# A Bayesian approach to parameter estimation in the presence of spatial missing data

Domenica Panzera, [ab] Roberto Benedetti,[a] Paolo Postiglione [a]

[a] Department of Economic Studies, University 'G. d'Annunzio' of Chieti-Pescara, Viale della Pineta 4, 65127 Pescara, Italy.
[b] Corresponding author, E-mail: domenicapanzera@yahoo.it

**Abstract** The missing data problem has been widely addressed in the literature. The traditional methods for handling missing data may be not suited for spatial data, which can exhibit distinctive structures of dependence and/or heterogeneity. As a possible solution to the spatial missing data problem, this paper proposes an approach that combines the Bayesian Interpolation method (Benedetti and Palma, 1994) with a multiple imputation procedure. The method is developed in a univariate and a multivariate framework, and its performance is evaluated through an empirical illustration based on data related to labour productivity in European regions.

**JEL classification**
C10; C11; O40; R10

**Keyword**
Spatial data; Missing values; Bayesian Interpolation Method; Multiple imputation

## 1. Introduction

Over the last several decades, statisticians have been intrigued by the issue of developing a satisfactory methodology for spatial data analysis (Cliff and Ord, 1973, 1981; Anselin, 1988; LeSage and Pace, 2009). Many scientific fields, such as agriculture, biology, geology, geography, and, more recently, economics and sociology, are interested in collecting and analyzing spatial data.

Spatial data contain information about attributes of interest and their locations, and such data typically exhibit structures of dependence and/or heterogeneity. To account for the multidirectional dependence and heterogeneity among the nearest sites, a general class of well-established models has been introduced in the statistical literature. For earlier developments, see, for example, the seminal papers by Besag (1974) and Strauss (1977); for a review of this topic, see Cressie (1993). The introduction of Markov Random Fields (MRFs) in the spatial data analysis context (Besag, 1974) has encouraged an increasing interest in the analysis of spatial dependence and the development of methods to estimate models under the assumption of an MRF data generating process.

A random field is a collection of univariate or multivariate random variables, $Y$, that are indexed by their locations, $s$; that is, $\{Y(s), s \in S\}$. The set of indices $S$ identifies the topology of data and can refer to a continuous surface, a finite set of discrete locations (i.e., regular or irregular lattice data), or continuous random locations (i.e., point process data). The spatial dependence effect can be imposed by modelling the random field according to different specifications, such as the Simultaneous Autoregressive models (Whittle, 1954), the Spatial Lag, the Spatial Error, and the spatial Durbin models (LeSage and Pace, 2009), the Conditional Autoregressive model (CAR, Besag, 1974), or the Spatial Moving Average model (Haining, 1978), among others. These specifications all introduce the spatial dependence effect in the

covariance structure of the process as a function of both a contiguity matrix and unknown autocorrelation parameters. The contiguity matrix, which expresses the interactions between any pair of spatial units, can be defined according to the geographical arrangement of the observations as well as their economic or social distances (Conley and Topa, 2002). For greater details about random field models, see Cressie (1993) and Gaetan and Guyon (2010). An alternative to spatial econometric modelling that accounts for spatial dependence among georeferenced data is eigenvector spatial filtering (Chun and Griffith, 2013). This method employs selected eigenvectors that are extracted from a spatial weight matrix as control variables in a model specification. These control variables identify and isolate the stochastic spatial dependencies among the georeferenced observations, thereby allowing model estimation to proceed as though the observations were independent.

Despite its growing popularity, spatial modelling has not provided a systematic treatment to the missing data problem. Missing data commonly arise in the spatial domain. Some examples, including as historical, physical geographical, geological, and ecological data, are reported in Bennett et al. (1984).

In statistics, missing values occur when, for some units, no data are recorded for some or all of the variables in the current observation. The presence of missing data can produce a significant effect on the analysis and, thus, on the conclusions that can be drawn from the data. To determine how to handle missing data, it is helpful to know why they are missing. According to a well-known classification, we distinguish three general possible *missingness mechanisms* (Little and Rubin, 1987). Data are *missing completely at random* (MCAR) if the probability of missingness is the same for all units or if the probability of an observation being missing does not depend on observed or unobserved data. Assuming the MCAR mechanism in a survey, subjects failing to respond is a consequence of selection by simple random sampling. A more general assumption is represented by the *missing at random* (MAR) mechanism, where the

probability of a value being missing depends only on the available information. The MAR assumption is linked to a sampling design that is more general than the simple random sampling, i.e., stratified simple random sampling. The population is split into strata, and in each stratum, an MCAR process acts. For example, in a survey, young subjects could be more reluctant to respond to a question than the elderly, but the non-response process is MCAR in each group. If the age is known for all subjects, the non-response mechanism is MAR (Longford, 2008). Finally, data are *missing not at random* (MNAR) if neither MCAR nor MAR hold. In other words, data are MNAR if the probability of an observation being missing is systematically related to the missing values. For example, consider a survey of drug use where occasional drug users fail to respond (Longford, 2008).

In a non-spatial context, missing data are handled in a variety of ways. A common practice is represented by the complete case analysis, which ignores units with any missing data on the variables included in the analysis, compared with the available case method, where all available cases are used for each analysis (in this case, the sample size can change from variable to variable). Alternative solutions involve the application of single imputation techniques based on the replacement of missing data by single values, such as the average of the observed data (i.e., mean substitution), random draws from the distribution of the observed values (i.e., hot deck imputation), or values predicted by a regression equation (i.e., regression-based imputation). Furthermore, missing data can be replaced by a set of likely values, as occurs in the multiple imputation procedure (Rubin, 1987), such that the uncertainty associated with the imputation is incorporated in the analysis. A solving approach based on the maximum likelihood principle was introduced by Dempster et al. (1977), who proposed an iterative procedure known as the Expectation Maximization (EM) algorithm. A comprehensive review of the main missing data methods can be found in the works of Little and Rubin (1987) and Schafer (1997).

Traditional approaches to missing values have also been applied to spatial data (see, e.g., Bennett et al. 1984; Haining, 2003, pp. 154-177). In the early literature, spatial mean imputation techniques were introduced by Kennedy and Tobler (1983) and Tobler and Kennedy (1985). They proposed replacing spatial missing data by a weighted mean of the values observed at the first-order neighbours' locations with the weights proportional to the length of the common shared boundaries. An iterative procedure to adapt this method to spatially clustered missing data has also been developed. A regression-based imputation method for geographical data was proposed by Oakes (1973), and Willmott and Wicks (1980) developed an approach based on the use of principal component analysis. A general theory of maximum likelihood estimation with spatial missing data was developed by Martin (1984), who derived the exact maximum likelihood estimators for the parameters of a Gaussian underlying spatial process when data are MAR. This approach is discussed for multivariate normal data satisfying the first-order spatial Markov property by Haining et al. (1984), and some of its applications can be found in the works of Haining et al. (1989) and Griffith et al. (1989). Birkin and Clarke (1989) introduced a method for generating estimates of small area income distributions at the micro-level. This technique is based on the combined use of iterative proportional fitting to estimate missing data and micro-simulation. Griffith and Layne (1999) described an SAR EM-type estimation method for missing data. A Bayesian approach to analyzing regression models with missing data in the response variable was proposed by Oh et al. (2002). The method was developed for regression models with spatially correlated errors and involves the Gibbs sampling algorithm in both missing value predictions and parameter estimation. LeSage and Pace (2004) proposed a method based on the application of the EM algorithm to the estimation of missing values in the dependent variable of different spatial econometrics models. An estimator for models with spatial lag in the dependent and independent variables, non-parametrically specified error terms and missing observations was proposed by Kelejian and Prucha (2010). In the special issue of

Geographical Analysis edited by Qiu and Cromley (2013), areal interpolation methods are also described as an appropriate approach for the case of incomplete coverage of a map (i.e., a missing data problem). Finally, Chun and Griffith (2013) described some methods to calculate imputations for missing values and evaluate the correspondence uncertainty.

In this paper, the spatial missing data problem is addressed by using a Bayesian approach. Sets of likely values for missing data are generated from a multivariate normal posterior density, whose parameters are derived by the application of the Bayesian Interpolation Method (BIM, Benedetti and Palma, 1994). The method allows for obtaining multiple versions of the complete data set. Parametric models can be estimated on each imputed data set, and the results can be combined by using the multiple imputation combining rules (Rubin 1987; Little and Rubin, 1987).

The Gaussian distribution of the data generating process is the only assumption required by the proposed method. A multivariate normal prior distribution is then assumed for the underlying spatial process, which is modelled through a CAR specification. The assumed specification does not entail any loss of generality because any Gaussian distribution on a finite set of sites can be expressed as a conditional autoregressive process (Ripley, 1981, p. 90). Through the assumed prior distribution, the spatial dependence effect is considered in the missing data imputation. Following Bayes' rule, the prior information on the complete data generating process is then combined with the evidence derived from the observed data.

Unlike the Bayesian approach developed by Oh et al. (2002), the proposed method is designed for irregular lattice data and is thus more suitable for socio-economic data analysis. The approach proposed by Oh et al. (2002) allows parameter estimates and missing value imputation to be obtained simultaneously. In contrast, our method focuses on the missing data imputation and then proceeds to estimate parametric models on the multiple imputed data sets.

The introduction of a multiple imputation procedure, which reflects the uncertainty associated with the missing data imputation, represents an additional advantage of our method.

The remainder of the paper is organized as follows. Section 2 describes the proposed method, which is presented in a univariate and a multivariate framework. In section 3, an empirical application of the method is provided. Section 4 concludes the paper by suggesting some possible extensions.

## 2. Methodology

Let $\{Z_i, i = 1,...,n\}$ be a spatial stochastic process whose finite realizations can be observed on $n$ areal units $\{s_1,...,s_n\}$ that form a lattice on the spatial domain $S$. We assume that the random vector $\mathbf{Z} = (Z_1,...,Z_n)'$ is distributed according to a Gaussian Markov Random Field (i.e., CAR model; Besag, 1974), which is specified by a set of full conditional distributions as

$$Z_i \mid \mathbf{Z}_{-\mathbf{i}} \sim N(\mu_i + \sum_{j=1}^{n} k_{ij}(Z_j - \mu_j), \sigma_i^2) \text{ for } i = 1,...,n \qquad (1)$$

where $\mathbf{Z}_{-\mathbf{i}} = \{Z_j : j = 1,...,n, j \neq i\}$, $\mu_i = E[Z_i]$, $\sigma_i^2 = VAR[Z_i \mid \mathbf{Z}_{-\mathbf{i}}]$, and $k_{ij}$ are spatial interaction parameters. The spatial interaction parameters are defined as $k_{ij} = \rho w_{ij}$, for $i, j = 1,...,n$, where $\rho$ is a scalar spatial autocorrelation parameter and $w_{ij}$ is the $(ij)-$th entry of a $n \times n$ contiguity matrix $\mathbf{W}$, which summarizes the neighbourhood structure of the areal units.

By Brook's lemma (Brook, 1964) and the Hammersley-Clifford theorem (see, e.g., Besag, 1974), the full conditional distributions in (1) uniquely determine the following joint distribution

$$\mathbf{Z} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{2}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_n)'$ and $\boldsymbol{\Sigma}$ is the covariance matrix, expressed as

$$\boldsymbol{\Sigma} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}\mathbf{T} \tag{3}$$

with $\mathbf{I}_n$ denoting an identity matrix of order $n$ and $\mathbf{T} = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)$.

A necessary and sufficient condition such that (2) is a valid joint distribution is that the covariance matrix in (3) is a symmetric and positive definite. The covariance matrix in (3) is symmetric if the identity $w_{ij}\sigma_j^2 = w_{ji}\sigma_i^2$ holds for all $i, j$=1,…,$n$. This can be achieved by conveniently specifying the contiguity matrix**.** A common choice consists of defining a symmetric $n$ x $n$ contiguity matrix **C**, with elements

$$c_{ij} = \begin{cases} 1 \text{ if } s_j \in N(i), \\ 0 \text{ if } i=j, \qquad \text{for } i, j = 1,...,n \\ 0 \text{ otherwise} \end{cases}$$

where $N(i)$ denotes the set of neighbours of the areal unit $s_i$. When the number of neighbours varies for geographical units, as often occurs for irregular lattice data, the specification of a normalized matrix, in which each row sums to 1, is recommended (see Clayton and Berardinelli, 1992; Wall, 2004). The matrix **W** in (3) is then specified as $\mathbf{W}=\mathbf{D}^{-1}\mathbf{C}$, where

$\mathbf{D} = diag(c_{1+}, c_{2+}, ..., c_{n+})$, with $c_{i+} = \sum_j c_{ij}$ denoting the cardinality of $N(i)$. When $\mathbf{W}$ is used, the conditional variances $\sigma_i^2$, $i$=1,…,$n$, on the main diagonal of $\mathbf{T}$ have to be proportional to $1/c_{i+}$ for all $i$. Then, by setting $\sigma_i^2 = \sigma^2/c_{i+}$ for all $i$, the covariance matrix in (3) becomes

$$\mathbf{\Sigma} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}\mathbf{T} = \sigma^2(\mathbf{D} - \rho\mathbf{C})^{-1}. \qquad (4)$$

The covariance matrix in (4) characterizes the specification that for the zero centred CAR is denoted by $CAR(\rho, \sigma^2)$, according to Gelfand and Vounatsou (2003), and a sufficient condition for its positive definiteness is $|\rho| < 1$.

Now, we define the process $\{Z_i, i = 1, ..., n\}$ as the *original process*, which is assumed to be the complete data generating process (Benedetti and Palma, 1994).

Suppose that data on the variable $\mathbf{Z}$ are recorded on $m < n$ areal units. Starting from the $n$-dimensional random vector $\mathbf{Z}$, the $m$-dimensional random vector $\mathbf{Z}^* = (Z_1^*, ..., Z_m^*)'$ can be derived by introducing a transformation operator $\mathbf{G}$ such that (Palma and Benedetti, 1998)

$$\mathbf{Z}^* = \mathbf{GZ}.$$

We refer to the process $\{Z_j^*, j = 1, ..., m\}$ as the *derived process*, which is assumed to be the incomplete data generating process (i.e., with missing data). Denoted by $r(j)$, the set of areal units for which the variable of interest is observed, the transformation operator $\mathbf{G}$ is defined as an $m \times n$ matrix with elements

$$g_{ji} = \begin{cases} 1 \text{ if } s_i \in r(j) \\ 0 \text{ otherwise} \end{cases}$$

for $i = 1,...,n$ and $j = 1,...,m$.

In the described approach, the solution to the spatial missing data problem corresponds to the restoration of the realizations of the original process starting from the knowledge of the realizations of the derived one. Formally speaking, we are interested in studying the conditional probability distribution $P(\mathbf{Z} \mid \mathbf{Z}^*)$ after the knowledge, based on data, of $P(\mathbf{Z}^* \mid \mathbf{Z})$ is obtained. Because $\mathbf{Z}$ is distributed according to $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\mathbf{Z}^* = \mathbf{GZ}$, with $\mathbf{G}$ denoting a linear operator, by recalling a well-known result (Anderson, 1958, p. 26), the conditional distribution of $\mathbf{Z}^* \mid \mathbf{Z}$ can be derived as

$$\mathbf{Z}^* \mid \mathbf{Z} \sim MVN(\mathbf{G}\boldsymbol{\mu}, \mathbf{G}\boldsymbol{\Sigma}\mathbf{G}'). \qquad (5)$$

The result in (5) identifies the likelihood function of $\mathbf{Z}$. From (2), any prior information on $\mathbf{Z}$ is summarized by a multivariate normal prior distribution, and, according to Bayes' rule, we obtain

$$P(\mathbf{Z} \mid \mathbf{Z}^*) \propto P(\mathbf{Z}) P(\mathbf{Z}^* \mid \mathbf{Z}).$$

The posterior distribution of $\mathbf{Z} \mid \mathbf{Z}^*$ is still a multivariate normal distribution for the Normality of the distributions of $\mathbf{Z}$ and $\mathbf{Z}^* \mid \mathbf{Z}$. In particular, by assuming known the covariance matrix of the original process, it follows that (see Pilz, 1991)

$$\mathbf{Z} \mid \mathbf{Z}^* \sim MVN(\hat{\mathbf{Z}}, \mathbf{V}_{\hat{\mathbf{Z}}})$$

where $\hat{\mathbf{Z}}$ and $\mathbf{V}_{\hat{\mathbf{Z}}}$ are the BIM estimates, defined as

$$V_{\hat{Z}} = [G'(G\frac{(D-\rho C)}{\sigma^2}G')^{-1}G + \frac{(D-\rho C)}{\sigma^2}]^{-1} \quad (6)$$

$$\hat{Z} = V_{\hat{Z}}[\frac{(D-\rho C)}{\sigma^2}\mu + G'(G\frac{(D-\rho C)}{\sigma^2}G')^{-1}Z^*]. \quad (7)$$

In this Bayesian framework, $\hat{\mathbf{Z}}$ represents the maximum a posterior (MAP) estimate of $\mathbf{Z}$ and is the mode of its posterior distribution.

Additional issues are raised from the *pycnophylactic* (or mass preserving) property. This property consists of finding an estimate of $\mathbf{Z}$ such that, by applying the transformation operator $\mathbf{G}$, the observed data $\mathbf{Z}^*$ are again obtained (Tobler, 1979). To preserve the pycnophylactic property, the multivariate normal distribution defined by (6) and (7) is conditioned to the linear constraint $\mathbf{G}\hat{\mathbf{Z}} = \mathbf{Z}^*$, thus obtaining the constrained versions of the BIM estimates, expressed as (Palma and Benedetti, 1998)

$$\tilde{\mathbf{Z}} = \hat{\mathbf{Z}} + \mathbf{V}_{\hat{\mathbf{Z}}}\mathbf{G}'[\mathbf{G}\mathbf{V}_{\hat{\mathbf{Z}}}\mathbf{G}']^{-1}(\mathbf{Z}^* - \mathbf{G}\hat{\mathbf{Z}}) \quad (8)$$

$$\mathbf{V}_{\tilde{\mathbf{Z}}} = \mathbf{V}_{\hat{\mathbf{Z}}} - \mathbf{V}_{\hat{\mathbf{Z}}}\mathbf{G}'[\mathbf{G}\mathbf{V}_{\hat{\mathbf{Z}}}\mathbf{G}']^{-1}\mathbf{G}\mathbf{V}_{\hat{\mathbf{Z}}}. \quad (9)$$

To impute the missing values, we perform $M \geq 2$ draws from the posterior probability distribution of $\mathbf{Z}$, with parameters defined by (8) and (9). The $M$ versions of the imputed data set can be analyzed using standard complete data methods, and the results can be combined by

simple rules such that a single inferential conclusion that incorporates the missing data uncertainty is obtained.

Suppose that a parametric model has been estimated and we are interested in making an inference on the parameter $\theta$. After performing the same analysis on each imputed data set, we obtain $M$ parameter estimates $\hat{\theta}_i$, $i = 1,...,M$ and the associated variance estimates $\hat{W}_i$, $i = 1,...,M$. Following Little and Rubin (1987), the overall estimate can be derived as follows

$$\bar{\theta}_M = M^{-1} \sum_{i=1}^{M} \hat{\theta}_i .$$

The total variance associated with $\bar{\theta}_M$ has two components that account for the variability within each imputation and the variability across imputations. The *within-imputation variance* is computed by averaging the variance estimates from each imputed data set as

$$\bar{W}_M = M^{-1} \sum_{i=1}^{M} \hat{W}_i .$$

The *between-imputation variance* is represented by the sample variance of the parameter estimates, that is,

$$B_M = (M-1)^{-1} \sum_{i=1}^{M} (\hat{\theta}_i - \bar{\theta}_M)^2 .$$

The total variability associated with $\bar{\theta}_M$ is then computed as follows

$$T_M = \bar{W}_M + (1 + M^{-1}) B_M$$

where the factor $(1+M^{-1})$ is an adjustment that is introduced to account for the finite number of used imputations. Note that $\overline{W}_M$ estimates the variance if the data were complete, whereas $B_M$ measures the increase in variance due to missing data. The between imputation component plays a central role in the analysis because it allows for quantifying the uncertainty introduced by the missing data imputation: a high value of $B_M$ is reported when the estimates vary greatly across the imputed data sets; in contrast, a lower value of $B_M$ denotes a low uncertainty of estimates due to the missing data imputation.

The proposed method can be generalized in a multivariate framework by considering the random vector $\mathcal{Z} = [\mathbf{Z_1}, \mathbf{Z_2}, ..., \mathbf{Z}_p]'$, where $\mathbf{Z}_k$, $k = 1, ..., p$, are $n \times 1$ vectors, with subscripts referring to distinct variables. Under the CAR assumption, the distribution of $\mathcal{Z}$ is multivariate normal with mean vector $\boldsymbol{\mu}_{\mathcal{Z}} = (\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, ..., \boldsymbol{\mu}_p)'$, where $\boldsymbol{\mu}_k = (\mu_{k1}, ..., \mu_{kn})'$ for $k = 1, ..., p$, and covariance matrix $\boldsymbol{\Sigma}_{\mathcal{Z}}$, which, under some simplifying assumptions, can be expressed as

$$\boldsymbol{\Sigma}_{\mathcal{Z}} = \left[\boldsymbol{\Lambda} \otimes (\mathbf{D} - \rho \mathbf{C})\right]^{-1} \qquad (10)$$

where $\boldsymbol{\Lambda}$ denotes a $p \times p$ symmetric and positive definite non-spatial precision (inverse covariance) matrix.

The covariance structure in (10) characterizes the specification that for the zero centred multivariate CAR is denoted by MCAR ( $\rho$ , $\boldsymbol{\Lambda}$ ), according to Carlin and Banerjee (2003) (see also Gelfand and Vounatsou, 2003). According to this specification, the covariance matrix is expressed as the Kronecker product of the univariate CAR form and $\boldsymbol{\Lambda}$ ; thus, the condition for its positive definiteness reduces to $|\rho| < 1$, as in the univariate setting.

This model can also be generalized to include different autocorrelation parameters for each variable, as shown by Gelfand and Vounatsou (2003) and Carlin and Banerjee (2003).

In this multivariate framework, the incomplete data generating process can be expressed by $\mathcal{Z}^* = [\mathbf{Z}_1^*,...,\mathbf{Z}_p^*]'$, where each $\mathbf{Z}_k^*$ is an $m \times 1$ vector. To define a correspondence between incomplete and complete data, the linear transformation operator $\mathcal{G}$ is now specified as follows

$$\mathcal{G} = \begin{bmatrix} \mathbf{G_1} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{G_2} & ... & \mathbf{0} \\ ... & ... & ... & ... \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{G}_p \end{bmatrix}$$

where $\mathbf{G}_k$, $k = 1,..., p$, are $m \times n$ binary matrices that express the missingness mechanism for each variable. Then, the random vector $\mathcal{Z}^*$ can be obtained as follows

$$\begin{bmatrix} \mathbf{Z}_1^* \\ \mathbf{Z}_2^* \\ ... \\ \mathbf{Z}_p^* \end{bmatrix} = \begin{bmatrix} \mathbf{G_1} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{G_2} & ... & \mathbf{0} \\ ... & ... & ... & ... \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{G}_p \end{bmatrix} \begin{bmatrix} \mathbf{Z_1} \\ \mathbf{Z_2} \\ ... \\ \mathbf{Z}_p \end{bmatrix}.$$

From the hypothesis on the distribution of $\mathcal{Z}$, we derive the conditional distribution of $\mathcal{Z}^* \mid \mathcal{Z}$ as in (5). As previously mentioned, the distribution of $\mathcal{Z}^* \mid \mathcal{Z}$ is multivariate normal and expresses the likelihood of $\mathcal{Z}$ based on the observed data. Given the multivariate normal prior distribution for $\mathcal{Z}$, the application of Bayes' rule yields the posterior probability distribution of $\mathcal{Z} \mid \mathcal{Z}^*$, which is again a multivariate normal distribution, with parameters specified according to the BIM estimates. Finally, the multiple imputation procedure is carried out by repeatedly drawing from the posterior distribution of $\mathcal{Z} \mid \mathcal{Z}^*$. Note that in particular situations (as may occur in border or isolated regions), the proposed approach could create local spatial outliers.

## 3. Empirical illustration

The proposed method is applied to data on labour productivity in European regions. The variables considered are the growth rate of labour productivity and the growth rate of output. The application concerns 187 NUTS-2 regions, and the period under investigation is 1980-2008. Data are derived from the REGIO database (Eurostat) and the Cambridge Econometrics data set.

The labour productivity is analyzed using Verdoorn's law (see, e.g., Piras et al. 2012), which states the existence of a linear relationship between the growth rate of labour productivity and the growth rate of output. The model in its classical non-spatial version can be expressed as

$$\mathbf{p} = \beta_0\mathbf{i} + \beta_1\mathbf{q} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(0,\sigma_\varepsilon^2) \qquad (11)$$

where $\mathbf{p}$ is a vector of observations on the growth rate of labour productivity, $\mathbf{i}$ is a vector of ones, $\mathbf{q}$ denotes the vector of observations on the growth rate of output, and $\boldsymbol{\varepsilon}$ is a vector of error terms. The parameters $\beta_0$ and $\beta_1$ represent the autonomous growth rate of labour productivity and Verdoorn's coefficient, respectively. As reported in the literature, a value of approximately 0.5 is usually found for the parameter $\beta_1$ when the growth rate of productivity and the growth rate of output refer to the manufacturing sector.

An alternative augmented specification of Verdoorn's law, obtained by introducing spatial effects in the model (11), is also considered in this application. Different spatial versions

of Verdoorn's law have been proposed in the literature (see, e.g., Pons-Novell and Viladecans-Marsal, 1999; Fingleton, 2001; Piras et al. 2012). The specification here assumed is based on the Spatial Durbin Model (SDM) that, as highlighted by LeSage and Pace (2009), represents the most general formulation for the spatial autoregressive model. The SDM, under some specified general conditions, subsumes both the spatial error and spatial lag models as special cases. As verified by LeSage and Pace (2009), this specification derives from two circumstances that are highly plausible in the spatial growth regression, i.e., the presence of spatial dependence in the disturbances and the existence of one or more omitted explanatory variables that are spatially correlated with one or more of the other variables that are included in the model. Additional theoretical motivations support the use of SDM against competing approaches. First, this specification yields unbiased coefficient estimates even if the true data generating process is a spatial lag or a spatial error model (Elhorst, 2010). Second, it allows for considering spillover effects without imposing prior restrictions on their magnitude (Elhorst, 2010). Following the SDM specification, Verdoorn's law is expressed as

$$\mathbf{p} = \lambda \mathbf{W}\mathbf{p} + \beta_0 \mathbf{i} + \beta_1 \mathbf{q} + \gamma \mathbf{W}\mathbf{q} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma_\varepsilon^2) \qquad (12)$$

where $\lambda$ and $\gamma$ are scalar parameters of spatial dependence and $\mathbf{W}$ is the normalized contiguity matrix. The model (12) introduces a spatial lag on the dependent variable, **Wp**, as well as a spatial lag on the explanatory variable, **Wq**.

In the application proposed here, starting from the complete data set, different missingness patterns are assumed. First, an MCAR mechanism is assumed: for regions that are selected according to simple random sampling, the observations on each variable are deleted by specifying two different rates of missingness (5% and 20%). Second, two patterns of spatially clustered missing data are obtained by deleting, in turn, all of the observations related

to French and Greek regions. Figures 1 and 2 display the selected missing units, which are coloured in grey.

The broad differences between the French and Greek regions motivate the choice of the patterns of spatially clustered missing data. First, the selected countries differ in their economic structure: the high growth rates characterizing the French economy contrast the lower values reported by the Greek regions. Second, the selected countries exhibit differences in their geographical position, which, for France, appears to be most connected to the other European regions.

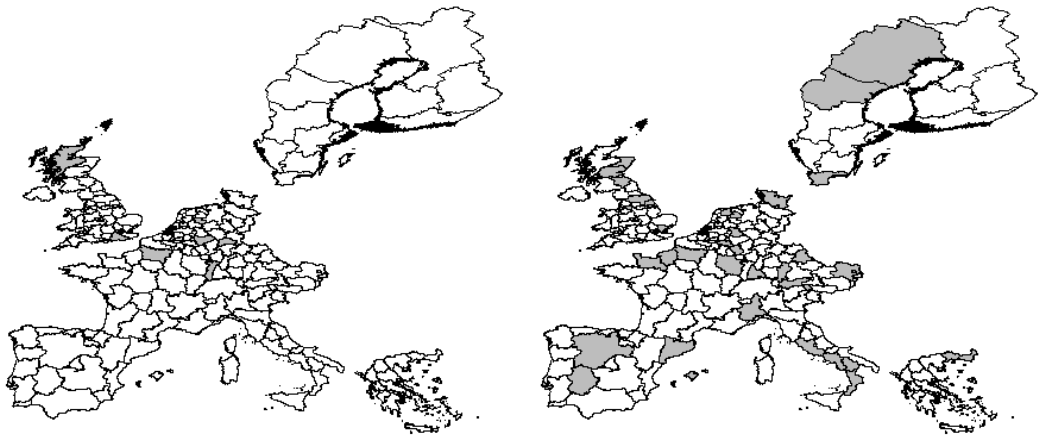

Figure 1: Data Missing Completely At Random. Rates of missingness: 5% and 20%

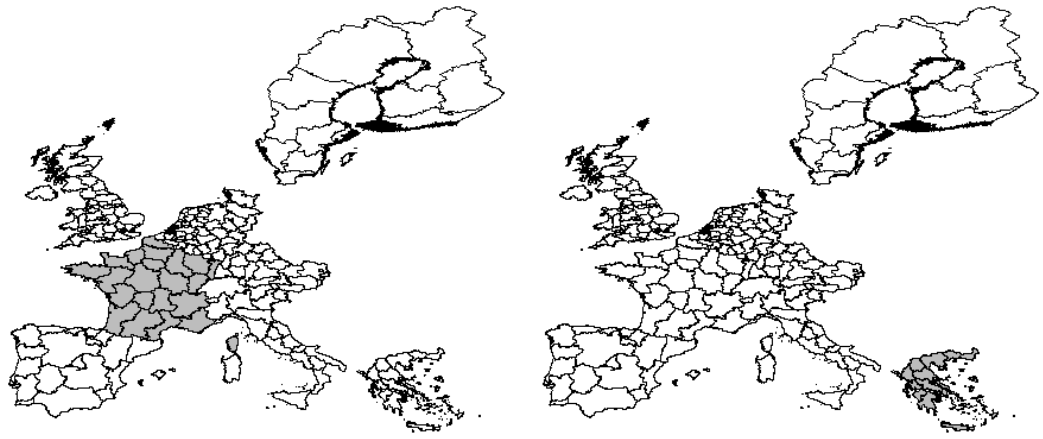

Figure 2: Spatially Clustered Missing Data. Missing observations for French and Greek Regions

In this paper, the proposed methodology is applied in its multivariate version, assuming a multivariate CAR specification for the data generating process. The imputed data sets are generated by 100 draws from the multivariate normal distribution with parameters given by the constrained versions of the BIM estimates. These estimates allow the pycnophylactic property to be preserved. We verified the Normality assumption through the Jarque-Bera test on model (11), estimated by only using the observed data.

In the computation of the BIM estimates, for each variable, a constant prior mean that equals the mean computed over the complete data is assumed for each variable. The covariance matrix of the process, which is assumed to be known, is computed over the complete data by specifying three different values for the autocorrelation parameter: $\rho = 0.06$, $\rho = 0.2$, and $\rho = 0.8$. These different values of $\rho$ are chosen to assess how the extent of spatial autocorrelation impacts the missing data imputation. The contiguity matrix $\mathbf{C}$ is specified according to the $k$-nearest neighbours criterion, with $k = 10$. The matrix is symmetrized by adding neighbours and is normalized. When dealing with European regions, the existence of islands does not allow a weights matrix to be defined by considering only simple geographical

contiguity; if this were the case, the islands would not be connected to the continent. Hence, the ten nearest neighbours based on distance are chosen. Distances between regions are computed as Euclidean distances between region centroids. Following this criterion, Greek regions are connected to South Italy, UK regions are linked to Ireland and continental Europe, and so on. The contiguity matrix has also been specified by assuming different numbers of neighbours, such as $k$=5 and $k$=7. However, such specifications did not produce significant differences in the results.

On the resulting 100 imputed data sets, the parametric models given in (11) and (12) are estimated. Conventional maximum likelihood estimation methods are used to estimate the model parameters. By applying the combining rules described in the previous section, the overall estimate and the associated total variability are derived for each parameter. As a benchmark to evaluate the proposed method, models (11) and (12) are also estimated using both the complete and incomplete data.

Note that the parameter estimates for the imputed data, when they are constrained to selected properties of the observed data, yield comparable parameter estimates. Further, the smoothing that is involved shrinks the variance of these estimates.

In the original Verdoorn's law specification (11), the parameters to be estimated are $\beta_0$, $\beta_1$, and $\sigma_\varepsilon^2$. Table 1 shows the overall estimate for Verdoorn's coefficient, $\hat{\beta}_1$, computed on the imputed data sets. The variability associated with the estimate is decomposed into the Within imputation ( $\overline{W}$ ) and the Between imputation ( $B$ ) components. The results are reported for the three different values of $\rho$ and characterize the covariance structure of the underlying spatial process.

Table 1: Parameters estimates and total variability for the non-spatial Verdoorn's law specification

| Missing | | $\rho = 0.06$ | $\rho = 0.2$ | $\rho = 0.8$ | Complete | Incomplete |
|---|---|---|---|---|---|---|
| 5% | $\hat{\beta}_1$ | 0.6397705 | 0.6397176 | 0.6391380 | 0.6430498 | 0.6395171 |
| | $\overline{W}_{\hat{\beta}_1}$ | 0.0027036 | 0.0027031 | 0.0027001 | 0.0027604 | 0.0028433 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | $B_{\hat{\beta}_1}$ | 0.0000000 | 0.0000000 | 0.0000004 | | |
| 20% | $\hat{\beta}_1$ | 0.6699252 | 0.6698544 | 0.6648343 | 0.6430498 | 0.6685362 |
| | $\overline{W}_{\hat{\beta}_1}$ | 0.0028924 | 0.0028937 | 0.0029450 | 0.0027604 | 0.0036270 |
| | $B_{\hat{\beta}_1}$ | 0.0000001 | 0.0000003 | 0.0000031 | | |
| 22 French regions | $\hat{\beta}_1$ | 0.6513170 | 0.6512049 | 0.6488600 | 0.6430498 | 0.6505908 |
| | $\overline{W}_{\hat{\beta}_1}$ | 0.0027471 | 0.0027474 | 0.0027545 | 0.0027604 | 0.0031224 |
| | $B_{\hat{\beta}_1}$ | 0.0000000 | 0.0000001 | 0.0000007 | | |
| 13 Greek regions | $\hat{\beta}_1$ | 0.6127962 | 0.6128118 | 0.6132057 | 0.6430498 | 0.6122349 |
| | $\overline{W}_{\hat{\beta}_1}$ | 0.0025812 | 0.0025813 | 0.0025861 | 0.0027604 | 0.0027778 |
| | $B_{\hat{\beta}_1}$ | 0.0000001 | 0.0000001 | 0.0000005 | | |

When data are MCAR at a 5% rate, the values of the overall estimate, $\hat{\beta}_1$, that are found using the imputed data sets are close to the estimate computed on the complete data. The three different values of $\rho$ do not seem to produce significant differences in the results. When a larger percentage of missing data is considered, the values of the overall estimates on the imputed data sets are slightly higher than the value computed for the complete data. However, the value of $\hat{\beta}_1$ slowly decreases when the value of the autocorrelation parameter $\rho$ increases.

Similar results are obtained considering spatially clustered missing data. When the missing observations are related to French regions, the overall estimate on the imputed data sets is closer to the estimate on the complete data, particularly for higher values of $\rho$. When missing observations are related to Greek regions, the values of $\hat{\beta}_1$ computed on the imputed data sets are smaller than the estimate on the complete data. However, the increasing of $\rho$ produces a slow increase in the value of the estimates.

Satisfactory results are also obtained for the variability associated with the estimates. For each pattern of missing data, the Within imputation component is close to the variability associated with the estimates on the complete data. This result is especially noticeable when

data are MCAR at a 5% rate and when the spatially clustered missing data are related to French regions. Furthermore, for each pattern of missing data, the Between imputations variability assumes values close to zero. This result reveals that the estimates do not vary greatly across the imputed data sets and thus denote a low uncertainty of estimates due to the missing data imputation.

Table 2 shows the results obtained when spatial effects are introduced in the original Verdoorn's law specification (see equation 12). With respect to the original Verdoorn's Law, in the Spatial Durbin model, additional parameters must be estimated, such as $\lambda$ and $\gamma$. The correct interpretation of the parameter estimates requires determining the direct and indirect effects (LeSage and Fischer, 2008; LeSage and Pace, 2009). Empirical illustrations of the estimation and interpretation of direct and indirect effects for spatial models of economic growth can be found in the works of LeSage and Fischer (2008), Fischer (2011), and Panzera and Postiglione (2014).

For the complete data, Verdoorn's coefficient, estimated for the spatial specification of the model, assumes a value close to the estimate obtained for the original specification. The spatial dependence parameter $\lambda$, which is associated with the dependent variable, assumes a positive value, whereas a negative value is reported for $\gamma$.

Table 2: Parameters estimates and total variability for the Spatial Durbin Model

| Missing | | $\rho = 0.06$ | $\rho = 0.2$ | $\rho = 0.8$ | Complete | Incomplete |
|---|---|---|---|---|---|---|
| 5% | $\hat{\beta}_1$ | 0.6343749 | 0.6342559 | 0.6344380 | 0.6398092 | 0.6327409 |
| | $\overline{W}_{\hat{\beta}_1}$ | 0.0023290 | 0.0023313 | 0.0023386 | 0.0023594 | 0.0024572 |
| | $B_{\hat{\beta}_1}$ | 0.0000011 | 0.0000009 | 0.0000001 | | |
| | $\hat{\lambda}$ | 0.7105628 | 0.7130806 | 0.7218097 | 0.7266684 | 0.7087666 |
| | $\overline{W}_{\hat{\lambda}}$ | 0.0056529 | 0.0055781 | 0.0053213 | 0.0051800 | 0.0057547 |
| | $B_{\hat{\lambda}}$ | 0.0000009 | 0.0000007 | 0.0000005 | | |
| | $\hat{\gamma}$ | -0.4510755 | -0.4524839 | -0.4591899 | -0.4686678 | -0.4440683 |
| | $\overline{W}_{\hat{\gamma}}$ | 0.0112934 | 0.0111905 | 0.0108317 | 0.0108931 | 0.0117341 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | $B_{\hat{\gamma}}$ | 0.0000033 | 0.0000039 | 0.0000012 | | |
| 20% | $\hat{\beta}_1$ | 0.6627899 | 0.6643763 | 0.6718414 | 0.6398092 | 0.6641202 |
| | $\overline{W}_{\hat{\beta}_1}$ | 0.0028478 | 0.0028325 | 0.0027683 | 0.0023594 | 0.0034675 |
| | $B_{\hat{\beta}_1}$ | 0.0000068 | 0.0000067 | 0.0000010 | | |
| | $\hat{\lambda}$ | 0.6467332 | 0.6631801 | 0.7251719 | 0.7266684 | 0.6619643 |
| | $\overline{W}_{\hat{\lambda}}$ | 0.0076423 | 0.0071129 | 0.0052210 | 0.0051800 | 0.0072060 |
| | $B_{\hat{\lambda}}$ | 0.0000073 | 0.0000065 | 0.0000032 | | |
| | $\hat{\gamma}$ | -0.4365654 | -0.4518131 | -0.5133435 | -0.4686678 | -0.4447774 |
| | $\overline{W}_{\hat{\gamma}}$ | 0.0142139 | 0.0134470 | 0.0107366 | 0.0108931 | 0.0135966 |
| | $B_{\hat{\gamma}}$ | 0.0000164 | 0.0000206 | 0.0000069 | | |
| 22 French regions | $\hat{\beta}_1$ | 0.6394492 | 0.6405765 | 0.6441587 | 0.6398092 | 0.6447132 |
| | $\overline{W}_{\hat{\beta}_1}$ | 0.0023844 | 0.0023831 | 0.0023801 | 0.0023594 | 0.0027115 |
| | $B_{\hat{\beta}_1}$ | 0.0000010 | 0.0000008 | 0.0000002 | | |
| | $\hat{\lambda}$ | 0.7204413 | 0.7226594 | 0.7333052 | 0.7266684 | 0.7075169 |
| | $\overline{W}_{\hat{\lambda}}$ | 0.0053610 | 0.0052963 | 0.0049890 | 0.0051800 | 0.0057292 |
| | $B_{\hat{\lambda}}$ | 0.0000007 | 0.0000006 | 0.0000008 | | |
| | $\hat{\gamma}$ | -0.4504587 | -0.4545097 | -0.4715823 | -0.4686678 | -0.4481005 |
| | $\overline{W}_{\hat{\gamma}}$ | 0.0112316 | 0.0111190 | 0.0106021 | 0.0108931 | 0.0121907 |
| | $B_{\hat{\gamma}}$ | 0.0000024 | 0.0000023 | 0.0000014 | | |
| 13 Greek regions | $\hat{\beta}_1$ | 0.6018369 | 0.6017911 | 0.6016224 | 0.6398092 | 0.6008760 |
| | $\overline{W}_{\hat{\beta}_1}$ | 0.0021742 | 0.0021739 | 0.0021715 | 0.0023594 | 0.0023327 |
| | $B_{\hat{\beta}_1}$ | 0.0000002 | 0.0000001 | 0.0000001 | | |
| | $\hat{\lambda}$ | 0.7472892 | 0.7474154 | 0.7484790 | 0.7266684 | 0.7499871 |
| | $\overline{W}_{\hat{\lambda}}$ | 0.0045934 | 0.0045899 | 0.0045602 | 0.0051800 | 0.0050118 |
| | $B_{\hat{\lambda}}$ | 0.0000001 | 0.0000001 | 0.0000001 | | |
| | $\hat{\gamma}$ | -0.4402774 | -0.4402676 | -0.4402747 | -0.4686678 | -0.4410248 |
| | $\overline{W}_{\hat{\gamma}}$ | 0.0094750 | 0.0094718 | 0.0094511 | 0.0108931 | 0.0101901 |
| | $B_{\hat{\gamma}}$ | 0.0000004 | 0.0000002 | 0.0000004 | | |

When data are MCAR at a 5% rate, the estimates computed on the imputed data sets for $\beta_1$, $\gamma$ and $\lambda$ are close to the estimates carried out on the complete data. Higher values of $\rho$ seem to determine an improvement of the estimates. As the percentage of missing data increases, the estimates of Verdoorn's coefficient, carried out on the imputed data sets, are larger than the

value obtained for the complete data. The estimates for $\lambda$ are closer to the value obtained for the complete data; this result is especially noticeable when the value of the autocorrelation parameter $\rho$ increases. For both percentages of missing data and higher values of $\rho$, the Within imputation variability is close to the variability associated with the parameter estimates carried out on the complete data. An increase in $\rho$ also produces a decrease in $B$, which, for all of the estimates, assumes values close to zero and thus denotes a low uncertainty due to the missing data imputation.

When the missing data are spatially clustered, particularly for French regions, the overall estimate, computed on the imputed data sets for each parameter in the model, is close to the estimate on the complete data. When missing data are related to Greek regions, values smaller than those estimated on the complete data are obtained for $\hat{\beta}_1$ and $\hat{\gamma}$, whereas higher values are obtained for $\hat{\lambda}$. For both the patterns of spatially clustered missing data, the Within imputation variability for each parameter assumes values close to the variability associated with the estimates on the complete data. The best results are reported when the missing data are related to French regions. For both patterns of missing data and for all of the parameter estimates, the Between imputation component is close to zero.

The empirical evidence reveals the best performances of the method when, for each site, more information on the neighbourhood is exploited, that is, when $\rho$ is higher. This finding means that the similarity between nearest sites contributes positively to the missing data imputation. The low similarity that the Greek regions exhibit with respect to their nearest neighbours thus gives a possible explanation for the worst results reported for this pattern of missing data.

As previously mentioned, the interpretation of the Spatial Durbin Model should be based on the estimated direct and indirect impacts rather than the estimated coefficients. Scalar summary measures for these effects were introduced by LeSage and Pace (2009). The average

direct impact expresses the impact that changes in the explanatory variable for a spatial unit have on the dependent variable of the same unit and includes feedback influences that arise as a result of impacts that, starting from a spatial unit, pass through neighbouring units and return to the unit itself. The average indirect impact reflects how a change in an explanatory variable in all regions $j$ ( $j \neq i$ ) influences the dependent variable of a particular region $i$. The sum of the average direct and indirect impacts measures the average total impact. The average impacts computed for the spatial Durbin model are presented in Table 3. A set of 1000 MCMC draws was used to produce the impact estimates. These average impacts were computed for the imputed data sets and for both the complete and incomplete data.

Table 3 : Average Impacts for Spatial Durbin Model

| Missing | | $\rho = 0.06$ | $\rho = 0.2$ | $\rho = 0.8$ | Complete | Incomplete |
|---|---|---|---|---|---|---|
| 5% | Average Direct Impact | 0.6342425 | 0.6342566 | 0.6342640 | 0.6392916 | 0.6333341 |
| | Average Indirect Impact | -0.0002975 | -0.0009934 | -0.0041673 | -0.0131602 | 0.0145057 |
| | Average Total Impact | 0.6339450 | 0.6332632 | 0.6300967 | 0.6261314 | 0.6478398 |
| 20% | Average Direct Impact | 0.6619818 | 0.6635643 | 0.6683012 | 0.6392916 | 0.6634259 |
| | Average Indirect Impact | -0.0218798 | -0.0345816 | -0.0913369 | -0.0131602 | -0.0145513 |
| | Average Total Impact | 0.6401020 | 0.6289827 | 0.5769643 | 0.6261314 | 0.6488746 |
| 22 French regions | Average Direct Impact | 0.6409389 | 0.6416603 | 0.6442141 | 0.6392916 | 0.6458421 |
| | Average Indirect Impact | 0.0345698 | 0.0297313 | 0.0033918 | -0.0131602 | 0.0263769 |
| | Average Total Impact | 0.6755087 | 0.6713916 | 0.6476059 | 0.6261314 | 0.6722190 |
| 13 Greek regions | Average Direct Impact | 0.6032855 | 0.6031989 | 0.6031625 | 0.6392916 | 0.6023090 |
| | Average Indirect Impact | 0.0361147 | 0.0363766 | 0.0379695 | -0.0131602 | 0.0370627 |
| | Average Total Impact | 0.6394002 | 0.6395755 | 0.6411320 | 0.6261314 | 0.6393717 |

A comparison between the direct impact estimate and the estimate of $\beta_1$ computed over the complete data set shows that these estimates are similar in magnitude. The differences between these estimates are due to feedback effects (Fischer, 2011). The average indirect impact

estimated for the complete data is negative, such as the coefficient associated with the spatially lagged dependent variable ($\hat{\gamma}$). The negative average indirect impact indicates that an increase in the growth rate of output for all regions would result in a decrease in the growth rate of labour productivity of a particular region.

When data are MCAR, the impact estimates computed over the imputed data sets are similar in magnitude to the estimates computed for the complete data. When missing data are spatially clustered, we found positive average indirect impacts.

For all of the patterns of missing data, the average total impacts computed over the imputed data sets are close to the average total impacts computed for the complete data. The results improve when the value of $\rho$ increases. This circumstance is especially noticeable when the rate of missingness is 5% and when the missing data are related to French regions.

## 4. Conclusions

The development of valid procedures for dealing with spatial missing data is a challenge in many scientific fields. Incomplete spatial data sets require imputation methods that account for the complex autocorrelation structure that characterizes this type of data. Most of the methods proposed in both the early and recent literature overcome the spatial missing data problem by applying single imputation techniques and likelihood-based approaches.

The method proposed in this paper extends the previous works in this area. This approach is developed in a Bayesian framework and is based on the assignment of a set of likely values to missing data. The spatial data generating process is modelled through a CAR specification, and each missing value is replaced by a set of values that are drawn from a multivariate normal distribution. The parameters of this distribution are derived by the application of the BIM, which was introduced by Benedetti and Palma (1994). The BIM theory, which was originally

developed to face the areal data disaggregation, is developed here to address the missing data problem and is extended to a multivariate framework.

The proposed method requires only the Normality assumption on the distribution of the underlying spatial process. As additional advantages, the method exploits the spatial dependence between observations in the imputation procedure and allows the uncertainty associated with the missing data imputation to be incorporated into the analysis.

In this paper, the method is applied to complete a data set whose missing observations have been generated according to different patterns. The considered data set is related to the growth rate of labour productivity in NUTS-2 European regions. Two parametric models, i.e., the original specification of Verdoorn's law and one of its spatial extensions, have been estimated on the imputed data sets. The results have been combined such that the uncertainty of imputations has been considered.

In the empirical exercise, the method seems to perform well for both data that are missing completely at random and spatially clustered missing data. For both the considered models, the parameter estimates carried out on the imputed data sets assume values close to the estimates obtained for the complete data. The variability associated with the estimates on the imputed data sets is close to that obtained on the complete data.

Regardless of the spatial configuration of missing data, the method revealed a low uncertainty associated with the missing data imputation. Such a result could likely be extended to different missing data patterns. The empirical exercise revealed that the performance of the method improved when the value of the spatial autocorrelation parameters increased. This result seems to confirm the importance of exploiting the spatial dependence effect in the missing spatial data imputation.

The proposed method has been performed by assuming total non-responses, which occur when, for a given unit, data are missing for all of the variables in the data set. The method can

be easily extended to address partial non-responses, which arise when, for a given unit, data are missing only for some of the variables in the data set.

The proposed method can also be generalized by assuming a separate spatial dependence parameter for each component of the multivariate process. Furthermore, the implementation of CAR models can be easily extended to the Simultaneous Autoregressive specification (Whittle, 1954). Regarding this, see Panzera and Postiglione (2014).

Further developments could involve extending the methodology to spatial-temporal data and addressing non-Gaussian random processes.

# References


Anderson, T. W. (1958) *An introduction to multivariate statistical analysis*, John Wiley, New York.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Benedetti R. & Palma D. (1994) Markov random field-based image subsampling method, *Journal of Applied Statistics*, 21(5), 495-509.

Bennett, R. J., Haining, R. P. & Griffith, D. A. (1984) The Problem of Missing Data on Spatial Surfaces, *Annals of the Association of American Geographers*, 74(1), 138-156.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society B*, 36, 192-225.

Birkin, M. & Clarke, M. (1989) The Generation of Individual and Household Incomes at the Small Area Level using Synthesis, *Regional Studies*, 23, 535-548.

Brook, D. (1964) On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems, *Biometrika*, 51, 481-483.

Carlin, B. P. & Banerjee, S. (2003) Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion), in *Bayesian Statistics 7* (Eds) J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid , D. Heckerman, A.F.M. Smith & M. West, Oxford University Press, Oxford, 45-63.

Chun, Y. & Griffith, D.A. (2013) *Spatial Statistics & Geostatistics*, SAGE, Thousand Oaks, CA.

Clayton, D. & Berardinelli, L. (1992) *Bayesian methods for mapping disease risk*, in *Geographical and Environmental Epidemiology: Methods for Small-Area Studies* (Eds.) P. Elliott, J. Cuzick, D. English & R. Stern, Oxford University Press, Oxford, 205-220.

Cliff, A. D. & Ord, J.K. (1973) *Spatial Autocorrelation,* Pion, London.

Cliff, A. D. & Ord, J.K. (1981) *Spatial Processes, Models and Applications*, Pion, London.

Conley, T. & Topa, G. (2002) Socio-economic distance and spatial patterns in unemployment, *Journal of Applied Econometrics*, 17(4), 303-327.

Cressie, N. (1993) *Statistics for Spatial data*, Wiley, New York.

Dempster, A. P., Laird, N. M. & Rubin, D. P. (1977) Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society Series B*, 39, 1-38

Elhorst, J. P. (2010) Applied Spatial Econometrics: Raising the Bar, *Spatial Economic Analysis*, 5(1), 9-28.

Fingleton, B. (2001) Equilibrium and economic growth: spatial econometric models and simulation, *Journal of Regional Science*, 41, 117-147.

Fischer, M. M. (2011) A spatial Mankiw-Romer-Weil model: theory and evidence, *The Annals of Regional Science*, 47, 419-436.

Gaetan, C. & Guyon, X. (2010) *Spatial Statistics and Modeling*, Springer-Verlag, New York.

Gelfand, A. E. & Vounatsou, P. (2003) Proper multivariate conditional autoregressive models for spatial data analysis, *Biostatistics*, 4(1), 11-25.

Griffith, D. A., Bennett, R. J. & Haining, R. P. (1989) Statistical Analysis of Spatial Data in the Presence of Missing Observations: A Methodological Guide and an Application to Urban Census Data, *Environment & Planning A*, 21, 1511-1523.

Griffith, D. & Layne, L. (1999) *A Casebook for Spatial Statistical Data Analysis: A Compilation of Analyses of Different Thematic Data Sets*, Oxford University Press, New York.

Haining, R. P. (1978) The Moving Average Model for Spatial Interaction, *Transactions of the Institute of British Geographers New Series,* 3(2), 205-225.

Haining, R. P. (2003) *Spatial Data Analysis: Theory and Practice*, Cambridge University Press, Cambridge.

Haining, R. P., Griffith, D. A. & Bennett, R. J. (1984) A statistical approach to the problem of missing spatial data using a First-order Markov Model, *Professional Geographer*, 36(3), 338-345.

Haining, R. P., Griffith, D. A. & Bennett, R. J. (1989) Maximum likelihood estimation with missing spatial data and with an application to remotely sensed data, *Communications in Statistics - Theory and Methods*, 18, 1875-1894.

Kelejian, H.H. & Prucha, I.R. (2010) Spatial models with spatially lagged dependent variable and incomplete data, *Journal of Geographical System*, 12, 241-257.

Kennedy, S. & Tobler, W.R. (1983) Geographic interpolation, *Geographical Analysis*, 15, 151-6.

LeSage, J. P. & Pace, R. K. (2004) Models for Spatially Dependent Missing Data, *Journal of Real Estate Finance and Economics*, 29, 233-254.

LeSage, J. P. & Fischer, M. M. (2008) Spatial Growth Regressions: Model Specification, Estimation and Interpretation, *Spatial Economic Analysis*, 3(3), 275-304.

LeSsage, J. P. & Pace, R. K. (2009) *Introduction to Spatial Econometrics*, Chapman & Hall/CRC Press, Boca Raton, Florida.

Little, R. J. A. & Rubin, D. B. (1987) *Statistical Analysis with Missing Data*, John Wiley and Sons, New York.

Longford, N. T. (2008) *Studying Human Populations. An Advanced Course in Statistics*, Springer, New York, 129-161.

Martin, R. J. (1984) Exact maximum likelihood for incomplete data from a correlated gaussian process, *Communications in Statistics - Theory and Method*, 13, 1275-1288.

Oakes, J. (1973) A forecasting approaches to the missing data problem, Discussion paper presented at the Institute of British Geographers, Historical Geography Research Group, Birmingham.

Oh, M. S., Shin, D. W. & Kim, H. J. (2002) Bayesian analysis of regression models with spatially correlated errors and missing observations, *Computational Statistics and Data Analysis*, 39, 387-400.

Palma, D., Benedetti, R. (1998). A transformational view of spatial data analysis. *Geographical Systems*, 5, 199-220.

Panzera, D. & Postiglione, P. (2014) Economic growth in Italian NUTS 3 provinces, *The Annals of Regional Science*, 53(1), 273-293.

Pilz, J. (1991) *Bayesian estimation and experimental design in linear regression models*, John Wiley, New York.

Piras, G., Postiglione, P. & Aroca, P. (2012) Specialization, R&D and productivity growth: evidence from EU regions, *The Annals of Regional Science*, 49(1), 35-51.

Pons-Novell, J. & Viladecans-Marsal, E. (1999) Kaldor's laws and spatial dependence: evidence for the European regions, *Regional Studies*, 33(5), 443-451.

Qiu, F. & Cromley, R. (2013) *Special Issue: Areal Interpolation and Dasymetric Modeling*, Geographical Analysis, 45(3), 213-344.

Ripley, B. D. (1981) *Spatial Statistics*, John Wiley, New York.

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*, John Wiley and Sons, New York.

Schafer, J. L. (1997) *Analysis of Incomplete Multivariate Data*, Chapman & Hall/CRC Press, Boca Raton, Florida.

Strauss, D. J. (1977) Clustering on colored lattice, *Journal of Applied Probability*, 14, 135-143.

Tobler, W. R. (1979) Smooth pycnophylactic interpolation for geographical regions, *Journal of the American Statistical Association*, 74, 519-530.

Tobler, W. R. & Kennedy, S. (1985) Smooth multi-dimensional interpolation, *Geographical Analysis*, 17, 251-257.

Wall, M. M. (2004) A close look at the spatial structure implied by the CAR and SAR models, *Journal of Statistical Planning and Inference*, 121, 311-324.

Whittle, P. (1954) On stationary process in the plane, *Biometrika*, 41, 434-449.

Willmott, C. & Wicks, D. E. (1980) An empirical method for the spatial interpolation of montly precipitation within California, *Physical Geography*, 1, 59-73.