Monitoring Network Design

Zhuowei Cheng, Lucy Lu, Evgeny Noi, Jing Xu

1 Introduction

As a continuation of our midterm project, this final project studies three different network design methods, non-adaptive, adaptive sampling design and preferential sampling. In some conventional geostatistical models, the locations used for inferring the spatial process are assumed to be noninformative. This assumption can be violated in many situations, for example, collecting data at locations thought to have a large or small value. Preferential sampling arises when the process that determines the data locations and the process being modelled are stochastically dependent. Diggle et al.[4] and Pati et al.[15] proposed two models to adjust for the bias caused by the informative sampling locations. Adaptive sampling is a sampling method that collects data in batches and the current batch is dependent on available data from previous batches. Important components and general steps of adaptive design are discussed followed by a specific adaptive sampling algorithm. One simulated experiment and one empirical study are provided to demonstrate how adaptive sampling works. In contrast to adaptive designs, non-adaptive designs imply a first-phase sampling campaigns, where the design of the network has to be selected prior to taking measurements. Three broad classes of exist to first-phase sampling, including complete random design, complete regular design and hybrid designs. To this end, Chipeta et al. [1] add inhibitory class of designs, which we further review and test on real-world data.

2 Methods

2.1 Non-Adaptive Geostatistical Designs [Evgeny]

Let $\mathcal{X}$ denote the desired sampling configuration. Then, in order to infer the unobserved spatial phenomenon $S = \{S(x) : x \in \mathcal{D} \subset \mathbb{R}\}$, we need to utilize known data measurements y_i in finite set of locations $x_i \in \mathcal{D}$ and optimize model parameters for both stochastic estimation and prediction at unmeasured locations, as well as a performance criterion inherent in the objective of the particular study.

Non-adaptive design implies that $\mathcal{X}$ must be chosen prior to any data gathering. Most common forms of non-adaptive designs are lattice and completely random designs (CRS). Generally, when model parameters are known, the lattice configurations enable efficient spatial prediction, albeit with a certain possibility of systematic bias. On the other hand, CRS has a better chance of incorporating wider range of inter-point distances, which provides a better estimate for the form of covariance function. The tradeoff is that complete randomness produces a spatial distribution that hinders efficient prediction on data. For these reasons,

some of the recent research on network sampling has been focused on hybrid approaches, which ensures comparatively even spatial spacedness between points by 1) imposing a restricted (hierarchical) randomization [18]; 2) pivotal method that restricts locations that are close to one another from appearing together in a sample [10]; 3) lattice-plus close pairs and lattice-plus infill design, in which a two-step process is used to first generate a regular lattice over the study area, and then supplement sample with additional points from the uniform distribution or simply more refined lattice [5]; 4) purely geometric (space-filling) algorithms [16].

2.1.1 Inhibitory Geostatistical Design

Chipeta et al. [1] propose a new class of inhibitory geostatistical design for spatial prediction, when model parameters are unknown. We put a stochastic process $\mathcal{S} = \{S(x) : x \in \mathcal{D} \subset \mathbb{R}\}$ into a statistical model $[S, Y] = [S][Y|S]$, where $Y = (Y_1, ..., Y_n)$ are the measured data values and $S = \{S(x_1), ..., S(x_n)\}$. Then following the Bayes theorem, we can infer estimates via conditional distribution:

$$[S|Y] = [S][Y|S] / \int [S][Y|S] ds \tag{1}$$

To assess performance of the model the authors estimate average prediction variance:

$$APV = \int_D Var\{S(x)|Y\} dx \tag{2}$$

2.1.2 Simple Inhibitory Design

In a simple case, we randomly select n locations from $\mathcal{D}$ within a distance δ to one another, where the resulting design $\mathcal{X}$ is a realization of pairwise interaction point process and all designs of $\mathcal{X}$ are equally likely to be picked. The sample space is controlled through *'packing density'*, which defines proportion of the region covered by n non-overlapping disks of diameter (δ):

$$\rho = \frac{n\pi\delta^2}{4|\mathcal{D}|} \tag{3}$$

The simple inhibitory design is defined as $\mathbf{SI}(n, \delta)$ and is drawn in the following order:

Step 1. Draw a sample of locations $x_i : i = 1, ..., n$ completely random in $\mathcal{D}$;
Step 2. Set $i = 1$;
Step 3. Calculate the minimum, d_{min}, of the distance from x_i to all x_j in the current sample;
Step 4. if $d_{min} \geq \delta$, increase i by 1 and return to step 3 if $i \leq n$, otherwise stop;
Step 5. if $d_{min} < \delta$, replace x_i by a new location drawn completely at random in $\mathcal{D}$ and return to **Step 4**.

2.1.3 Inhibitory Design With Close Pairs

In addition to δ (minimum distance between two points) and n (the total number of points), we add two additional controlling factors: k (the number of close pairs), and ζ (radius of the disk from the primary point within which to add a paired point). The design of $\mathbf{ICP}(n, k, \delta, \zeta)$ follows the outlined algorithm:

Step 1. Construct a simple inhibitory design $\mathbf{SI}(n-k, \delta)$;
Step 2. Sample k from $x_1, ..., x_{n-k}$ without replacement and call this set of locations x_j^*, $j = 1, ..., k$;
Step 3. For $j = 1, ..., k, x_{n-k+j}$ is uniformly distributed on the disk with center x_j^* and radius ζ.

Potential sampling locations are restricted to a finite set of points $\{X_i, i = 1, ..., N\}$. To ensure the same degree of spatial regularity, we must ensure the following conditions are met:

$$k \leq n/2 \tag{4}$$

$$\delta_{(k)} = \delta_0 \sqrt{n/(n-k)} \tag{5}$$

$$\zeta \leq \delta_{(k)}/2 \tag{6}$$

Finally, general purpose numerical optimizer known as control random search procedure is used to impose box constraints on the design parameters.

2.2 Adaptive Sampling [Jing]

An adaptive sampling collects data in batches and the current batch would be dependent on the available data from previous batches [6]. A formal representation of adaptive sampling is $\boldsymbol{X} = \{\boldsymbol{X_0}, \boldsymbol{X_1}, \ldots, \boldsymbol{X_n}\}$ where suffix 0 indicates an initial design and the rest suffices represent the batch numbers. The corresponding outcome is $\boldsymbol{Y} = (\boldsymbol{Y_0}, \boldsymbol{Y_1}, \ldots, \boldsymbol{Y_n})$. The outcomes of later batches $\boldsymbol{Y_k}$ $(k > 0)$ are affected by their sampling location $\boldsymbol{X_k}$, and thus in turn are influenced by earlier batch outcomes. There is a stochastic interplay between $\boldsymbol{Y_k}$ and $\boldsymbol{X_k}$, which will be introduced in Section 2.2.2. Same as the non-adaptive sampling, the selection criterion C is critical to the sampling design.

An important element in adaptive sampling is the batch size b. Two primary adaptive sampling methods are singleton sampling (when $b = 1$) and batch adaptive sampling (when $b > 1$) based on the batch size. In singleton sampling, $\boldsymbol{X_k} = (x_k)$ with only one location is added sequentially. x_{k+1} is dependent on data obtained at $x_0, x_1, \ldots, x_k$. Assume the selection criterion is to maximize a given C, x_{k+1} that maximizes C given the data at $x_0, x_1, \ldots, x_k$ is added in a singleton sampling. In batch adaptive sampling, $\boldsymbol{X_k} = (x_{k1}, x_{k2}, \ldots, x_{kb})$ with b location $(b > 1)$ are added at each stage. A naive extension of the selection in singleton sampling is to choose b locations that can maximize C. This tends to select a tight cluster of b locations that can maximize C more than any single of them since the corresponding underlying Gaussian process $(S(x))$ are highly correlated.

Next, important components of adaptive sampling and an algorithm of adaptive sampling will be discussed, which draw heavily on Chipeta et al. [2].

2.2.1 Geostatistical Model for Prevalence Data

Let Y_i be the number of positive outcomes out of n_i individuals tested at location x_i in a region $D \subset \mathbf{R}^2$ and $d(x_i) \in \mathbf{R}^p$ be a vector of associated covariates. The standard geostatistical model for prevalence data [7] assumes that $Y_i \sim Binomial(n_i, p(x_i))$ where $p(x_i)$ is the prevalence of disease at a location $x(i)$. Using the logit link function, the model assumes that

$$log\frac{p(x)}{1-p(x)} = d(x)'\beta + S(x) \tag{7}$$

where $S(x)$ is a stationary Gaussian process with zero mean, variance σ^2 and correlation function $\rho(u) = Corr\left\{(S(x), S(x')\right\}$ and u is the distance between x and x'.

A Matérn correlation structure [14] for $S(x)$ is assumed:

$$\rho(u;\phi;\kappa) = \left\{2^{\kappa-1}\Gamma(\kappa)\right\}^{-1}\left(\frac{u}{\phi}\right)^{\kappa}K_{\kappa}\left(\frac{u}{\phi}\right) \tag{8}$$

where $\phi > 0$ is a scale parameter that controls the rate at which correlation decays with increasing distance, $K_\kappa(\cdot)$ is a modified Bessel function of order $\kappa > 0$ and $S(x)$ is m times mean-sqaure differentiable if $\kappa > m$.

Monte Carlo methods can be used to fit the standard model but they are computationally intensive. Fortunately, software implementations are available such as R package **PrevMap** [9].

2.2.2 Likelihood-based Inference

In adaptive design, $\boldsymbol{X}$ and S are not independent but are conditionally independent given $\boldsymbol{Y}$. This simplifies the likelihood function. To show it, define $\boldsymbol{X_0}$ as the initial sampling design selected independently of S and $\boldsymbol{Y_0}$ is the corresponding outcome. Denote $\boldsymbol{X_1}$ the set of additional sampling design added considering the existing initial design and measured data pair $(\boldsymbol{X_0}, \boldsymbol{Y_0})$, $\boldsymbol{Y_1}$ is the resulting outcome associated with $\boldsymbol{X_1}$ and so on. After k stages, the complete data set is $\boldsymbol{X} = \boldsymbol{X_0} \cup \boldsymbol{X_1} \cup \cdots \cup \boldsymbol{X_k}$ and $\boldsymbol{Y} = (\boldsymbol{Y_0}, \boldsymbol{Y_1}, \ldots, \boldsymbol{Y_k})$. Let $[\cdot]$ be "the distribution of", the associated likelihood for the complete data set is

$$[\boldsymbol{X}, \boldsymbol{Y}] = \int_S [\boldsymbol{X}, \boldsymbol{Y}, S]dS \tag{9}$$

Consider a simple case with $k = 1$ which means only one stage after the initial sampling. The factorization of any multivariate distribution gives

$$[\boldsymbol{X}, \boldsymbol{Y}, S] = [S, \boldsymbol{X_0}, \boldsymbol{Y_0}, \boldsymbol{X_1}, \boldsymbol{Y_1}] = [S][\boldsymbol{X_0}|S][\boldsymbol{Y_0}|\boldsymbol{X_0}, S][\boldsymbol{X_1}|\boldsymbol{Y_0}, \boldsymbol{X_0}, S][\boldsymbol{Y_1}|\boldsymbol{X_1}, \boldsymbol{Y_0}, \boldsymbol{X_0}, S] \tag{10}$$

On the right hand side of Eq (10), by construction, $[\boldsymbol{X_0}|S] = [\boldsymbol{X_0}]$ and $[\boldsymbol{X_1}|\boldsymbol{Y_0}, \boldsymbol{X_0}, S] = [\boldsymbol{X_1}|\boldsymbol{Y_0}, \boldsymbol{X_0}]$. Thus, Eq (9) is

$$[\boldsymbol{X}, \boldsymbol{Y}] = [\boldsymbol{X_0}][\boldsymbol{X_1}|\boldsymbol{X_0}, \boldsymbol{Y_0}] \times \int_S [\boldsymbol{Y_0}|\boldsymbol{X_0}, S][\boldsymbol{Y_1}|\boldsymbol{X_1}, \boldsymbol{Y_0}, \boldsymbol{X_0}, S] dS \tag{11}$$

The first term on the right hand side of Eq (11) is the conditional distribution of $\boldsymbol{X} = \boldsymbol{X_0} \cup \boldsymbol{X_1}$ given $\boldsymbol{Y_0}$, i.e. $[\boldsymbol{X}|\boldsymbol{Y_0}]$. The second term simplifies to

$$[\boldsymbol{Y_0}|\boldsymbol{X_0}][\boldsymbol{Y_1}|\boldsymbol{X_1}, \boldsymbol{Y_0}, \boldsymbol{X_0}] = [\boldsymbol{Y_0}, \boldsymbol{Y_1}|\boldsymbol{X_0}, \boldsymbol{X_1}] = [\boldsymbol{Y}|\boldsymbol{X}] \tag{12}$$

Combing Eq (11) and Eq (12)

$$[\boldsymbol{X}, \boldsymbol{Y}] = [\boldsymbol{X}|\boldsymbol{Y_0}][\boldsymbol{Y}|\boldsymbol{X}] \tag{13}$$

Eq (13) shows that the conditional likelihood $[\boldsymbol{Y}|\boldsymbol{X}]$ can be used for inference though it may be inefficient depending on how $[\boldsymbol{X}|\boldsymbol{Y_0}]$ is specified. Eq (13) can be extended to $k > 1$.

2.2.3 A General Class of Adaptive Design

Step 1. Use a non-adaptive design to choose an initial set of sample locations, $\boldsymbol{X_0} = \{x_i \in D : i = 1, 2, \ldots, n_0\}$.

Step 2. Use the corresponding data $\boldsymbol{Y_0}$ to estimate the parameters of an assumed geostatistical model.

Step 3. Specify a selection criterion for the addition of one or more new sample locations to form an enlarged sample location set $\boldsymbol{X_0} \cup \boldsymbol{X_1}$.

Step 4. Repeat Steps 2 and 3 with augmented data $\boldsymbol{Y_1}$ at the points in $\boldsymbol{X_1}$.

Step 5. Stop when the required number of sample locations has been selected, a required performance criterion has been achieved or no more candidate sampling points are available.

2.2.4 Selection Criteria

Recall that $S = \{S(x) : x \in D\}$ denotes the realization of the process $S(x)$ over D. Let $T = T(S)$ represent the predictive target, presenting the property of S that is of primary interest. Define $\boldsymbol{X^*}$ as n^* potential sampling locations $x_i \in D$. Given an initial sampling design $\boldsymbol{X_0}$, the additional sampling location set is $\boldsymbol{A_0} = \boldsymbol{X^*} \setminus \boldsymbol{X_0}$. Two measures, predictive variance PV and exceedance probabilities EP, can be used to select sampling locations.

The prediction variance PV of each location x is defined as $PV(x) = Var(T|\boldsymbol{Y_0})$ (Diggle and Ribeiro 2007). Using PV, adaptive sampling selects locations with the largest values of PV singly or in batches (Chipeta et al 2016). Exceedance probability PE of each location x is $EP(x) = P(\{T(x) > t|\boldsymbol{Y_0}\} - 0.5)$ given a threshold t [9]. Adaptive sampling selects locations that minimize the $EP(x)$ singly or in batches.

2.2.5 Performance Criteria

A generic measure of the predictive accuracy of a design $\boldsymbol{X}$, the mean square error, is used:

$$MSE(\hat{T}) = E[(T - \hat{T})^2] \tag{14}$$

where $\hat{T} = E[T|\boldsymbol{Y}; \boldsymbol{X}]$ is the minimum mean square error predictor of T for any given design $\boldsymbol{X}$ in D.

2.2.6 An Adaptive Design Algorithm

An adaptive sampling algorithm based on PV is as follows.

Consider the following notation:

$\boldsymbol{X^*}$: the set of all candidate sampling locations, with number of elements n^*

$\boldsymbol{X_0}$: the initial sampling locations, with number of elements n_0

b: the batch size

$n = n_0 + kb$: the total sample size

$\boldsymbol{X_j}$ when $j \geq 1$: the set of locations added in the j^{th} batch, with number of elements b

$\boldsymbol{A_j} = \boldsymbol{X^*} \setminus (\boldsymbol{X_0} \cup \ldots, \cup \boldsymbol{X_j})$: the set of available locations after addition of the j^{th} batch

δ: the minimum distance between two sampling locations

Step 1. Use a non-adaptive design to determine $\boldsymbol{X_0}$.

Step 2. Set $j = 0$.

Step 3. For each $x \in \boldsymbol{A_j}$, calculate $PV(x)$:

(i) Choose $x^* = argmax_{\boldsymbol{A_j}} PV(x)$;

(ii) If $\| x^* - x_i \| > \delta$ for all $i = 1, \ldots, n_0 + jb$, add x^* to the design;

(iii) Otherwise, remove x^* from $\boldsymbol{A_j}$.

Step 4. Repeat Step 3 until b locations have been added to form the set $\boldsymbol{X_{j+1}}$.

Step 5. Set $\boldsymbol{A_j} = \boldsymbol{A_{j=1}} \setminus \boldsymbol{X_j}$ and j is updated to $j + 1$.

Step 6. Repeat Steps 3 to 5 until the total number of sampling locations is n or $\boldsymbol{A_j} = \phi$.

2.3 Preferential Sampling [Lucy and Zhuowei]

Geostatistics involves the fitting of spatially continuous models to spatially discrete data. Preferential sampling in geostatistics was coined by Diggle et al.[4] to mean that the locations at which observations are made may depend on the spatial process that underlines the correlation structure of the measurements.

Diggle, Menezes and Su (2010, henceforth DMS) used a parametric model to demonstrate some of the practical implications of strongly preferential sampling [4].Before we introduce DMS model, we discuss some stochastic processes which are important bases for constructing DMS model.

2.3.1 Stochastic Processes [Lucy]

2.3.1.1 Gaussian Processes

The class of Gaussian processes is one of the most widely used families of stochastic processes for modeling dependent data observed over time, or space, or time and space.

A real-valued stochastic process $X(t), t \in T$, where T is an index set, is a Gaussian process if all the finite-dimensional distributions have a multivariate normal distribution. That is, for any choice of distinct values $t_1, \cdots, t_k \in T$, the random vector $\boldsymbol{X} = \{X(t_1), \cdots, X(t_k)\}'$ has a multivariate normal distribution with mean vector μ=E$\boldsymbol{X}$ and covariance matrix Σ=cov$(\boldsymbol{X}, \boldsymbol{X})$, which will be denoted by $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$.

Provided the covariance matrix Σ is nonsingular, the random vector $\boldsymbol{X}$ has a Gaussian probability density function given by

$$f_{\boldsymbol{X}}(x) = (2\pi)^{-n/2}(det\Sigma)^{-1/2} \times \exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})) \tag{15}$$

In environmental applications, the subscript t will typically denote a point in time, or space, or space and time. A Gaussian process is completely determined by its mean and covariance functions which are defined by $\mu(t) = \mathrm{E}X(t)$ and $\gamma(s,t) = \mathrm{cov}(X(s), X(t))$ respectively.

A Gaussian process $X(t)$ is strictly stationary if all n-dimensional distributions of $X(t_1 + h), \cdots, X(t_n + h)$ are independent of h .

It is called weakly stationary if its mean is constant, $m(t) = \mathrm{E}X_t = \mu$, and if its covariance function $\gamma(t) = \gamma(t + h, t) = \mathrm{cov}(X(t + h), X(t))$ is a function only of the lag t.

2.3.1.2 Poisson Point Processes

The Poisson point processes is a type of random object that consists of randomly positioned points located on some underlying mathematical space [12]. Defined on the real line, the Poisson point process plays an important role in the field of queueing theory, where it is used to model random events happening in time such as the arrival of customers at a store or occurrence of earthquakes [13]. Defined on the Euclidean plane, the Poisson point process, also known as a spatial Poisson process, can represent scattered objects such as trees in a forest or transmitters in a wireless network [11].

The Poisson point process has the property that for a collection of disjoint bounded regions of the underlying space, the number of points of the Poisson point process in each bounded subregion will be completely independent from all the other points of the point process[3], which motivates the Poisson point process sometimes called a purely or completely random process.

Poisson Distribution of Point Counts

The Poisson point process is related to the Poisson distribution, which implies that the probability distribution of a Poisson random variable N is equal to n is given by:

$$P\{N = n\} = \frac{\Lambda^n}{n!} e^{-\Lambda} \tag{16}$$

where the parameter Λ determines the shape of the distribution. If a Poisson point process has a parameter of the form $\Lambda = v\lambda$, where λ is a constant and v is the Lebesgue measure which assigns length, area or volume to sets, then the resulting process is called a homogeneous or stationary Poisson point process. Otherwise, the parameter depends on its location in the underlying space, which leads to the inhomogeneous Poisson point process. The parameter, called rate or intensity, is related to the average number of Poisson point existing in some bounded region. In fact, the parameter can be interpreted as the average number of points per some unit of length, area or volume, depending on the underlying mathematical space, hence it is sometimes called the mean density [3].

Homogeneous Spatial Poisson Point Process

A spatial Poisson point process is a Poisson point process defined on the plane R^2. For its definition, consider a Borel measurable region $\mathcal{B}$ of the plane. $N(\mathcal{B})$ denotes the number of points of a point process N existing in this region $\mathcal{B} \subset R^2$. If the points belong to a homogeneous Poisson process with parameter $\lambda > 0$, then the probability of n points existing in $\mathcal{B}$ is given by:

$$P\{N(\mathcal{B}) = n\} = \frac{(\lambda|\mathcal{B}|)^n}{n!} e^{-\lambda|\mathcal{B}|} \tag{17}$$

where $|\mathcal{B}|$ denotes the area of $\mathcal{B}$.

Inhomogeneous Spatial Poisson Point process

The inhomogeneous Poisson point process is a Poisson point process with Poisson parameter defined as some location-dependent function in the underlying space on which the Poisson point process is defined. For Euclidean space R^2, this is achieved by introducing a locally integrable positive function $\lambda(x)$, where x is a 2-dimensional located point in R^2. For any bounded region $\mathcal{B} \subset R^2$, the integral of $\lambda(x)$ over region $\mathcal{B}$ which is denoted by $\Lambda(\mathcal{B})$ is finite:

$$\Lambda(\mathcal{B}) = \int_{\mathcal{B}} \lambda(x) dx < \infty \tag{18}$$

Then the probability of n points existing in $\mathcal{B}$ is given by:

$$P\{N(\mathcal{B}) = n\} = \frac{(\Lambda(\mathcal{B}))^n}{n!} e^{-\Lambda(\mathcal{B})} \tag{19}$$

Here, $\Lambda(\mathcal{B})$ has the interpretation of being the expected number of points of the Poisson process located in the bounded region $\mathcal{B}$, namely

$$\Lambda(\mathcal{B}) = E(N(\mathcal{B})) \tag{20}$$

Under a fixed design or a complete random design, the point process of the sampling locations is typically a homogenous Poisson point process with constant intensity λ . On the contrary, if preferential sampling occurred, the sampling design is dependent on the field process S and an inhomogeneous point process must be specified.

Cox Processes

The Poisson point process can be generalized in different ways. A Cox process is a generalisation of inhomogeneous Poisson point process where the intensity function $\lambda(x)$ is random. It can be constructed in two steps: in the first step, a non-negative random intensity function $\lambda(x)$ is generated. Conditional on this, an inhomogeneous Poisson with intensity function $\lambda(x)$ is constructed in the second step.

The intensity measure may be a realization of random variable. For example, if the logarithm of the intensity measure is a Gaussian field, then the resulting process is known as a log-Gaussian Cox process [17]. Here, $\lambda(x)$ is defined as

$$\lambda(x) = \exp(\alpha + \beta S(x)) \tag{21}$$

where $\{S(x) : x \in R^2\}$ is a Gaussian field.

2.3.2 Preferential Sampling Methodology [Lucy]

Geostatistics is a branch of spatial statistics that uses data that are obtained from sampling a spatially continuous phenomenon $\{S(\boldsymbol{X}) : \boldsymbol{X} \in R^2\}$ at a discrete set of locations $\boldsymbol{X} = \{x_i, i = 1, \cdots, n\}$, in a spatial region of interest $\mathcal{B} \subseteq R^2$. In general, the process S cannot be measured without error [7]. Here, if Y_i denotes the measured value at the location x_i, a simple model for the data takes the form

$$Y_i = \mu + S(x_i) + \epsilon_i, \qquad i = 1, \cdots, n \tag{22}$$

where ϵ_i are mutually independent, zero-mean random variables with so-called 'nugget variance' τ^2.

What we are interested in is performing inference on the parameters of the model, which include μ, τ and those in the covariance function of the process S. We denote $\boldsymbol{\theta}$ as the vector of all parameters in the model. Likelihood methods to perform inference for $\boldsymbol{\theta}$ based on observed data $\boldsymbol{Y}$ and $\boldsymbol{X}$ as

$$L(\boldsymbol{\theta}) = L(\boldsymbol{X}, \boldsymbol{Y}) = \log \int [S, \boldsymbol{X}, \boldsymbol{Y}] dS \tag{23}$$

where $[S, \boldsymbol{X}, \boldsymbol{Y}]$ denotes the joint distribution of S, $\boldsymbol{X}$ and $\boldsymbol{Y}$, and the integration with respect to S gives the log-likelihood function based on the data $\boldsymbol{Y}$ and $\boldsymbol{X}$.

A general factorisation of the joint distribution is

$$[S, \boldsymbol{X}, \boldsymbol{Y}] = [S][\boldsymbol{X}|S][\boldsymbol{Y}|\boldsymbol{X}, S] \tag{24}$$

In many geostatistical models, the observed locations $\boldsymbol{X}$ are assumed to be stochastically independent of the process S [4]. We refer to this as non-preferential sampling of geostatistical data. Under this assumption,$[\boldsymbol{X}|S] = [\boldsymbol{X}]$, and thus $[S, \boldsymbol{X}, \boldsymbol{Y}] = [S][\boldsymbol{X}][\boldsymbol{Y}|\boldsymbol{X}, S]$. In this case, it follows that

$$L(\boldsymbol{\theta}) = \log \int [S][\boldsymbol{Y}|\boldsymbol{X}, S]dS + \log[\boldsymbol{X}] \tag{25}$$

Conversely, preferential sampling refers to the situation where the observed locations $\boldsymbol{X}$ may depend on the unobserved process S, i.e. $[\boldsymbol{X}|S] \neq [\boldsymbol{X}]$. In this case, the sampling locations are deliberately concentrated in subregions of $\mathcal{B}$ where the underlying values of $S(x)$ are thought likely to be larger (or smaller) than average. In fact, air quality monitors are seldom randomly distributed in space and are often more commonly located where the level of pollution is higher.

Thus, for preferentail sampling, $[\boldsymbol{X}|S] \neq [\boldsymbol{X}]$. To perform inference on $\boldsymbol{\theta}$, one should use the full log likelihood function

$$L(\boldsymbol{\theta}) = \log \int [S][\boldsymbol{X}|S][\boldsymbol{Y}|\boldsymbol{X}, S]dS \tag{26}$$

Next, we further make distinction between weakly and strongly preferential sampling. We use an alternative factorisation to Eq (24) as

$$[S, \boldsymbol{X}, \boldsymbol{Y}] = [S][\boldsymbol{Y}|S][\boldsymbol{X}|\boldsymbol{Y}, S] \tag{27}$$

Under weakly preferential sampling [8], we separate $\boldsymbol{X}$ and S, i.e. $[\boldsymbol{X}|\boldsymbol{Y}, S] = [\boldsymbol{X}|\boldsymbol{Y}]$, and the log-likelihood becomes

$$L(\boldsymbol{\theta}) = L(\boldsymbol{X}, \boldsymbol{Y}) = \log \int [\boldsymbol{Y}|S][S]dS + \log[\boldsymbol{X}|\boldsymbol{Y}] \tag{28}$$

Therefore, if we ignore the term $\log[\boldsymbol{X}|\boldsymbol{Y}]$ in Eq (28) to make the inference for $\boldsymbol{S}$ or $\boldsymbol{Y}$, it will be valid, but may or may not be inefficient depending on the exact specification of $[\boldsymbol{X}|\boldsymbol{Y}]$. In contrast, under strongly preferential sampling, no factorisation of $[S, \boldsymbol{X}, \boldsymbol{Y}]$ separates $\boldsymbol{X}$ and S, hence valid inference requires the stochastic nature to be taken into account, i.e. $\boldsymbol{X}$ is not ignorable.

2.3.3 The Diggle, Menezes and Su model [Zhuowei]

This part is a review of the Diggle et al. 2010 paper [4].

As described above, preferential sampling refers to the situations when

$$[S, X] \neq [S][X]$$

In those situations, applying conventional geostatistical methods to preferentially sampled data can cause misleading inferences.

The scientific interest is that with the data (X, Y), we want to uncover the properties of S while protect against possible incorrect inferences caused by the stochastic dependence between S and X.

The Diggle, Menezes and Su (DMS) model defines a specific class of models through following assumptions:[4]

Assumption 1. S is a stationary Gaussian process with mean 0, variance σ^2 and correlation function $\rho(u;\phi) = \text{corr}\{S(x), S(x^{'})\}$ for any x and $x^{'}$ a distance u apart.

Assumption 2. Conditional on S, $\boldsymbol{X}$ is an inhomogeneous Poisson process with intensity

$$\lambda(x) = \exp\{\alpha + \beta S(x)\} \tag{29}$$

Assumption 3. Conditional on S and $\boldsymbol{X}$, $\boldsymbol{Y}$ is a set of mutually independent Gaussian variates with

$$Y_i \sim N\{\mu + S(x_i), \tau^2\} \tag{30}$$

2.3.3.1 Fitting the DMS model

Monte Carlo Maximum Likelihood

To fit the DMS model, Monte Carlo maximum likelihood estimation is used.

For a shared latent process model like in Eq (27), the likelihood function for data X and Y can be expressed as

$$L(\theta) = [X, Y] = E_s[[X|S][Y|X, S]], \tag{31}$$

θ represents the parameters for the model. To approximately evaluate $L(\theta)$, S is approximated by the set of values of S on a finely spaced lattice to cover A. The sampling locations X are also approximated by the nearest lattice points. Now, partition S into $S = \{S_0, \S_1$, where S_0 denotes the values of S where the lattice points are picked for sampling and S_1 denotes the remaining lattice points.

To evaluate $L(\theta)$, rewrite (31) as integral

$$\begin{aligned} L(\theta) &= \int [X|S][Y|X, S]\frac{[S|Y]}{[S|Y]}[S]dS \\ &= \int [X|S]\frac{[Y|S_0]}{[S_0|Y]}[S_0][S|Y]dS \\ &= E_{S|Y}\left[[X|S]\frac{[Y|S_0]}{[S_0|Y]}[S_0]\right] \end{aligned} \tag{32}$$

A Monte Carlo approximation is

$$L_{MC}(\theta) = m^{-1}\sum_{j=1}^{m}[X|S_j]\frac{[Y|S_{0j}]}{[S_{0j}|Y]}[S_{0j}], \tag{33}$$

where S_j are simulations of S conditional on Y.

[S|Y] Realizations Simulation

Let N be the number of prediction locations, X^*=$\{x_1^*, ...x_N^*\}$. Data locations X=$\{x_1, ...x_n\}$ is a subset of the X^*. Define C as the $n \times N$ matrix whose ith row has all 0s but a single 1 to identify the position of x_i. Unconditionally, $S \sim MVN(0, \Sigma)$, $Y \sim MVN(\mu, \Sigma_0)$ with $\Sigma_0 = C\Sigma C' + \tau^2 I$. Let Z denotes an independent random sample of size n from N(0, τ^2), y denotes the observed value of Y, S is a random draw from MVN(0, Σ). Given Y=y,

$$S + \Sigma C'\Sigma_0^{-1}(y - \mu + Z - CS), \tag{34}$$

has the required multivariate Gaussian distribution of S.

2.3.3.2 Goodness of Fit

Comparing the sample locations with realizations of the fitted Cox model for their unconditional distribution is a way of assessing the goodness of fit of the preferential sampling model.

K-function, which is the reduced second-moment measure is one of the standard diagnostic tool for stationary spatial point processes. K-function is defined as $\lambda K(s) = E[N_0(s)]$, where λ is the expected number of events of the process per unit area and $N_0(s)$ is the number of additional events of the process within distance s of an arbitrary event. Under the preferential sampling model, the marginal model for X is a log-Gaussian Cox process with intensity (29). For this process, K-function is

$$K(s) = \pi s^2 + 2\pi \int_0^s \gamma(u)udu, \tag{35}$$

where $\gamma(u) = exp\{\beta^2\sigma^2\rho(u;\phi)\} - 1$. The discrepancy between estimated and theoretical K-functions can be used as a goodness of fit statistics for Monte Carlo test.

2.3.3.3 Effect of Preferential Sampling

A simulation experiment is conducted by Diggle et al.[4] to illustrate the effect of preferential sampling. Here's how they simulate the data:

For each run of simulation:

1. On a finely spaced lattice, simulate a stationary Gaussian process on every unit square and the spatially continuous process $S(\cdot)$ is treated as constant within each unit square.

2. Sample the value of $S(\cdot)$ with additive Gaussian measurement error.

($\mu = 4$, $\sigma^2 = 1.5$, Matern correlation with scale parameter $\Phi = 0.15$, shape parameter $\kappa = 1$, nugget variance $\tau^2 = 0$)

3. Based on the design, sample X and Y accordingly

(1). For the completely random sampling design, a realization of $X = x_i : i = 1, ..., n$ is generated independently from uniform distribution on $\mathcal{B}$.

(2). For the preferential design, a realization of X is generated conditional on S by Eq (29) with $\beta = 2$. A realization of Y conditional on S and X is generated using Eq (30)

(3). For the clustered design, a realization of X is generated in the same way as in preferential design. However, a realization of Y is generated with a second, independent, realization of S.

Figure 1 shows the realizations of S and X. The background for (a), (b) and (c) is the realizations of $S(\cdot)$ while the dots represent the sampling locations. The figure shows the difference between different sampling designs. For (a), the sampling locations are scattered across the unit square. For (b), the locations are mostly at the shaded region(representing higher values). For (c), the locations picked are clustered, but not corresponding to the shown realization of S.

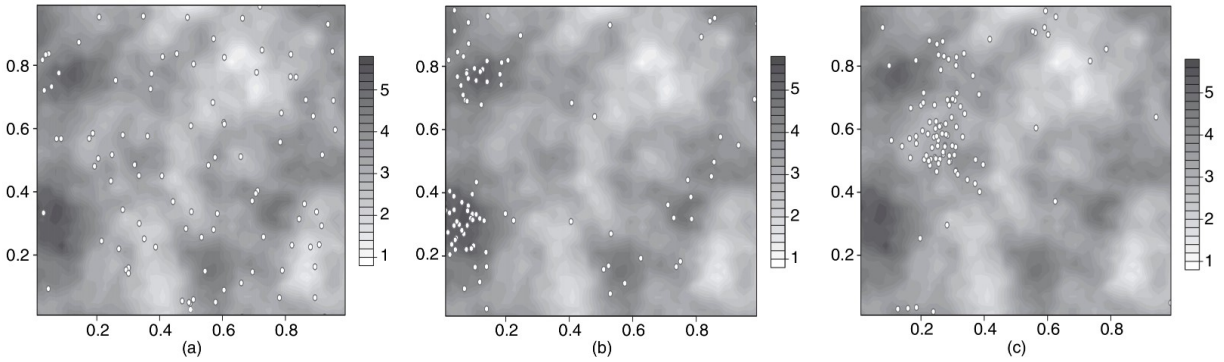


Figure 1: Underlying realization S and the sample locations for simulation. (a) completely random sampling. (b) preferential design (c) clustered design (from [4])

In spatial statistics, the theoretical variogram $V(u) = \frac{1}{2}varS(x) - S(x')$ described the degree of spatial dependence of a stationary spatial process S(x), where u denotes the distance between x and x'. With observed data (x_i, y_i), the empirical variogram ordinates can be calculated as $v_{ij} = (y_i - y_j)^2/2$. A scatterplot of v_{ij} against u_{ij} is called the variogram cloud. Under non-preferential design, each v_{ij} is an unbiased estimate of $V(u_{ij})$.

Using the three designs described above, more realizations can be simulated and variograms can be calculated for analysis. Here, 500 replicates are simulated for each of the three sampling designs. Fig.2(a) shows that preferential design has the largest bias while random and clustered design are approximately unbiased. This is not unexpected since in preferential design, sampling locations tend to span a reduced range of values of S(x). In this case, the expectations of the pairwise squared differences at any u are also reduced. Fig. 2(b) shows that clustered design has the highest standard deviation.

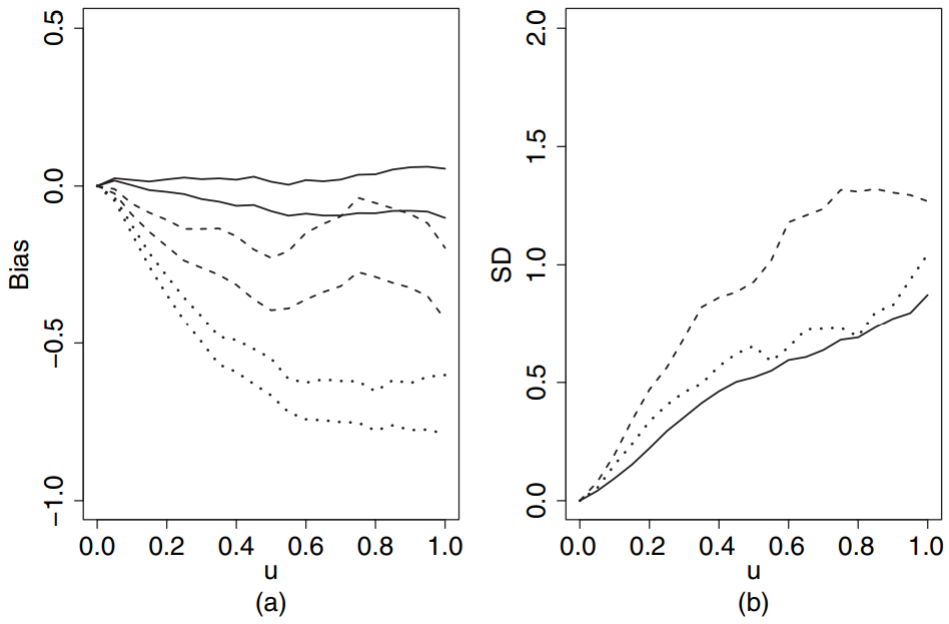


Figure 2: Estimated bias and standard deviation of the sample variogram under random (solid line), preferential(dotted line) and clustered(dashed line) design. (a) pointwise means plus and minus two pointwise standard errors (b) pointwise standard deviations (from [4])

We can also make predictions for the unmonitered locations. To predict for the value at $S(x_0)$, given sample data (x_i, y_i), maximum likelihood estimation under the assumed Gaussian model is used.

Analysis about predictions is also conducted under the three different designs mentioned above. Each simulation gives us an estimate of the bias $\hat{S}(x_0) - S(x_0)$ and the mean-square error $\{\hat{S}(x_0) - S(x_0)\}^2$, $\hat{S}(x_0)$ is the ordinary kriging predictor.

Table 1: 95% Confidence intervals for Bias and RMSE of the ordinary kriging predictor $\hat{S}(x_0)$ at x_0=(0.49, 0.49)

Model	Parameter	Completely random	Preferential	Clustered
1	Bias	(-0.014, 0.055)	(0.951, 1.145)	(-0.048, 0.102)
1	RMSE	(0.345, 0.422)	(1.387, 1.618)	(0.758, 0.915)
2	Bias	(0.003, 0.042)	(-0.134, -0.090)	(-0.018, 0.023)
2	RMSE	(0.202, 0.228)	(0.247, 0.292)	(0.214, 0.247)

Table 1 shows the 95% confidence intervals for bias and root-mean-square-error(RMSE) at the prediction location $x_0 = (0.49, 0.49)$. Again, 500 independent simulations are used here. Each sample consists of 100 locations on the unit square. In Model 1, the parameters are set the same as above, with $\mu = 4, \sigma^2 = 1.5, \Phi = 0.15, \kappa = 1, \beta = 2$ and $\tau^2 = 0$. In Model 2, the model values are set using 1997 Galicia biomonitering data(a preferential design), where $\mu = 1.515, \sigma^2 = 0.138, \Phi = 0.313, \kappa = 0.5, \beta = -2.198$ and $\tau^2 = 0.059$.

From model 1, the two non-preferential designs lead to approximately unbiased prediction but the bias is large and positive under preferential design. This prediction bias is caused by the bias in parameter estimation. Preferential design sets the sampling location mostly at the high value locations of S which causes the bias in the estimation of the model parameters. and then the bias in prediction. For RMSE, preferential design shows the biggest RMSE. Completely random and clustered designs have smaller RMSE. The larger RMSE for clustered design shows that clustered design is less efficient. From model 2, the difference between the three settings are smaller compared to model 1. This can be caused by two reasons. First off, the parameters set for model 2 have a weaker preferentiality than model 1. This can be measured by $\beta\sigma$, which is equal to 3 for model 1 and 0.815 for model 2. Secondly, the effect is also diluted by the non-zero nugget variance in model 2.

2.3.4 The Pati, Reich and Dunson model [Zhuowei]

The Pati, Reich and Duncon(PRD) model is an extension of the DMS model. A second Gaussian process is added as follows.[6]

X is a log-Gaussian process with intensity

$$\lambda(x) = \exp\{\alpha + S_1(x)\} \tag{36}$$

where S_1 is a Gaussian process with mean zero variance σ_1^2 and correlation function $\rho(u/\Phi_1)$.

The measurements $Y = \{y_1, ...y_n\}$ at locations x_i follow the model

$$y_i = \mu + \beta S_1(x_i) + S_2(x_i) + Z_i, \tag{37}$$

Here, S_2 is an isotropic Gaussian process, independent of S_1, with mean zero, variance σ_2^2 and correlation function $\rho(u/\Phi_2)$. Z_i are mutually independent N(0, τ^2) variates. The process S_2 represents a component of the spatial variation in the measurement process that is not linked to the sampling process.

3 Data

3.1 Simulated Binomial Data [Jing]

A realization of Gaussian process $S(x)$ on a 35 by 35 grid covering the unit square was conducted, giving a total of $n^* = 1225$ potential sampling locations. $S(x)$ is specified to have expectation $\mu = 0$, variance $\sigma^2 = 1$ and Matérn correlation function (8) with $\phi = 0.15$ and $\kappa = 1.5$, and no measurement error, i.e. $\tau^2 = 0$. Binomial observations, with 8 trails at each grid point and probabilities given by the anti-logit of the simulated values of the Gaussian process, constitute the response variable y. This data set is available in the R package **geosample** as **sim.data**.

3.2 Malaria Prevalence Data in Majete [Jing]

A survey sample for malaria prevalence mapping in an area surrounding Majete Wildlife Reserve with Chikwawa district, southern Malawi is provided in the R package **geosample**. There are in total 747 households as the candidate sampling locations (Figure 3). Data collected includes the outcome of a malaria rapid diagnostic test (a binary variable), age and gender of each individual and socioeconomic status of each household, digital elevation model (DEM), elevation and normalized difference vegetation index (NDVI).

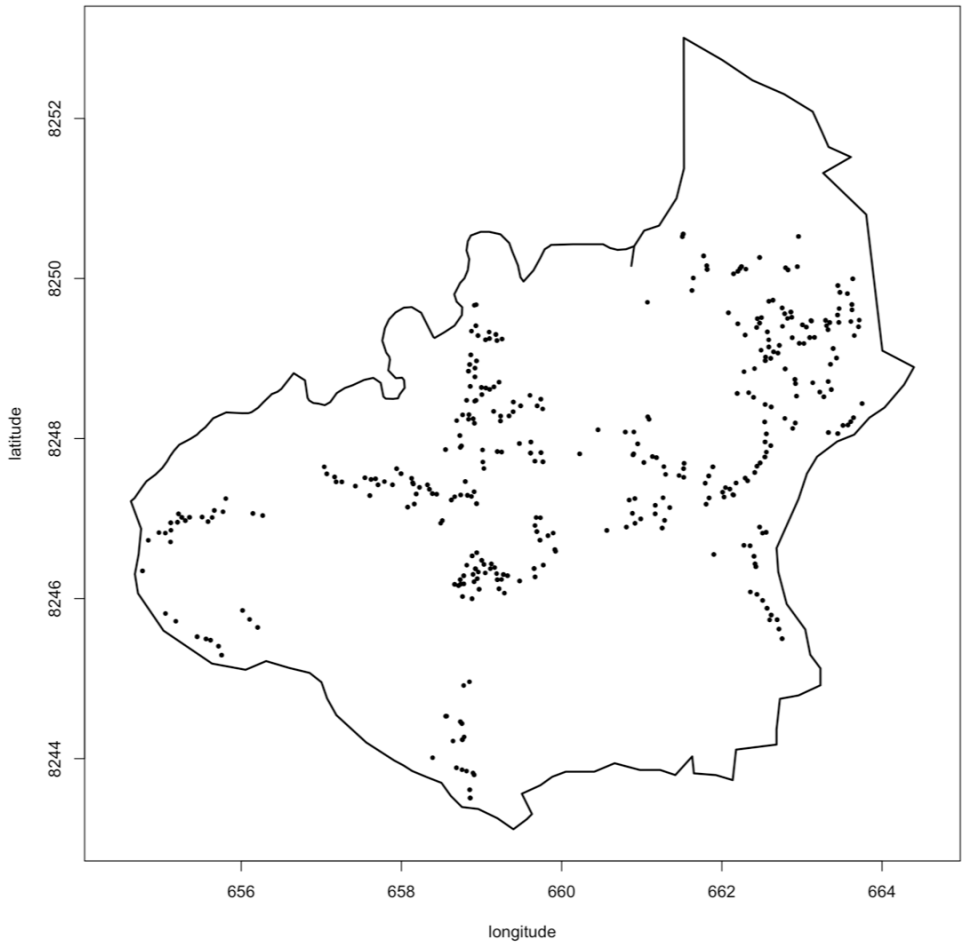


Figure 3: Candidate household sampling locations in Majete

3.3 Moscow Air Quality Monitoring Network [Evgeny]

According to MosEcoMonitoring (MEM)[1] the air quality network in Moscow has been in operation since 1995 and is now comprised of 52 stations, 3 meteorologic complexes and 2 mobile stations (denoted by blue dots on Figure 4). For regulatory purposes the stations are sited in various surroundings, including residential neighborhoods, industrial areas and

[1]Governmental agency in Moscow, tasked with regulatory environmental oversight.

mixed-type zones prone to industrial pollution. While MEM claims to monitor 32 different compounds, the data set had reference to only 18 pollutant species. Overall, only 3 measured compounds have the highest spatial coverage: carbon monoxide (49 stations), nitrogen oxide (44 stations) and nitrogen dioxide (49 stations). For the purposes of this study, only 47 stations were used.

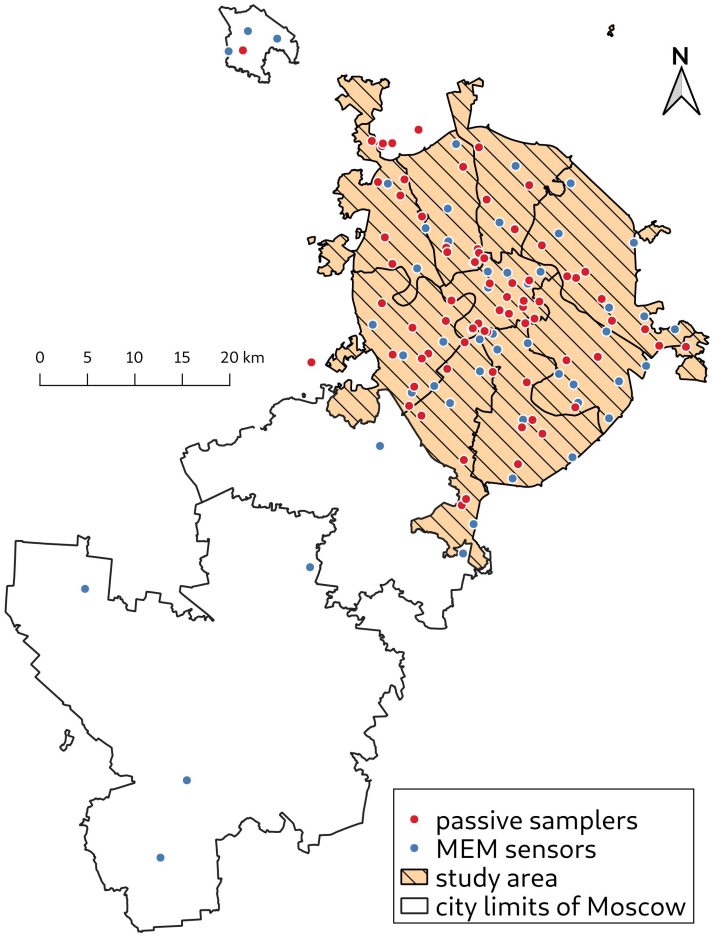


Figure 4: Existing Network in Moscow, Russia

4 Results

4.1 Non-adaptive Sampling [Evgeny]

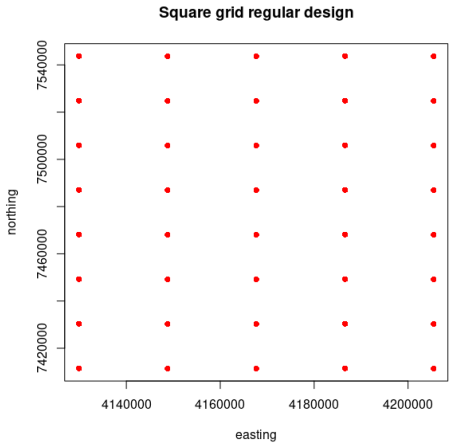
To look into non-adaptive sample designs, we decided to explore the existing monitoring network in Moscow, Russia and compare it to complete random design, complete regular design, probabilistic and inhibitory designs outlined in Section 2.1. Because there are 47 monitoring station in Moscow, and for comparison purposes, n was set to 47 to generate for all of the generated sampling designs.

First, we generated probabilistic, complete random and complete regular samples via *spsample* module in *sp* package:

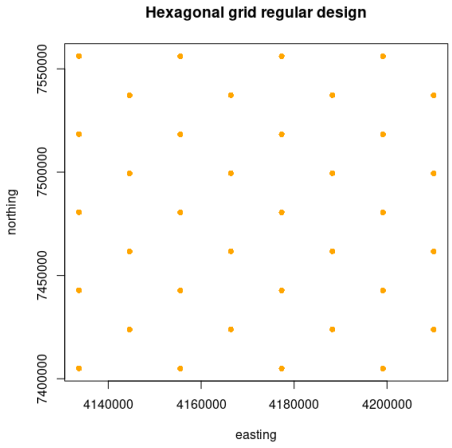
- square grid design (see Figure 5a)
- hexagonal regular design (see Figure 5b)
- clustered design (see Figure 5c)
- stratified design (see Figure 5d)
- non-aligned design (see Figure 5e)
- random design (see Figure 5f)

As we can see, square, hexagonal and non-aligned samples are regular in design, while clustered, stratified and random designs seem to have some aspect of randomness.

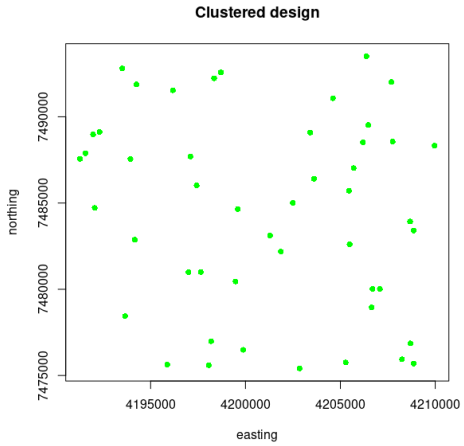
Second, simple inhibitory designs were explored (see Figure 6. Two maps were generated with $n = 47$ and $\delta_1 = 0.0045$ (500m) and $\delta_2 = 0.09$ (10000m). Since we have a fixed n and fixed study area, δ can only vary within a certain range to ensure that none of the generated points are closer than the specified threshold (δ).



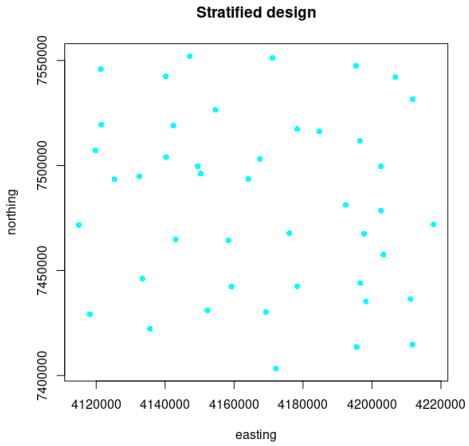
(a) Squared grid sample



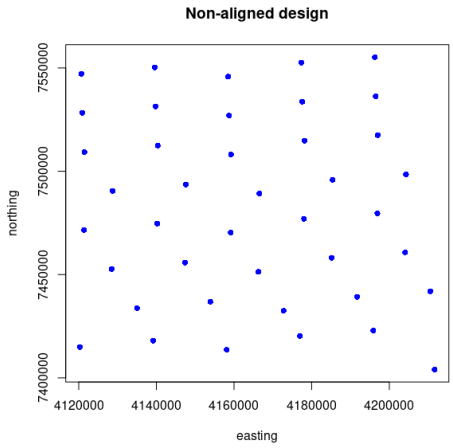
(b) Hexagonal regular sample



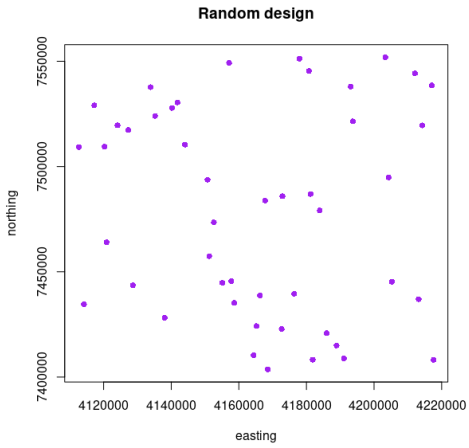
(c) Clustered sample



(d) Stratified random sample

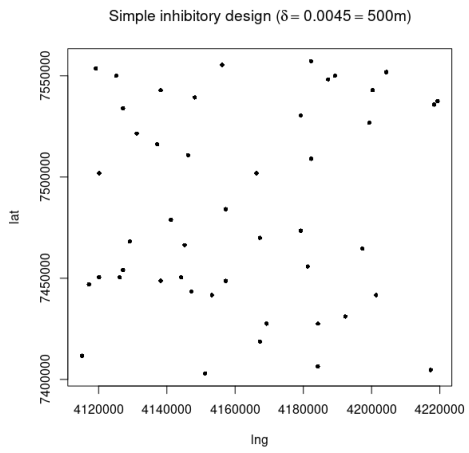


(e) Non-aligned sample

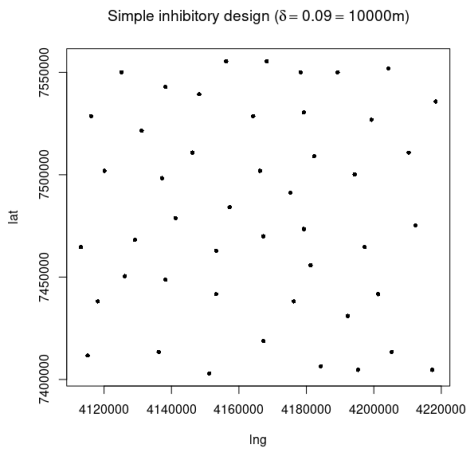


(f) Random sample

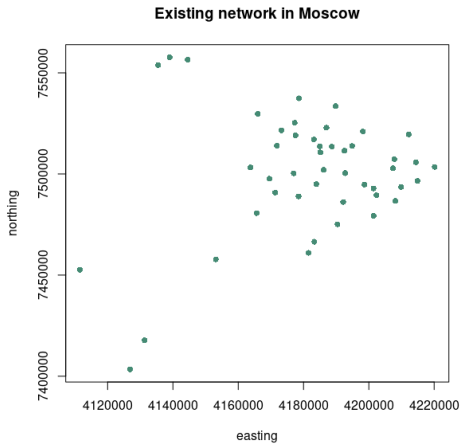
Figure 5: Comparison of Generated Sampling Designs in Moscow, Russia



(a) Simple inhibitory design $\delta = 0.009$



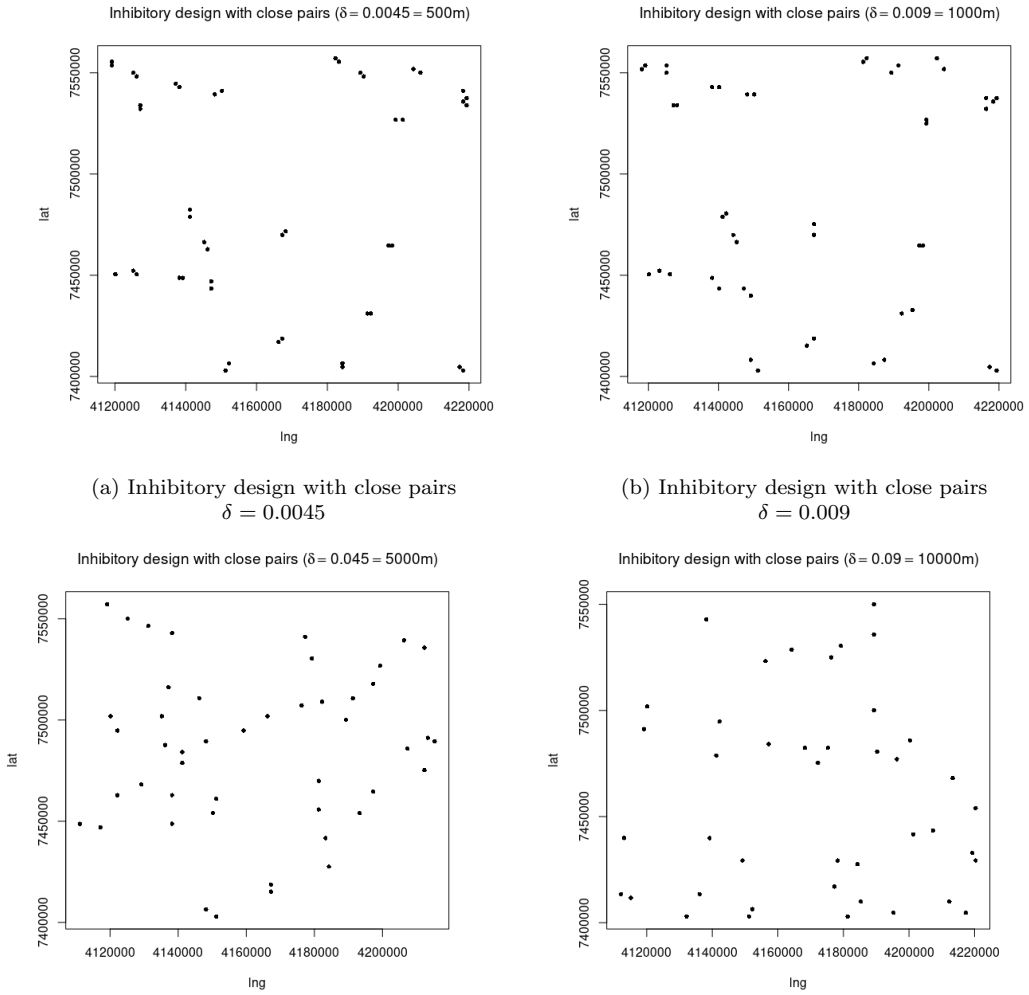
(b) Simple inhibitory design $\delta = 0.045$



(c) Existing network in Moscow

Figure 6: Simple Inhibitory Designs and Existing network

Third, for inhibitory design with close pairs, we prepared 4 different illustrations of generated networks with varying δ (0.0045,0.009, 0.045, 0.09), which roughly correspond to 500m, 1,000m, 5,000m, and 10,000m at the longitude/latitude where Moscow is located (see Figure 7). As we can see, as the δ increases the design tends to become more similar to geometric/space-filling designs, which is logical, since the minimum threshold is increasing.



(a) Inhibitory design with close pairs $\delta = 0.0045$

(b) Inhibitory design with close pairs $\delta = 0.009$

(c) Inhibitory design with close pairs $\delta = 0.045$

(d) Inhibitory design with close pairs $\delta = 0.09$

Figure 7: Comparison of Simple Inhibitory Designs with Close Pairs

4.2 Adaptive Sampling [Jing]

A simulated example using the simulated binomial data and an empirical study using Majete data are given to demonstrate adaptive sampling with the R packages **geosmaple**, **PrevMap** and other supporting packages. Detailed codes can be found in a separate R markdown file. In the simulated example, a two-stage adaptive sampling design is given. One stage of adaptive sample location selection is conducted with the empirical data.

4.2.1 Simulated Example

Initial Sampling: The initial sampling is conducted using discrete inhibitory sampling to sample $n_0 = 30$ locations with $\delta = 0.04$. The results are shown in Figure 8. Sampling locations are spread out in the unit square.

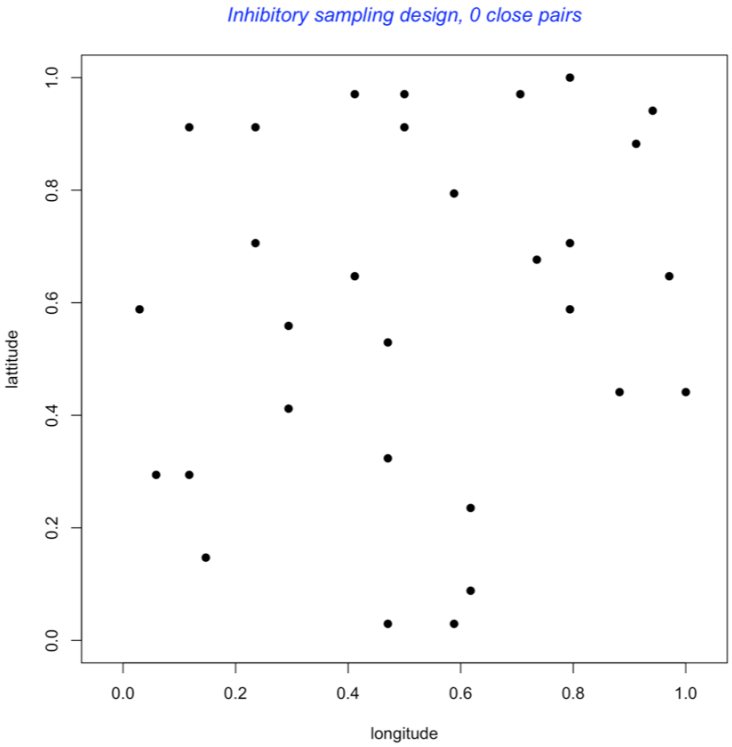


Figure 8: Inhibitory design with $n_0 = 30$ and $\delta = 0.04$

Model Estimates: The **PrevMap** package is used for model parameter estimation and spatial predictions [9]. Model (7) is fitted with an initial regression coefficient $\beta = 0.001$, the variance of Gaussian process $\sigma^2 = 1$, and the variance of the nugget effect $\tau^2 = 0$. Parameters of the Matérn function (8) are: the scale parameter of spatial correlation $\phi = 0.4$ and the shape parameter $\kappa = 1.5$. After computation, estimates are $\hat{\beta} = 0.216$, $\hat{\sigma}^2 = 1.053$ and $\hat{\phi} = 0.617$.

Spatial Prediction: The resulting binomial model is used to generate spatial predictions of prevalence at each of the 1225 sampling locations. The prediction results and exceedance probabilities with a threshold $t = 0.45$ at each location are shown in Figure 9.

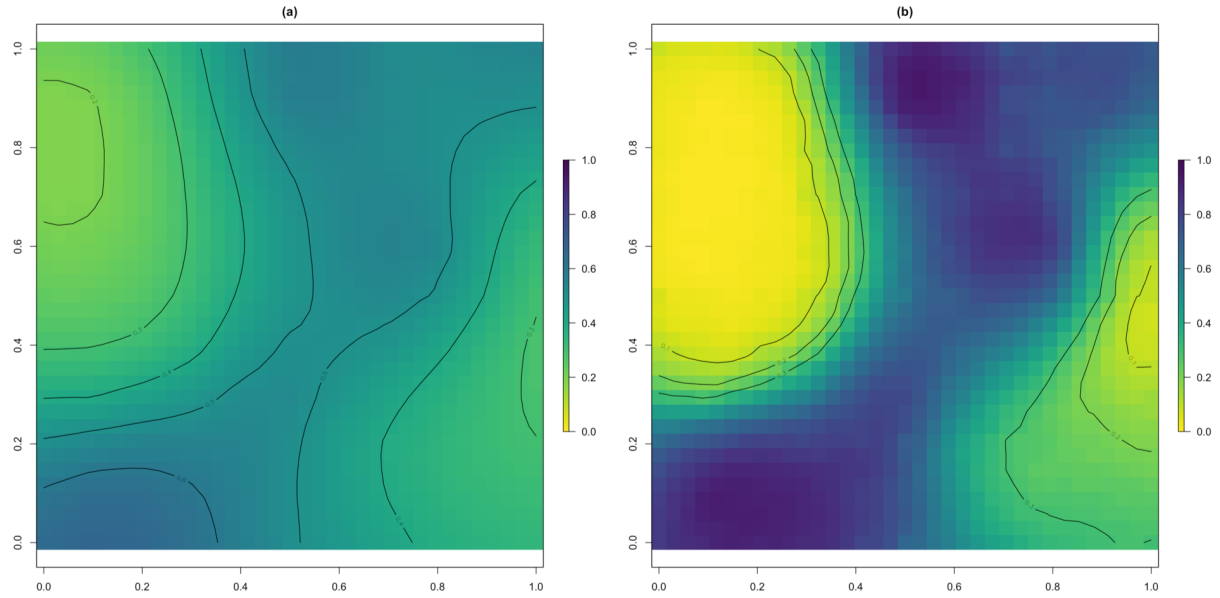


Figure 9: Spatial predictions: (a) Spatial predictions (b) Exceeedance probabilities $P(x; 0.45) = P(prev > 0.45 \, at \, location \, x)$

Adaptive Sampling – 1st Stage: In the first stage, 5 additional sampling locations are selected using the prediction variance selection criterion (Figure 10). The minimal sampling distance $\delta = 0.1$. The dark blue dots are the 30 initial sampling locations and the read dots are adaptive sampling locations added after analyzing data collected from the initial design. The new sampling locations are mostly in the right bottom corner in which there is no sampling location.

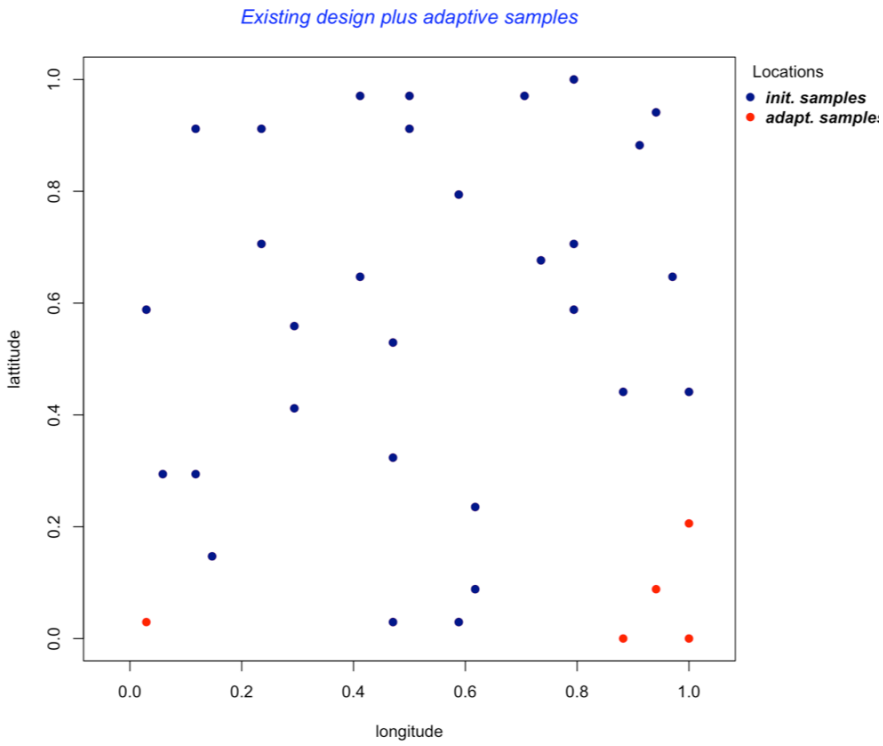


Figure 10: Adaptive sampling design with $\delta = 0.1$ and $b = 5$

Model Estimates and Prediction Updates: The binomial model are refitted with the current sampling locations (now 35 locations) and then spatial prediction is conducted based on the updated model. The initial values of model parameters are the estimated results of the previous binomial model. Now the updated model estimates are $\hat{\beta} = 0.390$, $\hat{\sigma}^2 = 0.843$ and $\hat{\phi} = 0.762$. The updated spatial predictions are given in Figure 11.

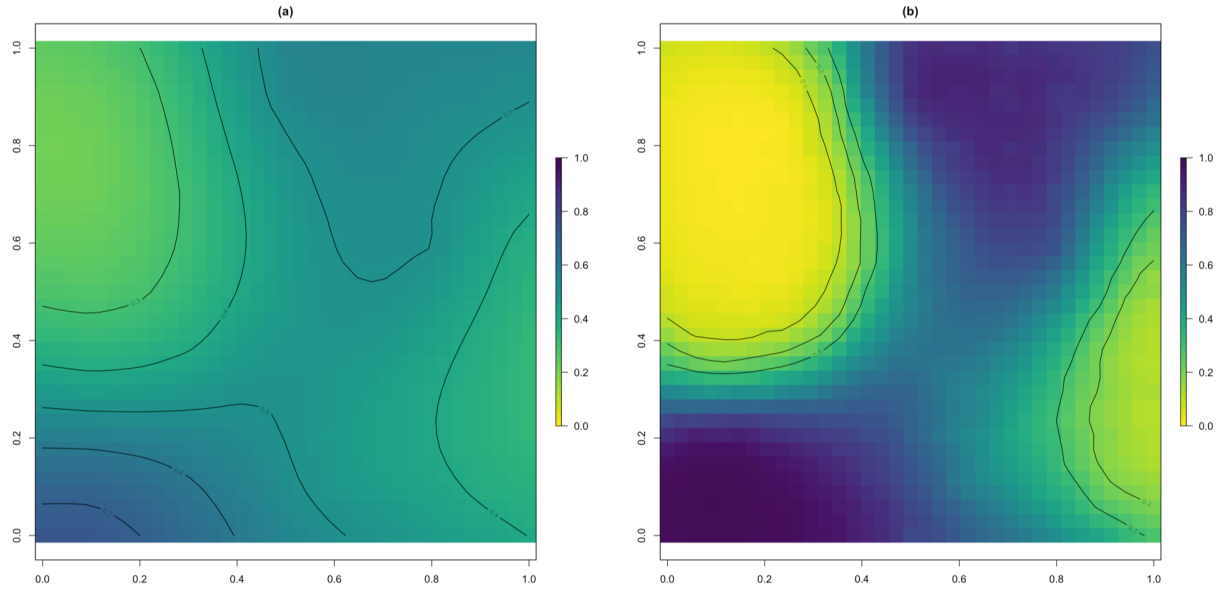


Figure 11: Updated spatial predictions: (a) Spatial predictions (b) Exceeedance probabilities $P(x; 0.45) = P(prev > 0.45 \, at \, location \, x)$

Adaptive Sampling – 2nd Stage: In the second stage, another 5 sampling locations are selected using the prediction variance selection criterion (Figure 12). The selected locations are quite spread out in the left half of the unit square.

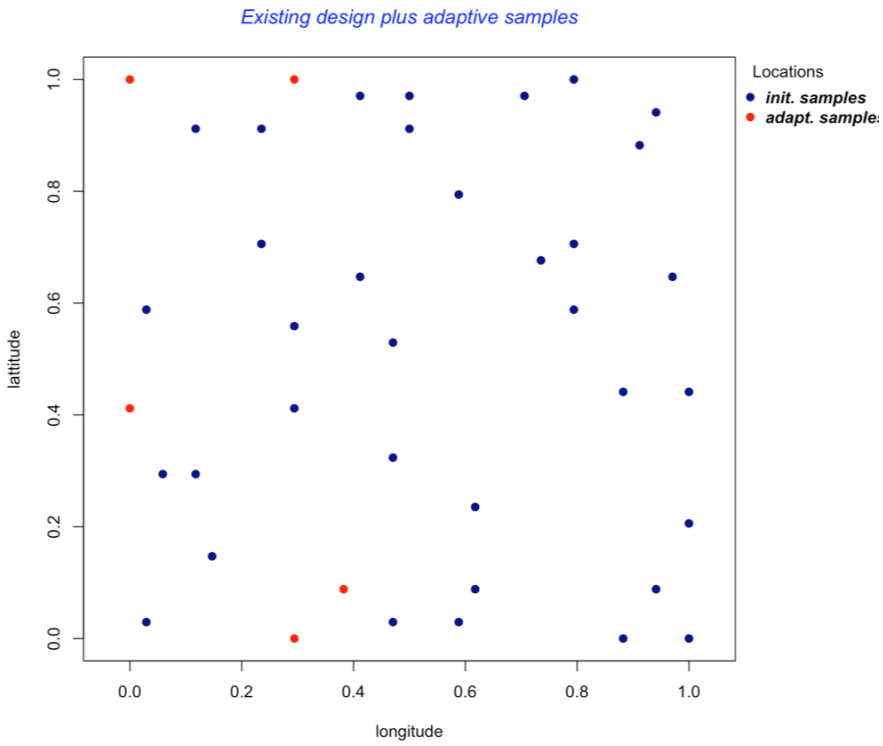


Figure 12: Adaptive sampling design with $\delta = 0.1$ and $b = 5$

4.2.2 Empirical Study

Initial Sampling: The initial sampling is conducted using discrete inhibitory sampling to sample $n_0 = 60$ households with $\delta = 400$ meters. The results are shown in Figure 13.

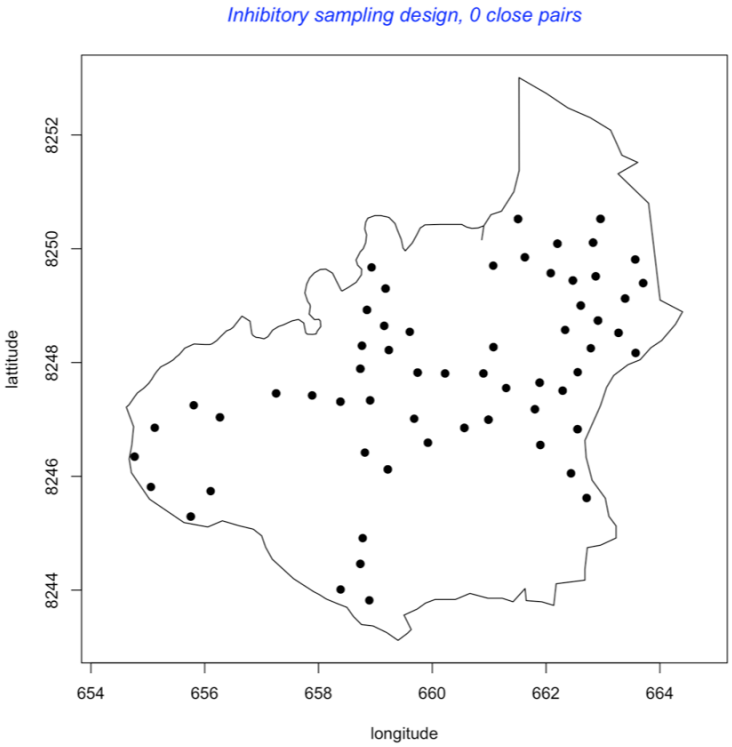


Figure 13: Inhibitory design with $n_0 = 60$ and $\delta = 400$

Model Estimates and Prediction: Here, the binomial model is studied with a simple intercept without using any covariate for demonstration. First, a generalized linear model (GLM) with logit function is fitted with collected outcomes and an intercept term. Then, the initial values of model parameters are obtained from the GLM: $\beta_0 = -1.437$, $\sigma_0^2 = 0.931$, $\tau^2 = 0$, $\phi_0 = 3.954$ and $\kappa = 0.5$. After computation, the estimates are $\hat{\beta} = -2.135$, $\hat{\sigma}^2 = 0.949$ and $\hat{\phi} = 3.913$. The resulting binomial model is used for spatial prediction.

Adaptive Sampling: Forty additional sampling locations are selected using the prediction variance selection criterion (Figure 14). The minimal sampling distance $\delta = 150$ meters. The dark blue dots are the 60 initial sampling locations and the read dots are adaptive sampling locations added after analyzing data collected from the initial design.

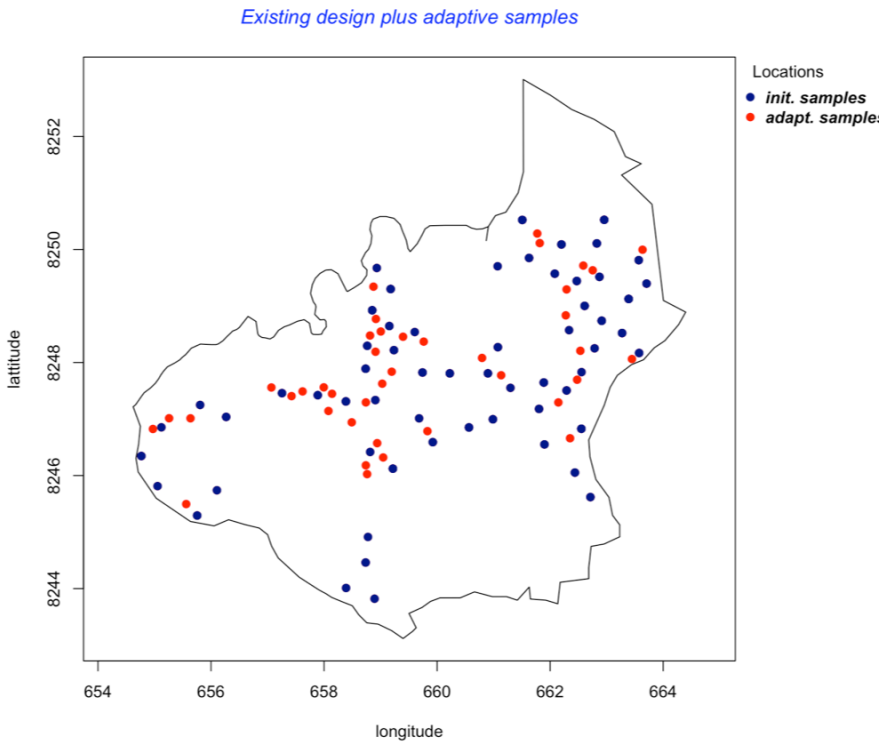


Figure 14: Adaptive sampling design with $\delta = 150$ and $b = 40$

Model Estimates and Prediction Updates: The model estimates and associated predictions are updated using the data collected on the current 100 sampling locations. Figure 15 shows the predicted malaria prevalence in Majete and associated standard errors.

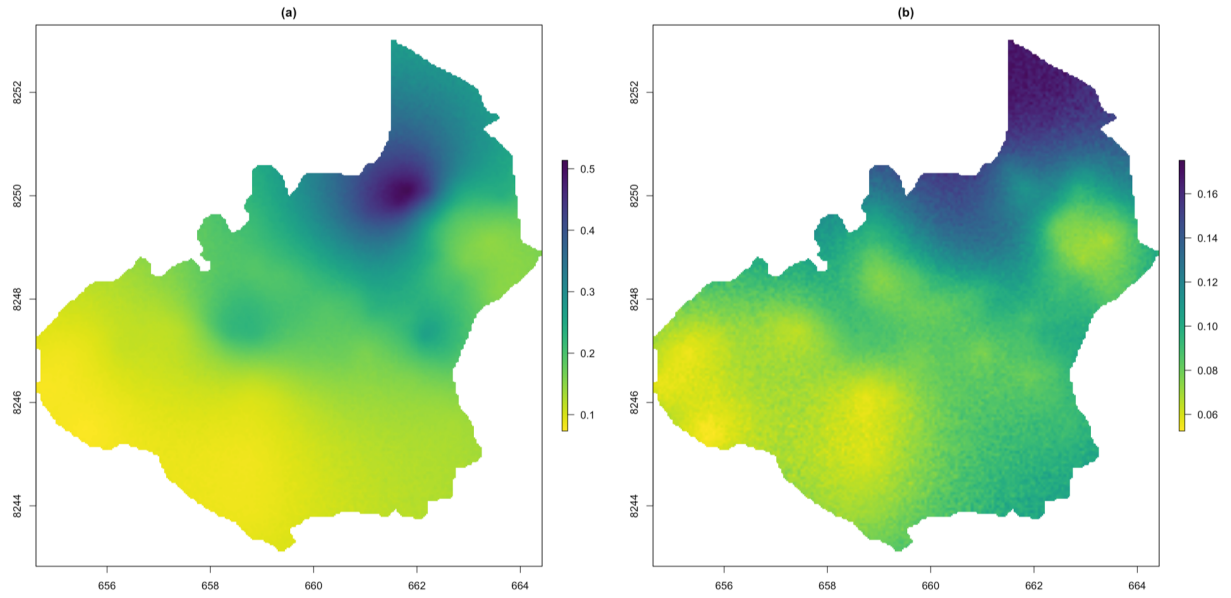


Figure 15: (a) Malaria prevalence predictions (b) Standard errors of predictions

5 Conclusion

Preferential sampling means that the process that generates the sampling locations is stochastically dependent on the spatial process S. The effect of preferential sampling can be big that ignoring the preferential nature of the sampling can lead to misleading inferences. DMS and PRD models take account of the preferential sampling to make the parameter estimation. In a potentially preferential setting, they are more suitable than conventional geostatistical models.

Adaptive sampling is in fact a sampling strategy and this study provides an adaptive sampling design strategy within a model-based geostatistics framework. The choice of selection criterion is vital. The Majete example uses average prediction variance due to the lack of prior knowledge about the spatial variation in prevalence. The criterion can be changed when more information is available. The initial sampling locations $\boldsymbol{X_0}$ are important as well. The initial sample size should be large enough to allow the fitting of a geostatistical model whose estimates decides the following sampling design.

Non-adaptive sampling, as we demonstrated can be a good starting point for first-phase sampling design, where no prior data is available. Here, inhibitory designs proposed by Chipeta et al.[1] can prove advantageous to simpler complete random and complete regular samples. There are, however, certain limitations to inhibitory design with close pairs. The main one being, the lack of control for allocating clusters of varying destiny. Very often, for monitoring and regulatory purposes it might be necessary to employ sensors at different density throughout the study area, which would not be possible with inhibitory design.

References

[1] Michael Chipeta, Dianne Terlouw, Kamija Phiri, and Peter Diggle. Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. *Environmetrics*, 28(1):e2425, 2017.

[2] Michael G Chipeta, Dianne J Terlouw, Kamija Phiri, and Peter J Diggle. Adaptive geostatistical design and analysis for sequential prevalence surveys. *arXiv preprint arXiv:1509.04448*, 2015.

[3] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes.*, volume I. Probability and its Applications (New York). Springer, 2003.

[4] Menezes R. Diggle, P. J. and T. Su. Geostatistical inference under preferential sampling. *Appl. Statist.*, 59:191–232, 2010.

[5] Peter Diggle and Søren Lophaven. Bayesian geostatistical design. *Scandinavian Journal of Statistics*, 33(1):53–64, 2006.

[6] Peter J Diggle and Emanuele Giorgi. *Model-based Geostatistics for Global Public Health: Methods and Applications*. Chapman and Hall/CRC, 2019.

[7] Tawn J. A. Diggle, P. J. and R. A. Moyeed. Model-based geostatistics. *Appl. Statist.*, 47:299–350, 1998.

[8] Daniel Dinsdale and Matias Salibian-Barrera. Methods for preferential sampling in geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68:181–198, 2018.

[9] Emanuele Giorgi, Peter J Diggle, et al. Prevmap: an r package for prevalence mapping. *J Stat Softw*, 78(8):1–29, 2017.

[10] Anton Grafström, Niklas LP Lundström, and Lina Schelin. Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514–520, 2012.

[11] M. Haenggi. *Stochastic geometry for wireless networks.* Cambridge University Press, 2012.

[12] J. F. C. Kingman. *Poisson processes*, volume 3. Oxford university press, 1992.

[13] L. Kleinrock. *Queueing systems*, volume 1. Wiley-interscience, 1975.

[14] Bertil Matérn. Spatial variation, volume 36 of. *Lecture Notes in Statistics*, 1960.

[15] D. Pati, B.J. Reich, and D.B. Dunson. Bayesian geostatistical modelling with informative sampling locations. *Biometrika*, 98(1):35–48, 2011.

[16] 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.

[17] W. S. Kendall S. N. Chiu, D. Stoyan and J. Mecke. *Stochastic geometry and its applications.* John Wiley Sons, 2013.

[18] Don L Stevens Jr and Anthony R Olsen. Spatially balanced sampling of natural resources. *Journal of the American statistical Association*, 99(465):262–278, 2004.