Stoch Environ Res Risk Assess (2018) 32:859–877
https://doi.org/10.1007/s00477-017-1419-7

REVIEW PAPER

Bayesian maximum entropy approach and its applications: a review

Junyu He[1] · **Alexander Kolovos**[2]

Published online: 28 April 2017

Abstract The present paper reviews the conceptual framework and development of the Bayesian Maximum Entropy (BME) approach. BME has been considered as a significant breakthrough and contribution to applied stochastics by introducing an improved, knowledge-based modeling framework for spatial and spatiotemporal information. In this work, one objective is the overview of distinct BME features. By offering a foundation free of restrictive assumptions that limit comparable techniques, an ability to integrate a variety of prior knowledge bases, and rigorous accounting for both exact and uncertain data, the BME approach was coined as introducing modern spatiotemporal geostatistics. A second objective is to illustrate BME applications and adoption within numerous different scientific disciplines. We summarize examples and real-world studies that encompass the perspective of science of the total environment, including atmosphere, lithosphere, hydrosphere, and ecosphere, while also noting applications that extend beyond these fields. The broad-ranging application track suggests BME as an established, valuable tool for predictive spatial and space–time analysis and mapping. This review concludes with the present status of BME, and tentative paths for future methodological research, enhancements, and extensions.

Keywords Bayesian maximum entropy · Applied stochastics · Spatial · Spatiotemporal · Geostatistics · Uncertainty

✉ Alexander Kolovos
alexander.kolovos@spacetimeworks.com

1 Institute of Islands and Coastal Ecosystems, Ocean College, Zhejiang University, Zhoushan, China

2 SpaceTimeWorks, San Diego, CA, USA

1 Introduction

Geostatistics is an applied stochastics field that uses spatial/spatiotemporal (S/ST) referenced information to provide insight about a natural attribute within an areal and/or temporal extent of interest. Geostatistics provides a collection of random field tools to predict[1] values of continuous variables and their associated uncertainty in uncharted fields based on prior knowledge that we have about the variables. It was initially formed and used in the mining sector by Krige (1952), and subsequently grew to become a high-efficiency, low-cost series of methodologies with successful application in fields like geography, environment, agronomy, oceanography, geology, hydrology, ecology, remote sensing, epidemiology, and public health. The range of its topics mainly includes environmental investigation/evaluation, resources exploration, policy/decision making and management (see, e.g., Matheron 1962, 1965; David 1977; Journel and Huijbregts 1978; Olea 1999; Webster and Oliver 2007). As geostatistics seeks to enhance understanding and detail in the description of our natural environment, developments in geostatistics involve proposing new methods or improvement of existing ones to achieve analytical results of higher resolution and increased accuracy.

Among the available geostatistical methods, the kriging family of techniques has been traditionally one of the most popular for the study of spatial characteristics of an attribute (Olea 1999). Kriging is more accurate than some

[1] In geostatistical literature, the terms "prediction" and "estimation" are often used interchangeably. In the following, we distinguish between the two terms by following the statistical vernacular. In it, prediction is the inference of random variable values, whereas estimation designates the inference tasks where unknown parameters are involved.

analytically rudimentary spatial analysis techniques such as Thiessen polygons and inverse distance interpolation (Tabios and Salas 1985) as well as remote sensing techniques (Lee et al. 2012). When applied to geostatistical mapping, kriging takes into account an attribute's mean (surface) trends, spatial covariance or semivariance, and observed attribute values. As such, it is a kriging framework limitation that kriging can handle only first and second order S/ST moments; that is, kriging ignores any knowledge bases of higher order S/ST moments that might be available, or other knowledge bases it is unable to integrate like physical laws and conceptual models. Kriging has also had limited results in incorporating uncertain data in a precise and efficient manner; see, for example, Savelyeva et al. (2010) whose approach on kriging with measurement errors requires assumptions about the measurement errors to consider, namely that the errors are uncorrelated among themselves and the measured value, and that the error variances are known. Such limitations reflect in the accuracy of kriging-based prediction and its ability to extrapolate (Christakos and Serre 2000).

The introduction of Bayesian maximum entropy (BME; Christakos 1990) provided a new approach to address this series of shortcomings through a more broadly designed epistemic framework for geostatistical analysis and mapping. BME is a knowledge-centered approach that can integrate informative content from different sources based on a rigorous theoretical support of considerable generality to achieve improved prediction accuracy (Christakos 2000, 2010). As an example, assessing the soil heavy metal distribution around a mine obviously presumes the collection of soil samples to measure the concentration of heavy metals. Accuracy in this analysis can be tentatively improved by including additional useful information about the soil types, soil moisture, rainfall, flowing direction and speed of surface runoff and groundwater, because these factors can influence the migration of heavy metals in the soil. BME provides the foundation to integrate such knowledge bases either as additional observed data or in general knowledge forms such as physical laws, basic principles, and empirical models (Kolovos et al. 2012). This is a particularly attractive feature in the contemporary data-driven analytical environments, because BME serves as an enabler to generate and integrate data from general knowledge bases that might be otherwise left unused (Kolovos et al. 2002). In addition to general knowledge sources, the BME framework uses measured data values as site-specific knowledge. Most prominently, BME extends our ability to use measured data beyond single-value observations. Specifically with respect to uncertainty, BME discriminates between two types of data; namely, the hard data and the soft data. Data whose observations are considered to be exact in the scope of an analysis are termed hard data. In contrast, soft data represent observed data that might carry nontrivial uncertainty in the context of a study. For example, soft data can be the result of obvious sources of uncertainty such as inaccuracy in measuring devices, modeling uncertainties, and human error. Alternatively, soft data can be derived as a result of inadequate knowledge, revised assessments, intuition estimates, and a host of other similar mechanisms that may provide uncertain informative content (Kolovos et al. 2016). The BME framework enables rigorous incorporation of both hard and soft data types. In the steps of BME, an additional effort that utilized a Bayesian approach to integrate soft data is the Bayesian data fusion formulation proposed by Bogaert and Fasbender (2007). BME is also free of additional limitations and restrictions that burden other geostatistical methods; for example, the BME analysis is independent of the data distribution (e.g., non-Gaussian distributions can be automatically incorporated), whereas kriging is unable to accurately handle heavy tailed data (Christakos et al. 2001). In addition, the BME theoretical foundation enables both univariate and multivariate prediction (Christakos 2000; Hristopulos and Christakos 2001). Per the detailed methods comparison in Christakos (2000), in the course of the decades since its introduction BME has been providing equally or more accurate prediction patterns than kriging (e.g., Lee et al. 2008, 2010; Bogaert et al. 2009; Li et al. 2013c; Gao et al. 2014; Shi et al. 2015a).

The Bayesian maximum entropy approach combines the maximum entropy theory with operational Bayesian statistics to construct its scientific mathematical framework for S/ST analysis and mapping (Christakos 1990). On one end, the maximum entropy theory enables handling general and site-specific multi-sourced knowledge bases and establishes a foundation to integrate them. In BME, maximization of information input serves as acquiring prior knowledge that can be filtrated, integrated and optimized by specified rules, and is a pivotal step to describe general attribute characteristics. On the other end, the Bayes rule updates this prior knowledge through sets of available site-specific hard and soft data to bring the model description closer to natural truth. By design, BME can be considered as part of the larger family of Bayesian hierarchical modeling (BHM) methods. BHM methods are well known for offering model-based analysis and handling different types of information and uncertain input. In addition, they can also process multiple data sets and accommodate exogenous information and uncertainty in their analysis. BHM has been accordingly applied in many disciplines, and offers foundation for both spatial and space–time analysis (e.g., Banerjee et al. 2004; Le and Zidek 2006). In this light, BME is a nonparametric instance of a BHM approach that applies an operational Bayesian rule, and furthermore uses maximum entropy as the conditionalization principle instead of making a distributional assumption. There have been no comparison studies between BME and other BHM methodologies, and this is presently ground for additional research. Conceptually, BME is more explicitly driven by physical considerations, while BHM methods typically rely on more elaborate statistical foundation. In this foundation, for example, application of the Bayes rule allows for non-informative priors, whereas BME priors always rely on background general knowledge.

Since its inception, BME has exhibited a robust ability for S/ST analysis through a wide spectrum of applications in environmental geology, environmental sciences, resources science, public health, ecology, remote sensing, energy, real estate research, etc. A sample of literature references includes Christakos and Hristopulos (1998), Christakos et al. (2004), Heywood et al. (2006), Law et al. (2006), Brus et al. (2008), Kolovos et al. (2010), Angulo et al. (2013), Shi et al. (2015b), Sun et al. (2015), Yang and Christakos (2015), Zagouras et al. (2015), Hayunga and Kolovos (2016).

In the following, we review contemporary BME research and discuss its application scope. First, the BME framework and analytical process are described. Then, we briefly emphasize the breakthrough role and contribution of BME to geostatistics, and review software tools for practical implementation of BME. In a subsequent step, recent BME applications and real-world case studies are reviewed. Finally, we bring forward potential future directions for BME research.

2 Basic theory of Bayesian maximum entropy

2.1 BME foundation

2.1.1 *The spatiotemporal random field model*

The spatiotemporal random field (STRF, Christakos 1991) is a powerful theoretical framework to study natural processes and phenomena that evolve in the spatial and/or space–time continuum. The STRF theory emerged from the largely non-deterministic character of natural processes in an effort to describe them in a stochastic framework. In the following, we present a summary of the main STRF concepts and characteristics that form the backbone of geostatistical S/ST analysis and BME.

In the STRF theory, a natural process that evolves in space–time, such as distribution of elements in soil or atmosphere, can be represented as an STRF that is designated by a variable $X(\boldsymbol{p})$. The vector $\boldsymbol{p} = [\boldsymbol{p_1}\ldots\boldsymbol{p_m}]$ indicates point locations in space and time, where each location $\boldsymbol{p_i} = (\boldsymbol{s_i},\boldsymbol{t_i})$, $i = 1, \ldots, m$, comprises of the spatial coordinate $\boldsymbol{s}_i$ in n spatial dimensions and the temporal coordinate t_i. These coordinates define locations in the Euclidean space–time domain $R^n \times T$, where R^n designates the spatial n dimensions and T stands for the temporal dimension. For example, for $n = 2$, $\boldsymbol{s}$ refers to planar geographical coordinates. The STRF $X(\boldsymbol{p})$ is a collection of individual random variables (RV). Sampling or simulation of a random variable $x(\boldsymbol{p})$ yields instances of $x(\boldsymbol{p})$; each one of these values is a realization $\chi(\boldsymbol{p})$ of the RV. Respectively, the vector $\boldsymbol{\chi} = [\chi_1\ldots\chi_m]$ represents a group of m sampled or simulated values of the random variables $\boldsymbol{x} = [x_1\ldots x_m]$ at space–time point locations $\boldsymbol{p} = [\boldsymbol{p_1}\ldots\boldsymbol{p_m}]$ (Christakos and Li 1998).

From a stochastic viewpoint, a random variable x_i is fully described by its probability density function (PDF). To define a PDF requires knowledge of the RV statistical characteristics given by its stochastic moments such as the mean value μ (first-order moment), variance σ^2 (second-order moment) and higher-order moments. Based on this perspective, one can compute the occurrence probability of an arbitrary realization value χ, which is given by Eq. 1 as follows:

$$P_X(\boldsymbol{\chi}) = \Pr\mathrm{ob}\left(\prod_{i=1}^{m} \chi_i \le x_i \le \chi_i + d\chi_i\right) = f_X(\boldsymbol{\chi})d\boldsymbol{\chi} \qquad (1)$$

where $f_X(\boldsymbol{\chi})$ is the multivariate PDF of points at locations $\boldsymbol{p} = [\boldsymbol{p_1}\ldots\boldsymbol{p_m}]$.

In the context of an STRF, the mean function $m_X(\boldsymbol{p})$ represents trends and systematic structures that exist in a larger scale than the study scale, and the covariance function $c_X(\boldsymbol{p}, \boldsymbol{p}')$ describes correlations and dependencies between two locations $\boldsymbol{p}$ and $\boldsymbol{p}'$ (Christakos 1991). The expressions for the mean and covariance functions are given by Eqs. 2 and 3, respectively:

$$m_X(\boldsymbol{p}) = E[X(\boldsymbol{p})] \qquad (2)$$

$$c_X(\boldsymbol{p},\boldsymbol{p}') = E[X(\boldsymbol{p}) - m_X(\boldsymbol{p})][X(\boldsymbol{p}') - m_X(\boldsymbol{p}')] \qquad (3)$$

A special case of interest is when the STRF is characterized by spatial and temporal stationarity.[2] In this case, it is only the relative distance between any couple of locations that affects the covariance function. Specifically then, the covariance function has the same value $c_X(\boldsymbol{p},\boldsymbol{p}') = c_X(\boldsymbol{r},\tau)$ for any location pair $(\boldsymbol{p}, \boldsymbol{p}')$ separated by the same spatial distance vector $\boldsymbol{r} = \boldsymbol{s}' - \boldsymbol{s}$ and same temporal distance lag $\tau = t' - t$ (Christakos and Serre 2000).

2.1.2 *Knowledge bases*

Other than a theoretical framework, to perform S/ST analysis we also need information about the natural process of interest. Mainstream geostatistical approaches follow a statistics-inspired data-driven workflow, where inference

[2] In geostatistical literature, the term "homogeneity" is commonly used to refer to what the statistical vernacular calls "spatial stationarity".

depends solely on information contained in observed values. Yet our knowledge about a natural process might extend beyond sampled observations and traditional data values. In fact, natural processes are commonly driven by physical laws, or governed by basic principles, or may be explained by scientific theories. If applicable in a given study, the informative content in such general forms of knowledge can steer confidently stochastic predictive tasks. Hence, using general knowledge relieves an analysis from having exclusive dependence on observed data. S/ST analyses can benefit significantly from this element because (*i*) there could be cases where an inadequate number of data might be available for conclusive analysis, and (*ii*) uncertainty in data can bias results, and might even lead to flawed analyses if unacknowledged.[3] To this end, general knowledge bases and the ability to integrate their informational content play a central role in BME, and Sect. 2.1.3 expands on this aspect. BME considers all available information about a natural process, and distinguishes between the general knowledge bases G, and the site-specific knowledge bases S. The site-specific knowledge S encompasses a variety of data forms that extends beyond single observed values, as detailed in Sect 2.1.4. BME defines the sum K of all available knowledge bases as $K = G \cup S$.

2.1.3 Information and entropy

To integrate general knowledge bases G, BME utilizes the Shannon information measure (IM) to quantify the informative content in G (Shannon 1948). The Shannon IM states that the higher the probability that a model predicts a process, the less informative this model is. The concept is intuitively sensible: A model that accounts for all possible outcomes will predict a process correctly, yet provides no valuable information about the process itself because no specific behavior is implied to explain how the process works. The Shannon IM is expressed as

$$Inf\left(\boldsymbol{x}_{map}\right) = -\log f_G\left(\boldsymbol{\chi}_{map}\right), \tag{4}$$

where $Inf(\boldsymbol{x}_{map})$ is the information about the STRF through the random vector $\boldsymbol{x}_{map}$ across the entire domain of interest, and $f_G(\boldsymbol{\chi}_{map})$ is the G-associated multivariate PDF that gives the probability of possible realizations $\boldsymbol{\chi}_{map}$. In practice, the expectation H of $Inf(\boldsymbol{x}_{map})$ is commonly used. In Information Theory, the quantity H is also known as entropy and is derived from Eq. 4 as

$$H = \overline{Inf(\boldsymbol{x}_{map})} = -\int d\boldsymbol{\chi}_{map} f_G(\boldsymbol{\chi}_{map}) \log f_G(\boldsymbol{\chi}_{map}). \tag{5}$$

Ideally, one would like to employ G that is relevant and accurate to the process. The more detailed G is, the thinner $f_G(\boldsymbol{\chi}_{map})$ becomes by excluding irrelevant outcomes, hence leading to a more informative model. BME uses the Shannon IM formalization in Eq. 5 to integrate G in a stochastic perspective. Furthermore, BME seeks to maximize the information if can get from G by maximizing the entropy H. This workflow, examined in more detail in Sect 2.2.1, explains the "maximum entropy" component in the BME name.

2.1.4 S/ST data types

A common goal in S/ST analysis is to obtain predicted values χ_k at unsampled locations $\boldsymbol{p}_k$, given a set of observed values $\boldsymbol{\chi}_{data} = [\chi_1 \ldots \chi_m]^T$ at locations $\boldsymbol{p}_i$, $i = 1,\ldots,m$, $i \neq k$. We designate as $\boldsymbol{\chi}_{map} = [\boldsymbol{\chi}_{data}\, \boldsymbol{\chi}_k]$ the entire set of values at locations $\boldsymbol{p}_{map} = [\boldsymbol{p}_i\, \boldsymbol{p}_k]^T$. With respect to uncertainty in the data values, $\boldsymbol{\chi}_{data}$ are discriminated into two groups, namely, the groups of hard data $\boldsymbol{\chi}_{hard}$ and soft data $\boldsymbol{\chi}_{soft}$, according to the guidelines presented in Sect. 1. Collectively, $\boldsymbol{\chi}_{data}$ constitute the BME site-specific knowledge S. In more detail:

$$\boldsymbol{\chi}_{data} = [\boldsymbol{\chi}_{hard}\, \boldsymbol{\chi}_{soft}]^T = [\chi_1 \cdots \chi_m]^T \tag{6}$$

- Hard data: $\boldsymbol{\chi}_{hard} = [\chi_1 \cdots \chi_{m_h}]^T$ at points $\boldsymbol{p}_i$, $i = 1, \ldots, m_h$. These observed single values can be derived from sources like devices, lab experiments, monitoring stations, etc. The probability of the value in each datum is 1.
- Soft data: $\boldsymbol{\chi}_{soft} = \left[\chi_{m_h+1} \ldots \chi_m\right]^T$ at points $\boldsymbol{p}_i$, $i = m_h + 1, \ldots, m$. Imprecision or vagueness in these uncertain observations is coded by intervals of values or probability density functions. Commonly encountered soft data forms are interval data, probabilistic data, and functional data (Christakos and Li 1998), as shown in Eqs. 7 in the following:

$$\chi_{soft} = \left\{[\chi_{m_h+1} \ldots \chi_m]^T | \chi_i \in I_i = [l_i, u_i],\ i = m_h + 1, \ldots, m\right\} \tag{7.1}$$

$$\chi_{soft} = \left\{[\chi_{m_h+1} \ldots \chi_m]^T | P(\chi_i \leq \zeta_i) = F_i(\zeta_i),\ i = m_h + 1, \ldots, m\right\} \tag{7.2}$$

$$\chi_{soft} = \left\{[\chi_{m_h+1} \ldots \chi_m]^T | P(\chi_i \leq I_i) = \vartheta_i,\ i = m_h + 1, \ldots, m\right\} \tag{7.3}$$

$$\chi_{soft} = \left\{[\chi_{m_h+1} \ldots \chi_m]^T | P[h(\chi_i) \leq \zeta_i] = F_i(\zeta_i, h),\ i = m_h + 1, \ldots, m\right\} \tag{7.4}$$

In particular, Eq. 7.1 shows that interval data values χ_i lie within a known interval I_i with lower and upper

[3] There are, of course, strong arguments in support of using observed data, too. For example, data-based results via well-designed experimental designs can be used to investigate potential flaws in unestablished theoretical rules and principles.

limits l_i and u_i, respectively; Eq. 7.2 displays probabilistic data for which the cumulative distribution functions (CDF) $F_i(\zeta_i)$ are known; Eq. 7.3 shows threshold soft data, where the probabilities ϑ_i for certain threshold values I_i are known; and Eq. 7.4 displays functional soft data, where the values can be derived from a given function h.

2.2 The BME workflow

The BME analysis is implemented in three stages, known as the prior stage, the metaprior stage, and the posterior stage (Fig. 1). The present section reviews the main characteristics of each stage in the following.

2.2.1 *Prior stage*

At the BME prior stage, general knowledge is collected, integrated, and analyzed to obtain the prior G-associated PDF $f_G(\boldsymbol{\chi}_{map})$. Subsection 2.1.2 indicated different sources of general knowledge G about an attribute. Subsection 2.1.3 explained the information principle used by BME to incorporate them. In this mechanism, a G-based mathematical formula that contains the attribute is used to construct expressions that give the statistical moments of the attribute. These moments constitute a series of constraints by which $f_G(\boldsymbol{\chi}_{map})$ is defined, as shown in the following.

First, the G-based constraints are built as

$$\overline{g_\alpha} = \int d\boldsymbol{\chi}_{map} f_G(\boldsymbol{\chi}_{map}) g_\alpha(\boldsymbol{\chi}_{map}) \tag{8}$$

where $\alpha = 1,\ldots,N_C$. For $\alpha = 0$, $g_0 = \overline{g_0} = 1$. The g_α are N_C functions that express the given general knowledge for an attribute. Their corresponding expectations $\overline{g_\alpha}$ act as

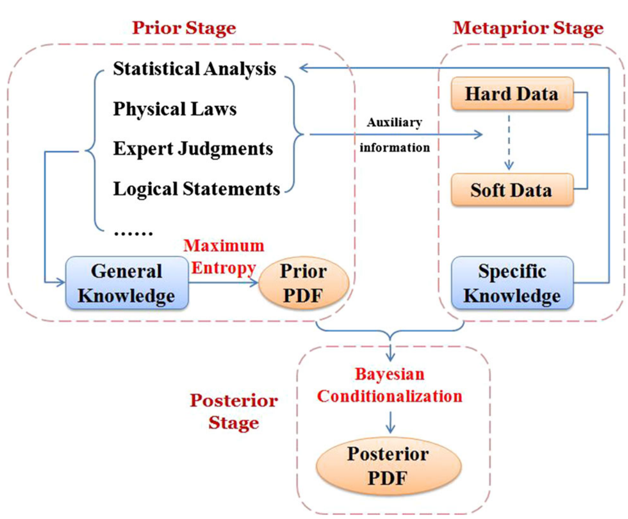

Fig. 1 Workflow of the BME analysis

informative constraints across $\boldsymbol{\chi}_{map}$ in the context of Eq. 5. Examples of $(g_\alpha, \overline{g_\alpha})$ pairs include forms of G such as the attribute mean $(\chi_i, \overline{x_i})$, covariance $\left((\chi_i - \overline{x_i})(\chi_j - \overline{x_j}), C_X(\boldsymbol{p}_i,\boldsymbol{p}_j)\right)$, semivariance $\left(\frac{1}{2}(\chi_i - \chi_j)^2, \gamma_X(\boldsymbol{p}_i,\boldsymbol{p}_j)\right)$, higher-order moments $\left((\chi_i - \overline{x_i})^\lambda, m_i^\lambda, \lambda = 1,\ldots,L\right)$, and even expressions of physical laws, for example $\left(\lambda b \chi^\lambda(t), \overline{dX^\lambda(t)}/dt, \lambda = 1,2\right)$. Given the g_α, the prior PDF $f_G(\boldsymbol{\chi}_{map})$ is expressed in the general form:

$$f_G(\boldsymbol{\chi}_{map}) = e^{\mu_0 + J}, J = \sum_{\alpha=1}^{N_C} \mu_\alpha(\boldsymbol{p}_{map}) g_\alpha(\boldsymbol{\chi}_{map}) \tag{9.1}$$

where μ_α are Lagrange multipliers (Ewing 1969), including the normalization-related μ_0.

According to Eq. 9.1, the prior PDF $f_G(\boldsymbol{\chi}_{map})$ is fully defined by the unknown μ_0 and μ_α coefficients. To compute them, we apply the maximum entropy approach by maximizing Eq. 5 with respect to $f_G(\boldsymbol{\chi}_{map})$, subject to the conditions $\overline{g_\alpha}$. For a set of N_C constraints, this task translates into solving the system of equations

$$\left.\begin{aligned} \overline{g_0} &= \int d\boldsymbol{\chi}_{map} f_G(\boldsymbol{\chi}_{map}) = 1 \\ \overline{g_1} &= \int d\boldsymbol{\chi}_{map} f_G(\boldsymbol{\chi}_{map}) g_1(\boldsymbol{\chi}_{map}) \\ &\vdots \\ \overline{g_{N_C}} &= \int d\boldsymbol{\chi}_{map} f_G(\boldsymbol{\chi}_{map}) g_{N_C}(\boldsymbol{\chi}_{map}) \end{aligned}\right\} \tag{9.2}$$

where $\overline{g_0}$ is used to define a normalization constant, and the unknowns are the μ_0 and μ_α coefficients. As a result, the prior stage produces a fully defined prior PDF $f_G(\boldsymbol{\chi}_{map})$. For illustration, Table 1 provides simple examples of conditions that are derived from a few selected general knowledge constraints. In particular, the table shows how to generate the corresponding functions $g_\alpha(\boldsymbol{\chi}_{map})$ in the cases where means, variances and covariances of $X(\boldsymbol{p})$ are known at $\boldsymbol{p}_i$, $i = 1,\ldots,n,k$.

Note that often, data-driven-only predictive techniques use the same observed data twice in their workflow: The first time is to compute empirically a trend and covariance of the attribute, and the second time is to condition the predictor. This is a self-referential workflow that is generally undesired, because the second task (prediction) depends on the outcome of the first task (trend and covariance analyses) by using the exact same data input. Although BME can replicate this workflow, it extends the features of mainstream predictive techniques by enabling specification of trends and covariances on the basis of other applicable G beyond reusing data.

Table 1 Condition functions from G-based constraints

General knowledge	α	g_α	$\bar{g}_\alpha$
Normalization	0	$g_0(\boldsymbol{\chi}_{map}) = 1$	$\bar{g}_0 = 1$
Means	$n+1\begin{cases} 1 \\ \vdots \\ n \\ n+1 \end{cases}$	$g_1(\chi_1) = \chi_1$ $\vdots$ $g_n(\chi_n) = \chi_n$ $g_{n+1}(\chi_k) = \chi_k$	$\bar{g}_1 = \bar{x}_1$ $\vdots$ $\bar{g}_n = \bar{x}_n$ $\bar{g}_{n+1} = \bar{x}_k$
Variances	$n+1\begin{cases} n+2 \\ \vdots \\ 2n+1 \\ 2n+2 \end{cases}$	$g_{n+2}(\chi_1,\chi_1) = (\chi_1 - \bar{x}_1)^2$ $\vdots$ $g_{2n+1}(\chi_n,\chi_n) = (\chi_n - \bar{x}_n)^2$ $g_{2n+2}(\chi_k,\chi_k) = (\chi_k - \bar{x}_k)^2$	$\bar{g}_{n+2} = \sigma_1^2$ $\vdots$ $\bar{g}_{2n+1} = \sigma_n^2$ $\bar{g}_{2n+2} = \sigma_k^2$
Covariances	$\frac{n(n+1)}{2}\begin{cases} 2n+3 \\ \vdots \\ \frac{(n+4)(n+1)}{2} \end{cases}$	$g_{2n+3}(\chi_1,\chi_2) = (\chi_1 - \bar{x}_1)(\chi_2 - \bar{x}_2)$ $\vdots$ $g_{\frac{(n+4)(n+1)}{2}}(\chi_n,\chi_k) = (\chi_n - \bar{x}_n)(\chi_k - \bar{x}_k)$	$\bar{g}_{2n+3} = C_X(\boldsymbol{p}_1,\boldsymbol{p}_2)$ $\vdots$ $\bar{g}_{\frac{(n+4)(n+1)}{2}} = C_X(\boldsymbol{p}_n,\boldsymbol{p}_k)$

2.2.2 *Metaprior stage*

The metaprior stage refers to collection of observed data, and additional data tasks performed as needed. These tasks may include data cleaning, transformations, and evaluation, identification, and discrimination into hard and soft data, as appropriate for the analysis.

2.2.3 *Posterior stage*

At the posterior stage, the prior PDF $f_G(\boldsymbol{\chi}_{map})$ is updated with the available site-specific knowledge S. In particular for given hard and soft data $\boldsymbol{\chi}_{data}$, the Bayes conditionalization rule is applied to $f_G(\boldsymbol{\chi}_{map})$ to produce the posterior PDF $f_K(\chi_k)$ as follows:

$$f_K(\boldsymbol{\chi}_k) = f_G(\chi_k|\boldsymbol{\chi}_{data}) = \frac{f_G(\chi_k,\boldsymbol{\chi}_{data})}{f_G(\boldsymbol{\chi}_{data})} \tag{10}$$

The specific form of the posterior PDF in Eq. 10 is determined on the basis of the available soft data types. For example, given interval (Eq. 7.1), probabilistic (Eq. 7.2) or functional (Eq. 7.4) soft data, the corresponding posterior PDFs are given, respectively, from the following equations:

$$f_K(\boldsymbol{\chi}_k) = \frac{\int_I f_G(\chi_k,\boldsymbol{\chi}_{data})d\boldsymbol{\chi}_{soft}}{\int_I f_G(\boldsymbol{\chi}_{data})d\boldsymbol{\chi}_{soft}} \tag{11.1}$$

$$f_K(\boldsymbol{\chi}_k) = \frac{\int_I f_G(\chi_k,\boldsymbol{\chi}_{data})dF(\boldsymbol{\chi}_{soft})}{\int_I f_G(\boldsymbol{\chi}_{data})dF(\boldsymbol{\chi}_{soft})} \tag{11.2}$$

$$f_K(\boldsymbol{\chi}_k) = \frac{\int_R dF(\zeta,h)\int_{I(\zeta)} f_G(\chi_k,\boldsymbol{\chi}_{data})d\boldsymbol{\chi}_{soft}}{\int_R dF(\zeta,h)\int_{I(\zeta)} f_G(\boldsymbol{\chi}_{data})d\boldsymbol{\chi}_{soft}} \tag{11.3}$$

where I, R, and $I(\zeta)$ are definition domains that depend on the soft data. Given the Eqs. 11.1–11.3, the following couple of examples illustrate how to obtain posterior PDFs in the case of interval and probabilistic soft data:

1. Interval soft data: Assume the dataset $\boldsymbol{\chi}_{hard} = (\chi_1,\chi_2,\chi_3)$ and $\boldsymbol{\chi}_{soft} = (\chi_4,\chi_5)$, where $\chi_i \in I_i = [l_i, u_i]$, $i = 4, 5$. Then, the posterior PDF can be calculated by Eq. 11.4:

$$f_K(\boldsymbol{\chi}_k) = \frac{\int_{I_4}\int_{I_5} f_G(\chi_k,\boldsymbol{\chi}_{data})d\chi_4 d\chi_5}{\int_{I_4}\int_{I_5} f_G(\boldsymbol{\chi}_{data})d\chi_4 d\chi_5} \tag{11.4}$$

2. Probabilistic soft data: Assume the dataset $\boldsymbol{\chi}_{hard} = (\chi_1,\chi_2,\chi_3)$ and $\boldsymbol{\chi}_{soft} = (\chi_4,\chi_5)$, where χ_4 and χ_5 have cumulative density functions $F(\chi_4)$ and $F(\chi_5)$, respectively, and $\chi_i \in I_i$, $i = 4, 5$. Then, the posterior PDF can be calculated by Eq. 11.5:

$$f_K(\boldsymbol{\chi}_k) = \frac{\int_{I_4}\int_{I_5} f_G(\chi_k,\boldsymbol{\chi}_{data})dF(\chi_4)dF(\chi_5)}{\int_{I_4}\int_{I_5} f_G(\boldsymbol{\chi}_{data})dF(\chi_4)dF(\chi_5)} \tag{11.5}$$

2.3 BME prediction and uncertainty assessment

Computation of the BME posterior PDFs $f_K(\boldsymbol{\chi}_k)$ at locations $\boldsymbol{p}_k$ provides complete statistical description of the predicted attribute, and enables us to extract statistical results as needed or as appropriate for a specific S/ST analysis. For example, measures such as the BME mean and BME mode values can be calculated from the posterior PDF to express the analysis attribute results according to different predictors; and prediction variance can be calculated to provide a measure of the BME prediction uncertainty. These measures are displayed in the following Eqs. 12.1, 12.2, and 12.3, respectively.

$$\hat{\chi}_{k,\mathrm{mean}} = \int \chi_k f_K(\chi_k) d\chi_k \tag{12.1}$$

$$\hat{\chi}_{k,\mathrm{mode}} = \max_{\chi_k} f_K(\chi_k) \tag{12.2}$$

$$\sigma^2 = \int (\chi_k - \hat{\chi}_{k,\mathrm{mean}})^2 f_K(\chi_k) d\chi_k \tag{12.3}$$

Most notably, consider the scenario where the following conditions are met concurrently:

1. The general knowledge G comprises only the attribute mean trend and covariance, that is the first 2 stochastic moments of the attribute
2. The site-specific knowledge S contains only hard data

In this case, Christakos (2000) shows that the BME equations reduce to kriging, thus rendering kriging a special case of BME prediction. In particular, the BME posterior PDF $f_K(\chi_k)$ is then fully defined by its mean and variance, and these measures correspond exactly to the kriging prediction and prediction error, respectively.

3 BME highlights and software implementation

3.1 Prior information

BME owes much of its strength to its versatile character that relies on key concepts from statistics (Bayes rule) and information theory (information maximization through maximum entropy). In the core of Bayesian statistics, prior information is a critical component for inference tasks, unlike the data-driven classical statistics approach that is based purely on sampling information (Ellison 2004).

One benefit of using prior information pertains to substantial savings in research time and costs associated with data collection and analysis; see McCarthy and Masters (2005), Kuhnert et al. (2010). Unlike classical statistics, the Bayesian approach can give value to existing information and ignore sampling that has not occurred. The value of this property is particularly evident in the analysis of natural processes that cannot be replicated to produce additional data samples, such as heavy metal concentration at a specific soil sampling point, or precipitation in a specific area. Often, in the presence of only few data samples, classical statistics might be unable to describe a random variable, whereas Bayesian statistics might perform precise inference with prior information (McCarthy and Masters 2005; Kuhnert et al. 2010).

Bayesian statistics also counts on probability as a direct measure of uncertainty to use prior information as a useful way to incorporate expert knowledge from life experience, education, skills or training (Shoemaker et al. 1999). For example, given the geological conditions at a specific area, a geologist might think there is 20% chance to discover groundwater in a depth of 5 meters. It is proved that in similar cases, prior information in the form of expert knowledge can make up for the deficiency or even entire lack of data to achieve accurate results within a Bayesian framework (Choy et al. 2009; Martin et al. 2012). In addition to expert knowledge, there may be other situations where physical laws apply, such as conservation principles, differential equations, multiple-point relationships, etc. It is then necessary to use these laws in building constraints for mathematical models that aspire accurate inference of a natural process (Christakos 2002). For example, Serre and Christakos (2002) showed that precise hydraulic head prediction is unattainable without utilizing physical laws.

To complement prior information in Bayesian statistics, the maximum entropy approach enables choosing useful knowledge and filtering out redundant knowledge. Maximum entropy displays strong ability to fuse multi-source prior information; for example, it can manage continuous and categorical data, and it can also address conditions with small, sparse or irregular data sets (Phillips et al. 2006; Urbani et al. 2015). In the presence of adequate applicable constraints, maximum entropy has been found to yield optimal approximation to a real world process (Jaynes 1957).

In essence then, BME uses prior information as auxiliary constraint information to guide towards more accurate S/ST prediction. Very commonly, this prior information may comprise of a physical law or principle that is applicable to the natural process of interest. As an example, Christakos and Hristopulos (1998) described differential equations built around S/ST dimensionality, which can depict spatiotemporal evolution of concentrations and fluxes. Also, Kolovos et al. (2002) presented a novel computational technique for a BME-based solution of an advection-reaction partial differential equation in a contaminated river environment. By using the physical law prior information, this technique can further improve prediction of the contaminant concentration that might have hereto relied only on data measurements. In a different example, Serre et al. (2003a) considered Darcy's law to solve an inverse problem for subsurface flow with BME that led to accurate prediction of hydraulic resistivity. In the above examples, prior information comes from physical laws in the form of differential equations that describe natural processes. Even though differential equations are generally used as deterministic expressions, their coefficients might vary in the S/ST domain. BME steps in as a capable tool to characterize this variation in a stochastic framework, and this enables using the physical law as integral component for the stochastic representation, simulation, and prediction of the natural process (Christakos and Hristopulos 1998).

3.2 Soft data

Another vital strength of the BME framework is rigorous handling of soft data. Commonly, exact measurements might cover inadequately the extent of a study area for the purpose of S/ST inference. In such cases, we might be able to enhance our information about a process if soft data observations are also available (e.g., Shi et al. 2015b). With respect to the soft data level of uncertainty, it is known that uncertainties can propagate from the model input to its output (Zhao and Kockelman 2002). Correspondingly for BME, research has shown that higher quality soft data, i.e., soft data that are characterized by relatively lower uncertainty and hence higher informative content than other soft data, can increase inference accuracy (Christakos and Serre 2000; Bogaert and Fasbender 2007; Lee et al. 2008; Adam-Poupart et al. 2014). Moreover, D'Or et al. (2001) showed that to a certain extent, increasingly larger ratios of soft-to-hard data can lead to higher accuracy BME prediction. It is also possible to experience scarcity or entire absence of hard data. Even in extreme situations where no hard data might be available at all, BME still enables prediction if soft data are present by using them alone (Hayunga and Kolovos 2016). Therefore, in the BME framework any informative content, even if imperfect in the form of soft data, is important for S/ST inference.

Due to the critical role played by the quality and amount of informational content in soft data, it is important that researchers can assess early on the value of any available uncertain information. One possible path to this assessment is cross-examination with secondary resources to consolidate a level of content integrity. Such resources can be, e.g., historical records, expert opinions, valid beliefs, laws and principles; see, for example the exhaustive detail in the detective work done by Christakos et al. (2005) to yield the site-specific datasets for their study. Clearly, there is no predictive value in using soft data with little or no informational content.

In the presence of soft data, non-BME techniques that are unable to account as rigorously or at all for soft data might find tempting to include soft information in some manner to improve prediction. This attempt is known as hardening of soft data, where a soft datum is replaced by a single value and thus soft information is misrepresented as exact measurements. In this scenario, results from such non-BME techniques and BME could be found of comparable prediction accuracy, especially so when all of the following conditions are met concurrently:

1. The BME general knowledge G input is only the attribute mean trend and covariance.
2. The soft data distribution type is symmetric with respect to its mean.
3. The non-BME technique uses the soft data as hard data by selecting the mean of the soft distribution as the exact value.

Per the earlier remarks in Sect. 2.3, point (1) indicates a condition where BME prediction reduces to kriging prediction. Conditions (1)–(3) might actually favor non-BME prediction over BME in this scenario, because prediction accuracy is expected to be similar and BME is further reasonably expected to exhibit higher prediction uncertainty by carrying over the additional uncertainty of soft data to the predicted values. Another, more practical reason one might favor non-BME prediction in the same scenario is performance. Namely, using soft data typically requires additional resources than corresponding computations with hard data, and this might cause measurable computational overhead if one should choose to apply BME prediction. Regardless, from a scientific standpoint BME offers overall improved prediction by correctly accounting for known existing uncertainty that is otherwise incorrectly disregarded when soft data are hardened.

Typically, soft data are derived as inaccurate observations of a natural attribute. However, in certain cases, such observations might be also unavailable. The following subsections examine common situations and patterns where soft data can be appropriately constructed from other available information sources in the context of an analysis.

3.2.1 Simple statistical methods

Some natural processes are characterized by recurring similarities or periodicities. This suggests the corresponding attribute might obey certain statistical rules at any given time. It is also common to find missing values in long-term data series and large data scale analyses. This is a scenario where missing values can be filled in with manufactured soft data by using existing observations.

One way to address this problem is to perform a statistical analysis on existing data. For instance, Bayet et al. (2013) used the non-missing data sample mean and variance to construct probabilistic soft data (as in Eq. 7.2) that followed a Gaussian distribution. A different suggestion is to aggregate the study time scale, if possible, by using daily data for a monthly-based analysis, or jumping from a monthly to an annual scale, etc. For instance, when faced with incomplete daily data, Lee et al. (2008) constructed probabilistic soft data of monthly average minimum temperatures by introducing finite population correction and by using the Student's *t* distribution.

3.2.2 Expert or experienced judgment

A different path for the construction of soft data is by sourcing expert advice, empirical handling of similar situations, intuition, and professional knowledge (Christakos and Serre 2000).

In one instance, monitoring stations of attributes such as concentrations of atmospheric particulate matter (PM) $PM_{2.5}$, heavy metals in soil, or precipitation, can provide data where values might fall below detection limits. In such cases, expert feedback can help construct soft data of interval type (as in Eq. 7.1) that range from zero to the detection limit or to a threshold determined by expert judgment (Serre et al. 2003b; Puangthongthub et al. 2007).

In a separate instance, it might be difficult to establish a network of monitoring stations for a natural attribute due to considerations like terrain, geology and economic factors. Such factors may pose a problem in acquiring attribute data suitable for an S/ST analysis across an area of interest. In this case, the analogy analysis can be invoked to describe the natural process throughout this area. For example, Yu and Wang (2013) presented a case where, over the period of one month, daily probabilistic soft data values (as in Eq. 7.2) were generated at non-monitoring locations on the basis of the data probability distribution of another existing monitoring station that featured similar natural characteristics for that month.

Additional examples of expert-based soft data construction can be found in Christakos and Kolovos (1999), Choi et al. (2008), and Hampton et al. (2011).

3.2.3 Statistical model estimates

Natural processes commonly exhibit high levels of sophistication with relationships of causality and interplay between participating attributes. For this reason, predictive models attempt to achieve accuracy by involving a variety of statistical approaches like correlation models, multiple regression models, artificial neuron networks, autoregressive moving average models, etc. Eventually, such models produce estimates that alongside with their uncertainty can be used as soft data. Constructing soft data in this manner can be regarded as a combination of advanced statistics and expert judgment. To this end, BME provides a unique framework to interface and indirectly integrate a plethora of mathematical, statistical models and auxiliary information.

For instance, Money et al. (2011) devised simple regression models to find the relationship between fish tissue mercury and water column mercury. The regression models provided estimates and estimation variance that were jointly used to construct probabilistic Gaussian soft data (as in Eq. 7.2). In a different example, Hu et al. (2016) regarded soil temperature as auxiliary data to construct interval soft data (as in Eq. 7.1) of soil respiration via the Arrhenius formula by fitting results at the 95% confidence level of the Student's *t* distribution. In the field of remote sensing, Li et al. (2013a) used the inversion model estimates with their uncertainties as a soft data generator. In a more elaborate example, Reyes and Serre (2014) built a land use regression model to estimate the concentration of $PM_{2.5}$. At locations with incomplete soft data, $PM_{2.5}$ uncertainty was described by a Gaussian probability distribution with mean and revised variance from historical data. This approach enabled construction of probabilistic Gaussian soft data (as in Eq. 7.2) by using the above revised variance with a revised mean calculated by the difference of the mean minus the land use model estimate without residual.

3.3 Advantages of applying BME in geostatistics

Geostatistical S/ST analysis often entails some fundamental issues that researchers must address, such as the following:

1. The information to work with may be multi-sourced and at different scales of S/ST resolution, integration, precision, or a combination of the above such as having observation scales changed by dataset sources. Sources may include historical data, current observed data detected by semi-quantitative or quantitative analysis, graphs, manuscripts, and expert knowledge with various types and styles. Assortments of such information may be less useful if considered independently, yet by merging them into a single entity could possibly fill gaps in a complementary manner and render the parts useful as a whole (Perez-Hoyos et al. 2012). To do so, however, one must answer the scientific question "how can one integrate multi-sourced data and information and make full use of them in a methodical manner".
2. The observed data set might be relatively limited with regards to the research area to allow for attribute characterization and prediction. Lack of adequate input information can lead to reduced accuracy description and inference about an attribute. The question posed by

this point is "how to best deal with relatively small or insufficient data sets to provide an as accurate as possible interpretation of a natural process comprehensively and accurately".

3. Purely data-driven patterns are insufficient to the needs of modern science (Christakos 2000). Instances where this occurs include samplings under extreme conditions where observed data might be rare occurrences and derived from small probability events, thus biasing an analysis. In addition, observed data are usually considered as hard values where ignoring their uncertainty can flaw an analysis. Such cases raise the important question "how to introduce and account for data uncertainties".
4. In contrast to the previous point, samplings might consistently provide observations that occur frequently at given locations or periods, and should be rather ignored to avoid repetitive patterns or misleadingly smooth behavior. In this case, the question is "whether it is possible to avoid artificial smoothing of a particular process in S/ST analysis".
5. In general, a natural process might obey one or more physical laws, principles and mechanisms that constitute general knowledge about the process. Hence, measured data can describe the evolution of a process from an observer's standpoint, but accounting for general knowledge can critically provide us with explanatory insight about the process itself. The question raised at this point is "how to best utilize general knowledge".

Classical geostatistical methods appear to face limitations in addressing in practice the above questions. However, Sect. 2 illustrated that BME can operate steadily across these issues and provide satisfactory answers through its framework. The powerful BME characteristics enable it to consider versatile types of general knowledge for its prior PDF, to merge multi-sourced data types and uncertainties as conditional probabilities, and to be free of restraints and assumptions that commonly limit the capabilities of mainstream geostatistical methods (such as data distributional assumptions, stationarity assumptions, etc.)

3.4 BME in software

Presently, BME is implemented computationally through a variety of software tools. All of these tools have their roots in the BME library BMElib, a Matlab-based compilation of functions to carry out S/ST analysis and mapping (Christakos et al. 2002). The Spatiotemporal Epistemic Knowledge Synthesis Graphical User Interface (SEKS-GUI[4]) is the first BMElib interface that emerged. SEKS-GUI is also built on the Matlab platform (Yu et al. 2007), and provides an application for interactive BME analysis, complete with visualization features. A second application that followed was the Python-based BME-GUI.[5] More recently, the Space–Time Analysis Rendering with BME (STAR-BME[6]) software was released as a Python-based plugin for the QuantumGIS open source software. STAR-BME has the advantage of rendering visualizations within a Geographic Information System (GIS) environment, and that makes it very attractive for mapping applications (Yu et al. 2012, 2016).

All the above BME tools are free software, and enable at minimum an S/ST analysis on the plane (2-D in space) plus time. Each tool features a guided analysis through a sequence of screens across the different analysis tasks that start at importing data, and extend all the way to visualization or results and exporting the analysis output.

4 Application of BME in science of the total environment

BME is a predictive technique that has been applied in many different scientific fields where the investigated processes feature S/ST characteristics. The present review focuses on applications related to studying the change or characteristics of the total environment, which is the broader field wherein numerous BME analyses have been performed to-date. This section summarizes and classifies BME-based investigations of natural phenomena exploration, ecological and environmental conversion, resources development and utilization, social and economic management, pathophoresis control, etc., that particularly relate to the Earth's atmosphere, lithosphere, hydrosphere, and ecosphere (Fig. 2).

4.1 Atmosphere

Ambient air is a major source through which humans ingest pollutants. Exposure of humans to air pollution can cause human disease directly; hence we are very interested in the study of ambient air from a human exposure perspective (He et al. 2016; Jerrett et al. 2016; Li et al. 2016). Air quality can vary locally due to spatial and space–time variation of emission sources, meteorological factors, society factors, seasons, etc. (Dionisio et al. 2010; Akita et al. 2012; Ma et al. 2014). Since the very first applications

[4] As of May 2017, available at http://seksgui.org.

[5] As of May 2017, available at http://www.unc.edu/depts/case/BMEGUI.

[6] As of May 2017, available at http://140.112.63.249/STARBME/index.html.

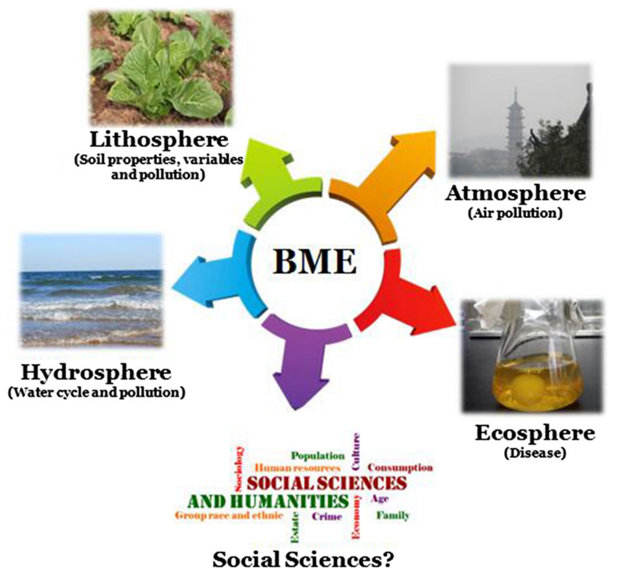

Fig. 2 Application of BME in science of the total environment

of BME in the study of spatiotemporal health impacts of ozone exposure (Christakos 1990; Christakos and Kolovos 1999), BME has developed itself into a robust and flexible approach that enables insightful in atmospheric applications, as shown in the following.

Initial publications were rather focused on introducing the BME theory, and case studies were presented to assist readers in comprehending the new framework and technique. With time, attention was drawn to the fact that model-driven analyses could provide keys to understand underlying principles. The ability to yield new data points by constructing soft data through model estimates evolved into a major BME feature. Expert judgment and fitting with trigonometric functions were use to build data and estimate particular matter distribution in North Carolina and California studies, respectively. Covariance composition was explained by short- and long-range interactions in space and time. The use of soft data led to more accuracy in BME performance than using only hard values (Christakos and Serre 2000; Christakos et al. 2001). Further, Puangthongthub et al. (2007) went on to point out that the uncertainties assessed by BME could be used to optimize the particulate matter monitoring network.

Prior knowledge, as a key BME point, has been brought to the forefront as means for a deeper understanding of data characteristics. These abilities were proven in related studies to broaden the application scale of BME. Specifically, Bogaert et al. (2009) consider the natural occurring phenomenon of the seasonal change of ozone as general knowledge to model the spatiotemporal distribution of ozone in California. De Nazelle et al. (2010) used soft data to cover an entire study area of high 4 × 4 km grid resolution in an application for the Environmental Protection Agency's (EPA) Community Air Quality modeling system. Nol et al. (2010) provided a fine example of BME's ability to assess uncertainties in a study of categorical spatial variables with a nitrogen oxide emission model. Yu and Wang (2010) explored in depth how BME can process uncertain, multi-sourced data and perform retrospective prediction in a $PM_{2.5}$ study in Taipei, for which the input was total suspended particulate matter and PM_{10} information.

The gradual addition of auxiliary variables and models enabled the BME framework to further improve prediction accuracy and analysis detail, while also improving methodological tools available for cognitive sciences. As an example in this category, land-use regression (LUR) was creatively employed to explore the association between environmental impact factors and ambient pollutant. For instance, Yu et al. (2011) used LUR to explain the relationship between $PM_{2.5}$ concentration and nine major classes of land usage. The model was the source of soft data observations that were consequently used for BME prediction. The LUR-BME model concept was further improved by having LUR estimate $PM_{2.5}$ values at selected location and then by BME predicting LUR residuals. Following this approach, Beckerman et al. (2013) selected land use auxiliary variables for use with the LUR-BME model that included land use types, traffic counts for roadways, and $PM_{2.5}$ estimates derived from remotely sensed data. For a similar analysis, Reyes and Serre (2014) used a different set of auxiliary variables that included SO_2, NH_3, $PM_{2.5}$, NO_X as stationary emission and total traffic, average congestion, and emission efficiency as road mobile sources of ambient pollutant. In the same work, the authors built further on this analysis to predict $PM_{2.5}$-related disease mortality as a topic of increased public interest. In a related analysis, Adam-Poupart et al. (2014) combined meteorological factors, location and land use to construct a generalized LUR model to predict ozone levels in Quebec. Finally, Akita et al. (2014) built six models to predict and compare average annual nitrogen oxide concentration in Catalunya, Spain. This study featured a chemical transport model (CTM), LUR, kriging, LUR-based BME, simple kriging with CTM, BME with CTM and LUR; the findings showed that the combination model of BME with CTM and LUR performed optimally, in a lowest root mean square error (RMSE) sense, among the six participants.

4.2 Lithosphere

The Earth's lithosphere comprises mainly of rock and soil. Both components are solids, and weathering can make rock transform into soil in a slow process. Also, both components have similar characteristics; namely they both have spatial continuity and are persistent in space and time. In addition to rock and soil, mineral resources are formed underground during the evolution of long geologic processes. In general, the lithospheric environment is stable and undergoes very little change, except for interference due to anthropogenic activity or geological upheavals such as crustal movement. From a spatiotemporal properties context, natural processes in lithosphere can be excellent candidates to study by means of S/ST geostatistical methods.

Soil studies about soil properties, composition and land utilization are important for countries with large populations, because soil is a key element in food production and as a receptor of residuals from human activities (Piao et al. 2010). Nonetheless, soil sampling and analysis is costly, and this makes S/ST geostatistics a valuable tool for analysis with auxiliary information. The growing concern about human health makes soil pollution a topic of increased interest, particularly with regard to agriculture soils.

BME has been broadly applied in the study of soil attributes, and in cases where auxiliary data are used to associate attributes of interest. In one example, Modis et al. (2013) used BME to describe the distribution of heavy metals in a lignite mining region in Greece. In another example, the focus was on increased soil salinity which is a negative factor in food production. Specifically, Douaik et al. (2005) analyzed spatial characteristics of electrical conductivity (EC), and found that soil salinity prediction with BME and soft interval data exhibited lower mean square error (MSE) than a corresponding kriging-based analysis, while validating the results with leave-one-out cross-validation (LOOCV).[7] Interestingly, even when considering the soft observations as hardened data by using the interval mid points, kriging led to partially unrealistic predicted EC values. In a different study, four methods were used to predict the spatial distribution of soil moisture, namely ordinary kriging (OK), regression kriging (RK), ordinary co-kriging (Co-OK) and BME by Gao et al. (2014). An extra dataset was employed alongside LOOCV and cross-validation to evaluate model performance, and prediction accuracy was quantified by computing the RMSE and correlation coefficient (CR). In the presence of auxiliary information from remote sensing, Co-OK, RK and BME produced more accurate prediction than OK, among which RK and BME performed best. BME was further found to slightly outperform RK when integrating probabilistic soft data. In Fan et al. (2015), four kinds of information were merged from a wireless sensor network. This study suggested that results tend to be more accurate upon introduction of additional auxiliary information. Finally, Orton and Lark (2007) used BME for soil depth prediction in a study that exhibited the BME value in accurate mapping of soil properties.

In soil studies, historical data are often available that contain nontrivial uncertainty. BME is a preferred tool in such cases for its ability to integrate soft information rigorously. For instance, Sun et al. (2015) updated digital soil maps with observations of uncertain organic matter concentration from historical data. By using the BME framework to account for the assessed means and variances in the observations, BME prediction led to comparatively increased accuracy and reduced prediction variance. In a similar study, Brus et al. (2008) used a traditional soil map of the Netherlands at scale 1:50,000 as soft information to predict soil categories at 8369 locations. 100 samples were selected by stratified simple random sampling for validation; their BME-based prediction led to a significant increase of map purity of soil classification by 15%. Soil texture maps can be sometimes imprecise tools for scientific research like modeling of soil processes and pedotransfer functions. D'Or and Bogaert (2003) used BME to predict the granular contents (sand, slit and clay) of the soil map for legend quantification; their results indicated that BME prediction exhibited lower average variance than methods with mapping unit mean. In the context of understanding the terrestrial carbon cycle and the climate change mechanism, Hu et al. (2016) focused on predicting the spatial distribution of soil respiration. To this end, they coupled BME with auxiliary information of soil temperature, selected 10% of collected data for cross validation and this led BME to outperform co-kriging and ordinary kriging in their analysis by exhibiting lower RMSE and larger CR. In the field of coal seam surface modeling, Li et al. (2013c) used BME to incorporate multi-sourced data that included information from boreholes, cross-sections and coal seam floor contour map. These data were obtained by geological surveys, and the study suggested BME could contribute to significant reduction of the number of soil sampling points.

4.3 Hydrosphere

Life is tied to water. Anthropogenic disturbance can influence the water cycle and resources in the short and long term, including effects on precipitation, runoff, etc. (Vorosmarty and Sahagian 2000; Huntington 2006). As a result, natural processes in the hydrosphere can exhibit elaborate variation in the spatial and temporal domains. BME has made considerable research contribution in this field, as shown in the following.

[7] In leave-one-out cross-validation, observations are removed from a dataset one at a time. The attribute is predicted at each vacated location from the remaining data, and the result is compared to the observed value. In general, it is recommended to keep validation data as a separate set from training observations, so that cross-validation produces a nearly unbiased estimate of the true error expected on an independent test set (e.g., Efron 1983; Varma and Simon 2006).

Drinking water safety is one of the research aspects in the domain of hydrosphere that enjoys sufficient attention by the public (Brahman et al. 2016, Schijven et al. 2016, Wang et al. 2016). In earlier research, groundwater quality assessment relied on the data uncertainty; see, e.g., Serre et al. (2003b) and Sanders et al. (2012). In more recent work, the use of auxiliary variables has led to improved prediction accuracy. For example, Messier et al. (2012) combined address geo-coding, land use regression (LUR) and BME to predict tetrachloroethylene concentration in a water supply quality study of groundwater in North Carolina. LUR was introduced as general knowledge to estimate a surface trend. The joint LUR-BME analysis led to higher accuracy than using simple kriging or BME, according to the LOOCV, MSE, and CR criteria. In follow-up studies, Messier et al. (2014, 2015) used LUR-BME to also investigate nitrates and radon concentrations in North Carolina groundwater. A major advantage of their approach was overcoming the limitation of relatively small data sizes by address geo-coding and soft data construction to augment understanding of contributing factors in the process and characteristics of the attribute spatiotemporal distribution. In addition to assessment of chemical composites in groundwater, BME has been used to analyze S/ST patterns of groundwater physical properties. In particular, Yu and Chu (2010) combined BME and empirical orthogonal functions to investigate the driving force behind spatiotemporal change of groundwater levels on the basis of piezometric heads, rainfall, pumping activities, and water exchange. Earlier on, Vyas et al. (2004) used BME for spatial interpolation of horizontal hydraulic conductivity.

An additional aspect of hydrosphere analyses focuses on health risk assessment due to transport of pollutants and contaminants through the food chain. For example, edible organisms like shellfish may accumulate pollutants, carry germs, or their growth might be impaired due to pollutants. To this end, water quality is often assessed by integrating monitoring data and modeling results (LoBuglio et al. 2007). Coulliette et al. (2009) constructed a linear regression model to predict fecal contamination through the levels of *Escherichia coli* (*E. coli*) from rainfall in an eastern North Carolina estuary for unmonitored space–time points. BME was used to integrate a mean trend model to study fecal contamination characteristics in different seasons to assist with water quality management decision-making. Money et al. (2009) investigated the fecal contamination using the *E. coli* indicator, too, along the Rartian River in New Jersey. A turbidity predictive model of *E. coli* was used to construct soft data for BME, and this secondary information reduced their analysis MSE by about 30% in LOOCV result comparisons.

Further in the hydrosphere domain, studying the S/ST distribution of precipitation is vital for environmental studies and risk assessment (Mahbub et al. 2010). In this specialization area, secondary variables are commonly introduced to the aid of precipitation spatio-temporal analysis. For example, by using elevation as a second attribute, Bayat et al. (2014) applied BME coupled with clustering techniques to improve prediction accuracy in assessing mean annual precipitation. In a more elaborate study, Zhang et al. (2016) used annual precipitation, average temperature, average water vapor pressure, elevation, and distance to coast to construct a multivariate BME framework for precipitation distribution mapping. The results showed that multivariate BME outperformed a univariate BME approach that only utilized the precipitation input.

Covering almost 71% of the Earth's surface area, oceans play a vital role in our planet's environmental cycle. Advances in remote sensing provide strong tools to assist studies of S/ST characteristics of participating attributes. In this field, BME has been used repeatedly in studies by integrating satellite data from multiple sources that carry different amounts of non-trivial uncertainties. See, for example, the analyses related to sea surface temperature by Li et al. (2013b) and Tang et al. (2015), and the ocean color products assessment in the chlorophyll-a study by Shi et al. (2015b).

4.4 Ecosphere

The ecosphere environment involves multiple and large categories of processes that are associated to and influenced by atmosphere, lithosphere and hydrosphere. As a general remark, the complexity of interactions in the ecosphere inevitably gives rise to uncertainties in the processes involved. The methodological advantages of BME in dealing with uncertainty render BME a strong contender for the analysis of S/ST characterized information in this subject. The following review highlights key BME contributions in this space.

Diseases constitute a prime topic of research. Whether infectious, acute, or their chronic variations, diseases are threats to human population, and more so are unexpected disease outbreaks that typically occur due to heterogeneities in the hosts and pathogens (Angulo et al. 2013). BME was found to increase significantly the accuracy in prediction of avian influenza by joint analysis with logistic regression modeling on the basis of auxiliary information such as migratory birds flyways, and environmental plaques like reservoirs, lakes, etc. (Cao et al. 2010). Choi et al. (2008) collected age-adjusted mortality data at county level to map BME influenza mortality rates, and to assess mortality risk in winter. Angulo et al. (2013) went deeper into investigating a disease mechanism analysis; first, a Susceptible-Infected-Recovered (SIR) model was adopted to quantify the exact number of people infected, people susceptible to get sick, and people recovering from the disease. The simulated results were then combined with BME to predict hand-foot-mouth population disease rates. This approach was favorably compared to using an extended Kalman filter. In fact, the SIR-BME model was consequently used again by Yu et al. (2014) to predict dengue fever cases by building on the model's strong ability to describe the disease spread process, and the disease spatio-temporal characteristics. In another study, Law et al. (2006) employed BME to describe qualitatively and quantitatively the propagation process of syphilis—before, during and after outbreaks—for the area of Baltimore between 1994 and 2002. The study findings suggested three core areas with high density of syphilis infection. A common challenge in the analysis of spatio-temporal disease population incidence is data uncertainty manifested in ways like small population effects. Yu et al. (2010) used BME to address this type of uncertainty by accounting for population size effects in a county-level study of oral cancer as an example of chronic disease. Assessment of the uncertainty in each county enabled its integration by BME, a decrease of the population size effects in the study, and subsequently higher precision identification of regions with higher risk of oral cancer incidence.

Some additional noteworthy topics in BME ecosphere analysis include the following examples: Yu et al. (2009) examined potential long-term impacts of climate/environmental quality change on humans. Specifically, two kinds of data processing methods were developed for this study; namely (1) data aggregation followed by BME prediction, and (2) BME prediction followed by data aggregation. These methods were applied in prediction tasks related to long-term $PM_{2.5}$ and ozone exposure. Money et al. (2011) used BME to predict the mercury concentration in fish tissue. Sedda et al. (2011) studied tree height in a young cork oak plantation with BME. Li et al. (2013a) used auxiliary information from coarse satellite data and exact land observation data for accurate BME prediction of leaf area index. Finally, Fox et al. (2015) applied BME to predict pedestrian mortality rates and reduce statistical noise.

5 Real-world case studies

Complementing the overview of examples presented in the previous section, the following subsections provide a closer look into selected novel studies that illustrate versatile applications of the BME method in real-world examples.

5.1 Mapping the black death epidemic in Europe

Christakos et al. (2005) undertook a systematic space–time mapping task of mortality and propagation of the infected areas in the great Black Death epidemic event across Europe during the years 1347–1351. Through painstaking collection of available information from a variety of surviving historical, civil, and ecclesiastical sources, a wealth of relevant informational content with various levels of uncertainty was gathered. This content was then encoded by generating soft data that contained all known information about the event. Thus, BME analysis made possible to create informative maps that provided an innovative insight to the characteristics of Black Death (temporal evolution, local and global geographical pattern, epidemic elasticity, propagation speeds, etc.). This case study made optimal use of the BME strengths while setting precedent for researchers in integrating a versatile multitude of interdisciplinary knowledge bases about an epidemic for disease mapping. The space–time nature of the Black Death epidemic spread was an ideal application ground for the BME stochastic spatiotemporal foundation, which facilitated establishing critical dependencies between different disease attribute characteristics in the space–time domain.

5.2 Merging satellite ocean color data

Remote sensing is a proven, potent approach to study natural attributes in larger scales. However, remote sensing relies on information provided by satellites located in space; as such, factors like the swath width, sun glint and clouds can influence the coverage of ground observations. To improve data coverage for further research, Shi et al. (2015b) employed BME to merge ocean color data (i.e. chlorophyll-a, Chl-a) collected from different satellites sensors such as MODIS, MERIS and SeaWiFS. Considering the variable measurement error across the sensors, probabilistic models among sensors were built to standardize and produce probabilistic soft data. In this framework, MODIS data were regarded to be the reference standard for the study. BME was employed to construct global chlorophyll-a maps from the data. The results showed a significantly higher, roughly sixfold improvement in spatial coverage compared to the individual measurements. In situ measurements of chlorophyll-a were used for prediction validation, and performance was evaluated by using the RMSE and correlation coefficient measures. The analysis showed that BME outperformed and exhibited higher accuracy than the corresponding results from the existing Garver-Siegel-Maritorena-01 method (GSM) and the weighted averaging method (AVW). This study highlighted the ability of BME to merge and handle uncertain data from multiple sources.

Additional studies have tested and found the method to be flexible and robust enough for similar tasks in remote sensing applications (e.g., Tang et al. 2015, 2016).

5.3 Retrospective prediction of spatiotemporal distribution of $PM_{2.5}$ concentration in Taipei

Long-term exposure to particulate matter, and especially PM with aerodynamic diameter <2.5 μm, has adverse effects on human health, including the cardiovascular and respiratory system. However, the research scope for epidemiology is limited by lack of long-term $PM_{2.5}$ measurements. In the study of $PM_{2.5}$ concentration in Taipei by Yu et al. (2010), PM_{10} and total suspended particles (TSP) concentrations had been locally monitored for a longer time span. Inspired by this information, the authors introduced ratios of $PM_{2.5}/PM_{10}$ and $PM_{2.5}$/TSP for retrospective prediction of $PM_{2.5}$ by assuming that the ratios remain invariant when $PM_{2.5}$ observations are absent in any given month. The predicted spatiotemporal distribution of $PM_{2.5}/PM_{10}$ ratios is considered yearly invariant. Thus, when PM_{10} or TSP information is available at any spatiotemporal location, distribution of $PM_{2.5}$ can be deducted through BME prediction. The study data refer to monthly values that were generated as probabilistic soft data from available daily observations. Overall, this work demonstrates the BME value in deriving and processing uncertain data from secondary sources, and in various temporal scales, to predict spatiotemporal distribution maps of natural attributes.

5.4 Updating digital soil maps with new data

Soil attributes can change continuously through materials and energy exchange in nature. To this end, scientists carry out regular investigations to understand the current state of such attributes. Existing work and observations can be very informative in this process; therefore it is often desirable to integrate past results in ongoing research. Government-run soil investigations could assist this research further, yet they take place at rather sparse time intervals every few years due to the typically slow rate of changes in the soil and the relatively high cost of surveys. In response to the previous, Sun et al. (2015) developed a BME-based method to update digital soil organic matter maps by using historical records from different years. In more detail, BME was employed to map the distribution of soil organic matter. Historical BME mapping output with the associated prediction error was then considered as soft data input for current mapping by using new data. Compared to ordinary kriging, BME exhibited less biased results in updating maps; this fact suggested that incorporating mapping uncertainty as soft data for BME prediction possibly contributes to improvement of the mapping accuracy. The case study provides an illustrative application of the Bayesian formula where prior knowledge is updated with new knowledge. This functionality was also utilized successfully by Yu et al. (2014) to update parameters in a Susceptible-Infected-Recovered (SIR) model in the context of their study of dengue fever incidence prediction.

6 Conclusions and prospects

Bayesian maximum entropy is regarded as a modern spatiotemporal geostatistics method; it is a powerful tool built within a rigorous theoretical framework that is used to represent, predict and map natural attributes at unsampled locations under conditions of in situ uncertainty. BME displays some compelling characteristics on both theoretical and practical grounds, which make it stand out among mainstream geostatistical methods. On the theoretical front, BME makes no restrictive assumptions when it comes to properties like normality, linearity, content-independency, and is capable of ingesting information beyond the commonly used low-order moments and hard data. On the practical front, BME is built with the ability to integrate a variety of knowledge bases, including general knowledge bases such as physical laws, mathematical models, expert knowledge, etc., and site-specific knowledge bases such as hard data, soft data, and secondary information.

Since its inception, the BME technique has enjoyed broad application in multiple disciplines, and has proven to be a most potent successor to classical geostatistics. In this ongoing course, efforts in the BME realm have brought to surface some noteworthy points of interest for forthcoming consideration by researchers, as shown in the following.

First, software implementation is important for the successful adoption and utilization of any computational technique, both in terms of a strategy and availability. In this domain, existing BME software tools can be further improved. Platform-wise, tools like SEKS-GUI are strictly tied to the data-processing ability of the MATLAB software foundation. To this end, progress can be made in the computational efficiency of calculations by taking advantage of the computing abilities offered by other contemporary languages, platforms and technologies. BME development can benefit greatly from utilizing scaled computing, adopting faster processing features, and tentatively exploring platforms with less processing overhead to (1) improve performance and availability of existing offerings, and (2) facilitate preparation of the next generation of BME software tools. The latter goal, in particular, connects technology aspects to future BME research, and is examined in more detail in the following point.

Second, literature indicates there is plenty of research latitude in tapping deeper into the rich theoretical BME foundation that allows for currently underused features such as multivariate prediction, and taking full advantage of physical laws as BME general knowledge. The rather limited scope of work performed in these BME areas thus far is partly justified by the generally more elaborate programming and computing tasks involved in them. Earlier implementations and advances in applications of the BME theory relied mostly on computational platforms prior to the establishment of more advanced scaling options available in contemporary technologies. As a result, most existing MATLAB and Python-based BME implementations focus on angles that have been previously easier to implement computationally, such as accounting for first and second-order statistical moments as general knowledge, in addition to hard and soft datasets. At the same time, these approaches left entirely on the side additional predictive potential that can be materialized by a systematic inclusion of valuable, applicable general knowledge bases like physical laws. In light of the ongoing developments in computing tools and technologies, all this material constitutes very fertile ground for numerical BME development and can become a fruitful sub-area of future BME research.

Third, the phenomenon of “buphthalmos” has been identified in several articles that results from low soft data quality, as defined earlier in the sense of low informative content. Hence, finding and selecting higher quality soft data is a key performance factor in the BME framework. Often, soft data are generated from raw sources in the course of research with the methods summarized in Sect. 3.2. In such cases, it might be possible to define soft data within degrees of flexibility; for example, in interval type soft data one might have the option to assign any width to an interval. It is the authors’ experience that unless dictated otherwise by the process characteristics, aiming to choose smaller interval widths can result in reduced datum uncertainty, and can thus lead to more accurate prediction with less prediction uncertainty. In a different example, improving the quality of soft data that are derived as estimates of a statistical model translates directly into seeking how to rather improve the model performance itself first, if possible. In general, BME accuracy improvement comes as a result of enabling integration of uncertain information that carries higher informative value. To this end, it can be constructive that BME users are informed and guided about generating soft data and choosing suitable soft data types in ways that BME analyses can be improved. In that light, nurturing closer communication between the BME research and user communities is a significant step towards maximizing the BME method benefits for the users.

An additional point of interest for further BME development is extrapolation, which includes focus areas such as forecasting and decision-making. To-date, there has been limited research on this capability; prime instances involve spatial mapping of water consumption (Lee and Wentz 2008), and spatiotemporal extrapolation for water demand (Lee et al. 2010). This is undoubtedly a direction of great interest, and a complementary one to the majority of hereto interpolation applications BME has been used for. In this context, it would be further useful to study how to choose auxiliary information effectively to improve BME accuracy and efficiency for both interpolation and extrapolation tasks.

Moreover, BME research can take direction hints from the ever-increasing availability of information sources and data, where BME prediction can play a significant role in harnessing S/ST big data. In a developing universe where progressively larger volumes of data are available, the aspects of data accuracy and informative content are as relevant as ever. In that sense, it would be prudent to counter-balance the excitement over the availability of larger data volumes with a reasonable level of caution in their usage. In this big data framework, the BME foundation enables addressing effectively the informative content challenge and sifting through uncertainty in the data sources; BME is therefore already present as a potent predictive methodology for the S/ST big data processing back-end.

Finally, this review pointed out primarily BME applications related to the total environment, while there still exist a host of other scientific and research fields that can benefit significantly from the BME methodology. Such fields include ocean sciences, ecology, economics, social sciences, crime and fraud analysis, etc. To this end, literature has been recently enriched with articles where BME manifested its capabilities for the first time in domains like renewable energy (Zagouras et al. 2015) and real estate research (Hayunga and Kolovos 2016). Regardless the domain of application, our work portrays BME as a highly sophisticated, inclusive, and expandable theoretical approach that is well capable, proven, and strongly positioned for the advanced analysis and predictive tasks with S/ST-characterized information.

Acknowledgements The work was supported by the National Natural Science Foundation of China (No. 529105-N11701ZJ). Alexander Kolovos received no funding.

References

Adam-Poupart A, Brand A, Fournier M, Jerrett M, Smargiassi A (2014) Spatiotemporal modeling of Ozone levels in Quebec

(Canada): a comparison of kriging, land-use regression (LUR), and combined bayesian maximum entropy-LUR approaches. Environ Health Perspect 122:970–976
Akita Y, Chen JC, Serre ML (2012) The moving-window Bayesian maximum entropy framework: estimation of PM2.5 yearly average concentration across the contiguous United States. J Eposure Sci Environ Epidemiol 22:496–501
Akita Y, Baldasano JM, Beelen R, Cirach M, de Hoogh K, Hoek G, Nieuwenhuijsen M, Serre ML, de Nazelle A (2014) Large scale air pollution estimation method combining land use regression and chemical transport modeling in a geostatistical framework. Environ Sci Technol 48:4452–4459
Angulo J, Yu H-L, Langousis A, Kolovos A, Wang J, Madrid AE, Christakos G (2013) Spatiotemporal infectious disease modeling: a BME-SIR approach. PLoS ONE 8:e72168
Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. CRC Press Inc, Boca Raton, FL
Bayat B, Zahraie B, Taghavi F, Nasseri M (2013) Evaluation of spatial and spatiotemporal estimation methods in simulation of precipitation variability patterns. Theoret Appl Climatol 113:429–444
Bayat B, Nasseri M, Naser G (2014) Improving Bayesian maximum entropy and ordinary Kriging methods for estimating precipitations in a large watershed: a new cluster-based approach. Can J Earth Sci 51:43–55
Beckerman BS, Jerrett M, Serre M, Martin RV, Lee SJ, van Donkelaar A, Ross Z, Su J, Burnett RT (2013) A hybrid approach to estimating national scale spatiotemporal variability of PM2.5 in the contiguous United States. Environ Sci Technol 47:7233–7241
Bogaert P, Fasbender D (2007) Bayesian data fusion in a spatial prediction context: a general formulation. Stoch Environ Res Risk Assess 21(6):695–709
Bogaert P, Christakos G, Jerrett M, Yu H-L (2009) Spatiotemporal modelling of ozone distribution in the State of California. Atmos Environ 43:2471–2480
Brahman KD, Kazi TG, Afridi HI, Baig JA, Arain SS, Talpur FN, Kazi AG, Ali J, Panhwar AH, Arain MB (2016) Exposure of children to arsenic in drinking water in the Tharparkar region of Sindh, Pakistan. Sci Total Environ 544:653–660
Brus D, Bogaert P, Heuvelink G (2008) Bayesian maximum entropy prediction of soil categories using a traditional soil map as soft information. Eur J Soil Sci 59:166–177
Cao C, Xu M, Chang C, Xue Y, Zhong S, Fang L, Cao W, Zhang H, Gao M, He Q, Zhao J, Chen W, Zheng S, Li X (2010) Risk analysis for the highly pathogenic avian influenza in Mainland China using meta-modeling. Chin Sci Bull 55:4168–4178
Choi KM, Yu HL, Wilson ML (2008) Spatiotemporal statistical analysis of influenza mortality risk in the State of California during the period 1997–2001. Stoch Env Res Risk Assess 22:S15–S25
Choy SL, O'Leary R, Mengersen K (2009) Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. Ecology 90:265–277
Christakos G (1990) A Bayesian/maximum-entropy view to the spatial estimation problem. Math Geol 22:763–777
Christakos G (1991) On certain classes of spatiotemporal random-fields with applications to space-time data-processing. IEEE Trans Syst Man Cybern 21:861–875
Christakos G (2000) Modern spatiotemporal geostatistics. Oxford University Press, New York
Christakos G (2002) On the assimilation of uncertain physical knowledge bases: bayesian and non-Bayesian techniques. Adv Water Resour 25:1257–1274
Christakos G (2010) Integrative problem-solving in a time of decadence. Springer, New York
Christakos G, Hristopulos DT (1998) Spatiotemporal environmental health modelling. Kluwer Acadamic Publication, Boston
Christakos G, Kolovos A (1999) A study of the spatiotemporal health impacts of ozone exposure. J Expo Anal Environ Epidemiol 9:322–335
Christakos G, Li X (1998) Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators? Math Geol 30:435–462
Christakos G, Serre ML (2000) BME analysis of spatiotemporal particulate matter distributions in North Carolina. Atmos Environ 34:3393–3406
Christakos G, Serre ML, Kovitz JL (2001) BME representation of particulate matter distributions in the state of California on the basis of uncertain measurements. J Geophys Res Atmos 106:9717–9731
Christakos G, Bogaert P, Serre M (2002) Temporal GIS: advanced functions for field-based applications. Springer, Berlin
Christakos G, Kolovos A, Serre ML, Vukovich F (2004) Total ozone mapping by integrating databases from remote sensing instruments and empirical models. IEEE Trans Geosci Remote Sens 42:991–1008
Christakos G, Olea RA, Serre ML, Wang LL, Yu HL (2005) Interdisciplinary public health reasoning and epidemic modelling: the case of black death. Springer, Berlin
Coulliette AD, Money ES, Serre ML, Noble RT (2009) Space/time analysis of fecal pollution and rainfall in an Eastern North Carolina Estuary. Environ Sci Technol 43:3728–3735
David M (1977) Geostatistical ore reverse estimation. Elsevier, Amsterdam
De Nazelle A, Arunachalam S, Serre ML (2010) Bayesian maximum entropy integration of ozone observations and model predictions: an application for attainment demonstration in North Carolina. Environ Sci Technol 44:5707–5713
Dionisio KL, Arku RE, Hughes AF, Vallarino J, Carmichael H, Spengler JD, Agyei-Mensah S, Ezzati M (2010) Air pollution in accra neighborhoods: spatial, socioeconomic, and temporal patterns. Environ Sci Technol 44:2270–2276
D'Or D, Bogaert P (2003) Continuous-valued map reconstruction with the Bayesian maximum entropy. Geoderma 112:169–178
D'Or D, Bogaert P, Christakos G (2001) Application of the BME approach to soil texture mapping. Stoch Environ Res Risk Assess 15:87–100
Douaik A, Van Meirvenne M, Toth T (2005) Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data. Geoderma 128:234–248
Efron B (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. J Am Stat Assoc 78(382):316–331
Ellison AM (2004) Bayesian inference in ecology. Ecol Lett 7:509–520
Ewing GM (1969) Calculus of variations with applications. W. W. Norton, New York
Fan L, Xiao Q, Wen JG, Liu Q, Jin R, You DQ, Li XW (2015) Mapping high-resolution soil moisture over heterogeneous cropland using multi-resource remote sensing and ground observations. Remote Sens 7:13273–13297
Fox L, Serre ML, Lippmann SJ, Rodriguez DA, Bangdiwala SI et al (2015) Spatiotemporal approaches to analyzing pedestrian fatalities: the case of Cali, Colombia. Traffic Injury Prevention 16:571–577
Gao S, Zhu Z, Liu S, Jin R, Yang G, Tan L (2014) Estimating the spatial distribution of soil moisture based on Bayesian maximum entropy method with auxiliary data from remote sensing. Int J Appl Earth Obs Geoinf 32:54–66
Hampton KH, Serre ML, Gesink DC, Pilcher CD, Miller WC (2011) Adjusting for sampling variability in sparse data: geostatistical approaches to disease mapping. Int J Health Geogr 10:1
Hayunga DK, Kolovos A (2016) Geostatistical space–time mapping of house prices using Bayesian maximum entropy. Intl J of Geog Inf Sci. doi:10.1080/13658816.2016.1165820

He T, Yang Z, Liu T, Shen Y, Fu X, Qian X, Zhang Y, Wang Y, Xu Z, Zhu S, Mao C, Xu G, Tang J (2016) Ambient air pollution and years of life lost in Ningbo, China. Sci Rep 6:22485

Heywood B, Brierley A, Gull S (2006) A quantified Bayesian maximum entropy estimate of Antarctic krill abundance across the Scotia Sea and in small-scale management units from the CCAMLR-2000 survey. Ccamlr Sci 13:97–116

Hristopulos DT, Christakos G (2001) Practical calculation of non-Gaussian multivariate moments in spatiotemporal Bayesian maximum entropy analysis. Math Geol 33(5):543–568

Hu JG, Zhou J, Zhou GM, Luo YQ, Xu XJ, Li PH, Liang JY (2016) Improving estimations of spatial distribution of soil respiration using the Bayesian maximum entropy algorithm and soil temperature as auxiliary data. PLOS ONE 11(1):e0146589

Huntington TG (2006) Evidence for intensification of the global water cycle: review and synthesis. J Hydrol 319:83–95

Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106:620–630

Jerrett M, Turner MC, Beckerman BS, Pope CA III, van Donkelaar A et al (2017) Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates. Environ Health Perspect 125(4):552–559

Journel A, Huijbregts C (1978) Mining geostatistics. Acadamic Press, London

Kolovos A, Christakos G, Serre ML, Miller CT (2002) Computational BME solution of a stochastic advection-reaction equation in the light of site-specific information. Water Resour Res 38(12):1318–1334

Kolovos A, Skupin A, Jerrett M, Christakos G (2010) Multi-perspective analysis and spatiotemporal mapping of air pollution monitoring data. Environ Sci Technol 44:6738–6744

Kolovos A, Angulo JM, Modis K, Papantonopoulos G, Wang J-F, Christakos G (2012) Model-driven development of covariances for spatiotemporal environmental health assessment. Environ Monit Assess. doi:10.1007/s10661-012-2593-1

Kolovos A, Smith LM, Schwab-McCoy A, Gengler S, Yu H-L (2016) Emerging patterns in multi-sourced data modeling uncertainty. Spat Stat Spec Issue Emerg Patterns 18A:300–317. doi:10.1016/j.spasta.2016.05.005

Krige D (1952) A statistical analysis of some of the borehole values in the orange free state goldfield. J Chem Metall Min Soc S Afr 53:47–70

Kuhnert PM, Martin TG, Griffiths SP (2010) A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecol Lett 13:900–914

Law DCG, Bernstein KT, Serre ML, Schumacher CM, Leone PA, Zenilman JM, Miller WC, Rompalo AM (2006) Modeling a syphilis outbreak through space and time using the Bayesian maximum entropy approach. Ann Epidemiol 16:797–804

Le ND, Zidek JV (2006) Statistical analysis of environmental space-time processes. Springer, Berlin

Lee SJ, Wentz EA (2008) Applying Bayesian maximum entropy to extrapolating local-scale water consumption in Maricopa County, Arizona. Water Resour Res 44:W01401. doi:10.1029/2007WR006101

Lee S-J, Balling R, Gober P (2008) Bayesian maximum entropy mapping and the soft data problem in urban climate research. Ann Assoc Am Geogr 98:309–322

Lee S-J, Wentz EA, Gober P (2010) Space–time forecasting using soft geostatistics: a case study in forecasting municipal water demand for Phoenix, Arizona. Stoch Environ Res Risk Assess 24:283–295

Lee S-J, Serre ML, van Donkelaar A, Martin RV, Burnett RT, Jerrett M (2012) Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient PM2.5 Concentrations across the Continental United States. Environ Health Perspect 120(12):1727–1732

Li AH, Bo YC, Chen L (2013a) Bayesian maximum entropy data fusion of field-observed leaf area index (LAI) and landsat enhanced thematic mapper plus-derived LAI. Int J Remote Sens 34:227–246

Li AH, Bo YC, Zhu YX, Guo P, Bi J, He YQ (2013b) Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method. Remote Sens Environ 135:52–63

Li X, Li P, Zhu H (2013c) Coal seam surface modeling and updating with multi-source data integration using Bayesian geostatistics. Eng Geol 164:208–221

Li L, Yang J, Song Y-F, Chen P-Y, Ou C-Q (2016) The burden of COPD mortality due to ambient air pollution in Guangzhou, China. Sci Rep 6:25900

LoBuglio JN, Characklis GW, Serre ML (2007) Cost-effective water quality assessment through the integration of monitoring data and modeling results. Water Resour Res 43(3):W03435. doi:10.1029/2006WR005020

Ma ZW, Hu XF, Huang L, Bi J, Liu Y (2014) Estimating ground-level PM2.5 in China Using satellite remote sensing. Environ Sci Technol 48:7436–7444

Mahbub P, Ayoko GA, Goonetilleke A, Egodawatta P, Kokot S (2010) Impacts of Traffic and rainfall characteristics on heavy metals build-up and wash-off from urban roads. Environ Sci Technol 44:8904–8910

Martin TG, Burgman MA, Fidler F, Kuhnert PM, Low-Choy S, Mcbride M, Mengersen K (2012) eliciting expert knowledge in conservation science. Conserv Biol 26:29–38

Matheron G (1962) Traité de géostatistique appliquée. Technip, Paris

Matheron G (1965) Les variables régionalisées et leur estimation. Masson, Paris

McCarthy MA, Masters P (2005) Profiting from prior information in Bayesian analyses of ecological data. J Appl Ecol 42:1012–1019

Messier KP, Akita Y, Serre ML (2012) Integrating address geocoding, land use regression, and spatiotemporal geostatistical estimation for groundwater tetrachloroethylene. Environ Sci Technol 46(5):2772–2780

Messier KP, Kane E, Bolich R, Serre ML (2014) Nitrate variability in groundwater of North Carolina using monitoring and private well data models. Environ Sci Technol 48:10804–10812

Messier KP, Campbell T, Bradley PJ, Serret ML (2015) Estimation of groundwater Radon in North Carolina using land use regression and Bayesian maximum entropy. Environ Sci Technol 49:9817–9825

Modis K, Vatalis KI, Sachanidis C (2013) Spatiotemporal risk assessment of soil pollution in a lignite mining region using a Bayesian maximum entropy (BME) approach. Int J Coal Geol 112:173–179

Money ES, Carter GP, Serre ML (2009) Modern space/time geostatistics using river distances: data integration of turbidity and *E. coli* measurements to assess fecal contamination along the Raritan river in New Jersey. Environ Sci Technol 43:3736–3742

Money ES, Sackett DK, Aday DD, Serre ML (2011) Using River distance and existing hydrography data can improve the geostatistical estimation of fish tissue mercury at unsampled locations. Environ Sci Technol 45:7746–7753

Nol L, Heuvelink GBM, Veldkamp A, de Vries W, Kros J (2010) Uncertainty propagation analysis of an N_2O emission model at the plot and landscape scale. Geoderma 159:9–23

Olea RA (1999) Geostatistics. Kluwer Acadamic Publication, Boston

Orton TG, Lark RM (2007) Estimating the local mean for Bayesian maximum entropy by generalized least squares and maximum likelihood, and an application to the spatial analysis of a censored soil variable. Eur J Soil Sci 58:60–73

Perez-Hoyos A, Garcia-Haro FJ, San-Miguel-Ayanz J (2012) A methodology to generate a synergetic land-cover map by fusion of different land-cover products. Int J Appl Earth Obs Geoinf 19:72–87

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259

Piao S, Ciais P, Huang Y, Shen Z, Peng S, Li J, Zhou L, Liu H, Ma Y, Ding Y, Friedlingstein P, Liu C, Tan K, Yu Y, Zhang T, Fang J (2010) The impacts of climate change on water resources and agriculture in China. Nature 467:43–51

Puangthongthub S, Wangwongwatana S, Kamens RM, Serre ML (2007) Modeling the space/time distribution of particulate matter in Thailand and optimizing its monitoring network. Atmos Environ 41:7788–7805

Reyes JM, Serre ML (2014) An LUR/BME framework to estimate PM2. 5 explained by on road mobile and stationary sources. Environ Sci Technol 48:1736–1744

Sanders AP, Messier KP, Shehee M, Rudo K, Serre ML, Fry RC (2012) Arsenic in North Carolina: public health implications. Environ Int 38:10–16

Savelyeva E, Utkin S, Kazakov S, Demyanov V (2010) Modeling spatial uncertainty for locally uncertain data. Geoenv VII Geostat Environ Appl 16:295–306

Schijven J, Forêt JM, Chardon J, Teunis P, Bouwknegt M, Tangena B (2016) Evaluation of exposure scenarios on intentional microbiological contamination in a drinking water distribution network. Water Res 96:148–154

Sedda L, Atkinson PM, Filigheddu MR, Cotzia G, Dettori S (2011) Spatio-temporal analysis of tree height in a young cork oak plantation. Int J Geogr Inf Sci 25:1083–1096

Serre ML, Christakos G (2002) BME-based hydrogeologic parameter estimation in groundwater flow modelling. Acta Univ Carol Geol 46:566–570

Serre ML, Christakos G, Li H, Miller CT (2003a) A BME solution of the inverse problem for saturated groundwater flow. Stoch Environ Res Risk Assess 17(6):354–369

Serre ML, Kolovos A, Christakos G, Modis K (2003b) An application of the holistochastic human exposure methodology to naturally occurring arsenic in Bangladesh drinking water. Risk Anal 23:515–528

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423

Shi T, Yang X, Christakos G, Wang J, Liu L (2015a) Spatiotemporal interpolation of rainfall by combining BME theory and satellite rainfall estimates. Atmosphere 6:1307–1326

Shi Y, Zhou X, Yang X, Shi L, Ma S (2015b) Merging Satellite Ocean Color Data With Bayesian Maximum Entropy Method. IEEE J Sel Top Appl Earth Obs Remote Sens 8:3294–3304

Shoemaker JS, Painter IS, Weir BS (1999) Bayesian statistics in genetics—a guide for the uninitiated. Trends Genet 15:354–358

Sun XL, Wu YJ, Lou YL, Wang HL, Zhang C, Zhao YG, Zhang GL (2015) Updating digital soil maps with new data: a case study of soil organic matter in Jiangsu, China. Eur J Soil Sci 66:1012–1022

Tabios GQ, Salas JD (1985) A comparative analysis of techniques for spatial interpolation of precipitation. JAWRA J Am Water Resour Assoc 21:365–380

Tang SL, Yang XF, Dong D, Li ZW (2015) Merging daily sea surface temperature data from multiple satellites using a Bayesian maximum entropy method. Front Earth Sci 9:722–731

Tang Q, Bo Y, Zhu Y (2016) Spatiotemporal fusion of multiple-satellite aerosol optical depth (AOD) products using Bayesian maximum entropy method. J Geophys Res Atmos 121(8):4034–4048

Urbani F, D'Alessandro P, Frasca R, Biondi M (2015) Maximum entropy modeling of geographic distributions of the flea beetle species endemic in Italy (Coleoptera: Chrysomelidae: Galerucinae: Alticini). Zoologischer Anzeiger 258:99–109

Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. Bioinformatics. doi:10.1186/1471-2105-7-91

Vorosmarty CJ, Sahagian D (2000) Anthropogenic disturbance of the terrestrial water cycle. Bioscience 50:753–765

Vyas VM, Tong SN, Uchrin C, Georgopoulos PG, Carter GR (2004) Geostatistical estimation of horizontal hydraulic conductivity for the Kirkwood–Cohansey aquifer. J Am Water Resour Assoc 40:187–195

Wang H, Wang N, Wang B, Zhao Q, Fang H, Fu C, Tang C, Jiang F, Zhou Y, Chen Y, Jiang Q (2016) Antibiotics in drinking water in shanghai and their contribution to antibiotic exposure of school children. Environ Sci Technol 50:2692–2699

Webster R, Oliver MA (2007) Geostatistics for environmental scientists. Wiley, New York

Yang Y, Christakos G (2015) Spatiotemporal characterization of ambient PM2. 5 concentrations in Shandong Province (China). Environ Sci Technol 49:13431–13438

Yu HL, Chu HJ (2010) Understanding space-time patterns of groundwater system by empirical orthogonal functions: a case study in the Choshui River alluvial fan, Taiwan. J Hydrol 381:239–247

Yu HL, Wang CH (2010) Retrospective prediction of intraurban spatiotemporal distribution of PM2.5 in Taipei. Atmos Environ 44:3053–3065

Yu HL, Wang CH (2013) Quantile-based bayesian maximum entropy approach for spatiotemporal modeling of ambient air quality levels. Environ Sci Technol 47:1416–1424

Yu HL, Kolovos A, Christakos G, Chen JC, Warmerdam S, Dev B (2007) Interactive spatiotemporal modelling of health systems: the SEKS-GUI framework. Stoch Environ Res Risk Assess 21:555–572

Yu HL, Chen JC, Christakos G, Jerrett M (2009) BME estimation of residential exposure to ambient PM10 and Ozone at multiple time scales. Environ Health Perspect 117:537–544

Yu HL, Chiang CT, Lin SD, Chang TK (2010) Spatiotemporal analysis and mapping of oral cancer risk in Changhua County (Taiwan): an application of generalized Bayesian maximum entropy method. Ann Epidemiol 20:99–107

Yu HL, Wang CH, Liu MC, Kuo YM (2011) Estimation of fine particulate matter in Taipei using landuse regression and Bayesian maximum entropy methods. Int J Environ Res Public Health 8:2153–2169

Yu H-L, Ku S-J, Kolovos A (2012) Advanced space-time predictive analysis with STAR-BME. In: Proceedings of the 20th international conference on advances in geographic information systems, pp. 593–596. ACM

Yu H-L, Angulo JM, Chen M-H, Wu J, Christakos G (2014) An online spatiotemporal prediction model for dengue fever epidemic in Kaohsiung (Taiwan). J Biom 56(3):428–440. doi:10.1002/bimj.201200270

Yu HL, Ku SC, Kolovos A (2016) A GIS tool for spatiotemporal modeling under a knowledge synthesis framework. Stoch Environ Res Risk Assess 30:665–679

Zagouras A, Kolovos A, Coimbra CFM (2015) Objective framework for optimal distribution of solar irradiance monitoring networks. Renew Energy 80:153–165. doi:10.1016/j.renene.2015.01.046

Zhang FS, Yang ZT, Zhong SB, Huang QY (2016) Exploring mean annual precipitation values (2003–2012) in a specific area (36 degrees N-43 degrees N, 113 degrees E-120 degrees E) using meteorological, elevational, and the nearest distance to coastline variables. Adv Meteorol 2016:2107908. doi:10.1155/2016/2107908

Zhao Y, Kockelman KM (2002) The propagation of uncertainty through travel demand models: an exploratory analysis. Ann Reg Sci 36(1):145–163