Spatial Lasso with Application to GIS Model Selection

F. Jay Breidt
Colorado State University

with Hsin-Cheng Huang, Nan-Jung Hsu, and Dave Theobald

September 2005

The work reported here was developed under STAR Research Assistance Agreements CR-829095 and CR-829096 awarded by the U.S. Environmental Protection Agency (EPA) to Colorado State University and Oregon State University. This presentation has not been formally reviewed by EPA. The views expressed here are solely those of the authors. EPA does not endorse any products or commercial services mentioned in this report.

Outline

- Regression for sparse spatial sample
  - layers and neighborhoods for GIS data
  - linear model formulation with many parameters
- Lasso
  - shrinkage and model selection
  - numerical experiment I
- Spatial Lasso
  - modifications for spatial smoothness
  - numerical experiment II
- Application: Prediction of soil moisture index

Responses and Covariates

- Data $\{Y(\boldsymbol{s}_i) : i = 1, \ldots, n\}$
  - observed at spatial locations $\boldsymbol{s}_i \in D \subset \mathbb{Z}^2$
  - D is a regular grid
- Covariate layers $\{x_k(\boldsymbol{s}) : \boldsymbol{s} \in D\}$; $k = 1, \ldots, p$
- **Goal:** find an appropriate model for $Y(\boldsymbol{s})$ as a linear function of $\{x_k(\boldsymbol{s}) : \boldsymbol{s} \in D\}$; $k = 1, \ldots, p$

Model Formulation

- Each layer has an associated neighborhood: $\mathcal{N}_k \subset \mathbb{Z}^2$ is a neighborhood set of $\mathbf{0}$

$$
\begin{aligned}
Y(\boldsymbol{s}_i) = & \sum_{j=1}^{J} a_j \phi_j(\boldsymbol{s}_i) \\
& + \sum_{k=1}^{p} \sum_{\boldsymbol{u} \in \mathcal{N}_k} b_k(\boldsymbol{s}_i, \boldsymbol{u}) x_k(\boldsymbol{s}_i + \boldsymbol{u}) + \varepsilon(\boldsymbol{s}_i)
\end{aligned}
$$

  - $\phi_1(\cdot), \ldots, \phi_J(\cdot)$ are known functions
  - $\{\varepsilon(\boldsymbol{s}_1), \ldots, \varepsilon(\boldsymbol{s}_n)\}$ are iid $\mathrm{N}(0, \sigma^2)$

Spatial Homogeneity

- For each layer k and pixel $\boldsymbol{u}$, expect some spatial homogeneity in $b_k(\cdot, \boldsymbol{u})$:

$$b_k(\cdot, \boldsymbol{u}) = \sum_{l=1}^{L_k} c_{k,l}(\boldsymbol{u}) \psi_{k,l}(\cdot); \quad k = 1, \ldots, p, \ \boldsymbol{u} \in \mathcal{N}_k \ ,$$

  where $\psi_{k,l}(\cdot)$'s are known functions possibly depending on some covariates.

  - can adjust for spatial features like flow directions
  - take $L_k \equiv 1$, $\psi_{k,1} \equiv 1$ in this discussion

$$b_k(\cdot, \boldsymbol{u}) = c_{k,1}(\boldsymbol{u})$$

Linear Model and Least Squares Estimation

- Rewrite as linear model:

$$\boldsymbol{Y} \equiv (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))' = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \,,$$

- Ordinary least squares estimates?
  - large number of parameters: $J + \sum_{k=1}^{p} |\mathcal{N}_k|$
  - OLS has low bias, large variance: **shrinkage**
  - OLS difficult to interpret: **model selection**

Lasso

- **Lasso** (Tibshirani, 1996): least absolute shrinkage and selection operator

  – standardize $\boldsymbol{X}$ as $\boldsymbol{X}^*$

- Minimize RSS subject to L_1 constraint:

$$(\boldsymbol{Y} - \boldsymbol{X}^*\boldsymbol{\beta}^*)'(\boldsymbol{Y} - \boldsymbol{X}^*\boldsymbol{\beta}^*),$$

  subject to $\sum |\beta_j^*| \leq t$, a tuning parameter

- **Key feature:** some estimated β_j^* can be exactly zero

Lasso, Continued

- Equivalently, minimize

$$(\boldsymbol{Y}-\boldsymbol{X}^{*}\boldsymbol{\beta}^{*})'(\boldsymbol{Y}-\boldsymbol{X}^{*}\boldsymbol{\beta}^{*})+\lambda\sum_{j=1}^{m}|\beta_j^{*}|,$$

  - if priors are independent Laplace, $\{\hat{\beta}_j^{*}\}$ are posterior modes
- **Least angle regression** (LARS, Efron et al., 2004) provides fast algorithm for Lasso
  - same computational order as OLS applied to full set of covariates
  - `lars` package in R (`http://cran.r-project.org`)

Numerical Experiment I: Mean Function

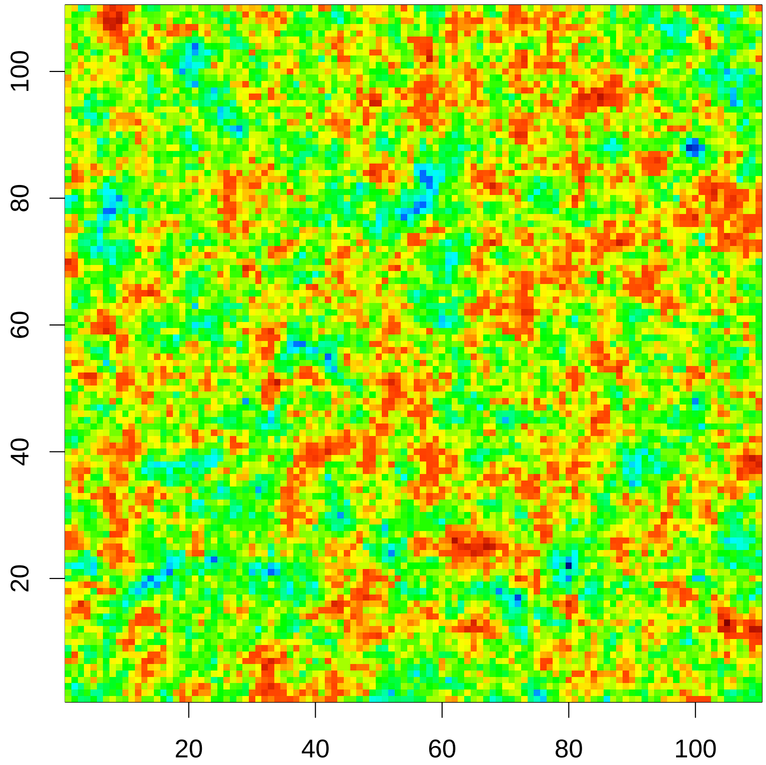
- Consider six covariate layers
  - simulated as Gaussian random fields
  - exponential covariance, strong dependence
- Construct true mean function:

$$\mu(\boldsymbol{s}) = 1 + \frac{1}{3}\left\{x_1(\boldsymbol{s}) + \sum_{j=-1}^{1} x_1(\boldsymbol{s} + (j,1)) + \sum_{j=-2}^{2} x_1(\boldsymbol{s} + (j,2))\right\} + \frac{1}{3}\sum_{j=-1}^{1}\sum_{k=-1}^{1} x_2(\boldsymbol{s} + (j,k))$$

- True neighborhoods:
  - x_1: inverted 5-3-1 pyramid
  - x_2: centered 3×3 block
  - x_3, x_4, x_5, x_6: empty

Simulated Covariate Layer

- Strong spatial dependence in $x_1(\cdot)$



Estimation and Prediction

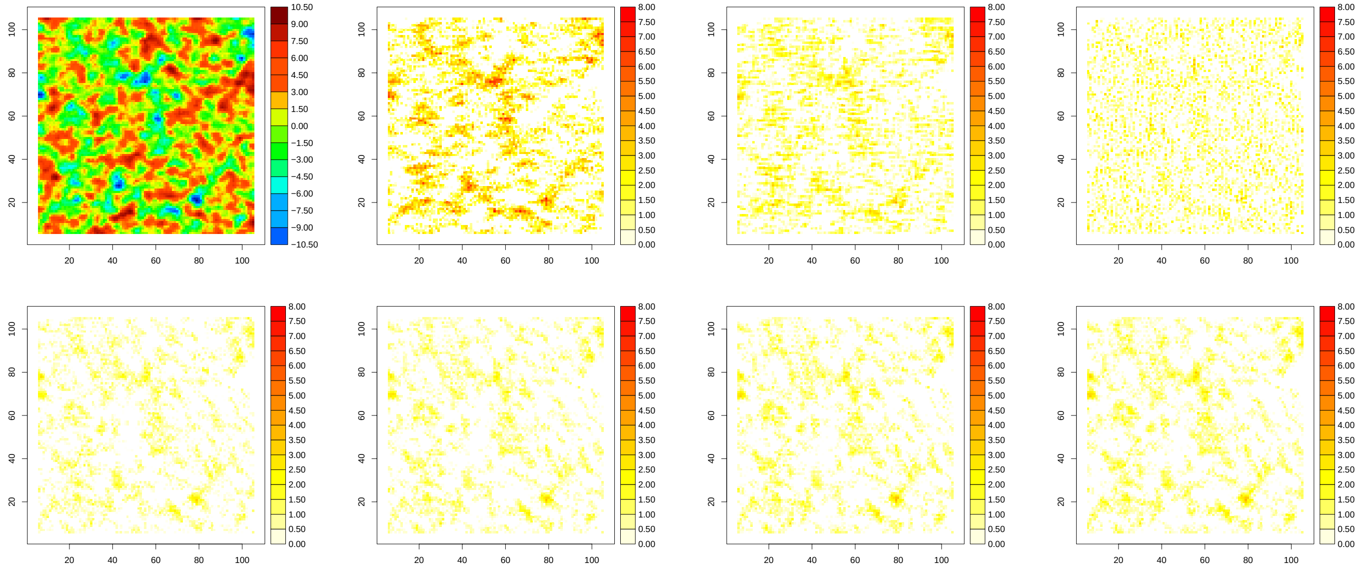
- Sample 100 sites from 100×100 spatial domain
  - observed response = mean function plus noise
- Estimate and predict using OLS or Lasso
  - OLS models with correct layers:

$$
\begin{aligned}
Y(\boldsymbol{s}_i) &= \beta_0 + \sum_{l=1}^{2} \beta_l x_l(\boldsymbol{s}_i) + \varepsilon(\boldsymbol{s}_i); \quad i = 1, \ldots, n, \\
Y(\boldsymbol{s}_i) &= \beta_0 + \sum_{l=1}^{2} \sum_{j=-1}^{1} \sum_{k=-1}^{1} \beta_{l,j,k} x_l(\boldsymbol{s}_i + (j,k)) + \varepsilon(\boldsymbol{s}_i); \quad i = 1, \ldots, n, \\
Y(\boldsymbol{s}_i) &= \beta_0 + \sum_{l=1}^{2} \sum_{j=-2}^{2} \sum_{k=-2}^{2} \beta_{l,j,k} x_l(\boldsymbol{s}_i + (j,k)) + \varepsilon(\boldsymbol{s}_i); \quad i = 1, \ldots, n,
\end{aligned}
$$

  - Lasso neighborhoods, $\mathcal{N}^{(2q+1)}$: $(2q+1) \times (2q+1)$ blocks

True Mean Function and Absolute Prediction Errors

- Qualitatively similar for weak dependence



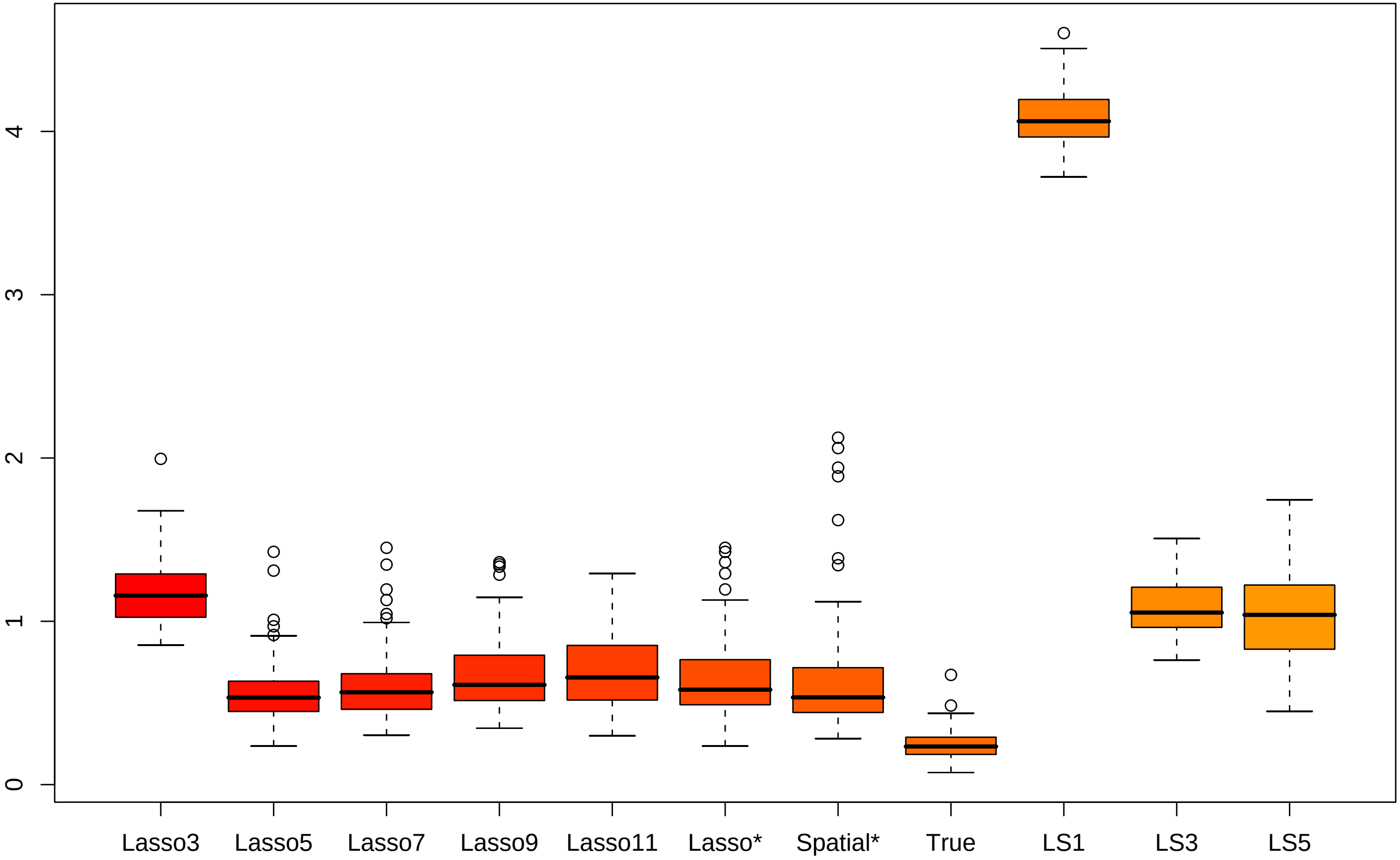
Average Squared Error

- ASE over 100×100 spatial domain:

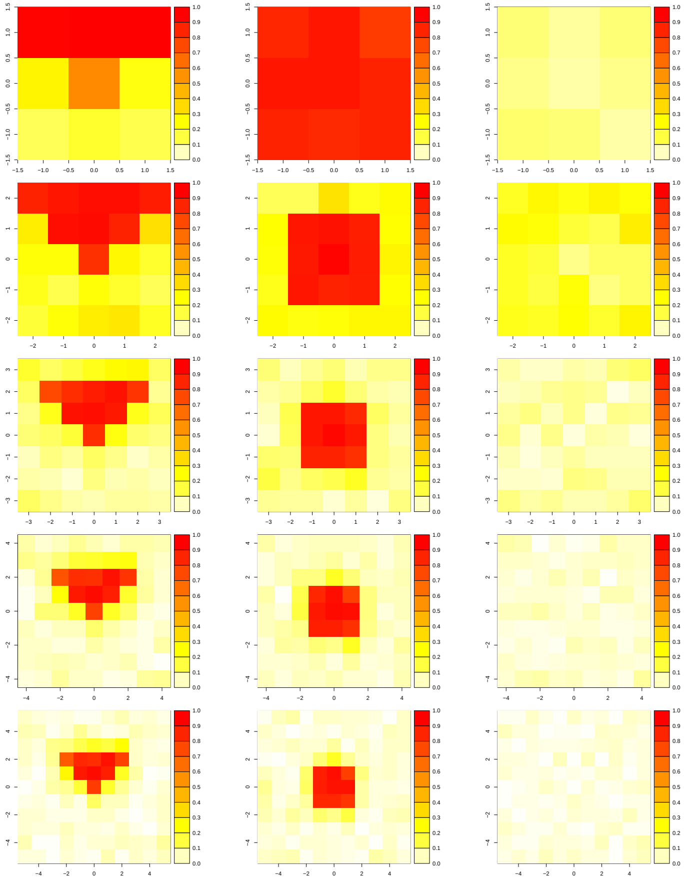
$$ASE = \frac{1}{(100)^2} \sum_{\boldsymbol{s} \in D} (\hat{\mu}(\boldsymbol{s}) - \mu(\boldsymbol{s}))^2,$$

- Conduct 100 simulation replicates
  - get 100 ASE's for LS1, LS3, LS5, Lasso3, . . . , Lasso11
  - produce boxplots of ASE's

Average Squared Error Under Strong Dependence



Estimated Neighborhoods Using Lasso (Strong)



Spatial Smoothness

- Above model assumes **spatial homogeneity** of regression coefficients *across* neighborhoods
- No assumption of **spatial smoothness** of regression coefficients *within* neighborhoods
  - in many applications, reasonable to assume smoothness of $\{c_{k,l}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{N}_k\}$
- Ordinary Lasso does not account for spatial smoothness

Spatial Lasso

- Allow for smoothness of coefficients
- Assume spatial dependence prior for $\boldsymbol{\beta}$:
  - $\boldsymbol{\Gamma}$ is prior correlation matrix of $\boldsymbol{\beta}$
  - $\boldsymbol{\beta}^{**} \equiv \boldsymbol{\Gamma}^{-1/2}\boldsymbol{\beta^*} \sim$ independent Laplace
- **Spatial Lasso** obtained by minimizing

$$(\boldsymbol{Y} - \boldsymbol{X}^{**}\boldsymbol{\beta}^{**})'(\boldsymbol{Y} - \boldsymbol{X}^{**}\boldsymbol{\beta}^{**}) + \lambda \sum_{j=1}^{m} |\beta_j^{**}|,$$

where $\boldsymbol{X}^{**} \equiv \boldsymbol{X}^{*}\boldsymbol{\Gamma}^{1/2}$

  - computation via modification of LARS: **equi-projection regression**

Numerical Experiment II

- Mean function: for $x(\boldsymbol{s})$ iid $\mathsf{N}(0,1)$

$$\mu(\boldsymbol{s}) = 1 + 5\sum_{j=-2}^{2}\sum_{k=-2}^{2} w_{j,k}\, x(\boldsymbol{s} + (j,k)); \quad \boldsymbol{s} \in D,$$

  and Gaussian weight function

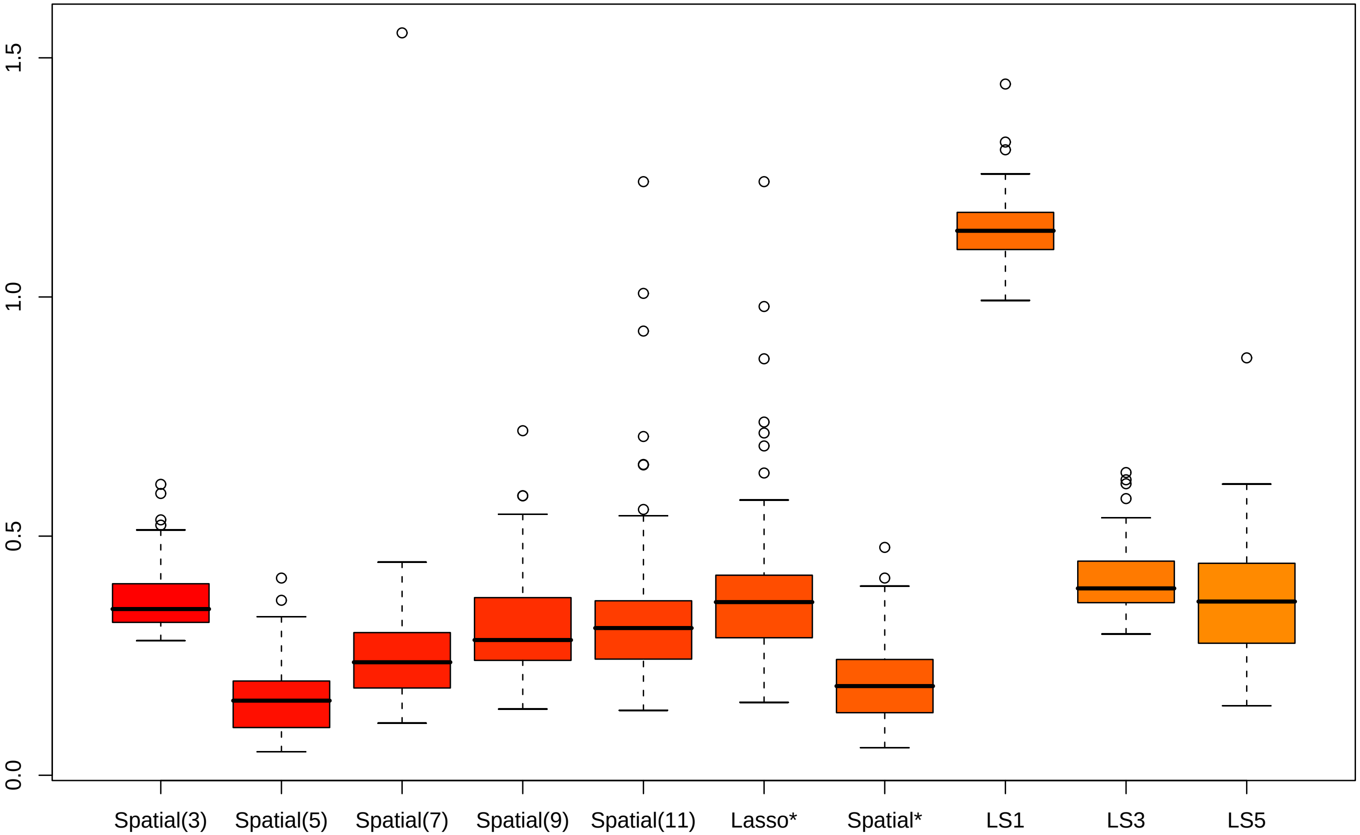
$$w_{j,k} \equiv \frac{\exp\left(-(j^2+k^2)/4\right)}{\sum_{j=-2}^{2}\sum_{k=-2}^{2}\exp\left(-(j^2+k^2)/4\right)}$$

- Sample $n = 100$ sites and generate $\{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\}$ by adding noise

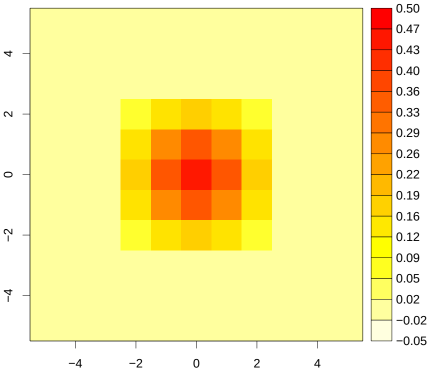
Spatial Lasso for Numerical Experiment II

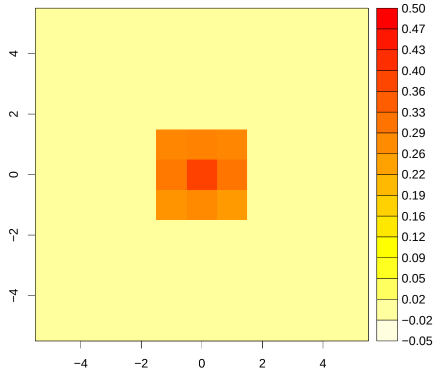
- Compare OLS, Lasso, Spatial Lasso
- Apply Spatial Lasso with exponential covariance structure
- Choose smoothness parameter and neighborhood size via grid search:
  - neighborhood sets $\{\mathcal{N}^{(2q+1)} : q = 1, \ldots, 5\}$
  - smoothness parameters $\gamma = 0, 1, \ldots, 5$
  - choose combination with best ten-fold cross-validation

Average Squared Error for Experiment II

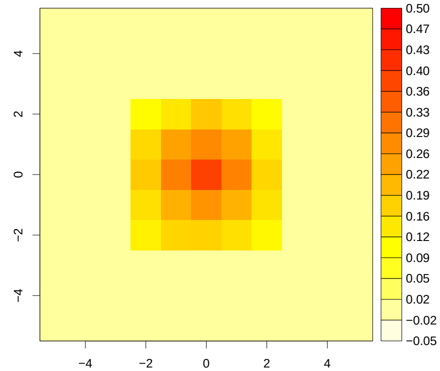


Estimated Coefficients Using Spatial Lasso

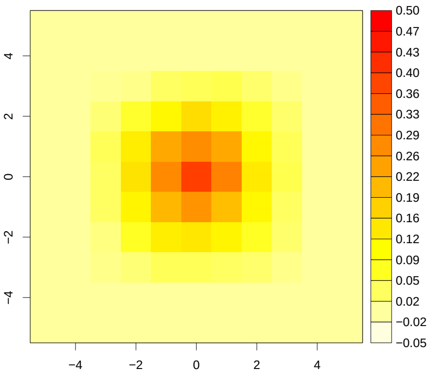




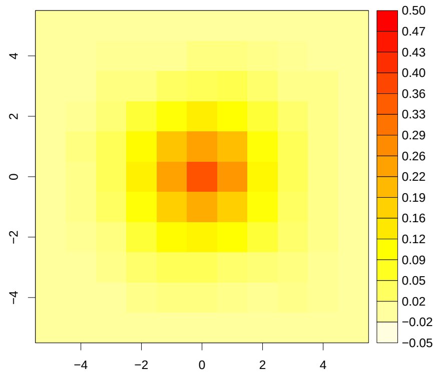
(b)



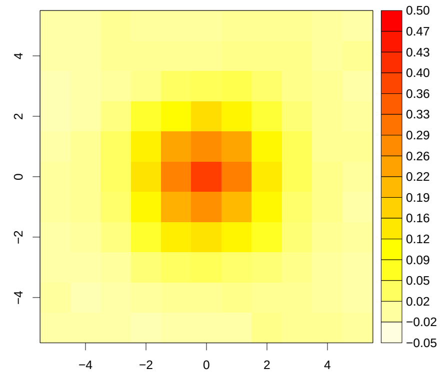
(c)



(d)



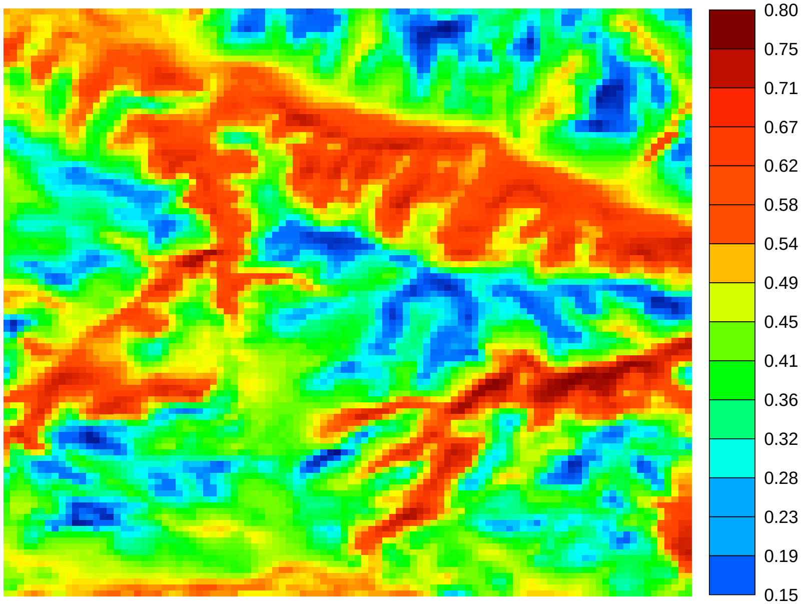
(e)



(f)

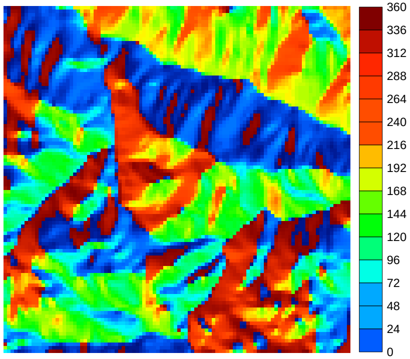
Application: Prediction of Soil Moisture Index

- Select $n = 100$ sites via simple random sampling without replacement

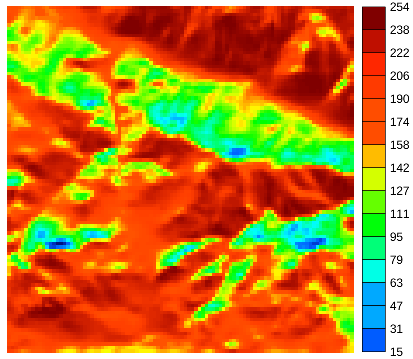


Covariates for Soil Moisture Prediction

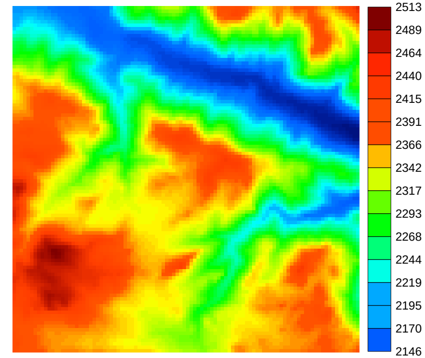
- Aspect, hill shade, elevation, slope, precipitation



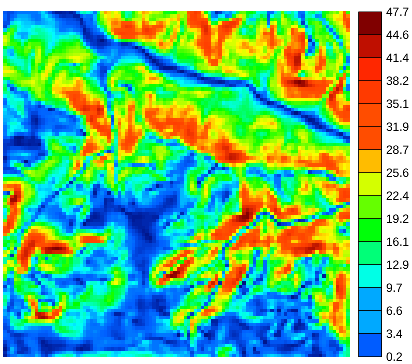
(a)



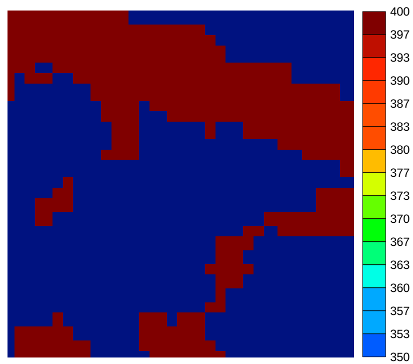
(b)



(c)



(d)



(e)

Covariates for Soil Moisture Prediction

- Response: soil moisture index
- Basic covariate layers:
  - aspect, hill shade, elevation, slope, precipitation
  - 100×100 regular grid
- Expanded covariate layers (14 total):
  - average elevation, slope, precipitation on 3×3 blocks
  - average elevation, slope, precipitation on 9×9 blocks
  - aspect*elevation, aspect*slope, aspect*precipitation

Neighborhoods for Soil Moisture Prediction

- Neighborhoods:
  - use single pixel for elevation, 3×3 elevation, 9×9 elevation
  - same for slope and precipitation
  - (up to 9×9 neighborhood, but highly constrained parameters)
  - for remaining variables, use

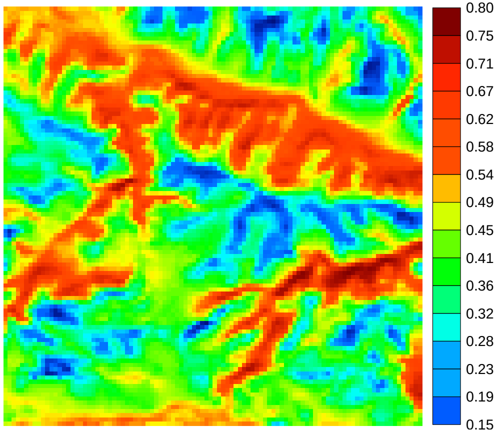
$$\mathcal{N}^{(2q+1)} = \{0, \pm 1, \pm 2, \ldots, \pm(2q+1)\}^2$$

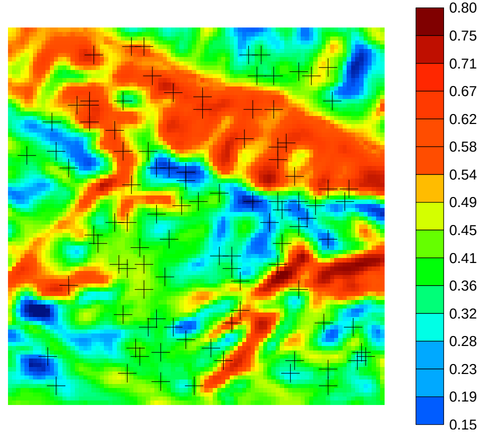
with $q = 0, 1, \ldots, 5$

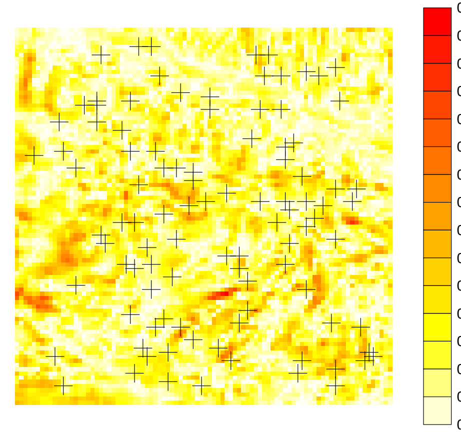
Spatial Lasso for Soil Moisture Prediction

- Apply spatial Lasso with exponential covariance structure
  - grid search on $\gamma = 0, 1, \ldots, 5$, crossed with neighborhoods above
  - 36 possible combinations
- Smallest ten-fold cross-validation value at $\gamma = 2$ and $q = 5$

True and Predicted Soil Moisture Index







Summary

- Lasso approach to GIS model selection and estimation
  - simultaneous layer selection, neighborhood selection, and estimation
  - dominates OLS for unknown layers/neighborhoods
  - extensible to neighborhood transformation
- Spatial Lasso
  - accounts for spatial smoothness within neighborhoods
  - dominates Lasso if smoothness is present
- Promising results in prediction of soil moisture index