Bayesian Inference for Finite Populations Under Spatial Process Settings

Alec M. Chan-Golston

UCLA Department of Biostatistics

aleccg@ucla.edu

Sudipto Banerjee

UCLA Department of Biostatistics

sudipto@ucla.edu

Mark S. Handcock

UCLA Department of Statistics

handcock@stat.ucla.edu

November 20, 2019

Abstract

We develop a Bayesian model-based approach to finite population estimation accounting for spatial dependence. Our innovation here is a framework that achieves inference for finite population quantities in spatial process settings. A key distinction from the small area estimation setting is that we analyze finite populations referenced by their geographic coordinates. Specifically, we consider a two-stage sampling design in which the primary units are geographic regions, the secondary units are point-referenced locations, and the measured values are assumed to be a partial realization of a spatial process. Estimation of finite population quantities from geostatistical models do not account for sampling designs, which can impair inferential performance, while design-based estimates ignore the spatial dependence in the finite population. We demonstrate using simulation experiments that process-based finite population sampling models improve model fit and inference over models that fail to account for spatial correlation. Furthermore, the process based models offer richer inference with spatially interpolated maps over the entire region. We reinforce these improvements and demonstrate scaleable inference for ground-water nitrate levels in the population of California Central Valley wells by offering estimates of mean nitrate levels and their spatially interpolated maps.

Key words: Finite population inference; Bayesian modeling; Spatial process; Two-stage sampling; Hierarchical models.

I. Introduction

Finite population survey sampling concerns statistical modeling and inference on finite populations from sampling designs; see, for example, Cochran (1977), Hartley and Sielken Jr. (1975), Royall (1970), and Horvitz and Thompson (1952). In this article, we will concern ourselves with Bayesian inference for finite populations when the sampling units are spatially oriented. For instance, one may consider estimating the total biomass in a forest given a sample of trees, the average income of a city given a sample of individuals and their addresses, or the total amount of air pollution attributable to cars on a freeway given a sample of pollution measurements. Additionally, these techniques could be used to conduct Bayesian inference for finite population survey sampling is discussed in great detail in Gelman (2007), Little (2004), Ghosh and Meeden (1997), and Ericson (1969). In this domain, there is a substantial literature on small area estimation for regionally aggregated data (see, e.g., Rao, 2003; Ghosh et al., 1998; Ghosh and Rao, 1994; Clayton and Kaldor, 1987), where interest lies in modeling dependencies across regions.

Unlike the aforementioned literature on small area estimation, where the sampling units are regions such as counties, states or census-tracts, spatial process models consider quantities that, at least conceptually, exist in continuum over the entire domain. The process assigns a probability law to an uncountable subset within a d-dimensional Euclidean domain. In general, spatial process modeling (Banerjee et al. 2014; Cressie and Wikle 2011; and Ripley 2004) follows the generic paradigm

$$[\text{data} \,|\, \text{process}] \times [\text{process} \,|\, \text{parameters}] \times [\text{parameters}] \,, \tag{1}$$

which accommodates complex dependencies and multiple sources of variation.

With regard to finite population sampling in spatial process settings, the literature appears to be considerably more scant than small area estimation. Here, Ver Hoef (2002) discuss connections between geostatistical models and classical design-based sampling and develop methods for executing model-based block kriging. Cicchitelli and Montanari (2012) present a spline regression model-assisted, design-based estimator of the mean for use on a random sample from both finite and infinite spatial populations. A linear spatial interpolator is used by Bruno et al. (2013) to create a design-based predictor of values at unobserved locations which outperforms non-spatial predictors. While related to these developments, none of these techniques have presented a Bayesian approach. In this manuscript, we pursue a fully Bayesian, model-based approach as in (1) and carry out inference on the finite population quantities and the spatial process.

Bayesian finite population survey sampling is essentially model-based (see, e.g., Little 2004). The population units are themselves assumed to be endowed with a probability distribution. In a Gaussian setting, Scott and Smith (1969) devised Bayesian hierarchical models for inferring with two-stage designs, while Malec and Sedransk (1985) extended this framework to general multi-stage (more than two-stages) models and also discussed handling unknown variances. Our current contribution focuses on incorporating survey sampling designs within (1). We extend this framework to spatial process settings under the context of ignorable sampling designs (Rubin 1976; Sugden and Smith 1984), where the probability of element selection is assumed independent of the measured outcome given the design variables. We specifically develop the distribution theory and algorithms for implementing (1) in the context of two-stage designs that encompass simple random, cluster and stratified sampling (as defined in Cochran, 1977) as special cases. Extension of this work to multi-stage present no new methodological difficulties, building upon Malec and Sedransk (1985).

The remainder of this paper evolves as follows. In Section II, we review a general framework for Bayesian modeling for multi-stage sampling and how simple, two-stage, and stratified random sampling designs arise as special cases. Section III presents modeling strategies for spatially

correlated data sampled using a two-stage design, the implementation of which, along with the model proposed by Scott and Smith (1969), is discussed in Section IV using Bayesian exact and Markov chain Monte Carlo (MCMC) sampling. These models are then applied to simulated data in Section V and then used in an analysis of nitrate levels in California groundwater in Section VI. The paper concludes with a brief discussion of the results in Section VII.

II. Bayesian modeling of multi-stage sampling

Suppose that n samples are randomly drawn from a finite population of size N, $n \leq N$ and for the i-th sampled unit, the outcome y_i is measured. Without loss of generality, let the set of outcomes from the finite population be stacked into a vector $y = [y_s^\top : y_{ns}^\top]^\top$, where $y_s = [y_1, \ldots, y_n]^\top$ and $y_{ns} = [y_{n+1}, \ldots, y_N]^\top$ are vectors of outcome values from the sampled and nonsampled units, respectively. This vector has corresponding design matrix $X = [X_s^\top : X_{ns}^\top]^\top$, which observed for the entire finite population and denotes group membership.

From a superpopulation perspective, we consider the finite population to be a random sample of size N from an infinitely large population. This superpopulation is assumed to follow a Gaussian distribution with mean ν and a covariance function defined by parameters θ. In general, we can construct the following linear regression model:

$$\begin{bmatrix} y_s \\ y_{ns} \end{bmatrix} = \begin{bmatrix} X_s \\ X_{ns} \end{bmatrix} \beta + \begin{bmatrix} \epsilon_s \\ \epsilon_{ns} \end{bmatrix} \;;\; \begin{bmatrix} \epsilon_s \\ \epsilon_{ns} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_s(\theta) & V_{s,ns}(\theta) \\ V_{ns,s}(\theta) & V_{ns}(\theta) \end{bmatrix} \right). \tag{2}$$

Bayesian specifications further model $\beta \sim N(A\nu, V_\beta)$, where A and ν are a vector with length equal to the number of groups and a scalar, respectively, and V_β is the variance of β. The hierarchy continues with probabilistic specifications on ν and θ.

Our goal is to estimate linear finite population quantities of the form $\alpha^\top y$, where α is a given, fixed vector of weights defined for the entire population. Suppose $\nu \sim N(0, \gamma^2)$. Define $V_{\beta \mid y_s} = [(\gamma^2 A A^\top + V_\beta)^{-1} + X_s^\top V_s^{-1} X_s]^{-1}$ and $Q = X_{ns} - V_{ns,s} V_s^{-1} X_s$. Fixing the variance parameters, the posterior expectation of the finite population quantity is:

$$E[\alpha^\top y \mid y_s] = \{\alpha_s^\top + \alpha_{ns}^\top [V_{ns,s} + \alpha_{ns}^\top Q V_{\beta \mid y_s} X_s^\top] V_s^{-1}\} y_s .$$

Defining $B_V = V_{ns,s} + Q V_{\beta \mid y_s} X_s^\top$, the variance of this expectation is:

$$\text{Var}[E[\alpha^\top y \mid y_s]] = \alpha_s^\top V_s \alpha_s + 2\alpha_s^\top B_V^\top \alpha_{ns} + \alpha_{ns}^\top B_V V_s^{-1} B_V^\top \alpha_{ns} .$$

Additionally, the posterior variance of $\alpha^\top y$ is:

$$\text{Var}[\alpha^\top y \mid y_s] = \alpha_{ns}^\top (Q V_{\beta \mid y_s} Q^\top + V_{ns} - V_{ns,s} V_s^{-1} V_{s,ns}) \alpha_{ns} .$$

For the special case of a census, in which all members of the population are sampled, e.g. $y_s = y$, the conditional expectation of the finite population quantity is finite population consistent, in the sense that $E[\alpha^\top y \mid y_s] = E[\alpha^\top y \mid y] = \alpha^\top y$. Different sampling designs can incorporated by appropriately structuring the sampled and nonsampled elements. We provide a few examples below. All derivations can be located in the supplementary materials.

Example 1. *Simple Random Sampling*
    In simple random sampling, n units are randomly drawn from a population of size N, where each unit in the population is independent and identically distributed with mean μ and variance σ^2.

To express this as in (2), define $y_s = [y_1, \ldots, y_n]^\top$ and $y_{ns} = [y_{n+1}, \ldots, y_N]^\top$, with corresponding design matrices $X_s = 1_n$ and $X_{ns} = 1_{N-n}$, respectively, where 1_n represents the $n \times 1$ vector of ones. Let $A = 1$ and take β to be a scalar μ with mean $\nu = 0$ and variance $V_\beta = \xi^2$. Additionally, let $V_s = \sigma^2 I_n$, $V_{ns} = \sigma^2 I_{N-n}$, and $V_{s,ns} = V_{ns,s}^\top = O$, where O is a matrix of zeroes of appropriate order. Define finite population weights $\alpha = [\alpha_1, \ldots, \alpha_N]^\top$. Fixing the variance parameters, the posterior expectation of $\alpha^\top y$ is:

$$\mathrm{E}[\alpha^\top y \,|\, y_s] = \sum_{i=1}^{n} \left(\alpha_i + \frac{\sum_{i=n+1}^{N} \alpha_i}{\frac{\sigma^2}{\xi^2} + n} \right) y_i \tag{3}$$

$\square$

Example 2. *Two-Stage Sampling*

In a more complex case, suppose that the population is divided into N distinct groups defined by geography or other characteristics, with the i-th group of size M_i. Assume that within the i-th group, each unit is independent and identically distributed with mean μ_i and variance σ_i^2. The group means $\mu_1, \ldots, \mu_N$ are independent and follow a normal distribution centered at ν with a variance of δ^2, hence $\beta = \mu = [\mu_1, \ldots, \mu_N]^\top$, $A = 1_N$ and $V_\beta = \delta^2 I_N$. Suppose that only n of the N groups are randomly sampled, where $n \leq N$. Without loss of generality, take the first n groups to be sampled, and then within the chosen i-th group, m_i units are randomly selected; $m_i \leq M_i$, $i = 1, \ldots, n$. As $N - n$ groups are not sampled, the number of observed units in these groups are zero, e.g. $m_i = 0$, $i = n + 1, \ldots, N$. Hence, we can define the number of sampled units as $k = \sum_{i=1}^{N} m_i = \sum_{i=1}^{n} m_i$, the number of unsampled units as $K = \sum_{i=1}^{N}(M_i - m_i)$, and the population total to be $T = K + k = \sum_{i=1}^{N} M_i$.

To examine (2) in the context of a two-stage design, define outcome vectors $y_s = [y_1^\top, \ldots, y_N^\top]^\top$ and $y_{ns} = [y_1^{*\top}, \ldots, y_N^{*\top}]^\top$, with i^{th} components $y_i = [y_{i1}, \ldots, y_{im_i}]^\top$ and $y_i^* = [y_{im_i+1}, \ldots, y_{iM_i}]^\top$, respectively. The design matrix for the sampled units can be modified by fixing the $k \times N$ matrix $X_s = [\oplus_{i=1}^{n} 1_{m_i} : O]$, reflecting that $N - n$ of the N sites are unobserved. Similarly, for the unobserved units, define X_{ns} as a block diagonal, $K \times N$ matrix with upper block $[\oplus_{i=1}^{n} 1_{M_i-m_i}]$ and lower block $[\oplus_{i=n+1}^{N} 1_{M_i-m_i}]$. For notational convenience, we also define $X_{s1} = [\oplus_{i=1}^{n} 1_{m_i}]$ and divide the group mean vector μ into sampled, $\mu_s = [\mu_1, \ldots, \mu_n]^\top$, and nonsampled, $\mu_{ns} = [\mu_{n+1}, \ldots, \mu_N]^\top$, components such that $\mu = [\mu_s^\top, \mu_{ns}^\top]^\top$. Note that distributional mean of y_s, $X_s 1_N \mu$, can be simplified to $X_{s1} 1_n \mu_s$, as the mean of the sampled units does not depend on μ_{ns}. Define the sampled and nonsampled covariance matrices to be $V_s = V_s^{(\sigma)} = [\oplus_{i=1}^{n} \sigma_i^2 I_{m_i}]$ and $V_{ns} = V_{ns}^{(\sigma)} = [\oplus_{i=1}^{N} \sigma_i^2 I_{M_i-m_i}]$, respectively, and set $V_{s,ns} = V_{ns,s}^\top = O$. Additionally, define

$$V^{(\sigma)} = \begin{bmatrix} V_s^{(\sigma)} & O \\ O & V_{ns}^{(\sigma)} \end{bmatrix}.$$

To make this model fully Bayesian, let $\nu \sim N(0, \gamma^2)$ and $\delta^2 \sim IG(a, b)$. As our interest lies in estimating $\alpha^T y$, we can derive the posterior distributions of $p(\delta^2 \,|\, y_s)$ and $p(\nu \,|\, y_s)$ for exact sampling of the superpopulation parameters, the details of which are provided in Section II. This approach yields results similar to those derived by Scott and Smith (1969), but has the added strength of including a prior distribution on ν, and Ghosh and Meeden (1997), who replaced distributional assumptions with the assumption of posterior linearity and fixed the variance parameters.

In the two-stage case, for a set of weights $\alpha = [\alpha_{11}, \ldots, \alpha_{NM_N}]^\top$, define the group mean of sampled units as $\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$, $i = 1, \ldots, n$, and the group weight of nonsampled units as $\alpha_i = \sum_{j=m_i+1}^{M_i} \alpha_{ij}$, $i = 1, \ldots, N$. Also, let $\tilde{\gamma}^2 = \gamma^2/\delta^2$ and define $\lambda_i = \delta^2/(\delta^2 + \sigma_i^2/m_i)$ if

4

$i \in \{1, \ldots, n\}$ and $\lambda_i = 0$ if $i \in \{n+1, \ldots, N\}$. Fixing all variance parameters, the expected value of the finite population estimate is

$$\mathbb{E}[\alpha^\top y \mid y_s] = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(\alpha_{ij} + \left[\alpha_i + \frac{\sum_{i=1}^{N} \alpha_i(1-\lambda_i)}{1/\tilde{\gamma}^2 + \sum_{i=1}^{n} \lambda_i} \right] \frac{\lambda_i}{m_i} \right) y_{ij} \tag{4}$$

Additionally, the two-stage case can be extended to a three-stage case by assuming that the jth element of the ith group has m_{ij}^* subelements. Malec and Sedransk (1985) derive posterior distributions for the means for a three-stage sampling scheme and provide a framework to extend this to data with t stages of sampling. $\square$

Example 3. *Stratified Random Sampling*

Stratified sampling is a special case of two-stage sampling where all groups are sampled (e.g. $n = N$ and $m_i > 0$, $i = 1, \ldots, N$), and, therefore, considering the same population described in Example 2, the number of sampled units is $k = \sum_{i=1}^{N} m_i$, the number of nonsampled units is $K = \sum_{i=1}^{N}(M_i - m_i)$, and the population total is again $T = K + k = \sum_{i=1}^{N} M_i$. Thus, to express this design as (2), let $y_s = [y_1^\top, \ldots, y_N^\top]^\top$ and $y_{ns} = [y_1^{*\top}, \ldots, y_N^{*\top}]^\top$, with i^{th} components, $y_i = [y_{i1}, \ldots, y_{im_i}]^\top$ and $y_i^* = [y_{im_i+1}, \ldots, y_{iM_i}]^\top$, respectively. To reflect a membership to one of N groups, we take $X_s = [\oplus_{i=1}^{N} 1_{m_i}]$, $X_{ns} = [\oplus_{i=1}^{N} 1_{M_i-m_i}]$, $\beta = \mu = [\mu_1, \ldots, \mu_N]^\top$, and $A = 1_N$. The variance components also reflect this and are defined as $V_\beta = \delta^2 I_N$, $V_s = [\oplus_{i=1}^{N} \sigma_i^2 I_{m_i}]$, $V_{ns} = [\oplus_{i=1}^{N} \sigma_i^2 I_{M_i-m_i}]$, and $V_{s,ns} = V_{ns,s}^\top = O$.

The posterior expectation of $\alpha^\top y$ is given by (4), noting that $n = N$ and $\lambda_i = \frac{\delta^2}{\delta^2 + \sigma_i^2/m_i}$, $i = 1, \ldots, N$, is well-defined as $m_i > 0$ for all i. In fact, if non-informative priors are taken for the means, e.g. $\gamma^2 \to \infty$ and $\delta^2 \to \infty$, then $\lambda_i \to 1$, $i = 1, \ldots, n$ and the stratified finite population mean is $\mathbb{E}\left[\frac{1}{T} 1_T^\top y \mid y_s \right] = \sum_{i=1}^{N} \frac{M_i}{T} \bar{y}_i$, (see, e.g. Little, 2004). $\square$

III. BAYESIAN SPATIAL PROCESS MODELING FOR MULTI-STAGE SAMPLING

Data believed to be correlated as a function of geographic distance is typically described using a spatial process model. The data is assumed to be a partial realization of a Gaussian process with dependencies between elements defined by an isotropic covariance function, $C(d)$, where d is the distance between any two points. Several choices for $C(d)$ are available (see, e.g. Banerjee et al., 2014), but a versatile family is the Matérn, defined as $C(d_{ab}) = \sigma^2 + \tau^2$ if $d_{ab} = 0$ and $C(d_{ab}) = \tau^2 \frac{2^{1-\eta}}{\Gamma(\eta)} \left(\sqrt{2\eta} d_{ab} \phi \right)^\eta K_\eta \left(\sqrt{2\eta} d_{ab} \phi \right)$ if $d_{ab} > 0$, where $K_\eta(\cdot)$ is the modified Bessel function, d_{ab} is the distance between two locations ℓ_a and ℓ_b. Here σ^2 captures variation due to measurement error or micro-resolution variation, τ^2 is the spatial variance, ϕ is a decay parameter which determines the rate of decline in spatial association, and η is a smoothness parameter. The exponential covariance function is a special case of Matérn when $\eta = 1/2$. In this specific instance, the decay parameter is used to calculate the effective spatial range, which is the distance where spatial correlation between two points drops below 0.05.

Extending Example 2 to a geographic context, our spatial domain comprises N regions. Let ℓ_{ij} denote the j-th location in region i. The finite population is described by values $y(\ell_{ij})$, $i = 1, \ldots, N$ and $j = 1, \ldots, M_i$. Let y_s be the $k \times 1$ vector corresponding to measurements from the sampled locations and y_{ns} be the $K \times 1$ vector of unsampled measurements. Consider the following spatial regression model for the two-stage finite population,

$$y(\ell_{ij}) = \mu(\ell_{ij}) + \omega(\ell_{ij}) + \epsilon(\ell_{ij}) \; ; \omega \sim N(0, \Omega) \; ; \epsilon \sim N(0, V^{(\sigma)}) \, , \tag{5}$$

where $\mu(\ell_{ij})$ is the mean of the outcome at ℓ_{ij}, $\omega = [\omega_s^\top : \omega_{ns}^\top]^\top$ and $\epsilon = [\epsilon_s^\top : \epsilon_{ns}^\top]^\top$ are $T \times 1$ vectors formed by stacking up $\omega(\ell_{ij})$'s and $\epsilon(\ell_{ij})$'s, respectively (analogous to y in Example 2). Here, ω accounts for spatial effects and Ω is the $T \times T$ spatial covariance matrix constructed with $C(d_{ab})$ and is partitioned as $\Omega = \begin{bmatrix} \Omega_s & \Omega_{s,ns} \\ \Omega_{ns,s} & \Omega_{ns} \end{bmatrix}$. Introducing spatial effects in Example 2 yields $\mu(\ell_{ij}) = \mu_i$, $V_s = \Omega_s + V_s^{(\sigma)}$, $V_{ns} = \Omega_{ns} + V_{ns}^{(\sigma)}$, and $V_{s,ns} = V_{ns,s}^\top = \Omega_{s,ns}$ in (2). This also accommodates spatial versions of Examples1 and 3 by setting $N = 1$ and $N = n$, respectively.

Analogous to (4), the posterior estimate of a linear function of the population values is

$$
\mathrm{E}[\alpha^\top y \,|\, y_s] = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(\alpha_{ij} + \alpha_{ns}^\top \left\{ \Omega_{ns,s} Q_s + [X_{ns} - \Omega_{ns,s} Q_s^{-1} X_s] \times \left(\tfrac{1}{\delta^2} I_N + X_s^\top Q_s^{-1} X_s \right)^{-1} \right.\right.
$$
$$
\left.\left. \times \left[X_s^\top Q_s^{-1} + \frac{\tfrac{1}{\delta^2} 1_N 1_N^\top X_s^\top (\delta^2 X_s X_s^\top + Q_s)^{-1}}{\tfrac{1}{\gamma^2} + \sum_{i=1}^{n} \lambda_i^*} \right] \right\} q_{ij} \right) y(\ell_{ij}) \,,
$$
(6)

where $\lambda^{*\top} = [\lambda_1^* \ldots, \lambda_N^*] = 1_N^\top X_s^\top (\delta^2 X_s X_s^\top + \Omega_s + V_s^{(\sigma)})^{-1} X_s$, $Q_s = \Omega_s + V_s^{(\sigma)}$, and q_{ij} is a set of k indicator vectors of length k, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. For $i = 1$, $q_{1j} = j$ and 0 elsewhere, and if $i > 1$, q_{ij} is 1 at element $\sum_{i=1}^{i-1} m_i + j$ and 0 elsewhere. This two-stage spatial model, (5), can be written as an intercept-only spatial model by setting $\mu(\ell_{ij}) = \mu$ and $\sigma_i^2 = \sigma^2$, $i = 1, \ldots, N$, i.e., simplifying $V^{(\sigma)}$ to $\sigma^2 I_T$. As region is not accounted for, the design matrices X_s and X_{ns} are replaced with 1_k and 1_K, respectively.

However, as the size of the finite population, T, grows, the scaleability of (5) diminishes due to an increased computational burden stemming from the inversion of the $T \times T$ matrix Ω. To address this, we also consider a more computationally efficient model which also allows for region specific means, but specifies that each region is defined by its own process parameters and is independent from all other regions. To reflect this regional independence, we specify the covariance function specifying the spatial process $\omega(\ell)$ in (5) to be 0 for any two points in different regions, and equal to the value of the Matérn covariance function for any two points within the same region.

Comparing the finite populations estimates given in (4) and (6), it is evident that accounting for spatial variation results in a more complex equation, as all observed and unobserved outcome values in a population can no longer be assumed to be independent conditional on the group means. This can also be seen in the calculation of the λ parameters, which in the two-stage model, are a simple ratio of variances. In the spatial case, however, the complexity of the parameters is increased by the addition of the spatial covariance matrix.

IV. Model Implementation and Assessment

I. General framework

A Bayesian linear model corresponding to the likelihood of the sampled data in (2) is

$$
p(\theta, \nu, \beta \,|\, y_s) \propto p(\theta) \times N(\nu \,|\, 0, V_\nu) \times N(\beta \,|\, A\nu, V_\beta) \times N(y_s \,|\, X_s \beta, V_s(\theta)) \,.
$$
(7)

We use Markov chain Monte Carlo algorithms (see, e.g. Robert and Casella, 2004) for sampling from (7). Subsequeny Bayesian inference for y_{ns} is available in posterior predictive fashion by drawing samples from

$$
p(y_{ns} \,|\, y_s) = \int p(y_{ns} \,|\, y_s, \theta, \nu, \beta) \times p(\theta, \nu, \beta \,|\, y_s) \, d\theta \, d\nu \, d\beta \,.
$$
(8)

Using the conditional independence of parameters in (2), we obtain $p(y_{ns} \mid y_s, \theta, \nu, \beta) = N(y_{ns} \mid \mu_{ns\mid s}, V_{ns\mid s})$, where $\mu_{ns\mid s} = X_{ns}\beta + V_{ns,s}(\theta)V_s(\theta)^{-1}(y_s - X_s\beta)$ and $V_{ns\mid s} = V_{ns}(\theta) - V_{ns,s}(\theta)V_s(\theta)^{-1}V_{s,ns}(\theta)$. Therefore, sampling from (8) is achieved by drawing one $\{\beta, \theta\}$ from (7) followed by one $y_{ns} \sim N(X_{ns}\beta, V_{ns}(\theta))$, for each posterior sample of $\{\beta, \theta\}$. The resulting samples provide inference on the nonsampled group means μ_{ns} and Bayesian imputation for the nonsampled population units, y_{ns}.

These samples from the posterior predictive distribution can be used to obtain posterior finite population estimates of the form $\alpha^\top y$. We consider four models using (7).

Model 1. Two-Stage

For the model provided in Example 2, we take $\theta = [\gamma^2, \delta^2, \sigma_1^2, \ldots, \sigma_N^2]^\top$, $p(\theta) = IG(\gamma^2 \mid a_\gamma, b_\gamma) \times IG(\delta^2 \mid a_\delta, b_\delta) \times \prod_{i=1}^N IG(\sigma_i^2 \mid a_{\sigma_i}, b_{\sigma_i})$ in (7), and follow the other specifications as in the two-stage setting in Example 2.

As the priors have been chosen to be fully conjugate, one can derive the full posterior conditional distributions for each of the parameters. Specifically, the variance parameters will have posterior distributions of the form $IG(a^*, b^*)$, while the rest of the parameters will have posterior distributions of the form $N(Mm, M)$. However, as only n of the N groups are observed, the variance terms of the unsampled groups, $\sigma_{n+1}^2, \ldots, \sigma_N^2$, must either be fixed or given informative priors. If not, draws from the posterior predictive distribution corresponding to units in the nonsampled groups will have arbitrary variability and could spuriously dominate the finite population estimates.

Model 2. Spatial

Under (7), the intercept-only spatial model defines $\theta = [\phi, \delta^2, \sigma^2, \tau^2]^\top$ with corresponding prior distribution $p(\theta) = p(\phi) \times IG(\delta^2 \mid a_\delta, b_\delta) \times IG(\sigma^2 \mid a_\sigma, b_\sigma) \times IG(\tau^2 \mid a_\tau, b_\tau)$ and $V_\beta = \delta^2$, $V_s = \Omega_s + \sigma^2 I_k$. As there are no group terms, replace X_s with 1_k and take $\nu = 0$ with probability 1, e.g. $V_\nu^{-1} = 0$.

Unlike model 1, regardless of the prior distribution placed on $p(\phi)$, a closed-form posterior distribution cannot be found for ϕ. In practice, ϕ is often fixed using an estimate found from a variogram and then full posterior conditional distributions can be found using the same techniques described for the non-spatial case. However, MCMC can still be implemented by specifying a prior distribution for ϕ (Banerjee et al. 2014), which is often taken to be a uniform distribution.

To recover the spatial effects ω absorbed into the variance parameter of y, note that $\beta = \mu_s$ and $p(\omega \mid y, \theta, \mu_s) \propto N(\omega \mid 0, \Omega) \times N(y \mid 1_T\mu_s + \omega, \sigma^2 I_T) \propto N(M_\omega m_\omega, M_\omega)$, where $m_\omega = \frac{1}{\sigma^2}(y - 1_T\mu_s)$ and $M_\omega = (\Omega^{-1} + \frac{1}{\sigma^2}I_T)^{-1}$. Thus, drawing one $\omega \sim N(M_\omega m_\omega, M_\omega)$ for each posterior sample of $\{\theta, \mu_s, y_{ns}\}$ will result in a set of posterior samples of ω.

Model 3. Two-Stage + Spatial

The spatial model in (5) can be rewritten using (7) by letting $\theta = [\phi, \gamma^2, \delta^2, \tau^2, \sigma_1^2, \ldots, \sigma_N^2]^\top$, with $p(\theta) = p(\phi) \times IG(\gamma^2 \mid a_\gamma, b_\gamma) \times IG(\delta^2 \mid a_\delta, b_\delta) \times IG(\tau^2 \mid a_\tau, b_\tau) \times \prod_{i=1}^N IG(\sigma_i^2 \mid a_{\sigma_i}, b_{\sigma_i})$, $V_\nu = \gamma^2$, $V_\beta = \delta^2 I_N$, $A = 1_N$, $V_s = \Omega_s + V_s^{(\sigma)}$, and $\beta = \mu$. After posterior samples of $\{\theta, \mu, y_{ns}\}$ are drawn as described in (8), posterior samples of the spatial effects can be recovered by sampling one $\omega \sim N(M_{\omega 2} m_{\omega 2}, M_{\omega 2})$ for each posterior sample of $\{\theta, \mu, y_{ns}\}$, where $m_{\omega 2} = V^{(\sigma)-1}(y - X\mu)$ and $M_{\omega 2} = (\Omega^{-1} + V^{(\sigma)-1})^{-1}$.

Model 4. Regional Spatial

To rewrite the region-specific spatial model given using (7), let $V_\nu = \gamma^2$, $V_\beta = \delta^2 I_N$, $A = 1_N$, $V_s = \Omega_{s*} + V_s^{(\sigma)}$, and $\beta = \mu$. Also take $\theta = [\phi_1, \ldots, \phi_N, \gamma^2, \delta^2, \tau_1^2, \ldots, \tau_N^2, \sigma_1^2, \ldots, \sigma_N^2]^\top$ with $p(\theta) = \prod_{i=1}^N p(\phi_i) \times IG(\gamma^2 \mid a_\gamma, b_\gamma) \times IG(\delta^2 \mid a_\delta, b_\delta) \times \prod_{i=1}^N IG(\tau_i^2 \mid a_{\tau_i}, b_{\tau_i}) \times \prod_{i=1}^N IG(\sigma_i^2 \mid a_{\sigma_i}, b_{\sigma_i})$. Similar to model 1, as not all locations are sampled, informative priors must be placed on the $\phi_{n+1}, \ldots, \phi_N$ spatial decay parameters. Additionally, to recover posterior samples of ω, sample

one $\omega \sim N(M_{\omega 3}m_{\omega 3}, M_{\omega 3})$ for each posterior sample of $\{\theta, \mu, y_{ns}\}$ drawn using (8), where $m_{\omega 3} = V^{(\sigma)-1}(y - X\mu)$ and $M_{\omega 3} = (\Omega_*^{-1} + V^{(\sigma)-1})^{-1}$.

To achieve computation efficiency, redefine $y = [y_1^\top, y_1^{*\top}, \ldots, y_n^\top, y_n^{*\top}, y_{n+1}^{*\top}, \ldots y_N^{*\top}]^\top$ so that the outcome is organized by region and then Ω_* becomes a $T \times T$ block diagonal matrix composed of N blocks. This allows us to instead invert N covariance matrices of size $M_1 \times M_1, \ldots, M_N \times M_N$, rather than one $T \times T$ matrix, in the estimation of ω.

II. Exact Monte Carlo Estimation

If we are able to provide reasonable fixed values of the parameters, (2) can be simplified into a conjugate Bayesian linear model resembling:

$$IG(\delta^2 \,|\, a, b) \times N(\nu \,|\, 0, \delta^2 \tilde{V}_\nu) \times N(\beta \,|\, A\nu, \delta^2 \tilde{V}_\beta) \times N(y_s \,|\, X\beta, \delta^2 \tilde{V}_s) \,. \tag{9}$$

For a model such as this, the components a, b, $\tilde{V}_\nu$, $\tilde{V}_\beta$, and $\tilde{V}_s$ are fixed, reducing the model to three unknown parameters, δ^2, ν, and β. Thus, we can avoid MCMC and sample from the joint posterior $p(\delta^2, \nu, \beta \,|\, y_s)$ using the following steps. First sample δ^2 from $p(\delta^2 \,|\, y - S) = IG(a^*, b^*)$ and then for each δ^2 drawn, draw a corresponding ν from $N(M_\nu m_\nu, \delta^2 M_\nu)$. Next, for each pair of $\{\delta^2, \nu\}$, draw β from $N(M_\beta m_\beta, \delta^2 M_\beta)$ (see the Supporting Information for details). As an example, we recast each model presented in Section I in the form of (9) and derive the posterior conditional distributions for model 1 and model 2, details of which are provided in the Supporting Information.

Model 1. Two-Stage

To create a conjugate Bayesian model such as (9) from the non-spatial model, define $A = 1_N$, $\tilde{V}_\nu = \tilde{\gamma}^2 = \frac{\gamma^2}{\delta^2}$, $\tilde{V}_\beta = I_N$, and $\tilde{V}_s = \tilde{V}_s^{(\sigma)} = [\oplus_{i=1}^n \frac{\sigma_i^2}{\delta^2} I_{m_i}]$. Noting that $p(\nu \,|\, y_s) \propto N(\nu \,|\, 0, \delta^2 \tilde{\gamma}^2) \times N(y_s \,|\, X_{s1} 1_n \nu, \delta^2 [X_{s1} X_{s1}^\top + \tilde{V}_s^{(\sigma)}])$ a little algebra reveals

$$\nu \,|\, y_s, \delta^2 \sim N(\nu \,|\, c, \delta^2 d) \,, \tag{10}$$

where $c = \frac{\sum_{i=1}^n \lambda_i \bar{y}_i}{\frac{1}{\tilde{\gamma}^2} + \sum_{i=1}^n \lambda_i}$ and $d = \left[\frac{1}{\tilde{\gamma}^2} + \sum_{i=1}^n \lambda_i\right]^{-1}$. The mean of the posterior distribution, c, is the weighted average of the sampled group means, where each mean is weighted by a function of each group's element-wise variance. Integrating out ν and μ from $p(\delta^2, \nu, \mu \,|\, y_s)$ yields $p(\delta^2 \,|\, y_s)$, which is:

$$\delta^2 \,|\, y_s \sim IG\left(a + \frac{k}{2}, b + \frac{1}{2}\left[y_s^\top (X_{s1} X_{s1}^\top + \tilde{V}_s^{(\sigma)})^{-1} y_s + \frac{c^2}{d}\right]\right) \,. \tag{11}$$

Taking the limits of c and d as $\tilde{\gamma}^2 \to \infty$ (e.g. $\gamma^2 \to \infty$) we recover the findings of Scott and Smith (1969), who assigned $p(\nu) \propto 1$:

$$\lim_{\tilde{\gamma}^2 \to \infty} c = \frac{\sum_{i=1}^n \lambda_i \bar{y}_i}{\sum_{i=1}^n \lambda_i} \quad \text{and} \quad \lim_{\tilde{\gamma}^2 \to \infty} \delta^2 d = \frac{\delta^2}{\sum_{i=1}^n \lambda_i} \,.$$

As $p(\mu_s \,|\, y_s, \delta^2, \nu) \propto N(\mu_s \,|\, \nu 1_n, \delta^2 I_n) \times N(y_s \,|\, X_{s1}\mu_s, \delta^2 \tilde{V}_s^{(\sigma)})$ we have that:

$$\mu_s \,|\, y_s, \nu, \delta^2 \sim N(\mu_s \,|\, c_*, \delta^2 d_*) \,, \tag{12}$$

where $c_* = \begin{bmatrix} (1-\lambda_1)\nu + \lambda_1 \bar{y}_1 \\ \vdots \\ (1-\lambda_n)\nu + \lambda_n \bar{y}_n \end{bmatrix}$ and $d_* = [\oplus_{i=1}^n (1-\lambda_i)]$. The posterior mean is appealing for interpretation, as its i-th element is the weighted average of the i-th group's sample mean and the

superpopulation mean estimate. Finally, note that $\mu_{ns} \mid y_s, \nu, \delta^2 \sim N(\mu_{ns} \mid \nu, \delta^2 I_{N-n})$, as y_s provides no information pertaining to the nonsampled groups.

Model 2. Spatial

The spatial model can be recast as (9) by defining $\tilde{V}_\nu^{-1} = 0$ and $\tilde{V}_s = \tilde{\Omega}_s = \frac{1}{\delta^2}\Omega_s + I_k$, where $\tilde{V}_\beta$ is fixed to 1. Defining $V_{\tilde{\Omega}_s} = (1 + 1_k^\top \tilde{\Omega}_s^{-1} 1_k)^{-1}$, the posterior conditionals are:

$$\delta^2 \mid y_s \sim IG \left[a + \frac{k}{2}, b + \frac{1}{2} y_s^\top \left(\tilde{\Omega}_s^{-1} - \tilde{\Omega}_s^{-1} 1_k V_{\tilde{\Omega}_s} 1_k^\top \tilde{\Omega}_s^{-1} \right) y_s \right] \text{ and} \tag{13}$$

$$\mu_s \mid y_s, \delta^2 \sim N \left[V_{\tilde{\Omega}_s} 1_k^\top \tilde{\Omega}_s^{-1} y_s, \delta^2 V_{\tilde{\Omega}_s} \right]. \tag{14}$$

Model 3. Two-Stage + Spatial

The form of (9) is achieved by defining $\tilde{V}_\nu = \tilde{\gamma}^2$, $\tilde{V}_\beta = I_N$, $A = 1_N$, and $\tilde{V}_s = \frac{1}{\delta^2}\Omega_s + \tilde{V}_s^{(\sigma)}$.

Model 4. Regional Spatial

The form of (9) is achieved by defining $\tilde{V}_\nu = \tilde{\gamma}^2$, $\tilde{V}_\beta = I_N$, $A = 1_N$, and $\tilde{V}_s = \frac{1}{\delta^2}\Omega_{s*} + \tilde{V}_s^{(\sigma)}$.

III.   Model Comparison and Assessment

Model fit was evaluated in two ways. In general, consider a sample of size k drawn from a population of size T with outcome $y = [y_s^\top : y_{ns}^\top]^\top$. Without loss of generality, say $y_h \in y_s$ if $h = 1, \ldots, k$ and $y_h \in y_{ns}$ if $h = k+1, \ldots, T$. First we evaluate the predictive accuracy of the models using the Watanabe-Akaike Information Criteria (WAIC), which is expressed as $WAIC = -2\widehat{elpd} = -2(\widehat{lpd} + \hat{p}_{WAIC})$ in Vehtari et al. (2017), where $\widehat{elpd}$ is the estimated expected log pointwise predictive density and is multiplied by -2 to be on the deviance scale. To calculate this, at each iteration, $l = 1, \ldots, L$, $p(y_h \mid \Theta^{(l)})$ is computed; the likelihood of each observed value conditional on that iteration's parameters. The estimated log pointwise predictive density is the sum of the log average likelihood for each observation, $\widehat{lpd} = \sum_{h=1}^k \log \left[\frac{1}{L} \sum_{l=1}^L p(y_h \mid \Theta^{(l)}) \right]$. The sample variance of the log-likelihood for each observation is $s_{lp(y_h)}^2 = \frac{1}{L-1} \sum_{l=1}^L \left[\log(p(y_h \mid \Theta^{(l)})) - \frac{1}{L} \sum_{l=1}^L \log(p(y_h \mid \Theta^{(l)})) \right]^2$ and the estimated effective number of parameters is the sum of these variances: $\hat{p}_{WAIC} = \sum_{h=1}^k s_{lp(y_h)}^2$. To calculate the standard error of the WAIC, rewrite $-2(\widehat{lpd} + \hat{p}_{WAIC}) = -2\sum_{h=1}^k \widehat{elpd}_h = \sum_{h=1}^k \left\{ \log \left[\frac{1}{L} \sum_{l=1}^L p(y_h \mid \Theta^{(l)}) \right] + s_{lp(y_h)}^2 \right\}$. Under the assumption that each $\widehat{elpd}_h$ is independent, the sample variance of each individual $\widehat{elpd}_h$ is $s_{elpd,ind}^2 = \frac{1}{N-1} \sum_{h=1}^k \left[\widehat{elpd}_h - \frac{1}{k} \sum_{h=1}^k \widehat{elpd}_h \right]$. Then $SE(WAIC) = \sqrt{Var(-2\sum_{h=1}^k \widehat{elpd}_h)} = 2\sqrt{nVar(\widehat{elpd}_h)} = 2s_{elpd,ind}\sqrt{n}$.

Second, for simulated data the true values $y = [y_s : y_{ns}]^\top$ are known and so we compare these with replicated datasets, $y_{rep}^{(l)} = [y_{rep,1}^{(l)} \ldots y_{rep,k}^{(l)}]^\top$, generated from the pointwise posterior predictive distribution at each iteration l. These are used to formulate the goodness of fit measurement $D = G + P$ described in Gelfand and Ghosh (1998), composed of an error sum of squares term and a penalty term for large predictive variances. For L iterations, $G = \sum_{h=1}^k (y_h - E[y_{rep,h} \mid y_s])^2$ and $P = \sum_{h=1}^k var(y_{rep,h} \mid y_s)$. We approximate $E[y_{rep,h} \mid y_s] \approx \frac{1}{L} \sum_{l=1}^L y_{rep,h}^{(l)}$ and $var(y_{rep,h} \mid y_s) \approx \frac{1}{L-1} \sum_{l=1}^L (y_{rep,h}^{(l)} - \frac{1}{L} \sum_{l=1}^L y_{rep,h}^{(l)})^2$. For non-simulated datasets, where y_{ns} is unknown, D can still be

calculated by restricting the replicate datasets to the observed units, y_s, e.g. $y_{rep}^{(l)} = [y_{rep,1}^{(l)} \cdots y_{rep,k}^{(l)}]^\top$, at the l-th iteration.

V. Simulation

I. Data Generation

To simulate spatial correlation and allow for two-stage random sampling, a unit square was divided into 100 equally sized square regions and 2,500 locations were randomly drawn from the unit square. Data was simulated from the intercept-only spatial model described in Model 2 with $\mu = 2$. A distance matrix for all points was constructed and used to create a covariance matrix that reflects an exponential covariance function described in Section III, where ϕ was assigned a value of 10, reflecting an effective spatial range of 3/10. The spatial variance, τ^2, was fixed at 9, while the non-spatial variance, σ^2, was set to 4. After a dataset was generated, a cluster random sampling scheme was implemented. 25 regions were randomly selected and then in each cluster, a random number of individuals were selected (the minimum and maximum percent of those selected from a region was set to be 20% and 90%, respectively). 20 datasets containing information of both the sampled and nonsampled units were generated in this way. To examine Models 1 and 4 for larger dataset, this process was then repeated with the same parameters to generate 20 datasets with 8,100 locations from 324 regions, where 81 regions were randomly sampled. All data generation and analyses were performed using R version 3.5.1 R Core Team 2018.

II. Exact Monte Carlo Simulation

To perform the two-stage procedure using the conditional distributions and methods described in Section II, sample means, $\hat{\mu}_i$, and sample variances, $\hat{\sigma}_i^2$, were calculated from each observed cluster, $i = 1, \ldots, n$. The variance matrix of the sampled units was fixed to be $V_s^{(\sigma)} = \left[\oplus_{i=1}^n \frac{\hat{\sigma}_i^2}{\text{Var}(\hat{\mu})} I_{m_i} \right]$, where $\text{Var}(\hat{\mu})$ represents the sample variance of the observed sample means. Similarly, the variance matrix of the nonsampled units was fixed at $V_{ns} = \left[\oplus_{i=1}^N \frac{\tilde{\sigma}_i^2}{\text{Var}(\hat{\mu})} I_{M_i - m_i} \right]$, where $\tilde{\sigma}_i^2 = \hat{\sigma}_i^2$ if $i \in \{1, \ldots, n\}$ and $\tilde{\sigma}_i^2 = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_i^2$ if $i \in \{n+1, \ldots, N\}$. The value of γ^2 was fixed to be half of the value of δ^2, reflecting the belief that there was less variability in the population mean than between group means. The prior distribution for δ^2 was assigned to be $IG(3,5)$. Sampling from the posterior was performed using the conditional distributions and methods described in Section II. At each iteration g, the population mean estimate for that iteration was then calculated as $\bar{y}^{(g)} = \frac{1}{T} \left(\sum_{i=1}^n \sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^N \sum_{j=m_i+1}^{M_i} y_{ij}^{(g)} \right)$. Details of this iterative procedure can be found in the supplementary materials.

To perform the spatial random effect procedure, ϕ was set to its true value of 10 and the ratio of δ^2/τ^2 to its true value of 4/9. The posterior conditionals of $\delta^2 \,|\, y_s$ and $\mu_s \,|\, y_s, \delta^2$, (13) and (14) respectively, were sampled as outlined in Section III. This sampling and the prediction of y_{ns} was performed using commands from the *spBayes* R package (Finley et al., 2015, 2007). The population mean estimate was calculated using the technique described in the non-spatial sampling case above.

Figure 1 plots population average-centered mean estimates and 95% credible intervals from both methods applied to the twenty simulated datasets. While the spatial cases consistently have a smaller credible interval, their point estimates are similar to the two-stage case. However, as the ratio of spatial and non-spatial variance is fixed for this method, this may result in a reduction in the overall variance of the population mean. Posterior mean estimates and their associated 95%
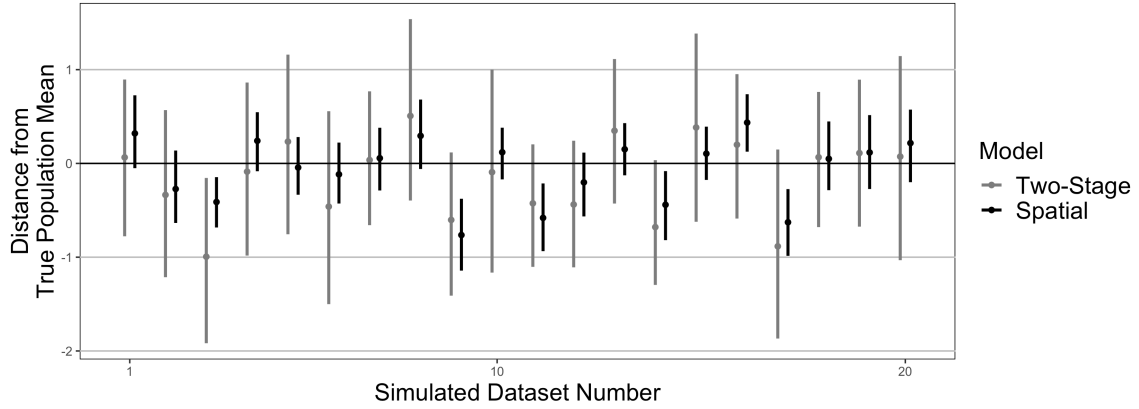
Figure 1: Centered Population Mean Estimates from 2 Exact Models with 95% CI

credible intervals of the superpopulation parameters and finite population mean, $\bar{y}$, from the first generated dataset are given in Table 1, along with the WAIC, its standard error, and D values. While both models have similar estimates for ν, the two-stage model overestimates the non-spatial variance. This is expected, as we know that there is additional variance due to spatial correlation that is not being accounted for otherwise in the model. Similarly, both measures of goodness of fit prefer the spatial model.

Table 1: Comparison of Parameter Estimation and Model Fit in Two Exact Models

	Two-Stage	Spatial
ν (2)	2.60 (1.55, 3.58)	2.79 (1.36, 4.23)
δ^2 (4)	6.36 (5.47, 7.45)	3.84 (3.35, 4.43)
τ^2 (9)	—	8.65 (7.53, 9.96)
$\bar{y}$ (2.60)	2.66 (7.53, 9.96)	2.92 (7.53, 9.96)
WAIC	1803.83 (25.52)	1686.7 (27.6)
D = G + P	66100.83=34203.26+31897.56	45335.54=22875.45+22460.08

III. Markov Chain Monte Carlo Simulation

To explore these findings further, we implemented the four models described in Section IV using the JAGS software in R on the same generated datasets. Models were run for 650 iterations with 50 burn-in, as examination of individual trace plots suggested sufficient mixing and convergence of the non-spatial parameters. At each iteration g, estimates of the nonsampled units were drawn and estimates for the population mean, $\bar{y}^{(g)} = \frac{1}{T} \left(\sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{N} \sum_{j=m_i+1}^{M_i} y_{ij}^{(g)} \right)$ were calculated. All variance parameters (σ^2, τ^2, and δ^2, as well as site-specific variances such as σ_i^2 and τ_i^2) were given an inverse-gamma prior with shape 2 and scale 10, reflecting a weakly-informative prior distribution with mean 10. There is a substantial literature in theoretical spatial statistics regarding the identifiability, or lack there of, of the spatial process parameters, hence, non-informative or completely flat priors are excluded from consideration. The prior families and specifics we use are fairly customary in spatial modeling. They exploit some information about the spatial domain

11

and the extent of spatial association that can be detected from finite samples using variograms. For example, in practice, given a real data set, we would pass the data through an exploratory analysis tool, such as a variogram, glean some information about the spatial variance component and the measurement error component, and use the weakly informative centered inverse-gamma priors to reflect these values. In addition, ν was given a flat prior to not inform the estimation of the mean and all ϕ parameters were given Uniform(5,15) priors to allow the spatial range to vary from 0.2 (3/15) to 0.6 (3/5). While our priors were chosen to be weakly-informative to be conservative in estimation, more informative priors could easily be added in a data analysis if additional information regarding the parameters was known. MCMC sampling was performed using the computer program JAGS (Plummer, 2017) in R.

When assessing model fit in the first realization of the data with WAIC, the spatial model performed slightly worse than the rest of the models with a value of 1,912.70 (SE = 26.36). This may be due to the additional variation which comes from varying the spatial range parameter. This was followed closely by the two-stage model with 1,870.26 (35.00) , which was outperformed by both the regional spatial model with 1,202.08 (38.99) and the two-stage + spatial model with 455.67 (17.03). It is interesting that while the data was generated by the spatial model and sampled by a two-stage framework, neither of these models perform better than the two models which take both the spatial correlation and study design into account. Additional comparisons of estimates from this first realization can be found in the supplementary materials.
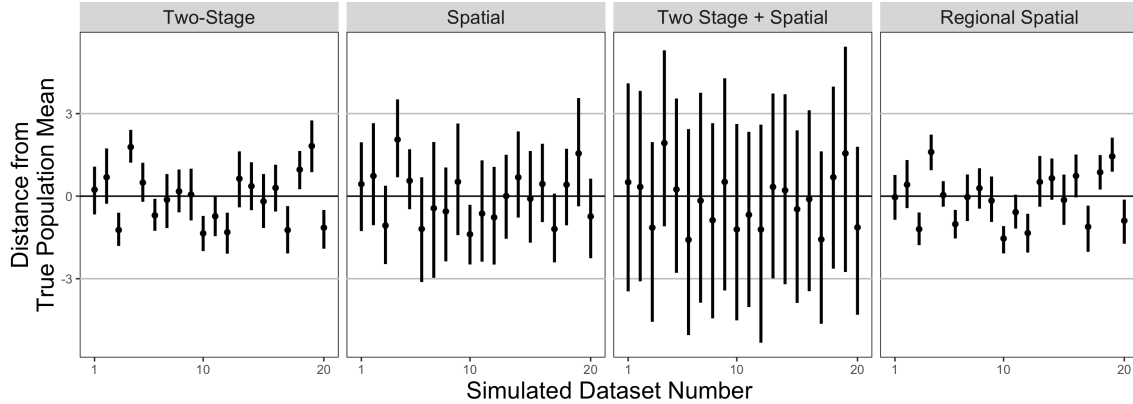

Figure 2: Centered Population Mean Estimates from 4 MCMC Models with 95% CI

Figure 2 shows the models' posterior mean estimates of the finite population mean, which are centered at the true population mean and presented with 95% credible intervals for the 20 simulated datasets. While point estimates remain similar across models, the best fitting model, Model 3, has the widest credible intervals for the population mean. Accounting for only regional effects results in tight credible intervals in Models 1 and 4, which are narrow compared to Model 2, which fails to take into account region specific variability. Similar results were found when applying Models 1 and 4 to the larger simulated datasets and are provided in the supplementary materials.

VI. Data Analysis: Nitrate in Central California Groundwater

In this section, we provide an analysis of groundwater nitrate content of the Tulare Lake Basin (TLB) in Central California from the California Ambient Spatio-Temporal Information on Nitrate

in Groundwater (CASTING) Database, which is described in Harter et al. (2017) and Boyle et al. (2012) and is available as the University of California, Davis, nitrate data in the data repository of the Groundwater Ambient Monitoring and Assessment Program (2019). Interest lies in identifying regions in which ground water nitrate levels exceed 45 mg/L, which is the maximum contaminant level established by the EPA Boyle et al. (2012). At high levels, infants and pregnant women are more susceptible to nitrate poisoning, which makes it more difficult for oxygen to be distributed to body and can be fatal to infants less than six months old. Besides human sources such as sewage disposal, many sources of nitrate are agricultural, such as fertilizer for crops and animal waste (Harter and Lund, 2012). Because of this, regions with high agricultural activity, such as the TLB, have experienced rising levels of nitrate over the past few decades. As groundwater, and therefore nitrate levels in groundwater, can be assumed to be present at all areas of the Central Valley, we can assume that water samples taken from wells come from a spatial field. Therefore, given a sample of readings from various wells, our primary goal is to estimate of the finite population average of all known wells, which represents an overall measure of water-health. Additionally, plots of posterior predictive distribution may be useful in identifying high-risk regions which exceed the maximum contaminant level.

The CASTING Database is an extensive collection of nitrate readings from the TLB and Salinas Valley collected by multiple agencies, over 70% of which were collected between 2000 and 2011. Of these, most wells had repeated measurements taken over the time. As the Salinas Valley and the TLB are geographically separate regions of California, only the TLB was included. In order to avoid associations over time, the data was restricted to a single year. The year 2009 was selected as variogram plots suggested nitrate levels followed a roughly exponential distribution. While directional variograms suggested that the measurements may be anisotropic, we continued with the methods presented above, recognizing that a model accounting for directional spatial dependence may provide a better fit to the data.

While the data roughly covers the TLB, we construct a likely sampling scenario in which the TLB is separated into distinct geographic regions and due constraints (perhaps time or financial), a random subset of these regions are sampled. In our scenario, zip codes were used as it is common to collect such geographic information in large scale health surveys, but many alternatives such as cities or grid-based approach could have also been used. A map of California zip codes tabulation areas obtained from the *tigris* R package (Walker, 2018) was overlaid on the approximate geographic locations of each of the sampled wells, effectively assigning each well to one specific region, defined by a zip code. 1) Only the most recent observation was taken from each well so that each well was only represented once. 2) If unique wells had the same geographic coordinates, one was chosen at random to be removed. 3) Sparse zip codes with less than 10 wells were excluded to ensure that each selected zip code would have a large sample size and to avoid overfitting when modeling. These restrictions resulted in a dataset with 6,117 unique wells among 63 zip codes. Nitrate level had a mean of 37.9 mg/L, standard deviation of 52.3 mg/L, and ranged from 0.0 to 903.1 mg/L.

In order to recreate a cluster sampling scenario, 21 of the zip codes were randomly chosen and 50-90% of the population in that zip code was randomly sampled. This resulted in an observed sample size of 489 with a mean nitrate level of 34.2 mg/L and a standard deviation of 40.0 mg/L. The nitrate level ranged from 0.0 to 269.6. A plot of these sampled and non-sampled zip codes is available in the supplementary materials. Using this sampled data, all four models in Section III were implemented and the results are shown in Table 2. In all models, ν was given a flat prior. For model 1, the regional variance parameters were given an inverse-gamma prior with shape 2 and scale 1600 to reflect the sample variance. For the spatial models, a variogram was fit and spatial variances were given an inverse-gamma prior with shape 2 and scale 1800. Similarly, non-spatial

variance terms were assigned an inverse-gamma prior with shape 2 and scale 1000. The variance of regional means was assigned an inverse-gamma prior with shape 2 and scale 10 to allow for small, localized deviations. All ϕ parameters were given Uniform(0.01,5) priors to reflect a spatial range varying between 0.6 km (3/5) and 300 km (3/0.01). MCMC sampling was performed using JAGS (Plummer, 2017) in R (R Core Team, 2018).

With respect to the estimate of the true mean nitrate level, only the intercept-only spatial model contained the true mean value within its 95% credible intervals. However, as evidenced by the larger mean, standard deviation, and range in the complete dataset, it appears that the sampled units did not capture some of the larger outliers, so it is unsurprising that the estimates of the population mean are lower than the truth. Comparing WAIC, we see results similar to those found in Section III. The spatial models which accounted for regional means had lower WAIC values than the two-stage model, which is evidence that this data is spatially correlated. However, the intercept-only spatial model did not fit the data as well as the two-stage model, which may be due to ignoring the study design. Additionally, the two-stage + spatial model again fits the model the best.
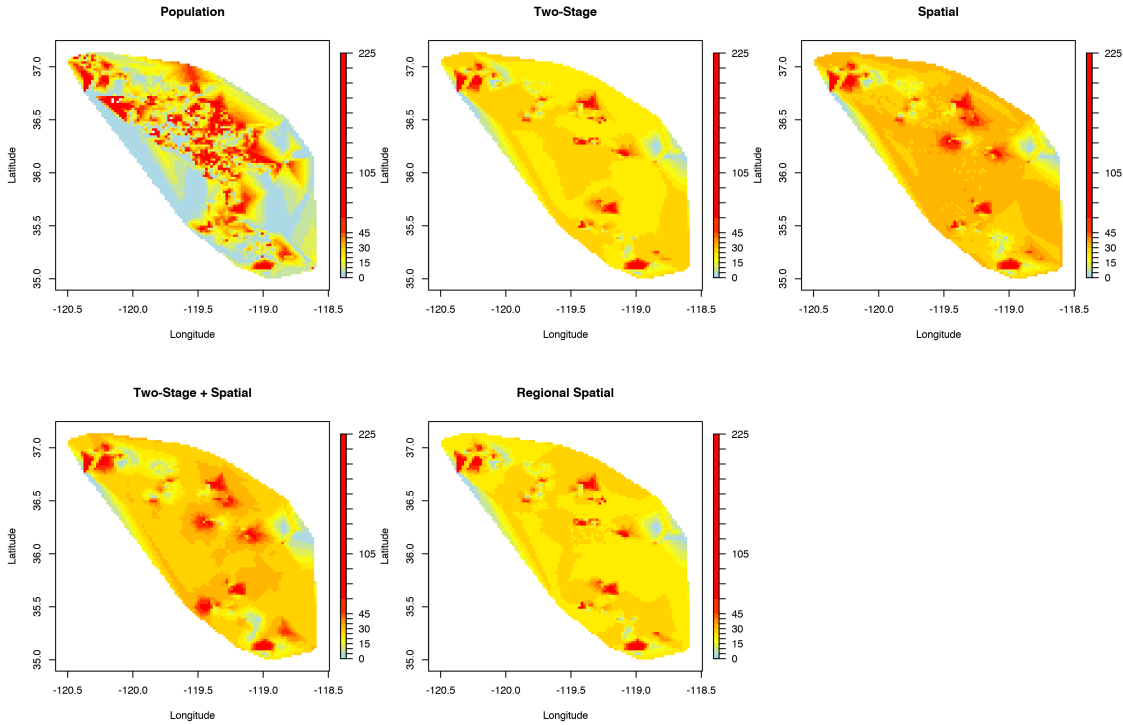

Figure 3: Interpolated surfaces using the population (truth) and posterior predictive samples from the four models.

Figure 3 shows the interpolated population surface from the complete sample and the interpolated surface from posterior predictive samples. While there are common regions at high risk

(nitrate level greater than 45 mg/L) in all the posterior predictive maps, the spatial and two-stage + spatial maps predict larger regions. Also seen in Table 2, it is clear that Model 3 estimates a population mean that is larger than Models 1 and 4, but smaller than Model 3. Spatial residual plots for Models 2, 3, and 4 are provided in the supplementary materials.

Table 2: Results of Data Analysis

Model	FP Mean (95% CI)	WAIC (SE)
1. Two-Stage	26.6 (19.2, 35.1)	4701.6 (78.6)
2. Spatial	32.6 (27.2, 38.7)	4858.4 (78.0)
3. Two-Stage + Spatial	31.2 (24.2, 37.3)	2697.0 (150.5)
4. Regional Spatial	26.2 (18.7, 34.4)	4536.6 (93.1)

VII. Discussion

This paper examines the implications of performing two-stage random sampling on point-referenced data which exists in a spatial field. While Scott and Smith (1969) and Malec and Sedransk (1985) provided a Bayesian model-based framework to account for such a study design, we have demonstrated that an analysis ignoring the underlying spatial correlation between locations or sampling design may lead to spurious inference and poorer model fit.

This work is a first step in developing an overarching framework for Bayesian finite population sampling from spatial process based populations. In our two-stage case, additional work may be done to further improve this model. For instance, CAR priors could be placed on regional parameters such as the μ_i's, the regional means, to induce additional spatial correlation in the model. Also, many of the models presented account for regional differences in variance but if other sources of heteroscedasticity are suspected, new approaches (such as Zangeneh and Little 2015) may be needed to account for this. While an exponential covariance function was employed in the analyses in this paper, other spatial covariance functions could be used to create similar simulations and data analyses, as well as account for anisotropy.

Future work is needed to establish a more general framework that can account for more sophisticated sampling designs in a spatial context. The sampling designs presented in this paper are said to be ignorable (Rubin 1976; Sugden and Smith 1984), which allows us to perform inference on the superpopulation parameters while ignoring the inclusion probability distribution. However, designs in which the data cannot be assumed to be missing at random or where parameters define both the outcome and inclusion distributions are referred to as nonignorable and must account for the inclusion probability distribution. One example of this in the spatial context is preferential sampling (Diggle et al. 2010, Gelfand et al. 2012), in which the measurement values and sampling strategy are assumed to stem from the same spatial process. While Pati et al. (2011) have analyzed such data using Bayesian hierarchical models, an overall framework is needed to account for this and other non-ignorable design types.

Additionally, the implications of study design on finite population estimates when sampling from a spatially correlation population over time are unknown. In order to better understand this, these Bayesian models must first be extended to account for both study design and spatio-temporal associations.

Finally, while this paper provided a scaleable model which can account for study design and spatial correlation in massive survey data by assuming regional independence, further work

should be done to incorporate recent strategies in modeling large spatial data (Heaton et al., 2018) when analyzing survey data with spatial correlations, such as nearest neighbor processes (Datta et al., 2016), covariance tapering (Furrer et al., 2006), and metakriging (Guhaniyogi and Banerjee, 2018). Finite population models would particularly benefit from such techniques, as computation increases as a function of the population total, T, rather than the sample size, k.

Supporting Information

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC, 2nd edition.

Boyle, D., King, A., Kourakos, G., Lockhart, K., Mayzelle, M., Fogg, G. E., and Harter, T. (2012). "Groundwater Nitrate Occurrence, Technical Report 4." Technical report, Center for Watershed Sciences, University of California, Davis, Davis, CA.

Bruno, F., Cocchi, D., and Vagheggini, A. (2013). "Finite population properties of individual predictors based on spatial pattern." *Environmental and Ecological Statistics*, 20(3): 467–494.

Cicchitelli, G. and Montanari, G. E. (2012). "Model-Assisted Estimation of a Spatial Population Mean." *International Statistical Review*, 80(1): 111–126.

Clayton, D. and Kaldor, J. (1987). "Empirical Bayes estimates of age-standardized relative risks for use in disease mapping." *Biometrics*, 43(3): 671–681.

Cochran, W. G. (1977). *Sampling Techniques*. Hoboken, NJ: John Wiley & Sons, 3rd edition.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Hoboken, NJ: John Wiley & Sons.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). "Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets." *Journal of the American Statistical Association*, 111(514): 800–812.

Diggle, P. J., Menezes, R., and Su, T.-L. (2010). "Geostatistical inference under preferential sampling." *Journal of the Royal Statistical Society: Series C*, 59(2): 191–232.

Ericson, W. A. (1969). "Subjective Bayesian models in sampling finite populations." *Journal of the Royal Statistical Society, Series B*, 31(2): 195–233.

Finley, A. O., Banerjee, S., and Carlin, B. P. (2007). "spBayes: An R Package for Univariate and Multivariate Hierarchical Point-Referenced Spatial Models." *Journal of Statistical Software*, 19(4): 1–24.

Finley, A. O., Banerjee, S., and E.Gelfand, A. (2015). "spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models." *Journal of Statistical Software*, 63(13): 1–28.

Furrer, R., Genton, M. G., and Nychka, D. (2006). "Covariance Tapering for Interpolation of Large Spatial Datasets." *Journal of Computational and Graphical Statistics*, 15(3): 502–523.

Gelfand, A. E. and Ghosh, S. K. (1998). "Model choice: a minimum posterior predictive loss approach." *Biometrika*, 85(1): 1–11.

Gelfand, A. E., Sahu, S. K., and Holland, D. M. (2012). "On the effect of preferential sampling in spatial prediction." *Environmetrics*, 23(7): 565–578.

Gelman, A. (2007). "Struggles with Survey Weighting and Regression Modeling." *Statistical Science*, 22(2): 153–164.

Ghosh, M. and Meeden, G. (1997). *Bayesian Methods for Finite Population Sampling*. London: Chapman & Hall.

Ghosh, M., Natarajan, K., Stroud, T. W. F., and Carlin, B. P. (1998). "Generalized Linear Models for Small-Area Estimation." *Journal of the American Statistical Association*, 93(441): 273–282.

Ghosh, M. and Rao, J. N. K. (1994). "Small Area Estimation: An Appraisal." *Statistical Science*, 9(1): 55–93.

Groundwater Ambient Monitoring and Assessment Program (2019). "GAMA Data Download." Available from https://gamagroundwater.waterboards.ca.gov/gama/datadownload [last accessed June 1, 2019].

Guhaniyogi, R. and Banerjee, S. (2018). "Meta-Kriging: Scalable Bayesian Modeling and Inference for Massive Spatial Datasets." *Technometrics*, 60(4): 430–444.

Harter, T., Dzurella, K., Kourakos, G., Hollander, A., Bell, A., Santos, N., Hart, Q., King, A., Quinn, J., Lampinen, G., Liptzin, D., Rosenstock, T., Zhang, M., Pettygrove, G., and Tomich, T. (2017). "Nitrogen Fertilizer Loading to Groundwater in the Central Valley, Final Report to the Fertilizer Research Education Program Projects 11-0301 and 15-0454." Technical report, California Department of Food and Agriculture and University of California Davis, Davis, CA.

Harter, T. and Lund, J. R. (2012). "Addressing Nitrate in California's Drinking Water with a Focus on Tulare Lake Basin and Salinas Valley Groundwater, Report for the State Water Resources Control Board report to the Legislature ? executive summary." Technical report, Center for Watershed Sciences, University of California, Davis, Davis, CA.

Hartley, H. O. and Sielken Jr., R. L. (1975). "A "Super-Population Viewpoint" for Finite Population Sampling." *Biometrics*, 31(2): 411–422.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A. (2018). "A Case Study Competition Among Methods for Analyzing Large Spatial Data." *Journal of Agricultural, Biological and Environmental Statistics*.

Horvitz, D. G. and Thompson, D. J. (1952). "A Generalization of Sampling Without Replacement From a Finite Universe." *Journal of the American Statistical Association*, 47(260): 663–685.

Little, R. J. (2004). "To Model or Not To Model? Competing Modes of Inference for Finite Population Sampling." *Journal of the American Statistical Association*, 99(466): 546–556.

Malec, D. and Sedransk, J. (1985). "Bayesian Inference for Finite Population Parameters in Multistage Cluster Sampling." *Journal of the American Statistical Association*, 80(392): 897–902.

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

Plummer, M. (2017). *JAGS Version 4.3.0 user manual*. International Agency for Research on Cancer, Lyon, France.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Rao, J. N. K. (2003). *Small Area Estimation*. Hoboken, NJ: John Wiley & Sons.

Ripley, B. D. (2004). *Spatial Statistics*. Hoboken, NJ: John Wiley & Sons.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York, NY: Springer, 2nd edition.

Royall, R. M. (1970). "On finite population sampling theory under certain linear regression models." *Biometrika*, 57(2): 377–387.

Rubin, D. B. (1976). "Inference and Missing Data." *Biometrika*, 63(3): 581–592.

Scott, A. and Smith, T. M. F. (1969). "Estimation in Multi-Stage Surveys." *Journal of the American Statistical Association*, 64(327): 830–840.

Sugden, R. A. and Smith, T. M. F. (1984). "Ignorable and Informative Designs in Survey Sampling Inference." *Biometrika*, 71(3): 495–506.

Vehtari, A., Gelman, A., and Gabry, J. (2017). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC." *Statistics and Computing*, 27(5): 1413–1432.

Ver Hoef, J. (2002). "Sampling and geostatistics for spatial data." *Écoscience*, 9(2): 152–161.

Walker, K. (2018). *tigris: Load Census TIGER/Line Shapefiles*. R package version 0.7.
URL https://CRAN.R-project.org/package=tigris

Zangeneh, S. Z. and Little, R. J. A. (2015). "Bayesian inference for the finite population total from heteroscedastic probability proportional to size sample." *Journal of Survey Statistics and Methodology*, 3: 162–192.

This section first provides the derivation of the empirical Bayesian estimators presented in Section II and then the finite population estimates presented in Sections II and III.

The values a^*, b^*, M_ν, m_ν, M_β, and m_β for the general case (9) in Section II are presented below.

$$a^* = a + \frac{n}{2}$$

$$b^* = b + \frac{1}{2}\left(y_s^\top y_s - y_s^\top \tilde{V}_s^{-1} X_s^\top [(\tilde{V}_\beta + A\tilde{V}_\nu A^\top)^{-1} + X_s^\top \tilde{V}_s^{-1} X_s]^{-1} X_s^\top \tilde{V}_s^{-1} y_s\right)$$

$$M_\nu = \left[\tilde{V}_\nu^{-1} - A^\top \tilde{V}_\beta^{-1}(\tilde{V}_\beta^{-1} + A^\top X_s^\top \tilde{V}_s^{-1} X_s A)^{-1}\tilde{V}_\beta^{-1} A^\top\right]^{-1}$$

$$m_\nu = A^\top \tilde{V}_\beta^{-1}(\tilde{V}_\beta^{-1} + A^\top X_s^\top \tilde{V}_s^{-1} X_s A)^{-1} A^\top X_s^\top \tilde{V}_s^{-1} y_s$$

$$M_\beta = \left[\tilde{V}_\beta^{-1} + A^\top X_s^\top \tilde{V}_s^{-1} X_s A\right]^{-1}$$

$$m_\beta = \tilde{V}_\beta^{-1} A\nu + A^\top X_s^\top \tilde{V}_s^{-1} y_s$$

The derivation of these values arise the conjugacy of the Normal and Inverse-Gamma distributions. We first derive (11) and (12). Take $\epsilon_s \sim N(0, \delta^2 \tilde{V}_s^{(\sigma)})$, $\epsilon_{ns} \sim N(0, \delta^2 \tilde{V}_{ns}^{(\sigma)})$, and $\nu \sim N(0, \delta^2 \tilde{\gamma}^2)$, where $\tilde{\gamma}^2 = \frac{1}{\delta^2}\gamma^2$, $\tilde{V}_s^{(\sigma)} = \frac{1}{\delta^2} V_s^{(\sigma)} = \left[\oplus_{i=1}^n \frac{\sigma_i^2}{\delta^2} I_{m_i}\right]$, and $\tilde{V}_{ns}^{(\sigma)} = \frac{1}{\delta^2} V_{ns}^{(\sigma)} = \left[\oplus_{i=1}^n \frac{\sigma_i^2}{\delta^2} I_{(M_i - m_i)}\right]$. Since the elements of y are independent conditional on μ, $V_{s,ns} = 0$ and $V_{ns,s} = 0$. Define observed group means as $\bar{y}_i = \frac{1}{m_i}\sum_{j=1}^n y_{ij}$ and the ratio of variances as $\lambda_i = \delta^2/(\delta^2 + \sigma_i^2/m_i)$ if $i \in (1, \ldots, n)$ and $\lambda_i = 0$ if $i \in (n+1, \ldots, N)$. Also define the vector of observed group variances to be $\tilde{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_n^2]^\top$. Recall $\mu_s = \nu 1_n + \epsilon_{\mu_s}; \epsilon_{\mu_s} \sim N(0, \delta^2 I_n)$, then $y_s = X_{s1}\mu_s + \epsilon_s = X_{s1}1_n\nu + e_s^*$, where $e_s^* \sim N(0, \delta^2[X_{s1}X_{s1}^\top + \tilde{V}_s^{(\sigma)}])$. Then we have that $p(\nu \mid y_s) \propto N(\nu \mid 0, \delta^2\tilde{\gamma}^2) \times N(y_s \mid X_{s1}1_n\nu, \delta^2[X_{s1}X_{s1}^\top + \tilde{V}_s^{(\sigma)}]) \propto N(\nu \mid Bb, \delta^2 B)$, where

$$b = 1_n^\top X_{s1}^\top (X_{s1}X_{s1}^\top + \tilde{V}_s^{(\sigma)})^{-1} y_s = [1_{m_1}^\top, \ldots, 1_{m_n}^\top]\left[\oplus_{i=1}^n \left(\frac{\delta^2}{\sigma_i^2}\right)\left(I_{m_i} - \frac{\frac{\delta^2}{\sigma_i^2}1_{m_i}1_{m_i}^\top}{1 + \frac{\delta^2}{\sigma_i^2}m_i}\right)\right] y_s \text{ and}$$

$$B^{-1} = \frac{1}{\tilde{\gamma}^2} + 1_n^\top X_{s1}^\top (X_{s1}X_{s1}^\top + \tilde{V}_s^{(\sigma)})^{-1} X_{s1}1_n = \frac{1}{\tilde{\gamma}^2} + \sum_{i=1}^n \left(\frac{\delta^2}{\sigma_i^2}\right) 1_{m_i}^\top \left(I_{m_i} - \frac{\frac{\delta^2}{\sigma_i^2}1_{m_i}1_{m_i}^\top}{1 + \frac{\delta^2}{\sigma_i^2}m_i}\right) 1_{m_i}.$$

Therefore, $B = \left[\frac{1}{\tilde{\gamma}^2} + \sum_{i=1}^n \lambda_i\right]^{-1}$ and $Bb = \frac{\sum_{i=1}^n \lambda_i \bar{y}_i}{\frac{1}{\tilde{\gamma}^2} + \sum_{i=1}^n \lambda_i}$. To solve $p(\delta^2 \mid y_s)$, split the posterior conditional distribution of the superpopulation parameters; $p(\delta^2, \nu \mid y_s) = IG(\delta^2 \mid a_\delta^*, b_\delta^*) \times N(\nu \mid Bb, \delta^2 B)$, where $a_\delta^* = a_\delta + \frac{k}{2}$ and $b_\delta^* = b_\delta + \frac{1}{2}[y_s^\top (X_{s1}X_{s1}^\top + \tilde{V}_s^{(\sigma)})^{-1} y_s + b^\top Bb]$. As $p(\mu_{ns} \mid y_s, \nu, \delta^2) = p(\mu_{ns} \mid \nu, \delta^2)$, $\mu_{ns} \mid \nu, \delta^2 \sim N(1_{N-n}\nu, \delta^2 I_{N-n})$. To solve $p(\mu_s \mid y_s, \nu, \delta^2)$, note that $p(\mu_s \mid y_s, \nu, \delta^2) \propto N(\mu_s \mid 1_n\nu, \delta^2 I_n) \times N(y_s \mid X_{s1}\mu_s, \delta^2 \tilde{V}_s^{(\sigma)}) \propto N(\mu_s \mid B_* b_*, \delta^2 B_*)$, where

$$b_* = \nu 1_n + X_{s1}^\top (\tilde{V}_s^{(\sigma)})^{-1} y_s = \left[\nu + \frac{\delta^2}{\sigma_1^2} m_1 \bar{y}_1, \cdots, \nu + \frac{\delta^2}{\sigma_n^2} m_n \bar{y}_n\right]^\top,$$

$$B_*^{-1} = I_n + X_{s1}^\top (\tilde{V}_s^{(\sigma)})^{-1} X_{s1} = \left[\oplus_{i=1}^n \frac{\sigma_i^2 + \delta^2 m_i}{\sigma_i^2}\right]; \quad B_* = \left[\oplus_{i=1}^n \frac{\sigma_i^2}{\sigma_i^2 + \delta^2 m_i}\right] = \left[\oplus_{i=1}^n (1 - \lambda_i)\right], \text{ and}$$

$$B_* b_* = \left[\oplus_{i=1}^n (1 - \lambda_i)\right]\left[\nu + \frac{\delta^2}{\sigma_1^2}m_1\bar{y}_1, \cdots, \nu + \frac{\delta^2}{\sigma_n^2}m_n\bar{y}_n\right]^\top$$

$$= \left[(1 - \lambda_1)\nu + \lambda_1\bar{y}_1, \cdots, (1 - \lambda_n)\nu + \lambda_n\bar{y}_n\right]^\top.$$

To derive (13) and (14) define $\tilde{V}_\nu^{-1} = 0$ and $\tilde{V}_s = \tilde{\Omega}_s = \frac{1}{\delta^2}\Omega_s + I_k$, and $\tilde{V}_{\mu_s} = 1$. Note that $p(\mu_s \mid y_s, \delta^2) \propto N(\mu_s \mid 0, \delta^2) \times N(y_s \mid 1_k \mu_s, \delta^2 \tilde{\Omega}_s) \propto N(\mu_s \mid B_{\mu_s} b_{\mu_s}, \delta^2 B_{\mu_s})$, where $B_{\mu_s} = (1 + 1_k^\top \tilde{\Omega}_s^{-1} 1_k)^{-1}$ and $b_{\mu_s} = 1_k^\top \tilde{\Omega}_s^{-1} y_s$. Splitting the posterior conditional distribution of the superpopulation parameters, $p(\delta^2, \mu_s \mid y_s) = IG(\delta^2 \mid a_\delta^{**}, b_\delta^{**}) \times N(\mu_s \mid B_{\mu_s} b_{\mu_s}, \delta^2 B_{\mu_s})$, where $a_\delta^{**} = a + \frac{k}{2}$ and $b_\delta^{**} = b + \frac{1}{2} y_s^\top \left\{ \tilde{\Omega}_s^{-1} - \tilde{\Omega}_s^{-1} 1_k \left(1 + 1_k^\top \tilde{\Omega}_s^{-1} 1_k \right)^{-1} 1_k^\top \tilde{\Omega}_s^{-1} \right\} y_s$.

We now continue by deriving the general cases presented in Section II. As $\beta \mid \nu \sim N(A\nu, V_\beta)$ and $\nu \sim N(0, \gamma^2)$, we have that $\beta \sim N(0, \gamma^2 A A^\top + V_\beta)$. Note that $p(\beta \mid y_s) \propto N(0, \gamma^2 A A^\top + V_\beta) \times N(X_s \beta, V_s) \propto N(V_{\beta \mid y_s} X_s^\top V_s^{-1} y_s, V_{\beta \mid y_s})$, where $V_{\beta \mid y_s} = [(\gamma^2 A A^\top + V_\beta)^{-1} + X_s^\top V_s^{-1} X_s]^{-1}$. Defining $B_V = V_{ns,s} + (X_{ns} - V_{ns,s} V_s^{-1} X_s) V_{\beta \mid y_s} X_s^\top$ and $Q = X_{ns} - V_{ns,s} V_s^{-1} X_s$, we have that:

$$
\begin{aligned}
\mathrm{E}[\alpha^\top y \mid y_s] &= \alpha_s^\top y_s + \alpha_{ns}^\top \mathrm{E}[\mathrm{E}[y_{ns} \mid \beta, y_s] \mid y_s] \\
&= \alpha_s^\top y_s + \alpha_{ns}^\top \mathrm{E}[X_{ns}\beta + V_{ns,s} V_s^{-1}(y_s - X_s \beta) \mid y_s] \\
&= \alpha_s^\top y_s + \alpha_{ns}^\top V_{ns,s} V_s^{-1} y_s + Q V_{\beta \mid y_s} X_s^\top V_s^{-1} y_s \\
&= \{\alpha_s^\top + \alpha_{ns}^\top [V_{ns,s} + \alpha_{ns}^\top Q V_{\beta \mid y_s} X_s^\top] V_s^{-1}\} y_s \, , \\
\mathrm{Var}[\mathrm{E}[\alpha^\top y \mid y_s]] &= \mathrm{Var}[(\alpha_s^\top + \alpha_{ns}^\top B_V V_s^{-1}) y_s] \\
&= \alpha_s^\top V_s \alpha_s + 2\alpha_{ns}^\top B_V \alpha_s + \alpha_{ns}^\top B_V V_s^{-1} B_V^\top \alpha_{ns} \, , \text{ and} \\
\mathrm{Var}[\mathrm{E}[\alpha^\top y \mid y_s]] &= \mathrm{Var}[\alpha_{ns}^\top y_{ns} \mid y_s] \\
&= \mathrm{Var}[\mathrm{E}[\alpha_{ns}^\top y_{ns} \mid \beta, y_s] \mid y_s] + \mathrm{E}[\mathrm{Var}[\alpha_{ns}^\top y_{ns} \mid \beta, y_s] \mid y_s] \\
&= \mathrm{Var}[\alpha_{ns}^\top Q\beta \mid y_s] + \mathrm{E}[\alpha_{ns}^\top (V_{ns} - V_{ns,s} V_s^{-1} V_{s,ns}) \alpha_{ns} \mid y_s] \\
&= \alpha_{ns}^\top (Q V_{\beta \mid y_s} Q^\top + V_{ns} - V_{ns,s} V_s^{-1} V_{s,ns}) \alpha_{ns} \, .
\end{aligned}
$$

To derive the estimate given in Example 1, note that $p(\mu \mid y_s) \propto N(\mu \mid 0, \xi^2) \times N(y_s \mid 1_n \mu, \sigma^2 I_n) \propto N(\mu \mid B_{srs} b_{srs}, B_{srs})$, where $B_{srs} = (\frac{1}{\xi^2} + \frac{n}{\delta^2})^{-1}$, $b_{srs} = \frac{1}{\sigma^2} 1_n^\top y_s$, and $B_{srs} b_{srs} = \frac{\frac{1}{\sigma^2} 1_n^\top y_s}{\frac{1}{\xi^2} + \frac{n}{\sigma^2}} = \frac{\sum_{i=1}^n y_i}{\frac{\sigma^2}{\xi^2} + n}$. Fixing the variance components, the finite population estimate is

$$
\begin{aligned}
\mathrm{E}[\alpha^\top y \mid y_s] &= \alpha_s^\top y_s + \alpha_{ns}^\top \mathrm{E}[\mathrm{E}[y_{ns} \mid \mu, y_s] \mid y_s] = \alpha_s^\top y_s + \alpha_{ns}^\top 1_{(N-n)} \mathrm{E}[\mu \mid y_s] \\
&= \sum_{i=1}^n \alpha_i y_i + \frac{\sum_{i=n+1}^N \alpha_i}{\frac{\sigma^2}{\xi^2} + n} \sum_{i=1}^n y_i = \sum_{i=1}^n \left(\alpha_i + \frac{\sum_{i=n+1}^N \alpha_i}{\frac{\sigma^2}{\xi^2} + n} \right) y_i \, .
\end{aligned}
$$

To derive (4), it is helpful to first make a note regarding X_{s1} vs X_s in the calculation of $p(\nu \mid y_s)$ and $p(\mu \mid \nu, y_s)$. Specifically, $p(\nu \mid y_s)$ does not change, since $X_s = [X_{s1} : 0]$, $b = 1_n^\top X_{s1}^\top (X_{s1} X_{s1}^\top + \tilde{V}_s^{(\sigma)})^{-1} y_s = 1_N^\top X_s^\top (X_s X_s^\top + \tilde{V}_s^{(\sigma)})^{-1} y_s$. Similarly, $B^{-1} = \frac{1}{\tilde{\gamma}^2} + 1_n^\top X_{s1}^\top (X_{s1} X_{s1}^\top + \tilde{V}_s^{(\sigma)})^{-1} X_{s1} 1_n = \frac{1}{\tilde{\gamma}^2} + 1_N^\top X_s^\top (X_s X_s^\top + \tilde{V}_s^{(\sigma)})^{-1} X_s 1_N$. However, while computing $p(\mu_s \mid \nu, y_s)$ using X_{s1} is computationally convenient for interpretation, employing X_s provides us the posterior distribution $p(\mu \mid \nu, y_s)$. We have that $p(\mu \mid y_s, \nu) \propto N(\mu \mid \nu 1_N, \delta^2 I_N) \times N(y_s \mid X_s \mu, \delta^2 \tilde{V}_s^{(\sigma)}) \propto N(\mu \mid B_{**} b_{**}, \delta^2 B_{**})$. Some algebra simplifies the expressions for b_{**} and $B_{**} b_{**}$ and matches the

conclusions found by deriving $p(\mu_s \mid \nu, y_s)$ and $p(\mu_{ns} \mid \nu, y_s)$ separately:

$$b^{**} = \nu 1_N + X_s^\top (\tilde{V}_s^{(\sigma)})^{-1} y_s = \left[\nu + \frac{\delta^2}{\sigma_1^2} m_1 \bar{y}_1, \ldots, \nu + \frac{\delta^2}{\sigma_n^2} m_n \bar{y}_n, \nu 1_{(N-n)}^\top\right]^\top ;$$

$$B_{**}^{-1} = I_N + X_s^\top (\tilde{V}_s^{(\sigma)})^{-1} X_s = \begin{bmatrix} \oplus_{i=1}^n \frac{\sigma_i^2 + \delta^2 m_i}{\sigma_i^2} & 0 \\ 0 & I_{(N-n)} \end{bmatrix} ; \ B_{**} = \left[\oplus_{i=1}^N (1 - \lambda_i)\right] ; \text{and}$$

$$B_{**} b_{**} = \left[(1 - \lambda_1)\nu + \lambda_1 \bar{y}_1, \ldots, (1 - \lambda_n)\nu + \lambda_n \bar{y}_n, \nu 1_{(N-n)}^\top\right]^\top .$$

Using these derivations, define $\lambda = [\lambda_1, \ldots, \lambda_N]^\top$ and $\bar{y} = [\bar{y}_1, \ldots, \bar{y}_n, 0_{(N-n)}]^\top$. Then fixing the variance components, we have:

$$\mathrm{E}[\alpha^\top y \mid y_s] = \alpha_s^\top y_s + \alpha_{ns}^\top \mathrm{E}[\mathrm{E}[\mathrm{E}[y_{ns} \mid \mu, \nu, y_s] \mid \nu, y_s] \mid y_s] = \alpha_s^\top y_s + \alpha_{ns}^\top \mathrm{E}[\mathrm{E}[X_{ns} \mu \mid \nu, y_s] \mid y_s]$$

$$= \alpha_s^\top y_s + \alpha_{ns}^\top X_{ns} \mathrm{E}[[(1 - \lambda_1)\nu + \lambda_1 \bar{y}_1, \ldots, (1 - \lambda_n)\nu + \lambda_n \bar{y}_n, \nu 1_{(N-n)}^\top]^\top \mid y_s]$$

$$= \alpha_s^\top y_s + \alpha_{ns}^\top X_{ns} [\oplus_{i=1}^N (1 - \lambda_i)] 1_N \frac{\sum_{i=1}^n \lambda_i \bar{y}_i}{1/\tilde{\gamma}^2 + \sum_{i=1}^n \lambda_i} + \alpha_{ns}^\top X_{ns} [\oplus_{i=1}^N \lambda_i] \bar{y}$$

$$= \alpha_s^\top y_s + [\alpha_1 (1 - \lambda_1), \ldots, \alpha_N (1 - \lambda_N)]^\top 1_N \frac{\sum_{i=1}^n \lambda_i \bar{y}_i}{1/\tilde{\gamma}^2 + \sum_{i=1}^n \lambda_i} + [\alpha_1 \lambda_1, \ldots, \alpha_N \lambda_N]^\top \bar{y}$$

$$= \sum_{i=1}^n \sum_{j=1}^{m_i} \alpha_{ij} y_{ij} + \left\{\sum_{i=1}^N \alpha_i (1 - \lambda_i)\right\} \frac{\sum_{i=1}^n \frac{\lambda_i}{m_i} \sum_{j=1}^{m_i} y_{ij}}{1/\tilde{\gamma}^2 + \sum_{i=1}^n \lambda_i} + \sum_{i=1}^n \alpha_i \frac{\lambda_i}{m_i} \sum_{j=1}^{m_i} y_{ij}$$

$$= \sum_{i=1}^n \sum_{j=1}^{m_i} \left(\alpha_{ij} + \left[\alpha_i + \frac{\sum_{i=1}^N \alpha_i (1 - \lambda_i)}{1/\tilde{\gamma}^2 + \sum_{i=1}^n \lambda_i}\right] \frac{\lambda_i}{m_i}\right) y_{ij}$$

Now consider the stratified case for estimating the population mean. Taking non-informative priors for the group means, μ, is equivalent to letting $\delta^2 \to \infty$ and $\gamma^2 \to \infty$. Therefore $\lambda_i = \frac{\delta^2}{\delta^2 + \sigma_i^2 / m_i} \to 1$, for all $i = 1, \ldots, N$. Note $\alpha_i = \sum_{j=m_i+1}^{M_i} \frac{1}{T} = \frac{M_i - m_i}{T}$, $i = 1, \ldots, n$. We have that:

$$\lim_{\delta^2, \gamma^2 \to \infty} \mathrm{E}\left[\frac{1}{T} 1_T^\top y \mid y_s\right] = \sum_{i=1}^n \sum_{j=1}^{m_i} \left(\frac{1}{T} + \left[\frac{M_i - m_i}{T} + 0\right] \frac{1}{m_i}\right) y_{ij} = \sum_{i=1}^n \frac{M_i}{T} \bar{y}_i$$

To derive (6), note $p(\nu \mid y_s, \tau^2, \Omega_s, V_s^{(\sigma)}) \propto N(y_s \mid X_s 1_N \nu, \delta^2 X_s X_s^\top + \Omega_s + V_s^{(\sigma)}) \times N(\nu \mid 0, \gamma^2) \propto N(\nu \mid B_{sp2} b_{sp2}, B_{sp2})$, where $B_{sp2} = (\frac{1}{\gamma^2} + 1_N^\top X_s^\top (\delta^2 X_s X_s^\top + \Omega_s + V_s^{(\sigma)})^{-1} X_s 1_N)^{-1}$ and $b_{sp2} = 1_N^\top X_s^\top (\delta^2 X_s X_s^\top + \Omega_s + V_s^{(\sigma)})^{-1} y_s$. Consider the non-spatial case and define $\lambda^\top = [\lambda_1, \ldots, \lambda_N] = 1_N^\top X_s^\top (X_s X_s^\top + \tilde{V}_s^{(\sigma)})^{-1} X_s$, then $B = \left[\frac{1}{\tilde{\gamma}^2} + 1_N^\top X_s^\top (X_s X_s^\top + \tilde{V}_s^{(\sigma)})^{-1} X_s 1_N\right]^{-1} = \left[\frac{1}{\tilde{\gamma}^2} + \lambda^\top 1_N\right]^{-1} = \left[\frac{1}{\tilde{\gamma}^2} + \sum_{i=1}^n \lambda_i\right]^{-1}$, which agrees with our previous findings.

Similarly, define $\lambda^{*\top} = [\lambda_1^*, \ldots, \lambda_N^*] = 1_N^\top X_s^\top (\delta^2 X_s X_s^\top + \Omega_s + V_s^{(\sigma)})^{-1} X_s$.

Then $B_{sp2} = \left[\frac{1}{\tilde{\gamma}^2} + \lambda^{*\top} 1_N\right]^{-1} = \left[\frac{1}{\tilde{\gamma}^2} + \sum_{i=1}^n \lambda_i^*\right]^{-1}$.

Additionally, $p(\mu \mid y_s, \nu) \propto N(\mu \mid \nu 1_N, \delta^2 I_N) \times N(y_s \mid X_s \mu, \Omega_s + V_s^{(\sigma)}) \propto N(\mu \mid B_{sp2*} b_{sp2*}, B_{sp2*})$. Here $B_{sp2*} = (\frac{1}{\delta^2} I_N + X_s^\top (\Omega_s + V_s^{(\sigma)})^{-1} X_s)^{-1}$ and $b_{sp2*} = \frac{1}{\delta^2} 1_N \nu + X_s^\top (\Omega_s + V_s^{(\sigma)})^{-1} y_s$.

Fixing the variance parameters, we have that:

$$
\begin{aligned}
\mathrm{E}[\alpha^\top y \mid y_s] &= \alpha_s^\top y_s + \alpha_{ns}^\top \mathrm{E}[\mathrm{E}[\mathrm{E}[y_{ns} \mid \mu, \nu, y_s] \mid \nu, y_s] \mid y_s] \\
&= \alpha_s^\top y_s + \alpha_{ns}^\top \mathrm{E}[\mathrm{E}[X_{ns}\mu + \Omega_{ns,s}(\Omega_s + V_s^{(\sigma)})^{-1}(y_s - X_s\mu) \mid \nu, y_s] \mid y_s] \\
&= \alpha_s^\top y_s + \alpha_{ns}^\top \Omega_{ns,s}(\Omega_s + \sigma^2 V_s^{(\sigma)})^{-1} y_s + \alpha_{ns}^\top[X_{ns} - \Omega_{ns,s}(\Omega_s + V_s^{(\sigma)})^{-1} X_s]\times \\
&\quad \mathrm{E}\left[\left(\tfrac{1}{\delta^2}I_N + X_s^\top(\Omega_s + V_s^{(\sigma)})^{-1}X_s\right)^{-1}\left(\tfrac{1}{\delta^2}1_N\nu + X_s^\top(\Omega_s + V_s^{(\sigma)})^{-1}y_s\right) \mid y_s\right] \\
&= \alpha_s^\top y_s + \alpha_{ns}^\top \Omega_{ns,s}(\Omega_s + \sigma^2 V_s^{(\sigma)})^{-1} y_s + \alpha_{ns}^\top[X_{ns} - \Omega_{ns,s}(\Omega_s + V_s^{(\sigma)})^{-1} X_s]\times \\
&\quad \left(\tfrac{1}{\delta^2}I_N + X_s^\top(\Omega_s + V_s^{(\sigma)})^{-1}X_s\right)^{-1} X_s^\top(\Omega_s + V_s^{(\sigma)})^{-1}y_s + \\
&\quad \alpha_{ns}^\top[X_{ns} - \Omega_{ns,s}(\Omega_s + V_s^{(\sigma)})^{-1} X_s]\times \\
&\quad \left(\tfrac{1}{\delta^2}I_N + X_s^\top(\Omega_s + V_s^{(\sigma)})^{-1}X_s\right)^{-1} \frac{1}{\delta^2} 1_N \frac{1_N^\top X_s^\top(\delta^2 X_s X_s^\top + \Omega_s + V_s^{(\sigma)})^{-1}y_s}{\frac{1}{\gamma^2} + \sum_{i=1}^n \lambda_i^*} \\
&= \sum_{i=1}^n \sum_{j=1}^{m_i}\left(\alpha_{ij} + \alpha_{ns}^\top \left\{\Omega_{ns,s}(\Omega_s + V_s^{(\sigma)}) + \right.\right. \\
&\quad [X_{ns} - \Omega_{ns,s}(\Omega_s + V_s^{(\sigma)})^{-1} X_s]\left(\tfrac{1}{\delta^2}I_N + X_s^\top(\Omega_s + V_s^{(\sigma)})^{-1}X_s\right)^{-1}\times \\
&\quad \left.\left.\left[X_s^\top(\Omega_s + V_s^{(\sigma)})^{-1} + \frac{\frac{1}{\delta^2}1_N 1_N^\top X_s^\top(\delta^2 X_s X_s^\top + \Omega_s + V_s^{(\sigma)})^{-1}}{\frac{1}{\gamma^2} + \sum_{i=1}^n \lambda_i^*}\right]\right\} q_{ij}\right) y_{ij} \, .
\end{aligned}
$$

APPENDIX B

To perform the two-stage procedure implemented in Section II, sample means, $\hat{\mu}_i$, and sample variances, $\hat{\sigma}_i^2$, were calculated from each observed region, $i = 1, \dots, 25$. The variance matrix of the sampled units was fixed to be $\tilde{V}_s^{(\sigma)} = \left[\oplus_{i=1}^n \frac{\hat{\sigma}_i^2}{\mathrm{Var}(\hat{\mu})} I_{m_i}\right]$ where $\mathrm{Var}(\hat{\mu})$ represents the sample variance of the observed sample means. Similar to fixing $V_s^{(\sigma)}$, the variance matrix of the nonsampled units was fixed at $V_{ns} = \left[\oplus_{i=1}^N \frac{\tilde{\sigma}_i^2}{\mathrm{Var}(\hat{\mu})} I_{M_i - m_i}\right]$, where $\tilde{\sigma}_i^2 = \hat{\sigma}_i^2$ if $i \in \{1, \dots, n\}$ and $\tilde{\sigma}_i^2 = \frac{1}{n}\sum_{i=1}^n \hat{\sigma}_i^2$ if $i \in \{n+1, \dots, N\}$. The value of γ^2 was fixed to be half of the value of δ^2, reflecting the belief that there was less variability in the population mean than between group means. The prior distribution for δ^2 was assigned to be $IG(3,5)$. Sampling from the posterior was performed using the conditional distributions and methods described in Section II. As we have fixed the ratios of all variance components, we have also fixed $\tilde{\lambda}_i = \mathrm{Var}(\hat{\mu})/(\mathrm{Var}(\hat{\mu}) + \hat{\sigma}_i^2/m_i)$ if $i \in (1, \dots, n)$ and $\tilde{\lambda}_i = 0$ if $i \in (n+1, \dots, N)$.. Define $c = \frac{\sum_{i=1}^n \tilde{\lambda}_i \bar{y}_i}{\frac{1}{2} + \sum_{i=1}^n \tilde{\lambda}_i}$, $d = \left[\frac{1}{2} + \sum_{i=1}^n \tilde{\lambda}_i\right]^{-1}$, $c^{*(g)} = \begin{bmatrix} (1 - \tilde{\lambda}_1)\nu^{(g)} + \tilde{\lambda}_1 \bar{y}_1 \\ \vdots \\ (1 - \tilde{\lambda}_n)\nu^{(g)} + \tilde{\lambda}_n \bar{y}_n \end{bmatrix}$, and $d^* = \left[\oplus_{i=1}^n (1 - \tilde{\lambda}_i)\right]$. The following procedure was implemented to produce posterior estimates of the population mean, $\bar{y}^{(g)}$, for G iterations.

for(g in 1:G){

$$\delta^{2(g)} \sim IG\left(3 + \frac{1}{2}\sum_{i=1}^{n} m_i, 5 + \frac{1}{2}\left[y_s^\top(\tilde{V}_s^{(\sigma)} + X_s X_s^\top)^{-1} y_s + \frac{c^2}{d}\right]\right)$$

$$\nu^{(g)} \sim N(c, \delta^{2(g)}d)$$

$$\mu_s^{(g)} \sim N(c^{*(g)}, \delta^{2(g)}d^*)$$

$$\mu_{ns}^{(g)} \sim N(\nu^{(g)}1_n, \delta^{2(g)}I_{N-n})$$

$$y_{ns}^{(g)} \sim N(X_{ns}\mu^{(g)}, \delta^{2(g)}\tilde{V}_{ns})$$

$$\bar{y}^{(g)} = \frac{1}{T}\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{N}\sum_{j=m_i+1}^{M_i} y_{ij}^{(g)}\right)$$

}

Figure 4 recreates the centered mean plots presented in Figure 2 for the larger data case, in which the number of regions is 324. As in the $N = 100$ case, the point estimates and 95% credible intervals are similar for the two models. Additionally, the regional spatial model still outperforms the two-stage model with a WAIC of 1,052 (SE = 38.89) compared to 1,869 (SE = 34.77). MCMC sampling was performed using JAGS (Plummer, 2017) in R (R Core Team, 2018).
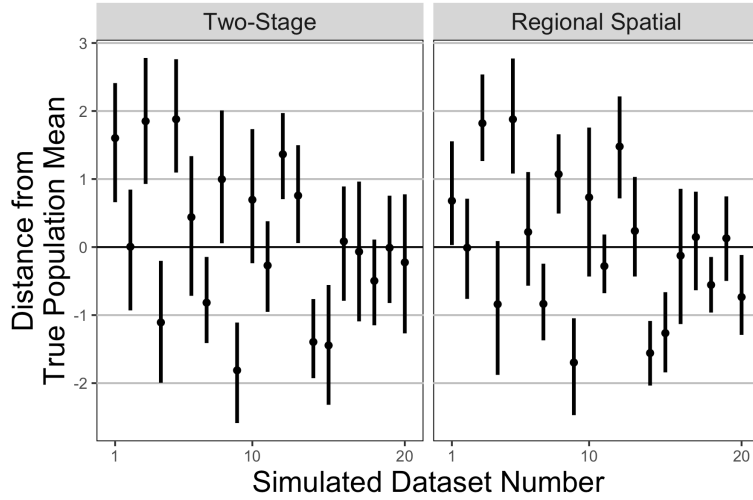

Figure 4: Centered Population Mean Estimates from 2 MCMC Models with 95% CI.

Table 3 compares the finite population mean estimate, ν estimate, and model fit for the first simulated dataset. Recall that $\nu = 2$ and the finite population mean was 2.60. Notice that while the true values are included in the credible intervals for all models, the credible intervals for ν are wider than those of the FP mean.

Table 3: Comparison of Estimates and Model Fit from MCMC

Model	FP Mean (95% CI)	ν (95% CI)	WAIC (SE)
1. Two-Stage	2.84 (1.93, 3.67)	2.84 (1.82, 3.82)	1870.26 (35.00)
2. Spatial	3.04 (1.33, 4.56)	3.24 (1.20, 5.01)	1912.70 (26.36)
3. Two-Stage + Spatial	3.11 (-0.86, 6.70)	3.22 (-1.23, 7.33)	455.67 (17.03)
4. Regional Spatial	2.56 (1.74, 3.37)	2.44 (1.49, 3.47)	1202.08 (38.99)

Appendix C

Figure 5 presents the 2010 zip code tabulation areas in the California Central Valley obtained from the *tigris* R package (Walker, 2018). Sixty-three zip codes were included in the analysis of a subset of observations from the UC Davis Nitrate Data in the data repository of the Groundwater Ambient Monitoring and Assessment Program (2019), presented in Section VI. These zip codes are denoted as either sampled or non-sampled, while all other zip codes are denoted as excluded.
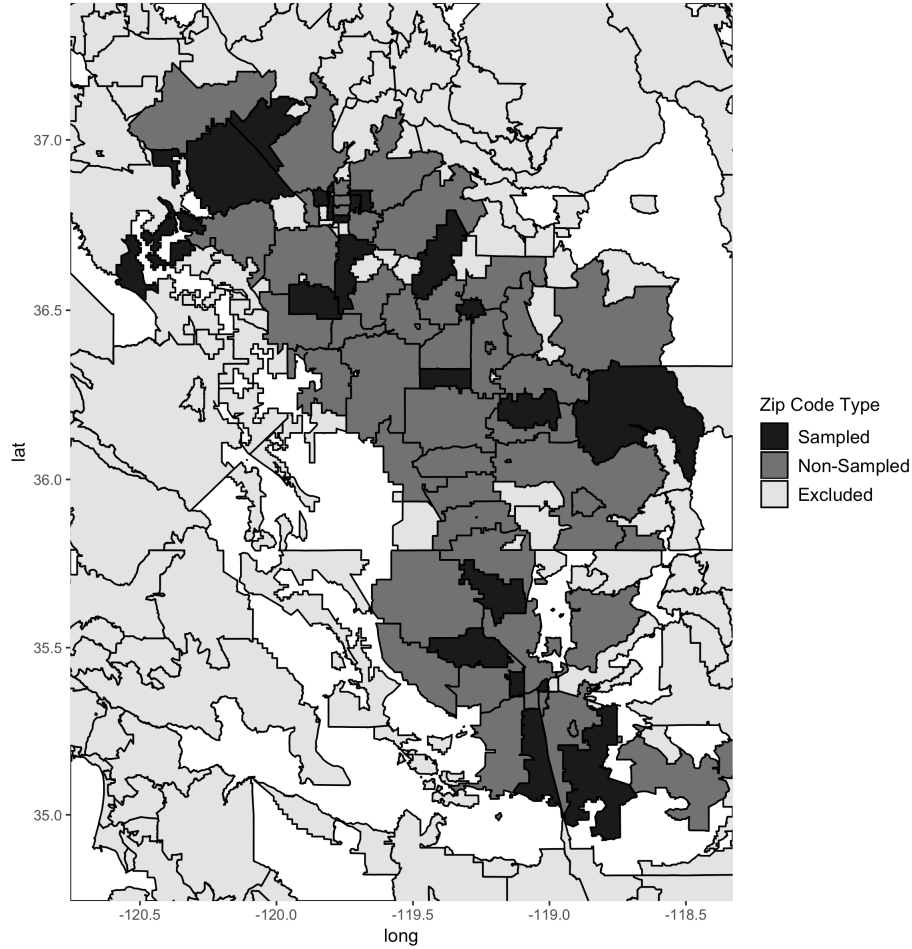

Figure 5: Plot of California zip code tabulation areas.

The variograms provided in Figure 6 were fitted to the entire population of wells (left) and sampled wells (right). For the population variogram, the estimated values of the nugget, partial sill, and range were 2527.4, 319.9, and 2.29, respectively. For the sample variogram, the estimated values of the nugget, partial sill, and range were 1808.6, 1014.4, and 335.9, respectively.

Figure 6: Variograms from Population and Sampled Data.

Figure 7 provides spatial residual plots arising from the three spatial models. The spatial model which does not account for regional effects sees the most dispersed spatial effects, while the two-stage + spatial and regional spatial models show more localized spatial variability. MCMC sampling was performed using JAGS (Plummer, 2017) in R (R Core Team, 2018).
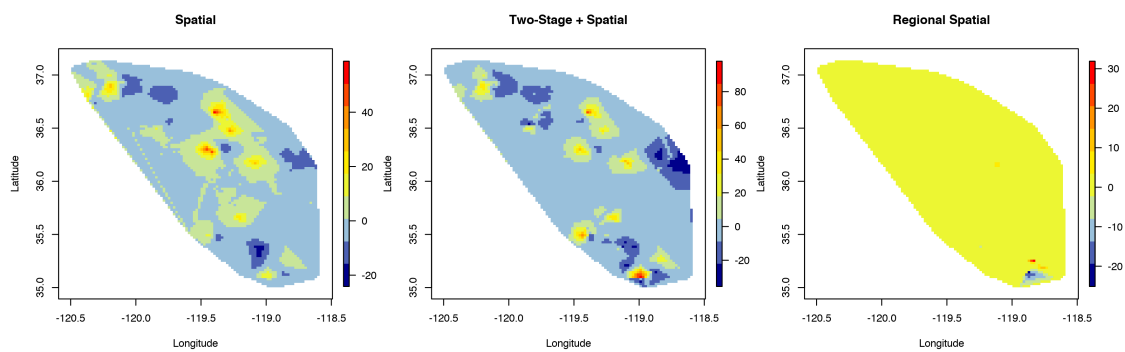
Figure 7: Spatial residual plots from the three spatial models.