Monitoring Network Design

Zhuowei Cheng, Lucy Lu, Evgeny Noi, Jing Xu

1 Introduction

The design of the monitoring network is critical, directly affecting the observed measures, the derived statistics and inferences, and eventually the decision making. This project aims to study different approaches of monitoring network design. The theories, implementation and empirical studies of three types of network design, including prediction based, multivariate warping based and space filling methods, are investigated. Prediction-based criteria models try to minimize parameters associated with fitting, either through minimization of the weighted variance or through maximization of the number of points-pairs contributing to each lag distance. Multivariate warping based method has been designed to use for multivariate responses and the general k-step staircase pattern of missing data. The method uses EM and Sampson-Guttorp methods to estimate hyperparameters of the Gaussian-generalized inverted Wishart model and spatial covariance matrix for gauged and ungauged sites based on the observed multivariate data. Then, with an entropy criterion for network extension, new sites can be added to the existing network. A space-filling method called the p-dispersion network design, which seeks to locate p sites to place monitors that maximize the minimum distance between any two sites, is studied. Two extensions to the model including extending an existing network and incorporating cost are conducted and evaluated with empirical studies.

Overall, these methods are compared across three different study areas: New York City ($n = 9$), Los Angeles ($n = 8$) and Santa Barbara County ($n = 8$). Interestingly, the complexity of the methods explored in this article can be compared in the following manner: space-filling design look into space domain only ($\vec{X}, \vec{Y}$), prediction-based criteria models add the estimated values, like ozone concentration ($\vec{X}, \vec{Y}, \vec{Z}$), and, finally, multivariate-warping method adds time ($\vec{X}, \vec{Y}, \vec{Z}, \vec{T}$). The former method is deterministic, while the latter two are probabilistic.

2 Methods

2.1 Prediction-based criteria [Evgeny]

Optimal configuration, particularly in initial sampling (also known as first-phase sampling), seeks to estimate the variogram by clustering sites within short distances between one another. However, when the covariance structure is unknown, space-filling models perform well [3, 13, 25, 26, 28], particularly those utilizing triangle or square sampling configurations [37]. Ideally, a second-stage sampling should capitalize on the ascertained variogram structure of

the network. This is usually accomplished by minimizing kriging variance [32], maximum likelihood estimation [7], coupling spatial weights with simulated annealing-based procedure [34, 27]. In practice, minimizing kriging variance would yield additional sampling sites at intermediate positions between existing samples, at the expense of underlying spatial variation of the modeled attribute. To remedy this incongruity, we incorporate weighted function into kriging variance estimation.

Assume we need to measure a certain parameter Y at m locations within the study area D. Given a grid set of nodes $s_g(g =, 1, 2, ...G)$, kriging variance can be calculated as follows:

$$(\sigma_K(S_g))^2 = \sigma^2 - c^T(S_g) \cdot C^{-1} \cdot c(S_g) \tag{1}$$

where C^{-1} is the covariance matrix C based on the covariogram function. The term c is a column vector and c^T is a corresponding row vector. The Total Kriging Variance (TKV) is obtained by integrating Eq (1) over D. Computationally, it is easier to discretize D over a fine grid of points (set G). The Average Kriging Variance (AKV) over D becomes:

$$TKV = \frac{\int_D (\sigma_K(S_g))^2}{|D|} \approx AKV = \frac{1}{[G]} \sum_{g \in G} (\sigma_K(S_g))^2 \tag{2}$$

At each iteration, the network with m samples is reconfigured by addition of n new samples to the total of N samples. Then, two AKVs are compared and absolute value of difference is calculated to account for negative kriging weights. The sampling objective $Q[S]$ is thus:

$$\underset{\{s_{m+1},\ldots,s_{m+n}\},m \in P}{maximize} Q[S] = AKV_{old} - AKV_{new} = \frac{1}{|G|} \sum_{g \in G} \left| (\sigma_k^{old}(S_g))^2 - \sigma_k^{new}(S_g))^2 \right| \tag{3}$$

where S is a sampling pattern, P is a set of p potential points added to the network, such that $m \neq n$, $\forall m \in M$, $\forall n \in N$.

To account for roughness of kriging estimates in nearby points, a moving circular filter ($S_j \in J$) is consecutively applied to all points (S_g) on the grid (G). Then squared difference is computed between interpolated values $\hat{y}(S_g)$ of the modeled compound at central grid point and the derived values of nodes $\hat{y}(S_j)$ within the filter search radius. The influence of the neighboring points is inversely weighted by the distance parameter $d(S_j, S_g)^{-\beta}$:

$$\lambda(S_g) = \sum_{j \in J, j \neq g} \frac{d(S_j, S_g)^{-\beta} \cdot (\hat{y}(S_j) - \hat{y}(S_g))^2}{\sum_{j=1, j \neq g}^{J} d(S_j, S_g)^{-\beta}} \tag{4}$$

By plugging in Eq (4) into Eq (3), we arrive at weighted objective function:

$$\underset{\{s_{m+1},\ldots,s_{m+n}\},m \in P}{maximize} Q[S] = \frac{1}{|G|} \sum_{g \in G} \left(\frac{\lambda(S_g)}{\underset{s_g \in G}{argmax} \lambda(s_g} \right)^{\alpha} \cdot \left| (\sigma_k^{old}(S_g))^2 - \sigma_k^{new}(S_g))^2 \right| \tag{5}$$

where α is the parameters controlling the importance of the weight factor in the objection function. This optimization approach has been applied by Delmelle Goovaerts [6] for second-phase sampling.

Because objective function is non-linear, the search of optimal sampling configuration must apply appropriate heuristic method, such as greedy algorithm (GA), tabu search, simulated annealing (SA) and genetic algorithms [5]. Greedy algorithm employs sequential addition of the new sampling sites and tends to produce sub-optimal solutions, but takes less time to converge. Simulated annealing utilizes simultaneous addition generating more optimal solution at the expense of computational time. In fact, many geostatistical use cases exploit annealing due to its inherent stochastic nature. Simulated annealing was first developed by Kirkpatrick et al. [14] and is often applied to problems on discretized space, particularly in soil sampling [33]. SA makes use of two tricks: first, it utilizes the so-called Metropolis algorithm that enables the 'bad' moves to explore possible space of solutions, even when there is no increase in the objective function values; second, it allows controlled decrease of the 'temperature', limiting the number of random 'bad' walks to concentrate on local optima.

Two packages, specifically designed for sampling re-configurations exist: *spsann* [30] and *geospt* [29]. The former utilizes *spatial simulated annealing* to optimize sampling configuration for the purposes of variogram estimation, spatial trend estimation, and spatial interpolation. We further describe the usage and metaheuristics employed in this packages in Chapter 4.1.

2.2 Multivariate Warping Based Method [Zhuowei and Lucy]

Multivariate warping based method has been designed to use for multivariate responses and the general k-step staircase pattern of missing data. Given the observed data and hyperparameters, the predictive distribution has been derived for the unobserved locations and times. The method uses EM to estimate hyperparameters in gauged sites and utilizes the Sampson-Guttorp method to estimate hyperparameters in ungauged sites. Then, with an entropy criterion for network extension, newsites can be added to the existing network.

2.2.1 Predictive Distribution [Lucy]

Predictive distribution for unobserved data is the basis in multivariate warping based method. The EM algorithm and Sampson-Guttorp method are based on it. Before deriving the predictive distribution, we define some parameters.

Consider the notation:

Parameters:

p: number of responses considered at each station

l: number of covariates (e.g. number of air pollutants we are interested)

n: number of timepoints (e.g. number of days)

u: number of stations with no monitors (i.e. ungauged sites)

g: number of stations with monitors (i.e. gauged sites)

Staircase Notation

Considering that not all of stations starts monitoring at the same time, so the data matrix has a lot of missing data. The stations can be organized into k blocks where the g_j $(j = 1, 2, ..., k)$ sites in the jth block have the same number of missing data m_j.Then these blocks are numbered with a staircase pattern of missing data [38], that is,

$$m_1 \geq m_2 \geq ... \geq m_k \geq 0$$

However, the mathematical work in the general case of several blocks entails a heavy burden of notation that obscures the main ideas. Thus, we present the simpler case of two blocks where we hope the key ideas are clearer.

Consider a special case where k=2, i.e., a two-block setting for the data. Let g_1 and g_2 denote the numbers of gauged sites in Blocks 1 and 2, respectively, with $g = g_1 + g_2$. In turn, m_1 and m_2 are the numbers of missing responses in each of the two blocks.

Response Variables

The response variables can accordingly be organized as

$$Y = [Y^{[u]}, Y^{[g]}]$$

$Y^{[u]}$:$n \times up$ denotes the unobserved responses at ungauged sites

$Y^{[g]}$:$n \times gp$ denotes the missing and observed responses at gauged sites. Further,

$$Y^{[g]} = [Y^{[g_1]}, Y^{[g_2]}] = \left[\begin{pmatrix} Y^{[g_1^m]} \\ Y^{[g_1^o]} \end{pmatrix}, \begin{pmatrix} Y^{[g_2^m]} \\ Y^{[g_2^o]} \end{pmatrix} \right]$$

Here the superscripts m and o denote the missing and observed responses, respectively, and the subscript indicates a particular block 1 and 2.

Covariates

$Z_t = (Z_{t1}, ..., Z_{tl})^T$ denotes l covariates at timepoint t. Let $Z = \begin{pmatrix} Z_1^T \\ \vdots \\ Z_n^T \end{pmatrix}$

Parameter Partition

Partition the $l \times (u + g)p$ coefficient matrix β corresponding l covariates and covariance matrix Σ of dimension $(u + g)p \times (u + g)p$ over gauged and ungauged sites as

$$\beta = (\beta^{[u]}, \beta^{[g_1]}, \beta^{[g_2]}) \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma^{[u,u]} & \Sigma^{[u,g]} \\ \Sigma^{[g,u]} & \Sigma^{[g,g]} \end{pmatrix} \tag{6}$$

Further partition the covariance matrix $\Sigma^{[g,g]}$ for the 2 gauged site blocks: $\Sigma^{[g,g]} = \begin{pmatrix} \Sigma^{[g_1,g_1]} & \Sigma^{[g_1,g_2]} \\ \Sigma^{[g_2,g_1]} & \Sigma^{[g_2,g_2]} \end{pmatrix}$

Staircase Model Specification

We assume that the multivariate responses have a joint Gaussian distribution. The generalizecd inverted Wishart (GIW) distribution is a conjugate prior for a Gaussian distribution and it is very flexible to deal with the staircase structure of the observed data [21]. Thus we assume the response matrix Y follows the Gaussian-GIW model. Specifically, using the notation described above,

$$\begin{cases} Y|\beta, \Sigma \sim N(Z\beta, I_n \otimes \Sigma), \\ \beta|\Sigma, \beta_0 \sim N(\beta_0, F^{-1} \otimes \Sigma), \\ \Sigma \sim GIW(\Theta, \delta) \end{cases} \tag{7}$$

Where $N(\cdot, \cdot)$ denotes the multivariate Gaussian distribution, $\otimes$ refers to the Kronecker product [31], β_0 is the $l \times (g+u)p$ hyperparameter mean matrix of β, F^{-1} is an $l \times l$ positive definite matrix representing the variance component of β between its l rows, and Z is the matrix of covariates. Θ is a collection of hyperparameters described later and $\delta = (\delta_0, \delta_1, \cdots, \delta_k)$(In two blocks case, $k = 2$) representing degrees of freedom.

Through the Bartlett decomposition [16], the Σ in Eq (6) can be expressed as

$$\Sigma = \begin{pmatrix} \Sigma^{[u,u]} & \Sigma^{[u,g]} \\ \Sigma^{[g,u]} & \Sigma^{[g,g]} \end{pmatrix} = \begin{pmatrix} \Gamma^{[u]} + (\tau^{[u]})^T \Sigma^{[g,g]} \tau^{[u]} & (\tau^{[u]})^T \Sigma^{[g,g]} \\ \Sigma^{[g,g]}(\tau^{[u]}) & \Sigma^{[g,g]} \end{pmatrix} \tag{8}$$

where $\Gamma^{[u]} = \Sigma^{[u,u]} - \Sigma^{[u,g]}(\Sigma^{[g,g]})^{-1}\Sigma^{[g,u]}$, $\tau^{[u]} = (\Sigma^{[g,g]})^{-1}\Sigma^{[g,u]}$. Then Σ can be transformed to a new set of variables $(\Sigma^{[g,g]}, \Gamma^{[u]}, \tau^{[u]})$.

That is, the above GIW distribution is defined, through the Bartlett decompostion, in a stepwise function starting with

$$\begin{cases} \Sigma^{[g,g]} \sim GIW(\Theta^{[g]}, \delta^{[g]}) \\ \Gamma^{[u]} \sim IW(\Lambda_0 \otimes \Omega, \delta_0) \\ \tau^{[u]}|\Gamma^{[u]} \sim N(\tau_{00}, H_0 \otimes \Gamma^{[u]}) \end{cases} \tag{9}$$

Λ_0, Ω,τ_{00}, H_0 are four parameters coming from ungauged sites. The Λ_0 represent the residual spatial covariance between sites while Ω reflects the covariance within sites (for example, between responses). The matrix τ_{00} is the hyperparameter of $\tau^{[u]}$; and the matrix H_0 is the variance component of $\tau^{[u]}$ between its rows.

Note that the GIW distribution is defined recursively. The stepwise definition then continues with $\Sigma^{[g,g]}$, i.e.,$\Sigma^{[g_1,g_2]}$ through the Bartelett transformation into a new set of variables $(\Sigma_{22}, \Gamma_1, \tau_1)$ following

$$\begin{cases} \Sigma_{22} \sim IW(\Lambda_2 \otimes \Omega, \delta_2) \\ \Gamma_1 \sim IW(\Lambda_1 \otimes \Omega, \delta_1) \\ \tau_1|\Gamma_1 \sim N(\tau_{01}, H_1 \otimes \Gamma_1) \end{cases} \tag{10}$$

where $\Sigma_{22} = \Sigma^{[g_2,g_2]}$,$\Gamma_1 = \Sigma^{[g_1,g_1]} - \Sigma^{[g_1,g_2]}(\Sigma^{[g_2,g_2]})^{-1}\Sigma^{[g_2,g_1]}$, $\tau_1 = (\Sigma^{[g_2,g_2]})^{-1}\Sigma^{(g_1,g_2)}$.

By Eq (7) Eq (9) and Eq (10), the hyperparameters involved in this two-block Gaussian-GIW model can be written as $\mathcal{H} = \{\beta_0, F, \Theta, \delta\}$, where $\Theta = \{\Omega, (\Lambda_0, \tau_{00}, H_0), (\Lambda_1, \tau_{01}, H_1), \Lambda_2\}$ and $\delta = (\delta_0, \delta_1, \delta_2)^T$

In general case, $\Theta = \{\Omega, \Lambda_k, (\Lambda_j, \tau_{0j}, H_j), j = 1, \cdots, k-1\}$ and $\delta = (\delta_0, \delta_1, \cdots, \delta_k)^T$

Then, the $\mathcal{H}$ would be

$$\mathcal{H} = \{\beta_0, F, \Omega, \Lambda_k, \delta_k, (\Lambda_j, \tau_{0j}, H_j, \delta_j), j = 0, \cdots, k-1\} \tag{11}$$

Among these, the hyperparameters for gauged sites would be

$$\mathcal{H}_g = \{\beta_0, F, \Omega, \Lambda_k, \delta_k, (\Lambda_j, \tau_{0j}, H_j, \delta_j), j = 1, \cdots, k-1\} \tag{12}$$

and the remaining hyperparameters for ungauged sites would be

$$\mathcal{H}_u = \{\Lambda_0, \tau_{00}, H_0, \delta_0\} \tag{13}$$

Spatial Predictive Distribution

Let D and Y_{unob} denote the data set and unobserved response, respectively. That is $D = \{Y^{[g_1^o]}, \cdots, Y^{[g_k^o]}\}$ and $Y_{unob} = \{Y^{[u]}, Y^{[g_1^m]}, \cdots, Y^{[g_k^m]}\}$. Under the model (7), the predictive distribution of the unobserved responses conditional on the observed data D and the hyperparemeter set $\mathcal{H}$ is given by

$$(Y_{unob}|D, \mathcal{H}) \sim (Y^{[u]}|Y^{[g_1^m, \cdots, g_k^m]}, D, \mathcal{H}) \prod_{j=1}^{k} (Y^{[g_j^m]}|Y^{[g_{j+1}^m, \cdots, g_k^m]}, D, \mathcal{H}) \times (Y^{[g_k^m]}|D, \mathcal{H}) \tag{14}$$

where the three components of the conditional distributions are specified as follows

$$(Y^{[g_k^m]}|D, \mathcal{H}) \sim t_{m_k * g_k p}(\mu^{[2]}_{(u|g)}, \Phi^{[2]}_{(u|g)} \otimes \Psi^{[2]}_{(u|g)}, \delta^{[2]}_{(u|g)})$$

$$(Y^{[g_j^m]}|Y^{[g_{j+1}^m, \cdots, g_k^m]}, D, \mathcal{H}) \sim t_{m_j * g_j p}(\mu^{[1]}_{(u|g)}, \Phi^{[1]}_{(u|g)} \otimes \Psi^{[1]}_{(u|g)}, \delta^{[1]}_{(u|g)})$$

$$(Y^{[u]}|Y^{[g_1^m, \cdots, g_k^m]}, Y^{[g_2^m]}, D, \mathcal{H}) \sim t_{n \times up}(\mu^{[u|g]}, \tfrac{\Phi^{[u|g]} \otimes \Lambda_0 \otimes \Omega_0}{\delta_0 - up + 1}, \delta_0 - up + 1)$$

Here $t_{m_j \times g_j}$ denotes a matric t-distribution [38] and for j =1,2

$$\mu^{[j]}_{(u|g)} = \mu^{[j]}_{(1)} + A^{[j]}_{12}(A^{[j]}_{22})^{-1}(Y^{[g_j^o]} - \mu^{[j]}_{(2)})$$

$$\Phi^{[j]}_{(u|g)} = \tfrac{\delta_j - g_j p + 1}{\delta_j - g_j p + n - m_j + 1}[A^{[j]}_{11} - A^{[j]}_{12}(A^{[j]}_{22})^{-1}A^{[j]}_{21}$$

$$\Psi^{[j]}_{(u|g)} = \tfrac{1}{\delta_j - g_j p + 1}[\Psi_j + (Y^{[g_j^o]} - \mu^{[j]}_{(2)})^T(A^{[j]}_{22})^{-1}(Y^{[g_j^o]} - \mu^{[j]}_{(2)})]$$

$$\delta^{[j]}_{(u|g)} = \delta_j - g_j p + n - m_j + 1$$

At the same time,

$$\mu^{[u|g]} = Z\beta_0^{[u]} + (Y^{[g]} - Z\beta_0^{[g]})\tau_{00},$$

$$\Phi^{[u|g]} = I_n + ZF^{-1}Z^T + (Y^{[g]} - Z\beta_0^{[g]})\tilde{\epsilon}^{[g]}H_0(Y^{[g]} - Z\beta_0^{[g]})^T$$

with

$$\begin{pmatrix} \mu^{[j]}_{(1)} \\ \mu^{[j]}_{(2)} \end{pmatrix} = Z\beta_0^{[g_j]} + \tilde{\epsilon}^{[g_{j+1},\cdots,g_k]}\tau_{0j},$$

$$\begin{pmatrix} A^{[j]}_{[11]} & A^{[j]}_{[12]} \\ A^{[j]}_{[21]} & A^{[j]}_{[22]} \end{pmatrix} = I_n + ZF^{-1}Z^T + \tilde{\epsilon}^{[g_{j+1},\cdots,g_k]}H_j(\tilde{\epsilon}^{[g_{j+1},\cdots,g_k]})^T,$$

where

$$\tilde{\epsilon}^{[g_{j+1},\cdots,g_k]} = \begin{cases} Y^{[g_{j+1},\cdots,g_k]} - Z\beta_0^{[g_{j+1},\cdots,g_k]} & j = 1,\cdots,k-1 \\ 0 & j = k \end{cases}$$

Under the Gaussian-GIW model for staircase structure, the resulting predictive distribution is a product of matric-t distributions. The means $\mu^{[j]}_{(u|g)}$ represents the linear predictor for the missing observations at gauged sites in the jth block based on the observed responses. Similarly, $\mu^{[u|g]}$ represents the linear predictors for the contaminant levels at the ungauged sites. The matrices $\Phi^{[j]}_{(u|g)}$ $\Psi^{[j]}_{(u|g)}$ represent the covariance structure of the predictive distribution. The variance structure from the gauged sites within the block contribute through $\Psi^{[j]}_{(\mu|g)}$ while the variance structure from other blocks contribute through $\Phi^{[j]}_{(u|g)}$. Moreover, for the ungauged sites, the observed data at the gauged sites contribute though $\Phi^{[u|g]}$.

2.2.2 Estimating gauged Site Hyperparameters [Lucy]

Conditional on the observed data and the hyperparameters, we derive the predictive distribution (14) for the unobserved locations and times. Then the hyperparameters in the model associated with the gauged sites $\mathcal{H}_g$ (12) can be estimated using the EM algorithm and those associated with the ungauged sites $\mathcal{H}_u$ (13) can be estimated via the Sampson-Guttorp method.

EM Algorithm

The EM algorithm facilitates the computation of the maximum likelihood for $\mathcal{H}_g$. Suppose at the pth iteration, the current estimate for gauged sites are

$$\mathcal{H}_g^{(p)} = \{\beta_0^{(p)}, F^{(p)}, \Omega^{(p)}, \Lambda_k^{(p)}, \delta_k^{(p)}, (\Lambda_j^{(p)}, \tau_{0j}^{(p)}, H_j^{(p)}), j = 1,\cdots,k-1\} \tag{15}$$

There are two steps in the EM algorithm at the $(p+1)$st iteration:

E-step: The computation of

$$Q(\mathcal{H}_g|\mathcal{H}_g^{(p)}) = E(log[f(Y^{[g]}, \beta, \Sigma|\mathcal{H}_g)]|D, \mathcal{H}_g^{(p)}) \tag{16}$$

given the previous parameter estimates $\mathcal{H}_g^{(p)}$ from iteration p.

M-step: Maximize the resulting $Q(\mathcal{H}_g|\mathcal{H}_g^{(p)})$ over $\mathcal{H}_g$ to obtain the updated estimate $\mathcal{H}_g^{(p+1)}$ of $\mathcal{H}_g$ at step $(p+1)$. The estimates for the hyperparameters can be obtained by iterating these EM steps until the marginal density given below converges.

Marginal Distribution

Assume the observed response $D = \{Y^{[g_1^o]}, \cdots, Y^{[g_k^o]}\}$ has the Gaussian-GIW model specified by (7), then its marginal distribution can be similarly derived as predictive distribution for unobserved responses (14) and we get

$$f(\{Y^{[g_1^o]}, \cdots, Y^{[g_k^o]}|\mathcal{H}_g)) = \prod_{j=1}^{k} t_{(n-m_j)\times g_j p}(\mu_o^{[j]}, \Phi_o^{[j]} \otimes \Psi_o^{[j]}, \delta_o^{[j]}) \tag{17}$$

where $t_{a\times b}$ denotes a matix-t distribution [38] and

$\mu_0^{[j]} = \mu_2^{[j]}$

$\Phi_o^{[j]} = A_{22}$

$\Psi_o^{[j]} = \frac{1}{\delta_j - g_j p + 1}[\Gamma_j] \otimes \Omega$

$\delta^{[j]} = \delta_j - g_j p + 1$

where $\mu_{(2)}^{[j]}$ and A_{22} have been defined in (14).

2.2.3 Estimating Ungauged Site Hyperparameters [Zhuowei]

To estimate the hyperparameters for ungauged sites, one can use Sampson-Guttorp method. This method extends the estimated spatial covariance associated with the gauged sites to include ungauged sites with a nonparametric technique.

Knowing the covariance matrix associated with the gauged sites, Sampson-Guttorp method can be applied to get estimates of the covariance matrix and the cross-covariance matrix associated with the ungauged sites. Then the estimates for Λ_0, τ_{00} can be obtained with Bartlett transformation.

The Sampson-Guttorp Method (SG Method)

Suppose data are available for N sampling stations at T time points. If Z_{it} and Z_{jt} represents observations at stations x_i and x_j at time t, a natural metric for the spatial covariance structure would be:

$$d_{ij}^2 = var(Z_{it} - Z_{jt}) = s_{ii} + s_{jj} - 2s_{ij}. \tag{18}$$

s_{ij} is the sample covariance between values at stations x_i and x_j.

$$s_{ij} = \frac{1}{T}\sum_{t=1}^{T}(Z_{it} - \bar{Z}_i)(Z_{jt} - \bar{Z}_j). \tag{19}$$

Here, d_{ij}^2 is the spatial dispersions. It is similar to sample the variogram, but often the variogram is associated with the assumption of stationarity. Here we do not impose an assumption of stationarity.

Let f: $R^2 \to R^2$ be a smooth nonlinear mapping from geographic space (G-space) to dispersion space (D-space). G-space is the plane of geographical coordinates x_i, for example, latitude and longitude from a map. The D-space representation y_i is determined from the space dispersion so that the space dispersion in D-space d_{ij}^2 can be approximated by $h_{ij} = |y_i - y_j|$ by a monotone function $d_{ij}^2 \approx g(|y_i - y_j|)$. Knowing a location x_i in G-space, one can obtain the corresponding location y_i in D-space by $y_i = f(x_i)$. Equivalently, x_i can be obtained by $x_i = f^{-1}(y_i)$.

The stationarity of the random field is not assumed for the G-space. But because the spatial dispersions in D-space are only dependant on h_{ij}, the spatial dispersion structure is both stationary and isotropic[38].

Sampson and Guttorp propose a two step method to estimate f and g[24]:

First, to compute the D-space representation, the Shepard-Kruskal multidimentional scaling algorithm[18] is used. This algorithm determines a monotone function g such that the matrix of $\delta(d_{ij}) \equiv \delta_{ij} \approx |y_i - y_j|$. Here, d_{ij}^2 are the sample dispersions between locations in G-space. Solving this relationship for d_{ij}^2 gives us $d_{ij}^2 \equiv (\delta^{-1}(\delta_{ij}))^2 \approx g(|y_i - y_j|)$. Now the algorithm searches for a configuration of y_i so the distances $h_{ij} = |y_i - y_j|$ minimize a stress criterion defined as

$$\min_{\delta} \sum_{i<j} \frac{(\delta(d_{ij}) - h_{ij})^2}{\sum_{i<j} h_{ij}^2} \tag{20}$$

The minimum is taken over all possible monotone functions.

Second, the thin-plate spline approach[10] is used to estimate the smooth mapping f that maps the coordinate representation of the sampling stations in the G-space to the scaled spatial dispersion coordinates in the D-space. Here, we consider bivariate problem. Let Cartesian coordinates in the G-space be denoted by $(x^{(1)}, x^{(2)})$, and those in the D-space $(y^{(1)}, y^{(2)})$

$$f(x) = \alpha_0 + \alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \sum_{i=1}^{N} \beta_i u_i(x) \tag{21}$$

where $u_i(x) = |x - x_i|^2 log|x - x_i|$. The parameters to be fitted are αs and βs. f is computed as two thin plate splines f_1 and f_2 for the two coordinates of y_i.

$f = (f_1, f_2)$ is computed to minimize

$$\sum_{j=1}^{2} \sum_{i=1}^{N} (y_i^{(j)} - f_j(x_i^{(j)}))^2 + \lambda(J_2(f_1) + J_2(f_2)), \tag{22}$$

for specified smoothing parameter λ, where λ can be defined by users. J_2 measures the smoothness of the functions defined as

$$J_2(f) = \int_{R^2} [(\frac{\partial^2 f}{\partial x^{(1)2}})^2 + 2(\frac{\partial^2 f}{\partial x^{(1)} \partial x^{(2)}})^2 + (\frac{\partial^2 f}{\partial x^{(2)2}})^2] dx^{(1)} dx^{(2)} \tag{23}$$

With estimated $\hat{f}$ and $\hat{g}$, the variogram between any two locations x_i and x_j can be estimated by:

1. Calculate the representation in D-space y_i and y_j with $y_i = \hat{f}(x_i)$, $y_j = \hat{f}(x_j)$;
2. Calculate the D-space distance $h_{ij} = |y_i - y_j|$;
3. Evaluate $2\gamma(x_i, x_j) = 2\hat{g}(h_{ij})$.

The covariance between locations can be estimated by

$$C(h) = C(0) - \hat{g}(h_{ij}), \tag{24}$$

where C(0) is the variance.

Hyperparameters Estimation

The spatial component covariance matrix can be written in terms of $\Sigma = \begin{pmatrix} \Sigma^{[u,u]} & \Sigma^{[u,g]} \\ \Sigma^{[g,u]} & \Sigma^{[g,g]} \end{pmatrix}$

Here, u stands for ungauged sites and g stands for gauged sites. $\Sigma^{[u,u]}$ represents the spatial covariance between ungauged sites. $\Sigma^{[g,g]}$ represents the spatial covariance between gauged sites. $\Sigma^{[u,g]}$ is the cross covariance which is given above as $\Psi^{[1,...,k]}$.

Let $\hat{\Lambda}_j$, $\hat{\tau}_{0j}$ and $\hat{\delta}_j$ denote the estimated hyperparameters associated with the gauged sites. $\Sigma^{[g,g]}$ can be estimated as in the previous section. Then, $\Sigma^{[u,u]}$ and $\Sigma^{[u,g]}$ can be estimated based on $\Sigma^{[g,g]}$ using SG method.

$$\Sigma^{[g,u]} = \Sigma^{[g,g]} \xi_0^{[u]}, \tag{25}$$

where $\tau_{0j} = \xi_{0j} \otimes I_p$. Hence $\tilde{\xi}_{00}$ can be estimated as

$$\tilde{\xi}_{00} = (\hat{\Sigma}^{[g,g]})^{-1} \tilde{M}^{[g,u]}. \tag{26}$$

Now τ_{00} and Λ_0 can be estimated as

$$\tilde{\tau}_{00} = (\hat{\Sigma}^{[g,g]})^{-1} \tilde{M}^{[g,u]} \otimes I_p \tag{27}$$

$$\tilde{\Lambda}_0 = \frac{\delta_0 - up - 1}{1 + tr((\Psi^{[1,...,k]} \otimes \Omega) H^{[u]})} (\tilde{M}^{[u,u]} - \tilde{\xi}_{00}^T \hat{M}^{[g,g]} \tilde{\xi}_{00}) \tag{28}$$

δ_0 can be estimated with a method proposed by Kibria et al[12] as $\tilde{\delta}_0 = min(\hat{\delta}_1, ..., \hat{\delta}_k)$ or $\tilde{\delta}_0 = \frac{\hat{\delta}_1 + ... + \hat{\delta}_k}{k}$, where $\delta_0 \geq up$.

2.2.4 Entropy in Network Design [Zhuowei]

Entropy measures the uncertainty. In the context of monitoring network design, the maximum amount of uncertainty reduction is one of the objectives. Thus, selecting sites to maximize entropy can be one of the criterion for network design. In this section, we would focus on network extension.

Entropy Expression

Suppose Y is a discrete random variable that takes a finite number of possible outcomes with probabilities p_i. Entropy H(Y) can be expressed as

$$H(Y) = \sum_{i=1}^{n} p_i log(p_i) \tag{29}$$

Now Let Y_t be a p-dimensional random row vector denoting the random field at time t. To obtain the entropy criterion, the predictive distribution for all locations is required. Y_t can be partitioned as $Y_t = Y_t^{(u)}, Y_t^{(g)}$. Again u means ungauged sites and g means gauged sites.

$$Y_t | z_t, B, \Sigma \sim N_p(z_t B, \Sigma) \tag{30}$$

Here, z_t is the covariates and B denotes a matrix of regression coefficients $B \equiv (B^{(u)} B^{(g)})$

$$B | B_o, \Sigma, F \sim N_{kp}(B_o, F^{-1} \otimes \Sigma) \tag{31}$$

$$\Sigma | \Psi, \delta \sim W_p^{-1}(\Psi, \delta) \tag{32}$$

The predictive distribution of $Y_f = (Y_f^{(u)'}, Y_f^{(g)'})$ given covariate vector z_f and the prior hyperparameters B_o and $(\psi_{gg}, \psi_{u|g}, \tau_o)$ is

$$Y_f^{(g)} | D \sim t_g(\mu_{(g)}, \frac{c}{l} \hat{\psi}_{gg}, l) \tag{33}$$

$$Y_f^{(u)} | Y_f^{(g)} = y_f^{(g)}, D \sim t_u(\mu_{(u)}, \frac{d}{q} \psi_{u|g}, q), \tag{34}$$

where t_r means the r-variate Student's t distribution, $l = \delta + n - u - g + 1$, $q = \delta - u + 1$, and

$$\hat{\Psi}_{gg} = \Psi_{gg} + S + (\hat{B}^{(g)} - B_o^{(g)})^T (A^{-1} + F^{-1})^{-1} (\hat{B}^{(g)} - B_o^{(g)}) \tag{35}$$

$$\mu^{(g)} = (1 - W) \hat{B}^{(g)} + W B_o^{(g)} \tag{36}$$

$$\mu^{(u)} = z_f B_o^{(u)} + \tau_o (y_f^{(g)} - z_f B_o^{(g)}). \tag{37}$$

Ψ_{gg} is the hypercovariance matrix between gauged sites and $\Psi_{u|g}$ is the residual hypercovariance matrix between ungauged sites.

Now, the total entropy can be expressed

$$H(Y_f|D) = H(Y_f^{(u)}|Y_f^{(g)}, D) + H(Y_f^{(g)}|D) = \frac{1}{2}log|\Psi_{u|g}| + c_u(u,q) + \frac{1}{2}log|\hat{\Psi}_{gg}| + c_g(g,l), \quad (38)$$

Criterion for Extension

Assume n_1 sites need to be added among n potential sites. The add sites are selected to maximize the entropy in Eq (38). The optimality criterion is

$$\max_{add}(\frac{1}{2}log|\Psi_{u|g}|)^{add} \quad (39)$$

This approach has been extended to deal with multivariate responses in staircase patterns of missing data [22, 23].

2.3 Space-filling Method – The *p*-Dispersion [Jing]

The space-filling design is another primary type of network design, which seeks to locate facilities such that a certain planning objective (or objectives) is maximized or minimized and the network satisfies the imposed constraints. Space-filling designs have been developed mainly for computer experiments but are also important to environmental monitoring studies. In environmental monitoring design, the goal is usually related to ensure no convex subregion of the study area with appreciable area lacks of monitoring points [1]. But note that the space-filling method differs with the probability-based design in the way that the locations of a space-filling design are determined by optimizing a mathematical objective rather than by randomization.The space-filling design also provides informative planning strategies when planners cannot define an appropriate model-based criterion or when there are multiple competing design objectives [1]. In addition, the space-filling design can be conducted with no prior knowledge about the model governing the observations and is generally more computationally manageable compared to the model-based design.

Most research attention has been devoted to five types of space-filling designs. The first one is called the p-dispersion (also named maximin distance design or Max-Min-Min) that seeks to locate p monitors such that the minimal distance (usually Euclidean distance in environmental monitoring) between two sited monitors [13, 17, 8]. In contrast to the the p-dispersion design, the minimax distance design minimizes the maximal distance from any point in the study area to its closest point in the design [20, 25]. A Latin hypercube design ensures that each one dimensional projection is a maximin distance design. The fourth type is the spatial coverage design that minimizes L_q-average of generalized distances between candidate points and the design [28]. Lastly, the regular grids design selects points among the gridpoints using the criteria of the p-dispersion or the minimax distance design.

Of particular interest of this section is the p-dispersion design. To monitor the whole study region, the p-dispersion design wishes to locate p monitors so that they are as far apart as

possible. Therefore, it is not surprising to maximize the distance between two closest sited monitors. Next, the mathematical formulation of the p-dispersion network design will be given followed by the modification to extend a network and inclusion of costs. Finally, the solution techniques of solving this type of spatial optimization problem are discussed.

2.3.1 The p-Dispersion Network Design

Consider the notation:

Parameters:

i: index of the candidate site

I: set of the candidate site, $|I| = n$

(x_i, y_i): the planar coordinates of location i

p: the number of monitors to site

M: a very large number

$|\cdot|$: the set size operator

Decision Variables:

Z_i: location variable, 1 if a monitor is sited at location i and 0 otherwise

d: the minimum Euclidean distance of two sited monitors

The p-dispersion problem for a new network design is formulated as the following.

$$\max \quad d \tag{40}$$

$$\text{s.t.} \quad \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} + M(1 - Z_i) + M(1 - Z_j) \geq d \quad \forall i, j \in I, j > i \tag{41}$$

$$\Sigma_{i \in I} Z_i = p \tag{42}$$

$$Z_i = \{0, 1\} \quad \forall i \in I \tag{43}$$

The objective (40) is to maximize d, the minimum Euclidean distance between sited monitors. Constraints (41) track the minimum distance between only sited monitors. The role of the big number M here is to control that no inter-monitor distance is considered if one of two sites or both is/are not selected. More specifically, if one of the site i and j is not selected to place a monitor, the left hand side of Constraints (41) will become the Euclidean distance between sites i and j plus M, which in fact imposes an ineffective upper bound to d since M is set to a very big number. Similarly, if both sites i and j are not selected, this upper bound is even larger, the distance plus $2M$. Therefore, d represents the minimum distance between sited monitors. Note that the distance here can be easily adjusted to other metrics such as the travel distance, Manhattan distance depending on the context. Exactly p monitors

can be sited according to Constraints (42). This can be viewed as budget and resource consideration. Finally, the decision variable Z_i is restricted to be a binary value indicating site i is selected to place a monitor or not. In total, there are $n+1$ decision variables and $\frac{n(n-1)}{2}+n+1$ constraints in the problem defined by (40)-(43). The set $D = \{i|Z_i = 1, \forall i \in I\}$ is the designed network.

2.3.2 The *p*-Dispersion Network Extension

There is an assumption in the above model that no monitor exists in the current network. That is to say, a brand new network is designed with the above model. However, this is usually not the case in practice. Often, there are already some monitors in the network and the problem is to extend the current monitor network by adding new monitors. To this end, some modification can be made.

A new notation E as the existing monitor set is needed and $|E| = m$. Constraints (41) are modified into Constraints (44) and (45):

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} + M(1 - Z_i) + M(1 - Z_j) \geq d \quad \forall i, j \in I, j > i \tag{44}$$

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} + M(1 - Z_i) + M(1 - Z_j) \geq d \quad \forall i \in I, j \in E \tag{45}$$

Constraints (44) track the distance between selected candidate sites while Constraints (45) track the distance between selected candidate sites and existing monitor sites. Now the network extension problem is (40), (42), (43), (44) and (45). There are totally $n+1$ decision variables and $\frac{n(n-1)}{2} + mn + n + 1$ constraints. The set $A = \{i|Z_i = 1, \forall i \in I\}$ contains sties to add and the network is $D = A \bigcup E$.

2.3.3 Cost

The p-dispersion model has been utilized to design a new network and extend an existing network so far, with a primary focus on "dispersing" monitors. An important element that has not been addressed yet is the cost of implementing the network in practice such as construction and maintenance fees. It is possible and easy to incorporate cost consideration into the p-dispersion model. Assume f_i is the cost of placing and maintaining a monitor at site i, then a second objective (46) can be written to minimize the total cost:

$$\min \quad \Sigma_{i \in I} f_i Z_i \tag{46}$$

Now the problem becomes a bi-objective optimization problem. Different approaches can be used to deal with multi-objective optimization problem [4, 9], among which two most prominent ways are weighting method and constraint method. The objective (47) is a weighted objective combining objectives (40) and (46) using weighting method. Here, α is a non-negative weighting and scaling factor to adjust the weights and scales between the

two competing objectives. Since the objective (46) is minimization, there is a negative sign connecting these two objectives.

$$\max \quad d - \alpha \Sigma_{i \in I} f_i Z_i \tag{47}$$

Converting the cost objective into a constraint (48) is an example of the constraint method. Here, B is an endogenously given budget.

$$\Sigma_{i \in I} f_i Z_i \leq B \tag{48}$$

In order to produce the trade-off curve to determine the Pareto-efficient solutions, weighting method is used in this research. Thus, the p-dispersion network design with cost consideration is (47) and (41)-(43) (or (47), (42), (43), (44) and (45) for the network extension).

2.3.4 Solution Techniques

Above models are 0-1 mixed integer linear programming problems, which are linear programming (LP) models with additional requirement that one or more decision variables have to be integers (e.g. Z_i here). Generally, the mixed integer linear programming problem is more computational intensive than the linear programming problem due to the integer constraints. It is not hard to find much research interest on solution techniques for a general or specified 0-1 mixed integer linear programming problem including exact and heuristic approaches [11, 36]. One common exact technique is to solve the relaxed LP problem and use branch-and-bound to reach the optimality of the integer programming problem, which has been incorporated in almost all linear programming solvers such as Gurobi, Xpress, Cplex and LINGO just to name a few. The network design models in this research are solved by the branch-and-bound technique provided by Xpress 8.4.

The focus here is the branch-and-bound tree that converts a hard integer problem to a collection of LP problems rather than how to solve a LP problem (more details can be found in [2]. So assume the LP problem can be solved optimally with some efficient methods (such as the Simplex method). Take the problem (40)-(43) as an example, it is natural to think about solving a problem that relaxes Constraints (43) with Constraints (49).

$$0 \leq Z_i \leq 1 \quad \forall i \in I \tag{49}$$

Thus, now the relaxed LP problem becomes (40)-(42) and (49). The optimal solution might give some Z_i with fraction values between 0 and 1 since no binary constraints are imposed the relaxed LP (denoted as LP1). Assume $Z_2 = 0.6$ and the optimal objective is $d_{LP1} = 100$. The algorithm creates two new problems based on Z_2: one problem (denoted as LP2) contains $Z_2 = 0$ (can be viewed as adding a bound) and the other (denoted as LP3) contains $Z_2 = 1$ and then solves the LP problems (Figure 1). The optimal objectives of LP2 and LP3 are 98.6 and 99, respectively. Note that as the constraints are added, the objective function,which is to maximize the minimum distance, decreases because the feasibility region is shrunk.

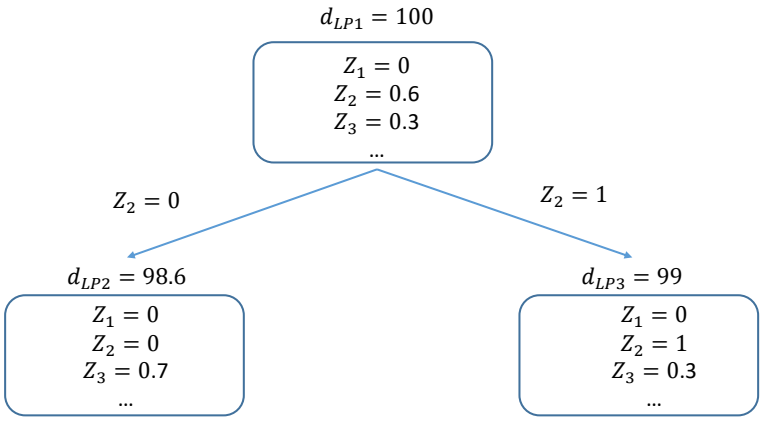


Figure 1: Branch-and-bound tree example

Then the LP2 and LP3 will be further branched into new LP problems since there are still non-integer variables in the solutions. This branching and bounding iteration continues until obtaining all-integer solutions or some predefined thresholds (such as iteration number, computation time) has been reached.

3 Data

Several different data sets have been used to assess the methods in this study and they were originally downloadeed from the EPA's air quality site (Note that the coordinates are projected coordinates by the Lambert projection functionality provided by the ENVIROSTAT package):

1. Los Angeles Ozone Hourly Data

This dataset consists of hourly Ozone levels(14:00-17:00) from eight stations in Los Angeles County in 2018 from March to October. The other months are omitted because of the large amount of missing data. Figure 2 shows the distribution of the existing eight sites and one hundred potential site locations.

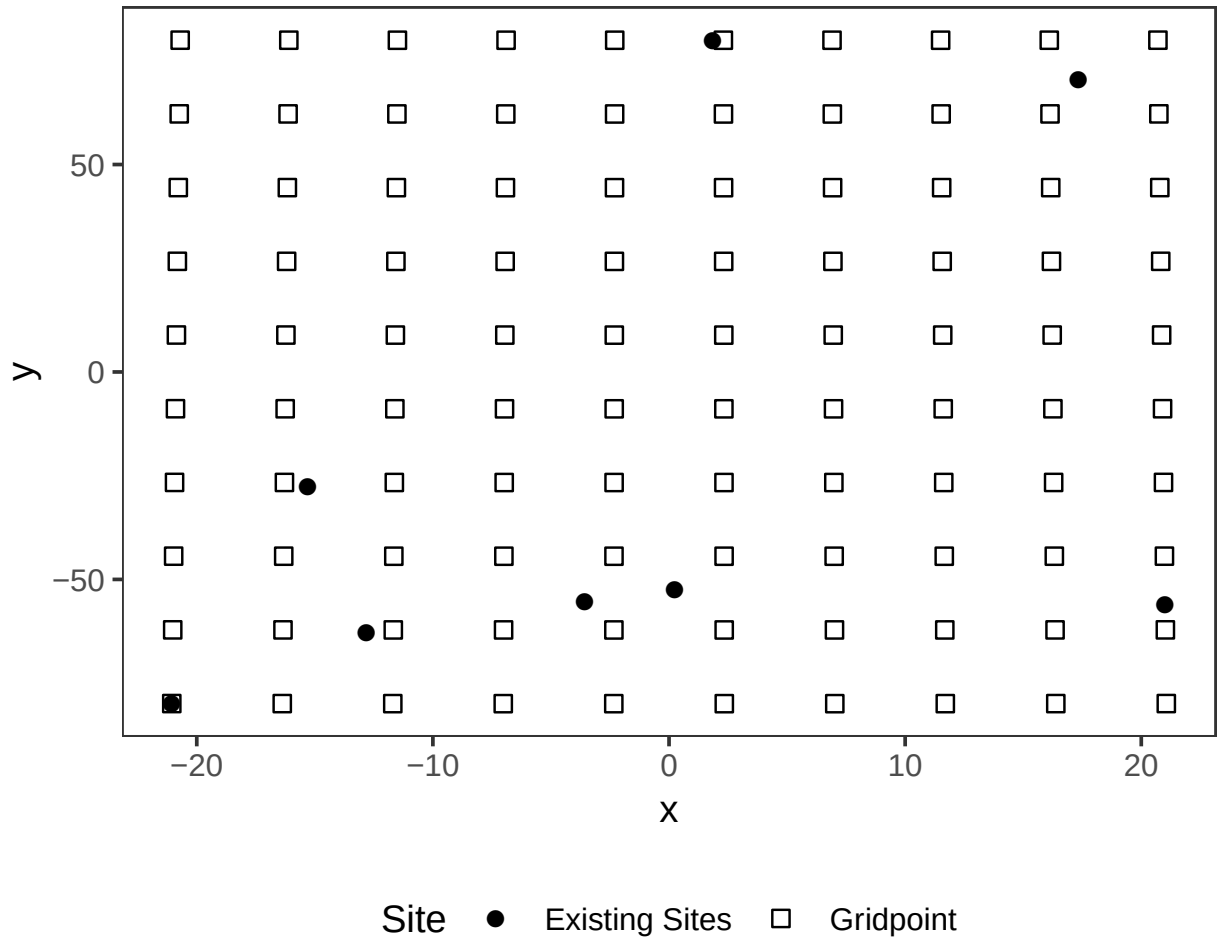


Figure 2: LA County Ozone Monitoring sites and potential new sites distribution

2. Santa Barbara Ozone Hourly Data

There are 8 existing gauged sites for monitoring Ozone data in Santa Barbara now. The dataset consists of hourly Ozone concentration levels from these eight stations at 15pm and 16pm everyday from January 1 to May 30, 2018. The other months are omitted because of the large amount of missing data. Figure 3 shows the distribution of the existing eight sites and one hundred potential site locations.

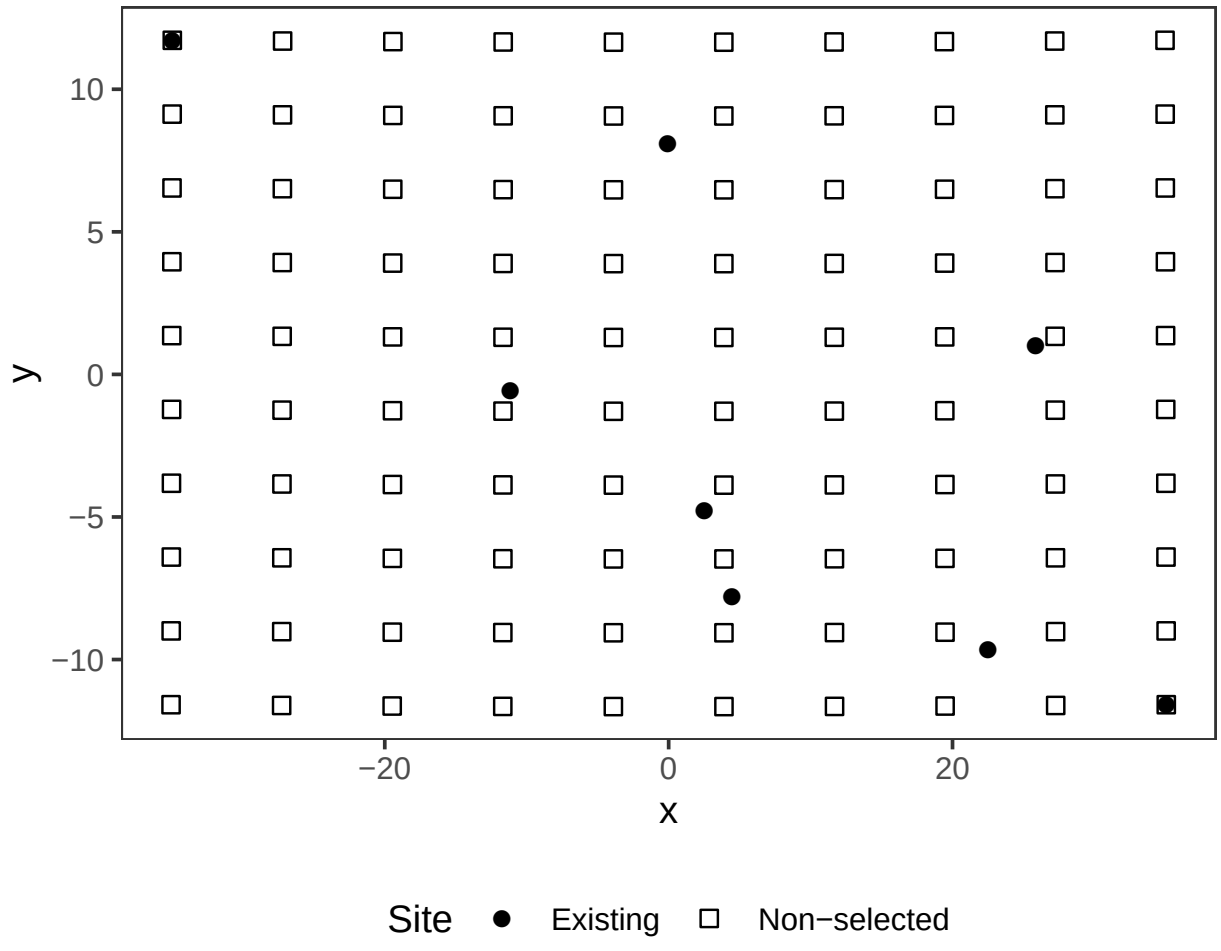


Figure 3: Santa Barbara County Ozone Monitoring sites and potential new sites distribution

3. New York Hourly Ozone Data

In New York area, there are 9 existing monitor sties and 100 evenly distributed gridpoints as the candidate sites for the network extension (Figure 4). Associated with the 9 sites, hourly Ozone concentration levels are available from 8am to 12pm everyday from April 1, 1995 to September 30, 1995.

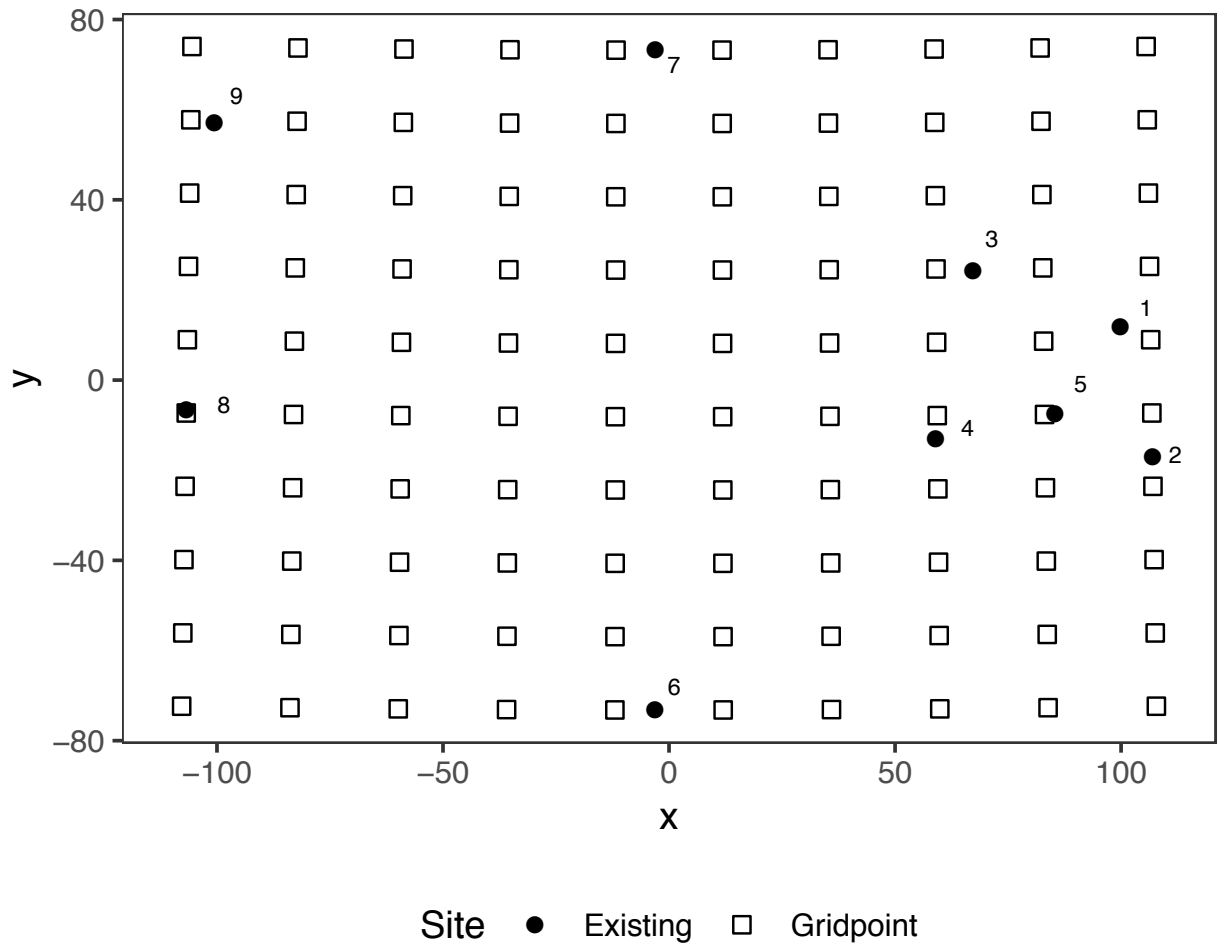


Figure 4: The existing 9 sites and evenly place 100 gridpoints

4 Results

4.1 Minimizing Kriging Variance [Evgeny]

Overall, we looked into two existing packages that reconfigure the sampling networks by minimizing kriging variance: *spsann* and *geospt*. These two packages were compared against two datasets with average ozone values for the specified observed period: Los Angeles ($n = 8$) and New York City ($n = 9$). Due to implementation specificity, for *geospt* the addition of 3 sampling sites was considered; for *spsann* a complete re-configuration was performed.

4.1.1 *geospt*: Optimization via Genetic Algorithms

geospt employs a genetic algorithm to minimize the value of average standard error (ASE) derived from the kriging interpolation:

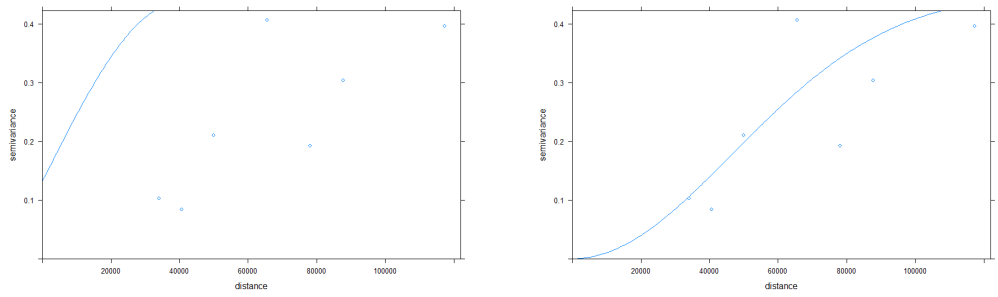
$$ASE = \frac{\sum_{i=1}^{n} \sigma_i}{n} \tag{50}$$

where σ_i is the square root of kriging variance and n is the number of sampling sites.

The optimization and re-sampling occurs in the following order:

1. Use *autofitVariogram* function from *automap* package to fit variogram.

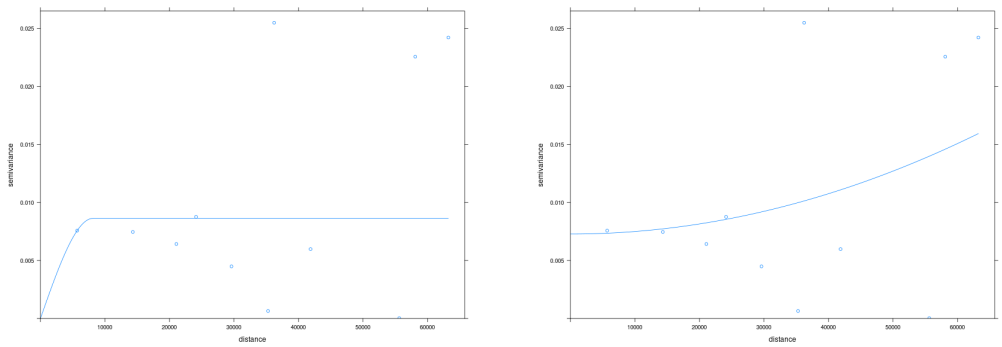
```
plot(ozvar <- variogram(oz~1, ozone.m))
auto.oz.var = autofitVariogram(oz~1,ozone.m)
oz.fit2 <- fit.variogram(ozvar, vgm(c("Exp", "Mat", "Sph", "Gau")),
    fit.kappa = TRUE)
```



(a) Variogram fitted via autofit

(b) Variogram with fitted Gaussian model

Figure 5: Fitted variograms for New York City ozone data



(a) Variogram fitted via autofit (b) Variogram with fitted Matheron model

Figure 6: Fitted variograms for Los Angeles ozone data

2. Use parameters from the previous step to fit in variogram into *gstat* fit object.

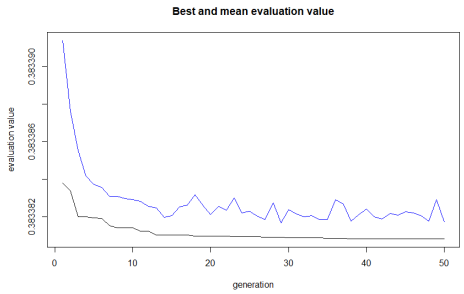
```
oz.fit3 <- fit.variogram(ozvar, vgm(sill=1.2, range=35338, 'Sph',
    nugget=.22))
```

3. Generate bounding box polygon to infer minimal and maximal values for longitude/latitude, these will be further used in Genetic Algorithm.

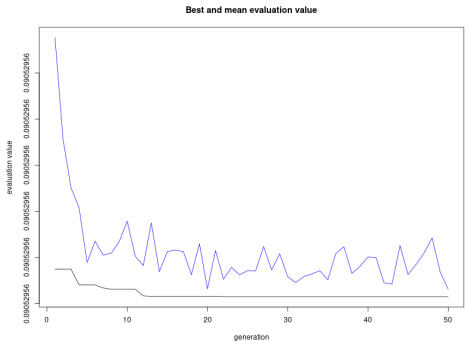
```
e <- as(raster::extent(xmin, xmax, ymin, ymax), "SpatialPolygons")
proj4string(e) <- CRS("+init=epsg:4326")
plot(e)
```

4. Initiate simultaneous additions to the network

```
add.points <- 3 # Number of additional points to be added to the
optnet.oz <- simPtsOptNet(oz~ 1, loc=~long+lat, df, oz.fit3, n=add.points,
    popSize=30, generations=50,
                          xmin=xmin, ymin=ymin, xmax=xmax, ymax=ymax,
                          plotMap=TRUE, spMap=e)
plot(optnet.oz)
```



(a) Evaluation value vs generations for GA: New York City data

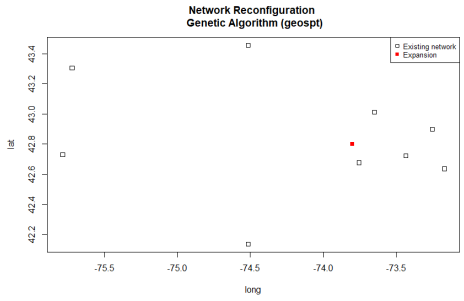


(b) Variogram with fitted Gaussian model

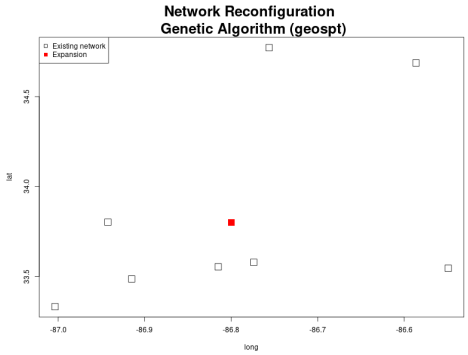
Figure 7: Evaluation value vs generations for GA: Los Angeles data

5. Extract best generations of solutions and plot with existing network. As we can see, three added points convert to one location, so that they become indistinguishable from one another.

```
(oz.bnet <- bestnet(optnet.oz))
plot(oz.bnet@coords, xlim=c(-73.85, -73.77), ylim=c(42.77, 42.85))

plot(df$long, df$lat, pch=0, xlab='long', ylab='lat',
     main=paste('Network Reconfiguration', '\n', 'Genetic Algorithm
         (geospt)'))
plot(oz.bnet, add=TRUE, pch=15, col='red') #col=alpha('red', .5))
legend("topright", pch=c(0,15), c("Existing network", "Expansion"),
       col=c("black", "red"), cex=0.8)
```



(a) Network re-configuration: New York City data



(b) Network re-configuration: Los Angeles data

Figure 8: Network re-configuration

As we can see the library tends to produce a somewhat questionable results. Furthermore, the performance of the optimization depends largely on the fitted variogram model. For some models that I tested (especially those with a singular fit), the library throws a **'singular covariance matrix'** error and never returns optimized sampling set. For most of the fitted

variogram that I fitted on the data, the algorithm tends to converge all three points to the same site. Whether this is by design or by implementation is impossible to trace, due to bad documentation for the library. Part of the reason is the small number of sampling locations in ozone data ($n_{NYC} = 9$, $n_{LA} = 8$), as well as the absence of spatial correlation

4.1.2 *spsann*: Optimization via Simulated Annealing

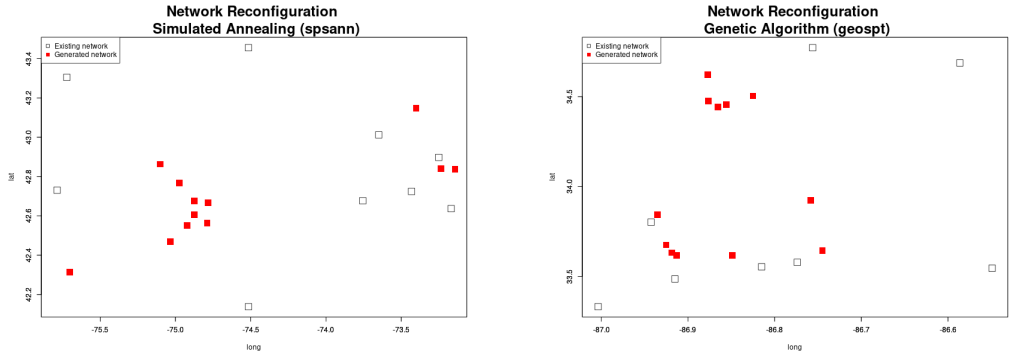
spsann package was designed for first-phase sampling. For New York City data, 12 stations were allocated (9 existing samples + 3 added as per other methods researched in this paper). For Los Angeles, 11 stations were allocated (8 existing + 3 added). *optimPPL* method is used to optimize a sample configuration for variogram identification and estimation. A criterion is defined so that the optimized sample configuration has a given (maximizied) number of points or point-pairs contributing to each lag-distance class (PPL).

1. Generate a 2d grid with 5000m step over the study area.

```
# set limits
xmin = min(df$long); xmax=max(df$long); ymin=min(df$lat); ymax=max(df$lat)

# generate raster via bounding box
e <- as(raster::extent(xmin, xmax, ymin, ymax), "SpatialPolygons")
proj4string(e) <- CRS("+init=epsg:4326")

# reproject and generate grid
e.m <- spTransform(e, CRS("+init=epsg:3857"))
grid <- makegrid(e.m, cellsize = 5000) # cellsize in map units!
grid <- SpatialPoints(grid, proj4string = CRS(proj4string(e.m)))

# prep the format for spsann
candi = as.data.frame(grid@coords)
names(candi) = c('x', 'y')
```

2. Run the optimization function and generate plots.

```
schedule <- scheduleSPSANN(initial.temperature = 500)
set.seed(2001)
# Execute the simulated annealing algorithm
res <- optimPPL(points = 12, candi = candi, pairs = TRUE, schedule =
    schedule,
                plotit = TRUE, boundary = e.m)
plot(res)
```



(a) Network re-configuration: New York City data (b) Network re-configuration: Los Angeles data

Figure 9: Network re-configuration

As we can see, the algorithm tends to cluster sampling sites in the center of the study areas for both New York City and Los Angeles with a somewhat smaller distances in between samples.

4.2 Network Extension of LA Ozone Data [Zhuowei]

Here, we apply the described multivariate warping based method to Los Angeles Ozone data with the help of ENVIROSTAT package in R [19]. There are eight existing sites in Los Angeles County and our goal is to add three sites among one hundred new locations to extend the network. The parameters we have for our data corresponding to the model in 2.2 are $p = 4, l = 1, n = 244, u = 100, g = 8$. Here, our interest is to interpolate the hourly Ozone levels at unobserved locations for a specific hour (17:00-18:00) of the day. The method described in 2.2 allows us to use a few preceding hours. In fact, predictive strength can be borrowed from the multivariate approach and substantial improvements are possible in some application [35]. Here, our data includes Ozone levels measured at from 14:00 to 17:00.

4.2.1 Pre-processing

The dataset also includes temporal information like day within month(from 1 to 31), month(from 1 to 8), weekday(from 1 to 7) information. Possible weekday and month effect can be assessed. A linear model with weekday and month as factors is fitted for each station separately.

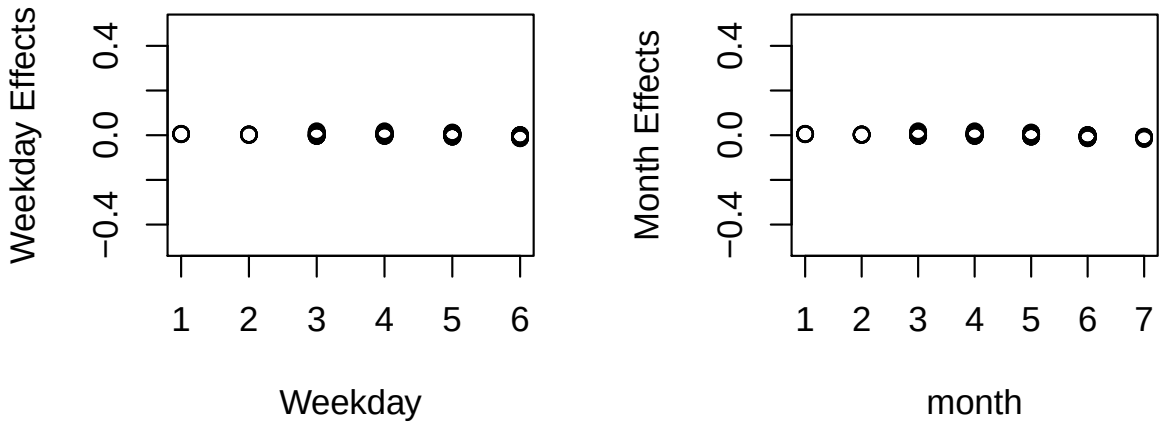


Figure 10: Estimated effects

Figure 10 shows consistent results of weekday and month effects for all stations, although the effect seems to be small. Next, the assumption of normality is checked. QQ-plots are obtained to check the fitted residuals. The results show that the normality assumption is satisfied. Here, Figure 11 shows the QQ-plots for one of the hours (17:00) at each station.

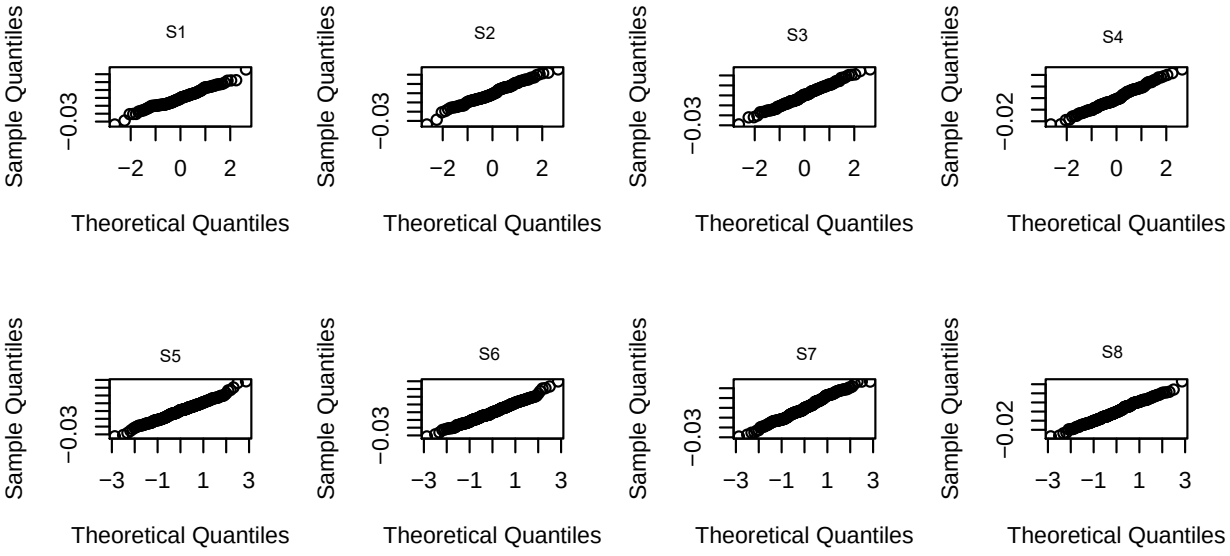


Figure 11: QQ plots of the fitted residuals for 17:00 at eight sites

There are in total 976 observations(244 days × 4 hours) for each station. Missing data consists of 1.4% of total observations in the dataset. With missing data, responses are organized

in a staircase structure as mentioned in 2.2.1. Staircase steps have a block structure. After arranged in staircase structure, our data has 6 blocks with one station in the first four blocks and two stations in the latter two.

The dataset includes the spatial location data for all sites(latitude and longitude). Since the earth's surface is curved, to project the latitude and longitude into a flat plane, Lambert project is used to obtain projected coordinates.

4.2.2 Hyperparameters Estimation

To estimate the intersite spatial covariance matrix for gauged sites, we use a staircase EM fit provided by the package. The multivariate response model is fitted and the hyperparameters for the gauged sites are estimated here.

Figure 12a checks the relationship between spatial correlation distance and the distance between sites. Figure 12b shows the dispersion curve against intersite distances with a fitted exponential variogram. Our results show that the spatial correlation is lower with longer distance. The dispersion goes up from close to zero when distance is close to zero to under two, which is expected.

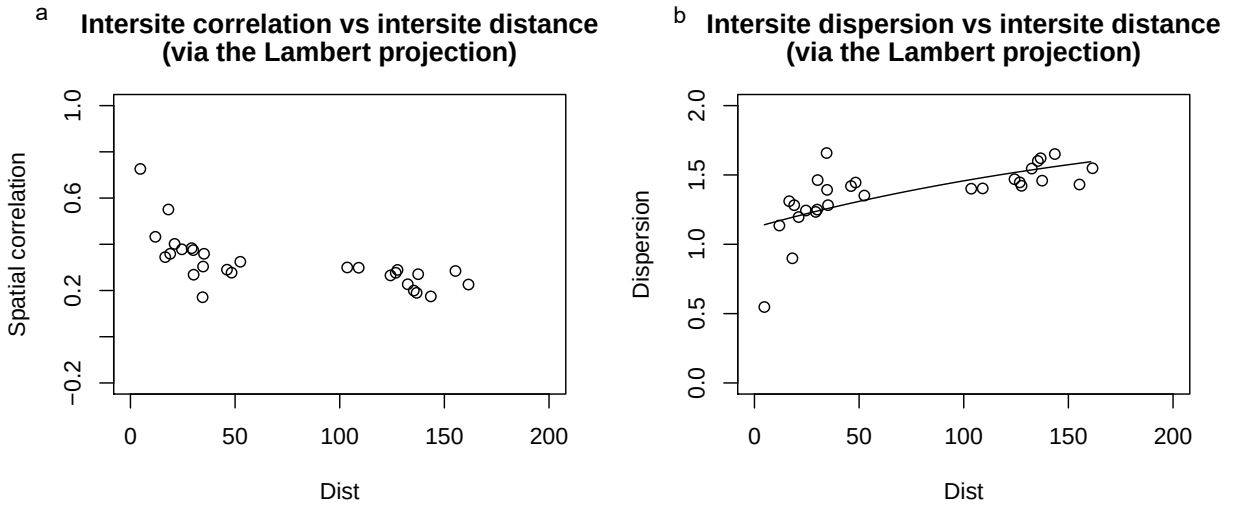


Figure 12: Spatial correlation and dispersion vs distances

Next, new potential sites are generated using the location of the existing sites. One hundred grid points are generated evenly within the minimum and the maximum range of the latitude and longitude of the existing sites. With these new locations, the hyperparameters associated with the ungauged sites are estimated with the SG method. The steps are described in the method section. First, a thin-plate spline mapping transformation is identified with function 'Falternate3'. Then, dispersion between new locations and existing sites is estimated. Next step is to estimate the hyperparameters at the new locations with 'staircase.hyper.est' function [19]. This estimates all hyperparameters for the predictive distribution and it can be used for spatial interpolation now.

Knowing the predictive distribution, its mean and variance can be derived. Interpolation can also be done with simulation. Realizations simulated from the distribution can be used to estimate the mean and covariance from the simulated data. Here, we pick day 200 as an example. 1000 realizations are simulated and contour plots for mean and covariance at hour 14:00-15:00, 15:00-16:00, 16:00-17:00 and 17:00-18:00 are generated. Figure 13 and Figure 14 show the estimated mean and variance for day 200.

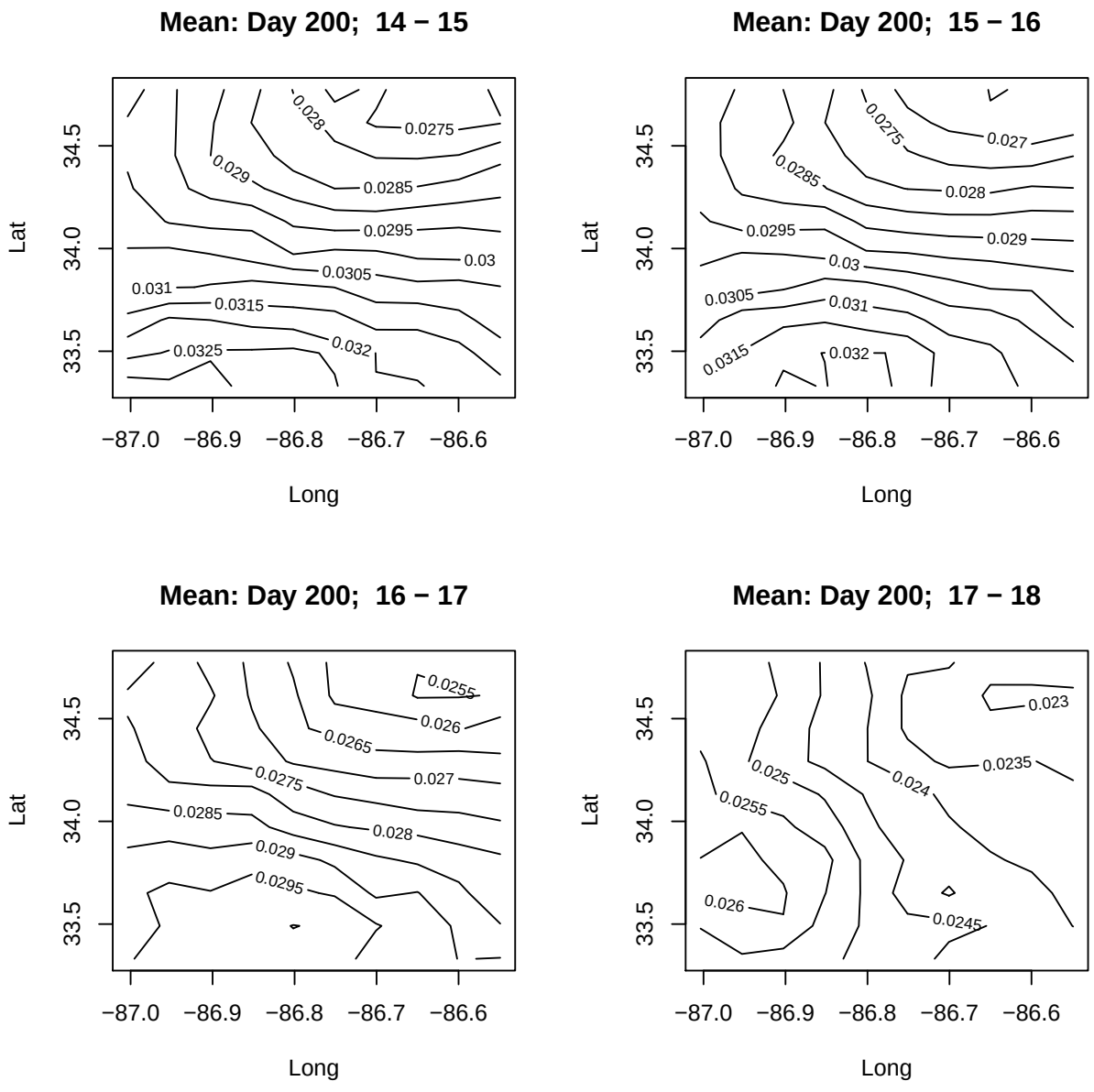


Figure 13: Contour plots of sample mean

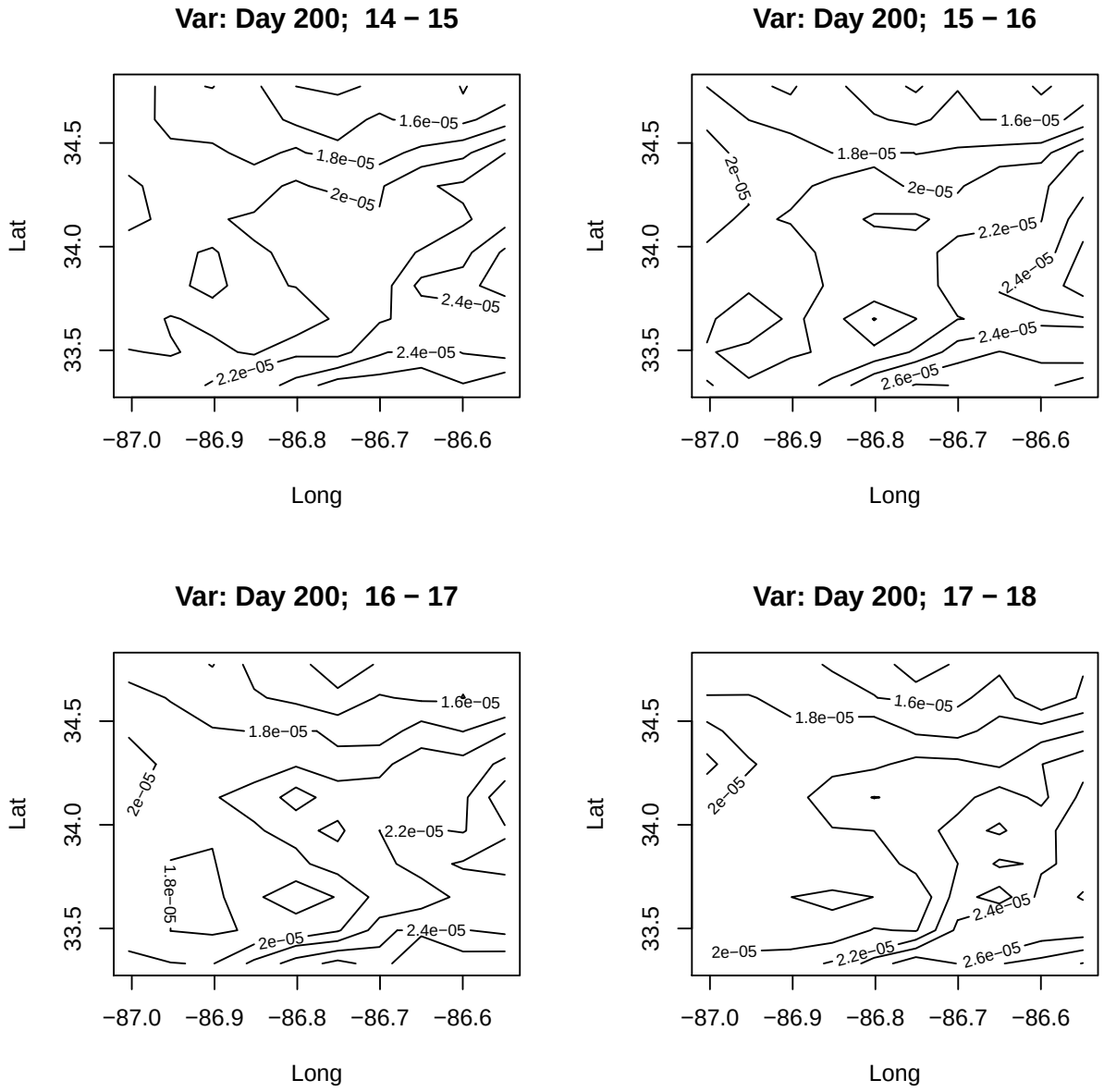


Figure 14: Contour plots of sample variance

4.2.3 Network Extension Results

Knowing the conditional covariance for the locations, we can now add the new sites into our existing network. 'ldet.eval()' function[19] evaluates the log determinants for all combinations of the number of potential sites as given in the criterion for extension (39).

Here, out of one hundred potential sites, sites number 61, 91 and 95 are added with the entropy criterion (Figure 15).

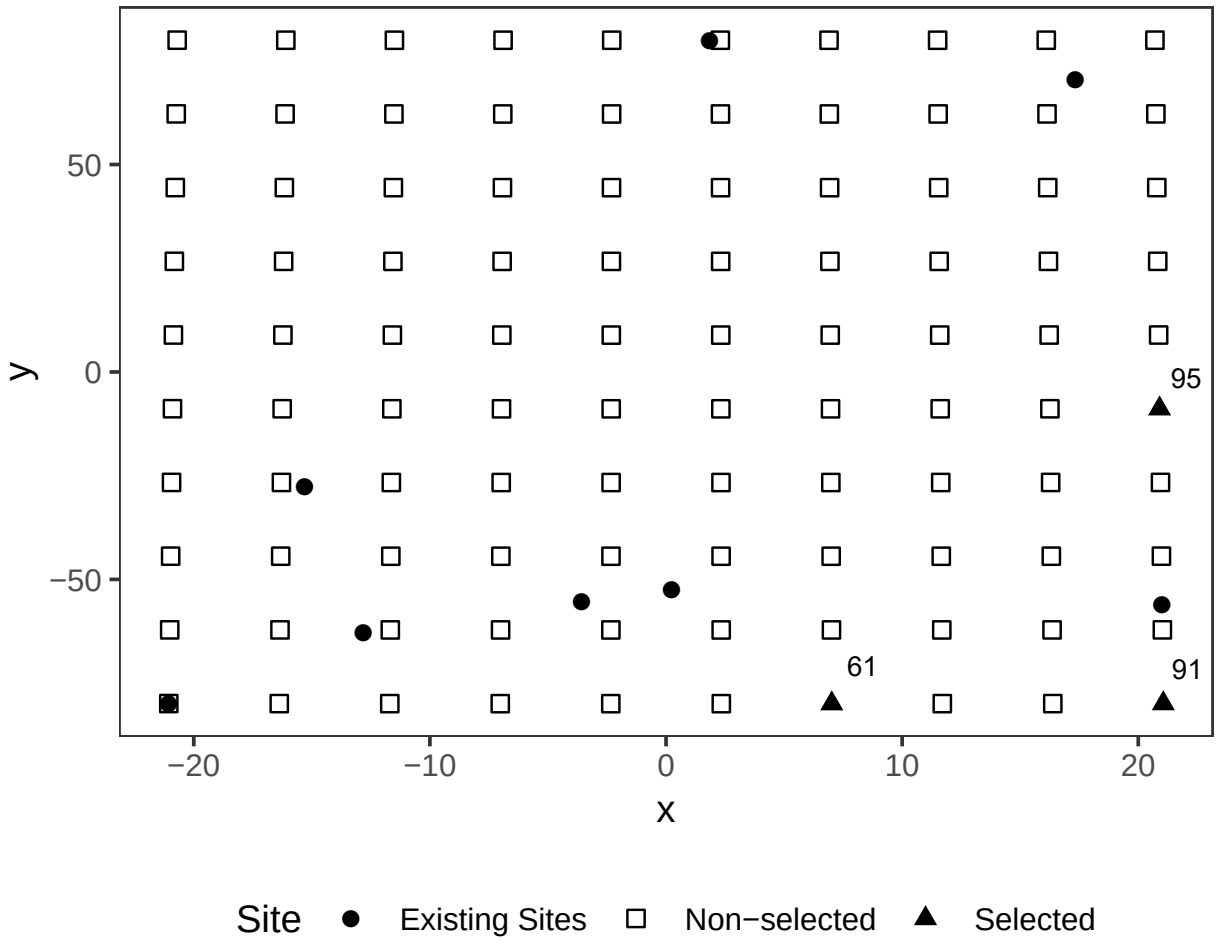


Figure 15: Three new selected sites to extend the network

The entropy criterion picked the sensors that are most uncertain about each other. From Figure 15, the new sites that get picked are all along the boundary. This shows a drawback of the entropy method. If the setting is in a closed space, placements on the boundary can waste information since sensor measurements are expected to be most precise in a region around it.

4.3 Network Extension of Santa Barbara Ozone Data [Lucy]

With the help of ENVIROSTAT package in R [19], the eight-station Santa Barbara Ozone data and 100 simulated gridpoints are used here to find new monitoring locations in network. The hourly Ozone levels in these eight sites are available from 15pm to 16pm everyday from January 1, 2018 to May 30, 2018. In this case, the values for parameters introduced in 2.2.1 are: $p = 2, l = 1, n = 151, u = 100, g = 8$.

4.3.1 Pre-processing

Considering that the larger temporal effects sometimes may influence spatial correlation, firstly, we examine the temporal trend of Ozone level by fitting a linear model with weekday and month as factors for each station separately.

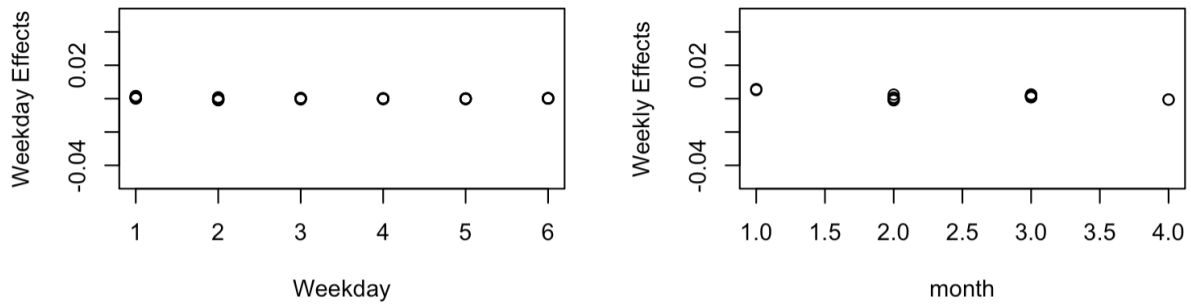


Figure 16: Estimated temporal component effects

The circle in Figure 16 represents the monitoring station. The graph shows that the weekday and month effects are small with the roughly consistent pattern for all stations. Thus, we don't need to subtract the temporal component effects before fitting the model.

Next we check normality, since in our Gaussian-GIW model (7), the multivariate response are assumed to have a joint Gaussian distribution. The results show that the normality assumptions are roughly satisfied.

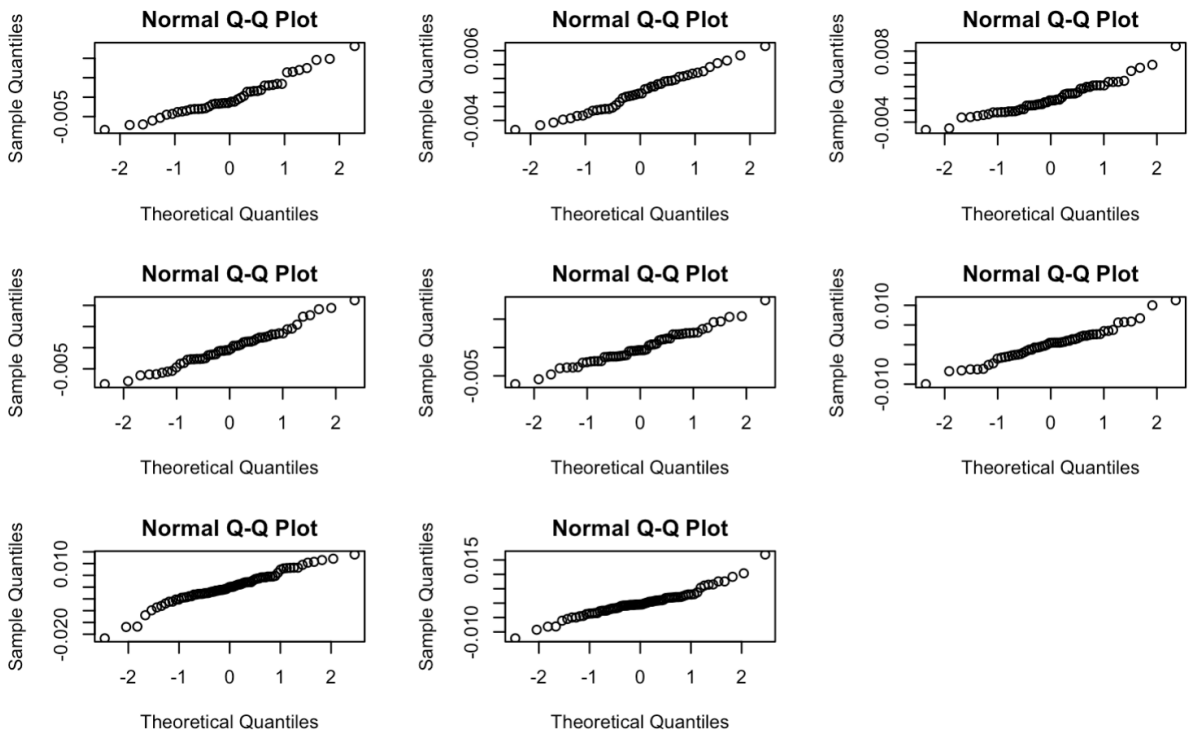


Figure 17: QQ plots of the fitted residuals for eight sites

The staircase EM algorithm requires the staircase structure for missing data. In Santa Barbara dataset, there are in total 2416 observations(151 days×2 hours×8 stations) with 52 missing observations in total. After appropriate reordering, the stations will be assigned to blocks with a staircase structure. Our data has 5 blocks with station arrangement as: Block 1: 1 station; Block 2: 2 stations; Block 3: 1 station; Block 4: 2 stations; Block 5: 2 stations. The blocks have the missing data in an descreasing staircase pattern.

Before estimating parameters for spatial predictive distribution (14), the spatial location data for all sites should be projected from the original coordinates (locations in geographical space) to a newly configured locations (flat plane space) by using Lambert project [21].

4.3.2 Hyperparameters Estimation

The Gaussian-GIW model (7) is fitted and the hyperparameters $\mathcal{H}_g$ for eight gauged sites are estimated by the staircase EM algorithm. For showing the intersite spatial correlation, we plot the spatial correlation vs distances plot and the dispersion curve vs distance plot below. Here, the dispersion is the term to distinguish the curve obtained from the spatial correlation, which doesn't assume stationarity from the variogram.

The Figure 18 shows that the further distance between sites, the lower spatial correlation, which make sense in real life. The Figure 18 shows that the dispersion vs distances curve starts near zero and rise to approach around 2, which satisfies the feature of dispersion cure

[19]. Besides, the line in this plot seems flat which means the nonstationarity here seems good.

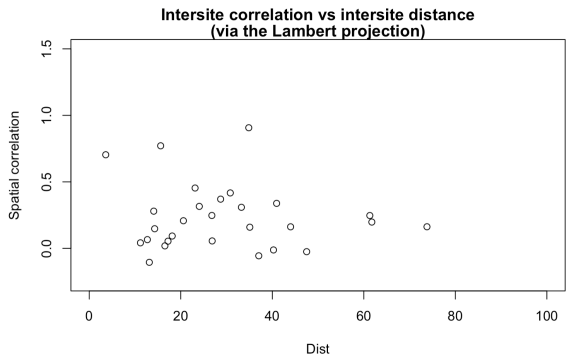


Figure 18: Intersite correlation vs intersite distance

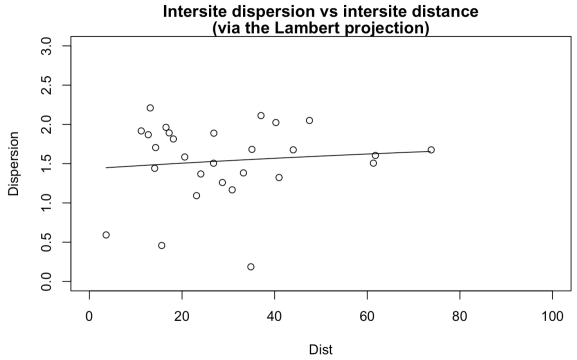


Figure 19: Intersite disperson vs intersite distance

Next the hyperparameters associated with the ungauged sites are estimated with SG method by using the 'staircase.hyper.est' function [19]. Given the predictive distribution (14), the spatial interpolation can also be done with simulation of this distribution. Here, a sample of $N = 1000$ replicates are generated for all new 100 locations, each within hour from 15:00pm to 16:00pm on day 88. The contours of the sample mean and variance surfaces are displayed in Figure 20 and 21.

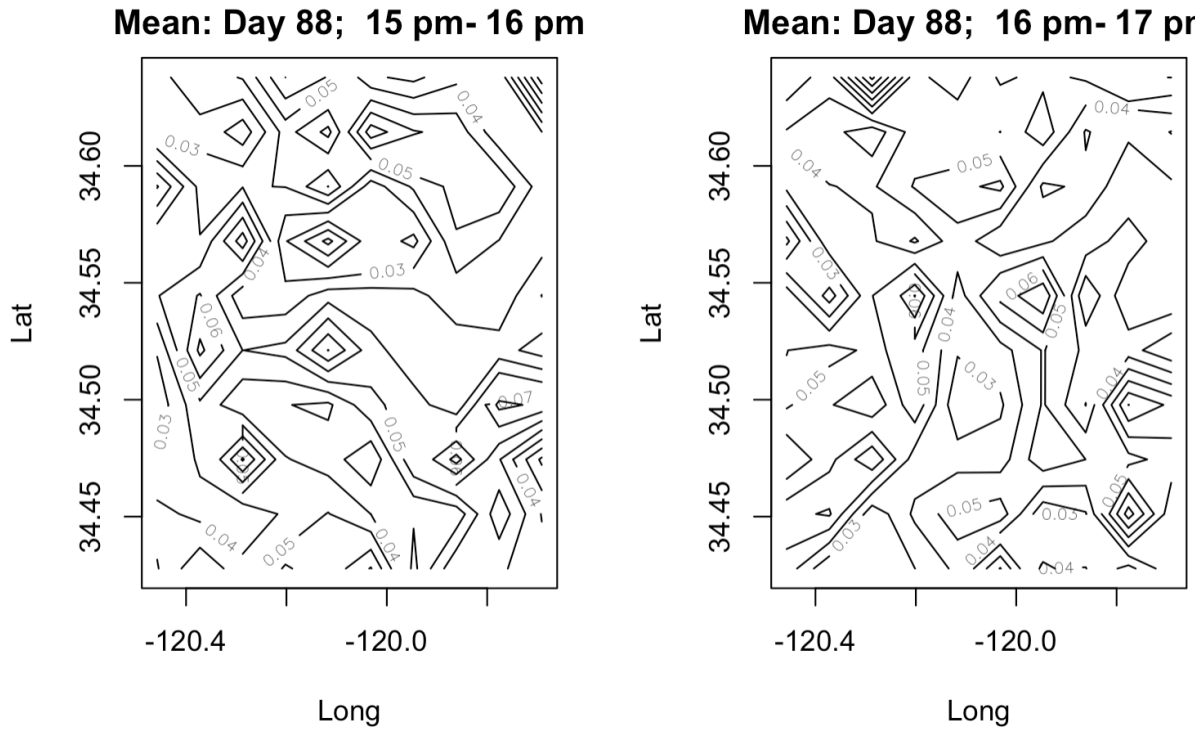


Figure 20: Contour plots of sample mean

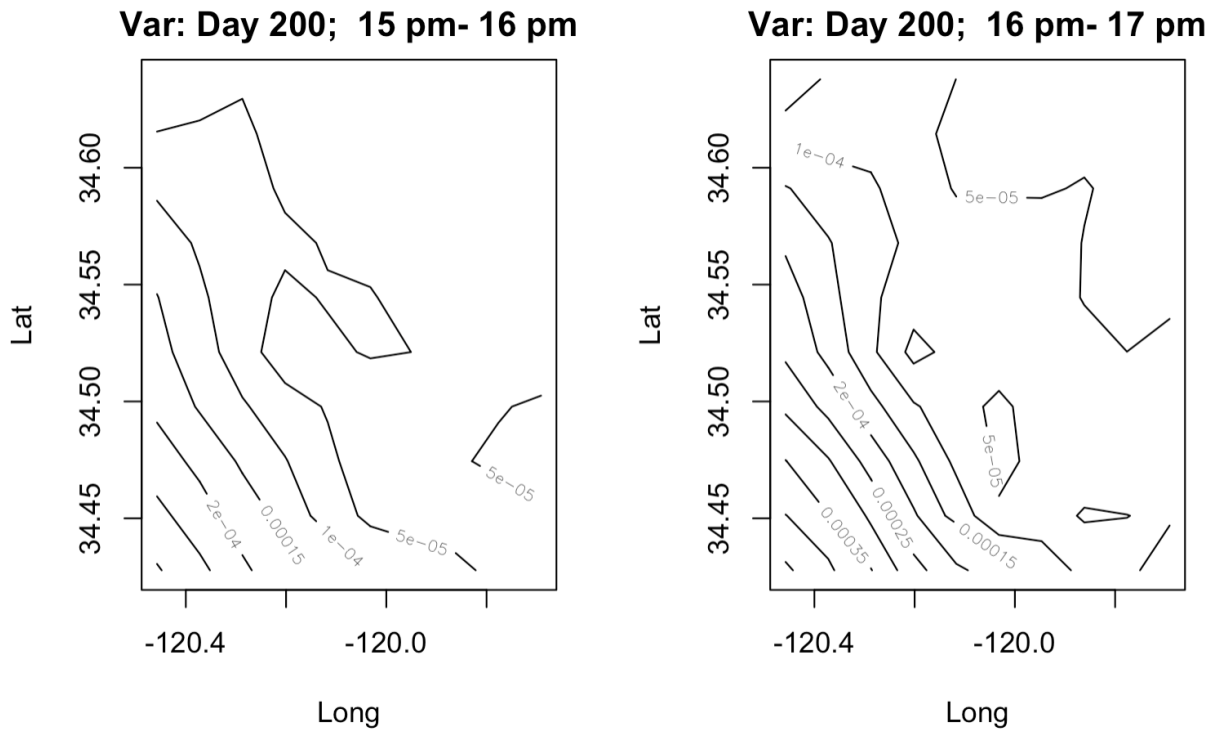


Figure 21: Contour plots of sample variance

4.3.3 Network Extension Results

Given the predictive distribution (14) of Ozone levels, new sites can be added to the existing network with an entropy criterion for network extension. Here, we use 'ldet.eval()' function [19] to select 3 new sites among 100 new locations from the above grid.

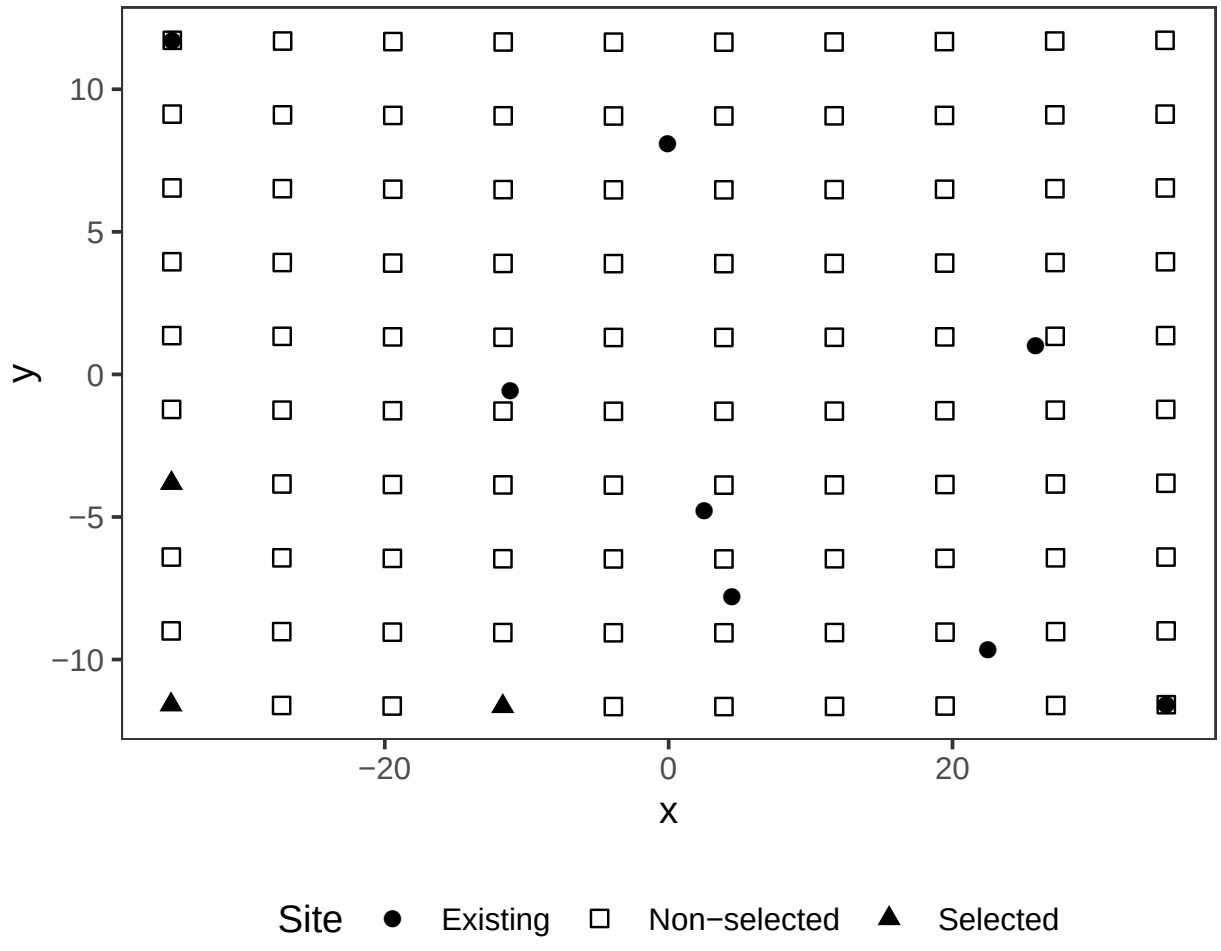


Figure 22: Three new selected sites to extend the network in Santa Barbara

Here, we can detect that the new stations are located in the bottom left corner in the graph, which makes sense since the most of existing gauged sites are located in middle and lower right corner of the graph. The new stations can monitor observations which can not be detected by existing stations. However, it may be surprising that the new sites are all located along the boundary, which has the same phenomenon as LA ozone data we have analyzed before. We can speculate that the entropy method prefer boundary location if the setting is in a closed space.

4.4 The p-Dispersion [Jing]

The nine-station NY data and 100 simulated gridpoints are used here. All empirical results are obtained using a desktop personal computer (Intel Xeon E5 CPU 2.30 GHz with 64 GB RAM and 4 cores).

4.4.1 The p-Dispersion Network Design

The p-dispersion network design that selects p sites among the 9 candidate NY location with p varying from 2 to 9 is evaluated (Figure 23). When selecting two sites, the two farthest points, Site 2 and 9, are selected with a objective value of 220.372. As p increases, the objective function decreases because it is harder to spread out sites with more and more sites. Also, the the decrease is marginal with large p. Figure 24 shows the selected three sites (Sites 1, 6 and 9) that can maximize the minimum distance among selected sites. Note that one characteristic of the p-dispersion optimal solution is that the selected sites tend to locate near the boundary of the study area.

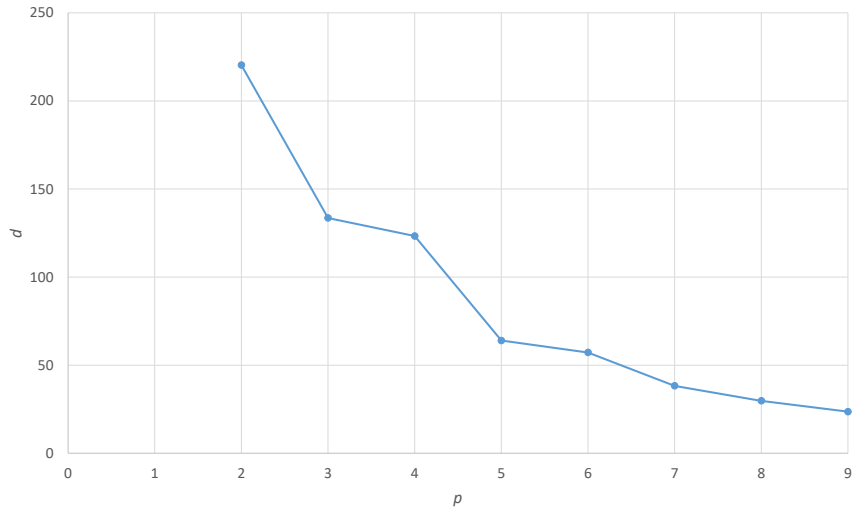


Figure 23: Tradeoff curve of d vs p

4.4.2 The p-Dispersion Network Extension

The p-dispersion network extension that adds p sites among the 100 gridpoints with p varying from 1 to 10 is evaluated (Figure 25). When $p = 1$, Site 36 is selected such that the minimum distance between Site 36 and existing 9 monitor stations is maximized. Similarly, increasing number of sites to select decreases the objective function. Figure 26 shows the optimal configuration of adding 3 stations to the existing network. Sites 11, 35 and 100 among the given 100 gridpoints are selected with a objective value of 62.355.

4.4.3 Cost

The p-dispersion network design with the second objective, minimizing the cost, with p varying from 1 to 9 is evaluated. Here, the 9 sites are given some random costs (f =

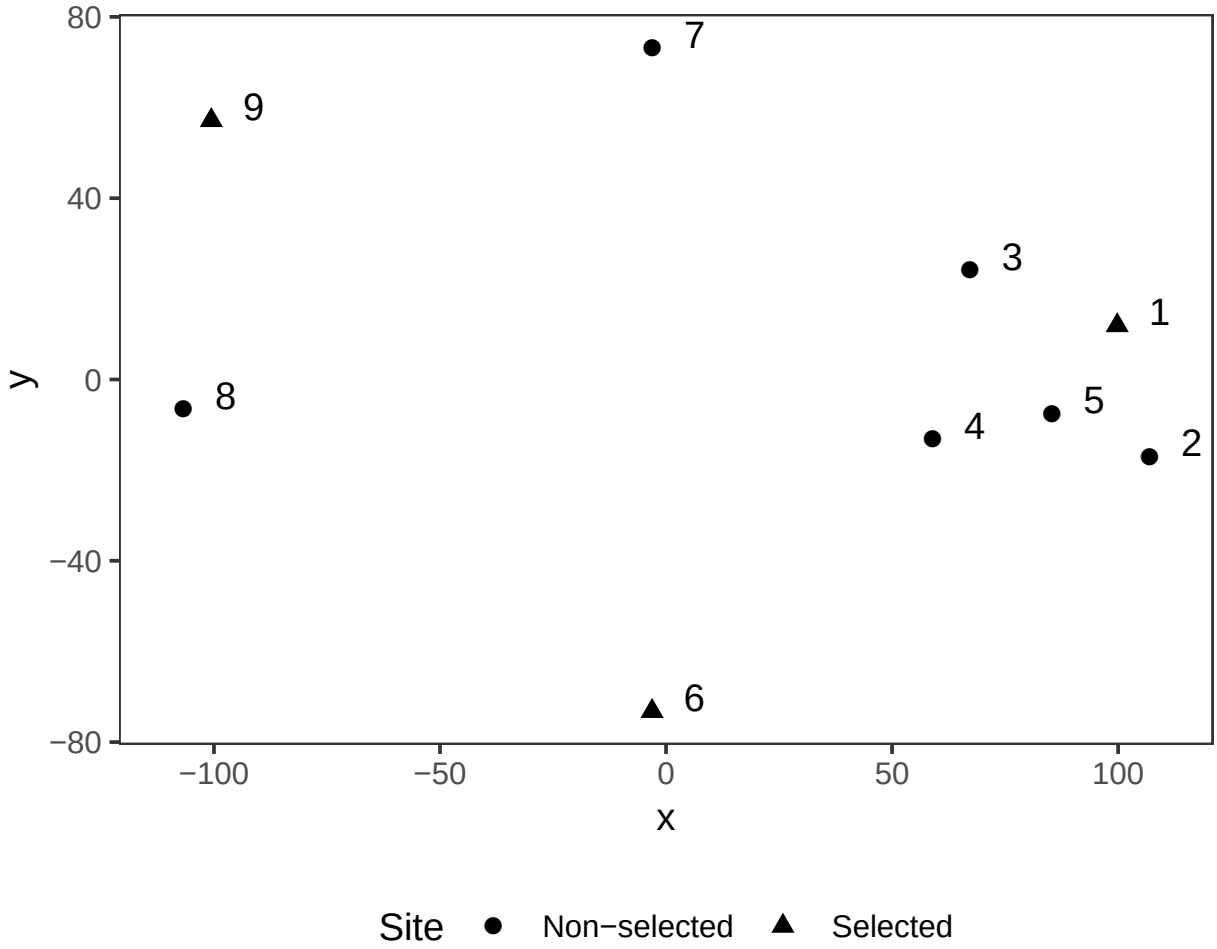


Figure 24: The p-dispersion network design result (p=3)

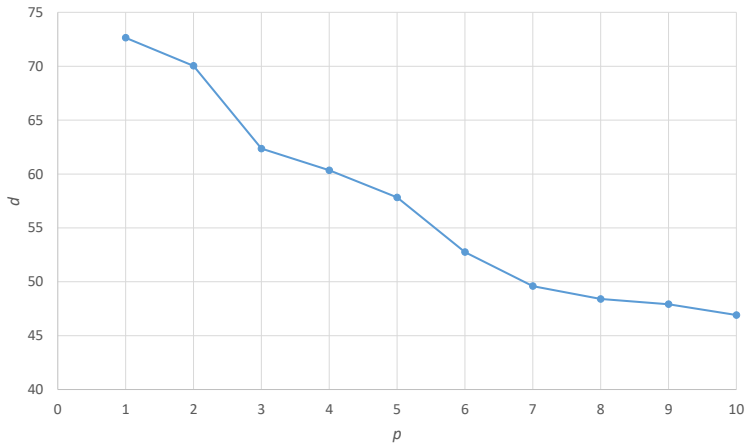


Figure 25: Tradeoff curve of d vs p

$(92, 72, 89, 25, 64, 95, 21, 37, 81))$ and α varies from 0 to 2.0 with an interval of 0.2. In total, $9 \times 11 = 99$ problems are computed. Selected results are shown in Table 1.

Table 1: Selected results of the p-dispersion with cost

α	$d - \alpha\Sigma_{i\in I} f_i Z_i$	d	$\Sigma_{i\in I} f_i Z_i$	Selected sites
0	133.534	133.534	268	1, 6, 9
0.2	104.837	130.837	130	2, 7, 8
0.4	78.8368	130.837	130	2, 7, 8
0.6	56.4173	106.217	83	4, 7, 8
0.8	39.8173	106.217	83	4, 7, 8
1.0	23.2173	106.217	83	4, 7, 8

The optimal network is impacted by the cost variability across different sites. When $\alpha = 0.2$, Sites 2, 7 and 8 are selected giving $d = 130.837$ rather than the Sites 1, 6 and 9 without considering the cost (i.e. $\alpha = 0$). Because the total cost of selecting Sites 1, 6 and 9 is 268, much higher than the cost of selecting Sites 2, 7 and 8 that is only 130. Similarly, the site configuration is different when α changes from 0.4 to 0.6. Increasing the importance of the cost in the planning process (i.e. α here) shifts the selected sites to sites with relatively low costs.

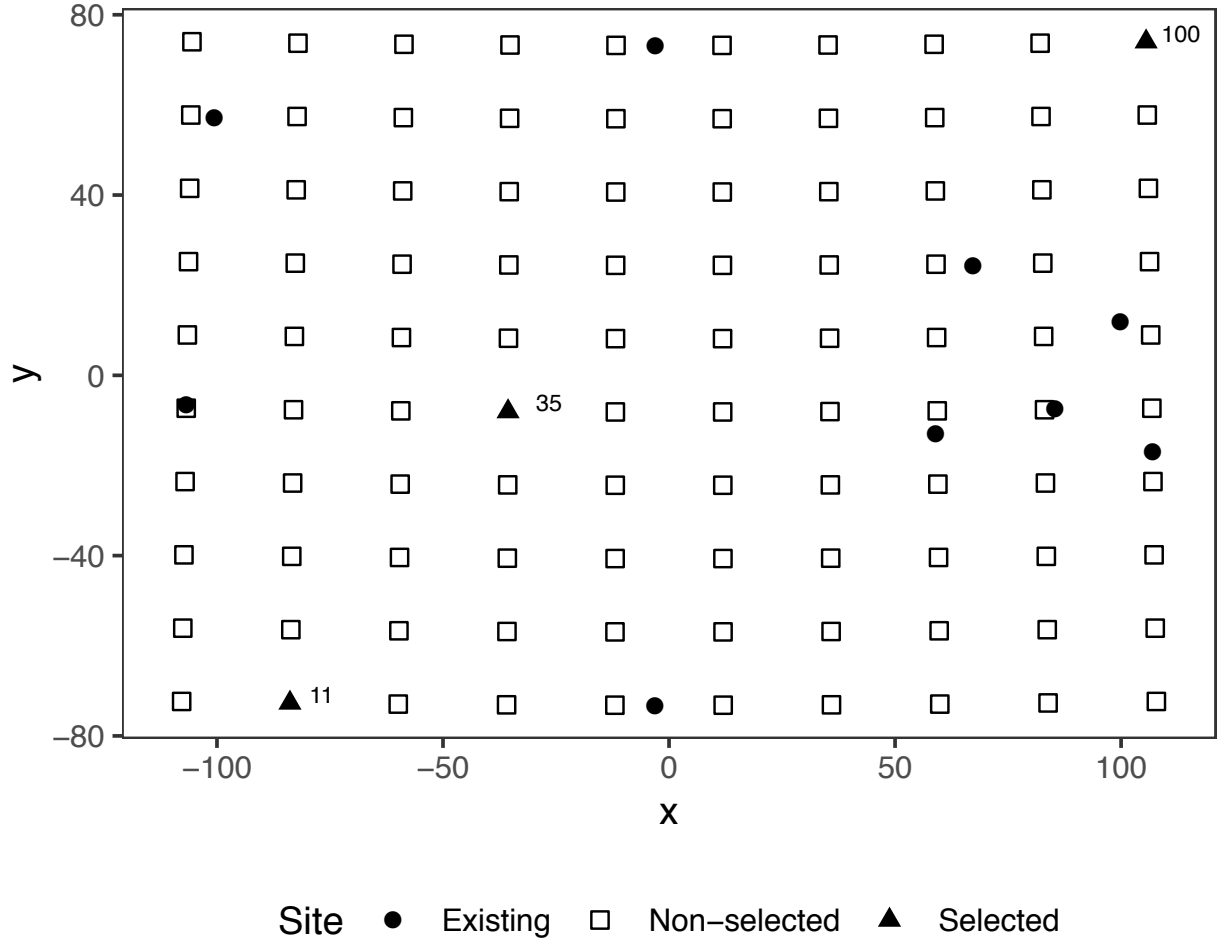


Figure 26: The p-dispersion network extension result (p=3)

5 Conclusion

In analyzing two prediction-based techniques, we generated both first-phase and second-phase sampling configurations for New York City and Los Angeles. We found out that simultaneous addition of three stations acquired via *geospt* tends to converge in the same geographical location for both for two study areas. This, as discussed, could be the result of the small number of sensors and the corresponding low variation of distances in fitted variogram functions. On the other hand, *spsann* only allowed first-phase sampling, so we visually compared enhanced generated networks to the existing ones. Overall, *spsann* favored centrally clustered samples.

Multivariate warping based method deals with multivariate responses with staircase missing data. Considering that not all of stations started its monitoring at the same time, so the data matrix has a lot of missing data. Filling in the missing values with real measurements would be a method but it still bring measurement error for our estimation. The staircase pattern of missing data can make sure that the each step of the staircase consists of stations with the same starting time, which reduces the uncertain information of missing data. The generalizecd inverted Wishart distribution is very flexible to deal with the staircase structure of the observed data [21]. Given the Gaussian-GIW model, the method then estimate the hyperparameters for the predictive distribution with EM and SG methods, then extend the existing network with entropy criterion. Entropy criterion is intuitive since it finds the sensors that are most uncertain about each other. One problem of this is it often finds the sites far apart at the boundary, which can cause a waste of information. Using mutual information criterion instead can be an improvement of the entropy criterion[15].

The p-dispersion network design locates p monitors such that they are as dispersed as possible. The empirical experiments show that the selected sites tend to locate near the boundary of the study region. The p-dispersion network design can be easily modified to conduct network extension and consider the spatial heterogeneity of costs. This method can provide planning strategies when it is hard to obtain prior knowledge or appropriate model-based criterion. Thus, this method is appealing when designing a brand new network without any prior knowledge. In addition, the p-dispersion network design is usually easier to solve compared to complex model-based network design. On the other hand, the primary limitation of this method is that it only considers the spatial location of candidate sites but does not utilize the spatial and temporal structure of the observed measures at the existing sites if any. The p-dispersion model is a typical spatial optimization model and can be conveniently extended to incorporate other objectives and constraints such as must covering the populated area and the minimal number of sites in each sub-region. Efficient heuristics such as simulated annealing, genetic algorithm and Lagrangian relaxation might be necessary for large instances of the p-dispersion model and modified complexed models.

References

[1] Jennifer A. Hoeting Richard L. Smith Alan Gelfand, Montse Fuentes, editor. *Handbook of Environmental and Ecological Statistics*.

[2] David R Anderson, Dennis J Sweeney, Thomas A Williams, Jeffrey D Camm, and James J Cochran. *An introduction to management science: quantitative approaches to decision making*. Cengage learning, 2015.

[3] DJ Brus, JJ De Gruijter, and JW Van Groenigen. Designing spatial coverage samples using the k-means clustering algorithm. *Developments in Soil Science*, 31:183–192, 2006.

[4] John Current, Hokey Min, and David Schilling. Multiobjective analysis of facility location decisions. *European Journal of Operational Research*, 49(3):295–307, 1990.

[5] Eric M Delmelle. Spatial sampling. *Handbook of regional science*, pages 1385–1399, 2014.

[6] Eric M Delmelle and Pierre Goovaerts. Second-phase sampling designs for non-stationary spatial variables. *Geoderma*, 153(1-2):205–216, 2009.

[7] Simone Di Zio, Lara Fontanella, and Luigi Ippoliti. Optimal spatial sampling schemes for environmental surveys. *Environmental and Ecological Statistics*, 11(4):397–414, 2004.

[8] Erhan Erkut and Susan Neuman. Comparison of four models for dispersing facilities. *INFOR: Information Systems and Operational Research*, 29(2):68–86, 1991.

[9] Reza Zanjirani Farahani, Maryam SteadieSeifi, and Nasrin Asgari. Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling*, 34(7):1689–1709, 2010.

[10] Wabba G and Wendelberger J. Some new mathematicl methods for variational objective analysis using splines and cross-validatoin. *Monthly Weather Rev*, 108:36–57, 1980.

[11] RS Garfinkel and GL Nemhauser. Integer programming. 1972. *John Wiley&Sons, New York*.

[12] Kibria GBM, Sun L, JV Zidek, and Le ND. Bayesian spatial prediction of random space-time fields with application to mapping pm2.5 exposure. *J Amer Statist Assoc*, 457:101–112, 2002.

[13] Mark E Johnson, Leslie M Moore, and Donald Ylvisaker. Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148, 1990.

[14] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[15] Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, pages 235–284, 2008.

[16] A. M. Kshirsagar. Bartlett decomposition and wishart distribution. *Ann. Math. Statist.*, 30(1):239.

[17] Michael J Kuby. Programming models for facility dispersion: The p-dispersion and maxisum dispersion problems. *Geographical Analysis*, 19(4):315–329, 1987.

[18] Mardia KV, Kent JT, and Bibby J. *Multivariate Analysis*. London: Academic, 1979.

[19] Nhu Le, Jim Zidek, Rick White, and Davor Cubranic. Envirostat: Statistical analysis of enviromental space-time processes. https://cran.r-project.org/web/packages/EnviroStat/index.html, 2015.

[20] Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[21] Jorge Mateu[U+FF0C] Werner G. Müller. *Spatio-temporal Design: Advances in Efficient Data Acquisition*. WILEY, 2012.

[22] Le ND and Zidek JV. Network designs for monitoring multivariate random spatial fields. pages 191–206, 1994.

[23] Le ND, Sun L, and JV Zidek. Spatial prediction and temporal backcasting for environmental fields having monotone data patterns. *Can J Statist*, 29:516–529, 2001.

[24] Sampson P and P Guttorp. nonparametric estimation of nonstationary spatial structure. *J Amer stat Assoc*, 87:108–119, 1992.

[25] Giovanni Pistone and Grazia Vicario. Comparing and generating latin hypercube designs in kriging models. *AStA Advances in Statistical Analysis*, 94(4):353–366, 2010.

[26] Luc Pronzato and Werner G Müller. Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3):681–701, 2012.

[27] Peter A Rogerson, Eric Delmelle, Rajan Batta, Mohan Akella, Alan Blatt, and Glenn Wilson. Optimal sampling design for variables with varying spatial importance. *Geographical Analysis*, 36(2):177–194, 2004.

[28] J Andrew Royle and Doug Nychka. An algorithm for the construction of spatial coverage designs with implementation in splus. *Computers & Geosciences*, 24(5):479–488, 1998.

[29] Alessandro Samuel-Rosa. geospt: Geostatistical analysis and design of optimal spatial sampling networks. https://cran.r-project.org/web/packages/geospt/geospt.pdf, 2019.

[30] Alessandro Samuel-Rosa. spsann: Optimization of sample configurations using spatial simulated annealing. https://cran.r-project.org/web/packages/spsann/index.html, 2019.

[31] Kathrin Sch¨acke. On the kronecker product. https://www.math.uwaterloo.ca/~hwolkowi/henry/reports/kronthesisschaecke04.pdf, 2013.

[32] Jan Willem Van Groenigen, W Siderius, and A Stein. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma*, 87(3-4):239–259, 1999.

[33] JW Van Groenigen. Spatial simulated annealing for optimizing sampling. In *geoENV I—Geostatistics for Environmental Applications*, pages 351–361. Springer, 1997.

[34] JW Van Groenigen, G Pieters, and A Stein. Optimizing spatial sampling for multivariate contamination in urban areas. *Environmetrics: The official journal of the International Environmetrics Society*, 11(2):227–244, 2000.

[35] Sun W, Le ND, Zidek JV, and Burnett R. Assessment of bayesian multivariate interpolation approach for health impact studies. *Envirometrics*, 9:445–457, 1998.

[36] Laurence A Wolsey and George L Nemhauser. *Integer and combinatorial optimization*, volume 55. John Wiley & Sons, 1999.

[37] Evangelos A Yfantis, George T Flatman, and Joseph V Behar. Efficiency of kriging estimation for square, triangular, and hexagonal grids. *Mathematical Geology*, 19(3):183–205, 1987.

[38] Nhu D. Le James V. Zidek. *Statistical Analysis of Environmental Space-Time Processes*. Springer Science+Business Media, 2006.