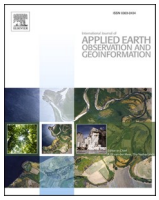
International Journal of Applied Earth Observations and Geoinformation 101 (2021) 102353

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Modeling tree canopy height using machine learning over mixed vegetation landscapes

Hui Wang [a,*], Travis Seaborn [b], Zhe Wang [c], Christopher C. Caudill [b], Timothy E. Link [d]

[a] Institute for Modeling Collaboration and Innovation, University of Idaho, Moscow, Idaho 83844, United States
[b] Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID 83844, United States
[c] Department of Geography, University of Idaho, Moscow, ID 83844, United States
[d] Department of Forest, Rangeland, and Fire Sciences, University of Idaho, Moscow, ID 83844, United States

ARTICLE INFO

Keywords:
Tree canopy height
LiDAR
Random forest
Spatial non-stationarity
Landsat

ABSTRACT

Although the random forest algorithm has been widely applied to remotely sensed data to predict characteristics of forests, such as tree canopy height, the effect of spatial non-stationarity in the modeling process is oftentimes neglected. Previous studies have proposed methods to address the spatial variance at local scales, but few have explored the spatial autocorrelation pattern of residuals in modeling tree canopy height or investigated the relationship between canopy height and model performance. By combining Light Detection and Ranging (LiDAR) and Landsat datasets, we used spatially-weighted geographical random forest (GRF) and traditional random forest (TRF) methods to predict tree canopy height in a mixed dry forest woodland in complex mountainous terrain. Comparisons between TRF and GRF models show that the latter can lower predefined extreme residuals, and thus make the model performance relatively stronger. Moreover, the relationship between model performance and degree of variation of true canopy height can vary considerably within different height quantiles. Both models are likely to present underestimates and overestimates when the corresponding tree canopy heights are high (>95% quantile) and low (<median), respectively. This study provides a critical insight into the relationship between tree canopy height and predictive abilities of random forest models when taking account of spatial non-stationarity. Conclusions indicate that a trade-off approach based on the actual need of project should be taken when selecting an optimal model integrating both local and global effects in modeling attributes such as canopy height from remotely sensed data.

1. Introduction

Vegetation features have large impacts on the physical, chemical and biological characteristics of aquatic ecosystems (Allan, 2004; Roth, Westhoff, Huwald et al., 2010). Among the functional properties of vegetation cover, tree canopy height controls heat exchange processes by regulating wind speed, relative humidity, and both shortwave and longwave radiation between the atmosphere and land surface (Hardy, Melloh, Koenig et al., 2004; Klos,Link, 2018; Tao, Guo, Li et al., 2016), and this can impact watershed-scale diurnal stream temperatures depending on shading effects from the riparian tree canopy (DeWalle, 2010; Loicq, Moatar, Jullian et al., 2018; Roth et al., 2010). Therefore, a high-quality and effective measurement of tree canopy height is essentially important for not only forest management but also stream habitats protection and restoration, such as large wood recruitment (Schuett-Hames,Roorbach,Conrad, 2012).

Methods for collecting data on tree canopy features vary widely depending on factors such as scale and extent of study, available resources, and physical or logistical restrictions. For example, field measurements provide direct and accurate quantification but are expensive and often limited in spatial extent and challenging in certain regions (Buckley,Isebrands,Sharik, 1999; McIntosh,Gray,Garman, 2012). In such cases, remote sensing (RS) techniques can be adopted to provide accurate, cost-effective and multi-spectral data at diverse spatial resolutions. Airborne Light Detection and Ranging (LiDAR) technology has been used to achieve 3D depiction of tree canopy cover, including both horizontal measurements and vertical structural information (Hudak, Lefsky, Cohen et al., 2002; Singh, Vogler, Shoemaker et al., 2012; Zald, Ohmann, Roberts et al., 2014). While this 3D depiction may lead to a good performance of forest cover assessment, LiDAR also has

* Corresponding author at: Institute for Modeling Collaboration and Innovation, 875 Perimeter Drive, MS-1122, Moscow, Idaho 83844, United States.
E-mail address: huiwang@uidaho.edu (H. Wang).

https://doi.org/10.1016/j.jag.2021.102353
Received 26 November 2020; Received in revised form 19 March 2021; Accepted 26 April 2021
Available online 7 May 2021

disadvantages including cost of collecting and processing data and limited spatial and temporal coverage. As characteristics of forest canopies, such as height and density, are correlated with spectral indexes of Landsat imagery (Huang, Yang, Wylie et al., 2001; Karlson, Ostwald, Reese et al., 2015), an important application of LiDAR-derived datasets is for the calibration and validation of models for Landsat datasets with larger spatiotemporal extents to achieve mutual complementarity. One potential method for calibrating and validating is through random forest models.

Random forest, a decision tree-based ensemble learning method, has been considered as an effective algorithm for both classification and regression modeling analysis, even with non-parametric and non-linear datasets (Belgiu,Drăguţ, 2016; Karlson et al., 2015). In addition to providing high accuracy, random forest algorithms can handle complex data with a variety of features. Therefore, numerous studies have investigated the possibility of using LiDAR-derived data to train random forest models to predict tree canopy height from satellite imagery features and have achieved reasonable results. For instance, Staben,Lucieer and Scarth (2018) predicted tree canopy height over different vegetation communities to understand their transitions between 1987 and 2016 by developing a random forest model with LiDAR-derived canopy height, vegetation indices and spectral band ratios from Landsat images. Ota, Ahmed, Franklin et al. (2014) integrated Landsat and LiDAR data to model mean canopy height for a tropical forest in Cambodia. The research demonstrated that predictive models can achieve the highest accuracy when including forest disturbance and recovery metrics from Landsat derived data in a novel random forest model. Ahmed et al. (2015) demonstrated that random forest models for canopy structure incorporating disturbance history information derived from LiDAR data outperformed multiple regression models in terms of three different forest classes, namely mature, young and their combination, with substantially lower statistical errors.

Although random forest algorithms have been widely recognized to have good performance and applied in a range of domains, such as urban planning and forest management, past attempts usually neglected spatially non-stationary effect in distributions of sample data. Such unweighted spatial modeling processes may make the result problematic given the potential for spatially-varying relationships between dependent and independent variables (i.e., non-stationarity). Previous research showed that non-stationarity in remote sensing-derived data, such as clustered distribution of land use/cover classification, may lead to a less accurate predictive models (Wang,Stephenson,Qu, 2019). Thus, training datasets may also be unable to properly reflect the relationship between predictive and explanatory variables at a local scale due to spatial heterogeneity. Moreover, the range of predicted values can be restricted by the selected training dataset, and variance of feature values may not be fully taken into consideration in the modeling process. This is due to the partial coverage of study area and thus a possible unequal distribution pattern of selected training data. Georganos, Grippa, Niang Gadiaga et al. (2019) developed an improved and straightforward model for address these challenges, namely Geographical Random Forest (GRF), to predict population density in Dakar, Senegal in Africa. The model uses Geographical Weighted Regression (GWR) as a reference and generates a local random forest model in each observational location. Then, the closest local model was used for its neighbors at validation and prediction steps.

While spatial non-stationarity has been recognized as a potential issue when applying random forest methods, an in-depth analysis of this issue integrated with tree canopy height prediction is still lacking (Sekulić, Kilibarda, Heuvelink et al., 2020; Staben et al., 2018). Specifically, assessments of the effect of non-stationarity on machine learning modeling outcomes of remotely sensed data would benefit an array of researchers interested in landscape level data where LiDAR is not readily available. Our overall objective was to assess the role of spatial non-stationarity in modeling tree canopy height from Landsat data by using traditional random forest (TRF) and GRF models trained and validated with independent subsets of LiDAR data. Our specific goals were: (1) to investigate whether the accuracy of predicted tree canopy height could be improved or not by taking into account spatial variance; (2) to detect spatial autocorrelation patterns of residuals in modeling tree canopy height by using TRF and GRF separately and (3) to discover whether the predictive ability of TRF and GRF models would be associated with tree canopy height or not. We hypothesized that GRF would improve model fit to some degree and that tradeoffs may exist between modeling effort and accuracy when using the two approaches in a mixed vegetation landscape.

2. Materials and methods

2.1. Study area

The Payette National Forest is one of the protected lands managed by the United States federal government, and covers approximately 9300 km^2 in the western USA (Fig. 1). It is specifically located in the Hitt Mountains of Payette National Forest in the state of Idaho, USA, and was selected because the area contains a mix of open, low, and tall canopy land covers in complex terrain. The distribution pattern of true canopy height for testing dataset was plotted in quantile format (Fig. 2). Less training data in high quantiles (95%: 19.53 m; 99%: 23.30 m) reinforced our suspicion that predictive ability of a random forest model would be impacted by the limited samples for training. The study area was in the upper reaches of the Mann Creek watershed within the Columbia River basin at altitudes of ca.1300 m to 1500 m. The landscape is composed of interspersed forest, sagebrush communities, high mountain ridges and deep canyons. Forested areas contain eight species of conifer trees, including ponderosa pine and Douglas-fir (~1000 m - ~1800 m) and whitebark pine (~2300 m) (Lund, 2004), with the forest mainly containing coast Douglas-fir reaching at a maximum height of 55 m (Carder, 1995). Riparian areas are dominated by evergreen forest, shrub/scrub, sagebrush, etc. Vegetation type and canopy height are of interest to conservation and management of this watershed because of the effects of riparian vegetation on stream water quality and temperature. Climate at the site is “Warm-summer Mediterranean” climate type under Köppen Climate Classification, with precipitation of ca. 1270 mm per year (Peel, Finlayson,McMahon, 2007).

2.2. Data acquisition

2.2.1. LiDAR point cloud data

Airborne LiDAR technology has been widely used to retrieve three-dimensional information from the features on the Earth’s surface by emitting and recording laser pulses. The Canopy Height Model (CHM) represents the height of objects above the ground that a LiDAR system observes, reflecting the distance between the Earth surface and the top of the trees in the present research (Khosravipour, Skidmore, Wang et al., 2015). Feature positional information, including coordinates (horizontal) and elevation (vertical), can be collectively stored in a point of data wherever the LiDAR laser pulse hits. A large amount of these individual points eventually forms a LiDAR point cloud dataset, providing a detailed three-dimensional basis for generating and rasterizing CHMs. In this study, the LiDAR point cloud data comprised of 4 (column) by 3 (row) image sets (Fig. 1) were obtained from the National Geospatial Program developed by the U.S. Geological Survey (USGS) in the American Society for Photogrammetry and Remote Sensing (ASPRS) LAS format. Original downloaded dataset was a zipped LAZ file and can be uncompressed through a laszip function embedded in the LAStools software (Isenburg, 2012). The point cloud data were collected from September 9th to October 14th, 2017. Traditional method of developing a CHM is to do a raster subtraction between the Digital Surface Model (DSM) representing top of the surface and Digital Terrain Model (DTM) representing ground elevation (Pitkänen et al., 2004). To remove the possible empty pixels in this subtraction, a triangulation-based pit-free

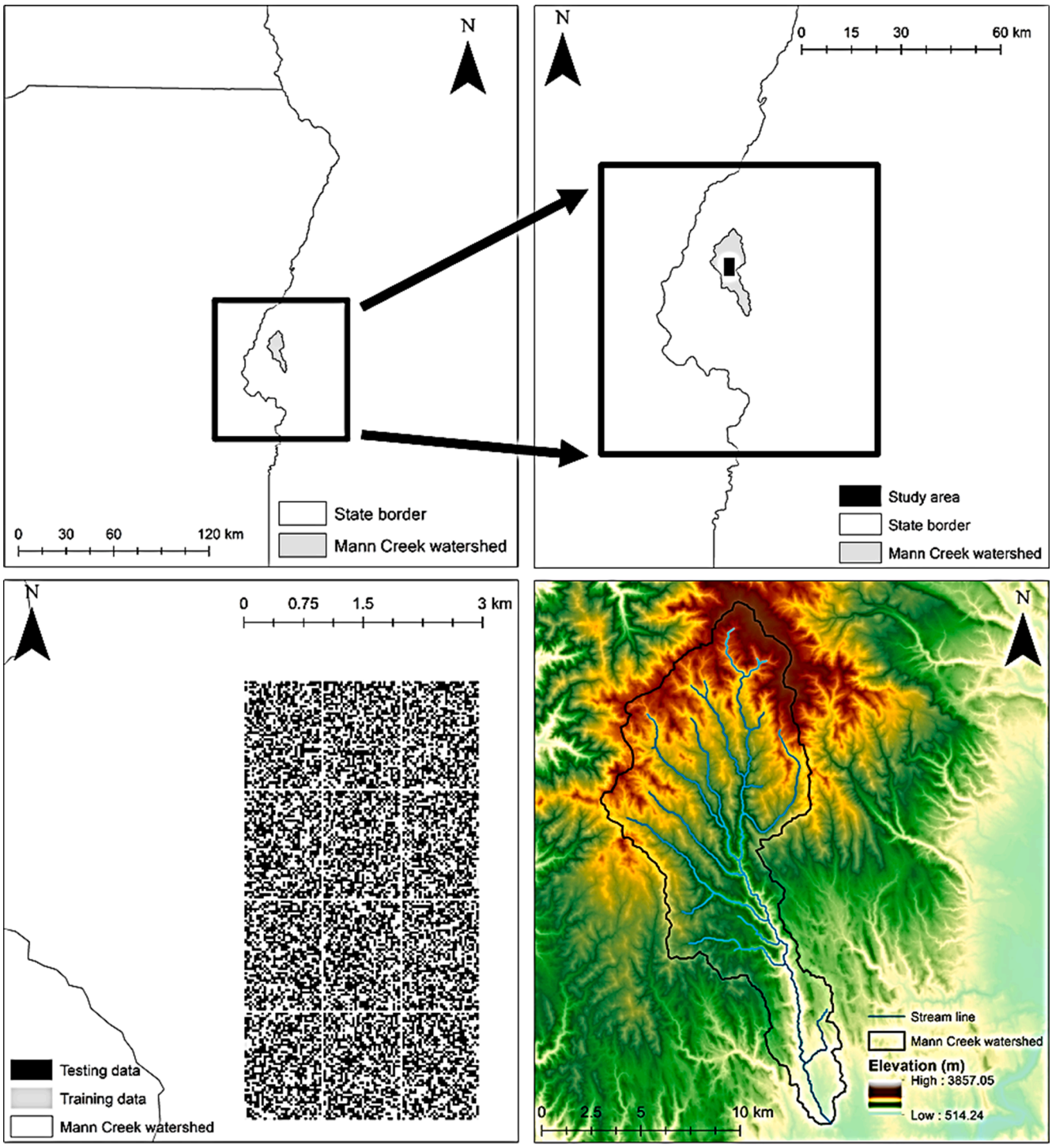


Fig. 1. Study area and spatial distribution of testing and training datasets.

algorithm of Khosravipour, Skidmore, Isenburg et al. (2014) was used to generate the rasterized CHM through the lidR package in R (v.3.6.3). Final resolution of the CHM derived from the point cloud data was 0.25 m by 0.25 m. We note that the outermost pixels for each point cloud derived raster were eliminated in the process of creating the CHM due to the edge effect of LiDAR data. We calculated the mean value of the LiDAR CHM for each 30 m × 30 m pixel corresponding to each Landsat pixel as a reference of true canopy height (Pascual et al., 2010). Since the true canopy height is defined as LiDAR-derived canopy height, it is worth mentioning that this reference cannot avoid some error associated with the derivation process and determination of the ground surface (Tinkham, Smith, Hoffman et al., 2012). Post-hoc, after modeling was complete, we excluded vegetation, primarily sagebrush which occurs in relatively homogenous stands in uplands, less than or equal to 1 m in further analysis.

2.2.2. *Landsat imagery*

The Landsat 8 Operational Land Imager (OLI) level 2 product, which has a spatial resolution of 30 m × 30 m, was obtained from the USGS Earth Explorer website (https://earthexplorer.usgs.gov/). The Landsat 8 OLI level 2 product which has been atmospherically corrected provides the land surface reflectance (LSF) data derived from the level 1 product

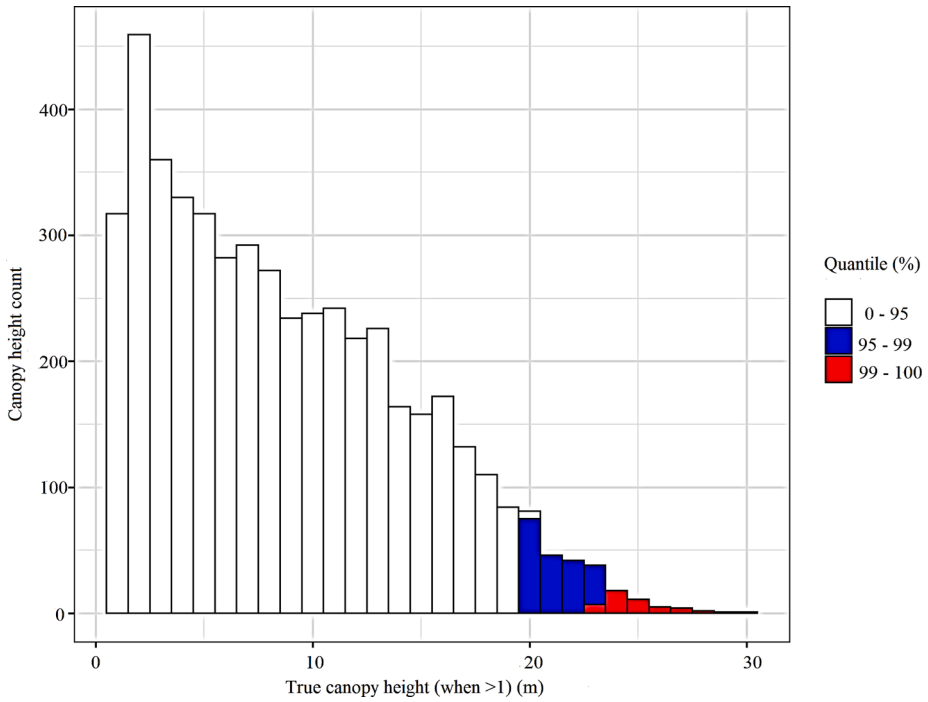


Fig. 2. Quantile distribution of true canopy height as derived from LiDAR for training dataset.

(Masek et al., 2006). The acquisition date was October 6th, 2017 in order to eliminate cloud effects and to be effectively synchronous with the date of the LiDAR data acquisition.

The reflectance of a vegetated surface depends on the structural and optical properties of the vegetation and underlying soil (Myneni et al., 1995). However, the blue band is sensitive to the aerosol conditions in the atmosphere (Flood, Danaher, Gill et al., 2013). Therefore, only green, red, near-infrared (NIR), short-wave infrared 1 (SWIR1), and short-wave infrared 2 (SWIR2) LSF bands in the acquired dataset were used to calculate the vegetation indices and band ratios. To cover as many spectral indices as possible for correlation and feature importance tests for running random forest models, a total of 27 parameters, including 5 LSF bands, 12 vegetation indices, and 10 band ratios calculated from a combination of 5 LSF bands were used as input predictor variables (Table 1) to predict the canopy height from Landsat data. Parameters (i.e., CVI, GDVI, Band Ratios and LSF bands) were then normalized to a range of −1 to 1 to reduce data redundancy and standardize values across parameters.

2.3. Random forest

2.3.1. Traditional random forest

As an aggregated machine learning algorithm, random forest algorithms are comprised of decision trees and Classification And Regression Trees (CART) on a basis of bootstrap aggregation ("bagging" for short) (Breiman, 2001). Compared with a single CART, random forest less likely suffers from the overfitting problem as it eventually takes the average of the predicted values across decision trees from the bootstrap iterations. Generally speaking, there are four basic stages when running the random forest model: (1) select n random samples from the given dataset for training; (2) build a decision tree for each selected sample by integrating features; (3) set up the optimal number of decision trees based on time cost and computational ability; (4) take the average of the predicted value from each tree. Before applying the algorithm, it is crucial to make a right decision on the selection of the training and validation datasets in order to proceed to the next step. Fig. 1 shows that the pixels used for training and validation were randomly distributed within the extent of our study area. Moreover, to reduce the underlying

Table 1
Vegetation indices, band ratios and land surface reflectance bands used in this study.

Vegetation Index	Formula
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - Red}{NIR + Red}$ (Tucker, 1979)
Green Soil Adjusted Vegetation Index (GSAVI)	$\frac{NIR - Green}{(NIR + Green + 0.5)*(1 + 0.5)}$ (Sripada et al., 2006)
Green Normalized Vegetation Index (GNDVI)	$\frac{NIR - Green}{NIR + Green}$ (Buschmann and Nagel, 1993)
Chlorophyll Vegetation Index (CVI)	$\frac{NIR}{Green} * \frac{Red}{Green}$ (Vincini et al., 2008)
Normalized Difference Greenness Index (NDGI)	$\frac{Green - Red}{Green + Red}$ (Bannari et al., 1995)
Normalized Burn Ratio SWIR2 (NBR)	$\frac{NIR - SWIR2}{NIR + SWIR2}$ (Ji et al., 2011)
Normalized Burn Ratio SWIR1 (NDII)	$\frac{NIR - SWIR1}{NIR + SWIR1}$ (Ji et al., 2011)
Green Difference Vegetation Index (GDVI)	NIR −Green (Sripada et al., 2006)
Modified Soil Adjusted Vegetation Index (MSAVI)	$(2*NIR + 1 - \sqrt{(2*NIR + 1)^2 - 8*(NIR - Red)})/2$ (Qi et al., 1994)
Difference Vegetation Index (DVI)	NIR −Red (Tucker, 1979)
Soil adjusted Vegetation index (SAVI)	$\frac{NIR - Red}{(NIR + Green + 0.5)*(1 + 0.5)}$ (Huete, 1988)
Modified Simple Ratio (MSR)	$[(NIR/RED) - 1]/[(\sqrt{NIR/RED}) + 1]$ (Chen, 1996)
Band Ratio	Red/Green; SWIR1/NIR NIR/Green; SWIR2/Green NIR/Red; SWIR2/Red SWIR1/Green; SWIR2/NIR SWIR1/Red; SWIR2/SWIR1
LSF bands	Green; Red; NIR; SWIR1; SWIR2

autocorrelation between these two datasets, pixels were also equally dispersed over the space. Additionally, to cover as many of the height samples as possible over different landscape units, we consequently selected the same number of pixels for training and validation datasets, resulting in a total of 8688 pixels for each. The TRF algorithm was implemented using the open source Scikit-learn module in Python (Pedregosa, Varoquaux, Gramfort et al., 2011).

Calibration of the model was divided into three main stages. First, as we previously mentioned, 27 vegetation indices and band ratios were applied in the present research following the study of Staben et al. (2018). One difference, however, is that a feature selection technique was employed to discover and omit those irrelevant predictor variables. It has been widely recognized that removing highly correlated features with less importance is able to reduce training time and overfitting issues (Rogers,Gunn, 2005). Therefore, a correlation matrix and feature importance tool provided by the Scikit-learn module were applied to detect collinearity and importance score (i.e., Gini importance or Mean Decrease Impurity) of variables, respectively (Louppe, Wehenkel, Sutera et al., 2013). Second, the number of decision trees used in the random forest algorithm (n_estimators) was optimized by assessing the relationship between generalization error and model complexity (Breiman, 2001; Nadeau,Bengio, 2000). Generally, a larger number of decision trees provides better training, but previous studies also demonstrated that predictive accuracy of the model will eventually converge when the number of decision trees continues to be increased while still increasing the computational time (Oshiro,Perez,Baranauskas, 2012; Pal, 2005). To find an optimal number of decision trees, therefore, users usually adjust the input value of a parameter (model complexity) to achieve the lowest generalization error. Similarly, at the last step, this research also investigated the best number of features at each split (max_features), the maximum depth of each tree (max_depth), the minimum number of samples to split each node (min_samples_split) and the minimum number of samples for a leaf node (min_samples_leaf) by obtaining and comparing the score of 10-fold cross-validation through sklearn package in Python (Ojala,Garriga, 2010). In addition, the automatic tuning process of these hyperparameters is on the basis of grid-search method over a parameter grid (Paper, 2020). After achieving the optimal model, the calibration result was evaluated, and validation data were tested by using the root mean squared error (RMSE, Equation (1)) and variance of error (VE, Equation (2)).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum (p_i - x_i)^2} \tag{1}$$

$$\mathrm{VE} = \frac{1}{n-1}\sum_{i=1}^{n} (e_i - \bar{e})^2 \tag{2}$$

In Equations (1) and (2), n represents the number of sample points, p_i and x_i were the predicted and observed or true canopy height at point i, respectively. e_i denotes the difference between p_i and x_i at point i while $\bar{e}$ is the average of errors for all sample points.

2.3.2. Geographical random forest

GRF is most appropriate when spatial heterogeneity is present in covariates and model predictions are biased because the residuals are spatially autocorrelated (Hengl, Nussbaum, Wright et al., 2018). Previous studies demonstrated that these issues can be addressed using local models with simulations taking account of neighboring effects (Wang et al., 2019). Therefore, this study applied an existing spatial model GRF integrated with the geographically weighted regression (GWR) model and random forest to detect spatial autocorrelation patterns of residuals in modeling tree canopy height. The residual was calculated by Equation (3):

$$x_i - p_i = \mathrm{Residual} \tag{3}$$

The core working mechanism of GRF is similar to that of the GWR, with respect to bandwidth and kernel selections (Fotheringham,Brunsdon,Charlton, 2003). In bandwidth selection, an adjustable kernel, the maximum radius at which a target point can reach, was used to fit in various spatial density of data points. Compared with the fixed kernel which uses a constant distance as the bandwidth, the bandwidth of the so-called adaptive kernel in GRF is equivalent to the distance to the n^{th} nearest neighbor (Wang et al., 2019). Due to the unequal distribution of vegetation over semi-arid mountainous watersheds, the adaptive kernel is able to capture as much information as possible to account for the aggregated tall vegetation on north-facing slopes along streams in the present research. In the case of random forest algorithm, a specific local model will be calculated for each training sample at location i. A n number of neighboring points were taken into consideration when calibrating the corresponding GRF for the sample point. Thus, the major difference between TRF (Equation (4)) and GRF (Equation (5)) is the dimensionality over space.

$$p_i = \beta_i x_i + \cdots + e_i \tag{4}$$

$$p_{i(a,b)} = \beta_{i(a,b)} x_i + \cdots + e_{i(a,b)} \tag{5}$$

In Equations (4) and (5), p_i represents the predicted canopy height at location i, the other contribution of selected Landsat-derived features besides $\beta_i x_i$ is omitted in the middle, e_i is the prediction error at the corresponding site and (a, b) represents the coordinates at location i. In the calibration of GRF, the two most important parameters regarding number of decision trees and selected features, namely n_estimators and max_features, follow the calibrated results of TRF applied to the same dataset. At the validation stage, a new target of sample pixels only uses the closest tuned GRF model to predict the corresponding mean canopy height. The GRF algorithm has been developed in R by Kalogirou and Georganos (2018) as the R package 'SpatialML' published as a version of 0.1.3 in 2019.

To avoid a highly biased and discrete prediction, the GRF is designed to allow for partial weighting of the local model versus global model. For example, an equal weighting fusion represents 50% weights from local model and 50% weights from global model (hereafter, "50% LM"). In this study, we examined the model fits of various weightings of TRF vs. GRF by comparing configurations of eleven levels of local weighting from 0% LM (TRF) to 100% LM at ten percent intervals. We evaluated the fit among models using the three statistical measurements used for TRF, along with Moran's I, calculated to evaluate the degree of spatial autocorrelation of residuals for each combination type (Moran, 1948).

3. Results

3.1. Parameter selection and model optimization

After feature selection and model optimization, Table 2 shows the ranking of importance of all initial 27 features and decisions of whether to retain or omit them. The extent of correlation between every two features was reflected by the Pearson coefficient in the correlation matrix with heatmap (Figure S1). In comparison, the screening process follows a criterion that features with a coefficient score equal or>0.96 and lower ranking of importance were ruled out. We eventually retained 12 features as the inputs when building decision trees for the random forest models. The Modified Simple Ratio (MSR), a function of Near Infrared (NIR) and Red band, was demonstrated to be the most important feature in the dataset, with an importance score of 0.1683. After weighting the correlation and importance attributes in pairs, 4 band ratios and 11 vegetation indexes were then excluded in the training process of the random forest procedure (Table 2). For tuning the most influential parameter, n_estimators, the initial value of this non-zero parameter was set to 1, leading to the lowest cross validation score of 0.7022. Testing numbers fall into the range between 1 and 1951, with an interval of 50. The variation of testing results tends to be stable when it reaches to the last 9 values ranging, from 1551 to 1951 (Fig. 3A). The

Table 2
Features retained and omitted for random forest to estimate canopy height for Landsat data after training with LiDAR data.

Feature	Importance score	R or O	Feature	Importance score	R or O
MSR	0.1683	R	NIR/Green	0.0092	O
MSAVI	0.1565	R	SWIR2/Red	0.0079	R
SAVI	0.1552	O	CVI	0.0075	R
NDVI	0.1483	O	NDGI	0.0074	O
NIR/Red	0.1012	O	Red/Green	0.0072	O
NDII	0.0556	O	SWIR2/Green	0.0070	R
SWIR1/NIR	0.0297	R	SWIR1/Red	0.0070	R
NBR	0.0259	O	NIR	0.0059	R
SWIR2/NIR	0.0201	O	SWIR2	0.0059	O
DVI	0.0122	R	GDVI	0.0057	O
SWIR2/SWIR1	0.0120	R	Red	0.0054	O
Green	0.0109	R	SWIR1	0.0054	O
SWIR1/Green	0.0106	R	GSAVI	0.0025	O
GNDVI	0.0093	O			

Note: R denotes retained feature and O denotes omitted feature.

cross-validation scores plateaued at values above ~ 1000. Fig. 3B shows the results of a smaller scope of values ranging from 1480 to 1519, with the highest cross validation score (0.84649) occurring when the number of decision trees was 1500 (Fig. 3B). Based on the fixed n_estimators at 1500, the optimal values of max_depth (9), min_samples_leaf (13), min_samples_split (2), and max_features (7) were then obtained individually by tuning the model complexity to reach the highest cross validation score of 0.85026.

3.2. Model validation and predictive performance

As a base model, we examined the ability of the TRF to predict canopy height using Landsat data by comparing the predictions for 8688 points to an independent subset of the LiDAR data (Fig. 4). Linear regression resulted in a slope and intercept of 0.9952 and 0.0446, respectively, reflecting a near 1:1 relationship. Moreover, the high coefficient of determination value ($R^2 = 0.8714$) implies that the plotted data were close to the fitted regression line, while a dispersive pattern indicates that the predicted and true canopy height dataset of the TRF model may have a slightly higher variance at higher height values.

TRF-GRF fusion models were generated after selecting a distance of 56 neighboring points as the adaptive kernel when calibrating the local models. Comparison of model fit among the fusion models revealed that RMSE, VE, and R^2 obtained their minimum (2.0338), minimum (4.1330) and maximum (0.8889) values at 50% LM level, respectively (Table 3). The assessment of spatial statistics, namely Moran's I and the Z score, shows that there were some discrepancies in the spatial distribution patterns of residuals between the model configurations, reflecting that the explanatory ability due to spatial non-stationarity was quite variable. Moreover, the extremely small P values further confirm that these positive spatial autocorrelations (clustered patterns) were significant (Table 3). In spite of being significantly aggregated, the extent of being spatially autocorrelated was apparently mitigated when the 100% LM was applied.

We compared the spatial distribution patterns of residuals between the TRF (0% LM) model and 100% LM model to visually evaluate the mitigating influence of the GRF on spatial non-stationarity (Fig. 5). As a whole, the degree of spatial autocorrelation becomes weaker when the 100% LM was employed (Table 3). Transitions can be more evident in the Section A and the first two rows of Section B, where more trees have grown along streams within the Mann Creek watershed. Notably, although the transition shows that most clustering can be mitigated by applying GRF, certain areas with extremely high or low residuals are barely changed. For example, areas with extremely low and high residuals were hardly mitigated even after applying the 100% LM. Visually, we note extreme residuals exist primarily in the areas with very high and low canopy heights and were moderately mitigated by models with high GRF weighting. To define the scope of extreme residuals, we used quantile statistics to visualize their distributions and extract those values in high quantiles. Fig. 6 depicts the quantile distribution of the residual for the TRF model and 100% LM. Table 4 further shows more details with respect to quantile statistics. The 95% and 99% quantiles residuals from the TRF model were 3.83 m and 6.97 m for true canopy height, respectively. It is worth mentioning that the middle bin of the histogram was omitted in the figure due to the large quantity (>4000) of 0 values. In this case, residuals less than the 5% quantile and greater than the 95% quantile were defined as the extremes in the present study. Similarly, to set thresholds for defining low and high canopy heights, quantiles of true canopy height were also detailed. As a result, the 1%, 5%, 95%, and 99% quantiles of true canopy height were 1.06 m, 1.38 m, 19.53 m and 23.30 m, respectively. Additionally, a five-number summary was used to provide more information regarding the distribution of the observations. The minimum, 1st quantile, median, 3rd quantile and maximum were 1 m, 3.72 m, 7.78 m, 12.88 m and 29.51 m, respectively. For convenience of further comparisons, those canopy heights lower than the median (<7.78 m) and higher than the 95% quantile (>19.53

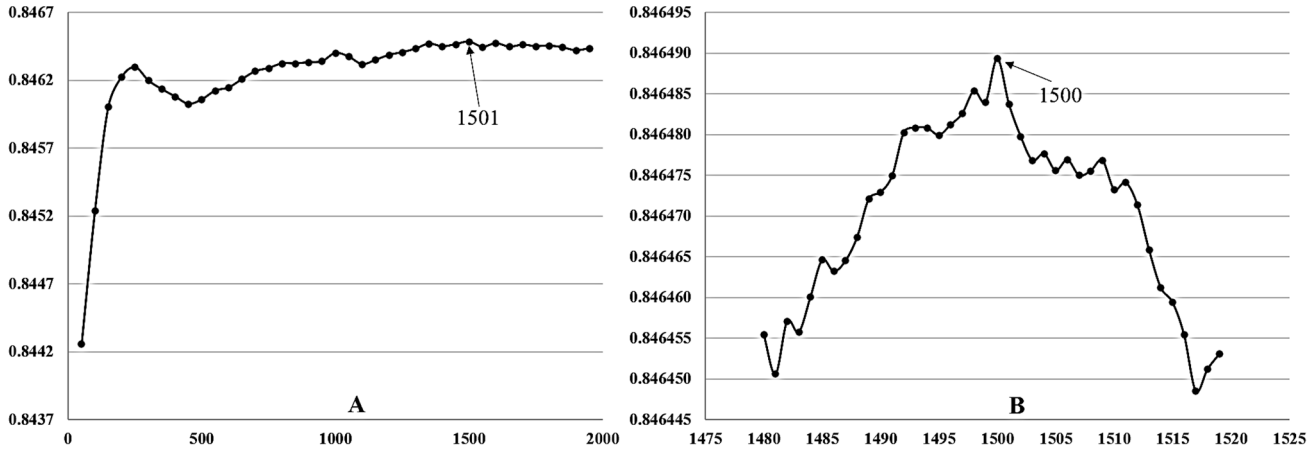


Fig. 3. Test results of n_estimators (number of decision trees) in calibration for TRF model (x axis represents the number of decision trees used for testing; y axis represents the cross-validation score). The initial testing value, n_estimators = 1, has been removed from Fig. 3A as its corresponding cross-validation score (0.7022) is out of the listed range.

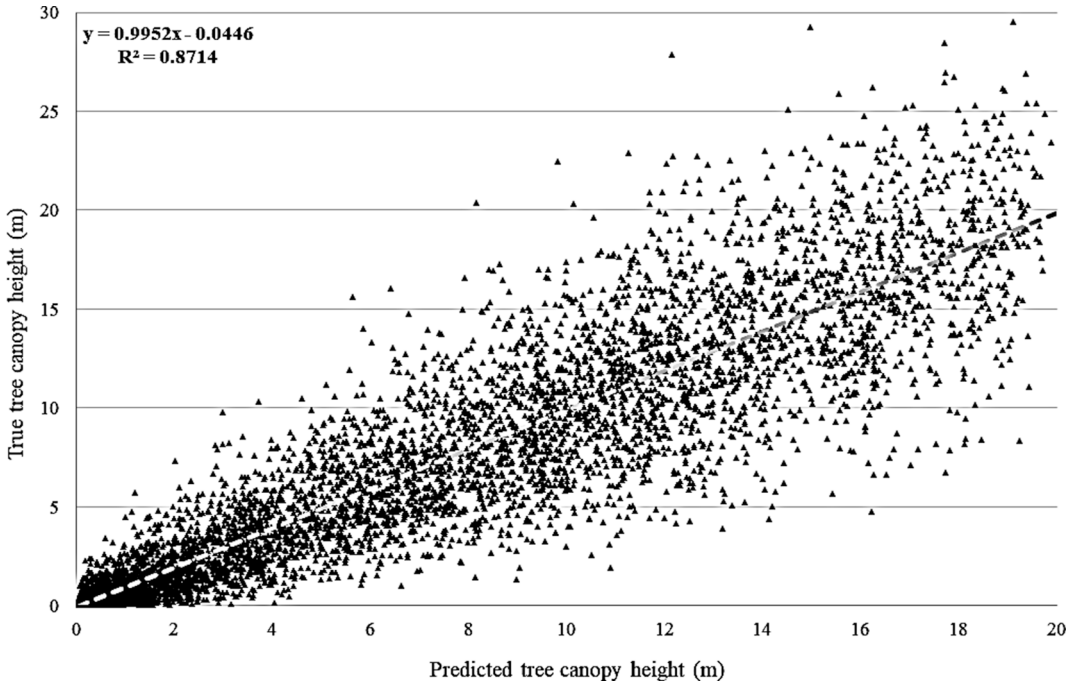


Fig. 4. Validation results of TRF model.

Table 3
Statistical analysis of different configuration formats of local and global models.

	RMSE	VE	R^2	Moran's I	Z score	P value
100% LM	2.1481	4.6127	0.8759	0.0957	18.4671	<0.01
90% LM	2.1037	4.4236	0.8809	0.0972	18.7682	<0.01
80% LM	2.0696	4.2812	0.8848	0.1010	19.5004	<0.01
70% LM	2.0464	4.1852	0.8875	0.1072	20.6834	<0.01
60% LM	2.0344	4.1359	0.8888	0.1156	22.3085	<0.01
50% LM	2.0338	4.1330	0.8889	0.1261	24.3368	<0.01
40% LM	2.0446	4.1767	0.8877	0.1384	26.7013	<0.01
30% LM	2.0667	4.2670	0.8852	0.1519	29.3142	<0.01
20% LM	2.0996	4.4038	0.8815	0.1663	32.0774	<0.01
10% LM	2.1430	4.5872	0.8765	0.1809	34.8935	<0.01
TRF model	2.1961	4.8170	0.8714	0.1953	37.6746	<0.01

m) are hereafter defined as the "low canopy" and "high canopy", respectively.

3.3. Statistical analysis of TRF and GRF models

Our results show patterns about how canopy height impacts the variation of residuals of the two different models (Fig. 7). Since a residual was equivalent to the difference between true value and its corresponding predicted value, a positive residual resulted from underestimated canopy height from the model. For both models shown in Fig. 7, although positive residuals span a larger extent than the negative residuals over true canopy height, the number of negative values (TRF: 5377, 100% LM: 5332) was much larger than that of the positive (TRF: 3311, 100% LM: 3356), indicating that both the TRF model and 100% LM tend to overestimate the LiDAR-derived values. The expanded portions of Fig. 7, focusing only on low and high canopy heights, shown in Figures S2 and S3 respectively, provide more details regarding predictive abilities of the models. The corresponding statistics for low canopy height data show that the number of negative residuals (TRF: 4383, 100% LM: 4373) was greater than that of the positives (TRF: 1873, 100% LM: 1883), revealing that the models were inclined to overestimate heights when the true canopy was low. A locally weighted regression (loess) curve was also plotted to reveal the trends of the correlation which was not evident in the parametric model (Cleveland and Devlin, 1988). The loess regression curve shows little bias and almost no fluctuation over this parameter space. Interestingly, this finding was opposite than what we found from the relationship between the high canopy values and predictive abilities of the TRF model and 100% LM. Specifically, the high-quantile canopy heights were almost all positive values, indicating that the models were very likely to underestimate height when the true canopy was relatively tall (Figure S3). Moreover, the loess regression curve also implies that there was a positive correlation between residuals and high canopy values. In other words, the error of prediction of two models has a tendency of becoming larger at the places where the true canopy was higher. This conclusion applies to both the TRF model and 100% LM.

Switching coordinate axes, we investigated the similarities and discrepancies between the TRF model and 100% LM to understand how residual quantiles were distributed relative to true canopy height. Fig. 8 clearly shows that this correlation pattern was mainly driven by the combination of low canopy heights and small residuals. The clustered points within the 5–95% percentiles indicate overall good predictive performance for both the TRF model and 100% LM. The loess regression curve on this plot show that the true canopy heights were either relatively high or relatively low for residuals in opposing directions (i.e., positive and negative). In other words, most of extreme residuals occur for conditions where canopy heights were either relatively high or low. For the two different models, the slope of loess regression curve of the 5–95% quantile was much steeper than that of the 1–5% and 99–100% quantiles. Hence, we can draw a conclusion that the relationship between residual and degree of variation of true canopy height was relatively sensitive within the 5–95% percentile. Beyond these quantiles, a small variation of residuals may lead to a small change in true canopy height, especially in the positive residual variable space.

4. Discussion

To the best of our knowledge, this is an important study shedding new light on how to investigate the relationship between tree canopy height and predictive abilities of TRF and GRF models. To achieve the goal, we built different random forest models in both global and local

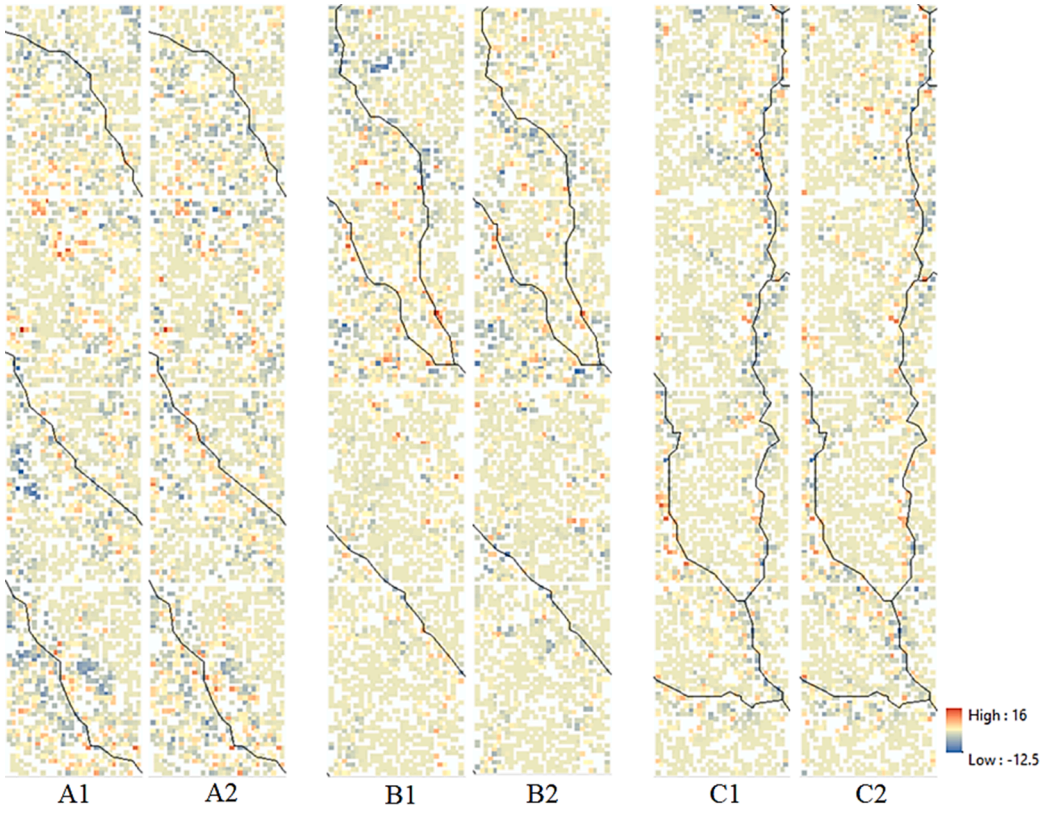


Fig. 5. Spatial distribution patterns of residuals for TRF model (0% LM) and 100% LM (A1, B1 and C1 are comprised of spatial distribution pattern of residual for TRF model; A2, B2 and C2 are comprised of spatial distribution pattern for 100% LM. Black lines indicate streams in the study area.)

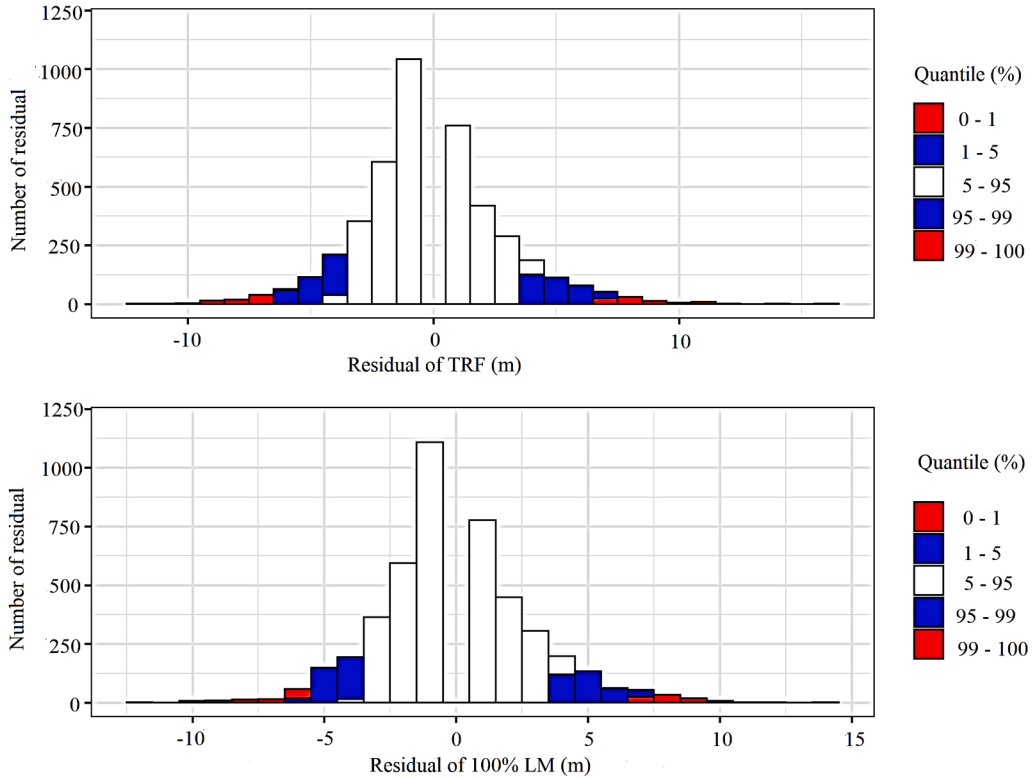


Fig. 6. Quantile distribution of residual for TRF model and 100% LM (Note: Residual 0 has been removed from the figure to better visualize quantiles).

Table 4
Statistical distribution of residuals for TRF model and 100% LM.

Quantile (%)	TRF model (m)	100% LM (m)
1	−6.40	−5.74
5	−3.63	−3.58
95	3.83	3.76
99	6.97	7.03

dimensions. The key findings of our study are: (1) the spatial autocorrelation can be mitigated in some degrees while the accuracy may not necessarily be improved by taking into account spatial variance; (2) the predictive abilities of TRF and GRF are closely associated with the tree canopy height over the target area. There are at least four primary findings from this work.

First, unlike most previous studies only focusing on one criterion for evaluation, here we used two, namely prediction accuracy and spatial predictive abilities, in order to select the optimal hybrid model for estimating canopy height (García et al., 2018; Ota et al., 2014). In this study, the 100% LM can be recognized as the best model to estimate the tree canopy height as our criterion primarily focuses on mitigation of spatial effect. Although there are discrepancies existing among the statistical indicators regarding numerical accuracy, these subtle differences can be negligible when only taking the spatial autocorrelation of residuals into consideration. Thus, a trade-off approach based on actual needs of a given project is highly desirable before making a selection on TRF-GRF fusion models. This approach covers a balance between maximizing model accuracy and minimizing spatial autocorrelation. In this study, for instance, a 50% LM would be chosen if the difference in Moran's I is negligible and researchers are less concerned with the mitigation of spatial clustering.

Second, models tend to produce under and overestimates when the corresponding true tree canopy heights are very high and low, respectively. The correlation between true canopy height and residuals also exhibits an overall increasing trend for both of the models. At the highest quantiles of canopy height, residuals of the TRF model and 100% LM likewise tend to increase as well, revealing that model performance is reduced at the most extreme height present in this dataset. In the meantime, however, neither an apparent increasing nor decreasing relationship is apparent for the lowest canopy heights, suggesting that the performance is relatively unpredictable at the extreme low end of the height distribution.

Third, the relationship between model performance and the degree of variation of true canopy height can vary between different quantiles. Based on our findings, the variation of true canopy height can possibly change either very slowly (e.g., the 99–100% quantiles) or rapidly (e.g., the 5–95% quantiles) in a certain scope of residual. This evidence clearly depicts how TRF and GRF models perform, in terms of its relationship with canopy height, within different quantiles. A steep change of true canopy height can probably be detected when models have relatively low errors, while the stable variation is associated with its weak performance, indicating that the elasticity of true canopy height to model performance varies across quantiles. Moreover, comparisons between the two models show a very similar overall trend of the loess regression curve, implying that mitigating the effect of spatial non-stationarity barely makes a difference in the degree of variation of tree canopy height.

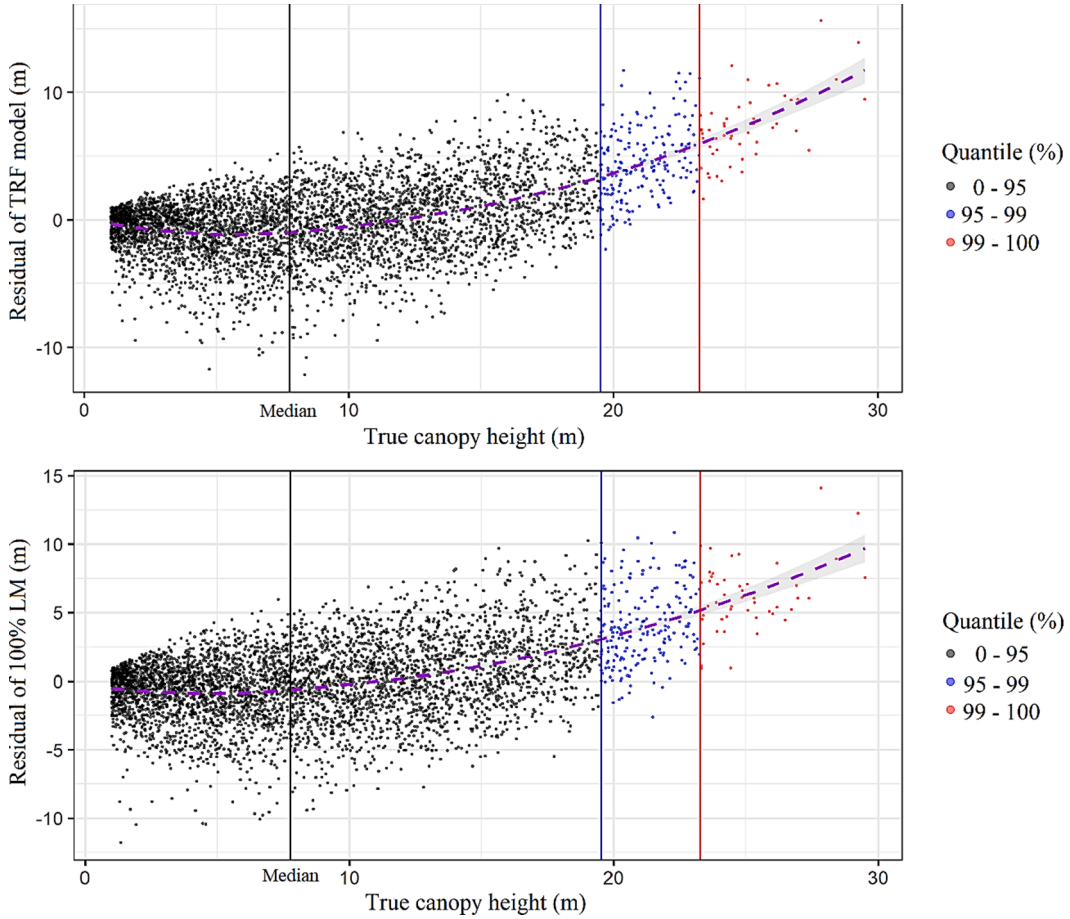


Fig. 7. Correlation between residual and true canopy height based on height quantile (the black, blue and red solid lines represent the borders of median, 95% quantile and 99% quantile of true canopy height, respectively. True canopy height minus predicted canopy height equals residual). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

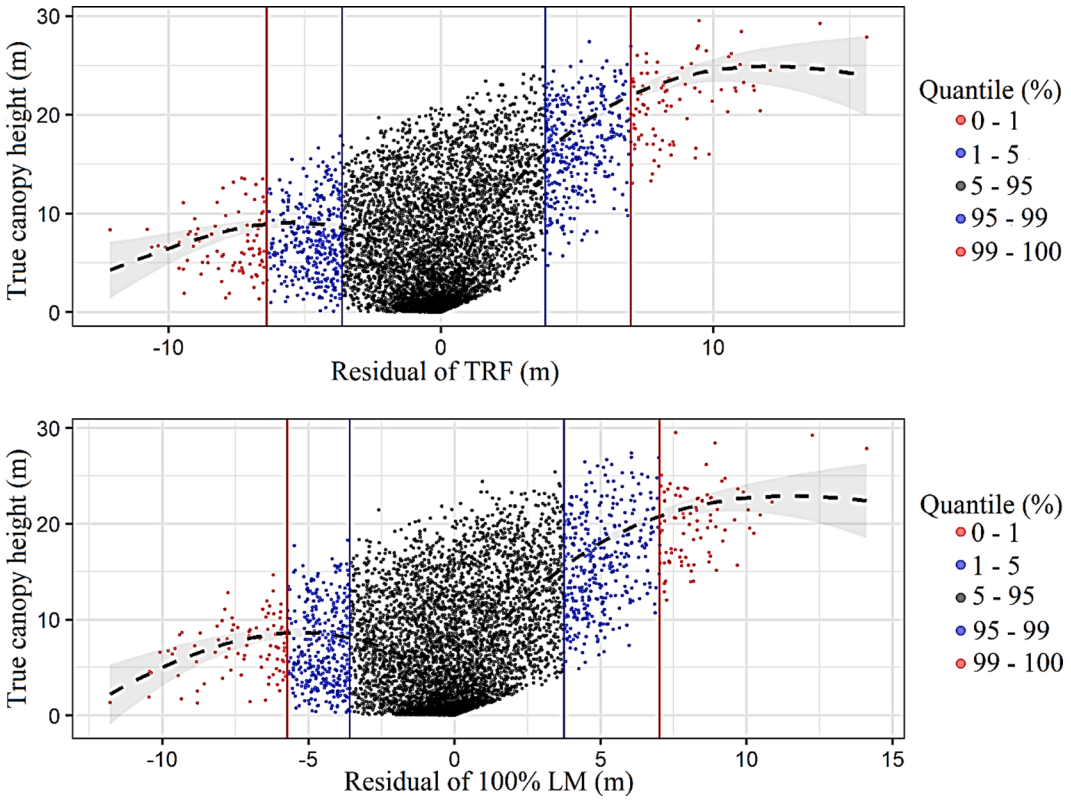


Fig. 8. Correlation between residual and true canopy height based on residual quantile (Solid lines from left to right represent the borders of 1%, 5%, 95% and 99% quantiles of residual. True canopy height minus predicted canopy height equals residual).

Finally, large residuals occur at the most extreme ends of the tree canopy heights when fitting TRF and GRF models. The GRF model is capable of reducing high residuals and hence improving performance, especially for those points falling within the 99% – 100% quantiles. This conclusion is especially beneficial for data preparation and model application. First, although not all of the high and low canopies lead to extreme residuals, it is still highly recommended to balance the sample size of various intervals of canopy heights within a study area. Second, management implications of model errors could differ radically. For instance, errors in the mid-canopy height might have minor effects on estimated shading of riparian areas, but if the data were being used to identify the habitats or potential shading from the highest trees, the model would perform poorly.

5. Conclusions

There are still certain limitations to our data preparation and model application. First, the algorithm uses the closest available GRF model to predict tree canopy height at a new spatial location. A major change of spatial distribution patterns in training and validation datasets may produce a different result. An aggregated pattern of training or validation pixel can make it difficult to mitigate the extent of spatial autocorrelation at new locations by using a misleading sub-model. Therefore, spatial arrangement of sample data is an important cornerstone in random forest model design. Second, this research is limited by its mountainous landscapes. Slopes have been demonstrated to cause error in LiDAR DEMs and hence carrying through to the derived CHM (Tinkham et al., 2012). We suggest that more generalized findings of taking account of spatial non-stationarity in random forest modeling are expected to be obtained over the areas with various landforms and vegetation characteristics.

Based on Landsat and LiDAR derived data, this study contributes to a deeper understanding of approaches to model tree canopy heights at local and global scales by using the existing GRF algorithm to bring in the concept of non-stationarity. The relationship between tree canopy height and predictive ability of the models implies that it is appropriate to assess model performance based on different quantiles. In our case, the lowest and highest canopies result in overestimated and underestimated height predictions, respectively. Comparisons between TRF and GRF models show that the latter can lower predefined extreme residuals, and thus make the model performance relatively stronger.

CRediT authorship contribution statement

Hui Wang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing - original draft. **Travis Seaborn:** Formal analysis, Investigation, Writing - review & editing. **Zhe Wang:** Data curation, Investigation, Writing - original draft. **Christopher C. Caudill:** Supervision, Writing - review & editing. **Timothy E. Link:** 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.

Acknowledgements

This publication was made possible by the NSF Idaho EPSCoR Program and by the National Science Foundation under award number OIA-1757324. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NSF. We also would like to thank Dr. Shijin Qu, the editor and two anonymous reviewers for their useful suggestions on an earlier version of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2021.102353.

References

Ahmed, O.S., et al., 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. ISPRS Journal of Photogrammetry Remote Sensing 101, 89–101.

Allan, J.D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annu. Rev. Ecol. Evol. Syst. 35, 257–284.

Bannari, A., et al., 1995. A review of vegetation indices. Remote sensing reviews 13 (1–2), 95–120.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry Remote Sensing 114, 24–31.

Breiman, L., 2001. Random forests. *Machine learning* 45 (1), 5–32.

Buckley, D.S., et al., 1999. Practical field methods of estimating canopy cover, PAR, and LAI in Michigan oak and pine stands. North. J. Appl. For. 16 (1), 25–32.

Buschmann, C., Nagel, E., 1993. In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. Int. J. Remote Sens. 14 (4), 711–722.

Carder, A., 1995. Forest giants of the world, past and present. Fitzhenry and Whiteside.

Chen, J.M., 1996. Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. Agricultural Forest Meteorology 80 (2–4), 135–163.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83 (403), 596–610.

DeWalle, D.R., 2010. Modeling Stream Shade: Riparian Buffer Height and Density as Important as Buffer Width 1. JAWRA Journal of the American Water Resources Association 46 (2), 323–333.

Flood, N., et al., 2013. An operational scheme for deriving standardised surface reflectance from Landsat TM/ETM+ and SPOT HRG imagery for Eastern Australia. Remote Sensing 5 (1), 83–109.

Fotheringham, A.S., et al., 2003. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.

García, M., et al., 2018. Modelling forest canopy height by integrating airborne LiDAR samples with satellite Radar and multispectral imagery. International journal of applied earth observation geoinformation 66, 159–173.

Georganos, S., et al., 2019. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto International 1–16.

Hardy, J., et al., 2004. Solar radiation transmission through conifer canopies. Agric. For. Meteorol. 126 (3–4), 257–270.

Hengl, T., et al., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518.

Huang, C., et al. (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.

Hudak, A.T., et al., 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. Remote Sens. Environ. 82 (2–3), 397–416.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25 (3), 295–309.

Isenburg, M. (2012). LAStools-efficient tools for LiDAR processing. Available at: http: http://www. cs. unc. edu/~ isenburg/lastools/.

Ji, L., et al., 2011. On the terminology of the spectral vegetation index (NIR− SWIR)/(NIR+ SWIR). Int. J. Remote Sens. 32 (21), 6901–6909.

Kalogirou, S., & Georganos, S. (2018). Spatial Machine Learning (Version 0.1.3) [Package].

Karlson, M., et al., 2015. Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian woodlands using Landsat 8 and random forest. Remote Sensing 7 (8), 10017–10041.

Khosravipour, A., et al., 2014. Generating pit-free canopy height models from airborne lidar. Photogrammetric Engineering Remote Sensing 80 (9), 863–872.

Khosravipour, A., et al., 2015. Effect of slope on treetop detection using a LiDAR Canopy Height Model. ISPRS Journal of Photogrammetry Remote Sensing 104, 44–52.

Klos, P.Z., Link, T.E., 2018. Quantifying shortwave and longwave radiation inputs to headwater streams under differing canopy structures. Forest Ecology Management 407, 116–124.

Loicq, P., et al., 2018. Improving representation of riparian vegetation shading in a regional stream temperature model using LiDAR data. Sci. Total Environ. 624, 480–490.

Louppe, G., et al., 2013. Understanding variable importances in forests of randomized trees. Paper presented at the Advances in neural information processing systems.

Lund, K. (2004). Geology of the Payette National Forest and vicinity, west-central Idaho: US Geological Survey.

Masek, J.G., et al., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. IEEE Geoscience Remote Sensing Letters 3 (1), 68–72.

McIntosh, A.C., et al., 2012. Estimating canopy cover from standard forest inventory measurements in western Oregon. Forest Science 58 (2), 154–167.

Moran, P.A., 1948. The interpretation of statistical maps. J. Roy. Stat. Soc. B 10 (2), 243–251.

Myneni, R.B., et al., 1995. The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience Remote Sensing 33 (2), 481–486.

Nadeau, C., & Bengio, Y. (2000). Inference for the generalization error. Paper presented at the Advances in neural information processing systems.

Ojala, M., Garriga, G.C., 2010. Permutation tests for studying classifier performance. Journal of Machine Learning Research 11 (6).

Oshiro, T. M., et al. (2012). How many trees in a random forest? Paper presented at the International workshop on machine learning and data mining in pattern recognition.

Ota, T., et al., 2014. Estimation of airborne lidar-derived tropical forest canopy height using landsat time series in Cambodia. Remote Sensing 6 (11), 10750–10772.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26 (1), 217–222.

Paper, D. (2020). Scikit-Learn Classifier Tuning from Simple Training Sets. Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python, 137-163.

Pascual, C., et al., 2010. Relationship between LiDAR-derived forest canopy height and Landsat images. Int. J. Remote Sens. 31 (5), 1261–1280.

Pedregosa, F., et al., 2011. Scikit-learn: Machine learning in Python. J. Machine Learning Res. 12, 2825–2830.

Peel, M. C., et al. (2007). Updated world map of the Köppen-Geiger climate classification.

Pitkänen, J., et al., 2004. Adaptive methods for individual tree detection on airborne laser based canopy height model. International Archives of Photogrammetry, Remote Sensing Spatial Information Sciences 36 (8), 187–191.

Qi, J., et al., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48 (2), 119–126.

Rogers, J., Gunn, S., 2005. Identifying feature relevance using a random forest. Paper presented at the International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection.

Roth, T.R., et al., 2010. Stream temperature response to three riparian vegetation scenarios by use of a distributed temperature validated model. Environmental science technology 44 (6), 2072–2078.

Schuett-Hames, D., et al., 2012. Results of the Westside Type N Buffer Characteristics, Integrity and Function Study Final Report. Report CMER 12–1201.

Sekulić, A., et al., 2020. Random Forest Spatial Interpolation. *Remote Sensing* 12 (10), 1687.

Singh, K.K., et al., 2012. LiDAR-Landsat data fusion for large-area assessment of urban land cover: Balancing spatial resolution, data volume and mapping accuracy. ISPRS Journal of Photogrammetry Remote Sensing 74, 110–121.

Sripada, R.P., et al., 2006. Aerial color infrared photography for determining early in-season nitrogen requirements in corn. Agron. J. 98 (4), 968–977.

Staben, G., et al., 2018. Modelling LiDAR derived tree canopy height from Landsat TM, ETM+ and OLI satellite imagery—A machine learning approach. International journal of applied earth observation geoinformation 73, 666–681.

Tao, S., et al., 2016. Global patterns and determinants of forest canopy height. Ecology 97 (12), 3265–3270.

Tinkham, W.T., et al., 2012. Investigating the influence of LiDAR ground surface errors on the utility of derived forest inventories. Can. J. For. Res. 42 (3), 413–422.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8 (2), 127–150.

Vincini, M., et al., 2008. A broad-band leaf chlorophyll vegetation index at the canopy scale. Precis. Agric. 9 (5), 303–319.

Wang, H., et al., 2019. Modeling spatially non-stationary land use/cover change in the lower Connecticut River Basin by combining geographically weighted logistic regression and the CA-Markov model. International Journal of Geographical Information Science 33 (7), 1313–1334.

Zald, H.S., et al., 2014. Influence of lidar, Landsat imagery, disturbance history, plot location accuracy, and plot size on accuracy of imputation maps of forest composition and structure. Remote Sens. Environ. 143, 26–38.