



sampling (SRS) algorithm, which does not explicitly incorporate spatial locations into sampling. We also found that model-based inference tends to outperform design-based inference, even for skewed data where the model-based distributional assumptions are violated. The performance gap between design-based inference and model-based inference is small when GRTS samples are used but large when SRS samples are used, suggesting that the sampling choice (whether to use GRTS or SRS) is most important when performing design-based inference.

4. There are many benefits and drawbacks to the design-based and model-based approaches for finite population spatial sampling and inference that practitioners must consider when choosing between them. We provide relevant background contextualizing each approach and study their properties in a variety of scenarios, making recommendations for use based on the practitioner's goals.

## Keywords

Design-based inference; Finite population block kriging (FPBK); Generalized random tessellation stratified (GRTS) algorithm; Local neighborhood variance estimator; Model-based inference; Restricted maximum likelihood (REML) estimation; Spatially balanced sampling; Spatial covariance

## 1. Introduction

When data cannot be collected for all units in a population (population units), data are collected on a subset of the population units – this subset is called a sample. There are two general approaches for using samples to make frequentist statistical inferences about a population: design-based and model-based. In the design-based approach, inference relies on randomly assigning some population

55 units to be in the sample (random sampling). Alternatively, in the model-based  
 56 approach, inference relies on distributional assumptions about the underlying  
 57 data-generating stochastic process (superpopulation). Each paradigm has a deep  
 58 historical context (Sterba, 2009) and its own set of benefits and drawbacks (Brus  
 59 and De Gruijter, 1997; Hansen et al., 1983). In this manuscript, we compare  
 60 design-based and model-based approaches for finite population spatial sampling  
 61 and inference.

62 Spatial data are data that have some sort of spatial index (usually specified  
 63 via coordinates). De Gruijter and Ter Braak (1990) and Brus and DeGruijter  
 64 (1993) give early comparisons of design-based and model-based approaches for  
 65 spatial data, quashing the belief that design-based approaches could not be  
 66 used for spatially correlated data. Since then, there have been several general  
 67 comparisons between design-based and model-based approaches for spatial data  
 68 (Brus and De Gruijter, 1997; Brus, 2021; Ver Hoef, 2002, 2008). Cooper (2006)  
 69 reviews the two approaches in an ecological context before introducing a “model-  
 70 assisted” variance estimator that combines aspects from each approach. In  
 71 addition to Cooper (2006), there has been substantial research and development  
 72 into estimators that use both design-based and model-based principles (see e.g.,  
 73 Sterba (2009) and Cicchitelli and Montanari (2012), and for Bayesian approaches,  
 74 see Chan-Golston et al. (2020) and Hofman and Brus (2021)).

75 While comparisons between design-based and model-based approaches have  
 76 been studied in spatial contexts, our contribution is comparing design-based  
 77 approaches specifically built for spatial data to model-based approaches. Though  
 78 the broad comparisons we draw between design-based and model-based ap-  
 79 proaches generalize to finite and infinite populations, we focus on finite popu-  
 80 lations. A finite population contains a finite number of population units (we  
 81 assume the finite number is known) – an example is lakes (treated as a whole

82 with the lake centroid representing location) in the conterminous United States.  
83 An infinite population contains an infinite number of population units – an  
84 example is locations within a single lake.

85 The rest of the manuscript is organized as follows. In Section 1.1, we introduce  
86 and provide relevant background for design-based and model-based approaches  
87 to finite population spatial sampling and inference. In Section 2, we describe  
88 how we intend to compare performance of the approaches using simulated and  
89 real data. The real data is from the United States Environmental Protection  
90 Agency’s 2012 National Lakes Assessment (NLA) (USEPA, 2012). In Section 3,  
91 we present analysis results for the simulated data and NLA data. And in Section  
92 4, we end with a discussion and provide directions for future research.

### 93 *1.1. Background*

94 The design-based and model-based approaches incorporate randomness in  
95 fundamentally different ways. In this section, we describe the role of randomness  
96 for each approach and the subsequent effects on statistical inferences for spatial  
97 data.

#### 98 *1.1.1. Comparing Design-Based and Model-Based Approaches*

99 The design-based approach assumes the population is fixed. Randomness is  
100 incorporated via the selection of population units according to a sampling design.  
101 A sampling design assigns a probability of selection to each sample (subset of  
102 population units). Some examples of commonly used sampling designs include  
103 simple random sampling, stratified random sampling, and cluster sampling.  
104 The inclusion probability of a population unit is calculated by summing each  
105 sample’s probability of selection over all samples that contain the population unit.  
106 Inclusion probabilities are often used when selecting samples and estimating  
107 population parameters.

108        When samples are chosen in a manner such that the layout of sampled units  
 109 reflects the layout of the population units, we call the resulting sample spatially  
 110 balanced. By “reflecting the layout of the population units”, we mean that if  
 111 population units are concentrated in specific areas, the units in the sample should  
 112 be concentrated in the same areas. Because spatially balanced samples reflect  
 113 the layout of the population units, they are not necessarily spread out in space  
 114 in some equidistant manner. One method of selecting spatially balanced samples  
 115 is the generalized random tessellation stratified (GRTS) algorithm (Stevens and  
 116 Olsen, 2004), which we discuss in more detail in Section 1.1.2. To quantify the  
 117 spatial balance of a sample, Stevens and Olsen (2004) proposed loss metrics  
 118 based on Voronoi polygons (i.e., Dirichlet Tessellations).

119        Fundamentally, the design-based approach combines the randomness of the  
 120 sampling design with the data collected via the sample to justify the estimation  
 121 and uncertainty quantification of fixed, unknown parameters of a population (e.g.,  
 122 a population mean). Treating the data as fixed and incorporating randomness  
 123 through the sampling design yields estimators having very few other assumptions.  
 124 Confidence intervals for these types of estimators are typically derived using  
 125 limiting arguments that incorporate all possible samples. Sample means, for  
 126 example, are asymptotically normal (Gaussian) by the central limit theorem  
 127 (under some assumptions). If we repeatedly select samples from the population,  
 128 then 95% of all 95% confidence intervals constructed from a procedure with  
 129 appropriate coverage will contain the true fixed population mean. Särndal et al.  
 130 (2003) and Lohr (2009) provide thorough reviews of the design-based approach.

131        The model-based approach assumes the population is a random realization of a  
 132 data-generating stochastic process. Randomness is formally incorporated through  
 133 distributional assumptions on this process. Strictly speaking, randomness need  
 134 not be incorporated through random sampling, though Diggle et al. (2010)

warn against preferential sampling. Preferential sampling occurs when the process generating the data locations and the process being modeled are not independent of one another. To guard against preferential sampling, model-based approaches can implement some form of random sampling. It is common, however, for model-based approaches to sample non-randomly. When model-based approaches do implement random sampling, the inclusion probabilities are ignored when analyzing the sample (in contrast to the design-based approach, which relies on these inclusion probabilities to analyze the sample).

Instead of estimating fixed, unknown population parameters, as in the design-based approach, often the goal of model-based inference is to predict the value of a realized variable. For example, suppose the realized mean of all population units (the realized population mean) is the variable of interest. Instead of a fixed, unknown mean, we are predicting the value of the mean, a random variable. Prediction intervals are then derived using assumptions of the data-generating stochastic process. If we repeatedly generate realizations from the same process and select samples, then 95% of all 95% prediction intervals constructed from a procedure with appropriate coverage will contain their respective realized means. Cressie (1993) and Schabenberger and Gotway (2017) provide thorough reviews of model-based approaches for spatial data. In Fig. 1, we provide a visual comparison of the design-based and model-based approaches (Ver Hoef (2002) and Brus (2021) provide similar figures). Fig. 1 contrasts the design-based approach with a fixed population and random sampling to the model-based approach with random populations and non-random sampling.

### *1.1.2. Spatially Balanced Design and Analysis*

We previously mentioned that the design-based approach can be used to select spatially balanced samples. Spatially balanced samples are useful because parameter estimates from these samples tend to vary less than parameter estimates

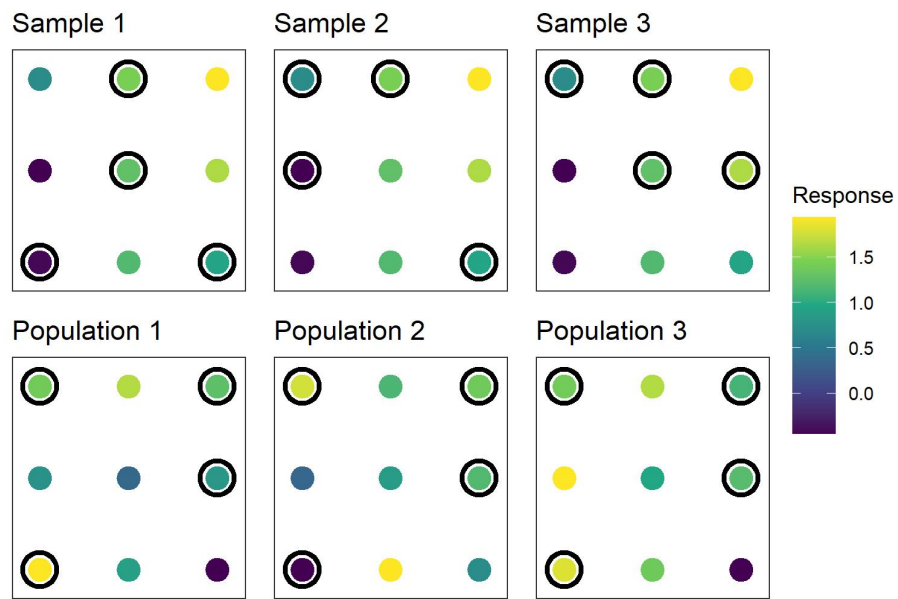
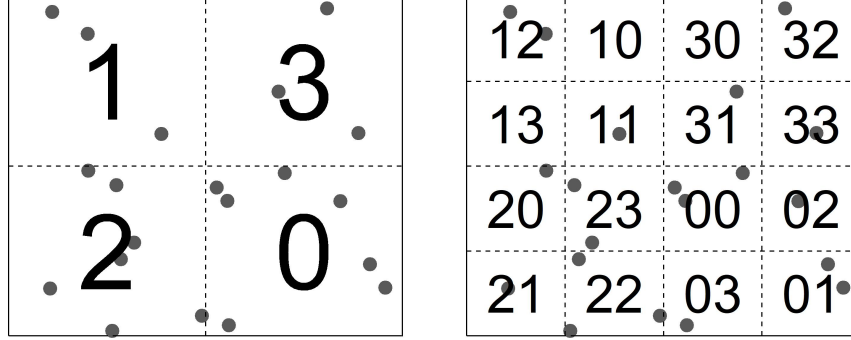


Figure 1: A visual comparison of the design-based and model-based approaches. In the top row, the design-based approach is highlighted. There is one fixed population with nine population units and three random samples of size four (points circled are those sampled). The response values at each site are fixed. In the bottom row, the model-based approach is highlighted. There are three realizations of the same data-generating stochastic process that are all sampled at the same four locations. The response values at each site are random.

from samples lacking spatial balance (Barabesi and Franceschi, 2011; Benedetti et al., 2017; Grafström and Lundström, 2013; Robertson et al., 2013; Stevens and Olsen, 2004; Wang et al., 2013). The first spatially balanced sampling algorithm to see widespread use was the generalized random tessellation stratified (GRTS) algorithm (Stevens and Olsen, 2004). After the GRTS algorithm was developed, several other spatially balanced sampling algorithms emerged, including stratified sampling with compact geographical strata (Walvoort et al., 2010), the local pivotal method (Grafström et al., 2012; Grafström and Matei, 2018), spatially correlated Poisson sampling (Grafström, 2012), balanced acceptance sampling (Robertson et al., 2013), within-sample-distance sampling (Benedetti and Piersimoni, 2017), and Halton iterative partitioning sampling (Robertson et al., 2018). In this manuscript, we select spatially balanced samples using the GRTS algorithm because it is readily available in the **spsurvey** **R** package (Dumelle et al., 2022) and naturally accommodates finite and infinite sampling frames, unequal inclusion probabilities, and replacement units. Replacement units are additional population units that can be sampled when a population unit originally selected can no longer be sampled. A couple of reasons why an originally selected site can no longer be sampled include its location being physically inaccessible or it is on private land that the researcher does not have permission to access.

The GRTS algorithm selects samples by utilizing a particular mapping between two-dimensional and one-dimensional space that preserves proximity relationships. First, the bounding box of the domain is split up into four distinct, equally sized squares called level-one cells. Each level-one cell is randomly assigned a level-one address of 0, 1, 2, or 3. The set of level-one cells is denoted by  $\mathcal{A}_1$  and defined as  $\mathcal{A}_1 \equiv \{a_1 : a_1 = 0, 1, 2, 3\}$ . Within each level-one cell, the inclusion probability for each population unit (which is





(a) Assignment of level-one cells to the spatial domain. Grey circles indicate population units. (b) Assignment of level-two cells to the spatial domain. Grey circles indicate population units.

Figure 2: Assignment of level-one and level-two cells to the spatial domain. In (a), each level-one cell is randomly given a level-one address of 0, 1, 2, or 3. In (b), each level-two cell within each level-one cell is randomly given a level-two address of 0, 1, 2, or 3.

pre-specified) is summed, and if any of these sums exceed one, a second level of cells is added. Then each level-one cell is split into four distinct, equally sized squares called level-two cells. Each level-two cell is randomly assigned a level-two address of 0, 1, 2, or 3. The set of level-two cells is denoted by  $\mathcal{A}_2$  and defined as  $\mathcal{A}_2 \equiv \{a_1 a_2 : a_1 = 0, 1, 2, 3; a_2 = 0, 1, 2, 3\}$ . The inclusion probabilities within each level-two cell are summed, and if any of these sums exceed one, a third level of cells is added. This process continues for  $k$  steps, until all level- $k$  cells have inclusion probability sums no larger than one. Then  $\mathcal{A}_k \equiv \{a_1 \dots a_k : a_1 = 0, 1, 2, 3; \dots; a_k = 0, 1, 2, 3\}$ . Figure 2 provides some intuition regarding the assignment of level-one and level-two cells.

After determining  $\mathcal{A}_k$ , the set is placed into hierarchical order. Hierarchical order is a numeric order that first sorts  $\mathcal{A}_k$  by the level-one addresses from smallest to largest, then sorts  $\mathcal{A}_k$  by the level-two addresses from smallest to largest, and so on. For example,  $\mathcal{A}_2$  in hierarchical order is the set  $\{00, 01, 02, 03, 10, \dots, 13, 20, \dots, 23, 30, \dots, 33\}$ . Because hierarchical ordering sorts by level-one cells, then level-two cells, and so on, population units that have

similar hierarchical addresses tend to be nearby one another in space. Next, each population unit is mapped to a one-dimensional line in hierarchical order where each population unit's inclusion probability equals its line-length. If a level- $k$  cell has multiple population units in it, they are randomly placed within the cell's respective line segment. A uniform random variable is then simulated in  $[0, 1]$  and a systematic sample is selected on the line, yielding  $n$  sample points for a sample size  $n$ . Each of these sample points falls on some population unit's line segment, and thus that population unit is selected in the sample. For further details regarding the GRTS algorithm, see Stevens and Olsen (2004).

After selecting a sample and collecting data, unbiased estimates of population means and totals can be obtained using the Horvitz-Thompson estimator (Horvitz and Thompson, 1952). If  $\tau$  is a population total, the Horvitz-Thompson estimator for  $\tau$ , denoted by  $\hat{\tau}_{ht}$ , is given by

$$\hat{\tau}_{ht} = \sum_{i=1}^n z_i \pi_i^{-1}, \quad (1)$$

where  $z_i$  is the value of the  $i$ th population unit in the sample,  $\pi_i$  is the inclusion probability of the  $i$ th population unit in the sample, and  $n$  is the sample size. An estimate of the population mean is obtained by dividing  $\hat{\tau}_{ht}$  by  $N$ , the number of population units.

It is also important to quantify the uncertainty in  $\hat{\tau}_{ht}$ . The Horvitz-Thompson (Horvitz and Thompson, 1952) and Sen-Yates-Grundy (Sen, 1953; Yates and Grundy, 1953) variance estimators are often used to estimate  $\text{Var}(\hat{\tau}_{ht})$ , but these estimators have two drawbacks. First, they rely on calculating  $\pi_{ij}$ , the probability that population unit  $i$  and population unit  $j$  are both in the sample – this quantity can be challenging if not impossible to calculate analytically for GRTS samples. Second, these estimators tend to ignore the spatial locations of the population units. To address these two drawbacks simultaneously, Stevens

and Olsen (2003) proposed the local neighborhood variance estimator. The local neighborhood variance estimator does not rely on  $\pi_{ij}$  and estimates the variance of  $\hat{\tau}$  conditional on the random properties of the GRTS sample – the idea being that this conditioning should yield a more precise estimate of  $\hat{\tau}$ . They show that the contribution from each sampled population unit to the overall variance is dominated by local variation. Thus the local neighborhood variance estimator is a weighted sum of variance estimates from each sampled population unit’s local neighborhood. These local neighborhoods contain the sampled population unit itself and its three nearest neighbors (among all other sampled population units). For more details, see Stevens and Olsen (2003).

### 1.1.3. Finite Population Block Kriging

Finite population block kriging (FPBK) is a model-based approach that expands the geostatistical Kriging framework to the finite population setting (Ver Hoef, 2008). Instead of developing inference based on a specific sampling design, we assume the data are generated by a spatial stochastic process. We summarize some of the basic principles of FPBK next – see Ver Hoef (2008) for technical details and see Higham, J. Ver Hoef, et al. (2021) for an extension to cases of imperfect detection among population units. Let  $\mathbf{z} \equiv \{z(s_1), z(s_2), \dots, z(s_N)\}$  be an  $N \times 1$  response vector at locations  $s_1, s_2, \dots, s_N$  that can be measured at the  $N$  population units. Suppose we want to use a sample to predict some linear function of the response variable,  $f(\mathbf{z}) = \mathbf{b}'\mathbf{z}$ , where  $\mathbf{b}'$  is a  $1 \times N$  vector of weights (e.g, the population mean is represented by a weights vector whose elements all equal  $1/N$ ). Denoting quantities that are part of the sampled population units with a subscript  $s$  and quantities that are part of the unsampled population units with a subscript  $u$ , let

$$\begin{pmatrix} \mathbf{z}_s \\ \mathbf{z}_u \end{pmatrix} = \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_u \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\delta}_s \\ \boldsymbol{\delta}_u \end{pmatrix}, \quad (2)$$

251 where  $\mathbf{X}_s$  and  $\mathbf{X}_u$  are the design matrices for the sampled and unsampled  
 252 population units, respectively,  $\boldsymbol{\beta}$  is the parameter vector of fixed effects, and  
 253  $\boldsymbol{\delta} \equiv [\boldsymbol{\delta}_s \ \boldsymbol{\delta}_u]'$ , where  $\boldsymbol{\delta}_s$  and  $\boldsymbol{\delta}_u$  are random errors for the sampled and unsampled  
 254 population units, respectively.

FPBK assumes  $\boldsymbol{\delta}$  in Equation (2) has mean-zero and a spatial dependence structure that can be modeled using a covariance function. This covariance function is commonly assumed to be non-negative, second-order stationary (depending only on the separation vector (e.g., distance) between population units), isotropic (independent of direction), and decays with distance between population units (Cressie, 1993). Henceforth, it is implied that we have made these same assumptions regarding  $\boldsymbol{\delta}$ . Chiles and Delfiner (1999), pp. 80-93 discuss covariance functions that are not second-order stationary, not isotropic, or not either. A variety of flexible covariance functions can be used to model  $\boldsymbol{\delta}$  (Cressie, 1993) – one example is the exponential covariance function. Cressie (1993) provides a thorough list of spatial covariance functions. The  $i, j$ th element of the exponential covariance matrix,  $\text{cov}(\boldsymbol{\delta})$ , is

$$\text{cov}(\delta_i, \delta_j) = \begin{cases} \sigma_1^2 \exp(-h_{i,j}/\phi) & h_{i,j} > 0 \\ \sigma_1^2 + \sigma_2^2 & h_{i,j} = 0 \end{cases}, \quad (3)$$

255 where  $\sigma_1^2$  is the variance parameter that quantifies the spatially dependent (cor-  
 256 related) variability,  $\sigma_2^2$  is the variance parameter the quantifies that spatially  
 257 independent (not correlated) variability,  $\phi$  is the distance parameter that mea-  
 258 sures the distance-decay rate of the covariance, and  $h_{i,j}$  is the Euclidean distance  
 259 between population units  $i$  and  $j$ . In geostatistical literature,  $\sigma_1^2$  is called the

260 partial sill,  $\sigma_2^2$  is called the nugget, and  $\phi$  is called the range. We denote  $\boldsymbol{\theta}$  as the  
 261 vector of covariance parameters that composes  $\boldsymbol{\delta}$ . In Equation 3,  $\boldsymbol{\theta} = \{\sigma_1^2, \sigma_2^2, \phi\}$ .

The parameters in Equation 2 can be estimated using a variety of techniques, but we focus on using restricted maximum likelihood (REML) (Harville, 1977; Patterson and Thompson, 1971; Wolfinger et al., 1994). REML is preferred over maximum likelihood (ML) because ML estimates can be badly biased for small sample sizes, due to the fact that ML makes no adjustment for the simultaneous estimation of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  (Patterson and Thompson, 1971). Minus twice the REML log-likelihood of the sampled sites is given by

$$\ln |\boldsymbol{\Sigma}| + (\mathbf{z}_s - \mathbf{X}_s \tilde{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}_{ss}^{-1} (\mathbf{z}_s - \mathbf{X}_s \tilde{\boldsymbol{\beta}}) + \ln |\mathbf{X}_s^T \boldsymbol{\Sigma}_{ss}^{-1} \mathbf{X}_s| + (n - p) \ln(2\pi), \quad (4)$$

262 where  $\tilde{\boldsymbol{\beta}} = (\mathbf{X}_s^T \boldsymbol{\Sigma}_{ss}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \boldsymbol{\Sigma}_{ss}^{-1} \mathbf{z}_s$  and  $\boldsymbol{\Sigma}_{ss}$  is the covariance matrix of the  
 263 sampled sites. Minimizing Equation 4 yields  $\hat{\boldsymbol{\theta}}_{reml}$ , the REML estimates of  
 264  $\boldsymbol{\theta}$ . Then  $\hat{\boldsymbol{\beta}}_{reml}$ , the REML estimate of  $\boldsymbol{\beta}$ , is given by  $(\mathbf{X}_s^T \hat{\boldsymbol{\Sigma}}_{ss}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \hat{\boldsymbol{\Sigma}}_{ss}^{-1} \mathbf{z}_s$ ,  
 265 where  $\hat{\boldsymbol{\Sigma}}_{ss}$  is  $\boldsymbol{\Sigma}_{ss}$  evaluated at  $\hat{\boldsymbol{\theta}}_{reml}$ .

266 With the model formulation in Equation 2, the best linear unbiased predictor  
 267 (BLUP) of  $f(\mathbf{b}'\mathbf{z})$  and its prediction variance can be computed. While details  
 268 of the derivation are in Ver Hoef (2008), we note here that the predictor and  
 269 its variance are both moment-based, meaning that they do not rely on any  
 270 distributional assumptions. Distributional assumptions are used, however, when  
 271 constructing prediction intervals.

272 Other approaches, such as k-nearest-neighbors (Fix and Hodges, 1989; Ver  
 273 Hoef and Temesgen, 2013) and random forest (Breiman, 2001), among others,  
 274 could also be used to obtain predictions for a mean or total from finite population  
 275 spatial data. Compared to the k-nearest-neighbors and random forest approach,  
 276 we prefer FPBK because it is model-based and relies on theoretically-based  
 277 variance estimators leveraging the model's spatial covariance structure, whereas

278 k-nearest-neighbors and random forests use ad-hoc variance estimators (Ver Hoef  
 279 and Temesgen, 2013). Additionally, Ver Hoef and Temesgen (2013) compared  
 280 FPBK, k-nearest-neighbors, and random forest in a variety of spatial data  
 281 contexts, and FPBK tended to perform best.

## 282 2. Materials and Methods

In this section we describe how we used simulated and real data to investigate performance between simple random sampling (SRS) and GRTS sampling as well as performance between design-based (DB) and model-based (MB) inference. In SRS and GRTS sampling, all population units had equal inclusion probabilities and were selected without replacement. The important distinction between SRS and GRTS is that SRS ignores spatial locations while sampling but GRTS explicitly incorporates them. Together, the two sampling plans (SRS and GRTS) combined with the two inference approaches (DB and MB) yielded four sampling-inference combinations: SRS-DB, SRS-MB, GRTS-DB, and GRTS-MB. For SRS-DB, the Horvitz-Thompson estimator (1) was used to estimate means and the commonly-used SRS variance formula (Lohr, 2009; Särndal et al., 2003) was used to estimate the variance. This variance formula is given by

$$\frac{f[\sum_{i=1}^n (z_i - \bar{z})^2]}{n(n-1)}, \quad (5)$$

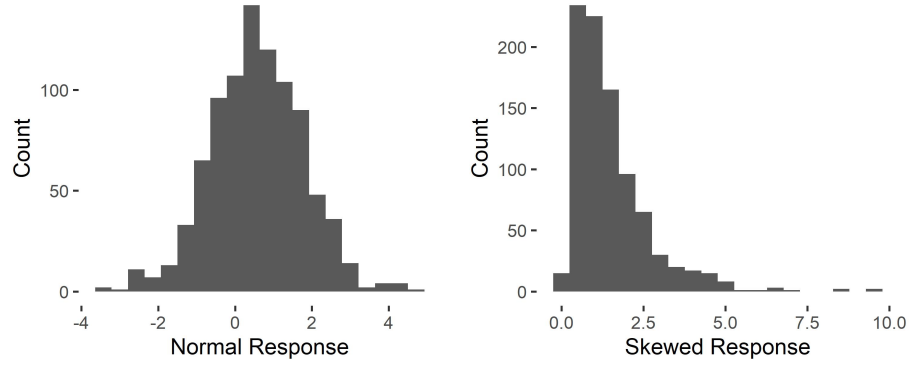
283 where  $z_i$  is the  $i$ th response value,  $\bar{z}$  is the mean of all  $z_i$ ,  $n$  is the sample  
 284 size,  $N$  is the population size, and  $f = (1 - n/N)$  ( $f$  is often called the finite  
 285 population correction factor). For GRTS-DB, the Horvitz-Thompson estimator  
 286 was used to estimate means and the local neighborhood variance was used to  
 287 estimate variances. For SRS-MB and GRTS-MB, FPBK was used to estimate  
 288 means and variances using restricted maximum likelihood. SRS, GRTS sampling,  
 289 and design-based inference were implemented using the `spsurvey` R package

(Dumelle et al., 2022). FPBK was implemented using the `sptotal` **R** package (Higham, Ver Hoef, et al., 2021).

We used simulated data to compare the sampling-inference combinations across many realized populations from the same data-generating stochastic process. We then used real data from the 2012 National Lakes Assessment to compare the sampling-inference combinations within a single realized population (which is typically the case in reality). With the simulated data, we were in control of the data-generating stochastic process and the random sampling process. With the real data, we were in control of only the random sampling process.

### 2.1. Simulated Data

We evaluated performance of the four sampling-inference combinations in 36 different simulation scenarios. The 36 scenarios resulted from the crossing of three sample sizes, two location layouts (of the population units), two response types, and three proportions of dependent random error (DRE). The three sample sizes ( $n$ ) were  $n = 50$ ,  $n = 100$ , and  $n = 200$ . Samples were always selected from a population size ( $N$ ) of  $N = 900$ . The two location layouts were random and gridded. Locations in the random layout were randomly generated inside the unit square ( $[0, 1] \times [0, 1]$ ). Locations in the gridded layout were placed on a fixed, equally spaced grid inside the unit square. The two response types were normal and skewed. For the normal response type, the response was simulated using mean-zero random errors with the exponential covariance (Equation 3) for three proportions of dependent random error (DRE): 0% DRE, 50% DRE, and 90% DRE. Recall the proportion of DRE is represented by  $\sigma_1^2/(\sigma_1^2 + \sigma_2^2)$ , where  $\sigma_1^2$  and  $\sigma_2^2$  are the DRE variance and independent random error (IRE) variance from Equation 3, respectively. The total variance,  $\sigma_1^2 + \sigma_2^2$ , was always 2. The distance parameter was always  $\sqrt{2}/3$ , chosen so that the correlation in the DRE decayed



(a) Histogram of a realized population for the normal response. (b) Histogram of a realized population for the skewed response.

Figure 3: Histograms of realized populations simulated for the normal and skewed responses using the random layout and 50% DRE.

to nearly zero at  $\sqrt{2}$ , the largest possible distance between two population units in the domain. For the skewed response type, the response was first simulated using the same approach as for the normal response type, except that the total variance was 0.6931 instead of 2. The response was then exponentiated, yielding a skewed random variable whose total variance was 2. The skewed responses were used to evaluate performance of the sampling-inference approaches for data that were not normal but were still estimated using REML, which relies on a normal log-likelihood. Figure 3 shows an example of a realized population for the normal and skewed responses using the random layout and 50% DRE.

In each of the 36 simulation scenarios, there were 2000 independent simulation trials. Within each simulation scenario and trial, IRS and GRTS samples were selected and then design-based and model-based inferences were used to estimate (design-based) or predict (model-based) the mean and construct 95% confidence (design-based) or 95% prediction (model-based) intervals. With model-based inference, covariance parameters and the mean were estimated (using REML) separately for each trial. After all 2000 trials, we summarized the long-run performance of the sampling-inference combination in each scenario by calculating



mean bias, root-mean-squared error, and interval coverage. Mean bias is taken as the average deviation between each trial's estimated (or predicted) mean ( $\hat{\mu}_i$ ) and its realized mean ( $\mu_i$ ):  $\frac{1}{n} \sum_{i=1}^{2000} (\hat{\mu}_i - \mu_i)$ , where  $i$  indexes simulation trials. Root-mean-squared error is taken as the square root of the average squared deviation between each trial's estimated (or predicted) mean and its realized mean:  $\sqrt{\frac{1}{n} \sum_{i=1}^{2000} (\hat{\mu}_i - \mu_i)^2}$ . Interval coverage is taken as the proportion of simulation trials where the realized mean is contained in its 95% confidence (or prediction) interval. These intervals are constructed using the normal distribution – justification comes from the asymptotic normality of means via the central limit theorem (under some assumptions). Quantifying these metrics is important because together, they give us an idea of the accuracy (mean bias), spread (RMSE), and validity (interval coverage) of the sampling-inference combinations.

## 2.2. National Lakes Assessment Data

The United States Environmental Protection Agency (USEPA), states, and tribes periodically conduct National Aquatic Research Surveys (NARS) to assess the water quality of various bodies of water in the conterminous United States. One component of NARS is the National Lakes Assessment (NLA), which measures various aspects of lake health and water quality. We focus on analyzing zooplankton multi-metric indices (ZMMI) and mercury concentrations in parts per billion (Hg ppb) from the 2012 NLA. For ZMMI, data were collected at 1035 unique lakes. At less than 10% of lakes, two ZMMI replicates were collected. These were averaged for the purposes of our study so that each lake had one measurement for ZMMI. For Hg ppb, data were collected at 995 unique lakes (there were no replicates). The ZMMI and Hg ppb data are shown as spatial maps and as histograms in Figure 4. The ZMMI data tend to be highest near the coasts, lowest in the Central United States, are relatively symmetric, and have a mean of 55.05. The Hg ppb data tend to be highest in the Northeastern United

States, lowest elsewhere, are skewed, and have a mean of 103.16 ppb. Also in Figure 4 are separate spatial semivariograms for ZMMI and Hg ppb. The spatial semivariogram quantifies the halved average squared differences (semivariance) of responses whose separation (distance) falls within some distance class. The spatial semivariance is closely related to the spatial covariance, and spatial semivariograms are often used to gauge the strength of spatial dependence in data. Both ZMMI and Hg ppb seem to have moderately strong spatial dependence (Figure 4), as the semivariance increases steadily with distance (meaning that observations nearby one another tend to be more similar than observations far apart from one another).

We studied performance of the four sampling-inference combinations by selecting 2000 random IRS and GRTS samples of size  $n = 50$ ,  $n = 100$ , and  $n = 200$  from the realized ZMMI and Hg ppb populations and then analyzing the samples using MB and DB inference. In total, there were six separate scenarios (two responses crossed with three sample sizes). For each sample from each scenario, parameters were estimated using the same approaches as described for simulated data. We also used the same evaluation metrics as for the simulated data: mean bias, RMSE, and interval coverage. Mean bias is taken as the average deviation between each sample's estimated (or predicted) mean ( $\hat{\mu}_i$ ) and the population mean ( $\mu$ ) (of ZMMI or Hg ppb):  $\frac{1}{n} \sum_{i=1}^{2000} (\hat{\mu}_i - \mu)$ , where  $i$  indexes simulation trials. Root-mean-squared error is taken as the square root of the average squared deviation between each sample's estimated (or predicted) mean and its population mean:  $\sqrt{\frac{1}{n} \sum_{i=1}^{2000} (\hat{\mu}_i - \mu)^2}$ . Interval coverage is taken as the proportion of simulation trials where the population mean is contained in its 95% confidence (or prediction) interval. These intervals are constructed using the normal distribution.

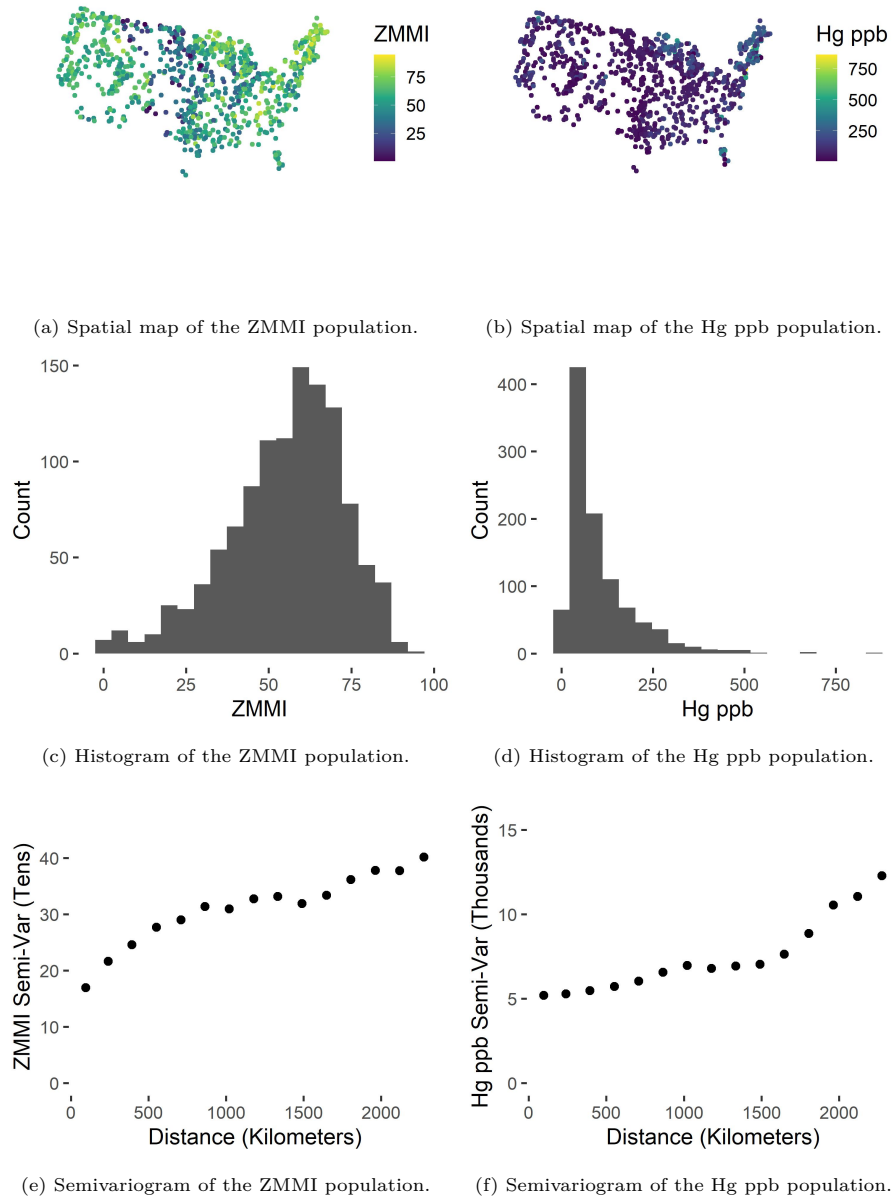


Figure 4: Exploratory graphics representing populations for the zooplankton multi-metric indices (ZMMI) and mercury concentration in parts per billion (Hg ppb) in the 2012 National Lakes Assessment (NLA) data.

### 3. Results

#### 3.1. Simulated Data

Mean bias is nearly zero for all four sampling-inference combinations in all 36 scenarios, so we omit a more detailed summary of those results here. Tables for mean bias in all 36 simulation scenarios are provided in the supporting information.

We define the relative RMSE as a ratio with numerator given by the RMSE for a sampling-inference combination and the denominator given by the RMSE for SRS-DB. Relative RMSEs for the random location layout are provided in Fig. 5. When there is no spatial covariance (Fig. 5, “DRE%: 0%”), the four sampling-inference combinations have approximately equal RMSE. In these scenarios, using GRTS sampling or model-based inference does not generally increase efficiency compared to SRS-DB. When there is spatial covariance (Fig. 5, “DRE%: 50%” and “DRE%: 90%”), GRTS-MB tends to have the lowest RMSE, followed by GRTS-DB, SRS-MB, and finally SRS-DB. As the strength of spatial covariance increases, the gap in RMSE between SRS-DB and the other sampling-inference combinations widens. Finally we note that when there is spatial covariance, SRS-MB has a much lower RMSE than SRS-DB, suggesting that the lack of efficiency from SRS is largely mitigated by model-based inference. These RMSE conclusions are similar to those observed in the grid location layout, so we omit a figure and discussion regarding the grid location layout here. Tables for RMSE in all 36 simulation scenarios are provided in the supporting information.

95% interval coverage for each of the four sampling-inference combinations in the random location layout is shown in Fig. 6. Within each simulation scenario, all sampling-inference combinations tend to have fairly similar interval coverage, though when  $n = 50$  or  $n = 100$ , GRTS-DB coverage is usually a

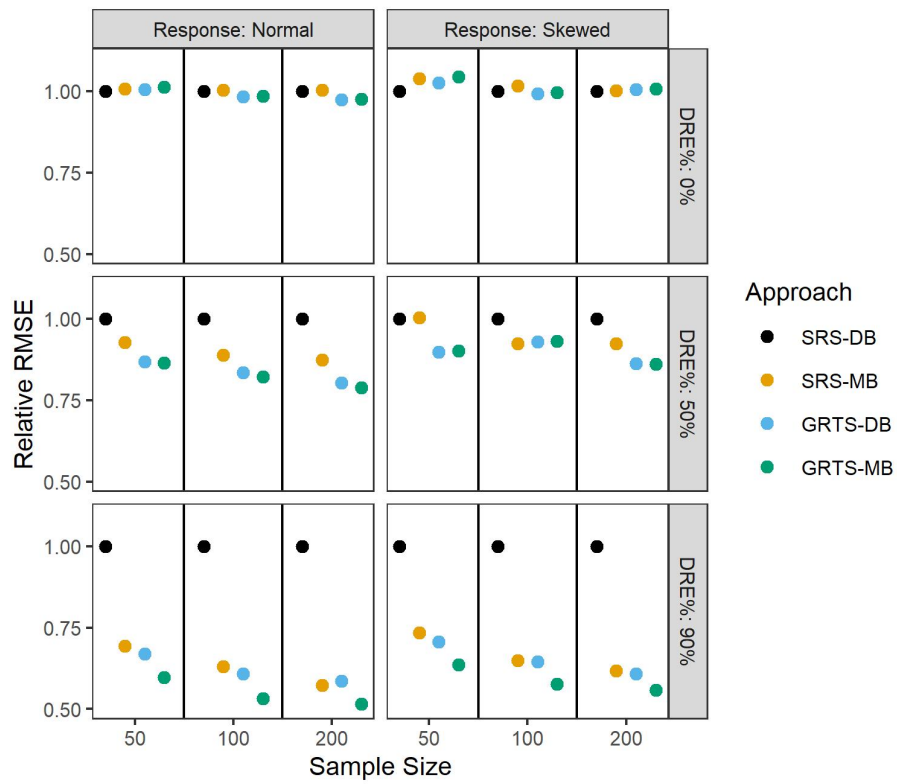


Figure 5: Simulated data relative RMSE for the four sampling-inference combinations and three sample sizes in the random location layout. The rows indicate the proportion of dependent error and the columns indicate the response type. The solid, black lines separate the sample sizes.

few percentage points lower than the other combinations, which suggests that the local neighborhood variance estimate may be slightly too small for small  $n$ . Coverage in the normal response scenarios is usually near 95%, while coverage in the skewed response scenarios usually varies from 90% to 95% but increases with the sample size. At a sample size of 200, all four sampling-inference combinations have approximately 95% interval coverage in both response scenarios for all DRE proportions. These interval coverage conclusions are similar to those observed in the grid location layout, so we omit a figure and discussion regarding the grid location layout here. Tables for interval coverage in all 36 simulation scenarios are provided in the supporting information.

### 3.2. National Lakes Assessment Data

Mean bias is nearly zero for all four sampling-inference combinations in all six scenarios, so we omit a more detailed summary of those results here. Tables for mean bias in all six simulation scenarios are provided in the supporting information.

The relative RMSE of both ZMMI (symmetric response) and Hg ppb (skewed response) for all four sampling-inference combinations are shown in Fig. 7. GRTS-MB has the lowest RMSE, followed by GRTS-DB, SRS-MB, and then SRS-DB. The difference in RMSE among GRTS-MB and GRTS-DB tends to be quite small. When  $n = 50$ , SRS-MB RMSE is approximately evenly between IRS-DB RMSE and GRTS-MB RMSE, but for the larger sample sizes ( $n = 100$ ,  $n = 200$ ), SRS-MB RMSE is closer to GRTS-MB RMSE. Lastly we note that GRTS-MB, GRTS-DB, and SRS-MB all have noticeably lower RMSE than SRS-DB. Tables for RMSE in all six scenarios are provided in the supporting information.

95% interval coverage of both ZMMI and Hg ppb for all four sampling-inference combinations is shown in Fig. 8. When  $n = 50$ , interval coverage for both responses is too low, though interval coverage is higher for ZMMI (symmetric

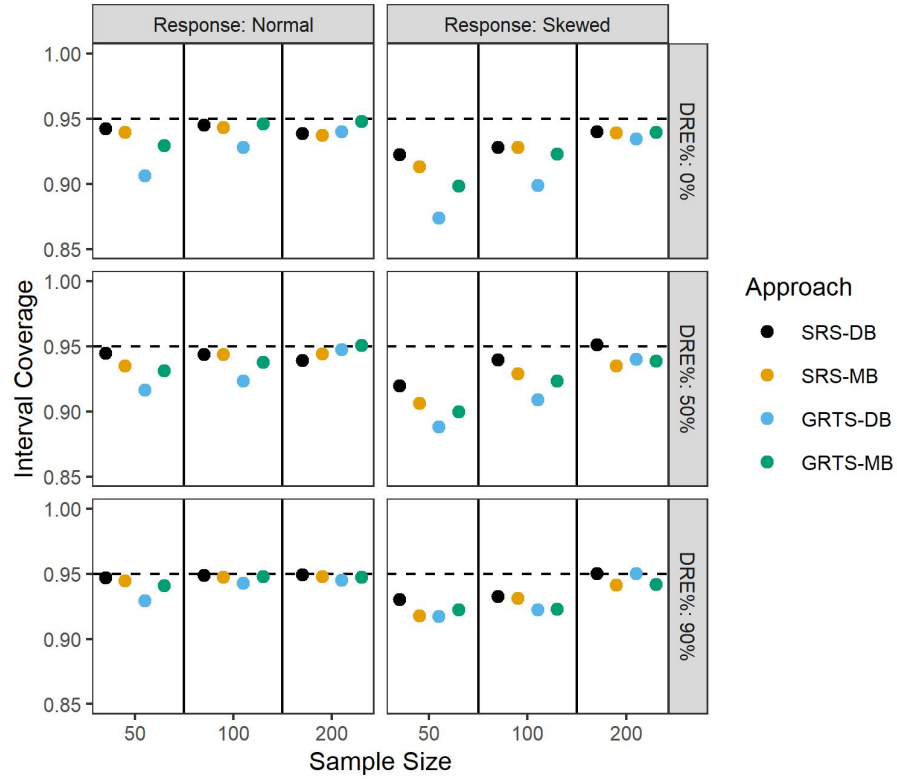


Figure 6: Simulated data interval coverage for the four sampling-inference combinations and three sample sizes in the random location layout. The rows indicate the proportion of dependent error and the columns indicate the response type. The solid black lines separate the sample sizes and the dashed black lines represent 95% coverage.

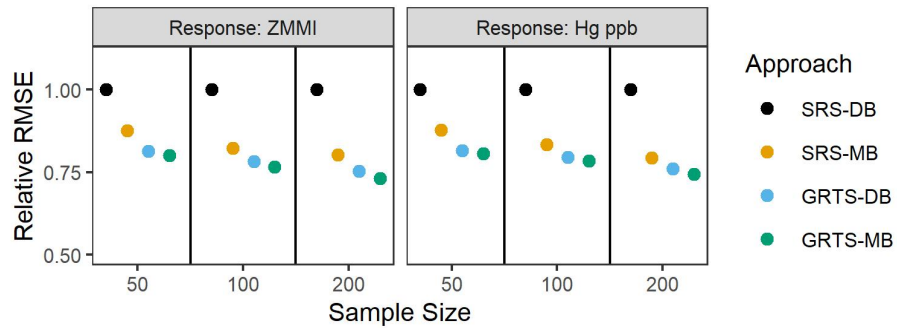


Figure 7: NLA data relative RMSE for the four sampling-inference combinations. The rows indicate the proportion of dependent error and the columns indicate the response type. The solid, black lines separate the sample sizes.

response) than for Hg ppb (skewed response). When  $n = 100$ , ZMMI interval coverage is approximately 95% except for GRTS-DB, which has coverage around 92%, while Hg ppb interval coverage ranges from approximately 90% (GRTS-DB) to 93% (GRTS-MB). When  $n = 200$ , ZMMI interval coverage is approximately 95% while Hg ppb interval coverage ranges from approximately 93% (GRTS-DB) to 95% (GRTS-MB). As with the simulated data, coverages for the NLA data tend to increase with the sample sizes, coverages tend to be higher for symmetric responses than for skewed responses, and the local neighborhood variance was slightly too small for small  $n$ , yielding slightly lower interval coverages than the other sampling-inference combinations. Recall that model-based inference defines interval coverage properties across realized populations. With the simulated data, we evaluated interval coverage across realized populations, but for the NLA data, we evaluated interval coverage within a single realized population for different samples. We did find that model-based coverages were similar to the design-based coverages, however, suggesting that for some realized populations it is reasonable to heuristically view data from separate random samples as being from approximately separate realized populations. But generally, if model-based intervals constructed from many random samples of a single realized population show improper coverage, this does not necessarily imply a deficiency in model-based inference. Tables for interval coverage in all six simulation scenarios are provided in the supporting information.

#### 4. Discussion

The design-based and model-based approaches to frequentist statistical inference rest on fundamentally different foundations. Design-based approaches rely on random sampling to estimate population parameters. Model-based approaches rely on distributional assumptions to predict realized values of a



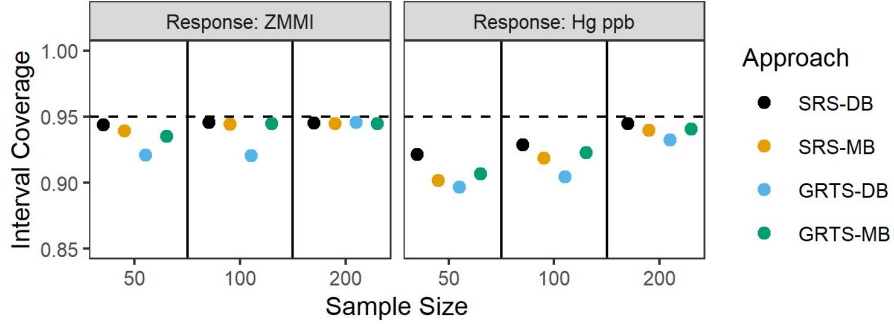


Figure 8: NLA data interval coverage for the four sampling-inference combinations. The rows indicate the proportion of dependent error and the columns indicate the response type. The solid black lines separate the sample sizes and the dashed black lines represent 95% coverage.

467 data-generating stochastic process. Though model-based approaches do not rely  
 468 on random sampling, it can still be beneficial as a way to guard against pref-  
 469 erential sampling. While design-based and model-based approaches have often  
 470 been compared in the literature from theoretical and analytical perspectives,  
 471 our contribution lies in studying them for finite population spatial data while  
 472 implementing GRTS sampling and the local neighborhood variance estimator.  
 473 Aside from the theoretical differences described throughout the manuscript, a  
 474 few analytical findings from the simulated and real data studies were particularly  
 475 notable. All sampling-inference combinations had approximately zero mean  
 476 bias. Independent of the inference approach, GRTS-DB and GRTS-MB had  
 477 lower RMSE than their SRS counterparts. Though GRTS-DB and GRTS-MB  
 478 generally had very similar RMSE, SRS-MB tended to have much lower RMSE  
 479 than SRS-DB, suggesting that the model-based inference mitigated much of the  
 480 inefficiency in RMSE from SRS. As the proportion of dependent random error  
 481 in the simulated data increased, SRS-MB, GRTS-DB, and GRTS-MB become  
 482 increasingly more efficient (lower RMSE) than SRS-DB. Interval coverage tended  
 483 to be higher for the symmetric responses than skewed responses and tended to  
 484 increase with the sample size. At a sample size of  $n = 200$ , generally all interval

485 coverages were near the desired value of 95%.

486     There are several benefits and drawbacks of the design-based and model-  
 487 based approaches for finite population spatial sampling and inference. Some we  
 488 have discussed, but others we have not, and they are worthy of consideration  
 489 in future research. First, we discuss advantages of the design-based approach.  
 490 Design-based inference is often computationally efficient, while model-based  
 491 inference can be computationally burdensome, especially for likelihood-based  
 492 estimation methods like REML that rely on the inverse of a covariance matrix.  
 493 Design-based inference easily handles binary data through a straightforward  
 494 application of the Horvitz-Thompson estimator. In contrast, analyzing binary  
 495 data using model-based inference generally requires a logistic mixed regression  
 496 model, the parameters of which can be difficult to estimate and interpret (Bolker  
 497 et al., 2009). An advantage of design-based inference is that interval coverage  
 498 is valid (has the proper coverage rate) as long as 1) the sample is sufficiently  
 499 large to ensure the statistic’s sampling distribution is approximately normal and  
 500 2) the variance estimator is consistent (Brus and De Gruijter, 1997; Särndal  
 501 et al., 2003). This is because with the design-based approach, the sampling  
 502 plan and inclusion probabilities are specified directly by the researcher. An  
 503 advantage of SRS-DB not previously mentioned is that it is likely to be valid  
 504 given the consistency of its variance estimator (Särndal et al., 2003). With  
 505 the model-based approach, however, interval coverage is unlikely to be valid  
 506 if the model assumptions made do not not accurately reflect reality. Whether  
 507 model assumptions accurately reflect reality can be a challenging and sometimes  
 508 impossible question to answer definitively.

509     Now, we discuss advantages of the model-based approach. The model-  
 510 based approach can more naturally quantify the relationship between covariates  
 511 (predictor variables) and the response variable than design-based approaches.

512 Model-based inference also yields estimated spatial covariance parameters, which  
513 help better understand the dependence structure of the process in study. Model  
514 selection is also possible using model-based inference and criteria such as cross  
515 validation, likelihood ratio tests, or AIC (Akaike, 1974). Model-based inference  
516 is capable of more efficient small-area estimation than design-based inference  
517 because model-based inference can leverage distributional assumptions in areas  
518 with few observed population units. Model-based approaches also accommodate  
519 unit-by-unit predictions at unobserved locations that can be used to construct  
520 informative visualizations like smoothed maps. Brus and De Gruijter (1997)  
521 provide a more thorough discussion regarding the benefits and drawbacks of the  
522 two approaches. In short, when deciding whether the design-based or model-  
523 based approach is more appropriate to implement, the benefits and drawbacks of  
524 each approach should be considered alongside the particular goals of the study.

525     There are many extensions of this research worthy of future consideration that  
526 include sampling with unequal inclusion probabilities, using different spatially  
527 balanced sampling approaches (instead of GRTS), using different spatial data  
528 configurations, using different spatial domains like stream networks (Ver Hoef  
529 and Peterson, 2010), using different response or covariance structures, and using  
530 spatial or external mean trends (which can be defined through covariates).

## 531 **Acknowledgments**

532     We would like to thank the editors and anonymous reviewers for hard work  
533 and time spent providing us with thoughtful, valuable feedback which greatly  
534 improved the manuscript.

535     The views expressed in this manuscript are those of the authors and do not  
536 necessarily represent the views or policies of the U.S. Environmental Protection  
537 Agency or the National Oceanic and Atmospheric Administration. Any mention

of trade names, products, or services does not imply an endorsement by the U.S. government, the U.S. Environmental Protection Agency, or the National Oceanic and Atmospheric Administration. The U.S. Environmental Protection Agency and National Oceanic and Atmospheric Administration do not endorse any commercial products, services, or enterprises.

#### **Conflict of Interest Statement**

There are no conflicts of interest for any of the authors.

#### **Author Contribution Statement**

All authors conceived the ideas; All authors designed the methodology; MD and MH performed the simulations and analyzed the data; MD and MH led the writing of the manuscript; All authors contributed critically to the drafts and gave final approval for publication.

#### **Data and Code Availability**

This manuscript has a supplementary **R** package that contains all of the data and code used in its creation. The supplementary **R** package is hosted on GitHub. Instructions for download are available at

<https://github.com/michaeldumelle/DvMsp>.

If the manuscript is accepted, this repository will be archived in Zenodo.

#### **Supporting Information**

In the supporting information, we provide tables of summary statistics for all 36 simulation scenarios and all six real data scenarios.

## References

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Barabesi, L., Franceschi, S., 2011. Sampling properties of spatial total estimators under tessellation stratified designs. *Environmetrics* 22, 271–278.
- Benedetti, R., Piersimoni, F., 2017. A spatially balanced design with probability function proportional to the within sample distance. *Biometrical Journal* 59, 1067–1084.
- Benedetti, R., Piersimoni, F., Postiglione, P., 2017. Spatially balanced sampling: A review and a reappraisal. *International Statistical Review* 85, 439–454.
- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in ecology & evolution* 24, 127–135.
- Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- Brus, D., De Gruijter, J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma* 80, 1–44.
- Brus, D.J., 2021. Statistical approaches for spatial sample survey: Persistent misconceptions and new developments. *European Journal of Soil Science* 72, 686–703.
- Brus, D.J., DeGruijter, J.J., 1993. Design-based versus model-based estimates of spatial means: Theory and application in environmental soil science. *Environmetrics* 4, 123–152.
- Chan-Golston, A.M., Banerjee, S., Handcock, M.S., 2020. Bayesian inference for finite populations under spatial process settings. *Environmetrics* 31, e2606.

- Chiles, J.-P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons, New York.
- Cicchitelli, G., Montanari, G.E., 2012. Model-assisted estimation of a spatial population mean. *International Statistical Review* 80, 111–126.
- Cooper, C., 2006. Sampling and variance estimation on continuous domains. *Environmetrics* 17, 539–553.
- Cressie, N., 1993. Statistics for spatial data. John Wiley & Sons.
- De Gruijter, J., Ter Braak, C., 1990. Model-free estimation from spatial samples: A reappraisal of classical sampling theory. *Mathematical Geology* 22, 407–415.
- Diggle, P.J., Menezes, R., Su, T.-l., 2010. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59, 191–232.
- Dumelle, M., Kincaid, T.M., Olsen, A.R., Weber, M.H., 2022. Spsurvey: Spatial sampling design and analysis.
- Fix, E., Hodges, J.L., 1989. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique* 57, 238–247.
- Grafström, A., 2012. Spatially correlated poisson sampling. *Journal of Statistical Planning and Inference* 142, 139–147.
- Grafström, A., Lundström, N.L., 2013. Why well spread probability samples are balanced. *Open Journal of Statistics* 3, 36–41.
- Grafström, A., Lundström, N.L., Schelin, L., 2012. Spatially balanced sampling through the pivotal method. *Biometrics* 68, 514–520.
- Grafström, A., Matei, A., 2018. Spatially balanced sampling of continuous populations. *Scandinavian Journal of Statistics* 45, 792–805.
- Hansen, M.H., Madow, W.G., Tepping, B.J., 1983. An evaluation of model-

613 dependent and probability-sampling inferences in sample surveys. *Journal of the*  
614 *American Statistical Association* 78, 776–793.

615 Harville, D.A., 1977. Maximum likelihood approaches to variance compo-  
616 nent estimation and to related problems. *Journal of the American Statistical*  
617 *Association* 72, 320–338.

618 Higham, M., Ver Hoef, J., Frank, B., Dumelle, M., 2021. Sptotal: Predicting  
619 totals and weighted sums from spatial data.

620 Higham, M., Ver Hoef, J., Madsen, L., Aderman, A., 2021. Adjusting a finite  
621 population block kriging estimator for imperfect detection. *Environmetrics* 32,  
622 e2654.

623 Hofman, S.C., Brus, D., 2021. How many sampling points are needed to  
624 estimate the mean nitrate-n content of agricultural fields? A geostatistical  
625 simulation approach with uncertain variograms. *Geoderma* 385, 114816.

626 Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling with-  
627 out replacement from a finite universe. *Journal of the American Statistical*  
628 *Association* 47, 663–685.

629 Lohr, S.L., 2009. *Sampling: Design and analysis*. Nelson Education.

630 Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information  
631 when block sizes are unequal. *Biometrika* 58, 545–554.

632 Robertson, B., Brown, J., McDonald, T., Jaksons, P., 2013. BAS: Balanced  
633 acceptance sampling of natural resources. *Biometrics* 69, 776–784.

634 Robertson, B., McDonald, T., Price, C., Brown, J., 2018. Halton iterative  
635 partitioning: Spatially balanced sampling via partitioning. *Environmental and*  
636 *Ecological Statistics* 25, 305–323.

637 Särndal, C.-E., Swensson, B., Wretman, J., 2003. *Model assisted survey*  
638 *sampling*. Springer Science & Business Media.

639 Schabenberger, O., Gotway, C.A., 2017. *Statistical methods for spatial data*

640 analysis. CRC press.

641 Sen, A.R., 1953. On the estimate of the variance in sampling with varying  
642 probabilities. *Journal of the Indian Society of Agricultural Statistics* 5, 127.

643 Sterba, S.K., 2009. Alternative model-based and design-based frameworks  
644 for inference from samples to populations: From polarization to integration.  
645 *Multivariate Behavioral Research* 44, 711–740.

646 Stevens, D.L., Olsen, A.R., 2003. Variance estimation for spatially balanced  
647 samples of environmental resources. *Environmetrics* 14, 593–610.

648 Stevens, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural  
649 resources. *Journal of the American Statistical Association* 99, 262–278.

650 USEPA, 2012. National lakes assessment 2012. [https://www.epa.gov/national-](https://www.epa.gov/national-aquatic-resource-surveys/national-results-and-regional-highlights-national-lakes-assessment)  
651 [aquatic-resource-surveys/national-results-and-regional-highlights-national-lakes-](https://www.epa.gov/national-aquatic-resource-surveys/national-results-and-regional-highlights-national-lakes-assessment)  
652 [assessment](https://www.epa.gov/national-aquatic-resource-surveys/national-results-and-regional-highlights-national-lakes-assessment).

653 Ver Hoef, J., 2002. Sampling and geostatistics for spatial data. *Ecoscience* 9,  
654 152–161.

655 Ver Hoef, J.M., 2008. Spatial methods for plot-based sampling of wildlife  
656 populations. *Environmental and Ecological Statistics* 15, 3–13.

657 Ver Hoef, J.M., Peterson, E.E., 2010. A moving average approach for spatial  
658 statistical models of stream networks. *Journal of the American Statistical*  
659 *Association* 105, 6–18.

660 Ver Hoef, J.M., Temesgen, H., 2013. A comparison of the spatial linear  
661 model to nearest neighbor (k-nn) methods for forestry applications. *PIOS ONE*  
662 8, e59129.

663 Walvoort, D.J., Brus, D., De Gruijter, J., 2010. An r package for spatial  
664 coverage sampling and random sampling from compact geographical strata by  
665 k-means. *Computers & geosciences* 36, 1261–1267.

666 Wang, J.-F., Jiang, C.-S., Hu, M.-G., Cao, Z.-D., Guo, Y.-S., Li, L.-F., Liu, T.-



- 667 J., Meng, B., 2013. Design-based spatial sampling: Theory and implementation.  
668 Environmental Modelling & Software 40, 280–288.
- 669     Wolfinger, R., Tobias, R., Sall, J., 1994. Computing gaussian likelihoods and  
670 their derivatives for general linear mixed models. SIAM Journal on Scientific  
671 Computing 15, 1294–1310.
- 672     Yates, F., Grundy, P.M., 1953. Selection without replacement from within  
673 strata with probability proportional to size. Journal of the Royal Statistical  
674 Society: Series B (Methodological) 15, 253–261.