A spline-based approach to spatial confounding in spatial linear regression methods for geostatistical data

Guilherme Ludwig
gvludwig@ime.unicamp.br

University of Campinas, Brazil
Joint work with Jun Zhu, Chun-Shu Chen, Perla Reyes and Shawn Conley

07-25-2017

Outline

Splines and properties of semiparametric estimators

Simulation study of penalized least squares estimates

Simulation study of a spatially confounded scenario

A case study in precision agriculture

Context and motivation

A data model for spatial linear regression is a spatial process with covariates and measurement error,

$$Y(\mathbf{s}) = \beta_0 + \beta_1 x_1(\mathbf{s}) + \ldots + \beta_p x_p(\mathbf{s}) + \eta(\mathbf{s}) + \varepsilon(\mathbf{s}).$$

Splines are included in spatial regression models when the investigator does not want to impose a structure to the spatial covariance function (Stroup, 2012).

Recent findings by Hodges and Reich, (2010) and Paciorek, (2010), among others, outlined the problem of spatial confounding: including a spatial random effect in a regression model often changes the estimates of $\boldsymbol{\beta}$ in ways that are difficult to anticipate (more on this later).

We describe how spatial confounding is present in spline-based regression, how the spline is connected with the spatial effect and how the spline tuning parameter can be used examine changes in coefficients.

Spatial data model

The data model is a spatial process with covariates and measurement error,

$$Y(\mathbf{s}) = \beta_0 + \beta_1 x_1(\mathbf{s}) + \ldots + \beta_p x_p(\mathbf{s}) + \eta(\mathbf{s}) + \varepsilon(\mathbf{s}).$$

If we assume $\eta(\mathbf{s})$ is L_2 continuous, then it has a *Karhunen-Loève expansion*,

$$= \beta_0 + \beta_1 x_1(\mathbf{s}) + \ldots + \beta_p x_p(\mathbf{s}) + \sum_{k=1}^{\infty} \xi_k^{1/2} Z_k \phi_k(\mathbf{s}) + \varepsilon(\mathbf{s}).$$

such that

- ξ_k and ϕ_k are eigenvalues and eigenfunctions, respectively, of the covariance function of η.
- $\xi_k^{1/2} \geq 0, k = 1, \ldots$ and $\sum_{k=1}^{\infty} \xi_k < \infty$,
- $\|\phi_k\| = 1$ and $\langle \phi_k, \phi_{k'} \rangle = 0$ whenever $k \neq k'$,
- if $\eta(\mathbf{s})$ is a Gaussian process, $\{Z_k\}_{k=1}^{\infty}$ are pairwise uncorrelated $N(0, 1)$.

Nature of $\eta(\mathbf{s})$

(A) We focus on the case where $\eta(\mathbf{s})$ represents spatial correlation, or a structured noise, among the observations. If replicates were available, new realizations of $\eta(\mathbf{s})$ would be different. This is the interpretation of Paciorek, (2010) and Hanks et al., (2015).

(B) It is also important to know that $\eta(\mathbf{s})$ can be used to account for missing covariates (or important but not measured covariates) and thus, it is more suitable to treat $\eta(\mathbf{s})$ as a fixed realization of some spatial process. As such the spline role is akin to the “random effects as a device to define a smoother" in Hodges and Reich, (2010). We do not address this scenario directly.

Spatial confounding is due, roughly speaking, multicollinearity between $x(\mathbf{s})$ and either of (A) the variance components of $\eta(\mathbf{s})$ or (B) the missing/unavailable covariates.

Thin-plate spline

Thin-plate splines (TPS) arise as a solution to the variational problem (Wahba, 1990)

$$\min_{f \in \mathcal{W}_2^2} \sum_{i=1}^{n} \{y_i - f(\mathbf{s}_i)\}^2 + \lambda J[f]$$

where $J[f]$ is the penalty

$$J[f] = \int_{\mathbb{R}^2} \left(\frac{\partial^2}{\partial s_x^2} f(\mathbf{s})\right)^2 + 2\left(\frac{\partial^2}{\partial s_x s_y} f(\mathbf{s})\right)^2 + \left(\frac{\partial^2}{\partial s_y^2} f(\mathbf{s})\right)^2 \mathrm{d}\mathbf{s}$$

TPS can be explicitly written as

$$f(\mathbf{s}) = \alpha_0 + \alpha_1 s_x + \alpha_2 s_y + \sum_{i=1}^{n} \theta_i \varphi_i(\mathbf{s})$$

where $\mathbf{s} = (s_x, s_y)$ are spatial coordinates, and

$$\varphi_i(\mathbf{s}) = \|\mathbf{s} - \mathbf{s}_i\|^2 \log \|\mathbf{s} - \mathbf{s}_i\|.$$

Fitting the spatial data with penalized least squares

Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$ be covariates, $\mathbf{f} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\Phi}\boldsymbol{\theta} = (f(\mathbf{s}_1), \ldots, f(\mathbf{s}_n))'$ be a thin-plate spline, then

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{f}} \end{pmatrix} = \arg\min_{\boldsymbol{\beta}, \mathbf{f}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{f}\|^2 + \lambda J[f]$$

The estimators $\hat{\mathbf{f}}$ and $\hat{\boldsymbol{\beta}}$ are

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{y} \\ \hat{\mathbf{f}} &= \mathcal{S}_\lambda(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \end{aligned}$$

where $\mathcal{S}_\lambda$ satisfies $\mathcal{S}_\lambda(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{T}\hat{\boldsymbol{\alpha}} + \boldsymbol{\Phi}\hat{\boldsymbol{\theta}}$ and is called the *smoother matrix*, $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\theta}}$ are the PLS estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$.

Mean and variance of parameter estimates

Unconditional expectation

$$\mathbb{E}(\hat{\boldsymbol\beta}) = \boldsymbol\beta$$

$$\mathbb{E}(\hat{\mathbf{f}}) = \mathbf{0}$$

Conditional expectation

$$\mathbb{E}(\hat{\boldsymbol\beta}|\boldsymbol\eta) = \boldsymbol\beta + \sum_{k=1}^{\infty} \xi_k^{1/2} Z_k \boldsymbol\psi_{\lambda,k}$$

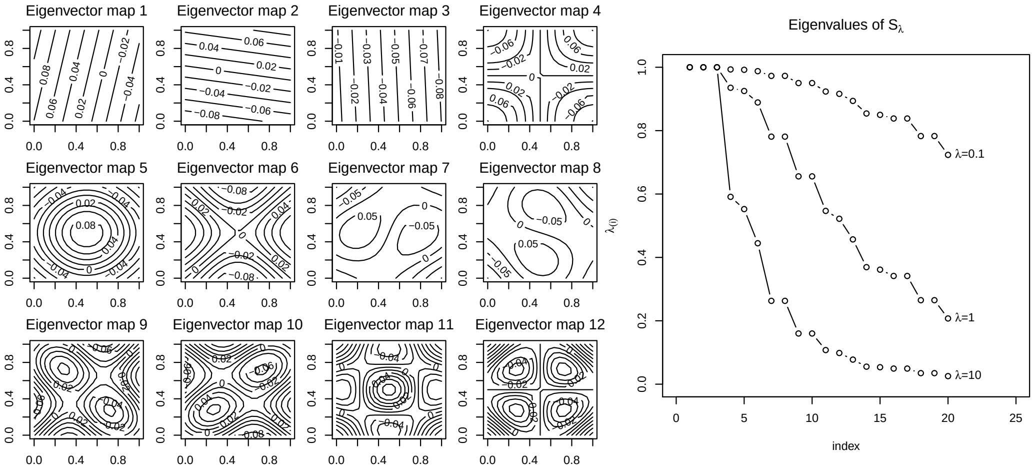
$$\mathbb{E}(\hat{\mathbf{f}}|\boldsymbol\eta) \neq \mathbf{0}$$

and

$$\begin{aligned}
\mathrm{Var}\,(\hat{\boldsymbol\beta}) &= \sigma^2\{\mathbf{X}'(\mathbf{I}-\mathcal{S}_\lambda)\mathbf{X}\}^{-1}\mathbf{X}'(\mathbf{I}-\mathcal{S}_\lambda)^2\mathbf{X}\{\mathbf{X}'(\mathbf{I}-\mathcal{S}_\lambda)\mathbf{X}\}^{-1} + \sum_{k=1}^{\infty} \xi_k \boldsymbol\psi_{\lambda,k}\boldsymbol\psi'_{\lambda,k} \\
&= \mathbb{E}\left\{\mathrm{Var}\,(\hat{\boldsymbol\beta}|\boldsymbol\eta)\right\} + \mathrm{Var}\,\left\{\mathbb{E}(\hat{\boldsymbol\beta}|\boldsymbol\eta)\right\} \\
&= \mathrm{Var}_{\mathbf{X}}(\hat{\boldsymbol\beta}) + \mathrm{Var}_{\eta}(\hat{\boldsymbol\beta})
\end{aligned}$$

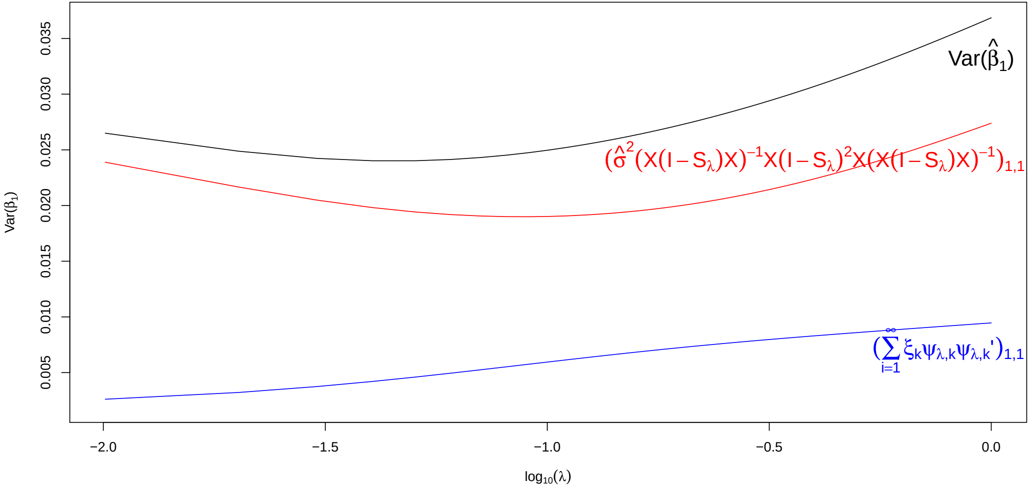
where $\boldsymbol\psi_{\lambda,k} = (\mathbf{X}'(\mathbf{I}-\mathcal{S}_\lambda)\mathbf{X})^{-1}\,\mathbf{X}'(\mathbf{I}-\mathcal{S}_\lambda)\boldsymbol\phi_k$ and $\boldsymbol\eta = \sum_{k=1}^{\infty} \xi_k^{1/2} Z_k \boldsymbol\phi_k$.

Eigendecomposition of $\mathcal{S}_\lambda = \sum_{j=1}^{n} \ell_j \mathbf{q}_j \mathbf{q}_j'$



Mean and variance of parameter estimates

$x_j(\mathbf{s})$ realization of independent Gaussian random field



Tuning parameter selection

Automatic selection of λ is usually based on prediction, such as when minimizing

$$\mathrm{GCV}(\lambda) = \frac{n\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\{n - p - \mathrm{tr}(\mathcal{S}_\lambda)\}^2}$$

We can minimize the mean squared error (MSE) of $\hat{\beta}$. Since

$$\mathrm{MSE}(\hat{\beta}) = \sigma^2 \mathrm{tr}\left(\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)^2\mathbf{X}\{\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{X}\}^{-2}\right) + \sum_{k=1}^{\infty}\xi_k\|\psi_{\lambda,k}\|_2^2,$$

we can approximate $\mathrm{MSE}(\hat{\beta})$ with

$$\mathrm{eMSE}(\hat{\beta}) = \hat{\sigma}^2 \mathrm{tr}\left(\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)^2\mathbf{X}\{\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{X}\}^{-2}\right) + \|\mathcal{P}_{\mathbf{X}}\hat{\mathbf{f}}\|^2$$

where $\mathcal{P}_{\mathbf{X}} = \{\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{X}\}^{-1}\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)$. Select λ that satisfies

$$\lambda_{\mathrm{opt}} = \arg\min_{\lambda \geq 0} \mathrm{eMSE}(\hat{\beta}).$$

Simulation scenario

- $n = 50$ locations sampled randomly over $[0,1] \times [0,1]$
- $x_1(\mathbf{s}), x_2(\mathbf{s})$ sampled from $N(0,1)$, *independently* of $\eta(\mathbf{s})$, and treated as fixed across simulations.
- $\beta_0 = \beta_1 = \beta_2 = 1$, and $\sigma^2 = 1$.
- The spatial process $\eta(\mathbf{s})$ has Matérn covariance function with parameters $\sigma_\eta^2 = 1$, $\kappa = 0.5, 1.5$ or 2.5 and $\rho = 0.02, 0.05, 0.2$ or 0.5.
- For comparison, we fitted the regression coefficients with ordinary least squares (OLS), restricted maximum likelihood using the Matérn covariance function (REML), penalized least squares with λ selected via GCV (PLS-GCV) or the proposed selection method (PLS-MSE).
- The simulations were repeated $S = 500$ times.

Regression coefficient β_2 estimate

κ	ρ	OLS	REML	PLS-GCV	PLS-MSE
0.5	0.02	1.005 (0.21)	1.004 (0.21)	1.004 (0.21)	1.004 (0.22)
0.5	0.05	1.012 (0.21)	1.013 (0.21)	1.010 (0.23)	1.010 (0.23)
0.5	0.20	1.010 (0.19)	1.011 (0.19)	1.005 (0.19)	1.004 (0.19)
0.5	0.50	0.992 (0.18)	0.987 (0.16)	0.989 (0.16)	0.989 (0.17)
1.5	0.02	1.008 (0.21)	1.008 (0.21)	1.004 (0.22)	1.005 (0.22)
1.5	0.05	0.999 (0.21)	1.001 (0.20)	0.999 (0.21)	0.997 (0.21)
1.5	0.20	0.998 (0.19)	1.003 (0.17)	1.005 (0.17)	1.006 (0.17)
1.5	0.50	1.010 (0.17)	1.008 (0.16)	1.010 (0.16)	1.011 (0.16)
2.5	0.02	0.996 (0.21)	0.998 (0.21)	0.995 (0.22)	0.995 (0.22)
2.5	0.05	1.001 (0.20)	1.002 (0.19)	1.003 (0.19)	1.002 (0.20)
2.5	0.20	0.992 (0.19)	0.996 (0.17)	0.997 (0.17)	0.996 (0.17)
2.5	0.50	1.003 (0.15)	0.999 (0.15)	1.002 (0.15)	1.003 (0.15)

Figure version

Simulation scenarios for confounding

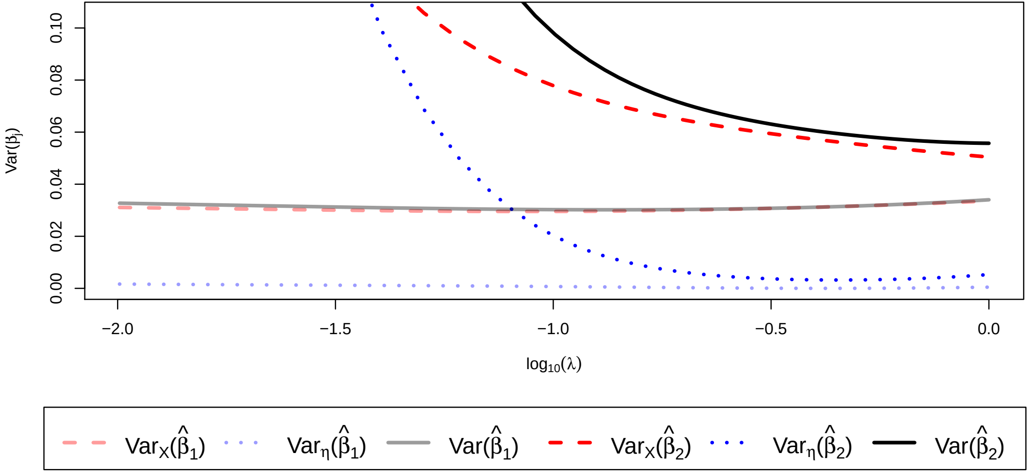
Goals:

- To study spatial confounding effect on regression coefficient estimates in semiparametric thin-plate spline regression.
- To evaluate the effect of tuning λ via eMSE and its impact on spatial confounding effects.

We impose a scenarios of spatial confounding where both the covariate and the error have spatial structures. In this case, $x_2(\mathbf{s})$ is a Gaussian random field with Matérn covariance function and parameters $\kappa = 2.5$, $\rho = 0.2$

Mean and variance of parameter estimates

$x_2(\mathbf{s})$ and $\eta(\mathbf{s})$ GRF with Matérn covariance ($\kappa = 2.5, \rho = 0.2$)



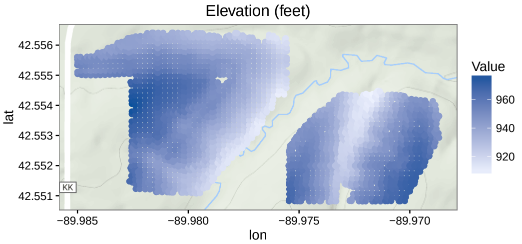
$x_2(\mathbf{s})$ Matérn covariance ($\kappa = 2.5, \rho = 0.2$): β_2 estimate

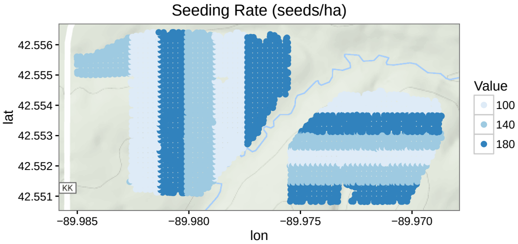
κ	ρ	OLS	REML	PLS-GCV	PLS-MSE
0.5	0.02	1.023 (0.60)	1.024 (0.61)	1.075 (1.55)	1.082 (1.46)
0.5	0.05	0.994 (0.64)	0.988 (0.64)	0.938 (1.64)	0.944 (1.51)
0.5	0.20	1.007 (0.87)	1.005 (0.83)	1.066 (1.61)	1.061 (1.52)
0.5	0.50	1.064 (0.94)	1.070 (0.90)	1.043 (1.71)	1.041 (1.60)
1.5	0.02	0.985 (0.57)	0.984 (0.58)	0.977 (1.44)	0.976 (1.36)
1.5	0.05	0.998 (0.89)	1.000 (0.87)	1.031 (2.15)	1.030 (1.99)
1.5	0.20	1.003 (1.09)	1.003 (1.02)	1.037 (1.85)	1.039 (1.83)
1.5	0.50	0.990 (0.79)	0.985 (0.74)	0.953 (1.05)	0.955 (1.01)
2.5	0.02	1.014 (0.62)	1.016 (0.61)	1.027 (1.91)	1.014 (1.73)
2.5	0.05	0.989 (0.87)	0.997 (0.84)	0.829 (3.87)	0.856 (3.62)
2.5	0.20	1.031 (0.96)	1.013 (0.92)	0.959 (1.49)	0.951 (1.47)
2.5	0.50	0.991 (0.79)	1.000 (0.77)	1.048 (1.08)	1.034 (1.04)

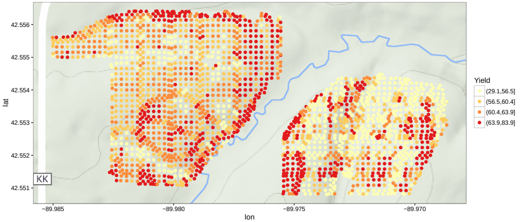
Figure version

Precision AG example - Covariates and Yield

Soybean growth in SW Wisconsin

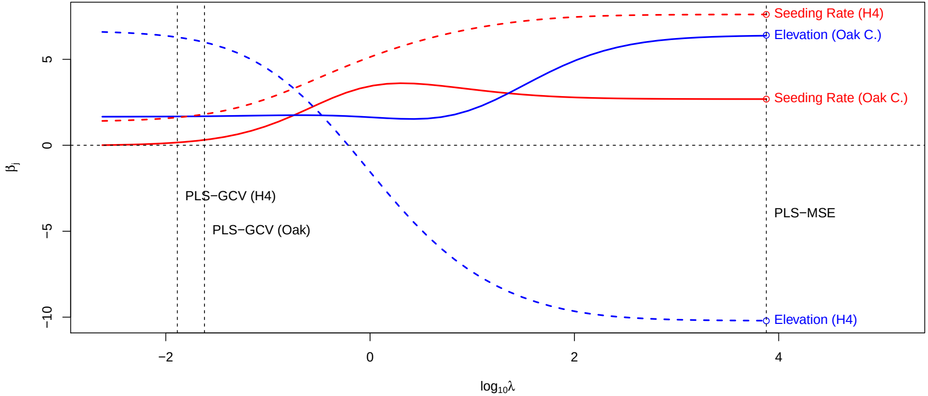






Precision AG example - $\hat{\beta}_j / s.e.(\hat{\beta}_j)$ as a function of λ

Soybean growth in SW Wisconsin



Precision AG example - Regression coefficients' estimates

Soybean growth in SW Wisconsin

H4

	OLS		PLS-GCV		PLS-MSE		REML	
Seeding Rate	0.028	*	0.009		0.008		0.013	*
Elevation	-0.109	*	0.288	*	0.309	*	0.097	*

Oak Creek

	OLS		PLS-GCV		PLS-MSE		REML	
Seeding Rate	0.022	*	0.003		0.000		0.028	*
Elevation	0.125	*	0.179		0.187		0.182	*

Summary of the results

- Including a thin-plate spline is equivalent to regressing the data on additional covariates and shrinking their coefficients.
- The Karhunen-Loève representation of the spatial data model is helpful for understanding the mechanisms of spatial confounding.
- Spatial confounding affects the estimation of regression coefficients in semiparametric methods. Estimation of $\hat{\beta}$ is better when $\eta(\mathbf{s})$ shows stronger dependence than the spatial covariates (i.e. spline is "greedy").
- It is a good strategy to look at changes in estimates as a function of λ, and think about the scale of variation of the covariates.

Future work

- For future work we will explore the spatial confounding and spline usage in generalized linear models.
- We are currently investigating semiparametric spline-based methods in spatial point processes' regression (joint work with Nancy Garcia, University of Campinas).

This work was supported by FAPESP grant 2016/09390-4, Brazil.

Thank you!

References I

Hanks, E. M., Schliep, E. M., Hooten, M. B., and Hoeting, J. A. (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics* Vol. 26., pp. 243–254.

Hodges, J. S. and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician* Vol. 64., pp. 325–334.

Paciorek, C. J. (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science* Vol. 25., pp. 107–125.

Smidt, E. R., Conley, S. P., Zhu, J., and Arriaga, F. J. (2016). Identifying Field Attributes that Predict Soybean Yield Using Random Forest Analysis. *Agronomy Journal* Vol. 108., pp. 637–646.

Stroup, W. W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Chapman & Hall. Baton Rouge.

Wahba, G. (1990). *Spline Models for Observational Data*. Vol. 59. SIAM. Philadelphia.

Regression coefficients β estimates

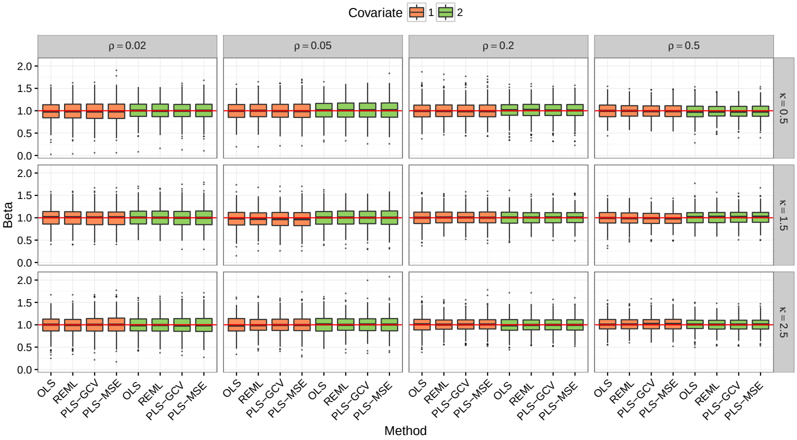


Table version

$x_2(\mathbf{s})$ Matérn covariance $(\kappa = 2.5, \rho = 0.2)$: β estimates

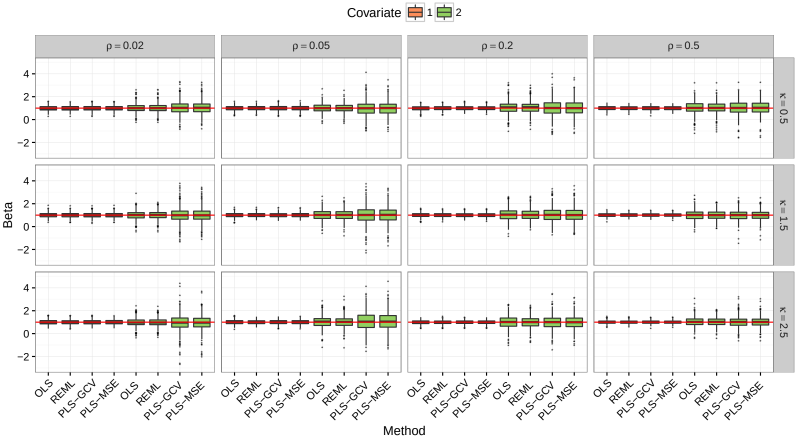
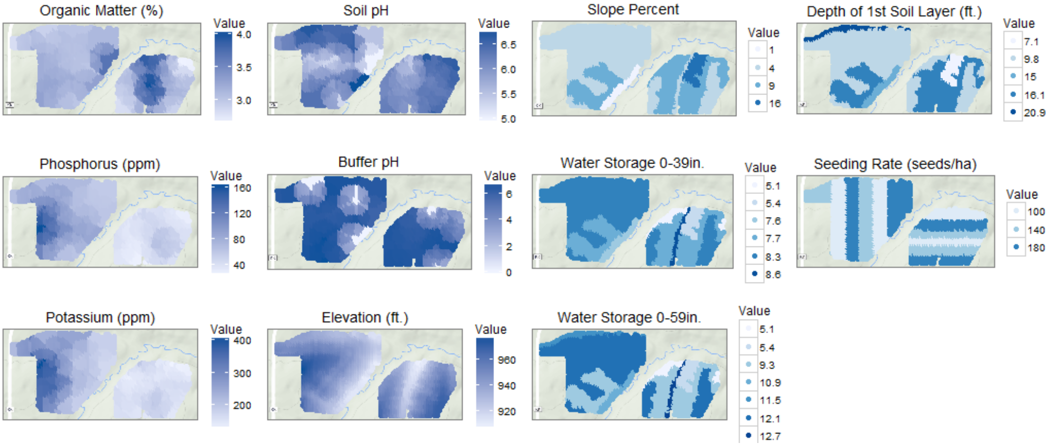


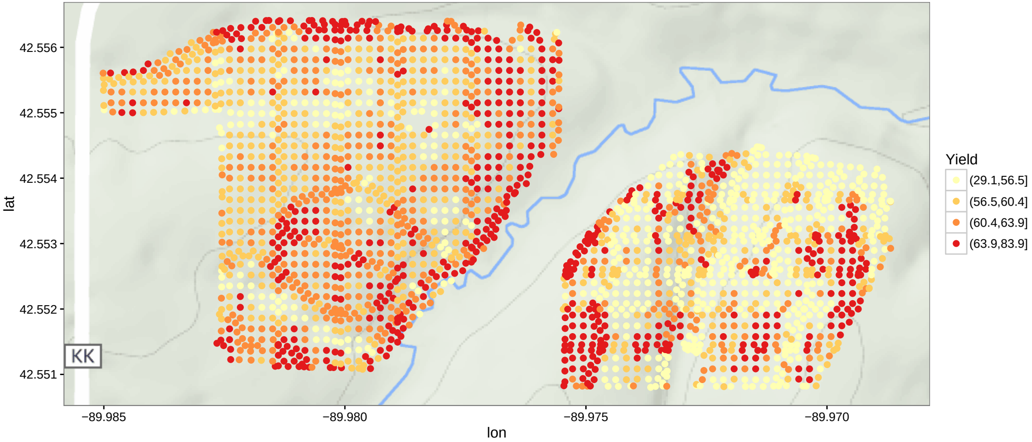
Table version

Precision AG example - covariates

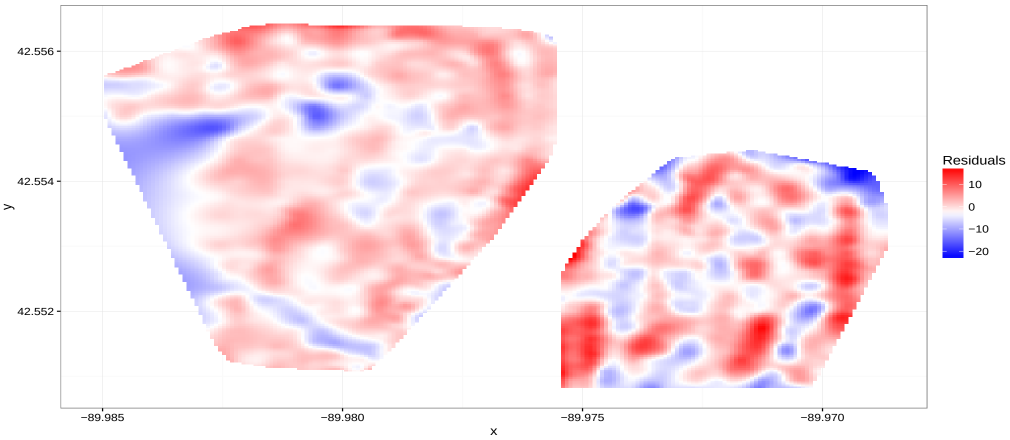
Consider now an example in precision agriculture, in the study by Smidt et al., (2016):



Precision AG example - yield



Precision AG example - residuals



Precision AG example - Regression coefficients' estimates

	OLS		PLS-GCV		PLS-MSE		REML	
Intercept for H4	-12650.74		-64.52		32.44		-61.74	
Oak Creek Increment	-2.91	*	-0.96		-2.82	*	-8.82	
Organic Matter	1.84	*	0.05		1.84	*	0.24	
Phosphorus	0.03		-0.05		0.03		-0.01	*
Potassium	-0.04	*	0.05	*	-0.03	*	0.01	*
Soil pH	1.72	*	-4.12	*	1.86	*	-1.13	
Buffer pH	0.73	*	-0.37	*	0.72	*	0.05	
Elevation	0.01		0.15	*	0.02		0.14	*
Slope Percent	-0.93	*	-0.77	*	-0.88	*	-0.44	*
Water Supply 0-39in.	0.95		3.16	*	0.56		0.72	
Water Supply 0-59in.	-1.06		-2.08	*	-0.86		-0.80	
Depth of 1st Soil Layer	0.03		-0.08		0.06		-0.01	
Seeding Rate	0.03	*	0.03	*	0.03	*	0.02	*

Spatial Point processes

Let X be a locally finite, pairwise interaction point process on $\mathcal{S} \subset \mathbb{R}^d$, with density

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \alpha(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \prod_{\xi \in \mathbf{x}} \phi_{\boldsymbol{\theta}_1}(\xi) \prod_{\{\xi,\eta\} \subset \mathbf{x}} h_{\boldsymbol{\theta}_2}(\xi, \eta), \tag{1}$$

with respect to a standard Poisson process, and indexed by a vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t)^t, \boldsymbol{\theta}_1 \in \mathbb{R}^p$, $\boldsymbol{\theta}_2 \in \mathbb{R}^q$. The process has Papangelou conditional intensity

$$\lambda_{\boldsymbol{\theta}}(u, \mathbf{x}) = \begin{cases} f_{\boldsymbol{\theta}}(\mathbf{x} \cup u)/f_{\boldsymbol{\theta}}(\mathbf{x}), & \text{if } u \notin \mathbf{x}, \text{ and} \\ f_{\boldsymbol{\theta}}(\mathbf{x})/f_{\boldsymbol{\theta}}(\mathbf{x} \backslash x_i), & \text{for } x_i \in \mathbf{x}. \end{cases}$$

Consistency of the parametric regression coefficient estimates has been established (Baddeley, 2000; Kurtz and Li, 2003), it remains an open research question how to obtain optimal coefficients' estimates, in terms of efficiency (see Guan, Jalilian, and R. Waagepetersen, 2015; Coeurjolly et al., 2016, and references therein). We expect that a semiparametric approach can be helpful in improving efficiency of the regression coefficients' estimates.

Estimating equations

Our objective is to consider a set of equations that include a parametric component to model $\phi_{\boldsymbol{\theta}} = \phi_{\boldsymbol{\theta}_1}$, and a non-parametric component to model $h_{\boldsymbol{\theta}_2}$. The parametric part of the method fits estimating equations such as

$$\mathcal{A}_{\boldsymbol{\theta}} e_k = \int_W z_k(u) \exp\{\boldsymbol{\beta}' \mathbf{z}(u)\} \mathrm{d}u + \sum_{i=1}^{n(\mathbf{x})} z_k(x_i), \quad k = 1, \ldots, p.$$

We propose to include a spline term in the fitting process, and tune its smoothness in a way that improves estimation for $\boldsymbol{\theta}$, the regression coefficients. We do so by augmenting the set of estimating equations with

$$\mathcal{A}_{\boldsymbol{\theta}} e_{\ell+p} = \int_W S_\ell(u) \exp\{\boldsymbol{\theta}' \mathbf{z}(u)\} \mathrm{d}u - \sum_{i=1}^{n(\mathbf{x})} (1 - c\gamma_i) S_\ell(x_i),$$

for $\ell = 1, \ldots, K$, with c a tuning parameter and γ_i being a regularization penalty.

Estimating equations

Following Heyde, (2008), the sensitivity matrix of an estimating equation $\mathcal{A}_{\boldsymbol{\theta}} e = 0$ is

$$A_n = \mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{A}_{\boldsymbol{\theta}} e(\boldsymbol{\theta})\right),$$

and the covariance of the estimating equation

$$B_n = \mathrm{Var}\left(\mathcal{A}_{\boldsymbol{\theta}} e(\boldsymbol{\theta})\right).$$

Then the Godambe information matrix is

$$\mathbf{G}_n = A_n^t B_n^{-1} A_n.$$

The Giorgii-Nguyen-Zessin formula states that, for $h \ : \ \mathbb{R}^d \times \mathrm{N}_{\mathrm{lf}} \to [0, \infty)$,

$$\mathbb{E}\left(\sum_{x_i \in X \cap W} h(x_i, X \backslash x_i)\right) = \int_W \mathbb{E}(\lambda_{\boldsymbol{\theta}}(u, X) h(u, X)) \mathrm{d}u,$$

see Møller and R. P. Waagepetersen, (2003, p. 95).

Estimating equations

The sensitivity matrix of an estimating equation e for a finite Gibbs process is

$$A_n = \mathbb{E}\left(\int_W e(u, X)\frac{\partial}{\partial \boldsymbol{\theta}}\lambda(u, X; \boldsymbol{\theta})\right)$$

and can be rewritten as

$$A_n = \begin{pmatrix} \int_W \mathbf{z}(u)\mathbf{z}^t(u)\rho(u)\mathrm{d}u & \mathbf{0}_{p\times K} \\ \int_W \mathbf{S}(u)\mathbf{z}^t(u)\rho(u)\mathrm{d}u & c\boldsymbol{\Gamma}_{K\times K} \end{pmatrix}$$

where ρ is the first order intensity function for the process, such that $\rho(u) = \mathbb{E}(\lambda_{\boldsymbol{\theta}}(u, X))$. Optimal estimation of $\boldsymbol{\theta}$ depends on the tuning parameter c and the choice of penalty $\boldsymbol{\Gamma}_{K\times K}$, and is a work in progress.

References I

Baddeley, A. (2000). Time-invariance estimating equations. *Bernoulli* Vol. 6., pp. 783–808.

Coeurjolly, J.-F., Guan, Y., Khanmohammadi, M., and Waagepetersen, R. (2016). Towards optimal Takacs–Fiksel estimation. *Spatial Statistics* Vol. 18, pp. 396–411.

Guan, Y., Jalilian, A., and Waagepetersen, R. (2015). Quasi-likelihood for spatial point processes. *Journal of the Royal Statistical Society: Series B* Vol. 77., pp. 677–697.

Heyde, C. C. (2008). *Quasi-likelihood and its Application: A General Approach to Optimal Parameter Estimation*. Springer. New York.

Kurtz, T. and Li, S. (2003). Time-invariance modeling and estimation for spatial point processes: general theory. *Res. Rep., University of Wisconsin-Madison. Available at http://www.math.wisc.edu/ kurtz/papers/shun.pdf*.

Møller, J. and Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall. Baton Rouge.