke Wetenschappen

Universiteit Utrecht. - Ill., tab.

Proefschrift Universiteit Utrecht. - Met lit. opg. - Met
samenvatting in het Nederlands.

ISBN 90-6266-127-0

Trefw.: grondwaterkwaliteit ; kartering ; Nederland

ISBN 90-6266-127-0 (Thesis)

ISBN 90-6809-216-2 (NGS)

Copyright © Faculteit Ruimtelijke Wetenschappen Universiteit Utrecht 1996

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van de uitgevers.

All rights reserved. No part of this publication may be reproduced in any form, by print or photoprint, microfilm or any other means, without written permission by the publishers.

Omslagontwerp: Mark Oonk

MAPPING GROUNDWATER QUALITY IN THE NETHERLANDS

HET KARTEREN VAN DE GRONDWATERKWALITEIT IN NEDERLAND
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE
UNIVERSITEIT UTRECHT OP GEZAG VAN DE RECTOR MAGNIFICUS,
PROF.DR. J.A. VAN GINKEL, INGEVOLGE HET BESLUIT VAN HET
COLLEGE VAN DECANEN IN HET OPENBAAR TE VERDEDIGEN OP
1 MAART 1996 DES MIDDAGS TE 14:30 UUR

DOOR

EDZER JAN PEBESMA

GEBOREN OP 3 JULI 1967, TE BEETSTERZWAAG

promotor: Prof.dr. P.A. Burrough Faculteit Ruimtelijke Wetenschappen
Universiteit Utrecht
co-promotor: Dr.ir. J.W. de Kwaadsteniet Rijksinstituut voor Volksgezondheid en Milieu

Dit proefschrift werd mogelijk gemaakt met financiële steun van het Rijksinstituut voor Volksgezondheid en Milieu (RIVM)

foar Anne en Andrys

Contents

Chapter 1 General introduction	1
1.1 Problem definition	1
1.2 Groundwater quality—what is it?	2
1.3 Monitoring groundwater quality	2
1.4 Modelling groundwater quality	4
1.5 Groundwater quality monitoring network optimization	4
1.6 Objectives and approach	5
1.7 Outline of this thesis	5
Chapter 2 Statistical Mapping	7
2.1 Introduction	7
2.2 Uncertainty, accuracy and statistics	9
2.3 Linear models	11
2.4 Change of support and error structure	15
2.5 Geostatistics	16
2.6 Confidence interval maps	19
2.7 Mapping groundwater quality	20
Chapter 3 Maps of the groundwater quality in the Netherlands at 5-17 metre depth in 1991	22
3.1 Introduction	22
3.2 An aggregated map comprising soil type and land use	23
3.3 Data selection	23
3.4 Spatial interpolation	25
3.5 The groundwater quality maps	29
3.6 The effects of stratification	32
3.7 The effects of monitoring network density	32
3.8 Discussion	41
Chapter 4 Maps of temporal changes in the groundwater quality in the Netherlands at 5-17 metre depth	44
4.1 Introduction	44
4.2 An aggregated map comprising soil type and land use	45

4.3	Data selection	45
4.4	Changes in groundwater quality variables at the well screens	46
4.5	Spatial interpolation	48
4.6	Results	51
4.7	Discussion	54
Chapter 5	Improving estimates with ancillary information	57
5.1	Introduction	57
5.2	Ancillary information	58
5.3	Spatial interpolation with ancillary information	59
5.4	Example A. Nitrate leaching from agricultural soils	61
5.5	Example B. Atmospheric deposition of zinc	66
5.6	Example C. Measurements of sulphate in shallow groundwater	71
5.7	The effects of ancillary information	74
5.8	Discussion	74
Chapter 6	Discussion and conclusions	80
6.1	Introduction	80
6.2	Optimizing the groundwater quality monitoring networks	80
6.3	General aspects of the discussions	81
6.4	Conclusions	83
6.5	Future directions	83
Summary		87
Samenvatting in het Nederlands		91
Appendix A	Tables	96
References		99

List of Figures

2.1	<i>Nitrate-N concentration measured in the groundwater monitoring networks at 5-17 m depth in 1991. Bullets are centred at measurement locations. Bullet size is proportional to the measured concentration</i>	8
3.1	<i>Soil type and land use map. Dominant soil-land use categories for 2 km × 2 km cells</i>	26
3.2	<i>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</i>	27
3.3	<i>Variograms of ${}^e\log(Al)$, stratified by soil type and land use. A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97</i>	28
3.4	<i>Map of aluminium in the groundwater. 95% Confidence intervals for 4 km × 4 km block median values related to four concentration levels</i>	30
3.5	<i>Map of aluminium in the groundwater (alternative display of Fig. 3.4). 95% Confidence intervals for 4 km × 4 km block median values related to four concentration levels</i>	31
3.6	<i>The effects of no stratification for aluminium. 95% Confidence intervals for 4 km × 4 km block median values when soil and land use information is ignored, related to four concentration levels. Only results that differ from Fig. 3.4 are shown</i>	33
3.7	<i>Variogram of ${}^e\log(Al)$, when stratification is omitted. Codes used for the variogram model (—) are explained in Table A3, page 97</i>	34
3.8	<i>Variograms of ${}^e\log(Al)$, stratified by land use only. A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97</i>	34
3.9	<i>The effects of stratification by land use only for aluminium. 95% Confidence intervals for 4 km × 4 km block median values when soil information is ignored, related to four concentration levels. Only results that differ from Fig. 3.4 are shown</i>	35
3.10	<i>Variograms of ${}^e\log(Al)$, stratified by soil type and land use, using only information from the national groundwater quality monitoring network. A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97</i>	37

3.11	<i>The effects of network density for aluminium (total). 95% Confidence intervals for 4 km × 4 km block median values based on measurements from the national groundwater quality monitoring network only (provincial monitoring network information is left out), related to four concentration levels. Only results that differ from Fig. 3.4 are shown</i>	38
3.12	<i>The effects of network density for aluminium (second order information). 95% Confidence intervals for 4 km × 4 km block median values based on (i) first order information from both monitoring networks, and (ii) second order information from the national groundwater quality monitoring network only, related to four concentration levels. Only results that differ from Fig. 3.4 are shown</i>	39
3.13	<i>The effects of network density (first order information). 95% Confidence intervals for 4 km × 4 km block median values based on (i) first order information from the national groundwater quality monitoring network only, and (ii) second order information from both networks, related to four concentration levels. Only results that differ from Fig. 3.4 are shown</i>	40
4.1	<i>Monitoring network locations. Sites of the national groundwater quality monitoring network that contributed to the maps of changes in groundwater quality</i>	47
4.2	<i>Schematic representation of ‘new observations’ and their accuracy (as error bars) at a single monitoring well screen (no actual measurements drawn)</i>	49
4.3	<i>Variograms of $^e\log(K)$, stratified by soil type and land use, obtained from the 1991 measurements (national and provincial monitoring networks). A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97</i>	50
4.4	<i>Map of potassium in the groundwater around 1980, where it differs from Fig. 4.5 (‘around 2000’). 95% Confidence interval for 4 km × 4 km block median values, related to four concentration levels</i>	52
4.5	<i>Map of potassium in the groundwater around 2000, where it differs from Fig. 4.4 (‘around 1980’). 95% Confidence interval for 4 km × 4 km block median values, related to four concentration levels</i>	53
5.1	<i>Map of nitrogen leaching rate from agricultural soils to shallow groundwater. Values are averaged from the original 500 m × 500 m model calculations to yield 2 km × 2 km mean values. Only cells in the category grassland on sand are shown</i>	62
5.2	<i>NO₃-N concentration in groundwater and nitrogen leaching rate. Plot of NO₃-N concentration measured in the groundwater versus nitrogen leaching rate from soils to shallow groundwater, for the category grassland on sand</i>	62
5.3	<i>Variograms of (a) $^e\log(\text{NO}_3\text{-N})$, and (b) residuals for the category grassland on sand. Calculation of the sample variograms is explained in Tables A2 (page 96) and A6 (page 97). Numbers reflect the number of (a) observation pairs or (b) residual pairs used for an estimate (+) of a variogram point. Codes used for a variogram model (—) are explained in Table A3, (page 97)</i>	63

- 5.4 *Map of NO₃-N in the groundwater (category grassland on sand). 95% Confidence intervals for 4 km × 4 km block median values, related to four concentration levels. Estimates are obtained by ordinary kriging (on log-scale), using information from the monitoring networks only* 64
- 5.5 *The effects of using information on nitrogen leaching rates. 95% Confidence intervals for 4 km × 4 km block median NO₃-N concentrations, related to four levels. Estimates are obtained by universal kriging (on log-scale), using both information from the monitoring networks and ancillary information (nitrate leaching rates). Only results that differ from Fig. 5.4 are shown* 65
- 5.6 *Map of atmospheric deposition of zinc. Values are interpolated from the original 10 km × 10 km model calculations to 2 km × 2 km cells. Only cells in the category semi-natural vegetation on sand are shown* 67
- 5.7 *Zinc concentration in groundwater and atmospheric deposition of zinc. Plot of zinc concentration measured in the groundwater versus atmospheric deposition of zinc, for the category semi-natural vegetation on sand. A bullet (•) marks a measurement at a seepage location* 67
- 5.8 *Variograms of (a) ^elog(Zn), and (b) residuals for the category semi-natural vegetation on sand, non-seepage locations. Numbers reflect the number of (a) observation pairs or (b) residual pairs, used for an estimate (+) of a variogram point. Codes used for a variogram model (—) are explained in Table A3, page 97. Calculation of the sample variograms is explained in Tables A2 (page 96) and A6 (page 97)* 68
- 5.9 *Map of zinc in the groundwater (category semi-natural vegetation on sand). 95% Confidence intervals for 4 km × 4 km block median values, related to four concentration levels. Estimates are obtained by ordinary kriging (on log-scale), using information from the monitoring networks only (results apply to non-seepage locations only)* 69
- 5.10 *The effects of using information on atmospheric deposition of zinc. 95% Confidence intervals for 4 km × 4 km block median zinc concentrations, related to four levels. Estimates are obtained by universal kriging (on log-scale), using both information from the monitoring networks and ancillary information (zinc atmospheric deposition rates). Only results that differ from Fig. 5.9 are shown (results apply to non-seepage locations only)* 70
- 5.11 *Map of sulphate measurement sites. Shallow groundwater measurement sites (+) and groundwater quality monitoring network sites (•) in the category semi-natural vegetation on sand* 72

- 5.12 Variograms of sulphate. Variogram for (a) ${}^e\log(\text{SO}_4)$ measurements from the groundwater quality monitoring networks (category semi-natural vegetation on sand), for (b) ${}^e\log(\text{SO}_4)$ from the shallow groundwater measurements, and (c) cross variogram for (a) and (b). A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97. Calculation of the sample variograms is explained in Tables A2 (page 96) and A7 (page 98) 73
- 5.13 Map of sulphate in the groundwater (category semi-natural vegetation on sand). 95% Confidence intervals for 4 km × 4 km block median values, related to four concentration levels. Estimates are obtained by ordinary kriging (on log-scale), using measurements from the groundwater quality monitoring networks only and the variogram of Fig. 5.12a 75
- 5.14 The effects of using shallow groundwater measurement information. 95% Confidence intervals for 4 km × 4 km block median sulphate concentrations, related to four levels. Estimates are obtained by cokriging (on log-scale), using measurements from shallow groundwater and from the groundwater quality monitoring networks, and the variograms of Fig. 5.12a-c. Only results that differ from Fig. 5.13 are shown 76
- 5.15 Local scatter plots of measurements versus ancillary variable for $\text{NO}_3\text{-N}$ (a) and zinc (b). For five arbitrarily chosen (but distinct) local kriging neighbourhoods **R**, scatter plots show concentration in groundwater (a: $\text{NO}_3\text{-N}$ in g/m^3 , b: zinc in mg/m^3) versus the ancillary variable (a: nitrate leaching rate in $\text{kg}/(\text{ha}\cdot\text{yr})$, b: atmospheric deposition of zinc in $\text{mol}/(\text{ha}\cdot\text{yr})$). Arrows link the location for which kriging neighbourhood information is shown to the corresponding scatter plot 78
- 6.1 Monitoring network optimization: total cost as a function of measurement intensity. The arrow points to an ‘optimal’ measurement intensity 81

List of Tables

- 3.1 *Basic information: (1) soil-land use category, (2) area covered by the category, (3) number of selected measurements from the national groundwater quality monitoring network for each category and (4) number of selected measurements from the provincial groundwater quality monitoring network for each category. Figs. 3.1 and 3.2 show the spatial information on (2), (3) and (4)* 24
- 3.2 *The effects of stratification for aluminium. 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 columns (b) and (c) are shown in Figs. 3.9 and 3.6 respectively* 36
- 3.3 *The effects of monitoring network density for aluminium. Number of cells for which the value of a 95% confidence interval for the 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 about variograms (second order information), and/or (b) loss of information about the mean level and location specific information (first order information). Changes from each one of these components are also shown. Changed cells are shown in Figs. 3.11 (total), 3.12 (a) and 3.13 (b)* 36
- 5.1 *The effects of ancillary information. For each variable and level the number and percentage of cells for which the addition of ancillary information changed the outcome (possible values: lower, higher and not distinguishable)* 77

Chapter 1

General introduction

1.1 Problem definition

Pollution from agricultural and industrial origin threatens the groundwater quality in the Netherlands (Van Duijvenbooden, 1989). Locally, this pollution is measured in the groundwater at tens of metres depth (CCRX, 1994, 1995). Since groundwater is the main source of fresh water, pollution causes a decrease in the long-term resources of water suitable for human consumption, and currently water companies have to increase their efforts for making groundwater suitable for drinking (Versteegh et al., 1995). Furthermore, because most groundwater finally flows to the surface, a deterioration of groundwater quality affects the ecohydrological conditions of the receiving seepage areas as well as the receiving seas.

In order to get insight in the current situation of groundwater quality and the systematic changes of groundwater quality over time, the national groundwater quality monitoring network (NGM) was established from 1978 to 1984 (Van Duijvenbooden et al., 1985). This network consists of 370 permanent wells spread fairly evenly over the country (Fig. 3.2, page 27), with screens at 8-10 and 23-25 metre below the soil surface. The network is sampled yearly, at first for 19 variables, and more recently for 25 variables. In recent years all provinces commenced the installation of similar monitoring network sites, the provincial groundwater quality monitoring networks (PGMs). Currently, they provide a doubling of the number of wells in the NGM.

Because the monitoring network is a major financial investment, questions like ‘do we have sufficient measurements?’ or ‘can a reduced monitoring network be sufficient?’ arise. Such questions can only be answered when we have answers to the questions ‘sufficient for what?’ and ‘what exactly do we need to know about groundwater quality?’ Clearly, this calls for the quantification of what can be inferred from available information about the quality of *all* the groundwater of interest. Therefore we need (a) to choose a suitable, effective and estimable measure of groundwater quality (what characteristic of groundwater quality do we need to know?), and (b) to handle the estimation uncertainty in a meaningful way (to what extent can we infer this characteristic from available information?). The uncertainty about the true groundwater quality that results from limited estimation accuracy is the key parameter in optimization studies, where a trade-off between physical cost and losses due to limited knowledge should be made: if we need to know everything about groundwater then we will have to measure everything, which will cost too much; if we do not want to spend money, we will have to live with complete uncertainty about the groundwater quality. Somewhere between these two extrema lies an optimal groundwater quality monitoring network.

1.2 Groundwater quality—what is it?

Groundwater may be defined as the subsurface water in soils and rocks that are fully saturated (Ward and Robinson, 1989). Due to a precipitation surplus, part of the infiltrated rainwater flows through the soil and reaches the saturated zone, where it becomes groundwater. In general, fresh groundwater flows in the direction of the lowest groundwater heads, usually the lower elements in the landscape (e.g., a sea, lake, river, depression or local drainage point), where it finally exfiltrates as surface water. The flow speed depends on the gradient (slope) in the groundwater table and the permeability (or resistance) of the soil. Conceptually, groundwater flow can be divided into vertical flow and horizontal flow. In the area with sandy soils (Fig. 3.1, page 26), the vertical, downward flow speed of groundwater near the groundwater table is estimated to be approximately 1 metre per year (Meinardi, 1994) in recharge (infiltration) areas. The horizontal flow may range from a few metres to hundreds of metres per year, depending on local circumstances.

When infiltrating water reaches the groundwater table it has a chemical composition that depends on the composition of the rainwater, on the condensation due to evapotranspiration, and on the absorption, desorption, chemical and biochemical reactions that occurred during the flow through the possibly contaminated soil. The chemical composition of groundwater is variable in space and time as a result of variability in the amount and in the chemical composition of rain, of variability in the chemical composition of infiltrating water (possibly resulting from agricultural or industrial pollution), variability in evapotranspiration, as a result of soil heterogeneity, of interaction of infiltrating water with the solid phase and of the heterogeneity of the geochemical and geohydrological environment. It is often established that the variation in chemical composition is very large in the most shallow groundwater (i.e. the upper metre of groundwater). During the flow of groundwater, mixing occurs, and it is expected that as a result of this mixing the large (horizontal and vertical) short-distance variation decreases with depth.

Groundwater quality is the suitability of groundwater for a certain purpose. Here, we will define groundwater quality as the suitability of groundwater for human consumption. This suitability depends mostly on the chemical composition of groundwater (note that in some cases microbiological components may be important as well). As far as groundwater quality is determined by chemical composition, it can be mapped by showing the chemical composition in relation to critical levels for human consumption, e.g. (multiples of) target concentrations or maximum tolerated concentrations.

1.3 Monitoring groundwater quality

The national groundwater quality monitoring network (NGM) was established from 1978 to 1984. The objectives of this network are (Van Duijvenbooden et al., 1985) to investigate (i) the quality of the groundwater in the upper aquifer, (ii) the extent of human influence on groundwater quality and (iii) the changes of groundwater quality over time; and to collect sufficient information (iv) to allow a good management of groundwater resources, and (v) for the development of groundwater quality management models.

Human influence on groundwater quality would be detectable at first in the most

shallow groundwater, and therefore it seems sensible to monitor only the upper few metres of groundwater. However, due to the large short-distance variation in shallow groundwater this would call for so many monitoring sites to be effective that the costs of such a network would be too high. For this reason, the upper screens of the NGM were installed somewhat deeper, at 8-10 m depth. In recharge areas the groundwater measured at this depth has infiltrated approximately 10 years before the time of measurement. Screens of 2 m length result in an integrated sample of a 2 m layer of groundwater, thus representing groundwater that infiltrated in a period of approximately 2 years. The screen length choice has been a compromise: shorter screens might have revealed variations at short depth intervals that result from (irrelevant) seasonal variation, whereas longer screens result in averaging over a longer time (depth) interval which would have hindered the detection of systematic changes of groundwater quality over time.

The monitoring well locations of the NGM were selected with the following criteria in mind (Van Duijvenbooden et al., 1985):

- (i) the locations should be spread rather evenly over the country (e.g. see Fig. 3.2)
- (ii) a well should be sited on the prevalent soil type and land use of the location
- (iii) a well site should, as far as possible, be on the off-stream side of a contiguous area with constant soil type and land use, thus enabling a non-ambiguous determination of soil type and land use of the influence area for well screens
- (iv) practical considerations, e.g. it must be possible to reach the location by car.

Monitoring wells are sampled yearly because only the structural change in groundwater quality over several years is of direct interest. In order to avoid possible seasonal fluctuations to show up in the time series for a well, each well is visited every year in the same month. The sample that is analysed in the laboratory is taken after the withdrawal of approximately 100 litres of groundwater. Therefore, at a sediment porosity of 0.3, this sample represents the groundwater present in a sediment volume of approximately one third of a cubic metre. The set of variables measured has grown over the years. Most of the variables measured in the first years of measuring are listed in Table A5 (page 97), the variables measured in 1991 are listed in Table A1. Obviously, the choice of variables that are measured is guided by the relevance of the variables for groundwater quality assessment.

The area of influence for a well screen is the area where groundwater sampled at that screen has infiltrated. Horizontal groundwater flow causes a horizontal displacement of the influence area relative to a measurement location.

It is expected that factors like land use, soil type and geohydrological situation have an influence on groundwater quality. For the mapping of groundwater quality—the estimation of current situation and systematic temporal changes of groundwater quality variables—these factors should, as far as possible, be taken into account.

1.4 Modelling groundwater quality

When we model groundwater quality, we try to summarize the spatial and temporal variation of groundwater quality variables with a limited set of concepts or mathematical equations. Because of the complexity of variation in groundwater quality and our limited knowledge of it, any model must be a simplification of reality.

Not all models are suitable for mapping purposes. For instance, a model that needs the chemical composition of the groundwater *as input* for assigning a certain characteristic to it (e.g. Frapporti et al., 1993) can only be applied at measurement sites, i.e. where this composition is known. Such a model is often a descriptive model: it summarizes (describes) the measurements. For mapping purposes it is necessary that a model is predictive: it must be capable of estimating the groundwater quality at locations where it was not measured. In addition to measurements, predictive models typically use one or more independent variables whose value is known at all estimation (mapping) locations.

To be complete, a physically and chemically based deterministic groundwater quality model would contain at least three-dimensional non-stationary convective and dispersive transport and account for geochemical reactions. For the transport modelling (e.g. Domenico and Schwartz, 1990) it would then be necessary to know for every location the initial groundwater table, the groundwater recharge as a function of time, and for every location and depth (or more precisely, for every spatial unit that discretizes the three-dimensional model space) the permeability and dispersion coefficients of the sediment. For the geochemical modelling (e.g. Appelo and Postma, 1993) it would then be necessary to know the initial chemical composition of groundwater, the chemical characteristics of the sediment, the groundwater temperature and many equilibrium constants, again for all model units. Collecting the necessary information—if possible—is a major effort. For practical reasons, a model on a national scale would have units the size of hundreds of metres to kilometres, and even if we had acceptable estimates of the necessary variables and parameters, it remains questionable whether the constants and equations used in such a model are still relevant when applied to spatial units this size (e.g., Beven, 1985).

As will be shown later, mapping groundwater quality is possible by using much simpler models that lump most of the above mentioned factors into a spatially dependent stochastic term.

1.5 Groundwater quality monitoring network optimization

The groundwater quality monitoring networks account for a substantial part of the total cost of environmental quality monitoring networks, and questions arise as to whether the way we currently monitor groundwater quality is worth the cost. Monitoring network optimization would at least involve the evaluation of the total cost of a network with respect to measurement intensity and parameter choice, where total cost includes cost of measuring and the loss due to limited knowledge of the groundwater quality. The specification of such a loss function is non-trivial, and is beyond the scope of this study.

Before we can make a judgment about the value of the current monitoring network, we have to establish to what extent we can infer the current situation and changes in time of groundwater

quality for the whole of the Netherlands. For this, we must map (estimate) the groundwater quality, using as much as possible relevant information: measurements from the national and provincial groundwater quality monitoring networks, map information on soil type and land use, hydrological parameters and deterministic process models. The estimation accuracy limits the possibility to infer groundwater *quality* from estimates.

1.6 Objectives and approach

The objectives of this study are to

- map groundwater quality in the Netherlands, using available measurements from the national and provincial groundwater quality monitoring networks and map information on soil type and land use
- show the effects of monitoring network density and the effects of using soil type and land use information on the resulting groundwater quality maps
- map the systematic, temporal changes in groundwater quality in a way similar to the mapping of groundwater quality
- show how groundwater quality maps can be improved by using relevant ancillary information in the estimation procedure, where ancillary information is obtained from deterministic process models or from other measured variables.

The primary aim is, with these objectives, to try to answer the basic questions of describing the current situation and systematic temporal changes of groundwater quality for the whole of the Netherlands.

In this study fairly simple models are used in which many of the factors (or processes) that could be used in a complex deterministic physical-chemical approach are lumped into a stochastic term that may be spatially dependent. The models used allow an explicit quantification of the accuracy of resulting estimates. This accuracy is taken into account in the resulting maps, anticipating the question about the value of the current monitoring networks for inferring current situation and systematic changes in time of groundwater quality.

This thesis explains the methods used, illustrated by results on a few groundwater quality variables. Complete results for all groundwater quality variables measured in the monitoring networks for maps of 1991 (Chapter 3) and for maps of changes in groundwater quality (1980 vs. 2000, Chapter 4) are found in separate reports (Pebesma and De Kwaadsteniet, 1994, 1995).

1.7 Outline of this thesis

The statistical mapping of quantitative spatial variables like groundwater quality variables is introduced in Chapter 2. It explains the basics of spatial variation, errors, uncertainty and statistics, why we use geostatistical models for the spatial estimation of variables, and why we present results with limited but known accuracy as confidence interval maps, and it concludes

with an outline of the methods used for mapping the groundwater quality.

Chapter 3 describes how maps were made for all groundwater quality variables that were measured at 5-17 m depth in 1991. Based on a stratification by soil type and land use, selected measurements, and an interpolation method that takes spatial dependence between measurements into account, estimates of 4 km \times 4 km local mean (block median) concentrations have been presented on maps in terms of the approximate 95% confidence intervals. The resulting maps and the effects of stratification and monitoring network density on these maps are illustrated with examples on aluminium concentration.

In Chapter 4 short-term predictions for two extrapolation moments (1980 and 2000), obtained at the monitoring sites from measurement series are used to obtain maps of statistical indications of time-trends. This is illustrated by results on potassium concentration.

Chapter 5 shows how maps of groundwater quality variables can be improved by using ancillary information obtained from deterministic process models or from other measured variables.

Chapter 6 discusses some general aspects of Chapters 3-5: it addresses monitoring network optimization, it shortly lists the main conclusions, and it gives some future directions. More specific aspects of monitoring network optimization, i.e. the value of results from this study for groundwater quality monitoring network optimization are addressed in Chapters 3 (monitoring density), 4 (monitoring frequency) and 5 (improving estimates with ancillary information). The aspect of uncertainty, resulting from limited sample size is addressed throughout this study.

Chapter 2

Statistical Mapping

2.1 Introduction

This chapter describes statistical mapping: why we need it, what statistical maps are, how we can create them, why we use geostatistics, why we present statistical maps as confidence intervals and how we make statistical maps of groundwater quality variables. Without claiming to be comprehensive, this chapter introduces the most obvious statistical approaches to mapping groundwater quality variables, i.e. variables that have a value at every location but which are sampled at a limited number of locations only. Although geostatistics was not developed from classical statistics (Cressie, 1990), here it is introduced as the obvious generalization of classical statistical ‘prediction’ for spatial applications. This chapter does not give a full introduction in statistics or geostatistics, but merely describes the methods used in the following chapters. More complete references are e.g. Cressie (1991), Draper and Smith (1981), Journel and Huijbregts (1978), and Christensen (1987, 1991).

A map is a drawing of some attribute of an area as it would appear if it was seen from above: it is a special type of graph that shows observations in geographical space mapped in two dimensions by making a scaled (and therefore usually simplified, generalized) image. Maps tell us *where* something happens. Maps often show *directly observed*, i.e. visually observed or measured phenomena that can for instance be discrete events such as the location of a city or a road, numbers such as number of inhabitants at the location of the city, or continuous variables such as the measured value of a groundwater quality variable at the measurement locations (Fig. 2.1). For technical reasons, elements on a map can only be displayed with limited accuracy. In order to show not more than what is known, this display accuracy should not exceed the extent to which the elements are known, and generally this is solved by choosing a proper scaling and display resolution. In the environmental sciences it is very common that the ‘observations’ shown on a map do not directly portray observed phenomena but quantities that are only known approximately, and in this case the need to limit the display accuracy becomes more important.

Errors in maps, the discrepancies between what the map shows and the part of reality aimed at, can be ascribed to locational errors and attribute errors. Errors that accrue from locational uncertainty will not be addressed here. Attribute error is the discrepancy between the value shown at a certain location on a map and the real, true value that the map aimed to show. Attribute errors usually stem from incomplete knowledge of the attribute in the map area, and these errors occur easily when we have to estimate the attribute from measurements: the measured sample is often difficult and expensive to obtain, and reflects only a small fraction of the population. Furthermore the spatial variation in the measured values can be large, as in Fig. 2.1.

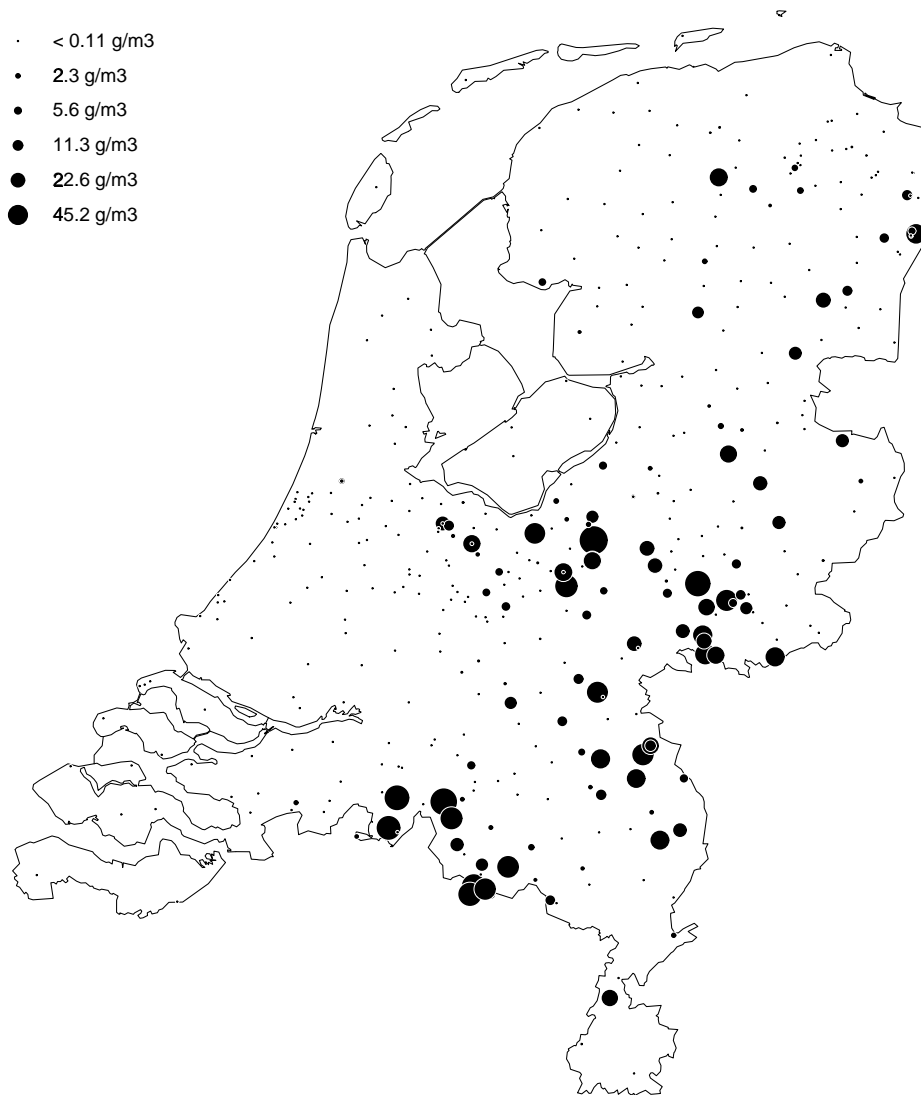


Figure 2.1 Nitrate-N concentration measured in the groundwater monitoring networks at 5-17 m depth in 1991. Bullets are centred at measurement locations. Bullet size is proportional to the measured concentration

The accuracy of a map is the closeness to reality of the map, and depends on the observations on which the map is based and on what exactly is being mapped. Fig. 2.1 is a map showing the Nitrate-N concentration *as measured* in the groundwater in the Netherlands (at 5-17 m depth, in 1991), and each measurement applies to a volume groundwater of at most a cubic metre. Consequently, the total volume of the measurements is a tiny fraction of the total volume from which the samples were taken. Since Fig. 2.1 does not tell us anything about the groundwater quality at unsampled locations, the value of it as a map of Nitrate-N in the groundwater for the whole of the Netherlands is rather limited. Although Fig. 2.1 shows some pattern, no claim is made about the concentration at unsampled locations. For mapping the groundwater quality for the whole of the Netherlands, simply showing the measurements on a map is not enough and we have to estimate the value of groundwater quality variables at all unsampled locations in the mapping area using measurements and, as we will see, other relevant information whenever possible.

2.2 Uncertainty, accuracy and statistics

If we want to know the groundwater quality at some location with maximum certainty—within the measurement precision—then we have to measure it. Often, we cannot measure it at all desired locations for practical reasons, and we have to be satisfied with an estimate of it, which will be an approximate, uncertain answer. We estimate the value at an unsampled location usually by averaging measurements from *similar* locations, i.e. locations where the conditions are similar to those of the estimation location. How far this estimate may depart from the true value—the estimation accuracy—can be derived from the number of and the variation in the measurements at similar locations.

If we are interested in the mean value of a groundwater quality variable over the whole country, simply taking more measurements (e.g. at randomly chosen locations) would increase the accuracy of the estimate of the mean. However, it would be rather naive to use this grand mean as an estimate for the value at a specific location: this estimate would be (i) not specific (it would be the same everywhere) and (ii) highly inaccurate, because we know that local conditions cause groundwater quality variables to deviate locally from this grand mean.

In order to obtain estimates that are more specific than this, for an estimate of a groundwater quality variable at a specific location we should select measurements that are obtained from locations with conditions that are similar to those at the estimation location. Only on the basis of external, independent information about local conditions can we decide which observations may be considered for this selection. For mapping purposes, the external, independent variables that define these conditions need to be known for the whole area of interest, i.e. at measurement locations as well as at estimation locations.

Ideally, ‘similar conditions’ for an unsampled location would be specified in such a way that the set of measurements at connected locations—having the same conditions—is (i) homogeneous (ii) large enough to yield good estimates (iii) dissimilar to measurement sets at other locations. This would result in estimates that are specific and accurate enough to reveal local differences in the mapping variable.

External, independent variables

External, independent variables that define similarity of conditions can be categorical variables, continuous variables, or combinations of them. An example of a categorical variable that would be useful for the mapping of many groundwater quality variables would for instance be the area with marine influence: if we knew this variable (assuming we could define it), none of the estimates for fresh groundwater would be ‘contaminated’ with measurements from the area with marine influence. An example of a continuous variable defining similarity of conditions could be the atmospheric deposition of a variable. If atmospheric deposition is the main source for a variable measured in the groundwater, the measurements at locations with deposition rates similar to the deposition rate at the estimation location are expected to yield the best estimate of the variable at the estimation location.

In general, independent, external variables that are suitable for the purpose of mapping groundwater quality are variables that cause (or explain) a major part of the spatial variation in groundwater quality, such as geochemical variables (e.g. demarcating the area with marine influence, or oxygen- or calcium-rich groundwater), or variables that define fluxes from human

supply (e.g. from manuring, pesticide application, atmospheric deposition from industrial origin, etc.). The constraint of being known at all locations in the mapping area limits the possibilities (Sections 3.1, 5.2).

Model structure

The mathematical formulation of the set of relations connecting independent, condition defining variables to the measurements is the model. Which independent variables are used, and how the measurement variable depends on them is the model structure.

The choice of the model structure should be guided both by availability of data, i.e. measurements and independent, condition defining variables, and by theory, the *a priori* ideas we have about the process we are modelling. Theory and data should be balanced when choosing the model structure. If a model is too complex it is over-specified, and this is inefficient because more parameters than necessary have to be estimated. In an extreme case some of the parameters cannot be identified by the data and it is likely that such a model will produce inadequate estimates which more reflect the creator's fantasy than the reality aimed at. On the other hand, if a model is too simple it is under-specified. This is inefficient because it will result in unnecessarily inaccurate estimates. Since the complexity of the model structure depends on the data availability, there is not one correct model, and more data will usually suggest and allow a more complex model structure. The compromise in model structure complexity we usually choose is the simplest adequate model that is parsimonious in its parameters.

Estimation

Only after choosing a model structure that is suitable for the problem at hand, it is possible to estimate the quantity of interest. This demands an explicit statement of the objective: which characteristic of a measured groundwater quality variable do we want to estimate? We can for instance be interested in the value that would be measured at a specific location, the value that would be measured at a location randomly chosen from a specified area, the mean or median value of all measurements in a specific area, or the fraction of an area that has values above a critical level. Not all models are equally suitable for estimating a specific characteristic. Given a set of relations that defines the model structure, we have to choose a statistical procedure to estimate the characteristic of interest.

Accuracy

When we estimate an unknown quantity, we never *know* the estimation error. However, from measurements we can retrieve information about the size of the error we can expect, as averaged over similar estimation problems. In some instances observations are so regular or so numerous that the accuracy of the variable we wish to map is of no concern: the accuracy is good enough. For groundwater quality variables the accuracies of estimates are usually not negligible, and we cannot ignore this when presenting estimates in maps. Why and how estimation accuracies are accounted for when mapping estimates will be explained in Section 2.6, but we will first explain some basic concepts of statistical and geostatistical models.

Statistics

Statistics is the discipline that provides theory and tools for deciding in the data-theory complexity

trade-off, for inferring (estimating) population characteristics from limited sample information and for assigning accuracy measures to the inference. Here, statistical mapping is defined as mapping results from an (inferential) statistical analysis, i.e. combining estimated value and accuracy measure in one map. The most commonly used models in statistics are some form of a linear model.

2.3 Linear models

For mapping groundwater quality—estimating groundwater quality variables at locations where they are not measured—we need a model. From all possible models we choose statistical models, and the family of linear models provides a comprehensive framework for most commonly used statistical models. A limitation of linear models is that they only allow additive effects. For groundwater quality variables (being non-negative and highly skewed) additivity of effects is often a reasonable assumption after log-transforming the measurements. Wider classes of problems can be formed, e.g. by non-linear transformation of variables (as in generalized linear models, McCullagh and Nelder, 1989), or by defining linear relations *locally* (approximating more complex relations with local linear models, as in generalized additive models, Hastie and Tibshirani, 1990).

Linear models are a means of expressing real-life problems in a mathematical form, and are therefore helpful in the abstraction of problems and in communication. Comprehensive treatments of linear models are found in Rao (1973), Searle (1971), Christensen (1987), and in Draper and Smith (1981). Cressie (1991) and Christensen (1991) extend these treatments to spatial (i.e. mapping) problems. A note on terminology: in classical statistics the word ‘estimation’ is reserved for assigning values to model parameters, whereas ‘prediction’ is used to denote the evaluation of random variables, but here I will use ‘estimation’ for both, following most of the traditional geostatistical literature (e.g. Matheron, 1971, 1989 (page 4-5), Journel and Huijbregts, 1978).

Three simple problems

Linear models provide a flexible way of expressing a wide range of problems in a compact notation. Consider the three following simple problems:

- (i) n groundwater quality measurements are collected randomly from a homogeneous spatial unit, and we want to estimate the mean value of the measured variable in this unit and its estimation variance
- (ii) n groundwater quality measurements are collected in two spatial units, q in the first, $n - q$ in the second, and we want to estimate the mean (and its estimation variance) of each group
- (iii) n nitrate concentration measurements are collected in a spatial unit, and for every measurement location the manuring rate is known. We assume that nitrate concentration increases linearly with manuring rate, and we want to know the rate of increase (the increase in nitrate concentration at a unit increase in manuring rate)

For these simple problems the observations $z(x_i)$ at locations x_i , $i = 1 \dots n$ can be written as

$$(i) \quad z(x_i) = m + d(x_i), \quad \forall i : i = 1 \dots n$$

$$(ii) \quad z(x_i) = m_1 + d(x_i), \quad \forall i : i = 1 \dots q$$

$$z(x_i) = m_2 + d(x_i), \quad \forall i : i = q + 1 \dots n$$

$$(iii) \quad z(x_i) = \beta_1 + r(x_i)\beta_2 + d(x_i), \quad \forall i : i = 1 \dots n$$

Here, *conceptually* the measurements are taken as the sum of a structural, systematic part and a residual, unsystematic part. The structural part consists of (i) m , the mean value of the spatial unit; (ii) m_1 and m_2 , the mean values of units 1 and 2; (iii) $r(x_i)$, the manuring rate at location x_i , β_2 the change in nitrate concentration at a unit change in manuring rate, and β_1 the nitrate concentration at no manuring rate. The unsystematic part in all three problems is $d(x_i)$: the deviation of the i -th measurement from the structural part.

In a linear model the observation $z(x_i)$ is represented by a random variable $Z(x_i)$, and $Z(x_i)$ is modelled as the sum of its expected value $E(Z(x_i)) = m(x_i)$ and a random deviation from $m(x_i)$, $e(x_i)$

$$Z(x_i) = m(x_i) + e(x_i), \quad E(e(x_i)) = 0,$$

(here e refers to ‘error,’ not in the sense of a mistake, but natural, residual variation that is not accounted for by $m(x)$). The expected value $m(x)$ is modelled as a linear function of p known, independent variables that have a causal influence upon $Z(x)$, and p unknown coefficients β_j that relate these independent variables to the observations

$$m(x_i) = \sum_{j=1}^p f_j(x_i)\beta_j,$$

which gives, using vector notation

$$Z(x_i) = f(x_i)\beta + e(x_i)$$

with $f(x_i) = (f_1(x_i), f_2(x_i), \dots, f_p(x_i))$ the row vector with the values of the independent variables at location x_i , and $\beta = (\beta_1, \dots, \beta_p)'$, the column vector with the unknown coefficients. Representing all observations, the model can be written as

$$Z(x) = F_x \beta + e(x)$$

with $Z(x) = (Z(x_1), \dots, Z(x_n))'$, $F_x = [f_j(x_i)]_{n \times p} = (f_1(x), \dots, f_p(x))$ with $f_j(x) = (f_j(x_1), \dots, f_j(x_n))'$, and $e(x) = (e(x_1), \dots, e(x_n))'$. When we define F_x and β accordingly, this model covers the three simple problems discussed above:

$$(i) \quad p = 1, f(x_i) = 1, i = 1 \dots n, \text{ and } \beta_1 = m$$

$$(ii) \quad p = 2, \text{ if } i \leq q \text{ then } f(x_i) = (1, 0) \text{ else } f(x_i) = (0, 1), \text{ and } \beta = (m_1, m_2)'$$

(iii) $p = 2, f(x_i) = (1, r(x_i)), i = 1 \dots n$, and $\beta = (\beta_1, \beta_2)'$.

Thus we can structure the problem by choosing F_x (and thus defining the size of β). The j -th column in $F_x, f_j(x)$ defines the structure of the relation of the measurement variable $z(x)$ to the j -th parameter β_j . If β_j is the overall mean as m in problem (i) or an intercept as β_1 in problem (iii), then $f_j(x)$ is a column of ones. If the model contains categories as problem (ii) and β_j is the mean of the j -th category, then $f_j(x)$ is the binary variable that for every observation denotes whether it belongs to the j -th category with a one, or not, with a zero. In case of a regression-type relation like in problem (iii), $f_j(x)$ contains the value of the regressor, the independent variable at the observation locations.

Before we can estimate β or the value of observations at unsampled locations, we have to specify the structure of the errors.

Linear models with independent, identically distributed errors

In the simplest case we assume that the errors $e(x)$ are independent and identically distributed (IID), resulting in the model

$$Z(x) = F_x \beta + e(x), \quad E(e(x)) = 0, \quad \text{Cov}(e(x)) = \sigma^2 I, \quad (2.1)$$

which leads to the ordinary least squares (OLS) estimate (provided that F_x has full rank)

$$\hat{\beta} = (F_x' F_x)^{-1} F_x' z(x), \quad (2.2a)$$

and estimation variances and covariances for $(\beta - \hat{\beta})$

$$\text{Cov}(\beta - \hat{\beta}) = (F_x' F_x)^{-1} \sigma^2, \quad (2.2b)$$

and where σ^2 is estimated by

$$s^2 = z(x)'(I - F_x(F_x' F_x)^{-1} F_x')z(x)/(n - R), \quad (2.3)$$

with R the rank (the number of columns) of F_x .

At an unsampled location x_0 , given this estimate $\hat{\beta}$, the value of $z(x_0)$ (and the mean value of r independent replications of $z(x_0)$) is estimated by

$$\hat{z}(x_0) = f(x_0) \hat{\beta}, \quad (2.4)$$

with $f(x_0)$ the value of the independent variables at location x_0 . The estimation variance of the estimator of $z(x_0)$ ($r = 1$) or the mean of r independent replications of $z(x_0)$ is given by

$$\sigma_r^2(x_0) = \left(\frac{1}{r} + f(x_0)(F_x' F_x)^{-1} f(x_0)'\right) \sigma^2. \quad (2.5)$$

If the assumption of IID errors is appropriate, the equations (2.2a)-(2.3) lead to the answer of the

three simple questions posed at the beginning of this section.

Linear models with dependent errors

A wider class of problems than the one with IID errors is obtained when the errors are allowed to be dependent

$$Z(x) = F_x \beta + e(x), \quad E(e(x)) = 0, \quad \text{Cov}(e(x)) = V \quad (2.6)$$

with $V = [\text{Cov}(e(x_i), e(x_j))]_{n \times n}$. This leads to weighted least squares (WLS) estimates of β

$$\hat{\beta}_* = (F_x' V^{-1} F_x)^{-1} F_x' V^{-1} z(x), \quad (2.7a)$$

with estimation covariances

$$\text{Cov}(\beta - \hat{\beta}_*) = (F_x' V^{-1} F_x)^{-1}. \quad (2.7b)$$

Under this model, given $\hat{\beta}_*$ by (2.7a), the estimate of $z(x_0)$ is

$$\hat{z}_{wls}(x_0) = f(x_0) \hat{\beta}_* + v_0' V^{-1} (z(x) - F_x \hat{\beta}_*) \quad (2.8)$$

where $v_0' = (\text{Cov}(e(x_1), e(x_0)), \dots, \text{Cov}(e(x_n), e(x_0)))$, and has estimation variance:

$$\sigma_{wls}^2(x_0) = C(0) - v_0' V^{-1} v_0 + (f(x_0) - v_0' V^{-1} F_x) (F_x' V^{-1} F_x)^{-1} (f(x_0) - v_0' V^{-1} F_x)' \quad (2.9)$$

with $C(0) = \text{Var}(e(x_0))$. In statistical terms, the value estimated in (2.8) is expressed as the sum of the best linear unbiased estimate (BLUE) of $m(x_0)$, $\hat{m}(x_0) = f(x_0) \hat{\beta}_*$ and the best linear unbiased predictor (BLUP) of the (correlated) error, $\hat{e}(x_0) = v_0' V^{-1} (z(x) - F_x \hat{\beta}_*)$.

Multivariable estimation

When s variables $Z_k(x)$, $k = 1 \dots s$ each follow a linear model $Z_k(x) = F_{k,x} \beta_k + e_k(x)$, and the $e_k(x)$ are correlated, then it makes sense to extend the weighted least squares model to allow multivariable estimation. Without loss of generality, assume $s = 2$. When $\mathbf{z}(x) = (z_1(x), z_2(x))'$ and $\mathbf{B} = (\beta_1, \beta_2)'$ are substituted for $z(x)$ and β , and when

$$\mathbf{f}(x_0) = \begin{bmatrix} f_1(x_0) & 0 \\ 0 & f_2(x_0) \end{bmatrix}, \quad \mathbf{F}_x = \begin{bmatrix} F_{1,x} & 0 \\ 0 & F_{2,x} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \quad \mathbf{v}_0 = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix},$$

with $V_{21} = [\text{Cov}(e_2(x_i), e_1(x_j))]$, $v_{21} = (\text{Cov}(e_2(x_1), e_1(x_0)), \dots, \text{Cov}(e_2(x_n), e_1(x_0)))'$, and 0 a conforming zero matrix or vector, are substituted for $f(x_0)$, F_x , V and v_0 , then the left-hand sides of both (2.8) and (2.9) yield the multivariable estimates: the left-hand side of (2.8) then becomes the estimate vector $\hat{\mathbf{z}}(x_0) = (\hat{z}_1(x_0), \hat{z}_2(x_0))'$, and the left-hand side of (2.9) becomes the (2×2) matrix with estimation covariances.

Confidence intervals

When the estimation error $z(x_0) - \hat{z}(x_0)$ is normally distributed with zero mean and variance $\sigma^2(x_0)$, confidence intervals can be constructed for $z(x_0)$ (from, depending on the model used, (2.4) and (2.5), or (2.8) and (2.9)), and the interval

$$[\hat{z}(x_0) - 2\sigma(x_0), \hat{z}(x_0) + 2\sigma(x_0)] \quad (2.10)$$

is a 95% confidence interval for $z(x_0)$.

2.4 Change of support and error structure

Usually a measurement is not a characteristic of a point in space, but an average over some area or volume, the *support* of the measurement (Journel and Huijbregts, 1978). For groundwater quality, the volume measured applies to something between the volume pumped from the well before a sample is taken (i.e. approximately hundred litres) and the volume that is actually measured (i.e. a few millilitres). We hope that enough mixing occurred during sampling and that the measurement on the small volume is representative for the larger volume. In practice the measurement support is determined by practical limitations: if it were easy we would have measured the relevant population characteristic (e.g. mean groundwater quality in a certain area) *directly*. The support of measurements also determines the variation in the values observed: measured means of larger spatial units show less variation, because variation within the units is averaged (Journel and Huijbregts, 1978).

Meaningfully mapping groundwater quality variables for units of the size of the measurements—estimating values that would actually be measured at unsampled locations—is not possible because the variation in the measurements is frequently too large and consequently we cannot estimate groundwater quality with a reasonable accuracy at that scale. At a lower spatial resolution (e.g. for mean values of adjacent or overlapping square cells on a grid) it is possible to estimate groundwater quality variables because the pattern of local average groundwater quality is smoother than the pattern of the measurements. For the estimation of this smooth pattern of local averages we need a model that allows estimation of spatial local average values.

It would be convenient to use the linear model with independent (IID) errors presented in the previous section because it is simple and it is supported by a rich body of research in classical statistics. However, if we want to obtain independence from a design-based argument (e.g. Hansen et al., 1983, De Gruijter and Ter Braak, 1990), then this model is only adequate for estimating the pattern of local mean values when (i) the observations are collected *randomly* from the areas with constant values for the $f(x_i)$ (i.e. from the categories distinguished or from an area with a specific value for the regressors) and (ii) the areas for which we want estimates of a spatial average *coincide* with these areas of constant $f(x_i)$. These conditions limit the suitability of this model for estimation *local* averages severely.

Therefore, for the mapping of groundwater quality variables, we will continue with models that allow errors to be spatially dependent. Moreover, although part of the residual error can usually be attributed to measurement error, it is very likely that in a linear model

intended for spatial estimation a large proportion of the residual, unexplained variation is caused by (unknown) spatially smooth factors, resulting in a spatially dependent error, and for mapping purposes we can use this spatial dependence to capture the spatial structure present in the measurements beyond the part explained by the independent variables. Average values of arbitrarily shaped elements can be estimated efficiently if we are willing to assume a model with a spatially dependent error structure. Estimating local averages using a model that allows spatial dependent errors is a typical problem in geostatistics.

2.5 Geostatistics

Geostatistics is the discipline that provides a set of models and tools for the estimation of block averages or local averages from sample observations, taking both large scale variation (the trend) and small scale variation (spatial correlation) into account. Typically, in a geostatistical model the observations $z(x)$ are considered as obtained from a realization of the random field $Z(x)$. For this random field we assume that the universal kriging model

$$Z(x) = m(x) + e(x) \quad (2.11)$$

holds, with expectation

$$E(Z(x)) = m(x) = \sum_{j=1}^p f_j(x)\beta_j = F_x\beta \quad (2.11a)$$

and error

$$E(e(x)) = 0, \quad \text{Cov}(e(x)) = V, \quad (2.11b)$$

with $m(x)$ the trend, the large scale fluctuation of $Z(x)$, and $e(x)$ the (possibly) spatially correlated error. Thus, the universal kriging model is a linear model with dependent errors, like (2.6). Ordinary kriging is the simplest form of universal kriging with $m(x) = m$, a constant mean. In the past universal kriging has been held synonymously with non-stationary kriging using polynomials of the spatial coordinates as the $f_j(x)$, but here we will only consider variables that carry a physical significance with respect to $Z(x)$ as clients for the independent variables $f_j(x)$, and we will account for regional differences (non-stationarity) by assuming that (2.11a) holds only *locally* (Journel and Rossi, 1989, Journel, 1992).

The challenge in geostatistical modelling is to find a meaningful structure for the trend and suitable values for the spatial correlation. Since it is impossible to obtain estimates of $\text{Cov}(e(x_i), e(x_j))$ experimentally from a single realization of $Z(x)$, additional assumptions have to be imposed on the error $e(x)$. Commonly used models for the error are the closely related second order stationary and intrinsically stationary processes. Second order stationarity implies that the error covariance between two locations x_i and x_j is a function of the separation vector $h = x_i - x_j$ only, written as

$$\text{Cov}(e(x_i), e(x_j)) = C_e(x_i, x_j) = C_e(h),$$

where $C_e(h)$ is the covariogram of $e(x)$. Intrinsic stationarity implies that the semivariance of the error process $\gamma_e(x_i, x_j) = \frac{1}{2}E(e(x_i) - e(x_j))^2$ is a function of the separation vector only

$$\gamma_e(x_i, x_j) = \gamma_e(x_i - x_j) = \gamma_e(h),$$

where $\gamma_e(h)$ is the variogram of $e(x)$. A further assumption that facilitates the estimation of covariances is isotropy. Isotropic processes have a covariogram or variogram that depends on the separation distance only, disregarding direction: $C_e(h) = C_e(\|h\|)$ and $\gamma_e(h) = \gamma_e(\|h\|)$. An advantage of using the variogram over the covariogram is that if the expectation of $Z(x)$ is a constant, the variogram can be directly expressed as a function of the observation variable $Z(x)$

$$\gamma_e(x_i, x_j) = \gamma_Z(x_i, x_j) = \frac{1}{2}E(Z(x_i) - Z(x_j))^2,$$

and hence it can be estimated directly from the measurements (Table A2, page 96), whereas covariance estimation additionally involves the estimation of the mean value.

For estimation, variogram values for any distance h may be needed. Therefore we have to choose a function that guarantees positive estimation variances (e.g. Table A3, page 97) and fit this function to the sample variogram $\hat{\gamma}_e(h_j)$, calculated from sample data at regular distance intervals h_j (e.g. Table A2, page 96 and Table A6, page 97).

If the trend contains an intercept (which will usually be true, e.g. an unknown mean value for each variable or an unknown mean for each category), and if a generalized covariance is defined as $K_e(x_i, x_j) = -\gamma_e(x_i, x_j) + d$ for an arbitrary constant d , then substituting $K_e(x_i, x_j)$ for $\text{Cov}(e(x_i), e(x_j))$ in (2.6) is sufficient for (2.8) to yield the kriging estimate of $z(x_0)$ and for (2.9) to yield the kriging variance $\sigma_k^2(x_0) = \sigma_{wls}^2(x_0)$ (Christensen, 1991, VI.3; Cressie, 1991, 5.4).

Averaged values $Z(B_0)$ for rectangular or otherwise shaped blocks B_0 are estimated by replacing the point-to-point semivariance $\gamma_e(x_i, x_0)$ with the point-to-block semivariance $\gamma_e(x_i, B_0)$, the mean of all point semivariances between x_i and the points that define B_0

$$\gamma_e(x_i, B_0) = |B_0|^{-1} \int_{B_0} \gamma_e(x_i, u) du,$$

with $|B_0|$ the area or volume of B_0 ; by replacing $\gamma_e(x_0, x_0)$ with the block-to-block semivariance $\gamma_e(B_0, B_0)$, the mean of all point semivariances between the pairs of points defining the block B_0

$$\gamma_e(B_0, B_0) = |B_0|^{-2} \int_{B_0} \int_{B_0} \gamma_e(u, v) dudv,$$

and by replacing $f(x_0)$ with $f(B_0) = (f_1(B_0), \dots, f_p(B_0))$, with $f_j(B_0) = |B_0|^{-1} \int_{B_0} f_j(u) du$, the average value of $f_j(\cdot)$ over block B_0 .

In Chapter 4 we have ‘new observations’ that are short-term predictions for an extrapolation moment, obtained from a linear regression model for each monitoring site. These ‘new observations’ are subject to a location-specific time-prediction error, and can therefore not be adequately represented by a stationary model. However, when these ‘new observations’ estimate an underlying variable $z(x)$ that can be modelled by a stationary model, then we can model

such ‘new observations’ $y(x)$ as the sum of a stationary component $Z(x)$ (as of (2.11)) and a non-stationary ‘measurement error’ $\varepsilon(x)$:

$$Y(x) = Z(x) + \varepsilon(x). \quad (2.12)$$

When we assume that (i) $\text{Var}(\varepsilon(x_i)) = \sigma_\varepsilon^2(x_i)$ is known and $\text{Cov}(\varepsilon(x_i), \varepsilon(x_j)) = 0, \forall i, j : i \neq j$, and (ii) $\text{Cov}(\varepsilon(x_i), e(x)) = 0, \forall i, x$, and (iii) the variogram $\gamma_e(h)$ is known, then, from ‘new observations’ $y(x_i), z(x_0)$ is estimated with (2.8) and $\sigma_k^2(x_0)$ with (2.9) when $K_e(0) + \sigma_\varepsilon^2(x_i)$ is substituted for $\text{Cov}(e(x_i), e(x_j))$ and generalized covariances (setting $d = 0$) are substituted for all other covariances, as described above (Delhomme, 1978).

When more than one measured variable is used for the estimation of the variable of interest and the variables are spatially inter-correlated, the multivariable extension of (2.8) and (2.9) can be used for multivariable estimation (Ver Hoef and Cressie, 1993). Inter-variable spatial dependence is usually modelled with the cross variogram (Section 5.3; Table A7, page 98) or the cross covariogram. Multivariable estimation of point or block averages is called cokriging.

Ordinary kriging within categories that subdivide the area with variograms modelled per category is called stratified kriging (Chapter 3). Stratified kriging can be seen as a special case of multivariable kriging (with the variables in the different categories taken as different variables and no cross variograms), and with a little more effort as a special case of universal kriging (with the trend modelled as a series of binary variables—a step trend—reflecting the categories, and a category-specific variogram that is only defined (i.e. not a constant) for pairs of observations from the same category).

An important adaption of the universal kriging model that leads to a wider and more flexible class of models is the *local* application of (2.11): the observations are assumed to follow (2.11a) and (2.11b) only within a local neighbourhood, e.g. defined by the number of nearest observations or the size of a neighbourhood area considered. This allows the local expectation to deviate from a globally fixed expectation, essentially weakening the stationarity hypothesis (2.11a) to local stationarity (i.e. global non-stationarity). Furthermore local estimation only calls for variogram values at distances that allow a meaningful estimation of the variogram (i.e. values at distances less than approximately half the dimensions of the area used for variogram calculation, Journel and Huijbregts, 1978, III.B.7; Matheron, 1989) and that are less affected by biased estimation of the trend parameters β (Cressie and Zimmerman, 1992).

In practice V and v_0 are not known but estimated with \hat{V} and \hat{v}_0 from sample data, causing $\hat{\sigma}_k^2(x_0) = f(\hat{V}, \hat{v}_0)$ to underestimate the true estimation error σ_k^2 (Harville, 1985, Christensen, 1991, VI.5). Thus, when $\hat{\sigma}_k^2(x_0)$ is substituted for $\sigma_k^2(x_0)$ in (2.10), the probability that the resulting intervals will cover the real value $z(x_0)$ is less than 95%. In this study, however, intervals obtained in this way will be used as first approximations to 95% confidence intervals.

Estimation using the multivariable extension of model (2.11) and variogram modelling were implemented in the GSTAT computer program (Pebesma, 1995), developed during this study. This program can be used as an extension to the GIS PCRASTER (Wesseling and Van Deursen, 1995, Van Deursen, 1995), and it allows a wide range of possibilities for the integration of point information (measurements) and map information for the mapping of the measured variable. It has been used for all geostatistical computations throughout this study, and PCRASTER has been

used for the processing and presentation of all raster maps.

2.6 Confidence interval maps

If the estimation accuracy (or more precisely, the estimation precision) of groundwater quality variables is low, we have to show this in maps because showing only the estimated value—silently pretending it is a good estimate of reality—is misleading. The estimation precision can be expressed as the estimation variance ((2.5), (2.9)), but if we present the estimation variance in a separate map, we split the information about estimate and precision, suggesting that they are two separate entities and inviting the reader to ignore the precision information.

The obvious solution to this problem is to combine estimate and estimation variance in a confidence interval (e.g., (2.10)) and to present this on a map. This can only be done if assumptions on the error distribution are made (which is not necessarily bad: the question arises what the value of an error variance is when the form of the error distribution is unknown). Confidence intervals show where the true value probably lies (mostly between the upper and lower bounds of the interval), and what the precision of the estimate is (the width of the interval). When the precision is high, the confidence interval map reduces to the classical map that shows the estimated value only, because both sides of the confidence interval contract to this value.

The confidence *level* of a confidence interval controls the fraction of intervals that cover the true value if all model assumptions are met (i.e., in a map with many independent, valid 95% confidence intervals, approximately 95% of the intervals will cover the true value). This level should be chosen with the application of the map in mind. At first sight, choosing a low confidence level (e.g. 50% or 70%) seems attractive because it leads to maps with narrower intervals, but the ‘precision’ achieved in this way is false: the fraction of matches is simply smaller. Moreover, choosing a low confidence level seriously undermines the intuitive meaning of the word ‘confidence,’ especially because the confidence level is valid only when all model assumptions are met, which will in practice only be true approximately. The choice for 95% confidence intervals for the groundwater quality maps in this study is a mixture of a matter of taste and tradition (e.g., Fisher, 1935).

The display of confidence intervals in maps raises the problem of showing two values (both sides of the interval) at one location. Splitting the interval and showing the two sides in separate maps (as in Pebesma, 1992) is not recommended, since the upper side of a confidence interval may seem alarmingly high while the lower side is very low, only meaning that we have no information about the true value. The notion of precision, the width of the interval is lost when only one side of the interval is considered.

Considering both sides of the interval simultaneously leads to the desired aim of a map: to show where we can distinguish estimates from specific values. For a single specific value, this can be easily achieved in a map: any interval is entirely *lower* than the specific value, is entirely *higher* than the value, or straddles the value (in which case the estimate and the specific value are *not distinguishable*, based on the information available). For example, Fig. 3.4 (page 30) gives a map where the position of the confidence intervals relative to four reference levels is shown in four separate sub-maps, one for each reference level. The disadvantage of this map is that each separate sub-map does not show the precision of the estimates: considering only one sub-map,

not distinguishable can be *close to* (precise) as well as *completely unknown*, and this information is only revealed by considering all sub-maps at once.

An attempt to achieve a combined presentation of both position and width of the intervals in one single map is shown in Fig. 3.5 (page 31). This map has solid coloured square cells—thus reducing to the ‘classical,’ deterministic map—where a legend class covers a confidence interval completely, and strongly contrasting upper and lower triangles in a squared cell when the confidence interval is wide (no information available). Again, this map might be misleading since red upper triangles in green-red cells may unjustly be taken as alarming.

2.7 Mapping groundwater quality

This section draws the line from theory, as set out in the previous sections, to the application of mapping groundwater quality which is the main objective of this study, described in detail in Chapters 3-5.

All groundwater quality variables considered in this thesis are concentrations of solutes and are therefore physically limited between zero and saturation level. In practice, relative to saturation level, only low concentrations occur, and the upper limit can be ignored. In general the distribution of the measurements is highly skewed: many measurements are low, few are very large. Because of the physical lower bound and the shape of the distributions, the natural scale to model these variables is the log-scale. This implies that on the log-scale effects are assumed to be additive and allows the use of linear models with stationary error distributions (e.g. Myers, 1989).

In the simplest geostatistical approach we assume that the observations in the area of interest come from a realization of an intrinsically stationary process. This means that a model with a locally constant mean value and homogeneous variation sufficiently characterizes the variable from which we have observations in the area. If this model is too simple to be realistic, we can for instance subdivide the area in several sub-areas for which this model is appropriate, or we can transform the observations non-linearly, or we can assume a more complex model for the mean value.

The groundwater quality variables measured in the monitoring networks cannot be considered as having a constant mean or having homogeneous variation throughout the Netherlands. Locally however, under similar conditions this may be a reasonable assumption. In order to divide the Netherlands completely in sub-areas that may be thought of as having a locally constant mean and being homogeneous, we need a suitable external, independent, criterion that is known at all locations in the mapping region (Section 2.3). In the simple approach of Chapter 3 we chose this criterion to be the soil-land use category of each location (Fig. 3.1, page 26).

Within soil-land use categories, from available measurements (in 1991, at 5-17 metre depth) we selected a suitable subset of measurements that could be marked as ‘similar’ measurements, allowing the use of a stationary model. From these measurements, within soil-land use categories, estimates of 4 km × 4 km block mean values, as obtained by ordinary kriging on the log-scale, were presented as approximate 95% confidence intervals for block median concentration on the original scale (Chapter 3, Pebesma and De Kwaadsteniet, 1994). The effects of stratification were studied by systematically leaving out either soil type or land use, or both; the effects of monitoring

network density were studied by leaving out a subset of the measurements.

For studying time-trends in groundwater quality (Chapter 4), attention was focused on the recognition of steady (i.e. monotonous) long-term components in changes in groundwater quality. Within soil-land use categories, using a regression model, short-term predictions for 1980 and 2000 were obtained at 250 monitoring sites from long-term (1987-1993) measurement series of the national groundwater quality monitoring network, and used in a spatial interpolation, taking time-prediction inaccuracies into account. For the spatial interpolation local level, the size of spatial variation, spatial dependence in spatial variation and location specific time-prediction variances were explicitly taken into account for each mapping variable by using a slightly adapted version of ordinary kriging, modified as described in Section 2.5 (2.12), and variograms obtained from the 1991-situation (Chapter 3).

Maps of groundwater quality variables can be improved by incorporating other relevant, ancillary information in the mapping procedure. This ancillary information may consist of either map information (e.g. obtained from deterministic process models), or point information (e.g. other measured variables). Chapter 5 explores the use of model (2.11) and its multivariable extension for this purpose, and gives several examples. Using the first type of ancillary information, map information from deterministic process models led to using universal kriging with the deterministic process model as the independent variable. The second type of ancillary information led to the use of cokriging.

Chapter 3

Maps of the groundwater quality in the Netherlands at 5-17 metre depth in 1991

(from: Pebesma and De Kwaadsteniet, 1994)

3.1 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 over the country, are sampled yearly at two depths (from well screens at about 10 and 25 metre below surface). At first samples were analysed for 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 A1, page 96) maps are made, using measurements from both monitoring networks. Uncertainties in predicted values are quantified and presented on the maps. Full results are presented in a separate report (Pebesma and De Kwaadsteniet, 1994). In this chapter we illustrate the methods used with results on one groundwater quality variable, namely aluminium 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). Soil type (Steur et al., 1985) and land use (Thunnissen et al., 1992) can be distinguished from the other factors because the data are available as digitized maps. 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 (at a fixed moment) by using kriging interpolation (Journel and Huijbregts, 1978) within categories.

Because a groundwater quality monitoring network that has a national or provincial extent by no means can 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 $4 \text{ km} \times 4 \text{ km}$ square elements were estimated for a $2 \text{ km} \times 2 \text{ km}$ 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 mind. The resulting maps of block mean values of groundwater quality (or, more precisely, block median values) can be regarded as a translation of point information obtained from the monitoring networks into information on spatial units, the size of units 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 originally available information from the NGM.

Sections 3.2, 3.3 and 3.4 describe the stratification, data selection and spatial interpolation. Section 3.5 explains the groundwater quality maps. Effects of stratification and monitoring density are discussed in sections 3.6 and 3.7. Section 3.8 concludes with a discussion.

3.2 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 analogous with considerations on the monitoring network design (Van Duijvenbooden et al., 1985), we aggregated classes to the 13 categories—apart from built-up areas—shown in Table 3.1. Here, semi-natural vegetation comprises forests, heaths and nature reserves; the category ‘dunes’ is strictly spoken not a soil-land use category, but was obtained from a map with ecological districts (Klijn, 1988). For each element on a $2 \text{ km} \times 2 \text{ km}$ grid the dominant soil-land use category was determined from an overlay of both aggregated maps. The result is shown in Fig. 3.1 (page 26). The soil type and land use information for measurements was derived from this map.

3.3 Data selection

For the generation of groundwater quality maps a number of measurements were left out because they seemed unsuitable for this purpose. For instance, monitoring sites from 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. 3.1) because they were influenced directly by water from rivers or the sea were left out too. Fig. 3.2 shows the locations of the 425 monitoring network sites that contributed to the groundwater quality maps.

The groundwater quality maps were based on measurements of the 25 groundwater quality variables available from 1991 (Table A1, page 96) in the selection described above, with monitoring screens completely between 5 and 17 metre below ground surface. Detection limit

Table 3.1 Basic information: (1) soil-land use category, (2) area covered by the category, (3) number of selected measurements from the national groundwater quality monitoring network for each category and (4) number of selected measurements from the provincial groundwater quality monitoring network for each category. Figs. 3.1 and 3.2 show the spatial information on (2), (3) and (4)

category	area		number	
	(%)	NGM	PGM	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

values were assigned to measurements below the detection limit (using the highest value when different detection limits occurred).

Some doubtful measurements were left out altogether. 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 for mapping only if (i) the measurement differed more than a factor 10 from all other measurements of that variable in the same monitoring screen in the period 1985-1992 and (ii) the factor 10 difference did not arise by 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 3.4). 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 mean value and the 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.

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 ten metres is similar to the spatial variation at a horizontal distance of a few kilometres).

3.4 Spatial interpolation

Maps of groundwater quality variables were obtained by means of ordinary block-kriging (Journel and Huijbregts, 1978) of log-transformed measurements within each of the soil-land use categories of Table 3.1, using the measurements available. For categories with only very few measurements a modified procedure was used, as described at the end of this section.

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:

$$Z(x) = m + e(x) \quad (3.1)$$

with

$$m \text{ constant,} \quad (3.1a)$$

$$E(e(x)) = 0, \quad \forall x \in \mathbf{R}, \quad (3.1b)$$

and

$$\frac{1}{2}E[(Z(x_1) - Z(x_2))^2] = \gamma(\|x_1 - x_2\|), \quad \forall x_1, x_2 \in \mathbf{R}. \quad (3.1c)$$

That is, within an area of size \mathbf{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 of the distance h between the measurement locations only. 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 locations of the area of size \mathbf{R}).

Depending on the sample variogram (Table A2, page 96), the form of the theoretical variogram $\gamma(h)$ was chosen from a nugget model, a spherical model, or a mixture of these two models (Table A3, page 97). 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. 3.3 shows the variograms for aluminium. The area size \mathbf{R} used was defined (in order of preference) as (i) the closest 20 observations when more than 20 observations were available within a distance of 100 km, (ii) 100 km when 10 to 20 observations

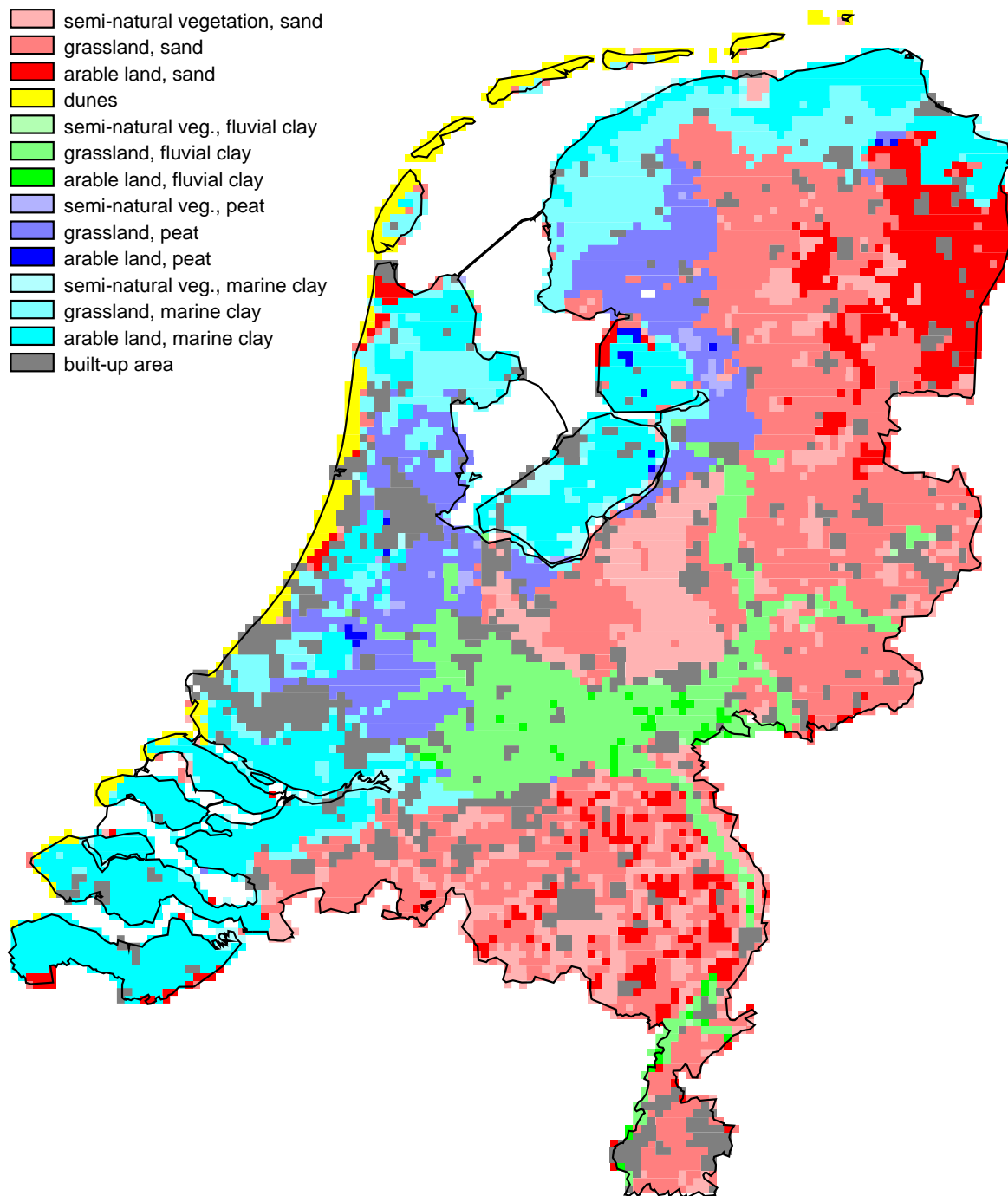


Figure 3.1 Soil type and land use map. Dominant soil-land use categories for 2 km × 2 km cells



Figure 3.2 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

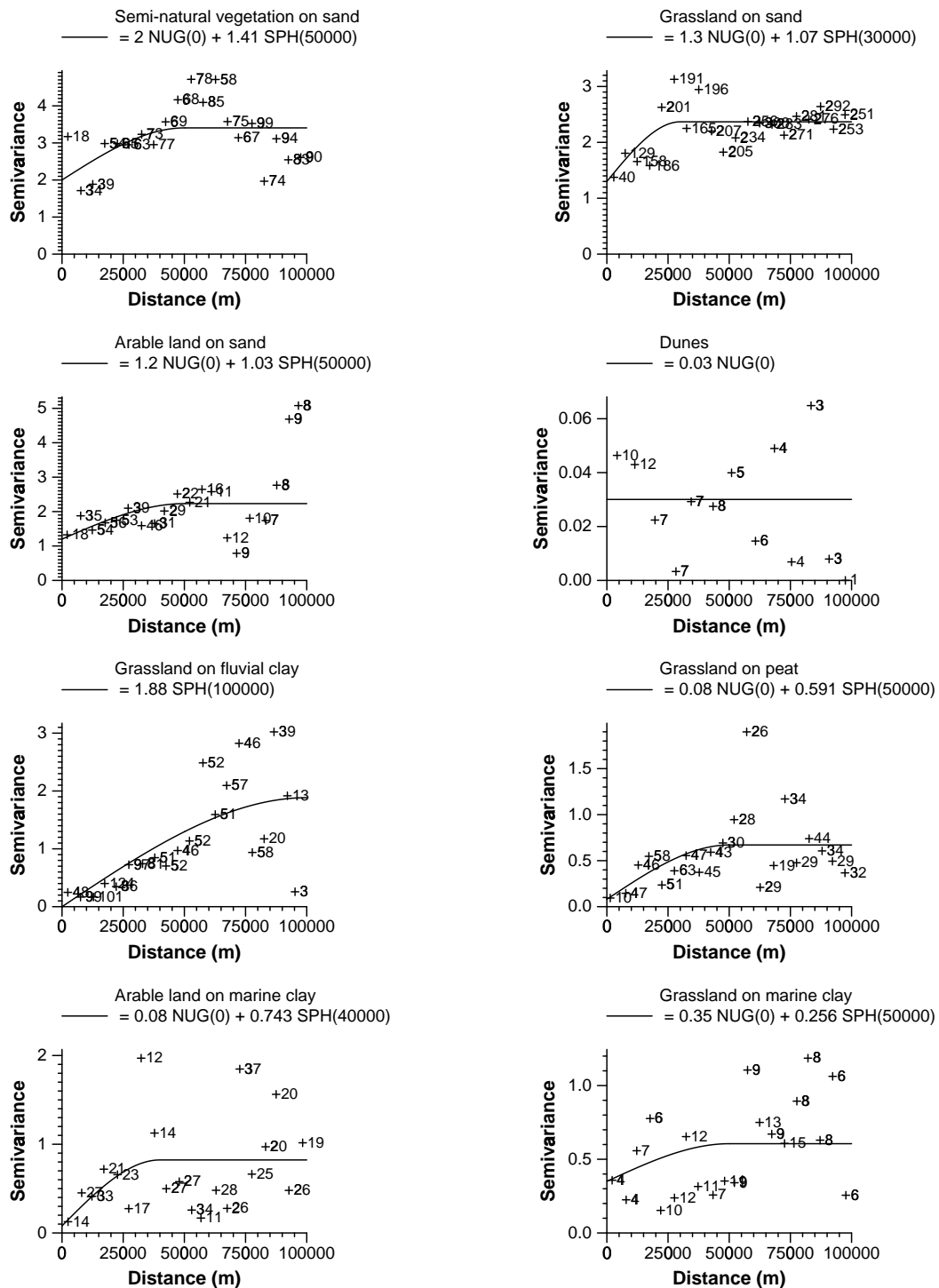


Figure 3.3 Variograms of $^e \log(AI)$, stratified by soil type and land use. A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97

were available within this distance, (iii) the 10 nearest observations.

The estimates given in the maps concern mean values on log-scale of $4 \text{ km} \times 4 \text{ km}$ blocks. Estimates are presented as approximate 95% confidence intervals calculated by

$$[\hat{z}(x_0) - 2\sigma_k(x_0), \hat{z}(x_0) + 2\sigma_k(x_0)] \quad (3.2)$$

with $\hat{z}(x_0)$ the point estimate (kriging estimate) on log-scale, and $\sigma_k(x_0)$ the estimation standard deviation (kriging standard deviation) of the estimate (Journel and Huijbregts, 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 $4 \text{ km} \times 4 \text{ km}$ blocks considered are centred at the cells of the $2 \text{ km} \times 2 \text{ km}$ grid of Fig. 3.1 (assuming that soil-land use dominance of $2 \text{ km} \times 2 \text{ km}$ cells still holds for the $4 \text{ km} \times 4 \text{ km}$ blocks). In this way two adjacent $4 \text{ km} \times 4 \text{ km}$ blocks have a 50% overlap, resulting in the finer spatial resolution of $2 \text{ km} \times 2 \text{ km}$ cells, while the limited estimation variance of $4 \text{ km} \times 4 \text{ km}$ block mean values remains relevant. The results have been presented in maps (e.g. see Figs. 3.4 and 3.5). Only the $2 \text{ km} \times 2 \text{ km}$ centres of blocks for which estimates were made are shown on the maps.

For the small soil-land use categories with less than 10 observations on 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 before. 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) (ii) the observations of the category from which the variogram was used 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, implying 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.

3.5 The groundwater quality maps

In the groundwater quality maps presented the 95% confidence intervals for $4 \text{ km} \times 4 \text{ km}$ 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 a block median value does not exceed a critical level, still many individual samples within the block may exceed it.

The maps are shown in two forms. In the base maps (the black-and-white maps, e.g. aluminium concentration in Fig. 3.4) 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*, based on the information available). In the colour maps (e.g. aluminium concentration in Fig. 3.5) four black-and-white sub-maps are merged into

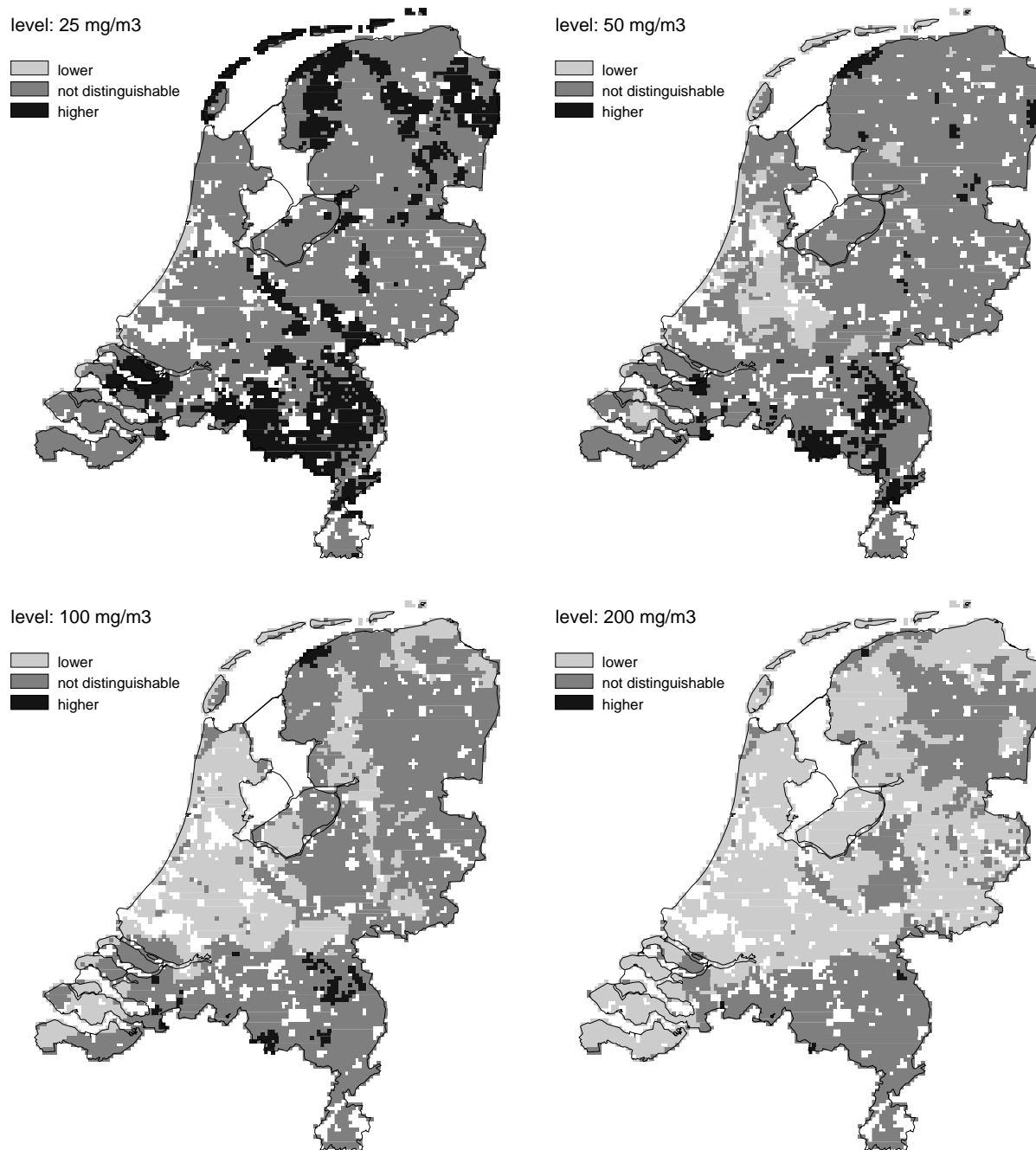


Figure 3.4 Map of aluminium in the groundwater. 95% Confidence intervals for 4 km × 4 km block median values related to four concentration levels

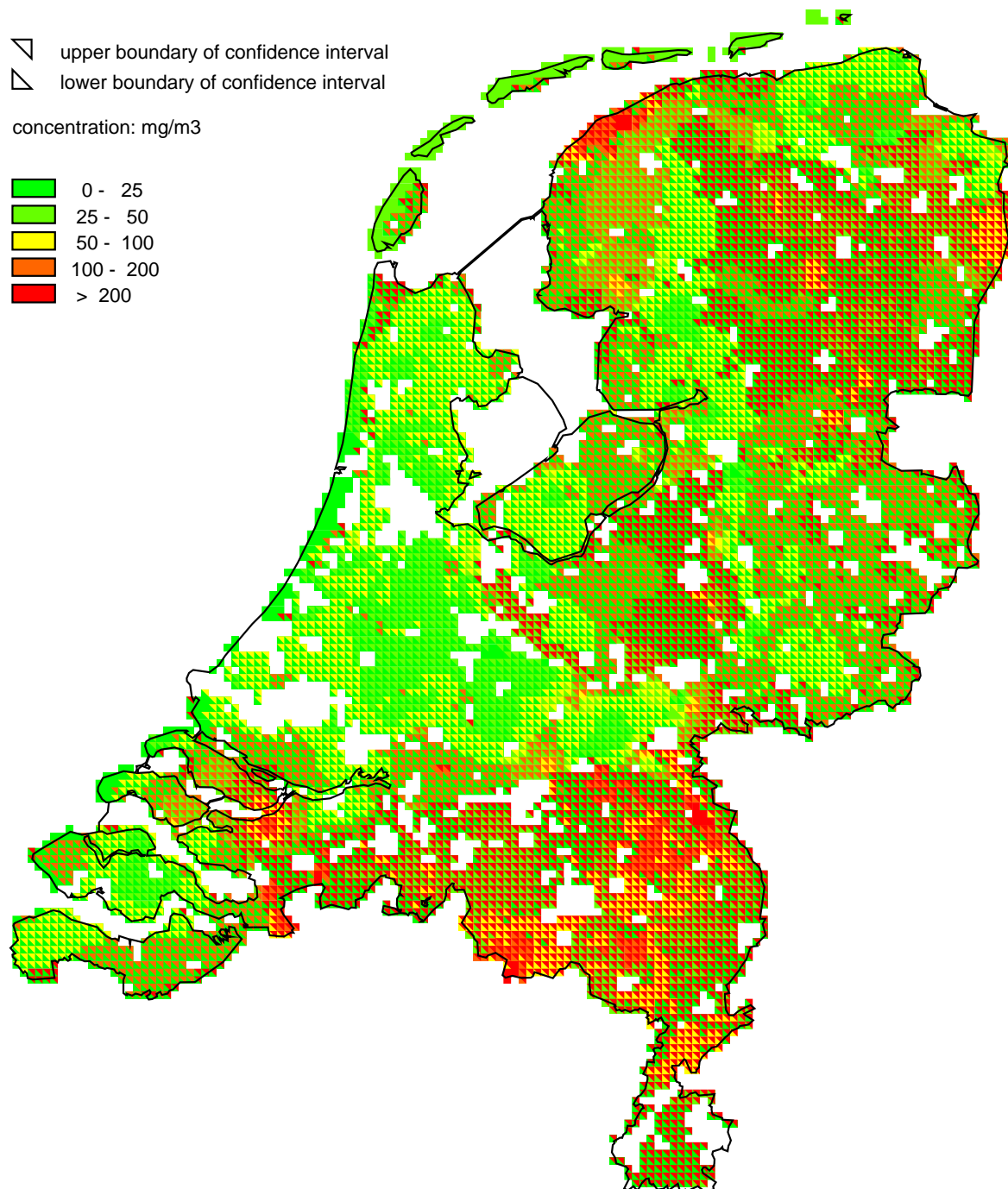


Figure 3.5 Map of aluminium in the groundwater (alternative display of Fig. 3.4). 95% Confidence intervals for 4 km × 4 km block median values related to four concentration levels

a single map (holding exactly the same information).

In general, the black-and-white maps are easier to read, but the colour maps immediately show the accuracy 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 considered. The doubling of the reference levels in the maps reflects the choice for equidistant levels on log-scale.

3.6 The effects of stratification

Groundwater quality maps can also be obtained while completely ignoring the geohydrological knowledge that resulted in stratification by soil type and land use. For nine groundwater quality variables (Table A4, page 97) the effects of stratification were investigated by making groundwater quality maps with partial stratification and without stratification. These maps are based on the same data as the main maps, using the same model (3.1) but now (a) for one category only (containing all 13 categories of Fig. 3.1), (b) for land use categories only, or (c) for soil type categories only. For aluminium the effects of no stratification are shown in Figs. 3.7 and 3.6, and the effects of stratification by land use only—ignoring soil type—are shown in Figs. 3.8 and 3.9. Figs. 3.6 and 3.9 only show the differences from the base map (Fig. 3.4). The results are summarized in Table 3.2.

In general it can be said that

- (i) The effects of omitting or partially applying stratification in this study 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 10 to 25% change in at least one of the four 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 like this (De Kwaadsteniet, 1990).
- (iv) The full stratification used for the base maps represents the most safe 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), one may prefer the corresponding map resulting from no (or partial) stratification.

3.7 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 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 A4, page 97). The results for

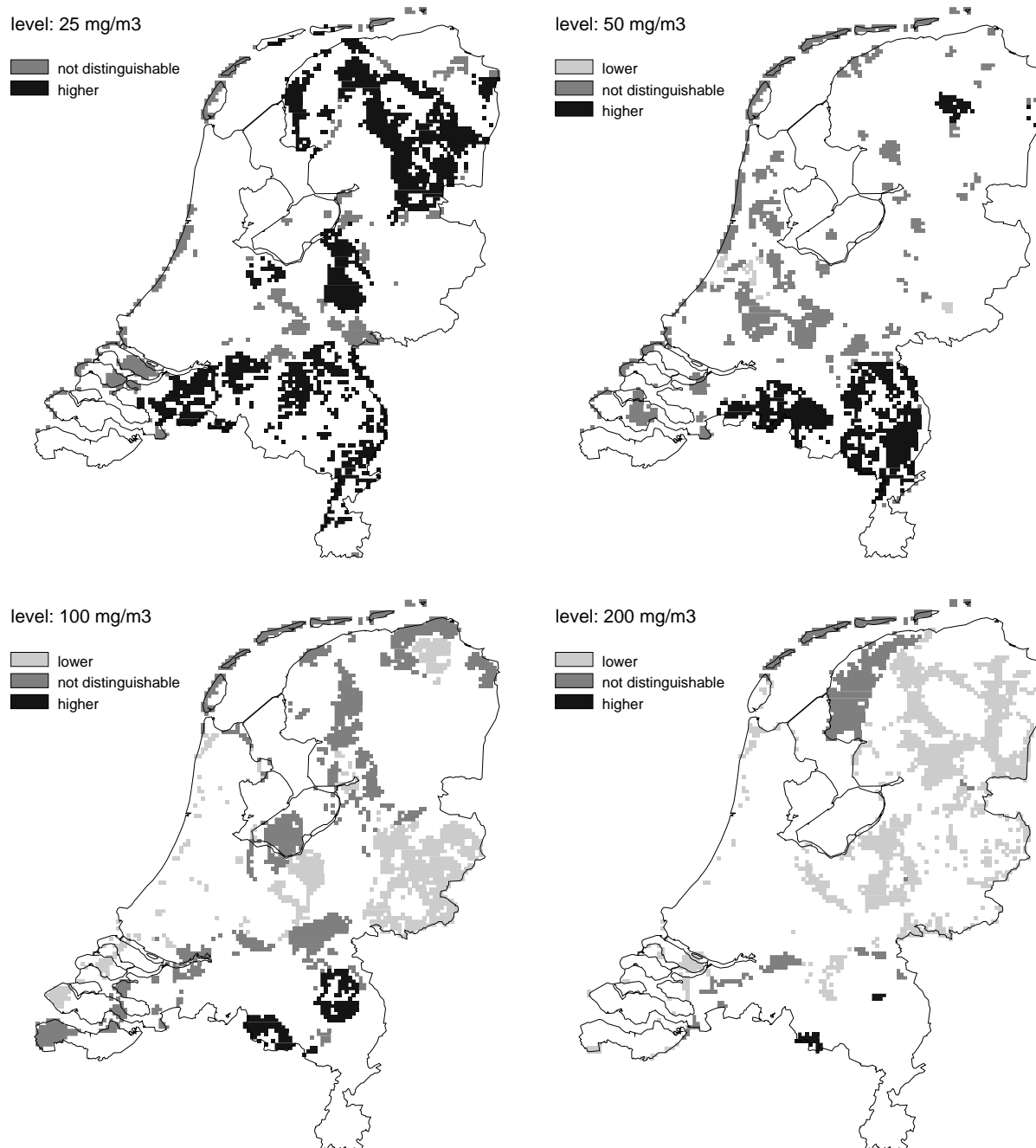


Figure 3.6 *The effects of no stratification for aluminium. 95% Confidence intervals for 4 km × 4 km block median values when soil and land use information is ignored, related to four concentration levels. Only results that differ from Fig. 3.4 are shown*

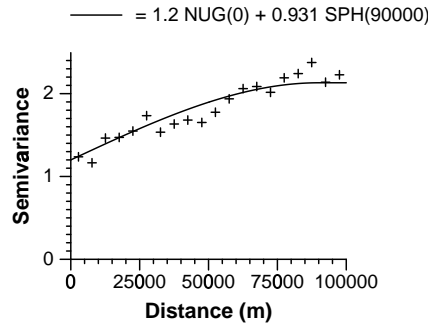


Figure 3.7 Variogram of $\log(AI)$, when stratification is omitted. Codes used for the variogram model (—) are explained in Table A3, page 97

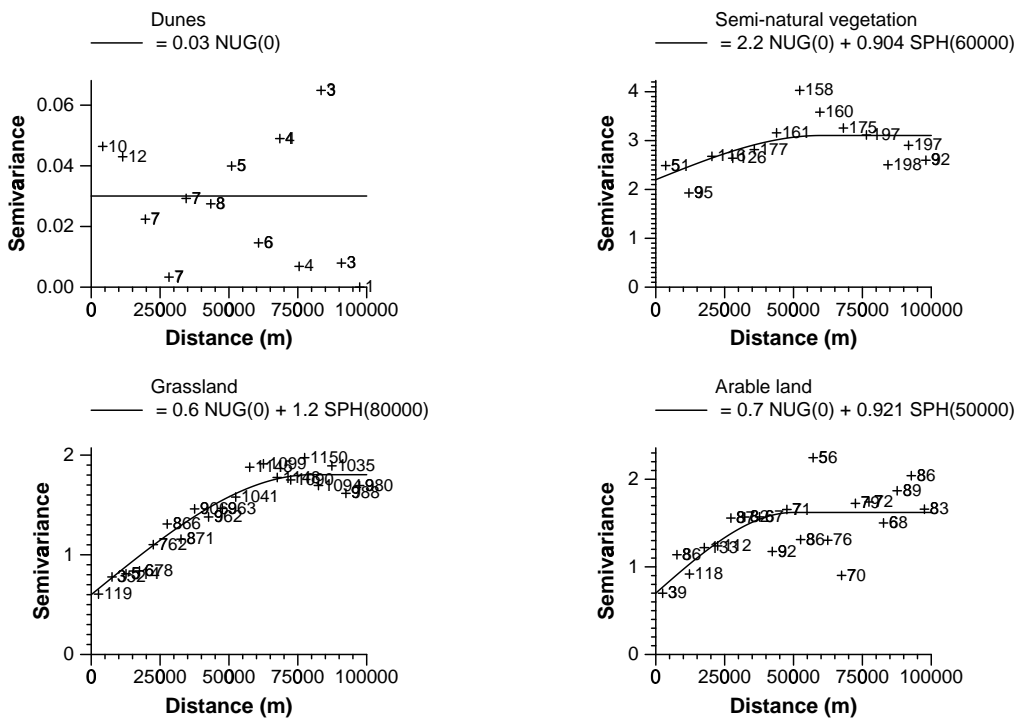


Figure 3.8 Variograms of $\log(AI)$, stratified by land use only. A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97

aluminium are presented in Figs. 3.10 and 3.11. Changes can be assigned to (i) loss of information about 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 Figs. 3.13 and 3.12 respectively. The results are summarized in Table 3.3.

It is evident that the extension of the NGM with the PGMs has resulted in important improvements of the groundwater quality maps (Fig. 3.11, ‘total’ column in Table 3.3). For a large

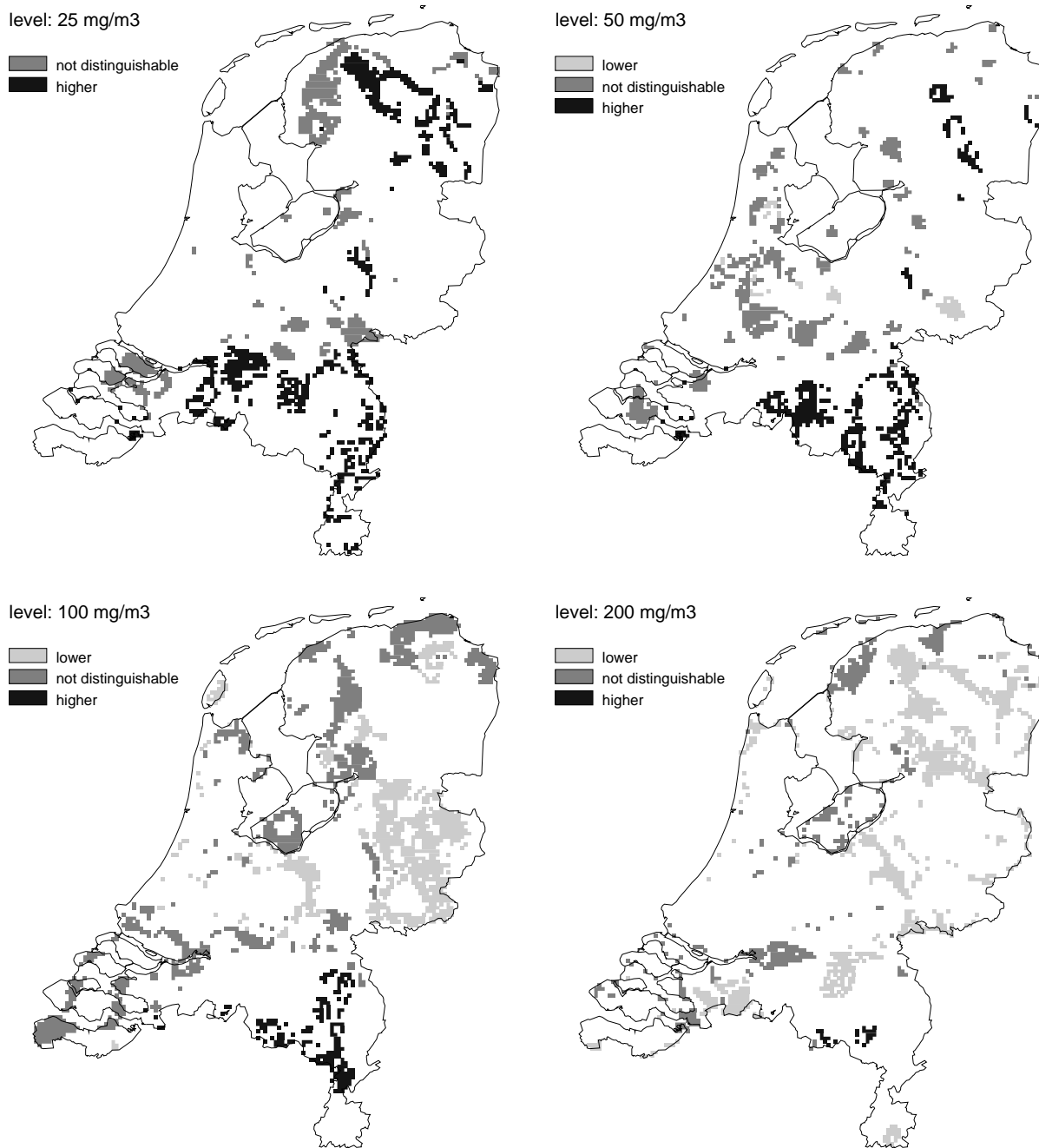


Figure 3.9 *The effects of stratification by land use only for aluminium. 95% Confidence intervals for 4 km × 4 km block median values when soil information is ignored, related to four concentration levels. Only results that differ from Fig. 3.4 are shown*

Table 3.2 *The effects of stratification for aluminium. 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 columns (b) and (c) are shown in Figs. 3.9 and 3.6 respectively*

level (mg/m ³)	number of changed cells as a result of omitting		
	land use	soil type	both
25	1276	1200	2227
50	975	1030	1668
100	959	1992	2314
200	825	1444	1898

Table 3.3 *The effects of monitoring network density for aluminium. Number of cells for which the value of a 95% confidence interval for the 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 about variograms (second order information), and/or (b) loss of information about the mean level and location specific information (first order information). Changes from each one of these components are also shown. Changed cells are shown in Figs. 3.11 (total), 3.12 (a) and 3.13 (b)*

level (mg/m ³)	number of changed cells		
	total	as a result of (a)	as a result of (b)
25	2765	2275	598
50	1649	1981	599
100	2643	2558	665
200	1828	1681	496

part these improvements result from the increase of spatial dependence information (Fig. 3.12 and column (a) in Table 3.3), especially from information about the short distance variation (compare Fig. 3.3 with Fig. 3.10). The effects of increased first order information also led to important improvements of the groundwater quality maps (Fig. 3.13 and last column of Table 3.3).

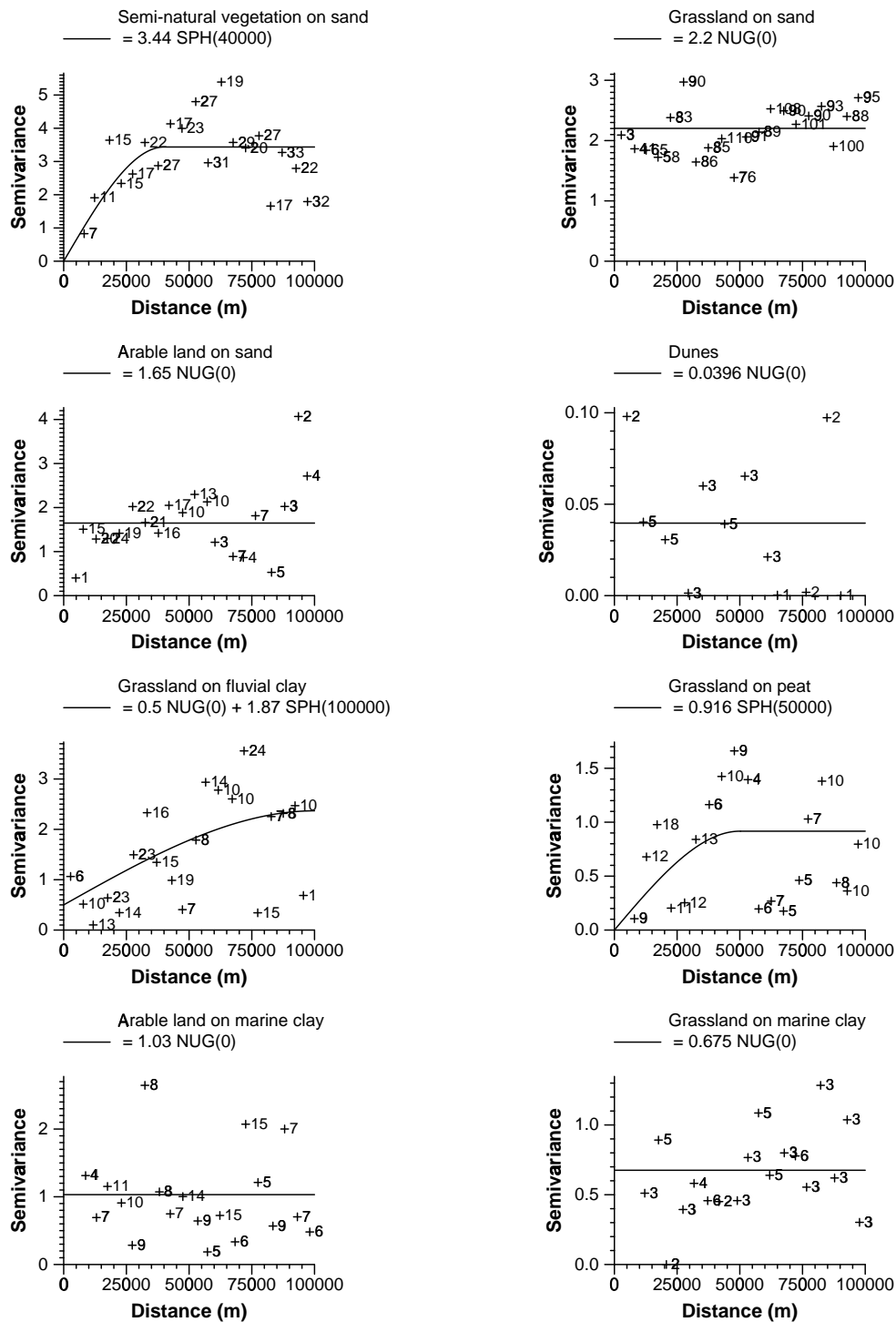


Figure 3.10 Variograms of $\ell \log(AI)$, stratified by soil type and land use, using only information from the national groundwater quality monitoring network. A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97

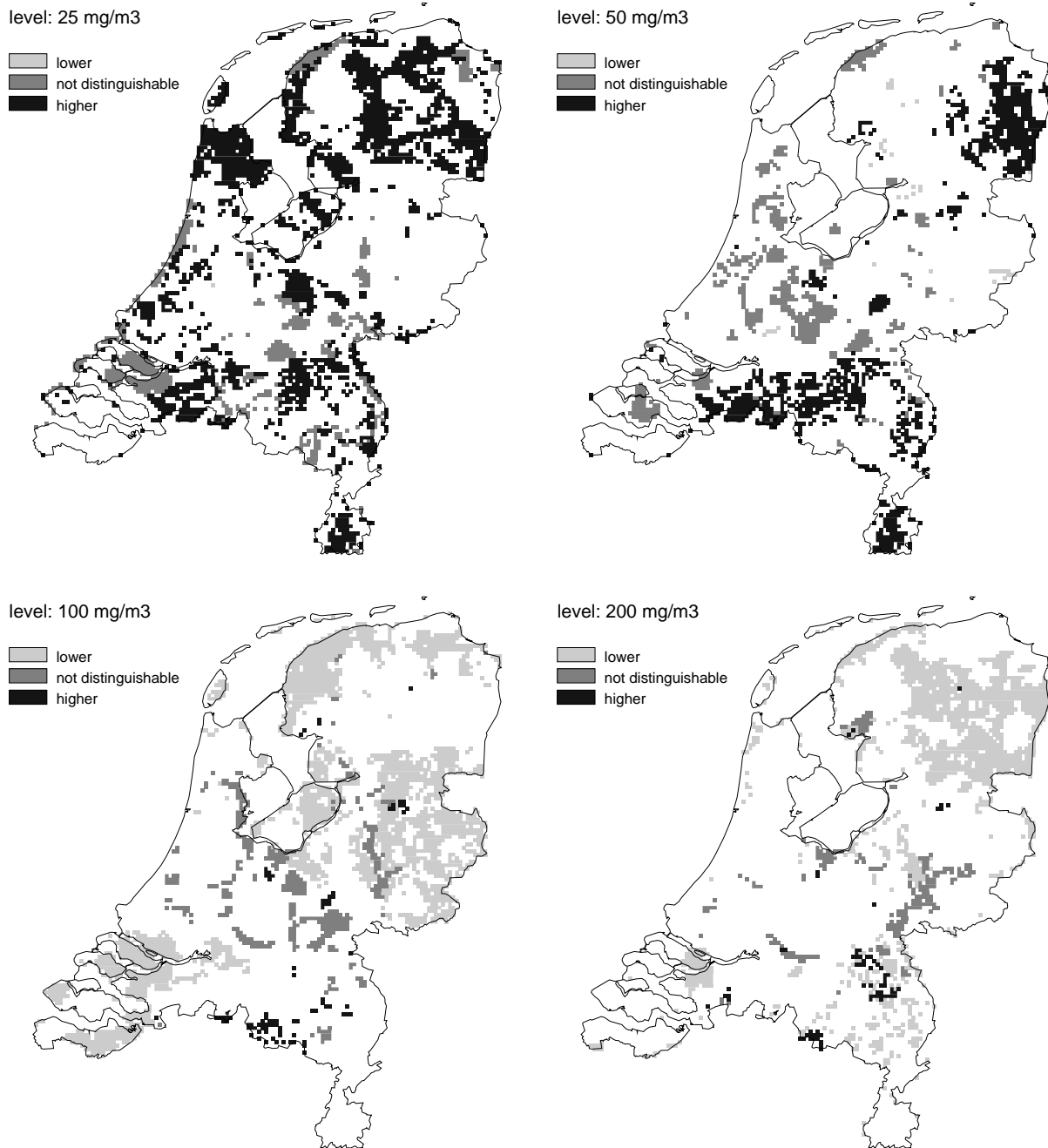


Figure 3.11 *The effects of network density for aluminium (total). 95% Confidence intervals for 4 km × 4 km block median values based on measurements from the national groundwater quality monitoring network only (provincial monitoring network information is left out), related to four concentration levels. Only results that differ from Fig. 3.4 are shown*

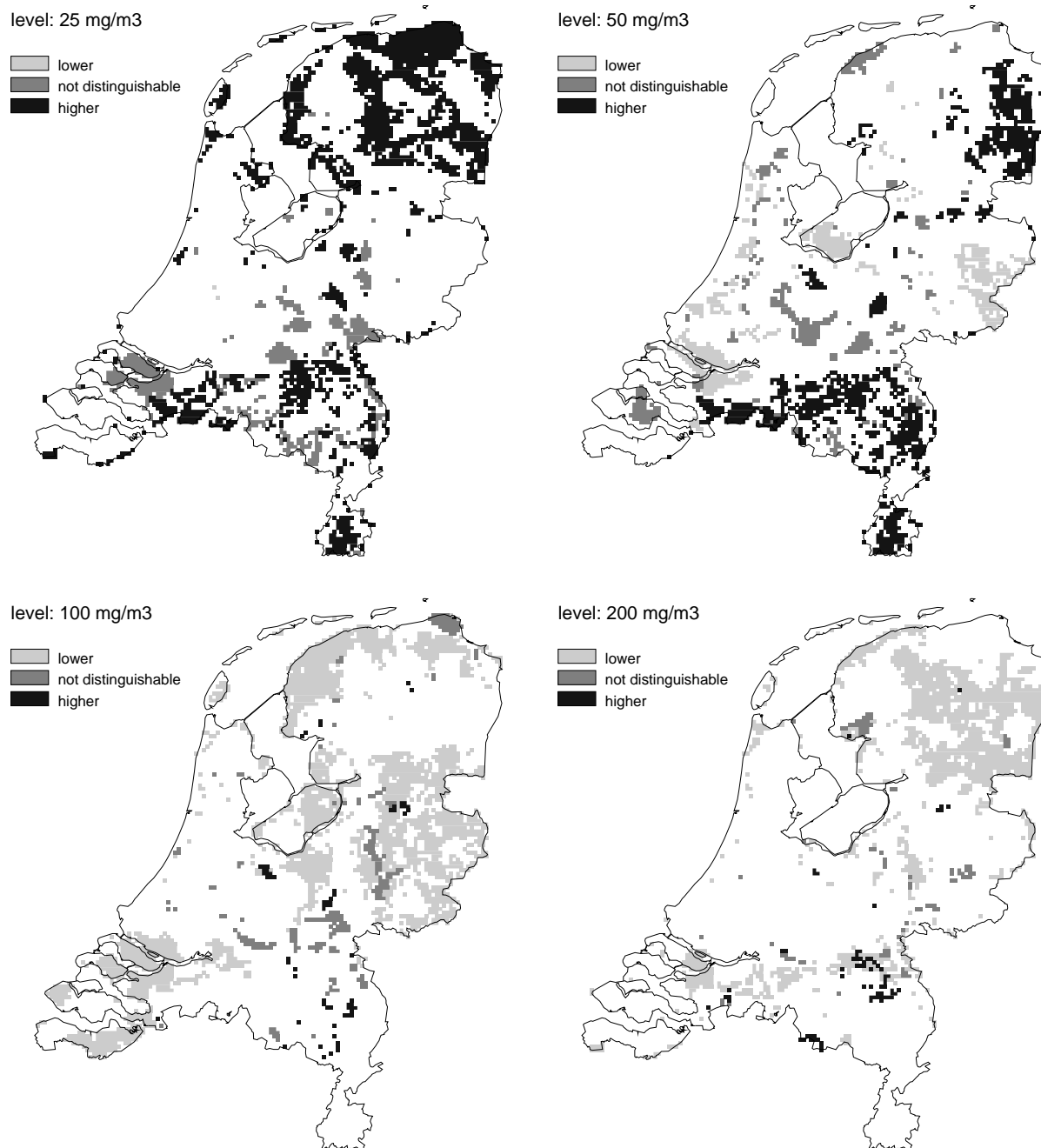


Figure 3.12 *The effects of network density for aluminium (second order information). 95% Confidence intervals for 4 km × 4 km block median values based on (i) first order information from both monitoring networks, and (ii) second order information from the national groundwater quality monitoring network only, related to four concentration levels. Only results that differ from Fig. 3.4 are shown*

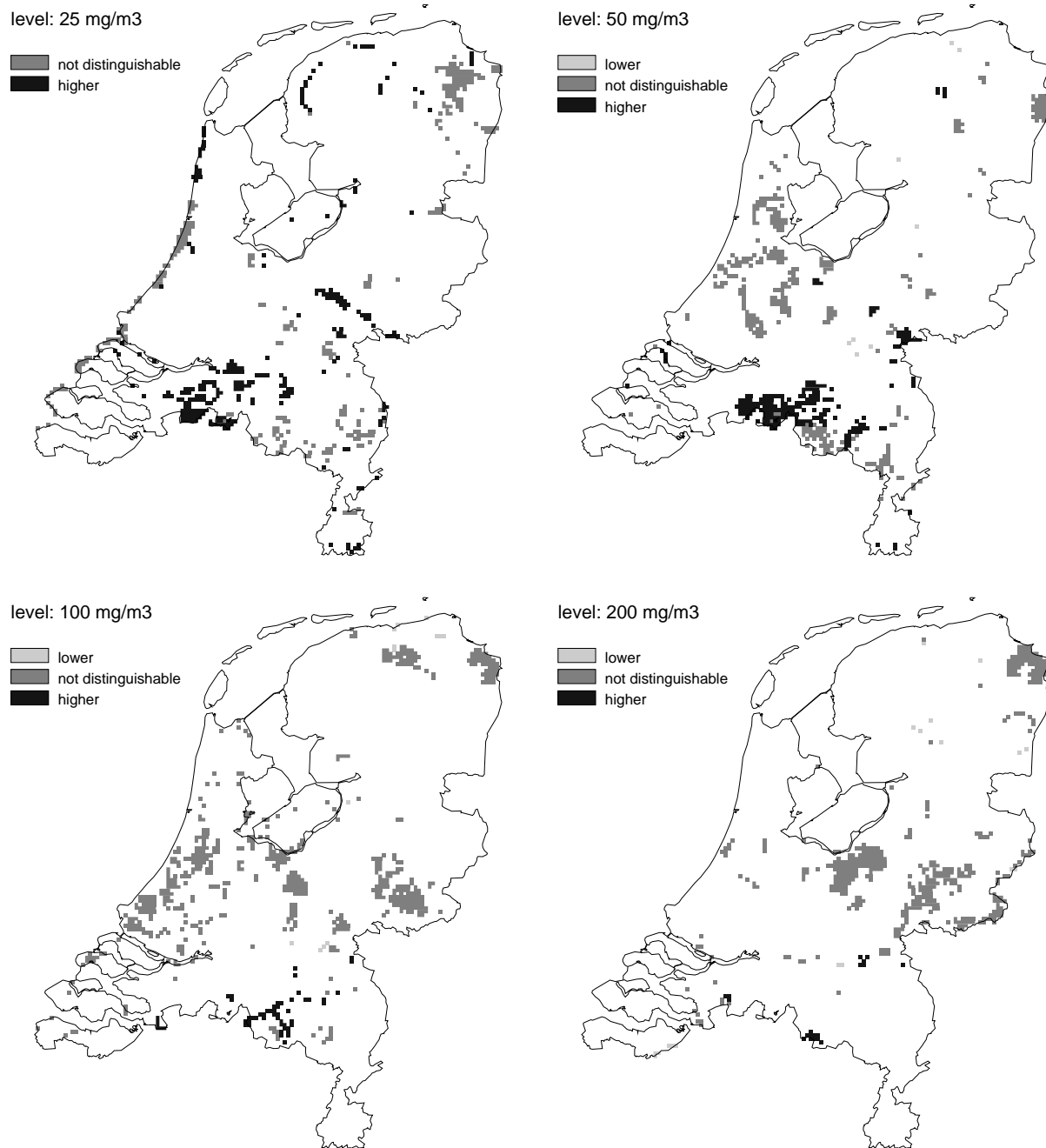


Figure 3.13 *The effects of network density (first order information). 95% Confidence intervals for 4 km × 4 km block median values based on (i) first order information from the national groundwater quality monitoring network only, and (ii) second order information from both networks, related to four concentration levels. Only results that differ from Fig. 3.4 are shown*

3.8 Discussion

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 where 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, monitoring network optimization 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 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 at the measurement size into spatial averages the size of spatial units, as 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: because of the spatial variation of soil type and land use, 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 (e.g. estimation of the fraction of individual samples in a region exceeding a critical level in the tail of the sample distribution). Provided that this latter approach includes both spatial and 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 it 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. 3.3.

More generally it can be said that the current approach closely follows a classical statistical

approach: we start with a stratification by soil type and land use. It improves on a purely classical statistical approach by explicitly taking account of location specific information and spatial dependence, by using stratified kriging. The effects of monitoring network density clearly demonstrate the added value of using location specific information and spatial dependence: especially the effect of short distance variation but also the effect of local information results in important modifications of the groundwater quality maps.

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 sub-optimal 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 relevant geohydrological attributes, of which no maps are available. When maps of relevant geohydrological attributes (e.g. on infiltration versus seepage) become available, it might be worthwhile to include them in the stratification, which would result in fewer measurements per category, but with smaller within-category variation more accurate estimates may be obtained.

Mathematical aspects were kept simple, primarily because we aim for (a) an optimal accessibility of the various methodological items mentioned, and (b) the avoidance of more complex assumptions than necessary, based on the objective of this study and the information available. The simplicity refers to the model used (3.1), the choice and fit of variogram models, the choice of \mathbf{R} , the way approximate confidence intervals for block median values were calculated, and the outlier procedure. Some general remarks about the mathematical aspects are:

- (i) Kriging as applied in this study (using model (3.1) for each variable for each soil-land use category) can be considered as a first step beyond and an improvement of the classical statistical approach (for each variable for each category) in its four forms based on spatial dependence or independence, and one mean per category or local means of regions within a category, respectively. The kriging approach is an improvement on these classical statistical approaches because local information is used more effectively.
- (ii) The (approximate) 95% confidence intervals presented in this study are consistent with interpolations on a log-scale: an obvious choice given the character of the measurements at hand. For a first quantification of uncertainties (the calculation of confidence intervals using (3.2)) the estimation error of block mean values on log-scale is assumed to be normally distributed, and uncertainty of the (kriging) standard deviation is ignored. This implies that only a rough indication of confidence is obtained, and that the width of the confidence intervals should be considered as the lower bound to more carefully obtained confidence intervals.
- (iii) 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_{\square} and block median M_{\square} , given by

$$m_{\square} = M_{\square} \exp(s_{\square}^2 / 2), \quad (3.3)$$

where s_{\square}^2 is the short distance variance, the short distance value of the variogram for the relevant variable and category. Relation (3.3) 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 \times 4 km block mean values are obtained when the legend class boundaries of a map are multiplied with the factor of M_{\square} from (3.3). For a groundwater quality variable, this will result in different legend class boundaries for different soil-land use categories.

Chapter 4

Maps of temporal changes in the groundwater quality in the Netherlands at 5-17 metre depth

(from: Pebesma and De Kwaadsteniet, 1995)

4.1 Introduction

The primary goal of the national groundwater quality monitoring network is to provide insight in the current situation and systematic changes in time of groundwater quality. For this purpose, about 370 monitoring sites, spread fairly evenly over the country, are sampled yearly at two depths (from well screens at about 10 and 25 metre below surface). At first samples were analysed for 19, and later 25, groundwater quality variables.

The current study aims at mapping groundwater quality variables, using point measurements obtained from the monitoring networks, in such a way that uncertainties in estimates are quantified statistically. In particular, the attention is focused on the description of changes in groundwater quality in time. Mapping groundwater quality variables at a measurement moment (in one year, 1991) was described in Chapter 3.

Because both starting level and time gradient are significant to the assessment of changes in groundwater quality, we chose to examine this subject in terms of short-term predictions. Using a simple regression model, extrapolation from observed time series yields short-term predictions ('expected observations') for the monitoring screens, which can be used in the spatial interpolation. Time-prediction uncertainties, expressed as estimation variances of 'expected observations' can be incorporated in the spatial interpolation.

Basically, the procedure for mapping 'expected observations' is similar to the procedure for mapping groundwater quality at a measurement moment (Chapter 3). As before, maps of soil type (Steur et al., 1985) and land use (Thunnissen et al., 1992) were aggregated to a set of soil-land use categories. For each category, for every groundwater quality variable, and for each extrapolation time attention was paid to the time-prediction uncertainty of, the level of, the variation in and the spatial dependence in 'expected observations' by using a variant of ordinary kriging (Journel and Huijbrechts, 1978) that was modified to account for non-constant time prediction uncertainties (Delhomme, 1978). For each extrapolation time, estimates of 4 km × 4 km block median values were presented on maps as approximate 95% confidence intervals.

It was not possible to create maps of changes in groundwater quality for all variables or to

use all measurements that were used for the maps of 1991 (Pebesma and De Kwaadsteniet, 1994, Table A1, Fig. 3.2) because (i) many measurement sites have been installed only recently, and (ii) for the remaining sites the monitoring of several groundwater quality attributes was stopped or started only recently. Therefore, this study concerns only 12 groundwater quality variables and is based on a much smaller set of observations: the measurements of the provincial groundwater quality monitoring networks had to be left out of this study. Full results are presented in a separate report (Pebesma and De Kwaadsteniet, 1995). In this chapter we illustrate the methods used with results on one groundwater quality variable, namely potassium concentration.

The maps of changes in groundwater quality can be used for (i) assessment of diffuse groundwater contaminations (because the measurements do not represent strictly local contaminations) (ii) the validation of policy-supporting regional or national groundwater quality models.

Sections 4.2, 4.3, 4.4 and 4.5 describe the stratification, the data selection, the changes in time at the well screens and the spatial interpolation. The results are discussed in Section 4.6 and this chapter concludes with a discussion in Section 4.7.

4.2 An aggregated map comprising soil type and land use

The soil-land use map that was used for the stratification and classification of the measurements is identical to the map described in Section 3.2, shown in Fig. 3.1 (page 26). Table 3.1 presents basic data on the categories.

4.3 Data selection

For the generation of maps of changes in groundwater quality several observations were left out because they seemed unsuitable for this purpose. As in Section 3.3, monitoring sites in 'built-up areas' were left out, as well as monitoring sites that were influenced directly by water from rivers or the sea.

Because the current study concentrates on the recognition of meaningful, steady, long-term components of the changes in time, the selection only comprises twelve groundwater quality variables for which long-term series were available (Table A5, page 97). As in the previous chapter, only monitoring screens between 5 and 17 metre below the ground surface were considered.

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 ('85 and '86) did not fit in the general pattern of the series for a substantial fraction of screens, leading to doubts about the quality of these measurements (the quality of sampling, sample treatment or analysis). We chose for a 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. Examination of the time series plots furthermore revealed that for some samples a considerable part (more than 3) of the variables did not fit in the series, which was interpreted as an indication of erroneous sampling or sampling treatment. These samples were considered as outliers, and they were left out altogether.

(Three samples were marked as outlier.)

Detection limit values were assigned to measurements below the detection limit, for each variable using the highest detection limit when a change in detection limit had occurred during the measurement period considered. In exceptional cases, when only a tiny fraction of the measurements were determined with a considerably higher detection limit, measurements below such a high detection limit were treated as missing values.

The measurements were finally subjected to a global quality control, in which, for each variable and monitoring screen, measurements that differed more than a factor 10 from the other measurements of the time series were considered as ‘doubtful’ and were left out of the selection. Finally, after applying all previously described selections, only the measurements from monitoring screens with at least 5 measurements in the period considered were used for mapping changes in groundwater quality.

The locations of the 250 monitoring screens used for mapping changes in groundwater quality are shown in Fig. 4.1. The number of measurements in each soil-land use category is shown in the third column of Table 3.1 (except for the category ‘grassland on peat,’ which has 22 measurements due to the additional selection criteria described above).

4.4 Changes in groundwater quality variables at the well screens

Because starting level and time gradient are both significant to the assessment of changes in groundwater quality, we chose to examine the subject of changes in groundwater quality in terms of short-term predictions. In particular, we are interested in the assessment of meaningful, steady (i.e. monotonous), long-term components in the change in time of groundwater quality variables.

For each groundwater quality variable and for each separate monitoring well screen, the variation of a concentration variable on the log-scale in time $y(t)$ is represented by a linear model with time, having independent, identically distributed errors

$$Y(t) = \alpha_0 + \alpha_1 t + e(t), \quad E(e(t)) = 0, \quad \text{Cov}(e(t)) = \sigma^2 I, \quad (4.1)$$

which is a special case of the linear model (2.1), written as $Y(t) = F_t \beta + e(t)$ when $f(t_i) = (1, t_i)$, $i = 1 \dots n$, is the i -th row of F_t , and $\beta = (\alpha_0, \alpha_1)'$. Thus, estimate and estimation variance for the mean value of a short-term prediction at extrapolation time t_{new} , given $\hat{\beta}$ (2.2a), using (2.4) and (2.5), are

$$\hat{y}(t_{new}) = f(t_{new}) \hat{\beta} = \hat{\alpha}_0 + \hat{\alpha}_1 t_{new} \quad (4.2)$$

and

$$\sigma^2(t_{new}) = f(t_{new})(F_t' F_t)^{-1} f(t_{new})' \sigma^2, \quad (4.3a)$$

and $\sigma^2(t_{new})$ is estimated by $s^2(t_{new})$ when, using (2.3),



Figure 4.1 *Monitoring network locations. Sites of the national groundwater quality monitoring network that contributed to the maps of changes in groundwater quality*

$$s^2 = y(t)'(I - F_t(F_t'F_t)^{-1}F_t')y(t)/(n - 2) \quad (4.3b)$$

is substituted for σ^2 in (4.3a). Using model (4.1) implies that on a log-scale we assume a constant increase or decrease of concentration in time: a first approach for a trend in which the deviation from a straight line is considered as ‘white noise’. Using the log-transformation not only guarantees predictions that are strictly positive on the original scale, it also yields predictions at the monitoring network sites on the most obvious scale for spatial interpolation (Section 2.7, Chapter 3).

For creating maps of changes in groundwater quality the short-term predictions from (4.2) were taken as ‘new observations’ at the monitoring network sites. These ‘new observations’ were interpolated, and location specific observation accuracy was derived from (4.3a) and (4.3b). More specifically, for two moments (i.e. time periods) 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 two ‘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 accuracies for both extrapolation moments (Fig. 4.2). The aim of choosing an extrapolation span of 1980-2000 is to emphasize trends that are visible despite the observation inaccuracies.

Quantifying the observation accuracy with (4.3a) implies that the ‘new observations’ only relate to the situation *around* 1980 and *around* 2000: for ‘new observations’ the short-term components of temporal variation, the year-to-year deviation from the regression line, are filtered out, within the context of the current model structure.

For a monitoring well with two screens in the depth interval considered, each ‘new observation’ used for the spatial interpolation of a groundwater quality variable for an extrapolation moment was calculated as a simple average of both ‘new observations’ $\hat{y}_1(t_{new})$ and $\hat{y}_2(t_{new})$

$$\hat{y}(t_{new}) = (\hat{y}_1(t_{new}) + \hat{y}_2(t_{new}))/2.$$

The estimated observation accuracy for such a monitoring well was calculated as

$$s^2(t_{new}) = (s_1^2(t_{new}) + s_2^2(t_{new}))/4,$$

where $s_1^2(t_{new})$ and $s_2^2(t_{new})$ were derived from (4.3a) and (4.3b).

4.5 Spatial interpolation

Maps of short-term predictions for groundwater quality variables were obtained by means of ordinary block-kriging of log-transformed data (Journel and Huijbregts, 1978) within each of the categories in Table 3.1, using the short-term predictions obtained for every well in a category: the ‘new observations’ obtained from (4.2) at $t = 1980.5$ and $t = 2000.5$.

For the spatial interpolation a model concept was adopted which basically considered observations of a mapping variable within an area \mathbf{R} inside a soil-land use category as a sample

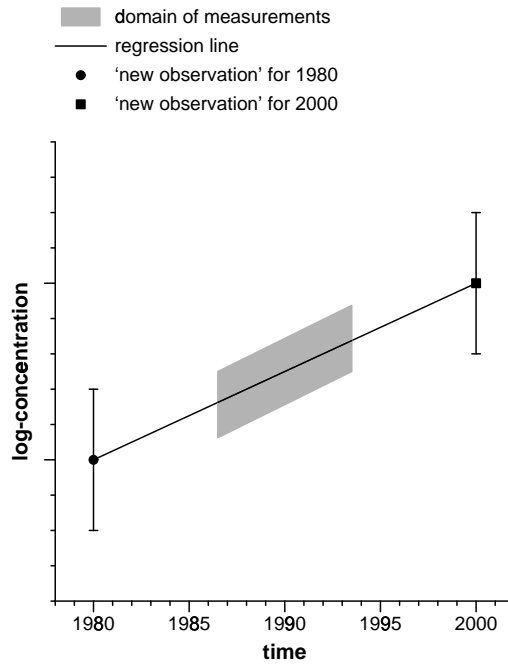


Figure 4.2 Schematic representation of ‘new observations’ and their accuracy (as error bars) at a single monitoring well screen (no actual measurements drawn)

from a realization of the random field $Z(x)$ with the following properties:

$$Z(x) = m + e(x) \tag{4.4}$$

with

$$m \text{ constant,} \tag{4.4a}$$

$$E(e(x)) = 0, \quad \forall x \in \mathbf{R}, \tag{4.4b}$$

and

$$\frac{1}{2}E[(Z(x_1) - Z(x_2))^2] = \gamma(\|x_1 - x_2\|), \quad \forall x_1, x_2 \in \mathbf{R}. \tag{4.4c}$$

That is, within an area of size \mathbf{R} inside a soil-land use category we used an additive model (on log-scale) with (i) a constant expected value and (ii) spatial dependence between measurements that is a function of the distance h between the measurement locations only. Additionally we assumed that in a soil-land use category for a groundwater quality variable only one function $\gamma(h)$ is needed (independent of \mathbf{R}), and that this function (the ‘variogram’) is identical to the variogram that was obtained from the measurements of 1991 (Section 3.4, compare Fig. 3.3), for the corresponding soil-land use category. Fig. 4.3 shows these variograms for potassium, as obtained from the 1991 measurements (Pebesma and De Kwaadsteniet, 1994).

The area size \mathbf{R} was defined (in order of preference) as (i) the closest 20 observations

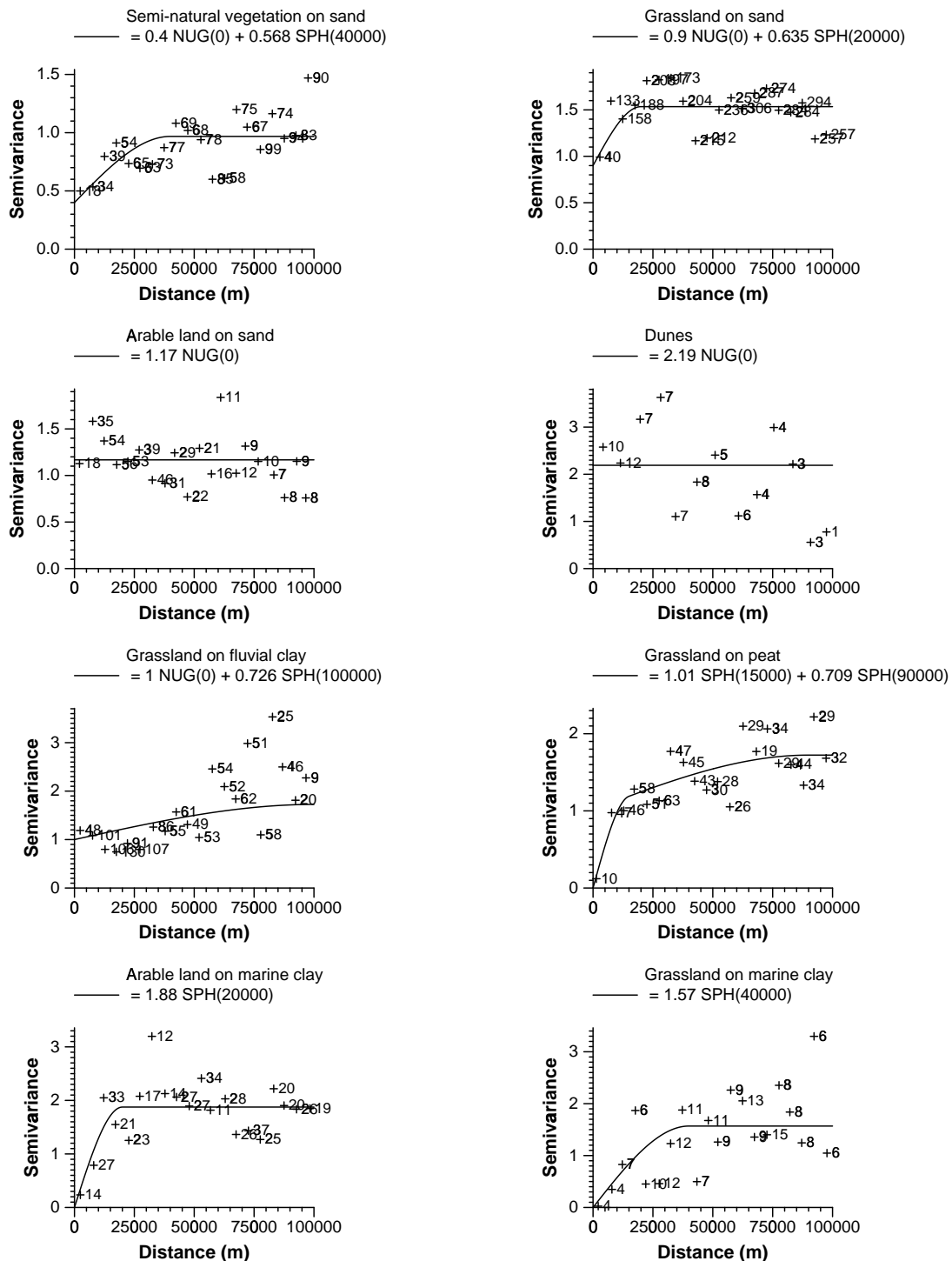


Figure 4.3 Variograms of $\ell \log(K)$, stratified by soil type and land use, obtained from the 1991 measurements (national and provincial monitoring networks). A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97

when more than 20 observations were available within a distance of 100 km, (ii) 100 km when 10 to 20 observations were available within this distance, (iii) the 10 nearest observations. The interpolation procedure that was used here differs from standard block-kriging based on (4.4) (Journel and Huijbregts, 1978) in that location specific uncertainty for each ‘new observation’ obtained with (4.3a) and (4.3b) was taken into account, as described in Section 2.5 (2.12), and Delhomme (1978).

The estimates concern mean values on log-scale of $4 \text{ km} \times 4 \text{ km}$ blocks. Estimates are presented as approximate 95% confidence intervals calculated by

$$[\hat{z}(x_0) - 2 \sigma_k(x_0), \hat{z}(x_0) + 2 \sigma_k(x_0)] \quad (4.5)$$

with $\hat{z}(x_0)$ the point estimate (kriging estimate) on the log-scale, and $\sigma_k(x_0)$ the corresponding estimation standard deviation (kriging standard deviation), using the model described above (Delhomme, 1978, Journel and Huijbregts, 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.

The display of values for $4 \text{ km} \times 4 \text{ km}$ blocks on the $2 \text{ km} \times 2 \text{ km}$ block centres, and the treatment of small soil-land use categories with less than 10 observations is identical to that of Section 3.4.

4.6 Results

In the maps of changes in groundwater quality, the 95% confidence intervals for $4 \text{ km} \times 4 \text{ km}$ block median values are related to four reference levels, as obtained from multiples of target levels or critical levels for human consumption (these target levels or critical levels are defined in relation to individual samples). For each reference level a sub-map shows where the confidence intervals are completely above or below the level, or where they straddle the level. In the latter case, block median values and reference level cannot be distinguished, based on the information available. The doubling of the reference levels in the maps reflects an equidistant (linear) choice of levels on the log-scale. For each variable changes in groundwater quality are revealed by showing maps for 1980 and 2000 side by side, and by only showing their mutual differences—where the maps are identical, no values are shown. Figs. 4.4 and 4.5 show the maps of changes in potassium concentration.

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 slope of all or part of the regression lines, as fitted to the measured time series: if the slopes are zero the maps are 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, trend magnitude cannot be derived from this shift, for instance, in case of a single legend class shift. It can be stated that all trend indications shown are of vital importance in the sense that they concern a shift in estimates of $4 \text{ km} \times 4 \text{ km}$ block median concentration that results in a different map for at least one of the reference levels.

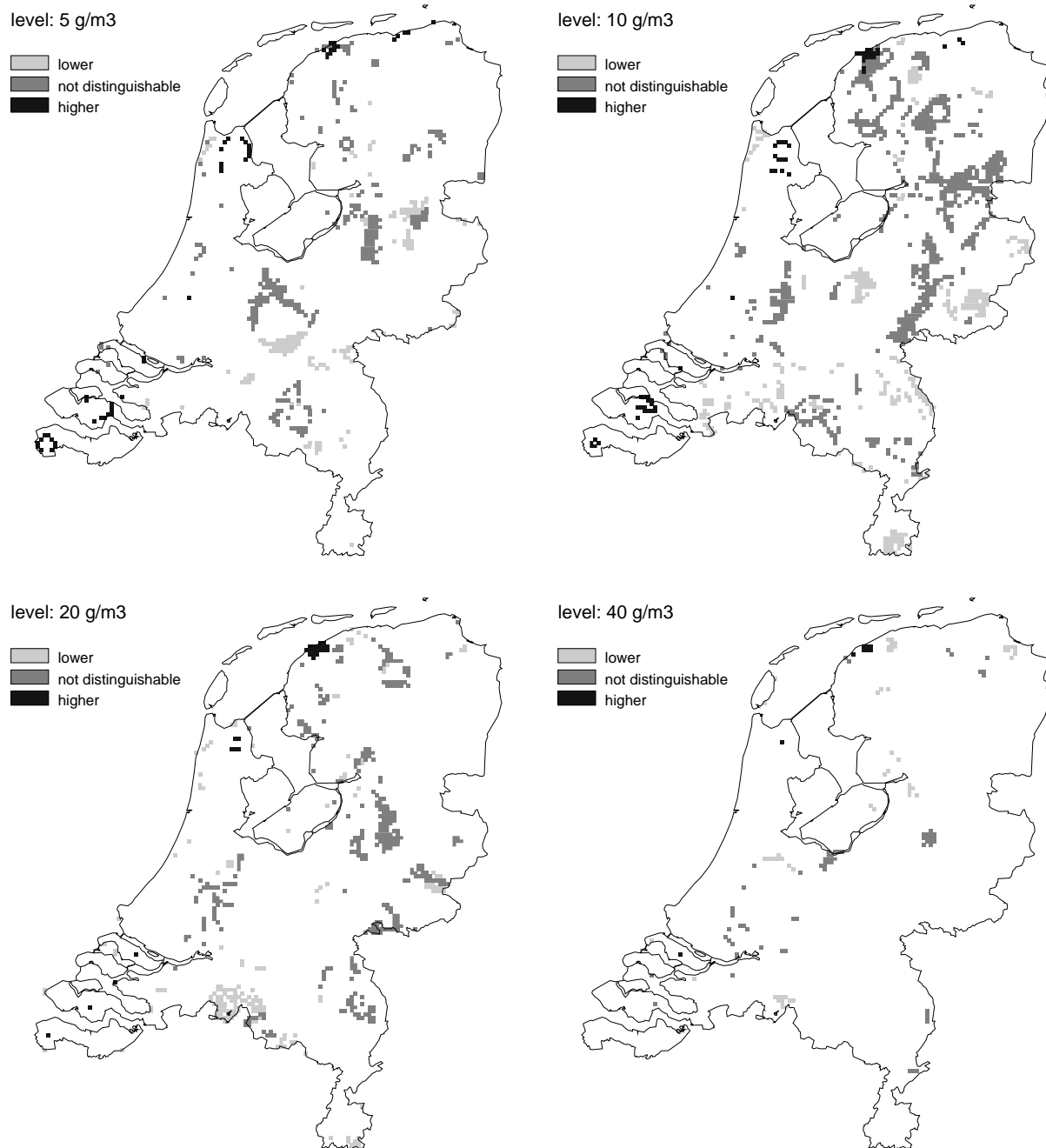


Figure 4.4 Map of potassium in the groundwater around 1980, where it differs from Fig. 4.5 ('around 2000'). 95% Confidence interval for $4 \text{ km} \times 4 \text{ km}$ block median values, related to four concentration levels

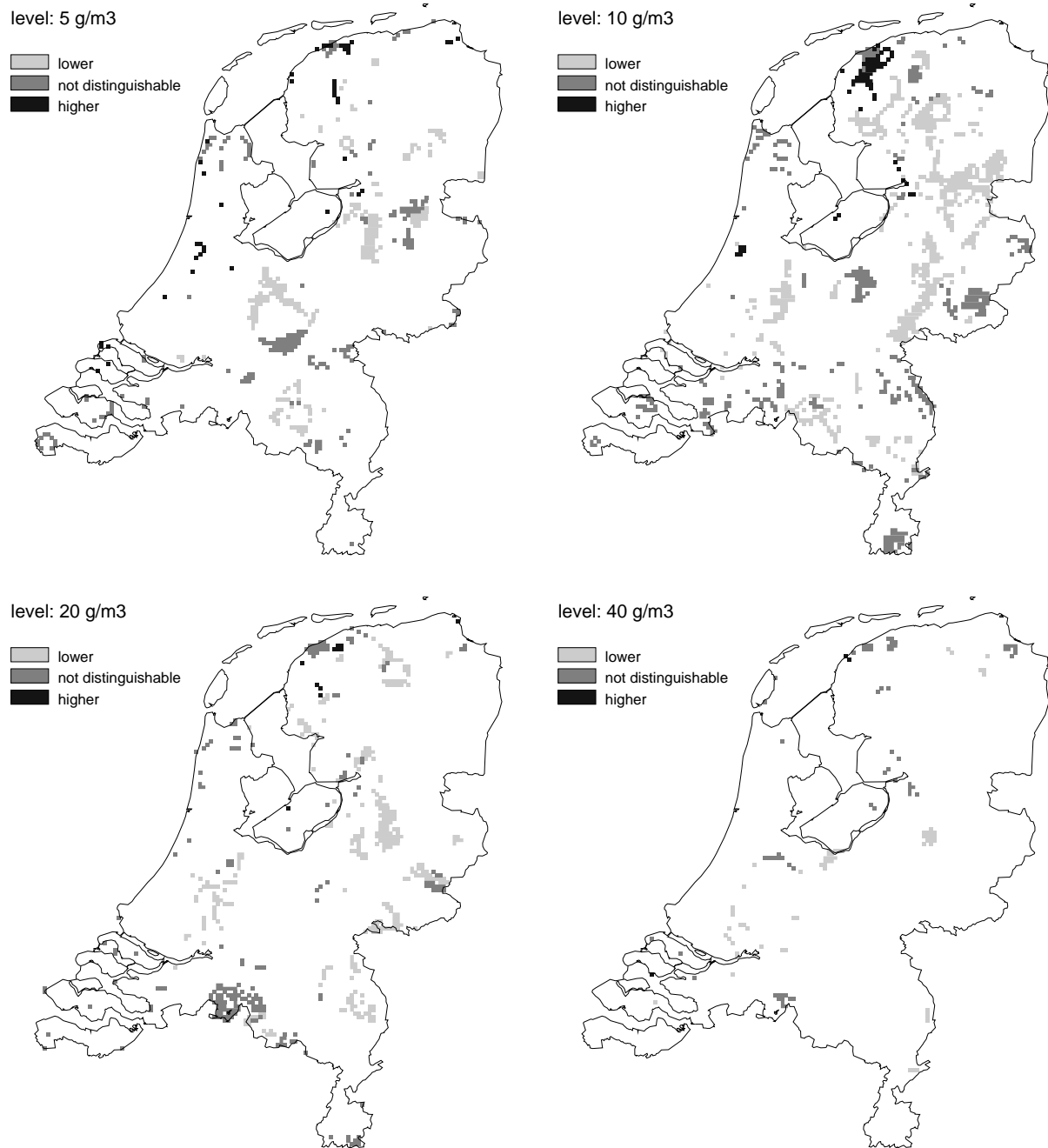


Figure 4.5 Map of potassium in the groundwater around 2000, where it differs from Fig. 4.4 ('around 1980'). 95% Confidence interval for 4 km × 4 km block median values, related to four concentration levels

The maps (Figs. 4.4 and 4.5; 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) cells with a pronounced trend can be distinguished from a neighbourhood with less pronounced trends, (ii) trends in adjacent or nearby cells (also within the same soil-land use category) can have opposite directions.

4.7 Discussion

The changes in predicted groundwater quality as presented on the maps arise from (i) the measurement series from 250 monitoring sites and (ii) the model structure used. Some further comments belong here.

For the description of temporal change in a groundwater quality variable we used a first order linear regression model for measurements on the log-scale with time, assuming that the deviation from a regression line could be considered as ‘white noise’.

Using the log-scale as the working scale yields (i) temporal predictions on the most obvious scale for spatial interpolation (Chapter 3), (ii) a uniformly quantified location specific measure of accuracy for the prediction at the monitoring sites, directly on the spatial interpolation scale and (iii) strictly positive predictions on the original scale.

Initially, the intuitive and, except for large factorial trends, most obvious and simple approach for modelling steady trends would be to use a linear model with time on the natural (i.e. measurement) scale. However, when we consider 10-year-extrapolations, this approach differs significantly from modelling linear trends on the log-scale only when factorial trends of at least a factor 2 occur and when prediction inaccuracies are ignored. For large factorial trends the log-scale is preferred on intuitive grounds, for small factorial trends the differences between both approaches fade when prediction inaccuracies are taken into consideration. The quantitative incorporation of location specific prediction accuracies is an essential aspect of the current study. As the working scale, the log-scale permits this.

The maps are based on a concept where the variable we want to map—the spatial variable $z(x)$ at an extrapolation time at an observation location x_i —is not observed, but where only approximate observations $\hat{y}(x_i)$ are available, which are subject to an unknown observation error. We can model such approximate observations as the sum of a stationary variable $Z(x_i)$ (4.4) that represents the variable of interest $z(x_i)$, and a measurement error $\varepsilon(x_i)$

$$Y(x_i) = Z(x_i) + \varepsilon(x_i), \quad (4.6)$$

and for this observation error $\varepsilon(x_i)$ we assume that it has no systematic component

$$E(\varepsilon(x_i)) = 0, \quad \forall i, \quad (4.7)$$

and that it has a location specific variance $\sigma_\varepsilon^2(x_i)$

$$\text{Var}(\varepsilon(x_i)) = \sigma_\varepsilon^2(x_i), \quad \forall i. \quad (4.8)$$

In this study, we assume that $\sigma_\varepsilon^2(x_i)$ is known, as it is derived from (4.3a) and (4.3b). Furthermore, we assume that the observation errors at two different locations are not correlated

$$\text{Cov}(\varepsilon(x_i), \varepsilon(x_j)) = 0, \quad \forall i \neq j, \quad (4.9)$$

and that the observation errors are not correlated with the spatial variable $Z(x)$

$$\text{Cov}(\varepsilon(x_i), Z(x)) = 0, \quad \forall i, \quad \forall x. \quad (4.10)$$

This implies that beyond the standard kriging assumptions the maps are 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 case of a normally distributed observation error) independently of everything else 95% confidence intervals for $z(x_i)$ are given by

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

The adaption of the standard kriging procedure to handle observations that are subject to an observation error that satisfies (4.6)-(4.10) was given in Section 2.5 and in Delhomme (1978). Compared to the situation where the observation error would have been ignored, applying this adaption 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 observation errors.

The statistical indication of steady time trends in block median values (steady trends in first order statistics) is the primary goal of this study. From a methodological point of view, in a first trend study (like this one) it is a correct first approach to work with time-invariant second order statistics (spatial variation, spatial dependence in variation, the variogram) within soil-land use categories. The decision for that is based on 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 that probably cannot be shown to be necessary, given the number of observations per soil-land use category, Table 3.1), and would represent a modelling approach in which maps no more 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 $Z(x)$ we used the variograms of the 1991 measurements that were obtained for the maps of the groundwater quality in 1991 (Chapter 3, Pebesma and De Kwaadsteniet, 1994). The primary reason for this is that for studying spatial variation of a groundwater quality variable within a soil-land use category, in the previous study we had a much larger observation set. Apart from the measurements from the national groundwater quality monitoring network we could use about 170 measurements from provincial monitoring networks. Distinguishing the 1991-situation and the 'around 1991-situation' with respect to the variogram specification was—with all other simplifications in mind—considered to be overdone.

In this study, estimated second order statistics (variograms and 'observation accuracies') are taken as deterministic quantities. This aspect, together with the contents and foundation

of the model structure implies that the maps only yield a first step in the direction of spatially differentiated statistical trends. Therefore, for the interpretation of the maps caution is required: only first statistical indications of meaningful steady trends in block median values were obtained.

Taking the estimated 'observation accuracy' 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 to 7 measurements, implies that in this first approach for the determination of the 'observation accuracy' the estimated variance $s^2(t_{new})$ (3), based on 3 degrees of freedom when 5 observations are available, was not distinguished from the true variance of (4.3a). This results in a serious underestimation of the 'observation accuracy,' 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 5 measurements, despite the shortcomings of the methods used, is the observation from practice that the influence of such series cannot be ignored.

More generally, it can be stated that in trend studies the accuracies 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 scale, the—now yearly—measurement frequency should not be reduced.

To be complete, it should be noticed that the maps do not pretend to show the actual changes that occur over the 20 year time span considered. The changes shown only concern 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 when these series can be incorporated in the study.

Finally, the current study should be considered as a supplement to the previous study (Chapter 3, Pebesma and De Kwaadsteniet, 1994). The majority of the items discussed in Section 3.8 apply to this study as well.

Chapter 5

Improving estimates with ancillary information

5.1 Introduction

The goal of both monitoring and modelling groundwater quality is to gain understanding of the spatial and temporal variation in groundwater quality. This variation is understood better when more measurements are available or with better models, i.e. models that explain the variation in the available measurements better. Consequently, the optimization of a groundwater quality monitoring network can be aimed at the number and frequency of measurements as well as at a better explanation of the variation in measurements with models. This chapter shows how we can use different types of relevant, ancillary information in models, to increase the accuracy of maps of groundwater quality.

Chapter 3 describes a simple approach for mapping groundwater quality in the Netherlands. Maps of groundwater quality variables were obtained by within soil-land use category interpolation of measurements obtained from the groundwater quality monitoring networks (Van Duijvenbooden et al., 1985), taking the spatial dependence between measurements within categories into account (stratified ordinary block kriging, Journel and Huijbregts, 1978). Estimates of $4 \text{ km} \times 4 \text{ km}$ block mean values (on log-scale, yielding estimates of block median values on the original scale) were presented as approximate 95% confidence intervals on maps having a $2 \text{ km} \times 2 \text{ km}$ grid.

In general the resulting maps turned out to be rather inaccurate with wide confidence intervals. If available, more groundwater quality measurements or better stratification criteria might be used to increase the accuracy of maps, by using the model of the simple stratification approach. Other potentially relevant information typically consisting of either measurements on other (related) variables or results from deterministic process models are available, but cannot be handled by this model. This chapter will show how this other, *ancillary* information can be incorporated in the mapping procedure to improve the accuracy of resulting groundwater quality maps. For three groundwater quality variables (nitrate, zinc and sulphate concentration) the size of the improvements will be shown by comparing the maps that use ancillary information with the maps from the simple stratification approach.

The nature of the ancillary information used in the examples is described in Section 5.2. The simple approach of Chapter 3 (stratification–data selection–interpolation) is the point of departure for the current study, and Section 5.3 describes how the interpolation procedure is adapted to use the ancillary information within a soil-land use category. Two examples of using map information obtained from models are presented: one example uses nitrogen leaching

rate from soils in the agricultural area (see Fig 3.1, page 26) to aid the estimation of nitrate concentration in deeper groundwater (Section 5.4), the other uses atmospheric deposition of zinc for estimating zinc concentration in groundwater in areas with semi-natural vegetation (Section 5.5). The example of using other measured variables is the application of sulphate concentration measured in shallow groundwater in areas with semi-natural vegetation for the estimation of sulphate concentration in deeper groundwater in that area, and this example is given in Section 5.6. Section 5.7 summarizes the size of the effects of ancillary information, and section 5.8 concludes with a discussion.

5.2 Ancillary information

In groundwater quality studies our major concern is groundwater pollution, and in this chapter we limit the study area to the area of sandy soils in the Netherlands (Fig. 3.1, page 26) that are not influenced by strictly local pollution. In these areas groundwater pollution occurs because pollutants applied to the soil surface can dissolve and infiltrate to a depth of about 10 m within a few decades (Meinardi, 1994). Furthermore, these areas allow the best connection between groundwater quality measurements at 10 m depth and soil surface characteristics at the measured location, because the horizontal displacement is generally small (which is necessary because information on the actual groundwater flow patterns is lacking). The main factors influencing groundwater quality that are available as maps (and therefore useful for the mapping of groundwater quality) are soil type (Steur et al., 1985) and land use (Thunnissen et al., 1992): they were used for the stratification in the simple approach described in Chapter 3.

Variations between measurements of groundwater quality as observed within soil-land use categories can be considered as stemming from (i) measurement error in the groundwater quality variable, (ii) classification errors (measurement error in the soil-land use information resulting in wrong assignment of measurements to categories), or (iii) true variation in groundwater quality. Estimates of groundwater quality can be improved by reducing the first two sources of variation, but here we focus on the third. True variation of groundwater quality cannot be reduced (although a change in the method of measurement may reduce the variation observed), but it can be modelled. If information on a process that causes the variation of the groundwater quality variable within a soil-land use category is available (e.g. from deterministic process models), this information may provide a better explanation of the variation in the measurements, and the subsequent reduction of the residual, unexplained variation may yield more accurate estimation. Also if measurements are available for other, associated variables, the variation of which is caused by the same processes that cause the variation of the given groundwater quality variable, they may help to improve its estimation.

Deterministic process models do not necessarily work with the same size spatial units as used in Chapter 3. In examples A and B the spatial units are 500 m and 10 km, and the model calculations were not obtained specifically for explaining variation in groundwater quality. When a large part of the variation of groundwater quality measurements occurs within areas the size of the spatial units of a deterministic process model, then this variation can never be explained by using this model information. It may however be possible to apply the same deterministic process model for smaller spatial units, and tuning the spatial unit size of these process models to the size

of groundwater quality measurements, i.e. downscaling the process models at the measurement sites to the measurement scale might further improve the value of these process models for the estimation of groundwater quality variables, which is discussed in section 5.8.

5.3 Spatial interpolation with ancillary information

In the simple approach (Chapter 3) we obtained maps of groundwater quality by ordinary kriging (Journel and Huijbregts, 1978) of the groundwater quality variables stratified by soil type and land use. For this, we adopted a model where log-transformed measurements $z(x)$ within a soil-land use category were considered as a sample from a realization of a locally intrinsic random field $Z(x)$ with the following properties:

$$Z(x) = m + e(x) \tag{5.1}$$

with

$$m \text{ constant,} \tag{5.1a}$$

$$E(e(x)) = 0, \quad \forall x \in \mathbf{R}, \tag{5.1b}$$

and

$$\frac{1}{2}E[(Z(x) - Z(x+h))^2] = \gamma_Z(\|h\|), \quad \forall x, x+h \in \mathbf{R}. \tag{5.1c}$$

The expected value of the variable $Z(x)$ within a local neighbourhood \mathbf{R} is m , the deviation $e(x)$ from this mean value might be spatially correlated, and the spatial correlation between two measurements is a function of the separation distance h only. Additionally we assumed that only one function $\gamma_Z(h)$ is needed per soil-land use category. The neighbourhood size \mathbf{R} was chosen as the nearest 10 to 20 measurements within a distance of 100 km.

When the ancillary information $f(x)$ is obtained from a deterministic process model (and is known at all locations x) we can extend (5.1) to the more general form:

$$Z(x) = g(f(x)) + e(x).$$

The expectation of Z , $g(f(x))$ is now a function of $f(x)$, the form of which must be specified. If $f(x)$ is a continuous variable and we assume a linear relation, the obvious choice for $g(\cdot)$ is the linear model, extending (5.1) with $f(x)$ to:

$$Z(x) = \beta_1 + f(x)\beta_2 + e(x) \tag{5.2}$$

with

$$\beta_1, \beta_2 \text{ constant,} \tag{5.2a}$$

$$E(e(x)) = 0, \quad \forall x \in \mathbf{R}, \tag{5.2b}$$

and

$$\frac{1}{2}E[(e(x) - e(x+h))^2] = \gamma_e(\|h\|), \quad \forall x, x+h \in \mathbf{R}. \quad (5.2c)$$

In (5.2), β_1 is the constant parameter (like m in (5.1)), and $f(x)\beta_2$ represents a linear contribution from the deterministic process model to the explanation of the measurements. Model (5.2) expresses that within a local neighbourhood \mathbf{R} the expectation of $Z(x)$ has the shape of $f(x)$. In the estimation procedure this shape will be fitted (locally) to the measurements by scaling $f(x)$ by a factor β_2 and shifting this with a value β_1 . When the scatter plot of the observations and $f(x)$ reveals nonlinearities, a non-linear transformation of the observations or of $f(x)$ possibly makes (5.2) more acceptable (e.g., in our examples the model (5.2) is used for measurements on log-scale), or, when justified by the measurements, a more complex model can be adopted.

The variogram $\gamma_e(\|h\|)$ can be modelled by calculating the sample variogram from ordinary least squares residuals $\hat{e}(x)$ (Table A6, page 97), and fitting a suitable variogram model to this sample residual variogram using weighted least squares fitting (Cressie, 1985). This variogram will be biased, but some recent papers (Cressie and Zimmerman, 1992, Kitanidis, 1993, Christensen, 1993) argue that this bias will not seriously disturb estimates based on this variogram, and we accept their arguments.

Estimation of block mean values of Z using (5.2) is done by universal kriging (Section 2.3 and 2.5; Matheron, 1971). Prerequisites for this are (i) $f(x)$ is known at all measurement locations and (ii) block average values of $f(x)$ are known for the estimation locations.

If ancillary information consists of measurements $z_2(x)$ on a variable which is related to our primary variable of interest $z_1(x)$, we cannot use it as a fixed, explanatory variable (like $f(x)$ in (5.2)) because it is known only at the measurement locations. If, apart from the measurements, no other ancillary information is available than the knowledge that both variables belong to the same soil-land use category, we can adopt a model similar to (5.1) for both variables in the category and allow for inter-variable spatial dependence. This dependence may actually improve the estimation of $z_1(x)$. Thus, we extend model (5.1) with a secondary variable:

$$Z_1(x) = m_1 + e_1(x), \quad Z_2(x) = m_2 + e_2(x) \quad (5.3)$$

with

$$m_1, m_2 \text{ constant}, \quad (5.3a)$$

$$E(e_1(x)) = 0, \quad E(e_2(x)) = 0, \quad \forall x \in \mathbf{R}, \quad (5.3b)$$

$$\frac{1}{2}E[(Z_1(x) - Z_1(x+h))^2] = \gamma_{11}(\|h\|), \quad \forall x, x+h \in \mathbf{R}, \quad (5.3c)$$

$$\frac{1}{2}E[(Z_2(x) - Z_2(x+h))^2] = \gamma_{22}(\|h\|), \quad \forall x, x+h \in \mathbf{R}, \quad (5.3d)$$

and

$$\frac{1}{2}E[((Z_1(x) - m_1) - (Z_2(x+h) - m_2))^2] = \gamma_{12}(\|h\|), \quad \forall x, x+h \in \mathbf{R}. \quad (5.3e)$$

The cross variogram of (5.3e) is used because it does not require measurements at identical

locations. Ordinary cokriging is used for estimation of the variable of interest $Z_1(x)$ (Section 2.5, Journel and Huijbregts, 1987, Myers, 1982). As originally proposed by Clark et al. (1989), Myers (1991) and Ver Hoef and Cressie (1993) discuss the use of the cross variogram (5.3e) in multivariable estimation.

To prevent misinterpretation of point estimates, estimated values on the groundwater quality maps are presented as approximate 95% confidence intervals, calculated (on log-scale) by

$$[\hat{z}(x_0) - 2\sigma_k(x_0), \hat{z}(x_0) + 2\sigma_k(x_0)] \quad (5.4)$$

with $\hat{z}(x_0)$ the estimated value (block kriging estimate), and $\sigma_k(x_0)$ the estimation standard deviation (kriging standard deviation). Back transformation to the original scale by taking the exponent of both sides yields approximate 95% confidence intervals for 4 km \times 4 km block median concentrations (the median of all point values in a 4 km \times 4 km block). Because of the difficulty of showing two values (both sides of the confidence interval) at one location in a map, every groundwater quality map is split into four sub-maps, each of them showing the position of the intervals relative to one of four reference levels. This relative position can be *lower* (the interval is completely below the reference level), *higher* (the interval is completely above the reference level), or *not distinguishable* (the interval straddles the reference level, and the estimated value cannot be distinguished from that level given the information available). The reference levels are chosen as (multiples of) critical levels or target levels. It should be noted that these critical levels or target levels usually apply to individual samples. If a block median concentration exceeds a critical level, more than half of the individual samples in the block exceeds this level. If a block median concentration is below a critical level, still many individual samples in the block may exceed this level. The doubling of the levels reflects equidistant (linear) levels on the log-scale.

5.4 Example A. Nitrate leaching from agricultural soils

Regional nitrate leaching rates in agricultural areas were estimated by the NLOAD model (Van Drecht, 1993a, 1993b), an empirical model that predicts steady state leaching of nitrate from the soil to the shallow groundwater as a function of manuring rate, soil type, type of agricultural land use and depth of the groundwater table. The finest spatial resolution of NLOAD calculations currently available is 500 m \times 500 m. Only the soil-land use category of grassland on sand will be considered in this example. Fig. 5.1 shows NLOAD model calculations for 1989 averaged to 2 km \times 2 km cells, derived from the original map in Maas (1992). Fig. 5.2 shows a scatter plot of the nitrate concentrations measured at the sites of the groundwater quality monitoring networks and the NLOAD calculations obtained from the 500 m \times 500 m cells at the corresponding locations. Fig. 5.2 shows that nitrate concentration in groundwater tends to be greater when nitrate leaching rates are higher. The nitrate measurements used for this study were obtained from the category 'grassland on sand,' and were selected from the monitoring network data according to the selection criteria of Section 3.3.

Fig. 5.3a shows the variogram of the (log) nitrate concentration measurements. Fig. 5.4 shows the map with estimates of nitrate concentration in the groundwater, based on these

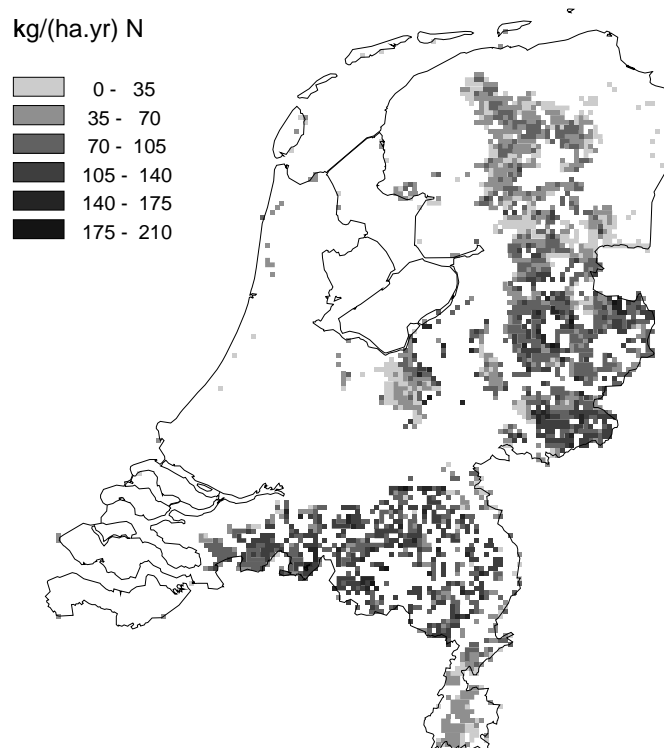


Figure 5.1 Map of nitrogen leaching rate from agricultural soils to shallow groundwater. Values are averaged from the original $500\text{ m} \times 500\text{ m}$ model calculations to yield $2\text{ km} \times 2\text{ km}$ mean values. Only cells in the category grassland on sand are shown

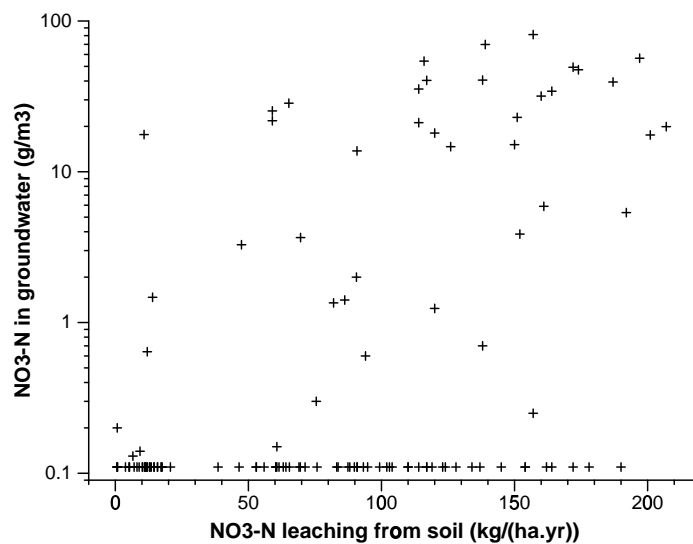


Figure 5.2 $\text{NO}_3\text{-N}$ concentration in groundwater and nitrogen leaching rate. Plot of $\text{NO}_3\text{-N}$ concentration measured in the groundwater versus nitrogen leaching rate from soils to shallow groundwater, for the category grassland on sand

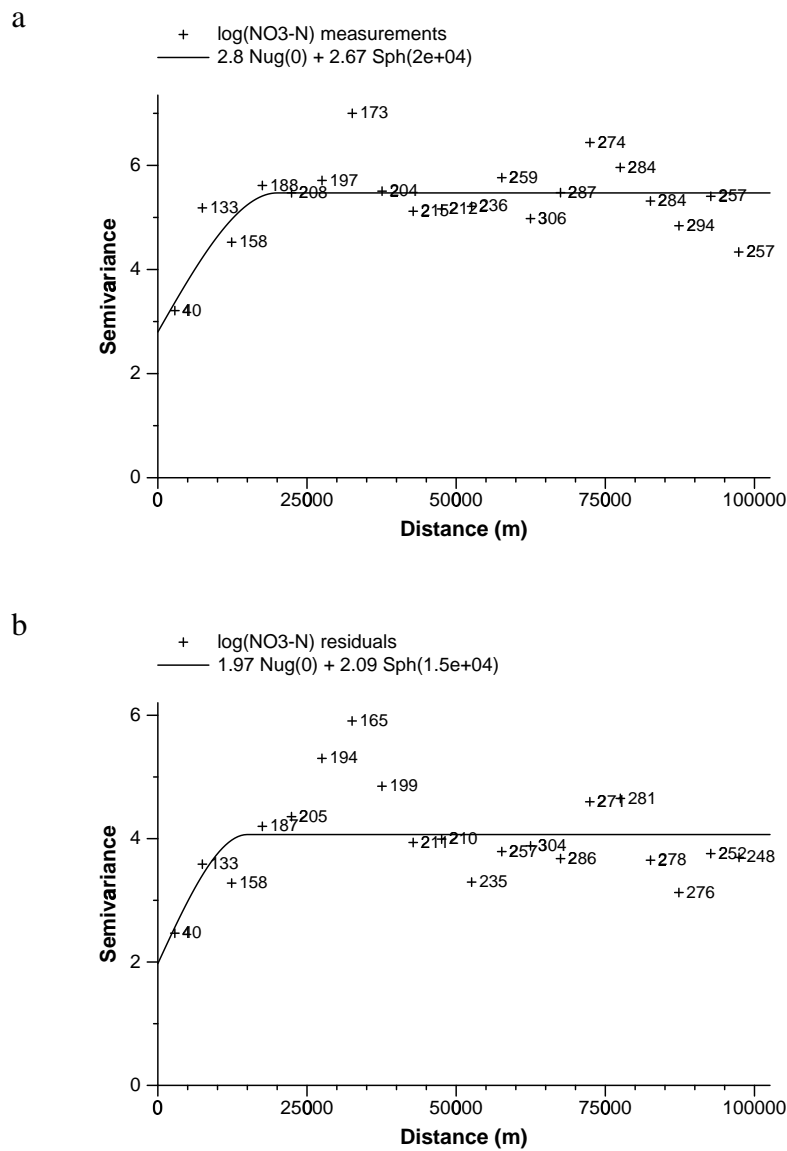


Figure 5.3 Variograms of (a) $\log(\text{NO}_3\text{-N})$, and (b) residuals for the category grassland on sand. Calculation of the sample variograms is explained in Tables A2 (page 96) and A6 (page 97). Numbers reflect the number of (a) observation pairs or (b) residual pairs used for an estimate (+) of a variogram point. Codes used for a variogram model (—) are explained in Table A3, (page 97)

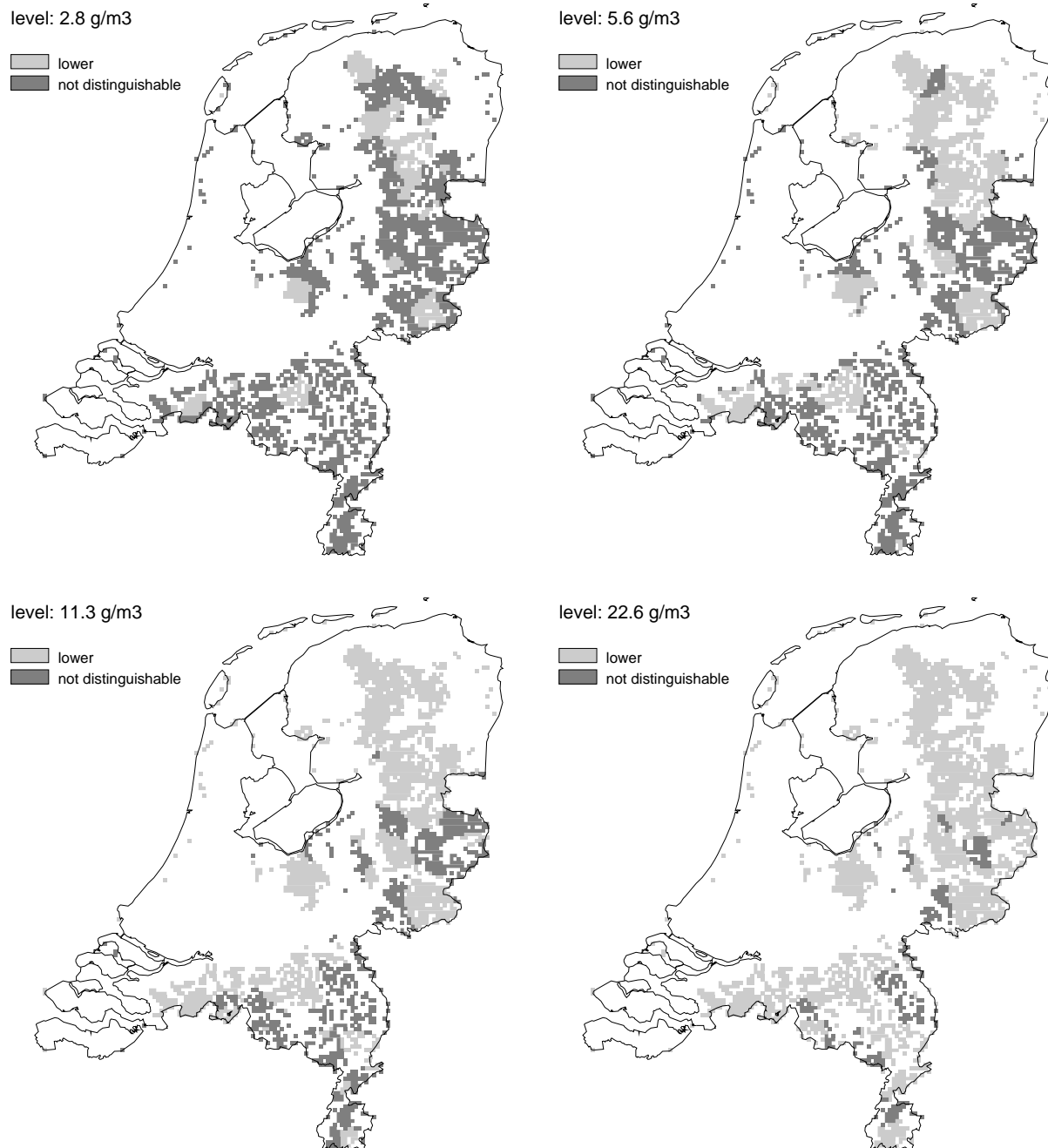


Figure 5.4 Map of $\text{NO}_3\text{-N}$ in the groundwater (category grassland on sand). 95% Confidence intervals for $4 \text{ km} \times 4 \text{ km}$ block median values, related to four concentration levels. Estimates are obtained by ordinary kriging (on log-scale), using information from the monitoring networks only

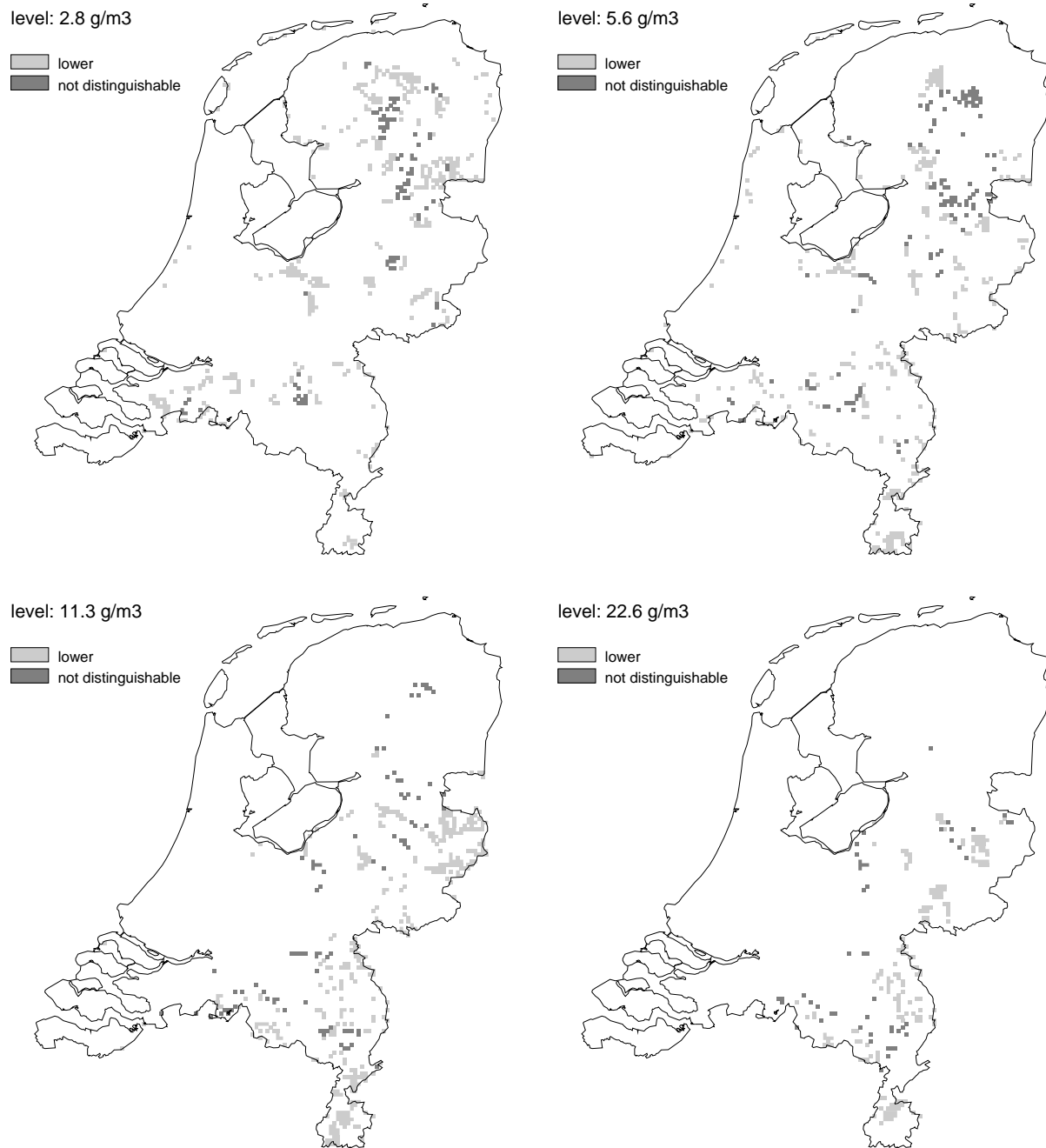


Figure 5.5 *The effects of using information on nitrogen leaching rates. 95% Confidence intervals for 4 km × 4 km block median $\text{NO}_3\text{-N}$ concentrations, related to four levels. Estimates are obtained by universal kriging (on log-scale), using both information from the monitoring networks and ancillary information (nitrate leaching rates). Only results that differ from Fig. 5.4 are shown*

measurements and this variogram, computed using ordinary kriging as in Section 3.4, as obtained from Pebesma and De Kwaadsteniet (1994).

Fig. 5.3b shows the variogram of the estimated residuals, obtained from simple (ordinary least squares) linear regression of these (log) measurements $Z(x)$ and nitrate leaching rates at corresponding locations, $L(x)$

$$Z(x) = \beta_1 + L(x)\beta_2 + e(x), \quad E(e(x)) = 0,$$

as described in Table A6 (page 97).

Fig. 5.5 shows the map of nitrate concentration in the groundwater when both measurements and N_{LOAD} calculations (Fig. 5.1) are used with the variogram of Fig. 5.3b (using universal kriging). To show the magnitude of the effects of the N_{LOAD} information, Fig. 5.5 shows only those cells that differ from Fig. 5.4.

5.5 Example B. Atmospheric deposition of zinc

Total atmospheric deposition of zinc in 1985 was modelled for $10 \text{ km} \times 10 \text{ km}$ blocks with the $TREND$ model (Van Jaarsveld, in prep.) by combining information about sources (emissions, mainly from areas south of the Netherlands and areas with heavy industry) and air transport. Because groundwater quality was modelled using soil type and land use information on $2 \text{ km} \times 2 \text{ km}$ cells, the $10 \text{ km} \times 10 \text{ km}$ model predictions have been interpolated to 2 km cell values (by simply considering $10 \text{ km} \times 10 \text{ km}$ block estimates as the value at the block centre, and using inverse distance interpolation with the nearest 4 block-centre values). Since the original $10 \text{ km} \times 10 \text{ km}$ map was very smooth, the interpolation to $2 \text{ km} \times 2 \text{ km}$ cells has not affected the original pattern, as shown in Fig. 5.6. We will focus on the area with semi-natural vegetation on sand, where air deposition is expected to be the main source for zinc in groundwater. The zinc measurements used for this study were obtained from the category ‘semi-natural vegetation on sand,’ and were selected from the monitoring network data according to the criteria of Section 3.3.

The scatter plot of zinc concentration measured in the groundwater and the modelled atmospheric deposition of zinc (Fig. 5.7) reveals that in general zinc concentration tends to be larger at locations with higher deposition rates. In Fig. 5.7 the measurement with the highest deposition rate disturbs this pattern, and this measurement turned out to be located in a seepage situation, which is a plausible reason for deviating from the general pattern (water sampled at that site is likely to have a different origin). Therefore the measurements at seepage sites (marked with a bullet in Fig. 5.7) were left out for the mapping of zinc concentration, both for the situation with and without the deposition information.

Fig. 5.8a shows the variogram for the (log) zinc concentration measurements selected from the monitoring network data. Fig. 5.9 shows the map of zinc concentration estimates, based on these measurements and this variogram, as obtained by ordinary kriging at the log-scale.

Fig 5.8b shows the variogram of estimated residuals obtained from simple (ordinary least squares) regression of these (log) measurements $Z(x)$ and atmospheric deposition rates, $D(x)$

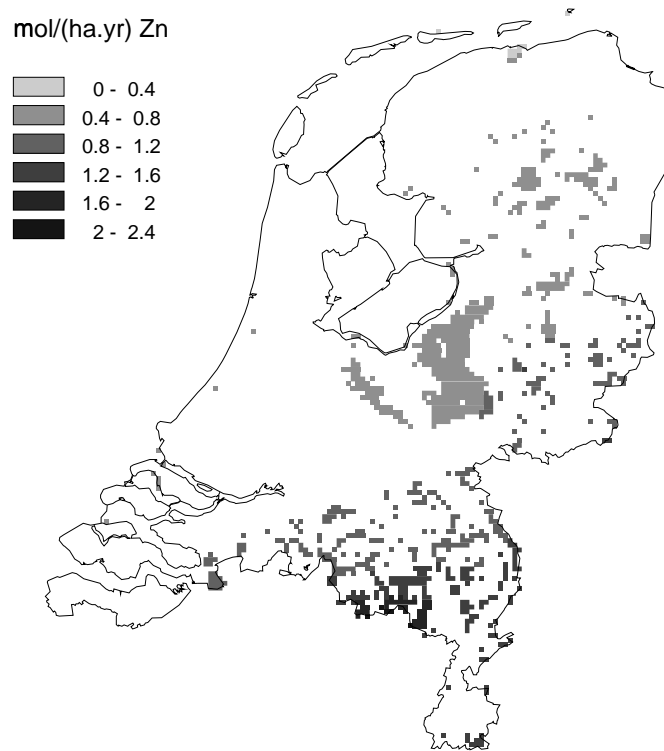


Figure 5.6 Map of atmospheric deposition of zinc. Values are interpolated from the original $10\text{ km} \times 10\text{ km}$ model calculations to $2\text{ km} \times 2\text{ km}$ cells. Only cells in the category semi-natural vegetation on sand are shown

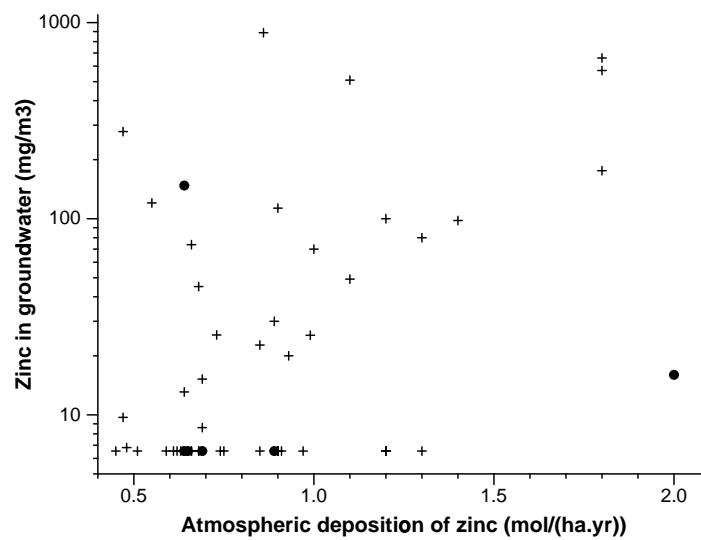


Figure 5.7 Zinc concentration in groundwater and atmospheric deposition of zinc. Plot of zinc concentration measured in the groundwater versus atmospheric deposition of zinc, for the category semi-natural vegetation on sand. A bullet (•) marks a measurement at a seepage location

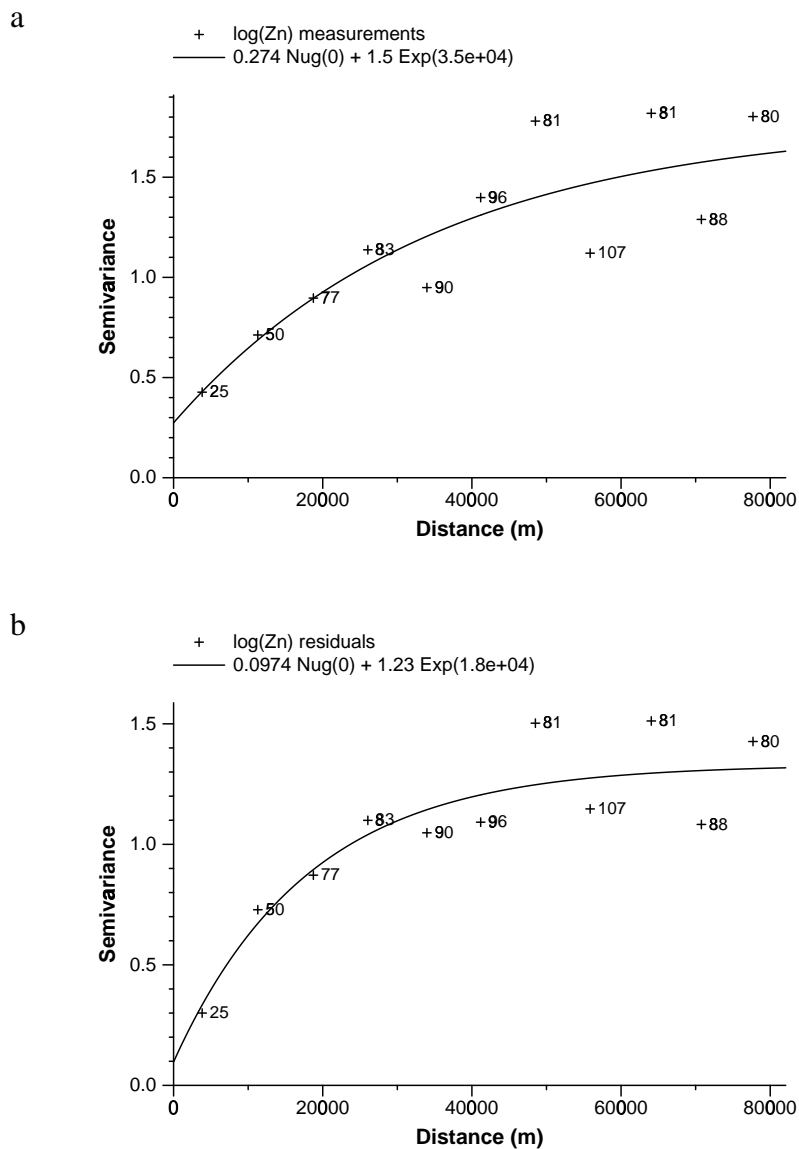


Figure 5.8 Variograms of (a) $\log(\text{Zn})$, and (b) residuals for the category semi-natural vegetation on sand, non-seepage locations. Numbers reflect the number of (a) observation pairs or (b) residual pairs, used for an estimate (+) of a variogram point. Codes used for a variogram model (—) are explained in Table A3, page 97. Calculation of the sample variograms is explained in Tables A2 (page 96) and A6 (page 97)

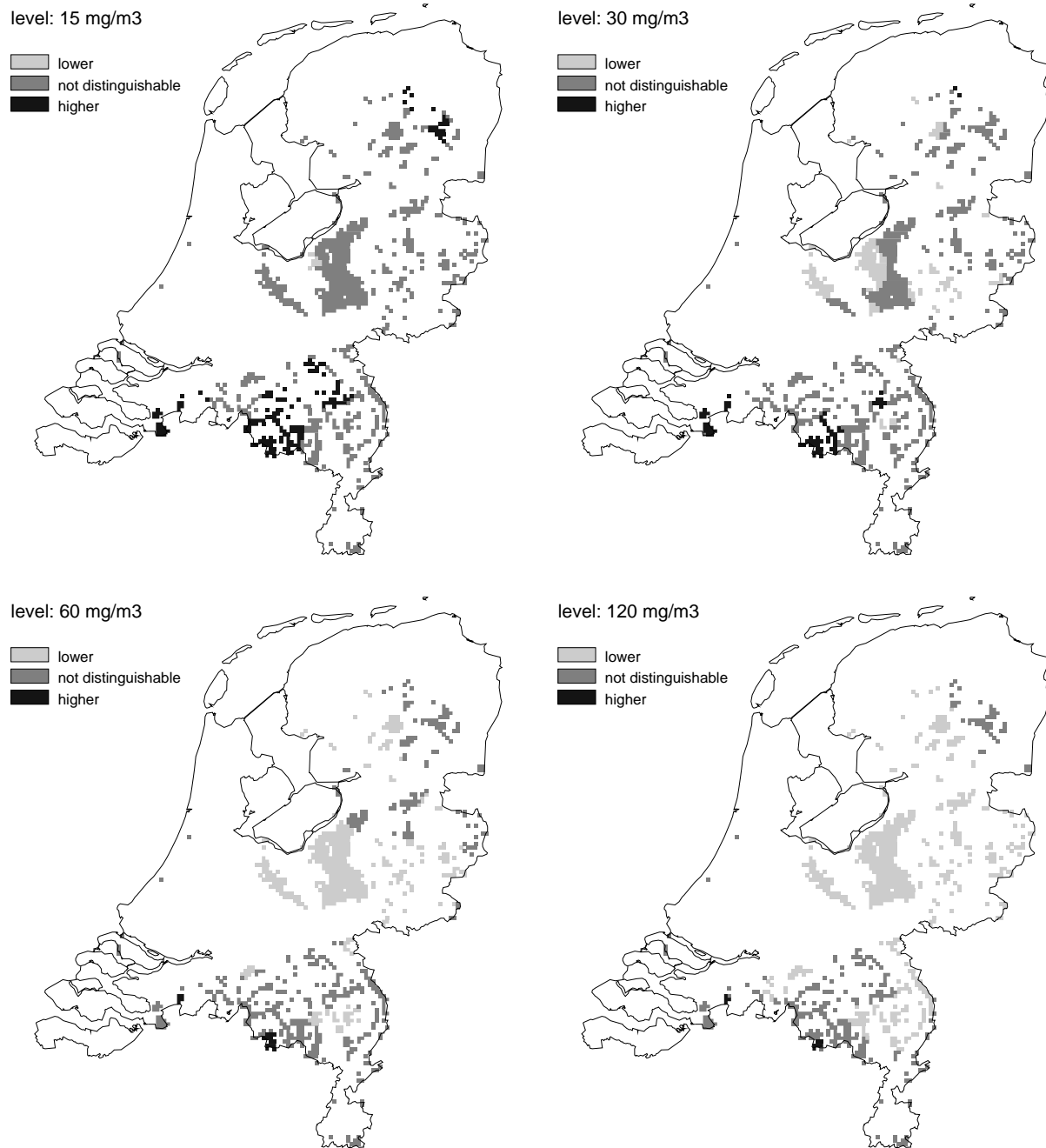


Figure 5.9 Map of zinc in the groundwater (category semi-natural vegetation on sand). 95% Confidence intervals for $4 \text{ km} \times 4 \text{ km}$ block median values, related to four concentration levels. Estimates are obtained by ordinary kriging (on log-scale), using information from the monitoring networks only (results apply to non-seepage locations only)

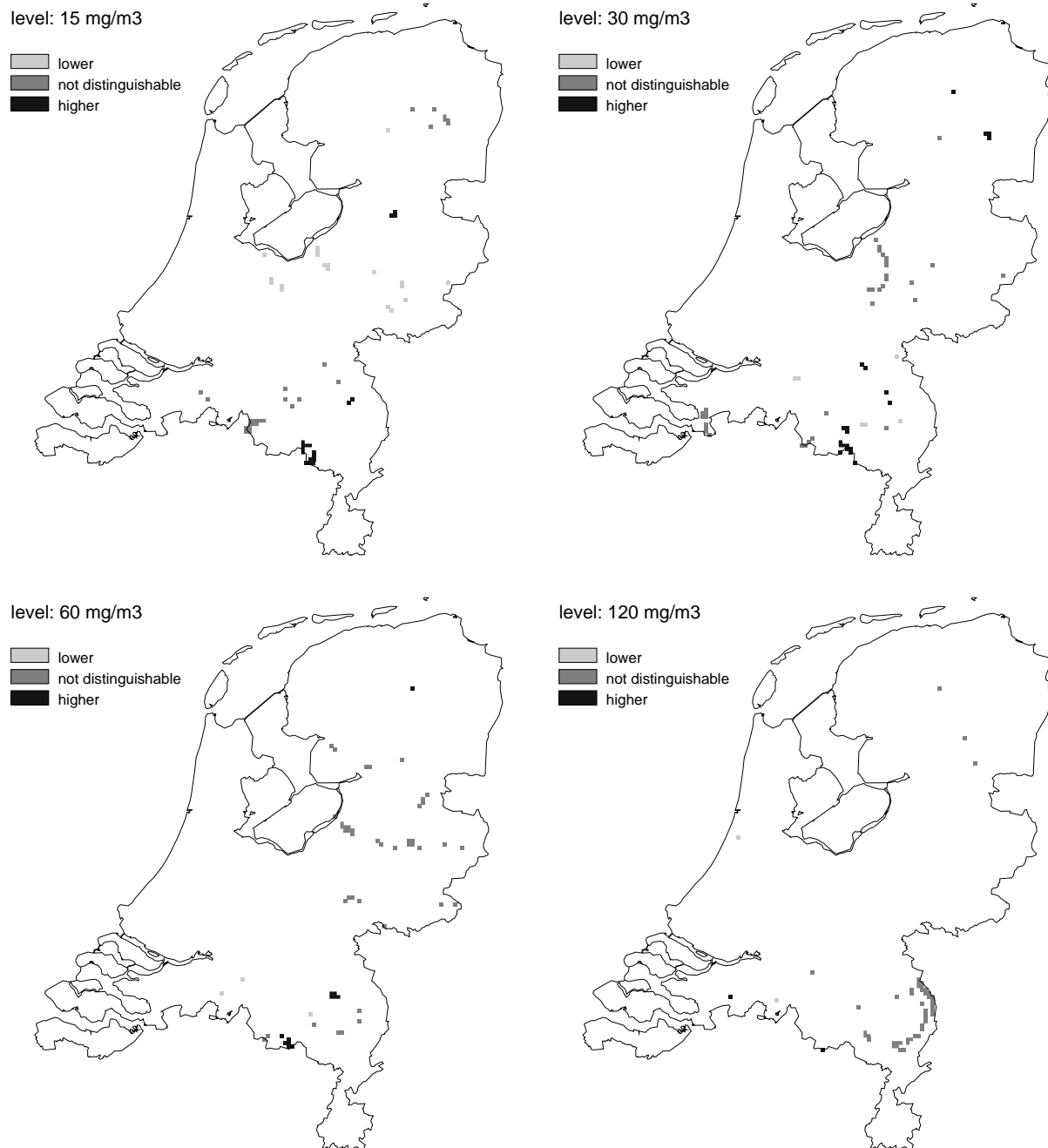


Figure 5.10 The effects of using information on atmospheric deposition of zinc. 95% Confidence intervals for 4 km × 4 km block median zinc concentrations, related to four levels. Estimates are obtained by universal kriging (on log-scale), using both information from the monitoring networks and ancillary information (zinc atmospheric deposition rates). Only results that differ from Fig. 5.9 are shown (results apply to non-seepage locations only)

$$Z(x) = \beta_1 + D(x)\beta_2 + e(x), \quad E(e(x)) = 0,$$

as described in Table A6 (page 97).

Fig. 5.10 shows the modifications to the zinc concentration map (obtained by universal kriging) that resulted from addition of the model information on atmospheric deposition of zinc. Estimates based on explicitly selected non-seepage sites only apply to non-seepage locations, and since all seepage locations are currently not known, Figs. 5.9 and 5.10 should be interpreted with care.

5.6 Example C. Measurements of sulphate in shallow groundwater

Sulphate concentration in shallow groundwater was measured in forest areas during a one-time investigation of soil and groundwater quality in semi-naturally vegetated areas in 1989-90 (Boumans and Beltman, 1991) as part of the national acidification research program. The measurements comprise 1560 individual samples, collected in 156 separate 500 m × 500 m squares (in each square 10 samples were taken on a 50 m inter-spaced transect). As before, the measurements are restricted to areas with semi-natural vegetation on sand. The locations of the 156 squares are shown in Fig. 5.11, along with the sites for the groundwater quality monitoring networks. The mean values for individual squares are used for the variogram modelling, since they will be used for the interpolation. Cross variogram and direct (sample) variograms for sulphate measurements are shown in Fig. 5.12. To assure valid estimates, the linear model of coregionalization (LMC) is used for modelling the variograms (Journel and Huijbregts, 1978). The LMC requires that for m variables all direct and cross variogram models are derived from a linear combination of n direct variogram functions $\gamma_i(h)$:

$$\gamma_{kk'}(h) = \sum_{i=1}^n b_{kk'}^i \gamma_i(h), \quad \text{with } b_{kk'}^i = b_{k'k}^i, \quad \forall k, k' : k, k' = 1 \dots m, \quad \forall i : i = 1 \dots n, \quad (5.5)$$

and that for every i , the coefficient matrix $[b_{kk'}^i]$ is positive semidefinite. For two variables it suffices to show that

$$|b_{12}^i| \leq \sqrt{b_{11}^i \cdot b_{22}^i}, \quad \forall i : i = 1 \dots n. \quad (5.6)$$

Since (i) any valid cross variogram plus a constant $\bar{\gamma}_{kk'} = \gamma_{kk'} + C$ gives identical estimates irrespective of the value of C (Myers, 1991), and (ii) the estimation of block mean values from two variables with no identical locations never requires the actual value of $\gamma_{kk'}(0)$, it is irrelevant that condition (5.6) does not hold for the 'nugget' variance. The apparent 'nugget' variance of Fig. 5.12c can be taken as consisting of a true nugget variance that conforms with (5.6) and an unknown constant that needs no explicit specification (e.g. resulting from the estimation error for the mean values m_1 and m_2). Calculation of variograms and cross variogram is explained in Tables A2 (page 96) and A6 (page 97).

Using the LMC limits the possible models for the cross variogram to symmetric cross variograms only (i.e. $\gamma_{kk'}(h) = \gamma_{kk'}(-h)$). In our example this limitation is not serious, since



Figure 5.11 Map of sulphate measurement sites. Shallow groundwater measurement sites (+) and groundwater quality monitoring network sites (●) in the category semi-natural vegetation on sand

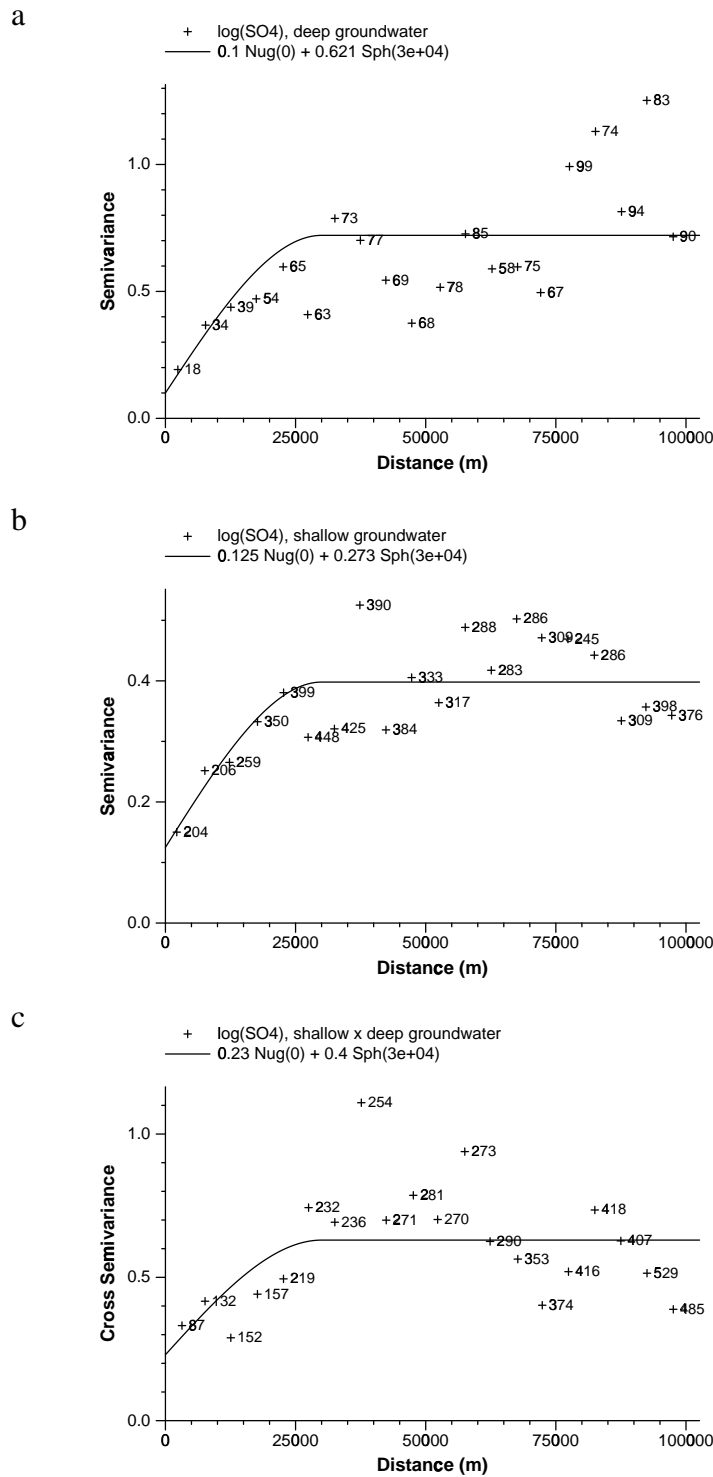


Figure 5.12 Variograms of sulphate. Variogram for (a) $\log(SO_4)$ measurements from the groundwater quality monitoring networks (category semi-natural vegetation on sand), for (b) $\log(SO_4)$ from the shallow groundwater measurements, and (c) cross variogram for (a) and (b). A number reflects the number of observation pairs used for an estimate (+) of a variogram point, codes used for a variogram model (—) are explained in Table A3, page 97. Calculation of the sample variograms is explained in Tables A2 (page 96) and A7 (page 98)

directional (asymmetric) sample cross variograms did not suggest a serious asymmetry.

Fig. 5.13 shows the groundwater quality map of sulphate concentration based on information from the groundwater quality monitoring networks only (using ordinary kriging and the variogram of Fig. 5.12a). Fig. 5.14 shows the modifications to this map resulting from the incorporation of the shallow groundwater measurements, using cokriging and the variograms of Figs. 5.12a-c.

5.7 The effects of ancillary information

The effects of adding ancillary information to the interpolation process can be seen by comparing the results with the maps resulting from the simple approach; i.e. compare Fig. 5.5 with Fig. 5.4 (addition of N_{LOAD} nitrate leaching information), Fig. 5.10 with Fig. 5.9 (addition of $TREND$ zinc deposition information), and Fig. 5.14 with Fig. 5.13 (addition of sulphate measurements from shallow groundwater). Table 5.1 shows the number and percentage of changed cells at every reference level for each example. In general, the modifications to the nitrogen map are considerable, whereas the modifications to the zinc and sulphate maps are rather modest (though not negligible).

At first sight the difference in magnitude of the modifications between nitrate and zinc maps may be surprising, since for both the strength of the relation (Figs. 5.2 and 5.7) and the difference between measurement and residual variograms (in Figs. 5.3 and 5.8) is similar. The spatial variation of the respective maps of nitrate leaching estimates and zinc deposition estimates (Figs. 5.1 and 5.6) however is very different, and the variation in the *local* relation between measurement variable and ancillary variable differs for nitrate (Fig. 5.15a) and zinc (Fig. 5.15b). For zinc the relation between measurements and ancillary variable is negligible in the northern two neighbourhoods of Fig. 5.15b, and local estimation of the mean (as in (5.1)) will account for an important part of the gradual change in mean zinc concentration (suggested by Figs. 5.6 and 5.7). On the contrary, the large short distance variation of the N_{LOAD} map (Figs. 5.1 and 5.15a) contributes locally more to the estimation of nitrate concentration.

The modifications to the sulphate map resulting from the shallow groundwater measurements (Fig. 5.14) obviously occur mainly at locations with many shallow groundwater measurements and few monitoring network locations nearby.

5.8 Discussion

Chapter 3 explained the procedure used to make maps of groundwater quality using a stratification by soil type and land use, relevant measurements from available groundwater quality monitoring networks, and an interpolation method that takes spatial dependence between measurements within strata into account. This study shows how ancillary information, consisting either of results from deterministic models of processes that cause variation in groundwater quality or of measurements on a variable the variation of which is caused by the same processes as the groundwater quality variable, can be incorporated in the mapping procedure to improve the accuracy of the groundwater quality maps. Three examples (on nitrate, zinc and sulfate

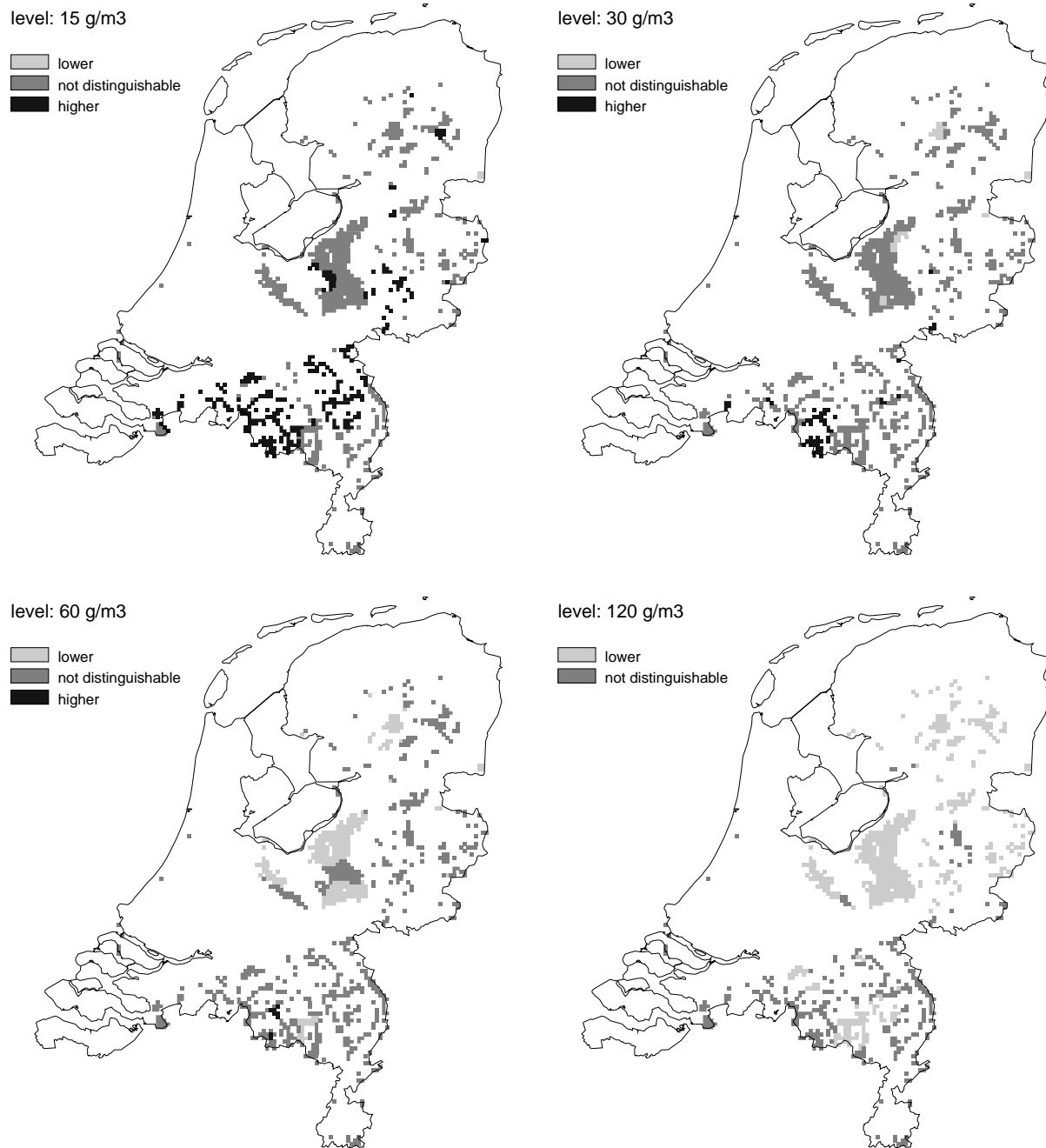


Figure 5.13 Map of sulphate in the groundwater (category semi-natural vegetation on sand). 95% Confidence intervals for $4 \text{ km} \times 4 \text{ km}$ block median values, related to four concentration levels. Estimates are obtained by ordinary kriging (on log-scale), using measurements from the groundwater quality monitoring networks only and the variogram of Fig. 5.12a

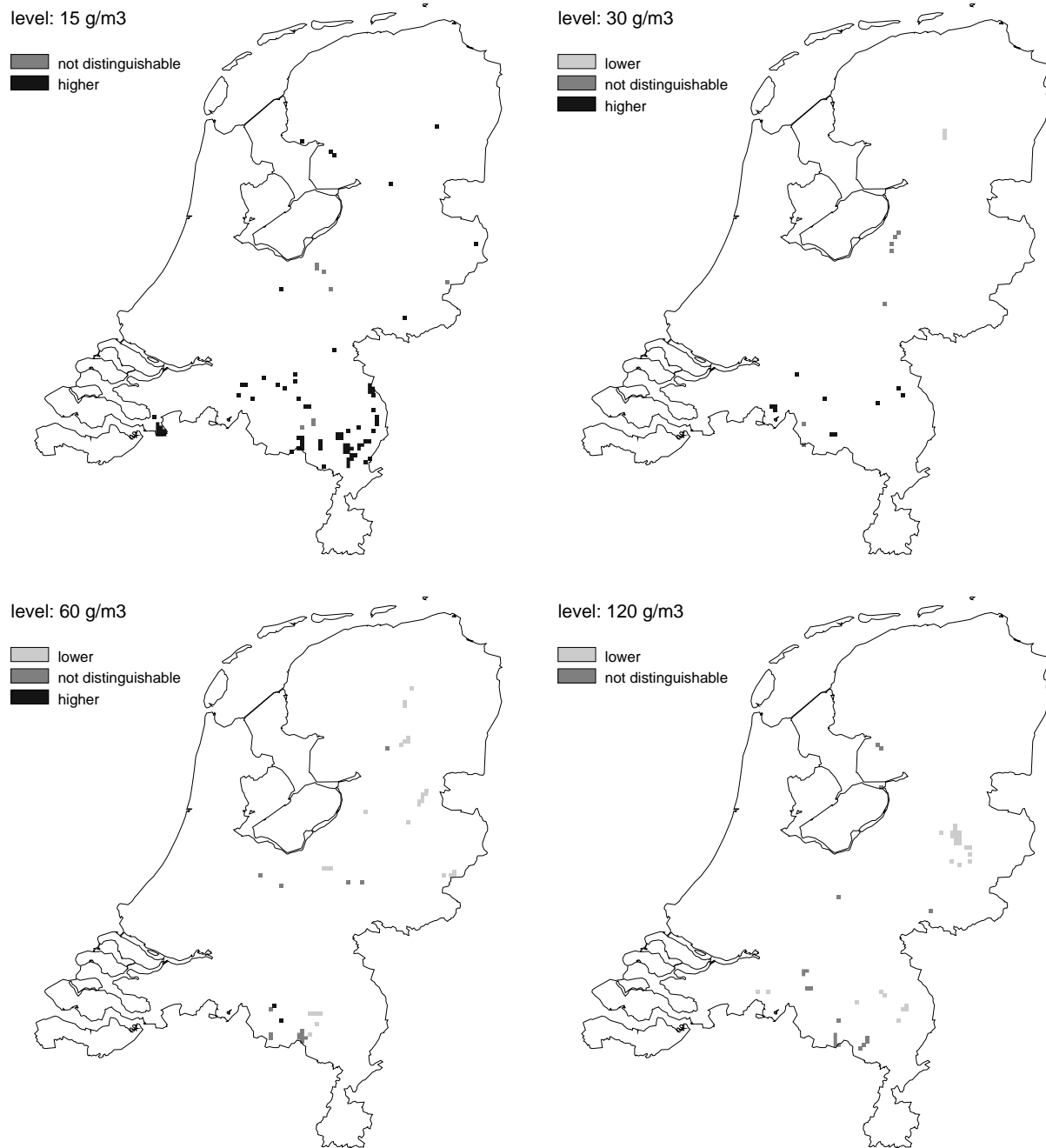


Figure 5.14 The effects of using shallow groundwater measurement information. 95% Confidence intervals for 4 km × 4 km block median sulphate concentrations, related to four levels. Estimates are obtained by cokriging (on log-scale), using measurements from shallow groundwater and from the groundwater quality monitoring networks, and the variograms of Fig. 5.12a-c. Only results that differ from Fig. 5.13 are shown

Table 5.1 *The effects of ancillary information. For each variable and level the number and percentage of cells for which the addition of ancillary information changed the outcome (possible values: lower, higher and not distinguishable)*

variable	level	% cells changed	number of cells changed
NO ₃ -N	2.8 g/m ³	22	505
	5.6	20	477
	11.3	20	471
	22.6	10	240
Zn	15 mg/m ³	7	64
	30	7	60
	60	6	57
	120	5	51
SO ₄	15 g/m ³	9	87
	30	2	20
	60	5	46
	120	5	48

concentration) have shown that the addition of ancillary information does improve the resulting maps in terms of the ability to distinguish estimated 4 km × 4 km block median values from critical or target levels and their multiples (compare Fig. 5.5 with 5.4, 5.10 with 5.9 and 5.14 with 5.13). The models presented in this chapter are extensions to the simple model of Chapter 3 that are easy to apply, provided that the relevant ancillary information is available. Since the improvement is considerable, it has been well worth the effort—only extra computing and data handling costs were involved.

Examples A, B and C have shown that we gained understanding of the spatial variation of groundwater quality variables by including new, relevant information in the model, resulting in a better explanation of the variation of the measurements and in improved maps (more accurate estimates) of groundwater quality variables. From the perspective of a monitoring network optimization it is important to know how many added measurements would have given an equivalent improvement (e.g. Burrough et al., 1995), but this is a topic for future study.

The methods presented here are generally applicable in environmental studies, and provide a convenient way for incorporating results from (often complex) deterministic process models and measurements of other variables in the estimation (spatial interpolation) of the variable of interest at non-measured locations. Both models (5.2) and (5.3) include model (5.1) as a special case and will reduce to (5.1) when the added information has no relation to the primary measurement variable. Therefore they are robust to overspecification of the model (only for (5.2) will overspecification cost some efficiency, since parameters that are zero will be estimated). Most applications of universal kriging use (polynomials of) the coordinates as explanatory variables (replacing $f(x)$ in (5.2), e.g. Stein et al., 1991). Examples A and B instead show how variables that carry a physical significance to the measurement variable (based on theory, i.e. the hydrological



Figure 5.15 Local scatter plots of measurements versus ancillary variable for $\text{NO}_3\text{-N}$ (a) and zinc (b). For five arbitrarily chosen (but distinct) local kriging neighbourhoods \mathbf{R} , scatter plots show concentration in groundwater (a: $\text{NO}_3\text{-N}$ in g/m^3 , b: zinc in mg/m^3) versus the ancillary variable (a: nitrate leaching rate in $\text{kg}/(\text{ha}\cdot\text{yr})$, b: atmospheric deposition of zinc in $\text{mol}/(\text{ha}\cdot\text{yr})$). Arrows link the location for which kriging neighbourhood information is shown to the corresponding scatter plot

cycle and chemical balances) can be utilized. We felt no need to further ‘prove’ the validity of the models (e.g. with external criteria) since both theory and observations (in particular Figs. 5.2 and 5.3, 5.7 and 5.8, and 5.12) support the models chosen.

The examples presented in this study can be improved further. In particular in examples A and B the process models concern averages of spatial ‘model units’ the size of 500 m × 500 m (nitrate leaching) or of 10 km × 10 km (atmospheric zinc deposition), and these averages are compared with groundwater measurements on a volume less than a cubic metre. These values on ‘model units’ can never explain the variation of the groundwater quality variable that occurs *within* these units. When the variation of the measurement variable within these ‘model units’ is relatively large and when this variation is mostly true variation (small measurement errors), then it may be worthwhile to modify (or simply apply) the process model in such a way that it yields values specifically for ‘model units’ that correspond to the area of influence (Section 1.3) of the monitoring screens. Such model values are specifically intended to explain variation in the measurements, and it is expected that using them will lead to a smaller residual variation and result in improved (more accurate) estimates of groundwater quality variables. For the NLOAD model such a modification may be feasible because the short-distance variation of the measurements is relatively large (Fig. 5.3), and all input variables are available at a finer spatial resolution (soil maps 1:50,000 (De Vries and Denneboom, 1992), land use information at a 25 m cell resolution (Thunnissen et al., 1992), and manuring rate information at municipality level). Obtaining zinc deposition estimates for smaller elements may require a different model structure, taking account of local effects (such as local landscape roughness and the position of the groundwater quality measurement site relative to forest edges exposed to the prevailing wind direction, Draaijers, 1993). When the size of the spatial units of the process model is adjusted to the size of the area of influence, the need to know the locations of the influence area (using local groundwater flow patterns) becomes more profound.

Transformation of the estimated fluxes in Examples A and B into estimated concentrations would simplify the interpretation of the regression slope in (5.2). Here, the regression slope is a purely conventional parameter of no interest by itself, and because the rainfall excess within a soil-land use category (as currently known) is nearly a constant (Meinardi, 1994), this transformation would hardly affect the concentration estimates.

In this study mathematical aspects were kept simple because the primary goal is to illustrate the effects of adding ancillary information to the estimation of groundwater quality. The simplicity refers to the models used (5.1-5.3), the estimation of the variograms (5.1c), (5.2c) and (5.3c-e), the choice and fit of variogram models, the choice of the linear model of coregionalization for modelling (5.3c-e), the choice of the local neighbourhood **R**, the way approximate 95% confidence intervals for block median values are estimated (5.4), and the data selection procedure. The approximation of 95% confidence intervals with (5.4) is known to be too optimistic (biased, cf. Rao, 1973), as explained in Sections 2.5 and 3.8.

Chapter 6

Discussion and conclusions

6.1 Introduction

In the context of a full monitoring network optimization this study can be considered as the first basic step: the inference of groundwater quality from available information. The relevance of this study for groundwater quality assessment, and some of the questions that need answers in a network optimization are discussed in Section 6.2. A number of enhancements that are worthwhile considering from the network optimization point of view or from a mathematical or statistical point of view are listed in Section 6.5. A number of methodological choices that are common throughout this study are discussed in Section 6.3, and Section 6.4 summarizes the main conclusions.

6.2 Optimizing the groundwater quality monitoring networks

The groundwater quality monitoring networks account for a substantial part of the total cost of the environmental quality monitoring network in the Netherlands. During the first years of operation the networks provided an initial survey of the spatial variation in groundwater quality, and some years later the first analysis of changes in time appeared. Now, after several years of measuring, the question arises whether the way we currently monitor groundwater quality is worth the cost. Aspects of a groundwater quality monitoring network that control the cost are labour and capital costs of boring and setting up the wells, measurement intensity (spatial well density, number of screens sampled per well, and monitoring frequency), choice and cost of analysis of the chemical variables measured, and modelling effort (the development of models that better explain variation in the measurements).

Optimizing the monitoring with respect to measurement intensity can be done by minimizing the total cost as a function of the measurement intensity. The total cost at a given measurement intensity could be calculated by summing the cost of measurement and the loss due to limited knowledge (e.g. Fig. 6.1). This would be fairly easy if the loss due to limited knowledge (the cost of 'not monitoring') were known as a function of the measurement intensity. Such a loss function would call for an explicit statement of the value of knowing groundwater quality to a certain extent. From this value we could determine the function of expected loss due to limited knowledge, which obviously increases when measurement intensity decreases. The measurement intensity for which the total cost is at its minimum value will yield the optimal (cheapest) measurement intensity. Parameter choice could be evaluated in a similar way.

Specifying a loss function for groundwater quality monitoring networks is not a trivial task, because it calls for the assignment of a value to a particular 'degree of knowledge'. Such

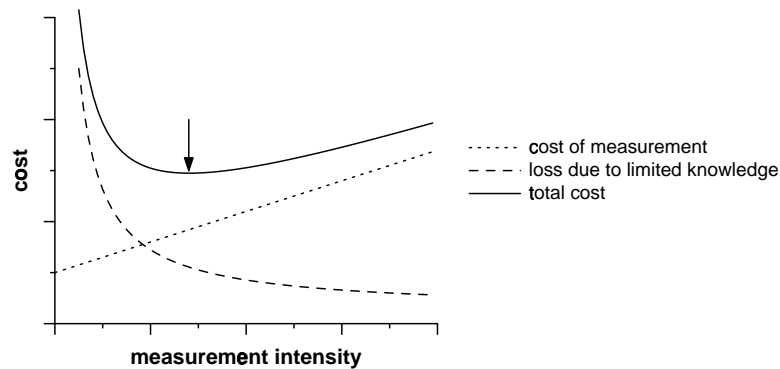


Figure 6.1 *Monitoring network optimization: total cost as a function of measurement intensity. The arrow points to an ‘optimal’ measurement intensity*

a function is only something we can agree on, and it may change in time and with specific application. Clearly this function depends on the uncertainty attached to estimates of groundwater quality, given a certain monitoring network. Because in this study ‘degree of knowledge’ is expressed as the ability to distinguish estimates from critical concentration levels (target levels or maximum tolerated levels for human consumption), it can directly be interpreted as the degree of knowledge of groundwater *quality*, and a meaningful loss function can be derived from it. Other measures of ‘degree of knowledge’ that have been commonly used in optimization studies (e.g. the estimation variance as in Rodríguez-Iturbe and Mejía (1974) and in Van Geer (1987), or the relative estimation variance on the log-scale as in Hoekstra and Heuberger (1995)) do not link estimates to quality assessment, but focus *exclusively* on estimation uncertainty. Therefore, with respect to the degree of knowledge of groundwater quality, and thus for groundwater quality monitoring network optimization, these other measures are inappropriate. Cressie (1991, 4.6.2) discusses a wider class of optimization criteria.

This rough sketch of a measurement intensity optimization assumes that the characteristic of groundwater quality we focus on and the method for obtaining this characteristic from available information have been fixed in advance. Chapter 5 showed that by incorporating relevant ancillary information in the interpolation procedure we can improve estimates of groundwater quality, thus extending the ‘degree of knowledge’ of groundwater quality without changing measurement intensity, and more improvement can be expected (Sections 3.8, 5.8, 6.5). Although models can never replace the crucial role of measurements (e.g. because measurements are needed for model development, validation and calibration), from a monitoring network optimization point of view it could be worthwhile to study to what extent an increase in explained (or described) variation obtained by improved modelling could compensate for a certain decrease in measurement intensity, when a certain ‘degree of knowledge’ of groundwater quality should be retained.

6.3 General aspects of the discussions

This section covers some aspects of the discussions that are relevant to the whole study. More specific aspects of the discussions on the topics of Chapters 3-5 have been placed at the end of these chapters.

The estimates of groundwater quality variables in this study concern block median concentrations of 4 km × 4 km blocks: a balanced choice in which the measurement variation, the spatial monitoring network density and the character of soil type and land use information available were considered. If, for some reason estimates for medians of larger or smaller areas are needed, the same methods can be used (but see also *Measurement error reduction* in Section 6.5). For estimates of other characteristics (e.g. mean values of blocks or quantiles of larger areas) some modifications would be necessary (Sections 3.8, 6.5), or an approach based on classical statistics (design-based inference) may be more appropriate (Section 3.8).

In this study, no information on groundwater flow or transport of solubles as obtained from deterministic process models has been used. If this information is available it can readily be used to improve estimates of measured variables, using the methods presented in Chapter 2, and demonstrated in Chapters 3 and 5. For the *development* (validation, calibration or input specification) of deterministic groundwater quality models, the maps obtained in this study can be considered as a translation of measurement (point) information into information on spatial units the size of units used in regional or national groundwater models (Sections 1.4, 3.8).

The models used in this study merge the advantages of the traditional, classical statistical approach (independent, non-measurement information is used to explain variation in the measured variable) and the geostatistical approach (local information is used by modelling the spatial dependence structure that is often present in the residual, unexplained variation), to estimate (map) the measurement variables.

In chapter 2 geostatistical methods were derived from a classical statistical point of view. This is not the way it evolved historically (Cressie, 1990), but it emphasized that geostatistical models can provide the obvious generalization of models with independent errors, when they are used for spatial estimation (mapping) purposes. In the context of mapping, when no appeal is made to independence of errors from a design-based argument (Section 2.4), the automatic assumption of independence of errors (e.g., Boumans, 1994, Alkemade and Van Esbroek, 1994, RIVM, 1995) is unjustified and can be highly inefficient.

Throughout this study there has been no validation of the model structure through some quantitative test criterion. However, the models used are (i) simple, (ii) based on both theory and observations and (iii) available relevant information is used as much as possible. If the models used are in some instances over-specified, then they contain the ‘true’ model as a special case, and this will result in possibly inefficient, but not invalid estimates (e.g. Cressie and Zimmerman, 1992, Sections 3.6, 5.8). But when we consider the reduced models evaluated in section 3.6, the variograms of the full model (Fig. 3.3) reveal already that a variogram pooled over several soil type or land use categories yields unnecessarily inadequate (i.e. non-specific) error estimates. Under-specification of models is inevitable when it is a consequence of limited data availability. If available data suggest that more complex models result in better estimates, then estimates from the simpler, under-specified model are not necessarily wrong, but are inefficient and should be considered as a ‘lost opportunity’ (Switzer, 1992). Considering the case studies in Chapter 5, the simple model is not necessarily wrong, but the more complex model is better—it results in more accurate estimates.

By definition, the exponent of the mean on the log-scale is the geometric mean on the original scale. However, when on the log-scale the mean and the median coincide, then the

geometric mean and median coincide on the original scale, and the exponent of a confidence interval for the mean on the log-scale is a confidence interval for the median on the original scale. This last condition is met when on the log-scale individual samples within a block are symmetrically (e.g. normally, Aitchison and Brown, 1957) distributed. In general, this symmetry condition on the log-scale seems reasonably mild, given the information at hand.

6.4 Conclusions

The main conclusions are:

- Despite the large estimation inaccuracies resulting from natural variation and limited data availability, groundwater quality variables that are measured in the national and provincial groundwater quality monitoring networks can be mapped at the spatial resolution of 4 km × 4 km blocks when (i) soil type and land use map information is taken into account, (ii) suitable measurements from available monitoring networks are selected, and (iii) attention is paid to local variation in level, size of spatial variation, and spatial dependence in spatial variation of measurements
- The effects of stratification by soil type and land use are profound, partial or no stratification results in essentially different maps
- The effects of monitoring network density, as shown by comparing maps resulting from a reduced monitoring network with the maps from the full network, are considerable
- Spatially differentiated statistical indications of long-term, steady changes in groundwater quality variables can be obtained when short-term predictions, obtained for each monitoring well screen by a linear regression model, are used in the interpolation procedure, and when location-specific time-prediction inaccuracies are taken into account
- Maps of groundwater quality can be improved by using ancillary information in the estimation procedure, where the ancillary information may consist of information from deterministic process models, or of measurements of related variables
- It is expected that maps of groundwater quality and of changes of groundwater quality can be improved further when the relevant information on groundwater flow is used in the mapping procedure

6.5 Future directions

This section lists a number of subjects that have not been studied in depth, but which are worthy of future research. Partly they are directly related to mapping groundwater quality, partly they are mathematical or statistical items, relevant for the methods used.

Incorporating geohydrological information

The most obvious extension of the present study is to use information on groundwater flow for mapping groundwater quality variables. This would reveal a better relation between measurements and maps with soil type and land use information, since groundwater flow causes the influence (infiltration) area (Section 1.3) to be somewhere ‘upstream’ from the monitoring well. With detailed flow information available for monitoring well locations as well as for mapping locations (Section 2.3), connections between deterministic models and measurements can be made at a finer spatial resolution (Section 5.8), and this would allow a better prediction (more accurate estimation) of the groundwater quality at unsampled locations, blocks or areas. Other relevant geohydrological information could be information about the subsurface sedimentary environment (in addition to soil maps describing the top soil) and information about chemical balances found there (e.g. redox potential, calcium or oxygen availability, pH).

Detection limits

Groundwater quality variables have a threshold below which measurements are unreliable. A measurement below this threshold is a ‘below detection limit measurement’ (BDL). Some groundwater quality variables measured in the groundwater quality monitoring networks have a high fraction (e.g. 50-75%) of BDLs. Several methods have been designed for choosing a value between zero and the detection limit so that estimation of the sample mean or variance is unbiased, usually assuming that the distribution of the measurements is known (e.g. Haas and Scheff, 1990). The problem of applying any of these methods in a geostatistical context is that the sample variogram will not reveal the spatial dependence of the true, underlying variable.

In this study the value of the detection limits is assigned to all BDLs because it is the highest value possible. The disadvantage of this choice is that it leads to underestimation of variogram values and consequently to too optimistic estimation errors, and it is not considered as a satisfactory solution. When variables are modelled on a log-scale, the effect of the values assigned to BDLs is profound: assigning a small fraction of the detection limit to BDLs will yield a variogram that is dominated by the spatial distribution of BDLs versus non-BDLs.

Measurement error reduction

Measurement error in a groundwater quality variable can be considered as a source of variation that causes part of the nugget variance, the variogram value at nearly-zero distance. It only needs explicit quantification when estimates of single measurements are needed at measurement locations (Christensen, 1991, VI.4), and therefore it is not explicitly considered in this study. Furthermore, it is unlikely that measurement error of this kind can be reduced.

The independent variables (e.g. soil type and land use in Chapter 3, nitrate leaching rate in section 5.4) are also prone to measurement error. A larger measurement error in the independent variables will always result in a weaker relation between groundwater quality variables and independent variables. If the independent variables are continuous and their measurement error is (i) independent, normally distributed with zero mean, (ii) independent of the true (underlying) values, and (iii) independent of the groundwater quality measurements, then this measurement error will not invalidate the estimates of the groundwater quality variables (Fuller, 1987). Similar conditions may be expressed for categorical independent variables. Reducing a measurement error of this kind will lead to models with smaller residual variations, and increase the accuracy

of estimates (Section 5.8).

Using dominant soil-land use for $4 \text{ km} \times 4 \text{ km}$ blocks instead of actual soil-land use induces a measurement error: when matching this new variable with the measurements, part of the measurements are misclassified (i.e. at measurement locations where dominant soil-land use differs from actual soil-land use), resulting in larger within-class variances than the within-class variances that would have been found when actual soil-land use were used (Section 3.8). An enhancement of this procedure that would eliminate most of these misclassifications would be to use the actual soil-land use at the monitoring well locations (preferably of the influence area of the monitoring well, at the time of infiltration) as obtained from the original soil type and land use maps, to separately estimate the groundwater quality variables for every soil-land use category in a $4 \text{ km} \times 4 \text{ km}$ block, and to calculate the final estimates by weighing separate estimates with the area fractions of each soil-land use category in a block. This approach will be especially useful in cases where dominant soil-land use loses its relevance (i.e. when estimates are needed for averages of areas that are larger or more heterogeneous than the $4 \text{ km} \times 4 \text{ km}$ blocks considered in Chapters 3-5), but may yield more accurate estimates for $4 \text{ km} \times 4 \text{ km}$ block median values as well.

Approximate confidence intervals

Within the context of a valid model structure, confidence intervals obtained by substituting $\sigma_k(x_0)$ for $\sigma(x_0)$ in (2.10) are valid 95% confidence intervals only when (i) the estimation error is normally distributed and (ii) Γ , γ_0 and $\gamma_e(B_0, B_0)$ are known. In practice, (i) cannot be checked experimentally because block averages on the log-scale cannot be measured and (ii) is not true when these quantities are estimated from sample data. This leads to the conclusion that in this case intervals calculated with (2.10) result in *approximate* 95% confidence intervals that have a true confidence level less than 95% (i.e. biased confidence intervals, cf. Rao, 1973).

In this study, we use these approximate 95% intervals because currently no efficient and generally accepted alternatives for obtaining ‘better’ 95% confidence intervals are available. Attempts in the direction of solving (ii) in a Bayesian framework are for instance given by Kitanidis (1986) and Handcock (1994), whereas Brus (1993) presents a design-based approach for this problem. Obviously modelling variograms from smaller samples decreases the true confidence level for intervals obtained with (2.10).

Change of support and non-linear transformations

Block averages can be estimated with linear universal kriging but estimating block averages from a non-linear model (e.g. a linear model for non-linearly transformed measurements) is not trivial (Cressie, 1991 (5.2), Cressie, 1992, Myers, 1994). In section 3.8 an approximate solution for lognormal variables is proposed, but to be effective it may be necessary to use either Cressie’s (1992) adjusted variance-unbiased estimator to obtain better estimates of block averages, or an approach based on averaging conditionally simulated point values in a block (e.g. Journel, 1992).

Space-time geostatistics and uncertain observations

In most geostatistical applications the starting point is the availability of measurements that are possibly prone to a measurement error having constant variance, allowing the use of a

stationary model. With time-predicted observations, we cannot use this as a starting point because time-prediction error variances are typically very diverse (non-stationary). In Chapter 4 simple methods for handling uncertain time-predictions in block kriging were proposed, essentially assuming that the time-prediction error has known variance and is independent from the underlying spatial variable. This should be considered as a starting point for more relevant approaches in space-time geostatistics.

Summary

1. *General introduction*

Groundwater quality is the suitability of groundwater for a certain purpose (e.g. for human consumption), and is mostly determined by its chemical composition. Pollution from agricultural and industrial origin threatens the groundwater quality in the Netherlands. Locally, this pollution is measured at tens of metres depth. Since groundwater is the main source for fresh water, this pollution causes a decrease in the long-term resources of water suitable for human consumption.

In order to get insight into the current situation of groundwater quality and systematic changes of groundwater quality over time, the national groundwater quality monitoring network was established from 1978 to 1984. This network consists of 370 permanent wells, spread fairly evenly over the country (Fig. 3.2, page 27), with screens at 8-10 and 23-25 metre below the soil surface. The well screens are sampled yearly. More recently, the provinces have installed similar monitoring networks that doubled the measurement density. Because the monitoring networks are a major financial investment, the question arises whether the information on groundwater quality, as obtained from the current monitoring networks, is sufficient. This calls for the quantification of what can be inferred from this information about the quality of *all* the groundwater of interest.

For the modelling of the spatial and temporal variation in groundwater quality, using a physically and chemically based deterministic model would call for information on many variables (e.g. initial and boundary conditions, model parameters), that are at present not available on a national scale. However, mapping groundwater quality is possible by using much simpler models, that lump much of the unknown factors into a spatially dependent stochastic term.

The objectives of this study are to

- map groundwater quality in the Netherlands, using available measurements from the national and provincial groundwater quality monitoring networks and map information on soil type and land use
- show the effects of monitoring network density and the effects of using soil type and land use information on the resulting groundwater quality maps
- map the systematic, temporal changes in groundwater quality in a way similar to the mapping of groundwater quality
- show how groundwater quality maps can be improved by using relevant ancillary information in the estimation procedure, where ancillary information is obtained from deterministic process models or from other measured variables.

The primary aim is, with these objectives, to answer the basic questions of describing the

current situation and systematic temporal changes of groundwater quality for the whole of the Netherlands. In this study we use fairly simple models that allow an explicit quantification of the accuracy of resulting estimates. This accuracy is taken into account in the resulting maps, anticipating the question about the value of the current monitoring networks for inferring current situation and systematic changes in time of groundwater quality.

2. *Statistical mapping*

Traditionally, maps show directly observed phenomena. When they do so, most often the observation accuracy does not exceed the display accuracy because of the scaling applied. In the environmental sciences however, it is very common that the ‘observations’ shown on the map do not directly portray observed phenomena but quantities that are only known approximately, e.g. estimates of some quantity based on limited sample information. In order not to suggest more ‘knowledge’ on a map than we have, in the latter case it is necessary to limit the display accuracy.

When observations are available, statistical models can be used for the estimation of the value of a variable at an unsampled location or the mean value in a prespecified area, and for the assessment of estimation accuracy. Geostatistical models—a particular class of linear models with dependent observations—are specifically suited for the mapping problem, i.e., they are suited for estimating the value at a *specific* location, or the mean value of a *specific* area. These models can be viewed as a generalization of the most commonly used statistical models, which are based on independent observations.

In order not to invite the reader of a map to ignore the accuracy information, estimate and estimation accuracy should be presented combined on the map as a confidence interval. This was done by showing the position of the confidence intervals relative to each of four reference levels in four separate sub-maps (e.g. Fig. 3.4, page 30). An alternative display form is to show both sides of the interval in a single map that reduces to the classical, ‘deterministic’ map where confidence intervals are completely covered by a legend class interval (e.g. Fig. 3.5, page 31).

The general approach to mapping groundwater quality variables is that within soil-land use categories, after a careful data selection, estimates for 4 km × 4 km blocks are obtained by local ordinary kriging on the log-scale. After back-transformation to the original scale, approximate 95% confidence intervals for block median concentration levels are shown in maps. This procedure is slightly modified to account for location specific (i.e. non-stationary) time prediction variances in Chapter 4, and to improve estimates by using ancillary information in Chapter 5.

3. *Maps of the groundwater quality in the Netherlands at 5-17 metre depth in 1991*

Maps of 25 groundwater quality variables were obtained by estimating 4 km × 4 km block median concentrations (Pebesma and De Kwaadsteniet, 1994). Estimates were presented as approximate 95% confidence intervals related to 4 concentration levels mostly obtained from critical levels for human consumption. All maps were based on 425 measurements from 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. Regional differences in mean level and in spatial dependence between measurements 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 measurements of the provincial monitoring networks. Important changes in resulting maps are assigned to loss of information on short distance variation, as well as loss of location specific information.

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 units used in regional groundwater models. The maps enable location specific network optimization. In general, the maps give little reason for reducing monitoring network density (wide confidence intervals). In Chapter 3 the methods used are illustrated by results on aluminium concentration in groundwater.

4. Maps of temporal changes in groundwater quality in the Netherlands at 5-17 metre depth

For 12 groundwater quality variables statistical maps of temporal changes were made, focusing on the long-term components of temporal changes (Pebesma and De Kwaadsteniet, 1995). For each monitoring well screen, short-term predictions were calculated from time series of yearly measurements over 5-7 years, by using a simple regression model. Within each soil-land use category these predictions were interpolated spatially, taking location specific prediction accuracies, local level, and the size and structure of spatial variation into account. Spatial estimates concern 4 km × 4 km block median concentrations. For two extrapolation times (1980 and 2000) estimates of block median values are shown in maps as 95% confidence intervals, related to four levels that were derived from critical levels for human consumption.

The resulting maps (e.g. Fig. 4.4 and 4.5 on pages 52 and 53) 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 cells with a pronounced trend can be distinguished from a neighbourhood with less pronounced trends, and because trends in adjacent or nearby cells (also within the same soil-land use category) can have opposite directions. Because of the contents and foundation of several aspects, some caution is required for the interpretation of the maps, and the maps reflect only a first step in the direction of spatially differentiated statistical time-trends. Reduction of the—now yearly—measurement frequency is discouraged. Serious modifications of the maps in this study may be expected when, after a few years, measurement series from the provincial groundwater quality monitoring networks become available and can be incorporated in the study. In Chapter 4 the methods used are illustrated by results on potassium concentration in groundwater.

5. Improving estimates with ancillary information

Maps with estimates of groundwater quality variables, as previously obtained by ordinary kriging within soil-land use categories can be improved by incorporating relevant ancillary information in the interpolation procedure. Typically, this ancillary information is obtained from deterministic models of the process causing the variation in the groundwater quality variable, or from measurements on a related variable, the variation of which was caused by the same process. Incorporation of the first type of ancillary information leads to the use of the universal kriging model with base functions that have a causal influence upon the measured variable, the second

type leads to the use of cokriging.

In three examples it is shown how estimates of zinc, nitrate and sulphate concentration in groundwater improved by adding information from a model of atmospheric deposition for zinc, information from a model of nitrate leaching from the soil, and sulphate measurements from shallow groundwater. Especially nitrate estimates improved considerably. Since the size of the elements modelled with deterministic process models is much larger than the size of the area of influence for monitoring screens, further improvement may be expected when, at the measurement sites, the process models are adjusted to the latter size.

6. *Discussion and conclusions*

Optimizing the monitoring network with respect to measurement intensity would require knowledge of the loss (or cost) due to 'not monitoring' as a function of monitoring intensity. Assessing such a function is not trivial: several parties would have to agree on it. Because in this study 'degree of knowledge' is expressed as the ability to distinguish estimates from critical concentration levels (target levels or maximum tolerated levels for human consumption), it can directly be interpreted as the degree of knowledge of groundwater *quality*, and a meaningful loss function can be derived from it. Chapter 5 showed that improving estimates with ancillary information may play a role that is similar to increasing the measurement intensity when the method for interpretation of the measurements is fixed.

Flow information or other information from deterministic groundwater quality models could readily be used in the estimation procedure to improve estimates. Alternatively the maps that resulted from this study could be used for the development of such models. Several aspects of the methods as used in this study can be improved. When estimates of groundwater quality variables are required at another spatial resolution than the current 4 km × 4 km blocks, slight modifications of the estimation procedure applied here may be required. The main conclusions are listed and directions for future research are given.

Samenvatting in het Nederlands

HET KARTEREN VAN DE GRONDWATERKWALITEIT IN NEDERLAND

1. Algemene inleiding

Grondwaterkwaliteit is de geschiktheid van grondwater voor een bepaald doel (bijvoorbeeld voor menselijke consumptie), en wordt hoofdzakelijk bepaald door de chemische samenstelling van grondwater. De grondwaterkwaliteit in Nederland wordt bedreigd door vervuiling van agrarische en industriële oorsprong. Lokaal wordt deze vervuiling gemeten op meer dan tien meter diepte. Omdat grondwater de belangrijkste bron is voor schoon water (drinkwater), veroorzaakt deze vervuiling op de lange termijn een afname van de voorraad grondwater dat geschikt is voor menselijke consumptie.

Om inzicht te krijgen in de huidige toestand van de grondwaterkwaliteit en de trendmatige veranderingen daarin, werd tussen 1978 en 1984 het landelijk meetnet grondwaterkwaliteit ingericht. Dit meetnet bestaat uit ongeveer 370 regelmatig over Nederland verspreide meetputten (figuur 3.2, pag. 27), die jaarlijks worden bemonsterd op ongeveer 10 en 25 meter beneden het maaiveld. In het begin van de jaren '90 zijn de provincies begonnen met de inrichting van vergelijkbare meetnetten, teneinde een verhoogde (thans verdubbelde) ruimtelijke dichtheid te bewerkstelligen. Omdat de meetnetten een aanzienlijke financiële investering betekenen, is het belangrijk te evalueren of de informatie over grondwaterkwaliteit die met de huidige meetnetten wordt verzameld, voldoende is. Dit vraagt om de kwantificering van de mate waarin de kwaliteit van *al* het grondwater waarin we geïnteresseerd zijn kan worden afgeleid uit deze informatie.

Voor het modelleren van ruimtelijke en temporele variatie in grondwaterkwaliteit met behulp van fysische en chemische deterministische modellen, is het noodzakelijk vele variabelen (bijvoorbeeld initiële voorwaarden en grensvoorwaarden, systeemparameters) te kennen. Op dit moment zijn deze variabelen niet beschikbaar op een nationale schaal. Het is echter mogelijk grondwaterkwaliteit te karteren door gebruik te maken van eenvoudigere modellen, waarbij de onbekende factoren worden samengevoegd tot een ruimtelijk afhankelijke stochastische term. Het doel van deze studie is om

- de grondwaterkwaliteit in Nederland te karteren, gebruik makende van metingen van nationale en provinciale meetnetten grondwaterkwaliteit en kaartinformatie aangaande landgebruik en bodemtype
- de effecten van meetnetdichtheid en de effecten van het gebruik van informatie omtrent landgebruik en bodemtype op de resulterende kaartbeelden te laten zien
- de monotone, langjarige systematische veranderingen in grondwaterkwaliteit in beeld te brengen op een wijze die nauw aansluit bij de wijze waarop de grondwaterkwaliteit in beeld

werd gebracht

- te laten zien hoe grondwaterkwaliteitskaarten kunnen worden verbeterd door relevante aanvullende informatie in de schattingsprocedure te betrekken, waarbij de aanvullende informatie afkomstig is van deterministische procesmodellen of van andere gemeten variabelen

Het primaire doel is om door deze doelen de hoofdvragen omtrent de huidige toestand van de Nederlandse grondwaterkwaliteit en de trendmatige veranderingen daarin te beantwoorden. In deze studie worden betrekkelijk simpele modellen gebruikt die het mogelijk maken om de nauwkeurigheid van schattingen expliciet te kwantificeren. Deze nauwkeurigheid is in de resulterende kaartbeelden betrokken, vooruitlopend op de vraag in hoeverre de kwaliteit van *al* het grondwater waarin we geïnteresseerd zijn kan worden afgeleid uit de huidige meetnetinformatie.

2. *Statistisch karteren*

Normaalgesproken tonen kaarten verschijnselen die direct waargenomen zijn. Wanneer dit zo is, dan overtreft de afbeeldingsnauwkeurigheid door de toegepaste schaling gewoonlijk niet de waarnemingsnauwkeurigheid. In milieukundige toepassingen is het echter normaal dat de in een kaart afgebeelde ‘waarnemingen’ niet betrekking hebben op direct waargenomen verschijnselen maar op verschijnselen die slechts bij benadering bekend zijn, zoals schattingen van een grootheid gebaseerd op beperkte steekproefinformatie. Om in dit geval niet meer ‘kennis’ te suggereren dan voorhanden is, is het noodzakelijk om de afbeeldingsnauwkeurigheid te beperken.

Wanneer waarnemingen beschikbaar zijn, dan kunnen statistische modellen worden gebruikt om een schatting te maken van de waarde van de waarnemingsvariabele op een niet-bemeten locatie of van het gemiddelde over een bepaald gebied, en om een schattingsnauwkeurigheid hieraan toe te kennen. Geostatistische modellen — een specifieke klasse van lineaire modellen met gecorreleerde waarnemingen — zijn speciaal geschikt voor het karteren van een variabele: het schatten van de waarde van een variabele op een specifieke locatie of van het gemiddelde over een specifiek gebied. Geostatistische modellen kunnen worden gezien als een generalisatie van de statistische modellen die zijn gebaseerd op onafhankelijke waarnemingen.

Om de lezer van een kaart niet uit te nodigen de nauwkeurigheidsinformatie te negeren is het noodzakelijk schatting en schattingsnauwkeurigheid in een kaart gecombineerd weer te geven als een betrouwbaarheidsinterval. Dit is gedaan door de relatieve ligging van het betrouwbaarheidsinterval ten opzichte van elk van vier referentieniveaus af te beelden in één van vier deelkaarten (zie het voorbeeld van aluminium, figuur 3.4, pag. 30). Een alternatief is om beide zijden van het interval in één enkele kaart af te beelden, die reduceert tot de klassieke, ‘deterministische’ kaart waar een betrouwbaarheidsinterval geheel binnen een legendaklasse valt (figuur 3.5, pag. 31).

In grote lijnen is de volgende procedure voor het in kaart brengen van de grondwaterkwaliteit gehanteerd. Binnen categorieën bodemtype en landgebruik werden vanuit zorgvuldig geselecteerde waarnemingen schattingen voor $4 \text{ km} \times 4 \text{ km}$ -ruitgemiddelden verkregen door gebruik te maken van *ordinary kriging* op de log-schaal. Na terugtransformatie naar de oorspronkelijk-

ke schaal werden benaderende 95% betrouwbaarheidsintervallen voor ruitmediane concentraties verkregen, welke zijn afgebeeld in kaarten. Deze procedure is enigszins aangepast om rekening te houden met locatiespecifieke (niet-stationaire) tijd-voorspellingsvarianties in hoofdstuk 4, en om schattingen te verbeteren door gebruik te maken van aanvullende informatie in hoofdstuk 5.

3. Kaarten van de grondwaterkwaliteit in Nederland op 5-17 meter diepte in 1991

Van 25 grondwaterkwaliteitsvariabelen werden kaarten gemaakt door ruitmediane concentraties voor $4 \text{ km} \times 4 \text{ km}$ -ruiten te schatten (Pebesma en De Kwaadsteniet, 1994). Schattingen werden gepresenteerd in de vorm van benaderende 95% betrouwbaarheidsintervallen, gerelateerd aan een viertal veelal aan normstelling ontleende referentieniveaus. De kaarten zijn gebaseerd op 425 metingen uit landelijk en provinciale meetnetten grondwaterkwaliteit. De schattingsprocedure is gebaseerd op een stratificatie naar bodemtype en landgebruik. Binnen elke categorie werden de waarnemingen geïnterpoleerd, rekening houdend met regionale verschillen in niveau en in ruimtelijke afhankelijkheid tussen waarnemingen. De gebruikte stratificatie is essentieel gebleken: het niet of slechts partieel uitvoeren van de betreffende stratificatie leidt tot wezenlijk afwijkende kaartbeelden. De effecten van meetnetdichtheid zijn onderzocht door kaartbeelden te genereren op basis van de meetpuntenverzameling van het landelijk meetnet grondwaterkwaliteit alleen. Wezenlijke aanpassingen voortkomend uit toevoeging van de provinciale meetnetten kunnen worden toegeschreven aan de toegenomen informatie aangaande korte-afstand-variatie en de toegenomen hoeveelheid locatiespecifieke informatie.

Beleidsmatig kunnen de kaarten (a) rechtstreeks worden benut enerzijds in relatie tot de meer diffuse vormen van grondwaterverontreiniging, anderzijds voor het verkrijgen van een ruimtelijk gedifferentieerd beeld van achtergrondwaarden waartegen strikt locatiespecifieke verontreinigingssituaties kunnen worden beoordeeld; (b) meer indirect worden benut als invoer- en toetsingsmateriaal voor de de beleidsonderbouwing ten dienste staande regionale en landelijke modellen voor grondwaterkwaliteitsontwikkeling. De kaarten kunnen worden gezien als een vertaling van de uit de meetnetten afkomstige puntinformatie naar informatie op een ruimtelijk schaalniveau waarop bijvoorbeeld gangbare regionale modellering aansluit. Meetnetoptimalisatie kan efficiënt en locatiespecifiek worden gericht op aspecten van het gegenereerde landelijke beeld. In het algemeen kan worden gesteld, dat de gepresenteerde kaarten weinig aanleiding geven tot meetnetreductie (brede betrouwbaarheidsintervallen). In hoofdstuk 3 wordt de gehanteerde methodiek geïllustreerd met resultaten voor de aluminiumconcentratie in het grondwater (figuur 3.4 en 3.5, pag. 30 en 31).

4. Kaarten van veranderingen in de grondwaterkwaliteit in Nederland op 5-17 meter diepte

Van een 12-tal grondwaterkwaliteitsparameters werden landsdekkende statistische kaartbeelden van veranderingen in de tijd gepresenteerd (Pebesma en De Kwaadsteniet, 1995). In het bijzonder ging de belangstelling daarbij uit naar het onderkennen van monotone langjarige componenten in het veranderingsbeeld. Aan de basis werden, aansluitend bij langjarige waarnemingstijdreeksen in een 250-tal meetpunten van het landelijk meetnet grondwaterkwaliteit, via regressieanalyse korte-termijn-voorspellingen in de betreffende meetpunten bepaald, die vervolgens ruimtelijk werden geïnterpoleerd. Schattingsonnauwkeurigheid van de korte-termijn-voorspellingen in de meetpunten werd mede in beschouwing genomen. Bij ruimtelijke interpolatie werd expliciet rekening gehouden met het (locatiespecifieke) niveau, de grootte van ruimtelijke variatie en de

ruimtelijke samenhang in de ruimtelijke variatie van een te interpoleren variabele. Ruimtelijke interpolatie vond plaats per categorie landgebruik/bodemtype. De ruimtelijke resolutie van de gepresenteerde kaartbeelden is beperkt tot $4 \text{ km} \times 4 \text{ km}$ -ruiten. In de kaarten worden, voor twee extrapolatietijdstippen (1980 en 2000), schattingen van ruitmediane waarden gegeven in de vorm van benaderende 95% betrouwbaarheidsintervallen, gerelateerd aan een 4-tal parameterafhankelijke, veelal aan normstelling ontleende niveaus.

De resulterende kaarten (kaliumconcentratie in figuren 4.4 en 4.5, pag. 52 en 53) laten een rijkdom aan trend-indicaties zien. Primair kan worden geconcludeerd dat een sterk ruimtelijk gedifferentieerde trend-benadering haalbaar en zinvol is; zinvol omdat (i) ruiten met een sterk geprononceerde trend daarbij kunnen worden onderscheiden van een minder geprononceerde omgeving en (ii) trends in onmiddellijk aangrenzende of nabijgelegen ruiten (ook binnen een zelfde categorie landgebruik/bodemtype) van tegengestelde richting kunnen zijn. Voorzichtigheid bij de interpretatie van de kaartbeelden is echter, gezien de inhoud en onderbouwing van een aantal aspecten, geboden; de kaartbeelden weerspiegelen slechts een eerste stap in de richting van een ruimtelijk gedifferentieerde statistische trend-benadering. In hoofdstuk 4 wordt de gehanteerde methodiek geïllustreerd met resultaten voor de kaliumconcentratie in het grondwater.

Het terugbrengen van de nu gangbare jaarlijkse bemonsteringsinspanning naar een meer-jaarlijkse dient te worden ontraden. Wezenlijke aanpassingen van het gepresenteerde landsdekkende beeld mogen worden verwacht, wanneer met de tijd langjarige meetreeksen in de meetpunten van de provinciale Meetnetten Grondwaterkwaliteit ter beschikking komen en deze mede in de beschouwingen worden betrokken.

5. *Het verbeteren van schattingen met behulp van aanvullende informatie*

Kaarten met schattingen voor grondwaterkwaliteitsvariabelen zoals deze eerder werden verkregen door middel van *ordinary kriging* binnen bodemtype/landgebruik categorieën kunnen worden verbeterd door relevante aanvullende informatie in de interpolatieprocedure te betrekken. Deze aanvullende informatie is ontleend aan deterministische modellen van het proces dat de variatie in de grondwaterkwaliteitsvariabele veroorzaakte, of aan metingen aan een gerelateerde variabele, van welke de variatie werd veroorzaakt door hetzelfde proces. Het betrekken van het eerste type aanvullende informatie in de interpolatieprocedure leidt tot het gebruik van *universal kriging* met basis-functies die een causale verband hebben met de gemeten variabele, het betrekken van het tweede type leidt tot het gebruik van *cokriging*.

In drie voorbeelden wordt getoond hoe schattingen van zink-, nitraat- en sulfaatconcentratie in het grondwater verbeterden door respectievelijk modelinformatie omtrent atmosferische depositie van zink, modelinformatie omtrent nitraatuitspoeling uit de bodem, en door meetinformatie omtrent sulfaat in ondiep grondwater te betrekken in de schattingsprocedure. Met name de nitraatschattingen verbeterden aanzienlijk. Omdat de omvang van de elementen waarop de deterministische procesmodellen betrekking hebben vele malen groter is dan de omvang van de invloedsgebieden voor waarnemingsfilters, is een verdere verbetering te verwachten wanneer op de waarnemingslocaties het procesmodel wordt aangepast aan de omvang van de invloedsgebieden.

6. *Discussie en conclusies*

Het optimaliseren van het meetnet met betrekking tot de meetintensiteit (ruimtelijke dichtheid, meetfrequentie) vergt de bepaling van de verlieskosten als gevolg van ‘niet meten’ als een functie van de meetintensiteit. Het vaststellen van een dergelijke functie is niet triviaal: verschillende partijen zouden hierover consensus moeten bereiken. Omdat in deze studie de ‘mate van kennis’ wordt uitgedrukt in termen van de mogelijkheid om schattingen te onderscheiden van kritische concentratieniveaus (streefwaarden, maximaal toelaatbare waarden), kan dit direct als de mate van kennis omtrent grondwaterkwaliteit worden geïnterpreteerd, en kan een betekenisvolle verlieskostenfunctie hiervan worden afgeleid. In hoofdstuk 5 werd getoond dat het verbeteren van schattingen met behulp van aanvullende informatie een rol kan vervullen die vergelijkbaar is met het verhogen van de meetintensiteit onder een vastgelegde wijze van interpretatie van de metingen.

Informatie omtrent grondwaterstroming of andere informatie vanuit deterministische procesmodellen zouden enerzijds eenvoudig in de schattingsprocedure kunnen worden betrokken om schattingen te verbeteren, anderzijds kunnen de kaarten uit deze studie worden gebruikt voor de ontwikkeling van dergelijke modellen. Verschillende aspecten van de gebruikte methoden kunnen worden verbeterd. Voor schattingen van grondwaterkwaliteitsvariabelen op een andere ruimtelijke resolutie dan de $4 \text{ km} \times 4 \text{ km}$ -ruiten kunnen kleine aanpassingen in de schattingsprocedure noodzakelijk zijn. De belangrijkste conclusies werden getoond en richtingen voor toekomstig onderzoek werden genoemd.

Appendix A

Tables

Table A1. 25 Groundwater quality variables

Al, As, Ba, Ca, Cd, Cl, Cr, Cu, DOC, EC, Fe, pH, HCO₃, K, Mg, Mn, Na, NH₄-N, Ni, NO₃-N, Pb, P-tot, SO₄, Sr, Zn

Table A2. The sample variogram

The sample variogram, $\hat{\gamma}_Z(h_i)$ is calculated from observations $z(x)$ by:

$$\hat{\gamma}_Z(h_i) = \frac{1}{2N_{h(i)}} \sum_{j,k} [z(x_j) - z(x_k)]^2 \quad (\text{A1})$$

where

$z(x_j)$ the observation at location x_j

and, for every i ($i = 1..q$):

$\sum_{j,k}$ sum for all observation pairs $z(x_j), z(x_k)$ in the category for which
 $h_{jk} = \|x_j - x_k\| \in [(i-1)\Delta x, i\Delta x)$

h_i the mean value of the corresponding h_{jk} 's

$N_{h(i)}$ number of observation pairs in the i -th summation

$q, \Delta x$ depending on the number of observations (in Chapter 3, $q = 20$ and $\Delta x = 5000$)

Table A3. Three variogram models

name	equation	code
nugget	$\gamma(h) = 0, h = 0$ $\gamma(h) = a, a \text{ constant}, h > 0$	a Nug(0)
spherical	$\gamma(h) = b(3h/2c - h^3/2c^3), b, c \text{ constant}, 0 \leq h \leq c$ $\gamma(h) = b, h > c$	b Sph(c)
exponential	$\gamma(h) = d(1 - \exp(-h/e)), h \geq 0$	d Exp(e)

Table A4. 9 Groundwater quality variables

Al, Ca, Cd, Cu, K, NO₃-N, Pb, P-tot, Zn

Table A5. 12 Groundwater quality variables

Ca, Cl, EC, pH, HCO₃, K, Mg, Na, NH₄-N, NO₃-N, P-tot, SO₄

Table A6. The residual sample variogram

When $z(x)$, the n measurements $z(x_j)$ can be represented by a linear model in p base functions $f_p(x)$

$$Z(x_j) = \sum_{i=1}^p f_i(x_j)\beta_i + e(x_j), \quad \forall j = 1..n,$$

ordinary least squares (OLS) residuals are calculated by

$$\hat{e}(x_j) = z(x_j) - f(x_j)\hat{\beta},$$

with $\hat{\beta} = (F_x'F_x)^{-1}F_x'z(x)$ where $\beta = (\beta_1, \dots, \beta_p)'$, $F_x = (f_1(x), \dots, f_p(x))$ and $f_i(x) = (f_i(x_1), \dots, f_i(x_n))'$. For model (5.2) and the examples A and B, $p = 2, f_1(x) = 1, \forall x$, and $f_2(x)$ is the $f(x)$ in (5.2). The sample residual variogram $\hat{\gamma}_{\hat{e}}$ is calculated as in Table A2 by substituting the OLS residuals $\hat{e}(x_j)$ for observations $z(x_j)$.

Table A7. The sample cross variogram

The sample cross variogram $\hat{\gamma}_{12}(h_i)$ for two observation variables $z_1(x)$ and $z_2(x)$ is calculated by

$$\hat{\gamma}_{12}(h_i) = \frac{1}{2N_{h(i)}} \sum_{j,k} [(z_1(x_j) - \hat{m}_1) - (z_2(x_k) - \hat{m}_2)]^2$$

with \hat{m}_1 and \hat{m}_2 the sample means of $z_1(x)$ and $z_2(x)$, the rest of the symbols as in Table A2.

References

- Aitchison, J., J.A.C. Brown (1957). The lognormal distribution. Cambridge University Press, Cambridge.
- Alkemade, J.R.M., M.L.P. van Esbroek (1994). Naar een effecten voorspellingsmodel voor de bodemfauna: BOEF (*Towards a prediction model of effects on soil fauna*). National institute for public health and environmental protection, report no. 712901001 (in Dutch).
- Appelo, C.A.J., D. Postma (1993). Geochemistry, groundwater and pollution. Balkema, Rotterdam.
- Beven, K. (1985). Changing ideas in hydrology—the case of physically-based models. *Journal of Hydrology* 105, 157-172.
- Boumans, L.J.M., W.H.J. Beltman (1991). Kwaliteit van het bovenste freatische grondwater in de zandgebieden van Nederland, onder bos en heidevelden (*Quality of the upper freatic groundwater in the sand area of the Netherlands, under forests and heather*). National institute of public health and environmental protection, Bilthoven. Report no. 724901001 (in Dutch).
- Boumans, L.J.M. (1994). Nitraat in het bovenste grondwater onder natuurgebieden op zandgrond in Nederland (*Nitrate in the upper groundwater, in semi-natural areas on the sandy area of the Netherlands*). National institute for public health and environmental protection, report no. 712300002 (in Dutch).
- Brus, D.J (1993). Incorporating models of spatial variation in sampling strategies for soil. Doctoral thesis, Wageningen Agricultural University, The Netherlands.
- Burrough, P.A., M. Rikken, R. van Rijn (1995). Spatial data quality and error analysis issues: GIS functions and environmental modelling. In: M.F. Goodchild, B.O. Parks, L.T. Seyaert, Proceedings 2nd international conference on environmental modelling and GIS, Breckenridge, Colorado, Sept. 1993. GIS World Publications (in press).
- CCRX (1994). Metingen in het milieu in Nederland 1992 (*Monitoring the environment in the Netherlands, 1992*). National institute for public health and environmental protection, Bilthoven (in Dutch).
- CCRX (1995). Metingen in het milieu in Nederland 1993 (*Monitoring the environment in the Netherlands, 1993*). National institute for public health and environmental protection, Bilthoven (in Dutch).
- Christensen, R. (1987). Plane answers to complex questions: the theory of linear models. Springer, New York.
- Christensen, R. (1991). Linear models for multivariate, time series and spatial data. Springer, New York.
- Christensen, R. (1993). Quadratic covariance estimation and equivalence of predictions. *Mathematical Geology* 25 (5), 541-558.
- Clark, I., K. Basinger, W. Harper (1989). MUCK—A novel approach to co-kriging. In: B.E.

- Buxton (Ed.), Proceedings of the conference on geostatistics, sensitivity and uncertainty: methods for ground-water flow and radionuclide transport modeling. Batelle Press, Columbus, 473-494.
- Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology* 17 (5), 563-586.
- Cressie, N. (1990). The origins of kriging. *Mathematical Geology* 22 (3), 239-252.
- Cressie, N. (1991). *Statistics for spatial data*. Wiley, New York.
- Cressie, N. (1992). Aggregation in geostatistical problems. In: A. Soares (ed.), *Geostatistics Troia '92* Volume 1, Kluwer, Dordrecht, 25-35.
- Cressie, N., D.L. Zimmerman (1992). On the stability of the geostatistical method. *Mathematical Geology* 24 (1), 45-59.
- De Gruijter, J.J., C.J.F. ter Braak (1990). Model-free estimation from spatial samples: a reappraisal of classical sampling theory. *Mathematical Geology* 22 (4), 407-415.
- De Kwaadsteniet, J.W. (1990). On some fundamental weak spots of kriging technique and their consequences. *Journal of Hydrology* 114, 277-284
- De Vries, F., J. Denneboom (1992). *De bodemkaart van Nederland digitaal (The digital soil map of the Netherlands)*. Technical Report 1, Winand Staring Centre for Integrated Land, Soil and Water Research (in Dutch).
- Delhomme, J.P. (1978). Kriging in the hydrosociences. *Advances in Water Resources* 1, 251-266
- Domenico, P.A., W. Schwartz (1990). *Physical and chemical hydrogeology*. Wiley, New York.
- Draaijers, G. (1993). The variability of atmospheric deposition to forests, the effects of canopy structure and forest edges. Ph.D. thesis, Utrecht University.
- Draper, N, and H. Smith (1981). *Applied regression analysis*, Second edition. Wiley, New York.
- Fisher, R.A. (1935). *The design of experiments*. Oliver and Boyd, Edinburgh.
- Frapporti, G., S.P. Vriend, P.F.M. van Gaans (1993). Hydrogeochemistry of the shallow Dutch groundwater, interpretation of the national groundwater quality monitoring network. *Water Resources Research* 9, 2993-3004.
- Fuller, W.A. (1987). *Measurement error models*. Wiley, New York.
- Haas, C.N., P.A. Scheff (1990). Estimation of averages in truncated samples. *Environmental Sciences Technology*, 24 (6), 912-919.
- Handcock, Mark S. (1994). Measuring the uncertainty in kriging. In: R. Dimitrakopoulos (ed.), *Geostatistics for the next century*, Kluwer, Dordrecht, 436-447.
- Hansen, M.H., W.G. Madow, B.J. Tepping (1983). An evaluation of model-dependent probability-sampling inferences in sample surveys. *Journal of the American Statistical Association* 78, 776-793.
- Harville, D.A. (1985). Decomposition of prediction error. *Journal of the American Statistical Association* 80 (389), 132-138.
- Hastie, T.J., R.J. Tibshirani (1990). *Generalized additive models*. Chapman and Hall, London.
- Hoekstra, J.A., P.S. Heuberger (1995). Reliability of environmental information obtained by modelling and monitoring, In: G.J. Hey, J.W. Erisman (eds.), *Acid Rain Research: Do we have enough answers?* Studies in environmental sciences 64, Elsevier Amsterdam.
- Journel, A.G., Ch.J. Huijbregts (1978). *Mining geostatistics*. Academic Press, London.
- Journel, A.G., M. Rossi (1989). When do we need a trend model in kriging? *Mathematical*

- Geology 21 (7), 715-739.
- Journel, A.G. (1992). Geostatistics: roadblocks and challenges. In: A. Soares (ed.), Geostatistics Troia '92 Volume 1, Kluwer, Dordrecht, 213-224.
- Kitanidis, P.K. (1986). Parameter uncertainty in estimation of spatial functions: bayesian analysis. *Water Resources Research* 22 (4), 499-507.
- Kitanidis, P.K. (1991). Orthonormal residuals in geostatistics: model criticism and parameter selection. *Mathematical Geology* 23 (5), 741-758.
- Kitanidis, P.K. (1993). Generalized covariance functions in estimation. *Mathematical Geology* 25 (5), 525-540.
- Klijn, F. (1988). Milieubeheergebieden. Deel A: indeling van Nederland in ecoregio's en ecodistricten (*Environment management areas. Part A: subdivision of the Netherlands in ecological regions and ecological districts*). Rijksuniversiteit Leiden, mededelingen no. 37 (in Dutch).
- Maas, R.J.M. (ed.) (1992). National environmental outlook 1990-2010. National institute of public health and environmental protection, Bilthoven.
- Matheron, G. (1971). The theory of regionalized variables and its applications. Les cahiers du centre morphologie mathématique de Fontainebleau no. 5. École nationale supérieure des mines de Paris.
- Matheron, G. (1989). Estimation and choosing. Springer, Berlin.
- McCullagh, P., J.A. Nelder (1989). Generalized linear models (2nd ed). Chapman and Hall, London.
- Meinardi, C.R. (1994). Groundwater recharge and travel times in the sandy regions of the Netherlands. Ph.D. Thesis, VU Amsterdam; also published as: National institute of public health and environmental protection, Bilthoven. Report no. 715501004.
- Myers, D. E. (1982). Matrix formulation of cokriging. *Mathematical Geology* 14 (2), 249-257.
- Myers, D.E. (1989). To be or not to be...stationary? That is the question. *Mathematical Geology* 21 (3), 347-362.
- Myers, D. E. (1991). Pseudo-cross variograms, positive-definiteness, and cokriging. *Mathematical Geology* 23 (6), 805-816.
- Myers, D.E. (1994). Change of Support and Transformations. In: R. Dimitrakopoulos (ed.), Geostatistics for the next century. Kluwer, Dordrecht, 253-258.
- Pebesma, E.J. (1992). Een landelijk beeld van grondwaterkwaliteitsvariabelen, deel 1: een onderzoek naar interpolatiemethoden (*Mapping groundwater quality variables, part 1: a study of interpolation methods*). National institute of public health and environmental protection, Bilthoven. Report no. 721801002 (in Dutch).
- Pebesma, E.J. (1995). GSTAT: geostatistical modelling, prediction and simulation. Computer program manual. Available from <ftp://pop.frw.ruu.nl/pcraster>
- Pebesma, E.J., J.W. de Kwaadsteniet (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 metre depth, in 1991*). National institute of public health and environmental protection, Bilthoven. Report no. 714810014 (in Dutch).
- Pebesma, E.J., J.W. de Kwaadsteniet (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 metre depth*). National institute of public health and the environment, Bilthoven. Report no. 714810015 (in Dutch).
- Rao, C.D. (1973). *Linear statistical inference and its applications*, 2nd ed. Wiley, New York.
- RIVM (1992). Milieurapportage 1993, I, Integrale rapportage stikstof (*Environmental report 1993: integral report on nitrogen*). National institute for public health and environmental protection, Bilthoven, report no. 482533001 (in Dutch).
- RIVM (1995). Milieubalans 95 (*Environmental balans 1995*). Samson, Alphen aan de Rijn (in Dutch).
- Rodríguez-Iturbe, I., J.M. Mejía (1974). The design of rainfall networks in time and space. *Water Resources Research*, 10 (4), 713-728
- Searle, S.R. (1971). *Linear Models*. Wiley, New York.
- Stein, A., I.G. Staritsky, J. Bouma, A.C. van Eijnsbergen, A.K. Brecht, (1991). Simulation of moisture deficits and areal interpolation by universal Cokriging. *Water Resources Research* 27 (8), 1963-1973.
- Steur, G.G.L., F. de Vries and C. van Wallenburg (1985). *Bodemkaart van Nederland 1:250 000 (Soil map of the Netherlands 1:250 000)*. Stichting voor Bodemkartering, Wageningen (in Dutch).
- Switzer, P. (1992). The spatial variability of prediction errors. In: A. Soares (ed.), *Geostatistics Troia '92 Volume 1*, Kluwer, Dordrecht, 261-272.
- Thunnissen, H., R. Olthof, P. Getz and L. Vels (1992). Grondgebruiksdatabase van Nederland vervaardigd met behulp van Landsat Thematic Mapper opnamen (*Land use database of the Netherlands constructed with the aid of Thematic Mapper images*). DLO-Staring Centrum, Wageningen. Report no. 168, (in Dutch).
- Van Deursen, W.P.A. (1995). *Geographic information systems and dynamic modelling*. Ph.D. thesis, Utrecht University.
- Van Drecht, G. (1993a). Modelling of regional scale nitrate leaching from agricultural soils, The Netherlands. *Applied Geochemistry*, supplementary issue 2, 175-178.
- Van Drecht, G. (1993b). Berekening van de nitraatbelasting van het grondwater, achtergrond-document bij de Nationale Milieverkenningen 2, 1990-2010 (*Calculation of the nitrogen load of the groundwater, background document for the National environmental outlook 2, 1990-2010*). National institute of public health and environmental protection, Bilthoven. Report no. 714901001 (in Dutch).
- Van Duijvenbooden, W., J. Taat, L.F.L. Gast (1985). Landelijk meetnet grondwaterkwaliteit: eindrapport van de inrichtingsfase (*The national groundwater quality monitoring network: final report of the installation*). National institute of public health and environmental protection, Bilthoven. Report no. 840382001 (in Dutch).
- Van Duijvenbooden, W. (ed.) (1989). *De kwaliteit van het grondwater in Nederland (The quality of the groundwater in the Netherlands)*. National institute for public health and environmental protection, Bilthoven, report. no. 728820001 (in Dutch).
- Van Geer, F.C. (1987). *Applications of Kalman filtering in the analysis and design of groundwater monitoring networks*. Ph.D. thesis, Delft University Press
- Van Jaarsveld, J.A. (1995). *Modelling the long-term atmospheric behaviour of pollutants on various spatial scales*. Ph.D. Thesis, Univ. of Utrecht, The Netherlands.

- Ver Hoef, J.M., N. Cressie (1993). Multivariable spatial prediction. *Mathematical Geology* 25 (2), 219-240.
- Versteegh, J.F.M., F.W. van Gaalen, A.J.H. van Breemen (1995). De kwaliteit van het drinkwater in Nederland, in 1993 (*The quality of the drinking water in the Netherlands, in 1993*). National institute for public health and environmental protection, Bilthoven, report no. 731011007 (in Dutch).
- Ward, R.C., M. Robinson (1989). Principles of hydrology. McGraw-Hill, London.
- Wesseling, C.W., W.P.A. van Deursen (1995). PCRaster user manual. Computer program manual. Available from <ftp://pop.frw.ruu.nl/pcraster>

DANKWOORD

Dit proefschrift was er niet geweest zonder de directe of indirecte hulp en steun van vele anderen, waarvan ik hier een aantal wil noemen. Mijn samenwerking met Hans de Kwaadsteniet gedurende dit project heb ik in veel opzichten als erg inspirerend ervaren. Peter Burrough wil ik bedanken voor de interesse, ondersteuning en kritiek tijdens het onderzoek. Gerard Heuvelink wil ik bedanken voor zijn kritische beoordeling van vroege en late versies van het manuscript, en Gerard Heuvelink en Marc Bierkens voor de vele discussies — waar of wanneer ook.

Verder gaat mijn dank uit naar Cees Wesseling voor de hulp bij het programmeren en naar Peter van Horssen voor de vele ‘plakavonden’. Ook gaat mijn dank uit naar Arthur van Beurden, Rob van de Velde and Jaap Willems voor het zitting nemen in mijn begeleidingscommissie, naar Cees Meinardi en Wil van Duijvenbooden voor de kritische opmerkingen bij hoofdstuk 1, naar Rob Alkemade, Leo Boumans, Gerard van Drecht, Herman Prins, and Hans Reijnders voor de vele discussies over meetnetten en statistiek, en naar Jan Kunst, Roland Lieste en Hans Verlouw voor de hulp bij de moeilijke delen van het GIS-werk. Tot slot bedank ik Ellen voor de voortdurende steun, met name gedurende de zomer van ’95.

CURRICULUM VITAE

Ik ben geboren op 3 juli 1967 te Beetsterzwaag, alwaar ik van 1971 tot 1979 het basisonderwijs genoot. Van 1979 tot 1985 volgde ik Atheneum-B aan 'RSG het Drachtster Lyceum' te Drachten. In 1985 begon ik in Utrecht aan de universitaire opleiding Fysische Geografie aan de toenmalige Interfakulteit Aardrijkskunde en Prehistorie aan de Rijksuniversiteit Utrecht, de huidige Faculteit der Ruimtelijke Wetenschappen aan de Universiteit Utrecht. In januari 1991 heb ik deze studie afgerond met als afstudeerrichting proceskunde. Mijn eindveldwerk betrof het onderwerp 'ruimtelijke en temporele variatie van bodemfysische parameters in de Ardèche,' mijn afstudeerscriptie betrof de steekproefonzekerheid van variogramschatters. Op 1 februari 1991 ben ik bij de vakgroep Fysische Geografie van de Universiteit Utrecht begonnen aan een AIO-onderzoek, gefinancierd door en uitgevoerd bij het Laboratorium voor Bodem- en Grondwateronderzoek van het toenmalige Rijksinstituut voor Volksgezondheid en Milieuhygiëne, thans het Rijksinstituut voor Volksgezondheid en Milieu (RIVM). In oktober 1992 werd mijn voltijdse dienstverband omgezet in een 8/10 dienstverband, waardoor dit doorliep tot 1 september 1995. Vanaf 1 september 1995 tot 1 februari 1996 werd mijn aanstelling verlengd door het RIVM, om een begin te maken met de optimalisatie van de provinciale meetnetten grondwaterkwaliteit. Vanaf 1 februari 1996 werk ik als post-doc bij de vakgroep fysische geografie en bodemkunde van de Universiteit van Amsterdam aan het tweejarige EG-project 'Assessment of cumulative uncertainty in spatial decision support systems: applications to examine the contamination of groundwater from diffuse sources.'