Bayesian spatial modeling (contd.)

Abhi Datta

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland

abhidatta.com @dattascience

Review of last lecture

- Using STAN package to run the MCMC for marginalized and latent spatial linear models
- MCMC diagnostics
- Spatial predictions from Bayesian output
- Model comparison using Bayesian output (DIC and WAIC)
- MCMC-free and fully Bayesian spatial analysis using spBayes package
- Hierarchical Bayesian spatial analysis beyond the spatial linear models
    - Example 1: Second-stage GP model for covariate imputation

Example 2: Discrete response

- In many applications, the spatial response of interest is binary or count-valued
- (Spatial LM for continuous data:) $E(y(s)) = x(s)'\beta + w(s)$
- (Spatial GLM for discrete data:) $g(E(y(s))) = x(s)'\beta + w(s)$ where g is a suitable link function (logit-link for binary data, log-link for count data)
- Other components stay the same, i.e., GP prior for $w(s)$, priors for β and covariance parameters

Spatial GLM for non-Gaussian response

- (Poisson Spatial GLM)

$$
\begin{aligned}
y(s) &\overset{\text{iid}}{\sim} \text{Poisson}\left(\exp(x(s)'\beta + w(s))\right) \\
w = (w(s_1), \ldots, w(s_n))' &\sim N(0, C(\theta)) \\
(\beta, \theta) &\sim p(\beta, \theta)
\end{aligned}
$$

- (Bernoulli Spatial GLM)

$$
\begin{aligned}
y(s) &\overset{\text{iid}}{\sim} \text{Bernoulli}\left(\frac{\exp(x(s)'\beta + w(s))}{1 + \exp(x(s)'\beta + w(s))}\right) \\
w = (w(s_1), \ldots, w(s_n))' &\sim N(0, C(\theta)) \\
(\beta, \theta) &\sim p(\beta, \theta)
\end{aligned}
$$

Eastern hemlock presence/absence data

- In many applications, the spatial response of interest is binary or count-valued

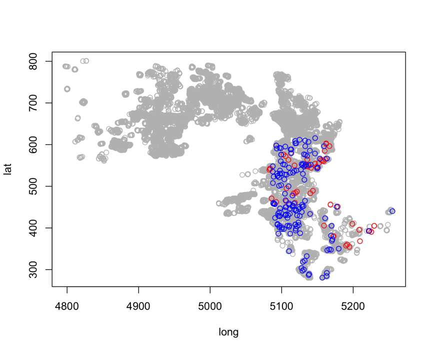


Figure: Figure: (Grey) All locations sampled for presence/absence of Eastern Hemlock species in Michigan. We will analyze a subset of locations with presence (red) and absence (blue).

Model

- Covariates:
  - Minimum winter temperature (MIN)
  - Maximum summer temperature (MAX)
  - Total precipitation in the coldest quarter of the year (WIP)
  - Total precipitation in the warmest quarter of the year (SUP)
  - Annual actual evapotranspiration
  - Annual climatic water deficit (DEF) (AET)
- Model[1]:

$$
\begin{aligned}
E(y(s)) =& \text{logit}\Big(\alpha + \beta_1 \text{MIN(s)} + \beta_2 \text{MAX(s)} + \\
& \beta_3 \text{WIP(s)} + \beta_4 \text{SUP(s)} + \beta_5 \text{AET(s)} + \beta_6 \text{DEF(s)} + w(s)\Big)
\end{aligned}
$$

[1] See https://arxiv.org/pdf/2001.09111.pdf for more details of the data and model. Dataset available in the spNNGP R-package.

Results using Stan

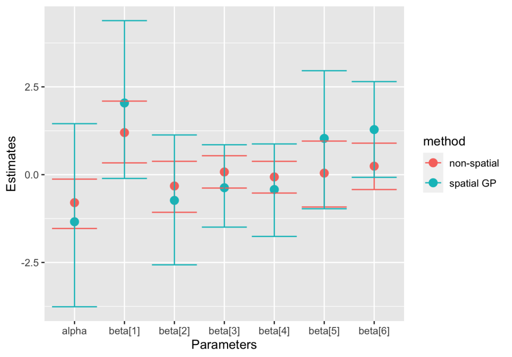


Figure: Regression coefficient estimates from spatial and non-spatial GLM.

Model	WAIC
Non-spatial GLM	206
Spatial GLM using GP	135

Table: Comparison of the spatial and non-spatial GLM.

Example 3: Spatially varying coefficient models

- The spatial linear model can be reparametrized as $y(s) = (\alpha + w(s)) + x(s)'\beta + \epsilon(s)$ where $\alpha + w(s)$ is the spatially-varying intercept (SVI) and the slope β is constant
- In many applications, the relationship between y and x can also vary over space and the assumption of a constant slope is inappropriate
- Such datasets can be modeled using spatially varying coefficient (SVC) models

Spatially varying coefficient models

- $y(s) = \Big(\alpha + w_\alpha(s)\Big) + x(s)'\Big(\beta + w_\beta(s)\Big) + \epsilon(s)$
- With p-covariates, $\beta + w_\beta(s)$ is a p-dimensional function modeling the space-varying slope
- $w(s) = (w_\alpha(s), w_\beta(s)')'$ is a $p+1$-dimensional function
- $w(s)$ can be modeled as either a collection of $p+1$-independent GPs or a $p+1$-dimensional multivariate GP
- `spSVC` function of the spBayes package offers fitting SVC model

Predicting air pollution in Central Europe

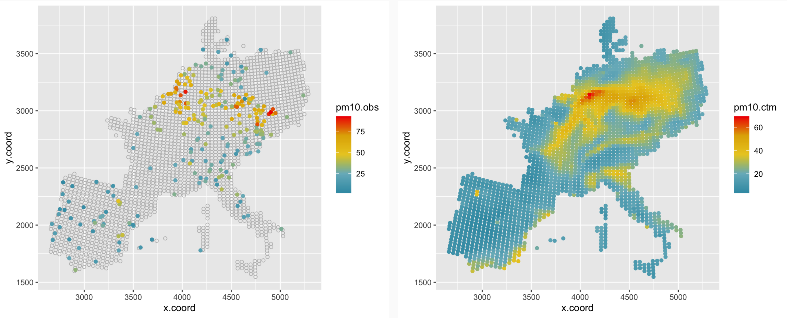
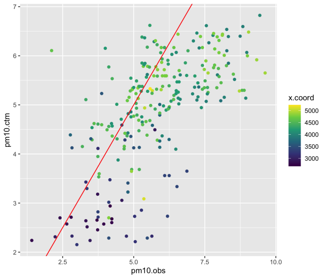


Figure: (Left) PM10 (in $\mu g/m^3$) measured at regulatory stations. (Right) PM10 output (in $\mu g/m^3$) from a chemistry transport model (CTM) at a grid over the area.

- Goal: Use the high resolution CTM output to predict PM10 at all the grid locations

Exploratory analysis



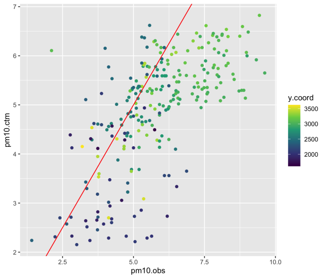


Figure: Scatterplots of square-root transformed PM10 measured by regulatory devices and predicted by CTM color-coded by longitude (left) and latitude (right).

Scatterplots reveal

- General under-prediction by CTM
- Variability in the relationship over space

SVC Model

$$
\begin{aligned}
\sqrt{PM10(s)} &= \beta_0 + w_0(s) + \sqrt{CTM(s)}(\beta_1 + w_1(s)) + \epsilon(s) \\
w_i(\cdot) &\sim GP(0, \sigma_i^2 R(\cdot, \cdot \,|\, \phi_i)), \text{ for } i = 0, 1 \\
\epsilon(s) &\overset{\text{iid}}{\sim} N(0, \tau^2) \\
\theta &= (\beta_0, \sigma_0^2, \sigma_1^2, \phi_0, \phi_1, \tau^2)' \sim p(\theta)
\end{aligned}
$$

Results

Parameters	50%	2.5%	97.5%
β_0	3.177	2.061	4.313
β_1	0.329	-0.085	0.737
σ_0^2	0.291	0.142	0.470
σ_1^2	0.104	0.070	0.151
ϕ_0	0.411	0.062	0.954
ϕ_1	0.002	0.001	0.002
τ^2	0.274	0.133	0.459

Results

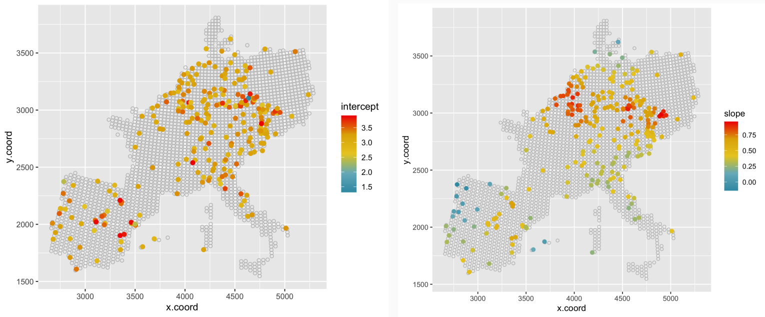
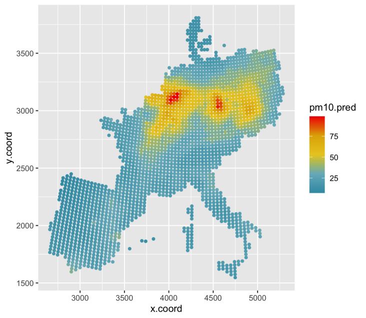


Figure: Estimates of the spatially-varying intercept (left) and spatially-varying slope (right) for the locations with regulatory data.

Results



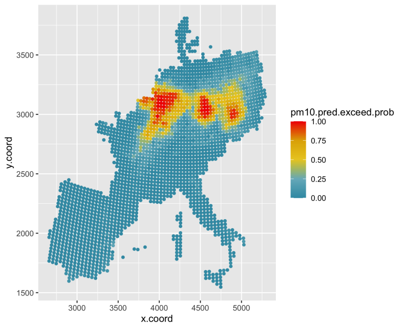


Figure: Predictions of PM10 (left) and predicted $P(PM10 > 50\mu g/m^3)$ (right) for the entire grid .

Review of lecture

- More examples of hierarchical spatial modeling:
    - Example 2: Spatial GLM for binary or count spatial data
    - Example 3: Spatially-varying coefficient model for heterogenous covariate effects