Stat547L: Spatial Statistics

Jim Zidek [2]

[2]University of British Columbia

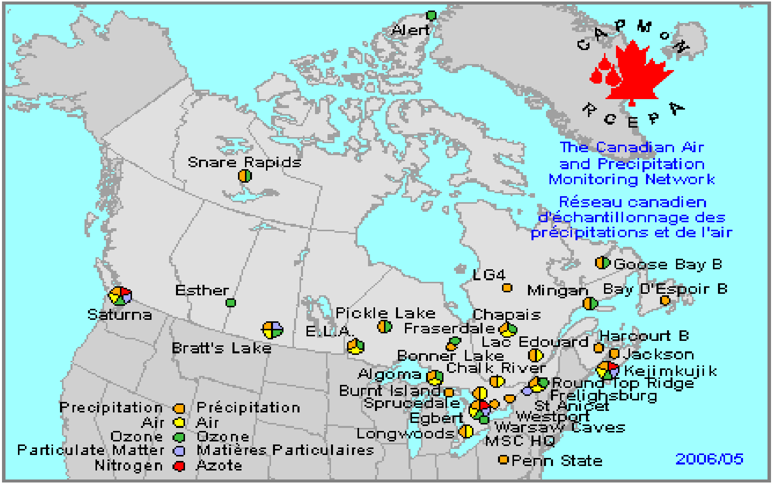
Term 1, 2017/18

Designing monitoring networks

In this lecture you will learn:

- Uncertainty, information & **the need for monitoring**
-
- That getting **more information can increase not decrease your uncertainty** for the phenomenon of interest
- The difference between **aleatory uncertainty** and **epistemic uncertainty**
- How to **optimally redesign a monitoring network**
- **New challenges facing designers**

Example # 1 Capmon network.



NOTES on Capmon network

- 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

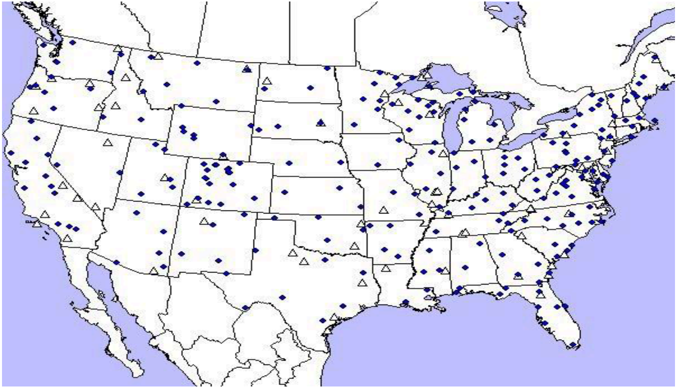
---

Lessons learned:

A network’s purposes often diverse & unforeseen.

Example #2. NADP/NTN network

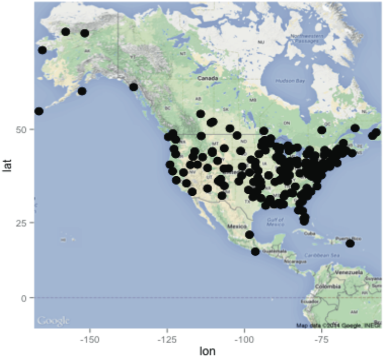
Monitors multivariate responses related to “acid precipitation”– another network merger with better defined siting rules!



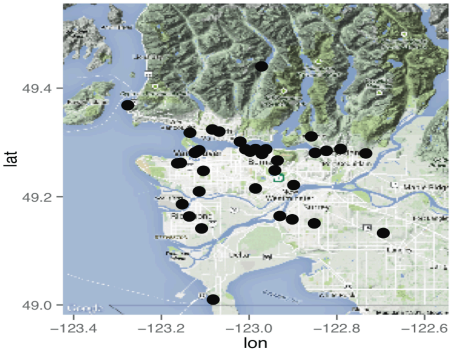
Rules governing siting and types of NDP/NTN monitors:

> *"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."*⁚

Example # 3. NA mercury monitoring sites



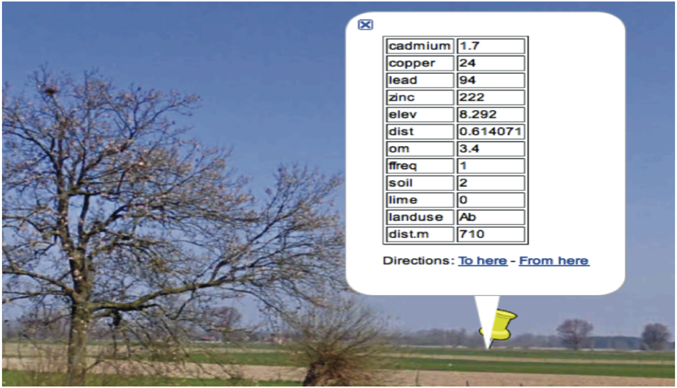
Example # 4. Vancouver air quality monitors



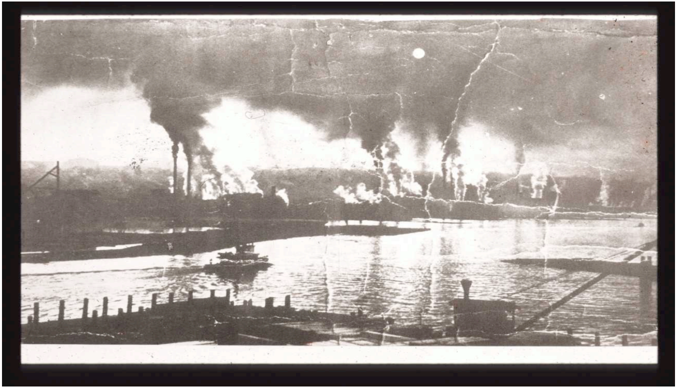
Example # 5. Meuse river bank soil sampling sites

The "street view"

Click on yellow tack:



False creek circa 1900 before monitoring



Courtesy of Professor Douw Steyn

False creek circa 1990

Courtesy of Professor Douw Steyn

What do monitoring sites look like?

At Kitsilano High School

Courtesy of Professor Douw Steyn

What do monitoring sites look like?

Near Robson Square

Courtesy of Professor Douw Steyn

Some specific objectives of monitoring

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

- **Non-compliance detection** given regulatory standards
- **Health risk analysis**
  - & provide good estimates of relative risk
  - determine how well sensitive sub-populations are protected
  - can include all life, not just human
- **Time trend analysis**
  - are things getting worse?
  - is climate changing?

General purposes

Explore/reduce uncertainty re the environment:

- One form of uncertainty (***aleatory***) is irreducible (e.g. outcome of fair die toss
- The other (***epistemic***) (e.g. whether the die is fair) can increase or decrease. Implication: even an optimum design must be regularly revisited

But what is “uncertainty”?

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

Exercises

8.1 Suppose $X \sim N(0,1)$. Prove that uncertainty about X, i.e. $Var(X \mid \; \mid X \mid < C)$ is increasing function of C.

8.2 Suppose $X \sim N(\eta, 1)$. Prove that uncertainty about X, i.e. $Var(X \; \mid\mid X \mid < C)$ is increasing function of C. Warning:Very hard! [1]

[1] I posed the problem years ago with a prize of $100 but it was not solved until Jiahua Chen (UBC)did so Chen et al. [2010].

Possible design criteria

Add monitors to (i.e. "Gauge") sites:

- **that maximally reduce uncertainty about their responses** at their space–time points [because then measuring their responses eliminates their uncertainty]
- **best minimize uncertainty** about ungauged site responses
- give best **process parameter estimates**
- best **detect non-compliers**

Designer challenges:

- **Multiplicity of objectives**
- **Unforeseen & changing objectives**
- **Multiple responses** at each site: which to monitor?
- Including **prior knowledge**
- **Good process models**
- **Fit with existing networks**
- **Fit with reality!!!**

Exercises

8.3 How might design criteria be arrived at in practice? Who should be responsible for setting them?

8.4 Monitor placement should recognize such things as the geographical distribution of impacted populations (eg trees or fish). How can an optimal design be determined in such a context? Research question!

8.5 Develop a design theory in a non–Gaussian context. Research question!

Approaches to design

A big field [Zidek and Zimmerman, 2009]

- **Space-filling designs**
- **Probability based designs**
  - simple random sampling
  - stratified, multistage designs
  - e.g. (1) EPA's survey of lakes; (2) the EMAP project
- **Model Based**
  - Regression model approach
    - eg to estimate the slope put 1/2 the data at each end of the data range
  - Random fields (prediction, e.g. entropy) approach
- **Other.** In particular Zhengyuan Zhu (UNC) incorporates both of the latter, prediction and parameter estimation.

Entropy based approach

"Gauges" sites with greatest "uncertainty"

- **uncertainty = entropy**
- **maximally reduces uncertainty about "ungauged" sites**
- **best estimates predictive posterior distribution under entropy utility**

By - passes specification of objectives

Long history[2], currently popular

[2] General: Good [1952], Lindley [1956], Shewry and Wynn [1987]. Network design: Caselton and Zidek [1984],Sebastiani and Wynn [2002],Zidek et al. [2000]

What is entropy?

Let **probability of uncertain event E** (e.g. heads on bent coin) be:

$$p = P(E)$$

That uncertainty becomes 0 when outcome becomes known.

Let **reduction in uncertainty** be

$$\phi(p) \textit{ if } E \textit{ occurs}$$
$$\phi(1-p) \textit{ if } E \textit{ does not occur}$$

Expected reduction in uncertainty:

$$p\phi(p) + (1-p)\phi(1-p$$

Simple assumptions imply

:

$$\boxed{\phi(p) = log(p)}$$

Conclusion: knowing E occurred reduces entropy by

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

Thus"**uncertainty**" about E can be quantified as the **entropy** of the two point distribution $(p, 1-p)$:

Relative entropy

But how much is that entropy?

Needs a **reference level.** Complete uncertainty about E would be the two point distribution $(q, 1-q)$ with $q = 1/2$. Thus the relative entropy would be

$$I(p, q) = p \log \left\{ \frac{p}{q} \right\} + (1-p) \log \left\{ \frac{(1-p)}{(1-q)} \right\}$$

This is f the **Kullback-Leibler** measure of the deviation of $(p, 1-p)$ from that reference level (the "state of equilibrium" in physics (thermodynamics).

Multiple events

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

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

$$I(f,g) = \int f(x) \log\left\{\frac{f(x)}{g(x)}\right\} dx > 0$$

.

Commonly $g \equiv 1(unitsoff)$. 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.

Entropy framework

Adopt a Bayesian framework.

- $\boldsymbol{\theta}$: process parameter vector.
- **Y** : process response vector at future time T+1 including all sites (gauged & ungauged) conditional on
- D_T set of all available data at time T
- h_1 & h_2: baseline reference densities against which to measure uncertainty.

- Finally compute the entropies **with θ random**

$$
\begin{aligned}
H(\mathbf{Y} \mid \theta) &= E[-\log\left\{\frac{f(\mathbf{Y} \mid \theta, D)}{h_1(\mathbf{Y})}\right\} \mid D_T] \\
H(\theta) &= E[-\log\left\{\frac{f(\theta \mid D_T)}{h_2(\theta)}\right\} \mid D_T]
\end{aligned}
$$

Then we get **fundamental entropy identity**

(**Exercise**):

$$
H(\mathbf{Y}, \theta) = H(\mathbf{Y} \mid \theta) + H(\theta)
$$

In other words, at time T

TOTAL UNCERTAINTY =
UNCERTAINTY ABOUT RESPONSE GIVEN MODEL
+ **MODEL UNCERTAINTY**

Design goal: add sites

Add (or subtract) sites at time $T + 1$. Let's focus on adding new sites to an existing network

- $\mathbf{Y} = (\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)})$ = all site responses, time T +1
- $\mathbf{Y}^{(2)}$ for site currently gauged (time T)
- $\mathbf{Y}^{(1)}$ for sites currently ungauged (time T)

DESIGN GOAL: Partition $\mathbf{Y}^{(1)} = (\mathbf{Y}^{(rem)}, \mathbf{Y}^{(add)})$ at time T so that

- $\mathbf{Y}^{(rem)}$: future ungauged sites
- $\mathbf{Y}^{(add)}$ future new network stations.

Entropy decomposition thm

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

Fundamental identity:

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

where

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

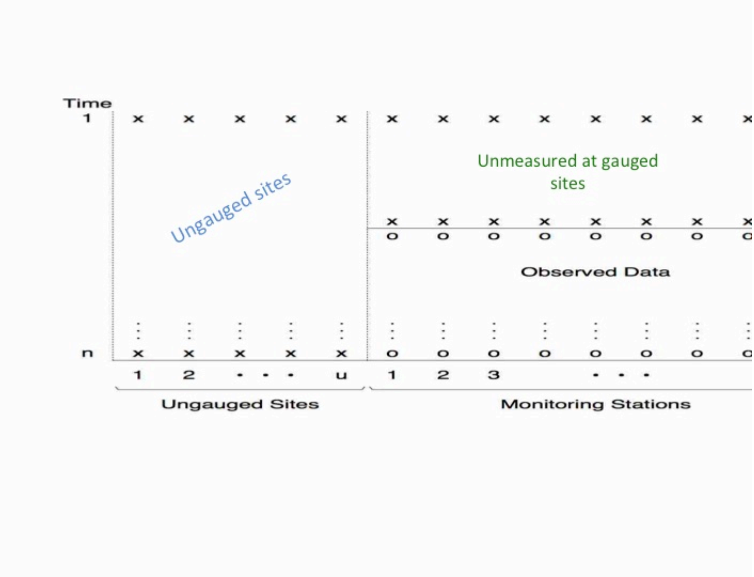
and

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

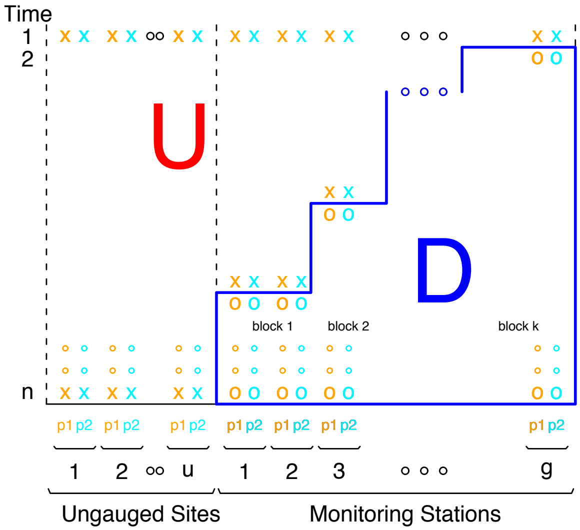
.

Theorem: Maximizing MEAS = Minimizing MODEL + PRED

The environmental process



The multivariate data staircase



Possible inferential objectives

- **Forecasting:** process values at time $T + 1$
- **Spatial prediction:** process values in U at time T
- **Hindcasting:** past values of the process at times $t < T$ both in D sites as well as U sites, e.g. cancer vs air pollution.

Design objective for expanding the network:

Choose sites from **U** to "add" to the existing network at time $T+1$.

NOTATION: u: ungauged sites at time T
g: gauged sites at time T $p = u + g$: total number of sites

The process model

Transform responses as necessary. **Remove regular temporal & trend components. Assume EnviroStat model**: at time $t = 1, \ldots, T+1$

$$\mathbf{Y}_t^{1\times p} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma} \overset{ind}{\sim} N(\mathbf{x}_t^{1\times k} \boldsymbol{\beta}^{k\times p}, \boldsymbol{\Sigma}^{p\times p})$$

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

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

where $\otimes$ denotes **Kronecker product**, F^{-1} covariance between rows; $\boldsymbol{\Sigma}$ covariance between columns.

The predictors

The x_t are assumed to be the same for all sites in the region

Question: what about site specific predictors?

- Random predictor at site j W^j of Y^j:
- model $[Y^j, W^j]$ then $[Y^j \mid W^j]$
- Nonrandom predictor w^j: fit $\beta_0^j = \gamma w_j$ via empirical Bayes

Matric-normal extension

:

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

NOTE: The IW distribution is the inverse χ^2 for matrices. It can be generalized to the GIW - it had different numbers of degrees of freedom for different steps in the data staircase for example.

Bartlett decomposition:

Reading Notation: "u" means "ungauged" sites; "g" means "gauged" sites (with $p = u + g$:

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

Bartlett decomposition:

$$\Sigma = T \Delta T'$$

where

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

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

Generalized Inverted Wishart Distribution

Reading

Let

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

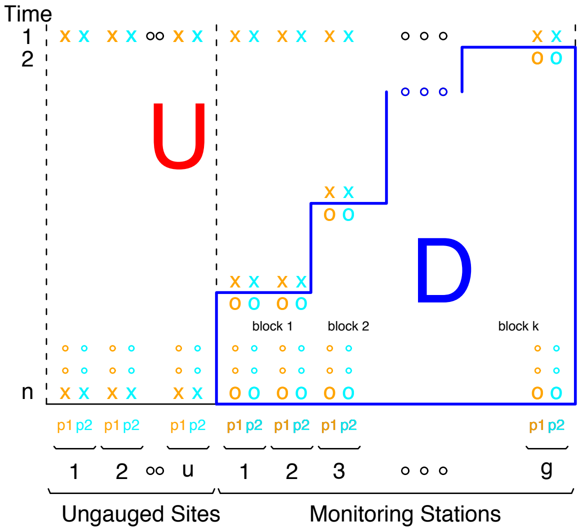
Definition: For $\boldsymbol{\Sigma} \sim GIW(\Psi, \delta)$, interate the following decomposition starting with $\boldsymbol{\Sigma}^{[g]}$ to get successive deg's freedom $\delta_0, \delta_1, \ldots$:

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

$$\Gamma^{[u]} \sim IW(\Lambda_0 \otimes \Omega, \delta_0)$$

$$\tau^{[u]} \mid \Gamma^{[u]} \sim N\left(\tau_{0\,u}, H_0 \otimes \Gamma^{[u]}\right)$$

Reminder: the multivariate data staircase



Additional notation

$\mathbf{g}_i^o$ denotes **observed responses at gauged site** i sites

$\mathbf{g}_i^m$ denotes **missing data at gauged site** i, $i = 1, \ldots, g$.

Then $D_T = \{\mathbf{g}_1^o, \ldots, \mathbf{g}_g^o\}$.

Then [Le and Zidek, 2006]:

$$\begin{aligned}\left[\mathbf{Y}^{[u]} \mid D_T, \mathcal{H}\right] \quad \sim \quad & \left[\mathbf{Y}^{[u]} \mid Y^{[g_1^m,\ldots,g_k^m]}, D_T, \mathcal{H}\right] \times \\ & \prod_{j=1}^{k-1} \left[\mathbf{Y}^{[g_j^m]} \mid \mathbf{Y}^{[g_{j+1}^m,\ldots,g_k^m]}, D_T, \mathcal{H}\right] \\ & \times \left[\mathbf{Y}^{[g_k^m]} \mid D_T, \mathcal{H}\right].\end{aligned}$$

and $\mathbf{Y}^{[u]}$ only appears in first factor:

$$\left(\mathbf{Y}^{[u]} \mid \mathbf{Y}^{[g_1^m,\ldots,g_k^m]}, D, \mathcal{H}\right) \sim$$

$$t_{n\times u\ p}\left(\mu^{[u|g]}, \text{Dispersion}, \delta_0 - u + 1\right).$$

$$\text{Dispersion} = (\delta_0 - u + 1)^{-1}\Phi^{[u|g]} \otimes (\Lambda_0 \otimes \Omega)$$

- Here Λ_0 represents spatial covariance that links U to G
- Ω represents within site covariance e.g. a four hour block of ozone concentrations or several different chemical species
- $\delta_0 - u + 1 > 0$ is required to avoid a degenerate t-distribution. No such thing as a free lunch. Must keep u to realistic size. Don't see this with purely Gaussian process models.

$$\text{Dispersion} = (\delta_0 - u + 1)^{-1} \Phi^{[u|g]} \otimes (\Lambda_0 \otimes \Omega)$$

Entropy calculation

The relevent entropy for $Y^{[u]}$ is from that first factor in the entropy decomposition:

$$H\left[Y^{[u]} \mid Y^{[g_1^m,\ldots,g_k^m]}, D\right] \quad = \quad \frac{p}{2}\log|\Lambda_0| + \text{irrelevant terms}$$

- Ungauged sites u must be partitioned into 'add' and 'rem' sites in optimal way.
- Applying the Bartlett decomposition tells us we must find the submatrix of $|\Lambda_0[add, add]|$ whose sub–determinant $\mid \Lambda_0 \mid$ corresponds to the 'add' sites in the partitioned $Y^{[u]}$.

NOTE: Will simultaneously minimize the entropy left in the 'rem' sites.

Finding the ‘add’ sites

- **NP-Hard:** No exact algorithms for big networks
- Inexact Methods:
  - **Greedy** - add (or subtract) one at a time
  - **Reverse Greedy** - subtract one at a time
- Exact Methods:
  - Complete **enumeration**
  - **Branch and bound**

How many sites?

Compute:

$$\frac{\text{Entropy}(n)}{n}$$

where n is the number of sites for success n's. The ratio will initially increase than decrease. Stop at the maximum (bang for the buck) n.

Application: Redesign Vancouver's hourly PM_{10} monitoring field.

- Increase **10** monitoring sites to **16**–add **6** new stations from among **20** possible sites [3].
- Use entropy approach with **Normal–GIW predictive distribution.**

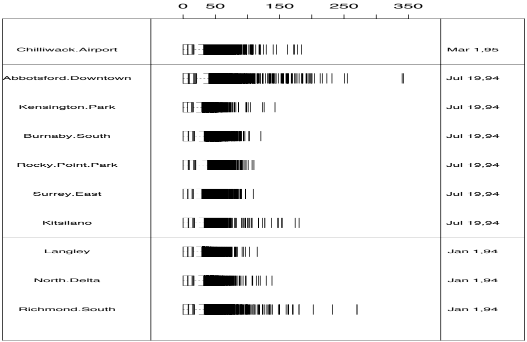
[3] Resembles actual study Nhu Le & I did [Ainslie et al., 2009]

Method (Will see details in the Practicum:

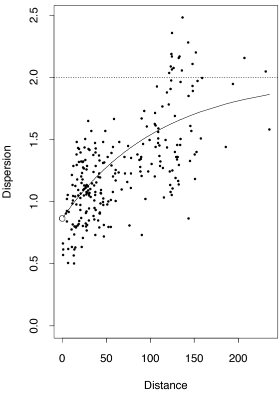
- **Transform data**
- **Remove all-site (common) spatio–temporal trends**
  This means fit the same model to all sites.
- **Whiten" residuals** with common 24 (hour) dimensional multivariate AR(1) model
  same idea - result is approximately unautocorrelated temporal process
- 10 different station startup times means **monotone ("staircase") data pattern** – use GIW distribution with different degrees of freedom (δs) for each staircase step
- **Select the 6 new stations** that jointly maximize conditional entropy

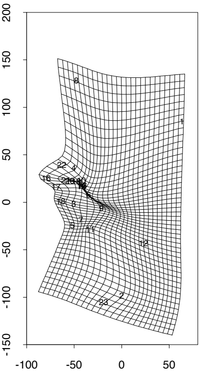
Preliminary data analysis:

PM_{10} levels at the 10 existing stations. **Note differing startup times.**



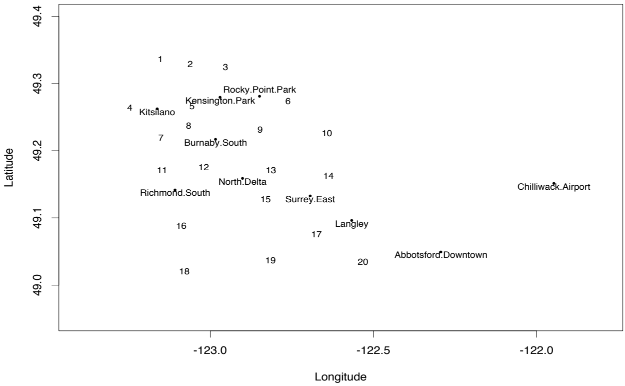
Vancouver’s ozone field is clearly non-stationary and the Sampson–Guttorp method was used to estimate the hypercovariance in the GIW matrix.



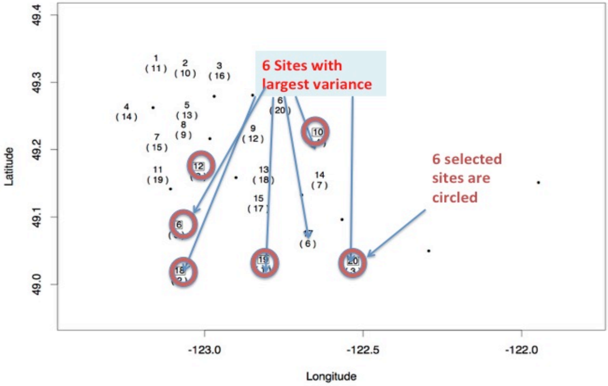


The original 10 PM_{10} monitoring sites

Also **20** possible new locations – subregions with no existing monitors & big populations.



Locations of the old and **newly selected 'add' sites (square brackets).** The ranks of the **20** sites by estimated variance is in curved brackets and those selected. Notice **6**^{t}*h*selected site.



Alternative design strategies: #1

Non-Compliance detection design: Select sites with max prob of finding sites out of compliance with regulatory standards.

Probability & best design is day–dependent!

- **which day?**
- **a simulated future day? Average day? Bad day?**

How to implement?

- **monitor sites most likely to noncomply?**
- **do not monitor sites least likely?**
- **what about existing sites?**

Example: What if the 6 sites had been selected to catch noncompliance rather than to maximize entropy?

Say we use 10 station hourly data for PM_{10}, February 28, 1999 & hierarchical Bayes predictive distribution

DETECTION CRITERION:

$$\textbf{argmax PR}\{\textbf{daily max } PM_{10}\ \mathbf{Y}^{6added} \geq \mathbf{50}\ (\mu g\ m^{-3})\}$$

RESULT: Same as for entropy based design!!!

NOTES:

1. Entropy does not work for detecting noncompliance nearly as well on August 1, 1998!
2. The selected new sites are now determined pretty much by their posterior estimated variances. That is because the spatial correlation is now quite small – strength hard to borrow.
3. Designing for noncompliance seems largely unexplored issue

Alternative design strategy #2

Site selection for fields of extremes: Very difficult since spatial dependence declines when looking say at monthly maxima rather than daily maxima.

Yet regulatory criteria metrics (risk) usually base on extremes!

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

Annual Arithmetic Mean: 50 μg m^{-3}

24 - hour Average: 150^{FN} μg m^{-3}

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

The bad news for fields of extremes

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**

The good news

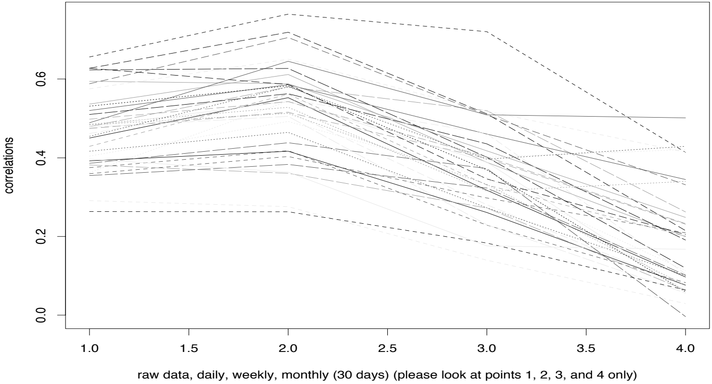
Joint distribution of extremes approximately a log multivariate t distribution. Hence can:

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**

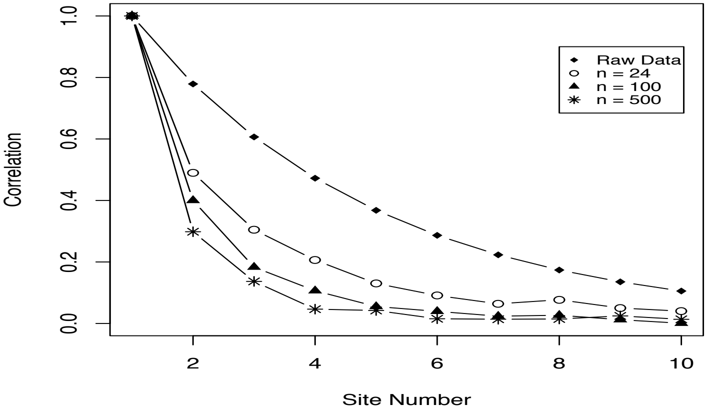
Some detail: Inter-site correlations

Inter–site dependence declines with time span of extremes for many but not all site pairs

Example: The figure shows Vancouver's PM_{10} intersite correlation between maxima computed for various time spans.



Simulation study*[4] *: *multinormal responses; maxima with varying ranges at 10 sites. Multivariate t results show smaller loss of dependence. Inter-site correlations for maxima for simulated fields of extremes. Big n = light tails.*



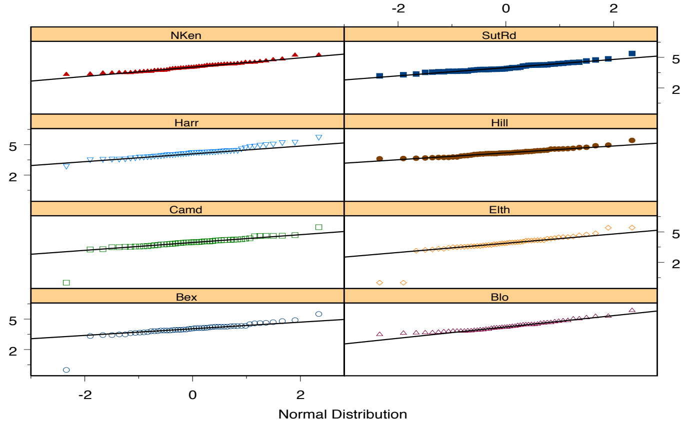
[4] Chang et al. [2007]

Exercises

23.6 Try the simulation experiment yourself and confirm that for heavier tails lead to increased intersite correlations

An approach to monitoring extremes

Empirical results $\mapsto$ 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**:



T approx approach cont'd

Empirical results $\rightarrow$ well-calibrated 95% (etc) prediction intervals. 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**

Room for optimism?

Use of log multivariate t distribution for extreme fields promising. But:

- **How far can approximation go? Need new theory**
- **Must test approximation on case–by–case basis**
- **need to compare regular - and extreme-entropy designs.**

Summary

In this lecture we have seen:

- Why environmental monitoring networks are needed and what they look like
- Why spatio–temporal theory is needed in their design
- A class of models with associated software that provide a platform for designing them.
- The entropy and other criteria for design.

Regulation and control

We saw that networks are needed for control and mitigation.
Also saw some issues arising:

- Current designs need to be revisited - fitness for use.
- Simpler control metrics than the current ones would be preferable for transparency and analysis.
- The current design criteria have not been well spelled out.

Some conclusions

- Current urban networks may be inadequate for surveillance of extremes. Much more attention needs to be paid
- But the MaxEnt strategy may be a way out of this problem.
- The EnviroStat approach needs an upgrade that takes account of recent theoretical developments .

References

B Ainslie, C Reuten, DG Steyn, Nhu Le, and James Zidek. Application of an entropy-based bayesian optimization technique to the redesign of an existing monitoring network for single air pollutants. *Journal of Environmental Management*, 90(8):2715–2729, 2009. ISSN 03014797. doi: 10.1016/j.jenvman.2009.02.016.

William F Caselton and James 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.

Jiahua Chen, Constance van Eeden, and James Zidek. Uncertainty and the conditional variance. *Statistics & probability letters*, 80(23):1764–1770, 2010.

Irving John Good. Rational decisions. *Journal of the Royal*

Statistical Society. Series B (Methodological), pages 107–114, 1952.

Nhu D Le and James V Zidek. *Statistical analysis of environmental space-time processes*. Springer Science & Business Media, 2006.

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

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.

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

J Zidek and D Zimmerman. Monitoring network designs. *Handbook of Spatial Statistics (ed. A. Gelfand, P. Diggle, M. Fuentes and P. Guttorp)*, pages 131–148, 2009.

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.