Université catholique de Louvain
Faculté d'Ingénierie Biologique, Agronomique et Environnementale
Département des Sciences du Milieu et de l'Aménagement du Territoire

Bayesian data fusion in environmental sciences : Theory and applications

Dominique Fasbender

Novembre 2008

Thèse présentée en vue de l'obtention
du grade de docteur en Sciences Agronomiques
et Ingénierie Biologique

Promoteur : P. Bogaert

Composition du jury :

Président :	Pr Bruno Delvaux (UCL, Belgique)
Promoteur :	Pr Patrick Bogaert (UCL, Belgique)
Lecteurs :	Pr Marnik Vanclooster (UCL, Belgique)
	Pr Dominique Peeters (UCL, Belgique)
	Pr George Christakos (SDSU, USA)
	Pr Dionissios Hristopulos (TUC, Grèce)

Remerciements (Acknowledgement)

Je profites de ces quelques lignes pour remercier mes proches ainsi que tous ceux qui m'ont soutenu, encouragé, conseillé, supporté et parfois même subit...

Je pense en premier à mes collègues qui m'ont entouré durant ces 4 années d'assistanat dans l'Unité d'Environnemétrie et Géomatique. Autant leur interêt pour les images n'a pas manqué d'influencer mes travaux et mes applications, autant j'espère que mon goût prononcé pour les équations a su leur apporter un côté plus quantitatif et analytique dans leurs propres travaux de recherche. D'autre part, je garderai des souvenirs impérissables des pauses de 10h30 (qui commencent parfois à 11h15 selon les jours) et des temps de midi à la cafétaria ENGE.

En particulier, mes remerciements vont à Julien Radoux, Luk Peeters, Devis Tuia et Olivier Brasseur qui m'ont chacun permis d'approcher une question environnementale différente. Grâce à leurs expertises propres à chacun de ces domaines d'application, j'ai pu avoir un apperçu des thematiques qu'ils traitent dans leurs recherches et surtout ils m'ont aidé à écrire les quatres articles qui composent la partie pratique de cette thèse. De plus, bien sûr, les applications présentes dans cette thèse n'auraient pas pû être étudiées sans le support de la Belgian Scientific Policy pour avoir fourni les images IKONOS dans le cadre du STEREO earth observation program (Chapitre 3), le support de Swiss National Foundation à travers le projet "Urbanization Regime and Environmental Impact: Analysis and Modeling of Urban Metamorphoses" (n°100012-113506; Chapitre 4), celui de la Vlaamse Milieu Maatschappij pour avoir fourni les données piézométriques (Chapitre 5) et finalement les supports combinés de la Flemish Environment Agency en Flandres, du Brussels Institute for the Management of the Environment à Bruxelles, de l'Institute for Public Service et du Ministère Wallon de l'Environment pour avoir fourni les différentes données qui ont servi pour la prédiction des concentrations en NO_2 dans l'air (Chapitre 6).

Je remercie aussi les Professeurs Vanclooster, Peeters, Christakos et Hristopulous pour avoir accepté d'être membre de mon jury de thèse. Leurs commentaires et leurs conseils ont certainement été un apport important tout au long de ce travail. En particulier, je suis reconnaisant au Prof. Vanclooster pour la confiance qu'il m'a portée en acceptant d'être co-promoteur de mon mémoire de licence, ainsi qu'au Prof. Christakos pour m'avoir permis de faire un stage d'échange dans le département de Géographie de la San Diego State University en Californie. Ce fut une expérience inoubliable...

Bien sûr, je ne peux oublier de citer le Professeur Patrick Bogaert. Je ne serais pas là où je suis sans lui qui m'a fait confiance d'abord en encadrant mon mémoire de licence, puis en me proposant une place d'assistant comme financement pour cette thèse. La rigueur et l'esprit critique (dont il fait preuve) sont (sans aucun doute) une référence pour moi. Mais plus qu'un promoteur, c'est aussi devenu un ami avec qui j'ai partagé des discussions sérieuses quand il le faut et plus détendues dès que cela nous était possible.

Je souhaite évidemment remercier toute ma famille (et assimilé). Plus particulièrement, Charles, Stéphanie, Hati, Gwenaël, François, Olivier, Cathy, Jonathan, Alexandre, Chantal, Thierry, Bon-Papa Joël, Bonne-Maman Claire, Bon-Papa André, Bonne-Maman Jenny, Papa et Maman étaient toujours là pour m'aider à tenir bon durant ces 9 ans à Louvain-la-Neuve (Et moi qui, petit, disait à qui veut l'entendre que je n'aimais pas l'école ! J'y suis resté plus longtemps que la moyenne.) Ils m'ont souvent permis de relativiser et de me détendre durant les périodes plus stressantes des examens ou des remises de documents de thèse.

Finalement, je dois mentioner celle qui partage ma vie depuis presque 9 ans maintenant. Celle qui a toujours été là pour moi et qui m'a fait l'honneur d'accepter d'être mon épouse il y a un peu plus d'un mois. Fanny, merci d'avoir été là en première ligne durant toutes ces années. Tout ceci n'aurait pas été possible sans toi. C'est donc à toi que je dédie cette thèse.

Contents

List of Figures

List of Tables

Introduction

Motivation for the research

During the last thirty years, new technologies have contributed to a drastic increase of the amount of data in environmental sciences. Monitoring networks, remote sensors, archived maps and large databases are just few examples of the possible information sources responsible for this growing amount of information. Moreover, this is becoming more and more critical since every day monitoring networks and remote sensors get real-time data. For obvious reasons, it might be interesting to account for all these information when dealing with a space-time prediction/estimation context.

In environmental sciences, measurements are very often sampled scarcely over space and time. Geostatistics is the field that investigates variables in a space-time context. It includes a large number of methods and approaches that all aim at providing space-time predictions (or interpolations) for variables scarcely known in space and/or in time by accounting for space-time dependence between these variables (e.g. Cressie, 1991). As a consequence, geostatistical methods are relevant when dealing with the processing and the analysis of environmental variables in which space and time play an important role, even if the final purpose of the space-time predictions for these variables vary along with the application at hand (e.g., estimation of natural resources or estimation/prediction of pollutants).

As direct consequence of the increasing amount of data, there is an important diversity in the information. Indeed, data generally do not share the same nature so that it is very common to deal jointly with continuous, categorical or even mixed continuous-categorical variables. In addition, even variables of same nature generally do not share the same quality, thus carrying different uncertainties. These issues have recently motivated the emergence of the concept of data fusion (e.g. Mitchell, 2007). Broadly speaking, the main objective of data fusion methods is to deal with various information sources in such a way that the final result is a single prediction that accounts for all the sources at once. This enables us thus to conciliate several and potentially contradictory sources instead of having to select only one of them because of a lack of appropriate methodology.

For most of existing geostatistical methods, it is quite difficult to account for a potentially large number of information sources at once. Furthermore, geostatistical methods are also generally dedicated to one particular kind of data. As a consequence, often one has to opt for only one information source among all the available sources. This of course leads to a dramatic loss of information. In order to avoid such choices,

it is thus relevant to bring together the concepts of data fusion and geostatistics in the context of environmental sciences.

It is worth noting that the idea of accounting for several information sources is not new in geostatistics. Several attempts have been proposed in the literature in order to solve these data assimilation issues. Unfortunately, solutions proposed so far by these methods are usually dedicated to specific issues at the price of complicated models or strong modeling hypotheses. As a consequence, they are rather difficult to generalize to other situations. Furthermore, most of these methods have major drawbacks which are discussed hereafter.

State-of-the-art

As mentioned above, geostatistics is a general term that includes a wide range of methods and approaches that are all dedicated to the analysis of space-time variables. Broadly speaking, there is a distinction between deterministic (e.g. Dempster-Shafer theory)and stochastic approaches. In the context of space-time analysis of environmental variables, stochastic approaches have shown their advantages with respect to the deterministic ones (e.g. Krige, 1951; Cressie, 1991; Christakos, 1992; Goovaerts, 1997), so that stochastic approaches only will be considered here in this thesis. Furthermore, among stochastic approaches, one can also distinguish between Bayesian methods (e.g. Bayesian Maximum Entropy, Bayesian kriging) and non-Bayesian ones (e.g. kriging, cokriging, kriging with external drift, residual kriging, stratified kriging). We describe hereafter a non-exhaustive list of these methods with their own advantages and drawbacks.

Initially proposed in the fifties, kriging is a stochastic prediction method widely used in geostatistics (e.g. Krige, 1951; Journel and Huijbregts, 1978; Cressie, 1990). Kriging approach has a lot of similarities with Generalized Linear Models (GLM; see e.g. Green and Silverman, 1994). Kriging is also known as the Best Linear Unbiased Predictor (BLUP) in the sense that it minimizes the variance of prediction among the family of linear predictors that respect the mean of the predicted variable (unbiased predictors). This principle is found in each of the kriging-related variations described hereafter. Further, kriging is known to be the global optimal predictor under Gaussian assumptions of the random fields. Though initially dedicated to the spatial prediction of a unique variable, kriging was rapidly generalized so that secondary information can be accounted for in the predictions.

Cokriging (Cressie, 1991) is a straightforward extension of kriging. It constitutes a family of methods (e.g. simple cokriging, universal cokriging) that use continuous secondary variables for the prediction of the variable of interest. Similarly to kriging, cokriging is also known to be the best predictor under multivariate Gaussian assumptions. Cokriging is theoretically sound as it accounts for the uncertainty of the relationships between both secondary and primary variables. Unfortunately, in cokriging, these relationships are assumed to be linear only, thus limiting the scope of the method when applied to more complex situations. Moreover, cokriging generally relies on the so-called Linear Model of Coregionalization (see Chilès and Delfiner, 1999 or Goovaerts, 1997) for the estimation of the space-time dependence, which is a rather strong modeling hypothesis.

The presence of gradients (also known in geostatistics as the trend) in the sampled variables often happens in environmental applications. This trend is generally difficult to infer from the raw observations only. Kriging with External Drift (Desbarats *et al.*, 2002) was proposed in order to circumvent this issue by modeling the trend of the variable of interest with several continuous secondary variables. Unfortunately, this method is not able to account naturally for the uncertainty of the estimated trend [1]. Furthermore, the method can not be applied if secondary variables are scarcely known in space or in time.

There is a classical distinction between continuous and categorical stochastic variables. Though kriging with external drift and cokriging were designed for continuous secondary variables, residual kriging (Goovaerts, 1997) and stratified kriging (Brus *et al.*, 1996) enable us to account for categorical secondary variables. Unfortunately, the main drawbacks of both approaches are similar to those of kriging with external drift, namely (i) categorical secondary variables must be known exhaustively and (ii) the uncertainty of the estimated trend is not taken into account.

Though kriging-related methods have successfully been applied in environmental sciences during the last fifty years (e.g. Krige, 1951; Yamamoto, 1999; Costa, 2003; Jaquet, 1989; Hoeksema *et al.*, 1989), Bayesian approaches were rapidly proposed as alternatives (Christakos, 1990, 1991). Among them, the Bayesian Maximum Entropy (BME) paradigm gathers several methods inside a same general framework. BME is a non-linear method that relies on the use of a *prior* distribution and on a *Bayesian conditionalization rule* for the assimilation of

[1] It is however possible to sum the uncertainties coming from the kriging prediction and from the estimated trend, assuming thus that these estimations are uncorrelated.

secondary information. BME enables us to successfully predict continuous variables (Christakos, 1992, 2000; Christakos and Li, 1998; Serre and Christakos, 1999), categorical variables (Bogaert, 2002; Bogaert and D'Or, 2002; D'Or and Bogaert, 2004) and even mixed random fields (Wibrin *et al.*, 2006) whatever the nature of the secondary information, making thus of BME a complete framework in the context of space-time prediction. Moreover, BME has the great advantage that it rebuilds the whole distribution instead of simply providing one or several of its moments as it is the case for kriging-related methods which only focus on the mean and the variance. Various adaptations of BME were also proposed when several information sources are available (e.g. mutivariate BME; see Christakos and Li, 1998; Christakos, 2000), but unfortunately they have the same kind of limitations as kriging-related approaches (e.g. relying on the linear model of coregionalization).

Objectives

The main objective of this thesis is to develop a sound theoretical framework that enables us to account for several information sources in a space-time prediction context. For this, this thesis focuses on a Bayesian data fusion (BDF) framework that gathers the concepts of both data fusion and geostatitics. This should enable us to account for as much information sources as available which is an important limitation for most of existing geostatistics methods.

It is crucial that the BDF framework is able to deal with diverse information sources. Indeed, we mentioned that geostatistics methods generally are devoted to one specific secondary information only and are hard to generalize for other situations. For this, the proposed BDF framework will rely on weak hypotheses and thus will be as general as possible regarding the potential information sources to be fused.

In order to be more conscious of the pros and cons of the proposed BDF framework, explicit hypotheses will be assumed along with the presentation of the method. Thanks to this, it will be potentially easier to adapt the proposed BDF framework to other situations where other hypotheses might hold true (e.g. additional relationships between variables that were neglected in previous hypotheses).

Finally, the second objective of this thesis is to illustrate how the proposed BDF framework can account for divers information sources in a space-time context. The method will thus be applied to a few environmental sciences applications for which (i) crucial available information

sources are typically difficult to account for or (ii) the number of secondary information sources is a limitation when using existing methods.

Outline of the thesis

This thesis is divided into six chapters where the two first ones refer to the theoretical developments of the proposed BDF framework and the four remaining ones are presenting various applications of the method.

The first chapter sets the theoretical background for the rest of the thesis. Chapter 2 emphasizes the importance of the set of hypotheses that are assumed in the BDF framework along with their influence on the prediction results. For this purpose, BDF and simple cokriging methods are compared in controlled conditions. Chapters 3 and 4 show how BDF can be suitable for the enhancement of coarser multispectral images with higher spatial resolution images, illustrating thus the case of data exhaustively known over space. Chapter 5 presents an application of BDF for the spatial estimation of water table depth. In this application, the real difficulty is to account for the river network position that does not straightforwardly fit in classical geostatistical methods. Chapter 6 shows how BDF can be extended to space-time applications. This last application was particularly interesting since up to 6 secondary information sources were available at some locations, illustrating thus very well the potential of BDF when dealing with numerous and very diverse information sources, thus emphasizing the generality and versatility of this methodology.

Part I

Theory

Chapter 1

Bayesian data fusion in a spatial prediction context : a general formulation

In this first chapter, we present the theory of a Bayesian Data Fusion framework in a spatial prediction context. This first work is an angular part of the thesis as it sets the theoretical background for the rest of the thesis.

1.1 Outline

In spite of the exponential growth in the amount of data that one may expect to provide greater modeling and prediction opportunities, the number and diversity of sources over which this information is fragmented is growing at an even faster rate. As a consequence, there is real need for methods that aim at reconciling them inside an epistemically sound theoretical framework. In a statistical spatial prediction framework, classical methods are based on a multivariate approach of the problem, at the price of strong modeling hypotheses. Though new avenues have been recently opened by focusing on the integration of uncertain data sources, to the best of our knowledge there have been no systematic attempts to explicitly account for information redundancy through a data fusion procedure.

Starting from the simple concept of measurement errors, this chapter proposes an approach for integrating multiple information processing as a part of the prediction process itself through a Bayesian approach. A general formulation is first proposed for deriving the prediction distribution of a continuous variable of interest at unsampled locations using

more or less uncertain (soft) information at neighboring locations. The case of multiple information sources is then considered, with a Bayesian solution to the problem of fusing multiple information that are provided as separate conditional probability distributions. Well-known methods and results are derived as limit cases. The convenient hypothesis of conditional independence is discussed in the light of information theory and the maximum entropy principle, and a methodology is suggested for the optimal selection of the most informative subset of information, if needed. Based on a synthetic case study, an application of the methodology is presented and discussed [1].

[1] With kind permission from Springer Science+Business Media (Bogaert and Fasbender, 2007).

1.2 Introduction

As emphasized by van der Putten *et al.* (2002), in spite of the exponential growth in the amount of data that one may naively expect to provide greater opportunities for improving modeling and predictions, reality is somewhat different. In many real world applications, the number and diversity of sources, over which this information is fragmented, grows at an even faster rate, thus making this information more and more difficult to combine. As a consequence, there is a growing need for methods that aim at reconciling different information sources inside a unique and sound theoretical framework.

Data fusion is a generic expression widely found in the literature, that conveys the idea of combining at best information coming from different sources in order to achieve improved accuracy and better inferences. In principle, fusion of multiple sources of information provides significant advantages over single source data, as improved estimate of the physical phenomena should be obtained via additional information. Although the general idea is rather simple, the way this goal is precisely achieved and its objectives are however diverse as well as widely dependent on the field of application.

In data mining applications for marketing purposes, data fusion techniques typically aim at achieving a complete data file from different sources (e.g. databases) which do not contain the same units (Cho *et al.*, 2003; Rässler, 2004), a method also called statistical matching in this context. In the field of applied computer sciences that involves multiple data sources such as sensor readings or model decision, data fusion is a main concern due to the widely different nature of the outputs that are generated. Among others, this involves for example biometrics verification systems (e.g., Duc *et al.*, 1997; Ross and Jain, 2003), surveillance systems (e.g., Jones *et al.*, 2003), robotics (e.g., Cremer *et al.*, 2001; Pradalier *et al.*, 2003), medical imagery (e.g., Song *et al.*, 2003) or military/civil engineering (e.g., Gros *et al.*, 1999; Sohn and Lee, 2003). Data fusion also has important applications in classification of remote sensing images (e.g., Costantini *et al.*, 1997; Melgani and Serpico, 2002; Simone *et al.*, 2002) and in environmental modeling.

Besides this diversity of objectives and applications, the way data fusion is precisely achieved covers a wide spectrum of methods and topics as diverse as neural networks, k-means clustering, fuzzy logic, Kalman filtering, Dempster-Shafer theory of evidence, or traditional Bayesian theory (just to quote few of them) that may involve information fusion at different levels (data-level, feature-level or decision-level; see Sohn and Lee, 2003 or Ross and Jain, 2003 for a description of these concepts).

However, as emphasized by Fassinut-Mombot and Choquel (2004), the most classical techniques of information fusion are based on probability theory associated with Bayesian decision theory. In a spatiotemporal study of wind speed over ocean surface, Wikle *et al.* (2001) shows how, for example, combining in a Bayesian procedure two different sets of measurements (namely satellite data and output from models based on synoptic measurements) that are indirect measurements of the same underlying true (but unknown) values may lead to serious improvements for prediction.

Although Bayesian methods – or variations around them – have been widely used for fusing collocated information (i.e., coregistered information in the remote sensing terminology), there have been little attempts to integrate multiple redundant information through a data fusion process in a spatial prediction framework, where what is typically sought for is (a distribution of) predicted values at unsampled locations, to be obtained from the knowledge of the same or different sampled variables at a set of nearby locations. Standard methods rather aim at using these variables into a classical multivariate framework, at the price of strong hypotheses for modeling their joint dependence (see, e.g., Cressie, 1993, pg 141; Goovaerts, 1997, pg 113; Wackernagel, 1995, pg 152; Chilès and Delfiner, 1999, pg 339 for detailed description and discussion about these methods) that may even lead to the use of dimensionality reduction techniques when the number of variables is too high. It is worth noting at this point that "multivariate" refers in our context to the case of multiple variables representing different attributes in the reality. We thus will consider later the distinction between the primary variables (i.e. the variables of direct interest for the prediction) and the secondary variables (i.e. any additional information sources linked somehow to the primary variables).

New avenues have been recently opened by Christakos (2000) and Christakos *et al.* (2002), who proposed the so-called set of Bayesian Maximum Entropy (BME) methods that aim at rigorously accounting for data uncertainty in the prediction process. A comparison of the BME methods with non Bayesian approaches can also be found in Christakos (2002). Although these methods have proved to be successful in a wide range of applications (see, e.g., Bogaert and D'Or, 2002; D'Or and Bogaert, 2004; Savelievaa *et al.*, 2005), data fusion as a way of accounting for information redundancy is not a part of the process.

Using the classical concept of measurement errors embedded into a Bayesian framework, this chapter is an attempt to show how an efficient general formulation of the spatial prediction problem can easily be achieved. The first part of the chapter mainly focuses on a presentation

of general theoretical results. Starting from an additive error model, it is first shown how uncertainty can be accounted for at the price of mild independence hypotheses. Connections with well-known methods and results that are obtained as singular cases are emphasized. With the help of information theory, the rationale behind some of these independence hypotheses is explained, and a procedure for selecting the most useful subset of information to be fused is suggested too. Using afterward the concept of dual notation for information, simpler equations involving only prior and conditional distributions are obtained. In the second part of the chapter, the concept of Bayesian data fusion is presented. It is shown how multiple collocated information can be merged into a single conditional distribution. Specific results for the Gaussian case are presented, as well as the possibility of defining an informativeness index, measuring the amount of information brought by any conditional distribution. Finally, the third part of the chapter aims at presenting a simplified illustration of some of these concepts based on a synthetic case study.

1.3 A prediction model with uncertain information

Assume that $\mathbf{Z} = (Z_0, \ldots, Z_n)'$ is a random vector sampled from a spatial random field (RF) $\mathfrak{Z} = \{Z(\mathbf{x}, \omega) : \mathbf{x} \subset D \subseteq \mathbb{R}^d, \omega \subset \Omega, Z(\mathbf{x}, \omega) \in \mathbb{R}\}$ (where Ω is the set of all possible random event) and that we know the joint multivariate probability distribution function (pdf) $f(\mathbf{z})$, where each Z_i is associated with a different spatial location $\mathbf{x}_i$ in our context. We will consider that it is not possible to directly observe z_i values for these variables, but that instead observed y_i values are available, where the Y_i's are functionally related to the Z_i's up to random errors E_i's. The Y_i's may possibly (but not necessarily) correspond partially or totally to different physical variables, so that the random vector $\mathbf{Y} = (Y_0, \ldots, Y_n)'$ should be merely regarded as a collection of (possibly correlated) random variables rather than as resulting from the sampling of a same spatial RF $\mathfrak{Y}$ at locations $\mathbf{x}_0, \ldots, \mathbf{x}_n$. It will be considered here that both Z_i and E_i (and consequently Y_i) are of the continuous type, though all results that will be presented can be extended to discrete or mixed random variables without problem.

1.3.1 The case of simple additive errors

Let us consider first hereafter the simple case where $Y_i = Z_i + E_i$, so that in vector notations $\mathbf{Y} = \mathbf{Z} + \mathbf{E}$ (the general case of an arbitrary relationship $\mathbf{Y} = \mathbf{g}(\mathbf{Z}) + \mathbf{E}$ will be presented later on). Given the fact that values $\mathbf{y} = (y_0, \ldots, y_n)'$ are observed, what is sought for is the conditional pdf $f(\mathbf{z}|\mathbf{y})$, where

$$f(\mathbf{z}|\mathbf{y}) = f(z_0, \ldots, z_n | z_0 + e_0, \ldots, z_n + e_n)$$

For deriving this pdf, let us make the hypothesis that $\mathbf{E} \perp \mathbf{Z}$, i.e. $\mathbf{E}$ can be considered as stochastically independent from $\mathbf{Z}$. Using Bayes theorem,

$$f(\mathbf{z}|\mathbf{y}) = \frac{f(\mathbf{y}|\mathbf{z})f(\mathbf{z})}{\displaystyle\int_{\mathbb{R}^n} f(\mathbf{y}|\mathbf{z})f(\mathbf{z})\mathrm{d}\mathbf{z}} = \frac{1}{A} f(\mathbf{y}|\mathbf{z})f(\mathbf{z}) \tag{1.1}$$

where A plays the role of a normalization constant. Because we assumed that $\mathbf{Y} = \mathbf{Z} + \mathbf{E}$ with $\mathbf{E} \perp \mathbf{Z}$, it is easy to see that

$$F(\mathbf{y}|\mathbf{z}) = P(\mathbf{Z} + \mathbf{E} < \mathbf{y}|\mathbf{z}) = P(\mathbf{E} < \mathbf{y} - \mathbf{z}) = F_{\mathbf{E}}(\mathbf{y} - \mathbf{z})$$

so that we have

$$f(\mathbf{y}|\mathbf{z}) = \frac{\partial}{\partial \mathbf{y}} F(\mathbf{y}|\mathbf{z}) = f_{\mathbf{E}}(\mathbf{y} - \mathbf{z}) \tag{1.2}$$

It is worth noting that Eq. (1.2) is directly linked to the classical convolution theorem for obtaining the joint distribution of a sum of random variables (see e.g. Papoulis, 1991, pg 136), extended here in the multivariate case, so that

$$f(\mathbf{y}) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{y}|\mathbf{z})f(\mathbf{z})\mathrm{d}\mathbf{z} = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{E}}(\mathbf{y} - \mathbf{z})f(\mathbf{z})\mathrm{d}\mathbf{z}$$

However, in our case we are interested in $f(\mathbf{z}|\mathbf{y})$ instead, so plugging Eq. (1.2) into Eq. (1.1) gives

$$f(\mathbf{z}|\mathbf{y}) = \frac{1}{A} f_{\mathbf{E}}(\mathbf{y} - \mathbf{z})f(\mathbf{z}) \tag{1.3}$$

As prediction is generally sought for the single z_0, we will focus on the marginal conditional pdf at location $\mathbf{x}_0$, that can be obtained by integrating Eq. (1.3) over other variables, so that

$$f(z_0|\mathbf{y}) = \frac{1}{A} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{E}}(\mathbf{y} - \mathbf{z})f(\mathbf{z})\mathrm{d}z_1 \cdots \mathrm{d}z_n \tag{1.4}$$

Further simplifications can be obtained by assuming the mutual independence $E_0 \perp \cdots \perp E_n$ for the errors, so that

$$f_{\mathbf{E}}(\mathbf{y}-\mathbf{z}) = \prod_{i=0}^{n} f_{E_i}(y_i - z_i) \quad \Longleftrightarrow \quad f(\mathbf{y}|\mathbf{z}) = \prod_{i=0}^{n} f(y_i|\mathbf{z}) \tag{1.5}$$

showing that $(Y_0 \perp \cdots \perp Y_n)|\mathbf{z}$, and we finally obtain

$$f(z_0|\mathbf{y}) = \frac{1}{A} f_{E_0}(y_0 - z_0) \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{z}) \prod_{i=1}^{n} f_{E_i}(y_i - z_i) \mathrm{d}z_1 \cdots \mathrm{d}z_n \tag{1.6}$$

In summary, Eq. (1.4) allows to derive the conditional pdf of Z_0 at an arbitrary location $\mathbf{x}_0$ if we have observed values for variables $\mathbf{Y}$ that are linearly related to $\mathbf{Z}$ up to random errors $\mathbf{E}$ assuming that these errors are independent from $\mathbf{Z}$, where Eq. (1.6) assumes additionally that these errors are mutually independent. Thanks to this hypothesis, one thus separates the problem in (i) the spatial dependence of the Z_i's variables (accounted for in the pdf $f(\mathbf{z})$) and (ii) the gain of information from the Y_i's (coded as the conditional distributions $f_{E_i}(y_i - z_i)$), while classical multivariate methods (e.g. cokriging ones) model the influence of all the Y_i's variables on the variable to be predicted Z_0 directly (see Fig. 1.1).

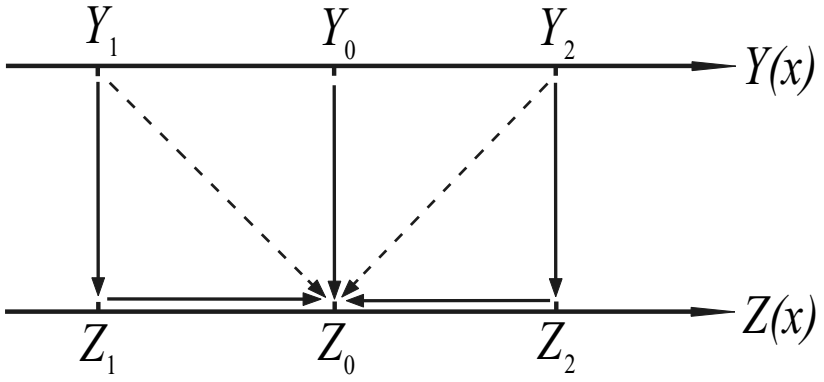


Figure 1.1: Plain lines illustrate the methodology used in the BDF approach when accounting for secondary variables whereas dashed lines correspond to the methodology used in classical multivariate methods (e.g. cokriging ones).

Though the previous developments have been conducted out of the context of classical methods in spatial statistics or geostatistics, it is easy to show that they are intimately related to well-known approaches. A

particular interpretation of Eq. (1.6) is to consider that each Y_i variable can be viewed as an indirect measurements of the true Z_i through, e.g., a same imperfect measuring device, so that each measurement Y_i includes an additive measurement error, leading to the relation $Y_i = Z_i + E_i$, where it is reasonable in general to assume that errors are mutually independent and do not depend on the true value. This is similar, e.g., to the hypothesis by e.g. Wikle *et al.* (2001) and Pradalier *et al.* (2003), who assume that, conditionally to the true values, multiple measurements of the same underlying unknown quantity can be conveniently viewed as independent from each others, thus leading to simple expressions involving the product of the corresponding various pdf's. Assuming now additionally that these errors are $N(0, \sigma_i^2)$ and that $\mathbf{Z} \sim N(\boldsymbol{\mu}, \Sigma)$, it can be shown that the conditional mean $\mathbb{E}[Z_0|\mathbf{y}]$ and the conditional variance $\mathbb{V}ar[Z_0|\mathbf{y}]$ are identical to the kriging predictor and variance as obtained from the so-called simple kriging with measurement errors (Cressie, 1993; see Appendices for the proof).

1.3.2 Generalization to arbitrary functionals

Eq (1.6) can be slightly generalized if one considers that the observable Y_i are linked to the Z_i's through various arbitrary functionals (see Fig. 1.2), so that $Y_i = g_i(Z_i) + E_i$ (we will restrict here the developments to the case were each Y_i is only functionally dependent of a single corresponding Z_i, though a similar approach can be applied when Y_i possibly depends on several variables in $\mathbf{Z}$). It is worth noting here that the Y_i's may refer to different physical variables so that the $g_i(.)$'s functions are also different in general. We thus have $\mathbf{Y} = \mathbf{g}(\mathbf{Z}) + \mathbf{E}$, where $\mathbf{g}(\mathbf{z}) = (g_0(z_0), \ldots, g_n(z_n))'$, and what is sought for is

$$f(\mathbf{z}|\mathbf{y}) = f(z_0, \ldots, z_n | g_0(z_0) + e_0, \ldots, g_n(z_n) + e_n)$$

Reasoning along the same lines as before with $\mathbf{Y} = \mathbf{g}(\mathbf{Z}) + \mathbf{E}$ and $\mathbf{E} \perp \mathbf{Z}$,

$$F(\mathbf{y}|\mathbf{z}) = P(\mathbf{g}(\mathbf{Z}) + \mathbf{E} < \mathbf{y}|\mathbf{z}) = P(\mathbf{E} < \mathbf{y} - \mathbf{g}(\mathbf{z})) = F_{\mathbf{E}}(\mathbf{y} - \mathbf{g}(\mathbf{z})) \tag{1.7}$$

so that we obtain similarly

$$f(\mathbf{y}|\mathbf{z}) = \frac{\partial}{\partial \mathbf{y}} F(\mathbf{y}|\mathbf{z}) = f_{\mathbf{E}}(\mathbf{y} - \mathbf{g}(\mathbf{z}))) \tag{1.8}$$

that can be plugged into Eq. (1.1). If the mutual independence hypothesis for the errors is assumed, this simplifies again to

$$f_{\mathbf{E}}(\mathbf{y} - \mathbf{g}(\mathbf{z})) = \prod_{i=0}^{n} f_{E_i}(y_i - g_i(z_i)) \quad \Longleftrightarrow \quad f(\mathbf{y}|\mathbf{z}) = \prod_{i=0}^{n} f(y_i|\mathbf{z})$$

and we finally obtain the equivalent of Eq. (1.6), with

$$f(z_0|\mathbf{y}) = \frac{1}{A} f_{E_0}(y_0 - g_0(z_0)) \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{z}) \prod_{i=1}^{n} f_{E_i}(y_i - g_i(z_i)) \mathrm{d}z_1 \cdots \mathrm{d}z_n \tag{1.9}$$

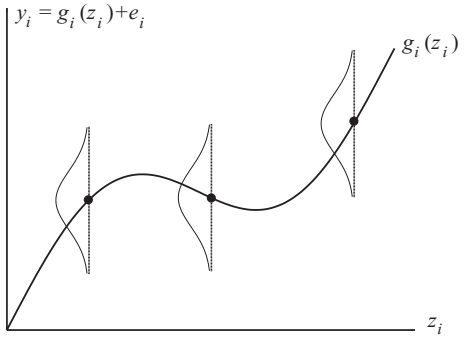


Figure 1.2: Arbitrary functional $y_i = g_i(z_i) + e_i$

As an illustration, Fig. 1.3 shows the graphs of the functions $f_{E_i}(y_i - g_i(z_i))$ with respect to z_i, computed for different observed values y_i and for the arbitrary functional $g_i(.)$ given in Fig. 1.2. One can remark that, because of the possibly non linear and non monotonic relationship between Y_i and Z_i, these functions may exhibit complex shapes.

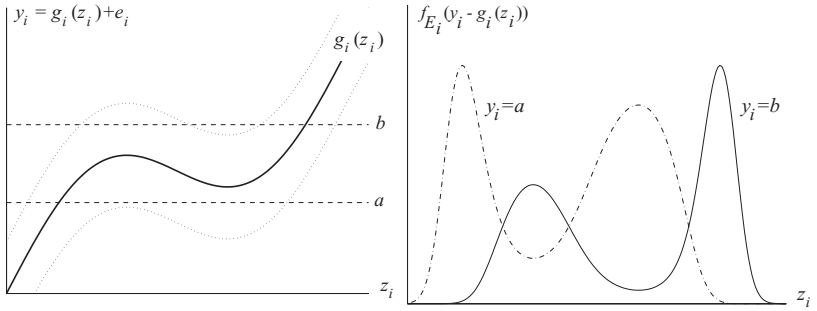


Figure 1.3: Left part shows the graph for the arbitrary functional $y_i = g_i(z_i) + e_i$ as in Fig. 1.2, along with the bounds of the $\pm 2\sigma$ probability interval for E_i assumed to be $N(0, \sigma^2)$. Right part shows the expression for $f_E(y_i - g_i(z_i))$ as a function of z_i when $y_i = a$ and $y_i = b$, with same scale for horizontal axis as on left part.

It is worth noting that the basic assumption for the model above is that the random vector $\mathbf{E}$ is considered as independent from the vector of interest $\mathbf{Z}$, or stated in other words the pdf of the errors $\mathbf{E}$ is the same whatever the choice for the conditioning $\mathbf{z}$ values. From a practical viewpoint, this assumption also implies that (i) $\mathbf{Y}$ can be modeled by a deterministic relation with $\mathbf{Z}$ and (ii) that this relation is accurately known, the remaining errors $\mathbf{E}$ being uncorrelated noise. Though this hypothesis proved to be convenient for deriving the previous results, it can be limiting as, e.g., (i) in the case of heteroskedasticity where the variance of errors depends on the $\mathbf{z}$ values or more generally (ii) under endogeneity problem between $\mathbf{Z}$ and $\mathbf{E}$ (i.e. when $\mathbf{Z}$ and $\mathbf{E}$ are correlated). One can remark however that arbitrary monotonic transformations can be applied on the Y_i's without affecting the generality of the results, as traditionally done in regression analysis.

1.3.3 Mixing soft and not so soft (hard) information

One can consider various situations with respect to the comparative amount of error attached with each observable Y_i. In some situations, the errors for a subset of these variables can be so high that they do not convey much information about the corresponding Z_i's, whereas in other situations these errors are so small that they can be neglected in the computations. We will consider both cases hereafter.

For the first case, let us write $\mathbf{Y} = (\mathbf{Y}'_a, \mathbf{Y}'_b)'$ where $\mathbf{Y}_a = (Y_0, \ldots, Y_m)'$ and $\mathbf{Y}_b = (Y_{m+1}, \ldots, Y_n)'$, where we assume that $\mathbf{E}_a \perp \mathbf{E}_b$. If the errors $\mathbf{E}_a$ tend to be very high compared to $\mathbf{g}(\mathbf{Z}_a)$, we can make the approximation $\mathbf{Y}_a = \mathbf{g}(\mathbf{Z}_a) + \mathbf{E}_a \simeq \mathbf{E}_a$ so that using Eq. (1.7) leads to

$$F(\mathbf{y}_a|\mathbf{z}_a) \simeq P(\mathbf{E}_a < \mathbf{y}_a|\mathbf{z}_a) = P(\mathbf{E}_a < \mathbf{y}_a) = F_{\mathbf{E}_a}(\mathbf{y}_a)$$

and we obtain $f(\mathbf{y}_a|\mathbf{z}_a) \simeq f_{\mathbf{E}_a}(\mathbf{y}_a)$. Plugging this result into Eq. (1.1) and using the fact that $(\mathbf{Y}_a \perp \mathbf{Y}_b)|\mathbf{z}$ because $\mathbf{E}_a \perp \mathbf{E}_b$, we obtain

$$f(\mathbf{z}|\mathbf{y}_a, \mathbf{y}_b) = \frac{1}{A} f(\mathbf{y}_a, \mathbf{y}_b|\mathbf{z}) f(\mathbf{z}) \simeq \frac{1}{A} f_{\mathbf{E}_a}(\mathbf{y}_a) f_{\mathbf{E}_b}(\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b)) f(\mathbf{z})$$

Integrating now this result over $z_1, \ldots, z_n$, we get

$$\begin{aligned} f(z_0|\mathbf{y}_a, \mathbf{y}_b) &\simeq \frac{1}{A} f_{\mathbf{E}_a}(\mathbf{y}_a) \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{E}_b}(\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b)) f(\mathbf{z}) \mathrm{d}z_1 \cdots \mathrm{d}z_n \\ &= \frac{1}{B} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{E}_b}(\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b)) f(\mathbf{z}_b) \mathrm{d}\mathbf{z}_b = f(z_0|\mathbf{y}_b) \end{aligned} \tag{1.10}$$

where $B = A/f_{\mathbf{E}_a}(\mathbf{y}_a)$ is the new normalization constant. As seen from Eq. (1.10), if the error $\mathbf{E}_a$ is assumed to be very high compared to $\mathbf{g}(\mathbf{Z}_a)$,

we can assume that $\mathbf{Y}_a$ does not bring any significant information on Z_0 and can be discarded for the prediction, as $f(z_0|\mathbf{y}_a, \mathbf{y}_b) \simeq f(z_0|\mathbf{y}_b)$.

For the second case, assume that the errors $\mathbf{E}_b$ tend to be very small compared to $\mathbf{g}(\mathbf{Z}_b)$, so that we can make the approximation $\mathbf{Y}_b = \mathbf{g}(\mathbf{Z}_b) + \mathbf{E}_b \simeq \mathbf{g}(\mathbf{Z}_b)$. Using Eq. (1.7) again leads to

$$F(\mathbf{y}_b|\mathbf{z}_b) \simeq P(\mathbf{g}(\mathbf{Z}_b) < \mathbf{y}_b|\mathbf{z}_b) = P(\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b) > 0) = \begin{cases} 0 \text{ if } \mathbf{y}_b < \mathbf{g}(\mathbf{z}_b) \\ \\ 1 \text{ if } \mathbf{y}_b \geq \mathbf{g}(\mathbf{z}_b) \end{cases}$$

so that $F(\mathbf{y}_b|\mathbf{z}_b)$ is the Heaviside step function, whose derivative is $f(\mathbf{y}_b|\mathbf{z}_b) = \delta(\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b))$, the Dirac delta function. Plugging this result into Eq. (1.1) and knowing that $(\mathbf{Y}_a \perp \mathbf{Y}_b)|\mathbf{z}$, we obtain

$$f(\mathbf{z}|\mathbf{y}_a, \mathbf{y}_b) \simeq \frac{1}{A} f_{\mathbf{E}_a}(\mathbf{y}_a - \mathbf{g}(\mathbf{z}_a))\delta(\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b))f(\mathbf{z})$$

and integrating now this result over $z_1, \ldots, z_n$, we get

$$f(z_0|\mathbf{y}_a, \mathbf{y}_b) \simeq \frac{1}{A} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{E}_a}(\mathbf{y}_a - \mathbf{g}(\mathbf{z}_a)) \left[\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \delta(\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b))f(\mathbf{z}) \mathrm{d}\mathbf{z}_b \right] \mathrm{d}\mathbf{z}_a \tag{1.11}$$

If the system of equations $\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b) = 0$ can be solved with respect to $\mathbf{z}_b$ so that $\mathbf{z}_b = \mathbf{g}^{-1}(\mathbf{y}_b)$ is the unique solution (this will happen, e.g., if for each Y_i we have $Y_i = g_i(Z_i)$ where the $g_i(.)$'s are monotonic so that $Z_i = g_i^{-1}(Y_i)$ $\forall i = m+1, \ldots, n$), we then have from the properties of the Dirac delta function

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \delta(\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b))f(\mathbf{z}) \mathrm{d}\mathbf{z}_b = J_{\mathbf{g}}(\mathbf{g}(\mathbf{z}_b))f(\mathbf{z}_a, \mathbf{z}_b)$$

with $\mathbf{z}_b = \mathbf{g}^{-1}(\mathbf{y}_b)$ known and where $J_{\mathbf{g}}(.)$ is the Jacobian of function $\mathbf{g}$. Plugging this result into Eq. (1.11) and writing $f(\mathbf{z}_a, \mathbf{z}_b) = f(\mathbf{z}_a|\mathbf{z}_b)f(\mathbf{z}_b)$,

$$\begin{aligned} f(z_0|\mathbf{y}_a, \mathbf{y}_b) &\simeq \frac{1}{C} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{E}_a}(\mathbf{y}_a - \mathbf{g}(\mathbf{z}_a))f(\mathbf{z}_a|\mathbf{z}_b) \mathrm{d}\mathbf{z}_a \\ &= f(z_0|\mathbf{y}_a, \mathbf{z}_b) \end{aligned} \tag{1.12}$$

where $C = AJ_{\mathbf{g}}(\mathbf{g}(\mathbf{z}_b))/f(\mathbf{z}_b)$. If the system of equations $\mathbf{y}_b - \mathbf{g}(\mathbf{z}_b) = 0$ admits several solutions $\mathbf{z}_{b,1}, \ldots, \mathbf{z}_{b,k}$, the event $\mathbf{Y}_b = \mathbf{y}_b$ is equivalent to $\mathbf{Z}_b = (\mathbf{Z}_b = \mathbf{z}_{b,1}) \cup \cdots \cup (\mathbf{Z}_b = \mathbf{z}_{b,k})$ where the elementary events $\mathbf{Z}_b = \mathbf{z}_{b,j}$ $(j = 1, \ldots, k)$ are mutually exclusive, so that

$$\begin{aligned} f(z_0|\mathbf{y}_a, \mathbf{y}_b) &= f(z_0|\mathbf{y}_a, (\mathbf{Z}_b = \mathbf{z}_{b,1}) \cup \cdots \cup (\mathbf{Z}_b = \mathbf{z}_{b,k})) \\ &= \frac{1}{\sum_{j=1}^{k} f(\mathbf{z}_{b,j})} \sum_{j=1}^{k} f(z_0|\mathbf{y}_a, \mathbf{z}_{b,j}))f(\mathbf{z}_{b,j}) \end{aligned}$$

From Eq. (1.12), one can see that if we assume the errors $\mathbf{E}_b$ to be negligible, this amounts to incorporating the knowledge brought by $\mathbf{Y}_a$ directly from the conditional pdf $f(\mathbf{z}_a|\mathbf{z}_b)$, with $f(\mathbf{z}_a|\mathbf{z}_b) = f(\mathbf{z}_a, \mathbf{z}_b)/f(\mathbf{z}_b)$, where the $\mathbf{z}_b$ values are computed from the observed $\mathbf{y}_b$ through the inverse relationship $\mathbf{z}_b = \mathbf{g}^{-1}(\mathbf{y}_b)$. If several solutions $\mathbf{z}_{b,j}$ exist, the procedure is repeated for each $\mathbf{z}_{b,j}$ and the results are weighted by the corresponding $f(\mathbf{z}_{b,j})$ values. It is worth noting that, by extension, Eq. (1.12) includes of course the more classical case where, for at least a subset of locations, values $\mathbf{z}_b$ are directly observed for the variable of interest. Finally, though both situations have been presented separately, they can of course be combined in an arbitrary way, and further simplifications are obtained by assuming the mutual independence for the errors.

1.3.4 Conditional independence and entropy

From eqs. (1.5) and (1.9), one can see that simple analytical results are obtained by assuming $E_0 \perp \cdots \perp E_n$, this corresponding to the very convenient conditional independence hypothesis $(Y_0 \perp \cdots \perp Y_n)|\mathbf{z}$, as it alleviates the need of inference for the joint pdf of errors in eqs. (1.2) and (1.8). Thus, when only a limited subset of Y_i is at hand, estimating this joint pdf may prove to be difficult. Clearly, substituting the joint pdf of errors by the product of the corresponding marginal pdf's will induce a loss of potentially valuable information. However, in the framework of information theory, one can prove that in absence of clear joint information, the best choice is precisely to assume this conditional independence. Indeed, from the maximum entropy principle, what is sought for is the joint pdf $(\mathbf{Y}, \mathbf{Z})$ for which the corresponding entropy $H(\mathbf{Y}, \mathbf{Z})$ is maximum, with

$$H(\mathbf{Y}, \mathbf{Z}) = -\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(\mathbf{y}, \mathbf{z}) \ln f(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{z}$$

We also have the relation $H(\mathbf{Y}, \mathbf{Z}) = H(\mathbf{Z}) + H(\mathbf{Y}|\mathbf{Z})$ where the marginal entropy $H(\mathbf{Z})$ is known as we have specified $f(\mathbf{z})$, with

$$H(\mathbf{Z}) = -\int_{\mathbb{R}^n} f(\mathbf{z}) \ln f(\mathbf{z}) \mathrm{d}\mathbf{z}$$

Since $H(\mathbf{Z})$ is known, the maximum of $H(\mathbf{Y}, \mathbf{Z})$ is reached when the conditional entropy $H(\mathbf{Y}|\mathbf{Z})$ is maximum, where

$$H(\mathbf{Y}|\mathbf{Z}) = -\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(\mathbf{y}, \mathbf{z}) \ln f(\mathbf{y}|\mathbf{z}) \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{z}$$

We also know that $H(\mathbf{Y}|\mathbf{Z}) \leq H(Y_0|\mathbf{Z}) + \cdots + H(Y_n|\mathbf{Z})$ (see e.g., Papoulis, 1991), where the marginal conditional entropies $H(Y_i|\mathbf{Z})$ are given by

$$H(Y_i|\mathbf{Z}) = -\int_{\mathbb{R}^n}\int_{\mathbb{R}} f(y_i|\mathbf{z}) \ln f(y_i|\mathbf{z}) \mathrm{d}y_i \mathrm{d}\mathbf{z} \qquad \forall i = 1, \ldots, n \tag{1.13}$$

If each Y_i depends on a unique Z_i so that $Y_i = g_i(Z_i) + e_i$, we can write that

$$f(y_i|\mathbf{z}) = f(y_i|z_i) = f_{E_i}(y_i - g_i(z_i)) \tag{1.14}$$

or stated in other words, knowing the complete set of realized values $\{z_0, \ldots, z_n\}$ does not bring any additional information on Y_i if we already know the single realized value z_i. As the error pdf's $f_{E_i}(e_i)$ are specified, one can evaluate $f_{E_i}(y_i - g_i(z_i))$ for any (y_i, z_i) so that plugging Eq. (1.14) into (1.13) shows that the various $H(Y_i|\mathbf{Z})$ are known too, with

$$H(Y_i|\mathbf{Z}) = H(Y_i|Z_i) = -\int_{\mathbb{R}}\int_{\mathbb{R}} f_{E_i}(y_i - g_i(z_i)) \ln f_{E_i}(y_i - g_i(z_i)) \mathrm{d}y_i \mathrm{d}z_i$$

As we finally know that

$$H(\mathbf{Y}|\mathbf{Z}) = H(Y_0|\mathbf{Z}) + \cdots + H(Y_n|\mathbf{Z}) \quad \Longleftrightarrow \quad (Y_0 \perp \cdots \perp Y_n)|\mathbf{z}$$

assuming the conditional independence for the Y_i's with respect to $\mathbf{z}$ is thus equivalent to maximizing $H(\mathbf{Y}, \mathbf{Z})$ when the pdf's $f(\mathbf{z})$ and $f_{E_i}(e_i)$ as well as the relationships $Y_i = g_i(Z_i) + E_i$ are specified.

It is also interesting to remark that when one only knows the expectation vectors $\boldsymbol{\mu}_{\mathbf{E}}$, $\boldsymbol{\mu}_{\mathbf{Z}}$ and the covariance matrices $\Sigma_{\mathbf{E}}$, $\Sigma_{\mathbf{Z}}$, the maximum for $H(\mathbf{Y}, \mathbf{Z})$ is reached when $\mathbf{Z} \sim N(\boldsymbol{\mu}_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$ and $\mathbf{E} \sim N(\boldsymbol{\mu}_{\mathbf{E}}, \Sigma_{\mathbf{E}})$ with $\mathbf{E} \perp \mathbf{Z}$, thus leading again to the conditional independence $(Y_0 \perp \cdots \perp Y_n)|\mathbf{Z}$. If only the variances $\sigma^2_{E_i}$ for the errors are known instead of $\Sigma_{\mathbf{E}}$, then the maximum entropy is obtained with $E_i \sim N(\mu_{E_i}, \sigma^2_{E_i})$ and $E_0 \perp \cdots \perp E_n$ (see Appendices for these proofs). In the case of linear relationships $Y_i = Z_i + E_i$ the conditional pdf $f(z_0|\mathbf{y})$ will always be Gaussian, with conditional mean and variance as given by simple kriging with measurement errors.

1.3.5 Dual notation for information

Up to now, all notations have been given in terms of the error pdf $f_{E_i}(.)$ along with the functionals $g_i(.)$. Both for practical and theoretical reasons, it is sometimes more convenient to reformulate these error pdf's

as functions of conditional pdf's on the Z_i's. Starting from Eq. (1.14) and using Bayes theorem again, we have

$$f_{E_i}(y_i - g_i(z_i)) = f(y_i|z_i) = \frac{f(z_i|y_i)}{f(z_i)} f(y_i) \tag{1.15}$$

Using Eq. (1.15) allows to express each $f_{E_i}(y_i - g_i(z_i))$ in Eq. (1.9) in terms of a conditional pdf on z_i, so that

$$f(z_0|\mathbf{y}) = \frac{1}{D} \frac{f(z_0|y_0)}{f(z_0)} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{z}) \prod_{i=1}^{n} \frac{f(z_i|y_i)}{f(z_i)} \mathrm{d}z_1 \cdots \mathrm{d}z_n \tag{1.16}$$

where $D = A/\prod_{i=0}^{n} f(y_i)$ is the new normalization constant, as all $f(y_i)$'s can be assimilated to constants in Eq. (1.16) due to the conditioning on the set $\{y_0, \ldots, y_n\}$. Eq. (1.16) is an alternative but equivalent way of incorporating the uncertainty that has the advantage of being more intuitive. Information conveyed by any Y_i is directly expressed through a conditional pdf for the corresponding Z_i. In this sense, Eq. (1.16) can be viewed as a complete recoding of the information brought by all the Y_i's, these information being now directly expressed through their influence on the Z_i variables.

From Eq. (1.16), it is also worth noting that the influence of a given y_i value on the final result is directly linked to how much $f(z_i|y_i)/f(z_i)$ is different from one as z_i varies. Stated in other words, the influence of knowing y_i can be measured by the discrepancies between the conditional pdf $f(z_i|y_i)$ and the marginal pdf $f(z_i)$. If a specific y_i is non informative with respect to Z_i, i.e. when $f(z_i|y_i)/f(z_i) = 1$ for some z_i, the corresponding factor will be filtered out from Eq. (1.16) (note that, for a fixed y_i value, the equality $f(z_i|y_i) = f(z_i)$ does not necessarily implies $Y_i \perp Z_i$, as there could be specific values of y_i for which this is true, whereas the equality is not verified in general). As a consequence, combining any non informative y_i values together with other informative values y_j $(i \neq j)$ will not cause any harm, as irrelevant information will be automatically filtered out during the prediction process. By the light of this last remark, it is thus useful to quantify the amount of information brought by any Y_i variable through some kind of informativeness index. This will be discussed in detail in Section 1.4.4.

1.4 Data fusion

Up to this point, the use of Eq. (1.9) is restricted to situations where a single measured value is available at each one of the $\mathbf{x}_0, \ldots, \mathbf{x}_n$ locations.

However, situations where multiple information are made available occurs commonly and need to be accounted for. This encompasses, e.g., the cases of (i) repeated measurements at location $\mathbf{x}_i$ of the same physical variable related to Z_i (with the same or with different measuring devices), (ii) different physical variables jointly measured at location $\mathbf{x}_i$, all of them being related in a different way to Z_i. As using more relevant information is expected to lead to more accurate predictions, it is thus of some concern to derive rules for fusing these information. Considerable efforts and literature have been devoted to this issue in spatial statistics. Most traditional approaches rely on a multivariate framework, where a joint distribution for the whole set of variables has to be inferred first and conditioned on observed values afterwards. However, due to the possibly highly different nature and properties of these variables (e.g., categorical vs continuous-valued variables) and potentially high number of variables, this can prove to be a very complex or even impossible task. We propose hereafter a different approach, that relies first on a full recoding of the information expressed as a unique conditional distribution on the variable of interest at every point where information is initially available, so that Eq. (1.16) can be used afterwards for combining them in order to get posterior distributions at arbitrary locations. A simple fusion rule is suggested and is discussed by the light of the Naive Bayes Fusion (NBF) rule.

1.4.1 Multiple collocated information

Let us assume without loss of generality that, for the last location $\mathbf{x}_n$, a set of m realized values $\{y_{n,1}, \ldots, y_{n,m}\}$ are available, all of them being related to the same z_n through the relations

$$Y_{n,j} = g_j(Z_n) + E_{n,j} \qquad \forall j = 1, \ldots, m \tag{1.17}$$

Let us denote $\mathbf{Y}_n = (Y_{n,1}, \ldots, Y_{n,m})'$ and $\mathbf{y}_a = (y_0, \ldots, y_{n-1})'$ (where the a subscript will always refer to the first $n-1$ elements of the corresponding vector), and let us write more simply Eq. (1.17) as $\mathbf{Y}_n = \mathbf{g}(Z_n) + \mathbf{E}_n$. What we are interested in is the conditional pdf $f(\mathbf{z}|\mathbf{y}_a, \mathbf{y}_n)$. Using Bayes' theorem,

$$f(\mathbf{z}|\mathbf{y}_a, \mathbf{y}_n) = \frac{f(\mathbf{y}_a, \mathbf{y}_n|\mathbf{z}) f(\mathbf{z})}{\displaystyle\int_{\mathbb{R}^n} f(\mathbf{y}_a, \mathbf{y}_n|\mathbf{z}) f(\mathbf{z}) \mathrm{d}\mathbf{z}} = \frac{1}{A} f(\mathbf{y}_a, \mathbf{y}_n|\mathbf{z}) f(\mathbf{z})$$

We will assume as before that $\mathbf{E}_a \perp \mathbf{Z}$ and $\mathbf{E}_n \perp \mathbf{Z}$, so that again

$$f(\mathbf{y}_a|\mathbf{z}) = f_{\mathbf{E}_a}(\mathbf{y}_a - \mathbf{g}(\mathbf{z}_a)) \quad ; \quad f(\mathbf{y}_n|\mathbf{z}) = f_{\mathbf{E}_n}(\mathbf{y}_n - \mathbf{g}(z_n))$$

Assuming moreover that $\mathbf{E}_a \perp \mathbf{E}_n$, we have the result

$$f(\mathbf{y}_a, \mathbf{y}_n|\mathbf{z}) = f_{\mathbf{E}_a}(\mathbf{y}_a - \mathbf{g}(\mathbf{z}_a)) f_{\mathbf{E}_n}(\mathbf{y}_n - \mathbf{g}(z_n))$$

so that the formula that combines all the information for prediction at location $\mathbf{x}_0$ becomes

$$f(z_0|\mathbf{y}_a, \mathbf{y}_n) = \frac{1}{A} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{z}) f_{\mathbf{E}_a}(\mathbf{y}_a - \mathbf{g}(\mathbf{z}_a)) f_{\mathbf{E}_n}(\mathbf{y}_n - \mathbf{g}(z_n)) \mathrm{d}z_1 \cdots \mathrm{d}z_n$$

Of course, this procedure can be easily generalized for any other set of locations where repeated measurements are made available. Further simplifications can be obtained by assuming the mutual independence for $\mathbf{E}_a$ and $\mathbf{E}_n$, so that along with the dual notation for information, where

$$f_{E_{n,j}}(y_{n,j} - g_j(z_n)) \propto \frac{f(z_n|y_{n,j})}{f(z_n)} \tag{1.18}$$

we get similarly to Eq. (1.16) the simpler result

$$\begin{aligned} f(z_0|\mathbf{y}) &\propto \frac{f(z_0|y_0)}{f(z_0)} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{z}) f(z_n)^{-m} \prod_{i=1}^{n-1} \frac{f(z_i|y_i)}{f(z_i)} \prod_{j=1}^{m} f(z_n|y_{n,j}) \mathrm{d}z_1 \cdots \mathrm{d}z_n \\ &\propto \frac{f(z_0|y_0)}{f(z_0)} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{z}) \phi(z_n|\mathbf{y}_n) \prod_{i=1}^{n-1} \frac{f(z_i|y_i)}{f(z_i)} \mathrm{d}z_1 \cdots \mathrm{d}z_n \end{aligned} \tag{1.19}$$

where $\phi(z_n|\mathbf{y}_n)$ denotes the posterior pdf resulting from the fusion operation for collocated information, i.e.

$$\phi(z_n|\mathbf{y}_n) \propto f(z_n)^{-m} \prod_{j=1}^{m} f(z_n|y_{n,j})$$

Through this operation, the information brought separately by each observed $y_{n,j}$ with respect to a same Z_n are thus fused into a single pdf $\phi(z_n|\mathbf{y}_n)$.

As a last remark, it is worth noting that in various spatial applications like, e.g., remote sensing, image analysis or cartography, a common situation that occurs is the need for fusing collocated information at the prediction location itself, discarding any other possible information at other locations (so that the spatial prediction part of the problem does not appear). If we define $\mathbf{Y}_0 = (Y_{0,1}, \ldots, Y_{0,m})'$ with $Y_{0,j} = Z_0 + E_{0,j}$ $\forall j = 1, \ldots, m$ and reasoning along the same lines as above by assuming mutual independence for the $E_{0,j}$'s, it is then easy to see that the posterior pdf simply becomes

$$f(z_0|\mathbf{y}_0) \propto \frac{1}{(f(z_0))^{m-1}} \prod_{j=1}^{m} f(z_0|y_{0,j}) \propto \phi(z_0|\mathbf{y}_0) \tag{1.20}$$

which corresponds to the so-called naive Bayes' (or Idiot's Bayes's) fusion rule of the individual posterior predictors as given by the $f(z_0|y_{0,j})$'s.

Criticisms about the use of a NBF rule can be raised on the ground that it implicitly relies on a conditional independence hypothesis of the variables. However, in a classification context, Kuncheva (2004) emphasized that NBF is experimentally observed to be surprisingly accurate and efficient, where the surprise comes from the fact that independence assumption is seldom true. Indeed, naive Bayes can often outperform more sophisticated classification methods (Altinçay, 2005; Lewis, 1998). To our opinion, by the light of the previously discussed relation between conditional independence and entropy maximization, this can be at least partially explained by the fact that NBF is precisely the choice that maximizes entropy if no additional information is incorporated about the joint dependence of these variables, so it will yield the posterior pdf which is the most likely to be observed (see Jaynes, 2003, for an enlightening discussion about this principle). Note however that if a joint pdf $f(z_n, \mathbf{y}_n)$ can be reasonably well inferred from data at hand (thus alleviating the need of a conditional independence hypothesis), the corresponding conditional pdf $f(z_n|\mathbf{y}_n)$ can be used into Eq. (1.19) without any problem, instead of relying on the simpler NBF rule-based pdf $\phi(z_n|\mathbf{y}_n)$.

1.4.2 Weighting information for confidence

Let us assume in Eq. (1.19) that among the $f(z_n|y_{n,j})$'s, there is a subset $J \subset \{1, \ldots, m\}$ of them that are identical, so that we can write $f(z_n|y_{n,j}) = f(z_n|y_{n,J})\ \forall j \in J$, i.e., the same information on Z_n is brought by each $y_{n,j}$ when $j \in J$. We have now

$$\begin{aligned}\phi(z_n|\mathbf{y}_n) \propto\ & f(z_n)^{-m} f(z_n|y_{n,J})^{m_J} \prod_{j \notin J} f(z_n|y_{n,j}) \\ =\ & f(z_n)^{-m} f(z_n|y_{n,J})^{w_J m} \prod_{j \notin J} f(z_n|y_{n,j})^{w_j m}\end{aligned}$$

where m_J is the cardinality of J, with weights $w_J = m_J/m$ and $w_j = 1/m\ \forall j \notin J$ so that their sum is equal to 1. This is equivalent to a relative weighting of the pdf's. By generalization, it also suggests that pdf's can be fused by accounting for the confidence one may have in their respective accuracy (e.g., when they are inferred from data of varying quality or quantity, or when they come from experts that do not get the

same credit), so that we could write

$$\phi(z_n|\mathbf{y}_n) \;\propto\; f(z_n)^{-m}\prod_{j=1}^{m} f(z_n|y_{n,j})^{w_j m} \qquad \text{with } \sum_j w_j = 1 \qquad (1.21)$$

Eq. (1.21) can be interpreted easily by remarking that if $w_j > 1/m$ (or $w_j < 1/m$), the corresponding pdf will count for more (or for less) than a single pdf among the set of m fused ones, as $w_j m > 1$ (or as $w_j m < 1$), whereas Eq. (1.19) is the case where the same credit is given to each pdf.

1.4.3 The Gaussian case

For obvious inferential reasons, a quite common hypothesis is to assume that all $f(z_n|y_{n,j})$'s are the pdf's of $N(\mu_j, \sigma_j^2)$ variables, as inference for these pdf's only requires the estimation of the conditional expectations $\mathbb{E}[Z_n|y_{n,j}] = \mu_j$ and conditional variances $\mathbb{V}ar[Z_n|y_{n,j}] = \sigma_j^2$ with $j = 1, \ldots, m$. It is then easy to prove that

$$\prod_{j=1}^{m} f(z_n|y_{n,j}) \;\propto\; \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\Big(\frac{z_n-\mu}{\sigma}\Big)^2\right)$$

so that the product of the $f(z_n|y_{n,j})$'s is proportional to the density of a Gaussian distribution with expectation μ and variance σ^2 that are given by

$$\mu = \sum_{j=1}^{m} \frac{1/\sigma_j^2}{\sum_{k=1}^{m}(1/\sigma_k^2)}\mu_j \qquad \sigma^2 = 1\Big/\sum_{j=1}^{m}\frac{1}{\sigma_j^2}$$

(see e.g. Papoulis, 1991, pg 187 for obtaining this result based on a least squares criterion, and Duc *et al.*, 1997, for the use of this result in a somewhat different context). It is worth noting that μ thus corresponds to a weighting of the μ_j's, where the weights are proportional to the inverse of the corresponding variances. This is consistent with the intuition, as the information brought by any particular $f(z_n|y_{n,j})$ decreases as its corresponding variance σ_j^2 increases. Moreover, one can also see that, as we have $\sigma_{m+1}^2 \geq 0\ \forall m = 1, 2, \ldots$, we have the inequality

$$\sum_{j=1}^{m+1}\sigma_j^2 \geq \sum_{j=1}^{m}\sigma_j^2 \quad\Longleftrightarrow\quad 1\Big/\sum_{j=1}^{m+1}\frac{1}{\sigma_j^2} \;\leq\; 1\Big/\sum_{j=1}^{m}\frac{1}{\sigma_j^2}$$

thus showing that σ^2 can only decrease as the number of $f(z_n|y_{n,j})$'s increases. It is worth noting that the final fused pdf is not Gaussian in general, as it also depends on the *a priori* $f(z_n)$ in Eq. (1.19), and

the influence of any specific $f(z_n|y_{n,j})$ on the final result will still depend on its distance from this *a priori* pdf. As an illustration, Fig. 1.4 shows the result of the fusion for three Gaussian pdf's. The left graph presents the Gaussian pdf's $f(z_n|y_{n,j})$ $(j = 1, 2, 3)$ to be fused as well as the corresponding product $\prod_{j=1}^{3} f(z_n|y_{n,j})$ (up to the multiplicative normalization constant). The right graph shows the fusion pdf $\phi(z_n|\mathbf{y}_n)$ for two different *a priori* pdf's. One can easily see the additional effect of this pdf on the final result.

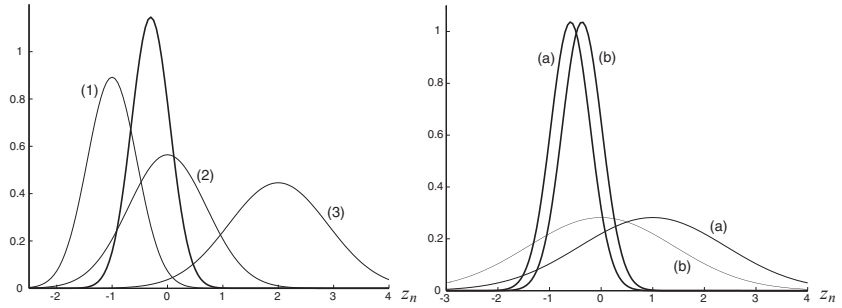


Figure 1.4: Left graph shows Gaussian pdf's $f(z_n|y_{n,j})$ labeled as (1),(2) and (3) with parameters $\mu_1 = -1, \mu_2 = 0, \mu_3 = 2, \sigma_1^2 = 0.2, \sigma_2^2 = 0.5, \sigma_3^2 = 0.8$ along with the normalized product of these 3 distributions, which is Gaussian with $\mu = -0.30$ and $\sigma^2 = 0.12$ (bold curve). Right graph shows the resulting fusion pdf's $\phi(z_n|\mathbf{y}_n)$ (bold curves) for two different choice of the *a priori* pdf's $f(z_n)$ (plain curves) that are respectively (a) $N(1, 2)$ and (b) $N(0, 2)$.

It is worth noting here that Gaussian distributions are not at all required within the BDF framework. However, as direct consequence of the above expressions, a significant increase in the computation speed can be seen when using Gaussian distribution as analytical results exist in those cases. It thus might be interesting to apply the BDF method on transformed variables so that Gaussian distributions can be assumed and analytical expressions can be used for the mean and variance of the resulting fused distribution and then to back-transformed this distribution to the original variable space in order to get the final fused distribution.

1.4.4 Informativeness index

As seen from Eq. (1.18), the amount of information brought by a specific $y_{n,j}$ can be quantified by measuring of how much the ratio $f(z_n|y_{n,j})/f(z_n)$ is different from one. As z_n may vary, thus leading

to different ratio values, one could be interested by how much, on the average, this ratio differs from one. Let us consider the natural logarithm of the ratio, so that its value is equal to 0 when the ratio is equal to one (as the logarithm is a monotonic function, the ordering of the ratios computed for $y_{n,1}, \ldots, y_{n,m}$ for the same z_n value will not be modified). The average discrepancy between $f(z_n|y_{n,j})$ and $f(z_n)$ with respect to Z_n and for a fixed $y_{n,j}$ value is thus given by

$$\mathbb{E}\left[\ln \frac{f(Z_n|y_{n,j})}{f(Z_n)}\right] = \int_{\mathbb{R}} f(z_n|y_{n,j}) \ln \frac{f(z_n|y_{n,j})}{f(z_n)} \mathrm{d}z_n \qquad \forall j = 1, \ldots, m \tag{1.22}$$

where this equation is by definition the Kullback-Leibler (KL) distance or relative entropy between $f(z_n)$ and $f(z_n|y_{n,j})$ when the $y_{n,j}$ value is observed (Kullback and Leibler, 1951). In practice, Eq. (1.22) can only be computed if $y_{n,j}$ is known. It is thus more convenient to define the expected KL distance with respect to $Y_{n,j}$, that can be used prior to any observation for $Y_{n,j}$. It is well known that this is by definition the mutual information $I(Z_n, Y_{n,j})$ (see e.g. Cover and Joy, 2006), with

$$\begin{aligned} I(Y_{n,j}, Z_n) &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(y_{n,j}, z_n) \ln \frac{f(y_{n,j}, z_n)}{f(y_{n,j}) f(z_n)} \mathrm{d}y_{n,j} \mathrm{d}z_n \\ &= \int_{\mathbb{R}} f(y_{n,j}) \int_{\mathbb{R}} f(z_n|y_{n,j}) \ln \frac{f(z_n|y_{n,j})}{f(z_n)} \mathrm{d}z_n \mathrm{d}y_{n,j} \\ &= \mathbb{E}\left[\int_{\mathbb{R}} f(z_n|Y_{n,j}) \ln \frac{f(z_n|Y_{n,j})}{f(z_n)} \mathrm{d}z_n\right] \end{aligned} \tag{1.23}$$

Moreover, we also have the relation

$$I(Y_{n,j}, Z_n) = H(Z_n) - H(Z_n|Y_{n,j}) \tag{1.24}$$

where $H(Z_n)$ is the entropy for Z_n and where $H(Z_n|Y_{n,j})$ is the conditional entropy, that are defined as

$$\begin{aligned} H(Z_n) &= -\int_{\mathbb{R}} f(z_n) \ln f(z_n) \mathrm{d}z_n \\ H(Z_n|Y_{n,j}) &= -\int_{\mathbb{R}} f(y_{n,j}, z_n) \ln f(z_n|y_{n,j}) \mathrm{d}y_{n,j} \mathrm{d}z_n \end{aligned}$$

Comparing Eqs. (1.23) and (1.24) thus shows that measuring how much the logarithm of the ratio $f(z_n|y_{n,j})/f(z_n)$ differs from zero on the average is equivalent to measuring the information that $Y_{n,j}$ conveys on Z_n. As the various mutual information $I(Y_{n,j}, Z_n)$ $(j = 1, \ldots, m)$ can

be computed with respect to the same Z_n, they can be compared and sorted, thus allowing the selection of the "best" $Y_{n,j}$ variable, i.e., the variable for which the corresponding $I(Y_{n,j}, Z_n)$ is maximum over the set $\{I(Y_{n,1}, Z_n), \ldots, I(Y_{n,m}, Z_n)\}$. For example, in presence of abundant potential information (i.e., a high m value), for practical reasons one may consider using only a limited subset of $m' < m$ useful variables to be included.

If we denote $Y_{n,[1]}, \ldots, Y_{n,[m]}$ as the $Y_{n,j}$ variables sorted in decreasing order of mutual information values so that $I(Y_{n,[1]}, Z_n) \geq I(Y_{n,[2]}, Z_n) > \cdots \geq I(Y_{n,[m]}, Z_n)$, a naive solution would be to use the subset of these m' first variables. Unfortunately, such a simple approach does not take into account the possible redundancy of information between the $Y_{n,[j]}$ variables (e.g., taken separately, $Y_{n,[1]}$ and $Y_{n,[2]}$ are the two most informative variables, but it is possible that $Y_{n,[2]}$ does not convey much more additional information on Z_n compared to what is already included when only using $Y_{n,[1]}$). A more satisfactory solution is obtained using the definition of the mutual information $I(\mathbf{Y}_n, Z_n,)$ for the whole set of variables $Y_{n,1}, \ldots, Y_{n,m}$ (see e.g. Papoulis, 1991, pg 562), with

$$I(\mathbf{Y}_n, Z_n) = \int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{y}_n, z_n)) \ln \frac{f(\mathbf{y}_n, z_n)}{f(\mathbf{y}_n) f(z_n)} \mathrm{d}y_{n,1} \cdots \mathrm{d}y_{n,m} \mathrm{d}z_n \quad (1.25)$$

Using the fact that

$$f(\mathbf{y}_n, z_n) = f(y_{n,1}, z_n) f(y_{n,2}, z_n | y_{n,1}) \cdots f(y_{n,m}, z_n | \{y_{n,1}, \ldots, y_{n,m-1}\})$$

$$f(\mathbf{y}_n) = f(y_{n,1}) f(y_{n,2} | y_{n,1}) \cdots f(y_{n,m} | \{y_{n,1}, \ldots, y_{n,m-1}\})$$

and substituting these results into Eq. (1.25), it is easy to show after some elementary manipulations that

$$\begin{aligned} I(\mathbf{Y}_n, Z_n) = {} & I(Y_{n,1}, Z_n) + I(Y_{n,2}, Z_n | Y_{n,1}) + \cdots \\ & + I(Y_{n,m}, Z_n | \{Y_{n,1}, \ldots, Y_{n,m-1}\}) \end{aligned} \quad (1.26)$$

where the various (conditional) mutual information in the right-hand side of Eq. (1.26) are defined as

$$\begin{aligned} I(Y_{n,1}, Z_n) &= H(Z_n) - H(Z_n | Y_{n,1}) \\ I(Y_{n,2}, Z_n | Y_{n,1}) &= H(Z_n | Y_{n,1}) - H(Z_n | \{Y_{n,1}, Y_{n,2}\}) \\ &\vdots \\ I(Y_{n,m}, Z_n | \{Y_{n,1}, \ldots, Y_{n,m-1}\}) &= H(Z_n | \{Y_{n,1}, \ldots, Y_{n,m-1}\}) - H(Z_n | \mathbf{Y}_n) \end{aligned}$$

From a strict point of view, selecting the best (optimal) subset $\mathbf{Y}_{opt} \subset \mathbf{Y}_n$ of m' variables would thus correspond to look for the m' variables such that, with respect to all other possible subsets of same size, the corresponding mutual information $I(\mathbf{Y}_{opt}, Z_n)$ is maximum. (see e.g. Fassinut-Mombot and Choquel, 2004, for a similar idea). Unfortunately, this is an optimization problem that involves combinatorics, with number of possible combinations given by $C_m^{m'}$, the binomial coefficient. However, mutual information decomposition as given by Eq. (1.26) suggests a simpler forward selection procedure for selecting sequentially this subset :

1. select $Y_{n,[1]}$ so that $I(Y_{n,[1]}, Z_n) \geq I(Y_{n,[j]}, Z_n)\ \forall j$

2. select $Y_{n,[2]}$ so that $I(Y_{n,[2]}, Z_n|Y_{n,[1]}) \geq I(Y_{n,[j]}, Z_n|Y_{n,[1]})\ \forall j \neq 1$,

3. repeat this for $i = 3, \ldots, m'$, where at the ith step the selected $Y_{n,[i]}$ is such that $\forall j \neq 1, \ldots, i-1$
$I(Y_{n,[i]}, Z_n|\{Y_{n,[1]}, \ldots, Y_{n,[i-1]}\}) \geq I(Y_{n,[j]}, Z_n|\{Y_{n,[1]}, \ldots, Y_{n,[i-1]}\})$.

this procedure being a direct analogy with the forward procedure in multiple regression, where the aim is the sequential selection of the most explanatory subset of variables to be included into a regression model (Neter *et al.*, 1996).

The proposed procedure enables us thus to automatically select which secondary information sources should first be accounted for when the number of those sources is large. However, it is worth noting that this procedure was not tested here in this thesis as the number of secondary variables was too small to yield numerical limitations for the BDF method.

1.5 A synthetic case study

In order to illustrate some of the general principle of the methodology, a simple synthetic case study will be presented. Clearly, due to space limitations, only some of the concepts presented in this chapter can be illustrated here. We will thus mainly focus on a situation where what is sought for is to fuse information sources that do not easily fit into a joint classical multivariate RF framework.

Let us assume that we are interested in the spatial prediction of a continuous random fields $\mathfrak{Z}$, with a typical realization as given by Fig. 1.5a. This corresponds to a smooth RF realization obtained by

taking the absolute value of a zero-mean unit-variance Gaussian RF $\mathfrak{Y}$ with Gaussian covariance function, i.e. $Z_i = |Y_i| \ \forall \mathbf{x}_i \in D$ where $Y_i \sim N(0,1)$. The aim of smoothness and absolute value is to create a network-like appearance on the map, the location of this network being part of the available information. Let us define this network as a binary RF $\mathfrak{N}$ for which $(N_i = 1) \equiv (|Y_i| < y_{0.55})$, with $y_{0.55}$ the 0.55 quantile of a $N(0,1)$ distribution, so that $P(N_i = 1) = 0.1$ and $P(N_i = 0) = 0.9$ (see Fig. 1.5b). Aside from this RF, we will consider another independent realization of $\mathfrak{Y}$ that will be used to define the binary RF $\mathfrak{C}$ where $(C_i = 1) \equiv (Y_i > 0)$ so that $P(C_i = 0) = P(C_i = 1) = 0.5$ (see Fig. 1.5c). This information is irrelevant for predicting values of Fig. 1.5a as $C_i \perp Z_j$ $\forall i, j$, but we will force its use in order to illustrate the automatic filtering properties of the method.

In order to rebuild Fig. 1.5a at best, we will consider that the only available information are a set of sparsely sampled values $\mathbf{Z} = (Z_1, \ldots, Z_{75})'$, along with the spatially exhaustive realization for $\mathfrak{N}$ and $\mathfrak{C}$. Clearly, this whole set of information does not lead to a nice and straightforward formulation in a traditional continuous multivariate RF approach if the aim is to predict Z_0 at unsampled locations $\mathbf{x}_0$, as (i) it requires to combine both continuous and categorical RF's, (ii) the information brought by the realized event $n_0 = 0$ or 1 at prediction location is considerably poorer than the information brought by the distance d_{0i} between $\mathbf{x}_0$ and the closest location $\mathbf{x}_i$ for which $n_i = 1$, (iii) unfortunately, this set of new random variables D_{0i} does not correspond to a second-order stationary RF.

Instead of relying on a rather complex multivariate approach, the idea is thus to recode the set of indirect information about $\mathfrak{Z}$ in terms of conditional distributions $f(z_0|d_{0i})$ and $f(z_0|c_0)$. Fig. 1.5c shows the estimated regression $\mathbb{E}[Z_0|d_{0i}]$ along with confidence bounds from the estimated $\mathbb{V}ar[Z_0|d_{0i}]$, where it will be assumed that $Z_0|d_{0i}$ is Gaussian distributed. Assuming (wrongly) that $(Z_0, \mathbf{Z})'$ is Gaussian distributed too, the conditional $Z_0|\mathbf{z}$ is also Gaussian. Finally, assuming again normality, the corresponding $f(z_0|c_0)$ are estimated too when $c_0 = 1$ or when $c_0 = 0$ (see Fig. 1.7). Using now a NBF rule at the prediction location itself, we obtain according to Eq. (1.20) the result

$$f(z_0|\mathbf{z}, d_{0i}, c_0) \propto \frac{f(z_0|\mathbf{z}) f(z_0|d_{0i}) f(z_0|c_0)}{f^2(z_0)} \tag{1.27}$$

where all $f(z_0|.)$'s on the right part of equality have been assumed Gaussian and where it is expected that $f(z_0|\mathbf{z}, d_{0i}, c_0) = f(z_0|\mathbf{z}, d_{0i})$ as knowing C_0 is irrelevant, and thus $f(z_0|c_0) = f(z_0)$ either when $c_0 = 1$ or when $c_0 = 0$.

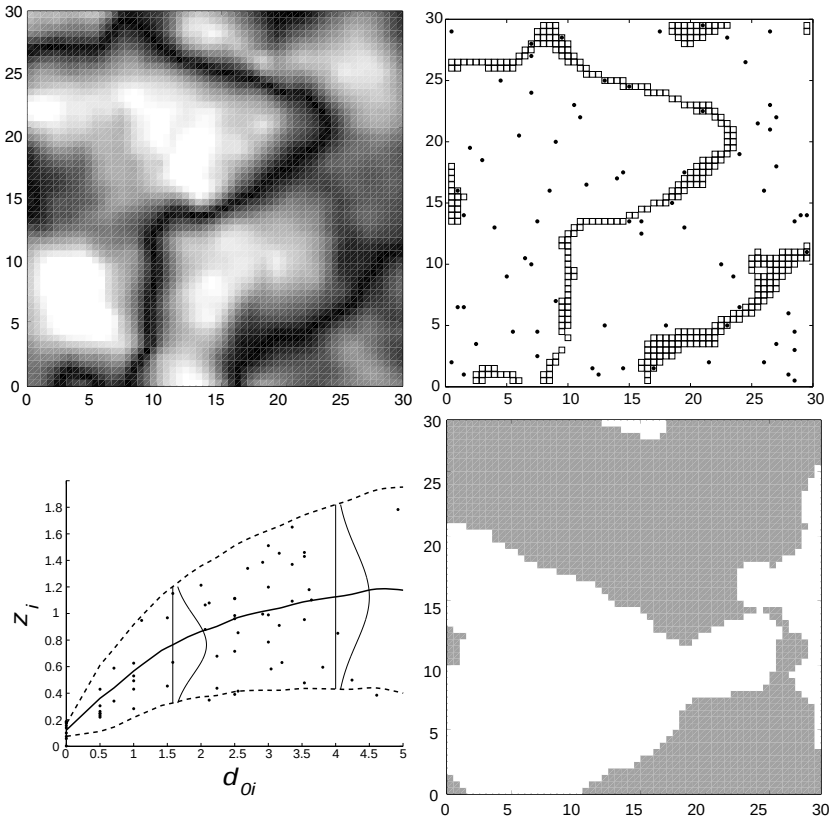


Figure 1.5: Synthetic case study. Part (a) shows a realization on a 60×60 grid of the $\mathfrak{Z}$ RF, where lower values appearing in dark have the meaning of a network. Part (b) shows locations for 75 sampled values z_i (black dots) as a subset of the realization, along with locations where the network binary RF $\mathfrak{N}$ takes value $n_j = 1$ (white squares). Part (c) shows the estimated regression $\mathbb{E}[Z_i|d_{ij}]$ (plain line) estimated from the (Z_i, d_{ij})'s (black dots) where d_{ij} is the Euclidean distance to the closest location where $n_j = 1$, along with the symmetric 0.95 confidence bounds (dashed lines) assuming normality. Part (d) shows the realization of the $\mathfrak{C}$ RF ($c_i = 1$ in white and $c_i = 0$ in gray).

The resulting map of predicted values $\mathbb{E}[Z_0|\mathbf{z}, d_{0i}]$ obtained using Eq. (1.27) is shown in Fig. 1.6b, along with the predicted map $\mathbb{E}[Z_0|\mathbf{z}]$ obtained by neglecting the network information. Clearly, accounting for the network yields significantly better results. A comparison between the maps for $\mathbb{E}[Z_0|\mathbf{z}, d_{0i}, c_0]$ and $\mathbb{E}[Z_0|\mathbf{z}, d_{0i}]$ (not shown for the sake of brevity) also confirms that making use or not of the C_0 variable does not have any significant impact on the results, as expected.

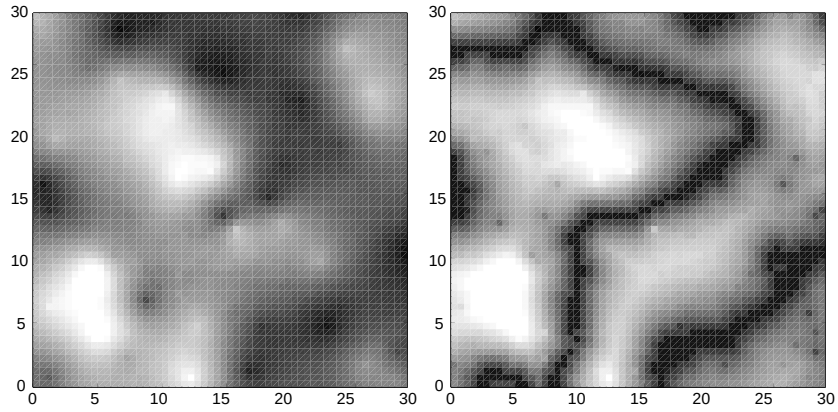


Figure 1.6: Prediction results. Part (a) shows the map of the $\mathbb{E}[Z_i|\mathbf{z}]$'s on the same 60×60 grid as for realization, thus neglecting network information (simple kriging). Part (b) shows the map of the $\mathbb{E}[Z_i|\mathbf{z}, d_{ij}]$'s as obtained from the NBF rule neglecting information about C_0. Map of the $\mathbb{E}[Z_i|\mathbf{z}, d_{ij}, c_0]$'s (not shown here) is similar.

More specifically, Fig. 1.7 illustrates the results obtained for two different prediction locations $\mathbf{x}_0$ that are respectively close or remote from the network. As we expect that $\lim_{d_{0i}\to\infty} f(z_0|d_{0i}) = f(z_0)$, the network does not have significant influence on the result when $\mathbf{x}_0$ is far from it, so the result is solely influenced by surrounding measurements $\mathbf{z}$, with $f(z_0|\mathbf{z}, d_{0i}) \simeq f(z_0|\mathbf{z})$. On the contrary, for $\mathbf{x}_0$ close to or inside the network, $f(z_0|d_{0i})$ may become much more informative than $f(z_0|\mathbf{z})$, especially if there are no close $\mathbf{z}$ surrounding measurements, so that $f(z_0|\mathbf{z}, d_{0i}) \simeq f(z_0|d_{0i})$, thus yielding a very good restitution of the network. From these figures, one can also see that, for the estimated distributions $f(z_0|c_0)$, we both have $f(z_0|c_0) \simeq f(z_0)$ when $c_0 = 1$ and when $c_0 = 0$, thus showing that these distributions will have no sensible effect on the final result for the fusion, i.e. $f(z_0|\mathbf{z}, d_{0i}, c_0) \simeq f(z_0|\mathbf{z}, d_{0i})$.

Finally, it is interesting to compare the semi-variograms of the errors of prediction made when using each information source separately

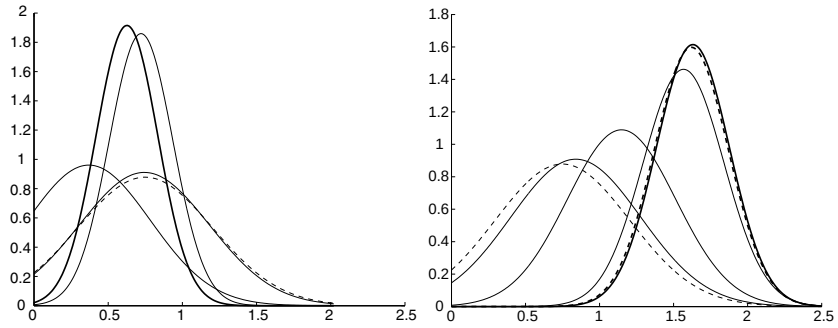


Figure 1.7: Influence of the information in the fusion process. Part (a) represents the pdf's to be fused along with the result using the NBF rule for a prediction location close to the network. Part (b) is the same for a remote location. Symbols for the curves are dashed line for *a priori* pdf $f(z_0)$, plain lines for (1) $f(z_0|d_{0i})$, (2) $f(z_0|\mathbf{z})$ and (3) $f(z_0|c_0)$, with $c_0 = 0$ in Part (a) and $c_0 = 1$ in Part (b). Bold lines correspond to $f(z_0|\mathbf{z}, d_{0i}, c_0)$ as obtained from NBF rule. Additionally, dashed bold lines correspond to $f(z_0|\mathbf{z}, d_{0i})$, i.e. neglecting the useless information about C_0. The two last curves cannot be distinguished from each other on the left graph.

or when using the BDF prediction. Its is clear from the results that the BDF prediction outperformed the other predictions as the semi-variogram variance is significantly lower for the BDF method (see Fig. 1.8).

1.6 Discussion and conclusions

In this chapter, starting from a simple and classical measurement errors context, it has been shown that a general formulation can be obtained for the prediction of a spatial RF at unsampled locations using various sources of direct and/or indirect measurements. Several well-known situations that arise as limit cases of the method have been presented, and connections between the convenient conditional independence hypothesis and maximum entropy properties have been emphasized too. Finally, with the help of information theory, it has been shown that a useful informativeness index can be derived with the aim of selecting the "best" subset of information to be used.

Clearly, the major interest of the proposed procedure is to allow the user to incorporate in a flexible way a potentially high number of secondary information sources that may exhibit high diversity, by avoiding at the same time major complications with respect to modeling hypothe-

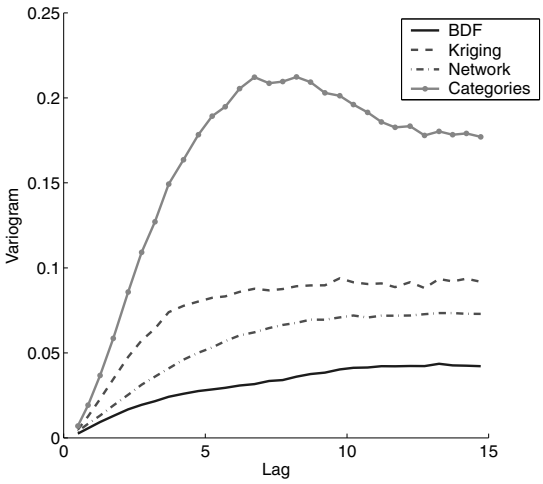


Figure 1.8: Semi-variograms of the errors for the three predictions separately built from each of the information sources and for the prediction built from the fusion of these sources within the BDF framework.

ses and computing time requirements. As all secondary information can be recoded (and fused if needed) as conditional distributions with respect to the variable of interest, this leads to a same generic formulation for prediction, that relies on subsequent multivariate integrations over a joint distribution.

Of course, this chapter should not be viewed and is by no way a prosecution against the use of sound multivariate methods, that aim at fully accounting for the joint spatial dependence of RF's. However, it is stressed that the need for such models is not as critical as it may appear at a first sight. Moreover, it is not uncommon situation that multivariate models may bring more problems than solutions. Indeed, these models typically rely on more or less demanding hypotheses that may prove to be impossible to fulfill in a reasonable way with data at hand, especially when number and diversity of information sources that need to be accounted for is high. As a consequence, the user is commonly facing non obvious choices like, e.g. (i) fooling the model by forcing it to use data that do not naturally match the required hypotheses, with potentially serious (and often poorly assessed) effects on the final results, (ii) using dimensionality reduction techniques or data transformation for improving the adequacy between data properties and model hypotheses, or (iii) discarding potentially valuable information based on the single fact that they do not nicely fit into the model framework. Even for situ-

ations were the need for such choices can be avoided, in most of the cases difficult practical issues remains to be addressed, as computing burden and modeling complexity is typically expected to quickly increases with number of information sources.

With respect to these aspects, the proposed methodology has nice features, as all efforts are concentrated on the initial optimal recoding of the information, thus cutting down the need for subsequent heavy computations and complex modeling issues. Clearly, the performance of the method is still to be evaluated and compared to more complex modeling approaches in specific circumstances. Among others, further work should be devoted to quantifying the impact on the final results of the information loss which is induced by neglecting the spatial correlation between RF's . However, based on the simple synthetic case study that has been presented, it can already be seen as a quite feasible and realistic alternative.

Chapter 2

Data fusion in a spatial multivariate framework : Trading off hypotheses against information

The previous chapter presented the Bayesian data fusion theory in detail and specified the necessary hypotheses. In this chapter, we emphasize the importance of the set of hypotheses when dealing with several information sources and we quantify the loss of precision generated when the hypotheses are not fulfilled.

2.1 Outline

Due to the exponential growth in the amount and diversity of data that one may expect to provide greater modeling and predictions opportunities, there is a real need for methods that aim at reconciling them inside a flexible and sound theoretical framework. In a geostatistical prediction context, beside more or less straighforward variations around univariate kriging (e.g., kriging with external drift), the most classical method (i.e. cokriging) is based on a multivariate random field approach of the problem, at the price of strong modeling hypotheses. However, there are expected practical situations where these hypotheses may be hard to fulfill or do not make sense from a modeling viewpoint.

This chapter proposes an alternative way of using secondary information for spatial prediction. Based on a data fusion perspective, a general theoretical procedure is proposed. Simple cokriging and Bayesian data fusion are compared both from theoretical and practical viewpoints.

Theoretical differences are first emphasized based on the corresponding modeling hypotheses. A case study based on synthetic data subsequently allows to compare both methods from a practical viewpoint. It is shown that, in spite of some simplifying hypotheses required by data fusion, the method provides quite comparable performances in situations where simple cokriging is expected to be the best possible predictor. Moreover, it offers a much greater flexibility and opens new avenues for incorporating a wide panel of very different and possibly numerous secondary information that, by nature, would not easily fit into a multivariate random field framework, as required by cokriging.

As the classical (co)kriging predictor relies on the knowledge of first and second-order moments for a given set of random fields (RF's), we will assume here that first and second-order stationarities can be assumed for all variables, and that the corresponding functions (i.e. the mean functions and the (cross-)covariance functions) can reasonably be inferred from the data. Without loss of generality and for the sake of simplicity, we will restrict here the discussion to variables with known mean functions so that simple cokriging can be used, though the methodology includes of course the case of non stationary mean and intrinsic stationarity as well (see e.g. Christakos, 1992).

In a first section, a short recall of simple cokriging (SCoK) formulation is presented. The hypotheses involved in this model are emphasized and modeling issues are pointed out with respect to both theoretical and practical viewpoints. Subsequently, the Bayesian Data Fusion (BDF) methodology is presented. For the sake of conciseness and without loss of generality, presentation will be restricted here to the particular case of bivariate Gaussian RFs. Pros and cons of the method are discussed too, and a synthetic case study is presented. The corresponding results indicate that BDF is an efficient alternative in the context of spatial prediction that need to account for additional secondary information which may not fit very well into a multivariate stationary RF framework[1].

[1]This chapter is an adaptation of the proceedings of a presentation given at geoENV2006 (Fasbender and Bogaert, 2006).

2.2 Simple Cokriging

In the case of several correlated RFs that are sampled over the same spatial domain, assuming that means are known everywhere, the classical geostatistical method used for conducting multivariate prediction is SCoK (Cressie, 1991). If $\mathfrak{Z} = \left\{Z(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^d, Z(\mathbf{x}) \in \mathbb{R}^1\right\}$ and $\mathfrak{Y} = \left\{Y(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^d, Y(\mathbf{x}) \in \mathbb{R}^1\right\}$ are two zero-mean second-order stationary RFs and $\mathbf{Z}_\alpha = Z_1, \ldots, Z_n$ and $\mathbf{Y}_\gamma = Y_{n+1}, \ldots, Y_{n+m}$ are corresponding random vectors sampled from them at locations $\mathbf{x}_i$ $(i = 1, \ldots, m+n)$, the SCoK predictor for Z_0 is then

$$Z_0^p = \left(\begin{array}{cc} \boldsymbol{\sigma}'_\alpha & \boldsymbol{\sigma}'_\gamma \end{array} \right) \left(\begin{array}{cc} \boldsymbol{\Sigma}_{\alpha\alpha} & \boldsymbol{\Sigma}_{\alpha\gamma} \\ \boldsymbol{\Sigma}_{\gamma\alpha} & \boldsymbol{\Sigma}_{\gamma\gamma} \end{array} \right)^{-1} \left(\begin{array}{c} \mathbf{z}_\alpha \\ \mathbf{y}_\gamma \end{array} \right) \tag{2.1}$$

where the $\boldsymbol{\sigma}$s and $\boldsymbol{\Sigma}$s are obtained from the partitioning of the covariance matrix $\boldsymbol{\Sigma}$ for the whole vector $(Z_0, \mathbf{Z}'_\alpha, \mathbf{Y}'_\gamma)$, with

$$\boldsymbol{\Sigma} = \left(\begin{array}{ccc} \sigma_0^2 & \boldsymbol{\sigma}'_\alpha & \boldsymbol{\sigma}'_\gamma \\ \boldsymbol{\sigma}_\alpha & \boldsymbol{\Sigma}_{\alpha\alpha} & \boldsymbol{\Sigma}_{\alpha\gamma} \\ \boldsymbol{\sigma}_\gamma & \boldsymbol{\Sigma}_{\gamma\alpha} & \boldsymbol{\Sigma}_{\gamma\gamma} \end{array} \right) \tag{2.2}$$

(where the delimitations between the blocks in $\boldsymbol{\Sigma}$ respect the same conventions as in vector $(Z_0, \mathbf{Z}'_\alpha, \mathbf{Y}'_\gamma)$) whereas the corresponding variance of prediction is given by

$$\sigma_0^p = \sigma_0^2 - \left(\begin{array}{cc} \boldsymbol{\sigma}'_\alpha & \boldsymbol{\sigma}'_\gamma \end{array} \right) \left(\begin{array}{cc} \boldsymbol{\Sigma}_{\alpha\alpha} & \boldsymbol{\Sigma}_{\alpha\gamma} \\ \boldsymbol{\Sigma}_{\gamma\alpha} & \boldsymbol{\Sigma}_{\gamma\gamma} \end{array} \right)^{-1} \left(\begin{array}{c} \boldsymbol{\sigma}_\alpha \\ \boldsymbol{\sigma}_\gamma \end{array} \right) \tag{2.3}$$

Clearly, SCoK is the best linear predictor, but not necessarily the best predictor. However, assuming multivariate gaussianity for $(Z_0, \mathbf{Z}'_\alpha, \mathbf{Y}'_\gamma)$, the best predictor is linear, and SCoK then corresponds to the result of a linear regression, with

$$Z_0^p = \mathbb{E}\left[Z_0 | \mathbf{z}_\alpha, \mathbf{y}_\gamma\right] \quad ; \quad \sigma_0^p = \mathbb{V}ar\left[Z_0 | \mathbf{z}_\alpha, \mathbf{y}_\gamma\right] \tag{2.4}$$

and where $Z_0 | \mathbf{z}_\alpha, \mathbf{y}_\gamma \sim N(Z_0^p, \sigma_0^p)$. In other words, SCoK can only be considered as the best predictor when full gaussianity holds. If this is not the case, SCoK is still a valuable predictor, but its prediction variance is not necessarily the smallest possible one and the true conditional probability distribution function (pdf) for $Z_0 | \mathbf{z}_\alpha, \mathbf{y}_\gamma$ is not Gaussian in general.

Though SCoK is a straighforward and well-known method that can be easily numerically implemented without difficulties, there are some

practical issues that need to be addressed. Clearly, obtaining the covariance matrix $\mathbf{\Sigma}$ for arbitrary locations rely on a multivariate modeling of the covariance functions, in order to ensure positive-definiteness of the final results. The most frequently used method is based on the so-called Linear Model of Coregionalization (LMC; see e.g. Chilès and Delfiner, 1999 or Goovaerts, 1997), which for a bivariate case is written as

$$\begin{pmatrix} C_{\alpha\alpha}(\mathbf{h}) & C_{\alpha\beta}(\mathbf{h}) \\ C_{\alpha\beta}(\mathbf{h}) & C_{\beta\beta}(\mathbf{h}) \end{pmatrix} = \sum_i \begin{pmatrix} a_{i,\alpha\alpha} & a_{i,\alpha\beta} \\ a_{i,\alpha\beta} & a_{i,\beta\beta} \end{pmatrix} c_i(\mathbf{h}) = \sum_i \mathbf{A}_i c_i(\mathbf{h}) \tag{2.5}$$

where all matrices $\mathbf{A}_i$ are positive definite and where the same elementary positive definite covariance models $c_i(\mathbf{h})$ must be used for modeling all (cross-)covariance functions. Clearly, this imposes an important (and rarely discussed) conceptual constraint on the method, as any secondary variable that need to be accounted for when predicting Z_0 must fit into the second-order stationary RF paradigm, i.e. the random vector $\mathbf{Y}_\gamma$ must be considered as a sample from a second-order stationary RF that can be characterized by a covariance function $C_{\beta\beta}(\mathbf{x}_i - \mathbf{x}_j)$ that only depends on $\mathbf{x}_i - \mathbf{x}_j$ but neither on $\mathbf{x}_i$ nor on $\mathbf{x}_j$. Unfortunately, it happens frequently that quite useful information does not fit well into this paradigm. As a simple example, in an environmental pollution context, the distance $\mathbf{x} - \mathbf{x}_j$ to a chemical industry located at $\mathbf{x}_j$ is likely to be a quite relevant measure for quantifying atmospheric deposition $Y(\mathbf{x})$, but $Y(\mathbf{x})$ cannot be considered as coming from a RF whose covariance function would be translation invariant, i.e., depending only on $\mathbf{h}$, of course. This in turn may seriously impair the prediction of a related variable of interest $Z(\mathbf{x})$ (e.g., soil pollution) using a cokriging approach, as corresponding covariance function estimates $\widehat{C}_{\beta\beta}(\mathbf{h})$ and $\widehat{C}_{\alpha\beta}(\mathbf{h})$ are meaningless.

2.3 Bayesian Data Fusion

By the light of the previous example, it appears that relying on a multivariate second-order stationary RF framework is quite arguable in some instances. In order to account for this issue, a new theoretical framework has recently been proposed. Its aim is to alleviate the need of second-order stationarity for secondary variables at the price of mild simplifying hypothesis. Due to space limitations, only main and most important results will be presented here. Additional details can be found and are discussed at length in Bogaert and Fasbender (2007).

Let us assume that $\mathfrak{Z}$ is a zero-mean second-order stationary RF of primary interest, where $\mathbf{Z}$ is random vector sampled from it. Let

us define a mapping $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbf{g}(\mathbf{Z}) = \{g(Z_i)\}$ is a new random vector, along with an arbitrary zero-mean random vector $\mathbf{E}$ of same size and independent from $\mathbf{Z}$. The basic assumption of BDF is to assume that, for any secondary variable, we have $\mathbf{Y} = \mathbf{g}(\mathbf{Z}) + \mathbf{E}$. Clearly, $\mathbf{Z}$ and $\mathbf{Y}$ are collocated random variables (i.e. corresponding to the same locations) but $\mathbf{Y}$ is not longer second-order stationary in general, as its properties also depend on the arbitrary $\mathbf{E}$. According to the independence assumption $\mathbf{E} \perp \mathbf{Z}$, it is also clear that the conditional pdf for $\mathbf{Z}|\mathbf{y}$ is given by

$$f_{\mathbf{z}|\mathbf{y}}(\mathbf{z}|\mathbf{y}) \propto f_{\mathbf{z}}(\mathbf{z}) f_{\mathbf{y}|\mathbf{z}}(\mathbf{y}|\mathbf{z}) = f_{\mathbf{z}}(\mathbf{z}) f_{\mathbf{e}}(\mathbf{y} - \mathbf{g}(\mathbf{z})) \tag{2.6}$$

where $f_{\mathbf{z}}$ is the *a priori* pdf for $\mathbf{Z}$ and $f_{\mathbf{e}}$ is the pdf for $\mathbf{E}$. A more intuitive interpretation of Eq. 2.6 is to consider that each Y_i can be viewed as an indirect measurements of the true Z_i, as Y_i is a functional $g(Z_i)$ up to an additive error E_i, where it is reasonable in general to assume these errors as independent from the true $\mathbf{Z}$.

Starting from this last very general relation, a straightforward formulation can be proposed in a spatial prediction context. Let us consider that what is sought for is the conditional pdf for Z_0 given a set of observed values $\mathbf{z}_\alpha = z_1, \ldots, z_n$ for the RF of interest $\mathfrak{Z}$ along with a set of observed values $\mathbf{y}_\gamma = y_{n+1}, \ldots, y_{n+m}$ for the auxiliary RF $\mathfrak{Y}$. Defining additionally the vector $\mathbf{Z}_\beta = Z_{n+1}, \ldots, Z_{n+m}$ of unobserved variables at the same location as $\mathbf{Y}_\gamma$, the conditional pdf for $Z_0|\mathbf{z}_\alpha, \mathbf{y}_\gamma$ is then given by

$$f_{z_0|\mathbf{z}_\alpha,\mathbf{y}_\gamma}(z_0|\mathbf{z}_\alpha, \mathbf{y}_\gamma) \propto \int_{\mathbf{R}^m} f_{z_0,\mathbf{z}_\alpha,\mathbf{z}_\beta}(z_0, \mathbf{z}_\alpha, \mathbf{z}_\beta) f_{\mathbf{e}}(\mathbf{y}_\gamma - \mathbf{g}(\mathbf{z}_\beta)) d\mathbf{z}_\beta \tag{2.7}$$

This is a very general and nonlinear formula for prediction that requires the knowledge of the joint pdf $f_{z_0,\mathbf{z}_\alpha,\mathbf{z}_\beta}$ as well as the joint pdf of errors $f_{\mathbf{e}}$. A classical approach would be to consider $\mathfrak{Z}$ as a Gaussian RF along with a mutual independence hypothesis for the vector $\mathbf{E}$, so that $f_{z_0,\mathbf{z}_\alpha,\mathbf{z}_\beta}$ is multivariate Gaussian and $f_{\mathbf{e}} = \prod_i f_{e_i}$. Using again Baye's theorem, we then have $f_{e_i}(y_i - g(z_i)) \propto f_{z_i|y_i}(z_i|y_i)/f_{z_i}(z_i)$, so that Eq. 2.7 simplifies to

$$f_{z_0|\mathbf{z}_\alpha,\mathbf{y}_\gamma}(z_0|\mathbf{z}_\alpha, \mathbf{y}_\gamma) \propto \int_{\mathbf{R}^m} f_{z_0,\mathbf{z}_\alpha,\mathbf{z}_\beta}(z_0, \mathbf{z}_\alpha, \mathbf{z}_\beta) \prod_i \frac{f_{z_i|y_i}(z_i|y_i)}{f_{z_i}(z_i)} d\mathbf{z}_\beta \tag{2.8}$$

As a consequence, Eq. 2.8 is still a nonlinear expression that requires multivariate integration, but it is now possible to express the covariance matrix of the Gaussian distribution $f_{z_0,\mathbf{z}_\alpha,\mathbf{z}_\beta}$ from the covariance function $C_{\alpha\alpha}(\mathbf{h})$ (i.e. the covariance function of RF $\mathfrak{Z}$) as well as to infer

the pdf's $f_{z_i|y_i}$ from the data set. A synthetical illustration of this can be found in Bogaert and Fasbender (2007). The real advantage of this formulation is that it is not longer needed to have any stationarity hypothesis about $\mathfrak{Y}$ compared to SCoK. Moreover, it can be shown that a similar reasonsing can be used in order to account for multiple secondary information without any difficulties.

2.4 Comparing data fusion and cokriging

Though BDF and SCoK may appear at the first sight as completely different (and thus difficult to compare) approaches for accounting for secondary information, it can however be shown that close analytical linear relations for expressing the conditional mean and variances can be obtained for BDF if an additional multivariate Gaussian hypothesis is assumed to hold.

It has been reminded that, for jointly Gaussian RF's $\mathfrak{Y}$ and $\mathfrak{Z}$, the corresponding conditional pdf $f_{z_0|\mathbf{z}_\alpha,\mathbf{y}_\gamma}$ is Gaussian with mean and variance given by Eq. 2.4, so that SCoK is the best predictor. On the other side, for BDF, one can also remark that assuming that $\mathfrak{Y}$ and $\mathfrak{Z}$ are jointly zero-mean Gaussian RF's leads to a zero-mean Gaussian vector $(Z_0, \mathbf{Z}_\alpha, \mathbf{Z}_\beta, \mathbf{Y}_\gamma)$ with covariance matrix $\mathbf{\Sigma}$ as given by

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_0^2 & \boldsymbol{\sigma}'_\alpha & \boldsymbol{\sigma}'_\beta & \boldsymbol{\sigma}'_\gamma \\ \boldsymbol{\sigma}_\alpha & \mathbf{\Sigma}_{\alpha\alpha} & \mathbf{\Sigma}_{\alpha\beta} & \mathbf{\Sigma}_{\alpha\gamma} \\ \boldsymbol{\sigma}_\beta & \mathbf{\Sigma}_{\beta\alpha} & \mathbf{\Sigma}_{\beta\beta} & \mathbf{\Sigma}_{\beta\gamma} \\ \boldsymbol{\sigma}_\gamma & \mathbf{\Sigma}_{\gamma\alpha} & \mathbf{\Sigma}_{\gamma\beta} & \mathbf{\Sigma}_{\gamma\gamma} \end{pmatrix} \tag{2.9}$$

Several simplifications will then occur. First, the functional g is now a linear mapping with $\mathbf{g}(\mathbf{z}_\beta) = \frac{\sigma_{yz}}{\sigma_0^2}\mathbf{z}_\beta$, where $\sigma_{yz} = \sigma_{zy} = Cov(Z(\mathbf{x}), Y(\mathbf{x}))$, $\forall \mathbf{x} \in \mathbb{R}^d$. Second, $\mathbf{E}$ is now Gaussian with covariance matrix equal to $\sigma^2_{\gamma|\beta}\mathbf{I}$ where $\sigma^2_{\gamma|\beta}$ is equal to $\sigma_y^2 - \frac{\sigma^2_{yz}}{\sigma_0^2}$ with $\sigma_y^2 = \mathbb{V}ar[Y(\mathbf{x})]$, $\forall \mathbf{x} \in \mathbb{R}^d$. Third, since $(Z'_0, \mathbf{Z}'_\alpha, \mathbf{Z}'_\beta)'$ and $\mathbf{E}$ are Gaussian vectors, the product of pdf's in Eq. 2.8 is proportional to a Gaussian pdf with mean vector $\mathbf{M}$ and covariance matrix $\mathbf{S}$ given by

$$\begin{cases} \mathbf{S}^{-1} &= \begin{pmatrix} \sigma_0^2 & \boldsymbol{\sigma}'_\alpha & \boldsymbol{\sigma}'_\beta \\ \boldsymbol{\sigma}_\alpha & \mathbf{\Sigma}_{\alpha\alpha} & \mathbf{\Sigma}_{\alpha\beta} \\ \boldsymbol{\sigma}_\beta & \mathbf{\Sigma}_{\beta\alpha} & \mathbf{\Sigma}_{\beta\beta} \end{pmatrix}^{-1} + \begin{pmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma^2_{\gamma|\beta}}\frac{\sigma^2_{yz}}{\sigma_0^4}\mathbf{I} \end{pmatrix} \\ \mathbf{M} &= \frac{1}{\sigma^2_{\gamma|\beta}}\mathbf{S}\begin{pmatrix} 0 \\ \mathbf{0} \\ \frac{\sigma_{yz}}{\sigma_0^2}\mathbf{y}_\gamma \end{pmatrix} \end{cases} \tag{2.10}$$

Finally, since the integrand of Eq. 2.8 is proportional to a Gaussian pdf, integrating over $\mathbf{z}_\beta$ and conditioning on $\mathbf{z}_\alpha$ leads to the conclusion that the conditional pdf $Z_0|\mathbf{z}_\alpha, \mathbf{y}_\gamma$ is univariate Gaussian, thus completely characterized by its mean and variance, with

$$\begin{cases} \mathbb{E}[Z_0|\mathbf{z}_\alpha, \mathbf{y}_\gamma] &= \frac{1}{\sigma^2_{\gamma|\beta}} \frac{\sigma_{yz}}{\sigma_0^2} \mathbf{s}'_\beta \mathbf{y}_\gamma + \mathbf{s}'_\alpha \mathbf{S}^{-1}_{\alpha\alpha} \left(\mathbf{z}_\alpha - \frac{1}{\sigma^2_{\gamma|\beta}} \frac{\sigma_{yz}}{\sigma_0^2} \mathbf{S}_{\alpha\beta} \mathbf{y}_\gamma \right) \\ \mathbb{V}ar[Z_0|\mathbf{z}_\alpha, \mathbf{y}_\gamma] &= s_0 - \mathbf{s}'_\alpha \mathbf{S}^{-1}_{\alpha\alpha} \mathbf{s}_\alpha \end{cases} \tag{2.11}$$

where s_0, $\mathbf{s}_\alpha$, $\mathbf{s}_\beta$, $\mathbf{S}_{\alpha\alpha}$ and $\mathbf{S}_{\alpha\beta}$ are defined by

$$\mathbf{S} = \begin{pmatrix} s_0 & \mathbf{s}'_\alpha & \mathbf{s}'_\beta \\ \mathbf{s}_\alpha & \mathbf{S}_{\alpha\alpha} & \mathbf{S}_{\alpha\beta} \\ \mathbf{s}_\beta & \mathbf{S}_{\beta\alpha} & \mathbf{S}_{\beta\beta} \end{pmatrix} \tag{2.12}$$

and where the delimitations between the blocks in $\mathbf{S}$ respect the same conventions as in vector $(Z_0, \mathbf{Z}'_\alpha, \mathbf{Z}'_\beta)$.

The gain of BDF on SCok in terms of inference is non negligible. Indeed, instead of inferring the multivariate covariance model (with LMC namely), we only need to estimate three simple quantities : (i) $C_{\alpha\alpha}(\mathbf{h})$ the covariance function of $\mathfrak{Z}$, (ii) σ_y^2 the variance of $Y(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^d$ and (iii) σ_{yz} the pointwise covariance of $Z(\mathbf{x})$ and $Y(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^d$.

Comparing Eqs. 2.1 and 2.3 with Eq. 2.11 is not straightforward (expressions in these equations are in fact equivalent under independence hypothesis of vector $\mathbf{Y}_\gamma$ conditionally to vector $\mathbf{Z}_\beta$), but it is worth noting that (i) BDF now provides expressions for the conditional expectation and variance that are linear with respect to the observed values $\mathbf{z}_\alpha$ and $\mathbf{y}_\gamma$ as for SCoK, and (ii) results for BDF and SCoK can be compared too on a common basis, as we fulfill the optimal conditions for using SCoK as the best possible predictor. It is worth remembering however that, in general, BDF and SCoK will obey distinct optimality properties : SCoK is optimal under the hypothesis of jointly Gaussian RF's $\mathfrak{Y}$ and $\mathfrak{Z}$, whereas BDF only relies on a Gaussian hypothesis for $\mathfrak{Z}$ and an independence hypothesis for $\mathbf{E}$, which is a somewhat milder hypothesis than assuming joint Gaussianity for both RF's $\mathfrak{Y}$ and $\mathfrak{Z}$.

2.5 A synthetic case study

In order to illustrate the similitudes of the results that are obtained using both approaches in a situation where SCoK is expected to give the best possible results, a synthetic case study is presented. The aim here is to show that, even under optimal conditions for SCoK, using SCoK instead

of BDF does not significantly increase the quality of predictions. Stated in other words, the loss of information due to the use of BDF instead of SCoK does not dramatically affect the quality of the predictions.

Let assume a smooth zero-mean unit-variance Gaussian RF $\mathfrak{Z}$ for which a realization over a 100×100 regular grid is given in Fig. 2.1a (covariance function is exponential with sill equal to 1 and range equal to 30). Let assume also a second Gaussian RF $\mathfrak{Y}$ (Fig. 2.1b) defined as a linear combination of the RF $\mathfrak{Z}$ and another zero-mean unit-variance Gaussian RF $\mathfrak{E}$ with the same covariance function. By taking $\mathfrak{Y} = 2\mathfrak{Z} + \mathfrak{E}$, the resulting RF $\mathfrak{Y}$ is thus a zero-mean Gaussian RF with variance equal to 5, and both RF's are jointly Gaussian with pointwise correlation equal to 0.894. Under these conditions, SCoK is thus the best possible predictor. For conducting predictions, two random samples $\mathbf{z}$ and $\mathbf{y}$ are extracted from these simulated grids. In order to be in a situation when SCoK would be interesting compared to simple kriging (i.e. the auxiliary $\mathbf{y}$ conveys valuable extra information compared to what is already known from $\mathbf{z}$), samples size have been chosen equal to 200 and 400 for $\mathbf{z}$ and $\mathbf{y}$, respectively. Predictions of $\mathfrak{Z}$ is then conducted at the nodes of the grid using $\mathbf{z}$ and $\mathbf{y}$ as observed values. It is worth noting too that sampling has been conducted so that there are no locations for which $\mathfrak{Z}$ and $\mathfrak{Y}$ are jointly observed.

In oder to compare relative differences between BDF and SCoK results compared to simple kriging results, a relative comparison approach has been used here. Clearly, using simple kriging with the n closest observations of $\mathbf{z}$ (that will be denoted as SK_n) provides a lower bound in terms of prediction quality, as it only makes use of the primary variable. On the other side, using simple kriging with the n closest observations of $\mathbf{z}$ along with the $\mathbf{z}$ observations at the n closest locations for the observed $\mathbf{y}$ (that will be denoted as SK_{2n}) provides an upper bound, as it assumes that the true values for the primary variable are available at locations where only the auxiliary variable is observed. As a consequence, in terms of quality predictions, results for BDF and SCoK will be located somewhere in between these two bounds. This also provides a way to compare the results for BDF and SCoK on a relative scale ranging from SK_n to SK_{2n}.

Figure 2.2 shows the predictions results obtained using the four previously described methods, namely SCoK (Fig. d), BDF (Fig. c) and the two extreme simple kriging situations (Figs. a and b). One can notice that BDF and SCoK provide visually very similar results. This observation is confirmed from the Root Mean Squared Error (RMSE) as computed between simulated and predicted values (see Table 2.1). As expected, SCoK and BDF gives intermediate results in between the two

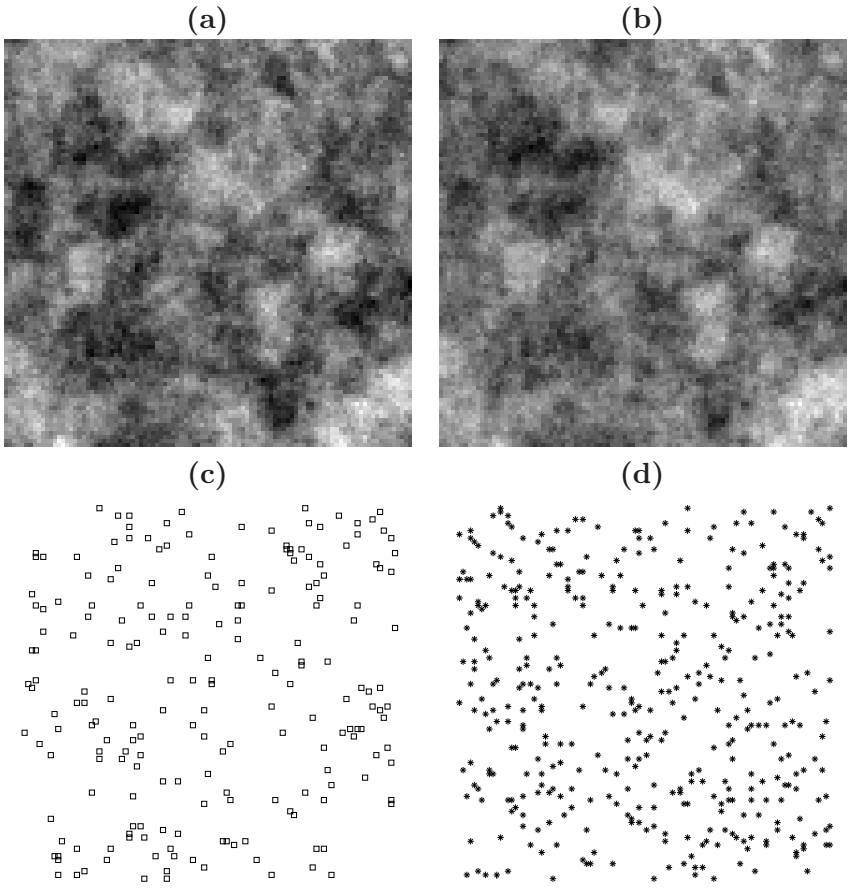


Figure 2.1: Simulations over a 100×100 regular grid for RF $\mathfrak{Z}$ **(a)** and RF $\mathfrak{Y}$ **(b)**, along with the random sample $\mathbf{z}$ at 200 locations **(c)** and the random sample $\mathbf{y}$ at 400 locations **(d)**.

kriging predictions, with only a slight advantage for SCoK when values are compared on a relative scale.

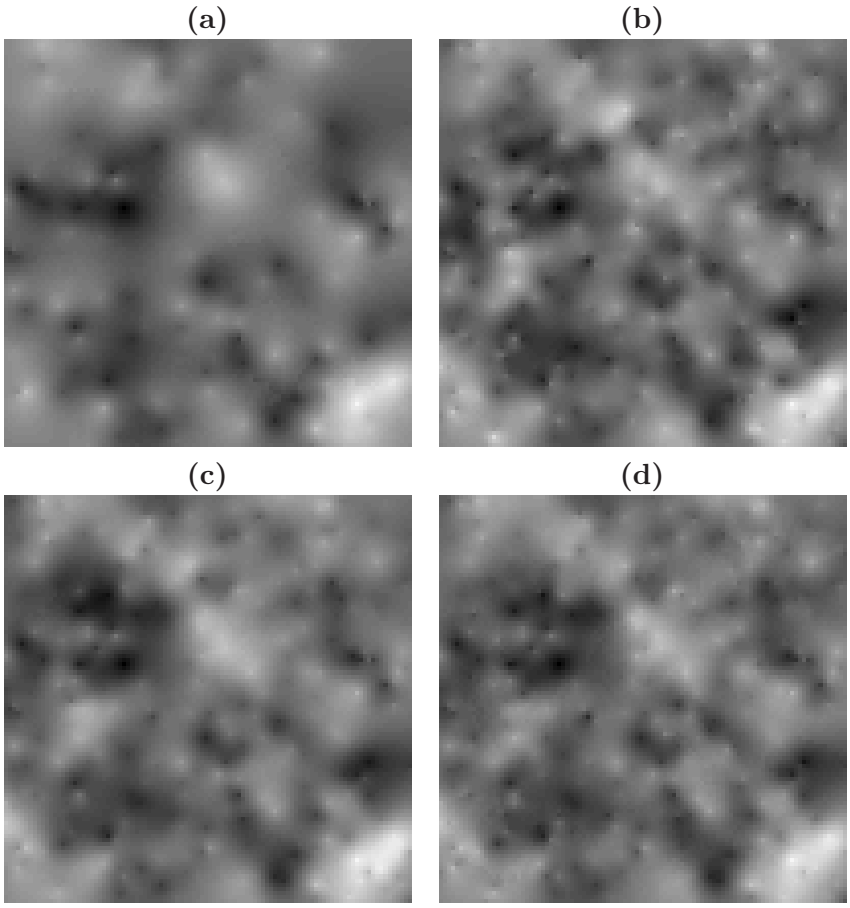


Figure 2.2: Results for the different predictions methods. **(a)** is simple kriging with 10 closest points, **(b)** is simple Kriging with 2×10 closest points, **(c)** is BDF with 10 closest points for each RF and **(d)** is SCoK with 10 closest points for each RF.

The above computations can been repeated by keeping everything identical except for the pointwise correlation between the two RF's $\mathfrak{Z}$ and $\mathfrak{Y}$. It can be seen from Fig. 2.3 that the relative difference between BDF and SCoK is null for high and low correlations (i.e., when secondary information is useless and when secondary information is equivalent to

Table 2.1: Quality assessment for the various prediction methods. RMSEs are computed as the root mean squared differences between simulated and predicted values. Relative RMSEs are RMSEs rescaled between 0 and 1 according to the bounds as provided by SK_n and SK_{2n} (value 1 is thus the best possible result whereas value 0 is the worst possible one).

	SK_n	SCoK	Bayesian Data Fusion	SK_{2n}
RMSE	0.63	0.53	0.55	0.48
Relative RMSE	0	0.68	0.58	1

primary one, respectively), so that biggest differences will be observed for intermediate values.

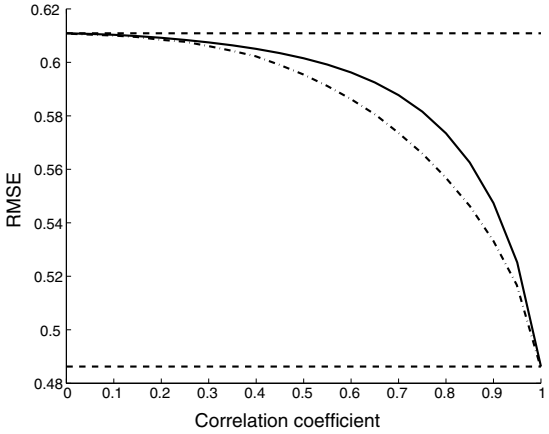


Figure 2.3: Evolution of the RMSE with respect to the pointwise correlation between $\mathfrak{Z}$ and $\mathfrak{Y}$ RF's. Dashed lines correspond to SK_n (higher value) and SK_{2n} (lower value), whereas plain line and dotted line correspond to BDF and SCoK, respectively.

2.6 Conclusions

In this chapter, a BDF approach has been proposed as an interesting alternative way to account for various secondary information in a spatial prediction context. Indeed, the method does not rely on a classical multivariate second-order stationary RF framework, as it is required for

cokriging. Because by nature BDF relies on a different set of assumptions, it is difficult to compare both methods from a general viewpoint in situations where both of them would be relevant. In order to overcome this difficulty, a comparison has been conducted in a situation where SCoK is known to be the best possible predictor, so that the loss of information caused by using BDF instead of SCoK can be assessed on an objective way. Results show that in terms of performances, differences are however quite limited.

Of course, there is no point in using data fusion instead of cokriging when one can reasonably assume that a second-order multivariate RF hypothesis holds, as SCoK is then by definition the right method to be used. However, the application field of BDF is quite more general, as it allows the user to account for secondary information that would not fit into this framework, permitting thus to deal with a much wider panel of situations that could not be meaningfully handled by SCoK. Hence, BDF can be viewed as a robust alternative to cokriging for multivariate prediction where cokriging hypotheses are known to be irrelevant or at least quite arguable. Moreover, it is also possible to deal with extreme deviation from Gaussianity by applying transformation on the raw data so that Gaussian distributions can be assumed (e.g. Box-Cox transformation). This should thus enlarge the scope of the results found in this chapter.

Finally, it is worth noting that the aim of this chapter was not to suggest that sound multivariate spatial prediction methods like cokriging, which have frequently proved to be quite useful, should be discarded or criticized when used under the appropriate hypotheses. It is rather the limited practical pertinency of these hypotheses which is at stake here. It is suggested that in situations where a multivariate modelling does not appear to be conceptually consistent with what is known from data, BDF is then a more reasonable choice by avoiding the need of using a multivariate model (e.g., as the LMC) at the price of mild simplifying hypotheses. Though rather simple in the case of a single auxiliary variable, this modelling problem is expected to become critical in situations where the number and diversity of auxiliary informations sources that need to be accounted for is increasing. This is a typical case where BDF is expected to be a much more flexible way of handling the problem than SCoK.

Part II

Applications

Preamble to the applications

In the first part of the thesis, several theoretical concepts and results have been presented. As the second part of the thesis is dedicated to applications, it is clear that the choice of these applications is particularly important in order to properly illustrate these theoretical aspects. Consequently, prior to really enter into the applications in details, it is necessary at this point to present a short motivation for each of the four situations that were studied within the framework of this thesis (see Table 2.2 for a summary of these motivations for each of the following applications).

In Chapter 3, multispectral satellite images were enhanced using panchromatic images of the same scene. The case of satellite images is particular as data are exhaustively known over space. In this application, IKONOS images were chosen as (i) multispectral and panchromatic images share the same spectral ranges and (ii) the IKONOS sensor gets coregistred image, thus avoiding the need of orthorectification. Further, as there were two information sources, it was interesting to use the weighting parameters proposed in Section 1.4.2 in order to choose to give more credit to one of the sources depending on the user's needs.

Chapter 4 is also dedicated to spatial enhancement of satellite images. However, the choice of ASTER images was made here in order to illustrate that the BDF methodology also works when the spectral ranges of the different images are not overlapping. Furthermore, the change of support issue was explicitly taken into account in this application.

The water table spatial mapping situation of Chapter 5 is the first application in which data are not exhaustively known over space. The main challenge of this application was how to deal with the information provided by a river network position that does not straightforwardly fit within classical random fields theory (as required in cokriging approaches).

Finally, the BDF methodology is applied to a space-time prediction context. The case of NO_2 concentrations was particularly interesting as up to 6 information sources were available to an extent that accounting for all these sources is very difficult in practice.

Table 2.2: Summary of the motivations for the choice of each of the applications in the thesis.

	Chap. 3	Chap. 4	Chap. 5	Chap. 6
Exhaustive data	X	X		
Weighting parameters	X			
Gaussian distributions	X	X	X	X
Spatial	X	X	X	
Space-Time				X
Number of inf. sources	2	2	3	6

Chapter 3

Bayesian Data Fusion for Adaptable Image Pansharpening

The previous chapters dealt with theoretical concerns about Bayesian data fusion. In this chapter, we show a first application in which Bayesian data fusion provides an interesting solution in the context of image pansharpening (i.e. spatial enhancement of remotely sensed images). Several images obtained by the IKONOS sensors were chosen here as illustration of the method.

3.1 Outline

Nowadays, most optical Earth Observation satellites carry both a panchromatic sensor and a set of lower spatial resolution multispectral sensors. In order to benefit from both sources of information, several pansharpening methods have been developed to produce a multispectral image at the spatial resolution of the panchromatic band. The aim of this study is to suggest a novel approach to the pansharpening problem within a Bayesian framework. This Bayesian data fusion method relies on statistical relationships between the various spectral bands and the panchromatic band without suffering from restricting modeling hypotheses. Furthermore, it allows the user to weight the spectral and panchromatic information with respect to either visual or quantitative criteria, which leads to adaptable results according to users' needs and study areas.
The performance of this approach was compared to existing methods based on markedly different subset images from very high spatial reso-

lution IKONOS images. Results show that Bayesian data fusion yield the highest spectral consistency. Furthermore, small details were adequately added to the pansharpened images with little artifact compared to those created using wavelet-based methods. Finally, the method is fast and easy to implement thanks to its straightforward formulation. As it does not have any intrinsic limitations on the type of data to be processed or the number of bands to be merged, it also appears to be very promising for optical/SAR or hyperspectral image fusion [1].

[1] ©2008 IEEE. Reprinted, with permission, from Fasbender *et al.* (2008b).

3.2 Introduction

There is currently a wide variety of satellite sensors collecting imagery with different spatial and spectral resolutions. Among them, several present and future platforms carry at the same time a fine spatial resolution panchromatic sensor and a set of coarser resolution spectral sensors. These two different but complementary sources of information are used in image fusion, also called pansharpening. This aims at merging the high spatial resolution of the panchromatic band with the high spectral resolution of the other bands to obtain a new image that retains the best of these two sources (i.e. both high spatial and high spectral resolutions).

Pansharpening can be a key pre-processing step in remote sensing. In the particular case of Very High spatial Resolution (VHR) satellite sensors (e.g. FORMOSAT-2, ORBVIEW-3, IKONOS, Pleiade or Quickbird), it consists of merging the four bands of the multispectral image with the four-times finer panchromatic band. (Laporterie-Déjean *et al.*, 2005) distinguished six applications, namely agriculture, forest, urban planning, coastal environment, defense and cartography. To give a few methodological examples, VHR image pansharpening has been performed to improve visual photointerpretation (Yamazaki *et al.*, 2005), digital surface model extraction in stereo images (Raggam *et al.*, 2005), texture analysis (Souza and Roberts, 2005) and the automated detection of linear or small objects such as, e.g., roads (Mohammadzadeh *et al.*, 2006) or wilting citrus trees (Fletcher, 2005).

Depending on users' needs and the land cover of the study areas, different pansharpening algorithms are commonly used. These include Brovey method, Intensity-Hue-Saturation (IHS), Principal Component Analysis (PCA) and wavelet-based Multi-Resolution Analyses (MRA). The first three are based on various linear combinations and substitutions of the original bands, whereas MRA are non linear. Detailed reviews of the numerous available algorithms can be found in (Pohl and van Genderen, 1998; Chavez *et al.*, 1991; Wang *et al.*, 2005; Ballester *et al.*, 2006) or (Laporterie-Déjean *et al.*, 2005). These methods will only be briefly described below in order to point out known limitations and issues that need to be addressed with respect to VHR satellite images.

The main hypothesis of the Brovey method is that the ith fused spectral band is proportional to the panchromatic one with a coefficient equal to the ratio of the ith initial bands and the sum of all the spectral bands. This method was designed for 3 bands but has been generalized to a larger number of bands as well (Pohl and van Genderen, 1998). However, assuming proportionality between the bands is arguable for

IKONOS images, as the spectral range of the panchromatic band and the union of the spectral ranges of the spectral bands do not correspond. Tu *et al.* (2001) shows that color distortion occurs when using Brovey transform, mainly due to a change in saturation.

The IHS method consists of a spectral transformation from an orthogonal three-dimensional color space into a conical IHS color space (Harrison and Jupp, 1990). In a pansharpening framework, the panchromatic band substitutes the intensity component of the multispectral image, and the reverse transformation is applied back to the original color space. For more than three bands, IHS can be adapted by selecting groups of three bands or by projecting the original color space into a three-dimensional color space (Chen *et al.*, 2003). IHS fusions for VHR satellite images can result in color distortions (Zhang, 1999), often toward more vivid colors (Chavez *et al.*, 1991).

Like the IHS method, the PCA method is based on a spectral transformation of the image but this time with no limitation on the number of spectral bands. It consists of the computation of the eigenvectors of the correlation or covariance matrix, leading to a linear combination of the initial spectral bands (the Principal Components). The first component, related to the greatest fraction of the variance, is then substituted by the panchromatic band. This is followed by a back transform of the principal components toward the original color space. Depending on the data structure of the area to be fused, the replacement of the component of highest variance by the panchromatic band may maximize the effect of the panchromatic image on the final result (Zhang, 1999), thus weakly accounting for the multispectral information.

Unlike these methods that rely on various linear combinations of the original bands, the numerous variations of the MRA approach consist of combining the original bands in a non-linear way. Wavelet coefficients are computed for both spectral and panchromatic bands. Lower resolution coefficients of the spectral bands are then combined with the higher resolution coefficients of the panchromatic band in order to define the wavelet coefficients of the fused spectral band. This leads to stable image fusions regardless of the spectral dimension of the multispectral images (e.g., Zhou *et al.*, 1998; Li *et al.*, 2002; Tu *et al.*, 2001; Du *et al.*, 2003; Garzelli and Nencini, 2005; Ranchin *et al.*, 2003; Simone *et al.*, 2002). However, since exactly the same procedure is applied to all spectral bands without encompassing the relationships between the spectral and the panchromatic bands, the wavelet methods may also suffer from color distortions (e.g. Zhang and Hong, 2005). Moreover, as the information in the panchromatic band is somewhat forced in the procedure, it may generate artifacts that are similar to the effect of a high-pass

filter (Zhang, 1999; Hirschmugl *et al.*, 2005). It may induce a loss of color information for small objects (Ranchin and Wald, 2000).

These methods are commonly used because of their relatively straightforward implementation and fast computation. There are also other advanced approaches such as variational methods (e.g. Ballester *et al.*, 2006), high-pass filter methods (e.g. Chavez *et al.*, 1991), generalized Laplacian filter methods (e.g. Aiazzi *et al.*, 2002a,b) and more recently Super-Resolution Variable-Pixel Linear Reconstruction (Merino and Nunez, 2007). Only wavelet was selected here from the advanced methods because of its growing interest in recent studies.

For users wishing to merge data while preserving most of the original spatial and spectral information, choosing one method among the many possibilities is not straightforward. Indeed, their performance may vary according to the specific data at hand, so that the choice to be made relies heavily on users' experience (Pohl and van Genderen, 1998; Hirschmugl *et al.*, 2005; Laporterie-Déjean *et al.*, 2005). Moreover, each method suffers from its own limitations with respect to the quality of the final result. In order to avoid these issues, a Bayesian approach to the problem is proposed in this chapter, with the aim of obtaining a general image fusion method that is adaptive, robust and easy to implement.

The first part of the chapter will present the specific issues of IKONOS imagery, which will be used to illustrate the new fusion approach. The second part will focus on a brief presentation of Bayesian data fusion methodology. The last part of the chapter will be devoted to quality assessments and discussions of this method compared to standard ones. Based on the chosen samples, the goal will be to assess how the same Bayesian data fusion approach performs under very different situations compared to other methods. Emphasis will thus be put on the adaptability and robustness of the proposed approach when used in markedly different study areas, compared to commonly used methods for which the optimality of the results is more case-specific.

3.3 Data and study area

In order to compare the various methods in a pansharpening context, four subsets (Fig. 3.1) of 500×500 pixels were extracted from three different IKONOS Geo Orthokit images. These subsets are illustrative of the three major land-cover classes (urban, forested and agricultural areas) in Northern France and Southern Belgium as well as a rural areas in Northern Morocco (see Table 3.1 subset description). These samples address markedly different issues: the urban subset includes impervi-

ous areas with many objects close to the sensors spatial resolution; the mixed forest displays temperate deciduous and coniferous forest stands with different textures; the agricultural area consists of homogeneous fields covered by crop or void depending on crop phenology at the acquisition date; and finally the rural area in Morocco is a patchwork of roads, houses, orchards, pastures and forests. Despite the small number of subsets chosen, they thus reflect the main situations that were highlighted as key concerns in (Laporterie-Déjean *et al.*, 2005), except for coastal areas.

Table 3.1: List of image subsets used for the study

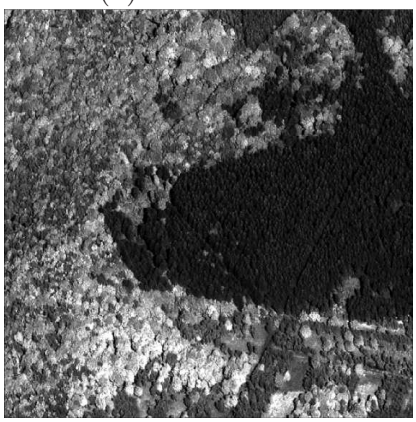
Subset	**Location**	**Acquisition date**
Urban	Northern France	29/06/2004
Forest	Southern Belgium	07/06/2004
Agricultural	Southern Belgium	07/06/2004
Mediterranean	Northern Morocco	25/04/2005

IKONOS images raise several pansharpening issues. The bundle product includes four spectral bands with 4-meter spatial resolution and one panchromatic band with 1-meter spatial resolution. There are thus 16 panchromatic pixels per multispectral one. Moreover, all IKONOS images have a relative spectral response above 50 percent in the regions 444.7 - 516 nm in blue, 506.4 - 595 nm in green, 631.9 - 697 nm in red and 757.3 - 852.7 nm in near-infrared, while it is between 525.8 - 928.5 nm in panchromatic (GeoEYE, 2006). Unlike Quickbird and Orbview-3, where the panchromatic band ranges from 450 to 900 nm and the near-infrared band ranges up to 900 nm, the spectral response of the panchromatic band extends above the near infra-red but is lower in the blue band (Fig. 3.2). IKONOS is therefore the most difficult case of VHR image fusion when prediction of the pansharpened images relies on the linear combination of existing bands.

The radiometric resolution of the images was 11 bits. Pansharpening was performed before orthorectification (as both the panchromatic and the multispectral images were taken at the same moment from the same sensor, thus having the same topographic distortion) and all multispectral images were resampled at the spatial resolution of the panchromatic bands using a bicubic convolution, as suggested by Ranchin *et al.* (2003) and Alparone *et al.* (2004). Bicubic convolution can be approximated by kernel smoothing method (with a specific kernel function) when speed is

(a) Urban area

(b) Forested area



(c) Agricultural area

(d) Mediterranean area

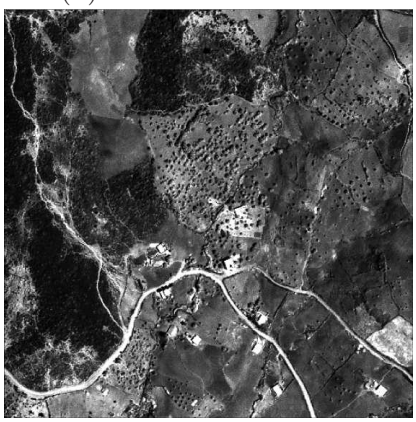


Figure 3.1: Panchromatic band of the four IKONOS 500 × 500 image subsets.

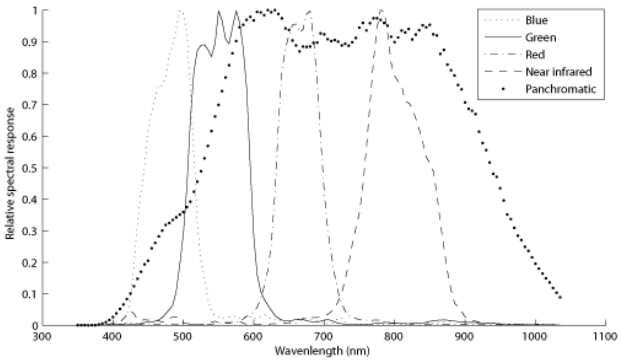


Figure 3.2: Relative spectral response of IKONOS sensors.

an issue even if bicubic convolution does not belong to the set of kernel smoothing methods.

3.4 Bayesian Data Fusion

Combining different sources of information into a single final result (i.e. data fusion) is a problem of general concern in a wide variety of applications, going far beyond satellite imagery and including a wide array of potential methods. Among them, Bayesian approaches have lead to interesting applications in various areas such as image surveillance (Jones *et al.*, 2003), object recognition (Chung and Shen, 2000), object localization (Pinheiro and Lima, 2004), robotics (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) and environmental modeling (Wikle *et al.*, 2001; Christakos, 2002), to quote just a few. The main advantage of a Bayesian approach is to put the problem of data fusion into a clear probabilistic framework. Recently, (Bogaert and Fasbender, 2007) suggested a Bayesian Data Fusion approach (BDF) to deal with spatial data in general. An adaptation of these general results will be given below, with the pansharpening of satellite images specifically in mind.

3.4.1 General formulation

The basic concept of BDF as presented by Bogaert and Fasbender (2007) relies on the idea that the variables of interest, denoted as vector $\mathbf{Z}$, cannot be directly observed. Instead, they are linked with the observable variables $\mathbf{Y}$ through an error-like model, with

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

where $g(\mathbf{Z})$ is a set of functionals and $\mathbf{E}$ is a vector of random errors that are stochastically independent of $\mathbf{Z}$. Using elementary probability calculus, it is possible to derive the conditional probability density function (pdf) of vector $\mathbf{Z}$ given the observed variables, with

$$f(\mathbf{z}|\mathbf{y}) \;\propto\; f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}}(\mathbf{y} - g(\mathbf{z})) \tag{3.1}$$

where $f_{\mathbf{Z}}(.)$ is the *a priori* pdf for $\mathbf{Z}$ and $f_{\mathbf{E}}(.)$ is the *a priori* pdf of the errors $\mathbf{E}$. In the context of pansharpening, the variables $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ correspond to a same target pixel (i.e., a multispectral pixel with a high spatial resolution). Besides, the set of observable variables is $\mathbf{Y} = (\mathbf{Y}'_S, Y_P)'$, where $\mathbf{Y}_S = (Y_{S,1} \ldots, Y_{S,n})'$ refers to the various spectral bands of the corresponding pixel on the low spatial resolution image, and where Y_P refers to the panchromatic pixel on the high spatial resolution image. Using the same partitioning, let us define $\mathbf{E} = (\mathbf{E}'_S, E_P)'$ as the vector of errors, assumed here to be zero-mean (as are the corresponding set of functionals $g_S(.)$ and $g_P(.)$) so that $\mathbf{Y}_S = g_S(\mathbf{Z}) + \mathbf{E}_S$ and $Y_P = g_P(\mathbf{Z}) + E_P$. Assuming independence between $\mathbf{E}_S$ and E_P, Eq. 3.1 can be written as

$$\begin{aligned} f(\mathbf{z}|\mathbf{y}_S, y_P) \;\propto\; & f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}_S}(\mathbf{y}_S - g_S(\mathbf{z})) \\ & \times f_{E_P}(y_P - g_P(\mathbf{z})). \end{aligned} \tag{3.2}$$

To account for the fact that $\mathbf{Y}_S$ and Y_P may carry information of quite different quality about $\mathbf{Z}$, Eq. 3.2 can be modified to obtain

$$\begin{aligned} f(\mathbf{z}|\mathbf{y}_S, y_P) \;\propto\; & f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}_S}(\mathbf{y}_S - g_S(\mathbf{z}))^{2(1-w)} \\ & \times f_{E_P}(y_P - g_P(\mathbf{z}))^{2w} \end{aligned} \tag{3.3}$$

where the parameter w (with $w \in]0, 1[$) can be interpreted as the weight to be given to the panchromatic information at the expense of the multispectral information. In other words, w corresponds to the preference that is to be given to the panchromatic image compared to the multispectral one in the pansharpening procedure. Setting $w = 0.5$ gives the same weight to both images, so leading back to Eq. 3.2,

whereas considering w close to either 0 or 1 is equivalent to discarding the panchromatic or the multispectral image, respectively. Clearly, large w values will correspond to an emphasis on spatial resolution whereas small w values will do the same for spectral information. This has the advantage of offering the user the possibility of balancing the two sources of information depending on the focus of interest.

3.4.2 Specific assumptions

The previous equations are very general, so additional assumptions are needed to make Eq. 3.3 tractable in the present context, from both practical and numerical viewpoints :

- As $\mathbf{Y}_S$ are the observed spectral bands, we may expect them to be directly related to the unknown $\mathbf{Z}$, so that using $g_S(\mathbf{z}) = \mathbf{z}$ (i.e., $\mathbf{Y}_S = \mathbf{Z} + \mathbf{E}_S$) seems to be a reasonable choice. Functional $g_S(\mathbf{z})$ could also be replaced by , for example, a modulation transfer function to account for the physical model for the image acquisition process. This was not implemented here as a simple implementation was preferred.

- For inferring the pdf of $\mathbf{E}_S$, let us assume that $\mathbf{E}_S \sim N(\mathbf{0}, \mathbf{\Sigma}_S)$ (i.e., $\mathbf{E}_S$ is multivariate Gaussian), where $\mathbf{\Sigma}_S$ is the estimated covariance matrix as computed from the observed multispectral image.

- Let us assume that $E_P \sim N(0, \sigma_P^2)$.

- Let us assume that the functional $g_P(\mathbf{z})$ can be expressed as a linear regression model that links the multispectral pixels to the panchromatic ones, as estimated from both observed images, so that $g_P(\mathbf{z}) = \alpha + \mathbf{z}'\boldsymbol{\beta}$ where α and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)'$ are regression parameters. It is worth noting that the IHS, PCA and Brovey methods all rely on a homologous linear modeling of the panchromatic band as a function of the spectral bands. Furthermore, similar expressions can be found in (Blum, 2005, 2006; Ballester *et al.*, 2006) and (Laporterie-Déjean *et al.*, 2005). Because of its simplicity, this choice also has the advantage of leading to simple analytical results.

- Finally, for the *a priori* pdf, let us assume a non-informative prior pdf with $f_{\mathbf{Z}}(\mathbf{z}) \propto 1$, as no clear information on this is available (see e.g. Gelman *et al.*, 2004).

It is worth noting here that even though the previous hypotheses are quite convenient, as they lead to a dramatic decrease in processing time,

they are by no means mandatory and could be modified, depending on circumstances or preferences.

Plugging these assumptions into Eq. 3.2 and using some simple algebraic manipulations (see the Appendix C for details), it can be shown that the conditional pdf of $\mathbf{Z}$ given the observed $\{\mathbf{y}_S, y_P\}$ is multivariate Gaussian, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ given by

$$\begin{cases} \boldsymbol{\mu} &= \boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}_S^{-1}\mathbf{y}_S + \frac{1}{\sigma_P^2}(y_P - \alpha)\boldsymbol{\beta}\right) \\ \boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Sigma}_S^{-1} + \frac{1}{\sigma_P^2}\boldsymbol{\beta}\boldsymbol{\beta}' \end{cases} .$$

Hence, the $\boldsymbol{\mu}$ vector is a first relevant estimator for the unknown $\mathbf{Z}$. Plugging the same results into Eq. 3.3, the conditional distribution is still multivariate Gaussian, but now with mean vector $\boldsymbol{\mu}_w$ and covariance matrix $\boldsymbol{\Sigma}_w$ given by

$$\begin{cases} \boldsymbol{\mu}_w &= \boldsymbol{\Sigma}_w\left(2(1-w)\boldsymbol{\Sigma}_S^{-1}\mathbf{y}_S + 2w\frac{1}{\sigma_P^2}(y_P - \alpha)\boldsymbol{\beta}\right) \\ \boldsymbol{\Sigma}_w^{-1} &= 2(1-w)\boldsymbol{\Sigma}_S^{-1} + 2w\frac{1}{\sigma_P^2}\boldsymbol{\beta}\boldsymbol{\beta}' \end{cases} .$$

As a result, $\boldsymbol{\mu}_w$ is a relevant estimator of the unknown $\mathbf{Z}$ when a weighting of the panchromatic and multispectral images is applied, and its value depends on the specific choice of w, thus leading to an adaptable BDF ($\boldsymbol{\mu}$ is of course a particular case of $\boldsymbol{\mu}_w$ when $w = 0.5$).

3.5 Quality assessment

All the state-of-the-art algorithms described in the introduction (namely Brovey, IHS and PCA) have been implemented in Matlab. Some methodological aspects need to be noticed: IHS pansharpened images were computed by groups of three spectral bands, namely NIR-Red-Green and Red-Green-Blue; the wavelet method was implemented in R software (R, 2006) using the Daubechies orthonormal compactly supported wavelet of length L=8 (Daubechies, 1992).

A comparative quality assessment of pansharpening methods can be based on various criteria, none of which is currently considered as standard in the literature. Broadly speaking, the evaluation of pansharpening performance can be divided into qualitative and quantitative approaches. Visual comparison between the merged images given by different pansharpening methods is the main qualitative approach (Laporterie-Déjean *et al.*, 2005; Zhang, 1999). It is relevant when the fusion aims at easing visual photointerpretation, its main drawback being

however its obvious dependency on the experts involved in the assessing the results. For these reasons, if the fusion is intended to be a part of an automated workflow, it may be preferable to rely on a quantitative approach, in order to preserve the characteristics that are the most useful for the final goal.

There are several spectral quality criteria for quantitative comparisons. The shift or relative shift of the mean computed over the entire study area (see e.g., Laporterie-Déjean *et al.*, 2005) is not particulary relevant here, since two images can share the same mean but still be completely different. Zhou *et al.* (1998) used the average difference between the merged image and the original ones. The main disadvantage of these approaches is that there are no reference values, thus preventing a direct comparison between spectral quality criteria. Among other alternatives, there is a large family of criteria based on correlation coefficients, to be computed either between the spectral bands (Pardo-Iguzquiza *et al.*, 2006) or between the hue bands of the corresponding merged and original images (Zhou *et al.*, 1998). Correlations have the advantage of always lying in [-1,1], thus enabling direct comparisons between quality criteria. In this study, the correlations between the merged and the original images were used, as it is clear that the ith fused band has to be close to the ith original band. Methods aiming at low color distortion should therefore yield high correlations between the original and fused bands. Recently, Alparone *et al.* (2004) proposed a generalization of the correlation coefficients metric and referred it as Q_4. It aims to encapsulate both spectral and radiometric distortion measurements by accounting for local measurements of (i) the correlation coefficient, (ii) the bias and (iii) the change in contrast, the three quantities being computed using quaternions algebra (Kantor and Solodnikov, 1989). However, a minimum spectral distortion may also arise from a lack of detail added to the fused image. It is therefore necessary to consider the spatial quality of the images as well. Spatial quality criteria are, however, much more rarely used, and are also often based on correlations. Zhou *et al.* (1998) used correlations (i) between high frequency components of the Laplacian decompositions of the merged and panchromatic image, or (ii) between the intensities of the near-infrared composite and the panchromatic image. In this study, an intensity image based on the four bands of the pansharpened image was used.

In addition to the choice of a specific quantitative quality criterion, the BDF method also needs to address the choice of w, the weighting coefficient which enables the multispectral and panchromatic information to be balanced, as shown in Eq. 3.3. The use of this possibility has advantages and disadvantages. A main advantage is the adaptability of

the method to the user's need. Indeed, w can be tuned with respect to the specificities of the case being studied. The main disadvantage is that it implies the identification of a criterion to optimize the value of w. Three options can be considered : (i) optimizing w with respect to a visual interpretation of the corresponding fused image; (ii) optimizing w using some given quantitative criteria; or (iii) choosing $w = 0.5$ so that Eq. 3.3 simplifies to Eq. 3.2, i.e. the non-weighted version. Optimization based on visual interpretation ensures that results will be visually pleasant for the user, but it is a highly subjective approach. In order to be more objective, a quantitative criterion may be optimized. As there is no obvious or unique possible choice of such a criterion, it is proposed here to rely on a combination of spectral and spatial quality criteria, called T, so that the optimal value w_{opt} is given by

$$\begin{aligned} w_{opt} &= \arg\max_{w} T \\ &= \arg\max_{w} \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} C_{1i} + \frac{1}{n}\sum_{i=1}^{n} C_{2i}\right) \end{aligned} \tag{3.4}$$

where C_{1i} is the correlation between the ith band of the original spectral image and the ith band of the fused image, and C_{2i} is the correlation between the panchromatic band and the ith band of the fused image. Despite the restrictions on the use of correlations as quality indices, this criterion is chosen here since (i) it synthesizes both spectral and spatial qualities, and (ii) the spectral and spatial qualities influence the criterion in the same way. Of course, one may argue about this specific choice, but it is worth noting that making a different choice would not affect the general method of the BDF approach as presented above. In order to assess the benefit of using this optimization criterion, we compared its results to those based on visual optimization and on the non-weighted version.

3.6 Results

Using the quantitative criterion T specified by Eq. 3.4, the computation of w_{opt} yielded different values for each study area (Table 3.2). Fig. 3.3 shows the changes in (i) the global criterion T, (ii) the spatial quality criterion and (iii) the spectral quality criterion as w is altered. As expected, the spatial quality criterion increases with w, while spectral quality criterion slowly decreases. This is consistent with visual interpretation. It can also be seen that BDF with the weighting parameter $w \in [0.4, 1[$ provides satisfactory results with respect to T. Although

the values of T are quite similar for this interval of w values, differences may be found between the results since (i) high values of w will lead to sharper details whereas (ii) low values of w will lead to a better conservation of color.

Table 3.2: Optimal values for w in various image contexts

	Urban	**Forest**	**Agricultural**	**Mediterranean**
w_{opt}	0.65	0.65	0.99	0.55

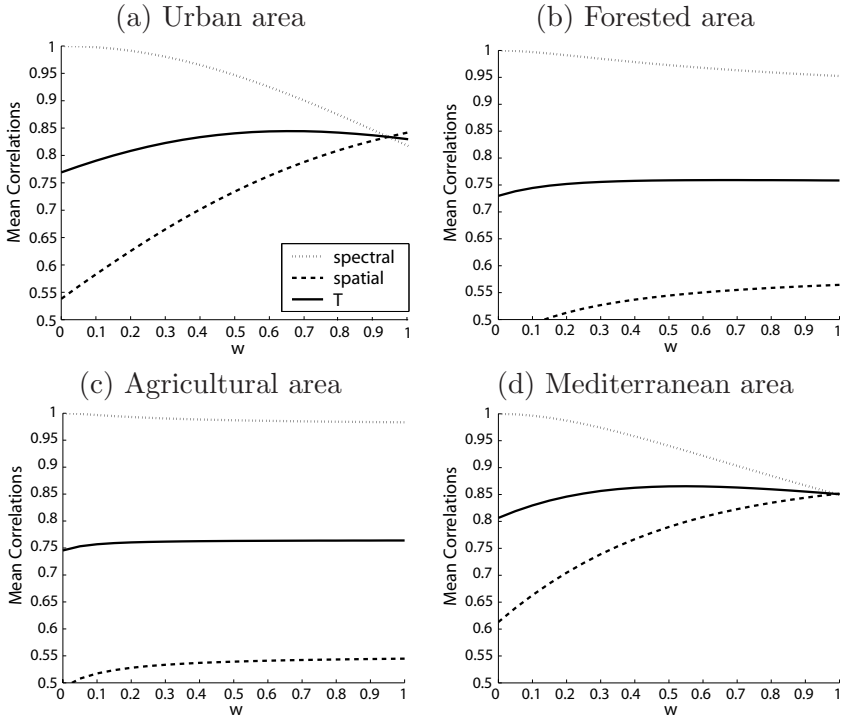


Figure 3.3: Changes in the mean spectral correlation, mean spatial correlation and criterion T (their mean) as the weighting parameter w changes.

As shown in Fig. 3.4, the correlations with the original spectral bands were above 0.9 for every study area. These results were at least as good as wavelet based fusion. This shows that the optimized BDF method performs properly on all the subsets selected in terms of spectral quality. This observation is confirmed by the Q_4 metric computed on 32×32

blocks as recommended by Alparone *et al.* (2004) (see Table 3.3). On this criterion, the BDF method outperformed all the other methods except the IHS fusion method for the densely forested image. This will be investigated below.

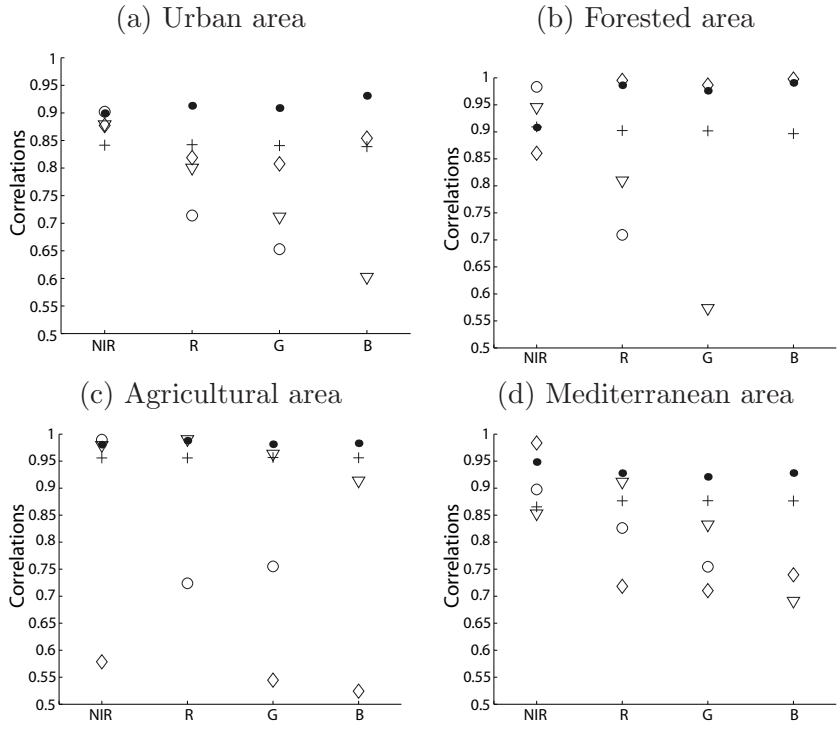


Figure 3.4: Spectral correlations between optimized Bayesian Data Fusion (•), wavelet fusion (+), IHS fusion (∘), PCA fusion (◊) and Brovey fusion (▽) and the spectral bands (i.e. Near-InfraRed, Red, Green and Blue). Correlation values below 0.5 are not shown in these figures.

Fig. 3.4 also shows that wavelet and Bayesian fusion were spectrally robust whatever the main land cover or the spectral bands. However, specific color distortions occurred for other methods. PCA fusion was found to be very sensitive to the context : its results ranged from large color distortions on open Mediterranean landscape (see Fig. 3.5b) to very good results on the densely forested area. IHS fusion exhibited low color distortions in the near-infrared, but its results in visible bands were sometimes inconsistent, especially in the blue band. As an extreme case,

Table 3.3: Q_4 metric computed for the different fusion methods

	Urban	**Forest**	**Agricul.**	**Mediter.**
IHS	0.73	0.79	0.69	0.73
PCA	0.78	0.63	0.70	0.77
Brovey	0.72	0.65	0.74	0.74
Wavelet	0.78	0.67	0.56	0.75
BDF w_{opt}	**0.86**	**0.67**	**0.76**	**0.86**

the IHS fused blue band in the densely forested image was negatively correlated with the original blue band (although this was not penalized in the Q_4 metric). Brovey fusion was more robust than IHS and PCA, but the blue band values were overestimated and thus weakly correlated with the original image. This led to the occurrence of blue areas where dark bodies were expected, in a similar way to the borders of the shaded areas in Fig. 3.5g.

Due to the spectral inconsistencies of the PCA, IHS and Brovey methods, and despite their ability to handle spatial information (see Table 3.4), the spatial quality assessment below will be mainly restricted to BDF and wavelet methods. In terms of spatial quality, it was clear that all the methods tested improved the local contrast on the fused image (see Fig. 3.5 and Table 3.4). Roads, trees, houses and other details were successfully added to the fused image. For the BDF and wavelet fusions, context had more influence on the spatial criteria than on the spectral criteria. However, all the values were higher than 0.89 and neither of these two methods clearly outperformed the other (see Table 3.4). Nevertheless, it appears that BDF fusion did not suffer from artifacts generation, unlike wavelets. Indeed, as mentioned in (Hirschmugl *et al.*, 2005), the profiles of roads may be noisy in wavelets' fused images (see Fig. 3.5).

If one were mainly concerned about visually better spatial quality, parameter w could be adjusted by visual interpretation. In that case, choosing w equal to 0.99 (i.e. giving the largest weight to the panchromatic information without completely deleting the contribution of the original image) noticeably improved image sharpness over all the study areas without leading to major color distortions. The color conservation of such visually adjusted images is illustrated in Fig. 3.5e. On the other hand, emphasis can be put on spectral quality. In this case, w could be set to 0.5 (i.e. equal weighting) or less. Obviously, improving spec-

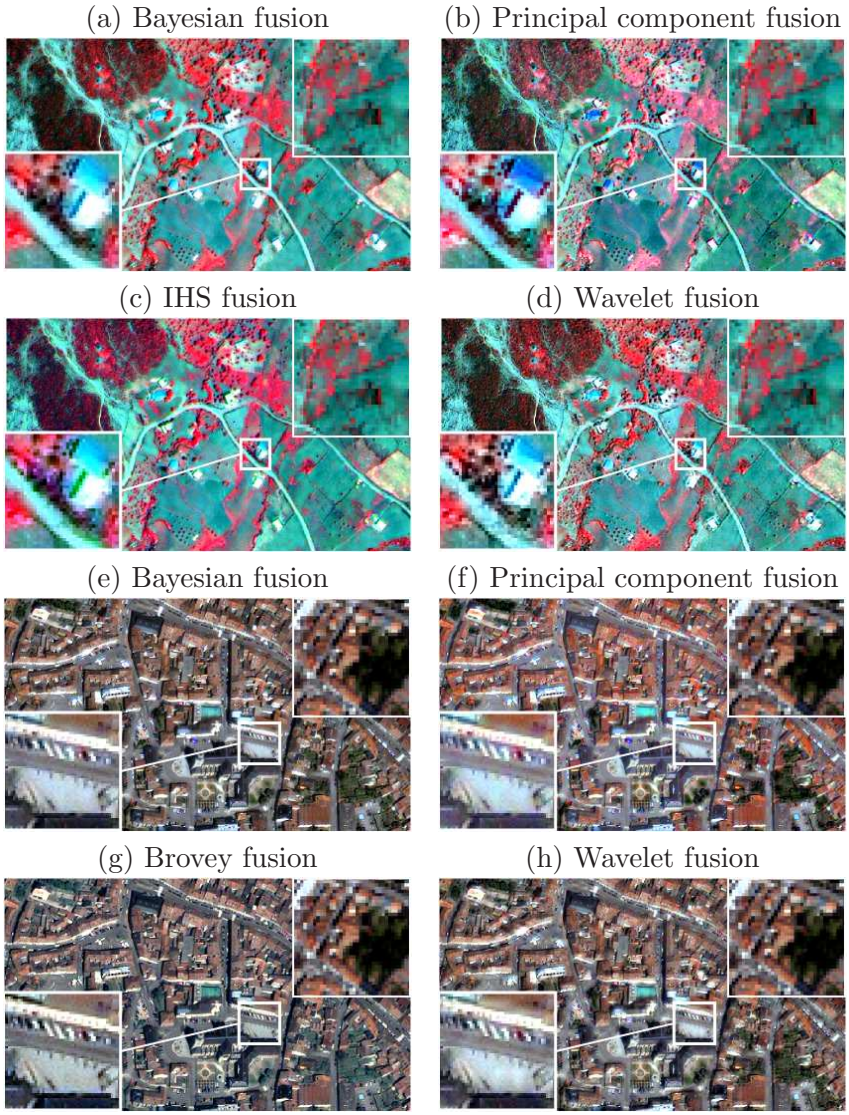


Figure 3.5: Results of the fusion represented in false colors for the Mediterranean area and in real colors for the urban area. In each figure, left boxes correspond to a zoom whereas right boxes replace the fused image by the original color image. Parameter w for BDF has been (i) optimized according to quantitative criterion given in Eq. 3.4, with $w = 0.55$, for the false colors result and (ii) visually adjusted for image sharpness, with $w = 0.99$, for the real colors result.

Table 3.4: Correlations between the original panchromatic band and the intensity of fused images

	Urban	**Forest**	**Agricul.**	**Mediter.**
Resampled	0.70	0.87	0.85	0.77
IHS	≈ 1	≈ 1	0.99	≈ 1
PCA	0.99	0.96	0.95	0.95
Brovey	1	1	1	1
Wavelet	0.95	0.98	0.89	0.94
BDF w_{opt}	**0.95**	**0.99**	**0.91**	**0.96**

tral quality reduces spatial quality, and *vice versa* (see also Fig. 3.3). However, as shown in Fig. 3.6 by comparison with wavelet fusion, both spectral and spatial quality remained consistent for these choices.

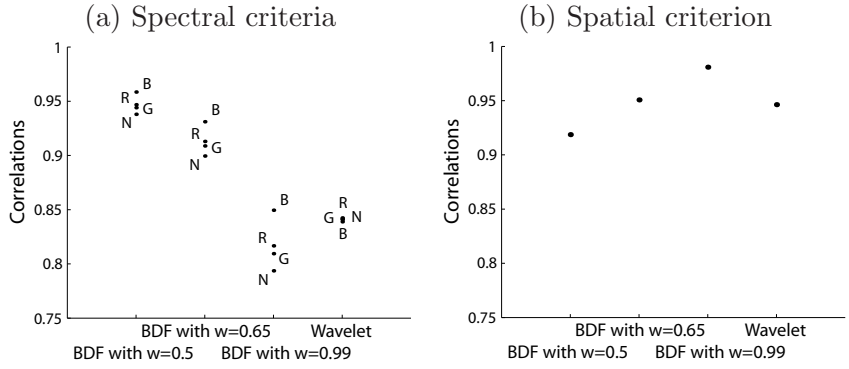


Figure 3.6: Comparison between wavelet fusion, BDF using equal weighting ($w = 0.5$), BDF using $w_{opt} = 0.65$ and BDF using $w = 0.99$ based on (a) spectral and (b) spatial criteria in the urban area. Symbols N, R, G and B stand for Near-Infrared, Red, Green and Blue bands, respectively.

3.7 Discussion and conclusions

The results demonstrate the ability of BDF to address IKONOS image pansharpening, which is particularly difficult due to the fact that the blue band is outside the spectral range of the panchromatic band. It is a

novel approach, although related methods have already been successfully applied in a range of different contexts.

One of BDF's most appealing features is its ability to use an optional weighting parameter for balancing spectral and spatial information, thus enhancing the versatility of the method depending on the context and users' needs. For example, photo-interpreters may wish to favor image sharpness with a weighting parameter close to 1, while automated procedures may require a better color consistency and thus a lower weight for panchromatic information. Considering the overall image quality with respect to w in the present IKONOS case study, we would suggest w_{opt} for optimal fusion with respect to a chosen quantitative criterion and using $w = 0.99$ for improving image sharpness from a qualitative viewpoint.

Correlation coefficients are at the basis of the quantitative quality criteria that were used here. Of course, other quality criteria could have been used instead, leading to different w_{opt} values. However, this choice proved to be consistent with the results of visual assessments of spectral quality, although it was less satisfactory for spatial quality, as it did not account for the presence of spatial artifacts due to amplified local contrast. Visual expert-based assessments (Laporterie-Déjean *et al.*, 2005) might still be useful to provide pansharpened images for photo-interpretation. The fact that the relative ranking of the various methods may differ with respect to either a quantitative criterion or a visual assessment also suggests that better quality indices should probably be developed. It also suggests that an adaptable method is more relevant, since there is probably no unique and universal optimal solution. The tuning ability of BDF pansharpening thus deserves to be emphasized in this respect.

A second appealing feature of the BDF method, freely accessible within the ORFEO Tool Box (ORFEO Accompaniment Program, 2007), is its speed, comparable to the wavelet-based approach when optimization is used for w. The most time-consuming (but optional) part of the process is the optimization of the weighting factor, so that BDF is less demanding than wavelets if optimization is not required (either because w is chosen from a visual assessment or because it can be guessed for a particular study). It is also worth noting that correlation coefficients, regression coefficients and covariance matrix do not need to be estimated using the entire image: using one or more representative sub-samples (in this case a set of 50,000 randomly chosen pixels) proved to be satisfactory for the fusion (results not shown here). However, the differences between the covariance matrices observed on the four image subsets indicate that the fusion process could be further improved if the land cover

classes and their spectral signatures were known *a priori*. Even so, these differences had little impact on the high overall quality of the globally optimized BDF.

Finally, although the BDF method was tested here with specific assumptions and for the specific case of IKONOS image pansharpening, there are no intrinsic limitations on the type of data to be processed or the number of bands to be merged. It could be used, for example, (i) for optical/SAR image fusion, because of its ability to handle noise in the process, and (ii) for hyperspectral fusion, because it can account for the relationships between an arbitrary and potentially large number of spectral bands. For all these reasons, it is our belief that the BDF method is very promising, although further work should be devoted to its application to other types of sensors and studies.

Chapter 4

Support-based implementation of Bayesian data fusion for spatial enhancement : Applications to ASTER thermal images

The previous chapter showed how Bayesian data fusion could be efficiently applied to the image pansharpening issue in the case of IKONOS images. In this chapter, we adapt the Bayesian data fusion approach to the case of a spatial enhancement of remotely sensed images in general. The case of ASTER thermal images of the Lausane region (Switzerland) was chosen as illustration of the method.

4.1 Outline

In this chapter, a general Bayesian Data Fusion (BDF) approach is proposed and applied to the spatial enhancement of ASTER thermal images. This method fuses information coming from the visible/near infrared bands (15×15 meter pixels) with the thermal infrared bands (90×90 meter pixels) by explicitly accounting for the change of support. By relying on linear multivariate regression assumptions, differences of support size for input images can be explicitly accounted for. Due to the use of locally varying variances, the BDF also avoids producing artefacts on the fused images. Based on a set of ASTER images over the Lausanne region (Switzerland), the advantages of this support-based approach are

assessed and compared to the Downscaling Cokriging approach recently proposed in the literature. Results show that improvements are substantial both with respect to visual and quantitative criteria. Though the method is illustrated here with a specific case study, it is versatile enough to be applied to the spatial enhancement problem in general. It thus opens new avenues in the context of remotely sensed images [1].

[1]

4.2 Introduction

Fusion methods for spatial enhancement or pansharpening methods are a set of techniques that aim at improving the spatial resolution of remotely sensed multispectral images (Varshney, 1997; Pohl and van Genderen, 1998; Simone *et al.*, 2002). Usually, information coming from a high spatial resolution panchromatic image is used to increase the spatial resolution of multispectral images. A wide range of methods can be found in the literature: the IHS (Chen *et al.*, 2003), PCA (Zhang, 1999), Wavelets (Zhou *et al.*, 1998) and recently Downscaling Cokriging (DSCK) (Pardo-Iguzquiza *et al.*, 2006) methods are the most frequently used and their performances have been compared in several papers (e.g. Chavez *et al.*, 1991; Laporterie-Déjean *et al.*, 2005). Most often, these methods are applied for pansharpening visible or near-infrared (VNIR) multispectral images using panchromatic images. As the spatial improvement is obtained by using images having similar or close range of wavelengths compared to the multispectral ones, they are expected to be highly correlated.

A less common problem is the spatial enhancement of multispectral images covering the thermal infrared (TIR) wavelengths. Thermal imagery provides information about surface temperature and may prove to be useful in, e.g. forest fires survey because thermal bands are less affected by smoke than visible bands, or archeology for the detection of buried structures (Rejas Ayuga *et al.*, 2006). Unfortunately, TIR images are often associated with a very poor spatial resolution compared to visible and near infrared ones. For instance, the ASTER sensor provides three bands in the VNIR wavelengths at 15 meters resolution, six short wave bands (SWIR) at 30 meters resolution, and five bands in the TIR wavelengths (ranging from 8.125 to 11.65 μm) at 90 meters resolution only. In addition to this poor spatial resolution, VNIR and TIR spectral bands do not share the same range of wavelengths, thus preventing the direct use of classical pansharpening methods when spatial enhancement is needed.

Some methods aiming at the spatial enhancement of images when spectra do not overlap can be found in the literature (Aiazzi *et al.*, 2005; Nishii *et al.*, 1996; Hardie *et al.*, 2004). Among them, a Bayesian approach that relies on the maximization of an *a posteriori* distribution computed using both the raw coarser images and a linear model accounting for the finer resolution image has been proposed in (Hardie *et al.*, 2004). As this method has the advantage of explicitly dealing with the difference of support between images, it proved to be efficient though its assessment was solely based on a synthetic case study. Recently, a

Bayesian data fusion (BDF) framework was proposed by Bogaert and Fasbender (2007). Initially developed in a spatial prediction context, it also provides a consistent framework for fusing an arbitrary large number of information sources that are related to a same variable of interest. Among other applications, it has already been successfully applied to pansharpening issues (Fasbender *et al.*, 2008b, 2007) and a first pointwise implementation was addressed to the spatial enhancement of TIR images (Tuia *et al.*, 2007). It is also worth noting that these fusion approaches have similarities with data fusion of super-resolution imaging (see e.g. Rajan and Chaudhuri, 2002; Fransens *et al.*, 2004)

In this chapter, an extension of the BDF approach is proposed to the spatial enhancement of ASTER images. The general BDF framework will be briefly presented in the context of the spatial enhancement of TIR bands. More specifically, the method will be illustrated by a case study of TIR bands enhancement using VNIR bands from an ASTER image over the region of Lausanne, Switzerland. It will be shown how results of Hardie *et al.* (2004) can be viewed as a particular case of the BDF methodology. Results will be compared to those obtained by DSCK (Pardo-Iguzquiza *et al.*, 2006) using both visual and quantitative criteria. The advantage of using a support-based implementation will be illustrated and finally the overall possibilities offered by the BDF approach for spatial enhancement will be discussed.

4.3 Bayesian data fusion for TIR bands

Combining different sources of information into a single final result (i.e. data fusion) is a problem of general concern for a large panel of applications, going far beyond satellite imagery and encompassing a wide array of potential methods. Among them, Bayesian approaches have led to interesting applications responding to various 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), image processing (Rajan and Chaudhuri, 2002), classification of remote sensing images (Simone *et al.*, 2002; Bruzzone *et al.*, 2002) and environmental modelling (Wikle *et al.*, 2001; Christakos, 2002), just to quote few. The main advantage of a Bayesian approach is that it sets the problem of data fusion in a clear probabilistic framework. The present chapter relies on a general Bayesian Data Fusion approach in the context of spatial data (Bogaert and Fasbender, 2007). Its specific implementation will focus here on the problem of change of support and

on the local adaptability to secondary information sources for predicting a variable of interest (section 4.3.2).

4.3.1 General formulation

The basic concept of BDF as presented in (Bogaert and Fasbender, 2007) relies on the idea that variables of interest, denoted as vector $\mathbf{Z}$, cannot be directly observed. Instead, they are linked to the observable variables $\mathbf{Y}$ through an error-like model, with

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

where $g(\mathbf{Z})$ is a set of functionals and $\mathbf{E}$ is a vector of random errors that are stochastically independent from $\mathbf{Z}$. Assuming that the E_j's of the random vector $\mathbf{E}$ are stochastically independent, it is easy to obtain the conditional probability density function (pdf) of the vector of interest given the observed variables, with

$$f(\mathbf{z}|\mathbf{y}) \;\propto\; f_{\mathbf{Z}}(\mathbf{z}) \prod_{j=1}^{n} f_{E_j}(y_j - g_j(z_j)) \tag{4.1}$$

where j corresponds to the channel number (see Bogaert and Fasbender, 2007 for more details).

Using Bayes theorem again for $f_{E_j}(y_j - g_j(z_j)) = f(y_j|z_j)$, Eq. 4.1 becomes

$$f(\mathbf{z}|\mathbf{y}) \;\propto\; f_{\mathbf{Z}}(\mathbf{z}) \prod_{j=1}^{n} \frac{f(z_j|y_j)}{f(z_j)} \tag{4.2}$$

where $f(z_j)$ is the *a priori* distribution of Z_j. According to the information sources at hand, intermediate stages between Eq. 4.1 and Eq. 4.2 are possible as well (e.g. an expression mixing both $f_{E_j}(.)$ and $f(.|y_j)$ distributions).

In our context of spatial enhancement of remotely sensed images, there are two sources of information, namely the VNIR image (high spatial resolution) and the coarser TIR image. In our implementation, the following notations will be used :

- $\mathbf{Z}(\mathbf{x})$ as the unknown TIR pixel on the high spatial resolution image at location $\mathbf{x}$
- $\mathbf{Y}_{TIR}(\mathbf{x})$ as the TIR pixel on the low spatial resolution image at location $\mathbf{x}$, where image has been resampled using a nearest neighbor method
- $\mathbf{Y}_{VNIR}(\mathbf{x})$ as the visible and near-infrared pixel on the high spatial resolution image at location $\mathbf{x}$

4.3.2 Support-based implementation

A convenient way to account for the change of support consists in associating with the jth TIR band the N high resolution pixels $Z_j(\mathbf{x})$ and $\mathbf{Y}_{VNIR}(\mathbf{x})$ that corresponds to the same low resolution pixel $Y_{TIR,j}(\mathbf{x})$. For this purpose, $\mathbf{Y}$ is defined such that the first element is $Y_{TIR,j}(\mathbf{x})$ and the last $N \times 3$ elements are the $Y_{VNIR}(\mathbf{x})$ pixels. In the case of ASTER images, N will be equal to 36 as there are 36 pixels at 15m resolution into a single 90m resolution pixel. The following relations will be assumed in order to account for both sources of information

$$Y_{TIR,j}(\mathbf{x}) = \frac{1}{||A||} \sum_{\mathbf{x}' \in A} Z_j(\mathbf{x}') + E_{TIR,j}(\mathbf{x}) \tag{4.3}$$

$$Z_j(\mathbf{x}) = \alpha_j + \boldsymbol{\beta}_j' \mathbf{Y}_{VNIR}(\mathbf{x}) + E_{VNIR,j}(\mathbf{x}) \tag{4.4}$$

where A refers to a particular low spatial resolution pixel (i.e. 90m pixel in the case of ASTER TIR images), $E_{TIR,j}(\mathbf{x})$ are zero-mean Gaussian errors with variances equal to $\sigma^2_{TIR,j}$, and $E_{VNIR,j}(\mathbf{x})$ are zero-mean Gaussian errors with variances equal to $\sigma^2_{VNIR,j}(\mathbf{x})$. Eq. 4.4 corresponds to a linear multivariate regression model linking each $Z_j(\mathbf{x})$ to the $\mathbf{Y}_{VNIR}(\mathbf{x})$, where α_j and $\boldsymbol{\beta}_j$ are respectively the intercept and the slope of this model. In order to estimate $\sigma^2_{VNIR,j}(\mathbf{x})$ for each $\mathbf{x}$, a set of P random perturbations $\boldsymbol{\varepsilon}_i$ are simulated (i.e. Monte Carlo simulations; in this chapter, $P = 100$ was used) and an estimator $\widehat{\sigma}^2_{VNIR,j}(\mathbf{x})$ of $\sigma^2_{VNIR,j}(\mathbf{x})$ is provided by

$$\begin{aligned} \widehat{\sigma}^2_{VNIR,j}(\mathbf{x}) &= \widehat{Var}\left[E_{VNIR}(\mathbf{x})\right] \\ &= \frac{1}{P} \sum_{i=1}^{P} \left(y_{TIR,j}(\mathbf{x}) - \alpha_j - \boldsymbol{\beta}_j' \left[\mathbf{y}_{VNIR}(\mathbf{x}) + \boldsymbol{\varepsilon}_i\right] \right)^2 \end{aligned}$$

According to these assumptions and as suggested before, a mixed version of Eq. 4.1 and Eq. 4.2 is proposed here for each coarse pixel and each TIR band j, with

$$f(\mathbf{z}_j | y_{TIR,j}, \mathbf{y}_{VNIR}) \propto f_{\mathbf{Z}_j}(\mathbf{z}_j) \frac{f(\mathbf{z}_j | \mathbf{y}_{VNIR})}{f_{\mathbf{Z}_j}(\mathbf{z}_j)} \tag{4.5}$$

$$\times \; f_{E_{TIR,j}}\left(y_{TIR,j} - \tfrac{1}{N}\mathbf{1}'\mathbf{z}_j\right)$$

$$= f(\mathbf{z}_j | \mathbf{y}_{VNIR}) \tag{4.6}$$

$$\times \; f_{E_{TIR,j}}\left(y_{TIR,j} - \tfrac{1}{N}\mathbf{1}'\mathbf{z}_j\right)$$

where the components of vector $\mathbf{z}_j$ are the unknown pixel values of the jth TIR band that corresponds to the same coarse pixel at 90m, and where $\mathbf{1} = (1 \dots 1)'$ with same length as $\mathbf{z}_j$.

Furthermore, as Gaussian distributions are assumed (which is in a good agreement with the observed distributions in the application presented in Section 4.4), the final distribution $f(\mathbf{z}_j|y_{TIR,j}, \mathbf{y}_{VNIR})$ as given by Eq. 4.6 is Gaussian too, with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$ given by

$$\begin{cases} \boldsymbol{\Sigma}_j &= \left(\frac{1}{N^2\sigma^2_{TIR,j}} \mathbf{1}\mathbf{1}' + \boldsymbol{\Sigma}^{-1}_{VNIR} \right)^{-1} \\ \boldsymbol{\mu}_j &= \boldsymbol{\Sigma}_j \left(\frac{y_{TIR,j}}{N\sigma^2_{TIR,j}} \mathbf{1} + \boldsymbol{\Sigma}^{-1}_{VNIR} \left(\alpha_j \mathbf{1} + \boldsymbol{\beta}'_j \mathbf{y}_{VNIR} \right) \right) \end{cases} \tag{4.7}$$

where $\boldsymbol{\Sigma}_{VNIR} = \begin{pmatrix} \ddots & & \mathbf{0} \\ & \sigma^2_{VNIR,j}(\mathbf{x}) & \\ \mathbf{0} & & \ddots \end{pmatrix}$, $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

and where all $\mathbf{x}$'s refer to the same coarse pixel. It is clear that mean $\boldsymbol{\mu}_j$ is a relevant candidate for the estimation of $\mathbf{Z}_j|y_{TIR,j}, \mathbf{y}_{VNIR}$ (i.e. the 36 unknown pixels within the same coarse pixel at 90m).

Using this set of specific assumptions, a comparison between this formulation and Hardie *et al.* (2004) shows that they are equivalent. However, Hardie *et al.* (2004) obtained these results for the case where only two information sources are to be accounted for, whereas Bogaert and Fasbender (2007) showed that BDF is relevant whatever the number of information sources to be fused. It is thus a more general approach of the problem, as other information sources like, e.g. a Digital Elevation Model, a land use map and/or an occupation map, could be accounted for as well whenever they are relevant for the specific problem at hand.

4.4 Application to the ASTER sensors

Support-based BDF has been applied to a set of ASTER thermal bands covering the region of Western Switzerland on August 21, 2005 (see Figure 4.1d). Images were pre-processed in Erdas Imagine 8.7 and BDF algorithms were implemented in MATLAB 6.5. The 8-bit format images were reprojected into the Swiss grid coordinate system (CH-1903).

In order to adapt the coarser bands to the 15m resolution required for computation, nearest neighbor resampling has been used. The resulting multispectral image is thus composed by 8 images that correspond to the spectral bands in the VNIR and TIR as explained in Section 4.3.2.

4.5 Quality assessment

For assessing the global agreement between the fused and the original TIR images, several spectral quality criteria were chosen, namely the coherence between images (see e.g. Pardo-Iguzquiza *et al.*, 2006), the ERGAS (see e.g. Nencini *et al.*, 2007), the SAM (in degrees; see e.g. Nencini *et al.*, 2007) and the Mutual Information (MI; see e.g. Qu *et al.*, 2002). Fusion results should show high values of both coherence (close to 1) and MI and low values (close to 0) of both ERGAS and SAM to ensure that the TIR information source was properly taken into account. It is worth noting that the coherence is the only criterion that properly accounts for the change of support (from 90m to 15m in this study case).

For assessing the agreement between the fused and 15m original VNIR images, only two quality criteria were chosen, namely the correlation coefficients between images and the Mutual Information (MI; see e.g. Qu *et al.*, 2002). Fusion results should get high values of these criteria (close to 1 for the correlation coefficients) to ensure that the VNIR information was properly taken into account.

Finally, a visual comparison was realized within two sub-regions of 150x150 pixels (see Fig. 4.1d). Such a comparison allows to assess the local agreement between observed features on the original and fused TIR images.

4.6 Results and discussion

Results of the support-based BDF are shown for the TIR 1 band in Fig. 4.1a whereas Fig. 4.1b shows these results for the DSCK algorithm. For comparison purposes, Fig. 4.1c and Fig. 4.1d show the original TIR 1 and VNIR 1 bands respectively.

Direct visual comparison (see Fig. 4.1) shows a good agreement between the original TIR image and the fused bands for both DSCK and support-based BDF implementations, DSCK seems to better reproduce the original TIR image. This is confirmed by the quality criteria as shown in Table 4.1: ERGAS, SAM and MI show the high similitude between the original TIR image and both algorithms but DSCK is related to slightly higher values. This is in agreement with the observations done by visual comparison. Finally, the coherence criterion indicates that both algorithms account properly with the change of support between TIR and VNIR images.

On the other side, if DSCK seems to outperform BDF in terms of reconstruction of the original TIR pattern, agreement with the VNIR

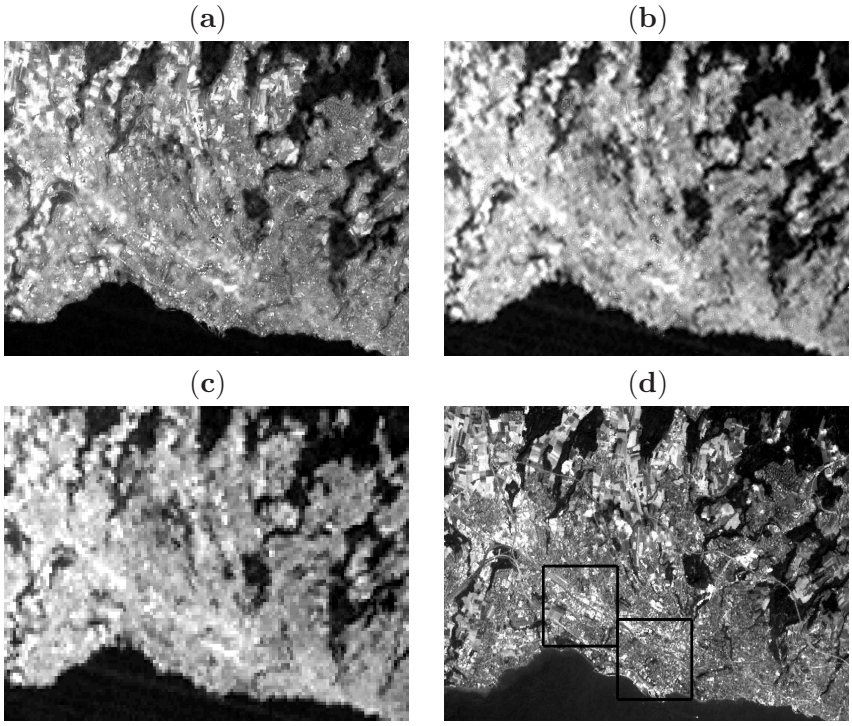


Figure 4.1: (**a**) Results on the Lausanne case study of support-based BDF (band TIR 1). (**b**) Results of DSCK (band TIR 1). (**c**) Original band TIR 1. (**d**) Original band VNIR 1. Black squares correspond to subregions highlighted by Fig. 4.2.

Table 4.1: Comparison between BDF and DSCK algorithms using spectral and spatial quality criteria.

	Spectral Quality			
	Coherence	SAM(deg)	ERGAS	MI
BDF	0.96	11.46	0.65	0.69
DSCK	0.99	1.72	0.20	0.99

	Spatial Quality					
	C_1	C_2	C_3	MI_1	MI_2	MI_3
BDF	0.80	0.84	0.31	0.50	0.42	0.20
DSCK	0.70	0.73	0.23	0.38	0.30	0.20

bands shows that BDF integrates more spatial information in the fused image (see Fig. 4.1 and Table 4.1). Spatial details are much sharper in Fig. 4.1a than in Fig. 4.1b where results remain rather blurry with only few details added from the VNIR information source.

The fact that support-based BDF better exploits the VNIR information source is linked to the locally adaptive properties of the method. For example, the use of locally estimated variances enables us to limit the amount of VNIR information that will be accounted for when regression yields poor results in Eq. 4.4 (i.e. when VNIR values are high in presence of low TIR values): in this case, the influence of the VNIR bands in the fusion process decreases, so that values for the 90m TIR band is dominating the result. As a consequence, it also avoids creating artifacts on the fused image.

Finally, a visual comparison for subregions of the image (Fig. 4.2) highlights the amount of information from the VNIR image. While DSCK does not enhance significantly the spatial behavior of the TIR image, support-based BDF correctly inserts patterns from the VNIR image in the fusion while preserving the TIR information source.

4.7 Conclusions

In this chapter, the general BDF algorithm as proposed by Bogaert and Fasbender (2007) and used by Fasbender *et al.* (2008b) for image pan-sharpening was extended to the spatial enhancement of thermal (TIR) bands from multispectral and multiresolution images. The algorithm proposed by Hardie *et al.* (2004) can be viewed as a particular case of it. It has been shown here how BDF can be used to infer the conditional

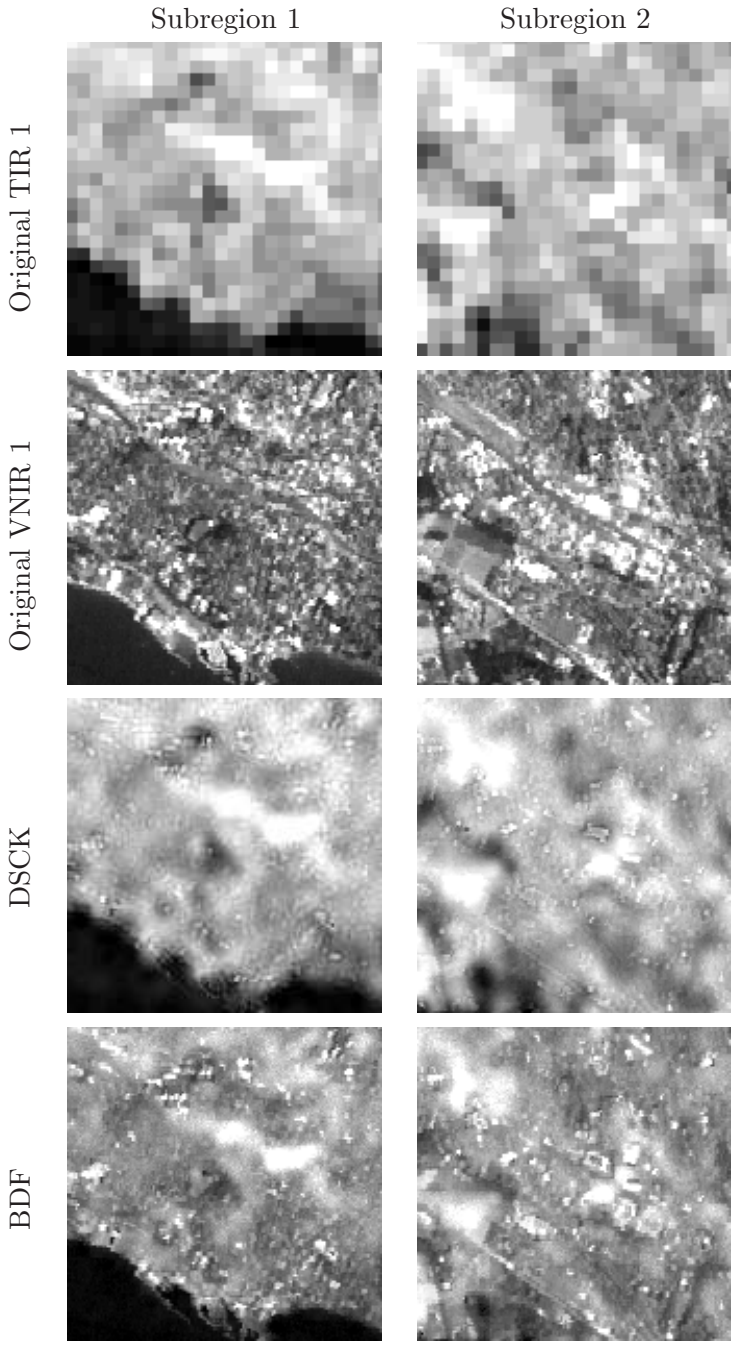


Figure 4.2: Comparison of the results of BDF and DSCK algorithms for both subregions highlighted in Fig. 4.1d.

probability distribution of a spatially enhanced image by accounting for a change of support through local linear multivariate regression, thus allowing the user to take advantage of spatial information as related to high spatial resolution VNIR bands. By considering the original TIR pixels (90m resolution) as the mean value of the pixels of the fused image located within a (6×6) window, the fused image preserves the local spectral values of the original image while at the same time adding spatial details as coming from the 15m VNIR bands.

Application to an ASTER image of Lausanne (Western Switzerland) illustrated the excellent performance of this support-based algorithm. Comparisons with the DSCK method as proposed in (Pardo-Iguzquiza *et al.*, 2006) showed that, even if performing similarly in terms of reproduction of the original TIR pattern, BDF allows a better integration of spatial details and avoids the artifacts and blurry results obtained with the DSCK.

Though the implementation of BDF we proposed here was dedicated to the spatial sharpening of thermal images, BDF has also been applied to image pansharpening using visible bands with the same benefit (Fasbender *et al.*, 2008b), as well as it has already been successfully applied to totally different issues (Bogaert and Fasbender, 2007). Modelling assumptions can be changed according to the problem, the objectives and the information at hand. BDF methodology is thus versatile enough to be used in a wide array of situations without major modifications. More specifically, for remote sensing applications, it thus opens new avenues for important issues like change of support, fusion of multiple spatial information sources or even multitemporal images processing.

Chapter 5

Bayesian data fusion applied to water table spatial mapping

The two previous chapters dealt with applications of the Bayesian data fusion approach to classical remote sensing issues. In this chapter, we apply the Bayesian data fusion approach to the spatial interpolation of a water table. The case of the Dijle basin in Belgium was chosen in order to illustrate the method.

5.1 Outline

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 chapter, 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 [1].

[1] This chapter is an adaptation of a paper accepted in Water Resources Research (Fasbender *et al.*, 2008a). Reproduced by permission of American Geophysical Union.

5.2 Introduction

Water table elevations can be directly obtained from piezometric heads measurements at wells and boreholes locations. Unfortunately, for most survey studies due to the associated costs, the number of these locations are mostly 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 emphasized again 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 categor-

ical 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 chapter, 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.

5.3 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.*, 2008b, 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 a straightforward way 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.

5.3.1 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

$$\mathbf{Y} \;=\; g(\mathbf{Z}) + \mathbf{E} \tag{5.1}$$

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

$$f(\mathbf{z}|\mathbf{y}) \;\propto\; f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}}(\mathbf{y} - \mathbf{g}(\mathbf{z})) \tag{5.2}$$

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

$$f(z_0|\mathbf{z}_S, \mathbf{y}) \;\propto\; \int f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}}(\mathbf{y} - g(z)) \mathrm{d}\mathbf{z}_U \tag{5.3}$$

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

$$f(z_0|\mathbf{z}_S,\mathbf{y}) \propto \prod_{i=0}^{m} f_{E_i}(y_i - g_i(z_i)) \times \int f_{\mathbf{Z}}(\mathbf{z}) \prod_{j=m+1}^{m+n} f_{E_j}(y_j - g_j(z_j))\mathrm{d}\mathbf{z}_U \quad (5.4)$$

$$\propto \prod_{i=0}^{m} \frac{f(z_i|y_i)}{f(z_i)} \int f_{\mathbf{Z}}(\mathbf{z}) \prod_{j=m+1}^{m+n} \frac{f(z_j|y_j)}{f(z_j)}\mathrm{d}\mathbf{z}_U \quad (5.5)$$

where Eqs. 5.4 and 5.5 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. 5.4 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 theoretically sound in that case as it is sufficient to remove the corresponding conditional pdf in Eq. 5.4 (resp. in Eq. 5.5). Particularly, if all conditional distributions at unsampled locations are not taken into account, the integrals in Eq. 5.4 and 5.5 reduce 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-related ones).

5.3.2 Specific assumptions

Though Eqs. 5.4 and 5.5 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. 5.5.

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

$$Z(\mathbf{x}_i) \;=\; DEM(\mathbf{x}_i) - g\Big(\, d_{DEM}(\mathbf{x}_i)\Big) + E(\mathbf{x}_i) \tag{5.6}$$

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 5.4.2 for a concrete example) and $E(\mathbf{x}_i)$ is a zero-mean random error with a variance $\sigma^2_E(\mathbf{x}_i)$ 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 (see Fig. 5.1 for an illustration of this behavior). 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 $\sigma^2_E(\mathbf{x}_i)$ 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. 5.5 (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

$$f(z_0|\mathbf{z}_S, DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0)) \propto \frac{f(z_0|\mathbf{z}_S)}{f(z_0)} f(z_0|DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0)) \tag{5.7}$$

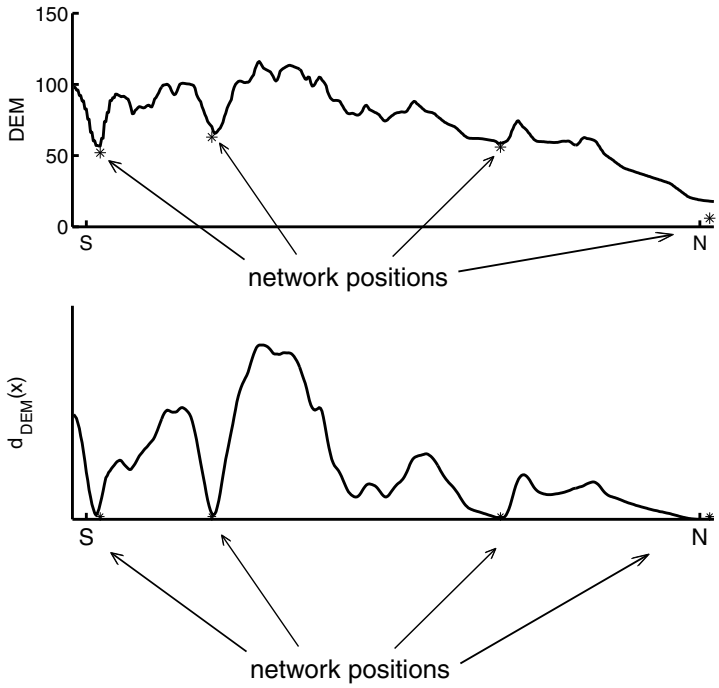


Figure 5.1: Illustration of the behavior of distance $d_{DEM}(\mathbf{x}_i)$ with respect to the topography imposed by the DEM values. The curves are profiles of the DEM values (upper graph) and of the distance $d_{DEM}(\mathbf{x}_i)$ (lower graph).

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 D for details)

$$\begin{cases} \mu_P &= \left(\frac{\mu_k}{\sigma_k^2} + \frac{\mu_d}{\sigma_E^2} - \frac{\mu_0}{\sigma_0^2}\right)\sigma_P^2 \\ \sigma_P^2 &= \left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_E^2} - \frac{1}{\sigma_0^2}\right)^{-1} \end{cases} \tag{5.8}$$

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).

5.4 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, with the Brabant Massif at the bottom 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.

5.4.1 Data

Through the monitoring network of the Flemish Region (Databank Ondergrond Vlaanderen; http://dov.vlaanderen.be), piezometric data where gathered from 135 locations in the Dijle basin (see Fig. 5.2). 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. 5.2 and 5.3). The histograms of these raw piezometric head measurements and of the corresponding DEM values do not show particular shape (see Fig. 5.4) so that no transformation (e.g. Box-Cox) can be applied here in order to correct the non-Gaussianity of the measurements

5.4.2 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. 5.5). As suggested by Fig. 5.5 and alike Linde *et al.* (2007), a spherical model was chosen with theoretical variance and range equal to 300 m^2 and 13 700 m, respectively. This high range value is reflecting a strong spa-

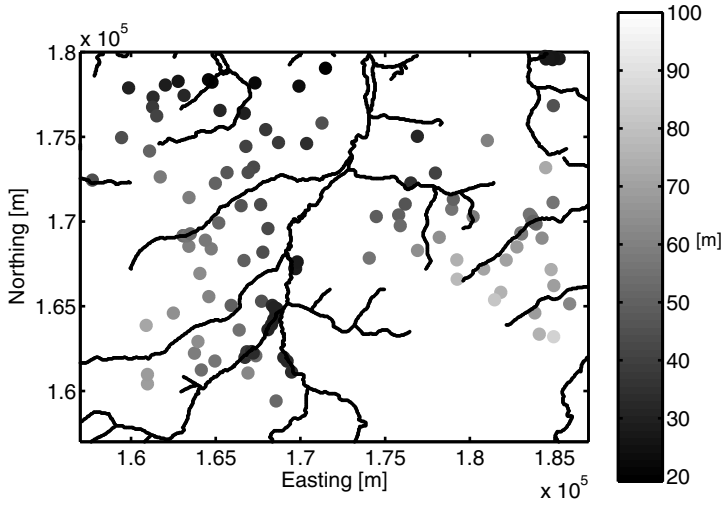


Figure 5.2: Sampled locations of the 135 piezometric heads values, with corresponding elevations above sea-level (in meters) as represented by color.

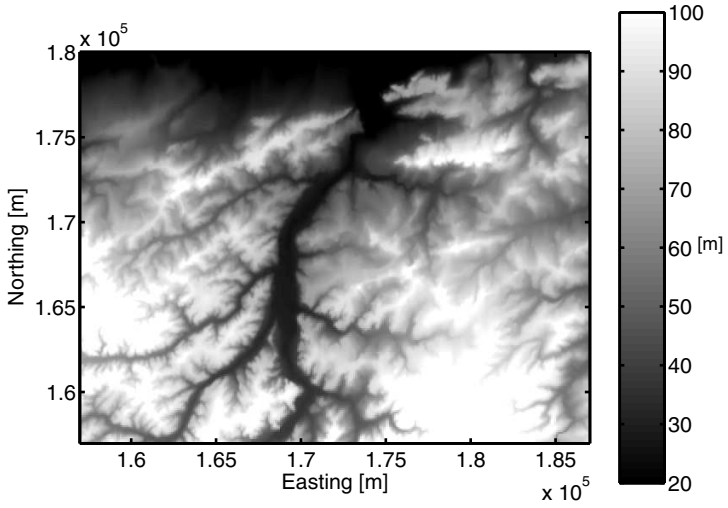


Figure 5.3: Digital Elevation Model of the study area (in meters above sea-level).

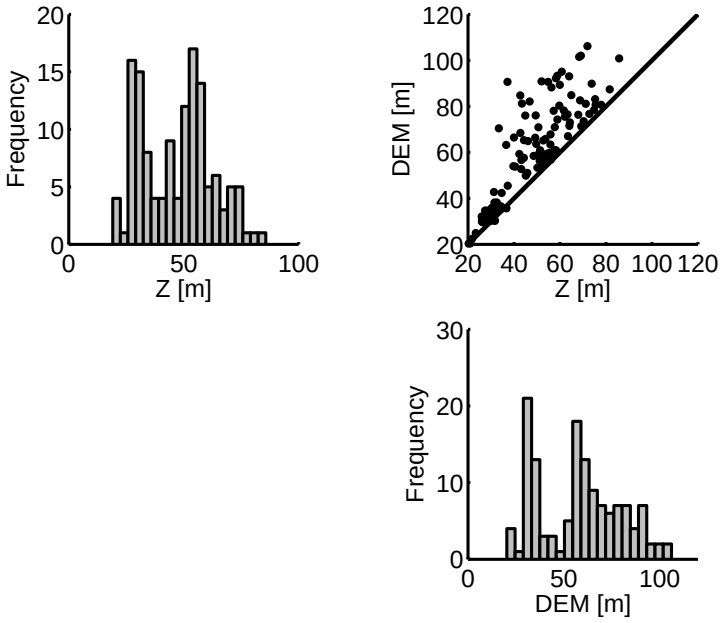


Figure 5.4: Histograms of the raw piezometric head measurements (upper left graph) and of the DEM values (lower right graph). Upper right graph represents the scatterplot between the variables.

tial 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 5.4). 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. Moreover, one could argue the choice of the spherical model for the semi-variogram of the water table as it is a non-differentiable model. However, as the results obtained when using the spherical model was satisfying, this model was kept in place of a differentiable model (e.g. Gaussian model).

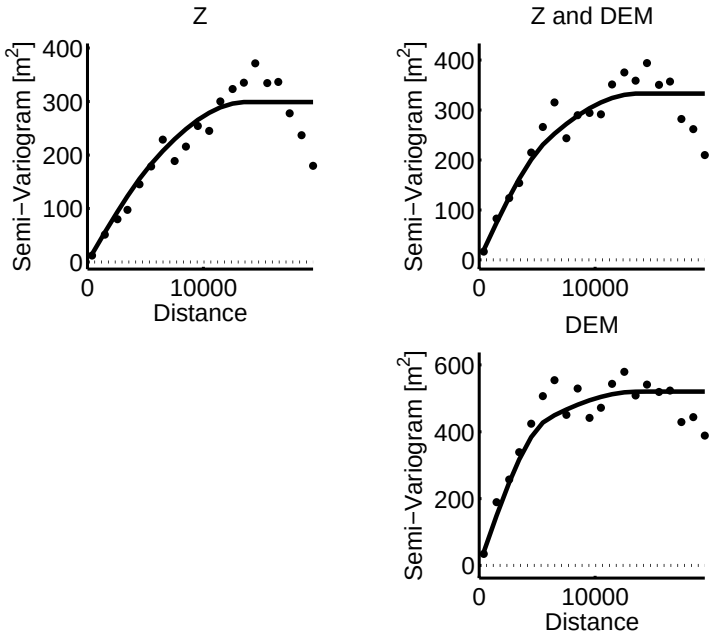


Figure 5.5: Experimental and modelled spatial semi-variogram based on the 135 raw piezometric measurements and on the corresponding DEM values.

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. 5.6; 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 6.3 m. 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.

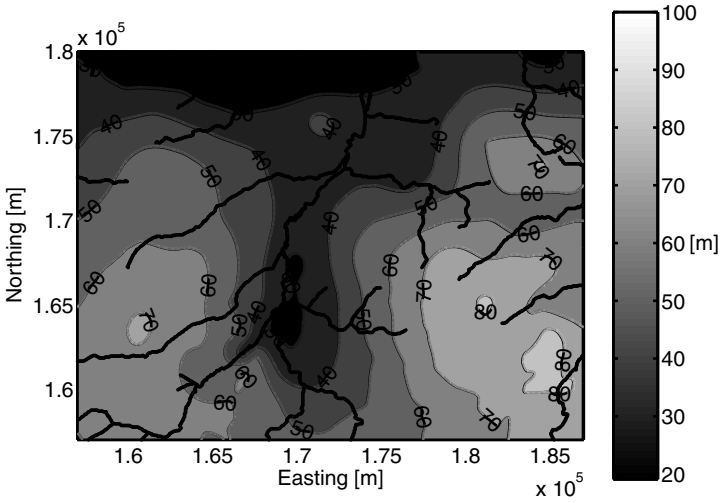


Figure 5.6: Prediction of the water table using ordinary kriging. The gray scale convention is the same as in Fig. 5.2. Bold lines represent the river network over the study area.

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 (lower graph in Fig. 5.5). 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 (13 700 m and 6 000 m) 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. 5.7), 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 6.4 m 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.

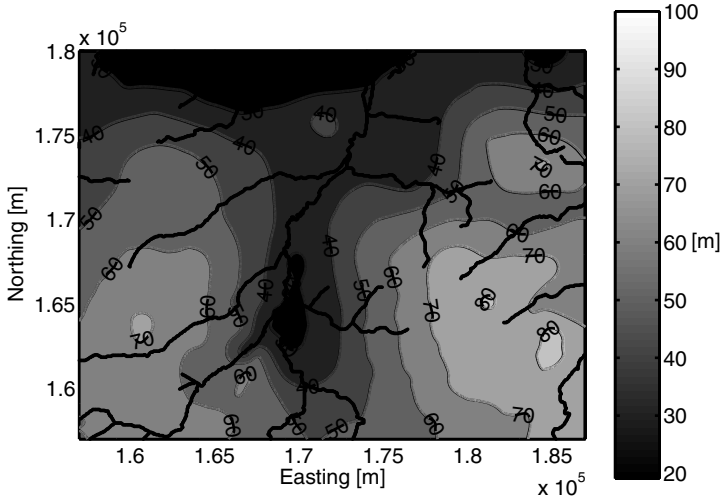


Figure 5.7: Prediction of the water table using ordinary cokriging. The grey scale convention is the same as in Fig. 5.2. Bold lines represent the river network over the study area.

In order to implement the BDF approach, a penalized distance $d_{DEM}(\mathbf{x}_i)$ (Section 5.3.2) 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. 5.8 and upper right graph in Fig. 5.4)

$$DEM(\mathbf{x}_i) - Z(\mathbf{x}_i)) = g\left(d_{DEM}(\mathbf{x}_i)\right) + E(\mathbf{x}_i) \tag{5.9}$$

A logistic-like functional $g(.)$ with equation

$$g(h) \;=\; a + b\left(1 - exp\left(\frac{-3h^2}{c^2}\right)\right) \tag{5.10}$$

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.

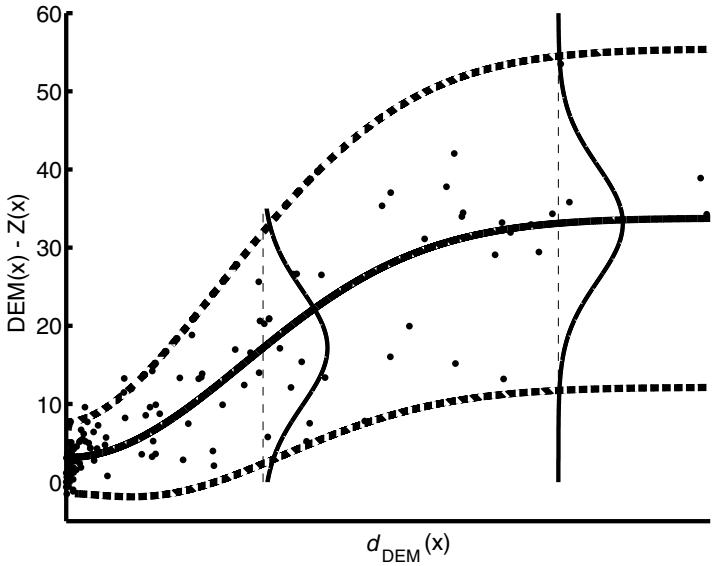


Figure 5.8: Graph of the groundwater depth $DEM(\mathbf{x}) - Z(\mathbf{x})$ as a function of the penalized distance $d_{DEM}(\mathbf{x})$ to the network. Dots represents the observed pair of values, plain line represents the fitted non-linear relationship $g(.)$ whereas dashed lines represent the 95% 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 $d_{DEM}(\mathbf{x})$.

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. 5.7, 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. 5.9, 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. 5.6 and 5.7, 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.

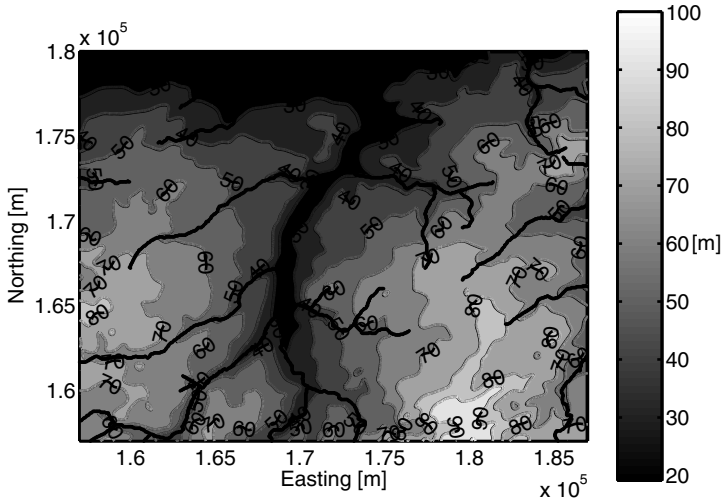


Figure 5.9: Prediction of the water table using the Bayesian data fusion approach. Bold lines represent the river network over the study area.

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 :

$$ME = \frac{1}{N}\sum_{i=1}^{N}\widehat{e}_i \tag{5.11}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\widehat{e}_i| \tag{5.12}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\widehat{e}_i{}^2} \tag{5.13}$$

where $\widehat{e}_i$ is the estimated error at sampled location $\mathbf{x}_i$, with $N = 135$. From Table 5.1 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. 5.10a to 5.10c 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. 5.8.

Table 5.1: 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

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 lack of samples there (i.e. neither network positions nor sample locations). This reflection is totally in accordance with the variances of predictions shown in Figs. 5.10a to 5.10c.

5.5 Discussion and conclusions

In this chapter, 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

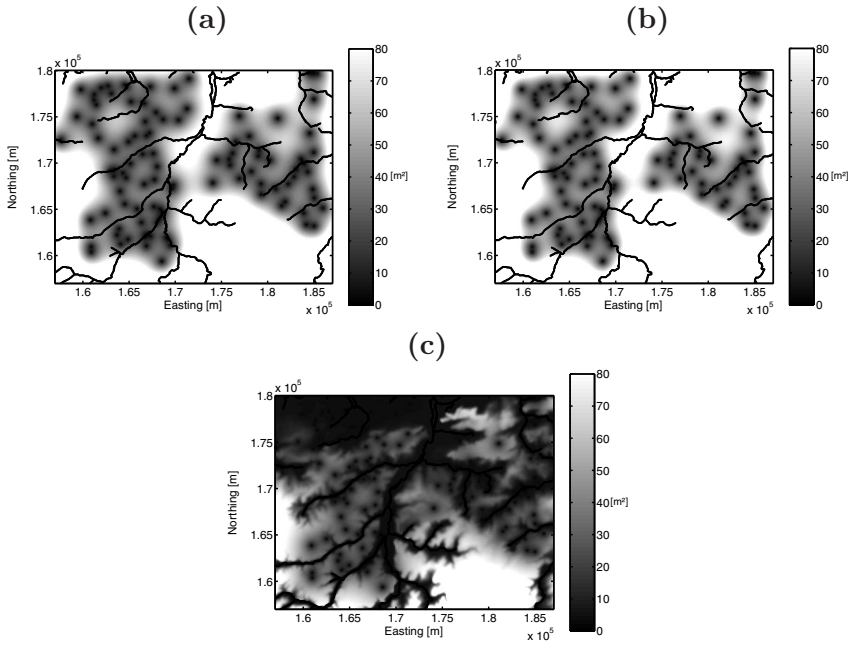


Figure 5.10: Variance of prediction of **(a)** ordinary kriging, **(b)** ordinary cokriging and **(c)** Bayesian data fusion methods. Bold lines represent the river network over the study area.

need of defining a spatial multivariate model, that may be to restrictive or demanding. Though we did not mention it in this chapter, 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 rationale and none of them can be generalized to the case of multiple secondary information sources.

In this chapter, the BDF method was implemented using (multivariate) Gaussian distributions. Besides of being particularly convenient (e.g. analytical expression for the final *posterior* distribution, fast implementation), this choice is also the Maximum Entropy solution when using only the first two moments (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. Bogaert and Fasbender, 2007 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 research at this point. Moreover, considering another distribution would not have induced any change for the general formulation of the BDF approach.

An originality of this work is also that it makes use of a "distance" to a network for defining an empirical 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 depend on 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 also be readily implemented 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.

Chapter 6

Bayesian data fusion for space-time prediction of air pollutants : the case of NO_2 in Belgium

In this last chapter, the Bayesian data fusion approach is extended to a space-time context. The application proposed here is the space-time prediction of air pollutants, illustrated here by the case of NO_2 concentrations in Belgium.

6.1 Outline

In the beginning of the 21st century, it is obvious that health and environmental matters are among the most important political and societal topics. The various kinds of pollution (e.g. air pollution, soil pollution, water contamination) are responsible for significant health and environmental degradation. In order to adequately deal with pollution issues, it is important to better understand the acting processes and to be able to account for specific knowledge about the pollutant. Thanks to this, it will be possible to forecast pollutant concentrations so that efficient actions can be rapidly taken. Based on a Bayesian data fusion (BDF) framework, the present chapter proposes a methodology for air pollutant forecasting using the space-time properties of the process and several secondary information sources that contribute to a better understanding of the pollutant behavior (e.g. meteorological variables and anthropogenic activities). Consequently, the present work can contribute to improving the representation and the forecast of pollutant fields. Moreover the

developed approach also permits to predict the probability of exceeding a given threshold, as required in official regulations for some pollutants (e.g. the European directives). The BDF framework is applied here to the case of space-time predictions of air concentrations of nitrogen dioxide (NO_2) in Belgium. After a detailed description of some specific assumptions, results showed that BDF is able to successfully account for secondary information sources, thus leading to meaningful NO_2 predictions.

6.2 Introduction

Since the industrial revolution, pollution and more specifically air pollution increased steadily first in Europe and in North America and then all over the world. Processes such as the burning of fossil fuels, industrial operations and forest clearing release various gazes into the atmosphere (e.g. carbon dioxide, methane and nitrous oxide), thus affecting air quality and inducing adverse effects on the environment (e.g. acid rains, smog, ozone depletion or global warming) and on public health (e.g. asthma, various types of cancers or other respiratory diseases). With this aim in view, European directives on ambient air quality become more and more stringent. The limit values fixed in these directives appear problematic for some pollutants (e.g. nitrogen dioxide, particulate matter), especially in urban and industrial zones. Consequently, Member States have to take all necessary measures in order to reduce emissions and respect the European norms. When pollution peaks occur, short-term action plans (traffic limitation, ...) have to be activated in order to reduce the risk and the duration of the population exposure to these peaks.

In order to make appropriate decisions, it is crucial to have an accurate knowledge on the pollution state both in space and in time and to be able to rely on sound and efficient models. Among the large set of modeling methods, one usually distinguishes between deterministic and stochastic approaches. Deterministic approaches consist in explicitly modeling the physical and chemical processes governing the spatial and temporal evolutions of pollutant concentrations. In general, such models jointly take into account meteorology (wind, temperature, turbulence,...) and emission sources (e.g. traffic, industry, domestic heating,...). The main disadvantage of deterministic models is the lack of parametrization schemes. Both physical and chemical aspects of the pollutant are very complex and thus are hardly accounted for in those models. In addition, the required computing resources increase significantly with the grid resolution. Stochastic approaches are potential alternatives to deterministic models since they generally require less computing resources than deterministic ones. Space-time statistics (namely geostatistics) have showed their ability to tackle loads of environmental issues within a space-time context : ore estimation (e.g. Krige, 1951; Yamamoto, 1999; Costa, 2003), oil estimation (e.g. Jaquet, 1989; Ren *et al.*, 2006), water table estimation (e.g. Hoeksema *et al.*, 1989; Linde *et al.*, 2007; Fasbender *et al.*, 2008a), soil map estimation (e.g. D'Or *et al.*, 2001), health (e.g. Christakos and Vyas, 1998b; Christakos and Kolovos, 1999) or air pollution (e.g. Christakos and Vyas, 1998a; Chris-

takos and Serre, 2000; Christakos *et al.*, 2001) are possible applications of geostatistics in environmental sciences.

Recently, Bogaert and Fasbender (2007) proposed a Bayesian data fusion (BDF) framework that aims at reconciling various secondary information about a same variable of interest in order to provide a unique spatial prediction. This BDF has already been successfully applied to various environmental sciences applications, namely, fusion of remotely sensed images (Fasbender *et al.*, 2008b,c) and water table estimation (Fasbender *et al.*, 2008a).

In this chapter, the BDF framework will be extended in a space-time context, with the objective of space-time air pollution predictions in mind. In the second part of the chapter and after a short description of the data set, the case of nitrogen dioxide (NO_2) in Belgium will be tackled. For illustration purpose and for sake of brevity, only Simple Kriging (SK) and BDF predictions will be presented here. Results from both methods will be compared and several validations processes will be performed in order to assess the quality of both predictions. By doing so, the two main objectives of this chapter will be achieved, namely, (i) to show that secondary information sources (such as meteorological variables) are potentially useful for air pollutant predictions and (ii) to show that the BDF framework is one possible way to account for these information sources.

6.3 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 ; it covers a wide variety of potential 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.*, 2008b,c, 2007) and environmental modeling (Wikle *et al.*, 2001; Christakos, 2002; Fasbender *et al.*, 2008a), just to quote a few of them. The main advantage of Bayesian approaches is to set the problem in a proper probabilistic framework. Lately, Bogaert and Fasbender (2007) proposed a general BDF formulation especially designed for spatial prediction problems. These general

results have already proved their relevance in several applications (see Fasbender *et al.*, 2008b,c, 2007, 2008a) and will be applied here for the space-time prediction of air pollutants. For the sake of brevity, it is not possible to present the whole underlying theory, so only theoretical results that are the most relevant for our application will be presented hereafter. The interested reader may refer to Bogaert and Fasbender (2007) for a detailed description of the theory.

6.3.1 General formulation

Let us define $\{\mathbf{x}_0, \ldots, \mathbf{x}_n\}$ as the set of space-time locations (i.e. each $\mathbf{x}_i$ is composed by a spatial component $\boldsymbol{\ell}_i$ and a temporal component t_i) where indirect observations $\mathbf{y}' = (\mathbf{y}'_0, \ldots, \mathbf{y}'_n)$ are available about a variable of interest Z and where $\mathbf{y}_i$ is a vector of n_i observations at same space-time location $\mathbf{x}_i$. 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 aims at reconciling the auxiliary variables $\mathbf{Y}$ to the primary variables $\mathbf{Z}$ through an error-like model, with

$$Y_{i,j} \;=\; g_{i,j}(\mathbf{Z}) + E_{i,j} \tag{6.1}$$

where $g_{i,j}$'s are functionals and where $E_{i,j}$'s are 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

$$f(\mathbf{z}|\mathbf{y}) \;\propto\; f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}}(\mathbf{y} - \mathbf{g}(\mathbf{z})) \tag{6.2}$$

where $f_{\mathbf{Z}}(.)$ is the *a priori* pdf for $\mathbf{Z}$ and $f_{\mathbf{E}}(.)$ is the pdf of the errors $\mathbf{E}' = (\mathbf{E}'_1, \ldots, \mathbf{E}'_n)$. In the context of an air pollutant prediction, one can write that $\mathbf{Z} = (Z_0, \mathbf{Z}'_S, \mathbf{Z}'_U)'$ where Z_0 refers to the pollutant concentration at prediction space-time location $\mathbf{x}_0$, $\mathbf{Z}_S$ refers to space-time locations $\mathbf{x}_S = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ where both Z_i's and $Y_{i,j}$'s are jointly sampled, and $\mathbf{Z}_U$ refers to space-time locations $\mathbf{x}_U = \{\mathbf{x}_{m+1}, \ldots, \mathbf{x}_n\}$ where only $Y_{i,j}$'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

$$f(z_0|\mathbf{z}_S, \mathbf{y}) \;\propto\; \int f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}}(\mathbf{y} - \mathbf{g}(\mathbf{z})) \mathrm{d}\mathbf{z}_U \tag{6.3}$$

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

$$f(z_0|\mathbf{z}_S,\mathbf{y}) \propto \prod_{i=0}^{m}\prod_{j=1}^{n_i} f_{E_{i,j}}(y_{i,j}-g_{i,j}(z_i)) \times \int f_{\mathbf{Z}}(\mathbf{z}) \prod_{i=m+1}^{m+n}\prod_{j=1}^{n_i} f_{E_{i,j}}(y_{i,j}-g_{i,j}(z_i))\mathrm{d}\mathbf{z}_U \quad (6.4)$$

$$\propto \prod_{i=0}^{m}\prod_{j=1}^{n_i} \frac{f(z_i|y_{i,j})}{f(z_i)} \times \int f_{\mathbf{Z}}(\mathbf{z}) \prod_{i=m+1}^{m+n}\prod_{j=1}^{n_i} \frac{f(z_i|y_{i,j})}{f(z_i)}\mathrm{d}\mathbf{z}_U \quad (6.5)$$

where Eqs. 6.4 and 6.5 are completely equivalent expressions as they are linked to each other using Bayes theorem, with

$$f_{E_{i,j}}(y_{i,j}-g_{i,j}(z_i)) = f(y_{i,j}|z_i) \propto \frac{f(z_i|y_{i,j})}{f(z_i)}$$

so that using either distributions of errors $f_{E_{i,j}}(.)$ or conditional distributions $f(z_i|y_{i,j})$ provides two possible way of incorporating different information sources.

As an interpretation and a consequence of the independence error hypothesis, this Bayesian approach separates the problem in two parts. The first one focuses on the space-time 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 $f(z_i|y_{i,j})$ 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 (e.g. Linear Model of Coregionalization in the case of cokriging methods; see e.g. Chilès and Delfiner, 1999). One may argue about the practical relevance of these equations as they rely on this independence hypothesis. However, one can show that, from an entropic viewpoint, this hypothesis corresponds to the minimum loss of information if no precise information about the joint distribution of the auxiliary variables is at hand (again, see Bogaert and Fasbender, 2007 for details about this topic).

6.3.2 Important considerations for air pollutants prediction

A first direct consequence of the important space-time variability of air pollutants is that stationarity hypotheses do not hold here. Indeed,

even first-order stationarity does not hold true since the mean of the concentrations $Z(\boldsymbol{\ell}, t)$ is expected to vary along with location (largely due, but to an unknown extent, to the proximity of emission sources; see Fig. 6.1 in the case of NO_2 in Belgium). Furthermore, the mean of the concentrations is also expected to vary along with time (due to human activities such as transportation and industries; see Fig. 6.2 in the case of NO_2 in Belgium). As a matter of fact, the first issue is to provide a coarse approximation of these local effects. This approximation is then to be used for computing the *prior* distributions $f_{\mathbf{Z}}(\mathbf{z})$ and $f(z_i)$ in Eq. 6.5. In this chapter, for each location, the time-varying mean of air pollutant concentrations $\mu(\boldsymbol{\ell}, t)$ was modeled using the expression

$$\mu(\boldsymbol{\ell}, t) = \mathbb{E}\left[Z(\boldsymbol{\ell}, t)\right] = W(\boldsymbol{\ell}, t) + S(\boldsymbol{\ell}, t) \tag{6.6}$$

where $W(\boldsymbol{\ell}, t)$ is the weekly component (here, the mean with respect to the day of the week) and $S(\boldsymbol{\ell}, t)$ is the annual component (here, a sine curve with 1-year periodicity). Contrary to the space-time mean, the variability $\sigma^2(\boldsymbol{\ell})$ was assumed to be constant in time, i.e.,

$$\sigma^2(\boldsymbol{\ell}) = \mathbb{V}ar\left(Z(\boldsymbol{\ell}, t)\right) = \mathbb{E}\left[\left(Z(\boldsymbol{\ell}, t) - \mu(\boldsymbol{\ell}, t)\right)^2\right], \qquad \forall t \in \mathbb{R} \tag{6.7}$$

One could criticize this stationarity assumption as variances of air pollutant concentrations are expected to be somehow related on other time-dependent phenomena (e.g. human activities). However, this assumption seems to be quite reasonable according to what was observed in the NO_2 application (see Fig. 6.1).

Now using Eq.6.6 and 6.7, it is clear that

$$R(\boldsymbol{\ell}, t) = \frac{Z(\boldsymbol{\ell}, t) - \mu(\boldsymbol{\ell}, t)}{\sigma(\boldsymbol{\ell})} \tag{6.8}$$

has a mean equal to 0 and a variance equal to 1 so that the second-order stationarity hypothesis holds true now. It is therefore easy to infer the space-time dependence using $R(\boldsymbol{\ell}, t)$ instead of using $Z(\boldsymbol{\ell}, t)$. Based on this estimation, the following space-time model (Bogaert, 1996) is fitted

$$\Gamma_R(\mathbf{x}_i, \mathbf{x}_j) = \gamma_{ST}(\Delta_{\boldsymbol{\ell}}, \Delta_t) + \gamma_S(\Delta_{\boldsymbol{\ell}}) + \gamma_T(\Delta_t) \tag{6.9}$$

where $\gamma_{ST}(.,.)$ is a non-separable space-time semi-variogram model, $\gamma_S(.)$ is a purely spatial semi-variogram model and $\gamma_T(.)$ is a purely temporal semi-variogram model, with $\Delta_{\boldsymbol{\ell}} = ||\boldsymbol{\ell}_i - \boldsymbol{\ell}_j||_2$ and $\Delta_t = |t_i - t_j|$. Thanks to this choice for the semi-variogram model, it is possible to account for a change of curvature along with the values of $\Delta_{\boldsymbol{\ell}}$ and Δ_t (such as can

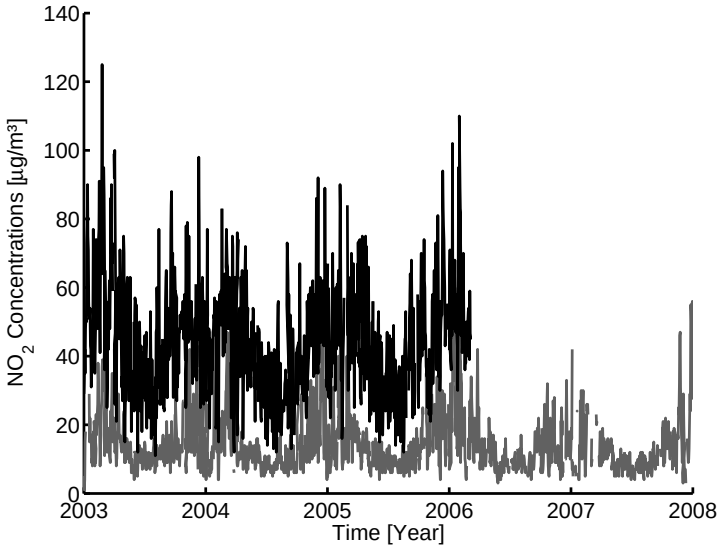


Figure 6.1: Illustration of the non-stationarity in space of the NO_2 concentrations in Belgium. The bold line and dotted line stand for monitoring devices located in the Brussels region (center of Belgium) and in Virton (far South of Belgium) respectively.

be seen in Fig. 6.5 for the NO_2 application). Using this final model, it is easy to compute the covariance for any couple (Z_i, Z_j) as

$$\mathbb{C}ov\left(Z_i, Z_j\right) = \sigma(\boldsymbol{\ell}_i)\sigma(\boldsymbol{\ell}_j) - \sigma(\boldsymbol{\ell}_i)\sigma(\boldsymbol{\ell}_j)\Gamma_R(\mathbf{x}_i, \mathbf{x}_j) \tag{6.10}$$

The *prior* distributions $f(z_i)$ are assumed here to be Gaussian distributions with mean equal to $\mu(\mathbf{x}_i)$ and variance equal to $\sigma(\boldsymbol{\ell}_i)$. Distribution $f_{\mathbf{Z}}(\mathbf{z})$ in Eq. 6.5 are also assumed to be Gaussian with mean vector equal to $(\mu(\mathbf{x}_0), \ldots, \mu(\mathbf{x}_{m+n}))'$ and covariance matrix with elements corresponding to $\mathbb{C}ov\left(Z_i, Z_j\right)$. One can contest the use of Gaussian distributions, but there are arguments in favor of this choice. First, on a theoretical basis, the (multivariate) Gaussian distribution is the Maximum Entropy solution under constraints on the first and second moments (see e.g. Shannon, 1948; Christakos, 1990, 1992 for more details). Moreover, as distributions $f(z_i)$ and $f_{\mathbf{Z}}(\mathbf{z})$ rely on the same moments constraints, the Gaussian distributions require that $f(z_i)$ agrees with $f_{\mathbf{Z}}(\mathbf{z})$ (see e.g. Papoulis, 1991; Sveshnikov, 1979). Finally, from a practical viewpoint, the restriction to Gaussian distributions in Eq. 6.4 and 6.5 proved to be computationally efficient with convincing results (see Fasbender *et al.*, 2008b,c,a).

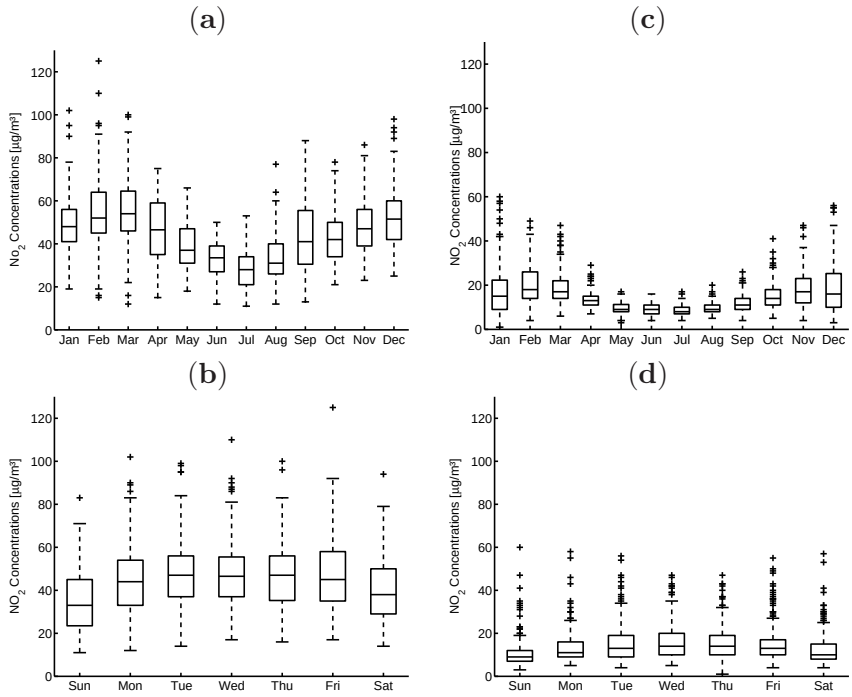


Figure 6.2: Illustration of the seasonal (**a** and **c**) and the weekly (**b** and **d**) effects. Right column and left show the behaviors of the monitoring devices in the Brussels region (same as in Fig. 6.1) and in Virton respectively.

As for the case of $f(z_i)$ with $f_{\mathbf{Z}}(\mathbf{z})$, distributions $f(z_i|y_{i,j})$ should be in accordance with $f(z_i)$. From a theoretical viewpoint, the relation

$$f(z_i) = \int_{\mathbb{R}} f(z_i|y_{i,j}) f(y_{i,j}) dy_{i,j}, \qquad \forall z_i \in \mathbb{R} \tag{6.11}$$

must hold true. As a direct consequence of this, we thus have

$$\mathbb{V}ar(Z_i) = \mathbb{E}_{Y_{i,j}}\left[\mathbb{V}ar_{Z_i}[Z_i|Y_{i,j}]\right] + \mathbb{V}ar_{Y_{i,j}}\left[\mathbb{E}_{Z_i}[Z_i|Y_{i,j}]\right] \tag{6.12}$$

This theoretical constraint might be difficult to satisfy in practice. It is worth noting here that a first consequence of Eq. 6.11 and 6.12 is that if one wants to model $f(z_i|y_{i,j})$ for a given $y_{i,j}$ only with its conditional mean $\mathbb{E}_{Z_i}[Z_i|y_{i,j}]$ and conditional variance $\mathbb{V}ar_{Z_i}[Z_i|y_{i,j}]$ (e.g. if $f(z_i|y_{i,j})$ is assumed to be Gaussian), then both expressions should be estimated in a consistent way. Appendix E shows how a lack of consistency could lead to disastrous results when using Eq.6.5. So, as Gaussian distributions will be used here for aforementioned reasons (see e.g. Bogaert and Fasbender, 2007; Fasbender *et al.*, 2008b,c,a for examples of applications), it is important to control this kind of issues in specific implementations. Appendix E also proposes an easy way that enables us to circumvent this issue in practice.

6.4 Data presentation

The pollution data at hand for this study come from the Belgian regional networks for air quality measurements. For the selected study period, NO_2 was measured at 62 telemetric stations spread unevenly over the three regions of Belgium (half in the Flanders region, a third in the Walloon region and a sixth in the Brussels region; see Fig. 6.3). Though data have been collected since the late seventies, the studied data were restricted to the period between January 1, 2003 and December 31, 2007 (i.e. 1826 days). Observed concentrations are available on a hourly basis but were aggregated as the daily mean for each station. Note that these daily values are computed from hourly concentrations, provided that at least 75% of hourly values are available. Of course, due largely to their respective period of activity, the time series were relatively scarce with unobserved daily values ranging from 30 up to 1501 depending on locations.

Besides the problem of missing values, monitoring devices present very different time series. Indeed, depending mainly on the human activities in their neighborhood and on the fact that the data are measured

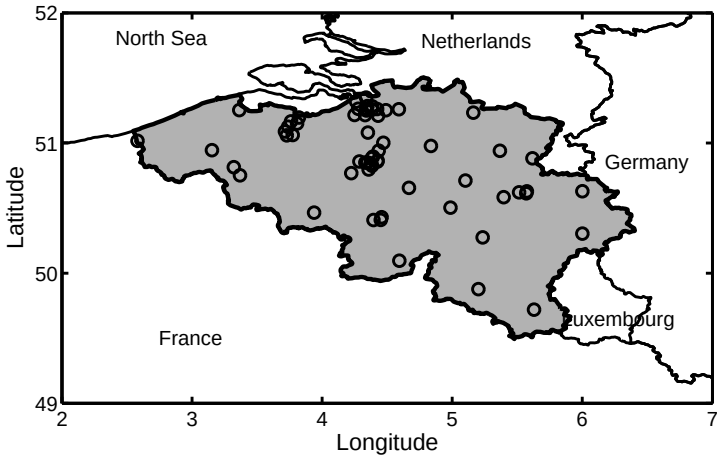


Figure 6.3: Visualization of the 62 sampled locations spread unevenly over Belgium. Approximately half of them are in the Flanders region (Northern part), a third are in the Walloon region (Southern part) and a sixth are in the Brussels region (central part).

with different sensors, time series may differ with respect to (i) their annual means, (ii) their seasonal amplitudes, (iii) their weekly amplitudes and (iv) their variances (see Fig. 6.1 and 6.2).

Broadly speaking, the spatio-temporal evolution of NO2 concentrations mostly depends on (i) emission sources and (ii) meteorological conditions related to the dispersion of pollutants. Concerning the meteorological aspect, three variables have been considered in this study: (i) 10 m above ground wind velocity (noted 10U), (ii) horizontal transport in the boundary layer (noted HT), and (iii) vertically-integrated turbulent kinetic energy (noted TKE). The variables 10U, HT and TKE were obtained or computed using the meteorological fields from ECMWF (European Center for Medium Range Weather Forecast) analyses. 10U allows to characterize the horizontal transport of pollutants in the surface layer. HT is computed as the product of the boundary layer height and the mean wind velocity in the boundary layer. High values of this variable mean an enhanced ability of the atmosphere to disperse pollutants and consequently to reduce concentrations. Compared to 10U, HT represents the horizontal dispersion of pollutants over a wider layer. TKE is based on turbulent kinetic energy and is representative of the intensity of turbulent motions in the atmosphere. This variable is probably the most appropriate to characterize the dispersion of pollutants, since it takes into account horizontal wind shear as well as buoyancy

(i.e. vertical exchanges). Moreover TKE is vertically integrated and therefore takes into account the boundary layer depth.

Having a clear description of the physical state in the troposphere is not sufficient. It is obvious that anthropogenic activities (e.g. transportation, house heating, industries) release pollutants in the air, and thus influence these concentrations. It is therefore useful to account for these information sources whenever they are available.

In this study, three datasets were available for characterizing the emission sources from anthropogenic activities. The first one is based on land cover data. Such data may be useful to better characterize the spatial distribution of NO_2 concentrations. Lately, Janssen *et al.* (2008) proposed an index based on land cover data and showed that there exists interesting relationships between this index and air pollutants such as NO_2. This index, designated by the authors as the β index, is computed using CORINE land cover data on local window with radius equal to 2 kilometers. In this chapter, this β index was computed on a 1001 by 60 regular grid with 0.005 degree between each consecutive nodes (see Fig. 6.4a).

The second dataset is about emissions from the car traffic. Indeed, it is well known that cars reject a variety of gases in the atmosphere (CO_2, SO_2, NO_2, etc) as well as small particles matters. In Belgium, traffic represents about 50% of the emissions of nitrogen oxides. Among the fleet of cars, diesel cars reject 3 to 10 times more NOx than gasoline cars. Further, approximately 20% to 25% of these NOx are in fact NO_2 (Brussels Environment, 2008). This source of information is particularly significant as the proportion of cars working with diesel in Belgium changed approximately from 30% in 1993 to 55% in 2007 (see FEBIAC, 2008 for more details on these statistics). It is also worth noting that, among the 5 millions of cars, approximately 60% of the fleet belong to the Flanders region, 30% to the Walloon region and 10% to the Brussels region. For this study, a shape file with 3 levels (namely freeways, beltways and highways) limited to the Belgian territory was converted to raster files with a spatial resolution of 0.006 degree (see Fig. 6.4b).

The third available dataset for this study was the population density. Humans do not emit NO_2 but population density is correlated with domestic energy consumption (e.g. heating systems) and also emissions from cars. Since the heating systems represents about 36% of nitrogen oxides emissions in large cities such as Brussels (Brussels Environment, 2008), it would be useful to couple population density with temperature fields in order to improve the representation of emissions from heating systems, in particular the fluctuations during wintertime. However, though population density was available at some locations (but not on a

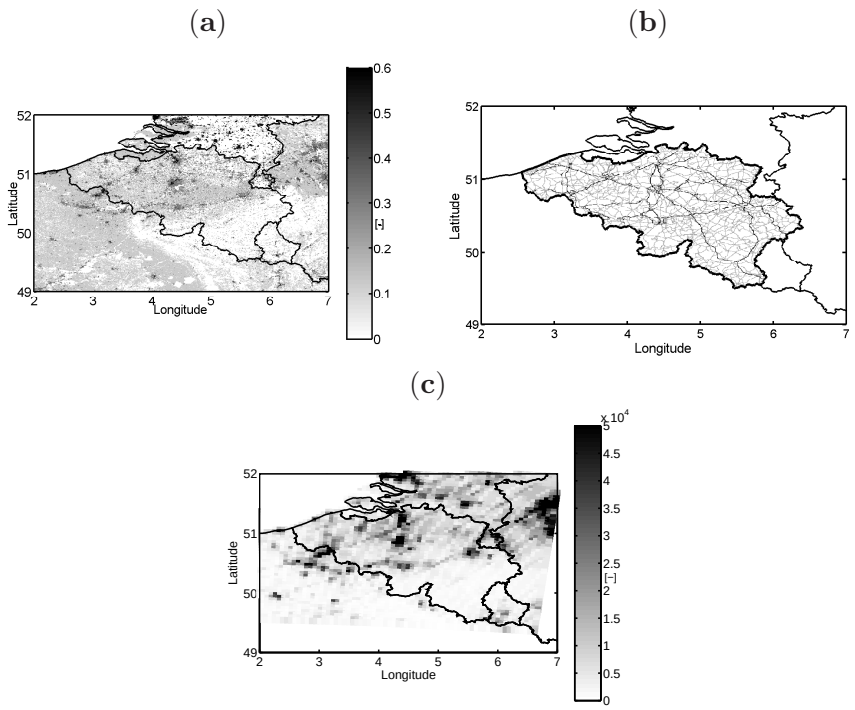


Figure 6.4: Spatial illustration of the information sources related to anthropogenic activities and available for the study : (a) the β index proposed in Janssen *et al.* (2008), (b) the road network and (c) the population density.

regular grid; see Fig. 6.4c), neither temperatures maps nor temperatures time series were known for this study.

Before making predictions, a first issue is the choice of the spatial resolution. As mentioned, the spatial resolutions of the information sources vary from 0.005 to 0.25 deg. Moreover, information sources unfortunately do not share the exact same area. Therefore, the area was chosen as the largest covered area (i.e. the area covered by the meteorological variables) with a spatial resolution equal to $\frac{0.25}{3}$ degree (i.e. an intermediate resolution). Of course, as this resolution did not correspond to any of the other resolution, finer information sources were aggregated (i.e. β index, population density and road network) and coarser information sources were downscaled using a nearest neighbor interpolation algorithm (e.g. for meteorological variables). The resulting prediction locations were thus located at the node of a 39 by 63 regular grid with $\frac{0.25}{3}$ degree between two consecutive nodes.

6.5 Results

As explained in Section 6.3.2, the first issue was to infer the space-time mean $\mu(\boldsymbol{\ell}, t)$ and the spatial variance $\sigma^2(\boldsymbol{\ell})$ from Eq. 6.6 and 6.7. In this chapter, this was achieved in 2 steps. First, for each of the 62 locations of the data set, $\mu(\boldsymbol{\ell}, t)$ was estimated using a linear model with 9 parameters (1 for each of the weekday and 2 for the amplitude of the sine curve), so that

$$\mu(\boldsymbol{\ell}, t) \;=\; \mu_{\boldsymbol{\ell}, d_1(t)} + C_{\boldsymbol{\ell}} \cos\left(\frac{2\pi d_2(t)}{365}\right) + S_{\boldsymbol{\ell}} \sin\left(\frac{2\pi d_2(t)}{365}\right)$$

where $d_1(t)$ and $d_2(t)$ are respectively the day of the week and the number of day in the year for date t (there are thus 7 possible values for $\mu_{\boldsymbol{\ell}, d_1(t)}$). Then, for the estimation of $\mu(\boldsymbol{\ell}, t)$ at other space-time locations, this first estimation was used in a inverse distance interpolation method. Similarly to $\mu(\boldsymbol{\ell}, t)$, the local variance $\sigma^2(\boldsymbol{\ell})$ was first estimated at each of the 62 sampled locations and then interpolated at other space-time locations with the same interpolation method.

Following the reasoning of Section 6.3.2 and using the estimated $\mu(\boldsymbol{\ell}, t)$ and $\sigma^2(\boldsymbol{\ell})$, the original observations were standardized using Eq. 6.8 and a space-time semi-variogram was estimated. Exponential semi-variogram models were chosen for $\gamma_{ST}(.,.)$, $\gamma_S(.)$ and $\gamma_T(.)$ and their parameters were estimated with a least squares minimization algorithm. Fig. 6.5 shows both the estimated semi-variogram and the fitted model.

On the basis of $\mu(\boldsymbol{\ell}, t)$ and $\mathbb{C}ov\,(Z_i, Z_j)$ (see Eq.6.10), a straightforward prediction method is provided by the kriging formalism. Indeed,

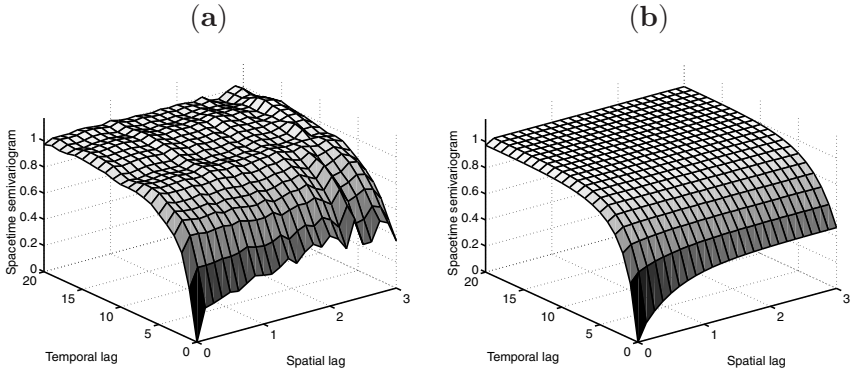


Figure 6.5: Space-time semi-variogram of the standardized values R_i : (**a**) estimation and (**b**) fitted model from Eq. 6.9.

kriging is an easy-to-use linear predictor that only relies on $\mu(\boldsymbol{\ell}, t)$ and $\mathbb{C}ov\,(Z_i, Z_j)$. Kriging is the Best Linear Unbiased Predictor (BLUP) in the sense that it minimizes the variance of prediction among the set of linear predictors (see e.g. Goovaerts, 1997; Cressie, 1991 for more details about the kriging formalism). Therefore, using $\mu(\boldsymbol{\ell}, t)$, $\mathbb{C}ov\,(Z_i, Z_j)$ and observations at dates $\{t, t-1, \ldots, t-4\}$ (i.e. at dates before the predicted day), Simple Kriging (SK) was used on the residuals for the prediction at date $(t+1)$. Predictions were then compared explicitly with the original values observed at $(t+1)$. Fig. 6.6 shows such predictions for several dates (right column). It can be seen that, though quite acceptable in general and correctly modeling the overall space-time process, the kriging predictions exhibit unnatural spatial patterns. This is of course due to the small number of sampled locations. Further, these predictions are often lacking in accuracy (see, e.g., Fig.6.6 on January 17th 2007) due to a lack of exogenous information. Clearly, as already mentioned, air pollutant concentrations such as NO_2 are influenced by (i) the physical state of the troposphere and (ii) the amount of emissions in the area. It might thus be interesting to use these secondary information sources within the BDF framework presented in Section 6.3.

6.5.1 Using physical properties of the troposphere

As mentioned previously, the concentrations of air pollutants are highly influenced by the dynamics of the atmosphere. More specifically the dispersion properties in the boundary layer are mainly influenced by wind intensity and thermal stability of the atmosphere. Though qualitatively understood, quantifying these effects is not straightforward. Examples

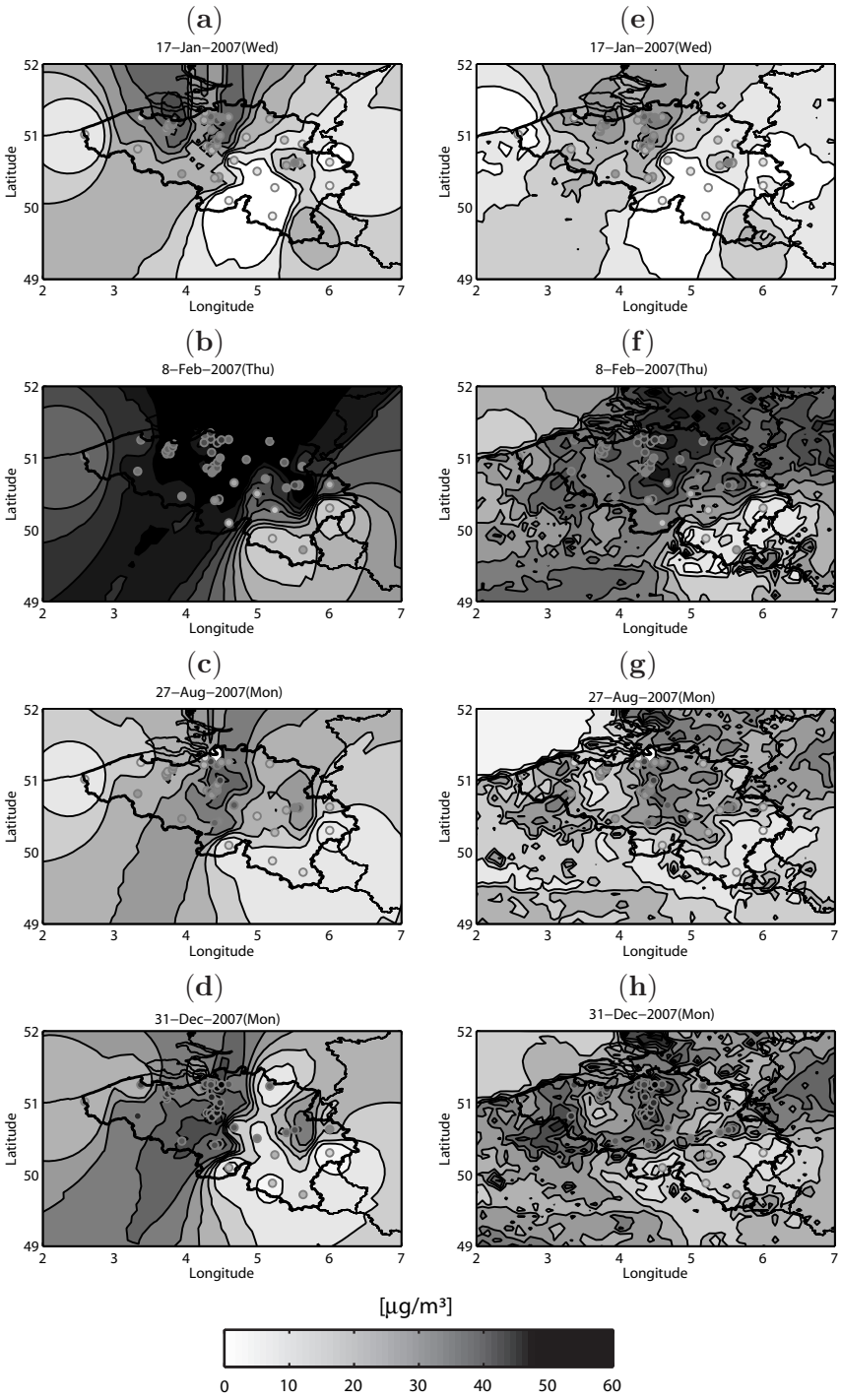


Figure 6.6: Space-time predictions of daily-mean NO_2 concentrations. On the left column are the Simple Kriging predictions whereas Bayesian data fusion ones are on the right column.

Table 6.1: Summary table for the estimated parameters of Eq. 6.13

	α	β_0	β_1	β_2	γ
$10U_i$	-1.28	0.68	-0.01	-0.02	-1.74

of modeling solutions are proposed hereafter in order to explicitly account for selected physical properties of the troposphere.

The first selected variable for characterizing pollutant dispersion is the wind speed at 10 meters height. Indeed, wind speed and air pollutant concentrations are expected to be somehow inversely proportional. However, if increasing wind speed contributes to reduce pollutant concentrations in some circumstances, the amplitude of the process is also influenced by the local emissions. Therefore, the influence of wind speed on pollutant concentrations was modeled using the modeling process proposed in Appendix E. After plotting the scatter plot between wind speed $10U_i$ and NO_2 standardized observations R_i (see Eq. 6.8 for definition), the following empirical assumptions were used

$$\left\{ \begin{array}{rcl} \mu_{R_i|10U_i} & = & \alpha + e^{\beta_0+\beta_1 10U_i+\beta_2 10U_i^2} \\ \sigma^2_{R_i|10U_i} & = & e^{\gamma\mu^2_{R_i|10U_i}} \end{array} \right. \tag{6.13}$$

$\mu_{r_i|10u_i}$ and $\sigma^2_{r_i|10u_i}$ were estimated sequentially based on non-linear least squares minimization algorithm in MATLAB (Matlab, 2002) (see Table 6.1 for the estimated values of these parameters). Fig. 6.7a shows a superposition of the upper-defined scatter plot, the estimation of $\mu_{r_i|10u_i}$ and an estimated 95% region based on estimation of both $\mu_{R_i|10U_i}$ and $\sigma^2_{R_i|10U_i}$ and using a Gaussian assumption for the conditional distribution.

Similarly to wind speed, horizontal transport in the boundary layer (HT) and vertically integrated turbulent kinetic energy (TKE) characterize the efficiency of pollutant dispersion. Again, these variables are thought to be inversely proportional to the concentrations of air pollutants (and also with standardized values R_i). Therefore, the same approach was used for the empirical estimation of their influences on NO_2, i.e.

$$\left\{ \begin{array}{rcl} \mu_{R_i|W_i} & = & a + e^{b_0+b_1 W_i} \\ \sigma^2_{R_i|W_i} & = & e^{c\mu^2_{R_i|W_i}} \end{array} \right. \tag{6.14}$$

where W_i is either equal to HT_i or TKE_i respectively (the parameters of $\mu_{R_i|HT_i}$ and $\sigma^2_{R_i|HT_i}$ being different from those of $\mu_{R_i|TKE_i}$ and $\sigma^2_{R_i|TKE_i}$, of course; see Table 6.2 for the estimated values of these parameters).

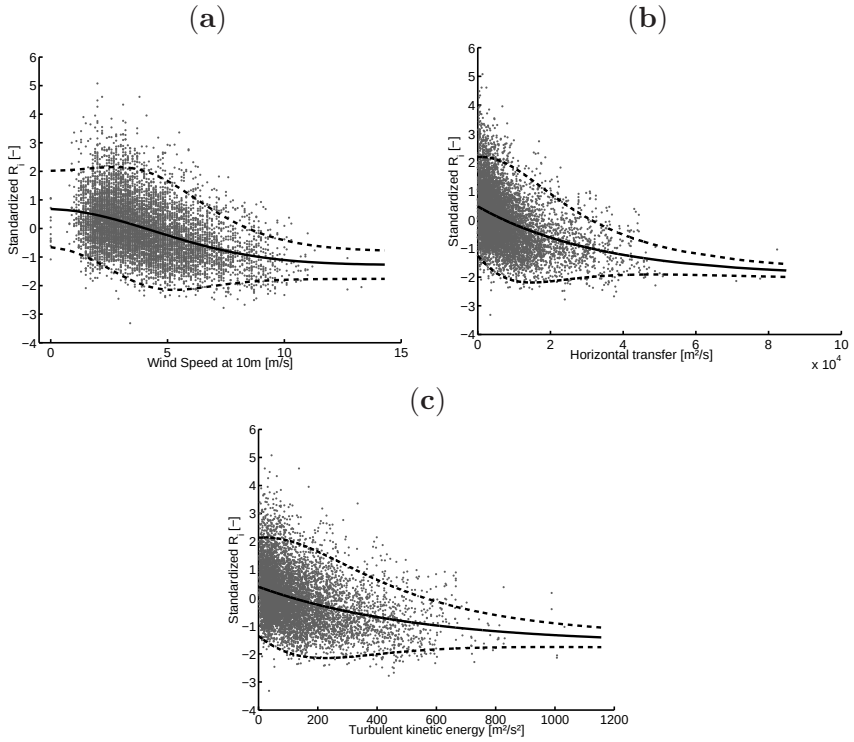


Figure 6.7: Model of the conditional mean and the conditional variance of standardized NO_2 observations with respect to (**a**) wind speed in the first 10 meters, (**b**) horizontal transfer and (**c**) turbulent kinetic energy. Bold line stands for estimated mean and dashed lines represent symmetric 95% intervals.

Table 6.2: Summary table for the estimated parameters of Eq. 6.14

	a	b_0	b_1	c
HT_i	-1.97	0.89	-2×10^{-5}	-1.40
TKE_i	-1.64	0.71	-2×10^{-3}	-1.75

Fig. 6.7b and c shows the resulting models for the horizontal transfer and the turbulent kinetic energy, respectively.

6.5.2 Using anthropogenic activities

Similarly to physical properties of the troposphere, the qualitative effects of indexes related to pollutant emissions (e.g. land use indicators, proximity to road network or population density, etc) are well identified. On the other hand, there is a lack of quantitative characterizations (Janssen *et al.*, 2008). Model hypotheses are proposed next in order to overcome this shortcoming.

As described in Section 6.4 and as proposed by Janssen *et al.* (2008), the β index is an interesting secondary information source for space-time prediction of air pollutant concentrations. Indeed, relying on land cover data, this spatial index represents somehow the human activity in the area. As suggested in Janssen *et al.* (2008), the β index can be used in order to better interpolate mean concentrations (i.e. $\mu(\boldsymbol{\ell}, t)$ in this chapter). Again, the interpolation of $\mu(\boldsymbol{\ell}, t)$ by means of this index has the advantage that the corresponding conditional distribution will be in accordance with the prior pdf $f(z_i)$. To do so, values of $\mu(\boldsymbol{\ell}, t)$ estimated at space-time sampled locations were first used to fit the parameters of a function such as in Eq. 6.14, then the fitted model was applied to the β values at other space-time locations where predictions are sought for. Secondly, conditional variance was estimated using the relation

$$\sigma^2_\beta(\boldsymbol{\ell}, t) = \sigma^2(\boldsymbol{\ell}, t) e^{-a(\mu_\beta(\boldsymbol{\ell},t)-\mu(\boldsymbol{\ell},t))^2} \tag{6.15}$$

with $a > 0$ and where $\mu_\beta(\boldsymbol{\ell}, t)$ is the modeled conditional mean with respect to the β index.

For the case of road network and population density, the approach chosen here was very close to the one of the β index. The main difference was that the models for the conditional means were linear with respect to explanatory variables (multivariate linear in the case of road network as there was three kind of roads). Again, for the conditional variance, Eq. 6.15 was used with another set of parameters for each information source.

Final prediction and validation

Using the six information sources as described in Sections 6.5.1 and 6.5.2 along with Eq. 6.5, predictions were made on a regular grid. Fig. 6.6 (right column) shows these results when using secondary information only at the prediction location. The benefits of these secondary information sources are clear. Contrary to SK predictions and thanks to these information sources, the BDF predictions exhibit relevant local patterns. The effect of anthropogenic information sources are non negligible, especially on the Northern part of the study area where one could explicitly see the Belgian and Dutch coastlines. More generally, one can observe these effects mostly on isolated locations (i.e. for locations situated far away from the sampled ones).

It is worth noting that influences of secondary information vary highly in time and in space. This is of course due to the fact that both the conditional means and the conditional variances were modeled using the space-time mean $\mu(\ell, t)$ that also varies in space and in time. As a consequence, on days characterized by efficient dispersion (e.g. on February 8 2007), concentrations predicted with the BDF approach are significantly lower than the SK ones (see Fig. 6.6b and f). This is even more obvious for week days since their means are generally higher than for weekend days (see Fig. 6.2). In such situations, meteorological data were able to correct the detrimental influence of the *prior* distribution on SK predictions. Furthermore, land-use related indexes provide finer scale information (i.e. 3 times finer than the meteorological variables) so that local patterns (and in particular the differentiation between rural and urban zones) can clearly be seen in the BDF predictions.

Another interesting observation is the fact that the BDF predictions in the adjacent countries (namely France, Netherlands, Luxembourg and Germany) are much more realistic than the SK ones. This is of course due to the fact that SK predictions rely only on data sampled in Belgium and does not account for the secondary information located outside Belgium. Thus, for locations in those countries, secondary information sources are quite influential on the final predictions giving less credit to farther sampled locations. BDF was thus able to account for the secondary information in order to improve the global knowledge at unsampled locations.

Using the SK and BDF daily predictions, it is straightforward to compute annual properties over the study area (see Fig. 6.8a and c). It is clear that both methods provide results in good agreement with the true annual mean concentrations estimated as means from the raw data (filled circles on the maps), though BDF produced more realistic

prediction maps (e.g. see littoral zone and local patterns). However, BDF outperformed SK on the number of days for which daily mean were above $80\mu g/m^3$ (see Fig. 6.8b and d). This threshold has been chosen in order to specifically select situations characterized by a greater exposure to NO_2. Indeed, SK overestimated that number of days at several locations, whereas BDF was much closer to the actual number of high concentration days computed from the raw data (again, see the filled circles on the maps).

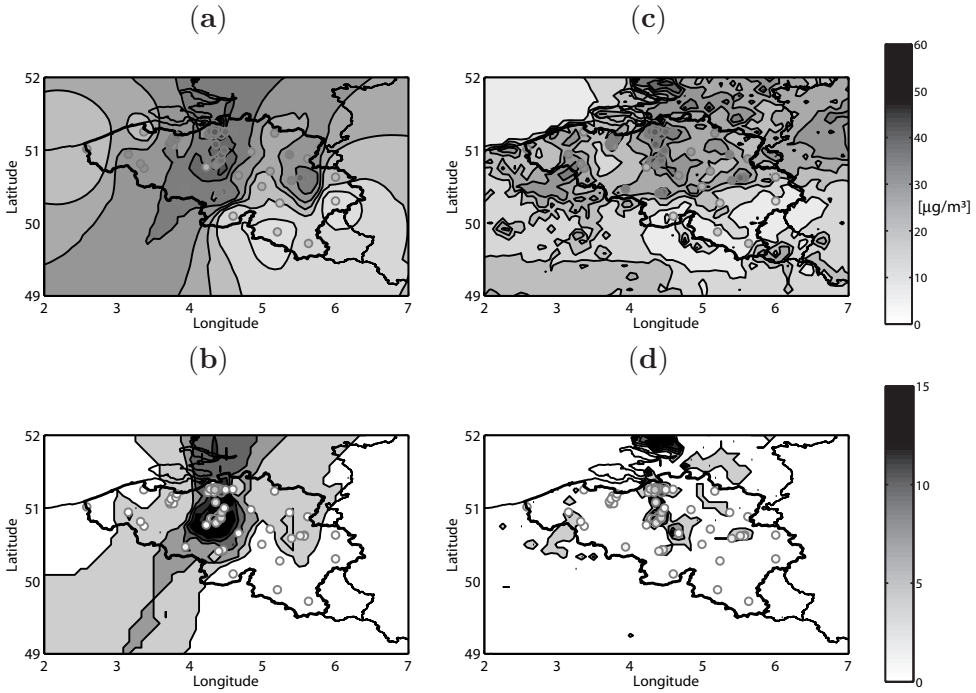


Figure 6.8: Mean concentrations in 2007 (a and c) and total number of high exposure days in 2007 (daily mean higher than chosen threshold of $80\mu g/m^3$; b and d) for both methods. (a) and (b) stand for Simple Kriging whereas (c) and (d) stand for Bayesian data fusion. Circles represent the observed mean concentrations and observed total number of high exposure days on the raw data.

Prediction variance is a common indicator for the quality assessment of space-time prediction methods. By definition, the variance of prediction is the variance of the difference between the unknown target variable and its prediction, namely $\mathbb{V}ar(Z_0 - \widehat{Z}_0)$. Therefore, if one compares two prediction methods, one should use the one with the smallest variance of prediction. Fig. 6.9 shows these variances of prediction for both SK and

BDF for the same dates as in Fig. 6.6. It is clear that BDF prediction yields smaller variances of prediction than SK. It is worth noting here that for days with particularly windy meteorological conditions, there is a significant drop in the BDF variance of prediction compared to the SK one (see Fig. 6.9e and f). Further, it can be seen on Fig. 6.9g and h how the number of secondary information sources affects the variance of prediction for some locations (South-West part of the study area). Indeed, since this number is smaller at those locations, the variance of prediction slightly increases. Variance of prediction thus shows clearly how sources of secondary information were able to increase the precision of the BDF predictions, thus increasing the knowledge on NO_2 predictions.

Although visual interpretations were clearly in favor of the BDF predictions, a validation was also performed for both methods. Predictions were made at the sampled space-time locations for different time-lags (i.e. predictions were performed at time $t+1$ day, $t+2$ days, ..., $t+10$ days using past observations up to time t only along with the complete meteorological information). Fig. 6.10 shows the evolution of the Root Mean Square Error (RMSE) of both methods for some of these time-lags. Though results are very similar for $t+1$ predictions, BDF outperformed SK for larger time-lags as SK performance drops steadily along with the time-lags. This observation is as clearer as the time-lag increases. This shows of course that though SK method is quite acceptable for short-term predictions, its accuracy and its precision drop steadily when the time-lag increases. Moreover, as the quality of SK decreases, the influence of meteorological variables increases and enables to better estimate NO_2 concentrations for longer term predictions.

As the whole NO_2 distribution is known for each location through the fact that Gaussian distributions were assumed and that both the mean and the variance are known, it is easy to estimate the probability that NO_2 concentrations will exceed a given threshold. Fig. 6.11 shows the probability map that NO_2 concentrations will be higher than 50 $\mu g/m^3$ for the same dates as in Fig. 6.6 and 6.9 and for both SK and BDF. Therefore, as the observations at prediction day were not accounted for in the predictions, this somehow validates if the two methods correctly predict the probability of threshold exceedance. Again, a simple comparison between columns in Fig. 6.11 leads to the conclusion that BDF outperformed SK. This is especially obvious on February 8, 2007 (Fig. 6.11b and f) where SK predicts high probability of exceedance all over in Belgium and in The Netherlands whereas BDF correctly predicts high probability of exceedance at only two locations (close to Antwerpen in the Northern Belgium and Brussels in the Central Belgium). Again, this is mainly due to the windy meteorological conditions for this date,

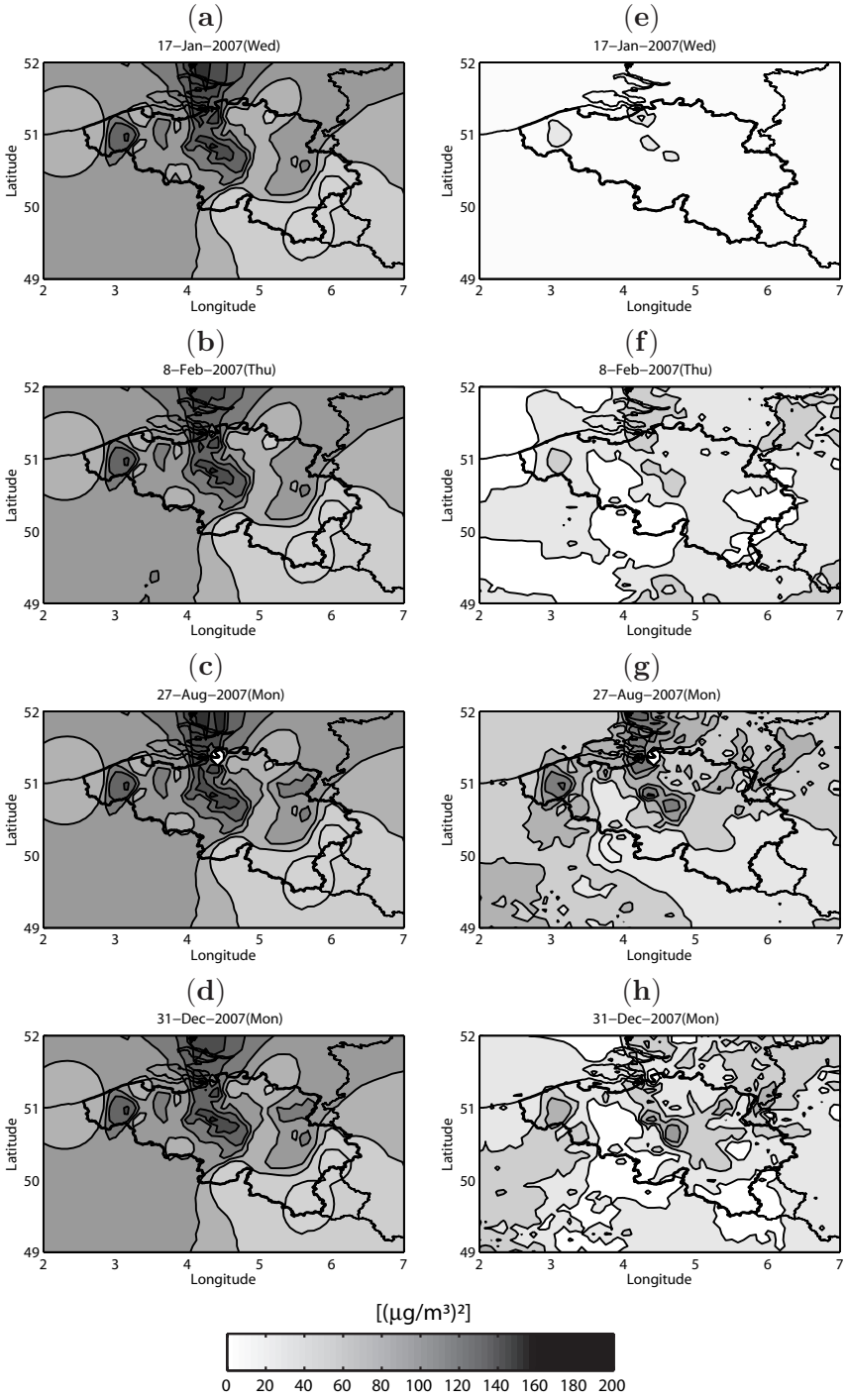


Figure 6.9: Space-time variances of predictions of daily-mean NO_2 concentrations. On the left column are the Simple Kriging variances whereas Bayesian data fusion ones are on the right column. Illustrated days are the same as in Fig. 6.6.

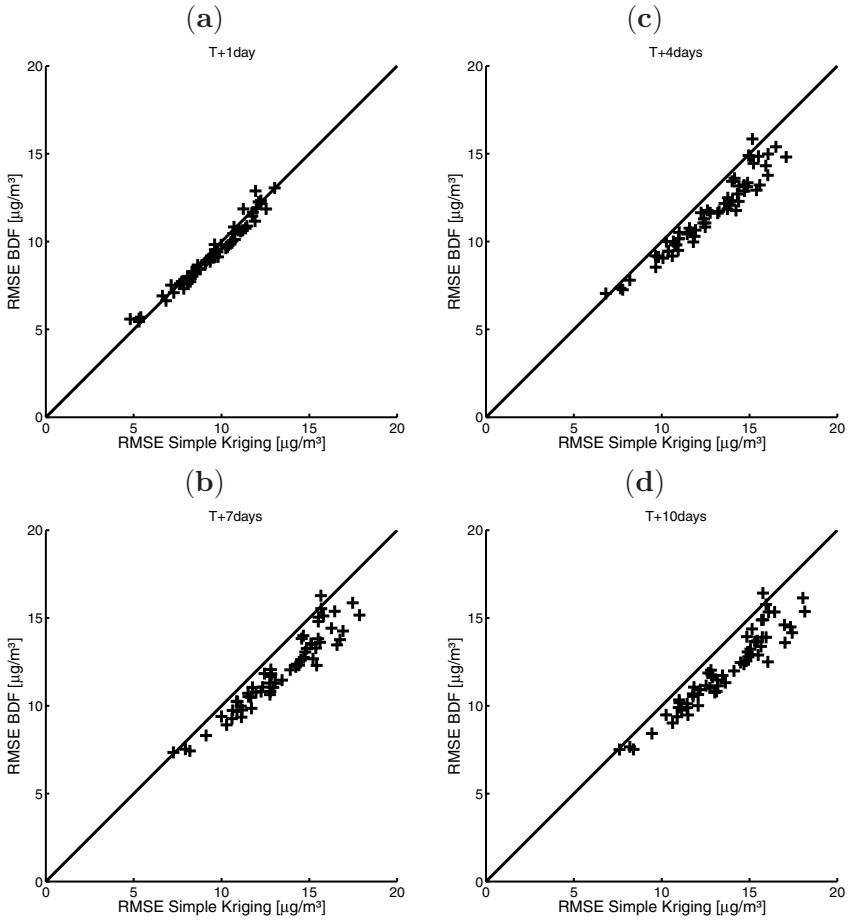


Figure 6.10: Comparison of the RMSE values of Simple Kriging and of BDF predictions with respect to different time-lags of prediction. Each + symbol refers to a different sampled location.

leading thus to a drop in the NO_2 concentrations except in the biggest cities.

6.6 Discussion and Conclusion

In this chapter, the Bayesian data fusion (BDF; Bogaert and Fasbender, 2007) method was applied to the space-time prediction of air pollutants, more specifically to the case of NO_2 in Belgium. The proposed method showed its ability to account for several secondary information sources (in this chapter, the number of secondary sources ranged from 3 to 6 depending on the location). These results led to the conclusion that the use of secondary information is relevant, particularly in case of windy meteorological conditions which lead to a drop in the NO_2 concentrations.

A particular effort was devoted to the presentation of some specific precautions that one should take when using the BDF formula. Indeed, though BDF is a sound theoretical framework, its implementation to real case study should be cautious. Several theoretical properties linking the conditional and marginal distributions should be respected in order to avoid numerical issues. After a short description of these theoretical properties, a solution was also proposed in order to avoid these pitfalls.

As a validation procedure, each prediction on the entire area was first overlaid to the raw values for the same day. By this, one could check that the BDF prediction was relevant and rather close to true observations. Based on predictions over the year 2007, annual means were computed for both BDF and Simple Kriging (SK) methods. These maps were quite acceptable for both methods as no particular shift was observed between the maps and the annual means of the sampled locations, even if the BDF map seemed to be much more relevant thanks to its local patterns. Similarly, for each location in the prediction map, the number of times that the prediction crossed the $80\mu g/m^3$ threshold over the year 2007 were computed for both methods. By comparison with the true number counted at sampled locations, SK overestimated this number whereas BDF results were in very good agreement. Variances of prediction of both methods were also compared. Results revealed that BDF prediction variance was always smaller or equal to the SK one, especially in the case of windy meteorological conditions. Probability maps of crossing a given threshold ($50\mu g/m^3$ here) were easily estimated using a normality assumption. Again, BDF results were better than the SK ones, as high values of the predicted day are generally associated with locations of higher probability, contrary to SK. Finally, predictions were made on

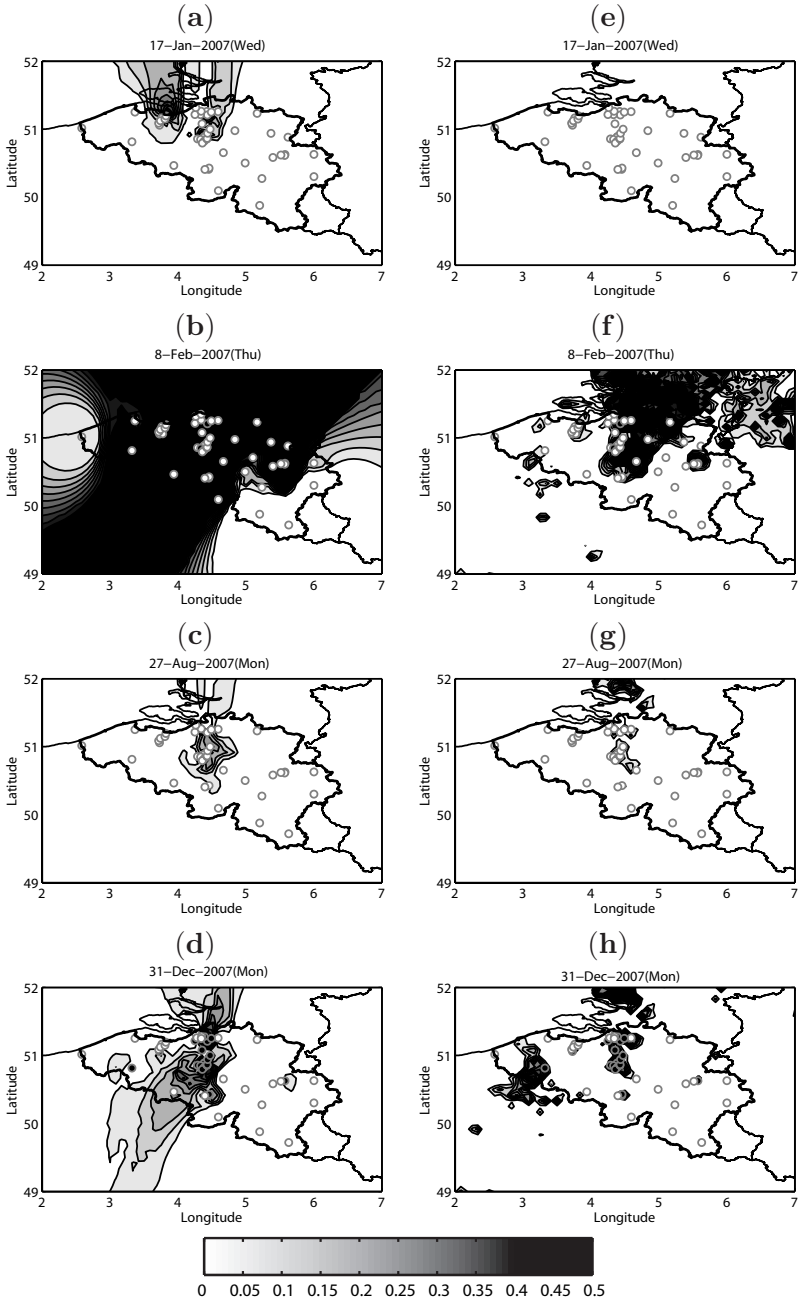


Figure 6.11: Space-time predictions for the probability of crossing the 50 μg/m^3 threshold. Black circles are for observed values crossing the threshold and white ones are for lower observed values. On the left column are the Simple Kriging variances whereas Bayesian data fusion ones are on the right column. Illustrated days are the same as in Fig. 6.6.

sampled locations and Root Mean Square Errors were computed for each sampled location, thus providing a quantitative validation of the results. Though results are quite close for short time-lags, BDF outperformed SK for larger time-lags (see Fig. 6.10). This was of course due to the fact that correlations between observations and variables to be predicted decrease as for the time-lag increases.

It is worth noting that, in this application, information sources did not share the same support (i.e. the same area). Indeed, though raw NO_2 observations were valid for rather small areas, meteorological variables, population density and land use indexes were provided on grids (regular or not) with different resolutions (from 0.005 to 0.25 degree). For sake of simplicity in this chapter, an intermediate resolution (0.25/3 degree) was chosen and information were downscaled (resp. upscaled) on the selected regular grid. Using the theory of regularization of semi-variograms, it should be possible to account explicitly for this change-of-support issue. This possibility was not investigated here and is left for further research on the topic, though accounting for this effect is not expected to drastically change the previous conclusions.

As mentioned in this chapter, BDF has the advantage that specific modeling assumptions can be considered separately for each secondary information source in a complete and sound theoretical probability framework, contrary to other existing methods. Indeed, as an example, cokriging methods enable theoretically to account for several secondary variables, but relationships between variables are assumed to be linear only and these methods generally rely in practice on strong hypotheses (e.g. Linear Model of Coregionalization), limiting thus the scope of theses methods when applied to more complex situations. The BDF framework is thus a significant improvement among geostatistics methods and should be implemented in future space-time applications.

Conclusion and Perspectives

Conclusion

In this thesis, we proposed a Bayesian data fusion (BDF) framework in the context of space-time prediction of environmental variables. Based on a simple error model and the use of Bayes' theorem, the BDF approach aims at reconciling several information sources that are all related to the same non directly observable variable into a unique prediction framework. The number and the nature of these information sources are not restricted in theory, as all possible combinations of continuous, categorical or mixed random variables can be accounted for, whether they are observed at the same location or not. For these reasons, the proposed BDF framework enables us to theoretically tackle a wide set of real-case situations in environmental sciences.

In contrast with classical geostatistical methods (such as kriging-related ones) but similarly to modern ones (e.g. Bayesian Maximum Entropy paradigm; Christakos, 1990, 1991), the BDF framework provides the whole *posterior* distribution and not just some of its statistical moments (i.e. the mean, the mode, the median, the variance,...). By achieving this, the information provided by the final prediction exceeds that of classical methods, as it is possible to directly derive every needed property from the distribution itself (e.g. statistical moments or intervals of prediction). Furthermore, although applications presented in this thesis assumed Gaussian distributions, the BDF approach is *distribution-free* (i.e. in theory BDF is not limited to any particular choice of distribution). However, the choice of Gaussian distributions still remains an important one as it has several interesting properties that can be useful for practical applications; e.g., analytical expressions for the mean and variance of the *posterior* distribution whereas results can only numerically be obtained if this hypothesis does not hold true.

The procedure proposed in the BDF framework has the advantage to be well-stated and presented with intuitive underlying hypotheses. The BDF framework starts from the simple idea that the variable of interest is not directly observable, but is related to some secondary observed variable through an error-like model. Using thus the Bayes theorem along with some independence hypotheses, the method provides an easy way to account for several information sources in a space-time prediction context. Furthermore, as the hypotheses are expressly stated at each step of the development, it is easy to exchange the set of hypotheses that will best fit with the application at hand. As an example, it is

therefore possible to gather secondary variables for which relationships between the variables of interest and the secondary information sources can not be neglected and assume independence hypotheses only conditionally to the primary variable for the remaining secondary variables (see Chapter 3 for a practical example).

As suggested in the introduction and as already expressed in this conclusion, we focused our attention on the consequence of the modeling hypotheses on the final prediction. In the proposed BDF framework, these hypotheses are rather weak : conditional independence of the secondary variables with respect to the variable of interest. Therefore, the BDF theory potentially encompasses a wide set of typical situations[1]. Furthermore, it was proven that assuming these weak hypotheses can be justified on the ground that they maximize the entropy in absence of clear joint information (see in Chapter 1 for the detailed proof). As a consequence, BDF is a relevant choice either because the application is in accordance with the BDF hypotheses or because the loss of information due to this choice is rather negligible. This was illustrated in Chapter 2 where it was shown that the loss induced by these weak hypotheses in BDF was not significant with respect to cokriging results, even in situations where cokriging is known to be the optimal predictor.

As discussed in Chapter 1, the BDF framework proposes also an easy and straightforward way to update previous prediction maps, as long as the uncertainty associated with these maps is known. Indeed, it is clear from the BDF equations and from the independence hypotheses that the update of existing maps only involves products of distributions. Furthermore, it was also shown that in the case of Gaussian distribution (or at least, in case where Gaussian distributions are assumed), the BDF formulation simplifies to the so-called Naive Bayes Fusion rule. As simple as it might appear at the first sight, this fusion rule for updating existing maps is experimentally observed to be surprisingly accurate and efficient, at least for a first approximation of the updated map.

In the second part of this thesis, the BDF framework was applied to four different applications in environmental sciences. The two first applications involved the spatial enhancement of remotely sensed images where for both cases the goal was to merge spectral information (i.e. spectral bands) at different spatial resolutions in order to get fused images at the highest possible spatial resolution. Remotely sensed images are singular examples of secondary information as measurements are exhaustively observed over space, with the specific and convenient

[1]This is in opposition with strong hypotheses that restrict typically a given theory to few particular cases.

property that interpolation is not even needed here. Results provided by the BDF implementation were encouraging compared to existing and widely used image fusion methods. Indeed, colors were best preserved by the BDF method and there were fewer artifacts in the BDF fused images than for other typically used methods such as wavelet-based ones (see e.g. Garzelli and Nencini, 2005) or downscaling cokriging (Pardo-Iguzquiza *et al.*, 2006). Furthermore, it was shown in Chapter 3 that the use of a weighting parameter is a significant and valuable contribution as it enables the user to tune the final results to his own objectives (i.e. color preservation *versus* higher spatial details).

The third application dealing with the spatial estimation of a water table illustrated very well the importance of modeling hypotheses in geostatistics. Indeed, it was shown that cokriging failed to account correctly for secondary information (a Digital Terrain Model here as suggested by Hoeksema *et al.*, 1989; DTM), mainly because cokriging is merely a linear prediction method (i.e. relations between variables are assumed to be linear). On the contrary, BDF was able to account for the non-linear relationships between the water table level, a DTM and the proximity to a river network. Thanks to the definition of a measure of the proximity to the river network, it was easier to code the information brought by the river network and to model its relation with the water table depth. This was motivated by the assumption that the water table elevation is expected to be close to the DTM one at proximity of the river network and then that the DTM information loosens its influence for further locations. Contrary to cokriging, the BDF framework was thus able to account for this diversity in the uncertainty associated jointly with the DTM and the river network, leading thus to more meaningful results (e.g. local patterns induced by the DTM were present in the BDF results).

Though previous applications were only refering to a spatial prediction problem, the last application in Chapter 6 illustrated the possibility of extending BDF to space-time prediction issues. Theory presented in Chapter 1 was explicitly adapted to the case of space-time prediction, more specifically to the case of concentrations of air pollutants. In this application (illustrated by the real-case study of NO_2 concentrations in Belgium), up to six secondary information sources were available at some locations (e.g. wind speed, turbulence index, land use index, road network, population density). From a practical viewpoint, results showed that the method leads to encouraging perspectives in terms of temporal predictions of NO_2 concentrations, which is of particular interest for real-time alert systems.

In the light of the results obtained for real-case applications, the BDF framework has a promising potential in environmental sciences. Indeed, similarly to the case of NO_2 air concentrations, most of the environmental applications typically involve several non-linear processes. Regarding the classical way to tackle this kind of issues and for aforementioned reasons, the BDF approach proposes a convenient alternative and can be expected to be routinely applied for some environmental applications.

Perspectives

We already mentioned that BDF is *distribution-free*. However, in our applications, we assumed Gaussian distributions both for the *prior* and the conditional distributions. An obvious perspective is thus to set aside these Gaussian assumptions by using other distributions. Bogaert and Fasbender (2006) investigated the possibility of estimating *prior* distributions using marginal distributions and correlation coefficients, but the gain with respect to the Gaussian *prior* remains unclear. However, non-Gaussian conditional distributions offer promising perspectives for more complex applications. Furthermore, it could be interesting to assess the real influence of those Gaussian assumptions on prediction maps in case of significant singularities in the conditional distributions (e.g. asymmetric distributions, semi-bounded domains, long tailed distributions, multimodal distributions, etc).

In this thesis, the BDF theory was only developed for the prediction of continuous variables. It is however possible to extent these developments to the case of categorical variables as well. Similarly to the case of continuous variables, this will enable us to potentially account for a wide diversity of information involved in the prediction of categorical variables. The production of soil maps, occupation maps and land use maps are just examples of typical situations for which the influence of the proximity to a river network, to cities or to industrial zones is a valuable information, so that similarities can be found with the water table application of Chapter 5. Furthermore, updating such maps is an important issue that can be easily tackled by the BDF approach.

In environmental applications, information sources might be associated with diverse uncertainty so that there is a distinction between uncertainty associated with the information sources (i.e. the shape of the conditional distribution) and the reliability of the sources themselves (i.e. the quality of the information). As an example, if we want to merge an existing map that contains 90% of errors with a second one that contains only 10%, the contribution of the first map to the final re-

sults should be rather minor compared to the contribution of the second one. Chapter 1 mentioned the possibility of weighting the information sources so that credit given to each of these sources can be accounted for in the fusion process. This idea was successfully applied in Chapter 3 in order to propose an adaptable image fusion algorithm. However, although this idea is very promising, the unanswered question is how to choose or estimate these weighting parameters in practice. Franken and Hupper (2005) already proposed an easy way for such estimation when fusing (multivariate) Gaussian distributions only, but there is still need for further investigations when using other distributions.

In this thesis, we assumed that both the *prior* and the conditional distributions were known or that they could be easily inferred. In the proposed applications, several parameters were thus estimated for those distributions. However, the uncertainty of the estimated parameters were never accounted for although this might lead to significant changes for the *posterior* distribution. Fortunately, the generalization of the BDF framework to classical Bayesian statistics (i.e. statistics that account for parameters uncertainty) should be easy. Indeed, accounting for uncertainty in Bayesian statistics is done by means of integrals and *prior* distribution about the parameters, a methodology which directly fits within the BDF framework. The impact of parameters uncertainty on *posterior* distributions will then not be neglected anymore.

The NO_2 application of Chapter 6 raised some issues about how to estimate jointly both the *prior* and the conditional distribution for avoiding consistency issues. Indeed, numerical instabilities were observed under some circumstances (see Chapter 6 for more details about this). There is thus a real need for sound theoretical methodologies that enable us to avoid this problem. Nevertheless, as we have not search exhaustively in the literature yet, such methods might already exist in other branches of statistics.

Although BDF can be viewed as sound from a theoretical viewpoint, there still are practical limitations when applying it in cases of secondary information that are non-exhaustively observed in space and time. Indeed, BDF equations rely on the computation of multivariate integrals, a task that becomes difficult and cumbersome as the number of integrals increases (except in some circumstances ; e.g., if Gaussian distributions are assumed). In that context, it could be valuable to write and gather in a library generic implementations of BDF for an easier diffusion in the scientific community. It is worth knowing that there already exists such a library for BME (BMELib, 2008). The new BDF library might thus rely on several programs of BME as both BDF and BME need similar integral calculus resources and also share some similarities.

From a practical viewpoint, due to a lack of methodologies, it might be necessary to restrain the number of locations for the secondary information since the computation of multivariate integrals becomes difficult as this number increases. Moreover, it is also worth noting that the influence of the secondary information on the prediction decreases along with its distance to the prediction location, so that accounting for further secondary information has almost no effect on the results in practice. It is thus rather sufficient to account for closer secondary information as additional information will generally be too far for significantly influencing the final results. In Chapter 1, we suggested a simple forward selection procedure for selecting sequentially the subset of secondary variables that maximizes its mutual information when combined with primary variables, but this procedure has not been tested in practice yet.

Appendices

Appendix A

Kriging with measurement errors

Assume that $\mathbf{Z}$ is a random vector sampled from a Gaussian second-order stationary spatial RF $\mathfrak{Z} = \{Z(\mathbf{x}) : \mathbf{x} \in D \subseteq \mathbb{R}^d, Z(\mathbf{x}) \in \mathbb{R}\}$ so that $\mathfrak{Z}$ is fully characterized by its mean function $\mu(\mathbf{x})$ and covariance function $C(\mathbf{h})$. Let us denote $\Sigma_{\mathbf{Z}} = \mathbb{C}ov[\mathbf{Z}]$, $\boldsymbol{\mu}_{\mathbf{Z}} = \mathbb{E}[\mathbf{Z}]$. Let us consider that $\mathbb{E}[\mathbf{E}] = \mathbf{0}$ and $\mathbb{C}ov[\mathbf{E}] = \sigma_E^2\mathbf{I}$, where $\mathbf{I}$ is an identity matrix of appropriate size. For $\mathbf{Y} = \mathbf{Z} + \mathbf{E}$, we thus have $\boldsymbol{\mu}_{\mathbf{Y}} = \boldsymbol{\mu}_{\mathbf{Z}}$ and $\Sigma_{\mathbf{Y}} = \Sigma_{\mathbf{Z}} + \sigma_E^2\mathbf{I}$ because $\mathbf{E} \perp \mathbf{Z}$. Let us now consider the following partitions

$$\mathbf{Y} = \begin{pmatrix} Y_0 \\ \mathbf{Y}_b \end{pmatrix} \quad ; \quad \boldsymbol{\mu}_{\mathbf{Z}} = \begin{pmatrix} \mu_0 \\ \boldsymbol{\mu}_b \end{pmatrix} \quad ; \quad \Sigma_{\mathbf{Z}} = \begin{pmatrix} \sigma_0^2 & \boldsymbol{\sigma}' \\ \boldsymbol{\sigma} & \Sigma_b \end{pmatrix}$$

The simple kriging with measurement errors predictor Z_0^p and the corresponding prediction variance $\sigma_0^{2,p}$ are then given by the expressions

$$Z_0^p = \boldsymbol{\lambda}'(\mathbf{Y}_b - \boldsymbol{\mu}_b) + \mu_0 \quad ; \quad \sigma_0^{2,p} = \sigma_0^2 - \boldsymbol{\lambda}'(\Sigma_b + \sigma_E^2\mathbf{I})\boldsymbol{\lambda} \tag{A.1}$$

where the vector of weights $\boldsymbol{\lambda}$ is the solution of a linear system of equations

$$(\Sigma_b + \sigma_E^2\mathbf{I})\boldsymbol{\lambda} = \boldsymbol{\sigma} \quad \Longleftrightarrow \quad \boldsymbol{\lambda}' = \boldsymbol{\sigma}'(\Sigma_b + \sigma_E^2\mathbf{I})^{-1}$$

Using Eq. (1.6) under the same hypotheses, it can then be shown that

$$Z_0^p = \mathbb{E}[Z_0|\mathbf{y}] \quad ; \quad \sigma_0^{2,p} = \mathbb{V}ar[Z_0|\mathbf{y}]$$

Indeed, from Eq. (1.5), it is clear that we have

$$f(\mathbf{y}, \mathbf{z}) = f(\mathbf{y}|\mathbf{z})f(\mathbf{z}) = f(\mathbf{z})\prod_{i=0}^{n} f_{E_i}(y_i - z_i) \tag{A.2}$$

Plugging the expression of the Gaussian pdf's into Eq. (A.2) gives

$$\begin{aligned} f(\mathbf{y},\mathbf{z}) \propto & \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_\mathbf{Z})'\Sigma_\mathbf{Z}^{-1}(\mathbf{z}-\boldsymbol{\mu}_\mathbf{Z}) - \frac{1}{2\sigma_E^2}(\mathbf{y}-\mathbf{z})'(\mathbf{y}-\mathbf{z})\right) \\ = & \exp(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_\mathbf{Z})'\left(\Sigma_\mathbf{Z}^{-1}+\frac{1}{\sigma_E^2}\mathbf{I}\right)(\mathbf{z}-\boldsymbol{\mu}_\mathbf{Z}) \\ & +\frac{1}{2\sigma_E^2}(\mathbf{y}-\boldsymbol{\mu}_\mathbf{Z})'(\mathbf{z}-\boldsymbol{\mu}_\mathbf{Z}) - \frac{1}{2\sigma_E^2}(\mathbf{y}-\boldsymbol{\mu}_\mathbf{Z})'(\mathbf{y}-\boldsymbol{\mu}_\mathbf{Z}) \\ = & \exp\left(-\frac{1}{2}\begin{pmatrix}\mathbf{z}-\boldsymbol{\mu}_\mathbf{Z}\\ \mathbf{y}-\boldsymbol{\mu}_\mathbf{Z}\end{pmatrix}'\begin{pmatrix}\Sigma_\mathbf{Z}^{-1}+\frac{1}{\sigma_E^2}\mathbf{I} & -\frac{1}{\sigma_E^2}\mathbf{I}\\ -\frac{1}{\sigma_E^2}\mathbf{I} & \frac{1}{\sigma_E^2}\mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{z}-\boldsymbol{\mu}_\mathbf{Z}\\ \mathbf{y}-\boldsymbol{\mu}_\mathbf{Z}\end{pmatrix}\right) \end{aligned}$$

and this final expression proves clearly that $(\mathbf{Y},\mathbf{Z})$ is multivariate Gaussian with $\mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbf{Z}] = \boldsymbol{\mu}_\mathbf{Z}$, where the inner matrix in the last line is the inverse of the joint covariance matrix $\Sigma_{(\mathbf{Y},\mathbf{Z})}$, so that taking its inverse gives

$$\Sigma_{(\mathbf{Y},\mathbf{Z})} = \begin{pmatrix}\Sigma_\mathbf{Z} & \Sigma_\mathbf{Z}\\ \Sigma_\mathbf{Z} & \Sigma_\mathbf{Z}+\sigma_E^2\mathbf{I}\end{pmatrix}$$

Therefore, we know from multivariate Gaussian distribution theory that $Z_0|\mathbf{y}$ follows a Gaussian distribution with conditional mean and variance given by

$$\mathbb{E}[Z_0|\mathbf{y}] = \boldsymbol{\sigma}'\left(\Sigma_b+\sigma_E^2\mathbf{I}\right)^{-1}(\mathbf{Y}_b-\boldsymbol{\mu}_b)+\mu_0 = \boldsymbol{\lambda}'(\mathbf{Y}_b-\boldsymbol{\mu}_b)+\mu_0$$

$$\mathbb{V}ar[Z_0|\mathbf{y}] = \sigma_0^2 - \boldsymbol{\sigma}'\left(\Sigma_b+\sigma_E^2\mathbf{I}\right)^{-1}\boldsymbol{\sigma}$$

these being precisely the expressions for Z_0^p and $\sigma_0^{2,p}$ as given by Eq. (A.1).

Appendix B

Maximum entropy with known moments

If we assume that we only know the vector of means $\boldsymbol{\mu}$ and covariance matrix Σ for $(\mathbf{E}', \mathbf{Z}')'$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{E}} \\ \boldsymbol{\mu}_{\mathbf{Z}} \end{pmatrix} \quad ; \quad \Sigma = \begin{pmatrix} \Sigma_{\mathbf{E}} & V \\ V & \Sigma_{\mathbf{Z}} \end{pmatrix}$$

then we know that the maximum entropy solution for its joint pdf will be $(\mathbf{E}', \mathbf{Z}')' \sim N(\boldsymbol{\mu}, \Sigma)$, with entropy $H(\mathbf{E}, \mathbf{Z})$ given by

$$H(\mathbf{E}, \mathbf{Z}) = \ln \sqrt{(2\pi)^{2n}} + \ln \sqrt{\det(\Sigma)}$$

so that $H(\mathbf{E}, \mathbf{Z})$ is maximum if $\det(\Sigma)$ is maximum. We also know that

$$\det(\Sigma) = \det(\Sigma_{\mathbf{Z}}) \det(\Sigma_{\mathbf{E}} - V\Sigma_{\mathbf{Z}}^{-1}V)$$

where $\det(\Sigma_{\mathbf{Z}})$ is known and where $\Sigma_{\mathbf{E}} - V\Sigma_{\mathbf{Z}}^{-1}V$ is the Schur's complement. From Fisher's inequality, we have

$$\det(\Sigma) \leq \det(\Sigma_{\mathbf{Z}}) \det(\Sigma_{\mathbf{E}})$$

with equality occurring for $V = \mathbf{0}$, thus showing that assuming the independence $\mathbf{E} \perp \mathbf{Z}$ leads to the maximum value if V is unknown. The maximum entropy solution is thus obtained with $\mathbf{Z} \sim N(\boldsymbol{\mu}_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$, $\mathbf{E} \sim N(\boldsymbol{\mu}_{\mathbf{E}}, \Sigma_{\mathbf{E}})$ and $\mathbf{E} \perp \mathbf{Z}$. From Hadamard's inequality, we also have

$$\det(\Sigma_{\mathbf{E}}) \leq \prod_{i=0}^{n} \sigma_{E_i}^2$$

and as we finally know that

$$\det(\mathbf{D}) = \prod_{i=0}^{n} \sigma_{E_i}^2 \qquad \text{with } \mathbf{D} = \begin{pmatrix} \sigma_{E_0}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{E_1}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{E_n}^2 \end{pmatrix}$$

this shows that $\det(\Sigma) \leq \det(\Sigma_{\mathbf{Z}}) \det(\mathbf{D})$. If only variances $\sigma_{E_i}^2$ are known instead of $\Sigma_{\mathbf{E}}$, the maximum entropy is thus reached by assuming additionally $E_0 \perp \cdots \perp E_n$.

Appendix C

Details for Chapter 3

Let us remember that the pdf $f(\mathbf{z})$ of a multivariate Gaussian vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by

$$f(\mathbf{z}) \;\propto\; e^{-\frac{1}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}+\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{z}}. \tag{C.1}$$

Using this expression in 3.3 along with the specific assumptions in Section 3.4.2yields the result

$$\begin{aligned}
f(\mathbf{z}|\mathbf{y}_S, y_P) \quad &\propto \quad f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{E}_S}(\mathbf{y}_S - g_S(\mathbf{z}))^{2(1-w)} \\
&\quad \times f_{E_P}(y_P - g_P(\mathbf{z}))^{2w} \\
&\propto \quad exp\left(-2(1-w)\frac{1}{2}\left(\mathbf{y}_S - \mathbf{z}\right)'\boldsymbol{\Sigma}_S^{-1}\left(\mathbf{y}_S - \mathbf{z}\right)\right) \\
&\quad \times exp\left(-2w\frac{1}{2\sigma_P^2}\left(y_P - \alpha - \mathbf{z}'\boldsymbol{\beta}\right)'\left(y_P - \alpha - \mathbf{z}'\boldsymbol{\beta}\right)\right) \\
&\propto \quad exp\left(-2(1-w)\frac{1}{2}\mathbf{z}'\boldsymbol{\Sigma}_S^{-1}\mathbf{z} + 2(1-w)\mathbf{y}_S'\boldsymbol{\Sigma}_S^{-1}\mathbf{z}\right) \\
&\quad \times exp\left(-2w\frac{1}{2\sigma_P^2}\mathbf{z}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{z} + 2w\frac{1}{\sigma_P^2}\left(y_P - \alpha\right)\boldsymbol{\beta}'\mathbf{z}\right) \\
&\propto \quad exp\left(-\frac{1}{2}\mathbf{z}'\left(2(1-w)\boldsymbol{\Sigma}_S^{-1} + 2w\frac{1}{\sigma_P^2}\boldsymbol{\beta}\boldsymbol{\beta}'\right)\mathbf{z}\right) \\
&\quad \times exp\left(\left(2(1-w)\mathbf{y}_S'\boldsymbol{\Sigma}_S^{-1} + 2w\frac{1}{\sigma_P^2}\left(y_P - \alpha\right)\boldsymbol{\beta}'\right)\mathbf{z}\right).
\end{aligned}$$

By identification of the terms in this expression with the terms in C.1, we find that $\boldsymbol{\mu}_w$ and $\boldsymbol{\Sigma}_w$ are given by

$$\left\{ \begin{array}{lcl} \boldsymbol{\mu}_w & = & \boldsymbol{\Sigma}_w \left(2(1-w)\boldsymbol{\Sigma}_S^{-1}\mathbf{y}_S \right. \\ & & \left. + 2w\dfrac{1}{\sigma_P^2}(y_P - \alpha)\boldsymbol{\beta} \right) \\ \boldsymbol{\Sigma}_w^{-1} & = & 2(1-w)\boldsymbol{\Sigma}_S^{-1} + 2w\dfrac{1}{\sigma_P^2}\boldsymbol{\beta}\boldsymbol{\beta}' \end{array} \right. .$$

Results for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in the non-weighted case are directly obtained by setting $w = 0.5$ in these expressions.

Appendix D

Details for Chapter 5

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

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

For the water table prediction study, it was shown that (see Eq. (5.7))

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

Assuming now that all these pdf's are Gaussian and plugging the corresponding expressions given by Eq. (D.1) into the previous equality, one obtains

$$\begin{aligned} & f(z_0|\mathbf{z}_S, DEM(\mathbf{x}_0), d_{DEM}(\mathbf{x}_0)) \\ & \propto e^{-\frac{1}{2\sigma_E^2}(z_0-\mu_d)^2} e^{\frac{1}{2\sigma_0^2}(z_0-\mu_0)^2} e^{-\frac{1}{2\sigma_k^2}(z_0-\mu_k)^2} \\ & \propto e^{-\frac{1}{2}\left(\frac{1}{\sigma_k^2}+\frac{1}{\sigma_E^2}-\frac{1}{\sigma_0^2}\right)z_0^2 + \left(\frac{\mu_k}{\sigma_k^2}+\frac{\mu_d}{\sigma_E^2}-\frac{\mu_0}{\sigma_0^2}\right)z_0} \end{aligned} \tag{D.2}$$

By direct identification between Eqs. (D.1) and (D.2), 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

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

Appendix E

Properties of the fused variance

Assuming that $f(z_i)$ (resp. $f(z_i|y_{i,j})$ for a given $y_{i,j}$) in Eq. 6.5 is a Gaussian distribution with mean equal to μ (resp. μ_y) and variance σ^2 (resp. σ_y^2), then the ratio $\dfrac{f(z_i|y_{i,j})}{f(z_i)}$ is proportional to a third Gaussian distribution with mean equal to μ_F and variance equal to σ_F^2 where

$$\left\{ \begin{array}{rcl} \sigma_F^2 & = & \dfrac{\sigma_y^2 \sigma^2}{\sigma^2 - \sigma_y^2} \\ \\ \mu_F & = & \dfrac{\sigma^2 \mu_y - \sigma_y^2 \mu}{\sigma^2 - \sigma_y^2} \end{array} \right. \tag{E.1}$$

Now recoding $\sigma_y^2 = a^2\sigma^2$ ($a^2 < 1$) in Eq. E.1, it is clear that $\mu_F = \dfrac{\mu_y - a^2\mu}{1 - a^2}$ so that

$$\lim_{\substack{a \to 1 \\ a < 1}} \mu_F = \left\{ \begin{array}{ll} +\infty & \text{if } \mu_y < \mu \\ \mu & \text{if } \mu_y = \mu \\ -\infty & \text{if } \mu_y > \mu \end{array} \right.$$

As a consequence, the only admissible situation is $\mu_y = \mu$, otherwise the use of Eq. 6.5 leads to meaningless results.

One way to circumvent this difficulty is to follow the four steps of this modeling process

1. define R_i as in Eq. 6.8
2. use $Y_{i,j}$ observations to model $\mu_{R_i|Y_{i,j}} (= \mathbb{E}_{R_i}[R_i|Y_{i,j}])$

3. use $\mu_{R_i|Y_{i,j}}$ values to model $\sigma^2_{R_i|Y_{i,j}} (= \mathbb{V}ar_{R_i}[R_i|Y_{i,j}])$ such that
   (a) $\lim_{\mu_{R_i|Y_{i,j}} \to 0} \sigma^2_{R_i|Y_{i,j}} = 1$
   (b) $\lim_{|\mu_{R_i|Y_{i,j}}| \to +\infty} \sigma^2_{R_i|Y_{i,j}} = 0$

4. back-transform the modeled $\mu_{R_i|Y_{i,j}}$ and $\sigma^2_{R_i|Y_{i,j}}$ to the original variable Z_i, i.e.,
   (a) $\mu_{Z_i|Y_{i,j}} = \sigma(\boldsymbol{\ell}_i)\mu_{R_i|Y_{i,j}} + \mu(\boldsymbol{\ell}_i, t_i)$
   (b) $\sigma^2_{Z_i|Y_{i,j}} = \sigma^2_{R_i|Y_{i,j}}\sigma^2(\boldsymbol{\ell}_i)$

Thanks to this, the mean μ_F (resp. variance σ^2_F) of the fused Gaussian distribution will tend to μ (resp. to $+\infty$), leading thus to a meaningful contribution for the ratio $\dfrac{f(z_i|y_{i,j})}{f(z_i)}$ in Eq. 6.5.

Bibliography

Aiazzi, B., Alparone, L., Baronti, S., and Garzelli, A. 2002a. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. IEEE Trans. Geosci. Remote Sens., 40(10):2300–2312.

Aiazzi, B., Alparone, L., Baronti, S., Pippi, I., and Selva, M. 2002b. Generalized Laplacian pyramid-based fusion of MS + P image data with spectral distortion minimization. In *ISPRS Int. Arch. Photogram. Remote Sens.*, volume 39, pages 3–6.

Aiazzi, B., Alparone, L., Baronti, S., Santurri, L., and Selva, M. 2005. Spatial resolution enhancement of ASTER thermal bands. In Bruzzone, L., editor, *Proceedings of SPIE - vol. 5982, Image and Signal Processing for Remote Sensing XI*.

Alparone, L., Baronti, S., Garzelli, A., and Nencini, F. 2004. A global quality measurement of pansharpened multispectral imagery. IEEE Geosci. Remote Sens. Letters, 1:313– 317.

Altinçay, H. 2005. On naive Bayesian fusion of dependent classifiers. Pattern Recogn. Lett., 26:2463–2473.

Ballester, C., Caselles, V., Igual, L., Verdera, J., and Rougé, B. 2006. A variational model for P+XS image fusion. Int. J. Comput. Vision, 69(1).

Blum, R. S. 2005. Robust image fusion using a statistical signal processing approach. Inf. Fusion, 6:119–128.

Blum, R. S. 2006. On multisensor image fusion performance limits from an estimation theory perspective. Inf. Fusion, 7:250–263.

BMELib. 2008. The Bayesian Maximum Entropy software for Space/Time Geostatistics, and temporal GIS data integration. `http://www.unc.edu/depts/case/BMELIB/`.

Bogaert, P. 1996. *Geostatistics applied to the analysis of space-time data*. Ph.D. thesis, Université catholique de Louvain.

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'Or, D. 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 Fasbender, D. 2006. Nonlinear spatial prediction with non-Gaussian data : a maximum entropy viewpoint. In *Proceedings of the 6th International Conference on Geostatistics for Environmental Applications. geoENV 2006*, pages 445–455. Rhodos, Greece.

Bogaert, P. and Fasbender, D. 2007. Bayesian data fusion in a spatial prediction context: A general formulation. Stoch. Env. Res. Risk A., 21(6):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.

Brus, D., de Gruijter, J., Marsman, B., Visscers, R., Bregt, A., Breeuwsma, A., and Bouma, J. 1996. The perfomance of spatial interpolation methods and choropleth maps to estimate properties at points: a soil survey case study. Environmetrics, 7:1–16.

Brussels Environment. 2008. Rapport sur les incidences environnementales du plan d'urgence en cas de pic de pollution. `http://documentation.bruxellesenvironnement.be/documents/RIE\_pic\_pollution\_20080528\_FR.PDF`. In French.

Bruzzone, L., Cossu, R., and Vernazza, G. 2002. Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images. Inf. Fusion, 3(4):289–297.

Chavez, P. S., Sides, S. C., and Anderson, A. 1991. Comparison of three diffrent methods to merge multiresolution and multispectral data : Landsat TM and SPOT panchromatic. Photogram. Eng. Remote Sens., 57(3):295–303.

Chen, C. M., Hepner, G. F., and Forster, R. R. 2003. Fusion of hyperspectral and radar data using the IHS transformation to enhance urban surface features. ISPRS J. Photogram. and Rem. Sensing, 58:19–30.

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

Cho, S., Beak, S., and Kim, J. 2003. Exploring artificial intelligence-based data fusion for conjoint analysis. Expert Syst. Appl., 24:287–294.

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 Grandy, W. T. and Schick, L. H., editors, *Maximum entropy and Bayesian methods*, pages 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 Kolovos, A. 1999. A study of the spatiotemporal health impacts of ozone exposure. J. Expos. Anal. Environ. Epidemiol., 9(4):322–335.

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

Christakos, G. and Serre, M. L. 2000. BME analysis of spatiotemporal particulate matter distributions in North Carolina. Atmosph. Environ., 34(20):3393–3406.

Christakos, G. and Vyas, V. M. 1998a. A composite space/time approach to studying ozone distribution over Eastern United States. Atmos. Environm., 32(16):2845–2857.

Christakos, G. and Vyas, V. M. 1998b. A novel method for studying population health impacts of spatiotemporal ozone distribution. Social Sci. Medicine, 47(8):1051–1066.

Christakos, G., Serre, M. L., and Kovitz, J. L. 2001. BME representation of particulate matter distributions in the state of California on the basis of uncertain measurements. J. Geophys. Res., 106(D9):9717–9731.

Christakos, G., Bogaert, P., and Serre, M. L. 2002. *Temporal GIS. Advanced Functions for Field-Based Applications*. Springer-Verlag, New York NY.

Chung, A. C. S. and Shen, H. C. 2000. Entropy-based Markov chains for multisensor fusion. J. Intell. Robot. Syst., 29(2):161–189.

Costa, J. 2003. Reducing the impact of outliers in ore reserves estimation. Math. Geol., 35(3):323–345.

Costantini, M., Farina, A., and Zirilli, F. 1997. The fusion of different resolution SAR images. In *Proceedings of the IEEE*, pages 139–146.

Cover, T. and Joy, A. 2006. *Elements of Information Theory*. Willey, 2nd edition. 748pp.

Cremer, F., Schutte, K., Schavemaker, J. G. M., and den Breejen, E. 2001. A comparison of decision-level sensor-fusion methods for anti-personnel land mine detection. Inf. Fusion, 2:187–208.

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.

Cressie, N. 1993. *Statistics for Spatial Data*. Wiley series in probability and mathematical statistics. John Wiley and Sons, Inc., New York, NY, revised edition. 928pp.

Daubechies, I. 1992. *Ten Lectures on Wavelets.*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Math.*. Philadelphia, SIAM edition.

Desbarats, A., Logan, C., Hinton, M., and Sharpe, D. 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 Bogaert, P. 2004. Spatial prediction of categorical variables with the Bayesian Maximum Entropy approach: the Ooypolder case study. Europ. J. Soil Science, 55:763–775.

D'Or, D., Bogaert, P., and Christakos, G. 2001. Application of the BME approach to soil texture mapping. Stoch. Env. Res. Risk A., 15(1):87–100.

Du, Y., Vachon, P. W., and van der Sanden, J. J. 2003. Satellite image fusion with multiscale wavelet analysis for marine applications: preserving spatial information and minimizing artifacts (PSIMA). Can. J. Remote Sens., 29:14–23.

Duc, B., Bigun, E. S., Bigun, J., Maitre, G., and Fischer, S. 1997. Fusion of audio and video information for multi modal person authentication. Pattern Recogn. Lett., 18:835–843.

Fasbender, D. and Bogaert, P. 2006. Data fusion in a spatial multivariate framework : Trading off hypotheses against information. In *Proceedings of the 6th International Conference on Geostatistics for Environmental Applications. geoENV 2006*, pages 457–466. Rhodos, Greece.

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

Fasbender, D., Peeters, L., Bogaert, P., and Dassargues, A. 2008a. Bayesian data fusion applied to water table spatial mapping. Water Resour. Res. Accepted for publication in 2008.

Fasbender, D., Radoux, J., and Bogaert, P. 2008b. Bayesian data fusion for adaptable image pansharpening. IEEE Trans. Geosci. Remote Sens., 46(6):1847–1857.

Fasbender, D., Tuia, D., Bogaert, P., and Kanevski, M. 2008c. Support-based implementation of Bayesian data fusion for spatial enhancement : Applications to ASTER thermal images. IEEE Geosci. Remote Sens. Letters, 6(4):598–602.

Fassinut-Mombot, B. and Choquel, J.-B. 2004. A new probabilistic and entropy fusion approach for management of information sources. Inf. Fusion, 5:35–47.

FEBIAC. 2008. Datadigest 2008. http://www.febiac.be/public/statistics.aspx?FID=23\&lang=FR. In French.

Fletcher, R. S. 2005. Evaluating high spatial resolution imagery for detecting citrus orchards affected by sooty mould. Int. J. Remote Sens., 26:495–502.

Franken, D. and Hupper, A. 2005. Improved fast covariance intersection for distributed data fusion. In *Proceedings of Information Fusion 2005*.

Fransens, R., Strecha, C., and Van Gool, L. 2004. A Probabilistic Approach to Optical Flow based Super-Resolution. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04), 2004*, page 191.

Garzelli, A. and Nencini, F. 2005. Interband structure modeling for Pan-sharpening of very high-resolution multispectral images. Inf. Fusion, 6(3):213–224.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. *Bayesian Data Analysis.*. Chapman and Hall, 2nd edition edition.

GeoEYE. 2006. IKONOS relative spectral response and radiometric calibration coefficient. http://www.geoeye.com/products/imagery/ikonos/spectral.html.

Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation.*. Oxford University Press, New York, NY.

Green, P. J. and Silverman, B. W. 1994. *Nonparametric Regression and Generalized Linear Model*. Number 58 in Monographs on Statistics and Applied Probabilité. Chapman and Hall, London.

Gros, X., Bousigue, J., and Takahashi, K. 1999. NDT data fusion at the pixel level. NDT & E Int., 32:283–292.

Hardie, R., Eismann, M., and Wilson, G. 2004. MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor. IEEE Trans. Image Process., 13(9):1174–1184.

Harrison, B. A. and Jupp, D. L. B. 1990. *Introduction to image processing.*. CSIRO Division of Water Resources, Canberra.

Hirschmugl, M., Gallaun, H., Perko, R., and Schardt, M. 2005. Pansharpening - Methoden fur digitale, sehr hoch auflösende Fenerkundungsdaten. In *Beiträge zum 17. AGIT Symposium*. Salzburg, Austria.

Hoeksema, R., Clapp, R., Thomas, A., Hunley, A., Farrow, N., and Dearstone, K. 1989. Cokriging Model for Estimation of Water Table Elevation. Water Resour. Res., 25:429–438.

IAHS. 2003. IAHS Decade on Predictions in Ungauged Basins (PUB). PUB Science and Implementation Plan 2003-2012.

Janssen, S., Dumont, G., Fierens, F., and Mensink, C. 2008. Spatial interpolation of air pollution measurements using CORINE land cover data. Atmos. Environm., 42(20):4884–4903.

Jaquet, O. 1989. Factorial kriging analysis applied to geological data from petroleum exploration. Math. Geol., 21(7):683–691.

Jaynes, E. 2003. *Probability Theory: the Logic of Science*. Cambridge University Press. 758pp.

Jones, G. D., Allsop, R. E., and Gilby, J. H. 2003. Bayesian analysis for fusion of data from disparate imaging systems for surveillance. Image and Vision Computing, 21(10):843–849.

Journel, A. and Huijbregts, C. 1978. *Mining Geostatistics*. Academic Press, New York.

Kantor, I. and Solodnikov, A. 1989. *Hypercomplex Numbers, an Elementary Introduction to Algebras*. Springer-Verlag, New York, NY.

Krige, D. 1951. A statistical approach to some basic mine valuation problems on the witwatersrand. J. Chem. Metall. Min. Soc. South Africa, 52(6):119–139.

Kullback, S. and Leibler, R. 1951. On information and sufficiency. Ann. Math. Stat., 22:79–86.

Kuncheva, L. 2004. *Combining Pattern Classifiers Methods and Algorithms*. Wiley & Sons. 350pp.

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

Laporterie-Déjean, F., de Boissezon, H., Flouzat, G., and Lefèvre-Fonollosa, M.-J. 2005. Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR images. Inf. Fusion, 6(3):193–212.

Lewis, D. 1998. Naive (Bayes) at forty: the independence assumption in information retrieval. In *Proceedings of the European Conference on Machine Learning*, pages 4–15. Springer, Berlin.

Li, S., J.T., K., and Wang, Y. 2002. Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images. Inf. Fusion, 3:17–23.

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

Matlab. 2002. Mathworks Inc. Version 6.5.

Melgani, F. and Serpico, S. B. 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.

Merino, M. and Nunez, J. 2007. Super-Resolution of remotely sensed images with Variable-Pixel Linear Reconstruction. IEEE Trans. Geosci. Remote Sens., 45:1446–1457.

Mitchell, H. 2007. *Multi-Sensor Data Fusion*. Springer-Verlag, Berlin.

Mohammadzadeh, A., Tavakoli, A., and Valadanzoej, M. 2006. Road extraction based on fuzzy logic and mathematical morphology from pansharpened IKONOS images. The Photogrammetric Record, 21:44–60.

Moshiri, B., Asharif, M. R., and Hosein Nezhad, R. 2002. Pseudo information measure : a new concept for extension of Bayesian fusion in robotic map building. Inf. Fusion, 3(1):51–68.

Nencini, F., Garzelli, A., Baronti, S., and Alparone, L. 2007. Remote sensing image fusion using the curvelet transform. Inf. Fusion, 8:143–156.

Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W. 1996. *Applied Linear Statistical Models*. McGraw-Hill/Irwin. 1408pp.

Nishii, R., Kusanobou, S., and Tanaka, S. 1996. Enhancement of low resolution image based on high resolution bands. IEEE Trans. Geosci. Remote Sens., 34:1151–1158.

ORFEO Accompaniment Program. 2007. ORFEO Tool Box. `http://smsc.cnes.fr/PLEIADES/A_prog_accomp.htm`.

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

Pardo-Iguzquiza, C., Chica-Olmo, M., and Atkinson, P. M. 2006. Downscaling cokriging for image pansharpening. Remote Sens. Environ., 102(1-2):86–98.

Pieczynski, W., Bouvrais, J., and Michel, C. 1998. Unsupervised Bayesian Fusion of Correlated Sensors. In *Proceedings of the First International Conference on Multisource-Multisensor Information Fusion (Fusion'98)*, pages 794–801. Las Vegas, Nevada.

Pinheiro, P. and Lima, P. 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.

Pohl, C. and van Genderen, J. L. 1998. Multisensor image fusion in remote sensing : concepts, methods and applications. Int. J. Remote Sens., 19(5):823–854.

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

Qu, G., Zhang, D., and Yan, P. 2002. Information measure for perfomance of image fusion. Electron. Lett., 38:313–315.

R. 2006. The R Project for Statistical Computing. `http://www.r-project.org/`.

Raggam, H., Franke, M., Ofner, M., and Gutjahr, K. 2005. Accuracy Assessment of Vegetation Height Mapping Using Spaceborne IKONOS as well as Aerial UltraCam Stereo Images. volume EARSeL workshop on 3D remote sensing.

Rajan, D. and Chaudhuri, S. 2002. Data fusion techniques for superresolution imaging. Inf. Fusion, 3(1):25–38.

Ranchin, T. and Wald, L. 2000. Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. Photogram. Eng. Remote Sens., 66:49–61.

Ranchin, T., Aiazzi, B., Alparone, L., Baronti, S., and Wald, L. 2003. Image fusion-the ARSIS concept and some successful implementation schemes. ISPRS J. Photogram. Remote Sens., 58(1-2):4–18.

Rejas Ayuga, J., Burillo Mozota, F., López, R., and Farjas Abadía, M. 2006. Application of hyperspectral remote sensing to the Celtiberian city of Segeda. In *Presented at the 2nd International Conference on Remote Sensing in Archeology, Rome, December 2006*.

Ren, W., Mclennam, J., Leuangthong, O., and Deutsch, C. 2006. Reservoir characterization of McMurray formation by 2D geostatistical modeling. Nat. Resour. Res., 15(2):111–117.

Ross, A. and Jain, A. 2003. Information fusion in biometrics. Pattern Recogn. Lett., 24:2115–2125.

Rässler, S. 2004. Data fusion: Identification problems, validity, and multiple imputation. Austrian Journal of Statistics, 33:153–171.

Savelievaa, E., Demyanova, V., Kanevski, M., Serre, M., and Christakos, G. 2005. BME-based uncertainty assessment of the Chernobyl fallout. Geoderma, 128:312–324.

Serre, M. L. and Christakos, G. 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.

Shannon, C. E. 1948. A mathematical theory of communication. Bell system Tech. J., 27:379–423.

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

Sohn, S. and Lee, S. 2003. Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea. Saf. Sci., 41:1–14.

Song, X., Abu-Mostafa, Y., Sill, J., Kasdan, H., and Pavel, M. 2003. Robust image recognition by fusion of contextual information. Inf. Fusion, 3:277–287.

Souza, C. M. and Roberts, D. 2005. Mapping forest degradation in the Amazon region with IKONOS images. Int. J. Remote Sens., 26:425–429.

Sveshnikov, A. 1979. *Problems in probability theory, mathematical statistics and theory of random functions*. Dover Publications, Inc., New York.

Tu, T.-M., Su, S.-C., Shyu, H.-C., and Huang, P. S. 2001. A new look at IHS-like image fusion methods. Inf. Fusion, 2:177–186.

Tuia, D., Fasbender, D., Kanevski, M., and Bogaert, P. 2007. Bayesian data fusion for image enhancement: an application for thermal infrared ASTER sensor. In *Presented at the URBAN/URS joint event 2007, Paris, France*.

van der Putten, P., Kok, J., and Gupta, A. 2002. Why the information explosion can be bad for data mining, and how data fusion provides a way out. In *Proceedings of the Second SIAM International Conference on Data Mining*. Arlington, USA.

Varshney, P. 1997. Multisensor data fusion. Electron. and Commun. Eng. J., 6(9):245–253.

Wackernagel, H. 1995. *Multivariate Geostatistics*. Springer-Verlag. 291pp.

Wang, Z., Ziou, D., Armenakis, C., Li, D., and Li, Q. 2005. A comparative analysis of image fusion methods. IEEE Trans. Geosci. Remote Sens., 43(6):1391–1402.

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

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. 2001. Spatial-temporal hierarchical Bayesian modeling: Tropical ocean surface winds. J. Am. Stat. Assoc., 96(454):382–397.

Yamamoto, J. 1999. Quantification of uncertainty in ore-reserve estimation : Applications to Chapada Copper Deposit, State of Goiàs, Brazil. Nat. Resour. Res., 8(2):153–163.

Yamazaki, F., Matsuoka, M., Warnitchai, P., Polngam, S., and Ghosh, S. 2005. Tsunami Reconnaissance Survey in Thailand Using Satellite Images and GPS. Asian J. Geoinformatics, 5:53–61.

Zhang, Y. 1999. A new merging method and its spectral and spatial effects. Int. J. Remote Sens., 20:2003–2014.

Zhang, Y. and Hong, G. 2005. An IHS and wavelet integrated approach to improve pansharpening visual quality of natural colour IKONOS and QuickBird images. Inf. Fusion, 6(3):225–234.

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

Zhou, J., Civco, D. L., and Silander, J. A. 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. Int. J. Remote Sens., 19(4):743–757.

Titres récents de la collection

2005

Département de biologie appliquée et des productions agricoles

- N°53. Nkunzimana, Tharcisse, Une filière agro-industrielle en mutation. Cas de la filière théicole au Burundi.
- N°55. Meunier, Alexandre, Detection of rhizomania in Belgium: Diversity and characterization of Beet necrotic yellow vein virus RNAs-2 and -3.
- N°56. Grissa, Kaouthar, Étude de la dynamique des populations et du contrôle des acariens ravageurs dans les vergers de pommiers tunisiens.
- N°62. Rubayiza, Aloys, Authentication and quality control of coffee by fluorescence and Raman spectroscopy.
- N°68. Odjo, Sunday Pierre, Modélisation en équilibre général non séparable et analyse de réformes agricoles et commerciales au Bénin.
- N°69. Célestin, Dario Styve, Stabilisation du secteur caféier en Haïti par la promotion de la concurrence et la mise en œuvre d'un plan d'assurance pour la gestion des risques agricoles.

Département de chimie appliquée et des bio-industries

- N° 57. Vermeulen, Catherine, Synthèse et caractérisation organoleptique de thiols polyfonctionnels Recherche de leur présence dans la bière.

- N°58. Van Wilder, Valérie, Phosphorylation des aquaporines de la membrane plasmique de maïs.

- N°64. Bertinchamps, Fabrice, Total oxidation of chlorinated VOCs on supported oxide catalysts.

- N°65. Demoulin, Olivier, Dynamic aspects of the surface of a palladium on alumina catalyst used in combustion of methane.

- N°67. Trincal, Mathieu, Grx5p, a mitochondrial glutaredoxin required for mitochondrial DNA maintenance.

- N°70. Dury, Frédéric, Coupling of the deoxygenation of benzoic acid with the oxidation of propylene as a new tool to elucidate the architecture of Mo-based oxide catalysts.

Département des sciences du milieu et de l'aménagement du territoire

- N°54. Romanowicz, Agnieszka Aleksandra, The impacts of environmental change on water resources: A case study in the Dyle catchment (Belgium) in support of the implementation of the Water Framework Directive.

- N°59. Delfosse, Thomas, Acid Neutralisation and Sulphur Retention in S-Impacted Andosols

- N°60. Titeux, Hugues, Transfert de carbone organique dissous et de métaux dans des sols forestiers acides: influence du fonctionnement de la litière et des réserves minérales du sol.

- N°61. Blaes, Xavier, Crop monitoring by advanced SAR remote sensing techniques : modelling and experimental analysis.

- N°63. Farcy, Christine, L'aménagement forestier à la croisée des chemins. Éléments de réponse au défi posé par les nouvelles attentes d'une société en mutation. N°66. Jonard, Mathieu, Dynamique des litières foliaires en peuplements purs et mélangés de chêne et de hêtre : Retombées foliaires et premières étapes de la décomposition.

2006

Département de biologie appliquée et des productions agricoles

- N°71. Muratori, Frédéric, Reconnaissance des facteurs cuticulaires de l'hôte et conséquences sur l'exploitation des colonies par le parasitoïde de pucerons, Aphidius rhopalosiphi (Hymenoptera: Aphidinae).

- N°72. Condori Ali, Paulino Bruno, Analyse comparative et modélisation de la croissance et du développement de tubercules andins dans les Andes en Bolivie.

- N°75. Doucet, Diane, Development of a model and similarities with a benyvirus.

- N°76. Nkwembe Unsital, Guy-Bernard, La problématique de la pauvreté des ménages agricoles ruraux et urbains dans la périphérie de la ville de Kinshasa. Essai d'analyse du phénomène et de ses implications sur la sécurité alimentaire.

- N°78. Ndayiragjie, Alexis, Relations entre métabolisme des polyamines et la résistance au stress salin chez le riz (Oryza sativa L.).

- N°80. Secheli Ringot, Diana, Risk analysis of ochratoxin A and developpement of a decontamination procedure based on the biosorption of ochratoxin A onto yeast industry derivatives.

- N°82. Abboudi, Tarik, The indispensable amino acid requirements for maintenance in Atlantic salmon fry.

- N°83. Martins da Silva, Evaldo, Polyphenols from the Amazonian plant Inga edulis: process optimization for the production of purified extracts with high antioxidant capacity.

- N°84. Fadlaoui, Aziz, Modélisation bioéconomique de la conservation des ressources génétiques animales.

- N°85. Kambale Valimunzigha, Charles, Étude du comportement physiologique et agronomique de la tomate (Solanum lycopersicum L.) en réponse à un stress hydrique précoce.

- N°86. Lepoint, Pascale, Speciation within the African Coffee Wilt Pathogen.

- N°88. Mushagalusa Nachigera, Gustave, Competitive relationships in potato/maize intercropping: involvement of root system architecture
- N°89. Nyamangyoku Ishibwela, Obedi, Ferrous iron toxicity: Mechanisms of resistance and impact on nutrient elements at the vegetative stage in cultivated rices (Oryza sativa L., Oryza glaberrima Steud and interspecific hybrids)

Département de chimie appliquée et des bio-industries

- N°73. Pety de Thozée, Cédric, Implication of the ER quality control in the degradation of the yeast plasma membrane Pdr5 L183P ABC transporter.
- N°79. Gurdak, Elzbieta, Supramolecular organization of collagen layers adsorbed on polymers.
- N°87. Janssens, Xavier, Monitoring and predicting elusive species colonisation - Application to the otter in the Cévennes National Park (France).

Département des sciences du milieu et de l'aménagement du territoire

- N°74. Kayitakire, François, Forest stand characterisation using very high resolutions satellite remote sensing
- N°77. Van der Velde, Marijn, Agricultural and climatic impacts on the groundwater resources of a small island: measuring and modelling water and solute transport in soil and groundwater on Tongatapu.
- N°81. Degen, Thomas, Dynamique initiale de la végétation herbacée et de la régénération ligneuse dans le cas de trouées, au sein d'une hêtraie (Luzulo-Fagetum).

2007

Département de biologie appliquée et des productions agricoles

- N°97, De Dorlodot, Sophie, Root system architecture : a genetic analysis in rice.

- N°99, Daoui, Khalid, Recherche de stratégies d'amélioration de l'efficience d'utilisation du phosphore chez la fève (Vicia faba L.) dans les conditions d'agriculture pluviale au Maroc.
- N°101, Vanloqueren, Gaëtan, Penser et gérer l'innovation en agriculture à l'heure du génie génétique . Contribution d'une approche systémique d'innovations scientifiques dans deux filières agroalimentaires wallonnes pour l'évaluation, la gestion et les politiques d'innovation.
- N°109, Pairon, Marie, Ecology and population genetics of an invasive forest tree species: Prunus serotina Ehrh.
- N°111, Manyonga, Madika, Modelling the competition for light in maize (Zea mays L.)/bean (Phaseolus vulgaris L.) intercropping with three-dimensional approach.
- N°115, Silva de Souza, Jesus Nazareno, Etude des propriétés antioxydantes in vitro d'extraits de feuilles de Byrsonima crassifolia et Inga edulis et caractérisation partielle des composés phénoliques.
- N°116, Colinet, Hervé, Une approche écologique et biochimique de la résistance au froid chez un parasitoïde de puceron Aphidius colemani (Hymenoptera: Aphidiinae).

Département de chimie appliquée et des bio-industries

- N°90. Eyers, Laurent, Characterization of bacterial populations of 2,4,6-trinitrotoluene (TNT) contaminated soils and isolation of a Pseudomonas aeruginosa strain with TNT denitration activities.
- N°91. Dupré de Boulois, Hervé, Role of arbuscular mycorrhizal fungi on the accumulation of radiocaesium by plants.
- N°92. Crouzet, Jérôme, Expression pattern, ectopic expression and gene silencing of two Nicotiana Pleiotropic Drug Resistance transporters.
- N°93, Drugmand, Jean-Christophe, Characterization of Insect Cell Lines Is Required for Appropriate Industrial Processes.
- N°94, De La Providencia, Ivan Enrique, Contribution of Anastomosis and Hyphal Healing Mechanism to the Extraradical Mycelium Architecture of Arbuscular Mycorrhizal Fungi.

- N°95, Sebari, Karima, Optimisation de la gestion conjointe des ressources en eau de surface et des ressources en eau souterraines dans une région semi aride - la Vallée du Drâa.
- N°100, Vandermeeren, Caroline, Quaternary Structure and Regulation of the Nicotiana plumbaginifolia Plasma Membrane Proton Pump ATPase.
- N°102, Morales Belpaire Isabel, Fate of the active form of proteins in selected environmental matrices. Investigation with green fluorescent protein, b-glucosidase, and amyloid fibrils in wastewater sludges, biofilms and soils.
- N°103, Trombik Tomasz, Characterization of two Pleiotropic Drug Resistance transporters from Nicotiana plumbaginifolia.
- N°104, Callemien, Delphine, Use of new methodologies to study phenolic compounds through beer storage.
- N°106, Van der Auwera, Géraldine, pAW63: A molecular wanderer in the Bacillus cereus gene pool.
- N°107, Mateos Pedrero, Cecilia, Synthesis and characterization of Rh supported on Ti-modified oxides: catalytic activity in the partial oxidation of methane.
- N°108, Voets, Liesbeth, Role of arbuscular mycorrhizal networks on plant interconnection and carbon transfer.
- N°110, Jerkovic, Vesna, Découverte du resvératrol dans les houblons. Impact de la variété, de l'année de récolte et du conditionnement en vue de la préparation d'extraits polyphénoliques enrichis en stilbènes.
- N°112, Dekeyser, Caroline, Elaboration and evaluation of nanostructured surfaces for the control of cell behavior.
- N°113, Nonckreman, Cristèle, Design of materials with nanostructured surface to improve hemocompatibility.
- N°114, Caillou, Samuel, Protein Adsorption: from Interfacial Interactions to Enzyme Activity.

Département des sciences du milieu et de l'aménagement du territoire

- N°96, Lefèvre, Isabelle, Investigation of three Mediterranean plant species suspected to accumulate and tolerate high cadmium and zinc levels: morphological, physiological and biochemical characterization under controlled conditions.
- N°98, Desclée, Baudouin, Automated object-based change detection for forest monitoring by satellite remote sensing: applications in temperate and tropical regions.
- N°105, André, Frédéric, Influence of the heterogeneity of canopy structure on the spatio-temporal variability of atmospheric deposition within a mixed oak-beech stand

2008

Département de biologie appliquée et des productions agricoles

- N°120, Chirinos Gallardo, Rosana Sonia, Polyphenols from the Andean mashua (Tropaeolum tuberosum) tuber: Evaluation of genotypes, extraction, chemical characterization and antioxidant properties.
- N°126, Romier, Béatrice, Modulation of intestinal inflammation by polyphenols.
- N°130, Vaïanopoulos, Céline, Widespread occurrence of soil-borne mosaic viruses in Belgium. Preferential associations of Polymyxa graminis and mosaic viruses on cereals.
- N°131, Tran, Thi Nang Thu, Nutritional value of sesame oil cake and lysine utilization efficiency in plant protein-based diets for rainbow trout.
- N°134, Ngirente, Edouard, Les marchés agricoles intérieurs et les mutations agro- économiques en milieu rural rwandais. Application à la filière pomme de terre

Département de chimie appliquée et des bio-industries

- N° 117, Zelazny, Enric, Regulation of maize aquaporin trafficking to the plasma membrane.
- N°122, Bobik, Krzysztof, The two major tobacco plasma membrane H+-ATPases, PMA2 and PMA4 are differentially phosphorylated on their penultimate threonine.
- N°123, Cabana, Hubert, Elimination des perturbateurs endocriniens nonylphénol, bisphénol A et triclosan à l'aide de laccase de Coriolopsis polyzona. C
- N°133, Delannoy, Mélanie, Identification of peptidases in the Nicotiana tabacum leaf intercellular fluid.
- N°132, De Palmenaer, Daniel, Genomic portrayal of IS4 family - Modular insertion sequences, mobile insertion cassettes and overall family snapshot.
- N°135, Manente, Myriam, Rta1, a yeast plasma membrane protein which confers resistance to fungicides.
- N°129, Srasra, Mondher, Preparation and characterization of new basic catalyst by nitridation of Y zeolites.
- N°137, Hachez, Charles, The expression pattern and activity of Zea mays plasma membrane aquaporins highlight their central role in the control of the cell membrane water permeability.

Département des sciences du milieu et de l'aménagement du territoire

- N°118, Van Cauwenbergh, Nora, Expert and local knowledge in decision support for natural resource management.
- N°119, Henriet, Céline, Silicon in banana (Musa spp.): a soil-plant system approach.C
- N°121, Tidjani, Adamou Didier, Erosion éolienne dans le Damagaram Est (Sud-Est du Niger) : Paramétrisation, quantification et moyens de lutte.
- N°124, Akponikpe, Pierre Bienvenu Irénikatché, Millet response to water and soil fertility management in the Sahelian Niger: experiments and modeling Erosion éolienne dans le Damagaram Est

(Sud-Est du Niger) : Paramétrisation, quantification et moyens de lutte.

- N°125, Vancutsem, Christelle, Revisiting satellite time series compositing for earth observation applications.
- N°127, Opfergelt, Sophie, Silicon cycle in the soil-plant system: biogeochemical tracing using Si isotopes. N°128, de Araujo Marinho Filho, Eduardo Celsio, Economics of Hunger: Several Methods, Levels and Scales.
- N°136, Saqalli, Mehdi, Populations, Farming systems & Social transitions in Sahelian Niger: An Agent-based modeling approach.

Pour plus d'informations, visitez http://edoc.bib.ucl.ac.be/