Small Area Estimation
Using Spatial Information.
The Rathbun Lake Watershed Case Study

Alessandra Petrucci,
Nicola Salvati

Università degli Studi di Firenze

Statistics
Abstract

The paper describes an application of a modified small area estimator to the data collected in the Rathbun Lake Watershed in Iowa (USA). Opsomer et al. (2003) estimated the average erosion per acre for 61 sub-watersheds within the study region using an empirical best linear unbiased predictor (EBLUP) and a composite estimator.

The proposed methodology considers an EBLUP estimator with spatially correlated error taking into account the information provided by neighboring areas.

KEY WORDS: Small area models; watershed erosion; spatial models; spatial EBLUP.

1 Introduction

The previous study (Opsomer et al., 2003) discussed small area models make use of explicit linking models based on random area specific effects that account for between areas variation beyond what is explained by auxiliary variables included in the model. The random area effects are considered independent, but in practice, especially in most of the applications on environmental data, it should be more reasonable to assume that the random area effects between the neighboring areas (for instance the neighborhood could be defined by a contiguity criterium) are correlated and the correlation decays to zero as distance increases.

The aim of this article is to estimate the average sub-watershed erosion taking into account the spatial dimension of the soil erosion data, collected on the Rathbun Lake Watershed, adapting a model with spatially correlated errors in the EBLUP estimator. As well the paper considers the possible gains from modelling the spatial correlation among small area random effects used to represent the unexplained variation of the small area target quantities are examined.

Section 2 introduces the small area models that include random area-specific effects and EBLUP estimator is showed. Section 3 reports the Spatial EBLUP. Section 4 shows the data and the results of the application of Spatial EBLUP to estimate the average sub-watershed erosion per acre on the Rathbun Lake Watershed (Iowa - USA).

2 Area Level Random Effect Models

Area level random effect models are used when auxiliary information is available only at area level. The basic area level model includes random area specific effects and the area specific auxiliary covariates \( x_i = (x_{i,1}, x_{i,2}, ..., x_{i,p}) \) are related to the parameters of inferential interest \( \theta_i \) (totals \( y_i \), means \( \bar{y}_i \)):

\[
\theta_i = x_i \beta + z_i u_i \quad \text{with} \quad i = 1...m
\]
where $z_i$ are known positive constants, $\beta$ is the regression parameters vector $p \times 1$, $u_i$ are independent and identically distributed random variables with mean 0 and variance $\sigma_u^2$. Moreover it assumes that direct estimators $\hat{\theta}_i$ are available and design-unbiased:

$$\hat{\theta}_i = \theta_i + e_i \quad (2)$$

where $e_i$ are independent sampling errors with mean 0 and known variance $\psi_i$. Combining (1) and (2) the obtained model is:

$$\hat{\theta}_i = x_i^t \beta + z_i u_i + e_i \quad \text{with } i = 1...m \quad (3)$$

that is a special case of the general linear mixed model with diagonal covariance structure. The covariance matrices $m \times m$ of $u$ and $e$ are:

$$G = \sigma_u^2 I \quad (4)$$

and

$$R = diag(\psi_i) \quad (5)$$

with $I$ is an identity matrix. Then the covariance matrix of the studied variable is:

$$V = R + ZGZ^T. \quad (6)$$

The Best Linear Unbiased Predictor (BLUP) estimator of $\theta_i$ is:

$$\tilde{\theta}_i(\sigma_u^2) = x_i^t \beta + b_i^T GZ^T V^{-1}(\hat{G} - X \hat{\beta}) \quad (7)$$

where $b_i^T$ is $1 \times m$ vector $(0,0,...0,1,0,...0)$ with 1 referred to $i$-th area and $\beta$ are estimated by generalized least squares: $\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \hat{G}$.

The BLUP estimator is a weighted average of the design-based estimator $\hat{\theta}_i$, and the regression-synthetic estimator $x_i^t \beta$; it can be given by:

$$\tilde{\theta}_i(\sigma_u^2) = \gamma_i \hat{\theta}_i + (1 - \gamma_i) x_i^t \beta \quad (8)$$

where $\gamma_i = \sigma_u^2 / \sigma_u^2 + \psi_i$ is a weight ($0 \leq \gamma_i \leq 1$), it is called shrinkage factor and it measures the uncertainty in modelling the $\theta_i$ (Ghosh and Rao, 1994).

The $MSE[\tilde{\theta}_i(\sigma_u^2)]$ depends on a variance parameter $\sigma_u^2$ and it is:

$$MSE[\tilde{\theta}_i(\sigma_u^2)] = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) \quad (9)$$

with

$$g_{1i}(\sigma_u^2) = b_i^T (G - GZ^T V^{-1} G) b_i = \sigma_u^2 z_i^2 \psi_i (\sigma_u^2 z_i^2 + \psi_i)^{-1} = \gamma_i \psi_i \quad (10)$$

and

$$g_{2i}(\sigma_u^2) = (x_i - b_i^T GZ^T V^{-1} X)(X^T V^{-1} X)^{-1}(x_i - b_i^T GZ^T V^{-1} X) =$$

$$= (1 - \gamma_i)^2 x_i \left[ \sum_{i=1}^m x_i^T x_i \frac{1}{(\sigma_u^2 z_i^2 + \psi_i)} \right]^{-1} x_i^T \quad (11)$$

where $g_{1i}(\sigma_u^2)$ due to the random effects and $g_{2i}(\sigma_u^2)$ accounts for the variability in the estimator $\tilde{\beta}$ (Rao, 2003).

In practical applications $\sigma_u^2$ is unknown and it is replaced by an estimator $\hat{\sigma}_u^2$, a two stage estimator $\tilde{\theta}(\hat{\sigma}_u^2)$ is obtained and it is called Empirical BLUP (EBLUP). It has some properties:

1. it is unbiased for $\theta$;
2. $E[\tilde{\theta}(\hat{\sigma}_u^2)]$ is finite;
3. $\hat{\sigma}_u^2$ is any translation invariant estimator of $\sigma_u^2$ (Kackar and Harville, 1984).
The variance of random effects can be estimated either by Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) methods, assuming normality, or by the method of fitting constants. The MSE of EBLUP estimator appears to be insensitive to the choice of the estimator $\hat{\sigma}_u^2$. Under normality of random effects

$$MSE[\hat{\theta}_i(\hat{\sigma}_u^2)] = MSE[\hat{\theta}_i(\sigma_u^2)] + E[\hat{\theta}_i(\hat{\sigma}_u^2) - \hat{\theta}_i(\sigma_u^2)]^2$$

where the last term is obtained as an approximation because is generally intractable:

$$E[\hat{\theta}_i(\hat{\sigma}_u^2) - \hat{\theta}_i(\sigma_u^2)]^2 \approx \text{tr} \left\{ \left[ \frac{\partial b_i^T GZ^T V^{-1}}{\partial \sigma_u^2} \right] V \left[ \frac{\partial b_i^T GZ^T V^{-1}}{\partial \sigma_u^2} \right]^T \tilde{V}(\sigma_u^2) \right\} = g_{3i}(\sigma_u^2) =$$

$$= \psi_i^2 z_i^4 (\psi_i + \sigma_u^2 z_i^2)^{-3} \tilde{V}(\sigma_u^2)$$

with $\tilde{V}(\sigma_u^2)$ denoting the asymptotic variance of $\sigma_u^2$ which can be approximated as $\tilde{V}(\hat{\sigma}_u^2)$. An approximation to the $MSE[\hat{\theta}_i(\hat{\sigma}_u^2)]$ is

$$MSE[\hat{\theta}_i(\hat{\sigma}_u^2)] \approx g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + g_{3i}(\sigma_u^2)$$

with $g_{2i}(\sigma_u^2)$ and $g_{3i}(\sigma_u^2)$ are of lower order than the term $g_{4i}(\sigma_u^2)$.

In practical application the estimator $\hat{\theta}_i(\hat{\sigma}_u^2)$ has to be associated with an estimator of $MSE[\hat{\theta}_i(\hat{\sigma}_u^2)]$. An approximately unbiased estimator of this mean square error is computed using the following expression:

$$mse[\hat{\theta}_i(\hat{\sigma}_u^2)] \approx g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + 2g_{3i}(\sigma_u^2)$$

when $\hat{\sigma}_u^2$ is obtained by REML method. Otherwise, if a ML procedure is used

$$mse[\hat{\theta}_i(\hat{\sigma}_u^2)] \approx g_{1i}(\sigma_u^2) - b_{i,ML}^T(\hat{\sigma}_u^2) \nabla g_{1i}(\hat{\sigma}_u^2) + g_{2i}(\sigma_u^2) + 2g_{3i}(\sigma_u^2)$$

where $b_{i,ML}^T(\hat{\sigma}_u^2) \nabla g_{1i}(\hat{\sigma}_u^2)$ is an extra term due to the bias $g_{1i}(\sigma_u^2)$ and it is of the same order as $g_{2i}(\sigma_u^2)$ and $g_{3i}(\sigma_u^2)$.

The area basic model considers the random area effects as independent. In practice, it should be more reasonable to assume that the random effects between the neighboring areas (for instance the neighborhood could be define by a distance criterium) are correlated and the correlation decays to zero as distance increases. Considering the spatial dimension of the data, a model with spatially autocorrelated errors has to be implemented, as it is shown in the next section.

3 Spatial Area Level Random Effect Models

In order to take into account the correlation between neighboring areas we regarded to the spatial models and how these models could be utilized in small area estimation (Cressie, 1991). In this study a standard linear regression is considered and the spatial dependence has been incorporated in the error structure ($E[v_i, v_j] \neq 0$). It can be specified in a number of different ways, and results in an error variance covariance matrix of the form:

$$E[v_i, v_j] = \Omega(\tau),$$

where $\tau$ is a vector of parameters, such as the coefficient in a Simultaneously Autoregressive (SAR) or Conditional Autoregressive (CAR) error process, and $v_i, v_j$ are the area random effects. A SAR error model is used:

$$y = X\beta + v$$

where $v = \rho W v + u$, $\rho$ is the spatial autoregressive coefficient, $W$ is the spatial weight matrix for $y$, $u \sim N(0, \sigma_u^2 I)$ is direct area effect and

$$v \sim (0, \sigma_v^2 [I - \rho W]^{-1} ) .$$
Spatial models are a special case of the general linear mixed model. Considering the spatial dimensions of the data, a new model with spatially correlated errors could be implemented and in matrix form it is:

\[
\theta = X\beta + Z(I - \rho W)^{-1}u
\]

\[
\hat{\theta} = \theta + e
\]

where \(\theta\) is the parameter of inferential interest, \(X\) is the matrix of area auxiliary information, \(\beta\) is the regression parameters vector \(p \times 1\), \(Z\) is a matrix of known positive constants, \(v\) is defined as in (18), \(\hat{\theta}\) is the vector of the direct estimators, \(e\) represents the sampling errors with mean \(0\) and known variance \(\text{diag}(\psi_i)\), \(u\) is a vector of independent and identically distributed random variables with mean \(0\) and variance \(\sigma_u^2 I\) and \(m\) is the number of small areas. The covariance matrices \(m \times m\) of \(v\) and \(e\) are:

\[
G = \sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}
\]

and

\[
R = \text{diag}(\psi_i).
\]

Then the covariance matrix of the studied variable is:

\[
V = R + ZGZ^T = \text{diag}(\psi_i) + Z\sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}Z^T
\]

with \(v\) and \(e\) independently distributed. Combining the first and the second model in formula (20) the Spatial BLUP estimator of \(\theta_i\) is:

\[
\hat{\theta}_i^S(\sigma_u^2, \rho) = x_i\hat{\beta} + b_i^T \{\sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}\} ZT \{\text{diag}(\psi_i) + \sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}\}^{-1} (\hat{\theta} - X\hat{\beta})
\]

where \(\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \hat{\theta}\) and \(b_i^T\) is \(1 \times m\) vector \((0, 0, ...0, 1, 0, ...0)\) with 1 in the \(i\)-th position.

The \(\text{MSE}[\hat{\theta}_i^S(\sigma_u^2, \rho)]\), depending on two parameters \((\sigma_u^2, \rho)\), can be expressed as:

\[
\text{MSE}[\hat{\theta}_i^S(\sigma_u^2, \rho)] = g_1(\sigma_u^2, \rho) + g_2(\sigma_u^2, \rho)
\]

with

\[
g_1(\sigma_u^2, \rho) = b_i^T \{\sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1} - \sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}Z^T \times \{\text{diag}(\psi_i) + Z\sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}Z^T\}^{-1}Z\sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}\} b_i
\]

and

\[
g_2(\sigma_u^2, \rho) = (x_i - b_i^T \sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}Z^T \times \{\text{diag}(\psi_i) + Z\sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}Z^T\}^{-1}X) \times\]

\[
(X^T \{\text{diag}(\psi_i) + Z\sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}Z^T\}^{-1}X)^{-1} \times \]

\[
(x_i - b_i^T \sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}Z^T \{\text{diag}(\psi_i) + Z\sigma_u^2 [(I - \rho W)(I - \rho W^T)]^{-1}Z^T\}^{-1}X)^T.
\]

The estimator \(\hat{\theta}_i^S(\sigma_u^2, \rho)\) depends on the variance components \(\sigma_u^2\) and \(\rho\), but in practice they will be unknown. Replacing the parameters with asymptotically consistent estimators \(\hat{\sigma}_u^2, \hat{\rho}\), a two stage estimator \(\hat{\theta}_i^S(\hat{\sigma}_u^2, \hat{\rho})\) is obtained and it is called Spatial EBLUP:

\[
\hat{\theta}_i^S(\hat{\sigma}_u^2, \hat{\rho}) = x_i\hat{\beta} + b_i^T \{\hat{\sigma}_u^2 [(I - \hat{\rho} W)(I - \hat{\rho} W^T)]^{-1}\} ZT \{\text{diag}(\psi_i) + \hat{\sigma}_u^2 [(I - \hat{\rho} W)(I - \hat{\rho} W^T)]^{-1}\}^{-1} (\hat{\theta} - X\hat{\beta})
\]

with \(b_i^T = (0, 0, ...0, 1, 0, ...0)\) and 1 referred to \(i\)-th area. Assuming normality, \(\sigma_u^2\) and \(\rho\) can be estimated both by ML and REML procedures. The ML estimators \(\hat{\sigma}_u^2_{ML}\) and \(\hat{\rho}_{ML}\) can be obtained iteratively using the "scoring" algorithm:

\[
\begin{bmatrix}
\hat{\sigma}_u^2 \\
\hat{\rho}
\end{bmatrix}^{(n+1)} = \begin{bmatrix}
\hat{\sigma}_u^2 \\
\hat{\rho}
\end{bmatrix}^{(n)} + [Z(\hat{\sigma}_u^2, \hat{\rho})]^{-1} \cdot s \left[\hat{\beta}(\hat{\sigma}_u^2, \hat{\rho}), \hat{\sigma}_u^2, \hat{\rho}\right]
\]

(29)
where $s \left[ \hat{\beta}(\sigma_u^{(n)}, \rho^{(n)}), \sigma_u^{(n)}, \rho^{(n)} \right]$ is the vector of the partial derivatives of log-likelihood function with respect to $\sigma_u^2$ and $\rho$, $\bar{T}^{-1}(\sigma_u^2, \rho)$ is the inverse matrix of expected second derivatives minus log-likelihood function with respect to the variance components and $n$ indicates the number of iteration.

The ML procedure to estimate $\sigma_u^2$ and $\rho$ does not consider the loss in degrees of freedom due to estimating $\beta$. This drawback involves the use of REML method (Cressie, 1992). The “scoring” algorithm (29) is used and at convergence the REML estimators are obtained and the asymptotic covariance matrix of $\hat{\beta}_R$, $\hat{\sigma}_u^2$ and $\hat{\rho}_R$ has a diagonal structure $\text{diag} \left[ \hat{\Sigma}(\hat{\beta}_R), \hat{\Sigma}(\hat{\sigma}_u^2, \hat{\rho}_R) \right] \approx \text{diag} \left[ \hat{\Sigma}(\hat{\beta}_ML), \hat{\Sigma}(\hat{\sigma}_u^2, \hat{\rho}_ML) \right]$ with

$$ \hat{\Sigma}(\hat{\beta}_R) \approx \hat{\Sigma}(\hat{\beta}_ML) = (X^T V^{-1} X)^{-1} $$

$$ \hat{\Sigma}(\hat{\sigma}_u^2, \hat{\rho}_R) \approx \hat{\Sigma}(\hat{\sigma}_u^2, \hat{\rho}_ML) = \bar{T}^{-1}(\sigma_u^2, \rho). $$

(30)

The ML and REML estimators are robust, in fact they may work well even under non normal distributions (Jiang, 1996).

The MSE of Spatial EBLUP $\tilde{\theta}_i^\delta(\hat{\sigma}_u^2, \hat{\rho})$ is:

$$ \text{MSE}[\tilde{\theta}_i^\delta(\hat{\sigma}_u^2, \hat{\rho})] = C_{i1}(\sigma_u^2, \rho) + C_{i2}(\sigma_u^2, \rho) + C_{i3}(\sigma_u^2, \rho) $$

(31)

where $C_{i4}(\sigma_u^2, \rho)$ is approximately

$$ C_{i4}(\sigma_u^2, \rho) \approx \text{tr} \left\{ \left[ b_i^T \left( C^{-1} Z^T V^{-1} + \sigma_u^2 C^{-1} Z^T (-V^{-1} Z C^{-1} Z^T V^{-1}) \right) \right] \times \left[ b_i^T \left( A Z^T V^{-1} + \sigma_u^2 C^{-1} Z^T (-V^{-1} Z A Z^T V^{-1}) \right) \right] \right\} $$

(32)

with $C = [(I - \rho W)(I - \rho W^T)]$ and $A = \sigma_u^2 [\sigma_u^2 C^{-1} (2\rho WW^T - 2W) C^{-1}]$. An estimator of $\text{MSE}[\tilde{\theta}_i^\delta(\hat{\sigma}_u^2, \hat{\rho})]$ can be expressed as:

$$ \text{mse}[\tilde{\theta}_i^\delta(\hat{\sigma}_u^2, \hat{\rho})] = C_{i1}(\hat{\sigma}_u^2, \hat{\rho}) + C_{i2}(\hat{\sigma}_u^2, \hat{\rho}) + 2C_{i3}(\hat{\sigma}_u^2, \hat{\rho}) $$

(33)

If $\hat{\sigma}_u^2$ and $\hat{\rho}$ are REML estimators. Otherwise, if ML procedure is used, the $\text{mse}[\tilde{\theta}_i^\delta(\hat{\sigma}_u^2, \hat{\rho})]$ is given by

$$ \text{mse}[\tilde{\theta}_i^\delta(\hat{\sigma}_u^2, \hat{\rho})] \approx g_{i1}(\hat{\sigma}_u^2, \hat{\rho}) - b_{iML}(\hat{\sigma}_u^2, \hat{\rho}) \nabla g_{i1}(\hat{\sigma}_u^2, \hat{\rho}) + g_{i2}(\hat{\sigma}_u^2, \hat{\rho}) + 2g_{i3}(\hat{\sigma}_u^2, \hat{\rho}) $$

(34)

with

$$ \nabla g_{i1}(\hat{\sigma}_u^2, \hat{\rho}) = b_i^T \left\{ \left( C^{-1} - [C^{-1} Z^T V^{-1} Z \sigma_u^2 C^{-1} + \sigma_u^2 C^{-1} Z^T (-V^{-1} Z C^{-1} Z^T V^{-1})] Z \sigma_u^2 C^{-1} + (A - [A Z^T V^{-1} Z \sigma_u^2 C^{-1} + \sigma_u^2 C^{-1} Z^T (-V^{-1} Z A Z^T V^{-1})] Z \sigma_u^2 C^{-1} + + \sigma_u^2 C^{-1} Z^T V^{-1} Z C^{-1}] \right) b_i \right\} $$

(35)

and

$$ b_{iML}(\hat{\sigma}_u^2, \hat{\rho}) = \frac{1}{2m} \left\{ \bar{T}^{-1}(\hat{\sigma}_u^2, \rho) \left[ \text{tr} \left( X^T V^{-1} X \right)^{-1} X^T (-V^{-1} Z C^{-1} Z^T V^{-1}) \right] \right\} \left( \bar{T}^{-1}(\hat{\sigma}_u^2, \rho) \left[ \text{tr} \left( X^T V^{-1} X \right)^{-1} X^T (-V^{-1} Z A Z^T V^{-1}) \right] \right)^{-1}.$$

(36)

If the term $b_{iML}(\hat{\sigma}_u^2, \hat{\rho}) \nabla g_{i1}(\hat{\sigma}_u^2)$ is ignored, the use of ML estimators could lead to underestimation of MSE approximation.
4 Data and results

In 2000 a survey designed to estimate the amount of erosion delivered to the streams in the Rathbun Lake watershed was completed. The watershed, located in southern Iowa (USA), covers more than 365000 acres (147715 ha) in six counties and is divided into 61 sub-watersheds.

The main sources of agricultural erosion are sheet and rill, ephemeral gullies, gullies, and streambanks. The sheet and rill erosion was expected to be a major contributor to total erosion.

In the application the data are the result of this design: each small area (domain) has been divided in plots (total 2146), each plot has been sequentially labelled and a systematic sampling of plots has been selected. The fractional interval has been fixed in order to select four units from each small area (domain). Not all these 4 \times 61 units have been included in the sample. From each domain a simple random sample of 3 units has been selected. Then within each sub-watershed, three 160-acre (64 ha) plots were selected, as is showed in Figure 1, and a sample of 183 units was obtained. The final sample can be reasonably assimilated to a simple random sample from the domains and the sampling variance $\psi_i$ at the domain level can be estimated by

$$\left(1 - \frac{n_i}{N_i}\right) \frac{\hat{\sigma}^2_i}{n_i},$$

where $n_i = 3$ and $N_i$ is the number of plots in the $i$-th area (for details Opsomer et al., 2003). The estimated variance $\hat{\psi}_i$ is then treated as a proxy to $\psi_i$. As result the $mse[\hat{\theta}_S(\hat{\sigma}^2_u, \hat{\rho}, \hat{\psi}_i)]$ is greater than $mse[\hat{\theta}_S(\hat{\sigma}^2_u, \hat{\rho}, \psi_i)]$.

Auxiliary data at the sub-watershed level were the land use and the topography that are considered major determinants of the erosion. Data related to these factor were available for the study region in the form of digital elevation and land use classification coverages. Hence, the Spatial EBLUP method is implemented to this data to estimate the average of watershed erosion in each of the 61 small area within the study region using SAR model. The neighborhood structure $W$ is defined as follows: spatial weight, $w_{ij}$, is 1 if area $i$ shares an edge with area $j$ and 0 otherwise. The value of the estimated spatial autoregressive coefficient $\hat{\rho}$ is 0.132 (s.e. = 0.0258) with ML procedure and 0.136 (s.e. = 0.0288) with REML method, which suggests a moderate spatial relationship. To summarize, Figure 2 displays the map of the Rathbun Lake Watershed with the Spatial EBLUP estimates for the average erosion per acre in only 17 small areas, which are an aggregation of sub-watersheds.

In order to assess the achieved results with the introduction of the spatial information in the small area estimation, the EBLUP estimator and the direct estimator are also calculated. In Table 1 are reported the average estimated standard errors and its variability per acre of Direct, EBLUP and Spatial EBLUP estimators. Table 1 shows also the average estimated of $MSE$ and
Small Area Estimation: Spatial EBLUP

Figure 2: The 17 HUC of Rathbun Lake

Table 1: Average Estimated Standard Errors (A.E.Se.) of Direct, EBLUP and Spatial EBLUP estimators.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>A.E.Se.</th>
<th>V[A.E.Se.]</th>
<th>A.E.MSE</th>
<th>A.E.(g₁)</th>
<th>A.E.(g₂)</th>
<th>A.E.(g₃)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta^S(\hat{\sigma}^2_u, \hat{\rho})$</td>
<td>0.501</td>
<td>0.025</td>
<td>44.33</td>
<td>36.03</td>
<td>5.49</td>
<td>1.38</td>
</tr>
<tr>
<td>$\theta^S(\hat{\sigma}^2_u, \hat{\rho}_R)$</td>
<td>0.510</td>
<td>0.027</td>
<td>45.96</td>
<td>36.92</td>
<td>5.65</td>
<td>1.68</td>
</tr>
<tr>
<td>$\theta(\hat{\sigma}^2_u)$</td>
<td>0.545</td>
<td>0.034</td>
<td>52.76</td>
<td>45.21</td>
<td>5.66</td>
<td>0.92</td>
</tr>
<tr>
<td>DIRECT $\theta$</td>
<td>0.554</td>
<td>0.036</td>
<td>54.84</td>
<td>47.09</td>
<td>5.75</td>
<td>1.00</td>
</tr>
</tbody>
</table>

An evaluation of the resulting model is performed by treating the standard residuals $r = \hat{\theta}^S(\hat{\sigma}^2_u, \hat{\rho}) - X\beta / (\text{diag}(V))^{1/2}$ as iid $N(0, 1)$. In particular, to check the normality of the standardized residuals $r$ and to detect outlier $r$, a normal q-q plot is examined (Figure 3). It can be noted that the outliers $r$ are few, which correspond to neighboring areas in the north-west of the watershed; they can be originated from a particular micro-climate which characterizes that region. Nothing else significant departures from the assumed model are observed.
In conclusion, considering the case study, the use of Spatial EBLUP methodology, which takes into account the SAR spatial model in the small area estimation, reduces the confidence interval.

Acknowledgements: the author thanks Jean Opsomer for the support providing the data and Prof. Chambers and Dr. Saei for their suggestions.
A  Annex

<table>
<thead>
<tr>
<th>Area Code</th>
<th>$\theta^2(\hat{s}^2_u, \hat{\rho}_R)$</th>
<th>s.q.m.</th>
<th>$\theta(\hat{s}^2_u)$</th>
<th>s.q.m.</th>
<th>$\theta$</th>
<th>s.q.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10280201040010</td>
<td>2.831</td>
<td>0.217</td>
<td>2.779</td>
<td>0.218</td>
<td>2.796</td>
<td>0.225</td>
</tr>
<tr>
<td>10280201040020</td>
<td>4.167</td>
<td>0.249</td>
<td>4.166</td>
<td>0.253</td>
<td>4.364</td>
<td>0.264</td>
</tr>
<tr>
<td>10280201040030</td>
<td>3.228</td>
<td>0.368</td>
<td>3.122</td>
<td>0.379</td>
<td>3.306</td>
<td>0.419</td>
</tr>
<tr>
<td>10280201040040</td>
<td>4.306</td>
<td>0.494</td>
<td>4.074</td>
<td>0.541</td>
<td>5.085</td>
<td>0.701</td>
</tr>
<tr>
<td>10280201040050</td>
<td>4.144</td>
<td>0.606</td>
<td>3.540</td>
<td>0.654</td>
<td>3.572</td>
<td>0.992</td>
</tr>
<tr>
<td>10280201040060</td>
<td>4.436</td>
<td>0.188</td>
<td>4.404</td>
<td>0.189</td>
<td>4.512</td>
<td>0.193</td>
</tr>
<tr>
<td>10280201040070</td>
<td>4.035</td>
<td>0.633</td>
<td>2.713</td>
<td>0.766</td>
<td>3.404</td>
<td>2.525</td>
</tr>
<tr>
<td>10280201040080</td>
<td>2.711</td>
<td>0.496</td>
<td>2.213</td>
<td>0.546</td>
<td>1.939</td>
<td>0.714</td>
</tr>
<tr>
<td>10280201040090</td>
<td>3.258</td>
<td>0.467</td>
<td>3.345</td>
<td>0.496</td>
<td>3.882</td>
<td>0.605</td>
</tr>
<tr>
<td>10280201041010</td>
<td>2.037</td>
<td>0.373</td>
<td>1.977</td>
<td>0.388</td>
<td>1.786</td>
<td>0.434</td>
</tr>
<tr>
<td>10280201041011</td>
<td>2.562</td>
<td>0.516</td>
<td>2.357</td>
<td>0.565</td>
<td>2.148</td>
<td>0.761</td>
</tr>
<tr>
<td>10280201041012</td>
<td>2.098</td>
<td>0.530</td>
<td>2.109</td>
<td>0.586</td>
<td>1.754</td>
<td>0.809</td>
</tr>
<tr>
<td>10280201041013</td>
<td>3.625</td>
<td>0.512</td>
<td>3.497</td>
<td>0.547</td>
<td>3.680</td>
<td>0.712</td>
</tr>
<tr>
<td>10280201041014</td>
<td>3.033</td>
<td>0.687</td>
<td>2.806</td>
<td>0.722</td>
<td>3.765</td>
<td>1.542</td>
</tr>
<tr>
<td>10280201041015</td>
<td>3.584</td>
<td>0.388</td>
<td>3.560</td>
<td>0.396</td>
<td>3.704</td>
<td>0.441</td>
</tr>
<tr>
<td>10280201041016</td>
<td>2.001</td>
<td>0.444</td>
<td>1.968</td>
<td>0.475</td>
<td>1.656</td>
<td>0.569</td>
</tr>
<tr>
<td>10280201041017</td>
<td>2.732</td>
<td>0.610</td>
<td>2.961</td>
<td>0.723</td>
<td>4.455</td>
<td>1.565</td>
</tr>
<tr>
<td>10280201041018</td>
<td>5.141</td>
<td>0.652</td>
<td>5.178</td>
<td>0.650</td>
<td>4.953</td>
<td>0.700</td>
</tr>
<tr>
<td>10280201041019</td>
<td>3.241</td>
<td>0.429</td>
<td>3.300</td>
<td>0.448</td>
<td>3.543</td>
<td>0.525</td>
</tr>
<tr>
<td>10280201042020</td>
<td>2.234</td>
<td>0.377</td>
<td>2.419</td>
<td>0.393</td>
<td>2.408</td>
<td>0.439</td>
</tr>
<tr>
<td>10280201042021</td>
<td>2.307</td>
<td>0.512</td>
<td>2.413</td>
<td>0.574</td>
<td>2.154</td>
<td>0.779</td>
</tr>
<tr>
<td>10280201042022</td>
<td>1.267</td>
<td>0.300</td>
<td>1.230</td>
<td>0.307</td>
<td>1.016</td>
<td>0.327</td>
</tr>
<tr>
<td>10280201042023</td>
<td>1.868</td>
<td>0.543</td>
<td>2.021</td>
<td>0.596</td>
<td>1.489</td>
<td>0.841</td>
</tr>
<tr>
<td>10280201042024</td>
<td>2.486</td>
<td>0.606</td>
<td>2.598</td>
<td>0.650</td>
<td>2.668</td>
<td>1.043</td>
</tr>
<tr>
<td>10280201042025</td>
<td>2.076</td>
<td>0.509</td>
<td>2.383</td>
<td>0.546</td>
<td>2.009</td>
<td>0.710</td>
</tr>
<tr>
<td>10280201042026</td>
<td>1.816</td>
<td>0.441</td>
<td>1.766</td>
<td>0.463</td>
<td>1.437</td>
<td>0.547</td>
</tr>
<tr>
<td>10280201042027</td>
<td>1.815</td>
<td>0.551</td>
<td>1.808</td>
<td>0.586</td>
<td>1.160</td>
<td>0.806</td>
</tr>
<tr>
<td>10280201042028</td>
<td>3.006</td>
<td>0.306</td>
<td>3.133</td>
<td>0.311</td>
<td>3.221</td>
<td>0.332</td>
</tr>
<tr>
<td>10280201042029</td>
<td>2.184</td>
<td>0.529</td>
<td>2.409</td>
<td>0.577</td>
<td>2.313</td>
<td>0.790</td>
</tr>
<tr>
<td>10280201050010</td>
<td>2.760</td>
<td>0.606</td>
<td>2.927</td>
<td>0.689</td>
<td>2.766</td>
<td>1.275</td>
</tr>
<tr>
<td>10280201050020</td>
<td>2.583</td>
<td>0.559</td>
<td>2.851</td>
<td>0.629</td>
<td>3.185</td>
<td>0.959</td>
</tr>
</tbody>
</table>

Table A-1: Estimate of the total watershed in each small area and estimated Standard Errors (E.Se.) of Spatial EBLUP, EBLUP and Direct estimators. REML estimators
<table>
<thead>
<tr>
<th>Area Code</th>
<th>$\hat{\theta}^s(\hat{\sigma}^2_{\mu}, \hat{\rho}_R)$</th>
<th>s.q.m.</th>
<th>$\hat{\theta}(\hat{\sigma}^2_{\mu})$</th>
<th>s.q.m.</th>
<th>$\hat{\theta}$</th>
<th>s.q.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10280201060010</td>
<td>3.166</td>
<td>0.666</td>
<td>2.951</td>
<td>0.747</td>
<td>4.366</td>
<td>1.961</td>
</tr>
<tr>
<td>10280201060020</td>
<td>2.738</td>
<td>0.492</td>
<td>2.695</td>
<td>0.524</td>
<td>2.721</td>
<td>0.666</td>
</tr>
<tr>
<td>10280201060030</td>
<td>3.656</td>
<td>0.350</td>
<td>3.603</td>
<td>0.351</td>
<td>3.615</td>
<td>0.366</td>
</tr>
<tr>
<td>10280201060040</td>
<td>3.281</td>
<td>0.501</td>
<td>3.167</td>
<td>0.523</td>
<td>2.801</td>
<td>0.644</td>
</tr>
<tr>
<td>10280201060050</td>
<td>4.197</td>
<td>0.584</td>
<td>4.358</td>
<td>0.617</td>
<td>4.500</td>
<td>0.787</td>
</tr>
<tr>
<td>10280201060060</td>
<td>3.012</td>
<td>0.600</td>
<td>3.118</td>
<td>0.690</td>
<td>3.462</td>
<td>1.257</td>
</tr>
<tr>
<td>10280201060070</td>
<td>3.098</td>
<td>0.665</td>
<td>3.117</td>
<td>0.747</td>
<td>4.498</td>
<td>1.865</td>
</tr>
<tr>
<td>10280201060080</td>
<td>3.028</td>
<td>0.430</td>
<td>2.974</td>
<td>0.438</td>
<td>2.985</td>
<td>0.504</td>
</tr>
<tr>
<td>10280201060090</td>
<td>3.236</td>
<td>0.671</td>
<td>3.531</td>
<td>0.777</td>
<td>4.780</td>
<td>1.608</td>
</tr>
<tr>
<td>10280201060100</td>
<td>2.801</td>
<td>0.586</td>
<td>2.710</td>
<td>0.659</td>
<td>2.316</td>
<td>1.091</td>
</tr>
<tr>
<td>10280201060110</td>
<td>2.566</td>
<td>0.523</td>
<td>2.659</td>
<td>0.551</td>
<td>2.556</td>
<td>0.726</td>
</tr>
<tr>
<td>10280201060120</td>
<td>2.510</td>
<td>0.642</td>
<td>2.801</td>
<td>0.702</td>
<td>3.447</td>
<td>1.359</td>
</tr>
<tr>
<td>10280201060130</td>
<td>2.132</td>
<td>0.256</td>
<td>2.117</td>
<td>0.259</td>
<td>2.044</td>
<td>0.270</td>
</tr>
<tr>
<td>10280201060140</td>
<td>2.271</td>
<td>0.618</td>
<td>2.761</td>
<td>0.666</td>
<td>2.939</td>
<td>1.130</td>
</tr>
<tr>
<td>10280201060150</td>
<td>3.107</td>
<td>0.572</td>
<td>3.097</td>
<td>0.632</td>
<td>2.819</td>
<td>0.952</td>
</tr>
<tr>
<td>10280201060160</td>
<td>2.727</td>
<td>0.642</td>
<td>2.869</td>
<td>0.745</td>
<td>4.093</td>
<td>1.903</td>
</tr>
<tr>
<td>10280201060170</td>
<td>3.916</td>
<td>0.664</td>
<td>4.264</td>
<td>0.750</td>
<td>5.931</td>
<td>1.518</td>
</tr>
<tr>
<td>10280201060180</td>
<td>2.275</td>
<td>0.428</td>
<td>2.338</td>
<td>0.447</td>
<td>2.121</td>
<td>0.521</td>
</tr>
<tr>
<td>10280201060190</td>
<td>1.849</td>
<td>0.394</td>
<td>1.777</td>
<td>0.410</td>
<td>1.423</td>
<td>0.464</td>
</tr>
<tr>
<td>10280201060200</td>
<td>2.300</td>
<td>0.564</td>
<td>2.491</td>
<td>0.619</td>
<td>2.421</td>
<td>0.911</td>
</tr>
<tr>
<td>10280201060210</td>
<td>4.584</td>
<td>1.311</td>
<td>4.927</td>
<td>1.407</td>
<td>6.024</td>
<td>3.131</td>
</tr>
<tr>
<td>10280201060220</td>
<td>2.802</td>
<td>0.454</td>
<td>2.877</td>
<td>0.248</td>
<td>2.870</td>
<td>0.257</td>
</tr>
<tr>
<td>10280201060230</td>
<td>1.900</td>
<td>0.549</td>
<td>2.130</td>
<td>0.617</td>
<td>1.633</td>
<td>0.910</td>
</tr>
<tr>
<td>10280201060240</td>
<td>1.971</td>
<td>0.426</td>
<td>1.999</td>
<td>0.453</td>
<td>1.728</td>
<td>0.533</td>
</tr>
<tr>
<td>10280201060250</td>
<td>1.758</td>
<td>0.270</td>
<td>1.750</td>
<td>0.274</td>
<td>1.621</td>
<td>0.288</td>
</tr>
<tr>
<td>10280201070010</td>
<td>2.274</td>
<td>0.457</td>
<td>2.313</td>
<td>0.476</td>
<td>2.083</td>
<td>0.571</td>
</tr>
<tr>
<td>10280201070020</td>
<td>3.161</td>
<td>0.552</td>
<td>3.308</td>
<td>0.606</td>
<td>3.836</td>
<td>0.878</td>
</tr>
<tr>
<td>10280201070030</td>
<td>2.786</td>
<td>0.616</td>
<td>2.689</td>
<td>0.703</td>
<td>2.776</td>
<td>1.376</td>
</tr>
<tr>
<td>10280201070040</td>
<td>2.635</td>
<td>0.566</td>
<td>2.618</td>
<td>0.625</td>
<td>2.800</td>
<td>0.937</td>
</tr>
<tr>
<td>10280201070050</td>
<td>2.415</td>
<td>0.594</td>
<td>2.542</td>
<td>0.670</td>
<td>2.391</td>
<td>1.134</td>
</tr>
</tbody>
</table>
References


Figure A-1: Map of the area code in the Rathbun Lake Watershed