remote sensing

MDPI

Article

Improving Estimates of Natural Resources Using Model-Based Estimators: Impacts of Sample Design, Estimation Technique, and Strengths of Association

John Hogland [1,*] and David L. R. Affleck [2]

[1] Rocky Mountain Research Station, U.S. Forest Service, Missoula, MT 59801, USA
[2] W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA; david.affleck@umontana.edu
* Correspondence: john.s.hogland@usda.gov; Tel.: +1-406-329-2138

Abstract: Natural resource managers need accurate depictions of existing resources to make informed decisions. The classical approach to describing resources for a given area in a quantitative manner uses probabilistic sampling and design-based inference to estimate population parameters. While probabilistic designs are accepted as being necessary for design-based inference, many recent studies have adopted non-probabilistic designs that do not include elements of random selection or balance and have relied on models to justify inferences. While common, model-based inference alone assumes that a given model accurately depicts the relationship between response and predictors across all populations. Within complex systems, this assumption can be difficult to justify. Alternatively, models can be trained to a given population by adopting design-based principles such as balance and spread. Through simulation, we compare estimates of population totals and pixel-level values using linear and nonlinear model-based estimators for multiple sample designs that balance and spread sample units. The findings indicate that model-based estimators derived from samples spread and balanced across predictor variable space reduce the variability of population and unit-level estimators. Moreover, if samples achieve approximate balance over feature space, then model-based estimates of population totals approached simple expansion-based estimates of totals. Finally, in all comparisons made, improvements in estimation were achieved using model-based estimation over design-based estimation alone. Our simulations suggest that samples drawn from a probabilistic design, that are spread and balanced across predictor variable space, improve estimation accuracy.

Keywords: sample design; model-based estimation; spread; balance

Citation: Hogland, J.; Affleck, D.L.R. Improving Estimates of Natural Resources Using Model-Based Estimators: Impacts of Sample Design, Estimation Technique, and Strengths of Association. *Remote Sens.* **2021**, *13*, 3893. https://doi.org/10.3390/rs13193893

Academic Editors: Georgios Mallinis and Charalampos Georgiadis

Received: 25 August 2021
Accepted: 25 September 2021
Published: 29 September 2021

1. Introduction

Managers of natural resources need accurate information describing the resources they manage to make informed decisions [1–3]. Owing to high acquisition costs, managers typically employ sampling and estimation techniques to describe various aspects of a given population. Additionally, to increase estimation accuracy, ancillary data, such as remotely sensed data, can be used to improve population estimates through specification and application of models that characterize how the resources of interest vary as a function of the ancillary data (e.g., [4–7]). Whether or not such models are employed or if simple expansion-based techniques are used to estimate population parameters, sample design plays an important role in estimation accuracy [7,8].

Within the context of estimating forest characteristics from remotely sensed data, less emphasis is generally placed on rigorous sample design to train models than when validating models [9]. In large part, this discrepancy stems from the conditions required for design versus model-based inference [4,5]. As Gregoire [8] describes, "in the design-based framework, the population is regarded as fixed whereas the sample is regarded as a realization of a stochastic process". Conversely the model-based framework assumes,

"that the values $y_i, \ldots, y_n$ are regarded as realizations of random variables $Y_i, \ldots, Y_n$ and hence the population is a realization of a random process". The primary difference being, "in the model-based approach [inference] stems from the model, not from the sampling design". While some rely on the availability of models as justification for deviating from probabilistic designs when drawing inferences, it is critical to remember that with natural systems, we seldom understand or can collect all the information needed to build models that completely describe the complexities of those systems. This can lead to model misspecification, localized deviations in relationships, and more generally to models that do not describe a specific population well. Therefore, sample designs used to calibrate models that include randomness in the selection process are critical to developing models that can be used to estimate population and subpopulation parameters. Moreover, probabilistic designs that ensure that samples are widely distributed across feature space can reduce the variability of population estimates.

However, many mapping projects employ non-probabilistic sample designs when calibrating predictive models [9] or use designs that may not fully capture the spread of predictor variables and that may be unbalanced with regard to the population. Furthermore, some authors have argued that sample units should only contain pure homogenous examples of a given class [10–15] and have developed sampling protocols to ensure that outcome. Compared to probabilistic sampling designs, samples of homogeneous units can be expected to estimate lower error rates when calibrating and validating models. However, such metrics can only be attributed to the class of units targeted for sampling and typically cannot be generalized to the rest of the population [16,17].

Stehman [16,18,19] describes multiple probabilistic sampling designs and outlines their associated strengths and weaknesses as they relate to map accuracy. Tille and Wilhelm [20], Grafstrom et al. [6], and Steven and Olsen [21] further describe the importance of probabilistic designs while highlighting concepts of balance and spread as they relate to the precision of estimators. Balance and spread in this case refer to characteristics of a sample with regards to auxiliary variables. Specifically, a balanced sample is one for which the estimated means or totals of the auxiliary variables equal the actual population means or totals of those variables [20]. A sample that is well spread in auxiliary space means that the sample units are widely dispersed across the auxiliary values of that population [21]. For population estimates, balance becomes appealing where there are potentially strong relationships between response and auxiliary variables; for balance then implies that a sample will provide an accurate estimate of the mean or total of an auxiliary variable, and hence of the true mean or total of the response variable. Likewise, a sample that is well spread across auxiliary variables has the advantage of being balanced [22] while also capturing the variability in auxiliary variables. While often described for expansion-based estimators, these arguments are also important in the context of sampling to develop models to support estimation [9,20,23–25]. Nevertheless, many researchers disregard these warnings and build models derived from unbalanced, ill spread, and/or non-probabilistic samples to estimate characteristics of a given population.

Reasons for deviating from a probabilistic study design typically stem from logistical constraints and cost of implementation. However, using predicted values derived from models developed from non-probabilistic designs to describe complex natural systems such as a forested landscape can be highly questionable [26–28]. Questions pertaining to how sample units should be spread across multiple dimensions of feature and geographical space and subsequent impacts on estimator accuracy and inference are at the very forefront of understanding how resource assessments and maps can be used to inform decision making. In this study, we investigate these questions through a series of simulations that compare and contrast estimates of population totals and pixel values obtained under varying sampling designs. We also evaluate the relative impact of sampling design on estimators derived without use of ancillary data and from various linear, non-linear, and non-parametric models. Specifically, in this study, we: (1) evaluate the relative accuracy of alternative model-based estimators of unit values and population totals across populations

where the form and/or strength of associations vary, (2) evaluate the impact of probabilistic designs that spread and balance samples over geographic or feature space on the accuracy of those estimators, (3) demonstrate the potential impact of using a probabilistic sampling design taken from a subset of geographic space on the performance of those estimators and (4) assess the impact of misspecified models on the accuracy of estimation. We anticipate that within our simulations, samples drawn from probabilistic designs that are spread and balanced across predictor variable space will produce better estimates of the population and subpopulations.

2. Theoretical Background

In this study we assess expansion and model-based estimation procedures applied to populations constructed according to distinct functional relationships between response and predictor variables but tempered by normally distributed errors of varying magnitudes. While fundamentally different, expansion and model-based estimation techniques can be used to estimate population parameters [29]. For expansion-based methods, parameter estimators are derived by weighting observed response values according to the relative frequency with which the corresponding sample units are selected under the design. In contrast, model-based estimators of population parameters are derived from a model of how the response variables change across the population. In this approach, the emphasis is on estimating model parameters, which can then be indirectly used to estimate either parameters of a given realization of the model, or the model-expectations thereof [29]. Contrary to the expansion-based methodology, the model-based approach is not necessarily tied to a specific population but instead assumes that a given population is a realization of some random process, typically involving known or mapped auxiliary variables. The emphasis of the model-based approach is often on the relationships between a variable of interest (response) and these other variable(s) (predictors). Given those relationships, the forms of deviations around them, and the predictor variable values within a specific population, one can then estimate model and population parameters.

While models do not necessarily need to be tied to a given population, it is the case that within natural systems the underlying shapes and forms of the relationships between response and predictor variables vary and may be unknown and noisy. Therefore, model development is often based on a specific population (e.g., [30,31]), making models uniquely calibrated to a given time and place. Though uniquely calibrated, different modeling techniques make different assumptions with regard to the relationships between response and predictor variables. For many natural systems, these assumptions may be difficult to meet and can lead to a model that is misspecified. To assess the impact of potential model misspecification on estimating population and subpopulation parameters, we use simulated landscapes with known functional relationships and introduced error (Figure 1). The functional shape of relationships evaluated in our simulations include linear, quadratic, and nonlinear distributions with normally distributed errors and various amounts of introduced noise (Section 3.1). Additionally, estimation techniques are evaluated under various sample designs to assess the impact of spread and balance on parameter estimation (Section 3.2). Expansion and model-based estimators were chosen based on their frequency of application, underlying assumptions, and flexibility in capturing various relationships among response and predictor variables. Modeling techniques evaluated include linear regression (LN), neural networks (NNs), support vector machines (SVMs), random forests (RF), and generalized additive models (GAMs).

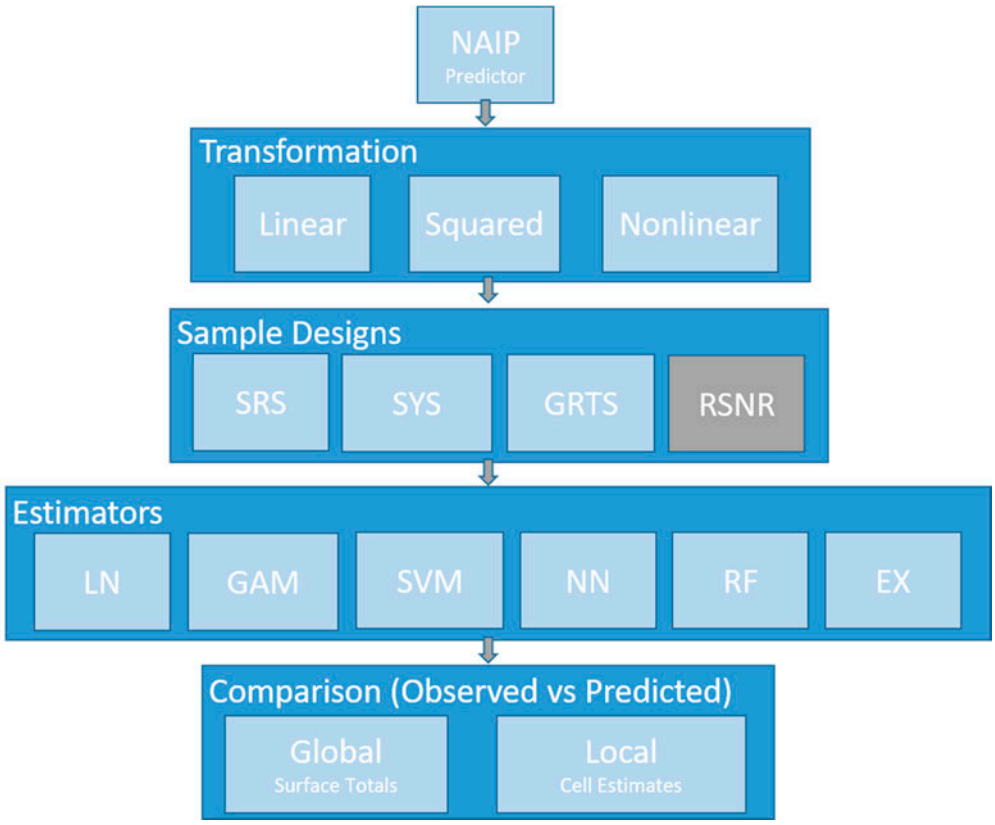


Figure 1. Diagram of simulations and workflow to determine the impact of sample designs on model and design-based estimates. The gray box in sample design denotes a probabilistic sample design for a subset of geographic space. NAIP = National Agriculture Imagery Program imagery; SRS = simple random sample, SYS = systematic random sample, GRTS = generalized random tessellation stratified, RSNR = simple random sample near roads, LN = linear model; GAM = generalized additive models, SVM = support vector machines, NN = neural networks, RF = random forests, EX = Horvitz Thompson expansion estimator.

In terms of complexity, expansion estimators are the simplest, followed by model-based estimators LN, NNs, SVMs, RF, and GAMs. Expansion estimators are commonly limited to a singular, population-wide estimate of mean or total for a population of interest [7]. Conversely, linear regression models can estimate the mean response value of pixels for given values of predictor variables but are obviously limited in their ability to describe nonlinear relationships and can produce estimates outside the range of values observed in a sample [32]. NNs, a machine learning technique, can represent nonlinear associations by introducing activation functions and weighting schemes of linear coefficients [33]. SVMs use kernel functions to identify subsets of the data within multidimensional feature space that describe the relationship between response and predictor variables [33]. RFs allow for discontinuities in the response function across feature space by implementing classification and regression trees (CART) and incorporating bagging and randomness into model training. RFs ultimately rely on ensembles of CART models to capture relationships between response and predictors but can yield only estimates within the range of observed response values [34]. GAMs are based on a flexible modeling technique that assumes additivity in predictor variables effects, and that response values change smoothly over feature space. GAMs allow for nonlinear, nonparametric relationships between response and predictor variables [35] and can explicitly account for non-Gaussian error distributions.

Of particular interest with regard to expansion and model-based estimators is that an expansion estimator can be thought of as a model-based estimator with only an intercept parameter if an equal probability design is used and if the presumed model incorporates an independently identically distributed error structure. That is, if the same sample is used

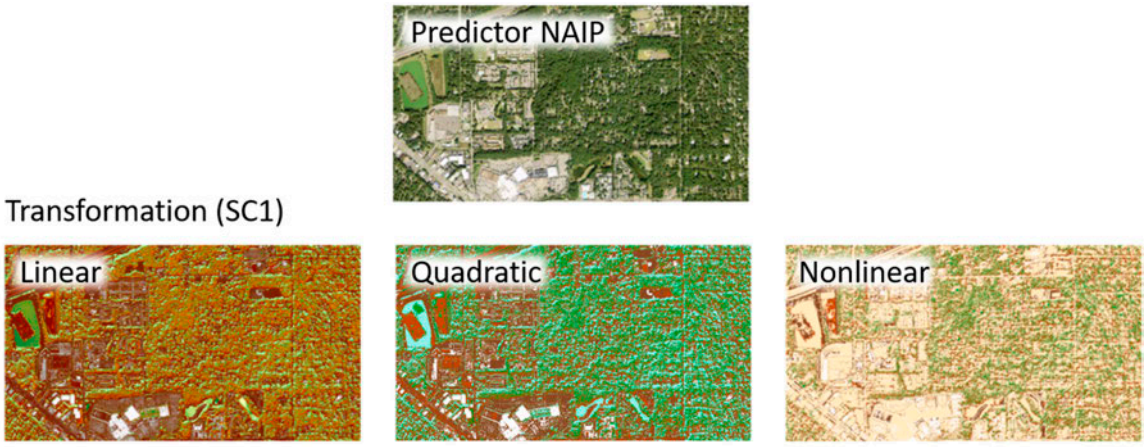
by an expansion estimator to calibrate an intercept only model, the population estimates from the expansion and model-based estimators will be the same. This breaks down if unequal probability designs area used, because information on sampling intensities under a design are not used in the model-based approach. Furthermore, if relationships between the response and potential predictors cannot be identified, then the only viable model is one describing variation around the mean, which suggests use of simple estimators like the sample mean. Owing to this, within our simulation, we predict that as the amount of model error (noise) introduced into the relationship between response and predictors is increased, model and expansion-based estimates will converge. Conversely, as the amount of noise introduced between response and predictor decreases, the estimated values from the model-based estimators will be less variable than the expansion-based estimates. Similarly, within our simulations we anticipate that when underlying model relationships are misspecified or calibrated in the absence of important correlated data, model-based estimates will converge with expansion-based estimates. Moreover, we predict that samples drawn from probabilistic designs that enhance spread and balanced across predictor variable space will produce better estimates of the population and subpopulations.

3. Materials and Methods

3.1. Simulated Raster Surfaces

A spatial subset extracted from a National Agricultural Imaging Program (NAIP) [36] image located in the panhandle of Florida, USA, serves as our base dataset for all derived raster surfaces used within our simulations. This subset was acquired in the summer of 2018 and covers approximately 6 km^2 of a mixed forested/urban landscape (Figure 2) with a nominal pixel resolution of 1 m^2. The four spectral bands of the NAIP image (x_i; i = 1, 2, 3, 4) were transformed to produce three distinct continuous surfaces (SC1, Figure 2). The first SC1 surface was obtained by averaging the band values for each pixel:

$$\mu_{1j} = 0.25 \sum_{i=1}^{4} x_{ij} \quad (1)$$



Response Surfaces Statistics

% Noise	Mean	SD	Min	Max	Moran's I
20	139.6	40.0	19.2	275.1	0.9
40	139.6	42.2	-12.6	305.5	0.8
60	139.6	45.7	-41.2	332.3	0.7
80	139.6	50.2	-77.8	375.1	0.6

% Noise	Mean	SD	Min	Max	Moran's I
20	21023.4	11228.9	-7261.4	69092.5	0.9
40	21026.6	11860.3	-16593.9	75489.2	0.8
60	21024.7	12837.7	-29005.1	84684.2	0.7
80	21025.2	14103.9	-34668.9	97319.5	0.6

% Noise	Mean	SD	Min	Max	Moran's I
20	62.6	78.5	-85.2	285.2	0.7
40	62.6	82.9	-154.0	350.3	0.6
60	62.6	89.8	-242.0	426.9	0.5
80	62.6	98.6	-331.9	480.1	0.4

Figure 2. Depiction of the NAIP, Linear, Quadratic, and Nonlinear transformations (SC1). Beneath each SC1 depiction, cell mean, standard deviation, minimum, maximum, and global Moran's I [37] (rook neighborhood) statistics are presented by the amount of random error (% Noise) introduced within each response surface.

The second SC1 surface was obtained as the square of the first,

$$\mu_{2j} = \mu_{1j}^2 \quad (2)$$

and thus is a quadratic function of the individual band values and their cross-products. Finally, a discontinuous, nonlinear SC1 surface was created by adding an additional logical query on the NAIP band 4 cell values:

$$\mu_{3j} = \lambda_j \mu_{1j} \quad (3)$$

where λ_j is an indicator variable taking the value 1 if the value of band 4 for pixel j falls between 170 and 210 and is 0 otherwise.

For each SC1 surface (Linear, Quadratic, and Nonlinear), normally distributed errors were added to create response surfaces $y_{kj} = \mu_{kj} + \varepsilon_{kj}$ (Figure 2), where the ε_{kj} are independent, identically distributed errors. Normally distributed errors were generated using a standard deviation proportional to the standard deviation of pixel values within a given SC1 and a mean of zero; the constants of proportionality were set to 20%, 40%, 60%, and 80% to provide an equally spaced range of noise added to each relationship and to evaluate the impact of noise on model estimation. The realized errors produced homoscedastic surfaces with error values centered at zero before being added to SC1 surfaces. In total, 12 unique continuous response surfaces were created, with various shapes and amounts of error (% Noise).

3.2. Sample Designs

To evaluate the impact of sample designs on estimation, we compared population and raster cell estimates for each of our response surfaces using four different sampling designs. Sample designs included simple random sampling (SRS), systematic random sampling (SYS), modified Generalized Random Tessellation Stratified sampling (GRTS; [21,38]), and randomly selecting sample units next to roads (RSNR). In our simulation, SRS selects locations across the spatial domain of the image, independently and uniformly at random. SYS spreads sample units evenly across the spatial domain of the image by distributing locations over a square grid of fixed orientation. The dimensions of each square grid cell was calculated as the square root of the total image area divided by a sample size of 50. SYS was achieved by selecting a random start within the bottom east square grid cell and systematically selecting locations based on the grid cell dimensions. Owing to the systematic nature of the design and the dimension of each grid cell, some SYS samples initially had fewer or more than 50 sample units. For SYS samples with more than 50 locations selected, locations were randomly removed until the sample size reached 50. For SYS samples with less than 50 locations selected, random locations across the spatial domain of the image were added to the sample until the sample size reached 50. GRTS spreads and balances samples in auxiliary space based on the NAIP spectral values (bands 1–4). To achieve this, we selected an initial SRS of 100,000 locations within the NAIP image, extracted raw spectral values from each band at those locations, and performed a principal components analysis (PCA) using the bands' correlation matrix. Using the loadings of the first two components of the PCA, we then transformed the 100,000 sampled NAIP cell values to bivariate component scores. These were then input into the GRTS algorithm [38] to select a sample of 50 observations spread and balanced within a two-dimensional PCA score space; response surface values were available only at these 50 locations. RSNR uses Tiger Road files [39] and a 50 m buffer around roads to subset the population and then randomly select sample unit locations within those subsets. Sample sizes of 50 for model calibration and parameter estimation were chosen to reflect common cost constraints associated with field data collection campaigns and were held constant across estimation techniques to control for sample intensity.

3.3. Model Calibration and Estimation

LN, GAM, SVM, NN, and RF models were trained separately for each sample drawn from the 12 response surfaces and 4 sample designs. For LN models, ordinary least squares regression was used to relate response and predictor variables [40]. For GAMs, the Gaussian family with an identity link and additive thin plate penalized regression splines were used to relate response and predictor variables [41]. SVM models used Epsilon Regression and a Radial Basis kernel (Gaussian) [42]. NNs fit a single layer hidden linear model (least squares) with a size parameter equal to half the sample size and a decay parameter equal to 1/10 the size parameter [43]. RF models averaged the results of 20 regression trees drawing on 66% of the data; owing to the small number of predictors involved (three or four), each tree randomly selected one predictor variable at each training node [44]. All model calibrations assumed independent, homoscedastic errors.

Model predictor variables initially included all four NAIP band cell values. That is, all information used to generate the response surfaces were made available in the modeling process as predictor variables. Below, we refer to these as fully-specified models. A second class of models (partially-specified models) utilized only the first three NAIP bands as predictors. This represents the common situation in which the available predictor variables do not fully describe the response function.

Once trained, model and NAIP cell values were used to create prediction surfaces of estimated raster cell values. Estimates of the population total ($\hat{\tau}_y$) were calculated by aggregation:

$$\hat{\tau}_y = \sum_{i=1}^{N} \hat{y}_i \tag{4}$$

where $\hat{y}_i$ is an estimated pixel value obtained from a calibrated model of one of the above types and N is the total number of pixels in the population.

Additionally, from the response values alone, each sample provided an expansion estimate of the population total obtained simply from the per-pixel sample mean expanded by the population size:

$$\hat{\tau}_y = \frac{N}{n} \sum_{i=1}^{n} y_i \tag{5}$$

where n is the sample size (50).

3.4. Evaluation

For each sample design (d) and response surface (k), 100 samples (iterations) were drawn and used to calibrate models and estimate surface characteristics. The spread of each sample (B) was measured using the approach described in [21,22]. Briefly, Mahalanobis distance [45] was used to identify Voronoi polytopes (p_i) around each selected pixel of a particular sample in multidimensional predictor space or in geographic space. Inclusion probabilities (n/N) were then summed within each polytope and the variance of these sums were then calculated:

$$B = \frac{1}{n} \sum_{i \in s} (v_i - 1)^2 \tag{6}$$

where v_i is the sum of inclusion probabilities in p_i.

For each response surface, sample design, and estimation technique, estimates were compared against the corresponding response surface total or against individual cell values. For a global perspective, estimates of the response surface totals ($\hat{\tau}_y$) were recorded and compared against actual totals ($\tau_y = \sum_{i=1}^{N} y_i$); differences were expressed relative to the

actual totals, on a percentage basis. For a local perspective, estimated cell values were recorded and the RMSE was calculated across cells as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (7)$$

RMSE was also calculated for the expansion estimator reflecting the fact that the expansion estimator (scaled by N) is the only available estimate of cell values as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\overline{y} - y_i)^2} \quad (8)$$

where $\overline{y}$ is the simple sample mean. While similar to the variance of the expansion estimator, this RMSE expression is evaluated over the entire population and is a measure of within response-surface variability. Using either expression, relative RMSE was then calculated as

$$\text{RRMSE} = \frac{100\text{RMSE}}{sd_y} \quad (9)$$

where sd_y is the actual population standard deviation.

4. Results

4.1. Allocation of Sample Units

Of all the sample designs considered, only GRTS utilized information from feature space. Although it used only the first two principal components of the four bands, these accounted for approximately 97% of the total variation in the NAIP image. Within feature space, GRTS sample means were approximately balanced, clustering near the population mean values for each band (Figure 3). By contrast, samples drawn by SRS, SYS, and RSNR designs were unbalanced, with individual sample means varying substantially from band averages (Figure 3). In geographic space, GRTS and SRS had similarly dispersed distributions of sample means while the distribution of sample means for the SYS design was tighter (Figure 3). However, the distribution of sample means for RSNR was skewed to upper right quadrant in Figure 3 owning to road densities (Figure 4). Imbalance of SYS samples in geographic space was higher than anticipated, possibly as a result of the sample adjustments made to ensure a constant sample size of 50 units.

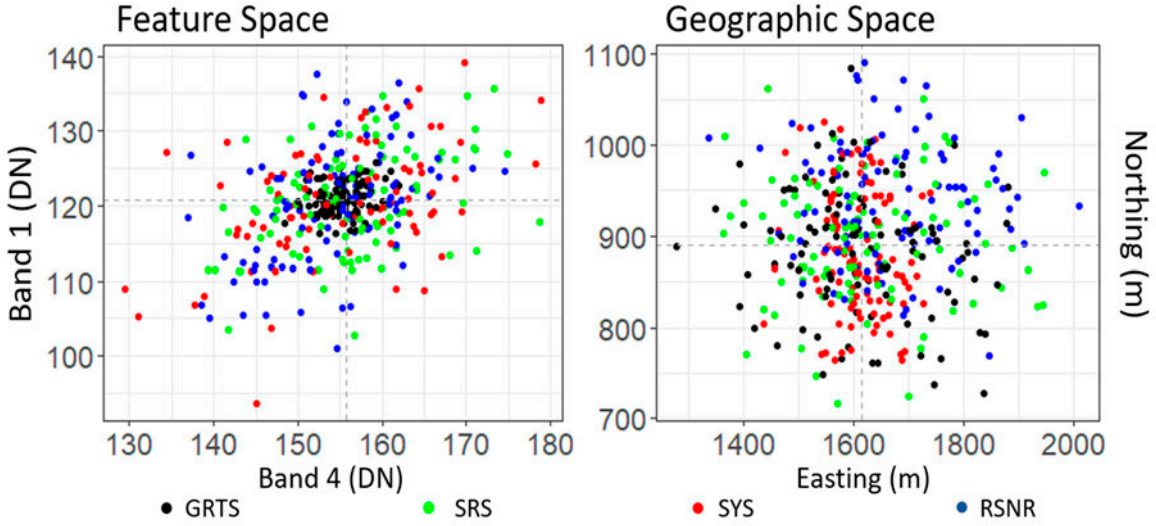


Figure 3. Distribution of sample mean digital number (DN) values of bands 1 and 4 (**left** panel; Feature Space) and sample mean northing and easting values (**right** panel; Geographic Space) for 100 samples obtained from generalized random tessellation stratified (GRTS), simple random Scheme 50. was used for each of the 100 samples.

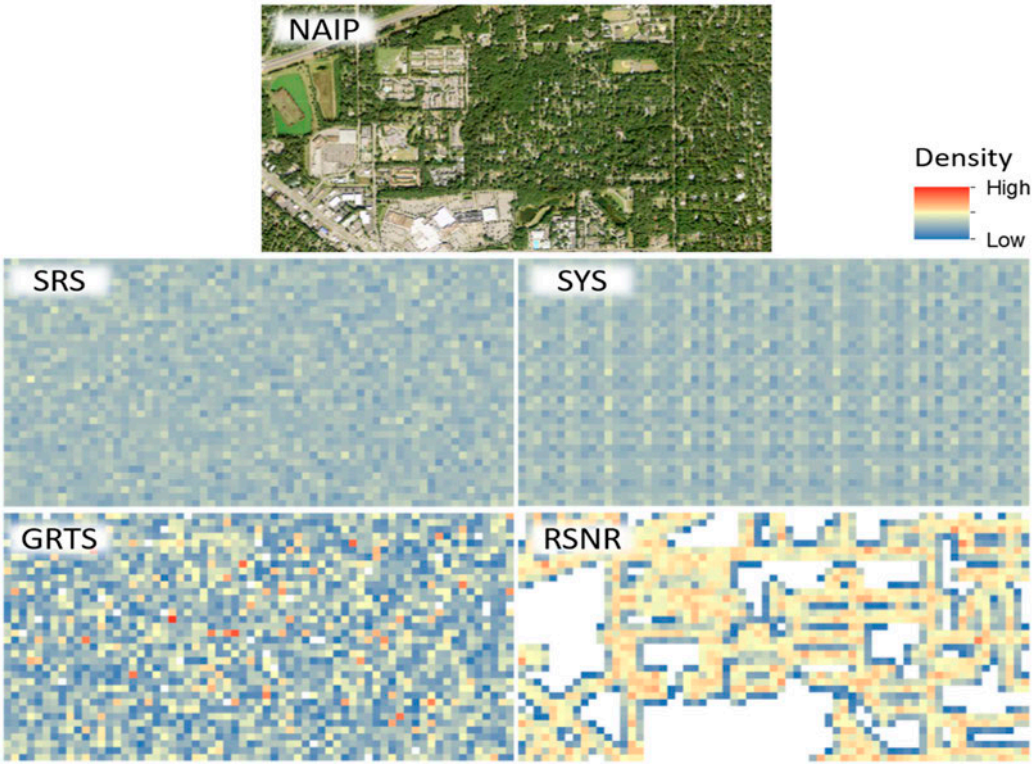


Figure 4. Illustration of spatial density of sample unit locations for each sample design across the study area. The NAIP image is used for reference; the density raster surfaces depict the frequency of selected cell locations for samples using the SRS, SYS, GRTS, and RSNR designs. To enhance the display of density surfaces, each surface was created by drawing 2000 random samples (sample size 50) for each design and counting the number of sample units falling within each density raster cell (50 m grain size). White cells within the density raster surfaces identify cells where no sample units were selected. SRS = simple random sample; SYS = systematic random sample; GRTS generalized random tessellation stratified; RSNR = random sample near roads.

While the RSNR design only selected sample units within 50 m of roads inside the study area and appeared to underrepresent impervious surfaces such as industrial areas and buildings (Figure 4), mean values of NAIP bands and northing and easting closely matched population parameters in feature space (Figure 3). Moreover, even though no sample units were selected further than 50 m from a road in the RSNR design, sample unit locations appeared to be generally spread across the study area in both feature and geographic space (Figure 3). While the areas within 50 m of a road represent a subset of the geographic domain of the study area, the spatial extent of that subset still spanned a broad distribution of feature space (Figure 3).

4.2. Model Calibration and Functional Relationships

While trends in population estimates were generally consistent across response surfaces, the most complex results occurred when estimators were applied to the nonlinear SC1 response surfaces. For these surfaces, the RF estimator would presumably have an advantage over expansion, LN, SVM, NN, and GAM estimators given that the latter techniques presume smooth response functions over feature space and are limited in their ability to capture the discontinuities in these response surfaces. Figure 5 depicts the averaged fitted estimates over the 100 SRS iterations for a slice of feature space (40% introduced normally distributed errors). Relatively speaking, fully-specified GAM, RF, and SVM estimators were able to reproduce the rise and drop of the response surface cell values. While NN estimators partially capture the rise in values, they were not able to capture the subsequent fall in those values in nonlinear relationships (Figure 5). As expected, the linear estimators

only allowed for constant rates of change. They could therefore describe patterns in the linear SC1 surfaces but failed to describe characteristics of the other two SC1 surfaces.

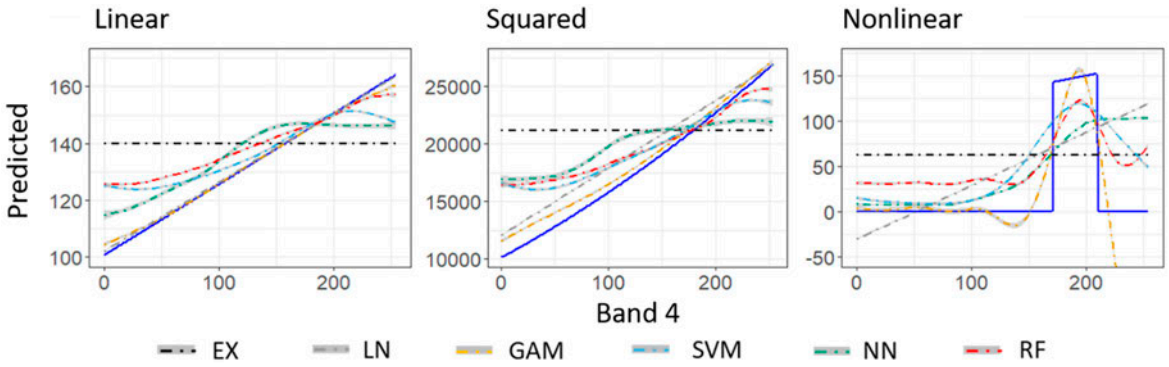


Figure 5. Depiction of functional relationships of fully-specified estimators and SC1 surface values (blue line) for a slice of predictor variable space where NAIP cell values were held constant at the mean for bands 1–3 while band 4 cell values (*x*-axis) were allowed to vary. The amount of random noise introduced into the relationships between predictor and SC1 surfaces was 40% of the SC1 surface standard deviation. Sample design used to generate models was simple random sampling. Dashed colored lines correspond to averaged expansion and model-based estimates. The grey shaded regions around each estimation technique correspond to the 99% empirically based confidence interval given the 100 iterations performed. Sample size for expansion estimates and model calibration was 50. EX = expansion; LN = linear; GAM = general additive model; SVM = support vector machine; NN = neural networks; RF = random forests.

These aspects of model performance were consistent across sample designs and levels of introduced noise. As anticipated, when more error was introduced into the response surfaces, variability in estimates increased. When comparing estimation techniques across these slices of SC1 response surface values, fully-specified GAMs, RFs, and SVMs consistently outperformed other estimation approaches (Figure 5).

4.3. Estimates Using NAIP Bands as Predictors

As expected, sample design had an impact on the accuracy of estimated pixel values and population totals (Figures 6 and 7). While we anticipated that the RSNR design would produce biased estimates, bias was minor in estimated values (Figure 7). Moreover, across all response surfaces, similar patterns in RMSE and variability in the estimated population totals were observed. Generally, as the proportion of noise increased in the response surfaces, RMSE and the variability in population estimates increased (Figures 6 and 7). Because results and trends were similar for linear, squared, and nonlinear response surfaces, we primarily present results for the nonlinear (most complex) response surfaces within this section.

For the expansion estimators, RMSE varied little across iterations within nonlinear response surfaces but consistently exceeded the RMSEs of the fully-specified model-based estimators (Figure 6). While there did appear to be minor improvements in RMSE for the GRTS design, it is difficult to clearly rank sample designs with regards to RMSE in these figures. Generally, though, GRTS and SYS designs had lower RMSE values than the other designs (Table 1).

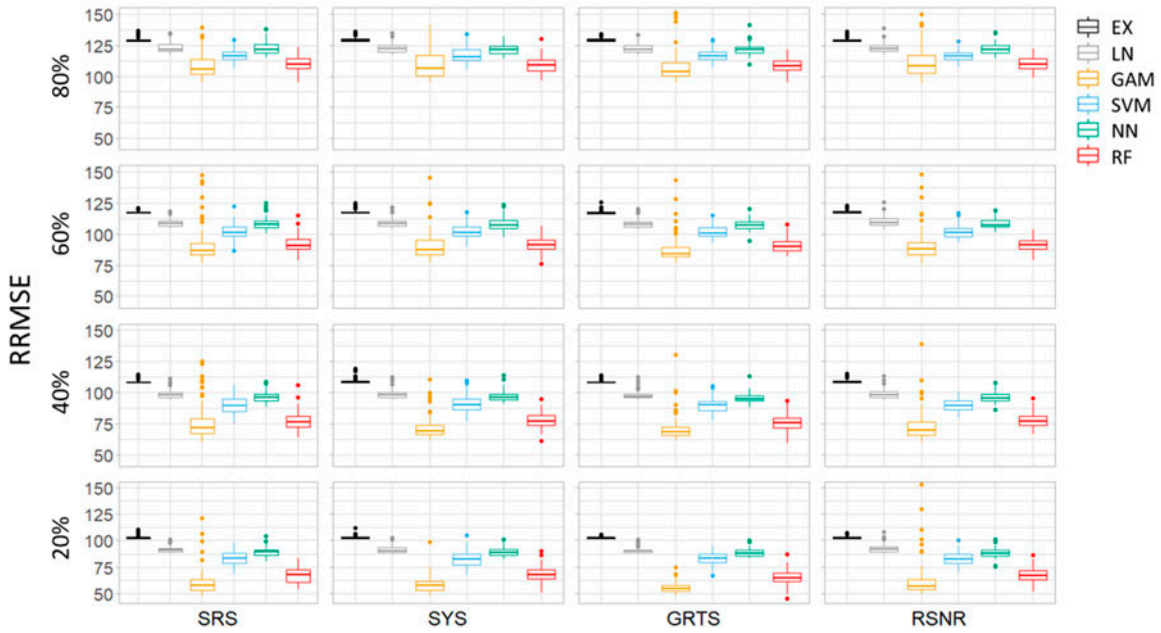


Figure 6. Relative RMSE (RRMSE) for expansion and fully-specified model-based estimates derived from 100 iterations of nonlinear trend response surfaces. Response surfaces incorporated random errors at 20%, 40%, 60%, and 80% of the total nonlinear SC1 image standard deviation. Column and row titles identify sample designs and proportion of SC1 standard deviation used in response surfaces. Ex = expansion; LN = linear; GAM = general additive model; SVM = support vector machine; NN = neural networks; RF = random forests; SRS = simple random sample; SYS = systematic random sample; GRTS generalized random tessellation stratified; and RSNR = random sample near roads.

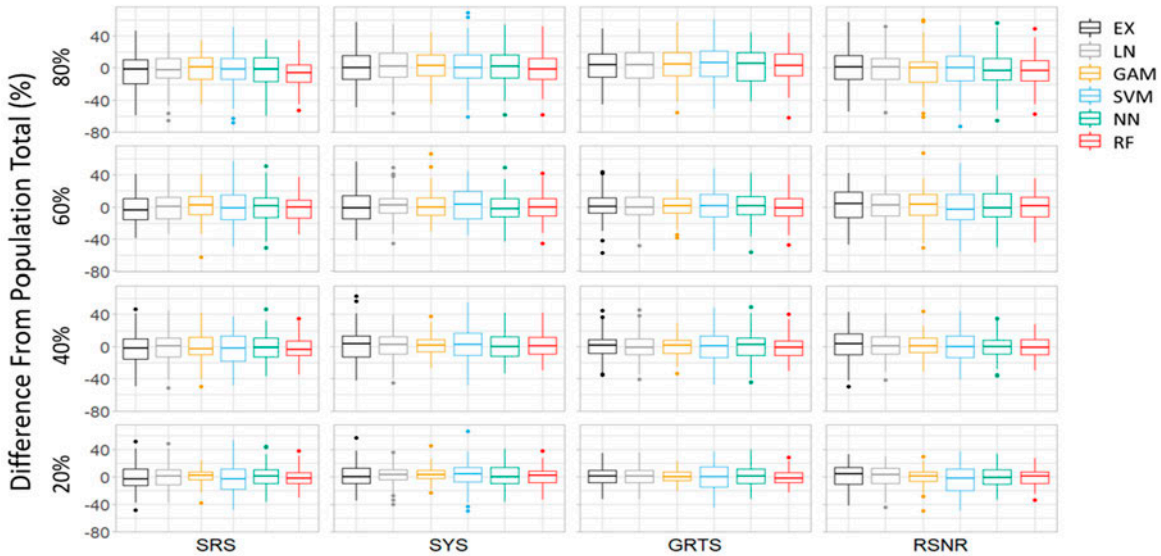


Figure 7. Distribution of estimation errors for response surface population totals. Estimation errors are expressed as a percentage of the population total for trend response surfaces with normally distributed errors based on 20%, 40%, 60%, and 80% of the total nonlinear SC1 image standard deviation. Column and row titles identify sample designs and proportion of SC1 standard deviation used in response surfaces. Ex = expansion; LN = linear; GAM = general additive model; SVM = support vector machine; NN = neural networks; RF = random forests; SRS = simple random sample; SYS = systematic random sample; GRTS generalized random tessellation stratified; and RSNR = random sample near roads.

Table 1. Design with smallest average root mean squared error (RMSE) across iterations by response surface distribution (SC1), fully-specified modeling technique, and amount of introduced noise (% Noise).

SC1	% Noise	GAM	RF	SVM	NN	LN	EX
Linear	20	SRS	GRTS ***	GRTS ***	GRT S *	SRS	GRTS ***
	40	GRTS	GRTS	GRTS ***	SRS	GRTS	GRTS **
	60	GRTS	GRTS *	GRTS ***	GRTS *	RSNR	GRTS ***
	80	SYS	SYS	GRTS **	GRTS	SYS	GRTS *
Squared	20	GRTS *	GRTS ***	GRTS ***	GRTS	GRTS ***	GRTS ***
	40	SYS	GRTS ***	GRTS **	GRTS	GRTS *	GRTS ***
	60	GRTS	GRTS *	GRTS *	SYS	GRTS	GRTS ***
	80	SRS	GRTS	GRTS*	GRTS	GRTS	GRTS *
Nonlinear	20	GRTS *	GRTS *	SYS	RSNR *	GRTS	GRTS *
	40	GRTS **	GRTS	GRTS	GRTS	GRTS	GRTS
	60	GRTS	GRTS	RSNR	GRTS	GRTS	GRTS
	80	GRTS	GRTS *	RSNR	GRTS	GRTS	GRTS

* statistically different than SRS at $\alpha = 0.05$; ** statistically different than SRS at $\alpha = 0.01$; *** statistically different than SRS at $\alpha = 0.001$; GAM = general additive model; RF = random forests; SVM = support vector machine; NN = neural network; LN = linear model; EX = expansion estimator; SRS = simple random sample; SYS = systematic random sample; GRTS generalized random tessellation stratified; RSNR = random sample near roads.

From the perspective of population totals, the GRTS designs had the largest positive impact on expansion-based estimators (Figure 7, reduction in variability of % bias). While the impact of sample designs was less apparent for model-based estimators, generally the GRTS design had the least variation in population estimates (Figure 7), while producing the smallest RMSEs (Table 1). Across all iteration, the mean difference between observed and estimated total, measured as a percentage of the response surface total, was close to zero (Figure 7), suggesting that estimator bias, if present, was minor.

On the basis of RMSE, the best performing estimator for all sample designs, transformations, and introduced errors was the GAM estimator followed by RF, SVM, NN, LN, and expansion estimators. Comparing results across sample designs for the GAM estimators, GRTS and SYS typically had the smallest RMSE and highest accuracy in population estimates. While we anticipated that expansion and model-based estimates would be slightly biased for the RSNR sample design, our results did not fully support that claim. This is likely due to the relatively large proportion of area within 50 m of a road in the NAIP image subset, the identical trend and error distributions applied across the study area, and the similar distribution of spectral values found within the road buffers and the image.

As anticipated, when evaluating designs, percent error, and model estimation techniques for various ranges of response variable values using nonlinear response surface cell values, we saw similar trends as described in Section 4.2 and as displayed in Figure 5. Figure 8 illustrates these relationships by presenting predicted versus observed smoothed trends for the GRTS sample designs, expansion, and model-based estimators, and 40% introduced noise into the relationship between response and predictors. Of particular note in Figure 8 is that all methods produced attenuated estimates and that the GAM, RF, and SVM procedures more closely match the one-to-one line depicted in Figure 8 within the most common areas in feature space (frequency graphs above graphics in Figure 8). These general trends also extended to partially-specified estimators in which only NAIP bands 1–3 were used as predictor variables.

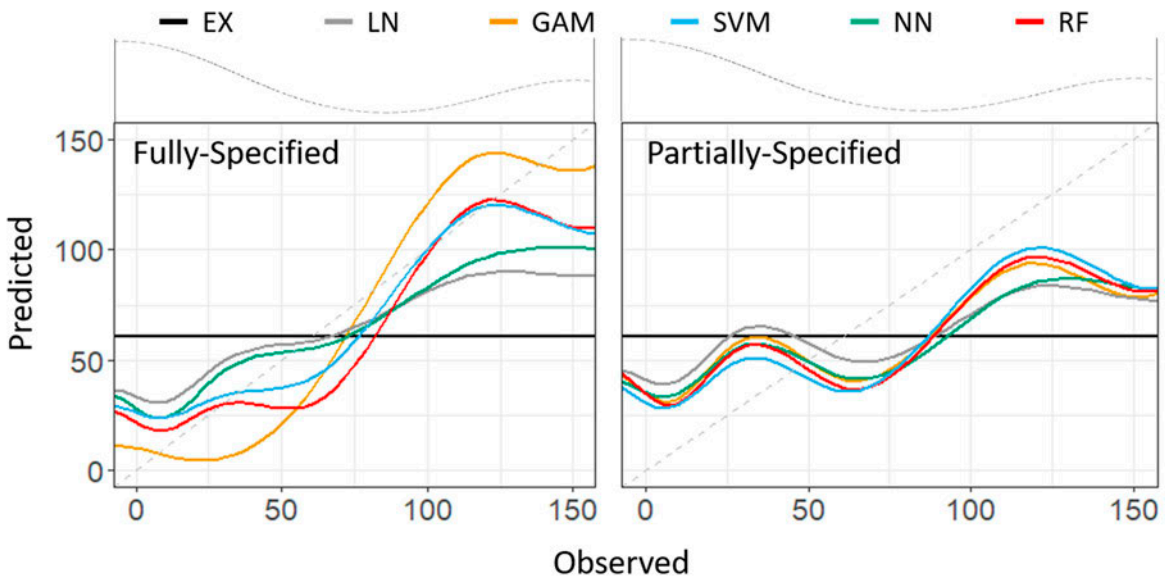


Figure 8. Smoothed trend of nonlinear response surface predicted values (*y*-axis) versus observed (*x*-axis) values for design and model-based estimators calibrated using generalized random tessellation stratified designs (GRTS) and image bands 1–4 (Fully-Specified) and image bands 1–3 (Partially-Specified) as predictor variables. Introduced noise into the nonlinear response surface was held constant at 40% of the nonlinear SC1 surface standard deviation. Sample size was held constant at 50 for each of the 100 iterations used in the study. The gray dashed lines within Fully-Specified and Partially-Specified graphs serve as a one-to-one reference line while the gray dotted lines above the graphs depict relative frequency of each observed value. Ex = expansion; LN = linear; GAM = general additive model; SVM = support vector machine; NN = neural networks; RF = random forests.

Moreover, all models, regardless of misspecification (model distribution or lack of predictors) tended to produce better estimates than expansion-based estimation alone. Similar to results in Section 4.2, the GRTS sample design coupled with the GAM-based estimator produced the best results and more closely followed the one-to-one line when weighted by the frequency of sampled values (Figure 8). Finally, in all comparisons, as percent error increased departures from the one-to-one line also increased (not shown).

4.4. Impact of Spreading Sample Units in Feature Space

While estimation approach had a larger effect on our results, sample design also impacted parameter estimates. Across all response surfaces, iterations, and modeling techniques the GRTS sample design provided estimates with the smallest average RMSEs (or was not statistically different than the smallest averaged RMSEs; Table 1). Additionally, the variability in estimated population totals tended to be smaller for the GRTS sample design (Figure 7). Moreover, for the top three estimation techniques (GAM, RF, and SVM) and the most complex response surfaces (nonlinear), spreading samples in feature space generally produced the smallest RMSEs (Figure 9, Table 1). Interestingly, estimates derived from samples spread in geographic space (SYS) often performed equally to samples spread in feature space (GRTS) even though response surfaces did not depend directly on geography (Equations (1)–(3)). In large part, this can be attributed to spatial correlation in NAIP bands (Figure 2, Moran's I statistic). Nevertheless, across all estimation techniques, estimators derived from samples spread in feature space tended to have the smallest RMSE (Figure 9). Moreover, the sample design that yielded the smallest average values for B (well spread samples) was GRTS.

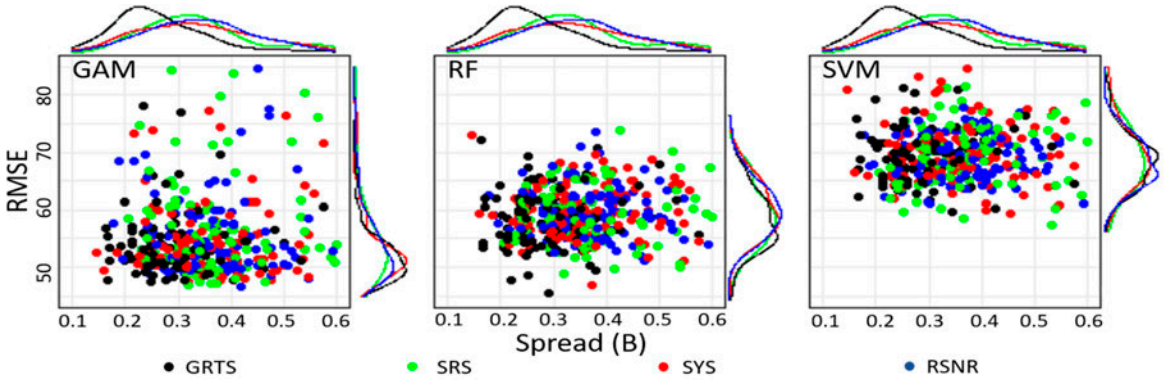


Figure 9. Root mean squared error (RMSE) versus spread values (B) for the top three estimation techniques given nonlinear response surface sample design with 40% introduced noise. Densities of RMSE and B values are presented to the right and above each graph. GAM = general additive model; RF = random forests; SVM = support vector machine; SRS = simple random sample; SYS = systematic random sample; GRTS generalized random tessellation stratified; and RSNR = random sample near roads.

5. Discussion

While sample design affected the performance of all estimators, the estimation approach had a larger impact on cell and population estimates. As expected, in all instances evaluated, regardless of the underlying relationship between response and predictor variables, model-based estimators produced better pixel-level estimates than expansion estimators (Figure 6, Table 1). Additionally, when sample units were spread and balanced in feature space, model-based estimators of population totals had relatively low bias and remained less variable than expansion estimators (Figure 7). This finding suggests that models derived from samples balanced and spread in feature space produce low-bias estimates with less variability than RSNR, SRS, and SYS designs.

Likewise, expansion estimators derived from samples spread and balanced across feature space (GRTS), produced estimates similar to model-based estimates of population totals with less variability than SYS, SRS, and RSNR expansion estimators; supporting the idea that spreading and balancing sample observations in feature space can substantially improve population estimates presented by others [6,20,46]. Within our simulations, the variability in estimated totals was on average always less for model-based estimators and expansion estimators derived from samples spread and balanced in feature space. In all instances the increased precision (reduction in RMSE) of using a model-based estimator over expansion estimators alone outweighed the impacts of the bias introduced by model-based estimators, even when models were misspecified.

Within our simulations, GAM and RF were the two best modeling techniques when all NAIP bands were used to calibrate models. Once calibrated, these estimators were able to better identify and track the underlying transformations of NAIP cell values better than other evaluated estimators. Additionally, when models were misspecified (e.g., a linear model applied to a nonlinear response), raster cell values approached expansion-based estimates. Furthermore, when only the first three NAIP bands were used to calibrate models, model-based cell estimates also approached expansion- based estimates. These findings suggest that using probability sampling, ancillary data, and models calibrated from probability samples improves pixel-level value estimates, even if the model is misspecified and especially if sample units are spread across the values of the ancillary data (feature space). This finding is especially relevant today with the availability of remotely sensed imagery and the correlations between land cover (e.g., forest ecosystems) attributes and spectral and textural image metrics (e.g., [30,31]). While it is tempting to view our results as confirmation of the robustness of GAM and RF modeling techniques, it is important to recognize that within our simulation, randomness played a critical role in all sample designs [20] and that error and SC1 transformations were consistent across the spatial domain. Foody et al. [11] points out that other relationships and error structures may

favor different modeling techniques—for example, non-Gaussian error distributions could advantage the GAM approach. However, on average, when samples were spread and balanced in feature space, pixel-level and population total estimates were as good as or better than samples that were not spread or balanced in feature space. This result is most likely because spreading and balancing sample units across feature space will more consistently allow for the detection of changes in complex relationship between response and predictors than samples that are not spread and balanced.

While we did not directly evaluate sample size in our comparisons, we assume that the strength of the relationship between response and predictor variables, represented in our study as added random noise, would have impacts on estimation similar to sample size. Specifically, we would expect that an increase in sample size would have a similar impact on estimation accuracy as a decrease in noise introduced into the response surfaces. This would suggest that as sample size increases, estimates should become more precise and measures of spread, such as the B statistic [21,22] will become smaller. Future investigations should look at quantifying tradeoffs in sample size, spread, and the strength of the relationship between response and predictor variables on estimation for modeling techniques such as LN, GAM, RF, SVM, and NN.

These simulated findings have practical relevance for natural resource managers who are interested in inventory, monitoring, and managing natural resources. Specifically, improvements in estimating characteristics of a natural resource for a given population (e.g., basal area in a forest) can be gained by incorporating models into the estimation process [4,30,31]. Moreover, indirect estimates of population subdomains (e.g., stands) can be made from model-based estimates [26]. In the case where observational units (e.g., pixels) cover a smaller spatial domain than the subdomain of interest (e.g., stands), those pixel estimates can be aggregated to the spatial extent of the stand. Conversely, when pixels have an extent larger than the stand of interest, model estimates can be attributed to the entire stand. In instances where pixel estimates partially cover the extent of the stand, estimates can be weighted and attributed to the stand based on the amount of overlapping area between pixels and the stand. However, models have estimation error, which should be incorporated into the standard errors of the estimates of the domain and subdomain characteristics [26,47,48].

In our simulation, we used iteration and sampling to quantify estimation error. Our top performing estimator (GAM) on average had the least amount of bias and the smallest RMSE across iterations. However, there were instances (iterations) within our simulations when the GAM model had, relatively speaking, large RMSE. These instances within our iterations identify the case in which a sample drawn from the population were not well balanced or spread. While the GRTS sample design minimized the occurrence of samples that were not well balanced or spread, it did not eliminate those types of samples. This means that while rare, some samples may have a disproportionately large number of extreme occurrences within feature space which could adversely affect the calibration of a given model and model estimates. In this situation, bagging or boosting could be used to iteratively draw random subsets from a given sample, calibrate multiple models, average model estimates, and empirically estimate standard error to reduce the impact of extreme observations [49,50]. Applying methods such as this should have a similar impact to the variation in RMSE values as what is displayed in Figure 9 for the RF modeling techniques, which uses bagging and model averaging [34].

For forest managers, this suggests that estimates of forest characteristics such as species composition, basal area, and tree counts can be improved for a given area by relating field measurements to remotely sensed imagery such as NAIP (e.g., [4,30]). Moreover, with the relative abundance of free remotely sensed data and newer software designed to facilitate these types of analyses (e.g., [51]), managers can estimate characteristics of the forests they manage with a greater level of detail and accuracy than was previously possible.

6. Conclusions

In this study we compared multiple estimators for a variety of response and predictor variable relationships, Gaussian error distributions, amounts of noise, and sample designs. Our findings indicated that balance and spread are important aspects of a sample, estimates of pixel-level and population totals can be improved by incorporating models, and spreading samples within feature space can improve estimates. Likewise, for those same samples, when the relationship between response and predictor variables is misspecified, missing key information, or is extremely noisy, expansion and model-based estimates of the population converge, suggesting that with regards to estimation, nothing is lost by using ancillary data. Moreover, for mapping endeavors attempting to spatially depict various characteristics of a landscape, samples used to calibrate predictive models should aim to spread and balance observational units across feature space.

Author Contributions: Conceptualization, J.H. and D.L.R.A.; data curation, J.H.; formal analysis, J.H.; investigation, J.H. and D.L.R.A.; methodology, J.H. and D.L.R.A.; software, J.H.; supervision, D.L.R.A.; validation, J.H.; visualization, J.H.; writing—original draft, J.H.; writing—review and editing, J.H. and D.L.R.A. Both authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Gulf Coast Ecosystem Restoration Council (RESTORE Council) through an interagency agreement with the U.S. Department of Agriculture (USDA) Forest Service (17-IA-11083150–001) for the Apalachicola Tate's Hell Strategy, and by the U.S. Forest Service, Southern Region, and Rocky Mountain Research Station. D.A.'s contributions were supported via agreements with the USDA Rocky Mountain Research Station (15-CS-11221636–199 and 20-JV-11221636–110).

Institutional Review Board Statement: This study does not involve human subjects, human material, human tissues, or human data.

Informed Consent Statement: This study does not involve human subjects, human material, human tissues, or human data.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thompson, M.P.; Wei, Y.; Calkin, D.E.; O'connor, C.D.; Dunn, C.J.; Anderson, N.M.; Hogland, J.S. Risk Management and Analytics in Wildfire Response. *Curr. For. Rep.* **2019**, *5*, 226–239. [CrossRef]
2. Hogland, J.; Dunn, C.J.; Johnston, J.D. 21st Century Planning Techniques for Creating Frie-Resilient forest in the American West. *Forests* **2021**, *12*, 1084. [CrossRef]
3. Hogland, J.; Affleck, D.L.R.; Anderson, N.; Seielstad, C.; Dobrowski, S.; Graham, J.; Smith, R. Estimating Forest Characteristics for Longleaf Pine Restoration Using Normalized Remotely Sensed Imagery in Florida USA. *Forests* **2020**, *11*, 426. [CrossRef]
4. McRoberts, R. Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sens. Environ.* **2010**, *114*, 1017–1025. [CrossRef]
5. McRoberts, R. Satellite image-based maps: Scientific inference or pretty pictures? *Remote Sens. Environ.* **2011**, *115*, 715–724. [CrossRef]
6. Grafström, A.; Shao, X.; Nylander, M.; Petersson, H. A new sampling strategy for forest inventories applied to the temporary clusters of the Swedish national forest inventory. *Can. J. For. Res.* **2017**, *47*, 1161–1167. [CrossRef]
7. Gregoire, T.; Valentine, H. *Sampling Strategies for Natural Resources and the Environment*; Chapman & Hall: Boca Raton, FL, USA; London, UK; New York, NY, USA, 2008; 474p.
8. Gregoire, T. Design-based and model-based inference in survey sampling: Appreciating the difference. *Can. J. For. Res.* **1998**, *28*, 1429–1447. [CrossRef]
9. Stehman, S.; Czaplewski, R. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sens. Environ.* **1998**, *64*, 331–344. [CrossRef]
10. Crowson, M.; Hagensieker, R.; Waske, B. Mapping Land cover change in northern Brazil with limited training data. *Int. J. Appl. Eaarth Obs. GeoInf.* **2019**, *78*, 202–214. [CrossRef]
11. Foody, G.; Mathur, A. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sens. Environ.* **2006**, *103*, 179–189. [CrossRef]

12. Xie, Y.; Lark, T.; Brown, J.F.; Gibbs, H. Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2019**, *155*, 136–149. [CrossRef]
13. Houborg, R.; McCabe, M. A hybrid training approach for leaf area index estimation via Cubist and random forests. *Machine-learning* **2018**, *135*, 173–188. [CrossRef]
14. Kavzoglu, T. Increasing the accuracy of neural network classifications using refined training data. *Environ. Model. Softw.* **2009**, *24*, 850–858. [CrossRef]
15. Lesparre, J.; Gorte, B. Using mixed pixels for the training of a maximum likelihood classification. *Int. Soc. Photogramm. Remote Sens.* **2006**, *36*, 6.
16. Stehman, S. Basic probability sampling designs for thematic map accuracy assessment. *Remote Sens.* **1999**, *20*, 2423–2441. [CrossRef]
17. Comber, A.; Fisher, P.; Brunsdon, C.; Khmag, A. Spatial analysis of remote sensing image classification accuracy. *Remote Sens. Environ.* **2012**, *127*, 237–246. [CrossRef]
18. Stehman, S. Model-assisted estimation as a unifying framework for estimating the area of land cover and land-cover change from remote sensing. *Remote Sens. Environ.* **2009**, *113*, 2455–2462. [CrossRef]
19. Stehman, S. Sampling designs for accuracy assessment of land cover. *Int. J. Remote. Sens.* **2009**, *3*, 5243–5272. [CrossRef]
20. Tille, Y.; Wilhelm, M. Probability Sampling Designs: Principles for Choice of Design and Balancing. *Stat. Sci.* **2017**, *32*, 176–189. [CrossRef]
21. Stevens, D.L.; Olsen, A.R. Spatially-balanced sampling of natural resources. *J. Am. Stat. Assoc.* **2004**, *99*, 262–277. [CrossRef]
22. Grafstrmöm, A.; Lundström, N.L.P. Why well spread probability samples are balanced. *Open J. Stat.* **2013**, *3*, 36–41. [CrossRef]
23. Edwards, T.; Cutler, D.; Zimmermann, N.E.; Geiser, L.; Moisen, G.G. Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecol. Modeling* **2006**, *199*, 132–141. [CrossRef]
24. Brus, D.J. Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma* **2019**, *338*, 464–480. [CrossRef]
25. McRoberts, R.E.; Walter, B. Statistical inference for remote sensing-based estimates of net deforestation. *Remote Sens. Environ.* **2012**, *124*, 394–401. [CrossRef]
26. Sarndal, C.E. Design-based and Model-based Inference in Survey Sampling. *Scand. J. Statist.* **1978**, *5*, 27–52.
27. Shi, Y.; Cameron, J.; Heckathorn, D.D. Model-Based and Design Based Inference: Reducing Bias Due to Differential Recruitment in Respondent-Driven Sampling. *Sociol. Methods Res.* **2016**, *48*, 13–33. [CrossRef]
28. Sterba, S. Alternative Model_Based and Design-Based Frameworks for Inference From Samples to Populations: From Polarization to Integration. *Multivar. Behav. Res.* **2009**, *44*, 711–740. [CrossRef]
29. Rao, J.N.K.; Molina, I. *Small Area Estimation*, 2nd ed.; Wiley and Sons: Hoboken, NJ, USA, 2015; p. 441.
30. Hogland, J.; Anderson, N.; St. Peters, J.; Drake, J.; Medley, P. Mapping Forest Characteristics at Fine Resolution across Large Landscapes of the Southeastern United States Using NAIP Imagery and FIA Field Plot Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 140. [CrossRef]
31. Hogland, J.; Anderson, N.; Affleck, D.L.R.; St. Peter, J. Using Forest Inventory Data with Landsat 8 imagery to Map Longleaf Pine Forest Characteristics in Georgia, USA. *Remote Sens.* **2019**, *11*, 1803. [CrossRef]
32. Neter, J.; Kutner, M.; Nachtsheim, C.; Wasserman, W. *Applied Linear Statistical Models*, 4th ed.; McGraw Hill: Boston, MA, USA, 1996; p. 1408.
33. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer Science and Business Media, LLC.: Singapore, 2006; p. 738.
34. Wood, S.N.; Augustin, N.H. GAMs with integrated model selection using penalized regression splines and applications to environmental modeling. *Ecol. Modeling* **2002**, *157*, 157–177. [CrossRef]
35. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
36. National Agriculture Imagery Program [NAIP]. National Agriculture Imagery Program (NAIP) Information Sheet. 2012. Available online: http://www.fsa.usda.gov/Internet/FSA_File/naip_info_sheet_2013.pdf (accessed on 14 May 2014).
37. Moran, P. Notes on Continuous Stochastic Phenomena. *Biometrika* **1950**, *37*, 17–23. [CrossRef] [PubMed]
38. Kincaid, T.M.; Olsen, A.R. Spsurvey: Spatial Survey Design and Analysis. R Package Version 4.1.0 2019. Available online: https://cran.r-project.org/web/packages/spsurvey/index.html (accessed on 27 September 2021).
39. TIGER 2017. TIGER/Line Shapefiles (Machine Readable Data Files)/Prepared by the U.S. Census Bureau. Available online: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html (accessed on 27 April 2019).
40. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014. Available online: http://www.R-project.org/ (accessed on 28 April 2018).
41. Wood, S.N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. (B)* **2011**, *73*, 3–36. [CrossRef]
42. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab—An S4 Package for Kernel Methods. *R. J. Stat. Softw.* **2004**, *11*, 1–20.
43. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; p. 495.
44. Liaw, A.; Wiener, M. Classification and Regression by random forest. *R News* **2002**, *2*, 18–22.
45. Johnson, R.; Wichern, D. *Applied Multivariate Statistical Analysis*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2002; p. 767.
46. Kermorvant, C.; D'Amico, F.; Bru, N.; Caill-Milly, N.; Robertson, B. Spatially balanced sampling designs for environmental surveys. *Environ. Monit. Assess.* **2019**, *191*, 524–531. [CrossRef]

47. Opsomer, J.D.; Breidt, F.J.; Moisen, G.G.; Kauermann, G. Model-Assisted Estimation of Forest Resources With Generalized Additive Models. *J. Am. Stat. Assoc.* **2007**, *102*, 400–416. [CrossRef]
48. Breidt, F.J.; Opsomer, J.E. Model-Assisted Survey Estimation with Modern Prediction Techniques. *Stat. Sci.* **2017**, *32*, 190–205. [CrossRef]
49. Opitz, D.; Maclin, R. Popular ensemble methods: An empirical study. *JAIR* **1999**, *11*, 169–198. [CrossRef]
50. Breiman, L. Stacked regressions. *Mach. Learn.* **1999**, *24*, 49–64. [CrossRef]
51. Hogland, J.; Anderson, N. Function Modeling Improves the Efficiency of Spatial Modeling Using Big Data from Remote Sensing. *Big Data Cogn. Comput.* **2017**, *1*, 3. [CrossRef]