Bayesian Spatial Model Selection for Detection and Identification in Chemical Plumes Based on Hyperspectral Imagery Data

Abel Rodriguez - UC, Santa cruz
Nicole Mendoza - UC, Santa Cruz

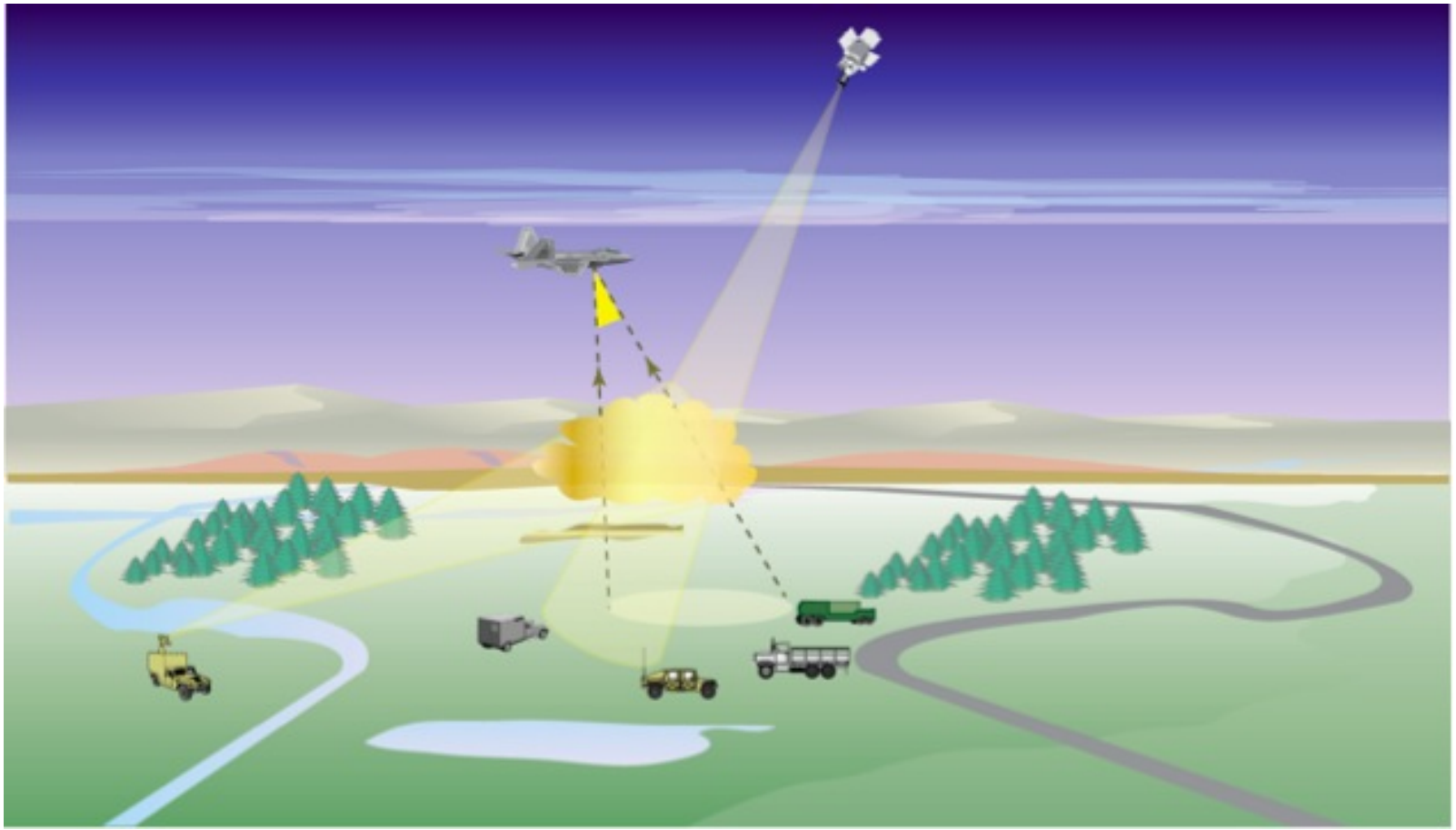
Conference on Applied Statistics in Defense 2016

1. Hyperspectral Imaging
2. Identification
3. Bayesian spatial selection
4. Semiparametric approaches
5. Discussion

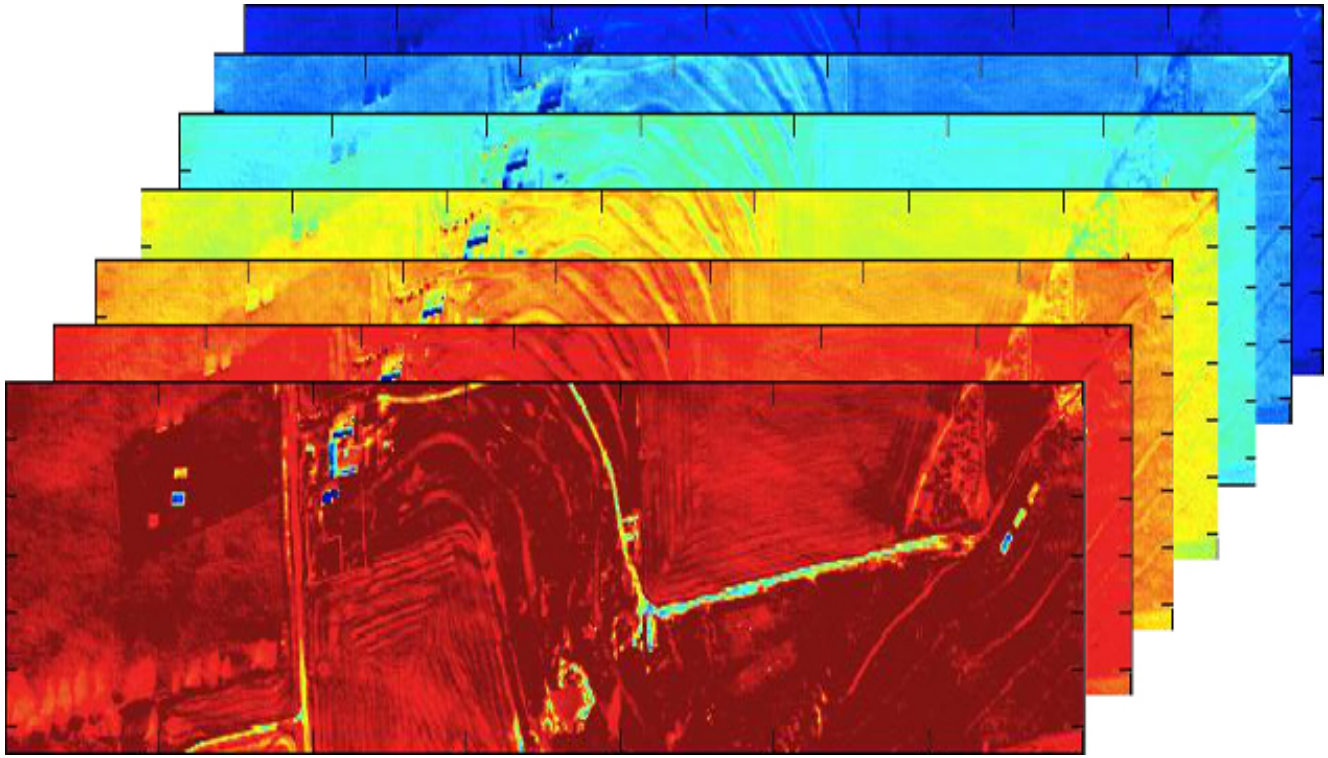
Hyperspectral sensors

- "Hyperspectral" = Radiance recorded on *dozens* of adjacent, narrow frequencies $\Rightarrow$ We focus on the LWIR spectrum (more on this later).
- Multiple applications in defense and national security that involve detection/identification/quantification of materials/chemicals that are invisible to the naked eye:
  - Recognizance.
  - Chemical plumes.
- A number of other civilian (e.g., agriculture) and research (e.g., monitoring of environmental processes) applications.
- "Supervised" vs. "unsupervised" problems $\Rightarrow$ Is a library of signatures available?

Hyperspectral sensors

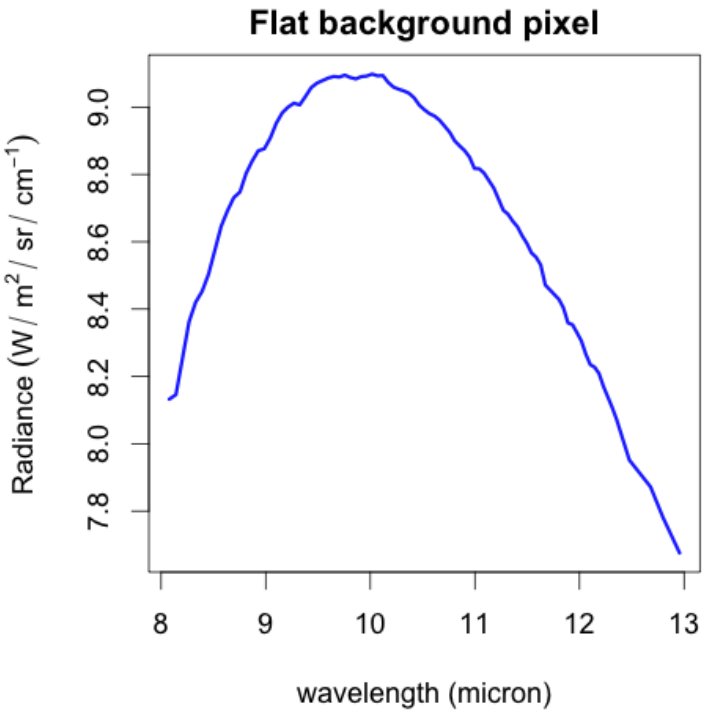


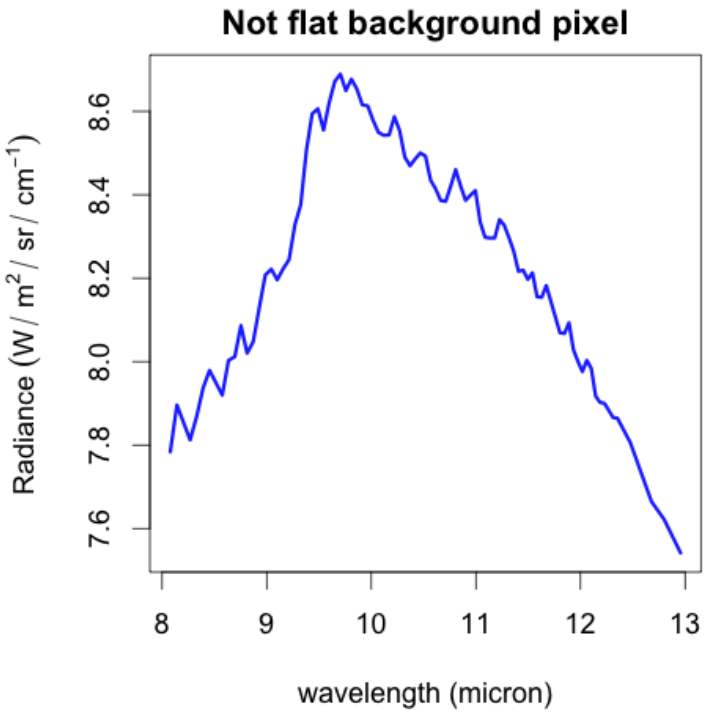
An example of a hyperspectral cube



- Radiance $y_{i,j}$ for $i = 1, \ldots, 89,600 pixels$ and $j = 1, \ldots, 85$ wavelengths in LWIR (8-13 microns).
- Topside image, rural area with a few man-made structures, no chemical plume.

Spectrum of a pixel





Hyperspectral models for plume detection and classification

- Three possible sources of radiance: Background, Atmosphere and Plume.
- Radiance at each pixel depends for a spectral frequency λ:
  - If plume is not present:

$$L_{off}(\lambda) = \underbrace{L_a(\lambda)}_{\text{Rad atmos}} + \underbrace{\tau_a(\lambda)}_{\text{Trans atmos}} \underbrace{L_b(\lambda)}_{\text{Rad backgr}}$$

  - If plume is present:

$$L_{on}(\lambda) = \underbrace{L_a(\lambda)}_{\text{Rad atmos}} + \underbrace{\tau_a(\lambda)}_{\text{Trans atmos}} \underbrace{L_p(\lambda)}_{\text{Rad plume}} + \underbrace{\tau_a(\lambda)}_{\text{Transm. atmos.}} \underbrace{\tau_p(\lambda)}_{\text{Trans plume}} \underbrace{L_b(\lambda)}_{\text{Rad. backgr.}}$$

Connecting physics and data: Assumptions

1. In the LWIR bands the atmosphere is transparent ($\tau_a(\lambda) \approx 1$).
2. Kirchhoffs law:

$$L_p(\lambda) = \{1 - \tau_p(\lambda)\} \underbrace{B(\lambda, T_p)}_{\text{Plank function}}$$

   where T_p is the temperature of the plume.

3. Beer's law

$$\tau_p(\lambda) = \exp\left\{-\sum_{m=1}^{M} \gamma_m x_m(\lambda)\right\} \overset{\text{Taylor}}{\approx} 1 - \sum_{m=1}^{M} \gamma_m x_m(\lambda)$$

   where $x_m(\lambda)$ and γ_m are the signature and the contribution of the m-th chemical.

4. If background is flat, $L_b(\lambda) \approx B(\lambda, T_b)$.
5. If $T_p - T_b$ is small then $B(\lambda, T_p) - B(\lambda, T_b) \approx C_B(T_p - T_b)$.

Connecting physics and data: A statistical model

- Under the five assumptions outlined above, the equations for the radiance $y_{i,j}$ in pixel i and spectral band j becomes a linear function of the (known) spectral signatures of chemicals in the cloud measured at the same locations, $x_{m,j}$,

$$y_{i,j} = \sum_{m=1}^{M} \beta_{i,m} x_{m,j} + \epsilon_{i,j}$$

where the $\epsilon_{m,j}$s represents an non-zero mean, non-independent errors that incorporate:

  - Measurement error in the sensor (assumed to be zero mean and often independent across frequencies).
  - Error introduced due to the approximation (maybe zero mean, and independent across frequencies.
  - The structure of the background (non-zero mean and correlated across frequencies).

Connecting physics and data: A statistical model

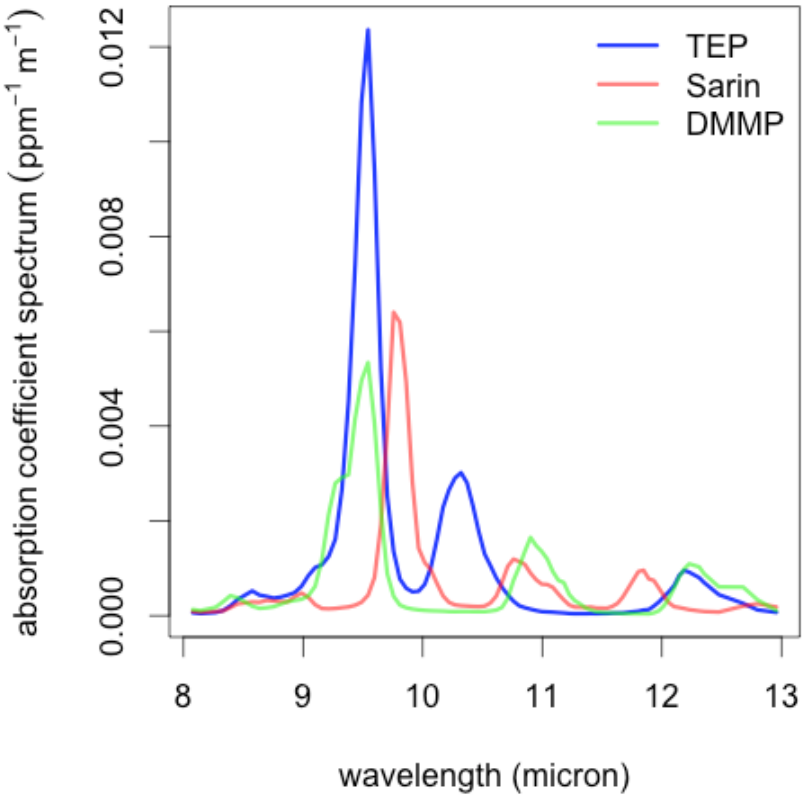
- Two final simplifications that are often made out of convenience:
  - Errors $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \cdots, \epsilon_{i,J})^T$ are normally distributed,

$$\boldsymbol{\epsilon}_i \sim \mathrm{N}(\boldsymbol{\mu}, \sigma_i^2 \boldsymbol{\Omega}).$$

  - Although in principle $\beta_{i,m} \geq 0$ for all m and $\sum_{m=1}^{M} \beta_{i,m} = 1$, when detection/identification is the focus the constraints are often ignored.
- Under these two simplifications we are in the context of a standard linear model:
  - Detection and identification become variable selection problems.
  - Quantification becomes an estimation problem.

Chemical signatures

- PNNL chemical spectral database (Sharpe et al., 2004) $\Rightarrow$ Over 400 compounds.
- High resolution signatures at predetermined frequencies $\Rightarrow$ Usually there is a need to interpolate and downsample.



1 Hyperspectral Imaging

2 Identification

3 Bayesian spatial selection

4 Semiparametric approaches

5 Discussion

Identification and variable selection

- Techniques for variable selection have been widely applied to detection and identification problems in supervised settings:
  - F tests/forward/backward t-tests.
  - Penalization (in particular L1/Lasso penalization).
  - Bayesian posterior probabilities.

- Sometimes it will be useful to write the model in matrix form:

$$\underbrace{\mathbf{Y}}_{I\times J} = \underbrace{\mathbf{B}}_{I\times M}\underbrace{\mathbf{X}}_{M\times J} + \underbrace{\mathbf{E}}_{I\times J}, \qquad \mathbf{E} \sim \text{MN}\left(\mathbf{1}_I \otimes \boldsymbol{\mu}^J, \boldsymbol{\Sigma}, \boldsymbol{\Omega}\right),$$

where $\boldsymbol{\Sigma} = \text{diag}\{\sigma_1^2, \ldots, \sigma_I^2\}$.

To standardize or not to standardize, that is the question!

- It is very common (particularly in frequentist approaches) to work with a standardized model

$$\mathbf{Y}^* = \mathbf{B}\mathbf{X}^* + \mathbf{E}^*, \qquad \mathbf{E}^* \sim \text{MN}\left(\mathbf{0}, \mathbf{\Sigma}, \mathbf{I}_J\right).$$

  where $\mathbf{Y}^* = \left\{\mathbf{Y} - \mathbf{1}_I \otimes \hat{\boldsymbol{\mu}}^J\right\} \hat{\mathbf{\Omega}}^{-1/2}$, $\mathbf{X}^* = \mathbf{X}\hat{\mathbf{\Omega}}^{-1/2}$ and $\mathbf{E}^* = \mathbf{E}\hat{\mathbf{\Omega}}^{-1/2}$.
- How to obtain the estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{\Omega}}$? Assuming the plume covers a small number of pixels:
  - $\hat{\boldsymbol{\mu}}$ is the average raw image.
  - $\hat{\mathbf{\Omega}}$ is regularized version of the covariance matrix of the image (treating pixels as independent), e.g., dominant mode rejection.
- Using the standardized model can be also helpful (for computational reasons) in a Bayesian context.

Penalization: Lasso

- L1 penalization is a popular tool for variable selection $\Rightarrow$ Fast, which is useful in large datasets such as ours!
- Although I focus on the Lasso, but other penalties that are non-differentiable at the origin could be used instead (elastic net, SCAD, etc).
- Estimates obtained as

$$\tilde{\mathbf{B}} = \arg\min_{\mathbf{B}} \text{tr}\left\{(\mathbf{Y}^* - \mathbf{B}\mathbf{X}^*)^T(\mathbf{Y}^* - \mathbf{B}\mathbf{X}^*)\right\} + |\mathbf{\Psi}\mathbf{B}|$$

  where $\mathbf{\Psi} = \text{diag}\{\psi_1, \ldots, \psi_i\}$ and $|\mathbf{H}| = \sum_{i,j} |h_{i,j}|$

- Because the penalty is non-differentiable, values of $\hat{\beta}_{i,m}$ can be identically zero $\Rightarrow$ Nice properties such as sparse-consistency.
- MAP under independent, double exponential priors.

Penalization: Lasso

- Computation in this case is relatively simple because this is a separable, convex optimization problem $\Rightarrow$ We can do computations for pixels in parallel, and a simple alternating conditional modes (ACM) algorithm will converge to the global maximum.
- ACM are easy to carry out in close form using the softthresholding operator:

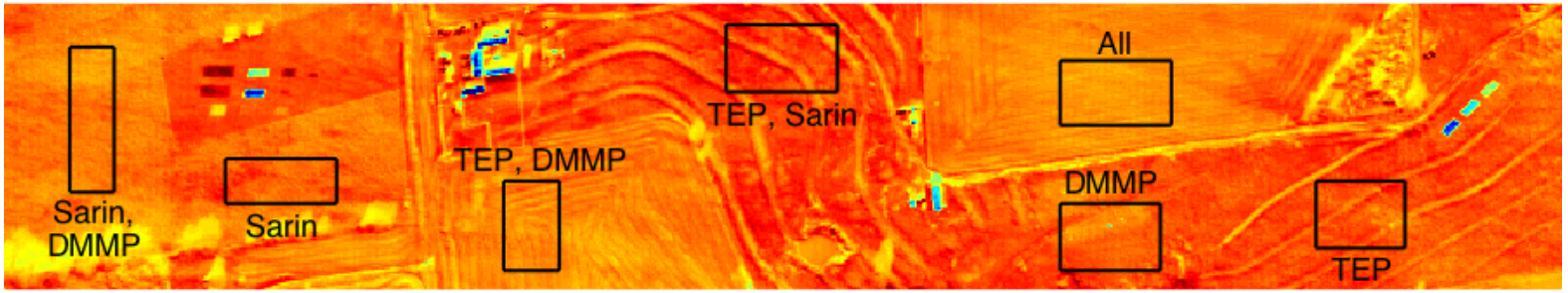
$$S(\beta, \psi) = \begin{cases} \text{sign}(\beta)\{|\beta| - \psi\} & |\beta| > \psi \\ 0 & |\beta| \leq \psi \end{cases}$$

- The value of the penalties $\psi_1, \ldots, \psi_I$.
- Interestingly, this approach with indpendent penalties works better than setting a common penalty $\psi_i = \psi$ for all i.

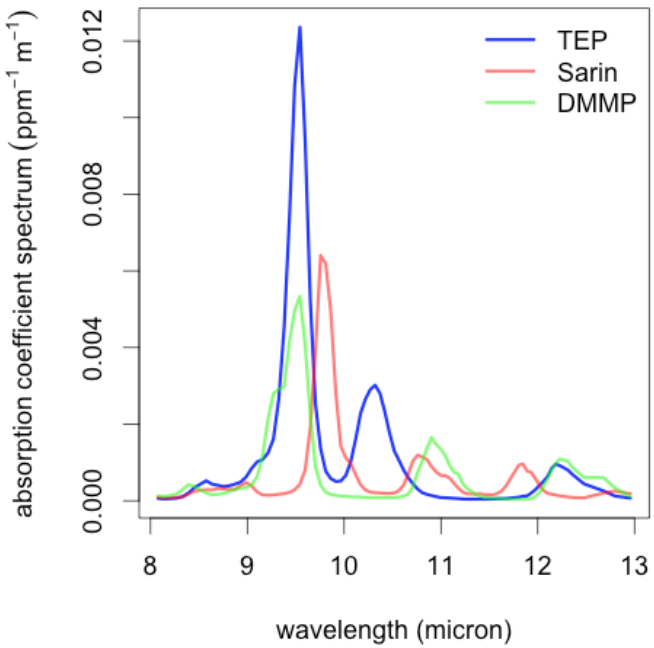
An illustration

- Hyperspectral data from real plume releases is hard to come by (even if you work with the government).
- We will work in part with semi-synthetic data.
  - We start with a real hyperspectral cube.
  - We use the physical, non-linear model to embed a plume in the a subset of pixels of the cube.
  - Physical parameters associated with the embedding can be varied to obtain a number of scenarios, including linear and non-linear embeddings.

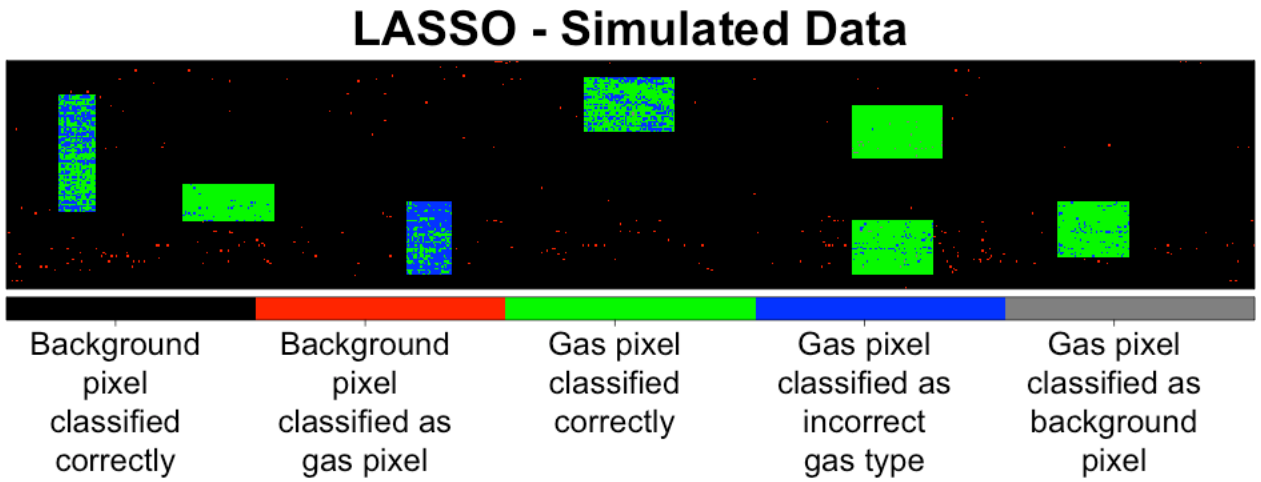
An illustration



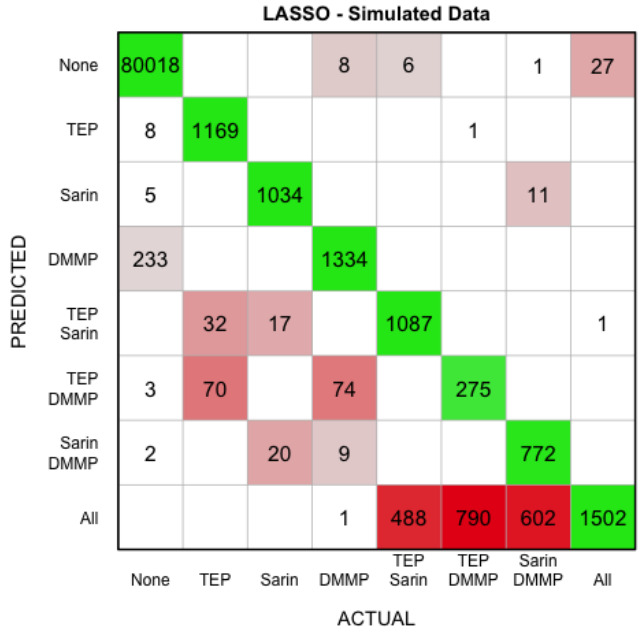
- 89,600 pixels over 85 wavelengths in LWIR.
- 7 plumes, all possible combinations of 3 chemicals: TEP (pesticides/fire retardants), Sarin (chem weapon), DMMP (fire retardants/chem weapons).
- Linear regime.



An illustration



Lasso has a particularly hard time when two chemicals are present (it tends to overestimate), probably because of the high autocorrelation across some chemicals.



LASSO's shortcomings

- Obviously, the linear model is a simple approximation, we could try to come up with a more flexible model. But linearity is extremely convenient, is there anything we can do before going back to drawing board?
- LASSO is, at its core, an estimation and not a variable selection procedure.
- LASSO treats coefficients in different pixels as independent, but we know that in cases where there is strong colinearity (as is the case here!) independent priors are suboptiomal.
- The approach also treats nearby pixels as independent, but we would expect some sort of spatial structure: plumes, by definition are made of adjacent pixels.

Accounting for collinearity: Bayesian model selection

- Let's make explicit which coefficients are non-zero using an $I \times M$ matrix of indicators $\mathbf{\Gamma}$ such that $\gamma_{i,m} = 0$ if $\beta_{i,m} = 0$, and $\gamma_{i,m} = 1$ otherwise:

$$\mathbf{Y}^* = \mathbf{B}_{\mathbf{\Gamma}}\mathbf{X}^* + \mathbf{E}^*, \qquad \mathbf{E}^* \sim \text{MN}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{I})$$

- Bayesian inference for this model involves placing a prior the unknown parameters of the model:

$$p(\mathbf{B}_{\mathbf{\Gamma}} \mid \mathbf{\Sigma}, \mathbf{\Gamma}) p(\mathbf{\Sigma}) p(\mathbf{\Gamma})$$

- In identification we are interested in posterior inferences for the matrix $\mathbf{\Gamma}$. In particular, if $\text{P}(\gamma_{i,m} = 1 \mid \mathbf{Y}) > t$ for some threshold t (e.g., $t = 0.5$) then we claim that chemical m has been identified in pixel i.

Bayesian model selection: Prior on $\boldsymbol{\Gamma}$

- One option is to use a uniform prior on all possible 2^M models $\Rightarrow$ heavily favors mid-sized models (because there are many more of them!)
- A better alternative if a special case of the Beta-Binomial prior:

$$P(\boldsymbol{\Gamma}) = \prod_{i=1}^{I} \underbrace{\frac{1}{M+1}}_{\text{Uniform over model size}} \underbrace{\frac{1}{\binom{M}{\sum_{m=1}^{M} \gamma_{i,m}}}}_{\text{Uniform of models of size } \sum_{m=1}^{M} \gamma_{i,m}} = \prod_{i=1}^{I} \int_0^1 \prod_{m=1}^{M} \theta_i^{\sum_{m=1}^{M} \gamma_{i,m}} (1-\theta_i)^{M - \sum_{m=1}^{M} \gamma_{i,m}} d\theta_i$$

(pixels independent from each other, but chemicals are dependent within a pixel)

- There are a few variants of this idea, but this one worked consistently better in our simulations.

Bayesian model selection: Prior on $\mathbf{B} \mid \boldsymbol{\Gamma}, \boldsymbol{\Sigma}$

- We treat $\beta_1^T, \ldots, \beta_I^T$, the rows of $\mathbf{B}$ (which describe the levels of all chemicals for a given pixel) as conditionally independent:

$$\underbrace{\prod_{i=1}^{I} \mathrm{N}\left(\beta_{i,\gamma_i} \mid \mathbf{0}, c_i \sigma_i^2 \left\{\mathbf{X}_{\gamma_j}^T \mathbf{X}_{\gamma_j}\right\}^{-1}\right)}_{\text{Joint normal prior for included coefficients}} \quad \underbrace{\prod_{m:\gamma_{i,m}=0} \delta_0(\beta_{i,m})}_{\text{Independent point masses at zero for coefficients not included}}$$

  where c_i is given a hyperprior, $c_i \sim \mathrm{IG}(0.5, 0.5J)$.
- These are independent mixtures g-priors: working with a scale mixture of Gaussians has a number of theoretical advantages (consistency, good small-sample behavior, etc).

Bayesian model selection: Prior on $\boldsymbol{\Sigma}$

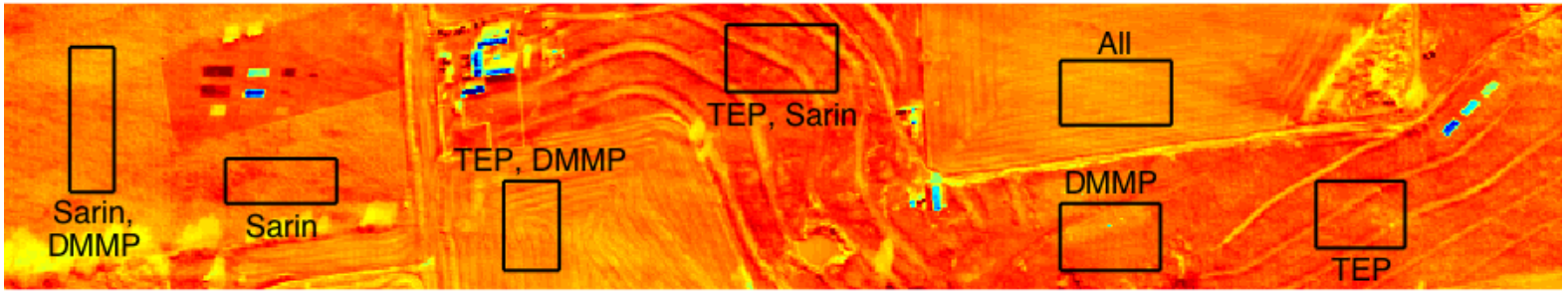
- Again, we treat the variances on each pixel as independent and use non-informative priors,

$$p(\sigma_j^2) \propto \frac{1}{\sigma_j^2}$$

- Note that the structure of these priors allows for a closed-form expression for $p(\boldsymbol{\Gamma}, \mathbf{c} \mid \mathbf{Y})$.

- This suggests an MCMC algorithm that alternate between
  1. Sampling $\{\gamma_i \mid \cdots\}$ (or $\gamma_{i,m} \mid \cdots$ if M is very large).
  2. Sampling $\{c_j \mid \cdots\}$.
- Because pixels are independent from each other we can naively parallelized the algorithm

Illustration revisited



- 89,600 pixels over 85 wavelengths in LWIR.
- 7 plumes, all possible combinations of 3 chemicals: TEP (pesticides/fire retardants), Sarin (chem weapon), DMMP (fire retardants/chem weapons).
- Linear regime.

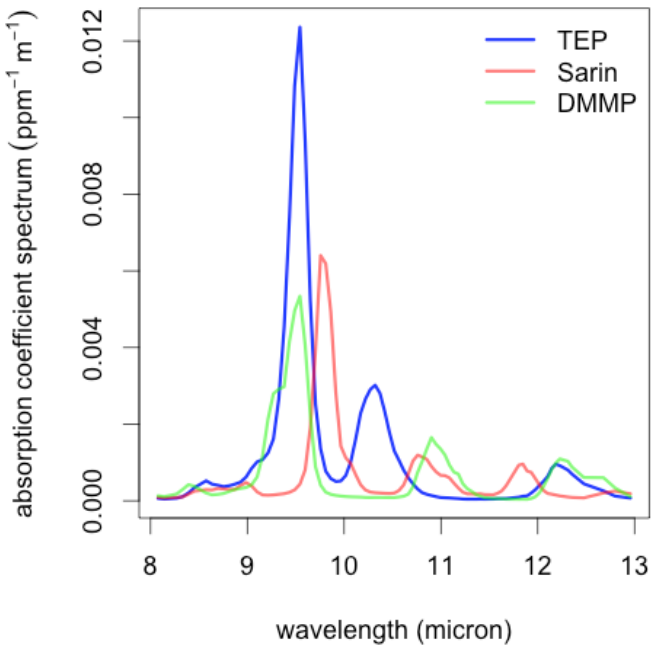
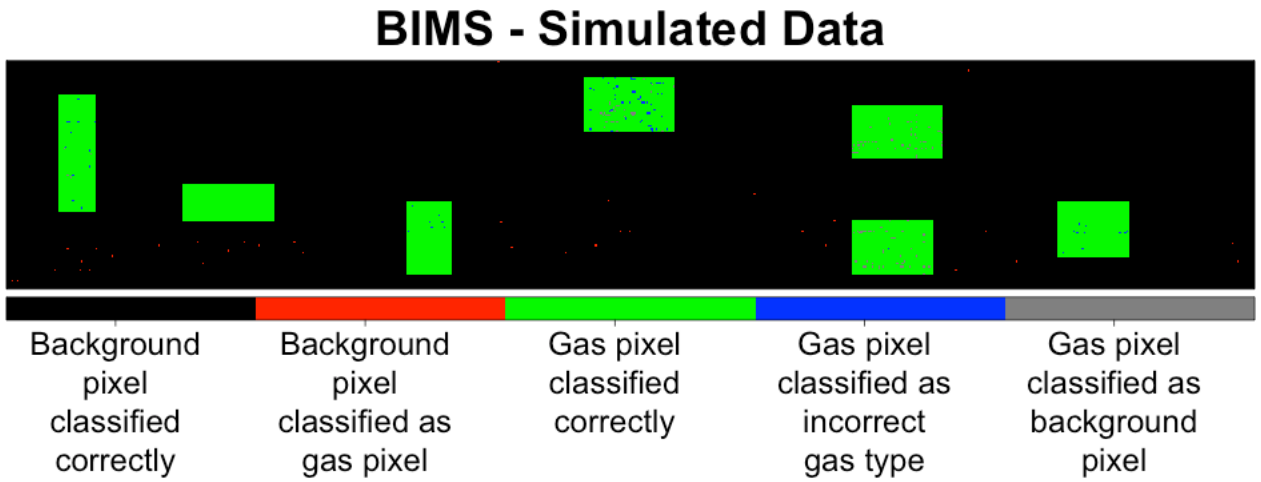
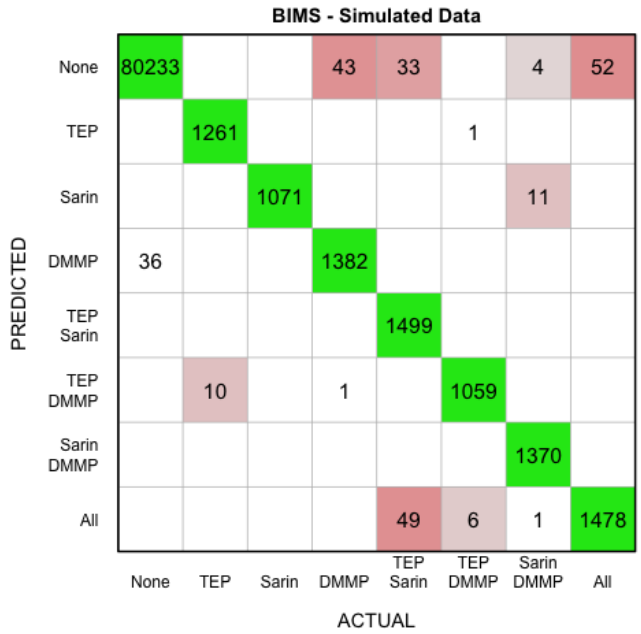


Illustration revisited



Performance is much improved over LASSO, although the rate of false negatives in terms of detection is a bit higher now (look at the first row of the confusion matrix).



Accounting for spatial structure: Fused LASSO

- The "fused" LASSO can incorporate spatial information by enforcing similarity among adjacent coefficents
  - Requires that you define a neighborhood structure $\Rightarrow$ We work with first-order neighborhoods: each pixel has at most four neighbors.
  - Define $\mathbf{L}$ as a $N \times I$: one row for each pair of neighbors, all entries in the row are zero except for a 1 and -1 associated with the coefficients of the two neighboring pixels.
- Fused Lasso is given by

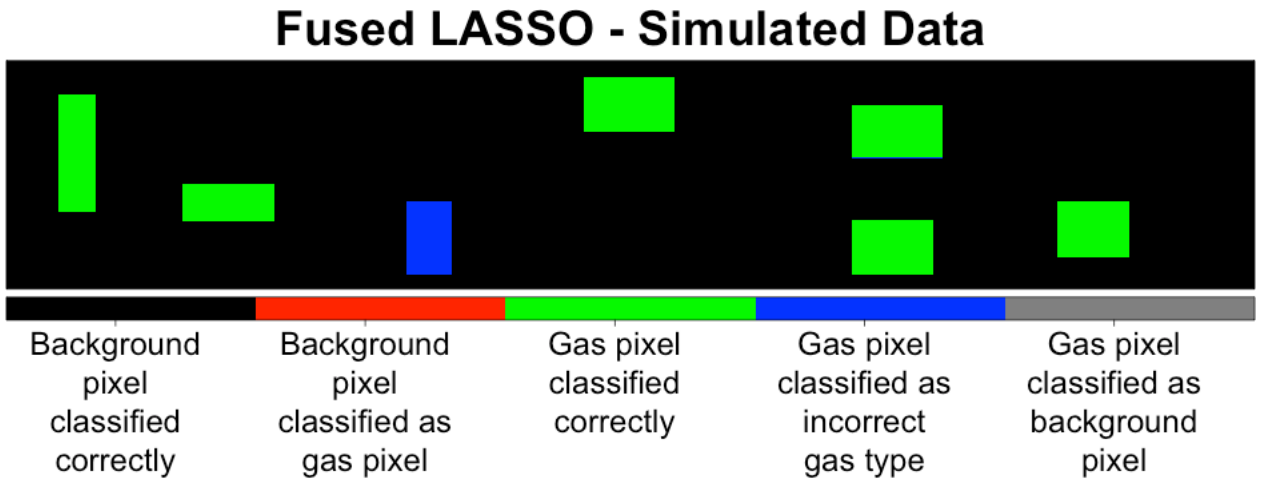
$$\tilde{\mathbf{B}} = \arg\min_{\mathbf{B}} \operatorname{tr}\left\{(\mathbf{Y}^* - \mathbf{B}\mathbf{X}^*)^T(\mathbf{Y}^* - \mathbf{B}\mathbf{X}^*)\right\} + \psi|\mathbf{B}| + \eta|\mathbf{L}\mathbf{B}|$$

- Special case of an (improper) Markov Random Field (MRF) prior!

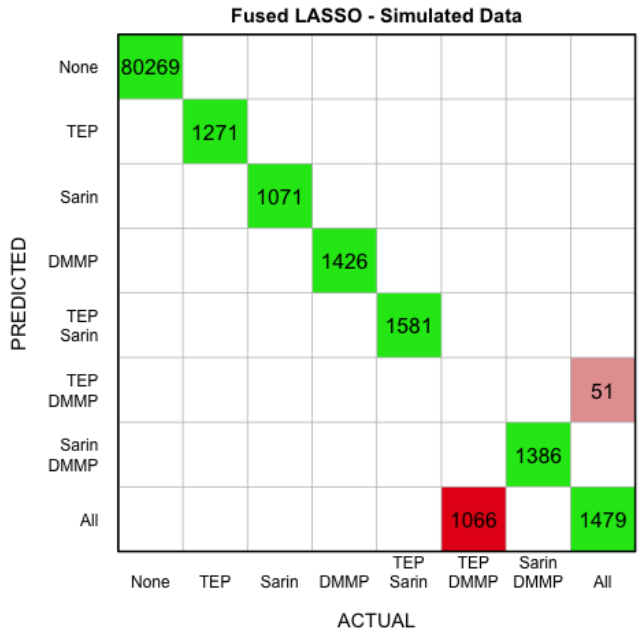
Accounting for spatial structure: Fused LASSO

- Pixels are not independent any more!
- Because the penalty is a function of the absolute value of the difference between coefficients for a given gas in neighboring pixels, the estimate $\tilde{\mathbf{B}}$ can contain identical values for neighboring coefficients.
- Unlike before, we have a common penalty ψ for all pixels.
- Computation is a bit trickier because the objective function is not convex! $\Rightarrow$ We use a split-Bregman algorithm.

Illustration revisited



Performance is much improved over LASSO, although the rate of false negatives in terms of detection is a bit higher now (look at the first row of the confusion matrix).



1 Hyperspectral Imaging

2 Identification

3 Bayesian spatial selection

4 Semiparametric approaches

5 Discussion

Bayesian spatial selection

- Moving into a Bayesian approach that accounts for correlations across gases in the prior or incorporating the assumption of cloud continuity both separately improved the accuracy of the algorithm. Can we do both simultaneously?
- We start with our Bayesian model and look for ways to create priors that borrow information across nearby pixels. There are at least a couple of ways in which this can be done:
  - Building spatial correlations between $\boldsymbol{\beta}_{i,\boldsymbol{\gamma}_i} \mid \sigma_i^2, \boldsymbol{\gamma}_i$ and $\boldsymbol{\beta}_{i',\boldsymbol{\gamma}_{i'}} \mid \sigma_{i'}^2, \boldsymbol{\gamma}_{i'} \Rightarrow$ This is quite hard (although we will discuss some ideas if we have time ..).
  - Building spatial correlations between the prior inclusion probabilities $\theta_{i,m} = \mathrm{P}(\beta_{i,m} \neq 0)$ and $\theta_{i',m} = \mathrm{P}(\beta_{i',m} \neq 0) \Rightarrow$ Much easier. Furthermore, it leads to "soft" rather than hard constraints on the coefficients.

Spatial probit models

- Traditional probit model: $\gamma_{i,m} \mid \alpha_m \sim \text{Ber}\left(\Phi\left(\alpha_m\right)\right)$ and $\alpha_m \sim \text{N}(\bar{\alpha}_m, \tau_m^2)$. $\bar{\alpha}_m = 0$, $\tau_m^2 = 1 \Rightarrow \Phi\left(\alpha_i\right) \sim \text{Uni}[0,1]$ (a version of our original model ...)

- Model can be rewritten using latent variables (particularly useful for computational purposes):

$$\gamma_{i,m} = \begin{cases} 1 & z_{i,m} \geq 0 \\ 0 & z_{i,m} < 0 \end{cases} \quad z_{i,m} \mid \alpha_m \sim \text{N}\left(\alpha_m, 1\right) \quad \alpha_m \sim \text{N}\left(\bar{\alpha}_m, \tau_m^2\right).$$

Spatial probit models

- This model can be extended to spatial settings by making the latent variables $\{z_{i,m}\}$ dependent $\Rightarrow$ "Clipped Gaussian Random Markov Field".

$$\begin{pmatrix} z_{1,m} \\ \vdots \\ z_{I,m} \end{pmatrix} \mid \alpha_m, \tau_m^2, \rho_m \sim \mathrm{N}\left(\alpha_m \mathbf{1}_I, \tau_m^2 \left\{\mathbf{D_W} - \rho_m \mathbf{W}\right\}^{-1}\right).$$

  where $\mathbf{W} = \mathbf{L}^T\mathbf{L}$ is the neighborhood matrix and $\mathbf{D_W}$ is a diagonal matrix whose entries are the sums of the rows of $\mathbf{W}$.

- Can also be rewritten as

$$z_{i,m} \mid \{z_{i',m}\}_{i'=1,i'\neq i}^{I}, \alpha_m, \tau_m^2, \rho_m \sim \mathrm{N}\left(\alpha_m + \rho_m \sum_{i'=1,i'\neq i}^{I} \frac{w_{i,i'}}{w_{i,+}} \left\{z_{i',m} - \alpha_m\right\}, \tau_m^2\right).$$

Bayesian spatial selection

- Using a proper MRF (i.e., $\rho_m < 1$) is quite important. Otherwise posterior is improper!!
- The model is simpler than the more standard spatial random-effects probit model,

$$z_{i,m} \sim \mathrm{N}(\upsilon_{i,m}, 1) \qquad \begin{pmatrix} \upsilon_{1,m} \\ \vdots \\ \upsilon_{1,m} \end{pmatrix} \sim \mathrm{CAR}(\alpha_m, \rho_m, \tau_m^2)$$

- Note that $\rho_m = 0$ leads to essentially the same independent-pixels model we discussed in the previous slide.
- Identifiability not an issue because we we are only interested in the distribution of $\mathbf{\Gamma}$ and not any of the hyperparameters.

Hyperspectral sensors

- Computation using MCMC algorithms.
- Because we use the same prior on $\mathbf{B}_{\mathbf{\Gamma}}$, we can still integrate both $\mathbf{B}_{\mathbf{\Gamma}}$ and $\mathbf{\Sigma}$.
- Because of the first-order neighborhood structure, it can be naively parallelized using a checkerboard scheme.
- Pre-standardization speeds up computation!

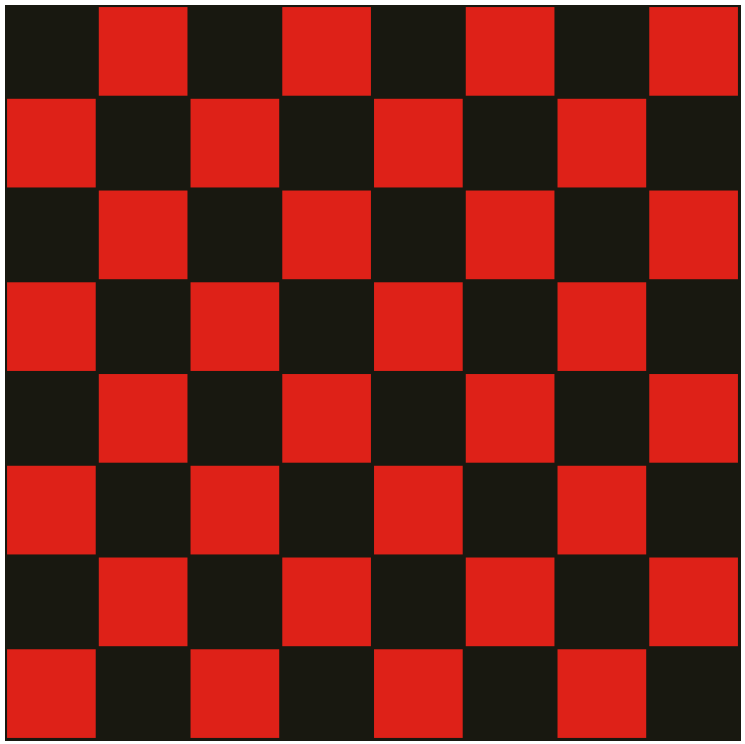
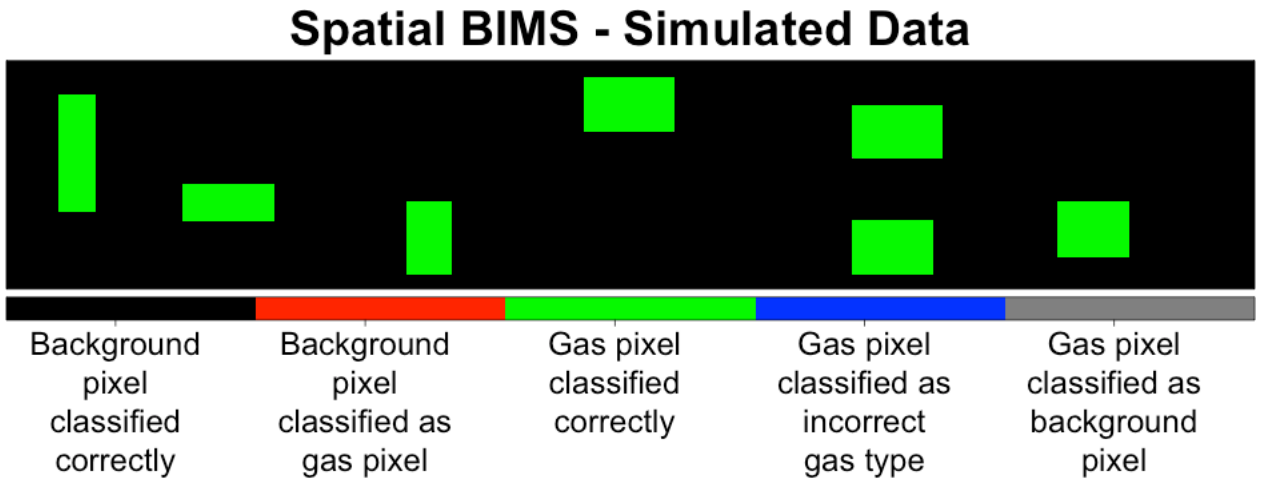
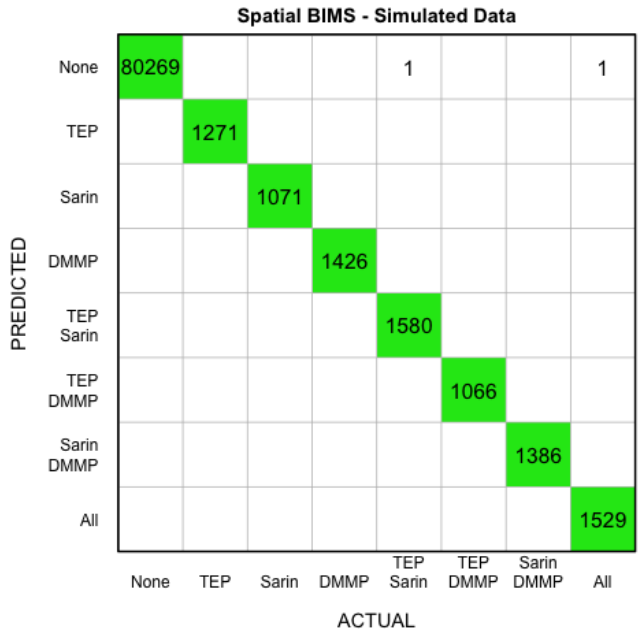


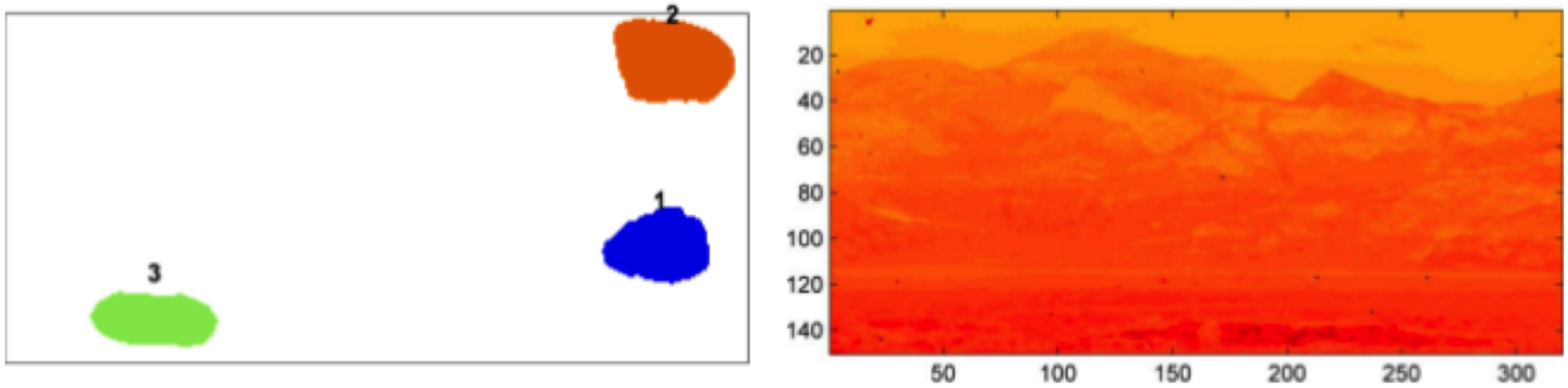
Illustration revisited



Surprisingly, the results are almost perfect in this case ....

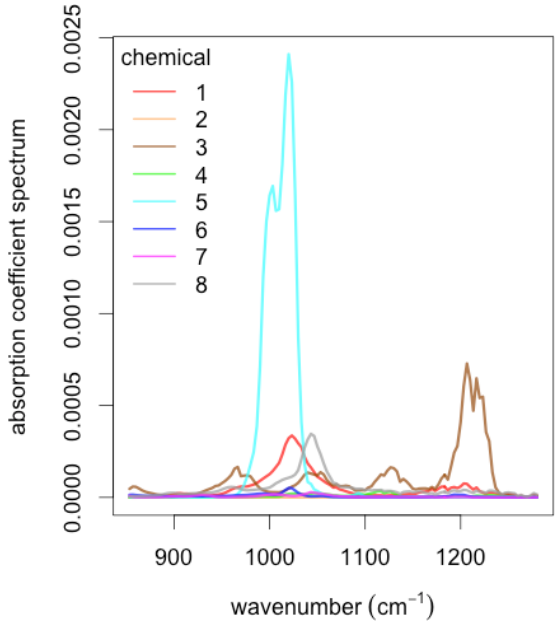


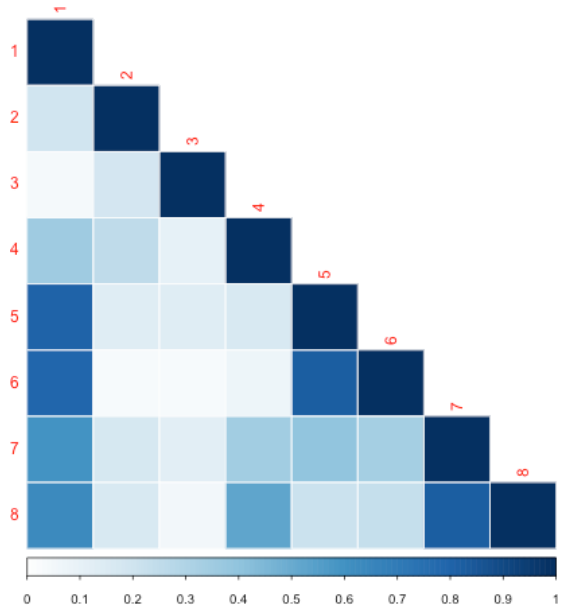
The MIT Lincoln Lab Challenge Data



- Various semi-synthetic datasets generated and released by MIT Lincoln Lab as part of the NSF/DTRA Algorithms for Threat Detection Program.
- Open challenge to all grantees of the program to test identification algorithms (phase II of the challenge).
- Varying degrees of difficulty, we focus here on the "hardest" one in the set.

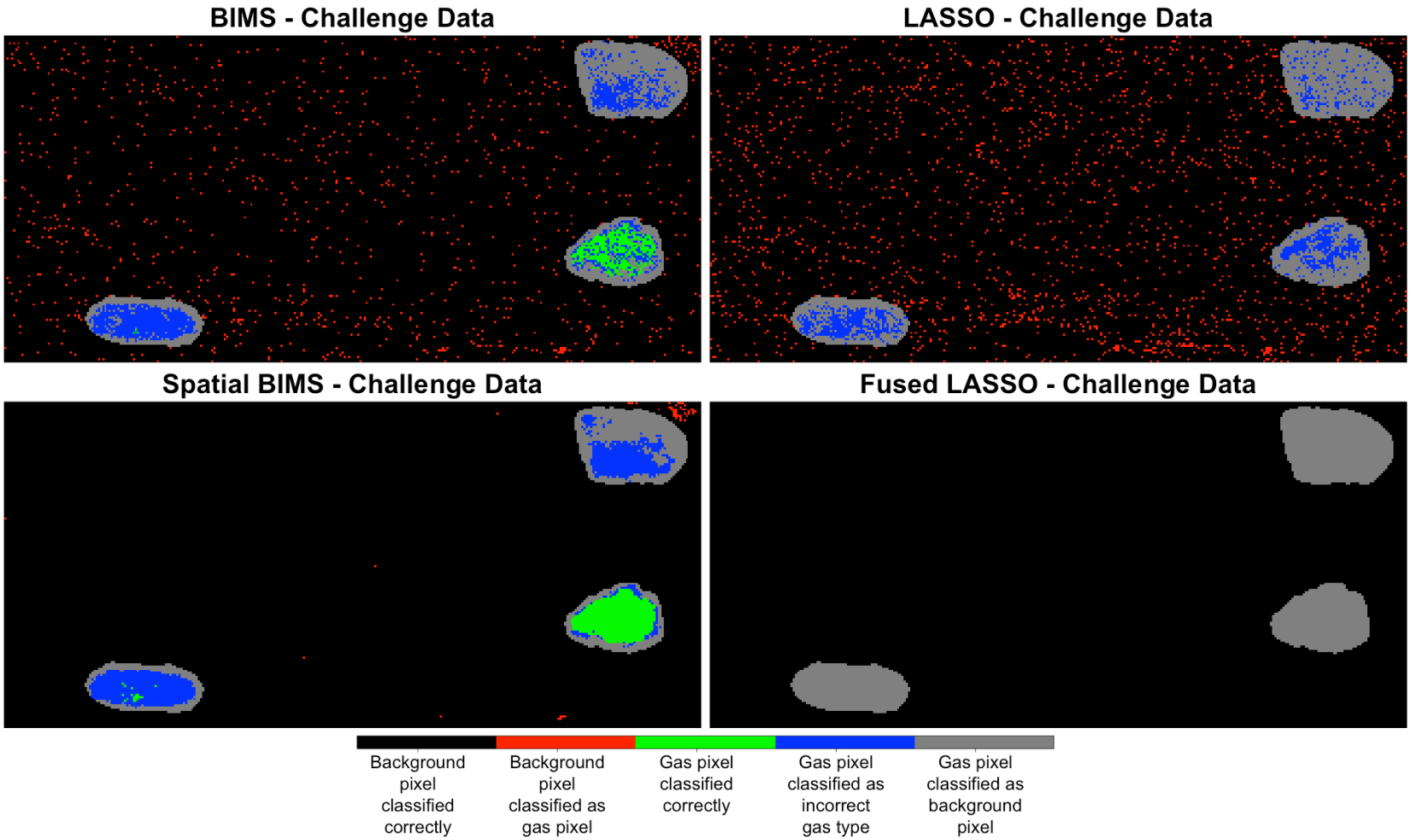
The MIT Lincoln Lab Challenge Data





- Side-facing data cube.
- Library of candidate chemicals includes 8 chemicals, some with highly correlated signatures.
- Three plumes, each containing a single chemical.

Illustration revisited



1 Hyperspectral Imaging

2 Identification

3 Bayesian spatial selection

4 Semiparametric approaches

5 Discussion

Going beyond linear regression models

- The previous approaches are motivated by a number of linearizations (e.g., of Beer's law).
- To relax this assumption, we borrow ideas from the uncertainty quantification literature and model the response function nonparametrically using a Gaussian process prior centered around the linear model:

$$y^*_{i,j} = f_i(\mathbf{x}^*_j) + \epsilon^*_{i,j} \qquad \epsilon^*_{i,j} \sim \mathsf{N}(0, \sigma^2_i)$$

where $f_i \sim \mathsf{GP}(\mathbf{x}_j^T \beta_i, \xi_i(\mathbf{x}, \mathbf{x}'))$.

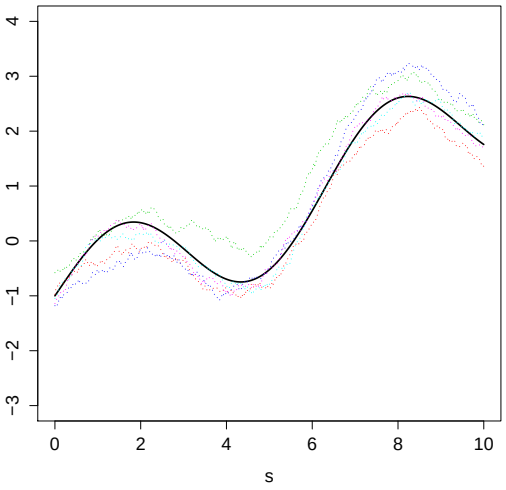
Gaussian process priors

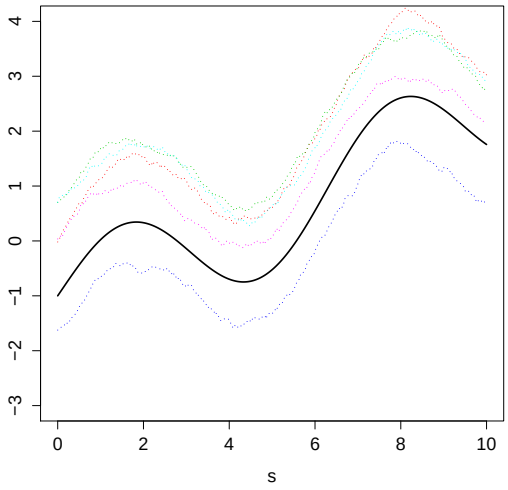
- A function $f : \mathcal{X} \to \mathbb{R}$ has a Gaussian process prior with mean function $g(\mathbf{x})$ and covariance function $\xi(\mathbf{x}, \mathbf{x}')$, denoted $f \sim \text{GP}(g(\mathbf{x}), \xi(\mathbf{x}, \mathbf{x}'))$ if for any n and set of values $\mathbf{x}_1, \ldots, \mathbf{x}_n$ we have

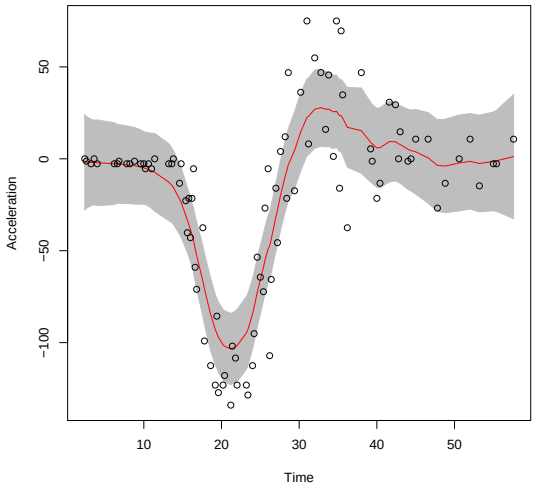
$$\begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_n) \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} g(\mathbf{x}_1) \\ \vdots \\ g(\mathbf{x}_n) \end{pmatrix}, \begin{pmatrix} \xi(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \xi(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \xi(\mathbf{x}_n, \mathbf{x}_1) & \cdots & \xi(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix} \right)$$

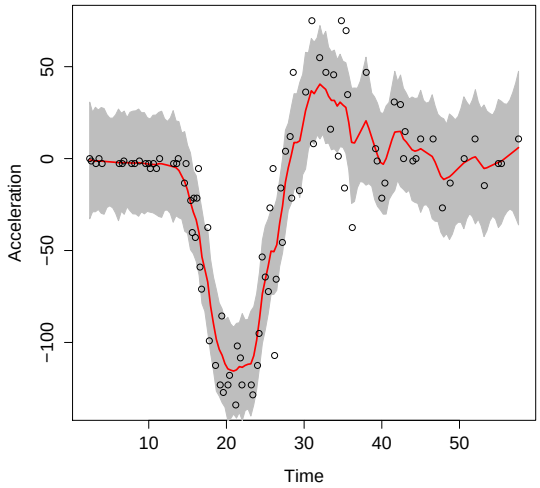
- Gaussian process priors combine flexibility and computational simplicity (if not scalability ...).

Gaussian process priors









Covariance functions

- We work with a product exponential covariance functions

$$\xi_i(\mathbf{x}, \mathbf{x}') = \kappa_i \sigma_i^2 \exp\left\{ - \sum_{m=1}^{M} \vartheta_{i,m} |x_m - x'_m| \right\}.$$

- Note that if $\kappa_i = 0$ then we are back in the linear model $\Rightarrow$ Prior favors relatively small values of κ_i, so that the posterior is close to the linear model.
- To perform model selection we need a prior $p(\boldsymbol{\beta}_i, \boldsymbol{\vartheta}_i \mid \boldsymbol{\gamma}_i)$ (both the mean and covariance depend of $\mathbf{x}$). Use independent priors, $p(\boldsymbol{\beta}_i, \boldsymbol{\vartheta}_i \mid \boldsymbol{\gamma}_i) = p(\boldsymbol{\beta}_i \mid \boldsymbol{\gamma}_i) p(\boldsymbol{\vartheta}_i \mid \boldsymbol{\gamma}_i)$.
- For these two priors we set:
  - For $p(\boldsymbol{\beta}_i \mid \boldsymbol{\gamma}_i)$, we use mixtures of g-priors as before.
  - $p(\boldsymbol{\vartheta}_i \mid \boldsymbol{\gamma}_i) = \prod_{m:\gamma_{i,m}=1} \mathsf{G}(\vartheta_{i,m} \mid a_\vartheta, b_\vartheta) \prod_{m:\gamma_{i,m}=0} \delta_0(\vartheta_{i,m})$.

Computation

- Computation using MCMC algorithms $\Rightarrow$ Quite a few Metropolis-Hastings steps that require some tuning.
- The unknown functions $f_1, \ldots, f_l$ can be integrated out of the model, along with $\mathbf{B}$ and the variances $\sigma_1^2, \ldots, \sigma_l^2$.
- Algorithms are much slower because they require repeated inversions of moderately-sized covariance matrices.
- Again, pre-standardization is quite helpful in somewhat speeding things up.

Illustration revisited

Coming soon ...

Discussion

- Flexible and powerful methods for detection and identification of gas plumes in hyperspectral images.
- There are a number of small tweaks that can be incorporated and would likely improve performance.
  - Spatial correlation on the matrix $\mathbf{B}$ (roughly speaking, correlation among the rows of $\mathbf{B}$, in addition to correlation among the columns).
  - More general mixtures of g-priors (e.g., with a different mixing parameter for each coefficient.
  - Correlation among the columns of $\mathbf{Z}$.
- Applications in a number of other disciplines.