WATER RESOURCES RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

Bayesian data fusion applied to water table spatial mapping

D. Fasbender, L. Peeters, P. Bogaert, and A. Dassargues

D. Fasbender, Dept. of Environmental Sciences and Land Use Planning, University catholic of Louvain, Croix du Sud 2/16, B-1348 Louvain-la-Neuve, Belgium. (Dominique.Fasbender@uclouvain.be)

L. Peeters, Dept. of Earth and Environmental Sciences, Katholieke Universiteit Leuven, Celestijnenlaan 200E, B-3001 Heverlee (Luk.Peeters@geo.kuleuven.be)

P. Bogaert, Department of Environmental Sciences and Land Use Planning, University catholic of Louvain, Croix du Sud 2/16, B-1348 Louvain-la-Neuve, Belgium. (Patrick.Bogaert@uclouvain.be)

A. Dassargues, Hydrogeology & Environmental Geology, ArGEnCo, University of Liege, Chemin des Chevreuils 1, B-4000 Liège 1, Belgium (Alain.Dassargues@ulg.ac.be)

also at: Dept. of Earth and Environmental Sciences, Katholieke Universiteit Leuven, Celestijnenlaan 200E, 3001 Heverlee

Abstract. Water table elevations are usually sampled in space using piezometric measurements that are unfortunately expensive to obtain and are thus scarce over space. Most of the time, piezometric data are sparsely distributed over large areas, thus providing limited direct information about the level of the corresponding water table. As a consequence, there is a real need for approaches that are able at the same time to (i) provide spatial predictions at unsampled locations and (ii) enable the user to account for all potentially available secondary information sources that are in some way related to water table elevations. In this paper, a recently developed Bayesian Data Fusion framework (BDF) is applied to the problem of water table spatial mapping. After a brief presentation of the underlying theory, specific assumptions are made and discussed in order to account for a digital elevation model as well as for the geometry of a corresponding river network. Based on a data set for the Dijle basin in the north part of Belgium, the suggested model is then implemented and results are compared to those of standard techniques like ordinary kriging and cokriging. Respective accuracies and precisions of these estimators are finally evaluated using a "leave-one-out" cross-validation procedure. Though the BDF methodology was illustrated here for the integration of only two secondary information sources (namely a digital elevation model and the geometry of a river network), the method can be applied for incorporating an arbitrary number of secondary information sources, thus opening new avenues for the important topic of data integration in a spatial mapping context.

Keywords: kriging, cokriging, data merging, Digital Elevation Model, DEM, river network.

1. Introduction

Water table elevations can be directly obtained from piezometric heads measurements at wells and boreholes locations. Unfortunately, for most survey studies and due to the associated costs, the number of these locations are most of the time limited to already existing wells and boreholes, that are typically scarce and sparsely distributed over space. As a result, using cheaper and/or more abundant auxiliary information that are in some way related to piezometric heads is of great interest for the prediction of the water table elevations, especially for predicting at locations that are far away from the sampled ones. More generally, there is a real need for methods that enable the user to account for multiple auxiliary information sources in a spatial prediction context. Though such methods exist since the early 1990s (e.g. *Christakos* [1990]), this was recently called to mind in [*IAHS*, 2003] and, accordingly, new methods are currently undergoing investigations.

Focusing on the single context of water table spatial mapping, [*Hoeksema et al.*, 1989] already tried to use a cokriging (CoK) approach (see e.g. [*Chilès and Delfiner*, 1999]) for the spatial mapping of a water table in the area of Oak Ridge (Tennessee). The secondary variable used in that study was the ground surface elevation because the underlying water table was supposed to be a smoothed replica of it. Though results were found more accurate than ordinary kriging (OK), the cokriging approach remains limited to linear predictions and the corresponding multivariate model (like, e.g., the linear model of coregionalization that relies on linear combinations of basis covariance functions models ; see e.g. [*Chilès and Delfiner*, 1999]) is in general hard to infer when there are several (and possibly numerous) information sources at hand. Lately, [*Linde et al.*, 2007] proposed a

Bayesian approach of the problem that enables the user to account for piezometric and self-potential measurements. Thanks to its Bayesian formulation, non-linear relationships were permissible, leading thus to a more flexible set of models. As expected, the influence and advantage of auxiliary self-potential measurements were noticeable in locations far away from piezometric heads measurements.

Over the last twenty years, Bayesian approaches have gained more and more credit in spatial statistics. Initially proposed by Christakos [*Christakos*, 1990, 1991], the Bayesian Maximum Entropy (BME) paradigm for example has proven in many cases its ability to account for additional information sources and their associated uncertainties in various space-time prediction contexts. Though originaly proposed in the case of continuous data [see e.g. *Christakos*, 1992; *Christakos and Li*, 1998; *Christakos*, 2000; *Serre and Christakos*, 1999], other cases like categorical data (see e.g. [*Bogaert*, 2002; *Bogaert and D'Or*, 2002; *D'Or and Bogaert*, 2004]) and even mixed continuous and categorical data [*Wibrin et al.*, 2006] were rapidly tackled, making BME a complete and unified framework. Very recently, [*Bogaert and Fasbender*, 2007] proposed a complementary Bayesian Data Fusion (BDF) framework that permits to account at the same time for several auxiliary information sources, where each of them is potentially improving the knowledge about a variable of interest. In theory, the number of secondary information sources that can be incorporated is not restricted. As emphasized by the authors, one of the main advantages of this approach compared to traditional multivariate ones (e.g. cokriging methods) is that it relieves the need of relying on spatial multivariate linear models, so that a much richer category of non-linear models can be accessed.

In this paper, an implementation of the BDF approach is proposed in the context of a water table spatial prediction. A Digital Elevation Model (DEM) and the geometry of the corresponding river network are used as secondary information sources in order to improve knowledge about water table elevations at unsampled locations. The general formulation of the method is first briefly described and several specific assumptions are proposed for its practical implementation. This method is then applied to the case study of the Dijle basin in the north part of Belgium. Finally, a discussion and some conclusions about the method, its results compared to other ones and its perspectives are provided.

2. Bayesian data fusion

Combining multiple information sources into a single final prediction (i.e. data fusion) is not a new problem and is not restricted to environmental sciences, as it covers a wide variety of applications. Among them, Bayesian approaches have provided convenient solutions to various interesting problems such as image surveillance [*Jones et al.*, 2003], object recognition [*Chung and Shen*, 2000], object localization [*Pinheiro and Lima*, 2004], robotic [*Moshiri et al.*, 2002; *Pradalier et al.*, 2003], image processing [*Pieczynski et al.*, 1998; *Zhang and Blum*, 2001; *Rajan and Chaudhuri*, 2002], classification of remote sensing images [*Melgani and Serpico*, 2002; *Simone et al.*, 2002; *Bruzzone et al.*, 2002], enhancement of remote sensing images [*Fasbender et al.*, 2008, 2007] and environmental modeling [*Wikle et al.*, 2001; *Christakos*, 2002], just to quote a few of them. The two main advantages of Bayesian approaches are (i) to set the problem in a proper probabilistic framework and (ii) to provide straightforward means to update existing probability density functions with new relevant information. Lately, [*Bogaert and Fasbender*, 2007] proposed a general BDF formulation especially designed for spatial prediction problems.

These general results will be applied here for the spatial mapping of water table elevations. For the sake of brevity, it is not possible to present the whole underlying theory, so only theoretical results that are the most relevant ones for our application will be presented hereafter. The interested reader may refer to [*Bogaert and Fasbender*, 2007] for a detailed description of the theory.

2.1. General formulation

Let us define $\{\mathbf{x}_0, \ldots, \mathbf{x}_n\}$ as the set of locations where indirect observations $\mathbf{y}' = (y_0, \ldots, y_n)$ are available about a variable of interest Z. Based on the idea that the corresponding random vector of interest $\mathbf{Z}' = (Z_0, \ldots, Z_n)$ cannot be directly observed at these locations, BDF as presented in [*Bogaert and Fasbender*, 2007] aims at reconciling the auxiliary variables $\mathbf{Y}$ to the primary variables $\mathbf{Z}$ through an error-like model, with

$$\mathbf{Y} = g(\mathbf{Z}) + \mathbf{E} \tag{1}$$

where $\mathbf{g}(.)$ are functionals and where $\mathbf{E}' = (E_1, \ldots, E_n)$ is a vector of random errors that are stochastically independent from $\mathbf{Z}$. Using classical probability calculus, it is possible to formulate the conditional probability density function (pdf) of the vector of interest given the observed variables as

$$f(\mathbf{z}|\mathbf{y}) \propto f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}}(\mathbf{y} - \mathbf{g}(\mathbf{z})) \tag{2}$$

where $f_{\mathbf{Z}}(.)$ is the *a priori* pdf for $\mathbf{Z}$ and $f_{\mathbf{E}}(.)$ is the pdf of the errors $\mathbf{E}$. In the context of a water table mapping, one can write that $\mathbf{Z} = (Z_0, \mathbf{Z}_S, \mathbf{Z}_U)'$ where Z_0 refers to the water table elevations at prediction location $\mathbf{x}_0$, $\mathbf{Z}_S$ refers to locations $\mathbf{x}_S = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ where both Z_i's and Y_i's are jointly sampled, and $\mathbf{Z}_U$ refers to locations $\mathbf{x}_U = \{\mathbf{x}_{m+1}, \ldots, \mathbf{x}_n\}$ where only Y_i's are sampled. As the final goal is to obtain a conditional pdf of $Z_0|\mathbf{z}_S, \mathbf{y}$, elementary probability theory leads to the expression

$$f(z_0|\mathbf{z}_S, \mathbf{y}) \propto \int f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}}(\mathbf{y} - g(z))\mathrm{d}\mathbf{z}_U \tag{3}$$

Furthermore, if stochastic independence of errors $\mathbf{E}$ can be assumed as well as the fact that each Y_i depends only on a single corresponding Z_i through a functional $g_i(.)$ (stated in other words, $Y_i = g_i(Z_i) + E_i$), then one can show that the final expression of the conditional pdf is given by

$$f(z_0|\mathbf{z}_S, \mathbf{y}) \propto \prod_{i=0}^{m} f_{E_i}(y_i - g_i(z_i)) \int f_{\mathbf{Z}}(\mathbf{z}) \prod_{j=m+1}^{m+n} f_{E_j}(y_j - g_j(z_j))\mathrm{d}\mathbf{z}_U \tag{4}$$

$$\propto \prod_{i=0}^{m} \frac{f(z_i|y_i)}{f(z_i)} \int f_{\mathbf{Z}}(\mathbf{z}) \prod_{j=m+1}^{m+n} \frac{f(z_j|y_j)}{f(z_j)} \mathrm{d}\mathbf{z}_U \tag{5}$$

where Eqs. 4 and 5 are completely equivalent expressions as they are linked to each other using Bayes theorem, with

$$f_{E_j}(y_j - g_j(z_j)) = f(y_i|z_i) \propto \frac{f(z_i|y_i)}{f(z_i)}$$

so that using either distributions of errors $f_{E_i}(.)$ or conditional distributions $f(z_i|y_i)$ provides two possible way of incorporating different information sources.

As an interpretation and as a consequence of the independence hypothesis for the errors, this Bayesian approach separates the problem into two parts. The first one is making use of the spatial dependence of the primary variable through the multivariate distribution $f_{\mathbf{Z}}(\mathbf{z})$, whereas the second one integrates the various auxiliary information sources through the univariate conditional distributions $f(z_i|y_i)$ on a per-location basis. As a consequence, a multivariate formulation is no longer needed and corresponding multivariate models do not need to be inferred, thus avoiding the restrictions imposed by multivariate models as used in cokriging. One may argue about the practical pertinency of these equations as they rely on this independence hypothesis, but one can show that, from an entropic viewpoint, it corresponds to the hypothesis that leads to the minimum loss of information (again, see [*Bogaert and Fasbender*, 2007] for details about this topic).

It is also worth noting that Eq. 4 is closely related to those obtained with the Bayesian Maximum Entropy (BME) method, although BME proposes a Maximum Entropy step for the choice of the *prior* distribution $f(\mathbf{Z})$ whereas BDF leaves this choice open to the user. BME and BDF can thus be viewed as complementary formulations of a same general Bayesian approach. The main difference between both approaches is that the distributions from the secondary information are either considered as likelihood functions $f(y_j|z_j)$ or as a conditional distribution $f(z_j|y_j)$, which leads thus to different results.

Finally, it is worth noting that for applications where secondary information are not exhaustively known, one could of course not use $f(z_j|y_j)$ (resp. $f(y_j|z_j)$) at these locations. Fortunately, the BDF framework is still sound theoretical in that case as it is sufficient to remove the corresponding conditional pdf in Eq. 4 (resp. in Eq. 5). Particularly, if all conditional distributions at unsampled locations are not taken into account, the integrals in Eq. 4 and 5 come down both to $f(z_0|\mathbf{z}_S)$ which simplifies substantially the final expression (e.g. $f(z_0|\mathbf{z}_S)$ could be computed separately from classical space-time prediction methods such as kriging ones).

2.2. Specific assumptions

Though Eqs. 4 and 5 are intended to be as general as possible, specific assumptions need to be made here in order to fit the specific problem of water table prediction, of course. For this purpose, it is also important to identify which secondary information sources may be potentially useful.

Using merely the sampled water table elevations, it is already possible to estimate the spatial dependence of this variable and to use it afterwards for kriging prediction (e.g. [*Chilès and Delfiner*, 1999; *Cressie*, 1990, 1991]). Kriging is known to provide a linear predictor that corresponds to the Best Linear Unbiased Predictor (BLUP) in the least-squares sense. Additionally, it is the best possible predictor when the random vector $\mathbf{Z}$ is assumed to be multivariate Gaussian. It is also well-known that, under constraints for the first two moments (i.e., the vector of the means and the covariance matrix), the joint distribution $f_{\mathbf{Z}}(\mathbf{z})$ that maximizes Shannon entropy is precisely the multivariate Gaussian one (see e.g. [*Papoulis*, 1991] or [*Christakos*, 1990]). For these reasons, an *a priori* multivariate Gaussian distribution $f_{\mathbf{Z}}(\mathbf{z})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as inferred from the data is relevant according to information at hand. One is not restricted to this choice, as it would be possible to include other information (i.e., skewness) that would lead to another maximum entropy distribution (see [*Papoulis*, 1991] or [*Christakos*, 1990]). However, we will show hereafter that assuming a multivariate Gaussian hypothesis is also a convenient choice as it will provide analytical formulas in some situations, thus decreasing significantly the computational burden induced by the multivariate integration in Eq. 5.

For our specific case study, possible auxiliary information are the DEM and the geometry of the river network. Indeed, the study area is a sandy aquifer with a high hydraulic conductivity and a draining river network. For these reasons, one thus expects to get water table elevation close to the DEM for locations that are close to a river, and the water table level is expected to be a smooth surface (because of the relative homogeneity and high conductivity of the aquifer) that (i) shows on the surface at the river network locations and (ii) that remains under the DEM at every other locations, of course. Accordingly, it is relevant to think about the water table elevation $Z(\mathbf{x}_i)$ as possibly modeled with

$$Z(\mathbf{x}_i) = DEM(\mathbf{x}_i) - g\left(d_{DEM}(\mathbf{x}_i)\right) + E(\mathbf{x}_i) \quad (6)$$

where $DEM(\mathbf{x}_i)$ is the DEM value at location $\mathbf{x}_i$, $d_{DEM}(\mathbf{x}_i)$ is a measure of the proximity of location $\mathbf{x}_i$ to the river network, $g(.)$ is an increasing non-negative function (see Section 3.2 for a concrete example) and $E(\mathbf{x}_i)$ is a zero-mean random error with a variance $\sigma_E^2(\mathbf{x}_i)$ that increases as the distance $d_{DEM}(\mathbf{x}_i)$ increases, i.e. the correspondence between $DEM(\mathbf{x}_i)$ and $Z(\mathbf{x}_i)$ is supposed to loosen as a location is further away from the network.

It is worth noting that $d_{DEM}(\mathbf{x}_i)$ is computed using an empirically defined penalized function on the changes of DEM along the path between the location and the network, in such a way that, for a same planar distance $d_{DEM}(\mathbf{x}_i)$ will be larger if the DEM varies highly along the path between $\mathbf{x}_i$ and the network than if the DEM is quite flat. The computation of this penalized function is similar to the computation of the distance that one should walk between the location and the network, except that vertical moves are highly penalized. As a consequence, $d_{DEM}(.)$ may be relatively small in areas where the DEM is quite constant even if these locations are quite far in a Euclidean viewpoint, whereas it may increase rapidly in areas where the DEM varies strongly, even if these points are quite close from an Euclidean viewpoint. Therefore, high DEM fluctuation areas will obtain high $d_{DEM}(\mathbf{x}_i)$ values, ensuring that this information will get less credit in the model since the corresponding variance $\sigma_E^2(\mathbf{x}_i)$ will be high.

Based on this model, we may thus predict water table elevations at any arbitrary location $\mathbf{x}_0$ as DEM values are exhaustively known over space and corresponding distances to the network are easily computed. Using only this information at location $\mathbf{x}_0$ in Eq. 5

(i.e. neglecting information at other surrounding locations $\mathbf{x}_U$), integration is not relevant anymore and the conditional pdf is then simply given by

$$f(z_0|\mathbf{z}_S, DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0)) \propto \frac{f(z_0|\mathbf{z}_S)}{f(z_0)} f(z_0|DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0)) \tag{7}$$

Assuming now that $f(z_0|DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0))$, $f(z_0)$ and $f(z_0|\mathbf{z}_S)$ are Gaussian distributed automatically implies that $f(z_0|\mathbf{z}_S, DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0))$ is also Gaussian distributed with a mean μ_P and a variance σ_P^2 that are given by (see Appendix for details)

$$\begin{cases} \mu_P &= \left(\dfrac{\mu_k}{\sigma_k^2} + \dfrac{\mu_d}{\sigma_E^2} - \dfrac{\mu_0}{\sigma_0^2} \right) \sigma_P^2 \\ \sigma_P^2 &= \left(\dfrac{1}{\sigma_k^2} + \dfrac{1}{\sigma_E^2} - \dfrac{1}{\sigma_0^2} \right)^{-1} \end{cases} \tag{8}$$

where μ_k and σ_k^2 are the mean and variance of distribution $f(z_0|\mathbf{z}_S)$ (i.e. the ordinary kriging prediction and variance of prediction, respectively), where $\mu_d = DEM(\mathbf{x}_0) - g(d_{DEM}(\mathbf{x}_0))$ and σ_E^2 are the mean and variance of distribution $f(z_0|DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0))$, and where μ_0 and σ_0^2 are the mean and variance of the *a priori* distribution $f(z_0)$.

Using only information at location $\mathbf{x}_0$, μ_p could be considered as a relevant choice for the predictor of the water table elevation at location $\mathbf{x}_0$ whereas σ_P^2 would be the associated prediction variance (remembering that $f(z_0|\mathbf{z}_S, DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0))$ is Gaussian distribution, μ_P is at the same time the mean, the median and the mode of this distribution, and μ_P along with σ_P^2 fully characterizes this pdf).

3. Application to the Dijle basin in Belgium

The study area is situated in Central Belgium where the geology is dominated by the Brussels Sands Formation, one of the main aquifers in Belgium for drinking water production. This Brussels Sands aquifer is of Middle Eocene age and consists of a heterogeneous alteration of calcified and silicified coarse sands [*Laga et al.*, 2001]. These sands are deposited on top of a clay formation of Early Eocene age, the Ieper Clay Formation, which forms the base of the aquifer in the study area. On the hilltops, younger sandy formations of Late Eocene to Early Oligocene age cover the Brussels Sands. These mainly consist of glauconiferous fine sands. The entire study area is covered with an eolian loess deposit of Quaternary age; in the north east of the study area, these deposits are more sandy loess.

The main river in the study area is the Dijle river and many of its tributaries cut through the Brussels Sands during the Quaternary, so in most of the valley floors the Brussels Sands are absent and groundwater flows in the alluvial deposits of the rivers on top the Ieper Clay formation. These alluvial deposits consist of gravels at the base, covered with peat and silt. In the river valleys a great number of springs drain the aquifer and provide the base flow for the river Dijle and its tributaries.

The hydraulic conductivity of the Brussels Sands varies between 6.9×10^{-5} m/s and 2.3×10^{-4} m/s, due to the heterogeneity of the Eocene aquifer [*Bronders*, 1989]. The calciferous parts of the aquifer have a lower conductivity than the silicified parts.

3.1. Data

Through the monitoring network of the Flemish Region (Databank Onderground Vlaanderen, http://dov.vlaanderen.be), piezometric data where gathered from 135 locations in the Dijle basin (see Fig. 1). The measuring frequency varies depending on the locations and most data are available from 2004 onwards. In this study, measurements between April 2005 and June 2005 were used. For locations where no measurements were available for this period in the year 2005, the average value for the April-June period was computed based on measurements in the preceding years. The piezometric measurements and the

DEM elevations are both expressed in elevation above sea-level. The whole data set thus consists of the 135 piezometric head measurements, planar coordinates and values, along with the DEM and the geometry of the river network over the area (see Figs. 1 and 2).

[Insert Figures 1 and 2 about here]

3.2. Results

According to the data at hand, several options can be considered for obtaining a map of predicted piezometric heads values over the area. The first one would be to only rely on piezometric measurements as classically done using ordinary kriging (OK). A second one would be to try improving the prediction by accounting at the same time for DEM values using ordinary cokriging (OCoK). Both of them are very classical ones, but it will be shown herafter that both fail to provide satisfactory results.

Using the 135 piezometric measurements, an experimental semi-variogram was computed and a semi-variogram was modeled (see Fig. 3). As suggested by Fig. 3 and alike [*Linde et al.*, 2007], a spherical model was chosen with theoretical variance and range equal to $300m^2$ and $13700m$, respectively. This high range value is reflecting a strong spatial dependence of the water table elevations, in close agreement with the geology and soil properties of the study area (i.e. coarse sand aquifer with high conductivity; see Section 3). It is also worth noting that fluctuations around the fitted model are only artifacts due to the relatively large distance lag in comparison with data set window.

[Insert Figure 3 about here]

Using only the 135 piezometric measurements, OK was conducted on a 525×525 regular grid covering the whole area in order to draw a map of piezometric head values (see Fig. 4; the 15 closest measurements where used at each prediction node). As emphasized by [*Hoeksema et al.*, 1989], water table elevations should of course never exceed DEM values, which is obviously not the case for OK prediction, especially at locations close to the network and far away from the sampled locations. In approximately 17% of the grid cells, the predicted head values are above the DEM-value and along the network the average overestimation of OK amounts to $6.3m$. One may think about correcting for this issue by replacing predicted values with the corresponding DEM elevations in this case. However, this option was discarded as it leads to obvious artifacts, like creating areas of constant piezometric heads as well as leading to non continuous derivatives of the piezometric heads along the border of these areas, which is in conflict with the expected physical behaviour of the aquifer.

[Insert Figure 4 about here]

As OK is only making use of the 135 piezometric head measurements, one may think about accounting for DEM information as well using OCoK. In order to do so, variogram and cross-variogram for DEM need to be estimated and modelled as well (not shown here). Because the spatial dependences of the DEM and the water table seemed to have different ranges, a combination of 2 spherical models with different ranges ($13700m$ and $6000m$) was needed. Using again the 15 closest measurement along with the DEM value at the corresponding prediction location, a OCoK prediction map was produced (see Fig. 5), very similar to the OK one. As for the OK map, approximately 17% of the predictions were above the corresponding DEM elevations with an average overestimation of $6.4m$ for the water table elevations along the network. Clearly, despite the fact that OCoK is using more information than OK, no real benefit can be observed, leading us to think that this multivariate approach is not particularly relevant here. Indeed, since OCoK belongs to the family of linear spatial predictors and since the relation between the water table and the DEM elevations is not linear (the DEM elevation is merely an upper bound for the water table), it explains easily why OCoK does not provide significant improvement in that study case and justifies the use of non linear approaches such as BDF.

[Insert Figure 5 about here]

In order to implement the BDF approach, a penalized distance $d_{DEM}(\mathbf{x}_i)$ (Section 2.2) was computed first at each of the 135 measurement locations. Plotting now $DEM(\mathbf{x}_i) - Z(\mathbf{x}_i)$ (i.e., groundwater depth) as a function of $d_{DEM}(\mathbf{x}_i)$ clearly shows that there exists on the average a non-linear relationship g(.) between these quantities, i.e. that (see Fig. 6)

$$DEM(\mathbf{x}_i) - Z(\mathbf{x}_i)) = g\left(d_{DEM}(\mathbf{x}_i)\right) + E(\mathbf{x}_i) \tag{9}$$

A logistic-like functional $g(.)$ was fitted from these observations, and a same logistic-like equation was used to model the way the variance of $E(\mathbf{x}_i)$ increases with the distance to the network. This choice for the function $g(.)$ is motivated by (i) the fact that the depth is expected to increase with the distance to the network and (ii) as the function $g(.)$ reaches a plateau for larger distances, it is also expected that for such distances one could only estimate a mean depth. Furthermore, as the variance was also modeled as a logistic-like function, the growth of this second function indicates that the information is loosing its influence on the fused pdf. As a consequence, the plateaus of the function $g(.)$ and of the conditional variance could be interpreted respectively as the unconditional mean depth and unconditional depth variance.

[Insert Figure 6 about here]

If we assume that $\mathbf{Z}$ is multivariate Gaussian distributed, the conditional pdf $f(z_0|\mathbf{z}_S)$ is Gaussian distributed too with a mean and a variance that correspond to OK prediction and variance of prediction, respectively. From Eq. 7, it is then easy to remark that in this case BDF amounts to updating the OK pdf by using information about the DEM and the river network as given by $f(z_0|DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0))$. As seen from Eq. 7, the BDF prediction map leads to much more satisfactory results. Contrary to OK or OCoK, only 0.023% of predicted values were above the corresponding DEM values. By comparison with Figs. 4 and 5, one can also notice that the DEM values and the network position were well accounted for, especially in locations close to the river network, of course, as the relationship between distance to network and groundwater depth is loosening up (i.e. variance of error increases) as this distance increases.

[Insert Figure 7 about here]

In order to validate our results, cross-validation was performed using a "leave-one-out" approach (see e.g. [*Chilès and Delfiner*, 1999]). For comparing the respective accuracies and precisions of the methods, the following indicators were chosen :

$$ME = \frac{1}{N}\sum_{i=1}^{N}\widehat{e_i} \tag{10}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\widehat{e_i}| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\widehat{e_i}^2} \tag{12}$$

where $\widehat{e_i}$ is the estimated error at sampled location $\mathbf{x}_i$, with $N = 135$. From Table 1 that summarizes the results, one can notice that the DEM values and the network position enabled us to increase the quality of the prediction since all indicators were found to be lower for BDF. The variance of prediction can also be used as an indicator of the expected local quality of the predicted map. Figs. 8a to 8c show this variance for the OK, OCoK and BDF predictions, respectively. A direct comparison of these figures indicates an important decrease of BDF variance especially in locations (i) that are close to the river or (ii) where DEM is flatter (northern area). This is a direct consequence of the information conveyed by the model as given in Fig. 8.

[Insert Table 1 about here]

[Insert Figure 8 about here]

Eventually, it is worth noting that none of the method (neither OK, OCoK nor BDF) provides pertinent predictions in the South-East area given the pour amount of information there (i.e. neither network positions nor sample locations). This reflection is totally in accordance with the variances of predictions shown in Figs. 8a to 8c.

4. Discussion and conclusions

In this paper, the recently developed spatial BDF technique as proposed by [*Bogaert and Fasbender*, 2007] was applied to the case study of water table elevation mapping. After a brief presentation of the method, specific assumptions were stressed in details and the method was illustrated with the case study of the north part of the Dijle basin in Belgium. A comparison of BDF with classically used spatial interpolation methods like ordinary (co)kriging showed that, to the contrary of cokriging, BDF is able to account for secondary information sources (namely a digital elevation model and the geometry of a river network in this case) both in a consistent and efficient way. Compared to standard multivariate methods (see e.g, [*Hoeksema et al.*, 1989] who used a multivariate model for the water table and ground elevations), BDF permits to avoid the need of defining a spatial multivariate model, that may be to restrictive or demanding. Though we did not mentioned it in this paper, several variations around kriging have been proposed to circumvent the limitations that were observed here (e.g., kriging with external drift was used by [*Desbarats et al.*, 2002]) for the water table prediction too). However, most of them lack sound theoretical rational and none of them can be generalized to the case of multiple secondary information.

Though more general, in this paper, the BDF method was implemented using (multivariate) Gaussian distributions. More than being particularly convenient (e.g. analytical expression for the final *posterior* distribution, fast implementation), this choice is also the Maximum Entropy solution when using the first two moments only (see e.g. [*Papoulis*, 1991]). On the other hand, by construction, the final *posterior* distribution suffers from several drawbacks (e.g. symmetric distribution, non-bounded support) which might influence unfavorably the results and the prediction maps for some applications. In particular, the multivariate Gaussian assumption might be replaced by other multivariate distributions that account for irregularities in the marginal distributions (see e.g. [?] for more details). However, in the present water table application, results were acceptable and significantly better than Ordinary (co-)kriging methods, so that these possible adaptations were left for further researches at this point.

One of the originality of this work was also to make use of a distance to a network for defining a non-linear relationship between a Digital Elevation Model (DEM) and water table elevations. Consequently, the proposed approach is less restrictive than cokriging as proposed by [*Hoeksema et al.*, 1989], in which the relation is purely linear by construction. There are also some similitudes with [*Linde et al.*, 2007] who proposed a Bayesian based solution for his specific data integration problem. However, the BDF framework as proposed here is intended to be more general and only rely on assumptions that are depending of the application at hand.

Finally, it is worth noting that the BDF methodology was illustrated here for the integration of only two secondary information sources, but the method can be readily implemented too in situations where more information sources might be available (see [*Bogaert and Fasbender*, 2007] for a detailed discussion about this topic), thus potentially improving the quality of the prediction and opening new avenues for the important topic of data integration in a spatial mapping context.

Acknowledgments

The authors would like to thank the Vlaamse Milieu Maatschappij for providing piezometric data. The authors also would like to express their gratitude towards the anonymous reviewers.

References

Bogaert, P. (2002), Spatial prediction of categorical variables: the bayesian maximum entropy approach, *Stoch. Env. Res. Risk A.*, *16*(6), 425–448.

Bogaert, P., and D. D'Or (2002), Estimating soil properties from thematic soil maps: the bayesian maximum entropy approach, *Soil Sci. Soc. Am. J.*, *66*(5), 1492–1500.

Bogaert, P., and D. Fasbender (2007), Bayesian data fusion in a spatial prediction context: A general formulation, *Stoch. Env. Res. Risk A.*, *21*, 695–709.

Bronders, J. (1989), Bijdrage tot de geohydrologie van midden-belgië door middel van geostatistische analyse en een numeriek model (contribution to the hydrogeology of central belgium by means of geostatistical analysis and numeric modelling), Ph.D. thesis, Vrije Universiteit Brussel.

Bruzzone, L., R. Cossu, and G. Vernazza (2002), Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images, *Inf. Fusion*, *3*, 289–297.

Chilès, J.-P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, John Wiley and Sons, Inc., New York, NY.

Christakos, G. (1990), A Bayesian/maximum-entropy view to the spatial estimation problem., *Math. Geol.*, *22*(7), 763–777.

Christakos, G. (1991), Some applications of the BME concepts in geostatistics., in *Maximum entropy and Bayesian methods*, edited by W. T. Grandy and L. H. Schick, pp. 215–229, Kluwer Acad. Publ.

Christakos, G. (1992), *Random Field Models in Earth Sciences*, Academic Press, San Diego, CA.

Christakos, G. (2000), *Modern Spatiotemporal Geostatistics*, Oxford University Press, New York.

Christakos, G. (2002), On the assimilation of uncertain physical knowledge bases: Bayesian and non-Bayesian techniques, *Adv. Water Resour.*, *25*(8-12), 1257–1274.

Christakos, G., and X. Y. Li (1998), Bayesian maximum entropy analysis and mapping: A farewell to kriging estimators?, *Math. Geol.*, *30*(4), 435–462.

Chung, A. C. S., and H. C. Shen (2000), Entropy-based Markov chains for multisensor fusion., *J. Intell. Robot. Syst.*, *29*, 161–189.

Cressie, N. (1990), The origings of kriging, *Math. Geol.*, *22*(3), 239–252.

Cressie, N. (1991), *Statistics for Spatial Data*, Wiley series in probability and mathematical statistics, John Wiley and Sons, Inc., New York, NY.

Desbarats, A., C. Logan, M. Hinton, and D. Sharpe (2002), On the kriging of water table elevations using collateral information from a digital elevation model, *J. Hydrology*, *255*, 25–38.

D'Or, D., and P. Bogaert (2004), Spatial prediction of categorical variables with the Bayesian Maximum Entropy approach: the Ooypolder case study, *Europ. J. Soil Science*, *55*, 763–775.

Fasbender, D., V. Obsomer, J. Radoux, P. Bogaert, and P. Defourny (2007), Bayesian data fusion : Spatial and temporal applications, in *Proceedings of the International Workshop on the Analysis of Multi-temporal Remote Sensing Images, 2007. MultiTemp 2007*, Leuven, Belgium.

Fasbender, D., J. Radoux, and P. Bogaert (2008), Bayesian data fusion for adaptable image pansharpening, *IEEE Trans. Geosci. Remote Sens.*, *46*, 1847–1857.

Hoeksema, R., R. Clapp, A. Thomas, A. Hunley, N. Farrow, and K. Dearstone (1989), Cokriging model for estimation of water table elevation, *Water Resour. Res.*, *25*, 429–438, doi:10.1029/88WR04044.

IAHS (2003), IAHS decade on predictions in ungauged basins (PUB), PUB Science and Implementation Plan 2003-2012.

Jones, G. D., R. E. Allsop, and J. H. Gilby (2003), Bayesian analysis for fusion of data from disparate imaging systems for surveillance., *Image and Vision Computing*, *21*, 843–849.

Laga, P., S. Louwye, and S. Geets (2001), Paleogene and neogene lithostratigraphic units (Belgium), *Geologica Belgica*, *4*, 135–152.

Linde, N., A. Revil, A. Bolène, C. Dagès, J. Castermant, B. Suski, and M. Voltz (2007), Estimation of the water table throughout a catchment using self-potential and piezometric data in a bayesian framework, *J. Hydrology*, *334*, 88–98.

Melgani, F., and S. B. Serpico (2002), A statistical approach to the fusion of spectral and spatiotemporal contextual information for the classification of remote-sensing images., *Pattern Recogn. Lett.*, *23*, 1053–1061.

Moshiri, B., M. R. Asharif, and R. HoseinNezhad (2002), Pseudo information measure : a new concept for extension of Bayesian fusion in robotic map building., *Inf. Fusion*, *3*, 51–68.

Papoulis, A. (1991), *Probability, random variables, and stochastic processes*, McGraw-Hill series in electrical engineering. Communication and signal processing., third ed., McGraw-Hill International Editions, Singapore.

Pieczynski, W., J. Bouvrais, and C. Michel (1998), Unsupervised Bayesian fusion of correlated sensors., in *Proceedings of the First International Conference on Multisource-Multisensor Information Fusion (Fusion'98)*, pp. 794–801, Las Vegas, Nevada.

Pinheiro, P., and P. Lima (2004), Bayesian sensor fusion for cooperative object localization and world modeling., in *Proceedings of the 8th Conference on Intelligent Autonomous Systems, IAS-8*, Amsterdam, The Netherlands.

Pradalier, C., F. Colas, and P. Bessiere (2003), Expressing bayesian fusion as a product of distributions: Application in robotics., in *Proceedings IEEE-RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*.

Rajan, D., and S. Chaudhuri (2002), Data fusion techniques for super-resolution imaging, *Inf. Fusion*, *3*(1), 25–38.

Serre, M. L., and G. Christakos (1999), Modern geostatistics: computational BME analysis in the light of uncertain physical knowledge - the equus beds study, *Stoch. Env. Res. Risk A.*, *13*(1-2), 1–26.

Simone, G., A. Farina, F. C. Morabito, S. B. Serpico, and L. Bruzzone (2002), Image fusion techniques for remote sensing applications, *Inf. Fusion*, *3*(1), 3–15.

Wibrin, M.-A., P. Bogaert, and D. Fasbender (2006), Combining categorical and continuous spatial information within the Bayesian maximum entropy paradigm., *Stoch. Env. Res. Risk A.*, *20*, 381–467.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner (2001), Spatial-temporal hierarchical Bayesian modeling: Tropical ocean surface winds., *J. Am. Stat. Assoc.*, *96*, 382–397.

Zhang, Z., and R. S. Blum (2001), A hybrid image registration technique for a digital camera image fusion application., *Inf. Fusion*, *2*, 135–149.

Appendix

Assuming that X is a Gaussian random variable with mean equal to b and variance equal to a^2, its pdf must be given by

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}a} e^{-\frac{1}{2a^2}(x-b)^2} \\ &\propto e^{-\frac{1}{2a^2}x^2 + \frac{b}{a^2}x} \end{aligned} \tag{13}$$

For the water table prediction study, it was shown that (see Eq. (7))

$$f(z_0|\mathbf{z}_S, DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0)) \propto \frac{f(z_0|\mathbf{z}_S)}{f(z_0)} f(z_0|DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0))$$

Assuming now that all these pdf's are Gaussian and plugging the corresponding expressions given by Eq. (13) into the previous equality, one obtains

$$f(z_0|\mathbf{z}_S, DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0)) \propto e^{-\frac{1}{2\sigma_E^2}(z_0-\mu_d)^2} e^{\frac{1}{2\sigma_0^2}(z_0-\mu_0)^2} e^{-\frac{1}{2\sigma_k^2}(z_0-\mu_k)^2}$$

$$\propto e^{-\frac{1}{2}\left(\frac{1}{\sigma_k^2}+\frac{1}{\sigma_E^2}-\frac{1}{\sigma_0^2}\right)z_0^2+\left(\frac{\mu_k}{\sigma_k^2}+\frac{\mu_d}{\sigma_E^2}-\frac{\mu_0}{\sigma_0^2}\right)z_0} \quad (14)$$

By direct identification between Eqs. (13) and (13), it is now easy to see that $f(z_0|\mathbf{z}_S, DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0))$ is also a Gaussian pdf with a mean μ_P and a variance σ_P^2 that are given by

$$\begin{cases} \mu_P &= \left(\frac{\mu_k}{\sigma_k^2}+\frac{\mu_d}{\sigma_E^2}-\frac{\mu_0}{\sigma_0^2}\right)\sigma_P^2 \\ \sigma_P^2 &= \left(\frac{1}{\sigma_k^2}+\frac{1}{\sigma_E^2}-\frac{1}{\sigma_0^2}\right)^{-1} \end{cases}$$

List of Tables

Table 1. Mean Error (ME), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of the leave-one-out procedure for ordinary kriging, ordinary cokriging and Bayesian data fusion predictions.

	ME [m]	MAE [m]	RMSE [m]
OK	-0.71	2.63	4.68
OCoK	-0.70	2.65	4.70
BDF	-0.31	2.45	3.94

List of Figures

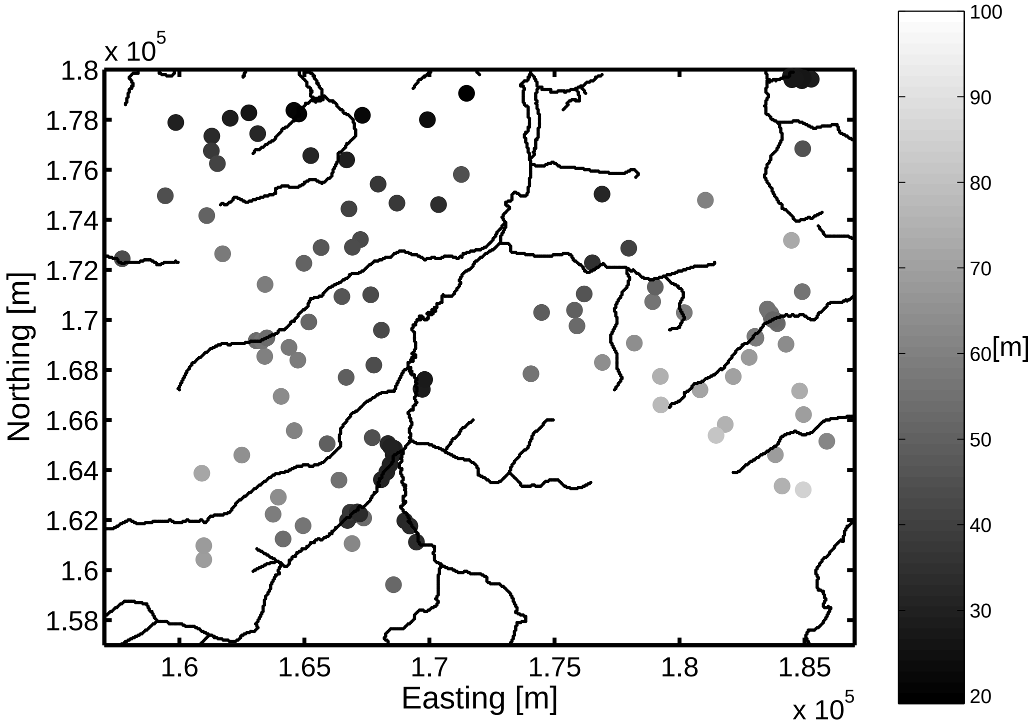

Figure 1. Sampled locations of the 135 piezometric heads values, with corresponding elevations above sea-level (in meters) as represented by color.

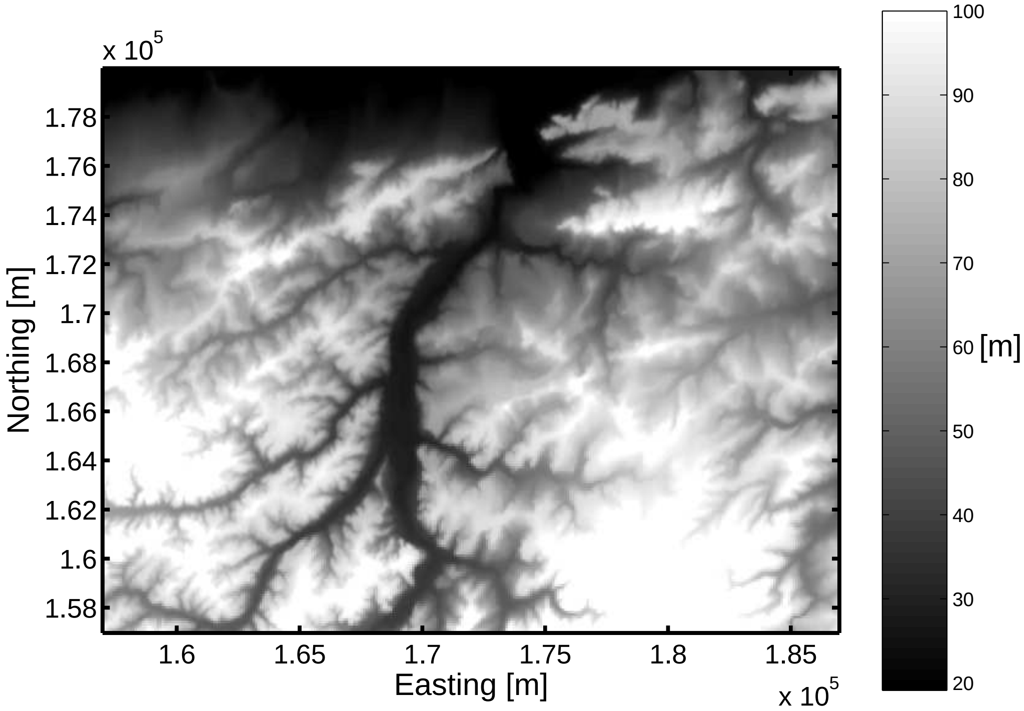

Figure 2. Digital Elevation Model of the study area (in meters above sea-level).

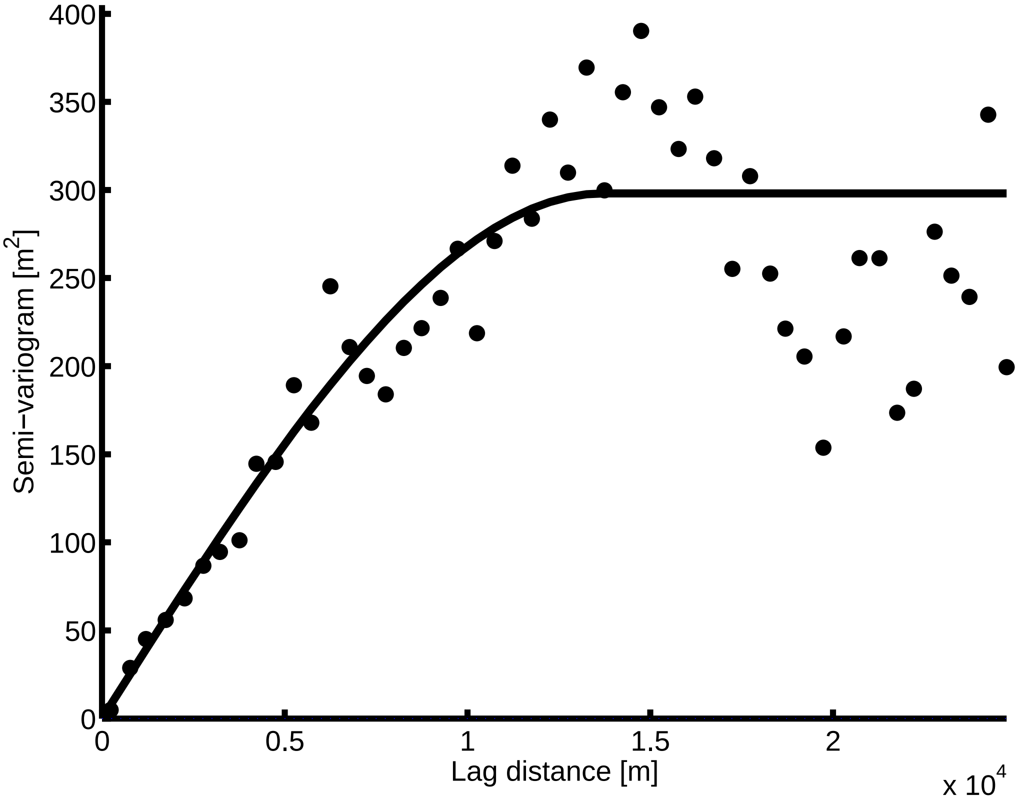

Figure 3. Experimental and modelled spatial semi-variogram based on the 135 raw piezometric measurements.

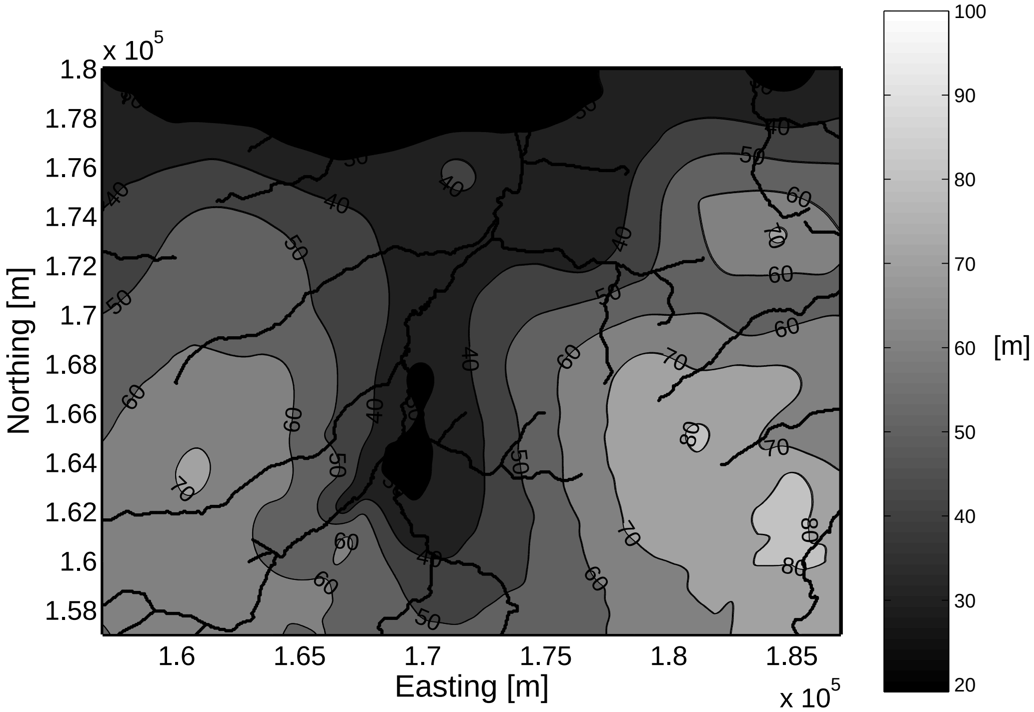

Figure 4. Prediction of the water table using ordinary kriging. The colormap convention is the same as in Fig. 1. Bold lines represent the river network over the study area.

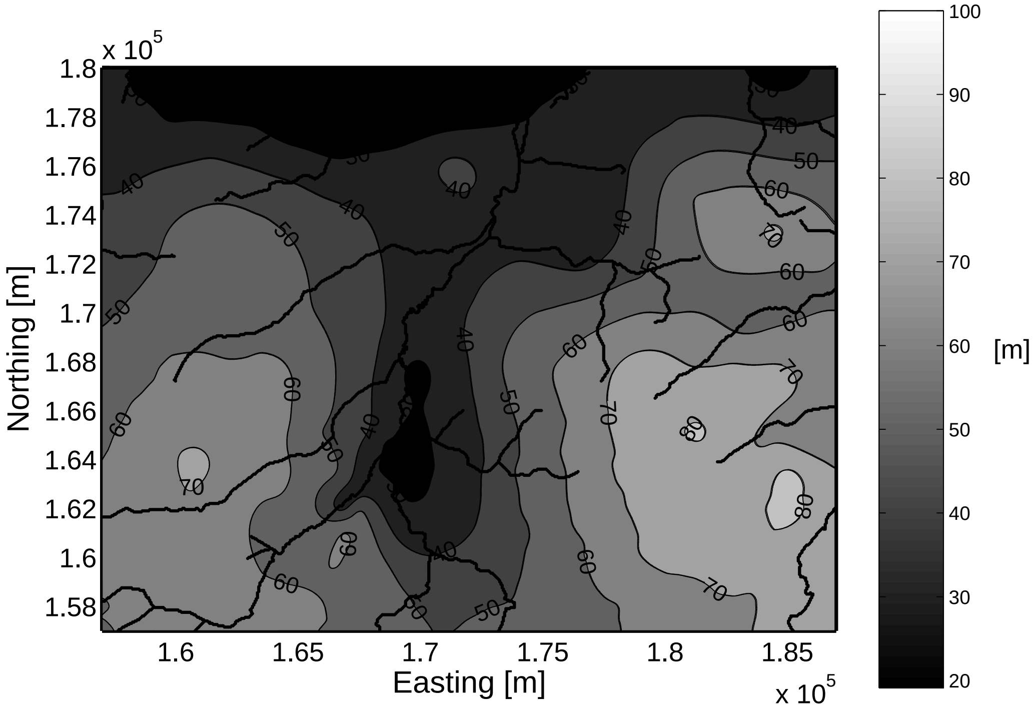

Figure 5. Prediction of the water table using ordinary cokriging. The colormap convention is the same as in Fig. 1. Bold lines represent the river network over the study area.

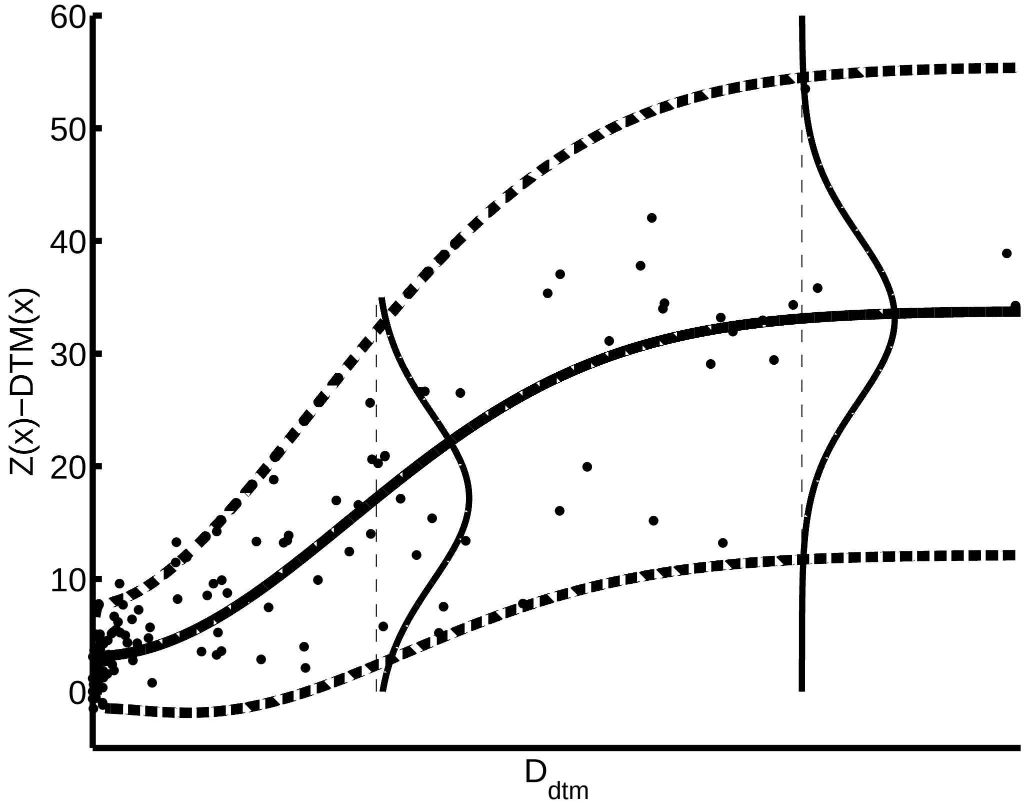

Figure 6. Graph of the groundwater depth $DEM(\mathbf{x}) - Z(\mathbf{x})$ as a function of the penalized distance $d_{DEM}(\mathbf{x})$ to the network. Dots represents the observed pair of values, plain line represents the fitted non-linear relationship $g(.)$ whereas dashed lines represent the 95% symmetric confidence interval based on a Gaussian distribution. The two Gaussian distributions overlayed on the graph illustrate the way variance of $E(\mathbf{x})$ is increasing with $d_{DEM}(\mathbf{x})$.

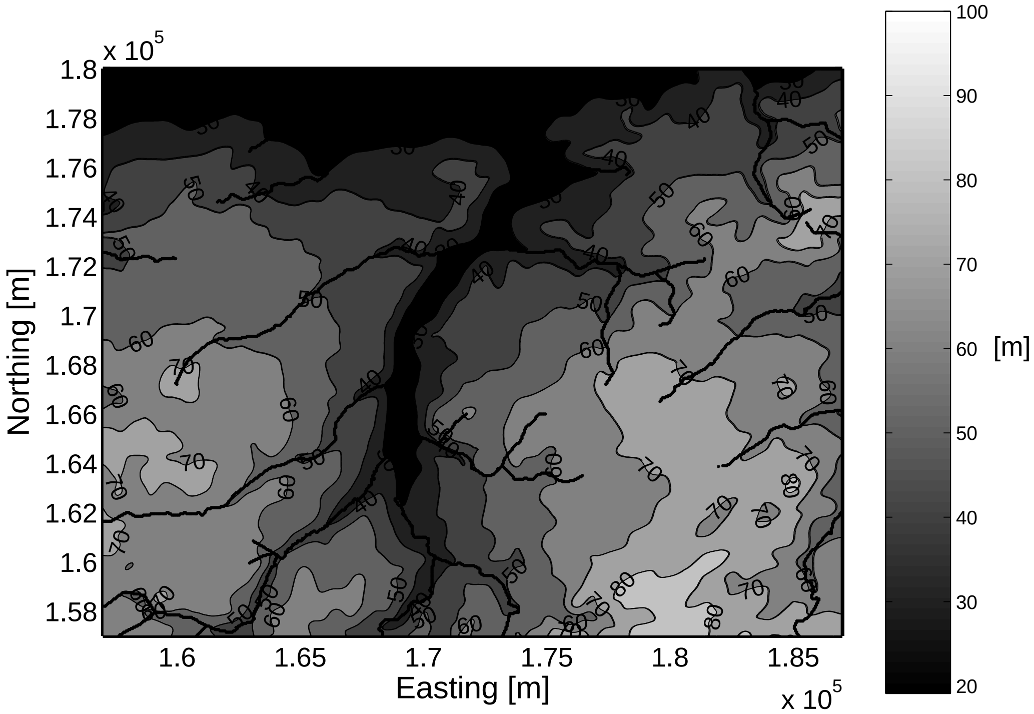

Figure 7. Prediction of the water table using the Bayesian data fusion approach. Bold lines represent the river network over the study area.

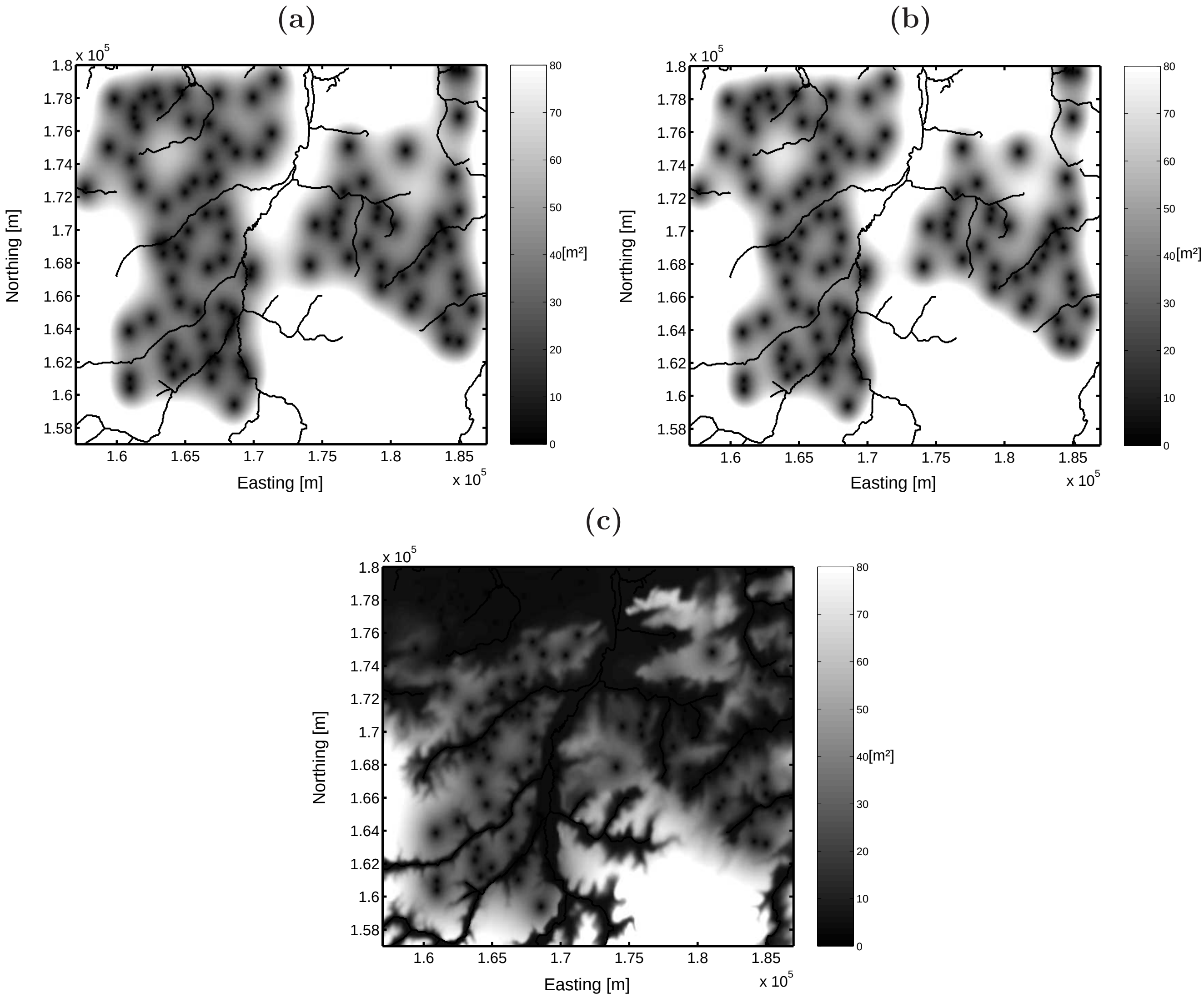

Figure 8. Variance of prediction of **(a)** ordinary kriging, **(b)** ordinary cokriging and **(c)** Bayesian data fusion methods. Bold lines represent the river network over the study area.

List of Captions

Figure 1. Sampled locations of the 135 piezometric heads values, with corresponding elevations above sea-level (in meters) as represented by color.

Figure 2. Digital Elevation Model of the study area (in meters above sea-level).

Figure 3. Experimental and modelled spatial semi-variogram based on the 135 raw piezometric measurements.

Figure 4. Prediction of the water table using ordinary kriging. The colormap convention is the same as in Fig. 1. Bold lines represent the river network over the study area.

Figure 5. Prediction of the water table using ordinary cokriging. The colormap convention is the same as in Fig. 1. Bold lines represent the river network over the study area.

Figure 6. Graph of the groundwater depth $DEM(\mathbf{x}) - Z(\mathbf{x})$ as a function of the penalized distance $d_{DEM}(\mathbf{x})$ to the network. Dots represents the observed pair of values, plain line represents the fitted non-linear relationship $g(.)$ whereas dashed lines represent the 95% symmetric confidence interval based on a Gaussian distribution. The two Gaussian distributions overlayed on the graph illustrate the way variance of $E(\mathbf{x})$ is increasing with $d_{DEM}(\mathbf{x})$.

Figure 7. Prediction of the water table using the Bayesian data fusion approach. Bold lines represent the river network over the study area.

Figure 8. Variance of prediction of **(a)** ordinary kriging, **(b)** ordinary cokriging and **(c)** Bayesian data fusion methods. Bold lines represent the river network over the study area.