Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

21-10

Small Area Estimation under Spatial Nonstationarity

Hukum Chandra, Nicola Salvati, Ray Chambers and Nikos Tzavidis

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Small Area Estimation under Spatial Nonstationarity

Hukum Chandra[1], Nicola Salvati[2], Ray Chambers[3] and Nikos Tzavidis[4]

Abstract

In this paper a geographical weighted pseudo empirical best linear unbiased predictor (GWEBLUP) for small area averages is proposed, and two approaches for estimating its mean squared error (MSE), a conditional approach and an unconditional one, are developed. The popular empirical best linear unbiased predictor (EBLUP) under the linear mixed model and its associated MSE estimator are obtained as a special case of the GWEBLUP. Empirical results using both model-based and design-based simulations, with the latter based on two real data sets, show that the GWEBLUP predictor can lead to efficiency gains when spatial nonstationarity is present in the data. A practical gain from using the GWEBLUP is in small area estimation for out of sample areas. In this case the efficient use of geographical information can potentially improve upon conventional synthetic estimation.

Key words: Borrowing strength over space; Geographical weighted regression; Out of sample small area estimation; Spatial analysis.

[1]Centre for Statistical and Survey Methodology, University of Wollongong, Wollongong, NSW, 2522, Australia. E-mail: hchandra@uow.edu.au

[2] Dipartimento di Statistica e Matematica Applicata all'Economia, University of Pisa, Italy, E-mail: salvati@ec.unipi.it

[3]Centre for Statistical and Survey Methodology, University of Wollongong, Wollongong, NSW, 2522, Australia. Email: ray@uow.edu.au

[4]Social Statistics and S3RI, University of Southampton, United Kingdom. E-mail: n.tzavidis@soton.ac.uk

1. Introduction

Small area estimation is widely used for producing estimates of population parameters for areas (domains) with small, or even zero, sample sizes. In the case of small domain-specific sample sizes direct estimation that only relies on domain-specific observations may lead to estimates with large sampling variability (Rao, 2003). When direct estimation is not possible, one has to rely upon alternative model-based methods for producing small area estimates. One popular approach uses mixed (random) effects models for small area estimation (Fay and Herriot, 1979; Battese *et. al*., 1988). A mixed effects model consists of a fixed effects part and a random effects part with the latter accounting for between area variations beyond that explained by the auxiliary variables included in the fixed part of the model.

In small area estimation it is customary to assume that population units in different small areas are uncorrelated. However, in practice the boundaries that define a small area are arbitrarily set and hence there appears to be no good reason why population units that belong to neighbouring small areas and are close to the boundary between them should not be correlated. This may be the case, for example, with agricultural, environmental, economic and epidemiological data where units that are spatially close may be more related than units that are further apart, although they may belong to different small areas. It is therefore often reasonable to assume that the effects of neighbouring areas, defined by a contiguity criterion, are correlated. Extensions of the mixed effects model to allow for spatially correlated random effects using for example simultaneous autoregressive (SAR) models (Anselin, 1992) have been considered in the small area literature among others by Singh *et al.* (2005) and Pratesi and Salvati (2008). These models define the dependence between areas by using a contiguity matrix and allow for spatial correlation in the error structure while the fixed effects parameters are spatially invariant. SAR models offer only one possible way of borrowing strength over space. Alternative and potentially more flexible approaches, based on non-parametric extensions of the mixed effects model, have been also recently proposed in the small area estimation literature by Opsomer *et al*. (2008) and Ugarte *et al.* (2009).

An alternative approach for incorporating the spatial information in the model is by assuming that the regression coefficients vary spatially across the geography of interest. Models of this type can be fitted using geographical weighted regression (GWR), and are suitable for modelling spatial nonstationarity (Brunsdon *et al*., 1998, Fotheringham *et al*., 2002). The use of geographically weighted predictors in small area estimation has been only very recently investigated by Salvati *et al.* (2010) who proposed a GWR extension to predictors based on the M-quantile small area model (Chambers and Tzavidis, 2006). In the present paper we propose a similar extension to the widely used empirical best linear unbiased predictor or EBLUP that is often used for small area estimation under a linear mixed model. This is referred to below as the Geographical Weighted pseudo-Empirical Best Linear Unbiased Predictor or GWEBLUP, and is based on a mixed model that allows for spatially non-stationary linear fixed effects as well as random area effects. It is obtained by local linear fitting of a linear mixed model, using weights that are a function of the distance between the sample data points. Parameter estimation for the GWEBLUP is performed by extending the maximum likelihood estimation of the conventional linear mixed model in order to incorporate the geographical information contained in these distances.

The paper is organised as follows. In Section 2 we review the linear mixed model (LMM) and present the EBLUP of the small area average under this model. In Section 3 we present a spatially non-stationary extension to the LMM and define the GWEBLUP of the small area average under this model. MSE estimation for the GWEBLUP is considered in Section 4. In particular, two approaches for MSE estimation are discussed, a conditional approach that is based on the pseudo-linearization approach proposed by Chambers *et al.* (2009) and an unconditional approach which is similar in spirit to that of the Prasad and Rao (1990) MSE estimator. In Section 5 we discuss estimation for out of sample areas, i.e. small areas that contain no sample points. In Section 6 we empirically evaluate the performance of the GWEBLUP and of its associated MSE estimators using both model-based and design-based simulation studies, with the latter based on two real datasets. Finally, in Section 7 we conclude the paper with some summary comments.

2. Linear Mixed Effects Models for Small Area Estimation

Let us assume that the target population *U* of size *N* is made up of *A* non-overlapping small areas. We index the population units by *j* and the small areas by *i*. Each small area *i* contains a known number N_i of units. Let y_{ij} denote the value of the variable of interest *y* for unit *j* ($j = 1,\ldots,N_i$) in small area *i* ($i = 1,\ldots,A$) and let $\mathbf{x}_{ij}$ denote the vector of values of the *p* unit level auxiliary variables associated with this unit. Moreover, $\mathbf{z}_{ij}$ is a *q*-vector of auxiliary variables whose values are known for all units in the population. We also assume that there is a linear relationship between y_{ij} and $\mathbf{x}_{ij}$. A sample *s* of size *n* units is drawn from this population and n_i units belong to area *i*. That is, the total number of units in the population is $N = \sum_{i=1}^{A} N_i$, with corresponding total sample size $n = \sum_{i=1}^{A} n_i$. We also assume that the sample data are obtained via a non-informative sampling method. The aim is to use these data to predict the small area average of y. The most popular method used for this purpose employs linear mixed models (Rao, 2003). Let **y**, **X** and **Z** denote the population level vector and matrices defined by y_{ij}, $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$, respectively. Then,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{\beta}$ is a *p* vector of regression coefficients regression, $\mathbf{a} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ denotes a Aq-vector of area specific random effects and $\boldsymbol{\varepsilon} \sim (0, \sigma_\varepsilon^2 \mathbf{I}_N)$ is vector of *N* specific individual random errors with $\mathbf{I}_N$ the identity matrix of order *N*. In the simplest case, **Z** is given by a matrix whose *i*-th column, for $i = 1\ldots A$, is an indicator variable that takes the value 1 if a unit is in area *i* and is zero otherwise. The two error terms are assumed to be mutually independent, both across individuals as well as across areas, so that the covariance matrix of the vector **y** is given by

$$V(\mathbf{y}) = \mathbf{V}(\boldsymbol{\theta}) = \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}^T + \sigma_\varepsilon^2 \mathbf{I}_N,$$

where $\boldsymbol{\theta}$ are typically referred to as the variance components of (1).

Model (1) is a model both for sampled and non-sampled population units. It follows that we can partition $\mathbf{y}$, $\mathbf{X}$, $\mathbf{Z}$ and $\boldsymbol{\varepsilon}$ into components defined by the n sampled and N-n non-sampled population units, denoted by subscripts of s and r, respectively. We can therefore write (1) as follows:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_s \\ \mathbf{y}_r \end{bmatrix} = \begin{bmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_s \\ \mathbf{Z}_r \end{bmatrix} \mathbf{a} + \begin{bmatrix} \boldsymbol{\varepsilon}_s \\ \boldsymbol{\varepsilon}_r \end{bmatrix},$$

with variance of $\mathbf{y}$ given by

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{bmatrix}.$$

Thus, $\mathbf{X}_s$ represents the matrix defined by the n sample values of the auxiliary variable vector, while $\mathbf{V}_{rr}$ is the matrix of covariances of the response variable among the $N - n$ non-sampled units. We use subscript of i to denote restriction to small area i, so that s_i (r_i) denotes the set of sample (non-sample) population units from area i, and $U_i = s_i \cup r_i$ denotes the set of population units making up small area i. The variance components in (1) are estimated using Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) (see Harville, 1977). We use a "hat" to denote an estimated quality. Given the estimated values $(\hat{\boldsymbol{\Omega}}, \hat{\sigma}_\varepsilon^2)$ of the variance components we can obtain the estimated covariance matrix $\hat{\mathbf{V}}$ and the empirical best linear unbiased estimator (EBLUE) of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_s^T \hat{\mathbf{V}}_{ss}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \hat{\mathbf{V}}_{ss}^{-1} \mathbf{y}_s, \tag{2}$$

and the EBLUP of $\mathbf{a}$ is

$$\hat{\mathbf{a}} = \hat{\boldsymbol{\Omega}} \mathbf{Z}_s^T \hat{\mathbf{V}}_{ss}^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}}). \tag{3}$$

Under model (1), and using the estimated fixed and random effects, the estimator of the average of y in small area i is

$$\hat{m}_i^{EBLUP} = N_i^{-1} \left(\sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^T \hat{\mathbf{a}}_i \right). \tag{4}$$

Estimator (2) is commonly known as the EBLUP of m_i (Henderson, 1975; Rao, 2003).

3. Geographically Weighted Mixed Effects Models for Small Area Estimation

Under (1) we assume that the fixed effect parameters $\boldsymbol{\beta}$ are spatially invariant. There are situations, however, where the relationship between *y* and *x* is not constant over the study area, a phenomenon referred to as spatial nonstationarity. Geographical weighted regression (GWR) is a method that is widely used for fitting data exhibiting spatial nonstationarity (Brunsdon *et al*., 1998, Fotheringham *et al*., 2002). The model underpinning GWR is a local linear model, i.e. a linear model for the conditional expectation of *y* given *x* at location u_0. Salvati *et al.* (2010) have recently proposed an M-quantile extension of GWR for small area estimation and show that this approach represents a promising alternative for flexibly incorporating the available spatial information in small area estimation. Note that under GWR the data are assumed to follow a location specific or local regression function, with the geographical weights used for estimation of the parameters of this local regression function. In this Section we use the GWR concept to fit a local mixed model and we consider small area estimation under this model. In a slight abuse of notation, we refer below to this local mixed model as a geographically weighted linear mixed model (GWLMM). Let u_{ij} denotes the coordinates or the spatial location (longitude and latitude) of unit *j* in area *i*. The GWLMM is expressed as follows,

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}(u_{ij}) + \mathbf{z}_{ij}^T \mathbf{a}_i + \varepsilon_{ij}\,,\ j = 1,\ldots,N_i\,,\ i = 1,\ldots,A\,, \tag{5}$$

where $\boldsymbol{\beta}(u_{ij})$ is a parameter of *p* unknown fixed effects at location u_{ij}. Here $\mathbf{a}_i$ and ε_{ij} are the area-specific and individual-specific random errors, which are assumed to be normally distributed. That is, $\mathbf{a}_i \sim N(\mathbf{0},\boldsymbol{\Omega})$, $\varepsilon_{ij} \sim N(0,\sigma_e^2)$ and $\mathbf{a}_i$, ε_{ij} are assumed to be independent. At population level, on a point-wise basis model (5) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}(u) + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}\,. \tag{6}$$

The GWR method can be used to fit (6) by assigning a weight to every sample unit that depends on its distance from the location *u*. A similar partition of various quantities into sample and non-sample components as in Section 2 follows directly. Under the GWLMM (6) and following Henderson *et al.* (1959) we maximize the 'geographically weighted joint maximum likelihood' function (see Appendix 1 for the detailed development) to obtain the geographically weighted BLUE of $\boldsymbol{\beta}(u_{ij})$ at a location u_{ij} as

$$\tilde{\boldsymbol{\beta}}(u_{ij}) = \left(\mathbf{X}_s^T \mathbf{V}_{ss}^{-1}(u_{ij}) \mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \mathbf{V}_{ss}^{-1}(u_{ij}) \mathbf{y}_s, \tag{7}$$

and the geographically weighted pseudo-BLUP of the random area effects at a location u_{ij} as

$$\tilde{\mathbf{a}}(u_{ij}) = \boldsymbol{\Omega} \mathbf{Z}_s^T \mathbf{V}_{ss}^{-1}(u_{ij}) \left(\mathbf{y}_s - \mathbf{X}_s \tilde{\boldsymbol{\beta}}(u_{ij})\right), \tag{8}$$

where $\mathbf{V}_{ss}^{-1}(u_{ij}) = \left(\mathbf{Z}_s \boldsymbol{\Omega} \mathbf{Z}_s^T + \sigma_e^2 \mathbf{W}_s^{-1}(u_{ij})\right)^{-1}$. When unknown parameters in (7) and (8) are replaced by sample estimates, we refer to (7) as defining the Empirical BLUE (EBLUE) of $\boldsymbol{\beta}(u_{ij})$ and (8) as defining the geographically weighted pseudo-EBLUP of the random area effects at u_{ij}. It is worth noting that the GWR estimate of the function $\boldsymbol{\beta}(u)$ in Brunsdon *et al*. (1998) and Fotheringham *et al.* (2002) is a special case of (7) when the underlying model (6) does not include the term for the random area effect. As we shall see later, the predictor of the area effect is obtained by averaging these location specific predictors of area affects. Here $\mathbf{W}_s(u_{ij})$ is a matrix of weights that are specific to location *j* in area *i* such that observations nearer to location *j* are given greater weight than observations further away. This matrix is referred to as the spatial weighting matrix for location *j*. In particular, $\mathbf{W}_s(u_{ij}) = diag\left\{w_{ijk}; k = 1,...,n\right\}$, where w_{jk} is the weight given to observation k with respect to point j. There are various approaches to define w_{jk}. In this paper we use a Gaussian specification for defining such a weighting function,

$$\mathbf{W}_s(u_{ij}) = w_{ijk} = \exp\left\{-0.5(d_{j,k}/b)^2\right\}, \tag{9}$$

where $d_{j,k}$ denotes the Euclidean distance between point *j* and *k* and *b* is the bandwidth, which can be optimally defined using a least squares criterion (Fotheringham *et al*., 2002). As the distance between point *j* and *k* increases the spatial weight decreases exponentially. If *j* and *k* coincide, the weighting of the data at that point will be equal to one. If $w_{jk} = 0.5$ and $w_{lk} = 0.25$ then observations at location *j* have twice the weight in determining the fit at location *k* compared with observations at location *l*. That is, the weighting system is based on the concept of distance decay and it works by means of a weight function that reduces the influence of distant units in the estimation for location *j*. As mentioned above, the weighting function (9) depends on a unknown the bandwidth *b*. The bandwidth is a measure of how quickly the weighting function decays with increasing distance. For computing the bandwidth we use a cross validation (CV) procedure that mimimizes the following CV criterion

$$CV = \sum_{i=1}^{A} \left\{ \mathbf{y}_{s_i} - \hat{\mathbf{y}}_{s_{-i}}(b) \right\}^2 , \tag{10}$$

where $\hat{\mathbf{y}}_{s_{-i}}(b)$ is the vector of the predicted values of $\mathbf{y}_{s_i}$ using bandwidth *b* with the observations of area *i* omitted from the model fitting process. The value of *b* that minimizes (10) is then selected. It should be noted that alternative weighting functions, e.g. the bi-square function, can also be used. The results of GWR are relatively insensitive to the choice of weighting function but they are sensitive to the choice of bandwidth and hence obtaining the optimal value of the bandwidth is crucial.

Let m_i denote the area *i* average of the y_{ij}. Under (6), the geographically weighted EBLUP type predictor of m_i (Henderson, 1975; Rao, 2003) is then

$$\hat{m}_i^{GWEBLUP} = N_i^{-1} \left(\sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \left\{ \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}(u_{ij}) + \mathbf{z}_{ij}^T \hat{\mathbf{a}}_i(u_{ij}) \right\} \right) . \tag{11}$$

Here $\hat{\boldsymbol{\beta}}(u_{ij}) = \left(\mathbf{X}_s^T \hat{\mathbf{V}}_{ss}^{-1}(u_{ij}) \mathbf{X}_s \right)^{-1} \mathbf{X}_s^T \hat{\mathbf{V}}_{ss}^{-1}(u_{ij}) \mathbf{y}_s$ is the geographically weighted EBLUE of $\boldsymbol{\beta}$ at location *j* in area *i*, $\hat{\mathbf{a}}(u_{ij}) = \hat{\boldsymbol{\Omega}} \mathbf{Z}_s^T \hat{\mathbf{V}}_{ss}^{-1}(u_{ij}) \left(\mathbf{y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}}(u_{ij}) \right)$ is the geographically weighted pseudo-

EBLUP of $\mathbf{a}_i$ at location j in area i and $(\hat{\boldsymbol{\Omega}}, \hat{\sigma}_e^2)$ are the estimated values of the variance components. Predictor (11) under model (6) is a pseudo-EBLUP for the small area mean and we refer to it as the geographical weighted empirical best linear unbiased predictor or GWEBLUP. Note that computation of (11) requires estimates for $\hat{\boldsymbol{\beta}}(u_{ij})$ and $\hat{\mathbf{a}}(u_{ij})$ for $j \in r_i$. We distinguish two cases:

(i) When the spatial locations or the coordinates of the nonsampled units are known. In this case one can compute $\mathbf{W}_s(u_{ij})$ by using (9) where the distances are those between unit *j* and each unit in the sample. This gives a $n \times n$ spatial weight matrix that can be used to estimate $\hat{\boldsymbol{\beta}}(u_{ij})$ and $\hat{\mathbf{a}}(u_{ij})$ for $j \in r_i$ $(i = 1,...., A)$.

(ii) When the spatial coordinates for the nonsampled units are not known. In this case one can use the centroids of each small area to estimate the fixed and the random effects. In other words, it has assigned as spatial position of each unit belonging to area *i* the spatial coordinates of the centroid of area *i*. That is, there are the same vectors of area-specific coefficients and area-specific random effects, $\hat{\boldsymbol{\beta}}(u_{ij}) = \hat{\boldsymbol{\beta}}(u_i)$, $\hat{\mathbf{a}}(u_{ij}) = \hat{\mathbf{a}}(u_i)$, for all nonsampled units belonging to area *i*, where u_i denotes the spatial coordinates of centroid of area $i\,(i = 1,..., A)$.

Maximum likelihood (ML) estimation is used to estimate the model parameters in (6). In particular, an iterative algorithm for computing the ML estimates of $\boldsymbol{\Omega}$, σ_e^2 and $\boldsymbol{\beta}(u_{ij})$ is implemented. The steps are as follows:

1. Compute the distance matrix of the sample locations;
2. Compute the optimal bandwidth by the CV criterion (10);
3. Compute the spatial weights matrix $\mathbf{W}_s(u_{ij})$ for each sample location;
4. Assign starting values to the variance components $\boldsymbol{\Omega}$ and σ_e^2;
5. Use these starting values to calculate $\mathbf{V}_{ss}(u_{ij})$;

6. Update $\boldsymbol{\beta}(u_{ij})$ for each sample unit by using $\hat{\boldsymbol{\beta}}(u_{ij}) = \left(\mathbf{X}_s^T \mathbf{V}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \mathbf{V}_{ss}^{-1}(u_{ij})\mathbf{y}_s$;
7. Estimate $\boldsymbol{\Omega}$ and σ_e^2 by numerically maximising the log likelihood using for example the Nedler-Mead method (Nedler and Mead, 1965);
8. Return to step 5 and repeat the procedure until convergence.

The convergence is achieved when the difference between the estimated model parameters obtained from two successive iterations is less than a very small value. R code (R Development Core Team, 2010) has been developed for fitting model (6).

4. Mean Squared Error Estimation

The MSE of the EBLUP predictor (4) can be estimated by using the Prasad and Rao (1990) MSE estimator (hereafter denoted by PR). An alternative approach to MSE estimation for the EBLUP has been proposed by Chambers *et al.* (2009) (hereafter denoted by CCT). The PR is an unconditional MSE estimator while the CCT is a conditional one. Unconditional methods of MSE estimation for small area EBLUPs are based on averaging over the distribution of the random area effects. In contrast, conditional methods are based on conditioning on the realised values of the area effects (see also Longford, 2007). In what follows we propose two approaches to MSE estimation for the GWEBLUP predictor (11), a conditional and an unconditional MSE estimator. The conditional estimator is based on the pseudo-linearization approach to MSE estimation proposed by Chambers *et al.* (2009). On the other hand, the unconditional estimator is a second order approximation of the MSE based on Henderson's BLUP theory (Henderson, 1975), followed by the approximations proposed by Kackar and Harville (1984), Prasad and Rao (1990) and Datta and Laihiri (2000). Hereafter, the conditional and unconditional MSE estimators are respectively denoted by the MSE_C and MSE_U.

4.1 The conditional MSE of the GWEBLUP

The conditional approach to MSE estimation is motivated by first re-expressing the GWEBLUP predictor (11) in a pseudo-linear form as a weighted sum of the sample values of *y*, and then applying heteroskedasticity-robust prediction variance estimation methods that treat these weights, which typically depend on estimated variance components, as known. The GWEBLUP can be expressed as (see Appendix 2 for details) as

$$\hat{m}_i^{GWEBLUP} = \sum_{j\in s} w_{ij}^{GWEBLUP} y_j = \left(\mathbf{w}_{is}^{GWEBLUP}\right)^T \mathbf{y}_s,\ i \in 1\ldots A,$$

with

$$\left(\mathbf{w}_{is}^{GWEBLUP}\right)^T = N_i^{-1}\left\{\mathbf{d}_{is} + (N_i - n_i)\left(\bar{\mathbf{B}}_{jr}^T + \bar{\mathbf{C}}_{jr}^T\right)\right\} \tag{12}$$

where $\mathbf{d}_{is}$ (*i*=1, …, *A*) is the *n*-vector with *j*th component takes the value 1 if a unit is in area *i* and is zero otherwise. Here, $\bar{\mathbf{B}}_{jr}^T$ and $\bar{\mathbf{C}}_{jr}^T$ are $n\times 1$ vectors defined as follows:

- $\bar{\mathbf{B}}_{jr}^T = (N_i - n_i)^{-1}\left(\sum_{j\in r_i} \mathbf{B}_{(j)}^T\right)$ with $\mathbf{B}_{(j)}^T = \mathbf{x}_{ij}^T\left(\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)^{-1}\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}(u_{ij}) = \mathbf{x}_{ij}^T\mathbf{H}_{s(j)}^T$
- $\bar{\mathbf{C}}_{jr}^T = (N_i - n_i)^{-1}\left(\sum_{j\in r_i} \mathbf{C}_{(j)}^T\right)$ with $\mathbf{C}_{(j)}^T = \mathbf{z}_{ij}^T\hat{\boldsymbol{\Omega}}\mathbf{Z}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})\left(\mathbf{I}_s - \mathbf{H}_{s(j)}^T\mathbf{X}_s^T\right)^T$ and
- $\mathbf{H}_{s(j)}^T = \left(\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)^{-1}\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}(u_{ij})$.

The estimated MSE of (11) is then

$$\hat{MSE}(\hat{m}_i^{GWEBLUP}) = \hat{Var}(\hat{m}_i^{GWEBLUP}) + \left\{\hat{Bias}(\hat{m}_i^{GWEBLUP})\right\}^2. \tag{13}$$

Let $I(j\in i)$ denote the indicator for whether unit *j* is in area *i*. An estimator of the conditional prediction variance is

$$\hat{Var}(\hat{m}_i^{GWEBLUP}) = N_i^{-2}\sum_{j\in s}\left\{\delta_{ij}^2 + (N_i - n_i)n^{-1}\right\}\hat{\lambda}_j^{-1}(y_j - \hat{\mu}_j)^2, \tag{14}$$

where, $\delta_{ij} = N_i w_{ij}^{GWEBLUP} - I(j\in i)$ and $\hat{\lambda}_j = (1-\phi_{jj})^2 + \sum_{k\in s(-j)}\phi_{kj}^2 = \left\{1 - 2\phi_{jj} + \sum_{k\in s}\phi_{kj}^2\right\}$. Here $\hat{\mu}_j = \sum_{k\in s}\phi_{kj}y_k$ is an unbiased linear estimator of the conditional expected value $\mu_j = E(y_j\,|\,\mathbf{x}_j, \mathbf{a})$, ϕ_{kj} are weights that are defined implicitly by the expression for $\hat{\mu}_j$. Under (6) we have

$$\hat{\mu}_j = \mathbf{B}_{(j)}^T \mathbf{y}_s + \mathbf{z}_j^T \hat{\boldsymbol{\Omega}} \mathbf{Z}_s^T \hat{\mathbf{V}}_{ss}^{-1}(u_{ij}) \left(\mathbf{I}_s - \mathbf{H}_{s(j)}^T \mathbf{X}_s^T \right)^T \mathbf{y}_s . \quad (15)$$

However, because of the well-known shrinkage effect associated with BLUPs, this specification leads to biased estimation of the prediction variance under the conditional model. For this reason, Chambers *et al*. (2009) recommend that $\hat{\mu}_j$ be computed as the 'unshrunken' version of the BLUP for μ_j. That is, following Chambers *et al.* (2009) we use

$$\hat{\mu}_j = \mathbf{B}_{(j)}^T \mathbf{y}_s + \mathbf{z}_j^T \left(\mathbf{Z}_s^T \mathbf{Z}_s \right)^{-1} \mathbf{Z}_s^T \left(\mathbf{I}_s - \mathbf{H}_{s(j)}^T \mathbf{X}_s^T \right)^T \mathbf{y}_s . \quad (16)$$

Note that $\hat{\lambda}_j = 1 + O(n^{-1})$ in this case so that $\hat{\lambda}_j$ will be very close to one in most practical applications. This suggests that there is little to be gained by not setting $\hat{\lambda}_j \equiv 1$ when calculating the conditional prediction variance.

The simple 'plug-in' estimator of bias is

$$\hat{Bias}(\hat{m}_i^{GWEBLUP}) = \sum_{j \in s} w_{ij}^{GWEBLUP} \hat{\mu}_j - N_i^{-1} \sum_{j \in U_i} \hat{\mu}_j \quad (17)$$

with $\hat{\mu}_j$ defined above. Using (14) and (17), we derive the estimator of the conditional MSE of the GWEBLUP (11). Note that this MSE estimator ignores the extra variability associated with estimation of the variance components, and is therefore a heteroskedasticity-robust first order approximation to the actual conditional MSE of the GWEBLUP. Since use of the GWEBLUP (11) will typically require a large overall sample size, we expect that any consequent underestimation of the conditional MSE of the GWEBLUP will be small. The extent of this underestimation will depend on the small area sample sizes and the characteristics of the population of interest, particularly the strength of the small area effects. Finally, we also expect that for very small domain sample sizes the conditional MSE estimator will be unstable.

4.2 The unconditional MSE of the GWEBLUP

Using the development in Appendix 3 and assuming that the sampling faction $f_i = n_i N_i^{-1}$ is non-negligible, a second order approximation to the MSE of the GWEBLUP (11) is given by

$$MSE\left(\hat{m}_i^{GWEBLUP}\right) = E\left(\hat{m}_i^{GWEBLUP} - m_i\right)^2 = M_{1i}(\boldsymbol{\theta}) + M_{2i}(\boldsymbol{\theta}) + M_{3i}(\boldsymbol{\theta}) + M_{4i}(\boldsymbol{\theta}) \tag{18}$$

where

$$M_{1i}(\boldsymbol{\theta}) = N_i^{-2}\left(\sum_{j\in r_i}\mathbf{z}_{ij}^T\left(\boldsymbol{\Omega} - \boldsymbol{\Omega}\mathbf{Z}_s^T\mathbf{V}_{ss}^{-1}(u_{ij})\mathbf{Z}_s\boldsymbol{\Omega}\right)\mathbf{z}_{ij}\right),$$

$$M_{2i}(\boldsymbol{\theta}) = N_i^{-2}\left(\sum_{j\in r_i}\mathbf{b}_{ij}^T\left(\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)^{-1}\mathbf{b}_{ij}\right),$$

$$M_{3i}(\boldsymbol{\theta}) \approx N_i^{-2}\sum_{j\in r_i}\left[tr\left\{\left(\frac{\partial\mathbf{c}_i^T(u_{ij})}{\partial\boldsymbol{\theta}}\right)\mathbf{V}_{ss}(u_{ij})\left(\frac{\partial\mathbf{c}_i^T(u_{ij})}{\partial\boldsymbol{\theta}}\right)^T Var(\hat{\boldsymbol{\theta}})\right\}\right], \text{ and}$$

$$M_{4i}(\boldsymbol{\theta}) = N_i^{-1}(1-f_i)\sigma_e^2.$$

Here $\mathbf{b}_{ij}^T = \left(\mathbf{x}_{ij}^T - \mathbf{z}_{ij}^T\boldsymbol{\Omega}\mathbf{Z}_s^T\mathbf{V}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)$, $\mathbf{c}_i^T(u_{ij}) = \mathbf{z}_{ij}^T\boldsymbol{\Omega}\mathbf{Z}_s^T\mathbf{V}_{ss}^{-1}(u_{ij})$ and $Var(\hat{\boldsymbol{\theta}})$ is asymptotic covariance matrix of the variance components and obtained as the inverse of the Fisher information matrix $I(\hat{\boldsymbol{\theta}})$ with respect to the variance components. It is common practice to estimate the MSE of the predictors by replacing the unknown parameters including components of the variance by their respective estimators. However, such practice leads to severe underestimation of the true MSE.

An approximately unbiased estimator of (18) is

$$\hat{MSE}(\hat{m}_i^{GWEBLUP}) = M_{1i}(\hat{\boldsymbol{\theta}}) + M_{2i}(\hat{\boldsymbol{\theta}}) + 2M_{3i}(\hat{\boldsymbol{\theta}}) + M_{4i}(\hat{\boldsymbol{\theta}}) - M_{5i}(\hat{\boldsymbol{\theta}}), \tag{19}$$

where $M_{ti}(\hat{\boldsymbol{\theta}})$, $t = 1,...,4$ are obtained from $M_{ti}(\boldsymbol{\theta})$ replacing $\boldsymbol{\theta}$ by its estimate $\hat{\boldsymbol{\theta}}$. For large *A*, $M_{5i}(\hat{\boldsymbol{\theta}}) = \text{-}\mathbf{B}_i^T(\hat{\boldsymbol{\theta}})\nabla M_{1i}(\hat{\boldsymbol{\theta}})$, where $\nabla M_{1i}(\hat{\boldsymbol{\theta}})$ is the first order derivative of $M_{1i}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The $M_{5i}(\hat{\boldsymbol{\theta}})$ can be ignored when REML or method of fitting constant is used for estimating variance components since in this case $\mathbf{B}_i^T(\hat{\boldsymbol{\theta}}) = o(A^{-1})$ is negligible. However, this term is non-negligible when variance components are estimated via maximum likelihood (ML) method. For ML method, $\mathbf{B}_i^T(\hat{\boldsymbol{\theta}}) = N_i^{-2}\sum_{j\in r_i}\left[(2A)^{-1}\left\{I^{-1}(\hat{\boldsymbol{\theta}})\underset{1\le k\le q}{col}\, tr\left(\left(\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)^{-1}\left(\mathbf{X}_s^T\mathbf{V}_{ss}^{(k)}(u_{ij})\mathbf{X}_s\right)\right)\right\}\right]$, with

$\mathbf{V}_{ss}^{(k)}(u_{ij}) = \dfrac{\partial \mathbf{V}_{ss}^{-1}(u_{ij})}{\partial \theta_k} = -\mathbf{V}_{ss}^{-1}(u_{ij})\left(\dfrac{\partial \mathbf{V}_{ss}(u_{ij})}{\partial \theta_k}\right)\mathbf{V}_{ss}^{-1}(u_{ij})$ and the inverse of Fisher information matrix , $I^{-1}(\hat{\boldsymbol{\theta}})$ with the $(k,l)-th$ element given by $I_{kl}^{-1}(\hat{\boldsymbol{\theta}}) = 0.5tr\left(\mathbf{V}_{ss}^{-1}\dfrac{\partial \mathbf{V}_{ss}}{\partial \theta_k}\mathbf{V}_{ss}^{-1}\dfrac{\partial \mathbf{V}_{ss}}{\partial \theta_l}\right)$; $k,l = 1,..,q$. Here q is the number of variance component parameters in the model. See Datta and Lahiri, (2000). The MSE estimator (19) is an approximately model unbiased in the sense that its bias is of order $o(A^{-1})$. That is, $E\left\{\hat{MSE}(\hat{m}_i^{GWEBLUP})\right\} = MSE(\hat{m}_i^{GWEBLUP}) + o(A^{-1})$. Both the EBLUP predictor (4) and the PR MSE estimator for the EBLUP can be obtained as special case of predictor (11) and the MSE estimator (19) respectively.

5. Geographically Weighted Synthetic Prediction

In real applications of small area estimation domains may be unplanned. This may result in target small areas with zero sample sizes also referred to as out of sample areas. Estimation for out of sample areas is, however, as important as estimation for in sample areas is. The conventional approach for estimating the area average in this case is synthetic estimation (Rao, 2003, page 46) and is based on the mixed effects model (1) estimated with data from sampled areas. This is equivalent to setting the area effect for the out of sample area equal to zero. Under model (1), the synthetic EBLUP predictor for the small area average for out of sample area *i* is

$$\hat{m}_i^{EBLUPSYN} = \overline{\mathbf{x}}_i^T \hat{\boldsymbol{\beta}} \ . \tag{20}$$

A similar approach can be followed with the GWLMM (6). When geo-referenced population location data are available, this model has the potential to improve conventional synthetic prediction for out of sample areas. We note that with GWLMM-based synthetic estimation all variation in the area-specific predictions comes from the area-specific auxiliary information, including the locations of the population units in the area. We expect that when a truly spatially non-stationary process underlies the data, use of geographically weighted synthetic estimator will

lead to improved efficiency relative to more conventional synthetic mean predictors. The geographically weighted synthetic predictor (GWSYN) for the average of small area *i* is defined by

$$\hat{m}_i^{GWSYN} = N_i^{-1} \sum_{j \in U_i} \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}(u_{ij}) . \tag{21}$$

The unconditional estimator of the MSE for the (21) is

$$\hat{MSE}(\hat{m}_i^{GWSYN}) = N_i^{-2} \sum_{j \in U_i} \left\{ \mathbf{x}_{ij}^T \hat{Var}\left\{ \hat{\boldsymbol{\beta}}(u_{ij}) \right\} \mathbf{x}_{ij} + \mathbf{z}_{ij} \hat{\boldsymbol{\Omega}} \mathbf{z}_{ij}^T \right\} \tag{22}$$

where $\hat{Var}\left\{ \hat{\boldsymbol{\beta}}(u_{ij}) \right\} = \left(\mathbf{X}_s^T \hat{\mathbf{V}}_{ss}^{-1}(u_{ij}) \mathbf{X}_s \right)^{-1}$. When only the centroids of the areas are known, then the geographically weighted synthetic predictor for area *i* is $\hat{m}_i^{GWSYN} = \bar{\mathbf{x}}_i^T \hat{\boldsymbol{\beta}}(u_i)$ with estimates of MSE given by $\hat{MSE}(\hat{m}_i^{GWSYN}) = \bar{\mathbf{x}}_i^T \hat{Var}\left\{ \hat{\boldsymbol{\beta}}(u_i) \right\} \bar{\mathbf{x}}_i + \bar{\mathbf{z}}_i \hat{\Omega} \bar{\mathbf{z}}_i^T$, where $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{z}}_i$ are the vectors of population means of **X** and **Z** respectively in area *i*. Here, u_i denotes the coordinates of centroid of area *i*. The conditional MSE estimator (13) can be used by expressing GWSYN in the pseudo-linear form as:

$$\hat{m}_i^{GWSYN} = N_i^{-1} \sum_{j \in U_i} \mathbf{B}_{(j)}^T \mathbf{y}_s = \bar{\mathbf{B}}_i^T \mathbf{y}_s = \left(\mathbf{w}_{is}^{GWSYN} \right)^T \mathbf{y}_s \tag{23}$$

with weights $\left(\mathbf{w}_{is}^{GWSYN} \right)^T = (w_{ij}^{GWSYN}) = \bar{\mathbf{B}}_i^T$, where $\mathbf{B}_{(j)}^T = \mathbf{x}_{ij}^T \left(\mathbf{X}_s^T \mathbf{V}_{ss}^{-1}(u_{ij}) \mathbf{X}_s \right)^{-1} \mathbf{X}_s^T \mathbf{V}_{ss}^{-1}(u_{ij})$. Estimator (17) of the area-specific bias cannot be used here since this is an out of sample area. Let us denote by *A* and A^0 as the number of sampled and out of sampled areas respectively. Then under model (6)

$$E(\hat{m}_i^{GWSYN} - m_i) = \sum_{d=1}^{A} \sum\nolimits_{j \in s_d} w_{ij}^{GWSYN} \left\{ \mathbf{x}_{dj}^T \boldsymbol{\beta}(u_{dj}) + \mathbf{z}_{dj}^T \mathbf{a}(u_{dj}) \right\} - \bar{\mathbf{x}}_i^T \boldsymbol{\beta}(u_i) - \bar{\mathbf{z}}_i^T \mathbf{a}(u_i) ; \; i = 1, ..., A^0 .$$

The conditional expectation of the square of this expected bias, given the area effects for the sampled areas, is

$$E\left\{ E^2(\hat{m}_i^{GWSYN} - m_i) \middle| \mathbf{X}, \mathbf{a} \right\} = \left[\sum_{d=1}^{A} \sum\nolimits_{j \in s_d} w_{ij}^{GWSYN} \left\{ \mathbf{x}_{dj}^T \boldsymbol{\beta}(u_{dj}) + \mathbf{z}_{dj}^T \mathbf{a}(u_{dj}) \right\} - \bar{\mathbf{x}}_i^T \boldsymbol{\beta}(u_i) \right]^2 + \bar{\mathbf{z}}_i^T \boldsymbol{\Omega} \bar{\mathbf{z}}_i .$$

Then for a non-sampled area *i* $(i = 1, ..., A^0)$ the estimate of the squared bias of the GWSYN predictor (22) is given by

$$\left[\hat{Bias}(\hat{m}_i^{GWSYN})\right]^2 = \left\{\sum_{d=1}^{A}\sum\nolimits_{j\in s_d} w_{ij}^{GWSYN}\left(\mathbf{x}_{dj}^T\hat{\boldsymbol{\beta}}(u_{dj}) + \mathbf{z}_{dj}^T\tilde{\hat{\mathbf{a}}}(u_{dj})\right) - \overline{\mathbf{x}}_i^T\hat{\boldsymbol{\beta}}(u_i)\right\}^2 + \overline{\mathbf{z}}_i^T\hat{\boldsymbol{\Omega}}\overline{\mathbf{z}}_i . \quad (24)$$

Here $\tilde{\hat{\mathbf{a}}}(u_{dj})$ is the 'unshrunken' estimated effect for the sampled area *d* at location *j* given by (16).

6. Empirical Studies

In this section we present empirical results from simulation studies designed to contrast the performance of the small area estimators described in previous the previous sections. Two types of simulation studies are carried out namely, model-based and design-based simulations. In model based simulations a synthetic population is generated at each simulation run under alternative model specifications and a sample is drawn from this population. Design based simulations are based on realistic population structures obtained from real survey data. Two real survey datasets are used for these simulations. The first dataset comes from the Australian Agricultural Grazing Industry Survey (AAGIS) conducted by the Australian Bureau of Agricultural and Resource Economics in year 1995-96 while the second dataset comes from the Environmental Monitoring and Assessment Program (EMAP) that forms part of the Space Time Aquatic Resources Modelling and Analysis Program (STARMAP) at Colorado State University. In the design-based simulations the survey data are first used to generate a synthetic population. The synthetic fixed population is then kept fixed and within area random samples of size equal to the area-specific sizes in the original sample, are drawn.

The small area estimators we contrast in the simulations are the EBLUP (4) and the GWEBLUP (11) for in sample areas and the synthetic EBLUP (SYN) predictor (20) and the geographically weighted synthetic (GWSYN) predictor (21) for out of sample areas. The estimators we considered in our empirical evaluations are summarised in Table 1.

The performance of different small area estimators is evaluated by computing for each small area the Average Relative Bias (*AvRBias*), the Average Relative Root MSE (*AvRRMSE*) and the Average Coverage Rate (*AvCR*) of nominal 95 per cent confidence intervals defined as follows:

$$AvRBias_i = \left(T^{-1}\sum_{t=1}^{T} m_{it}\right)^{-1}\left\{T^{-1}\sum_{t=1}^{T}\left(\hat{m}_{it} - m_{it}\right)\right\}\times 100\,,$$

$$AvRRMSE_i = \left(T^{-1}\sum_{t=1}^{T} m_{it}\right)^{-1}\left\{\sqrt{T^{-1}\sum_{t=1}^{T}\left(\hat{m}_{it} - m_{it}\right)^2}\right\}\times 100\,,\text{ and}$$

$$AvCR_i = \frac{1}{T}\sum_{t=1}^{T} I\left(\left|\hat{m}_{it} - m_{it}\right| \le 2\hat{M}SE_{it}^{1/2}\right)\times 100\,.$$

Here m_{it} denotes the actual average for area *i* at simulation *t*, with $\hat{m}_{it}$ denoting the estimated small area average and $\hat{M}SE_{it}$ denoting the area *i* estimated MSE in simulation *t*. Note that in the design-based simulation study $m_{it} = m_i$ since the population is kept fixed over simulations.

6.1. Model based simulations

Model-based simulations are commonly used for evaluating the performance of estimation procedures. Here we fix the number of small areas at $A = 20$ and use the following two types of models to generate the population values of y_{ij}. In particular, the first method of simulation generates population values of y and x according to a two-level model $y_{ij} = 100 + 1.5x_{ij} + \mathrm{a}_i + e_{ij}$, where $x_{ij} \sim Chi^2(20)$, $j = 1,....,N_i$ and $i = 1,...,A$, with random area effects a_i generated as independent realizations from $N(0, 23.52)$ and e_{ij} distributed as $N(0, 94.09)$, which corresponds to an intra area correlation equal to 0.20. This simulation set up corresponds to the stationary process. The second method of simulation generates population values random effects simulated as in the case of the stationary simulation procedure but now also for the intercept and the slope of the linear model for y to vary according to the longitude and latitude (Salvati *et al*., 2010). This leads to a non-stationary process. That is, the two-level model is $y_{ij} = \beta_{0ij} + \beta_{1ij}x_{ij} + \mathrm{a}_i + e_{ij}$ with

$$\beta_{0ij} = 95 + 0.1\times longitude_{ij} + 0.1\times latitude_{ij};\ \beta_{1ij} = 0.2\times longitude_{ij} + 0.2\times latitude_{ij}$$

and the location coordinates ($longitude_{ij}, latitude_{ij}$) for each unit of the population are independently generated from $U[0,50]$. In the model-based simulations we assume that we know

the spatial coordinates for the sampled units but only the centroids for the out of sample units are known. The small area population sizes N_i are randomly drawn from a uniform distribution on [450,500] and kept fixed over the simulations. A sample of size $n = 400$ is selected from each simulated population with small area sample sizes proportional to the fixed small area population sizes, resulting in an average area sample size of $n_i = 20$. These area specific sample sizes n_i are fixed in the simulations by treating the small areas as strata and carrying out stratified random sampling. A total of $T = 1000$ simulations are then carried out. In each simulation, the average of each small area is estimated by using the predictors outlined in Table 1. Estimates of the corresponding MSEs of these estimators are also calculated using the MSE estimators in Table 1.

Table 2 shows the mean and summaries (minimum, first quartile, median, third quartile and maximum) of the distribution of values of *AvRBias, AvRRMSE* and *AvCR* over simulations. In Table 2 we show the corresponding performance of EBLUP and GWEBLUP in the stationary and non stationary processes. In the stationary case, as one would expect, the EBLUP has lower average and median relative bias and relative RMSE than the GWEBLUP. If one looks at the distribution of the relative biases and the relative RMSEs the EBLUP is performs better. However, things change when we look at the results for the non- stationary process. In this case, we see a substantial gain in terms of relative root mean squared error for the GWEBLUP when compared to the EBLUP.

In Figure 1 we show how the MSE estimator, using (13) and (19) and averaging over simulations, tracks the true MSE of the GWEBLUP. In Figure 1 we see that the proposed MSE estimators provide a good approximation to the true MSE. In the case of non-stationary data, the conditional MSE estimator (13) traces the variability in the true MSE, however, the unconditional estimator (19) leads an average estimate of true MSE. Further, the unconditional estimator (19) is slightly underestimating the true MSE for the GWEBLUP. Furthermore, the proposed methods of MSE estimation provide confidence intervals with good coverage performance.

6.2. Design based simulations

From a practical perspective, design-based simulations are more interesting than model-based simulations since they constitute a more realistic representation of the small area problem. Here design based simulations are based on two real survey data. The first dataset is the AAGIS data for the year 1995-96. In the original sample there are 759 farms from 12 regions, which are defined as the target small areas for the purposes of this study. For this data set there are available the centroids of the 12 regions. We have assigned at each unit belonging to area i the spatial coordinates of the centroid of area i. Using the original sample data and the survey weights we generated a synthetic population of size 39562. A total of 200 independent random samples, each of size n = 759, are then drawn from this fixed population by simple random sampling without replacement from within the 12 regions. The variable of interest is the total cash costs (TCC) and the target is to estimate the average value of TCC in each target area. A range of potential explanatory variables is available for building a working small area model. The covariates used in the fixed part of our working model provide an R^2 value of 0.40; they are the land area, four identifiers for the five industries (i.e. specialist croppers, mixed livestock croppers, sheep specialists, beef specialists, and mixed sheep beef farms), the number of closing stock-beef, the number of closing stock-sheep and the quantity of harvested wheat. In addition, we use an ANOVA test proposed by Brundson *et al.* (1999) for testing the nonstationarity of the model parameters. The nonstationarity test for the original AAGIS sample data indicates that the assumption of stationarity of the model parameters is rejected.

The second dataset comes from the U.S. Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP) Northeast lakes survey (Larsen *et al*., 2001). Between 1991 and 1995, researchers from the U.S. Environmental Protection Agency (EPA) conducted an environmental health study of the lakes in the north-eastern states of the U.S.A. For this study, a sample of 334 lakes -or more accurately, lake locations- was selected from the population of 21,026 lakes in these states using a random systematic design. The lakes making up

this population are grouped into 113 8-digit Hydrologic Unit Codes (HUCs), of which 64 contained less than 5 observations and 27 did not have any observations. Here we define lakes grouped by 8-digit Hydrologic Unit Code (HUC) as our small areas of interest. The variable of interest was Acid Neutralizing Capacity (ANC), an indicator of the acidification risk of water bodies and we are interested in estimating the average of ANC for in and out of sample HUCs. Since some lakes were visited several times during the study period and some of these were measured at more than one site, the total number of observed sites was 349 with a total of 551 measurements. In addition to ANC values and associated survey weights for the sampled locations, the EMAP data set also contained the elevation and geographical coordinates of the centroid of each lake in the target area. In our simulations we use elevation in the fixed part of the working small area model.

In the case of the EMAP data a synthetic population of ANC individual values is nonparametrically simulated using a nearest-neighbour imputation algorithm that retains the spatial structure of the observed ANC values in the EMAP sample data. The algorithm is the same as used in Salvati *et al*. (2010) and defined as follows: (1) we first randomly order the non-sampled locations in order to avoid list order bias and give each sampled location a 'donor weight' equal to the integer component of its survey weight minus 1; (2) taking each non-sample location in turn, we choose a sample location as a donor for the j^{th} non-sample location by selecting one of the ANC values of the EMAP sample locations with probability proportional to $w(u_j, u) = \exp\left[-d^2_{u_j,u} / 2b^2\right]$. Here $d_{u_j,u}$ is the Euclidean distance from the j^{th} non-sample location u_j to the location u of a sampled location and *b* is the GWR bandwidth estimated from the EMAP data; and (3) we reduce the donor weight of the selected donor location by 1. The synthetic population of ANC values is then kept fixed over the Monte-Carlo simulations. A total of 200 independent random samples of lake locations are then taken from the population of 21,026 lake locations by randomly selecting locations in the 86 HUCs that containing EMAP sampled lakes, with sample sizes in these HUCs set to the greatest of five and the original EMAP sample size.

Lakes in HUCs not sampled by EMAP are also not sampled in the simulation study. This results in a total sample size of 652 locations selected within the 86 'EMAP' HUCs. The synthetic ANC values at these 652 sampled locations are then noted. Similar to the AAGIS data, the ANOVA test (Brundson *et al.*, 1999) rejects the null hypothesis of stationarity of the model parameters in the EMAP data. That is, there evidence of nonstationarity in the data. In both design-based simulations we assume to know the spatial coordinates of the centroids for non-sample units.

The results for the AAGIS data are reported in Table 3 and Figure 2. The simulation results for the EMAP data set are presented in Table 4 and Figure 3. Note that the EMAP data have both sampled and out of sampled areas so the results in Table 4 corresponds to both the 86-sampled and the 27-out of sampled areas. In Figure 3 the vertical line in the plots distinguishes the results for sampled and out of sampled areas. The values on the left and on the right side of the vertical line correspond to sampled and out of sample areas respectively.

The GWEBLUP has both smaller relative biases and relative root MSE than the EBLUP predictor for the AAGIS data in Table 3. Two things stand out from Figure 2. Firstly, the proposed unconditional and conditional MSE estimators for the GWEBLUP are performing in exactly the same way as the corresponding MSE estimators for the EBLUP. Secondly, a reduction in the true MSE can be seen for the GWEBLUP. Furthermore, the unconditional MSE estimator gives an average estimate of true MSE while the conditional MSE estimator captures the variability in estimating the true MSE. That is, MSE_C provides better estimates for the area-specific MSE. However, as discussed by Chambers *et al.* (2009) the conditional MSE estimator may be unstable for areas with small sample sizes. In terms of overall coverage properties neither of the two MSE estimators performs overall better.

Turing now to the results in Table 4 for the EMAP data we see that the GWEBLUP performs better than the EBLUP. More interestingly, the results for out of sample areas, reported in the lower part of the Table 4, indicate the advantage of using the proposed methods of small area estimation since the use of the GWSYN reduces the relative bias and increases the efficiency relative to the

SYN. The geographical weighting based predictors have overall better coverage properties. In Table 4, the minimum coverage rates of the SYN for both unconditional and conditional MSE estimators are noteworthy. In this case, two regions have zero coverage rates. We noticed that these two regions have large values of true MSE. This was due to large conditional biases of the SYN predictor itself. The MSE estimates of sampled regions presented on the left side of vertical line in both plots in Figure 3 have same conclusion as those of Figure 2 for the AAGIS data. The results for out of sample regions shown on the right side of the vertical line of the plots in Figure 2 are of more interest to see. For SYN predictor, both conditional and unconditional MSE estimators perform almost identical. For GWSYN predictor, the conditional MSE estimator (23) provides an average estimate of true MSE. However, the unconditional estimator (22) overestimates the true MSE in many out of sample regions. An investigation to the results shows these poor performing out of sample regions are in the one corner of the lakes. The distances from the sampled regions are quite big and this leads to very small weights for these regions. This inflates the variance of fixed effect parameter, a term in the unconditional MSE estimator (22).

7. Concluding Remarks

In this paper we propose a geographically weighted extension of the popular EBLUP, which we refer to as the GWEBLUP, for the small area average under the local linear mixed model (6). In addition, we propose two methods for estimating its MSE. The empirical results provide evidence that the GWEBLUP can be used for efficiently borrowing strength over space in the presence of spatial nonstationarity in the data. Moreover, the use of the GWEBLUP can significantly improve synthetic estimation for out of sample areas. It is worth noting that in this paper all empirical studies are carried out by using the centroids of the small areas. This seems a realistic scenario since in practice the geographic locations of non-sample units will be unknown. Nevertheless, we expect that the gains from using the GWEBLUP will be further enhanced if information on unit level spatial coordinates is available for entire population.

References

Anselin, L. (1992). *Spatial econometrics: Method and models*. Kluwer Academic Publishers, Boston.

Battese, G., Harter, R. and Fuller, W. (1988). An Error-Components Model for Prediction of County Crop Areas using Survey and Satellite Data. *Journal of the American Statistical Association*, **83**, 28-36.

Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1998). Geographically weighted regression - modelling spatial non-stationarity. *Journal of the Royal Statistical Society*, *Series D*, **47(3)**, 431-443.

Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science,* **39**, 497-524.

Chambers, R., Chandra, H. and Tzavidis, N. (2009). On Bias-Robust Mean Squared Error Estimation for Linear Predictors for Domains, Working Papers, 11-09, Centre for Statistical and Survey Methodology, The University of Wollongong, Australia. (Available from: http://cssm.uow.edu.au/publications).

Chambers, R. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, **93,** 255-268.

Datta, G. S. and Lahiri , P. (2000). A Unified Measure of Uncertainty of Estimates for Best Linear Unbiased Predictors in Small Area Estimation Problem. *Statistica Sinica*, 10, 613-627.

Fay, R. E. and Herriot, R. A. (1979). Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistics Association*, **74**, 269-277.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002). *Geographically Weighted Regression*. John Wiley & Sons, West Sussex.

Harville, D.A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, **72**, 320–338.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model, *Biometrics*, **31**, 423-447.

Henderson, C. R., Kempthorne, O., Searle, S. R., and KrosigkSource, C. M. (1959). The Estimation of Environmental and Genetic Trends from Records Subject to Culling. *Biometrics*, **15**, 192-218.

Kackar, R. and Harville, D. A. (1984). Approximations for standard errors of estimators for fixed and random effects in mixed models, *Journal of the American Statistical Association,* **79**, 853-862.

Larsen, D. P., Kincaid, T. M., Jacobs, S. E. and Urquhat, N. S. (2001). Designs for evaluating local and regional scale trends, *Bioscience*, **51**, 1049-1058.

Longford, N.T. (2007). On Standard Errors of Model-Based Small-Area Estimators. *Survey Methodology*, **33**, 69-79.

Nelder, J. and Mead, R. (1965). A Simplex Method for Function Minimization. *Computer Journal*, **7**, 308—313.

Opsomer, J.D., Claeskens, G., Ranalli, M.G., Kauermann, G. and Breidt, F.J. (2008). Nonparametric Small Area Estimation Using Penalized Spline Regression. *Journal of the Royal Statistical Society, Series B*, **70**, 265-286.

Prasad, N.G.N. and Rao, J.N.K. (1990). The Estimation of the Mean Squared Error of Small Area Estimators. *Journal of the American Statistical Association*, **85**, 163-171.

Pratesi, M. and Salvati, N. (2008). Small Area Estimation: the EBLUP estimator based on spatially correlated random area effects, *Statistical Methods & Applications*, **17**, 114-131.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: Wiley.

Singh, B.B., Shukla, G.K. and Kundu, D. (2005) Spatio-Temporal Models in Small Area Estimation, *Survey Methodology*, **31**, *2*, 183-195.

Salvati, N., Tzavidis, N., Pratesi, M., Chambers, R. (2010). Small Area Estimation Via M-quantile Geographically Weighted Regression, forthcoming in *TEST,* DOI 10.1007/s11749-010-0231-1.

Ugarte, M.D., Goicoa, T., Militino, A.F. and Durbán, M. (2009). Spline Smoothing in Small Area Trend Estimation and Forecasting. *Computational Statistics and Data Analysis,* **53,** 3616-3629.

Table 1. Definitions of small area predictors and their MSE used in simulations studies.

Predictors	MSE Estimators
EBLUP: Predictor (4) under model (1)	MSE_C by CCT MSE MSE_U by PR MSE
GWEBLUP: Predictor (11) under model (6)	MSE_C by (13) with weights (12) MSE_U by (19)
SYN: Predictor (20) under model (1)	MSE_C by CCT MSE MSE_U by PR MSE
GWSYN: Predictor (21) under model (6)	MSE_C by (13) with weights (23) and bias (24) MSE_U by (22)

Table 2. Summary of results from model based simulation. Values are given in percentage.

Predictor	Indicator	Summary of across small areas distribution					
		Min	Q1	Median	Mean	Q3	Max
n_i		18	19	20	20	21	22
Stationary process							
EBLUP	RB	-0.067	-0.032	-0.010	0.008	0.053	0.131
	RRMSE	1.351	1.404	1.425	1.425	1.450	1.491
	CR_U	94	95	95	95	96	97
	CR_C	93	93	94	94	94	95
GWEBLUP	RB	-0.071	-0.034	-0.010	0.008	0.058	0.125
	RMSE	1.356	1.412	1.438	1.434	1.460	1.505
	CR_U	94	95	95	95	95	96
	CR_C	92	92	93	93	93	94
Non stationary process							
EBLUP	RB	-0.551	-0.205	0.038	0.257	0.260	2.807
	RRMSE	2.496	2.672	3.021	4.903	4.586	22.432
	CR_U	81	92	98	95	99	100
	CR_C	93	94	94	94	94	95
GWEBLUP	RB	-0.955	-0.498	-0.032	0.500	0.817	7.864
	RRMSE	1.693	2.000	2.502	3.329	3.546	12.945
	CR_U	85	91	93	93	94	95
	CR_C	92	93	93	94	94	95

Table 3. Summary of results from design based simulation using AAGIS data. Values are given in percentage.

Predictor	Indicator	Summary of across small areas distribution					
		Min	Q1	Median	Mean	Q3	Max
n_i		42	46	61	63	78	88
EBLUP	RB	-8.19	-3.09	0.94	0.31	4.40	7.52
	RRMSE	7.16	9.16	10.27	11.51	13.51	18.21
	CR_U	64	84	96	90	99	100
	CR_C	65	84	95	89	97	100
GWEBLUP	RB	-6.85	-1.89	0.20	0.20	3.06	6.08
	RRMSE	6.92	7.73	9.11	10.49	12.63	17.29
	CR_U	68	90	98	93	100	100
	CR_C	67	82	93	88	94	98

Table 4. Summary of results from design based simulation using EMAP data. Values are given in percentage.

Predictor	Indicator	Summary of across areas distribution					
		Min	Q1	Median	Mean	Q3	Max
		86 sampled HUCs					
n_i		5	5	5	8	8	34
EBLUP	RB	-23.31	0.39	10.79	12.55	21.43	83.22
	RRMSE	14.20	23.95	35.18	38.05	49.49	99.00
	CR_U	47	89	96	93	100	100
	CR_C	78	91	95	94	98	100
GWEBLUP	RB	-10.65	-2.96	-0.80	0.61	3.45	28.28
	RRMSE	4.58	22.16	32.97	29.39	41.07	89.96
	CR_U	68	91	98	95	100	100
	CR_C	56	79	84	83	90	100
		27 non-sampled HUCs					
SYN-EBLUP	RB	-72.50	-57.29	-36.59	-2.47	38.14	288.11
	RRMSE	5.75	40.14	53.76	60.44	62.21	288.61
	CR_U	0	100	100	89	100	100
	CR_C	0	100	100	89	100	100
GWSYN-EBLUP	RB	-36.99	-16.24	-2.38	-2.83	7.83	66.53
	RRMSE	10.01	18.37	22.17	29.51	30.11	133.68
	CR_U	99	100	100	100	100	100
	CR_C	98	100	100	100	100	100

Figure 1. Area specific values of actual RMSE (solid line) and average estimated RMSE (dashed line and dotted line) in the model-based simulations. Values for the MSE_U estimator are indicated by the dotted line and by △ while those for the MSE_C estimator are indicated by the dashed line and by ○. The plots show the results for the EBLUP (left) and the GWEBLUP (right) predictors. The plots in the top show the results under the stationary scenario, whereas the plots in the bottom show the results under the non-stationary case.

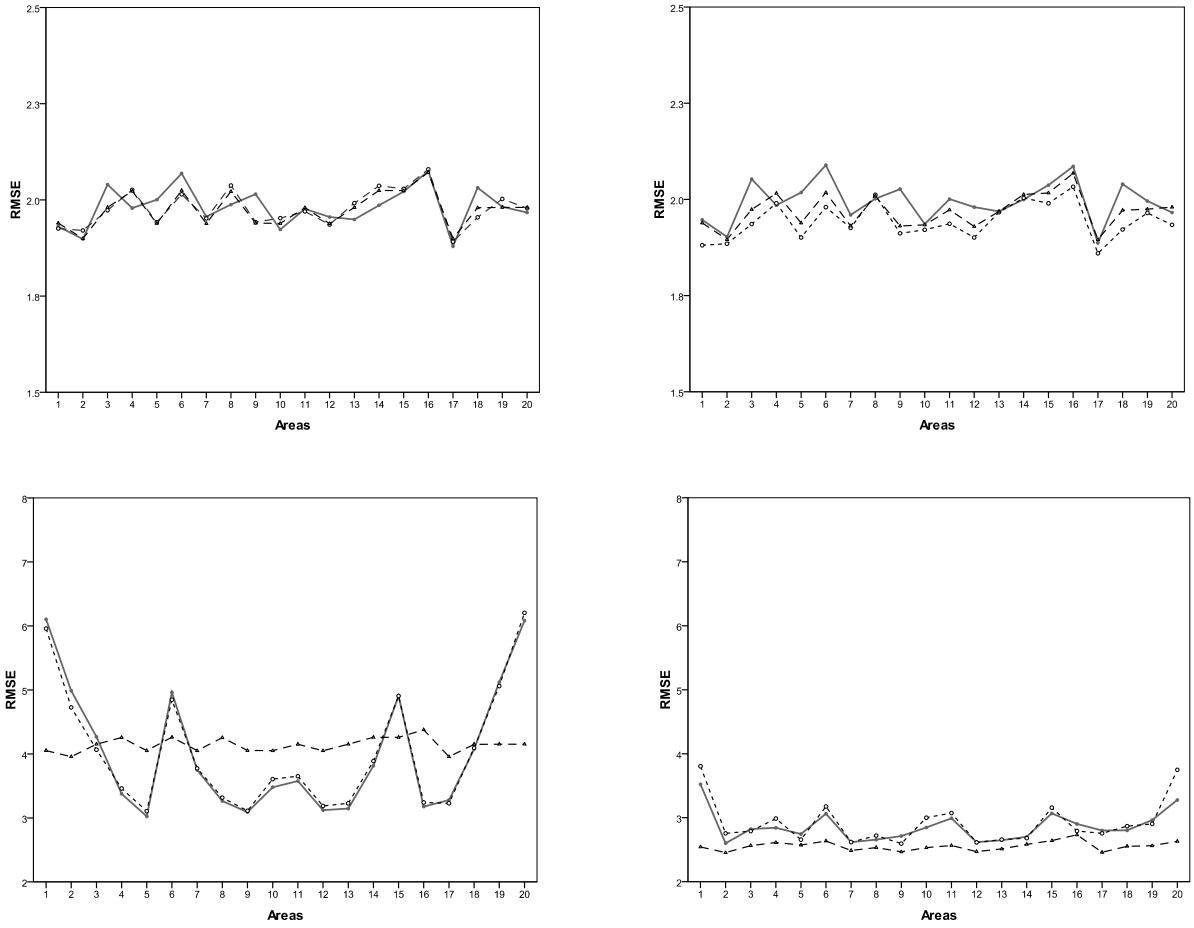


Figure 2. Region specific values of actual RMSE (solid line) and average estimated RMSE (dashed line and dotted line) obtained in the design-based simulations using AAGIS data. Values for the MSE_U estimator are indicated by the dotted line and by ▵ while those for the MSE_C estimator are indicated by the dashed line and by ○. The plots show the results for the EBLUP (left) and the GWEBLUP (right) predictors.

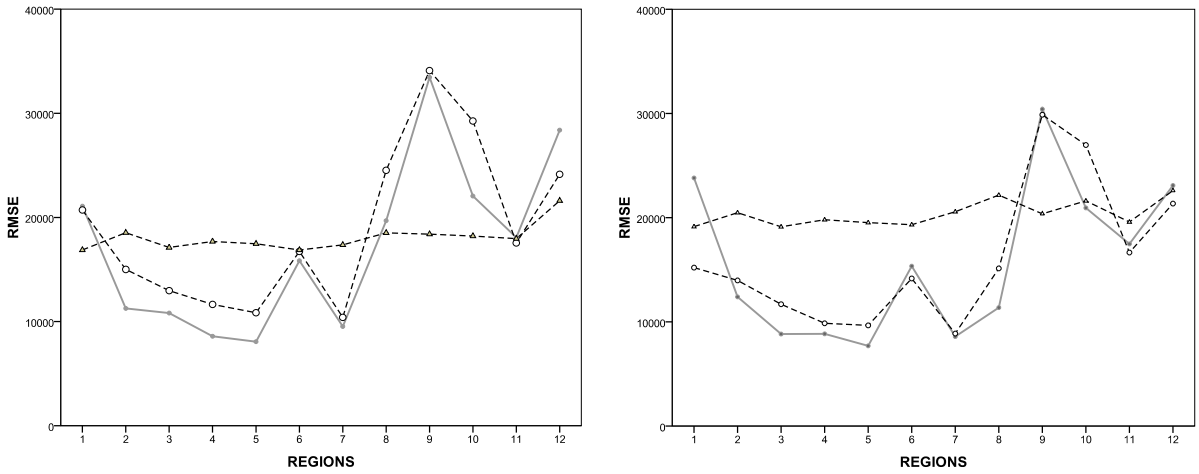
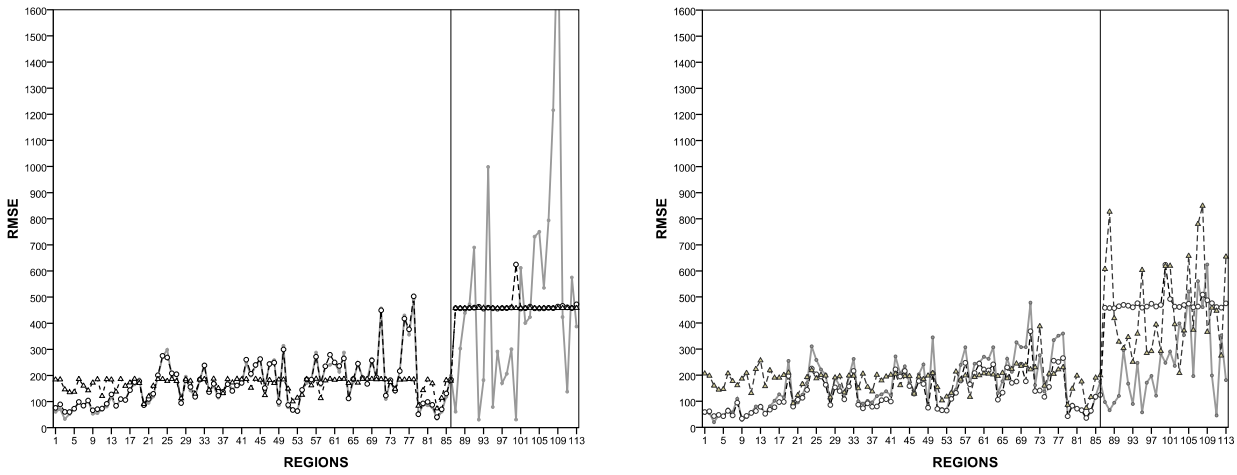


Figure 3. Region specific values of actual RMSE (solid line) and average estimated RMSE (dashed line) obtained in the design-based simulations using EMAP data. Values for the MSE_U estimator are indicated by the dotted line and by ▵ while those for the MSE_C estimator are indicated by the dashed line and by ○. The plots show the results for the EBLUP (left) and the GWEBLUP (right) predictors.



Appendix 1

For a given specific location u_0 in geographical space, the 'joint weighted maximum likelihood estimation' to obtain the estimates of $\boldsymbol{\beta}$ and $\mathbf{a}$ are as follows. Under model (1), following the Henderson *et al.,* (1959), the joint maximum likelihood is

$$f(\mathbf{y}_s,\mathbf{a}) = f(\mathbf{y}_s \mid \mathbf{a})\, f(\mathbf{a}) \approx \exp\left\{-\frac{1}{2}\left(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta} - \mathbf{Z}_s\mathbf{a}\right)^T \mathbf{R}_s^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta} - \mathbf{Z}_s\mathbf{a}\right) - \frac{1}{2}\mathbf{a}^T\boldsymbol{\Omega}^{-1}\mathbf{a}\right\}$$

where $\mathbf{R}_s = \sigma_e^2\mathbf{I}_n$, $f(\mathbf{y}_s \mid \mathbf{a}) \approx \exp\left\{-\frac{1}{2}\left(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta} - \mathbf{Z}_s\mathbf{a}\right)^T \mathbf{R}_s^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta} - \mathbf{Z}_s\mathbf{a}\right)\right\}$ and

$f(\mathbf{a}) \approx \exp\left\{-\frac{1}{2}\mathbf{a}^T\boldsymbol{\Omega}^{-1}\mathbf{a}\right\}$. Using the geographical weights $\mathbf{W}(u_0)$ with respect to a location u_0, the joint log density function is given by

$$\approx \left\{-\frac{1}{2}\begin{pmatrix}\mathbf{a}\\ \mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta} - \mathbf{Z}_s\mathbf{a}\end{pmatrix}^T \begin{bmatrix}\mathbf{I}_{A\times A} & \mathbf{0}\\ \mathbf{0} & \mathbf{W}(u_0)\end{bmatrix}\begin{bmatrix}\boldsymbol{\Omega} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_s\end{bmatrix}^{-1}\begin{pmatrix}\mathbf{a}\\ \mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta} - \mathbf{Z}_s\mathbf{a}\end{pmatrix}\right\}.$$

Differentiating this function with respect to $\boldsymbol{\beta}$ and $\mathbf{a}$, and equating the derivatives to zero gives the Henderson's mixed model equation of form:

$$\mathbf{X}_s^T\mathbf{W}(u_0)\mathbf{R}_s^{-1}\mathbf{X}_s\boldsymbol{\beta} + \mathbf{X}_s^T\mathbf{W}(u_0)\mathbf{R}_s^{-1}\mathbf{Z}_s\mathbf{a} = \mathbf{X}_s^T\mathbf{W}(u_0)\mathbf{R}_s^{-1}\mathbf{y}_s$$

$$\mathbf{Z}_s^T\mathbf{W}(u_0)\mathbf{R}_s^{-1}\mathbf{X}_s\boldsymbol{\beta} + \left[\mathbf{z}_s^T\mathbf{W}(u_0)\mathbf{R}_s^{-1}\mathbf{Z}_s + \boldsymbol{\Omega}^{-1}\right]\mathbf{a} = \mathbf{Z}_s^T\mathbf{W}(u_0)\mathbf{R}_s^{-1}\mathbf{y}_s.$$

Following the usual steps as in Henderson *et al.* (1959) leads to geographically weighted BLUE estimate of regression 'function' $\boldsymbol{\beta}(u_0)$ at an arbitrary location u_0

$$\boldsymbol{\beta}(u_0) = \left(\mathbf{X}_s^T\left[\mathbf{Z}_s\boldsymbol{\Omega}\mathbf{Z}_s^T + \mathbf{W}^{-1}(u_0)\mathbf{R}_s\right]^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s^T\left[\mathbf{Z}_s\boldsymbol{\Omega}\mathbf{Z}_s^T + \mathbf{W}^{-1}(u_0)\mathbf{R}_s\right]^{-1}\mathbf{y}_s,$$

and the geographically weighted BLUP of the random area effects

$$\mathbf{a}(u_0) = \boldsymbol{\Omega}\mathbf{Z}_s^T\left[\mathbf{Z}_s\boldsymbol{\Omega}\mathbf{Z}_s^T + \mathbf{W}^{-1}(u_0)\mathbf{R}_s\right]^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta}(u_0)\right).$$

Appendix 2

Under the geographical weighted regression model (6) the GWEBLUP (11) is

$$\hat{m}_i^{GWEBLUP} = N_i^{-1}\left(\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i}\left\{\mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}(u_{ij}) + \mathbf{z}_{ij}^T\hat{\mathbf{a}}_i(u_{ij})\right\}\right) = N_i^{-1}\left[\sum\nolimits_{j\in s_i} y_{ij} + \sum_{j\in r_i}\left\{\mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}(u_{ij})\right\} + \sum_{j\in r_i}\left\{\mathbf{z}_{ij}^T\hat{\mathbf{a}}_i(u_{ij})\right\}\right]$$

$$= N_i^{-1}\left[\sum\nolimits_{j\in s_i} y_{ij} + \sum_{j\in r_i}\left\{\mathbf{x}_{ij}^T\left(\mathbf{X}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)^{-1}\mathbf{X}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})\mathbf{y}_s\right\} + \sum_{j\in r_i}\left\{\mathbf{z}_{ij}^T\hat{\boldsymbol{\Omega}}\mathbf{Z}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})\left(\mathbf{I}_s - \hat{\mathbf{H}}_{s(j)}^T\mathbf{X}_s^T\right)^T\mathbf{y}_s\right\}\right]$$

$$= N_i^{-1}\left[\sum\nolimits_{j\in s_i} y_{ij} + \left\{\sum_{j\in r_i}\mathbf{B}_{(j)}^T\right\}\mathbf{y}_s + \left\{\sum_{j\in r_i}\mathbf{C}_{(j)}^T\right\}\mathbf{y}_s\right] = N_i^{-1}\left[\left\{I(j\in i) + (N_i - n_i)\left(\bar{\mathbf{B}}_{jr}^T + \bar{\mathbf{C}}_{jr}^T\right)\right\}\mathbf{y}_s\right]$$

$$= \left(\mathbf{w}_{is}^{GWEBLUP}\right)^T\mathbf{y}_s$$

with $\left(\mathbf{w}_{is}^{GWEBLUP}\right)^T = N_i^{-1}\left\{\mathbf{d}_{is} + (N_i - n_i)\left(\bar{\mathbf{B}}_{jr}^T + \bar{\mathbf{C}}_{jr}^T\right)\right\}; i \in 1\ldots A$. Here,

$\mathbf{B}_{(j)}^T = \mathbf{x}_{ij}^T\left(\mathbf{X}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)^{-1}\mathbf{X}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})$, $\hat{\mathbf{V}}_{ss}^{-1}(u_{ij}) = \left(\mathbf{Z}_s\hat{\boldsymbol{\Omega}}\mathbf{Z}_s^T + \mathbf{W}_s^{-1}(u_{ij})\hat{\sigma}_e^2\right)^{-1}$,

$\mathbf{C}_{(j)}^T = \mathbf{z}_{ij}^T\hat{\boldsymbol{\Omega}}\mathbf{Z}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})\left(\mathbf{I}_s - \hat{\mathbf{H}}_{s(j)}^T\mathbf{X}_s^T\right)^T$, and $\hat{\mathbf{H}}_{s(j)}^T = \left(\mathbf{X}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})\mathbf{X}_s\right)^{-1}\mathbf{X}_s^T\hat{\mathbf{V}}_{ss}^{-1}(u_{ij})$.

Appendix 3

To motivate the unconditional MSE estimation of the GWEBLUP (11) described in Section 4.2, we first consider the situation when the variance components $\boldsymbol{\theta} = (\sigma_e^2, \boldsymbol{\Omega})$ are known. In this case, following Henderson (1975), the MSE of the geographically weighted BLUP (GWBLUP) is

$$MSE(\hat{m}_i^{GWBLUP}) = M_{1i}(\boldsymbol{\theta}) + M_{2i}(\boldsymbol{\theta}), \qquad \text{(A3.1)}$$

where $M_{1i}(\boldsymbol{\theta})$ and $M_{2i}(\boldsymbol{\theta})$ are given below equation (18). The first term, $M_{1i}(\boldsymbol{\theta})$ is due to the estimation of random effects, shows the variability in small area predictor when all the model parameters are known and is of order $O(1)$. The second term, $M_{2i}(\boldsymbol{\theta})$ due to estimating the fixed effects parameter, is of order $o(A^{-1})$ for large *A*. The expression (A3.1) assumes that variance components are known, which is not the case in practice. Let us consider that the variance components $\boldsymbol{\theta} = (\sigma_e^2, \boldsymbol{\Omega})$ are estimated by the $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_e^2, \hat{\boldsymbol{\Omega}})$. Then the MSE of the GWEBLUP (11) is obtained as

$$MSE(\hat{m}_i^{GWEBLUP}) = MSE(\hat{m}_i^{GWBLUP}) + E\left(\hat{m}_i^{GWEBLUP} - \hat{m}_i^{GWBLUP}\right)^2 + E\left(\hat{m}_i^{GWEBLUP} - \hat{m}_i^{GWBLUP}\right)\left(\hat{m}_i^{GWBLUP} - m_i\right). \qquad \text{(A3.2)}$$

Here *E* is the expectation under the model (6). The first term on the right hand side of (A3.2) is given by (A3.1). A naïve estimator of MSE of the GWEBLUP is obtained by replacing the unknown variance components in the $MSE(\hat{m}_i^{GWBLUP})$ by some suitable estimators. However, this naïve MSE estimator of the GWEBLUP severely underestimates the true MSE as the variability due to the estimation of the variance components is ignored. Following the pioneering work of Prasad and Rao (1990), we obtain a second order approximation to the MSE of GWEBLUP, with the assumption of large *A* and by neglecting all the terms of the order $o(A^{-1})$. We assume the regularity conditions similar to as given in Prasad and Rao (1990) and Datta and Lahiri (2000), hereafter, PR and DL respectively. See also Rao (2003, page 99-104). Here these regularity conditions are satisfied. Under the normality assumption of two random errors and translation invariance of $\hat{\boldsymbol{\theta}}$,

from the Kackar and Harville (1984), the cross-product term in (A3.2) is negligible. Therefore, we have

$$MSE(\hat{m}_i^{GWEBLUP}) \approx MSE(\hat{m}_i^{GWBLUP}) + E\left(\hat{m}_i^{GWEBLUP} - \hat{m}_i^{GWBLUP}\right)^2 = M_{1i}(\boldsymbol{\theta}) + M_{2i}(\boldsymbol{\theta}) + M_{3i}(\boldsymbol{\theta}) + o(A^{-1}). \qquad \text{(A3.3)}$$

The term $M_{3i}(\boldsymbol{\theta})$ comes from estimating the unknown variance components from the sample data and this term is generally intractable except the special cases, see Rao (2003, page 103). A second order approximation to this term is obtained by using Taylor series linearization method; see for example PR and DL. Since the Taylor series linearization approach is fairly well known in small area estimation literature we are omitting the technical details for this approximation. Following the PR and DL, a second order approximation of the $M_{3i}(\boldsymbol{\theta})$ is given by

$$M_{3i}(\theta) \approx N_i^{-2} \sum_{j \in r_i} \left[tr\left\{ \left(\frac{\partial \mathbf{c}_i^T(u_{ij})}{\partial \boldsymbol{\theta}} \right) \mathbf{V}_{ss}(u_{ij}) \left(\frac{\partial \mathbf{c}_i^T(u_{ij})}{\partial \boldsymbol{\theta}} \right)^T Var(\hat{\boldsymbol{\theta}}) \right\} \right] \qquad \text{(A3.4)}$$

where $\mathbf{c}_i^T(u_{ij}) = \mathbf{z}_{ij}^T \boldsymbol{\Omega} \mathbf{Z}_s^T \mathbf{V}_{ss}^{-1}(u_{ij})$ and $Var(\hat{\boldsymbol{\theta}})$ is asymptotic covariance matrix of $\hat{\theta}$. Note that the $M_{3i}(\boldsymbol{\theta})$ is of the same order $o(A^{-1})$ as that of $M_{2i}(\boldsymbol{\theta})$. When the sampling faction $f_i = n_i N_i^{-1}$ is non-negligible, under model (6) the MSE of the GWEBLUP (11), with bias of order $o(A^{-1})$, is

$$MSE\left(\hat{m}_i^{GWEBLUP}\right) = M_{1i}(\boldsymbol{\theta}) + M_{2i}(\boldsymbol{\theta}) + M_{3i}(\boldsymbol{\theta}) + M_{4i}(\boldsymbol{\theta}) \qquad \text{(A3.5)}$$

where $M_{4i}(\boldsymbol{\theta}) = N_i^{-2} Var\left(\sum_{j \in r_i} \varepsilon_{ij} \right) = N_i^{-1}(1 - f_i)\sigma_e^2$. Following DL, we note that

$E\left\{M_{1i}(\hat{\boldsymbol{\theta}})\right\} = M_{1i}(\boldsymbol{\theta}) + M_{5i}(\boldsymbol{\theta}) - M_{3i}(\boldsymbol{\theta}) + o(A^{-1})$ with $M_{5i}(\boldsymbol{\theta}) = -\mathbf{B}_i^T(\boldsymbol{\theta}) \nabla M_{1i}(\boldsymbol{\theta})$ for large A,

$E\left\{M_{2i}(\hat{\boldsymbol{\theta}})\right\} = M_{2i}(\boldsymbol{\theta}) + o(A^{-1})$, and $E\left\{M_{3i}(\hat{\boldsymbol{\theta}})\right\} = M_{3i}(\boldsymbol{\theta}) + o(A^{-1})$.

This leads to approximately, unbiased estimate of the MSE of the predictor $\hat{m}_i^{GWEBLUP}$ as

$$\hat{MSE}(\hat{m}_i^{GWEBLUP}) \approx M_{1i}(\hat{\boldsymbol{\theta}}) + M_{2i}(\hat{\boldsymbol{\theta}}) + 2M_{3i}(\hat{\boldsymbol{\theta}}) + M_{4i}(\hat{\boldsymbol{\theta}}) - M_{5i}(\hat{\boldsymbol{\theta}}). \qquad \text{(A3.6)}$$

The order of the bias being $o(A^{-1})$ since $M_{2i}(\hat{\theta})$ and $M_{3i}(\hat{\theta})$ have biases of order $o(A^{-1})$.