STATISTICS IN TRANSITION new series, June 2023
Vol. 24, No. 3, pp. 77–94, https://doi.org/10.59170/stattrans-2023-037
Received – 17.05.2023; accepted – 31.05.2023

Spatial Prediction in Small Area Estimation

Martin Vogt[1], Partha Lahiri[2], Ralf Münnich[3]

ABSTRACT

Small area estimation methods have become a widely used tool to provide accurate estimates for regional indicators such as poverty measures. Recent research has provided evidence that spatial modelling still can improve the precision of regional and local estimates. In this paper, we provide an intrinsic spatial autocorrelation model and prove the propriety of the posterior under a flat prior. Further, we show using the SAIPE poverty data that the gain in efficiency using a spatial model can be essentially important in the presence of a lack of strong auxiliary variables.

Key words: Fay-Herriot, CAR, poverty estimation, spatial models.

1. Introduction

International programmes such as the United Nations Sustainable Development Goals (SDG), United States Small Area Income and Poverty (SAIPE), the strategy for combating poverty in the European Union need poverty estimates at disaggregated levels for making public policies. Survey data to provide the necessary information on indicators for poverty and social exclusion are generally constructed at regional rather than local levels. Due to budgetary constraints, it is generally not feasible to allocate samples for all conceivable small areas in which different stakeholders may be interested. The estimation for these unplanned small areas may become problematic if the survey does not provide any sample information for these local areas. A standard solution for this problem is to employ a regression method that exploits a possible relationship between the variable of interest and a set of predictor variables available for both planned and unplanned areas. The method essentially generates synthetic estimates that are subject to considerable bias since the method does not use any direct information on the variable of interest for small areas. One potential way to reduce the bias is to utilize data on the variable of interest from neighbouring areas. This can be achieved by incorporating small area specific random effects, which are then linked by a spatial model; see Saei and Chambers (2005). The method is indeed much more complex than the regression method and its success depends on the ability to define a suitable spatial neighbourhood, specification of a spatial model and estimation of additional parameters of the spatial model.

[1]Trier University of Applied Sciences, Germany.
E-mail: vogt@hochschule-trier.de. ORCID: https://orcid.org/0009-0003-2934-1415.
[2]University of Maryland, College Park, MD 20742, USA.
E-mail: plahiri@umd.edu. ORCID: https://orcid.org/0000-0002-7103-545X.
[3]Economics, Economic and Social Statistics Department, Trier University, Germany.
E-mail: muennich@uni-trier.de. ORCID: https://orcid.org/0000-0001-8285-5667.

In the context of mapping the risk from a disease for granular levels, spatial models have been implemented by both empirical Bayes (see, e.g., Clayton and Kaldor, 1987) and hierarchical Bayes (see, Maiti, 1998) methods. Researchers also considered empirical best prediction approach to implement various extensions of the well-celebrated Fay-Herriot small area model that incorporates spatial correlations; see Saei and Chambers , 2005, Singh et al. (2005), Petrucci et al. (2005), Petrucci and Salvati (2006), Petrucci and Salvati (2008), Petrucci and Salvati (2009), and others. For the hierarchical Bayes approach to the spatial Fay-Herriot models, see You and Zhou (2011) and Chung and Datta (2022).

The estimation methodologies developed in the papers cited in the preceding paragraph do not cover the intrinsic CAR model described in Besag et al. (1991) because these papers exclude models with spatial correlation 1. In this paper, we develop a hierarchical Bayes methodology for an extension of the Fay-Herriot model (Fay and Herriot, 1979) that incorporates spatial neighbourhood using a intrinsic CAR model. Thus the proof of posterior propriety for our model does not follow as a corollary of Chung and Datta (2022). Our research is following the PhD thesis of (Vogt , 2010) and is closely linked to the research of Ghosh et al. (1998) and Sun et al. (1999). Note that Sun et al. (1999) extended the research of Ghosh et al. (1998) and show the propriety of the posterior distribution of hierarchical models using CAR(1) distributions. We apply the same spatial structure and the same prior distributions considered by Sun et al. (1999). However, Sun et al. (1999) assume the sampling variances to be *unknown*, but *equal* and, therefore, the model does not include the Fay-Herriot type model with *known* but *unequal* sampling variances. Hence, their theorem does not ensure the propriety of posterior for the model considered in this paper. We adapt the theorem of Sun et al. (1999) to include the spatial general linear mixed model with known but unequal sampling variances.

We apply the proposed model to data of the SAIPE program in the United States. Our application shows that especially if non-sampled areas are present, the incorporation of spatial neighbourhood improves the estimation. The usefulness of the spatial model can also be observed when the quality of auxiliary information is not good, which is often the case in many applications. However, even if the spatial information is already included in the auxiliary variables, the inclusion of the spatial structure does not worsen the results.

The paper is structured as follows. In the next Section, we first obtain a closed-form expression for the Bayes estimator of the small area mean when no sample from the area is available. For this part of the paper, we consider a non-intrinsic CAR model with known hyperparameters. Our analytical calculations allow us to interpret the Bayes estimator and compare it with the alternative synthetic estimator. We then propose an extension of the Fay-Herriot model (see Fay and Herriot, 1979) that incorporates spatial correlations using an intrinsic CAR model. We show the propriety of the posterior for this proposed model under certain regularity conditions. Using the Small Area Income and Poverty (SAIPE) data of the United States Census Bureau (see Bell and Franco , 2017), we demonstrate in the subsequent Section that spatial correlation could considerably improve on the associated hierarchical Bayes methodology if the area specific auxiliary data are either weak or not available. Our investigation reveals that the need for complex spatial models diminishes as strong area specific predictor variables become available.

2. Theory

In this section, the Fay-Herriot model is extended by including prior distributions on the regression coefficient and the variance component. Afterward, the independence assumption of the random intercepts is replaced by the conditional autoregressive structure. Then, a formula for the mean of the unsampled area is derived, and finally the propriety of the posterior distribution is proved for our intrinsic CAR Fay-Herriot model.

2.1. Spatial Hierarchical Extension of the Fay-Herriot Model

The Fay-Herriot model is given by:

$$\begin{aligned} Y_i|\theta_i &\overset{\text{ind}}{\sim} N(\theta_i\ ,\ \sigma^2_{\varepsilon,i}) \\ \theta_i &= X_i\beta + u_i \\ u_i &\overset{\text{ind}}{\sim} N(0\ ,\ \sigma^2_u I)\ , \end{aligned} \tag{1}$$

where Y_i is an estimate of the true small area mean θ_i, the sampling variances $\sigma^2_{\varepsilon,i}$ are assumed to be known. X_i is $s \times 1$ vector of known auxiliary variables, β is $s \times 1$ vector of unknown regression coefficients, $i = 1, \cdots, m$.

We also consider the following extension of the above model that incorporates possible spatial correlations:

$$\begin{aligned} Y_i|\theta_i &\overset{\text{ind}}{\sim} N(\theta_i\ ,\ \sigma^2_{\varepsilon,i}), \\ \theta_i &= X_i\beta + u_i \\ u &\sim N(0\ ,\ \sigma^2_u (I - p\tilde{Q})^{-1} W) \end{aligned} \tag{2}$$

where $W = \text{diag}(1/\sum_{j=1}^{k} Q_{i,j})$ and $\tilde{Q}_{i,j} = \frac{Q_{i,j}}{\sum_{j=1}^{k} Q_{i,j}}$ are the weights suggested by Banerjee et al. (2004, p. 79). The neighborhood matrix Q is symmetric with $Q_{ii} = 0$.

The overall goal of this section is to analyze how effective the spatial hierarchical Fay-Herriot model (2) is, in terms of prediction for one unsampled area, compared to the corresponding hierarchical Bayes methodology without spatial correlations. To do this a formula for the mean of the unsampled area is derived in the following section.

2.2. The Predicted Mean of One Unsampled Area

In order to derive a formula for the mean of the unsampled area under model (2), the conditional distribution of Y_1 representing the unsampled area given the other areas Y_2 is needed. Assume that the hyperparameters of model (2), i.e., β , σ^2_u and p, are known. Then, model (2) may be written in the form:

$$Y \sim N(\mu\ ,\ \Sigma) \Leftrightarrow \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} X_1\beta \\ X_2\beta \end{pmatrix} , \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) , \tag{3}$$

where $\Sigma := \sigma^2_{\varepsilon,i} I + (I - p\tilde{Q})^{-1} W \sigma^2_u$. Using standard results, we have:

$$Y_1 \mid (Y_2 = y_2) \sim N(\bar{\mu} \; , \; \bar{\Sigma}) \; ,$$

$$\text{where } \bar{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2)$$

$$\text{and } \bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \; .$$

The conditional mean $\bar{\mu}$ may be used to predict one unsampled area. Using the block matrix inversion formula:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \Rightarrow M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \tag{4}$$

an alternative formulation of $\bar{\mu}$ can be derived. The new specification shows explicitly how the spatial correlations enter the model.

Lemma 1 *Consider the model (2) with known β , $\sigma^2_{\varepsilon,i}$, and σ^2_u in the form (3). Then, the mean $\bar{\mu}$ of the conditional distribution of the unsampled area Y_1 given the other areas Y_2, may be written as:*

$$\bar{\mu} = X_1\beta - \sigma^2_u B \left[\sigma^2_{\varepsilon,22} I_{22} W_{22}^{-1}(D - CA^{-1}B) + \sigma^2_u I_{22}\right]^{-1}(y_2 - X_2\beta) \; , \tag{5}$$

where A is the $1,1$ element, B the $1,2:n$ elements, C the $2:n,1$ and D the $2:n,2:n$ elements of $\Sigma_T := I - p\tilde{Q}$. Thus, A represents the variance of the first area, B and C the correlation between the unsampled and the sampled areas, and D includes the spatial dependence parameter p between the sampled areas.

Proof 1 *Model (3) follows with the block matrix inversion formula (4):*

$$\begin{aligned} \Sigma &= \sigma^2_{\varepsilon,i} I + \Sigma_T^{-1} W \sigma^2_u \\ &= \begin{bmatrix} \sigma^2_{\varepsilon,11} + \left((A - BD^{-1}C)^{-1}\right) W_{11}\sigma^2_u & -A^{-1}B(D - CA^{-1}B)^{-1}W_{22}\sigma^2_u \\ -D^{-1}C(A - BD^{-1}C)^{-1}W_{11}\sigma^2_u & \sigma^2_{\varepsilon,22} I_{22} + (D - CA^{-1}B)^{-1}W_{22}\sigma^2_u \end{bmatrix} \\ &=: \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} . \end{aligned}$$

And thus:

$$\begin{aligned} \bar{\mu} &= X_1\beta + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - X_2\beta) \\ &= X_1\beta - A^{-1}B(D - CA^{-1}B)^{-1}W_{22}\sigma^2_u \left[\sigma^2_{\varepsilon,22} I_{22} + (D - CA^{-1}B)^{-1}W_{22}\sigma^2_u\right]^{-1}(y_2 - X_2\beta) \; . \end{aligned}$$

Using the fact that $M^{-1}N^{-1} = (NM)^{-1}$ with $M = (D - CA^{-1}B)^{-1}W_{22}$,

$N = \left[\sigma^2_{\varepsilon,22} I_{22} + (D - CA^{-1}B)^{-1} W_{22} \sigma^2_u\right]^{-1}$ *and* $A^{-1} = 1$ *it follows that:*

$$\begin{aligned}\bar{\mu} &= X_1\beta - \sigma^2_u A^{-1} B \left[(\sigma^2_{\varepsilon,22} I_{22} + (D - CA^{-1}B)^{-1} \sigma^2_u W_{22}) \cdot W_{22}^{-1} (D - CA^{-1}B)\right]^{-1} (y_2 - X_2\beta) \\ &= X_1\beta - \sigma^2_u B \left[\sigma^2_{\varepsilon,22} I_{22} W_{22}^{-1} (D - CB) + \sigma^2_u I_{22}\right]^{-1} (y_2 - X_2\beta) \ .\end{aligned}$$

□

The following example clarifies the meaning of formula (5).

Example 1 *In this example the mean* $\bar{\mu}$ *of formula (5) will be calculated for a situation with 3 areas, where the first area is unsampled. A nearest neighbor structure is assumed, which means that the first area is a neighbor of the second, the second area is a neighbor of the first and the third area, and finally the third area has got area two as a neighbor (see Figure 1).*

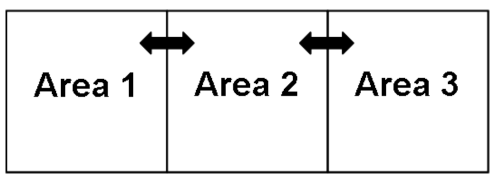

Figure 1: Nearest neighbor structure of 3 areas in a row.

Therefore, the neighborhood matrix Q is as follows:

$$Q = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} .$$

Dividing each row of Q by the number of neighbors ($\tilde{Q}_{i,j} = \frac{Q_{i,j}}{\sum_{j=1}^3 (Q_{i,j})}$) *yields:*

$$\tilde{Q} = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} .$$

The weight matrix $W = diag(\frac{1}{\sum\limits_{j=1}^{3} (Q_{i,j})})$ *is given by:*

$$W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

In the following the mean of one unsampled area shall be calculated using formula (5). Since the first area is unsampled, W_{22} are the weights for the second and third area, given by

$$W_{22} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} .$$

Now $\bar{\mu}$ may be calculated using formula (5). First:

$$\sigma^2_{\varepsilon,22} I_{22} W_{22}^{-1} = \begin{bmatrix} 2\sigma^2_{\varepsilon,2} & 0 \\ 0 & \sigma^2_{\varepsilon,3} \end{bmatrix} \tag{6}$$

and:

$$I - p\tilde{Q} = \begin{bmatrix} 1 & -p & 0 \\ -\frac{p}{2} & 1 & -\frac{p}{2} \\ 0 & -p & 1 \end{bmatrix} =: \begin{bmatrix} A & B \\ C & D \end{bmatrix} , \tag{7}$$

where $A = 1$, $B = \begin{bmatrix} -p & 0 \end{bmatrix}$, $C = \begin{bmatrix} -\frac{p}{2} \\ 0 \end{bmatrix}$ *, and* $D = \begin{bmatrix} 1 & -\frac{p}{2} \\ -p & 1 \end{bmatrix}$.

Using formula (7) yields:

$$D - CA^{-1}B = \begin{bmatrix} 1 - p^2/2 & -p/2 \\ -p & 1 \end{bmatrix} . \tag{8}$$

With (6) and (8) it follows that:

$$\sigma^2_{\varepsilon,22} I_{22} W_{22}^{-1} (D - CA^{-1}B) + \sigma^2_u I_{22} = \begin{bmatrix} 2\sigma^2_{\varepsilon,2}(1 - p^2/2) + \sigma^2_u & -\sigma^2_{\varepsilon,2} p \\ -p\sigma^2_{\varepsilon,3} & \sigma^2_{\varepsilon,3} + \sigma^2_u \end{bmatrix} .$$

Therefore:

$$(\sigma^2_{\varepsilon,22} I_{22} W_{22}^{-1} (D - CA^{-1}B) + \sigma^2_u I_{22})^{-1} = \frac{1}{m} \begin{bmatrix} \sigma^2_{\varepsilon,3} + \sigma^2_u & \sigma^2_{\varepsilon,2} p \\ p\sigma^2_{\varepsilon,3} & 2\sigma^2_{\varepsilon,2}(1 - p^2/2) + \sigma^2_u \end{bmatrix} ,$$

where

$$m := (2\sigma^2_{\varepsilon,2}(1 - p^2/2) + \sigma^2_u)(\sigma^2_{\varepsilon,3} + \sigma^2_u) - \sigma^2_{\varepsilon,2}\sigma^2_{\varepsilon,3} p^2 .$$

Now all the necessary parts to calculate $\bar{\mu}$ are derived. Thus:

$$\begin{aligned} \bar{\mu} &= X_1\beta - \sigma^2_u B \left[\sigma_{\varepsilon,22} I_{22} W_{22}^{-1} (D - CA^{-1}B) + \sigma^2_u I_{22}\right]^{-1} (y_2 - X_2\beta) \\ &= X_1\beta - \sigma^2_u \begin{bmatrix} -p & 0 \end{bmatrix} \frac{1}{m} \begin{bmatrix} \sigma^2_{\varepsilon,3} + \sigma^2_u & \sigma^2_{\varepsilon,2} p \\ p\sigma^2_{\varepsilon,3} & 2\sigma^2_{\varepsilon,2}(1 - p^2/2) + \sigma^2_u \end{bmatrix} (y_2 - \mu_2) . \end{aligned}$$

Matrix calculation yields:

$$\bar{\mu} = X_1\beta + \frac{p}{m} \begin{bmatrix} (\sigma^2_{\varepsilon,3} + \sigma^2_u) & p\sigma^2_{\varepsilon,2} \end{bmatrix} (y_2 - X_2\beta) . \tag{9}$$

Out of formula (9) the following things may be observed.

1. *The resulting estimate is a linear combination between the synthetic estimate* $X_1\beta$ *and information of the other areas* $(y_2 - X_2\beta)$.
2. *Weight is given to neighbors (area 2)* **and** *to non-neighbors (area 3).*
3. *Since* $\sigma_u^2 > 0$ *and* $p < 1$ *it follows that if* $\sigma_{\varepsilon,3}^2$ *and* $\sigma_{\varepsilon,2}^2$ *are of equal size, more power is given to the neighborhood area* 2.
4. *If* $\sigma_{\varepsilon,2}^2$ *is large and thus the information of the second area is low, then more strength is taken from area* 3 *and vice versa.*
5. *If* $p = 0$ *and thus independence is assumed, just the synthetic estimate will be used.*

2.3. Propriety of the Posterior Distribution

In applications the spatial correlation term p is frequently assumed to equal 1. Unfortunately, this leads to an improper prior distribution on the random effects u, the so-called intrinsic CAR model (cf. Besag et al. , 1991 and Besag and Kooperberg , 1995). Thus, the propriety of the posterior distribution is not ensured.

Sun et al. (1999, p. 346) stated the propriety for the intrinsic CAR model with unknown, but equal sampling variances. Let $Y = (Y_1, \cdots, Y_n)$ be the vector of n observations and let X and Z be the $n \times r$ and $n \times k$ design matrices. The least squares estimator for (β', u') is given by $(\hat{\beta}, \hat{u}) = ((X,Z)'(X,Z))^{-}(X,Z)'Y$, where $((X,Z)'(X,Z))^{-}$ is a generalised inverse of $(X,Z)'(X,Z)$. Finally, let $SSE = Y'\{I_n - (X,Z)((X,Z)'(X,Z))^{-}\}Y$ be the sum of squared errors. Then, the following theorem holds (cf. Sun et al. , 1999, p. 346).

Theorem 1 *Consider the linear mixed model:*

$$Y = X\beta + Zu + \varepsilon \ ,$$

where $\varepsilon \sim N(0\ ,\ \sigma_\varepsilon^2 I)$. *Assume the prior distributions* $f(u) \propto \exp\left(-\frac{1}{2\sigma_u^2} u'Bu\right)$, *where* B *is nonnegative definite,* $f(\beta) \propto 1$, $g_\varepsilon(\sigma_\varepsilon^2) \propto (\sigma_\varepsilon^2)^{-(a_\varepsilon+1)} \exp(-b_\varepsilon/\sigma_\varepsilon^2)$ *and* $g_u(\sigma_u^2) \propto (\sigma_u^2)^{-(a_u+1)} \exp(-b_u/\sigma_u^2)$. *The variance components are assumed to be a priori independent. Assume the following conditions:*

- $rank(X) = r$ *and* $rank(u'R_1u + B) = k$, *where* $R_1 = I_n + X(X'X)^{-1}X'$
- $a_u > 0$ *and* $b_u > 0$
- $n - r - k - 2a_\varepsilon > 0$ *and* $SSE + 2b_\varepsilon > 0$

Then, the joint posterior distribution of $(\beta, Z, \sigma_\varepsilon^2, \sigma_u^2)$ *given* Y *is proper.*

In this theorem the sampling variances are assumed to be *unknown*, but *equal*. Therefore, this theorem does not ensure the propriety for a Fay-Herriot type model with *known* but *unequal* sampling variances. However, Theorem 1 can be adapted to the spatial general linear mixed model with known, unequal sampling variances, which includes the spatial hierarchical Fay-Herriot model (2).

Theorem 2 *Consider the linear mixed model* $Y = X\beta + Zu + \varepsilon$, *where* $\varepsilon \sim N(0\ ,\ \Sigma_\varepsilon)$ *with known sampling variance matrix* Σ_ε . *In addition, the following prior distributions are assumed:* $f(u) \propto \exp\left(-\frac{1}{2\sigma_u^2}u'Bu\right)$, *where B is nonnegative definite,* $f(\beta) \propto 1$ *and* $g(\sigma_u^2) \propto (\sigma_u^2)^{-(a+1)}\exp(-b/\sigma_u^2)$. *Assume the following conditions:*

- *rank*$(X) = r$ *and rank*$(u'R_1u + B) = k$, *where* $R_1 = \Sigma_\varepsilon^{-1} + \Sigma_\varepsilon^{-1}X(X'\Sigma_\varepsilon^{-1}X)^{-1}X'\Sigma_\varepsilon^{-1}$
- $a > 0$ *and* $b > 0$.

Then, the joint posterior distribution of (β, u, σ_u^2) *given* Y *is proper.*

Proof 2 *The idea is to integrate the joint posterior density of* (β, u, σ_u^2) *with respect to the three variables:* β, u, *and* σ_u^2 . *The joint posterior density is proportional to:*

$$G := (\sigma_u^2)^{-\frac{1}{2}k} \cdot \exp\{-\frac{1}{2}(Y - X\beta - Zu)'\Sigma_\varepsilon^{-1}(Y - X\beta - Zu) - \frac{u^T Bu}{2\sigma_u^2}\} \cdot g(\sigma_u^2)\ .$$

The proof is split up into six parts. In parts 1 and 2 the joint posterior is transformed for better handling and integrated with respect to β. *Afterward, in parts 2 and 3 the integrated posterior is rearranged to allow for an easier integration with respect to* u. *Finally, in parts 5 and 6 the joint posterior is bounded and thus the existence is shown.*

1. *First, G is transformed since this helps to better handle the integration with respect to* β. *This is done by adding and subtracting* $X\hat\beta$ *and* $Z\hat u$. *It follows that:*

$$\begin{aligned} & (Y - X\beta - Zu)'\Sigma_\varepsilon^{-1}(Y - X\beta - Zu) \\ = & (Y - X\hat\beta - Z\hat u - X(\beta - \hat\beta) - Z(u - \hat u))'\Sigma_\varepsilon^{-1} \cdot \\ & \cdot(Y - X\hat\beta - Z\hat u - X(\beta - \hat\beta) - Z(u - \hat u)). \end{aligned} \tag{10}$$

Expanding (10) yields:

$$\begin{aligned} & (Y - X\beta - Zu)'\Sigma_\varepsilon^{-1}(Y - X\beta - Zu) \\ = & \quad e'\Sigma_\varepsilon^{-1}e \\ - & \quad e'\Sigma_\varepsilon^{-1}X(\beta - \hat\beta) + \\ + & \quad (\beta - \hat\beta)'X'\Sigma_\varepsilon^{-1}X(\beta - \hat\beta) - \\ - & \quad (\beta - \hat\beta)'X'\Sigma_\varepsilon^{-1}e + \\ + & \quad (\beta - \hat\beta)'X'\Sigma_\varepsilon^{-1}Z(u - \hat u) + (u - \hat u)'Z'\Sigma_\varepsilon^{-1}X(\beta - \hat\beta) - \\ - & \quad e'\Sigma_\varepsilon^{-1}Z(u - \hat u) - \\ - & \quad (u - \hat u)'Z'\Sigma_\varepsilon^{-1}e + \\ + & \quad (u - \hat u)'Z'\Sigma_\varepsilon^{-1}Z(u - \hat u), \end{aligned} \tag{11}$$

where $e := Y - X\hat\beta - Z\hat u$.

2. *Now all of the factors in (11) containing* $(\beta - \hat\beta)$ *are collected:*

$$(Y - X\beta - Zu)'\Sigma_\varepsilon^{-1}(Y - X\beta - Zu) = (\beta - \hat\beta - C_0 - C_1)'X'\Sigma_\varepsilon^{-1}X(\beta - \hat\beta - C_0 - C_1) + Const_0,$$

where

$$C_0 = (X'\Sigma_\varepsilon^{-1}X)^{-1}X'\Sigma_\varepsilon^{-1}e,$$
$$C_1 = (X'\Sigma_\varepsilon^{-1}X)^{-1}X'\Sigma_\varepsilon^{-1}Z(u-\hat{u}).$$

Note that $X'\Sigma_\varepsilon^{-1}X$ *is symmetric.* $Const_0$ *is a constant which contains all the factors independent of* β :

$$\begin{aligned} Const_0 &= (u-\hat{u})'Z'\Sigma_\varepsilon^{-1}Z(u-\hat{u}) - (u-\hat{u})'Z'\Sigma_\varepsilon^{-1}e - \\ &- e'\Sigma_\varepsilon^{-1}Z(u-\hat{u}) - C_1'X'\Sigma_\varepsilon^{-1}XC_0 - \\ &- C_1'X'\Sigma_\varepsilon^{-1}XC_1 - C_0'X'\Sigma_\varepsilon^{-1}XC_1 - C_0'X'\Sigma_\varepsilon^{-1}XC_0 \end{aligned}$$

Integrating G with respect to β *yields:*

$$\int_{\mathbb{R}^p} G \; d\beta = \frac{(2\pi)^{\frac{1}{2}p}|X'\Sigma_\varepsilon^{-1}X|^{-\frac{1}{2}}}{(\sigma_u^2)^{\frac{1}{2}k}} \exp\left\{-\frac{1}{2}Const_0 - \frac{u'Bu}{2\sigma_u^2}\right\} \cdot g(\sigma_u^2). \tag{12}$$

3. *Now, the integration with respect to u is prepared. Therefore, the exponential function (12) is calculated and transformed by collecting the terms containing u in* $Const_0$:

   (a) $C_0'X'\Sigma_\varepsilon^{-1}XC_0$ *which is independent of all integration variables and will be seen as a constant.*
   (b) $C_0'X'\Sigma_\varepsilon^{-1}XC_1 = ((X'\Sigma_\varepsilon^{-1}X)^{-1}X'\Sigma_\varepsilon^{-1}e)'X'\Sigma_\varepsilon^{-1}Z(u-\hat{u})$
   (c) $C_1'X'\Sigma_\varepsilon^{-1}XC_1 = (u-\hat{u})'Z'\Sigma_\varepsilon^{-1}X(X'\Sigma_\varepsilon^{-1}X)^{-1}X'\Sigma_\varepsilon^{-1}Z(u-\hat{u})$
   (d) $C_1'X'\Sigma_\varepsilon^{-1}XC_0 = -((X'\Sigma_\varepsilon^{-1}X)^{-1}X'\Sigma_\varepsilon^{-1}Z(u-\hat{u}))'X'\Sigma_\varepsilon^{-1}e$
   (e) $e'\Sigma_\varepsilon^{-1}Z(u-\hat{u})$
   (f) $(u-\hat{u})'Z'\Sigma_\varepsilon^{-1}e$
   (g) $(u-\hat{u})'Z'\Sigma_\varepsilon^{-1}Z(u-\hat{u})$

   This leads to:

$$Const_0 = (u-\hat{u}-C_2)'Z'R_1Z(u-\hat{u}-C_2) + Const_1, \tag{13}$$

   where

$$C_2 = (Z'R_1Z)^{-1}Z'(R_1+\Sigma_\varepsilon^{-1})e$$

   includes the above terms with one $(u-\hat{u})$. *The constant* $Const_1$ *contains all of the terms independent of u which arise by including* C_2 *in the formula.*

4. *Now,* $\frac{u'Bu}{\sigma_u^2}$ *is considered. Therefore, formula (13) of the previous step is adapted:*

$$Const_0 + \frac{u'Bu}{\sigma_u^2} = (u-\hat{u}-C_2)'Z'R_1Z(u-\hat{u}-C_2) + Const_1 + \frac{u'Bu}{\sigma_u^2}.$$

This leads to

$$u'Z'R_1Zu + \frac{u'Bu}{\sigma_u^2} + \text{other terms}. \tag{14}$$

A rearrangement of terms in (14) yields:

$$(u - C_3)'R_2(u - C_3) - (C_3)'R_2C_3, \tag{15}$$

where $R_2 = Z'R_1Z + \frac{B}{\sigma_u^2}$ *and* $C_3 = R_2^{-1}Z'R_1Z(\hat{u} + C_2)$.
Using the integrated G in (12) together with (15), it follows that:

$$\int_{\mathbb{R}^k}\int_{\mathbb{R}^r} G \; d\beta du = \frac{(2\pi)^{\frac{1}{2}(r+k)}|X'\Sigma_\varepsilon^{-1}X|^{-\frac{1}{2}}}{(\sigma_u^2)^{\frac{1}{2}k+a+1}|R_2|^{\frac{1}{2}}} \exp\left\{-(C_3)'R_2C_3 - \frac{b}{\sigma_u^2}\right\} . \tag{16}$$

Integration leads to the factor $|R_2|^{-\frac{1}{2}}$ *in front of the exponential function. Since* $Z'R_1ZR_2^{-1}Z'R_1Z$ *in* $(C_3)'R_2C_3$ *is nonnegative definite this term can be bounded by* 0 *and thus discarded out of the integral.*

5. *Using arguments similar to Sun et al. (1999, p. 346), we get:*

$$\begin{aligned} |R_2|^{-\frac{1}{2}} &\leq \{min(1, (\sigma_u^2)^{-1})^k \cdot |Z'R_1Z + B|\}^{-\frac{1}{2}} \\ &< (1 + (\sigma_u^2)^{\frac{k}{2}}) \cdot |Z'R_1Z + B|^{-\frac{1}{2}} \ . \end{aligned} \tag{17}$$

6. *Finally, combining (16) and (17), we have:*

$$\int_{\mathbb{R}^k}\int_{\mathbb{R}^r} G \; d\beta du \leq (2\pi)^{r+k}|X'\Sigma_\varepsilon^{-1}X|^{-\frac{1}{2}}|Z'R_1Z + B|^{-\frac{1}{2}}(J_1 + J_2) \ ,$$

where

$$J_1 = \frac{1}{(\sigma_u^2)^{\frac{1}{2}k+a+1}} \exp\left(-\frac{b}{\sigma_u^2}\right) \ ,$$

and

$$J_2 = \frac{1}{(\sigma_u^2)^{a+1}} \exp\left(-\frac{b}{\sigma_u^2}\right) \ .$$

Since $a > 0$ *and* $b > 0$ *the integrals* J_1 *and* J_2 *exist with respect to* σ_u^2 .

This completes the proof. □

Remark:

1. In the theorem presented in Sun et al. (1999), the sum of squared errors SSE arises in the assumptions. This term is not important in our case, since it depends only on the known variance-covariance matrix Σ_ε .

2. In the original proof, all terms containing e cancel out of formula (11), since SSE is orthogonal to e . This needs not be the case if Σ_ε is included in the term.

3. An empirical comparison of the intrinsic CAR Fay-Herriot model and non-spatial Fay-Herriot models

In the following, we compare the Fay-Herriot model (1) with the intrinsic CAR Fay-Herriot model (2) (i.e. with $p = 1$) in terms of prediction using real data. To implement the intrinsic CAR Fay-Herriot model, we used priors described in Theorem 2. As for the implementation of the Fay-Herriot model (1), we considered the following prior distributions proposed by Sun et al. (1999):

$$\beta \propto \text{Uniform}(\mathbb{R}^s),\ \mathbb{R}^s \text{ being the s-dimensional Euclidean space},$$
$$1/\sigma_u^2 \sim \Gamma(0.5\ ,\ 0.005).$$

3.1. Data

We consider the data from the Small Area Income and Poverty Estimates of the U.S. Census Bureau for the years 1989 and 1993 . For details on the data, see Bell and Franco (2017). Four covariates are available: Internal Revenue Service (IRS) pseudo child poverty rate ($x1$), IRS non-filer rate ($x2$), food stamp participation rate ($x3$) and census residuals ($x4$). The known sampling variances are denoted by d . In this application the official Small Area Income and Poverty Estimates (SAIPE) estimates are treated as a gold standard. In the application 48 contiguous United States and the District of Columbia are considered. Every area is left out once and is predicted by means of spatial and non-spatial models. This procedure is repeated for different numbers of covariates $(0-4)$ and for the years 1989 and 1993 . The estimation results are compared to the official estimates (treated as gold standard in this paper).

3.2. Results

Since the results are similar for the years 1989 and 1993 the following interpretation will just be for the year 1993. Table 1 contains the simulation results for the year 1993 and Table 2 for the year 1989 . Column 1 of Table 1 shows different measures of comparison, based on the squared deviance between the estimator and the official value, the absolute deviance and the maximum of the deviance. Since each of the 49 states is left out once, the deviances are averaged over all states. The deviances are constructed for the model containing all of the four covariates, no covariate or each of the covariates alone. Columns 2 and 3 contain the corresponding values for the spatial and non-spatial models. The last two columns compare the deviances of the spatial and non-spatial model with each other (difference and ratio).
The following observations can be made:

1. If no covariates are included, the deviances for the spatial and non-spatial models are large. These values decrease as the quality and number of covariates increases. The lowest value is reached, when all covariates are included.

2. The ratio of the non-spatial deviance compared to the spatial is large, if no or weak covariates are included. The ratio decreases, if the quality of the covariates improves.

Table 1: Simulation results for the year 1993 .

	spatial	**non-sp.**	**spatial − non-sp.**	$\frac{\textbf{non-sp.}}{\textbf{spatial}}$
$\frac{1}{49} \cdot \sum_{i=1}^{49} ((\textbf{Estimator}_i - \textbf{Official}_i)^2)$				
all	1,08	1,08	0,00	1,00
x1	7,76	12,79	-5,03	1,65
x2	17,04	27,44	-10,40	1,61
x3	4,74	4,63	0,11	0,98
x4	23,42	33,55	-10,13	1,43
without	25,14	34,01	-8,86	1,35
$\frac{1}{49} \cdot \sum_{i=1}^{49} (\lvert\textbf{Estimator}_i - \textbf{Official}_i\rvert)$				
all	0,79	0,79	0,00	1,00
x1	2,20	2,84	-0,65	1,29
x2	3,00	4,04	-1,05	1,35
x3	1,76	1,78	-0,02	1,01
x4	3,76	4,85	-1,09	1,29
without	3,85	4,89	-1,04	1,27
$\textbf{max}(\lvert\textbf{Estimator} - \textbf{Official}\rvert)$				
all	0,07	0,07	0,00	1,00
x1	0,12	0,17	-0,05	1,38
x2	0,24	0,30	-0,05	1,21
x3	0,12	0,12	0,00	0,99
x4	0,26	0,29	-0,03	1,12
without	0,29	0,27	0,02	0,93

3. If all covariates are included there is no gain by using the spatial model.

The same effects can be seen, in Figures 2, 4 and 5. These figures compare the predicted values of the spatial and non-spatial models with varying numbers of covariates. Figure 2 shows that if all covariates are included, there is no visible difference between the spatial and non-spatial models. However, if no covariate is included, then the predicted values of the non-spatial model compared to the official values are almost constant. But the spatial model improves the relationship. The same effect can be observed if covariates of different quality are included (Figures 4 and 5). Figure 3 underlines these results, by showing the squared deviance of the spatial and non-spatial models for all 4 covariates (upper plots) and no covariates (lower plots) on the map. If no covariates are included the spatial model performs better than the non-spatial model. This effect diminishes if all covariates are included.

Table 2: Simulation results for the year 1989 .

	spatial	**non-sp.**	**spatial – non-sp.**	$\frac{\textbf{non-sp.}}{\textbf{spatial}}$
$\frac{1}{49} \cdot \sum_{i=1}^{49} ((\textbf{Estimator}_i - \textbf{True}_i)^2)$				
all	0,97	0,95	0,02	0,98
x1	4,53	4,81	-0,28	1,06
x2	12,14	21,83	-9,69	1,80
x3	3,89	5,14	-1,25	1,32
x4	17,27	27,38	-10,11	1,59
without	16,11	26,57	-10,46	1,65
$\frac{1}{49} \cdot \sum_{i=1}^{49} (\lvert\textbf{Estimator}_i - \textbf{True}_i\rvert)$				
all	0,82	0,81	0,01	0,98
x1	1,51	1,55	-0,04	1,02
x2	2,64	3,42	-0,79	1,30
x3	1,58	1,81	-0,23	1,15
x4	3,13	4,04	-0,91	1,29
without	3,16	3,97	-0,81	1,26
$\textbf{max}(\lvert\textbf{Estimator} - \textbf{True}\rvert)$				
all	0,05	0,05	0,00	1,00
x1	0,14	0,14	0,00	1,02
x2	0,25	0,33	-0,08	1,30
x3	0,09	0,10	-0,01	1,07
x4	0,30	0,34	-0,04	1,12
without	0,25	0,34	-0,09	1,34

4. Conclusion

In this paper, we have developed a hierarchical Bayes methodology for an extension of the well-celebrated Fay-Herriot model that incorporates spatial correlation using an intrinsic CAR model. We have proved the propriety of the posterior distribution for our proposed model. We have tested the effect of covariates on the estimation results. An application to SAIPE data revealed that modeling spatial correlation can considerably improve on the associated hierarchical Bayes methodology if the area specific auxiliary data are either weak or not available. This effect diminishes if the quality of the area specific covariates improves. Like Besag et al. (1991) we also assumed proper prior for the variance component. As for the future research, it will be of interest to study the sensitivity of such proper prior and to compare our hierarchical Bayes estimator with various empirical best predictors and hierarchical Bayes estimators proposed in the literature that use non-intrinsic CAR model extensions of the Fay-Herriot model.

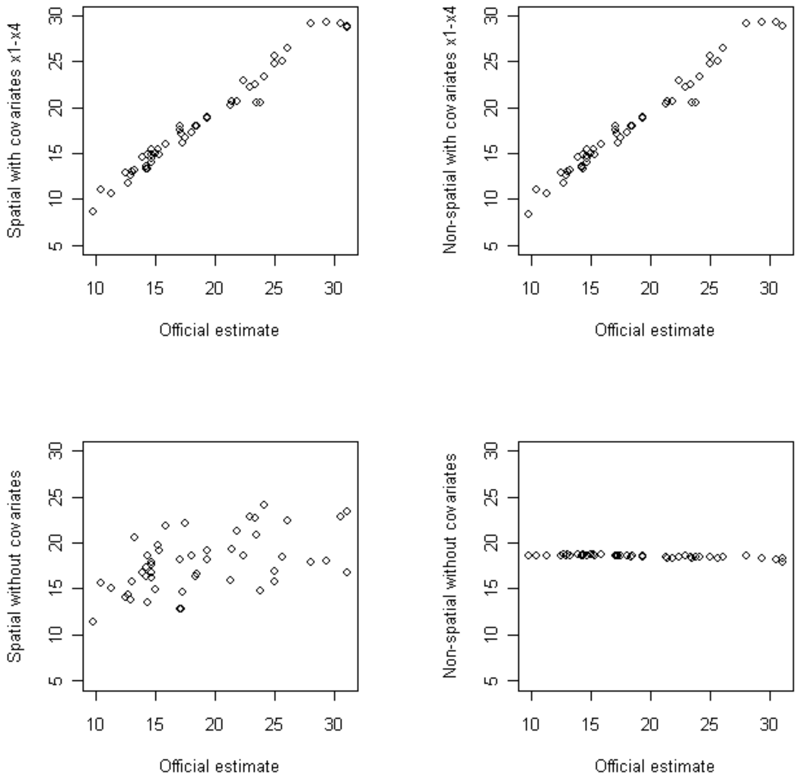

Figure 2: Predicted values of the spatial and non-spatial FH model compared to the official estimates 1993: 4 and no covariates

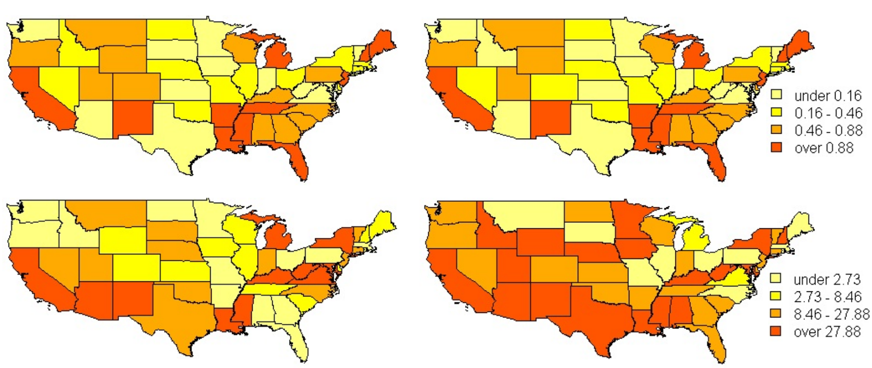

Figure 3: Squared deviance of the spatial and non-spatial FH model for 4 (upper plots) and no (lower plots) covariates.

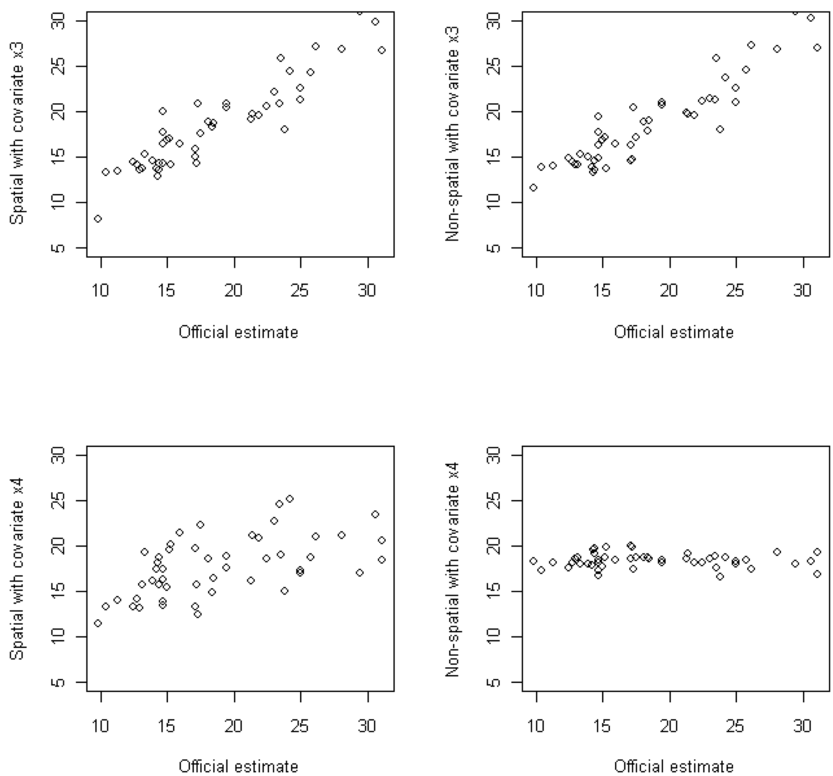

Figure 4: Predicted values of the spatial and non-spatial FH model compared to the official estimates 1993: covariates x3, x4

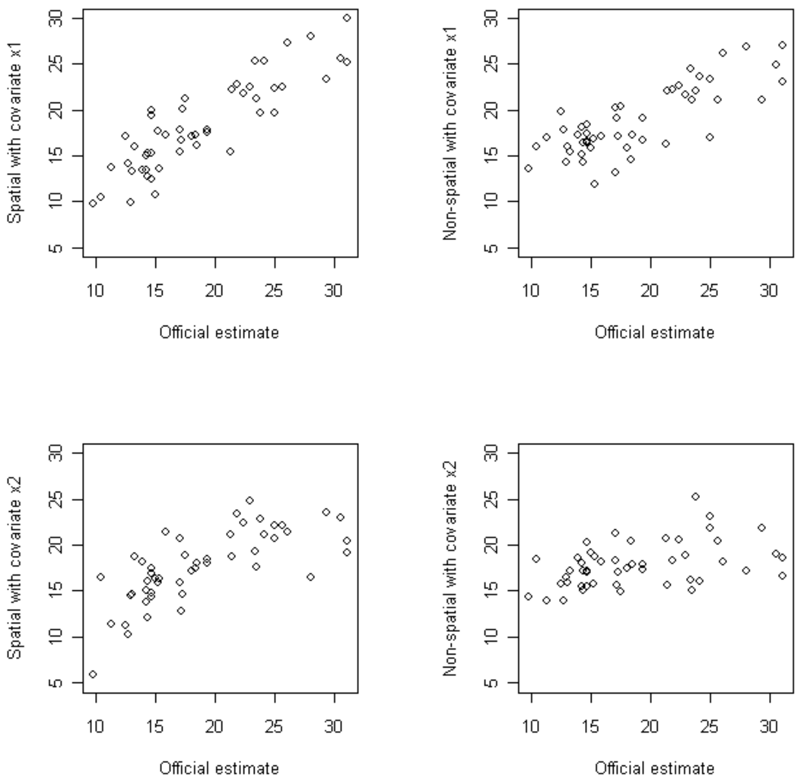

Figure 5: Predicted values of the spatial and non-spatial FH model compared to the official estimates 1993: covariates x1, x2

References

Banerjee, S., Carlin, B. and Gelfand, A. E., (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall/CRC.

Bell, W. R., and Franco, C., (2017). Small Area Estimation – State Poverty Rate Model Research Data Files. Available at https:// www.census.gov/srd/csrmreports/byyear.html [accessed October 22, 2018]

Besag, J., York, J. and Mollie, A., (1991). Bayesian image restoration with two applications in spatial statistics, Annals of the Institute of Statistical Mathematics, 43, 1–59 (With discussion).

Besag, J. and Kooperberg, C. L., (1995). On conditional and intrinsic autoregressions, Biometrika, 82, 733–746.

Chung, H. C. and Datta, G. S., (2022). Bayesian spatial models for estimating means of sampled and non-sampled small areas, Survey Methodology, 48, 2, 463–489.

Clayton, D. and Kaldor, J., (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, Biometrics, 43, 671–681.

Cressie, N. A. C., (1993). Statistics for spatial data, New York: John Wiley & Sons.

Fay, R. E. and Herriot, R. A., (1979). Estimates of Income for Small Places: An Application of James Stein Procedures to Census Data, Journal of the American Statistical Association, 74 (366), 269–277.

Geobugs user manual version 1.2, (2004).

Ghosh, M., Natarajan, K., Stroud, T. W. F. and Carlin B. P., (1998). Generalized Linear Models for Small-Area Estimation, Journal of the American Statistical Association, 93, 273–282.

Kass, R. E. and Wassermann, L., (1996). The selection of prior distributions by formal rules, Journal of the American Statistical Association, 91, 1343–1370.

Kelsall, J. E. and Wakefield, J. C., (1999). Discussion of Bayesian models for spatially correlated disease and exposure data. by Best et al. in Bayesian Statistics 6. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds), Oxford: Oxford University Press

Jiang, J. and Lahiri, P., (2006). Mixed Model Prediction and Small Area Estimation, Test, 15 (1), 1–96.

Maiti, T., (1998). Hierarchical Bayes estimation of mortality rates for disease mapping, Journal of Statistical Planning and Inference, 69, 339–348.

Petrucci, A. Pratesi, M. and Salvati, N., (2005). Geographic Information in Small Area Estimation: Small Area Models and Spatially Correlated Random Area Effects, Statistics in Transition, 3.

Petrucci, A. and Salvati, N., (2006). Small Area Estimation for Spatial Correlation in Watershed Erosion Assessment, Journal of Agricultural, Biological, and Environmental Statistics, 11, 2, 169–182.

Petrucci, A. and Salvati, N., (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects, Statistical Methods and Applications , 17, 113–141.

Petrucci, A. and Salvati, N., (2009). Small Area Estimation in the Presence of Correlated Random Area Effects, Journal of Official Statistics, 25, 1, 37–53.

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

Saei, A. and Chambers, R., (2005). Out of Sample Estimation for Small Areas using Area Level Data, Southampton Statistical Sciences Research Institute Methodology Working Paper M05/11.

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

Sun, D., Tsutakawa, R. K. and Speckman, P. L., (1999). Posterior Distribution of hierarchical models using CAR(1) distributions, Biometrika, 86, 341–350.

You, Y. and Zhou, Q. M., (2011). Hierarchical Bayes small area estimation under a spatial model with application to health survey data, Survey Methodology, 37, 1, 25–37.

Vogt, M., (2010). Bayesian Spatial Modeling: Propriety and Applications to Small Area Estimation with Focus on the German Census 2011. PhD thesis, Universität Trier. DOI: 10.25353/ubtr-xxxx-2ba6-6f2e/.