Forest Ecosystems 11 (2024) 100164

Contents lists available at ScienceDirect

Forest Ecosystems

journal homepage: www.keaipublishing.com/cn/journals/forest-ecosystems

Why ecosystem characteristics predicted from remotely sensed data are unbiased and biased at the same time – and how this affects applications

Göran Ståhl [a,*], Terje Gobakken [b], Svetlana Saarela [b], Henrik J. Persson [a], Magnus Ekström [a], Sean P. Healey [c], Zhiqiang Yang [c], Johan Holmgren [a], Eva Lindberg [a], Kenneth Nyström [a], Emanuele Papucci [a], Patrik Ulvdal [a], Hans Ole Ørka [b], Erik Næsset [b], Zhengyang Hou [d], Håkan Olsson [a], Ronald E. McRoberts [e]

[a] Department of Forest Resource Management, Swedish University of Agricultural Sciences, 901 83, Umeå, Sweden
[b] Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway
[c] USDA Forest Service, Rocky Mountain Research Station, Ogden, UT, USA
[d] The Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, Beijing, 100083, China
[e] Department of Forest Resources, University of Minnesota, St. Paul, MN, USA

ARTICLE INFO

Keywords
Bias
Model-based inference
Design-based inference

ABSTRACT

Remotely sensed data are frequently used for predicting and mapping ecosystem characteristics, and spatially explicit wall-to-wall information is sometimes proposed as the best possible source of information for decision-making. However, wall-to-wall information typically relies on model-based prediction, and several features of model-based prediction should be understood before extensively relying on this type of information. One such feature is that model-based predictors can be considered both unbiased and biased at the same time, which has important implications in several areas of application. In this discussion paper, we first describe the conventional model-unbiasedness paradigm that underpins most prediction techniques using remotely sensed (or other) auxiliary data. From this point of view, model-based predictors are typically unbiased. Secondly, we show that for specific domains, identified based on their true values, the same model-based predictors can be considered biased, and sometimes severely so.

We suggest distinguishing between *conventional model-bias*, defined in the statistical literature as the difference between the expected value of a predictor and the expected value of the quantity being predicted, and *design-bias of model-based estimators*, defined as the difference between the expected value of a model-based estimator and the true value of the quantity being predicted. We show that model-based estimators (or predictors) are typically design-biased, and that there is a trend in the design-bias from overestimating small true values to underestimating large true values. Further, we give examples of applications where this is important to acknowledge and to potentially make adjustments to correct for the design-bias trend. We argue that relying entirely on conventional model-unbiasedness may lead to mistakes in several areas of application that use predictions from remotely sensed data.

1. Introduction

Remotely sensed (RS) data are widely used in ecosystem surveys for predicting and mapping characteristics of interest (e.g., Tomppo et al., 2008; Heckel et al., 2020). A standard procedure is, first, to acquire data from field plots to serve as reference data. Metrics from the RS sensor are then derived for the same plots to establish a dataset from which prediction models are specified and estimated (e.g., McRoberts et al., 2015). The models may be used either for classification or for predicting continuous variables, such as biomass. In this article, we focus on prediction of continuous variables. Several statistical methods are available for developing prediction models. Standard methods include parametric models, e.g. linear and non-linear regression models, and non-parametric models, such as random forests and k-nearest neighbour imputation (e.g., Penner et al., 2013). All of them share the common feature that when properly applied, they provide approximately unbiased predictions

* Corresponding author. Swedish University of Agricultural Sciences, 901 83, Umeå, Sweden.
E-mail address: goran.stahl@slu.se (G. Ståhl).

https://doi.org/10.1016/j.fecs.2023.100164
Received 2 August 2023; Received in revised form 21 November 2023; Accepted 23 December 2023

conditional on the RS metrics used as explanatory variables.

However, bias is a concept with several flavours. First, we may distinguish between design-bias and model-bias, relating to which statistical inference framework is applied. In design-based inference (e.g., Gregoire, 1998; Arnab, 2017) an estimator of a population quantity (such as the total biomass in a study area) is unbiased if its expected value coincides with the true value, which is fixed but unknown. In model-based inference (ibid.), model-unbiasedness occurs when the expected value of a predictor of a population quantity coincides with the expected value of the quantity; in model-based inference, this quantity is assumed to be a random variable with properties determined by a model known as a super-population model (e.g., Cassel et al., 1977). It should thus be noted that model-bias and design-bias are two different concepts. Further, note that we use the terms *estimation* and *estimator* when the target quantity is fixed (e.g. the population parameters of interest in design-based inference), and *prediction* and *predictor* when the target quantity is a random variable (e.g. the population random quantities of interest in model-based inference), following the conventional use of these terms in statistical literature. Moreover, we use the term bias in its strict sense, as a property of an estimator or predictor, whereas in general language the term is often used in many other contexts.

Second, we need to acknowledge that certain conditions may be required for an estimator or predictor to be unbiased. In model-based inference, to which standard regression analysis and many similar techniques belong, predictors are model-unbiased conditional on the input explanatory variables (if the model is correctly specified and the parameters have been estimated using unbiased estimators). However, if we alternatively assess the performance of predictors conditional on some arbitrary set of true values, we may reach a different conclusion. For example, the mean biomass value of the 10% of plots with the largest biomasses from a forest inventory typically deviates substantially from the mean of the biomass predictions for these plots. Thus, conditional on the true state for some set of population units, model-based predictors can be conceived of as biased (which will be discussed and defined more thoroughly later on in this article). This is often noted as a tendency towards the mean for model-based predictions, i.e., small true values are over-predicted whereas large true values are under-predicted, leading to less empirical variability among predicted values than among observed values.

The term *regression fallacy* has been coined for situations where analysts fail to acknowledge the random nature of the response variable (e.g., Quah, 1993). For example, medical studies that select patients for treatment with the largest observed levels of some bodily substance that varies rapidly across time and observe smaller levels after treatment must be cautious about concluding that the treatment has an effect. The reason is that the levels of the substance could have decreased even without treatment, because it might have been patients with temporarily greater levels that were selected for the treatment (e.g., Barnett et al., 2005). In this discussion article, we suggest that another kind of regression fallacy is also important to understand. This fallacy is related to failure to acknowledge that for populations that remain more or less stable across a long period of time (such as forests), model-based predictors may systematically under- or over-predict the true values for certain population units, although the predictors are model-unbiased. This kind of "inverse" fallacy is rarely discussed but leads to problems in several applications.

Many users of regression analysis have been surprised to learn that when graphing residuals versus true values of the response variable a strong trend is typically observed across the range of true values, although the predictor is model-unbiased. In some areas of application, such as chemistry, calibration methods are applied which remove this trend (e.g., Shukla, 1972; Tellinghuisen, 2000). In other areas, researchers and practitioners rely on the model-unbiasedness property of model-based predictors, deliberately or by convention. Remote sensing of forest ecosystems is an area of the latter kind.

In this discussion article, we first describe and illustrate different bias concepts. Then, we show how different scientifically valid perspectives lead to different conclusions regarding whether model-based predictors (or estimators) are unbiased. Further, we discuss how potential bias adversely affects the results in several areas of application, and we briefly discuss different methods available for correcting for the bias.

2. Bias formalism

Loosely speaking, bias occurs when the expected value of an estimator or predictor does not coincide with the true value of the quantity of interest. More specifically, for a fixed target quantity, Y, an estimator $\widehat{Y}$ is unbiased if its expected value (over all possible samples) coincides with the fixed true value, i.e. if $E(\widehat{Y}) = Y$. If we are instead interested in predicting a random quantity, the value of which varies depending on the realisation of some random process, the convention (e.g., Cassel et al., 1977) is to define unbiasedness as the case where the expected value of a predictor $\widehat{Y}$ coincides with the expected value of Y, i.e. $E(\widehat{Y}) = E(Y)$. In this case, we formally evaluate the expectations over an infinite number of realisations of the random process, normally conditional on a given sample that does not have to be a probability sample.

In the following, we first give an overview of bias in model-based inference. Secondly, we move to design-based inference and suggest how methods similar to model-based prediction can be applied under this framework.

2.1. Model-based inference

In model-based inference (e.g., Cassel et al., 1977; Gregoire, 1998; Arnab, 2017) our target population is considered a realisation from a model known as a superpopulation model, indicating that it has the capacity to create a variety of realised populations that all conform with the specified model assumptions.

The state of our target population at a given time point is a single realisation from the superpopulation model. Model-based inference treats the generated values as random (e.g., Chambers and Clark, 2012) until they are observed. This means that population quantities of interest, such as means and totals, are also random variables because in practice we will rarely make observations on all units in a population. This has implications for the definition of bias in model-based inference, because there is no single fixed population value to compare with. Instead, in model-based inference the expected value of a predictor is compared with the expected value of the quantity being predicted (which may be related to an individual population unit, a certain domain, or the entire population). Model-unbiasedness occurs when the expected value of the predictor coincides with the expected value of the random quantity being predicted (e.g., Cassel et al., 1977; Thompson, 2012). That is, unbiasedness is a property that we, theoretically, assess in relation to an infinite number of realisations from the superpopulation model.

Model-based prediction is typically conducted under the umbrella of model-based inference. To further demonstrate and illustrate the concepts, we assume a simple linear superpopulation model between some RS metric, X (assumed to be fixed in this article), and plot level biomass, Y. Thus, the model is $Y = \alpha_0 + \alpha_1 X + \epsilon$. Besides specifying the model form, we also need to specify the properties of the error terms, ϵ. In this article, for simplicity, we assume that these terms are independent, normally distributed, with constant variance, and that they have zero expectation. However, in many real-world cases, the variance of the error terms depends on X, and spatial autocorrelation between them is often present.

The model parameters can be estimated in several ways, e.g. through ordinary least squares regression analysis or maximum likelihood methods. We thus obtain an estimated model, $\widehat{Y} = \widehat{\alpha}_0 + \widehat{\alpha}_1 X$. Under the given assumptions, the parameter estimates will be unbiased and thus it can be formally shown that, for any given X, $E(\widehat{Y}|X) = \alpha_0 + \alpha_1 X = E(Y|X)$. Thus, the expected value of the predictor coincides with the expected value of Y, which implies model-unbiasedness. In this case, we

addressed prediction at the level of individual population units. By aggregating across population units, we can show that model-unbiasedness holds also for population totals and means in case it holds for all individual units. A straightforward predictor of the (random) population total $\sum_{i=1}^{N} Y_i$, for an area tessellated into N grid-cells, would be $\sum_{i=1}^{N} \widehat{Y}_i$. Because the predictor for each population unit is model-unbiased it follows that $E(\sum_{i=1}^{N} \widehat{Y}_i) = E(\sum_{i=1}^{N} Y_i)$, i.e. the predictor of the population total is also model-unbiased.[1]

In this article, we denote the above kind of unbiasedness as *conventional model-unbiasedness* to separate it from another kind of model-related unbiasedness, which will be introduced later. In summary, conventional model-unbiasedness means that on average across an infinite number of realisations from a superpopulation model, the mean of our predictions coincides with the average realised value of the predicted quantity.

The principle of conventional model-unbiasedness is illustrated in Fig. 1, which is based on a large number of realisations from a superpopulation model, out of which a limited number of realisations are displayed with different colours. That is, in the same figure different realised true values exist for each population unit. However, the general patterns that we observe would remain the same for a single realisation of true population values or for a large sample from some population.

The upper left part of Fig. 1 (a) shows how the explanatory variable (the variable X in our model) is related to the realised true values in different hypothetical populations. The red line in Fig. 1 (a) is the superpopulation model, without the error term. The model we would estimate based on a sample from any realised population would approximate the superpopulation model, if the regression model is correctly specified. Thus, intuitively, the red line in Fig. 1 (a) also approximates the regression line from a single sample, passing through the mean value of the realised population values at all levels of the explanatory variable.

The upper right part of Fig. 1 (b) illustrates that model-based predictions are (conventional) model-unbiased, because for all levels for the expectation of predicted values (on the horizontal axis) the mean value of the error terms is zero (the dashed line). That is, the expected value of the predictions coincides with the expectation of the true value. This corresponds to checking graphs of residuals, which are predictions of error terms, versus predicted values in case regression analysis is conducted using a sample from a single realisation of true values.

The lower left part of Fig. 1 (c) shows a scatterplot of realised true values versus expectations of predicted values. From this graph, it can be seen that the range of predicted values is narrower than the range of true values, i.e. model-based prediction leads to less variability among predicted values compared to the variability among true values. This can be observed as the well-known tendency that the predictions are closer to the mean than the true values (e.g., Galton, 1886) (The dashed line displays a 1:1 relationship).

Finally, in the lower right part of Fig. 1 (d) the true values are graphed versus the error term values. Here, a trend can be observed that the average value of the error terms deviates substantially from zero in case the true values are either large or small. Intuitively, this appears to indicate that model-based predictions are biased, although we have shown that they are in fact model-unbiased (at the level of individual population units, as well as for domains, totals and means).

Scatterplots like the one in the lower right part of Fig. 1 (d) frequently cause confusion among remote sensing researchers and practitioners, when graphing residuals versus true values. How can the predictors be model-unbiased although we observe a strong trend for the residuals across the range of true values? Studies like the ones by Barth et al. (2012), Gilichinsky et al. (2012) and Lindgren et al. (2022) have proposed methods to adjust for this apparent "bias" of predictors based on RS data.

In this article, we suggest that a core issue of concern is whether we apply model-based inference for populations with realised values of the response variable that vary quickly across short periods of time (i.e. we frequently obtain new realisations from the superpopulation model) or if we apply it for populations that remain more or less stable across long periods of time, such as many forests (i.e. a given realisation from the superpopulation model remains the same for a longer time). In the first case, if we repeat a survey at several occasions across time, the average of our predictions for a given population unit would tend to coincide with the average of the realised true values. In this case it is intuitively straightforward to accept the model-based predictions as the "best" predictions we can obtain at the level of individual population units. However, in the latter case our model-based predictors would repeatedly over- or under-predict the realised true values for many population units. Whenever we address a certain population unit with a new prediction using a certain type of RS data, we are likely to obtain a similar systematic error. Indeed, it is commonly observed that small true values of biomass or growing stock volume tend to be systematically over-predicted and large true values under-predicted (e.g., Gilichinsky et al., 2012; Ehlers et al., 2018; Persson and Ståhl, 2020). This occurs although the predictors are model-unbiased in the conventional sense.

We argue that using methods and definitions that are valid for conventional model-based inference could be questioned in case we are addressing populations that remain stable across time. For many applications based on predictions from RS data, we need to acknowledge that systematic over- or under-prediction occurs and that this might affect the usefulness of the RS-based predictions. In the next section, we propose an alternative view to model-based assessment where we assume that the realised population is fixed. We suggest that framing the problem in this way, i.e. in a design-based context, leads to results that might be more straightforward and relevant to practitioners.

2.2. Use of models in design-based inference

In design-based inference (e.g., Gregoire, 1998; Arnab, 2017), probability samples from populations are selected and the observations for these units, as well as the population quantities of interest, such as the population mean or total, are treated as fixed values. Estimators are specified for estimating the population quantities of interest; the estimators are random variables because the input to them emanates from randomly selected samples. For illustration and comparison with model-based inference, we assume that the study area has been tessellated into N units. With some random sampling design (without replacement), n units are randomly selected for the design-based inference. A Horvitz-Thompson estimator (e.g., Wu and Thompson, 2020) of the population total is $\widehat{\tau} = \sum_{i=1}^{n} y_i/\pi_i$, where y_i is the value for the i:th unit[2] and π_i is the unit's inclusion probability. The expected value of the estimator is $E(\widehat{\tau}) = E(\sum_{i=1}^{n} y_i/\pi_i) = E(\sum_{i=1}^{N} I_i y_i/\pi_i) = \sum_{i=1}^{N} E(I_i) y_i/\pi_i = \sum_{i=1}^{N} y_i$; because the expected value coincides with the true value the estimator is unbiased. In the derivation, a random inclusion indicator, I_i, which takes the value 1 if the unit is included in the sample and 0 otherwise, is used; the expected value of the indicator coincides with the inclusion probability, i.e. $E(I_i) = \pi_i$. In design-based inference, unbiasedness is a property that we, theoretically, assess in relation to all possible samples under the given design. Note that in this case the values

[1] However, in large-area surveys based on RS data model-based predictors are rarely evaluated for bias, because it would involve substantial efforts to collect field reference data from all parts of the target population to check if the assumed model is valid. Many studies indicate that model-bias arises due to incorrectly specified models or lack of field data for model fitting from some ecosystem types (e.g. Réjou-Méchain et al., 2019). Thus, model-unbiasedness is a theoretical concept that is typically difficult to verify in practical surveys.

[2] Note that, for simplicity, we use the same notation for the ith value in the sample and the ith value in the population.

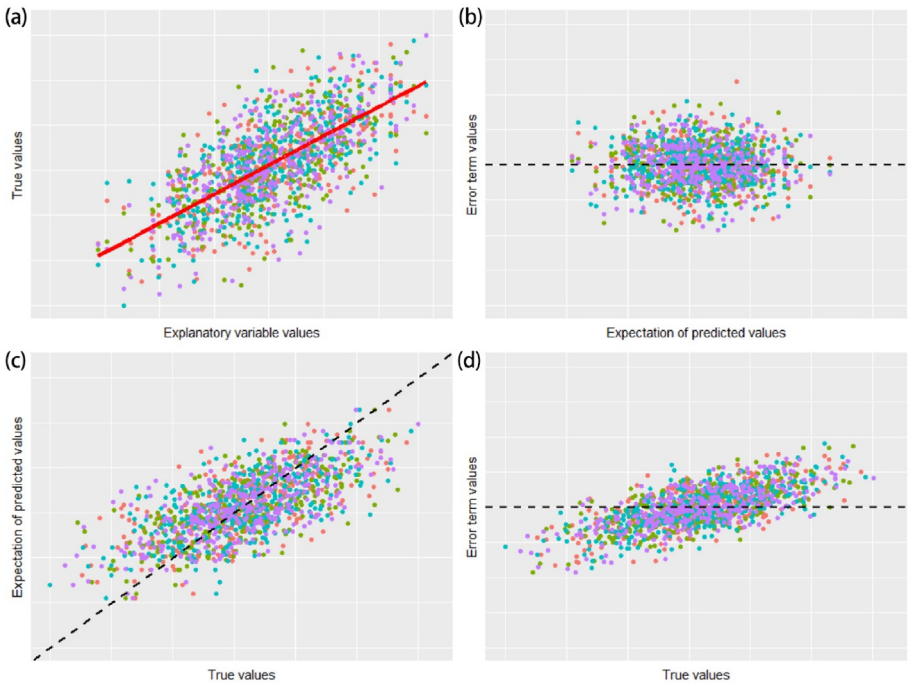

Fig. 1. An illustration of different characteristics of model-unbiasedness. In (a), the explanatory variable values are graphed versus realised true values (the red line is the superpopulation model, without the error terms). In (b), the expectations of predicted values are graphed versus the error terms. In (c), true values are graphed versus expectations of predicted values (the dashed line is the 1:1-line). In (d), true values are graphed versus the error terms. The different colours display different realised populations. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

of the variables of interest for individual population elements are treated as fixed, but unknown unless the population elements are included in the sample.

A standard method of applying models in design-based inference is to use model-assisted estimators (Särndal et al., 2003). With such estimators, predicted values for all units in the population (or for a large sample of auxiliary data) are obtained using models. A probability sample is used for estimating the total (or mean) of the deviations between model predictions and observed values. This estimate is added to the total (or mean) of the model-based predictions for all population units to obtain an approximately design-unbiased estimator of the population quantity of interest. Several studies have demonstrated the usefulness of model-assisted estimation based on combinations of RS and field data (e.g., Andersen et al., 2011; Gregoire et al., 2011; Næsset et al., 2011). The assisting models can be developed in different ways, and the procedure assures approximate design-unbiasedness of estimators although the models applied may be incorrectly specified or "biased" in other ways.

However, we will not address model-assisted estimation further but instead a model-based estimation counterpart to the model-based prediction discussed in the previous section. We have argued that assessing properties of predictors across a large number of random realisations from a superpopulation model may be conceived of as non-intuitive for populations that remain stable across time. Thus, we now instead adopt a design-based view to the problem and conceptualize that all target quantities (such as plot or grid-cell level biomasses) are fixed but unknown, just as in the case of standard design-based inference. We argue that, compared to the assumptions underpinning model-based inference, practitioners would more easily understand this assumption and it would be more useful in applications where they are interested in properties of the single population confronting them rather than properties of hypothetical realisations from a superpopulation model.

Standard regression analysis (and similar prediction techniques) is based on the same assumptions as those underpinning model-based inference. For fixed populations, design-based methods for estimating the parameters of models of similar kind as those used in model-based inference are available (e.g., Heeringa et al., 2017). Alternatively, linear relationships between explanatory and response variables can be estimated fully from the design (e.g., Särndal et al., 2003). With simple random sampling from a population, following the methods proposed by Heeringa et al. (2017), the design-based parameter estimators are similar to those used in standard regression analysis. Thus, for the purpose of this discussion article, we conclude that models of similar kind as those previously discussed can be estimated and applied also when we frame our problem in a design-based context. However, because the quantities we assess are now fixed, the proper nomenclature would be *model-based estimators* rather than model-based predictors. Although we change the terminology, the models used for assessing the values of the variables of interest remain more or less the same as before.

A major difference compared to the case of model-based inference is that we treat the true values (as well as totals and means) for the population units as fixed when we assess whether or not an estimator is biased. *Design-unbiasedness of a model-based estimator* occurs when the expectation of an estimator $\widehat{Y}$ coincides with the fixed true value, i.e. if $E(\widehat{Y}) = Y$, when evaluated over all possible samples (collected for estimating the model parameters). Whereas our focus is estimators for the values of the variables of interest for individual population element, the definition holds in general for estimators of any other population parameter, such as the population mean and total.

Returning to Fig. 1, we would observe similar patterns in the case of model-based estimators of the response variable for individual population units, since the models would be developed in a similar way as in model-based inference. However, in the design-based case, all estimators with expected values which do not coincide with the true value are formally biased.

Fig. 2 shows the results of a design-based approach to estimating the value of the variable of interest for each unit in the population (corresponding to Fig. 1). In this case, however, (i) a simple random sample (without replacement) of 50 units was selected; (ii) the model parameters were estimated based on the sample, and (iii) the estimated model was applied for estimating the value of the variable of interest for all population units. This procedure was repeated 500,000 times (for a single realisation from the superpopulation model), to obtain mean values for each population element that would be good approximations of the corresponding expected values.

Fig. 2 (a) shows the same general pattern as Fig. 1 (c), i.e. that the range of estimated values is narrower than the range of true values. Fig. 2

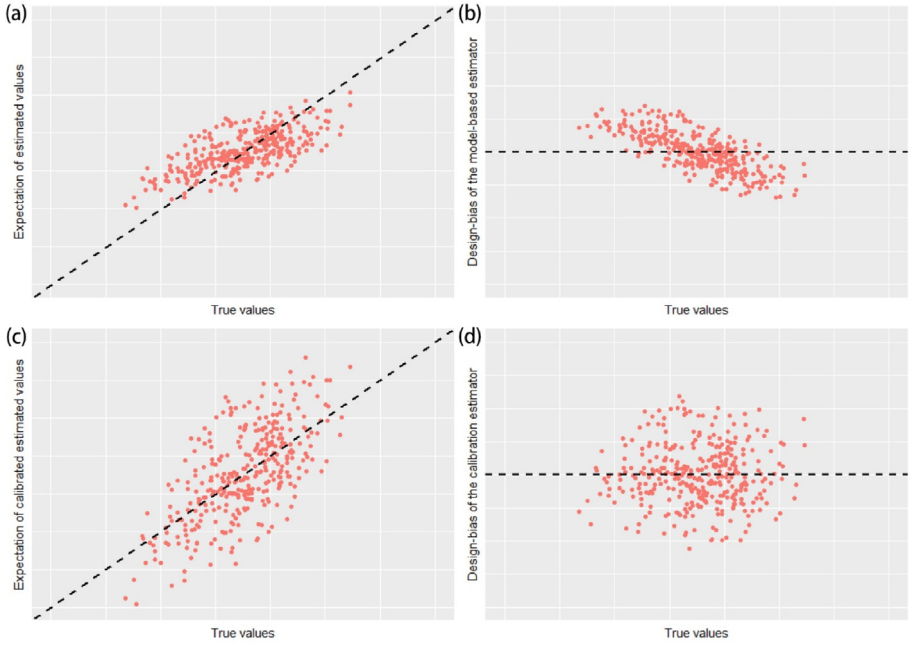

Fig. 2. Model-based estimation at the level of individual population units and the effects of calibration. In (a) true values are graphed versus the expectation of estimated values; in (b) the design-bias of standard model-based estimators is shown; in (c) true values are graphed versus expectations of calibrated estimated values; in (d) the design-bias of the calibration estimator is shown.

(b) shows that most estimators of the value of the variable of interest at the level of population units are design-biased. Thus, standard regression techniques typically result in design-biased estimates at the level of individual population units. Especially, there is a problematic trend from positive bias for small true values to negative bias for large true values (Fig. 2b).

As suggested by Tian et al. (2016) and Persson and Ståhl (2020), the simple linear model $\tilde{Y} = \beta_0 + \beta_1 Y$ may be used for characterising properties of estimates or predictions based on RS data. In this model, $\tilde{Y}$ is the estimate obtained for an individual population unit from a model-based estimator, and the βs are model parameters. Following principles for calibration developed in chemometrics (e.g., Shukla, 1972; Tellinghuisen, 2000), this model can be rearranged to obtain a calibration estimator that removes the *trend* of the bias for standard model-based estimators (although estimates for individual units remain biased). A simple classical calibration estimator is $\hat{Y}_{\text{cal}} = \frac{\tilde{Y} - \hat{\beta}_0}{\hat{\beta}_1}$. It has previously been applied to model-based predictions from RS data by Lindgren et al. (2022).

Fig. 2 (c) and (d) show the effects of applying the calibration estimator. Fig. 2 (c) shows that it increases the range of calibrated values, which tends to be similar to the range of true values. Fig. 2 (d) shows that calibration removes the bias trend. It is important to note that through calibration we cannot make estimators at the level of individual population elements design-unbiased, but we can remove the trend of the design-bias.

Further, based on the simple linear models presented earlier in the article, it is possible to deduce what would be the linear relationship between the explanatory variable and the response variable, both for the case of a standard linear regression model and for a calibration model. In Fig. 3, we demonstrate the difference between the two models for two cases, which differ in terms of the correlation between the explanatory variable and the response variable. It can be seen that in the case of a 0.95 correlation (Fig. 3a) the difference between the calibration model and the standard regression model is minimal. However, in case of a weaker correlation (0.7; Fig. 3b), the difference between the two models is substantial.

In summary, we argue that a design-based inference perspective rather than a model-based inference perspective would often be more

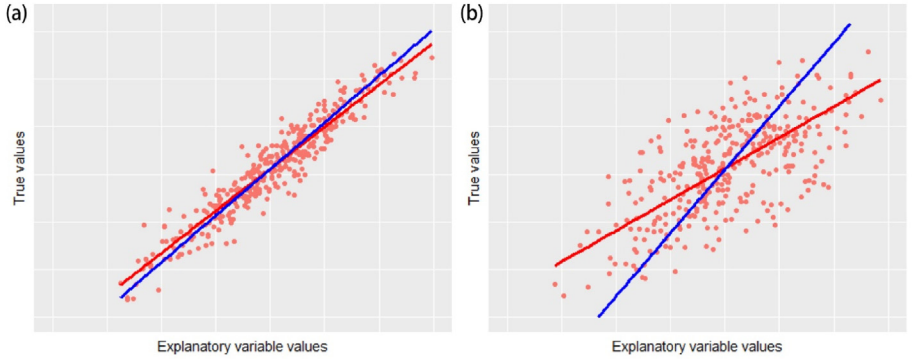

Fig. 3. The standard regression model (red line) and the calibration model (blue line) in case of a 0.95 correlation between the explanatory variable and the response variable (a) and in case of a 0.70 correlation (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

relevant when we model the response variable for individual units in populations that remain more or less stable across long periods of time, such as many forests. In this case, standard methods of model-based prediction may lead to substantial trends of the design-bias across the range of true values. This trend can be very strong in case the correlation between the explanatory variable and the response variable is weak.

3. Consequences for applications

The consequences of design-bias of model-based estimators of response variable values for individual population units differ between different areas of application. In this section, we give a brief overview of the consequences of this type of bias. We restrict the discussion to populations that remain relatively stable across a long period of time, such as forest ecosystems.

3.1. Assessment at the level of individual population units

When the main interest in applications is to make assessments at the level of individual population units, conventional model-based prediction (without calibration) leads to smaller mean square errors (cf., Tellinghuisen, 2000) of predicted values compared to when classical calibration is applied. Thus, we suggest that in this case a straightforward approach would be to apply prediction models obtained through standard regression analysis, or similar techniques, without attempting to remove the bias trend. However, in some cases it might be important to remove this trend, in which case adjustments could be considered.

3.2. Assessment of population totals and means

Model-based prediction of population totals and means has a strong theoretical underpinning in the literature (e.g., Cassel et al., 1977; Chambers and Clark, 2012). Predictors of population totals and means can be shown to be conventional model-unbiased if the models have been correctly specified and estimated. For large populations, the relative difference between the expected value of the quantity being predicted and the realised value is likely to be small (e.g., McRoberts et al., 2018). Thus, in assessing population totals and means we can rely on conventional model-based inference, not least since model-based predictors are likely to be approximately design-unbiased as well. The design-unbiasedness follows from balancing positive bias for predictions of small true values with negative bias for predictions of large true values. As a consequence, it might be prudent to apply sampling schemes which are balanced in the explanatory variable(s), when models are estimated (cf., Chambers and Clark, 2012).

However, it should be noted that model-based prediction of population totals and means is sensitive to conventional model-bias, due to incorrectly specified models or lack of data from certain subpopulations (cf., Réjou-Méchain et al., 2019).

3.3. Assessment for domains

Although conventional model-based inference is model-unbiased also for domains (e.g., Chambers and Clark, 2012), domain estimators based on conventional model-based inference would typically be design-biased. For example, if the domain of interest is the 10% of a forest region with the largest biomass density, the design-bias of a model-based predictor for this domain is likely to be substantial unless corrections are made. Thus, our suggestion is to consider calibration or similar adjustment if the interest is assessment for domains. However, note that many other techniques are available for domain estimation, using either model-based or design-based inference (e.g., Hou et al., 2022).

3.4. Mapping

Increasingly, national, regional and global maps of ecosystem conditions are produced. One important example is biomass maps, supporting reporting of greenhouse gas fluxes to the UN Convention on Climate Change (e.g., Langner et al., 2014; Dubayah et al., 2022). Although the biomass maps may be estimated using methods that ensure model-unbiasedness, they may turn out to be very different depending on which type of RS data are applied (e.g., Mitchard et al., 2013). With RS data strongly correlated with biomass, the actual variability of biomass across the landscape would be fairly well depicted, whereas if data weakly correlated with biomass have been applied the map would display substantially less variability (e.g., Saarela et al., 2023). Thus, our suggestion is to consider corrections to reduce the design-bias of model-based estimators in case the purpose is to present maps that display the real variability of the feature of interest in the landscape. This conclusion concerns maps displaying continuous rather than categorical variables. However, note that calibration will not remove the bias at the level of individual population elements, only the bias trend.

3.5. Scenario modelling and planning

Scenario modelling is conducted in many countries as a means of assessing the possible outcomes of certain ecosystem management policies, before policy decisions are made (e.g., Mohren, 2003; Lämås et al., 2023). Increasingly, scenario modellers and planners require spatially explicit wall-to-wall data for modelling features that cannot be adequately assessed using sample data. An important example is biodiversity, where habitat models often require spatially explicit data (e.g., Ruckelshaus et al., 1997). Further, a current trend is to integrate modelling of several ecosystem services using the same set of data and to move towards precision management where treatments are allocated to small parcels of land, based on the available information (e.g., Wilhelmsson et al., 2021).

Trends in the design-bias of model-based estimators, or predictors, may be a substantial problem in scenario modelling, because “extreme” values are often the most important ones. For example, old forests with large growing stock volumes are of particular interest, both from the point of view of timber harvesting and biodiversity preservation in forestry scenario modelling. Thus, our suggestion in this case is to consider making modifications to reduce the design-bias trend, before data are entered into scenario modelling or planning systems.

3.6. Data assimilation and composite estimation

In data assimilation, as well as composite estimation, several predictions of the same feature within a geographical area are combined to improve the precision of a composite predictor, which uses all the available information. For example, in data assimilation a series of predictions for a given area will be linked through an updating model, and combined based on the precision of the update and the new prediction (e.g., Ehlers et al., 2013). In this case, design-bias trends pose problems because the magnitude of the bias will differ depending on which RS data source is used for the predictions. As shown by Lindgren et al. (2022), correcting for design-bias trends has the potential to decrease the mean square error of composite data assimilation predictors.

4. Discussion

The main conclusion from this article is that standard predictors (or estimators) of the values of interest for individual population elements, such as pixel-level forest biomass predicted from RS data using regression analysis or machine learning algorithms, are design-biased, although in the mean time they are typically model-unbiased. Potentially, this has negative consequences in several areas of application. Negative consequences mainly occur when the goodness-of-fit of estimated models is intermediate or poor. This can be seen in Fig. 3, where the difference between a non-calibrated and a calibrated estimator (or predictor) is small when there is strong correlation between the explanatory variable

and the response variable, but large when the correlation is weaker. However, we argue that many models for forest resources and ecosystems assessment are of the second kind. Models for biomass assessment based on optical satellite or radar data (e.g., Gao et al., 2018) are important examples.

The interest in forest ecosystem information is currently increasing and wall-to-wall mapping is increasingly being proposed as a strong alternative for acquiring the information. With maps, geographically explicit information is obtained, and totals can be obtained by summing estimates or predictions from individual map elements. Changes can be assessed by comparing maps from different time points (e.g., Hansen et al., 2013) and with wall-to-wall information, forest scenario models can incorporate features that cannot be retrieved from sample data (e.g., Wilhelmsson et al., 2021). In this development, we suggest that it is becoming increasingly important to understand the difference between model-based and design-based inference, and the assumptions underlying each of the concepts.

Model-based inference has a strong theoretical underpinning (e.g., Cassel et al., 1977; Chambers and Clark, 2012) and it is widely applied, deliberately or by tradition, for mapping and formal inference about population totals or means. In this discussion article, we do not question the theoretical soundness of model-based inference or the toolbox of methods linked to it. However, we suggest that for populations that remain more or less stable across time, in many cases it could be more relevant to treat the population quantities of interest as fixed but unknown rather than as random variables. In many applications, we are only interested in the features of the single realisation of the superpopulation model that constitutes the real world. We are less interested in the hypothetical worlds that the superpopulation could have created, but did not. We suggest that for populations and applications of this kind, design-based inference could be a conceptually more appropriate framework than model-based inference. Design-based inference treats the features of the world we face as fixed but unknown. We assess them by collecting random samples where we record the conditions.

As shown in this article, model-based predictors typically are design-biased, and the use of such predictors will typically lead to trends in the design-bias across the range of true values. In case extreme values are of special interest, which they often are, the design-bias of model-based predictors tends to be especially large for such population units.

Interestingly, although model-based prediction techniques are applied, many remote sensing studies evaluate the results in a design-based perspective by collecting a validation sample based on which the properties of the model-based predictions are assessed (e.g., Persson and Ståhl, 2020). A typical conclusion is that small true values are overestimated and large true values are underestimated (e.g., Gao et al., 2018). As previously described in this article, this conclusion should not come as a surprise, because it is an inherent property of model-based prediction when evaluated in a design-based context.

Several studies have addressed adjustment for the bias we propose to be termed design-bias of model-based estimators. Barth et al. (2012) proposed a routine for imputation that ensured that the wall-to-wall imputed values had the same proportions in the imputed population as in the sample. Gilichinsky et al. (2012) adopted histogram matching for achieving a similar result. Lindgren et al. (2022) applied classical calibration for removing the design-bias trend. A related technique is empirical best linear prediction (E-BLUP) which makes adjustments for all units belonging to the same group based on observations of some units in the group (e.g., Breidenbach and Astrup, 2012; Hou et al., 2019). This will also reduce the design-bias, but not remove it.

In chemometrics, for a long time, an important challenge has been to "calibrate" measurement instruments, so that the instrument reading corresponds to the actual concentration of some compound being analysed (e.g., Shukla, 1972; Tellinghuisen, 2000). From a known concentration, Y, of some compound, the instrument provides the response, $\tilde{Y}$. The objective is then to calibrate the instrument so it provides a reading corresponding to Y rather than $\tilde{Y}$. The classical calibration approach applied by Lindgren et al. (2022) to predictions based on RS data emanates from approaches devised in chemometrics.

In this article, we have assumed that a certain RS sensor would repeatedly provide more or less the same response for a certain forest area, leading to the same systematic error being repeatedly committed when estimating the conditions in that area (if a survey is repeated several times). In reality, many factors influence the sensor response and assuming it to be constant is a simplification, e.g. due to seasonal variability (e.g., Wang et al., 2017). However, as shown by Ehlers et al. (2018) for a boreal forest ecosystem, the error correlations between repeated assessments of forest characteristics using a certain sensor are typically very strong, which supports the assumption made in this article.

Lastly, we suggest that there is room for further studies on how to construct model-based estimators for individual population units that have small design-bias, and that lead to minimal trends of the design-bias across the range of true values. Classical calibration and histogram matching appear to be relevant approaches, but there are probably other approaches that could be considered as well, such as quantile regression (e.g., Hao and Naiman, 2007), orthogonal regression, or total least squares regression (cf., Solberg et al., 2010).

5. Conclusions

We have shown that standard regression or machine learning predictors, which are approximately model-unbiased, are at the same time design-biased. For populations that remain more or less stable across long periods, such as forests, this implies that the same type of systematic errors repeatedly will be observed when forest ecosystem characteristics are assessed using data from a certain RS sensor type. This has important negative implications in several areas of application.

The discussion about the relevance of model-based inference in comparison to design-based inference is far from new (e.g. Brewer, 2013; Dumelle et al., 2022). However, conclusions tend to vary depending on area of application. In this article, we have discussed and highlighted important issues to consider in studies applying RS data for assessing forest ecosystem characteristics. In some cases, calibration to remove design-bias trends of model-based predictors should be considered.

CRediT authorship contribution statement

Göran Ståhl: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing, Visualization. **Terje Gobakken:** Conceptualization, Funding acquisition, Project administration, Resources, Writing – review & editing. **Svetlana Saarela:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – review & editing. **Henrik J. Persson:** Conceptualization, Writing – review & editing. **Magnus Ekström:** Conceptualization, Formal analysis, Methodology, Writing – review & editing. **Sean P. Healey:** Conceptualization, Writing – review & editing. **Zhiqiang Yang:** Conceptualization, Writing – review & editing. **Johan Holmgren:** Conceptualization, Funding acquisition, Project administration, Resources, Writing – review & editing. **Eva Lindberg:** Conceptualization, Funding acquisition, Project administration, Writing – review & editing. **Kenneth Nyström:** Conceptualization, Writing – review & editing. **Emanuele Papucci:** Conceptualization, Writing – review & editing. **Patrik Ulvdal:** Conceptualization, Writing – review & editing. **Hans Ole Ørka:** Conceptualization, Writing – review & editing. **Erik Næsset:** Conceptualization, Funding acquisition, Methodology, Project administration, Writing – review & editing. **Zhengyang Hou:** Conceptualization, Methodology, Writing – review & editing. **Håkan Olsson:** Conceptualization, Methodology, Writing – review & editing. **Ronald E. McRoberts:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is part of the programme Mistra Digital Forests and of the Center for Research-based Innovation SmartForest: Bringing Industry 4.0 to the Norwegian forest sector (NFR SFI project no. 309671, smartforest.no).

References

Arnab, R., 2017. Survey Sampling Theory and Applications. Academic Press.
Andersen, H.E., Strunk, J., Temesgen, H., 2011. Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the Upper Tanana Valley of Interior Alaska. West. J. Appl. Finance 26 (4), 157–164.
Barnett, A.G., Van Der Pols, J.C., Dobson, A.J., 2005. Regression to the mean: what it is and how to deal with it. Int. J. Epidemiol. 34 (1), 215–220.
Barth, A., Lind, T., Ståhl, G., 2012. Restricted imputation for improving spatial consistency in landscape level data for forest scenario analysis. For. Ecol. Manag. 272, 61–68.
Breidenbach, J., Astrup, R., 2012. Small area estimation of forest attributes in the Norwegian National Forest Inventory. Eur. J. For. Res. 131, 1255–1267.
Brewer, K., 2013. Three controversies in the history of survey sampling. Surv. Methodol. 39 (2), 249–263.
Cassel, C.M., Särndal, C.E., Wretman, J.H., 1977. Foundations of Inference in Survey Sampling. Wiley, New York.
Chambers, R., Clark, R., 2012. An Introduction to Model-Based Survey Sampling with Applications. Oxford University Press, New York.
Dubayah, R., Armston, J., Healey, S.P., Bruening, J.M., Patterson, P.L., Kellner, J.R., Duncanson, L., Saarela, S., Ståhl, G., Yang, Z., Tang, H., Blair, J.B., Fatoyinbo, L., Goetz, S., Hancock, S., Hansen, M., Hofton, M., Luthcke, S., 2022. GEDI launches a new era of biomass inference from space. Environ. Res. Lett. 17 (9), 095001.
Dumelle, M., Higham, M., Ver Hoef, J.M., Olsen, A.R., Madsen, L., 2022. A comparison of design-based and model-based approaches for finite population spatial sampling and inference. Methods Ecol. Evol. 13 (9), 2018–2029.
Ehlers, S., Grafström, A., Nyström, K., Olsson, H., Ståhl, G., 2013. Data assimilation in stand-level forest inventories. Can. J. For. Res. 43 (12), 1104–1113.
Ehlers, S., Saarela, S., Lindgren, N., Lindberg, E., Nyström, M., Persson, H.J., Olsson, H., Ståhl, G., 2018. Assessing error correlations in remote sensing-based estimates of forest attributes for improved composite estimation. Rem. Sens. 10 (5), 667.
Galton, F., 1886. Regression towards mediocrity in hereditary stature. J. Anthropol. Inst. G. B. Ireland 15, 246–263.
Gao, Y., Lu, D., Li, G., Wang, G., Chen, Q., Liu, L., Li, D., 2018. Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. Rem. Sens. 10 (4), 627.
Gilichinsky, M., Heiskanen, J., Barth, A., Wallerman, J., Egberth, M., Nilsson, M., 2012. Histogram matching for the calibration of k NN stem volume estimates. Int. J. Rem. Sens. 33 (22), 7117–7131.
Gregoire, T.G., 1998. Design-based and model-based inference in survey sampling: appreciating the difference. Can. J. For. Res. 28 (10), 1429–1447.
Gregoire, T.G., Ståhl, G., Næsset, E., Gobakken, T., Nelson, R., Holm, S., 2011. Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. Can. J. For. Res. 41 (1), 83–95.
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J., 2013. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853.
Hao, L., Naiman, D.Q., 2007. Quantile Regression (No. 149). Sage, London.
Heckel, K., Urban, M., Schratz, P., Mahecha, M.D., Schmullius, C., 2020. Predicting forest cover in distinct ecosystems: the potential of multi-source Sentinel-1 and-2 data fusion. Rem. Sens. 12 (2), 302.
Heeringa, S.G., West, B.T., Berglund, P.A., 2017. Applied Survey Data Analysis. CRC press. https://doi.org/10.1201/9781315153278.
Hou, Z., Mehtätalo, L., McRoberts, R.E., Ståhl, G., Tokola, T., Rana, P., Siipilehto, J., Xu, Q., 2019. Remote sensing-assisted data assimilation and simultaneous inference for forest inventory. Remote Sens. Environ. 234, 111431.
Hou, Z., McRoberts, R.E., Zhang, C., Ståhl, G., Zhao, X., Wang, X., Li, B., Xu, Q., 2022. Cross-classes domain inference with network sampling for natural resource inventory. For. Ecosyst. 9, 100029.
Langner, A., Achard, F., Grassi, G., 2014. Can recent pan-tropical biomass maps be used to derive alternative Tier 1 values for reporting REDD+ activities under UNFCCC? Environ. Res. Lett. 9 (12), 124008.
Lindgren, N., Nyström, K., Saarela, S., Olsson, H., Ståhl, G., 2022. Importance of calibration for improving the efficiency of data assimilation for predicting forest characteristics. Rem. Sens. 14 (18), 4627.
Lämås, T., Sängstuvall, L., Öhman, K., Lundström, J., Årevall, J., Holmström, H., Nilsson, L., Nordstrom, E., Wikberg, P., Wikstrom, P., Eggers, J., 2023. The multi-faceted Swedish Heureka forest decision support system: context, functionality, design, and 10 years experiences of its use. Front. For. Glob. Change 6, 1163105.
McRoberts, R.E., Næsset, E., Gobakken, T., Bollandsås, O.M., 2015. Indirect and direct estimation of forest biomass change using forest inventory and airborne laser scanning data. Remote Sens. Environ. 164, 36–42.
McRoberts, R.E., Næsset, E., Gobakken, T., Chirici, G., Condés, S., Hou, Z., Saarela, S., Chen, Q., Ståhl, G., Walters, B.F., 2018. Assessing components of the model-based mean square error estimator for remote sensing assisted forest applications. Can. J. For. Res. 48 (6), 642–649.
Mitchard, E.T., Saatchi, S.S., Baccini, A., Asner, G.P., Goetz, S.J., Harris, N.L., Brown, S., 2013. Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. Carbon Bal. Manag. 8, 1–13.
Mohren, G.M.J., 2003. Large-scale scenario analysis in forest ecology and forest management. For. Pol. Econ. 5 (2), 103–110.
Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G., Weydahl, D., 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: a case study from a boreal forest area. Remote Sens. Environ. 115 (12), 3599–3614.
Penner, M., Pitt, D.G., Woods, M.E., 2013. Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario. Can. J. Rem. Sens. 39 (5), 426–443.
Persson, H.J., Ståhl, G., 2020. Characterizing uncertainty in forest remote sensing studies. Rem. Sens. 12 (3), 505.
Quah, D., 1993. Galton's fallacy and tests of the convergence hypothesis. Scand. J. Econ. 427–443.
Réjou-Méchain, M., Barbier, N., Couteron, P., Ploton, P., Vincent, G., Herold, M., Mermoz, S., Saatchi, S., Chave, J., de Boissieu, F., Feret, J., Takoudjou, S.M., Pélissier, R., 2019. Upscaling forest biomass from field to satellite measurements: sources of errors and ways to reduce them. Surv. Geophys. 40, 881–911.
Ruckelshaus, M., Hartway, C., Kareiva, P., 1997. Assessing the data requirements of spatially explicit dispersal models. Conserv. Biol. 11 (6), 1298–1306.
Saarela, S., Varvia, P., Korhonen, L., Yang, Z., Patterson, P.L., Gobakken, T., Næsset, E., Healey, S.P., Ståhl, G., 2023. Three-phase hierarchical model-based and hybrid inference. MethodsX 11, 102321.
Solberg, S., Weydahl, D.J., Nasset, E., 2010. Simulating X-band interferometric height in a spruce forest from airborne laser scanning. IEEE Trans. Geosci. Rem. Sens. 48 (9), 3369–3378.
Särndal, C.E., Swensson, B., Wretman, J., 2003. Model Assisted Survey Sampling. Springer, Berlin, Heidelberg.
Shukla, G.K., 1972. On the problem of calibration. Technometrics 14 (3), 547–553.
Tellinghuisen, J., 2000. Inverse vs. classical calibration for small data sets. Fresen. J. Anal. Chem. 368, 585–588.
Tian, Y., Nearing, G.S., Peters-Lidard, C.D., Harrison, K.W., Tang, L., 2016. Performance metrics, error modeling, and uncertainty quantification. Mon. Weather Rev. 144 (2), 607–613.
Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., Katila, M., 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sens. Environ. 112 (5), 1982–1999.
Thompson, S.K., 2012. Sampling, vol. 755. John Wiley & Sons, New Jersey.
Wang, R., Chen, J.M., Liu, Z., Arain, A., 2017. Evaluation of seasonal variations of remotely sensed leaf area index over five evergreen coniferous forests. ISPRS J. Photogrammetry Remote Sens. 130, 187–201.
Wilhelmsson, P., Sjödin, E., Wästlund, A., Wallerman, J., Lämås, T., Öhman, K., 2021. Dynamic treatment units in forest planning using cell proximity. Can. J. For. Res. 51 (7), 1065–1071.
Wu, C., Thompson, M.E., 2020. Sampling Theory and Practice. Springer, Cham.