

# **Area-to-point parameter estimation with geographically weighted regression**

Daisuke Murakami

Center for Global Environmental Research, National Institute for Environmental Studies,

16-2 Onogawa, Tsukuba-shi, Ibaraki, 305-8506, Japan

murakami.daisuke@nies.go.jp

Morito Tsutsumi

Faculty of Engineering, Information and Systems, University of Tsukuba,

1-1-1 Tennodai, Tsukuba-shi, Ibaraki, 305-8573, Japan

tsutsumi@sk.tsukuba.ac.jp

Abstract: The modifiable areal unit problems (MAUP) is a problem by which aggregated units of data influence the results of spatial data analysis. Standard GWR, which ignores aggregation mechanisms, cannot be considered to serve as an efficient countermeasure of MAUP. Accordingly, this study proposes a type of GWR with aggregation mechanisms, termed area-to-point (ATP) GWR herein. ATP GWR, which is closely related to geostatistical approaches, estimates the disaggregate-level local trend parameters by using aggregated variables. We examine the effectiveness of ATP GWR for mitigating MAUP through a simulation study and an empirical study. The simulation study indicates that the method proposed herein is robust to the MAUP when the spatial scales of aggregation are not too global compared with the scale of the underlying spatial variations. The empirical studies demonstrate that the method provides intuitively consistent estimates.

JEL codes: C43, C21, R12

Keywords: modifiable areal unit problem, geographically weighted regression, change of support problem, geostatistics

## 1. Introduction

Differences among the units in spatial data often complicate their analyses, and spatial unit conversions (aggregation and disaggregation) often become necessary. Spatial unit conversions are strongly related to both areal interpolation problems and modifiable areal unit problems (MAUP; Openshaw and Taylor 1979). The former refers to the problems of conducting accurate spatial unit conversions and the latter to the problems of addressing biases in the parameter estimates that follow such conversions. While both these problems are important topics in geography, we primarily focus on MAUP.

MAUP is mainly caused by two factors (Wong 2009). The first is the underlying spatial pattern of the data: MAUP has a large influence if the data are positively spatially dependent and a small influence if the data are negatively dependent (e.g., Reynolds 1998). The second factor is the aggregation process: when aggregated units are large, large variability can be canceled out by aggregation and MAUP could then become serious.

According to Swift et al. (2008), at least five approaches have been proposed to address MAUP. The first approach is the use of disaggregated data (e.g., Tagashida and Okabe 2002); the second is estimating aggregate-level parameters by considering disaggregate-level structures in a variance–covariance matrix (e.g., Tranmer and Steel 1998); the third is optimizing the zoning system by minimizing intra-zone variances and maximizing variances between the zones (Openshaw 1984); the fourth is performing a sensitivity analysis (e.g., Odoi et al. 2003; Swift et al. 2008); and the fifth is applying geographically weighted regression (GWR; e.g., Fotheringham et al. 2002; Páez et al. 2002; Wheeler and Páez 2010). Since GWR models spatial patterns in the data, which are a source of MAUP, GWR is believed to be robust to MAUP. However, because GWR does not consider aggregation mechanisms, it cannot be considered to be a measure that mitigates MAUP (Fotheringham et al. 2002; Wong 2009). As an exception, Young and Gotway (2007) show that the consideration of an aggregation mechanism in a GWR-based analysis can change the analysis results. However, they consider the aggregation mechanism to predict the explanatory variables in their GWR model, and their GWR model in itself does not consider it.

The aforementioned approaches for MAUP have some drawbacks, and theoretically sufficient measures for mitigating MAUP do not yet exist (Siffel et al. 2006; Butkiewicz and Ross 2010).

In geostatistics, MAUP is regarded as a change of support problem (COSP; e.g., Cressie 1996; Gotway and Young 2002; Schabenberger and Gotway 2005; Fuentes and Raftery 2005; Lee et al. 2009; Young et al. 2009a, b; Berrocal et al. 2010; Sahu et al. 2010; Gelfand 2010, 2012; Nagle et al. 2011; Berrocal et al. 2012). The general framework of COSP discusses changing spatial supports (e.g., units, data locations, etc.). Yet, whereas the majority of COSP studies focus on changing supports (e.g., areal interpolations), some discuss the biases in parameter estimates from doing so, including MAUP (e.g., Gotway and Young 2002; Gelfand 2010; Nagle et al. 2011).

Today, COSP is a topic of current interest in geostatistics, and the effectiveness of geostatistical approaches to COSP has been demonstrated in a number of studies. However, geographical MAUP studies and geostatistical COSP studies have been discussed practically independently thus far (see Haining et al. 2010)<sup>1</sup>. Combining the discussions on MAUP in geography and geostatistics would thus be useful step to advancing this research stream. In order to bridge this gap, this study extends GWR, which has been discussed in geographical MAUP studies (e.g., Fotheringham et al. 2002; Wong 2009), and proposes a method termed area-to-point (ATP) GWR. In particular, we focus on a specific type of MAUP, namely when regression parameter estimates change depending on the spatial scale of aggregations (i.e., the scale problem).

The remainder of this paper is organized as follows. We briefly explain GWR in Section 2, and the ATP GWR approach is developed in Section 3. The effectiveness of ATP GWR for mitigating the MAUP is examined by using a simulation study in Section 4; the method proposed herein is then applied to two empirical studies in Section 5. We next show the theoretical relationships between ATP GWR and a representative geostatistical approach in Section 6. Finally, Section 7 summarizes our discussion and concludes.

---

<sup>1</sup> As an exception, Vichiensan et al. (2008) discuss a COSP using GWR. However, different from our study, they do not consider the volume-preserving property (Lam 1983) that the aggregation of disaggregated variables must be equal to the actual aggregate variable values (see §3.1).

## 2. The GWR model

GWR is a spatial extension of the standard linear regression model that allows parameter values to vary continuously in geographical space. The basic model for site  $s_i$  on  $D \subset \mathfrak{R}^2$  is described as follows (see also Fotheringham et al. 2002):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon} \quad E[\boldsymbol{\varepsilon}] = \mathbf{0} \quad Var[\boldsymbol{\varepsilon}] = \sigma^2\mathbf{I}, \quad (1)$$

where  $\mathbf{y}$  is a vector of the explained variables,  $\mathbf{X}$  is a matrix of the explanatory variables,  $\boldsymbol{\varepsilon}$  is a vector of disturbances,  $\mathbf{0}$  is a vector of zeros,  $\mathbf{I}$  is an identity matrix, and  $\sigma^2$  is a variance parameter.  $\boldsymbol{\beta}_i$  is a vector of the spatially varying parameters at  $s_i$ .

The parameters in  $\boldsymbol{\beta}_i$  are estimated by assigning more weight to nearby observations than to more distant ones. The estimates are given as follows:

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}'\mathbf{G}_i\mathbf{X})^{-1}\mathbf{X}'\mathbf{G}_i\mathbf{y}, \quad (2)$$

where  $'$  denotes the matrix transpose,  $\mathbf{G}_i$  is a diagonal matrix whose  $i'$ -th element is  $g_{i,i'}$ , the strength of the connectivity between  $s_i$  and  $s_{i'}$ , and  $g_{i,i'}$  is given by using a distance-decay function. The Gaussian function is one of the most commonly used distance-decay functions:

$$g_{i,i'} = \exp\left(-\frac{h_{i,i'}^2}{b^2}\right), \quad (3)$$

where  $h_{i,i'}$  is the distance between  $s_i$  and  $s_{i'}$  and  $b$  is a bandwidth parameter. The bandwidth parameter is calibrated, for example, by cross-validation.

## 3. GWR for the MAUP

There are two types of variables: extensive (or volume) variables and intensive (including density, ratio, and averaged) variables. The former increases in proportion to the size of the spatial unit, whereas the latter is independent of the size. For example, population and area are extensive variables, while population density and the ratio of residential zones are intensive variables. Hereafter, we assume that intensive variables are explained variables to which GWR has typically been applied. Note that, as shown in Appendix 1, the

discussion presented in this section is easily extended for extensive variables. This section develops a GWR-based approach that mitigates the MAUP by estimating disaggregate-level parameters, which are essentially free from aggregation mechanisms, by using aggregated variables.

### 3.1. Model

We assume a model composed of two sub-models. The first is disaggregate-level GWR, which is given by the standard GWR model in Eq. (1) as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_d + \boldsymbol{\varepsilon} \quad E[\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\boldsymbol{\varepsilon}] = \sigma^2\mathbf{I}, \quad (4)$$

where  $d$  is an index of the disaggregated units and  $\boldsymbol{\beta}_d$  is a vector of the spatially varying parameters in the  $d$ -th disaggregated unit. The second model is aggregate-level GWR, which is automatically obtained by multiplying aggregation matrix  $\mathbf{A}$  by the left-hand side of Eq. (4):

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\bar{\boldsymbol{\beta}}_d + \bar{\boldsymbol{\varepsilon}} \quad E[\bar{\boldsymbol{\varepsilon}}] = \mathbf{0} \quad \text{Var}[\bar{\boldsymbol{\varepsilon}}] = \sigma^2\mathbf{A}\mathbf{A}^t, \quad (5)$$

where  $\bar{\mathbf{y}}$  is a vector of the aggregate-level explained variables,  $\bar{\mathbf{X}} = \mathbf{A}\mathbf{X}$ ,  $\bar{\boldsymbol{\varepsilon}} = \mathbf{A}\boldsymbol{\varepsilon}$ , and  $\bar{\mathbf{0}} = \mathbf{A}\mathbf{0}$ . This study assumes that  $\mathbf{y}$  is unknown, while  $\bar{\mathbf{y}}$  and  $\mathbf{X}$  (and  $\bar{\mathbf{X}}$ ) are known.

$\mathbf{A}$  must be determined by considering the volume-preserving property (Lam 1983) that the aggregation of disaggregated variables,  $\mathbf{A}\mathbf{y}$ , must be equal to the actual aggregate variable values,  $\bar{\mathbf{y}}$ . In other words,  $\bar{\mathbf{y}} = \mathbf{A}\mathbf{y}$  must hold. For example, suppose that the population density in an aggregated unit  $a$  is  $\bar{y}_a$  and unit  $a$  comprises two disaggregated units  $d$  and  $d'$ ; then,  $\bar{\mathbf{y}} = \mathbf{A}\mathbf{y}$  is expressed as

$$\bar{y}_a = \begin{bmatrix} A_{a,d} & A_{a,d'} \end{bmatrix} \begin{bmatrix} Y_d/w_d \\ Y_{d'}/w_{d'} \end{bmatrix},$$

$$\frac{Y_d + Y_{d'}}{w_d + w_{d'}} = A_{a,d} \frac{Y_d}{w_d} + A_{a,d'} \frac{Y_{d'}}{w_{d'}}, \quad (6)$$

where  $A_{a,d}$  is the  $(a, d)$ -th element of  $\mathbf{A}$ ,  $Y_d$  and  $Y_{d'}$  are the populations in the disaggregated units, and  $w_d$  and  $w_{d'}$  are the areas of these units. Eq. (6) merely states that  $\bar{y}_a$  must be equal to the weighted average of

population densities in the disaggregated units; hence,  $Y_d/w_d$  and  $Y_{d'}/w_{d'}$  are satisfied by defining  $A_d$  and  $A_{d'}$  with  $w_d/(w_d+w_{d'})$  and  $w_{d'}/(w_d+w_{d'})$ , respectively.

In general,  $\bar{\mathbf{y}} = \mathbf{A}\mathbf{y}$  is fulfilled by defining  $A_{a,d}$  as follows:

$$A_{a,d} = \begin{cases} w_d / \sum_{d \subseteq a} w_d & \text{if } d \subseteq a \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

where  $d \subseteq a$  denotes the disaggregated units in the  $a$ -th aggregated unit,  $A_{a,d}$  takes a value between 0 and 1, and  $\sum_{d \subseteq a} A_{a,d} = 1$ .

By definition, intensive variables are always expressed as  $Y_d/w_d$ , as assumed in Eq. (6), and our framework is applicable to any intensive variables as long as data on  $w_d$  are available. For example, when  $\bar{\mathbf{y}}$  consists of variables divided by area (e.g., population density),  $A_{a,d}$  must be given by Eq. (7) whose  $w_d$  is the area of the  $d$ -th unit (see Eq. (6)). When  $\bar{\mathbf{y}}$  consists of the variables divided by population (e.g., per capita income),  $A_{a,d}$  must be given by Eq. (7) whose  $w_d$  is the population of the  $d$ -th unit. In the case that  $\bar{\mathbf{y}}$  consists of average housing prices,  $A_{a,d}$  can be given by Eq. (7) with its  $w_d$  being defined by the variables, which are proportional to the number of dwellings, such as the number of households. Further, although for the sake of brevity this point is not discussed in detail, our framework is modifiable for extensive (or volume) variables, too (see Appendix 1).

By substituting Eq. (7) for  $\mathbf{A}\mathbf{A}^t$ , this matrix becomes a diagonal matrix with its  $a$ -th element  $\sum_{d \subseteq a} A_{a,d}^2$ , which indicates a value between 0 and 1; then, the  $a$ -th element of  $\bar{\boldsymbol{\varepsilon}} \sim N(\bar{\mathbf{0}}, \sigma^2 \mathbf{A}\mathbf{A}^t)$  in Eq. (5),  $\bar{\varepsilon}_a$ , is expressed as

$$E[\bar{\varepsilon}_a] = 0 \quad \text{Var}[\bar{\varepsilon}_a] = \sigma^2 \sum_{d \subseteq a} A_{a,d}^2. \quad (8)$$

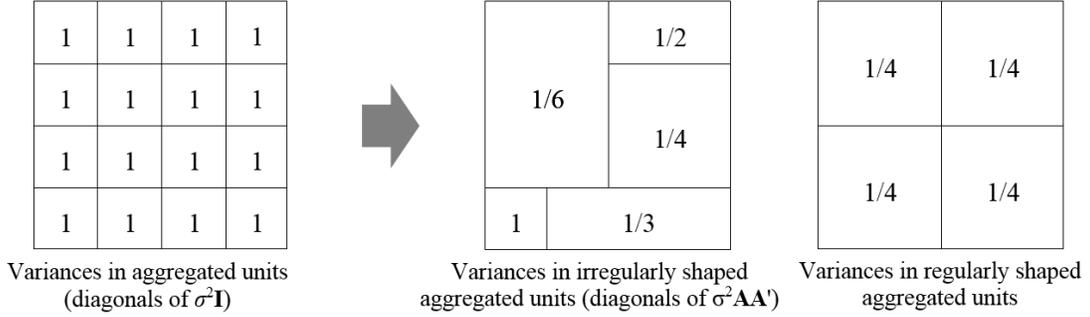


Figure 1: Variance deflation due to aggregation

Eq. (8) indicates that aggregate-level GWR evaluates the variances in each aggregated unit by deflating the disaggregate-level variance,  $\sigma^2$ , with  $\sum_{d \subseteq a} A_{a,d}^2$ . The deflation is large (i.e.,  $\sum_{d \subseteq a} A_{a,d}^2$  is small) for the aggregated units including many disaggregated units. In addition, as illustrated in Figure 1, the deflations are uniform (not uniform) when the aggregated units are regularly (irregularly) shaped. Thus, our model explicitly considers the variance deflation caused by aggregation, which is a principal source of the MAUP.

Moreover, the GWR model in Eq. (4) can mitigate the influence of spatial dependence, which is another principal source of the MAUP, especially the scale problem (see §1). Scale dependency is introduced by delineating the spatially dependent process and, accordingly, zoning (or delineation) optimization is a potential approach to mitigate scale dependency (e.g., Openshaw and Rao 1994). Kernel-based local weighting in GWR can be interpreted as fuzzy zoning or sliding neighborhoods (Páez et al. 2008). Specifically, GWR estimates  $\beta_s$  by using geographically weighted variables in a fuzzy zone, whose zone size is calibrated by  $b$ . As suggested by Fotheringham et al. (2002), calibrated fuzzy zoning depicts the spatial-dependent patterns related to the MAUP, and mitigate scale dependency.

### 3.2. Parameter estimation

The aggregate-level model, Eq. (5), is identical to the standard GWR model whose variances are scaled by the diagonals of  $\mathbf{AA}^t$ ,  $\sum_{d \subseteq a} A_{a,d}^2$ . Hence, the estimator of  $\beta_d$  is given, as with the standard GWR, as

$$\hat{\beta}_d = (\bar{\mathbf{X}}^t \bar{\mathbf{G}}_d^{1/2} (\mathbf{AA}^t)^{-1} \bar{\mathbf{G}}_d^{1/2} \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}^t \bar{\mathbf{G}}_d^{1/2} (\mathbf{AA}^t)^{-1} \bar{\mathbf{G}}_d^{1/2} \bar{\mathbf{y}}. \quad (9)$$

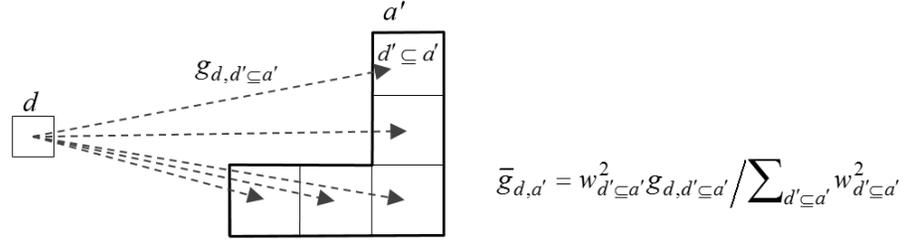


Figure 2: Image of spatial connectivity between  $d$  and  $a'$ :  $\bar{g}_{d,a'}$ .  $\bar{g}_{d,a'}$  considers shapes of aggregated units utilizing  $g_{d,d' \subseteq a'}$ , which are described by arrows.

Suppose that  $\mathbf{G}_d$  is a diagonal matrix whose  $d'$ -th element is  $g_{d,d'}$ , which is defined by the function of distance between the disaggregated units  $d$  and  $d'$ , while  $\bar{\mathbf{G}}_d = \mathbf{A}\mathbf{G}_d\mathbf{A}'$  is a diagonal matrix whose  $a'$ -th diagonal  $\bar{g}_{d,a'} = w_{d' \subseteq a'}^2 g_{d,d' \subseteq a'} / \sum_{d' \subseteq a'} w_{d' \subseteq a'}^2$  is a weighted average of the spatial connectivity between the  $d$ -th disaggregated unit and disaggregated units in the  $a'$ -th aggregated unit,  $g_{d,d' \subseteq a'}$ , and  $\bar{\mathbf{G}}_d^{1/2}$  is a diagonal matrix whose diagonals are the square roots of the diagonals of  $\bar{\mathbf{G}}_d$ . As shown in Figure 2, this specification enables us to consider the shapes of the aggregated units. In short, Eq. (9) is a weighted least squares estimator that considers variance deflations by aggregations with  $\mathbf{A}\mathbf{A}'$  (see Section 3.1) and the shapes of the aggregated units with  $\bar{\mathbf{G}}_d$ .

Our aggregate-level model, Eq. (5), is identical to standard GWR. Therefore, the variance–covariance matrix of  $\hat{\boldsymbol{\beta}}_d$  is given as (see Fotheringham et al. 2002)

$$\text{Cov}[\hat{\boldsymbol{\beta}}_d] = \hat{\sigma}^2 \mathbf{V}_d \mathbf{V}_d', \quad (10)$$

$$\mathbf{V}_d = (\bar{\mathbf{X}}' \bar{\mathbf{G}}_d^{1/2} (\mathbf{A}\mathbf{A}')^{-1} \bar{\mathbf{G}}_d^{1/2} \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}' \bar{\mathbf{G}}_d^{1/2} (\mathbf{A}\mathbf{A}')^{-1} \bar{\mathbf{G}}_d^{1/2},$$

where  $\hat{\sigma}^2$  denotes the estimates of  $\sigma^2$ . By substituting Eq. (9) into Eq. (5), the fitted values of  $\bar{\mathbf{y}}$  are given by  $\hat{\bar{\mathbf{y}}} = \mathbf{A}\mathbf{L}\bar{\mathbf{y}}$ , where  $\mathbf{L}$  is a matrix whose  $d$ -th row is  $\mathbf{x}'_d \mathbf{V}_d$  and  $\mathbf{x}_d$  is a vector of the explanatory variables observed in the  $d$ -th disaggregated unit.  $\hat{\sigma}^2$  is given as (see Cressie 1998)

$$\hat{\sigma}^2 = \frac{(\bar{\mathbf{y}} - \mathbf{A}\hat{\mathbf{y}})'(\bar{\mathbf{y}} - \mathbf{A}\hat{\mathbf{y}})}{\text{tr}\{(\mathbf{I} - \mathbf{A}\mathbf{L})(\mathbf{I} - \mathbf{A}\mathbf{L})'\}}, \quad (11)$$

where  $\text{tr}\{\bullet\}$  is a trace operator and  $\hat{\mathbf{y}}$  is a vector whose  $d$ -th element is  $\mathbf{x}'_d \hat{\boldsymbol{\beta}}_d$ . The significance of  $\boldsymbol{\beta}_d$  can be tested by using the diagonal elements of Eq. (7).

$\hat{\boldsymbol{\beta}}_d$  is estimated by calibrating the bandwidth parameter  $b$  in  $\bar{\mathbf{G}}_d$  and substituting the estimated bandwidth into Eq. (9). The calibration above can be performed by applying  $m$ -fold cross-validation in the following five steps: (a) the elements in  $\bar{\mathbf{y}}$  are randomly divided into  $m$  subsets; (b) under a given  $b$ ,  $1/m$  subsamples of  $\bar{\mathbf{y}}$  are predicted by using the remaining  $(m-1)/m$  subsamples; (c) step (b) is performed for all  $m$  cases; (d) the aggregate-level squared error ( $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$  in Eq. (12)) is evaluated; and (e) steps (a) to (d) are iterated by varying  $b$ , and the optimal  $b$ ,  $\hat{b}$ , satisfying Eq. (12) is identified:

$$\hat{b} = \underset{b}{\text{arcmmin}} \left[ \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \right] \quad \hat{\boldsymbol{\varepsilon}} = (\mathbf{A}\mathbf{A}' )^{-1/2} (\bar{\mathbf{y}} - \hat{\mathbf{y}}). \quad (12)$$

$\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{y}}$ , where  $\hat{\mathbf{y}}$  is a vector whose  $d$ -th element is  $\mathbf{x}'_d \hat{\boldsymbol{\beta}}_{d(4/5)}$ . Further,  $\mathbf{x}_d$  is a vector of the explanatory variables in the  $d$ -th disaggregated unit and  $\hat{\boldsymbol{\beta}}_{d(4/5)}$  is given based on Eq. (9) as

$$\hat{\boldsymbol{\beta}}_{d(4/5)} = (\bar{\mathbf{X}}_{(4/5)}' \bar{\mathbf{G}}_{d(4/5)}^{1/2} (\mathbf{A}\mathbf{A}')_{(4/5)}^{-1} \bar{\mathbf{G}}_{d(4/5)}^{1/2} \bar{\mathbf{X}}_{(4/5)}^t)^{-1} \bar{\mathbf{X}}_{(4/5)}^t \bar{\mathbf{G}}_{d(4/5)}^{1/2} (\mathbf{A}\mathbf{A}')_{(4/5)}^{-1} \bar{\mathbf{G}}_{d(4/5)}^{1/2} \bar{\mathbf{y}}_{(4/5)}, \quad (13)$$

where  $\bar{\mathbf{X}}_{(4/5)}$  is  $\bar{\mathbf{X}}$  whose 1/5 rows are discarded,  $\bar{\mathbf{y}}_{(4/5)}$  is  $\bar{\mathbf{y}}$  whose 1/5 elements are discarded, and  $\bar{\mathbf{G}}_{d(4/5)}$  and  $(\mathbf{A}\mathbf{A}')_{(4/5)}$  are  $\bar{\mathbf{G}}_d$  and  $\mathbf{A}\mathbf{A}'$  whose 1/5 rows and columns are discarded. The discarded 1/5 elements must include the  $d$ -th element.

Unlike standard GWR, which does not consider aggregation mechanisms, the elements of  $\hat{\boldsymbol{\beta}}_d$  denote the disaggregate-level spatially varying parameters (note that the parameters  $\boldsymbol{\beta}_d$  in Eqs. (4) and (5) are identical). We term our approach ATP GWR, after ATP kriging (Kyriakidis 2004), a closely related geostatistical approach to COSP.

## 4. Simulation study

### 4.1. Simulation design

This section examines the effectiveness of ATP GWR for mitigating the 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; Páez et al. 2011). For example, Páez 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 us to control collinearity among the spatially varying parameters, which is a critical factor that determines the effectiveness of GWR (Wheeler and Tiefelsdorf 2005). The second approach assumes the presence of spatially varying parameters that obey spatial processes,  $\boldsymbol{\beta}_d \sim N(\mathbf{0}, \tau^2 \mathbf{C})$  (e.g., Finley 2011), where  $\mathbf{C}$  is a covariance matrix whose elements are parameterized by a distance-decay function  $c(h_{d,d'})$  and  $\tau^2$  is a variance parameter.  $c(h_{d,d'})$  can be modeled by using the following Gaussian covariance function:

$$c(h_{d,d'}) = \exp\left(-\frac{h_{d,d'}^2}{r^2}\right), \quad (14)$$

where  $r$  is a parameter called range. This function is consistent with the Gaussian kernel function (Eq. (3)) used in GWR. Unlike the eigenvector-based approach, it enables us to control the spatial scales of the spatially varying parameter distributions by tuning  $r$ . Spatial scale is an essential factor that determines the seriousness of MAUP. Besides, the impacts of collinearity on GWR have already been studied extensively by Wheeler and Tiefelsdorf (2005) and Páez et al. (2011). Hence, we conduct a simulation study of the latter type, while paying attention to the collinearity among the spatially varying parameters.

In our simulation, we first generate the disaggregate-level explained variables and explanatory variables on  $50 \times 50$  grids. The explanatory variables include one intercept and two variables,  $x_{1,d}$  and  $x_{2,d}$ , generated independent of  $N(0,1)$ , respectively. The explained variables are generated from Eq. (15):

$$y_d = \alpha_d + x_{1,d}\beta_{1,d} + x_{2,d}\beta_{2,d} + \varepsilon_d \quad \varepsilon_d \sim N(0, \sigma^2), \quad (15)$$

Table 1: Shares of variances of each component (%)

	$\alpha_d$	$x_{1,d}\beta_d$	$x_{2,d}\beta_{2,d}$	$\varepsilon_d$
Case 1: $\sigma^2=1.0$	36.4	9.10	36.4	18.2
Case 1: $\sigma^2=4.0$	23.5	5.89	23.5	47.1

where  $\alpha_d$ ,  $\beta_{1,d}$ , and  $\beta_{2,d}$  are the spatially varying parameters generated by using Gaussian processes whose covariance functions are shown in Eq. (14), where the  $\tau^2$  values for both  $\alpha_d$  and  $\beta_{2,d}$  are 2.0 and that for  $\beta_{1,d}$  is 0.5. The intercept and  $x_{2,d}$ , which correspond to  $\tau^2=2.0$ , explain  $y_d$  effectively, whereas  $x_{1,d}$  does not. Our simulations are performed for six cases, by altering  $\sigma^2 = \{1.0, 4.0\}$ ,  $r = \{5, 10, 20\}$ .

The sum of two independent normal distributions  $N(0, \sigma_1^2)$  and  $N(0, \sigma_2^2)$  obeys  $N(0, \sigma_1^2 + \sigma_2^2)$ , and their product is approximated by  $N(0, \sigma_1^2\sigma_2^2)$  (Ware and Lad 2003); by using this property, we summarize the variance shares of each term in Eq. (15) and present the results in Table 1. This simulation examines whether our approach properly identifies the spatially varying effects, whose variance shares are 0.364, 0.235, 0.091, and 0.059, respectively. On the contrary, the assumed values of  $r$  indicate that the effective ranges (e.g., Cressie 1993) at which 95% of spatial dependence vanishes are set at 8.66, 17.32, and 34.64.<sup>2</sup> Thus, our simulation covers the likely ranges of values of both the contribution ratios and the distance ranges.

In the simulation, we first generate the true distributions of  $\alpha_{i,d}$ ,  $\beta_{1,d}$ , and  $\beta_{2,d}$  for each of the six ( $\sigma^2 = \{1.0, 4.0\} \times r = \{5, 10, 20\}$ ) cases (the true distributions when  $\sigma^2 = 1.0$  and  $r = 5$  or 20 are shown in Figure 3). Then, in each of these six cases, the following steps are iterated 100 times: (i) the disaggregate-level variables  $x_{1,d}$ ,  $x_{2,d}$ , and  $y_d$  are generated; (ii) they are aggregated into  $A$  aggregated units, which are generated by using the Voronoi tessellation; (iii) the disaggregate-level parameters,  $\alpha_{i,d}$ ,  $\beta_{1,d}$ , and  $\beta_{2,d}$ , are estimated by using Eq. (5), which is defined by the aggregated variables; and (iv) the accuracies of these estimates are measured by comparing them with their true values. The Voronoi tessellation is used in step (ii) because it is a standard partitioning algorithm whose use seems to increase transparency of our simulation. Besides, Nagle et al. (2011) apply the Voronoi tessellation in a similar simulation. If the aggregate-level model (Eq. (5)) effectively recovers the disaggregate-level parameters irrespective of the aggregated units, we can say that the method is

<sup>2</sup> The effective range of the Gaussian function is defined by  $r\sqrt{3}$ .

robust to the MAUP. Considering Following the suggestion of Páez et al. (2011) that the data applied for GWR should not be small, we have 2,500 grids (i.e., 50×50 grids) aggregated into 400 units.

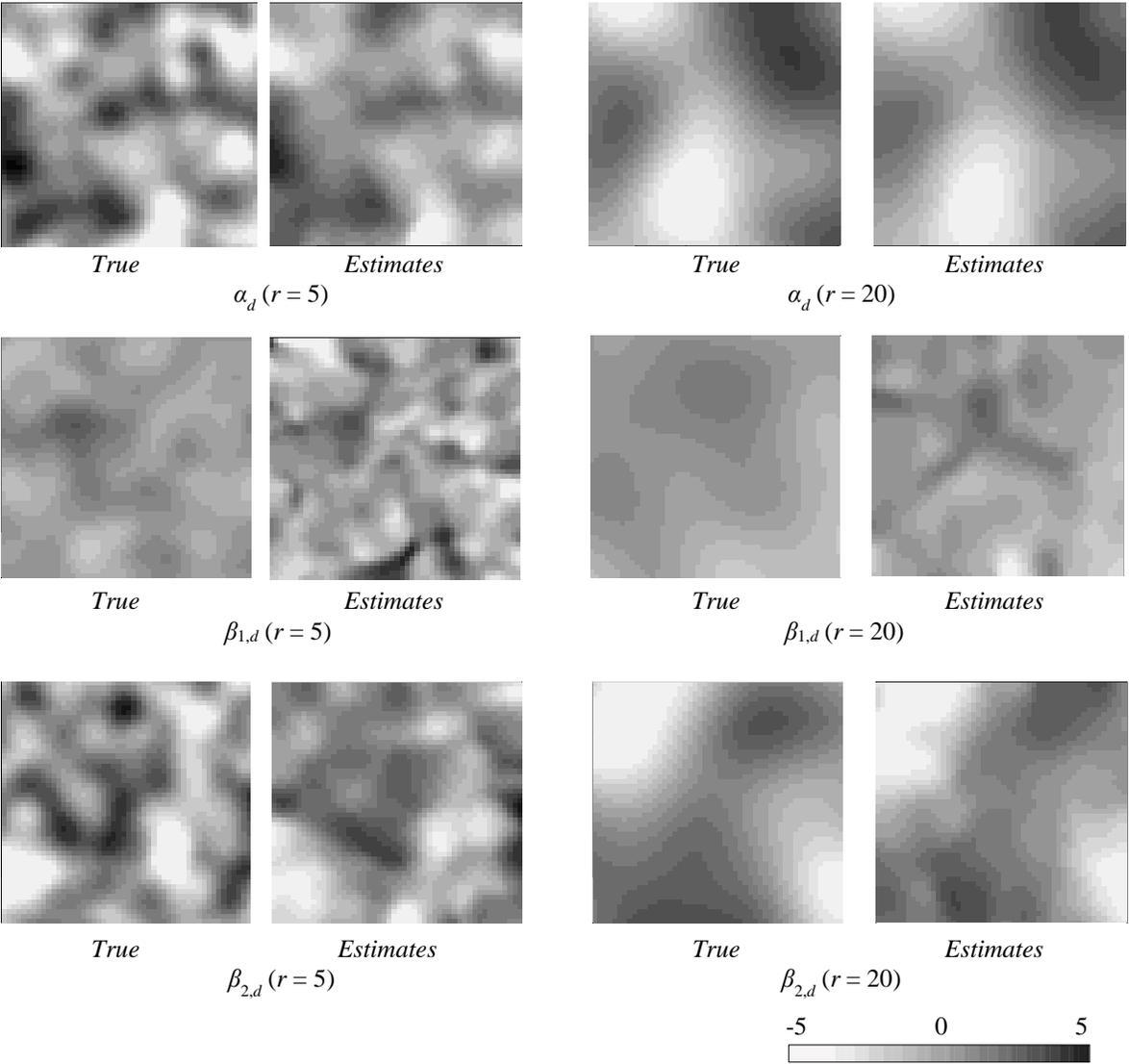


Figure 3: Plots of estimated  $\alpha_d$ ,  $\beta_{1,d}$ , and  $\beta_{2,d}$ , and their true values when  $\sigma^2 = 1.0$

Regular grids (e.g., 50×50 grids) are rarely used in GWR simulation studies, as they do not represent real-world geographical topologies (Farber et al. 2009). However, since the objective of COSP simulation studies is to mitigate the influences of spatial supports (shape, size, etc.), most such studies discuss the modeling of continuous spatial processes, which are free from such spatial supports, and approximate the

continuous processes by using a discrete spatial process on regular grids (e.g., Kyriakidis and Yoo 2005; Nagle et al. 2011). Thus, our assumption of 50×50 grids is consistent with standard assumptions in COSP studies.

## 4.2. Results

The estimates of  $\alpha_{i,d}$ ,  $\beta_{1,d}$ , and  $\beta_{2,d}$  given in each of the first attempts with  $\sigma^2 = 1.0$  and  $r = \{5, 20\}$  are plotted in Figure 3. The results seem to indicate that (i) the estimates of  $\alpha_d$  and  $\beta_{2,d}$ , which explain  $y_d$  well, are more accurate than those of  $\beta_{1,d}$  and (ii) the estimates are accurate when the variations of the spatially varying parameters are global.

In the next step, we assess the accuracy of the parameter estimates by using the root mean square error (RMSE) and R-square ( $R^2$ ) measures. Figure 4 plots the average RMSEs and  $R^2$  values evaluated in each attempt as well as those values obtained from disaggregate-level standard GWR (GWR\_DA) for comparison purposes. Note that since ATP GWR is an aggregate-level model, the RMSEs ( $R^2$  values) of the estimates must be larger (smaller) than those of GWR\_DA. Figure 4 shows that the RMSEs and  $R^2$  values in our method change significantly depending on  $r$  and that they perform poorly when  $r$  is small. The poor performance occurs because when  $r$  is small,  $y_d$  has local scale or highly frequent spatial variations within each aggregated unit (i.e., the within-aggregated unit variances are large), which disappear following aggregation. The RMSEs and  $R^2$  values also change depending on the explanation capabilities of the explanatory variables. Specifically, the average  $R^2$  values of  $\alpha_d$  and  $\beta_{2,d}$ , which explain  $y_d$  well, change between 0.4 and 1.0, whereas those of  $\beta_{1,d}$  change between 0.1 and 0.6. By contrast, the impact of  $\sigma^2$  is relatively small. In summary, ATP GWR provides accurate disaggregate-level parameter estimates if the parameters have significant impacts and their spatial variations are not too local compared with their aggregation scales.

The standard deviations of the RMSEs and  $R^2$  values shown in Figure 4 indicate that the accuracy of the parameter estimates is not sensitive to changes in the aggregated units, again under the condition that these parameters explain  $y_d$  well and  $r$  is not small.

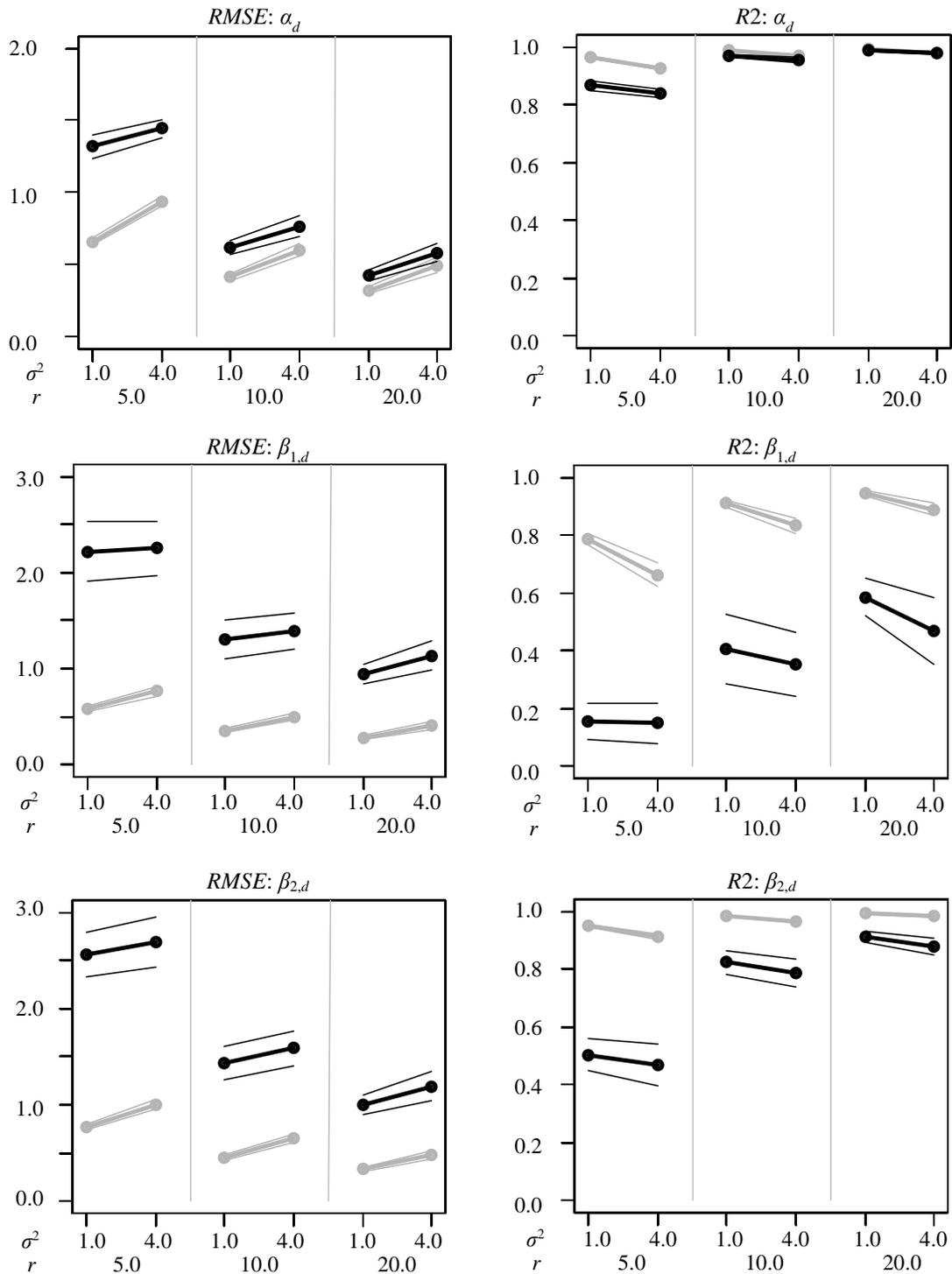


Figure 4: RMSEs and R2s of the estimates of  $\alpha_d$ ,  $\beta_{1,d}$ , and  $\beta_{2,d}$ . Here, the averages of the RMSEs and R2s are plotted (black line: ATP GWR; gray line: GWR\_DA). The bold lines represent averages, and the gaps between the bold lines and the thin lines around them represent the standard deviations of the RMSEs or R2s.

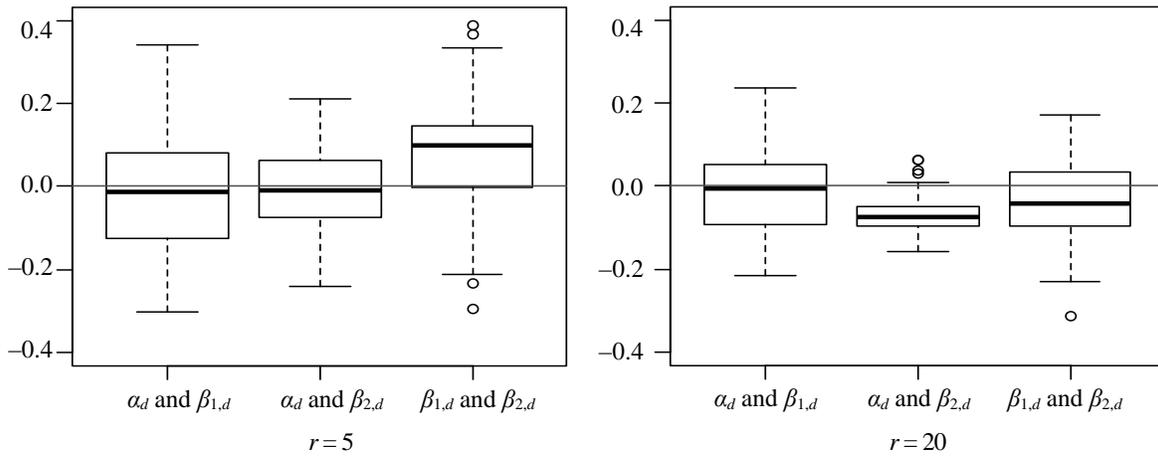


Figure 5: Correlation coefficients among spatially varying parameters ( $\sigma^2 = 1.0$ )

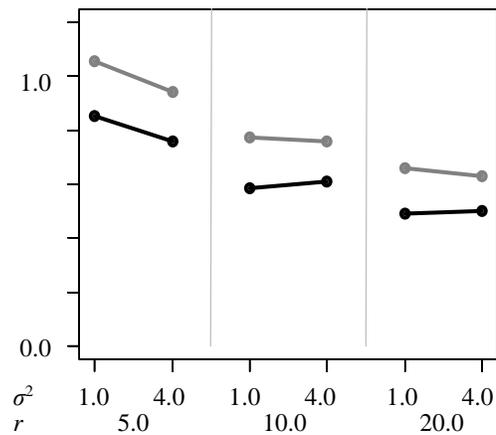


Figure 6: RMSEs of the bandwidth parameter estimates (black line: ATP GWR; dark gray line: GWR\_A). Here, the bandwidth parameter estimates of GWR\_DA are considered as the true values, and gaps between the true values and the estimates of ATP GWR and GWR\_A are evaluated using RMSE.

The estimates of GWR are easily collinear even when the explanatory variables are uncorrelated (Wheeler and Tiefelsdorf 2005; Wheeler 2007); therefore, it is important to examine whether multicollinearity is present in our estimates. The correlations among the estimated parameters when  $\sigma^2 = 1.0$  and  $r = \{5, 20\}$  summarized in Figure 5 suggest that no serious multicollinearity issues arise in our simulation.

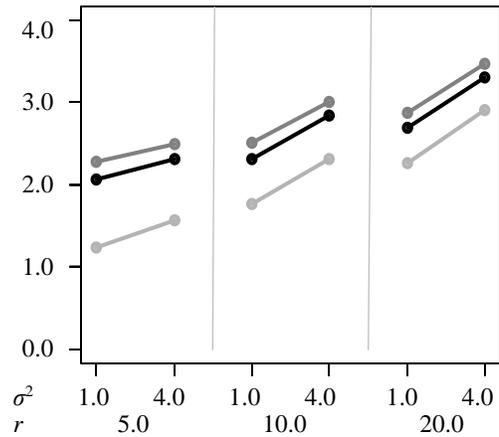


Figure 7: Averages of the bandwidth parameter estimates (black line: ATP GWR; dark gray line: GWR\_A; light gray line: GWR\_DA).

We then compare the bandwidth parameter estimates for two aggregate-level models: ATP GWR and aggregate-level standard GWR (GWR\_A). GWR\_A is an aggregate-level variable-based conventional GWR wherein the geometric centers of each aggregated unit are used to calculate the degree of spatial connectivity. Figure 6 plots the average RMSEs of their bandwidth parameters. Here, the estimates of GWR\_DA are regarded as their true values. As shown in this figure, on average, the estimates of ATP GWR are more accurate than those of GWR\_A in all cases. In each of the six cases, at least 91% of the attempts indicate that ATP GWR has greater efficiency than GWR\_A. However, the estimates of ATP GWR are still upwardly biased, and this bias is particularly prominent when  $r$  is small (see Figure 7). Future studies should thus aim to reduce this bias.

# 5. Empirical study

## 5.1. Outline

Building on the results of our simulation study presented in Section 4, we now turn to the empirical study. There are two important planning issues relating the elderly, who are typically vulnerable road users: moving them to areas with fewer hazard risks; and moving them to city centers, which allow them to achieve greater quality of life without relying on intensive car use. As a first step to tackle these issues, we apply both ATP GWR and GWR\_A to aging rate data on 249 municipalities in the Tokyo metropolitan area derived from the population Census of 2005 (Figure 8), and examine influences of flood risks and city center locations on aging rates. The explained variables are inundation depths under river flood (*Flood*) [m] and distance to the nearest station (*Station*) [km], with data taken from the National Land Numerical Information service. The coefficient of *Flood*,  $\beta(Flood)$ , quantifies to what extent the elderly live in flood-prone areas, while the coefficient of *Station*,  $\beta(Station)$ , quantifies to what extent the elderly reside in city centers. Here, ATP GWR estimates the parameters in 10,247 minor municipal districts by using the 249 samples, whereas GWR\_A estimates the parameters in 249 municipalities by using the same samples. Further, GWR\_DA, which estimates the parameters in minor municipal districts by using data on these districts, is also applied for comparison purposes.

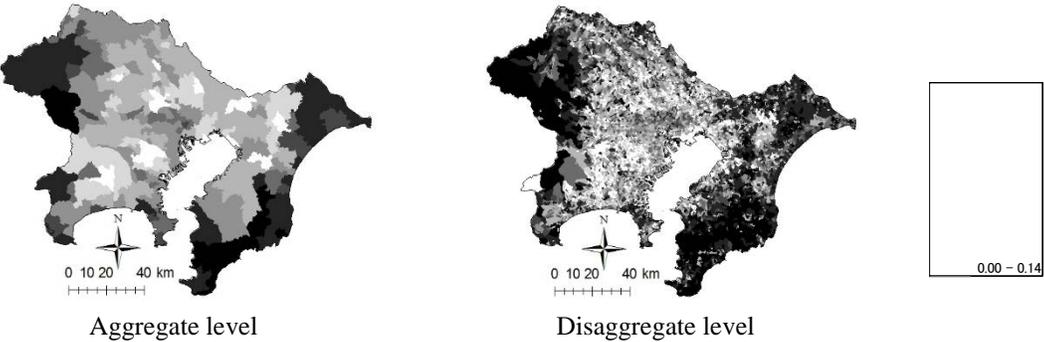


Figure8: Aging ratio

## 5.2. Results

The estimated bandwidth of GWR\_DA is 14.1 km, suggesting the existence of medium-scale spatial variation in disaggregate-level aging rates, whereas those of GWR\_A and ATP GWR are 20.0 and 17.9 km, respectively. Thus, as with the simulation results shown in the previous section, while both GWR\_A and ATP GWR overestimate the bandwidth, the estimate of ATP GWR is closer to that of GWR\_DA.

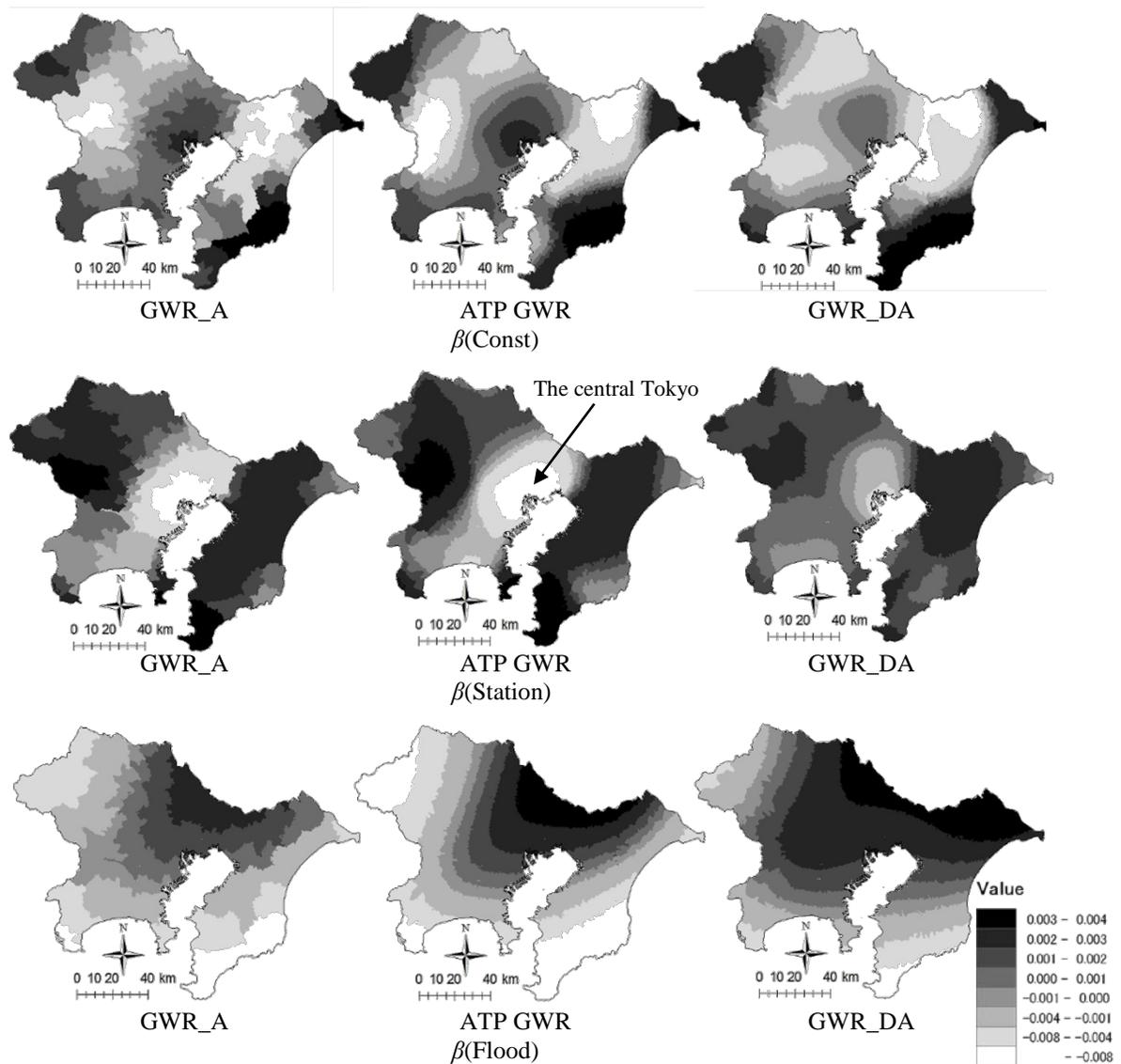


Figure 9: Parameter estimates

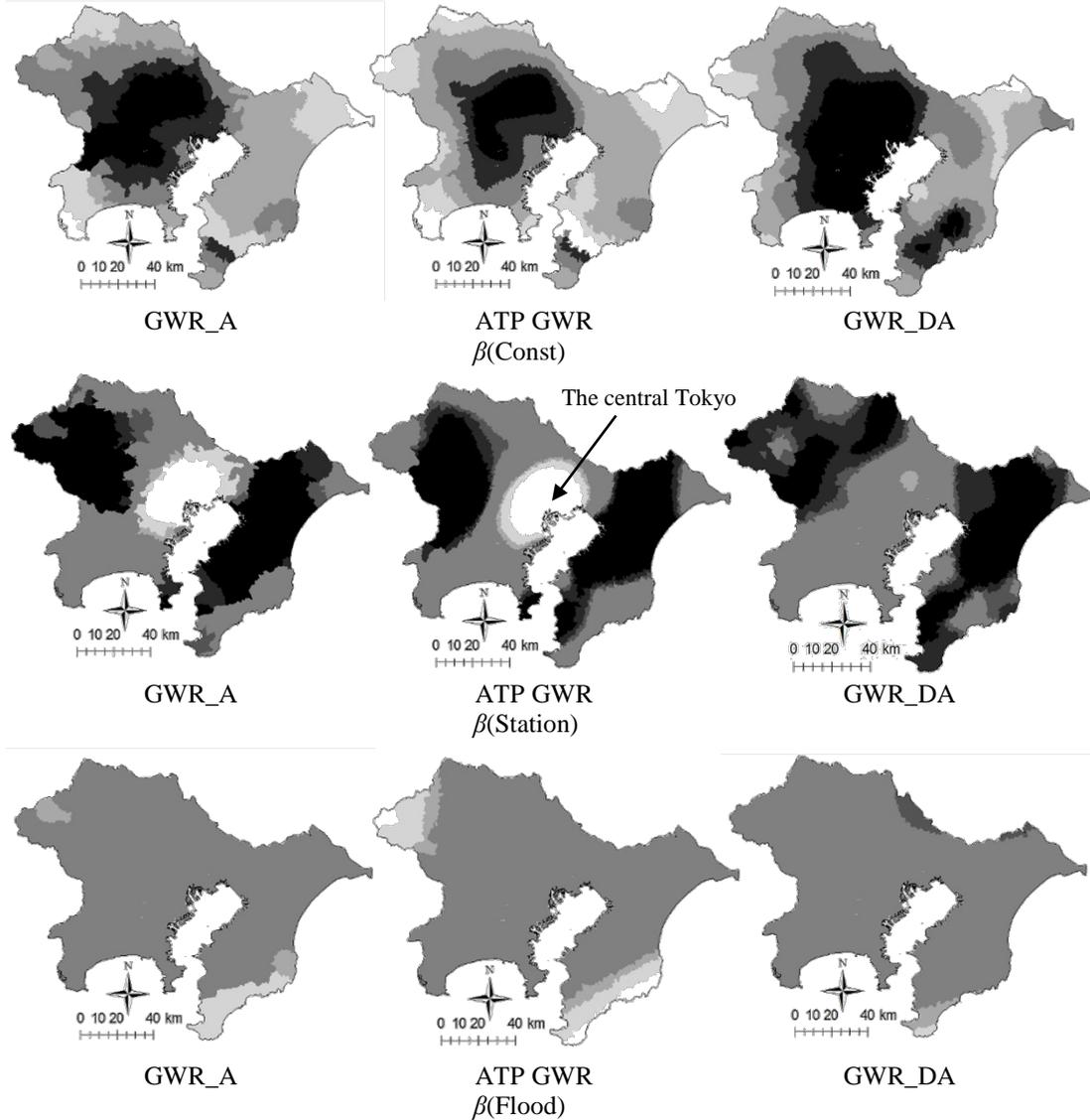


Figure 10: Significance of the parameter estimates: Since  $\beta(\text{Const})$  in each model is positively significant at the 1 % level across the target area, instead of the significance, the  $t$ -values are plotted.

Figure 9 plots  $\beta(\text{Const})$  (the coefficient of the constant),  $\beta(\text{Station})$ , and  $\beta(\text{Flood})$ . The greatest absolute correlation coefficient among these parameters is  $|-0.46| = 0.46$  (between  $\beta(\text{Const})$  and  $\beta(\text{Station})$ ), with no serious multicollinearity noted. The estimates of ATP GWR are similar to those of GWR\_DA, suggesting the robustness of our approach to the MAUP. However, the estimates of ATP GWR and GWR\_DA still show some differences; for example, the magnitudes of their  $\beta(\text{Station})$  values differ in the central Tokyo area. It shows one of the limitations of our approach that it cannot solve the MAUP, only

mitigate it. On the other hand, the presented estimates from the proposed method are spatially smoother than those of GWR\_A. Since GWR estimates are usually interpreted visually, providing spatially smooth estimates from aggregated variables that concur with the theory of COSP, would be an advantage of ATP GWR.

Figure 10 plots the significance levels of the estimated coefficients. The significance levels of ATP GWR are generally similar to those of GWR\_DA, again verifying the effectiveness of ATP GWR for addressing MAUP. Indeed, in each model,  $\beta(Const)$  is positively significant across all regions,  $\beta(Station)$  is negatively significant in suburban areas, and  $\beta(Flood)$  is not significant except for in certain border areas. The significance of  $\beta(Station)$  suggests that the elderly prefer to live outside city centers despite the inconveniences for vulnerable road users.

Some differences between ATP GWR and GWR\_DA are also found. For instance, in the central Tokyo area,  $\beta(Station)$  of ATP GWR is negatively significant at the 1% level, whereas  $\beta(Station)$  of GWR\_DA is negatively significant at the 10% level (at most) and the area is limited compared with ATP GWR. Such differences alter the interpretation of the results, which for ATP GWR now suggest that the elderly tend to prefer living close to stations (i.e., city centers) in the central Tokyo area. By contrast, the results derived from GWR\_DA suggest that the preference is not statistically significant.

## 6. Discussion

It is important to clarify the relationship between the approach presented in this paper and geostatistical approaches. In general, Gaussian geostatistical models (e.g., Kyriakidis 2004; Gotway and Young 2007; Gelfand 2010) of COSP, including the model of ATP kriging, can be expressed as

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad E[\boldsymbol{\varepsilon}] = \mathbf{0} \quad Var[\boldsymbol{\varepsilon}] = \tau^2 \mathbf{C} + \sigma^2 \mathbf{I}, \quad (16)$$

$$\bar{\mathbf{y}} = \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\varepsilon}} \quad E[\bar{\boldsymbol{\varepsilon}}] = \mathbf{0} \quad Var[\bar{\boldsymbol{\varepsilon}}] = \mathbf{A}(\tau^2 \mathbf{C} + \sigma^2 \mathbf{I})\mathbf{A}', \quad (17)$$

where  $\boldsymbol{\mu}$  is a vector of the trend components,  $\mathbf{C}$  is a matrix whose elements are given by a covariance function (e.g., Eq. (14)), and  $\bar{\boldsymbol{\mu}} = \mathbf{A}\boldsymbol{\mu}$ . Eq. (16) describes the spatially dependent disaggregate-level process by modeling its error term using  $\sigma^2$ , which measures non-spatial variation,  $\tau^2$ , which measures spatial variation,

and  $r$  in  $\mathbf{C}$ , which is a range parameter. Eq. (17) describes the aggregate-level spatial process by aggregating the disaggregate-level process. Our models defined by Eqs. (4) and (5) can be obtained from Eqs. (16) and (17) by replacing  $\boldsymbol{\mu}$  and  $\mathbf{C} + \sigma^2 \mathbf{I}$  with  $\mathbf{X}\boldsymbol{\beta}_d$  and  $\sigma^2 \mathbf{I}$ , respectively. Thus, our model is strongly related to geostatistical models.

The effects of  $\sigma^2 \mathbf{I}$  and  $\mathbf{C}$  are indistinguishable, while  $\sigma^2$  in the geostatistical model cannot be estimated properly (Goovaerts 2008). Hence, our approach might be preferable for making statistical inferences that require  $\sigma^2$ . Further, whereas typical geostatistical approaches assume  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ , where  $\boldsymbol{\beta}$  is a vector of the trend parameters, our approach allows the trend parameters to vary across space. On the contrary, our approach ignores spatial dependence, whereas geostatistical approaches do not. Because ignorance can degrade the validity of statistical inferences (e.g., LeSage and Pace 2009), the consideration of spatial dependence is an important task for future studies. Combining ATP GWR and ATP kriging is one possible solution to this problem (see also Harris et al. 2011).

## 7. Concluding remarks

This study developed ATP GWR, a type of GWR with aggregation mechanisms. From the viewpoint of the MAUP, ATP GWR is superior to standard GWR in that it explicitly estimates disaggregate-level parameters, which are essentially free from aggregation mechanisms. In a simulation study, we confirmed that the method is effective for estimating parameters in the face of the MAUP and discussed its practical applications.

Our method presents three main limitations. First, the simulation study in Section 4 indicates the ineffectiveness of our method when true spatially varying parameters have local spatial patterns. Fisher and Langford (1995) find that disaggregate-level local spatial patterns can be effectively captured by considering spatially fine auxiliary data (e.g., high-resolution land use data) in aggregation mechanisms. Thus, our method needs to be extended to consider spatially fine auxiliary data in  $\mathbf{A}$ . The second limitation is multicollinearity. As with standard GWR, ATP GWR seems to suffer from the issue of multicollinearity, particularly when the number of explanatory variables is large. Applying a penalized form of GWR, such as a geographically

weighted ridge regression or geographically weighted lasso regression model (Wheeler, 2007; 2009), might be useful to tackle this problem. The third problem is the ignorance of spatial dependence. The integration of ATP GWR and ATP kriging might be useful to overcome this problem.

In terms of future research avenues, while we extended GWR by combining geographical (i.e., GWR-based modeling) and geostatistical (i.e., aggregation-specific geostatistical) approaches, geographical MAUP studies and geostatistical COSP studies have developed practically independently thus far. The integration of these study areas would thus be an important step toward developing more sophisticated measures to tackle MAUP.

## Acknowledgments

We thank Dr. Haoying Wang (Pennsylvania State University, United States) for his helpful comments at the VII World Conference of the Spatial Econometrics Association. This work was supported by two Grants-in-Aid for Scientific Research from JSPS (Research Project Numbers: 23360219 and 24241053). Further, this paper fulfils the criteria of the Global Climate Risk Management Strategies (S10) project.

## Appendix 1: ATP GWR for the extensive variables

$\mathbf{A}$  must be determined by considering the volume-preserving property for the extensive variables. In other words,  $\bar{\mathbf{y}} = \mathbf{A}\mathbf{y}$  must hold under the condition that the elements in  $\mathbf{y}$  are extensive variables. For example, suppose that the population in an aggregated unit  $a$  is  $\bar{y}_a$  and unit  $a$  comprises two disaggregated units  $d$  and  $d'$ ; then,  $\bar{\mathbf{y}} = \mathbf{A}\mathbf{y}$  is expressed as

$$\bar{y}_a = \begin{bmatrix} A_{a,d} & A_{a,d'} \end{bmatrix} \begin{bmatrix} Y_d \\ Y_{d'} \end{bmatrix},$$

$$Y_d + Y_{d'} = A_{a,d}Y_d + A_{a,d'}Y_{d'}. \quad (\text{A-1})$$

Eq. (6) merely states that  $\bar{y}_a = Y_d + Y_{d'}$  must equal  $A_{a,d}Y_d + A_{a,d'}Y_{d'}$ . This is satisfied by defining  $A_d$  and  $A_{d'}$  as ones. In general,  $\bar{\mathbf{y}} = \mathbf{A}\mathbf{y}$  is fulfilled by defining  $A_{a,d}$  as follows:

$$A_{a,d} = \begin{cases} 1 & \text{if } d \subseteq a \\ 0 & \text{otherwise} \end{cases} \quad (\text{A-2})$$

ATP GWR for the extensive variables is thus defined by defining  $A_{a,d}$ , which was defined in Eq. (7), as shown in Eq. (A-2). In this case, the variance of the error term is scaled by the number of disaggregated units in each aggregated unit.

## Reference

- Berrocal VJ, Gelfand AE, Holland DM (2010) A spatio-temporal downscaler for outputs from numerical models. *J Agric Biol Environ Stat* 15(2):176–197
- Berrocal VJ, Gelfand AE., Holland DM (2012) Space-time data fusion under error in computer model output: an application to modeling air quality. *Biom* 68(3):837–848
- Butkiewicz T, Ross K (2010) Alleviating the modifiable areal unit problem within probe-based geospatial analyses. *Comput Graph Forum* 29(3):923–932
- Cressie, N (1996) Change of support and the modifiable areal unit problem. *Geogr Sys* 3(2-3):159–180
- Cressie N (1998) Fundamentals of spatial statistics. In: Müller WG (ed) *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*. Springer-Verlag, Berlin, pp.11–42
- Farber S, Páez A, Volz E (2009) Topology and dependency tests in spatial and network autoregressive models. *Geogr Anal* 41(2):158–180
- Finley AO (2011) Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods Ecol Evol* 2(2):143–154
- Fisher PF, Langford M (1996) Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation. *Prof Geogr* 48(3):299–309
- Fotheringham S, Brunson C, Charlton M (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley
- Fuentes M, Raftery AE (2005) Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biom* 61(1):36–45

- Gelfand AE. (2010) Misaligned spatial data: the change of support problem. In: Gelfand AE, Diggle PJ, Guttorp P, Fuentes M (eds) *Handbook of Spatial Statistics*. CRC Press, Boca Raton, pp.517–539
- Gelfand AE (2012) Hierarchical modeling for spatial data problems. *Spat Stat* 1: 30–39
- Goovaerts P (2008) Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathemat Geol* 40(1):101–128
- Gotway CA, Young LJ (2002) Combining incompatible spatial data. *J Am Stat Assoc* 97(458):632–648
- Gotway CA, Young LJ (2007) A geostatistical approach to linking geographically aggregated data from different sources. *J Comput Graph Stat* 16(1):115–135
- Haining RP, Kerry R, Oliver MA (2010) *Geography, spatial data analysis, and geostatistics: An overview*. *Geogr Anal* 42(1):1–31
- Harris P, Brunson C, Fotheringham AS (2011) Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stoch Environ Res Risk Assess* 25(2):123–138
- Kyriakidis PC (2004) A geostatistical framework for area-to-point spatial interpolation. *Geogr Anal* 36(3):259–289
- Kyriakidis PC, Yoo E-H (2005) Geostatistical Prediction and Simulation of Point Values from Areal Data. *Geogr Anal* 37(2):124–151
- Lam N-S (1983) Spatial Interpolation Methods: A Review. *Am Cartogr* 10(2):129–149
- Lee SJ, Yeatts KB, Serre ML (2009) A Bayesian maximum entropy approach to address the change of support problem in the spatial analysis of childhood asthma prevalence across North Carolina. *Spat Spatiotemporal Epidemiol* 1(1):49–60
- LeSage JP, Pace RK (2009) *Introduction to Spatial Econometrics*. Chapman and Hall/CRC
- Nagle NN, Sweeney SH, Kyriakidis PC (2011) A geostatistical linear regression model for small area data. *Geogr Anal* 43(1):38–60
- Odoi A, Martin W, Michel P, Holt J, Middleton D, Wilson J (2003) Geographical and temporal distribution of human giardiasis in Ontario Canada. *Int J Health Geogr* 2:5
- Openshaw S (1984) *The modifiable areal unit problem*. Norwich, UK, Geo Books

- Openshaw S, Rao L (1994) Algorithms for reengineering 1991 Census geography. *Environ Plan A* 27(3):425-446
- Openshaw S, Taylor P (1979) A Million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley N (ed) *Statistical Methods in the Spatial Sciences*. Pion, London, pp.127–144
- Páez A, Farber S, Wheeler D (2011) A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environ Plan A* 43(12):2992–3010
- Páez A, Long F, Farber S (2008) Moving window approaches for hedonic price estimation: An empirical comparison of modeling techniques. *Urban Stud* 45(8):1565–1581
- Páez A, Uchida T, Miyamoto K (2002) A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environ Plan A* 34(4):733–754
- Reynolds HD (1998) The modifiable areal unit problem: empirical analysis by statistical simulation. Thesis, University of Toronto
- Sahu SK, Gelfand AE, Holland DM (2010) Fusing point and areal level space–time data with application to wet deposition. *J R Stat Soc C Appl Stat* 59(1):77–103
- Schabenberger O, Gotway CA (2005) *Statistical Methods for Spatial Data Analysis*. Chapman and Hall/CRC
- 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(11):825–833
- Swift A, Liu L, Uber J (2008) Reducing MAUP bias of correlation statistics between water quality and GI illness. *Comput Environ Urban Systems* 32(2):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. *Geogr Anal* 34(1):1–19
- Tranmer M, Steel D (1998) Using census data to investigate the causes of the ecological fallacy. *Environ Plan A* 30(5):817–831
- Vichiensan V, Páez A, Kawai K, Miyamoto K (2006) Nonstationary spatial interpolation method for urban model development. *Transp Res Rec* 1977:103–111

- Ware R, Lad F (2003) Approximating the distribution for sums of products of normal variables, University of Canterbury
- Wheeler D (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environ Plan A* 39(10):2464–2481
- Wheeler D (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environ Plan A* 41(3):722–742
- Wheeler D (2010) Visualizing and diagnosing coefficients from geographically weighted regression. In: Jiang B, Yao X (eds) *Geospatial Analysis and Modeling of Urban Structure and Dynamics*. Springer, Netherlands, pp.415–436
- Wheeler D. 2013. Geographically weighted regression. In: Fischer MM, Nijkamp P (eds) *Handbook of Regional Science*. Springer, Berlin, pp.1435–1459
- Wheeler D, Páez A (2010) Geographically weighted regression. In: Fischer MM, Getis A (eds) *Handbook of Applied Spatial Analysis*, Springer, Berlin, pp.461–486
- Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *J Geogr Sys* 7(2):161–187
- Wong D (2009) The modifiable areal unit problem (MAUP). In: Fotheringham AS, Rogerson PA (eds) *The SAGE Handbook of Spatial Analysis*, SAGE, London, pp.105–124
- Young LJ, Gotway CA (2007) Linking spatial data from different sources: the effects of change of support. *Stoch Environ Res Risk Assess* 21(5):589–600
- Young LJ, Gotway CA, Kearney G, Duclos C (2009a) Assessing uncertainty in support-adjusted spatial misalignment problems. *Commun Stat Theory Methods* 38(16-17):3249–3264
- Young LJ, Gotway CA, Yang J, Kearney G, Duclos C (2009b) Linking health and environmental data in geographical analysis: it's so much more than centroids. *Spat Spatiotemporal Epidemiol* 1(1):73–84