Lecture 23

Spatio-temporal Models

Colin Rundel

04/17/2017
Spatial Models with AR time dependence
Example - Weather station data

Based on Andrew Finley and Sudipto Banerjee’s notes from National Ecological Observatory Network (NEON) Applied Bayesian Regression Workshop, March 7 - 8, 2013 Module 6

**NETemp.dat** - Monthly temperature data (Celsius) recorded across the Northeastern US starting in January 2000.

```r
library(spBayes)
data("NETemp.dat")
ne_temp = NETemp.dat %>%
  filter(UTMX > 5.5e6, UTMY > 3e6) %>%
  select(1:27) %>%
tbl_df()

ne_temp
```

### A tibble: 34 × 27

<table>
<thead>
<tr>
<th>elev</th>
<th>UTMX</th>
<th>UTMY</th>
<th>y.1</th>
<th>y.2</th>
<th>y.3</th>
<th>y.4</th>
<th>y.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;int&gt;</td>
<td>&lt;dbl&gt;</td>
<td>&lt;dbl&gt;</td>
<td>&lt;dbl&gt;</td>
<td>&lt;dbl&gt;</td>
<td>&lt;dbl&gt;</td>
<td>&lt;dbl&gt;</td>
<td>&lt;dbl&gt;</td>
</tr>
<tr>
<td>1</td>
<td>6094162</td>
<td>3195181</td>
<td>-6.38889</td>
<td>-3.611111</td>
<td>3.722222</td>
<td>6.777778</td>
<td>12.55556</td>
</tr>
<tr>
<td>2</td>
<td>6245390</td>
<td>3262354</td>
<td>-6.277778</td>
<td>-4.111111</td>
<td>2.611111</td>
<td>6.555556</td>
<td>11.38889</td>
</tr>
<tr>
<td>3</td>
<td>6157302</td>
<td>3484043</td>
<td>-11.111111</td>
<td>-9.444444</td>
<td>-0.388889</td>
<td>3.944444</td>
<td>9.888889</td>
</tr>
<tr>
<td>5</td>
<td>6004871</td>
<td>3275456</td>
<td>-12.611111</td>
<td>-9.055556</td>
<td>-1.500000</td>
<td>2.944444</td>
<td>9.000000</td>
</tr>
<tr>
<td>6</td>
<td>6051946</td>
<td>3225830</td>
<td>-9.111111</td>
<td>-6.388889</td>
<td>1.222222</td>
<td>4.944444</td>
<td>10.88889</td>
</tr>
<tr>
<td>7</td>
<td>6099462</td>
<td>3184587</td>
<td>-7.944444</td>
<td>-6.055556</td>
<td>2.055556</td>
<td>5.555556</td>
<td>11.11111</td>
</tr>
<tr>
<td>9</td>
<td>6174891</td>
<td>3455064</td>
<td>-9.944444</td>
<td>-8.944444</td>
<td>-0.277778</td>
<td>3.555556</td>
<td>9.611111</td>
</tr>
<tr>
<td>10</td>
<td>6005282</td>
<td>3327413</td>
<td>-12.277778</td>
<td>-9.444444</td>
<td>-1.500000</td>
<td>2.944444</td>
<td>9.000000</td>
</tr>
</tbody>
</table>

# ... with 24 more rows, and 19 more variables: y.6 <dbl>, y.7 <dbl>, y.8 <dbl>, y.9 <dbl>, y.10 <dbl>, y.11 <dbl>, y.12 <dbl>, y.13 <dbl>,
Dynamic Linear / State Space Models (time)

\[
y_t = F_t \theta_t + v_t \\
\theta_t = G_t \theta_{t-1} + \omega_t
\]

observation equation

\[
v_t \sim \mathcal{N}(0, V_t) \\
\omega_t \sim \mathcal{N}(0, W_t)
\]
DLM vs ARMA

ARMA / ARIMA are a special case of a dynamic linear model, for example an $AR(p)$ can be written as

$$F_t' = (1, 0, \ldots, 0)$$

$$G_t = \begin{pmatrix}
\phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & 0 \\
0 & 0 & \cdots & 1 & 0
\end{pmatrix}$$

$$\omega_t = (\omega_1, 0, \ldots, 0), \quad \omega_1 \sim \mathcal{N}(0, \sigma^2)$$
ARMA / ARIMA are a special case of a dynamic linear model, for example an AR$(p)$ can be written as

\[ F_t' = (1, 0, \ldots, 0) \]

\[ G_t = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \]

\[ \omega_t = (\omega_1, 0, \ldots 0), \quad \omega_1 \sim \mathcal{N}(0, \sigma^2) \]

\[ y_t = \theta_t + v_t, \quad v_t \sim \mathcal{N}(0, \sigma_v^2) \]

\[ \theta_t = \sum_{i=1}^{p} \phi_i \theta_{t-i} + \omega_1, \quad \omega_1 \sim \mathcal{N}(0, \sigma_\omega^2) \]
Dynamic spatio-temporal models

The observed temperature at time $t$ and location $s$ is given by $y_t(s)$ where,

$$y_t(s) = x_t(s)\beta_t + u_t(s) + \epsilon_t(s)$$

$$\epsilon_t(s) \sim \mathcal{N}(0, \tau_t^2)$$

$$\beta_t = \beta_{t-1} + \eta_t$$

$$\eta_t \sim \mathcal{N}(0, \Sigma_\eta)$$

$$u_t(s) = u_{t-1}(s) + w_t(s)$$

$$w_t(s) \sim \mathcal{N}(0, \Sigma_t(\phi_t, \sigma_t^2))$$
The observed temperature at time $t$ and location $s$ is given by $y_t(s)$ where,

$$y_t(s) = x_t(s)\beta_t + u_t(s) + \epsilon_t(s)$$

$$\epsilon_t(s) \sim \mathcal{N}(0, \tau_t^2)$$

$$\beta_t = \beta_{t-1} + \eta_t$$

$$\eta_t \sim \mathcal{N}(0, \Sigma_\eta)$$

$$u_t(s) = u_{t-1}(s) + w_t(s)$$

$$w_t(s) \sim \mathcal{N}(0, \Sigma_t(\phi_t, \sigma_t^2))$$

Additional assumptions for $t = 0$,

$$\beta_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$$

$$u_0(s) = 0$$
**Data**:  

```r
coords = ne_temp %>% select(UTMX, UTMY) %>% as.matrix() / 1000
y_t = ne_temp %>% select(starts_with("y.")) %>% as.matrix()

max_d = coords %>% dist() %>% max()
n_t = ncol(y_t)
n_s = nrow(y_t)
```

**Parameters**:  

```r
n_beta = 2
starting = list(
    beta = rep(0, n_t * n_beta),
    phi = rep(3/(max_d/2), n_t),
    sigma.sq = rep(1, n_t),
    tau.sq = rep(1, n_t),
    sigma.eta = diag(0.01, n_beta)
)

tuning = list(phi = rep(1, n_t))

priors = list(
    beta.0.Norm = list(rep(0, n_beta), diag(1000, n_beta)),
    phi.Unif = list(rep(3/(0.9 * max_d), n_t), rep(3/(0.05 * max_d), n_t)),
    sigma.sq.IG = list(rep(2, n_t), rep(2, n_t)),
    tau.sq.IG = list(rep(2, n_t), rep(2, n_t)),
    sigma.eta.IW = list(2, diag(0.001, n_beta))
)
```
Fitting with spDynLM from spBayes

n_samples = 10000
models = lapply(paste0("y.", 1:24, "~elev"), as.formula)

m = spDynLM(
  models, data = ne_temp, coords = coords, get.fitted = TRUE,
  starting = starting, tuning = tuning, priors = priors,
  cov.model = "exponential", n.samples = n_samples, n.report = 1000)

save(m, ne_temp, models, coords, starting, tuning, priors, n_samples,
  file="dynlm.Rdata")

## ----------------------------------------
## General model description
## ----------------------------------------
## Model fit with 34 observations in 24 time steps.
## Number of missing observations 0.
## Number of covariates 2 (including intercept if specified).
## Using the exponential spatial correlation model.
## Number of MCMC samples 10000.
## ...
Posterior Inference - $\beta$s

Lapse Rate
Posterior Inference - $\theta$
Posterior Inference - Observed vs. Predicted

The graph shows the comparison of observed values (y_obs) and predicted values (y_hat). The data points align closely along a 45-degree line, indicating a strong correlation between the observed and predicted values.
**Prediction**

*spPredict* does not support *spDynLM* objects.

```r
r = raster(xmn=575e4, xmx=630e4, ymn=300e4, ymx=355e4, nrow=20, ncol=20)
pred = xyFromCell(r, 1:length(r)) %>%
  cbind(elev=0, ., matrix(NA, nrow=length(r), ncol=24)) %>%
  as.data.frame() %>%
  setNames(names(ne_temp)) %>%
  rbind(ne_temp, .) %>%
  select(1:15) %>%
  select(-elev)
```

```r
models_pred = lapply(paste0("y.", 1:n_t, "~1"), as.formula)
n_samples = 5000
m_pred = spDynLM(
  models_pred, data = pred, coords = coords_pred, get.fitted = TRUE,
  starting = starting, tuning = tuning, priors = priors,
  cov.model = "exponential", n.samples = n_samples, n.report = 1000)
```

```r
save(m_pred, pred, models_pred, coords_pred, y_t_pred, n_samples,
     file="dynlm_pred.Rdata")```
Spatio-temporal models for continuous time
In general, spatiotemporal models will have a form like the following,

\[ y(s, t) = \mu(s, t) + e(s, t) \]

mean structure \hspace{2cm} error structure

\[ = x(s, t) \beta(s, t) + w(s, t) + \epsilon(s, t) \]

Regression \hspace{2cm} Spatiotemporal RE \hspace{2cm} Error
In general, spatiotemporal models will have a form like the following,

\[ y(s, t) = \mu(s, t) + e(s, t) \]

\[ = x(s, t) \beta(s, t) + w(s, t) + \epsilon(s, t) \]

The simplest possible spatiotemporal model is one where we assume there is no dependence between observations in space and time,

\[ w(s, t) = \alpha(t) + \omega(s) \]

these are straightforward to fit and interpret but are quite limiting (no shared information between space and time).
Spatiotemporal Covariance

Let's assume that we want to define our spatiotemporal random effect to be a single stationary Gaussian Process (in 3 dimensions\textsuperscript{*}),

\[ w(s, t) \sim \mathcal{N}(0, \Sigma(s, t)) \]

where our covariance function depends on both $\|s - s'\|$ and $|t - t'|$,

\[ \text{cov}(w(s, t), w(s', t')) = c(\|s - s'\|, |t - t'|) \]

- Note that the resulting covariance matrix $\Sigma$ will be of size $n_s \cdot n_t \times n_s \cdot n_t$.
  - Even for modest problems this gets very large (past the point of direct computability).
  - If $n_t = 52$ and $n_s = 100$ we have to work with a $5200 \times 5200$ covariance matrix.
One solution is to use a separable form, where the covariance is the product of a valid 2d spatial and a valid 1d temporal covariance / correlation function,

$$\text{cov}(w(s, t), w(s', t')) = \sigma^2 \rho_1(\|s - s'\|; \theta) \rho_2(|t - t'|; \phi)$$
Separable Models

One solution is to use a separable form, where the covariance is the product of a valid 2d spatial and a valid 1d temporal covariance / correlation function,

$$\text{cov}(w(s, t), w(s', t')) = \sigma^2 \rho_1(\|s - s\|; \theta) \rho_2(|t - t'|; \phi)$$

If we define our observations as follows (stacking time locations within spatial locations)

$$w^t_s = (w(s_1, t_1), \ldots, w(s_1, t_{n_t}), w(s_2, t_1), \ldots, w(s_2, t_{n_t}), \ldots, \ldots, w(s_{n_s}, t_1), \ldots, w(s_{n_s}, t_{n_t}))$$

then the covariance can be written as

$$\Sigma_w(\sigma^2, \theta, \phi) = \sigma^2 H_s(\theta) \otimes H_t(\phi)$$

where $H_s(\theta)$ and $H_t(\phi)$ are $n_s \times n_s$ and $n_t \times n_t$ sized correlation matrices respectively and their elements are defined by

$$\{H_s(\theta)\}_{ij} = \rho_1(\|s_i - s_j\|; \theta)$$

$$\{H_t(\phi)\}_{ij} = \rho_1(|t_i - t_j|; \phi)$$
Kronecker Product

Definition:

\[
\begin{pmatrix}
A \\
B
\end{pmatrix}_{[m \times n] \times [p \times q]} =
\begin{pmatrix}
a_{11}B & \cdots & a_{1n}B \\
\vdots & \ddots & \vdots \\
a_{m1}B & \cdots & a_{mn}B
\end{pmatrix}_{[m \cdot p \times n \cdot q]}
\]
Kronecker Product

Definition:

\[ A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix} \]

Properties:

- \( A \otimes B \neq B \otimes A \) (usually)
- \( (A \otimes B)^t = A^t \otimes B^t \)
- \( \det(A \otimes B) = \det(B \otimes A) \)
  \[ = \det(A)^{\text{rank}(B)} \det(B)^{\text{rank}(A)} \]
- \( (A \otimes B)^{-1} = A^{-1}B^{-1} \)
If we have a spatiotemporal random effect with a separable form,

\[ w(s, t) \sim \mathcal{N}(0, \Sigma_w) \]

\[ \Sigma_w = \sigma^2 H_s \otimes H_t \]

then the likelihood for \( w \) is given by

\[
- \frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_w| - \frac{1}{2} w^t \Sigma_w^{-1} w
\]

\[
= - \frac{n}{2} \log 2\pi - \frac{1}{2} \log [(\sigma^2)^{n_t \cdot n_s} |H_s|^{n_t} |H_t|^{n_s}] - \frac{1}{2} w^t \frac{1}{\sigma^2} (H_s^{-1} \otimes H_t^{-1}) w
\]
Non-separable Models

- Additive and separable models are still somewhat limiting

- Cannot treat spatiotemporal covariances as 3d observations

- Possible alternatives:
  - Specialized spatiotemporal covariance functions, i.e.
    \[
    c(s - s', t - t') = \sigma^2 (|t - t'| + 1)^{-1} \exp \left( -\|s - s'\|(|t - t'| + 1)^{-\beta/2} \right)
    \]
  - Mixtures, i.e. \( w(s, t) = w_1(s, t) + w_2(s, t) \), where \( w_1(s, t) \) and \( w_2(s, t) \) have separable forms.