Hierarchical Dynamical Spatio-Temporal Models: *Statistics for Spatio-Temporal Data*, Ch. 7

John Paige

Statistics Department
University of Washington

January 29, 2018

Hierarchical Dynamical Spatio-Temporal Models (DSTMs): Data Model (Sec. 7.1)

$$[\{Z(\mathbf{x}; r) : \mathbf{x} \in D_s, r \in D_t\} \mid \{Y(\mathbf{s}; t) : \mathbf{s} \in \mathcal{N}_x, t \in \mathcal{N}_r\}, \boldsymbol{\theta}_D]$$

- $Z(\mathbf{x}, r)$: observations at location $\mathbf{x}$, time r
- $Y(\mathbf{s}; t)$: latent process at location $\mathbf{s}$, time t
- $\boldsymbol{\theta}_D$: data model parameters, possibly varying in space/time
- $\mathcal{N}_x$, $\mathcal{N}_r$: neighborhoods of $\mathbf{x}$ and r in space and time
- D: Data model

Process Model

$$\left[Y(\mathbf{s};t) \middle| \left\{ Y(\mathbf{w};t-\tau_1) : \mathbf{w} \in \mathcal{N}_s^{(1)} \right\}, ..., \left\{ Y(\mathbf{w};t-\tau_p) : \mathbf{w} \in \mathcal{N}_s^{(p)} \right\}, \boldsymbol{\theta}_P \right]$$

- $\mathcal{N}_s^{(1)}, ..., \mathcal{N}_s^{(p)}$: neighborhoods of location **s** at time lags $0, \tau_1, ..., \tau_p$
- $\boldsymbol{\theta}_P$: process model parameters, possibly varying in space/time
- P: process model

Parameter Model

$$[\boldsymbol{\theta}_D, \boldsymbol{\theta}_P | \boldsymbol{\theta}_h]$$

- $\boldsymbol{\theta}_D$, $\boldsymbol{\theta}_P$, $\boldsymbol{\theta}_h$: data, process, and hyperparameters

Linear Mappings with Equal Dimensions

$$\mathbf{Z}_t = \mathbf{Y}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}\right) \qquad (7.8)$$

$$Z(\mathbf{s};t) = a + hY(\mathbf{s};t) + \epsilon(\mathbf{s};t) \qquad E[\epsilon(\mathbf{s};t)] = 0 \qquad (7.9)$$

$$\mathbf{Z}_t = \mathbf{a}_t + \text{diag}(\mathbf{h}_t)\mathbf{Y}_t + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}\right) \qquad (7.10)$$

$$\mathbf{Z}_t = \mathbf{a}_t + \mathbf{H}_t\mathbf{Y}_t + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}\right) \qquad (7.11)$$

$$\mathbf{Z}_t = \mathbf{a}_t + \mathbf{H}_t\mathbf{Y}_t + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_t\right)$$

- (7.9): a, h are additive and multiplicative bias terms
- (7.10), (7.11): since $\mathbf{a}_t, \mathbf{h}_t, \mathbf{H}_t$ vary in space and time, requires simplifying assumptions about the latent process and data models
- (7.11): $\mathbf{R}_t$ can be modeled using a standard spatial covariance model

Linear Mappings with Unequal Dimensions: Intro

$$\mathbf{Z}_t = \mathbf{H}_t \mathbf{Y}_t + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim (\mathbf{0}, \mathbf{R}_t) \qquad (7.12)$$

- $\mathbf{H}_t : m_t \times n$
- $\boldsymbol{\epsilon}_t$ are independent

What form can $\mathbf{H}_t$ take?

Linear Mappings with Unequal Dimensions: Incidence Matrices

Say we have 3 observation locations $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ and two process locations, $\{\mathbf{s}_1, \mathbf{s}_2\}$, with $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{s}_1$ and $\mathbf{x}_3 = \mathbf{s}_2$. We could then write $\mathbf{H}_t$ as:

$$\mathbf{H}_t = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This is an **incidence matrix**.

Linear Mappings with Unequal Dimensions: Change of Support

Say we have 3 observation (areal) locations $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ and two process (areal) locations, $\{\mathbf{s}_1, \mathbf{s}_2\}$. We could then write $\mathbf{H}_t$ as:

$$\mathbf{H}_t = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ h_{31} & h_{32} \end{pmatrix} \tag{7.15}$$

$$h_{ij} = \frac{|\mathbf{x}_i \cap \mathbf{s}_j|}{|\mathbf{x}_i|} \tag{7.16}$$

where $|\mathbf{x}_i|$ represents the area of region $\mathbf{x}_i$.

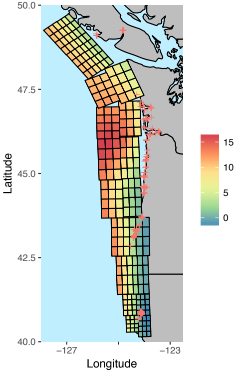
- Wikle and Berliner (2005) show this is optimal under ‘minor’ assumptions

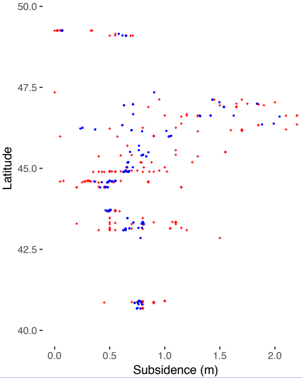
Linear Mappings with Unequal Dimensions: Earthquakes

For an earthquake at time **t**, we might want to model how the ground sinks (subsidence), $\mathbf{Z}_t$, for an earthquake $\mathbf{Y}_t$:

$$\mathbf{Z}_t = \mathbf{G}_t \mathbf{Y}_t + \epsilon_t$$

where **G** is a matrix determining subsidence resulting from an earthquake





Linear Mappings with Unequal Dimensions: Dimension Reduction

$$\mathbf{Y}_t = \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\nu}_t \quad (7.24)$$

$$\mathbf{Z}_t = \mathbf{H}_t\boldsymbol{\Phi}\boldsymbol{\alpha}_t + \underbrace{\mathbf{H}_t\boldsymbol{\nu}_t + \boldsymbol{\epsilon}_t}_{\gamma_t} \quad (7.25)$$

- Replacing with γ_t leads to replacing process $\mathbf{Y}_t$ with reduced dimension process $\boldsymbol{\alpha}_t$.
- Examples choices of basis matrix $\boldsymbol{\Phi}$:
  - Spectral representation
  - Empirical Orthogonal Functions (EOFs)
  - Dynamical system dependent approaches
  - Smoothing kernels
- Without assumed structure for $\boldsymbol{\Phi}$, model identifiability difficult

Dimension Reduction (kinda): Spectral Representation

$$\mathbf{Z}_t = \mathbf{H}_t \boldsymbol{\Phi} \boldsymbol{\alpha}_t + \underbrace{\mathbf{H}_t \boldsymbol{\nu}_t + \boldsymbol{\epsilon}_t}_{\gamma_t} \qquad (7.25)$$

- Assume $\mathbf{Y}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$
- Take $\boldsymbol{\Phi}$ $n \times n$ so that $\boldsymbol{\alpha}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\alpha)$
- If $\boldsymbol{\Phi}$ is orthogonal, often the case that $\boldsymbol{\Phi}$ has decorrelating effect:
$$\mathbf{R}_\alpha = \boldsymbol{\Phi}' \mathbf{R} \boldsymbol{\Phi} \approx \mathrm{diag}(\mathbf{d})$$
- Can use multiresolution wavelet basis functions if data is on a lattice (possibly using $\mathbf{H}_t$ incidence matrix for missing data)

Dimension Reduction: EOFs

$$\mathbf{Y}_t = \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\nu}_t \qquad (7.24)$$

$$\mathbf{Z}_t = \mathbf{H}_t\boldsymbol{\Phi}\boldsymbol{\alpha}_t + \underbrace{\mathbf{H}_t\boldsymbol{\nu}_t + \boldsymbol{\epsilon}_t}_{\gamma_t} \qquad (7.25)$$

- Take eigendecomposition of $\boldsymbol{\Phi}$, use the first p_α eigenvectors
- Then covariance in $\boldsymbol{\nu}_t$ could be characterized using next p_ν eigenvectors:

$$\boldsymbol{\Sigma}_\nu = c\mathbf{I} + \sum_{k=p_\alpha+1}^{p_\alpha+p_\nu} \lambda_k \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k'$$

- Problems: missing data, low number of temporal replicates, sensitivity to geometry of spatial domain, basis might poorly represent the dynamics

Process Models for the DSTM: Linear Models (Sec. 7.2)

We will consider vector autoregressive processes (VARs) of order one:

$$\mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t; \;\; t = 1, \; 2, \ldots \quad E[\boldsymbol{\eta}_t] = \mathbf{0} \quad \mathrm{Var}\,(\boldsymbol{\eta}_t) = \boldsymbol{\Sigma}_\nu$$

where $\boldsymbol{\eta}_t$ independent of $\mathbf{Y}_{t-1}$, $E[\mathbf{Y}_t] = \mathbf{0}$, and $\mathrm{Var}\,(\mathbf{Y}_t) = \boldsymbol{\Sigma}_Y$.

Process Models for the DSTM: Lagged Nearest-Neighbor Model

We will consider vector autoregressive processes (VARs) of order one:

$$Y_t(s_i) = \sum_{j \in \mathcal{N}_i} m_{ij} Y_{t-1}(s_j) + \eta_t(s_i)$$

- $\mathcal{N}_i$: neighborhood of s_i
- $\eta_t \stackrel{iid}{\sim} \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_t\right)$
- Dynamics could determine m_{ij}, $\mathcal{N}_{ij}$. Larger lags could be considered

Process Models for the DSTM: PDE-Based Parameterizations

One-dimensional diffusion equation:

$$\frac{\partial Y}{\partial t} = \frac{\partial}{\partial x}\left(b(x)\frac{\partial Y}{\partial x}\right)$$

$$\frac{\partial Y}{\partial x} \approx \frac{Y(x+\Delta_x;t) - Y(x-\Delta_x;t)}{2\Delta_x}$$

$$\frac{\partial^2 Y}{\partial x^2} \approx \frac{Y(x+\Delta_x;t) - 2Y(x;t) + Y(x-\Delta_x;t)}{\Delta_x^2}$$

$$\frac{\partial Y}{\partial t} \approx \frac{Y(x;t+\Delta_t) - Y(x;t)}{\Delta_t}$$

- $x \in [0, L]$: location
- $b(x)$: diffusion coefficients
- Boundary conditions: $Y(0;t) = Y_0$, $Y(L;t) = Y_L$, $\{Y(x;0) : 0 \le x \le L\}$ (known or have prior distribution)

Process Models for the DSTM: PDE-Based Parameterizations

For three locations, x_1, x_2, x_3, this yields:

$$
\begin{aligned}
Y(x; t+\Delta_t) &\approx \theta_1(x) Y(x;t) + \theta_2(x) Y(x+\Delta_x; t) \\
&\quad + \theta_3(x) Y(x-\Delta_x; t), \\
\Rightarrow \begin{pmatrix} Y(x_1; t+\Delta_t) \\ Y(x_2; t+\Delta_t) \\ Y(x_3; t+\Delta_t) \end{pmatrix} &\approx \begin{pmatrix} \theta_1(x_1) & \theta_2(x_2) & 0 \\ \theta_3(x_2) & \theta_1(x_2) & \theta_2(x_2) \\ 0 & \theta_3(x_3) & \theta_1(x_3) \end{pmatrix} \begin{pmatrix} Y(x_1;t) \\ Y(x_2;t) \\ Y(x_3;t) \end{pmatrix} \\
&\quad + \begin{pmatrix} \theta_3(x_1) & 0 \\ 0 & 0 \\ 0 & \theta_2(x_3) \end{pmatrix} \begin{pmatrix} Y_0 \\ Y_L \end{pmatrix}
\end{aligned}
$$

- $\theta_i(x)$ are known functions of $\Delta_x, \Delta_t, b(x), b(x-\Delta_x)$, and $b(x+\Delta_x)$

Process Models for the DSTM: PDE-Based Parameterizations

Add a stochastic term:

$$\mathbf{Y}_{t+\Delta_t} = \mathbf{M}(\boldsymbol{\theta})\mathbf{Y}_t + \mathbf{M}^{(b)}(\boldsymbol{\theta})\mathbf{Y}_t^{(b)} + \boldsymbol{\eta}_{t+\Delta_t}, \quad \boldsymbol{\eta}_t \overset{iid}{\sim} \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_\eta\right)$$

- $\theta_i(x)$ are known functions of $\Delta_x, \Delta_t, b(x), b(x - \Delta_x)$, and $b(x + \Delta_x)$
- This is a lagged nearest-neighbor model!
- Estimation of $b(x)$ and $\mathbf{Q}_\eta$ is nontrivial, may require simplifying assumptions

Process Models for the DSTM: Nonlinear Models (Sec. 7.3)

Nonlinear autoregressive model:

$$\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}, \boldsymbol{\eta}_t; \boldsymbol{\theta}_t), \qquad (7.62)$$

- $\mathcal{M}$: nonlinear Markovian function control process evolution

Nonlinear Models: Local Linear Approximations

Take a Taylor expansion (delta method) of $\mathcal{M}(\cdot)$ in (7.62):

$$\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}) + \boldsymbol{\eta}_t \tag{7.63}$$

$$\mathcal{M}(\mathbf{Y}_t) \approx \mathcal{M}(\widehat{\mathbf{Y}}_{t-1}) + \mathbf{M}_t(\mathbf{Y}_{t-1} - \widehat{\mathbf{Y}}_{t-1})), \quad E[\mathbf{Y}_t] = \widehat{\mathbf{Y}}_t \tag{7.64}$$

$$(\mathbf{M}_t)_{ij} = \left.\frac{\partial \mathcal{M}_i(\mathbf{Y})}{\partial \mathbf{Y}_t(s_j)}\right|_{\mathbf{Y}_t = \widehat{\mathbf{Y}}_{t-1}} \tag{7.65}$$

We can therefore write $\mathbf{Y}_t$ using the form:

$$\mathbf{Y}_t = \mathbf{c}_t + \mathbf{M}_t \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t \tag{7.66}$$

- Often times must determine $\mathcal{M}$ by estimating $\boldsymbol{\theta}_t$

Nonlinear Models: General Quadratic Nonlinearity

Take a second-order Taylor expansion of $\mathcal{M}(\cdot)$ in (7.62):

$$\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}) + \boldsymbol{\eta}_t \tag{7.63}$$

$$\begin{aligned}\mathcal{M}(\mathbf{Y}_t) &\approx \mathcal{M}(\widehat{\mathbf{Y}}_{t-1}) + \mathbf{M}_t(\mathbf{Y}_{t-1} - \widehat{\mathbf{Y}}_{t-1})) \\ &\quad + \frac{1}{2}(\mathbf{I} \otimes (\mathbf{Y}_t - \widehat{\mathbf{Y}}_t)')\mathbf{H}_t(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t)\end{aligned} \tag{7.80}$$

$$\mathbf{H}_t \equiv \mathbf{H}_t(\widehat{\mathbf{Y}}_t) = \begin{pmatrix} \mathbf{H}_{1t}(\mathbf{Y}_t) \\ \vdots \\ \mathbf{H}_{nt}(\mathbf{Y}_t) \end{pmatrix} \Bigg|_{\mathbf{Y}_t = \widehat{\mathbf{Y}}_t} \tag{7.81}$$

$$(\mathbf{H}_{it}(\widehat{\mathbf{Y}}_t))_{kl} = \frac{\partial^2 \mathcal{M}_i(\mathbf{Y}_t)}{\partial Y(s_k) \partial Y_t(s_l)} \Bigg|_{\mathbf{Y}_t = \widehat{\mathbf{Y}}_t} \tag{7.82}$$

Multivariate DSTMs (Sec. 7.4)

What if we want to model multiple spatio-temporal processes that are co-dependent (e.g. temperature, salinity, current speed in the ocean)?

$$\mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q}_\eta) \qquad (7.98)$$

$$\mathbf{Y}_t = (\mathbf{Y}_t^{(1)'} \ldots \mathbf{Y}_t^{(K)'})'$$

- The key is in reducing the dimensionality of this problem

Multivariate DSTMs: Reduced Rank Approach

$$\mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q}_\eta) \tag{7.98}$$

$$\mathbf{Y}_t = (\mathbf{Y}_t^{(1)'} \ \ldots \ \mathbf{Y}_t^{(K)'})'$$

$$\mathbf{Y}_t^{(k)} = \mathbf{\Phi}^{(k)}\boldsymbol{\alpha}_t^{(k)} + \boldsymbol{\nu}_t^{(k)}$$

- Use reduced rank representation of $\mathbf{\Phi}^{(k)}$
- Can choose covariance structure of $\boldsymbol{\nu}_t^{(k)}$ depending on approximation of $\mathbf{\Phi}^{(k)}$

Multivariate DSTMs: Modeling Via Common Processes

Now assume:

$$\mathbf{Y}_t = \int_{D_s} \mathbf{H}(s, x; \boldsymbol{\theta}) \boldsymbol{\alpha}_t(x)\ dx + \boldsymbol{\gamma}_t(s) \qquad (7.103)$$

$$\mathbf{Y}_t \equiv (\mathbf{Y}_t^{(1)}\ \dots\ \mathbf{Y}_t^{(K)})'$$

$$\boldsymbol{\alpha}_t(s) \equiv (\alpha^{(1)}(s),\ \dots,\ \alpha^{(J)}(s))'$$

$$\boldsymbol{\gamma}_t(s) \equiv (\gamma_t^{(1)}(s),\ \dots,\ \gamma_t^{(K)}(s))'$$

for kernel matrix $\mathbf{H}(\cdot, \cdot; \cdot)$

- Hence, we represent the K processes in terms of J processes with $J < K$
- If the $\alpha_t^{(i)}(\cdot)$ have simple structure (e.g. independent AR(1) processes with unit variances), and if $\mathbf{H}$ has assumed simple structure, this can be helpful

Multivariate DSTMs: Modeling Via Common Processes

Approximate the integral with sums:

$$\mathbf{Y}_t = \int_{D_s} \mathbf{H}(s, x; \boldsymbol{\theta})\boldsymbol{\alpha}_t(x)\ dx + \boldsymbol{\gamma}_t(s) \qquad (7.103)$$

$$\Rightarrow Y_t^{(k)}(s) \approx \sum_{i=1}^{p_\alpha} \sum_{j=1}^{J} h^{(kj}(s, x_i; \boldsymbol{\theta})\alpha_t^{(j)}(x_i) \qquad (7.104)$$

$$\Rightarrow \mathbf{Y}_t(s) \approx \mathbf{H}(s)\boldsymbol{\alpha}_t + \boldsymbol{\gamma}_t(s) \qquad (7.105)$$

for $s, x_1, x_2, ..., x_{p_\alpha} \in D_s$.

Questions?