Approaches to ST modeling

This module sketches some current theory and research directions in modelling space time fields. The amount of interest in this topic has escalated dramatically in the past few years. A lot of different approaches have been and are being proposed.

1 Introduction

London fog

1952: The most infamous environmental space-time field.

London fog

The most (in-) famous example

London fog

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

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

London fog

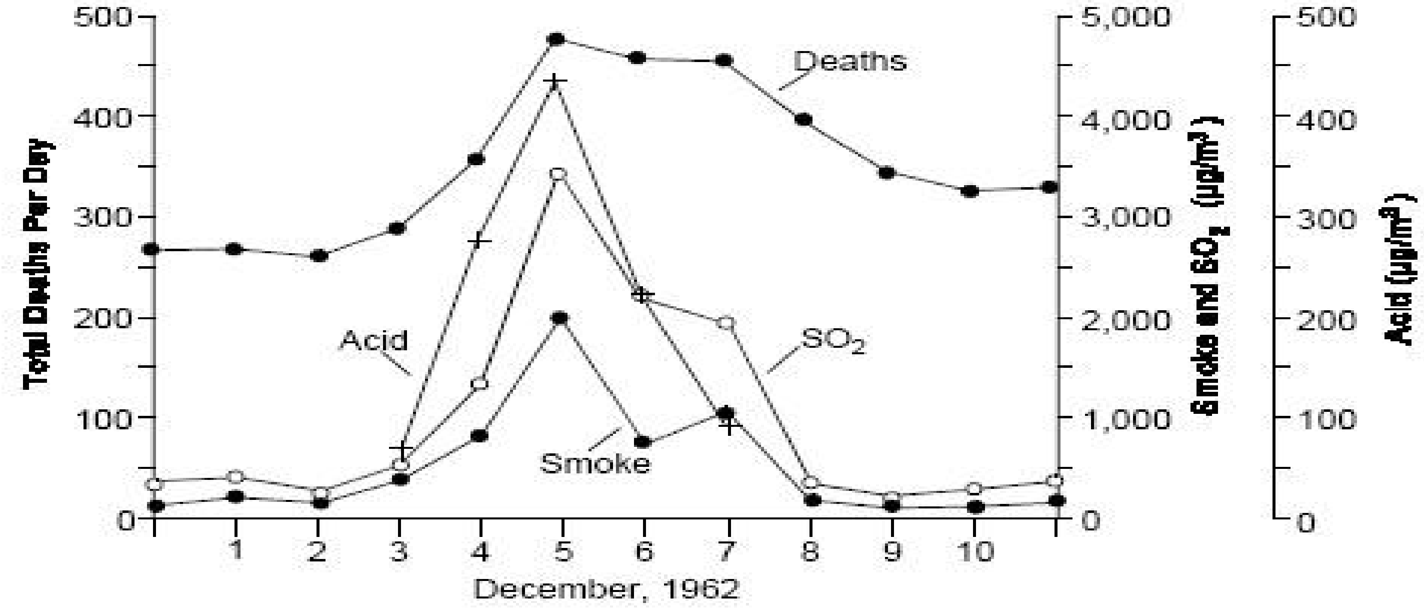


Figure 12-10. December 1962, London pollution episode.

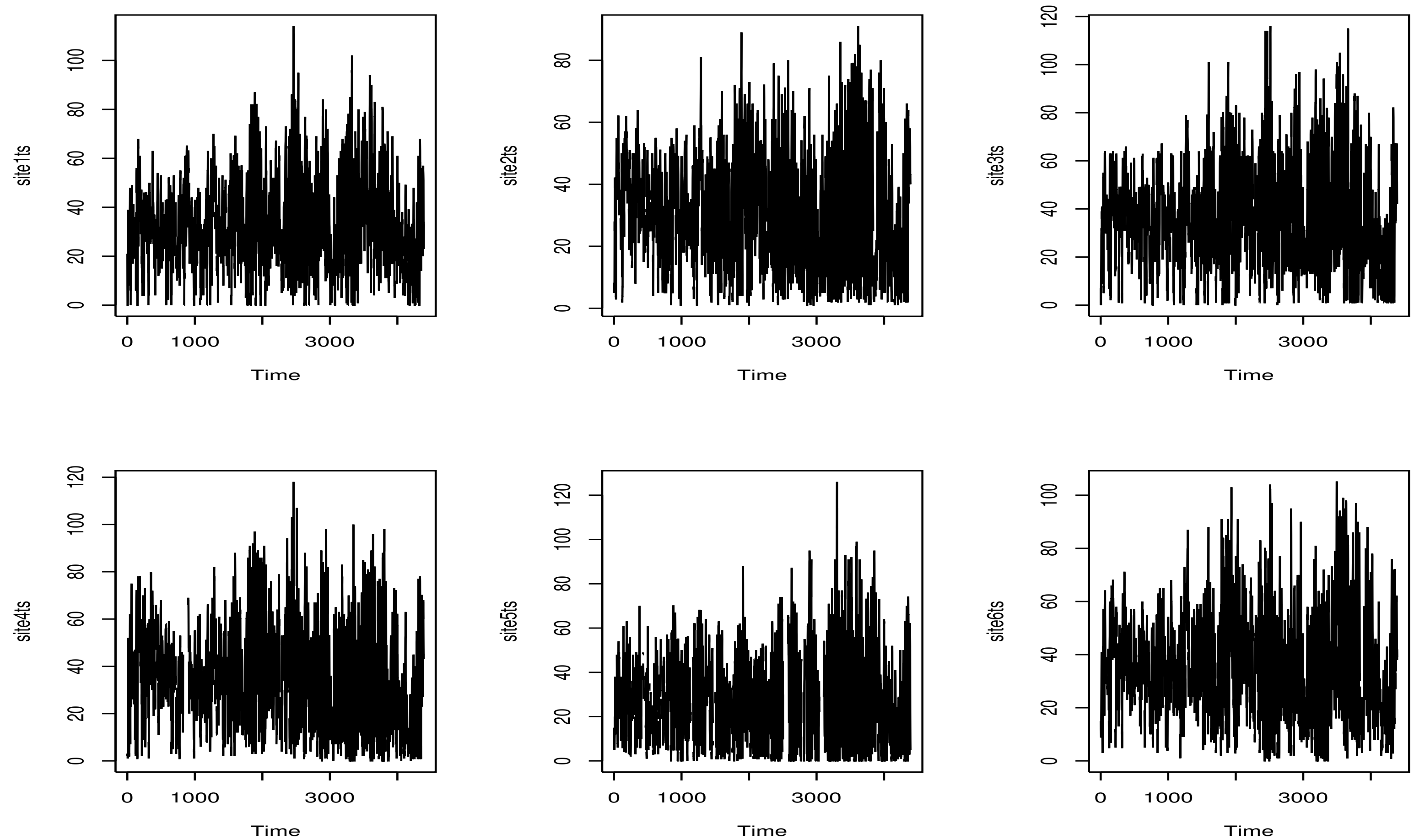
Space-time fields

Occur in a great variety of forms in the environment: E.g.

- **toxins** in fresh water bodies
- **fog**
- **air pollution**
- **acid precipitation** including pH a measure of acidity
- **meteorological variables** such as wind direction, temperature
- **death counts** over space and time

Example: ozone fields

Time series plots of 6 sampled O3 sites in the Eastern USA



Example (cont'd)

Footnotes:

- little randomly missing data
- marked daily cycles
- amplitude varies dynamically over season
- stations started up at varying times
- not all measure same species
- physical model data also available but at different resolution
- time variation similar from site-to-site

Modelling Space-Time fields, X

X massively **multivariate** being over $\text{time} \times \text{space} \times \text{species}$
Time & space usually discretized:

- **time** may mean, hour, day, month, etc
- **space** can refer to a point with a latitude and longitude or a region eg a county
- **species:**
  - may be a number of different chemical species, gases, aerosols, etc
  - generally, any vector of dependent variables like say 24 hourly values within day

Why model ST fields?

- to **impute** unmeasured responses
  - temporal forecasting
  - spatial prediction eg of systematically unmeasured responses eg species at certain sites
- to **smooth** noisy data as in disease mapping
- to detect spatial or temporal **gradients or trends**
- for u**nderstanding** of environmental processes
- to **optimize locations** of monitoring stations to be added or deleted
- to **integrate** physical and statistical models
- to integrate **misaligned** responses measurements
- for input in environmental **health risk** analysis model

Models: physical vs statistical

Large numerical deterministic **computer models** are bit to represent the physical world and used to predict random space-time fields, eg. ozone concentrations, on a coarse grid. Models are needed to combine these with statistical models for site specific responses.

Misaligned data

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

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

Health impacts model

Example

$$E[Y|\mathbf{X}] = \exp\left(\beta^T \mathbf{X}\right)$$

where

- $Y =$ daily number of hospital admissions for respiratory morbidity;
- $\mathbf{X}$ = vector of pollutants, co-pollutants and potential confounders. Need space time model for $\mathbf{X}$

2 Approaches to modelling ST fields

Modelling X's joint distribution

- **APPROACH 1.** Combine marginal distributions eg using **coppulas**:

$$F(x_1, x_2) = C[F_{X_1}(x_1), F_{X_2}(x_2)]$$

- Since $U_i \doteq F_{X_i}(X_i) \sim U[0,1]$, $C[U_1, U_2]$ on uniforms can be used to specify the dependence without regard to the marginals
  - nice separation of joint aspects from marginal aspects
  - but little used in ST modeling for some reason

Modelling X's joint distribution

- **APPROACH 2.** Products of conditional distributions:
  $f(x, y, z) = f(x|y, z)f(y|z)f(z)$
  - most common approach
  - conceptually simple
  - embraces Bayesian hierarchical models
  - computationally advantageous
    - each factor univariate density
    - fits well with Gibbs in MCMC

Modelling X's joint distribution

APPROACH 3. Treat timepoints separately - use geostatistics (kriging) **PROS**:

- mature: theory & applications
- simple: optimal linear predictors
- robust against misspecified temporal features
- flexible, interpretable
- allows prediction & design

CONS:

- needs lots of monitoring sites
- wastes past data
- oversimplifies spatial covariance
- ignores uncertainty in estimated covariance

Modelling X's joint distribution

APPROACH 4. Transform, prefilter: achieve near normality, remove systematic effects & autocorrelation (next lecture)
PROS:

- yields general Gaussian multivariate design and prediction theory
- fully admits parameter uncertainty
- yields predictive distribution for input into impact assessment and non-Gaussian kriging models
- good empirical performance
- admits systematically missing data
- R package available

Modelling X's joint distribution

CONS:

- complex
- challenged by non-separable space time series
- has not been fully Bayesianized - work underway

Modelling X's joint distribution

APPROACH 5. Dynamic state space modelling.

Example: Huerta, Sanso & Stroud (to appear) model the hourly $\sqrt{(O_3)}$ field over Mexico City - data from 19 monitors in Sep 1997. For time t and site i, they assume the "measurement model":

$$X_{it} = \beta_t^y + S_t' \alpha_{it} + Z_{it} \gamma_t + \epsilon_{it}^y$$

$S_t : 2 \times 1$ has sin's and cos's, α has their amplitudes, Z the temperature, ϵ_{it}^y the un-autocorrelated error that has an isotropic exponential spatial covariance.

Modelling X's joint distribution

Example (continued). The parameter/process model lets the parameters change dynamically:

$$
\begin{aligned}
\beta_t^y &= \beta_t^y + \omega_t^y \\
\alpha_{it} &= \alpha_{it-1} + \omega_t^{\alpha_i} \\
\gamma_t^y &= \gamma_t^y + \omega_t^\gamma
\end{aligned}
$$

They model Z (temp) in terms of height. The resulting model is hierarchical Bayes.

Modelling X's joint distribution

Approach 5 (continued)

PROS:

- intuitive, flexible and powerful
- allows for the incorporation of physical/prior knowledge
- leads to optimal designs that change from time to time - their value unclear

Modelling X's joint distribution

CONS:

- computationally intensive - may not yield practical design objective functions
- non - unique model specification - finding good one can be difficult
- unrealistic covariance in above example
- unclear if space - time non-separability is overcome with the approach above

Modelling X's joint distribution

APPROACH 6. Orthogonal series representations

$X_{it} = \sum_{k=1}^{K} a_k(t)\phi_k(i)$
are non-random orthonormal basis functions.

- the ϕs can be eigenvectors from decomposition of spatial covariance.
- incomplete - represents spatial dispersion not temporal
- long history, used in many ways (e.g. EOFs = empirical orthogonal functions.

Modelling X's joint distribution

EOFs: Suppose for n spatial sites $\mathbf{X}_t : n \times 1, \; t = 1, \ldots, T$ are independent copies of a r.vector $\mathbf{X}_0$. Let $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_T]$ and estimate the spatial covariance by

$$\hat{\Sigma}_{\mathbf{X}} = \frac{1}{T-n}\mathbf{X}\mathbf{X}' = \frac{1}{T-n}\sum \mathbf{X}_t\mathbf{X}_t'$$

Diagonalize it

$$\hat{\Sigma}_{\mathbf{X}} = \mathbf{Q}'\mathbf{D}^2\mathbf{Q}, \text{ with } D_1 > \cdots > D_p$$

Then $\mathbf{W}' \doteq \mathbf{D}^{-1}\mathbf{Q}$ is orthogonal ($\mathbf{W}'\mathbf{W} = I_n$),
$\mathbf{P} \doteq \mathbf{W}\mathbf{X}/(T-n) : n \times T$ is orthonormal ($\mathbf{P}\mathbf{P}' = I_n$).
Furthermore,

Modelling X's joint distribution

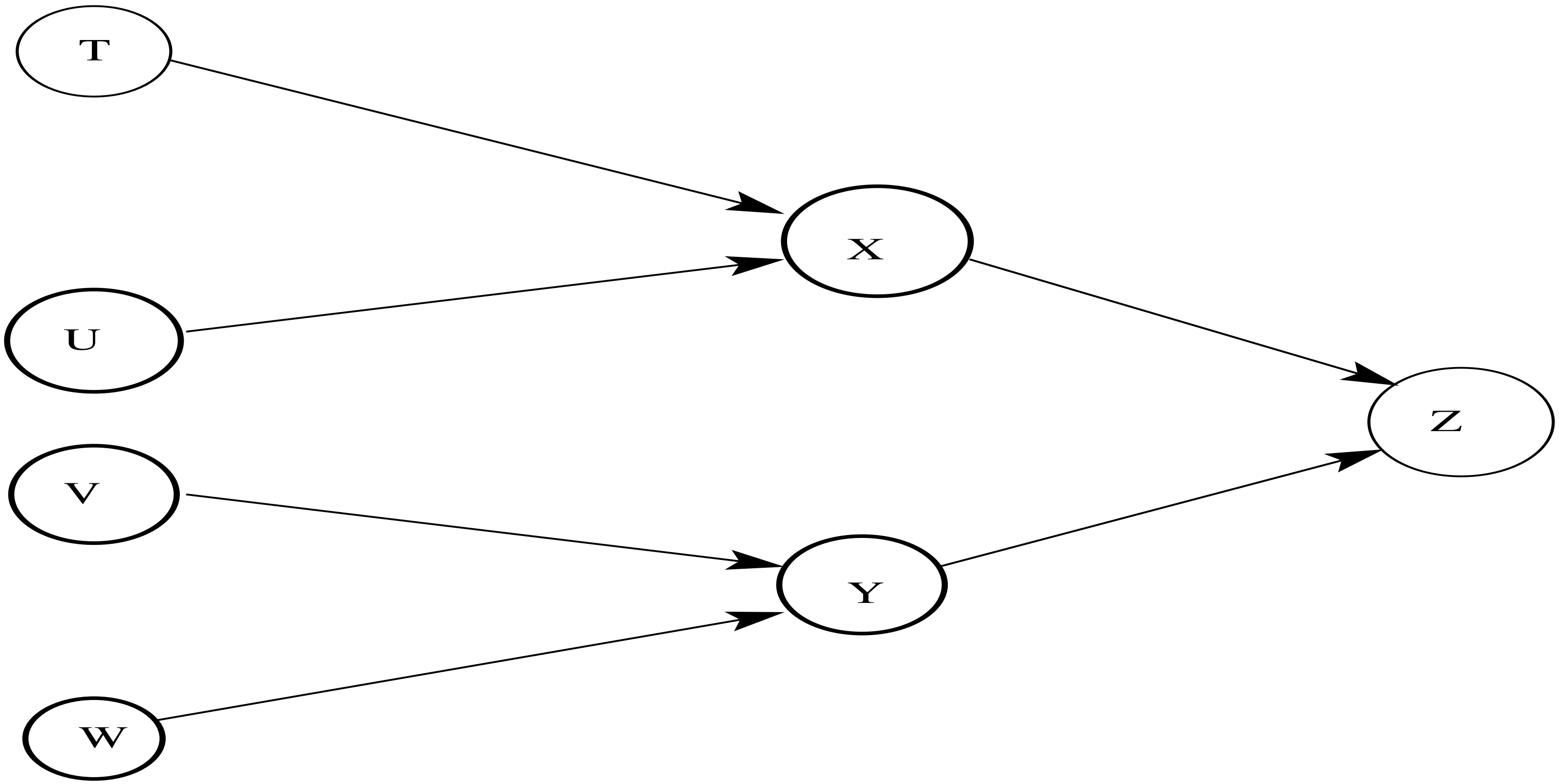
$$\begin{aligned} \mathbf{X} &\equiv \mathbf{WP} \\ &= [\mathbf{W}_1, \ldots, \mathbf{W}_n][\mathbf{P}'_1, \ldots, \mathbf{P}'_n]' \\ &= \sum_1^n \mathbf{W}_i \mathbf{P}_i \sim \mathbf{W}_1\mathbf{P}_1 + \mathbf{W}_2\mathbf{P}_2 \end{aligned}$$

The $\{\mathbf{W}_i\}$ are called the **EOFs** as they capture the spatial variation. For example $\mathbf{W}_1$, the most important component of spatial variation might represent the northern - southern hemisphere component of temperature variation across space. The $\{\mathbf{P}_i\}$ are called the **"principal components"** altho this means something different to statisticians. The approximation using just two EOFs would mean a big reduction in the data file.

Modelling X's joint distribution

APPROACH 7: Computer Graphical Modelling

Example:



Modelling X's joint distribution

The joint density can be expressed as

$$\begin{aligned} f(t,u,v,w,x,y,z) &= f(z|pa(z))f(x|pa(x)) \\ &\times f(y|pa(y))f(pa(x)f(pa(y)) \end{aligned}$$

with pa(z) = {x,y }, pa(x) = {t,u}, pa(y) = {v,w} as implied by the "causal diagram" or "directed acyclical graph". Powerful tool when causal relationships exist. But not used yet for ST processes.

Modelling X's joint distribution

APPROACH 8: Markov Random Fields (MRF's)
Like approach 1, but for site i, pa(i) = {neighbours of i }
PROS:

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

CONS:

- compatible joint distribution may not exist
- neighbours may be hard to specify
- a new site may not have neighbours for spatial prediction!
- conditionals may be hard to specify eg when "sites" are regions

Modelling X's joint distribution

APPROACH 9. Latent variables

$$X_{it} = \sum_j a_{ijt} Z_{jt}$$

where the Z_js are independent for any t. **EG:** co-regionalization in geostatistics.

PROS

- powerful representation yielding spatial covariance for example

CONS

- may be hard to implement in particular
- latent variables $\neq$ physical quantities at physical quantities

Modelling X's joint distribution

Example: MRF

Crown die back in birch trees. (Cressie and Kaiser 2002). Features:

- Single timepoint, t.
- X_{it} = prob a tree's crown dies back in region i with m_{it} trees.
- Y_{it} = number of trees with die back $\sim Bin(m_{it}, X_{it})$.
- pa(i) = all regions within 48 km of i. Conditional on pa(i), X_{it} has beta with parameters depending on responses in neighbours thru 3 parameters.
- very parsimonius model but unclear how time would be handled

3 Current directions

Current directions

DIRECTION 1. Mapping Fields of **Extreme Values**

Of fundamental importance in environmental risk assessment. 100-yr floods can cause property damage or human mortality. Extreme ozone levels associated with acute respiratory morbidity or even mortality. In fact, risk is generally associated with extremes.

Current directions

Example: EPA'S PM_{10} criterion

- Particle diameters $\leq$ 10 micrometers:
  - 24 - hour Average: 150$^{\mathrm{FN}\mu g\ m^{-3}}$

  actually based on extremes (annual 98th % iles of daily averages).

Quite an extreme value!

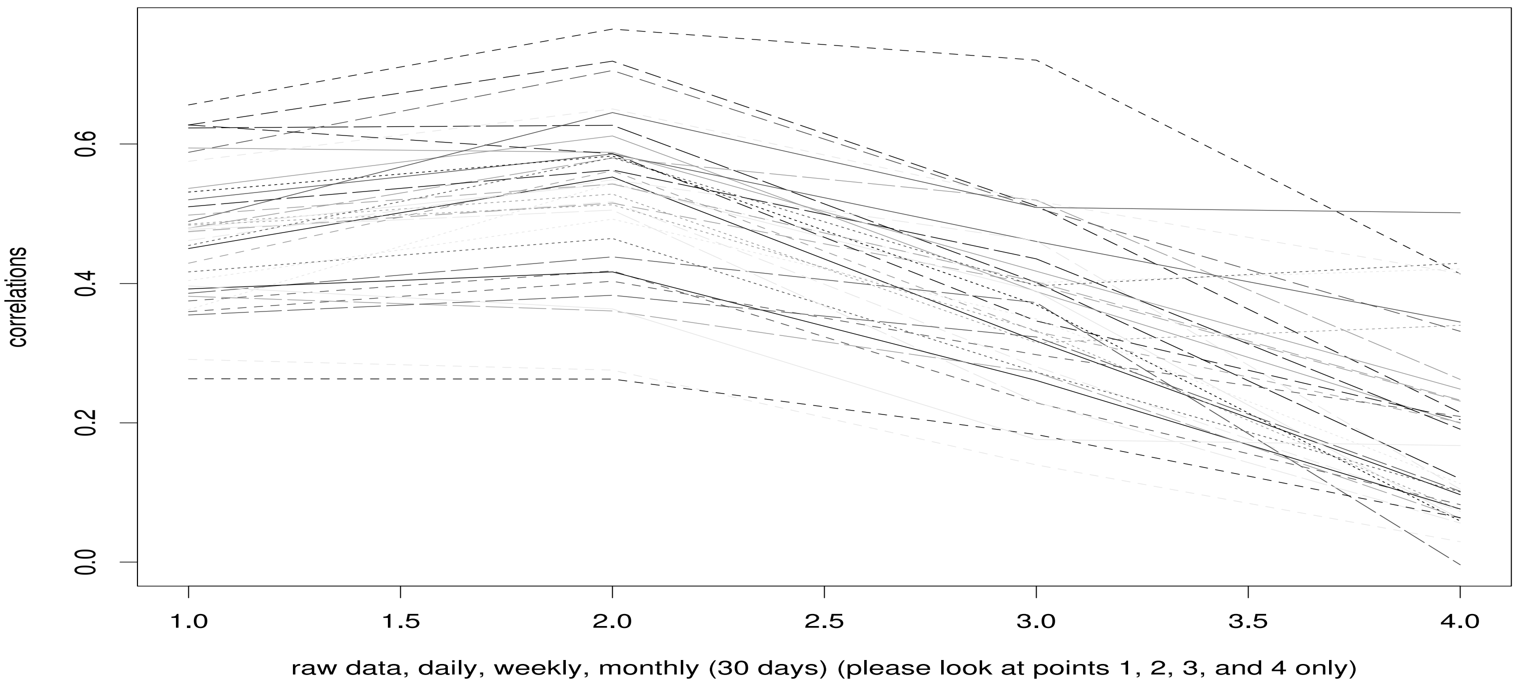
Current directions

Extremes: $\mathcal{BADNEWS}$ SUMMARY

1. **Insufficient data, spatial and temporal.**
2. **Extremes have small inter - site dependence**
   - **between some site pairs, not others**
3. **Conventional approaches fail**
4. **Multivariate extreme value distributions - not tractable**
   - **conditional computation (e.g. entropy) difficult**
   - **simulating extreme fields hard**
5. **Elusive design objective**

Current directions

Small inter - site correlations *Inter- site dependence declines with increases in extreme's"range" for many, not all site pairs [London and Vancouver analyses]*



Current directions

Extremes: $\mathcal{GOODNEWS}$ **SUMMARY**

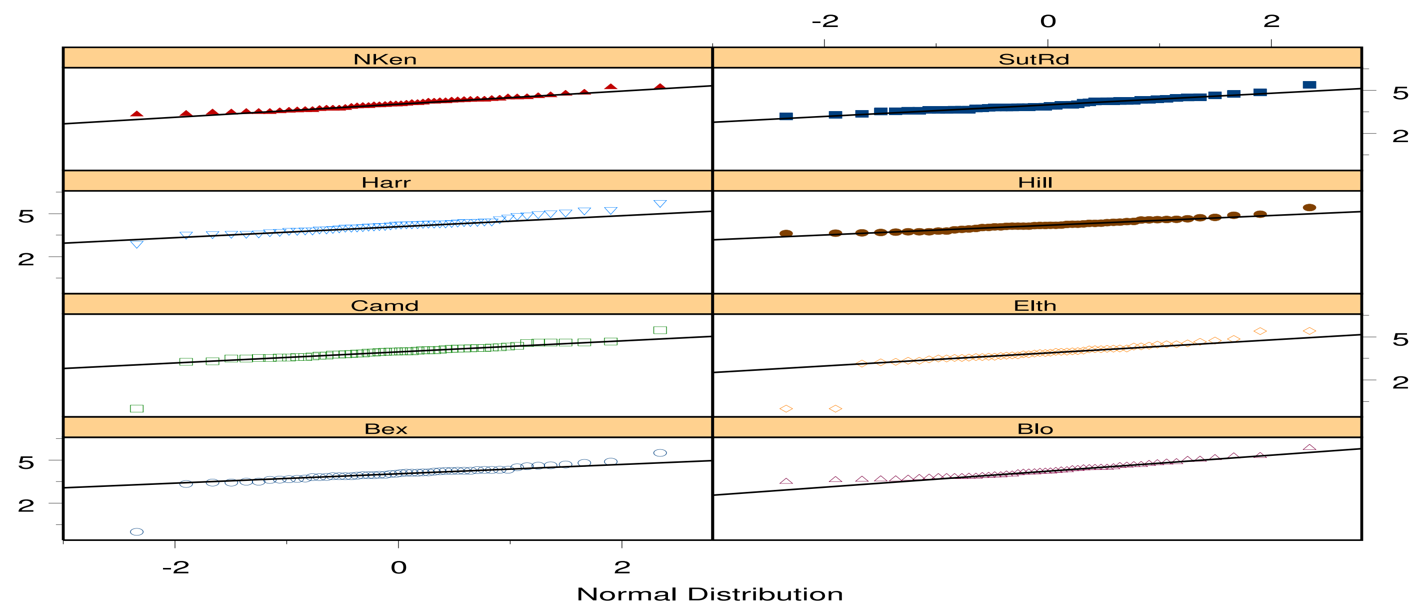
Joint distribution of extremes approximately a log multivariate t distribution. Hence can:

1. **have convenient conditional, marginal distributions**
2. **accommodate existing sites and historical data**
3. **permit simulation of complex metric distributions**
4. **have explicitly computable entropy's, regression models, etc**
5. **can enable "elusive objectives issue" to be bypassed**

Current directions

A practical approach to modelling extreme fields

Empirical results $\Rightarrow$ log multivariate - t distribution as approximation to joint distribution of extremes field.



QQplots for weekly maxima of hourly $\log PM_{10}$ London 1997 data $\Rightarrow$ **marginal** normality of extremes

Current directions

DIRECTION 2. Addressing space-time interaction.

Implication: filtering out temporal correlation will remove spatial correlation as well. Related concept: non-separability of time and space

Defn: Space-time separability. With **X** $: n \times T$ defined as above $Cov(\mathbf{X}_{rowi}, \mathbf{X}_{rowj}) = c_{ij}\Sigma_T$ while $Cov(\mathbf{X}_{coli}, \mathbf{X}_{colj}) = d_{ij}\Sigma_S$. In other words each row of **X** has same temporal structure, each column the same spatial. More simply

$$Cov(\mathbf{X}) = \Sigma_S \otimes \Sigma_T$$

Current directions

- space time interaction arises with short time aggregates like hourly data, not long like, weekly
- hourly measurements hold a lot more information about their neighbours than do daily
- filtering out autocorrelation reduces spatial correlation, accuracy of spatial predictors

Current directions

Direction 2 (continued). **Illustration:**
Vancouver's log transformed hourly PM_{10} field.

- Step 1.
  $T(t) = \mu + Hr(t) + Day(t) + Lin(t) + Seas(t) + Meteor(t)$
  fitted across sites and removed. Result: $E_{it} = X_{it} - T(t)$
- Next AR(3) model fitted across all sites and whitened residuals obtained: E'_{it}.

Current directions

Direction 2: Example (continued). A problem!!!!
Intersite correlation (*100) dies!:

BEFORE AND AFTER CORRELATION MATRIX $\times 100$

for E_{it} (**BEFORE**):

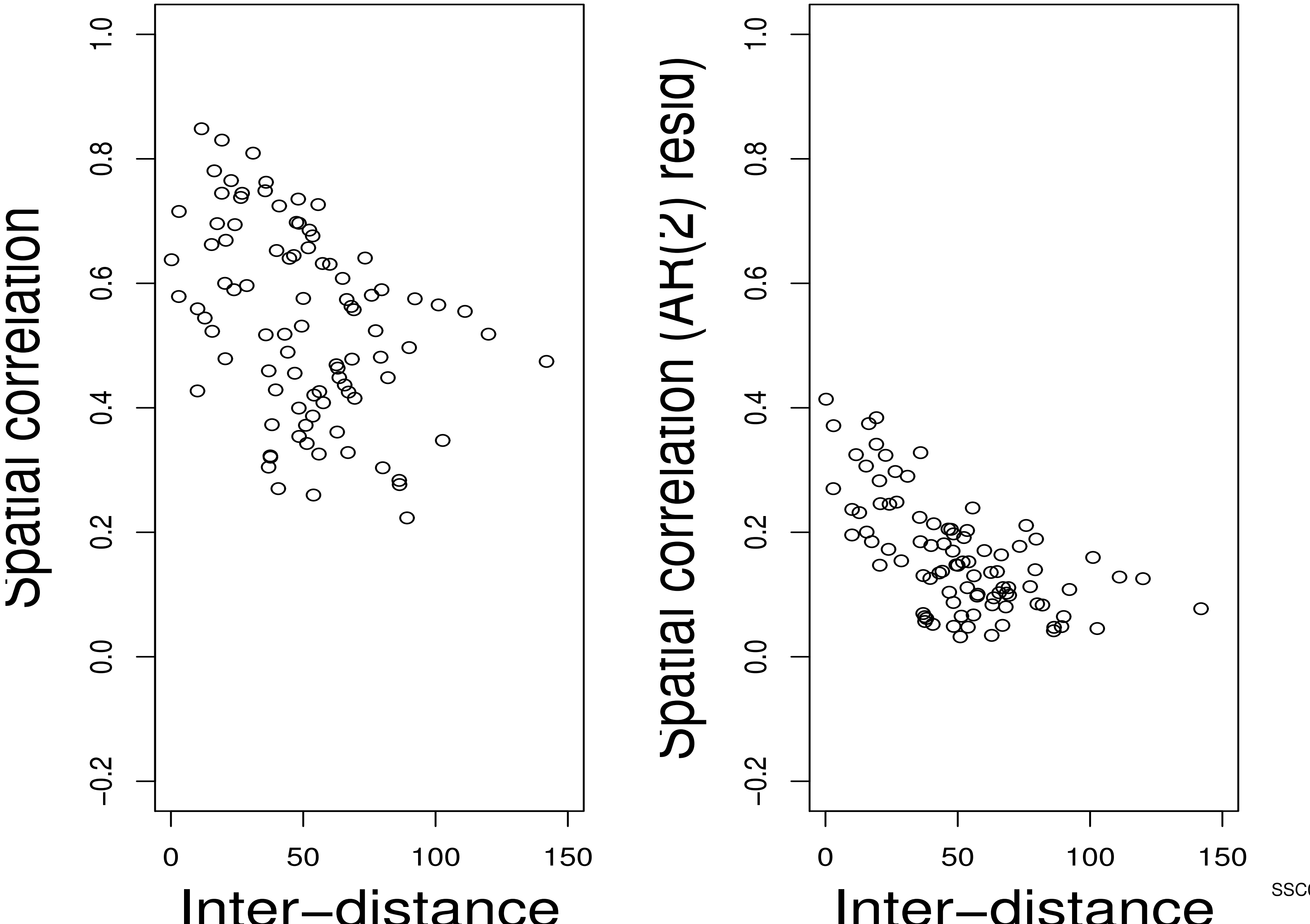
Site	1	2	..	10
1	100	41	..	37
2	41	100	..	25
..	..	..	..	..
10	37	25	..	100

for E'_{it} (**AFTER**):

Site	1	2	..	10
1	100	21	..	16
2	21	100	..	12
..	..	..	..	..
10	16	12	..	100

Current directions

Spatial Correlation Leakage



Current directions

Spatial Correlation Leakage - Simple Case (Zidek et al 2002) AR(1) Model: $y_t(s) = \alpha y_{t-1}(s) + \epsilon_t(s)$

$\epsilon_t(s)-$ time independent with spatial correlation

Spatial correlation:

$$cor(\epsilon_t(s), \epsilon_t(s')) = cor(y_t(s), y_t(s')) \ -$$

$$\frac{\alpha}{\sqrt{(1-\alpha^2)}}[cor(y_{t-1}(s), \epsilon_t(s') + cor(y_{t-1}(s'), \epsilon_t(s)]$$

Cross-corr $= 0 \rightarrow cor(y_t(s), y_t(s')) = cor(\epsilon_t(s), \epsilon_t(s'))$

Correlation leakage occurs since sample ones $\neq$ 0

- substantial when α is large.

Current directions

Direction 2: Example (continued).
Possible solutions:

- avoid modelling small time scale autocorrelations:
  - stack 24 hourly responses into a 24×1 random vector -
    use MAR(1) (multivariate autoregressive model) over days
  - gain robustness against mis-specification
  - avoid deleterious effects of mis-specification
  - but requires a good multivariate spatial predictor
- ??use dynamic state space model like Huerta et al above - value unclear at this time

Current directions

Direction 3: Big space-time domains
EG climatology

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

Current directions

Direction 3 (continued):

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

Current directions

Direction 3 (continued):
May be strength in unity but:

- big gulf between two cultural “attitudes”
- communication between camps strained
- approaches very different
- route to reconciliation unclear

General framework:

- measurement model
- process model
- parameter model

Current directions

Direction 3 (continued):
Approaches to reconciliation - depend on: purpose; context; # of (differential) equations; etc.
Can address measurement, process or parameter.

With many equations (e.g. 100):

- build a better predictive response density f(field response|deterministic model outputs)
  e.g. input model value as prior mean
- view model output itself as response and build joint density
  $f(fieldresponse, modeloutput) =$
  $\int f(fieldresponse|\lambda) f(modeloutput|\lambda) \times \pi(\lambda|data) d\lambda$

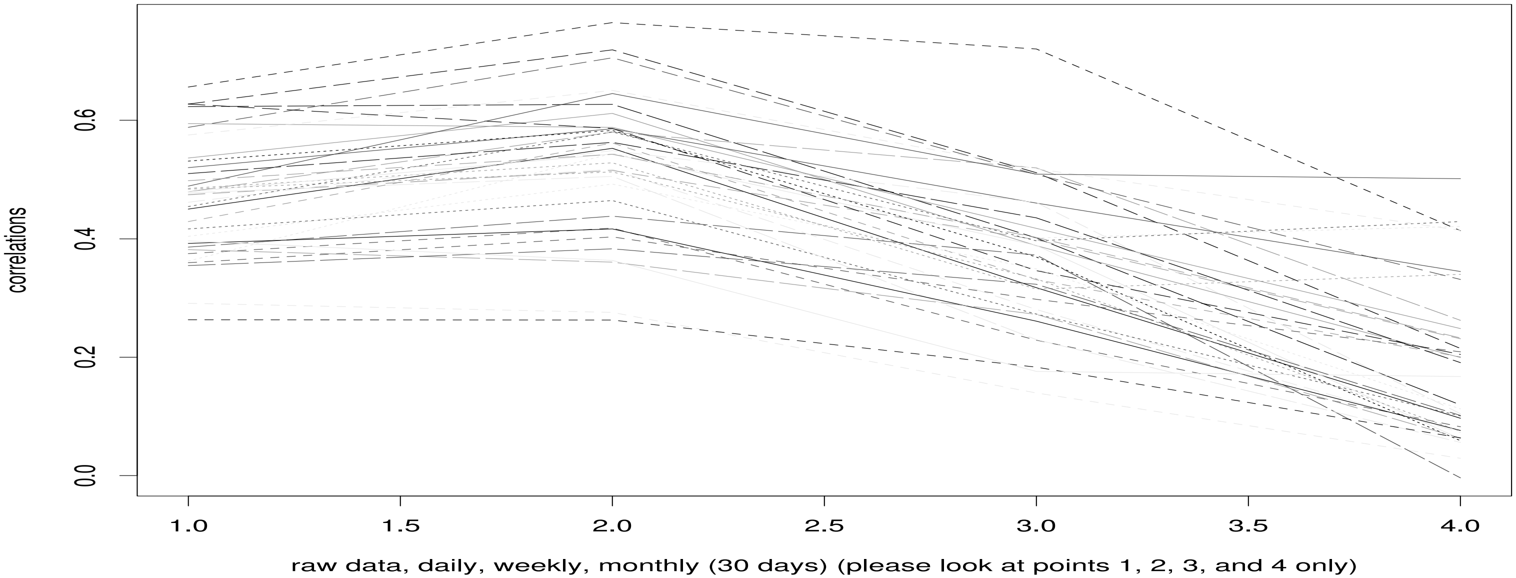
Current directions

Direction 3 (continued): Few equations (de's)
Example: $dX(t)/dt = \lambda X(t)$.

- solve it and make known or unknown constants uncertain (i.e. random):
  $X(t) = \beta_1 \exp \lambda t + \beta_0$
- discretize the de and add noise to get a state space model: $X(t+1) = (1+\lambda)X(t) + \epsilon_t$
- use functional data analytic approach - incorporate de through a penalty term as in splines
  $\sum_t (Y_t - X_t)^2 + (smoothing\ parameter) \int (DX - \lambda X)^2 dt$

Diminishing intersite correlation

Inter-site correlations for Vancouver's PM_{10} decline with span in days of the maximum



Other directions

- better state space models
- big datasets
  - multi-resolution analysis
  - ensemble state space modelling
  - reducing dimension through orthogonal expansions
- misaligned data and other such structural problems
- spatial covariance modelling
- non- Gaussian Kriging

Non-Gaussian kriging

Example: The spatial process is binary. Model it by

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

where X is a space-time process modeled by any of the methods described above.

The mis-specification of exposure in health impacts studies can be modeled in a similar way as shown in the next sub module shows.

4. Measurement error

Measurement error

Space- time modelling can mitigate unpredictable, pernicious effects in environmental epidemiology. This sub-module reviews error types, effects & modelling

Error types

- missing data
- classical & Berkson $\Rightarrow$ **2.1**
- non-differential & differential
  - **non-differential** if conditional on true value x, health outcome Y independent of measured predictor X - otherwise **differential**
- structural & functional
  - **structural** means x is random, otherwise **functional**
- misclassification - error in a binary response

2.1 Classical and Berkson types

- Classical: x = "true" value measured with error to get X:

$$X = x + \epsilon$$

  x, ϵ **uncorrelated**

- **Berkson:** e.g. experimentor sets “control” at level X but output is (unmeasured) x satisfying

$$x = X + \epsilon$$

  equivalently

$$X = x + \epsilon$$

  (assuming symmetry of ϵ’s distribution) - now x and ϵ are **correlated**!!

⇐**Back**

Effects

- effects of binary exposure variables, x: reduction in apparent effect if *non-differential*
- same with linear regression & continuous exposures (classical error model) but not with *Berkson*
- generally effects vary, hard to predict - best: reduce measurement error by good design
- for nonlinear models effects more subtle

Error effects: nonlinear models

Suppose

- $(Y, x, X) \sim$ normal
- $E[Y \mid x] = \exp[\beta x]$

. Thus

- $E[Y \mid X] = E[\exp[\beta x] \mid X] = \exp[\beta\beta_{xX}X + \beta^2\sigma_{x\cdot X}/2]$ if Y,X independent given x

Residual variance $\sigma_{x\cdot X}$ = precision of X.

- if 0 fit $Y = \exp bX$ bias-correct $\hat{\beta} = b/\beta_{xX}$ like linear case
- if $\neq 0$ bias wants to inflate b, imprecision to deflate b

(Large residual variance puts fitted β close to 0.)

Transfer of causality

Y given $x \sim$ Poisson $(\exp \alpha_0 + \alpha_1 x)$, x = hazard, w = covariate, both measured non-differentially as X, W with correlation ρ & σ_X^2»σ_W^2.

- Fitted model: $a_0 + a_1 X + a_2 W$.
- Result: a_1 nonsignificant; a_2 significant if $\rho = 0.9$ while $\sigma_X^2 > 0.5$ when $\sigma_W^2 = 0$, for example

Conclusion: Although x causes Y, W is identified instead. Causality 'transferred' through measurement error & collinearity. [**Zidek, Wong, Le, Burnett (1996)**]

Role of spatial prediction

- Basic building blocks: uncorrelated clusters i
- health outcomes (eg deaths) $\{Y_{it}\}$, for timepoint t (eg day), & cluster i
- pollution concentration (& covariate) vectors $\{X_{it} = (X_{it1}, \ldots, X_{itk})\}$ may be hi-pass filtered to unmask blip effects
- effects model

$$E[Y_{it} \mid X_{it}, \mathbf{a_i}] = m_{it} \exp\left(\mathbf{a}_i^T X_{it}\right)$$

  - m_{it} a fixed factor accounting for population size, day of week & low frequency seasonal components

2.3 Clusters

Clusters	**Data**
subjects eg mice	repeated measures eg tumors
hospitals	daily admission counts
Census	auto-correlated
Subdivisions (CSD)	daily death counts
years	spatially correlated
	CSD school absences

⇐Back

Borrowing strength

Small insignificant effects for each cluster i can be significant in the aggregate if pattern consistent. $\Rightarrow$**2.5**

How to aggregate?

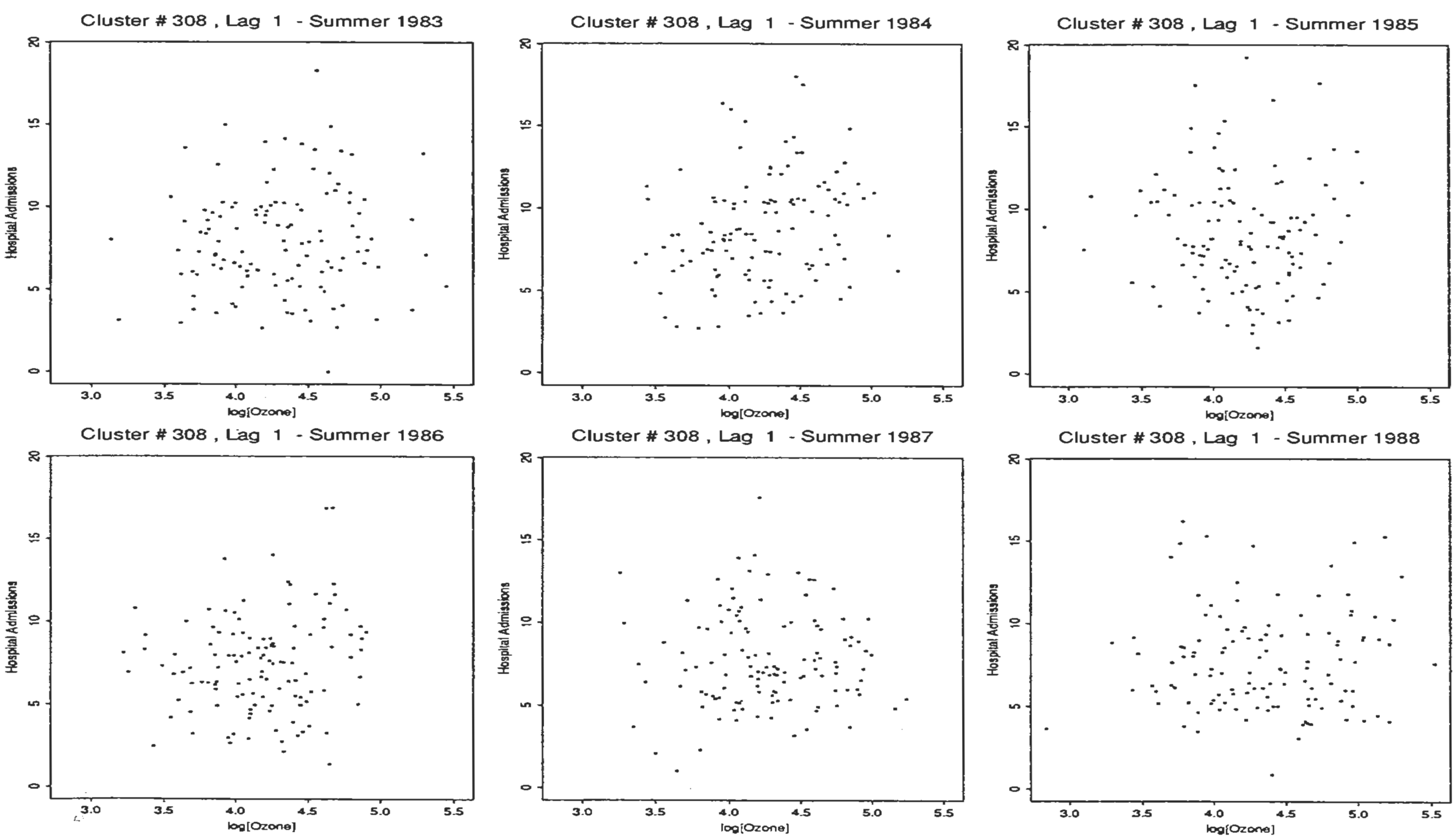
- Use random effects model:

$$\mathbf{a}_i = \beta + \mathbf{b}_i,\ \mathbf{b}_i^{vector} \sim N(0, D)$$

- Here $\mathbf{b}_i$ is a random effects vector, the deviations for cluster i from the population levels, β

2.5 Subtle effects

Not apparent: strong association between ozone and admissions in plots of daily $\log O_3$ vs hospital admissions in this census subdivision of Toronto, 1983 (upper left)-1988



Ambient monitoring

Not all monitoring sites measure the same pollutants.

- networks set up for different purposes often amalgamated
- some purposes not foreseen & hence no gauges originally attached (eg importance of $PM_{2.5}$ only recently recognized)

Ambient pollution levels are unmonitored over many large urban areas Hence:

ambient levels can be a poor surrogate for exposure

$\Rightarrow$ need for spatial predictive methodology

Using spatial prediction

Assumptions:

- $E[Y_{it} \mid X_{it}, \mathbf{a}_i] = \zeta(\mathbf{a}_i^T X_{it})$
- $Var[Y_{it} \mid X_{it}, \mathbf{a}_i] = \phi\zeta(\mathbf{a}_i^T X_{it})$
  - ϕ = overdispersion parameter

- Y_{it_1},Y_{it_2} $t_1 \neq t_2$ independent conditional on $\mathbf{a}_i$,$X_{it_1} X_{it_2}$ a ("working assumption")

$\Rightarrow$**2.6**

2.6 GEE approach

Only the spatial predictive distribution's mean and variance are needed in this approach!! [**Zeger, Liang (1986)**]

- By a similar linearization of the nonlinear mean and variance functions of $\{a_k\}$ additional simplification is gained. Now only their means and variances are needed!
- Assuming $\{Y_{kt}\}$ are (quasi) normal (**GEE approach!!**) enables estimation of the β, $\{b_k\}$ etc.

Covariance approximation

$$
\begin{aligned}
Cov\left(Y_{it_1}, Y_{it_2} \mid \mathbf{a_i}\right) &\approx \Lambda_{it_1t_2}(\mathbf{a}_i) \text{ where} \\
\Lambda_{it_1t_2}(\mathbf{a}_i) &= \delta_{it_1t_2}\phi E\left(Y_{it_1} \mid \mathbf{a}_i\right) \\
&\quad + \zeta'\left(\mathbf{a}_i^T \mathbf{z}_{it_1}\right)\zeta'\left(\mathbf{a}_i^T \mathbf{z}_{it_2}\right)\mathbf{a}_i^T \mathbf{G}_{it_1t_2}\mathbf{a}_i.
\end{aligned}
$$

[Zidek, Le, Wong, and Burnett (1998)]

Lindstrom-Bates Approximation

Suppose $\mathbf{a}_i \simeq \mathbf{a}_i^{o\ fixed} \equiv \beta_i^o$. Then

$$E(Y_{it}|\mathbf{a}_i) \simeq \zeta(\mathbf{a}_i^{oT} z_{it}) + \hat{Z}_{it}(\mathbf{a}_i - \mathbf{a}_i^o) + \cdots \equiv \eta(\mathbf{a}_i)$$

where [**Lindstrom,Bates (1990)**]

- $\hat{Z}_{it} = \zeta'(\mathbf{a}_i^{oT} z_{it}) z_{it}^T$

Further with $E(\mathbf{a}_i) = \beta$ & $\mathrm{Cov}(\mathbf{a}_i) = \mathrm{D}$,

$$E(Y_{it}) \simeq \mu(\mathbf{a}_i^o) \equiv \zeta(\mathbf{a}_i^{oT} z_{it}) + \hat{Z}_{it}(\beta - \mathbf{a}_i^o)$$

$$+\frac{1}{2}\zeta''(\mathbf{a}_i^{oT} z_{it}) \left\{ z_{it}^T [D + (\beta - \mathbf{a}_i^o)(\beta - \mathbf{a}_i^o)^T] z_{it} \right\}$$

Similar approximations for conditional & unconditional covariances. ⇒2.7

Computing random effects estimates

To solve the estimating equations iteratively using "Fisher's scoring algorithm" requires the gradient of $\mathcal{W}_i$ w.r.t. $\mathbf{b}_i$, i.e. $\mathbf{A} = -\hat{Z}_i^T \Lambda_i \hat{Z}_i - D^{-i}$. Getting the next value of $\mathbf{b}_i$ in an iterative solution of the estimating equation involves finding $\mathbf{b}_i^*$ as solution of

$$\mathcal{W}_i + \mathbf{A}[\mathbf{b}_i^* - \mathbf{b}_i] = 0$$

That is

$$\mathbf{b}_i^* = D\hat{Z}_i^T \Sigma_i^{-1} \hat{\mathbf{r}}_\mathbf{i}$$

where $\hat{\mathbf{r}}_\mathbf{i} = y_i - \eta(\beta + \hat{\mathbf{b}}_i) + \hat{Z}_i \hat{\mathbf{b}}_i$

Estimating β

After estimating the $\{\mathbf{b}_i\}$ for fixed β at stage K, update the latter's estimated by "marginalizing out" the $\{\mathbf{b}_i\}$ & maximizing the marginal posterior. Result: estimating equations solved numerically [analogous to the random effects] The result:

$$\hat{\beta}^* = \hat{\beta} + [\sum_i \hat{X}_i^T \Sigma_i^{-1} \hat{X}_i]^{-1} [\sum_i \hat{X}_i^T \Sigma_i^{-1} \mathbf{r}_i]$$

is fixed & used for the $\{\beta_i^o\}$ in stage $K+1$ to get revised versions of the $\{\mathbf{b}_i\}$ & so on.

Likewise **D** and ϕ may be estimated by the maximizing the quasi log likelihood. [**Burnett,Krewski (1994)**]. Robust estimates of the covariance matrix of the coefficient estimates vector can also be found.

2.4 Effects significant?

Central issue: Is $a_{kj} = 0$ for pollutant j? for some specific k? All k?

NOTES:

- Effects subtle not significant for specific k
- X_{kt} unmeasured for many k-**need spatial prediction!!**

⇐**Back**

2.7 Random effect estimates

Assume conditional on $\{\mathbf{a}_i\}$, $\{Y_i\}$(vector form) normally distributed. Given $Y_i = y_i$ (observed for all i) & β, log posterior density of the $\mathbf{b}_i$ is

$$\propto -\frac{1}{2}\mathbf{r}_i^T \Lambda_i^{-1} \mathbf{r}_i - \frac{1}{2}\mathbf{b}_i^T D^{-1} \mathbf{b}_i$$

with $(\mathbf{r}_i \equiv y_i - \eta_i(\beta + \mathbf{b}_i))$. Its mode solves

$$\mathcal{W}_i \equiv \hat{Z}_i^T \Lambda_i^{-1} \mathbf{r}_i + D^{-1} \mathbf{b}_i = 0$$

to yield $\hat{\mathbf{b}}_i$. [Λ fixed at previous iteration & η linear]

⇐ **Back**

Summary

- explosion of interest in space time modelling
- reconciliation of physical and statistical modelling approaches vital - long way to go
- large datasets - immense challenges
- spatial covariance modelling has begun - finally!
- good multivariate models vitally needed
- spatial models for non-Gaussian processes need more work
- work on extreme fields has only just begun
- **vital need to integrate space time process theory and environmental health risk analysis**

2.2 Error effects: nonlinear models

Suppose

- $(Y, x, X) \sim$ normal
- $E[Y \mid x] = \exp[\beta x]$

. Thus

- $E[Y \mid X] = E[\exp[\beta x] \mid X] = \exp[\beta \beta_{xX} X + \beta^2 \sigma_{x \cdot X}/2]$ if Y,X independent given x

Residual variance $\sigma_{x \cdot X}$ = precision of X.

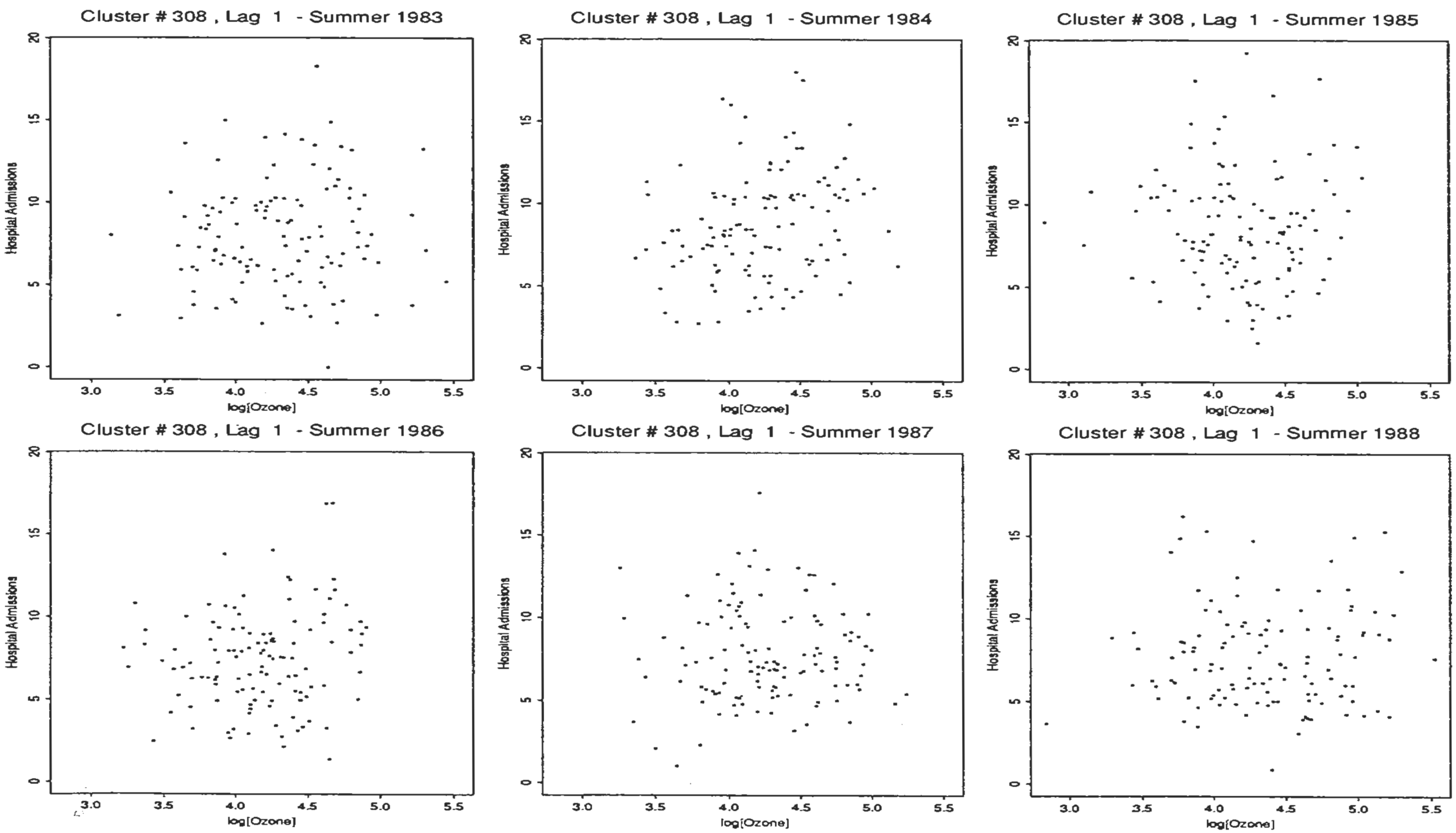
- if 0 fit $Y = \exp bX$ bias-correct $\hat{\beta} = b/\beta_{xX}$ like linear case
- if $\neq 0$ bias wants to inflate b, imprecision to deflate b

(Large residual variance puts fitted β close to 0.)

⇐**Back**

2.5 Subtle effects

Not apparent: strong association between ozone and admissions in plots of daily $\log O_3$ vs hospital admissions in this census subdivision of Toronto, 1983 (upper left)-1988



2.6 GEE approach

Only the spatial predictive distribution's mean and variance are needed in this approach!! [**Zeger, Liang (1986)**]

- By a similar linearization of the nonlinear mean and variance functions of $\{a_k\}$ additional simplification is gained. Now only their means and variances are needed!
- Assuming $\{Y_{kt}\}$ are (quasi) normal (**GEE approach!!**) enables estimation of the β, $\{b_k\}$ etc.

2.7 Random effect estimates

Assume conditional on $\{\mathbf{a}_i\}$, $\{Y_i\}$(vector form) normally distributed. Given $Y_i = y_i$ (observed for all i) & β, log posterior density of the $\mathbf{b}_i$ is

$$\propto -\frac{1}{2}\mathbf{r}_i^T \Lambda_i^{-1} \mathbf{r}_i - \frac{1}{2}\mathbf{b}_i^T D^{-1} \mathbf{b}_i$$

with $(\mathbf{r}_i \equiv y_i - \eta_i(\beta + \mathbf{b}_i))$. Its mode solves

$$\mathcal{W}_i \equiv \hat{Z}_i^T \Lambda_i^{-1} \mathbf{r}_i + D^{-1} \mathbf{b}_i = 0$$

to yield $\hat{\mathbf{b}}_i$. [Λ fixed at previous iteration & η linear]

⇐ **Back**