Spatial variable selection

Brian Reich

Department of Statistics
North Carolina State University

December 13, 2015

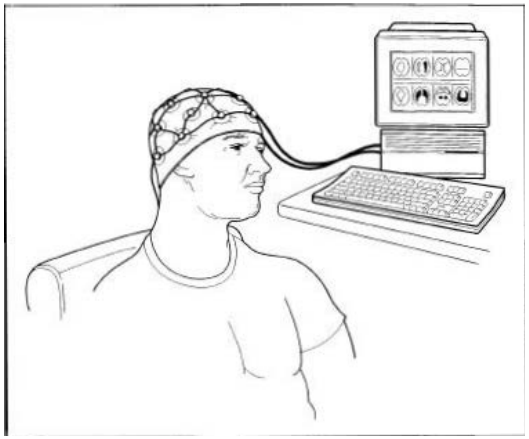
Joint with: Jian Kang (UMich) and Ana-Maria Staicu (NCSU)

Potential applications of spatial variable selection

- Let $\beta(s)$ be the regression coefficient at location s
- There are many examples where you might want a separate slope for each location:
  - $\beta(s)$ is the climate change effect at s
  - $\beta(s)$ is the time trend in air pollution level at s
  - $\beta(s)$ is the health effect of particulate matter at s
- In all cases, we might assume β is smooth and sparse:
  - **Smooth**: the spatial process $\beta(s)$ is continuous in s
  - **Sparse**: $\beta(s) = 0$ in many locations

Example: EEG study of alcoholism

Goal: Study the relationship between the brain activity as measured through EEG signals and genetic predisposition to alcoholism



EEG signals record the electrical activity in the brain by measuring the current flows produced when the neurons are activated

EEG study of alcoholism

Study: UCI KDD
https://kdd.ics.uci.edu/databases/eeg/eeg.data.html

- Data: 77 alcoholic subjects + 45 controls
- 64 electrodes sampled at 128Hz

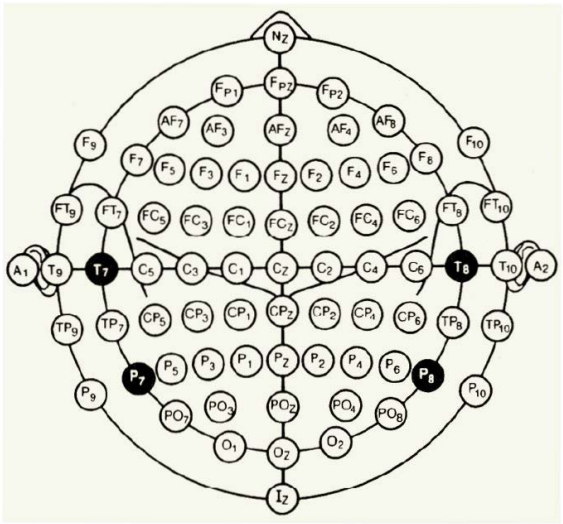
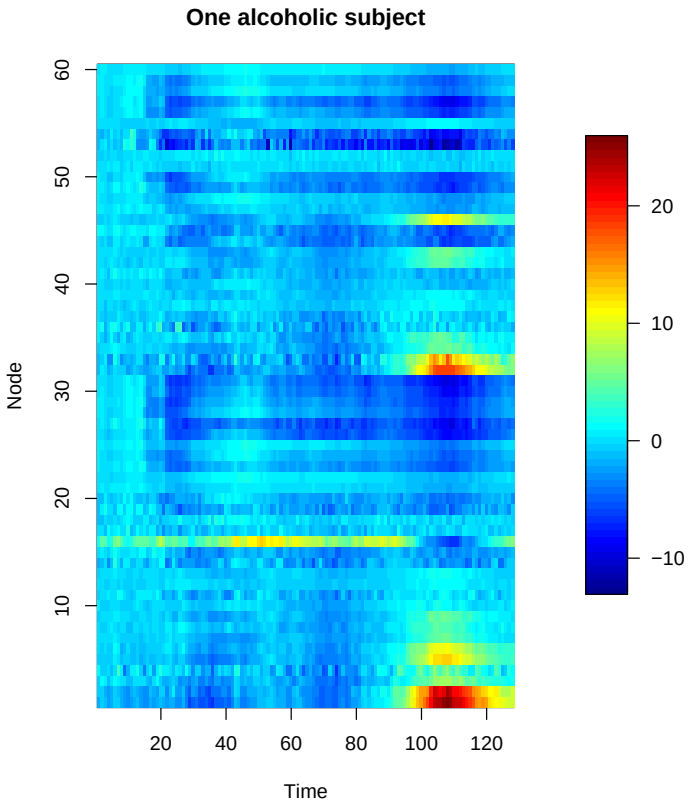
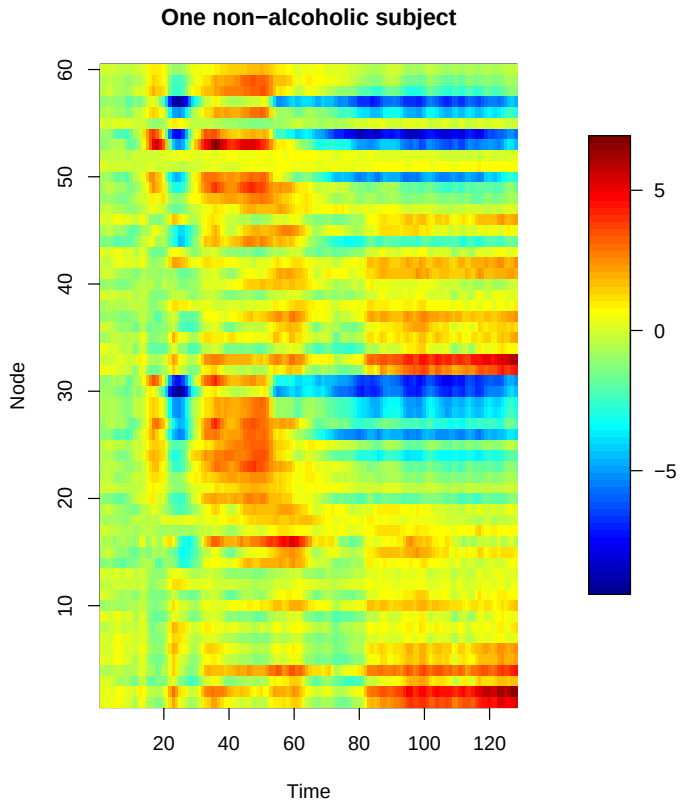


FIG.1. Modified combinatorial nomenclature for the 10-10 system.

- **Goal**: Identify regions most predictive of alcoholism

EEG study of alcoholism





Scaler-on-image regression framework

- Observed data for i^{th} subject
  - Y_i is scalar outcome
  - $X_i = (X_{i1}, ..., X_{ip})^T$ is an image/array

  Model: $Y_i = \sum_{j=1}^{p} \beta(s_j) X_{ij} + \epsilon_i$

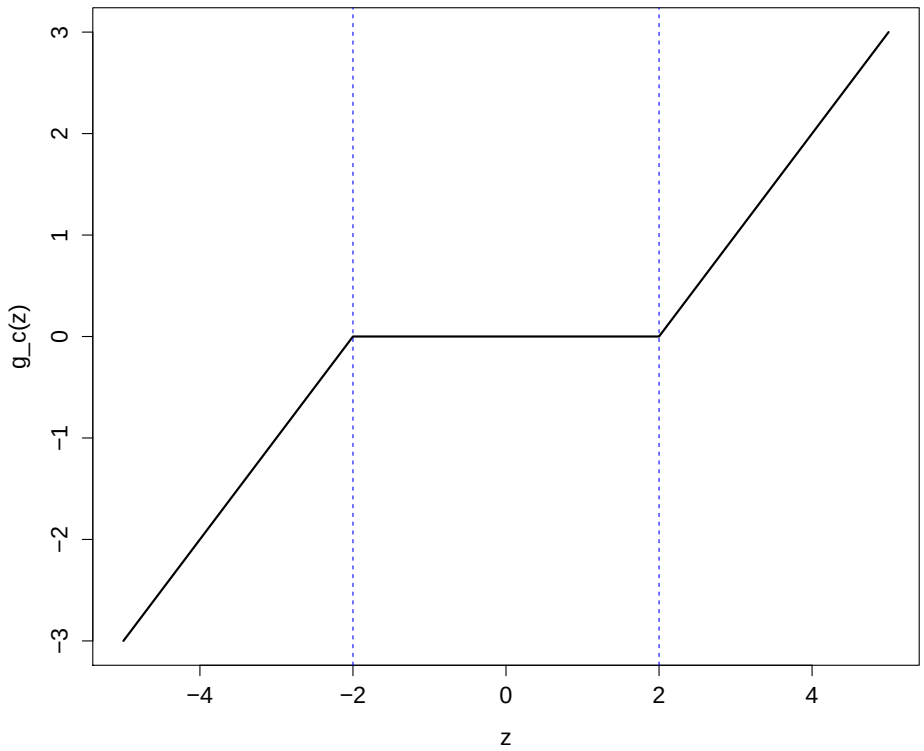
- β is the coefficient image
- Assumption: β is a sparse and piecewise smooth function
- $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$
- **Goal:** Estimation and inference of $\beta(s)$

Literature on scaler-on-image regression

- Frequentist approaches: Tibshirani (JRSSB1996); Tibshirani et al. (JRSSB05), Tibshirani & Taylor (AoS11); Reiss & Ogden (Bcs10); Wang & Zhu (Bka15)
  - No inferential methods available using this approach
- Bayesian approaches: Goldsmith et al. (JCGS14); Li et al. (AoAS15)
  - Stability issues due to using two latent processes to model the coefficient image
  - No smooth transition between the zero areas and non-zero parts

Soft Thresholded Gaussian Process ($\mathcal{STGP}$)

- Bayesian approach: assume $\beta(s) = g_c\{Z(s)\}$
  - $Z(s)$ is latent Gaussian Process with zero mean and continuous covariance function
  - g_c is real-valued function with $g_c(z) = sign(z)(|z| - c)$ for some specified threshold c



Low-rank spatial model representation

Higdon et al (1999):

$$Z(s) = K_h(s - \tau_1)a_1 + \ldots + K_h(s - \tau_L)a_L$$

- K_h is local kernel function with Kernel bandwidth h (e.g. tapered Gaussian kernels with bandwidth h)
- The τ_l's are fixed spatial knots
- Dimension is reduced from p to L, which makes it possible to handle large images

CAR model

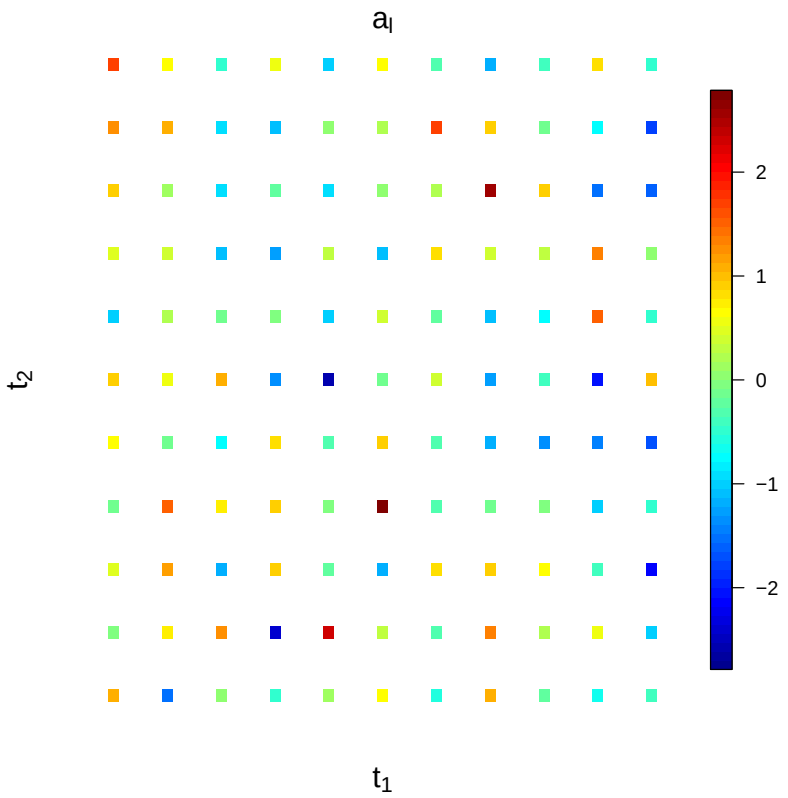
- Use conditionally autoregressive (CAR) prior to account for large-scale spatial dependence in the a_l (Nychka, et al JCGS15)

- The CAR prior can be defined by the full conditional distribution of one site given all others:

$$a_l | a_k \text{ for all } k \neq l \sim N(\rho \bar{a}_l, \sigma_a^2 / m_l),$$

  - $\bar{a}_l$ is the mean of a at site l's m_l neighbors
  - $\rho \in (0, 1)$ controls the strength of spatial dependence
  - σ_a^2 controls the variance

Illustration - Kernel smoothing ($a \rightarrow Z$)



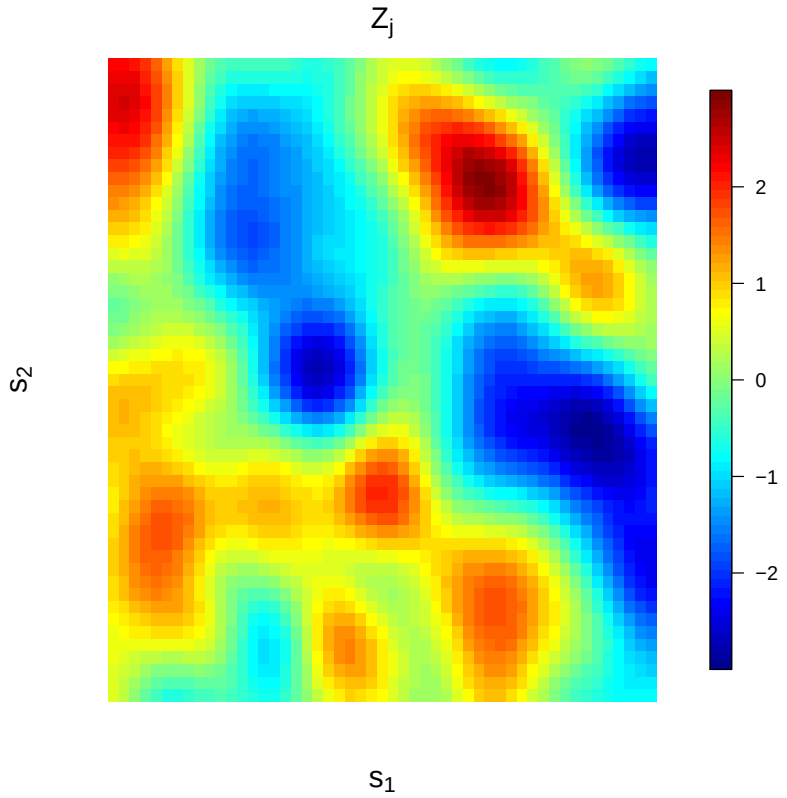


Illustration - Soft thresholding ($Z \rightarrow \beta$)

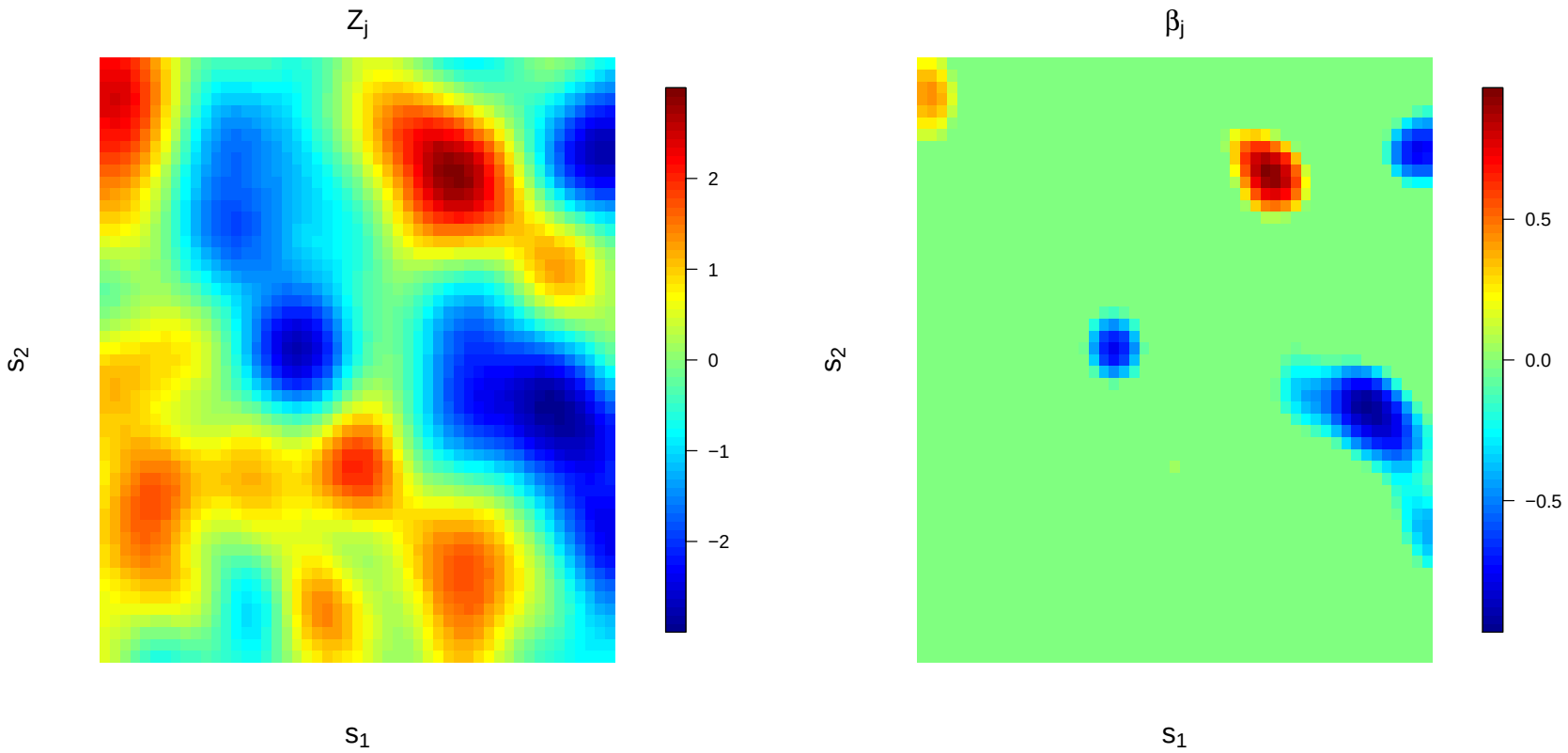
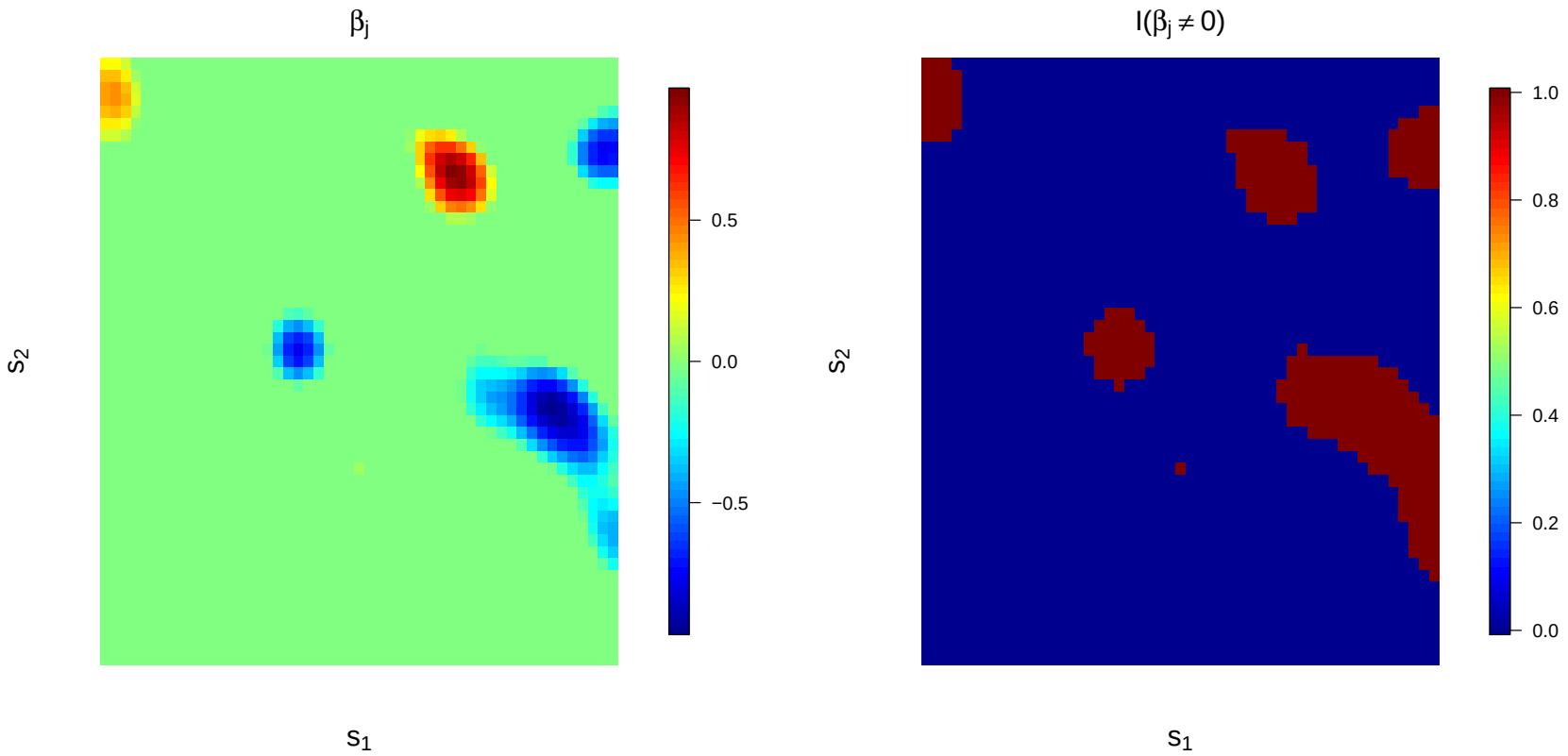


Illustration - Sparsity



Full Model

The full model can be written:

$$
\begin{aligned}
\mathrm{Y}|\mathrm{X},\beta,\sigma^2 &\sim N(\mathrm{X}\beta,\sigma^2 I_n) \\
\beta(\mathrm{s}) &= g_c\{Z(\mathrm{s})\} \\
Z(\mathrm{s}) &= K_h(\mathrm{s}-\tau_1)a_1+\ldots K_h(\mathrm{s}-\tau_L)a_L \\
\mathrm{a} &\sim N_L(0,\sigma_a^2(\mathrm{M}-\rho\mathrm{A})^{-1})
\end{aligned}
$$

CAR prior: $\mathrm{M}=\mathrm{diag}(m_1,\ldots,m_L)$ and A is the adjacency matrix with (k,l) element equal 1 if $k\sim l$ and zero otherwise

Advantages/novelties

- The proposed method uses a single spatial process to control both sparsity and smoothness
- As a result there is a gradual transition between zero and non-zero regions
- Allows full inference and stable computations
- It allows us to study theoretical properties
- Easily extended to incorporate additional covariates or generalized responses

Theoretical properties

Proof of large support: Assume the true signal $\beta^0(s)$ is (i) piecewise smooth, (ii) sparse, and (iii) continuous. If there exists a latent process $Z(s)$ such that $\beta^0(s) = g_c\{Z(s)\}$. Then the STGP $\beta(s)$ satisfies

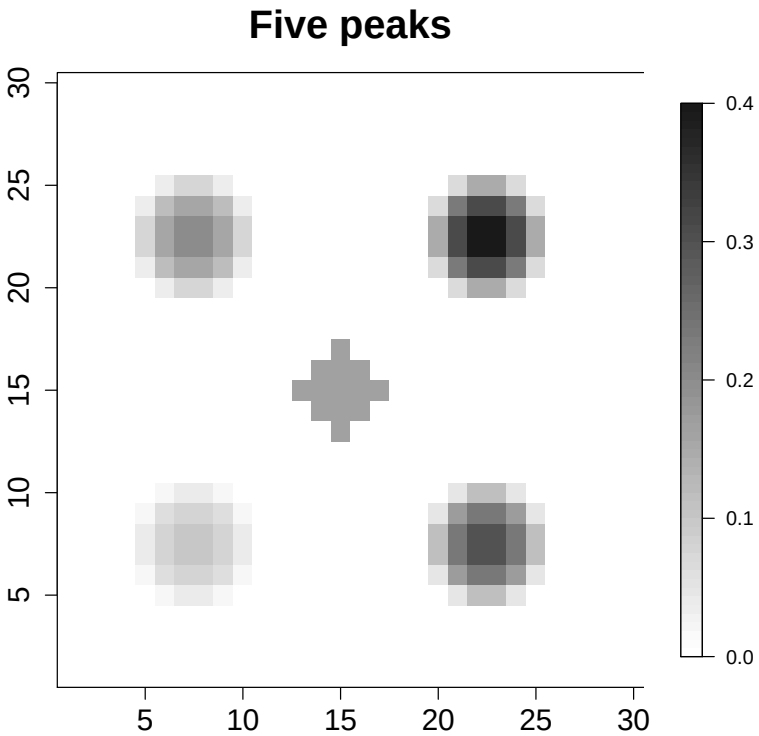
$$\Pi\left(\|\beta(s) - \beta^0(s)\|_\infty < \epsilon\right) > 0 \text{ for all } \epsilon > 0$$

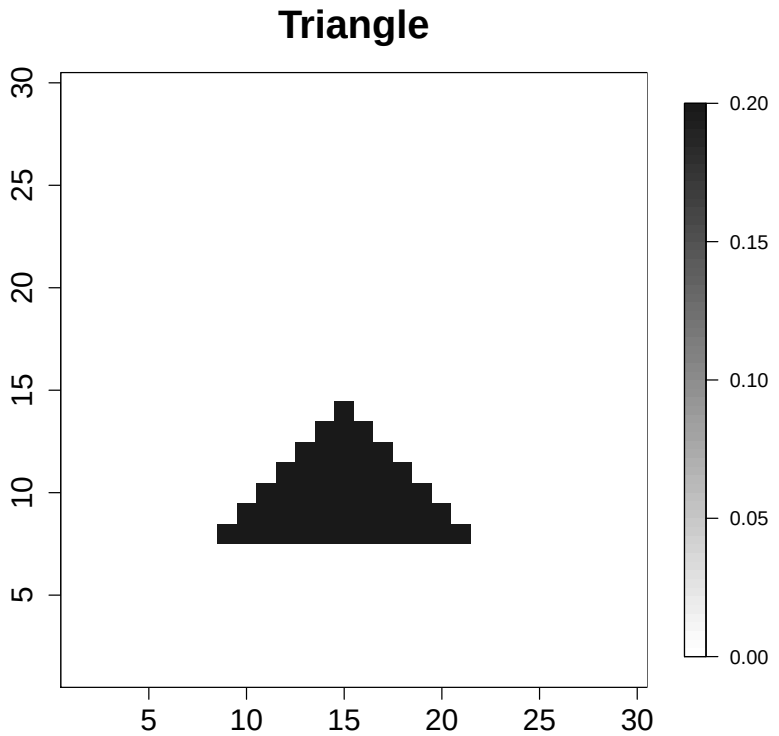
Posterior consistency: Assume regularity conditions for the design matrix of X_i's, and of kernel K and that true signal β^0 is as above. The number of spatial locations p is such that $\log(p) = o(n)$. Then as $n \to \infty$, the posterior distribution satisfies

$$\Pi\left[\|\beta(s) - \beta^0(s)\|_\infty < \varepsilon \mid Y, X\right] \to 1$$

Simulation study set-up

- Model: $Y_i \sim \text{Normal}\left(\sum_{j=1}^{p} \beta(s_j) X_{ij}, \sigma^2\right)$
- Images are generated on a 30×30 grid so $p = 900$
- True signal $\beta(s)$ is either:





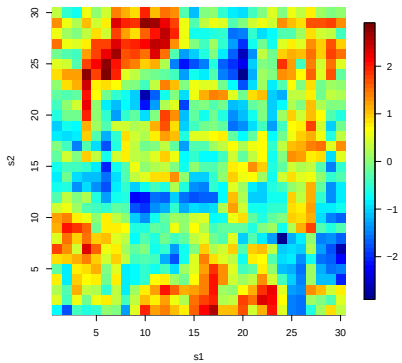
Set-up (cont'd)

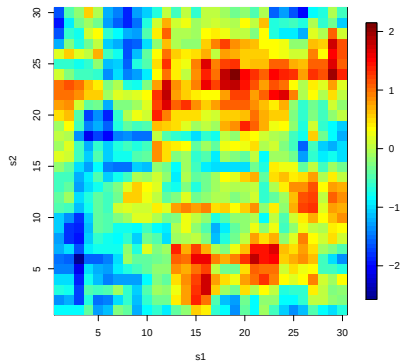
The covariates are generated at the p locations using either an exponential correlation or to share structure ("SS") with β

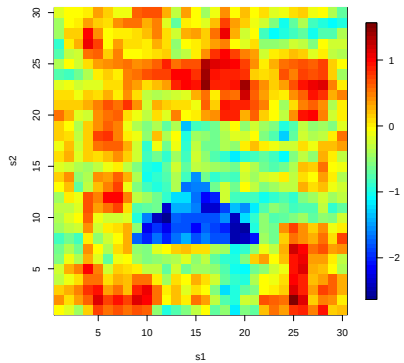
A1. $X_i \sim \text{GP}(0, \text{Exp}(\rho_X))$; 'Exp(3)' or 'Exp(6)'

A2. $X_i(\text{s}) = e_i \beta(\text{s}_j) + 0.5 U_i(\text{s})$

$e_i \sim N(0, \tau^2)$, $U_i \sim \text{GP}(0, \text{Exp}(3))$; 'SS (2)' or 'SS(4)'







Performance evaluation

- Assess performance using MSE and computing time
- We compare $\mathcal{STGP}$ with:
  - **Lasso** (Tibshirani, JRSSB1996)
  - **Fused lasso** (Tibshirani et al JRSSB05, Tibshirani & Taylor AoS11)
  - **fPCR**: smoothing the image X_i first (Xiao et al JRSSB13) and then doing functional principal component regression
  - **Ising**: Bayesian Scalar on Image regression (Goldsmith et al JCGS14)
  - **GP**: the non-threshold GP prior

Results: MSE ($\times$ 1,000)

Sample size $n = 100$, standard deviation of conditional response $\sigma = 5$. Results based on 100 simulations.

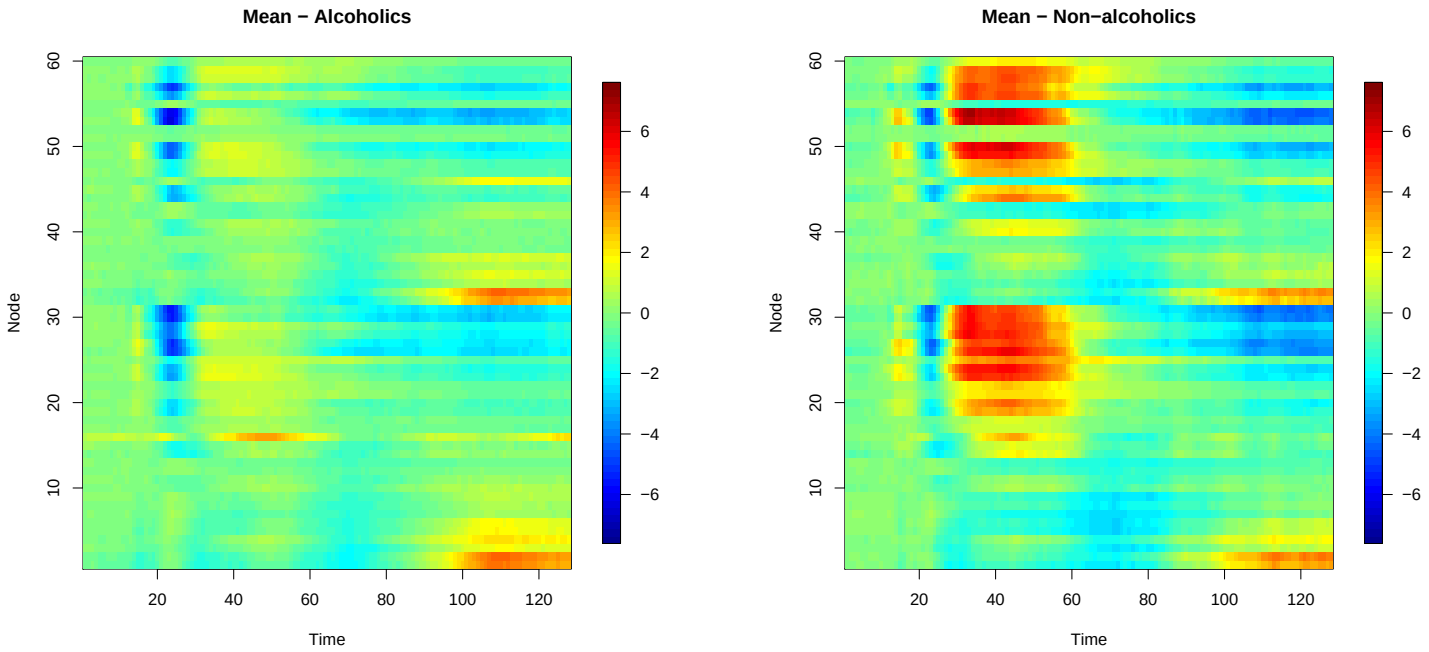
True β	Cov(X)	Fused lasso	fPCR	Ising	$\mathcal{GP}$	$\mathcal{STGP}$
5 peaks	Exp(3)	18.48	3.67	4.44	2.63	1.65
	Exp(6)	2.66	3.33	4.14	2.07	1.93
Triangle	Exp(3)	18.08	1.83	2.75	1.80	0.82
	Exp(6)	4.32	1.63	2.64	1.76	0.88
	SS(2)	70.65	0.98	2.77	3.28	1.40
	SS(4)	71.23	0.34	3.18	3.39	1.81

Results: Time (minutes) when $n = 100$

True β	Cov (X)	Fused lasso	fPCR	Ising	$\mathcal{GP}$	$\mathcal{STGP}$
5 peaks	Exp(3)	16.77	5.40	27.61	4.81	17.69

Recall EEG data

- $Y_i = 1$ for alcoholics and $Y_i = 0$ otherwise
- X_i is a 60×128 image



- Goal: Study EEG correlates of genetic predisposition to alcoholism

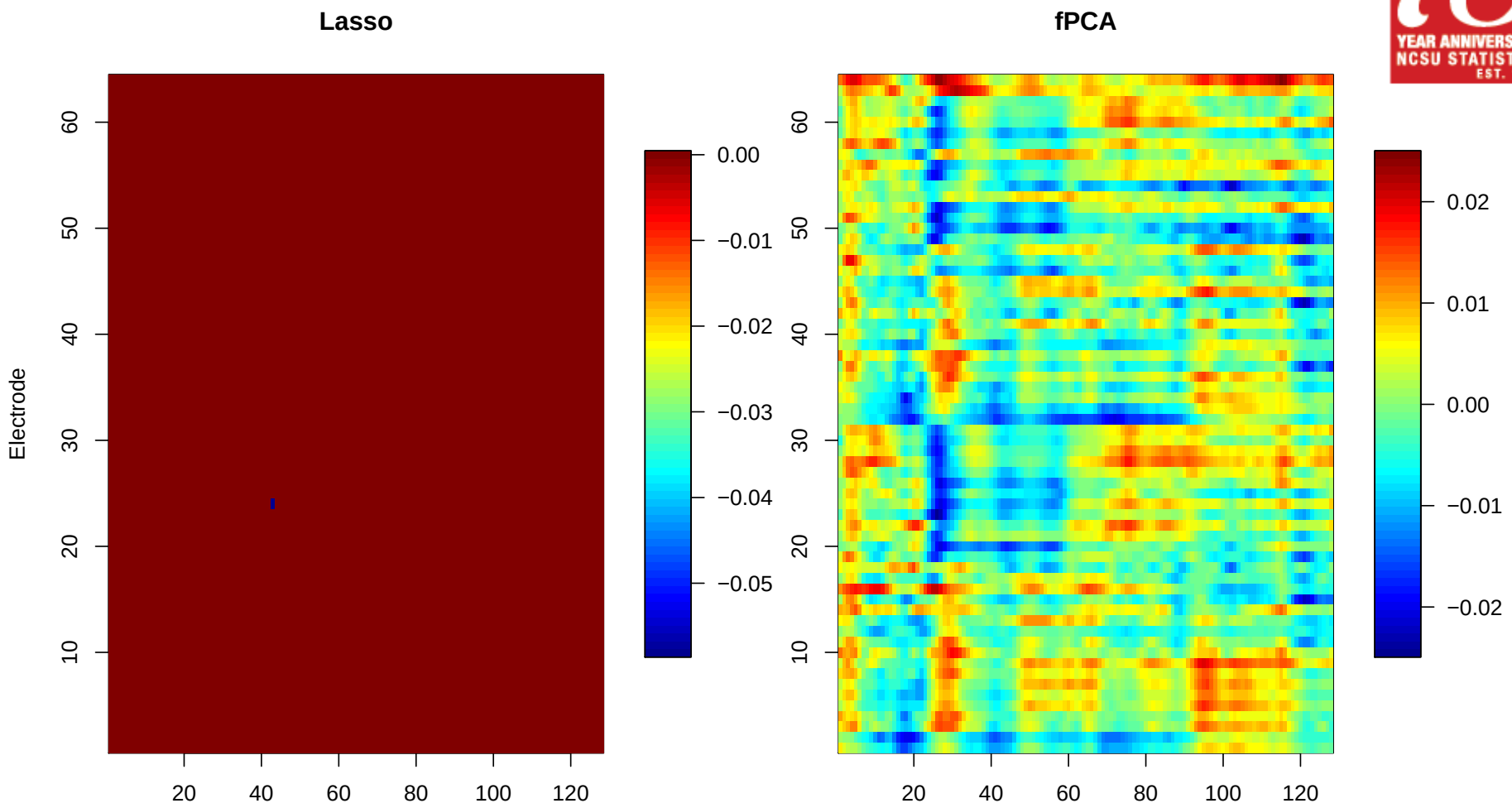
Implementation details

- Probit regression: $\text{Prob}(Y_i = 1 | X_i, \beta) = \Phi\left[\sum_j X_{ij}\beta(s_j)\right]$
- We use knots in every other column and row, with a different CAR dependence parameter in each direction
- The prior for the threshold c is somewhat informative,

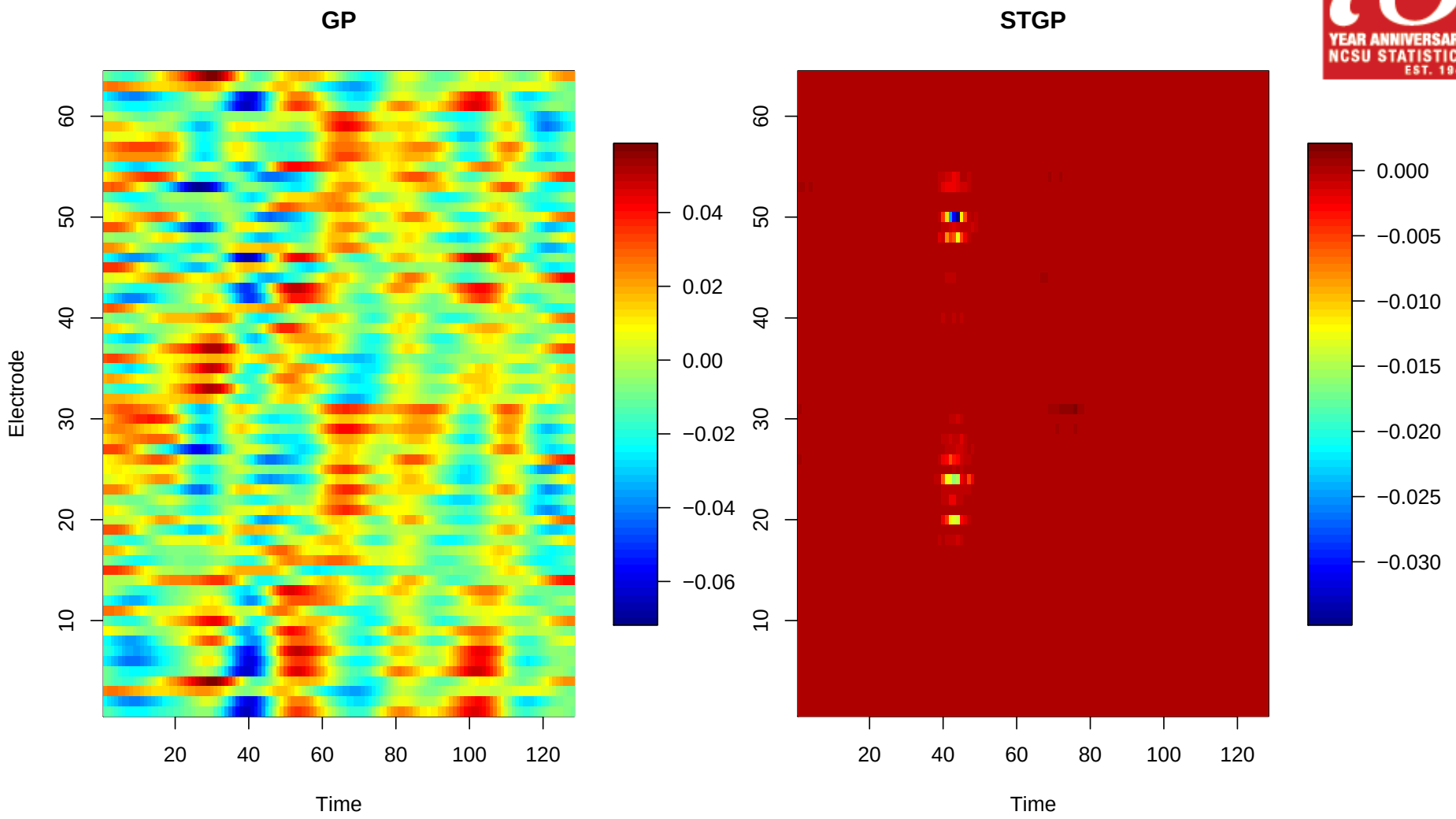
$$c \sim \text{Uniform}(1.43, 1.96)$$

- This gives about 5-15% inclusion probability

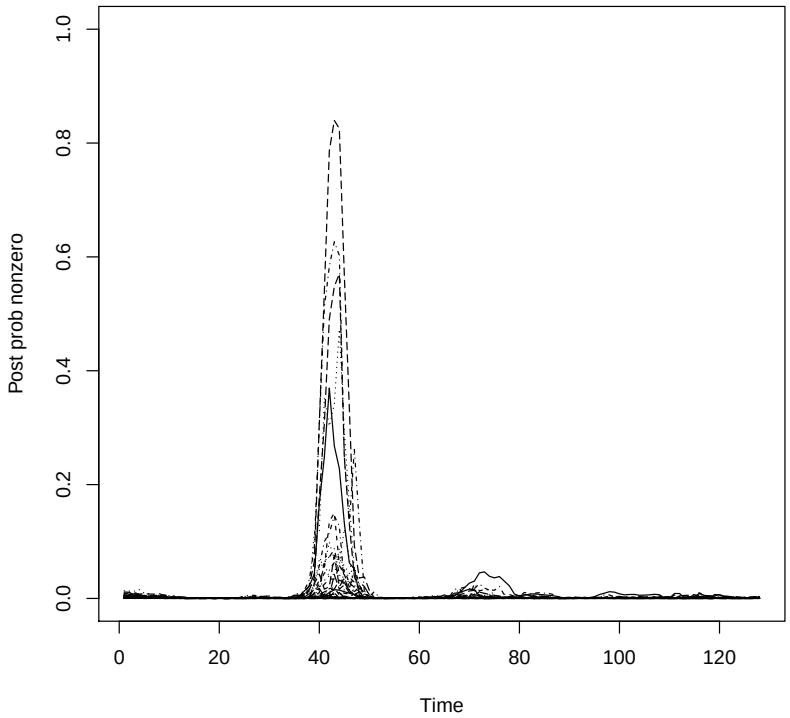
Estimated $\beta(s)$

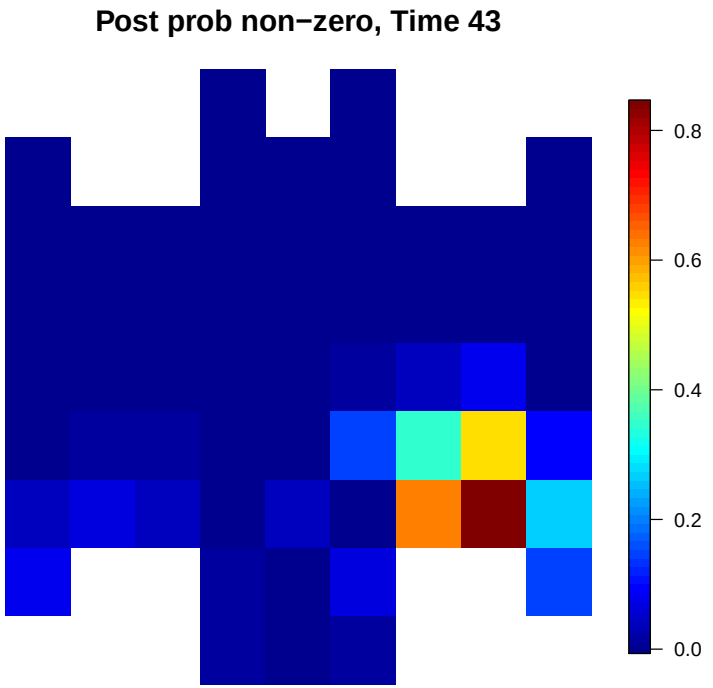


Estimated $\beta(s)$

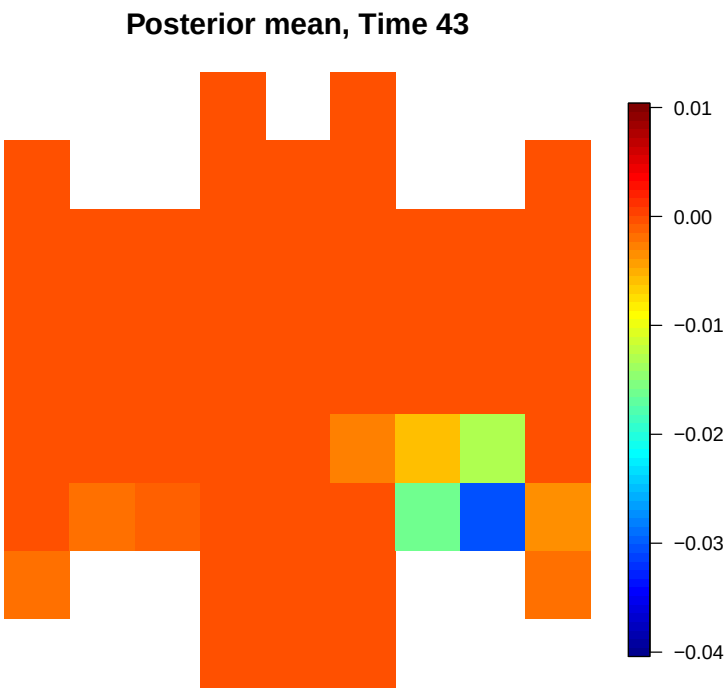


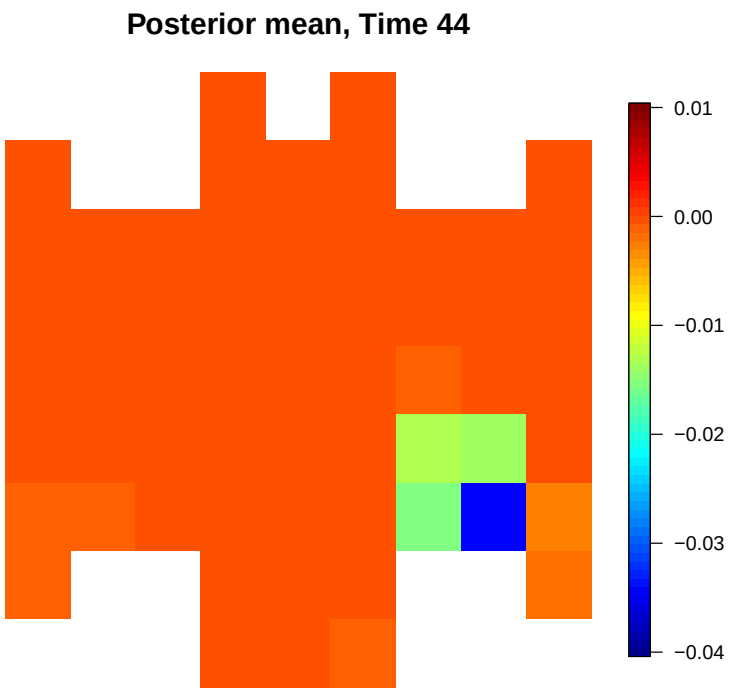
STGP estimates



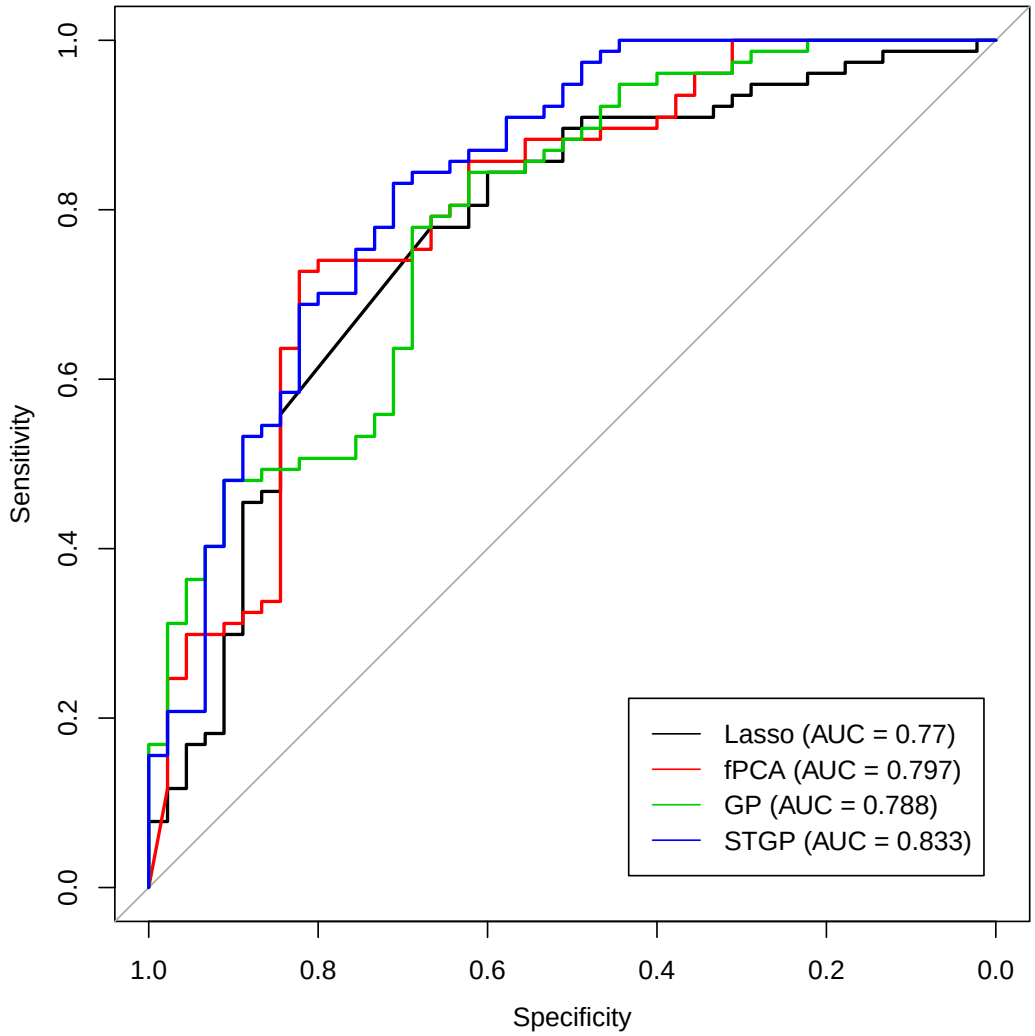


STGP estimates





ROC from 5-fold CV



Discussion

- Soft Thresholded Gaussian Process-based modeling for high dimensional regression, where the signal is sparse and piecewise smooth
- Single process to control the smoothness and sparsity of the signal has computational advantages and allows to study theoretical properties
- Low rank representation of the latent process allows the method to be applicable for high-dimensional predictors
- We have also applied this method in applications with multiple covariates, and responses at each spatial location

THANK YOU !

For comments or questions, please contact me at

brian_reich@ncsu.edu

Thanks to NIH, NSF, and EPA for financial support