Environ Ecol Stat (2009) 16:321–339
DOI 10.1007/s10651-007-0052-x

External drift kriging of NO_x concentrations with dispersion model output in a reduced air quality monitoring network

Jan van de Kassteele · Alfred Stein · Arnold L. M. Dekkers · Guus J. M. Velders

Received: 1 February 2004 / Revised: 1 April 2006 / Published online: 24 October 2007

Abstract In the mid nineteen eighties the Dutch NO_x air quality monitoring network was reduced from 73 to 32 rural and city background stations, leading to higher spatial uncertainties. In this study, several other sources of information are being used to help reduce uncertainties in parameter estimation and spatial mapping. For parameter estimation, we used Bayesian inference. For mapping, we used kriging with external drift (KED) including secondary information from a dispersion model. The methods were applied to atmospheric NO_x concentrations on rural and urban scales. We compared Bayesian estimation with restricted maximum likelihood estimation and KED with universal kriging. As a reference we also included ordinary least squares (OLS). Comparison of several parameter estimation and spatial interpolation methods was done by cross-validation. Bayesian analysis resulted in an error reduction of 10 to 20% as compared to restricted maximum likelihood, whereas KED resulted in an error reduction of 50% as compared to universal kriging. Where observations were sparse, the predictions were substantially improved by inclusion of the dispersion model output and by using available prior information. No major improvement was observed as compared to OLS, the cause presumably being that much good information is contained in the dispersion model output, so that no additional spatial residual random field is required to explain the data. In all, we conclude that reduction in the

J. van de Kassteele (✉) · A. L. M. Dekkers · G. J. M. Velders
National Institute for Public Health and the Environment - RIVM, P.O. Box 1, 3720 BA Bilthoven, The Netherlands
e-mail: Jan.van.de.kassteele@rivm.nl

J. van de Kassteele · A. Stein
Mathematical and Statistical Methods Group - Biometris, Wageningen University, P.O. Box 100, 6700 AC Wageningen, The Netherlands

A. Stein
International Institute for Geo-Information Science and Earth Observation - ICT, Enschede, The Netherlands

monitoring network could be compensated by modern geostatistical methods, and that a traditional simple statistical model is of an almost equal quality.

Keywords Air pollution · Bayesian inference · Cross validation · Different sample densities · Gaussian model-based geostatistics

1 Introduction

Accurate and spatially highly resolved maps of NO_x levels are essential to assessing individual human exposures to NO_x. It is well known that NO_x in high concentrations causes respiratory problems for humans (EPA 1998; WHO2003). In the Netherlands, NO_x maps are based on a limited number of monitoring stations. In the mid-eighties the Dutch NO_x air quality monitoring network (Van Elzakker 2001) was reduced from 73 to 32 stations for budgetary reasons. The increased distance between stations has caused a substantial loss of information and resulted in higher uncertainties in the maps. The combination of measurements with related external information from a dispersion model is likely to result in more accurate maps. For that purpose, we considered kriging with external drift (KED). This method merges two sources of information: a primary variable that is precise but only available at a small number of locations, and a secondary variable that covers the full domain on a fine-mazed grid. KED has been applied in the past in environmental mapping of sparsely sampled data using dense external information. Examples include combining rainfall data with a digital elevation model (DEM) as a covariate (Goovaerts 2000), combining rainfall data with satellite imagery or radar data (Grimes et al. 1999; Cassiraga et al. 1997), combining temperature with a DEM (Hudson and Wackernagel 1994), and combining temperature and land use for application in a crop growth model (Monestiez et al. 2001). In soil science and hydrology KED is applied in mapping soil horizons using a DEM (Bourennane et al. 1996), soil variables (Hengl et al. 2004), erosion mapping (Goovaerts 1999), or water table depths (Desbarats et al. 2002). Applications in air quality mapping are found in ozone mapping using a DEM (Pauly and Drueke 1996) or using dispersion model output for mapping ozone concentration around Paris (Bertino and Wackernagel 2002) or for urban air quality measurements (Genikhovich et al. 2002).

This study concerns mapping of yearly average atmospheric NO_x concentrations on rural and urban scales in the Netherlands. The purpose of the Dutch air quality monitoring network is to monitor air quality in the Netherlands on a continual basis (Van Elzakker 2001). Measurements from this network provide a general description of national, regional and local air quality, along with information on smog episodes; measurements are also tested against international air quality standards. The size of the area is about 35,000 km^2. Secondary information was provided by the Operational Priority Substances (OPS) dispersion model (Van Jaarsveld 1995), which calculates yearly average concentrations and deposition on the basis of emissions, dispersion, transport, chemical conversion, and wet and dry deposition. OPS also accounts for transport from adjacent countries. The model output consists of a national map with a spatial resolution of 5×5 km.

We explored the use of external drift kriging with the OPS model output in a reduced monitoring network. A comparison was made with universal kriging before (1983) and after (1987) network reduction. Parameter estimation was carried out by means of restricted maximum likelihood and Bayesian inference. Our hypothesis was that Bayesian inference would show lower prediction uncertainties. By cross-validation, we quantified the relationship between the number of stations and occurring errors. A range of errors resulted from selecting several random configurations of different station densities from the 1983 configuration, describing explicitly the effect of the number of stations, and implicitly the effect of the station configuration.

2 Materials and methods

2.1 Observations

The air quality monitoring network has undergone several changes in the past 25 years. A major reorganization took place in 1985/1986, resulting into a reduction in the number of monitoring sites for SO_2 and NO_x measurements. In this study we used observations from 1983, 2 years before the major reduction in 1985. A total of 85 yearly average NO_x concentrations were available. Twelve street stations were omitted since they were not representative on the scale considered in this study. Three regional stations had to be omitted because of non-representative values due to missing data, leaving 64 rural background and six city background stations (Fig. 1a). High concentrations (in ppb) occurred in the western part of the Netherlands, around the major cities of Rotterdam, The Hague, Amsterdam and Utrecht, and near roadways. High concentrations were also found in the south-east, under influence of the German industrial Ruhr area, 50 km east of the Dutch-German border.

After the reduction, the total number of rural, city and street stations in 1987 came to 22, 5 and 7, respectively. A few regional stations had moved. For our analysis of the station configuration in 1987, we matched the existing locations of the stations with

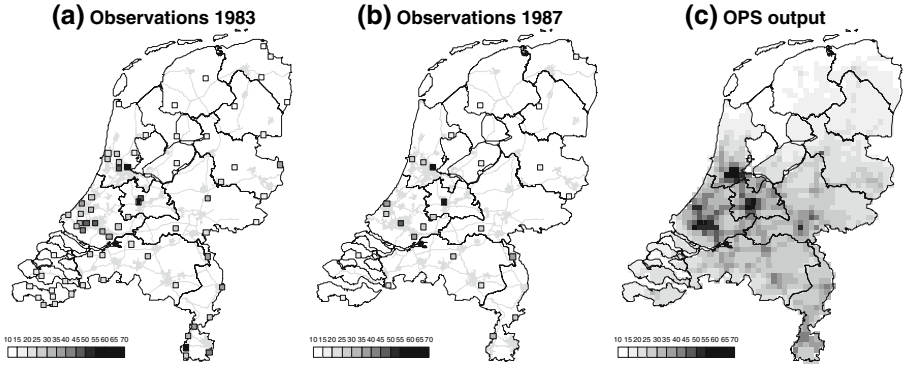

Fig. 1 Measured yearly average NO_x concentrations (ppb) in 1983 (**a**), in 1987 (**b**) and OPS model output for 1983 (**c**). The black lines indicate provincial boundaries, and gray lines and patches major roads and cities

those of 1983 and, where necessary, by selecting the nearest station (Fig. 1b). One regional station was excluded, since there was no possible match with a 1983 regional station. We maintained the concentrations of 1983.

A further change in the network took place in 1994, leading to a limited extension and re-adjustment of some of its nodes. We do not consider effect of thins change in the current article.

2.2 The OPS dispersion model

The OPS model calculates average atmospheric concentrations and deposition from the atmosphere on the basis of emissions within the Netherlands and Europe. The model is suitable for numerous pollutants for which the behavior can be described by first-order linear chemical reactions; it cannot be used, for example, for describing ozone concentrations (Van Jaarsveld 1995).

The processes described by the OPS model are emission, dispersion, transport, conversion, and wet and dry deposition. It uses the Gaussian plume model for dispersion at local scales and operates as a Lagrangian trajectory model for long-distance transport. The OPS model is driven by actual meteorological observations and is statistical in the sense that dispersion is distributed over specific classes according to transport direction, atmospheric stability and scale of transport. Accompanying dispersion parameters are determined according to properties of all trajectories within that class. Yearly average concentration and deposition fields are found by weighting all classes according to the frequency of occurrence. Computationally speaking, this procedure is relatively rapid.

Input consists of emissions from sources into the atmosphere. Source properties like emission height are determinative for the dispersion. Since a detailed emission inventory for 1983 was not available for this study, emissions from 1995 were taken and scaled proportionally to known total emissions per source group for 1983. Output is represented by a concentration field on a 5×5 km grid (Fig. 1c).

The OPS model is described in Van Jaarsveld (1991), Van Jaarsveld and De Leeuw (1993), and Van Jaarsveld (1995). In Van Jaarsveld (1995), the model is compared with measurements on different levels, e.g. process descriptions such as mixing height and descriptions of vertical dispersion. The model has played a role in international comparison studies (Derwent et al 1989). It also generates data at the monitoring station locations (Fig. 2) and does well at predicting yearly average NO_x concentrations, although predictions are systematically higher than observations.

2.3 Universal kriging versus external drift kriging

For kriging measured NO_x concentrations, we applied the model-based approach of Diggle et al. (1998), Ribeiro and Diggle (1999) and Diggle and Ribeiro (2002). After a Box–Cox transformation, the n observations $y(\mathbf{x})$ were interpreted as a realization of a Gaussian random variable $Y(\mathbf{x})$ at location $\mathbf{x}$. This random variable can be decomposed into

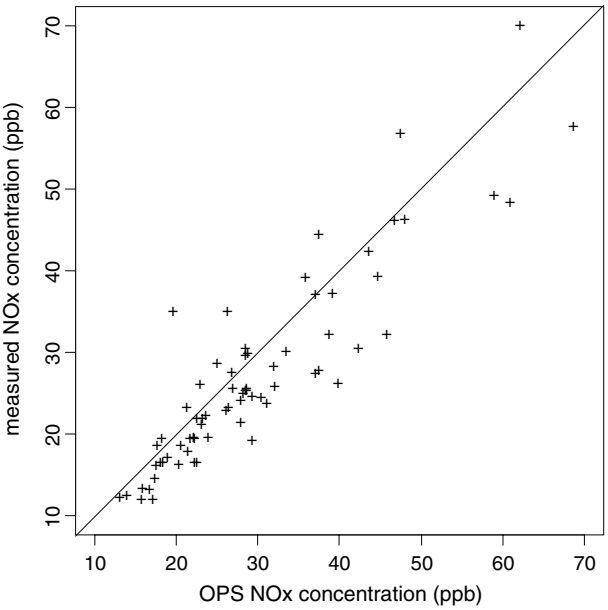

Fig. 2 Measured yearly average NO_x concentration (ppb) versus modeled OPS NO_x concentration (ppb) at the 70 monitoring stations in 1983

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + S(\mathbf{x}) + \varepsilon,$$

where $\mu(\mathbf{x}) = \mathbf{X}\boldsymbol{\beta}$ is the deterministic trend component, $\mathbf{X}$ an $n \times p$ matrix consisting of p known trend components at each location and $\boldsymbol{\beta}$, a vector with p unknown trend parameters. $S(\mathbf{x})$ is a zero-mean stationary Gaussian random process with a partial sill variance, σ^2. The process $S(\mathbf{x})$ accounts for spatial correlation between observations by means of a spatial correlation function $\rho(.)$, with range parameter, ϕ. Finally, ε is an error term with variance τ^2 (nugget). We will use $\tau^2_{rel} = \tau^2/\sigma^2$, the relative nugget, in our further calculations.

Kriging with external drift (KED) is a particular case of universal kriging (UK) (Bourennane et al. 2000). The difference between UK and KED lies in the trend component. For UK, $\mathbf{X}$ is a function of the coordinates x_1 and x_2 in two orthogonal directions, whereas for KED, $\mathbf{X}$ is a function of the OPS model output at location $\mathbf{x}$, i.e. $\mathbf{X} = [1\ x_1\ x_2]$ and $\mathbf{X} = [1\ OPS(\mathbf{x})]$, for UK and KED, respectively. Note that $S(\mathbf{x})$ describes different processes for UK and KED.

2.4 Restricted maximum likelihood versus Bayesian inference

A common method to estimate covariance parameters is fitting the parametric correlation function $\rho(\cdot)$ to the empirical variogram, obtained by binning and averaging the squared differences between residuals after de-trending. Instead, we applied restricted maximum likelihood (RML) and Bayesian inference which estimate parameters directly from the data.

Maximum likelihood estimation provides joint estimation of trend and covariance parameters, but it introduces a bias if the number of observations, n, is small compared to the number of covariates, p. With RML, the trend parameters are integrated out of the likelihood function, leaving unbiased estimators for the covariance parameters (Kitanidis and Shen 1996). RML estimation requires the data to be a realization of a multivariate Gaussian distribution (see Sect. 2.5). If the observations $\mathbf{y}$ and covariates $\mathbf{X}$ are regarded as fixed, the restricted likelihood function will depend only on the unknown parameters σ^2, ϕ, and τ^2_{rel}. The likelihood function is numerically optimized for ϕ, and τ^2_{rel}, because no analytical expression exists for these two parameters.

Bayesian inference treats the parameters as unknown stochastic variables and, as such, incorporates the uncertainty of all parameters (Gelman et al. 2003; Berger et al. 2001). Bayesian inference it attractive in the case of sparse data, since additional prior information can be used, thereby improving the estimation accuracy (see, for example, Cui et al. 1995). A joint prior distribution $p(\boldsymbol{\beta}, \sigma^2, \phi, \tau^2_{rel})$ is assigned to the parameters. The prior is updated by observations with the use of the likelihood function resulting in the posterior parameter distribution. The posterior distribution can be written in analytical form if we use conjugate priors or flat priors for $\boldsymbol{\beta}$ and σ^2, and if ϕ and τ^2_{rel} are known (Ribeiro and Diggle, 1999). If ϕ and τ^2_{rel} are unknown, the posterior distribution is to be factorized to obtain an expression for the joint posterior for ϕ and τ^2_{rel}. The posterior densities for $\boldsymbol{\beta}$ and σ^2 are then easily evaluated by plugging the obtained draws for ϕ and τ^2_{rel} in the analytical expression for the posterior density $p(\sigma^2|\mathbf{y}, \phi, \tau^2_{rel})$, a scaled inverse-$\chi^2$ distribution, and the posterior density $p(\boldsymbol{\beta}|\mathbf{y}, \sigma^2, \phi, \tau^2_{rel})$, a multivariate normal distribution. After drawing from these distributions, the joint posterior parameter distribution is fully evaluated.

The predictive distribution $p(\mathbf{y}_0|\mathbf{y}, \boldsymbol{\beta}, \sigma^2, \phi, \tau^2_{rel})$ is a multivariate normal distribution, that is evaluated analogously by plugging in the obtained parameter samples and, subsequently, drawing samples from it. In conventional geostatistics, the above expression is that for simple kriging, since all parameters are considered known. However, parameter uncertainty is incorporated directly by the draws from the posterior parameter distribution. Prediction using RML only accounts for uncertain trend parameters, but RML and Bayesian inference are similar if a uniform prior for $\boldsymbol{\beta}$ is chosen and the other parameters are kept fixed (i.e. are known).

Bayesian geostatistics are described by Handcock and Stein (1993) and Ribeiro and Diggle (1999). More details on UK and KED can be found for example in Chilès and Delfiner (1999).

2.5 Set-up of the kriging models

The Gaussian model-based approach requires residuals after de-trending to be stationary and normally distributed. We applied the Box–Cox transformation (Box and Cox 1964) of the original data $\mathbf{y}_{org}$ to obtain such residuals:

$$\mathbf{y} = g(\mathbf{y}_{org}) = \begin{cases} (\mathbf{y}^{\lambda}_{org} - 1)/\lambda & (\lambda \neq 0) \\ \log(\mathbf{y}_{org}) & (\lambda = 0) \end{cases}.$$

Here $g(\cdot)$ is the transformation function and λ is the Box–Cox parameter. Maximum likelihood estimates, without taking into account spatial effects, yielded Box–Cox parameters of $\lambda = -0.63$ for UK and $\lambda = 0.34$ for KED. We could have applied a Bayesian estimation of λ, but as in Ribeiro and Diggle (2001), we kept them known throughout. If not, each λ would have changed the location and scale of the transformed data, as well as the correlation structure (De Oliveira et al. 1997). Other transformations, for example, the Normal Score transformation (Lehmann 1975), could also have been applied.

The assumption of normality of the residuals after transformation is strong, since the transformation does not guarantee that the transformed data are normally distributed. There may be a kurtosis for example. Normality cannot be tested in practice because there is only one realization of the random field. However, to get more insight into this, we employed a Shapiro–Wilks test to check for evidence against normality. Spatial correlation in the residuals is not taken into account. The test yielded a P-value = 5.038E-7 and a P-value = 1.470E-4 for UK and KED, respectively, before transformation, and a P-value = 0.2371 and a P-value = 0.01461 for UK and KED, respectively, after transformation. This provided strong evidence (p-value < 0.01) that the residuals of both UK and KED before transformation were not normally distributed. It also showed that there was little or no real evidence (P-value > 0.1) that the UK residuals after transformation were not normally distributed and, finally, that there was moderate evidence ($0.01 <$ P-value < 0.05) that the KED residuals after transformation were not normally distributed. Note that the above test may be questionable since the observations can have strong spatial correlation.

The methods that we described in the previous section produce spatial predictions on a transformed scale. One cannot apply the inverse Box–Cox function directly to the expectations and variances because it would introduce biased predictions. This can be avoided by back-transforming the conditional simulations of the predictive distribution $p(\mathbf{y}_0|\mathbf{y})$. The most common predictor is the expectation of the back-transformed conditional simulations, but for many back-transformations of a Gaussian random field, the mean may not exist. Therefore, just as De Oliveira et al. (1997), we used the median as predictor and a fourth of the 95% credible interval as standard deviation:

$$\begin{aligned} \text{Med}[\mathbf{y}_{0,org}|\mathbf{y}_{org}] &= q_{0.50}[g^{-1}(\mathbf{y}_0|\mathbf{y})], \\ \text{sd}[\mathbf{y}_{0,org}|\mathbf{y}_{org}] &= \frac{q_{0.975}[g^{-1}(\mathbf{y}_0|\mathbf{y})] - q_{0.025}[g^{-1}(\mathbf{y}_0|\mathbf{y})]}{4} \end{aligned}$$

where $q(\cdot)$ produces sample quantiles corresponding to the given probabilities. A probability of exceeding a threshold level can be obtained analogously (Ribeiro and Diggle 2001). An exponential correlation function suited both UK and KED for the spatial covariance structure. The effective range or correlation distance equals 3ϕ.

The Bayesian inference requires specification of prior parameter distributions. These should be independent of the observations. For UK, prior information was deduced from the OPS model output. First, the OPS model output was transformed using the Box-Cox parameter we found earlier for UK. We took flat priors for the trend parameters $\boldsymbol{\beta}$, the inverse of the partial sill $1/\sigma^2$, the inverse of the range $1/\phi$ and the relative nugget τ^2_{rel} and then estimated the posterior distributions from the

OPS model output. For computational stability we limited the number of transformed OPS values to 200, at randomly chosen from the whole OPS grid. The posterior distributions of the trend and covariance parameters thus obtained were used as priors for the UK analysis. We fitted a multivariate Normal distribution to $(\boldsymbol{\beta}|OPS)$ and a scaled inverse-χ^2 distribution to $(\sigma^2|OPS)$. No standard statistical distributions were fitted to $(\phi|OPS)$ and $(\tau^2_{rel}|OPS)$, because these samples were used directly in the UK analysis. The priors are

$$
\begin{aligned}
p(\boldsymbol{\beta}) &= N\left[\begin{pmatrix} 1.53\text{E}+0 \\ 1.80\text{E}-4 \\ -5.18\text{E}-4 \end{pmatrix}, \sigma^2 \begin{pmatrix} 4.48\text{E}+0 & -2.01\text{E}-3 & -7.72\text{E}-3 \\ -2.01\text{E}-3 & 2.21\text{E}-5 & -3.43\text{E}-6 \\ -7.72\text{E}-3 & -3.43\text{E}-6 & 1.79\text{E}-5 \end{pmatrix}\right], \\
p(\sigma^2) &= \text{Inv}-\chi^2(11, 2.27\text{E}-3), \\
p(\phi) &= p(\phi|OPS), \\
p(\tau^2_{rel}) &= p(\tau^2_{rel}|OPS).
\end{aligned}
$$

The small (co)variances for $\boldsymbol{\beta}$ and σ^2 are due to the Box-Cox transformation. The values of ϕ and τ^2_{rel} were evaluated over a grid defined by (0, 300] km and [0, 0.3] respectively.

Since no prior information was available for the KED parameters, we started with non-informative flat priors, similar to those we used for analyzing the OPS output. The priors for the KED parameters are

$$
\begin{aligned}
p(\boldsymbol{\beta}) &\propto 1 \\
p(\sigma^2) &\propto 1/\sigma^2 \\
p(\phi) &\propto 1/\phi \\
p(\tau^2_{rel}) &\propto 1.
\end{aligned}
$$

Here ϕ and τ^2_{rel} were evaluated over a grid defined by (0, 20,000] km and [0, 0.5], respectively. The range parameter should be large, since the range for KED was expected to grow considerably (see Sect. 3.1). The bounds of ϕ and τ^2_{rel} were also applied to the RML parameter estimation.

Summarizing, we apply RML estimation to UK and KED. Besides RML estimation, we also apply Bayesian inference to UK using prior parameter information deduced from OPS output, and to KED using non-informative priors. KED uses OPS only as explanatory variable. The eventual parameter estimates are discussed in Sect. 3.1.

2.6 Validation procedure

A cross-validation by 'leaving one out' was carried out to study the performance of the models, given a number of stations n_s and network configurations. As in Cui et al. (1995) we used three error measures for validation: mean error (ME or bias), unbiased

root mean squared error (URMSE) and mean squared standardized error (MSSE):

$$ME = \frac{1}{n_s}\sum_{i=1}^{n_s}(y_{0,org,i} - y_{org,i}),$$

$$URMSE = \sqrt{\frac{1}{n_s}\sum_{i=1}^{n_s}(y_{0,org,i} - y_{org,i})^2 - ME^2},$$

$$MSSE = \frac{1}{n_s}\sum_{i=1}^{n_s}\frac{(y_{0,org,i} - y_{org,i})^2}{\mathrm{sd}_i^2}.$$

The ME indicates the bias of the predictions $\mathbf{y}_{0,org}$ to the original observations $\mathbf{y}_{org}$ and should be close to zero. The URMSE indicates the bias corrected standard deviation of the model and should be close to zero. We used the URMSE because it can be close to zero even in presence of a bias. The MSSE compares the squared differences with the kriging model variance, sd^2, and yields a value that should be close to one.

Unlike other studies, for example, in Cui et al. (1995) and Bourennane et al. (2000), where only one random set of different sizes is selected from the full set, we studied the effect of different configurations by selecting several random configurations of n_s stations ($n_s = 10, 20, \ldots, 70$) from the complete set of 70. Cross-validation was performed for each selection. The rationale for selecting several configurations is that every configuration leads to a different value for the ME, URMSE and MSSE. Since many combinations were possible, new configurations of n_s stations were selected until the 0.025, 0.50 and 0.975-quantiles of the outcomes of ME, URMS and MSSE reached stability. These values indicate the sensitivity of the models to the network configurations considered. If we had just selected one random configuration for every station sample density, we could have ended with accidental low or high values for some station sample density. Note that different configurations cannot be independent of each other if they share one or more stations within the correlation distance.

3 Results

3.1 Parameters before and after network reduction

Figure 3 shows the Bayesian posterior parameter densities (curves) and the RML parameter estimates (arrows). The 1983 posterior trend parameters $\boldsymbol{\beta}$ are more precise for both UK and KED; i.e. they show a density curve that is steeper and higher than curves for 1987 because of the larger number of stations. The trend parameters appear to be very precise, but this is due to the Box–Cox transformation. The RML estimates correspond well with the posterior modes.

The covariance parameters show different results. The posterior for the partial sill σ^2 for UK shows larger values than the prior based on OPS output. For KED, the posterior σ^2 is totally determined by the observations because of the non-informative prior. Posterior σ^2 obtained from the 1983 or the 1987 data are largely similar.

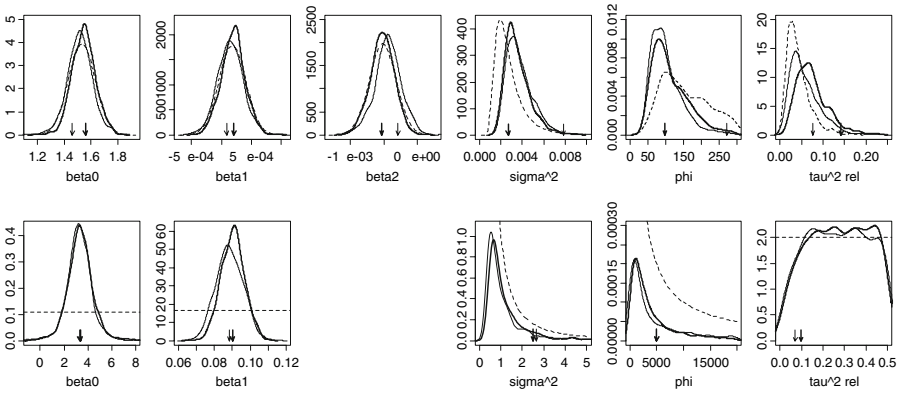

Fig. 3 Bayesian posterior parameter densities (curves) and the RML parameter estimates (arrows) for both UK (upper 6 panels) and KED (lower 5 panels). The bold solid lines represent 1983 posteriors and the fine solid lines 1987 posteriors. The dashed lines represent the prior distributions. The priors for UK are the posteriors based on OPS output, while priors for KED are non-informative

For UK, the posterior for the range parameter ϕ also differs from the prior. Based on OPS output, the mode is approximately 100 km, whereas the posterior mode for both 1983 and 1987 is approximately 80 km. The RML estimates differ considerably. For KED, both 1983 and 1987 posterior ϕ look similar, but have high values. The RML estimates for ϕ are the same. For UK, the posterior relative nugget, τ^2_{rel}, is higher than its prior, which is more pronounced due to more available stations. For KED, the posterior τ^2_{rel} is very flat, just like its prior. The RML estimates are approximately equal to 0.1 for both years.

The ϕ and τ^2_{rel} parameters are related (Fig. 4). For UK, a maximum occurs at approximately $\phi = 100$ km and $\tau^2_{rel} = 0.15$. For KED, this maximum is very flat, stretching out over a whole range of possible ϕ and τ^2_{rel} values. For KED, it is therefore difficult to estimate these parameters using RML, of which the maximum is found at approximately $\phi = 5000$ km and $\tau^2_{rel} = 0.1$. Notice that this value is very large as compared to the size of the country and the extent of the network. For the 1987 data, both profile likelihood surfaces were flatter due to fewer observations (figure not shown).

Figure 4 explains the shape of the posterior density curves of ϕ and τ^2_{rel} in Fig. 3. High values for both UK and KED are shown in Fig. 4 in the banana-shaped area between low values for ϕ and corresponding high values for τ^2_{rel}, and high values for ϕ and corresponding low values for τ^2_{rel}. If ϕ is low, the corresponding values for τ^2_{rel} will be high. For UK this results in a density curve similar to the posterior for τ^2_{rel} in Fig. 3. For KED this results in a flat posterior for τ^2_{rel}. The banana-shaped area also explains the differences between the RML estimates and the posterior modes. Note prior information is included for UK in Fig. 3.

For KED, ϕ extends to values far beyond the largest area of the Netherlands (approximately 400 km). Furthermore, ϕ has to be multiplied by three to compute the effective range of the exponential model. For such large values, the exponential model for the variogram approaches a linear model for distances inside the Netherlands. The

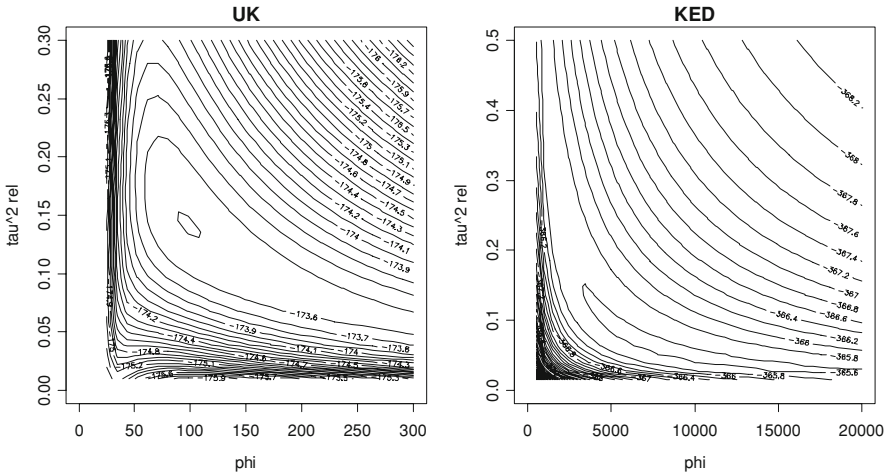

Fig. 4 Profile likelihood as a function of ϕ and τ^2_{rel} for UK and KED for the 1983 data

correlation between stations is no longer a function of their separation distance. Apparently, the information from the OPS model is so good, that it takes most of the spatial correlation in the residuals away. In this case, KED almost becomes similar to ordinary linear regression.

3.2 Spatial predictions and standard deviations before and after network reduction

We made predictions for (1) UK and KED, (2) based on the 1983 and 1987 data, (3) parameter estimates by RML and Bayesian inference, and (4), for the 1987 data, keeping the 1983 parameters fixed and re-estimating the parameters for the 1987 data. We call the respective predictions: UK 83 RML, UK 83 Bayes, UK 87 RML fix, UK 87 Bayes fix, UK 87 RML re, UK 87 Bayes re, KED 83 RML, KED 83 Bayes, KED 87 RML fix, KED 87 Bayes fix, KED 87 RML re and KED 87 Bayes re.

Figure 5 illustrates the spatial predictions and corresponding standard deviations of NO_x concentrations with UK and KED on the basis of 1983 data and parameter estimation with RML. For UK the predicted concentrations vary gradually in space, whereas for KED the predicted NO_x concentrations show more spatial variation in terms of alternating higher and lower values. Individual cities and highways are clearly visible as a result of using the OPS output as external trend. For UK, the standard deviations look very similar to the predicted expectation and show more variation in space than the KED standard deviations. Because of the back-transformation, the standard deviations are correlated with the predicted expectations. In fact, the predictive distribution at each location is skewed.

The spatial patterns of the other predictions look similar and the corresponding figures are therefore not shown. Instead, Fig. 6 shows box plots based on the 1,405 individual locations with the minimum, 0.25-quantile, median, 0.75-quantile and maximum values for the twelve predictions and standard deviations.

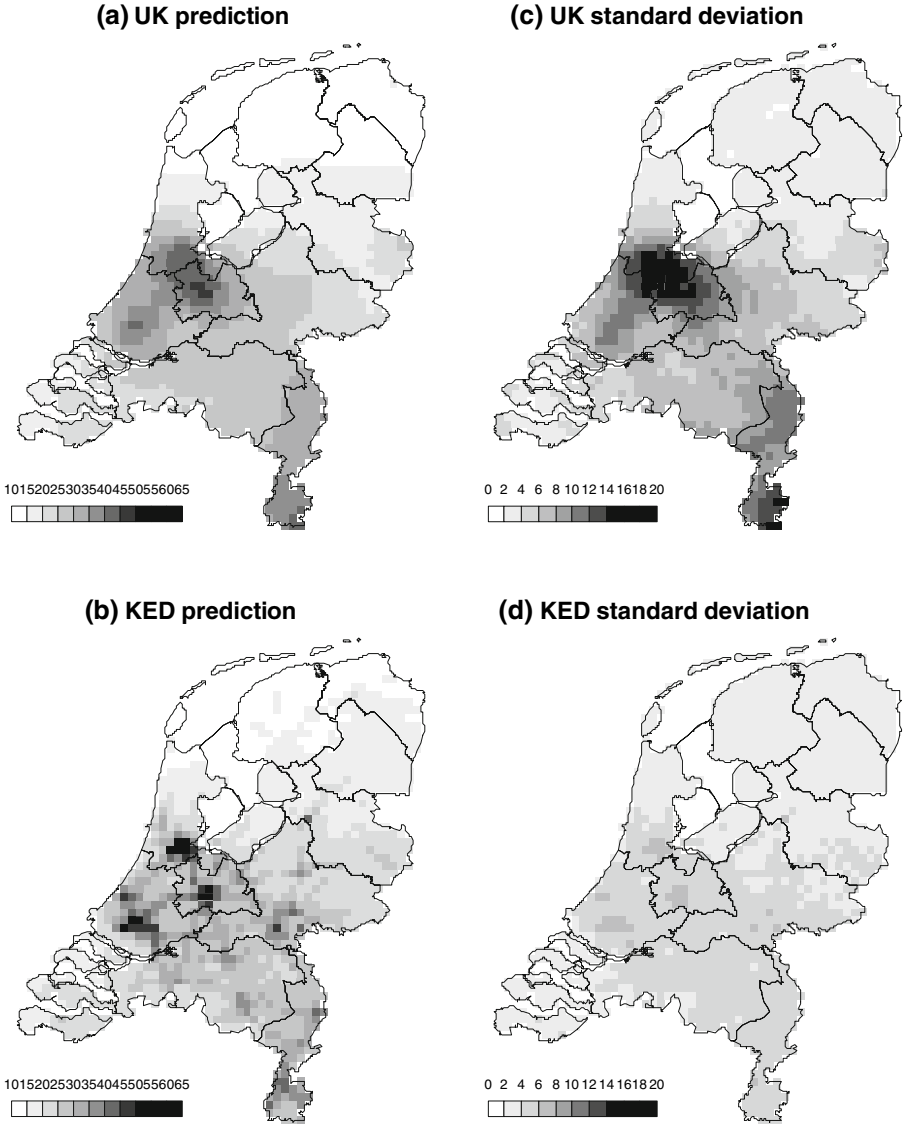

Fig. 5 Spatial predictions (**a, c**) and their standard deviations (**b, d**) for NO_X concentrations with UK (**a, b**) and KED (**c, d**), based on 1983 data and parameter estimation by RML

Based on Figs. 5 and 6, UK and KED can be said to differ considerably. We will therefore discuss UK and KED separately. Of interest are the differences before and after network reduction (1983 and 1987), the method of parameter estimation and, for 1987, keeping parameters fixed or not.

The median for UK 87 is systematically 9% lower than for UK 83. Re-estimating the parameters gives the same result as keeping them fixed. Bayesian inference shows

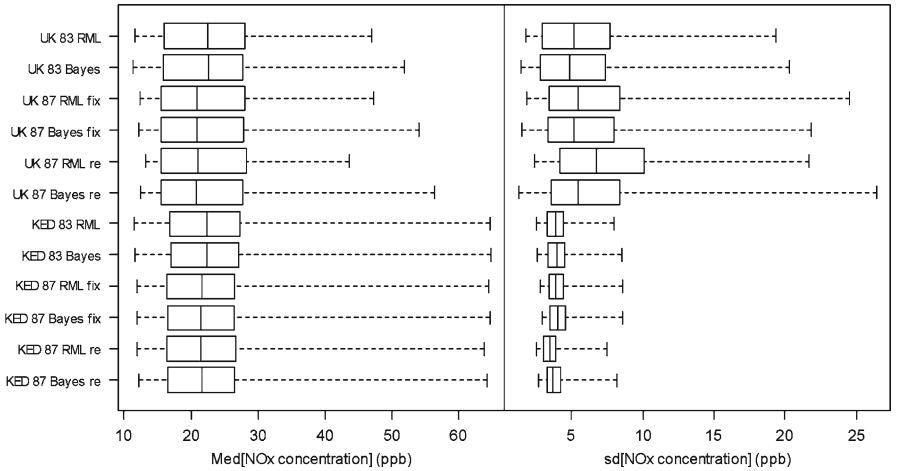

Fig. 6 Box plots based on the 1405 individual locations with the minimum, 0.25-quantile, median, 0.75-quantile and maximum values for the 12 predictions and standard deviations

higher extremes than RML. The standard deviations for UK 87 are higher than for UK 83 due to fewer observations. Re-estimating the parameters yields higher standard deviations, since parameter uncertainty has emerged.

The median for KED 87 is systematically 4% lower than the median for KED 83. Extremes for KED predictions are similar for all approaches. Apparently, the external trend determines the predicted values. The standard deviations look very similar, except for KED 87 re. The range of the KED standard deviations is almost equal to the 50% box of the UK standard deviations. The OPS output explains the spatial variation in the observations.

3.3 Validation

The core study results are reflected in Fig. 7. Validations were made for: (1) UK and KED and OLS, (2) parameter estimates by RML and Bayesian inference, and (3) keeping the 1983 parameters fixed and re-estimating the parameters for each station configuration. We called the respective validations: UK RML fix, UK Bayes fix, UK RML re, UK Bayes re, KED RML fix, KED Bayes fix, KED RML re, KED Bayes re, OLS fix and OLS re. The last step was a validation of the OPS model and we present, in succession, the validation results of OPS, UK, KED and OLS. The values are back-transformed to their original scale.

OPS is positively biased, with an average of 2 ppb (Fig. 7a). The median is independent of the number of observations, because it represents a model validation without using observations. The ME range increases with fewer observations. The URMSE (Fig. 7b) of OPS is on average 5 ppb. The range, as seen in Fig. 7a, is almost symmetric around the median and increases as well for fewer observations. The MSSE of OPS is not available since the OPS model does not provide a standard deviation.

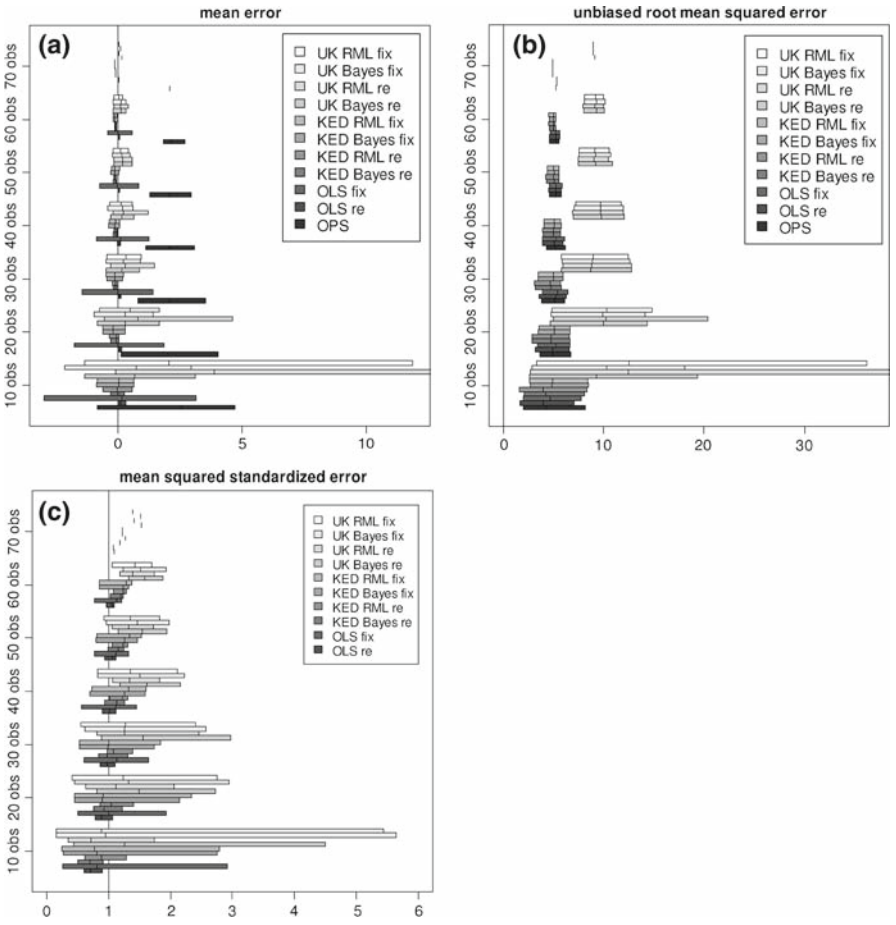

Fig. 7 The 95% intervals of outcomes of (**a**) mean errors (ppb), (**b**) unbiased root mean-squared errors (ppb), and (**c**) mean-squared standardized errors (−) for the eight spatial prediction methods, OLS and the OPS model output (ME and URMSE only), taken over random subsets of 10, 20, . . ., 60 points from the total test set of 70 stations

For UK, the median ME for 70–40 observations shows almost no bias as we were working here with observations only. Note the advantage of keeping parameters fixed (for RML) for sparse observations (20–10 observations), or the advantage of prior information over RML. This can also be observed from Fig. 7b. For UK RML and UK Bayes similar MSSE ranges for fixed parameters occur (Fig. 7c). If parameters are re-estimated, the values are closer to one. This indicates that the absolute prediction error is almost equal to the predicted kriging standard deviation. The 0.975-quantiles that lie outside Figs. 7a and 7b are 22.9 ppb and 64.3 ppb, respectively.

KED outperforms OPS and UK since observations are combined with OPS output. The ME deviance (Fig. 7a) is smaller and closer to zero for KED, even for sparse observations. KED RML and KED Bayes are very similar because of the non-informative

priors. Re-estimating the parameters causes a smaller deviance in the ME. For 40 or more observations, the URMSE values for KED are systematically lower than those for UK (Fig. 7b). Somewhat lower URMSE values are observed by re-estimating the parameters. Compared to UK, the MSSE values (Fig. 7c) remain closer to one. Re-estimating the parameters causes smaller MSSE values than keeping them fixed.

Finally, we applied OLS as a comparison. We found the relations observed $NO_x = 3.28 + 0.0928 \cdot$ OPS for 1983 (MSE = 0.300) and observed $NO_x = 3.25 + 0.0915 \cdot$ OPS for 1987 (MSE = 0.228). Note that we applied the Box–Cox transformation using the parameter we found earlier for KED. We notice that OLS fix and OPS show the same wide range in outcomes, but that OLS fix values are centered around 0 (Fig. 7a). Clearly, OLS fix yields unbiased predictions, similar to UK and KED, but different from the OPS calculations. It has a wider range in deviations than UK and KED, though. The range of outcomes for OLS re is similar to that of KED re. Fig. 7b shows similar root mean squared errors for OLS as compared to KED and OPS. Hence, in terms of squared deviations, these methods have the same outcomes. Finally, Fig. 7c shows that OLS fix has a similar range in mean squared standard errors as KED fix. OLS re has similar outcomes as KED re, but the range of outcomes is smaller than that of OLS fix. We notice, that OLS in this study yields predictions of a similar precision as those from advanced geostatistical methods. An explanation would be that the OPS model is of such a quality that, although biased, it removes the spatial dependencies. Similar situations may occur in universal kriging, where inclusion of a trend results in reduction of spatial dependence (see, e.g., Stein et al. 1991).

4 Discussion

In our comparison of several estimation and interpolation methods, we observed KED to be superior to both UK and OPS. With reference to UK, KED emphasizes the details of the OPS model, as shown in Fig. 5. It is less sensitive to the network configuration, resulting in smaller intervals for ME values, URMSE values and MSSE values (Fig. 7). In contrast to the OPS model, KED is almost unbiased due to the use of observations. Although the OPS model gives biased predictions, a combination of observations with OPS model outputs is superior to UK. One may note, however, that the KED results are somewhat optimistic since uncertainty in OPS calculations are only treated as part of the nugget effect. For UK we used a linear trend. We could have chosen a more complex model, like a second order trend. This could have been more realistically reflecting the data field, and may hence have resulted in lower prediction error variances. As our goal was to focus on parsimonious models, a linear trend was of a sufficient quality initially to approximate the observations.

The OPS model output explains much of the variation in the observations, resulting in a flat covariance structure of the KED model. In most cases, the effective range becomes large and in combination with a large relative nugget, the KED covariance structure is nearly constant within the horizontal scales of the Netherlands. Nevertheless, spatial prediction using the OPS model is still beneficial, as it reduces the uncertainties commonly occurring in UK models when predicting data beyond the correlation distance.

A next issue concerns estimation of the parameters with either RML or Bayesian inference. In the case of many observations (> 20), a choice for any is irrelevant. If prior information is available, however, Bayesian inference leads to lower prediction error standard deviations. In case of sparse observations (10–20 observations), Bayesian inference with prior information is more robust than RML estimation. Similarly, we may either keep parameters fixed or we could re-estimate. A low number of observations corresponds to a large parameter uncertainty. In that case, re-estimated parameters are to be preferred, allowing incorporation of specific uncertainties. Re-estimation of parameters brings MSSE values close to one. On the other hand, when parameters are kept fixed, no estimation problem occurs with sparse observations. The effect of parameter uncertainty on predictions, however, appears to be small relative to the effect of network configuration, i.e. the number and location of stations. This can have a substantial effect on interpolation accuracy and requires some further optimization.

KED allows spatial prediction of a primary variable, accounting for the dense secondary variable. Collocated co-kriging could have been used as an alternative. We preferred the use of KED, requiring a less demanding variogram analysis. Furthermore, comparison studies (Pardo-Igúzquiza 1998; Goovaerts 2000) show KED interpolation to perform better than collocated co-kriging. KED is a form of data assimilation, but differs from Kalman filtering, for example, which results in an optimal estimation by combining observations and model output weighted by their inverse variances (Jazwinski 1970; Meinhold and Singpurwalla 1983). KED is a regression-based interpolation method, where the model output is regarded as a covariate and, as such, partly explains the variation in the observations. This allows the model to have a systematic error (bias) whereas other external information may be included as well. It is limited in the sense that covariates are considered to be deterministic.

In this study we did not address the change-of-support problem (Lajaunie and Wackernagel 2000; Bertino and Wackernagel 2002). We applied point kriging. In a separate study (results not shown) we applied the OPS model on a 1×1 km resolution and aggregated this to 5×5 km resolution. Small differences occurred in the rural areas (-1% to $+1\%$), whereas larger differences occurred in and near urban areas ($+2\%$ to $+5\%$) because of the large number of point emissions.

The OPS model output contains uncertainties as well. Dispersion model uncertainties, however, can be large, since the model input is uncertain, along with uncertain model parameters and uncertainty introduced by model simplification. These were not considered further in this study. Neither were uncertainties in the observations and station coordinates. These issues have been addressed elsewhere (Van de Kassteele and Stein 2006).

5 Conclusions

We conclude from this study that a combination of observations and a deterministic dispersion model provides a successful, model-based geostatistical interpolation procedure. The combination leads to more precise spatial interpolation results than ignoring either the observations or the model output. This applies in particular to those

parts of the study area that are outside the sampling area. Compared to universal kriging, the dispersion model output applied as an external drift provides more detail in spatial maps and results in smaller standard deviations. Its inclusion leads to removal of spatial dependence, and to interpolation that is similar to ordinary least squares at the scale of study.

Cross-validation was done by repeatedly selecting different subsets from a set of test data, allowing a comparison of reliability intervals of the interpolation results. It showed that inclusion of the deterministic OPS model leads to a substantial improvement in the predictions. In particular we notice that external drift kriging leads to a much lower spread in mean error values than UK.

Bayesian estimation methods have advantages over RML methods, as its use in interpolation improves the predictions, and leads to more robust predictions in the case of sparse observations.

Data reduction, as occurred for the air quality monitoring network in the Netherlands in the nineteen eighties, can be overcome by skillful modeling with a deterministic model using model-based geostatistical interpolation methods. Further use of improved prior distributions chosen and more detailed modeling needs to be further investigated in the future. A further optimization of the location of network stations is as yet to be carried out and may lead to a further reduction of spatial uncertainties.

Acknowledgements The authors would like to thank Jan de Ruiter of the Netherlands Environmental Assessment Agency—RIVM for providing the OPS model output data, and the referees for their comments and suggestions, which were all very helpful to improve the quality of this article.

References

Berger JO, De Oliveira V, Sanso B (2001) Objective bayesian analysis of spatially correlated data. J Am Stat Assoc 96:361–1374

Bertino L, Wackernagel H (2002) Case studies of change-of-support problems. technical report N–21/02/G, ENSMP—ARMINES, Centre de Géostatistique, Fontainebleau, France

Bourennane H, King D, Couturier A (2000) Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. Geoderma 97(3–4):255–271

Bourennanne H, King D, Chery P, Bruand A (1996) Improving the kriging of a soil variable using slope gradient as external drift. Eur J Soil Sci 47(4):473–483

Box GEP, Cox DR (1964) An analysis of transformations. J Royal Stat Soc B 26:211–246

Cassiraga EF, Gomez-Hernandez JJ, Soares A, Froidevaux R (1997). Improved rainfall estimation by integration of radar data: a geostatistical approach. Proceedings GeoENV I—Geostatistics for environmental applications. Lisbon, Portugal, 18–19 November 1996. Kluwer Academic Publishers, Dordrecht, The Netherlands. pp 363–374

Chilès JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Cui H, Stein A, Meyers DE (1995) Extension of spatial information, Bayesian kriging and updating of prior variogram parameters. Environmetrics 4:373–384

Derwent RG, Hov O, Asman WAH, Van Jaarsveld JA, De Leeuw FAAM (1989) An intercomparison of long-term atmospheric transport models; the budgets of acidifying species for the Netherlands. Atm Environ 23:1893–1909

Desbarats AJ, Logan CE, Hinton MJ, Sharpe DR (2002) On the kriging of water table elevations using collateral information from a digital elevation model. J Hydrol 255(1–4):25–38

De Oliveira V, Kedem B, Short DA (1997) Bayesian prediction of transformed gaussian random fields. J Am Stat Assoc 92:1422–1433

Diggle PJ, Ribeiro Jr.PJ (2002) Bayesian inference in gaussian model-based geostatistics. Geograph Environ Model 6(2):129–146
Diggle PJ, Tawn JA, Moyeed RA (1998) Model-based geostatistics (with discussion). Appl Stat 47:299–350
EPA (1998). National air quality and emissions trends. Report 1997, U.S. environmental protection agency, Triangle Park N.C., America
Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Bayesian data analysis, 2nd edn. Chapman & Hall, New York
Genikhovich E, Filatova E, Ziv A (2002) A method for mapping the air pollution in cities with the combined use of measured and calculated concentrations. Int J Environ Poll 18(1):56–63
Goovaerts P (1999) Using elevation to aid the geostatistical mapping of rainfall erosivity. Catena 34(3–4): 227–242
Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J Hydrol 228(1–2):113–129
Grimes DIF, Pardo-Igúzquiza E, Bonifacio R (1999) Optimal areal rainfall estimation using raingauges and satellite data. J Hydrol 222(1–4):93–108
Handcock MS, Stein ML (1993) A bayesian analysis of kriging. Technometrics 35(4):403–410
Hengl T, Heuvelink GBM, Stein A (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120:75–93
Hudson G, Wackernagel H (1994) Mapping temperature using kriging with external drift: theory and an example from Scotland. Int J Climatol 14(1):77–91
Jazwinski AH (1970) Stochastic processes and filtering theory. Mathematics in Science and Engineering, Vol 64. Academic Press, New York
Kitanidis PK, Shen KF (1996) Geostatistical estimation of chemical concentration. Advances in Water Resources 19(6):369–378
Lajaunie C, Wackernagel H (2000) Geostatistical approaches to change of support problems, theoretical framework. technical report N–30/01/G, ENSMP—ARMINES, Centre de Géostatistique, Fontainebleau, France
Lehmann EL (1975) Nonparametrics: statistical methods based on ranks. Holden-Day, San Francisco
Meinhold RJ, Singpurwalla ND (1983) Understanding the kalman filter. The Am Stat 37(2):123–127
Monestiez P, Courault D, Allard D, Ruget F (2001) Spatial interpolation of air temperature using environmental context: Application to a crop model. Environ Ecolo Stat 8(4):297–309
Pardo-Igúzquiza E (1998) Comparison of geostatistical methods for estimating the areal average climatological rainfall mean using data on precipitation and topography. Int J Climatol 18(9):1031–1047
Pauly M, Drueke M (1996) Mesoscale spatial modelling of ozone immissions. An application of geostatistical methods using a digital elevation model. Gefahrstoffe reinhaltung der luft 56(6):225–230
Ribeiro Jr.PJ, Diggle PJ (1999) Bayesian inference in Gaussian model-based geostatistics. Technical report ST-99-08, Department of Mathematics and Statistics, Lancaster University, Lancaster, United Kingdom
Stein A, Van Eijnsbergen AC, Barendregt LG (1991) Cokriging nonstationary data. Mathe Geol 23(5): 703–719
Van de Kassteele J, Stein A (2006) A model for external drift kriging with uncertain covariates applied to air quality measurements and dispersion model output. Environmetics 17(4):309–322
Van Elzakker BG (2001) Monitoring activities in the Dutch national air quality monitoring network in 2000 and 2001. RIVM report 723101055, National Institute of Public Health and the Environment, Bilthoven, The Netherlands
Van Jaarsveld JA, De Leeuw FAAM (1993) OPS: An operational atmospheric transport model for priority substances. Enviro Software 8:93–100
Van Jaarsveld JA (1991) An Operational atmospheric transport model for Priority Substances; specification and instructions for use. RIVM report 222501002, National Institute of Public Health and the Environment, Bilthoven, Netherlands
Van Jaarsveld JA (1995) Modelling the long-term atmospheric behaviour of pollutants on various spatial scales. Ph.D. thesis, Utrecht University, the Netherlands
Wackernagel H (1995) Multivariate geostatistics. Springer, New York
WHO (2003) Health aspects of air pollution with particulate matter, Ozone and Nitrogen Dioxide. Report on a WHO Working Group EUR/03/5042688, Bonn, Germany

Author Biographies

Jan van de Kassteele is a statistician at the National Institute for Public Health and the Environment (RIVM) in Bilthoven (the Netherlands). He received an MSc in meteorology in 2001 and a PhD in statistical air quality mapping in 2006, both at Wageningen University. He did his PhD work at both the Mathematical and Statistical Methods Group, Biometris, and the Netherlands Environmental Assessement Agency in Bilthoven. He worked on optimization of uncertain air quality information by combining observations and dispersion model output. His research interests are (Bayesian) spatial statistics, boundary layer meteorology and air-quality modeling.

Alfred Stein is professor of mathematical and statistical methods for geodata at the International Institute for Geo-Information Science and Earth Observation (ITC) and chair of the Mathematical and Statistical Methods Group, Biometris, at Wageningen University. Professor Stein received his MSc in mathematics and information science, with a specialization in applied statistics from Eindhoven University of Technology (The Netherlands). In 1991 he was awarded a PhD in spatial statistics at Wageningen University. His research interests focus on statistical aspects of spatial data, including monitoring data in the broadest sense. Subjects of special interest are optimal sampling, image analysis, spatial statistics, Bayesian statistics, data quality, and fuzzy methods.

Arnold Dekkers is senior statistician at the National Institute for Public Health and the Environment (RIVM). He received his MSc in mathematical statistics from the University of Amsterdam. In 1991 he received his PhD in extreme value statistics from the Erasmus University Rotterdam. His research interests are time series analysis, (Bayesian) spatial statistics, environmental and disease modeling.

Guus Velders is senior scientist at the Netherlands Environmental Assessment Agency in Bilthoven. He received an MSc in physics in 1987 and a PhD in quantum chemistry in 1992, both from the University of Twente (The Netherlands). He worked on ozone layer depletion and climate change for ten years and is involved in EU, UNEP and IPCC assessments on these issues. His current research focuses on data assimilation and Kalman filtering for air quality applications.