Exploring Latent Structure in Multivariate Spatial Temporal Processes Using Process Convolutions

Catherine Calder

Institute of Statistics & Decision Sciences

Duke University

Abstract

Bayesian process convolution models provide an appealing approach for modeling spatial temporal data. Their structure can be exploited to significantly reduce the dimensionality of a complex spatial temporal process. Dynamic process convolution models can easily be extended to model multivariate spatial time-series. Instead of specifying the cross-covariance structure directly, we construct an underlying dynamic factor model that provides insight into the covariance structure. By constructing a factor model, we further reduce the model's dimension temporally. Each of the factors evolves over time and the data are modeled as a smoothed weighted average of these underlying factor processes. Inference procedures remain computationally tractable due to the additional reduction in the dimensionality of the model. We illustrate this model using multivariate pollutant data taken from the EPA's Clean Air Status and Trends Network (CASTNet).

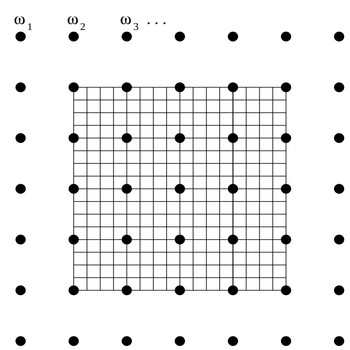
Background: Discrete Process Convolutions

(Higdon, 1998)

A Gaussian Process can be created by convolving a convolution kernel κ with a continuous white noise process x over a region D.

$$\psi(s) = \int_D \kappa(\omega - s)x(\omega)d\omega, \text{ for } s \in D.$$

This process can be approximated using a discrete white noise process on a lattice instead of a continuous process.



- Define a lattice over the field D with M points: $\omega_1, \ldots, \omega_M$.
- Create a white noise process at the ω's, $x(\omega_1), \ldots, x(\omega_M)$ where $x(\omega_i) \sim N(0, \lambda_x)$.
- The value of the field at any point $s \in D$ is then defined by the convolution of the white noise process with a kernel κ (with the level adjusted by a mean parameter μ),

$$\psi(s \mid x) = \sum_{i=1}^{M} \kappa(\omega_i - s) x(\omega_i) + \mu.$$

  Thus, the vector of values of the process at all sites on the grid can be written as

$$\boldsymbol{\psi} = \mathbf{K}\boldsymbol{x} + \boldsymbol{\mu}$$

  where $\boldsymbol{x} = [x(\omega_1), x(\omega_2), ..., x(\omega_M)]'$ and $\mathbf{K}$ is a matrix with rows $\mathbf{K}(s) = [\kappa(\omega_1 - s), \kappa(\omega_2 - s), ..., \kappa(\omega_M - s)]$ for all $s \in D$.

Note: All kernels κ will be **Gaussian** in the examples given here. For examples of non-Gaussian kernels see Higdon et al. (1999) and Kern (2000).

General Univariate Dynamic Process Convolution Model

Define $x(\omega_i, t)$ to be a Gaussian random walk for $t \in T$ and i.i.d. for all $\omega_i \in D$. A space-time process can be defined as

$$Y(s,t|x) = \sum_{i=1}^{M} \kappa(\omega_i - s) x(\omega_i, t) + \mu + \epsilon_{s,t}$$
$$x(\omega_i, t) = x(\omega_i, t-1) + \nu_{i,t}, \text{ for all } i$$

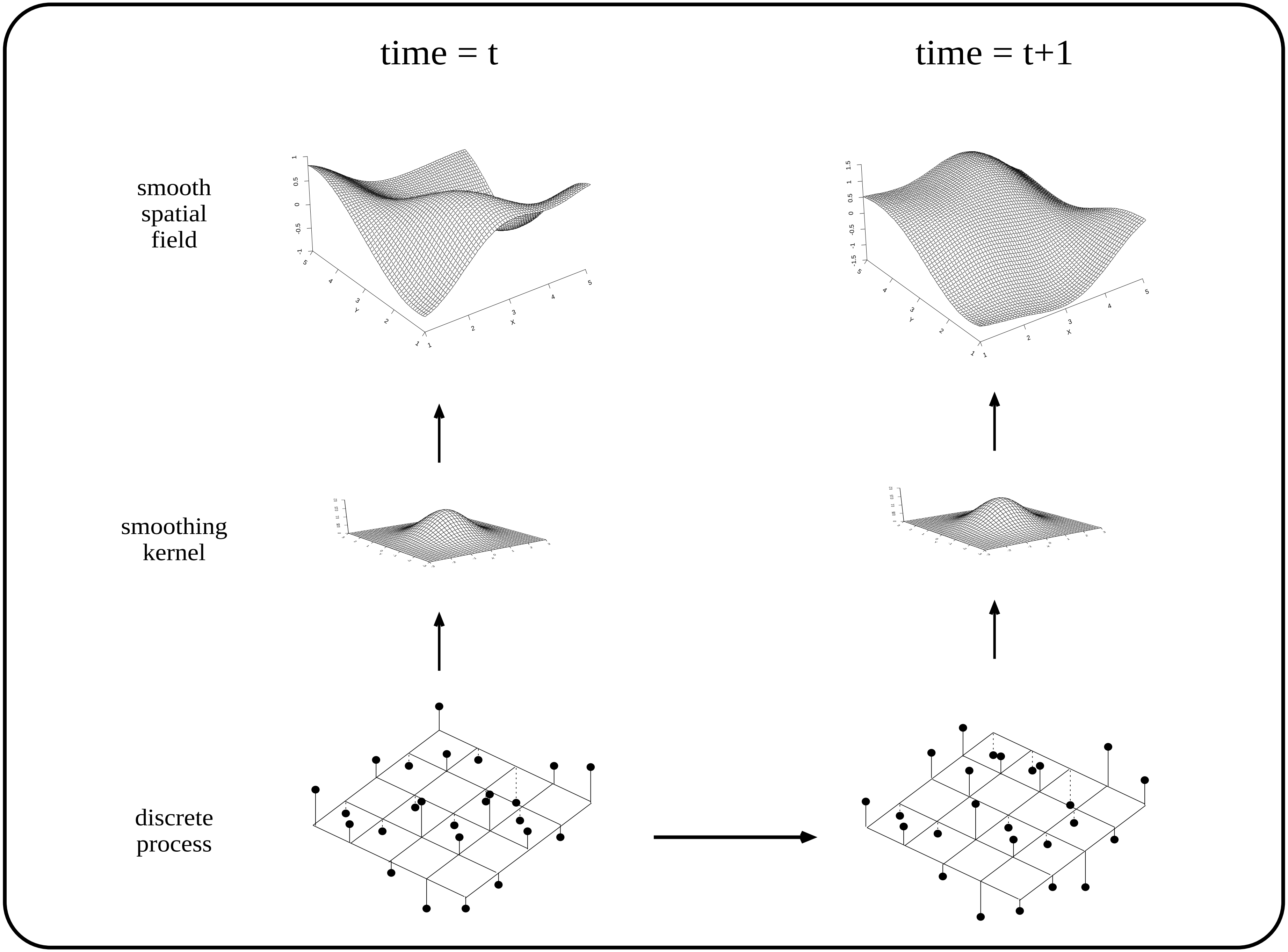
where $\epsilon_{s,t} \overset{i.i.d.}{\sim} N(0, \lambda_\epsilon)$ and $\nu_{i,t} \overset{i.i.d.}{\sim} N(0, \lambda_\nu)$.

This process can be written as

$$\underset{(N\times 1)}{\boldsymbol{Y}_t} = \underset{(N\times M)}{\mathbf{K}} \underset{(M\times 1)}{\boldsymbol{x}_t} + \underset{(N\times 1)}{\boldsymbol{\mu}} + \underset{(N\times 1)}{\boldsymbol{\epsilon}_t}$$
$$\underset{(M\times 1)}{\boldsymbol{x}_t} = \underset{(M\times 1)}{\boldsymbol{x}_{t-1}} + \underset{(M\times 1)}{\boldsymbol{\nu}_t}$$

where $\boldsymbol{x}_t$ and $\mathbf{K}$ have the same definition as above and N is the number of spatial locations.

$\Rightarrow$ Reduced-Dimension Space-Time Kalman Filter (Cressie and Wikle, 2002)


time = t
time = t+1
smooth
spatial
field
smoothing
kernel
discrete
process

Motivation

- Flexible Models
  - general model for both univariate and multivariate spatial temporal processes
  - learn about underlying spatial or temporal trends
  - learn about the covariance structure of the process
  - straightforward to model non-stationarity, anisotropy, misaligned data, etc.
- Reduction of Computational Expense
  - reduction of spatial dimension (process convolutions)
  - reduction of 'multivariate' dimension (dynamic factor analysis)

Multivariate Dynamic Factor PC Model

(Based on models from Aguilar and West (2000), Lopes and West (1998), etc.)

Measurement Equation:

$$\underset{(NI\times 1)}{Vec(\boldsymbol{Y}_t)} = (\underset{(N\times M)}{\mathbf{K}} \otimes \mathbf{I}_{I\times I})(\boldsymbol{I}_{M\times M} \otimes \underset{(I\times K)}{\boldsymbol{F}})\underset{(MK\times 1)}{Vec(\mathbf{X}_t)} + \underset{(NI\times 1)}{Vec(\boldsymbol{\mu}_t)} + \underset{(NI\times 1)}{Vec(\boldsymbol{\epsilon}_t)}$$

$$= (\mathbf{K} \otimes \boldsymbol{F})Vec(\mathbf{X}_t) + Vec(\boldsymbol{\mu}_t) + Vec(\boldsymbol{\epsilon}_t)$$

$$\underset{(N\times I)}{\boldsymbol{Y}_t} = \underset{(N\times M)}{\mathbf{K}}\underset{(M\times K)}{\mathbf{X}_t}\underset{(K\times I)}{\boldsymbol{F}}{}' + \underset{(N\times I)}{\boldsymbol{\mu}_t} + \underset{(N\times I)}{\boldsymbol{\epsilon}_t}$$

Temporal Process Equation:

$$\underset{(M\times K)}{\mathbf{X}_t} = \underset{(M\times K)}{\mathbf{X}_{t-1}} + \underset{(M\times K)}{\boldsymbol{\nu}_t}$$

- $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{N\times N}, diag(\boldsymbol{\lambda}_\epsilon)_{I\times I})$, where $\boldsymbol{\lambda}_\epsilon = (\lambda_{\epsilon_1}, \lambda_{\epsilon_2}, ..., \lambda_{\epsilon_I})$
- $\boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{M\times M}, diag(\boldsymbol{\lambda}_\nu)_{K\times K})$, where $\boldsymbol{\lambda}_\nu = (\lambda_{\nu_1}, \lambda_{\nu_2}, ..., \lambda_{\nu_K})$

N - # of data points	I - # of data types
M - # of lattice points	K - # of factors

Note: $K < I$ (dimension reduction)

- Constraint on the factor loadings matrix $\boldsymbol{F}$:

$$F = \begin{bmatrix} 1 & 0 & \dots & & & 0 \\ w_{2,1} & 1 & 0 & \dots & & 0 \\ w_{3,1} & w_{3,2} & 1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ w_{k,1} & w_{k,2} & \dots & & & 1 \\ w_{k+1,1} & w_{k+1,2} & \dots & & & w_{k+1,K} \\ \vdots & & & & & \vdots \\ w_{I,1} & w_{I,2} & \dots & & & \dots w_{I,K} \end{bmatrix}$$

  This restriction insures identifiability of the model.
  (Aguilar and West, 2000, Geweke and Zhou, 1996)

- Restriction on the number of factors: see slide 14

- Order of the time-series impacts:
  - interpretation of the factors
  - choice of K
  - model fit

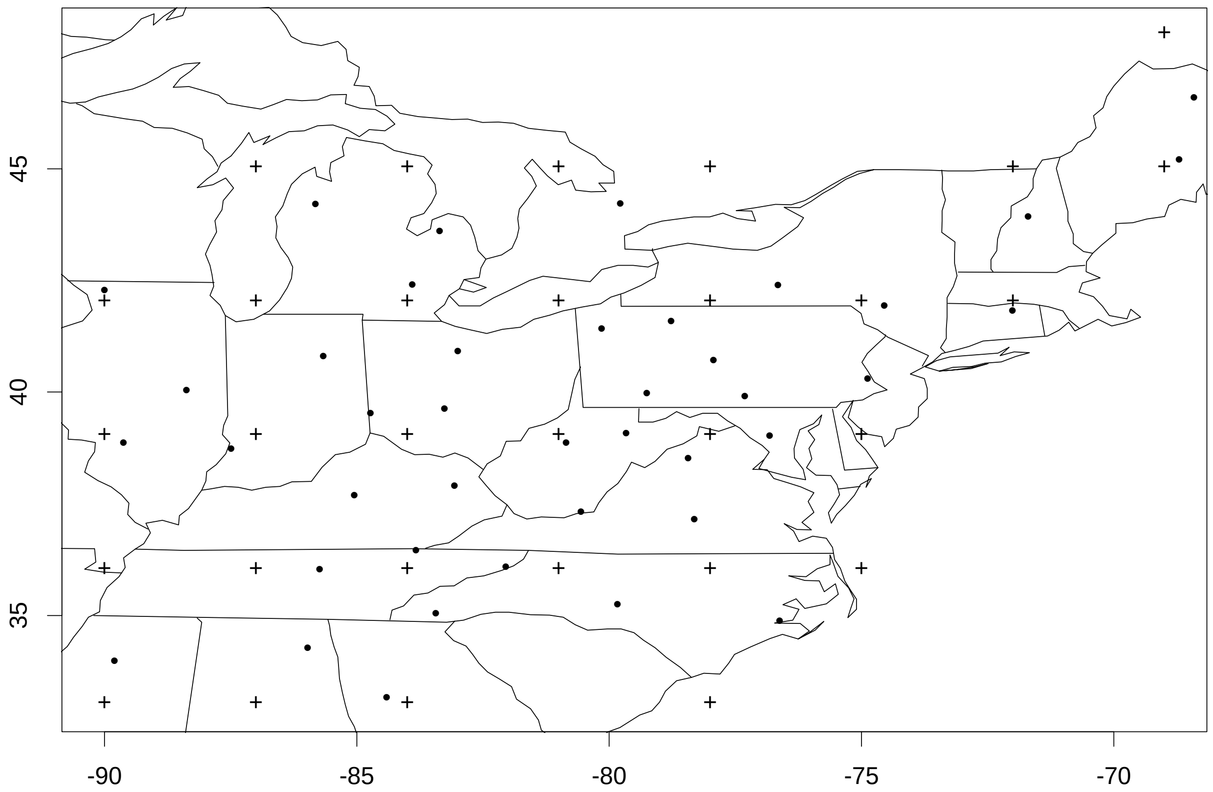
Example: CASTNet Data

Clean Air Status and Trends Network (EPA) - mandated by the 1990 Clean Air Act Amendments (CAAA) to determine the effectiveness of emission reductions by detecting and quantifying trends in pollutants

Summary of Relevant Data

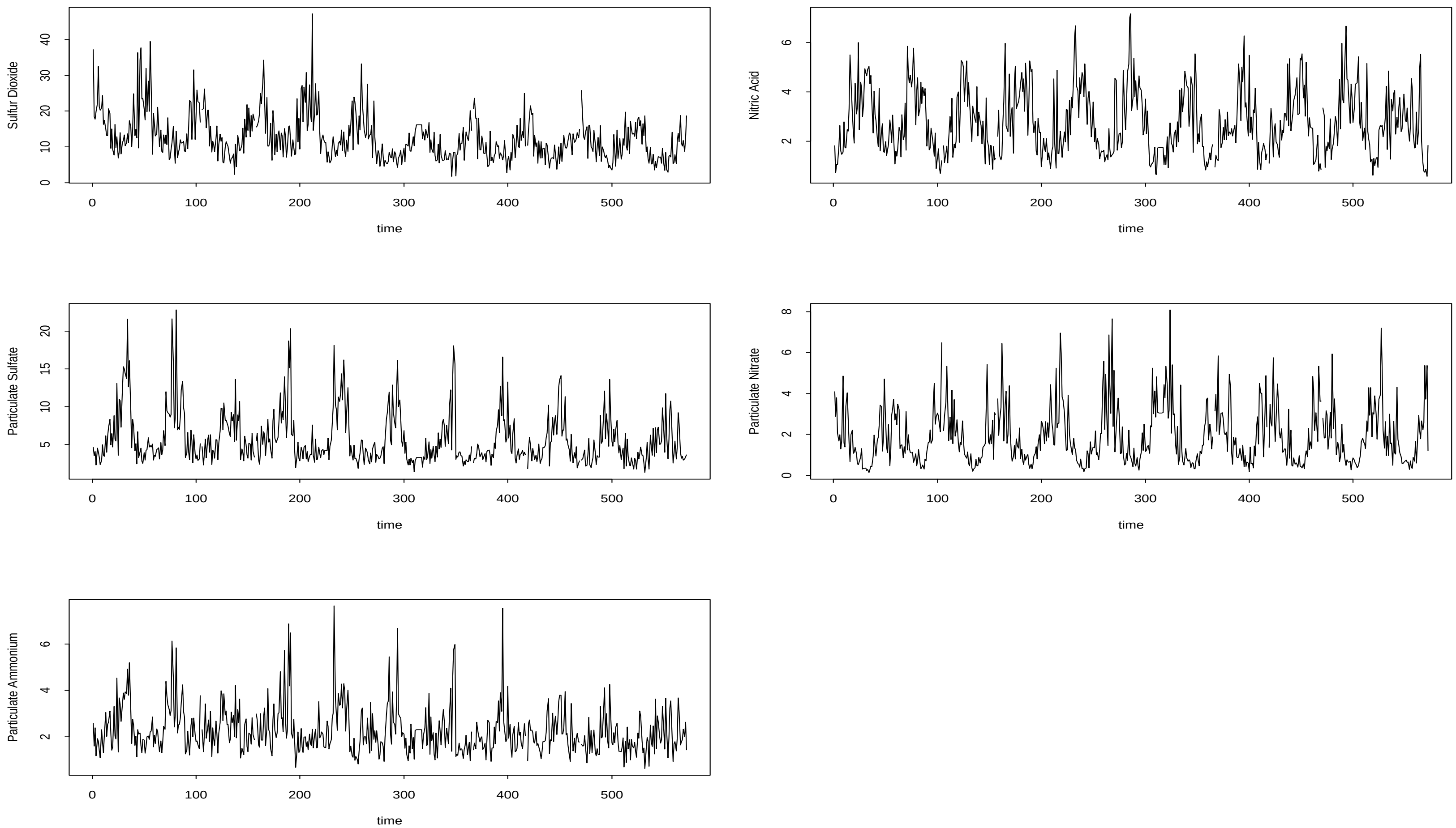
- Pollutants: Sulfur Dioxide (SO_2), Nitric Acid (HNO_3), Particulate Sulfate (SO_4^{2-}), Particulate Nitrate (NO_3^-), and Particulate Ammonium (NH_4^+) concentrations – weekly measurements
- Meteorological Measurements: Temperature, Δ Temperature, Relative Humidity, Solar Radiation, Precipitation, Wind Speed, Wind Direction – hourly measurements
- 42 monitoring locations in the Eastern United States out of 101 locations (100 in United States, 1 in Canada)
- 1990-2001 (572 Weeks)

Locations of the Eastern U.S. Monitoring Stations (42)



$+ =$ locations of the underlying process (30)

Pollution Data from a site in northeastern Alabama (lat=39.5, lon=-84.7)
January 1990 - December 2001



- Define $\boldsymbol{Y}_t$ be the log of the weekly pollution concentration readings.
- Mean Specification:

$$\mu_t(n,i) = \beta_0(n,i) + \beta_1(n,i)\sin(2\pi(t + c_1(n,i))/52)) + \beta_2(n,i)\cos(2\pi(t + c_2(n,i))/26))$$

  $\beta_1(n,i)$, $\beta_1(n,i)$, $c_1(n,i)$ and $c_2(n,i)$ are chosen using least-squares
  Note: eventually want to incorporate meteorological variables

- λ_{ν_k} is fixed at 0.1
- Priors:
  - Independent conjugate priors on λ_{ϵ_i} and X_0
  - Flat priors on $\beta_0(n,i)$ and $w_{i,k}$ (elements of F)
- Missing Data: imputed within Gibbs sampler
- MCMC algorithm (Gibbs sampler): straightforward since all full conditional distributions can be found in closed form
  Note: use dynamic linear modeling techniques to sample the $\mathbf{X}_t$'s

Choosing K

Restriction on the value of K

For dynamic factor model (for a multivariate time-series):

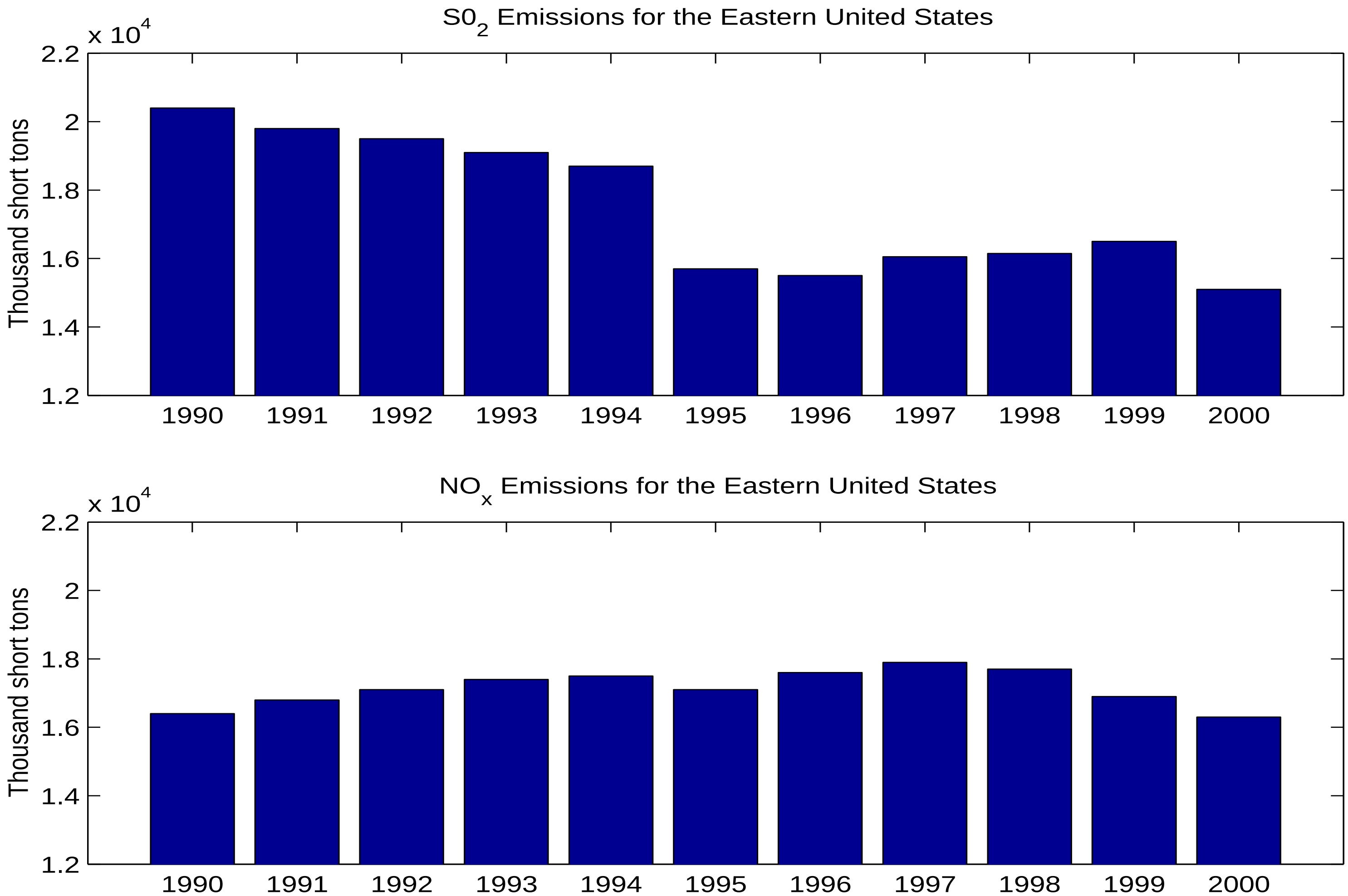
$$\frac{1}{2}I(I+1) - [(IK - \frac{1}{2}K(K+1)) + K + I] \geq 0$$
$$I = 5 \quad \Rightarrow \quad K <= 2$$

For PC dynamic factor model: no practical restriction on K for large N, but there may be numerical problems depending on the standard deviation of the convolution kernel and the difference between N and M

Methods for Choosing K

1. Interpretation (our method): **two factors** representing SO_2 and $N0_x$ emissions
2. Exploratory analysis (principal component decomposition)
3. Cross-Validation
4. Variance of Underlying Process (λ_ν)
5. Model Selection Criteria (see Lopes and West (1998))
6. Reversible Jump MCMC (Green, 1995)
   Algorithm developed in Lopes and West (1998) - independent parallel MCMC output to create proposal distributions

SO_2 *and* $N0_x$ *Emissions*



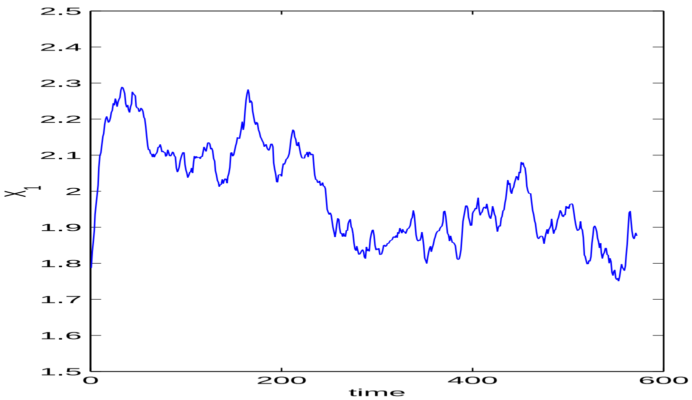
Total Annual Emissions over the Eastern United States

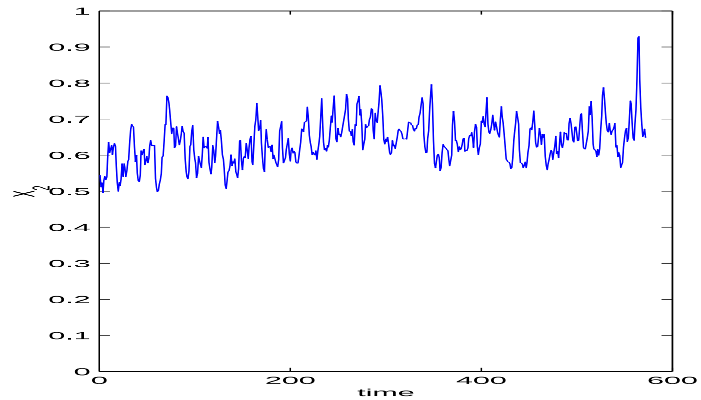
Results

Order of Pollutants = (SO_2,HNO_3,SO_4^{2-},NO_3^-,NH_4^+)

$$\hat{\boldsymbol{F}} = \begin{bmatrix} 1 & 0 \\ 0.224\ (0.178, 0.270) & 1 \\ 0.678\ (0.644, 0.715) & 0.686\ (0.590, 0.792) \\ 0.268\ (0.235, 0.307) & 0.883\ (0.716, 0.920) \\ 0.148\ (0.0871, 0.190) & 1.03\ (0.904, 1.18) \end{bmatrix}$$

$$\hat{\boldsymbol{\lambda}}_\epsilon = (3.01\ (2.80, 3.23), 0.540\ (0.455, 0.623), 1.83\ (1.65, 1.99), \\ 0.551\ (0.476, 0.630), 0.419\ (0.368, 0.476))$$





Multiresolution Models

General Framework

$$\psi(s,t) = \sum_{r=1}^{R} \psi^{(r)}(s,t).$$

Motivation

1. Modeling at different resolutions allows insight into the dynamics of the underlying process at various scales.
2. At fine resolutions, the underlying process may not exhibit temporal dependence yet the predictive ability of the model can be enhanced by including a fine resolution, temporally independent process.

Multiresolution Dynamic Factor PC Model

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \sum_{a=1}^{R_t} \mathbf{K}^{(r_a^t)} \mathbf{X}_t^{(r_a^t)} \boldsymbol{F}^{(r_a^t)\prime} + \sum_{b=1}^{R_s} \mathbf{K}^{(r_b^s)} \mathbf{X}_t^{(r_b^s)} \boldsymbol{F}^{(r_b^s)\prime} + \boldsymbol{\epsilon}_t$$

$\mathbf{X}_t^{(r_a^t)} = \mathbf{X}_{t-1}^{(r_a^t)} + \boldsymbol{\nu}_t^{(r_a^t)}$ for all a.

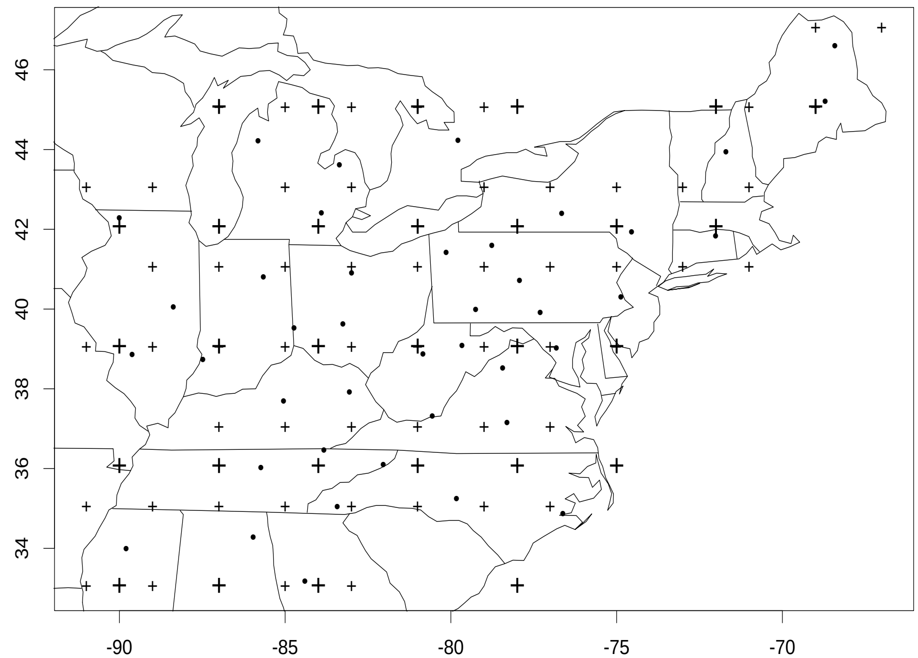
$\mathbf{X}_t^{(r_b^s)}$'s are temporally independent for all b.

Model for CASTNet data:

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \mathbf{K}^{(r_1^t)}\mathbf{X}_t^{(r_1^t)}\boldsymbol{F}^{(r_1^t)\prime} + \mathbf{K}^{(r_1^s)}\mathbf{X}_t^{(r_1^s)}\boldsymbol{F}^{(r_1^s)\prime} + \boldsymbol{\epsilon}_t$$

$$\mathbf{X}_t^{(r_1^t)} = \mathbf{X}_{t-1}^{(r_1^t)} + \boldsymbol{\nu}_t^{(r_1^t)}$$

$\Longrightarrow$ $\mathbf{X}_t^{(r_1^s)}$'s are temporally independent



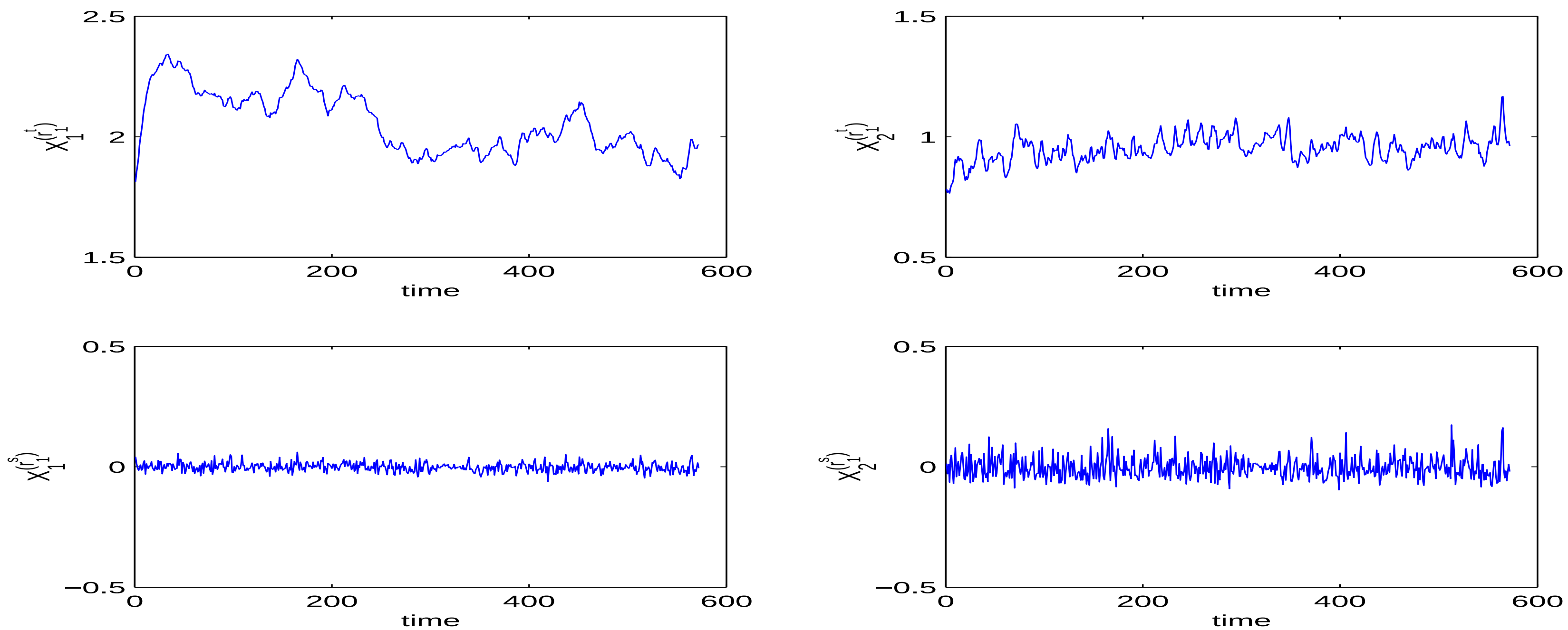
$+$ = locations of the coarse, temporally dependent underlying process (30)

$_+$ = locations of the fine, temporally independent underlying process (55)

Results

$$\hat{\boldsymbol{F}}^{(r_1^t)} = \begin{bmatrix} 1 & 0 \\ 0.107\ (0.044, 0.205) & 1 \\ 0.624\ (0.567, 0.688) & 0.574\ (0.475, 0.675) \\ 0.171\ (0.109, 0.253) & 0.837\ (0.748, 0.926) \\ 0.037\ (-0.033, 0.134) & 0.983\ (0.871, 1.081) \end{bmatrix} \quad \hat{\boldsymbol{F}}^{(r_1^s)} = \begin{bmatrix} 1 & 0 \\ 0.073\ (0.002, 0.137) & 1 \\ 0.105\ (-0.003, 0.205) & 0.420\ (0.318, 0.522) \\ -0.047(-0.110, 0.011) & -0.008\ (-0.072, 0.051) \\ 0.083\ (0.014, 0.149) & 0.240\ (0.170, 0.305) \end{bmatrix}$$

$$\hat{\boldsymbol{\lambda}}_\epsilon = (3.297\ (3.117, 3.472), 0.717\ (0.650, 0.787), 1.914\ (1.773, 2.0438), 0.598\ (0.534, 0.664), 0.447\ (0.402, 0.498))$$



Conclusions and Extensions

Conclusions:

- Dynamic PC models can capture structure in data which is not visible using ordinary exploratory analysis techniques.
- Inference procedures (MCMC) are realistic for large data sets.

Extensions:

- Mean Structure
- Spatial Prior on Factor Loadings Matrix
- Evolution Process of the Factors
- Visualization and Uncertainty Estimates