The Annals of Applied Statistics
2021, Vol. 15, No. 1, 102–125
https://doi.org/10.1214/20-AOAS1388

AN APPROXIMATE BEST PREDICTION APPROACH TO SMALL AREA ESTIMATION FOR SHEET AND RILL EROSION UNDER INFORMATIVE SAMPLING

BY EMILY BERG[1,*] AND JAE-KWANG KIM[1,†]

[1]*Iowa State University,* [*]*emilyb@iastate.edu;* [†]*jkim@iastate.edu*

The National Resources Inventory, a longitudinal survey of characteristics related to natural resources and agriculture on nonfederal U.S. land, has increasingly received requests for substate estimates in recent years. We consider estimation of erosion in subdomains of the Boone-Raccoon River Watershed. This region is of interest for its proximity to intensively cropped areas as well as important waterbodies. The NRI application requires a small area prediction approach that can handle nonlinear relationships and appropriately incorporate survey weights that may have nontrivial relationships to the response variable. Because of the informative design, the conditional distribution required to define a standard empirical Bayes predictor is unknown. We develop a prediction approach that utilizes the approximate distribution of survey weighted score equations arising from a specified two-level superpopulation model. We apply the method to construct estimates of mean erosion in small watersheds. We investigate the robustness of the procedure to an assumption of a constant dispersion parameter and validate the properties of the procedure through simulation.

1. Introduction. The National Resources Inventory (NRI) is a longitudinal survey of nonfederal land in the United States. The NRI has traditionally published estimates at the state level (USDA, 2018). In recent years, requests for substate estimates have become more frequent. The Boone-Raccoon River watershed, which is primarily located in Central Iowa and Southern Minnesota, has been of interest from water quality and conservation perspectives. We consider prediction of soil loss due to sheet and rill erosion for the year 2007 in nine domains defined by hydrologic units in the Boone-Raccoon River watershed. We develop a small area prediction procedure that accounts for the complexity of the NRI survey and appropriately models nonlinear relationships observed in the data. In this Introduction we first provide further background on the motivating NRI application. We then explain why existing methods for small area estimation under informative sampling are not immediately applicable to our problem. Finally, we overview our approach and the organization of the rest of this article.

1.1. *NRI design and estimation procedures.* The NRI objectives necessitate complex design and estimation procedures. We provide an overview to motivate the importance of the sample weights in our study. We refer the reader to Nusser and Goebel (1997) for further detail.

A "foundation sample" of approximately 300,000 primary sampling units supported the publication of estimates every five years from 1982 to 1997. A typical primary sampling unit in the NRI is defined as a 0.5 mile^2 area of land. The segments are grouped into strata defined by a rectangular grid derived from the Public Land Survey System (PLSS) of the United States. The sampling rates within strata are unequal and typically range from 2% to 4%. Three point locations are selected from each segment. The "restricted randomization"

Received April 2020; revised August 2020.
Key words and phrases. EM algorithm, Fay–Herriot model.

procedure used to select the points maintains geographic spacing between sample points, thereby improving efficiency relative to a simple random sample. To ensure unbiased estimation, the restricted randomization procedure maintains the basic property of a probability sample; namely, every location in the segment has a chance to be selected into the sample of three points (Lohr, 2009). The same 300,000 segments in the foundation sample are observed every five years during 1982–1997. The technique of revisiting the same sampling units over time supports NRI's objective of providing reliable estimates of both gross and net change.

In the year 2000, the NRI began collecting data annually instead of every five years. The goals that motivated this change were to distribute the workload more evenly and to enable construction of annual estimates. Collecting data for all 300,000 segments in the foundation sample on an annual basis is too resource intensive. Therefore, the transition to annual data collection necessitated a change to the sample design, as detailed in Breidt and Fuller (1999).

To facilitate the annual data collection, the NRI transitioned to a supplemented panel design in 2000. The general idea of a supplemented panel design is to observe a portion of sampling units every year and to collect data for a different subset of units periodically. The set observed annually improves estimators of change over time. The supplementary segments observed periodically improve estimators of level for a particular year. The application of the supplemented panel design to the NRI involves dividing the foundation panel into two parts. A subsample of approximately 40,000 segments, designated as the "core" panel, is observed annually. Each year, a different subset of approximately 30,000 segments is selected for observation and designated as the "rotation panel" (or supplement). The union of the core and rotation samples comprise the annual sample of approximately 70,000 segments. The core and rotation samples are each selected from the foundation using stratified sampling. Characteristics observed for all segments in the foundation sample in the 1997 NRI define the strata boundaries. Strata representing categories that are judged to have a high probability of change or that are of interest to the USDA have relatively high sampling rates. The NRI design creates an unbalanced data structure. Segments in the core panel have data collected every year, while segments in the roation panels have planned missing values.

The NRI estimation procedures convert the raw, collected data to a user-friendly database with intuitive properties. Imputation procedures create a complete time series for every point in the final estimation data set. Raking ratio adjustments (Deville, Särndal and Sautory, 1993) are applied to inverse inclusion probabilities to preserve both external and internal control totals. Administrative sources supply the total area of the state, the federal area, the area in certain waterbodies and the agricultural area enrolled in the Conservation Reserve Program (CRP). The adjusted NRI sample weights approximately preserve these external control totals. The 1997 NRI is the last year in which all foundation segments were simultaneously observed. Therefore, the weights are adjusted to approximately maintain a subset of estimates obtained in the 1997 NRI. Following the transition to annual data collection in 2000, the NRI estimation procedures begin by constructing a set of state-level estimates that account for the temporal correlation induced by the supplemented panel design. A further set of adjustments preserves these state-level estimates.

The NRI uses the grouped jackknife (Wolter, 2007) for variance estimation. The jackknife maintains NRI's basic objective of providing a user-friendly data set, as the procedure transfers the burden of variance estimation from the data user to the data producer. Each record in the final estimation database is equipped with 29 replicate weights for variance estimation. Constructing the replicate weights is an intensive process that involves repeating the basic steps of estimation after omitting a subset of segments. After the difficult process of constructing the replicates is complete, a user can estimate the variance of nearly any parameter of interest with the supplied replicates. The resulting jackknife variance estimates appropriately reflect the NRI sample design and estimation procedures.

TABLE 1
Number of PSUs in each eight-digit hydrologic unit in the Boone-Raccoon River Watershed

	1	2	3	4	5	6	7	8	9
# PSU	295	120	129	171	85	249	117	198	133

Ultimately, each record in the final estimation data set has a complete time series, a single estimation weight and a set of 29 replicate weights for variance estimation. The final weights appropriately reflect the NRI design as well as the imputation and reweighting operations used in estimation. As such, the final NRI sample weight is the relevant weight for our analysis.

1.2. *Sheet and rill erosion in the NRI.* One of the primary characteristics of interest in the NRI is a measure of sheet and rill erosion defined by the Universal Soil Loss Equation (USLE). The USLE, which provides the response variable y for our analysis, is a product of five factors (Wischmeir and Smith, 1965). Specifically,

$$y = RK(LS)CP, \tag{1}$$

where R is a rainfall factor, K is a soil erodibility index, LS is a measure of slope length and steepness, C represents crop managements and P represents conservation practices. The estimate of mean erosion and the corresponding coefficient of variation for the full Boone-Raccoon River Watershed are 3.03 tons/acre and 7.35%, respectively. Watersheds are arranged hierarchically, and the small areas of interest are nine eight-digit hydrologic units nested in the four-digit Boone-Raccoon River Watershed. (The digits refer to a hierarchical labeling system, the details of which are irrelevant for our work.) As shown in Table 1, the number of primary sampling units in each eight-digit hydrologic unit ranges from 85 to 295. Despite the relatively large sample sizes for the eight-digit hydrologic units, several of the estimated CVs exceed 10%. Our goal for small area estimation will be to construct improved estimates of erosion for eight-digit hydrologic units in the Boone-Raccoon River Watershed.

Auxiliary variables related to the R and K factors of equation (1) are available for the full population from the Soil Survey of the Natural Resources Conservation Service. Although these auxiliary variables, called "rfactdcp" and "kffactdcp," are not exactly the USLE R and K factors, we refer to them as USLE R and K factors for short. Figure 1 shows a plot of y against K with symbols representing the five different values of R in the data set. The line in the plot is the lowess line relating y to K. The mean erosion increases with K, as expected. The variation in erosion values increases with the mean. The plot also indicates a right skew in the conditional distribution of erosion for a given value of K.

The nature of the NRI application introduces several challenges. First, nonlinear relationships between erosion and K make linear models inadequate. As we discuss in more detail in Section 3, evidence exists of nontrivial relationships between the erosion response variables and the NRI estimation weight, after accounting for the information in the R and K covariates. Therefore, accounting for the possibly informative nature of the complex NRI survey sample is important.

1.3. *Related literature*: *Nonlinear small area models and informative sampling.* The small area procedure that we define in Section 2 combines characteristics of the two main categories of small area models, unit-level models and area-level models. The foundational works of Battese, Harter and Fuller (1988) and Fay and Herriot (1979) use linear models for

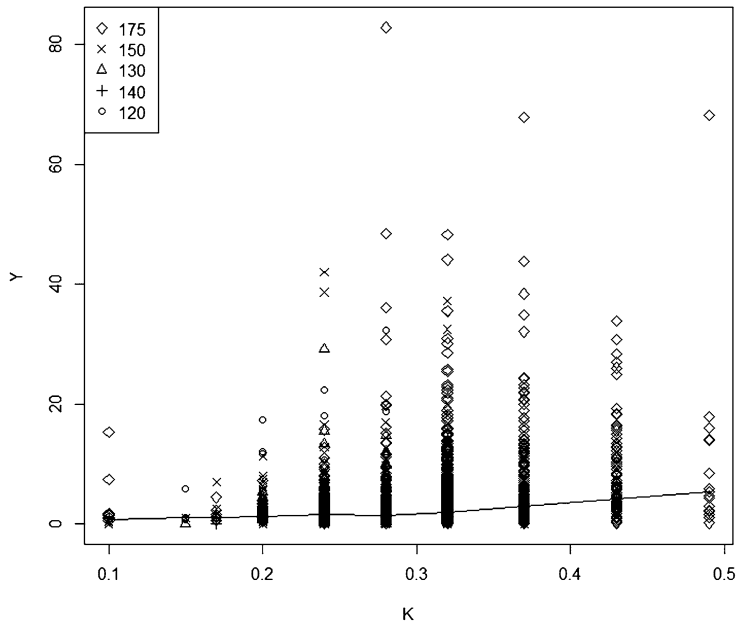

FIG. 1. *Plot of USLE erosion against K with symbols corresponding to R. Solid line is lowess fit of y on K.*

the unit-level and area-level model approaches, respectively. Numerous extensions of these approaches to nonlinear models have been developed. A comprehensive review is outside the scope of this paper but can be found in Jiang and Lahiri (2006), Pfeffermann (2013), Rao (2015) and Rao and Molina (2015). In the context of the NRI, Nandram and Sedransk (2002) and Erciulescu and Fuller (2019) define small area predictors of the proportion of land where an erosion tolerance threshold is exceeded using Bayesian and frequentist procedures, respectively. These methods do not consider the effect of the complex NRI design. The scatterplot of the NRI erosion data exhibits a right skewness, suggesting a skewed distribution for the unit-level responses, such as gamma or log-normal. Berg, Chandra and Chambers (2016) compare small area predictions based on log-normal and gamma distributions through simulation and conclude that the gamma distribution appears to be more robust. Dreassi, Petrucci and Rocco (2014) use a zero-inflated gamma distribution to model grape production in Italy. Jiang (2003) develops empirical Bayes small area predictors under a generalized linear mixed model.

The sample design is said to be informative for a specified small area model if the model for the population does not hold conditionally for the selected sample. If the sample design is informative for the specified small area model, then estimators that ignore the sampling design are biased. Small area procedures under an informative sampling design include You and Rao (2002), Pfeffermann and Sverchkov (2007) and Verret, Rao and Hidiroglou (2015).

You and Rao (2002) define a pseudo-EBLUP for a linear mixed effects model. Their approach begins with the linear unit-level model of Battese, Harter and Fuller (1988). They then modify the optimal predictor for the unit-level model to incorporate the sampling weights, replacing unweighted sums with corresponding survey-weighted sums. The You and Rao (2002) predictor of the small area mean is design-consistent as the small area sample size increases. The You and Rao (2002) estimator of the regression coefficient is design consistent, and the weight assigned to the direct estimator in the convex combination that defines the predictor approaches one as the area sample size increases. The You and Rao (2002) procedure is presented under the unit-level linear model. Ghosh and Maiti (2004) incorporate survey weights in predictors developed under an assumption that the distribution of the response is in the natural exponential quadratic variance function family, but their approach

only permits area level covariates. Our approach is closely related to You and Rao (2002) for the linear model but extends to nonlinear models. Unlike Ghosh and Maiti (2004), our prediction approach allows for unit-level covariates.

Pfeffermann and Sverchkov (2007) exploit relationships between the mean of the selection probabilities, the sample distribution, the population distribution and the inclusion probability to define a small area procedure for an informative design. Their procedure applies to nonlinear models but relies on an interpretation of the weight as an inverse inclusion probability. Because the NRI weight results from complex estimation operations, the approach of Pfeffermann and Sverchkov (2007) does not directly apply. Additionally, the approach of Pfeffermann and Sverchkov (2007) uses a parametric model for the inverse inclusion probability as a function of the response variable. An advantage of our procedure is that a model for the weight is not required.

Verret, Rao and Hidiroglou (2015) augment the unit-level model of Battese, Harter and Fuller (1988) with an additional covariate defined as a function of the selection probabilities. The augmented model approach can handle an informative sampling design successfully, but it requires the population mean of the selection probabilities. The approaches of Pfeffermann and Sverchkov (2007) and You and Rao (2002) only require the weights for sampled elements. Because the NRI weight is only available for sampled elements, the method of Verret, Rao and Hidiroglou (2015) does not directly apply.

You and Rao (2002), Pfeffermann and Sverchkov (2007), and Verret, Rao and Hidiroglou (2015) focus on the unit-level linear mixed effects model. Zimmermann and Munnich (2018) consider the unit-level log-normal model of Berg and Chandra (2014) under informative sampling. They adapt the procedures of Verret, Rao and Hidiroglou (2015) and of You and Rao (2002) to the log-normal model. As discussed above, these procedures are too restrictive to meet the needs of the NRI application.

The approach of composite likelihood has been used to construct estimators of regression parameters for linear and nonlinear mixed models in the presence of informative sampling (Asparouhov, 2006; Yi, Rao and Li, 2016). The composite likelihood does not furnish a natural predictor of a random effect because the appropriate conditional distribution to use is unclear. We solve this problem through the use of an approximate predictive distribution derived from the large-sample distribution of a set of estimating equations.

1.4. *Overview and innovation of proposed methodology.* The NRI application motivates us to develop a small area estimation procedure that permits a general distributional form and simultaneously accounts for an informative survey design. The crux of our approach is the large sample distribution of appropriately weighted estimating equations that arise from a specified model for the superpopulaton. We then define predictors of small area parameters based on an approximate predictive distribution that derives from the combination of the specified distribution of the area random effects and the approximate distribution of the weighted estimating equations. We formalize the approximate predictive distribution precisely in Section 2.

The small area procedure that we propose builds on the method developed in Kim, Park and Lee (2017) to estimate the parameters of a generalized linear mixed model under informative cluster sampling. While Kim, Park and Lee (2017) focus on estimating fixed model parameters, we develop a procedure to predict random small area parameters. A direct extension of Kim, Park and Lee (2017) to the small area estimation context would apply only to a particular type of cluster sample design in which primary sampling units are the same as the small areas. In practice, however, the small areas of interest may differ from design clusters and may even be defined after the sample is selected. The small area procedure that we propose applies to a general sample design, including an element sample design or a cluster design in which primary sampling units differ from small areas. In our NRI application

the NRI segments have no connection to the watersheds for which small area estimates are desired. A more subtle innovation of our work, relative to Kim, Park and Lee (2017), relates to variance estimation. We define an estimator of the MSE of the small area predictor that decomposes into a sum of two terms, analogous to the widely-used "g1" and "g2" of Prasad and Rao (1990). The first term is the conditional variance of the small area parameter given the observed estimating equation, calculated under the approximate predictive distribution. The second term accounts for the variance of the estimator of the fixed model parameter. We provide explicit expressions for the asymptotic variance of the fixed parameter estimator for gamma and Gaussian response distributions. The expressions for the variance estimator that we provide are simpler (and, therefore, more readily applicable) than the variance expression in Kim, Park and Lee (2017). A further innovation of our study, over Kim, Park and Lee (2017), is that we examine the robustness of our estimator, MSE estimator, and variance estimator to departures from the model assumptions. In particular, we consider the effects of a nonconstant dispersion parameter and the departure from an assumption of normally distributed random effects.

We use the approximate prediction procedure to construct small area estimates of erosion rates for watersheds nested in the Boone-Raccoon River watershed. We describe the procedure in general terms in Section 2. We apply a specific version of the approach that uses a gamma response distribution with a log link to the NRI data in Section 3. We validate the properties of the procedures through simulation in Section 4. Details of variance estimators are presented in the Appendix. A supplementary document contains additional simulations and details of the data analysis that are omitted from the main document for brevity.

2. Two-level model for small area estimation under informative sampling.

2.1. *Super-population, finite population, and samples for small area prediction.* Let y_{ij} be associated with the characteristic of interest for unit j in area i. Assume the population distribution for y_{ij} satisfies

$$y_{ij} \mid (v_i, \boldsymbol{x}_{ij}) \overset{\text{ind}}{\sim} f_1(y_{ij} \mid v_i, \boldsymbol{x}_{ij}, \boldsymbol{\psi}_1), \quad j = 1, \ldots, N_i, \tag{2}$$

where v_i is an area-level random effect with distribution

$$v_i \overset{\text{ind}}{\sim} f_2(v_i \mid \boldsymbol{z}_i, \boldsymbol{\psi}_2), \quad i = 1, \ldots, D, \tag{3}$$

$\boldsymbol{x}_{ij}$ is a vector of covariates that are known for $j = 1, \ldots, N_i$ and $i = 1, \ldots, D$ and $\boldsymbol{z}_i$ is a vector of area-level covariates. The $\boldsymbol{z}_i$ may include N_i, which implies that the area size can be informative (Neuhaus and McCulloch, 2011). The model should be specified such that $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ are identifiable, meaning that $(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \neq (\boldsymbol{\psi}_1', \boldsymbol{\psi}_2')$ implies

$$f(\boldsymbol{y} \mid \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \neq f(\boldsymbol{y} \mid \boldsymbol{\psi}_1', \boldsymbol{\psi}_2')$$

for at least one $\boldsymbol{y}$ in the sample space, where

$$f(\boldsymbol{y} \mid \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) = \prod_{i=1}^{D} \int_{-\infty}^{\infty} \prod_{j=1}^{N_i} f_1(y_{ij} \mid v_i, \boldsymbol{x}_{ij}, \boldsymbol{\psi}_1) f_2(v_i \mid \boldsymbol{z}_i, \boldsymbol{\psi}_2) \, dv_i.$$

For example, if the model has an intercept, then the intercept should be contained in $\boldsymbol{\psi}_1$ or $\boldsymbol{\psi}_2$ but not both. In the specific models that we work with, $\boldsymbol{\psi}_1$ contains regression parameters and a variance parameter, and f_2 is a normal distribution with mean 0 and variance ψ_2.

We assume a finite population is generated from the super-population model (2)–(3). The finite population consists of $\{y_{ij} : j = 1, \ldots, N_i\}$ for $i = 1, \ldots, D$. We assume that the unit-level covaraite $\boldsymbol{x}_{ij}$ is available for the full population of elements in area i and that the covariate $\boldsymbol{z}_i$ is observed for $i = 1, \ldots, D$. The finite population mean of interest is

$$\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}.$$

We now assume that a probability sample is selected from the finite population and let A_{2i} be the index set of the sample elements in area i. We observe y_{ij} for $j \in A_{2i}$ and $i \in A_1$, where A_1 denotes the set of areas with at least one sampled element. The precise definition of A_1 is $A_1 = \{i : |A_{2i}| > 0, i = 1, \ldots, D\}$. We assume that the survey furnishes a set of sample weights w_{ij} for $j \in A_{2i}$ and $i \in A_1$. Note that the sample weights are only required for elements in the sample (not for the full population). A simple weight is the inverse inclusion probability, $w_{ij} = \pi_{ij}^{-1}$. For designs in which the areas are the same as primary sampling units, π_{ij}^{-1} is of the form $\pi_{ij}^{-1} = \pi_{Ii}^{-1}\pi_{j|i}^{-1}$, where $\pi_{j|i}$ is the probability of selecting element j given area i and π_{Ii} is the inclusion probability of area i. For some surveys, such as the motivating NRI application, the weight w_{ij} may incorporate adjustments for missing data or calibration to control totals. A direct estimator of $\bar{Y}_i$, called a Hajek estimator, is defined as

$$\hat{\bar{Y}}_{wi} = \frac{\sum_{j \in A_{2i}} w_{ij} y_{ij}}{\sum_{j \in A_{2i}} w_{ij}}.$$

If the area sample size is small, then $\hat{\bar{Y}}_{wi}$ may have an undesirably large variance.

Our aim is to exploit the super-population model (2)–(3) to obtain a predictor of $\bar{Y}_i$ that is more efficient than $\hat{\bar{Y}}_{wi}$. To accomplish this goal, we consider the parameter defined by

$$\theta_i = \frac{1}{N_i} \sum_{j=1}^{N_i} E(y_{ij} \mid v_i, \boldsymbol{x}_{ij}, \boldsymbol{\psi}_1), \tag{4}$$

where the expectation is with respect to the distribution f_1 in (2). Since

$$(\bar{Y}_i - \theta_i) \mid v_i = O_p\left(\frac{1}{\sqrt{N_i}}\right),$$

we can safely ignore the error term in $\bar{Y}_i - \theta_i$. The small area parameter θ_i is a function of v_i, the area-specific random effect for area i. The best prediction of v_i, given the sample under squared error loss, is

$$v_i^* = E\big(v_i \mid \boldsymbol{x}_i^{(s)}, \boldsymbol{y}_i^{(s)}, z_i; \boldsymbol{\psi}_1, \boldsymbol{\psi}_2\big) = \frac{\int v_i f_1(\boldsymbol{y}_i^{(s)} \mid \boldsymbol{x}_i^{(s)}, v_i; \boldsymbol{\psi}_1) f_2(v_i \mid \boldsymbol{\psi}_2)\, dv_i}{\int f_1(\boldsymbol{y}_i^{(s)} \mid \boldsymbol{x}_i^{(s)}, v_i; \boldsymbol{\psi}_1) f_2(v_i \mid \boldsymbol{\psi}_2)\, dv_i},$$

where $\boldsymbol{x}_i^{(s)}$ and $\boldsymbol{y}_i^{(s)}$ are the sampled part of x_{ij} and y_{ij}, respectively, in area i. As discussed in Section 1.3, without additional assumptions we cannot obtain the joint density $f_1(\boldsymbol{y}_i^{(s)} \mid \boldsymbol{x}_i^{(s)}, v_i; \boldsymbol{\psi}_1)$ of the sample elements in area i from the population model. A new approach is proposed in the next subsection.

2.2. *Approximate best prediction.* To motivate the proposed method, we note that v_i are treated as fixed in (2), the level 1 model in the multilevel model structure. If the area random effects were treated as fixed effects, an estimating equation for v_i is given by

$$S_{1i}(v_i, \boldsymbol{\psi}_1) = \sum_{j \in A_{2i}} w_{ij} S_{1ij}(v_i, \boldsymbol{\psi}_1), \tag{5}$$

where

$$S_{1ij}(v_i, \boldsymbol{\psi}_1) = \frac{\partial}{\partial v_i} \log\{f_1(y_{ij} \mid v_i, \boldsymbol{x}_{ij}, \boldsymbol{\psi}_1)\}. \tag{6}$$

Let $\hat{v}_i(\boldsymbol{\psi}_1)$ satisfy $S_{1i}(\hat{v}_i(\boldsymbol{\psi}_1), \boldsymbol{\psi}_1) = 0$ for $i = 1, \ldots, D$. As shorthand, we use $\hat{v}_i(\boldsymbol{\psi}_1) = \hat{v}_i$, with the understanding that $\hat{v}_i$ depends on $\boldsymbol{\psi}_1$. The $\hat{v}_i$ is the profile pseudo maximum likelihood estimator (PMLE) of v_i, and it will play the central role in obtaining the approximate prediction model under informative sampling. We may assume

$$\hat{v}_i \mid v_i \sim g(\hat{v}_i \mid v_i, \boldsymbol{\psi}_1), \tag{7}$$

where g is a specified distribution for the sampling distribution of $\hat{v}_i$. We motivate a normal approximation for the distribution of $\hat{v}_i(\boldsymbol{\psi}_1)$ from an asymptotic framework in which $|A_{2i}| \to \infty$ as $N_i \to \infty$. Under common regularity conditions on the design and super-population model (Fuller, 2011), the Central Limit Theorem implies that

$$\hat{v}_i \mid v_i \dot{\sim} N(v_i, V_{1i}), \tag{8}$$

where $V_{1i} = \{H_i(v_i, \boldsymbol{\psi}_1)\}^{-1} V_{\pi i}\{S_{1i}(v_i, \boldsymbol{\psi}_1) \mid v_i\}\{H_i(v_i, \boldsymbol{\psi}_1)\}^{-1}$, $H_i(v_i, \boldsymbol{\psi}_1) = E\{\partial S_{1i}(v_i, \boldsymbol{\psi}_1)/(\partial v_i)\}$ and $V_{\pi i}\{S_{1i}(v_i, \boldsymbol{\psi}_1) \mid v_i\}$ is the design variance of $S_{1i}(v_i, \boldsymbol{\psi}_1)$ in (5), treating v_i as fixed. The specific form for $V_{\pi i}\{S_{1i}(v_i, \boldsymbol{\psi}_1) \mid v_i\}$ depends on the type of sample design under consideration as well as the definition of the sample weights. We discuss estimators of V_{1i} more specifically in Section 2.3 below.

If $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ were known, a predictor of a small area parameter θ_i would be defined by $E(\theta_i \mid \hat{v}_i, \boldsymbol{z}_i, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$, where, for any function $h(\cdot)$,

$$E\{h(v_i) \mid \hat{v}_i, \boldsymbol{z}_i, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2\} = \frac{\int_{-\infty}^{\infty} h(v_i) g(\hat{v}_i \mid v_i, \boldsymbol{\psi}_1) f_2(v_i \mid \boldsymbol{z}_i, \boldsymbol{\psi}_2)\, dv_i}{\int_{-\infty}^{\infty} g(\hat{v}_i \mid v_i, \boldsymbol{\psi}_1) f_2(v_i \mid \boldsymbol{z}_i, \boldsymbol{\psi}_2)\, dv_i}. \tag{9}$$

For predicting the small area mean, the function $h(\cdot)$ is defined as a function of v_i in (4). From the approximate normal distribution (8) for the survey-weighted score equation, one may take $g(\hat{v}_i \mid v_i) = \phi((\hat{v}_i - v_i)/\sqrt{V_{1i}})$. Since the choice for $g(\hat{v}_i \mid v_i)$ from (8) is based on approximate normality, we can call the prediction distribution in (9) as the approximate prediction distribution. The resulting predictor is then the approximate best predictor.

2.3. *Parameter estimation.* We now discuss parameter estimation. The basic idea is to find the marginal likelihood function of the parameters to obtain the maximum likelihood estimates. Since the marginal likelihood function involves integration using a random effect, a version of the EM algorithm can be developed.

The distributions (3) and (7) define an area level model for $\hat{v}_i$. The area level model is

$$\begin{aligned} \hat{v}_i \mid (v_i, \boldsymbol{\psi}_1) &\sim g(\hat{v}_i \mid v_i, \boldsymbol{\psi}_1), \\ v_i \mid (\boldsymbol{z}_i, \boldsymbol{\psi}_2) &\sim f_2(v_i \mid \boldsymbol{z}_i, \boldsymbol{\psi}_2). \end{aligned} \tag{10}$$

Note that we have now converted the problem of estimation for the unit-level model in (2) to the problem of estimation for the area-level model in (10). The profile PMLE $\hat{v}_i$ serves the role of direct estimator in the typical area level model.

Because the model (10) is a Fay–Herriot model for the case in which f_2 is a linear model with normally distributed random terms, we call the estimation procedure defined in this section the EM-FH algorithm. To use the model (10) for parameter estimation, we require an estimate of V_{1i} defined following (8). For a given value of $\boldsymbol{\psi}_1$, an estimate of V_{1i} is given by

$$\hat{V}_{1i}(\boldsymbol{\psi}_1) = \{\hat{H}_i(\hat{v}_i, \boldsymbol{\psi}_1)\}^{-1} \hat{V}_{\pi i}\{S_{1i}(\hat{v}_i, \boldsymbol{\psi}_1)\}\{\hat{H}_i(\hat{v}_i, \boldsymbol{\psi}_1)\}^{-1},$$

where $\hat{H}_i(\hat{v}_i, \boldsymbol{\psi}_1) = \partial S_{1i}(v_i, \boldsymbol{\psi}_1)/\partial v_i \mid_{v_i=\hat{v}_i}$ and $\hat{V}_{\pi i}\{S_{1i}(\hat{v}_i, \boldsymbol{\psi}_1)\}$ is a design-consistent estimator of $V_{\pi i}\{S_{1i}(v_i, \boldsymbol{\psi}_1) \mid v_i\}$ evaluated at $v_i = \hat{v}_i$. If $w_{ij} = \pi_{ij}^{-1}$, then one can use

$$\hat{V}_{\pi i}\{S_{1i}(v_i, \boldsymbol{\psi}_1) \mid v_i\} = \sum_{j \in A_{2i}} \sum_{k \in A_{2i}} \left(\frac{\pi_{ij,ik} - \pi_{ij}\pi_{ik}}{\pi_{ij,ik}\pi_{ij}\pi_{ik}} \right) S_{1ij}(v_i, \boldsymbol{\psi}_1) S_{1ik}(v_i, \boldsymbol{\psi}_1), \tag{11}$$

where $\pi_{ij,ik}$ is the joint inclusion probability defined as $\pi_{ij,ik} = P\{(ij) \in A \text{ and } (ik) \in A\}$ (Särndal, Swensson and Wretman, 2003), A denotes the set of sampled clusters and elements and $S_{1ij}(v_i, \boldsymbol{\psi}_1)$ is defined in (6). For the data analysis, replicate weights are used for variance estimation. Let $w_{ij}^{(r)}$ be the weight for unit (i, j) in replicate r. We define $\hat{V}_{1i}(\boldsymbol{\psi}_1)$ by

$$\hat{V}_{1i}(\boldsymbol{\psi}_1) = \sum_{r=1}^{R} (\hat{v}_i^{(r)} - \hat{v}_i)^2, \tag{12}$$

where $\hat{v}_i^{(r)}$ satisfies $S_{1i}^{(r)}(\hat{v}_i^{(r)}, \boldsymbol{\psi}_1) = 0$, with $S_{1i}^{(r)}(v_i, \boldsymbol{\psi}_1) = \sum_{j \in A_{2i}} w_{ij}^{(r)} S_{1ij}(v_i, \boldsymbol{\psi}_1)$.

Before presenting the proposed algorithm for estimating $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, we provide motivation. For a given value of $\boldsymbol{\psi}_1$, the marginal log-likelihood for $\boldsymbol{\psi}_2$ is defined by

$$\ell_w(\boldsymbol{\psi}_2 \mid \boldsymbol{\psi}_1) = \sum_{i \in A_1} w_i \log\left\{ \int_{-\infty}^{\infty} g(\hat{v}_i \mid v_i, \hat{V}_{1i}(\boldsymbol{\psi}_1)) f_2(v_i \mid \boldsymbol{z}_i, \boldsymbol{\psi}_2)\, dv_i \right\}, \tag{13}$$

where $w_i = \pi_{Ii}^{-1}$ for a cluster sample with areas as clusters and $w_i = 1$ for designs in which all areas have at least one sampled element. The area weight reflects the representativeness of the sample areas to the population of areas. In many cases, f_2 is the density of a normal distribution and so the marginal distribution in (13) is also normal.

In the proposed method we use the EM algorithm to estimate the level-one parameter $\boldsymbol{\psi}_1$ but use a direct maximization of the marginal log-likelihood in (13) to estimate the level-two parameter $\boldsymbol{\psi}_2$. The second step will be called FH-Step. The proposed EM-FH algorithm is as follows. For $t = 1, 2, \ldots, T$ complete the following steps:

[E-step]: Using the current estimates $\hat{\boldsymbol{\psi}}^{(t)} = (\hat{\boldsymbol{\psi}}_1^{(t)}, \hat{\boldsymbol{\psi}}_2^{(t)})$, compute the conditional expectation of $\boldsymbol{S}_{1ij}(\boldsymbol{\psi}_1; \boldsymbol{x}_{ij}, y_{ij}, v_i)$ conditional on $\hat{v}_i^{(t)}$ and $\boldsymbol{z}_i$ evaluated at $\hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\psi}}^{(t)}$, where

$$\boldsymbol{S}_{1ij}(\boldsymbol{\psi}_1; \boldsymbol{x}_{ij}, y_{ij}, v_i) = \frac{\partial}{\partial \boldsymbol{\psi}_1} \log\{f_1(y_{ij} \mid v_i, \boldsymbol{x}_{ij}, \boldsymbol{\psi}_1)\}.$$

The reference distribution for the E-step is

$$p(v_i \mid \hat{v}_i^{(t)}, \boldsymbol{z}_i; \hat{\boldsymbol{\psi}}^{(t)}) = \frac{g(\hat{v}_i^{(t)} \mid v_i; \hat{\boldsymbol{\psi}}_1^{(t)}) f_2(v_i \mid \boldsymbol{z}_i; \hat{\boldsymbol{\psi}}_2^{(t)})}{\int g(\hat{v}_i^{(t)} \mid v_i; \hat{\boldsymbol{\psi}}_1^{(t)}) f_2(v_i \mid \boldsymbol{z}_i; \hat{\boldsymbol{\psi}}_2^{(t)})\, dv_i}, \tag{14}$$

where $g(\hat{v}_i^{(t)} \mid v_i; \hat{\boldsymbol{\psi}}_1^{(t)})$ is the sampling distribution of $\hat{v}_i^{(t)}$ which is a normal distribution with mean v_i and variance $\hat{V}_{1i}(\hat{\boldsymbol{\psi}}_1^{(t)})$. The resulting mean score equation is

$$\bar{\boldsymbol{S}}_1(\boldsymbol{\psi}_1 \mid \hat{\boldsymbol{\psi}}^{(t)}) = \sum_{i \in A_1} \bar{\boldsymbol{S}}_{1i}(\boldsymbol{\psi}_1 \mid \hat{\boldsymbol{\psi}}^{(t)}),$$

where

$$\bar{\boldsymbol{S}}_{1i}(\boldsymbol{\psi}_1 \mid \hat{\boldsymbol{\psi}}^{(t)}) = \sum_{j \in A_{2i}} w_{ij} E\{\boldsymbol{S}_{1ij}(\boldsymbol{\psi}_1; \boldsymbol{x}_{ij}, y_{ij}, v_i) \mid \hat{v}_i^{(t)}, \boldsymbol{z}_i, \hat{\boldsymbol{\psi}}_1^{(t)}, \hat{\boldsymbol{\psi}}_2^{(t)}\}.$$

[M-step]: The updated $\hat{\boldsymbol{\psi}}_1^{(t+1)}$ satisfies $\bar{\boldsymbol{S}}(\hat{\boldsymbol{\psi}}_1^{(t+1)} \mid \hat{\boldsymbol{\psi}}_1^{(t)}) = \boldsymbol{0}$.

[FH-step]: The parameter $\boldsymbol{\psi}_2$ is updated by finding the maximizer of $\ell_w(\boldsymbol{\psi}_2 \mid \hat{\boldsymbol{\psi}}_1^{(t+1)})$ defined in (13) with respect to $\boldsymbol{\psi}_2$. That is, $\hat{\boldsymbol{\psi}}_2^{(t+1)} = \arg\max_{\boldsymbol{\psi}_2} \ell_w(\boldsymbol{\psi}_2 \mid \hat{\boldsymbol{\psi}}_1^{(t+1)})$.

[Direct Update]: We update the direct estimator of v_i and corresponding direct variance estimator with $\hat{v}_i^{(t+1)} = \hat{v}_i(\hat{\boldsymbol{\psi}}_1^{(t+1)})$ and $\hat{V}_{1i}(\hat{\boldsymbol{\psi}}_1^{(t+1)})$, respectively.

We denote the final parameter estimators by $\hat{\boldsymbol{\psi}}_1$ and $\hat{\boldsymbol{\psi}}_2$. The final direct estimator of v_i is denoted by $\hat{v}_i$. Elashoff and Ryan (2004) discusses convergence of EM-type algorithms for procedures based on estimating equations that are not necessarily score functions of likelihoods.

2.4. *Inference for small area parameters.* We define the small area predictor of θ_i by

$$\hat{\theta}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} E\{E(y_{ij} \mid v_i, \boldsymbol{x}_{ij}, \hat{\boldsymbol{\psi}}_1) \mid \hat{v}_i, z_i; \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2\}, \tag{15}$$

where the reference distribution for the inner expectation is f_1 of (2) and the reference distribution for the outer expectation is the approximate predictive distribution (14). The expectation under the approximate predictive distribution is defined in (9).

We estimate the MSE by

$$\widehat{mse}_i = \hat{V}\{\theta_i \mid \hat{v}_i, z_i; \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2\} + \hat{E}\{(\hat{\theta}_i - \tilde{\theta}_i)^2\}, \tag{16}$$

where $\hat{V}\{\theta_i \mid \hat{v}_i, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2\}$ estimates $V\{\theta_i \mid \hat{v}_i, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2\}$ and the second term, $\hat{E}\{(\hat{\theta}_i - \tilde{\theta}_i)^2\}$ estimates $E\{(\hat{\theta}_i - \tilde{\theta}_i)^2\}$ with $\tilde{\theta}_i = N_i^{-1} \sum_{j=1}^{N_i} E\{E(y_{ij} \mid v_i, \boldsymbol{x}_{ij}, \boldsymbol{\psi}_1) \mid \hat{v}_i, z_i; \boldsymbol{\psi}_1, \boldsymbol{\psi}_2\}$. Regardless of the sampling design, we estimate the leading term in the MSE by

$$\hat{V}\{\theta_i \mid \hat{v}_i, z_i; \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2\} = E(\theta_i^2 \mid \hat{v}_i, z_i; \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2) - \{E(\theta_i \mid \hat{v}_i, z_i; \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2)\}^2, \tag{17}$$

using the predictive distribution (14). The estimator of the second term in the sum (16), which accounts for parameter estimation, depends on the sample design. For designs in which clusters are not the same as areas, we use linearization variance estimation. For simplicity, we ignore uncertainty in the estimator of $\boldsymbol{\psi}_2$. Let $\hat{V}\{\hat{\boldsymbol{\psi}}_1\}$ be an estimator of the variance of $\hat{\boldsymbol{\psi}}_1$, such as that defined in Appendix B. Let $\partial\hat{\theta}_i/(\partial\boldsymbol{\psi}_1)$ denote the partial derivatives of $\hat{\theta}_i$ with respect to $\boldsymbol{\psi}_1$. Then, an estimate of the second term in the MSE is

$$\hat{E}\{(\hat{\theta}_i - \tilde{\theta}_i)^2\} = \{\partial\hat{\theta}_i/(\partial\boldsymbol{\psi}_1)\}' \hat{V}(\hat{\boldsymbol{\psi}}_1) \partial\hat{\theta}_i/(\partial\boldsymbol{\psi}_1). \tag{18}$$

We study the properties of the MSE estimator (18) through simulation in Section 4 and apply the estimator to the NRI data in Section 3. For a design in which the areas are clusters, we recommend the jackknife (Rao and Molina, 2015, page 301). In the Supplementary Material (Berg and Kim, 2021) we present simulation studies validating the properties of the jackknife MSE estimator for a cluster sample design. Because the clusters are not the same as areas for the NRI study, we defer a definition of the jackknife procedure to the supplement.

We considered alternatives to the MSE estimators suggested in this section. We found that the jackknife MSE estimator did not work well for the element sample design used in the simulation and, therefore, suggest the linearization based MSE estimator instead. The parametric bootstrap is not well suited to this problem because the design may be informative, and we do not specify a model for the relationship between the weights, the response variable and the covariates. Torabi and Rao (2010) compare a linearization based estimator of the MSE to a parametric bootstrap estimator of the MSE for the You and Rao (2002) predictor. In their simulations, the design is noninformative, as the weight is a function of the model covariate. If

we specify a model for the relationship between the weight, the response variable and the covariate, then one can define a parametric bootstrap estimator of the variance of the parameter estimators (Berg and Lee, 2019). If one does not specify a model for the weights and ignores the correlation between the weights and the response variables (i.e., implicitly assuming that the weight is uncorrelated with the response variable), then the resulting estimators of the variances of the parameter estimators can be biased. An advantage of the general small area model and procedure described in this section is that specifying a model for the weights is not required.

2.5. *Illustration for unit-level linear mixed model.* As an illustration, we present the steps of the EM-FH procedure for the widely used unit-level linear model (Battese, Harter and Fuller, 1988). To simplify the notation, we consider a univariate covariate x_{ij}. The two levels of the hierarchical model are defined as:

1. Level one model: $y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}, e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$,
2. Level two model: $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$

For the linear model we define $\hat{v}_i(\boldsymbol{\beta})$ as a function of $\boldsymbol{\beta} = (\beta_0, \beta_1)$ by

$$\hat{v}_i(\boldsymbol{\beta}) = \Big(\sum_{j \in A_{2i}} \pi_{ij}^{-1} \Big)^{-2} \sum_{j \in A_{2i}} \pi_{ij}^{-1} (y_{ij} - \beta_0 - \beta_1 x_{ij}).$$

For the linear model and a general design with $w_{ij} = \pi_{ij}^{-1}$, the estimator of the design variance of $\hat{v}_i$ is $\hat{V}_{1i}(\boldsymbol{\beta})$ given by

$$\hat{V}_{1i}(\boldsymbol{\beta}) = \Big(\frac{1}{\sum_{j \in A_{2i}} \pi_{ij}^{-1}} \Big)^2 \sum_{j \in A_{2i}} \sum_{k \in A_{2i}} \Big(\frac{\pi_{ij,ik} - \pi_{ij}\pi_{ik}}{\pi_{ij,ik}} \Big) \frac{\hat{e}_{ij}(\boldsymbol{\beta}) \hat{e}_{ik}(\boldsymbol{\beta})}{\pi_{ij}\pi_{ik}},$$

where $\hat{e}_{ij}(\boldsymbol{\beta}) = y_{ij} - \hat{v}_i(\boldsymbol{\beta}) - \beta_0 - \beta_1 x_{ij}$. The log likelihood for the marginal density of $\hat{v}_i$ is

$$\ell_w(\boldsymbol{\psi}_2 \mid \boldsymbol{\psi}_1) = -\frac{1}{2} \sum_{i \in A_1} w_i \Big\{ \frac{(\hat{v}_i(\boldsymbol{\beta}))^2}{(\sigma_v^2 + \hat{V}_{1i}(\boldsymbol{\beta}))^2} + \log\big(\sigma_v^2 + \hat{V}_{1i}(\boldsymbol{\beta})\big) \Big\}, \tag{19}$$

where $\boldsymbol{\psi}_2 = \sigma_v^2$, and $\boldsymbol{\psi}_1 = \boldsymbol{\beta}$. We ignore σ_e^2 because an estimate of σ_e^2 is not required for prediction or for variance estimation. Maximizing (19) with respect to $\boldsymbol{\psi}_2 = \sigma_v^2$ is a univariate optimization problem, and we use the `R` function `optimize`. For the E-step of the EM-FH algorithm, the mean score equation reduces to

$$\sum_{i \in A_1} \sum_{j \in A_{2i}} \pi_{ij}^{-1} \big\{ y_{ij} - \beta_0 - \beta_1 x_{ij} - E\big(v_i \mid \hat{v}_i; \hat{\boldsymbol{\beta}}^{(t)}\big) \big\} (1, x_{ij})' = (0, 0),$$

where $E(v_i \mid \hat{v}_i; \hat{\boldsymbol{\beta}}^{(t)}) = (\sigma_v^2 + \hat{V}_{1i}(\hat{\boldsymbol{\beta}}^{(t)}))^{-1} \sigma_v^2 \hat{v}_i(\hat{\boldsymbol{\beta}}^{(t)})$.

2.6. *Special case for exponential dispersion family and canonical link.* The observation that an estimate of σ_e^2 is not required for the case of the Gaussian response distribution motivates us to consider the role of the dispersion parameter for the more general class of generalized linear models. A common choice of distribution is the generalized linear mixed effects model, where f_1 is in the exponential dispersion family and f_2 is a normal distribution with mean 0 and variance ψ_2. Specifically, suppose

$$f_1(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{\beta}, \phi) = \exp\big[\phi\big\{y_{ij}\theta_{ij} - b(\theta_{ij})\big\} + c(y_{ij}, \phi)\big],$$

where $h(\theta_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + v_i$, $h\{b'(\theta_{ij})\} = \theta_{ij}$, and $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ (Jiang, 2003). For this model, $\boldsymbol{\psi}_1 = (\boldsymbol{\beta}', \phi)$, and $\psi_2 = \sigma_v^2$. The mean score equation, evaluated at the tth parameter estimate, is of the form

$$\bar{\boldsymbol{S}}_1(\boldsymbol{\psi}_1 \mid \hat{\boldsymbol{\psi}}^{(t)}) = \sum_{i \in A_1} \bar{\boldsymbol{S}}_{1i}(\boldsymbol{\psi}_1 \mid \hat{\boldsymbol{\psi}}^{(t)}),$$

where

$$\bar{S}_{1i}(\boldsymbol{\psi}_1 \mid \hat{\boldsymbol{\psi}}) = \phi \sum_{j \in A_{2i}} w_{ij}\{y_{ij} - E(b'(\theta_{ij}) \mid \hat{v}_i^{(t)}, \hat{\boldsymbol{\psi}}^{(t)})\}\boldsymbol{x}_{ij}$$

and θ_{ij} is implicitly a function of $\boldsymbol{\psi}_1$. In the data analysis, we use a gamma distribution for f_1 with a noncanonical log link function. For the simulation, we consider the gamma model used for the data analysis as well as the unit-level linear mixed model of Section 2.5. We present the specific forms for the score equation and EM-FH algorithm corresponding to the gamma model in Section 3.4 and in the Appendix.

In applications the assumption that ϕ is constant may not hold and may be difficult to verify. Because the mean score equation is a multiple of ϕ, the solution does not depend on ϕ. For a linear model the ordinary least squares estimator of the regression coefficients is unbiased even if the error variances are not constant. We investigate the robustness of the EM-FH algorithm and corresponding small area prediction procedure to the assumption of a constant ϕ through a simulation study in Section 4.

3. Application to NRI erosion data. As discussed in the Introduction, the objective is to estimate the mean sheet and rill erosion for domains defined by eight-digit hydrologic units in the Boone-Raccoon River Watershed. The response variable y is the sheet and rill erosion as measured by the USLE in (1). Four of the NRI erosion values are zero. We set these four zero values to be 0.005 to permit the use of distributional forms (such as gamma) that do not include zero in the support. The four zero values represent 0.47% of the NRI sample for our study. We expect the effect of 0.47% of the data to be negligible. Further, estimating the parameters of an appropriate model (such as a zero-inflated model) for only four observed zeros is difficult. The covariates are related to the USLE R and K factors. As we discuss in more detail in Section 3.1, the NRI data exhibit nonlinear relationships to covariates as well as nonconstant variances. Additionally, as we discuss in Section 3.2, the weight appears to be related to the response after accounting for the information in the model covariates. The EM-FH algorithm defined in Section 2 is well suited for the NRI application because it allows us to model nonlinear associations in the data and appropriately incorporates the NRI weight.

3.1. *Selection of distributions for NRI data.* In the right panel of Figure 2, $\log(y)$ is plotted against $\log(K)$ with symbols representing the five different values of R in the NRI data set. The smallest K value is 0.02 and the second smallest is 0.1. Note that a K-factor value of 0.02 corresponds to $\log(K) = -3.9$ on the horizontal axis of the right panel of Figure 2. The gap between these two values is large in the log scale, and the pattern in the right plot suggests that the nature of the association between $\log(K)$ and $\log(y)$ may change with $\log(K)$. Our experience is that the K-factor is subject to a modest amount of measurement error, and truncating small values of the K-factor may guard against extrapolation to erroneous K-factor values. We change all K values of 0.02 to 0.1. We expect the effect of this change on the predicted values to be small. As shown in the right panel of Figure 2, relatively few locations have a K-factor of 0.02, and the estimated mean erosion based on lowess for a K-factor value of 0.02 is relatively close to that for a K-factor of 0.1. The left panel of Figure 2 shows the $\log(y)$ plotted against the truncated values of $\log(K)$. The lowess line indicates a possibility

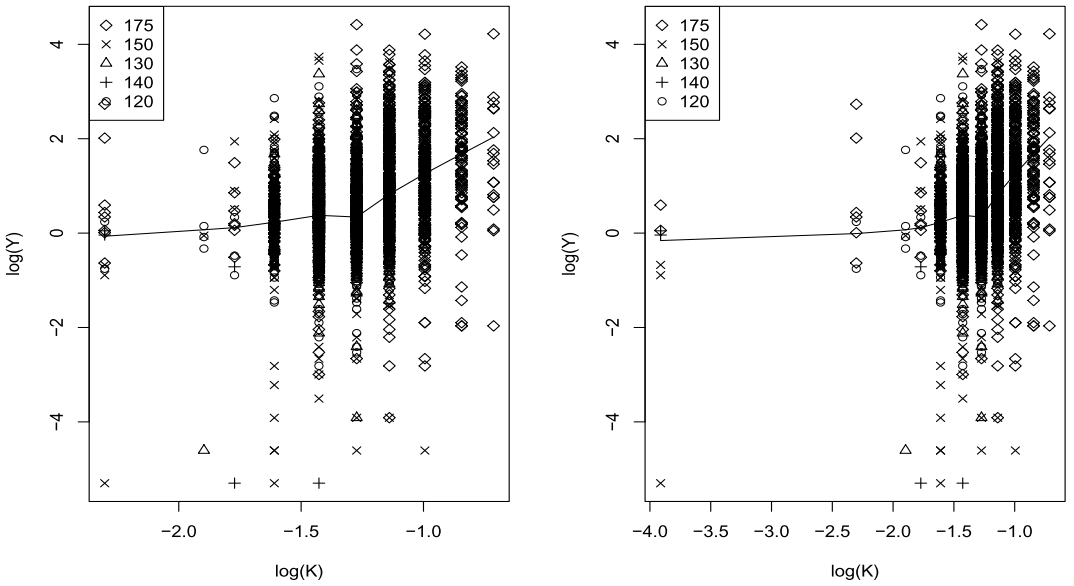

FIG. 2. *Plots of log USLE erosion against the log of the K-factor (right) and log of the truncated K-factor (left) with symbols corresponding to the R-factor and lowess lines.*

of a change in the slope when the K is 0.28. We fit a model where the slope changes at K = 0.28 and found no evidence of a change in the slope. (See the Supplementary Material for details.) We prefer to truncate the K-factor values at 0.1.

To select a distributional form, we conduct a Box–Cox type of analysis. In anticipation of using K as a covariate, we calculate the mean and standard deviation of erosion for groups defined by the 10 distinct values of K (after truncation to 0.1). We then regress the logs of the standard deviations on the logs of the means. The slope of the regression is 1.02 (SE 0.12) which is consistent with the mean variance relationship of a gamma distribution. We consider models with a gamma distribution defining the response distribution.

3.2. *Effect of survey weight.* The premise for our approach is that incorporating the survey weight in estimation guards against a bias that would result if the sample design is informative for the specified model and the weight is ignored. We assess the relationship between the survey weight and erosion, after accounting for the other covariates. To account for the effect of the NRI survey design on variances, we include NRI primary sampling units as random effects. Let j index the NRI point within PSU k and HUC-8 i. We consider a model where

$$
\begin{aligned}
&y_{ikj} \mid \mu_{ikj} \sim \text{gamma}(\mu_{ikj}, \alpha),\\
&\log(\mu_{ikj}) = \beta_{0i} + x_{1ikj}\beta_1 + x_{2ikj}\beta_2 + w_{ikj}\beta_3 + \delta_{ik} + \epsilon_{ikj},
\end{aligned}
\tag{20}
$$

μ_{ikj} and α, respectively, are the mean and shape parameter of the gamma distribution and $\delta_{ik} \overset{\text{iid}}{\sim} N(0, \sigma_\delta^2)$ independent of $\epsilon_{ikj} \overset{\text{iid}}{\sim} N(0, \sigma_\epsilon^2)$. The coefficients β_{0i} for $i = 1, \ldots, 9$ are fixed effects associated with the nine eight-digit hydrologic units (small areas). The covariate x_{1ikj} is the log K-factor, x_{2ikj} is the log R-factor, and w_{ikj} is the final NRI estimation weight for point j in HUC-8 i and PSU k. We fit the model (20) using the `glmer R` function. As discussed in Section 1.1, w_{ikj} is a modification to the original inclusion probability intended to preserve specified controls. As shown in Table 2, the coefficient associated with the weight (β_3) differs significantly from zero, indicating that a relationship between erosion and the weight exists, after accounting for the other model covariates.

TABLE 2
Estimates of the parameters of the model (20) and associated standard errors (SE), T-statistics and p-values

Parameter	Estimate	SE	T	P-value
β_1	0.801	0.089	8.971	$\leq$2e−16***
β_2	1.281	0.481	2.660	0.008**
β_3	−0.008	0.002	−3.918	0.000***
β_{01}	−4.745	2.334	−2.033	0.042*
β_{02}	−4.568	2.411	−1.895	0.058.
β_{03}	−4.558	2.410	−1.891	0.059.
β_{04}	−4.600	2.458	−1.871	0.061.
β_{05}	−4.676	2.422	−1.931	0.054.
β_{06}	−4.593	2.446	−1.878	0.060.
β_{07}	−4.201	2.494	−1.684	0.092.
β_{08}	−4.369	2.492	−1.753	0.080.
β_{09}	−4.129	2.493	−1.656	0.098.

3.3. *Small area model for NRI data.* In the estimation procedure we use the NRI replicate weights to appropriately reflect variability due to the sample design and estimation procedures. We, therefore, specify a model for NRI points in eight-digit hydrologic units. Let y_{ij}, x_{1ij} and x_{2ij} denote the erosion, the log of the K-factor and the log of the R-factor for NRI point j in HUC-8 i. Assume

$$y_{ij} \mid v_i \overset{\text{ind.}}{\sim} \text{gamma}(\mu_{ij}, \alpha),$$

where $E(y_{ij} \mid v_i) = \mu_{ij}$ and α is the shape parameter, assumed constant. The model for μ_{ij} is

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + v_i, \tag{21}$$

where $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$. Under this parametrization the small area parameter to predict is

$$\theta_i = \exp(\gamma_i) \frac{1}{N_i} \sum_{j=1}^{N_i} \exp(\beta_1 x_{1ij} + \beta_2 x_{2ij}),$$

where $\gamma_i = \beta_0 + v_i$.

3.4. *EM-FH algorithm for NRI data.* We apply the estimation procedure of Section 2 to the NRI data. For the NRI model, $\boldsymbol{\psi}_1 = (\beta_0, \beta_1, \beta_2)'$, and $\boldsymbol{\psi}_2 = \sigma_v^2$. For a given value of $\boldsymbol{\psi}_1$, the direct estimator of $\hat{v}_i(\boldsymbol{\psi}_1) = \hat{v}_i$ satisfies $S_{1i}(\hat{v}_i, \boldsymbol{\psi}_1) = 0$, where

$$S_{1i}(v_i, \boldsymbol{\psi}_1) = \sum_{j \in A_{2i}} w_{ij} \{ y_{ij} \exp(-\beta_0 - \beta_1 x_{1ij} - \beta_2 x_{2ij} - v_i) - 1 \}.$$

As discussed in Section 1.1, the NRI supplies 29 sets of replicate weights for variance estimation. The replicates are used to estimate $\hat{V}_{1i}(\boldsymbol{\beta})$, the design variance of $\hat{v}_i$, as explained in (12). The M-step of the EM-FH algorithm finds a root of the mean score equation defined by

$$\bar{\boldsymbol{S}}_1(\boldsymbol{\psi}_1 \mid \sigma_v^2) = \sum_{i \in A_1} \sum_{j \in A_{2i}} w_{ij} \{ y_{ij} \exp(-\tilde{\eta}_{ij}) - 1 \} (1, x_{1ij}, x_{2ij})',$$

where $\tilde{\eta}_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \tilde{\gamma}_i \hat{v}_i + 0.5 \tilde{\gamma}_i \hat{V}_{1i}(\boldsymbol{\beta})$ and $\tilde{\gamma}_i = \sigma_v^2 (\sigma_v^2 + \hat{V}_{1i}(\boldsymbol{\beta}))^{-1}$. This closed-form expression for the mean score equation assumes that $\text{Cov}(y_{ij}, \exp(v_i) \mid \hat{v}_i) \approx 0$ and uses the Moment Generating Function (MGF) of a normal distribution as in Slud and

Maiti (2006). The FH step of the EM algorithm finds a maximum of the log-likelihood defined by

$$\ell_w(\sigma_v^2 \mid \boldsymbol{\psi}_1) = \sum_{i \in A_1} \log\left\{ \frac{1}{\sqrt{(\sigma_v^2 + \hat{V}_{1i}(\boldsymbol{\psi}_1))}} \phi\left(\frac{\hat{v}_i}{\sqrt{\sigma_v^2 + \hat{V}_{1i}(\boldsymbol{\psi}_1)}} \right) \right\}.$$

We use $T = 5$ steps of the EM-FH algorithm and find that the algorithm converges quickly. The forms for the MSE and variance estimators for the gamma response distribution and the log link are presented in Appendix B.2.

3.5. *Parameter estimates for NRI data.* The loglinear form of the USLE equation would suggest that $\beta_1 = \beta_2 = 1$. As discussed, the K and R variables, defining x_{1ij} and x_{2ij}, are not exactly the same as the USLE R and K factors. Additionally, the LS, C and P factors are omitted. Therefore, β_1 and β_2 may differ from each other and may differ from 1. We assess this by fitting alternative models.

We first fit the full model (21), where we estimate both β_1 and β_2. We refer to the full model (21) as Model 1. The estimates of the regression coefficients are $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (-4.734, 1.146, 1.437)$ with corresponding SEs of $(2.310, 0.138, 0.464)$. The estimate of σ_v^2 is $\hat{\sigma}_v^2 = 0.0500$.

We calculate a t-test statistic of the null hypothesis $H_o : \beta_1 = \beta_2$ as

$$T = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\hat{V}(\hat{\beta}_1 - \hat{\beta}_2)}},$$

where $\hat{V}(\hat{\beta}_1 - \hat{\beta}_2) = \hat{V}(\hat{\beta}_1) + \hat{V}(\hat{\beta}_2) - 2\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$ and the estimated covariance matrix of the estimators of regression coefficients is defined in Appendix B.2. The value of the test statistic is $T = 0.634$. We fail to reject the null hypothesis and fit the model under the assumption that $\beta_1 = \beta_2$. The reduced model (Model 2) is defined by replacing the model for μ_{ij} in (21) with the reduced model defined as

$$\log(\mu_{ij}) = \beta_0 + \beta_1(x_{1ij} + x_{2ij}) + v_i. \tag{22}$$

With this restriction, $(\hat{\beta}_0, \hat{\beta}_1) = (-3.345, 1.167)$ with a corresponding estimated standard error of $(0.545, 0.146)$, and the estimate of σ_v^2 increases slightly to $\hat{\sigma}_v^2 = 0.0573$.

Interestingly, the estimate of β_1 from Model 2 does not differ significantly from 1. The coefficient associated with $x_{ij} = x_{1ij} + x_{2ij}$ does not differ significantly from 1, even though the R and K covariates are not identical to the USLE R and K factors and the model omits C and P factors. The restriction that $\beta_1 = \beta_2 = 1$ increases the estimate of σ_v^2, substantially, to $\hat{\sigma}_v^2 = 0.376$. This increase in the estimate of σ_v^2 leads to properties of predictors and estimated mean square errors that seem unreasonable. Therefore, we do not pursue the model with $\beta_1 = \beta_2 = 1$ further. We compare predictions and estimated root mean square errors for Model 1 and Model 2.

3.6. *Small area predictions and estimated RMSEs for mean erosion.* Table 3 contains small area predictors (est), estimated root mean squared errors and estimated coefficients of variation for Models 1 and 2. For comparison, the estimates, standard errors and coefficients of variation for the direct estimators of mean erosion are included as well. The rows labeled est-t give the prediction for model t for $t = 1, 2$. The rows labeled rmse-t give the estimated root mean square errors for model t, and the additional suffix of LT indicates that the estimated root mean square error is the square root of the estimated leading term in the mean square error. The bottom panel of the table compares the estimated means of the covariates

TABLE 3

Top panel: *Estimates*, *standard errors* (*SE*) *and CVs of direct estimators of erosion*. *Second panel*: *Small area predictors based on Model* 1, *standard errors and CVs*. *Third panel*: *Small area predictors based on Model* 2, *standard errors and CVs*. *Bottom panel*: *Population mean* (*kbarpop*) *and direct estimated mean* (*kbarhat*) *of K-factor as well as population mean* (*rbarpop*) *and direct estimated mean* (*rbarhat*) *of R-factor*

	1	2	3	4	5	6	7	8	9
					Direct Estimators				
Est	1.79	2.13	2.51	2.16	1.68	2.23	4.84	5.30	6.71
SE	0.19	0.19	0.38	0.39	0.33	0.31	0.78	0.63	1.76
CV	0.106	0.090	0.149	0.178	0.199	0.137	0.160	0.119	0.262
					Model 1 Predictions & RMSE				
est-1	3.554	2.922	2.747	3.437	2.373	2.359	3.462	3.859	5.190
rmse-1-LT	0.181	0.380	0.398	0.280	0.291	0.088	0.097	0.401	0.230
cv-1-LT	0.051	0.130	0.145	0.081	0.123	0.037	0.028	0.104	0.044
rmse-1	0.540	0.418	0.410	0.360	0.312	0.095	0.185	0.411	0.250
cv-1	0.152	0.143	0.149	0.105	0.131	0.040	0.053	0.106	0.048
					Model 2 Predictions & RMSE				
est-2	3.326	2.853	2.747	3.377	2.322	2.357	3.536	3.916	5.235
rmse-2-LT	0.169	0.380	0.409	0.269	0.290	0.086	0.100	0.412	0.231
cv-2-LT	0.051	0.133	0.149	0.080	0.125	0.036	0.028	0.105	0.044
rmse-2	0.336	0.401	0.419	0.340	0.304	0.092	0.145	0.420	0.246
cv-2	0.101	0.140	0.153	0.101	0.131	0.039	0.041	0.107	0.047
					NRI Estimates and Population Means of Covariates				
kbarhat	0.263	0.251	0.243	0.250	0.245	0.259	0.297	0.340	0.365
kbarpop	0.334	0.282	0.264	0.337	0.273	0.272	0.254	0.287	0.304
rbarhat	4.834	5.002	4.997	5.075	5.011	5.061	5.165	5.165	5.165
rbarpop	5.043	5.073	5.016	5.137	5.060	5.047	5.064	5.077	5.112

from the survey to the corresponding population means. The rows labeled kbarhat and rbarhat contain the NRI estimates of the means of $\log(K)$ and $\log(R)$ for the nine domains. The rows labeled kbarpop and rbarpop are the corresponding population means.

The small area procedure shrinks the small area estimates of mean erosion toward the overall mean. The procedure shrinks the standard errors similarly. For either Model 1 or Model 2, the estimated RMSE of the predictor is typically smaller (larger) than the estimated standard error of the direct estimator if the model based predictor of mean erosion is smaller (larger) than the direct estimator. Because the standard errors are related to the means, we focus on the estimated coefficients of variation. The estimated coefficients of variation from Model 2 are typically close to the estimated coefficients of variation from Model 1. For area 1 the assumption of a common slope for the R and K factors leads to an important decrease in the estimated coefficient of variation. The assumption of a common slope is expected to reduce the effect of the variances of parameter estimators on the prediction MSE. The effect of parameter estimation on the estimated MSE varies across the areas. For area 1 the effect of parameter estimation from Model 1 accounts for approximately 90% of the MSE. This area has a relatively large sample size, and $(\hat{\sigma}_v^2 + \hat{V}_{1i})^{-1}\hat{\sigma}_v^2 = 0.93$. For Model 2 the variance due to parameter estimation accounts for 75% of the total MSE of the predictor for area 1. Model 2 is a compromise between the full Model 1 and the simplest model that forces the slopes for both covariates to equal 1. Therefore, we prefer Model 2. The estimated coefficients of variation for Model 2 are smaller than those for the estimated direct estimators for all areas, except for areas 2 and 3. The estimated CV of the predictor for area 3 is close to that of the direct estimator. The increase in the estimated CV for area 2 is substantial. To better understand effect of small area modeling on the estimated standard error, we calculate a

version of a direct standard error in a different way. Define

$$\hat{V}_{\mathrm{dir},i,2} = \left[\frac{1}{N_i}\sum_{j=1}^{N_i}\exp\{\hat{\beta}_0 + \hat{\beta}_1(x_{1ij} + x_{2ij})\}\right]^2 \times \left[\exp\{2\hat{v}_i(\hat{\boldsymbol{\beta}}) + 2\hat{V}_{1i}(\hat{\boldsymbol{\beta}})\} - \exp\{2\hat{v}_i(\hat{\boldsymbol{\beta}}) + \hat{V}_{1i}(\hat{\boldsymbol{\beta}})\}\right].$$

The $\hat{V}_{\mathrm{dir},i,2}$ is the conditional variance of $[N_i^{-1}\sum_{j=1}^{N_i}\exp\{\hat{\beta}_0 + \hat{\beta}_1(x_{1ij} + x_{2ij})\}]\exp\{\hat{v}_i(\hat{\beta}_1)\}$ given v_i. The $\hat{V}_{\mathrm{dir},i,2}$ for the nine areas in order by index label are

$$(0.172, 0.440, 0.533, 0.278, 0.312, 0.086, 0.101, 0.483, 0.238).$$

For all areas except for areas 2 and 3, the $\hat{V}_{\mathrm{dir},i,2}$ is smaller than or equal to the direct estimator of the standard error of the mean erosion for the area. For areas 2 and 3 the $\hat{V}_{\mathrm{dir},i,2}$ are considerably larger than the direct standard error for mean erosion. Smoothing the $\hat{V}_{1i}(\hat{\boldsymbol{\beta}})$ may improve the standard errors for these two areas; however, the number of PSUs per area is relatively large, so we expect that $\hat{V}_{1i}(\hat{\boldsymbol{\beta}})$ is reliable enough.

A comparison of the means of the covariates for the survey to the corresponding population means illustrates why small area prediction increases the estimates for some areas and decreases the estimates for others. For areas 1–5 the population means of the covariates exceed the corresponding survey based estimators, and small area prediction leads to an increase in the estimated mean erosion relative to the direct estimator. For area 6, the population mean for the K-factor is larger than the direct estimator of the K-factor, while the opposite is true for the R-factor. The small area predictor for area 6 is relatively close to the direct estimator. Area 6 has a relatively large sample size, and the direct estimate of erosion for area 6 is close to the overall mean erosion in the Boone-Raccoon river watershed. For areas 8 and 9 the population means of the covariates are smaller than the survey based means, and small area predictors are smaller than the direct estimators.

3.7. *Model assessment.* The model (2) is a model for the super-population, so the plausibility of model (2) is difficult to assess directly using standard diagnostic tools. Because our estimation procedure is based on estimating equations, the difficulty of diagnosing the goodness of fit of the entire super-population model does not prevent us from defining a model assessment procedure. One of the critical assumptions that underlies the estimation procedure is the Fay–Herriot type of model given in (8) and (3). To evaluate if the assumption of the Fay–Herriot type of model in (8) and (3) is plausible for the NRI data; we define a residual by

$$r_i = \frac{\hat{v}_i(\hat{\boldsymbol{\beta}})}{\sqrt{\hat{\sigma}_v^2 + \hat{V}_{1i}(\hat{\boldsymbol{\beta}})}}.$$

If the model assumption holds, then r_i has a normal distribution. Figure 3 shows a normal probability plot of r_i for $i = 1, \ldots, 9$ based on Model 2. The p-value from a Shapiro–Wilk test of the null hypothsis that r_i has a normal distribution is 0.23, providing no evidence against normality.

4. Simulations. We consider Gaussian and gamma distributions for the population model. For the Gaussian distribution

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}, \tag{23}$$

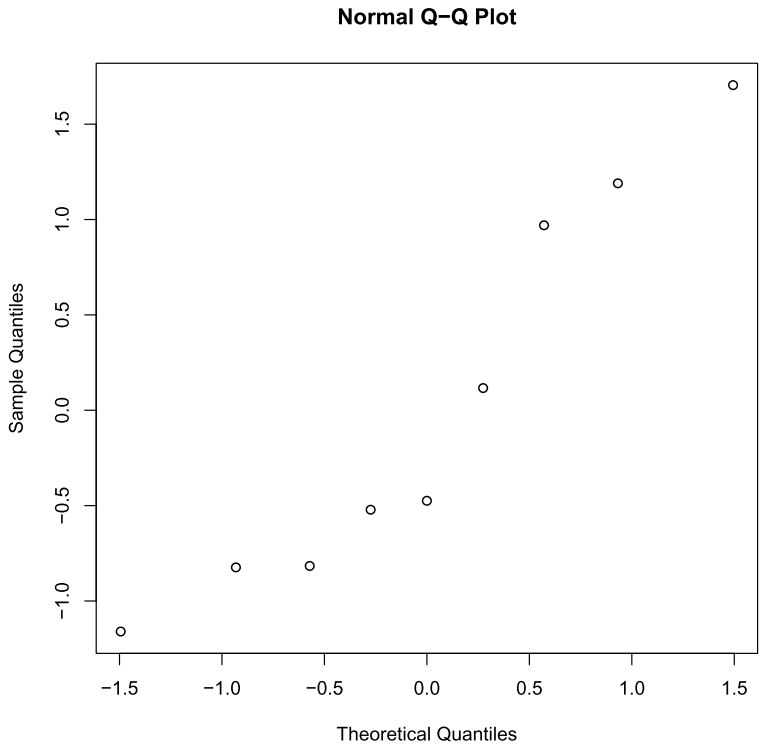

FIG. 3. *Normal probability plot of r_i for $i = 1, \ldots, 9$.*

where $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ and $e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$. For the gamma distribution

$$y_{ij} \mid \mu_{ij} \sim \text{gamma}(\mu_{ij}, \alpha), \tag{24}$$

where μ_{ij} and α denote the mean and shape parameter, respectively, of a gamma distribution. We define $\mu_{ij} = \exp(\beta_0 + \beta_1 x_{ij} + v_i)$, where $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$. For both distributions we set $\beta_0 = -4$, $\beta_1 = 2$, $\sigma_v^2 = 0.5$, and $\sigma_e^2 = 1$. We generate $x_{ij} \overset{\text{iid}}{\sim} \text{Unif}(0, 1)$.

We generate a population of 10,000 units spread evenly across 50 areas. The sample design is Poisson sampling with inclusion probability for element j in area i defined as

$$\pi_{ij} = \frac{0.02 N_i \exp\{0.2(z_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta}) + 0.2\delta_{ij}\}}{\sum_{j=1}^{N_i} \exp\{0.2(z_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta}) + 0.2\delta_{ij}\}},$$

where $z_{ij} = y_{ij}$ for the Gaussian distribution and $z_{ij} = \log(y_{ij})$ for the gamma distribution. The δ_{ij} are $i.i.d.$ standard normal random variables, truncated to the interval $[-2, 2]$. When we implement the EM-FH algorithm, we use $T = 2$ steps of the iteration.

We also examine the robustness to two assumptions. First, to assess sensitivity to the assumption of a constant σ_e^2 and α, we consider a simulation configuration where the dispersion parameter varies across the areas. For the Gaussian model we set $\sigma_e^2 = 0.5$ for areas 1–25 and set $\sigma_e^2 = 1.5$ for areas 26–50. For the gamma model we set $\alpha = 1$ for areas 1–25 and set $\alpha = 5$ for areas 26–50. Second, we assess sensitivity to the assumption of normally distributed random effects by generating the area random effects as multiples of a t distribution with five degrees of freedom. Specifically, we generate $v_i \overset{\text{iid}}{\sim} \sqrt{3/5}\sigma_v t_{(5)}$. The multiplication by $\sqrt{3/5}\sigma_v$ ensures that the generated random effects have variance of σ_v^2.

Table 4 allows a comparison of the average Monte Carlo (MC) MSEs of the predictors of the small area means based on the EM-FH algorithm of Section 2 to the average MSEs

TABLE 4
Average MC MSE of direct estimator and of predictor as well as MC mean of estimated MSE of predictor

	Common α, $v_i \sim N$		Nonconstant α/σ_e^2, $v_i \sim N$		Common α, $v_i \sim t_5$	
	Gamma	Gaussian	Gamma	Gaussian	Gamma	Gaussian
MC MSE of Direct.	0.000376	0.0374	0.00035	0.0384	0.000438	0.0372
MC MSE of Pred.	0.000219	0.0266	0.00021	0.0272	0.000323	0.0270
MC Mean of Est. MSE	0.000205	0.0266	0.00020	0.0272	0.000270	0.0271

of direct Hajek estimators. The EM-FH algorithm for the Gaussian distribution is defined in Section 2.5, and the EM-FH algorithm for the gamma distribution is defined in Appendix A. We define the MSE in the standard way as $\text{MSE}(\hat{\theta}_i) = R^{-1} \sum_{r=1}^{R} (\hat{\theta}_i^{(r)} - \bar{y}_{N_i}^{(r)})^2$, where $\hat{\theta}_i^{(r)}$ is either a predictor or a direct estimator obtained in MC replication r and $\bar{y}_{N_i}^{(r)}$ is the population mean of y_{ij} for area i simulated in MC replication r. We present the results for the simulation with nonconstant α or σ_e^2 below the heading "Nonconstant α/σ_e^2" in Table 4. The final row of Table 4 contains the MC means of the linearization based estimators of the MSEs of the predictors, defined in Section 2.3 with details in Appendix B.2. For the left panel of Table 4, the assumed model holds, and the results are consistent with the theory. For both the gamma and Gaussian models the predictors are more efficient than the direct estimators, and the MSE estimator is a good approximation for the MSE of the predictor. The prediction procedure continues to attain efficiency gains relative to the direct estimator when the assumption of a constant σ_e^2 or α is violated. The estimated MSEs of the predictors are reasonable, even if the assumption of a constant dispersion parameter does not hold. The $t_{(5)}$ distribution has heavier tails than the assumed normal distribution for v_i. For the gamma distribution the use of a $t_{(5)}$ distribution for v_i inflates the MSEs of the predictors slightly. For the Gaussian distribution the results for the case where the true distribution of v_i is $t_{(5)}$ are nearly the same as the case where the assumed normal distribution holds. Overall, the results for the $t_{(5)}$ distribution for v_i indicate that the procedure is resistant to outliers in the distribution of the area random effects.

5. Conclusion. We develop a small area procedure based on survey weighted estimating equations arising from a specified super-population model. The method allows us to model nonlinear relationships and incorporate the effect of an informative sampling design without making model assumptions on the sampling weights. The basic idea is to obtain the approximate predictive distribution using the asymptotic normality of the profile pseudo maximum likelihood estimator of the area effect. The profile PMLE is used to construct an area-level model from the original unit level model. The approximate predictive distribution is also used to construct the E-step of the EM algorithm for parameter estimation. The proposed method is applied to estimation of mean erosion in watersheds using NRI data. The estimated CVs of the predictors are less than or equal to those of the direct estimators for all but two areas.

One of the advantages of our approach is robustness. We do not require a model for the sampling weights. The simulation shows that the procedure is somewhat robust to departures from the assumption of a constant dispersion parameter.

The robustness of our procedure may come at the expense of a loss of efficiency if the area sample sizes are too small. As the area sample sizes decrease, adding stronger model assumptions may be necessary. For the NRI application the area sample sizes are moderately large. The procedure appropriately reflects the complex survey process and the relationships in the data.

APPENDIX A: EM-FH PROCEDURE FOR THE GAMMA MODEL AND THE ELEMENT SAMPLE DESIGN

We define the EM-FH algorithm for the gamma distribution used in the simulation. For this model, $\boldsymbol{\psi}_1 = (\beta_0, \beta_1)$, and $\psi_2 = \sigma_v^2$. The direct estimator of v_i is defined such that $S_{1i}(\hat{v}_i, \boldsymbol{\psi}_1) = 0$, where

$$S_{1i}(v_i, \boldsymbol{\psi}_1) = \sum_{j \in A_{2i}} \pi_{ij}^{-1} \{ y_{ij} \exp(-\beta_0 - \beta_1 x_{ij} - v_i) - 1 \}.$$

The direct estimator of the variance for area i in the simulation is defined as

$$\hat{V}_{1i}(\boldsymbol{\psi}_1) = \left\{ \sum_{j \in A_{2i}} \pi_{ij}^{-1} y_{ij} \exp(-\beta_0 - \beta_1 x_{ij} - v_i) \right\}^{-2} \hat{V}_{\pi i}, \tag{25}$$

where $\hat{V}_{\pi i} = \sum_{j \in A_{2i}} (\pi_{ij,ik} - \pi_{ij}\pi_{ik}) \pi_{ik}^{-1} \pi_{ij}^{-1} \{ y_{ij} \exp(-\beta_0 - \beta_1 x_{ij} - \hat{v}_i) - 1 \}^2$. The M-step finds a root of the mean score equation defined by

$$\bar{S}_1(\boldsymbol{\psi}_1 \mid \sigma_v^2) = \sum_{i=1}^{D} \sum_{j \in A_{2i}} \pi_{ij}^{-1} \{ y_{ij} \exp(-\tilde{\eta}_{ij}) - 1 \} (1, x_{ij})',$$

where $\tilde{\eta}_{ij} = \beta_0 + \beta_1 x_{ij} + \tilde{\gamma}_i \hat{v}_i + 0.5 \tilde{\gamma}_i \hat{V}_{1i}(\boldsymbol{\beta})$, and $\tilde{\gamma}_i = \sigma_v^2 (\sigma_v^2 + \hat{V}_{1i}(\boldsymbol{\beta}))^{-1}$.

APPENDIX B: LINEARIZATION VARIANCE ESTIMATION FOR THE ELEMENT SAMPLE DESIGNS

We define variance estimators appropriate for the NRI study design. In the NRI application, all small areas have at least one sampled element. Therefore, we consider a situation in which $A_1 = \{1, \ldots, D\}$:

$$\bar{\boldsymbol{S}}_1(\boldsymbol{\psi}_1 \mid \psi_2) = \frac{\int \boldsymbol{S}_1(\boldsymbol{\psi}_1 \mid v_i) g_2(\hat{v}_i \mid v_i) f_2(v_i \mid \boldsymbol{\psi}_2) \, dv_i}{\int g_2(\hat{v}_i \mid v_i) f_2(v_i \mid \boldsymbol{\psi}_2) \, dv_i} = \sum_{i=1}^{D} \bar{\boldsymbol{S}}_{1i},$$

where

$$\bar{\boldsymbol{S}}_{1i} = \frac{\int \hat{\boldsymbol{S}}_{1i}(\boldsymbol{\psi}_1 \mid v_i) g_2(\hat{v}_i \mid v_i) f_2(v_i \mid \boldsymbol{\psi}_2) \, dv_i}{\int g_2(\hat{v}_i \mid v_i) f_2(v_i \mid \boldsymbol{\psi}_2) \, dv_i}$$

and

$$\hat{\boldsymbol{S}}_{1i} = \sum_{j \in A_{2i}} \pi_{ij}^{-1} \frac{\partial}{\partial \boldsymbol{\psi}_1} \log\{ f(y_{ij} \mid \boldsymbol{x}_{ij}, v_i, \boldsymbol{\psi}_1) \}.$$

An estimator of the variance of $\boldsymbol{\psi}_1$ is

$$\hat{V}\{\hat{\boldsymbol{\psi}}_1\} = \bar{\boldsymbol{D}}^{-1} \hat{\boldsymbol{V}}_{ss11} (\bar{\boldsymbol{D}}^{-1})',$$

where $\bar{\boldsymbol{D}} = \sum_{i=1}^{D} \partial \bar{\boldsymbol{S}}_{1i} / (\partial \boldsymbol{\psi}_1)$ and $\hat{\boldsymbol{V}}_{ss11} = \sum_{i=1}^{D} \bar{\boldsymbol{S}}_{1i} \bar{\boldsymbol{S}}_{1i}'$. We derive the specific forms of $\bar{\boldsymbol{D}}$ for the gamma and Gaussian distributions in the following two subsections.

B.1. Gaussian model. The mean score equation is given by

$$\bar{\boldsymbol{S}}_1(\beta_0, \boldsymbol{\beta}_1 \mid \hat{v}_i) = \sum_{i=1}^{D} \bar{\boldsymbol{S}}_{1i},$$

where

$$\bar{\boldsymbol{S}}_{1i} = \sum_{j\in A_{2i}} w_{ij}\{y_{ij} - \beta_0 - \boldsymbol{\beta}_1'\boldsymbol{x}_{ij} - \gamma_i(\bar{y}_{wi} - \beta_0 - \bar{\boldsymbol{x}}_{wi}'\boldsymbol{\beta}_1)\}(1, \boldsymbol{x}_{ij}')',$$

$(\bar{\boldsymbol{x}}_{wi}', \bar{y}_{wi})' = w_{i.}^{-1}\sum_{j\in A_{2i}} w_{ij}(\boldsymbol{x}_{ij}', y_{ij})'$, $w_{i.} = \sum_{j\in A_{2i}} w_{ij}$ and $\gamma_i = \sigma_v^2(\sigma_v^2 + \hat{V}_{1i}(\boldsymbol{\beta}))^{-1}$ with $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$. Then, the matrix $\bar{\boldsymbol{D}}$ is given as

$$\bar{\boldsymbol{D}} = \begin{pmatrix} D_{11} & \boldsymbol{D}_{12}' \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{pmatrix},$$

where $D_{11} = \sum_{i=1}^{D} w_{i.}(\gamma_i - 1)$, $\boldsymbol{D}_{12} = \sum_{i=1}^{D} w_{i.}\bar{\boldsymbol{x}}_{wi}\gamma_i - \sum_{i=1}^{D}\sum_{j\in A_{2i}} w_{ij}\boldsymbol{x}_{ij} = \boldsymbol{D}_{21}$ and $\boldsymbol{D}_{22} = \sum_{i=1}^{D}\gamma_i w_{i.}\bar{\boldsymbol{x}}_{wi}\bar{\boldsymbol{x}}_{wi}' - \sum_{i=1}^{D}\sum_{j\in A_{2i}} w_{ij}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}'$. The estimated MSE of the predictor is then

$$\hat{\mathrm{MSE}}\{\hat{\theta}_i\} = \gamma_i\hat{V}_{1i}(\hat{\boldsymbol{\beta}}) + (1-\gamma_i, 1-\gamma_i\bar{\boldsymbol{x}}_{wi}')\hat{V}(\hat{\boldsymbol{\beta}})(1-\gamma_i, 1-\gamma_i\bar{\boldsymbol{x}}_{wi}')',$$

where

$$\hat{V}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{D}}^{-1}\left\{\sum_{i=1}^{D}\bar{\boldsymbol{S}}_{1i}\bar{\boldsymbol{S}}_{1i}'\right\}(\hat{\boldsymbol{D}}^{-1})'$$

and $\hat{\boldsymbol{D}}$ is obtained by replacing the unknown parameters defining $\bar{\boldsymbol{D}}$ with estimators.

B.2. Gamma model. The mean score equation for $\boldsymbol{\beta} = \boldsymbol{\psi}_1 = (\beta_0, \boldsymbol{\beta}_1')'$ is given by

$$\bar{\boldsymbol{S}}_1(\boldsymbol{\beta} \mid \hat{v}_i) = \sum_{i=1}^{D}\bar{\boldsymbol{S}}_{1i},$$

where

$$\bar{\boldsymbol{S}}_{1i} = \sum_{j\in A_{2i}} w_{ij}\{y_{ij}\exp(-\beta_0 - \boldsymbol{\beta}_1'\boldsymbol{x}_{ij} - \gamma_i\hat{v}_i - 0.5\gamma_i\hat{V}_{1i}) - 1\}(1, \boldsymbol{x}_{ij}')',$$

$\gamma_i = \sigma_v^2(\sigma_v^2 + \hat{V}_{1i}(\boldsymbol{\beta}))^{-1}$, $\hat{v}_i$ satisfies $S_{1i}(\hat{v}_i, \boldsymbol{\beta}) = 0$,

$$S_{1i}(v_i, \boldsymbol{\beta}) = \sum_{j\in A_{2i}} w_{ij}\{y_{ij}\exp(-\beta_0 - \boldsymbol{\beta}_1'\boldsymbol{x}_{ij} - v_i) - 1\}$$

and $\hat{V}_{1i}(\boldsymbol{\beta})$ is defined in (25). To calculate $\bar{\boldsymbol{D}}$, we require the derivative of $\hat{v}_i$ with respect to $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$. Using implicit differentiation of $\hat{\boldsymbol{S}}_{1i}(\hat{v}_i, \boldsymbol{\beta})$, we obtain

$$0 = \sum_{j\in A_{2i}} w_{ij}y_{ij}\exp(-\beta_0 - \boldsymbol{\beta}_1'\boldsymbol{x}_{ij} - \hat{v}_i)\left(-1 - \frac{\partial\hat{v}_i}{\partial\beta_0}\right)$$

and

$$0 = \sum_{j\in A_{2i}} w_{ij}y_{ij}\exp(-\beta_0 - \boldsymbol{\beta}_1'\boldsymbol{x}_{ij} - \hat{v}_i)\left(-x_{ijk} - \frac{\partial\hat{v}_i}{\partial\beta_{1k}}\right),$$

where x_{ijk} and β_{1k} are the kth components of $\boldsymbol{x}_{ij}$ and $\boldsymbol{\beta}_1$, respectively. Therefore,

$$\frac{\partial\hat{v}_i}{\partial\beta_0} = -1,$$

and

$$\frac{\partial\hat{v}_i}{\partial\beta_{1k}} = -\frac{\sum_{j\in A_{2i}} w_{ij}y_{ij}\exp(-\beta_0 - \beta_1 x_{ij} - \hat{v}_i)x_{ijk}}{\sum_{j\in A_{2i}} w_{ij}y_{ij}\exp(-\beta_0 - \beta_1 x_{ij} - \hat{v}_i)}.$$

We can now calculate $\bar{\boldsymbol{D}}$ as

$$\bar{\boldsymbol{D}} = \begin{pmatrix} D_{11} & \boldsymbol{D}_{12}' \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{pmatrix},$$

where

$$D_{11} = \sum_{i=1}^{D} \sum_{j \in A_{2i}} w_{ij} y_{ij} \exp(-\eta_{ij})(\gamma_i - 1),$$

$$\boldsymbol{D}_{12} = \sum_{i=1}^{D} \sum_{j \in A_{2i}} w_{ij} y_{ij} \exp(-\eta_{ij}) \left(-\boldsymbol{x}_{ij} - \gamma_i \frac{\partial \hat{v}_i}{\partial \boldsymbol{\beta}_1} \right),$$

$$\boldsymbol{D}_{21} = \sum_{i=1}^{D} \sum_{j \in A_{2i}} w_{ij} y_{ij} \boldsymbol{x}_{ij} \exp(-\eta_{ij})(\gamma_i - 1),$$

$$\boldsymbol{D}_{22} = \sum_{i=1}^{D} \sum_{j \in A_{2i}} w_{ij} y_{ij} \boldsymbol{x}_{ij} \exp(-\eta_{ij}) \left\{ -\boldsymbol{x}_{ij}' - \left(\gamma_i \frac{\partial \hat{v}_i}{\partial \boldsymbol{\beta}_1} \right)' \right\}$$

and $\eta_{ij} = \beta_0 + \boldsymbol{\beta}_1' \boldsymbol{x}_{ij} + \gamma_i \hat{v}_i + \gamma_i \hat{V}_{1i}(\boldsymbol{\beta})/2$. The estimated MSE of the predictor for the gamma model is then

$$\hat{\mathrm{MSE}}\{\hat{\theta}_i\} = \hat{M}_{1i} + \hat{M}_{2i},$$

where

$$\hat{M}_{1i} = \left\{ \frac{1}{N_i} \sum_{j=1}^{N_i} \exp(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1' \boldsymbol{x}_{ij}) \right\}^2 \exp(2\gamma_i \hat{v}_i) [\exp(\hat{\gamma}_i \hat{V}_{1i}(\hat{\boldsymbol{\beta}})) \{\exp(\hat{\gamma}_i \hat{V}_{1i}(\hat{\boldsymbol{\beta}})) - 1\}],$$

$$\hat{M}_{2i} = (\hat{d}_{\mathrm{pred},i,1}, \hat{\boldsymbol{d}}_{\mathrm{pred},i,2}') \hat{V}(\hat{\boldsymbol{\beta}}) (\hat{d}_{\mathrm{pred},i,1}, \hat{\boldsymbol{d}}_{\mathrm{pred},i,2}')',$$

$$\hat{d}_{\mathrm{pred},i,1} = N_i^{-1} \sum_{j=1}^{N_i} \exp(\hat{\eta}_{ij})(1 - \hat{\gamma}_i),$$

$$\hat{\boldsymbol{d}}_{\mathrm{pred},i,2} = N_i^{-1} \sum_{j=1}^{N_i} \exp(\hat{\eta}_{ij}) \left(\boldsymbol{x}_{ij} + \hat{\gamma}_i \frac{\partial \hat{v}_i}{\partial \hat{\boldsymbol{\beta}}_1} \right),$$

$\partial \hat{v}_i / (\partial \hat{\boldsymbol{\beta}}_1)$ denotes the derivative of $\hat{v}_i$ with respect to $\boldsymbol{\beta}_1$ evaluated at the estimated parameter and $\hat{V}(\hat{\boldsymbol{\beta}})$ is defined an a manner analogous to the linear model.

APPENDIX C: VALIDATION OF LINEARIZATION VARIANCE ESTIMATION

The prediction MSE is dominated by the leading term. Therefore, an analysis of the estimated MSE of the predictor can mask problems with the estimated variances of the estimators of the regression coefficients. In this section we assess the estimators of the variances of the estimators of the regression coefficients.

Table 5 contains the MC variances of the estimators of the regression coefficients along with the MC means of the variance estimators for the six simulation configurations. The MC means and MC variances are mutiplied by 100. For cases where σ_e^2 or α is constant across the areas, we observe overestimation of the variance for β_0 and underestimation for β_1. For both the Gaussian and gamma distributions the procedure is robust to the assumption of a constant dispersion parameter. The variance estimators remain reasonable if the true distribution of the random effects is a t-distribution.

TABLE 5

MC means of estimators, MC variances of estimators of (β_0, β_1) *and MC means of variance estimators for six simulation configurations. MC variances are multiplied by* 100. *True values are* $(\beta_0, \beta_1) = (-4, 2)$

Distribution	Statistic	β_0	β_1
Gaussian	MC Var of Est.	1.138	0.664
Gaussian	MC Mean of Var. Est.	1.369	0.574
Gaussian	MC Mean of Est.	−4.021	2.005
Gaussian $\neq \sigma_e^2$	MC Var of Est.	1.233	0.699
Gaussian $\neq \sigma_e^2$	MC Mean of Var. Est.	1.648	0.671
Gaussian $\neq \sigma_e^2$	MC Mean of Est.	−4.001	2.002
Gamma	MC Var of Est.	1.028	0.684
Gamma	MC Mean of Var. Est.	1.358	0.568
Gamma	MC Mean of Est.	−4.009	2.011
Gamma $\neq \alpha$	MC Var of Est.	1.180	0.287
Gamma $\neq \alpha$	MC Mean of Var. Est.	1.770	0.344
Gamma $\neq \alpha$	MC Mean of Est.	−4.013	1.986
Gaussian $v_i \sim t_5$	MC Var of Est.	1.129	0.850
Gaussian $v_i \sim t_5$	MC Mean of Var. Est.	1.273	0.674
Gaussian $v_i \sim t_5$	MC Mean of Est.	−3.998	1.993
Gamma $v_i \sim t_5$	MC Var of Est.	1.165	0.550
Gamma $v_i \sim t_5$	MC Mean of Var. Est.	1.322	0.580
Gamma $v_i \sim t_5$	MC Mean of Est.	−4.021	2.005

Acknowledgments. The authors are grateful for the constructive comments from the reviewers and the Associate Editor. The research was partially supported by NSF Grant MMS-1733572 and a cooperative agreement between Iowa State University and USDA-NRCS.

SUPPLEMENTARY MATERIAL

Supplementary material for "An approximate best prediction approach to small area estimation for sheet and rill erosion under informative sampling" (DOI: 10.1214/20-AOAS1388SUPP; .pdf). Supplementary material provides additional simulations and details about the data analysis.

REFERENCES

ASPAROUHOV, T. (2006). General multi-level modeling with sampling weights. *Comm. Statist. Theory Methods* **35** 439–460. MR2274063 https://doi.org/10.1080/03610920500476598

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

BERG, E. and CHANDRA, H. (2014). Small area prediction for a unit-level lognormal model. *Comput. Statist. Data Anal.* **78** 159–175. MR3212164 https://doi.org/10.1016/j.csda.2014.03.007

BERG, E., CHANDRA, H. and CHAMBERS, R. (2016). Small area estimation for lognormal data. In *Analysis of Poverty Data by Small Area Estimation* (M. Pratesi, ed.) 279–298.

BERG, E. and KIM, J.-K (2021). Supplement to "An approximate best prediction approach to small area estimation for sheet and rill erosion under informative sampling." https://doi.org/10.1214/20-AOAS1388SUPP

BERG, E. and LEE, D. (2019). Small area prediction of quantiles for zero-inflated data and an informative sample design. *Stat. Theory Relat. Fields* **3** 114–128. MR4028310 https://doi.org/10.1080/24754269.2019.1666243

BREIDT, F. J. and FULLER, W. A. (1999). Design of supplemented panel surveys with application to the National Resources Inventory. *J. Agric. Biol. Environ. Stat.* **4** 391–403. MR1816375 https://doi.org/10.2307/1400497

DEVILLE, J. C., SÄRNDAL, C. E. and SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *J. Amer. Statist. Assoc.* **88** 1013–1020.

DREASSI, E., PETRUCCI, A. and ROCCO, E. (2014). Small area estimation for semicontinuous skewed spatial data: An application to the grape wine production in Tuscany. *Biom. J.* **56** 141–156. MR3152707 https://doi.org/10.1002/bimj.201200271

ELASHOFF, M. and RYAN, L. (2004). An EM algorithm for estimating equations. *J. Comput. Graph. Statist.* **13** 48–65. MR2044870 https://doi.org/10.1198/1061860043092

ERCIULESCU, A. L. and FULLER, W. A. (2019). Bootstrap prediction intervals for small area means from unit-level nonlinear models. *J. Surv. Statist. Methodol.* **7** 309–333.

FAY, R. E. III and HERRIOT, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277. MR0548019

FULLER, W. A. (2011). *Sampling Statistics* (*Vol.* 560). Wiley, New York.

GHOSH, M. and MAITI, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika* **91** 95–112. MR2050462 https://doi.org/10.1093/biomet/91.1.95

JIANG, J. (2003). Empirical best prediction for small-area inference based on generalized linear mixed models. *J. Statist. Plann. Inference* **111** 117–127. MR1955876 https://doi.org/10.1016/S0378-3758(02)00293-8

JIANG, J. and LAHIRI, P. (2006). Mixed model prediction and small area estimation. *TEST* **15** 1–96. MR2252522 https://doi.org/10.1007/BF02595419

KIM, J. K., PARK, S. and LEE, Y. (2017). Statistical inference using generalized linear mixed models under informative cluster sampling. *Canad. J. Statist.* **45** 479–497. MR3729982 https://doi.org/10.1002/cjs.11339

LOHR, S. L. (2009). *Sampling*: *Design and Analysis*. Nelson Education.

NANDRAM, B. and SEDRANSK, J. (2002). Bayesian predictive inference for the proportion of eroded land in a small area in Iowa. *Environ. Ecol. Stat.* **9** 221–236. MR1917877 https://doi.org/10.1023/A:1016236809152

NEUHAUS, J. M. and MCCULLOCH, C. E. (2011). Estimation of covariate effects in generalized linear mixed models with informative cluster sizes. *Biometrika* **98** 147–162. MR2804216 https://doi.org/10.1093/biomet/asq066

NUSSER, S. M. and GOEBEL, J. J. (1997). The national resources inventory: A long-term multi-resource monitoring programme. *Environ. Ecol. Stat.* **4** 181–204.

U.S. DEPARTMENT OF AGRICULTURE (2018). *Summary Report*: 2015 *National Resources Inventory*, *Natural Resources Conservation Service*, *Washington*, *DC*, *and Center for Survey Statistics and Methodology*. Iowa State Univ., Ames, IA. Available at http://www.nrcs.usda.gov/technical/nri/15summary.

PFEFFERMANN, D. (2013). New important developments in small area estimation. *Statist. Sci.* **28** 40–68. MR3075338 https://doi.org/10.1214/12-STS395

PFEFFERMANN, D. and SVERCHKOV, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *J. Amer. Statist. Assoc.* **102** 1427–1439. MR2412558 https://doi.org/10.1198/016214507000001094

PRASAD, N. G. N. and RAO, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.* **85** 163–171. MR1137362

RAO, J. N. K. (2015). Inferential issues in model-based small area estimation: Some new developments. *Statist. Transit. New Ser.* **16** 491–510.

RAO, J. N. K. and MOLINA, I. (2015). *Small Area Estimation*, 2nd ed. *Wiley Series in Survey Methodology*. Wiley, Hoboken, NJ. MR3380626 https://doi.org/10.1002/9781118735855

SÄRNDAL, C. E., SWENSSON, B. and WRETMAN, J. (2003). *Model Assisted Survey Sampling*. Springer, Berlin.

SLUD, E. V. and MAITI, T. (2006). Mean-squared error estimation in transformed Fay–Herriot models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 239–257. MR2188984 https://doi.org/10.1111/j.1467-9868.2006.00542.x

TORABI, M. and RAO, J. N. K. (2010). Mean squared error estimators of small area means using survey weights. *Canad. J. Statist.* **38** 598–608. MR2753004 https://doi.org/10.1002/cjs.10078

VERRET, F., RAO, J. N. K. and HIDIROGLOU, M. A. (2015). Model-based small area estimation under informative sampling. *Survey Methodology* **41** 333–347.

WISCHMEIR, W. H. and SMITH, D. D. (1965). Predicting rainfall erosion losses from cropland east of the Rocky Mountains: Guide for selection for practices for soil and water conservation. U.S. Department of Agriculture. Agricultural Handbook NO. 282.

WOLTER, K. M. (2007). *Introduction to Variance Estimation*, 2nd ed. *Statistics for Social and Behavioral Sciences*. Springer, New York. MR2288542

YI, G. Y., RAO, J. N. K. and LI, H. (2016). A weighted composite likelihood approach for analysis of survey data under two-level models. *Statist. Sinica* **26** 569–587. MR3497760

YOU, Y. and RAO, J. N. K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Canad. J. Statist.* **30** 431–439. MR1944372 https://doi.org/10.2307/3316146

ZIMMERMANN, T. and MUNNICH, R. T. (2018). Small area estimation with a log-normal mixed model under informative sampling. *J. Off. Stat.* **34** 523–542.