UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
E METODI MATEMATICI°

QUADERNI DEL DIPARTIMENTO

Department of Information Technology and Mathematical Methods

Working Paper

Series "*Mathematics and Statistics*"

n. 9/MS – 2011

Geostatistical modelling of spatial potentials

by

Francesco Finazzi

° Viale Marconi. 5, I – 24044 Dalmine (BG), ITALY, Tel. +39-035-2052339; Fax. +39-035-562779

[§] L'accesso alle *Series* è approvato dal Comitato di Redazione. I *Working Papers* della Collana dei Quaderni del Dipartimento di Ingegneria dell'Informazione e Metodi Matematici costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

Geostatistical modelling of spatial potentials

Francesco Finazzi*

Abstract

Discrete systems often manifest themselves as spatially continuous phenomena over a region of space. A system is studied by measuring the corresponding spatial phenomenon at some locations over space and by drawing conclusion from the estimated spatial surface. In many cases, the discrete nature of the system has to be taken into account as well as the interaction between the system and the measuring instruments used to collect the data. As a consequence, neither the classic geostatistical models nor the classic methods provided by the point process statistics are suitable for estimating the phenomenon spatial surface. In this paper, the general problem is studied by introducing a novel paradigm which justifies the definition of a new class of geostatistical models able to address both the discreteness of the system and its interaction with the measuring instruments. The proposed class of models find application in many scientific fields including both the natural and the social sciences. The model estimation problem is solved in detail and a case study related to a geomarketing application is addressed.

Keywords: spatial statistics, spatial interaction, potential estimation, discrete systems, geomarketing, missing data.

*Corresponding author: Department of Information Technology and Mathematical Methods, University of Bergamo, Viale Marconi, 5, 24044, Dalmine (BG), Italy. francesco.finazzi@unibg.it

1 Introduction

Depending on the data generation mechanism, spatial point data can be analyzed with the tools provided by geostatistics or point process statistics. Geostatistics assumes an underline spatially continuous process which is observed at the spatial locations defined by the points. Usually, the aim is to recover the process in terms of its realization at some locations of interest. On the other hand, point process statistics focuses on the point locations and it mostly tries to explain their spatial pattern.

Within the point process approach, the marked point processes (MPPs) are considered when points are related to scalar or vector values, the marks, that describe some features of the spatial locations at the points. Although, in principle, the marks could be modeled by means of spatially continuous processes, the idea is that the marks are not defined "outside" the sampling locations. As a consequence, it is conceptually wrong to estimate the mark value at some new location. For this reason, the geostatistical and the point process statistical approaches remain different in their purposes. Indeed, the misuse of one approach or the other may lead to inconsistent results. Nevertheless, in some cases, it is not immediately clear which approach is best for the available data and an appropriate statistical model is chosen only after careful analysis.

1.1 Point process models and biological competition

Suppose, for example, that $y(\mathbf{s}_i)$ represents the annual number of items of a given product sold by a store at the spatial location $\mathbf{s}_i \in \mathcal{D} \subset \mathbb{R}^2$. In the notation of the MPPs, $y(\mathbf{s}_i) \equiv \{[\mathbf{s}_i; y(\mathbf{s}_i)]\}$, where $\mathbf{s}_i$ is the location of the point and $y(\mathbf{s}_i)$ is the mark value. The spatial support of the variable "sold items" is clearly not continuous since it is defined only at the spatial locations $\mathcal{S} = \{\mathbf{s}_1, ..., \mathbf{s}_N\}$ of all

the stores in $\mathcal{D}$. Moreover, it may be assumed that $y(\mathbf{s}_i) = f(\mathbf{s}_i, \mathcal{S}_{-i})$, namely the number of items sold at the spatial location $\mathbf{s}_i$ is a function of both $\mathbf{s}_i$ and the positions $\mathcal{S}_{-i}$ of all the other stores. In other words, it is assumed that a "biological competition" is in force and that the number of items sold at $\mathbf{s}_i$ depends on the presence of other stores nearby $\mathbf{s}_i$.

In the first instance, the point process statistics provides useful models for the analysis of the sales data. For example, the *geostatistical marking* provides allows to model spatially correlated marks by considering a stationary random field $z(\mathbf{s})$ and by assuming the relationship $y(\mathbf{s}_i) = z(\mathbf{s}_i)$ between the mark values and the random field. As pointed out in [1], however, the geostatistical marking fails to model point interaction at short distances, which is a key aspect in many applications.

An alternative approach is represented by the so called *statistical mark construction model*. Constructed marks are marks deterministically evaluated from the realization of the point process. For example, $y(\mathbf{s}_i) = d(\mathbf{s}_i)$, with $d(\mathbf{s}_i)$ the distance between $\mathbf{s}_i$ and its nearest neighbor in $\mathcal{S}_{-i}$. Staying with $d(\mathbf{s}_i)$, the statistical mark construction model assumes

$$y(\mathbf{s}_i) = f(d(\mathbf{s}_i)) + \varepsilon(\mathbf{s}_i) \tag{1}$$

where $f : \mathbb{R}^+ \longrightarrow \mathbb{R}$ is a generic function and $\varepsilon(\mathbf{s}_i)$ is the error term. Although (1) is able to handle point interaction, it is not enough flexible for modelling the complexity of real application data since $y(\mathbf{s}_i)$ is described only in terms of the point pattern.

A more flexible solution is represented by the intensity-weighted log-Gaussian Cox (ILGC) model which assumes the following local density of the point process

$$\Lambda(\mathbf{s}) = \exp(w(\mathbf{s})) \tag{2}$$

where $w(\mathbf{s})$ is a Gaussian random field. The mark model is

$$y(\mathbf{s}_i) = a + b\Lambda(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i) \tag{3}$$

with a, b constants and $\varepsilon(\mathbf{s}_i)$ a random error. Positive values of b correspond to higher mark values in high density areas while negative values correspond to lower mark values in high density areas.

1.2 Towards a new class of geostatistical models

Going back to the sales data example, even the ILGC model appears too rigid. This is due to the fact that a and b are constant over space, which may not be realistic. Nevertheless, force a relationship between the marks and a latent random field seems the natural way to proceed.

In the sales data example, though it might be interesting to explain the relationship between the spatial density of the stores and the sales data, a problem worth to be solved is how to relate the sales data to a spatially continuous latent random field $w(\mathbf{s})$ representing the *spatial market potential*, which may be defined as the number of items that could be sold if a store is placed at the spatial location $\mathbf{s}$. Indeed, in this as in other applications, there might be no simple relationship between the mark values and the spatial density of the sampling locations $\mathcal{S}$. Thus, on the contrary of what happens when the ILGC is considered, we do not want to impose one. Note that, though the actual stores are located at $\mathcal{S}$, the spatial marked potential (market potential henceforth) is defined for each $\mathbf{s} \in \mathcal{D}$.

Now, if $q(\mathbf{s}) \equiv w(\mathbf{s})$ is the market potential, we are interested in estimating $q(\mathbf{s})$ at each point of $\mathcal{D}$. For instance, we may want to locate the maxima of $q(\mathbf{s})$ to be sure that we have a store near that location. If we want to open a new store, on the other hand, we may want to evaluate the market potential conditioned on the presence of the actual stores. Moreover, we may want to do all this even if we

know a store is located at $\mathbf{s}$ but we don't know $y(\mathbf{s})$ (for instance because the store is our competitor). With this in mind, it can be noted that, though the observed data may be considered as the realization of a MPP, the above problems are better solved by following a geostatistical approach. In particular, a hierarchical spatial models (see [2]) seems to be appropriate for handling a latent random field and missing data, where the missing data problem should not be confused with the edge effect problem (see [3]) typical of the point process data. In fact, what might be missing is a subset of the mark values related to the points rather than the points themselves.

Although the geostatistical approach seems to be appealing, it is immediately clear that a classic geostatistical approach cannot be adopted for the sales data as they are. Due to the interaction between the stores, a small value of $y(\mathbf{s}_i)$ does not necessarily mean a low market potential at the spatial location $\mathbf{s}_i$. Thus, it would make no sense to consider the sales data (the number of items sold at $\mathcal{S}$) and to evaluate a kriging surface over $\mathcal{D}$. Such a surface, besides being conceptually wrong, would not represent by no means the spatial market potential. To be more general, any relationship between the spatial locations S and the mark values, such as the one induced by a biological competition, will be modeled as an interaction between the generic "system" which is observed and the measuring instruments. In particular, the random field $q(\mathbf{s})$ will be directly related to the state of the system one is interested in.

Ultimately, the data we are to consider are so characterized:

- they are a realization of a MPP;
- the marks reflect both the underlying random field $q(\mathbf{s})$ and the interaction between the observed system and the measuring instruments;
- the spatial pattern of the MPP is independent from $q(\mathbf{s})$;

- some mark values might be missing;

and we want to make inference on $q(\mathbf{s})$.

2 Interaction modelling

The effects of the interaction between a system to be measured and a measuring instrument are often neglected since characterized by a low order of magnitude if compared to that of the variability of what is measured. The role of this section is to shed some light on the nature of the above mentioned interaction and to lay the foundations for the analysis of spatial data arising from the interaction. By the end of the section the reader should be convinced that, for many practical applications, the interaction cannot be ignored and that the results obtained by the spatial analysis of such data make sense only if the interaction is properly modeled.

2.1 Introduction

Let $\mathcal{O}$ be a generic system which can be either a physical system or an abstract system. The system is characterized by an internal state π which is not directly measurable. The state π is supposed to be the equilibrium state of $\mathcal{O}$ at a given time t and the equilibrium is supposed to be stable, that is, the state π will not change unless external forces intervene.

Although π is not observable, $\mathcal{O}$ manifests itself over the region of space $\mathcal{D} \subset \mathbb{R}^d$, $d \geq 1$ as a spatially continuous variable $q(\mathbf{s}) : \mathbf{s} \in \mathcal{D}$. The variable $q(\mathbf{s})$ can be seen as a model for the equilibrium state π and I call $q(\mathbf{s})$ a *potential*, though the term does not refer to any particular property of the variable.

The set $\mathcal{M}(\mathcal{S}) = \{m(\mathbf{s}_1), ..., m(\mathbf{s}_N)\}$ is a finite collection of N measuring instruments located at $\mathcal{S}$ and it is called here the *measuring system*. For simplicity,

I assume that the measuring instruments are all of the same type, namely they share the same properties.

If the system $\mathcal{O}$ is to be measured, $\mathcal{O}$ and $\mathcal{M}$ must interact in some way. The interaction is made possible by what I call mediators. In particular, a *mediator* is an entity (for the moment abstract) that can move freely in $\mathcal{D}$, that is, it can reach any point of $\mathcal{D}$ with non-zero probability. The mediators are supposed to permeate $\mathcal{D}$ though their spatial density may differ across space. A mediator has no preference to interact with a particular measuring instrument but it tends to interact with the nearest ones. This possibly implies the existence of some kind of attraction between $\mathcal{M}$ and the mediators. The establishment of such an attraction, however, is not supposed, in itself, to alter the state π of $\mathcal{O}$. The set of all mediators is denoted by $\mathcal{P}$ and it is assumed to be a finite set.

To be more formal, the following definition are introduced.

Definition 1 *the potential* $q(\mathbf{s})$ *is the expected observed value when the system* $\mathcal{O}$ *is measured at the spatial location* $\mathbf{s} \in \mathcal{D}$.

Definition 2 *the conditional potential* $q(\mathbf{s} \,|\mathcal{S})$ *is the expected observed value when the system* $\mathcal{O}$ *is measured at the spatial location* $\mathbf{s} \in \mathcal{D}$ *given that it is concurrently measured at the set of locations* $\mathcal{S} = \{\mathbf{s}_1, ..., \mathbf{s}_N\}$, $\mathbf{s}_i \in \mathcal{D}$, $N \geq 1$.

In the above definitions, nothing is said about the relationship between the potential $q(\mathbf{s})$ and the conditional potential $q(\mathbf{s} \,|\mathcal{S})$ so that, in general, they may differ. Moreover, the locations $\mathcal{S}$ are not constrained to be different, neither between each other nor from $\mathbf{s}$.

From a probabilistic point of view, the potential $q(\mathbf{s})$ is modelled by a real-valued spatial random field $\{q(\mathbf{s})\}$ indexed by $\mathbf{s} \in \mathcal{D}$. Given an equilibrium state π and a realization q_π^0 of the potential, the aim is to recover q_π^0 in terms of its spatial surface over $\mathcal{D}$. Following the geostatistical approach, the potential is measured

at $\mathcal{S}$ and the measurements $\mathbf{y}(\mathcal{S}) = (y(\mathbf{s}_1), ..., y(\mathbf{s}_N))$ are analyzed in order to obtain an estimate $\hat{q}$ of q_π^0. As stated in the Definition 2, the N measurements are supposed to be concurrent and instantaneous. If not, additional considerations are to be made.

2.2 Opaqueness and transparency

Given a system $\mathcal{O}$ and a measuring system $\mathcal{M}$, I say that $\mathcal{O}$ is *opaque* with respect to $\mathcal{M}$ if, for each $\mathcal{S} \subset \mathcal{D}$, there exists a region $\mathcal{D}_0 \subset \mathcal{D}$ such that $q(\mathbf{s}) \neq q(\mathbf{s} \,|\mathcal{S})$ for each $\mathbf{s} \in \mathcal{D}_0$, that is, the action of measuring $\mathcal{O}$ at $\mathcal{S}$ alters the measure of q in a region $\mathcal{D}_0$ of non-zero measure. On the other hand, if $q(\mathbf{s}) = q(\mathbf{s} \,|\mathcal{S})$ for each $\mathbf{s} \in \mathcal{D}$ and each $\mathcal{S} \subset \mathcal{D}$, then $\mathcal{O}$ is said to be *transparent* with respect to $\mathcal{M}$. Given a particular choice of $\mathcal{D}$ and $\mathcal{S}$, the *opaqueness* of $\mathcal{O}$ with respect to $\mathcal{M}$ is measured by the following quantity

$$\int_{\mathcal{D}} \left(q(\mathbf{s}) - q(\mathbf{s} \,|\mathcal{S})\right)^2 d\mathbf{s}$$

which is equal to zero when $\mathcal{O}$ is transparent with respect to $\mathcal{M}$. It is important to note that a generic $\mathcal{O}$ can be transparent with respect to $\mathcal{M}$ but opaque with respect to a different measuring system $\mathcal{M}'$. Since the potential q represents what can be observed about $\mathcal{O}$, then q inherits the property of $\mathcal{O}$ so that q can be either opaque or transparent with respect to $\mathcal{M}$.

Opaque potentials are in contrast with most applications where spatial random fields are considered. As an example, consider the measurement of the air temperature in a region $\mathcal{D}$ through a thermometer $m(\mathbf{s})$ located at $\mathbf{s}$. When the air temperature is measured at $\mathbf{s}$, it is supposed that the action of measuring the air temperature does not modify the air temperature at and nearby $\mathbf{s}$. Actually, if the initial temperature of the measuring instrument is different from the air temperature at $\mathbf{s}$, then the instrument alters the air temperature (first locally and

then globally). In fact, what is read by the instrument (after some time) is a sort of mass-weighted average temperature between the air and the instrument temperature. If a second thermometer $m(\mathbf{s}')$ is placed at $\mathbf{s}'$, then it can be said that what is read by $m(\mathbf{s}')$ is influenced by the presence of $m(\mathbf{s})$. In the same manner, $m(\mathbf{s}')$ "feels" the presence of $m(\mathbf{s})$. When compared to the air mass, however, the mass of the measuring instruments is so small that the alteration of the air temperature is negligible, that is, much smaller than the instrument error. Thus, for all practical purposes, the random field q representing the temperature field is transparent with respect to $\mathcal{M}$.

Depending on the particular case, the interaction between $\mathcal{O}$ and $\mathcal{M}$ can affect or not the equilibrium state π of $\mathcal{O}$. In the air temperature example, for instance, the system reaches a different equilibrium state π' after the system is measured. Since the transformation is irreversible, the realization of q^0_π related with π is lost and a new realization $q^0_{\pi'}$ comes in. Even in the case q^0_π can be recovered from the measurements $\mathbf{y}(\mathcal{S})$, q^0_π no longer represents the current state of the system. In general, this kind of interactions are not desired. If measuring $\mathcal{O}$ alters its state in a significant way, either we are dealing with some kind of quantum system or we are considering an unsuitable measuring system. Instead, in this work, I am to consider interactions that do not affect the equilibrium state π and hence the realization q^0_π.

2.3 The discrete nature of systems

Ultimately, the opaqueness of a potential q is related to the discrete nature of the system $\mathcal{O}$, which is in contrast with the continuous model represented by q. In the air temperature example, q is just a model of π in the sense that q does not exist physically. In fact, the air temperature is related to the kinetic energy of the air molecules. When the thermometers come into play, the air molecules interact with

the molecules of the thermometers, bringing $\mathcal{O}$ to the new state π'. Indeed, we can say that the molecules are the mediators of the interaction between $\mathcal{O}$ and $\mathcal{M}$. Again, the number of molecules of $\mathcal{O}$ is much higher than the number of molecules of $\mathcal{M}$ and the interaction is negligible.

In the sales data example of the previous section, on the other hand, the interaction is more relevant. The market potential q is, by definition, continuous in space but, in a way, it is measured through the people that reach the stores, with the stores representing the measuring system $\mathcal{M}$ and the people the mediators. In contrast to the air temperature example, the mediators that interact with $\mathcal{M}$ represent a relevant part of the set $\mathcal{P}$ of all the mediators. Indeed, in practice, the number $N_{\mathcal{P}}$ of elements of $\mathcal{P}$ may be small. The opaqueness of a system $\mathcal{O}$ is directly related to $N_{\mathcal{P}}$ and the smaller the number of mediators the higher the degree of opaqueness of $\mathcal{O}$ with respect to $\mathcal{M}$.

When an interaction between $\mathcal{O}$ and $\mathcal{M}$ is needed to measure $\mathcal{O}$, depending on the relationship between $N_{\mathcal{P}}$ and N, three important cases can be distinguished:

- $N_{\mathcal{P}} \gg N$: the interaction between $\mathcal{O}$ and $\mathcal{M}$ can be neglected.
- $N_{\mathcal{P}} \not\ll N$: the interaction between $\mathcal{O}$ and $\mathcal{M}$ has to be taken into account.
- $N_{\mathcal{P}} < N$: the system $\mathcal{O}$ cannot be appropriately measured. The measures collected by $\mathcal{M}$ do not reflect the potential q and hence the state π of the system $\mathcal{O}$.

The sales data example helps explain in which sense the attraction between the measuring instruments (the stores) and the mediators (the people) does not change the state π of the system $\mathcal{O}$ (the market potential) and why the condition $N_{\mathcal{P}} \not\ll N$ should be satisfied. Clearly, people are attracted by the stores and people move to them. As a consequence, the market potential seems to arise where the stores are

and not where the market potential really is (for instance where the people live). If the same stores are moved at some new locations in space, a different market potential is measured so that it may seem that the market potential depends on $\mathcal{M}$. Actually, the market potential exists beside $\mathcal{M}$ and the state π is altered neither by the presence of $\mathcal{M}$ nor by the movement of the mediators. In fact, the market potential measured at $\mathbf{s}$ is also a measure of the willingness of the mediators to reach the location $\mathbf{s}$, which is inherent in π. Indeed, $\mathcal{M}$ can only affect the estimation of the market potential in the sense that, as in every other geostatistical application, the way the market potential is appropriately estimated over $\mathcal{D}$ depends on N and the spatial pattern of the measuring instruments. In a classic geostatistical application, the higher the N the better q is estimated. Since $N_{\mathcal{P}}$ is finite, however, q is not necessarily better estimated by increasing N, in particular if the condition $N_{\mathcal{P}} < N$ is reached (more stores than people). Note that, even if $N_{\mathcal{P}}$ is very small, the market potential is still well defined for each $\mathbf{s} \in \mathcal{D}$ and it is still well described by a spatially continuous (maybe discrete value) random field. If the number of measuring instruments increases toward infinity, however, the probability that a given measuring instrument m does not interact with any mediator increases as well. As a consequence, the observed market potential will be different from zero only over a finite set of points of $\mathcal{D}$. This cannot happen, for example, in the measure of the air temperature where the number of thermometers can be increased as desired, though, at some point, the measured temperature will be the temperature of the joint system air-thermometers rather than the air temperature.

2.4 Interaction types

Although many kind of interaction could be conceived and studied, the following types of interactions are defined:

- *absorption interaction*: the potential $q(\mathbf{s})$ is absorbed by the measuring system $\mathcal{M}$ if

$$\forall \mathcal{S} \subset \mathcal{D}, \exists \mathcal{D}_0 : q(\mathbf{s} \,|\mathcal{S}) < q(\mathbf{s}), \forall \mathbf{s} \in \mathcal{D}_0$$

that is, the measurable potential at the location $\mathbf{s}$ is lower if the potential is already measured at $\mathcal{S}$.

The sales data example falls in this category though many others can be considered. A common problem in assessing the impact of an airborne pollutant on the population health is to find a relationship between the spatial distribution of the pollutant concentration and the spatial occurrence of respiratory diseases (see [4] and [5]). While it is relatively simple to estimate the spatial distribution of the pollutant concentration ([6]), the same cannot be said about estimating the spatial incidence of the respiratory diseases. This is especially true when, for privacy reasons, the available data are aggregated; for instance, only the monthly/yearly number of patients with respiratory disease for each family doctor could be available. Note that, in this case, there is no correlation between the distribution of the pollutant concentration and the spatial locations where the doctors practice and the use of the ILGC model defined in (2-3) would be inappropriate. Although with some exceptions, it is common that people have their own family doctor near the place where they live or work, that is, near the location where they have been possibly exposed to the pollutant. Of course, the interaction between the family doctors and their patients is of the absorption type since a patient has no more than one family doctor. Given the number of patients with respiratory disease for each doctor, it might be useful to estimate $q(\mathbf{s})$ in order to compare, one-to-one, the spatial distribution of the pollutant concentration with the potential $q(\mathbf{s})$, intended as the expected number of people with a respiratory disease that would be visited by a doctor if the

doctor was located at $\mathbf{s}$.

- *alteration interaction*: the potential $q(\mathbf{s})$ is altered by the measuring system $\mathcal{M}$ if

$$\forall \mathcal{S} \subset \mathcal{D}, \exists \mathcal{D}_0 : q(\mathbf{s} \,|\mathcal{S}) \neq q(\mathbf{s}), \forall \mathbf{s} \in \mathcal{D}_0$$

  This is the case of the air temperature example where the temperature alteration due to the presence of the thermometers depends on the initial temperature of the thermometers.

- *reflection interaction*: the potential $q(\mathbf{s})$ is reflected by the measuring system $\mathcal{M}$ if

$$\forall \mathcal{S} \subset \mathcal{D}, \exists \mathcal{D}_0 : q(\mathbf{s} \,|\mathcal{S}) > q(\mathbf{s}), \forall \mathbf{s} \in \mathcal{D}_0$$

  As an example, consider the yearly average income of the pubs of a large city, Dublin for instance. Fleet street is a famous street of Dublin, the popularity of which is mainly due to the high number of pubs and restaurants that are destination for many Dubliners and tourists. The pubs do not act like absorbers in the sense that the typical tourist visits more than one pub during the same night. As a consequence, the income of a pub in fleet street is higher than the income the same pub would experience if it was the only pub in the street. On the contrary, a pub far from fleet street and far from other pubs might not experience a high income even if located in a highly populated area. Indeed, due to the customer behavior, it might be worthwhile to open a new pub a few meters away from an existing pub rather than far from it.

2.5 Point aggregated data

To all intents, the data of the previous examples can be considered a special type of aggregate data. When describing the role of the spatial statistics in epidemiological

studies, [7] identifies two main types of data: point data representing the exact locations where people with a particular disease live and aggregated data describing the number of ill people within each sub-region of a studied area. In the latter case, the data are related with the sub-regions as a whole rather than to points in space and any inferential problem should be solved with the statistical methods for areal data (see [8]).

The epidemiological data considered in the example of the previous paragraph (number of patients with respiratory diseases for each family doctor), on the other hand, can be associated with a third type of data that I here call *point aggregated data*. Indeed, the locations where the doctors practice correspond to precise points in space and this information should be retained when the data are analyzed. Although point aggregated data can be seen as the realization of a MPP, they are conceptually different from classic MPP data where the mark corresponds to a feature measured on a single entity (for instance, the diameter of a tree in a forest).

That said, the general problem addressed in this paper might be restated as the spatial modelling of point aggregated data, without invoking the interaction between the observed system and the measuring system. To pose the problem in terms of interaction, however, allows to explain the data generation mechanism of the point aggregated data and it will be useful in defining the functional and parametric form of the new class of geostatistical models. Without loss of generality, in order to keep the notation simple, the case $d = 2$ will be considered.

3 The geostatistical potential model

In this section, the geostatistical potential model (GPM) is introduced as the main statistical tool for the analysis of spatial data with the characteristic reported in

the list of paragraph 1.2. In its general form, the GPM is described by the following equations

$$\begin{aligned} y(\mathbf{s}|\mathcal{S}) &= h_{\boldsymbol{\xi}}\left(u\left(\mathbf{s}\right);\mathcal{M}\left(\mathcal{S}\right)\right) \\ u\left(\mathbf{s}\right) &= q(\mathbf{s})+\varepsilon(\mathbf{s}) \\ q(\mathbf{s}) &= \mu+\mathbf{x}(\mathbf{s})\boldsymbol{\beta}+\gamma w(\mathbf{s}) \end{aligned} \tag{4}$$

At the first stage of (4), $h_{\boldsymbol{\xi}}:\mathbb{R}\longrightarrow\mathbb{R}$ is the *interaction function* which is parametrized by the parameter vector $\boldsymbol{\xi}$. The interaction function defines how the measuring system $\mathcal{M}\left(\mathcal{S}\right)$ interacts with the potential and, depending on its analytical form, the interaction can be classified as absorption, alteration or reflection. At the second stage, $\varepsilon(\mathbf{s})$ represents an error component which is assumed to be *i.i.d.* $N(0,\sigma_{\varepsilon}^{2})$ and is supposed to capture both the measuring error and the model error. Finally, at the third stage, the potential $q(\mathbf{s})$ is modelled by three terms, where μ is the mean, $\mathbf{x}(\mathbf{s})$ is a vector of covariates, $\boldsymbol{\beta}$ is the vector of coefficient, $w(\mathbf{s})$ is a zero-mean latent Gaussian process and γ is a scale parameter. The covariance function of $w(\mathbf{s})$ is $cov\left(w(\mathbf{s}),w(\mathbf{s}')\right)=\rho_{\boldsymbol{\theta}}\left(\mathbf{s},\mathbf{s}'\right)$, with $\rho_{\boldsymbol{\theta}}\left(\mathbf{s},\mathbf{s}'\right)$ a valid correlation function parametrized by the vector $\boldsymbol{\theta}$. Note that the potential is allowed to be negative though it can be easily constrained to be positive if necessary. The model parameter vector is $\Psi=(\mu,\boldsymbol{\beta}',\sigma_{\varepsilon}^{2},\gamma,\boldsymbol{\theta}',\boldsymbol{\xi}')$.

In order to have a better insight into the role of the interaction function $h_{\boldsymbol{\xi}}$, the following family of interaction functions is adopted

$$h_{\boldsymbol{\xi}}\left(u\left(\mathbf{s}\right);\mathcal{S}\right)=u\left(\mathbf{s}\right)\cdot\left(1+\sum_{\mathbf{s}'\in\mathcal{S}}f_{\boldsymbol{\vartheta}}\left(\mathbf{s},\mathbf{s}'\right)\right)^{\delta}=u\left(\mathbf{s}\right)\cdot g_{\boldsymbol{\xi}}(\mathbf{s}\mid\mathcal{S}) \tag{5}$$

where $f_{\boldsymbol{\vartheta}}\left(\mathbf{s},\mathbf{s}'\right):\mathbb{R}^{d}\times\mathbb{R}^{d}\longrightarrow\mathbb{R}^{+}$ is a generic non-negative binary function and $\boldsymbol{\xi}=(\boldsymbol{\vartheta},\delta)'$. The parameter $\delta\in\{-1,0,1\}$ defines the interaction type and in particular $\delta=-1$ corresponds to an absorption interaction, $\delta=0$ corresponds to no interaction (the system $\mathcal{O}$ is transparent with respect to $\mathcal{M}$) while $\delta=1$ corresponds to a reflection interaction. In this work, the value of δ is not estimated

since the interaction type is supposed to be known from the particular case the GPM is applied to.

The function $f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}')$ can be any continuous function but, for practical applications, it should be monotonic and such that

$$\begin{aligned} \lim_{\|\mathbf{s}-\mathbf{s}'\|\to 0} f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}') &= 1 \\ \lim_{\|\mathbf{s}-\mathbf{s}'\|\to \infty} f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}') &= 0 \end{aligned}$$

For instance,

$$f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}') = f_{\boldsymbol{\vartheta}}(\|\mathbf{s}-\mathbf{s}'\|) = exp\left(-\frac{\|\mathbf{s}-\mathbf{s}'\|}{\phi}\right)^{\alpha} \tag{6}$$

where $\|\cdot\|$ is the Euclidean distance, $\phi, \alpha > 0$ are the function parameters and $\boldsymbol{\vartheta} = (\phi, \alpha)'$ is the parameter vector. In (6), ϕ define the strength of the interaction while α is a shape parameter.

Note that

$$\begin{aligned} y(\mathbf{s}|\mathcal{S}) &= u(\mathbf{s}) \cdot g_{\boldsymbol{\xi}}(\mathbf{s} \mid \mathcal{S}) \\ &= [q(\mathbf{s}) + \varepsilon(\mathbf{s})] \cdot g_{\boldsymbol{\xi}}(\mathbf{s} \mid \mathcal{S}) \\ &= q(\mathbf{s} \mid \mathcal{S}) + \varepsilon(\mathbf{s} \mid \mathcal{S}) \end{aligned}$$

namely the observed potential is equal to the conditional potential $q(\mathbf{s} \mid \mathcal{S})$ plus a transformation of the error $\varepsilon(\mathbf{s})$.

The term $g_{\boldsymbol{\xi}}(\mathbf{s} \mid \mathcal{S}) \equiv g_{\boldsymbol{\xi}}(\mathbf{s})$ is the key aspect of the interaction function and deserves more explanation. If, as an example, the function (6) is considered within the absorption interaction case ($\delta = -1$) and $\mathcal{S} \equiv \emptyset$, namely if there are no measuring instrument, then $g_{\boldsymbol{\xi}}(\mathbf{s}) = 1$ since the summand in (5) cannot be evaluated and it is equal to zero by definition. When a measuring instrument is added, $\mathcal{S} = \{\mathbf{s}_1\}$, the measured potential at $\mathbf{s}$ is a function of the distance between $\mathbf{s}$ and $\mathbf{s}_1$. In particular, if $\mathbf{s} = \mathbf{s}_1$ then $g_{\boldsymbol{\xi}}(\mathbf{s}) = 0.5$. On the contrary, if $\|\mathbf{s} - \mathbf{s}_1\| = \infty$ then $g_{\boldsymbol{\xi}}(\mathbf{s}) = 1$. This reflects the fact that the action of absorbing the potential at site $\mathbf{s}_1$ influences the measure at site $\mathbf{s}$. It is worth noting that $\mathbf{s}$ and $\mathbf{s}_1$ are exchangeable in the sense that absorbing and measuring the potential are equivalent

actions and the potential cannot be measured without being absorbed. Indeed, the "virtual" measuring instrument at $\mathbf{s}$ is a sort of test charge that can move freely in $\mathcal{D}$. The moment the measuring instrument is fixed in space, however, it becomes a measuring instrument belonging to the set $\mathcal{M}(\mathcal{S})$. Finally, note that, if $\mathbf{s} = \mathbf{s}_i$ for each $\mathbf{s}_i \in \mathcal{S}$, namely all the $N+1$ measuring instruments are located at $\mathbf{s}$, then

$$y(\mathbf{s}|\mathcal{S}) = \frac{1}{1+N} q(\mathbf{s}) + \frac{1}{1+N} \varepsilon(\mathbf{s})$$

that is, the potential $q(\mathbf{s})$ is equally absorbed by the $N+1$ measuring instruments. The error component $\varepsilon(\mathbf{s})$ undergoes the same transformation so that the variance of the error is high at the spatial locations where the observed potential $q(\mathbf{s}\,|\mathcal{S})$ is expected to be high and it is low where the observed potential is expected to be low.

In this work, the measuring instruments are supposed to satisfy the two following properties:

1. *Property of equi-effectiveness*: two measuring instruments $m(\mathbf{s}_i)$ and $m(\mathbf{s}_j)$ are said to be equally effective if $g(\mathbf{s}_i|\{\mathbf{s}_j\}) = g(\mathbf{s}_j|\{\mathbf{s}_i\})\ \forall \|\mathbf{s}_i - \mathbf{s}_j\|$.

2. *Property of full effectiveness*: the measuring instrument $m(\mathbf{s}_i)$ is fully effective if there is no upper bound on the number of mediators the instrument can interact with.

Property 1 is satisfied if the binary function $f_{\vartheta}(\mathbf{s}, \mathbf{s}')$ is commutative, which is the case of (6). In practical applications, the property may not be satisfied in the sense that a measuring instrument $m(\mathbf{s}_i)$ might be more effective in absorbing the potential than a second absorber $m(\mathbf{s}_j)$ close to it. Property 1, however, simplify the model and any discrepancy from it is accounted by the error term ε. Note that the measure of effectiveness is strictly related to the the measure of attractiveness of the *spatial behavior of consumers models* typical of the geomarketing literature

(see [9]). Property 2 assumes that there are no upper bound to $y(\mathbf{s}|\mathcal{S})$. Going back to the sales data example, the latter property can be restated as "the market fully satisfies the customer demand".

To better understand the notions of potential and conditional potential, the following examples are considered. As a first example, suppose that two measuring instruments $m(\mathbf{s}_1)$ and $m(\mathbf{s}_2)$ are located at $\mathbf{s}_1 = (200, 200)$ and $\mathbf{s}_2 = (800, 800)$. The GPM considered is

$$\begin{aligned} y(\mathbf{s}|\mathcal{S}) &= u(\mathbf{s}) \cdot g_{\boldsymbol{\xi}}(\mathbf{s} \mid \mathcal{S}) = q(\mathbf{s} \mid \mathcal{S}) \\ u(\mathbf{s}) &= q(\mathbf{s}) \\ q(\mathbf{s}) &= w(\mathbf{s}) \end{aligned} \tag{7}$$

namely it is supposed that the conditional potential is observed without error. Furthermore, suppose that

$$\rho_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{s}') = \rho_{\theta}(\|\mathbf{s} - \mathbf{s}'\|) = exp\left(-\frac{\|\mathbf{s} - \mathbf{s}'\|}{800}\right) \tag{8}$$

$$f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}') = f_{\vartheta}(\|\mathbf{s} - \mathbf{s}'\|) = exp\left(-\frac{\|\mathbf{s} - \mathbf{s}'\|}{300}\right) \tag{9}$$

that $\delta = -1$ (absorption interaction) and that $y(\mathbf{s}_1|\mathcal{S} \setminus \mathbf{s}_1) = y(\mathbf{s}_2|\mathcal{S} \setminus \mathbf{s}_2) = 10$. The potential and the conditional potential for this example are reported in the left and in the right part of Figure 1 respectively. Regarding the potential, its maximum value is $10.6 > 10$. The measuring instrument $m(\mathbf{s}_1)$ absorbs/measures a lower potential, 10, since a fraction of it is absorbed by $m(\mathbf{s}_2)$ and *vice-versa*. Indeed, the potential $q(\mathbf{s}_1) = 10.6$ would be absorbed by $m(\mathbf{s}_1)$ if $m(\mathbf{s}_2)$ were not present. The conditional potential, as expected, has its maximum value (5.991) halfway between $\mathbf{s}_1$ and $\mathbf{s}_2$ and it decreases approaching the measuring instrument locations. The second example consists of 4 measuring instruments with spatial locations depicted in Figure 2. The model considered is the same of (7) as well as the correlation and the interaction functions defined in (8) and (9) respectively. The measured potential is, again, $y(\mathbf{s}_i|\mathcal{S} \setminus \mathbf{s}_i) = 10$, $i = 1, ..., 4$. In this case, the

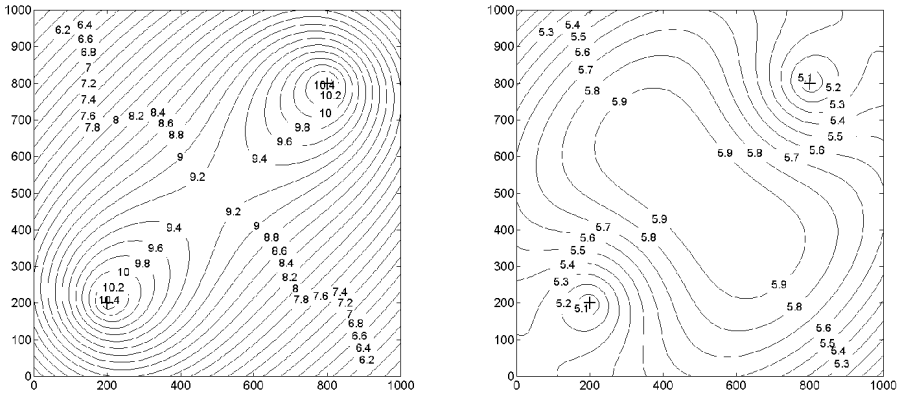

Figure 1: (left) potential $q(\mathbf{s})$; (right) conditional potential $q(\mathbf{s} \mid \mathcal{S})$.

maximum value for the potential is equal to 13.66 and it is attained at the centre of gravity of the measuring instrument locations (left part of Figure 2). Clearly, $y(\mathbf{s}_i|\mathcal{S}\setminus\mathbf{s}_i)$ being fixed, the higher the number of measuring instruments the higher the value of the potential q.

4 Model estimation and inference

Let $\mathbf{y} \equiv \mathbf{y}(\mathcal{S})$ be the $N \times 1$ vector of data collected at the sampling sites $\mathcal{S}$. The measurement equation for the vector $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{w} + \boldsymbol{\varepsilon}) \tag{10}$$

where $\mathbf{1}$ is the $N \times 1$ vector of ones, $\mathbf{X} \equiv \mathbf{X}(\mathcal{S})$ is the $N \times b$ matrix of covariates, $\mathbf{w} \equiv \mathbf{w}(\mathcal{S})$ is the latent Gaussian process at $\mathcal{S}$ with variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{w}} \equiv \boldsymbol{\Sigma}_{\mathbf{w}}(\mathcal{S}, \boldsymbol{\theta})$ and $\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}(\mathcal{S})$ is the measurement error at $\mathcal{S}$ with diagonal variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma_{\varepsilon}^2 I_N$. Finally, $\mathbf{G} \equiv \mathbf{G}_{\boldsymbol{\xi}}(\mathcal{S})$ is the $N \times N$ diagonal matrix whose diagonal vector is

$$\mathbf{g} = (g_{\boldsymbol{\xi}}(\mathbf{s}_1|\mathcal{S}\setminus\mathbf{s}_1), ..., g_{\boldsymbol{\xi}}(\mathbf{s}_N|\mathcal{S}\setminus\mathbf{s}_N))$$

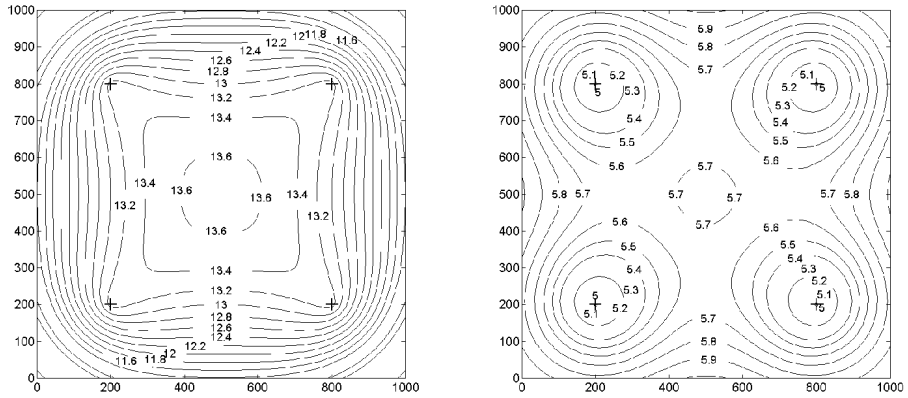

Figure 2: (left) potential $q(\mathbf{s})$; (right) conditional potential $q(\mathbf{s} \mid \mathcal{S})$.

Suppose now that $\mathcal{S}$ is partitioned as $\left\{\mathcal{S}^{(1)}, \mathcal{S}^{(2)}\right\}$, where $\mathcal{S}^{(1)}$ is the set of sites where the data are available and $\mathcal{S}^{(2)}$ is the set of sites where the data are missing. According to this, the vector $\mathbf{y}$ is partitioned in the following way

$$\mathbf{y}^* = \left(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}\right)'$$

where $\mathbf{y}^{(1)} = \mathbf{L}\mathbf{y}$ is the sub-vector of the non-missing data and $\mathbf{L}$ is the appropriate elimination matrix. The vector $\mathbf{y}^*$ is a permutation of $\mathbf{y}$ and $\mathbf{y} = \mathbf{D} \cdot \mathbf{y}^*$, with $\mathbf{D}$ the proper commutation matrix. The partitioned measurement equation become

$$\mathbf{y}^{(i)} = \mathbf{G}^{(i)} \left(\mathbf{1}^{(i)}\mu + \mathbf{X}^{(i)}\boldsymbol{\beta} + \gamma \mathbf{w}^{(i)} + \boldsymbol{\varepsilon}^{(i)}\right);\ i = 1, 2$$

and the variance-covariance matrix of the permuted errors is conformably partitioned as

$$Var\left[\left(\boldsymbol{\varepsilon}^{(1)}, \boldsymbol{\varepsilon}^{(2)}\right)'\right] = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}$$

In the sequel, given $\mathbf{b}$ a generic vector and $\mathbf{B}$ a generic matrix, $\mathbf{b}^{(1)}$ and $\mathbf{B}^{(1)}$ will stand for $\mathbf{L}\mathbf{b}$ and $\mathbf{L}\mathbf{B}\mathbf{L}'$ respectively, bearing in mind that, in general $\mathbf{L}\mathbf{B}^{-1}\mathbf{L}' \neq (\mathbf{L}\mathbf{B}\mathbf{L}')^{-1}$.

Given the data vector $\mathbf{y}$ and considering the GPM defined in (4), the following inferential problems are of interest:

1. provide an estimate of the model parameter vector Ψ;
2. provide confidence intervals for the elements of $\hat{\Psi}$;
3. estimate the potential $q(\mathbf{s})$ over the region $\mathcal{D}$ and its uncertainty;
4. estimate the conditional potential $q(\mathbf{s} \,|\mathcal{S})$ over the region $\mathcal{D}$ and its uncertainty;

4.1 Model estimation

Problem 1. is tackled here following the maximum likelihood (ML) approach. Being $w(\mathbf{s})$ a latent process and due to possible missing data, the expectation-maximization (EM) algorithm is adopted to find the ML estimate $\hat{\Psi}$ of Ψ.

The EM algorithm is based on the complete-data likelihood function $L_{\Psi}(\mathbf{y}, \mathbf{w})$ and it provides an iterative procedure to update the model parameter estimate from $\hat{\Psi}^{(k)}$ to $\hat{\Psi}^{(k+1)}$ until convergence. In particular, for each iteration of the algorithm, the E-step computes the conditional expectation

$$Q\left(\Psi, \hat{\Psi}^{(k)}\right) = E_{\hat{\Psi}^{(k)}}\left[L_{\Psi}(\mathbf{y}, \mathbf{w}) \mid \mathbf{y}^{(1)}\right]$$

while, at the M-step, the following maximization is carried out

$$\hat{\Psi}^{(k+1)} = \arg\max_{\Psi} Q\left(\Psi, \hat{\Psi}^{(k)}\right)$$

which is equivalent to solve the equation

$$\frac{\partial Q\left(\Psi, \hat{\Psi}^{(k)}\right)}{\partial \Psi} = \mathbf{0} \tag{11}$$

Closed form updating formulas are provided for $\mu, \boldsymbol{\beta}, \sigma_\varepsilon^2$ and γ while the remaining model parameters are updated by numerical optimization, namely

$$\left(\hat{\boldsymbol{\theta}}^{(k+1)}, \hat{\boldsymbol{\xi}}^{(k+1)}\right) = \arg\max_{\boldsymbol{\theta},\boldsymbol{\xi}} Q\left(\Psi, \left(\hat{\mu}^{(k+1)}, \hat{\boldsymbol{\beta}}^{(k+1)}, \left(\hat{\sigma}_\varepsilon^2\right)^{(k+1)}, \hat{\gamma}^{(k+1)}, \hat{\boldsymbol{\theta}}^{(k)}, \hat{\boldsymbol{\xi}}^{(k)}\right)\right) \tag{12}$$

Note that, if both the correlation function $\rho_{\boldsymbol{\theta}}$ and the interaction function $h_{\boldsymbol{\xi}}$ have analytical form of the first and second derivative with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ respectively, the update defined in (12) can be carried out by adapting the updating algorithm given in [10] (Proposition 4.4).

The closed form updating formulas are

$$\hat{\mu}^{(k+1)} = \frac{\operatorname{tr}\left[\left(\hat{\mathbf{e}}^{(1)} + \mu^{(k)}\mathbf{1}^{(1)}\right)\left(\mathbf{1}^{(1)}\right)'\right]}{N - N_m} \tag{13}$$

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \left[\left(\mathbf{X}^{(1)}\right)'\mathbf{X}^{(1)}\right]^{-1}\left(\mathbf{X}^{(1)}\right)' \cdot \left(\hat{\mathbf{e}}^{(1)} + \mathbf{X}^{(1)}\boldsymbol{\beta}^{(k)}\right) \tag{14}$$

$$\left(\hat{\sigma}_\varepsilon^2\right)^{(k+1)} = \frac{1}{N}\operatorname{tr}\begin{pmatrix} \hat{\mathbf{e}}^{(1)} \cdot \left(\hat{\mathbf{e}}^{(1)}\right)' + \left(\gamma^{(k)}\right)^2 \hat{\mathbf{A}}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} \end{pmatrix} \tag{15}$$

$$\hat{\gamma}^{(k+1)} = \frac{\operatorname{tr}\left[\left(\hat{\mathbf{e}}^{(1)} + \gamma^{(k)}\hat{\mathbf{w}}^{(1)}\right)\left(\hat{\mathbf{w}}^{(1)}\right)'\right]}{\operatorname{tr}\left[\hat{\mathbf{w}}^{(1)}\left(\hat{\mathbf{w}}^{(1)}\right)' + \hat{\mathbf{A}}^{(1)}\right]} \tag{16}$$

where $\hat{\mathbf{e}}^{(1)} = \left(\left(\mathbf{G}^{(1)}\right)^{-1}\mathbf{y}^{(1)} - \mu^{(k)}\mathbf{1}^{(1)} - \mathbf{X}^{(1)}\boldsymbol{\beta}^{(k)} - \gamma^{(k)}\hat{\mathbf{w}}^{(1)}\right)$, N_m is the number of missing data and

$$\hat{\mathbf{w}} = E_{\Psi^{(k)}}\left(\mathbf{w} \mid \mathbf{y}^{(1)}\right) \tag{17}$$

$$\hat{\mathbf{A}} = Var_{\Psi^{(k)}}\left(\mathbf{w} \mid \mathbf{y}^{(1)}\right) \tag{18}$$

are the estimated latent variable and the estimation variance, respectively. The proof for the updating formulas (13-16) is reported in Appendix A while the evaluation of (17) and (18) is reported in Appendix B.

4.2 Parameter confidence intervals

As known, the EM algorithm does not provide information about the uncertainty of the estimated parameter vector $\hat{\Psi}$. In this paragraph, two methods are proposed to solve problem 2. of the above list, namely to provide confidence intervals for the elements of $\hat{\Psi}$.

The first method is based on the fact that the maximum likelihood estimator has asymptotically normal distribution $N(\Psi_0, \mathbf{I}^{-1})$, with Ψ_0 the "true" value of Ψ and $\mathbf{I}$ the Fisher information matrix. An approximation of the information matrix for multivariate normal variables can be evaluated as

$$\begin{aligned} \tilde{\mathbf{I}}_{ij} &= \partial_i \boldsymbol{\epsilon}' \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_j \boldsymbol{\epsilon} + \frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_j \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\right) \\ &\quad + \frac{1}{4} \operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_j \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\right) \end{aligned} \tag{19}$$

(see [11]), where $\partial_i \boldsymbol{\epsilon}$ and $\partial_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ are short notations for $\partial \boldsymbol{\epsilon}(\Psi) / \partial \Psi_i$ and $\partial \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\Psi) / \partial \Psi_i$ respectively and $1 \leq i, j \leq |\Psi|$.

In the case of the GPM, the following vector

$$\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta}) \tag{20}$$

is normally distributed with variance-covariance matrix

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} &= \operatorname{Var}(\mathbf{y} - \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta})) && (21) \\ &= \operatorname{Var}(\gamma \mathbf{G}\mathbf{w} + \mathbf{G}\boldsymbol{\varepsilon}) \\ &= \mathbf{G}\left(\gamma^2 \boldsymbol{\Sigma}_{\mathbf{w}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\right) \mathbf{G}' && (22) \\ &= \mathbf{g}\mathbf{g}' \odot \left(\gamma^2 \boldsymbol{\Sigma}_{\mathbf{w}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\right) \end{aligned}$$

where $\odot$ is the Hadamard product operator. The solution for the derivatives $\partial_i \boldsymbol{\epsilon}$ and $\partial_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is reported in Appendix C. In the presence of missing data, (19) is still valid but $\boldsymbol{\epsilon}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ have to be replaced with $\boldsymbol{\epsilon}^{(1)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{(1)}$ respectively.

With $\tilde{\mathbf{I}}$ available, approximated confidence intervals for the elements of $\hat{\Psi}$ are immediately provided by considering $N\left(\hat{\Psi}, \tilde{\mathbf{I}}^{-1}\right)$. Note, however, that $N\left(\hat{\Psi}, \tilde{\mathbf{I}}^{-1}\right)$ is a good approximation of the distribution $[\Psi \mid \mathbf{y}(\mathcal{S})]$ only when N is large, which may not be the case in practical applications. In particular, the marginal distributions on the elements of Ψ are not Normal for small N. To solve this problem, a second method based on Monte Carlo simulation is considered. Although computationally more demanding, the method can be adopted even for small values of N and the accuracy of the result is directly related to the number M of simulation runs.

Let $\hat{\Psi}$ be the estimated parameter vector. For each simulation run m, the vector $\mathbf{y}_{(m)} = \mathbf{D}\left[\begin{array}{cc} \mathbf{y}_{(m)}^{(1)} & \mathbf{y}^{(2)} \end{array}\right]'$ is considered, where

$$\mathbf{y}_{(m)}^{(1)} = \mathbf{L} \cdot \mathbf{G}_{\hat{\boldsymbol{\xi}}}(\mathcal{S}) \cdot \left(\mathbf{1}\hat{\mu} + \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\gamma}\tilde{\mathbf{w}} + \tilde{\boldsymbol{\varepsilon}}\right)$$

and $\tilde{\mathbf{w}}$ and $\tilde{\boldsymbol{\varepsilon}}$ are realizations from the multivariate normal distributions $N\left(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\left(\hat{\sigma}_{\varepsilon}^{2}\right)\right)$ and $N\left(0, \boldsymbol{\Sigma}_{\mathbf{w}}\left(\hat{\boldsymbol{\theta}}\right)\right)$ respectively. Note that $\mathbf{y}_{(m)}$ preserves the missing data pattern of the observed $\mathbf{y}$. The simulated $\mathbf{y}_{(m)}$ is used to estimate a new parameter vector $\hat{\Psi}_{(m)}$ through the EM algorithm and the set

$$\hat{\Psi}_s = \left\{\hat{\Psi}_{(1)}, ..., \hat{\Psi}_{(M)}\right\} \tag{23}$$

is considered as a sample from the distribution $[\Psi \mid \mathbf{y}(\mathcal{S})]$. If M is large enough, then $\hat{\Psi}_s$ can be analyzed to derive approximated confidence intervals for the elements of $\hat{\Psi}$. Note that the latter method requires to run the EM algorithm M times while the former method only requires to evaluate (19).

4.3 Potential estimation

The potential $q(\mathbf{s})$ represents one of the main result of the data analysis based on the GPM. Following a plug-in approach, the estimated potential is obtained as

$$q_{\hat{\Psi}}(\mathbf{s}) = \hat{\mu} + \mathbf{x}(\mathbf{s})\,\hat{\boldsymbol{\beta}} + \hat{\gamma}\hat{\mathbf{w}}(\mathbf{s}),\ \mathbf{s} \in \mathbb{R}^2 \tag{24}$$

where $\hat{\mathbf{w}}(\mathbf{s}) = E_{\hat{\Psi}}(\mathbf{w}(\mathbf{s}) \mid \mathbf{y})$ is the kriging estimate of $\mathbf{w}(\mathbf{s})$, which is evaluated analogously to $\hat{\mathbf{w}}$ in().

It is worth noting that $q_{\hat{\Psi}}(\cdot)$ does not involve the matrix $\mathbf{G}$ since, as already mentioned, the potential $q_{\hat{\Psi}}(\mathbf{s})$ represents the potential that would be observed by an absorber located at $\mathbf{s}$ if no other absorber was present in $\mathcal{D}$.

The uncertainty of $q_{\hat{\Psi}}(\mathbf{s})$ is directly related to the uncertainty of $\hat{\Psi}$ which is expressed by $[\Psi \mid \mathbf{y}(\mathcal{S})]$. Again, approximated confidence intervals on $q_{\hat{\Psi}}(\mathbf{s})$ can be provided by repeatedly estimating $q_{\Psi}(\mathbf{s})$ with Ψ extracted either from $N\left(\hat{\Psi}, \tilde{\mathbf{I}}^{-1}\right)$ or from the set $\hat{\Psi}_s$ of (23).

4.4 Conditional potential estimation

The estimated conditional potential is simply given by

$$q_{\hat{\Psi}}(\mathbf{s} \mid \mathcal{S}) = q_{\hat{\Psi}}(\mathbf{s}) \cdot g_{\hat{\boldsymbol{\xi}}}(\mathbf{s} \mid \mathcal{S})$$

with $\mathcal{S}$, as always, the set of locations of the absorbers. Approximated confidence intervals on $q_{\hat{\Psi}}(\mathbf{s} \mid \mathcal{S})$ are provided following the same approach defined in Paragraph 4.3 for $q_{\hat{\Psi}}(\mathbf{s})$.

5 Case study

The GPM introduced in Section 3 is applied here to sales data of a daily newspaper for the city of Bergamo, northern Italy. The data represent the yearly average daily number of copies sold during the working days by $N = 75$ newsstands located over the Bergamo territory. The newsstand spatial locations are showed in Figure 3 along with their code and the average daily number of copies sold. The sales data of 5 newsstands are unavailable though their location is known.

By considering the theory of Section 2, it can be said that the system $\mathcal{O}$ is represented by the market potential, the measuring system $\mathcal{M}$ by the newsstands while the mediators by the customers. In particular, the population $\mathcal{P}$ of the mediators is represented by the people, living in or visiting Bergamo, that are potential buyers of the daily newspaper. The interaction between $\mathcal{O}$ and $\mathcal{M}$ is clearly of the absorption type. In fact, once the customer has bought a copy of the newspaper, it is absorbed in the sense that the same customer will not buy (during the same day) the same copy of the newspaper, neither at the same nor at a different newsstand. Moreover, the interaction between $\mathcal{O}$ and $\mathcal{M}$ does not change the state π of $\mathcal{O}$ since the measurement (the newsstands selling the newspaper copies to the customers) does not change, obviously, the customer behavior or the customer interest toward the newspaper. The market potential $\mathcal{O}$ is modelled by a potential $q(\mathbf{s})$ and in particular by a GPM. The choice of modelling $\mathcal{O}$ through a continuous spatial process makes sense since, for each location in space $\mathbf{s}$, the market potential, intended as the number of copies that could be sold by placing a newsstand at $\mathbf{s}$, is well defined for each $\mathbf{s} \in \mathcal{D}$.

From the point of view of the data analysis, the aim of the study is twofold: to estimate the market potential for the newspaper and to identify profitable spatial locations where new newsstands could be opened. In particular, the following questions are to be answered

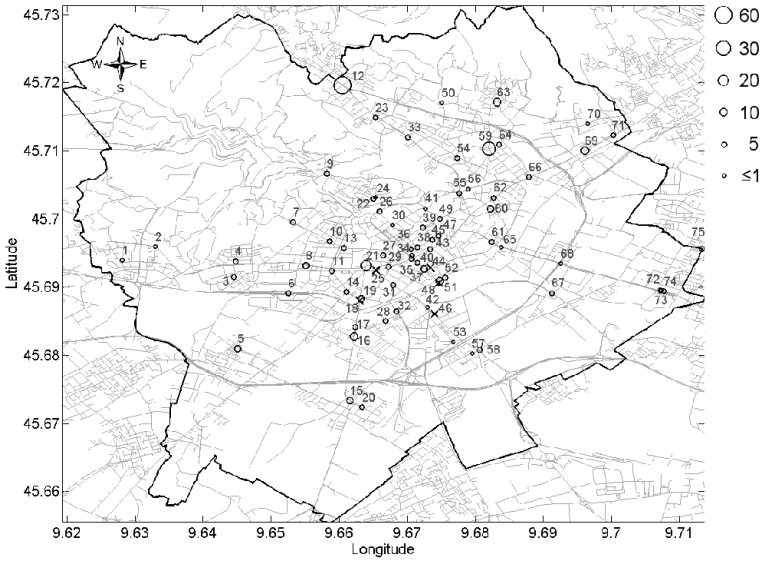

Figure 3: Newsstand locations and circle plot of the working day average daily number of copies sold.

- Which are the areas of the city with the highest market potential?
- Which are the areas of the city where it would be profitable to open a newsstand?
- What is the uncertainty of these areas in terms of potential and spatial location?

In this case, the classic geostatistical approach cannot be adopted since the sales data, as they are, cannot be considered as observations of a continuous phenomenon. Indeed, it does not make sense to estimate the number of copies sold outside the 75 newsstands. The point process statistics approach can provide information about, on the one hand, the spatial pattern of the newsstands, and on the other, the relationship between the spatial density of the newsstands and the number of copies sold. This kind of information, however, does not help answer

the above questions adequately. Moreover, the missing data problem is not easily solved within this approach.

Instead, the GPM model is adopted as it can provide the needed information by receiving as input the sales data.

The GPM model considered is the same of (4) but with $\mu \equiv 0$. The spatial correlation function of the latent component w is chosen to be

$$\rho_\theta(\mathbf{s}, \mathbf{s}') = exp\left(-\frac{\|\mathbf{s} - \mathbf{s}'\|}{\theta}\right) \tag{25}$$

while (5) is considered as interaction function, with $\delta = -1$ and

$$f_\phi(\mathbf{s}, \mathbf{s}') = exp\left(\frac{-\|\mathbf{s} - \mathbf{s}'\|}{\phi}\right) \tag{26}$$

As it is obviously related to the sold number of copies, the population density for the city of Bergamo, depicted in Figure 4, is considered as a covariate. The model parameter vector Ψ is estimated by means of the EM algorithm as discussed in 4.1. The estimation result is reported in Table 1 with the confidence intervals evaluated by following the simulation approach discussed in 4.2 and $M = 10,000$. Namely, 95% confidence intervals are obtained by evaluating empirical distributions on $\hat{\Psi}_s = \left\{\hat{\Psi}_{(1)}, ..., \hat{\Psi}_{(M)}\right\}$. The empirical variance-covariance matrix of $\hat{\Psi}$ is reported in Table 2 and it can be compared with the approximated Hessian matrix evaluated by (19). The Hessian matrix underestimate the variances related to the elements of $\hat{\Psi}$ and, in this case, it is unsuitable for deriving confidence intervals on $\hat{\Psi}$.

As expected, the $\hat{\beta}$ coefficient related to the population density is positive in sign and significantly different from zero. The value of $\hat{\theta}$ ($\simeq 200m$) suggests that the potential q is spatially correlated at the city-district level. In other words, net of the covariate effect, newsstands belonging to the same district (the city of Bergamo consists of 25 districts) experience a similar potential. As supported by the value of $\hat{\phi}$ ($\simeq 200m$), the competition between nearby newsstands is quite strong.

	$\hat{\beta}$	$\hat{\sigma}_{\varepsilon}^2$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\phi}$
estimated	20.48	26.38	14.73	198.91	200.37
LCL	13.32	1.60	10.70	48.43	164.49
UCL	48.42	162.16	36.36	471.23	469.30

Table 1: Estimated model parameter and 95% confidence interval

	β	σ_{ε}^2	γ	θ	ϕ
β	160.61	245.33	97.42	−78.72	1005.06
σ_{ε}^2		2867.73	138.07	2536.20	1797.01
γ			82.66	−75.32	800.66
θ				13889.89	−682.77
ϕ					8646.81

Table 2: Empirical variance-covariance matrix of $\hat{\Psi}$ based on 10,000 Monte Carlo simulation runs

	β	σ_{ε}^2	γ	θ	ϕ
β	83.66	−2.35	7.12	−65.10	117.01
σ_{ε}^2		2435.75	−93.25	2880.35	−28.92
γ			9.12	−110.67	87.79
θ				11247.41	−802.59
ϕ					1442.53

Table 3: Approximated Hessian matrix for $\hat{\Psi}$

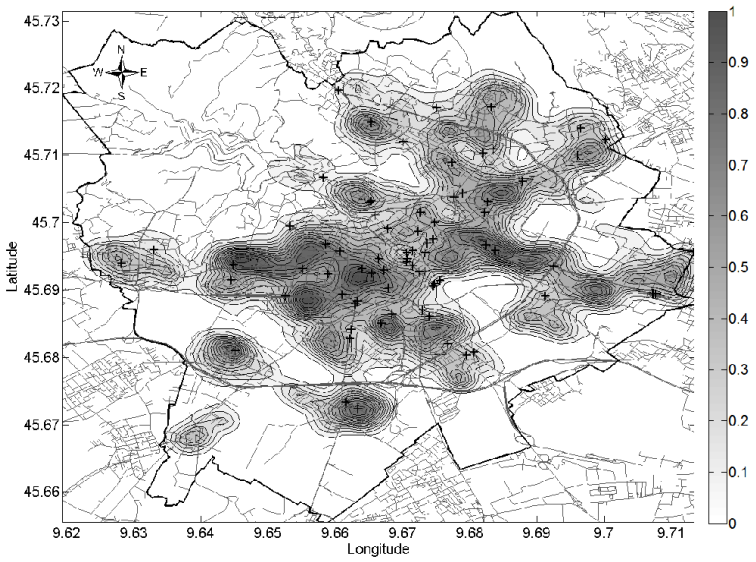

Figure 4: Scaled population density for the area of Bergamo.

Given $\hat{\Psi}$, the potential $q_{\hat{\Psi}}(\mathbf{s})$ is estimated by (24) over the city of Bergamo as depicted in Figure 5. The surface of $q_{\hat{\Psi}}(\mathbf{s})$ is characterized by many maxima the most relevant of which correspond to a potential equal to $\{69.21, 50.04, 49.97, 43.34\}$. This is the daily average newspaper copies that would be sold by a newsstand if placed at these maxima without other newsstands nearby. The conditional potential $q_{\hat{\Psi}}(\mathbf{s} \mid \mathcal{S})$, on the other hand, is depicted in Figure 6 and its surface represents the number of copies that would be sold by a (new) newsstand if placed at the generic $\mathbf{s}$. The maxima of $q_{\hat{\Psi}}(\mathbf{s} \mid \mathcal{S})$ represent the locations where it would be profitable to open new newsstands. A comparison between the potential and the conditional potential over a zoomed area is given in Figure 7.The standard deviation of $q_{\hat{\Psi}}(\mathbf{s} \mid \mathcal{S})$, representing its reliability, is depicted in Figure 8 and it has been evaluated by estimating $q_{\hat{\Psi}_{(i)}}(\mathbf{s} \mid \mathcal{S})$ for each $\hat{\Psi}_{(i)} \in \hat{\Psi}_s$.

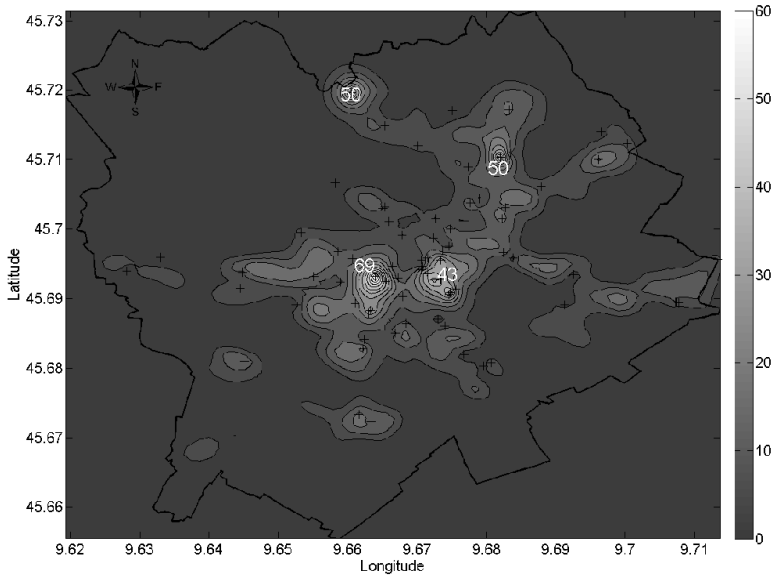

Figure 5: Estimated potential $q_{\hat{\Psi}}(\mathbf{s})$ (average daily number of copies) over the Bergamo area. Newsstand locations marked by the + symbol.

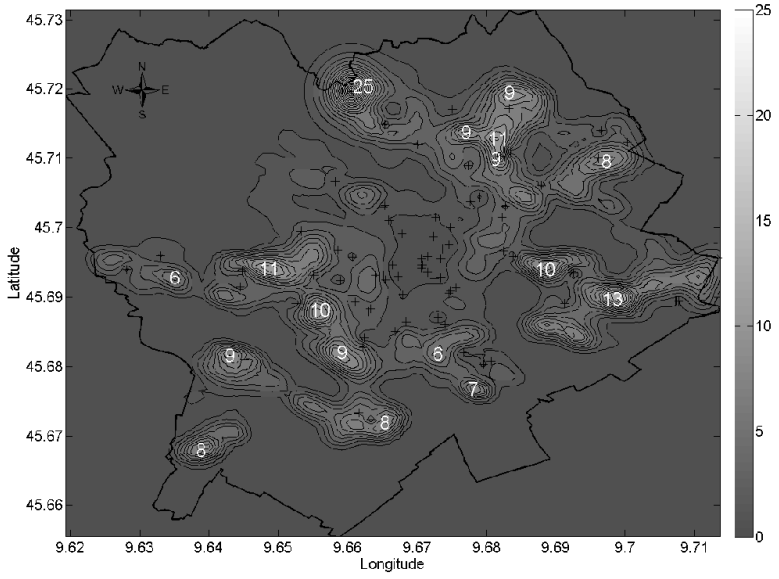

Figure 6: Estimated conditional potential $q_{\hat{\Psi}}(\mathbf{s} \mid \mathcal{S})$ (average daily number of copies) over the Bergamo area. Newsstand locations marked by the + symbol.

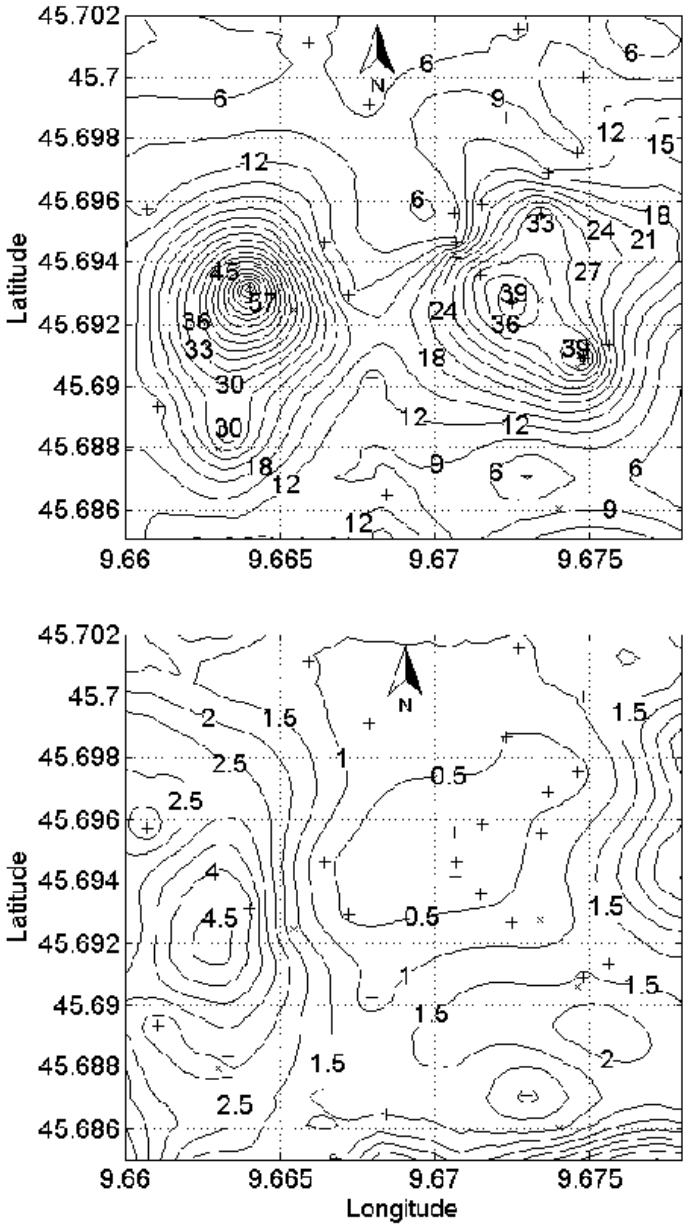

Figure 7: Comparison between the potential (upper) and the conditional potential (lower) for an enlarged area of Bergamo. Newstand locations marked by the + symbol (available data) and the × symbol (missing data).

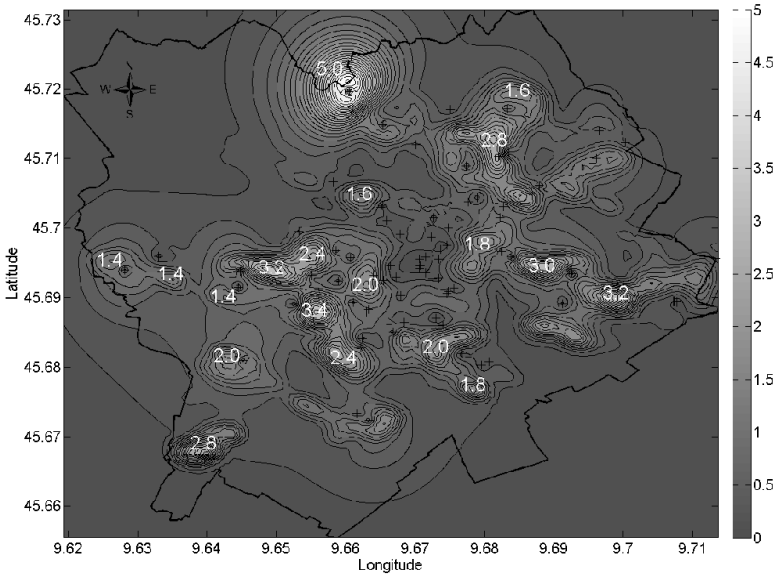

Figure 8: Standard deviation map for the estimated conditional potential $q_{\hat{\Psi}}(\mathbf{s} \mid \mathcal{S})$. Newsstand locations marked by the + symbol.

6 Conclusions

In this paper, the data generation mechanism that characterize the spatial data collected in many social and scientific fields has been explained in terms of an interaction occurring between the observed system and the set of measuring instruments used to acquire the data. The spatial data are assumed to be related to an underlying spatially continuous phenomenon which, in turn, is related to the state of the system. Thus, knowing the realization of the spatial phenomenon, and in particular its spatial surface, means to know the state of the system.

Depending on the properties of both the system and the measuring instruments, the effects of the interaction might have to be modeled. In such a case, the phenomenon spatial surface cannot be recovered neither by following a classic geostatistical approach nor by adopting a point process statistics approach. The problem has been solved by introducing the geostatistical potential model

in which the interaction is explicitly taken into account. The estimation of the model parameters is based on the ML approach and it is carried out by means of the EM algorithm. The missing data problem and the evaluation of both the model parameter and the model output uncertainty are solved within the framework developed through the paper. The geostatistical potential model has been successfully applied in the solution of a geomarketing problem and in particular in the estimation of the spatial market potential of a product from the sales data of spatially distributed stores.

References

[1] J. Illian, *Statistical analysis and modelling of spatial point patterns*. Statistics in practice, John Wiley, 2008.

[2] C. K. Wikle, "Hierarchical models in environmental science," *International Statistical Review*, vol. 71, no. 2, pp. 181–199, 2003.

[3] F. Goreaud and R. Pelissier, "On explicit formulas of edge effect correction for ripley's k-function," *Journal of Vegetation Science*, vol. 10, no. 3, pp. pp. 433–438, 1999.

[4] N. Kunzli, R. Kaiser, S. Medina, M. Studnicka, O. Chanel, P. Filliger, M. Herry, F. H. Jr, V. Puybonnieux-Texier, P. Quenel, J. Schneider, R. Seethaler, J.-C. Vergnaud, and H. Sommer, "Public-health impact of outdoor and traffic-related air pollution: a european assessment," *The Lancet*, vol. 356, no. 9232, pp. 795 – 801, 2000.

[5] D. Lee, C. Ferguson, and R. Mitchell, "Air pollution and health in scotland: a multicity study.," *Biostatistics*, vol. 10, pp. 409–423, 2009.

[6] A. Fassò and F. Finazzi, "Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data," *Environmetrics*, vol. 22, no. 6, pp. 735–748, 2011.

[7] P. Goovaerts, "Medical geography: A promising field of application for geostatistics," *Mathematical Geosciences*, vol. 41, pp. 243–264, 2009. 10.1007/s11004-008-9211-3.

[8] X. Jin, B. P. Carlin, and S. Banerjee, "Generalized hierarchical multivariate car models for areal data," *Biometrics*, vol. 61, no. 4, pp. 950–961, 2005.

[9] G. Cliquet, *Geomarketing: methods and strategies in spatial martketing.* Geographical Information Systems Series, ISTE USA, 2006.

[10] K. Xu and C. K. Wikle, "Estimation of parameterized spatio-temporal dynamic models," *Journal of Statistical Planning and Inference*, vol. 137, no. 2, pp. 567 – 588, 2007.

[11] R. Shumway and D. Stoffer, *Time series analysis and its applications, with R examples.* Springer, New York, 2006.

Appendix A

As stated in Section 4, the estimation of the model parameters is based on the maximum likelihood approach. In this appendix, the closed form estimation formulas for the parameter vector Ψ are derived starting from the complete-data likelihood and by applying the theory of the expectation-maximization algorithm.

The complete-data likelihood function $L_{\Psi}(\mathbf{y}, \mathbf{w})$ can be restated as

$$L_{\Psi}(\mathbf{y}, \mathbf{w}) = L_{\Psi_{\mathbf{y}}}(\mathbf{y} \mid \mathbf{w}) \cdot L_{\Psi_{\mathbf{w}}}(\mathbf{w})$$

and the following distributions hold

$$\begin{aligned} \mathbf{y} \quad &| \quad \mathbf{w} \sim N\left(\mathbf{G}\left(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{w}\right), \mathbf{G}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{G}'\right) \\ \mathbf{w} \quad &\sim \quad N\left(0, \boldsymbol{\Sigma}_{\mathbf{w}}\right) \end{aligned}$$

The complete-data log-likelihood function $-2l(\Psi; \mathbf{y}, \mathbf{w})$ is given by the summands

$$\begin{aligned} -2l\left(\Psi_{\mathbf{y}}; \mathbf{y} \mid \mathbf{w}\right) &= \log\left|\mathbf{G}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{G}'\right| + \qquad (27) \\ &\quad \left[\mathbf{y} - \mathbf{G}\left(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{w}\right)\right]' \left(\mathbf{G}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{G}'\right)^{-1} \left[\mathbf{y} - \mathbf{G}\left(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{w}\right)\right] \\ -2l\left(\Psi_{\mathbf{w}}; \mathbf{w}\right) &= \log\left|\boldsymbol{\Sigma}_{\mathbf{w}}\right| + \mathbf{w}'\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{w} \qquad (28) \end{aligned}$$

The E-step of the algorithm is defined by the following conditional expectation

$$\begin{aligned} Q\left(\Psi, \Psi^{(k)}\right) &= E_{\Psi^{(k)}}\left[-2l\left(\Psi; \mathbf{y}, \mathbf{w}\right) \mid \mathbf{y}^{(1)}\right] \\ &= E_{\Psi^{(k)}}\left[E_{\Psi^{(k)}}\left[-2l\left(\Psi; \mathbf{y}, \mathbf{w}\right) \mid \mathbf{y}^{(1)}, \mathbf{w}\right] \mid \mathbf{y}^{(1)}\right] \end{aligned}$$

The inner conditional expectation is

$$\begin{aligned} E_{\Psi^{(k)}}\left[-2l\left(\Psi; \mathbf{y}, \mathbf{w}\right) \mid \mathbf{y}^{(1)}, \mathbf{w}\right] &= \log\left|\mathbf{G}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{G}'\right| + tr\left[\left(\mathbf{G}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{G}'\right)^{-1}\left(\mathbf{e}\cdot\mathbf{e}' + \boldsymbol{\Lambda}\right)\right] + \\ &\quad \log\left|\boldsymbol{\Sigma}_{\mathbf{w}}\right| + tr\left[\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\left(\mathbf{w}\cdot\mathbf{w}'\right)\right] \end{aligned}$$

where

$$
\begin{aligned}
\mathbf{e} &= E\left[\mathbf{y}-\mathbf{G}\left(\mathbf{1}\mu+\mathbf{X}\boldsymbol{\beta}+\gamma\mathbf{w}\right)\mid\mathbf{y}^{(1)},\mathbf{w}\right] \\
&= \mathbf{D}\begin{pmatrix} \mathbf{y}^{(1)}-\mathbf{G}^{(1)}\left(\mathbf{1}^{(1)}\mu+\mathbf{X}^{(1)}\boldsymbol{\beta}+\gamma\mathbf{w}^{(1)}\right) \\ \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\left(\mathbf{y}^{(1)}-\mathbf{G}^{(1)}\left(\mathbf{1}^{(1)}\mu+\mathbf{X}^{(1)}\boldsymbol{\beta}+\gamma\mathbf{w}^{(1)}\right)\right)\end{pmatrix} \\
&= \mathbf{D}\begin{pmatrix} \mathbf{y}^{(1)}-\mathbf{G}^{(1)}\left(\mathbf{1}^{(1)}\mu+\mathbf{X}^{(1)}\boldsymbol{\beta}+\gamma\mathbf{w}^{(1)}\right) \\ \mathbf{0}\end{pmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{\Lambda} &= Var\left[\mathbf{y}-\mathbf{G}\left(\mathbf{1}\mu+\mathbf{X}\boldsymbol{\beta}+\gamma\mathbf{w}\right)\mid\mathbf{y}^{(1)},\mathbf{w}\right] \\
&= \mathbf{D}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}\left(\mathbf{R}_{22}-\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\right)\mathbf{G}'\end{pmatrix}\mathbf{D}' \\
&= \mathbf{D}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}\mathbf{R}_{22}\mathbf{G}'\end{pmatrix}\mathbf{D}'
\end{aligned}
$$

By applying the outer conditional expectation it follows

$$
\begin{aligned}
E_{\Psi^{(k)}}\left[E_{\Psi^{(k)}}\left[-2l\left(\Psi;\mathbf{y},\mathbf{w}\right)\mid\mathbf{y}^{(1)},\mathbf{w}\right]\mid\mathbf{y}^{(1)}\right] &= \log\left|\mathbf{G}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{G}'\right|+tr\left[\left(\mathbf{G}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{G}'\right)^{-1}\cdot\boldsymbol{\Omega}\right]+ \\
&\quad \log\left|\boldsymbol{\Sigma}_{\mathbf{w}}\right|+tr\left[\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\cdot\left(\hat{\mathbf{w}}\cdot\hat{\mathbf{w}}'+\hat{\mathbf{A}}\right)\right] \\
&= Q\left(\Psi_{\mathbf{y}},\Psi_{\mathbf{y}}^{(k)}\right)+Q\left(\Psi_{\mathbf{w}},\Psi_{\mathbf{w}}^{(k)}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\Omega} &= E_{\Psi^{(k)}}\left[\mathbf{e}\cdot\mathbf{e}'+\boldsymbol{\Lambda}\mid\mathbf{y}^{(1)}\right]=E_{\Psi^{(k)}}\left[\mathbf{e}\cdot\mathbf{e}'\mid\mathbf{y}^{(1)}\right]+\boldsymbol{\Lambda} \\
&= \mathbf{D}\begin{pmatrix} \boldsymbol{\Omega}^{(11)} & \boldsymbol{\Omega}^{(11)}\mathbf{R}_{11}^{-1}\mathbf{R}_{21} \\ \mathbf{R}_{21}\mathbf{R}_{11}^{-1} & \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\boldsymbol{\Omega}^{(11)}\mathbf{R}_{11}^{-1}\mathbf{R}_{21}+\left(\mathbf{R}_{22}-\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\right)\end{pmatrix}\mathbf{D}' \\
&= \mathbf{D}\begin{pmatrix} \boldsymbol{\Omega}^{(11)} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}\mathbf{R}_{22}\mathbf{G}'\end{pmatrix}\mathbf{D}' \\
\hat{\mathbf{w}} &= E_{\Psi^{(k)}}\left(\mathbf{w}\mid\mathbf{y}^{(1)}\right) \\
\hat{\mathbf{A}} &= Var_{\Psi^{(k)}}\left(\mathbf{w}\mid\mathbf{y}^{(1)}\right)
\end{aligned}
$$

and

$$\mathbf{\Omega}_t^{(11)} = E_{\Psi^{(k)}}\left[\mathbf{e}^{(1)} \mid \mathbf{y}^{(1)}\right] \cdot E_{\Psi^{(k)}}\left[\mathbf{e}^{(1)} \mid \mathbf{y}^{(1)}\right]' + Var_{\Psi^{(k)}}\left[\mathbf{e}^{(1)} \mid \mathbf{y}^{(1)}\right]$$

with

$$\begin{aligned}
E_{\Psi^{(k)}}\left[\mathbf{e}^{(1)} \mid \mathbf{y}^{(1)}\right] &= E_{\Psi^{(k)}}\left[\mathbf{y}^{(1)} - \mathbf{G}^{(1)}\left(\mathbf{1}^{(1)}\mu + \mathbf{X}^{(1)}\boldsymbol{\beta} + \gamma\mathbf{w}^{(1)}\right) \mid \mathbf{y}^{(1)}\right] \\
&= \mathbf{y}^{(1)} - \mathbf{G}^{(1)}\left(\mathbf{1}^{(1)}\mu + \mathbf{X}^{(1)}\boldsymbol{\beta} + \gamma\hat{\mathbf{w}}^{(1)}\right) \\
&= \tilde{\mathbf{e}}^{(1)} \\
Var_{\Psi^{(k)}}\left[\mathbf{e}^{(1)} \mid \mathbf{y}^{(1)}\right] &= Var_{\Psi^{(k)}}\left[\mathbf{y}^{(1)} - \mathbf{G}^{(1)}\left(\mathbf{1}^{(1)}\mu + \mathbf{X}^{(1)}\boldsymbol{\beta} + \gamma\mathbf{w}^{(1)}\right) \mid \mathbf{y}^{(1)}\right] \\
&= Var_{\Psi^{(k)}}\left[\gamma\mathbf{G}^{(1)}\mathbf{w}^{(1)} \mid \mathbf{y}^{(1)}\right] \\
&= \gamma^2\mathbf{G}^{(1)}Var_{\Psi^{(k)}}\left[\mathbf{w}^{(1)} \mid \mathbf{y}^{(1)}\right]\left(\mathbf{G}^{(1)}\right)' \\
&= \gamma^2\mathbf{G}^{(1)}\hat{\mathbf{A}}^{(1)}\left(\mathbf{G}^{(1)}\right)'
\end{aligned}$$

The M-step is characterized by the following minimization

$$\begin{aligned}
\Psi^{(k+1)} &= \arg\min_{\Psi} Q\left(\Psi, \Psi^{(k)}\right) \\
&= \arg\min_{\Psi} Q\left(\Psi_{\mathbf{y}}, \Psi_{\mathbf{y}}^{(k)}\right) + Q\left(\Psi_{\mathbf{w}}, \Psi_{\mathbf{w}}^{(k)}\right) \\
&= \left\{\arg\min_{\Psi_{\mathbf{y}}} Q\left(\Psi_{\mathbf{y}}, \Psi_{\mathbf{y}}^{(k)}\right), \arg\min_{\Psi_{\mathbf{w}}} Q\left(\Psi_{\mathbf{w}}, \Psi_{\mathbf{w}}^{(k)}\right)\right\} \qquad (29) \\
&= \left\{\Psi_{\mathbf{y}}^{(k+1)}, \Psi_{\mathbf{w}}^{(k+1)}\right\}
\end{aligned}$$

Note that $\Psi_{\mathbf{w}}^{(k+1)} \equiv \boldsymbol{\theta}^{(k+1)}$ and that the minimization of $Q\left(\Psi_{\mathbf{w}}, \Psi_{\mathbf{w}}^{(k)}\right)$ is tackled by a numerical iterative algorithm. With regard to $Q\left(\Psi_{\mathbf{y}}, \Psi_{\mathbf{y}}^{(k)}\right)$, the following minimization is carried out

$$\arg\min_{\Psi_{\mathbf{y}}}\left(N\log\left(\sigma_\varepsilon^2\right) + \log|\mathbf{G}\mathbf{G}'| + \frac{1}{\sigma_\varepsilon^2}\,\mathrm{tr}\begin{pmatrix}\tilde{\mathbf{e}}^{(1)}\cdot\left(\tilde{\mathbf{e}}^{(1)}\right)' + \gamma^2\hat{\mathbf{A}}^{(1)} & 0 \\ 0 & \mathbf{R}_{22}\end{pmatrix}\right) \qquad (30)$$

where $\hat{\mathbf{e}}^{(1)} = \left(\mathbf{G}^{(1)}\right)^{-1}\mathbf{y}^{(1)} - \mathbf{1}^{(1)}\mu - \mathbf{X}^{(1)}\boldsymbol{\beta} - \gamma\hat{\mathbf{w}}^{(1)}$.

A closed form solution can be obtained for the model parameters $\{\mu, \sigma_\varepsilon^2, \boldsymbol{\beta}, \gamma\}$ by evaluating the following derivatives

$$
\begin{aligned}
&\tfrac{\partial Q\left(\Psi,\Psi^{(k)}\right)}{\partial \mu} = 0 \Rightarrow \\
&\partial[1/\sigma_\varepsilon^2 \cdot \operatorname{tr}(\mu^2 \mathbf{1}^{(1)} \left(\mathbf{1}^{(1)}\right)' - \mu \left(\mathbf{G}^{(1)}\right)^{-1} \mathbf{y}^{(1)} \left(\mathbf{1}^{(1)}\right)' - \mu \mathbf{1}^{(1)} \left(\mathbf{y}^{(1)}\right)' \left(\mathbf{G}^{(1)}\right)^{-1} + \mu \mathbf{1}^{(1)} \boldsymbol{\beta}' \left(\mathbf{X}^{(1)}\right)' + \\
&\mu \mathbf{X}^{(1)} \boldsymbol{\beta} \left(\mathbf{1}^{(1)}\right)' + \mu\gamma \mathbf{1} \left(\hat{\mathbf{w}}^{(1)}\right)' + \mu\gamma \hat{\mathbf{w}}^{(1)} \left(\mathbf{1}^{(1)}\right)')]/\partial\mu = 0 \Rightarrow \\
&\tfrac{2N\mu}{\sigma_\varepsilon^2} + \tfrac{1}{\sigma_\varepsilon^2} \operatorname{tr}\left(-2 \left(\mathbf{G}^{(1)}\right)^{-1} \mathbf{y}^{(1)} \left(\mathbf{1}^{(1)}\right)' + 2\mathbf{X}^{(1)} \boldsymbol{\beta} \left(\mathbf{1}^{(1)}\right)' + 2\gamma \hat{\mathbf{w}}^{(1)} \left(\mathbf{1}^{(1)}\right)'\right) = 0 \Rightarrow \\
&\mu = \tfrac{\operatorname{tr}\left[\left(\left(\mathbf{G}^{(1)}\right)^{-1} \mathbf{y}^{(1)} - \mathbf{X}^{(1)} \boldsymbol{\beta} - \gamma \hat{\mathbf{w}}^{(1)}\right) \mathbf{1}'\right]}{N - N_m}
\end{aligned}
\tag{31}
$$

$$
\begin{aligned}
&\tfrac{\partial Q\left(\Psi,\Psi^{(k)}\right)}{\partial \boldsymbol{\beta}} = 0 \Rightarrow \\
&\partial[1/\sigma_\varepsilon^2 \cdot \operatorname{tr}(\mathbf{X}^{(1)} \boldsymbol{\beta} \boldsymbol{\beta}' \left(\mathbf{X}^{(1)}\right)' - \left(\mathbf{G}^{(1)}\right)^{-1} \mathbf{y}^{(1)} \boldsymbol{\beta}' \left(\mathbf{X}^{(1)}\right)' - \mathbf{X}^{(1)} \boldsymbol{\beta} \left(\mathbf{y}^{(1)}\right)' \left(\mathbf{G}^{(1)}\right)^{-1} + \\
&\mu \mathbf{1}^{(1)} \boldsymbol{\beta}' \left(\mathbf{X}^{(1)}\right)' + \mu \mathbf{X}^{(1)} \boldsymbol{\beta} \left(\mathbf{1}^{(1)}\right)' + \gamma \mathbf{X}^{(1)} \boldsymbol{\beta} \left(\hat{\mathbf{w}}^{(1)}\right)' + \gamma \left(\hat{\mathbf{w}}^{(1)}\right) \boldsymbol{\beta}' \left(\mathbf{X}^{(1)}\right)')]/\partial\boldsymbol{\beta} = 0 \Rightarrow \\
&2 \left(\mathbf{X}^{(1)}\right)' \mathbf{X}^{(1)} \boldsymbol{\beta} - 2 \left(\mathbf{X}^{(1)}\right)' \left(\mathbf{G}^{(1)}\right)^{-1} \mathbf{y}^{(1)} + 2\mu \left(\mathbf{X}^{(1)}\right)' \mathbf{1}^{(1)} + 2\gamma \left(\mathbf{X}^{(1)}\right)' \hat{\mathbf{w}}^{(1)} = 0 \Rightarrow \\
&\boldsymbol{\beta} = \left[\left(\mathbf{X}^{(1)}\right)' \mathbf{X}^{(1)}\right]^{-1} \left(\mathbf{X}^{(1)}\right)' \cdot \left(\left(\mathbf{G}^{(1)}\right)^{-1} \mathbf{y}^{(1)} - \mu \mathbf{1}^{(1)} - \gamma \hat{\mathbf{w}}^{(1)}\right);
\end{aligned}
\tag{32}
$$

$$
\begin{aligned}
&\tfrac{\partial Q\left(\Psi,\Psi^{(k)}\right)}{\partial \sigma_\varepsilon^2} = 0 \Rightarrow \\
&\tfrac{N}{\sigma_\varepsilon^2} - \tfrac{1}{(\sigma_\varepsilon^2)^2} \operatorname{tr}\begin{pmatrix} \tilde{\mathbf{e}}^{(1)} \cdot \left(\tilde{\mathbf{e}}^{(1)}\right)' + \gamma^2 \hat{\mathbf{A}}^{(1)} & 0 \\ 0 & \mathbf{R}_{22} \end{pmatrix} = 0 \Rightarrow \\
&\sigma_\varepsilon^2 = \tfrac{1}{N} \operatorname{tr}\begin{pmatrix} \tilde{\mathbf{e}}^{(1)} \cdot \left(\tilde{\mathbf{e}}^{(1)}\right)' + \gamma^2 \hat{\mathbf{A}}^{(1)} & 0 \\ 0 & \mathbf{R}_{22} \end{pmatrix};
\end{aligned}
\tag{33}
$$

$$\frac{\partial Q\left(\Psi,\Psi^{(k)}\right)}{\partial\gamma}=0\Rightarrow$$
$$\partial[1/\sigma_{\varepsilon}^{2}\cdot tr(\gamma^{2}\hat{\mathbf{w}}^{(1)}\left(\hat{\mathbf{w}}^{(1)}\right)'-\gamma\left(\mathbf{G}^{(1)}\right)^{-1}\mathbf{y}\left(\hat{\mathbf{w}}^{(1)}\right)'-\gamma\hat{\mathbf{w}}^{(1)}\left(\mathbf{y}^{(1)}\right)'\left(\mathbf{G}^{(1)}\right)^{-1}+\mu\gamma\mathbf{1}^{(1)}\left(\hat{\mathbf{w}}^{(1)}\right)'+$$
$$\mu\gamma\hat{\mathbf{w}}^{(1)}\left(\mathbf{1}^{(1)}\right)'+\gamma\mathbf{X}^{(1)}\boldsymbol{\beta}\left(\hat{\mathbf{w}}^{(1)}\right)'+\gamma\hat{\mathbf{w}}^{(1)}\boldsymbol{\beta}'\left(\mathbf{X}^{(1)}\right)'+\gamma^{2}\hat{\mathbf{A}}^{(1)})]/\partial\gamma=0\Rightarrow$$
$$\frac{2\gamma}{\sigma_{\varepsilon}^{2}}tr\left[\hat{\mathbf{w}}^{(1)}\left(\hat{\mathbf{w}}^{(1)}\right)'+\hat{\mathbf{A}}^{(1)}\right]+$$
$$\frac{1}{\sigma_{\varepsilon}^{2}}\operatorname{tr}\left[-2\left(\mathbf{G}^{(1)}\right)^{-1}\mathbf{y}^{(1)}\left(\hat{\mathbf{w}}^{(1)}\right)'+2\mu\hat{\mathbf{w}}^{(1)}\left(\mathbf{1}^{(1)}\right)'+2\mathbf{X}^{(1)}\boldsymbol{\beta}\left(\hat{\mathbf{w}}^{(1)}\right)'\right]=0\Rightarrow$$
$$\gamma=\frac{\operatorname{tr}\left[\left(\left(\mathbf{G}^{(1)}\right)^{-1}\mathbf{y}^{(1)}-\mu\mathbf{1}^{(1)}-\mathbf{X}^{(1)}\boldsymbol{\beta}\right)\left(\hat{\mathbf{w}}^{(1)}\right)'\right]}{\operatorname{tr}\left(\hat{\mathbf{w}}^{(1)}\left(\hat{\mathbf{w}}^{(1)}\right)'+\hat{\mathbf{A}}^{(1)}\right)};\tag{34}$$

Appendix B

The latent variable $\mathbf{w}$ is estimated by applying the usual formulas of the multivariate normal distribution. In particular

$$\begin{aligned}\hat{\mathbf{w}} &= E_{\Psi^{(k)}}\left(\mathbf{w}\mid\mathbf{y}\right)\\ &= \boldsymbol{\Sigma}_{\mathbf{wy}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\left[\mathbf{y}-E\left(\mathbf{y}\right)\right]\\ &= \boldsymbol{\Sigma}_{\mathbf{wy}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\left[\mathbf{y}-\mathbf{G}\left(\mathbf{1}\mu+\mathbf{X}\boldsymbol{\beta}\right)\right]\end{aligned}\tag{35}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{y}} &= \operatorname{Var}\left[\mathbf{G}\left(\mathbf{1}\mu+\mathbf{X}\boldsymbol{\beta}+\gamma\mathbf{w}+\boldsymbol{\varepsilon}\right)\right]\\ &= \mathbf{G}\operatorname{Var}\left[\gamma\mathbf{w}+\boldsymbol{\varepsilon}\right]\mathbf{G}'\\ &= \mathbf{G}\left(\gamma^{2}\boldsymbol{\Sigma}_{\mathbf{w}}+\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\right)\mathbf{G}'\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{wy}} &= E\left[\left(\mathbf{w}-\mathbf{0}\right)\cdot\left[\mathbf{y}-E\left(\mathbf{y}\right)\right]'\right]\\ &= E\left[\mathbf{w}\cdot\left(\gamma\mathbf{G}\mathbf{w}\right)'\right]\\ &= \gamma\boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{G}'\end{aligned}$$

The variance of the estimated $\hat{\mathbf{w}}$ is given by

$$\begin{aligned}\hat{\mathbf{A}} &= Var_{\Psi^{(k)}}(\mathbf{w} \mid \mathbf{y}) \\ &= \boldsymbol{\Sigma}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{wy}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}(\boldsymbol{\Sigma}_{\mathbf{wy}})' \end{aligned} \tag{36}$$

When $\mathbf{y}$ is characterized by missing data, (35) and (36) become

$$\begin{aligned}\hat{\mathbf{w}} &= (\boldsymbol{\Sigma}_{\mathbf{wy}}\mathbf{L}')(\mathbf{L}\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{L}')^{-1}[\mathbf{L}(\mathbf{y} - \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta}))] \\ \hat{\mathbf{A}} &= \boldsymbol{\Sigma}_{\mathbf{w}} - (\boldsymbol{\Sigma}_{\mathbf{wy}}\mathbf{L}')(\mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}')^{-1}(\mathbf{L}\Sigma_{\mathbf{wy}})\end{aligned}$$

Appendix C

The evaluation of the approximate Fisher information matrix defined in (19) requires the computation of the vector derivatives $\partial\boldsymbol{\epsilon}(\Psi)/\partial\Psi_i$ and the matrix derivatives $\partial\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\Psi)/\partial\Psi_i$, $1 \leq i \leq |\Psi|$, with $\boldsymbol{\epsilon}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ defined in (20) and (21) respectively.

In the case of the spatial correlation defined in (25) and the interaction function defined in (26), the following derivatives hold

$$\frac{\partial\boldsymbol{\epsilon}(\Psi)}{\partial\Psi_i} = \begin{cases} -\mathbf{g} & if \quad \Psi_i = \mu \\ -\mathbf{G}\mathbf{x}_j & if \quad \Psi_i = \beta_j \quad ; 1 \leq j \leq b \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{37}$$

$$\frac{\partial\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\Psi)}{\partial\Psi_i} = \begin{cases} \mathbf{g}\mathbf{g}' \odot I_N & if \quad \Psi_i = \sigma_\varepsilon^2 \\ 2\gamma\mathbf{g}\mathbf{g}' \odot \boldsymbol{\Sigma}_{\mathbf{w}} & if \quad \Psi_i = \gamma \\ \gamma^2\mathbf{g}\mathbf{g}' \odot \frac{\mathbf{H}}{\theta^2} \odot \boldsymbol{\Sigma}_{\mathbf{w}} & if \quad \Psi_i = \theta \\ \tilde{\mathbf{G}} \odot (\gamma^2\boldsymbol{\Sigma}_{\mathbf{w}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) & if \quad \Psi_i = \phi \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{38}$$

where $\mathbf{x}_j$ is the $j-th$ column of the matrix $\mathbf{X}$ and $\mathbf{H}$ is the distance matrix based on $\mathcal{S}$.

Finally, the $(ij)-th$ element of the matrix $\tilde{\mathbf{G}}$ is given by $\partial g_i \cdot g_j + g_i \partial g_j$, where g_i is the $i-th$ element of the vector $\mathbf{g}$ and

$$\partial g_i = \frac{\partial g_i}{\partial \phi} = -\frac{\sum_{j=1}^{N} \frac{h_{ij}}{\phi^2} exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{\phi}\right)}{\left[\sum_{j=1}^{N} exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{\phi}\right)\right]^2} \tag{39}$$

with h_{ij} the $(ij)-th$ element of the matrix $\mathbf{H}$.