Local Variable Selection and Parameter Estimation of Spatially Varying Coefficient Regression Models

Wesley Brooks

---

1. Introduction

Whereas the coefficients in traditional linear regression are scalar constants, the coefficients in a varying coefficient regression (VCR) model are functions - often *smooth* functions - of some effect modifying variable (Hastie and Tibshirani, 1993). When the effect modifying variable represents location in a spatial domain, a VCR model implies a spatially varying coefficient regression (SVCR) model wherein that the regression coefficients vary over space. Statistical inference for the coefficients as functions of location in an SVCR model is more complicated than estimating the coefficients in a traditional linear regression model where the coefficients are constant across the spatial domain. My research concerns the development of new methodology for the analysis of spatial data using SVCR.

The methodology described herein is applicable to geostatistical data and areal data. Let $\mathcal{D}$ be a spatial domain on which data is collected. For geostatistical data, let $\boldsymbol{s}$ denote a location in $\mathcal{D}$. Let a univariate spatial process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ and a possibly multivariate spatial process $\{\boldsymbol{X}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ denote random fields of the response and the covariates, respectively. For $i = 1, \ldots, n$, let $\boldsymbol{s}_i$ denote the sampling location in $\mathcal{D}$ of the ith observation of the response and the covariates. Let the observed data be denoted $\{y(\boldsymbol{s}_i), \boldsymbol{x}(\boldsymbol{s}_i)\}$, $i = 1, \ldots, n$. Then the data are a realization of the random fields at the sampling locations $\{Y(\boldsymbol{s}_i), \boldsymbol{X}(\boldsymbol{s}_i)\}$ for $i = 1, \ldots, n$.

For areal data, the spatial domain $\mathcal{D}$ is partitioned into n regions $\{D_1, \ldots, D_n\}$ such that $\mathcal{D} = \bigcup_{i=1}^{n} D_i$. In the case of areal data, the random variables $\{Y(D_i), \boldsymbol{X}(D_i)\}$ are defined for regions instead of for point locations; population and spatial mean temperature are examples of areal data. The analytical method described herein can be applied to areal data if they are recast as geostatistical data by assuming that the data are point-referenced to a representative location of each region, such as the centroid. That is, $\{\boldsymbol{X}(\boldsymbol{s}_i), Y(\boldsymbol{s}_i)\}$ where $\boldsymbol{s}_i$ is the centroid of D_i for $i = 1, \ldots, n$.

Common practice in the analysis of geostatistical and areal data is to model the response variable with a spatial linear regression model consisting of the sum of a fixed mean function, a spatial random effect, and random error all on domain $\mathcal{D}$, as in:

$$Y(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})'\boldsymbol{\beta} + W(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \tag{1}$$

where $\boldsymbol{X}(\boldsymbol{s})'\boldsymbol{\beta}$ is the mean function consisting of a vector of covariates $\boldsymbol{X}(\boldsymbol{s})$, and a vector of regression coefficients $\boldsymbol{\beta}$. The random error $\varepsilon(\boldsymbol{s})$ denotes white noise such that the errors are independent and identically distributed with mean zero and variance σ^2, while the random component $W(\boldsymbol{s})$ denotes a mean-zero, second-order stationary random field that is independent of the random error. The mean function captures the large-scale systematic trend of the response, the spatial random field $W(\boldsymbol{s})$ can be thought of as a small-scale spatial random effect, and the error term $\varepsilon(\boldsymbol{s})$ captures micro-scale variation (Cressie, 1993).

It is common to pre-specify the form of a covariance function for the spatial random effect $W(\boldsymbol{s})$ (Diggle and Ribeiro, 2007). For example, the exponential covariance function (a special case of the Matérn class of covariance functions) has the form

$$\mathrm{Cov}(W(\boldsymbol{s}), W(\boldsymbol{t})) = \sigma^2 \exp\left\{-\phi^{-1}\delta(\boldsymbol{s}, \boldsymbol{t})\right\} \tag{2}$$

where σ^2 is a variance parameter, ϕ is a range parameter, and $\delta(\boldsymbol{s}, \boldsymbol{t})$ is the Euclidean distance between locations $\boldsymbol{s}$ and $\boldsymbol{t}$. The general form of a covariance function in the Matérn class is

$$\text{Cov}(W(\boldsymbol{s}), W(\boldsymbol{t})) = \left\{\Gamma(\nu)2^{\nu-1}\right\}^{-1} \left\{\delta(\boldsymbol{s}, \boldsymbol{t})\phi^{-1}\sqrt{2\nu}\right\}^{\nu} K_\nu \left(\delta(\boldsymbol{s}, \boldsymbol{t})\phi^{-1}\sqrt{2\nu}\right) \tag{3}$$

where ν denotes the degree of smoothness, K_ν denotes the modified Bessel equation of the second kind, and as before ϕ denotes a range parameter and $\delta(\boldsymbol{s}, \boldsymbol{t})$ the Euclidean distance between locations $\boldsymbol{s}$ and $\boldsymbol{t}$. The exponential covariance function corresponds to a Matérn class covariance function with $\nu = 1/2$.

A random field is said to be stationary if the joint distribution of a the response at a finite set of locations does not change when the set of locations are all shifted in space by a fixed spatial lag. That is, letting $\{T(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ be a random field on spatial domain $\mathcal{D}$ that takes value $T(\boldsymbol{s}_i)$ at location $\boldsymbol{s}_i \in \mathcal{D}$ for $i = 1, \ldots, n$, the random field $T(\boldsymbol{s})$ is stationary if $F_n\left(T(\boldsymbol{s}_1), \ldots, T(\boldsymbol{s}_n)\right) = F_n\left(T(\boldsymbol{s}_1 + \boldsymbol{h}), \ldots, T(\boldsymbol{s}_n + \boldsymbol{h})\right)$ where $F_n(\cdot)$ is the joint distribution of a length n sample from $T(\boldsymbol{s})$ and $\boldsymbol{h}$ is a fixed spatial lag. The random field $\{T(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ is second-order stationary if the following are satisfied:

$$\begin{aligned} E\left\{T(\boldsymbol{s})\right\} &= \mu \text{ for all } \boldsymbol{s} \in \mathcal{D} \\ \text{var}\left\{T(\boldsymbol{s})\right\} &= \sigma^2 < \infty \text{ for all } \boldsymbol{s} \in \mathcal{D} \\ \text{cov}\left\{T(\boldsymbol{s}), T(\boldsymbol{s} + \boldsymbol{h})\right\} &= C(\boldsymbol{h}) \end{aligned} \tag{4}$$

where the function $C(\cdot)$ depends only on the spatial lag $\boldsymbol{h}$ and not on the location $\boldsymbol{s}$.

The coefficient vector $\boldsymbol{\beta}$ in (1) is a fixed constant. The model can be made more flexible if the

coefficients are described by a stationary random field. Such a model is written

$$Y(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})'\boldsymbol{\beta}(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \tag{5}$$

where $\boldsymbol{\beta}(\boldsymbol{s})$ is a random coefficient field with a Matérn-class covariance function and the spatial random effect $W(\boldsymbol{s})$ included in the intercept $\beta_0(\boldsymbol{s})$. The random coefficient field $\boldsymbol{\beta}(\boldsymbol{s})$ can be estimated by Markov Chain Monte Carlo (MCMC) methods under the assumption that $\boldsymbol{\beta}(\boldsymbol{s})$ is stationary (Gelfand et al., 2003).

Alternatively, kernel-based and spline-based methods can be considered for fitting VCR models without assuming the coefficients are described by a stationary random field.

Coefficients for a spline-based VCR model are estimated by maximizing a penalized global likelihood, with the penalty calculated from the wiggliness of the coefficient surface (Wood, 2006). This contrasts to kernel-based estimates of the coefficients in a VCR model, which maximize a local likelihood to estimate the local coefficients at each sampling location (Loader, 1999). Fan and Zhang (1999) demonstrated that the optimal kernel bandwidth estimate for a VCR model can be found via a two-step technique.

Model selection in VCR models may be local or global. Global selection means including or excluding variables everywhere in the spatial domain, while local selection means including or excluding variables at individual locations within the spatial domain. For global model selection in spline-based VCR models, Wang et al. (2008) proposed a SCAD penalty (Fan and Li, 2001) for variable selection in spline-based VCR models with a univariate effect-modifying variable. Antoniadas et al. (2012) used the nonnegative Garrote penalty (Breiman, 1995) in P-spline-based VCR models having a univariate effect-modifying variable.

Wavelet methods for fitting SVCR models were explored by Shang (2011) and Zhang and Clayton (2011). Sparsity in the wavelet coefficients is achieved either by ℓ_1-penalization (also known as the Lasso (Tibshirani, 1996)) (Shang, 2011) or by Bayesian variable selection (Zhang and Clayton, 2011). Sparsity in the wavelet domain does not imply sparsity in the covariates, though, so neither method is suitable for local variable selection.

Geographically weighted regression (GWR) is a kernel-based method for estimating the coefficients of an SVCR model where the kernel weights are based on the distance between sampling locations (Brundson et al., 1998; Fotheringham et al., 2002). At each sampling location, traditional GWR estimates the local regression coefficients by the local likelihood (Loader, 1999). As a kernel-based smoother for regression coefficients, traditional GWR tends to exhibit bias near the boundary of the region being modeled (Hastie and Loader, 1993). One way to reduce the boundary-effect bias is to model the coefficient surface as locally linear rather than locally constant by including coefficient-by-location interactions (Wang et al., 2008).

Traditional GWR relies on *a priori* global model selection to decide which variables should be included in the model. The idea of using Lasso regularization for local variable selection in a GWR model appears in the literature as the geographically weighted Lasso (GWL) (Wheeler, 2009). The GWL applies the Lasso for local variable selection and uses a jackknife criterion for selection of the Lasso tuning parameters. Because the jackknife criterion can only be computed at sampling locations where the response variable is observed, the GWL cannot be used to impute missing values of the response variable nor to interpolate the coefficient surface and/or the response variable between sampling locations.

Lasso regularization for model selection, while popular, can leave relevant covariates out of the

model when they are correlated with other covariates, and the predictive performance of the Lasso may be dominated in such a case by ridge regression, which does not allow for local model selection (Tibshirani, 1996). The elastic net is a regularization method that combines a ℓ_1 (Lasso) and a ℓ_2 (ridge) penalty on the estimated coefficients, overcoming these drawbacks of the Lasso (Zou and Hastie, 2005).

Additionally, Lasso regularization does not generally produce consistent estimates of the relevant covariates (Leng et al., 2006). The adaptive Lasso (AL) (Zou, 2006) is an improvement to the Lasso that does produce consistent estimates of the coefficients and has been shown to have appealing properties for automating variable selection, which under suitable conditions include the "oracle" property of asymptotically selecting exactly the correct set of covariates for inclusion in a regression model.

The remainder of this document is organized as follows. In Section **??**, a simulation study is conducted to assess the performance of the GWEN in variable selection and coefficient estimation. An application to real data is presented in Section **??**.

2. Spatially varying coefficients regression

2.1. Model

Consider n data observations, taken at sampling locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ in a spatial domain $D \subset \mathbb{R}^2$. For $i = 1, \ldots, n$, let $y(\boldsymbol{s}_i)$ and $\boldsymbol{x}(\boldsymbol{s}_i)$ denote the univariate response variable, and a $(p+1)$-variate vector of covariates measured at location $\boldsymbol{s}_i$, respectively. At each location $\boldsymbol{s}_i$, assume that the outcome is related to the covariates by a linear model where the coefficients $\boldsymbol{\beta}(\boldsymbol{s}_i)$ may be spatially-

varying and $\varepsilon(\boldsymbol{s}_i)$ is random error at location $\boldsymbol{s}_i$. That is,

$$y(\boldsymbol{s}_i) = \boldsymbol{x}(\boldsymbol{s}_i)'\boldsymbol{\beta}(\boldsymbol{s}_i) + \varepsilon(\boldsymbol{s}_i). \tag{6}$$

Further assume that the error term $\varepsilon(\boldsymbol{s}_i)$ is normally distributed with zero mean and variance σ^2, and that $\varepsilon(\boldsymbol{s}_i)$, $i = 1, \ldots, n$ are independent. That is,

$$\varepsilon(\boldsymbol{s}_i) \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right). \tag{7}$$

In order to simplify the notation, let $\boldsymbol{x}(\boldsymbol{s}_i) \equiv \boldsymbol{x}_i \equiv (1, x_{i1}, \ldots, x_{ip})'$, $\boldsymbol{\beta}(\boldsymbol{s}_i) \equiv \boldsymbol{\beta}_i \equiv (\beta_{i0}, \beta_{i1}, \ldots, \beta_{ip})'$, and $y(\boldsymbol{s}_i) \equiv y_i$. Equations (6) and (7) can now be rewritten as

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta}_i + \varepsilon_i \text{ and } \varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right). \tag{8}$$

Further, let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ and $\boldsymbol{y} = (y_1, \ldots, y_n)'$. Thus, conditional on the design matrix $\boldsymbol{X}$, observations of the response variable at different locations are independent of each other.

An SVCR model that estimates the regression coefficients as locally constant, as in the class of Nadaraya-Watson kernel smoothers (Härdle, 1990), suffers the problem of biased estimation that is common to that class of models - particularly where there is a gradient to the coefficient surface at the boundary of the domain (Hastie and Loader, 1993).

In the context of nonparametric regression, the boundary-effect bias can be reduced by local polynomial modeling, usually in the form of a locally linear model (Fan and Gijbels, 1996). Here, locally linear coefficients are estimated by augmenting the local design matrix with covariate-by-location interactions in two dimensions as proposed by Wang et al. (2008). The augmented local design

matrix at location $\boldsymbol{s}_i$ is

$$\boldsymbol{Z}_i = (\boldsymbol{X} \;\; L_i\boldsymbol{X} \;\; M_i\boldsymbol{X}) \tag{9}$$

where $\boldsymbol{X}$ is the unaugmented matrix of covariates, $L_i = \text{diag}\{s_{i',x} - s_{i,x}\}$ and $M_i = \text{diag}\{s_{i',y} - s_{i,y}\}$ for $i' = 1, \ldots, n$.

2.2. Estimation

The total log-likelihood of the observed data is the sum of the log-likelihood of each individual observation:

$$\ell(\boldsymbol{\beta}_i) = -(1/2) \sum_{i'=1}^{n} \left\{ \log \sigma_i^2 + \sigma_i^{-2} \left(y_{i'} - \boldsymbol{z}'_{i'} \boldsymbol{\beta}_i \right)^2 \right\}. \tag{10}$$

Since there are a total of $n \times 3(p+1)$ free parameters for n observations, the model is not identifiable and it is not possible to directly maximize the total likelihood.

The values of the local coefficients $\boldsymbol{\beta}_i$ are estimated at $\boldsymbol{s}_i$ by the weighted likelihood

$$\mathcal{L}_i \left(\boldsymbol{\beta}_i \right) = \prod_{i'=1}^{n} \left[\left(2\pi\sigma_i^2 \right)^{-1/2} \exp \left\{ -1/2\sigma_i^{-2} \left(y_{i'} - \boldsymbol{z}'_{i'} \boldsymbol{\beta}_i \right)^2 \right\} \right]^{w_{ii'}}, \tag{11}$$

where the weights are calculated by a kernel function $K_h(\cdot)$ such as the Epanechnikov kernel:

$$\begin{aligned} w_{ii'} &= K_h(\delta_{ii'}) = h^{-2} K \left(h^{-1} \delta_{ii'} \right) \\ K(x) &= \begin{cases} (3/4)(1 - x^2) & \text{if } \delta_{ii'} < h, \\ 0 & \text{if } \delta_{ii'} \geqslant h. \end{cases} \end{aligned} \tag{12}$$

Thus, the local log-likelihood function is, up to an additive constant:

$$\ell_i\left(\boldsymbol{\beta}_i\right) = -(1/2)\sum_{i'=1}^{n} w_{ii'}\left\{\log\sigma_i^2 + \sigma_i^{-2}\left(y_{i'} - \boldsymbol{z}_{i'}'\boldsymbol{\beta}_i\right)^2\right\}. \tag{13}$$

From (13), the maximum local likelihood estimate $\hat{\sigma}_i^2$ is:

$$\hat{\sigma}_i^2 = \left(\sum_{i'=1}^{n} w_{ii'}\right)^{-1}\sum_{i'=1}^{n} w_{ii'}\left(y_{i'} - \boldsymbol{z}_{i'}'\hat{\boldsymbol{\beta}}_i\right)^2 \tag{14}$$

3. Local variable selection and parameter estimation

3.1. Local variable selection

The adaptive group lasso (AGL) is explored as a penalty function for local variable selection in SVCR models. The proposed local variable selection with AGL penalty is an ℓ_1 regularization method for variable selection in regression models (Wang and Leng, 2008; Zou, 2006). The adaptive group lasso selects groups of covariates for inclusion or exclusion in the model. For an SVCR model, each variable group is a covariate and its interactions on location.

3.1.1. Local variable selection and coefficient estimation with the adaptive group lasso

The objective function for the AGL at $\boldsymbol{s}_i$ consists of the local log-likelihood and an additive penalty:

$$\begin{aligned}\mathcal{S}(\boldsymbol{\beta}_i) &= -2\ell_i\left(\boldsymbol{\beta}_i\right) + \mathcal{J}_1(\boldsymbol{\beta}_i)\\ &= \sum_{i'=1}^{n} w_{ii'}\left\{\log\sigma_i^2 + \sigma_i^{-2}\left(y_{i'} - \boldsymbol{z}_{i'}'\boldsymbol{\beta}_i\right)^2\right\} + \lambda_i\sum_{j=1}^{p}\|\beta_{ij}\|/\gamma_{ij}\end{aligned} \tag{15}$$

where $\sum_{i'=1}^{n} w_{ii'}\left(y_{i'} - \boldsymbol{z}_{i'}'\boldsymbol{\beta}_i\right)^2$ is the weighted sum of squares minimized by traditional GWR, and $\mathcal{J}_1(\boldsymbol{\beta}_i) = \lambda_i\sum_{j=1}^{p}\|\beta_{ij}\|/\gamma_{ij}$ is the AGL penalty. With the vector of unpenalized local coefficients $\boldsymbol{\gamma}_i$, the AL penalty for the jth group of coefficients β_{ij} at location $\boldsymbol{s}_i$ is λ_i/γ_{ij}, where $\lambda_i > 0$ is a the

local tuning parameter applied to all coefficients at location $\boldsymbol{s}_i$ and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{ip})'$ is the vector of adaptive weights at location $\boldsymbol{s}_i$.

3.2. Tuning parameter selection

A local tuning parameter λ_i is required for the variable selection step of fitting each local model by the AGL. The corrected AIC is used to select λ_i (Hurvich et al., 1998):

$$\begin{aligned} \mathrm{AIC}_{\mathrm{c},i} &= -2 \sum_{i'=1}^{n} \ell_{ii'} + 2\nu_i + \frac{2\nu_i(\nu_i+1)}{\sum_{i'=1}^{n} w_{ii'} - \nu_i - 1} \\ &= \sum_{i'=1}^{n} w_{ii'} \left\{ \log(2\pi) + \log \hat{\sigma}_i^2 + \hat{\sigma}_i^{-2} \left(y_{i'} - \boldsymbol{x}_{i'}' \hat{\boldsymbol{\beta}}_{i'} \right)^2 \right\} + 2\nu_i + \frac{2\nu_i(\nu_i+1)}{\sum_{i'=1}^{n} w_{ii'} - \nu_i - 1} \end{aligned} \tag{16}$$

The local AIC_c is calculated by adding a penalty to the local likelihood, with the sum of the weights around $\boldsymbol{s}_i$, $\sum_{i'=1}^{n} w_{ii'}$, playing the role of the sample size and the "degrees of freedom" (ν_i) at $\boldsymbol{s}_i$ given by

$$\nu_i = \sum_{j=1}^{p} I\left\{ \|\hat{\boldsymbol{\beta}}_j(\boldsymbol{s}_i)\| > 0 \right\} + \sum_{j=1}^{p} \frac{\|\hat{\boldsymbol{\beta}}_j(\boldsymbol{s}_i)\|}{\|\tilde{\boldsymbol{\beta}}_j(\boldsymbol{s}_i)\|}(d_j - 1) \tag{17}$$

where d_j is the number of variables in group j and $\tilde{\boldsymbol{\beta}}_j(\boldsymbol{s}_i)$ is the ordinary SVCR coefficient estimate for $\boldsymbol{\beta}_j(\boldsymbol{s}_i)$. The equation for ν_i is due to (Wang and Leng, 2008).

3.3. Bandwidth parameter estimation

The bandwidth parameter is estimated to minimize an information criterion. It is common in nonparametric regression to select the bandwdth to minimize a corrected AIC where the degrees of freedom are given by the trace of a smoothing matrix (Hurvich et al., 1998). Because ℓ_1 penalization procedures like the AGL are not linear smoothers, there is no smoothing matrix for the AGL. There is need for a procedure to estimate the degrees of freedom of a nonparametric AGL model.

The degrees of freedom used to estimate the $\hat{y}_i$ are estimated by $\nu_i w_{ii} / \sum_{j=1}^{n} w_{ij}$.

4. References

References

Antoniadas, A., I. Gijbels, and A. Verhasselt (2012). Variable selection in varying-coefficient models using p-splines. *Journal of Computational and Graphical Statistics 21*, 638–661.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics 51*, 373–384.

Brundson, C., S. Fotheringham, and M. Charlton (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A 30*, 1905–1927.

Cressie, N. (1993). *Statistics for Spatial Data (Revised Edition)*. Wiley, New York.

Diggle, P. and P. Ribeiro (2007). *Model-Based Geostatistics*. Springer New York.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modeling and its Applications*. Chapman and Hall, London.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*, 1348–1360.

Fan, J. and W. Zhang (1999). Statistical estimation in varying coefficient models. *Annals of Statistics 27*, 1491–1518.

Fotheringham, A., C. Brunsdon, and M. Charlton (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, West Sussex, England.

Gelfand, A. E., H.-J. Kim, C. F. Sirmans, and S. Banerjee (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association 98*, 387–396.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Boston MA.

Hastie, T. and C. Loader (1993). Local regression: automatic kernel carpentry. *Statistical Science 8*, 120–143.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society Series B 55*, 757–796.

Hurvich, C. M., J. S. Simonoff, and C.-L. Tsai (1998). Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society Series B 60*, 271–293.

Leng, C., Y. Lin, and G. Wahba (2006). A note on the lasso and related procedures in model selection. *Statistica Sinica 16*, 1273–1284.

Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York.

Shang, Z. (2011). *Bayesian Variable Selection*. Ph. D. thesis, University of Wisconsin-Madison.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B 58*, 267–288.

Wang, H. and C. Leng (2008). A note on adaptive group lasso. *Computational Statistics and Data Analysis 52*, 5277–5286.

Wang, L., H. Li, and J. Z. Huang (2008). Variable selection in nonparametric varying-coefficient

models for analysis of repeated measurements. *Journal of the American Statistical Association 103*, 1556–1569.

Wang, N., C.-L. Mei, and X.-D. Yan (2008). Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique. *Environment and Planning A 40*, 986–1005.

Wheeler, D. C. (2009). Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and Planning A 41*, 722–742.

Wood, S. (2006). *Generalized Additive Models: An Introduction With R*. Chapman and Hall, Boca Raton, FL.

Zhang, J. and M. K. Clayton (2011). Functional concurrent linear regression model for images. *Journal of Agricultural, Biological, and Environmental Statistics 16*, 105–130.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*, 1418–1429.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B 67*, 301–320.