

## Bayesian data fusion for space–time prediction of air pollutants: The case of NO<sub>2</sub> in Belgium

D. Fasbender <sup>a,\*</sup>, O. Brasseur <sup>b</sup>, P. Bogaert <sup>a</sup>

<sup>a</sup> Department of Environmental Sciences and Land Use Planning, Université catholique de Louvain, Croix du Sud 2/16, B-1348 Louvain-la-Neuve, Belgium

<sup>b</sup> Interregional Cell for the Environment (IRCEL), Kunstlaan 10-11, B-1210 Brussels, Belgium

### ARTICLE INFO

#### Article history:

Received 3 December 2008

Received in revised form

16 May 2009

Accepted 18 May 2009

#### Keywords:

Data merging

Kriging

Stochastic method

### ABSTRACT

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 paper 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<sub>2</sub>) 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<sub>2</sub> predictions.

© 2009 Elsevier Ltd. All rights reserved.

### 1. 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 gases 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. Some approaches such as the Kalman filter (see e.g. Kalman, 1960) or the Fokker–Planck equations (see e.g. Risken, 1996) provide alternative solutions to physically-based models by building the state of the process (either the state itself or

\* Corresponding author.

E-mail addresses: [Dominique.Fasbender@uclouvain.be](mailto:Dominique.Fasbender@uclouvain.be) (D. Fasbender), [Brasseur@irceline.be](mailto:Brasseur@irceline.be) (O. Brasseur), [Patrick.Bogaert@uclouvain.be](mailto:Patrick.Bogaert@uclouvain.be) (P. Bogaert).

its probability density function (pdf) with deterministic dynamic equations. These kinds of approaches generally provide accurate results even with quite simple hypotheses. Stochastic approaches are also 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; Christakos and Serre, 2000; Christakos et al., 2001) are possible applications of geostatistics in environmental sciences. It is also important to note that the border between deterministic and stochastic approaches is not sharp as it often happens that stochastic approaches include deterministic part in their equations (e.g. regression methods) and reciprocally (e.g. Fokker–Planck equations).

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 paper, 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 paper and after a short description of the data set, the case of nitrogen dioxide ( $\text{NO}_2$ ) 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 paper 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.

## 2. Bayesian data fusion

Combining multiple information sources into a single final prediction (i.e. data fusion) is not a new problem and is not restricted to environmental sciences; 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., 2008a,b,c, 2007) 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.

### 2.1. General formulation

Let us define  $\{\mathbf{x}_0, \dots, \mathbf{x}_n\}$  as the set of space-time locations (i.e. each  $\mathbf{x}_i$  is composed by a spatial component  $\mathbf{z}_i$  and a temporal component  $t_i$ ) where indirect observations  $\mathbf{y}' = (\mathbf{y}'_0, \dots, \mathbf{y}'_n)$  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  $\mathbf{Z}' = (Z_0, \dots, Z_n)$  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

$$Y_{ij} = g_{ij}(\mathbf{Z}) + E_{ij} \quad (1)$$

where  $g_{ij}$ 's are functionals and where  $E_{ij}$ '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

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

where  $f_{\mathbf{Z}}(\cdot)$  is the *a priori* pdf for  $\mathbf{Z}$  and  $f_{\mathbf{E}}(\cdot)$  is the pdf of the errors  $\mathbf{E}' = (\mathbf{E}'_1, \dots, \mathbf{E}'_n)$ . 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  $\mathbf{x}'_S = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$  where both  $Z_i$ 's and  $Y_{ij}$ 's are jointly sampled, and  $\mathbf{Z}'_U$  refers to space-time locations  $\mathbf{x}'_U = \{\mathbf{x}_{m+1}, \dots, \mathbf{x}_n\}$  where only  $Y_{ij}$ '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

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

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

$$\begin{aligned} f(z_0|\mathbf{z}'_S, \mathbf{y}) &\propto \prod_{i=0}^m \prod_{j=1}^{n_i} f_{E_{ij}}(y_{ij} - g_{ij}(z_i)) \int f_{\mathbf{Z}}(\mathbf{z}) \prod_{i=m+1}^{m+n} \\ &\times \prod_{j=1}^{n_i} f_{E_{ij}}(y_{ij} - g_{ij}(z_i)) d\mathbf{z}'_U \end{aligned} \quad (4)$$

$$\propto \prod_{i=0}^m \prod_{j=1}^{n_i} \frac{f(z_i|y_{ij})}{f(z_i)} \int f_{\mathbf{Z}}(\mathbf{z}) \prod_{i=m+1}^{m+n} \prod_{j=1}^{n_i} \frac{f(z_i|y_{ij})}{f(z_i)} d\mathbf{z}'_U \quad (5)$$

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

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

so that using either distributions of errors  $f_{E_{ij}}(\cdot)$  or conditional distributions  $f(z_i|y_{ij})$  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  $f_Z(\mathbf{z})$ , whereas the second one integrates the various auxiliary information sources through the univariate conditional distributions  $f(z_i|y_{ij})$  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).

Finally, it is also worth noting that, in most of the applications, both the conditional pdf and the *prior* distribution are not known and should therefore be estimated. There is thus uncertainties linked to these estimations which should ideally be accounted for in the total uncertainty of the model. Incidentally, Bayesian frameworks theoretically enable to account for the uncertainty of parameters by multiplying with a *prior* distribution for the parameters and by integrating the product over the possible values of the parameters, similarly to unobserved variables  $\mathbf{Z}_U$  in Eqs. (4) and (5). As a consequence, the conditional pdf  $f(z_0|\mathbf{z}_S, \mathbf{y})$  in Eqs. (4) and (5) become respectively

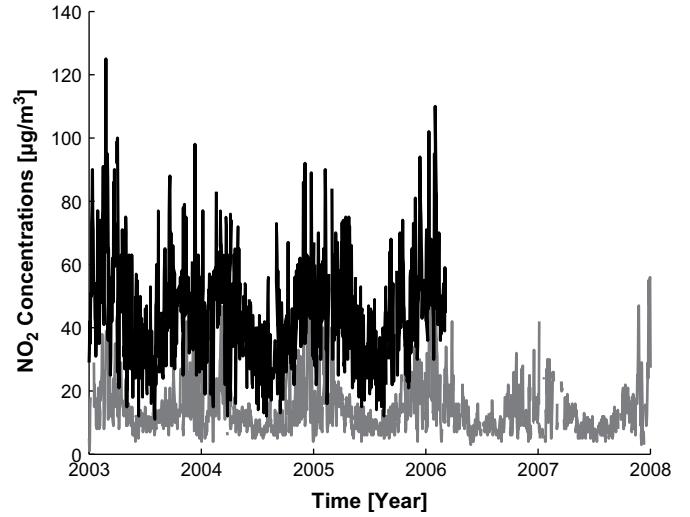
$$f(z_0|\mathbf{z}_S, \mathbf{y}) \propto \int \prod_{i=0}^m \prod_{j=1}^{n_i} f_{E_{ij}}(y_{ij} - g_{ij}(z_i)|\chi) \int f_Z(\mathbf{z}|\chi) \prod_{i=m+1}^{m+n} \times \prod_{j=1}^{n_i} f_{E_{ij}}(y_{ij} - g_{ij}(z_i)|\chi) f_\chi(\chi) d\mathbf{z}_U d\chi \quad (6)$$

$$\propto \int \prod_{i=0}^m \prod_{j=1}^{n_i} \frac{f(z_i|y_{ij}, \chi)}{f(z_i|\chi)} \int f_Z(\mathbf{z}|\chi) \prod_{i=m+1}^{m+n} \times \prod_{j=1}^{n_i} \frac{f(z_i|y_{ij}, \chi)}{f(z_i|\chi)} f_\chi(\chi) d\mathbf{z}_U d\chi \quad (7)$$

where  $\chi$  denotes the vector of the whole set of estimated parameters involved in the model. However, although Eqs. (6) and (7) are more general than Eqs. (4) and (5), the latter were implemented here in this paper as the objectives of this paper were to provide an alternative approach for accounting for several information sources in the context of air pollutants space-time prediction, leaving thus this uncertainty issue to further researches on the subject. As a final note, there exist interesting approaches in the literature that use Bayesian framework for accounting for parameters uncertainty (see e.g. Sahu et al., 2007 and other related works) which will provide a strong starting point for coping with this issue within the proposed BDF framework.

## 2.2. 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  $Z(\ell, t)$  is expected to vary along with location (largely due, but to an unknown extent, to the proximity of emission sources; see Fig. 1 in the case of  $\text{NO}_2$  in Belgium).



**Fig. 1.** Illustration of the non-stationarity in space of the  $\text{NO}_2$  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.

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. 2 in the case of  $\text{NO}_2$  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  $f_Z(\mathbf{z})$  and  $f(z_i)$  in Eq. (5). In this paper, for each location, the time-varying mean of air pollutant concentrations  $\mu(\ell, t)$  was modeled using the expression

$$\mu(\ell, t) = \mathbb{E}[Z(\ell, t)] = W(\ell, t) + S(\ell, t) \quad (8)$$

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

$$\sigma^2(\ell) = \text{Var}(Z(\ell, t)) = \mathbb{E}[(Z(\ell, t) - \mu(\ell, t))^2], \quad \forall t \in \mathbb{R} \quad (9)$$

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  $\text{NO}_2$  application (see Fig. 1).

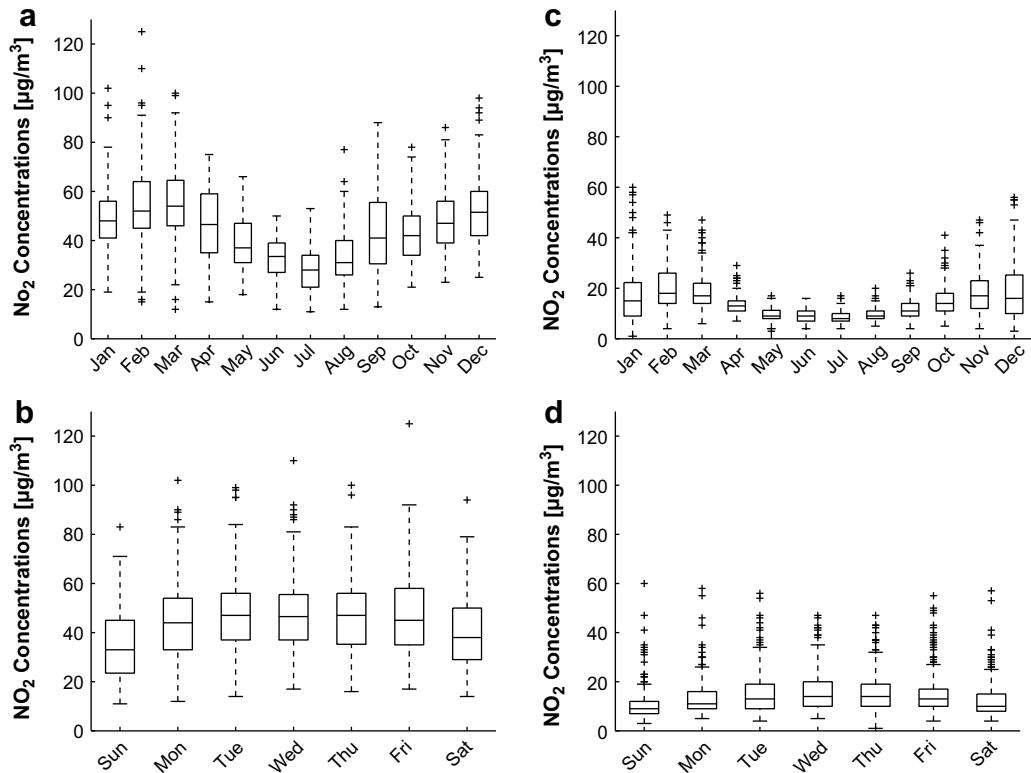
Now using Eqs. (8) and (9), it is clear that

$$R(\ell, t) = \frac{Z(\ell, t) - \mu(\ell, t)}{\sigma(\ell)} \quad (10)$$

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  $R(\ell, t)$  instead of using  $Z(\ell, t)$ . Based on this estimation, the following space-time model (Bogaert, 1996) is fitted

$$\Gamma_R(\mathbf{x}_i, \mathbf{x}_j) = \gamma_{ST}(\Delta_\ell, \Delta_t) + \gamma_S(\Delta_\ell) + \gamma_T(\Delta_t) \quad (11)$$

where  $\gamma_{ST}(\cdot, \cdot)$  is a non-separable space-time semi-variogram model,  $\gamma_S(\cdot)$  is a purely spatial semi-variogram model and  $\gamma_T(\cdot)$  is a purely temporal semi-variogram model, with  $\Delta_\ell = \|\ell_i - \ell_j\|_2$  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  $\Delta_\ell$  and  $\Delta_t$  (such as can be seen in Fig. 5 for the  $\text{NO}_2$  application). Using this final model, it is easy to compute the covariance for any couple  $(Z_i, Z_j)$  as



**Fig. 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. 1) and in Virton respectively.

$$\text{Cov}(Z_i, Z_j) = \sigma(\ell_i)\sigma(\ell_j) - \sigma(\ell_i)\sigma(\ell_j)\Gamma_R(\mathbf{x}_i, \mathbf{x}_j) \quad (12)$$

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  $\sigma(\ell_i)$ . Distribution  $f_z(\mathbf{z})$  in Eq. (5) are also assumed to be Gaussian with mean vector equal to  $(\mu(\mathbf{x}_0), \dots, \mu(\mathbf{x}_{m+n}))'$  and covariance matrix with elements corresponding to  $\text{Cov}(Z_i, Z_j)$ . 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  $f_z(\mathbf{z})$  rely on the same moments constraints, the Gaussian distributions require that  $f(z_i)$  agrees with  $f_z(\mathbf{z})$  (see e.g. Papoulis, 1991; Sveshnikov, 1979). Finally, from a practical viewpoint, the restriction to Gaussian distributions in Eqs. (4) and (5) proved to be computationally efficient with convincing results (see Fasbender et al., 2008a,b,c).

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

$$f(z_i) = \int_{\mathbb{R}} f(z_i|y_{ij})f(y_{ij}) dy_{ij}, \quad \forall z_i \in \mathbb{R} \quad (13)$$

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

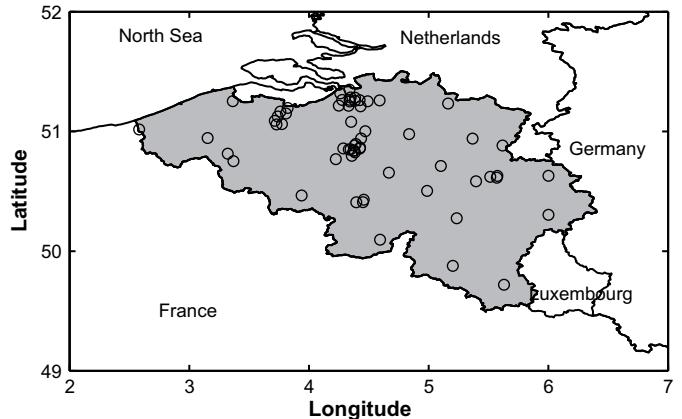
$$\text{Var}(Z_i) = \mathbb{E}_{Y_{ij}} [\text{Var}_{Z_i}[Z_i|Y_{ij}]] + \text{Var}_{Y_{ij}} [\mathbb{E}_{Z_i}[Z_i|Y_{ij}]] \quad (14)$$

This theoretical constraint might be difficult to satisfy in practice. It is worth noting here that a first consequence of Eqs. (13) and (14) is that if one wants to model  $f(z_i|y_{ij})$  for a given  $y_{ij}$  only with its conditional mean  $\mathbb{E}_{Z_i}[Z_i|Y_{ij}]$  and conditional variance  $\text{Var}_{Z_i}[Z_i|Y_{ij}]$  (e.g. if  $f(z_i|y_{ij})$  is assumed to be Gaussian), then both expressions

should be estimated in a consistent way. Appendix A shows how a lack of consistency could lead to disastrous results when using Eq. (5). So, as Gaussian distributions will be used here for aforementioned reasons (see e.g. Bogaert and Fasbender, 2007; Fasbender et al., 2008a,b,c for examples of applications), it is important to control this kind of issues in specific implementations. Appendix A also proposes an easy way that enables us to circumvent this issue in practice.

### 3. Data presentation

The pollution data at hand for this study come from the Belgian regional networks for air quality measurements. For the selected



**Fig. 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).

study period,  $\text{NO}_2$  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. 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 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 Figs. 1 and 2).

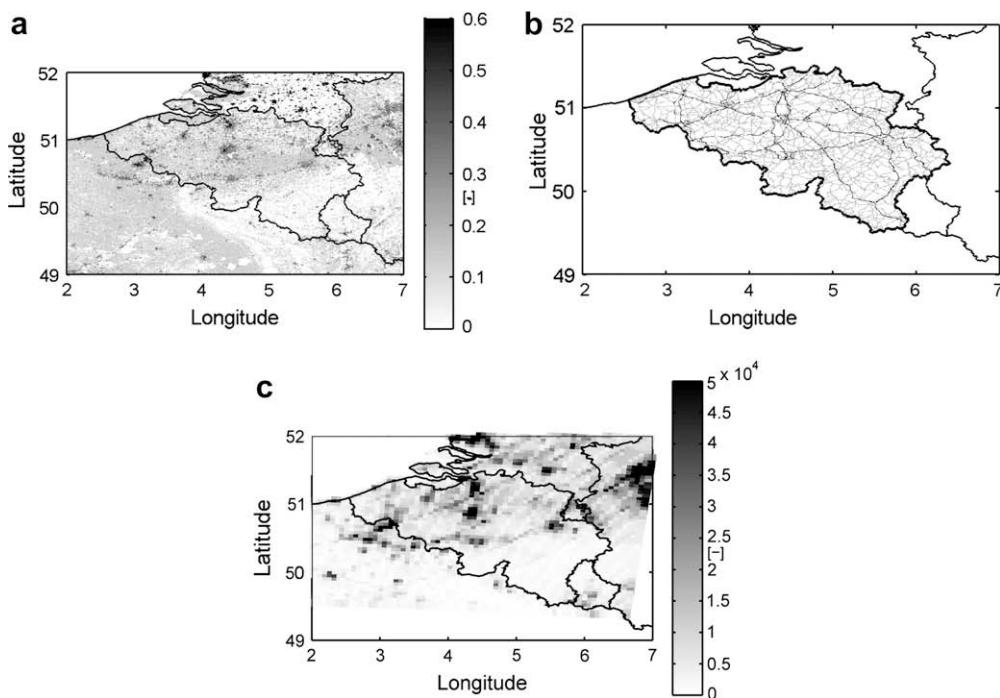
Broadly speaking, the spatio-temporal evolution of  $\text{NO}_2$  concentrations mostly depends on (i) emission sources and (ii) meteorological conditions related to the dispersion of pollutants. Concerning the meteorological aspect, three variables have been considered in this study: (i) 10 m above 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  $\text{NO}_2$  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  $\text{NO}_2$ . This index, designated by the authors as the  $\beta$  index, is computed using CORINE land cover data on local window with radius equal to 2 km. In this paper, this  $\beta$  index was computed on a 1001 by 60 regular grid with  $0.005^\circ$  between each consecutive nodes (see Fig. 4a).

The second data set is about emissions from the car traffic. Indeed, it is well known that cars eject a variety of gases in the atmosphere ( $\text{CO}_2$ ,  $\text{SO}_2$ ,  $\text{NO}_2$ , 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 eject 3–10 times more  $\text{NO}_x$  than gasoline cars. Further, approximately 20%–25% of these  $\text{NO}_x$  are in fact  $\text{NO}_2$  (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 1993–55% 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,



**Fig. 4.** Spatial illustration of the information sources related to anthropogenic activities and available for the study: (a) the  $\beta$  index proposed in Janssen et al. (2008), (b) the road network and (c) the population density.

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  $0.006^\circ$  (see Fig. 4b).

The third available data set for this study was the population density. Humans do not emit  $\text{NO}_2$  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 regular grid; see Fig. 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  $0.005^\circ$  to  $0.25^\circ$ . 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  $0.25/3$  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.  $\beta$  index, population density and road network) and coarser information sources were down-scaled 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  $0.25/3$  degree between two consecutive nodes.

#### 4. Results

As explained in Section 2.2, the first issue was to infer the space-time mean  $\mu(\ell, t)$  and the spatial variance  $\sigma^2(\ell)$  from Eqs. (8) and (9). In this paper, this was achieved in 2 steps. First, for each of the 62 locations of the data set,  $\mu(\ell, t)$  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

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

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  $\mu_{\ell, d_1(t)}$ ). Then, for the estimation of  $\mu(\ell, t)$  at other space-time locations, this first estimation was used in an inverse distance interpolation method. Similarly to  $\mu(\ell, t)$ , the local variance  $\sigma^2(\ell)$

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 2.2 and using the estimated  $\mu(\ell, t)$  and  $\sigma^2(\ell)$ , the original observations were standardized using Eq. (10) and a space-time semi-variogram was estimated. Exponential semi-variogram models were chosen for  $\gamma_{ST}(\cdot, \cdot)$ ,  $\gamma_S(\cdot)$  and  $\gamma_T(\cdot)$  and their parameters were estimated with a least squares minimization algorithm. Fig. 5 shows both the estimated semi-variogram and the fitted model.

On the basis of  $\mu(\ell, t)$  and  $\text{Cov}(Z_i, Z_j)$  (see Eq. (12)), a straightforward prediction method is provided by the kriging formalism. Indeed, kriging is an easy-to-use linear predictor that only relies on  $\mu(\ell, t)$  and  $\text{Cov}(Z_i, Z_j)$ . 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  $\mu(\ell, t)$ ,  $\text{Cov}(Z_i, Z_j)$  and observations at dates  $\{t, t-1, \dots, t-4\}$  (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 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 on January 17th 2007) due to a lack of exogenous information. Clearly, as already mentioned, air pollutant concentrations such as  $\text{NO}_2$  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 2.

#### 4.1. 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 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 m height. Indeed, wind speed and air pollutant concentrations are expected to be somewhat inversely

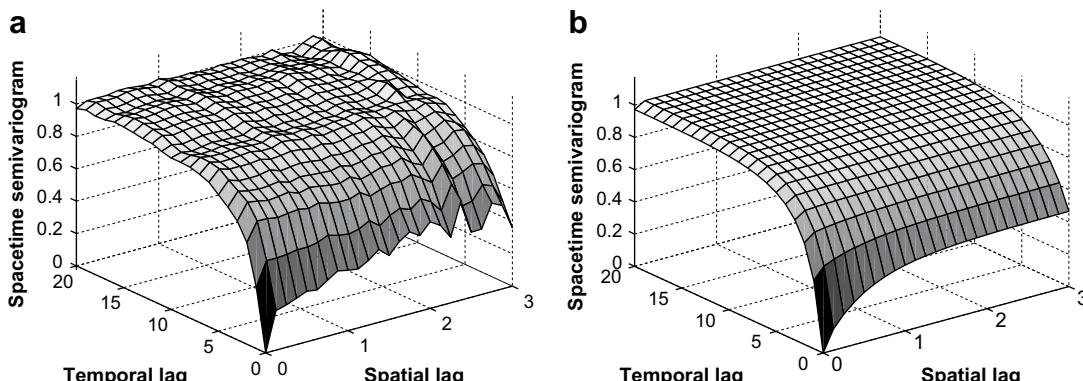
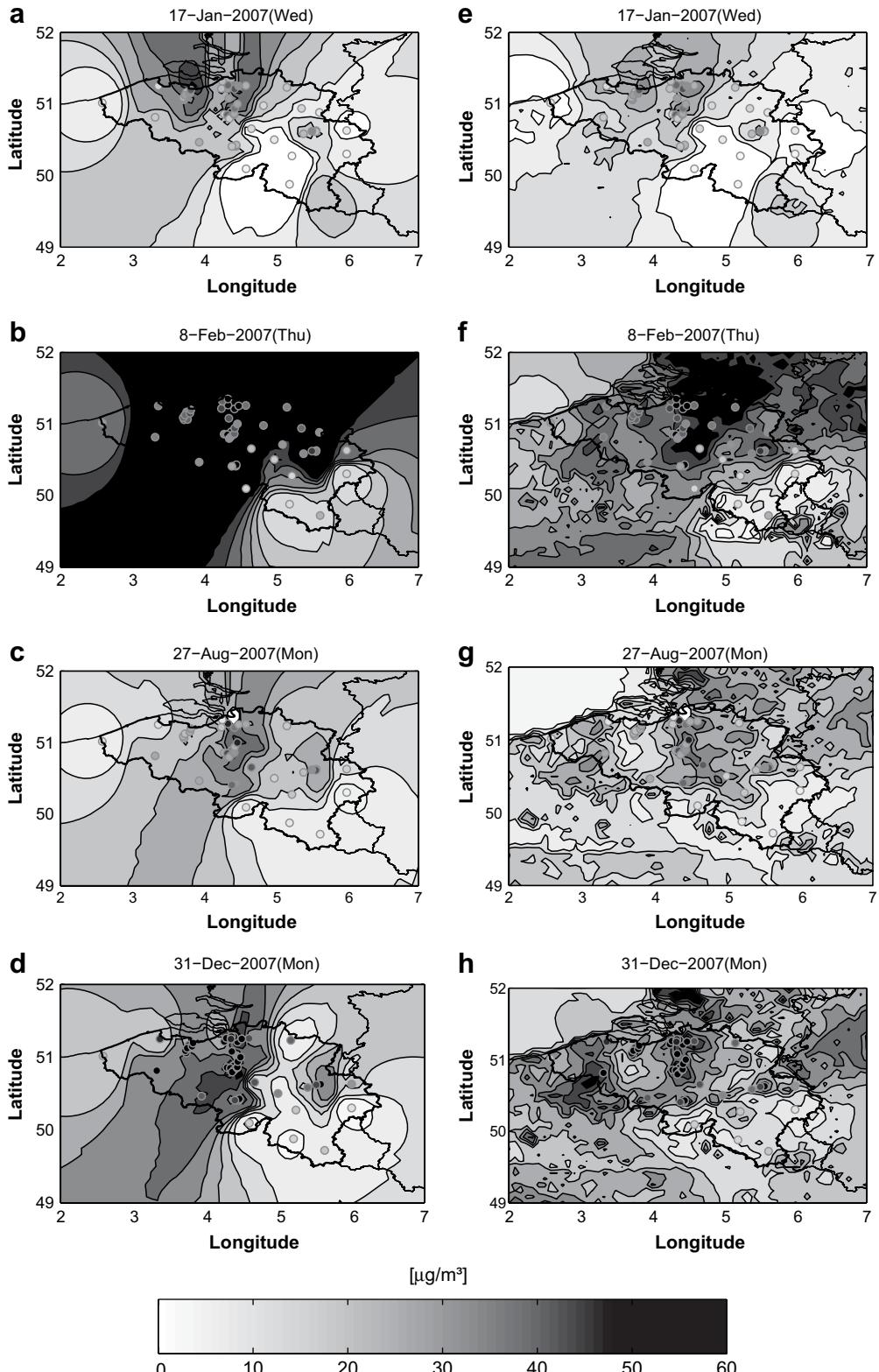


Fig. 5. Space-time semi-variogram of the standardized values  $R_i$ : (a) estimation and (b) fitted model from Eq. (11).



**Fig. 6.** Space–time predictions of daily-mean NO<sub>2</sub> concentrations. On the left column are the Simple Kriging predictions whereas Bayesian data fusion ones are on the right column.

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 A. After plotting the scatter plot between wind speed  $10U_i$  and NO<sub>2</sub> standardized observations  $R_i$  (see Eq. (10) for definition), the following empirical assumptions were used

**Table 1**

Summary table for the estimated parameters of Eq. (15).

|         | $\alpha$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\gamma$ |
|---------|----------|-----------|-----------|-----------|----------|
| $10U_i$ | -1.28    | 0.68      | -0.01     | -0.02     | -1.74    |

$$\begin{cases} \mu_{R_i|10U_i} = \alpha + e^{\beta_0 + \beta_1 10U_i + \beta_2 10U_i^2} \\ \sigma_{R_i|10U_i}^2 = e^{\gamma \mu_{R_i|10U_i}^2} \end{cases} \quad (15)$$

$\mu_{R_i|10U_i}$  and  $\sigma_{R_i|10U_i}^2$  were estimated sequentially based on non-linear least squares minimization algorithm in MATLAB (Matlab, 2002) (see Table 1 for the estimated values of these parameters). Fig. 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  $\sigma_{R_i|10U_i}^2$  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<sub>2</sub>, i.e.

$$\begin{cases} \mu_{R_i|W_i} = a + e^{b_0 + b_1 W_i} \\ \sigma_{R_i|W_i}^2 = e^{c\mu_{R_i|W_i}^2} \end{cases} \quad (16)$$

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

**Table 2**

Summary table for the estimated parameters of Eq. (16).

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

Fig. 7b and c shows the resulting models for the horizontal transfer and the turbulent kinetic energy, respectively.

#### 4.2. 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 3 and as proposed by Janssen et al. (2008), the  $\beta$  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  $\beta$  index can be used in order to better interpolate mean concentrations (i.e.  $\mu(\ell, t)$  in this paper). Again, the interpolation of  $\mu(\ell, t)$  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  $\mu(\ell, t)$  estimated at space-time sampled locations were first used to fit the parameters of a function such as in Eq. (16), then the fitted model was applied to the  $\beta$  values

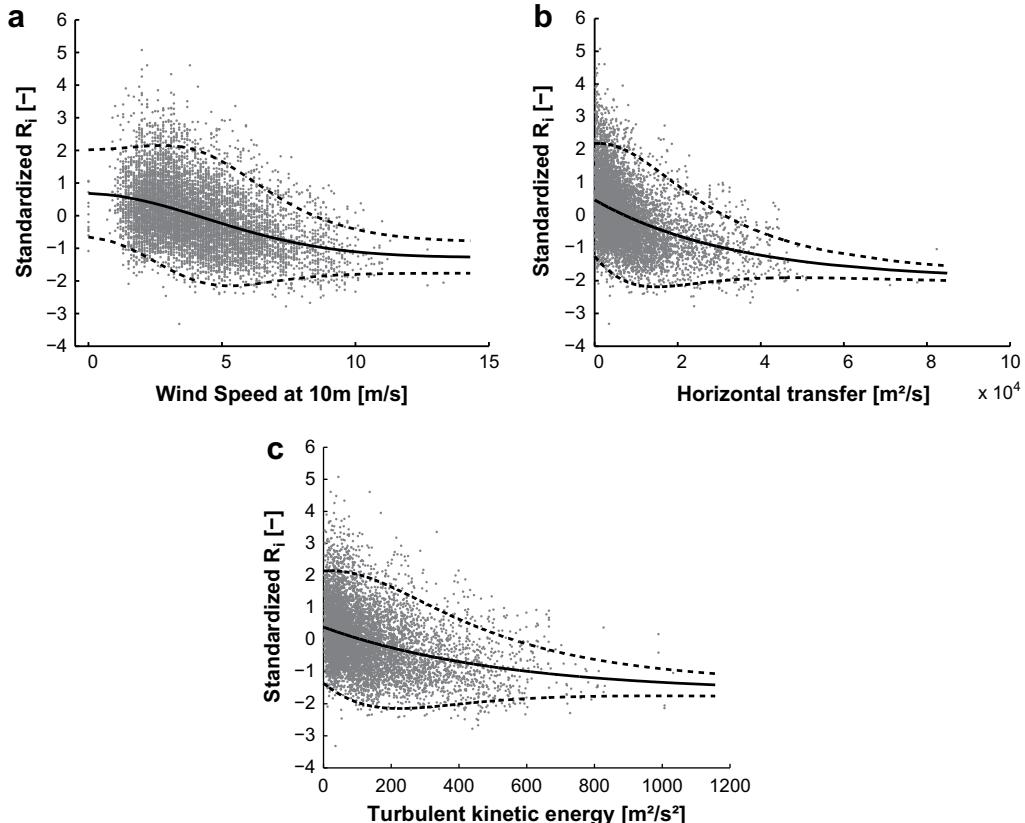


Fig. 7. Model of the conditional mean and the conditional variance of standardized NO<sub>2</sub> observations with respect to (a) wind speed in the first 10 m, (b) horizontal transfer and (c) turbulent kinetic energy. Bold line stands for estimated mean and dashed lines represent symmetric 95% intervals.

at other space-time locations where predictions are sought for. Secondly, conditional variance was estimated using the relation

$$\sigma_{\beta}^2(\ell, t) = \sigma^2(\ell, t)e^{-a(\mu_{\beta}(\ell, t) - \mu(\ell, t))^2} \quad (17)$$

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

For the case of road network and population density, the approach chosen here was very close to the one of the  $\beta$  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 were three kind of roads). Again, for the conditional variance, Eq. (17) was used with another set of parameters for each information source.

#### 4.2.1. Final prediction and validation

Using the six information sources as described in Sections 4.1 and 4.2 along with Eq. (5) and the Gaussian distribution hypotheses, the conditional mean  $M$  and the conditional variance  $V$  are

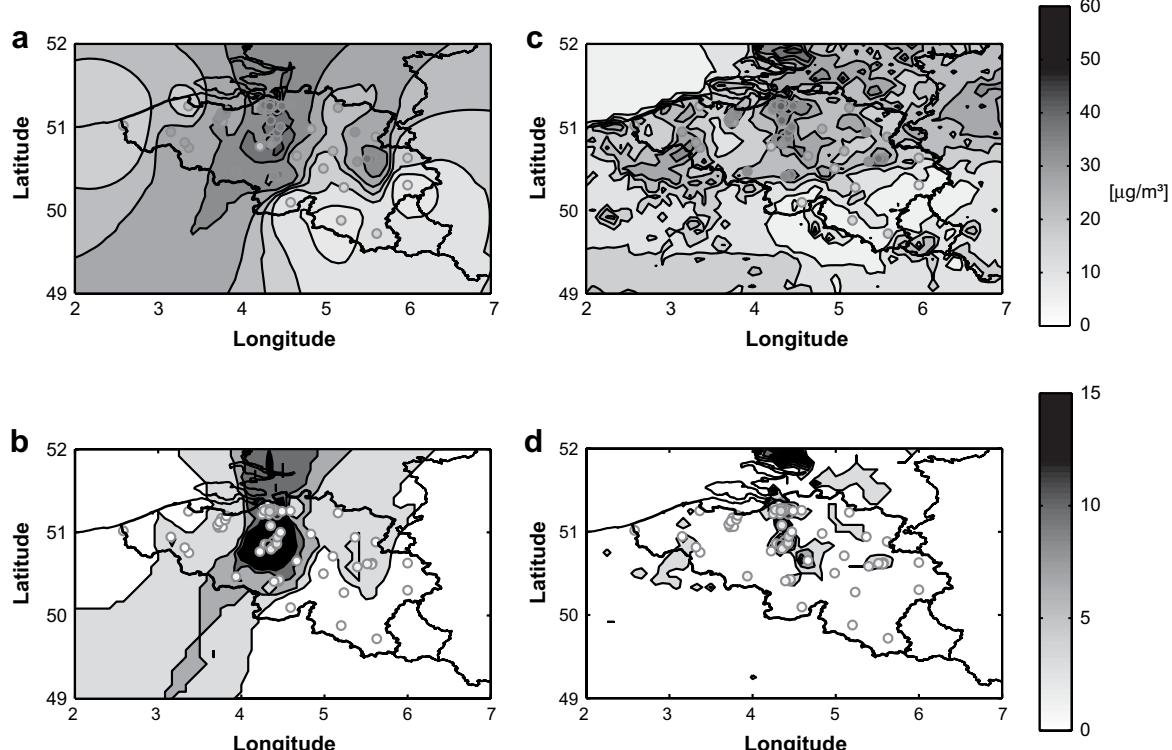
$$\left\{ \begin{array}{l} V(\ell_0, t_0) = \frac{1}{1/\sigma_{\text{krig}}^2(\ell_0, t_0) + \sum_{j=1}^{n_0} 1/\sigma_j^2(\ell_0, t_0) - n_0/\sigma^2(\ell_0)} \\ M(\ell_0, t_0) = \frac{\mu_{\text{krig}}(\ell_0, t_0) + \sum_{j=1}^{n_0} \frac{\mu_j(\ell_0, t_0)}{\sigma_j^2(\ell_0, t_0)} - \frac{n_0\mu(\ell_0, t_0)}{\sigma^2(\ell_0)}}{1/\sigma_{\text{krig}}^2(\ell_0, t_0) + \sum_{j=1}^{n_0} 1/\sigma_j^2(\ell_0, t_0) - n_0/\sigma^2(\ell_0)} \end{array} \right. \quad (18)$$

where  $\mu_{\text{krig}}(\ell_0, t_0)$  and  $\sigma_{\text{krig}}^2(\ell_0, t_0)$  are the SK prediction and its variance while  $\mu_j(\ell_0, t_0)$  and  $\sigma_j^2(\ell_0, t_0)$  are the mean and the variance of conditional distribution given the  $j$ th secondary information source (i.e. either one of the meteorological variables or one of the anthropogenic ones). It is worth noting that secondary information was accounted for at prediction location only. Predictions were made on a regular grid using the formula of  $M(\ell_0, t_0)$  in Eq.

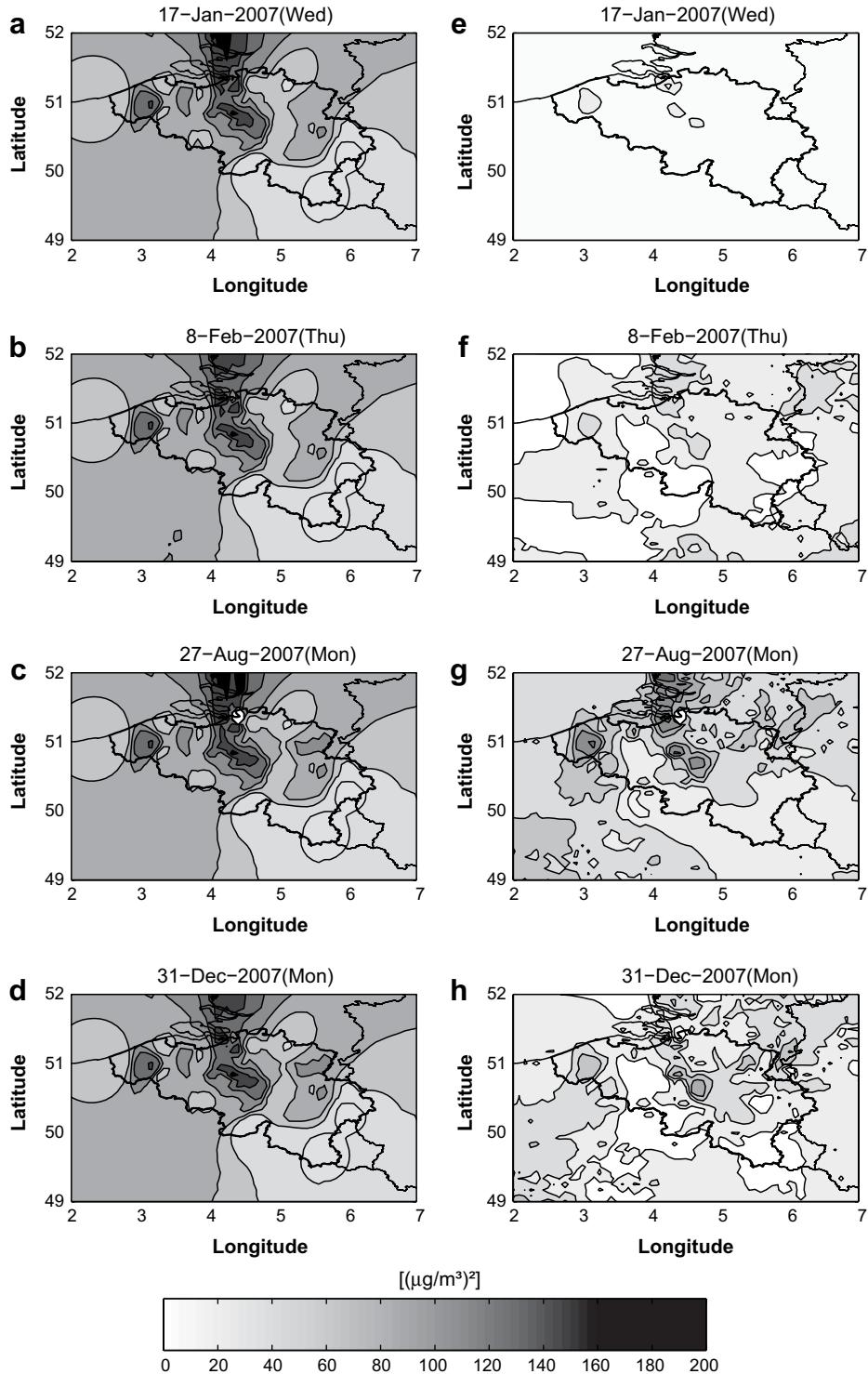
(18). Fig. 6 (right column) shows these results when using secondary information only at the prediction location. The benefits of these secondary information sources are clear. Contrary to SK predictions and thanks to these information sources, the BDF predictions exhibit relevant local patterns. The 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. 6b and f). This is even more obvious for week days since their means are generally higher than for weekend days (see Fig. 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



**Fig. 8.** Mean concentrations in 2007 (a and c) and total number of high exposure days in 2007 (daily mean higher than chosen threshold of  $80 \mu\text{g m}^{-3}$ ; 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.

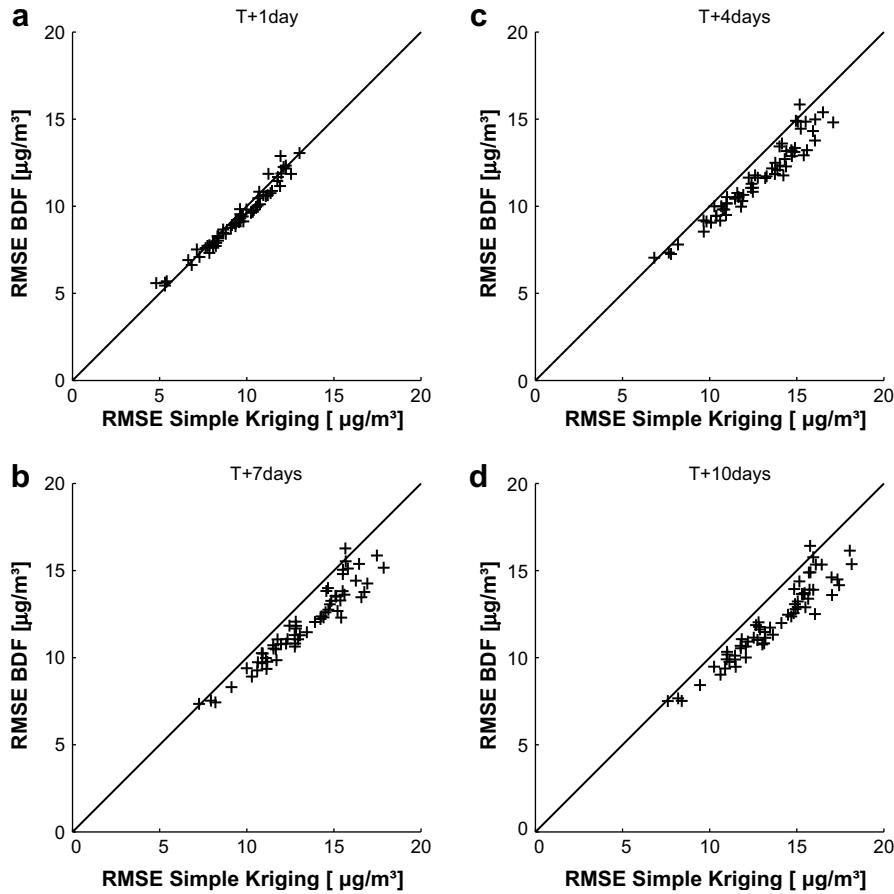


**Fig. 9.** Space-time variances of predictions of daily-mean NO<sub>2</sub> concentrations. On the left column are the Simple Kriging variances whereas Bayesian data fusion ones are on the right column. Illustrated days are the same as in Fig. 6.

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. 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  $\mu\text{g m}^{-3}$  (see Fig. 8b and d). This threshold has been chosen in order to specifically select situations characterized by a greater exposure to NO<sub>2</sub>. Indeed, SK overestimated that number of days at several locations, whereas BDF was much closer



**Fig. 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.

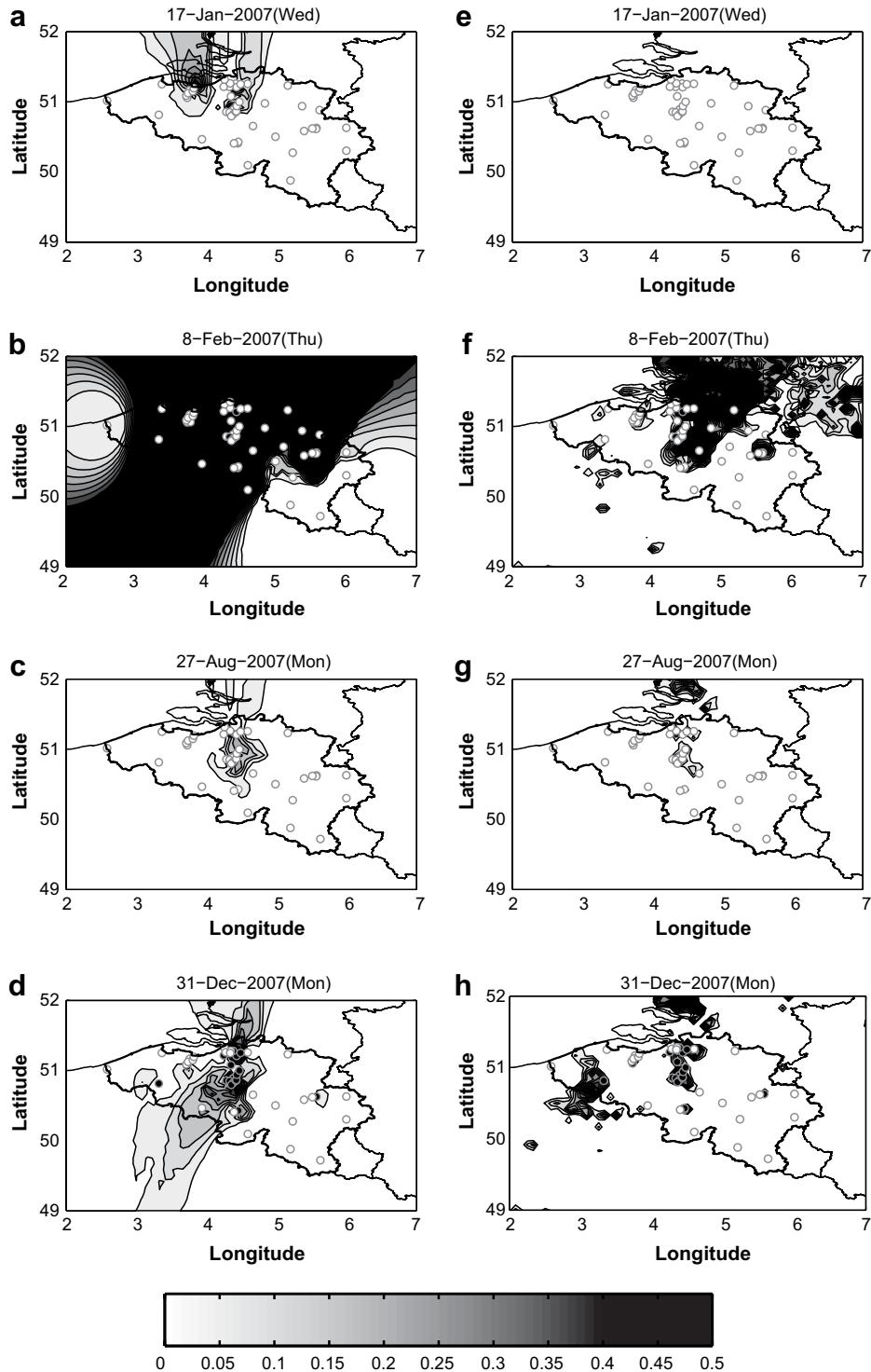
to the actual number of high concentration days computed from the raw data (again, see the filled circles on the maps).

Both the kriging and the BDF approaches provide the variance of the predicted conditional distribution (see Eq. (18) for the expression of the BDF conditional variance). Because nothing proves that these estimated variances are depicting the true conditional variances (i.e. covariance consistency; see e.g. Uhlmann, 2003), one cannot simply use the maps of variance as quality criteria. However, as BDF can be viewed as a Bayesian mixing between the kriging prediction and the secondary information, differences between conditional variances indicate the effect of the influence of these secondary information on the conditional distributions, and thus on the total knowledge about the process. Fig. 9 shows these conditional variances for both SK and BDF for the same dates as in Fig. 6. It is clear that BDF prediction yields smaller conditional variances than SK. It is worth noting here that for days with particularly windy meteorological conditions, there is a significant drop in the BDF variance compared to the SK one (see Fig. 9e and f). Further, it can be seen in Fig. 9g and h how the number of secondary information sources affects the variance for some locations (South-West part of the study area). Indeed, since this number is smaller at those locations, the variance slightly increases. Conditional variance thus shows clearly how sources of secondary information were able to increase the precision of the BDF predictions, thus increasing the knowledge on NO<sub>2</sub> 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. 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<sub>2</sub> concentrations for longer term predictions.

As the whole NO<sub>2</sub> 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<sub>2</sub> concentrations will exceed a given threshold. Fig. 11 shows the probability map that NO<sub>2</sub> concentrations will be higher than 50 µg m<sup>-3</sup> for the same dates as in Figs. 6 and 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. 11 leads to the conclusion that BDF outperformed SK. This is especially obvious on February 8, 2007 (Fig. 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,



**Fig. 11.** Space–time predictions for the probability of crossing the  $50 \mu\text{g m}^{-3}$  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.

leading thus to a drop in the  $\text{NO}_2$  concentrations except in the biggest cities.

## 5. Discussion and conclusion

In this paper, 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  $\text{NO}_2$  in Belgium. The proposed method showed its ability to account for several secondary information sources (in this paper, 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  $\text{NO}_2$  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  $80 \mu\text{g m}^{-3}$  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 ( $50 \mu\text{g m}^{-3}$  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 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. 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<sub>2</sub> 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°). For sake of simplicity in this paper, an intermediate resolution (0.25/3°) was chosen and information were downsampled (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.

Bayesian frameworks are known to be efficient in accounting for uncertainty such as parameter estimation. Depending on the relations involved with the distributions and on the influence of the parameters on these relations, the uncertainty of the parameters could be quite important with respect to the total uncertainty of the model. Belonging to the family of Bayesian approaches, BDF also theoretically enables to account for this kind of uncertainty. However, in order to focus the scope of the paper on the mixing properties of BDF, this total Bayesian framework was just mentioned in the text. It is clear that such considerations are worth considering but this was left for further researches on BDF.

As mentioned in this paper, 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 these 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.

## Acknowledgment

Pollution data presented in this paper were collected from the telemetric networks of the three Belgian regions. These networks are managed by the Flemish Environment Agency (VMM) in Flanders, the Brussels Institute for the Management of the Environment (Brussels Environment) in Brussels, and the Institute for Public Service (ISSEP) and the Ministry of the Environment (DGRNE) in Wallonia. The authors are grateful to these regional institutes and administrations. Also, the valuable comments of the anonymous reviewers noticeably contributed to the improvement of the manuscript.

## Appendix A. Properties of the fused variance

Assuming that  $f(z_i)$  (resp.  $f(z_i|y_{ij})$  for a given  $y_{ij}$ ) in Eq. (5) is a Gaussian distribution with mean equal to  $\mu$  (resp.  $\mu_y$ ) and variance  $\sigma^2$  (resp.  $\sigma_y^2$ ), then the ratio  $(f(z_i|y_{ij})) / f(z_i)$  is proportional to a third Gaussian distribution with mean equal to  $\mu_F$  and variance equal to  $\sigma_F^2$  where

$$\begin{cases} \sigma_F^2 = \frac{\sigma_y^2 \sigma^2}{\sigma^2 - \sigma_y^2} \\ \mu_F = \frac{\sigma^2 \mu_y - \sigma_y^2 \mu}{\sigma^2 - \sigma_y^2} \end{cases} \quad (19)$$

Now recoding  $\sigma_y^2 = a^2 \sigma^2 (a^2 < 1)$  in Eq. (19), it is clear that  $\mu_F = (\mu_y - a^2 \mu) / (1 - a^2)$  so that

$$\lim_{a \rightarrow 1^-} \mu_F = \begin{cases} +\infty & \text{if } \mu_y < \mu \\ \mu & \mu_y = \mu \\ -\infty & \text{if } \mu_y > \mu \end{cases}$$

As a consequence, the only admissible situation is  $\mu_y = \mu$ , otherwise the use of Eq. (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. (10)
2. use  $Y_{ij}$  observations to model  $\mu_{R_i|Y_{ij}}$  ( $= \mathbb{E}_{R_i}[R_i|Y_{ij}]$ )
3. use  $\mu_{R_i|Y_{ij}}$  values to model  $\sigma_{R_i|Y_{ij}}^2$  ( $= \text{Var}_{R_i}[R_i|Y_{ij}]$ ) such that
  - (a)  $\lim_{\mu_{R_i|Y_{ij}} \rightarrow 0} \sigma_{R_i|Y_{ij}}^2 = 1$
  - (b)  $\lim_{|\mu_{R_i|Y_{ij}}| \rightarrow +\infty} \sigma_{R_i|Y_{ij}}^2 = 0$
4. back-transform the modeled  $\mu_{R_i|Y_{ij}}$  and  $\sigma_{R_i|Y_{ij}}^2$  to the original variable  $Z_i$ , i.e.,
  - (a)  $\mu_{Z_i|Y_{ij}} = \sigma(\ell_i) \mu_{R_i|Y_{ij}} + \mu(\ell_i, t_i)$
  - (b)  $\sigma_{Z_i|Y_{ij}}^2 = \sigma_{R_i|Y_{ij}}^2 \sigma^2(\ell_i)$

Thanks to this, the mean  $\mu_F$  (resp. variance  $\sigma_F^2$ ) of the fused Gaussian distribution will tend to  $\mu$  (resp. to  $+\infty$ ), leading thus to a meaningful contribution for the ratio  $(f(z_i|y_{ij})) / f(z_i)$  in Eq. (5).

## References

- Bogaert, P., 1996. Geostatistics Applied to the Analysis of Space-time Data. Ph.D. thesis, Université catholique de Louvain.
- Bogaert, P., Fasbender, D., November 2007. Bayesian data fusion in a spatial prediction context: a general formulation. *Stoch. Env. Res. Risk A* 21 (6), 695–709.
- Brussels Environnement, 2008. Rapport sur les incidences environnementales du plan d'urgence en cas de pic de pollution. [http://documentation.bruxellesenvironnement.be/documents/RIE\\_pic\\_poll%20tution\\_20080528\\_FR.PDF](http://documentation.bruxellesenvironnement.be/documents/RIE_pic_poll%20tution_20080528_FR.PDF) (in French).
- Bruzzone, L., Cossu, R., Vernazza, G., December 2002. Combining parametric and non-parametric algorithms for a partially unsupervised classification of multi-temporal remote-sensing images. *Inf. Fusion* 3 (4), 289–297.
- Chilès, J.-P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. John Wiley and Sons, Inc., New York, NY.
- Christakos, G., 1990. A Bayesian/maximum-entropy view to the spatial estimation problem. *Math. Geol.* 22 (7), 763–777.
- Christakos, G., 1992. Random Field Models in Earth Sciences. Academic Press, San Diego, CA.
- 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., Kolovos, A., 1999. A study of the spatiotemporal health impacts of ozone exposure. *J. Expos. Anal. Environ. Epidemiol.* 9 (4), 322–335.
- Christakos, G., Serre, M.L., 2000. BME analysis of spatiotemporal particulate matter distributions in North Carolina. *Atmos. Environ.* 34 (20), 3393–3406.
- Christakos, G., Serre, M.L., 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., Vyas, V.M., 1998a. A composite space/time approach to studying ozone distribution over eastern United States. *Atmos. Environ.* 32 (16), 2845–2857.
- Christakos, G., Vyas, V.M., 1998b. A novel method for studying population health impacts of spatiotemporal ozone distribution. *Soc. Sci. Med.* 47 (8), 1051–1066.
- Chung, A.C.S., Shen, H.C., October 2000. Entropy-based Markov chains for multi-sensor 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.
- Cressie, N., 1991. Statistics for Spatial Data. In: Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., New York, NY.
- D'Or, D., Bogaert, P., Christakos, G., 2001. Application of the BME approach to soil texture mapping. *Stoch. Env. Res. Risk A* 15 (1), 87–100.
- Fasbender, D., Obsomer, V., Radoux, J., Bogaert, P., Defourny, P., 2007. Bayesian data fusion: spatial and temporal applications. In: Proceedings of the International Workshop on the Analysis of Multi-temporal Remote Sensing Images, 2007. MultiTemp 2007. Leuven, Belgium.
- Fasbender, D., Peeters, L., Bogaert, P., Dassargues, A., 2008a. Bayesian data fusion applied to water table spatial mapping. *Water Resour. Res.* 44.
- Fasbender, D., Radoux, J., Bogaert, P., June 2008b. Bayesian data fusion for adaptable image pansharpening. *IEEE Trans. Geosci. Remote Sens.* 46 (6), 1847–1857.
- Fasbender, D., Tuia, D., Bogaert, P., Kanevski, M., October 2008c. Support-based implementation of Bayesian data fusion for spatial enhancement: applications to ASTER thermal images. *IEEE Geosci. Remote Sens. Lett.* 6 (4), 598–602.
- FEBIAC, 2008. Datadigest 2008. <http://www.febiac.be/public/statistics.aspx?FID=23&lang=FR> (in French).
- Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, NY.
- Hoeksema, R., Clapp, R., Thomas, A., Hunley, A., Farrow, N., Dearstone, K., 1989. Cokriging model for estimation of water table elevation. *Water Resour. Res.* 25, 429–438.
- Janssen, S., Dumont, G., Fierens, F., Mensink, C., 2008. Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmos. Environ.* 42 (20), 4884–4903.
- Jacquet, O., 1989. Factorial kriging analysis applied to geological data from petroleum exploration. *Math. Geol.* 21 (7), 683–691.
- Jones, G.D., Allsop, R.E., Gilby, J.H., September 2003. Bayesian analysis for fusion of data from disparate imaging systems for surveillance. *Image Vis. Comput.* 21 (10), 843–849.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *Trans. ASME-J. Basic Eng.* 82 (Series D), 35–45.
- 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.
- Linde, N., Revil, A., Bolène, A., Dagès, C., Castermant, J., Suski, B., Voltz, M., 2007. Estimation of the water table throughout a catchment using self-potential and piezometric data in a Bayesian framework. *J. Hydrol.* 334, 88–98.
- Matlab, 2002. Mathworks Inc. Version 6.5.
- Melgani, F., 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.
- Moshiri, B., Asharif, M.R., Hosein Nezhad, R., March 2002. Pseudo information measure: a new concept for extension of Bayesian fusion in robotic map building. *Inf. Fusion* 3 (1), 51–68.
- Papoulis, A., 1991. Probability, Random Variables, and Stochastic Processes, third ed.. In: McGraw-Hill Series in Electrical Engineering, Communication and Signal Processing McGraw-Hill International Editions, Singapore.
- Pieczyński, W., Bouvrain, J., Michel, C., 1998. Unsupervised Bayesian fusion of correlated sensors. In: Proceedings of the First International Conference on Multisource-Multisensor Information Fusion (Fusion'98). Las Vegas, Nevada, pp. 794–801.
- Pinheiro, P., 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.
- Pradalier, C., Colas, F., 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).
- Rajan, D., Chaudhuri, S., March 2002. Data fusion techniques for super-resolution imaging. *Inf. Fusion* 3 (1), 25–38.
- Ren, W., McIennam, J., Leuangthong, O., Deutsch, C., 2006. Reservoir characterization of McMurray formation by 2D geostatistical modeling. *Nat. Resour. Res.* 15 (2), 111–117.
- Risken, H., September 1996. The Fokker-Planck Equation: Methods of Solutions and Applications. Springer Series in Synergetics. Springer. URL <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/354061530X>.
- Sahu, S., Gelfand, A., Holland, D., December 2007. High resolution space-time ozone modeling for assessing trends. *J. Am. Stat. Assoc.* 102 (480), 1221–1234.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.
- Simone, G., Farina, A., Morabito, F.C., Serpico, S.B., Bruzzone, L., 2002. Image fusion techniques for remote sensing applications. *Inf. Fusion* 3 (1), 3–15.
- Sveshnikov, A., 1979. Problems in Probability Theory, Mathematical Statistics and Theory of Random Functions. Dover Publications, Inc., New York.
- Uhlmann, J.K., 2003. Covariance consistency methods for fault-tolerant distributed data fusion. *Inf. Fusion* 4 (3), 201–215.
- Wikle, C.K., Milliff, R.F., Nychka, D., Berliner, L.M., June 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 Goiás, Brazil. *Nat. Resour. Res.* 8 (2), 153–163.
- Zhang, Z., Blum, R.S., 2001. A hybrid image registration technique for a digital camera image fusion application. *Inf. Fusion* 2, 135–149.