Model-Driven Development of Covariances for Spatiotemporal Environmental Health Assessment

Alexander Kolovos[1,*], José Miguel Angulo[2], Konstantinos Modis[3], George Papantonopoulos,[3] Jin-Feng Wang[4], and George Christakos.[5]

[1] *SpaceTimeWorks, LLC, 255 G St #105, San Diego, CA 92101, USA.*

[2] *Departamento de Estadistica e Investigación Operativa, Universidad de Granada, Granada, Spain.*

[3] *School of Mining and Metallurgical Engineering, National Technical University of Athens, Greece.*

[4] *LREIS, Institute of Geographical Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.*

[5] *Department of Geography, San Diego State University, San Diego, CA 92182-4493, USA.*

* Corresponding Author: Email: alexander.kolovos@spacetimeworks.com. Ph: +1-619-559-2497, Fax: +1-619-342-7164

Abstract

Known conceptual and technical limitations of mainstream environmental health data analysis have directed research to new avenues. The goal is to deal more efficiently with the inherent uncertainty and composite space-time heterogeneity of key attributes, account for multi-sourced knowledge bases (health models, survey data, empirical relationships etc.), and generate more accurate predictions across space-time. Based on a versatile, knowledge synthesis methodological framework, we introduce new space-time covariance functions built by integrating epidemic propagation models and we apply them in the analysis of existing flu datasets. Within the knowledge synthesis framework, the Bayesian maximum entropy (BME) theory is our method of choice for the spatiotemporal prediction of the ratio of new infectives (RNI) for a case study of flu in France. The space-time analysis is based on observations during a period of 15 weeks in 1998-1999. We present general features of the proposed covariance functions, and use these functions to explore the composite space-time RNI dependency. We then implement the findings to generate sufficiently detailed and informative maps of the RNI patterns across space and time. The predicted distributions of RNI suggest substantive relationships in accordance with the typical physiographic and climatologic features of the country.

Keywords

Spatiotemporal; environmental assessment; prediction; covariance models; BME.

1. Introduction

Stochastic spatiotemporal analysis is widely used in the study of environmental attributes, including the field of environmental health assessment (e.g., Christakos and Kolovos 1999; Choi et al. 2007). In point-referenced data analysis the goal is to predict unknown values of attributes of interest at unsampled spatial locations and temporal instances (e.g., Banerjee et al. 2004; Bossak and Welford 2009; Kuo et al. 2009). Information from spatiotemporal inference is often represented and visualized in the form of maps; hence, this type of analysis can be particularly important for prediction and decision-making in environmental health and risk studies, management and planning, etc.

Due to the inherent uncertainty in most natural attributes, stochastic methodologies have been introduced that can address the uncertainty issues in a scientific and meaningful manner at the stages of both data acquisition and subsequent processing (Christakos and Hristopulos 1998, Christakos et al. 2002; Escudero et al. 2009). The need to address the concerns of stochastic analysis led to the development of software and computational tools to address associated numerical tasks or provide related functionality. A common theme shared by many of these tools is the use of mainstream spatial statistics and geostatistics methods (Christakos 1992; Deutsch and Journel 1992; Heuvelink 1998; Bivand et al. 2008; Wong and Lee 2005) that were first introduced a few decades ago, and have since developed to the extent of their capabilities and weathered successfully the test of time. However, these techniques traditionally serve spatial-only analysis purposes (e.g., Banerjee et al. 2004; Arrow et al. 2003; Moore and

Carpenter 1999; Schabenberger and Gotway 2005; Wu et al. 2004). In addition, they are known to have a variety of inherent limitations, e.g. linear models are used to represent nonlinear phenomena, multi-sourced observation uncertainty is not handled in a conceptually rigorous and unified manner, restrictive assumptions are typically made regarding the probability distribution and geographical heterogeneity of the data (Bayraktar and Turalioglu 2005; Chen et al. 2004; Liao et al. 2006). In order to address these limitations, mathematical concepts and techniques of composite space-time analysis under conditions of uncertainty and spatiotemporal heterogeneity were introduced the last two decades that led to the development of modern spatiotemporal statistics and geostatistics (e.g., Choi et al. 2003; Christakos 1991a, 1991b, 2000; Douaik et al. 2004; Law et al. 2006; Serre et al. 2003).

The shift in analysis introduced by the latter developments was based on the cognitive framework of knowledge synthesis (KS) in order to integrate all relevant knowledge bases (core and site-specific) about the attribute under study, and generate inference in a mathematically rigorous and epistemically sound manner. KS allows for a variety of methods to be applied in the spatiotemporal inference problem and tackles successfully the aforementioned classical geostatistics concerns. Its appealing characteristics and usefulness have been already demonstrated, e.g., by means of introducing advanced functions for spatiotemporal analysis (Christakos et al. 2002; Christakos 1998; Lee et al. 2008). These functions are implemented in terms of KS methods such as Bayesian maximum entropy (BME) and non-Bayesian stochastic logic (NBSL) methods (Christakos et al. 2002; Christakos 1991b, 2000, 2010).

Exploring physical dependencies and structural correlations in the observed data plays a predominant role in spatiotemporal analysis. Most typically, these correlations are modeled in terms of permissible covariance functions that express mathematically the underlying structure in the stochastic field that represents the natural phenomenon, and provide the foundation to make predictions at unsampled locations. Permissible functional forms such as the exponential, the Gaussian, and the spherical are broadly used to fit empirical descriptions of the correlation in the space-time composite domain. Beyond those generic forms, additional permissible covariance models have been suggested for this purpose; see, e.g., Gneiting (2002), Ma (2008), Porcu et al. (2008). Among the variety of functions that can serve as suitable covariance models, Kolovos et al. (2004) proposed a series of elaborate, nonseparable spatiotemporal models that are based on physical laws and principles. In general, it is desirable to derive spatiotemporal correlations founded on such principles, because physical laws enable a science-based justification of these correlations rather than a fit of generic technical functions.

The present work builds further on this premise, and we present an example where new, model-based spatiotemporal covariances are used for prediction. Specifically, we derive covariances from well-known epidemic models, which we extend into the space-time domain so that they adequately describe the dynamic propagation of the rate of new infectives in an environmental health context. As a case study, we present an example of flu spread across France. In terms of the KS framework, the proposed covariances are part of the core knowledge bases that are used by the Bayesian maximum entropy (BME)

methodology for spatiotemporal inference and mapping. In particular, we use the BME approach to study the space-time flu distribution in France based on a specific dataset and the relevant epidemic law that is used to derive the covariances.

The BME approach may be viewed as a substantive extension of the classical geostatistics commonly used in environmental health research (Atkinson 1993, Saito et al. 2005; Stein et al. 1999). According to BME, due to the considerable multi-sourced uncertainty (caused by physiographic features, disease variability, meteorological conditions, etc.) the space-time distribution of the fraction of the population that is infected is represented as a spatiotemporal random field $X_p = X_{s,t}$ (S/TRF; Christakos and Hristopulos 1998) at each space-time point $p = (s,t)$ in the natural continuum with spatial coordinates $s = (s_1, s_2)$ and time t. Specification of the X_p values at all points in a space-time continuum determines a realization of the S/TRF. Randomness manifests itself as an ensemble of possible realizations regarding the X_p distribution. Similarly, let $Y_p = Y_{s,t}$ and $Z_p = Z_{s,t}$ be the S/TRFs that represent the distributions, respectively, of the fraction of the population that is susceptible of becoming infected and the fraction that has recovered and is immune.

We subsequently take advantage of software tools specifically created to implement the unique BME features and also integrate mapping capabilities. Tools such as the Spatiotemporal Epistematics Knowledge Synthesis-Graphical User Interface (SEKS-GUI; Kolovos et al. 2006; for more details, readers are encouraged to visit http://www.spacetimeworks.com) are commonly used in real-world studies, as they have

been crafted to perform all different stages of space-time analyses (exploratory, correlation analysis, prediction, mapping) and take advantage of the particular BME predictive features. In the present study, we use some of the SEKS-GUI components for the purpose of trend estimation and prediction, whereas custom software modules are also introduced to carry out the correlation analysis of flu propagation.

2. Methods

2.1. The Knowledge Bases

In principle, BME analysis distinguishes between the core (or general) knowledge bases (KB) *G*-KB, and the site-specific (or specificatory) *S*-KB. In this work, in particular, the following KB are considered.

2.1.1. The G-KB

One can find in the literature a considerable number of deterministic and stochastic epidemic models, most of which involve ordinary difference or differential equations that focus on the temporal variation of a disease (Hethcote 1994; Allen and Burgin 2000; Angulo *et al*. 2012). The following space-time stochastic flu model has been suggested by Angulo and Christakos (2010) to describe disease spread,

$$
\left.\begin{array}{l}
X_{s,t+1} = q(1-a+bf_{s,t})X_{s,t} + \int_{R^2} du\, \kappa_{s-u}(1-a+bf_{u,t})X_{u,t} \\
Y_{s,t+1} = q(Y_{s,t} - bf_{s,t}X_{s,t}) + \int_{R^2} du\, \kappa_{s-u}(Y_{u,t} - bf_{u,t}X_{u,t}) \\
Z_{s,t+1} = q(Z_{s,t} + aX_{s,t}) + \int_{R^2} du\, \kappa_{s-u}(Z_{u,t} + aX_{u,t})
\end{array}\right\} \tag{1}
$$

The set of Eqs. (1) is considered in the present work to yield a representation of the space-time flu variation. In Eqs. (1), q is the fraction of the population that stays at $p = (s,t)$, and 1-q is the fraction that moves during the time period $t \to t+1$; $a \in (0,1]$ denotes the probability that an infected individual recovers and becomes immune, $b \in (0,1]$ is the probability of transmission of infection in an encounter involving one infected and one susceptible individual (one may chose to assume that the probability of transmission varies across space-time in which case the b in Eqs. (1) should be replaced by $b_{s,t}$); $f_{s,t}$ is a monotonically decreasing function of time t with sufficient flexibility to represent what one would expect to be the behavior of the fraction of the population who is susceptible to infection; and κ_s is a time-independent kernel controlling population movement across space and such that $\int_{R^2} ds\, \kappa_s = 1-q$ (e.g., may be a Gaussian kernel with finite variance). Clearly, the following initial conditions (IC) are valid:

$$
\left.\begin{array}{l}
X_p + Y_p + Z_p = 1 \\
Y_{s,0} = 1 - X_{s,0} \\
Z_{s,0} = 0
\end{array}\right\} \tag{2}
$$

The space-time disease model (1)-(2) carries important information in the computation of explicit forms via recursion for the variables X_p, Y_p, Z_p and their (cross) covariances,

since these expressions involve iterated integration accumulated over time. Hence, randomness in Eqs. (1) can be introduced in terms solely of $X_{s,0}$, which can be assumed to be a spatially homogeneous random field (corresponding to the relevant assumptions about the kernel and population movement).

To gain valuable insight, in the following we consider a special case of the stochastic flu model (1)-(2) in which it is assumed that, (a) $f_{s,t} = f_t$, i.e. spatially independent, (b) $q = 1$ (i.e., the population does not move spatially during the duration of the disease), in which case the κ_s (time-independent kernel that controls population movement across space) has no significance (the integral terms in Eqs. (1) can be neglected), and (c) the random IC $X_{s,0}$ is spatially homogeneous (or isotropic). Under the assumptions a-c, Eqs. (1) reduce to

$$
\left.
\begin{aligned}
X_{s,t+1} &= A_t X_{s,0} \\
Y_{s,t+1} &= 1 - (1 + B_t) X_{s,0} \\
Z_{s,t+1} &= C_t X_{s,0}
\end{aligned}
\right\}
\tag{3}
$$

with IC as in Eq. (2), and the factors A_t, B_t, and C_t are defined as follows:

$$
A_t = \prod_{i=0}^{t}(1 - a + b f_i)
\tag{4}
$$

$$
B_t = \begin{cases} b f_0, & t = 0 \\ \sum_{j=0}^{t} b f_j \prod_{i=0}^{j-1}(1 - a + b f_i), & t \geq 1 \end{cases}
\tag{5}
$$

$$
C_t = \begin{cases} a, & t = 0 \\ a \sum_{j=0}^{t} \prod_{i=0}^{j-1}(1 - a + b f_i), & t \geq 1 \end{cases}
\tag{6}
$$

The corresponding mean, covariance, and cross-covariance functions are shown in Table 1 with IC, $c_{X;s-s',0}$ (covariance of $X_{s,0}$ and $X_{s',0}$). The expressions in Table 1 indicate that the proposed covariances are non-stationary in time. For example, the cross-covariance of $X_{s,t+1}$ and $Y_{s',t'+1}$ is $c_{X,Y;s,t+1;s',t'+1} = -A_t(1+B_{t'})c_{X;s-s',0}$, which depends explicitly on the specific temporal instances t and t'.

As noted earlier, f_t is a monotonically decreasing function of t. It may include at least two parameters that allow variation at $t=0$ and for different rates of decay as t increases, and possibly a parameter so that at $t=0$, $f_0 \in (0,1)$, and at $t \to \infty$, $f_\infty \geq 0$. In the present study, we use the following exponential form

$$f_t = \exp(-\gamma t^\varepsilon)$$ (7)

where the parameters $\gamma, \varepsilon > 0$. Other f_t forms could be considered, depending on the known features of the epidemic (Angulo *et al.* 2012).

2.1.2. The S-KB

The ratio of new infectives (RNI) $\rho_{s,t}$ is an attribute defined as follows: The RNI $\rho_{s,t}$ in a geographical region s and at the time period t is the ratio of the number of new infected individuals over the population of the region. Note that $\rho_{s,t}$ is a quantity similar to the incidence crude rate used in epidemiological studies; however, $\rho_{s,t}$ is dimensionless, whereas the crude rate is typically measured in person-weeks units. In our work, we use the RNI to study the number of new flu infectives per week in each one of the 21 regions in the French mainland (Fig. 1). The collected information consists of

aggregated observations recorded by general physicians through the Sentinel project in France over a time period of 53 weeks between August 1998 and August 1999.

In the RNI dataset, the $\rho_{s,t}$ values vary between 0 and $4.8 \cdot 10^{-3}$, where the high values indicate a fairly contagious flu distribution. The actual survey contains a significant number of zero new infectives at the finer district scale. It is known that some of these zeros in our data may be noise, e.g., an occasional lack of a physician at a particular location may have caused a potentially nonzero record to be recorded as zero. For this reason, we choose to present this flu study at the aggregated level of larger regions; this choice has a smoothing effect on the finer scale inconsistencies. The RNI observations are assigned the coordinates of the centroids of the corresponding French regions, and they form a time-series of 53 weekly values at each region. See Fig. 2 for a plot of the study data time series for five selected regions. Fig. 3 shows an example of the spatial distribution of RNI observations for the second of the weeks included in our analysis, as defined later in this section.

The study of aggregated information justifies the choice of RNI over the actual number of new infectives as follows: The original records of aggregate numbers of new infectives for each region constitute areal data. However, the population within a region is more probably spread across its area, rather than being concentrated at a single location (a selection of population models can be found in the literature; see, e.g., Renshaw 1993). In that sense, the newly infected people could be possibly grouped at many different locations across the region, such as towns and cities. Consequently, in the context of

areal data, single observations of the number of new infectives at the region centroid are most likely unrepresentative of the spatial distribution of the new infectives. Instead, the ratio of new infectives (RNI) is an attribute that can be sensibly considered at any arbitrary location as a point datum. In this perspective, an RNI value represents a quantitative characteristic that is valid for any single location across a region. Accordingly, we can use geostatistics to analyze the spatiotemporal structure of this characteristic and provide localized values on a space-time grid. In the following, we investigate the spatiotemporal structure directly on the basis of the RNI observations available at the regional scale. It is possible to further refine the structure analysis by means of a downscaling approach; for an example, see the work of Choi et al. (2003).

In effect, using the RNI attribute enables approximation of the areal data coming from regional aggregation with point-referenced data at the region centroids. Our work adopts this simplified approach in the interest of weighting on the proposed spatiotemporal covariances development. Often, approximating areal data with observations at centroids is a necessary mode of action to protect sensitive data at finer local scales. Nonetheless, as commonly as this approximation might be used, it also entails some concerns. These relate to the potential impact of the geographical units characteristics that are used in the analysis, as pointed out by Goovaerts (2008); the potentially dubious assumption of using similar spatial resolutions at the areal scale of the observations and the prediction grid scale; and the issue of ignoring variability within each of the studied areal units. Yu et al. (2011) discusses these concerns and recommends further study of their potential biasing impacts on prediction.

We study the RNI attribute in a rectangular domain that surrounds France. In geographic coordinates, the rectangle spans from 5°W to 8°30'E longitude, and from 42°N to 51°30'N latitude. An area this large can distort the calculations of eucleidian distances due to the Earth curvature. To account for this distortion, our study uses the Lambert Conformal Conic (LCC) projection for France, where distances are measured in meters. We consider selected locations within this domain to predict unsampled ratios ρ_p across space-time.

In the temporal scale, we focus on the 15 weeks that span from the 19th week of flu data starting on December 21, 1998 (that we will refer to as $t = 1$) until the end of the 33rd week of data that starts on March 29, 1999 (referred to as $t = 15$). Overall, we have a total of $21 \times 15 = 315$ observations over the 15-week period. We selected the specific time frame because it coincides with the most active period (rise, peaking and waning of the ratios of new infectives) of the flu season, according to the data time series. Our goal is to predict the spatiotemporal pattern of $\rho_{s,t}$ across France throughout the specified 15-week time interval.

The parameters a, b, γ and ε in our flu model Eqs. (3-7) regulate characteristics of the flu propagation, and their values are determined by case-specific information. To preserve the stochastic nature of Eqs. (3-7), we consider these parameters as random variables. In particular, the mean values for the b, γ and ε parameters are derived by fitting them to the RNI $\rho_{s,t}$ observations by using Eqs. (3-4) and (7) across all time

instances; the fitted values are the corresponding means shown in Table 2. For the probability a that an infected individual recovers and becomes immune, assuming that the cold lasts for an average of 7 days, one can deduce the expected value $E[a] = 1/7 \cong 0.143$; we assume this to be the mean of a. This rationale implies that, on average, a newly infected person gets well within one week; hence, in the context of our study the S/TRF $X_{s,t}$ of the fraction of population that is infected during week t coincides with the RNI $\rho_{s,t}$ of newly infected people at the same week t. For this reason, we represent $\rho_{s,t}$ through the S/TRF $X_{s,t}$. Since we have data only for RNI, this study performs inference analysis for the S/TRF $X_{s,t}$. If there are observations available for the fraction of the population that is susceptible of becoming infected and the fraction that has recovered and is immune, it is straightforward to extend the analysis in terms of $Y_{s,t}$ and $Z_{s,t}$ S/TRF, respectively.

By assuming that a and b are random variables, we also satisfy the realistic requirement of the flu study that a and b vary across space-time as follows: We consider a and b to follow beta distributions according to Table 2. Then, we manifest variation of a and b in space-time by selecting random values from their distributions to compute the covariance between each pair of space-time locations. Observe that no systematic space-time dependence is assumed for the values of each parameter, hence the parameters appear in the flu model equations without s or t subscripts. For comparison purposes, we perform two experiments in our study: (A) We consider a standard deviation of 10% in the values of each one of a and b, and (B) We consider a standard deviation of 2% in the values of each one of a and b.

For practical purposes, the *S*-KB is described in terms of the point vector $\boldsymbol{p}_{\text{data}}$ that

consists, respectively, of the hard data point vector $\boldsymbol{p}_{\text{hard}}$ (exact measurements), and the

soft data vector $\boldsymbol{p}_{\text{soft}}$ (where incomplete yet valuable $\rho_{s,t}$ information is available). In

the absence of information about uncertainty in the aggregation process, we consider that

the whole dataset consists of $\rho_{s,t}$ observations with exact values (hard data); hence, we

consider only the hard data vector $\boldsymbol{p}_{\text{hard}}$ of new infectives ratios to be the $\boldsymbol{p}_{\text{data}}$ vector.

For inference purposes, we need to identify and remove mean (surface) trends in the $\rho_{s,t}$

data so that we can explore the dependence among the data residuals in space and time.

2.2. BME Methodology

As noted earlier, we implement BME analysis to study the flu RNI distributions in

France. A solid presentation of the BME theory can be found in the literature (e.g.,

Christakos 2000, 2008). In a brief and concise summary, the fundamental BME

equations of spatiotemporal dependence analysis and mapping are as follows (Christakos

2008, 2010)

$$
\left.
\begin{aligned}
\int d\boldsymbol{\chi}\,(\boldsymbol{g} - \bar{\boldsymbol{g}})e^{\boldsymbol{\mu}^T \boldsymbol{g}} &= 0 \\
\int d\boldsymbol{\chi}\,\boldsymbol{\xi}_S\,e^{\boldsymbol{\mu}^T \boldsymbol{g}} - A f_K(\boldsymbol{p}) &= 0
\end{aligned}
\right\}
\tag{8}
$$

where the bar denotes stochastic expectation, $\boldsymbol{g}$ is a vector of g_i-functions ($i =1, 2, ...$)

that represents stochastically the available general knowledge bases G-KB described

above, $\boldsymbol{\mu}$ is a vector of μ_i-coefficients that depends on the space-time coordinates and is

associated with $\boldsymbol{g}$ (i.e., the μ_i express the relative significance of each g_i-function in the

composite solution sought), $\boldsymbol{\chi}$ stands for the vector of actual values of the attribute, the

$\boldsymbol{\xi}_S$ represents the available site-specific knowledge bases S-KB also described earlier, A

is a normalization parameter, and f_K is the attribute probability density function (PDF) at

each spatiotemporal point (the subscript K means that f_K is based on the blending of the

core and site-specific KB). The $\boldsymbol{g}$ and $\boldsymbol{\xi}_S$ are the inputs in Eqs. (8), whereas the

unknown are the $\boldsymbol{\mu}$ and f_K across space and time. Eqs. (8) establish a spatiotemporal

analysis context within which the G-KB refers to the entire $\boldsymbol{p}$-domain of interest; the

domain consists of the space-time point vector $\boldsymbol{p}_k$ where $\rho_{s,t}$ predictions are sought, and

the point vector $\boldsymbol{p}_{\text{data}}$ where S-KB is available. The model f_K describes the probability

distribution of the $\rho_{s,t}$ values at each point $\boldsymbol{p}_k$ in view of the total knowledge, K-KB.

Given the f_K at $\boldsymbol{p}_k$, different $\rho_{s,t}$ predictions can be derived at each spatiotemporal node

of the mapping grid (most probable, error minimizing etc.), depending on the objectives

of the study.

The BME framework is based on the above foundations to assimilate in a rigorous

manner uncertain but valuable data sources that were previously unaccounted for. It does

not make any of the common yet often questionable assumptions of classical statistics,

since BME can consider non-Gaussian probability laws, non-linear predictors and

spatiotemporally dependent attribute values with heterogeneous features. Notably, in the presence of only means and covariances *G*-KB and hard data *S*-KB, the BME predictor that is derived from Eqs. (8) reduces to the ordinary kriging predictor, and this renders BME as a superset of classical geostatistics methods (Christakos 2000). In terms of the analysis output, the complete prediction PDF is generated at each space-time point rather than a single estimate. Thus, by design, the BME mapping accuracy is superior to that of other techniques such as geostatistical Kriging and statistical regression; see, e.g., Christakos et al. (2004) and Serre and Christakos (1999).

3. Results

3.1. Preliminary Analysis and Spatiotemporal Dependence Structure of Flu

A first look at the spatiotemporally averaged histogram of the $\rho_{s,t}$ data reveals that the ratios are highly skewed (see Fig. 4). Indeed, only a small percentage of the population get the flu every week; most of the time no people get sick, which frequently leads to a near-zero ratio. The histogram is also in agreement with the fact that population risk ratios such as the RNI typically follow the Poisson law (Clayton and Hills 1993). The 15-week sample average is $\bar{x} \approx 5.2 \cdot 10^{-4}$, and the figure histogram resembles closely the Poisson distribution $Poisson(\bar{x})$.

We perform spatiotemporal analysis at the nodes of a regular grid surrounding the French mainland with spatial node distance equal to 20 Km and temporal distance 1 week between consecutive instances. The dynamic analysis is performed across space and between the instances $t=1$ and $t=15$, as defined earlier. The composite spatiotemporal grid has 53×52 spatial and 15 temporal nodes resulting in a total of 41340 nodes.

A spatiotemporal trend analysis is performed first to remove surface trends from the data. We estimate spatiotemporal trends by applying a spherical smoothing kernel with a spatial range of 400 Km and a temporal range of 4 weeks. These trends are temporarily removed to prevent larger-scale spatiotemporal variations from interfering with analysis at the study scale. After the prediction stage, the trends are reinstated to produce maps of the predicted RNI. The plots in Fig. 5 display examples of the estimated trends at the instances of weeks 2, 6, 10 and 14.

The residual RNI are then subjected to spatiotemporal dependence analysis to discover how the observed ratios $\rho_{s,t}$ are correlated in space-time, and then use this information for inference. The present study introduces the covariance functions in Table 1 based on the space-time flu model in Eqs. (3). These covariances are non-stationary in time, therefore computations must be made explicitly for every pair of temporal instances. For the S/TRF $X_{s,t}$ we calculate the covariances for every pair of spatiotemporal locations $(s,t+1)$ and $(s',t'+1)$ by computing the factors A_t and $A_{t'}$ based on Eqs. (4) and (7) starting at $t=0$. For the spatial covariance $c_{X;s-s',0}$ covariance factor

at $t = 0$, we fit a spherical model to the RNI observations at $t = 0$ that has a sill of $3.67 \cdot 10^{-8}$ (in RNI variance units) and a range of 120 Km (Fig. 6a).

In mainstream covariance studies, the correlation between any two space-time locations is calculated with the use of permissible covariance functions, such as the spherical form used earlier for the spatial covariance $c_{X; s-s', 0}$. For case-specific sill and range parameters, the correlation measure is then expressed in terms of distances from a spatial and a temporal origin. Hence, one can plot the covariance function for a series of spatial and temporal distances to visualize how correlation evolves as these distances increase; this is the space-time extension of the steps we took to create the spatial covariance $c_{X; s-s', 0}$ in Fig. 6a. Due to non-stationarity in time, covariance functionals in Table 1 depend on the selection of $t = 0$, and mainstream covariance modeling and plotting in terms of spatial and temporal distances is inapplicable in this case.

However, one can plot the factors A_t, B_t, and C_t in Eqs. (4-6) as functions of time to gain some insight into the behavior of each one of the factors; see plots in Fig. 6b-d. The factors are plotted for a long sequence of instances up to $t = 100$. The plots indicate that for the flu model mean values of coefficients a, b, γ and ε and the selection of f_t in Eq. (7), these factors exhibit an asymptotic behavior. The spatial covariance factor $c_{X; s-s', 0}$ governs dampening of the products between pairs of the A_t, B_t, and C_t at instances t and t', as shown in Table 1. In addition, the plots in Fig. 6b-d illustrate vividly the effect of considering random a, b coefficients in the flu model of Eqs. (4)-(6). As is intuitively expected, larger levels of variance in the coefficients cause the

19

factors to fluctuate more intensely in time (solid and dot-dashed plots in Fig. 6b-d) compared to the behavior of constant a and b (dashed line plots in Fig. 6b-d). The temporal behavior of each individual factor A_t, B_t, and C_t illustrated in Figure 6 could be possibly linked by disease specialists to specific flu spread physical characteristics. For example, Fig. 6b indicates peaking and attenuation in the temporal evolution of A_t in a single cycle that lasts for about 50 weeks. This cycle is governed by the parameters a, b, γ and ε and the selection of f_t in Eqs. (4) and (7), and can be classified as manifestation of specific epidemic behavior. Then, in a different epidemiological event governed by a possibly different combination of parameters a, b, γ, ε and function f_t, a similar temporal evolution of A_t can indicate similarity of the two events on the basis of the A_t characteristics. That is, the model-based covariances we introduce on the foundation of Eqs. (1) could be potentially play an important role as classification tools for disease spread events.

3.2. Spatiotemporal Mapping of Flu

The predictions of the flu RNI $\rho_{s,t}$ values are obtained by means of BME analysis, using the available $\boldsymbol{p}_{\text{data}}$ information, the $\rho_{s,t}$ mean trends, and covariances of $X_{s,t+1}$ and $X_{s',t'+1}$ in Table 1. At every space-time grid node, BME generates the prediction PDF, f_K, of $\rho_{s,t}$ at that location; thus, BME provides a full stochastic description of the ratios $\rho_{s,t}$ throughout the specified spatiotemporal grid. The series of prediction PDFs f_K across space-time contain a wealth of information about the spatiotemporal distribution

20

of $\rho_{s,t}$ because various RNI values can be derived from f_K. A typical $\rho_{s,t}$ prediction value is obtained in terms of the mean of f_K (*BMEmean*).

Fig. 7 displays the predicted RNI $\rho_{s,t}$ *BMEmean* for the experiment A (random coefficients a and b with 10% standard deviation) across space at selected weeks throughout the time-period of interest. Each map portrays predicted RNI values at the output grid locations based on each location's spatial and temporal data neighbors, in accordance to the specified spatiotemporal correlation model in Table 1. The maps in Fig. 7 are selected snapshots of the weekly prediction sequence at intervals of one month, and they show clearly how the RNI rates increase and subside during the high-activity segment of the flu season. All RNI prediction maps are identically color-scaled to facilitate comparisons between different temporal instances. Predicted values are overall low and range between 0 and about 0.0015 in the southern regions during the flu season peak.

Fig. 7 also shows that throughout the study duration the model predicts noticeable spikes of $\rho_{s,t}$ values at individual locations among much lower RNI values. These spikes appear as darker points on the maps because of the dense gradient between them and the values at the surrounding grid nodes. The actual values of these high RNI predictions might exceed the scale used in the maps; for example, at $t = 2$ in Fig. 7a, 5 of the 7 darker points have values that exceed the color scale of the plot, and the maximum of these predicted $\rho_{s,t}$ values is 0.0193. We attribute these isolated high values to a combination of two reasons: First, this is a natural result of having strong variability in

the observations at the corresponding and neighboring time instances; one may observe, for example, the high spatial variability in the RNI values at $t = 2$ (Fig. 3). The covariance models we use manifest this variability by naturally causing some local RNI predictions to exhibit higher values. Second, in some cases the magnitude of these few high RNI predictions exceeds even the highest observed RNI values; for example, at $t = 2$ (Fig. 3), the RNI data values remain below 0.0005. This behavior suggests that a few unreasonably high RNI values might be isolated numerical artifacts of the analysis. In the light of the previous two remarks, the predicted spikes in the RNI $\rho_{s,t}$ distribution should be rather seen as indications of the possibility to have comparatively higher RNI values in the corresponding spatial neighborhoods than be accounted as actual predictions of acute flu RNI.

The prediction results of experiment B (random coefficients a and b with 2% standard deviation) were found to be very similar to the results of experiment A, both in the actual predicted values and the spatiotemporal patterns produced by each different approach. Fig. 8 illustrates the *BMEmean* of the flu RNI $\rho_{s,t}$ prediction from the experiment B at the same four weekly instances as Fig. 7. A direct comparison of predicted RNI $\rho_{s,t}$ values at corresponding spatial locations is unfeasible due to the randomness infused in the prediction at each space-time location by the coefficients a and b; see also the discussion in the following paragraphs. However, the RNI prediction maps in experiment B are color-scaled using the same scale as those of the experiment A, and a visual comparison reveals the similarity of important prediction characteristics of the two experiments; namely, predictions in both cases appear to have similar levels of

spatial variation, patterns of higher or lower RNI predictions are similar, and these patterns are located at comparable spatial locations in both cases.

As a measure of prediction error we compute the standard deviation of the prediction PDF f_K at each spatiotemporal location (see an example in Fig. 9). Error values were distributed relatively evenly in space at every temporal instance for each one of the two experiments A and B. Errors were also found to be expectedly smaller close to data locations; however, this effect was less pronounced in experiment A than in experiment B, which is attributed to the higher standard deviations of the random coefficients a, b in experiment A. In addition, the error values exhibited increasing values in both experiments towards the final weeks in the study, and this is related to the progressively increasing variability in the values of the factors A_t, B_t, and C_t in Eqs. (4-6), as was illustrated in Fig. 6b-d. These findings reflect that BME incorporates efficiently in the output the uncertainty of the stochastic components in the G-KB, in addition to the uncertainty attributed to the prediction process itself.

The previous experiments A and B suggest that different levels of variation in the random coefficients a and b of the covariance expressions affect only mildly the predictive behavior of the BME model, despite the notable effect observed in the individual covariance factors in Fig. 6b-d and the effects on the prediction errors. The fact that we assume realistic changes in the random coefficients a and b for inference at different spatial and temporal locations leads to another interesting observation. At any given spatiotemporal node the relevant computations produce different results from any

other repetition of computations at the same node, and this is due to the constantly varying values of a and b. In light of this remark, it is possible to consider the space-time flu predictions as possible realizations of a simulation process that is controlled by the natural randomness in a and b. For this investigation, we performed the following experiment: We picked the time instance of week $t = 7$ to perform 100 iterations of BME flu predictions for each one of the settings in experiments A and B. Then, for each one of these settings we averaged the RNI *BMEmean* values at each node and produced a map of the averaged RNI *BMEmean*. These two maps are shown in Fig. 10. For comparison purposes, Fig. 10c shows the *BMEmean* map for experiment A at $t = 7$. Fig. 10d depicts a map of the absolute differences $\overline{BMEmean}_A - \overline{BMEmean}_B$ of the averaged *BMEmean* values for the A and B experiment settings.

Note that the absolute differences in Fig. 10d range in their vast majority between near-zero and around -0.005. The near-zero values are explained by the fact that in both experiments *BMEmean* tends to be very close to zero at the same locations throughout all iterations; hence, the averaging effect produces near-zero values for both experiment A and B settings, and their absolute difference is also a near-zero value at corresponding locations. The uniformity in the absolute differences across the rest of the spatial prediction grid has a similar explanation rooted in the averaging effect. Most interestingly, the absolute differences plot in Fig. 10d indicates that the averaged *BMEmean* values based on experiment B consistently overshoot their experiment A counterparts. This behavior suggests that higher standard deviations in the random distributions of the coefficients a and b possibly lead to reduced *BMEmean* predictions

24

of the flu RNI $\rho_{s,t}$. This is a very useful feature that could enable fine-tuning of the coefficients a and b in real-world studies based on the proposed models in Eq. (1).

4. Discussion

Spatiotemporal BME analysis is known to tackle successfully certain shortcomings of mainstream spatial statistics and geostatistics. This work takes advantage of one aspect of this analysis, namely the BME ability to integrate physical-based general knowledge bases. Specifically, using main features of the KS framework behind BME theory, we introduce a family of new covariance models that are constructed from epidemic propagation models, and we illustrate how to account for them as general knowledge bases for environmental health spatiotemporal prediction. As a case study, we examined the spatiotemporal rate of new infectives in a flu spread across France during a period of 15 weeks. Our investigation indicates that the KS methodology embraces the creation of an enhanced selection of knowledge bases to choose from and apply in different tasks as appropriate, as opposed to using standard mathematical constructs that might lack the elements of physical depth and foundation. This feature enables more accurate description of an attribute, and provides the freedom to experiment with a broader range of scenarios in different environmental health study designs.

The ratio of new infectives RNI is modeled as a stochastic S/TRF to study its behavior with the proposed new covariances. BME analysis produces maps of the predicted RNI PDF means around the time of the highest flu activity in the year,

according to the available observations. The main features of the new model-based covariances are explored by investigating the behavior of their components and the corresponding coefficients with the help of illustrative examples and carefully designed experiments. The generated maps and associated space-time correlation analysis offer informative representations of the RNI attribute and can play a strong supporting role to environmental assessment specialists in different ways. For example, the predicted RNI can indicate the spatial behavior of the propagation ratio during a particular season of the year; it can lead experts to identify possible causal mechanisms (such as the role of geographical location in outbursts, and the event propagation speed across space-time) and their interrelations; the generated spatiotemporal patterns can be useful for disentangling the influence of explanatory factors, or even for identifying some influential ones that could have been otherwise overlooked.

In sum, and in a more general context, a knowledge synthesis framework that combines efficient methodologies of integrating new knowledge bases (that can broaden the range of informative resources and improve the inference process) with practical software tools (for the physically meaningful implementation of the theory) could point to new and fruitful directions of research in environmental assessment and mapping.

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest, nor has this ever been an issue in question during the entire period of research for this paper.

References

Allen, L.J.S., & Burgin, A.M. (2000). Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Mathematical Biosciences* **163**: 1-33.

Angulo, J.M., & Christakos, G. (2010). A stochastic model of spatiotemporal infectious disease spread. *SDSU Medical Geography Res Rep* 12-09.

Angulo, J.M., Yu, H-L, Langousis, A., Madrid, D., and Christakos, G. (2012). Modeling of space-time infectious disease spread under conditions of uncertainty. *International J of Geographical Information Science.* In press.

Arrow, K., Dasgupta, P., & Maler, K.-G. (2003). Evaluating projects and assessing sustainable development in imperfect economics. *Environmental & Resource Economics* **26**: 647-85.

Atkinson, P.M. (1993). The effect of spatial resolution on the experimental variogram of airborne MSS imagery. *International Journal of Remote Sensing* **14**: 1005–11.

Banerjee, S., Carlin, B.P., & Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data (Monographs on Statistics and Applied Probability).* Boca Raton, FL: Chapman & Hall/CRC Press.

Bayraktar, H., & Turalioglu, F.S. (2005). A Kriging-based approach for locating a sampling site—in the assessment of air quality. *Stochastic Environmental Research and Risk Assessment* **19**: 301-5.

Bivand, R.S., Pebesma, E.J., Gómez-Rubio, V. (2008). *Applied Spatial Analysis with R*. New York, NY: Springer.

Bossak, B.H., & Welford, M.R. (2009). Did medieval trade activity and a viral etiology control the spatial extent and seasonal distribution of Black Death mortality? *Medical Hypotheses* **72**: 749-752.

Chen, C.-C., Wu, K.-Y., & Chang, M.J.W. (2004). A statistical assessment on the stochastic relationship between biomarker concentrations and environmental exposures. *Stochastic Environmental Research and Risk Assessment* **18**: 377-85.

Choi, K.-M., Serre, M.L., & Christakos, G. (2003). Efficient mapping of California mortality fields at different spatial scales. *J of Exposure Analysis and Environmental Epidemiology* **13**: 120-33.

Choi, K.M., Yu, H.-L., & Wilson, M.L. (2007). Spatiotemporal statistical analysis of influenza mortality risk in the State of California during the period 1997–2001. *Stochastic Env Res and Risk Assessment* **22**(Supplement 1): 15-25.

Christakos, G. (1991a). On certain classes of spatiotemporal random fields with application to space-time data processing. *IEEE Trans Systems, Man, and Cybernetics* **21**: 861-75.

Christakos, G. (1991b). Some applications of the BME concept in Geostatistics. In: *Fundamental Theories of Physics*, 215-229. Amsterdam, The Netherlands: Kluwer Acad Publ.

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

Christakos, G. (1998). Spatiotemporal information systems in soil and environmental sciences. *Geoderma* **85**: 141-79.

Christakos, G. (2000). *Modern Spatiotemporal Geostatistics*. New York, NY: Oxford Univ Press.

Christakos, G. (2008). Bayesian Maximum Entropy. In: Kanevski M, editor. *Advanced Mapping of Environmental Data: Geostatistics, Machine Learning, and Bayesian Maximum Entropy.* New York, NY: J. Wiley & Sons, p. 247-306.

Christakos, G. (2010). *Integrative Problem-Solving in a Time of Decadence.* New York, NY: Springer-Verlag.

Christakos, G., Bogaert, P., & Serre, M.L. (2002). *Temporal GIS*. New York, NY: Springer-Verlag.

Christakos G., Kolovos A., Serre M.L., & Vukovich F. (2004). Total Ozone Mapping by Integrating Data Bases From Remote Sensing Instruments and Empirical Models. *IEEE Trans on Geosc and Rem Sensing* **42**:991-1008.

Christakos, G., & Hristopulos, D.T. (1998). *Spatiotemporal Environmental Health Modelling: A Tractatus Stochasticus*. Boston, MA: Kluwer Acad Publ.

Christakos, G., & Kolovos, A. (1999). A study of the spatiotemporal health impacts of ozone exposure. *J of Exposure Anal & Env Epidem* **9**: 322-335.

Clayton, D., & Hills, M. (1993). *Statistical Models in Epidemiology*. New York, NY: Oxford Univ Press.

Deutsch, C.V., & Journel, A.G. (1992). *Geostatistical Software Library and User's Guide*. Oxford, UK: Oxford Univ Press.

Douaik, A., van Meirvenne, M., Toth, T., & Serre, M.L. (2004). Space-time mapping of soil salinity using probabilistic BME. *Stochastic Environmental Research and Risk Assessment* **18**: 219-27.

Escudero, L.F., Ortega, E.M., & Alonso, J. (2009). Variability comparisons for some mixture models with stochastic environments in biosciences and engineering. *Stochastic Env Res and Risk Assessment* **24**: 199-209.

Gneiting T. (2002). Nonseparable, Stationary Covariance Functions for Space-Time Data. *JASA* **97**(458):590–600.

Goovaerts P. (2008). Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences* **40**: 101-128.

Hethcote, H.W. (1994). A thousand and one epidemic models. In: Levin SA, editor. *Frontiers in Mathematical Biology, Lecture Notes in Biomathematics*. Berlin: Springer, Vol. 100.

Heuvelink, G.B.M. (1998). *Error Propagation in Environmental Modelling with GIS*. London, UK: Taylor & Francis Ltd.

Kolovos, A., Christakos, G., Hristopulos, D.T., & Serre, M.L. (2004). Methods for Generating Non-separable Spatiotemporal Covariance Models With Potential Environmental Applications. *Adv in Water Res* **27**: 815-30.

Kolovos, A., Yu, H.-L., & Christakos, G. (2006). *SEKS-GUI v.0.6 User Manual*. San Diego, CA: Dept of Geography, San Diego State University.

Kuo, H.I., Lu, C.-L., Tseng, W.-C., & Li, H.-A. (2009). A spatiotemporal statistical model of the risk factors of human cases of H5N1 avian influenza in South-east Asian countries and China. *J Public Health* **123**: 188-193.

Law, D.C.G., Bernstein, K., Serre, M.L., Schumacher, C.M., Leone, P.A., et al. (2006). Modeling an Early Syphilis Outbreak through Space and Time Using the Bayesian Maximum Entropy Approach. *Annals of Epidemiology* **16**:797-804.

Lee, S.-J., Balling, R., & Gober, P. (2008). Bayesian Maximum Entropy mapping and the soft data problem in urban climate research. *Annals of the Association of American Geographers* **98**: 309-22.

Liao, D., Peuquet, D.J., Duan, Y., Whitsel, E.A., Dou, J., et al. (2006). GIS approaches for the estimation of residential-level ambient PM concentrations. *Environmental Health Perspectives* **114**:1374-80.

Ma C. (2008). Recent developments on the construction of spatio-temporal covariance models. *Stoch Env Res and Risk Assess* 22:39–47.

Moore, D.A., & Carpenter, T.E. (1999). Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiology Reviews* **21**: 143-61.

Porcu E, Mateu J, & Saura F. (2008). New classes of covariance and spectral density functions for spatio-temporal modelling. *Stoch Env Res and Risk Assess* 22:65–79.

Renshaw E. (1993). *Modelling Biological Populations in Space and Time*, Cambridge, UK: Cambridge University.

Saito, H., McKenna, S.A., Zimmerman, D.A., & Coburn, T.C. (2005). Geostatistical interpolation of object counts collected from multiple strip transects: Ordinary Kriging versus finite domain Kriging. *Stochastic Environmental Research and Risk Assessment* **19**: 71-85.

Schabenberger, O., & Gotway, C.A. (2005). *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC Press.

Serre, M.L., Kolovos, A., Christakos, G., & Modis, K. (2003). An application of the holistochastic human exposure methodology to naturally occuring Arsenic in Bangladesh drinking water. *Risk Analysis* **23**: 515-28.

Serre, M.L., & Christakos, G. (1999). Modern geostatistics: computational BME analysis in the light of uncertain physical knowledge—the Equus Beds study. *Stochastic Environmental Research and Risk Assessment* **13**: 1-26.

Stein, A., van der Meer, F., & Gorte, B. (1999). *Spatial Statistics for remote Sensing*. Dordrecht, The Netherlands: Kluwer Acad Publ.

Yu, H.-L., Ku, S.-C., Yang, C.-H., Cheng, T.-J., & Chen, L. (2011). Assessment of Areal Average Air Quality Level over Irregular Areas: A Case Study of PM10 Exposure Estimation in Taipei (Taiwan). In: Nejadkoorki F., editor. *Advanced Air Pollution*. InTech, Available from: http://www.intechopen.com/articles/show/title/assessment-of-areal-average-air-quality-level-over-irregular-areas-a-case-study-of-pm10-exposure-est

Wong, D.W.S., & Lee, J. (2005). *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. New York, NY: Wiley.

Wu, J.L., Wang, J.F., Meng, B., Chen, G., Pang, L., et al. (2004). Exploratory spatial data analysis for the identification of risk factors to birth defects. *BMC Public Health* **4**:23.

Tables

	$\overline{X}_{s,t+1} = A_t \overline{X}_{s,0}$	$\overline{Y}_{s,t+1} = 1 - (1+B_t)\overline{X}_{s,0}$	$\overline{Z}_{s,t+1} = 1 - C_t \overline{X}_{s,0}$

	$X_{s',t'+1}$	$Y_{s',t'+1}$	$Z_{s',t'+1}$
$X_{s,t+1}$	$A_t A_{t'} c_{X;s-s',0}$	$-A_t(1+B_{t'})c_{X;s-s',0}$	$A_t C_{t'} c_{X;s-s',0}$
$Y_{s,t+1}$	$-A_{t'}(1+B_t)c_{X;s-s',0}$	$(1+B_t)(1+B_{t'})c_{X;s-s',0}$	$-(1+B_t)C_{t'} c_{X;s-s',0}$
$Z_{s,t+1}$	$A_{t'} C_t c_{X;s-s',0}$	$-(1+B_{t'})C_t c_{X;s-s',0}$	$C_t C_{t'} c_{X;s-s',0}$

Table 1: Means, covariances, and cross-covariances of X_p, Y_p, and Z_p.

| Coefficient | Mean | Beta Distribution Shape Parameters | | | |
| | | 2% Standard Deviation | | 10% Standard Deviation | |
		α	β	α	β
a	0.143	2142.2	12833	85.4353	507.0212
b	0.534725093	1162.7	1011.8	46.0741	40.2198
γ	0.058320202				
ε	1.024037215				

Table 2: Coefficient values used in the study. The means of the coefficients b, γ, and ε are fitted values, whereas the mean of the coefficient a is based on a physical assumption, as explained in the text. α and β are the shape parameters of the beta distribution for each of the parameters a and b in the two experiments of 2% and 10% standard deviation examined in the study.

Figure Legends

Fig. 1 Map of France and its 21 mainland regions

Fig. 2 Time series of the flu ratio of new infectives (RNI) $\rho_{s,t}$ (vertical axes) at the x (Easting) and y (Northing) centroid coordinates of five selected regions in France for the 53-week period starting on Aug. 17, 1988. Coordinate units are in m. The vertical axes show the weekly RNI values

Fig. 3 Locations of the flu ratio of new infectives (RNI) $\rho_{s,t}$ data across France and $\rho_{s,t}$ values at $t = 2$. Each one of the color-filled circles is located at the corresponding region centroid, and the color indicates the aggregate ratio $\rho_{s,t}$ at $t = 2$

Fig. 4 Histogram of the flu RNI $\rho_{s,t}$ data set values distribution

Fig. 5 Mean trend of the flu RNI $\rho_{s,t}$ at selected time instances: (a) week $t = 2$, (b) week $t = 6$, (c) week $t = 10$, and (d) week $t = 14$

Fig. 6 Plot of the spatial spherical covariance at $t = 0$ (subplot a). Plots of the factors A_t (subplot b), B_t (subplot c), and C_t (subplot d) with time. The black dashed line of the function plot designates the factor values by assuming exact values (no variance) for the coefficients a and b in the factors expressions. The dot-dashed line (mostly overlapping the dashed line) and the solid line designate factor values by assuming random, beta-distributed coefficients with 2% and 10% standard deviation, respectively

Fig. 7 *BMEmean* of the flu RNI $\rho_{s,t}$ prediction PDFs at selected time instances: (a) week $t = 2$, (b) week $t = 6$, (c) week $t = 10$, and (d) week $t = 14$. The RNI values are based on the settings for the experiment A

Fig. 8 *BMEmean* of the flu RNI $\rho_{s,t}$ prediction PDFs at selected time instances: (a) week $t = 2$, (b) week $t = 6$, (c) week $t = 10$, and (d) week $t = 14$. The RNI values are based on the settings for the experiment B

Fig. 9 Standard deviation of the flu RNI $\rho_{s,t}$ prediction PDFs at selected time instances. For the week $t = 2$: (a) experiment A, (b) experiment B. For the week $t = 6$: (c) experiment A (d) experiment B. The color scale is different for each plot to enable viewing the maximum standard deviation values in each case. For comparison purposes, it is relatively easy to see that at the same instances the majority of error values are very similar for both experiments. Specifically, at $t = 2$ the standard deviations range at about $3 \cdot 10^{-4}$, whereas at $t = 6$ they range below $5 \cdot 10^{-4}$

Fig. 10 Average $\overline{BMEmean}$ from 100 flu RNI $\rho_{s,t}$ predictions at week $t = 7$ for the settings of (a) experiment A, and (b) experiment B. (c) *BMEmean* of the flu RNI $\rho_{s,t}$ prediction PDFs at week $t = 7$, based on the experiment A settings. (d) Map of the absolute differences $\overline{BMEmean}_A - \overline{BMEmean}_B$
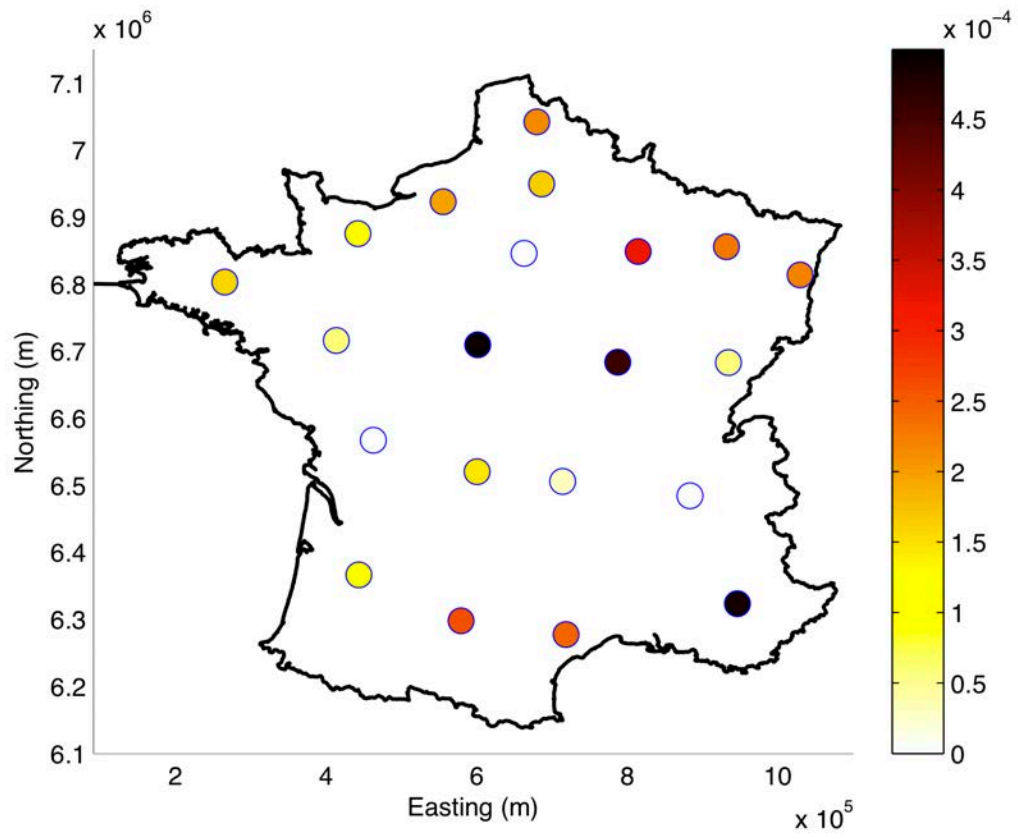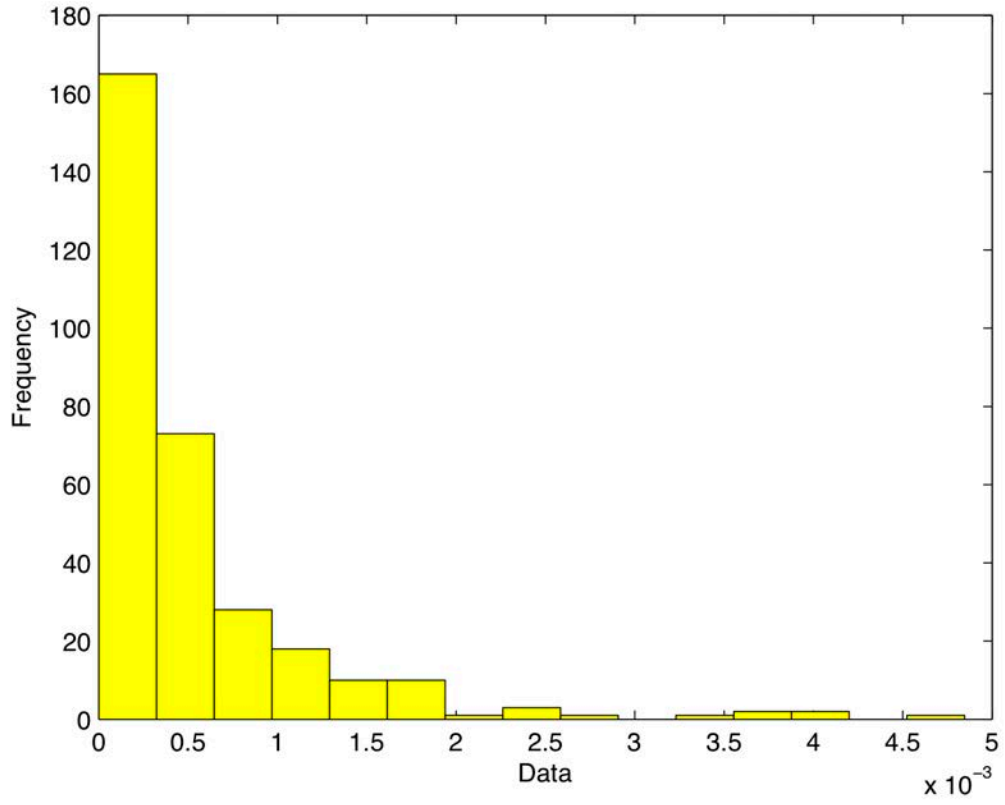
Fig. 1

Fig. 2
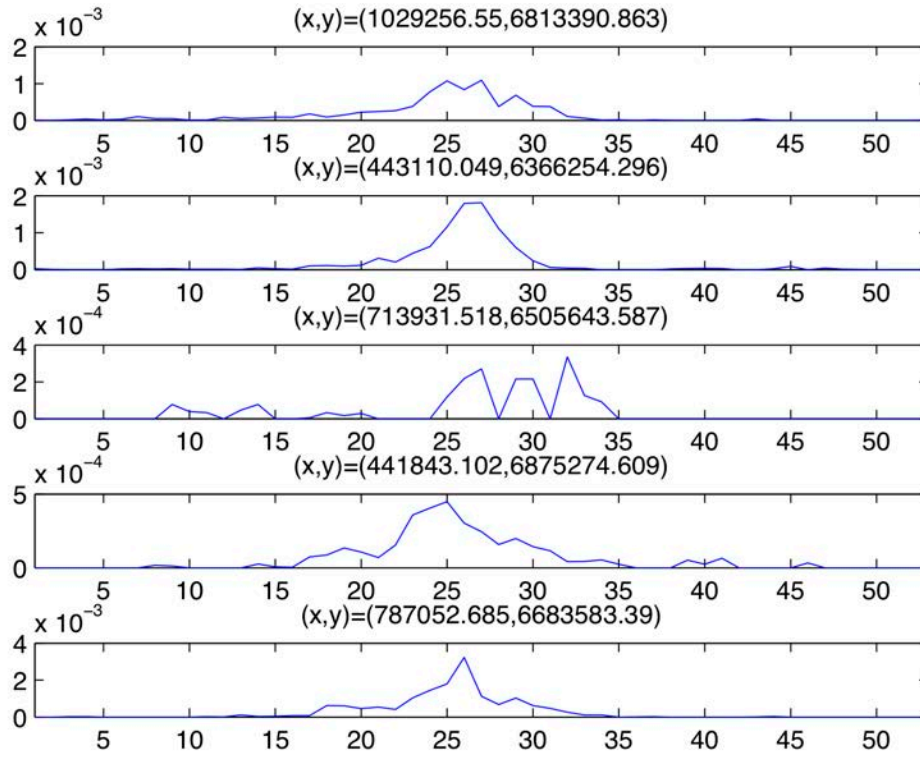
Fig. 3

Fig. 4

Fig. 5
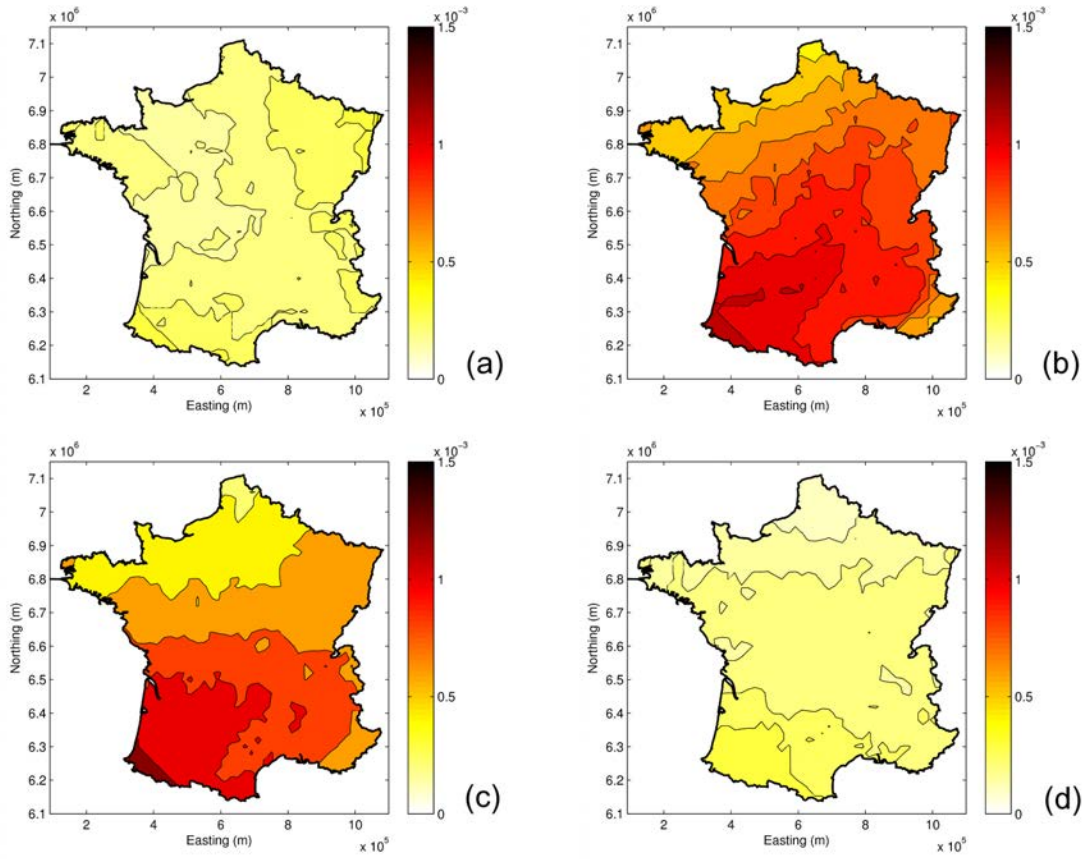
Fig. 6

Fig. 7

Fig. 8

Fig. 9

Fig. 10