

Approximating Prediction Uncertainty for Random Forest Regression Models

John W. Coulston , Christine E. Blinn , Valerie A. Thomas , and Randolph H. Wynne

Abstract

Machine learning approaches such as random forest have increased for the spatial modeling and mapping of continuous variables. Random forest is a non-parametric ensemble approach, and unlike traditional regression approaches there is no direct quantification of prediction error. Understanding prediction uncertainty is important when using model-based continuous maps as inputs to other modeling applications such as fire modeling. Here we use a Monte Carlo approach to quantify prediction uncertainty for random forest regression models. We test the approach by simulating maps of dependent and independent variables with known characteristics and comparing actual errors with prediction errors. Our approach produced conservative prediction intervals across most of the range of predicted values. However, because the Monte Carlo approach was data driven, prediction intervals were either too wide or too narrow in sparse parts of the prediction distribution. Overall, our approach provides reasonable estimates of prediction uncertainty for random forest regression models.

Introduction

Remote sensing scientists have a rich set of standard methods with which the uncertainty of (inherently categorical) thematic maps derived from remotely-sensed data can be estimated (e.g., Congalton and Green, 2008). For the most part, resulting uncertainty estimates are (a) independent of the analytical method used for the categorical data analysis, and (b) contain information on category-specific accuracy but not pixel specific accuracy. Methods with which to estimate the uncertainty of mapped continuous fields are, in contrast, much less standardized. Category-specific accuracy, of course, is no longer relevant, but the means by which uncertainty of continuous variables is estimated is often tied to the technique used. Examples abound, including use of RMSE in classical regression oriented approaches (Fernandes *et al.*, 2004) and cross-validation-derived PRESS (sum of squares of the prediction residuals) RMSE (Popescu *et al.*, 2004). Cross-validation approaches are also widely used in regression tree analyses of remotely sensed data (Bacini *et al.*, 2007). The cross-validation can estimate many prediction error statistics, including residual sum of squares. However, increasingly cross-validation is used primarily for model selection and (usually non-parametric) bootstrapping is used once the model is “fixed” (see, e.g., Molinaro, 2005). These methods have been extended to random forest implementations, but the resulting estimates of prediction uncertainty are aggregated (i.e., global) and do not produce pixel-specific uncertainties required for use in subsequent spatial modeling.

The use of machine learning techniques has increased substantially in remote sensing and geospatial data development. For example, Homer *et al.* (2004) used regression trees for the

development of a categorical land cover map for the United States, and Coulston *et al.* (2012) used random forests to develop a continuous field map of percent tree canopy cover. Other techniques that have been proposed and tested include artificial neural networks, support vector machines, stochastic gradient boosting, and K nearest neighbor (Moisen and Frescino, 2002; Wieland and Pittore, 2014). Machine learning approaches have become particularly attractive because they are well suited to recognize patterns in high-dimension data (Cracknell and Reading, 2014). Further, several of these approaches allow for modeling either categorical response variables or continuous response variables (e.g., random forests, support vector machines/support vector regression). However unlike traditional parametric approaches (e.g., multiple regression), information about prediction error (standard error of a prediction for a new data point) is not readily available.

Broad scale raster maps of continuous variables have been developed for percent impervious surface (Homer *et al.*, 2007), percent tree canopy (Huang *et al.*, 2001; Coulston *et al.*, 2012), forest biomass (Blackard *et al.*, 2008), and forest carbon (Wilson *et al.*, 2013) among other examples. These efforts all relied on machine learning approaches and used either Landsat or MODIS imagery for predictor variables. Each pixel within these modeled raster maps contains a predicted value yet, per-pixel uncertainty is rarely expressed along with the predictions. Understanding the pixel-level uncertainty is critical to understanding the utility of the data. Furthermore, many geospatial datasets (such as those mentioned above) are used in subsequent modeling applications. For example, the 2001 NLCD tree canopy cover dataset (Huang *et al.*, 2001) was a major component of forest fire behavior and fuel models (Rollins and Frame, 2006). Clearly the uncertainty around this fire behavior model is related to the uncertainty in the underlying data, such as the 2001 NLCD percent tree canopy cover. Our intent is to provide guidance on quantifying prediction uncertainty at the pixel level.

While there are numerous machine learning techniques, here we focus on random forest because it is straightforward to train, computationally efficient, and provides stable predictions (Cracknell and Reading, 2014). Random forest is an ensemble method that uses bootstrap aggregating (bagging) to develop multiple models to improve prediction (Breiman, 2001). Along with bagging, random forests also relies on random feature selection to develop a forest of independent CART models. This technique has been used by Powell *et al.* (2010) and Bacini *et al.* (2008) to predict forest biomass, Evans and Cushman (2009) to predict species occurrence probability, Hernandez *et al.* (2008) to predict faunal species distributions, and Moisen *et al.* (2012) to predict percent tree canopy cover. Though there have been numerous studies describing and using random forests, there is a lack of information regarding prediction

John W. Coulston is with the USDA Forest Service, Southern Research Station, Blacksburg, VA (jcoulston@fs.fed.us).

Christine E. Blinn , Valerie A. Thomas , and Randolph H. Wynne are with Virginia Polytechnic Institute and State University, Department of Forest Resources and Environmental Conservation, Blacksburg, VA.

Photogrammetric Engineering & Remote Sensing
Vol. 82, No. 3, March 2016, pp. 189–197.
0099-1112/16/189–197

© 2016 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.82.3.189

uncertainty. The objective of this study is to develop a technique to approximate prediction uncertainty for random forest models of continuous data. In our case we consider prediction uncertainty to be the uncertainty around a future prediction for a new observation (i.e., pixel-level uncertainty).

We further present a case example using predicted percent tree canopy cover in Georgia, USA. Portraying map uncertainty is an important component for geospatial data developers to consider. In some cases, prediction uncertainty is a central component to developing a final geospatial dataset. For example, the 2001 and 2011 NLCD percent tree canopy cover layers strived to mask out areas where there is clearly no tree canopy cover but canopy cover models predict low levels of tree canopy cover. In the 2001 NLCD percent tree canopy cover layer, the mask was created by creating a “liberal” tree cover map and hand editing (Huang *et al.*, 2001). However, the techniques presented here facilitate a more parsimonious approach.

Methods

Through the methods section we use standard matrix and bootstrap notation. Bold lower case letters (e.g., $\mathbf{y}$) represent a vector. Bold upper case letters (e.g., $\mathbf{X}$) represent a matrix. A $*$ superscript followed by b (e.g., $\mathbf{y}^{*b}$) refers to the b^{th} bootstrap sample and a $*$ superscript followed by $-b$ (e.g., $\mathbf{y}^{*-b}$) denotes the portion of the original data that was not part of the b^{th} bootstrap sample. Greek letters represent parameters (e.g., τ) and a vector of parameters or matrices of parameters are bold as described above.

Random Forest Overview

We provide a brief overview of Random Forest but point the interested reader to Breiman (2001) for more details. Random forest is an ensemble approach that relies on CART models. The goal of CART is to understand (learn) the relationship between a dependent variable ($\mathbf{y}$) and a set of predictor variables ($\mathbf{X}$). The learning algorithm employs recursive partitioning in which splits in the $\mathbf{X}$ variables are selected to create homogeneous groupings of $\mathbf{y}$. The recursive partitioning continues until either the subset of $\mathbf{y}$ at each node is the same value or further splitting adds no further improvement. Random forest differs from the CART procedure by (a) employing bootstrap resampling (Efron and Tibshirani, 1993), and (b) random variable selection. Consider a regression tree which is made up of splits and nodes. With random forest a random subset of $\mathbf{X}$ variables (selected without replacement) is used to determine the split for each node. Bootstrap resampling is used to develop replicates of the CART model. For continuous variables the ensemble estimate is the mean of the predicted values across trees ($\hat{\mathbf{y}}$) and the variance across trees is $\text{var}(\hat{\mathbf{y}})$.

Methods Overview

Generally speaking, our method to approximate prediction uncertainty for random forest regression models has five main steps (Figure 1). Step 1 is to fit a random forest model based on all available data. Step 2 is to use bootstrap resampling to parameterize a large number of random forest models (Figure 1B). Bootstrap resampling generally results in ~37 percent of the observations not being selected. Step 3: for each bootstrap replicate of the random forest (RF) model the observed values and predicted values are retained for observations not included in the bootstrap sample (Figure 1C). Step 3 yields an error assessment dataset. In Step 4 the properties of the prediction error are quantified using the error assessment dataset (Figure 1D). Step 5 is to make a prediction, including error, for a new observation (Figure 1E).

Bootstrap Resampling

The bootstrap is one tool that can also be used to approximate prediction uncertainty of a RF model (Figure 1B). Consider the response and predictor variables ($\mathbf{y}$, $\mathbf{X}$) where a bootstrap sample of ($\mathbf{y}$, $\mathbf{X}$) is ($\mathbf{y}^{*b}$, $\mathbf{X}^{*b}$). Suppose we draw $B = 2000$ bootstrap

samples to create $B = 2000$ bootstrap datasets. Using the bootstrap datasets we construct RF^{*1} , RF^{*2} , ..., RF^{*2000} random forest models, and then for each replicate quantify the prediction error for each observation in ($\mathbf{y}^{*-b}$) based on the corresponding RF^{*b} replicate. The error assessment is constructed for each observation based on the distribution of predicted values when the observation was not part of the bootstrap sample (Figure 1C). The prediction error is $\sqrt{\text{MSE}}$ where MSE is the mean square error for each observation. This technique allows one to quantify prediction error for each element of a holdout dataset but does not directly apply to predictions based on a new $\mathbf{X}$. However, because random forest relies on bootstrap sampling to construct the ensemble, a random forest model contains information that we can use to quantify prediction uncertainty for new locations (i.e., new $\mathbf{X}$ data are available, Figure 1D and 1E).

Prediction Uncertainty

In traditional parametric models (e.g., multiple regression), the prediction error for a new observation is a function of the mean squared error (MSE) and the variability in $\mathbf{X}$. Recall that a random forest model is an ensemble of CART models and the ensemble estimate is the mean across the set of CART model predictions. Each of the CART models is considered a weak learner. The predictions from these weak learners inherently capture the variability in the relationship between $\mathbf{X}$ and $\mathbf{y}$. We can calculate the variance among predictions, for each observation in $\mathbf{X}$, which represents the variability of predictions among CART models. However, we need to scale between $\text{var}(\hat{\mathbf{y}})$ and $(\mathbf{y} - \hat{\mathbf{y}})^2$ to approximate prediction uncertainty (Figure 1D). This is because to approximate the prediction uncertainty for a new observation only $\text{var}(\hat{\mathbf{y}})$ will be

available. A measure such as $\tau = \sqrt{\frac{(\mathbf{y} - \hat{\mathbf{y}})^2}{\text{var}(\hat{\mathbf{y}})}}$ provides such a

scaling. To implement this approach, for each bootstrap dataset ($\mathbf{y}^{*b}$, $\mathbf{X}^{*b}$) a random forest model is constructed RF^{*1} , RF^{*2} , ..., RF^{*B} . For each observation in $\mathbf{y}^{*-1}$, $\mathbf{y}^{*-2}$, ..., $\mathbf{y}^{*-B}$, a prediction ($\hat{\mathbf{y}}$) is made from the corresponding model RF^{*1} , RF^{*2} , ..., RF^{*B} . Subsequently $\mathbf{T} = [\tau^{*-1}, \tau^{*-2}, \dots, \tau^{*-B}]$ is calculated. The $\hat{\tau}$ for a 95 percent confidence interval can be estimated either using a bootstrap approach or a Monte Carlo approach (Figure 1D). For the Monte Carlo approach, the value of $\hat{\tau}$ such that 95 percent of the predictions lie within $\hat{\tau} \cdot \text{sd}(\hat{\mathbf{y}})$ of the true value is estimated by taking the 95th percentile across all elements in $\mathbf{T}$. For the bootstrap approach, the value of $\hat{\tau}$ such that 95 percent of the predictions lie within $\hat{\tau} \cdot \text{sd}(\hat{\mathbf{y}})$ of the true value is found by taking the 95th percentile for each τ^{*-b} in $\mathbf{T}$, and the bootstrap estimate of $\hat{\tau}$ is then the average across the B replicates. For this analysis we used the Monte Carlo approach.

Simulation to Generate Known Populations

We used a simulation approach to examine our proposed method to approximate prediction uncertainty in random forest models, so that we could evaluate our technique for known populations. We constructed six populations representing $\mathbf{X}$ variables. Each was constructed using Gaussian random fields (Schlather *et al.*, 2014) with different levels of spatial correlation (Figure 2). Gaussian random fields were used to simulate $\mathbf{X}$ variables because they offer a framework to develop normally distributed data within the spatial domain. Each map was 1,000 by 1,000 pixels. From the $\mathbf{X}$ maps we constructed three different $\mathbf{Y}$ populations (Figure 2):

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 + \mathbf{X}_4 + \mathbf{X}_5 + \mathbf{X}_6 + \mathbf{N}(0,2)$$

$$\mathbf{Y}_2 = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 + \mathbf{X}_4 + \mathbf{X}_5 + \mathbf{X}_6 + \mathbf{N}(0,1)$$

$$\mathbf{Y}_3 = \mathbf{X}_1 \cdot \mathbf{X}_2 + \mathbf{X}_3 + (\mathbf{X}_4 + \mathbf{X}_5)^2 + \mathbf{X}_6 + \mathbf{N}(0,1)$$

where $\mathbf{N}$ (mean, standard deviation) is additional random noise drawn from a normal distribution. The Normal High

population (Y_1) was developed to replicate a population where the value was a linear combination of the X values and the assumptions required for multiple linear regression were valid. Likewise, the Normal Low (Y_2) population was equivalent to the Normal High population except that less random noise was added. Both the Y_1 and Y_2 populations were normally distributed and error under linear regression was normal and homoscedastic. The Model Misspecification (Y_3) population was developed based on a non-linear combination of the X values that would violate the assumptions of multiple linear regression. The Y_3 population was non-normal and skewed left, and the error distribution under multiple linear regression was non-normal and heteroscedastic.

Quantifying Prediction Uncertainty from the Known (Simulated) Populations

To approximate common sampling rates, we drew 500 sample locations (0.05 percent sample) and extracted values for each dependent variable (y_1, y_2, y_3) and independent variable ($X = [x_1, x_2, \dots, x_6]$). First, for each dependent variable (y_1, y_2, y_3), we constructed a random forest model of the form $y = f(X) + \epsilon$ (Figure 1A). Call these models RFy₁, RFy₂, and RFy₃. For each random forest model we used 500 regression trees; two

independent variables were randomly selected for determining the splits at each node, and the models were bias corrected (Liaw and Wiener, 2002). Next we employed bootstrap resampling (Figure 1B) and we used $B = 2,000$ bootstrap samples to create 2,000 random forest models, using the same model specification as described above, for y_1, y_2 , and y_3 . We then predicted values ($\hat{y}$), calculated the variance across tree-level predictions ($\text{var}(\hat{y})$) (Figure 1C), calculated the squared prediction error $((y - \hat{y})^2)$, and estimated $\hat{\tau}$ for each RFy₁, RFy₂, and RFy₃ based on observations that were not part of each bootstrap sample (Figure 1D). To estimate the $\hat{\tau}$ value for each model (RFy₁, RFy₂, and RFy₃) we selected the 95th percentile of T values for each model (Figure 1D).

The random forest models RFy₁, RFy₂, and RFy₃ were used to predict each variable spatially (i.e., applied to all pixels in the map). Additionally for each of the predicted maps, the variance across tree-level predictions ($\hat{y}$) was calculated and the width of the 95 percent prediction intervals were then $\pm \hat{\tau} \cdot \text{sd}(\hat{y})$ (Figure 1E). We then compared the predicted values, and the 95 percent prediction interval to the true population values for Y_1 (the Normal High population), Y_2 (Normal Low population), and Y_3 (Model Misspecification population) by examining

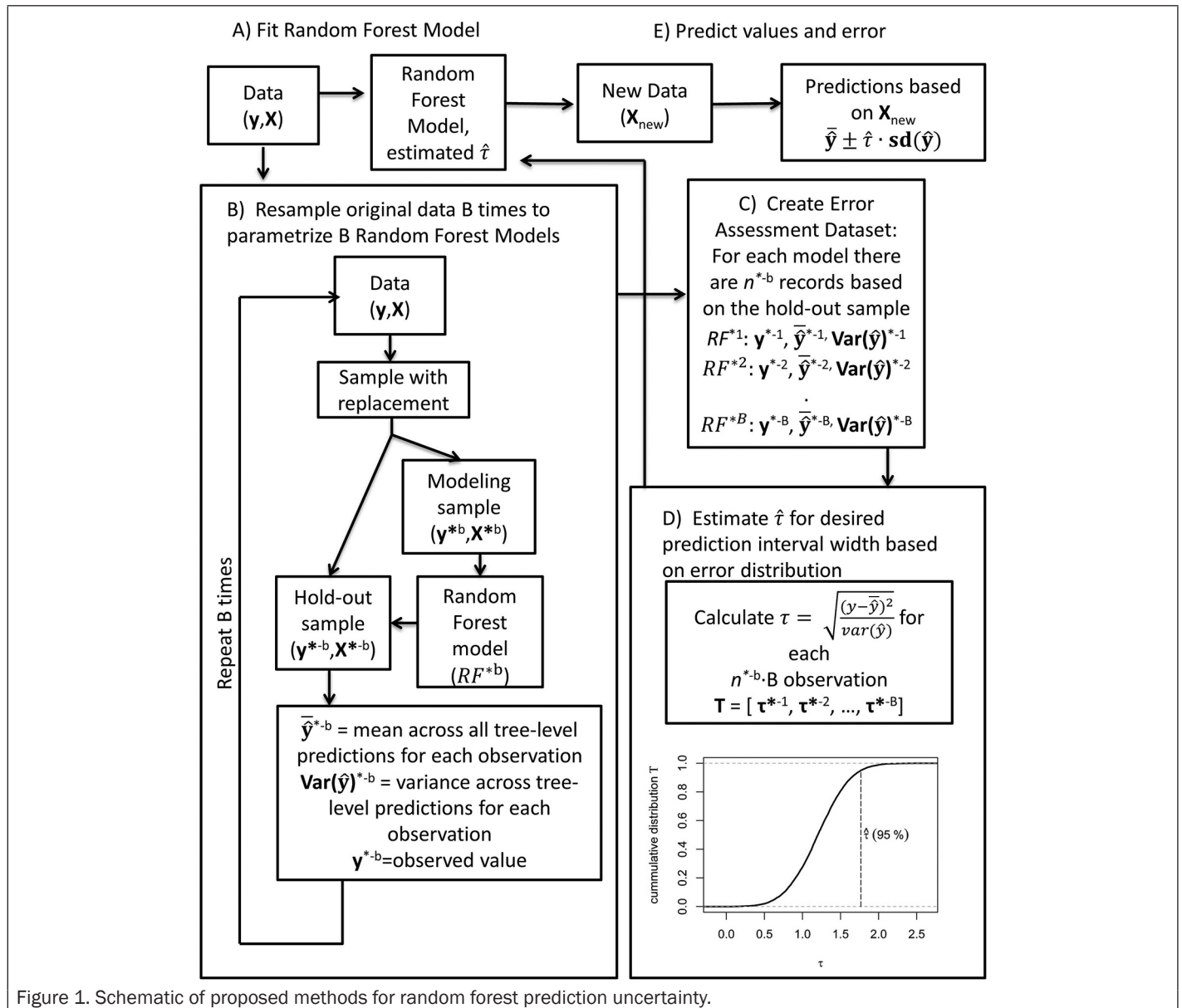


Figure 1. Schematic of proposed methods for random forest prediction uncertainty.

the percentage of observations that fell within the prediction interval for each integer value covering the range in predicted values.

Prediction Error from Linear Regression

To develop context for our approach, we also examined the behavior of prediction error based on multiple regression. The purpose for this is so readers can compare the behavior of our approach to prediction to that of multiple regression. Deviations in the results between our approach for random forest and the multiple regression approach helps to identify areas where our approach could be improved.

Using the sample data (500 observations), we parameterized linear regression models for y_1 , y_2 , and y_3 . The model form was $y = X \cdot \beta + \epsilon$. We predicted values and 95 percent prediction intervals for Y_1 (the Normal High population), Y_2 (Normal Low population), and Y_3 (Model Misspecification population) based on standard equations (see Draper and Smith, 1981 for background). For each population we also examined the percentage of observations that fell within the prediction interval for each integer value covering the range of predicted values.

Case Example Using Real (Unsimulated) Data from Georgia

We provide a case example for percent tree canopy cover

mapping based on data from Coulston *et al.* (2012) for Georgia. The dependent variable was percent tree canopy cover estimated by photo interpretation of 2009 National Agriculture Imagery Program (NAIP) photography for approximately 4,160 sample locations. The independent variables were based on leaf-on Landsat-5 TM imagery from 2008 to 2009 and elevation data. The six reflective Landsat-5 TM bands, normalized difference vegetation index, and tasseled cap were also used. Slope and cosine of aspect were derived from the elevation data and also used as predictor variables. Based on these data, we fit a random forest model which had a pseudo- R^2 (Liaw and Wiener, 2002) of 0.85 (RMSE = 13.2) and created a predicted surface of percent tree canopy cover (example area in Plate 1). We note that there is a substantial amount of area where, based on the NAIP imagery, there were no trees and hence the percent tree canopy cover should be zero. Many of these areas have predicted values in the 0.5 to 10 percent range. Our goal is to use the previously described Monte Carlo technique to mask out areas where the 95 percent prediction interval includes zero.

To implement our Monte Carlo procedure for this problem, note that we are interested in the uncertainty when the true value is zero and the predicted value is >0 . In this specific

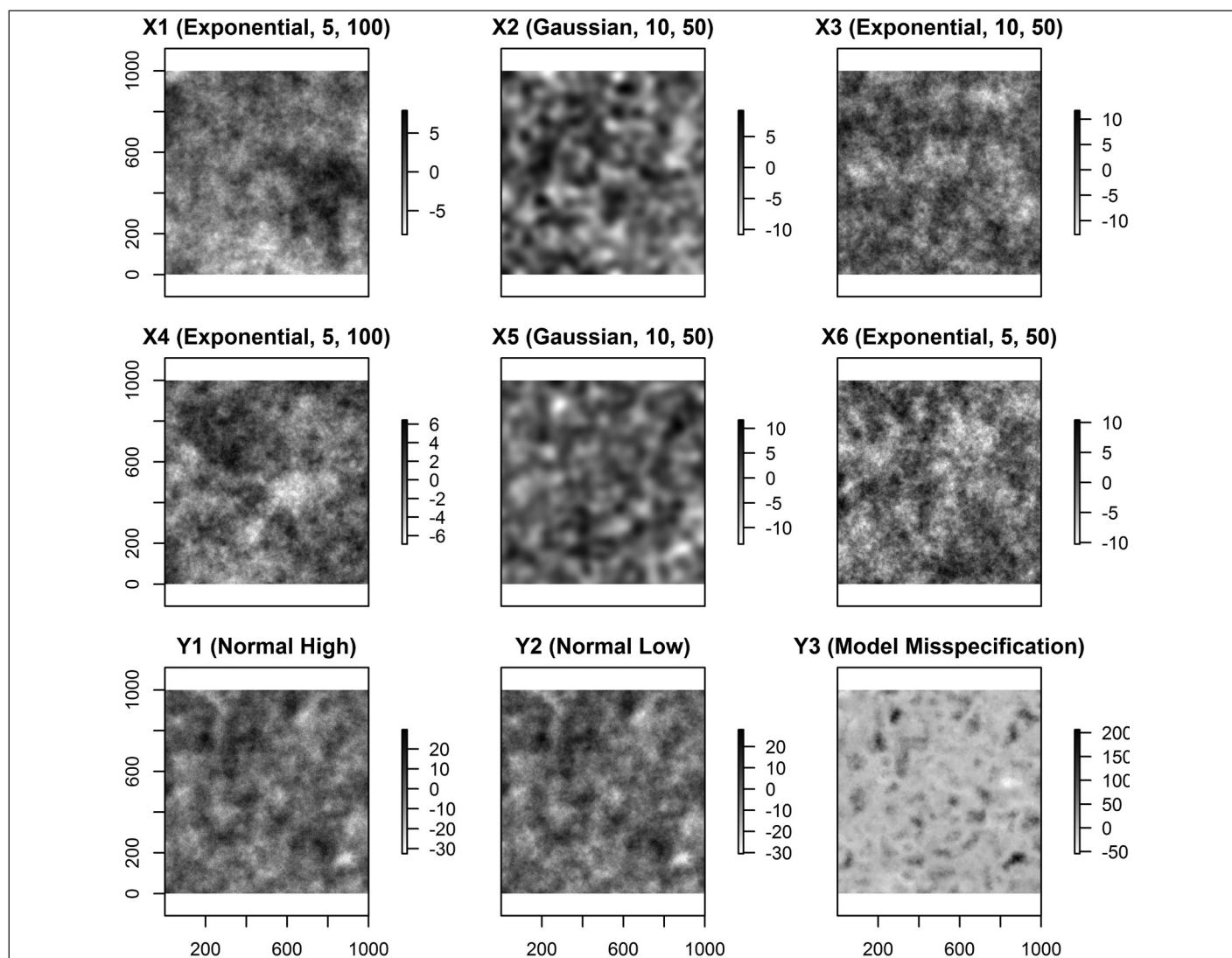


Figure 2. Simulated independent (X) and dependent (Y) variables. For the independent variables the variogram model, the variance parameter, and the scale parameter are denoted in parentheses. For background on variogram models see Isaaks and Srivastava (1989) for an example.

situation we want to estimate the value of $\hat{\tau}$ such that 95 percent of the true zero values are within $\hat{\tau} \cdot sd(\hat{y})$ percent of zero. The masked version is then developed by setting all predictions to zero when $\hat{y} - \hat{\tau} \cdot sd(\hat{y}) \leq 0$. This approach approximates a 95 percent certainty for areas of no canopy cover. To implement, we used 2,000 bootstrap replicates in our Monte Carlo assessment. To estimate $\hat{\tau}$ we simply restricted the Monte Carlo assessment dataset to the observation where the true value was zero and then estimated $\hat{\tau}$. Additionally, we developed 95 percent prediction intervals for all predicted values.

Results

Random Forest Models

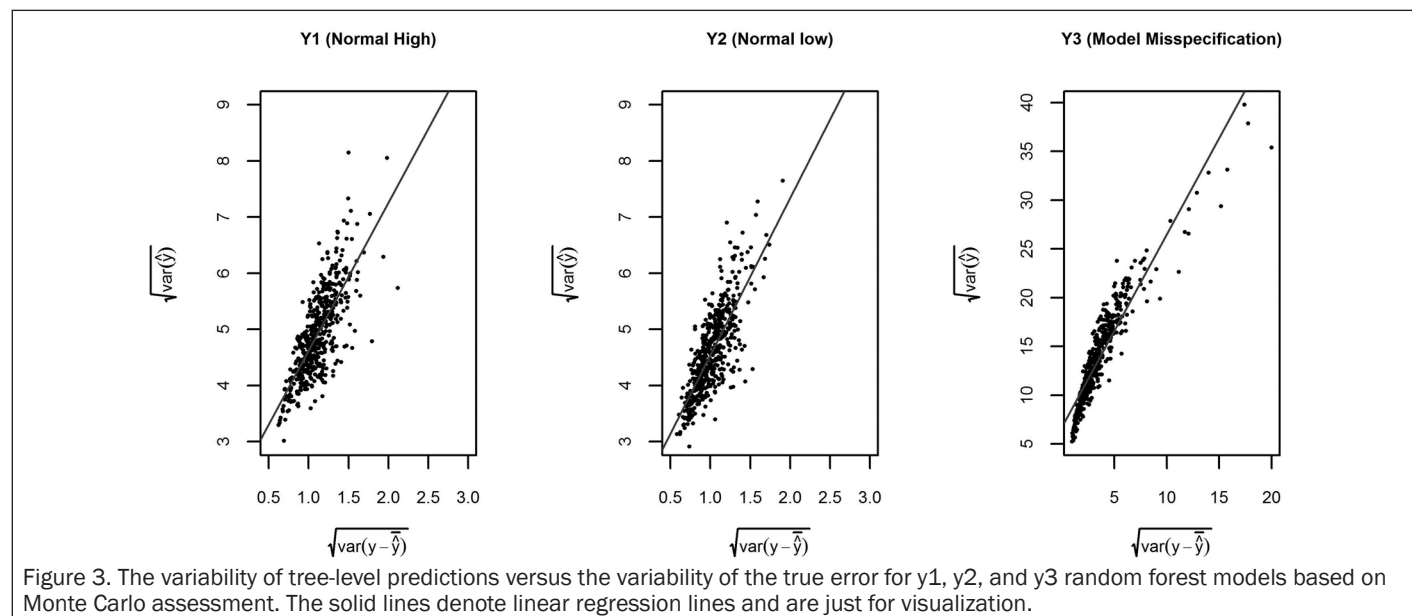
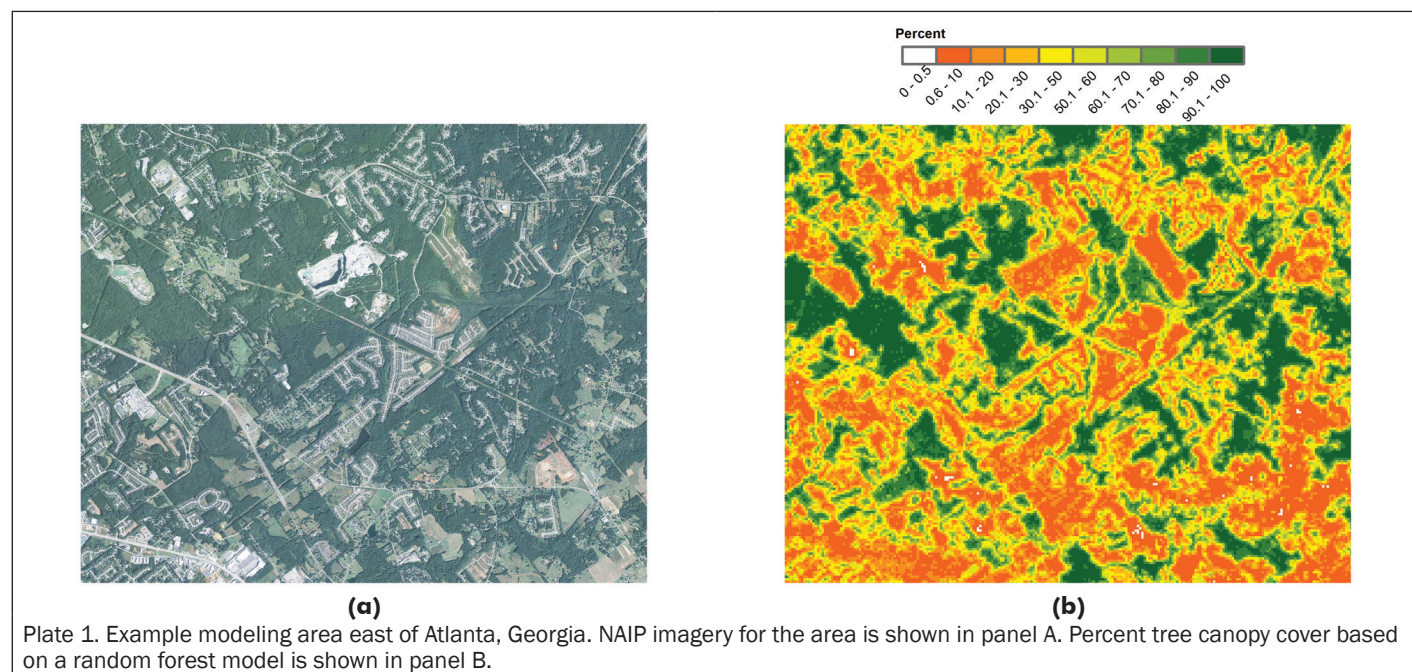
Using the 500 samples we fit random forest models for y_1 , y_2 , and y_3 . The pseudo- R^2 and root mean square error (RMSE) for the RFy1 model were 0.85 and 2.85, respectively. For the

RFy2 model the pseudo- R^2 was 0.90 and the RMSE was 2.28. The RFy3 random forest model had a pseudo- R^2 of 0.73 and a RMSE of 10.20.

Monte Carlo Error

The 500 original samples were sampled via the bootstrap 2,000 times which yielded a Monte Carlo assessment dataset where each of the 500 observation had ~740 observed and predicted values for each of the dependent variables (y_1 , y_2 , y_3). Figure 3 shows the relationship between the variability in individual tree-level predictions and the variability of the true error for each dependent variable. The correlations between these two quantities summarized for the 500 observations were 0.75, 0.78, and 0.93 for RFy1, RFy2, and RFy3, respectively. The values for constructing approximate 95 percent prediction intervals were 1.36, 1.18, and 1.31 for RFy1, RFy2, and RFy3, respectively. These steps are illustrated in Figure 1B and 1D.

Simulated Data: Predicting Values for New Observations and Uncertainty



Predicted maps were made and 95 percent prediction intervals were examined for Normal High (Y_1), Normal Low (Y_2), and Model Misspecification (Y_3) populations. For example, we randomly selected 100 pixels from Normal High (Y_1) population and examined the observed, predicted, and 95 percent prediction interval (Figure 4). Three out of 100 predictions (3 percent) fell outside the prediction interval (denoted by the open circles in Figure 4). However, across all draws of 100 observations in the population we would expect 5 percent of the true values, on average, to fall outside the prediction interval (95 percent would be in the prediction intervals). We would further expect the proportion of true values to fall outside the 95 percent prediction interval for any subregion of predicted values (e.g., predicted values < -10 in Figure 4) to be 0.05.

Overall, 0.97 of predictions for the Normal High (Y_1) population fell within the 95 percent prediction interval. For Normal Low (Y_2) and Model Misspecification (Y_3) populations 0.98 and 0.96 of the observations fell within the 95 percent prediction interval, respectively. We further examined the behavior of our uncertainty approach by examining all pairs of observed and predicted values and the frequency at which observed values were contained in 95 percent prediction intervals (Plate 2). The frequency was determined by examining the proportion of prediction intervals that contained the observed value by integer bins of predicted values (i.e., the predicted values were round to the nearest integer). We expected the proportion in the 95 percent prediction interval to be 0.95. We found that prediction intervals were generally conservative for 99 percent of predicted values in the Monte Carlo assessment. Results were generally adequate within this range of predicted values although for the Model Misspecification (Y_3) population there was an underestimation of width of the prediction interval in the right tail of the distribution. The uncertainty for predictions that had little or no representation in the Monte Carlo analysis were somewhat spurious. In some cases prediction interval width was underestimated while in other cases it was overestimated leading to prediction intervals that were either too narrow or too wide.

Multiple Regression Results

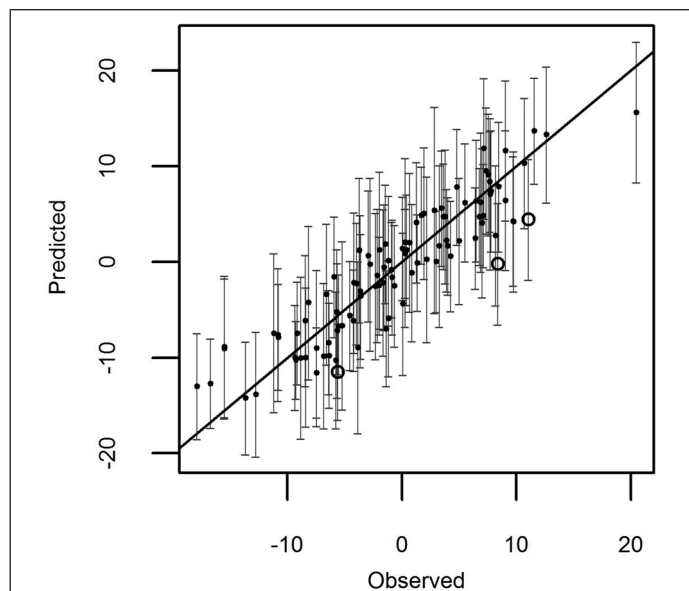


Figure 4. Predicted versus observed and prediction intervals for 100 randomly selected predictions of Y_1 . Prediction intervals are denoted by the grey error bars. The open circles represent predictions whose 95 percent prediction interval does not contain the observed value.

We also examined the behavior of multiple regression models for the three populations. The y_1 , y_2 , and y_3 models had r^2 of 0.93, 0.98, and 0.03, respectively. The RMSE was 1.98, 0.99, and 19.12 for y_1 , y_2 , and y_3 , respectively. These fit statistics were not based on cross-validation. Overall, 0.95 of predictions for the Normal High (Y_1) population fell within the 95 percent prediction interval. For the Normal Low (Y_2) and Model Misspecification (Y_3) populations 0.95 and 0.94 of the observations fell within the 95 percent prediction interval, respectively. As noted above, the expected proportion of true values in the 95 percent prediction intervals was 0.95. For both the Normal High (Y_1) and Normal Low (Y_2) populations this expectation held within the 99 percent quantile of predicted values (Plate 3). Outside the 99 percent quantile results varied slightly. The Model Misspecification (Y_3) population behaved somewhat differently because the assumptions regarding the distribution of errors for multiple regression were intentionally violated. This resulted in predicted values that were relatively close to the mean and prediction intervals that were generally too narrow outside the 99 percent quantile of predicted values (Plate 3).

Case Example in Georgia

We developed 95 percent prediction intervals for the study area. The half-width of the 95 percent prediction interval ranged from 0.27 percent to 90 percent (Plate 4A). Generally speaking, the prediction interval width was wider in heterogeneous areas such as edges between treed areas and non-treed areas. We also developed a “masked” version of the percent tree canopy cover map and the procedure provided reasonable results (Plate 4B). Generally speaking, areas that were clearly agriculture and un-vegetated developed areas were readily masked out from having canopy cover predictions.

Discussion

Since the late 1980s a substantial amount of research has focused on uncertainty in spatial products (Foody and Atkinson, 2002), though, as previously noted, there is much more methodological maturation with respect to categorical variables (particularly for parametric classifiers) than for continuous variables. With (a) the increasing prevalence of continuous field mapping (e.g., leaf area index, tree canopy cover, biomass, water turbidity, and the like), (b) the production of continuous field maps using non-parametric approaches, and (c) the use of these maps in subsequent geospatial modeling, there is a clear and present need for robust methods by which pixel-specific uncertainty can be estimated. The prior literature, while sparse, does clearly make the case for estimation (Wang *et al.*, 2005) and visualization (Dungan *et al.*, 2003) of pixel-specific uncertainty. Further, the potential role of simulation has long been established (Englund, 1993), absent the specifics needed for operational implementation in our particular use-case, namely continuous field estimation using random forest.

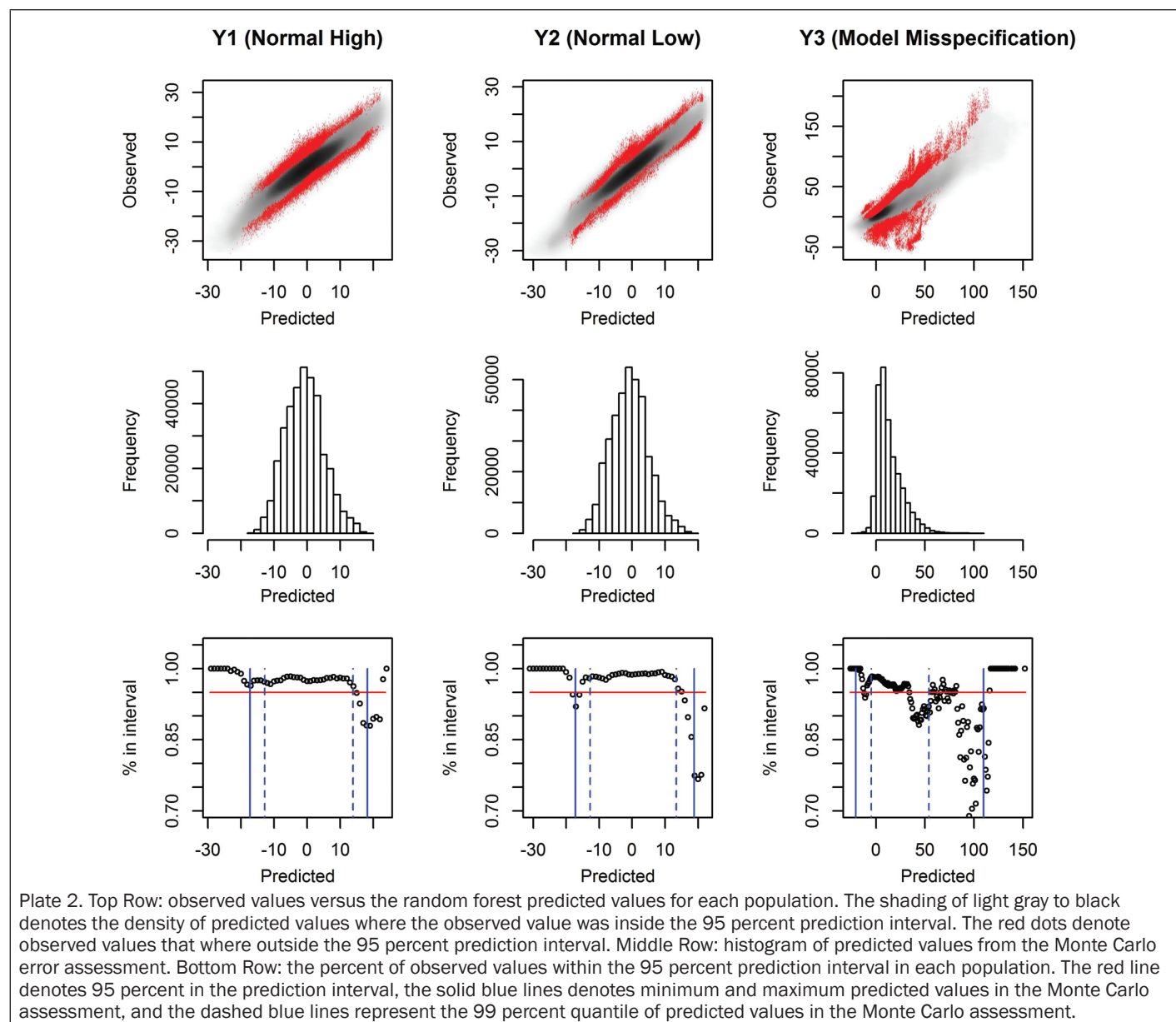
The Monte Carlo approach presented here is data driven and generally provided conservative prediction interval widths (i.e., wider than needed) within the 99 percent quantile of predicted values. Outside the 99 percent quantile, prediction interval widths could be too wide or too narrow. There are several ways in which this could have occurred. Our approach was data driven and therefore underperforms in sparse areas of the distribution. For example, for our test we drew 500 sample of the population (0.05 percent). Increasing the sample to 5,000 observations (0.5 percent sample) improved results significantly. Further, in the sparse parts of the distribution, observations can be rare enough that the way we analyzed the results (proportion of observations within their 95 percent prediction intervals (e.g., Plate 2), may not have sufficient information to estimate the proportion. For example, in the Normal High (Y_1) population fewer than 20

observations had predicted values < -24 . This means that the proportion within the 95 percent prediction intervals could increment in steps of 0.05 or greater. In short, the technique presented here assumes that the sampled data provide enough information about the population that uncertainty can be quantified. This assumption is common to multiple regression where practitioners are concerned with predicting values beyond the values in the sample dataset (Weisberg, 1985). Also, our approach assumed a random equal probability sample of the underlying populations. Modifications to the bootstrap protocol would be needed for other sampling schemes such as stratified random sampling.

Our case example was based on masking out areas that were likely to have 0 percent tree canopy cover in a portion of Georgia. Our technique offers a straightforward way to accomplish this however; masking procedures that only operate on part of the distribution may increase other types of errors. For example, the original percent tree canopy cover had an error rate of 0.2 percent (predicting 0 percent tree canopy cover when the true value was > 0 percent). The masking procedure increased this error rate to 9.7 percent. Decisions

to mask final products should be made with caution. Masking procedures may decrease some error rates while at the same time increasing other error rates, as is the case here. Further, if a model was originally unbiased (e.g., mean error = 0) then manipulating predictions in one part of the distribution will result in overall bias (mean error $\neq 0$).

The ability to obtain a spatially-explicit understanding of uncertainty and error of predictions provides valuable information for any application where continuous variable mapping is desired. For complex spatial models, such as the LANDFIRE Prototype Project referred to previously (Rollins and Frame, 2006), this information can be used to understand the sensitivity of the modeled fire output and help to define the conditions for which model outputs are most reliable. Further, maps of uncertainty and error can inform field validation efforts, which could significantly reduce costs of monitoring efforts. One example of this would be in the design of sampling schemes to validate models of above ground biomass in remote areas such as Alaska or some areas in the tropical rainforests, where permanent inventory plots are lacking.



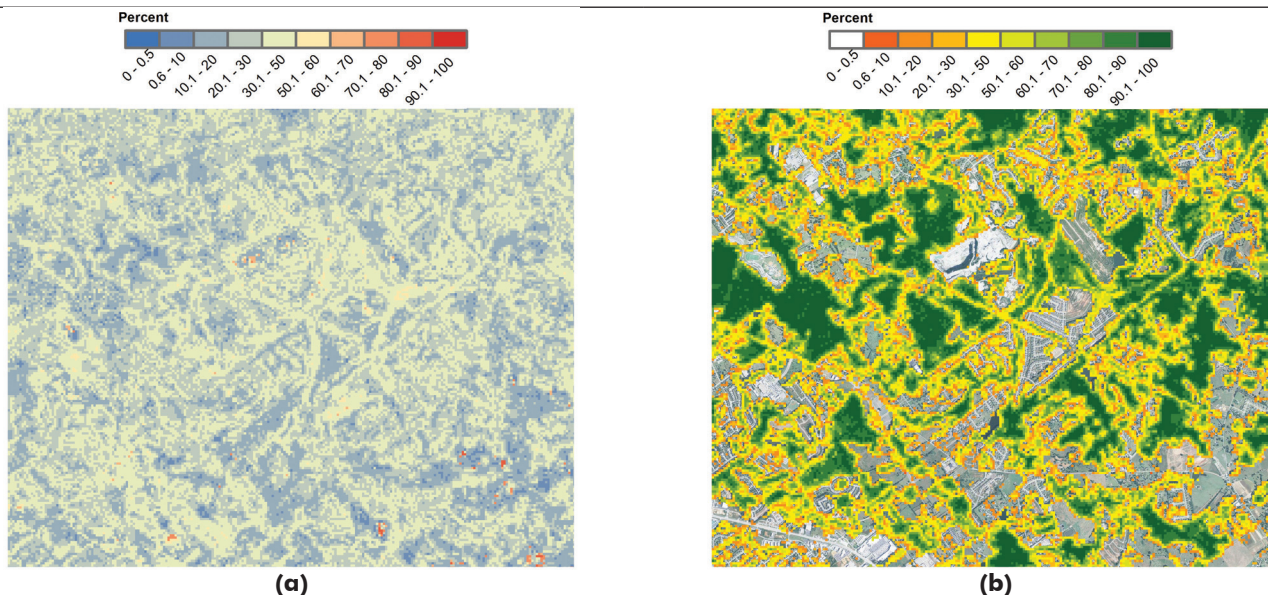
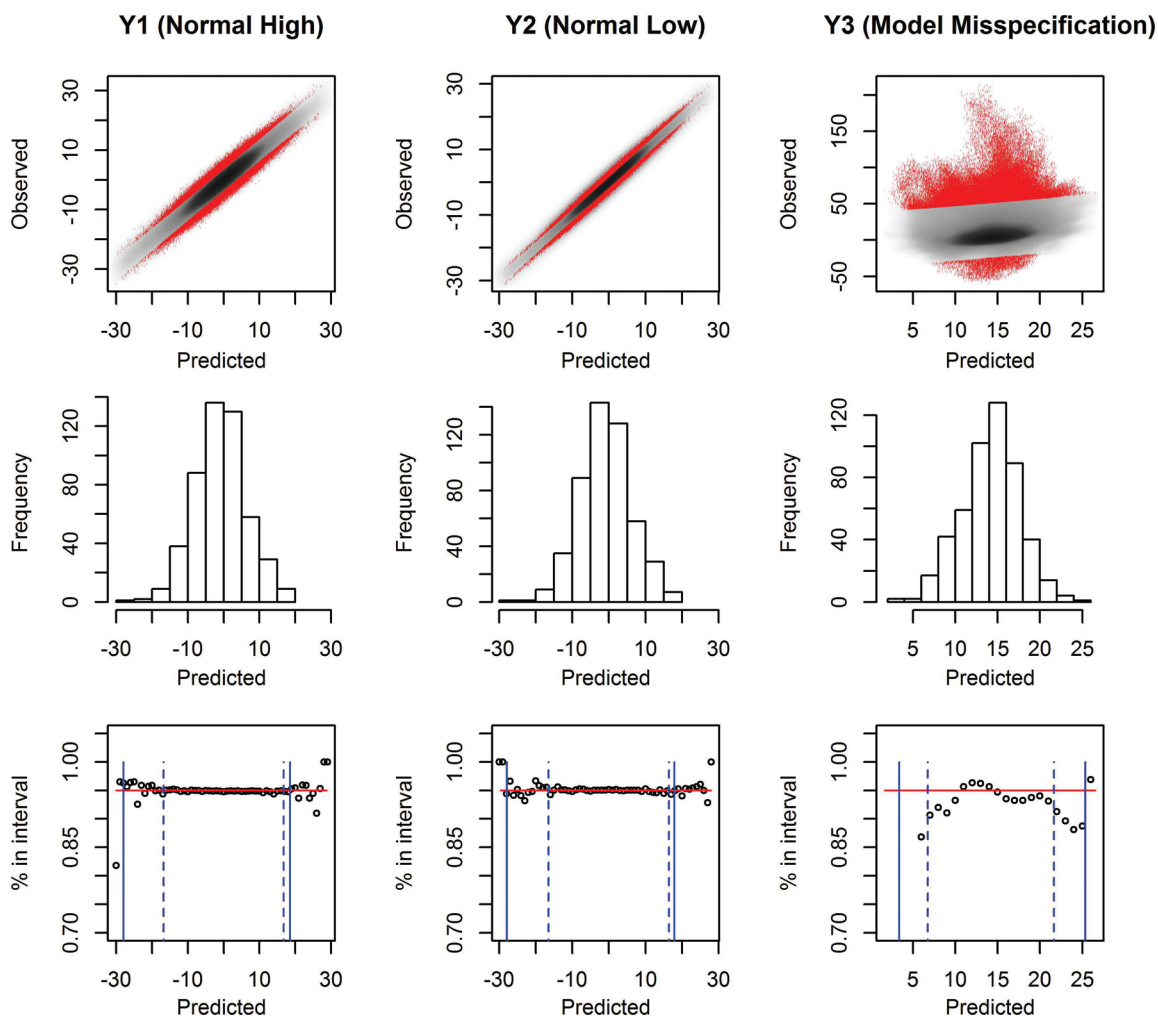


Plate 4. Half-width of the 95 percent prediction interval for percent tree canopy cover for a portion of (A) the Georgia study area, and masked predicted percent tree canopy cover with 5 percent error rate for area of no canopy cover overlaid on (B) the NAIP imagery.

Conclusions

Uncertainty remains an important subject area in remote sensing, but the tools used to quantify uncertainty must keep pace with the development and application of new modeling approaches. With the increased use of non-parametric modeling and classification approaches increased effort is required to provide uncertainty approaches for machine learning techniques. We developed a relatively straightforward approach to approximate prediction uncertainty for continuous maps developed from random forest models, tested the approach in a simulation environment and provided a case example. The results were reasonable but the method typically provided conservative confidence intervals for new observation. The approach is applicable to a broad range mapping efforts that use random forest models. This general approach may also be applicable to other ensemble modeling techniques.

References

- Baccini, A.N., Laporte, S.J., Goetz, M., Sun, and H. Dong, 2008. A first map of tropical Africa's above-ground biomass derived from satellite imagery, *Environmental Research Letters*, 3:045011, 9 p.
- Bacini, A., M.A. Friedl, C.E. Woodcock, and Z. Zhu, 2007. Scaling field data to calibrate and validate moderate spatial resolution remote sensing models, *Photogrammetric Engineering & Remote Sensing*, 73(8):945–954.
- Blackard, J.A., M.V. Finco, E.H. Helmer, G.R. Holden, M.L. Hoppus, D.M. Jacobs, A.J. Lister, G.G. Moisen, M.D. Nelson, R. Riemann, B. Ruefenacht, D. Salajanu, D.L. Weyermann, K.C. Winterberger, T.J. Brandeis, R.L. Czaplewski, R.E. McRoberts, P.L. Patterson, and R.P. Tymcio, 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information, *Remote Sensing of Environment*, 112(4):1658–1677.
- Breiman, L., 2001. Random forests, *Machine Learning*, 45(1):5–32.
- Congalton, R.G., and K. Green, 2008. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Second edition, CRC Press, 183 p.
- Coulston, J.W., G.G. Moisen, B.T. Wilson, M.V. Finco, W.B. Cohen, and C.K. Brewer, 2012. Modeling percent tree canopy cover: a pilot study, *Photogrammetric Engineering & Remote Sensing*, 78(7):715–727.
- Cracknell, M.J., and A.M. Reading, 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their responses to variation in the spatial distribution of training data and the use of explicit spatial information, *Computers & Geosciences*, 63:22–33.
- Draper, N., and H. Smith, 1981. *Applied Regression Analysis*, Second edition, John Wiley & Sons, 709 p.
- Dungan, J.L., D.L. Kao, and A. Pang, 2003. Modeling and visualizing uncertainty in continuous variables predicted using remotely sensed data, *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, 5:3017–3019.
- Efron, B., and R.J. Tibshirani, 1993. *An Introduction to the Bootstrap*, Chapman & Hall, 436 p.
- Englund, E., 1993. Spatial simulations: Environmental applications, *Environmental Modeling with GIS*, Oxford University Press, pp. 432–437.
- Evans, J.S., and S.A. Cushman, 2009. Gradient modeling of conifer species using random forests, *Landscape Ecology*, 24: 673–683.
- Fernandes, R.A., J.R. Miller, J.M. Chen, and I.G. Rubinstein, 2004. Evaluating image-based estimates of leaf area index in boreal conifer stands over a range of scales using high-resolution CASI imagery, *Remote Sensing of Environment*, 89(2):200–216.
- Footy, G.M., and P.M. Atkinson, 2002. *Uncertainty in Remote Sensing and GIS*, John Wiley & Sons, 326 p.
- Hernandez, P.A., I. Franke, S.K. Herzog, V. Pacheco, L. Paniagua, H.L. Quintana, A. Soto, J.J. Swenson, C. Tovar, T.H. Valqui, J. Vargas, and B.E. Young, 2008. Predicting species distributions in poorly-suited landscapes, *Biodiversity and Conservation*, 17:1353–1366.
- Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan, 2004. Development of a 2001 National Land Cover Database for the United States, *Photogrammetric Engineering & Remote Sensing*, 70(7):829–840.
- Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J.N. VanDriel, and J. Wickham, 2007. Completion of the 2001 National Land Cover Database for the conterminous United States, *Photogrammetric Engineering & Remote Sensing*, 73(4):337–341.
- Huang, C., L. Yang, B. Wylie, and C. Homer, 2001. A strategy for estimating tree canopy density using Landsat-7 ETM and high resolution images over large areas, *Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry*, 05-07 November, Denver, Colorado, unpaginated CD-ROM.
- Isaaks, E.H., and R.M. Srivastava, 1989. *An Introduction to Applied Geostatistics*, Oxford University Press, pp 561.
- Liaw, A., and M. Wiener, 2002. Classification and regression by RandomForest, *R News*, 2(3):18–22.
- Moisen, G.G., J.W. Coulston, B.T. Wilson, W.B. Cohen, and M.V. Finco, 2012. Choosing appropriate subpopulations for modeling tree canopy cover nationwide, *Proceedings of Monitoring Across Borders: 2010 Joint Meeting of the Forest Inventory and Analysis (FIA) Symposium and the Southern Mensurionists* (W. McWilliams and F.A. Roesch editors), e-General Technical Report, SRS-157, Asheville, NC, U.S. Department of Agriculture, Forest Service, Southern Research Station, pp. 195–200.
- Moisen, G.G., and T.S. Frescino, 2002. Comparing five modelling techniques for predicting forest characteristics, *Ecological Modelling*, 157:209–225.
- Molinaro, A.M., 2005. Prediction error estimation: A comparison of resampling methods, *Bioinformatics*, 21(15):3301–3307.
- Popescu, S.C., R.H. Wynne, and J.A. Scrivani, 2004. Fusion of small-footprint lidar and multispectral data to estimate plot level volume and biomass in deciduous and pine forests in Virginia, USA, *Forest Science*, 50(4):551–565.
- Powell, S.L., S.P. Healey, W.B. Cohen, R.E. Kennedy, G.G. Moisen, K.B. Pierce, and J.L. Ohmann, 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches, *Remote Sensing of Environment*, 114(5):1053–1068.
- Rollins, M.G., and C.K. Frame (editors), 2006. *The LANDFIRE Prototype Project: Nationally Consistent and Locally Relevant Geospatial Data for Wildland Fire Management*, USDA Forest Service, General Technical Report RMRS-GTR-175, Rocky Mountain Research Station, Fort Collins, Colorado, 416 p.
- Schlather, M., A. Malinowski, M. Oesting, D. Boecker, K. Strokorb, S. Engelke, J. Martini, F. Ballani, P.J. Menck, S. Gross, U. Ober, K. Burmeister, J. Manitz, P. Ribeiro, R. Singleton, B. Pfaff, and R Core Team, 2014. *RandomFields: Simulation and Analysis of Random Fields*, R package version 3.0.44, URL: <http://CRAN.R-project.org/package=RandomFields> (last date accessed: 06 January 2016)
- Wang, G., G.Z. Gertner, S. Fang, and A.B. Anderson, 2005. A methodology for spatial uncertainty analysis of remote sensing and GIS Products, *Photogrammetric Engineering & Remote Sensing*, 71(12):1423–1432.
- Weisberg, S., 1985. *Applied Linear Regression*, John Wiley & Sons, 324 p.
- Wieland, M., and M. Pittore, 2014. Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images, *Remote Sensing*, 6:2912–2939: doi:10.3390/rs6042912
- Wilson, B.T., C.W. Woodall, and D.M. Griffith, 2013. Imputing forest carbon stock estimates from inventory plots to a nationally continuous coverage, *Carbon Balance and Management*, 8:1, URL: <http://www.cbmjournal.com/content/8/1/1> (last date accessed: 06 January 2016).

(Received 25 January 2015; accepted 12 May 2015; final version 21 September 2015)