New methodologies for the change of support problems: development of spatial statistical models

Graduate School of Systems and Information Engineering
University of Tsukuba

March 2014

Daisuke Murakami

Abstract

Spatial data have become more diverse as geographic information systems (GIS) have developed, and as a result, so have the spatial supports for these data (e.g., aggregation units, data locations, and so on). This has created opportunities where spatial supports of data at hand do not compatible with the spatial supports that the users want. Thus, developing methodologies for change of spatial support problems (COSPs), such as the problem of how to convert prefectural population data into municipal population data is a critical issue in geographical information sciences.

COSPs consist of two sub-problems: (i) the problem of changing the spatial support itself; and (ii) the problem related to (i). If we consider the two major types of spatial data, namely areal data (or lattice data) and point data (or geo-referenced data), we can sub-divide the problems in (i) into two sub-problems: (i-1) changing the spatial support for areal data (e.g., a gridded population interpolation using municipal population data); and (i-2) changing the spatial support for point data (e.g., weather data interpolation on gridded points using data from monitoring stations; i.e., changing the support from the monitoring station sites to the gridded points). A possible solution to sub-problem (i-1) applies the areal interpolation technique, and a solution to sub-problem (i-2) applies the point interpolation technique. Therefore, discussing interpolation problems is important from the viewpoint of COSPs.

The sub-problems (i-1) and (i-2) each have their own concomitant problems. The modifiable areal unit problem (MAUP) is a problem related to (i-1). The MAUP refers to the problem of bias in the model parameters due to aggregation. A typical example of the MAUP is that the correlation coefficient between two aggregated variables changes drastically, depending on their aggregation (or areal) unit. Therefore, changing the spatial support for areal data (or areal data interpolation) must consider both the interpolation accuracy and the influence of the MAUP, especially when the interpolated data are used for secondary analyses. On the other hand, the sampling design problem is a problem related to (i-2). An example of a sampling design problem is efficient weather station allocation. This problem is closely related to changes in the support of point data. Here, point data interpolation must consider both the interpolation accuracy and the efficiency of the interpolated site allocation.

Thus, this study focuses on four types of COSPs: areal interpolation and its related problem, the MAUP, and point interpolation and its related sampling design problem.

While these COSPs are currently popular topics in geostatistics, non-geostatistical spatial statistical models, such as geographically weighted regression models and spatial filtering models, have rarely been applied to these problems. Discussing all relevant spatial models would be extremely helpful in constructing sophisticated methodologies to address COSPs. Hence, this study discusses the four COSPs by applying a wide variety of spatial statistical models.

The outline of this study is as follows. Chapter 1 introduces my discussion. In particular, this chapter indicates that developing new methodologies for COSPs is important, considering the recent diversification of spatial and spatiotemporal data, and that spatial statistical models offer a potentially useful set of solutions. Chapter 2 summarizes the basic spatial statistical models, including geostatistical models, spatial filter models, and the geographically weighted regression model. The subsequent chapters discuss COSPs. Chapters 3 and 4 discuss the areal interpolation problem and the MAUP, which are COSPs for areal data. Then, Chapters 5 and 6 discuss the point interpolation problem and the sampling design problem, which are COSPs for point data.

Chapter 3 extends the geographically weighted regression (GWR) model for areal interpolation. The GWR model captures spatial heterogeneity by allowing coefficients to vary across space. The extended GWR-based model has the following advantages: it captures spatial heterogeneity in the same way as the conventional GWR model; it provides the best unbiased linear predictor, as do existing geostatistical areal interpolation models; it satisfies the volume preserving property (e.g., the sum of the interpolated municipal populations must equal the actual prefectural-level population), which is the most basic property that must be satisfied in areal interpolation. In addition, I discuss how a non-negative constraint is imposed on the interpolated values.

The effectiveness of the GWR-based method is examined by applying it to a simulation study. The simulation results reveal that this method outperforms conventional non-statistical areal interpolation methods, including the areal weighting interpolation method and the dasymetric method. On the other hand, the results also show that the accuracy of proposed method is unstable compared to the conventional methods, and its accuracy possibly be worse than them.

To examine the effectiveness of the GWR-based method in a practical application, it is also applied to an empirical study of interpolating municipal building stocks for various categories (wooden/non-wooden, residential/non-residential, completion year) in Japan. The results again show the effectiveness of the proposed method from the viewpoint of interpolation accuracy and the ability to explain the spatially dependent component.

In contrast to the standard GWR model, the extended GWR model explicitly considers an aggregation mechanism, and offers a solution to the MAUP. In fact, this is an aggregate-level model that furnishes unbiased, consistent, efficient, and asymptotically normal estimators of non-aggregate-level parameters. Hence, Chapter 4 examines the effectiveness of the method from the viewpoint of the MAUP. This chapter first describes a simulation study, and reveals that the model effectively copes with the MAUP as long as the spatial scale of the aggregation is not coarser than the spatial scale of the underlying spatial heterogeneity. Then, the GWR-based model is applied to a criminal analysis. The results confirm that the model provides intuitively reasonable non-aggregate-level parameter estimates using aggregated variables.

Thus, Chapters 3 and 4 discuss the COSPs for areal data. Next, Chapters 5 and 6 discuss the COSPs for point data.

Chapter 5 focuses on the point interpolation problem. This problem has been discussed extensively among geostatisticians. However, geostatistical point interpolation methods generally have the following drawbacks. First, the methods are not necessarily simple to implement, and spatial adjustments are required to extend them (e.g., for non-Gaussian data). Then, they can easily become computationally intractable, particularly when interpolating spatiotemporal data.

The eigenvector spatial filtering (ESF) method, which models spatially dependent components using eigenvectors of a proximity matrix, is relatively straightforward to implement and extend. Thus, Chapter 5 extends ESF for point interpolation, while considering both simplicity and computational efficiency. Note that, ESF has already been extended for interpolating lattice data, which are point data with sample sites that are fixed and finite. On the other hand, this study focuses on interpolating geo-referenced data, which are another type of point data, with sample sites that are distributed in a continuous spatial region. Here, ESF is extended while ensuring consistency with both the standard ESF method and standard geostatistics.

The ESF-based extended method is applied to a land price analysis. The results show that its point interpolation accuracy is almost the same as the standard geostatistical method. In addition, the results demonstrate the method’s efficiency for multiscale spatial component extraction, estimation in the presence of spatial dependence, and variance partitioning analysis.

The proposed method is then extended to include spatiotemporal modeling. The analysis results show that its interpolation accuracy is the same as the standard spatiotemporal geostatistical method, while its computation time is substantially less than the geostatistical method.

Chapter 6 focuses on the sampling design problem, which is another COSP of point data. Whereas point interpolations have been well researched, studies on sampling designs are still relatively limited and, even for the most standard geostatistical approach, their effectiveness is still unclear. In particular, since the geostatistical approach has been applied mainly to natural science data, how appropriate it is for social economic data (e.g., land price data) remains unclear.

Hence, this chapter discusses the land price assessment site reduction problem (a sampling design problem) in Japan. Since Japan is planning to gradually reduce the number of land price assessment sites, discussing this problem is important. I first extend the standard geostatistical approach to consider the properties that must be included in the reduction problem. In particular, I discuss how the land price assessment site allocation criteria launched by the government, the uses of land prices, and so on, are considered in the geostatistical method. The method is then applied to the reduction problem in the Ibaraki prefecture, Japan. The results suggest that the extended geostatistical method provides intuitively reasonable reduction results.

Finally, the discussions are summarized and concluded in chapter 7.

Contents

List of Figures

List of Tables

List of abbreviations

AICc	: Corrected Akaike information criteria
ATP GWR	: Area-to-point GWR (geographically weighted regression)
ATP kriging	: Area-to-point kriging
AW	: Areal weighting interpolation method
BLUP	: Best linear unbiased predictor
COSP	: Change of support problem
DA	: Dasymetric method
dESF	: Distance based ESF (eigenvector spatial filtering)
EM algorithm	: Expectation-maximization algorithm
EOF	: Empirical orthogonal function
ESDA	: Exploratory spatial data analysis
ESF	: Eigenvector spatial filtering

EX_AICc	: AICc-minimization-based ESF with $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
EX_MC	: MC-minimization-based ESF with $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
E_AICc	: AICc-minimization-based ESF with $\mathbf{M} = \mathbf{I} - \mathbf{1}\mathbf{1}'/n$
E_MC	: MC-minimization-based ESF with $\mathbf{M} = \mathbf{I} - \mathbf{1}\mathbf{1}'/n$
GIS	: Geographic information systems
GISc	: Geographic information sciences
GLS	: Generalized least squares
GS	: Geostatistical model
GWR	: Geographically weighted regression
GWR_Ag	: Aggregate-level GWR
GWR_NAg	: Non-aggregate-level GWR
LM	: Linear regression model
MAE	: Mean absolute error
MAE_den	: Mean absolute error for density variables
MAPE	: Mean absolute percentage error
MAUP	: Modifiable areal unit problem
MC	: Moran coefficient
MC^{+}	: Moran coefficient on a continuous space
MEM	: Moran's eigenvector mapping
MLIT	: Ministry of Land, Infrastructure, Transport and Tourism, Japan
MSPE	: Mean square prediction error
NLNI	: National Land Numerical Information download service
OLS	: Ordinary least squares
RMSE	: Root mean square error
RMSE_den	: Root mean square error for density variables
RMSPE	: Root mean square percentage error
SEM	: Spatial error model
SLM	: Spatial lag model
SPE	: Square prediction error
WLS	: Weighted least squares

1. Introduction

1.1. Development of GIS and spatial data diversification

In accordance with the development of Geographic Information Systems (GIS), the interdisciplinary use of geographical information, which refers to “information about places on the Earth’s surface, knowledge about where something is, knowledge about what is at a given location” (Goodchild, 1997), has become widespread. The evolution of GIS can be summarized as follows (see, Goodchild, 2010):

Between the 1960s and the late 1980s:

> The term GIS was coined and GIS evolved into a widely adopted software application.

Between the late 1980s and the early 1990s:

> Discussions began about the science of GIS, and Geographic Information Sciences (GISc), which is a research field for “the development and use of theories, methods, technology, and data for understanding geographic processes, relationships, and patterns” (Mark, 2000; Goodchild, 2010), was established.

After the early 1990s:

> GIS and GISc advanced rapidly in accordance with the development of computer technology, and now, GIS is widely adopted, not only for research, but also in practical applications.

Based on such backgrounds, developed technologies relating GIS and GISc can be classified in four categories (Goodchild, 2009): (i) systems for positioning (e.g., Global Positioning Systems: GPS); (ii) systems for data acquisition (e.g., satellite and airborne remote sensing), (iii) systems for data dissemination (e.g., National Land Numerical Information download services (NLNI): http://nlftp.mlit.go.jp/ksj/; Google Maps: https://maps.google.com/), and (iv) systems for analysis (e.g., ArcGIS, provided by ESRI: http://www.esri.com/). Due to the developments of (i) and (ii), spatial data (i.e., collections of geographical information) have diversified dramatically, whereas the development of (iii) has increased the opportunities for handling such diversified spatial data. Thus, diversified spatial data handling has become increasingly important.

The diversification of spatial data introduces diversification of their spatial supports (e.g., areal units, data locations, and so on). In fact, areal (aggregated) units take various spatial scales. For example, in Japan, areal units of social economic data take either prefectural units, municipal units, or minor municipal units. Furthermore, areal units change depending on the study field (e.g., many natural science data are aggregated into grids, whereas many social economic data are aggregated into administrative units). On the other hand, in the case of point (non-aggregated) data, data locations differ depending on the data source. For example, in Japan, land price assessment sites and weather observation sites are incompatible.

The diversity of spatial data increases opportunities where the spatial supports of the data at hand do not accord with the spatial supports for which the users want. For example, one might wish to analyze gridded population data while only municipal population data are available. In another example, one might require weather data in each minor municipal unit while the weather data are available only at their specific monitoring stations. To cope with such difficulties, discussing techniques of changing spatial supports is increasingly important.

1.2. The change of support problems

Spatial interpolation, including point interpolation and areal interpolation, is a useful technique for changing spatial supports. Point interpolation, which refers to the point data interpolation using point data with different sites, has been discussed thoroughly in geostatistics (e.g., Cressie, 1993; Cressie and Wikle, 2011), whose origin is a point interpolation study in mining (see, §2.2). On the other hand, areal interpolation, which refers to the areal data interpolation using areal data with different spatial scale of aggregation, has been discussed mainly in geography rather than geostatistics. This is partly because an approach depending on the nature of aggregation (i.e., areal unit) is consistent with geographic common sense (Openshaw and Taylor, 1981), whereas such dependency is not necessarily preferred in (geo-)statistical literatures (e.g., Tobler, 1979). In the other words, an areal interpolation problem that considers areal units explicitly is a geographical problem rather than a (geo-)statistical problem (Gelfand, 2010).

On the other hand, as GIS develops, the importance of discussing changing spatial support is being increasingly recognized and an increasing number of geostatisical studies have started discussing spatial interpolation problems under a framework called the change of support problem (COSP: e.g., Gotway and Young, 2002; Cressie and Wikle, 2011). The major sub-problems in COSP can be summarized as shown in Table 1-1 (see, Gotway and Young, 2002).

While the point interpolation problem and areal interpolation problem are the central COSP problems (see Table 1-1), the COSPs also include the following related problems: the modifiable areal interpolation problem (MAUP: e.g., Openshaw, 1984) and the sampling design problem (e.g., Wang *et al*., 2012). In the MAUP, which is related to the areal interpolation problem, is the problem that the change of aggregation units changes the spatial data analysis results. For instance, Openshaw and Taylor (1979) showed that the correlation coefficient between two variables changes between -0.97 and 0.99, depending on the aggregation units. Areal interpolation must be performed considering not only interpolation accuracy but also its influences on the MAUP, especially when the interpolated data are used for secondary analyses. On the other hand, the sampling design problem discusses, for example, efficient weather monitoring sites allocation. This problem is closely related to point interpolation; point interpolation must be performed considering not only the interpolation accuracy but also efficiency of the interpolated site allocation. In fact, even if the interpolation model is accurate, when the allocation of interpolated sites is not good, the resulting interpolated data quality might be poor.

In summary, change of spatial support must be discussed considering not only interpolation problems themselves, but also the MAUP and the sampling design problem.

Table 1-1: Example of COSPs

Data before conversion	Data after conversion	Related problems
Areal data	Point data	Point interpolation Sampling design
	Areal data	Block interpolation (including averaging)
Point data	Point data	Areal interpolation Modifiable areal unit problem (including ecological fallacy)
	Areal data	

[1)] The figure is constructed while referring to Gotway and Young (2002)

1.3. Problems in COSP studies

Geostatistics is a sub-field within spatial statistics (see §2.1); a study area of discussing statistical spatial data analysis (Haining *et al.*, 2010). Spatial statistical models, including geostatistical models, spatial filtering models (e.g., Griffith, 2003), and geographically weighted regression model (GWR: e.g., Fotheringham *et al.*, 2002), have been discussed extensively in recent years.

While the COSPs have been discussed extensively in recent geostatistics, the COSPs have less focused in the other spatial statistics. This could be due to the fact that geostatistics originated as an interpolation (or COSP) study, whereas the other spatial statistics did not (see §2.2). However, as I will show later, both of these spatial interpolation approaches (or approaches of changing spatial support) are essentially identical. Hence, it is significant to examine effectiveness of non-geostatistical spatial statistical models from the perspective of the COSPs.

Another problem is the lack of interdisciplinary discussions of the COSPs. Although the COSPs have been studied both in geography and geostatistics[1], their interdisciplinary discussions seem insufficient. Such a tendency is particularly prominent in areal interpolation studies (see, chapter 3). Because advantages and disadvantages of geographical and geostatistical approaches are quite different, interdisciplinary discussions would be effective to develop more sophisticated methodologies.

[1] The term COSP itself is used only among geostatisticians.

In summary, the COSPs must be discussed from a broader perspective while referring to geography, geostatistics, and non-geostatistical spatial statistics (see §2.1).

1.4. Outline of this study

The objective of this study is developing new methodologies for the COSPs mainly focusing on spatial statistics, while paying attention to geographical literatures too.

Fig.1-1 organizes the chapters in this study. In the next chapter, I discuss the spatial statistical models, including the geostatistical model, the spatial filtering models, and the GWR model. Then, between chapter 3 and 6, I discuss the COSPs while referring to both spatial statistics and geography. Concretely, chapter 3 proposes a GWR-based areal interpolation model, and compares its efficiency with the other geographical and spatial statistical models. Chapter 4 applies the GWR-based model for the MAUP. Chapter 5 proposes a spatial filtering-based point interpolation method, and compares it with the standard point interpolation methods, and chapter 6 discusses a sampling design problem of land prices using a geostatistical approach. Finally, I summarize my whole discussion in chapter 7.

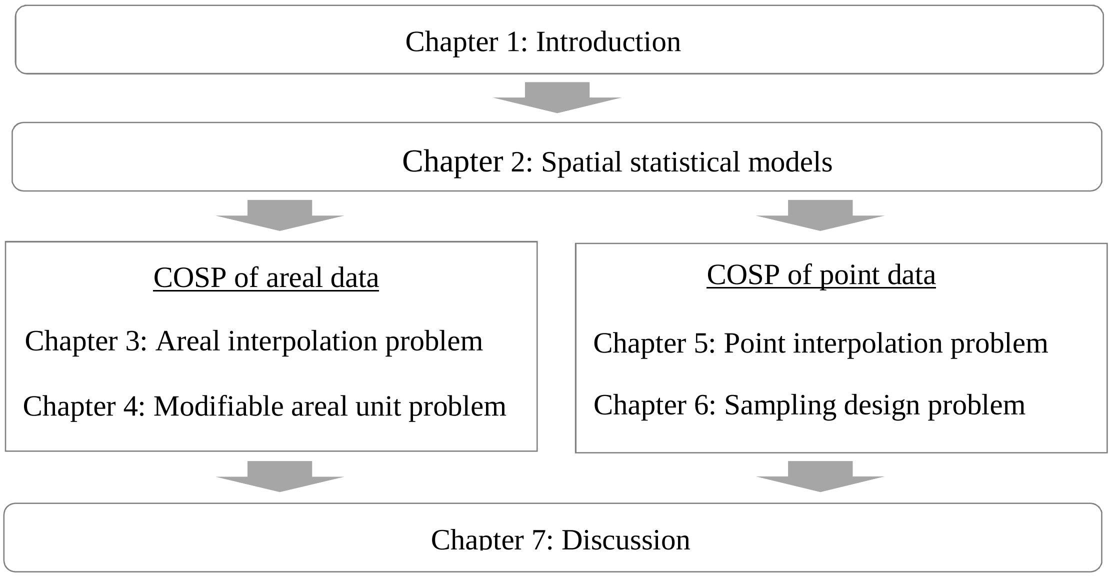

Chapter 1: Introduction
Chapter 2: Spatial statistical models
COSP of areal data
Chapter 3: Areal interpolation problem
Chapter 4: Modifiable areal unit problem
COSP of point data
Chapter 5: Point interpolation problem
Chapter 6: Sampling design problem
Chapter 7: Discussion

Figure 1-1: Outline of this study

2. Spatial Statistical Models

2.1. Spatial analysis and spatial statistics

According to Haining (2003), the term spatial analysis, which refers to an analysis performed by applying techniques and models that use explicitly the spatial referencing associated with each data value or object that is specified within the system under study, can be traced back to at least the 1950s (see Berry and Marble, 1968). Spatial analysis includes several distinct elements; however, the statistical analysis of spatial data, which is referred to by statisticians as spatial statistics (Ripley, 1981; Haining *et al*., 2010), is an element that has been discussed widely.

Spatial statistics is distinct from non-spatial (i.e., standard) statistics in that it considers fundamental properties of spatial data: spatial dependence and spatial heterogeneity (e.g., Anselin, 1988). Spatial dependence is the property that dictates that attribute values located closely in geographic space are similar. This property is also known as the first law of geography; "Everything is related to everything else, but near things are more related to each other" (Tobler, 1970). The consideration of spatial dependence is important, for example, in modeling spatial data accurately (or appropriately) (see, e.g., Cressie, 1993) and to test statistical significant appropriately (see, e.g., LeSage and Pace, 2009). Thus, spatial dependence modeling is one of the primal topics in spatial statistics. The geostatistical model, which is discussed in §2.2 and the spatial filter model, which is discussed in §2.3, are models representative of describing spatial dependence (Griffith and Paelinck, 2011).

On the other hand, spatial heterogeneity is a special case of observed or unobserved heterogeneity and is a familiar problem, e.g., in standard econometrics (Anselin, 2010). Different from spatial dependence, spatial statistical models are not necessarily required to capture spatial heterogeneity. In fact, some sort of heterogeneity can be captured by applying a standard linear regression model. Spatial statistics is helpful in capturing spatial heterogeneity, with its features vary gradually over space. GWR, which is discussed in §2.4, allows such heterogeneity (i.e., spatially continuous heterogeneity) to be captured by applying spatially varying parameters. Note that spatial dependence and spatially continuous heterogeneity are difficult to separate, and that modeling this type of heterogeneity is useful to capture spatial dependence (Fotheringham *et al*., 2002).

This chapter discusses the primal spatial statistical approaches of addressing spatial dependence and spatial heterogeneity.

2.2. Geostatistics

Geostatistics, which is a sub-field within spatial statistics, originates from a series of studies by D.G. Krige, a professor at the University of Witwatersrand, South African (Diggle, 2010). He promoted using statistical methods for mineral explorations (Krige, 1951). Motheron (1963) refined his work into a theory of stochastic process with covariograms and semivariograms (see §2.2.1). In addition, based on this theory, he proposed a best linear unbiased prediction (BLUP) methodology for spatial data, which he termed kriging in honor of D.G. Krige. The Motheron methodology has been discussed and extended, particularly by applied scientists and mathematicians (Haining *et al.*, 2010), and a field of study discussing the methodology is now known as geostatistics.

Such an evolution of geostatistics is very different from that of other methods of spatial statistics, which have been developed mainly in regional sciences and quantitative geography. As a result, it has often been mentioned that geostatistics is distinctive (e.g., Haining *et al.*, 2010).
Seq.2.2 briefly discusses geostatistics.

2.2.1. Basic assumptions

The geostatistical model describes a stochastic process that is a family or collection of random variables. The members of the collection can be identified or indexed by a set of locations **s** $\in D \subset \Re^d$, where *d* takes the value 2 or 3 in most cases (Schabenberger and Gotway, 2005). When *d* is greater than 1, a stochastic process is also called a random field.

Standard geostatistics models a stochastic process that satisfies stationarity and ergodicity. Stationarity is an assumption that the properties of the stochastic process remain unchanged, depending on the locations. This lack of change enables us to model the stochastic process using only one model. On the other hand, ergodicity is an assumption that the dependency between two samples asymptotically goes to zero as the distance between the two samples increases (see, e.g., Arbia, 2006). Ergodicity ensures that the true mean and covariance of a stochastic process are the same as their estimates (Arbia, 2006). Since most stationary stochastic processes satisfy ergodicity (the random work process is an exception: Gaetan and Guyon, 2010), hereafter, I discuss stationary stochastic processes without paying further attention to ergodicity.

Let $Z(s_i)$ be a real-valued stochastic process defined on a domain, D. $Z(s_i)$ is called a strict (or strong) stationary process if its (finite dimensional) joint distributions are invariant under spatial shifts (Gneiting and Guttorp, 2010); that is,

$$F[Z(s_i)] = F[Z(s_i + h_{i,j})], \tag{2-1}$$

where $F[\bullet]$ is the distribution function and $h_{i,j} \in \Re^d$ is a lag distance that separates sites s_i and s_j. Eq.(2-1) implies that all moments of the stochastic process are unchanged throughout the domain D.

While strict stationarity is a condition of the distribution, the second-order (or weak) stationarity is a weaker stationarity that conditions the first and second moments of $Z(s_i)$ using Eqs.(2-2) and (2-3):

$$E[Z(s_i)] = 0, \tag{2-2}$$

$$Cor[Z(s_i), Z(s_i + h_{i,j})] = c(h_{i,j}), \tag{2-3}$$

where $c(h_{i,j})$ is a distance function called a covariogram (or covariance function). A strict stationary process is always a second-order stationary process, but the reverse is not always true. As an exception, when $Z(s_i)$ is a second-order Gaussian process whose expectation and covariance obey Eq.(2-2) and Eq.(2-3), respectively, it is also a strict stationary (Gaussian) process.

Matheron (1973) proposes the other type of stationarity, called intrinsic stationary. This imposes stationarity on the increments $Z(s_i) - Z(s_i + h_{i,j})$, using Eqs.(2-4) and (2-5):

$$E[Z(s_i) - Z(s_i + h_{i,j})] = 0, \tag{2-4}$$

$$\frac{1}{2} Var[Z(s_i) - Z(s_i + h_{i,j})] = \gamma(h_{i,j}), \tag{2-5}$$

where $\gamma(h_{i,j})$ is a distance function called a semivariogram (and $2\gamma(h_{i,j})$ is called a variogram). The intrinsic stationary process is an analog of the stationary increment process used frequently in time series analyses (Schabenberger and Gotway, 2005). Eq.(2-5) is expanded as

$$\begin{aligned}\gamma(h_{i,j}) = \frac{1}{2}Var[Z(s_i) - Z(s_i + h_{i,j})] &= \frac{1}{2}\{Var[Z(s_i)] + Var[Z(s_i + h_{i,j})] - 2Cov[Z(s_i), Z(s_i + h_{i,j})]\},\\ &= \frac{1}{2}\{2Var[Z(s_i)] - 2c(h_{i,j})\},\\ &= c(0) - c(h_{i,j}). \qquad (2\text{-}6)\end{aligned}$$

Eq.(2-6) suggests that a second-order stationary process that includes $c(h_{i,j})$ corresponds to an intrinsic stationary process with $\gamma(h_{i,j})$ (Cressie, 1993). In contrast, because $c(0)$ is undefined for some $\gamma(h_{i,j})$ (e.g., $c(0)$ is undefined for the linear semivariogram model Eq.2-13), the reverse is not necessarily true. In short, second-order stationarity includes intrinsic stationarity.

While the aforementioned stochastic processes are isotropic in the sense that $h_{i,j}$ in $c(h_{i,j})$ or $\gamma(h_{i,j})$ does not depend on directions, a stochastic process is anisotropic if $h_{i,j}$ does depend on directions. There are at least two types of anisotropies: the geometric anisotropy and the zonal anisotropy. Geometric anisotropy applies a linear transformation (or rotation) of the 2D coordinate system. More precisely, suppose that **h** is a vector whose elements are given by $h_{i,j}$, and **B** is a matrix for the liner transformation, then, the covariogram and semivariogram are defined using the elements in **Bh**. Zonal anisotropy uses a linear combination of an isotropic model, ($c(h_{i,j})$ or $\gamma(h_{i,j})$), and a model depending only on a lag distance in one direction (i.e., $c(h_{i,j}) + c(h_1)$, where $h_1 \in \Re^1$). For more details about anisotropy, see e.g., Zimmerman (1993).

2.2.2. Models for covariogram and semivariogram

Suppose that **y** is a response variable vector with elements that obey the second-order stationary process (i.e., $E[\mathbf{y}]=\mathbf{0}$, $E[\mathbf{yy}'] = \mathbf{C}$, where **0** is a vector of zeros, and **C** is a variance-covariance matrix are given by $c(h_{i,j})$). Then, the standard geostatistics describes the spatial process of **y** using $\boldsymbol{\lambda}'\mathbf{y}$, where $\boldsymbol{\lambda}$ is a vector of weights and **y** is a vector of variables observed in D.

To ensure the validity of the spatial process, $Z(s_i) = \boldsymbol{\lambda}'\mathbf{y}$, the variance of $Z(s_i)$, $Var[Z(s_i)]$, must be non-negative (e.g., Armstrong and Diamond, 1984); that is,

$$Var[\boldsymbol{\lambda}'\mathbf{y}] \geq 0. \qquad (2\text{-}7)$$

Eq.(2-7) can be expanded as

$$Var[\boldsymbol{\lambda}'\mathbf{y}] = E[(\boldsymbol{\lambda}'\mathbf{y})(\boldsymbol{\lambda}'\mathbf{y})'] = \boldsymbol{\lambda}'E[\mathbf{y}\mathbf{y}']\boldsymbol{\lambda},$$

$$= \boldsymbol{\lambda}'\mathbf{C}\boldsymbol{\lambda} \geq 0. \qquad (2\text{-}8)$$

The second line of Eq.(2-8) is identical to the positive semi-definite condition for **C** (or $c(h_{i,j})$). Thus, to ensure that $Var[Z(s_i)]$ is non-negative, $c(h_{i,j})$ must be defined by a positive semi-definite function, or equally, $\gamma(h_{i,j})$ must be defined by a negative semi-definite function (it is readily derived from Eq.2-6). According to Cressie (1993), the positive definiteness of $c(h_{i,j})$ is the only necessary and sufficient condition for a valid spatial process model.

Various positive definite functions for $c(h_{i,j})$ have been proposed. The standard functions are as follows:

Exponential model $$c(h_{i,j}) = \begin{cases} \tau^2 + \sigma^2 & if \;\; h_{i,j} = 0 \\ \tau^2 \exp\left(-\dfrac{h_{i,j}}{r}\right) & otherwise \end{cases}, \qquad (2\text{-}9)$$

Gaussian model $$c(h_{i,j}) = \begin{cases} \tau^2 + \sigma^2 & if \;\; h_{i,j} = 0 \\ \tau^2 \exp\left(-\dfrac{h_{i,j}^2}{r^2}\right) & otherwise \end{cases}, \qquad (2\text{-}10)$$

Spherical model $$c(h_{i,j}) = \begin{cases} \tau^2 + \sigma^2 & if \;\; h_{i,j} = 0 \\ \tau^2\left(\dfrac{3}{2}\dfrac{h_{i,j}}{r} - \dfrac{1}{2}\dfrac{h_{i,j}^3}{r^3}\right) & if \;\; 0 < h_{i,j} < r \\ 0 & otherwise \end{cases}, \qquad (2\text{-}11)$$

Matérn model $$c(h_{i,j}) = \begin{cases} \tau^2 + \sigma^2 & if \;\; h_{i,j} = 0 \\ \tau^2 \dfrac{1}{2^{\nu-1}\Gamma(\nu)}\left(\sqrt{2\nu}\dfrac{h_{i,j}}{r}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\dfrac{h_{i,j}}{r}\right) & otherwise \end{cases}, \qquad (2\text{-}12)$$

where σ^2, τ^2, and r are parameters called nugget, partial-sill, and range, respectively. σ^2 denotes the variance of the micro-scale spatial variation and/or measurement error, τ^2 denotes the variance of the spatially dependent component, and r measures the range of the spatial dependence. Here, r in the spherical model Eq.(2-11) is interpreted as the distance at which the spatially dependent component vanishes (i.e., $c(h_{i,j}) = 0$). On the other hand, interpretations of r in the other models, in which $c(h_{i,j})$ becomes zero only asymptotically, are not necessarily straightforward. Hence, the effective range, r^*, denoting the distance at which 95% of the spatially dependent component vanishes, has often been applied. It was numerically clarified that $r^* = 3r$ when the exponential model Eq.(2-9) is used, and r^* =

$r\sqrt{3}$ when the Gaussian model Eq.(2-10) is used (Zimmerman and Stein, 2010). In the Matérn model shown in Eq.(2-12), $\Gamma(v)$ is the gamma function and $K_v(\bullet)$ is the modified Bassel function. The Matérn model is $v - 1$ times differentiable. In the other words, v controls the smoothness of this model. The exponential model is a particular case of the Matérn model with $v = 0.5$, and the Gaussian model is an extreme case of the Matérn model, with $v \to \infty$ (Hoeting *et al.*, 2006).

Because of such generality, use of the Matérn model has been encouraged, for example, by Stein (1999). On the other hand, the spherical model, which provides a sparse covariance matrix, is computationally more efficient (Gneiting and Guttorp, 2010).[1] Note that all of the models assume $h_{i,j}$ to be a Euclidean distance, and positive definiteness is not guaranteed when non-Euclidean distance measures are applied (Curriero, 2006). In addition, the spherical model is valid only on a Euclidean space with a dimension below 3 (Fuentes and Reich, 2010).

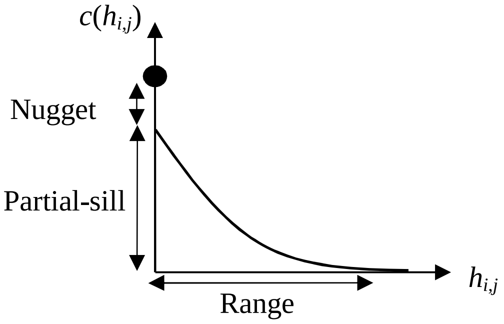


Figure 2-1: Image of covariogram: $c(h_{i,j})$

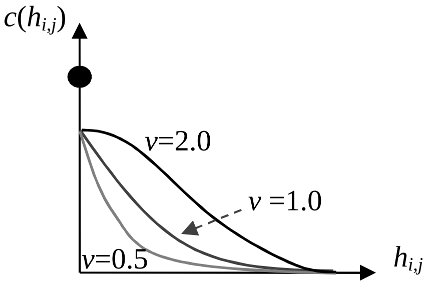


Figure 2-2: Image of the smoothing parameter: v.

Note: This diagram illustrates three covariograms, each with the same range, but with different smoothing parameters.

[1] Since the spherical function forces zero values for all $c(h_{i,j})$ with $h_{i,j} > r$, **C** becomes sparse if r is small.

On the other hand, various negative definite functions have also been proposed for the $\gamma(h_{i,j})$. The standard functions are as listed below:

Linear model
$$\gamma(h_{i,j}) = \begin{cases} 0 & if \quad h_{i,j} = 0 \\ \sigma^2 - \tau^2 h_{i,j} & otherwise \end{cases}, \tag{2-13}$$

Exponential model
$$\gamma(h_{i,j}) = \begin{cases} 0 & if \quad h_{i,j} = 0 \\ \sigma^2 - \tau^2 \exp\left(-\dfrac{h_{i,j}}{r}\right) & otherwise \end{cases}, \tag{2-14}$$

Gaussian model
$$\gamma(h_{i,j}) = \begin{cases} 0 & if \quad h_{i,j} = 0 \\ \sigma^2 - \tau^2 \exp\left(-\dfrac{h_{i,j}^2}{r^2}\right) & otherwise \end{cases}, \tag{2-15}$$

Spherical model
$$\gamma(h_{i,j}) = \begin{cases} 0 & if \quad h_{i,j} = 0 \\ \sigma^2 - \tau^2 \left(\dfrac{3}{2}\dfrac{h_{i,j}}{r} - \dfrac{1}{2}\dfrac{h_{i,j}^3}{r^3}\right) & if \quad 0 < h_{i,j} < r \\ \sigma^2 + \tau^2 & otherwise \end{cases}, \tag{2-16}$$

Matérn model
$$\gamma(h_{i,j}) = \begin{cases} 0 & if \quad h_{i,j} = 0 \\ \sigma^2 - \tau^2 \dfrac{1}{2^{v-1}\Gamma(v)}\left(\sqrt{2v}\dfrac{h_{i,j}}{r}\right)^v K_v\left(\sqrt{2v}\dfrac{h_{i,j}}{r}\right) & otherwise \end{cases}. \tag{2-17}$$

Eqs.(2-14), (2-15), (2-16), and (2-17) are given by substituting Eqs.(2-9), (2-10), (2-11), and (2-12) into Eq.(2-6). These models describe both the second-order stationary process, which is described by $c(h_{i,j})$, and the intrinsic stationary process, which is described by $\gamma(h_{i,j})$. In contrast, Eq.(2-13) does not have a corresponding $c(h_{i,j})$, and the linear model, which describes the intrinsic stationary process, cannot describe the second-order stationary process.

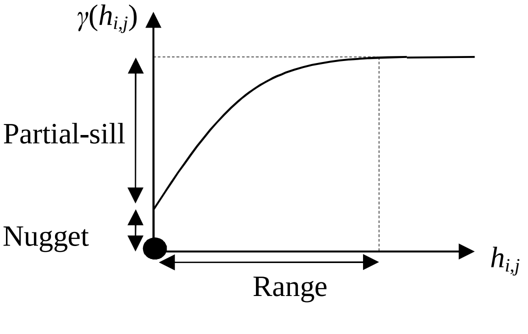


Figure 2-3: Image of semivariogram: $\gamma(h_{i,j})$

2.2.3. Geostatistical model

Geostatistics models the response variables, **y**, as

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \tag{2-18}$$

where $\boldsymbol{\mu}$ = $E[\mathbf{y}]$ is a deterministic non-spatial mean function, and $\boldsymbol{\varepsilon}$ is a spatial stochastic process. In many cases, $\boldsymbol{\mu}$ is given by a linear function and $\boldsymbol{\varepsilon}$ is given by the second-order (or strict) Gaussian process as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{C}), \tag{2-19}$$

where **X** is a matrix of explanatory variables, and $\boldsymbol{\beta}$ is a vector of parameters.

2.2.4. Estimation

While a number of estimation methods have been proposed for geostatistical models, including the maximum likelihood method (Mardia and Marshall, 1984), the restricted maximum likelihood method (e.g., Stein, 1999), the Bayesian estimation method (e.g., Handcock and Stein, 1993), and the estimation function-based methods (e.g., Schabenberger and Gotway, 2005), the weighted least squares (WLS)-based method (Cressie, 1985) is one of the most standard approaches. In what follows, I explain the WLS-based parameter estimation for Eq.(2-19).

I first consider the case that $\mathbf{X}\boldsymbol{\beta}$ is known. In this case, $\boldsymbol{\varepsilon}$ is given by $\hat{\boldsymbol{\varepsilon}}$ = $\mathbf{y}$ – $\mathbf{X}\boldsymbol{\beta}$. Consider fitting $\gamma(\mathbf{h})$ to $\{\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\}^2$ (see Eq.2-5), where **s** and **h** are vectors whose elements are s_i and $h_{i,j}$, respectively, and $\boldsymbol{\varepsilon}(\mathbf{s})$ is $\boldsymbol{\varepsilon}$ with its locations are specified using **s**. It is known that $\{\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\}^2$ has a large variance, and the variance makes the fitting inefficient (Cressie, 1993). To cope with this problem, $\gamma(\mathbf{h})$ is fitted after the elements in $\{\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\}^2$ are averaged (Schabenberger and Gotway, 2005). The averaging is performed in each of the lag-distance zones, which are pre-determined based on **h**. For example, the elements in $\{\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\}^2$, with **h** between 0 m and 10 m, are averaged; the elements in $\{\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\}^2$, with **h** between 10 m and 20 m, are averaged; and so on. Typically, to stabilize the fitting, the lag-distance zones are decided so as to include at least 30 location pairs in each lag. In addition, the elements in $\{\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\}^2$ with **h** more than half the maximum lag-distance (i.e., max[**h**]) are discarded.

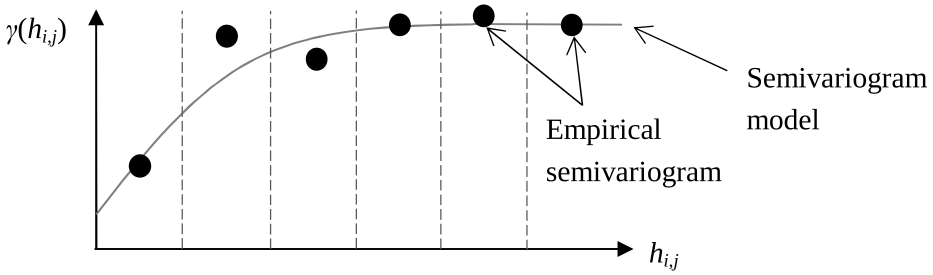


Figure 2-4: Image of the empirical semivariogram and semivariogram model

There are several averaging equations. For example, Matheron (1963) proposes Eq.(2-20):

$$\hat{\gamma}(h_l) = \frac{1}{2N_l} \sum_l \{\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\}^2 \ , \tag{2-20}$$

where l {= 1,... L} is the index of the lag-distance zones, h_l is their corresponding distance, and N_l is the number of pairs in the l-th zone. Cressie and Hawkins (1980) indicate that Matheron's estimator, which contains a squared term, is not robust for outliers, and propose the following robust averaging equation:

$$\hat{\gamma}(h_l) = \frac{\frac{1}{2N_l}\left\{\sum_l |\,\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\,|^{1/2}\right\}^4}{0.457 + \frac{0.494}{N_l}} \ . \tag{2-21}$$

As a result of replacing $\Sigma\{\,\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\,\}^2$ with $\{\Sigma|\,\hat{\boldsymbol{\varepsilon}}(\mathbf{s}) - \hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\,|^{1/2}\}^4$, the Cressie and Hawkins's estimator is more robust than Matheron's estimator (Cressie, 1993). In both estimators, $\hat{\gamma}(h_l)$ is called an empirical semivariogram.

The empirical semivariogram is defined only for L lag-distances: h_l (see Fig.2-4). Hence, to model semivariograms on arbitrary distances, a semivariogram model (e.g., the exponential model) must be fitted to the empirical semivariogram. Cressie (1985) suggests fitting a semivariogram model, $\gamma(h_l)$, using the non-linear WLS that minimizes Eq.(2-22):

$$\sum_l \frac{N_l}{2\gamma(h_l)^2} \sum_l \{\gamma(h_l) - \hat{\gamma}(h_l)\}^2 \ . \tag{2-22}$$

Eq.(2-22) can be minimized using standard non-linear statistics packages.

In summary, when **Xβ** in Eq.(2-19) (or **μ** in Eq.2-18) is known, a semivariogram model can be fitted using the following steps: (i) Calculate $\hat{\boldsymbol{\varepsilon}}$ = **y** – **Xβ**; (ii) Estimate the empirical semivariogram of $\hat{\boldsymbol{\varepsilon}}$ by averaging $\{\hat{\boldsymbol{\varepsilon}}(\mathbf{s})-\hat{\boldsymbol{\varepsilon}}(\mathbf{s}+\mathbf{h})\}^2$ using either Eq.(2-20) or Eq.(2-21); and (iii) Fit a semivariogram model for the empirical semivariogram by minimizing Eq.(2-22).

On the other hand, when **Xβ** (or **μ**) is unknown, **β** and **θ** must be estimated simultaneously. This is because **β** depends on **θ**, and vice versa. The iterative-reweighted least squares (IRLS) method is applicable for the simultaneous estimation. The estimation procedure is described as follows (Schabenberger and Gotway, 2005):

1: Estimate the ordinary least squares (OLS) estimates of **β**, $\hat{\boldsymbol{\beta}}$.

2: Calculate $\hat{\boldsymbol{\varepsilon}}$ = $\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}$.

3: Estimate the empirical semivariogram using Eq.(2-20) or Eq.(2-21), and fit a semivariogram model using the non-linear WLS estimation.

4: Construct **C**, with elements **c**(**h**), given by substituting **γ**(**h**) estimated in step 3 into Eq.(2-6).

5: Update $\hat{\boldsymbol{\beta}}$ using its generalized least squares (GLS) estimator, which is defined as

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{y}\,. \tag{2-23}$$

6: Iterate steps 2 to 5 until the parameter values converge.

The variance-covariance matrix of the resulting $\hat{\boldsymbol{\beta}}$ is given as

$$Var[\hat{\boldsymbol{\beta}}]=(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\,. \tag{2-24}$$

Eq.(2-24) is useful when testing the significance of $\hat{\boldsymbol{\beta}}$.

2.2.5. Kriging

The best linear unbiased prediction that applies a geostatistical model is called kriging. This section explains kriging, in line with Schabenberger and Gotway (2005).

Suppose that y_0 is the unobserved response variable at site s_0. The basic equation for $y(s_0)$, based on Eq.(2-19), is

$$y_0 = \mathbf{x}_0'\boldsymbol{\beta} + \varepsilon_0, \qquad Var[\varepsilon_0] = \sigma^2, \qquad Cor[\boldsymbol{\varepsilon}, \varepsilon_0] = \mathbf{c}, \tag{2-25}$$

where $\mathbf{x}_0$ is a vector of explanatory variables, ε_0 is the disturbance with variance σ^2, and $\mathbf{c}$ is a vector of covariances between ε_0 and $\boldsymbol{\varepsilon}$.

The objective of kriging is to find the best linear unbiased predictor (BLUP), $\hat{y}_0$, of y_0 that satisfies the following conditions:

Best (minimum variance) $\qquad \underset{\hat{y}(s_0)}{\arg\min}\ E[(y_0 - \hat{y}_0)^2]$, (2-26)

Linearity $\qquad \hat{y}_0 = \boldsymbol{\lambda}'\mathbf{y}$, (2-27)

Unbiasedness $\qquad E[y_0] = E[\hat{y}_0]$. (2-28)

$\hat{y}_0$ is identified by identifying $\boldsymbol{\lambda}$ that satisfies above conditions. Since $E[y_0] = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$, the unbiasedness suggests that $\mathbf{x}_0'\hat{\boldsymbol{\beta}} = E[\hat{y}_0]$, whereas the linearity suggests that $E[\hat{y}_0] = E[\boldsymbol{\lambda}'\mathbf{y}] = \boldsymbol{\lambda}'\mathbf{X}\hat{\boldsymbol{\beta}}$. These two conditions imply that $\mathbf{x}_0'\hat{\boldsymbol{\beta}} = \boldsymbol{\lambda}'\mathbf{X}\hat{\boldsymbol{\beta}}$, or equivalently, $\mathbf{x}_0' = \boldsymbol{\lambda}'\mathbf{X}$.

On the other hand, $E[(y_0 - \hat{y}_0)^2]$ in Eq.(2-26) can be expanded using Eqs.(2-25) and (2-27) as

$$\begin{aligned} E[(y_0 - \hat{y}_0)^2] &= E[(y_0 - \boldsymbol{\lambda}'\mathbf{y})^2], \\ &= Var[y_0] + Var[\boldsymbol{\lambda}'\mathbf{y}] - Cov[\boldsymbol{\lambda}'\mathbf{y}, y_0], \\ &= \sigma_0^2 + \boldsymbol{\lambda}'\mathbf{C}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\mathbf{c}. \end{aligned} \tag{2-29}$$

After all, the BLUP can be identified by minimizing Eq.(2-29) on condition that $\mathbf{x}_0' = \boldsymbol{\lambda}'\mathbf{X}$. The problem is solved by minimizing the Lagrangian, which is defined as

$$L = \sigma_0^2 + \boldsymbol{\lambda}'\mathbf{C}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\mathbf{c} + 2\mathbf{m}'(\mathbf{X}'\boldsymbol{\lambda} - \mathbf{x}_0), \tag{2-30}$$

where $\mathbf{m}$ is a vector of Lagrange multipliers. By differencing Eq.(2-30) with respect to $\boldsymbol{\lambda}$ and $\mathbf{m}$, the following first-order conditions are given:

$$\frac{\partial}{\partial \boldsymbol{\lambda}} L = 2\mathbf{C}\boldsymbol{\lambda} - 2\mathbf{c} + 2\mathbf{X}\mathbf{m} = \mathbf{0}, \tag{2-31}$$

$$\frac{\partial}{\partial \mathbf{m}} L = 2\mathbf{m}'(\mathbf{X}'\boldsymbol{\lambda} - \mathbf{x}_0) = \mathbf{0}\,. \tag{2-32}$$

By solving Eqs.(2-31) and (2-32), $\boldsymbol{\lambda}$ is given as

$$\boldsymbol{\lambda} = \mathbf{C}_X^{-}\mathbf{c} + \mathbf{C}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{C}^{-1}\mathbf{x}_0\,, \tag{2-33}$$

$$\mathbf{C}_X^{-} = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\,.$$

Consequently, $\hat{y}(s_0)$ is given by substituting Eq.(2-33) into Eq.(2-27), as follows:

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} + \mathbf{c}'\mathbf{C}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\,, \tag{2-34}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{y}\,.$$

The mean square prediction error (MSPE) of $\hat{y}_0$, which is called kriging variance, is also given by substituting Eq.(2-33) into Eq.(2-29), as follows:

$$MSPE[\hat{y}_0] = \sigma^2 + \mathbf{c}'\mathbf{C}^{-1}\mathbf{c} + (\mathbf{x}_0' - \mathbf{c}'\mathbf{C}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}(\mathbf{x}_0' - \mathbf{c}'\mathbf{C}^{-1}\mathbf{X})\,. \tag{2-35}$$

Thus, the interpolated value of y_0 is given by $\hat{y}_0$, and its uncertainty can be measured using Eq.(2-35).

For more details about geostatistics, see, for example, Cressie (1993), Schabenberger and Gotway (2005), Gelfand *et al*. (2010), and Cressie and Wikle (2011).

2.3. Spatial filtering

While conventional geostatistics models spatial dependence uses a distance function (i.e., $c(\mathbf{h})$ or $\gamma(\mathbf{h})$), the spatial filtering approach (e.g., Griffith, 2010) models it with map pattern variables (Getis and Griffith, 2002). There are two main spatial filtering approaches: the approach of Getis (1990), which defines the map pattern variables by applying the G_i statistics (Getis and Ord, 1992); and the approach of Griffith (1996), which is called eigenvector spatial filtering (ESF), which defines the variables based on the Moran coefficient (MC: Moran, 1950). One of the biggest advantages of these spatial filtering approaches is simplicity (Griffith, 2003; Getis, 2010). Their basic models are identical to the standard linear regression model, and accordingly, their implementations, estimations, and extensions are straightforward. In addition, the effectiveness of the spatial filtering approaches has been recognized in various purposes including parameter estimations in the presence of spatial dependence (e.g., Tiefersdorf and Griffith, 2007; Thayn and Simanis, 2013), and exploratory spatial data analysis (ESDA), such as spatial pattern analysis and spatial interpolation (e.g., Griffith, 2003; Legendre and Legendre, 2012).

This section explains the G_i statistics-based spatial filtering approach and the MC-based approach.

Note that the term "spatial filtering" is also known as a technique that separates an image into signals (or a de-noised image) and noise, in the study field of image analysis (e.g., Russ, 2006). For example, the moving average filter separates an image into a de-noised image, which is defined by the moving average of adjacent pixel values, and noise. Similar interpretation is possible for the spatial statistical spatial filtering techniques too: they decompose underlying process in spatial data into pure (or de-noised) spatial dependent component and noise that could not be explained by the spatial dependent component. Thus, spatial filtering in spatial statistics and spatial filtering in image analysis are compatible.

2.3.1. G_i statistics-based spatial filtering

The G_i statistics is a local indicator of spatial association (LISA: Anselin, 1995) that detects local spatial clusters, and is defined as

$$G_i = \frac{\sum_{j \neq i} w(h_{i,j}) y_j}{\sum_{j \neq i} y_j}, \tag{2-36}$$

where $w(h_{i,j})$ represents the spatial connectivity between s_i and s_j, and y_i is a response variable. Here, $w(h_{i,j})$ can be defined, for example, by applying a distance-decay function or a function that takes 1 if $h_{i,j}$ is less than a threshold, and 0 otherwise (see, e.g., Getis, 2010). By design, y_i must be positive. The value of G_i is large if higher values are clustered nearby to s_i, and small (close to zero) if lower values are clustered nearby to s_i. That is, a high G_i suggests that s_i is a hot spot, and a low G_i suggests that s_i is a cool spot.

The significance of hot or cool spots is tested using the expectation and variance of G_i, which are given under the randomized hypothesis, as

$$E[G_i] = \frac{\sum_{j \neq i} w(h(s_i, s_j))}{N-1}, \tag{2-37}$$

$$Var[G_i] = \frac{\sum_{j \neq i} w(h(s_i, s_j)) \left[N - 1 - \sum_{j \neq i} w(h(s_i, s_j)) \right]}{(N-1)^2 (N-2)} \frac{Y_{i2}}{Y_{i1}}. \tag{2-38}$$

$$Y_{i1} = \frac{\sum_{j \neq i} y_j}{N-1} \qquad Y_{i2} = \frac{\sum_{j \neq i} y_j^2}{N-1} - Y^2{}_i$$

For more details about G_i statistics, see Getis and Ord (1992) and Ord and Getis (1995).

The G_i statistics-based spatial filtering approach filters the spatially dependent component in y_i using $G_i / E[G_i]$. Here, $G_i / E[G_i]$ describes the deviation of the spatial pattern of y_i from the hypothesized spatially randomized distribution ($G_i / E[G_i]$ is greater than 1 if the response variables nearby s_i are larger than expected, and $G_i / E[G_i]$ is smaller than 1 if the values are smaller than expected). In other words, $G_i / E[G_i]$ captures the spatial patterns of y_i. Accordingly, by dividing y_i by $G_i / E[G_i]$, the spatial patterns in y_i are filtered out, and the spatially random (or independent) component in y_i, $y^{NS}{}_i$, is given as

$$y_i^{NS} = \left(\frac{G_i}{E[G_i]} \right)^{-1} y_i. \tag{2-39}$$

In addition, the spatially non-random (or dependent) component in y_i, $y^S{}_j$, is given by $y_i - y^{NS}{}_i$.

The G_i statistics-based spatial filtering approach is useful for spatial dependence modeling. For example, Getis (2010) estimates the linear regression model, which is defined as

$$y_i = \alpha + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \varepsilon_i , \qquad \varepsilon_i \sim N(0,\sigma^2) , \tag{2-40}$$

by applying the following procedure: (i) Spatial dependences in y_i, $x_{i,1}$, and $x_{i,2}$ are tested using the Moran coefficient (see §2.3.1.1); (ii) Variables with significant spatial dependence (MC) are decomposed into spatially dependent and independent components by applying the G_i statistics-based spatial filtering technique. In this case, when spatial dependences in y_i and $x_{i,1}$ are significant, the model is modified as follows:

$$y_i = \alpha + y_i^S \beta^S + x_{i,1}^S \beta_1^S + x_{i,1}^{NS} \beta_1^{NS} + x_{i,2}\beta_2 + \varepsilon_i , \qquad \varepsilon_i \sim N(0,\sigma^2) ; \tag{2-41}$$

(iii) The parameters in the modified model are estimated using OLS. Getis and Griffith (2002) and Getis (2010) demonstrate that this simple procedure effectively removes spatial dependence in residuals and improves model accuracy.

2.3.2. MC-based spatial filtering

2.3.2.1. MC-based eigenfunctions

The Moran coefficient (MC), which is defined by the following equation, is a spatial dependence diagnostic statistics:

$$MC = \frac{N}{\mathbf{1}'\mathbf{W}\mathbf{1}} \frac{\mathbf{z}'\mathbf{MWMz}}{\mathbf{z}'\mathbf{Mz}} , \tag{2-42}$$

where **1** is a vector of ones, **z** is a vector of diagnosed variables, **W** is a binary and symmetric connectivity matrix with diagonals of 0, and **M** is a projection matrix. Two types of the projection matrix **M** have been applied, namely $\mathbf{I}-\mathbf{11}'/N$ and $\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (e.g., Anselin and Rey, 1991). Here, $\mathbf{I}-\mathbf{11}'/N$ is used if **z** contains raw data, and $\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is used if **z** is a residual vector of a linear regression model. The expectation and variance of *MC* given under the randomized hypothesis is as follows (see Tiefelsdorf and Griffith, 2007):

$$E[MC] = \frac{tr[\mathbf{MWM}]}{N-K} , \tag{2-43}$$

$$Var[MC] = 2\frac{(N-K)tr[(\mathbf{MWM})^2] - tr[\mathbf{MWM}]^2}{(N-K)^2(N-K+2)} . \tag{2-44}$$

where *K* is the number of variables in **X**. In this case, *MC* > E[*MC*], *MC* = E[*MC*], *MC* < E[*MC*] imply positive spatial dependence, no spatial dependence, and no spatial dependence, respectively.

The significance of *MC* can be tested using Eqs.(2-43) and (2-44).

Consider the eigen-decomposition of $\mathbf{MWM}$ (i.e., $\mathbf{MWM} \rightarrow \mathbf{E}_{full}\mathbf{\Lambda}_{full}\mathbf{E}_{full}'$, where $\mathbf{\Lambda}_{full}$ is the diagonal matrix of the eigenvalues $\lambda_1,\ldots\lambda_l,\ldots\lambda_N$ and $\mathbf{E}_{full} = \{\mathbf{e}_1,\ldots\mathbf{e}_l,\ldots\mathbf{e}_N\}$ is a matrix of the eigenvectors). When $\mathbf{M} = \mathbf{I}-\mathbf{11}'/N$, only one eigenvalue in $\mathbf{\Lambda}_{full}$ indicates 0, and the eigenvectors in $\mathbf{E}_{full}$ are mutually orthogonal ($\mathbf{e}_l'\mathbf{e}_l = 1$ and $\mathbf{e}_l'\mathbf{e}_l = 0$) and orthogonal to $\mathbf{1}$ ($\mathbf{1}'\mathbf{e}_l = \mathbf{e}_l'\mathbf{1} = \mathbf{0}$). On the other hand, when $\mathbf{M} = \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, K (rank of $\mathbf{X}$) eigenvectors indicate 0, and the eigenvectors in $\mathbf{E}_{full}$ are mutually orthogonal and orthogonal to $\mathbf{X}$ ($\mathbf{X}'\mathbf{e}_l = \mathbf{e}_l'\mathbf{X} = \mathbf{0}$) (Griffith, 2003).

In both cases, $\mathbf{e}_l'\mathbf{1} = \mathbf{0}$ is satisfied, provided that $\mathbf{X}$ includes a constant. This means that the orthogonality also implies no correlation among $\{\mathbf{e}_1,\ldots\mathbf{e}_l,\ldots\mathbf{e}_N\}$. More precisely, using $\mathbf{e}_l'\mathbf{1} = \mathbf{0}$, the numerator of the correlation coefficient between $\mathbf{e}_l$ and $\mathbf{e}_m$ results in the equation representing orthogonality: $(\mathbf{e}_l - \mathbf{1}\mathbf{e}_l'\mathbf{1})'(\mathbf{e}_m - \mathbf{1}\mathbf{e}_m'\mathbf{1}) \rightarrow \mathbf{e}_l'\mathbf{e}_m$. Thus, the eigenvectors are both orthogonal and uncorrelated.

Calculate *MC* of $\mathbf{e}_l$ as

$$MC[\mathbf{e}_l] = \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}}\frac{\mathbf{e}_l'\mathbf{MWMe}_l}{\mathbf{e}_l'\mathbf{Me}_l}$$

$$= \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}}\frac{\mathbf{e}_l'\mathbf{E}_{full}\mathbf{\Lambda}_{full}\mathbf{E}_{full}'\mathbf{e}_l}{(\mathbf{Me}_l)'(\mathbf{Me}_l)}$$

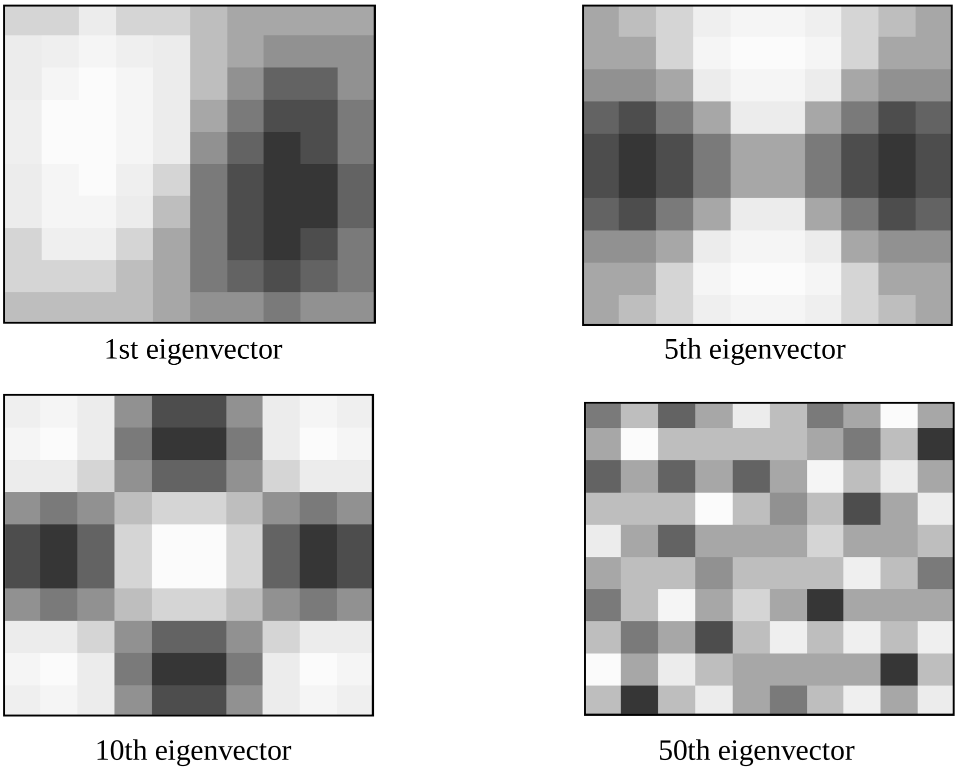
$$= \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}}\frac{[0,\cdots\mathbf{e}_l'\mathbf{e}_l,\cdots 0]\begin{bmatrix}\lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_l & & \\ & & & \ddots & \\ & & & & \lambda_N\end{bmatrix}\begin{bmatrix}0\\ \vdots\\ \mathbf{e}_l'\mathbf{e}_l\\ \vdots\\ 0\end{bmatrix}}{\mathbf{e}_l'\mathbf{e}_l}$$

$$= \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}}\lambda_l \,. \tag{2-45}$$

Here, we use the orthogonality among $\mathbf{e}_1,\ldots\mathbf{e}_l,\ldots\mathbf{e}_N$ (i.e., $\mathbf{e}_l'\mathbf{e}_l = 1$ and $\mathbf{e}_l'\mathbf{e}_l = 0$) and the property that $\mathbf{Me}_l = \mathbf{e}_l$[1]. Eq.(2-45) suggests that the *MC*s of $\mathbf{e}_1,\ldots\mathbf{e}_l,\ldots\mathbf{e}_N$ are proportional to their corresponding eigenvalues, $\lambda_1,\ldots\lambda_l,\ldots\lambda_N$.

[1] When $\mathbf{M} = \mathbf{I}-\mathbf{11}'/N$, $\mathbf{Me}_l = (\mathbf{I}-\mathbf{11}'/N)\mathbf{e}_l = \mathbf{e}_l - (\mathbf{11}'\mathbf{e}_l)/N = \mathbf{e}_l$. Here, $\mathbf{1}'\mathbf{e}_l = \mathbf{0}$ is used for the expansion. On the other hand, when $\mathbf{M} = \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{Me}_l = (\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{e}_l = \mathbf{e}_l - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}_l = \mathbf{e}_l$. Here, $\mathbf{X}'\mathbf{e}_l = \mathbf{0}$ is used for the expansion. Thus, $\mathbf{Me}_l = \mathbf{e}_l$.

Consequently, the eigenvectors provide distinct (i.e., orthogonal and uncorrelated) map pattern descriptions of latent spatial dependence, with each level being indexed by an *MC* that is proportional to its corresponding eigenvalue (Griffith, 2003). Specifically, $\mathbf{e}_1$ is the set of numerical values with the largest positive *MC* (maximum positive spatial dependence) achievable by any set of real numbers for the spatial arrangement defined by **C**. Then, $\mathbf{e}_2$ is the set of values with the largest positive *MC* uncorrelated with, and orthogonal to, $\mathbf{e}_1$, and $\mathbf{e}_N$ is the set of numerical values with the largest negative *MC* (maximum negative spatial dependence) achievable that is uncorrelated with, and orthogonal to, $\mathbf{e}_1$,... $\mathbf{e}_l$,... $\mathbf{e}_{N-1}$ (see Fig.2-5).



1st eigenvector

5th eigenvector

10th eigenvector

50th eigenvector

Figure 2-5: Images of the eigenvectors

Here, the 1st, 5th, 10th, and 50th eigenvectors of **MWM** ($\mathbf{M} = \mathbf{I} - \mathbf{11}'/N$) defined on a 10 by 10 gridded space are plotted.

Suppose that $\mathbf{E\gamma} = \mathbf{e}_1\gamma_1 +\dots \mathbf{e}_l\gamma_l + \dots \mathbf{e}_L\gamma_L$, where $\mathbf{E}$ is a matrix composed of L-eigenvectors in $\mathbf{E}_{full}$ ($L < N$), where l is the index of the L-eigenvectors, and γ_l is the weight for the l-th eigenvector, then, the MC of $\mathbf{E\gamma}$ is given, using Eq.(2-45), as

$$MC[\mathbf{E\gamma}] = \frac{N}{\mathbf{1'W1}}\frac{(\mathbf{e}_1\gamma_1 + \cdots \mathbf{e}_L\gamma_L)'\mathbf{MWM}(\mathbf{e}_1\gamma_1 + \cdots \mathbf{e}_L\gamma_L)}{(\mathbf{e}_1\gamma_1 + \cdots \mathbf{e}_L\gamma_L)'\mathbf{M}(\mathbf{e}_1\gamma_1 + \cdots \mathbf{e}_L\gamma_L)}$$

$$= \frac{N}{\mathbf{1'W1}}\frac{(\mathbf{e}_1\gamma_1 + \cdots \mathbf{e}_L\gamma_L)'\mathbf{E}_{full}\mathbf{\Lambda}_{full}\mathbf{E}'_{full}(\mathbf{e}_1\gamma_1 + \cdots \mathbf{e}_L\gamma_L)}{(\mathbf{e}_1\gamma_1 + \cdots \mathbf{e}_L\gamma_L)'\mathbf{M}(\mathbf{e}_1\gamma_1 + \cdots \mathbf{e}_L\gamma_L)}$$

$$= \frac{N}{\mathbf{1'W1}}\frac{[\gamma_1\mathbf{e}_1'\mathbf{e}_1 \cdots \gamma_L\mathbf{e}_L'\mathbf{e}_L, 0 \cdots 0]\begin{bmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_L & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & \lambda_N \end{bmatrix}\begin{bmatrix} \gamma_1\mathbf{e}_1'\mathbf{e}_1 \\ \vdots \\ \gamma_L\mathbf{e}_L'\mathbf{e}_L \\ 0 \\ \vdots \\ 0 \end{bmatrix}}{\gamma_1^2\mathbf{e}_1'\mathbf{e}_1 + \cdots \gamma_L^2\mathbf{e}_L'\mathbf{e}_L},$$

$$= \frac{N}{\mathbf{1'W1}}\frac{\sum_l \gamma_l^2\lambda_l}{\sum_l \gamma_l^2} = \frac{1}{\sum_l \gamma_l^2}\sum_l \gamma_l^2 MC[\mathbf{e}_l], \tag{2-46}$$

where $\lambda_1,\dots \lambda_l,\dots, \lambda_L$ are the eigenvalues corresponding to the L eigenvectors. Eq.(2-46) shows that the MC of $\mathbf{E\gamma}$ is given by the weighted average of the MCs of the eigenvectors. In other words, not only the eigenvectors themselves, but also their linear combination, $\mathbf{E\gamma}$, describes the map pattern description of latent spatial dependence explained by MC.

Note that decompositions of MC, such as Eq.(2-46), have often been discussed. For instance, the local MC (or local Moran's I statistics), defined by Eq.(2-47), is a local indicator of spatial association (LISA: Anselin, 1995), and can be considered as a decomposition of the (global) MC.

$$MC_i = \frac{N(z_i - \bar{z})}{\sum_i (z_i - \bar{z})^2}\sum_j w_{i,j}(z_j - \bar{z}). \tag{2-47}$$

The local MC is used to test whether local spatial dependence is present around the i-th sample. The local MC is expressed using matrix notation as follows (Tiefelsdorf, 1998):

$$MC_i = N\frac{\mathbf{z'M}(s_i\mathbf{W}_i)\mathbf{Mz}}{\mathbf{z'Mz}}, \tag{2-48}$$

where $\mathbf{W}_i$ is $\mathbf{W}$ with all elements replaced with zeros, except for the i-th row and column, and s_i is the scaling parameter for the i-th sample. The global MC is given by the local MCs as

$$\frac{1}{\mathbf{1}'\mathbf{W}\mathbf{1}}\sum_i MC_i = \frac{N}{\mathbf{1}'\mathbf{W}\mathbf{1}}\sum_i \frac{\mathbf{z}'\mathbf{M}(s_i\mathbf{W}_i)\mathbf{M}\mathbf{z}}{\mathbf{z}'\mathbf{M}\mathbf{z}},$$

$$= MC. \qquad (2\text{-}49)$$

Thus, similar to Eq.(2-46), the local *MC* is a decomposition of the global *MC*.

2.3.2.2. Eigenvector spatial filtering

The MC-based spatial filtering approach, called eigenvector spatial filtering (ESF), captures spatial dependence using **Eγ**. The basic linear model of ESF is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}). \qquad (2\text{-}50)$$

Since Eq.(2-50) is identical to the standard linear regression model, **β** and **γ** can be estimated using the OLS estimation. Provided that **M** = **I**–**11**′/*N*, the estimates of **β** and **γ** are as follows:

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{pmatrix} = \begin{pmatrix}\mathbf{X}'\mathbf{X} & \mathbf{E}'\mathbf{X}\\ \mathbf{X}'\mathbf{E} & \mathbf{I}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{E}'\mathbf{y}\end{pmatrix}, \qquad (2\text{-}51)$$

and their variances are given as

$$Var\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{pmatrix} = \sigma^2\begin{pmatrix}\mathbf{X}'\mathbf{X} & \mathbf{E}'\mathbf{X}\\ \mathbf{X}'\mathbf{E} & \mathbf{I}\end{pmatrix}^{-1}. \qquad (2\text{-}52)$$

On the other hand, if **M** = **I**–**X(X′X)**$^{-1}$**X′**, which imposes **X′e**$_i$ = **e**$_i$**′X** = **0**, their estimators and variances yield

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{pmatrix} = \begin{pmatrix}\mathbf{X}'\mathbf{X} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{E}'\mathbf{y}\end{pmatrix}$$

$$= \begin{pmatrix}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\\ \mathbf{E}'\mathbf{y}\end{pmatrix}, \qquad (2\text{-}53)$$

$$Var\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{pmatrix} = \sigma^2\begin{pmatrix}\mathbf{X}'\mathbf{X} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}^{-1}$$

$$= \sigma^2\begin{pmatrix}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}. \qquad (2\text{-}54)$$

Eq.(2-52) suggests that, if **M** = **I**–**11**′/*N*, the correlation between **X** and **E** inflates the variances of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$. Hence, **M** = **I**–**X(X′X)**$^{-1}$**X′**, which imposes no correlation between **X** and **E**, seems helpful, for example, when identifying a model accurately (without aggravating the instability of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$). On the other hand, considering the spatially dependent component correlated with **X** is useful in reducing the omitted variable bias, which is the bias due to factors that cannot be considered in the model (see LeSage and Pace, 2009). Thus, the assumption of $\mathbf{M} = \mathbf{I} – \mathbf{11}'/N$ seems helpful when testing the significance of $\hat{\boldsymbol{\beta}}$ (and $\hat{\boldsymbol{\gamma}}$), considering the omitted variable bias problem.

ESF is implemented as follows: (i) $\mathbf{E}_{full}$ and $\boldsymbol{\Lambda}_{full}$ are extracted from **MWM**; (ii) The eigenvectors responding to small eigenvalues, that is, small *MC*s (see Eq.2-45) are removed; and, (iii) Significant eigenvectors are chosen by applying an OLS-based stepwise variable selection procedure for Eq.(2-50). Step (ii) is conducted by removing the eigenvectors with eigenvalues that are small or of the wrong nature. For example, the criterion $MC[\mathbf{e}_l]/MC[\mathbf{e}_1] > 0.25$ has been used to analyze positive spatially dependent components (Griffith, 2003), while $MC[\mathbf{e}_l]/MC[\mathbf{e}_1] < –0.25$ has also been applied when analyzing negative spatial dependent components (e.g., Griffith, 2006). Step (iii) is done by maximizing the model accuracy (e.g., adjusted R^2 maximization) or minimizing residual spatial dependence (measure: *MC*). In each step of the stepwise selection procedure, the eigenvector that maximizes the accuracy or minimizes the residual spatial dependence is introduced into the model. In the latter case, a stopping rule is needed. Tiefelsdorf and Griffith (2007) set this rule using $|\mathrm{MC}[\boldsymbol{\varepsilon}]| < 0.01$ (i.e., the procedure is conducted until $|MC|$ of $\boldsymbol{\varepsilon}$ is less than 0.01).

As in the G_i statistics-based approach, the ESF is also simple (Griffith, 2003). The basic ESF model is identical to the standard linear model, and is therefore easy to implement and extend, for example, by combining it with standard non-spatial models (e.g., Poisson and logistic regression: see, e.g., Griffith, 2002; 2004a). Another advantage of the ESF is its effectiveness in capturing spatial dependence. Tiefelsdorf and Griffith (2007) show that the ESF effectively removes spatial dependence from residuals, and Griffith (2006) demonstrates its effectiveness in analyzing negative spatial dependent components hidden by dominant positive spatial dependence. Thayn and Simanis (2013) show that the ESF reduces spatial misspecification errors, increases the strength of a model fit, frequently increases the normality of the model residuals, and can increase the homoscedasticity of model residuals. Hughes and Haran (2013) demonstrate the usefulness of an ESF-based generalized linear mixed model for a spatial dependence analysis that considers spatial confounding. Spatial confounding occurs when variance inflation, caused by collinearity, is introduced between spatial processes in the explanatory variables and the spatial process in the response variables (see e.g., Paciorek, 2010).

Applications of ESF are increasing because of its practicality, expandability, and other appealing factors. For instance, Chun (2008) and Griffith (2009) use ESF for spatial interaction analyses (see also, Tiefelsdorf, 2003). Pecci and Pontarollo (2010), Patuelli *et al*. (2011), and Cuaresma and Feldkircher (2013) employ it for economic analyses. Griffith and Peres-Neto (2006) and Jacob *et al*. (2008) use it for ecological analyses, and Moniruzzaman and Paez (2012) use it for urban design analyses. Thus, ESF has become a popular way to address spatial dependence (Pace *et al*., 2011).

2.4. Geographically weighted regression (GWR)

While the geostatistical model and the spatial filtering models capture spatially dependence, GWR captures spatial heterogeneity by using spatially varying parameters. The basic model is given as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \tag{2-55}$$

where $\boldsymbol{\beta}_i$ is a vector of parameters that depend on location s_i. GWR estimates $\boldsymbol{\beta}_i$ by imposing the constraint that the $\boldsymbol{\beta}_i$s in each site are strongly related to nearby observations. In particular, the estimates for site s_i are given by Eq.(2-56). Here, a WLS estimator with larger weights is assigned for nearby samples, and smaller weights are assigned for more distant samples:

$$\boldsymbol{\beta}_i = (\mathbf{X}'\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_i\mathbf{y}, \tag{2-56}$$

where $\mathbf{W}_i$ is a diagonal matrix in which the *j*-th element describes the weight of the *j*-th sample for $\boldsymbol{\beta}_i$. The weight is modeled by a distance-decay function, $k(h_{i,j})$. The standard weighting functions are as follows:

Gaussian model $$k(h_{i,j}) = \exp\left(-\frac{h_{i,j}^2}{r^2}\right), \qquad (2\text{-}57)$$

Bi-square model $$k(h_{i,j}) = \begin{cases} \left[1 - \frac{h_{i,j}^2}{r^2}\right]^2 & \mathit{if}\ h_{i,j} < r \\ 0 & \mathit{otherwuse} \end{cases}, \qquad (2\text{-}58)$$

Tri-cube model $$k(h_{i,j}) = \begin{cases} \left[1 - \frac{h_{i,j}^3}{r^3}\right]^3 & \mathit{if}\ h_{i,j} < r \\ 0 & \mathit{otherwuse} \end{cases}, \qquad (2\text{-}59)$$

where r is a bandwidth parameter. Small r means the existence of small-scale spatial heterogeneity, and as r increases, the estimates of $\boldsymbol{\beta}_i$ asymptotically converge on the standard OLS estimates (Fotheringham *et al*., 2002).

The value of r is estimated by cross-validation or by minimizing the corrected Akaike information criterion (AICc). The cross-validation minimizes the cross-validation score, which is defined as

$$\sum_i [y_i - \hat{y}_{-i}]^2, \qquad (2\text{-}60)$$

where $\hat{y}_{-i}$ is given as

$$\hat{y}_{-i} = \mathbf{x}_i' \boldsymbol{\beta}_{-i}. \qquad (2\text{-}61)$$

Here, $\mathbf{x}_i$ is a vector of explanatory variables at s_i, and $\boldsymbol{\beta}_{-i}$ is defined as

$$\boldsymbol{\beta}_{-i} = (\mathbf{X}_{-i}' \mathbf{W}_{-i} \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}' \mathbf{W}_{-i} \mathbf{y}_{-i}, \qquad (2\text{-}62)$$

where $\mathbf{X}_{-i}$, $\mathbf{y}_{-i}$, and $\mathbf{W}_{-i}$ are $\mathbf{X}$, $\mathbf{y}$, and $\mathbf{W}_i$, respectively, but without their i-th elements. On the other hand, the AICc minimization of the GWR model is defined as

$$AICc = 2N \log(\hat{\sigma}^2) + N \log(2\pi) + N\left(\frac{N + trace[\mathbf{H}]}{N + 2 - trace[\mathbf{H}]}\right), \qquad (2\text{-}63)$$

where $\mathbf{H}$ is a matrix in which the i-th row, $\mathbf{h}_i$, is defined as

$$\mathbf{h}_i = \mathbf{x}_i' (\mathbf{X}' \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}_i, \qquad (2\text{-}64)$$

and $\hat{\sigma}^2$ is

$$\hat{\sigma}^2 = \frac{\sum_i (y_i - \mathbf{x}_i' \boldsymbol{\beta}_i)}{N - [2trace[\mathbf{H}] - trace[\mathbf{H}'\mathbf{H}]]}. \qquad (2\text{-}65)$$

By minimizing either Eq.(2-60) or Eq.(2-63), we can identify the optimal *r*. In both cases, *r* is estimated numerically.

For more details about GWR, see e.g., Fotheringham *et al.* (2002) and Wheeler and Paez (2010).

3. Areal Interpolation Problem: A GWR-based Approach

Chapter 3 and 4 discuss change of support problems for areal data. Specifically, chapter 3 discusses the areal interpolation problem, and the discussion is arranged for the MAUP in chapter 4.

This chapter establishes a GWR-based areal interpolation method by combining GWR with a standard geostatistical areal interpolation approach. The effectiveness of the constructed method is examined by applying a simulation study. After that, unknown municipal-level road stock data, which possibly form an important index in achieving a stock-type society, are interpolated by applying this method for the known prefectural-level road stock data. The advantages and disadvantages of the methods of statistical areal interpolation, including the proposed method, are discussed in this empirical study, and an additional GWR-based approach is presented based on this discussion.

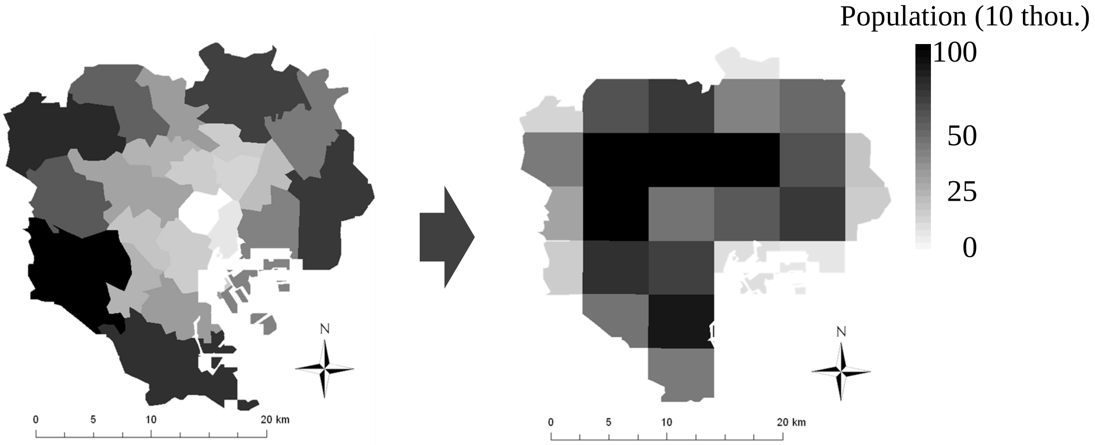


Figure 3-1: Image of areal interpolation

3.1. Introduction

3.1.1. Review of areal interpolation studies

Areal interpolation (aggregation unit conversion) has been discussed extensively among geographers (e.g., Wright, 1936; Tobler, 1979; Goodchild and Lam, 1980). Primal geographical areal interpolation methods are as follows (see also, Fig. 3-2): the areal weighting interpolation method (Wright, 1936) that interpolates data by a proportional allotment using areal weights; the point-in-polygon method (e.g., Sadahiro, 2000) that aggregates areal data that are replaced with point data; the dasymetric method (Wright, 1936; Fisher and Langford, 1995) that applies proportional allotment whose allotment weights are determined using supplementary data (e.g., population distribution only for residential area); the pycnophylactic method (Tobler, 1979), which models data using a spatially smooth function first, and aggregates the smoothed data after that; and the regression-based methods (Flowerdew and Green, 1989, 1992, 1994), which are based on the Expectation Maximization (EM) algorithm. Images of these methods are summarized in Fig.3-2. Among these methods, the dasymetric method, the pycnophilactic method, and the regression-based methods have been developed significantly.

The dasymetric method has been discussed in quantitative geography, particularly, after its efficiency was recognized in some comparative studies in the 1990s (e.g., Fisher and Langford, 1995, 1996; Mrozinski and Cromley, 1999). Extensions of the dasymetric method has been discussed extensively (e.g., Xie, 1995; Eicher and Brewer, 2001; Mennis and Hultgren, 2006; Reibel and Agrawal, 2007; Kim and Yao, 2010; Zhang and Qui, 2011; Schroeder and Riper, 2013; Langford, 2013), and some of them reveal that areal interpolation accuracy heavily depends on the supplementary data quality. Thus, the dasymetric method has been discussed with the focus on how supplementary data are considered.

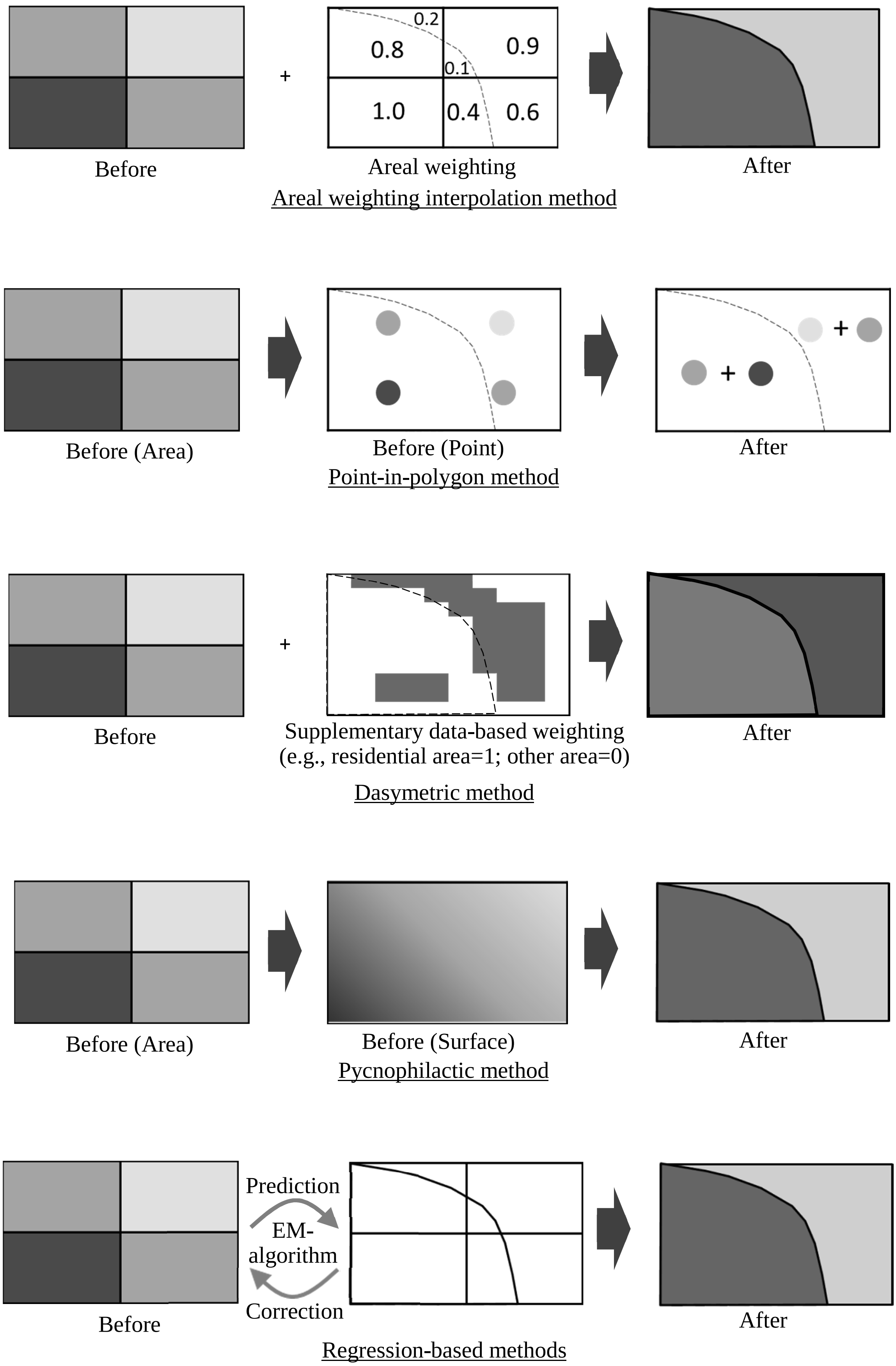


Figure 3-2: Images of the primal areal interpolation methods

The pycnophilactic method, which conducts spatially smooth interpolation, has been discussed mainly focusing on its theoretical aspects rather than use of supplementary data (e.g., Brillinger, 1990, 1994; Muller *et al*., 1997). In somewhat deferent context, Kyriakidis (2004) proposes a geostatistical method, which is called area-to-point (ATP) kriging. While ATP kriging also provides spatially smooth interpolation result, it is superior to the pycnophilactic method in that, as with conventional kriging, it minimizes MSPE. Some extensions of ATP kriging and the other kriging-based methods have been discussed in geostatistics (e.g., Kyriakidis, 2004; Yoo and Kyriakidis, 2006; Gooverts, 2006; Gotway and Young, 2007; Yoo, *et al*., 2010; Murakami and Tsutsumi, 2012). As an interesting finding, Yoo *et al*. (2010) shows that the pycnophilactic method and the ATP kriging solve similar problems, and, as a result, their interpolation results possibly be very similar.

Finally, the regression-based method has been extended for hierarchical Bayesian modeling (e.g., Mugglin and Carlin, 1998; Mugglin *et al*., 1999, 2000). The hierarchical modeling-based approach is quite flexible. For example, Sahu *et al*. (2010) consider both spatial dependence and spatial heterogeneity, and, also, they consider multiple supplementary data. Thus, the hierarchical Bayesian areal interpolation is a recent hot topic in geostatistics (e.g., Gelfand, 2010; Sahu *et al*., 2010; Cressie and Wikle, 2011; Berrocal *et al*., 2012).

3.1.2. Fundamentals of areal interpolation

This section discusses the fundamentals of areal interpolation. Here, I assume that unobserved variables in non-aggregate level units are interpolated using observed variables given in each aggregate level unit.

Areal interpolation is defined as a spatial interpolation that considers an aggregation mechanism, which is defined as

$$\bar{\mathbf{y}}^{volume} = \mathbf{N}^{volume}\mathbf{y}^{volume}. \tag{3-1}$$

Here, $\mathbf{y}^{volume}$ is a vector of the unknown count/volume (extensive) variables given in the non-aggregate level units, $\bar{\mathbf{y}}^{volume}$ is a vector of the known count variables given in the aggregate level units, and $\mathbf{N}^{volume}$ is an aggregation matrix. Eq.(3-1) implies that the aggregations of the unknown variables in

$\mathbf{y}^{volume}$ must be equal to the known variables in $\bar{\mathbf{y}}^{volume}$ (for example, the aggregations of municipal populations must be equal to actual prefectural populations). This property is called the volume preserving property (or the pycnophilactic property/the mass balance property), and is one of the most basic properties that must be considered in areal interpolation (Lam, 1983).

Eq.(3-1) can be expanded to describe the aggregation mechanism of density (intensive) variables, as follows:

$$\bar{\mathbf{y}} = \mathbf{N}\mathbf{y}\,, \tag{3-2}$$

$$\bar{\mathbf{y}} = \bar{\mathbf{M}}^{-1}\bar{\mathbf{y}}^{volume}, \qquad \mathbf{y} = \mathbf{M}^{-1}\mathbf{y}^{volume}, \qquad \mathbf{N} = \bar{\mathbf{M}}^{-1}\mathbf{N}^{volume}\mathbf{M}\,,$$

where $\mathbf{M}$ is a diagonal matrix, the elements of which are weights for the non-aggregate level units, and $\bar{\mathbf{M}}$, which has the same elements, but for the aggregate level units. Eq.(3-2) means that the aggregated values of $\mathbf{y}$ (the non-aggregate level density variables) must be equal to $\bar{\mathbf{y}}$ (the aggregate level density variables). For example, when the elements in $\mathbf{y}$ are population densities, the elements in $\mathbf{M}$ must be the areas of each unit. Because Eq.(3-1) and Eq.(3-2) are identical, the volume preserving property is satisfied if either equation is satisfied. As I will discuss later, the dasymetric method uses Eq.(3-1) for the volume preserving property, while many of the geostatistical studies use Eq.(3-2).

The interpolation equation of the dasymetric method (and the areal weighting interpolation method) is given by multiplying the generalized inverse matrix (e.g., Menke, 1989) of $\mathbf{N}^{volume}$, which minimizes the norm of $\mathbf{y}^{volume}$, from the left side of Eq.(3-1), as follows:

$$\hat{\mathbf{y}}^{volume} = \mathbf{N}^{volume\prime}(\mathbf{N}^{volume\prime}\mathbf{N}^{volume})^{-1}\bar{\mathbf{y}}^{volume}, \tag{3-3}$$

where the generalized inverse matrix, $\mathbf{N}^{volume\prime}(\mathbf{N}^{volume\prime}\mathbf{N}^{volume})^{-1}$, becomes a matrix describing a proportional distribution ratio. Eq.(3-3) shows that the dasymetric method provides a minimum-length solution to Eq.(3-1) (Kyriakidis and Yoo, 2005). It also implies that extensions of the dasymetric method are based on the minimum-length solution.

A problem of the minimum-length solution is that it cannot consider spatial dependence.[1] A possible extension that can do so is to model $\mathbf{y}$ in Eq.(3.2) using a geostatistical model, as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{C})\,. \tag{3-4}$$

Substituting Eq.(3-4) into Eq.(3-2) yields the following equation:

[1] The dasymetric method captures (spatial) heterogeneity by applying supplemental data.

$$\begin{pmatrix}\mathbf{y}\\ \bar{\mathbf{y}}\end{pmatrix}=\begin{pmatrix}\mathbf{X\beta}\\ \mathbf{NX\beta}\end{pmatrix}+\begin{pmatrix}\boldsymbol{\varepsilon}\\ \mathbf{N}\boldsymbol{\varepsilon}\end{pmatrix}, \qquad \begin{pmatrix}\boldsymbol{\varepsilon}\\ \mathbf{N}\boldsymbol{\varepsilon}\end{pmatrix}\sim N\left[\begin{pmatrix}\mathbf{0}\\ \bar{\mathbf{0}}\end{pmatrix}\begin{pmatrix}\mathbf{C} & \mathbf{CN}'\\ \mathbf{NC} & \mathbf{NCN}'\end{pmatrix}\right], \tag{3-5}$$

where $\bar{\mathbf{0}}$ (=**N0**) is a vector of zeros. The conditional expectation of **y** is given based on Eq.(3-5) as

$$\hat{\mathbf{y}}=\mathbf{X\beta}+\mathbf{CN}'(\mathbf{NCN}')^{-1}(\bar{\mathbf{y}}-\mathbf{NX\beta}). \tag{3-6}$$

Thus, **y** can be interpolated using Eq.(3-6). As with the standard geostatistics, this equation minimizes the MSPE. In addition, $\hat{\mathbf{y}}$, as given by Eq.(3-6), satisfies the volume preserving property. This is easily confirmed by substituting $\hat{\mathbf{y}}$ into Eq.(3-2):

$$\begin{aligned}\bar{\mathbf{y}}&=\mathbf{N}(\mathbf{X\beta}+\mathbf{CN}'(\mathbf{NCN}')^{-1}(\bar{\mathbf{y}}-\mathbf{NX\beta}))\\ &=\mathbf{NX\beta}+\mathbf{NCN}'(\mathbf{NCN}')^{-1}(\bar{\mathbf{y}}-\mathbf{NX\beta})\\ &=\mathbf{NX\beta}+(\bar{\mathbf{y}}-\mathbf{NX\beta}),\\ &=\bar{\mathbf{y}}.\end{aligned} \tag{3-7}$$

Many geostatistical areal interpolation methods, including the ATP kriging and some hierarchical Bayesian models, use Eq.(3-6) as their basic interpolation equation.

Thus, this study considers developing an areal interpolation method that is consistent with the aforementioned fundamentals.

3.2. GWR-based areal interpolation

3.2.1. Background

While geostatistics is a sub-field in spatial statistics, non-geostatistical spatial statistical models, including the GWR and the spatial filter models, have rarely been applied for areal interpolation. Exceptionally, Lo (2008) and Lin *et al*. (2011) applied GWR for areal interpolation. However, their methods do not minimize MSPE. In the other words, their methods are insistent with the discussions in geostatistics (see §3.1.2).

This section develops a GWR-based areal interpolation method that explicitly minimizes MSPE.

3.2.2. Model

Following geostatistical approach, which applies a standard geostatistical model for the non-aggregate level model, this study applies the standard GWR model Eq.(2-55) for the non-aggregate level model. Namely, I assume Eq.(3-8) as the non-aggregate level model:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{M}), \tag{3-8}$$

$$\boldsymbol{\mu} = \left\langle \mathbf{x}_k' \boldsymbol{\beta}_k \right\rangle, \tag{3-9}$$

where $\mathbf{x}_k$ is a vector of the explanatory variables in the k-th non-aggregate level unit, $\boldsymbol{\beta}_k$ is a parameter vector for the k-th unit, and $\langle x \rangle$ denotes a vector whose elements are given by x. In order to consider the weights (e.g., areas) of each unit, variances of $\boldsymbol{\varepsilon}$ are weighted by **M**. By substituting Eq.(3-8) into Eq.(3-2), my full-model is derived as

$$\begin{pmatrix} \mathbf{y} \\ \bar{\mathbf{y}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{N}\boldsymbol{\mu} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \mathbf{N}\boldsymbol{\varepsilon} \end{pmatrix}, \qquad \begin{pmatrix} \boldsymbol{\varepsilon} \\ \mathbf{N}\boldsymbol{\varepsilon} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \bar{\mathbf{0}} \end{pmatrix} \begin{pmatrix} \mathbf{M} & \mathbf{MN}' \\ \mathbf{NM} & \mathbf{NMN}' \end{pmatrix}\right]. \tag{3-10}$$

BLUP of **y**, which minimizes MSPE, is given, as same as the geostatistical models, as

$$\hat{\mathbf{y}} = \boldsymbol{\mu} + \mathbf{MN}'(\mathbf{NMN}')^{-1}(\bar{\mathbf{y}} - \mathbf{N}\boldsymbol{\mu}). \tag{3-11}$$

While geostatistics controls spatial dependence by parameterizing its variance-covariance matrix **C** using a distance function, the GWR-based model controls spatial heterogeneity by the non-aggregate level spatially varying coefficient $\boldsymbol{\beta}_k$ in $\boldsymbol{\mu}$. Note that, as Fotheringham *et al.* (2002) pointed out, GWR effectively captures spatial dependence, and, accordingly, the proposed GWR-based method would capture spatial dependence too.

The proposed model Eq.(3-10) can be summarized as $\bar{\mathbf{y}} = \mathbf{N}\mathbf{y}_{GWR}$, where $\mathbf{y}_{GWR}$ is **y** that is given by Eq.(3-8), and the predictor Eq.(3-11) is a MSPE-based solution of $\bar{\mathbf{y}} = \mathbf{N}\mathbf{y}_{GWR}$. On the other hand, the dasymetric method provides the minimum-length solution of $\bar{\mathbf{y}} = \mathbf{N}\mathbf{y}$ (or Eq.3-1). Accordingly, the proposed model can be considered as an extension of the dasymetric method that considers spatial heterogeneity and minimizes MSPE.

The predictor $\hat{\mathbf{y}}$ satisfies the volume preserving property. It is proved as

$$
\begin{aligned}
\bar{\mathbf{y}} &= \mathbf{N}(\boldsymbol{\mu} + \mathbf{MN}'(\mathbf{NMN}')^{-1}(\bar{\mathbf{y}} - \mathbf{N}\boldsymbol{\mu})) \\
&= \mathbf{N}\boldsymbol{\mu} + \mathbf{NMN}'(\mathbf{NMN}')^{-1}(\bar{\mathbf{y}} - \mathbf{N}\boldsymbol{\mu}) \\
&= \mathbf{NM} + (\bar{\mathbf{y}} - \mathbf{NM}) \\
&= \bar{\mathbf{y}}\,. \qquad (3\text{-}12)
\end{aligned}
$$

3.2.3. Estimation

Parameters in our model Eqs.(3-10) must be estimated on condition that **y** is unknown and $\bar{\mathbf{y}}$ is known. Some geostatistical studies prove consistency of aggregate level model-based parameter estimation (e.g., Nagle *et al*., 2011). Hence, this study also considers estimating parameters using the aggregate level model in Eq.(3-10), i.e.,

$$\bar{\mathbf{y}} = \mathbf{N}\boldsymbol{\mu} + \mathbf{N}\boldsymbol{\varepsilon}, \qquad \mathbf{N}\boldsymbol{\varepsilon} \sim N(\bar{\mathbf{0}}, \mathbf{NMN}'), \qquad (3\text{-}13)$$

The GLM estimator of $\boldsymbol{\beta}_k$ is given as

$$\hat{\boldsymbol{\beta}}_k = (\bar{\mathbf{X}}_k'(\mathbf{NMN})^{-1}\bar{\mathbf{X}}_k)^{-1}\bar{\mathbf{X}}_k'(\mathbf{NMN})^{-1}\bar{\mathbf{y}}_k\,. \qquad (3\text{-}14)$$

$$\bar{\mathbf{X}}_k = \overline{\mathbf{W}_k}^{-1/2}\mathbf{NX}, \qquad \bar{\mathbf{y}}_k = \overline{\mathbf{W}_k}^{-1/2}\mathbf{N}\bar{\mathbf{y}}, \qquad \overline{\mathbf{W}_k} = \mathbf{NW}_k\mathbf{N}'.$$

$\hat{\boldsymbol{\beta}}_k$ is identical to the estimator of the standard GWR. Chapter 4 discusses properties of the estimator from the viewpoint of MAUP.

The proposed method is implemented as follows: (i) The optimal bandwidth parameter in $\mathbf{W}_k$ is estimated via a cross-validation based on Eq.(3-13); (ii) $\hat{\boldsymbol{\beta}}_k$ is estimated by substituting the calibrated bandwidth into Eq.(3-14); (iii) $\hat{\mathbf{y}}$ is predicted by substituting the estimated $\hat{\boldsymbol{\beta}}_k$ into Eq.(3-11).

Following ATP kriging (see §3.1), I refer to the proposed method ATP GWR.

3.2.4. Non-negative constraint

One of the drawbacks of ATP GWR is that it allows negative interpolated values whereas negative interpolated values are physically impossible in most cases (Yoo *et al.*, 2010). For example, interpolated population must be non-negative. Thus, this section considers introducing a non-negative constraint in our model.

The proposed method minimizes MSPE of **y**, which is modeled by Eq.(3-8), on condition that Eq.(3-2). The minimization problem is written as

$$\arg\min_{\mathbf{y}} \quad (\mathbf{y}-\boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{y}-\boldsymbol{\mu}), \tag{3-15}$$

$$\text{s.t.} \quad \mathbf{Ny} = \overline{\mathbf{y}}. \tag{3-16}$$

Hence, by solving the problem of minimizing Eq.(3-15) on condition that Eq.(3-16) and Eq.(3-17), a non-negative constraint can be introduced.

$$\mathbf{y} > \mathbf{0} \tag{3-17}$$

Let expand Eq.(3-15), and rewrite the minimization problem as

$$\arg\min_{\mathbf{y}} \quad \mathbf{y}'\mathbf{M}^{-1}\mathbf{y} - 2\boldsymbol{\mu}'\mathbf{M}^{-1}\mathbf{y} + \boldsymbol{\mu}'\mathbf{M}^{-1}\boldsymbol{\mu}, \tag{3-18}$$

$$\text{s.t.} \quad \mathbf{Ny} = \overline{\mathbf{y}}, \qquad \mathbf{y} > \mathbf{0}.$$

Eq.(3-18) is identical to the basic form of the quadratic programming problem (e.g., Nocedal and Wright, 2006), which is given as

$$\arg\min_{\mathbf{y}} \quad \frac{1}{2}\mathbf{y}'\mathbf{Qy} - \mathbf{c}'\mathbf{y} + \mathit{const.}, \tag{3-19}$$

$$\mathbf{Ay} \le \mathbf{a}, \qquad \mathbf{By} = \mathbf{b}.$$

where **Q**, **A**, and **B** are known matrixes, and **a**, **b**, and **c** are known vectors. In our setting, **Q** = (1/2)$\mathbf{M}^{-1}$, **A** is an identity matrix, **B** = **N**, **a** is a vector of –1s, **b** = $\overline{\mathbf{y}}$, **c** = 2$\mathbf{M}^{-1}\boldsymbol{\mu}$, and *const.* =$\boldsymbol{\mu}'\mathbf{M}^{-1}\boldsymbol{\mu}$ (compare Eq.3-18 and Eq.3-19). After all, the non-negative is introduced by replacing the interpolation applying Eq.(3-11) with the interpolation by solving the quadratic programing problem.

3.3. A simulation study

3.3.1. Outline

This section compares accuracies of areal interpolation methods by employing the Monte Carlo simulation proposed by Fisher and Langford (1995). This simulation evaluates interpolation accuracies to non-aggregate level units repeatedly by varying aggregate level units *P* times. This study applies the minor municipal districts in Ibaraki prefecture for the non-aggregate level units, and the data interpolated are the employee numbers in 2006 (sample size = 4,800: Fig.3-3).

The aggregated level units are generated by iterating the following procedure *P* times: (i) *N* minor municipal districts are randomly chosen; (ii) Each of the *N* units are expanded by merging them with their adjacent minor municipal districts; (iii) If a minor municipal district is included in more than one expanded units, one of the expanded unit including the minor municipal district is selected randomly and the minor municipal district is merged with it; (iv) Steps (ii) and (iii) are repeated until all minor municipal districts are included in any of the expanded units. Following Cockings *et al.* (1997), we set *N* to 50, and *P* to 300. Namely, areal interpolations from the synthetic 50 units to the 4,800 minor municipal districts are iterated 300 times.

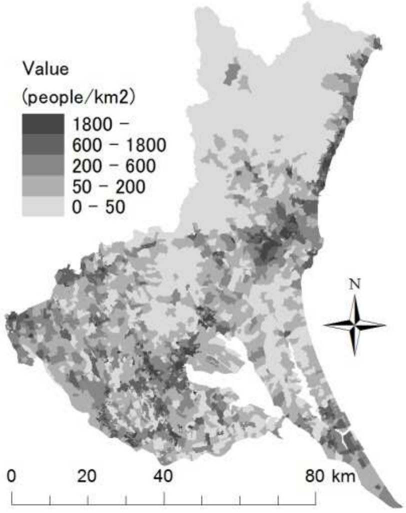


Figure 3-3: The employee numbers in the minor municipal units

This simulation examines whether or not the proposed GWR-based method (ATP GWR) outperforms the standard areal interpolation methods: the areal weighting interpolation method (AW) and the dasymetric method (DA).

The proportional distribution ratios (i.e., the elements in $\mathbf{N}^{volume}$ or $\mathbf{N}$) used in DA and ATP GWR are given by the building land areas in the minor municipal units. In the other words, DA and ATP GWR distribute the employee numbers only for building lands. Besides, ATP GWR considers the following explanatory variables: ratio of urban area; total length of roads per unit area (km/km^2); averages of the railway distances (km) from the nearest station to the Tokyo and Mito stations that are weighted by their numbers of annual passengers (people). Note that the Mito station is the central station in Ibaraki prefecture. The tri-cube function Eq.(2-59) is used in ATP GWR to model spatial heterogeneity, and the bandwidth parameter r is estimated via the cross-validation.

Calculations in this study were performed using R 2.11.1 (provided by CRAN), and visualization was done using ArcGIS 10.1 (provided by ESRI).

Table 3-1: Response Variables and Explanatory Variables

Variables	Description	Source
Employee Numbers	Numbers of employees in each municipal unit or minor municipal district	Ministry of Internal Affairs and Communications Statistics Bureau
Building land	Building land areas that are calculated by aggregating the indicator variables by 100m × 100m grids, indicating 1 if the grid is a building land, and 0, otherwise	NLNI
Urban_ratio	Ratio of urban area	
Road_density	Length of roads per unit area	
TM_dist	Averages of the railway distances from the nearest station to the Tokyo and Mito stations that is weighted by their numbers of annual passengers	NLNI; East Japan Railway Company

NLNI: National Land Numerical Information download service

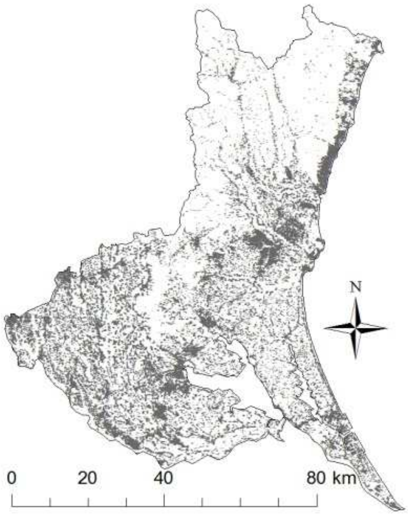


Figure 3-4: Spatial distribution of the building land areas

3.3.2. Result

The root mean square error (RMSE: Eq.3-20) and the mean absolute error (MAE: Eq.3-21), which are applicable even if some elements in **y** are zeros, are used for accuracy evaluations.

$$RMSE = \sqrt{\frac{1}{4{,}800}\sum_{k=1}^{4{,}800}(y_k - \hat{y}_k)^2} \ , \tag{3-20}$$

$$MAE = \frac{1}{4{,}800}\sum_{k=1}^{4{,}800}| \ y_k - \hat{y}_k \ |, \tag{3-21}$$

where y_k is the actual employee number in k-th minor municipal district and $\hat{y}_k$ is the interpolated employee numbers in that district.

The RMSEs and MAEs of AW, DA, and ATP GWR are summarized in Tables 3-2 and 3-3. The result indicates that the RMSEs and MAEs of the proposed method are better than those of AW and DA on average. However, these differences between ATP GWR and DA are not so large. Hence, I test their difference using the Tukey's test (Tukey, 1977), a test for multiple comparison. Table 3-4 summarizes the test results. The table shows that the accuracies of DA and ATP GWR are significantly different at the 1% level.

On the other hand, the maximums and the standard deviations of the RMSEs and MAEs of GWR are greater than those of DA. Thus, the performance of GWR is unstable. To overcome this problem, applying Bayesian estimation, which is a sort of shrinkage estimation, might be helpful.

Subsequently, MAE for each unit, which is defined as

$$MAE_k = \frac{1}{300}\sum_{iter=1}^{300} | y_k - \hat{y}_k^{(iter)} |, \tag{3-22}$$

where *iter* is the index of the iteration numbers, are plotted in Fig. 3-5. This figure shows that the accuracy of AW is worse, and that the accuracy of ATP GWR is superior to DA, particularly in the middle and southeastern areas.

Table 3-2: Summary statistics of RMSE

Statistics	AW	DA	ATP GWR
Mean	769	555	549
Median	767	555	545
Standard deviation	27.0	13.2	19.8
Maximum	883	618	629
Minimum	702	512	512

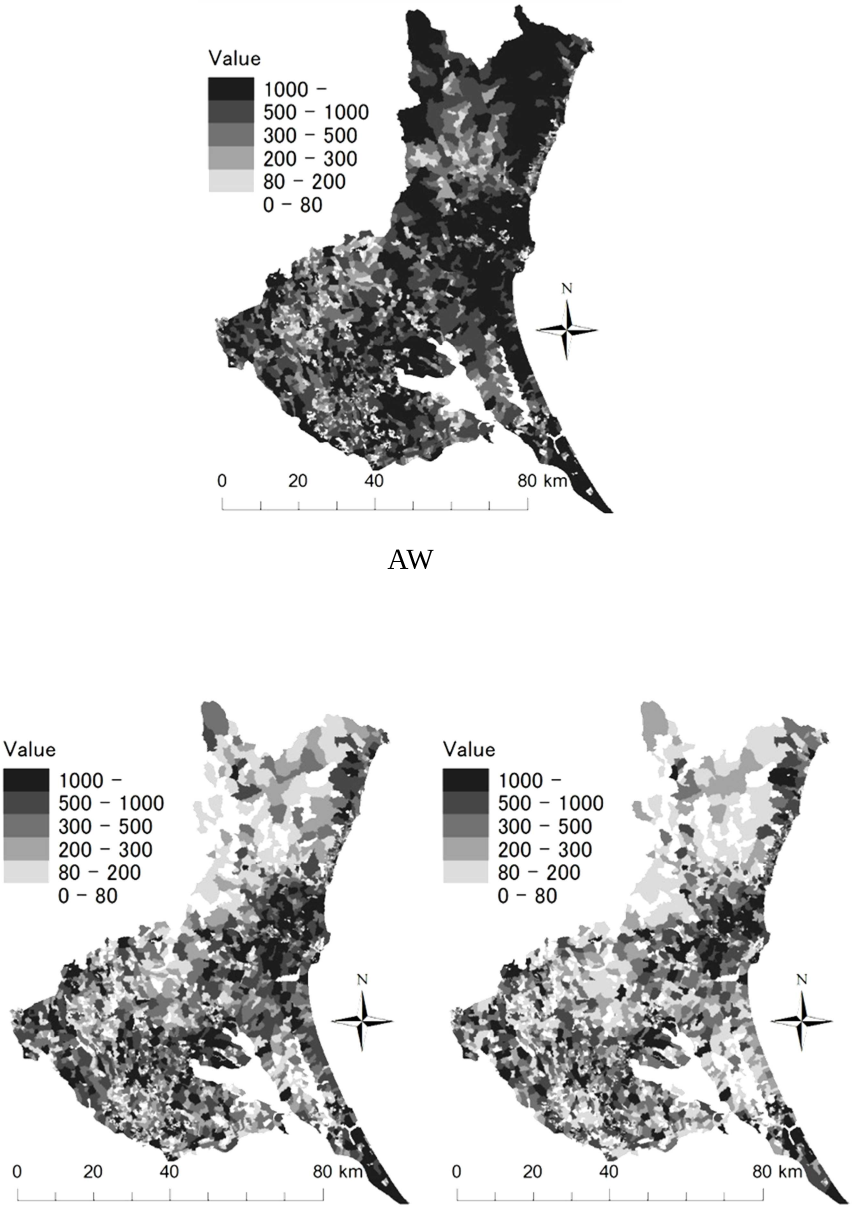
Table 3-3: Summary statistics of MAE

Statistics	AW	DA	ATP GWR
Mean	370	264	257
Median	370	264	255
Standard deviation	8.28	2.92	6.43
Maximum	401	272	287
Minimum	346	255	249

Table 3-4: Test result of the differences among the methods using Tukey’s test

	RMSE		MAE	
Pair	t-value	Significance	t-value	Significance
AW – DA	139	***	238	***
AW – ATP GWR	143	***	224	***
DA – ATP GWR	3.69	***	14.5	***

*, ** and *** represent significant levels (10%, 5%, and 1%, respectively)



AW

DA

ATP GWR

Figure 3-5: MAE of each method

Finally, interpolation results of the three methods are plotted in Fig. 3-7. Here, for easy understanding, the employee numbers are interpolated using the municipal level employee numbers (Sample size: 48; Fig.3-6). The results suggest that the interpolation results of DA and ATP GWR, which consider supplementary data, are much more similar to the true distribution (Fig. 3-3) than the result of AW. This result agrees with geographical studies that emphasize the importance of considering supplementary data (e.g., Fisher and Langford ,1995; 1996). The result of DW appears to be smoothed overly, whereas the result of ATP GWR is less smooth, which is more similar to the true distribution. The over-smoothness of DA would be due to its strong assumption that the employees are distributed evenly in building lands. Thus, effectiveness of GWR is verified from the viewpoint of avoiding such an over-smoothed result.

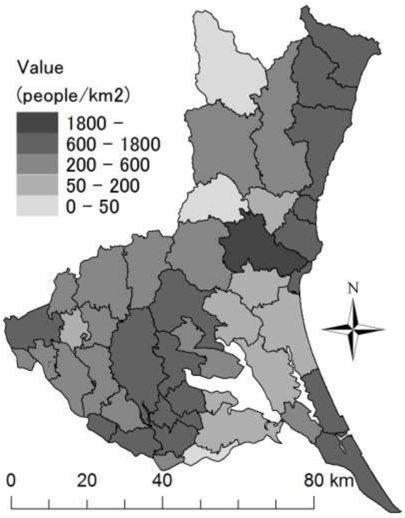
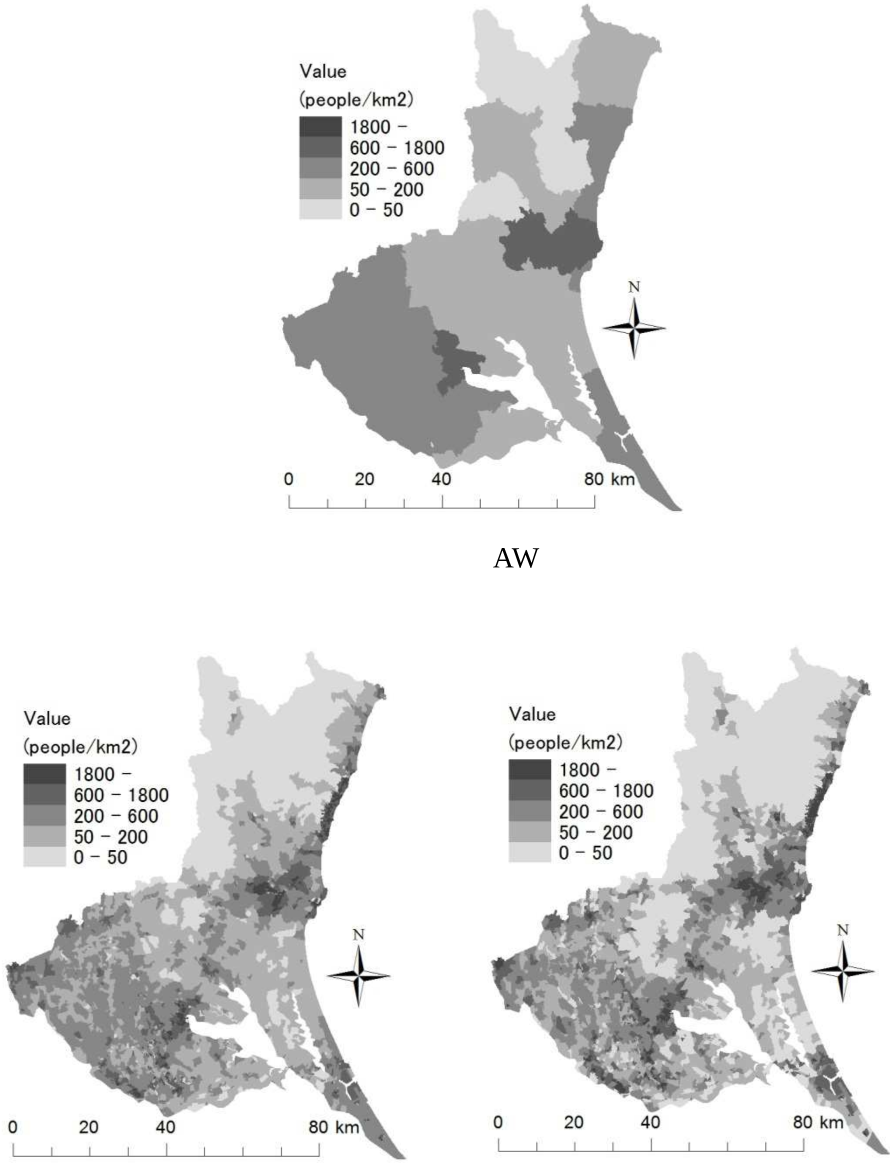


Figure 3-6: Municipal level employee numbers



AW

DA

ATP GWR

Figure 3-7: Interpolation results

3.4. An empirical study

3.4.1. Backgrounds

Building stock (total floor area) data have often been used as “basic units” in economic analysis. For instance, the Flood Control Project Economic Assessment Manual issued by the Ministry of Land Infrastructure, Transport, and Tourism (MLIT) estimates flood damage costs to buildings using building stock data (flood damage cost = total floor area × appraised value of buildings per unit area), and energy consumption is estimated using building stock data in many cases (e.g., Yamagata *et al.*, 2013). In addition, building stock data would be required to achieve a stock-type society.

Given this background, the importance of building stock data has been recognized, and, since 2010, Building Stock Statistics have been provided by the MLIT each year. These statistics estimate the stock amounts using the Housing and Land Survey, Corporations Survey on Buildings, Statistics Survey on Construction, and so on. They provide building stock amounts in each category (residence/non-residence, wooden/non-wooden, completion period). However, the data are only available at the prefecture level, whereas more spatially detailed (e.g., municipal level) stock data are required for, for example, compact city planning and climate change adaptation planning. Accordingly, this study considers constructing municipal level building stock data.

A bottom-up approach that estimates stocks by compiling micro level data, such as GIS data of buildings, would be an effective way to construct accurate municipal building stock data. However, performing such an approach for all municipalities in Japan would be prohibitively costly. On the other hand, a top-down approach that interpolates the municipal building stocks using the prefectural data would be a far more efficient way to construct the data.

Thus, this section considers applying areal interpolation techniques, including ATP GWR, to municipal-level building stock estimation.

3.4.2. Review of municipal level building stock estimation

The Property Tax Ledger (issued by the Fixed Property Tax Division, Local Tax Bureau) and the Basic Survey of City Planning (issued by each municipality) are the prime sources of municipal level stock data (Sakata and Yoshikawa, 2001).

The Property Tax Ledger is based on site surveys by municipalities, and is updated every year. Since the objective of this ledger is taxation, it is quite accurate, and therefore, using the data is the best way to accurately construct the municipal level stock data (Sakata and Yoshikawa, 2001). However, completing the municipal data for multiple years using the Property Tax Ledger would seem to be difficult.

The Basic Survey of City Planning comprises GIS data, and is constructed every five years based on the Fundamental Land Classification Survey. The data are based on aerial surveys and site surveys. As investigated by Miyagi (2009) and Tsutsumi *et al*. (2012), Chiba and Kanagawa are the only prefectures in the Tokyo metropolitan area that provide GIS data. In other words, this data is not yet available for all prefectures.

With regard to residential building stocks, municipal level building stock data are assessed by the Housing and Land Survey, which is a basis for the Building Stock Statistics. However, this survey does not provide non-residential building stocks. To the best of the author’s knowledge, no attempt has yet been made to provide detailed non-residential stock data across Japan.

This section applies areal interpolation methods to the prefecture level data of the Building Stock Statistics, and estimates the municipal level residential and non-residential stock amounts. §3.4.3 compares the effectiveness of the areal interpolation methods from the viewpoint of the building stock estimation, and §3.4.4 performs the building stock estimation based on the results of the comparison.

3.4.3. Comparative analysis of the building stock estimation

3.4.3.1. Data and models

In this section, the ATP GWR model is used to estimate the municipal residential stocks that were completed by 2005 (sample size: 1,803) using the prefectural stock data (Building Stock Statistics). Then, the accuracy of the estimates is measured by comparing the results to the actual data. This section also compares broader several methods: the areal weighting interpolation method (AW), the dasymetric method (DA), the geostatistical method, with predictive equation given by Eq.(3-6) (GS2: see Gotway and Young, 2007), and the ATP GWR model. In addition, I apply the geostatistical method implemented in ArcGIS 10.2 (GS1). GS1 is the standard form of ATP kriging, and has the following predictive equation:

$$\hat{\mathbf{y}} = \alpha\mathbf{1} + \mathbf{CN}'(\mathbf{NCN}')(\bar{\mathbf{y}} - \alpha\mathbf{N1}) , \tag{3-23}$$

where α is a parameter. GS1 is a geostatistical model that does not consider explanatory variables. Among the methods compared, AW and GS1 are easily implemented using ArcGIS, and DA is also a simple proportional distribution. Thus, AW, DA, and GS1 are practical.

DA, GS2, and ATP GWR have the advantage that the distribution ratios (the elements of N) can be arranged using supplementary data, which I do in this study using the building land area in each municipality (owing to a limitation of ArcGIS, GS1 does not have this advantage). In addition, the number of railway stations and the densities of buildings are used as explanatory variables in GS2 and ATP GWR. Furthermore, to avoid negative building stock estimates, I introduce non-negative constraints into these methods. Note that AW and DA estimates are always non-negative. On the other hand, owing to a limitation of ArcGIS, non-negative constraints cannot be introduced into GS1. GS1 and GS2 use the exponential covariogram model Eq.(2-9) to capture spatial dependence, and ATP GWR uses the Gaussian kernel model Eq.(2-57) to capture spatial heterogeneity. Finally, each of the methods is summarized in Table 3-5.

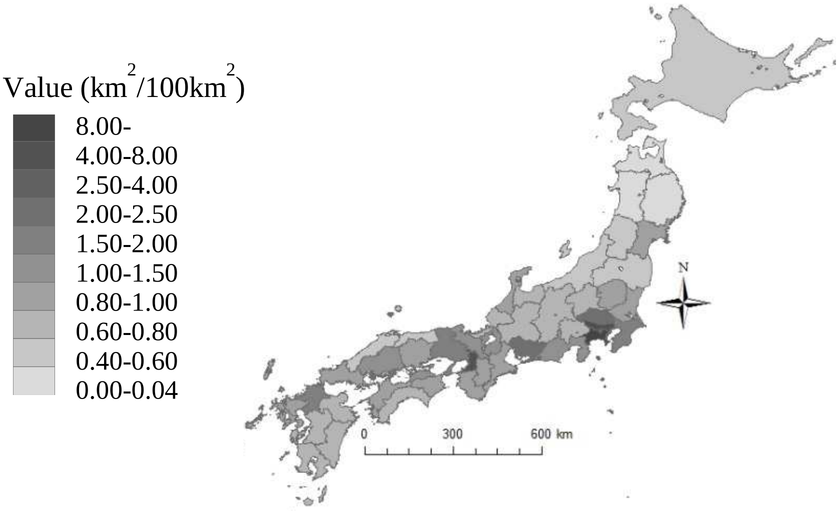


Figure 3-8: Prefectural residential stock densities (2005)

Table 3-5: Summary of the areal interpolation methods

Method	Distribution ratio	Explanatory variables	Spatial dependence	Spatial heterogeneity
AW	Area	NA		
DA	Building land area			
GS1	Area		×	
GS2	Building land area	Numbers of railway stations per unit area Road densities	×	
ATP GWR				×

[1] Data source: National Land Numerical Information download services

Municipal level residential stock data are required for accuracy verification, and are provided by the Housing and Land Survey (see §3.4.2). However, the data differ to the Building Stock Statistics, which I use for estimation in that they do not consider shared spaces in apartments, which account for about 20% of total residential stock. Accordingly, this study evaluates the accuracy of the estimates using the Property Tax Ledger data (2005) in 241 municipalities in the Tokyo metropolitan area, which are corrected by Miyagi (2009) and Tsutsumi *et al.* (2012). However, these data have some limitations. The first is the region. In this regard, I consider that the target region includes both urban areas (e.g., central Tokyo area) and non-urban areas (e.g., Tama area), and we can grasp some sort of tendency using the data in this region. Secondly, Building Stock Statistics includes data on residences for public servants, which the Property Tax Ledger does not. This may introduce discord between these data in some municipalities. However, I consider this influence to be sufficiently small, because the Building Stock Statistics data and the prefectural level aggregations of the Property Tax Ledger are quite similar (see Fig. 3-9). Their error ratios are 7.1% at maximum.

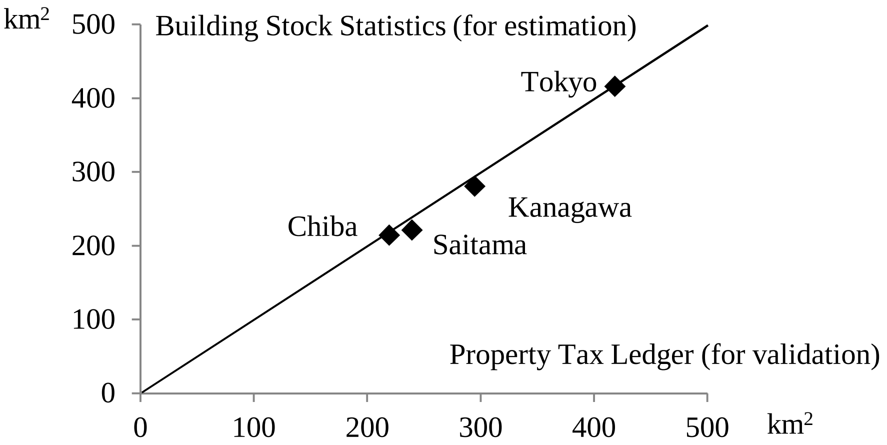


Figure 3-9: Comparison of the Building Stock Statistics data and Property Tax Ledger data

3.4.3.2. Parameter estimation result

Table 3-6 summarizes the parameter estimation results of GS2 and ATP GWR. As shown in this table, in both models, the numbers of railway stations are significant at the 1% level, while the road densities are not significant. The lack of significance of the road densities might be because the effect of the road densities is already explained by the building land areas, which I use as distribution ratios.

The range parameter of GS2 is relatively small. This indicates that the building stock data have local scale spatial dependence. On the other hand, the bandwidth parameter of ATP GWR is quite large, which implies an absence of spatial heterogeneity (i.e., $\boldsymbol{\beta}_i$ are constant across municipalities). ATP GWR considers heterogeneity by weighting each municipality based on spatial adjacency, which describes spatial heterogeneity, and building land area, which describes non-spatial heterogeneity (see Eq.3-14). Hence, the result implies that heterogeneity across municipalities is well captured by building land area only.

Table 3-6: Parameter estimates of GS2 and GWR

	GS2			GWR				
	Estimates	Std.err.	Signif.	Estimates		Std.err.		Signif.
				Max	Min	Max	Min	
Const	15.5	9.99×10^{-1}	***	15.6	15.4	1.00	9.67×10^{-1}	***
Num. of railway stations	36.7	6.27	***	36.7	36.5	5.00	4.99	***
Road density	1.90×10^{-2}	2.17×10^{-1}		2.34×10^{-2}	6.40×10^{-3}	2.07×10^{-1}	2.04×10^{-1}	
range (km)		31.3						
bandwidth (km)						2867		

[1)] *** denote 1% significant levels

3.4.3.3. Accuracy comparison result

The accuracy of each model is assessed using the following six measures:

$$\mathrm{RMSE} = \sqrt{\frac{1}{1{,}803}\sum_i (\hat{y}_i^C - y_i^C)^2}\ , \tag{3-24}$$

$$\mathrm{RMSE\_den} = \sqrt{\frac{1}{1{,}803}\sum_i (\hat{y}_i - y_i)^2}\ , \tag{3-25}$$

$$\mathrm{RMSPE} = \sqrt{\frac{1}{1{,}803}\sum_i \left(\frac{\hat{y}_i^C - y_i^C}{y_i^C}\right)^2}\ , \tag{3-26}$$

$$\mathrm{MAE} = \frac{1}{1{,}803}\sum_i |\ \hat{y}_i^C - y_i^C\ |\ , \tag{3-27}$$

$$\mathrm{MAE\_den} = \frac{1}{1{,}803}\sum_i |\ \hat{y}_i - y_i\ |\ , \tag{3-28}$$

$$\mathrm{MAPE} = \frac{1}{1{,}803}\sum_i \left|\ \frac{\hat{y}_i^C - y_i^C}{y_i^C}\ \right|\ . \tag{3-29}$$

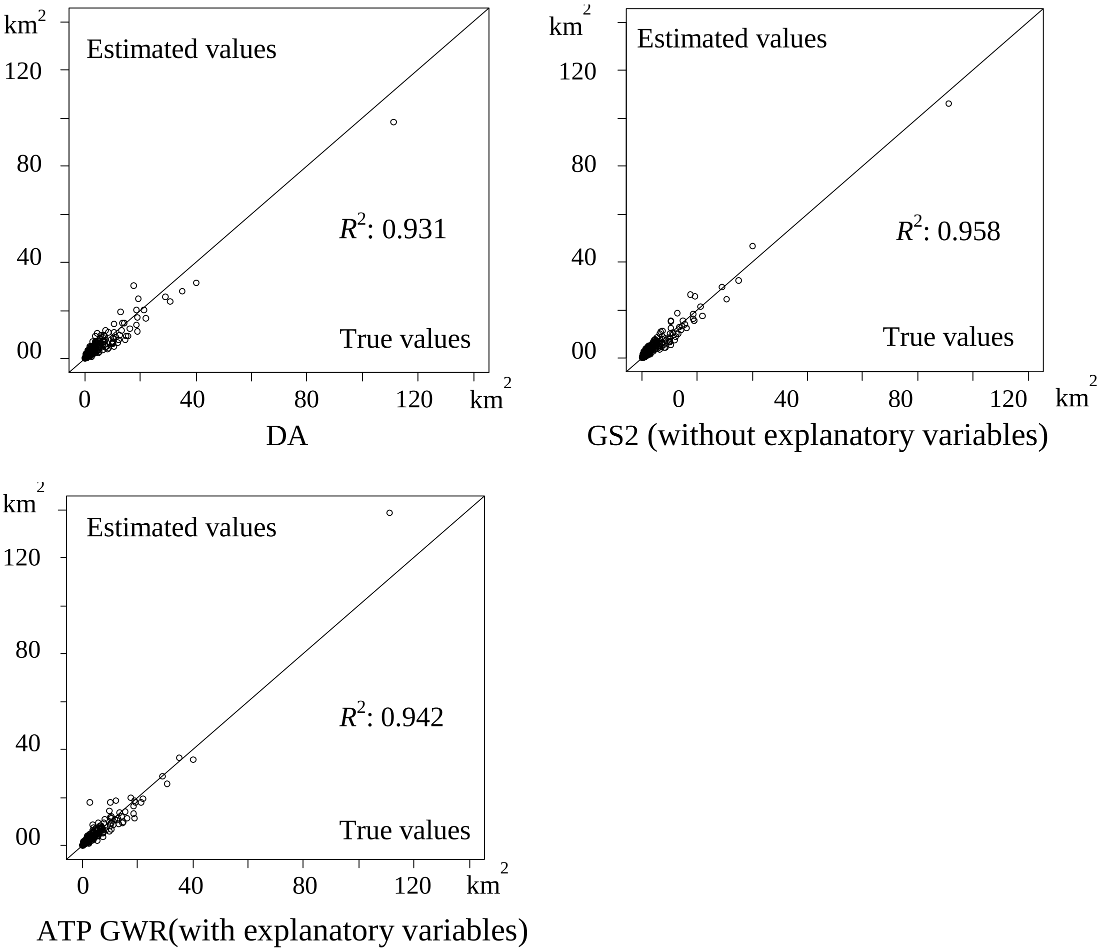
RMSE and MAE are sensitive to errors at municipalities that have large amounts of stock. RMSE_den and MSE_den are sensitive to errors at municipalities with large stock densities. RMSPE and MAPE are standardizations of RMSE and MAE, respectively. Large values of RMSPE and MAPE indicate that the errors are large compared to the amounts of stock (or to stock densities).

Here, two cases are assumed for GS2 and ATP GWR: the case with explanatory variables and the case without explanatory variables (constants only). Table 3-7 summarizes the results of the accuracy comparison. The table suggests that AW and GS1, which do not consider supplementary data as distribution ratios (elements of N) or explanatory variables, are inefficient. This confirms the importance of considering supplementary data in areal interpolation.

The two spatial statistical methods, GS2 and ATP GWR, outperform DA, the efficiency of which has been demonstrated. In addition, GS2 is more accurate than ATP GWR. This could be because the stock data have spatial dependence, but no spatial heterogeneity (see §3.4.3.2). However, the accuracy of GS2 with explanatory variables is worse than GS2 without explanatory variables. This result is intuitively inconsistent. ATP GWR does not have such strange results, and so may be better than GS2 at capturing the influence of the explanatory variables.

Table 3-7: Accuracy comparison result (gray: better than DA; bold: best)

	AW	DA	GS1	Without explanatory variables		With explanatory variables	
				GS2	GWR	GS2	GWR
RMSE	7.97×10^{6}	2.62×10^{6}	6.85×10^{6}	**1.93×10^{6}**	4.52×10^{6}	2.19×10^{6}	2.76×10^{6}
RMSE_den.	1.44×10^{5}	8.50×10^{4}	1.68×10^{5}	**7.58×10^{4}**	1.08×10^{5}	1.24×10^{5}	1.22×10^{5}
RMSPE	1.47	8.35×10^{-1}	7.43	**5.39×10^{-1}**	6.57×10^{-1}	6.27×10^{-1}	6.22×10^{-1}
MAE	4.26×10^{6}	1.70×10^{6}	3.71×10^{6}	1.26×10^{6}	1.93×10^{6}	**1.18×10^{6}**	1.27×10^{6}
MAE_den.	9.07×10^{4}	5.12×10^{4}	9.53×10^{4}	**4.34×10^{4}**	5.47×10^{4}	4.49×10^{4}	4.51×10^{4}
MAPE	3.29	5.74×10^{-1}	1.92	3.75×10^{-1}	4.59×10^{-1}	**3.56×10^{-1}**	3.60×10^{-1}



DA

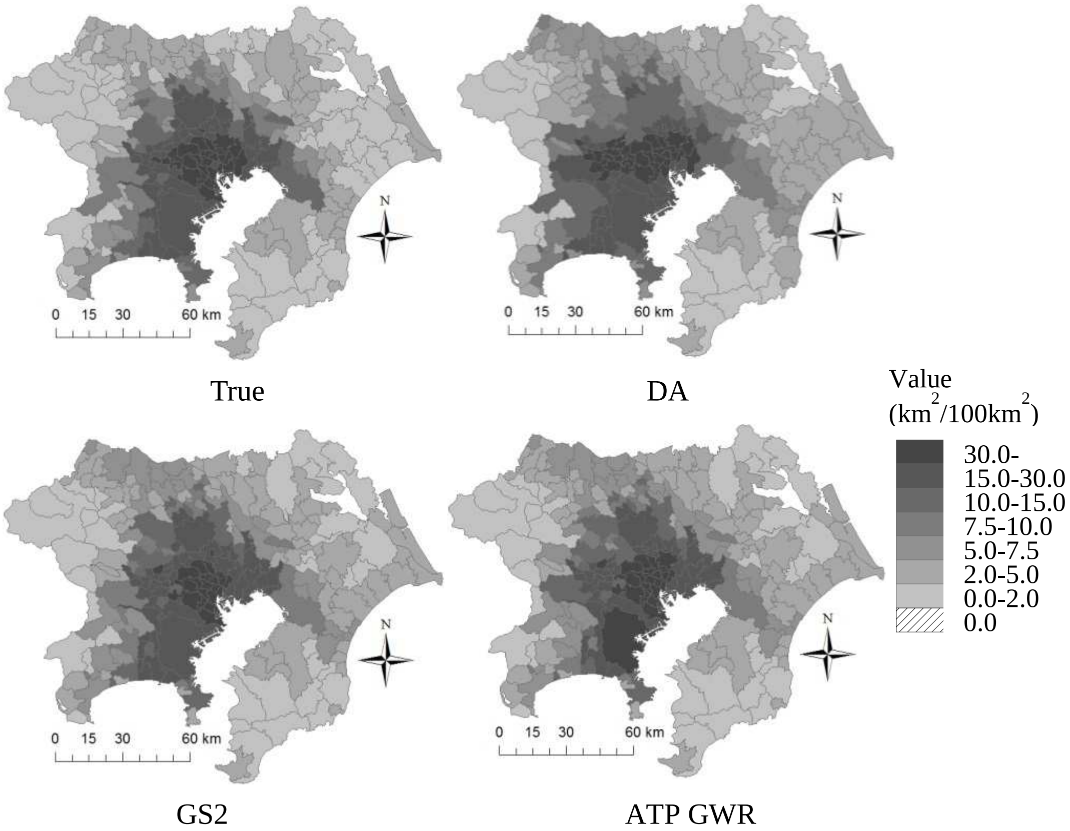
GS2 (without explanatory variables)

ATP GWR(with explanatory variables)

Figure 3-10: Comparison of the estimated values and the true values

Fig.3-10 compares the stock amounts estimated by DA, GS2 (without explanatory variables whose accuracy is best), and ATP GWR (with explanatory variables) with the true stock amounts. The figure shows that GS2 and ATP GWR are more accurate than DA in many of the municipalities. Specifically, GS2 outperforms DA in 63.9% (171/241) of municipalities, whereas ATP GWR outperforms DA in 71.0% (171/241) of municipalities.

Fig.3-11 plots the interpolation results of DA, GS2, and ATP GWR. This figure suggests that each of the results is visually quite similar to the true values. In addition, it seems that GS2 and ATP GWR capture the stock values of the central Tokyo area better than DA.


True

DA

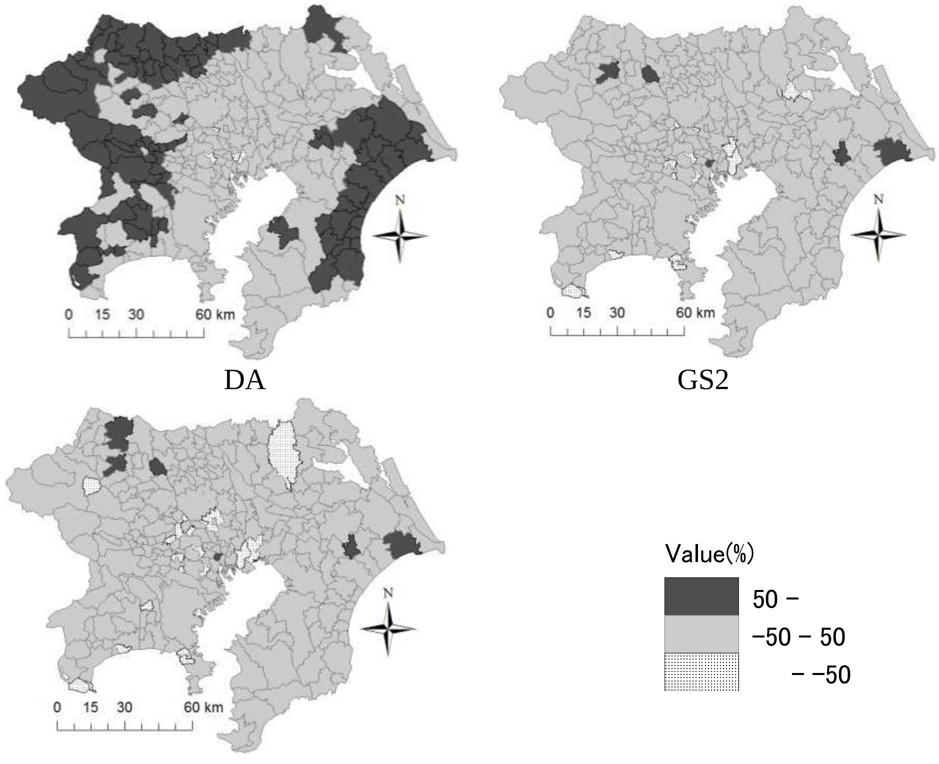
GS2

ATP GWR

Figure 3-11: True residential stock densities (Property Tax Ledger) and estimated stock densities of DA, GS2, and ATP GWR (with explanatory variables)

Then, the error ratios of DA, GS2, and ATP GWR are plotted in Fig.3-12. This figure suggests that DA overestimates the stocks in non-urban areas. Such a tendency is not seen in the results of GS2 and ATP GWR. Note that, because of the volume preserving property, the overestimation in non-urban areas implies an underestimation in urban areas.

Local residual spatial dependence is tested using the local MC (see §2.3.2). Fig.3-13 summarizes the tests results. The white dots in the figure represent municipalities whose residual spatial dependence is significantly positive (i.e., municipalities whose residual values are similar to their surrounding municipalities), and black dots represent municipalities with significant negative dependence (i.e., municipalities whose residual values are dissimilar to their surrounding municipalities). This figure demonstrates that the spatial dependence component, which could not be captured by DA, is captured well by GS and ATP GWR. However, the residuals of GS and ATP GWR still show significant positive spatial dependence in the central Tokyo area. As a result, their global MCs (see §2.3.2) are positively significant at the 1% level. Thus, future studies need to extend statistical areal interpolation methods to capture spatial dependence more adequately.

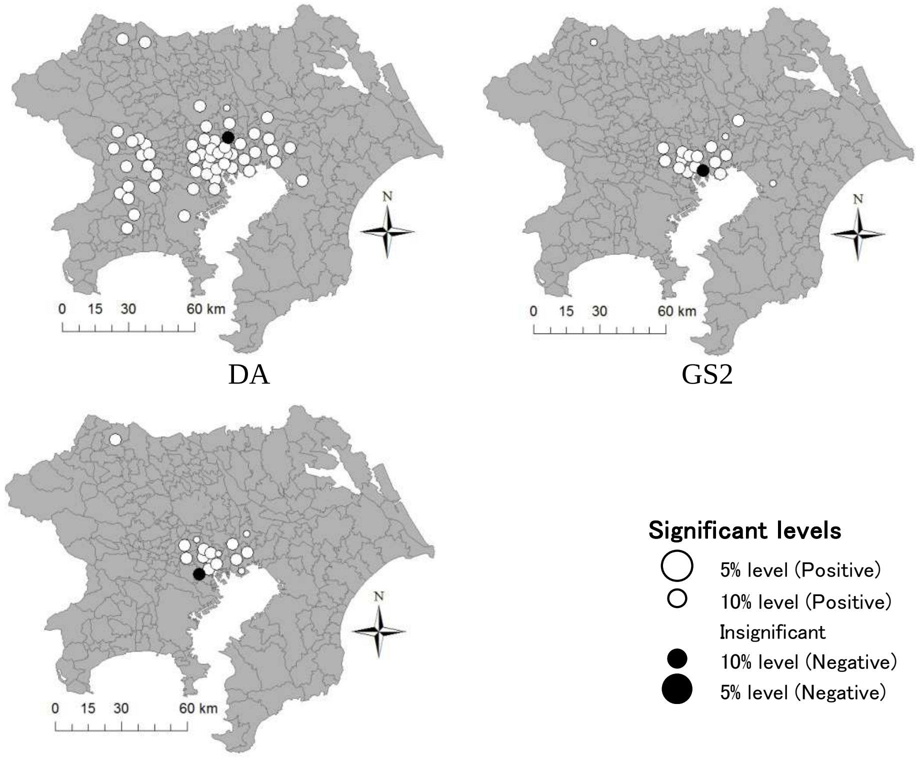


DA

GS2

ATP GWR

Figure 3-12: Error ratios of DA, GS2, and ATP GWR (with explanatory variables)



DA

GS2

ATP GWR

Figure 3-13: Significance of residual local MC for DA, GS2, and ATP GWR (with explanatory variables)

3.4.4. Municipal building stock estimation results

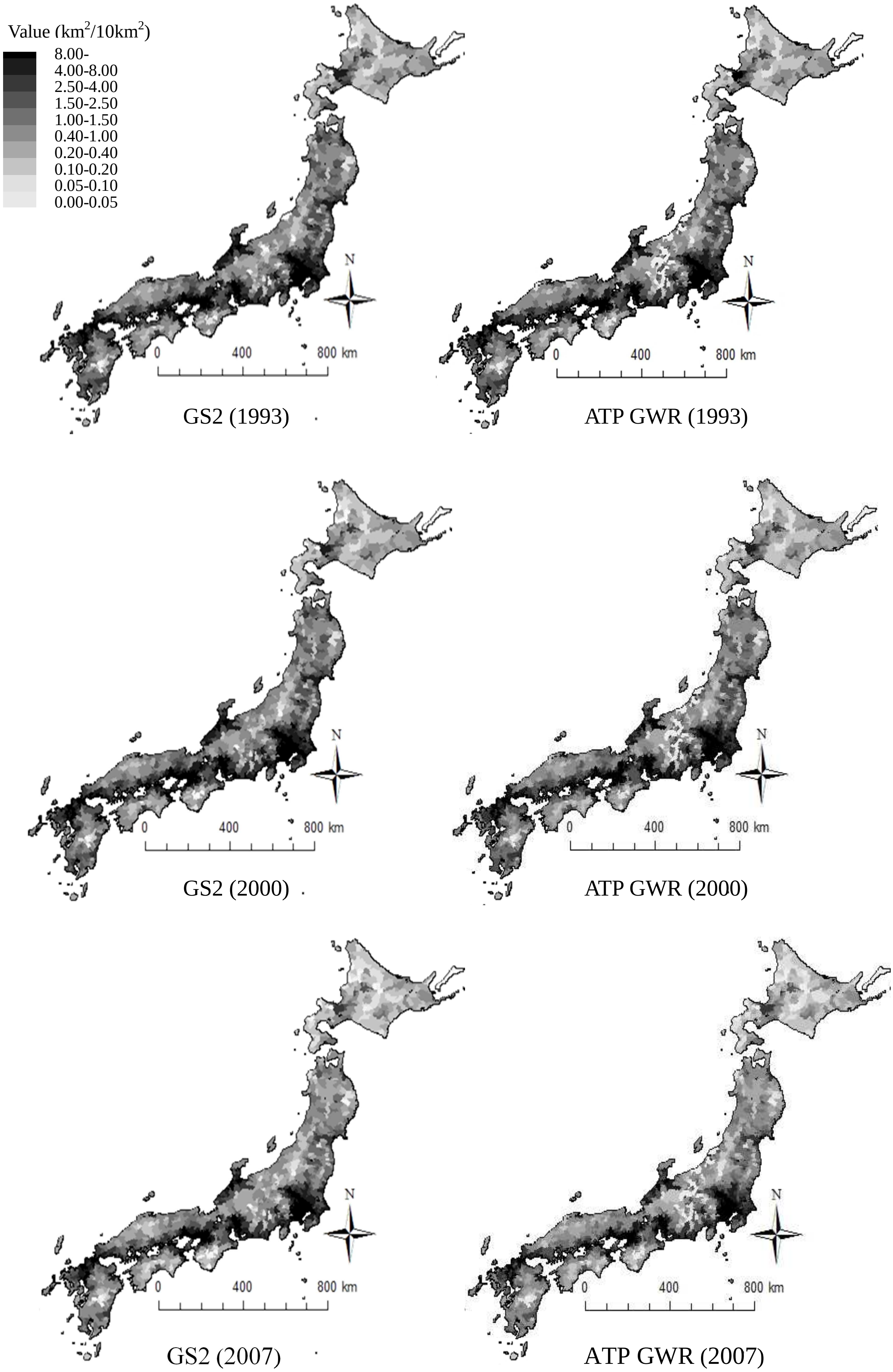
Building stocks in each type (residence/non-residence, wooden/non-wooden, completion years) are estimated using ATP GWR with explanatory variables, and using GS2 without explanatory variables, which was the most accurate. The estimation results in 1991, 2000, and 2007 are shown in Fig.3-14 (wooden residential stocks), Fig.3-15 (non-wooden residential stocks), Fig.3-16 (wooden non-residential stocks), and Fig.3-17 (non-wooden non-residential stocks).

The results of the two methods are visually similar, and roughly speaking, the estimation results are intuitively reasonable. The results all successfully describe concentrations of stocks in urban areas. The concentration is particularly prominent in the non-wooden stocks.

On the other hand, these results include some strange points. For example, the non-wooden residential stocks estimated by GS2 indicate extremely small values in the North Kanto area. In addition, the stock amount estimates in some municipalities indicate 0. Accordingly, the spatial statistical methods must be developed further to remove such odd results.

3.4.5. Discussion

This section compares the effectiveness of the areal interpolation methods by applying them to building stock estimation. As a result, the accuracy of the spatial statistical methods, including GS2 and ATP GWR, is confirmed. I also verify the importance of considering supplementary data as distribution ratios and explanatory variables. This discussion is significant when needing to estimate the building stocks required in a compact city policy and climate change adaptation policy effectively.



GS2 (1993)

ATP GWR (1993)

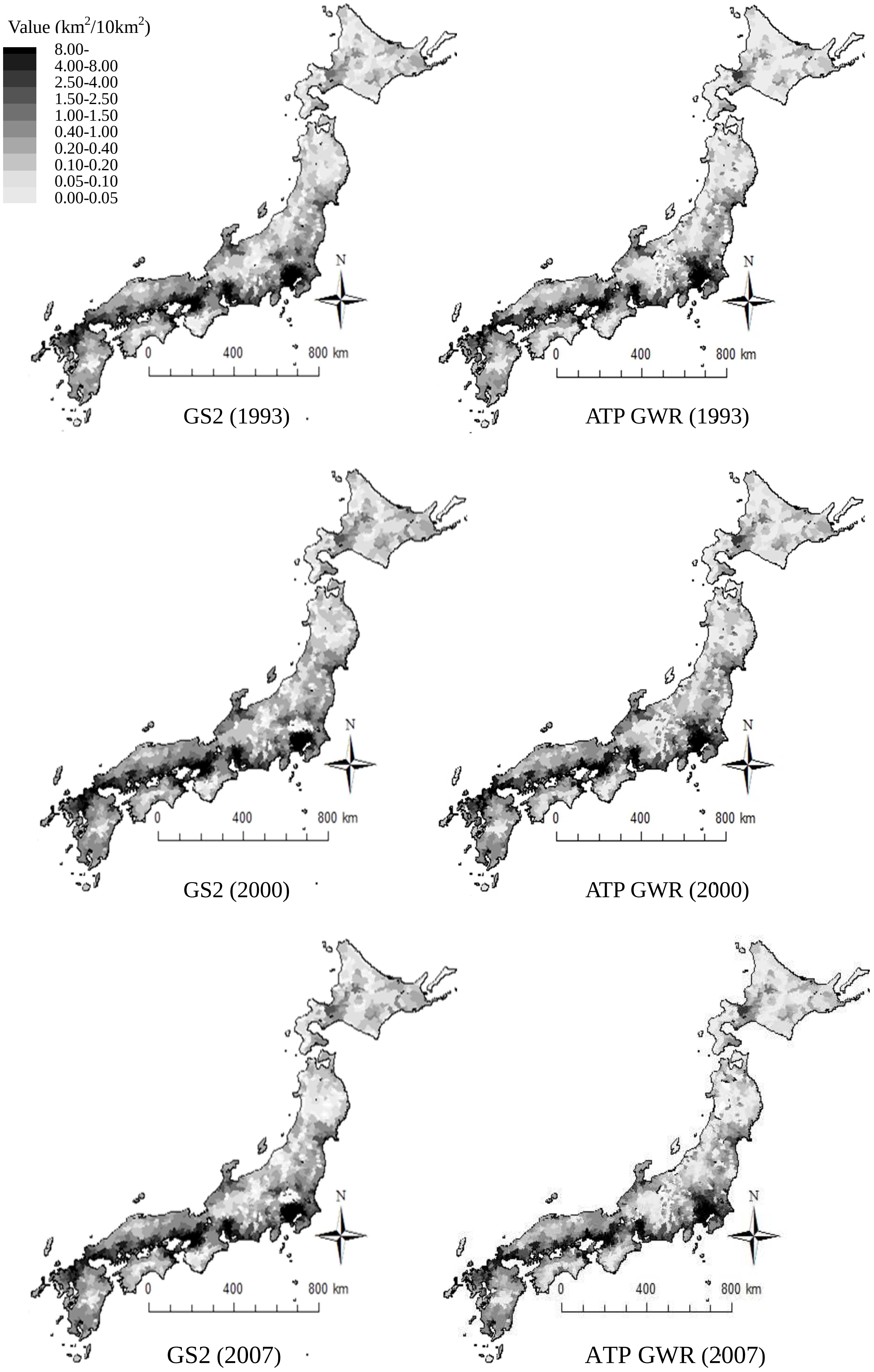
GS2 (2000)

ATP GWR (2000)

GS2 (2007)

ATP GWR (2007)

Figure 3-14: Estimated wooden residential building stock densities



GS2 (1993)

ATP GWR (1993)

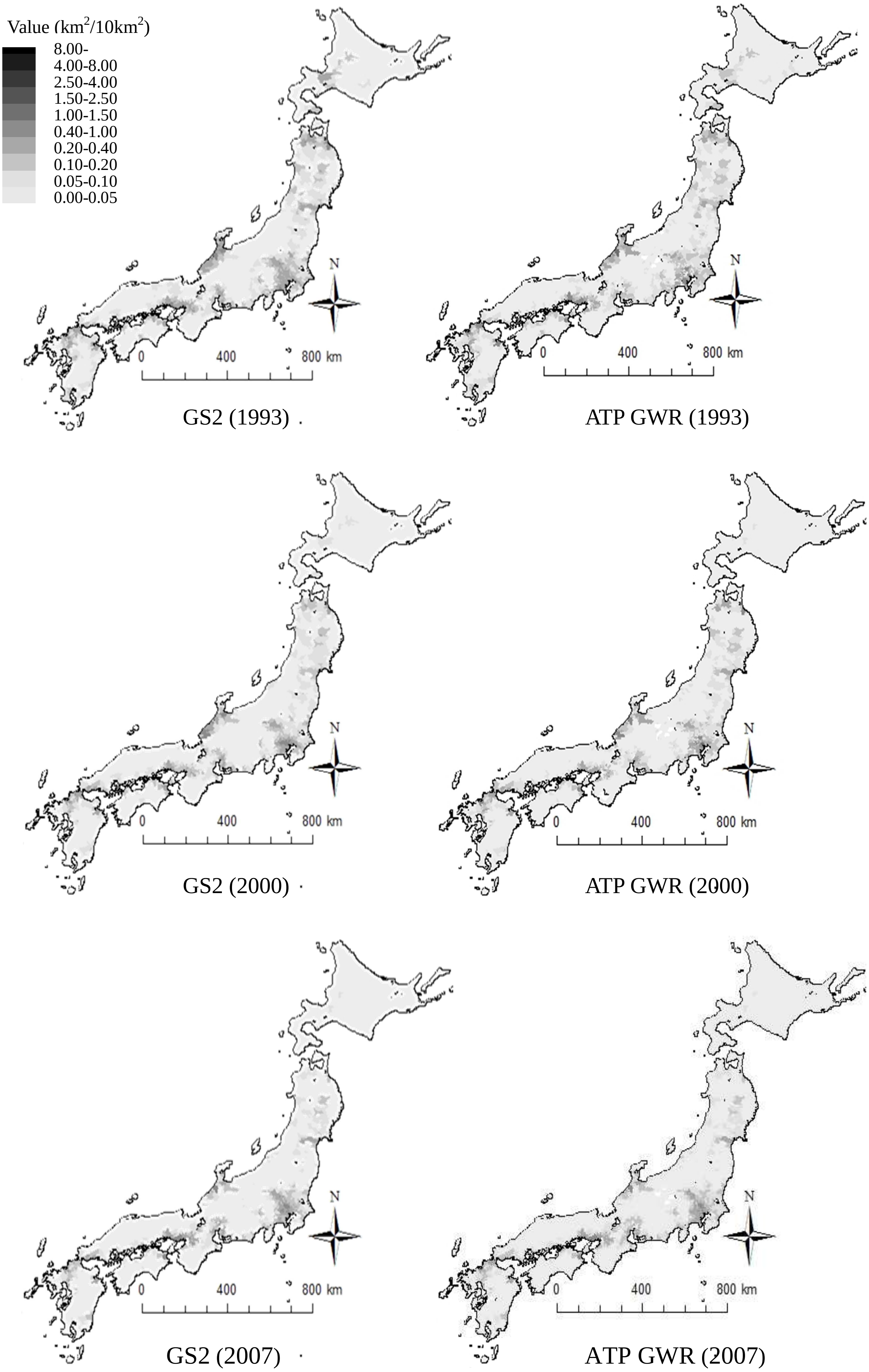
GS2 (2000)

ATP GWR (2000)

GS2 (2007)

ATP GWR (2007)

Figure 3-15: Estimated non-wooden residential building stock densities



GS2 (1993)

ATP GWR (1993)

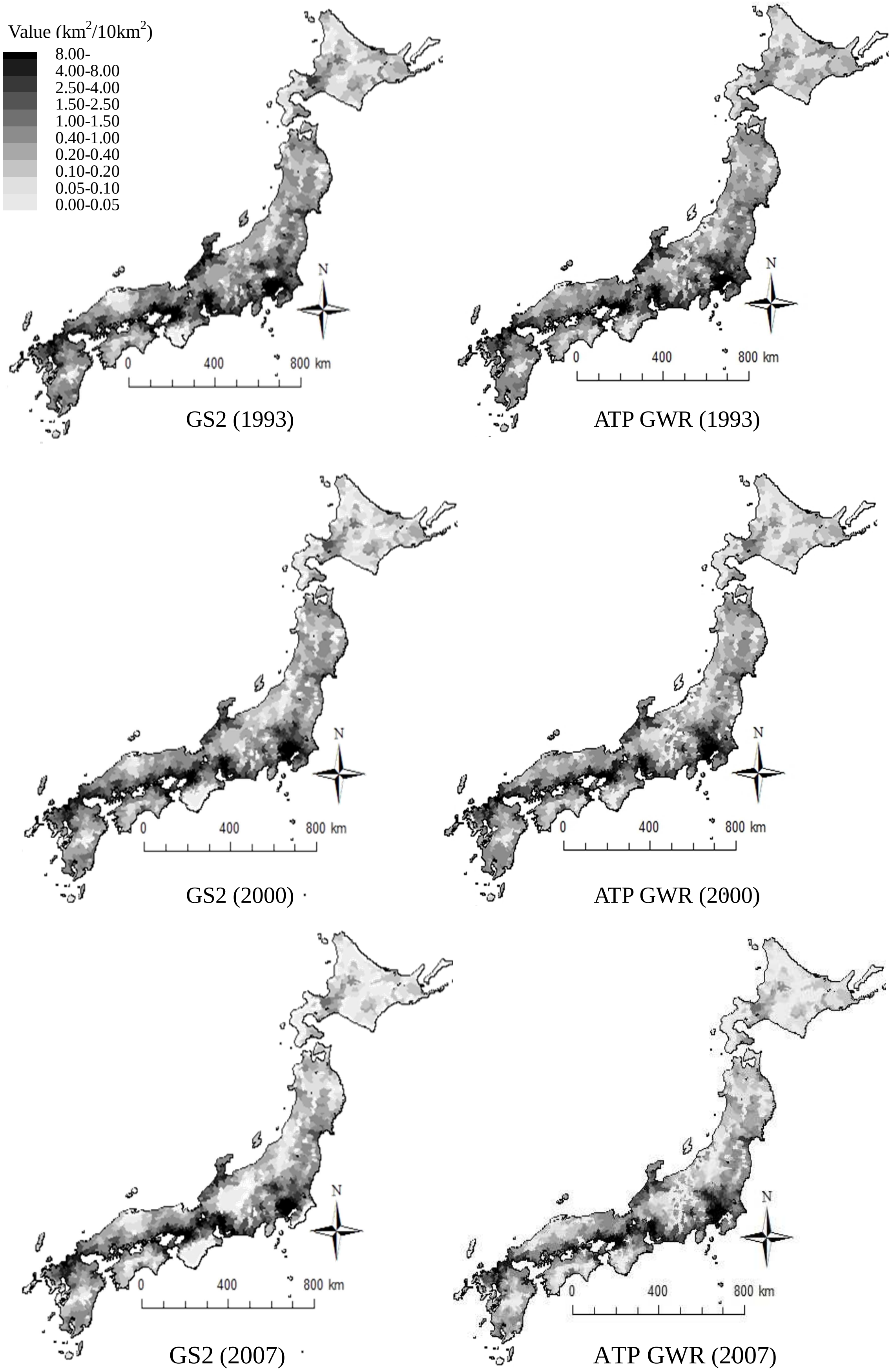
GS2 (2000)

ATP GWR (2000)

GS2 (2007)

ATP GWR (2007)

Figure 3-16: Estimated wooden non-residential building stock densities



GS2 (1993)

ATP GWR (1993)

GS2 (2000)

ATP GWR (2000)

GS2 (2007)

ATP GWR (2007)

Figure 3-17: Estimated non-wooden non-residential building stock densities

3.5. Summary

This chapter constructs GWR-based areal interpolation methods, and then, we confirmed their effectiveness by comparing them with standard areal interpolation methods. This method is consistent with geographical studies in that they can be considered as an extension of the dasymetric method whose accuracy has been shown (see §3.2.2). Besides, this method is consistent with geostatistics in that they give their interpolation equations using conditional expectations (or minimize MSPE). To the best of my knowledge, the proposed method is the only GWR-based areal interpolation method that explicitly minimizes MSPE.

I also found that statistical areal interpolation methods possibly be inaccurate if assumptions in these methods (e.g., assumption of the Gaussian distributed disturbance) are inconsistent with data distributions. This finding is consistent with the studies that pointed out inefficiency of statistical areal interpolation methods (e.g., Cromley *et al*., 2012). On the other hand, I also showed that statistical methods are accurate if the method applied is selected judiciously. It would be an important finding for further discussions of statistical areal interpolations. Especially, clarifying effectiveness and limitations of the statistical methods would be very important.

Areal interpolation (or changes in the support of areal data) must be discussed while paying attention to the MAUP, particularly when the interpolated areal data are used for secondary analyses (see §1.2). Accordingly, the next chapter considers how to apply the GWR-based model to the MAUP.

4. Modifiable Areal Unit Problem: A GWR-based Approach

4.1. Introduction

Areal interpolation models have often been used to cope with the problem of bias in parameter estimates due to aggregations (see, e.g., Wong, 2009; Gelfand, 2010), which is known as the modifiable areal unit problem (MAUP; Openshaw and Taylor, 1979). While a number of efficient areal interpolation methods have been proposed (e.g., Fisher and Langford, 1995; Xie, 1995; Eicher and Brewer, 2001; Mennis and Hultgren, 2006; Reibel and Agrawal, 2007; Kim and Yao, 2010; Zhang and Qui, 2011), there are, as yet, no theoretically sufficient solutions for MAUPs (Siffel *et al*., 2006; Butkiewicz and Ross, 2010).

There are two factors that affect the seriousness of the MAUP (Wong, 2009). The first is the underlying spatial pattern of the data. The MAUP becomes serious if the data are positively spatially dependent, while its influence is small if the data are negatively dependent (e.g., Reynolds, 1998). The second is the aggregation process. Since large variability can be canceled out, the MAUP also becomes serious when the aggregation units are large.

According to Swift *et al*. (2008), in geography, at least five approaches have been proposed to address the MAUP. The first is by applying GWR. Since GWR captures spatial patterns of data, which is a source of MAUPs, GWR is believed to be robust in dealing with the MAUP. However, GWR does not consider aggregation mechanisms, and so is not a solution to MAUPs (Fotheringham *et al*., 2002; Wong, 2009). The second approach is to apply non-aggregated data (e.g., Tagashida and Okabe, 2002). The third approach estimates aggregate-level parameters by considering non-aggregate-level structures in a variance-covariance matrix (e.g., Tranmer and Steel, 1998). The fourth approach optimizes the zoning system by minimizing intra-zone variances and maximizing the variances between zones (Openshaw, 1984). Finally, the fifth approach applies a sensitivity analysis (e.g., Odoi *et al*., 2003; Swift *et al*., 2008).

In geostatistics, the MAUP is considered a sub-problem of the change of support problem (COSP) (see §1.2). Whereas the majority of COSP studies focus on interpolation problems, some discuss the MAUP and its related problems (e.g., Gotway and Young, 2002; Gelfand, 2010; Nagle *et* *al*., 2011). However, geographical MAUP studies and geostatistical COSP studies have been discussed almost independently (see also, Haining *et al*., 2010). Combining the discussions of the MAUP from geography and geostatistics would be an important step in advancing discussions on the MAUP. As mentioned previously, GWR is believed to be robust to the MAUP in the geography field. Therefore, extending GWR based on geostatistical studies of COSP would be significant. Fortunately, I have developed a GWR-based areal interpolation model (ATP GWR), and this model is constructed in a geostatistical manner. Hence, this chapter applies the ATP GWR for MAUP.

4.2. MAUP and the GWR-based model

The ATP GWR model (Eq.3-10), which is constructed in §3.2.1, is given as

$$\begin{pmatrix}\mathbf{y}\\ \bar{\mathbf{y}}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\mu}\\ \mathbf{N}\boldsymbol{\mu}\end{pmatrix}+\begin{pmatrix}\boldsymbol{\varepsilon}\\ \mathbf{N}\boldsymbol{\varepsilon}\end{pmatrix}, \qquad \begin{pmatrix}\boldsymbol{\varepsilon}\\ \mathbf{N}\boldsymbol{\varepsilon}\end{pmatrix}\sim N\left[\begin{pmatrix}\mathbf{0}\\ \bar{\mathbf{0}}\end{pmatrix} \quad \begin{pmatrix}\mathbf{M} & \mathbf{MN}'\\ \mathbf{NM} & \mathbf{NMN}'\end{pmatrix}\right], \tag{4-1}$$

$$\boldsymbol{\mu}=\langle \mathbf{x}'_k\boldsymbol{\beta}_k\rangle,$$

where its aggregate level model, which is used for parameter estimation, is

$$\bar{\mathbf{y}}=\mathbf{N}\boldsymbol{\mu}+\mathbf{N}\boldsymbol{\varepsilon} \qquad\qquad \mathbf{N}\boldsymbol{\varepsilon}\sim N(\mathbf{0} \quad \mathbf{NMN}'). \tag{4-2}$$

Based on Eq.(4-2), the estimates of $\boldsymbol{\beta}_k$ is

$$\hat{\boldsymbol{\beta}}_k=(\bar{\mathbf{X}}'\bar{\mathbf{W}}_k^{1/2}(\mathbf{NN}')^{-1}\bar{\mathbf{W}}_k^{1/2}\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\bar{\mathbf{W}}_k^{1/2}(\mathbf{NN}')^{-1}\bar{\mathbf{W}}_k^{1/2}\bar{\mathbf{y}}, \tag{4-3}$$

where $\bar{\mathbf{X}}$ = **NX** (Eq.4-3 equals to Eq.3-14). Eq.(4-3) is a generalized least squares (GLS) estimator with its weighting matrix is $\bar{\mathbf{W}}_k^{1/2}(\mathbf{NN}')^{-1}\bar{\mathbf{W}}_k^{1/2}$. $\bar{\mathbf{W}}_k$ = $\mathbf{N}\mathbf{W}_k\mathbf{N}'$ is a diagonal matrix whose *i*-th diagonal represents spatial connectivity between *k*-th non-aggregate level unit and *i*-th aggregate level unit (the average connectivity between the *k*-th non-aggregate level unit and each non-aggregate level unit in the *i*-th aggregate level unit). As illustrated in Fig.4-1, the *i*-th diagonal considers the shapes of the *i*-th unit. **NN**′, which is another matrix in $\bar{\mathbf{W}}_k^{1/2}(\mathbf{NN}')^{-1}\bar{\mathbf{W}}_k^{1/2}$, is also a diagonal matrix whose *i*-th diagonal is large when the *i*-th spatial unit is small. In short, $\bar{\mathbf{W}}_k$ considers the shape of spatial units and **NN**′ considers the size of spatial units. Since $\bar{\mathbf{W}}_k$ and **NN**′ are diagonal matrixes, $\bar{\mathbf{W}}_k^{1/2}$ and $(\mathbf{NN}')^{-1}$ are computed efficiently.

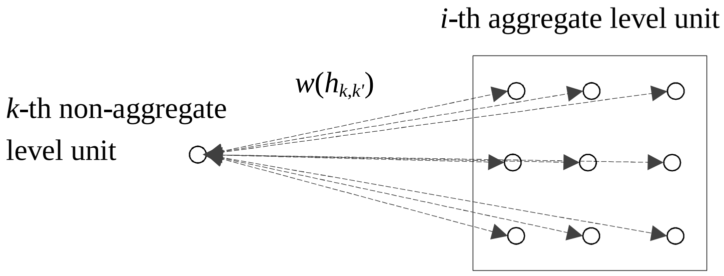


Figure 4-1: Image of spatial connectivity

Our aggregate-level model Eq.(4-2) is identical to standard GWR. Therefore, the variance–covariance matrix of $\hat{\boldsymbol{\beta}}_k$ is given as (see Fotheringham *et al.*, 2002)

$$Cov[\hat{\boldsymbol{\beta}}_k] = \hat{\sigma}^2 \mathbf{V}_k \mathbf{V}'_k, \tag{4-4}$$

$$\mathbf{V}_k = (\overline{\mathbf{X}}'\overline{\mathbf{W}}_k^{1/2}(\mathbf{NN}')^{-1}\overline{\mathbf{W}}_k^{1/2}\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\overline{\mathbf{W}}_k^{1/2}(\mathbf{NN}')^{-1}\overline{\mathbf{W}}_k^{1/2}$$

where $\hat{\sigma}^2$ denotes the estimates of σ^2. By substituting Eq.(4-3) into Eq.(4-2), the fitted values of $\overline{\mathbf{y}}$ are given by $\hat{\overline{\mathbf{y}}}$ =**NL** $\overline{\mathbf{y}}$, where **L** is a matrix whose *i*-th row is $\mathbf{x}'_k\mathbf{V}_k$, and $\mathbf{x}_k$ is a vector of explanatory variables observed at s_k. Using this property, $\hat{\sigma}^2$ is given as (see Cressie, 1998)

$$\hat{\sigma}^2 = \frac{(\overline{\mathbf{y}} - \mathbf{N}\hat{\boldsymbol{\mu}})'(\overline{\mathbf{y}} - \mathbf{N}\hat{\boldsymbol{\mu}})}{tr\{(\mathbf{I} - \mathbf{NL})(\mathbf{I} - \mathbf{NL})'\}}, \tag{4-5}$$

where $\hat{\boldsymbol{\mu}} = \left\langle \mathbf{x}'_k \hat{\boldsymbol{\beta}}_k \right\rangle$. Significance of $\boldsymbol{\beta}_k$ can be tested using diagonal elements of Eq.(4-4).

This method estimates non-aggregate-level parameters $\boldsymbol{\beta}_k$ irrespective of the aggregation units of data at hand. Besides, the estimators of $\boldsymbol{\beta}_k$, which are identical to the standard GLM estimators, are unbiased, consistent, efficient, and asymptotically normal. In other words, unlike the standard GWR, which does not consider aggregation mechanisms, ATP GWR can be considered a solution to MAUP (see, also, §4.1).

4.3. A simulation study

4.3.1. Outline

This section examines the effectiveness of ATP GWR for MAUP by applying a simulation study. There are at least two simulation approaches for GWR. The first utilizes the eigenvectors of a double-centered proximity matrix (see Wheeler and Tiefelsdorf, 2005; Paez *et al*., 2011). For example, Paez *et al*. (2011) apply the first, third, and fourth eigenvectors of a proximity matrix for their first, second, and third spatially varying parameters, respectively. This approach enables controlling collinearity among spatially varying parameters, which is a critical factor that determines the effectiveness of GWR (Wheeler and Tiefelsdorf, 2005).

The second approach models spatially varying parameters by using spatial processes (e.g., Finley, 2011). For instance, the spatial process whose covariance is modeled, based on the Gaussian covariance function Eq. (2-10), as

$$c(k,k') = \tau^2 \exp\left(-\frac{h_{k,k'}^2}{r^2} \right), \tag{4-6}$$

This function is consistent with the Gaussian kernel function Eq. (2-57), which is commonly used in GWR. Unlike the eigenvector-based approach, this approach enables us to control the spatial scales of spatially varying parameter distributions by tuning r.

Spatial scale is an essential factor determines the seriousness of MAUP. Besides, the influence of collinearity has already been discussed well in Wheeler and Tiefelsdorf (2005) and Paez *et al*. (2011). Hence, we conduct a simulation study of the latter type, focusing on MAUP and spatial scales and paying attention to the collinearity among spatially varying parameters.

In our simulation, we first generate non-aggregate-level response variables and explanatory variables on 50 × 50 sites. The explanatory variables include one intercept and two variables, $x_{k,1}$ and $x_{k,2}$, generated independently from $N(0,1)$, respectively. The response variables are generated using Eq. (4-7):

$$y_k = \alpha(k) + x_{k,1}\beta_1(k) + x_{k,2}\beta_2(k) + \varepsilon_k \qquad \varepsilon_k \sim N(0,\sigma^2), \tag{4-7}$$

where $\alpha(k)$, $\beta_1(k)$, and $\beta_2(k)$ are spatially varying parameters. They are generated using Gaussian processes whose means are zeros and covariance functions are Eq. (4-6), where the τ^2 values for both

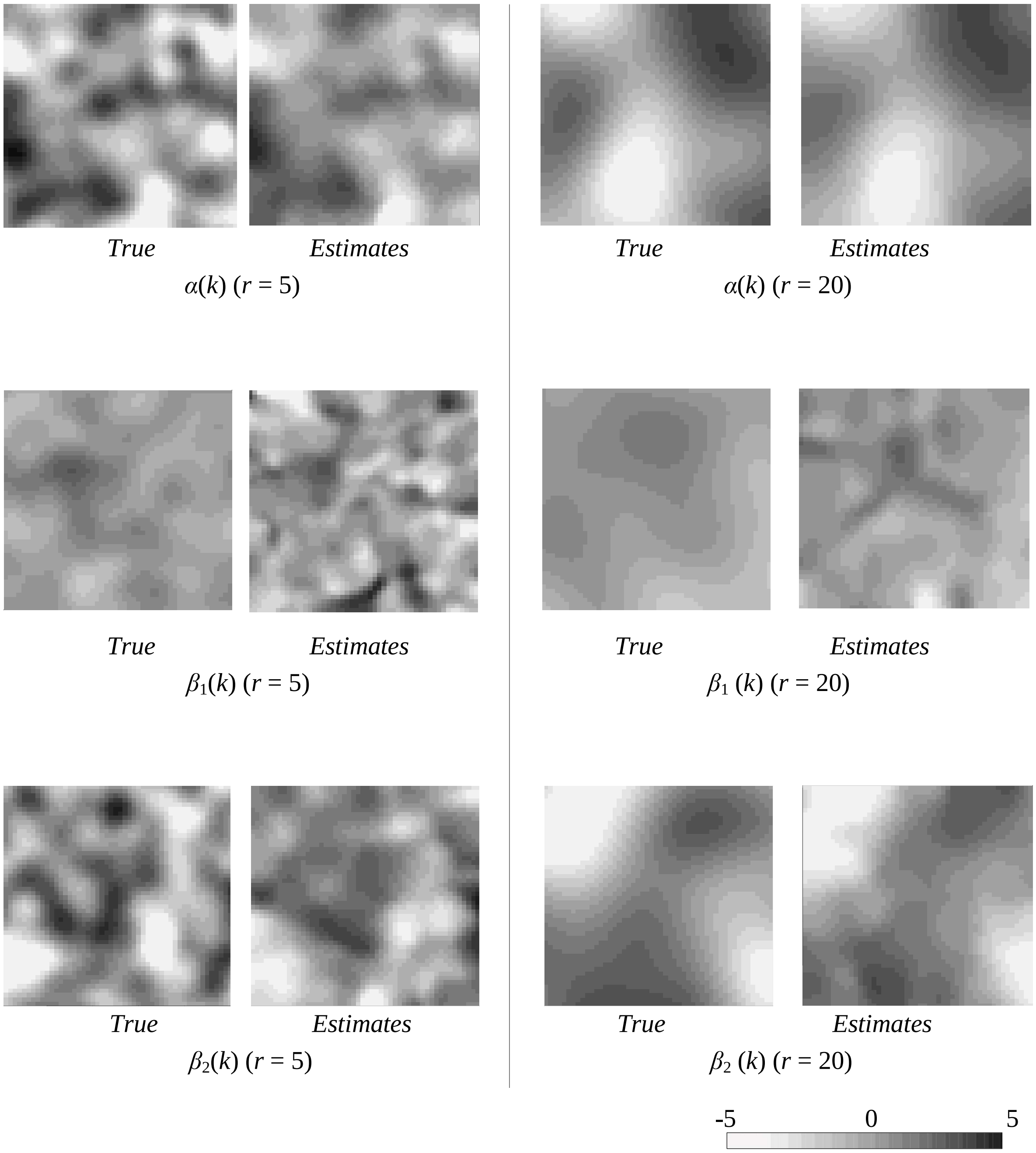
$\alpha(k)$ and $\beta_2(k)$ are 2.0 and that for $\beta_1(k)$ is 0.5. Note that the zero means of the parameter does not imply insignificance of them, and they possibly be significant at some sites in the assumed space. The zero means are also assumed in Paez *et al*. (2011). The intercept and $x_{k,2}$ corresponding to $\tau^2 = 2.0$ explain y_k effectively, whereas $x_{k,1}$ does not. Our simulations are performed by altering $\sigma^2 = \{1.0, 4.0\}$, $r = \{5, 10, 20\}$.

Under the above settings, we first generate the true distributions of $\alpha(k)$, $\beta_1(k)$, and $\beta_2(k)$ for each of the six (= 2 × 3) cases (the true distributions when $\sigma^2 = 1.0$ and $r = 5$ or 20 are shown in Fig. 4-2). Then, in each of the six cases, the following steps are iterated 100 times: (i) the non-aggregate-level variables $x_{k,1}$, $x_{k,2}$, and y_k are generated; (ii) $x_{k,1}$, $x_{k,2}$, and y_k are aggregated into M aggregation units, which are generated by Voronoi tessellation; (iii) the non-aggregate-level parameters in ATP GWR are estimated by using the aggregated variables; and (iv) the accuracies of the non-aggregate-level parameter estimates are measured by comparing them with their true values. If our (aggregate-level) model effectively recovers the non-aggregate-level parameters irrespective of the M aggregation units, we can say that the method is robust for MAUP. Considering the suggestion of Paez *et al*. (2011) that data applied for GWR should not be small, we assume $M = 400$.

Regular lattices (e.g., 50 × 50 grids) are usually not assumed much in GWR simulation studies. One of the reasons is that regular lattices do not appear to be representative of real-world geographical topologies (Farber *et al*., 2009). However, since the objective of COSP studies is to mitigate the influences of spatial supports (shape, size, etc.), most COSP simulation studies discuss the modeling of continuous spatial process, which is free from such spatial supports, and the continuous space is approximated by using a discrete spatial process on regular points (e.g., Kyriakidis and Yoo, 2005; Nagle *et al*., 2011). Thus, our assumption of 50 × 50 sites is consistent with the standard assumption in COSP studies.

4.3.2. Result

The estimates of $\alpha(k)$, $\beta_1(k)$, and $\beta_2(k)$ given in each of the first attempts with $\sigma^2 = 1.0$ and $r = \{5, 20\}$ are plotted in Fig.4-2. The results show the tendency of the accuracies of $\alpha(k)$ and $\beta_2(k)$, which explain y_k well, to be good and of the accuracy of $\beta_1(k)$ to be poor. Also, the estimates obtained when $r = 20$ are more accurate than the estimates obtained when $r = 5$.


True Estimates

$\alpha(k)$ (r = 5)

True Estimates

$\alpha(k)$ (r = 20)

True Estimates

$\beta_1(k)$ (r = 5)

True Estimates

β_1 (k) (r = 20)

True Estimates

$\beta_2(k)$ (r = 5)

True Estimates

β_2 (k) (r = 20)

Figure 4-2: Plots of estimated $\alpha(k)$ $\beta_1(k)$, and $\beta_2(k)$, and their true values (right) when $\sigma^2 = 1.0$

We measure the accuracies of the parameter estimates by using the root mean square error (RMSE) and R-squared (R^2). Since 2,500 (50 × 50) parameters are estimated in each attempt, the RMSEs and R^2s given in each attempt are averaged and plotted as shown in Fig.4-3. In this figure, the average RMSEs and R^2s obtained by the non-aggregate-level standard GWR (GWR_NAg) are also plotted for comparison. Note that since ATP GWR is an aggregate-level model, the results must be worse than the GWR_Nag results. This figure shows that the RMSEs and R^2s in our method change significantly depending on r, and that the change is particularly large when r is small. This result indicates that our method can be inefficient when the spatial process of spatially varying coefficient is too local. The RMSEs and R^2s also change depending on the explanation capabilities of the explanatory variables. Specifically, the average R^2s of $\alpha(k)$ and $\beta_2(k)$, which explain y_k well, are between 0.4 and 1.0, whereas those of $\beta_1(k)$ are between 0.1 and 0.6. This result suggests that the parameter estimates of our model should be discussed only when they are significant. In contrast, the impact of σ^2 is relatively small. In summary, ATP GWR effectively recovers the non-aggregate-level parameter (i.e., robust for MAUP), when the explanatory variables are significant and their spatial variations are not too local compared to their aggregation scales.

To examine collinearity among the estimated parameters, the correlations among the estimated parameters when σ^2 = 1.0 and r = {5, 20} are summarized in Fig.4-4. This figure suggests that any serious spurious correlation, which could occur even when the explanatory variables are uncorrelated, is not aroused in our simulation.

We then compare the bandwidth parameter estimates between two aggregate-level models: ATP GWR and the aggregate-level standard GWR (GWR_Ag: the GWR that models the aggregate-level variables; the geometric centers of each aggregation unit are used to calculate spatial connectivity). The average RMSEs of their bandwidth parameters are evaluated and plotted in Fig.4-5. Here, the estimates of GWR_NAg are regarded as their true values. As shown in this figure, the estimates of ATP GWR are more accurate than those of GWR_Ag in all cases. In each of the six cases, at least 91% of attempts indicate efficiency of ATP GWR over GWR_Ag. However, the estimates of ATP GWR are still upwardly biased, and this bias is particularly prominent when r is small (see Fig.4-6). We need to discuss the reduction of this bias in a future study.

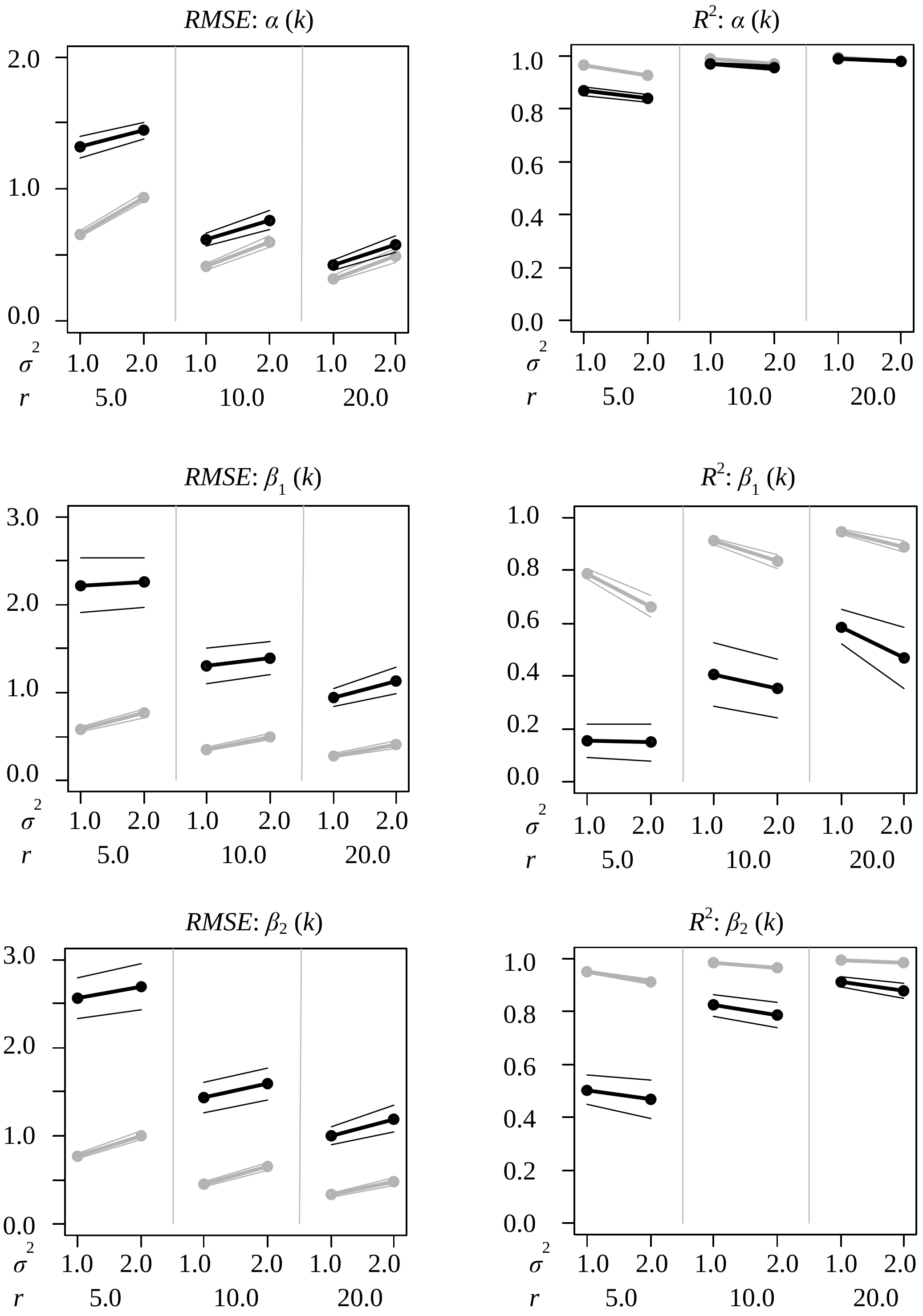


Figure 4-3: RMSEs and R^2s of the estimates of $\alpha(k)$, $\beta_1(k)$, and $\beta_2(k)$

Note: Here, the averages of the RMSEs and R^2s are plotted (black line: ATP GWR; gray line: GWR_NAg). The bold lines represent averages, and the gaps between the bold lines and the thin lines near them represent the standard deviations of the *RMSE*s or R^2s.

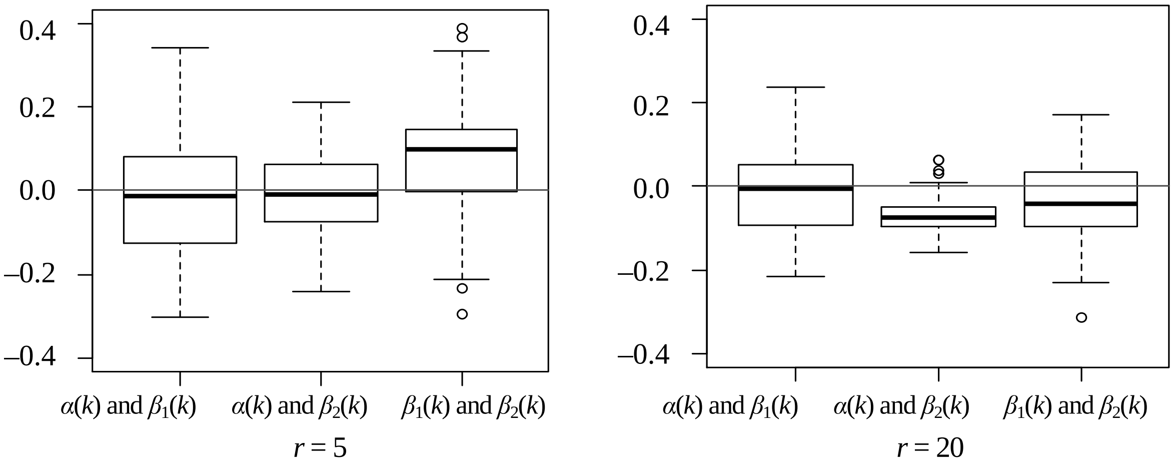


Figure 4-4: Correlation coefficients among spatially varying parameters (σ^2 = 1.0)

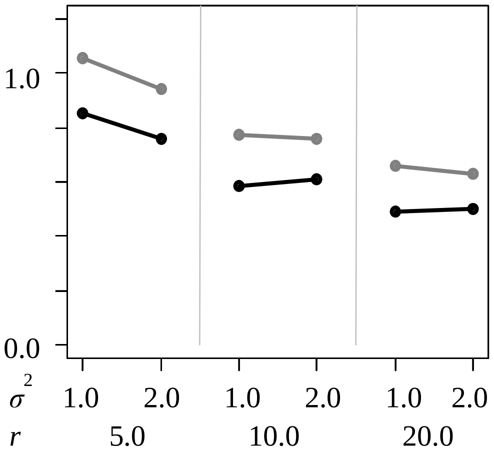


Figure 4-5: RMSEs of the bandwidth parameter estimates

Note: Black line: ATP GWR; Dark gray line: GWR_Ag. The true bandwidth parameter values are given by the estimates of GWR_Nag.

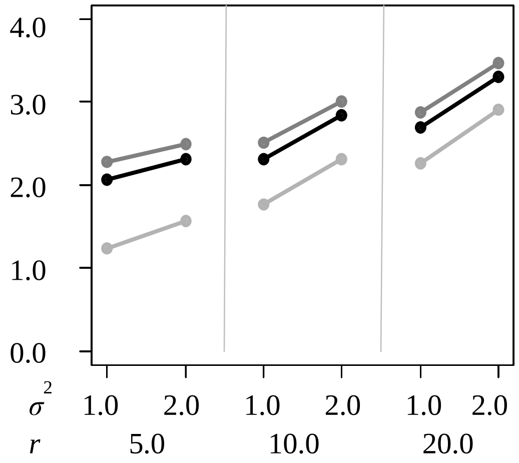


Figure 4-6: Averages of the bandwidth parameter estimates

Note: Black line: APGWR; Dark gray line: GWR_Ag; Light gray line: GWR_NAg.

4.4. An empirical study

4.4.1. Outline

In this section, we apply ATP GWR and GWR_Ag for the 2005 municipal-level crime data (sample size: 249; source: Criminal statistics, 2007) of the Tokyo metropolitan area, as shown in Fig.4-7. Our response variables are the number of crimes per km^2 (Fig.4-7), which we refer to as crime density. Since utilizing many explanatory variables in GWR could introduce serious multicollinearity (Wheeler and Tiefelsdorf, 2005), we apply only two explanatory variables: the constant and the population densities (thousand people par 1km^2; source: Population census, 2005), which are shown in Fig.4-8. In this analysis, GWR_Ag estimates the parameters in 249 municipal unit-level variables and ATP GWR estimates the parameters in (the geometric centers of) 10,247 minor municipal districts from the 249 samples.

We use R, a free statistical software provided by The Comprehensive R Archive Network (http://cran.r-project.org/), for computation, and ArcGIS, provided by ESRI Inc. (http://www.esri.com/), for mapping.

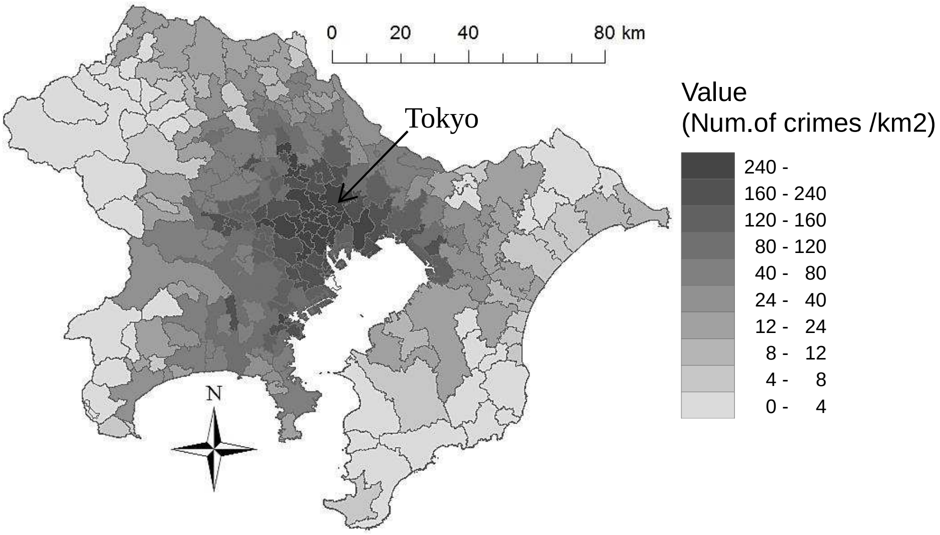


Figure 4-7: Crime densities in the municipal units

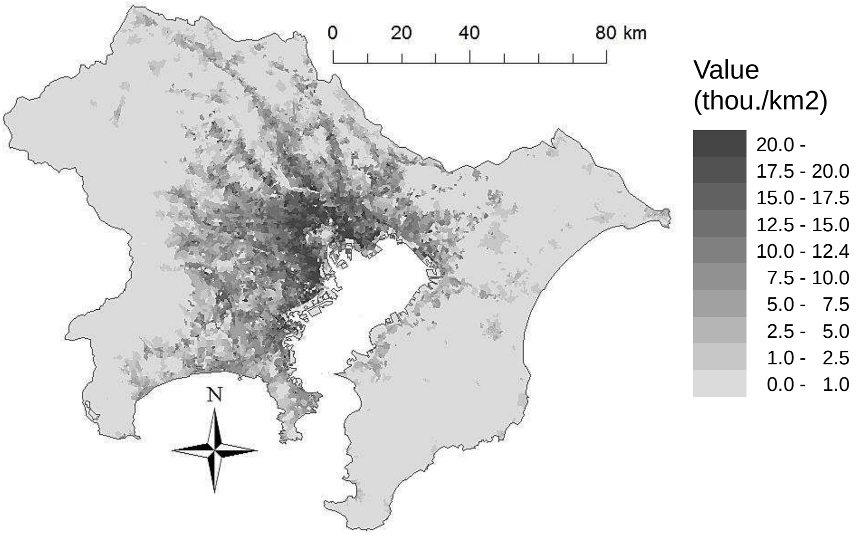
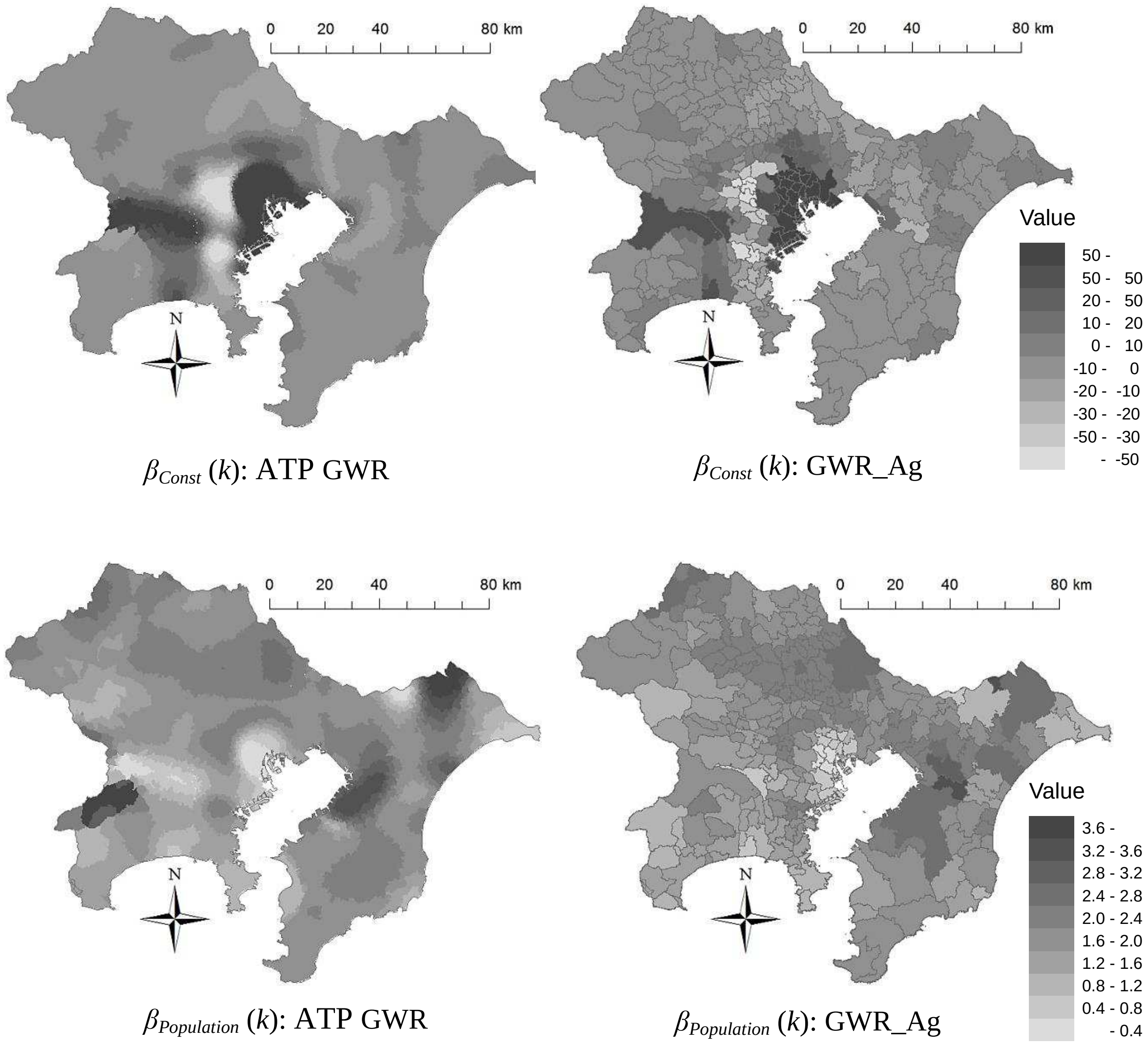


Figure 4-8: Population densities in the municipal units

4.4.2. Result

The bandwidth parameter estimates of ATP GWR and GWR_Ag are 4.89 km and 5.01 km, respectively. These results suggest that crime densities have local spatial variation. Fig.4-9 shows the spatial plots of the local trend parameter estimates, which we refer to as $\beta_{Const}(k)$ and $\beta_{Population}(k)$, and Fig.4-10 summarizes their significance levels. The result of ATP GWR is consistent with that of GWR_Ag. Besides, the estimates of ATP GWR, which are spatially smooth, seem to appear more natural. Since many GWR studies have discussed the spatial plots of their own parameter estimates, providing a seemingly natural result would be important.



β_{Const} (k): ATP GWR

β_{Const} (k): GWR_Ag

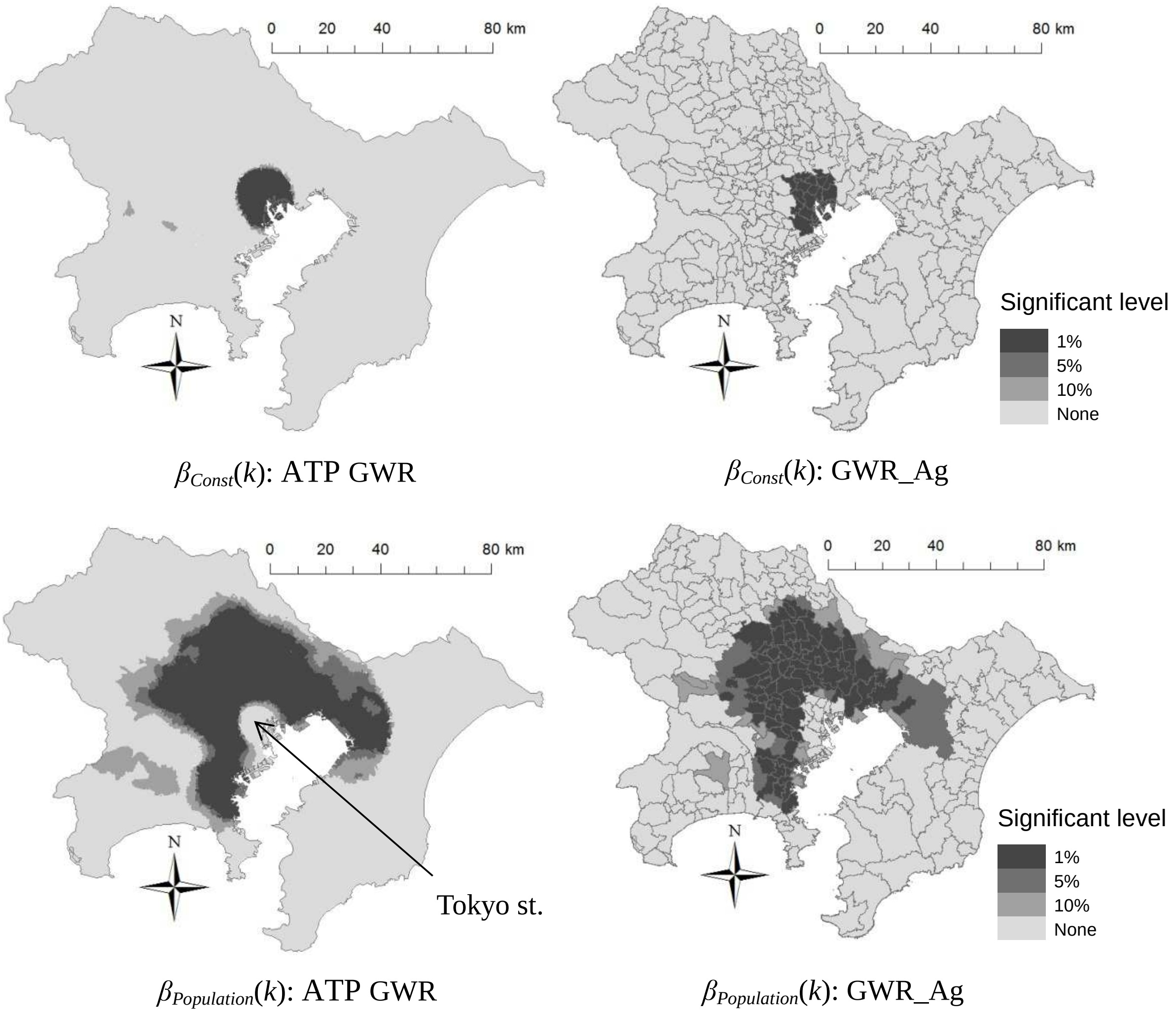
$\beta_{Population}$ (k): ATP GWR

$\beta_{Population}$ (k): GWR_Ag

Figure 4-9: Local trend parameter estimates

In each model, the estimates of $\beta_{Const}(k)$s are significantly high in the central Tokyo area. This result seems to indicate heterogeneity of this area. This result is intuitively consistent. On the other hand, the estimates of $\beta_{Population}(k)$s are significantly positive in the suburban areas of Tokyo, whose distance from the Tokyo station is between 10 km and 40 km, with significance particularly prominent in the northern area of Tokyo. Roughly, the significant area agrees with the commutable area of Tokyo with many populations (see Fig.4-8). Accordingly, our result could indicate the danger of having heavily populated areas in the commutable areas.

In summary, ATP GWR, which effectively mitigates MAUP, is useful for both simulation data and actual data.



$\beta_{Const}(k)$: ATP GWR

$\beta_{Const}(k)$: GWR_Ag

$\beta_{Population}(k)$: ATP GWR

$\beta_{Population}(k)$: GWR_Ag

Figure 4-10: Significance of the local trend parameter estimates

4.5. Summary

This study discussed the effectiveness of ATP GWR by focusing on MAUP. While several studies have often indicated that MAUP is yet to be resolved (e.g., Butkiewicz and Ross, 2010), ATP GWR, whose non-aggregate-level parameter estimates are unbiased, consistent, efficient, and asymptotically normal, can be considered a method to resolve MAUP. We confirmed the effectiveness of the method for MAUP in a simulation and an empirical study. Since the original GWR model, which does not consider aggregation mechanism, is not a solution to MAUP, my discussion of extending GWR for a solution to MAUP is significant.

However, our method still has some problems. First, our simulation study indicates the ineffectiveness of the method when spatially varying parameters have local spatial patterns. Some studies (e.g., Fisher and Langford, 1995) have shown that non-aggregate-level spatial patterns are in aggregate-level variables and can be effectively captured when detailed auxiliary data (e.g., high-resolution land use data) are considered in aggregation mechanisms. The aggregation mechanism in our model can easily be extended by modifying **N**. Hence, it is important that we consider the detailed auxiliary data in **N** to make our method more effective. Another problem is multicollinearity. As with the standard GWR, ATP GWR too seems to suffer from multicolliearity in many cases, particularly when the number of explanatory variables is large. To tackle this problem, applying a penalized form of GWR such as geographically weighted ridge regression (Wheeler, 2007) or geographically weighted lasso regression (Wheeler, 2009) model might be useful.

We have discussed MAUP while referring to COSP studies in geostatistics. Since the primary objective of COSP studies is to change spatial supports such as point interpolation and areal interpolation, MAUP has not been discussed in COSP literature sufficiently. Discussing MAUP in terms of COSP studies would be an important step toward developing more sophisticated solutions for MAUP.

In short, chapter 3 and 4 discussed two main COSPs for areal data: the areal interpolation problem and the MAUP, and showed that the proposed ATP GWR deals with these two problems effectively.

5. Point Interpolation Problem: An Eigenvector Spatial Filtering-based Approach

While the previous chapters discussed the COSPs for areal data, Chapters 5 and 6 discuss the COSPs for point data. Chapter 5 discusses the point interpolation problem, which has been discussed actively in the field of geostatistics. However, geostatistical methods have a number of drawbacks. First, they are not necessarily straightforward to implement and extend. Second, the methods can easily become computationally intractable, particularly when spatiotemporal data are interpolated.

Thus, I extend the ESF, which is simple and possible to model spatiotemporal data computationally efficiently, for the point interpolation problem. The effectiveness of the extended method is examined by applying it to land price interpolations. Note that the usability of the extended method is not restricted within the point interpolation problem. Hence, the extended method is also applied for several other purposes, including parameter estimation in the presence of spatial dependence, spatial component extraction, and fast computation.

Two types of point data are appeared in this section: continuous spatial (or geo-referenced/ point-referenced/geostatistical) data, i.e., the data distributed on $\Re^d$, and discrete spatial (or lattice) data; the data distributed on a discrete space.

5.1. Introduction

5.1.1. Review of point interpolation studies

Kriging (see §2.2.5) is a standard point interpolation method. There are variations in kriging. For instance, simple kriging (kriging with a known mean), ordinary kriging (kriging with an unknown and constant mean), university kriging (kriging with coordinates as explanatory variables), and regression kriging (kriging with explanatory variables) are prime linear geostatistical models. Then, log-normal kriging (kriging with log-transformed response variables), trans-Gaussian kriging (kriging with Box-Cox transformed response variables), and disjunctive kriging (kriging with non-linear transformed response variables) are prime non-linear geostatistical models (see, e.g., Cressie, 1993). These methods have the following features: (i) they minimize the MSPE; and (ii) they interpolate continuous spatial data. Feature (i) ensures that the point interpolation is accurate. On the other hand, since a complete observation of continuous spatial data is generally not possible, the interpolation of continuous spatial data is important. Thus, the feature (ii) increases the importance of kriging studies (see also, Longley *et al*., 2010).

Point interpolation problems have also been discussed in non-geostatistical spatial statistics. For instance, Martin (1983), Bennett *et al*. (1983), and LeSage and Pace (2004) discuss interpolation based on the SLM (see LeSage and Pace, 2009) and SEM models, Griffith and Paelinck (2011) discuss ESF-based interpolation, and Leung *et al*. (2000) and Harris *et al*. (2010; 2011) propose GWR-based interpolation techniques. The SLM/SEM-based and ESF-based approaches interpolate discrete spatial data by considering spatial dependence, and the GWR-based approach interpolates continuous spatial data by considering spatial heterogeneity.

5.1.2. Fundamentals of point interpolation

While the basic assumptions of the aforementioned point interpolation methods differ, their basic models (except for non-linear geostatistical models) are essentially identical. Their basic models are formulated as

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_0\end{pmatrix}=\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\mu}_0\end{pmatrix}+\begin{pmatrix}\boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon}_0\end{pmatrix}, \qquad \begin{pmatrix}\boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon}_0\end{pmatrix}\sim N\left[\begin{pmatrix}\mathbf{0}\\ \mathbf{0}_0\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{\Sigma}_0\\ \boldsymbol{\Sigma}_0' & \boldsymbol{\Sigma}_{00}\end{pmatrix}\right], \tag{5-1}$$

where the subscript “$_0$” indicates missing sites; $\boldsymbol{\mu}$ denotes a deterministic trend component; and $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_0$, and $\boldsymbol{\Sigma}_{00}$ are the matrix of covariance among observation sites, between observation sites and missing sites, and among missing sites, respectively. The predictors of each approach are defined by the conditional expectation of $\mathbf{y}_0$, $\hat{\mathbf{y}}_0$, which is given based on Eq.(5-1) as

$$\hat{\mathbf{y}}_0=\hat{\boldsymbol{\mu}}_0+\boldsymbol{\Sigma}_0'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\hat{\boldsymbol{\mu}}). \tag{5-2}$$

Eq.(5-2), which is identical to the kriging predictor given in Eq.(2-34), minimizes the MSE. In the other words, the kriging and non-geostatistical interpolation methods are essentially identical.

On the other hand, their implementation procedures are different between the methods for continuous spatial data, including kriging and the geographically weighted regression (GWR) approach, and the methods for discrete spatial data, including approaches based on the spatial lag model (SLM), the spatial error model (SEM), and the eigenvector spatial filtering (ESF). The continuous spatial data methods estimate parameters using observed data only (i.e., $\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$ in Eq.5-1 is used for parameter estimation), and perform an interpolation by substituting the estimated parameters into Eq.(5-2). The discrete spatial data methods use the EM-algorithm-based iterative calculation procedure, which is summarized as follows: (i) the initial values are set for the unobserved data, $\mathbf{y}_0$; (ii) the parameters are estimated using both observed and unobserved data (i.e., Eq.5-1 is used for parameter estimation); (iii) the unobserved data are updated by substituting the estimated parameters into Eq.(5-2); and (iv) iterate steps (ii) and (iii) until the unobserved data converge. Furthermore, the continuous spatial data methods model a continuous stochastic process, while the discrete spatial data methods model the spatial equilibrium that forms among observed and unobserved data (see Fig.5-1).

Although continuous spatial data interpolation is particularly important, as discussed previously, the SLM/SEM and ESF cannot be defined on a continuous space, owing to the algebraic limitations. More precisely, the ESF requires an eigen-decomposition of a matrix that describes the connectivity among all given sites. Since eigen-decomposition is tractable only for a finite dimensional matrix, the number of sites must be finite. Thus, the ESF is essentially a method in a discrete space. Similarly, since the SLM and SEM require an inversion of a proximity matrix, which is tractable only if the dimension of the proximity matrix is finite, they are also models in a discrete space. Consequently, they cannot interpolate a continuous spatial process.

Overcoming such a limitation is significant not only for point interpolation problems. It is also important to be able to apply the ESF or SLM/SEM to other problems, which have been discussed by modeling continuous spatial process, including the sampling design problem (e.g., Wang *et al*., 2012), gradient analysis (e.g., Banerjee, 2010), and block prediction (e.g., Cressie, 1993).

This chapter considers extending the ESF to continuous spatial data modeling. The extended model is then applied to spatial and spatiotemporal interpolation, as well as other purposes.

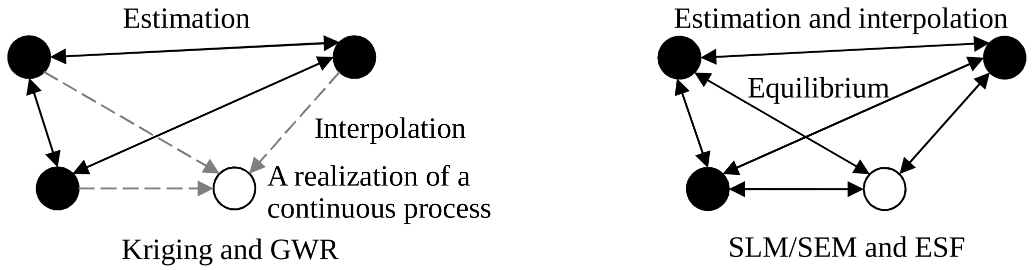

Estimation
Interpolation
A realization of a continuous process
Kriging and GWR
Estimation and interpolation
Equilibrium
SLM/SEM and ESF

Figure 5-1: Images of the point interpolation approaches

Note: Black circles denote observed sites and white circles denote unobserved sites

5.2. ESF on continuous space

This section extends ESF for modeling continuous spatial data. In §5.2.1, we briefly discuss the eigenfunction-based specification of the standard geostatistical model, which has actively been applied for continuous spatial data modeling. In §5.2.2, an eigenvector-based model (Eq.5-6), which describes continuous spatial phenomena, is constructed based on the geostatistical model. Then, we show that the model is a valid geostatistical model, and that it can be considered as a MC-based ESF model.

The continuous space model requires an eigen-decomposition of an infinite dimensional kernel matrix, which is computationally intractable. Hence, §5.2.3 and 5.2.4 discuss how this eigen-decomposition is performed. §5.2.3 shows that the infinite dimensional matrix can be divided into blocks (Eq. 5-21), given certain assumptions (Eqs.5-13, 5-14), and then §5.2.4 shows that the eigenfunctions of the infinite dimensional matrix can be approximated using an approximation technique called the Nyström extension (e.g., Drineas and Mahoney, 2005).

Based on discussions in §5.2.3 and 5.2.4, §5.2.5 modifies our model (Eq.5-6) to a tractable form (Eq.5-24). Subsequently, implementation of the model is discussed, focusing on accurate model identification problems and residual spatial dependence reduction problems. Finally, §5.2.6 compares the constructed method with the other eigenfunction-based spatial methods.

5.2.1. Geostatistics and eigenfunctions

The standard geostatistical model Eq.(2-18) can be expended, by decomposing **ε** into spatially dependent component and spatially independent component, as(e.g., Gneiting and Guttorp, 2010):

$$y_i = \mu_i + \eta_i + u_i \,, \tag{5-3}$$

where s_i ($i = 1,... n$) is a site in D, and μ_i is an element in $\boldsymbol{\mu}$. $\eta_i + u_i$ is an element in $\boldsymbol{\varepsilon}$ where η_i is a spatially dependent component, and $u_i \sim N(0, \sigma^2)$. The term η_i is modeled using a covariogram. For instance, η_i can be modeled using the exponential model Eq.(2-9) as

$$\begin{aligned} \mathrm{cov}(\eta_i, \eta_j) &= \tau^2 \exp(-h_{i,j} / r) \\ &= \tau^2 k(s_i, s_j) \,. \end{aligned} \tag{5-4}$$

As shown in the second line of Eq.(5-4), covariance functions are defined by the product of the variance parameter (partial-sill) τ^2 and a kernel function $k(s_i,s_j)$. Employing the eigen-decomposition for $k(s_i,s_j)$, η_i in Eq.(5-3) can be expanded as follows (e.g., Pintore and Holmes, 2004):

$$\eta_i = \sum_{l=1}^{\infty} e_{l,i} \gamma_l \,, \tag{5-5}$$

where $e_{l,i}$ is the l-th eigenfunction of $k(s_i, s_j)$. If $\sigma^2 > 0$, then geostatistical models are valid if and only if $k(s_i, s_j)$ is a positive semidefinite function (Cressie, 1993).

5.2.2. The eigenfunction-based model

Suppose that the region D is filled by an infinite number of points, including N observation sites. Eq.(5-3) may be rewritten using matrix notation as

$$\mathbf{y}^+ = \mathbf{X}^+\boldsymbol{\beta} + \mathbf{E}^+\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^+ \,, \qquad \boldsymbol{\varepsilon}^+ \sim N(\mathbf{0}^+, \sigma^2\mathbf{I}^+) \,, \tag{5-6}$$

where $\mathbf{y}^+$, $\mathbf{X}^+$, $\boldsymbol{\varepsilon}^+$, $\mathbf{0}^+$, and $\mathbf{I}^+$ respectively are **y**, **X**, **ε**, **0**, and **I** with dimensions of infinity, and $\mathbf{E}^+$ is a matrix of eigenfunctions extracted from a kernel matrix. $\mathbf{X}^+\boldsymbol{\beta}$, $\mathbf{E}^+\boldsymbol{\gamma}$, and $\boldsymbol{\varepsilon}^+$ in Eq.(5-6) correspond to μ_i, η_i (which is defined using Eq.5-5), and ε_i in Eq.(5-3), respectively. Following ESF, I use $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ for the $n^+ \times n^+$ kernel matrix, where $\mathbf{K}^+$ is an infinite dimensional standard kernel matrix, and $\mathbf{M}^+$ is $\mathbf{I}^+ - \mathbf{1}^+\mathbf{1}^{+\prime}/n^+$ or $\mathbf{I}^+ - \mathbf{X}^+(\mathbf{X}^{+\prime}\mathbf{X}^+)^{-1}\mathbf{X}^{+\prime}$.

Just like for the geostatistical model Eq.(5-3), $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ must be a positive semidefinite matrix. In other words, $|\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}| = |\mathbf{M}^{+}|^{2}|\mathbf{K}^{+}|$ must be non-negative. We can satisfy this condition by defining the elements in $\mathbf{K}^{+}$ using a positive semidefinite function. Following Eq.(5-4), we apply an (positive definite) exponential function $k(s_i, s_j) = \exp(-h_{i,j}/r)$, and following some studies of distance-based spatial filtering (e.g., Griffith and Peres-Neto, 2006; Dray *et al*, 2006; Griffith, 2010), r in the function is the longest distance in the minimum spanning tree covering the N observation sites distributed in D.

By design, the diagonals of $\mathbf{K}^{+}$ are not zero ($k(s_i, s_i) = \exp(-0/r) \neq 0$), and as a result, $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$ does not explain spatially dependent components, but rather a mixture of spatially dependent components and self-dependent components. This point is inconsistent with standard ESF, which models spatially dependent components only. However, the unneeded self-dependent components in $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$ can be detached as

$$\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+} = \mathbf{M}^{+}\mathbf{K}_0^{+}\mathbf{M}^{+} + \mathbf{M}^{+}, \tag{5-7}$$

where $\mathbf{K}_0^{+}$ is $\mathbf{K}^{+}$ with its diagonals replaced with zeros. $\mathbf{M}^{+}\mathbf{K}_0^{+}\mathbf{M}^{+}$ explains the spatially dependent components, and $\mathbf{M}^{+}$ (= $\mathbf{M}^{+}\mathbf{I}^{+}\mathbf{M}^{+}$) explains the self-dependent components.

Thus, the eigenfunctions of $\mathbf{M}^{+}\mathbf{K}_0^{+}\mathbf{M}^{+}$, rather than of $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$, should be used in a spatial dependence analysis. Fortunately, eigenfunctions of these matrixes are identical (Griffith, 2003). Furthermore, the diagonal matrix of eigenvalues of $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$, $\mathbf{\Lambda}^{+}$, and the same matrix of $\mathbf{M}^{+}\mathbf{K}_0^{+}\mathbf{M}^{+}$, $\mathbf{\Lambda}_0^{+}$, have the following relationship:

$$\begin{aligned}
\mathbf{\Lambda}^{+} &= \mathbf{E}_{full}^{+\prime}\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}\mathbf{E}_{full}^{+} \\
&= \mathbf{E}_{full}^{+\prime}\mathbf{M}^{+}\mathbf{K}_0^{+}\mathbf{M}^{+}\mathbf{E}_{full}^{+} + \mathbf{E}_{full}^{+\prime}\mathbf{M}^{+}\mathbf{E}_{full}^{+} \\
&= \mathbf{\Lambda}_0^{+} + \begin{pmatrix} \mathbf{E}_{full-K}^{+\prime} \\ \mathbf{E}_K^{+\prime} \end{pmatrix} \mathbf{M}^{+} \begin{pmatrix} \mathbf{E}_{full-K}^{+} & \mathbf{E}_K^{+} \end{pmatrix} \\
&= \mathbf{\Lambda}_0^{+} + \begin{pmatrix} \mathbf{I}_{full-K}^{+} & \mathbf{0}^{+\prime} \\ \mathbf{0}^{+} & \mathbf{0}_K \end{pmatrix},
\end{aligned} \tag{5-8}$$

where $\mathbf{E}_K^{+}$ is the subset composed of K eigenfunctions whose eigenvalues are zeros, $\mathbf{E}_{full-K}^{+}$ is the subset composed of the other eigenfunctions (i.e., $\mathbf{E}^{+}_{full} = [\mathbf{E}^{+}_{full-K}, \mathbf{E}^{+}_{K}]$), $\mathbf{I}_{full-K}^{+}$ is an identity matrix, $\mathbf{0}^{+}$ and $\mathbf{0}_K$ are matrixes of zeros. Because $\mathbf{M}^{+}$ induces K zero eigenvalues that are the same for matrixes

$\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ and $\mathbf{M}^+\mathbf{K}_0{}^+\mathbf{M}^+$, Eq.(5-8) suggests that the $n-K$ remaining eigenvalues of $\mathbf{M}^+\mathbf{K}_0{}^+\mathbf{M}^+$ are their counterparts for $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ minus 1.

After all, if only the eigenvalues of $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ are replaced with the eigenvalues of $\mathbf{M}^+\mathbf{K}_0{}^+\mathbf{M}^+$ using Eq.(5-8), Eq.(5-6) can be considered as a model that describes pure spatial dependence (without self-dependence). Precisely, $\mathbf{E}^+\boldsymbol{\gamma}$ in Eq.(5-6) furnishes distinct map pattern descriptions of latent spatial dependence that is explained by MC^+, which is defined for D as[1]

$$MC^+ = \lim_{n\to\infty} \frac{n}{\mathbf{1}'\mathbf{K}_0\mathbf{1}} \frac{\mathbf{z}'\mathbf{M}\mathbf{K}_0\mathbf{M}\mathbf{z}}{\mathbf{z}'\mathbf{M}\mathbf{z}}$$

$$= \lim_{n\to\infty} \frac{n}{\sum_i \sum_{j\neq i} k(s_i, s_j)} \frac{\sum_i \sum_{j\neq i} k(s_i, s_j)\tilde{z}_i \tilde{z}_j}{\sum_i \tilde{z}_i^2} \quad , \tag{5-9}$$

where $\tilde{z}_i$ is the i-th element of $\mathbf{Mz}$. By construction, mean of $\tilde{z}_i$ is 0. Because $\sum_i \tilde{z}_i{}^2 \big/ n$ represents the variance of $\tilde{z}_i$, Eq.(5-9) can be expanded as

$$MC^+ = \lim_{n\to\infty} \frac{1}{\sigma_z^2} \frac{\sum_i \sum_{j\neq i} k(s_i, s_j)\tilde{z}_i \tilde{z}_j}{\sum_i \sum_{j\neq i} k(s_i, s_j)} \quad , \tag{5-10}$$

where $\sigma_z{}^2 = \sum_i \tilde{z}_i{}^2 \big/ n$. Under the assumption of infill asymptotics, which fills D by an infinite number of missing sites, Eq.(5-10) may be further expanded as

$$MC^+ = \frac{1}{\sigma_z^2} \frac{\int_{j\neq i}\int_i k(s_i, s_j)\tilde{z}_i \tilde{z}_j ds_i ds_j}{\int_{j\neq i}\int_i k(s_i, s_j) ds_i ds_j} \quad . \tag{5-11}$$

We assume a finite number of missing sites (i.e., $n \to \infty$), whereas the number of observation sites N is unchanged. Hence, r in $k(s_i, s_j) = \exp(-h_{i,j}/r)$, which is determined based on the N observation sites, also is unchanged.

When $Cov[\tilde{z}_i, \tilde{z}_j] = E[\tilde{z}_i \tilde{z}_j] = 0$, the expectation of MC^+ yields

$$E[MC^+] = \frac{1}{\sigma_z^2} \frac{\int_{j\neq i}\int_i k(s_i, s_j) E[\tilde{z}_i, \tilde{z}_j] ds_i ds_j}{\int_{j\neq i}\int_i k(s_i, s_j) ds_i ds_j} = 0 \quad . \tag{5-12}$$

[1] Eq.(5-9) can be written as $MC^+ = \frac{n^+}{\mathbf{1}^{+\prime}\mathbf{K}_0^+\mathbf{1}^+} \frac{\mathbf{z}^{+\prime}\mathbf{M}^+\mathbf{K}_0^+\mathbf{M}^+\mathbf{z}^+}{\mathbf{z}^{+\prime}\mathbf{M}^+\mathbf{z}^+}$, too.

On the other hand, MC^+ is large when the co-variations $\tilde{z}_i \tilde{z}_j$ are large and/or the co-variations are positively related to $k(s_i, s_j)$ (i.e., the co-variations are explained by $k(s_i, s_j)$).

In summary, this section defines our model Eq.(5-6), which is based on a geostatistical model, and shows that it captures spatial dependence described by MC^+, MC defined in a continuous study region.

5.2.3. The double centered kernel matrix in continuous space

Our model Eq.(5-6) requires an eigen-decomposition of $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ (or $\mathbf{M}^+\mathbf{K}_0{}^+\mathbf{M}^+$), which is computationally intractable. To achieve it, in this subsection, I expand $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ to a tractable form. After some assumptions are imposed in §5.2.3.1, we approximate $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ using a finite number of observations (§5.2.3.2). The result is used in §5.2.4 for the eigen-decomposition approximation.

5.2.3.1. Assumptions

Because of the existence of the projection matrix $\mathbf{M}^+$, expressing similarities among arbitrary sites in D using $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ directly is difficult. Hence, the following assumptions are imposed:

$$\int k(s_i, s_j) p_j ds_j \approx \frac{1}{N}\sum_{J=1}^{N} k(s_i, s_J)\,, \tag{5-13}$$

$$\int k(s_i, s_j) p_i ds_i \approx \frac{1}{N}\sum_{I=1}^{N} k(s_I, s_j)\,, \tag{5-14}$$

where p_i is a probably density function, and s_I (I: 1,...N) and s_J (J: 1,...N) are observation sites. Combining Eqs.(5-13) and (5-14) yields

$$\int\int k(s_i, s_j) p_i p_j ds_i ds_j \approx \frac{1}{N^2}\sum_{I=1}^{N}\sum_{J=1}^{N} k(s_I, s_J)\,, \tag{5-15}$$

Eqs.(5-13), (5-14), and (5-15) assume that the average similarities among arbitrary sites are approximated by the average similarities among observation sites.

5.2.3.2. The double centered kernel matrix

§5.2.3.2 discusses expansion of $\mathbf{M}^+\mathbf{K}^+\mathbf{M}^+$ to a tractable matrix. First, the (i, j)-th element of

$\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$ given by $k^{*}(s_i, s_j)$ is approximated using our assumptions defined in §5.2.3.1. Then, the tractable version of $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$ (Eq.5-21) is derived using the approximated $k^{*}(s_i, s_j)$ (Eq.5-17). The result is used in §5.2.4 to calculate eigenfunctions.

Suppose $\mathbf{M}^{+}= \mathbf{I}^{+} - \mathbf{1}^{+}\mathbf{1}^{+\prime}/n^{+}$, the (i, j)-th element of $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$, is expressed under infill asymptotics as

$$k^{*}(s_i,s_j) = k(s_i,s_j) - \int k(s_i,s_j)p_j ds_j - \int k(s_i,s_j)p_i ds_i + \int\int k(s_i,s_j)p_i p_j ds_i ds_j \tag{5-16}$$

Eq.(5-16) is written using matrix notation as $\mathbf{K}^{+} - (\mathbf{1}^{+}\mathbf{1}^{+\prime}/n^{+})\mathbf{K}^{+} - \mathbf{K}^{+}(\mathbf{1}^{+}\mathbf{1}^{+\prime}/n^{+}) + (\mathbf{1}^{+}\mathbf{1}^{+\prime}/n^{+})\mathbf{K}^{+}(\mathbf{1}^{+}\mathbf{1}^{+\prime}/n^{+})$. Eq.(5-16) is approximated using the assumptions Eqs.(5-13), (5-14), and (5-15) as

$$k^{*}(s_i,s_j) \approx k(s_i,s_j) - \frac{1}{N}\sum_{J=1}^{N} k(s_i,s_J) - \frac{1}{N}\sum_{I=1}^{N} k(s_I,s_j) - \frac{1}{N^2}\sum_{I=1}^{N}\sum_{J=1}^{N} k(s_I,s_J)\,. \tag{5-17}$$

Based on Eq.(5-17), the similarity between observation sites s_I and s_J is

$$k^{*}(s_I,s_J) \approx k(s_I,s_J) - \frac{1}{N}\sum_{J=1}^{N} k(s_I,s_J) - \frac{1}{N}\sum_{I=1}^{N} k(s_I,s_J) - \frac{1}{N^2}\sum_{I=1}^{N}\sum_{J=1}^{N} k(s_I,s_J)\,, \tag{5-18}$$

whereas the similarity between an arbitrary site s_i and an observation site s_J is

$$k^{*}(s_i,s_J) \approx k(s_i,s_J) - \frac{1}{N}\sum_{J=1}^{N} k(s_i,s_J) - \frac{1}{N}\sum_{I=1}^{N} k(s_I,s_J) - \frac{1}{N^2}\sum_{I=1}^{N}\sum_{J=1}^{N} k(s_I,s_J)\,. \tag{5-19}$$

On the one hand, the $N \times N$ matrix whose (I, J)-th element is given by Eq.(5-18) equals **MKM**, the kernel matrix regarding the observation sites. On the other hand, a $1\times N$ vector with its J-th element given by Eq.(5-19) results in $\mathbf{k}^{*}$, a vector representing similarities between an arbitrary site s_i and each observation site:

$$\begin{aligned}\mathbf{k}^{*}{}_{i} &= \mathbf{k}_i - \mathbf{k}_i\mathbf{1}\mathbf{1}'/N - \mathbf{1}'\mathbf{K}/N + \mathbf{1}'\mathbf{K}\mathbf{1}\mathbf{1}'/N^2 \\ &= \mathbf{k}_i(\mathbf{I} - \mathbf{1}\mathbf{1}'/N) - \mathbf{1}'\mathbf{K}/N(\mathbf{I} - \mathbf{1}\mathbf{1}'/N) \\ &= (\mathbf{k}_i - \mathbf{1}'\mathbf{K}/N)(\mathbf{I} - \mathbf{1}\mathbf{1}'/N)\,,\end{aligned} \tag{5-20}$$

where $\mathbf{k}_i$ is a $1\times N$ vector whose J-th element is $k(s_i, s_J)$. In short, Eqs.(5-13), (5-14), and (5-15) and $\mathbf{M}^{+}=\mathbf{I}^{+}-\mathbf{1}^{+}\mathbf{1}^{+\prime}/n^{+}$ imply that $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$ is

$$\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+} = \begin{pmatrix} \mathbf{MKM} & - \\ \mathbf{m}_i\mathbf{M} & - \\ - & - \end{pmatrix}. \tag{5-21}$$

where $\mathbf{m}_i = \mathbf{k}_i - \mathbf{1}'\mathbf{K}/N$, and “–” in Eq.(5-21) represents unspecified sub-matrixes.

When $\mathbf{M}^{+} = \mathbf{I}^{+} - \mathbf{X}^{+}(\mathbf{X}^{+\prime}\,\mathbf{X}^{+})^{-1}\mathbf{X}^{+\prime}$, by performing a similar analysis, $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$ is given by Eq.(5-21) with $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{m}_i = \mathbf{k}_i - \mathbf{x}_i(\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'\,\mathbf{K}$, where $\mathbf{x}_i$ is a $1\times K$ vector of K explanatory variables observed at an arbitrary site s_i. Eq.(5-21) implies that $\mathbf{MKM}$ is a part of $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$, according to our assumptions.

5.2.4. Eigenfunction extraction using the Nyström extension

Eq.(5-21) is helpful to approximate its eigenfunctions. Let $\begin{pmatrix} \mathbf{A} & \mathbf{B}' \\ \mathbf{B} & \mathbf{C} \end{pmatrix}$ be a $(A+B) \times (A+B)$ matrix for which $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are matrixes whose sizes are $A\times A$, $B\times A$, and $B\times B$. Suppose that the first A-columns in the $A+B$ columns are randomly selected. Then, the eigenfunctions of $\begin{pmatrix} \mathbf{A} & \mathbf{B}' \\ \mathbf{B} & \mathbf{C} \end{pmatrix}$ are approximated by the Nyström extension as

$$\begin{pmatrix} \mathbf{E}_A \\ \mathbf{B}\mathbf{E}_A\mathbf{\Lambda}_A^{-1} \end{pmatrix}, \tag{5-22}$$

where $\mathbf{A} = \mathbf{E}_A\mathbf{\Lambda}_A\mathbf{E}_A'$. The dimension of Eq.(5-22) is not $(A+B) \times (A+B)$ but rather $(A+B) \times A$. The Nyström extension performs a low-rank approximation (see, e.g., Cressie and Wikle, 2011).

In our case, the eigenfunctions of $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$, $\mathbf{E}^{+}_{full}$, are approximated, under the assumption that N observation sites are randomly sampled in D, as

$$\mathbf{E}^{+}_{full} = \begin{pmatrix} \mathbf{E}_{full} \\ \mathbf{e}_{i,full} \\ - \end{pmatrix}, \tag{5-23}$$

where $\mathbf{e}_{i,full} = \mathbf{k}_i^{*}\mathbf{E}_{full}\mathbf{\Lambda}^{-1}$ and $\mathbf{MKM} = \mathbf{E}_{full}\mathbf{\Lambda}\mathbf{E}_{full}'$. Eq.(3-27) suggests that the eigenfunctions at any site in D are approximated by $\mathbf{e}_{i,\,full}$.

5.2.5. Implementation of the method

Based on discussions in §5.2.3 and 5.2.4, we first expand our basic model Eq.(5-6) to a tractable form (Eq.5-24). Then, its basic parameter estimation procedure is explained. Details of the procedure are discussed with a focus on accurate model identification (§5.2.5.2) and residual spatial dependence reduction (§5.2.5.3).

5.2.5.1. Estimation steps

Eq.(5-6) may be rewritten using Eq.(5-23) as

$$\begin{pmatrix} \mathbf{y} \\ y_i \\ - \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{x}_i \\ - \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{E} \\ \mathbf{e}_i \\ - \end{pmatrix} \boldsymbol{\gamma} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \varepsilon_i \\ - \end{pmatrix}, \tag{5-24}$$

where {$\mathbf{y}'$, y_i, $-'$ }$'$ = $\mathbf{y}^+$, {$\mathbf{X}'$, $\mathbf{x}_i'$, $-'$ }$'$ = $\mathbf{X}^+$, {$\mathbf{E}'$, $\mathbf{e}_i'$, $-'$ }$'$ = $\mathbf{E}^+$, and {$\boldsymbol{\varepsilon}'$, ε_i, $-'$ }$'$ = $\boldsymbol{\varepsilon}^+$ (see Eq.5-6), $\mathbf{E}^+$ is a subset of L eigenfunctions in $\mathbf{E}^+_{full}$, $\mathbf{y}$, $\mathbf{X}$, $\mathbf{E}$, and $\boldsymbol{\varepsilon}$ are matrixes/vectors defined on N observation sites, and y_i, $\mathbf{x}_i$, $\mathbf{e}_i$, and ε_i are variables defined on an unobserved site. Eq.(5-24) contains the following sub-model regarding observation sites:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}). \tag{5-25}$$

Estimation of parameters in Eq.(5-25) may be done using the sub-model. The estimation procedure is as follows: (i) Extract $\mathbf{E}_{full}$ from **MKM**; (ii) Select eigenfunctions in $\mathbf{E}_{full}$ whose eigenvalues are greater than some threshold value, and construct **E**; and, (iii) Apply an OLS-based stepwise selection procedure for the sub-model. Following studies of ESF (e.g., Tiefelsdorf and Griffith, 2007), we recommend the forward selection stepwise method.

The purpose of estimation is identifying Eq.(5-25), which has eigenfunctions of $\mathbf{M}^+\mathbf{K}_0^+\mathbf{M}^+$. Hence, the eigenfunction selection step (ii) must be performed using eigenvalues $\lambda^+_{l\_0}$ of $\mathbf{M}^+\mathbf{K}_0^+\mathbf{M}^+$. $\lambda^+_{l\_0}$ cannot be evaluated directly. However, the following relationship holds between $\lambda^+_{l\_0}$ and the eigenvalues λ_{l_0} of $\mathbf{MK}_0\mathbf{M}$ (see Fig.5-2):

$$\lambda^+_{l\_0} = \begin{cases} 0 & if\ \lambda_{l\_0} = 0 \\ \alpha(\lambda_{l\_0} + 1) - 1 & otherwise \end{cases}, \tag{5-26}$$

where $\alpha = \lim_{n \to \infty} n/N$, which is introduced because of the Nyström extension (see Williams and Seeger, 2001). Eq.(5-26) suggests that $\lambda^{+}_{l\_0}$ is proportional to λ_{l_0}. Hence, $\lambda^{+}_{l\_0}$ can be evaluated using λ_{l_0}.

Because $\alpha > 1$ and $\lambda_{l\_0}+1 > 0$ ($\lambda_{l\_0}+1$ equal the eigenvalues of **MKM**, which are positive semidefinite: see Eq.5-8), Eq.(5-26) also means that $\lambda^{+}_{l\_0}$ corresponding to λ_{l_0} is always positive, unless $\lambda_{l\_0} = 0$. After all, eigenfunctions corresponding to any λ_{l_0} describe positive spatial dependence in *D* (see Fig. 5-2). This result is consistent with an indication in Griffith (2006) that negative spatial dependent is not of interest.

We now discuss details of steps (ii) and (iii) assuming two purposes for applying our method: accurate model identification (Case 1); and, residual spatial dependence reduction (Case 2). The model given in Case 1 might be useful for an exploratory spatial data analysis (ESDA) such as spatial interpolation and spatial pattern analysis. Case 2 is helpful when avoiding bias in parameter estimates and/or their standard errors due to spatial dependence (e.g., LeSage and Pace, 2009).

§5.2.5.2 and §5.2.5.3 discuss Case 1 and 2, respectively.

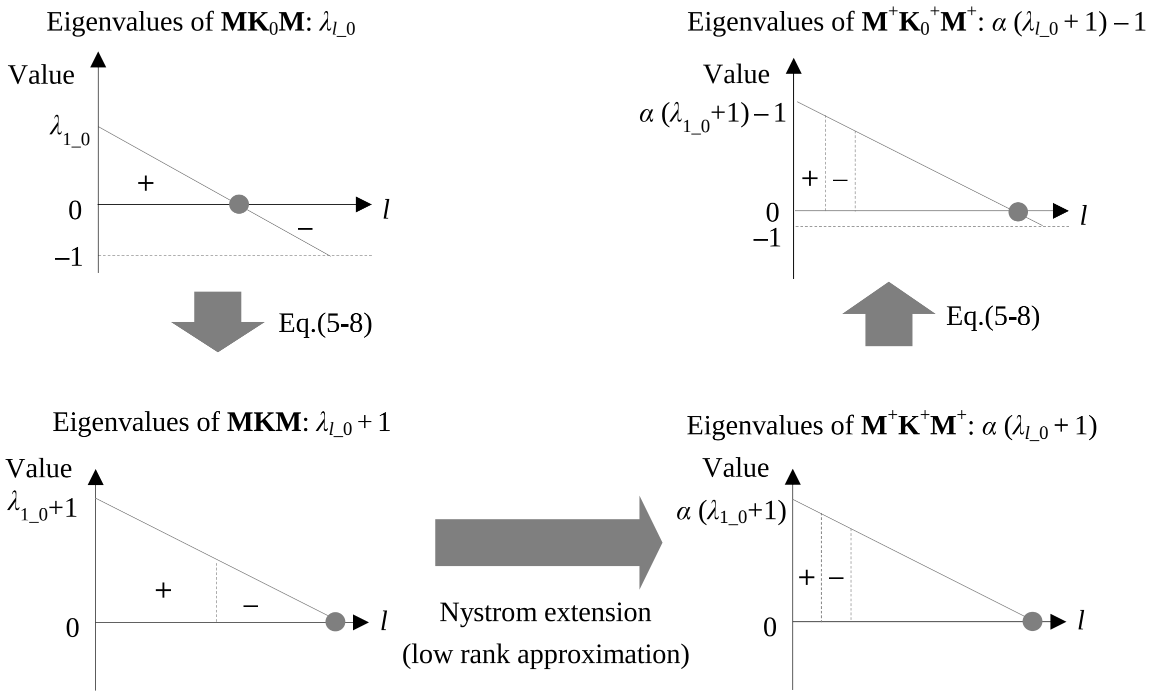

Eigenvalues of $\mathbf{MK}_0\mathbf{M}$: λ_{l_0}
Value
λ_{1_0}
+
0
–
–1
l
Eq.(5-8)
Eigenvalues of $\mathbf{M}^{+}\mathbf{K}_0^{+}\mathbf{M}^{+}$: $\alpha(\lambda_{l\_0}+1)-1$
Value
$\alpha(\lambda_{1\_0}+1)-1$
+ –
0
–1
l
Eq.(5-8)
Eigenvalues of **MKM**: $\lambda_{l\_0}+1$
Value
$\lambda_{1\_0}+1$
+
–
0
l
Nystrom extension
(low rank approximation)
Eigenvalues of $\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}$: $\alpha(\lambda_{l\_0}+1)$
Value
$\alpha(\lambda_{1\_0}+1)$
+ –
0
l

Figure 5-2: Relationship among eigenvalues

5.2.5.2. Estimation for accurate model identification

Considering eigenfunctions with not only large eigenvalues but also small eigenvalues is important for accurate model identification (Aubry *et al*., 1993; Cressie and Wikle, 2011). Hence, with regard to Case 1, this study removes no eigenfunctions in step (ii).

In contrast, in the subsequent step (iii), a large number of eigenfunctions must be considered in the forward stepwise procedure. We apply the AICc Eq.(5-27), which is robust in such a situation (Burnham and Anderson, 2002), for the objective function of our stepwise variable selection:

$$AICc = N\left[\log\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{N}\right) + 1 + \log 2\pi\right] + \frac{2K(K+L+1)}{N-K-L-1}\,. \tag{5-27}$$

The AICc-minimization-based forward stepwise regression technique can become computationally intensive. To cope with this problem, the AICc-minimization is replaced with an efficient algorithm. Suppose **M**= **I**–**X** (**X**′ **X**)$^{-1}$**X**′; then the following equation holds:

$$\left(\mathbf{X}', \mathbf{E}', \mathbf{e}_0'\right)\begin{pmatrix}\mathbf{X}\\ \mathbf{E}\\ \mathbf{e}_0\end{pmatrix} = \begin{pmatrix}\mathbf{X}'\mathbf{X} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & 1\end{pmatrix}, \tag{5-28}$$

where **E** is a subset of eigenvectors that is selected in earlier steps, and $\mathbf{e}_0$ is a candidate eigenvector to be entered into **E**, If Eq.(5-28) holds, the decrease in **ε**′**ε** by introducing $\mathbf{e}_0$ is always $||\mathbf{e}'_0\mathbf{y}||$ (Schott, 2005). Hence, in each stepwise selection step, the eigenvector with the greatest AICc improvement is the eigenvector that decreases **ε**′**ε** (=$||\mathbf{e}'_0\mathbf{y}||$) the most. Consequently, forward stepwise regression can be replaced with a simple algorithm that introduces eigenvectors in a decreasing order of $||\mathbf{e}_0\mathbf{y}||$ until AICc is minimized (also see Griffith, 2004b).

This algorithm is not available when **M** = **I**–**11**′/*n* and **X** contains explanatory variables other than an intercept. In this case, an exhaustive search is complicated by multicolinearity within **X** and among **X** and **E**. Thus, this study performs AICc minimization using **M**= **I**–**X** (**X**′ **X**)$^{-1}$**X**′ only.

5.2.5.3. Estimation for residual spatial dependence reduction

In Case 2, which seems helpful to avoid bias in parameter estimates due to spatial dependence, we apply both **M**= **I**–**X** (**X**′ **X**)$^{-1}$**X**′ and **M** = **I**–**11**′/*n*. To save degrees of freedom, we select eigenfunctions satisfying λ_{l_0} > 0 prior to the stepwise regression step.

Tiefelsdorf and Griffith (2007) demonstrate, using standard ESF, that two types of forward stepwise methods are useful for spatial dependence reduction: the MC-based method, and the accuracy-based method. The former selects eigenfunctions until the standardized MC of the residuals, z(*MC*), decreases to within a small absolute deviation from zero, δ, and the latter is a standard forward stepwise procedure that maximizes measure of model accuracy. Here, we demonstrate applying these two approaches for spatial dependence reduction, too.

5.2.6. Relationships among methods

At least three methods model spatial data using MC-based eigenvectors (eigenfunctions): ESF, Moran's eigenvector maps (MEM: e.g., Legendre and Legendre, 2012), and the proposed method. MEM has several variants, including principal coordinate analysis of neighbor matrices (Borcard and Legendre, 2002; Dray *et al*., 2006) and asymmetric eigenvector maps (Blancheta *et al*., 2008).

Table 5-1 summarizes properties of these three methods. ESF is a topology-based method, which describes spatial connectivity using an adjacency matrix, while MEM and our method are distance-based methods, which describe spatial connectivity using a distance matrix (see also Griffith and Peres-Neto, 2006). Another difference is that ESF and MEM model discretized spatial phenomena, whereas our method models continuous spatial phenomena over an area *D* using observations randomly distributed in *D*. Despite such differences, interpretations of ESF and MEM are strictly equivalent (Dray *et al*., 2006), and our method, which is an extension of ESF, is also an extension of MEM for continuous space. More specifically, the distance matrixes ($\mathbf{MK}_0\mathbf{M}$ and $\mathbf{M}^{+}\mathbf{K}_0^{+}\mathbf{M}^{+}$) essentially are identical. Furthermore, our eigenfunction selection criterion in Case 1 is also identical to the criterion for MEM. Specifically, both our method and MEM select all eigenfunctions representing positive eigenvalues. Meanwhile, our criterion in Case 2 also is similar to the criterion in MEM (see Table 5-1). Such similarities between our method and MEM are natural because both of them are distance-based approaches.

Our assumptions are similar to the standard assumptions in geostatistics that a continuous spatial process is modeled using a distance function. In addition, our model is derived from a geostatistical model. Hence, our method seems important from the perspective of linking discussions of the MC's eigenvector-based approaches and discussions in geostatistics. Actually, our method can be considered as a basis function-based method that has been developed in geostatistics for both dimension and flexible model construction reductions (e.g., Cressie and Johannesson, 2008; Matsuo *et al.*, 2011; Ren and Banerjee, 2013). Our method is distinctive from these methods in that OLS is applicable for the parameter estimation, and, therefore, ours might be useful as a simple method for geostatistical data modeling.

Many studies of ESF discuss estimation problems in the presence of spatial dependence (e.g., Tiefelsdorf and Griffith, 2007; Griffith, 2003; 2006), whereas MEM has been applied mainly for spatial component analysis of ecological data (e.g., Borcard and Legendre, 2002; Peres-Neto *et al.*, 2006; Legendre and Legendre, 2012). Our method may be applicable not only for these purposes, but also for purposes that have been discussed in geostatistics (e.g., spatial interpolations, change of supports, sampling designs).

Table 5-1: Comparison of approaches applying MC-based eigenvectors

Method	Proposed method	ESF	MEM
Classification	Distance-based	Topology-based	Distance-based
Connectivity Matrix	$\mathbf{M}^{+}\mathbf{K}_0{}^{+}\mathbf{M}^{+}(=\mathbf{M}^{+}\mathbf{K}^{+}\mathbf{M}^{+}-\mathbf{M}^{+})$	$\mathbf{MWM}$	$\mathbf{MK}_0\mathbf{M}$
Space	Continuous	Discrete	Discrete
Eigenvector truncation criterion	$\lambda^{+}{}_{l\_0} > 0$ (Case 1) $\lambda_{l\_0} > 0$ (Case 2)	Variable ($\lambda_{l\_0} > 0.25$ is standard)	$\lambda_{l\_0} > 0$
Principal use	N/A	Estimations of spatial models	Spatial component analysis

5.3. An empirical study

5.3.1. Outline

This section utilizes the proposed method to analyze land prices in the Ibaraki prefecture of Japan. The response variable is the logarithm of officially assessed residential land prices in 2009 (sample size: 587; Fig.5-3), which are provided by the Ministry of Land, Infrastructure and Transport (MLIT). Table 5-2 lists the explanatory variables. We apply four types of the proposed method: MC-based approaches whose $\mathbf{M}^+$ equals $\mathbf{I}^+–\mathbf{1}^+\mathbf{1}^{+\prime}/n^+$ and $\mathbf{I}^+–\mathbf{X}^+(\mathbf{X}^{+\prime}\,\mathbf{X}^+)^{-1}\mathbf{X}^{+\prime}$ (E_MC and EX_MC), respectively, and AICc-based approaches whose $\mathbf{M}^+$ equals $\mathbf{I}^+–\mathbf{1}^+\mathbf{1}^{+\prime}/n^+$ and $\mathbf{I}^+–\mathbf{X}^+(\mathbf{X}^{+\prime}\,\mathbf{X}^+)^{-1}\mathbf{X}^{+\prime}$ (E_AICc and EX_AICc), respectively. Fig.5-4 portrays the 1st, 10th, and 100th eigenvectors (Note: they are selected only for the illustrative purpose, and all eigenvectors are considered in the subsequent analyses). Following Tiefelsdorf and Griffith (2007), eigenvectors in E_MC and EX_MC are selected until |z(MC)| of the residuals is less than 0.1.

Results are compared with the standard linear regression model (LM), and the standard geostatistical model (GS), whose model is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}\,, \qquad \boldsymbol{\eta} \sim N(\mathbf{0}, \tau^2\mathbf{K})\,, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})\,. \tag{5-29}$$

where $\mathbf{K}$ is a covariance (kernel) matrix whose elements are given by the exponential function in Eq.(2-9) (Note: Eq.5-29 is identical to Eq.2-19 with its $\mathbf{C}$ is replaced with $\tau^2\mathbf{K} + \sigma^2\mathbf{I}$).

Subsequent results are from the R implementation provided by The Comprehensive R Archive Network (http://cran.r-project.org/), and mappings are from ArcGIS provided by ESRI Inc. (http://www.esri.com/).

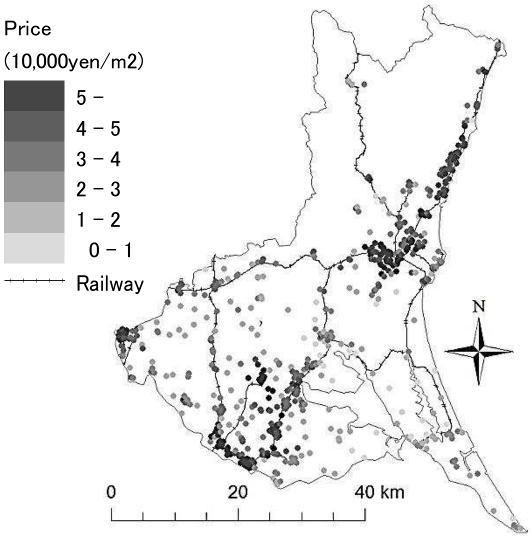
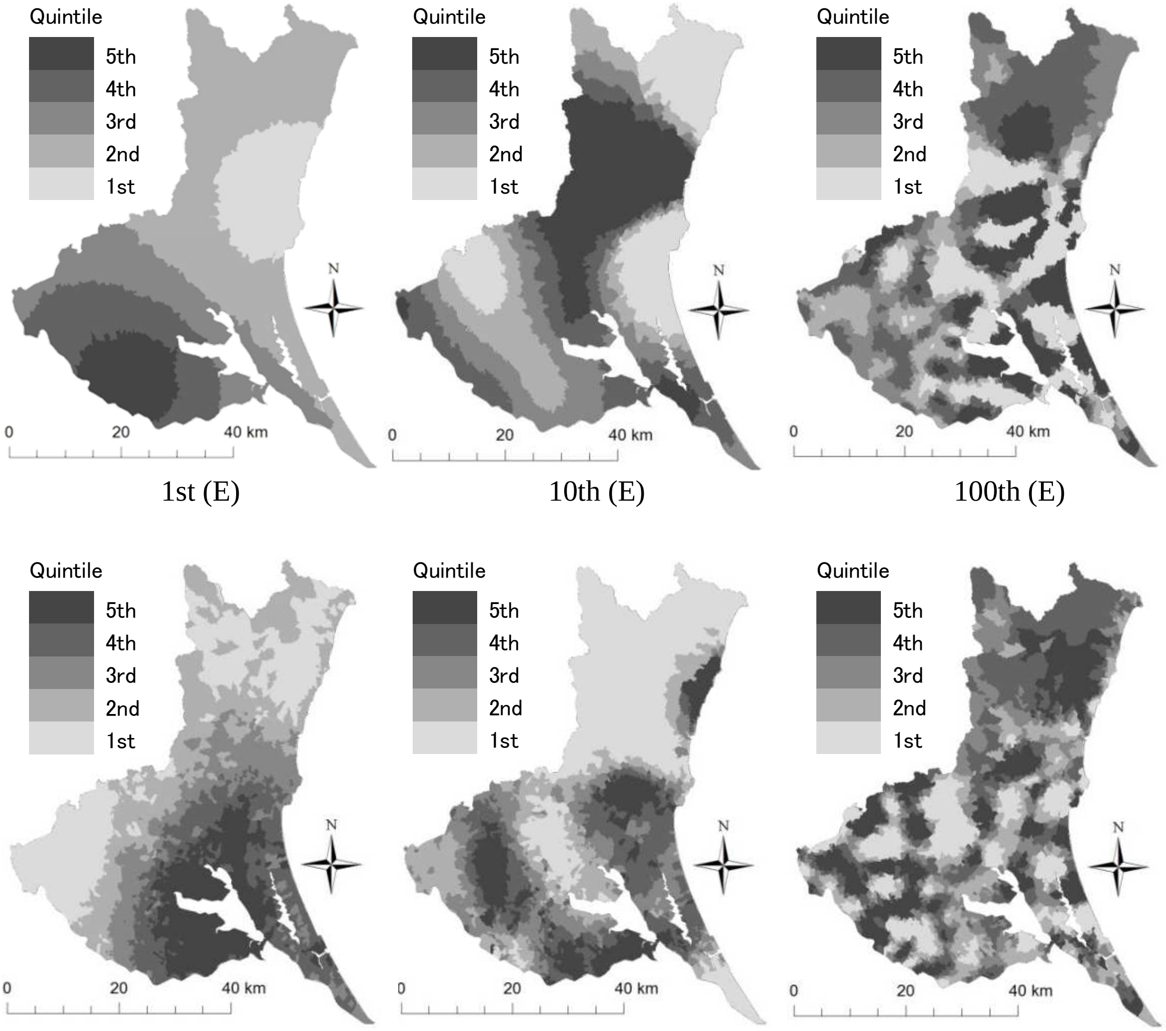


Figure 5-3: Land prices in Ibaraki prefecture

Table 5-2: Explanatory variables

Variables	Description	Unit
Tokyo dist.	Minimum railway distance from the nearest station to Tokyo station	Km
Station	Distance to the nearest station	
Urban	Dummy indicating 1 if a site is in an urbanized area	0 or 1
Agriculture	Area of agricultural land	km^2 par unit area
Forest	Area of forest	
Wasteland	Area of wasteland	
Traffic	Area of trunk transportation land	
Other land	Area of other land (e.g., athletic stadium, port district)	
Golf	Area of golf course	
River	Area of river and lake	
Sea	Area of beach and body of seawater	

[1)] Data source: National Land Numerical Information download service



1st (E) 10th (E) 100th (E)

1st (EX) 10th (EX) 100th (EX)

Figure 5-4: 1st, 10th, and 100th eigenvectors.

Note: Top: the eigenvectors in E_MC and E_AICc; Bottom: those in EX_MC and EX_AICc.

5.3.2. Parameter estimation

Following the discussion in §5.2.5, here, eigenfunctions satisfying λ_{l_0}>0 were selected. Behavior of *z*(*MC*)s during the stepwise selection procedures in E_MC, E_AICc , EX_MC, and EX_AICc are plotted in Fig.5-5. E_MC and EX_MC remove residual spatial dependences effectively using 10/45 and 11/42 eigenfunctions, respectively (see Fig.5-6). Interestingly, these are the selected eigenfunctions even if exhaustive searches are employed. E_AICc and EX_AICc also reduce *z*(*MC*)s substantially, although reductions of their *z*(*MC*)s are slower than those of E_MC and EX_MC

(E_AICc selects 30/45 eigenfunctions, and EX_AICc selects 27/42 eigenfunctions). In short, while both the MC-based approaches and the AICc-based approaches reduce *z*(*MC*) sufficiently, the former is more effective.

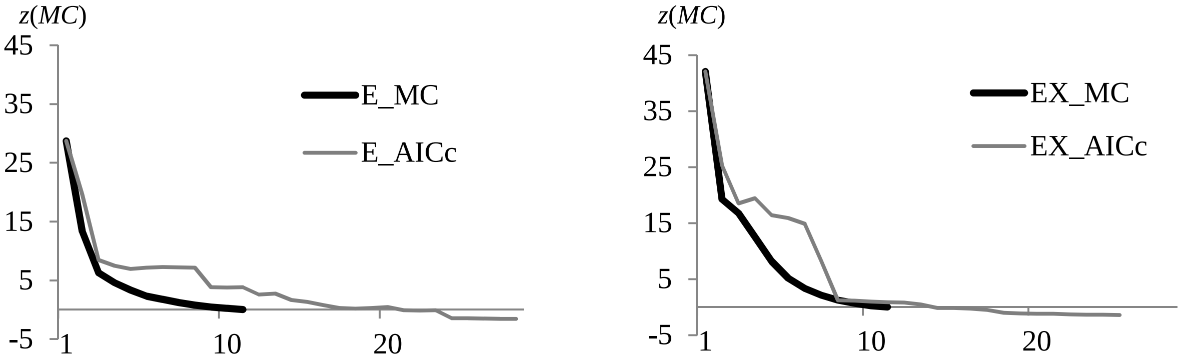


Figure 5-5: Behavior of *z*(*MC*) in variable selections

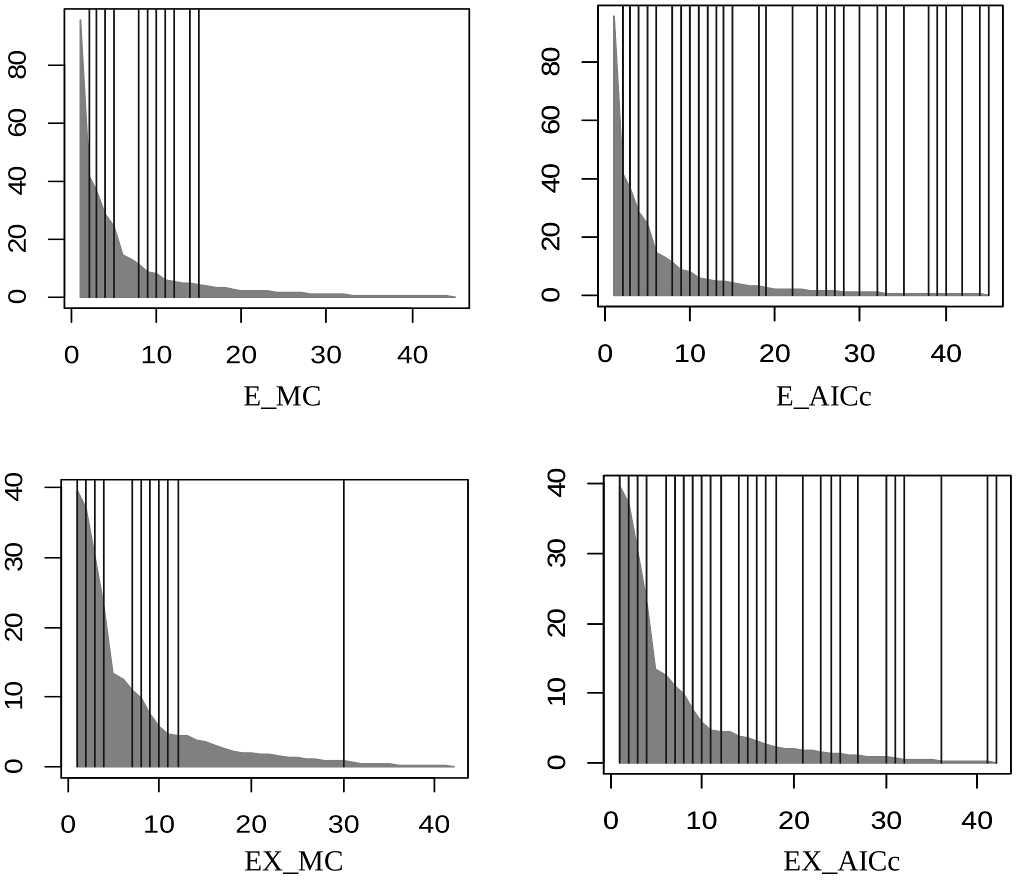


Figure 5-6: Spectrum of eigenvalues (gray) of $\mathbf{MK_0M}$ and selected eigenfunctions (black lines)

Table 5-3 summarizes estimation results for E_MC, E_AICc, EX_MC, EX_AICc, and LM and GS. EX_MC and EX_AICc, whose eigenvectors are uncorrelated with **X**. These implementations do not consider variance inflations due to spatial dependence (they consider variance deflation only), and, consequently, standard errors of their parameters are likely to be underestimated. In theory, standard errors of the coefficients in EX_MC and EX_AICc are always smaller than those in LM, whose underestimation by ignoring spatial dependence has been demonstrated (LeSage and Pace, 2009). Thus, applying EX_MC or EM_AICc for parameter estimation is not necessarily preferred. Our result is counter to those for some studies that suggest removing variance inflation due to spatial dependence prior to estimation (e.g., Paciorek, 2010; Hughes and Haran, 2013).

Estimation results of E_MC and E_AICc, which consider variance inflation due to spatial dependence, are similar to the results of GS, whose effectiveness in parameter estimation has been demonstrated (e.g., Tsutsumi and Seya, 2009). In E_MC and E_AICc, Station (–), Urban dum (+), Agriculture (–), Forest (–), Other land (+), River (–), and Ocean (–) are significant at the 0.05 level, and Traffic (+) in E_MC also is significant at the 0.10 level. These results indicate that land prices are high at sites with substantial urban facilities (Station, Urban dum, Other land, and Traffic), while low at sites with non-urban land uses (Agriculture, Forest, River, and Ocean).

The partial-sill and nugget estimates from GS indicate that the variance of the spatial component is far greater than the variance of the non-spatial component. The range parameter estimate indicates that the distance that spatial dependence spans (effective range) is 19.8 (6.6×3) km. In other words, land prices have small scale spatial variation.

Table 5-3: Parameter estimates

Variables	LM			E-MC			E-AICc		
	Coef.	St.dev.	Signif.	Coef.	St.dev.	Signif.	Coef.	St.dev.	Signif.
Const	10.520	0.073	***	10.320	0.058	***	10.240	0.054	***
Tokyo dist.	-0.001	0.000	**	0.000	0.000		0.000	0.000	
Station	-0.048	0.006	***	-0.030	0.005	***	-0.039	0.006	***
Urban_dum	0.476	0.039	***	0.479	0.032	***	0.532	0.029	***
Agrculture	-0.979	0.078	***	-0.799	0.062	***	-0.639	0.060	***
Forest	-0.473	0.133	***	-0.398	0.108	***	-0.300	0.097	***
Wasteland	-0.961	0.647		-0.299	0.527		-0.493	0.475	
Traffic	2.614	1.307	**	1.744	1.018	*	1.306	0.913	
Otherland	0.729	0.318	**	0.562	0.247	**	0.549	0.225	**
Golf	0.186	0.434		-0.231	0.337		-0.025	0.301	
River	-0.754	0.157	***	-0.534	0.123	***	-0.368	0.111	***
Ocean	-1.066	0.329	***	-0.663	0.262	**	-0.681	0.234	***
nugget									
partial-sill									
range									
z(MC)	42.09			0.029			-1.557		
AICc	454.7			157.9			25.3		

Variables	EX-MC			EX-AICc			GS		
	Coef.	St.dev.	Signif.	Coef.	St.dev.	Signif.	Coef.	St.dev.	Signif.
Const	10.520	0.054	***	10.520	0.051	***	10.070	0.182	***
Tokyo dist.	-0.001	0.000	***	-0.001	0.000	***	0.000	0.001	
Station	-0.048	0.004	***	-0.048	0.004	***	-0.059	0.009	***
Urban_dum	0.476	0.029	***	0.476	0.027	***	0.587	0.030	***
Agrculture	-0.979	0.057	***	-0.979	0.054	***	-0.366	0.056	***
Forest	-0.473	0.097	***	-0.473	0.092	***	-0.232	0.094	**
Wasteland	-0.961	0.473	**	-0.961	0.447	**	-0.076	0.472	
Traffic	2.614	0.957	***	2.614	0.903	***	0.727	0.759	
Otherland	0.729	0.233	***	0.729	0.220	***	0.366	0.201	*
Golf	0.186	0.318		0.186	0.300		-0.062	0.276	
River	-0.754	0.115	***	-0.754	0.108	***	-0.365	0.099	***
Ocean	-1.066	0.241	***	-1.066	0.227	***	-0.471	0.208	**
nugget							0.018		
partial-sill							0.117		
range							6.570		
z(MC)	-0.017			-1.418			[2]		
AICc	100.6			50.8			6.6		

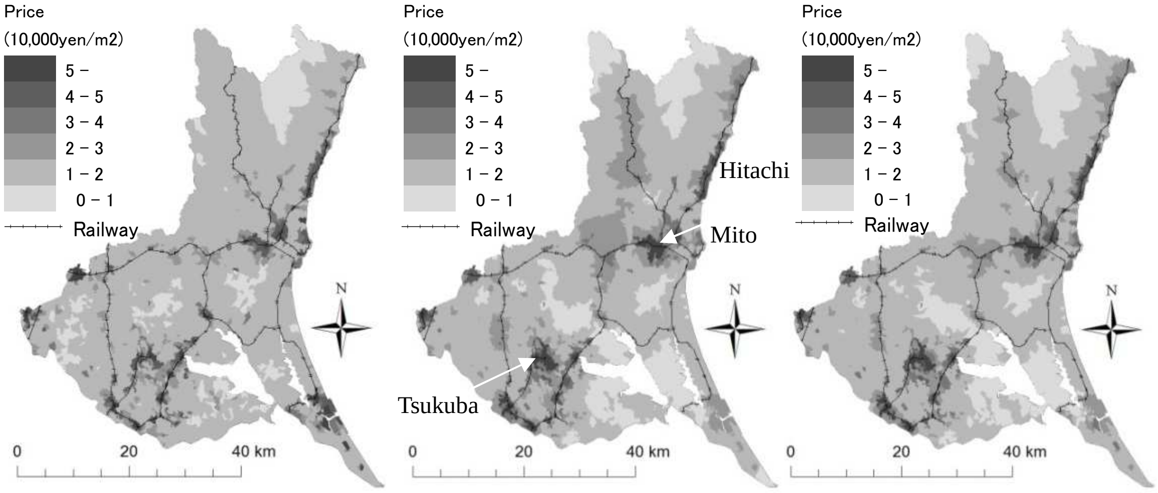
[1] *, **, *** denote significant levels (10%, 5% and 1%)

[2] Because residuals of GM are always 0, z(MC) cannot be used for GM

5.3.3. Exploratory spatial data analysis

On the basis of the discussion in §5.2.5.1, no eigenfunctions are omitted before the stepwise selection procedure begins. EX_AICc is used here because: (i) The efficient eigenfunction selection algorithm (see §5.2.5.2) is applicable; (ii) E_MC and EX_MC cannot capture small scale variations (no eigenfunction whose $\lambda_{l\_0} \leq 0$ is selected even if exhaustive searches are performed); and, (iii) when performing ESDA, which describes spatial patterns in data, variance inflation between **X** and **E** in E_AICc is problematic.

Land prices in each geometric center of the minor municipal units (number of units: 3,175) are interpolated using LM, EX_AICc, and GS (Fig.5-7). Although we use each minor municipal unit for mapping, because these units are at a fine spatial resolution, impacts of the shapes or sizes of these units on the resulting maps are sufficiently small. The results of EX-AICc and GS are quite similar, and both indicate high values nearby Mito, the capital of the Ibaraki prefecture, and Tsukuba and Hitachi, primal cities in this region. Such a feature is less clear in the LM result. Also, the distributions of low price areas in the LM result are quite different from those in the EX_AICc and GS results.



LM EX_AICc GS

Figure 5-7: Interpolatoin results

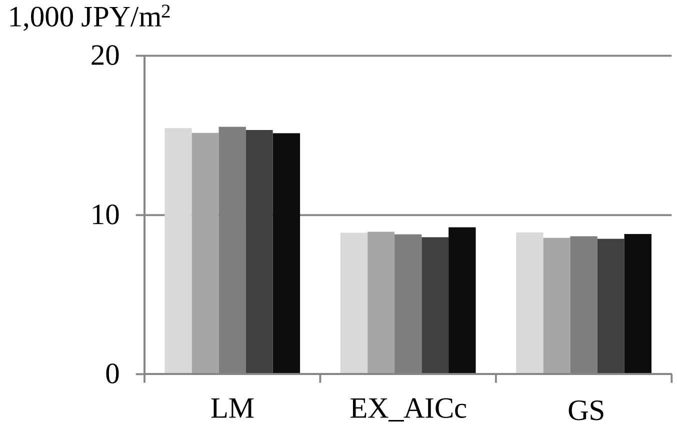


Figure 5-8: RMSE of the methods obtained using five 5-fold-cross validations

A 5-fold cross-validation is iterated 5 times to compare the accuracy of each result. This 5-fold cross-validation procedure is as follows: (i) Samples are randomly partitioned into 5 equal-size subsamples; (ii) A model is identified using 80% of the subsamples; (iii) Accuracy of a model is measured by fitting it to the remaining 20% of the subsamples; and, (iv) (ii) and (iii) are performed for all 5 cases. Root mean square error (RMSE) is used to evaluate model accuracy. Fig.5-8 summarizes results of the cross-validations. RMSE for EX_AICc (average: 8,889) and GS (average: 8,686) are about half of the RMSE for LM (average: 15,328). Thus, the importance of considering spatial dependence is confirmed. In addition, the accuracy of our OLS-based simple method is comparable with that of GS.

One of the advantages of our method is that the estimated spatial components are decomposable. Here, the extracted spatial component–that is, the linear combination of all significant eigenfunctions–is decomposed into a linear combination of eigenfunctions satisfying (s1) $\lambda_{l\_0}/\lambda_{1\_0} \geq 0.5$, (s2) $0.5 > \lambda_{l\_0}/\lambda_{1\_0} \geq 0.25$, (s3) $0.25 > \lambda_{l\_0}/\lambda_{1\_0} \geq 0$, and (s4) $0 > \lambda_{l\_0}/\lambda_{1\_0}$, respectively (Fig.5-9). Roughly speaking, s1, s2, s3, and s4 describe components whose spatial scales are coarse, mid-coarse, medium, and fine, respectively. The coarser component, s1, indicates high values in the southwestern part of the landscape, which is nearby Tokyo, and the northwestern area, which is nearby Mito or Hitachi. Thus, s1 might indicate significant impacts of these primal cities. The mid-coarser component, s2, is slight, and any prominent spatial pattern does not seem to materialized at this scale. The medium scale component, s3, indicates high values around some cities, including Mito, Tsukuba, Toride, Koga, and Inashiki. These places are well-developed compared to their surroundings, and, accordingly, s3 can be leveled as the local spatial pattern induced by these cities. Finally, land prices are strongly affected by the finer component, s4. It might be associated with local components that we cannot consider, such as living environment and geographical features.

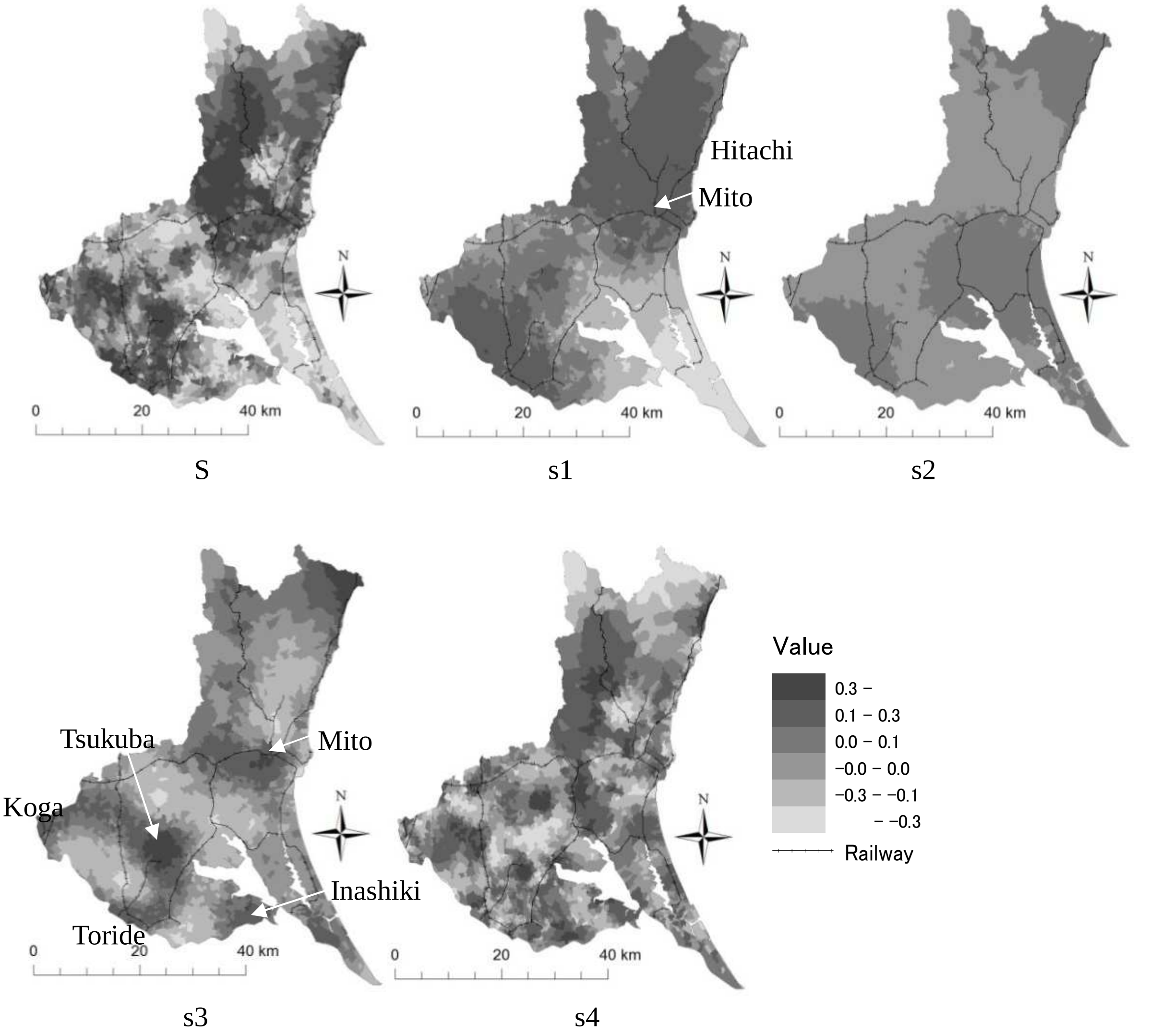


Figure 5-9: Plots of linear combinations of the eigenvectors selected from EX_AICc

S shows the linear combination of all selected eigenvectors, and s1, s2, s3, and s4 show linear combinatons of eigenvectors satisfying $\lambda_{l\_0}/\lambda_{1\_0} \geq 0.5$, $0.5 > \lambda_{l\_0}/\lambda_{1\_0} \geq 0.25$, $0.25 > \lambda_{l\_0}/\lambda_{1\_0} \geq 0$, and $0 > \lambda_{l\_0}/\lambda_{1\_0}$, respectively.

The extracted components constitute variance partitioning (see Legendre and Legendre, 2012, for more details). We divide the variance of land prices into non-spatial components (X), spatial components at each scale (s1, s2, s3, s4), and the disturbance (ε). Because these components are uncorrelated, and do not overlap, this partitioning differs from the conventional variance partitioning whose results are summarized using a 2-dimentional graph. Our result can be summarized using a 1-dimentional graph. The result shown in Fig.5-10 suggests that the spatial component, s1 + s2 + s3 + s4, explains 32.7% of the variation, with 8.7% being attributed to coarse component s1, 0.3% being attributed to mid-coarse components s2, 9.9% being attributed to medium scale components s3, and 13.8% being attributed to finer scale component s4. The significance of the fine scale variation is consistent with the small range parameter obtained with GS. The fine scale component is ignored if the eigenfunctions are selected among functions satisfying λ_{l_0}> 0, as both E_MC and EX_MC assume.

One interesting result is that s1 and s3 are significant, whereas s2 is not. For comparative purposes, EX_AICc is fitted to the land prices in the 23 wards of Tokyo, and the variance partitioning is performed (Fig.5-10). The result differs from that for the Ibaraki prefecture; the most significant component is s1, followed in order by s2, s3, and s4. Thus, the absence of s2 is a feature specific to the Ibaraki prefecture.

In summary, the proposed method is useful for both parameter estimation accounting for spatial dependence, and ESDA.

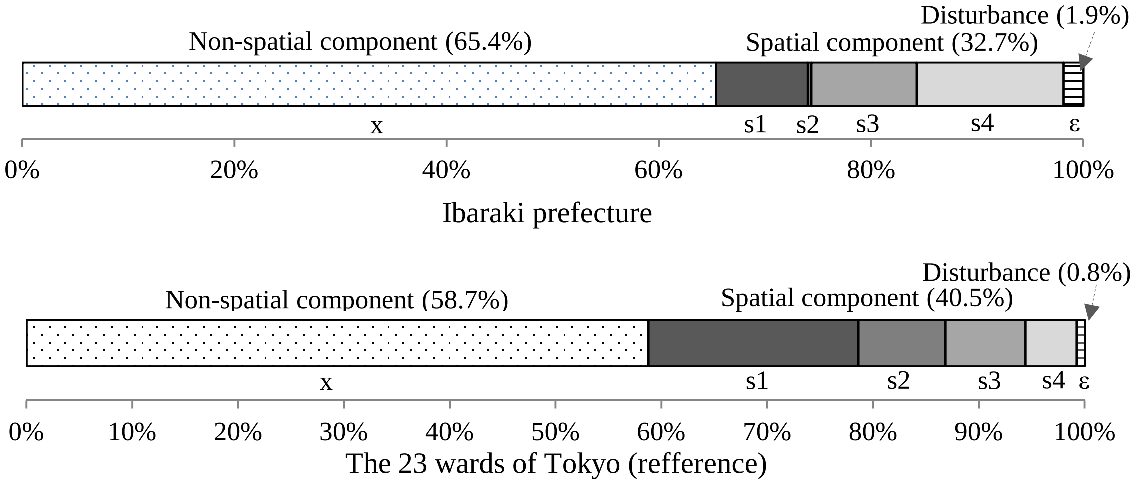


Figure 5-10: Result of variance partitioning

x: Non-spatial components; s1, s2, s3, s4: spatial components whose λ_{l_0} /λ_{1_0} are within 1.00–0.50, 0.50–0.25, 0.25–0.00, and below, 0.00 respectively; ε: disturbance.

5.4. A spatiotemporal extension

5.4.1. Introduction

Spatiotemporal statistical models, which have been discussed extensively in geostatistics (e.g., Cressie, 1993; Cressie and Wikle, 2011), are classified into dynamic and non-dynamic models. The advantage of dynamic models is that they can model causation, they are generally computationally efficient, and the validity of their models (e.g., positive definiteness of variance-covariance matrix) is easily proved (Cressie et al., 2010). As a result, dynamic modeling has recently begun attracting increasing attention. On the other hand, non-dynamic models are still important as descriptive or exploratory tools (Cressie and Wikle, 2011). This study focuses on the latter group of models.

Standard non-dynamic geostatistics describes space-time processes by parameterizing the covariance using a function of distance and time lag. The effectiveness of geostatistical models has been demonstrated in spatiotemporal interpolation studies. However, they are not necessarily flexible as descriptive models. For instance, they cannot reveal space-time components in data, such as global spatial components, time-invariant spatial components, and so on.

The empirical orthogonal function (EOF) analysis (e.g., Wilks, 2006) is another geostatistical approach that extracts such space-time components using an eigen-decomposition, and studies have demonstrated its effectiveness for space-time descriptive analyses, including multiscale spatial component analyses, visualization, and multivariate analyses (see Cressie and Wikle, 2011). However, this approach applies only to discrete spatial data (lattice data), whereas a descriptive analysis is particularly important for continuous spatial data (geo-referenced data), in which complete observations are generally not possible.

As same as the EOF analysis, ESF extracts spatial components in spatial data, and its effectiveness for spatial descriptive analysis has been clarified (e.g., Griffith and Peres-Neto, 2006; Legendre and Legendre, 2012). However, ESF is also not for continuous spatial data but for discrete spatial data.

On the contrary, my extended ESF, which I call dESF (distance-based ESF) hereafter, is for continuous spatial data. Hence, its spatiotemporal extension might bring a sophisticated space-time descriptive model. Thus, this section extends dESF for spatiotemporal data.

5.4.2. Model

This section extends the distance-based ESF for spatiotemporal descriptive analyses. §5.4.2.1 discusses standard spatiotemporal geostatistics, and §5.4.2.2 extends the dESF for spatiotemporal data based on the discussion in §5.4.2.1. This study proposes a model for longitudinal data with a sample size of *NT*, where *N* denotes the number of observation sites on $s_I \in D \subset \Re^2$, and *T* denotes the number of observation times on $t\in$ {1,... *T* }, in which the intervals are not necessarily uniform.

5.4.2.1. Spatiotemporal geostatistical model

The standard space-time geostatistical model is defined as follows (e.g., Gneiting and Guttorp, 2010):

$$y_{i,t} = \mu_{i,t} + \eta_{i,t} + \varepsilon_{i,t}\,, \tag{5-30}$$

where $y(s_i, t)$ are the response variables, $\mu(s_i, t)$ is a deterministic non-spatial component, $\eta(s_i, t)$ is a stochastic spatial component, and $\varepsilon(s_i, t)$~$N(0, \sigma^2)$.

The term $\eta\,(s_i, t)$ is modeled by parameterizing its covariance using a function of distance and time lag. For example, the product-sum model (De Cesare et al., 2001), which is defined by Eq.(5-31), is one of the most common functions:

$$\mathrm{cov}[\eta_{i,t}, \eta_{i,t'}] = c_s(s_i, s_j) + c_t(t, t') + c_s(s_i, s_j)c_t(t, t')\,, \tag{5-31}$$

where $c_s(s_i,\ s_j)$ and $c_t(t, t')$ are functions describing spatial dependency and temporal dependency, respectively.

When the model is applied to longitudinal data, Eq.(5-31) can be expressed using a matrix notation as

$$\mathbf{C}_{st} = \mathbf{C}_s \otimes \mathbf{I}_t + \mathbf{I}_s \otimes \mathbf{C}_t + \mathbf{C}_s \otimes \mathbf{C}_t\,, \tag{5-32}$$

where $\mathbf{C}_s$ $(N \times N)$ and $\mathbf{C}_t$ $(T \times T)$ are spatial and temporal covariance matrices, respectively, $\mathbf{I}_s$ $(N \times N)$ and $\mathbf{I}_t$ $(T \times T)$ are identity matrices, and $\otimes$ is the Kronecker product operator

5.4.2.2. ESF-based spatiotemporal eigenfunctions

First, let us decompose $\mathbf{C}_s$ and $\mathbf{C}_t$ into $\mathbf{E}_s\mathbf{\Lambda}_s\mathbf{E}_s'$ and $\mathbf{E}_t\mathbf{\Lambda}_t\mathbf{E}_t'$, respectively, using eigen-decompositions. Then, Eq.(11) is expanded as

$$\mathbf{C}_{st} = (\mathbf{E}_s\mathbf{\Lambda}_s\mathbf{E}_s') \otimes \mathbf{I}_t + \mathbf{I}_s \otimes (\mathbf{E}_t\mathbf{\Lambda}_t\mathbf{E}_t') + (\mathbf{E}_s \otimes \mathbf{E}_t)(\mathbf{\Lambda}_s \otimes \mathbf{\Lambda}_t)(\mathbf{E}_s \otimes \mathbf{E}_t)'. \qquad (5\text{-}33)$$

Eq.(5-33) models the spatial component, temporal component, and spatiotemporal component by weighting their corresponding eigenvectors, $\mathbf{E}_s$, $\mathbf{E}_t$, and $\mathbf{E}_s \otimes \mathbf{E}_t$, using their own eigenvalues (i.e., the diagonals of $\mathbf{\Lambda}_s$, $\mathbf{\Lambda}_t$, and $\mathbf{\Lambda}_s \otimes \mathbf{\Lambda}_t$), respectively.

We apply the eigenvectors of **MKM** and $\mathbf{M}_t\mathbf{K}_t\mathbf{M}_t$ to $\mathbf{E}_s$ and $\mathbf{E}_t$, respectively, where $\mathbf{K}_t$ is a matrix describing temporal connectivity, and $\mathbf{M} = \mathbf{I}_t - \mathbf{1}_t\mathbf{1}_t'/T$. The elements in $\mathbf{K}_t$ are given by $k(t, t') = \exp(-|t-t'|/r_t)$, where r_t is the longest time interval among the observations (see Dray et al., 2006). Since the eigenvectors of **MKM** (or $\mathbf{M}_t\mathbf{K}_t\mathbf{M}_t$) and **K** (or $\mathbf{K}_t$) are essentially identical,[1] this assumption implies that we replace $\mathbf{C}_s$ and $\mathbf{C}_t$ with **K**, and $\mathbf{K}_t$, respectively.

The first, second, and third terms in Eq.(5-33) are described by $\mathbf{E}_s$, $\mathbf{E}_t$, and $\mathbf{E}_s \otimes \mathbf{E}_t$, respectively, and the elements explained by these terms can be summarized as $\mathbf{E}_{st} = \{\mathbf{E}_s \otimes \mathbf{1}_t, \mathbf{1}_s \otimes \mathbf{E}_t, \mathbf{E}_s \otimes \mathbf{E}_t\}$, where $\mathbf{1}_s$ is a vector of ones. We can easily show that the vectors in $\mathbf{E}_{st}$ are mutually orthogonal, i.e., $\mathbf{E}'_{st}\mathbf{E}_{st} = \mathbf{I}$. Besides, because the means of the eigenvectors in $\mathbf{E}_s$ and $\mathbf{E}_t$ are uniformly zeros, the means of the vectors in $\mathbf{E}_{st}$ are also zeros. Consequently, the vectors in $\mathbf{E}_{st}$ are both orthogonal and uncorrelated (see also, Griffith, 2003). Thus, $\mathbf{E}_{st}$ furnishes distinct (i.e., orthogonal and uncorrelated) map pattern descriptions of latent space-time dependence.

$\mathbf{E}_s \otimes \mathbf{1}_t$ in $\mathbf{E}_{st}$ explains spatial components in each time, which is explained by MC^+ (see §5.2.2). As with the standard ESF, the vectors in $\mathbf{E}_s \otimes \mathbf{1}_t$ corresponding to large eigenvalues of $\mathbf{M}_s\mathbf{K}_s\mathbf{M}_s$ (diagonal elements in $\mathbf{\Lambda}_s$) explain global scale spatial components, and the vectors corresponding to small eigenvalues explain local components. Similarly, $\mathbf{1}_s \otimes \mathbf{E}_t$ explains temporal components, which are described by Eq.(5-34):

$$MC_t^+ = \frac{1}{\sigma_{z_t}^2} \int_{j \neq i} \int_i w_t(t,t')\tilde{z}_t(t)\tilde{z}_t(t')dtdt', \qquad (5\text{-}34)$$

$$w_t(s_i, s_j) = \frac{k(t,t')}{\int_{j \neq i} \int_i k(t_i, t')dtdt'},$$

[1] The eigenvectors of **MKM** are the eigenvectors of **K** after an axis rotation using **M**.

where $z_t(t)$ denotes the variables defined on $\Re^1$, $\tilde{z}_t(t)$ denotes the variables that center $z_t(t)$, and $\sigma^2_{z_t}$ is the variance of $z_t(t)$. The eigenvectors in $\mathbf{I}_{s\,.}$ $\mathbf{E}_t$ describe those temporal components with scales that are signified by the eigenvalues, or MC_t^+, of $\mathbf{M}_t\mathbf{K}_t\mathbf{M}_t$ (i.e., the diagonals of $\mathbf{\Lambda}_t$). Finally, $\mathbf{E}_s \otimes \mathbf{E}_t$ describes the spatiotemporal components that have scales signified by the eigenvalues of $\mathbf{MKM} \otimes \mathbf{M}_t\mathbf{K}_t\mathbf{M}_t$, which are equal to the diagonals of $\mathbf{\Lambda}_s \otimes \mathbf{\Lambda}_t$ (see Schott, 2005).

5.4.2.3. ESF-based spatiotemporal model

This study proposes the following model

$$\mathbf{y} = \alpha\mathbf{1} + (\mathbf{E}_s \otimes \mathbf{1}_t)^c \boldsymbol{\gamma}_s + (\mathbf{1}_s \otimes \mathbf{E}_t)^c \boldsymbol{\gamma}_t + (\mathbf{E}_s \otimes \mathbf{E}_t)^c \boldsymbol{\gamma}_{st} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \qquad (5\text{-}35)$$

where “ ^{c} ” represents the complementary set, and $\boldsymbol{\gamma}_s$, $\boldsymbol{\gamma}_t$, and $\boldsymbol{\gamma}_{st}$ are parameter vectors. The parameters are estimated by the following steps. First, assume that all vectors in $\mathbf{E}_s \otimes \mathbf{1}_t$,, $\mathbf{1}_s \otimes \mathbf{E}_t$, and $\mathbf{E}_s \otimes \mathbf{E}_t$ corresponding to non-zero eigenvalues are candidates to be entered into $(\mathbf{E}_s \otimes \mathbf{1}_t,)^c$, $(\mathbf{1}_s \otimes \mathbf{E}_t)^c$, and $(\mathbf{E}_s \otimes \mathbf{E}_t)^c$. Second, substitute the candidate vectors into Eq.(5-35) sequentially in decreasing order of the absolute value of the correlation coefficients between $\mathbf{y}$ and each of the vectors, until the AICc is minimized. The OLS technique is used for the AICc calculations. Since the vectors are mutually orthogonal, the OLS estimates of $\boldsymbol{\gamma}_s$, $\boldsymbol{\gamma}_t$, and $\boldsymbol{\gamma}_{st}$ results in $(\mathbf{E}_s \otimes \mathbf{1}_t,)^{c\,\prime}\mathbf{y}$, $(\mathbf{1}_s \otimes \mathbf{E}_t)^{c\,\prime}\mathbf{y}$, and $(\mathbf{E}_s \otimes \mathbf{E}_t)^{c\,\prime}\mathbf{y}$, respectively.

The model can be defined on unobserved sites too. Suppose that $\mathbf{y}_0$ is a vector of unobserved response variables at time points $t \in \{1,\ldots T\}$. Then, $\mathbf{y}_0$ is modeled as

$$\mathbf{y}_0 = \alpha\mathbf{1} + (\mathbf{E}_{s0} \otimes \mathbf{1}_t)^c \boldsymbol{\gamma}_s + (\mathbf{1}_{s0} \otimes \mathbf{E}_t)^c \boldsymbol{\gamma}_t + (\mathbf{E}_{s0} \otimes \mathbf{E}_t)^c \boldsymbol{\gamma}_{st} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \qquad (5\text{-}36)$$

where $\mathbf{E}_{s0}$, which is given by Eq.(5-37), is the eigenvector matrix approximated using the Nyström extension:

$$\mathbf{E}_{s0} = (\mathbf{K}_0 - \mathbf{1}_0\mathbf{1}'\mathbf{K} / N)(\mathbf{I} - \mathbf{1}\mathbf{1}' / N)\mathbf{E}_s(\mathbf{\Lambda}_s + \mathbf{I}(\mathbf{\Lambda}_s))^{-1}, \qquad (5\text{-}37)$$

where $\mathbf{I}(\mathbf{\Lambda}_s)$ is a diagonal matrix with an I-th element of 0 if the I-th diagonal of $\mathbf{\Lambda}_s$ is zero, and 1 otherwise. Similarly, when the time points of $\mathbf{y}_0$ are not consistent with $t \in \{1,\ldots T\}$, Eq.(5-36) is modified by replacing $\mathbf{E}_t$ with $\mathbf{E}_{t0}$, which is defined as

$$\mathbf{E}_{t0} = (\mathbf{K}_0 - \mathbf{1}_{0t}\mathbf{1}_t'\mathbf{K}_t / T)(\mathbf{I}_t - \mathbf{1}_t\mathbf{1}_t' / T)\mathbf{E}_t(\mathbf{\Lambda}_t + \mathbf{I}(\mathbf{\Lambda}_t))^{-1}, \tag{5-38}$$

where $\mathbf{I}(\mathbf{\Lambda}_t)$ is defined in the same way as $\mathbf{I}(\mathbf{\Lambda}_s)$.

The proposed model seems useful for descriptive analyses. For instance, response variables on arbitrary sites and times can be interpolated using Eq.(5-36). Furthermore, this model measures the significances of spatial, temporal, and spatiotemporal components in all space/temporal scales using $\boldsymbol{\gamma}_s$, $\boldsymbol{\gamma}_t$, and $\boldsymbol{\gamma}_{st}$, respectively. These extracted components can be visualized by mapping the estimated linear combinations (e.g., estimate of $(\mathbf{E}_s \otimes \mathbf{1}_{t,})^c \boldsymbol{\gamma}_s$).

Simplicity is also an advantage of this approach. As discussed previously, this method applies the OLS-based simple calculation procedure. In addition, in contrast to the spatiotemporal geostatistical model, which requires an inversion of the spatiotemporal covariance matrix (NT × NT), the proposed method does not explicitly manipulate such a large matrix. Instead, the proposed model imposes eigen-decompositions of $\mathbf{K}$ (N × N) and $\mathbf{K}_t$ (T × T). Hence, as long as neither N nor T is too large, this method is computationally efficient. For example, the proposed method might be suitable for analyzing data with a sample size of 1,000,000, where N = 1,000 and T = 1,000. Note that the assumed eigenvector selection procedure also makes my method computationally efficient.

5.4.3. An empirical study

5.4.3.1. Outline

This sub-section analyzes residential land prices, as officially assessed between 1995 and 2006 in Tokyo, Japan (source: Ministry of Land, Infrastructure, and Transport). Since my method assumes longitudinal data, I use samples at 2,010 sites that had land prices assessed during the target period. The resulting sample size is 24,120. Table 1 summarizes the descriptive statistics of the land prices in each year. The table suggests that the land prices are, on average, decreasing over time. The land prices in 2000 are plotted in Fig.1. The eastern area showing high land prices is the central Tokyo area.

I first apply the standard linear regression model (LM). The response variables are the log-transformed land prices. The explanatory variables are constant (Const), the distance to the nearest railway station (Station), the railway network distance from the nearest station to Tokyo station (Tokyo), and the area of each land use type in 1 km × 1 km grids, including the sample sites (Paddy, Agriculture, Forest, Wasteland, Traffic, Other land, River, Golf) (see Table 5-5). To consider the temporal variation of the regression coefficients, the LMs are fitted in each year independently.

Then, the residuals of the LMs are fitted to the distance-based ESF (dESF) and the spatiotemporal geostatistical model (GS), given as (see Eq.5-32):

$$\mathbf{y} = \alpha\mathbf{1} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\eta} \sim N(\mathbf{0}, \mathbf{C}_s \otimes \mathbf{I}_t + \mathbf{I}_s \otimes \mathbf{C}_t + \mathbf{C}_s \otimes \mathbf{C}_t), \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \qquad (5\text{-}39)$$

where the elements in $\mathbf{C}_s$ and $\mathbf{C}_t$ are given using the exponential model, as with the dESF (i.e., exp(–$d(s_i,\ s_j)/r$) and exp(–|t–t'|/r_t), respectively). Note that, since temporal variations disappear after applying the LMs in each year independently, the dESF estimation becomes strictly identical to the estimation result of the dESF with no pure temporal components:

$$\mathbf{y} = \alpha\mathbf{1} + (\mathbf{E}_s \otimes \mathbf{1}_t)^c \boldsymbol{\gamma}_s + (\mathbf{E}_s \otimes \mathbf{E}_t)^c \boldsymbol{\gamma}_{st} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}). \qquad (5\text{-}40)$$

Therefore, $\boldsymbol{\gamma}_t$, shall not be discussed further in this paper.

The subsequent results are from the R implementation provided by The Comprehensive R Archive Network (http://www.r-project.org/index.html), and mappings are from ArcGIS, provided by ESRI Inc. (http://www.esri.com/).

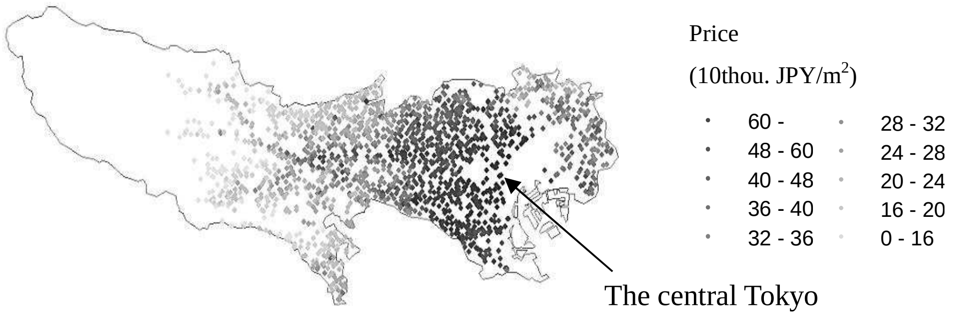


Figure 5-11: Land prices in 2000

Table 5-4: Summary statistics of the land prices (10 thou. JPY/m^2)

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Mean	59.4	52.7	48.7	46.7	43.6	40.9	39.1	37.6	36.3	35.5	35.2	36.0
Median	44.3	42.4	41.0	39.9	37.3	35.2	33.8	32.4	31.3	30.5	30.3	30.5
Std.dev.	64.7	47.2	37.5	34.9	32.2	30.4	29.3	28.3	27.7	27.6	28.1	31.0
Min.	8.05	7.90	7.82	7.60	6.80	6.00	5.70	5.10	4.75	4.30	4.05	3.95
Max	830	615	493	490	470	468	450	446	460	488	514	650

Table 5-5: Explanatory variables

Variables	Description	Unit
Tokyo dist.	Minimum railway distance from the nearest station to Tokyo station	Km
Station	Distance to the nearest station	
Urban	Dummy indicating 1 if a site is in an urbanized area	0 or 1
Agriculture	Area of agricultural land	km^2 par unit area
Forest	Area of forest	
Wasteland	Area of wasteland	
Traffic	Area of trunk transportation land	
Other land	Area of other land (e.g., athletic stadium, port district)	
Golf	Area of golf course	
River	Area of river and lake	
Sea	Area of beach and body of seawater	

Data source: National Land Numerical Information download service

5.4.3.2. Parameter estimation

In each year, the coefficients of Const and Traffic are positive, and those of Tokyo, Station, Paddy, Agriculture, Forest, Wasteland, Other land, River, and Golf are negative. Roughly speaking, the results indicate that adjacency to transportation facilities (Tokyo, Station, and Traffic) inflates land prices, whereas non-urban land uses (Paddy, Agriculture, Forest, Wasteland, River, and Golf) deflate prices. These results are intuitively reasonable.

Transition of the regression coefficient estimates are plotted in Fig.5-12. Here, the estimates in each year are standardized by dividing them by their estimates in 1995 (i.e., the values β_t/β_{1995} are plotted). Since the signs of all of coefficients are unchanged over time, $\beta_t/\beta_{1995} > 1.0$ implies an increase of β_t, and $\beta_t/\beta_{1995} < 1.0$ implies a decrease. Fig.5-12 shows that both the positive influences of the transportation facilities (Tokyo, Station, and Traffic: solid lines) and the negative influences of the major non-urban land uses (Forest and Wasteland: dashed lines) increase gradually. This may indicate that the gap between land prices in urban areas with many transportation facilities and the prices in non-urban areas has gradually increased.

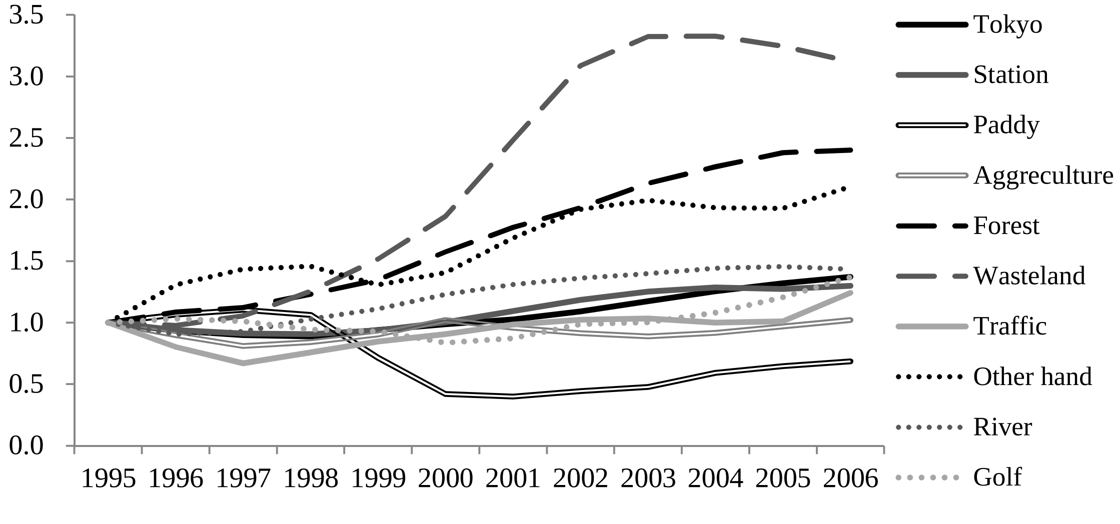


Figure 5-12: Transition of the regression coefficients

5.4.3.3. Interpolation

The land prices in the geometric centers of the minor municipal units in each year (66,060 points = 5,505 geometric centers × 12 years) are interpolated using the LM and dESF. The interpolation results in 1995, 2000, and 2005 are displayed in Fig.5-13. While the results of the LM and dESF are visually quite different, the result of the dESF seems better. For instance, the dESF succeeds in capturing the high land prices in central Tokyo and other major cities, including Kichijoji and Denenchofu.

To compare the accuracy of the LM and dESF, a five-fold cross-validation is iterated five times. This five-fold cross-validation procedure is as follows: (i) Sample sites are randomly divided into five sub-sample sites; (ii) Models are estimated using the 4/5 sub-samples observed on the 4/5 sites; (iii) The remaining 1/5 sub-sample values are interpolated using the estimated models; (iv) The interpolation accuracies are evaluated; and, (v) Steps (ii), (iii), and (iv) are performed for all five cases. The root mean square error (RMSE: Eq.21) is used to evaluate the model accuracy. Fig.5-14 (a) summarizes the resulting RMSEs. The average RMSE of the dESF is 0.219, and the average of the LM is 0.289. Thus, the interpolation accuracy of the dESF is better than that of the LM.

I then compare the accuracies of the dESF and GS. Note that the GS is not available for all the samples because of its computational complexity. Hence, the 24,120 samples are divided randomly into five 4,824 sub-samples, and five-fold cross validations are applied for each of the sub-samples. The results are summarized in Fig.5-14 (b). Interestingly, the RMSEs of the dESF (average: 0.257) and GS (average: 0.257) are almost the same in all five cases. This is because our method is essentially identical to that of the GS, as discussed in §5.4.2.

In contrast, the effectiveness of these methods is quite different from the viewpoint of computational cost. GS requires 32.53 seconds to interpolate 965 land prices using 3,859 samples. The dESF requires only 1.69 seconds to perform the same calculation. In both cases, the computer was a 64-bit laptop with 4.0 GB RAM. Furthermore, even in the original problem of interpolating 66,060 land prices using the 24,120 samples, the dESF required only 36.36 seconds.

In summary, the dESF is as accurate as the GS, and computationally more efficient.

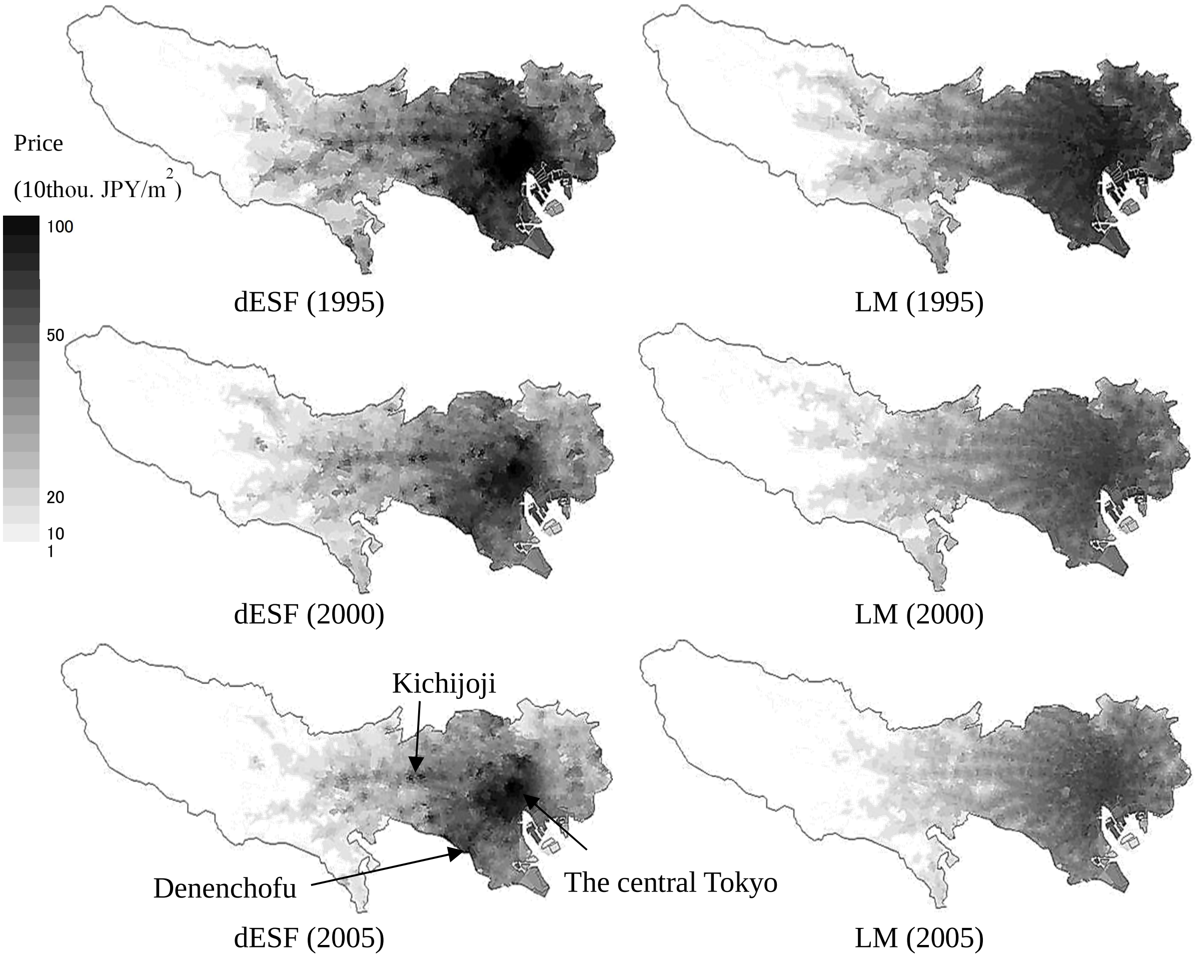


Figure 5-13: Interpolation results

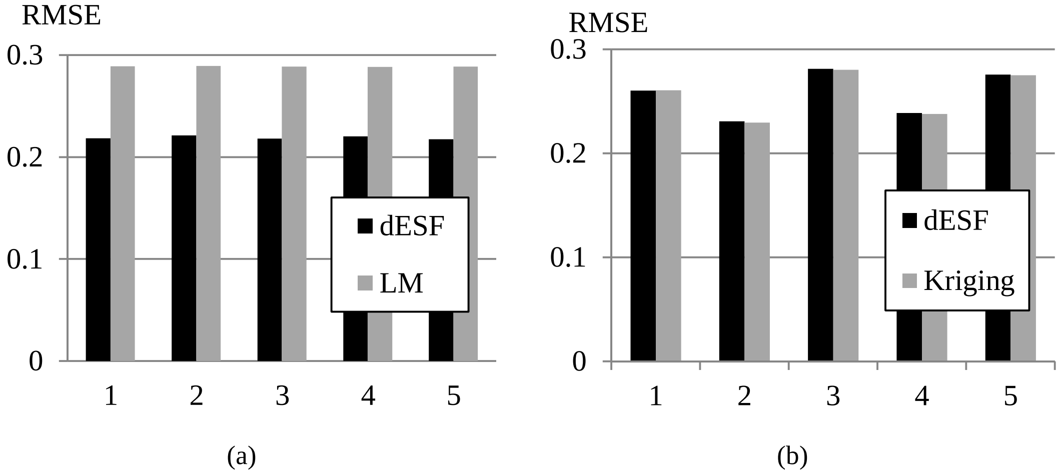


Figure 5-14: Comparison of the interpolation accuracies

Note: (a) RMSEs of LM and dESF given by the 5-fold cross-validation using the full samples. The cross-validation is conducted 5 times. (b) RMSEs of dESF and GS obtained by the 5-fold cross-validation using 1/5 sub-samples, which are obtained by dividing the full samples randomly. The RMSEs are calculated in each sab-sample.

5.4.3.4. Spatial component analysis

The spatiotemporal components estimated by the dESF (i.e., $(\mathbf{E}_s \otimes \mathbf{1}_t)^c \hat{\boldsymbol{\gamma}}_s + (\mathbf{E}_s \otimes \mathbf{E}_t)^c \hat{\boldsymbol{\gamma}}_{st}$) are plotted on the left side of Fig.5-15. The figure shows, for example, that the spatial component inflates prices in central Tokyo and its southwestern area, as well as in the area along the Chuo-line, a prime railway route. These areas are all popular residential areas, so the result is intuitively consistent.

The estimated spatiotemporal component can be decomposed in each spatial and temporal scale. For example, the component including spatial eigenvectors with large eigenvalues (or MC^+s) describes global scale spatial component. I define spatial scales of the spatial eigenvectors in $\mathbf{E}_s$ as

Global scale : $1.00 > MC_l^+/MC_1^+ > 0.50$,

Regional scale : $0.50 > MC_l^+/MC_1^+ > 0.25$,

Local scale : $0.25 > MC_l^+/MC_1^+ > 0.00$,

where MC_l^+ is the MC^+ value of the l-th spatial eigenvector. Since the MC_l^+s are proportional to λ_ls, MC_l^+/MC_1^+ can be evaluated by λ_l^+/λ_1^+. By definition, the estimated spatiotemporal component is the sum of the global, regional, and local components. Similarly, the temporal scales of the temporal eigenvectors in $\mathbf{E}_t$ are defined as

Long-term : $1.00 > MC_{t,\,l}^+/MC_{t,1}^+ > 0.50$,

Short-term : $0.50 > MC_{t\,l}^+/MC_{t,1}^+ > 0.00$,

where $MC_{t,l}^+$ is the MC_t^+ value of the l-th temporal eigenvector. While $(\mathbf{E}_s \otimes \mathbf{E}_t)^c \hat{\boldsymbol{\gamma}}_{st}$ is the time-variant component that explains the long-term and short-term temporal components, $(\mathbf{E}_s \otimes \mathbf{1}_t)^c \hat{\boldsymbol{\gamma}}_s$, which do not depend on $\mathbf{E}_t$, is the time-invariant component.

The estimated global components are plotted on the right side of Fig.5-15, and the regional and local components are shown in Fig.5-16. In 1995, the global component is significant in three areas: the central Tokyo area and areas around two cities, Kichijoji and Tachikawa. According to the questionnaire by NEXT Co. Ltd. in 2007 (http://www.next-group.jp/en/index.html), Kichijoji is the most popular residential city in Tokyo. On the other hand, Tachikawa is a major city that owns Tachikawa station, which has the greatest number of passengers of the stations in Tokyo, outside of the 23 wards including the central area (East Japan Railway Company: http://www.jreast.co.jp/e/). The global component seems to describe the positive influence of these prime urban areas. The three hot spots gradually merged, until in 2005, they became incorporated into one large hot spot. This is evidence that the urban areas in Tokyo have been combined over time (i.e., conurbation has taken place).

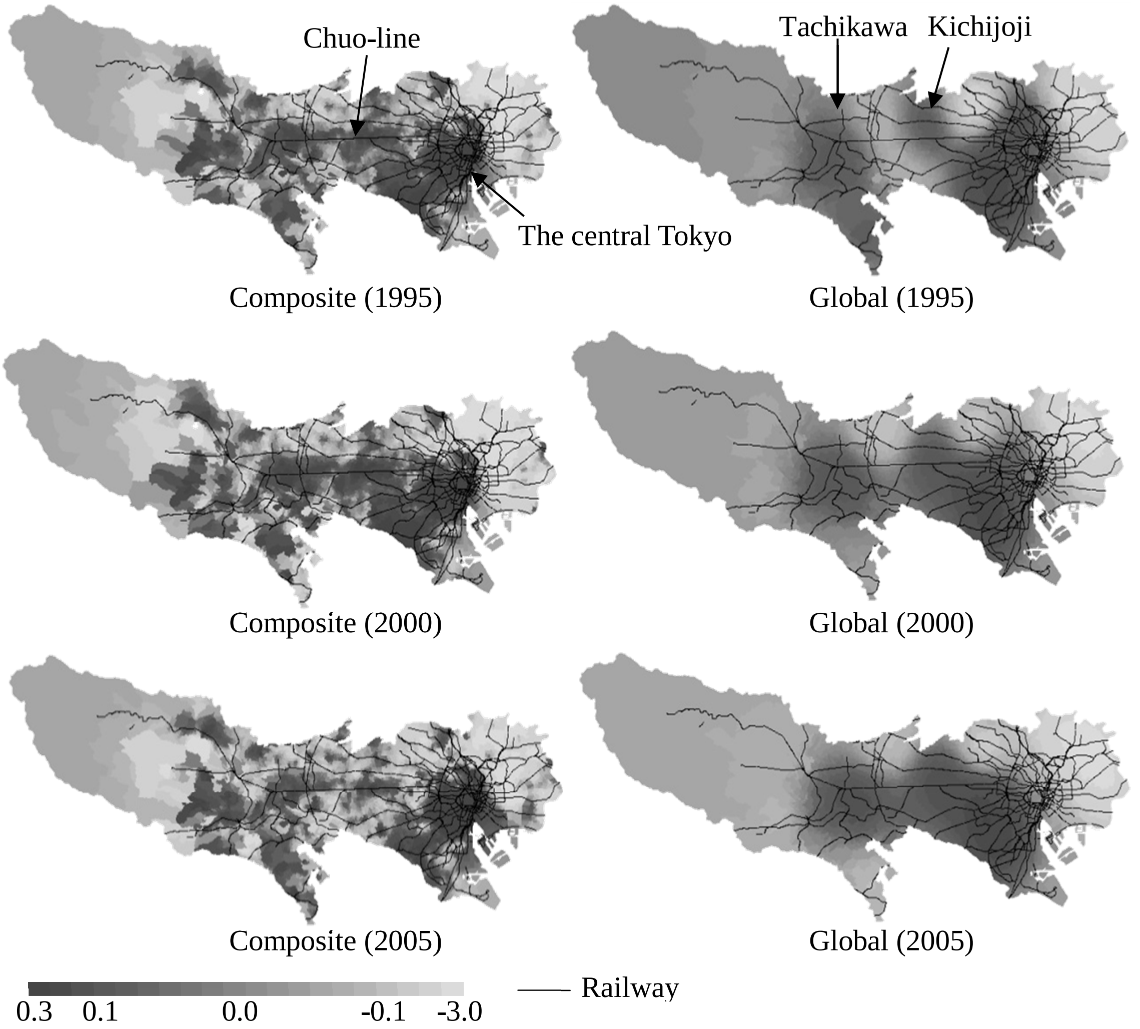


Figure 5-15: Extracted spatial components (Composite and Global component)

Note: The composite component is defined by the sum of the global, regional, and local component.

The regional scale component is again prominent around central Tokyo and Kichijoji. Beside, this component is also high in areas around the other cities, including Denenchofu, Hachioji, and Machida, which shows that these cities have regional scale influences. On the other hand, the local scale component seems to describe local heterogeneity. For instance, this component indicates high values in the area along the Chuo-line, which is a popular residential area, but indicates low values in the torus-shaped area around central Tokyo. The low values are somewhat unexpected. Thus, the multiscale decomposition is helpful to reveal hidden properties in spatiotemporal data. Furthermore, the regional and local components are relatively stable over time, in contrast to the global component.

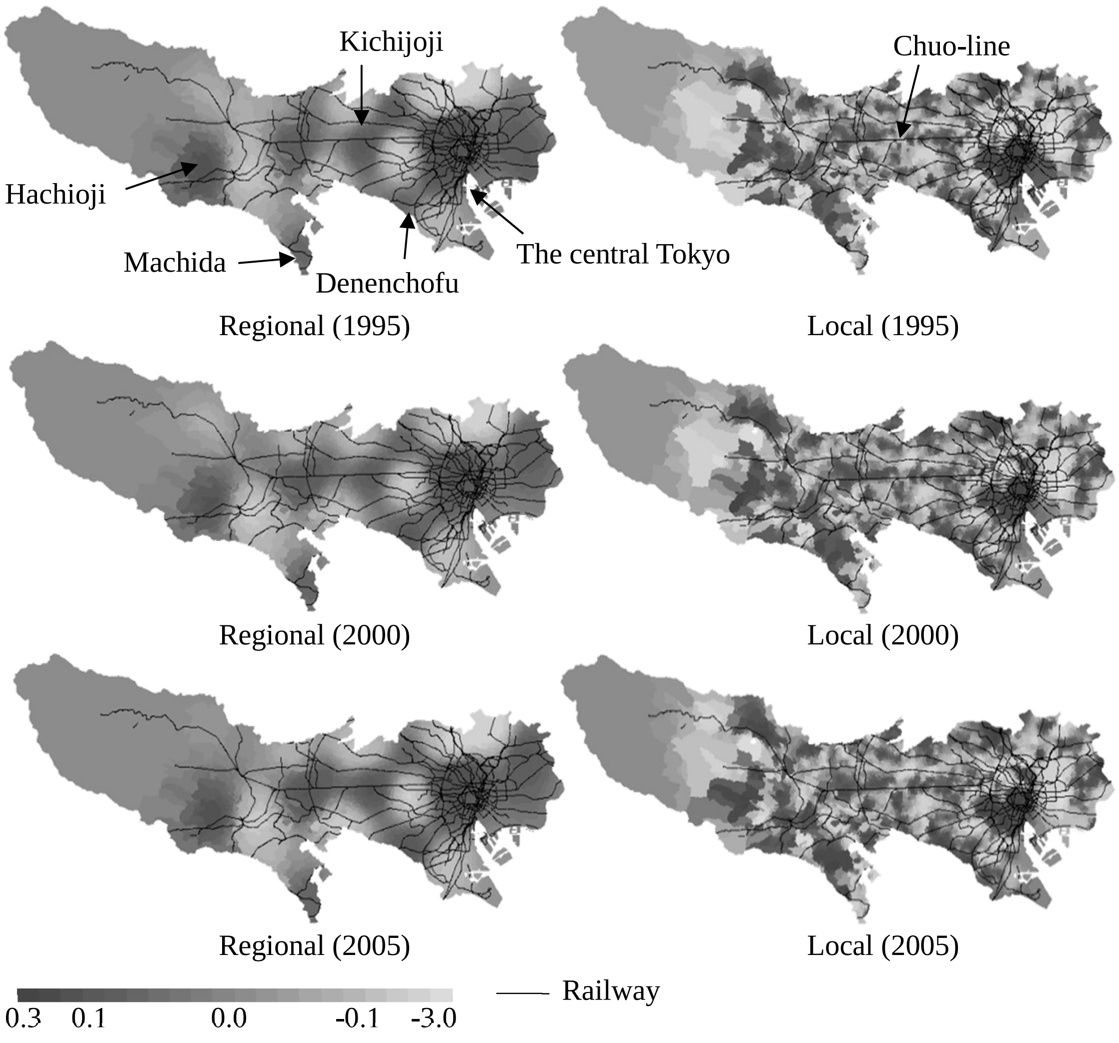
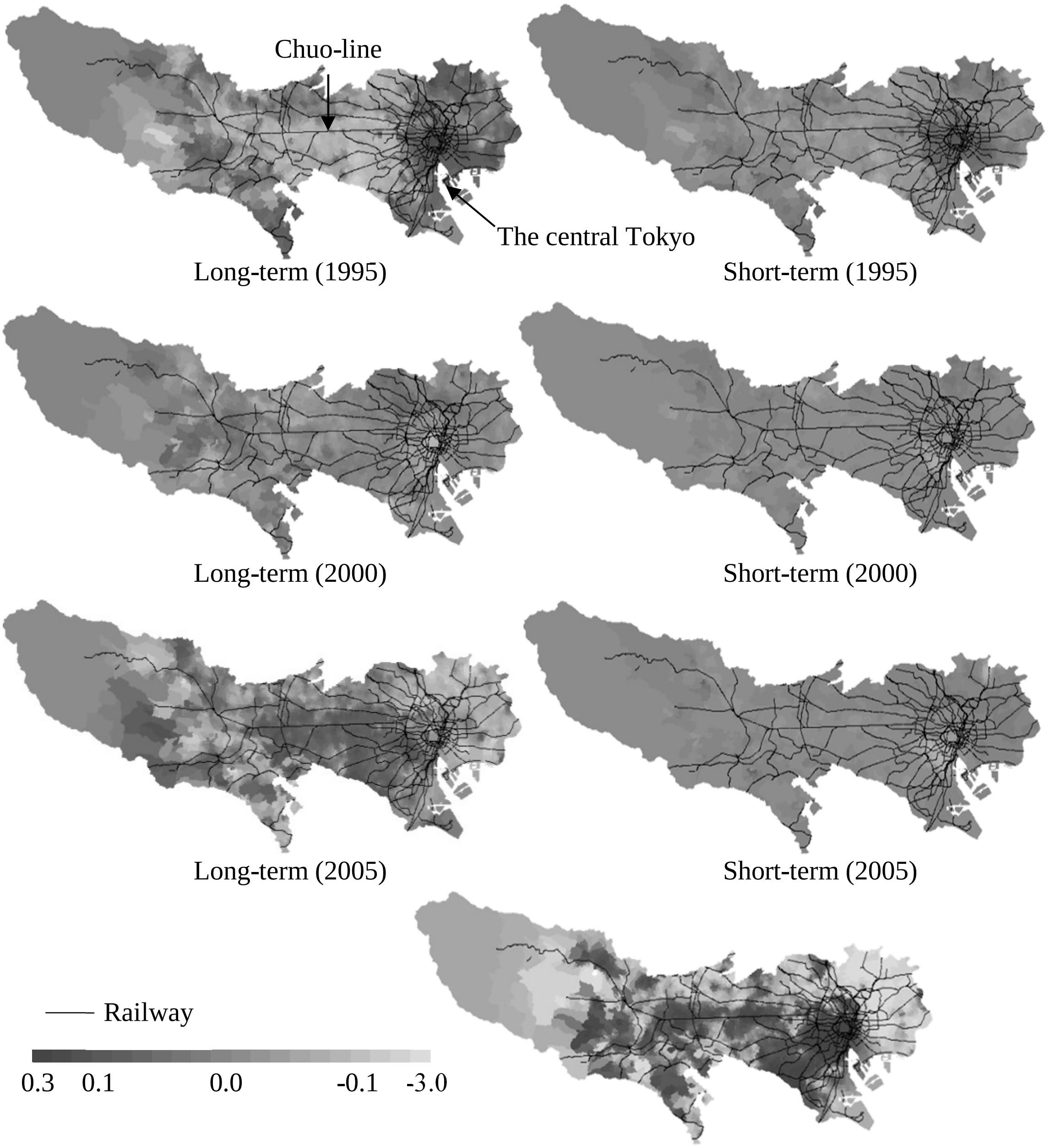


Figure 5-16: Extracted spatial components (Regional and local component)

The long-term, short-term, and time-invariant components are plotted in Fig.5-17. The long-term component in the mid-area, including the Chuo-line, and the southwest part of central Tokyo increased over time. This suggests that the land prices in these areas have become inflated compared to other areas. On the other hand, the estimated short-term component is quite small, although it displays relatively large variation in the central Tokyo area. This seems to imply heterogeneity in the central area. Finally, the time-invariant component suggests that land prices in central Tokyo and its southwestern area, as well as in the area along the Chuo-line are constantly high.



Long-term (1995)

Short-term (1995)

Long-term (2000)

Short-term (2000)

Long-term (2005)

Short-term (2005)

Time-invariant component

Figure 5-17: Extracted spatial components

Since the spatiotemporal eigenvectors are all orthogonal, the following equation holds (see, Legendre and Legendre, 2012):

$$R^2 = R^2_{g,0} + R^2_{g,L} + R^2_{g,S} + R^2_{r,0} + R^2_{r,L} + R^2_{r,S} + R^2_{l,0} + R^2_{l,L} + R^2_{l,S}\,, \tag{5-41}$$

where the subscripts *g*, *r*, and *l* denote spatial scales (global, regional, and local), the subscripts 0, *L*, and *S* denote temporal scales (long-term and short-term), R^2 is the *R*-squared of the dESF model, and $R^2_{A,B}$ is the *R*-squared of the dESF model with selected eigenvectors that have spatial scales of *A* and temporal scales of *B*.

The contribution of each spatiotemporal component to the model accuracy (i.e., R^2) can be evaluated using $R^2_{A,B}/R^2$. Table 4 summarizes the values of $R^2_{A,B}/R^2$s, and shows that the spatial components are prominent in the order of the local, global, and regional components. Thus, it is verified that the land prices have prominent local scale spatial variations. On the other hand, about 93% of the components are present in the time invariant component, indicating that land prices are stable over time. In addition, the long-term component is stronger than the short-term component.

Table 5-6: Contributions (%) of each component

		Spatially variant component		
		Global	Regional	Local
Time invariant component		24.17	11.06	57.79
Time variant component	Long-term	1.646	0.365	4.500
	Short-term	0.092	0.052	0.033

5.4.4. Discussion

This study extends the distance-based ESF for space-time modeling. The extended method, which is based on both the Moran coefficient and the standard geostatistical model, is suited for descriptive analysis, including spatiotemporal interpolation, spatial component extraction, and variance partitioning.

The proposed method is superior to the standard geostatistical method in some respects. Firstly, the method is computationally more efficient while its interpolation accuracy is almost same with the standard geostatistical model. A number of computationally efficient geostatistical methods, which would be faster than my eigen-decomposition-based method, have been proposed (see, e.g., Sun *et al*., 2012). However, they generally impose some approximations, and, generally, their predictive accuracies are worse than the standard geostatistical model. The proposed method also performs an approximation by removing insignificant eigenvectors, though, since the approximation is performed by an AICc-minimization, the approximation would never brought accuracy deterioration. The proposed method, which requires calculating spatial eigenvectors, would be slow when *N* is large. On the contrary, the temporal eigenvectors, which is defined on a 1-dimensional space, can be approximated using the sine function (Griffith, 2000; Borcard *et al*., 2004). In short, the proposed method is particular efficient for long-term spatiotemporal data.

Secondly, the proposed method reveals multiscale spatiotemporal structures in data, which cannot be captured by the standard geostatistical method. Spatial eigenvector-based approach have actively been discussed in ecology (see e.g., Legendre and Legendre, 2012), and, accordingly, like MEMs (see §5.2.6), my method might also be suited for ecological analysis. However, the proposed method is distinctive in that it models a continuous spatial process described by MC^+ while the ecological approach models a discrete spatial process described by *MC*.

The third advantage is simplicity. Its parameter estimation is conducted by an ordinary least squared (OLS)-based simple procedure and, and all conventional diagnostic statistics for linear regression model can be applicable directly.

On the other hand, the method also has a number of drawbacks. Firstly, computational time of its eigenvector selection would be large when a non-OLS estimation method is used. This is because the efficient eigenvector selection algorithm (see §5.2.5) is only for the OLS-based model. Applying penalized regression methods, including the lasso and ridge regression, might be helpful to cope with this problem. Secondly, the model has a limitation that it is only for longitudinal data. An EM algorithm-based approach that considers unbalanced longitudinal data as balanced longitudinal data with missing observations (e.g., LeSage and Pace, 2004) might be useful to overcome this limitation.

5.5. Summary

This study extends ESF to a form paralleling geostatistical data modeling. The formulated model is based on both a valid spatial process model in geostatistics and standard ESF methodology, and expresses spatial patterns described by MCs. Furthermore, because ESF specifications approximate spatial ecomonetric models (i.e., SLM/SEM), which are for discrete spatial data (see Tiefelsdorf and Griffith, 2007), the proposed model also can be considered an extension of spatial econometric models to continuous space. The usefulness of the method presented in this paper is confirmed by utilizing it for parameter estimation and ESDA.

As with standard ESF, an advantage of the proposed model is simplicity. Parameters in the proposed linear model are estimated using OLS, and values of eigenfunctions for arbitrary sites are obtained using a simple equation (Eq.5-23). The model is easily combined with other statistical models, such as those for logistic regression, Poisson regression, and mixed effects (see Griffith and Paelinck, 2011). One of its drawbacks is the exhaustive search needed for ESDA. The efficient selection algorithm discussed in §5.2.5.1 cannot be used for non-Gaussian models. Hence, efficient algorithms for eigenfunction selections need to be developed.

Continuous spatial model have been used to address various problems, which we do not discuss in this paper. Hence, examining the effectiveness of our method for more general problems is important. In addition, theoretical relationships between the proposed method and geostatistical models–for example, relationships between eigenfunctions extracted using the proposed method and variograms estimated using a geostatistical model–also must be clarified in future studies.

6. Sampling Design Problem: A Geostatistical Approach for Land Price Assessed Site Reduction

Changing the spatial support for point data requires considering the efficiency of both the point interpolation and interpolation site allocation. Thus, this chapter discusses another COSP for point data, namely the sampling design problem.

This chapter discusses how the spatial statistical sampling design approach can be applied to the land price-assessed site reduction problem in Japan. As the assessed sites are going to be reduced gradually after 2013, discussing this issue is important. However, the spatial statistical approach has never been applied to land price data. Accordingly, this chapter first extends the standard geostatistical sampling design approach for land price assessment data in Japan. Then, the effectiveness of the extended method is examined by applying it to actual land price data. Finally, the reduction problem is discussed using this method.

6.1. Methodology

6.1.1. Review of spatial sampling studies

Environmental/socio-economic data are monitored for various purposes. For instance, in Japan, concerns about weather have led to the measurement of weather data, and concerns about land transactions have led to the official assessment of land prices. Maintaining these data can be expensive. For example, the cost of the official land price assessment in Japan in 2010, which assesses land prices at 26,000 sites, is 3.74 billion JPY (Source: MLIT: http://www.mlit.go.jp/common/000213810.pdf), and similar amounts required each year. To use this investment effectively, the assessed site allocation must be determined judiciously, to help land transactions and other uses. Thus, discussing sample site allocation (assessed site allocation) or sampling design is important.

Methodologies for spatial sampling design are classified into design-based methods, which use a pre-determined scheme, and model-based methods, which use a model (Wang *et al*., 2012). The former group includes simple random sampling, in which sample sites are decided randomly, systematic sampling, in which samples are selected based on a given and preset order, stratified random sampling, which performs simple random sampling in each pre-determined, non-overlapping group (e.g., sub-region, age group), and two-step sampling, which selects a group randomly and performs simple random sampling on that group. Since these methods are for independent and identically distributed (i.i.d.) samples (stratified random sampling and two-step sampling assume i.i.d. for samples in each group), samples must be homogeneous. Systematic sampling outperforms the other methods when no prior knowledge is available about the samples (Ripley, 1981; Dunn and Harrison, 1993), whereas stratified random sampling is efficient when attributes in samples have strong spatial dependence (Ripley, 1981).

The model-based methods perform an optimization using a Monte Carlo-type simulation technique, including the Markov chain Monte Carlo (MCMC) method (Gelfand and Smith, 1990) and the simulated annealing (SA) method (Kirkpatrich *et al*., 1983). These methods have been well discussed in geostatistics (see §6. 2).

According to Wang *et al*. (2012), the design-based methods are suitable for "how much" problems, including estimating a global mean and standard deviation (of a population) whereas the model-based methods are more suitable for "where" problems, including the sample (or assessed) site relocation problem. Since this study focuses on the latter problem, the model-based approach will be discussed from here on.

6.1.2. Model-based sampling design

The model-based approach has been applied to two types of problems: the sampling design optimization problem for accurate spatial process description, and the optimization problem for efficient parameter estimation (Zimmerman, 2006; Zhu and Stein, 2005). Generally, the problem for accurate spatial process description minimizes either of the following objective functions (Zhu and Stein, 2005):

$$C_1(S) = \int_D V(\mathbf{s}_0; S) w(\mathbf{s}_0) d\mathbf{s}_0 \ , \tag{6-1}$$

$$C_2(S) = \max[V(\mathbf{s}_0; S) w(\mathbf{s}_0)], \tag{6-2}$$

where $\mathbf{s}_0 \in D \subset \Re^2$ denotes arbitrary sites in a study region D, $S = \{s_1, ... s_N\} \subset D$ denotes the sampling design (i.e., a collection of sample sites), $V(\mathbf{s}_0; S)$ denotes a squared prediction error (SPE) at location $\mathbf{s}_0$ under design S, and $w(\mathbf{s}_0)$ denotes the weight assigned to location $\mathbf{s}_0$. Eq.(6-1) provides a mini-sum solution, which is suitable when the quality of the overall sampling design must be maximized (i.e., total loss must be minimized). Eq.(6-2) provides a mini-max solution, which is suitable when the maximum loss at arbitrary sites in D must be minimized.

On the other hand, the optimization problem for efficient parameter estimation finds the optimal design that makes parameter estimation efficient. Efficient semivariogram model estimation has been discussed in geostatistics. For instance, Russo (1984) shows that the sampling design with unified numbers of location pairs within each of the lag-distance zones (see Fig.2-4) provides efficient semivariogram estimates. Warrick and Myers (1987) extended Russo’s (1984) idea to consider directions. On the other hand, Muller and Zimmerman (1999) propose a design minimizing the MSE of parameter estimators, which is defined as

$$\mathbf{M}(\boldsymbol{\theta}) = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'], \tag{6-3}$$

where $\boldsymbol{\theta}$ denotes the true parameter values and $\hat{\boldsymbol{\theta}}$ denotes their estimators. Since $\boldsymbol{\theta}$ is known, Eq.(6-3) cannot be applied directly. Hence, they minimize Eq.(6-3) by minimizing the inverse information matrix of the estimators that asymptotically converge to Eq.(6-3) (Zimmerman, 2006). The sampling design that minimizes Eq.(6-3) provides an efficient estimator.

Interestingly, it is known that the approaches for accurate spatial process description and approaches for efficient parameter estimation provide opposite results. More precisely, the former provides a spatially spread sampling design, while the latter provides a spatially clustered design (Zimmerman, 2006). However, these two approaches should not necessarily be discussed independently. In fact, the instability of $\boldsymbol{\theta}$ (MSE of $\boldsymbol{\theta}$), which is minimized in the latter group, is likely to influence the model accuracy (or SPE), which is maximized in the former group. Thus, Zhu and Stein (2005) and Zimmerman (2006) discuss the accuracy maximization problem while considering the instability of $\boldsymbol{\theta}$ in non-Bayesian fashions, whereas Diggle and Lophaven (2006) discussed the same in a Bayesian fashion.

More recently, preferential sampling (Diggle *et al.*, 2010) has been discussed extensively. This method allows dependency between sample values and sample allocations; for example, sample sites are densely located in areas where sample values are large (e.g., Olea, 2007; Diggle and Ribeiro, 2007; Gelfand, 2012). The Bayesian technique is required for preferential sampling. Applying the preferential sampling technique is important when some secondary statistical analyses, including parameter estimation and spatial prediction, are needed using the samples (e.g., Diggle *et al.*, 2010; Gelfand *et al.*, 2012).

Since sampling design optimization requires finding the global optimum from among many local optimum, the aforementioned methods require a Monte Carlo method (e.g., the simulated annealing method), which is computationally expensive. The complexity is particularly serious when the Bayesian approaches are applied (see, Diggle *et al.*, 2010; Zidec and Zimmerman, 2010).

The concern in this chapter is how to reduce the land price assessments sites. The reduction must be conducted to maintain the quality of the land price data. In other words, the resulting reduced design must describe land prices in the region well. Thus, I consider applying the accuracy maximization-based (or SPE-based) approach that finds the design with maximum descriptive capability. As a result of the computational expensiveness, this study does not consider either the instability of **θ** or the preference in sampling. The rationale for this is as follows. Ignoring the instability of **θ** on the accuracy maximization result is small (Zhu and Stein, 2005). Then, since the prime uses of the land price data (see §6.3.2) do not include statistical analyses, there is no clear advantage to applying preferential sampling.

6.1.3. Accuracy maximization-based geostatistical sampling design

Generally, the objective functions in Eqs.(6-1) and (6-2) are minimized after changing them into a tractable discretized form, as follows

$$C_1(S) = \sum_{\mathbf{s}_0 \in D} V(\mathbf{s}_0; S) w(\mathbf{s}_0), \tag{6-4}$$

$$C_2(S) = \max[V(\mathbf{s}_0; S) w(\mathbf{s}_0)]. \tag{6-5}$$

Eq.(6-4) or Eq.(6-5) are minimized, for example, by applying the simulated annealing method. The simulated annealing method is a heuristic algorithm that uses the following optimization procedure:

(i) Set an initial design, S_0, and an initial value for a parameter T, T_0.

(ii) Iterate (ii-1) and (ii-2) *iter* times, alternately.

(ii-1) Let S_i be the sampling design given at the i-th iteration, and let S_i' be the design given by randomly replacing a sample site in S_i with an un-sampled site in $\mathbf{s}_0$.

(ii-2) Calculate the values of the objective function, $C_g(S)$, for S_i and S_i', where $g \in \{1, 2\}$ (i.e., either Eq.6-4 or Eq.6-5). Then, if $C_g(S_i') \leq C_g(S_i)$, S_{i+1} is S_i'. On the other hand, if $C_g(S_i') > C_g(S_i)$, S_{i+1} is S_i', with the following probability:

$$\exp\left(-\frac{C_g(S_i') - C_g(S_i)}{T}\right). \tag{6-6}$$

Otherwise, S_{i+1} is S_i. Eq.(6-6) implies that the modified design, S_i', is accepted with the probability given in the equation, even if the modification worsens the objective function. This acceptance is required to reach the global optimum. The probability given by Eq.(6-6) is controlled by the parameter T, with a greater value of T indicating a larger acceptance ratio.

(iii) Replace T with pT, where p ($0 < p < 1$) is a fixed parameter that expresses the decreasing ratio of T.

(iv) Iterate (ii) and (iii) until S_i converges.

T_0, S_0, *iter*, and p must be determined a priori. Eq.(6-7) is a standard assumption for T_0 (e.g., Brus and Heuvelink, 2007):

$$T_0 = \frac{C_g(S_+^*) - C_g(S_0)}{N \log(0.8)}, \tag{6-7}$$

where S_+^* is the optimal design given by the simulated annealing method in which step (ii-2) accepts only improvements (i.e., if $C_g(S_i') \leq C_g(S_i)$, S_{i+1} is given by S_i', otherwise, S_{i+1} is given by S_i). Following Brus and Heuvelink (2007), this study sets S_0 randomly, and T_0, *iter*, and p are given by Eq.(6-7), 100, and 0.95, respectively.

6.2. Geostatistics for the land price assessed site reduction problem

6.2.1. Background

Japan has a huge deficit, and improving its financial soundness is a critical issue (Ministry of Finance, Japan: URL: http://www.mof.go.jp/english/index.htm). For the soundness, the efficiency/reduction of the land price assessment systems has been discussed, and it was decided that to reduce the number of assessed sites gradually after 2014 (YOMIURI ONLINE: http://www.yomiuri.co.jp/atmoney/news/20130109-OYT1T01049.htm: 2013/1/22 final access).

To maintain the quality of the assessment, the reduction in the number of assessed sites needs to be managed carefully. While applying the aforementioned sampling design techniques seems helpful to this reduction problem, I was not able to find any studies that use them for land price data. Therefore, in this study, I construct a methodology for this reduction problem, after discussing the details of the land price assessment systems in Japan.

6.2.2. Land prices assessment systems

There are two prime land prices in Japan: the officially assessed land price and the prefectural land price. The officially assessed land price is assessed by the Land Appraisal Committee under the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) at the beginning of the year, and is based on the Land Market Value Publication Act. This assessment provides standard market values of land per square meter for standard sites. Here, the standard land market values are the prices that would be formed in an assumed transaction without any extraordinary incentives that induce participants to sell off or buy aggressively (Land and Property in Japan: http://tochi.mlit.go.jp/english/: 2013/9/2 final access). Recently, land prices at the 26,000 standard sites are assessed every year. The prices in each of the sites are examined by more than two real-estate appraisers.

On the other hand, the prefectural land price is assessed by prefectures on July 1, based on the Enforcement Order for the National Land Use Planning Act. In this assessment, two or more real estate appraisers are placed on each site to assess the standard market value of the land per square meter of a standard site (e.g., in 2012, 22,264 sites were assessed). The prefectural land price plays a complementary role to the officially assessed land price.

The sites to be assessed are chosen from those in an urban planning area or from areas in which a certain number of land transactions is expected. In addition, the following criteria are used to decide on the assessed sites:

Representativeness: The sites must represent the land price level of the surrounding area.

Moderation : Occupancy condition, environment, land register, and so on, at the sites must be moderate.

Stability : The occupancy condition must be stable.

Certainty : The sites must be identifiable using, for example, land registers, buildings, and so on.

The adequacy of the assessed sites is checked every year, and sites that violate any of these criteria are replaced by more suitable sites. These criteria are basically for the officially assessed land prices, although many prefectures adopt the criteria for the prefectural land prices too. In addition, to complement the officially assessed land price from the viewpoint of space, more prefectural land prices are assessed outside of urban planning areas. The prefectural land prices also complement the officially assessed land prices from the viewpoint of time. This is because their assessments are conducted just half a year after the officially assessed land prices.

These land price data are provided by the National Land Numerical Information Download Service. The data have three main uses: (i) as a reference for the usual land transactions; (ii) as a reference for land acquisition and compensation by administration; and (iii) as a reference for taxation (e.g., inheritance tax and property tax).

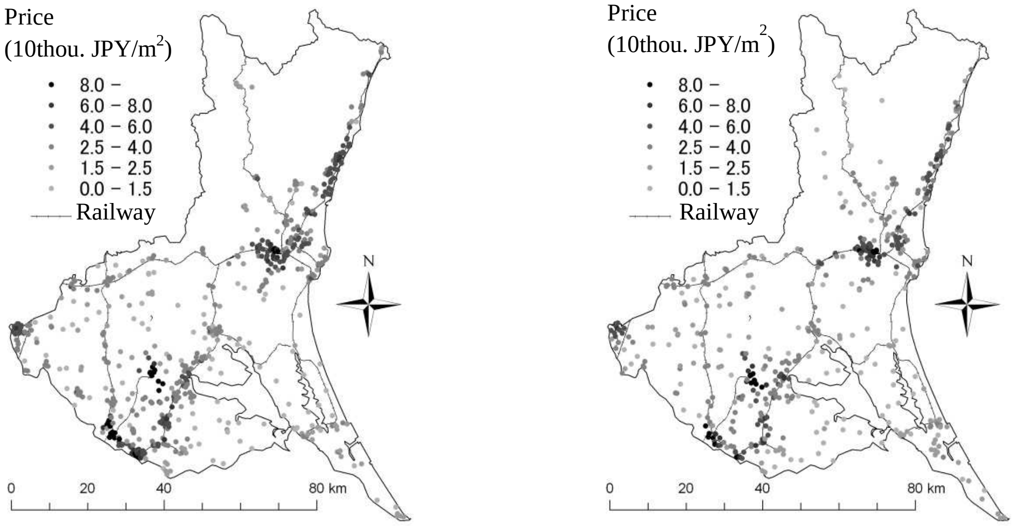
6.3. Model for the reduction problem

6.3.1. Assessed site reduction criteria

Generally, spatial sampling is discussed in terms of data quality and assessment cost, so I do so here as well. The data quality is maintained by applying the accuracy maximization-based approach. Although this approach cannot consider the cost, this is not necessarily needed when data acquisition costs are uniform over space. The assessment cost of land prices seems near uniform. Hence, this study assumes uniformity of cost, and does not consider this aspect further.

In addition to the data quality, considering the discussion in §6.2.2, I must consider the allocation criteria (including the four criteria) and diversity of use. The assessed site allocation criteria, which was discussed in §6.2.2, are summarized as follows: (i) Assessed sites must be allocated in areas with a certain number of expected transactions; (ii) An assessed site must have representativeness, moderation, stability, and certainty; and (iii) The prefectural land price data must complement the officially assessed land price data from the viewpoint of space and time. Since the prefectural land prices are assessed six months after the officially assessed land prices, temporal complementarity is automatically satisfied. In addition, because the accuracy maximization-based approach provides a sampling design with good coverage over a study area (Zidek and Zimmerman, 2010), complementarity over space is also satisfied if the geostatistical approach is used. However, satisfying the four allocation criteria in point (ii) by applying the geostatistical approach is not necessarily straightforward. Although, since the existing assessed sites have been determined to be consistent with the criteria, the criteria are also fulfilled as long as the sites to be assessed are chosen from among the existing sites.

I have clarified the following points: data quality and the complementarity on space and time are considered by applying the accuracy maximization-based approach; cost can be ignored by assuming uniformity of the assessment cost; the four allocation criteria are satisfied as long as the sites to be assessed are selected from among the existing assessed sites. In contrast, I am not sure how the expected number of land transactions and the diversity of use can be considered. Note that land transactions form one of the uses that falls under diversity of use. Hence, hereafter, I discuss only diversity of use.



Officially assessed land price

Prefectural land price

Figure 6-1: Residential land price in Ibaraki prefecture (2009)

6.3.2. Geostatistical assessed site reduction considering the diversity of use

To consider data quality, I reduce the assessed sites by applying the accuracy maximization-based approach. We can select an objective function from Eq.(6-1) or Eq.(6-2), and can set the weights, $w(\mathbf{s}_0)$, in these equations. §6.3.2 considers the most appropriate objective function and weights for each use (land transactions; land acquisition/compensation; taxation; see §6.2.2).

For land transactions, the assessed sites must fall within areas in which many transactions are expected. Such a preference can be considered by applying the expected transaction numbers (per unit area) to $w(\mathbf{s}_0)$. On the other hand, for land acquisitions/compensations or taxation, the assessed sites need to contain many households. This is because land acquisitions/compensations and taxation are conducted per household. In short, in the case of land transactions, $w(\mathbf{s}_0)$s are given by the expected numbers of transactions per unit area, while in the case of land acquisitions/compensations and taxation, they are given by the number of households per unit area.

Subsequently, the objective functions must be selected (either Eq.6-1 or Eq.6-2) for each use. The mini-sum function (Eq.6-1) minimizes the overall loss (the mean of the SPE, i.e., MSPE), but may include areas with a singularly large SPE. On the other hand, the mini-max function (Eq.6-2) avoids introducing such large SPE values, but its resulting design does not necessarily minimize the overall SPE or MSPE.

Land transactions reference both land price data and other data (e.g., real estate appraisal data and information on neighborhood transactions). The same holds for land acquisitions/compensations (e.g., expected profits and circumstances are checked). Accordingly, in these two uses, a certain level of loss in land price data can be covered by the other information referenced. However, to maintain the credibility of the land price data as an indicator of land transactions or land acquisitions/compensations, the inefficiency of including arbitrary sites must be reduced as much as possible. Accordingly, this study applies the mini-max approach for land transactions and land acquisitions/compensations, as it avoids introducing singularly large inefficiencies on arbitrary sites.

On the other hand, in Japan, land prices are directly related to taxation. For example, it has been established that the property tax valuation, which is a basis for property tax, must be about 0.7 times the spatially adjacent officially assessed land price values. In addition, land assessments for inheritance tax purposes must be about 0.8 times the spatially adjacent land price values. Thus, to adequately conduct taxation, the quality of the land price data must be maintained as much as possible. Hence, when considering taxation, this study uses the mini-sum function, which maximizes the overall data quality. The resulting objective functions for each use are summarized in Table 6-1.

Table 6-1: Objective functions for each use

	Land transaction	land acquisitions / compensations	Taxation
Objective function	$\max[V(\mathbf{s}_0;S)w(\mathbf{s}_0)]$	$\max[V(\mathbf{s}_0;S)w(\mathbf{s}_0)]$	$\sum_{\mathbf{s}_0\in D} V(\mathbf{s}_0;S)w(\mathbf{s}_0)$
$V(\mathbf{s}_0; S)$	SPE (squared prediction error)		
$w(\mathbf{s}_0)$	Expected transaction numbers per unit area	Household numbers per unit area	Household numbers per unit area

6.3.3. Computation of the weights ($w(\mathbf{s}_0)$)

The expected transaction numbers per unit area and the household numbers per unit area are used for the weights (see Table 6-1). The densities of householder numbers are calculated using data provided by E-Stat, a portal site provided by the Statistic Bureau, Ministry of Internal Affairs, and Communications, Japan (http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do).

This study estimates the expected number of transactions based on the land transaction data provided by Land General Information System, a portal site provided by the MLIT (http://www.land.mlit.go.jp/webland/). Using this data, the number of transactions in each minor municipal unit is easily obtained. However, these data are based on voluntary answers to a questionnaire, and so might be unreliable. In addition, the transaction numbers per minor municipal unit are small compared to the number of the minor municipal units (e.g., in the Ibaraki prefecture, [the number of transactions in residential lands]/[the number of minor municipal units] is less than 1), which would also make the data unstable.

To cope with such unreliability and instability, this study applies the Poisson-Gamma model (Bethlehem *et al.*, 1990), which is defined as

$$P(s_0) \sim Poisson(\theta(s_0) \times \overline{P}(s_0)), \tag{6-8}$$

$$\theta(s_0) \sim Gamma\,(a, b)\,, \tag{6-9}$$

where a and b are parameters, $P(s_0)$ is the number of transactions in the minor municipal unit s_0, and $\overline{P}(s_0)$ is the population that generates $P(s_0)$. This study gives $\overline{P}(s_0)$ by the number of households in s_0. Under these assumptions, $\theta(s_0)$ becomes the expected transaction number per household, and its estimator is given as

$$\hat{\theta}(s_0) = \frac{d(s_0) + \hat{a}}{\overline{d}(s_0) + \hat{b}}\,. \tag{6-10}$$

Eq.(6-10) can be considered an empirical Bayesian estimator of $\theta(s_0)$, with its prior distribution given by Eq.(6-9). Hence, as with the other Bayesian estimators, $\hat{\theta}(s_0)$ can be considered a shrinkage estimator. In the other words, $\hat{\theta}(s_0)$ is an estimator that copes with the aforementioned unreliability and instability. Finally, the estimator for the expected number of transactions is given by $\hat{\theta}(s_0)\overline{d}(s_0)$.

6.4. An empirical study

6.4.1. Outline

In this section, I describe how to use the proposed approach to reduce the number of sites assessed for the residential land prices (the officially assessed land price data + the prefectural land price data) in the Ibaraki prefecture in 2009 (sample size: 1,084; see Fig.6-2). In this study, reductions are conducted for each of the three uses, and, for each use, the number of sites is reduced by 108 (10% of the sample size), 325 (30% of the sample size), and 542 (50% of the sample size), respectively. In these reductions, 108, 325, or 542 sites are chosen randomly from among the existing 1,084 assessed sites. Then, the best design is identified by applying the simulated annealing method. Here, $\mathbf{s}_0$ is given by the geometric centers of 3,943 minor municipal units.

The SPEs are calculated using the standard geostatistical model given in Eq.(2-19). The response variables are the land prices (JPY/m^2). The explanatory variables are the Euclidean distance to the nearest station (Station: km), the railway network distance from the nearest station to the Tokyo station (Tokyo dist. km), the railway network distance from the nearest station to the Mito station (Mito dist. km), and the area of each land use type (Paddy, Agriculture, Forest, Wasteland, Railway, Road, Other land, Golf, River, Beach, Ocean) per 1 km^2 (see Table 6-2). The covariogram is given by the spherical model, Eq.(2-11), and parameters are estimated using the IRLS-based method (see §.2.2.4).

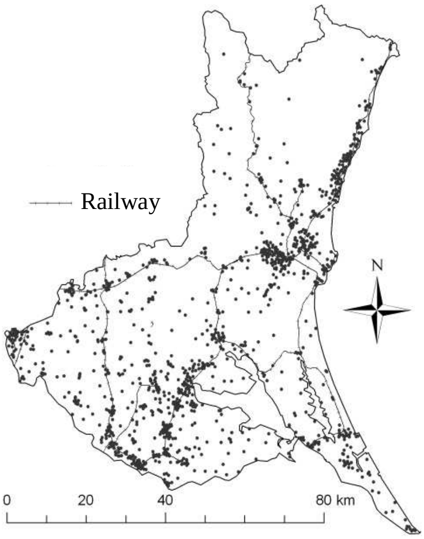


Figure 6-2: Assessed site allocation

Table 6-2: Variables applied in this empirical study

Variables	Description	Unit	Source
Tokyo dist.	Minimum railway distance from the nearest station to Tokyo station	km	NLNI* (2009)
Mito dist.	Minimum railway distance from the nearest station to Mito station		
Station	Distance to the nearest station		
Paddy	Area of paddy field	km^2 per unit area	
Agriculture	Area of agricultural land		
Forest	Area of forest		
Wasteland	Area of wasteland		
Railway	Area of railway		
Road	Area of road		
Other land	Area of other land		
Golf	Area of golf course		
River/Lake	Area of river/lake		
Beach	Area of beach		
Ocean	Area of beach and body of seawater		
Household Number	Household numbers in each minor municipal unit	Household	National census (2005)
Transaction numbers	Transaction numbers of residential lands in each minor municipal unit	Transaction number	Land General Information System (2009)

* NLNI: National Land Numerical Information download service

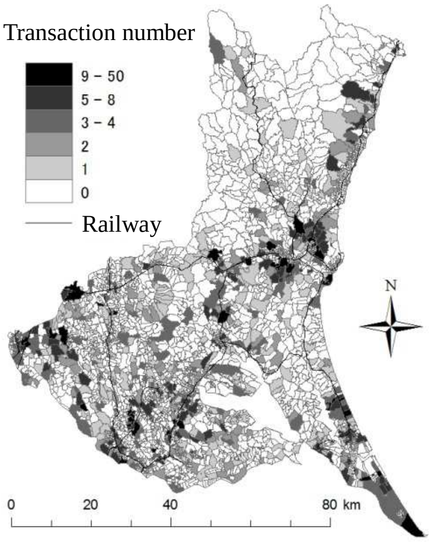


Figure 6-3: Transaction numbers

6.4.2. Parameter estimation

Station and Tokyo dist. are negatively significant at the 1% level, and Mito dist. is negatively significant at the 10% level. These results suggest that railways are an important factor in determining land prices. On the other hand, Paddy, Agriculture, Forest, and River/Lake are negatively significant at the 1% level. This suggests that these non-urban land uses have a negative impact.

The estimates of the partial-sill and nugget are 1.15 × 10^{8} and 6.92 × 10^{7}, respectively. These results indicate that 62.5 [= {1.15 × $10^{8/}$(1.84 × 10^{8} + 6.92 × 10^{7})} × 100] % of the disturbance is explained by spatial dependence. The estimated range is 6.48 km, which suggests that the land prices have local spatial variation.

The accuracy of the constructed model is checked by applying a five-fold-cross-validation. I first compare the resulting predicted values and their actual values using a 45° plot (see Fig.6-4). The comparison results suggest that the predicted values are similar to the actual values. The RMSE of the predicted values is 8,734 JPY/m^{2}, which is sufficiently small compared to the standard deviation of the land prices (see Table 6-3). Therefore, the constructed model is sufficiently accurate.

Table 6-3: Parameter estimation results

Variables	Estimates	*t* -values	
Const.	8.88×10^{4}	6.92	***
Tokyo dist	$–9.69\times10^{3}$	–3.73	***
Mito dist	$–1.24\times10^{3}$	–1.95	*
Station	$–4.75\times10^{3}$	–8.21	***
Paddy	$–2.09\times10^{–2}$	–8.98	***
Agriculture	$–2.69\times10^{–2}$	–8.59	***
Forest	$–1.60\times10^{–2}$	–4.77	***
Wasteland	$–6.21\times10^{–3}$	–1.32	
Road	$–1.57\times10^{–2}$	$–4.70\times10^{–1}$	
Railway	$1.59\times10^{–1}$	1.61	
Other land	$–3.24\times10^{–3}$	$–4.63\times10^{–1}$	
River/Lake	$–2.21\times10^{–2}$	–5.76	***
Beach	$–4.10\times10^{–2}$	–1.37	
Ocean	$–1.07\times10^{–2}$	–1.45	
Golf	$–6.61\times10^{–3}$	$–6.35\times10^{–1}$	
Nugget	6.92×10^{7}		
Partial-sill	1.15×10^{8}		
Range	6.48		

*, **, *** represent 10%, 5%, and 1% significance levels, respectively

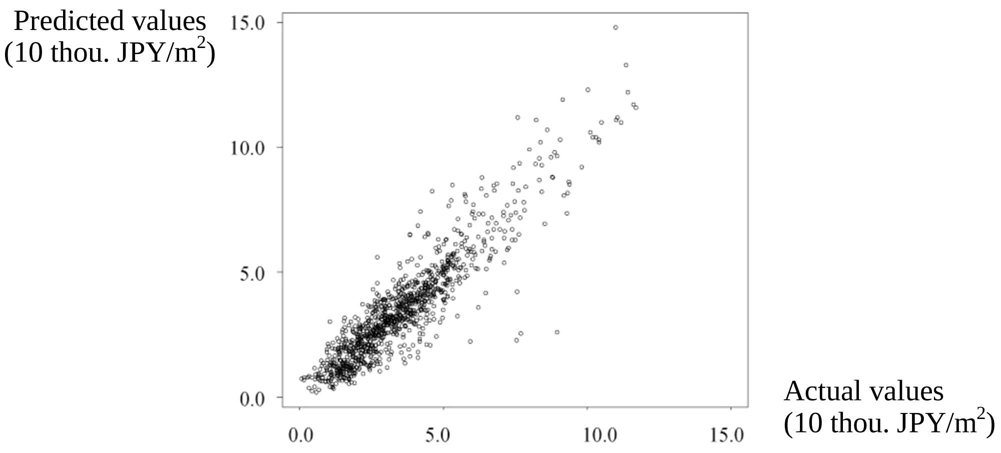


Figure 6-4: Comparison of the actual and predicted land price values

6.4.3. Result of the assessed sites reductions

The SPEs and the two types of weights, namely the number of households and the expected number of transactions, are plotted in Fig.6-5 and Fig.6-6, respectively. The SPEs are high in non-urban areas, while the number of households and expected number of transactions are high in urban areas. In the other words, the SPE recommends having more assessed sites in non-urban areas, while the expected number of transactions and number of households recommend having more sites in urban areas.

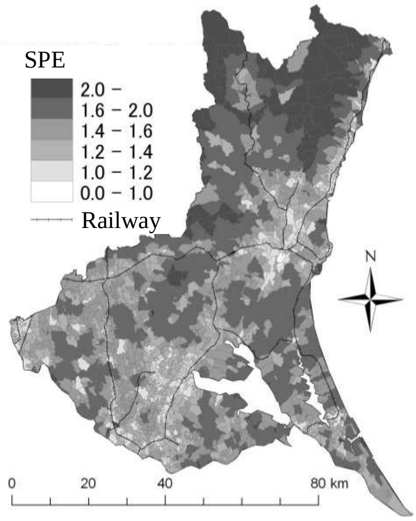


Figure 6-5: Squared prediction errors

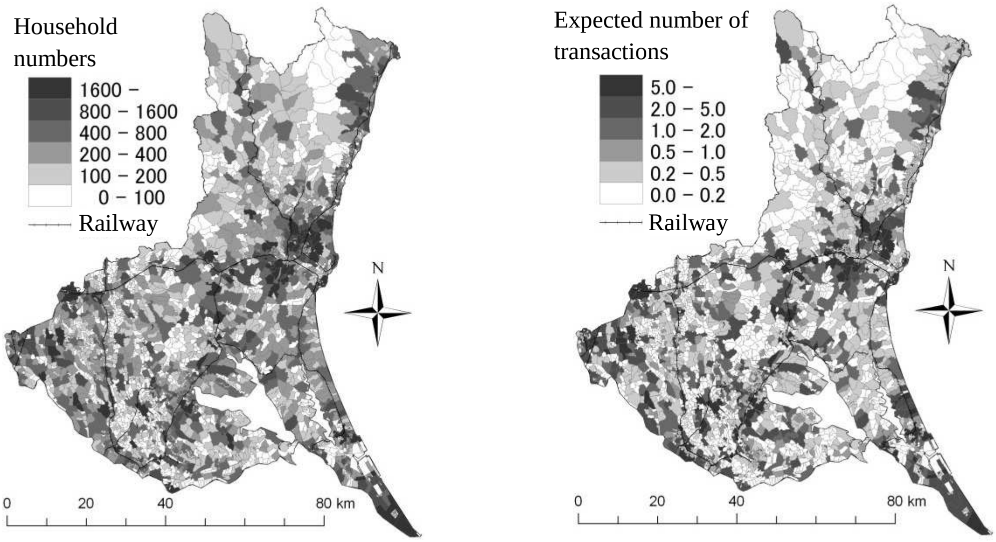


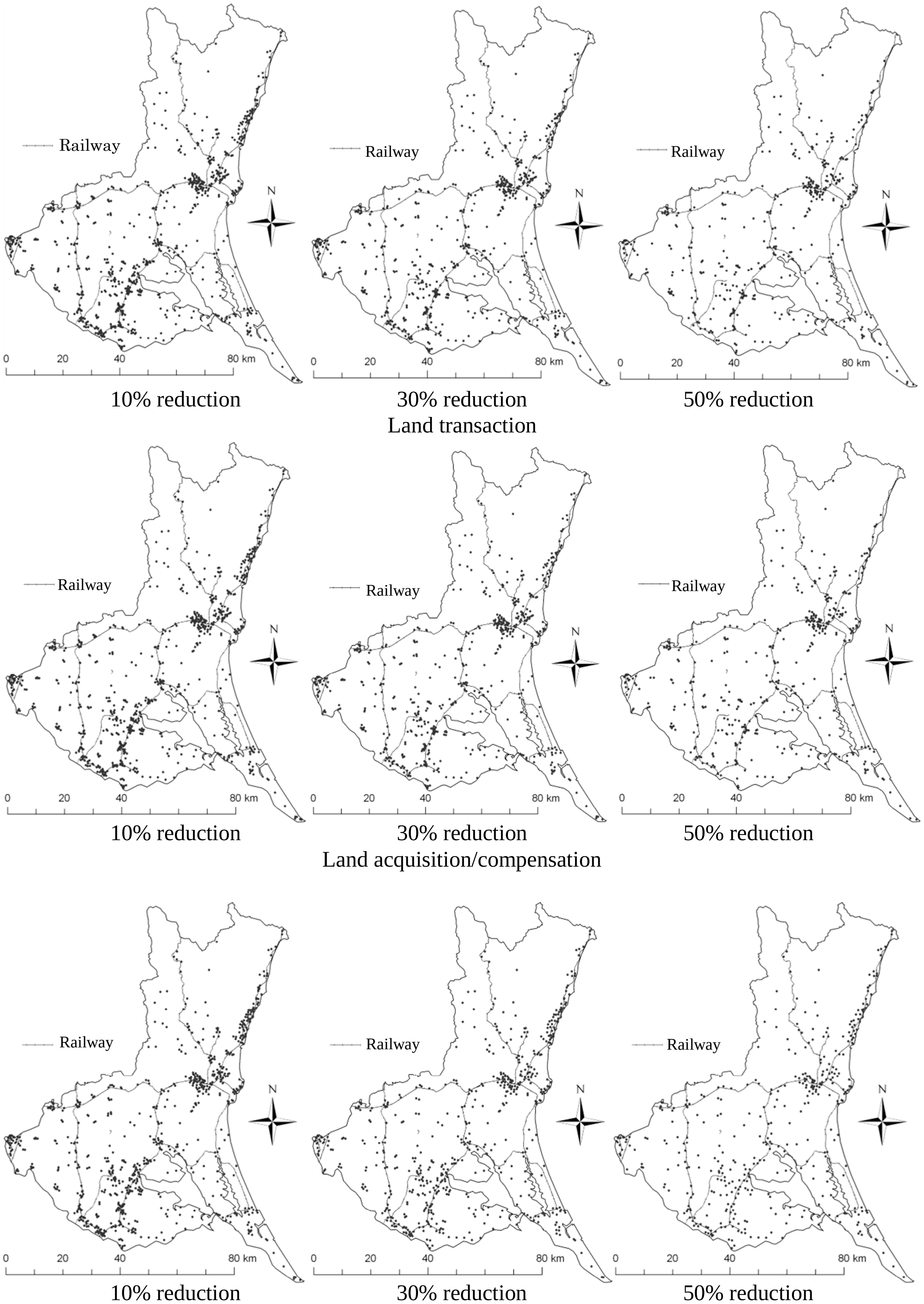
Figure 6-6: Household numbers (left) and expected number of transactions (right)

The results of the reduction in the number of assessed sites for land transactions, land acquisitions/compensations, and taxation are plotted in Fig.6-7. The results indicate the following features (see also Fig.6-2). Many assessed sites in non-urban areas have been maintained. Many assessed sites around Mito city have been maintained. Many assessed sites have been removed around the second-largest cities, including Tsukuba, Tsuchiura, and Hitachi. The latter result indicates that assessed sites in the second-largest cities should be removed a priori.

To discuss the similarities in the results, the ratios of commonly reduced assessed sites are calculated for each pair of uses. Table 6-4 summarizes the ratios given a 50% reduction. This table suggests that the results for land transactions and land acquisitions/compensations are relatively similar, but are less similar to the result for taxation. This dissimilarity is due to the difference in objective functions (see Table 6-1). The resulting allocation for taxation is more dispersed in the area that includes Tsukuba, Tsuchiura, and Toride, as well as the area around Hitachi, when compared to the results for land transactions and land acquisitions/compensations (particularly when 50% of the sites are reduced: see Fig.6-7). Considering such differences would be important to reduce the assessed sites properly.

Table 6-4: Ratio of sites that are commonly removed (50% reduction)

	Land transaction	Land acquisition/ compensation	Taxation
Land transaction	1.00	0.93	0.57
Land acquisition/ compensation	0.93	1.00	0.55
Taxation	0.57	0.55	1.00



10% reduction

30% reduction

50% reduction

Land transaction

10% reduction

30% reduction

50% reduction

Land acquisition/compensation

10% reduction

30% reduction

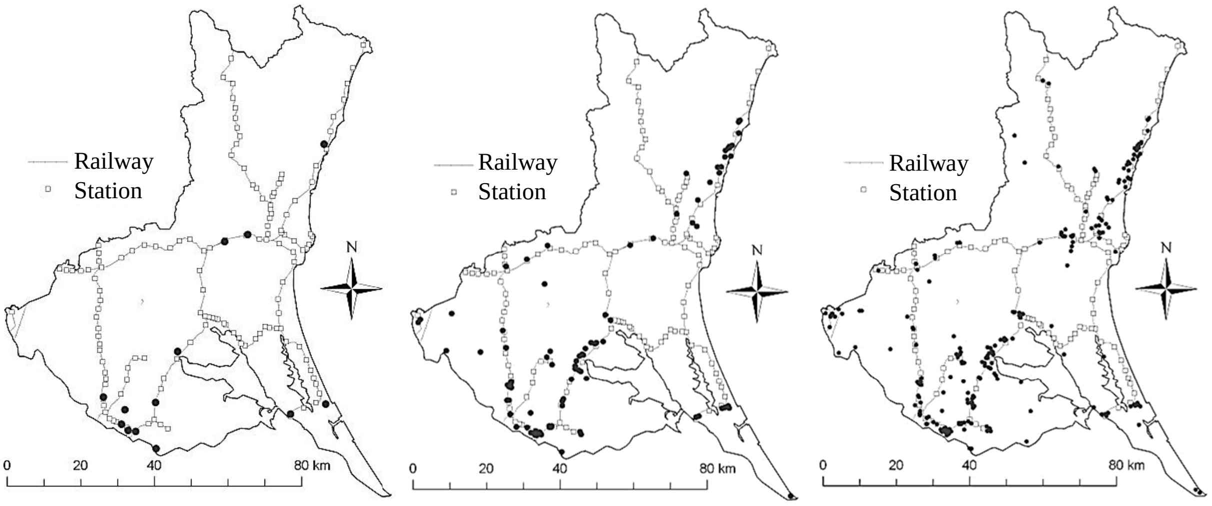
50% reduction

Taxation

Figure 6-7: Assessed site reduction results

Fig.6-8 plots the assessed sites that are removed in all three cases; in other words, the sites whose reductions are particularly recommended. This figure shows that, locally, a reduction in the number of assessed sites nearby to railway stations is recommended. However, globally, a reduction in the number of assessed sites around Toride, Tsuchiura, and Hitachi is recommended (the 10% reduction recommends a reduction in the number of sites around Toride only).

Finally, the realizations of the objective functions are summarized in Table 6-5. The figures in the table show that changes in the objective functions due to the reductions are small in all cases. In particular, the values before the reduction and the values after a 10% reduction are almost same. In addition, the decrease in the quality of the data is still small even when 50% of the sites are reduced. Thus, while smaller reductions are preferable to maintain the data quality, the data quality remains good even after a greater level of reduction.



10% reduction 30% reduction 50% reduction

Figure 6-8: Commonly reduced assessed sites

Table 6-5: Resulting objective function values

Sample size	Land transaction	Land acquisition/ compensation	Taxation
Full	5.510×10^{9}	1.026×10^{12}	2.866×10^{10}
10% reduction	5.510×10^{9}	1.026×10^{12}	2.870×10^{10}
30% reduction	5.511×10^{9}	1.027×10^{12}	2.934×10^{10}
50% reduction	5.517×10^{9}	1.028×10^{12}	3.092×10^{10}

6.5. Summary

This chapter extended the geostatistical approach to include land price data in Japan, and applied the proposed approach to the assessed site reduction problem. The results indicate which assessed sites are recommended for reductions, the decrease in data quality after the reductions, and so on. In addition, the discussion shows the effectiveness of the geostatistical approach for the reduction problem.

On the other hand, this study still has the following limitations. Firstly, it is not able to consider non-residential land prices, such as commercial land prices and industrial land prices. Non-residential land price data might have play a complementary roles to the residential land price data (e.g., an insufficiency of residential land price data might be covered by nearby non-residential land price data). To make my discussion more significant, non-residential land prices should be considered. In addition, the optimal allocations may possibly change drastically if a different geostatistical model is used (e.g., Fuentes et al., 2007). Hence, the reliability of the results must be verified. Finally, constructing a geostatistical model that is more suited to the reduction problem would be important.

7. Summary and Future Directions

This study proposed spatial statistical methods for COSPs. Chapters 3 and 4 discussed the main COSPs for areal data, namely the areal interpolation problem and the MAUP. Chapters 5 and 6 discussed the main COSPs for point data, namely the point interpolation problem and the sampling design problem. All of these discussions are important to conduct changing spatial support effectively.

Chapter 3 proposed a GWR-based areal interpolation method and clarified its effectiveness by comparing it to other geographical and geostatistical areal interpolation methods. I also conducted an empirical study of building stock estimation. Here, I verified that, while statistical methods have been accused of being ineffective in quantitative geography, the spatial statistical areal interpolation methods are efficient. Chapter 4 described the effectiveness of the GWR-based method for the MAUP. Chapter 5 discussed the extension to the ESF to handle continuous spatial data and applied the extended method to point interpolation, spatial component analysis, and so on. I also extended this method for spatiotemporal modeling. The results confirmed that the effectiveness of the proposed spatial and spatiotemporal models is comparable with standard geostatistical models. Chapter 6 develops the geostatistical sampling design approach for the land price assessment site reduction problem, showing that this approach provides intuitively reasonable reduction results.

Each of the chapters has revealed the effectiveness of spatial statistics for COSPs. However, considering the recent developments in GIS, we still have a lot of problems that must discuss. Firstly, considering the recent diversification of spatiotemporal data (Goodchild, 2010), the interpolation of spatiotemporal data must be discussed more thoroughly. The ESF-based spatiotemporal model proposed in Chapter 5 is significant in this regard. On the other hand, the GWR-based areal interpolation method must also be extended for spatiotemporal data. The extension would be useful, for example, when constructing municipal-level panel data between 1990 and 2010 in Japan, when many municipal units were merged. In addition, spatiotemporal modeling would be helpful in resolving the land price assessment site reduction problem from a long-term perspective.

On the other hand, spatiotemporal data have got larger and larger in recent years. The ESF-based spatiotemporal model might be helpful for large spatiotemporal data because of its computational efficiency. Furthermore, since the GWR-based areal interpolation method does not require inversion of the covariance matrix differ from the geostatistical models, it is also computationally more efficient than the standard geostatistical methods. However, these computational efficiencies would still be insufficient to handle so called “big data” (e.g., real-time twitter data). Thus, these methods must be made computationally more efficient.

Although I have discussed COSPs in each section somewhat independently, they should not necessarily be discussed in this way. For example, a researcher might need to estimate parameters considering the MAUP in an analysis using both areal data interpolated by an areal interpolation and point data interpolated by a point interpolation. Some geostatistical studies discuss such integrated problems by applying hierarchical Bayesian models (e.g., Sahu *et al*., 2010; Gelfand, 2010). Extending my methods to such problems is important.

While this study focuses mainly on model constructions, discussing these applications is also important. In this sense, the discussions of the building stock estimation in Chapter 3 and the land price assessment site reduction in Chapter 6 are meaningful. Currently, in a project in the National Institute of Environmental Studies, Japan, we are considering constructing a detailed population dataset using an areal interpolation technique. Such population data are already provided by the SEDAC (http://sedac.ciesin.columbia.edu/), although their data seem somewhat strange. In particular, their data on developing countries appear to be constructed using the simple areal weighting interpolation method, and their population values are disconnected at borders of regions/prefectures (see Fig.7-1). We have already confirmed that detailed populations can be estimated more accurately using my spatial statistical areal interpolation model. Completing our population data construction to provide this data is important.

Besides, providing calculation codes for the proposed methods is also important. I coded my results using R (http://cran.r-project.org/), a free statistical software package. We can upload packages (collections of R functions/codes) on R for free. Thus, uploading a package that collects my functions/codes is one way to achieve this.

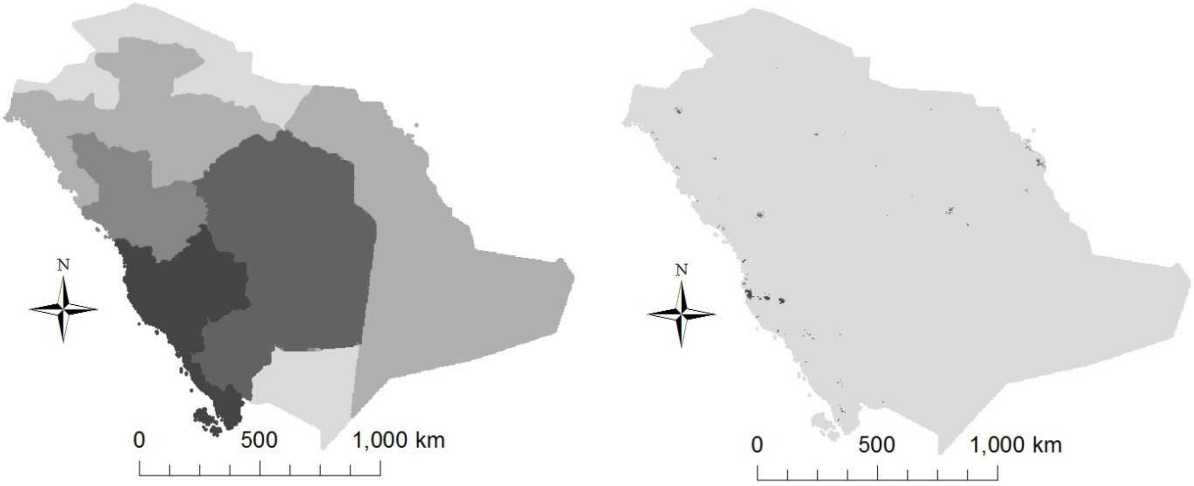


Figure 7-1: Detailed population densities in Saudi Arabia (2.5 arc-minute grid cells)

Note: Left: data provided by the SEDAC; Right: our estimates.

Acknowledgements

This dissertation would not have been completed without the help of many people. Firstly, my supervisor Prof. Morito Tsutsumi supported me generously in many ways. He educated me very thoughtfully considering my aptitude, my career path, and so on. In addition, he provided me with many opportunities, including the opportunity to present at international conferences, to study abroad, and more. This study life was stimulating, and enabled me to grow as a researcher. I would particularly like to thanks to him. Note that this dissertation is based on my master's study, as he was the original author of the idea.

Secondly, this study was supported by the colleges in the real estate & spatial statistics laboratory (or Tsutsumi lab). Dr. Hajime Seya, an OB of this lab, taught me a great deal about spatial statistics, and many of what I understand is thanks to him. In addition, his introduction to the book "Spatial Autocorrelation and Spatial Filtering" was the beginning of my interest in Prof. Daniel A. Griffith's study, subsequent to which I decided to visit him. Dr. Kazuki Tamesue, a PhD student in the Tsutsumi lab, also provided me with many suggestions that helped make my study more sophisticated. The other members also supported me in every aspect, which I needed to complete my study. The support of Secretary Hitomi Takahashi in clerical aspects was also very helpful.

Thirdly, professors at the University of Tsukuba contributed significantly to this dissertation. The advisors for this paper, Prof. Tsutomu Suzuki, Prof. Yoshiaki Osawa, Prof. Kunihiko Yoshino, and Prof. Yikio Sadahiro (University of Tokyo), carefully read my dissertation, and provided many helpful suggestions. Their comments greatly improved the quality of the dissertation. In addition, Prof. Haruo Oshida, Prof. Naohisa Okamoto, Prof. Mamoru Taniguchi, Prof. Ayako Taniguchi, and Prof. Yoshinori Kondo provided comments at the PhD seminars that were held jointly with the laboratories relating to transportation studies.

Prof. Daniel A. Griffith (The University of Texas, Dallas) was also an important contributor to my dissertation. He generously allowed me to study under him, despite my sudden request. He taught me a great deal about spatial statistics. I would not have been able to complete my discussion

on eigenvector spatial filtering without his guidance and support. Being able to discuss and author a paper with him will remain a lifetime memory. My study life at the University of Texas at Dallas was totally supported by many people, including Prof. Yongwan Chun, Dr. Monghyeon Lee, and Dr. Parmanand Sinha.

Many other people supported my study. Professor Yohiki Yamagata (National Institute of Environmental Study) provided me with opportunities to join research projects, during which time I had many good experiences. In addition, he let me join the S10 project and provided me with the opportunity to apply my work to the project. Prof. Makoto Tsukai (Univ. of Hiroshima) discussed my study with me often, and gave me much suggestive advice, particularly regarding areal interpolation. In addition, I was motivated by his and his college’s related works. I also had generous support from Prof. Yasushi Asami (Univ. of Tokyo) and Prof. Eihan Shimizu (Univ. of Tokyo). Furthermore, I received helpful comments from Prof. Yasumasa Matsuda (Univ. of Tohoku), Prof. Arthur Getis (San Diego State University), among others.

My study life was totally supported by my family. I want to give a special thanks to them. Besides, this dissertation was supported by a great many people, including others who I have not been able to list here. I would like to thank them all.

Reference

- Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer, Dordrecht.
- Anselin, L. (1995) Local indicators of spatial association–LISA, *Geographical Analysis*, 27: 93–115.
- Anselin, L. (2010) Thirty years of spatial econometrics, *Papers in Regional Science*, 89: 3–25.
- Anselin, L. and Rey, S. (1991) Properties of tests for spatial dependence in linear regression models, *Geographical Analysis*, 23: 112–131.
- Arbia, G. (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Growth Convergence*, Springer, New York.
- Armstrong, P. Diamond, M. (1984) Testing variograms for positive-definiteness. *Journal of the International Association for Mathematical Geology*, 16: 407–421.
- Aubry, N., Lian, W-Y., Titi, E.S. (1993) Preserving Symmetries in the Proper Orthogonal Decomposition. *SIAM Journal on Scientific Computing*, 14: 483–505.
- Banerjee, S. (2010) Spatial gradients and wombling. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (eds) *Handbook of Spatial Statistics,* CRC press.
- Bennett, R.P., Haining, R.P., Griffith, D.A. (1984) The problem of missing data on spatial surfaces. *Annals of the Association of American Geographer*, 74: 138–156.
- Berrocal, V.J., Gelfand, A.E., Holland, D.M. (2012) Space-time data fusion under error in computer model output: An application to modeling air quality. *Biometrics*, 68: 837–848.
- Berry, B.J.L., Marble, D.F. (1968) *Spatial Analysis: A Reader in Statistical Geograph*. Practice Hall.
- Bethlehem, J.G., Keller, W.J., Pannekoek, J. (1990) Disclosure control of micro data. *Journal of the American Statistical Association*, 85: 38-45.
- Borcard, D., Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbor matrices. *Ecological Modelling*, 153: 51-68.
- Borcard, D., Legendre, P., Avois-Jacquet, C., Tuosimoto, H. (2004) Dissecting the spatial

structure of ecological data at multiple scales. *Ecology*, 85: 1826–1832.

- Brillinger, D.R. (1990) Spatial-temporal modeling of spatially aggregate birth data. *Survey Methodology*, 16: 255–269.
- Brillinger, D.R. (1994) Examples of scientific problems and data analyses in demography, neurophysiology, and seismology. *Journal of Computational and Graphical Statistics*, 3: 1–22.
- Burnham, K.P., Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag.
- Butkiewicz T, Ross K (2010) Alleviating the modifiable areal unit problem within probe-based geospatial analyses. *Comput Graph Forum*, 29: 923–932.
- Chun, Y. (2008) Modeling network autocorrelation within migration flows by eigenvector spatial filtering. Journal of Geographical Systems, 10: 317–344.
- Cressie, N.A.C. (1985) Fitting variogram models by weighted least squares, *Mathematical Geology*, 17: 563–586.
- Cressie, N.A.C. (1993) *Statistics for Spatial Data, Revised Edition*, Wiley, New York.
- Cressie, N.A.C. and Hawkins, D.M. (1980) Robust estimation of the variogram, *Mathematical Geology*, 12: 115–125.
- Cressie, N., Johannesson, G. (2008) Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70: 209-226.
- Cressie, N.A.C. and Wikle, C.K. (2011) *Statistics for Spatio-Temporal Data*, Wiley, New York.
- Cromley, R., Hanink, D.M. and Bentley, G.C. (2012) A quantile regression approach for areal interpolation. *Annals of the Association of American Geographers*, 102: 763–777.
- Cuaresma, C.J., Feldkircher, M. (2013) Spatial filtering, model uncertainty and the speed of income convergence in Europe. *Journal of Applied Econometrics*, *28*: 720–741.
- Curriero F.C. (2006) On the use of non-Euclidean distance measures in geostatistics, *Mathematical Geology*, 38: 907–926.
- De Cesare, L., Myers, D.E., Posa, D. (2001) Product-sum covariance forspace-time modeling:

an environmental application. *Environmetrics*, 12: 11-23.

- Diggle, P.J. (2010) Historical introduction. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (eds) *Handbook of Spatial Statistics*, CRC press.
- Diggle, P.J., Lophaven, S. (2006) Bayesian geostatistical design. *Scandinavian. Journal of Statistics*, 33: 55–64.
- Diggle, P.J. and Ribeiro Jr., P.J. (2007) *Model-based Geostatistics*. Springer, New York.
- Diggle, P.J., Menezes, R. and Su, T.L. (2010) Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society, Series C*, 59: 191-232.
- Dray, S., Legendre, P., Peres-Neto, P.R. (2006) Spatial modeling: A comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecological modeling*, 196: 483–493.
- Dunn, R., Harrison, A.R. (1993) Two-dimentional systematic sampling of land use. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 42: 585-601.
- Eicher, C.L., Brewer, C.A. (2001) Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science*, 28: 125–138.
- Finley, A.O. (2011) Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2: 143–154.
- Fisher, P.F., Langford, M. (1995) Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, 27: 211–224.
- Fisher, P.F., Langford, M. (1996) Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation. *The Professional Geographer*, 48: 299–309.
- Flowerdew, R., Green, M. (1989) Statistical methods for inference between incompatible zonal systems, In: Goodchild, M., Gopal, S. (eds) *Accuracy of Spatial Databases*, Taylor and Francis, London.
- Flowerdew, R., Green, M. (1992) Developments in areal interpolation methods and GIS. *Annals of Regional Science*, 26: 67–78.

- Flowerdew, R. and Green, M. (1994) Areal interpolation and types of data. In: Fotheringham, S., Rogers, P. (eds) *Spatial Analysis and GIS*, Taylor and Francis, London.
- Fotheringham, S., Brunsdon, C., Charlton, M. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.
- Fuentes, M., Beich, B. (2010) Spectral domain. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (eds) *Handbook of Spatial Statistics*, CRC press.
- Gaetan, C. and Guyon, X. (2010) *Spatial Statistics and Modeling*, Springer, New York.
- Gelfand, A.E. (2010) Misaligned spatial data: The change of support problem. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (eds) *Handbook of Spatial Statistics*, CRC press.
- Gelfand, A.E. (2012) Hierarchical modeling for spatial data problems. *Spatial Statistics*, 1: 30–39.
- Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P., (2010) *Handbook of Spatial Statistics*, CRC press.
- Gelfand, A.E., Sahu, S.K., Holland, D.M. (2012) On the effect of preferential sampling in spatial prediction. *Environmetrics*, 23: 565-578.
- Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85: 398–409.
- Getis, A. (1990) Screening for spatial dependence in regression analysis. *Papers of the Regional Science Association*, 69: 69–81.
- Getis, A. (2010) Spatial Filtering in a Regression Framework: Examples Using Data on Urban Crime, Regional Inequality, and Government Expenditures. In: Fischer, M.M., Snickars, F., van Dijk, J-C., Westlund, H. (eds) *Advances in Spatial Science*, Springer.
- Getis, A., Griffith, D.A. (2002) Comparative spatial filtering in regression analysis. *Geographical Analysis*, 34: 130–140.
- Getis, A. and Ord, J.K. (1992) The analysis of spatial association by use of distance statistics, *Geographical Analysis*, 24: 189–206.
- Gneiting, T., Guttorp, P. (2010) Continuous parameter stochastic process theory. In: Gelfand,

A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (eds) *Handbook of Spatial Statistics*, CRC press.

- Goodchild, M. F (1992) Geographical Information Sciences. *International Journal of Geographical Information Systems*, 6 (1), 31-45.
- Goodchild, M. F (1997) What is Geographic Information Science? NCGIA Core Curriculum in GIScience.
- Goodchild, M.F. (2010) Twenty yours of progress: GIScience in 2010. *Journal of Spatial Information Science*, 1, 3-20.
- Goodchild, M.F. and Lam, N-S. (1980) Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing*, 1: 297–312.
- Goovaerts, P. (2006) Geostatistical analysis of disease data: Accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. *International Journal of Health Geographics*, 5: 52.
- Gotway, C.A., Young, L.J. (2002) Combining incompatible spatial data, *Journal of the American Statistical Association*, 97: 632–648.
- Gotway, C.A., Young, L.J. (2007) A geostatistical approach to linking geographically aggregated data from different sources. *Journal of Computational and Graphical Statistics*, 16: 115–135.
- Griffith, D.A. (1996) Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying georeferenced data. *Canadian Geographer*, 40: 351–367.
- Griffith, D.A. (2000) Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra and its Applications*, 321: 95-112.
- Griffith, D.A. (2002) A spatial filtering specification of the auto-Poisson model. *Statistics &Probability Letters*, 58: 245–251.
- Griffith, D.A. (2003) *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. Berlin: Springer–Verlag.
- Griffith, D.A. (2004a) A spatial filtering specification for the autologistic model. *Environment and Planning A*, 36: 1791–1811.

- Griffith, D.A. (2004b) Distributional properties of georeferenced random variables based on the eigenfunction spatial filter. *Journal of Geographical Systems*, 6: 263-288.
- Griffith, D.A. (2006) *Hidden negative spatial autocorrelation*. *Journal of Geographical Systems*, 8: 335–355.
- Griffith, D.A. (2009) Modeling spatial autocorrelation in spatial interaction data. *Journal of Geographical Systems*, 11: 117-140.
- Griffith, D.A. (2010) Spatial filtering. In: Fischer, M.M., Getis, A. (eds) *Handbook of Applied Spatial Analysis*. Springer, Berlin.
- Griffith, D.A., Paelinck, J.H.P. (2011) *Non-standard Spatial Statistics and Spatial Econometrics*, Springer, Berlin.
- Griffith, D.A., Peres-Neto, P.R. (2006) Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses in exploiting relative location information, *Ecology*, 87 (10), 2603–2613.
- Haining, R. (2003) *Spatial Data Analysis: Theory and Practice*, Cambridge University Press, Cambridge.
- Haining, R., Kerry, R. and Oliver, M.A. (2010) Geography, spatial data analysis, and geostatistics: An overview, *Geographical Analysis*, 42: 7–31.
- Handcock, M.S., Stein, M.L. (1993) A Bayesian analysis of kriging, *Technometrics*, 35:, 403–410.
- Harris, P., Brunsdon, C., Fotheringham, A.S. (2011a) Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor, *Stochastic Environmental Research and Risk Assessment*, 25: 123–138.
- Harris, R., Singleton, A., Grose, D., Brunsdon, C., Longley, P. (2010) Grid-enabling geographically weighted regression: A case study of participation in higher education in England, *Transactions in GIS*, 14: 43–61.
- Hoeting, J.A., Davis, R.A., Merton, A.A. and Thompson, S.E. (2006) Model selection for geostatistical models, *Ecological Applications*, 16: 87–98.

- Hughes, J., Haran, M. (2013) Dimension reduction and alleviation of confounding for spatial generalized linear mixed models, *Journal of the Royal Statistical Society B (Statistical Methodology)*, 75: 139–159.
- Jacob, G.B., Muturi, E.J., Caamano, E.X., Gunter, J.T. *et al*. (2008) Hydrological modeling of geophysical parameters of arboviral and protozoan disease vectors in Internally Displaced People camps in Gulu, Uganda. *International Journal of Health Geographics*, 7: 11.
- Kim, H. and Yao, X. (2010) Pycnophylactic interpolation revisited: integration with the dasymetric-mapping method. *International Journal of Remote Sensing*, 31: 5657–5671.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P. (1983) Optimizaton by simulated annealing. *Science*, 220: 671-680.
- Krige, D.G. (1951) A statistical approach to some basic mine valuation problems on the witwatersrand, *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 5: 119–139.
- Kyriakidis, P.C. (2004) A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36: 259–289.
- Kyriakidis, P.C., and Yoo, E.-H. (2005) Geostatistical prediction and simulation of point values from areal data. *Geographical Analysis*, 37: 124−151.
- Lam, N-S. (1983) Spatial Interpolation Methods: A Review. *The American Cartographer*, 1: 129–149.
- Langford, M. (2013). An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data. *Geographical Analysis*, 45: 324–344.
- Legendre, P., Legendre, L. (2012) *Numerical Ecology, 3rd edition*, Elsevier Science BV, Amsterdam.
- LeSage, J.P., Pace, R.K. (2009) *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, Boca Raton, Florida.
- LeSage, J.P. and Pace, R.K. (2004) Models for spatially dependent missing data, *The Journal of Real Estate Finance and Economics*, 29: 233–254.

- Leung, Y., Mei, C.-L. and Zhang, W.-X. (2000a) Statistical tests for spatial nonstationarity based on the geographically weighted regression model, *Environment and Planning A*, 32: 9–32.
- Lin J., Cromley, R. and Zhang. C. (2011) Using geographically weighted regression to solve the areal interpolation problem. *Annals of GIS*, 17: 1–14.
- Lo, C.P. (2008) Population estimation using geographically weighted regression. *GIScience & Remote Sensing*, 45: 131–148.
- Longley, P.A., Goodchild, M., Maguire, D.J., Rhind, D.W. (2010) *Geographic Information Systems and Science*. Wiley.
- Mardia, K.V., Marshall, R.J. (1984) Maximum likelihood estimation of models for
- residual covariance in spatial regression. *Biometrika*, 71: 135–146.
- Martin, R.J. (1984). Exact maximum likelihood for incomplete data from a correlated Gaussian process, *Communications in Statistics Theory and Methods*, 13: 1275–1288.
- Matheron, G. (1963) Principles of geostatistics, *Economic Geology*, 58: 1246–1266.
- Matheron, G. (1973) The intrinsic random functions and their applications. *Advances in Applied Probability*, 5: 439-468.
- Matsuo, T., Nychka, D.W., Paul, D. (2011) Nonstationary covariance modeling for incomplete data: Monte Carlo EM approach. *Computational Statistical & Data Analysis*, 55: 2059–2073.
- Menke, W. (1989) *Geophysical Data Analysis: Discrete Inverse Theory (International Geophysics)*. Academic Press.
- Mennis, J., Hultgren, T. (2006) Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, 33: 179–194.
- Miyagi, T, (2009) Study on practical use of Computable Urban Economic Model considered developer's behavior. Master’s thesis, Graduate School of Systems and Information Engineering, Univ. Tsukuba, Japan [Japanese].
- Moran, P.A.P. (1950) A test for the serial dependence of residuals, *Biometrika*, 37 (1-2), 178–181.
- Mrozinski, R.D., Cromley, R.G. (1999) Singly- and doubly-constrained methods of areal

interpolation for vector-based GIS. *Transactions in GIS*, 3: 285–301.

- Mugglin, A.S., Bradley, P., Carlin B P. and Gelfand, A.E. (2000) Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association*, 95: 877–887.
- Mugglin, A., Carlin, B.P. (1998) Hierarchical modeling in geographic information systems: Population interpolation over incompatible zones. *Journal of Agricultural, Biological, and Environmental Statistics*, 3: 111-130
- Mugglin, A.S., Carlin, B.P., Zhu, L. and Conlon, E. (1999) Bayesian areal interpolation, estimation, and smoothing: An inferential approach for geographic information systems: Population interpolation over incompatible zones. *Journal of Agricultural, Biological, and Environmental Statistics*, 3: 117–130.
- Muller, W.G., and Zimmerman, D.L. (1999) Optimal designs for variogram estimation. *Environmetrics*, 10: 23–37.
- Murakami, D., Tsutsumi, M. (2012) Practical spatial statistics for areal interpolation. *Environment and Planning B: Planning and Design*, 39: 1016–1033.
- Nagle, N.N., Sweeney, S.H., Kyriakidis, P.C. (2011) A geostatistical linear regression model for small area data. *Geographical Analysis*, 43: 38–60.
- Nocedal, J., Wright, S. (2006) *Numerical Optimization*. Springer.
- Odoi, A., Martin, W., Michel, P., Holt, J., Middleton, D., Wilson, J. (2003) Geographical and temporal distribution of human giardiasis in Ontario Canada. International Journal of Health Geographics: 5.
- Olea, R.A. (2007) Declustering of clustered preferential sampling for global histogram and semivariogram inference. *Mathematical Geology*, 39: 453-467
- Openshaw, S. (1984) *The modifiable areal unit problem*. Norwich, UK: Geo Books.
- Openshaw, S., Taylor, P. (1979) A Million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley, N. (eds) *Statistical Methods in the Spatial Sciences*, London: Pion: 127–144.
- Pace, P.K., LeSage, J.P., Zhu, J. (2011) Interpretation and computation of estimates from

regression models using spatial filtering. 5-th International Conference of the Spatial Econometrics Association, Toulouse, France.

- Paciorek, C.J. (2010) The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, 25: 107–125.
- Páez, A., Farber, S., Wheeler, D. (2011) A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43: 2992–3010.
- Pecci F, Pontarollo N (2010) The application of spatial filtering technique to the economic convergence of the European regions between 1995 and 2007. In: Taniar D, Gervasi O, Murgante B, Pardede E, Apduhan B (ed) Computational Science and its Applications—ICCSA 2010, Lecture notes in computer science, Berlin, Springer–Verlag: 46–61
- Petuelli, R., Griffith, D.A., Tiefelsdorf, M., Nijkamp, P. (2011) Spatial filtering and eigenvector stability: space-time models for German unemployment data. *International Regional Science Review*, 34: 253-280.
- Peres-Neto, P.R., Legendre, P., Dray, S., Borcard, D. (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecological Society of America*, 87: 2614–2625.
- Pintore, A., Holmes, C.C. (2004) Non-stationary covariance functions via spatially adaptive spectra. Technical Report, University of Oxford.
- Reibel, M. and Agrawal, A. (2007) Areal interpolation of population counts using pre-classified land cover data. *Population Research and Policy Review*, 26: 619–633.
- Ren, Q., Banerjee, S. (2013) Hierarchical factor models for large spatially misaligned data: A low–rank predictive process approach. *Biometrics*, 69: 19–30.
- Reynolds, H.D. (1998) The modifiable areal unit problem: empirical analysis by statistical simulation. Thesis, University of Toronto.
- Ripley, B.D. (1981) *Spatial Statistics*, Wiley, New York.

- Russ, J.C. (2006) *The image processing handbook*. CRC Press.
- Russo, D. (1984) Design of an optimal sampling network for estimating the variogram. *Soil Science Society of American Journal*, 52: 708–716
- Sadahiro, Y. (2000) Accuracy of count data estimated by the point-in-polygon method. *Geographical Analysis*, 32: 64–89.
- Sahu, S.K., Gelfand, A.E., Holland, D.M. (2010) Fusing point and areal level space–time data with application to wet deposition. *Journal of the Royal Statistical Society C (Applied Statistics)*, 59: 77–103.
- Sakata, T, Yoshikawa, T. (2001) An analysis of the consistency in building usage between the city planning basic survey and the map data of fixed property tax. *Theory and Applications of GIS*, 9: 9–18 [Japanese].
- Schabenberger, O., Gotway, C.A. (2005) *Statistical Methods for Spatial Data Analysis*. Chapman and Hall/CRC.
- Schott, J.R. (2005) *Matrix Analysis for Statistics (Wiley Series in Probability and Statistics)*. Wiley, New York, USA.
- Siffel C, Strickl MC, Gardner BR, Kirby RS, Correa A (2006) Role of geographic information systems in birth defects surveillance and research. Birth Defects Res A: Clin Mol Teratol 76: 825–833.
- Stein, M.L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York.
- Sun, Y., Li, B., Genton, M.G. (2012) Geostatistics for large datasets, *Advances and Challenges in Space-time Modelling of Natural Events* (eds. Montero, J.M., Porcu, E. and Schlather, M.), Springer, Berlin.
- Swift. A., Liu, L., Uber, J. (2008) Reducing MAUP bias of correlation statistics between water quality and GI illness. *Comput Environ Urban Systems*, 32: 134–148.
- Tagashira, N., Okabe, A. (2002) The modifiable areal unit problem in a regression model whose independent variable is a distance from a predetermined point. *Geographical Analysis*, 34: 1–19.

- Thayn, J.B., Simanis, J.M. (2013) Accounting for spatial autocorrelation in linear regression models using spatial filtering with eigenvectors. *Annals of the Association of American Geographers*, 103: 47–66.
- Tiefelsdorf, M. (1998) Some practical applications of Moran's Is exact conditional distribution. *Papers in Regional Science*, 77: 101–129.
- Tiefelsdorf, M. (2003) Misspecification in interaction model distance decay. *Journal of Geographical Systems*, 5: 25–50.
- Tiefelsdorf M, Griffith, D.A. (2007) Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environment and Planning A*, 39: 1193–1221.
- Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region, *Economic Geography*, 46: 234–240.
- Tobler, W.R. (1979) Smooth pycnophylactic interpolation for geographical, regions. *Journal of the American Statistical Association*, 74: 519–530.
- Tranmer, M., Steel, D. (1998) Using census data to investigate the causes of the ecological fallacy. *Environment and Planning A*, 30: 817–831.
- Tsutsumi, M., Seya, H. (2009) Hedonic approaches based on spatial econometrics and spatial statistics: Application to evaluation of project benefits, *Journal of Geographical Systems*, 11: 357–380.
- Tsutsumi, M., Miyagi, T., Yamasaki, K. (2012) Potential of computable urban economic model formalizing building market. *Journal of JSCE, Division D: Infrastructure Planning and Management*, 68: 333–343 [Japanese].
- Tukey, J.W. (1977) *Exploratory Data Analysis*. Addison-Wesley.
- Wang, J-F., Stein, A., Gao, B-B., Ge, Y. (2012) Review of spatial sampling. *Spatial Statistics*, 2: 1-14.
- Warrick A.W., Myers, D.E. (1987) Calculations of error variances with standardized variograms. *Soil Science Society of America*, 5: 265-268.
- Wheeler, D.C. (2007) Diagnostic tools and a remedial method for collinearity in geographically

weighted regression, *Environment and Planning A*, 39: 2464–2481.

- Wheeler, D. (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and Planning A*, 41: 722–742.
- Wheeler, D., Páez, A. (2010) Geographically weighted regression. In: Fischer, M.M., Getis, A. (eds) *Handbook of Applied Spatial Analysis*, Springer.
- Wheeler, D., Tiefelsdorf, M. (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7: 161–187.
- Wilks, D.S. (2006) *Statistical Methods in the Atmospheric Sciences, Volume 100, 2nd Edition (International Geophysics)*. Academic Press.
- Williams, C.K.I., Seeger, M. (2001) Using the Nyström method to speed up kernel machines. In: Leen, T.K., Diettrich, T.G., Tresp, V. (eds) *Advances in Neural Information Processing Systems 13,* MIT Press.
- White, H. (1980) A heteroskedastic-covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, 48 (4), 817–838.
- Whittle, P. (1954). On stationary process in the plane, *Biometrika*, 41 (3-4), 434–449.
- Wikle, C.K. and Berliner, L.M. (2005) Combining information across spatial scales, *Technimetrics*, 47 (1), 80–91.
- Wong, D. (2009) The modifiable areal unit problem (MAUP). In: Fotheringham, A.S., Rogerson, P.A. (eds) *The SAGE Handbook of Spatial Analysis*, SAGE.
- Wright, J.K. (1936) A method of mining densities of population with Cape Cod as an example. *Geographical Review*, 26: 103–110.
- Xie, Y. (1995) The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems,* 19: 287–306.

- Yamagata, Y., Seya, H., Nakamichi, K. (2013) Creation of future urban environmental scenarios using a geographically explicit land-use model: A case study of Tokyo. *Annals of GIS*, 19: 153–168.
- Yamagata, Y., Seya, H. (2013) Simulating a future smart city: An integrated land use-energy model. *Applied Energy*, 112, 1466–1474.
- Yoo, E-H., Kyriakidis, P.C. (2006) Area-to-point kriging with inequality-type data. *Journal of Geographical Systems*, 8: 357–390.
- Yoo, E-H., Kyriakidis, P.C., Tobler, W. (2010) Reconstructing population density surfaces from areal data: A comparison of Tobler's pycnophylactic interpolation method and area-to-point kriging. *Geographical Analysis*, 42: 78–98.
- Zhang, C., Qiu, F. (2011) A Point-Based Intelligent Approach to Areal Interpolation. *The Professional Geographer*, 63: 262–276.
- Zhu, Z., Stein, M. (2005) Spatial sampling design for parameter estimation of the covariance function. *Journal of Statistical Planning and Inference*, 134: 583–603.
- Zidec, J.V., Zimmerman, D.L. (2010) Monitoring network design, In: Gelfand, A.E., Diggle, P.J., Fuentes, M. and Guttorp, P. (eds) *Handbook of Spatial Statistics*, 45–56, CRC Press.
- Zimmerman, D.L. (1993) Another look at anisotropy in geostatistics, *Mathematical Geology*, 25: 453–470.
- Zimmerman, D.L. (2006) Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, 17: 635-652.
- Zimmerman, D.L., Stein, M. (2010) Classical geostatistical methods, In: Gelfand, A.E., Diggle, P.J., Fuentes, M. and Guttorp, P. (eds) *Handbook of Spatial Statistics*, 45–56, CRC Press.