Spatio-Temporal Statistics

Noel Cressie*

Program in Spatial Statistics and Environmental Statistics

The Ohio State University

Christopher K. Wikle

Department of Statistics

University of Missouri, Columbia

Slides are based on the book, *"Statistics for Spatio-Temporal Data"*

by Cressie and Wikle, 2011, Wiley, Hoboken, NJ

*ncressie@stat.osu.edu

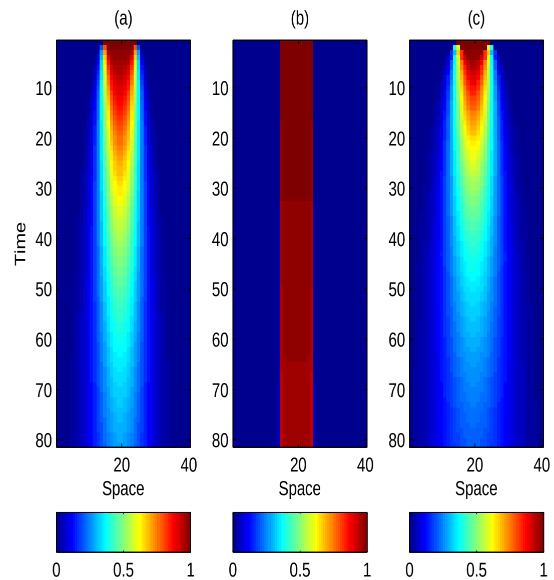
Spatio-Temporal Statistics

- There is no history without geography (and v.v.). We consider space and time together
- The dynamical evolution (time dimension) of spatial processes means that we are able to reach more forecefully for the "Why" question. The problems are clearest when there is **no aggregation**; henceforth consider processes at **point-level support**
- Consider the deterministic **1-D space** $\times$ **time**, reaction-diffusion equation:

$$\frac{\partial Y(s;t)}{\partial t} = \beta \frac{\partial^2 Y(s;t)}{\partial s^2} - \alpha Y(s;t);$$

β is the diffusion coefficient

Reaction-Diffusion Plots



$$Y(s,0) = I(15 \leq s \leq 24)$$

(a) $\alpha = 1$, $\beta = 20$; (b) $\alpha = 0.05$, $\beta = 0.05$; (c) $\alpha = 1$, $\beta = 50$

Stochastic Version

Consider the **stochastic** PDE:

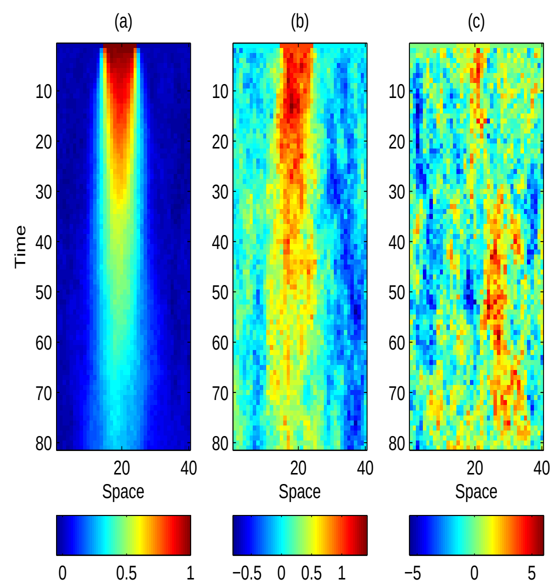
$$\frac{\partial Y}{\partial t} - \beta \frac{\partial^2 Y}{\partial s^2} + \alpha Y = \delta \,,$$

where $\{\delta(s;t) \colon s \in \mathbb{R}, t \geq 0\}$ is a zero-mean random process. Here we assume **white noise** for δ:

$$E(\delta(s;t)) \equiv 0$$

$$\text{cov}(\delta(s;t), \delta(u;r)) = \sigma^2 I(s = u, t = r)$$

Stochastic Reaction-Diffusion Plots



$$Y(s,0) = I(15 \leq s \leq 24)$$

$$\alpha = 1\,, \beta = 20$$

(a) $\sigma = 0.01$; (b) $\sigma = 0.1$; (c) $\sigma = 1$

Spatio-Temporal Covariance Function

- The stochastic reaction-diffusion equation implies a (stationary in space and time) **covariance function**:

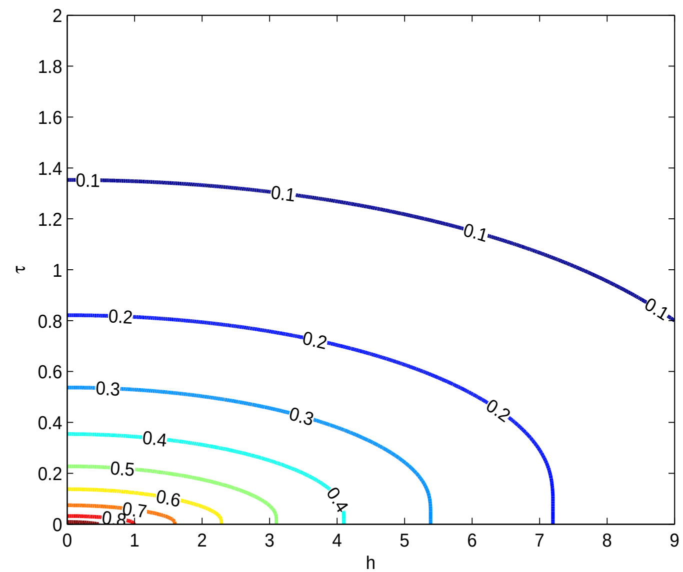
$$C(h;\tau) \equiv \mathrm{cov}(Y(s;t), Y(s+h;t+\tau))$$

and **correlation function**:

$$\rho(h;\tau) \equiv C(h;\tau)/C(0;0)$$

Heine (1955) *Biometrika*, gives a **closed-form** solution for $\rho(\cdot;\cdot)$

Contour Plot of Spatio-Temporal Correlation Function



$\rho(h;\tau)$ for the **stochastic reaction-diffusion equation**

Separability of Spatio-Temporal Covariance Functions

- Stochastic PDEs are built from dynamical physical considerations and they **imply** covariance functions
- Covariance functions have to be **positive-definite (p-d)**. So, specifying classes of spatio-temporal covariance functions to **describe** the dependence in spatio-temporal data is not all that easy
- Suppose the **spatial** $C^{(1)}(\mathbf{h})$ is p-d and the **temporal** $C^{(2)}(\tau)$ is p-d. Then the **separable** class:

$$C(\mathbf{h};\tau) \equiv C^{(1)}(\mathbf{h}) \cdot C^{(2)}(\tau)$$

  is guaranteed to be p-d
- **Separability is unusual in dynamical models**; it says that temporal evolution proceeds independently at each spatial location

Stochastic Reaction-Diffusion and Separability

- **If $C(h;\tau) = C^{(1)}(h) \cdot C^{(2)}(\tau)$, then**

$$C(h;0) = C^{(1)}(h)C^{(2)}(0)$$
$$C(0;\tau) = C^{(1)}(0)C^{(2)}(\tau)\,,$$

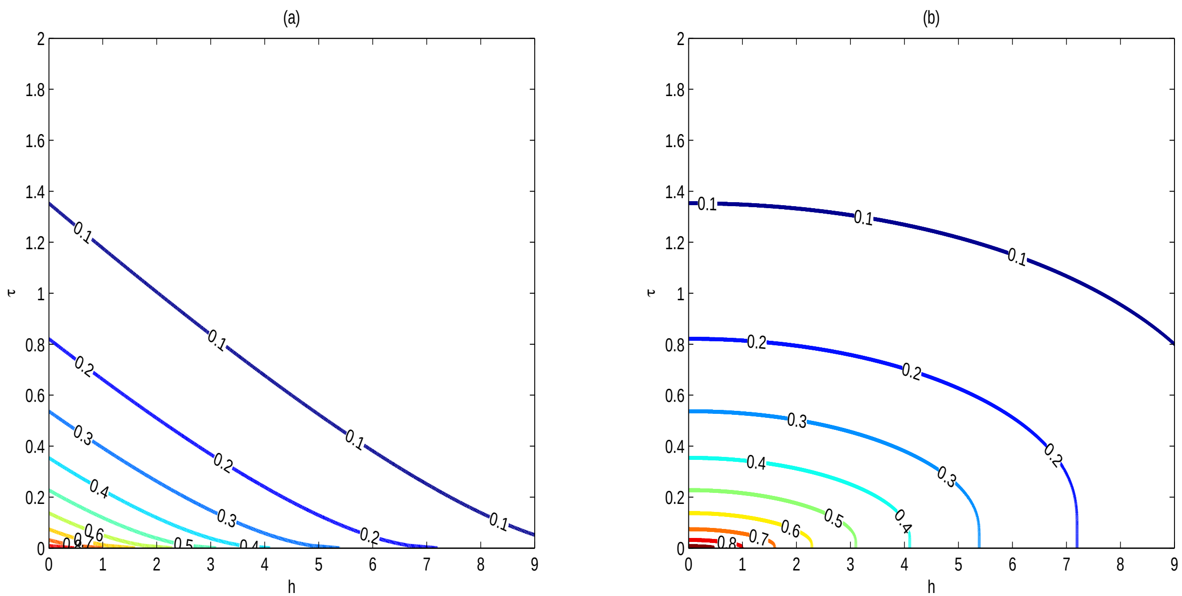
and hence

$$\rho(h;\tau) = \frac{C^{(1)}(h) \cdot C^{(2)}(\tau)}{C(0;0)} = \frac{C(h;0) \cdot C(0;\tau)}{C(0;0) \cdot C(0;0)} = \rho(h;0) \cdot \rho(0;\tau)$$

- Is this true for the stochastic reaction-diffusion equation? Plot

$$\rho(h;0) \cdot \rho(0;\tau) \quad \text{versus} \quad (h,\tau)$$
$$\rho(h;\tau) \quad \text{versus} \quad (h,\tau)$$

Contour Plots of Spatio-Temporal Correlation Functions

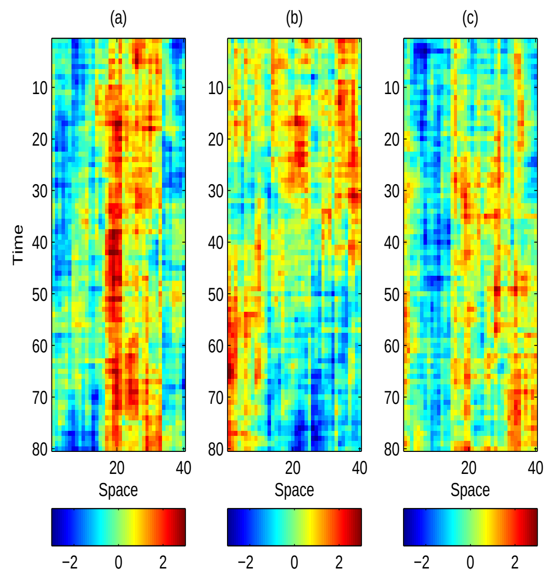


(a) $\rho(h;0) \cdot \rho(0;\tau)$; (b) $\rho(h;\tau)$

The **difference** in correlation functions is **striking**. Hence $\rho(\cdot;\cdot)$ is **not separable**. Can we see the difference between separability and non-separability in their **realizations**?

Non-Separable Realizations in Space-Time

Three realizations of $Y(s;t)$

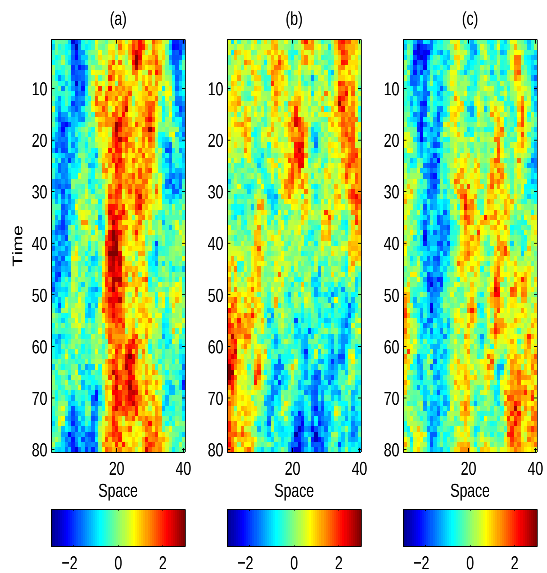


Realizations are generated from a stationary Gaussian process with the

non-separable, reaction-diffusion correlation function, $\rho(h;\tau)$

Separable Realizations in Space-Time

Three realizations of $Y(s;t)$



Realizations are generated from a stationary Gaussian process with **separable** correlation function, $\rho(h;0) \cdot \rho(0;\tau)$

Inference on a Hidden Spatio-Temporal Process

- We could ignore the dynamics and treat time as another "spatial" dimension. Write the **data** as:

$$\mathbf{Z} = (Z(\mathbf{s}_1; t_1), \ldots, Z(\mathbf{s}_n; t_n))' \,,$$

  which are observations taken at known space-time "locations". The data are noisy and not observed at all locations of interest

- Assume a **hidden ("true") process** $\{Y(\mathbf{s};t) \colon \mathbf{s} \in D \subset \mathbb{R}^d \,;\; t \geq 0\}$, which is not observable due to measurement error and missingness. Write

$$\mathbf{Z} = \mathbf{Y} + \boldsymbol{\varepsilon} \,,$$

  where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, var$(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 I$. We wish to **predict** $Y(\mathbf{s}_0; t_0)$ **from data Z**

Spatio-Temporal Kriging

Predict $Y(\mathbf{s}_0; t_0)$ with the linear predictor $\boldsymbol{\lambda}'\mathbf{Z}$:
For simplicity, assume $E(Y(\mathbf{s}; t)) \equiv 0$. Then minimize w.r.t. $\boldsymbol{\lambda}$, the **mean squared prediction error**,

$$E(Y(\mathbf{s}_0; t_0) - \boldsymbol{\lambda}'\mathbf{Z})^2 \,.$$

This results in the simple **kriging predictor**:

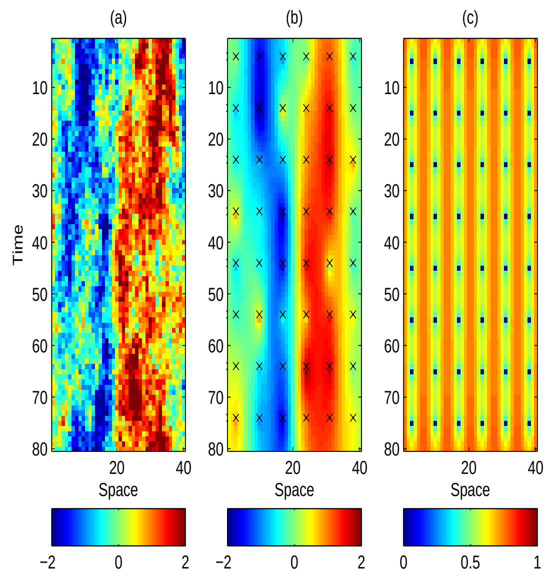
$$\widehat{Y}(\mathbf{s}_0; t_0) = \mathbf{c}(\mathbf{s}_0; t_0)'\Sigma_Z^{-1}\mathbf{Z}\,,$$

where $\Sigma_Z \equiv \text{var}(\mathbf{Z})$ and $\mathbf{c}(\mathbf{s}_0; t_0) = \text{cov}(Y(\mathbf{s}_0; t_0), \mathbf{Z})$

The simple **kriging standard error (s.e.)** is:

$$\sigma_k(\mathbf{s}_0; t_0) = \{\text{var}(Y(\mathbf{s}_0; t_0)) - \mathbf{c}(\mathbf{s}_0; t_0)'\Sigma_Z^{-1}\mathbf{c}(\mathbf{s}_0; t_0)\}^{1/2}$$

Kriging for the Stochastic Reaction-Diffusion Equation

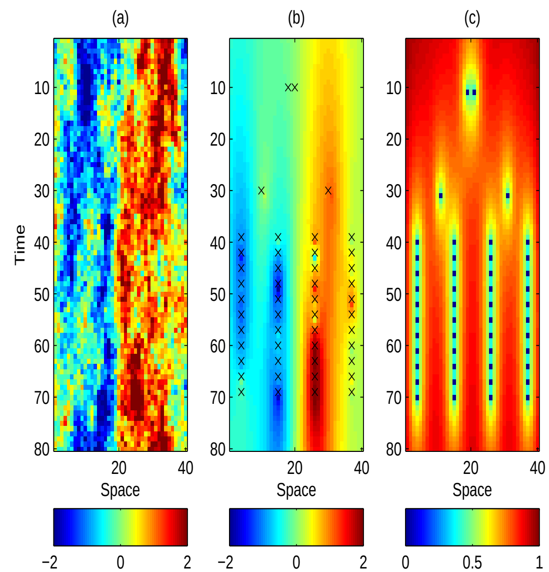


(a) Full realization; $\boldsymbol{\varepsilon} = \mathbf{0}$

(b) **Crosses** show $\{(\mathbf{s}_i; t_i)\}$ superimposed on the **kriging predictor** map, $\{\widehat{Y}(\mathbf{s}_0; t_0)\}$

(c) **Kriging s.e.** map, $\{\sigma_k(\mathbf{s}_0; t_0)\}$

Kriging for the Stochastic Reaction-Diffusion Equation, ctd.



(a) Same full realization; $\boldsymbol{\varepsilon} = \mathbf{0}$

(b) **Crosses** show **different** $\{(\mathbf{s}_i; t_i)\}$ superimposed on the **kriging predictor** map, $\{\widehat{Y}(\mathbf{s}_0; t_0)\}$

(c) **Kriging s.e.** map, $\{\sigma_k(\mathbf{s}_0; t_0)\}$

Emphasize the Dynamics

Approximate the **differentials** in the reaction-diffusion equation:

$$\frac{\partial Y}{\partial t} = \beta \frac{\partial^2 Y}{\partial s^2} - \alpha Y$$

with **differences**:

$$\frac{Y(s;t+\Delta_t) - Y(s;t)}{\Delta_t} = \beta \left\{ \frac{Y(s+\Delta_s;t) - 2Y(s;t) + Y(s-\Delta_s;t)}{\Delta_s^2} \right\} - \alpha Y(s;t)$$

Define $\mathbf{Y}_t \equiv (Y(\Delta_s;t), \ldots, Y(79-\Delta_s;t); \mathbf{Y}_t^B \equiv (Y(0;t), Y(79;t))'$. Then the **stochastic** version of the **difference equation** is:

$$\mathbf{Y}_{t+\Delta_t} = M\mathbf{Y}_t + M_B \mathbf{Y}_t^B + \boldsymbol{\delta}_{t+\Delta_t} \ ,$$

where $M_B \mathbf{Y}_t^B$ represents given boundary effects. The difference equation is a good approximation to the differential equation, provided $\alpha\Delta_t < 1$ and $2\beta\Delta_t/\Delta_s^2 < 1$

Emphasize the Dynamics, ctd.

Importantly,

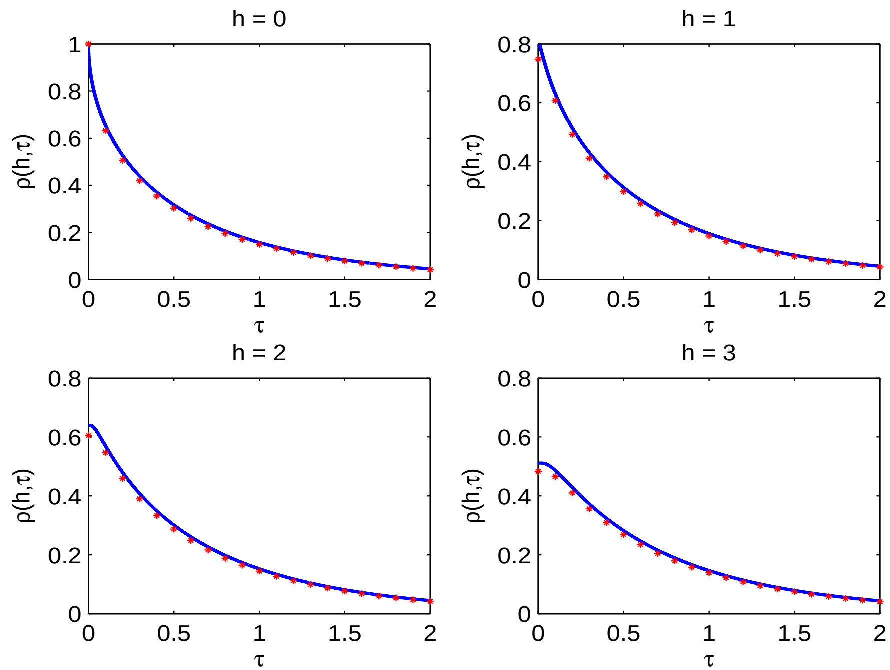
$$M = \begin{bmatrix} \theta_1 & \theta_2 & 0 & \dots & 0 \\ \theta_2 & \theta_1 & \theta_2 & \dots & \vdots \\ 0 & \theta_2 & \theta_1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \theta_2 \\ 0 & 0 & \dots & \theta_2 & \theta_1 \end{bmatrix},$$

where $\theta_1 = (1 - \alpha\Delta_t - 2\beta\Delta_t/\Delta_s^2)$, $\theta_2 = \beta\Delta_t/\Delta_s^2$. This can be viewed as the propagator matrix of a VAR(1) process. The matrix *is defined* by the dynamics. In other words, in a model of spatio-temporal dependence, M **has structure** (and is **sparse**). Conditional on the boundary effects, we see that the **lagged covariances** are given by,

$$C_Y^{(m)} = M^m C_Y^{(0)}\,,$$

where $C_Y^{(m)} \equiv \text{cov}(\mathbf{Y}_t, \mathbf{Y}_{t+m\Delta_t})$; $m = 0, 1, 2, \dots$

Comparison of Differential and Difference Equations



Spatio-temporal correlations; $\alpha = 1$, $\beta = 20$, $\Delta_s = 1$, and $\Delta_t = 0.01$

Solid blue line: from **differential** equations

Red dots: from **difference** equations

The Dynamics in the Difference Equation

Think of a spatial process at time t rather than a spatio-temporal process. Call it the vector $\mathbf{Y}_t$. Then describe the dynamics by a VAR(1):

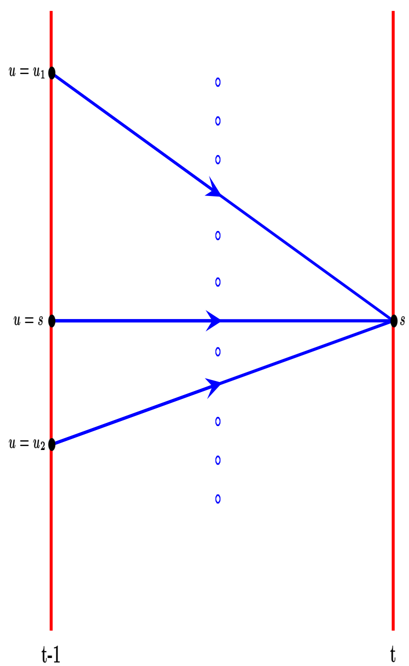
$$\mathbf{Y}_t = M\mathbf{Y}_{t-1} + \boldsymbol{\delta}_t$$

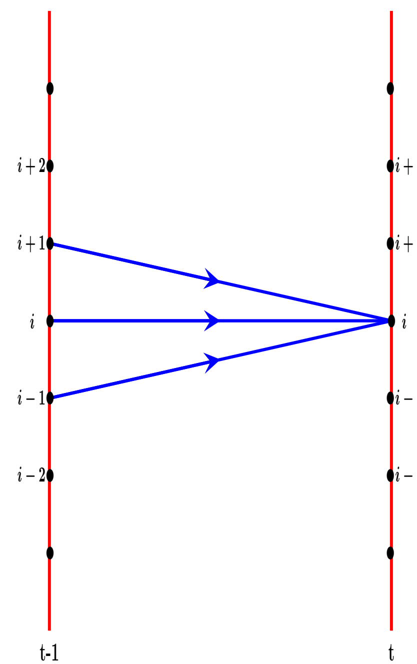
The choice of M is crucial. Define $M \equiv (m_{ij})$ **"spatially"**, that is, where the m_{ij} corresponding to **nearby** locations $\mathbf{s}_i$ and $\mathbf{s}_j$ are **non-zero**, and are zero when locations are far apart

This applies the **"First Law of Geography"** (Tobler; *cf.* Fisher and wheat yields) to the

dynamical evolution of the process

Structure of M





General M

M defined **"spatially"**

Instantaneous Spatial Dependence (ISD)

To capture the process' behavior at small temporal scales between time t and time $t+1$, we need a component of variation that is modeled as **instantaneous spatial dependence (ISD)**:

$$\mathbf{Y}_t = B_0 \mathbf{Y}_t + B_1 \mathbf{Y}_{t-1} + \boldsymbol{\nu}_t \; ,$$

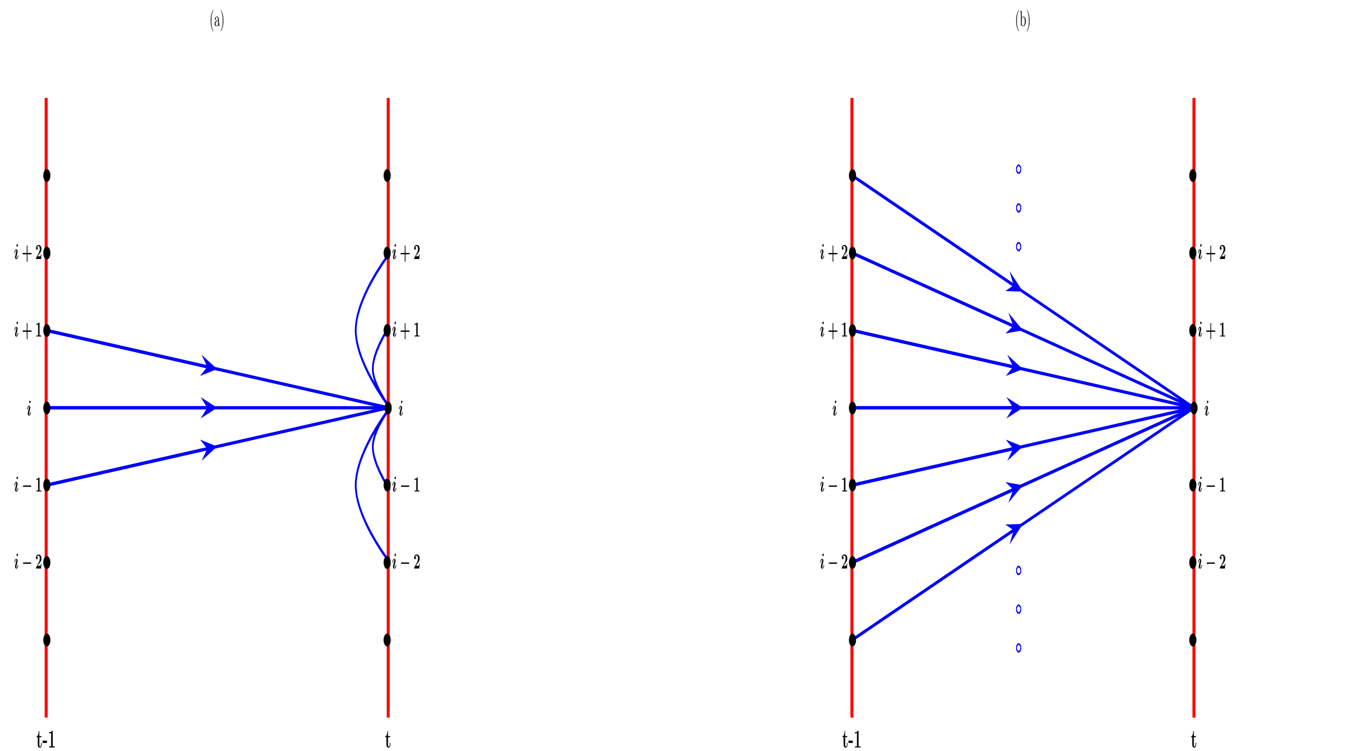
where B_0 has zero down its diagonal. Model B_0 and B_1 **"spatially"**; see the figure below. This implies

$$\mathbf{Y}_t = M \mathbf{Y}_{t-1} + \boldsymbol{\delta}_t \; ,$$

where $M = (I - B_0)^{-1} B_1$ and $\boldsymbol{\delta}_t = (I - B_0)^{-1} \boldsymbol{\nu}_t$. What use are B_0 and B_1? They

have **dynamic structure** and are **sparse**!

ISD in Graphical Form



(a) Graph structure **(sparse)** showing relationships that are defined “spatially”

(b) Equivalent directed graph structure (**non-sparse**)

Nonstationarity

Stationarity can be an unrealistic assumption. **Descriptive approaches** to spatio-temporal modeling, expressed in terms of covariance functions, almost demand it.

Dynamical approaches are much more forgiving. Consider the **nonstationary** VAR(1) process:

$$\mathbf{Y}_t = M_t \mathbf{Y}_{t-1} + \boldsymbol{\delta}_t$$

For example,

$$M_t = f(t) \cdot M \,,$$

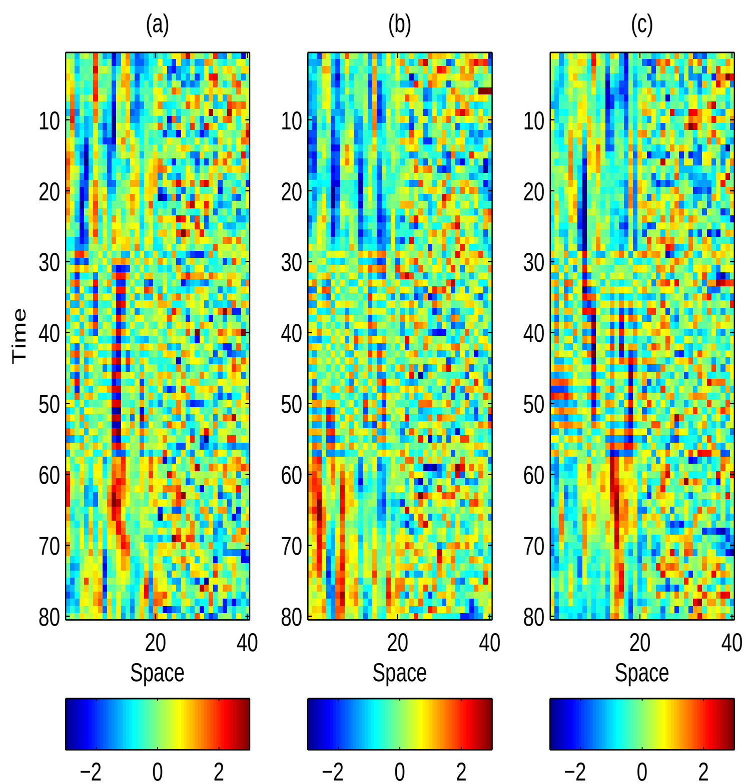
where

$$f(t) = \begin{cases} 1 & 0 \leq t \leq 29 \\ -1 & 30 \leq t \leq 59 \\ 1 & 60 \leq t \leq 79 \,, \end{cases}$$

and M is tridiagonal but has different parameters for $0 \leq s \leq 19$ and for $20 \leq s \leq 39$

Realizations for Nonstationary Process

Three realizations of $Y(s;t)$



$$\mathbf{Y}_t = M_t \mathbf{Y}_{t-1} + \boldsymbol{\delta}_t$$

Space-Time

... the **NEXT** frontier

(with apologies to Gene Roddenberry and Trekkies)