An Overview of Models and Methods for Spatio–temporal Data Analysis

Jim Zidek- *

*U British Columbia, Vancouver, Canada

May 30, 2012

Outline

1. Introduction
2. Processes
   - Temporal
   - Spatial: spatial; lattice (areal); point
   - Spatio-temporal
3. Wrap-up

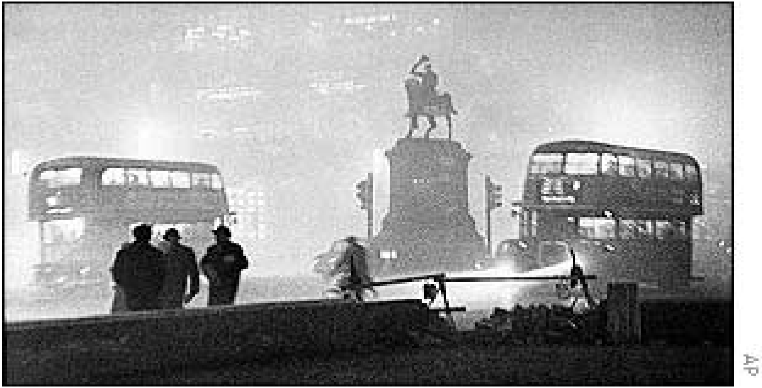
1 Introduction

1.1 London fog

1952: The most infamous environmental space-time field.

1.2 London fog

The most (in-) famous example



1.3 London fog

Barbara Fewster recalls her 16-mile walk home - in heels - guiding her fiancé's car"

> *"It was the worst fog that I'd ever encountered. It had a yellow tinge & a strong, strong smell strongly of sulphur, because it was really pollution from coal fires that had built up. Even in daylight, it was a ghastly yellow colour.*

1.4 London fog

Figure 12-10. December 1962, London pollution episode.

1.5 Ensuing developments

1952...: Environmental cleanup begins in Britain

1970: USA's Clean Air Act

1971: USA EPA formed

1973: First SIMS group set up; Stanford & Paul Switzer + others

1980s: Acid rain

1990s: Air pollution

2000s: Climate change

2010s: Environmental risk management

- Agroclimate risk management; crop yields; phenological events.
- Long term monitoring; lumber properties; forest fires
- Water quality and quantity

1.6 Current directions

- **Uncertainty quanitification**
  - Combining physical & statistical modeling
- **High dimensional random response vectors**
  - Eg. At 1000s of spatial sites
  - Methods like MCMC don't work
  - INLA - Laplace approximation under active development
- **Model–based geostatistics**
- **Multivariate extreme value theory** for high dimensions
- **Nonstationary spatio - temporal covariance structures**
- **Design** of monitoring networks
- **Spatio-temporal point processes**
- **Preferential sampling** & network design

1.7 ST modeling applications

- Relationship between **deaths & atmospheric particulate concentrations** [e.g. London Fog]
- **Climate modeling** - 1000s of sites for temperature or precipation
- Location, location, location: **house prices**
- **Used car prices**
- Strain gauges on the **space station**
- **Fires** in tall wooden buildings
- **Lightning strikes** & forest fires
- **Acid rain**

1.8 ST modeling: General approach

Hierarchical modeling:

- Measurement model
- Process model
- Parameter model

1.9 ST modeling: General approach

Hierarchical modeling: Alternate formulation; **[X] = distribution of X**

- $[measurement|process, parameters]$
- $[process|parameters]$
- $[parameters]$

1.10 ST modeling: General data categories

Time - usually discrete index, $t = 1, \ldots, T$.
Spatial locations indexed by $s \in D$.

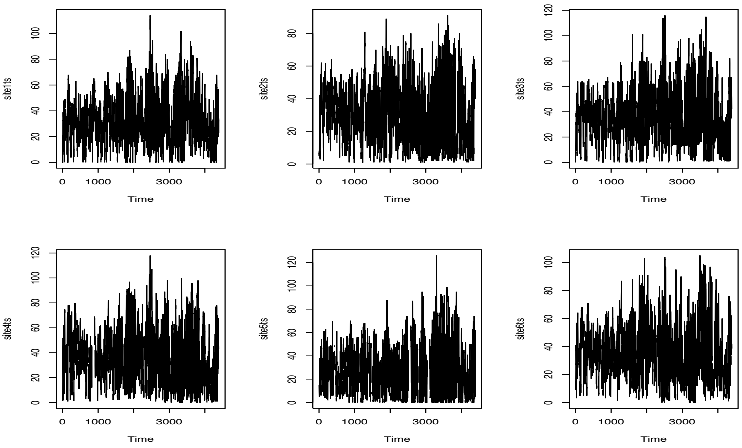
- **Point referenced data:** D = continuum or dense spatial grid; measurements made at irregular network of locations.
  E.g: ozone field
- **Lattice processes:** D = not necessarily regular grid of areal regions or specified locations D where meaurements are made.
  E.g: death counts per county; centroids = lattice points
- **Point processes:** Measurements or “marks”. made at randomly selected points in continuum D
  E.g: lightning strikes

Selected references: [Schabenberger and Gotway, 2005], [Le and Zidek, 2006], [Banerjee et al., 2003], [Cressie and Wikle, 2011]

2. Temporal processes

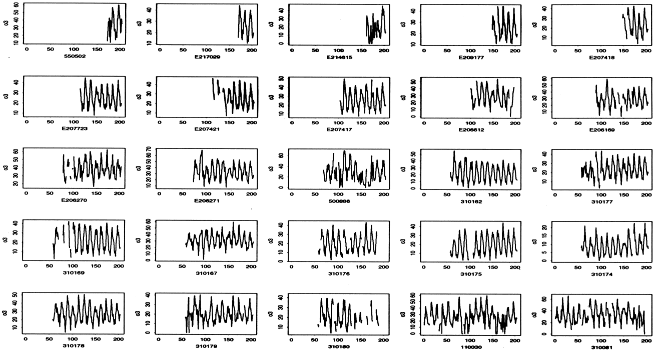
2.1 Example- ozone fields in US

Time series plots: Hourly concentrations at 6 O3 monitoring sites, Eastern USA Note **24 hour cycles.**



2.2 Example - ozone fields in BC

Time series plots: Monthly measurements at 25 O3 sites in BC. Note **seasonality** and different start dates.



2.3 Examples - lessons learned

- Monitoring start times different - staircase pattern in monitoring data
- Systematic patterns across space - trends, seasonality, daily cycles

2.4 Autoregressive models

AR(1) process. For time t & fixed spatial location s

$$X(s,t) = \alpha X(s,t-1) + W(s,t),\ t = 1, \dots,$$

Here $\alpha = corr[X(s,t), X(s,t-1)]$ for all t (stationary process);
$\{W(t,s)\}$ iid zero mean sequence
Multivariate version MAR(1).

$$\mathbf{X}(s,t) = \boldsymbol{\alpha}\mathbf{X}(s,t-1) + \mathbf{W}(s,t),\ t = 1, \dots,$$

2.5 Dynamic linear models

Generalize the AR process. At fixed spatial location s
measurement model:

$$X(s,t) = F_t \beta(s,t) + \epsilon(s,t),\ \epsilon(t,s) \sim N(0,V)$$

process model:

$$\beta(s,t) = G_t \beta(s,t-1) + \omega(t,s), \omega(s,t) \sim N(0,W)$$

parameter model:

$$[\beta(0,s), V, W]$$

2.6 Beyond linearity

Approaches to nonlinearity:

- Nonlinearize linear models e.g. with link functions.
- Purpose build them from ”ground - up”
  - Next few slides illustrate this approach

2.7 Markov chain models

Time series of binary outcomes. **Theorem:** Hosseini et al. [2011b]. For $X(s,t) \in \{0,1\}$ an r-th order Markov chain & g arbitrary, monotone, then uniquely:

$$g^{-1}\{\frac{P(X(s,t)=1|X(s,t-1),\cdots,X(s,0)}{P(X(s,t)=0|X(s,t-1),\cdots,X(s,0))}\} =$$

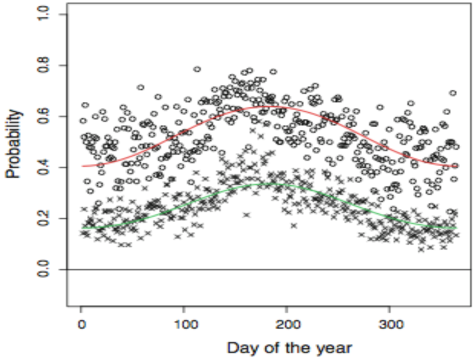
$$\alpha_0^t + \sum_{i=1}^{r} X(s,t-i)\alpha_i^t + \cdots +$$

$$\sum_{1\le i_1<i_2<\cdots<i_k\le r} \alpha_{i_1,\cdots,i_k}^t X(s,t-i_1)\cdots X(s,t-i_k) + \cdots +$$

$$\alpha_{12\cdots r}^t X(s,t-1)X(s,t-2)\cdots X(s,t-r).$$

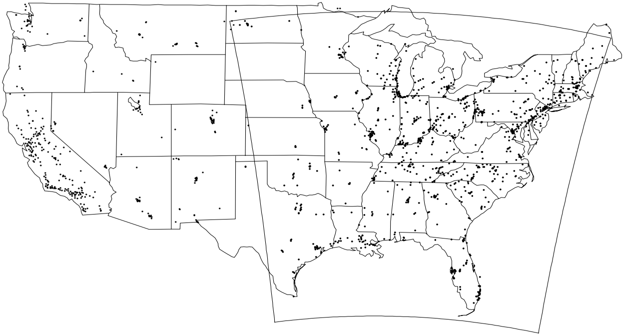
2.8 Application: Markov chain models

Canadian Prairie droughts: **Agroclimate risk management** needs stochastic models for non-precipitation days ($X = 0$). Model as Markov chain. Resulting one step transition model fits to empirical data [Hosseini et al., 2011a] for **Calgary**. Top curve (red) is for precip yesterday = 1.



3. Point referenced processes

3.1 Example: US Ozone monitoring sites



3.2 Moments and variograms

$X \sim F$: random vector field. (Fixed time t omitted in sequel).
For locations $\{s_1, \ldots, s_g\}$ for any g

$$F_{s_1,\ldots,s_g}(x_1, \ldots, x_g) \equiv P\{X(s_1) \leq x_1, \ldots, X(s_g) \leq x_g\}.$$

$F_{s_1,\ldots,s_g}(x)$ is joint distribution distribution (DF)

- **Moment** of k^{th}*-order:*

$$E[X(s)]^k \equiv \int x^k dF_s(x)$$

- **Expectation**: If exists, defined as the 1^{st}-order moment for any s

$$\mu(s) \equiv E[X(s)]$$

- **Variance:**

$$Var[X(s)] \equiv E[X(s) - \mu(s)]^2.$$

- **Covariance** between locations s_1 & s_2,

$$C(s_1, s_2) \equiv E[(X(s_1) - \mu(s_1))(X(s_2) - \mu(s_2))]$$

  - **NOTE:** $C(s_1, s_1) \equiv Var[X(s_1)]$

- **Variogram:** Between any 2 locations, s_1 & s_2:

$$\begin{aligned} 2\gamma(s_1, s_2) &\equiv var[X(s_1) - X(s_2)] \\ &= E[X(s_1) - X(s_2) - (\mu(s_1) - \mu(s_2))]^2. \end{aligned}$$

  - $\gamma(s_1, s_2)$ is called *semi-variogram*.

3.3. Stationarity

An important concept in characterizing the random field Y

- **Strict stationarity**
  X *strictly stationary* if:

$$F_{s_1,\ldots,s_n}(x) = F_{s_1+h,\ldots,s_n+h}(x)$$

  for any vector h & an arbitrary n

- **Second–order stationarity**
  X is *second-order stationary* if:

$$\begin{array}{lclcl} \mu(s) & = & E[X(s)] & = & \mu \\ C(s, s+h) & = & C(s+h-s) & = & C(h) \end{array}$$

  - when $h = 0:\ \ Var[X(s)] \ = \ C(s,s) \ = \ C(0)$
    ie. ***Mean, Variance do not depend on location***

- **Second–order stationarity - cont'd**
  - $C(h)$: *covariogram* (or *autocovariance* in time series)
  - Implies ***Intrinsic Stationarity*** *(weaker)*

$$\begin{aligned} Var[X(s) - X(s+h)] &= Var[X(s)] + Var[X(s+h)] \\ &\quad -2Cov[X(s), X(s+h)] \\ &= C(0) + C(0) - 2C(h) \\ &= 2[C(0) - C(h)]. \end{aligned}$$

or equivalently semi-variogram

$$\gamma(h) = C(0) - C(h).$$

3.4 Properties of C(h)

X second-order stationary process with covariance function $C(h)$.

- **Positive Definiteness (PD):** If $\Sigma = \{C(h_{ij})\}$ being covariance matrix of random vector $(X(s_1), \ldots, X(s_n))$ makes it PD implying for any vector a that:

$$\sum_i \sum_j a_i a_j C(h_{ij}) > 0$$

- **Anisotropy**: $C(h)$ - function of length & direction
- **Isotropy**: $C(h)$ - function only of length $|h|$

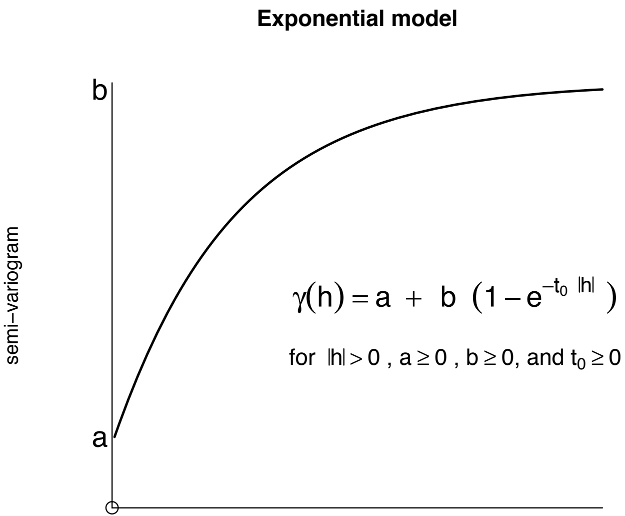
3.5 Isotropic Semi-Variogram Models

Second order stationarity implies $\gamma(h) = C(0) - C(h) \rightarrow \gamma(0) = 0$

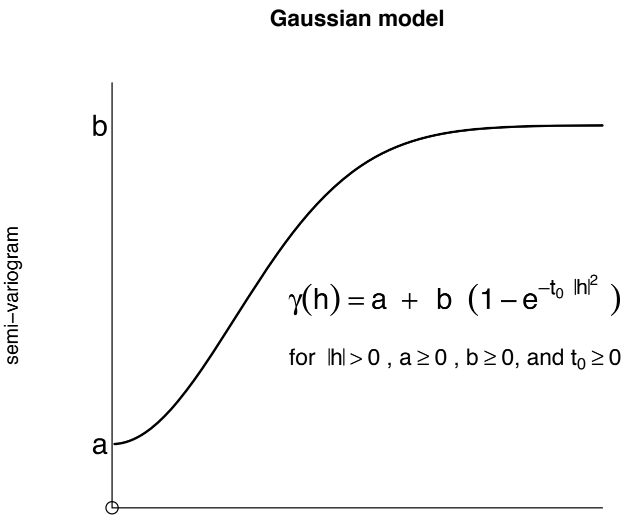
- But often $\lim_{h\rightarrow 0} \gamma(h) \neq 0$. Discontinuity called ***nugget effect***.
- When $\gamma(h) \rightarrow B$ as $h \rightarrow \infty$, B called a ***sill***

Note: Few functions satisfy positive definiteness condition - only certain ones (eg. variogram)

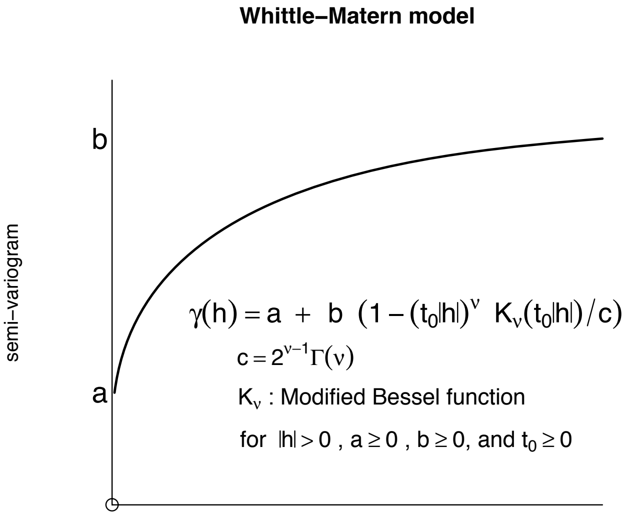
3.6 Common isotropic models



3.7 Common isotropic models

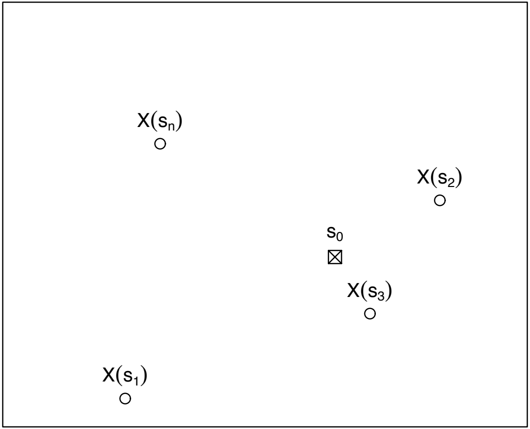


3.8 Common isotropic models



3.9 Spatial prediction

Problem: Estimate at location s_0 given observed levels $X(s_i)$?



3.10 Ordinary Kriging

Problem: Predict $X(s_0)$ given observations $x_1, \ldots, x_n$ at locations $s_1, \ldots, s_n$

- Assume $X(s) = \mu + Z(s)$ - intrinsic stationary, ie.

$$\begin{aligned} E[X(s)] &= \mu \\ Var[X(s) - X(s+h)] &= 2\gamma(|h|) \end{aligned}$$

- Kriging Predictor $X^*(s_0) = \sum_{i=1}^{n} \alpha_i X(s_i)$

  Choose the $\{\alpha\}$ to get unbiasedness and minimum prediction error, $\sigma^2_{s_0} \equiv E\left[X^*(s_0) - X(s_0)\right]^2$

Kriging predictor: Best linear unbiased predictor (BLUP)

References: [Krige, 1951] & [Matheron, 1963]

3.11 Ordinary Kriging system

- $E[X^*(s_0)] = E\left[\sum_{i=1}^n \alpha_i X(s_i)\right] = \mu \sum_{i=1}^n \alpha_i \qquad (1)$
  Thus $\sum_{i=1}^n \alpha_i = 1$ required.
- Prediction error (Kriging variance)

$$
\begin{aligned}
\sigma_{s_0}^2 &\equiv E\left[X^*(s_0) - X(s_0)\right]^2 = E\left[\sum_{i=1}^n \alpha_i (X(s_i) - X(s_0))\right]^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j E[X(s_i) - X(s_j)]^2/2 \\
&\quad - \sum_{i=1}^n \alpha_i E[X(s_i) - X(s_0)]^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma(|h_{ij}|) - 2\sum_{i=1}^n \alpha_i \gamma(|h_{i0}|) \qquad (2)
\end{aligned}
$$

α's chosen to minimize (2) & satisfy (1)

3.12 Ordinary Kriging System

- Solution for α's:

$$\begin{cases} \partial f/\partial \alpha_i = 0 & i = 1, \ldots, n \\ \partial f/\partial \lambda = 0 & \end{cases}$$

where $f(\alpha_1, \ldots, \alpha_n, \lambda) = \sigma^2_{s_0} + 2\lambda\left(\sum_{i=1}^n \alpha_i - 1\right)$

- $\Longrightarrow$ *ordinary Kriging system*

$$\begin{cases} \sum_{j=1}^n \alpha_j \gamma(|h_{ij}|) + \lambda = \gamma(|h_{i0}|) \\ \sum_{j=1}^n \alpha_j = 1 \end{cases}$$

for $i = 1, \ldots, n$; h_{ij}: distance between s_i & s_j

3.13 Implementation

- Select suitable semi-variogram model & estimate $\hat{\gamma}(.)$ using the data
- Solve the *Kriging system* to obtain $\hat{\alpha}$'s
- **Kriging interpolator** & estimated **Kriging variance**

$$\begin{aligned} \hat{X}^*(s_0) &= \sum_{i=1}^{n} \hat{\alpha}_i x_i \\ \hat{\sigma}^2_{s_0} &= \sum_{i=1}^{n}\sum_{j=1}^{n} \hat{\alpha}_i \hat{\alpha}_j \hat{\gamma}(|h_{ij}|) - \sum_{i=1}^{n} \hat{\alpha}_i \hat{\gamma}(|h_{i0}|) \end{aligned}$$

3.14 Remarks

- $X \sim$ Gaussian implies 95% prediction interval:

$$[X^*(s_0) - 1.96\sigma_{s_0}, X^*(s_0) + 1.96\sigma_{s_0}]$$

- Kriging predictor is **exact interpolator**;
(interpolator = observed value at that location)
- $\sigma^2_{s_0}$ is

$$\sigma^2_{s_0} = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j C(s_i, s_j) - 2\sum_{i=1}^{n}\alpha_i C(s_i, s_0) + Var(X(s_0))$$

- Stationarity required only because cannot otherwise estimate the covariance.

3.15 Universal Kriging

Random fields with non-constant means

- Let $X(s) = \mu(s) + Z(s)$
  $Z(s)$: 2^{nd}-order stationary with mean $= 0$

- $\mu(s)$, the *drift*, assumed to be $\sum_{l=1}^{k} a_l f_l(s)$
  $\{f_l(s), l = 1, \ldots, k\}$: known functions with parameters a_l
- Universal Kriging Estimator

$$X^*(s_0) = \sum_{i=1}^{n} \alpha_i X(s_i)$$

Weights α's chosen to get unbiased estimate with smallest prediction error

3.16 Universal Kriging

Derivation is similar to the ordinary Kriging

- Non-Bias Condition: $E[X^*(s_0)] = E[X(s_0)]$, or

$$\mu(s_0) - \sum_{i=1}^{n} \alpha_i \mu(s_i) = 0$$

Equivalently $\sum_{l=1}^{k} a_l (f_l(s_0) - \sum_{i=1}^{n} \alpha_i f_l(s_i)) = 0$
Since a_l's are non zero, the condition becomes

$$f_l(s_0) = \sum_{i=1}^{n} \alpha_i f_l(s_i) \quad \text{for } \; l = 1, \dots, k \qquad (3)$$

- Universal Kriging variance: same form as (2)
Hence α's chosen to minimize (2) & satisfy (3)

- Ordinary Kriging is a special case
  eg. $f_1 = 1$ & $f_2 = \ldots = f_l = 0$
- Like ordinary Kriging, stationarity not necessary

3.17 Other Kriging methods

- Multivariate Kriging - ***coKriging***
- ***Trans-Gaussian Kriging***(TGK): applying the Kriging method on Box-Coxed X - (***indicator or probability Kriging***)
- Non-linear Kriging: ***disjunctive Kriging***

$$X^*_{DK}(s_0) = \sum_{i=1}^{n} f_i(X(s_i))$$

f_i's: selected to minimize $E[X(s_0) - X^*DK(s_0)]^2$

References: [Cressie, 1993], [Wackernagel, 2003]

3.18 Other Kriging methods

- ***Model based Kriging***

  Example: Binary spatial process modeled by

$$\log \frac{p}{1-p} = \beta X$$

  where X is spatial process modeled by methods described above. Observations are counts & X a latent Gaussian field

References: [Diggle and Ribeiro Jr, 2010]

3.19 Deficiencies of Kriging

- Optimal only if covariances known. In practice, they are estimated & plugged into the interpolators, thereby underestimating the uncertainty.
- Generally requires isotropic variogram models - not realistic for environmental problems. Can be achieved by **spatial warping** or by **dimension expansion**

3.20 The Sampson-Guttorp method: Warping

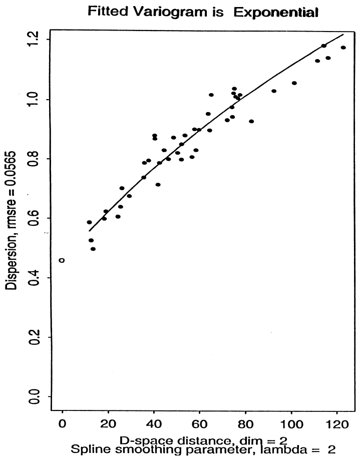
Nonparametric method for modelling spatial covariance structure without assuming stationarity [Sampson and Guttorp, 1992]

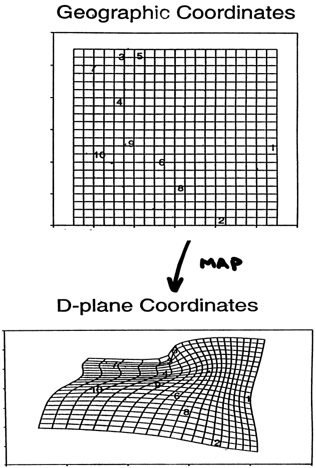
- **BASIC IDEA:** Map geographic space (G-Space) into dispersion space (D-space) where isotropy assumption valid. That is find $f : G \to D$ with

$$z_i = f(s_i) \text{ or } s_i = f^{-1}(z_i)$$

- Estimate (isotropic) semi-variogram, $\hat{\gamma}_D$, using D-distances (ie. between z_i) & estimated dispersion ($v_{ij} = 2 - 2co\hat{r}r_{ij}$)

3.21 Warping for Hourly PM_{10} in Vancouver - 1994-1999





3.22 The SG-method: Warping

- Correlation c_{ij} between s_i & s_j, obtained by:
  - getting D-distance, d_{ij} between z_i & z_j
  - evaluating $c_{ij} = 1 - \hat{\gamma}_D(d_{ij})$
- The SG-approach ensures constructed **correlation matrix,** $\{c_{ij}\}$**, non-negative definite** – based on a variogram.

3.23 SG-method: Construction of f

A two-step procedure using the observed dispersion (v_{ij}):

- Using the multidimensional scaling to find a configuration of the locations, s_i, so that their new inter-distances are 'close' to the corresponding dispersions,
  ie.
  $$min_\delta \sum_{i<j} \frac{(\delta(v_{ij}) - d_{ij})^2}{\sum d_{ij}^2}$$
  over all monotone functions

3.24 SG-method: Construction of f

- Fitting a thin-plate spline mapping, f, between new locations z_i & original locations s_i,
  ie.

$$f(s) = \alpha_0 + \alpha_1 s^{(1)} + \alpha_2 s^{(2)} + \sum_{i=1}^{n} \beta_i u_i(s)$$

where $\quad u_i(s) = |s - s_i|^2 log|s - s_i|$

Find α's & β's by minimizing

$$\sum_{j=2}^{2} \sum_{i=1}^{n} (z_i^{(j)} - f_j(s_i^{(j)}))^2 + \lambda(J_2(f_1) + J_2(f_2))$$

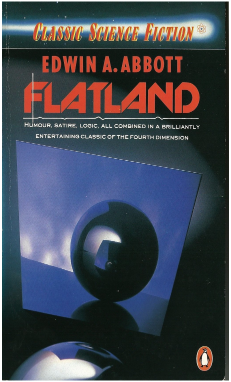
Smoothing parameter $\lambda \to \infty$ leads to $\beta \to 0$

3.25 SG-method: Implementation

- Need to estimate λ in the construction of f
- By trial – &– error or cross-validation to best estimate of dispersion while avoiding the folding of G space

3.26 New approach to nonstationarity: dimension expansion

An old idea actually (Abbott 1884). Now picked up by physicists in **string theory** who claim we live in 10 dimensional world.

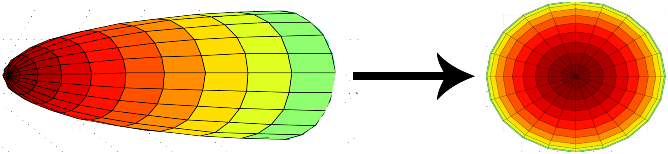


“ Place a penny on one of your tables in space; and leaning over look down upon it. It will appear as a circle. But now, drawing back to the edge of the table, gradually lower your eye....and you will find the penny becoming more and more oval...until you have placed your eye exactly on at the edge of the table [when] ...it will become a straight line.

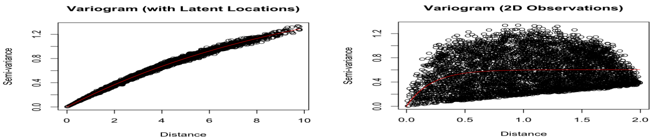
Edwin Abbott Abbott (1884)”

Example: Gaussian spatial process on half-ellipsoid.

Observations projected onto a 2-D disk.



Variogram plots



3.27 Dimension Expansion:

Embed original field in space of higher dimension for easier modeling.

- Original monitoring site coordinate vectors $\boldsymbol{s_1}, \ldots, \boldsymbol{s_g}$ each of dimension d
- Augment these coordinate vectors to get new site coordinate vectors $[\boldsymbol{s_1}, \boldsymbol{z_1}], \ldots, [\boldsymbol{s_g}, \boldsymbol{z_g}]$ each of dimension $d + p$.
- Goal: $Y([\boldsymbol{x}, \boldsymbol{z}])$ is now stationary with variogram $\gamma_\phi([\boldsymbol{s_i}, \boldsymbol{z_i}] - [\boldsymbol{s_j}, \boldsymbol{z_j}])$.

3.28 Theoretical support

Perrin and Schlather [2007]: Proves (subject to moment conditions) that for any Gaussian process Z on $\mathcal{R}^d$ there exists a stationary Gaussian field Z^* on $\mathcal{R}^{d+p}, p \geq 2$ such that Z on $\mathcal{R}^d$ is a realization of Z^* .

Existence theorem only. Construction of Z^* is not given.

3.29 Finding the coordinates

Could find the $\boldsymbol{z_1}, \ldots, \boldsymbol{z_s}$

$$\hat{\phi}, \boldsymbol{Z} = \underset{\phi, \boldsymbol{Z'}}{\operatorname{argmin}} \sum_{i<j} (v^*_{i,j} - \gamma_\phi(d_{i,j}([\boldsymbol{S}, \boldsymbol{Z'}])))^2$$

Here v^*_{ij} is an estimate of variogram (spatial dispersion between sites i and j). E.g.

$$v^*_{ij} = \frac{1}{|\tau|} \sum_{\tau} |X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j)|^2,$$

with $\tau > 1$ indexing some relevant observations.

Given matrix $\boldsymbol{Z} \in \mathcal{R}^d \times \mathcal{R}^p$ construct an f with $f(\boldsymbol{S}) \approx \boldsymbol{Z}$.

- Could follow Sampson and Guttorp (1992 the original space warpers) & use thin plate spline with smoothing parameter λ_2.
- Then f^{-1} carries us from the manifold in $\mathcal{R}^{d+p}$ defined by $(\boldsymbol{S}, f(\boldsymbol{S})), \boldsymbol{S} \in \mathcal{R}^d$ back to the original space.
- In other words, $f^{-1}(\boldsymbol{Z}) = \boldsymbol{S}$ so no issues arise around the bijectivity of f as in e.g. space warping.

3.30 Finding the # of new coordinates

- Could use cross-validation or model selection to determine Z's dimension.
- But for parsimony and to regularize (avoid overfitting) in the optimization step we instead solve

$$\hat{\phi}, \boldsymbol{Z} = \underset{\phi, \boldsymbol{Z'}}{\operatorname{argmin}} \sum_{i<j} (v^{*}_{i,j} - \gamma_{\phi}(d_{i,j}([\boldsymbol{S}, \boldsymbol{Z'}])))^2 + \lambda_1 \sum_{k=1}^{p} ||\boldsymbol{Z'}_{\cdot,k}||_1$$

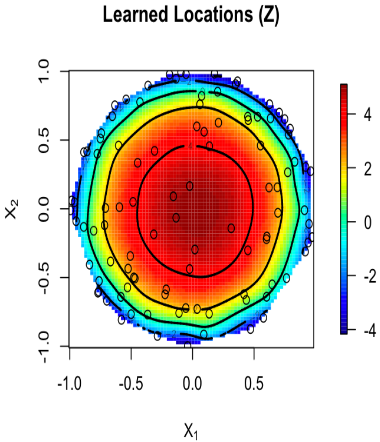
- λ_1 regularizes estimation of Z and may be estimated through cross-validation. But other model fit diagnostics or prior information could be used.

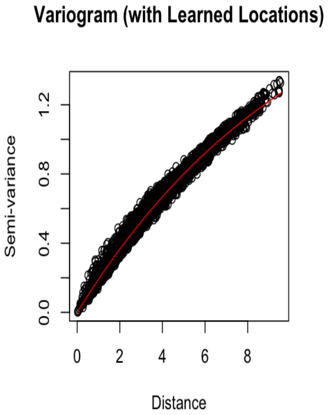
3.31 Solving the Optimization Problem

- As with traditional multi-dimensional scaling, first objective function does not have unique maximum. But learned locations unique up to rotation, scaling, and sign.
- Optimization problem more regularized, due to penalty function. Result: optimization is unique (up to sign and indices of zero/non-zero dimensions).
- We use gradient projection method of [Kim et al., 2006] to do the optimization.

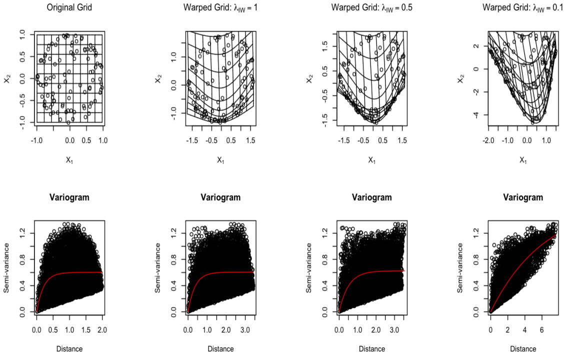
3.32 Ellipsoid application revisited

- **Dimension expansion on ellipsoid simulation yields**





In contrast, warping does not work well.



3.33 Bayesian Kriging

Prediction at u new locations given observations at g current monitoring sites

- Let $X(s) = \mu(s) + Z(s)$ with

$$\mu(s) = \sum_{l=1}^{k} a_l f_l(s), \text{ (universal Kriging setting)}$$
$$Z(s) \sim \text{Gaussian mean} = 0$$

- Vector notation:

$$\begin{aligned} X^{[u]} &= \mathbf{X}^{[u]}\boldsymbol{\beta} + Z^{[u]} \\ X^{[g]} &= \mathbf{X}^{[g]}\boldsymbol{\beta} + Z^{[g]} \end{aligned}$$

where $\boldsymbol{\beta} = (a_1, \ldots, a_k)^T$ and **X** = function of f's

- Let $\boldsymbol{\Sigma} = Cov(Z) = \frac{1}{\theta}\begin{pmatrix} \Sigma^o_{uu} & \Sigma^o_{ug} \\ \Sigma^o_{gu} & \Sigma^o_{gg} \end{pmatrix}$

3.34 Bayesian Kriging

Note: If Σ known, Kriging estimator & variance are mean and variance of $(X^{[u]} \mid X^{[g]})$ (Gaussian case)

Kitanidis [1986]: Assume Σ^o's known; put priors on $\boldsymbol{\beta}$ & θ

- Conjugate priors for $\boldsymbol{\beta}$ and θ:

$$\begin{aligned} \boldsymbol{\beta} \mid \theta &\sim N_k\left(\beta_0, (\theta F)^{-1}\right) \\ \theta &\sim Gamma\left(\frac{\nu}{2}, \frac{\nu q}{2}\right) \end{aligned}$$

- Predictive distribution:

$$(X^{[u]} \mid X^{[g]}) \sim t_u(\mu_{u|g}, \Psi_{u|g}, \nu + g)$$

where $\mu_{u|g}$ and $\Psi_{u|g}$ are functions of Σ^o matrices

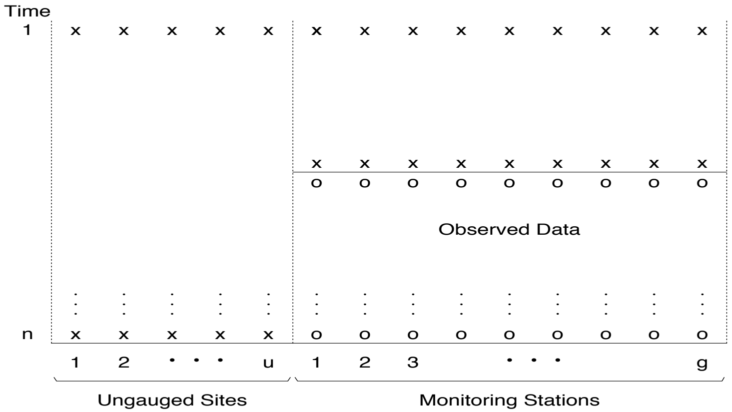
3.35 Remarks

- Kriging a special case – no uncertainty in β and θ
- Important theory but not practical – need known Σ^o's
- Handcock and Stein [1993]: Assume further $\Sigma^o = \{q_{ij}\}$
  - $q_{ij} = \gamma(|s_1 - s_2|)$ - Whittle-Matern model (isotropic)
    ie. $\gamma(x) = a + \frac{b}{2^{\nu-1}\Gamma(\nu)}\left(1 - (t_0 x)^{\nu}\,\kappa_{\nu}\,(t_0 x)\right)$
  - Obtain t–distribution for known ν and t_0
  - Plug–in estimates in applications
  - Extended with recent advents in MCMC, eg. [De Oliveira et al., 1997], [Gaudard et al., 1999]
- Isotropy assumption still needed !!

3.36 Hierarchical Bayesian Kriging - BSP method

A fast Bayesian alternative to Kriging [Le and Zidek, 2006].

Consider a simple setting:



3.37 Hierarchical Bayesian Kriging - BSP method

Model construction:

- Model: $X_t \mid z_t, B, \Sigma \quad \sim \quad N_p(z_t B, \Sigma)$
- Prior: Conjugate

$$\begin{aligned} B \mid B_o, \Sigma, F &\sim N_{kp}\left(B_o, F^{-1} \otimes \Sigma\right) \\ \Sigma \mid \Psi, \delta &\sim W_p^{-1}(\Psi, \delta) \qquad \text{(inverted Wishart)} \end{aligned}$$

- Predictive distribution - D observed data

$$\begin{aligned} X_m^{(g)} \mid D &\sim t_g\left(\mu_{gg}, \hat{\Psi}_{gg}, \delta + n - u - g + 1\right) \\ X_m^{(u)} \mid X_m^{(g)}, D &\sim t_u\left(\mu_{u|g}, \hat{\Psi}_{u|g}, \delta - u + 1\right) \end{aligned}$$

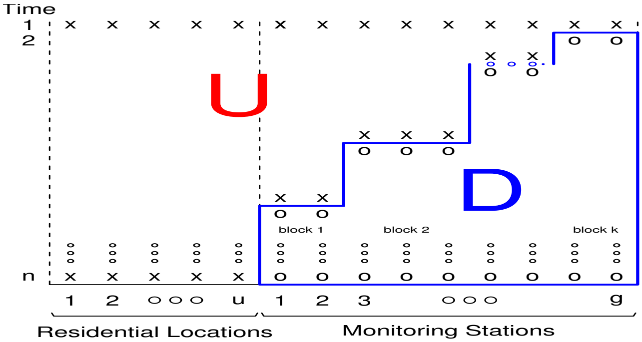
3.38 Remarks

- $\mu_{gg}, \mu_{u|g}, \hat{\Psi}_{gg}, \hat{\Psi}_{u|g},$: Functions of hyperparameters
- The predictive distribution is not a standard distribution but a product of two multivariate Student t distributions - completely characterized if hyperparameters are known
- Σ unstructured with its uncertainty (and B's) incorporated through prior distribution - reflected in the predictive distribution.
- Hyperparameters estimated using the type-II MLE

  ie. $\max \; f(D|\Psi, B_o, \delta)$
  - Empirical Bayes
  - Estimated Ψ_{gg} extended using SG method to estimate Ψ - avoiding isotropy assumption

3.39 Staircase pattern

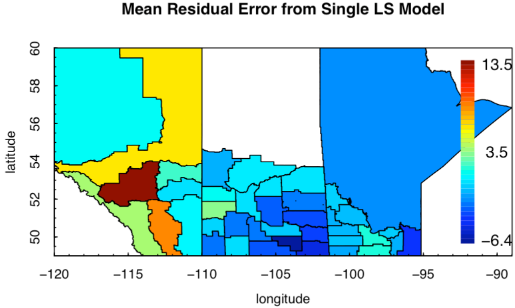
BSP handles staircase data patterns with little computational expense.



5. Lattice processes

5.1 Example

Annual Canadian prairie crop yield residuals by agrodistrict after linear regression on water stress index. Bornn and Zidek [2012]



5.2 Autogressive model analog; the CAR approach

Space unlike time not ordered. Conditional autogressive approach (CAR) is one way of emulating the AR model for fixed time t. Let:

- $D = \{s_1, \ldots, s_m\}$ be the lattice
- $X(s_i, t)$ be a response of interest
- $\mathbf{X}_i$ be all responses but $X(s_i, t)$
- $N(s_i)$ be s_i neighbourhood

The CAR model:

$$X(s_i, t) \sim N\left(\mu_i, \sigma_i^2\right), \text{ for all i}$$

with

$$E(X(s_i, t)|\mathbf{X}_i) = \sum_{s_j \in N(s_i)} c_{ij} X(s_j, t), \; Var(X(s_i, t)|\mathbf{X}_i) = \tau_i^2$$

5.3 The CAR approach

Does CAR necessarily determine a joint distribution

$$[X(s_i,t),\ldots,X(s_m,t)]?$$

Answer: Yes under reasonable conditions. [Besag, 1974]

5.4 CAR in process model

The following hierarchical model induces a CAR structure [Cressie and Wikle, 2011].

- **Measurement model:**

$$Y(s_i, t) \sim \ ind\ Poi(\exp[X(s_i, t])$$

- **Process model:**

$$[\mathbf{X}|\boldsymbol{\beta}, \tau^2, \phi] = Gau(\mathbf{Z}\boldsymbol{\beta}, \Sigma[\tau^2, \phi])$$

  where **Z** represents site specific covariates or factors & $\Sigma[\tau^2, \phi]$ the CAR neighbourhood structure.

- **Parameter model:** $[\boldsymbol{\beta}, \tau^2, \phi]$

5.5 Markov random field (MRF)

As before time t is fixed &

- $D = \{s_1, \ldots, s_m\}$ be the lattice
- $X(s_i, t)$ be a response of interest
- $\mathbf{X}_i$ be all responses but $X(s_i, t)$
- $N(s_i)$ be s_i neighbourhood

MRF models:

$$[X(s_i, t) | \{X(s_j, t),\ s_j \in N(s_i)\}] \text{ for all } i$$

When do the local MRF models determine

$$[X(s_1, t), \ldots, X(s_m, t)]?$$

Hammersley - Clifford Theorem: Gives necessary and sufficient conditions involving the *Gibbs distributions*.

5.6 Markov random fields: Example

Example: Crown die back in birch trees [Kaiser et al., 2002].
Features:

- Single timepoint, t.
- $X(s_i, t)$ = probability a tree's crown dies back in region i with $m(s_i, t)$ trees in it.
- $Y(s_i, t)$ = # of trees with die back $\sim Bin(m(s_i, t), X(s_i, t)$.
- $N(s_i)$ = all regions within 48 km of i. Conditional on $N(s_i)$, $X(s_i, t)$ has beta distribution with parameters depending on responses in neighbours.
- parsimonious model but unclear how to include time

5.7 Markov random fields: Assessment

PROS:

- elegant, simple mathematics + computational power
- may be useful component in hierarchical model

CONS:

- compatible joint distribution may not exist
- neighbours may be hard to specify
- a new site may not have neighbours for spatial prediction!
- conditional distributions may be hard to specify when “sites” are regions

5.8 Note on misaligned data

Different responses measured at monitoring sites in a systematic way. We call unmeasured complements at each site **systematically missing**. Often these unmeasured values are predicted from the others at different sites.

Change of support means data measured at different resolutions, e.g. some at a county level, some at point locations.

[Banerjee et al., 2003] provides extensive discussion.

5.9 Notes on areal data

Sometimes areal data can profitably be modeled as an aggregate of individual data.

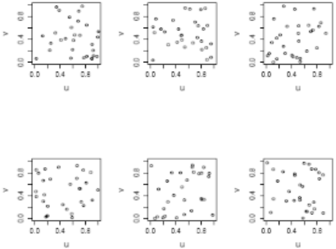
- Can reflect greater uncertainty due to variation within areas [Zidek et al., 1998]
- Was used to explore the ecological effect and develop model that avoids it [Wakefield and Shaddick, 2006].

6. Spatial point processes

6.1 Point process patterns

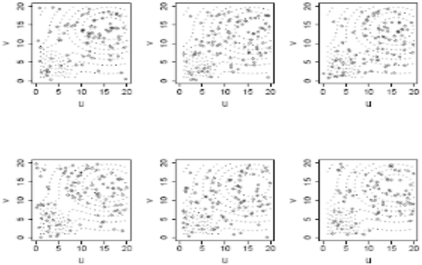
Illustrations from Gelfand (2009). SAMSI lecture.

spatial homogeneity

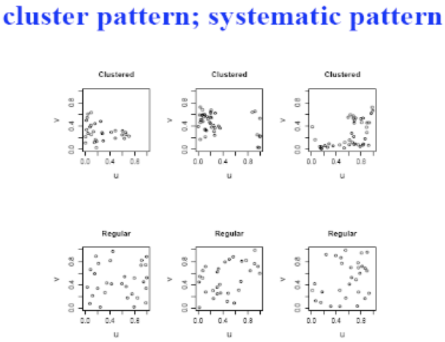


6.2 Point process patterns

spatial heterogeneity



6.3 Point process patterns



6.4 Point process model

Poisson spatial point process (PSPP)

Let $A \subset R^2$ & $X(A,t)$ = # points in A.

Assume

- $X(A_1,t)$ and $X(A_2,t)$ are independent if $A_1 \cap A_2 = \phi$
- $X(A,t) \sim Poi(\int_A \lambda[s,t]ds)$

The $X(\cdot,t)$ has a PSPP with intensity function $\lambda[\cdot,t]$.

Homogeneous if $\lambda[s,t] \equiv \lambda_t$

6.5 Point process properties

Suppose $X(\cdot, t)$ has a PSPP with intensity function $\lambda[\cdot, t]$.

Then

- $E[X(A,t)] = Var[X(A,t)] = \lambda[A,t] \int_A \lambda[s,t]ds$
- If A is small $P[X(A,t) = 0] \cong 1 - P[X(A,t) = 1]$

where $\lambda[A,t] = \int_A \lambda[s,t]ds$

6.6 Point process - inference

Partition $D = \cup_{i=1}^{M} D_i$. Then conditional on $X(D,t) = n$,

$$[(X(D_1,t),\ldots,X(D_M,t))] = multinomial(n,\mathbf{p})$$

with $\mathbf{p} = (p_1,\ldots,p_M)$ and $p_i = \lambda[D_i,t]/\lambda[D,t]$. But if the $\{D_i\}$ are small

- each will have 0 or 1 counts.
- $\lambda[D_i,t] \cong \lambda[s_i,t]ds_i$

So $\text{density of } [s_i,\ldots,s_n|X(D,t)=n] = \Pi_{i=1}^{n}\lambda[s_i,t]/(\lambda[D,t])^n$

6.7 Point process - inference

Conclusion: Given points $\{s_i^o\}$ at which events occur the **likelihood function** is

$$\frac{\Pi_{i=1}^n \lambda[s_i^o, t]}{(\lambda[D, t])^n} \times \frac{\lambda[D, t])^n \exp\left(-\lambda[D, t]\right)}{n!}$$

Example: $\lambda[s, t] = \exp \xi_0 + \xi_1 Z(s)$ where Z is observable covariate process e.g. 'temperature'. Then the likelihood can be used to estimate these parameters with integral approximated.

6.8 Cox process

- **Measurement model:** $X(A,t)|\lambda \sim Poi(\int_A \lambda[s,t]ds),\ \text{for all } A$
- **Process model:** $\log\lambda[\cdot,t]$ is a Gaussian process on R^2 with expectation and covariance

$$E[\log\lambda[s,t]] = \mathbf{Z}(s,t)\boldsymbol{\beta}$$
$$C_t[s_1,s_2|\boldsymbol{\phi}] = Cov[\log\lambda[s_1,t],\ \log\lambda[s_2,t]]$$

- **Parameter model:** $[\boldsymbol{\beta}, \phi]$

Then marginal distribution $[X]$ called **Cox process**

7 Spatio–temporal processes

7.1 Spatio–temporal modeling

Incorporating time.

- Depends on random response paradigm: point referenced; lattice; point process.
- Active area of current development

7.2 General approaches to incorporating time

Approach 1: Treat continuous time as like another spatial dimension with stationarity assumptions. Eg. Spatio–temporal Kriging. [Bodnar and Schmid, 2010]. **NOTE:** Constructing covariance models is more involved [Fuentes et al., 2008]

Approach 2: Integrate spatial fields over time. Eg. Given a spatial lattice let **X(t)** $: m \times 1$ be vectors of spatial responses at lattice points. Eg. use multivariate autoregression.

Approach 3: Integrate times series across space. For a temporal lattice let **X(s)** $: 1 \times T$ be vector of temporal responses at - use multivariate spatial methods. Eg.co–Kriging; BSP.

7.3 Specialized approaches

Approach 4: Build a statistiical framework on physical models that describe the evolution of physical processes over time

7.4 Example: the DLM

Combine dynamic linear models across space to get spatial predictor & temporal forecastor Huerta et al. [2004].

Result: model for hourly $\sqrt{(O_3)}$ field over Mexico City - data from 19 monitors in Sep 1997.

Measurement model:

$$X(s,t) = \beta(t) + S'(t)\alpha(s,t) + Z(s,t)\gamma(t) + \epsilon(s,t)$$

where

- $S_t : 2 \times 1$ has sin's and cos's;
- α has their amplitudes, Z temperature covariate
- $\epsilon(s,t)$: un-autocorrelated error with isotropic exponential spatial covariance.

7.5 Specialized approaches: Eg DLM

Process model:

$$\begin{aligned}\beta(t) &= \beta(t-1) + \omega^{\beta}(t)\\ \alpha(s,t) &= \alpha(s,t-1) + \omega^{\alpha}(s,t)\\ \gamma(t) &= \gamma(t-1) + \omega^{\gamma}(t)\end{aligned}$$

7.6 Specialized approaches: Eg DLM

PROS:

- intuitive, flexible
- allows incorporation of physical/prior knowledge

CONS:

- computationally intensive - maximum of 10 measurement sites
- non - unique model specification - finding good one can be difficult
- unrealistic covariance
- empirical tests suggest simpler multivariate BSP works better for spatial prediction Dou et al. [2010] and temporal forecasting [Dou et al., 2012] but much less computationally demanding, Eg. 300 measurement sites

7.7 Physical statistical modeling

- physical models needed for background
  - prior knowledge often expressed by differential equations (de's)
  - can lead to big computer models
  - yield deterministic response predictions
  - can encounter difficulties:
    - butterfly effect
    - nonlinear dynamics
    - lack of relevant background knowledge
    - lack of sufficient computing power

7.8 Physical statistical modeling

- statistical models also desirable
  - prior knowledge expressed by statistical models
  - often lead to big computer models
  - yield predictive distributions
  - can encounter difficulty:
    - off-the-shelf-models too simplistic
    - lack of relevant background knowledge
    - lack of sufficient computing power

7.9 Physical statistical modeling

May be strength in unity but:

- big gulf between two cultures
- communication between camps difficult
- approaches different
- route to reconciliation unclear

7.10 Physical statistical modeling

Approach to reconciliation - depends on: purpose; context; # of (differential) equations; etc.

With many equations (e.g. 100):

- build a better predictive response density for [field response — deterministic model outputs]
  eg. input model value as prior mean
- view model output as response and create joint density for
  $[\text{field response}, \text{model output}] =$
  $\int [\text{field response}|\lambda][\text{model output}|\lambda] \times \pi(\lambda|\text{data})d\lambda$

References: Fuentes and Raftery [2005], Liu et al. [2011]

7.11 Physical statistical modeling

With a few differential equations (de's)

Example: $dX(t)/dt = \lambda X(t)$.

Option 1: solve it and make known or unknown constants uncertain (i.e. random):
$X(t) = \beta_1 \exp \lambda t + \beta_0$

Option 2: discretize the de and add noise to get a state space model: $X(t+1) = (1+\lambda)X(t) + \epsilon(t)$

Option 3: use functional data analytic approach - incorporate de through a penalty term as in splines
$\sum_t (Y_t - X_t)^2 + (smoothing\ parameter) \int (DX - \lambda X)^2 dt$

7.13 Downscaling physical models

Regression – like approaches may be used:

$$X(s,t) = \alpha_{st} + \beta_{Mst} M(S,T) + \beta_{st} Z^{\text{covariates}}(s,t)\delta(s,t)$$

where M is physical model output, $s \in S^{\text{grid cell}}$ & $t \in T^{\text{Time Interval}}$.

References: Berrocal et al. [2010a], Zidek et al. [2012]

Wrapup

- Spatio–temporal modeling and data analysis has expanded rapidly in past 10 years. Lots of:
  - papers
  - books
  - jobs
  - conference presentations applications
- New directions are emerging:
  - Bayesian hierarchical modeling
  - Large datasets
  - Large domains
    - climate change
    - INLA
- Lots of research opportunities

Contact information

- Jim Zidek, Dept Statistics, UBC
  - email: jim@stat.ubc.ca
  - internet: http://www.stat.ubc.ca/ jim
  - **Copy of long version of this lecture:** www.stat.ubc.ca/ jim/talks.html

References

S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2003.

V.J. Berrocal, A.E. Gelfand, and D.M. Holland. A spatio-temporal downscaler for output from numerical models. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(2): 176–197, 2010a. ISSN 1085-7117.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B*, pages 192–236, 1974.

O. Bodnar and W. Schmid. Nonlinear locally weighted kriging prediction for spatio-temporal environmental processes. *Environmetrics*, 21:365–381, 2010.

L. Bornn and J.V. Zidek. Efficient stabilization of crop yield prediction in the canadian prairies. *Agricultural and Forest Meteorology*, 152: 223–232, 2012.

N. Cressie. *Statistics for Spatial Data*. John Wiley and Sons, 1993.

N. Cressie and C.K. Wikle. *Statistics for spatio-temporal data*, volume 465. Wiley, 2011.

V. De Oliveira, B. Kedem, and D.A. Short. Bayesian prediction of transformed gaussian random fields. *Journal of the American Statistical Association*, pages 1422–1433, 1997.

P.J. Diggle and P.J. Ribeiro Jr. *Model based geostatistics*. Springer Verlag, 2010.

Y. Dou, N. D Le, and J. V Zidek. Modeling hourly ozone concentration fields. *The Annals of Applied Statistics*, 4(3):1183–1213, 2010.

YP Dou, ND Le, and J.V. Zidek. Temporal prediction with a bayesian spatial predictor: an application to ozone fields. *Advances in Meteorology*, page To appear, 2012.

M. Fuentes and A.E. Raftery. Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models. *Biometrics*, 61:36–45, 2005.

M. Fuentes, L. Chen, and J.M. Davis. A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics*, 19(5):487–507, 2008.

M. Gaudard, M. Karson, E. Linder, and D. Sinha. Bayesian spatial prediction. *Environmental and Ecological Statistics*, 6(2):147–171, 1999.

M.S. Handcock and M.L. Stein. A bayesian analysis of kriging. *Technometrics*, pages 403–410, 1993.

R. Hosseini, N. Le, and J. Zidek. Selecting a binary markov model for a precipitation process. *Environmental and Ecological Statistics*, 18 (4):795–820, 2011a.

R. Hosseini, N.D. Le, and J.V. Zidek. A characterization of categorical markov chains. *Journal of Statistical Theory and Practice*, 5(2): 261–284, 2011b.

G. Huerta, B. Sansó, and J.R. Stroud. A spatiotemporal model for mexico city ozone levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(2):231–248, 2004.

M.S. Kaiser, N. Cressie, and J. Lee. Spatial mixture models based on exponential family conditional distributions. *Statistica Sinica*, 12(2): 449–474, 2002.

Y. Kim, J. Kim, and Y. Kim. Blockwise sparse regression. *Statistica Sinica*, 16(2):375, 2006.

P.K. Kitanidis. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water resources research*, 22(4):499–507, 1986.

D.G. Krige. *A Statistical Approach to Some Mine Valuation and Allied Problems on the Witwatersrand: By DG Krige*. PhD thesis, University of the Witwatersrand, 1951.

Nhu Le and James Zidek. *Statistical Analysis of Environmental Space-Time Processes (Springer Series in Statistics)*. Springer, 1 edition, 2006. ISBN 0387262091.

Zhong Liu, Nhu Le, and James Zidek. An empirical assessment of bayesian melding for mapping ozone pollution. *Environmetrics*, 22 (3):340–353, 2011. doi: 10.1002/env.1054.

G. Matheron. Principles of geostatistics. *Economic geology*, 58(8): 1246–1266, 1963.

O. Perrin and M. Schlather. Can any multivariate gaussian vector be inter preted as a sample from a stationary random process? *Statist. Prob. Lett.*, 77:881–4, 2007.

P.D. Sampson and P. Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87:108–119, 1992.

O. Schabenberger and CA Gotway. Statistical methods for spatial data analysis, chapmann and hall. *CRC, Florida*, 2005.

H. Wackernagel. *Multivariate geostatistics: an introduction with applications*. Springer Verlag, 2003.

Jon Wakefield and Gavin Shaddick. Health-exposure modeling and the ecological fallacy. *Biostatistics*, 7(3):438–455, 2006. ISSN 1465-4644. doi: 10.1093/biostatistics/kxj017.

James Zidek, Nhu Le, and Zhong Liu. Combining data and simulated data for space–time fields: application to ozone. *Environmental and Ecological Statistics*, 19(1):37–56, 2012. ISSN 1352-8505. doi: 10.1007/s10651-011-0172-1.

J.V. Zidek, R. White, N.D. Le, W. Sun, and R.J. Burnett. Imputing unmeasured explanatory variables in environmental. *Can. Jour. Statist.*, 26:537–548, 1998.