New frontiers: advanced modelling in space and time

Gavin Shaddick, Daniel Simpson,
University of Bath

James V Zidek,
University of British Columbia

Carlos Diaz
UNAM

2^{nd} - 4^{th} June 2016

OUTLINE

Friday, June 2

- *10:00 - 10:30* Coffee
- *10:30 - 11:00* Introduction to the BUC workshop series
- *11:00 - 12:00* Strategies for space-time modelling
- *12:00 - 13:00* Lunch
- *13:00 - 14:30* Better exposure measurements through better design
- *14:30 - 15:00* Coffee
- *15:00 - 17:30* Young researchers' conference

OUTLINE

Saturday, June 3

- *10:00 - 10:30* Coffee
- *10:30 - 12:00* Modelling point patterns
- *12:00 - 13:00* Lunch
- *13:00 - 14:30* Applications of space-time modelling
- *14:30 - 15:00* Coffee
- *15:00 - 17:30* Lab session

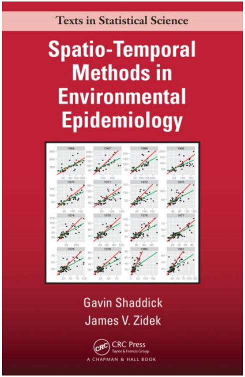
Outline

Sunday, June 4

- *10:30 - 13:00* Roman Baths
- *13:00 - 15:00* Sunday Lunch at the Westgate
- *15:00 - 16:00* Walk around Bath

Course Textbook

Title: Spatio-Temporal Methods in Environmental Epidemiology
Authors: Gavin Shaddick and Jim Zidek
Publisher: CRC Press
Resource Website:
`http://www.stat.ubc.ca/~gavin/STEPIBookNewStyle/`



Contact Information

Dr. Gavin Shaddick, University of Bath

- Email: G.Shaddick@bath.ac.uk
- Webpage: `http://people.bath.ac.uk/masgs/`

Dr. Daniel Simpson, University of Bath

- Email: D.Simpson@bath.ac.uk

Prof. Jim Zidek, University of Bath

- Email: jim@stat.ubc.ca

Prof. Carlos Diaz, UNAM

- Email: carlos@sigma.iimas.unam.mx

Strategies for space-time modelling

Overview of spatio–temporal modelling

- In recent years there has been an explosion of interest in spatio–temporal modelling.
- One major area where spatio-temporal is developing is environmental epidemiology, where interest is in the relationship between human health and spatio–temporal processes of exposures to harmful agents.

Overview of spatio–temporal modelling

- An example of is the relationship between deaths and air pollution concentrations or future climate simulations, the latter of which may involve 1000's of monitoring sites that gather data about the underlying multivariate spatio–temporal field of precipitation and temperature.

The need for Spatio-Temporal modelling

- Spatial epidemiology is the description and analysis of geographical data, specifically health data in the form of counts of mortality or morbidity and factors that may explain variations in those counts over space.
- These may include demographic and environmental factors together with genetic, and infectious risk factors.
- It has a long history dating back to the mid-1800s when John Snow's map of cholera cases in London in 1854 provided an early example of geographical health analyses that aimed to identify possible causes of outbreaks of infectious diseases.

Example: John Snow's Cholera Map

Figure: John Snow's map of cholera cases in London 1854. Red circles indicate locations of cholera cases and are scaled depending on the number of reported cholera cases.Purple taps indicate locations of water pumps.

THE NEED FOR SPATIO-TEMPORAL MODELLING

- Advances in statistical methodology together with the increasing availability of data recorded at very high spatial and temporal resolution has lead to great advances in spatial and, more recently, spatio–temporal epidemiology.
- These advances have been driven in part by increased awareness of the potential effects of environmental hazards and potential increases in the hazards themselves.

Example: Spatial Correlation in the UK

- An example of spatial correlation can be seen in the next slide which shows the spatial distribution of the risk of hospital admission for chronic obstructive pulmonary disease (COPD) in the UK.
- There seem to be patterns in the data with areas of high and low risks being grouped together suggesting that there may be spatial dependence that would need to be incorporated in any model used to examine associations with potential risk factors.

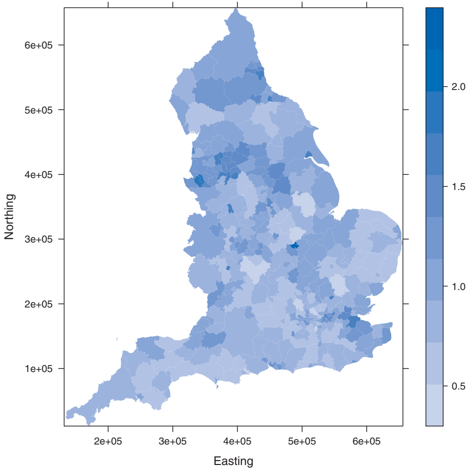


Figure: Map of the spatial distribution of risks of hospital admission for a respiratory condition, chronic obstructive pulmonary disease (COPD), in the UK for 2001. The shades of blue correspond to standardised admission rates, which are a measure of risk. Darker shades indicate higher rates of hospitalisation allowing for the underlying age–sex profile of the population within the area.

EXAMPLE: DAILY MEASUREMENTS OF PARTICULATE MATTER

An example of temporal correlation in exposures can be seen below, which shows daily measurements of particulate matter over 250 days in London in 1997. Clear auto-correlation can be seen in this series of data with periods of high and low pollution.

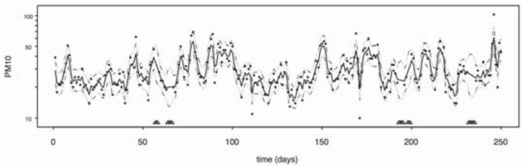


Figure: Time series of daily measurements of particulate matter (PM_{10}) for 250 days in 1997 in London. Measurements are made at the Bloomsbury monitoring site in central London. Missing values are shown by triangles. The solid black line is a smoothed estimate produced using a Bayesian temporal model and the dotted lines show the 95% credible intervals associated with the estimates.

EXAMPLE: EFFECT OF WILDCAT DRILLING IN ALASKA

- In this example, $N_T = 2$ and $t = 1, 2$ represent times before and after the startup of exploratory drilling in Harrison Bay, Alaska, the Beaufort Sea oil field having already been established.
- Interest in this case was on human welfare rather than human health, namely on the effect of such drilling on the food chain of the indigenous people who lived in that area.

EXAMPLE: EFFECT OF WILDCAT DRILLING IN ALASKA

- Clearly the risk of this drilling would depend on how wind and sea currents carried the plume of drilling mud which is used to lubricate the drill stem as it digs into the earth.
- Experts were asked to independently draw boundaries of what they saw to be the zones of equitable risk.
- There was surprising agreement amongst the experts.

Example: Effect of wildcat drilling in Alaska

- This led in the end to a model of the form

$$
\begin{array}{rcl}
Y_{st} & = & Z_{st} + v_{st},\ t = 1, 2,\ s \in S \\
Z_{st} & = & \mu_{st} + \omega_{st}, \\
\mu_{st} & = & \mu + \beta_s + \gamma_t x_t \\
\beta_s & ind \sim & N(0, \sigma^2_\beta)
\end{array}
$$

  where the dummy variable is $x_t = I\{t = 2\}$ and b_s puts s into its zone of equitable risk.

- This simple model was chosen, in part because of its simplicity; it resulted in a paired t–test like analysis to detect change.

EXAMPLE: EFFECT OF WILDCAT DRILLING IN ALASKA

- However to allow fully for uncertainty, random effects are assigned to the risk zones.
- The spatial domain S consisted of a geographic grid superimposed on the risk zones.
- The eventual design was based on that knowledge that the National Oceanic and Atmospheric Agency (NOAA), which was overseeing the project, would prefer a simple method of analysis for assessing the impact.

EXAMPLE: EFFECT OF WILDCAT DRILLING IN ALASKA

- The Bayesian elements were used to find the expected value of the uncertain non-centrality parameter for the test. This depended on the subset S that was to be selected and so was optimised to find the optimal design.
- That led to maximising the contrast in the field, with the optimal sites distributed between low and high risk zones.
- That in turn led to a theoretical paper that generalised this approach (Schumacher & Zidek, 1993)

Example: Modelling pollution fields

- It has long been recognised that particulate air pollution is associated with adverse health impacts in humans.
- Thus it is now a criteria air pollutant that is regulated to ensure air quality.
- Of particular concern is $PM_{2.5}$, consisting of small particles formed from gaseous emissions, for example, from the burning of wood.
- Both their mass μgm^{-3} as well as their counts *ppm* are considered important since a large number of tiny particles in the $PM_{2.5}$ mix,
  - for example those of size less than 1 micron in diameter, can penetrate deeply into the lung.

Example: Modelling pollution fields

- Primary interest lies in the spatial prediction of the $PM_{2.5}$ field. However, in general a spatio–temporal modelling approach is preferable since the quasi replicates of the spatial field over time enables better parameter estimation. The following model for the underlying spatio–temporal mean term was proposed

$$\mu_{st} = \beta_0 + \beta_1 p_s + \beta \alpha_s \times p_s + \sum_{l=2}^{12} \zeta_l u_{tl}$$

where the dummy $u = I\{t \text{ is in month m}\}$ tells us the month in which the time (week) $t = 1, \ldots, 52$ is located.

EXAMPLE: MODELLING POLLUTION FIELDS

- Although in some ways an appealing and simple way of handling seasonality when temporal replicates are available it comes at a cost of eleven degrees of freedom.
- These are used in estimating the set ofthe $\{\zeta_l\}$ coefficients.
- Here p_s denotes human population density while α_s is an rural–urban indicator function.

EXAMPLE: MODELLING POLLUTION FIELDS

- The process model used was

$$Z_{st} = \mu_{st} + \omega_{st} + p_s v_{st}$$

  where the the sum of the last two terms is thought of as representing a spatially varying temporal trend.
- To complete the model description for this example, we need a covariance structure and it is often assumed that space and time are separable

STRATEGIES FOR SPATIO–TEMPORAL MODELLING

- There are many ways in which space and time can be incorporated into a statistical model and we now consider a selection. One must first choose the model's space–time domain.
- Is it to be a continuum in both space and time?
- Or a discrete space with a finite number of locations at which measurements may be made?

STRATEGIES FOR SPATIO–TEMPORAL MODELLING

- Time is obviously different than space.
- For one thing, it is directed, whereas any approach to adding direction in space is bound to be artificial.
- A major challenge in the development of spatio–temporal theory has been combining these fundamentally different fields in a single modelling framework.
- Much progress has been made in this area over the last three or four decades to meet the growing need in applications of societal importance, including those in epidemiology.

STRATEGIES FOR SPATIO–TEMPORAL MODELLING

- There are competing advantages to using finite (discrete) and continuous domains.
- Indeed a theory may be easier to formulate over a continuous domain, but practical use may entail projecting themonto a discrete domain.
- Time is regarded as discrete because measurements are made at specified, commonly equally spaced, time points.
- The precise methodology will be determined by the nature of the data that is available over space, for example is it point-referenced or collected on a lattice?

STRATEGIES FOR SPATIO–TEMPORAL MODELLING

Some general approaches to incorporating time are as follows:

Approach 1: Treat continuous time as another spatial dimension,

- For example, spatio–temporal Kriging
- There is extra complexity in constructing covariance models compared to purely spatial process modelling and possible reductions in the complexity based on time having a natural ordering (unlike space) are not realised.

STRATEGIES FOR SPATIO–TEMPORAL MODELLING

Approach 2: Represent the spatial fields represented as vectors $\mathbf{Z}_t : N_S \times 1$, and combine them across time to get a multivariate time series.

Approach 3: Represent the time series as vectors, $\mathbf{Z}_s : 1 \times N_T$, and use multivariate spatial methods

- For example, co-kriging

Approach 4: Build a statistical framework based on deterministic models that describe the evolution of processes over space and time.

STRATEGIES FOR SPATIO–TEMPORAL MODELLING

- Approach 1 may appeal to people used to working in a geostatistical framework.
- Approach 2 may be best where temporal forecasting is the inferential objective while Approach 3 may be best for spatial prediction of unmeasured responses.
- Approach 4 is an important new direction that has promise because it includes background knowledge through numerical computer models.

STRATEGIES FOR SPATIO–TEMPORAL MODELLING

- If the primary aim is spatial prediction then you would want to preserve the structure of the spatial field.
- However if the primary interest is in forecasting this would lead to an emphasis in building time series models at each spatial location.
- The exact strategy for constructing a spatio–temporal model will also depend on the purpose of the analysis.

STRATEGIES FOR SPATIO–TEMPORAL MODELLING

- Interest may lie in forecasting an ambient measurement twenty-four hours ahead of time. Or to spatially predict such levels at unmonitored sites to get a better idea of the exposure of susceptible school children in a school far from the nearest ambient monitor.
- In deciding how to expand or contract an existing network of monitoring sites in order to improve prediction accuracy or to save resources, a spatio–temporal model will be required together with a criterion on which to evaluate the changes you recommend.

Bayesian Hierarchical Models

Bayesian hierarchical models are an extremely useful and flexible framework in which to model complex relationships and dependencies in data and they are used extensively throughout the course. In the hierarchy we consider, there are three levels;

(1) The observation, or measurement, level; $Y|Z, X_1, \theta_1$.
Data, Y, are assumed to arise from an underlying process, Z, which is unobservable but from which measurements can be taken, possibly with error, at locations in space and time. Measurements may also be available for covariates, X_1. Here θ_1 is the set of parameters for this model and may include, for example, regression coefficients and error variances.

Bayesian Hierarchical Models

(2) The underlying process level; $Z|X_2, \theta_2$.
The process Z drives the measurements seen at the observation level and represents the true underlying level of the outcome. It may be, for example, a spatio–temporal process representing an environmental hazard. Measurements may also be available for covariates at this level, X_2. Here θ_2 is the set of parameters for this level of the model.

(3) The parameter level; $\theta = (\theta_1, \theta_2)$.
This contains models for all of the parameters in the observation and process level and may control things such as the variability and strength of any spatio–temporal relationships.

SPATIO–TEMPORAL PROCESSES

- A spatial–temporal random field, $Z_{st}, s \in \mathcal{S}, t \in \mathcal{T}$, is a stochastic process over a region and time period.
- This underlying process is not directly measurable, but realisations of it can be obtained by taking measurements, possibly with error.
- Monitoring will only report results at N_T discrete points in time, $T \in \mathcal{T}$ where these points are labelled $T = \{t_1, \ldots, t_{N_T}\}$.
- The same will be true over space leading to a discrete set of N_S locations $S \in \mathcal{S}$ with corresponding labelling, $S = \{s_1, \ldots, s_{N_S}\}$.

Spatio–temporal Processes

We can represent the space–time random field Z_{st} in terms of a hierarchical model for the measurement and process models

$$\begin{aligned} Y_{st} &= Z_{st} + v_{st} \\ Z_{st} &= \mu_{st} + \omega_{st} \end{aligned}$$

where

- v_{st} represents independent random measurement error.
- μ_{st} is a spatio–temporal mean field (trend) that is often represented by a model of the form $\mu_{st} = x_{st}\beta_{st}$.
- ω_{st} is the underlying spatio–temporal process

SPATIO–TEMPORAL PROCESSES

- For many processes the mean term μ_{st} represents the largest source of variation in the responses.
- Over a broad scale it might be considered as deterministic if it can be accurately estimated,
  - An average of the process over a very broad geographical area.
- However where there is error in modelling μ_{st} the residuals ω_{st} play a vital role in capturing the spatial and temporal dependence of the process.

Spatio–temporal Processes

- The spatio–temporal process modelled by ω can be broken down into separate components representing space, m, time, γ and the interaction between the two, κ.

$$\omega_{st} = m_s + \gamma_t + \kappa_{st}$$

- Here, $\mathbf{m}$ would be a collection of zero mean, site-specific deviations (spatial random effects) from the overall mean, μ_{st} that are common to all times.
- For time, γ would be a set of zero mean time-specific deviations (temporal random effects) common to all sites.
- The third term κ_{st} represents the stochastic interaction between space and time.

SPATIO–TEMPORAL PROCESSES

- For example, the effect of latitude on temperature depends on the time of year.
- The mean term, μ_{st} may constitute a function of both time and space but the interaction between the two would also be manifest in κ_{st}.
- This would capture the varying intensity of the stochastic variation in the temperature field over sites which might also vary over time. In a place such as California the temperature field might be quite flat in summer but there will be great variation in winter.
- It is likely that there will be interaction acting both through the mean and covariances of the model.

SEPARABLE MODELS

- In most applications, modelling the entire spatial–temporal structure will be impractical because of high dimensionality.
- A number of approaches have been suggested to deal with this directly and we now discuss the most common of these, that of assuming that space and time are **separable**.
- This is in contrast to cases where the space–time structure is modelled jointly which are known as **non-separable** models.

SEPARABLE MODELS

- Separable models impose a particular type of independence between space and time components. It is assumed the correlation between Z_{st} and $Z_{s't}$ is $\rho_{ss'}$ at every time point t while the correlation between Z_{st} and $Z_{t's}$ is $\rho_{tt'}$ at all spatial time points s.
- The covariance for a separable process is therefore defined as

$$Cov(Z_{st}, Z_{s't'}) = \sigma^2 \rho_{ss'} \rho_{tt'}$$

for all $(s,t), (s',t') \in \mathcal{S} \times \mathcal{T}$.

SEPARABLE MODELS

- Expressed in matrix form, for Gaussian processes, we get the Kronecker product for the covariance matrix,

$$\boldsymbol{\Sigma}^{Tp \times Tp} = \sigma^2 \boldsymbol{\rho}_T^{N_T \times N_T} \otimes \boldsymbol{\rho}_S^{N_S \times N_S}.$$

where $\boldsymbol{\rho}_1$ is the between row temporal autocorrelations and $\boldsymbol{\rho}_2$ is the between column spatial correlations.

SEPARABLE MODELS

- Kronecker product models assumes temporal correlations are the same at every site.
- Likewise, the spatial correlations are the same at every point in time.
- These are strong assumptions which greatly simplify things but they do seem to be reasonable in a lot of applications, for example through cross-validation yields good results.

SEPARABLE MODELS

- Due to the reduction in computational burden that comes with this approach, the majority of work on space–time modelling tends to be based on analysing the temporal and spatial aspects separably, and then to combine the chosen models in a single separable model.

Example: A hierarchical model for spatio–temporal exposure data

- We give details of a hierarchical model described by (Shaddick & Wakefield 2002).
- There are three stages to the model: (i) the observation, or data, model; (ii) the process model which in this case now describes the form of the underlying spatial and temporal processes and (iii) assigning prior distributions to the unknown parameters.
- The model is designed for cases where there are multiple pollutants being measured at a number of monitoring sites.
- The model allows for a temporal–pollutant interaction and a spatial–pollutant interaction, with the spatial model being constant across time, isotropic and stationary.
- The model and its assumptions are now described.

Example: Stage 1 - Observation model

- At the first stage, the measurements of each pollutant $(p = 1, ..., P)$ over time $(t = 1, ..., T)$ at each monitoring site $(s = 1, ..., S)$ are modelled as a function of the true underlying level of the pollutant with a site adjustment and a pollutant-site specific error term.

$$y_{stp} = z_{stp} + v_{stp}$$

where y_{stp} denotes the observed level of the pollutant p, $p = 1, ..., N_P$ at time t and location s for $t = 1, .., N_T$ where N_T is the number of time points and $s = 1, ..., N_S$ where N_S is the number of monitoring sites. In this model v_{spt} represents the measurement errors which are assumed i.i.d. $N(0, \sigma^2_{sp})$.

Example: Stage 2 - Process model

- In this stage the underlying levels of the exposure, Z_{stp}, are assumed to comprise an underlying trend, μ_{stp} together with a separable spatio–temporal process, ω_p for each pollutant.

$$
\begin{aligned}
z_{stp} &= \mu_{stp} + \omega_{stp} \\
\omega_{stp} &= m_{sp} + \gamma_{tp} \\
\gamma_t &= \alpha\gamma_{t-1} + w_{stp}
\end{aligned} \tag{1}
$$

- Here $\mu_{stp} = \beta X$ where β is a vector of regression coefficients and X_{spt} represents an explanatory variables that may change temporally (for example, temperature), and spatially.

Example: Stage 2 - Process model

- The latter may represent, for example, spatial characteristics of the site that may be constant across time such as latitude and longitude (which could be used to remove any trend), or characteristics of the monitor, for example roadside or elevation.
- The subscript p allows these effects to be pollutant specific. and γ_t is a multivariate temporal process that induces temporal and pollutant dependence and m_{sp} represents the spatial effect of being at site s (for pollutant p).

EXAMPLE: STAGE 2 (A) - SPATIAL/POLLUTANT MODEL

- The collection of random effects $m_p = (m_{p1}, ..., m_{pN_S})'$, $p = 1, ..., P$, is assumed to arise from the multivariate normal distribution

$$m_p \sim MVN(0_{N_S}, \sigma^2_{pm}\Sigma_{pm})$$

where 0_{N_S} is an $N_S \times 1$ vector of zeros, σ^2_{pm} the between-site variance for pollutant p and Σ_{pm} is the $N_S \times N_S$ correlation matrix, in which element (s, s') represents the correlation between sites s and s', $s, s' = 1, ..., N_S$, for pollutant p.

EXAMPLE: STAGE 2 (B) - TEMPORAL/POLLUTANT MODEL

- We assume that

$$\gamma_{pt} = \gamma_{p,t-1} + w_{pt} \tag{2}$$

for $p = 1, ..., N_P$. Here $w_t = (w_{1t}, ..., w_{N_{Pt}})'$ are i.i.d. multivariate normal random variables with zero mean and variance-covariance matrix Σ_{N_P}. This matrix contains variances σ^2_{wp} thus allowing different pollutants to have different amounts of temporal dependence, and $N_P(N_P - 1)/2$ covariance terms reflecting the dependence (more precisely the covariance) between each of the pollutants, conditional on the previous values.

Example: Stage 3 - Hyperpriors

- A normal prior $N(c, C)$ is assumed for β, where c is a $q \times 1$ vector and C a $q \times q$ variance-covariance matrix.
- Gamma priors are specified for the precisions,specifically $\sigma_{sp}^{-2} \sim Ga(a_v, b_v)$.
- The variance-covariance matrix, $\Sigma_P^{-1} \sim W_P(D, d)$ where $W_P(D, d)$ denotes a $P-$dimensional Wishart distribution with mean D and precision parameter d.
- Unless there is specific information to the contrary, i.e. that a monitor with different characteristics is used at a particular site, it is assumed $\sigma_{vs}^{-2} \sim Ga(a_v, b_v)$, $s = 1, ..., S$.
- A uniform prior may be used for ϕ_p, with the limits being based on beliefs about the relationship between correlation and distance.

Example: Model assumptions

The assumptions of the model include the following:

- The measurement error variance σ^2_{sp} does not depend on time, although the model is easily extendable to situations in which the measurement error may change as a function of t.
- The relationship between the pollutants is constant over time.
- The relationship between the pollutants is spatially constant.
- The temporal and spatial components are independent.

Non-separable models

- Non-separable processes will often be more difficult to understand than when separation processes can be assumed for space and time and as a consequence modelling is often complex.
- In particular, dealing with the Kronecker products that define covariances poses technical challenges if the wrong approach is taken.
- To illustrate, consider the simple problem of showing that $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$.

NON-SEPARABLE MODELS

- This problem proves to be very difficult if we ignore the algebraic roots of the Kronecker product as a linear operator and instead use the matrix definition which for simplicity in thecase of 2×2 matrices is the 4×4 matrix given by:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} \end{pmatrix}$$

GAUSSIAN PROCESSES

- So why is all this important for modelling spatio–temporal Gaussian processes?
- There, the domain where measurements will be taken is $(s, t) \in \mathcal{S} \times \mathcal{T}$ where $\mathcal{S} \times \mathcal{T}$ denotes what is called the 'product space' of $\mathcal{S}$ and $\mathcal{T}$.
- Over that domain responses for a separable Gaussian process can be represented by a random matrix with a matrix normal distribution:

$$\mathbf{Z} \sim N_{N_S \times N_T}[\mu, \sigma^2 \rho_S \otimes \rho_T]$$

GAUSSIAN PROCESSES

- So if the temporal auto correlation matrix were known, we could easily reduce the process to another with independent replicates over time as follows.

$$\mathbf{Z}^* = (\mathbf{I} \otimes \rho_T^{-1/2})\mathbf{Z} \sim N_{N_S \times N_T}[(\mathbf{I} \otimes \rho_T^{-1/2})\mu, \sigma^2 \rho_S \otimes \mathbf{I})]$$

- Even if ρ_T is unknown, in some cases it may be possible to estimate it well, for
example when it has a simple parametric form and there are many time points.

NON-SEPARABLE MODELS

- The complexity of non-separable spatio–temporal processes often combined with computational issues has resulted in the development of a number of different approaches to modelling them.
- We now provide a brief description of a selection of the available approaches.

NON-SEPARABLE MODELS

- The complexity of non-separable spatio–temporal processes often combined with computational issues has resulted in the development of a number of different approaches to modelling them.
- We now provide a brief description of a selection of the available approaches.

Non-separable models

- A spatio–temporal model for hourly ozone measurements was developed by Carroll *et al.* (1997).
- The model,

$$Z_{st} = \mu_t + \omega_{st}$$

  combines a trend term incorporating temperature and hourly/monthly effects,

$$\mu_t = \alpha_{hour} + \beta_{month} + \beta_1 temp_t + \beta_2 temp_t^2,$$

  which is constant over space, and an error model in which the correlation in the residuals was a nonlinear function of time and space.

Non-separable models

- In particular the spatial structure was a function of the lag between observations,

$$COV(v_{st}, v_{s't'}) = \sigma^2 \rho(d, v),$$

where d is the distance between sites and $v = |t' - t'|$ is the time difference, with the correlation being given by

$$\rho(d, v) = \begin{cases} 1 & d = v = 0 \\ \phi_v^d \psi_v & d \text{ otherwise} \end{cases}$$

where

$$\log(\psi_v) = a_0 + a_1 v + a_2 v^2 \text{ and } \log(\phi_v) = b_0 + b_1 v + b_2 v^2$$

NON-SEPARABLE MODELS

- The correlation of the random field is thus a product of two factors, the first, ψ_v^d depends on both the time and space, the second only on the time difference.
- Unfortunately, as Carroll *et al.* (1997) pointed out, this correlation function is not positive definite.
- Using results from the model, there were occasions when

$$Cov(Z_{st}, Z_{s't}) > Cov(Z_{st}, Z_{st}).$$

- This highlights a genuine lack of a rich set of functions that can be used as space–time correlation functions.

SUMMARY

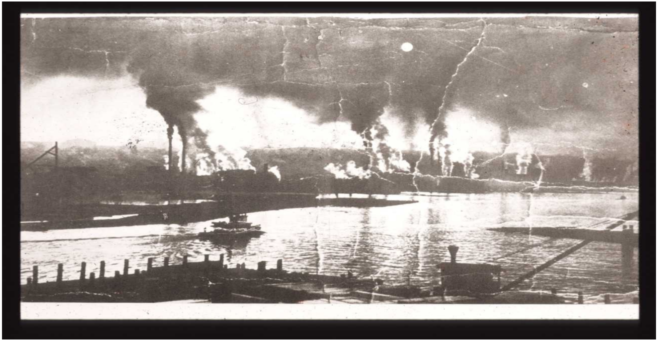
In this section we have seen the many ways in which the time can be added to space in order to characterise random exposure fields. In particular we have looked at the following topics:

- Additional power that can be gained in an epidemiological study by combining the contrasts in the process over both time and space while characterising the stochastic dependencies across both space and time for inferential analysis.
- Criteria that good approaches to spatio–temporal modelling should satisfy.
- General strategies for developing such approaches.
- Separability and non-separability in spatio–temporal models, and how these could be characterised using the Kronecker product of correlation matrices.
- Examples of the use of spatio–temporal models in modelling environmental exposures.

Better exposure measurements through better design

1. What do random environmental process fields look like?

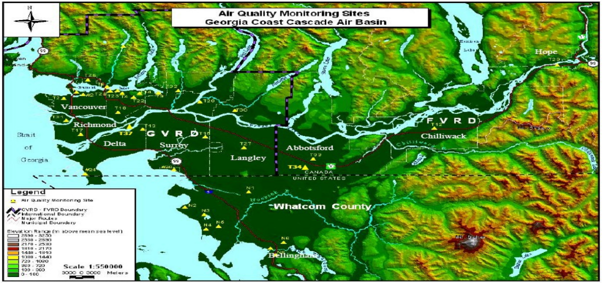
Vancouver's North Shore – Early 1900s



Vancouver's North Shore-Circa 1990

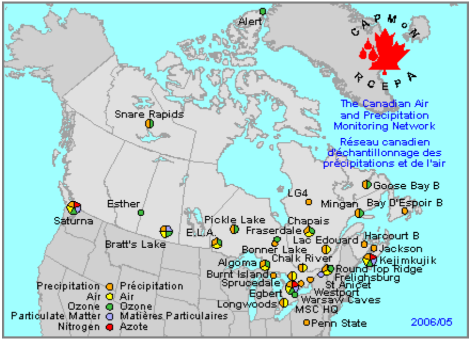
2. What do monitoring networks look like?

Metro Vancouver monitoriing network



NOTE: Detailed elevation data for the area north of the FVRD was unavailable at the time of map preparation. This gives the topography a flat appearance.

CAPMON NETWORK: ACID RAIN THEN AIR POLLUTION

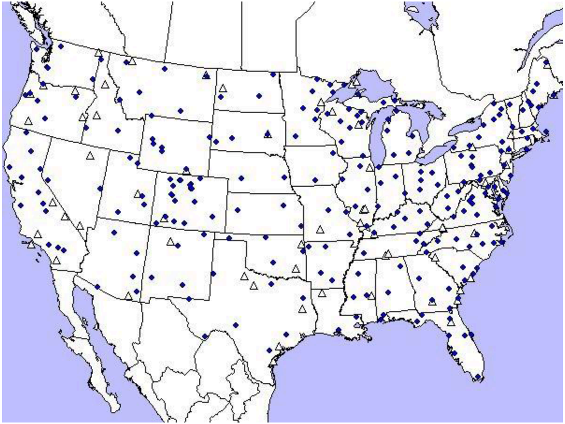


NOTES on Capmon:

- No sense an ‘optimal’ network for monitoring the environment.
- For administrative simplicity Capmon was a merger of three networks, each setup to monitor acid precipitation when that topic was fashionable.
- For simplicity, the sites were then adopted for other things, e.g, air pollution

THE NADP/NTN NETWORK: ACID PRECIPITATION

Monitors multivariate responses related to "acid precipitation"– another merger–better defined siting rules!



Rules governing siting and types of NDP/NTN monitors:

> *" The COLLECTOR should be installed over undisturbed land on its standard 1 meter high aluminum base. Naturally vegetated, level areas are preferred, but grassed areas and up or down slopes up to 15% will be tolerated. Sudden changes in slope within 30 meters of the collector should also be avoided. Ground cover should surround the collector for a distance of approximately 30 meters. In farm areas a vegetated buffer strip must surround the collector for at least 30 meters."* ⋮

3. What do monitors look like?

Monitoring shed in school yard

Monitoring Shed in a Park

4. Why do we need monitoring networks? To reduce uncertainty![1]

[1][Le and Zidek(2006), **?**]

WHY MONITOR?

General objectives:

- Measure process responses at critical points:
  - Near a new smelter using arsenic
- Enable predictions of unmeasured responses
- Enable forecasts of future responses
- Provide process parameter estimates
  - physical model parameters
  - stochastic model parameters eg. covariance parameters
- Address societal concerns

Specific objectives:

- Detect non-compliance with regulatory standards
- Enable health effect assessments to be made
  - & provide good estimates of relative risk
  - determine how well sensitive sub-populations are protected
  - can include all life, not just human
- To assess temporal trends
  - are things getting worse?
  - is climate changing?

Overall objective:

- To explore/reduce uncertainty
  - About aspects of environmental processes
  - One form of uncertainty (***aleatory***) cannot be reduced (outcome of fair die toss
  - the other (***epistemic***) (whether the die is fair) can increase or decrease. Implication: even an optimum design must be regularly revisited

Why monitoring design?

As air quality monitoring networks have proliferated over the past, so have doubts arisen as to where, when and what to monitor" (A.C. Stern, 1976).

Not much has changed in 40 years!

5. But what is uncertainty?

‘UNCERTAINTY’?

- Laplace: **Probability is the language of uncertainty**
- DeFinetti: **In life uncertainty is everything**
- Statisticians: **variance** or **standard error**
- Kolmorov & Renyi: **Entropy**

6. Designing networks with entropy

POSSIBLE DESIGN CRITERIA

"Gauge" (add monitors to) sites that

- maximally reduce uncertainty at their space-time points
  - measuring their responses eliminates their uncertainty
- best minimize uncertainty about unmeasured responses
- best inform about process parameters
- best detect **non-compliers**

APPROACHES TO DESIGNING MONITORING NETWORKS

Space-filling designs[2]**:** spead them out maximally over the domain of interest

Probability based designs [3] **:** pick locations at random

- simple random sampling
- stratified, multistage designs
- e.g. (1) EPA's survey of lakes; (2) the EMAP project

[2] [Nychka and Saltzman(1998)]
[3] [Mitch(1990)]

Model based designs:

- Regression modelling[4]
  - e.g. to estimate the slope of line put 1/2 the data at each end of the data range
- Random fields (prediction, e.g. entropy) approach

[4] [Müller(2007)]

Entropy based designs

- 'Gauges' sites with greatest 'uncertainty'
  - uncertainty = entropy
  - maximally reduces uncertainty about 'ungauged' sites
  - best estimates predictive posterior distribution under entropy utility
- Bypasses specification of objectives
- Has a long history[5]

Other designs. E.g. incorporate both of the latter, prediction and parameter estimation [6].

[5] General: [Good(1952)], [Lindley(1956)], [Shewry and Wynn(1987)]. Network design: [Caselton and Zidek(1984)],[Sebastiani and Wynn(2002)],[Zidek et al.(2000)Zidek, Sun, and Le], currently popular

[6] [Zhu and Stein(2006)]

CHALLENGES FACING THE DESIGNER:

- A multiplicity of valid objectives
- Unforeseen & changing objectives
- Multiple responses at each site: which to monitor?
- Must include prior knowledge & prior uncertainty
- Should use realistic process models. (How?)
- Must be integrated with existing networks
- Must deal with reality, e.g politicians, committees!!!

Use entropy approach–can deal with many challenges

7. Entropy basics

WHAT'S ENTROPY?

Let $p = P(E)$= probability an uncertain event E occurs (e.g. heads on possibly bent coin). That uncertainty reduces to 0 when outcome becomes known. Let the size of reduction be for some ϕ:

$\phi(p)$ if E occurs
$\phi(1-p)$ if not.

The expected reduction in uncertainty is then
$p\phi(p) + (1-p)\phi(1-p)$

Simple assumptions imply:

$$\phi(p) = -log(p)$$

Conclusion: reduction in uncertainty due to knowledge of E's occurrence (i.e. "**uncertainty**" about E) is the **entropy** for the two point distribution $(p, 1-p)$:

$$-p\log(p) - (1-p)\log(1-p)$$

RELATIVE ENTROPY

When is an entropy a "big" entropy?

Need a reference point. Complete uncertainty about the coin (how its to be tossed and so on) suggests a two point reference distribution $(q, 1-q)$ with $q = 1/2$.

Define the relative entropy as **Kullback-Leibler's measure of deviation** of $(p, 1-p)$ from its reference level $(q, 1-q) = (1/2, 1/2)$:

$I(p,q) = -p \log(p/q) - (1-p) \log\{(1-p)/(1-q)\}$

The reference distribution corresponds to a "state of equilibrium" in physics (thermodynamics).

ENTROPY FOR MULTIPLE EVENTS

$$I(p, q) = \sum_i p_i \log \{p_i / q_i\}$$

ENTROPY FOR CONTINUOUS VARIABLES

Start with $p_i \sim f(x_i)dx_i$ & $q_i \sim g(x_i)dx_i$ as approximations. Then as $dx_i \rightarrow 0$, this entropy converges to

$$I(f,g) = \int f \log (f/g)$$

.

Commonly $g \equiv 1$ (*units of* f). In any event, f/g is a unitless quantity. Moreover Jacobean cancels under transformations of x making entropy an "intrinsic" measure of uncertainty – not scale dependent.

8. Using entropy in design

USING ENTROPY IN DESIGN

Best handled in a Bayesian framework. Let:

- Y_f = process response vector at future time T+1 including all sites (gauged & ungauged)
- D = set of all available data upon which to condition and get posterior distributions
- h_1 & h_2 be baseline reference densities against which to measure uncertainty.
- Finally:

$$\begin{aligned} H(Y_f \mid \theta) &= E[-\log(f(Y_f \mid \theta, D)/h_1(Y) \mid D] \\ H(\theta) &= E[-\log(f(\theta \mid D)/h_2(\tilde{\theta})) \mid D] \end{aligned}$$

Then we get fundamental identity (**Exercise**):

$$H(Y_f, \theta) = H(Y_f \mid \theta) + H(\theta)$$

DESIGN GOAL

Add or subtract sites from an existing network. Let us **add new sites to an existing network**

- $Y_f = (Y_f^{(1)}, Y_f^{(2)})$ = all site responses at future time T +1
- $Y_f^{(2)}$ for sites currently gauged at time T
- $Y_f^{(1)}$ for sites currently ungauged at time T

DESIGN GOAL: Partition $Y_f^{(1)} = (Y_f^{(rem)}, Y_f^{(add)})$ at future time (T+1). Let

- $Y_f^{(rem)} \equiv U$: future ungauged sites
- $Y_f^{(add)} \equiv G$ future new network stations.

ENTROPY DECOMPOSITION THEOREM

Let $U = Y_f^{(rem)}; G = (Y_f^{(add)}, Y_f^{(2)}); Y_f = [U, G]$

Fundamental identity:

$$\text{TOT} = \text{PRED} + \text{MODEL} + \text{MEAS}$$

where

$$\begin{aligned} PRED &= E[-\log(f(U \mid G, \theta, D)/h_{11}(U)) \mid D], \\ MODEL &= E[-\log(f(\theta \mid G, D)/h_2(\theta)) \mid D], \end{aligned}$$

and

$$MEAS = E[-\log(f(G \mid D)/h_{12}(G)) \mid D].$$

Theorem: Maximizing MEAS=Minimizing MODEL + PRED

THE RESPONSE DISTRIBUTION

The model below is used in an R package: EnviroStat.
To apply it:

- transform the response if necessary to make its distribution symmetric
- subject out the regular components such as regional trends and regional seasonality
- fit a regional time series model to remove autocorrelation and leave whitened residuals

Let Y denote the resulting residuals. These now carry the spatial patterns.

Now assume for $T+1$ times and p sites both gauged and ungauged:

$$\mathbf{Y}^{(T+1)\times p} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma} \sim N(X^{(T+1)\times k}\boldsymbol{\beta}^{k\times p}, I_{(T+1)} \otimes \boldsymbol{\Sigma})$$

$$\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \boldsymbol{\beta}_0, F \sim N(\boldsymbol{\beta}_0, F^{-1} \otimes \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} \sim GIW(\Psi, \delta) \, \# \, \text{Generalized Inverted Wishart distribution}$$

THE FAMOUS BARTLETT DECOMPOSITION

With u meaning ungauged and g, gauged let:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma^{[u]} & \Sigma^{[ug]} \\ \boldsymbol{\Sigma}^{[gu]} & \boldsymbol{\Sigma}^{[g]} \end{pmatrix}$$

The Bartlett decomposition: $\boldsymbol{\Sigma} = \Xi \boldsymbol{\Delta} \Xi'$ where

$$\boldsymbol{\Xi} = \begin{pmatrix} I & \boldsymbol{\Sigma}^{[ug]}(\boldsymbol{\Sigma}^{[g]})^{-1} \\ 0 & I \end{pmatrix}$$

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Sigma}^{[u]} - \boldsymbol{\Sigma}^{[ug]}(\boldsymbol{\Sigma}^{[g]})^{-1}\boldsymbol{\Sigma}^{[gu]} & 0 \\ 0 & \boldsymbol{\Sigma}^{[g]} \end{pmatrix}$$

IMPLICATIONS OF THE GIW DISTRIBUTION

Let

$$\begin{aligned}\boldsymbol{\Sigma}^{[u|g]} &= \boldsymbol{\Sigma}^{[u]} - \boldsymbol{\Sigma}^{[ug]}(\boldsymbol{\Sigma}^{[g]})^{-1}\boldsymbol{\Sigma}^{[gu]} \\ \boldsymbol{\tau}^{[u]} &= (\boldsymbol{\Sigma}^{[g]})^{-1}\boldsymbol{\Sigma}^{[gu]}.\end{aligned}$$

Then $\boldsymbol{\Sigma} \sim GIW(\boldsymbol{\Psi}, \delta)$ implies with appropriate hyperparameters:

$$\boldsymbol{\Sigma}^{[g]} \sim GIW(\boldsymbol{\Psi}^{[g]}, \delta^{[g]})$$

$$\boldsymbol{\Sigma}^{[u|g]} \sim IW(\boldsymbol{\Psi}^{[u|g]}, \delta_0)$$

$$\boldsymbol{\tau}^{[u]} \mid \boldsymbol{\Sigma}^{[u|g]} \sim N\left(\boldsymbol{\tau}_{0\,u}, \boldsymbol{\Sigma}^{[u|g]} \otimes (\boldsymbol{\Psi}^{[g]})^{-1}\right)$$

PREDICTIVE DISTRIBUTION

For certain constaints c, d, l

$$(G \mid D, \mathcal{H}) \sim t_g \left(\mu^{[g]}, \frac{c}{l} \Psi_g, l \right).$$

$$(U \mid G, D, \mathcal{H}) \sim t_u \left(\mu^{[u|g]}, \frac{d}{q} \mathbf{\Psi}_{u|g}, q \right).$$

Conditional entropy for $U = Y_f^{[u]}$ that has to be partitioned into 'add' and 'rem' sites is

$$H[U \mid G, D] \quad = \quad \frac{p}{2} \log |\Psi_{u|g}| + \text{irrelevant terms}$$

Bartlett decomposition means choice of 'add' sites at time T+1 obtain from maximizing $|\Psi_{u|g}[add, add]|$, the sub–determinant of $\mid \Psi_{u|g} \mid$ corresponding to the 'add' sites in the partitioned U.

That will simultaneously minimize the entropy left in the 'rem' sites.

The "Add" computation

- **NP-Hard:** No exact algorithms for big networks
- **Inexact Methods:**
  - Greedy
  - Greedy + Swap
- **Exact Methods:**
  - Complete enumeration
  - Branch and bound

How many sites?

Compute:

$$Entropy/(Numberofsites)$$

as the number of sites varies. Eventually this reaches a maximum (bang for the sampling buck) and then declines. Indicates when to stop on redesign.

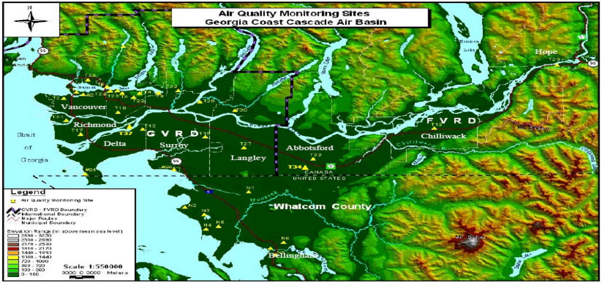
9. Case study in entropy design

REDESIGNING VANCOUVER'S AIR QUALITY NETWORK

Suppose hypothetically, Vancouver wishes to redesign its hourly PM_{10} **contentration field. The existing 10 monitoring sites are to be increased to 16 by selecting 6 new stations from among 20 possible sites.** Use the entropy approach.

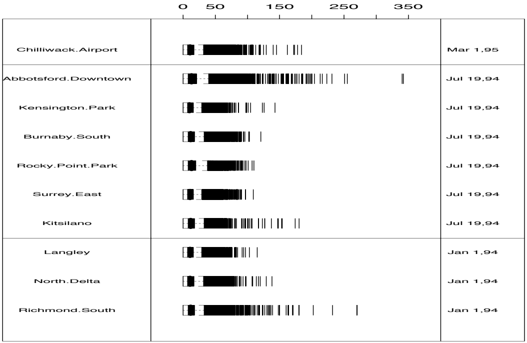
INTRODUCING THE SITES:

Metro Vancouver monitoriing network

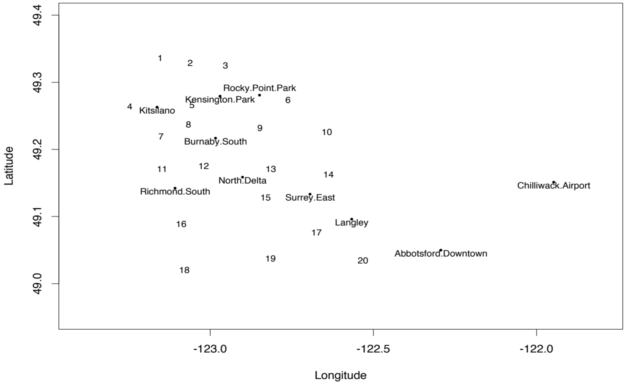


NOTE: Detailed elevation data for the area north of the FVRD was unavailable at the time of map preparation. This gives the topography a flat appearance.

The PM_{10} levels at the 10 existing stations in Metro Vancouver. Note differing startup times.



The 10 PM_{10} monitoring site locations (the ones with names) & prospective new locations.



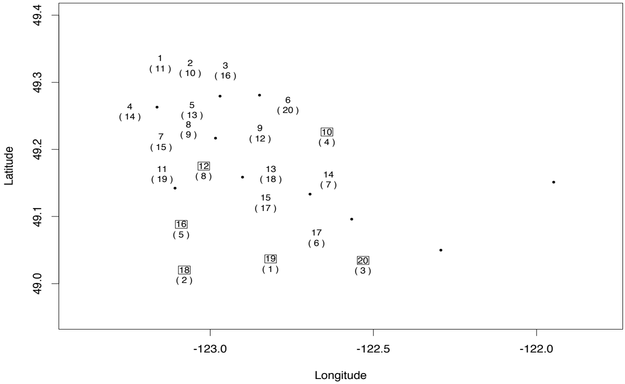
TECHNICALITIES

The response Y is the log-transformed hourly PM_{10} concentration.

The normal-generalized inverted Wishart predictive distribution is used

- **needs "whitened" residuals; space - time interaction $\rightarrow$ use of 24 (hour) dimensional multivariate AR(1) model**
- **different 10 - station startups $\rightarrow$ monotone ("staircase") data structure $\rightarrow$ generalized inverted Wishart distribution $\rightarrow$ different d.f. for each staircase step**
- **select the 6 new stations with jointly maximum conditional entropy**

Locations of the old and newly selected 'add' sites (square brackets). The ranks of the 20 sites by estimated variance is in curved brackets.



10. New Frontiers: Preferentially selection of sites

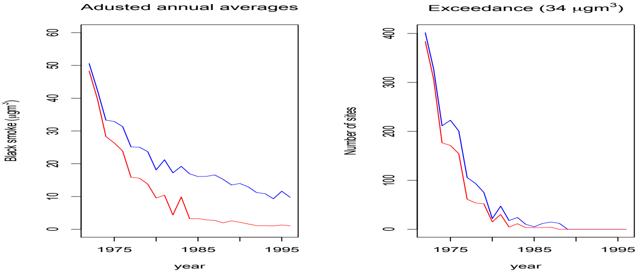
PREFERENTIAL SELECTION

Preferential selection: occurs when site locations are determined stochastically by the environmental field responses they are supposed to measure e.g. where the responses are expected to be high.[7]

Preferential site selection can be undesirable: when the data are meant to represent the random environment field of interest. For then estimates will be biased.

[7] [Diggle et al.(2010)Diggle, Menezes, and Su]

Example: There is conclusive evidence [8] that as the United Kingdom's network for monitoring black smoke in the United Kingdom was reduced in size to reflect the decreased burning of coal cuts were made preferentially [9] . The Figure shows how much aggregate estimates of levels were overestimated.



[8] [Shaddick and Zidek(2014)]
[9] [Zidek et al.(2014)Zidek, Shaddick, and Taylor]

But prefential site selection can also be desirable: When the objective is specific, e.g. near "sources" to measure mercury emissions[10]

[10] [Schmeltz et al.(2011)Schmeltz, Evers, Driscoll, Artz, Cohen, Gay, Haeuber, Krabbenhoft, M

Example: noncompliance with regulatory standards

Networks may need to be set up to detect unduly high levels of a air pollution field.[11].

But that objective is not well–defined[12]**.**

What does it mean? The field is random so best design may vary from day–to–day.

- Should you take a simulated future day? An average day? A bad day?

What design strategy should you use? Should we:

- monitor the sites most likely to comply?
- or do not monitor the sites least likely to comply? (leads to a different design)
- what do you do with existing sites?

[11] [Guttorp and Sampson(2010)]
[12] [Chang et al.(2007)Chang, Fu, Le, and Zidek]

How well does an entropy designs work for this purpose?

Example: how well would Vancouver's 6 site, entropy-based, addition compare to an optimal noncompliance based addition?

Using data from the 10 stations with hourly data for PM_{10} on February 28, 1999 we can use the EnviroStat package to predict/simulate the potential new sites in Metro Vancouver repeatedly and hence the distribution of the daily maximum PM_{10} level over the region. Thus we can find the the six sites that maximize the probability of detecting noncompliance over the region:
argmax PR{daily max PM_{10} $\mathbf{Y}^{6added} \geq$ 50 $(\mu g\ m^{-3})$}

Results: The top 7 choices of six new sites are identical to the top 7 chosen by the entropy criterion on Feb 28, 1999!

But the entropy criterion does not work as well on Aug 1, 1998 for detecting noncompliance that day.

The entropy criterion was not meant to be best for every specified purpose. But then the criterion above gives different designs for different days!

11. New frontiers: Designing for extreme values

MONITORING EXTREME VALUES

Regulatory criteria usually involve moderate extremes like those in the previous example. EnviroStat can be used to simulate/predict them however complex.

Example: EPA'S PM_{10} criterion:
For particles of diameters of 10 micrometers or less:

Annual Arithmetic Mean: $\leq 50\mu g/m^3$

24 - hour Average: $\leq 150\ \mu\text{g m}^{-3}$

Meaning: three year average of 98-th annual percentiles of 24 hour averages must be ≤ 150 ($\mu g\ m^{-3}$) at all sites in an urban area. Complex metric $\Rightarrow$ need predictive distribution to simulate its distribution!

SOME CHALLENGES WITH EXTREMES

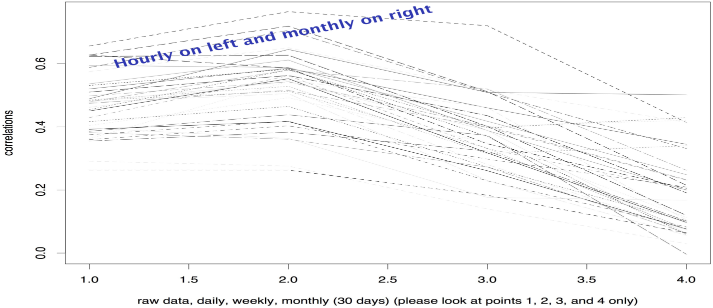
- **Seldom enough data:** spatial and temporal.
- **Generally extremes have small inter–site dependence** than raw reponses–spatial prediction is hard.
- **But some site pairs have strong intersite correlation:** marks it hard to model with classical extreme value theory
- **Multivariate extreme value distributions are not tractable** for large geographical domains with lots of sites:
  - conditional computation (e.g. entropy) difficult
  - simulating extreme fields hard
- **Design objective is elusive.**

Example: Intersite correlations - Vancouver PM_{10} field

Call n in $M_n = max\{Y_1, \ldots, Y_n\}$ the span of the maximum. Vancouver (and London) analyses show that **as n increases from $n = 1$ hour, the inter–site dependence declines with most site pairs but not all of them**.

EXAMPLE: INTERSITE CORRELATIONS - VANCOUVER PM_{10} FIELD

Demonstrated by the following Figure that shows results for Vancouver's PM_{10} **intersite correlations for successive maxima time spans n: hour, 24 hours, week of hours and a month of hours:**

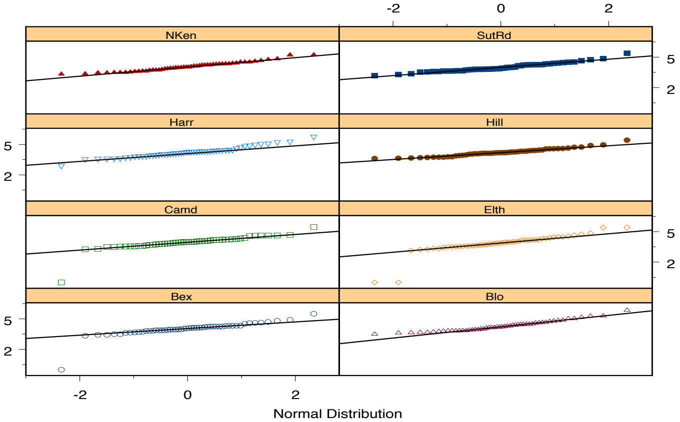


A POSSIBLE SOLUTION

Approximate the joint distribution of extremes by a log multivariate-t distribution. Hence:

- has convenient conditional, marginal distributions
- can accommodate existing sites and historical data
- can permit simulation of complex metric distributions
- is tractable with computable entropy's, regression models, etc
- bypasses need to specify specific design objectives

Empirical results: log multivariate–t distribution as approximation to joint distribution of extremes field. QQplots for weekly maxima of hourly log PM_{10} London 1997 data: **here even the marginal normal to approach extreme value distribution works well**:



Another example involving Canadian climate extremes:

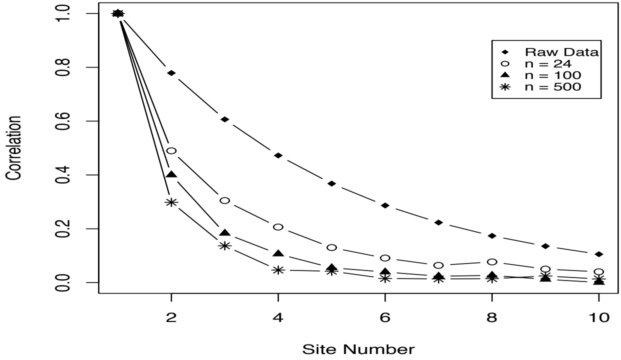
Empirically shows well–calibrated 95% (etc) prediction intervals and supports use of multivariate approximation.

Credibility Level	Mean	Median
30%	35	35
95%	96	97
99.9%	99.9	1

Table: **Summary of coverage probabilities at different credibility levels for the simulated precipitation data over 319 grid cells, Canadian Climate Model**

EFFECT OF SPAN FOR THE T-DISTRIBUTION[13]:

Simulation study: 10 sites with decreasing intersite correlation. Multi–normal responses. Varying span of maxima computed at individual sites. (Result for t is similar.)



13[Chang et al.(2007)Chang, Fu, Le, and Zidek]

SUMMARY

Design theory:

- **Design-a much neglected subject** in statistics and one that cannot be done by software alone, thus one of the few remaining domains where statisticians still rule.
- **Research opportunities** abound in spatial design.
- **The entropy design** is the one robust approach when goals cannot be specified precisely or future uses cannot be anticipated. Needs to be extended to non–Gaussian distributions.
- More work is needed on designing networks for **fields of extremes**. Also when interest lies in the **extremes of fields**.

Designing to measure the "metrics"

- Complex metrics seen in environmental epidemiology are complex for a reason. Need a predictive spatial distributions to simulate the distributions of these metrics, which often reflect levels north of whichhuman health is threatened.
- Complex metrics involve moderate extremes whose intersite correlations will tend to small. Hence the current urban monitoring networks are likely to be insufficiently dense to monitor the field of extremes.
- Urban monitoring sites seem likely to have selected preferentially where response levels are high. More assessment of these networks is needed.

Monitoring for extremes

- Metrics involve moderate but not ‘extreme’ extremes. **Designs for monitoring these extremes** is an important but unstudied topics.
- More attention to **design criteria for the field of extremes** needed

Modelling point patterns

Log Gaussian Cox Processes

SPATIAL POINT PROCESS MODELLING

Aim

- describe and model the spatial structure formed by the locations of individuals in space

specifically:

- develop methodology that is suitable for complex biological data

A SPATIAL PATTERN...

- are the daisies randomly distributed in the lawn of our garden???

issues

- what do we mean by “random"? formal description?
- what if they are not random?
- how should we describe and model non-random patterns?

SPATIAL POINT PROCESS MODELLING

Aim

describe and model the spatial structure formed by the locations of individuals in space

- is a pattern likely to be “random"?
- find a formal description of “randomness" – a suitable statistical **model**
- find a formal description of “non-randomness"

What do you mean by “random pattern"?

complete spatial randomness (CSR)::

points are **independently** scattered in space

SPATIAL POINT PROCESS MODEL

- captures the characteristics of spatial point patterns in a finite number of parameters
- provides a mechanism for generating **spatial point patterns**; all have the same spatial characteristics
- the locations of any objects can be modelled – plants, animals, stars, cells, cities...

Poisson process

a homogeneous Poisson process X with constant intensity λ_0 (number of points per unit area) has two properties:

(1) density of points is **constant**

(2) the location of any point in the pattern is independent of the location of any of the other points

$\Rightarrow$ formal way of describing "random" patterns; complete spatial randomness (CSR)

SPATIAL POINT PROCESS MODELS

- summary statistics have been applied in the literature to describe deviations from CSR
- applications of models much rarer – but often better suited for complex data sets
- statistical models for patterns that deviate from the Poisson case

$\Rightarrow$ more general spatial point process models

SPATIAL POINT PROCESSES

- complicated mathematical object
- why?
  point pattern differs from a standard dataset – number of observed points is random
- mathematically complicated concept of a **random measure** has to be used.

SPATIAL POINT PROCESSES

generalisations of the Poisson case

- inhomogeneous Poisson process – inhomogeneous intensity $\lambda(s), s \in \mathbb{R}^2$
- Cox processes – random intensity $\Lambda(s), s \in \mathbb{R}^2$
- Markov point processes – local interaction among individuals

We will focus on **Cox processes**!

SPATIAL MODELLING – A CLASS OF MODELS

Cox processes are spatial point processes with **random** intensity (density of points);
log Gaussian Cox processes depend on a (continuous) random field

$$\Lambda(s) = \exp\{Z(s)\},$$

where $\{Z(s) : s \in \mathbb{R}^2\}$ is a Gaussian random field.

- very flexible class of models
- used to be **really hard** to fit

but: given the random field (i.e. a latent field !), the points are independent (Poisson process)

$\Rightarrow$ we can use INLA (hurrah!)
$\Rightarrow$ for INLA the continuous field is approximated by a (discrete) Gauss Markov Random Field

The likelihood

The likelihood (in the most boring case) is

$$\log(\pi(Y|\eta)) = |\Omega| - \int_{\Omega} \Lambda(s)\, ds + \sum_{s_i \in Y} \Lambda(s_i),$$

where Y is the set of observed locations and $\Lambda(s) = \exp(Z(s))$, and $Z(s)$ is a Gaussian random field.

- likelihood is analytically intractable; requires the integral of the intensity function, which cannot be calculated explicitly
- the integral can be computed numerically; computationally expensive

INLA AND COX PROCESSES

- nicely, we can fit Cox processes elegantly with INLA – and it's fast
- we will see how this is done using some simple examples now
- followed by some examples to give you a taste of what is coming...
- later we will discuss how more complex models can be fitted with INLA (and it is still fast!)

GRIDDING

- to fit data a log Gaussian Cox process in INLA we need to grid the data (regular grid)[14]
- the observation window is discretised into $N = n_{row} \times n_{col}$ grid cells $\{s_{ij}\}$ with area $|s_{ij}|$, $i = 1, \ldots, n_{row}, j = 1, \ldots, n_{col}$
- y_{ij} the observed number of points in grid cell $\{s_{ij}\}$
- conditional on the latent field η_{ij} counts are independent

$$y_{ij}|\eta_{ij} \sim \text{Poisson}(|s_{ij}| \exp(\eta_{ij})),$$

[14] We can also fit point process models with an SPDE models (see Dan's lectures) but this is beyond the scope of this course.

A SIMPLE POINT PROCESS MODEL...

- model the point pattern as Poisson counts on a grid
- a very simple model assumes an underlying (but unobserved) spatial trend with an error term
- no covariates
- the latent field is then modelled by

$$\eta_{ij} = \alpha + f_s(s_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, n_{\text{row}}, \quad j = 1, \ldots, n_{\text{col}},$$

  α is an intercept
- $f_s(s_{ij})$ denotes a spatially structured effect reflecting the value of an unobserved spatial trend in grid cell s_{ij}
- models spatial autocorrelation

So what is the magical $f_s(s_{ij})$, i.e. the spatial model here?

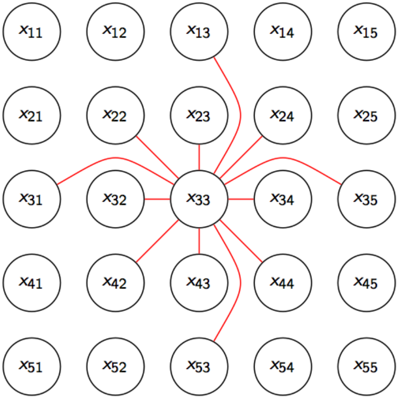
TWO-DIMENSIONAL RANDOM WALK (RW2D)

- in INLA point processes fitted on a grid uses a two-dimensional random walk model as spatially structured effect
- formally, it is an intrinsic Gauss Matkov random field (IGMRF)
- it has a Markov property; the value in each cell that only depends on the values in neighbouring cells
- *only* works on a *lattice*

```
inla(formula = y~...+f(index, nrow=100, ncol=100), ...)
```

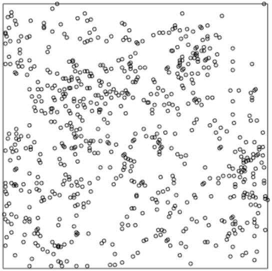
TWO-DIMENSIONAL RANDOM WALK (RW2D)

- constructed to produce a smooth field by penalizing local deviation from a plane
- conditional mean is the weighted average of neighbours
- closer neighbours get a larger weight than those further away



FITTING A SIMPLE MODEL IN INLA

fit the simple model to the spatial point pattern formed by the species *Andersonia heterophylla*

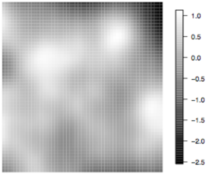


FITTING A SIMPLE MODEL IN INLA

The model is specified in `R-INLA` using the formula-call

```
formula = Y ~ 1
+ f(index, model="rw2d", nrow=nrow, ncol=ncol, ...)
+ f(J, model="iid",hyper=hyper.error)
```

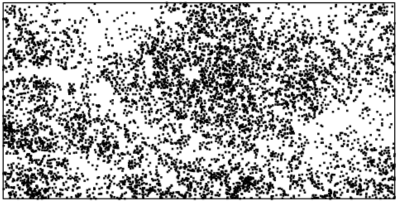
Estimated spatially structured effect:



- smoothness of the spatial field? – prior choice!
- how do we include covariates?

RAINFOREST DATA- SPATIAL PATTERN RELATIVE TO COVARIATES

- 50 ha forest dynamics plot at Pasoh Forest Reserve (PFR), Peninsular Malaysia; never logged
- e.g. species *Aporusa microstachya*; 7416 individuals
- we can model dependence of tree occurrence on soil or topography covariates

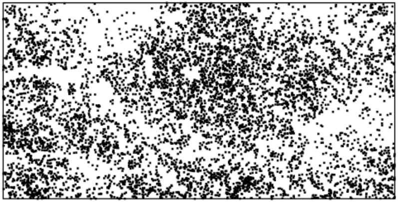


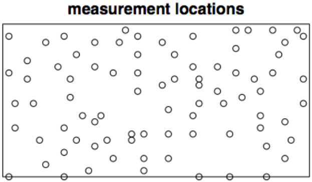
RAINFOREST DATA- SPATIAL PATTERN RELATIVE TO COVARIATES

we can also fit

- a **joint model** to two (or more) spatial patterns (accounting for shared environmental preferences)
- or a joint model of covariates AND the pattern (accounting for measurement error)

$\Rightarrow$ INLA with multiple likelihoods





KOALA DATA – MARKED POINT PATTERN DATA

- study conducted at the Koala Conservation Centre on Phillip Island, near Melbourne, Australia, 1993 - 2004
- $\approx$ 20 koalas present in the reserve at all times throughout study; reserve enclosed by a koala-proof fence
- koalas feed on eucalyptus leaves which are toxic to most animals; koalas have adapted to this

Do the koalas **feed selectively**, i.e. do they choose trees with the least toxic/ most nutritious leaves?

COMPLEXITY – MARKED POINT PATTERN

marks for each tree:

- mark 1: leaf samples taken from each eucalyptus tree and analysed for palatability
- mark 2: tree use by individual koalas collected at monthly intervals between 1993 and March 2004

fit joint model to:

- tree locations depend on (unobserved) soil nutrients levels and local clustering
- palatability depends on spatial pattern (through soil nutrients levels)
- koala visitation depends on spatial pattern, palatability

$\Rightarrow$ INLA with multiple likelihoods

Thymus carnosus – REPLICATED AND MARKED POINT PATTERN

- perennial herbaceous herb
- endemic to the Iberian Peninsula
- endangered and protected species
- aim: understand threats to survival of thymus plants to support conservation

INCLUDE MARKS IN ANALYSIS

- marks may provide additional information on underlying dynamics
- use **joint model** of point pattern and marks (health status) to distinguish short and long term survival

objective: use marks to distinguish processes operating at different **temporal scales**

ANIMAL STUDIES – VARYING EDGES

- edges in rainforest data have been arbitrarily chosen
- often different for data on animals
- tendency to use "natural borders" as edges

here: locations muskoxen herds in study area in Greenland

- assumption of "artificial" edges does not hold
- some edges are "true" edges; **impact** on the pattern

⇒ point pattern with varying edges...

SPATIO-TEMPORAL POINT PROCESS MODELS

- main interest in the muskoxen study (and many other studies) is to understand the changes in spatial behaviour over time – climate change
- (the few) spatio-temporal point processes discussed in the literature assume independence of space and time for computational reasons
- recent work (Lindgren et al. 2011) allows the fitting of spatio-temporal point processes that do **not** make this assumption
- in combination with INLA approach, this is computationally feasible

PREFERENTIAL SAMPLING

Preferential sampling occurs when the sampling locations are **stochastically dependent** on the thing that you are sampling.

- Not all inhomogeneous designs are preferential
- Classic example: Fish counts using commercial trawlers. They will only look where they believe there are fish.
- We can consider the sampling locations X as a realisation of a **point process** with intensity that depends on the unknown object of interest.
- Ignoring preferential sampling can be lead to biased estimates!

THE PREFERENTIAL SAMPLING MODEL OF DIGGLE ET AL. 2010

- $S(x), x \in \mathcal{W}$ is a stationary Gaussian process
- $X|S$ is a non-homogeneous Poisson process with intensity

$$\lambda(x) = \exp\{\alpha + \beta\, S(x)\}$$

- Conditional on S and X, Y is a set of mutually independent Gaussian variates:

$$Y_i|S(X) \sim \mathcal{N}(\mu + S(x_i), \tau^2)$$

Notation

We divide the domain $\mathcal{W}$ into N disjoint cells
Then for each cell i, $i = 1, \ldots, N$

$$
\begin{aligned}
S_i &= \text{Value of } S \text{ in cell } i \\
x_i &= \begin{cases} 1 & \text{if cell } i \text{ is observed} \\ 0 & \text{otherwise} \end{cases} \\
y_i &= \begin{cases} \text{Observed value at cell } i & \text{if cell } i \text{ is observed} \\ \text{NA} & \text{otherwise} \end{cases}
\end{aligned}
$$

LIKELIHOOD

We consider an extended data set $(y_1, \ldots, y_N, x_1, \ldots, x_N)$ with likelihoods:

$$\begin{aligned} x_i|\eta_{1i} &\sim \text{Po}\{\exp(\eta_{1i})\} \\ y_i|x_i = 1, \eta_{2i} \neq 0 &\sim \mathcal{N}(\eta_{2i}, \tau^2) \end{aligned}$$

where the linear predictors η_{1i}, η_{2i} are defined as

$$\begin{aligned} \eta_{1i} &= \alpha + \beta\, S_i^* \\ \eta_{2i} &= \mu + S_i \end{aligned}$$

LATENT GAUSSIAN FIELD

We assume $\boldsymbol{S} = (S_1, \ldots, S_N)$ to be a GMRF representation of a Matérn field with fixed range and unknown precision τ_S:

$$\boldsymbol{S} \sim \text{GMRF}(\tau_S)$$

While $\boldsymbol{S}^* = (S_1^*, \ldots, S_N^*)$ is a “copy” of $\boldsymbol{S}$

$$\boldsymbol{S}^* = \boldsymbol{S} + \boldsymbol{\epsilon}; \boldsymbol{\epsilon} \sim \mathcal{N}(0, 10^{-6}\boldsymbol{I})$$

We assume also

$$\mu, \alpha \sim \mathcal{N}(0, 10^{-6})$$

So that

$$(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \alpha, \mu, \boldsymbol{S}, \boldsymbol{S}^*)|\beta, \tau_S \text{ is Gaussian}$$

The preferential sampling model as a latent Gaussian model

- The data

$$(y_1, \ldots, y_N, x_1, \ldots, x_N)$$

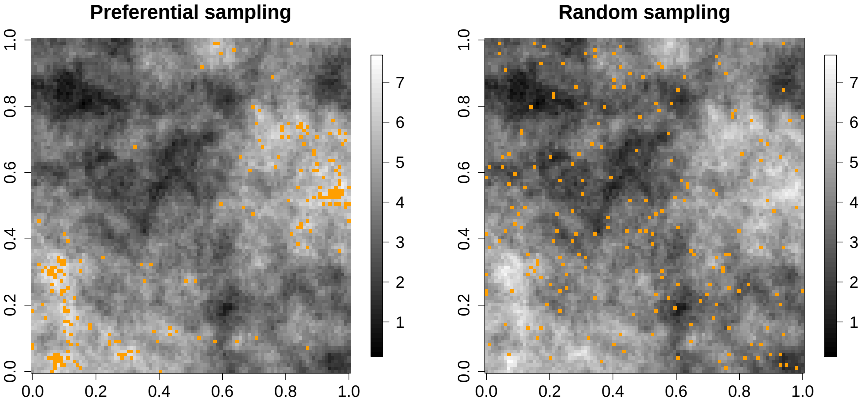
- The latent Gaussian model

$$(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \alpha, \mu, \boldsymbol{S}, \boldsymbol{S}^*)$$

- The hyperparameter vector

$$\boldsymbol{\theta} = (\beta, \tau, \tau_S)$$

A simulated experiment on a 100 × 100 grid



The value for the preferential parameter is $\beta = 2$.

IMPLEMENTING THE MODEL USING INLA
THE DATA FRAME

```
> yy                         > mu > alpha > ii   > jj
               [,1] [,2]
 [1,]            NA   NA      1      0      1      NA
 [2,]            NA   NA      1      0      2      NA
 [3,]       6.41564   NA      1      0      3      NA
 [4,]            NA   NA      1      0      4      NA
   .
   .
[10001,]         NA    0      0      1      NA     1
[10002,]         NA    0      0      1      NA     2
[10003,]         NA    1      0      1      NA     3
[10004,]         NA    0      0      1      NA     4

> data = list(yy=yy, mu=mu, ii=ii, jj=jj, alpha=alpha)
```

Implementing the model using INLA
The R code - Fitting a RW2D

```
formula = yy ~ alpha + mu +
      f(ii, model = "rw2d", nrow=nrow, ncol=ncol,
      constr=TRUE, bvalue=1) +
      f(jj, copy="ii", fixed=FALSE) -1

model = inla(formula, family = c("gaussian", "poisson"),
             data = data, verbose = TRUE)
```

Implementing the model using INLA
The R code - Fitting a Matérn Field

```
formula1 = yy ~ alpha + mu +
     f(ii, model = "matern2d", nrow=nrow, ncol=ncol,
       initial = c(3, log(10)),
       fixed=c(FALSE,TRUE),constr=TRUE) +
     f(jj, copy="ii", fixed=FALSE) -1

model1 = inla(formula1, family = c("gaussian", "poisson"),
              data = data, verbose = TRUE)
```

Range of the Matérn field is fixed at 10.

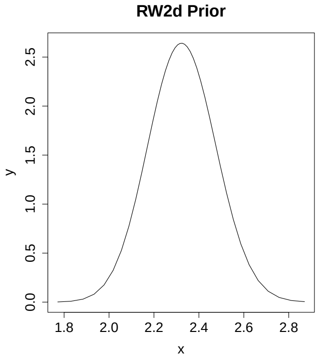
RESULTS

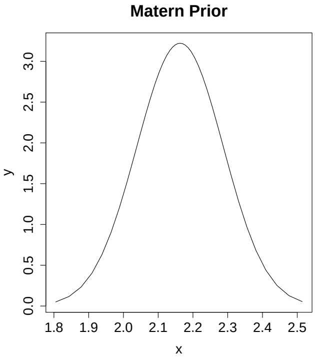
Computational time

- from about 2 to 10 minutes depending on the approximation type and machine.
- Just seconds for a 50×50 grid.

RESULTS

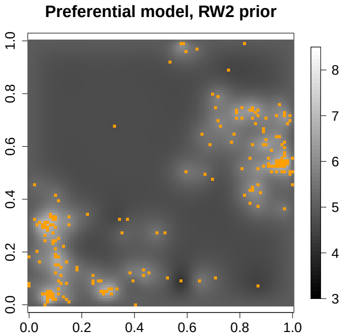
Posterior estimate for β





RESULTS

Posterior mean for the S field



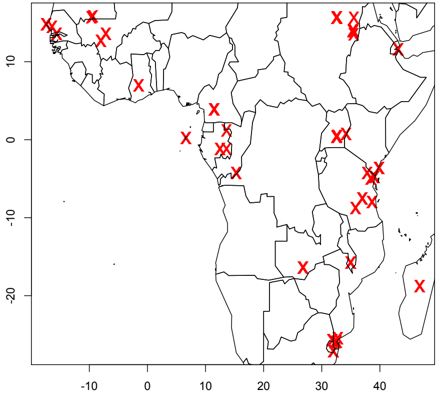
AN UNFORTUNATE FACT

At risk of disappointing you...

Just because you have data and a question, doesn't mean that the data can answer that question!

- The *best* statistics infers the answer to a question from data *specifically and carefully* collected to answer that question
- This is obviously not always possible, but we should do our best!
- For easy problems (differences of means, ANOVAs etc), there are well-known ways to do this
- In this session, we will have a look at some simple (and some practical) aspects of spatial experimental design

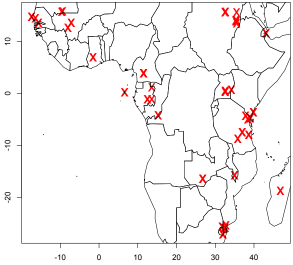
THE BAD NEWS: UNDESIGNED SPATIAL DATA MAY NOT ANSWER THE QUESTION



Question: Can we build a spatial map of sero-prevelance of a certain strain of malaria throughout Africa.
Answer: No.

What went wrong

- Data: $(n_{\text{test}}, n_{\text{present}})$
- Model: Binomial (low information!)
- Sampling locations are far apart
- Essentially uncorrelated!
- Low power, high uncertainty.



How this manifested

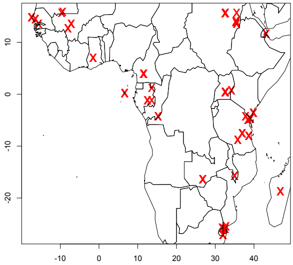
The Folk Theorem

If your computation breaks, the problem is usually your model.

- INLA assumes that there is enough information in your model to resolve all of the parameters
- If there isn't, it can break!
- That's what happened here!

What does a good spatial design look like?

- Sampling locations cover region of interest **(needed for prediction)**
- Sampling locations are close enough together that there is correlation **(hard to know beforehand)**
- Sampling locations are clustered (**needed for parameter estimation**)



THIS CAN BE HARD!

Partial answer:— **Sequential design**

- Begin with an initial set of sampling locations
- Compute the posterior
- Add a new location in the **best** un-sampled location
- Best = "lowest variance", "locally lowest variance", "most valuable-of-information"

NB: The over-all design here is preferential!

Joint modelling for point patterns

Outline

- Describe framework in fitting a joint model to a set of point patterns.
- Discuss three different case studies.

Example

1. Marked point patterns for the plant species*Thymus Carnosus*, observed in six different plots.
2. Koalas!
3. Muskoxen at Greenland, observed at different time points within the same study area.

FRAMEWORK

- Assume a set of spatial point patterns

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T,$$

observed within bounded regions $\Omega_t \in \mathbb{R}^2$.

- Each pattern

$$\boldsymbol{x}_t = \{x_{t1}, \ldots, x_{tn_t}\}.$$

is regarded as a realisation from a random spatial point process $\mathbf{X_t}$, where n_t is the number of points.

THE LOG-GAUSSIAN COX PROCESS

- Define random intensities

$$\Lambda_t(s) = \exp\{\eta_t(s)\}$$

where $\{\eta_t(s) : s \in \Omega_t \in \mathbb{R}^2\}$ is a Gaussian random field.

- Conditional on the random intensities

$$\boldsymbol{X_t} \mid \Lambda_t(s) \sim \text{Poisson}(\exp(\eta_t(s)))$$

THE LATTICE-BASED APPROACH

- Define

$$\begin{aligned} s_{ti} &: \text{ Grid cell } i \text{ in } \Omega_t \\ y_{ti} &: \text{ Number of points in grid cell } s_{ti} \text{ for pattern } \mathbf{x}_t \\ \eta_{ti} &: \text{ Representative value of the Gaussian field for pattern } \mathbf{x}_t \\ &\quad\ \text{in cell } s_{ti}. \end{aligned}$$

- Point patterns are assumed conditionally independent

$$y_{ti}|\eta_t(s_{ti}) \sim \text{Poisson}(|s_{ti}|\exp(\eta_t(s_{ti}))).$$

- Special case: $\Omega_t = \Omega$ for all t such that $s_{ti} = s_i$.

Fitting a joint model

- Each point pattern might be too small to make sensible model.
- Fit joint model to several point patterns:

$$\eta_{ti} = \alpha_t + \sum_{j=1}^{n_\beta} \beta_j z_{tji} + \sum_{k=1}^{n_f} f_k(c_{tki}) + \epsilon_{ti}, \quad t = 1, \ldots, T.$$

Estimation based on several point patterns

Use all of the point patterns to:

- Estimate fixed linear effects of covariates, that is the parameters $\beta_1, \ldots, \beta_{n_\beta}$.
- Estimate non-linear random effects of covariates, that is the underlying smooth functions $f_1, \ldots, f_{n_f}$.
- Account for dependencies/variation between different patterns.

ESTIMATION BASED ON SEVERAL POINT PATTERNS

Use all of the point patterns to:

- Estimate fixed linear effects of covariates, that is the parameters $\beta_1, \ldots, \beta_{n_\beta}$.
- Estimate non-linear random effects of covariates, that is the underlying smooth functions $f_1, \ldots, f_{n_f}$.
- Account for dependencies/variation between different patterns.

In `R-INLA`:

The joint model is fitted just stacking the responses and covariate terms in vectors.

UNDERSTANDING MARKED POINT PATTERNS

distinguish

a) different types of marks
b) different roles of marks

a) is obvious

- qualitative marks (species, age-groups, infected vs. non-infected...)
- quantitative marks (size, age, chemical properties...)

b) is harder...

MARKED POINT PATTERNS

different roles of marks

(i) models of the *pattern* that take the marks into account:
aim is to use marks to “explain" the pattern

(ii) models of the *marks* in a point pattern:
aim is to model the marks – often along with the pattern (!)

UNDERSTANDING MARKED POINT PATTERNS

for qualitative marks

(i) models of the *pattern* that take the marks into account:

"superposition"

- consider several (sub-)patterns formed by different types of points
- different subpatterns have been generated by separate (but not necessarily independent) mechanisms
  example: pattern formed by a multi-species plant community

(ii) models of the *marks* in a point pattern

"labelling"

- consider a single pattern with different (qualitative) characteristics
- some underlying mechanisms have lead to different qualitative properties of the points
  example: pattern formed by a single species but individuals have been affected or not affected by a disease

UNDERSTANDING MARKED POINT PATTERNS

for quantitative marks

(i) models of the *pattern* that take the marks into account:
   - very rarely looked at
   - difficult...

(ii) models of the *marks* in a point pattern
   - we'll discuss an example tomorrow
   - we'll actually discuss an example with several (non-independent) marks

EXAMPLES

qualitative marks

- superposition – revisiting the rainforest
- superposition – a joint model
- labelling – replicated patterns

quantitative marks

- revisiting the joint model – and the koalas

THE RAINFOREST REVISITED...

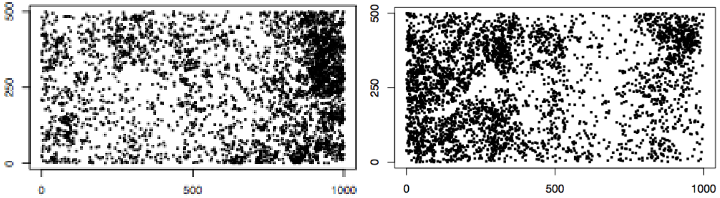
superposition

- bivariate point pattern of two different rainforest tree species, the species *Protium tenuifolium* (“species 1”) and *Protium panamense* (“species 2’)
- model the spatial pattern formed by pairs of species given spatial covariates
- model both species within a single model

THE RAINFOREST REVISITED...

superposition

- bivariate point pattern of two different rainforest tree species, the species *Protium tenuifolium* and *Protium panamense*



TWO SPECIES – CO-OCCURENCE

- the co-occurrence of the two species might be due to interaction between individual trees (local interaction) or due to shared environmental preferences
- these two effects are likely to be strongly confounded
- approach:
  - fit a **joint model** to the two species
  - use a **shared spatial effect**, representing the shared environmental preferences

A JOINT MODEL

- have two latent fields

$$\eta_{ij}^{(1)} = \alpha_1 + \sum_{p \in \mathcal{I}} \beta_{1p} z_p(s_{ij}) + \beta_{s1} f_s(s_{ij}),$$

$$\eta_{ij}^{(2)} = \alpha_2 + \sum_{p \in \mathcal{I}} \beta_{2p} z_p(s_{ij}) + \beta_{s2} f_s(s_{ij})$$

- note that $f_s(s_{ij})$ appears twice!

A JOINT MODEL – SPECIFICATION IN R−INLA

- a joint model for the two species basically means that we have **two response variables**

- R−INLA requires that these are stored as the column vectors of a $2n \times 2$ **matrix** B, where $n = n_{cells}$ is the number of grid cells
- first column: counts for the first species in entries $1, \ldots, n$ and NAs otherwise
- second column: NAs in entries $1, \ldots, n$ and counts for the second species in entries $n + 1, \ldots, 2n$
- we will revisit the principle a few times...

A JOINT MODEL – SPECIFICATION IN `R-INLA`

- a joint model for the two species basically means that we have **two response variables**

- we also have an area vector E of length $2n$ with the general intensity of point pattern 1, in entries $1, \ldots, n$ and that of species 2 in entries $n + 1 \ldots, 2n$
- also separate vectors $i.spat1$, $i.spat2$ for the spatial effects, containing NAs in entries $n + 1, \ldots, 2n$ for species 1 and in entries $1, \ldots, n$ for species 2
- similarly for the covariates
- and a vector `mu` with `1` in $1, \ldots, n$ and `2` in entries $n + 1 \ldots, 2n$ to get separate intercepts

A JOINT MODEL

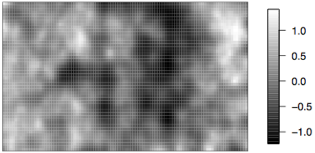
Specification in `R-INLA` becomes a bit more lengthy...

```
formula = B ~ mu - 1 + z11 + z12 + z21 + z22
+ f(i.spat1, model = "rw2d", ...)
+ f(i.spat2, copy = "i.spat1", fixed=FALSE, ...)
```

- `-1` tells INLA to not fit a joint intercept but to estimate two separate intercepts
- specify the two different likelihoods when calling `inla`:

```
result = inla(formula, family=c("poisson","poisson"),...
```

spatial effect (for both species)



INTERIM CONCLUSIONS I

- fitting a **joint model** in `R-INLA` can be done easily
- and it still doesn't take too long...
- we fitted a joint model to the two species
- the shared spatial effect accounted for shared environmental preferences
- this is still a rather boring model...

INTERIM CONCLUSIONS II

$$\eta_{ij}^{(1)} = \alpha_1 + \sum_{p \in \mathcal{I}} \beta_{1p} z_p(s_{ij}) + \beta_{s1} f_s(s_{ij}); \;\; \eta_{ij}^{(2)} = \alpha_2 + \sum_{p \in \mathcal{I}} \beta_{2p} z_p(s_{ij}) + \beta_{s2} f_s(s_{ij})$$

- two species have a correlation of +1 or -1 everywhere in space we can construct more complex multi-type models:
  - models where the two patterns have a correlation $\rho \in [-1, 1]$ but otherwise similar spatial autocorrelation range
  - models where the two patterns have a correlation $\rho \in [-1, 1]$ and **different** spatial autocorrelation range
- joint models can be very useful for multi-type patterns... but also for other marked patterns...

CONSERVATION STUDY

Aim: prevention of decline in abundance of koalas in Australia

$\Rightarrow$ need to determine properties of an optimal (or suitable) habitat

- study conducted at the Koala Conservation Centre on Phillip Island, near Melbourne, Australia
- run from 1993 to 2004
- ≈ 20 koalas present in the reserve at all times throughout study
- reserve enclosed by a koala-proof fence
- koalas feed on eucalyptus leaves which are toxic to most animals; koalas have adapted to this
- main interest is to assess if koalas **feed selectively**, i.e. if they choose trees with the least toxic/ most nutritious leaves

FOLIAGE COLLECTION AND ANALYSIS

- all 915 trees in woodland individually numbered and mapped

spatial autocorrelation:

- trees may cluster locally due to seed dispersal mechanisms (small spatial scale)
- trees are likely to aggregate in areas where soil nutrient levels are good (large spatial scale)

FOLIAGE COLLECTION AND ANALYSIS

- leaf samples taken from each eucalyptus tree and analysed for palatability

palatability: combination of toxins and nutrients based on previous studies

spatial autocorrelation: palatability likely to not be independent of **spatial pattern**:

- in areas with high soil nutrient levels, nutrients in leaves high

KOALA TREE VISITATION

- tree use by individual koalas collected at monthly intervals between 1993 and March 2004
- entire reserve searched for koalas
- identities of all koalas found and of the trees occupied were recorded

spatial autocorrelation: koala visits likely to not be independent of **spatial pattern** and **palatability**:

- koalas move very little and are more likely to favour areas with higher tree density
- koalas are likely to favour trees with high palatability

in summary this suggest:

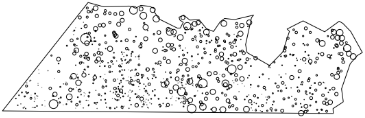
- tree locations depend on (unobserved) soil nutrients levels and local clustering
- palatability depends on spatial pattern (through soil nutrients levels)
- koala visitation depends on spatial pattern, palatability
- spatial point pattern data, to be modelled with a (marked) **spatial point process**

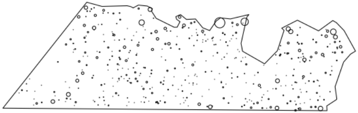
two types of marks:

1. palatability of leaves (“leaf marks")
2. koala use of trees (depends on palatability) (“frequency marks")

$\Rightarrow$ quantitative marks and another joint model

The spatial pattern with the leaf marks and the frequency marks





THE MODELLING APPROACH

- for the pattern of trees we have

$$\eta_{ij} = \alpha_1 + \beta_1 \cdot f_s(s_{ij}) + u_{ij},$$

- the marks $\mathbf{m}_1$ depend on the pattern through a joint spatially structured effect

$$\kappa_{ijk_{ij}} = \alpha_2 + \beta_2 \cdot f_s(s_{ij}) + v_{ijk_{ij}},$$

where $v_{ijk_{ij}}$ is another error term

- the marks $\mathbf{m}_2$ depend both on the spatial pattern through a joint spatial effect and on the marks $\mathbf{m}_1$

$$\nu_{ijk_{ij}} = \alpha_3 + \beta_3 \cdot f_s(s_{ij}) + \beta_4 \cdot m_1(\xi_{ijk_{ij}}) + w_{ijk_{ij}},$$

where $w_{ijk_{ij}}$ denotes another error term

DEPENDENT MARKS – SPECIFICATION IN `R-INLA`

- now we have **three response variables**
- `R-INLA` requires that these are stored as the column vectors of a $3n \times 3$ **matrix** `B`, in analogy to before
- three separate vectors $i.spat1$, $i.spat2$, $i.spat3$ for the spatial effects; similarly for error terms
- and a covariate `z` containing the values of the leaf marks only in the rows referring to the frequency marks

DEPENDENT MARKS – SPECIFICATION IN R−INLA

Specification in R−INLA is very similar to what we saw for the rainforest data - even though the models are very different

```
formula = B ~ mu - 1 + z
+ f(i.spat1, model = "rw2d", ...)
+ f(i.spat2, copy = "i.spat1", fixed=FALSE,...)
+ f(i.spat3, copy = "i.spat1", fixed=FALSE, ...)
+ f(i.error1, model = "iid", ...)
+ f(i.error2, copy = "iid", ...)
+ f(i.error3, copy = "iid", ...)
```

We specify the three different likelihoods when calling inla

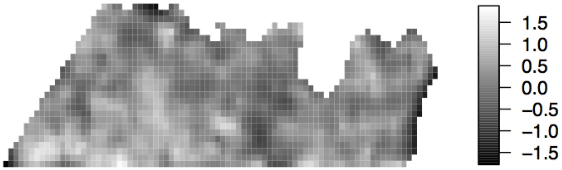
```
result =
inla(formula,...,family=c("poisson","normal","poisson"),
```

RESULTS – MODEL COMPARISON

Model	Terms	DIC	Time (s)
1.	Only error terms	11308	4
2.	Add intercepts	8362	4
3.	Add fixed covariate (β_4)	7640	5
4.	Add spatial effect		
	only for pattern	7511	25
	for pattern and leaf marks	7312	71
	for pattern and frequency marks	7193	61
	for pattern and both marks	6943	142

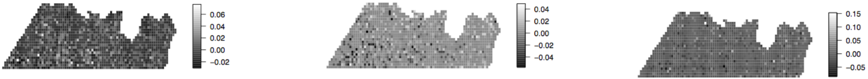
Table: DIC values and computation time for different fitted models for the koala data.

RESULTS – SPATIALLY STRUCTURED EFFECT



RESULTS – SPATIALLY UNSTRUCTURED EFFECTS

estimated error fields for the pattern, the leaf marks and the frequency marks



RESULTS

Parameter	Mean	95% credible interval
β_2	-1.18	[-1.39,-0.96]
β_3	1.72	[1.45, 1.98]
β_4	1.38	[1.24, 1.52]

Table: Posterior means and 95% credible intervals for parameters in the koala model.

- the full model shows the best fit
- negative β_2: palatability is low where the trees are aggregated (competition for soil nutrients?)
- positive β_3: koalas are more likely to be present in areas with higher intensity
- positive β_4: positive influence of palatability on the frequency of koala visits to the trees

CASE STUDY II: *Thymus carnosus*

- Evergreen coastal shrub, up to 0.5m high.
- Endemic to the southwestern of the Iberian Peninsula coastal dunes
- Protected species, in danger of extinction!

BACKGROUND

- High mortality of *Thymus carnosus* observed in 2008.
- Possibly due to severe drought in 2005, water table dropped sharply.
- Spatial pattern of mortality/decline in health not homogeneous, higher mortality in lower areas of the dunes.

BACKGROUND

- High mortality of *Thymus carnosus* observed in 2008.
- Possibly due to severe drought in 2005, water table dropped sharply.
- Spatial pattern of mortality/decline in health not homogeneous, higher mortality in lower areas of the dunes.

Aim:

Understand threats to survival of *Thymus Carnosus* to support conservation.

THREE DIFFERENT STUDY AREAS

1. High herbivory plots
   - Easily accessible for livestock.
   - Outside protected area.
2. Low herbivory plots
   - Limited livestock access
   - Outside protected area
3. Non herbivory plots
   - Not accessible to livestock.
   - Within protected area

COMPETING SPECIES: *Retama monosperma*

- Leafless leguminous shrub, up to 3.5m high.
- Planted in the El Rompido spit in the 1930s as a dune stabilizing species.
- Threat to dune landscapes, suppresses natural vegetation.

Descriptive summary

Plot	Number of *thymus* plants	Dead/unhealthy plants (%)	Retama cover (%)	Altitude (m)
High12	1309	65.62	23.84	[70, 246]
Low1	741	3.91	31.00	[49, 172]
Low2	526	11.03	26.00	[71, 246]
Nat1	206	13.11	54.88	[62, 101]
Nat2	79	30.38	62.24	[70, 107]

- High12 denotes two adjacent plots with high livestock pressure, analysed as one plot
- The other plots are in different areas and assumed independent.

Marked point patterns

Covariates:

- Different levels of livestock pressure
- Distance to water table (altitude)
- Coverage of competing species *Retama monosperma*.

Recorded marks:
All plants marked on a scale from 0 (dead) to 4 (very healthy).
Categorized as:

0 - 2: Poor health

3 - 4: Alive and healthy

Notation

- Point patterns for t different (rectangular/square) plots

$$\mathbf{x}_1, \ldots, \mathbf{x}_t$$

- Fixed covariates

$$z_1(s_{ti}), \ldots, z_{n_\beta}(s_{ti})$$

  where s_{ti} denotes cell i in plot t.

- Response variables

$$\begin{array}{lcl} y_{ti} & : & \text{Observed number of points in grid cell } s_{ti}. \\ m_{ti} & : & \text{Number of healthy plants in grid cell } s_{ti}. \end{array}$$

Conditional distributions for pattern and marks

- Number of points in each cell:

$$y_{ti}|\eta_{ti}^{(1)} \sim \text{Poisson}(|s_{ti}|\exp(\eta_{ti}^{(1)})),$$

- Number of healthy plants:

$$m_{ti}|\eta_{ti}^{(2)} \sim \text{Binomial}(y_{ti}, p_{ti}),$$

where

$$p_{ti} = \frac{\exp(\eta_{ti}^{(2)})}{1+\exp(\eta_{ti}^{(2)})}$$

is the probability of plants being healthy.

Joint model for pattern locations and marks

Need additional tools in `R-INLA`:

multiple likelihoods: Define response matrix

$$\begin{bmatrix} \text{counts} & \text{NA} \\ \text{NA} & \text{marks} \end{bmatrix}$$

and specify the families in the `inla`-call.

JOINT MODEL FOR PATTERN LOCATIONS AND MARKS

Need additional tools in `R-INLA`:

multiple likelihoods: Define response matrix

$$\begin{bmatrix} \text{counts} & \text{NA} \\ \text{NA} & \text{marks} \end{bmatrix}$$

and specify the families in the `inla`-call.

`replicate`: Define conditionally independent replicates of the same latent model, given hyperparameters.

JOINT MODEL FOR PATTERN LOCATIONS AND MARKS

Need additional tools in `R-INLA`:

multiple likelihoods: Define response matrix

$$\begin{bmatrix} \text{counts} & \text{NA} \\ \text{NA} & \text{marks} \end{bmatrix}$$

and specify the families in the `inla`-call.

`replicate`: Define conditionally independent replicates of the same latent model, given hyperparameters.

`copy`: Create a copy of a model component, when it is used more than once for each observation.

FORMULA SPECIFICATION IN R−INLA

- Assume a common spatial effect for counts and marks in each plot.
- Use all plots to estimate fixed effects for counts and marks.

Formula specification:

```
> formula = y.mat
            ~ -1 + beta.counts + beta.marks
            + cov.counts + cov.marks  + ...
            + f(I1, model="rw2d", nrow=nrow, ncol=ncol,
                    replicate=plot.location, hyper=...)
            + f(I2, copy="I1",
                    replicate=plot.marks, fixed=F)
```

RUNNING THE MODEL

Run `inla`:

```
> result = inla(formula, family =c("poisson","binomial"),
                data = data.frame, Ntrials = Ntrials,
                E = Area)
```

where `family =c("poisson","binomial")` specifies the two different distributions for the response variables and

`E`	:	The area of each of the cells
`Ntrials`	:	The number of plants in each cell.

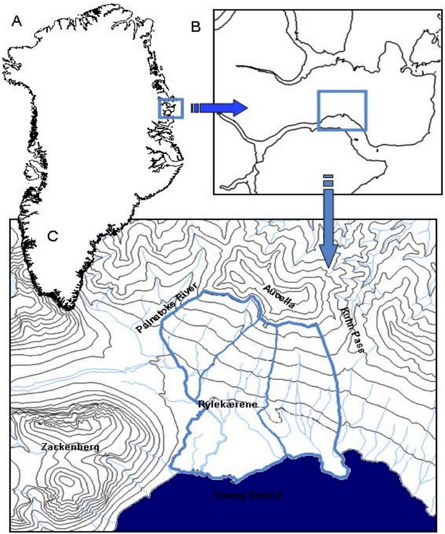
Results in:

Illian, J.B, Martino, S., Sørbye, S. H., Gallego-Fernandez, J. B., Zunzunegui, M., Paz Esquivias, M. and Travis, J. (2013). Fitting complex marked point patterns with integrated nested Laplace approximation (INLA). *Methods in Ecology and Evolution*, **4**, 305–315.

Case study I: Muskoxen in Zackenberg valley, Greenland

Dataset: Details the spatial location of muskoxen in a 45 km^2 census area, observed in summer months since 1996.

Zackenberg valley, Northeast Greenland



Environmental monitoring program

Zackenberg Basic:

- Large-scale monitoring programs, used to study the dynamics and potential climate effects on a high-arctic ecosystem:

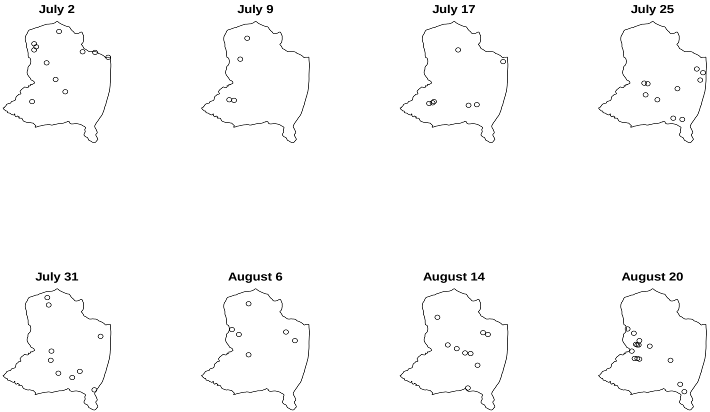
  http://www.zackenberg.dk/monitoring/

- Monitoring programs for a large variety of organisms, including the muskoxen.
- Provides long-term time series on vegetation and climate variables.

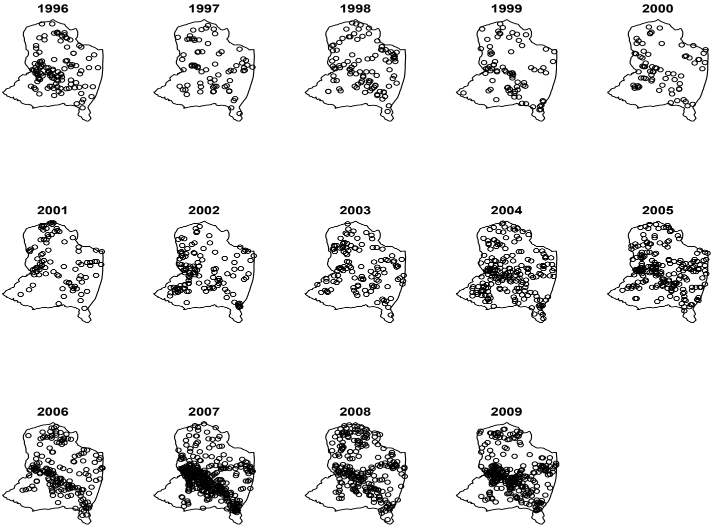
FACTS FOR THE MUSKOXEN DATA: 1996 - 2009

- Data collected on 144 different days.
- A total of 2249 observed muskoxen herds
- Herd size ranges from 1 to 59 individuals, giving a total number of 9975 individuals
- Each animal has been approached and categorized according to gender and age.

MUSKOXEN HERDS OBSERVED IN THE YEAR 2000



1996 - 2009: Superimposed patterns of muskoxen herds



Purpose of a statistical analysis

Long-term aim:

Investigate potential effects of climate changes.

Other relevant biological issues

Understand the dynamics of the spatial distribution of muskoxen:

- Influence of environmental covariates.
- Seasonal and annual variation.
- Interaction between different categories of the muskoxen.

INPUT

- Time variables: Days and years
- Fixed covariates:
  - Altitude: Can be regarded as a proxy for vegetation quality.
  - Ndvi: Normalized differential vegetation index, measure of vegetation greenness.
  - Snow: Binary variable for snow cover.
- Spatially structured effect:
  Account for spatial autocorrelation not explained by covariates.

A JOINT MODEL FOR ALL PATTERNS

For $t = 1, \ldots, T$:

$$\eta_{ti} \quad = \quad \beta_{year} + \sum_{j=1}^{n_\beta} \beta_j z_{tj}(s_i) + f_s(s_i) + v(t), \quad i = 1, \ldots, N$$

where

$$\begin{array}{rcl} \beta_{year} & : & \text{Factor for years} \\ \beta_j & : & \text{Fixed effects of covariates} \\ f_s(.) & : & \text{Common spatial effect} \\ v(t) & : & \text{Random iid variation in intensity on different days.} \end{array}$$

Dimension: $T \times N$ = Days $\times$ grid cells.

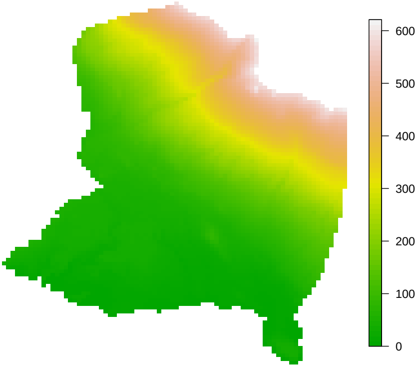
Subset: NDVI and snow cover only available for six years

Year	No. of days	No. of herds	Average: Herds pr. day
1998	13	99	7.62
1999	8	78	9.75
2000	8	70	8.75
2002	9	123	13.67
2004	14	202	14.43
2005	10	208	20.80
All years	62	780	12.58

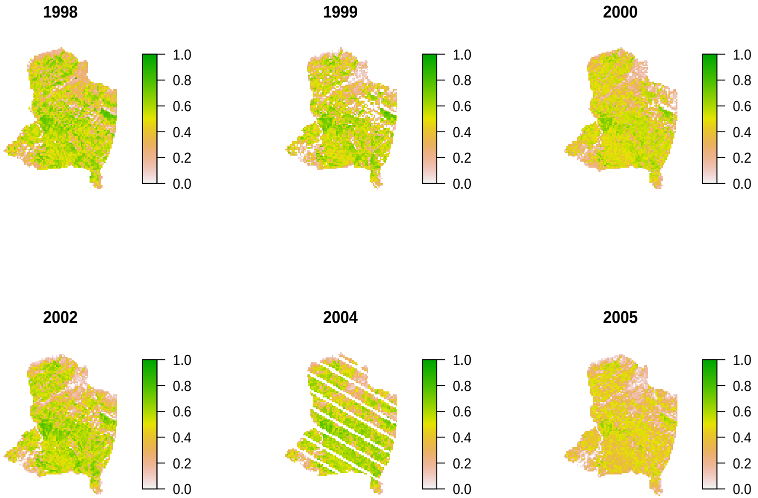
Dimension: $T \times N = 62 \times 4533 = 281046$

ALTITUDE

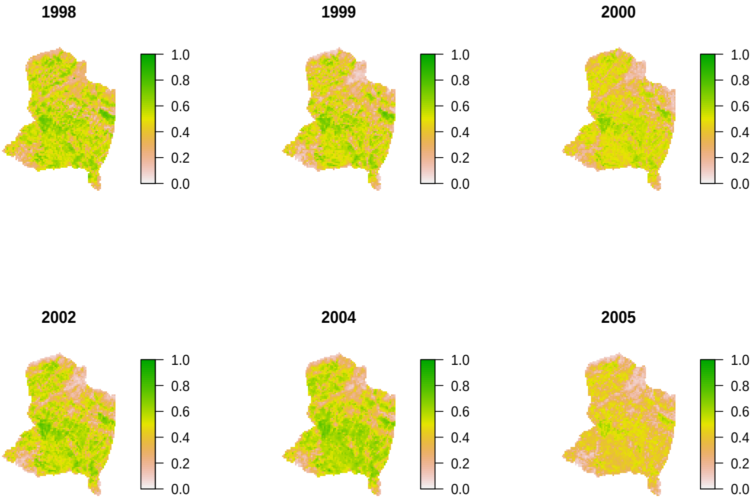
Ranges from sea level to 600 meters above sea level:



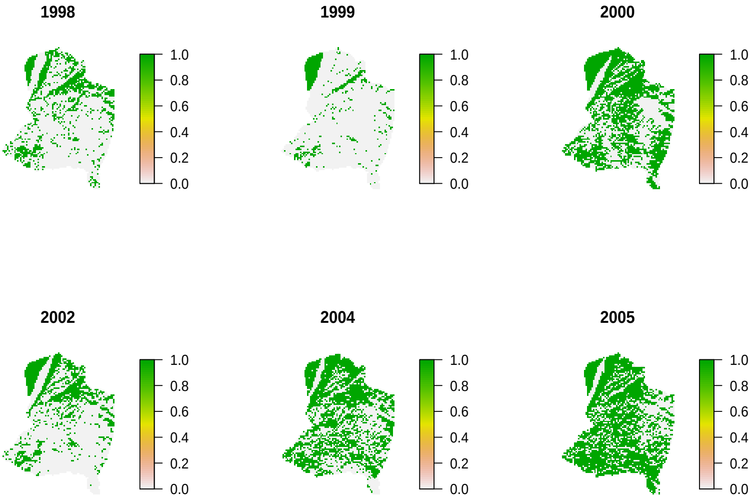
Normalized differential vegetation index



SMOOTHED!



Snow cover


1998
1.0
0.8
0.6
0.4
0.2
0.0
1999
1.0
0.8
0.6
0.4
0.2
0.0
2000
1.0
0.8
0.6
0.4
0.2
0.0
2002
1.0
0.8
0.6
0.4
0.2
0.0
2004
1.0
0.8
0.6
0.4
0.2
0.0
2005
1.0
0.8
0.6
0.4
0.2
0.0

COMMENTS: NDVI AND SNOW MEASUREMENTS

- Requires intensive pre-processing
  - Corrections, removal of artifacts etc.
- Other problems
  - Missing data, varying accuracy etc.
- Only measured once a year, around August 1.

More detailed covariate information?

- Ongoing work with the databases:
  - Based on satellite data, estimates of snow and vegetation will be available in every pixel for every day since 2003.
- Incorporating this in the model:
  - Estimate quantitative indices for vegetation quality, for example based on time since snow melt.

TWO MAIN CATEGORIES OF MUSKOXEN HERDS

- Each muskoxen herd have been categorized based on gender and age (a total of 17 categories)
- Current analysis:

  Males: Single old males and herds of young males.

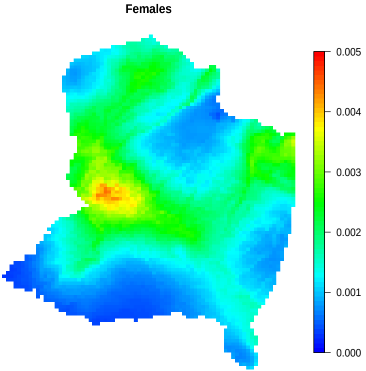
  Females: Herds with females, with or without calves.

Analysis of the two categories

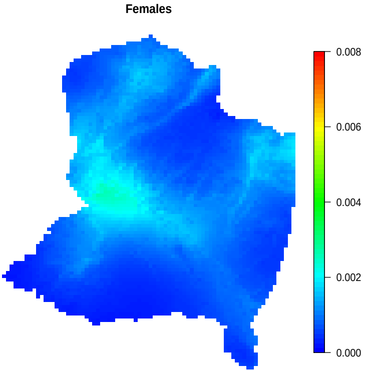
Hypotheses:

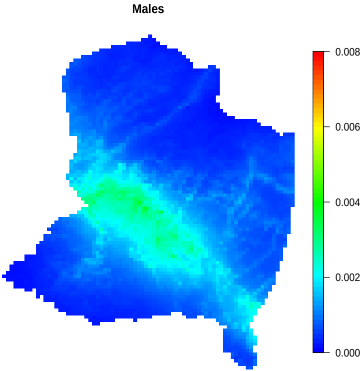
- Females and males have different spatial distributions.
  - Females need high-quality food, change area from year to year based on food availability.
  - Males keep to the same area, need enough food.
- Distribution between females and males changes during the summer.

Estimated intensity, all years

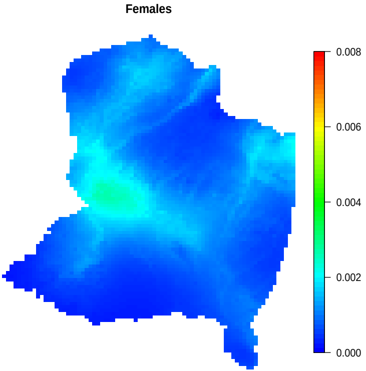


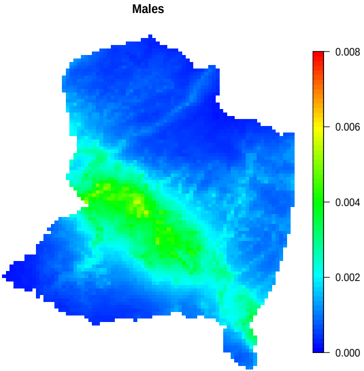
Estimated intensity 1998



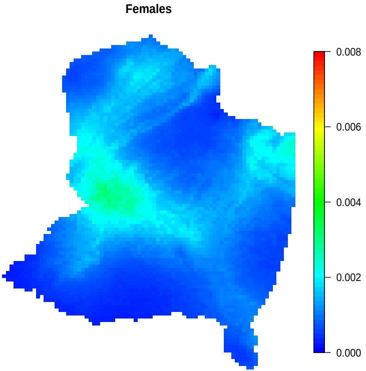


Estimated intensity 1999

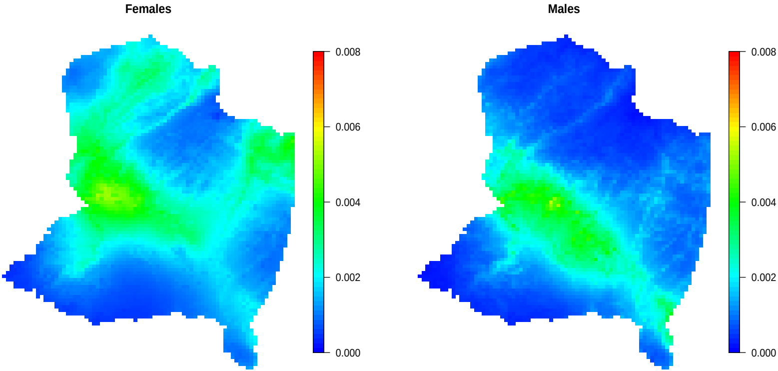




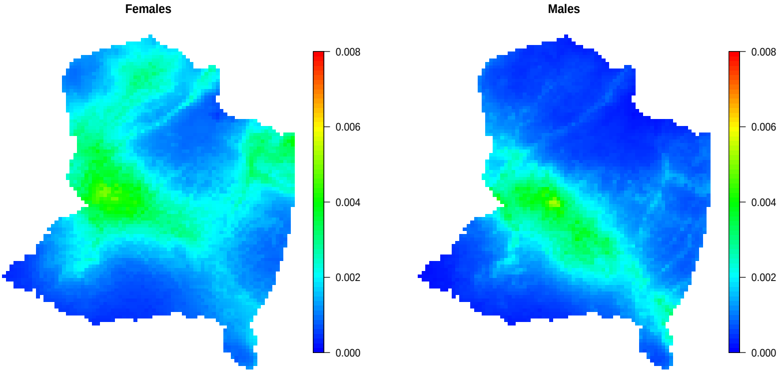
ESTIMATED INTENSITY 2000



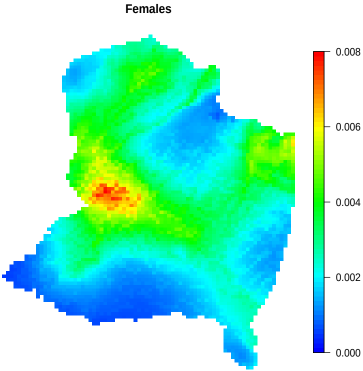
ESTIMATED INTENSITY 2002



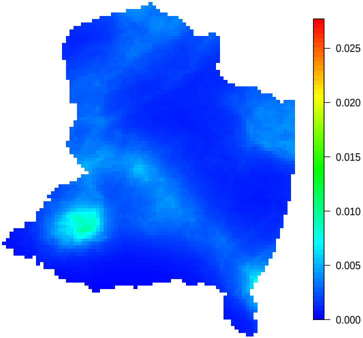
Estimated intensity 2004

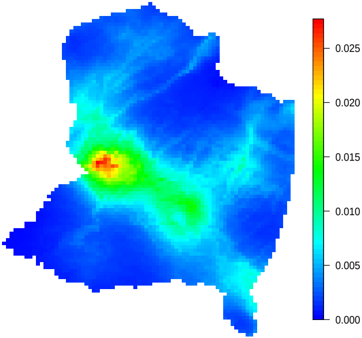


ESTIMATED INTENSITY 2005

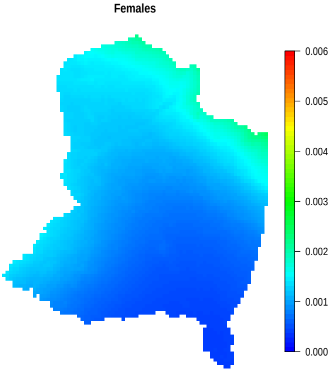


Estimated intensity pre- and post- season

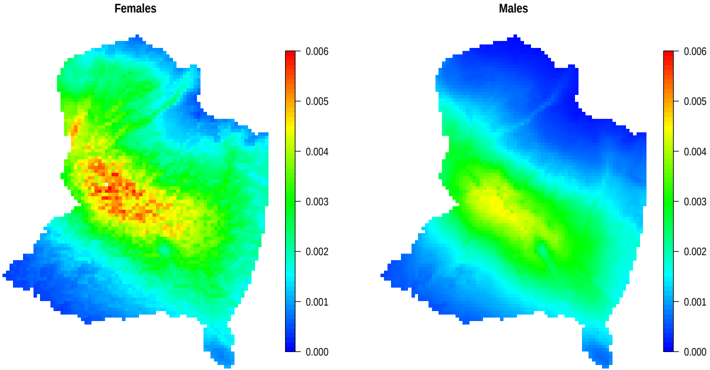




Seasonal variation? Estimated intensities pre-season



Seasonal variation? Estimated intensities post-season



SUMMARY MUSKOXEN

- Want to assess differences between different groups of muskoxen herds and study seasonal variation.
- Need joint model as many of the individual patterns are small.

SUMMARY MUSKOXEN

- Want to assess differences between different groups of muskoxen herds and study seasonal variation.
- Need joint model as many of the individual patterns are small.

In general:

Sequences of point patterns are easily analysed in `R-INLA`, just stacking all observations and covariate values in vectors.

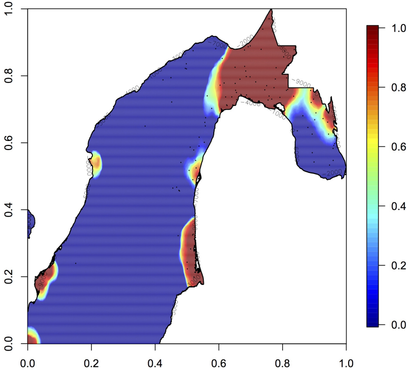
- Include some commonly estimated terms.
- Allow for variation between different patterns.

SUMMARY

- Joint models are easily adapted to analyse different sets of point patterns.
- Use information given by several point patterns to give more accurate estimates.
- Use commonly estimated terms to model dependency structures between point patterns, or between pattern locations and marks.

Applications

Modelling spatial phenomena has become increasingly important in recent years as scientific questions arise in fields such as Ecology,



Earth Sciences

Water resources planning

Figure 2

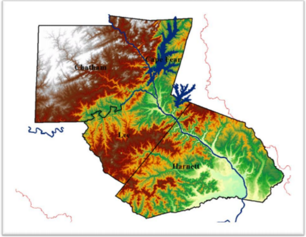
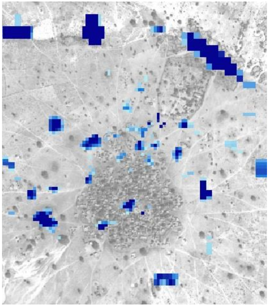
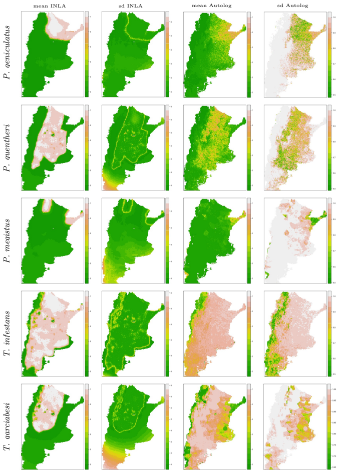


Figure 3



Health sciences



The main tasks is diverse, ranges from descriptive to predictive, mainly driven by industrial, health or conservationist concerns.

In spatial problems, a sensible approach is to assume that the observations come from a spatial process

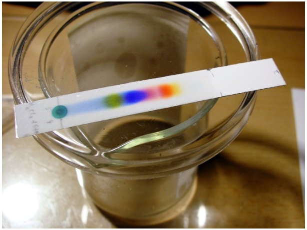
$$\mathcal{X} = \{X_s : s \in \mathcal{D}\}$$

indexed by a spatial set $\mathcal{D}$ with X_s taking values in a state space $\mathcal{E}$.

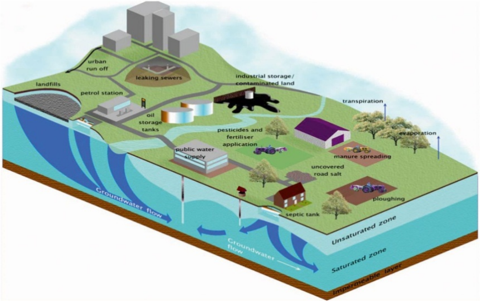
The positions of observation sites $s \in S$ may be fixed in advance or may be a random set.

Tipically, S is a 2-dimensional subset, $S \subset \mathbb{R}^2$.

S is 1-dimensional for applications such as chromatography, crop trials along rows, pollution along river beds



Or a subset of $\mathbb{R}^3$ in fields like mineral prospection, earth science, 3D imaging).



However, fields such as meteorology, Bayesian statistics and simulation may even require spaces $\mathcal{S}$ of dimension $d > 3$.

In some studies, data are collected both in space and in time, adding a temporal dimension to the statistical problem.

Data collected in space and time are becoming widely available as the result of new monitoring technologies (satellites, and continuous record monitoring stations), adding a new dimension to the statistical modelling problems

Space-time applications arise regardless of the support of the data (aggregated, point or continuos) and statistical methodologies have been developed for each problem type.

We now show a couple of applied examples.

Markov Random Fields
In some applications, the data represent aggregates over areas. The mesurement is attached to a point inside each area,

$$\mathcal{D} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$$

are the locations to which we associate a value of Z.

Define $Z(\mathbf{s}_i) = Z_i$ y $\mathbf{Z} = (Z_1, \ldots, Z_n)$

Markov Random Fields
In some applications, the data represent aggregates over areas. The mesurement is attached to a point inside each area,

$$\mathcal{D} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$$

are the locations to which we associate a value of Z.

Define $Z(\mathbf{s}_i) = Z_i$ y $\mathbf{Z} = (Z_1, \ldots, Z_n)$

Markov Random Fields
In some applications, the data represent aggregates over areas. The mesurement is attached to a point inside each area,

$$\mathcal{D} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$$

are the locations to which we associate a value of Z.

Define $Z(\mathbf{s}_i) = Z_i$ y $\mathbf{Z} = (Z_1, \ldots, Z_n)$

Definition

$\{Z_1, Z_2, \ldots, Z_n\}$ form a Markov Random Field $\iff Pr(Z_i|\mathbf{z}_{-i})$ depends only on the neighbouring sites to location i.

To make inferences we need to know $Pr(Z_1, Z_2, \ldots, Z_n)$. For this, we need the following assumption :

Positivity condition

$$Pr(Z_i = 0) > 0 \implies Pr(\mathbf{Z} = 0) > 0$$

The joint distribution of $\{Z_1, Z_2, \ldots, Z_n\}$ is found in the following steps:

1.- **Factorization Theorem** (Besag, 1974). Suppose that $\{Z_1, Z_2, \ldots, Z_n\}$ satisfy the positivity condition. Then

$$\frac{Pr(\mathbf{z})}{Pr(\mathbf{y})} = \prod_{i=1}^{n} \frac{Pr(z_i|z_1, , z_{i-1}, y_{i+1}, , y_n)}{Pr(y_i|z_1, , z_{i-1}, y_{i+1}, , y_n)}$$

where **z** y **y** are two valid realizations of the process.

2.-Define the potential energy

$$Q(\mathbf{z}) = log\left\{\frac{Pr(\mathbf{z})}{Pr(\mathbf{0})}\right\}$$

This implies

$$\frac{Pr(\mathbf{z})}{Pr(\mathbf{0})} = exp\{Q(\mathbf{z})\}$$

$\Longrightarrow$

$$Pr(\mathbf{z}) \propto exp\{Q(\mathbf{z})\}$$

and therefore

$$Pr(\mathbf{z}) = \frac{exp\{Q(\mathbf{z})\}}{\sum_{\mathbf{y}} exp\{Q(\mathbf{y})\}}$$

3.- $Q(\mathbf{z})$ satisfies: a)

$$Q(z) = \sum_{i=1}^{n} z_i G_i(z_i) + \sum_{i<j} z_i z_j G_{ij}(z_i, z_j) + \cdots$$
$$+ z_1 z_2 \cdots z_n G_{12\cdots n}(z_1, z_2, \cdots, z_n)$$

The G functions model the trend and the interactions between the sites.

b)

$$\frac{Pr(z_i|\{z_j : j \neq i\})}{Pr(0_i|\{z_j : j \neq i\})} = \frac{Pr(\mathbf{z})}{Pr(\mathbf{z}_i)} = exp\{Q(\mathbf{z}) - Q($$

$z_i)\}$

Note: The G satisfy

$$z_i G_i(z_i) = \frac{Pr(z_i|\{0_j : j \neq i\})}{Pr(0_j|\{0_j : j \neq i\})} \quad \text{y}$$

$$z_i z_j G_{ij}(z_i, z_j) = \frac{Pr(z_i|z_j, \{0_k : k \neq i, j\}) Pr(0_i|\{0_k : k \neq i\})}{Pr(0_i|z_j, \{0_k : k \neq i, j\}) Pr(z_i|\{0_k : k \neq i\})}$$

Note: The G satisfy

$$z_i G_i(z_i) = \frac{Pr(z_i|\{0_j : j \neq i\})}{Pr(0_j|\{0_j : j \neq i\})} \quad \text{y}$$

$$z_i z_j G_{ij}(z_i, z_j) = \frac{Pr(z_i|z_j, \{0_k : k \neq i, j\}) Pr(0_i|\{0_k : k \neq i\})}{Pr(0_i|z_j, \{0_k : k \neq i, j\}) Pr(z_i|\{0_k : k \neq i\})}$$

Definition

$\mathbf{s}_i$ y $\mathbf{s}_j$ son vecinos $\iff$ $Pr(z_i|\{z_j : j \neq i\})$ depende de z_j

Note: The G satisfy

$$z_i G_i(z_i) = \frac{Pr(z_i|\{0_j : j \neq i\})}{Pr(0_j|\{0_j : j \neq i\})} \quad \text{y}$$

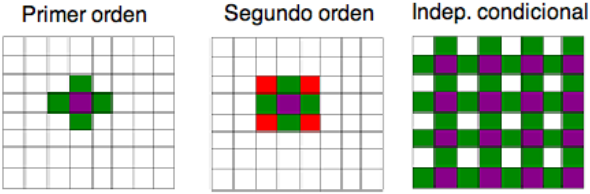
$$z_i z_j G_{ij}(z_i, z_j) = \frac{Pr(z_i|z_j, \{0_k : k \neq i, j\}) Pr(0_i|\{0_k : k \neq i\})}{Pr(0_i|z_j, \{0_k : k \neq i, j\}) Pr(z_i|\{0_k : k \neq i\})}$$

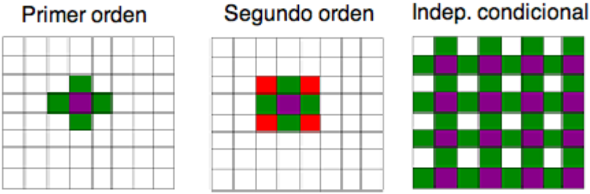
Definition

$\mathbf{s}_i$ y $\mathbf{s}_j$ son vecinos $\iff$ $Pr(z_i|\{z_j : j \neq i\})$ depende de z_j

Definition

A clique is composed either of a single site or by a set of mutual neighbouring sites.


Primer orden
Segundo orden
Indep. condicional



Hammersley-Clifford Theorem:

Suppose that the Markov Random field $\{z_1, \ldots, z_n\}$ satisfies the positivity condition. Then $G_{12\cdots p}(z_1, z_2, \cdots, z_p) = 0$ unless sites $\mathbf{s}_1, \ldots, \mathbf{s}_p$ form a clique

5.- **Automodels**. Suppose that

$$Pr(z_i|\{z_j : j \neq i\}) = exp\left\{A_i(\{z_j : j \neq i)B_i(z_i) + C_i(z_i) + D_i(\{z_j : j \neq i\})\right\}$$

and that there are only firs order interactions among sites. Then

$$A_i(\{z_j : j \neq i) = \alpha_i + \sum_{j=1}^{n} \theta_{ij} B_j(z_j); i = 1, \ldots, n$$

where $\theta_{ii} = 0; \theta?ij = \theta_{ji}$ y $\theta_{ij} = 0$ unless $\mathbf{s}_i$ y $\mathbf{s}_j$ are neighbours.

5.- **Example:** Autologistic model.- The binomial distribution can be expressed as

$$Pr(z_i) = exp\left\{z_i log\left(\alpha_i + \sum_{i=1}^{n}\theta_{ij}z_j\right) + log\left(\begin{array}{c} n_i \\ z_i \end{array}\right)\right.$$
$$+ n_i log(1 - exp\{\alpha_i + \sum_{i=1}^{n}\theta_{ij}z_j\})$$

(3)

Covariate information can be easily incorporated.

Space time analysis of forest fire ignitions in Oregon

Fire plays a key role in plant communities:

- Recycling of minerals.
- Trigers germination of dormant seeds.
- Removes plant material from surface $\rightarrow$ growth of new plants.
- Promotes habitat dynamics.

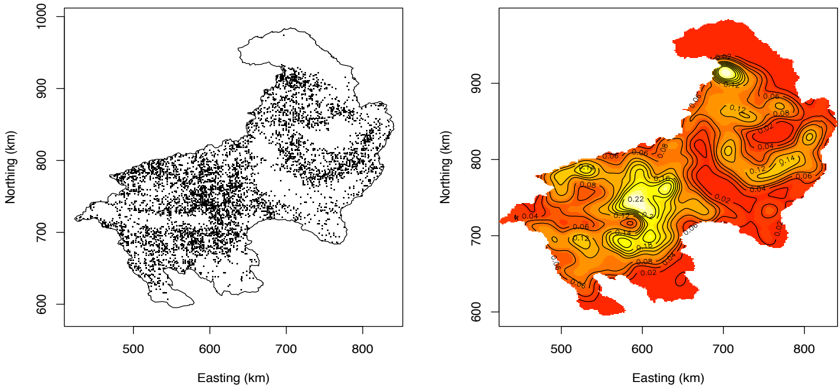

© EPA

Conditions for fire:

Fuel —— Ignition source

Oxigen

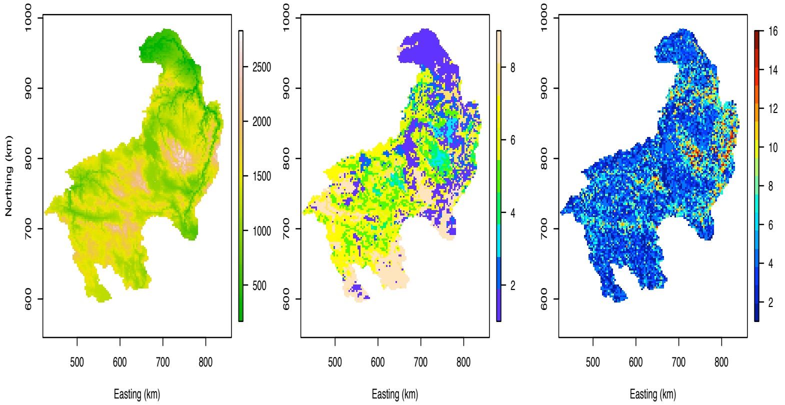
THE DATA SET

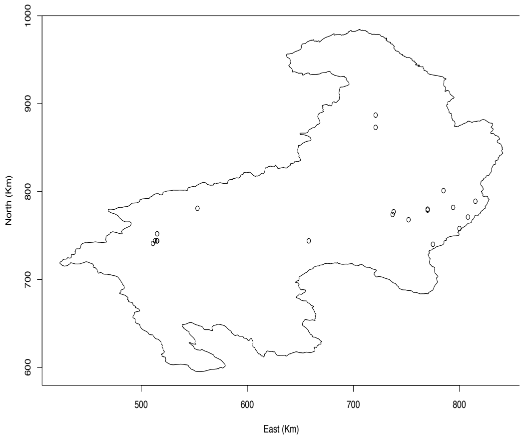


Northing (km)
Easting (km)
500
600
700
800
600
700
800
900
1000
0.22
0.14
0.12
0.08
0.06
0.04
0.02

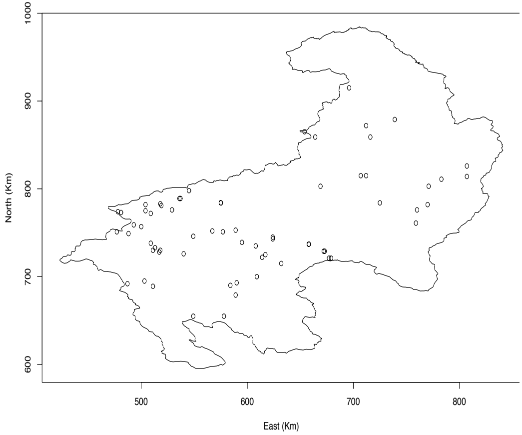
COVARIATES

For each site, we have the following covariate information

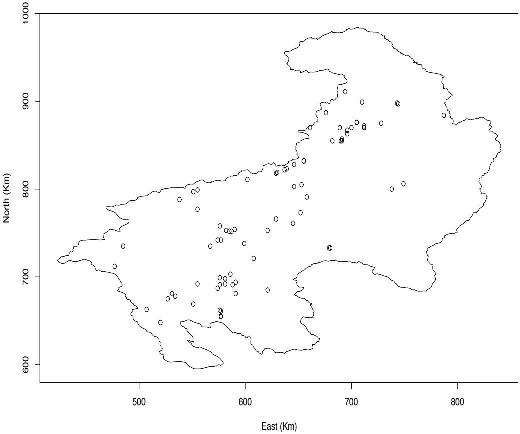



North (Km)
600
700
800
900
1000
500
600
700
800
East (Km)

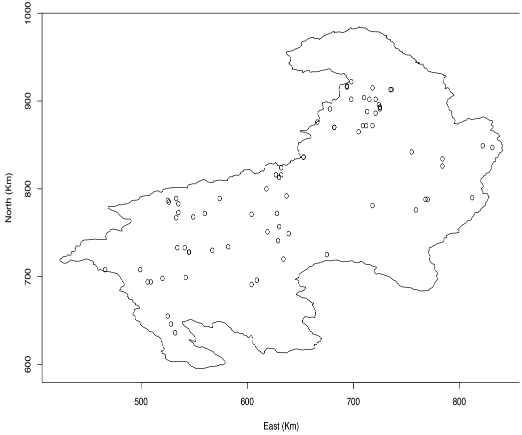
THE DATA SET



THE DATA SET



THE DATA SET



The time unit is a quarter, to avoid too much zeroes if a monthly basis is used

There goal is to assess the risk factors associated to fire ignitions and to get a predictive model.

The problem may be approached using space-time point process methods or by rasterizing the study area and use a MRF approach.

Consider an area $\mathcal{D}$ divided in n area units (pixels) and a set of observation times equally spaced (quarters). Let

$$Y_i = \begin{cases} 1 & \text{at least a fire ignition at site i and quarter t,} \\ 0 & \text{otherwise} \end{cases}$$

Given the binary nature of the data y_{it}, we fitted an autologistic regression model of the form

$$\xi_{it} = logit(p_{it}) = V_i + E_i + S_i + R_{it} + \psi_{it}$$

to the fire presence-absence data.

The spatial term ψ_{it} can be taken as a surrogate for unobserved variables that are correlated in space and time (Besag , 1995).

We assumed a flat, noninformative prior distribution (Box and Tiao, 1973) for the nonspatial parameters in our model

$$\pi(\psi) \propto \lambda^{0.5N} |W|^{0.5} \exp\{-0.5\lambda\psi' W\psi\}$$

$W_{ii} = \nu_i$, $W_{ij} = 1$ if pixels i and j are neighbors and $W_{ij} = 0$ otherwise.

For the precision λ we assumed a $G(1,1)$ prior density

The posterior density of the parameters is proportional to

$$\left(\prod_{i=1}^{N} \frac{\exp\left\{y_i x_i' \beta + \alpha R_t + \psi_{it}\right\}}{1 + \exp\left\{y_i x_i' \beta + \alpha R_t + \psi_{it}\right\}}\right) \times \lambda^{0.5N} |W|^{0.5} \exp\{-0.5 \lambda \psi' W \psi\}$$
$$\times \lambda^{a-1} \exp\{-b\lambda\}$$

Full conditionals

$$\pi(\beta, \alpha|\cdot) \propto \left(\prod_{i=1}^{N} \frac{\exp\left\{y_i x_i' \beta + \alpha R_t + \psi_{it}\right\}}{1 + \exp\left\{y_i x_i' \beta + \alpha R_t + \psi_{it}\right\}} \right)$$

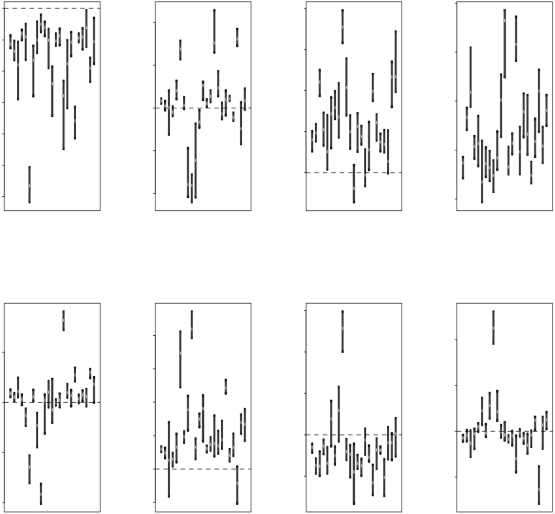
$$\pi(\psi|\cdot) \propto \left(\prod_{i=1}^{N} \frac{\exp\left\{y_i x_i' \beta + \alpha R_t + \psi_{it}\right\}}{1 + \exp\left\{y_i x_i' \beta + \alpha R_t + \psi_{it}\right\}} \right) \times \exp\{-0.5\lambda\psi' W \psi\}$$

$$\pi(\lambda|\cdot) \sim \Gamma(a + 0.5N, b + \sum_{i=1}^{N} \nu_i(\psi_{it} - \bar{\psi})^2)$$

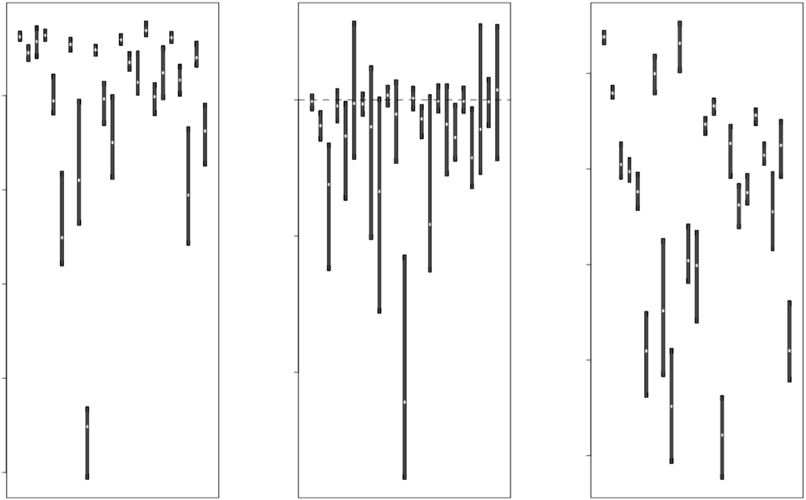
Implementation Vegetation was shrinked into 9 categories:

Level	Vegetation	Quarter
1	02,03	03/01-05/31
2	4,17	06/01-08/31
3	5,6	09/01-11/30
4	10,11,18	12/01-02/28
5	12	-
6	15	-
7	26	-
8	30,33,36,46,47	-
9	39,40,41,43	-

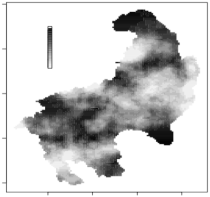
Results : Probability intervals for the covariate coefficients related to vegetation

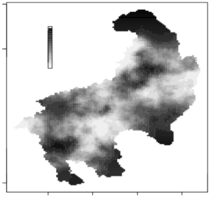


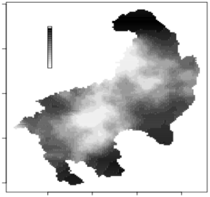
Results : Probability intervals for the covariate coefficients related to elevation, slope and rainfall by quarter



Spatial effect 86-87







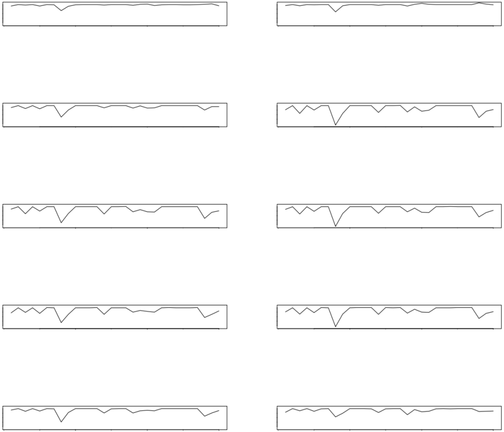
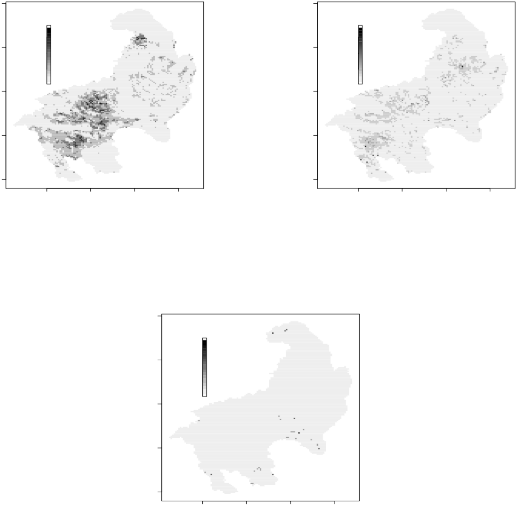
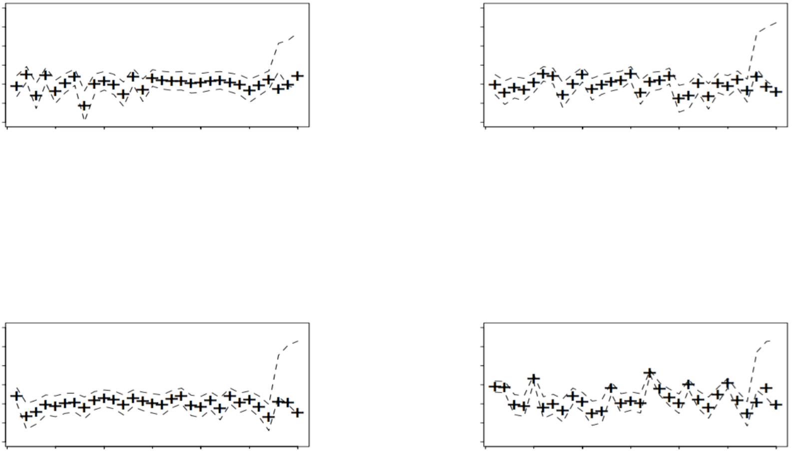


Figure 5.14: **Time series plot for the spatial effect in a sample of pixels.**





A DIFFERENT APPROACH FOR THE SAME PROBLEM:SPACE TIME POINT PROCESS.

Consider a spatio temporal point process $\mathbf{X} = \{\mathbf{X}_t : t \in \mathbb{Z}\}$, driven by an intensity functon $\Lambda = \{\Lambda_t : t \in \mathbb{Z}\}$

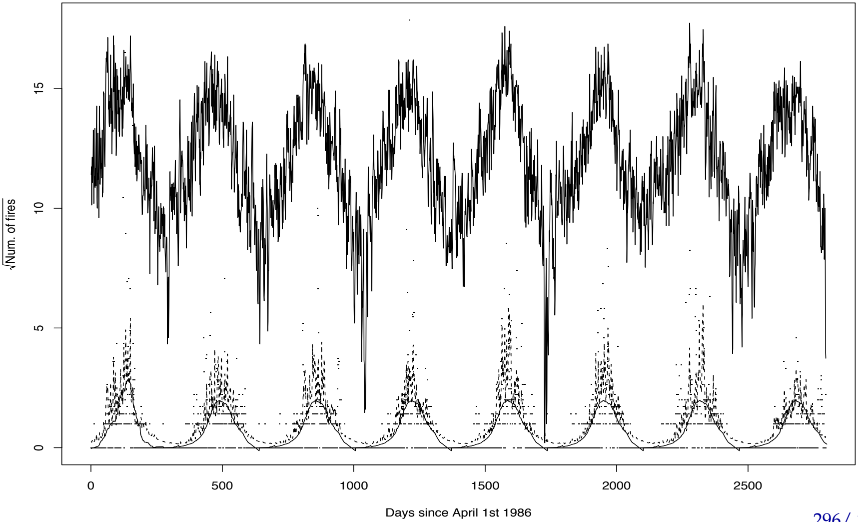
We further assume

$$\lambda(u,t) = \lambda_1(u)\lambda_2(t)S(u,t), \quad \mathrm{E}S(u,t) = 1, \quad (u,t) \in \mathbb{R}^2 \times \mathbb{Z}$$

where

$$log[\lambda_1(u)] = \mathbf{Z}_1(u) \cdot \theta_{1,1} \qquad log[\lambda_2(t)] = \mathbf{Z}_2 \cdot \theta_{1,2}$$

We include a time dependent covariate



MODEL

We let

$$\log \lambda_1(u) = c(\beta^{V,C,E}) + \sum_{i=1}^{9} \beta_i^V \mathbf{1}[V(u) = i] + \sum_{j=1}^{16} \beta_j^C \mathbf{1}[C(u) = j] + \beta_1^E E(u) + \beta_2^E E(u)^2 \quad (4)$$

and

$$\log \lambda_2(t) = \beta_0 + \beta_1^S \cos(\eta_t t) + \beta_2^S \sin(\eta_t t) + \beta_3^S \cos(2\eta_t t) + \beta_4^S \sin(2\eta_t t) + \beta_1^T T(t) + \beta_2^T T(t)^2 + \beta_3^T T(t)^3 + \beta_4^T T(t)^4 + \beta_5^T T(t)^5 \quad (5)$$

MODEL

For the shot-noise component we also assume a separable kernel

$$S(u,t) = \delta \sum_{s=-\infty}^{\infty} \sum_{y \in \breve{\ }_s} \phi(u-y, t-s)$$

$$\varphi(u,t) = \phi^{(2)}_{\sigma^2}(u)\chi_\zeta(t), \quad (u,t) \in \mathbb{R}^2 \times \mathbb{Z}. \tag{6}$$

$$\phi^{(2)}_{\sigma^2}(u) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|u\|^2}{2\sigma^2}\right)$$

MODEL

$\chi_\zeta(t)$ is concentrated on $0, \ldots, t^* - 1$ with $t^* = 20$

$$\chi_\zeta(t) = \zeta(t^* - t), \quad t = 1, \ldots, t^* - 1, \tag{7}$$

is a decreasing linear function so that χ_ζ becomes a probability density function.

Note that

$\chi_\zeta(0) = 1 - \zeta t^*(t^* - 1)/2$ and $0 \leq \zeta \leq 2/[t^*(t^* - 1)]$.

ESTIMATION

Regression parameters are obtained by maximum composite likelihood, ie,

$$max(\theta_{1,1}) \left\{ \sum_{u \in \mathbf{x}_W} \log \lambda_1(u; \theta_{1,1}) - \int_W \lambda_1(u; \theta_{1,1}) \, \mathrm{d}u \right\}$$

and

$$max(\theta_{2,2}) \left\{ \sum_{t \in T} n_t \log \lambda_2(t; \theta_{1,2}) - \sum_{t \in T} \lambda_2(t; \theta_{1,2}) \right\}$$

The rest of the parameters are estimated by minimum contrast., ie,

minimize

$$\sum_{t=1}^{t^*-1} \left[\mathrm{Corr}(t;\zeta) - \widehat{\mathrm{Corr}}(t)\right]^2$$

where

$$\mathrm{Corr}(t;\zeta) = \chi_\zeta * \tilde{\chi}_\zeta(t)/\chi_\zeta * \tilde{\chi}_\zeta(0); \quad \chi_\zeta(t) = \zeta(t^* - t), \quad t = 1, \ldots, t^* - 1$$

and

$$\int_0^a \left(K_{\cup T}(r;\sigma,\delta)^b - \hat{K}_{\cup T}(r)^b\right)^2 \, \mathrm{d}r$$

$$K_{\cup T}(r;\sigma,\delta) = \pi r^2 + \delta' \left[1 - \exp\left(-r^2/(4\sigma^2)\right)\right]$$

For the spatio-temporal margins we use

$$\hat{K}_{\cup T}(r) = \sum_{u_1 \in \mathbf{x},\, u_2 \in \mathbf{x}}^{\neq} \frac{\mathbf{1}[\|u_1 - u_2\| \leq r] w_{u_1,u_2}}{\rho_{\cup T}(u_1)\rho_{\cup T}(u_2)} \tag{8}$$

and compare

$$\mathrm{Corr}(t) = \frac{\chi * \tilde{\chi}(t)}{\chi * \tilde{\chi}(0)}, \quad t = 0, 1, \ldots . \tag{9}$$

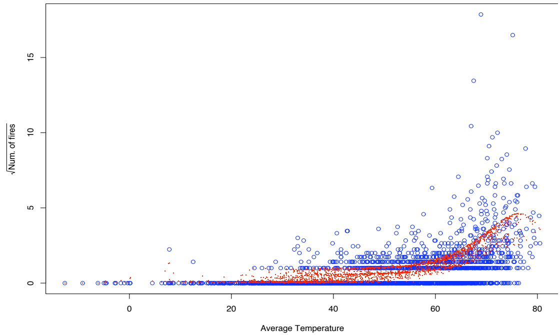
to

$$\widehat{\mathrm{Corr}}(t) = \frac{\sum_{s=1}^{m-t} \left[(n_s n_{s+t})/(\lambda_2(s)\lambda_2(s+t)) - 1\right]}{\sum_{s=1}^{m} \left[(n_s/\lambda_2(s))^2 - 1\right]}. \tag{10}$$

ESTIMATES: $\hat{\delta} = 2672.72$, $\hat{\sigma}^2 = 81.05$

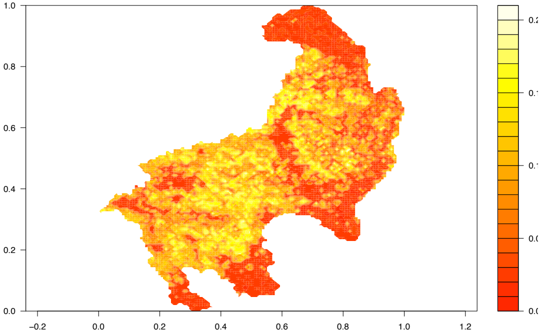
Table: Estimated regression parameters.

Index	$\hat{\beta}_i^S$	$\hat{\beta}_i^T$	$\hat{\beta}_i^V$	$\hat{\beta}_i^C$	$\hat{\beta}_i^E$
1	-1.6700	0.0527	-0.9769	0.0150	-0.0293
2	2.0360	-0.0059	0.1050	0.1479	-0.6040
3	0.1631	-9.04E-6	-0.1273	-0.0243	
4	0.1187	4.38E-6	0.5856	-0.0548	
5		-4.69E-8	0.6260	0.0836	
6			0.6810	0.2055	
7			0.2307	0.1208	
8			-0.0529	0.0971	
9			-1.0713	0.0370	
10				0.1892	
11				-0.1148	
12				-0.0569	
13				-0.2364	
14				0.0826	
15				-0.3104	
16				-0.1811	



Average temperature vs number of ignitions

ESTIMATES



Parametric estimate of $\hat{\lambda}_1(u)$

Model Control is done by simulation, and computing sample versions of the 2nd order properties:

$$\hat{K}(r,t) = \frac{1}{m-t}\sum_{t'=1}^{m-t}\hat{K}(r,t',t'+t), \tag{11}$$

where

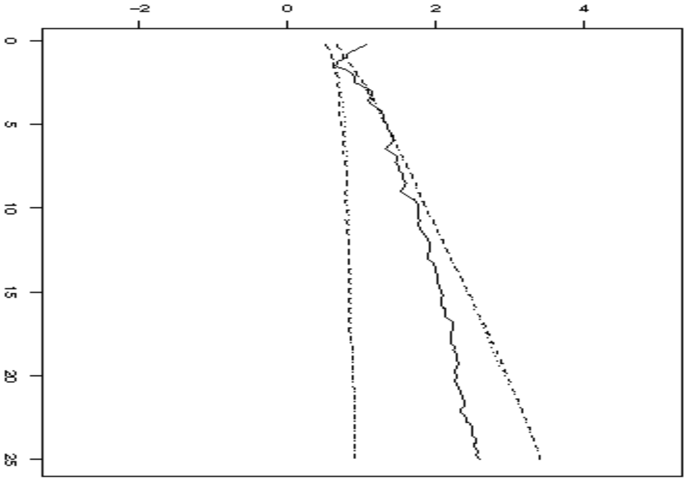
$$\hat{K}(r,t_1,t_2) = \sum_{u_1\in\mathbf{x}_{t_1},\,u_2\in\mathbf{x}_{t_2}}^{\neq} \frac{\mathbf{1}[\|u_1-u_2\|\le r]w_{u_1,u_2}}{\rho(u_1,t_1)\rho(u_2,t_2)} \tag{12}$$

Assuming $\lambda_1^{\max} = \sup_{u\in W}\lambda_1(u) < \infty$ and letting $\nu(s) = \sum_{t\in T}\lambda_2(t)\chi(t-s)$ for $s\in\tilde{T}$, the steps are as follows

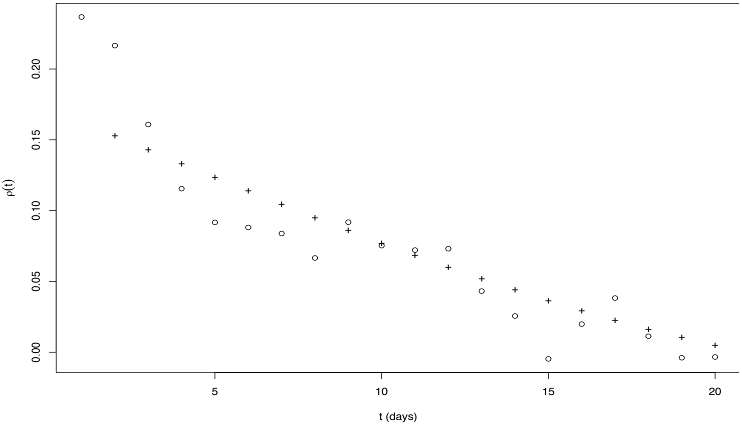
- Generate the mother processes $\tilde{\Phi}_s$, $s \in \tilde{T}$.
- For each $s \in \tilde{T}$ and $y \in \tilde{\breve{\ }}_s$,
  - (i) generate a realization $n(y, s)$ from a Poisson distribution with parameter $\lambda_1^{\max}\nu(s)\delta$;
  - (ii) generate $n(y, s)$ i.i.d. points with density $\phi(u - y)$, $u \in \mathbb{R}^2$;
  - (iii) make an independent thinning, where we retain each point u from (ii) with probability $(\lambda_1(u)/\lambda_1^{\max})\mathbf{1}[u \in W]$;
  - (iv) to each retained point u from (iii) associate a time t_u generated from the density $p_s(t) = \lambda_2(t)\chi(t - s)/\nu(s)$, $t \in T$.
- For each $t \in T$, return all retained points u with $t_u = t$ (no matter which $s \in \tilde{T}$ and $y \in \tilde{\breve{\ }}_s$ are associated to u).

These points constitute the approximate simulation of $\mathbf{X}_t \cap W$.

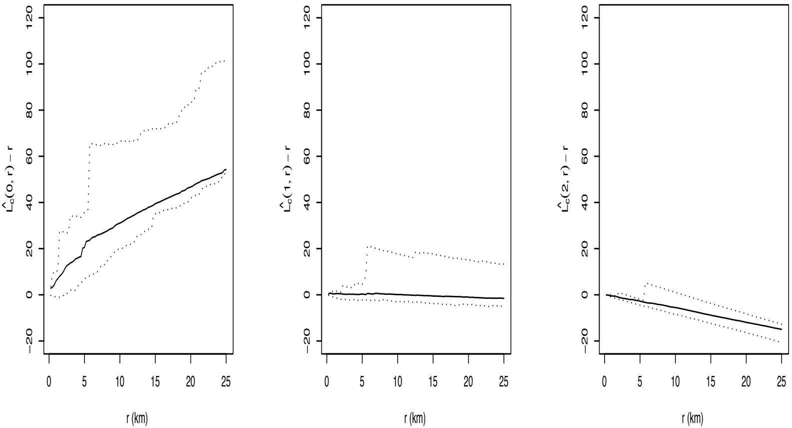
Results: Confidence bands for $\hat{K}_{\cup T}(r)$



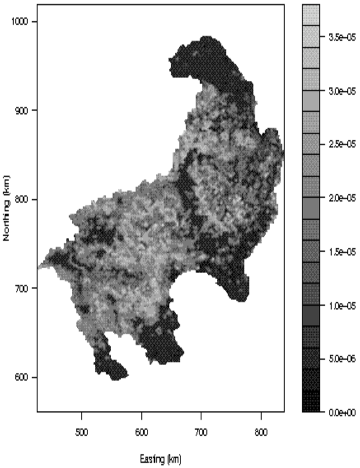
RESULTS: CORRELATION FUNCTION.

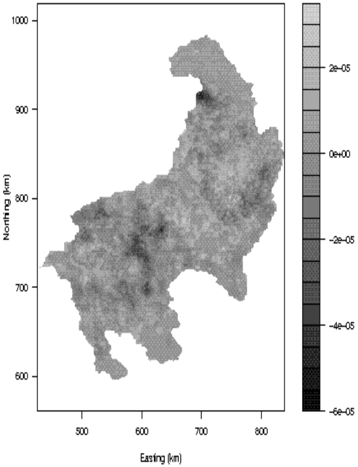


RESULTS: $\hat{K}(r, t, t+l), l = 0, 1, 2$



RESULTS:INTENSITY FUNCTION AND RESIDUALS.





CHEMICAL SPILL IN A RIVER

On 2014 there was an accidental spill from industry in a river in México

The government sent a team to make chemical analyses to monitor the extent of the damage

Samples were taken during 60 consecutive days at 32 locations along the river, covering 220 Km

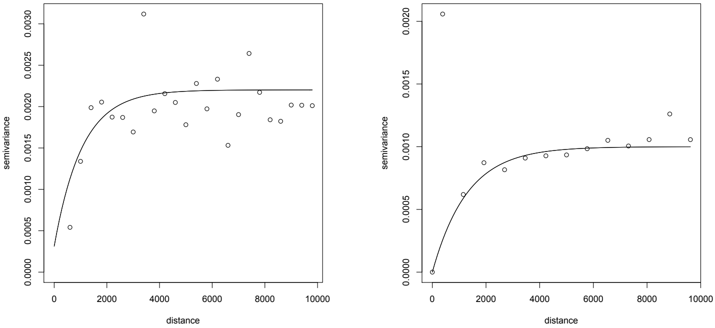
Data may be considered as a random field in $\mathbb{R} \times [0, T]$

A simple kriging analysis comes handy to give a quick answer to the question Is the pollution diluting along time and distance?

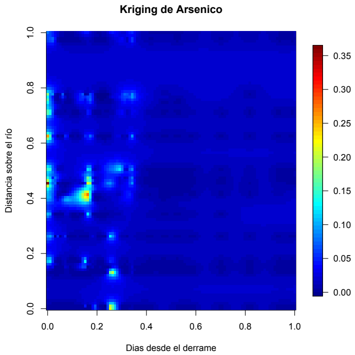
Metals under study were As, Al, Cd, Cu, Cr, Mn, Ni, Pb.

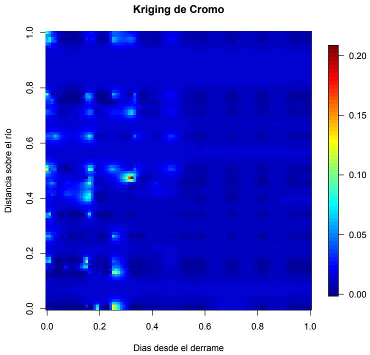
Variograms along time and space (distance) were computed and a prediction grid was used to obtain the kriging "map".

Selected variograms: As, Ni.

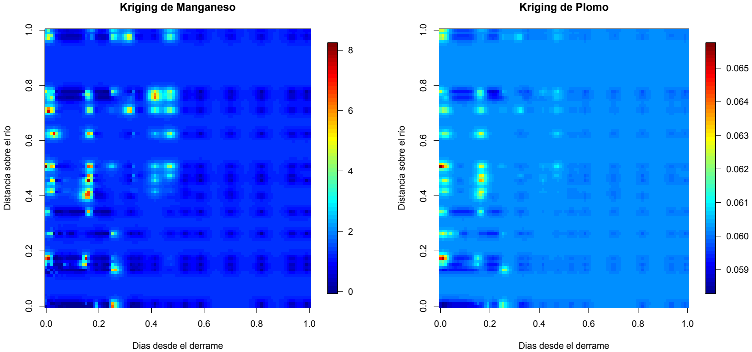


Results: As, Cr





Results:


Kriging de Manganeso
Distancia sobre el río
Dias desde el derrame
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
Kriging de Plomo
Distancia sobre el río
Dias desde el derrame
0.059
0.060
0.061
0.062
0.063
0.064
0.065

- Space time models provide a wide set of tools to analyze different kind of problems.
- Depending on the problem kind, fitting methods may be simple or computing intensive (Simple kriging or MCMC)
- Care must be taken to check the validity of the assumptions made (conditional independence)
- Separability or not is an importan issue in order to provide adequate answers to the scientific problem under study.
- Space-time models are still a rich subject of study, with many open problems waiting for a solution. Ongoing research is expected to provide better modeling approaches in the near future.

W.F. Caselton and J.V. Zidek.
Optimal monitoring network designs.
Statistics & Probability Letters, 2(4):223–227, 1984.

Howard Chang, Audrey Qiuyan Fu, Nhu D Le, and James V Zidek.
Designing environmental monitoring networks to measure extremes.
Environmental and Ecological Statistics, 14(3):301–321, 2007.

P.J. Diggle, R. Menezes, and T. Su.
Geostatistical inference under preferential sampling.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 59(2):191–232, 2010.

Irving John Good.
Rational decisions.
Journal of the Royal Statistical Society. Series B (Methodological), pages 107–114, 1952.

P Guttorp and P Sampson.

Discussion of geostatistical inference under preferential sampling by diggle, pj, menezes, r. and su, t.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 59(2):191–232, 2010.

N.D. Le and J.V. Zidek.
Statistical analysis of environmental space-time processes.
Springer Verlag, 2006.

Dennis V Lindley.
On a measure of the information provided by an experiment.
The Annals of Mathematical Statistics, pages 986–1005, 1956.

Mark E Mitch.
National stream survey database guide.
US Environmental Protection Agency, Office of Research and Development, 1990.

W.G. Müller.
Collecting spatial data: Optimum design of experiments for random fields, Third Edition.
Physica-Verlag, Heidelberg, 2007.

Douglas Nychka and Nancy Saltzman.
Design of air-quality monitoring networks.
In *Case studies in environmental statistics*, pages 51–76. Springer, 1998.

David Schmeltz, David C Evers, Charles T Driscoll, Richard Artz, Mark Cohen, David Gay, Richard Haeuber, David P Krabbenhoft, Robert Mason, Kristi Morris, et al.
Mercnet: a national monitoring network to assess responses to changing mercury emissions in the united states.
Ecotoxicology, 20(7):1713–1725, 2011.

Paola Sebastiani and Henry P Wynn.
Maximum entropy sampling and optimal bayesian experimental design.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62(1):145–157, 2002.

G. Shaddick and J.V. Zidek.
A case study in preferential sampling: Long term monitoring of air pollution in the uk.
Spatial statistics, page To appear, 2014.

Michael C Shewry and Henry P Wynn.
Maximum entropy sampling.
Journal of Applied Statistics, 14(2):165–170, 1987.

Zhengyuan Zhu and Michael L Stein.
Spatial sampling design for prediction with estimated parameters.
Journal of Agricultural, Biological, and Environmental Statistics, 11 (1):24–44, 2006.

James V Zidek, Weimin Sun, and Nhu D Le.
Designing and integrating composite networks for monitoring multivariate gaussian pollution fields.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 49(1):63–79, 2000.

J.V. Zidek, G. Shaddick, and C.G. Taylor.
Reducing estimation bias in adaptively changing monitoring networks with preferential site selection.
Ann Applied Statistics, page To appear, 2014.