Spatio–temporal methods in environmental epidemiology

Lecture 12

Overall outline

1. Temporal processes
2. Spatial processes or “fields”
    - Point referenced data
    - Area data
    - Point process data
3. Spatio–temporal processes

Review

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 measurements 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

Hierarchical modeling: Alternate formulation with
[X] = probability distribution of X

- $[\textit{parameters}] = [\theta]$
- $[\textit{process}|\textit{parameters}] = [Y \mid \theta]$
- $[\textit{measurement}|\textit{process}, \textit{parameters}] = [Z \mid Y, \theta]$

Review: What if process seems nonstationary?

2. Adopt nonstationary modeling approach, convolution approach:

Represent the residual as

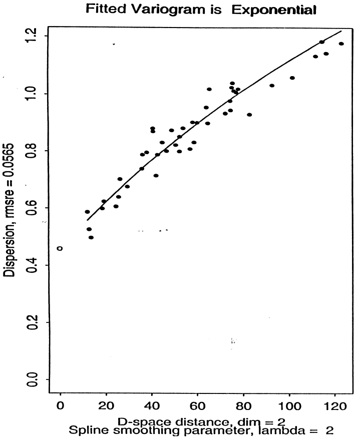
$$W(s) = \int K(s - s')W^*(s')ds'$$

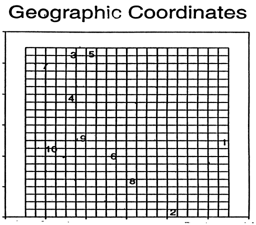
where W^* is stationary.

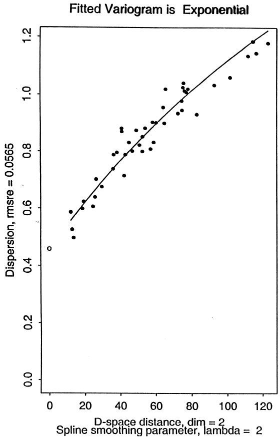
NOTE: Allows only modest degree of nonstationarity.

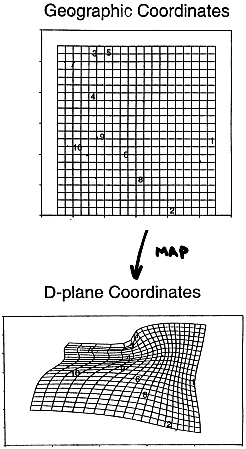
3. Warping: The famous Sampson–Guttorp approach warps the geographic space into dispersion space so that strongly correlated sites are moved closet together, uncorrelated ones further apart.
4. Dimension expansion: Keep the geographic space as is but add additional dimensions.

Review: Hourly PM_{10} in Vancouver -1994-1999



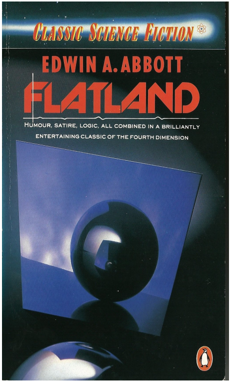



Fitted Variogram is Exponential
Dispersion, rmsre = 0.0565
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
20
40
60
80
100
120
D-space distance, dim = 2
Spline smoothing parameter, lambda = 2


Geographic Coordinates
MAP
D-plane Coordinates

New approach: dimension expansion[1]

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



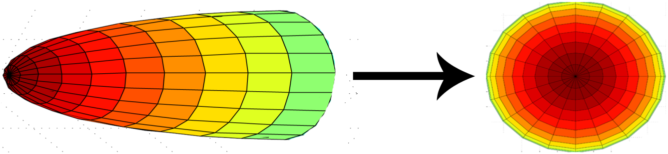
[1] Bornn et al. [2012]

" 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.

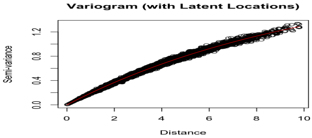
Edwin Abbott Abbott (1884)"

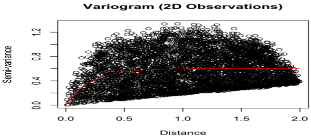
Example: Gaussian spatial process on half-ellipsoid.

Observations projected onto a 2-D disk.



Variogram plots





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 $\gamma_\phi([\boldsymbol{s_i}, \boldsymbol{z_i}] - [\boldsymbol{s_j}, \boldsymbol{z_j}])$.

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.

Finding the coordinates

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

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

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

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

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.

Finding the # of new coordinates

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

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

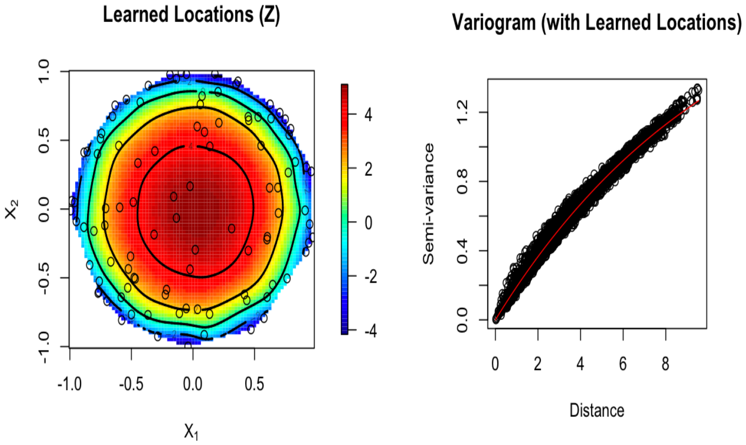
- λ_1 regularizes estimation of $\boldsymbol{Z}$ and may be estimated through cross-validation. But other model fit diagnostics or prior information could be used.

Solving the Optimization Problem

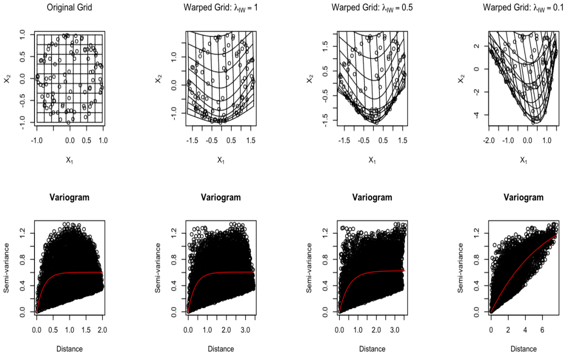
- As with traditional multi-dimensional scaling, first objective function does not have unique maximum. But learned locations unique up to rotation, scaling, and sign.
- Optimization problem more regularized, due to penalty function. Result: optimization is unique (up to sign and indices of zero/non-zero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

Ellipsoid application revisited

- **Dimension expansion on ellipsoid simulation yields**

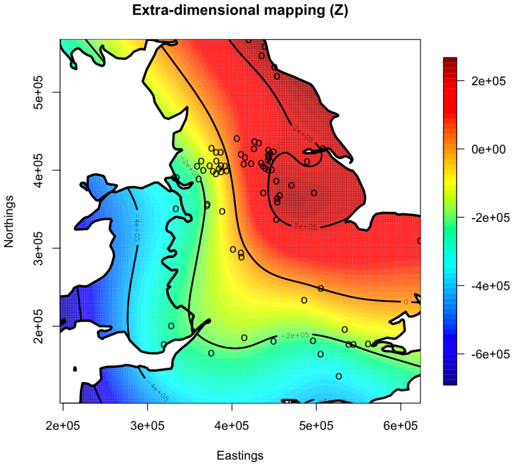


In contrast, warping does not work well.



Blacksmoke data in the UK

Black smoke very non-stationary field.



Wrapup on Kriging: other methods

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

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

f_i's: selected to minimize $E[Y(s_0) - Y^*DK(s_0)]^2$
Solutions difficult but less restrictive then TGK

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

- ***Model based Kriging***[2]
  Example: Binary spatial process modeled by

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

  where Y is spatial process modeled by methods described above. Observations are counts & Y a latent Gaussian field. Here p could be likelihood that disease is present as you move away from a hazardous waste site.

[2] Diggle and Ribeiro Jr [2010]

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**

These deficiencies have spawned Bayesian kriging. Discussion deferred to lectures on spatio–temporal processes.

Areal - Lattice Processes

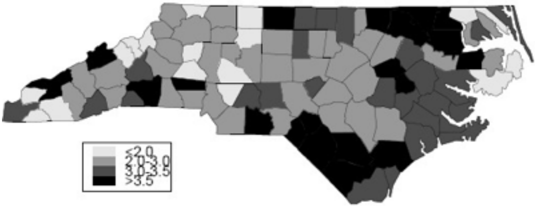
These processes cover all points in a discrete spatial domain $D = \{s_1, \ldots, s_m\}$. Can represent blocks or points on an irregular grid.

Why model lattice processes?

- To spot spatial patterns such as elevated disease counts near hazardous waste sites.
- To smooth data across space by borrowing strength - small units may not contain much data

Example: SIDS data

This well known data were treated in Cressie's 1993 text on spatial statistics. The represent counts of the sudden death infant syndrome. A plot of the counts and their counts is given in the figure. This exemplifies data obtained from records representing administrative regions like cities. Concerns about cause in high count regions.



Proximity matrices

Play fundamental role in analyzing such data. Form: $W = \{w_{ij}\}$ with $w_{ii} = 0$ represents the proximity to one another of two locations or regions i, j.

- Examples:
  - $w_{ij} = 1$ if and only they have common boundary.
  - $w_{ij} =$ inverse distance between units
  - $w_{ij} = 1$ if distance between units is $\leq K$
  - $w_{ij} = 1$ for all m of i's nearest neighbours j
- W is typically symmetric, but need not be
- $\tilde{W}$: is standardized so rows sum to one but symmetry lost
- W's elements called "weights"
- Can be used to define neighbours of i

Moran's I

W can be used to define clustering indices such as Moran's I for n regions:

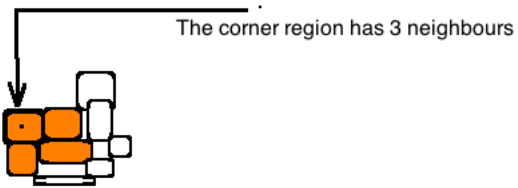
$$I = \frac{\sum_i \sum_j w_{ij}/w_{\{-i\}\cdot}(Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_i (Z_i - \bar{Z})^2/n}.$$

Here $w_{\{-i\}\cdot} = \sum_{i \neq j} w_{ij}$ & I large means that nearby points are similar. Good exploratory tool for cluster detection.

NOTE: W can be used to construct smoothers.

Markov random field (MRF)

Markov random fields focus on local modeling of spatial relationships through conditional distributions.



Neighbourhoods

- $D = \{s_1, \ldots, s_m\}$ be the lattice indices (e.g centroids)
- $Y(s_i)$ be a process of interest
- $\mathbf{Y}_{-i}$: all responses but $Y(s_i)$
- Define $N(s_i) \subset \{s_1, \ldots, s_m\}$ as s_i neighbourhood if $[Y(s_i) \mid Y_{-i}] = [Y(s_i) \mid Y(s_j),\ s_j \in N(s_i)]$

Local dependence

Specify local spatial dependencies by:

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

$\boxed{\text{Do these determine joint distribution } [Y(s_1), \ldots, Y(s_m)]?}$

If yes field is MRF.

Brook's Lemma

Brook's lemma [Brook, 1964][3] says "YES" if $N(s_i) \equiv D_{-i}$ for all i. More precisely it says if $m = 2$ for simplicity and we pick fix (y_{10}, y_{20}), for any (y_1, y_2).

$$[y_1, y_2] = \frac{[y_1 \mid y_2][y_2 \mid y_{10}]}{[y_{10} \mid y_2][y_{20} \mid y_{10}]}[y_{10}, y_{20}]$$

Left hand side proper means integration determines normalizing constant.

But doesn't answer question for all MRFs. Need some new concepts.

[3] Cressie and Wikle [2011] states correct citation is Besag [1974]

General result

- **Definition:** A **clique** is a set of cells or lattice indices such that each element is a neighbour of every other element
- **Definition:** A **potential function** Q of order k is a function of k arguments that is exchangeable in these arguments
  - **Example:** For binary (i.e. 0,1) data and k = 2, we take $Q(y_i, y_j) = I\{(y_i = y_j\} = y_i y_j + (1 - y_i)(I - y_j)$
- **Definition:** $p(y_1, \ldots, y_m)$ is a **Gibbs distribution** if [as function of $\{y_i\}$] it's product of potentials on cliques

Local modeling

- All cliques of size 1 $\Leftrightarrow$ implies independence
- For cliques of size 2 $\Leftrightarrow$ common choice is

$$p(y_1, \ldots, y_m) \propto \exp\left[-\frac{1}{2\tau^2}\sum_{i,j}(y_i - y_j)^2 I\{i \sim j\}\right]$$

  and therefore $[y_i \mid y_{-i}] = N(\sum_{j \in N(s_i)} y_j/m_i, \tau^2/m_i)$ where $m_i = \mid N(s_i) \mid$ is the number of neighbours of i
- **Hammersley-Clifford Theorem:** If MRF (i.e. $[y_i \mid y_j \ , j \in N(s_{-i})\ ]$ uniquely determines $p(y_1, \ldots, y_m)$) then the latter must be a Gibbs distribution
- **Geman and Geman:** The converse: if we have a joint Gibbs distribution, then we have an MRF [4].

[4] Geman and Geman [1984]

Markov random fields - Example

Example: [5] Features:

- $Y(s_i)$ = probability a health event for any individual in a region i with $m(s_i)$ susceptibles.
- $Z(s_i)$ = # of infecteds $\sim Bin(m(s_i), Y(s_i))$.
- $N(s_i)$ = all regions within fixed distance (e.g.48 km) of i. Conditional on $N(s_i)$, $Y(s_i)$ has beta distribution with parameters depending on counts in neighbours.
- parsimonious model but unclear how to include time

[5] Introduced by Kaiser et al. [2002] for crown dieback in birch trees

Markov random fields: Notes

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

Conditional autoregressive model (CAR)

Space not ordered like time. The conditional autoregressive approach (CAR) tries to emulate the AR approach. An MRF form. As before:

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

CAR model (Gaussian case):

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

with

$$E(Y(s_i)|\mathbf{Y}_{-i}) = \sum_{s_j \in N(s_i)} b_{ij} Y(s_j, t), \; Var(Y(s_i)|\mathbf{Y}_{-i}) = \tau_i^2$$

Does CAR necessarily determine a joint distribution

$$[Y(s_i), \ldots, Y(s_m)]?$$

Answer: Yes under reasonable conditions.[6]

[6] [Besag, 1974]

Implications

Brook's lemma implies (**Lab 4 exercise**):

$$p(y) = e^{-\frac{1}{2}y'D^{-1}(I-B)y}$$

with $y = y_{1:m}$ where $D = diag\{\tau_1^2, \ldots, \tau_m^2\}$ & $B = \{b_{ij}\}$.

Note that $D^{-1}(I-B)$ must be symmetric so for all i, j

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2}$$

meaning that B is not symmetric! Also note that $Cov(Y) = (I-B)^{-1}D$.

IAR: intrinsic autoregression

Much flexibility exists in choice of B. But natural choice is $B = W$ with $w_{ij} = 0$ *or* 1 for an adjacency matrix. Yet that would not be an allowable. Curiously $b_{ij} = w_{ij}/w_{i+}$ works & gives

$$p(y_i \mid y_{-i}) = N(\sum_j w_{ij} y_j / w_{i\cdot}, \tau_i^2 / w_{i+})$$

with $w_{i\cdot} = \sum_j w_{ij}$ while (**Exercise**)

$$p(y) = e^{-\frac{1}{2} y' (D_w - B) y}$$

where $D = diag\{w_{1+}, \ldots, w_{m+}\}$ and hence $Cov(Y)^{-1} = D_w - B$

However

$$(D_w - B) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 0$$

so:

1. the inverse of the covariance matrix is singular
2. the covariance is undetermined
3. the probability distribution is not integrable.

More explicitly (**Exercise**)

$$p(y) \propto \exp\left[-\frac{1}{2\tau^2}\sum_{i,j}(y_i - y_j)^2 w_{ij}\right]$$

which is non-integrable. An example where natural & proper local dependence models do not yield proper joint distribution. Meaning Y does not have stochastic generator, MCMC cannot be used, and so on. This model has been called the *intrinsic autoregression model* which de facto means a model concentrated on a lower dimensional space say where $Y_{\cdot} = 0$. Modellers use it despite issues.

Fixing the IAR

Choose $D_w - \rho W$ instead[7] with $\rho < 1$.

But now (**Exercise**)

$$p(y_i \mid y_{-i}) = N(\rho \sum_j w_{ij} y_j / w_{i+}, \tau_i^2 / w_{i+})$$

so conditional mean is fraction of neighbourhood mean. Makes interpretation and inference challenging(ρ is an extra parameter). Further even with ρ large say 0.95, Moran's I is small (around 0.25) in simulated samples[8]. So fix is unappealing.

Situation resembles AR(1) as the autocorrelation goes to 1 – model flips from AR (a stationary process) to a random walk (a non-stationary process).

[7] [Banerjee et al., 2003a]
[8] Banerjee et al. [2003a]

The revised IAR[9]

PROS:

- makes distribution proper
- adds parametric flexibility
- $\rho = 0$ interpretable as independence

CONS:

- hard to rationalize model with Y_i's conditional expectation a fraction of neighbour average – spatial interpretation?
- interpretation of ρ? As correlation seems tenuous since
  - $\rho = 0.80$ yields $0.1 < Moran's I < 0.15$
  - $\rho = 0.90$ yields $0.2 < Moran's I < 0.25$
  - $\rho = 0.99$ yields $Moran's I < 0.5$

[9] **Comments in a talk by Alan Gelfand**

CAR Note:

Spatial prediction with CAR is ad hoc using:

$$p(y_0 \mid y) = N(\sum_j w_{0j} y_j / w_{0+}, \tau^2 / w_{0+})$$

Well defined but not a CAR! That is it could not arise by application of Brook's lemma. (**Exercise**).

CAR in the non-Gaussian case

The CAR theory extends to the non–Gaussian case as the following example shows.

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

- **Measurement model:**

$$Z(s_i) \sim \ \mathit{ind}\ \mathit{Poi}(\exp\left[Y(s_i)\right])$$

- **Process model:**

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

  where $\mathbf{X}$ represents site specific covariates or factors & $\Sigma[\tau^2, \phi]$ the CAR neighbourhood structure.
- **Parameter model:** $[\boldsymbol{\beta}, \tau^2, \phi]$

Simultaneous autoregression (SAR)

This natural model [10] is like a CAR:

$$Y(s_i) - \mu(s_i) = \sum_j b_{ij}(Y(s_j) - \mu(s_j)) + \epsilon_i$$

where $\epsilon_i \sim indN(0, \sigma_i^2)$. In vector matrix form:

$$Y - \mu = B(Y - \mu) + \epsilon$$

or

$$Y = \mu + \epsilon^*$$

where $\epsilon^* \sim N_m(0, (I - B)^{-1}\Sigma(I - B')^{-1})$ with $\Sigma = diag\{\sigma_1^2, \ldots, \sigma_m^2\}$.

This model capture spatial independence through the mean structure - a moving average of the $\{\epsilon_i\}$.

[10] [Whittle, 1954]

SAR in extended form

Data model: $[Z(s_i) \mid Y(s_i), \sigma_\epsilon] = indN(Y(s_i), \sigma_\epsilon)$

Process model:
$[Y \mid \beta, \sigma^2, \rho] = N(X\beta, \sigma^2(I - \rho W')(I - \rho W))$ where W has zeros down the diagonal but need not be the adjacency matrix.

Parameter model: Prior distribution on the parameters.

A large class of models. Can see the affect of covariates on the process Y. CAR can also incorporate the $X\beta$ type model.

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., 2003b] provides extensive discussion.

Notes on areal data

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

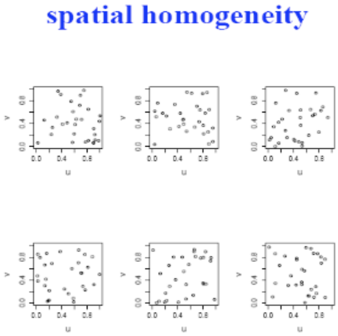
- Can reflect greater uncertainty due to variation within areas[11]
- Was used to explore the ecological effect and develop model that avoids it[12] .

[11] Zidek et al. [1998]
[12] Wakefield and Shaddick [2006]

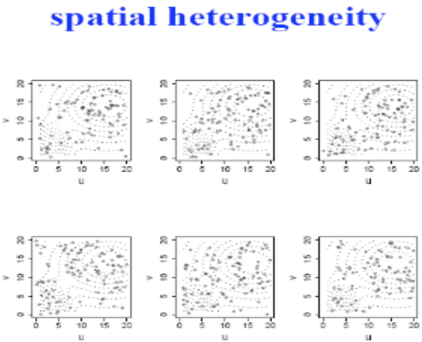
Spatial point proccesses

Point process patterns

Illustrations from Gelfand (2009). SAMSI lecture.

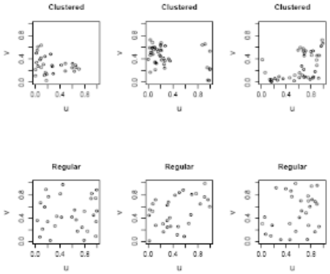


Point process patterns



Point process patterns

cluster pattern; systematic pattern



Point process model

Poisson spatial point process (PSPP)
Let $A \subset R^2$ & $Y(A)$ = # points in A.

Assume

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

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

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

Point process properties

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

Then

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

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

Point process - inference

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

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

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

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

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

Point process - inference

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

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

Example: $\lambda[s] = \exp \xi_0 + \xi_1 Z(s)$ where Z is observable covariate process e.g. 'temperature'. Then the likelihood can be used to estimate these parameters with integral approximated.

Cox process

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

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

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

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

S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical modeling and analysis for spatial data*, volume 101. Chapman & Hall/CRC, 2003a.

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

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

L. Bornn, G. Shaddick, and J.V. Zidek. Modeling nonstationary processes through dimension expansion. *Journal of the American Statistical Association*, 107(497):281–289, 2012.

D. Brook. On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, 51(3/4):481–483, 1964.

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.

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

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741, 1984.

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.

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.

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.

P. Whittle. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.

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.