Bayesian dynamic modeling for large space-time weather datasets using Gaussian predictive processes

Andrew O. Finley and Sudipto Banerjee

1 Department of Forestry & Department of Geography, Michigan State University, Lansing Michigan, U.S.A.
2 Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, U.S.A.

March 8, 2013
1. Opportunities and challenges in spatio-temporal data analysis
2. Typical examples
3. Dynamic space-time modeling
4. Dimension reduction and predictive spatial process models
5. Data analysis
6. Extensions to multivariate setting
Challenges in spatio-temporal data analysis

Data sets often exhibit:

- missingness and misalignment among outcomes
- space- and time-varying impact of covariates
- complex residual dependence structures
- nonstationarity among multiple outcomes across locations
- unknown time and perhaps space lags between outcomes and covariates
Challenges in spatio-temporal data analysis

Data sets often exhibit:

- missingness and misalignment among outcomes
- space- and time-varying impact of covariates
- complex residual dependence structures
- nonstationarity among multiple outcomes across locations
- unknown time and perhaps space lags between outcomes and covariates

We seek modeling frameworks that:

- incorporate many sources of space and time indexed data
- accommodate structured residual dependence
- propagate uncertainty through to predictions
- scale to effectively exploit information in massive datasets
Typical examples

Discrete-time, continuous-space settings

Data is viewed as arising from a time series of spatial processes.

For example:

- US Environmental Protection Agency’s Air Quality System which reports pollutants’ mean, minimum, and maximum at 8 and 24 hour intervals

- soil sensor array that measures moisture and CO$_2$ at 1 minute intervals

- climate model outputs of weather variables generated on hourly or daily intervals

- remotely sensed landuse/landcover change recorded at annual or decadal time steps
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

1530 stations that recorded temperature between 1/1/2000 – 1/1/2005. The outcome variable is mean monthly temperature.
NCDC weather station data from North-eastern US

Of the $n \times N_t = 1530 \text{ station } \times 61 \text{ months} = 93,330$ possible observations, 15,107 had missing outcomes.
Dynamic space-time modeling

West and Harrison (1997); Tonellato; (1997); Stroud et al. (2001); Gelfand et al. (2005); ... 

- **Measurement equation**: Regression specification + space-time process + serially and spatially uncorrelated zero-centered Gaussian disturbances as measurement error

- **Transition equation**: Markovian dependence in time
  - temporal component: akin to usual state-space model
  - space-time component: modeled using Gaussian processes
Dynamic space-time model

\[ y_t(s) = x_t(s)' \beta_t + u_t(s) + \epsilon_t(s), \quad t = 1, 2, \ldots, N_t \]
Dynamic space-time modeling

\[ y_t(s) = x_t(s)' \beta_t + u_t(s) + \epsilon_t(s), \quad t = 1, 2, \ldots, N_t \]

\[ \beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \overset{iid}{\sim} N(0, \Sigma_\eta) ; \]
Dynamic space-time model

\[
y_t(s) = x_t(s)\beta_t + u_t(s) + \epsilon_t(s), \quad t = 1, 2, \ldots, N_t
\]

\[
\beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \overset{iid}{\sim} N(0, \Sigma_\eta);
\]

\[
u_t(s) = u_{t-1}(s) + w_t(s), \quad w_t(s) \overset{ind}{\sim} GP\left(0, C_t(\cdot, \theta_t)\right),
\]
Dynamic space-time model

\[ y_t(s) = x_t(s)' \beta_t + u_t(s) + \epsilon_t(s), \quad t = 1, 2, \ldots, N_t \]

\[ \beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim iid N(0, \Sigma_\eta) ; \]

\[ u_t(s) = u_{t-1}(s) + w_t(s), \quad w_t(s) \sim ind GP(0, C_t(\cdot, \theta_t)) , \]

\[ \epsilon_t(s) \sim ind N(0, \tau_t^2). \]

Assume \( \beta_0 \sim N(m_0, \Sigma_0) \) and \( u_0(s) \equiv 0 \)
Consider settings where the number of locations is large but the number of time points is manageable.

For any $\mathcal{I} = \{s_1, s_2, \ldots, s_n\} \subset \mathbb{R}^d$.

Then, $w_t = (w_t(s_1), w_t(s_2), \ldots, w_t(s_n))' \sim N(0, C_t(\theta_t))$, where $C_t(\theta_t) = \{C_t(s_i, s_j; \theta_t)\}$.
Consider settings where the number of locations is large but the number of time points is manageable.

For any $\mathcal{S} = \{s_1, s_2, \ldots, s_n\} \subset \mathbb{R}^d$.

Then, $w_t = (w_t(s_1), w_t(s_2), \ldots, w_t(s_n))' \sim N(0, C_t(\theta_t))$, where $C_t(\theta_t) = \{C_t(s_i, s_j; \theta_t)\}$.

Model parameter estimation requires iterative matrix decomposition for $C_t(\theta_t)^{-1}$ and determinant, i.e., computation of order $O(n^3)$ for every MCMC iteration!

Conducting full Bayesian inference is computationally onerous!
Modeling large spatial datasets

- Covariance tapering (Furrer et al. 2006; Zhang and Du 2007; Du et al. 2009; Kaufman et al., 2009)

- Spectral domain: (Fuentes 2007; Paciorek 2007)

- Approximate using MRFs: INLA (Rue et al. 2009)

- Approximations using cond. indep. (Vecchia 1988; Stein et al. 2004)

Predictive process models


- Parent process: \( w_t(s) \sim GP(0, C_t(\cdot; \theta_t)) \)
Predictive process models


- **Parent process**: \( w_t(s) \sim GP(0, C_t(\cdot; \theta_t)) \)

- **“Knots”**: \( \mathcal{S}^* = \{s_1^*, s_2^*, \ldots, s_n^*\} \) with \( n^* << n \)

- **\( w_t^* \)**: \( (w_t(s_1^*), w_t(s_2^*), \ldots, w_t(s_n^*))' \sim N(0, C_t^*(\theta_t)) \), where \( C_t^*(\theta_t) = \{C_t(s_i^*, s_j^*; \theta_t)\} \)

- **\( \tilde{w}_t(s) \)**: \( \tilde{w}_t(s) = E[w_t(s) | w_t^*] = c_t(s; \theta_t)'C_t^*(\theta_t)^{-1}w_t^* \), where \( c_t(s; \theta_t)' = \{C_t(s, s_j^*; \theta_t)\} \)

Replace \( u_t(s) \) by \( \tilde{u}_t(s) = \tilde{u}_{t-1}(s) + \tilde{w}_t(s) \). Now we only have order \( O(n^*^3) \) complexity for every MCMC iteration.
Illustration of biased estimates of variance parameters from the predictive process (synthetic data).
True parameter values $\sigma^2=5$ and $\tau^2=5$. Increasing knot intensity reduces bias.
Bias-adjusted predictive process

$$\tilde{w}_\epsilon(s) \sim \text{ind } N(\tilde{w}(s), C(s, s; \theta) - c(s; \theta)'C^*(\theta)^{-1}c(s; \theta)).$$
Bias-adjusted predictive process

\[ \tilde{w}_\epsilon(s)^{\text{iind}} \sim N(\tilde{w}(s), C(s, s; \theta) - c(s; \theta)'C^*(\theta)^{-1}c(s; \theta)). \]

Rectified predictive process model estimates (synthetic data)
Bias-adjusted predictive process

\[ \tilde{w}_{\tilde{\epsilon}}(s) \overset{ind}{\sim} N(\tilde{w}(s), C(s, s; \theta) - c(s; \theta)' C^*(\theta)^{-1} c(s; \theta)). \]

Rectified predictive process model estimates (synthetic data)

So, in our current setting replace \( u_t(s) \) by

\[ \tilde{u}_t(s) = \tilde{u}_{t-1}(s) + \tilde{w}_t(s) + \tilde{\epsilon}_t(s), \]

where again,

\[ \tilde{\epsilon}(s) \overset{ind}{\sim} N(0, C_t(s, s; \theta) - c_t(s; \theta)' C^*_t(\theta_t)^{-1} c_t(s; \theta_t)). \]
Hierarchical dynamic predictive process models

\[
\prod_{t=1}^{N_t} p(\theta_t) \times \prod_{t=1}^{N_t} IG\left(\tau_t^2 \mid a_\tau, b_\tau\right) \times N(\beta_0 \mid m_0, \Sigma_0) \times IW(\Sigma_\eta \mid r_\eta, \Upsilon_\eta)
\]

\[
\times \prod_{t=1}^{N_t} N(\beta_t \mid \beta_{t-1}, \Sigma_\eta) \times \prod_{t=1}^{N_t} N(\mathbf{w}_t^* \mid 0, \mathbf{C}_t^*(\theta_t))
\]

\[
\times \prod_{t=1}^{N_t} N(\tilde{\mathbf{u}}_t \mid \tilde{\mathbf{u}}_{t-1} + \mathbf{c}_t(\theta_t)' \mathbf{C}_t^*(\theta_t)^{-1} \mathbf{w}_t^*, \mathbf{D}_t)
\]

\[
\times \prod_{t=1}^{N_t} N(y_t \mid X_t/\beta_t + \tilde{u}_t, \tau_t^2 I_n).
\]
How do we choose $\mathcal{S}^*$

Knot selection – Regular grid? More knots near locations we have sampled more?

- formal spatial design paradigm: maximize information metrics
- geometric considerations: space-filling designs; various clustering algorithms
- adaptive modeling of knots using point processes, see, e.g., Guhaniyogi et al. (2011)
How do we choose $\mathcal{I}^*$

Knot selection – Regular grid? More knots near locations we have sampled more?

- formal spatial design paradigm: maximize information metrics
- geometric considerations: space-filling designs; various clustering algorithms
- adaptive modeling of knots using point processes, see, e.g., Guhaniyogi et al. (2011)

In practice:

- compare performance of estimation of range and smoothness by varying knot intensity
- usually inference is quite robust to $\mathcal{I}^*$.
- more important to capture loss of variability due to low rank approximation
NCDC weather station data analysis

Candidate models:
- non-spatial, i.e., $\tilde{u}_t$'s = 0, but temporally dependent $\beta_t$'s
- full model with spatial predictive process using 5, 10, 25, and 50 knot intensities

Assessment:
- model fit – minimum posterior predictive loss approach
- predictive performance based on 1000 holdout observations – RMSPE and CI coverage rate
Elevation covariate
Knot locations
Dynamic estimation of the intercept $\beta_{0,t}$ (25 knot model)
Dynamic estimation of the impact of elevation $\beta_{1,t}$ (25 knot model)
Data analysis

Non-spatial Predictive process knots 25

<table>
<thead>
<tr>
<th></th>
<th>Non-spatial</th>
<th>Predictive process knots 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Sigma_{\eta}[1,1]$</td>
<td>24.41 (17.97, 36.65)</td>
<td>23.61 (16.83, 35.45)</td>
</tr>
<tr>
<td>$\Sigma_{\eta}[2,1]$</td>
<td>0.001 (-0.050, 0.054)</td>
<td>0.001 (-0.050, 0.055)</td>
</tr>
<tr>
<td>$\Sigma_{\eta}[2,2]$</td>
<td>0.002 (0.001, 0.002)</td>
<td>0.002 (0.001, 0.002)</td>
</tr>
</tbody>
</table>

$\Sigma_{\eta}$ parameter credible intervals, 50 (2.5, 97.5) percentiles, for the non-spatial and 25 knot predictive process models.
Dynamic estimation of spatial range $\theta_1$ (25 knot model)
Dynamic estimation of partial sill $\theta_2$ (25 knot model)
Dynamic estimation of nugget $\tau_t^2$ (25 knot model)
Data analysis

Fitted values

Non-spatial model

Full model (25 knots)
Using a squared error loss function, $G$ lack of fit, $P$ predictive variance, $D = G + P$ lower is better (Gelfand and Ghosh, 1998).

<table>
<thead>
<tr>
<th></th>
<th>Non-spatial</th>
<th>Predictive process knots</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>$G$</td>
<td>699174</td>
<td>282</td>
</tr>
<tr>
<td>$P$</td>
<td>699036</td>
<td>8333</td>
</tr>
<tr>
<td>$D$</td>
<td>1398210</td>
<td>8616</td>
</tr>
<tr>
<td>RMSPE</td>
<td>8.4</td>
<td>0.54</td>
</tr>
<tr>
<td>CI cover %</td>
<td>96</td>
<td>97</td>
</tr>
</tbody>
</table>
Cross-validation and temporal imputation
Data analysis

Cross-validation and temporal imputation

Years

Temp.

2000 2001 2002 2003 2004 2005

−10 0 10 20 30

2003 2004
Extending this model to consider $m$ outcome variables, e.g., temperature and precipitation, at each location.

Simply choose your favorite approach to modeling

$$w_t(s) = (w_{t,1}(s), w_{t,2}(s), \ldots, w_{t,m}(s))^l.$$

- Constructive
- LMC
- Multivariate Matérn
- Kernel Convolution
- Latent Dimension
- Characterize
Summary

Challenge – to meet modeling needs:
- predictive process models for large datasets
- use model-based adjustment to compensate for over-smoothing

Future work:
- extend to multivariate response and spatially-varying coefficients
- make these models available in the R package \textit{spBayes}

This work was supported by NSF Grant DMS 0706870.