Modelling Spatially Varying Coefficients via Sparsity Priors

Abstract. Sparsity inducing priors are widely used in Bayesian regression analysis, and seek dimensionality reduction to avoid unnecessarily complex models. An alternative to sparsity induction are discrete mixtures, such as spike and slab priors. These ideas extend to selection of random effects, either *iid* or structured (e.g. spatially structured). In contrast to sparsity induction in mixed models with *iid* random effects, in this paper we apply sparsity priors to spatial regression for area units (lattice data), and to spatial random effects in conditional autoregressive priors. In particular, we consider the use of global-local shrinkage to distinguish areas with average predictor effects from areas where the predictor effect is amplified or diminished because the response-predictor pattern is distinct from that of most areas. The operation and utility of this approach is demonstrated using simulated data, and in a real application to diabetes related deaths in New York counties.

Keywords. Sparsity Inducing Prior; Spatially Varying Coefficients; Shrinkage; Bayesian; Conditional Autoregressive

Introduction

Sparsity inducing priors are widely used in Bayesian regression analysis, and seek dimensionality reduction to avoid unnecessarily complex models that may predict poorly. This involves estimating relatively sparse models where some of the regression coefficients are effectively zero, or have credible intervals including zero. For example, the horseshoe prior (Carvalho et al., 2009) for a collection of regression coefficients is an example of a global-local shrinkage prior, with a global hyperparameter that tends to shrink all parameters towards zero, while predictor specific local hyperparameters (following heavy-tailed half-Cauchy priors) allow some coefficients to avoid the shrinkage (Piironen and Vehtari, 2017). An alternative to sparsity induction are discrete mixtures, or spike and slab priors, which place a positive prior probability on zero values of coefficients. Here binary retention indicators δ_h specific to predictors X_h determine whether such predictors take zero values (Kuo and Mallick, 1998), or have priors that confine predictors to effectively zero values (George and McCulloch, 1993). These retention indicators are in turn defined by retention probabilities ξ_h, which may be preset or unknowns.

These ideas extend to selection of random effects. For example, Frühwirth-Schnatter and Wagner (2011) propose discrete mixture selection to unit-specific random effects with the aim of identifying units that are "average" in the sense of not deviating significantly from the overall mean. This application involves *iid* random effects. By contrast, van der Pas et al. (2014) consider sparsity inducing priors as a way of selecting distinct random effects from average effects, again for *iid* random effects. To illustrate, suppose b_i are normal *iid* random unit intercepts in a normal linear mixed model for univariate longitudinal responses y_{it} for units i and times t, with subject level model

$$y_{it} \sim N(X_{it}\gamma + b_i, \sigma^2). \qquad (1)$$

Here y_{it} is a scalar, $X_{it} = (X_{1it}, ..., X_{pit})$ is a row $p-$vector of predictors (excluding an intercept), γ is a column $p-$vector of regression coefficients, b_i is a normal random effect (random intercept), and the residual variance is σ^2. For the full sample $i = 1, ..., n$ of units, and at time t, the model may be stated as

$$y_t \sim N(X_t\gamma + b, \sigma^2), \qquad (2)$$

where y_t is a column n-vector, X_t is a $n \times p$ matrix, and b is a column n-vector.

One may aim to seek out significant predictor or random effects by shrinking the relevant regression coefficients or random effects towards zero. Applied to the random effects, the sparsity inducing prior (which shrinks less relevant effects towards zero) for the above model is

$$b_i \sim N(0, \tau_b^2 \lambda_i^2), \qquad (3)$$

with global shrinkage parameter τ_b^2, and local shrinkage parameters with half Cauchy priors,

$$\lambda_i \sim C^+(0, 1). \qquad (4)$$

For the global parameter τ_b in (3), a half Cauchy prior may also be used (Carvalho et al., 2009).

The Scope of this Study

Dimensionality reduction is also relevant to *structured* random effects, namely effects dependent over space, time, or space-time. Spatial random effects are the focus in this paper, and in particular sparsity priors aimed at spatially varying regression coefficients in area (spatial lattice) data, as opposed to continuous spatial data (Saveliev et al., 2007). The model framework is therefore considerably distinct from that in equations (1) to (2).

Regarding related papers, Choi and Lawson (2016) consider a disease mapping framework with disease count responses for sets of areas. They propose a discrete mixture selection approach in models allowing spatially varying regression coefficients (SVC models). Thus predictor selection probabilities ξ_{hi} for areas i and predictor h determine whether local regression coefficients γ_{hi} are zero (or effectively zero). The ξ_{hi} follow a spatially structured prior.

By contrast, a continuous shrinkage prior is proposed by Jhuang et al. (2019), for a spatial process framework (Gelfand et al., 2005), with continuous outcomes and continuous space (e.g. point data), and so using spatial priors relevant to that form of spatial data. The shrinkage prior uses the method of Carvalho et al. (2009).

In this paper, we consider spatially structured random effects, sparsity selection, and a discrete spatial framework for spatial lattice data (e.g. disease counts for administrative small areas). Sparsity inducing priors are proposed to distinguish areas with average predictor effects from areas where the predictor effect is amplified or diminished because the response-predictor pattern is distinct from that of most areas. The operation and utility of this approach is demonstrated using simulated data, and a real application to diabetes

related deaths in New York counties.

Spatially Varying Coefficients in Regressions for Spatial Lattice Data

For concreteness, consider a spatially defined disease outcome (for lattice data). The data often consist of disease counts y_i (for areas $i = 1, ..., n$), which are assumed Poisson distributed with means $\mu_i = E_i \rho_i$, where E_i are pre-calculated Poisson offsets for expected event totals, and the ρ_i represent unknown relative disease risk in areas i. The expected event totals are obtained using demographic methods, with national disease rates applied to the populations of the n areas; for example, see Vranckx et al. (2019) and Richardson et al., 2004 (page 1017). The interest is often in the impacts of p ecologically defined area predictors $(X_{1i}, ..., X_{pi})$ on relative risks, and the coefficients are usually assumed constant over space. However spatially varying coefficients may be assumed, this being an acknowledgement of one form of spatial heterogeneity (e.g. Graif and Sampson, 2009).

Assume a Poisson model with means $\mu_i = E_i \rho_i$, and area-specific coefficients $\gamma_i = (\gamma_0, \gamma_{1i}, ..., \gamma_{pi})$ for a spatially varying coefficient (SVC) model (Assunçao, 2003). The relative risk is then defined as
$$log(\rho_i) = \gamma_0 + \gamma_{1i} X_{1i} + ... + \gamma_{pi} X_{pi}, \qquad (5)$$
where γ_0 is an intercept. The same framework may be applied to binomial illness outcomes $y_i \sim Bin(N_i, \pi_i)$, where N_i is the population at risk, and a logit regression links disease probabilities π_i to a spatially varying coefficient specification:
$$logit(\pi_i) = \gamma_0 + \gamma_{1i} X_{1i} + ... + \gamma_{pi} X_{pi}. \qquad (6).$$
In actual estimation, varying coefficients such as γ_{hi} (for area i and predictor h) may be represented, and estimated, as a sum of the central fixed effect (say Γ_h) and zero mean spatial random effects.

The coefficients γ_{hi} ($h = 1, ..., p$) are assumed to be spatially correlated following a suitable prior, typically of conditional autoregressive or CAR form (De Oliveira, 2012). A particular form of this type of prior is proposed by Besag and co-authors (Besag et al., 1991) - see equation (10) below. One may also assume spatially varying intercepts in (5) or (6), for example
$$\gamma_{0i} = \gamma_0 + s_i, \qquad (7)$$
where s_i is a zero mean spatial residual, typically a CAR random effect. However, models which allow for heterogeneity in regression impacts may account for most of the spatial correlation in residuals (Fotheringham et al., 2002, p 117).

Sparsity Prior for Spatially Varying Coefficients

The SVC approach of Assuncao (2003) assumes spatially smooth coefficient effects (with a homogenous smoothing mechanism across all areas). However, in some situations, one may have subregions with distinct outcome-predictor patterns, e.g. where there is ethnic or

social segregation, localized to some areas, may affect predictor effects. In such situations the usual approaches to varying coefficients may be subject to oversmoothing, not allowing for spatial discontinuity (Xu et al., 2017). Lee and Mitchell (2013) refer to step changes in the spatial pattern, although their model applies to spatial residuals. Consider a predictor which is a positive risk factor (one that enhances area disease risk). Then unusual outcome-predictor patterns might be highly elevated risk combined with a below average predictor value; highly elevated disease risk combined with an unusually high predictor value (implying a stronger than average regression effect); or low relative disease risk, but a relatively high predictor value. A suitable shrinkage mechanism would shrink the regression effect γ_{hi} to the average effect Γ_h (areas i and predictors h) for typical areas, but for atypical areas one might seek local adaptivity to express stronger or weaker effects.

Here we propose a model linking local shrinkage parameters to spatially structured effects to allow local adaptivity in regression coefficients. Thus for a single predictor X_{1i} ($p = 1$) and areas i, we assume a regression coefficient prior

$$\gamma_{1i} \sim N(\Gamma_1, \tau_{\gamma_1}^2 \lambda_{1i}^2), \qquad (8)$$

where τ_{γ_1} is a global shrinkage parameter, the λ_{1i} are local shrinkage parameters, and Γ_1 is the region-wide predictor effect. The local parameters are half Cauchy according to

$$\lambda_{1i} \sim C^+(r_{1i}, 1), \qquad (9)$$

where $r_1 = (r_{11}, ... r_{1n})$ is a vector of spatially structured random effects. Here the spatial effects follow the prior of Besag et al. (1991, page 8) with conditional normal form

$$r_{1i} \sim N(R_{1i}, \tau_{r_1}/L_r) \qquad (10)$$

where τ_{r1} is a conditional variance, L_r is the number of areas adjacent to area i (termed the locality of area i), and R_{1i} is the average of the effects r_{1j} in this locality.

Adaptivity to unusual outcome-predictor patterns implies larger λ_{1i}^2 in some areas (and hence a more diffuse prior for γ_{1i}). Lower λ_{1i}^2 apply to areas following the typical regression pattern, with the regression coefficient shrunk towards Γ_1. So the parameters λ_{1i} add extra information to the usual inferences made with disease data and observed predictors, providing a measure of outlyingness in the outcome-predictor space.

Equivalently, using the normal/gamma representation of a Cauchy variable

$$\lambda_{1i}^2 = \eta_{1i}^2/\phi_{1i}^{0.5}, \qquad (11)$$
$$\eta_{1i} \sim N^+(r_{1i}, 1), \qquad (12)$$
$$\phi_{1i} \sim Gamma(a_{\phi_1}, b_{\phi_1}), \qquad (13)$$

with $a_{\phi_1} = b_{\phi_1} = 0.5$. Note that the assumed parameterisation of the gamma density with random variable u is

$$p(u|a, b) = \frac{b^a}{\Gamma(a)} u^{a-1} e^{-bu}. \qquad (14)$$

Under a Bayesian hierarchical approach (Gelfand, 2012), the likelihood is the first stage of a three stage specification. For the Poisson approach with a single predictor, area relative risks defined by

$$log(\rho_i) = \gamma_0 + \gamma_{1i}X_{1i}, \qquad (15)$$

and means $\rho_i E_i$, the likelihood for area i conditions on the data (E_i, X_{1i}), the hyperparameters $\theta = (\beta, \Gamma_1, \tau_{\gamma_1}, \tau_{r_1}, \gamma_0)$, and random effects $(\gamma_{1i}, \lambda_{1i}, r_{1i})$ as in $(8) - (10)$, namely

$$p(y_i|E_i, X_{1i}, \theta, \gamma_{1i}, \lambda_{1i}, r_{1i}) = \qquad e^{-\rho_i E_i}(\rho_i E_i)^{y_i}/y_i! \qquad (16)$$

For binomial data with disease probability

$$logit(\pi_i) = \gamma_0 + \gamma_{1i}X_{1i}, \qquad (17)$$

the relevant conditional likelihood is

$$p(y_i|N_i, X_{1i}, \theta, \gamma_{1i}, \lambda_{1i}, r_{1i}) = \binom{N_i}{y_i}\pi_i^{y_i}(1 - \pi_i)^{N_i - y_i} \qquad (18)$$

For multiple predictors X_{hi} $(h = 1, ..., p)$ and areas i, one may similarly specify global-local shrinkage parameters specific to predictor h, and area-predictor specific local hyperparameters, namely

$$\gamma_{hi} \sim N(\Gamma_h, \tau_{\gamma_h}^2 \lambda_{hi}^2), \qquad (19)$$
$$\lambda_{hi} \sim C^+(r_{hi}, 1), \qquad (20)$$

with p sets of CAR spatial effects r_{hi} determining the level of local adaptivity. As above, the spatial effects follow the CAR prior of Besag et al. (1991) with conditional normal prior

$$r_{hi} \sim N(R_{hi}, \tau_{r_h}/L_r) \qquad (21)$$

where τ_{r_h} is a conditional variance, and R_{hi} is the average of the effects r_{hj} in the locality of area i.

Application to Simulated Data

We consider a simulation based on the spatial structure of the 62 counties of New York state and a Poisson likelihood structure as in equations (15) and (16). Expected disease counts E_i for counties i are simulated as gamma with shape 30 and rate 2, and values of a single predictor X_{1i} are simulated as N(0,1). The simulation introduces contamination into the generation of spatially varying coefficients. Thus the regression term defining area relative risks is obtained as

$$log(\rho_i) = -0.5 + (\gamma_{1i} + a_i)X_{1i}, \qquad (22)$$

where $\gamma_0 = -0.5$, and the γ_{1i} are CAR random spatial effects as per Besag et al. (1991), with mean $\Gamma_1 = 1$ and precision $1/\tau_{\gamma 1} = 50$. Additive effects a_i are uniform U(-0.75,0.75) and apply only to a random subsample of counties. So for areas not in the subsample, the relative risk is specified as $log(\rho_i) = -0.5 + \gamma_{1i}X_{1i}$. Two contamination rates are considered to obtain the random subsamples: 20% and 33%. So in this subset of counties, the impact of the covariate X_1 is shifted away from the normal CAR terms γ_{1i} by additive uniform effects. Disease counts are then simulated as $y_i \sim Poisson(E_i\rho_i)$.

We simulate 100 datasets and compare three model formulations with regard to (a) fit to each simulated dataset, measured by the leave one out information criterion (LOO-IC) of Vehtari et al. (2017) (b); recovery of the underlying parameters, specifically the mean $\overline{\gamma}_1$ of the simulated area regression parameters $\gamma_{1i} + a_i$; thus the true value is taken as recovered if the 95% credible interval of the estimated parameter contains the true value; and (c) predictive discrepancies, namely tail values (over 0.95 or under 0.05) for

$Pr(y_{rep,i} > y_i|y) + 0.5 \Pr(y_{rep,i} = y_i|y),$ where $y_{rep,i}$ are posterior predicted counts (Marshall and Spiegelhalter, 2003).

In a first analysis of the simulated data, we consider three forms of model addressing regression coefficient estimation. The first is a simple Poisson regression (model A), with a homogenous predictor effect:

$$log(\rho_i) = \gamma_0 + \gamma_1 X_{1i}, \qquad (23)$$

with regression coefficients assigned N(0,10) priors. This option is denoted as HPE. The second (model B) allows spatial variation in the predictor effect, as per the usual spatially varying coefficient model (Assuncao, 2003) with varying coefficients γ_{1i}, but with no sparsity or shrinkage mechanism:

$$log(\rho_i) = \gamma_0 + \gamma_{1i} X_{1i}. \qquad (24)$$

Here the γ_{1i} follow the CAR prior of Besag et al. (1991) with precision parameter $1/\tau_{\gamma_1}$ assigned a Ga(0.5,0.0005) prior. This is denoted as the SVC option.

The third formulation (model C) uses a sparsity inducing prior (SIP for short), as per:

$$log(\rho_i) = \gamma_0 + \gamma_{1i} X_{1i}, \qquad (25)$$
$$\gamma_{1i} \sim N(\Gamma_1, \tau_{\gamma_1}^2 \lambda_{1i}^2),$$
$$\tau_{\gamma_1} \sim C^+(0, 1),$$
$$\lambda_{1i}^2 = \eta_{1i}^2 / \phi_{1i}^{0.5},$$
$$\phi_{1i} \sim Gamma(0.5, 0.5),$$
$$\eta_{1i} \sim N^+(r_{1i}, 1),$$
$$r_{1i} \sim BYM - CAR(\omega_{r_1}),$$
$$\omega_{r_1} \sim Ga(0.5, 0.0005),$$

with the r_{1i} are zero-mean spatial effects following the CAR prior of Besag et al. (1991) (BYM for the authors) - see equation (10) - and $\omega_{r_1} = 1/\tau_{r_1}$ is the precision parameter.

In a second form of analysis of the simulated datasets, we allow, as in (7), for varying intercepts using random effects. In the above options (HPE, SVC, SIP) the residuals follow a scheme proposed by Besag et al. (1991, page 7) involving both a spatially correlated CAR effect and an unstructured (*iid*) random error, with the latter mainly serving to account for Poisson overdispersion, and the former representing spatially correlated but unobserved risk factors. This scheme is known as the convolution prior (Rodrigues and Assunção, 2012). Thus model A (HPE) becomes:

$$log(\rho_i) = \gamma_0 + \gamma_1 X_{1i} + s_i + u_i, \qquad (26)$$

where the s_i have a Besag et al. (1991) CAR prior, with precision parameter ω_s assigned a Ga(0.5,0.0005) prior. The u_i are iid random effects, with $u_i \sim N(0, 1/\omega_u)$, and precision ω_u assigned a Ga(0.5,0.0005) prior. The model B (SVC) option becomes

$$log(\rho_i) = \gamma_0 + \gamma_{1i} X_{1i} + s_i + u_i, \qquad (27)$$

where both the γ_{1i} and s_i follow the Besag et al. CAR (1991) prior, with precision parameters assigned Ga(0.5,0.0005) priors. The model C (SIP) option becomes:

$$log(\rho_i) = \gamma_0 + \gamma_{1i} X_{1i} + s_i + u_i, \qquad (28)$$

$$\gamma_{1i} \sim N(\Gamma_1, \tau_{\gamma 1}^2 \lambda_{1i}^2),$$
$$\tau_{\gamma 1} \sim C^+(0, 1),$$
$$\lambda_{1i}^2 = \eta_{1i}^2 / \phi_{1i}^{0.5},$$
$$\phi_{1i} \sim Gamma(0.5, 0.5),$$
$$\eta_{1i} \sim N^+(r_{1i}, 1),$$
$$r_{1i} \sim BYM - CAR(\omega_{r1}),$$
$$\omega_{r1} \sim Ga(0.5, 0.0005),$$
$$s_i \sim BYM - CAR(\omega_s),$$
$$\omega_s \sim Ga(0.5, 0.0005).$$

where r_{1i} and s_i are zero-mean spatial effects following the Besag et al. (1991) CAR prior, and the u_i are *iid* random normal with mean 0.

Table 1 compares the performance of the three model formulations (HPE,SVC,SIP), with coefficient estimation using spatially varying coefficient models (upper panel) and varying coefficient plus spatial residuals and iid random effects (lower panel). Estimation uses rstan (Carpenter et al., 2017; Morris et al., 2019), with two chains and 2000 iterations in estimation of each model in each of the 100 simulations.

It can be seen that the SIP option performs relatively well on all evaluation criteria across the board, with better parameter recovery, fewer poorly predicted disease counts, and better fit as measured by LOO-IC criteria. The performance of the SVC and HPE models is much improved when varying residuals are added. But for models for spatially varying effects alone, as under (23)-(25), without the assistance of varying residuals or iid effects, the SIP model clearly performs better. The SIP option retains the lower LOO-IC even when varying residuals are added to the models, but for the SIP option there is virtually no change in performance between the varying coefficients option in (25) and the model with varying residuals in (28). For example, the LOO-IC is very similar. Hence the sparsity inducing prior, as applied only to regression coefficients, is competitive with more heavily parameterised model variants.

Real Data Application: Diabetes and Poverty in New York Counties

A real data application involves binomial data $y_i \sim Bin(N_i, \pi_i)$ on diabetes related deaths y_i (International Classification of Diseases, ICD10 E10-E14) in populations N_i of 62 New York counties for the period 2013-17 (see Figure 1). The predictor is the 2015 county percentage poverty rate. Two models are compared to explain diabetes mortality rate variations between counties. The first is an SVC model with

$$logit(\pi_i) = \gamma_0 + \gamma_{1i} X_{1i} + s_i + u_i, \qquad (29)$$

where γ_{1i} and s_i follow zero-mean CAR priors as per Besag et al. (1991), and the u_i are random normal *iid*. The second uses the sparsity inducing prior with

$$logit(\pi_i) = \gamma_0 + \gamma_{1i} X_{1i} + s_i + u_i, \qquad (30)$$

$$\gamma_{1i} \sim N(\Gamma_1, \tau_1^2 \lambda_{1i}^2),$$
$$\tau_1 \sim C^+(0,1),$$
$$\lambda_{1i}^2 = \eta_{1i}^2 / \phi_{1i}^{0.5},$$
$$\phi_{1i} \sim Gamma(0.5, 0.5),$$
$$\eta_{1i} \sim N^+(r_{1i}, 1),$$
$$r_{1i} \sim BYM - CAR(\omega_r),$$
$$\omega_r \sim Ga(0.5, 0.0005),$$
$$s_i \sim BYM - CAR(\omega_s),$$
$$\omega_s \sim Ga(0.5, 0.0005),$$
$$u_i \sim N(0, 1/\omega_u),$$
$$\omega_u \sim Ga(0.5, 0.0005).$$

Both modesl show a generally positive association between poverty and diabetes mortality over the 62 counties. The New York state-wide parameter Γ_1 has posterior mean (95% CRI) of 3.24 (1.80, 4.90) under the SIP model, and 2.87 (1.16, 4.75) under the SVC model. However, there is considerable variation in county specific coefficients γ_{1i}.

Table 2 shows the estimated coefficients γ_{1i} under the SIP and SVC models in according to descending order of the estimated λ_{1i}^2 in the SIP model. The λ_{1i}^2 are highest for counties where the outcome-predictor pattern is unusual compared to the regression for typical counties (e.g. a high diabetes rate coupled with a relatively low poverty, given the overall positive relationship between the disease and risk factor). Diabetes death rates in Table 2 are the ratios y_i/N_i times 100,000. The respective LOO-IC for the SVC and SIP models are 559 and 554.

Distinct predictor-outcome profiles are associated with spatial discontinuites in the pattern of the γ_{1i}, corresponding to areas with high λ_{1i}^2. Unusual predictor-outcome profiles can be assessed by comparing county poverty and mortality to the state wide averages: the average poverty percentage across all counties is 14.5%, and the average mortality rate is 24.5 per 100,000. The highest λ_{1i}^2 is for Tompkins county, where poverty is considerably above average, but despite the excess poverty, diabetes mortality is below average. The regression coefficient for Tompkins under the SIP model is accordingly shrunk towards zero, with posterior mean for γ_{1i} of 0.94, as compared to a posterior mean under the SVC model of 2.67. The second highest λ_{1i}^2 is for Monroe county with around average poverty, but below average diabetes mortality. The third and fourth highest λ_{1i}^2 (for Herkimer and Genesee counties) have below average poverty, but much above diabetes mortality. Accordingly posterior means for γ_{1i} are adjusted upward under the SIP model to reflect the observed diabetes levels in these counties. The LOO-IC casewise values, the individual area contributions to the LOO-IC (Vehtari et al., 2017), for the five counties with highest λ_{1i}^2 are all lower under the SIP model.

By contrast, low values of λ_{1i}^2 apply to "typical" counties (in terms of the predictor-outcome relationship), and for these counties γ_{1i} values tend to be shrunk towards the global value

Γ_1 under the SIP prior. For example, the counties with the lowest five λ_{1i}^2 have posterior mean γ_{1i} under the SIP prior between 3.17 and 3.38, and the standard deviation of these five means is 0.08. By contrast, the SVC model has mean γ_{1i} varying from 2.73 to 3.03 for these counties, and their standard deviation is 0.11. Figure 2 shows how λ_{1i}^2 values are spatially clustered, while Figures 3 and 4 show the contrasting pattern in the posterior mean γ_{1i} under the SVC and SIP models.

Conclusion

Impacts of risk factors on spatial variations in disease may be important for policy purposes. For example, assume that the overall relationship between a disease and certain risk factor, such as poverty, is positive: high levels of poverty are associated with higher disease levels. In some areas, however, there may be high disease risk but relatively low poverty. Such areas may be disadvantaged if interventions are based simply on poverty rates.

In terms of implications for spatial models and disease mapping, spatially varying coefficient models with a single global shrinkage parameter may be insufficiently adaptive to such unusual patterning in outcome-risk factor associations. By contrast, an adaptive global-local shrinkage mechanism, as part of a sparsity inducing prior, has the potential to identify distinctive predictor-outcome associations not apparent under the usual spatially varying coefficients methodology. This paper has developed a methodology for small areas, applicable using rstan, that implements an adaptive procedure for such situations.

References

Assunçao, R (2003) Space varying coefficient models for small area data. Environmetrics, 14(5), 453-473.

S. Banerjee, B.Carlin, A Gelfand (2014) Hierarchical Modeling and Analysis for Spatial Data. Second Edition. CRC Press/Chapman & Hall.

Besag J., York J, Mollié A (1991) Bayesian image restoration with two applications in spatial statistics. Ann. Inst. Statist. Math., 43 (1991), pp. 1-59

Carpenter, B., Gelman, A., Hoffman, M, Lee, D., Goodrich, B., Betancourt, M.,Brubaker M, Guo J, Li P,Riddell, A. (2017). Stan: A probabilistic programming language. Journal of Statistical Software, 76(1)

Carvalho, C, Polson N, Scott J (2009) Handling Sparsity via the Horseshoe. Proceedings of Machine Learning Research 5, 73–80

De Oliveira, V. (2012) Bayesian analysis of conditional autoregressive models. Annals of the Institute of Statistical Mathematics, 64(1), 107-133.

Fotheringham A, Brundson C, Charlton M. (2002) Geographically Weighted Regression: the Analysis of Spatially Varying Relationships. West Sussex, England: Wily

Frühwirth-Schnatter S, Wagner H (2011) Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data, In: J. Bernardo, J Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West (eds) Bayesian Statistics 9,pp 165-200.

Gelfand, A (2012). Hierarchical modeling for spatial data problems. Spatial Statistics, 1, 30-39.

Gelfand, A., Banerjee, S., Gamerman, D. (2005) Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics, 16(5), 465-479

George, E, McCulloch, R (1993) Variable selection via Gibbs sampling. Journal of the American Statistical Association 88 881-889.

Jhuang, A, Fuentes, M., Jones, J, Esteves, G., Fancher, C, Furman, M., Reich, B (2019) Spatial signal detection using continuous shrinkage priors. Technometrics, 61(4), 494-506

Graif, C., Sampson, R (2009) Spatial heterogeneity in the effects of immigration and diversity on neighborhood homicide rates. Homicide Studies, 13(3), 242-260.

Kuo, L., Mallick, B. (1998). Variable selection for regression models. Sankhyā: The Indian Journal of Statistics, Series B, 65-81.

Lee, D., Mitchell, R. (2013) Locally adaptive spatial smoothing using conditional autoregressive models. Journal of the Royal Statistical Society: Series C (Applied Statistics), 62(4), 593-608.

Leroux B, Lei X, Breslow N (2000) Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence. In M Halloran, D Berry (eds.), Statistical Models in Epidemiology, the Environment and Clinical Trials, pp. 179-191. Springer-Verlag, New York

Marshall, E, Spiegelhalter, D. J. (2003). Approximate cross-validatory predictive checks in disease mapping models. Statistics in Medicine, 22(10), 1649-1660.

Morris M., Wheeler-Martin K, Simpson D., Mooney, S, Gelman, A, DiMaggio, C. (2019) Bayesian hierarchical spatial models: Implementing the Besag-York-Mollié model in stan. Spatial and Spatio-temporal Epidemiology, 31, 100301

Piironen, J., Vehtari, A. (2017) Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics, 11(2), 5018-5051

Rodrigues, E, Assunção, R (2012) Bayesian spatial models with a mixture neighborhood structure. Journal of Multivariate Analysis, 109, 88-102.

Richardson, S, Thomson, A, Best, N, Elliott, P (2004) Interpreting posterior relative risk estimates in disease-mapping studies. Environmental Health Perspectives, 112(9), 1016-1025.

Saveliev, A., Mukharamova, S, Zuur, A (2007) Analysis and modelling of lattice data. In Analysing Ecological Data (pp. 321-339). Springer, New York, NY.

Van der Pas, S, Kleijn, B, Van Der Vaart, A (2014) The horseshoe estimator: Posterior concentration around nearly black vectors. Electronic Journal of Statistics, 8(2), 2585-2618

Vehtari, A., Gelman, A, Gabry, J. (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing. 27(5), 1413–1432

Vranckx, M., Neyens, T., Faes, C. (2019) Comparison of different software implementations for spatial disease mapping. Spatial and Spatio-temporal Epidemiology, 31, 100302.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research, 11, 3571-3594.

Table 1 Performance Evaluation on Simulated Data

Models with Spatially Varying Coefficients Only	Contamination Rate	
	20%	33%
Model	% of Simulations with mean γ_1 recovered	% of Simulations with mean γ_1 recovered
HPE	54	50
SVC	92	93
SIP	99	98
	Average Number of Observations (from 62) with Tail Predictive p-values	Average Number of Observations (from 62) with Tail Predictive p-values
HPE	16.0	11.9
SVC	3.3	3.4
SIP	1.3	1.0
	Leave One Out Information Criterion (LOO-IC)	Leave One Out Information Criterion (LOO-IC)
HPE	429.6	434.6
SVC	333.1	336.7
SIP	328.7	331.2

Varying Coefficients and Intercepts	Contamination Rate	
	20%	33%
Model	% of Simulations with mean γ_1 recovered	% of Simulations with mean γ_1 recovered
HPE	96	96
SVC	95	95
SIP	99	97
	Average Number of Observations (from 62) with Tail Predictive p-values	Average Number of Observations (from 62) with Tail Predictive p-values
HPE	1.4	1.0
SVC	1.6	1.2
SIP	0.7	0.7
	Leave One Out Information Criterion (LOO-IC)	Leave One Out Information Criterion (LOO-IC)
HPE	334.3	337.8
SVC	331.9	335.3
SIP	328.7	331.4

Table 2 Model Estimates Compared, Diabetes Mortality in New York Counties

County Name	County ID	y_i, Diabetes Deaths	Mortality rate (per 100 Thousand)	Poverty Rate (Percent)	λ^2_{1i}	SVC estimate of $\gamma_{1,}$	SIP estimate of $\gamma_{1,}$	$y_{rep,i}$ (SVC model)	$y_{rep,i}$ (SIP model)	Piecewise LOO, SVC model	Piecewise LOO, SIP model
Tompkins	55	71	13.6	20.1	12.4	2.67	0.94	82.7	75.9	11.9	9.8
Monroe	28	605	16.2	14.8	8.2	2.74	1.88	614.4	612.7	11.4	10.5
Herkimer	22	108	34.2	12.7	4.2	3.04	4.07	101.4	100.1	10.3	10.1
Genesee	19	110	37.4	13.4	4.1	3.04	3.93	102.2	103.7	10.9	10.4
Albany	1	270	17.5	12.6	3.6	2.81	2.94	275.8	275.3	10.1	9.5
Livingston	26	55	17.1	13.2	3.4	2.84	2.72	61.6	62.1	9.2	9.6
Rockland	44	223	13.7	14	3.3	2.60	2.89	228.2	227.3	9.3	9.9
Hamilton	21	*	*	11.1	3.3	2.96	3.04	5.1	5.2	4.4	4.5
Chemung	8	78	17.9	14.9	3.3	2.76	2.90	82.0	82.7	8.3	8.9
Otsego	39	120	39.5	16.8	3.2	3.02	3.77	112.3	113.2	11.3	10.1
Clinton	10	95	23.4	17.5	3.2	2.87	3.02	98.6	100.0	8.0	8.8
Niagara	32	368	34.6	15.5	3.1	3.03	3.44	363.2	363.1	10.1	10.1
Essex	16	72	37.5	12.2	3.1	3.08	3.78	65.0	64.4	9.4	10.4
Wyoming	61	62	30.2	13.3	2.8	3.02	3.44	60.0	60.1	7.5	7.2
Delaware	13	75	32.6	16.9	2.8	2.94	3.56	71.1	70.3	8.3	8.3
Wayne	59	134	29.3	12.2	2.7	2.95	3.54	128.8	128.6	9.9	9.3
Allegany	2	100	42.1	16.8	2.7	3.05	3.66	93.1	93.2	9.8	9.8
Orleans	37	44	21.1	14.2	2.5	2.87	3.03	46.7	47.3	7.0	7.9
Saratoga	46	188	16.6	6.4	2.4	2.93	3.02	192.8	193.7	9.5	9.4
Schenectady	47	196	25.3	12	2.4	2.95	3.35	193.2	191.4	9.0	9.3
Lewis	25	43	31.9	14	2.3	2.96	3.47	39.4	39.5	7.4	7.8
New York	31	1404	17.1	17.6	2.3	2.68	3.09	1406.7	1398.5	11.2	12.4
Seneca	50	40	22.9	14.1	2.2	2.85	3.21	40.1	41.4	6.7	6.5
Warren	57	103	31.8	11.3	2.2	3.07	3.46	98.1	98.0	9.0	9.0
Westchester	60	697	14.3	10.1	2.2	2.64	3.22	698.2	699.4	10.5	10.7
Suffolk	52	1090	14.6	7.8	2.1	2.67	3.24	1090.4	1096.4	11.3	11.9
Jefferson	23	108	18.5	14.1	2.1	2.80	3.05	112.7	114.7	8.7	9.0
Cayuga	6	70	17.9	12.7	2.1	2.79	3.00	73.6	74.7	7.9	8.2
Greene	20	48	20.1	16.4	2.1	2.82	3.12	51.2	52.0	7.4	7.4
Ontario	35	112	20.4	9.6	2.0	2.88	3.23	112.2	111.8	8.4	7.9
Putnam	40	69	13.9	6	2.0	2.70	3.19	70.7	68.8	7.6	7.7
Schuyler	49	27	29.6	12.5	2.0	2.88	3.38	23.4	23.0	7.2	7.3
Columbia	11	65	21.1	13.6	2.0	2.84	3.18	66.0	67.0	7.4	7.2
Tioga	54	57	23.1	11.7	2.0	2.85	3.31	55.4	55.8	7.5	7.0
Yates	62	31	24.7	13.6	2.0	2.89	3.25	30.5	30.1	6.2	6.2
Richmond	43	571	24.0	14.2	1.9	2.89	3.40	568.9	569.2	10.4	10.3
Washington	58	97	31.2	13.3	1.9	3.03	3.38	93.8	91.5	8.4	8.5
St. Lawrence	45	140	25.3	18.5	1.9	2.88	3.15	142.6	142.9	8.1	8.7
Nassau	30	918	13.5	6.1	1.9	2.65	3.30	919.6	920.7	10.9	10.5
Sullivan	53	102	27.0	18.1	1.9	2.86	3.35	100.4	100.2	8.0	8.1
Oneida	33	351	30.2	18	1.9	2.93	3.26	349.2	350.5	10.5	9.3
Oswego	38	156	26.0	17.4	1.8	2.86	3.21	156.4	154.7	8.8	8.3
Orange	36	331	17.5	12.1	1.8	2.75	3.18	332.3	332.3	9.5	9.7
Chautauqua	7	174	26.6	17.2	1.8	2.92	3.13	176.3	178.6	8.9	8.8
Franklin	17	81	31.7	18.3	1.8	2.95	3.29	79.7	79.6	8.0	7.5
Steuben	51	130	26.6	15.4	1.8	2.90	3.28	129.2	129.0	8.5	8.4
Montgomery	29	94	37.9	18.4	1.8	3.00	3.44	89.3	90.8	9.2	8.8
Dutchess	14	249	16.8	10.4	1.8	2.76	3.13	251.1	254.0	9.2	9.5
Madison	27	73	20.3	12.9	1.8	2.86	3.12	75.1	74.4	7.9	7.5
Ulster	56	218	24.2	14	1.7	2.87	3.35	216.2	214.0	9.0	9.0
Bronx	3	1777	24.5	30.3	1.7	2.71	3.04	1780.7	1778.3	11.7	11.4
Rensselaer	42	193	24.1	12	1.7	2.94	3.30	191.6	190.6	9.1	9.0
Fulton	18	74	27.4	17.9	1.7	2.93	3.17	75.0	76.1	7.6	7.7
Broome	4	287	29.3	17.7	1.7	2.90	3.32	284.7	283.1	9.8	9.4
Cortland	12	55	22.7	16.2	1.7	2.82	3.17	56.0	56.5	7.1	6.9
Onondaga	34	505	21.6	14.6	1.7	2.83	3.12	506.6	507.0	10.1	9.8
Erie	15	1354	29.4	15.6	1.7	2.98	3.23	1351.6	1354.3	11.3	11.1
Schoharie	48	33	21.0	13.7	1.6	2.88	3.18	34.8	35.4	6.6	6.3
Chenango	9	54	22.1	14.3	1.6	2.87	3.17	55.3	55.8	7.1	6.9
Cattaraugus	5	148	37.9	18	1.6	3.03	3.38	144.4	142.8	9.0	9.1
Queens	41	2090	17.9	13.9	1.5	2.73	3.27	2090.8	2089.2	12.2	11.4
Kings	24	3221	24.5	22.3	1.5	2.78	3.24	3220.2	3223.9	12.2	12.2

* Count under 10

Figure 1 Counties in New York State

Figure 2 Posterior Mean λ_{1i}^2

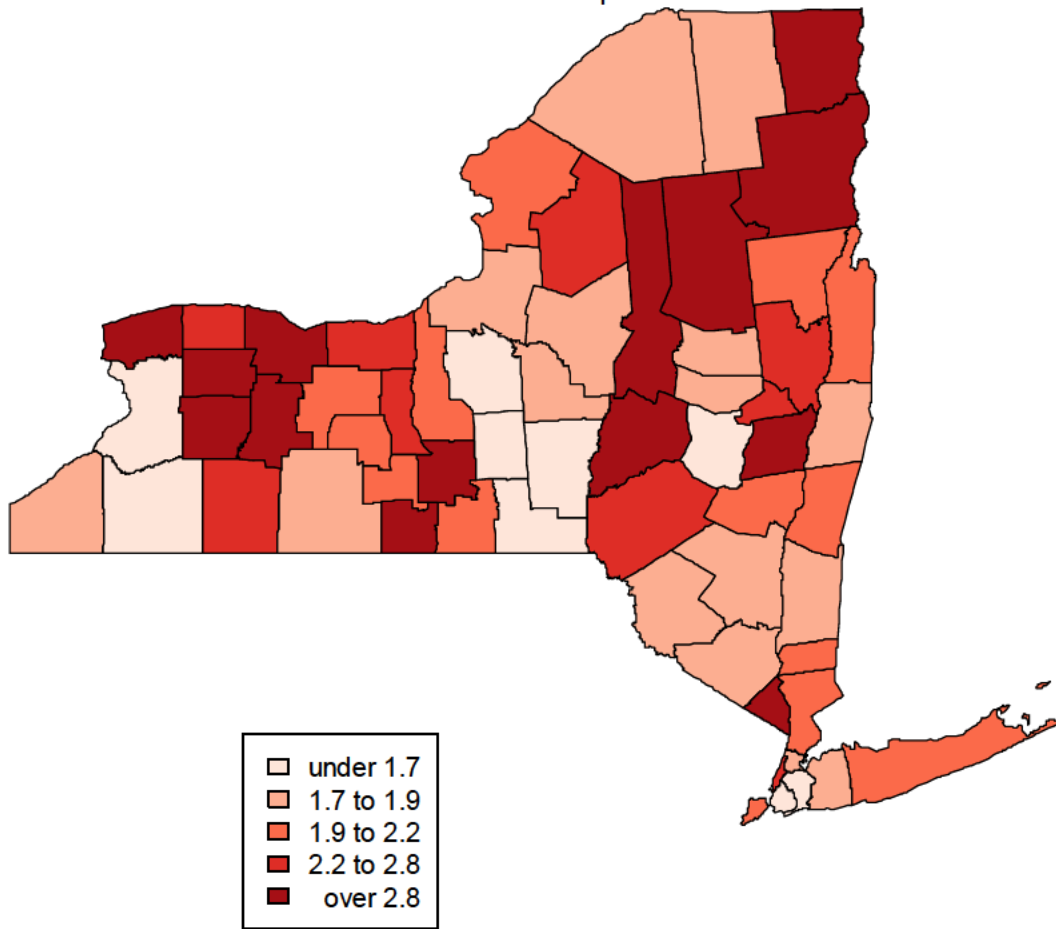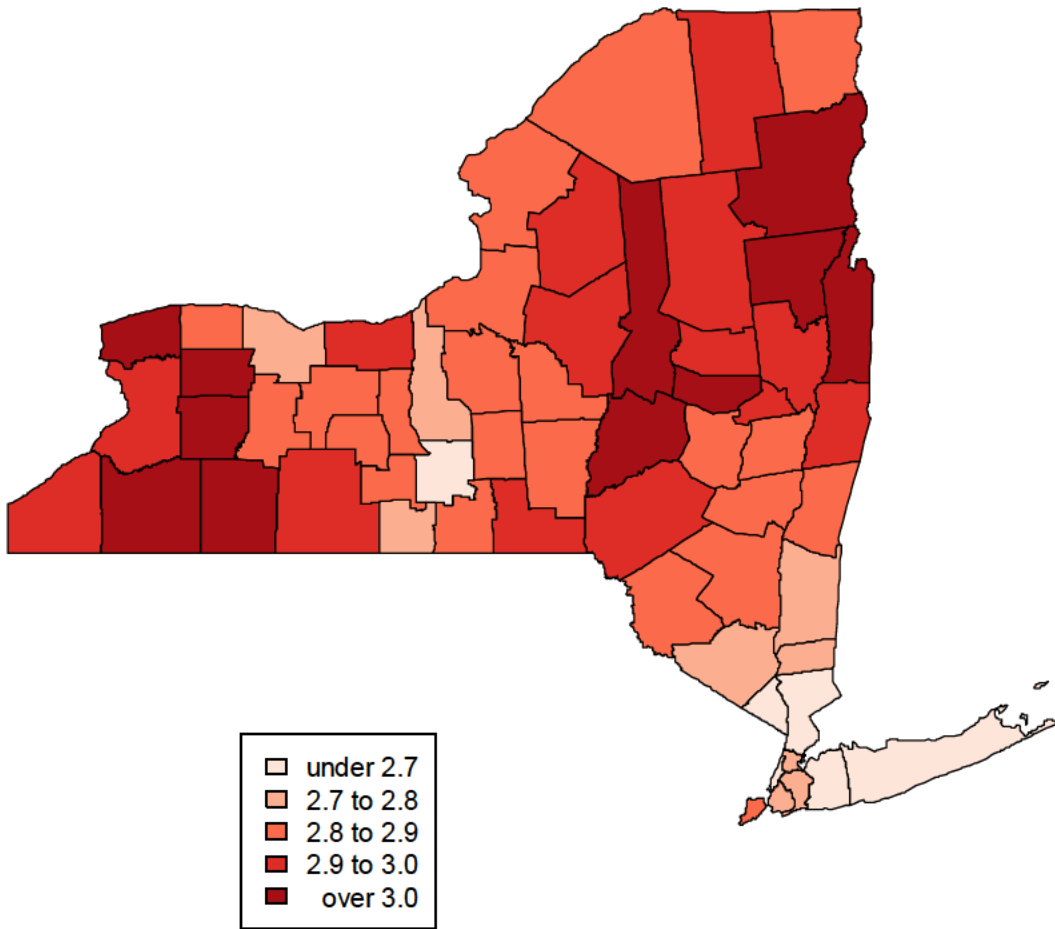|	
□	under 1.7
	1.7 to 1.9
	1.9 to 2.2
	2.2 to 2.8
	over 2.8

Figure 3 Posterior Mean SVC γ_{1i}

Figure 4 Posterior Mean SIP γ_{1i}

☐	under 3.1
	3.1 to 3.2
	3.2 to 3.3
	3.3 to 3.4
	over 3.4