Designing Environmental Monitoring Networks for Measuring Extremes

Howard Chang, Audrey Qiuyan Fu, Nhu D. Le, James V. Zidek

Technical Report #2005-4
March 26, 2005

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

Designing Environmental Monitoring Networks for Measuring Extremes.*

Howard Chang[1], Audrey Qiuyan Fu[2], Nhu D Le[3,4], James V Zidek[4]

[1] Department of Biostatistics, The Johns Hopkins University
[2] Department of Statistics, University of Washington
[3] Cancer Control Research, BC Cancer Agency
[4] Department of Statistics, University of British Columbia

March 26, 2005

*The work was supported partly by the Natural Science & Engineering Research Council of Canada and partly by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Abstract

This paper discusses some of the problems and solutions associated with the monitoring the extremes generated by a random environmental field. One of these problems is the loss of spatial dependence (*i.e.* asymptotic independence) encountered in going from the original series of measurements to extreme values computed over fixed time intervals. The paper exhibits this loss in a case study involving particulate air pollution in Vancouver. That loss is characterized through simulation studies where the effect of increasing tail weight in the original is investigated. Overall, the problem raises concerns about the adequacy of modern environmental monitoring systems intended to protect human health against the impact of hazardous substances.

This need for protection points to the difficult conceptual problem of selecting an appropriate design criterion. Seemingly natural ways of interpreting that need lead to different objective functions and hence different designs. Side stepping this ambiguity by adopting a naive entropy maximization criterion as an objective function leads to designs that agree with the noncompliance optima in some cases but not in others.

A less naive approach is described as a way out of the morass. In that approach a multivariate Gaussian-Inverse Wishart hierarchical Bayesian distribution is proposed for suitably transformed extremes with data-based estimates of the hyperparameters. The adequacy of this model is empirically assessed by finding the relative coverage frequency of the predictive credibility ellipsoid implied by the posterior distribution. The favorable results obtained suggest this posterior adequately describes the field of transformed extremes reasonably well. Hence in conjunction with the entropy it may provide a useful foundation for design in this context. We use in a hypothetical extension of a monitoring network.

Keywords and Phrases: extreme values; Bayesian hierarchical models; space-time models; generalized inverted Wishart; multivariate extremes; spatial design; maxent; maximum entropy.
MSC 2000: Primary 62H11, 62M20, 62M40.

1 Introduction

This paper presents challenges confronted in designing monitoring networks for measuring fields of extreme responses over a not necessarily regularly spaced discrete set of sites in a geographical region. We also suggest a way of confronting some of those challenges.

Such a network may be designed to meet any one of a variety of purposes, one being the detection of noncompliance in a regulatory setting. The metrics used may involve such complex combinations of statistics as to make their distributions unavailable in explicit form. In other words they may not be the sort of extremes to which extreme value theory has been devoted. That calls for a predictive distribution of the type developed in this paper that can be used to simulate the distribution of the metric.

The high cost of establishing and maintaining a network of monitoring sites like those described in the example, along with their technical limitations, means that few are found in a typical regulatory domain. That in spite of their vital role of protecting human health. It seems reasonable to ask if their density is sufficiently high that they can play their primary role. The answer may well be negative.

After all classical design approaches emphasize inference about the response field itself rather than the extremes it generates. (See Le, Sun and Zidek 2003, 2005 for surveys.) Section 2 shows by example that one such approach, based on the entropy of the joint conditional response distribution (given the measured values) to bypass the need to specify a unique design objective, can be unsatisfactory.

Section 2 also highlights a difficulty, that spatial dependence in the response field may be lost in the extremes of that field. This poses a problem since spatial designs rely on that dependence to transfer information from the monitors to unmonitored locations. Any loss of dependence points to a need for an increased density of monitoring sites.

To conclude we note that an extensive theory has been developed for extreme values (Gumbel 1958, Leadbetter *et al* 1983, Coles 2001 and Embrechts 1997), going at least as far back as the celebrated work of Fisher and Tippett (1928). Moreover the theory of multivariate extremes is undergoing current development (Smith 2004). However that theory has proven much more challenging than its univariate counterpart.

An important property explored in that theory, which presents itself in the next section, is *asymptotic dependence*. Two variables X_1 and X_2, have that property if $\lim_{q\to 1} Pr\{F_{X_2}(X_2) > q|F_{X_1}(X_1) > q\} \neq 0$ where F denotes their cumulative distribution functions. It has been recognized that some pairs of variables in a multivariate response vector can have that property while others are independent (Coles and Pauli 2002).

While progress has been made in addressing that issue, the current theory as a whole does not yet seem implementable in the analysis of complicated response metrics from space-time fields over the large domains encountered in modern environmetrics. (A survey of that theory and the difficulties it encounters can be found in Le and Zidek 2005). Thus we have not tried to adopt it as a foundation for developing a practical approach to design.

2 Design challenges.

This section describes difficulties associated with the design of networks for monitoring extreme values of a random environmental process over a geographical domain.

2.1 Loss of spatial dependence.

Here interest focusses on $Y_{i(r+1)} = \max_{j=k}^{k+n-1} X_{ij}$ for $k = 1 + rn,\ r = 0, \ldots,,\ i = 1, \ldots, I$, where X_{ij}, denotes site i's response at time j, $i = 1, \ldots, I,\ j = 1, \ldots,$. Of particular importance are $Cov(Y_{i(r+1)}, Y_{i'(r+1)})$ for $i \neq i'$ as r grows large, an admittedly simplistic description of dependence in this context but one that is easy to understand. An example in the next subsection shows covariance generally decreases as r increases.

Vancouver's PM_{10} field.

The data are measured hourly ambient $\log PM_{10}$ concentrations recorded at ten monitoring sites during the 240 week period beginning Jul 20, 1994. However only the nine with no missing data are used, these being sufficient for our purposes. Figure 1 shows the site locations in the Greater Vancouver Regional District (GVRD). In this example temporal effects were removed as in Sun *et al.* (2000) so that spatial structure alone could be expressed through the maxima computed for the series over varying time spans. (Another example in Le and Zidek (2005) demonstrates the phenomenon seen below when those effects are left in.)

To remove the temporal effects a single model was fitted to all nine stations in the analysis and computing the "detrended residuals" at site i and time j as

$$E_{ij} = X_{ij} - T_j,$$

where

$$T_j = \mu + H_j + D_j + L_j + S_j + M_j.$$

In that model μ denotes the overall mean found by averaging across all sites and hours. The "hour effect", $H_j,\ j = 1, \ldots$ cycles through 24 values, the averages for each hour of the log concentration over days and sites in the study. The "day-of-the-week effect", D, linear time trend, L, and the seasonality effect, S, were found in a similar way. (Details are omitted for brevity.) The meteorological effect, M, obtains from regressing on various meteorological variables, the residuals obtained after subtracting from the response, the other terms in the model (Sun *et al.* 2000). Next the auto-correlation was removed (again with a single model over all sites) to validate the assumptions that justify the predictive distribution. for successive time intervals ("spans") of varying length of these detrended, whitened residuals.

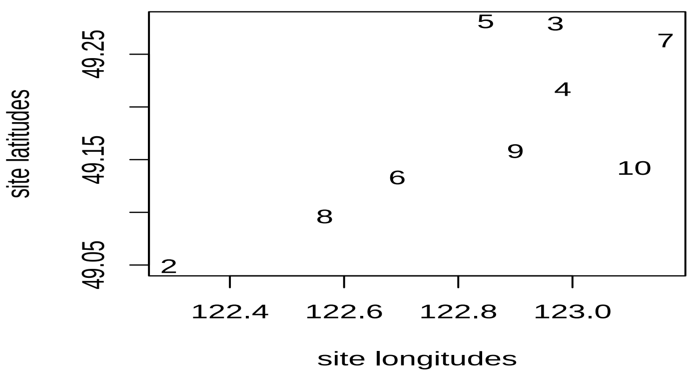

Figure 1: Locations of nine PM_{10} monitoring sites in the Greater Vancouver Regional District.

The intersite correlations are shown in Figure 2. Notice how they decline for most site pairs but persist for a few.

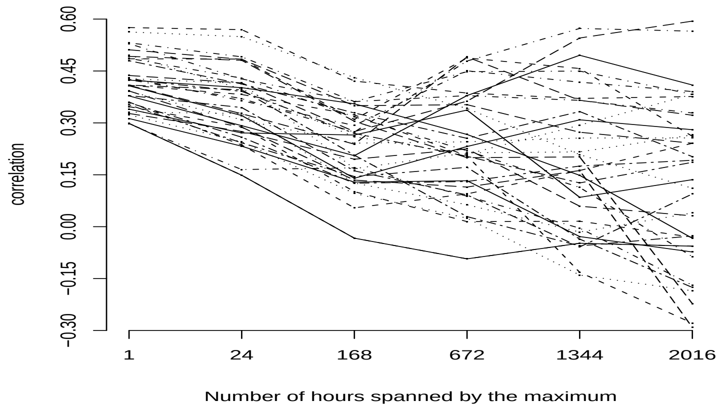

Figure 2: Intersite correlations for the maxima over time spans of between 1 and 2016 hours of $\log PM_{10}$ concentration residuals. These were obtained from nine monitoring sites in the Greater Vancouver Regional District between 1994 and 1999.

To examine this phenomenon more closely we list those correlations in Table 1 for the maxima at the nine sites computed over a very long time span of 84 days (2016 hours). The Table presents them in increasing order from left to right and down its rows. A glance at the table shows site pair, (3,6) to be the winner. That pair's correlation actually seems to increase as the time span increases (although this increase could be spurious, a product of sampling error). However, (8,9) is a close competitor. Figure 1 shows the sites in both pairs to be in relatively close proximity. Contrast the large correlations for these pairs to that of (6,9), a pair of sites with even smaller intersite distance yet much smaller intersite correlation. Likewise for (2,8) - they have a negative intersite correlation.

In fact site #2 seems to be uncorrelated with most of the other sites, 5,8,6,3,9, and 10. That does not seem surprising. Site #2 is situated near a major roadway with fairly heavy traffic volumes during rush hours that can generate relatively [large particulate concentrations. Mysteriously, site #2 does have a positive, albeit small, association with site #4.]

Roughly speaking we see evidence of asymptotic dependence in the following pairs: (8,9); (8,9). Weaker evidence of such dependence is found for: (4,9); (3,5);(7,10);(5,6); (3,8). However we know of no substantive evidence that would support for those findings.

(2,7)	(7,9)	(4,6)	(2,5)	(2,8)	(2,6)	(4,5)	(2,3)
-0.29	-0.28	-0.22	-0.19	-0.18	-0.17	-0.09	-0.07
(2,9)	(2,10)	(5,9)	(5,10)	(9,10)	(8,10)	(3,10)	(6,10)
-0.06	-0.06	-0.03	-0.03	-0.03	0.03	0.04	0.09
(6,10)	(3,4)	(7,8)	(3,7)	(5,8)	(4,7)	(2,4)	(6,9)
0.09	0.11	0.14	0.19	0.19	0.20	0.24	0.24
(6,9)	(6,7)	(3,9)	(4,10)	(4,8)	(5,7)	(6,8)	(4,9)
0.24	0.26	0.26	0.27	0.28	0.32	0.33	0.38
(3,5)	(7,10)	(5,6)	(3,8)	(8,9)	(3,6)		
0.38	0.38	0.39	0.41	0.57	0.59		

Table 1: In ascending order by site pair, intersite correlations of successive maxima over 2016 hours (84 days) for detrended, whitened log PM_{10} concentrations. In all, nine Vancouver sites were studied during the 1994-1999 period.

Although conclusions, being based on an analysis for just a single example with a simplistic measure of dependence, must be viewed as tentative, they do show a need for models such as that of Coles and Pauli (2002) and the one in Section 3, which allow flexibility in specifying the intersite dependence.

This problem of decreasing intersite field dependence has important implications for prediction and design. In particular it seems quite possible that some urban areas are undermonitored. A fairly dense grid of monitoring stations may well be needed to ensure extreme values over the region are reliably detected.

Simulation studies.

The results of a simulation study reported in this section shed further light on the issue raised in the previous subsection. The work of Fu (2002) as well as Fu et al (2003) suggested use in this study of a matric - t sampling distribution to characterize spatial fields. Responses were assumed to come from ten aligned monitoring sites $i = 1, \ldots, 10$ with mean 0 and variance 1. $N = 5000$ replicates of n responses yielded the maximum in each case. Of specific interest were the intersite correlations between the extreme values for the aligned site relative to the first site in line, # 1, for varying n and degrees of freedom.

The algorithm of Kennedy and Gentle (1980, page 231 - 232) generated random covariance matrices, Σ, from an inverted Wishart distribution, *i.e.* $\Sigma \sim IW(\Psi, m)$ with varying degrees of freedom, m, and isotropic (hyper-) covariance kernel, Ψ, $\Psi_{ij} = \exp\{-\alpha|i-j|\}$, $i,\ j = 1, \ldots, 10$. Then n random vectors $X_k = (X_{k1}, \ldots, X_{k,10}) \sim N_{10}(0, \Sigma),\ k = 1, \ldots, n$ were sampled. From these $Y_i = (Y_{i1}, \ldots, Y_{i,10})$ were found with $Y_{ij} = \max_{k=1}^{n} X_{kj}$, $j = 1, \ldots, 10$. The process was repeated N times.

Generally increasing the number of degrees of freedom of the inverted Wishart (thereby lightening the tails of the resulting t distribution) made the intersite correlation decline. Increasing n produced a similar result. Figure 3 shows those results for $\alpha = 0.25$.

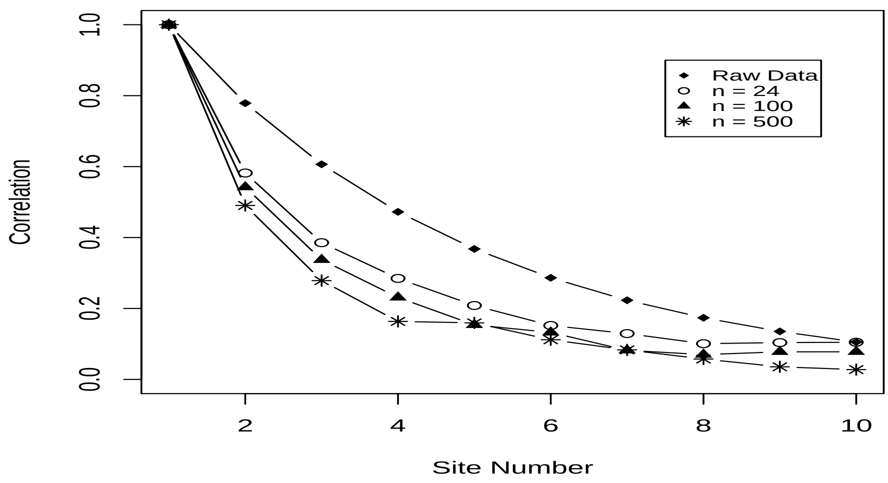

Figure 3: Inter - site correlations between site i=1 and the other sites for varying n, $\alpha = 0.25$, and 20 degrees of freedom

Figure 4 shows analogous results when the number of degrees of freedom becomes infinite yielding a 10 dimensional multi - normal distribution.

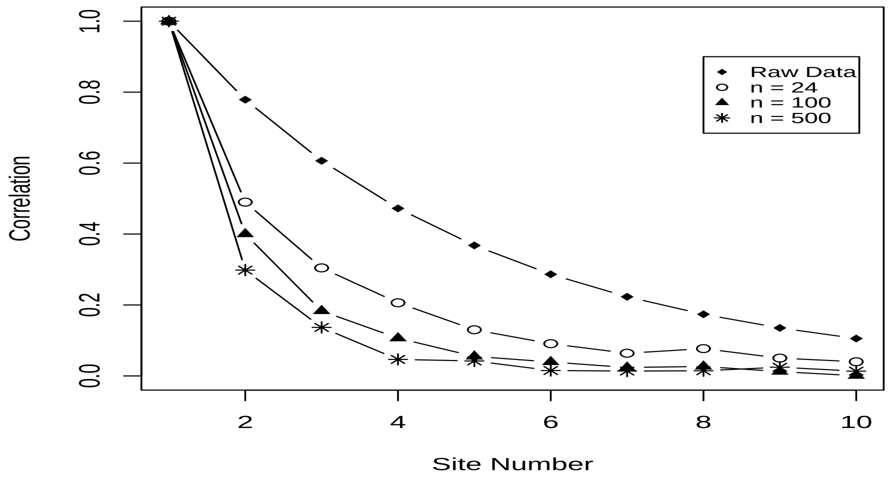

Figure 4: Inter - site correlations between Site # 1 and the other sites for varying n, $\alpha = 0.25$, and a multi - normal sampling distribution

As n increases the intersite correlation declines faster for a large number of degrees of freedom than a small number. In fact an empirical "power law" describes well the relationship between the intersite correlations for extremes versus that for the original response field:

$$Cor_{ext} = \beta(Cor_{raw})^{\gamma} + \delta,\ \gamma > 0,$$

where Cor_{ext} denotes the intersite correlation for extremes and Cor_{raw} that for the original response field. Fitting the power law equation for varying n and degrees of freedom to the simulated data led to the results in Table 2. Since the fitted coefficients are subject to sampling error, those results are regarded as descriptive and conclusions suggested by them, tentative.

Subject to that caveat note that large values of γ (the "power") in the power law represent rapidly declining intersite correlations. Thus extremes for any pair of sites have a weaker linear association than when that power is small. In fact when $\gamma = 1$ no loss of correlation accrues going from the original responses to the extremes field.

That suggests smaller powers are associated with heavier tails (that is, have a smaller number of degrees of freedom). At the same time empirical studies show the log matric - t air pollution predictive distributions fit observed fields well (Sun 1998; Sun et al 1998). Thus the simulation study suggests, subject to confirmation, that unmeasured extremes are better predicted in practice than if, for example, the field were log - Gaussian as is commonly assumed. If correct that finding would constitute good news for network designers.

These findings also suggest the power will decline with n (the span of the maxima). If rigorously confirmed this finding would have implications for developing realistic compromise air quality criteria and designing the associated monitoring networks.

DF	n	Power Law
20	24	$Cor_{ext} = 0.89Cor_{raw}^{2.28} + 0.10$
	100	$Cor_{ext} = 0.94Cor_{raw}^{2.83} + 0.05$
	500	$Cor_{ext} = 0.93Cor_{raw}^{2.98} + 0.06$
30	24	$Cor_{ext} = 0.94Cor_{raw}^{2.49} + 0.05$
	100	$Cor_{ext} = 0.94Cor_{raw}^{3.05} + 0.06$
	500	$Cor_{ext} = 0.96Cor_{raw}^{3.38} + 0.03$
40	24	$Cor_{ext} = 0.94Cor_{raw}^{3.04} + 0.05$
	100	$Cor_{ext} = 0.96Cor_{raw}^{3.06} + 0.01$
	500	$Cor_{ext} = 0.97Cor_{raw}^{3.85} + 0.02$
∞	24	$Cor_{ext} = 0.93Cor_{raw}^{2.80} + 0.07$
	100	$Cor_{ext} = 0.97Cor_{raw}^{3.51} + 0.02$
	500	$Cor_{ext} = 0.98Cor_{raw}^{4.40} + 0.01$

Table 2: Empirical power laws relating intersite correlations for extremes (denoted Cor_{ext}) against original responses (Cor_{raw}) for varying numbers of degrees of freedom (DF).

Finally note that the range of powers in the power law is larger when the degrees of freedom is large relative to the range when it is small.

2.2 Limitations of conventional approaches.

Model-based approaches to spatial design focus either on (1) predicting the unmeasured responses or (2) estimating parameters of a regression function. (Le *et al.* and Zidek 2003; Le and Zidek 2005). However none seem to be concerned with the prediction of the extreme values generated by a random field.

This section studies one (information-based) approach in Category (1) because it at least aspires to bypass the need to specify a monitoring objective. Surely it, if any, would produce networks that would satisfactorily monitor extremes. This section shows mixed results in an application.

The approach studied here stems from the challenge offered by the multiplicity of purposes, foreseen and unforeseen, for a environmental monitoring network. The latter come from new knowledge and resulting societal concerns (Caselton *et al.* 1992; Zidek *et al.* 2000). The information based approach sidesteps these difficulties (Caselton and Husain 1980; Caselton and Zidek 1984). It "gauges" sites that maximize the information for "ungauged" sites. Equivalently for the sites to be gauged, it selects those that maximize the entropy of the (joint) response distribution, a popular approach (Shewry and Wynn 1987; Wu and Zidek 1992; Le and Zidek 1994; Guttorp et al 1993; Bueso et al 1998, 1999a, 1999b; Sebastiani and Wynn 2000; Angulo et al 2001) and goes back at least as far as Lindley (1956). Müeller (2001) provides a recent review.

Vancouver example.

Now take extremes to be the responses of concern. How well does this "purpose-free" approach to design really work? The GVRD's hourly PM_{10} ($\mu g\ m^{-3}$) concentration field introduced in Section 2 suggests an answer that question. However, here we can consider all ten rather than nine monitoring sites by taking advantage of theory for monotone (staircase) data patterns (Le et al 2001; Kibria et al 2002). Figure 5 shows their spatial configuration. The data from these monitoring sites were processed as in the previous section to yield approximately white residuals, the subject of analysis below.

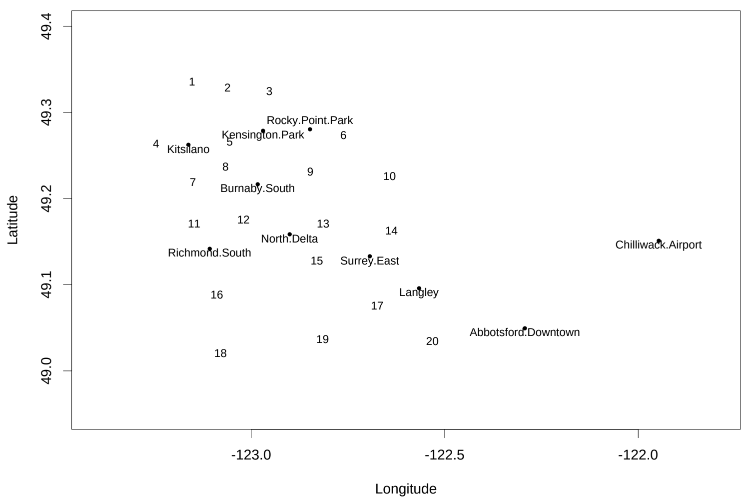

Figure 5: Configuration of hourly PM_{10} ($\mu g\ m^{-3}$) concentration monitoring network stations in the Greater Vancouver Regional District. Numbers label the potential sites for gauging.

As a hypothetical exercise six more sites are to be gauged among the twenty depicted in Figure 5, the centroids of census tracts with sizable populations. The analysis needs the hypercovariance matrix (hereafter called the "covariance" for simplicity with little risk of confusion) of the predictive distribution of the $\log PM_{10}$ ($\mu g\ m^{-3}$) residuals describe above. To reduce the dimensions of the parameter space to manageable levels, assume the covariance factors as $\Lambda \otimes \Omega$, $\Omega : 24 \times 24$, representing within - site correlations between hours and $\Lambda : 30 \times 30$, the spatial covariance between the $10 + 20 = 30$ sites where

$$\Lambda = \begin{pmatrix} \Lambda_{uu} & \Lambda_{ug} \\ \Lambda_{gu} & \Lambda_{gg} \end{pmatrix}.$$

Here "u" means "ungauged", (sites with no monitors) and "g", currently "gauged". Thus Λ_{uu} means the covariance matrix of the twenty site available for gauging, six to be chosen. Let $\Lambda_{u \cdot g} = \Lambda_{uu} - \Lambda_{ug}\Lambda_{gg}^{-1}\Lambda_{gu}$ be the conditional spatial covariance of the ungauged sites given observations at the gauged sites. All elements of the covariance can be estimated (Kibria *et al.* 2002). Finally ignoring irrelevant terms and factors, the entropy of any proposed $6 + 10 = 16$ station network (including the original ten) equals the log of the determinant of $\hat{\Lambda}_{a \cdot g}$, the 16×16 submatrix of the estimated $\Lambda_{u \cdot g}$ corresponding to that proposed network. Here "a" denotes the "added" stations. So the six new sites, "a", must be chosen to maximize this determinant and thereby find the maximum entropy optimal design.

Figure 6 shows all the sites with their (conditional) variance ranks in the predictive distributions. The existing sites are dots while the potential sites are numbered. The potential site with the largest conditional predictive covariance of its log transformed predictive concentrations is potential site # 19. It lies well outside the cluster of existing sites so its large conditional variance is not surprising. Therefore it seems a likely candidate for membership in selected the subset of six new stations.

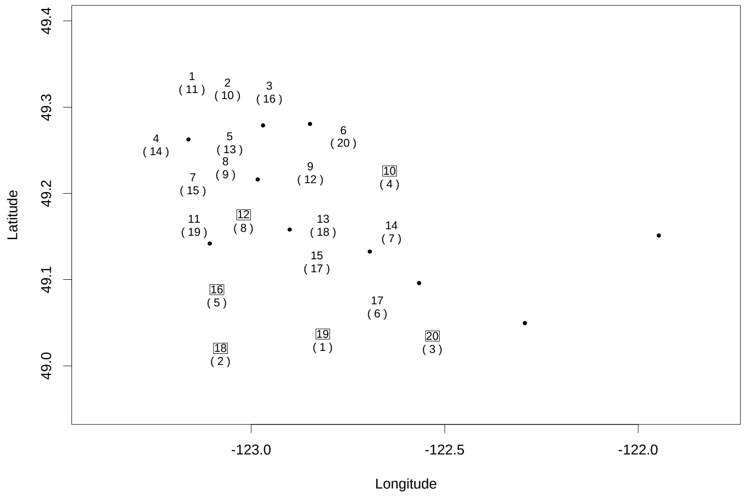

Figure 6: The numbered locations represent sites that might potentially be gauged to create six new monitoring stations. The numbers appearing in brackets indicate their ranks according to the size of their conditional predictive variances in the estimated covariance, $\Lambda_{u \cdot g}$.

In fact Le *et al.* (2003) show # 19 must be selected. Four of the other top six variance - ranked sites also make it: #'s 10, 16, 18, 20. However, the remaining selected site, # 12, turns out to be a surprise. It is selected ahead of the 6^{th} and 7^{th} ranked sites, # 14 and 17, presumably since responses at the latter will be predictable from sites # 19 and 20 once these are added.

Monitoring for compliance on Feb 28, 1999.

How well would this new network monitor extremes? The answer depends on which criterion we use to assess performance. Let us adopt one consistent with the most common purpose of urban air quality monitoring networks, the need to enforce compliance with national or local standards. More specifically suppose the six new "add" sites must be selected as those most likely to be in "non-compliance". The remainder, "rem", would be more likely than those in the "add" subgroup to be in compliance.

As an illustration for PM_{10} adopt the criterion

$$add = \arg\max_{add'} Prob[\max_{i \in add', j \in \{hours\}} X_{ij} > 50 \ (ppb)], \tag{1}$$

where "Prob" means computed using the joint conditional predictive probability for unmonitored responses given those at the gauged sites. The complexity of our metric requires that we exploit fully our predictive distribution and simulate the field of twenty unmeasured values every hour over the entire day and then computing the maximum over all 24 hours and six sites in "add". Thus the required probability can be found as the fraction of the outcomes leading to "noncompliance" as defined above. Indeed this approach can handle even more complicated "metrics" such as the "largest 8 hour moving average over the day".

However our conditional predictive distribution also requires the specification of a day. Unlike the criterion based on entropy, which depends only on the covariance, it depends on the conditional mean. That, in turn, depends on the values at the gauged sites which change from day - to - day, making the design day - dependent. How robust is the optimum design over days?

To find out we first picked February 28, 1999 somewhat arbitrarily when Vancouver's particulate air pollution levels can be high. Table 3 shows in ranked order the top ten choices of the six "add" sites based on the entropy criterion. For each choice we also give the probability defined above that they will be in non-compliance that day, along with their order among all the 38, 760 possible subsets of six sites we could have chosen.

MaxEnt Order	Selected Sites						100*Prob	Order
1	10	12	16	18	19	20	45.9	64
2	10	14	16	18	19	20	46.1	55
3	10	12	14	18	19	20	44.6	145
4	8	10	16	18	19	20	45.0	123
5	2	10	16	18	19	20	45.1	109
6	2	10	12	18	19	20	43.6	222
7	8	10	14	18	19	20	43.7	212
6	2	10	14	18	19	20	43.9	200
9	10	16	17	18	19	20	49.5	1
10	1	10	16	18	19	20	45.1	109

Table 3: The top ten choices among 38, 760 available subsets of six new sites to be added to Vancouver's existing network of ten sites gauged to measure PM_{10}. Here "Prob" means the probability of noncompliance while "Order" means order with respect with respect to the noncompliance probability criterion.

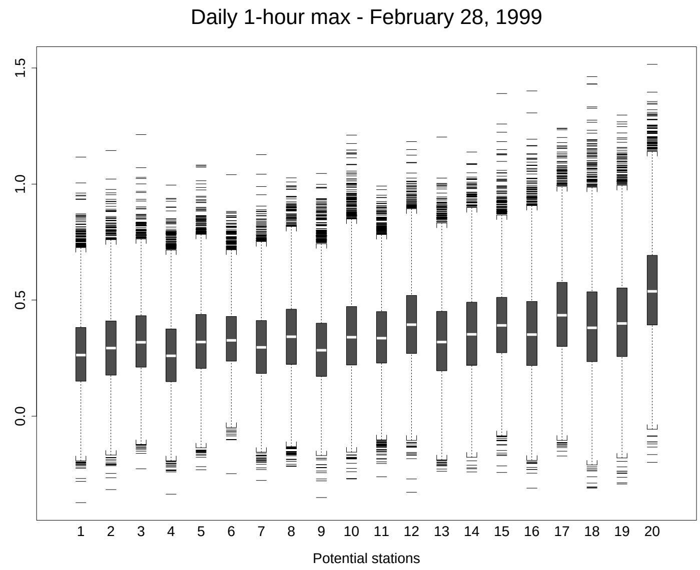

Figure 7: The site-wise distribution of simulated daily maxima $\log PM_{10}(\mu g\ m^{-3})$ responses for February 28, 1999.

Figure 7 suggests the entropy criterion does reasonably well for extremes, at least for the selected day. Its 9^{th} ranked candidate turns out to be the compliance ranked best. Moreover the chosen sites have either a large response level or dispersion. Finally the noncompliance probabilities for the top ten entropy choices are similar. (Although small, these differences could be important in a regulatory environment where financial and other penalties for noncompliance can be large.) Sites # 10, 18, 19 and 20 are always in the top ten because of their large conditional response variances as manifest in the boxplots of Figure 7. Intersite correlations are essentially ignored, no surprise in the light of the results of the previous section showing that for extreme values they tend to be small. The 9^{th} ranking choice proves best by the noncompliance criterion, that choice substituting site #17 for #site 12 since the former's daily PM_{10} maxima tend to be larger while their variances are similar and large (see Figure 7).

MaxEnt design for Aug1, 1998.

However the entropy criterion does not do so well on other days such as August 1, 1998. Figure 8, like Figure 7, indicates the distribution of daily $\log PM_{10}$ distributions for potential new sites.

Notice that the level of $\log PM_{10}$ at site #19 is much lower than it was above and hence it is not a contender by the noncompliance criterion. Sites #16 and # 18 have also dropped out. Site #7 looks likely to be in noncompliance that day.

We see that the optimum non-compliance design is not stable; it can vary from day - to -day. This last result implies that any fixed design, no matter how chosen, would be less than optimal on lots of days.

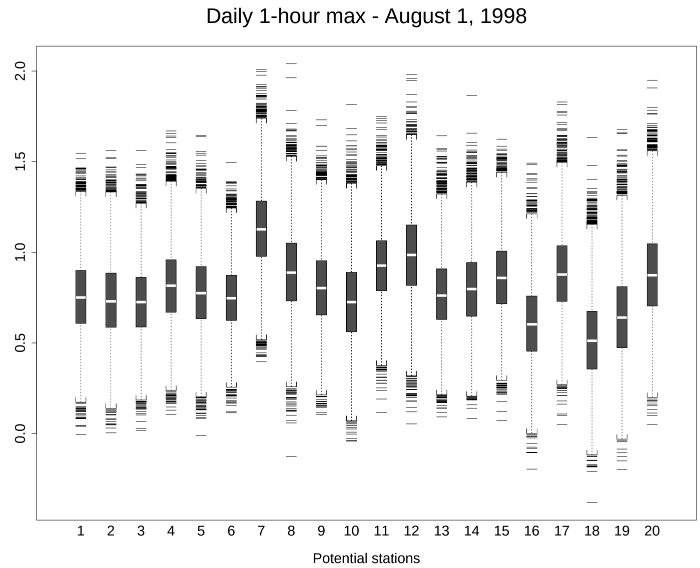

Figure 8: Site-wise distribution of simulated daily maxima $\log PM_{10}(\mu g\ m^{-3})$ values for August 1, 1998.

Multiday criterion.

Because of its sensitivity to change, the above method for finding that fixed design seems impractical. A preferable alternative, based on multi-criteria optimization considerations, may be to use a weighted average over days, of the criterion "Prob"s in Equation 1.

The problem of selecting the averaging weights now presents itself. Equal weights have some appeal since the objective function could then be interpreted as the expected probability for a randomly selected day. But that would weight equally, days when noncompliance was unlikely with those when it was likely. Why should the former be accorded any role in picking the winning design? Thus implementation of this approach faces some serious hurdles.

Simulated monitoring criterion.

This subsection presents one resolution of issues raised above. (A second follows in the next subsection.) It avoids reliance on the measured values at gauged sites however days are selected. Instead those values are sampled from their predictive distribution thereby representing both past responses at gauged sites and future ones as well.

Rather than simply repeat our earlier analysis in this new setting we illustrate the flexibility of the approach in this demonstration by choosing a different criterion metric. That metric, the 99^{th} percentile of 365 average daily PM_{10} concentrations in a simulated year, comes close to one actually used in air quality criteria.

Figure 9 shows the results obtained as follows: (1) for day 1 of 1998 generate a random vector from the marginal predictive distribution for the gauged sites on that day; (2) conditional on that vector generate another from the conditional predictive distribution for ungauged sites; (3) compute the daily average $\log PM_{10}(\mu g\ m^{-3})$ concentration for that day at every site; (4) repeat steps 1 - 3 for each day of 1998; (5) finally at each site find the 99th percentile of daily averages and take its antilogarithm.

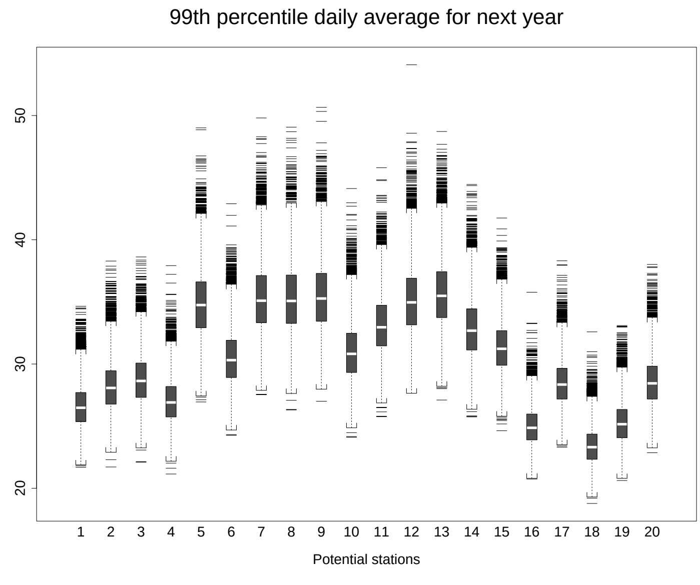

Figure 9: Site-wise distribution of simulated annual 99th percentiles of daily average $\log PM_{10}(\mu g\ m^{-3})$ concentrations.

If a critical level of $50(ppb)$ were assumed for this metric the figure suggests stations #5, 7, 8, 9, 12, and 13 would be the most likely to noncomply, quite a change from our previous optima. The figure also reveals major differences amongst the potential sites with respect to the distribution this metric. We turn now to an alternative resolution resolution of the issues.

Noncompliance event detection criteria.

Approaches taken above to formulating design criteria rely on the values observed (real or simulated) at the gauged sites. Analyses are conditional on those values. However regulatory action relies not on them but rather on the events of compliance or noncompliance. This fact suggests other ways of building noncompliance criteria for network design. Moreover the predictive distribution can be used to estimate conditional and marginal probabilities of those events, they being R, A and G, "compliance of the remaining sites", "compliance of the 'add' sites", and "compliance of the gauged sites", respectively.

To describe the process suppose a particular set of say six candidate sites among twenty have been selected. Then for any time series of simulated responses at gauged and ungauged sites, the occurrence - nonoccurrence sequence of each event can be determined. After a sufficiently large number of replicates have been simulated the probabilities for these events can be estimated as their relative frequency of occurrence. We illustrate these calculations below albeit with a simple model that makes extensive simulation feasible.

Even with this new approach a number of possible design objective functions present themselves, namely select the six add sites:

- to maximize $P(\bar{A}|G)$;
- to maximize $P(\bar{A}|\bar{G})$ where $\bar{A}$ means A is not in compliance.
- to maximize $P(\bar{A})$. The identity, $P(\bar{A}) = P(\bar{A}|G)P(G) + P(\bar{A}|\bar{G})P(\bar{G})$, makes clear that the first two are related to, but different from the third. In fact the third could be interpreted as a multi-criteria optimization problem that combines the objective functions from the first two.
- as the complement of the fourteen sites that minimize $P(R|G)$, $P(R|\bar{G})$, and $P(R) = P(R|G)P(G) + P(\bar{R}|\bar{G})P(\bar{G})$, respectively.

Original		Threshold: log 2		Threshold: log 4		Threshold: log 7	
Site Rank	100*$\hat{Var}$	Site Rank	100*$\hat{Var}$	Site Rank	100*$\hat{Var}$	Site Rank	100*$\hat{Var}$
1	52	1	44.6	1	47.9	1	49.8
4	55	4	45.8	4	49.9	4	52.3
2	57	3	47.0	2	51.8	2	54.0
3	59	2	47.7	3	52.0	3	55.7
16	61	6	47.9	6	55.2	16	58.6
18	63	15	49.9	15	56.4	6	59.9
6	65	5	50.4	16	56.7	15	60.7
17	65	11	52.6	5	57.4	17	61.2
15	66	16	53.3	17	58.1	18	61.4
19	66	13	53.5	18	60.5	5	62.3
5	68	17	53.7	11	60.6	19	63.7
7	70	7	54.2	7	61.0	11	65.6
11	70	8	55.2	13	61.1	7	65.9
14	71	14	56.0	14	61.8	14	66.1
8	72	12	56.9	8	61.9	8	66.5
10	73	9	57.0	19	62.6	13	68.4
20	74	18	59.1	12	64.1	10	68.8
12	75	10	60.4	9	64.6	12	69.3
13	75	19	60.7	10	64.8	9	70.7
9	77	20	64.5	20	67.7	20	70.8

Table 4: The ranked order of the twenty ungauged sites involved in the simulation study. The first two columns show the sites ranked by their variances with realistic values from a recent study involving two of the authors. The remaining three pairs of columns shows these same sites ranked by the variances their response distribution conditional on the ten gauged sites being in compliance with thresholds $\log(2)$, $\log(4)$ and $\log(7)$, respectively.

- to maximize $P(\bar{A}|G, R)$.

Numerous other options must exist.

What principles could be invoked to compel a choice of a single criterion for designing a network that best meets the regulators goals of detecting noncompliance? That remains an unanswered question. Yet the different choices available would lead to different fixed designs.

To see this we rely on another simulation study like that in Section 2.1. Here we take a total of thirty sites to correspond to those in Figure 5, sites #1-20 being the currently ungauged sites, #21-30, the gauged ones (the black dots in Figure 5).

We assume the joint 5 dimensional response matrix has a matric -t distribution with 35 degrees of freedom and generate these responses as in the earlier simulation study above. All marginal site response distributions are taken to have mean zero (after re-centering). To make the simulation study realistic we take the actual intersite covariances to be those estimated by Le et al (2001). From this distribution 24 joint response vectors were generated and the maximum at each site calculated as in Subsection 2.1 to represent its daily maximum.

Table 3 shows the twenty ungauged sites ranked by their variance, originally for their unconditional joint distribution, then conditionally on the ten gauged sites being in compliance with respective to various hypothetical thresholds. The variance order tends to change from that of the real data. When we choose a small threshold, $\log(2)$, the conditioning event G that the gauged sites are in compliance imposes a substantial constraint on the simulated field, thereby diminishing its dispersion and lowering its variances. More to the point, the complex interaction of this constraint with the pattern of intersite correlations means that the ranking of the site variances can be changed by the conditioning event.

We define "compliance" for any one response at any site to mean having a daily maximum below various thresholds indicated in Table 3. Compliance over a subset of sites would mean all daily maxima below that threshold for all sites.

We study the problem of selecting six new sites from among the twenty ungauged sites. Table 5 presents the results for two of the many compliance related criteria discussed above. These are A: add six new sites with a maximal value of $P(A \mid G)$ and B: find the fourteen sites that minimize $P(R \mid G)$ and select the remaining siz for inclusion in the network.

As expected, the results show that the optimal design does depend on the criterion thereby forcing a designer to select from among the myriad of seemingly plausible compliance criteria. However, in this case at least, only small differences are seen in the resulting criterion probabilities among the top 5 designs, thus offering some hope that at least at the practical level, the eventual choice of a winner among the leading contenders will not be critical. However, more analysis will be required to determine if this hope is realistic.

In conclusion, this section has presented a variety of seemingly plausible choices for spatial design objectives without finding a clear - cut choice among them. Nor are we aware of any accepted principles to help in that selection among the plethora of reasonable candidates available. That suggests revisiting the option of an information - based (entropy) approach to sidestep this issue. The results of the previous subsection that criterion cannot be applied directly to the response field, so a new approach is needed, one like that described in the next section.

Threshold	Sample Size	Criterion	Sites Selected						Prob*100	SD*10^4
log (2)	497,561	A	1	10	16	18	19	20	54.00	7.1
			4	10	16	18	19	20	54.16	
			2	10	16	18	19	20	54.18	
			2	4	10	18	19	20	54.52	
			1	4	10	18	19	20	54.56	
		R	4	16	17	18	19	20	53.13	7.1
			10	16	17	18	19	20	52.96	
			4	10	16	18	19	20	52.73	
			4	11	16	18	19	20	52.33	
			1	4	16	18	19	20	52.33	
log (4)	1,214,797	A	10	11	12	18	19	20	82.66	3.4
			8	10	11	18	19	20	82.74	
			10	12	16	18	19	20	82.80	
			10	12	14	18	19	20	82.81	
			10	11	16	18	19	20	82.81	
		R	10	16	17	18	19	20	77.26	3.8
			11	16	17	18	19	20	77.11	
			9	10	16	18	19	20	77.07	
			10	11	16	18	19	20	77.07	
			10	14	16	18	19	20	77.04	
log (7)	1,009,983	A	9	10	12	18	19	20	95.71	1.5
			9	10	11	12	19	20	95.75	
			10	11	12	18	19	20	95.75	
			10	12	14	18	19	20	95.75	
			7	10	12	18	19	20	95.77	
		B	9	10	14	18	19	20	94.17	2.3
			9	10	12	18	19	20	94.13	
			9	10	16	18	19	20	94.12	
			9	10	17	18	19	20	94.11	
			9	10	13	18	19	20	94.11	

Table 5: This table gives the five best choices of six new stations to gauge out of twenty possible for various criteria and various compliance thresholds. The criterion, A, means select the six sites with a maximal value of $P(A \mid G)$ while R means select the fourteen sites with a minimal value of $P(R \mid G)$, then gauge the six sites that remain. The probability (of compliance or noncompliance in the case of A and R respectively) is that for the selected subsets. The estimated SDs show that the simulation sample size is large enough to ensure that those probabilities are accurate to at least their third decimal place, thereby ensuring the ranking of the subsets is valid.

.

3 An entropy based approach.

Various approaches could be taken, the most obvious being the use of a multivariate extreme value distribution. However, as noted in the Introduction, this approach would not be practical. Another would use a predictive distribution for the response field to estimate the required conditional probability density functions by a Monte Carlo approach. That of Kibria et al (2002), as developed within a design framework by Le et al (2003), offers one such possibility. However, that approach also fails, due to the “curse of dimensionality”. This section considers a third approach suggested by Fu et al (2003).

Although neither of these last cited works were directly concerned with the design problem, they demonstrate the suitability of the log matric -t distribution as an approximation to the joint distribution of an extremes field. Although that distribution proves to be more complex than the log - Gaussian distribution, it has two of the latter’s important advantages, namely it has conditional distributions in the same family and it has an easily computable entropy. The latter proves valuable since the spatial design’s combinatorial optimization problem known to be “N_p Hard”.

However, the simulated data considered by Fu et al (2003) was generated by the (CGCM1) global climate (computer) model, leaving doubt about the applicability of this approach to real data. Thus, this section focuses on model assessment and the demonstration of its applicability to designing networks for monitoring the extreme values of environmental space-time processes.

3.1 The log matric t approximation.

We return to the data introduced in Subsection 2.1. Since the unconditional log matric t process distribution generated by the hierarchical Bayesian approach can have arbitrarily heavy tails, Figure 10 suggests that the (site-wise) conditional sampling distributions of the weekly maxima may be taken as Gaussian. In estimating hyperparameters of the prior distribution, we adopt an isotropic spatial correlation structure and fit an exponential semivariogram with the result:

$$\gamma(h) = \begin{cases} 0.2 + 0.1(1 - \exp(-h/0.2)), & \text{if } h > 0 \\ 0, & \text{if } h = 0. \end{cases}$$

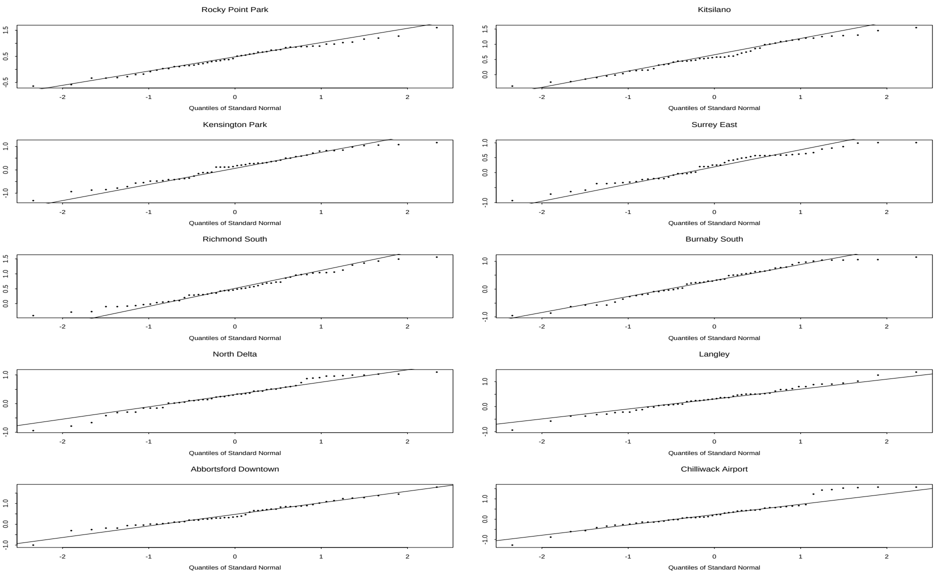

Figure 10: QQ plots of the residuals of 1996 weekly maxima of PM_{10} hourly concentration data for 10 sites. Logarithms of these maxima were taken and temporal components along with autocorrelation were removed as in Section 2.

[On a technical point we estimated $\hat{F}^{-1} = 0$, m = 34, and c = 13, in the notation of Fu et al (2003).]

To assess the multivariate process distribution generated by our approach we used two - deep cross validation, two out of ten sites being removed repeatedly at random to play the role of the "ungauged sites". Their values were then predicted and we determined how often the 95% and other credibility intervals obtained from the model contained the observed values at the two omitted sites. Table 6 shows the coverage probabilities are close to the corresponding credibility levels. Figure 11 shows the distribution of coverage fractions over the weeks of 1996 after running our cross validation experiment week - by - week. We conclude that the log matric - t distribution gives a reasonable approximation to the extremes examined here.

Credibility Level	Mean	Median
95%	0.93	0.94
80%	0.8	0.81
50%	0.53	0.54

Table 6: Summary of coverage probabilities at different credibility levels for weekly maxima obtained after transforming 1996 hourly PM_{10} concentration data.

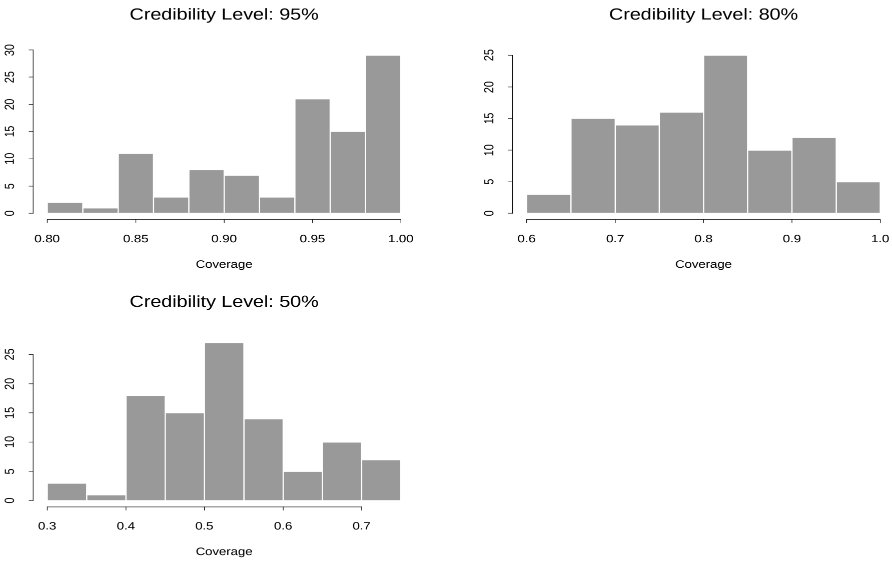

Figure 11: Histograms of coverage probabilities at different credibility levels for 1996 weekly maxima of hourly PM_{10} concentration data after transformation and removing temporal components.

In unpublished work, a second example was considered. That analysis involved London's 1997 hourly $PM_(10)$ data as described by Zidek et al (2003) in a different context. The log Gaussian seemed promising as an approximate sampling distribution model for the logarithm of weekly maximum over 52 weeks of data.

3.2 Redesigning the GVRD network.

This section extends the network considered in Subsection 2.2, relying on the justification provided above to use the log matric t distribution approximation. The data for maxima over successive 144 hour periods, *i.e.* over weeks, were pre-processed in the manner described in that section. The predictive distribution for the residuals and its entropy were found. Finally the analysis of Subsection 2.2 was repeated for that distribution and the best six of twenty potential sites found.

The optimal six are depicted in Figure 12

Not surprisingly the optimal six are now those having the highest predictive variance, # 10,16,17,18,19, and 20. The weakened spatial correlation no longer has a role to play in this case. That leaves open the question of how many sites should now be included, a subject under current investigation.

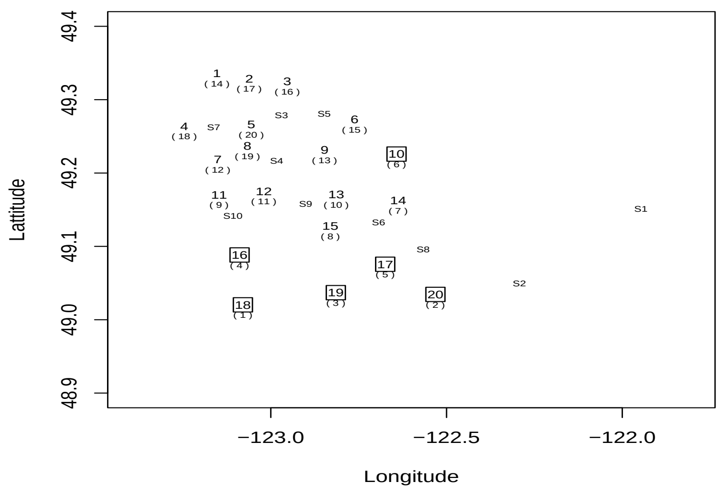

Figure 12: The optimal subset of six among twenty potential sites are the ones with squares around their numbers. The bracketed numbers are the predictive distribution's variance ranking of the sites. The numbers with an "S" in front are the current monitoring sites.

4 Conclusions.

Environmental monitoring criteria have been based on extremes because of their perceived association with risk. Consequently networks have been established to monitor such things as air pollution and acidic deposition and their extremes. Surprisingly no attempt seems to have been made to define explicit technical design objective functions that express societal concern about the risk associated with extremes. These functions would accommodate and embrace, knowledge, concern, risk perceptions, and political demands.

As a consequence, the specification of such criteria has been left to the designers. This paper points to some of the technical challenges they face. Amongst other things we have delineated a large number of seemingly credible alternatives, all of which are consistent with the goal of detecting non - compliance and thereby enforcing standards.

This complacency may derive from a belief that the layout of a network is not critical. That would certainly be true, for example, of designs for monitoring London's PM_{10} field since it is quite flat (Zidek et al 2003). Even a single station would characterize that field quite well no matter where it was placed.

However such complacency may not be warranted where extremes are concerned. An important finding in this paper is the diminished intersite correlations among extremes compared to the response fields from which they derive. The finding leads us to wonder how well current monitoring programs work especially in guarding the susceptible and sensitive such as the old and young in urban areas, if indeed extremes are important determinants of risk. In fact the apparent near independence of extremes between sites in certain areas would suggest the need for a dense network of monitors to adequately protect their associated population. In fact the high cost of setting up and maintaining monitors has severely restricted their numbers.

This paper also offers a practical way out of the difficulty presented by the multiplicity of objectives confronting the designer. That approach is based on an entropy criterion and uses a joint matric t distribution as an approximation to the joint distribution of spatial extremes. However the support for our approach is both limited and empirical. More validation and testing would clearly be desirable.

References

[1] Angulo, JM, and Bueso, MC (2001). Random perturbation methods applied to multivariate spatial sampling design. *Environmetrics*, 12, 631 - 646.

[2] Bueso, MC, Angulo, JM, and Alonso, FJ (1998). A state-space model approach to optimum spatial sampling design bad on entropy. *Environmental and Ecological Statist*, 5, 29-44.

[3] Bueso, MC, Angulo, JM, Qian, G and Alonso, FJ (1999a). Spatial sampling design based on stochastic complexity. *J Mult Anal*, 71, 94-110.

[4] Bueso, MC, Angulo, JM, Curz-Sanjuliàn, J and García-Aróstegui, JL (1999b). Optimal spatial sampling design in a multivariate framework. *Mathematical geology*, 31, 507 - 525.

[5] Caselton, WF and Husain, T (1980). Hydrologic networks: information transmission. *Water Resources Planning and Management Division, ASCE*, 106(WR2), 503-520.

[6] Caselton, WF and Zidek, JV (1984). Optimal monitoring network designs. *Statist & Prob Letters*, 2, 223-227.

[7] Caselton, WF, Kan, L & Zidek, JV (1992). Quality data networks that minimize entropy. In *Statistics in the Environmental & Earth Sciences.* (Eds) P Guttorp & A Walden. London: Griffin.

[8] Coles, S (2001). *An introduction to statistical modelling of extreme values.* UK:Springer.

[9] Coles, S & Pauli, F (2002). Models and inference for uncertainty in extremal dependence. *Biometrika*, 89, 183-196.

[10] Embrechts, P, Klüppelberg, C & Mikosch, T (1997). *Modelling extremal events for insurance and finance.* Berlin: Springer-Verlag, Berlin.

[11] Fisher, RA & Tippett, LHC (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc Cambridge Phil Soc*, 24, 180-190. Reproduced in Fisher (1950, Paper 15).

[12] Fu, A (2002). Case study: inference for extreme spatial random rainfall fields. MSc Thesis. Department of Statistics, University of British Columbia.

[13] Fu, AQ, Le, ND & Zidek, JV (2003). A statistical characterization of a simulated Canadian annual maximum rainfall field. TR 2003-17, Statistical and Mathematical Sciences Institute, RTP, NC.

[14] Guttorp, P, Le, ND, Sampson, PD and Zidek, JV (1993). Using entropy in the redesign of an environmental monitoring network. In *Multivariate Environmental Statistics*, (Eds) GP Patil, CR Rao and NP Ross. New York: North Holland/Elsevier Science, 175-202.

[15] Gumbel, EJ (1958). *Statistics of extremes.* New York: Columbia University Press.

[16] Heffernan, JE & Tawn, JA (2004). A conditional approach for multivariate extreme values (with discussion). *J Roy Statist Soc, Ser B*, 66, 497-546.

[17] Kennedy, WJ and Gentle, JE (1980). *Statistical computing.* New York: Marcel Dekker, Inc.

[18] Kibria GBM, Sun L, Zidek V, Le, ND. (2002). Bayesian spatial prediction of random space-time fields with application to mapping $PM_{2.5}$ exposure. *J Amer Stat Assoc*, 457, 101-112.

[19] Le, ND and Zidek, JV (1994). Network designs for monitoring multivariate random spatial fields. In *Recent Adv in Stat and Prob*, (Eds) JP Vilaplana and ML Puri, 191-206.

[20] Le, ND, Sun, L and Zidek, JV (2003). Designing networks for monitoring multivariate environmental fields using data with monotone pattern. TR 2003-5, Statistical and Applied Mathematical Sciences Institute, RTP, NC.

[21] Le, ND and Zidek, JV (2005). *Statistics in Environmental Process Analysis.* Forthcoming.

[22] Leadbetter, MR, Lindgren, G & Rootzén, H(1983). *Extremes and related properties of random sequences and series.* New York: Springer.

[23] Lindley, DV (1956). On the measure of the information provided by an experiment. *Ann Math Statist* , 27, 968-1005.

[24] Müller, WG (2001). *Collecting spatial data: optimum design of experiments for random fields, Second Edition.* Hiedelberg: Physica-Verlag.

[25] Sebastiani, P, and Wynn, HP (2000). Maximum entropy sampling and optimal Bayesian experimental design. *JR Statist Soc, Ser B*, 62, 145-157.

[26] Shewry, M, and Wynn, H (1987). Maximum entropy sampling. *J Applied Statist*, 14, 165-207.

[27] Smith, RL (2004). Contribution to the discussion of Heffernan and Tawn (2004).

[28] Sun, L, Zidek, JV, Le, ND and Özkaynak, H. (2000) Interpolating Vancouver's daily ambient PM_{10} field. *Environmetrics*, 11, 651-663.

[29] Sun, W (1998). Comparison of a co-kriging method with a Bayesian alternative. *Environmetrics*, 9, 445-457.

[30] Sun, W, Le, ND, Zidek, JV & Burnett, R (1998). Assessment of Bayesian multivariate interpolation approach for health impact studies. *Environmetrics*, 9, 565-586.

[31] Wu, S and Zidek, JV (1992). An entropy based review of selected NADP/NTN network sites for 1983-86. *Atmospheric Environment.*, 26A, 2089-2103.

[32] Zidek, JV, Sun, W and Le, ND. (2000). Designing and integrating composite networks for monitoring multivariate gaussian pollution fields. *Appl Statist*, 49, 63-79.

[33] Zidek, JV, Sun, L, Le, ND and Özkaynak, H (2002). Contending with Space-time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM_{10} field. *Environmetrics*, 13, 595- 613

[34] Zidek, JV, Meloche, J, Shaddick, G, C hatfield, C, and White, R (2003). A computational model for estimating personal exposure to air pollutants with application to London's PM_{10} in 1997. TR 2003-3, Statistical and Applied Mathematical Studies Institute, Research Triangle Park, NC.