Dynamic Models for Spatio-Temporal Data

Jonathan Stroud
Department of Statistics - University of Chicago
Environmental Research Division - Argonne National Laboratory

Peter Müller
ISDS - Duke University

Bruno Sansó
CESMa - Universidad Simón Bolívar

Examples of Spatio-Temporal Data

- Rainfall Data
- Oceanographic Data
- Pollution Monitoring
- Satellite Data
- Disease Mapping
- Image Analysis
- EEG Data
- Real Estate Prices

Difficulties in Modeling Space-Time Data

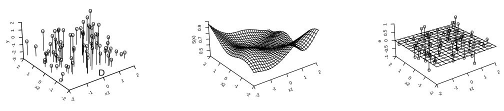
- Huge datasets (25GB)
- Unequally-spaced locations
- Stations that move over time
- Nonstationary components (trends, seasonality)
- MCMC often too slow!!

Methods/Applications

Authors	Method	Application
Carroll *et al.* (1997)	Separable Correlation	Hourly Ozone
Handcock & Wallis (1994)	Kriging	Global Warming
Haslett & Raftery (1989)	ARIMA/Kriging	Irish Wind Power
Niu & Tiao (1997)	STAR	Satellite Ozone
Meiring *et al.* (1994)	Space Deformation	Ground-Level Ozone
Higdon (1997)	Process Convolution	Ocean Temps.
Nychka, Wikle & Royle	Wavelet Bases	Rainfall
Wikle, Berliner & Cressie	Hierarchical Models	Atmospheric Data

Our Spatial Model

Data = Spatial Process + Noise

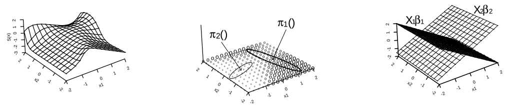


$$Y(x) = S(x,\boldsymbol{\beta}) + \epsilon(x)$$

- Sampling locations: $\boldsymbol{x} = (x_1, \ldots, x_n)$, $x_i \in D$.
- Observations at $\boldsymbol{x}$: $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, $Y_i \in \Re$.
- Gaussian noise: $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{V})$.
- Spatial process: $S(x, \boldsymbol{\beta}) \in \Re$.

Locally-Weighted Mixture Model

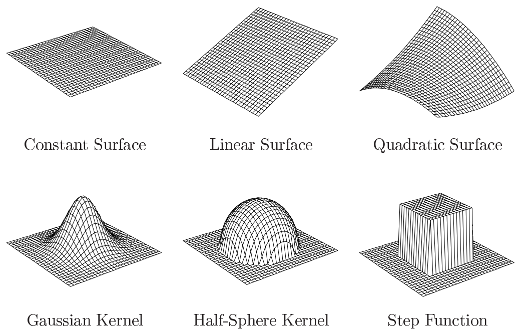
Spatial Process | Weighting Kernels | Component Surfaces



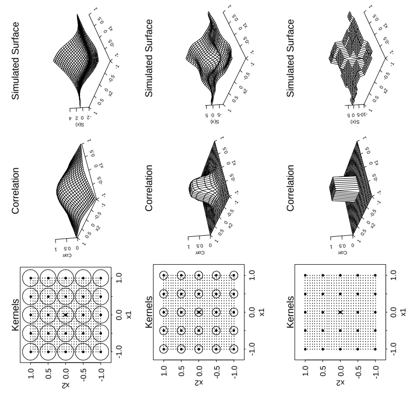
$$S(\boldsymbol{x},\boldsymbol{\beta}) = \sum_{j=1}^{2} \pi_j(\boldsymbol{x}) \times \boldsymbol{X}_j\boldsymbol{\beta}_j$$

- Weight function $\pi_j(\boldsymbol{x}) \propto N(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, $j = 1, 2$.
- Component design matrix $\boldsymbol{X}_j = [\mathbf{1}, \boldsymbol{x}_1, \boldsymbol{x}_2]$, $j = 1, 2$.
- Component regression vector $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \beta_{j2})$, $j = 1, 2$.
- Design matrix $\boldsymbol{X} = [\mathrm{diag}(\pi_1(\boldsymbol{x}))\boldsymbol{X}_1, \mathrm{diag}(\pi_2(\boldsymbol{x}))\boldsymbol{X}_2]$.
- Regression vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$.

Surface and Kernel Choice



Kernel Location and Size



Space-Time Model in State-Space Form

- Observation $\boldsymbol{Y}_t$: Data at $\boldsymbol{x}_t = (x_{t1}, \dots, x_{tn_t})$.
- State vector $\boldsymbol{\theta}_t$: Surface coefficients $\boldsymbol{\beta}_t$.
- Observation matrix $\boldsymbol{F}_t$: Locally-weighted design matrix $\boldsymbol{X}_t$.
- Observation variance $\boldsymbol{V}_t$: I.I.D. or Variogram-Based.
- Evolution matrix $\boldsymbol{G}_t$: Random walk, $\boldsymbol{G}_t = \boldsymbol{I}$.
- Evolution variance $\boldsymbol{W}_t$: Often defined through discount factor α:

$$\boldsymbol{W}_t = \alpha \ \mathrm{Var}(\boldsymbol{\theta}_{t-1} | \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{t-1}).$$

Forward-Filtering Backward-Sampling

Goal: Sample from smoothing distribution. $p(\boldsymbol{\theta}|Y^T)$.

$$p(\boldsymbol{\theta}|\boldsymbol{Y}^T) \propto \underbrace{p(\boldsymbol{\theta}_0) \prod_{t=1}^{T} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})}_{\text{prior}} \underbrace{\prod_{t=1}^{T} p(\boldsymbol{Y}_t|\boldsymbol{\theta}_t)}_{\text{likelihood}}$$

Method: Write posterior as

$$p(\boldsymbol{\theta}|\boldsymbol{Y}^T) = p(\boldsymbol{\theta}_T|\boldsymbol{Y}^T) \prod_{t=0}^{T-1} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{Y}^t)$$

Algorithm:

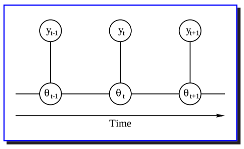
1. Compute moments of $p(\boldsymbol{\theta}_t|\boldsymbol{Y}^t)$, $\quad t = 1, \dots, T$.
2. Sample $\tilde{\boldsymbol{\theta}}_T$ from $p(\boldsymbol{\theta}_T|\boldsymbol{Y}^T)$.
3. Sample $\tilde{\boldsymbol{\theta}}_t$ from $p(\boldsymbol{\theta}_t|\tilde{\boldsymbol{\theta}}_{t+1}, \boldsymbol{Y}^t)$, $\quad t = T-1, \dots, 0$.

Gaussian State-Space Models

At times $t = 1, \dots, T$,

- Observation: $\boldsymbol{Y}_t = \boldsymbol{F}_t \boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{V}_t)$
- Evolution: $\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t)$
- Initial Prior: $\boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$

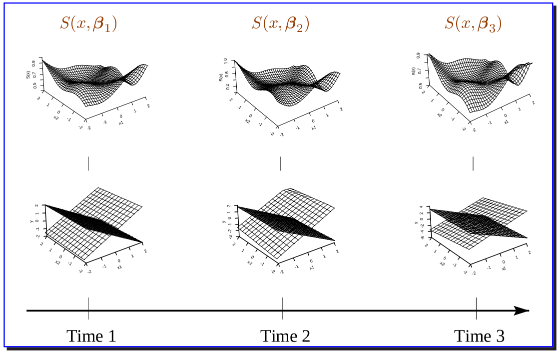
- $\boldsymbol{Y}_t$ Data
- $\boldsymbol{\theta}_t$ Unobserved State
- $\boldsymbol{F}_t$ Observation Matrix
- $\boldsymbol{V}_t$ Observation Variance
- $\boldsymbol{G}_t$ Evolution Matrix
- $\boldsymbol{W}_t$ Evolution Variance



Implementation Tools

Algorithm (Year)	Type of Inference
Kalman Filter (1960)	Recursive algorithm to compute moments.
Analytic Smoother (1962)	Smoothing analog of Kalman Filter.
FFBS (1994)	Efficient block simulation.
Single-State Gibbs (1992)	MCMC smoothing for nonlinear models.
Particle Filter (1993)	Sequential Importance Sampling.

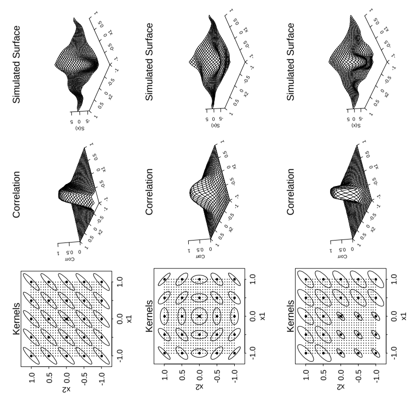
A Dynamic Spatial Process



Inference for Space-Time Models

Type	Distribution	Use
Filtering	$p(\boldsymbol{\theta}_t|\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_t)$	"Nowcasting"
Smoothing	$p(\boldsymbol{\theta}_t|\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_T)$	"Hindcasting"
Interpolation	$p(\boldsymbol{Y}_t^*|\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_T)$	"Kriging"
Prediction	$p(\boldsymbol{\theta}_s|\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_t)$, $s > t$	Predictive Maps
Forecasting	$p(\boldsymbol{Y}_s|\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_t)$, $s > t$	Point Forecasts

Kernel Size and Orientation



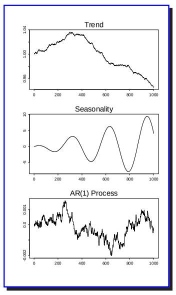
Additional Latent Components

Other possible components

- Temporal trends
- Seasonal effects
- AR(p) processes
- Regressions

Each component represented by a block

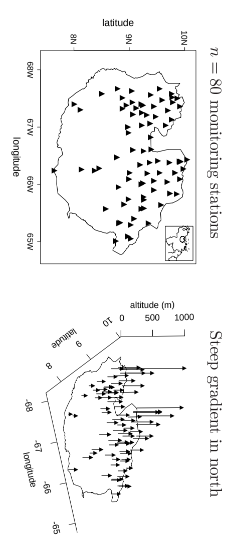
$$\begin{aligned}
\boldsymbol{\theta}_t &= (\boldsymbol{\beta}_t, \boldsymbol{\theta}_{t2}, \dots, \boldsymbol{\theta}_{tk}) \\
\boldsymbol{F}_t &= [\boldsymbol{X}, \boldsymbol{F}_{t2}, \dots, \boldsymbol{F}_{tk}] \\
\boldsymbol{G}_t &= \mathrm{BD}(\boldsymbol{I}, \boldsymbol{G}_{t2}, \dots, \boldsymbol{G}_{tk}) \\
\boldsymbol{\alpha} &= (\alpha, \alpha_2, \dots, \alpha_k)
\end{aligned}$$



Example: Guárico Rainfall Data

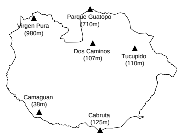
$n = 80$ monitoring stations

Steep gradient in north

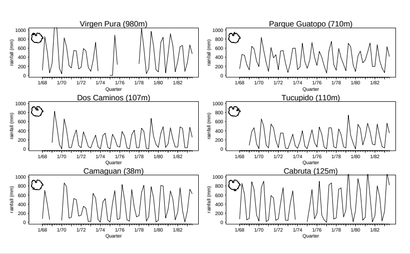


- 150km south of Carribean
- Located in Los Llanos (plains)
- Agricultural state
- Hills in north
- Flat savanna in south
- Water runoff into Orinoco River

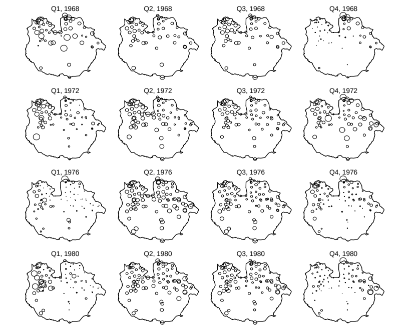
Seasonality/Annual Cycles



- Wet Season: May–Oct. (Q2–Q3).
- Dry Season: Nov.–Apr. (Q4–Q1).
- Rainfall varies from 0–1200mm.
- High Levels in north and south.
- Intertropical convergence effect. (S$\Longleftrightarrow$N)

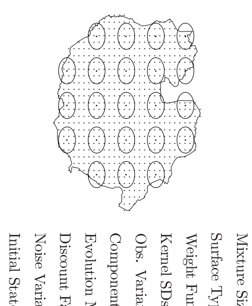


Space-Time Patterns



- High levels in northwest.
- Lowest levels in central plains.
- Parque Guatopo: high rainfall in dry season.

Model Summary

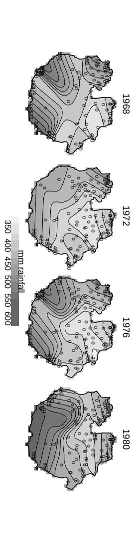


Mixture Size:	$J = 22$
Surface Types:	Constant
Weight Functions:	Gaussian Kernels
Kernel SDs:	0.4° lon/0.25° lat
Obs. Variance:	$\boldsymbol{V} = \sigma^2 \boldsymbol{I}$
Components:	Spatial/Seasonal
Evolution Matrix:	$\boldsymbol{G} = \mathrm{BD}(\boldsymbol{I}, \boldsymbol{G}^*)$
Discount Factor:	$\boldsymbol{\alpha} = (0.1, 0.05)$
Noise Variance:	$\sigma^2 \sim IG(1/2, 1000/2)$
Initial State:	$\boldsymbol{\theta}_0 \sim N(0, 100^2\boldsymbol{I})$

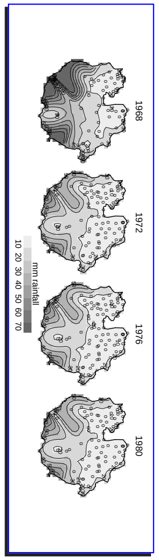
- Equally-spaced grid.
- East-west kernel orientation.

Guárico Rainfall Maps

Mean fields, 3rd quarter:

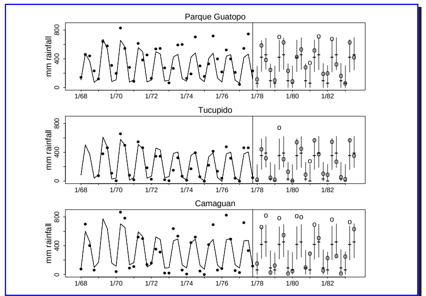


Uncertainty fields, 3rd quarter:



Filtering and Forecasting

Filtered means and 95% forecast intervals:



- Slightly underpredicts dry season at Parque Guatopo.
- Accurate predictions at Tucupido.
- Slightly underpredicts wet season in Camaguan.

Kriging/Interpolation

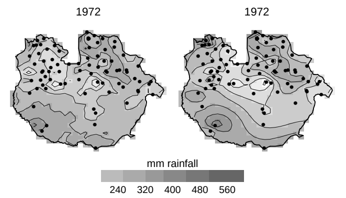
- Kriging Model:

$$\boldsymbol{Y}_t = \boldsymbol{S}(\boldsymbol{x}, \boldsymbol{\beta}_t) + \boldsymbol{g}(\boldsymbol{\gamma}_t) + \boldsymbol{\epsilon}_t, \ \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{V}(\boldsymbol{x})).$$

- Exponential Variogram:

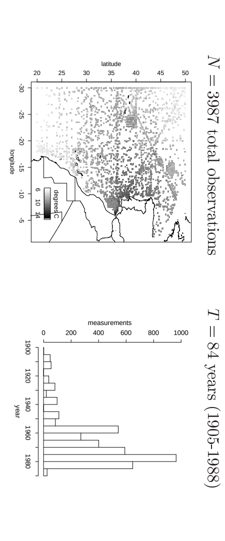
$$V_{ij} = \sigma^2 \exp(-d_{ij}/\lambda).$$

- Bayesian Kriging Distribution: $p(\boldsymbol{Y}_t^* | \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T)$.
- Two-step MCMC (FFBS $\Longleftrightarrow$ Metropolis).
- Kriged Mean Field (Quarter 3, 1972):



Example: Atlantic Water Temperatures

$N = 3987$ total observations $\quad T = 84$ years (1905-1988)



- Data collected by vessels
- Archived at NODC
- Data from constant density
- Warm water from Mediterranean.
- Uneven temporal distribution
- Goal: Produce temperature maps

The Kalman Filter

Goal: Sequential inference for Gaussian State-Space models.

Algorithm:

1. Initial step. Start with Gaussian initial distribution $p(\boldsymbol{\theta}_0)$.
2. Evolution step. Obtain $p(\boldsymbol{\theta}_1)$ by convolving $\boldsymbol{\theta}_0$ with $\boldsymbol{\omega}_1$.
3. Update step. Given $\boldsymbol{Y}_1$, obtain posterior $p(\boldsymbol{\theta}_1|\boldsymbol{Y}_1)$ (Bayes' theorem).
4. Repeat steps 2 and 3 for $t = 2, \ldots, T$.

Summary: Provides moments for filtering distribution.

The Analytic Smoother

Goal: Full posterior inference for GSS models.

Define: $\boldsymbol{Y}^t = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t)$.

Algorithm:

1. Initial step. Start with $p(\boldsymbol{\theta}_T|\boldsymbol{Y}^T) = N(\boldsymbol{a}_T, \boldsymbol{R}_T)$.
2. Backwards. Obtain $p(\boldsymbol{\theta}_t|\boldsymbol{Y}^T) = N(\boldsymbol{a}_t, \boldsymbol{R}_t)$, for $t = T-1, \ldots, 1$.

$$p(\boldsymbol{\theta}_t|\boldsymbol{Y}^T) = \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{Y}^T)\ \underbrace{p(\boldsymbol{\theta}_{t+1}|\boldsymbol{Y}^T)}_{\text{known}}\ d\boldsymbol{\theta}_{t+1}.$$

Compute moments:

- $\boldsymbol{a}_t = E[E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{Y}^T)]$
- $\boldsymbol{R}_t = V[E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{Y}^T)] + E[V(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{Y}^T)]$

The Particle Filter

Goal: Sequential inference for *general* state-space models.

- Observation: $\boldsymbol{Y}_t \sim p(\boldsymbol{Y}_t|\boldsymbol{\theta}_t)$
- Evolution: $\boldsymbol{\theta}_t \sim p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$
- Initial Prior: $\boldsymbol{\theta}_0 \sim p(\boldsymbol{\theta}_0)$

Algorithm:

1. Initial step. Generate $\boldsymbol{\Omega}_0 = \{\boldsymbol{\theta}_0^1, \ldots, \boldsymbol{\theta}_0^N\}$ from $p(\boldsymbol{\theta}_0)$.
2. Convolution step. Obtain $\boldsymbol{\Lambda}_1 = \{\boldsymbol{\theta}_1^1, \ldots, \boldsymbol{\theta}_1^N\}$ from $p(\boldsymbol{\theta}_1)$.
   (Add random noise to points in $\boldsymbol{\Omega}_0$.)
3. Update step. Given $\boldsymbol{Y}_1$, obtain $\boldsymbol{\Omega}_1 = \{\boldsymbol{\theta}_1^1, \ldots, \boldsymbol{\theta}_1^N\}$ from $p(\boldsymbol{\theta}_1|\boldsymbol{Y}_1)$.
   (Resample points from $\boldsymbol{\Lambda}_1$ with prob $\propto p(\boldsymbol{Y}_1|\boldsymbol{\theta}_1^i)$.)
4. Repeat steps 2 and 3 for times $t = 2, \ldots, T$.

Seasonal Effects

Single Harmonic. (Frequency $\gamma \in (0, \pi)$):

- $\boldsymbol{F} = [\mathbf{1}, \mathbf{0}]$
- $\boldsymbol{\theta}_t = (a_t, b_t)$
- $\boldsymbol{\alpha} = \alpha$
- $\boldsymbol{G}(\gamma) = \begin{pmatrix} \cos(\gamma) & \sin(\gamma) \\ -\sin(\gamma) & \cos(\gamma) \end{pmatrix}$

Multiple Harmonics. (Frequencies $\gamma_1, \ldots, \gamma_r \in (0, \pi)$):

- $\boldsymbol{F} = [\mathbf{1}, \mathbf{0}, \ldots, \mathbf{1}, \mathbf{0}]$
- $\boldsymbol{\theta}_t = (a_{t1}, b_{t1}, \ldots, a_{tr}, b_{tr})$
- $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$
- $\boldsymbol{G} = \mathrm{BD}(\boldsymbol{G}(\gamma_1), \ldots, \boldsymbol{G}(\gamma_r))$

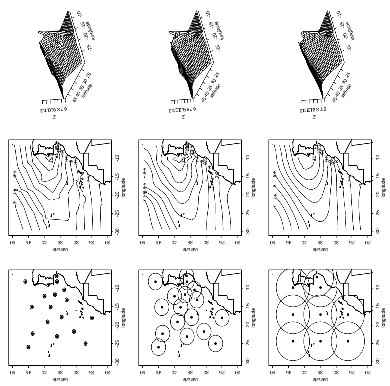
Model Summary

Mixture Size:	$J = 17$
Surface Types:	Linear
Weight Functions:	Gaussian kernels
Kernel SDs:	9° Lat/ 9° Lon
Obs. Variance:	$\boldsymbol{V} = \sigma^2\boldsymbol{I}$
Model Components:	Trend/Spatial
Evolution Matrix:	$\boldsymbol{G} = BD(1, \boldsymbol{I})$
Discount Factor:	$\alpha = c(0.1, 0.2)$
Noise Variance:	$\sigma^2 \sim IG(1/2, 1/200)$
Initial State:	$\theta_0 \sim N(0, 100^2\boldsymbol{I})$
CPU Time:	10 MINUTES

- More kernels in Med mouth.
- Linear Surfaces for flexibility.

Model Comparison

1980 Mean Fields



Extensions and Current Work

- Weight kernels based on wind velocity, rivers, or stream locations.
- Coupling with physical models.
- Physically-based $\boldsymbol{G}_t$ matrix (Lake Michigan chlorophyll).
- Fast particle filtering for non-conjugate models (Mexico City ozone monitoring).

Conclusions

- Simple, flexible models.
- No assumption of stationarity/isotropy.
- No assumption of space-time separability.
- Mean field/uncertainty maps.
- Probabilistic inference for model quantities.
- Computationally fast and efficient.

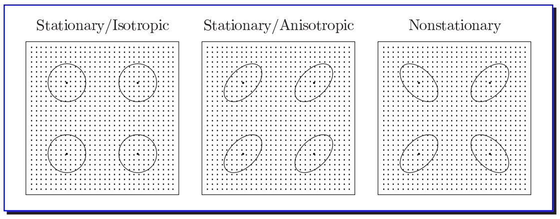
Stationarity and Isotropy

- Strict Stationarity:

$$p[Z(x_1),\dots,Z(x_n)] \;=\; p[Z(x_1+h),\dots,Z(x_n+h)], \quad \forall n, x_i, h$$

- Isotropy:

$$\mathrm{Cov}[Z(x_1),Z(x_2)] \;=\; K_\theta(||x_1-x_2||)$$



Autoregressive Components

First-order: AR(1).

- $\boldsymbol{F} = \mathbf{1}$
- $\boldsymbol{\theta}_t = z_t$
- $\boldsymbol{\alpha} = \alpha$
- $\boldsymbol{G} = \phi_1$

General case: AR(p).

- $\boldsymbol{F} = [\mathbf{1},\mathbf{0},\dots,\mathbf{0}]$
- $\boldsymbol{\theta}_t = (z_t,\dots,z_{t-p+1})$
- $\boldsymbol{\alpha} = (\alpha,0,\dots,0)$
- $\boldsymbol{G} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$

Polynomial Trends

0th order – local level:

- $\boldsymbol{F} = \mathbf{1}$
- $\boldsymbol{\theta}_t = f_t^{(0)}$
- $\boldsymbol{\alpha} = \alpha$
- $\boldsymbol{G} = 1$

$p-1$st order:

- $\boldsymbol{F} = [\mathbf{1},\mathbf{0},\dots,\mathbf{0}]$
- $\boldsymbol{\theta}_t = (f_t^{(0)}, f_t^{(1)},\dots,f_t^{(p-1)})$
- $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_p)$
- $\boldsymbol{G} = \begin{pmatrix} 1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$