Optical Remote Sensing and Machine Learning Enable Scalable Carbon Mapping in Himalayan Forests Despite Spectral Saturation Challenges

Namgay Wangchuk[1] and Yonten Dorji[1*]

[1]Department of Forest Science, College of Natural Resources, Royal University of Bhutan[1]

* *Correspondence to Ydorji.cnr@rub.edu.bt*

Abstract

Accurate aboveground biomass (AGB) estimation is critical for climate change mitigation and carbon accounting in mountain regions, yet spatially explicit mapping remains limited by reliance on conventional field methods and underexplored machine learning approaches in the Himalaya. Bhutan, a carbon-negative nation maintaining 70% forest cover, faces urgent demand to develop robust monitoring, reporting, and verification (MRV) systems for carbon market participation under the Paris Agreement Article 6. This study developed a random forest (RF) model that integrates Sentinel-2A multispectral imagery, nine vegetation indices, topographic variables, and field-measured AGB from 99 plots across 244.6 hectares in the Kyentshen Community Forest. Using recursive feature elimination, 10 key predictors were identified from 27 candidate variables: Band 11, SLAVI, Band 12, WDRVI, NDVI, elevation, NDWI, GNDVI, band ratio (r3), and Band 9. The optimized RF model achieved $R^2 = 0.34$ and RMSE = 67.98 ± 19.79 t/ha on training data (Ntree = 500, Mtry = 3). However, test-set performance declined substantially ($R^2 = 0.42$, RMSE = 86.14 t/ha), indicating overfitting and reflecting limitations of spectral saturation in high-biomass stands (>300 t/ha). A Wilcoxon signed-rank test revealed systematic overestimation (predicted median: 118.36 t/ha vs. observed: 105.15 t/ha; $p = 0.03$). Despite moderate accuracy, this study demonstrates that freely available Sentinel-2 data combined with machine learning offers a cost-effective, scalable alternative to traditional field mapping for rapid AGB assessment across Bhutan's temperate forests. Vegetation indices, particularly SLAVI and WDRVI, proved superior to raw spectral bands, offering improved sensitivity to dense canopies. The spatial-transect methodology and 10-predictor framework provide a replicable foundation for advancing Bhutan's forest carbon accounting systems. We recommend integrating synthetic aperture radar (SAR) or LiDAR to overcome optical saturation, enabling higher accuracy for robust MRV compliance and carbon credit monetization. This work directly supports the maintenance of Bhutan's carbon-neutral status and establishes the technical baseline for scaling AGB monitoring across mountain regions where engineered infrastructure capacity is limited, but climate action urgency is paramount.

Keywords: Machine Learning, Aboveground Biomass, Sentinel-2, Random Forest, Carbon Monitoring, Himalayan Forests, Remote Sensing, MRV Systems, Carbon Credits

Introduction

Forests are among the largest terrestrial carbon sinks (Yangka et al., 2019). They are crucial for maintaining ecosystem services and, therefore, essential to ensuring the survival of life and addressing climate change (Upadhyay & Singh, 2024). Trees in the forest provide vital supporting and provisioning services (Singh et al., 2024) and play a significant role in regulating the global climate (Bhandari et al., 2024). They serve as critical indicators of climate change (Khandu et al., 2022).

To comprehend these dynamics, forest biomass is an important variable that enables us to assess the ability of forest ecosystems to sequester and maintain their carbon balance capacity (Y. Li et al., 2020). Bhutan's forests exemplify this by storing 239.41 million tons of aboveground carbon [AGC] with a density of 89.45 t/ha (97% is contributed by trees) (FMID, 2023). In nature, biomass exists in various forms, and as per the Intergovernmental Panel on Climate Change [IPCC] (2006) the terrestrial carbon pool is classified into AGB, belowground biomass [BGB], soil organic matter [SOM], dead wood [DW], and litter. Of which, 70%-90% is credited to AGB (Houghton et al., 2009).

AGB underpins carbon stock calculations, as 47–50% of AGB constitutes AGC (IPCC, 2006). It also dominates the terrestrial carbon pool (Bhavsar et al., 2024) and regulates atmospheric carbon through its dynamic fluxes (Houghton et al., 2009). Consequently, the Global Climate Observing System [GCOS] designates AGB as an essential climate variable [ECV] (Purohit et al., 2021). The ecosystem's structure is influenced by AGB (Wang et al., 2023). Likewise, changes in vegetation health, forest succession, and ecosystem services are often more readily detectable through AGB dynamics (Houghton et al., 2009), suggesting that physical complexity and habitat diversity are directly proportional to AGB (Mensah et al., 2020).

Research indicates that subtropical and temperate forests in the Eastern Himalaya possess significant carbon sequestration potential (Devi & Lepcha, 2023), prompting interest in biomass and carbon studies in the region (Sengupta, 2018). Moreover, Kumar & Mutanga (2017) emphasised the need to quantify biomass as it enables us to better understand the impacts of climate change. Accurate measurements of carbon sequestration rates enable the commercialisation of carbon stocks, (Sasidharan et al., 2016), especially in Bhutan, a country well known for its carbon-neutral status and its pledge to maintain forest cover of 60% for all times to come (Yangka et al., 2019). Intriguingly, researchers have extensively studied Bhutan's biomass and carbon (Dorji et al., 2015; Tashi et al., 2017; Sengupta, 2018; Suberi et al., 2018; Chhetri et al., 2021; Tshering & Rinzin, 2022; FMID, 2023; Tshering, 2024; Chhetri et al., 2025). Except for the pilot study carried out by, Wangdi et al. (2018) in which they employ a conventional linear model that fails to capture the complex non-linear relationship between AGB and RS data (Li et al., 2020).

Hence, spatially explicit AGB mapping using Machine learning algorithms [MLAs] and RS fusion remains unexplored in Bhutan. MLAs such as RF address the issue of linearity by modelling non-linear patterns, thereby improving accuracy (Breiman, 2001). Additionally, optical imagery, such as S2A multispectral optical imagery [MSI], is freely available from the European Space Agency [ESA] and has proven effective for spatially estimating AGB (Y. Liu et al., 2019; Bhatti et al., 2023; M. Singh et al., 2024).

This study, in particular, aims to lay down a foundational approach to spatially estimating AGB in Bhutan by leveraging the potential synergy between S2A MSI, its bands, derived vegetation indices (VIs), and band ratios, and topographic and geophysical variables with MLA RF. This not only connects with the opportunities proposed by Phuntsho et al. (2015, 2016), but also serves as a proactive measure to maintain our carbon-neutral status (Yangka & Newman, 2019). This will directly support Bhutan’s recent commitment to Article 6 of the Paris Agreement (Dolkar, 2024) and contribute to the development of a robust monitoring, reporting, and verification [MRV] system. A relatively new perspective to AGB estimation that covers large areas both economically and efficiently (Dahy et al., 2019) when compared to TFM (Purohit et al., 2021). Studies have proven that AGB estimation is possible when RS data is integrated with field inventory data (Yadav & Nandy, 2015; Tian et al., 2023). Therefore, using KCF as a model system, this study aims to (1) assess and identify important predictors in estimating AGB of KCF and (2) generate and validate an AGB map for KCF.

Materials and Methods

Study Area

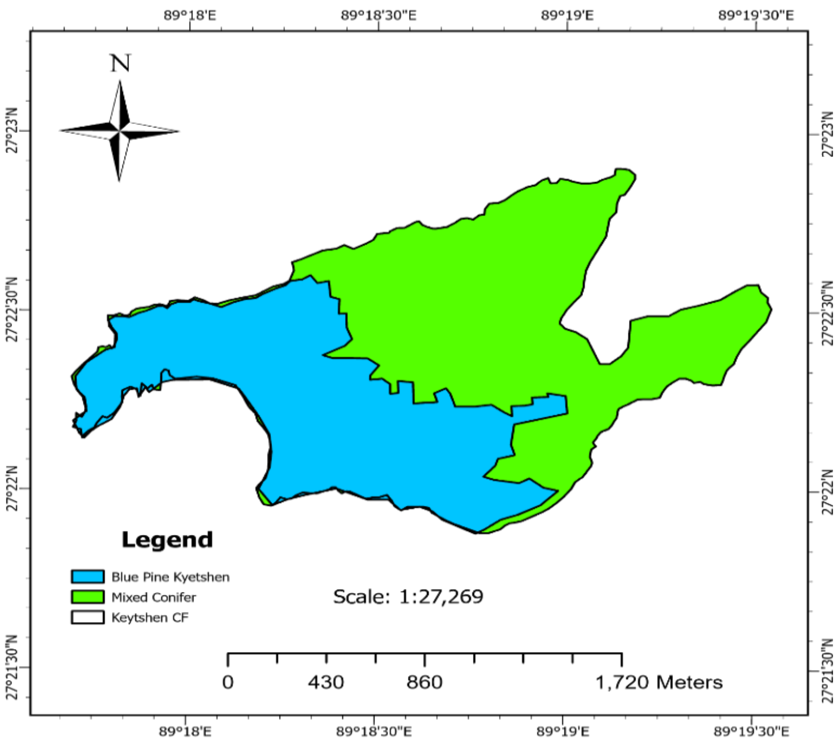


Figure 1: Forest types prevalent in KCF

The study was conducted at KCF located in Katsho Gewog under Haa Dzongkhag (Figure 1). Within elevations of 2710 masl to 3693 masl, the CF boasts a selection of species such as *Larix griffithii* Hook.F., the Himalayan larch (*Zaa-shing*), *Picea spinulosa* A.Henry Spruce (*Sayshing*), *Tsuga dumosa* (D.Don) Eichler. Hemlock (*Ba-shing*), *Pinus wallichiana* A.B Jacks. Blue pine (*Tongphu shing*), *Juniperus recurva* Buch.-Ham. Ex D.Don. Juniper (*Shup shing*), and *Quercus semecarpifolia* Sm. Oak (*Bji shing*) is its dominant tree species. The forest in the CF falls under the pole stage (63.3%), followed by the mature stage (30%), with the remaining area classified as scrubland (1.7%). It was also reported that the CF comprises Blue pine and mixed conifer forests covering areas of 122 ha and 149.6 ha, respectively (Figure 1). It is an ideal site for timber production and has a secondary roadway connecting the two Dzongkhags of Paro and Haa (Figure 2). While winter is cold and comparatively dry, summer is wet, humid, and gloomy (DOFPS, 2023).

Study Design

The road falling within the CF was buffered 30 m on each side in ArcGIS Pro 3.1.0 using the Buffer tool to avoid edge effects (Qasim et al., 2023). This reduced the study area from 271.6 ha to 244.6 ha. A stratified systematic random sampling using the equation n = (t^2 x $CV\%^2$)/SE^2 (where n = number of sample plots, t^2 = confidence level of 95% (1.96), CV = coefficient of variation, SE = Allowable error of 20%) was adopted to compute the number of plots required (DoFPS 2018). This resulted in 71 sample plots; however, to increase sampling intensity, the number was increased to 99 sample plots (45 for blue pine and 55 for mixed conifer), with a plot-to-plot distance of 156.2 m, yielding a sampling intensity of 3.28% (Figure 2). A circular design for the sample plots was adopted with a radius of 16.92 m, covering an area of 900 m^2 that represents 9 pixels of a 10 m resolution S2A imagery.

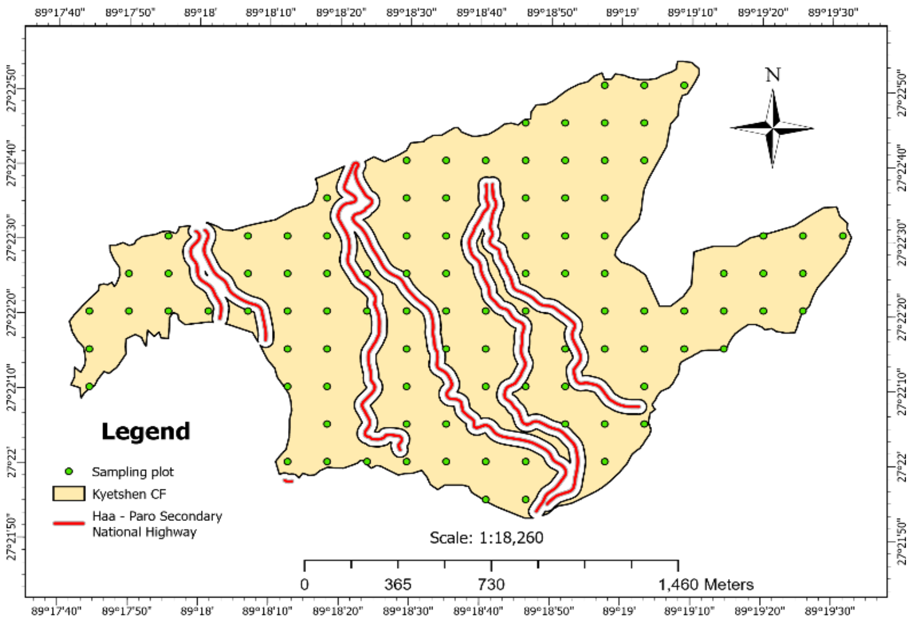


Figure 2: Map showing the plots laid out in the study area utilising the fishnet tool in ArcGIS Pro. Plot to plot distance is 156.2 m and the plot radius is 16.92 m.

Forest Inventory

From inside each sample plot, the geo-coordinates (Degrees Minutes Seconds [DMS]) using Garmin GPS that assures accuracy of ±15 m 95% of the time and generally between ±5 m to ±10 m under normal conditions (Garmin, 2023), height (m) employing Suunto clinometer and trigonometric relationships (West, 2004) with errors between 0.1 m and 0.5 m, and diameter at breast height [DBH] (>5 cm) of each living tree were recorded using a diameter tape along with other geophysical parameters following Kopio & Samanta (2024). The DBH threshold was adopted in line with the recommendations of Searle & Chen (2017). Subsequently, data sheets were used to record all information from each plot and were later compiled in Microsoft Excel (2013).

Satellite Data Acquisition

The digital elevation model (12.5 m) was obtained from NLCS, forest type map from (DOFPS, 2023), canopy height (10 m) from the global tree canopy height model developed by Lang et al. (2023), and the Sentinel 2 series Harmonised Sentinel 2 MSI: MultiSpectral Instrument, Level-2A product that has 13 bands in the visible, near-infrared [NIR], and shortwave-infrared [SWIR] regions of the spectrum (Kirches et al., 2024) were the RS datasets for this study. All of these datasets were uploaded/sourced/accessed and pre-processed in Google Earth Engine (GEE) (Gorelick et al., 2017). The S2A imagery acquisition dates were aligned with the field data collection dates (01/10/2024 to 31/01/2025) to avoid discrepancies (Bhandari et al., 2024). Then, cloud masking was performed using Google's Cloud Score+ [CS+] to filter out scenes with cloud cover greater than 10% (Pasquarella, 2024), resulting in a composite image ready for analysis. Pixel values were converted to reflectance values. All the datasets were resampled to 10 m using GEE's reduceRegion function (Chaaouan, 2025) and projected to EPSG:5266.

Predictor Variables

Vegetation indices [VIs] are effective metrics for assessing vegetation vigour and productivity (Steven et al., 2003). AGB estimation using VIs has proven to give good results (Kerebeh et al., 2024). Similarly, S2A spectral bands (Pandit et al., 2018) and band ratios (Tran et al., 2023) also exhibit good relationships with AGB. On the other hand, some studies reported that topography influences microclimate, affecting species composition and AGB (Ensslin et al., 2015). Forest canopy height is directly proportional to AGB and hence an important variable in AGB calculations (X. Liu et al., 2024). Therefore, a total of nine VIs (NDVI, GNDVI, EVI, AVI, SAVI, SLAVI, NDWI, VARI, and WDRVI), four S2A derived band ratios (r1 SWIR1/SWIR2, r2 Red/NIR, r3 NIR/Green, r4 NIR/SWIR) and ten spectral reflectance bands (Blue, Green, Red, Red edge 1, 2, and 3, NIR, SWIR1, and 2), elevation, slope, aspect, and tree canopy height were added were used in the study, making a total of 27 predictor variables (Figure 10 and Table 2 & 3). All of the above-mentioned variables were computed in GEE. Finally, GEE was called into R via the rgee package (Aybar, 2023) to develop the model (Figure 4).

AGB Calculation

The compiled inventory data were fed to the allometric equation developed by Chave et al. (2005), following the methodology of Suberi et al. (2018), to compute the observed AGB. It is expressed as $AGB = 0.0509 \times \rho \times D^2 \times H$ (where H is the height of the tree, DBH is the diameter of the tree at 1.3 m above the ground, and ρ is the wood specific gravity (FSI, 1996; Donegan et al., 2014; Chhetri et al., 2025).). These were then converted to shapefile [shp.] and raster via the XY Table to Point tool at EPSG:5266 and the Point to Raster tool, using sum as the cell assignment type, at 10m resolution, in ArcGIS Pro 3.1.0, respectively. To address spatial autocorrelation in the dataset (Ploton et al., 2020; Roberts et al., 2016), the dataset was spatially and environmentally separated by randomly assigning 5 folds overlaid with a 100 m resolution grid for each block using the blockCV package (Valavi et al., 2019) in R (Figure 3). This provides an unbiased estimate of model generalization, mitigates spatial overfitting, and ensures high methodological quality in AGB estimation (Roberts et al., 2016). Following best practices, the dataset was randomly split into a training and test set (70/30) to train and validate the model (Freeman et al., 2016; Gholamy et al., 2018).

AGB Prediction

The importance of careful predictor selection in enhancing model performance is well documented in the literature (Ibrahim et al., 2024). Prior to model development, 27 key predictors were selected using recursive feature elimination [RFE] via the caret package in R (Kuhn, 2024). It is a wrapper-based feature selection method that pinpoints the most relevant predictor variables for modelling by systematically removing those that contribute the least (Theng & Bhoyar, 2024). Hyperparameter tuning for the model was done to determine the best number of trees and variables (*Ntree, Mtry, and min.node.size*). RF internally ranked the

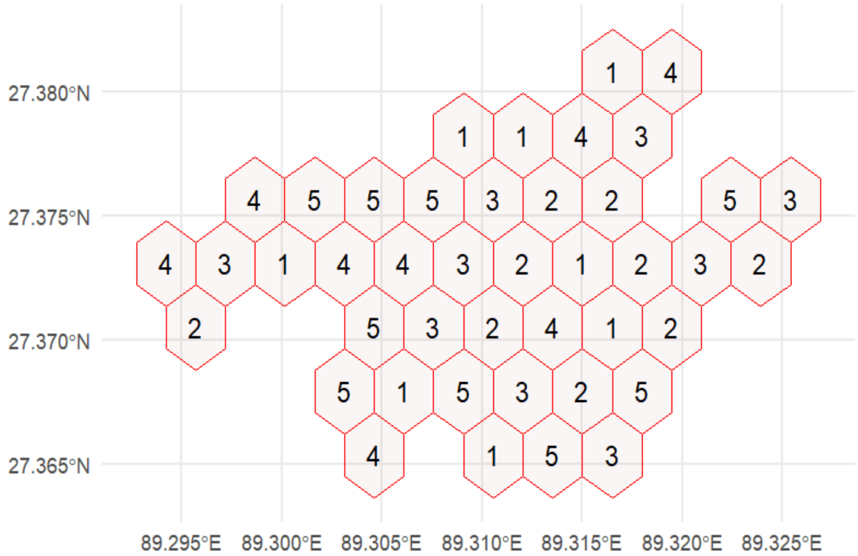


Figure 3: Spatial and environmental separation of samples, performed leveraging blockCV, R package

selected predictors based on the permutation [*%IncMSE*], it is stronger and predictive loss-based (Strobl et al., 2007). These predictors were then employed to train the RF model via the ranger package in R (Wright & Ziegler, 2018) to extract the final predicted AGB map. 10-fold spatial CV was employed throughout the modelling process to maintain a strict separation between validating and training data at all times (Valavi et al., 2019).

Model Validation

R^2 represents the proportion of variance in observed AGB explained by the model (values closer to 1 indicate a better fit), while *RMSE* measures the average prediction error in the same units as AGB (lower *RMSE* indicates higher accuracy) (Gao, 2024). While predictor ranking was based on *%IncMSE*, predictor selection and model validation were quantified based on the highest coefficient of determination [R^2] and the lowest root mean squared error *RMSE,* following best practices in ML and forestry studies (Chicco et al., 2021). Finally, a Wilcoxon signed-rank test was performed to assess systematic bias between the predicted and observed AGB values (Seely et al., 2025). All the results for this study were reported at a significance level $\alpha = 0.05$.

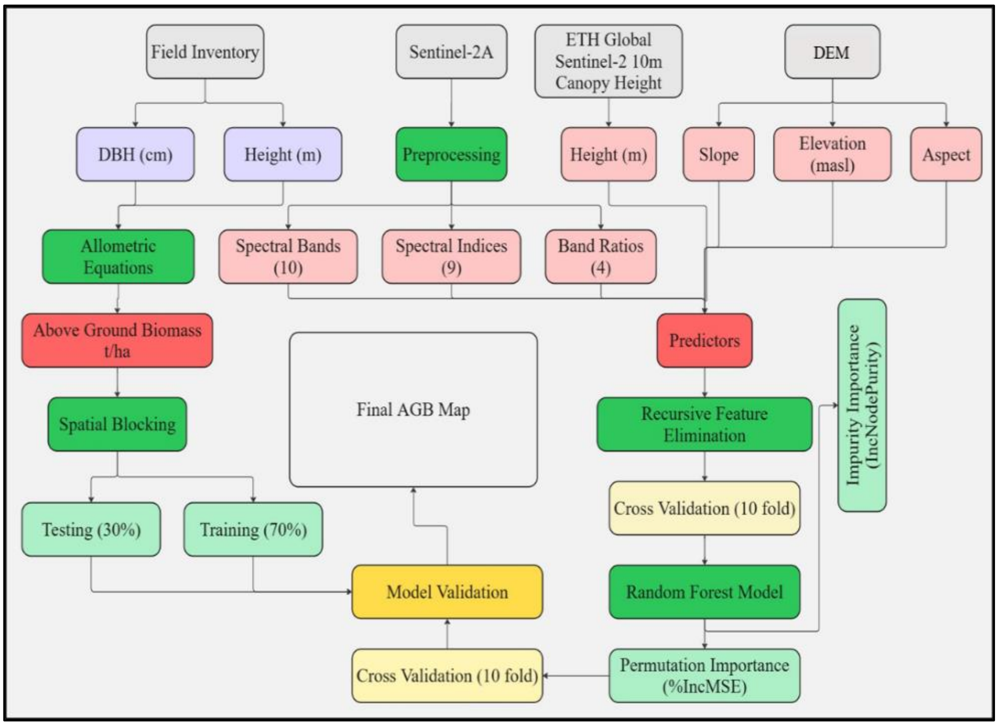


Figure 4: Methodological Framework of the present study

Results and Discussion

Field Observed AGB, Tree Height and Diameter

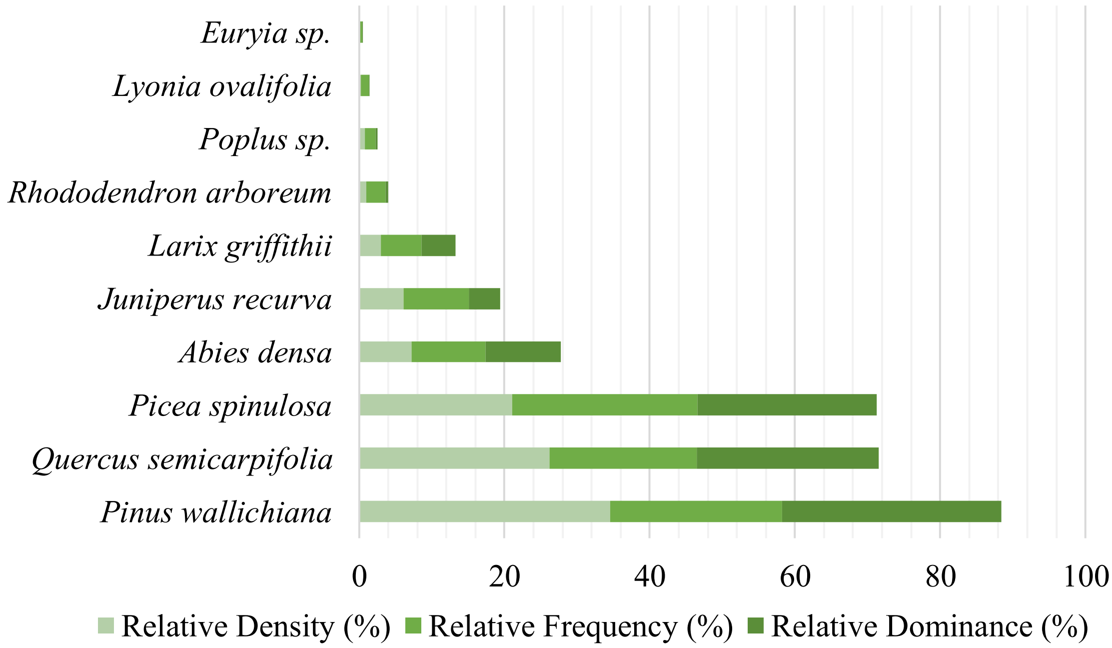


Figure 5. Species IVI chart showing their relative density, relative frequency, and relative dominance

The inventory results from 99 plots recorded 10 tree species in the CF, with *P wallichiana*, *Q semecarpifolia,* and *P spinulosa* emerging as the most dominant, followed by *A densa*, *J recurva*, *L griffithii*, *R arboreum*, *Populus* sp., *Lyonia ovalifolia* (Wall.) Drude, and *Eurya* sp. (Figure 5). This dominance describes the Himalayan mixed conifer forests, where *Pinus* and *Quercus* species often prevail due to their ecological adaptability and competitive advantages in nutrient-poor soils (Naudiyal & Schmerbeck, 2017). Additionally, the inferior nature of *L ovalifolia* and *Eurya* sp. illustrates their adaptability in unique ecological niches (Dhyani et al., 2019). Tree height ranged from 0.6 m to 46.6 m (*M* = 14.6 ±7.36 m), while diameter at breast height (DBH) varied between 5.50 cm and 151.30 cm (*M* = 32.50 ±19.00 cm) (Table 1).

Table 1: Descriptive statistics of height and DBH of the tree along with the observed and predicted AGB

Variables	Min	Max	Mean	Median	StdDev
Height (m)	0.6	46.6	14.6	13	7.36
DBH (cm)	5.50	151.30	32.50	28.00	19.00
Observed AGB (t/ha)	3.90	541.10	126.19	105.19	96.6
Predicted AGB (t/ha)	48.28	275.10	129.20	118.39	45.89

The wide DBH range, including exceptionally large trees (>150 cm), suggests a multi-aged forest with remnant old-growth individuals, a characteristic feature of less disturbed Himalayan forests (Singleton, 1999). Such structural heterogeneity supports diverse microhabitats and carbon storage potential (Aponte et al., 2020). Based on the forest inventory data, KCF accumulates AGB ranging from 3.90 to 541.10 t/ha, with an average of 126.19 ± 96.6 t/ha. A high standard deviation (*SD*) indicates spatial heterogeneity in AGB distribution, which can be attributed to microtopographical variations or species-specific allometric differences. AGB accumulation was higher in KCF than in studies conducted in similar forest types (Sengupta, 2018; Tshering, 2024). This may be due to the presence of diverse dominant species with varying wood densities (Coomes et al., 2014; Chhetri et al., 2025), highlighting the influence of species composition on AGB accumulation. The presence of open canopies with scattered regeneration, fuelled by a sanitation program (bark beetle infestation) that harvested a total of 30, 000 cft over the last decade, has also contributed to the heterogeneity (DOFPS, 2023).

Selecting the Key Predictors with RFE

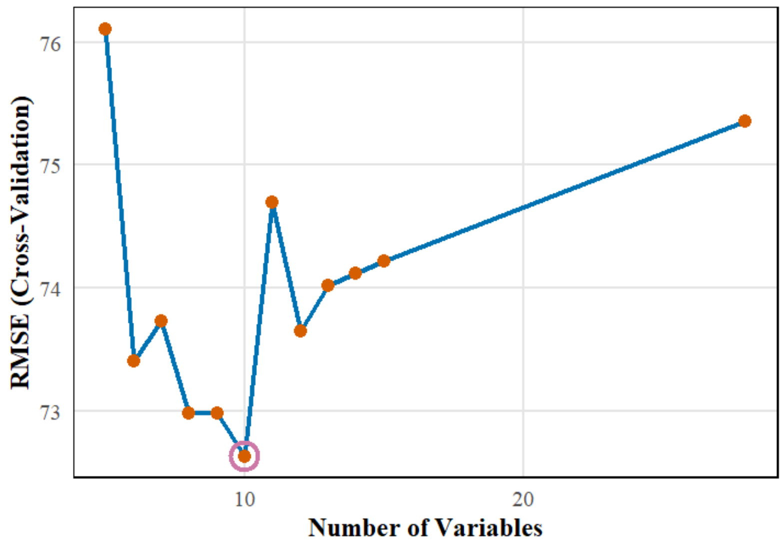


Figure 6: Variable importance selection based on the least *RMSE* for tenfold cross-validation

The RFE model incorporating ten predictors yielded R^2 = 0.28 and *RMSE* = 72.63 ± 28.49 t/ha, and the key predictors emerged (Table 4). These key predictors, Elevation, Band 11, Band 12, NDVI, SLAVI, WDRVI, GNDVI, NDWI, Band 9, and r3, were iteratively selected from an initial pool of 27 (Table 3). The selection of VIs such as NDVI and elevation is endorsed by Fan et al. (2022) as it has proven to boost model performance. WDRVI has also been shown to provide more accurate characterization of vegetation biophysical properties and land surface conditions under high biomass scenarios (Gitelson, 2004). In a similar vein, the NDWI exhibits high sensitivity to water content in closed canopies, particularly at elevated leaf area index [LAI] levels (Purohit et al., 2021). Moreover, SLAVI is less susceptible to cloud shadows, making it a valuable predictor variable in numerous studies focused on AGB prediction (Karlson et al., 2015; Koju et al., 2019; Li et al., 2019).

The inclusion of B9, 11, and 12 is comparable with the findings of Li et al., (2020), further reinforcing the selection of key predictors in this study. In addition, band ratios are widely used for AGB estimation and monitoring, particularly in areas with dense vegetation cover (Xue & Su, 2017; Shen et al., 2019; Luo et al., 2024). Although the R^2 value may appear low, prior research Bhatti et al., (2023) and Luo et al. (2024) demonstrates that models emphasizing strategic feature selection can still generate meaningful spatial predictions, particularly in heterogeneous forests. These studies affirms that, by integrating elevation data, reflectance bands, band ratios, and VIs can help mitigate the variability introduced by multispectral data, ultimately enhancing the robustness of biomass estimation models (Zurqani, 2025).

Ranking key predictors and model optimisation

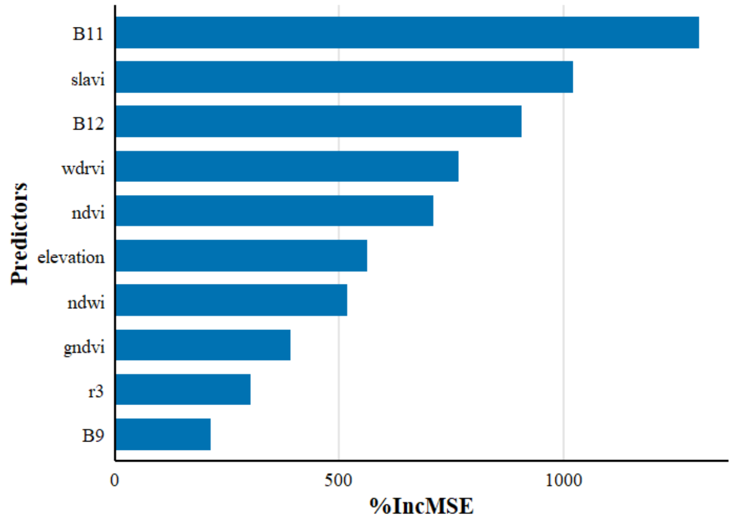


Figure 7: Visualising best predictors ranking based on *%IncMSE*

Configuring RF with *Ntree* = 500 and *Mtry* = 3 produced the best combination for AGB prediction, yielding R^2 = 0.34 and *RMSE* = 67.98 ± 19.79 t/ha (Table 5). The key predictors arranged in descending importance are B11 > SLAVI > B12 > WDRVI > NDVI > Elevation > NDWI > GNDVI > r3 > B9 (Figure 7). Consistent with our results, Chrysafis et al. (2017) reported B11 as the top predictor in their study. The placement of WDRVI above NDVI supports its sensitivity to moderate-to-high vegetation densities (Gitelson, 2004). Contrary to expectations (Wang et al., 2007; Otsu et al., 2019), NDVI outperformed GNDVI despite GNDVI's known sensitivity to high chlorophyll (Gitelson, 2004), likely due to NDVI's stronger response to low chlorophyll levels (Gitelson et al., 1996).

Elevation was ranked as the fifth most important variable, highlighting its notable role in shaping AGB distribution patterns (Ensslin et al., 2015). NDWI can detect changes in vegetation content across both broadleaf and coniferous species (Brown et al., 2022). Seasonal variation in photosynthetic activity may also have influenced the results (Zhang et al., 2023). The inclusion of both NDVI and GNDVI in the model, however, enables capture of chlorophyll variability across a broader range of concentrations. While established relationships and several studies (Kumar et al., 2017; Adam et al., 2020; Bhandari et al., 2024) have emphasized the relevance of tree height in AGB modelling, it was not a significant predictor in the current analysis. Finally, these predictors were fed to the RF algorithm to produce the final AGB map.

Model Validation and its implications

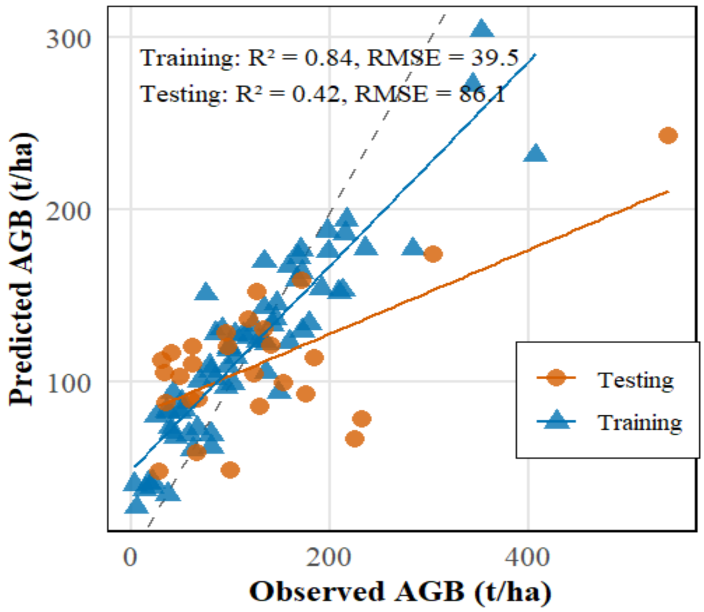


Figure 8: Relationship between observed and predicted AGB in t/ha

The RF algorithm exhibited strong predictive capabilities during the training phase, achieving R^2 = 0.84, *RMSE* = 39.52 t/ha. These metrics indicate that the model effectively captured the complex relationships among spectral indices, spectral bands, band ratios, and topographic variables in estimating AGB. However, its performance dropped substantially on the test dataset, with R^2 falling to 0.42 and *RMSE* rising to 86.14 t/ha (Figure 8). This notable decline reflects difficulties in model generalization to previously unseen data (Raschka, 2020). The significant difference between training and testing results (ΔR^2 = 0.42; $\Delta RMSE$ = 46.62 t/ha) indicates potential overfitting, in which the model may have captured noise specific to the training data rather than learning broader, transferable patterns (Song et al., 2022; Wei et al., 2024).

The result aligns with findings by Barreñada et al. (2024), who highlighted overfitting as a common issue in complex models applied to ecologically diverse forest landscapes. While the high training R^2 value indicates the model's effectiveness within the known data boundaries, the decline in testing performance highlights challenges in applying optical data to unfamiliar environments, especially where spectral saturation is prominent (Aliferis & Simon, 2024). To address overfitting, techniques such as regularization or the use of less complex models should be considered (Aliferis & Simon, 2024). Moreover, studies by Ali et al. (2024) and Bhandari et al. (2024) emphasize the value of incorporating texture variables to enhance model performance.

The findings are similar to those of Sainuddin et al. (2023), who recorded $R^2 = 0.46$ and $RMSE = 42.56$ t/ha and highlighted spectral saturation as a limitation in optical-based biomass models. The primary reason could be the limited penetrative capacity of optical imagery, which captures only surface features (Myneni et al., 2001). Employing non-saturating spectral indices, including WDRVI and SLAVI, alongside texture metrics, improves sensitivity to biomass variability (Gitelson, 2004). Sainuddin et al. (2023) Further suggest integrating synthetic aperture radar [SAR] or light detecting and ranging [LiDAR] to capture vertical forest structure and mitigate saturation issues. By evaluating the potential of optical data for RF modelling, this research also addresses the gap highlighted by Wangdi et al. (2018) regarding the applicability of optical remote sensing to Bhutan's unique forest ecosystems. These findings lay a foundation for enhancing Bhutan's forest carbon accounting systems and support evidence-based forest management and policy development.

Final AGB Map

The Wilcoxon signed-rank test revealed a statistically significant discrepancy between observed and predicted AGB values ($V = 1797$, $p = .03$, $Z = -2.08$), with a small effect size ($r = 0.21$). The *predicted AGB Mdn* of 118.36 t/ha exceeded the observed *Mdn* of 105.15 t/ha, suggesting systematic overestimation in the model outputs. This may be due to spectral saturation in high-biomass temperate forests, a well-documented limitation of optical remote sensing (Kumar & Mutanga, 2017). These findings carry important implications for Bhutan's carbon monitoring systems. This foundational approach aligns with Phuntsho et al.'s (2016) the call for developing more robust geospatial information systems for Bhutan. Furthermore, it

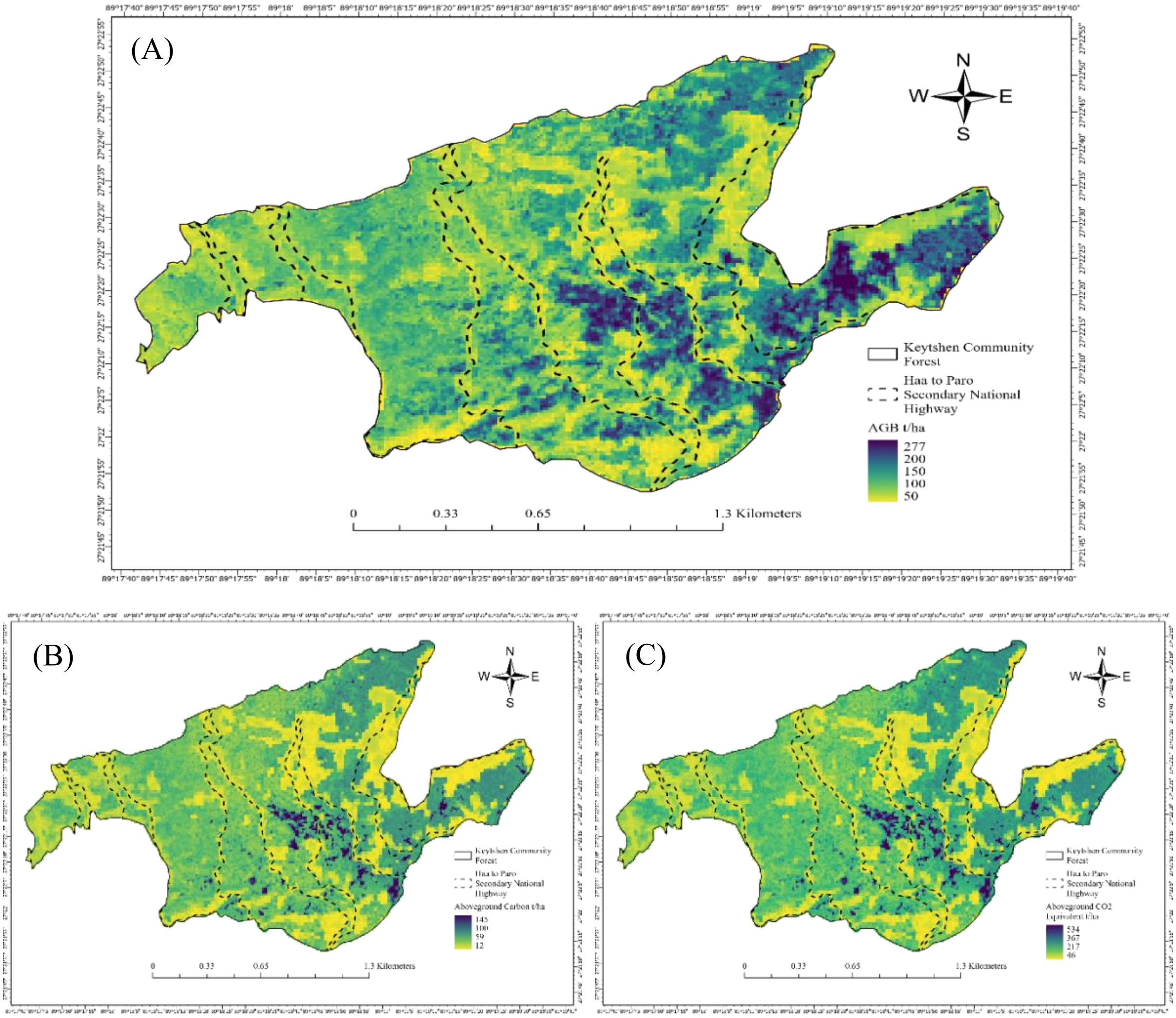


Figure 9: (A) AGB map of KCF along with (B) AGC, and (B) Aboveground CO_2 Equivalent [$AGCO_2$] maps derived based on the predicted AGB where, AGC (AGB x 0.47) and $AGCO_2$ (AGB x 3.67).

underscores the urgency of Yangka et al. (2019) the recommendation for proactive measures to maintain Bhutan's carbon-neutral status. The demonstrated need for model refinement supports Bhandari et al. (2024) the assertion that iterative improvement is fundamental to successful ML applications in ecological research. Further, the carbon and CO2-equivalent map was derived as shown in Figure 9.

Conclusion and Recommendations

Reliable AGB quantification is fundamental for evidence-based forest management and climate change mitigation at regional to global scales. This study demonstrates the potential of Sentinel-2 data combined with RF modelling to estimate AGB in Bhutan's temperate forests, using field-measured data as reference. While the model achieved moderate explanatory power (R^2 = 0.42) Using 10 key predictors derived from 27 candidate variables, the systematic difference between observed and predicted values highlights both the promise and the limitations of optical remote sensing in the complex Himalayan terrain.

Our findings establish three critical insights: First, integrating moisture-sensitive spectral indices (NDWI, SLAVI) with topographic variables can effectively capture biomass patterns despite the challenges posed by spectral saturation in dense canopies. Second, ML approaches offer a cost-effective alternative to TFM, reducing fieldwork requirements while leveraging freely available satellite data. Third, the developed framework provides a foundational baseline for advancing AGB estimation in Bhutan, though higher accuracy is needed to fully support robust MRV systems for REDD+ and Article 6 implementation.

These results carry important implications for climate policy and research: (1) Policymakers can utilize such spatially explicit AGB maps not only to inform mitigation strategies and carbon market participation but also for efficient forest resource management, and (2) researchers should prioritize multi-sensor fusion such as LiDAR/SAR with optical data, continuous field validation, and model optimization to address current accuracy limitations. Future work must focus on overcoming topographic complexity and high-biomass saturation effects - challenges particularly relevant to Himalayan ecosystems.

As Bhutan works to maintain its carbon-neutral status, this study contributes to the growing scientific foundation for monitoring forest carbon stocks. The methodology demonstrates how ML can enhance national carbon accounting systems while highlighting the need for ongoing refinement to meet international reporting standards. This work ultimately advances our capacity to quantify ecosystem services in Bhutan, where accurate biomass assessment remains both critically important and technically challenging. The following are some of the recommendations that arose from this study:

i. Future studies may address spectral saturation by integrating LiDAR or SAR to capture vertical forest structure.

ii. Integrate regional species-specific allometric models to reduce biases in different forest types.

iii. Scaling this approach nationwide and validating it across different forest types could enhance Bhutan’s forest carbon MRV.

iv. Assess how species-specific spectral reflectance patterns correlate with AGB across different forest types.

Acknowledgement

The authors gratefully acknowledge the College of Natural Resources and the Royal University of Bhutan for providing essential academic and institutional support. The authors acknowledge the National Land Commission Secretariat for access to ArcGIS Pro software critical for spatial analysis, and NASA's Applied Remote Sensing Training (ARSET) program for training. We are grateful for the cooperation of the Wangtsa Village community and the dedicated field assistance of Mr. Shacha Dorji, Mr. Wangdi Rigsel, Mr. Sangay Thinley Dorji, Mr. Tsheltrim Zangpo, Mr. Nima Tshering, and Mr. Karma Loday.

References

Ali, H., Mohammadi, J., & Shataee Jouibary, S. (2024). Deep and machine learning prediction of forest above-ground biomass using multi-source remote sensing data in coniferous planted forests in Iran. *European Journal of Forest Research*. https://doi.org/10.1007/s10342-024-01721-w

Aliferis, C., & Simon, G. (2024). Overfitting, Underfitting and General Model Overconfidence and Under-Performance Pitfalls and Best Practices in Machine Learning and AI. In G. J. Simon & C. Aliferis (Eds.), *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls* (pp. 477–524). Springer International Publishing. https://doi.org/10.1007/978-3-031-39355-6_10

Aybar, C. (2023). *rgee: R Bindings for Calling the 'Earth Engine' API* (p. 1.1.7) [Dataset]. https://doi.org/10.32614/CRAN.package.rgee

Barreñada, L., Dhiman, P., Timmerman, D., Boulesteix, A.-L., & Van Calster, B. (2024). Understanding overfitting in random forest for probability estimation: A visualization and simulation study. *Diagnostic and Prognostic Research*, *8*(1), 14. https://doi.org/10.1186/s41512-024-00177-1

Bhandari, K., Srinet, R., & Nandy, S. (2024). Forest Height and Aboveground Biomass Mapping by synergistic use of GEDI and Sentinel Data using Random Forest Algorithm in the Indian Himalayan Region. *Journal of the Indian Society of Remote Sensing*, *52*(4), 857–869. https://doi.org/10.1007/s12524-023-01792-z

Bhatti, S., Ahmad, S. R., Asif, M., & Farooqi, I. ul H. (2023). Estimation of aboveground carbon stock using Sentinel-2A data and Random Forest algorithm in scrub forests of the Salt Range, Pakistan. *Forestry: An International Journal of Forest Research*, *96*(1), 104–120. https://doi.org/10.1093/forestry/cpac036

Bhavsar, D., Das, A. K., Chakraborty, K., Patnaik, C., Sarma, K. K., & Aggrawal, S. P. (2024). Above Ground Biomass Mapping of Tropical Forest of Tripura Using EOS-04 and ALOS-2 PALSAR-2 SAR Data. *Journal of the Indian Society of Remote Sensing*, *52*(4), 801–811. https://doi.org/10.1007/s12524-024-01838-w

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., & Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, *9*(1), 251. https://doi.org/10.1038/s41597-022-01307-4

Chaaouan, J. (2025). Calculating Statistics with reduceRegion in Google Earth Engine. *GEE Academy*. https://gee.geojamal.com/2025/04/calculating-statistics-with.html

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.-P., Nelson, B. W., Ogawa, H., Puig, H., Riéra, B., & Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, *145*(1), 87–99. https://doi.org/10.1007/s00442-005-0100-x

Chhetri, Y. R., Tshering, C., Dukpa, D., & Tshering, D. (2025). Region-specific Average Wood Densities of Selected Tree Species of Bhutan and Their Comparison with the Global Database. *Asian Journal of Environment & Ecology*, *24*(4), 238–247. https://doi.org/10.9734/ajee/2025/v24i4690

Chhetri, Y., Suberi, B., & Katel, O. (2021). Simple Allometry To Estimate The Aboveground Tree Biomass For Five Cool-Broadleaved Species Of Bhutan Himalaya. *International Journal of Progressive Sciences and Technologies*, *29*, 58. https://doi.org/10.52155/ijpsat.v29.2.3715

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Chrysafis, I., Mallinis, G., Siachalou, S., & Patias, P. (2017). Assessing the relationships between growing stock volume and Sentinel-2 imagery in a Mediterranean forest ecosystem. *Remote Sensing Letters*, *8*, 508–517. https://doi.org/10.1080/2150704X.2017.1295479

Coomes, D. A., Flores, O., Holdaway, R., Jucker, T., Lines, E. R., & Vanderwel, M. C. (2014). Wood production response to climate change will depend critically on forest composition and structure. *Global Change Biology*, *20*(12), 3632–3645.

Dahy, B., Issa, S., Ksiksi, T., & Saleous, N. (2019). *Non-Conventional Methods as a New Alternative for the Estimation of Terrestrial Biomass and Carbon Sequestered– A Mini Review*. *4*. https://doi.org/10.33552/WJASS.2019.04.000579

Devi, N. B., & Lepcha, N. T. (2023). Carbon sink and source function of Eastern Himalayan forests: Implications of change in climate and biotic variables. *Environmental Monitoring and Assessment*, *195*(7), 843. https://doi.org/10.1007/s10661-023-11460-x

Dhyani, S., Maikhuri, R. K., & Dhyani, D. (2019). Impact of anthropogenic interferences on species composition, regeneration and stand quality in moist temperate forests of Central Himalaya. *Tropical Ecology*, *60*(4), 539–551. https://doi.org/10.1007/s42965-020-00054-0

DoFPS. (2018). *Field Forestry Manual for Bhutan – Community Forestry* (3rd ed.). y Department of Forests and Park Services, Ministry of Agriculture and Forests, Royal Government of Bhutan.

DOFPS. (2023). *Kyentshen Community Forest Management Plan* [Community Forest Management Plan]. Department of Forest and Park Services.

Dolkar, D. (2024). Bhutan to sell carbon credits to Singapore – Druk Green Power Corporation Limited. *Kuensel*. https://www.drukgreen.bt/bhutan-to-sell-carbon-credits-to-singapore/

Donegan, E., Sola, G., Cheng, Z., Birigazzi, L., Gamarra, J., Henry, M., Vieilledent, G., & Chiti, T. (2014). *GlobAllomeTree's wood density database*.

Dorji, Y., Godbold, D., & Ahmed, I. (2015). *Effects of Experimental Drought on Fine Root Biomass in Two Different Altitudinal Forests of Himalayan Bhutan*. https://doi.org/10.13140/RG.2.2.25596.87680

Ensslin, A., Rutten, G., Pommer, U., Zimmermann, R., Hemp, A., & Fischer, M. (2015). Effects of elevation and land use on the biomass of trees, shrubs and herbs at Mount Kilimanjaro. *Ecosphere*, *6*(3), art45. https://doi.org/10.1890/ES14-00492.1

Fan, X., He, G., Zhang, W., Long, T., Zhang, X., Wang, G., Sun, G., Zhou, H., Shang, Z., Tian, D., Li, X., & Song, X. (2022). Sentinel-2 Images Based Modeling of Grassland Above-Ground Biomass Using Random Forest Algorithm: A Case Study on the Tibetan Plateau. *Remote Sensing*, *14*(21), Article 21. https://doi.org/10.3390/rs14215321

FMID. (2023). *National Forest Inventory: State of Forest Carbon Report: Vol. II*. Department of Forest and Park Services.

Freeman, E. A., Moisen, G. G., Coulston, J. W., & Wilson, B. T. (2016). Random forests and stochastic gradient boosting for predicting tree canopy cover: Comparing tuning processes and model performance. *Canadian Journal of Forest Research*, *46*(3), 323–339. https://doi.org/10.1139/cjfr-2014-0562

FSI. (1996). *Volume equations for forests of India, Nepal, and Bhutan.* Forest Survey of India, Ministry of Environment & Forests, Govt. of India.

Gao, J. (2024). R-Squared (R2) – How much variation is explained? *Research Methods in Medicine & Health Sciences*, *5*(4), 104–109. https://doi.org/10.1177/26320843231186398

Garmin. (2023). *GPS Accuracy | Garmin Customer Support*. https://support.garmin.com/en-US/?faq=aZc8RezeAb9LjCDpJplTY7

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Departmental Technical Reports (CS)*. https://scholarworks.utep.edu/cs_techrep/1209

Gitelson, A. A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*, *161*(2), 165–173. https://doi.org/10.1078/0176-1617-01176

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, *58*(3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Houghton, R. A., Hall, F., & Goetz, S. J. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research: Biogeosciences*, *114*(G2). https://doi.org/10.1029/2009JG000935

Ibrahim, S., Balzter, H., & Tansey, K. (2024). Machine learning feature importance selection for predicting aboveground biomass in African savannah with landsat 8 and ALOS PALSAR data. *Machine Learning with Applications*, *16*, 100561. https://doi.org/10.1016/j.mlwa.2024.100561

IPCC. (2006). *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. IGES. https://www.ipcc.ch/report/ar5/wg3/

Karlson, M., Ostwald, M., Reese, H., Sanou, J., Boalidioa, T., & Mattsson, E. (2015). Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest. *Remote Sensing*, *7*, 10017–10041. https://doi.org/10.3390/rs70810017

Kerebeh, H., Forkel, M., & Zewdie, W. (2024). Above ground biomass estimation in the upper Blue Nile basin forests, North-Western Ethiopia. *Environmental Systems Research*, *13*(1), 48. https://doi.org/10.1186/s40068-024-00376-1

Khandu, Y., Polthanee, A., & Isarangkool Na Ayutthaya, S. (2022). Ecological Dynamics and Regeneration Expansion of Treeline Ecotones in Response to Climate Change in Northern Bhutan Himalayas. *Forests*, *13*(7), 1062. https://doi.org/10.3390/f13071062

Kirches, G., Brockman, C., Preusker, R., Gamba, P., & Hagolle, O. (2024). *Sentinel-2 Next Generation Phase 0 Scientific Support Study* (S2NG-SSS_FinalReport_v1.0). European Space Agency. https://esamultimedia.esa.int/docs/EarthObservation/S2NG_Phase_0_Science_Support_Study_Final_Report.pdf

Koju, U. A., Zhang, J., Maharjan, S., Zhang, S., Bai, Y., Vijayakumar, D. B. I. P., & Yao, F. (2019). A two-scale approach for estimating forest aboveground biomass with optical remote sensing images in a subtropical forest of Nepal. *Journal of Forestry Research*, *30*(6), 2119–2136. https://doi.org/10.1007/s11676-018-0743-1

Kuhn, M. (2024). *caret: Classification and Regression Training*. https://github.com/topepo/caret/

Kumar, L., & Mutanga, O. (2017). Remote Sensing of Above-Ground Biomass. *Remote Sensing*, *9*(9), Article 9. https://doi.org/10.3390/rs9090935

Lang, N., Jetz, W., & Wegner, J. (2023). A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution*, *7*, 1–12. https://doi.org/10.1038/s41559-023-02206-6

Li, B., Wang, W., Bai, L., Chen, N., & Wang, W. (2019). Estimation of aboveground vegetation biomass based on Landsat-8 OLI satellite images in the Guanzhong Basin, China. *International Journal of Remote Sensing*, *40*(10), 3927–3947. https://doi.org/10.1080/01431161.2018.1553323

Li, X., Huang, C., & Jiang, X. (2020). Spatiotemporal occurrence of Mishmi takin *Budorcas taxicolor* in Dulongjiang Region, southwestern China. *Mammalia*, *84*(6), 513–519. https://doi.org/10.1515/mammalia-2019-0114

Li, Y., Li, M., Li, C., & Liu, Z. (2020). Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific Reports*, *10*(1), 9952. https://doi.org/10.1038/s41598-020-67024-3

Liu, X., Neigh, C. S. R., Pardini, M., & Forkel, M. (2024). Estimating forest height and above-ground biomass in tropical forests using P-band TomoSAR and GEDI observations. *International Journal of Remote Sensing*, *45*(9), 3129–3148. https://doi.org/10.1080/01431161.2024.2343134

Liu, Y., Gong, W., Xing, Y., Hu, X., & Gong, J. (2019). Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *151*, 277–289. https://doi.org/10.1016/j.isprsjprs.2019.03.016

Luo, M., Anees, S. A., Huang, Q., Qin, X., Qin, Z., Fan, J., Han, G., Zhang, L., & Shafri, H. Z. M. (2024). Improving Forest Above-Ground Biomass Estimation by Integrating Individual Machine Learning Models. *Forests*, *15*(6), Article 6. https://doi.org/10.3390/f15060975

Mensah, S., Salako, V. K., & Seifert, T. (2020). Structural complexity and large-sized trees explain shifting species richness and carbon relationship across vegetation types. *Functional Ecology*, *34*(8), 1731–1745. https://doi.org/10.1111/1365-2435.13585

Myneni, R. B., Dong, J., Tucker, C. J., Kaufmann, R. K., Kauppi, P. E., Liski, J., Zhou, L., Alexeyev, V., & Hughes, M. K. (2001). A large carbon sink in the woody biomass of Northern forests. *Proceedings of the National Academy of Sciences*, *98*(26), 14784–14789. https://doi.org/10.1073/pnas.261555198

Naudiyal, N., & Schmerbeck, J. (2017). The changing Himalayan landscape: Pine-oak forest dynamics and the supply of ecosystem services. *Journal of Forestry Research*, *28*(3), 431–443. https://doi.org/10.1007/s11676-016-0338-7

Otsu, K., Pla, M., Duane, A., Cardil, A., & Brotons, L. (2019). Estimating the Threshold of Detection on Tree Crown Defoliation Using Vegetation Indices from UAS Multispectral Imagery. *Drones*, *3*(4), Article 4. https://doi.org/10.3390/drones3040080

Pandit, S., Tsuyuki, S., & Dube, T. (2018). Estimating Above-Ground Biomass in Sub-Tropical Buffer Zone Community Forests, Nepal, Using Sentinel 2 Data. *Remote Sensing*, *10*(4), Article 4. https://doi.org/10.3390/rs10040601

Pasquarella, V. (2024). All Clear with Cloud Score+. *Google Earth and Earth Engine*. https://medium.com/google-earth/all-clear-with-cloud-score-bd6ee2e2235e

Phuntsho, P., Tshering, K., Rai, A., Na, T., Choden, T., Wangdi, S., Dorji, S., Tenzin, J., delma, S., Gyelteshen, N., & Forest. (2016). *Multi-Scale Forest Biomass Assessment and Monitoring in the Hindu Kush Himalayan Region: A Geospatial Perspective Special Publication 30 Multi-Scale Forest Biomass Assessment and Monitoring in the Hindu Kush Himalayan Region: A Geospatial Perspective Bhutan's Geospatial Information System for Forest Biomass Assessment*.

Phuntsho, P., Tshering, K., Rai, A., Tshering, Choden, T., Wangdi, S., Dorji, S., Tenzin, J., delma, S., & Gyelteshen, N. (2015). Bhutan's Geospatial Information System for Forest Biomass Assessment. *Special Publication of Science, ICIMOD*.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, *11*(1), 4540. https://doi.org/10.1038/s41467-020-18321-y

Purohit, S., Aggarwal, S. P., & Patel, N. R. (2021). Estimation of forest aboveground biomass using combination of Landsat 8 and Sentinel-1A data with random forest regression algorithm in Himalayan Foothills. *Tropical Ecology*, *62*(2), 288–300. https://doi.org/10.1007/s42965-021-00140-x

Qasim, M., Csaplovics, E., & Villegas, M. H. S. (2023). Forest biomass assessment combining field inventorying and remote sensing data. *Open Geosciences*, *15*(1), 20220553. https://doi.org/10.1515/geo-2022-0553

Raschka, S. (2020). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning* (arXiv:1811.12808). arXiv. https://doi.org/10.48550/arXiv.1811.12808

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2016). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Nordic Society Okios*. https://doi.org/10.1111/ecog.02881

Sainuddin, F. V., Chirakkal, S., Asok, S. V., Das, A. K., & Putrevu, D. (2023). Evaluation of multifrequency SAR data for estimating tropical above-ground biomass by employing radiative transfer modeling. *Environmental Monitoring and Assessment*, *195*(9), 1102. https://doi.org/10.1007/s10661-023-11715-7

Sasidharan, S., Sivaram, M., Henry, M., Birigazzi, L., & RINI, G. (2016). *TREE ALLOMETRIC EQUATIONS IN SOUTH ASIA*.

Searle, E. B., & Chen, H. Y. H. (2017). Tree size thresholds produce biased estimates of forest biomass dynamics. *Forest Ecology and Management*, *400*, 468–474. https://doi.org/10.1016/j.foreco.2017.06.042

Seely, H., Coops ,Nicholas C., White ,Joanne C., Montwé ,David, & and Ragab, A. (2025). Forest aboveground biomass estimation using deep learning data fusion of ALS, multispectral, and topographic data. *International Journal of Remote Sensing*, *46*(10), 3874–3912. https://doi.org/10.1080/01431161.2025.2492412

Sengupta, A. (2018). Carbon Storage Potential of Temperate Mixed Coniferous Trees in Community Forests of Bumthang. *Journal of the Bhutan Ecological Society*, *3*.

Shen, X., Cao, L., Yang, B., Xu, Z., & Wang, G. (2019). Estimation of Forest Structural Attributes Using Spectral Indices and Point Clouds from UAS-Based Multispectral and RGB Imageries. *Remote. Sens.*, *11*, 800. https://doi.org/10.3390/RS11070800

Singh, M., Arshad, A., Bijlwan, A., Tamang, M., Shahina, N. N., Biswas, A., Bhowmick, A., Vineeta, Banik, G. C., Nath, A. J., Shukla, G., & Chakravarty, S. (2024). Mapping tree carbon density using sentinel 2A sensor on Google Earth Engine in Darjeeling Himalayas: Implication for tree carbon management and climate change mitigation. *Physics and Chemistry of the Earth, Parts A/B/C*, *134*, 103569. https://doi.org/10.1016/j.pce.2024.103569

Song, H., Kim, M., Park, D., Shin, Y., & Lee, J.-G. (2022). Learning from noisy labels with deep neural networks: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(11), 8135–8153. https://ieeexplore.ieee.org/abstract/document/9729424/

Steven, M. D., Malthus, T. J., Baret, F., Xu, H., & Chopping, M. J. (2003). Intercalibration of vegetation indices from different sensor systems. *Remote Sensing of Environment*, *88*(4), 412–422. https://doi.org/10.1016/j.rse.2003.08.010

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. https://doi.org/10.1186/1471-2105-8-25

Suberi, B., Tiwari, K., Gurung, D., Bajracharya, R., & Sitaula, B. (2018). *Effect of Harvesting and Non-Harvested Forest Management on Carbon Stocks*. https://doi.org/10.9734/IJECC/2018/41984

Tashi, S., Keitel, C., Singh, B., & Adams, M. (2017). Allometric equations for biomass and carbon stocks of forests along an altitudinal gradient in the eastern Himalayas. *Forestry: An International Journal of Forest Research*, *90*(3), 445–454. https://doi.org/10.1093/forestry/cpx003

Tian, L., Wu, X., Tao, Y., Li, M., Qian, C., Liao, L., & Fu, W. (2023). Review of Remote Sensing-Based Methods for Forest Aboveground Biomass Estimation: Progress, Challenges, and Prospects. *Forests*, *14*(6), Article 6. https://doi.org/10.3390/f14061086

Tran, M. D., Vantrepotte, V., Loisel, H., Oliveira, E. N., Tran, K. T., Jorge, D., Mériaux, X., & Paranhos, R. (2023). Band ratios combination for estimating chlorophyll-a from sentinel-2 and sentinel-3 in coastal waters. *Remote Sensing*, *15*(6), 1653. https://www.mdpi.com/2072-4292/15/6/1653

Tshering, S. (2024). Terrestrial tree biomass and carbon stock for forests of Thimphu district, western Bhutan Himalaya. *Bhutan Journal of Natural Resources and Development*, *11*(1), Article 1. https://doi.org/10.17102/cnr.2024.92

Tshering, S., & Rinzin, P. (2022). Terrestrial Biomass and Carbon Stock in Broad-leaved Forests of Punakha District, Western Bhutan. *Nature Environment and Pollution Technology*, *21*, 175–181. https://doi.org/10.46488/NEPT.2022.v21i01.019

Upadhyay, P., & Singh, T. S. (2024). The Importance of Conserving Existing Forest Areas and Protecting Biodiversity in Addressing Climate Change Issues. In H. Singh (Ed.), *Forests and Climate Change: Biological Perspectives on Impact, Adaptation, and Mitigation Strategies* (pp. 605–623). Springer Nature. https://doi.org/10.1007/978-981-97-3905-9_29

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, *10*(2), 225–232. https://doi.org/10.1111/2041-210X.13107

Wang, F., Huang, J., Tang, Y., & Wang, X. (2007). New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. *Rice Science*, *14*(3), 195–203. https://doi.org/10.1016/S1672-6308(07)60027-4

Wang, Y., Song, Z., Zhang, X., & Wang, H. (2023). Effects of Tree Diversity, Functional Composition, and Large Trees on the Aboveground Biomass of an Old-Growth Subtropical Forest in Southern China. *Forests*, *14*(5), 994. https://www.mdpi.com/1999-4907/14/5/994

Wangdi, D., Phuntsho, Y., Rai, A., & Zangpo, D. (2018). *Aboveground Tree Biomass Density Mapping for Pilot Sites of Thimphu and Paro Dzongkhag, Bhutan Using NFI and ALOS PALSAR-2 mosaic data*. DoFPS, FRMD.

Wei, J., Zhang, Y., Zhang, L. Y., Ding, M., Chen, C., Ong, K.-L., Zhang, J., & Xiang, Y. (2024). *Memorization in deep learning: A survey* (arXiv:2406.03880). arXiv. https://doi.org/10.48550/arXiv.2406.03880

West, P. W. (2004). *Tree and Forest Measurement*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-05436-9

Wright, M. N., & Ziegler, A. (2018). *ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R*. https://doi.org/10.18637/jss.v077.i01

Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, *2017*(1), 1353691. https://doi.org/10.1155/2017/1353691

Yadav, B. K. V., & Nandy, S. (2015). Mapping aboveground woody biomass using forest inventory, remote sensing and geostatistical techniques. *Environmental Monitoring and Assessment*, *187*(5), 308. https://doi.org/10.1007/s10661-015-4551-1

Yangka, D., & Newman, P. (2019). Happiness, Environment and Wealth: What Can Bhutan Show Us about Resolving the Nexus? *Modern Economy*, *10*(8), Article 8. https://doi.org/10.4236/me.2019.108120

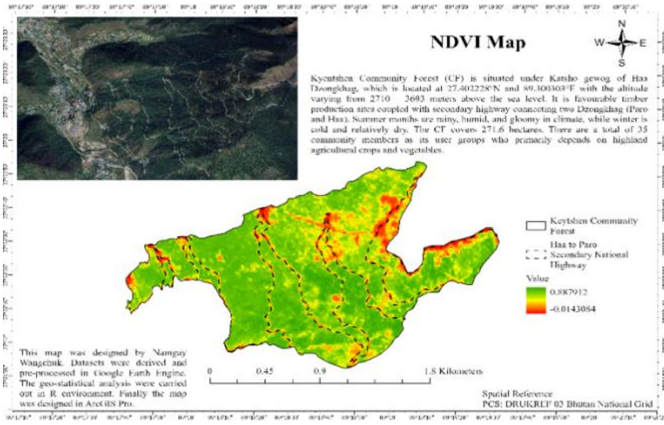
Yangka, D., Rauland, V., & Newman, P. (2019). Carbon neutral policy in action: The case of Bhutan. *Climate Policy*, *19*(6), 672–687. https://www.tandfonline.com/doi/abs/10.1080/14693062.2018.1551187

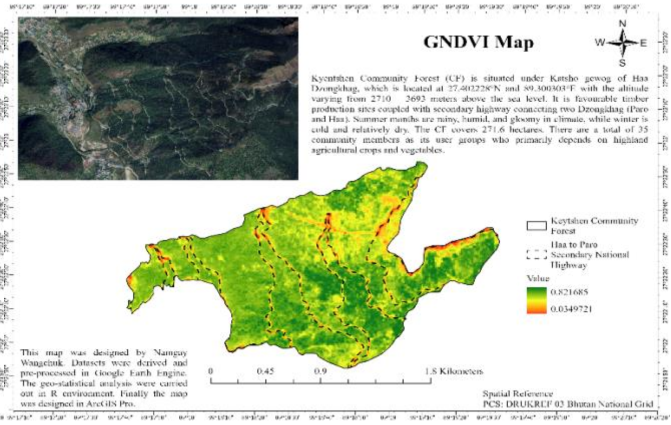
Zurqani, H. A. (2025). A multi-source approach combining GEDI LiDAR, satellite data, and machine learning algorithms for estimating forest aboveground biomass on Google Earth Engine platform. *Ecological Informatics*, *86*, 103052. https://doi.org/10.1016/j.ecoinf.2025.103052

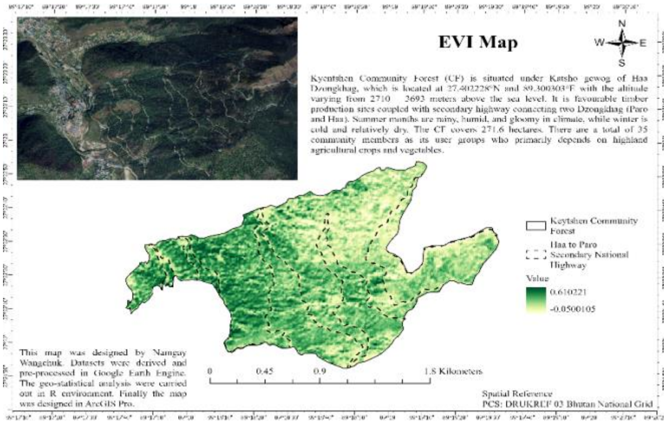
Annexes:

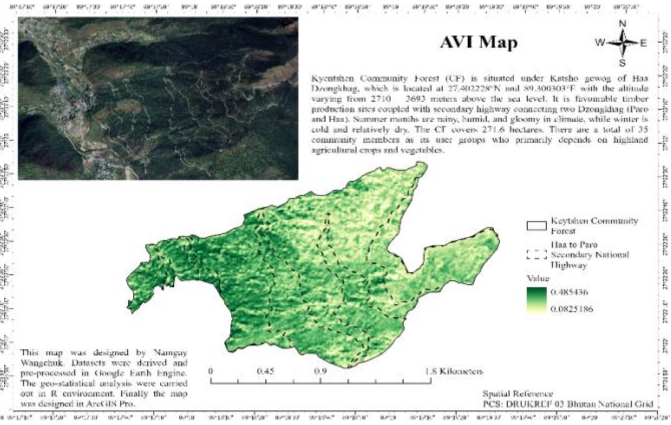
ANNEX A: Spatial Distribution of Predictor Variables

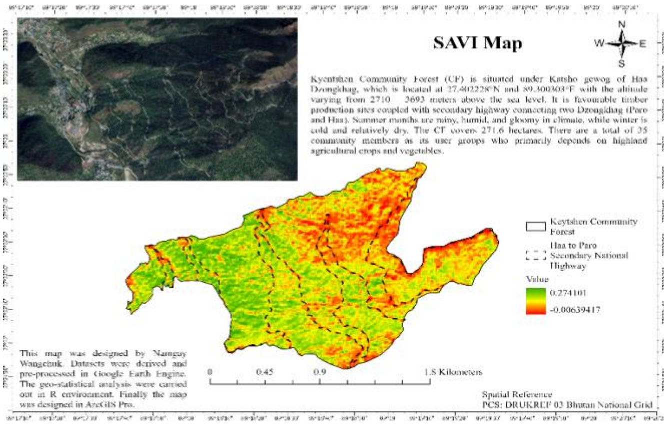
Contains: Figure 10 (maps of all 27 predictor variables)

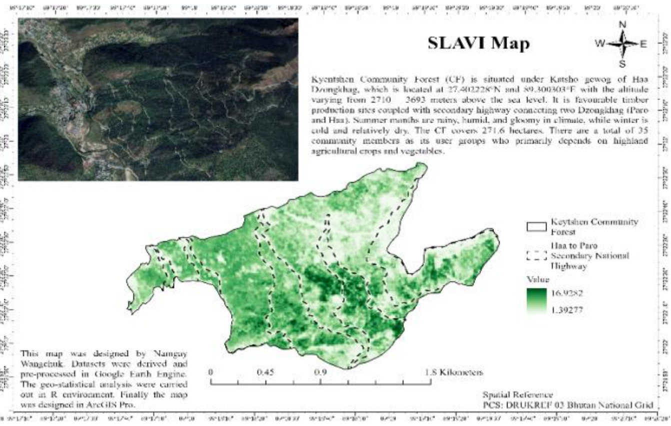


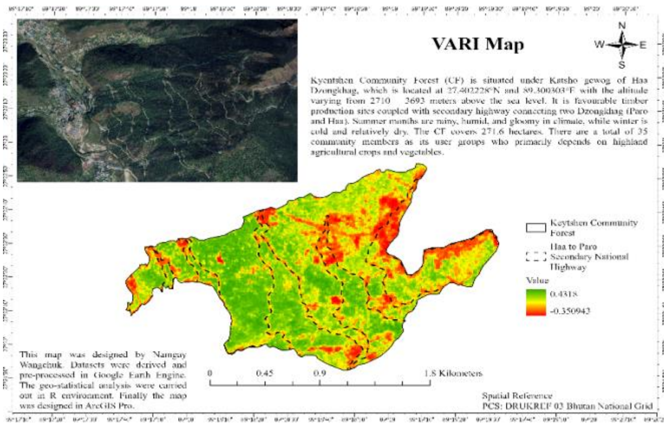


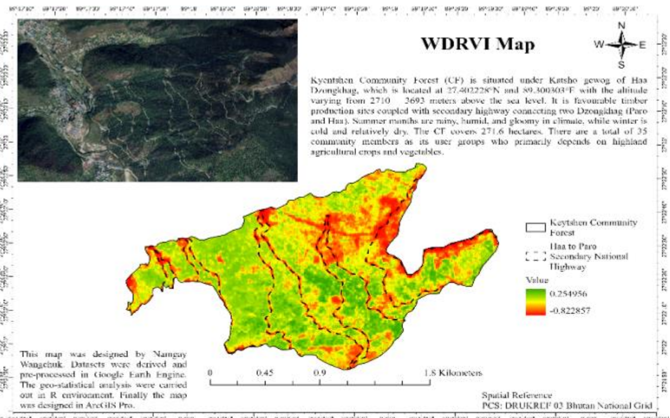


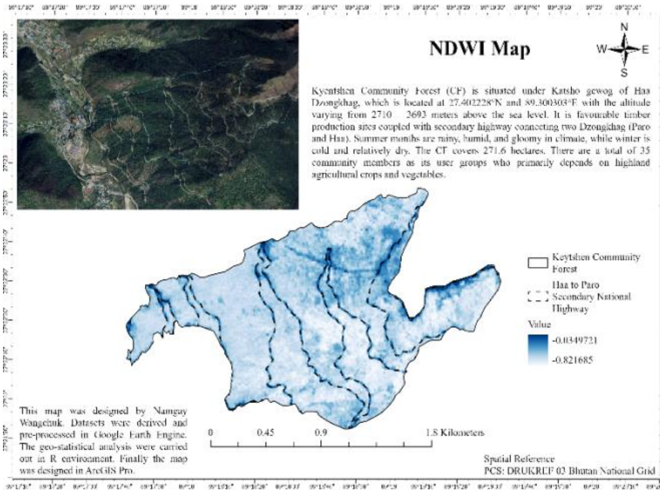


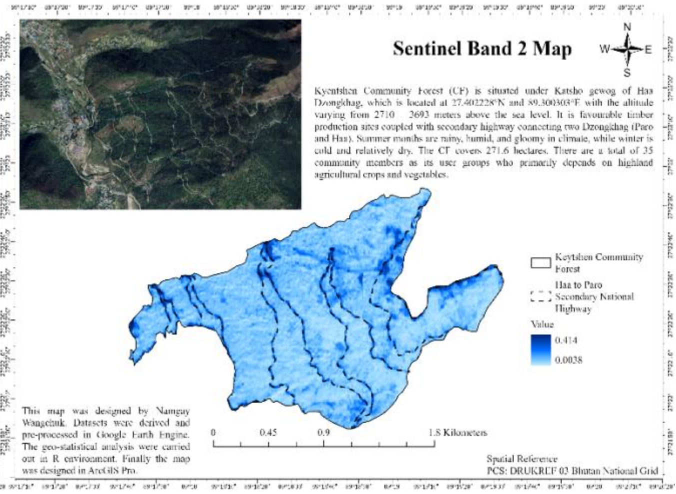


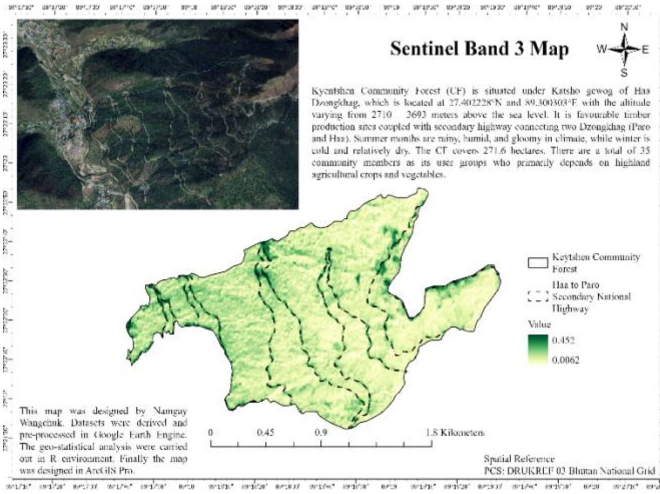


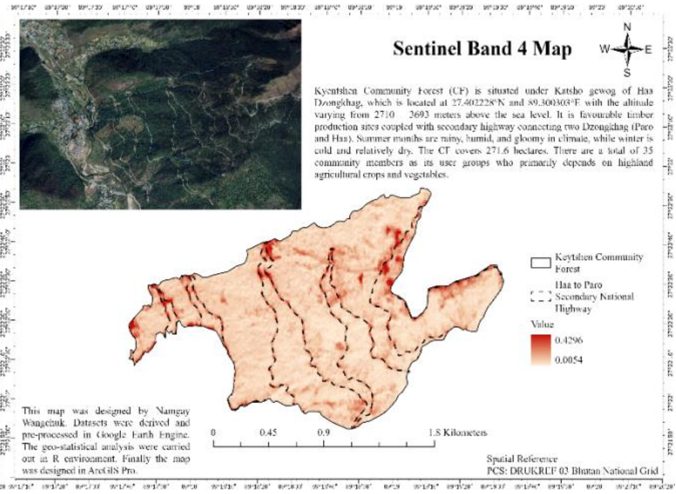


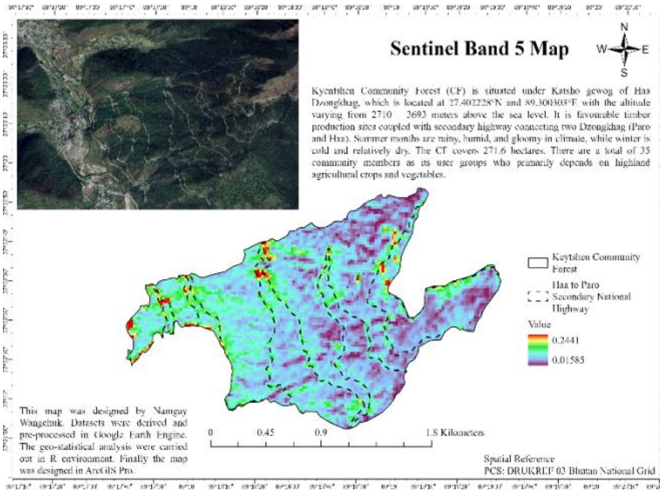


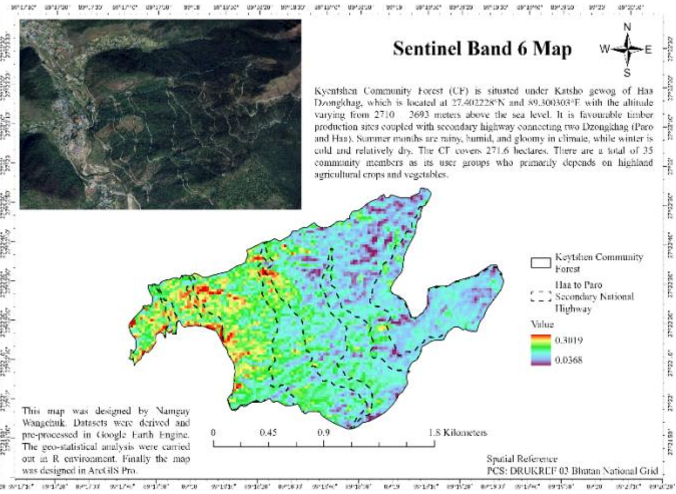

NDWI Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 2693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CT covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
-0.0349721
-0.821685
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

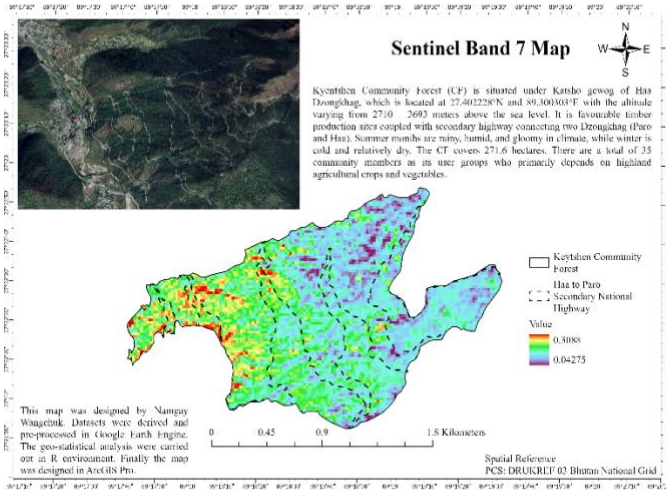

Sentinel Band 2 Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 2693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CT covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.414
0.0038
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

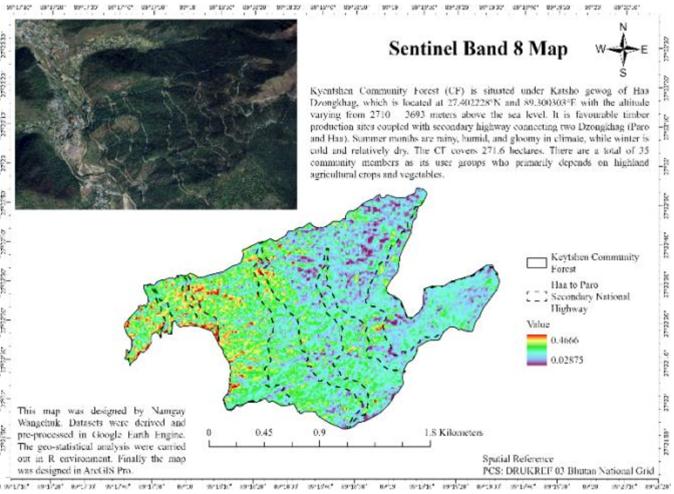

Sentinel Band 3 Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 2693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CT covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.452
0.0062
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

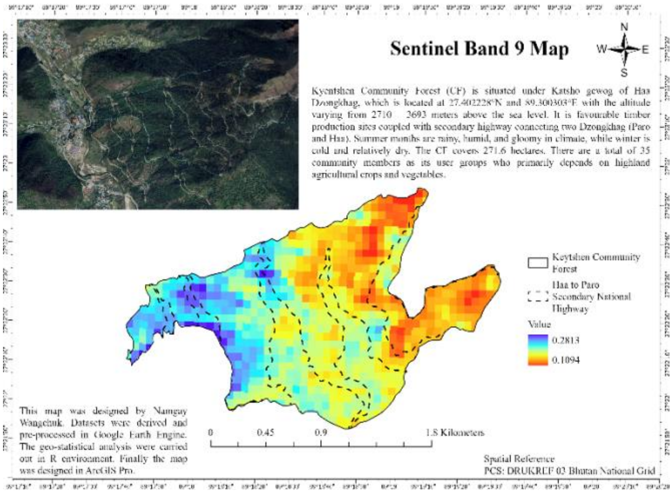

Sentinel Band 4 Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 2693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CT covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.4296
0.0054
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

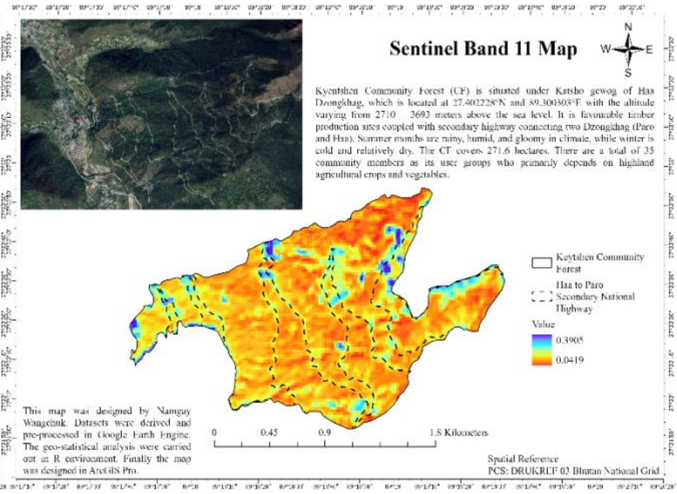

Sentinel Band 5 Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 2693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CT covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.2441
0.01585
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

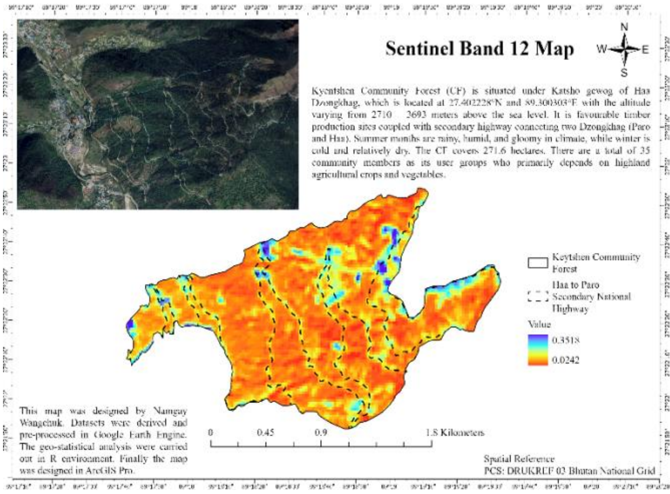

Sentinel Band 6 Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 2693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CT covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.3019
0.0368
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

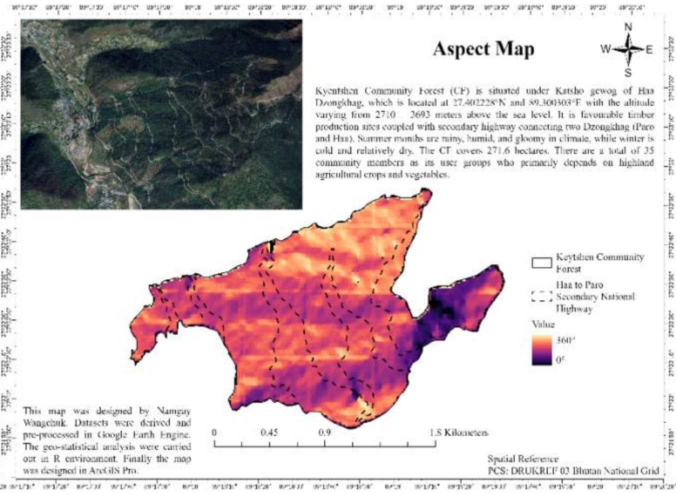

Sentinel Band 7 Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 2693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CT covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.3888
0.04275
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

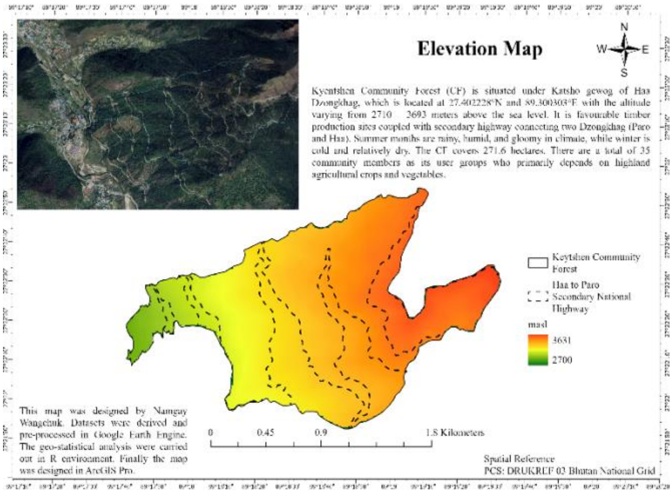

Sentinel Band 8 Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 2693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CT covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.4666
0.02875
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

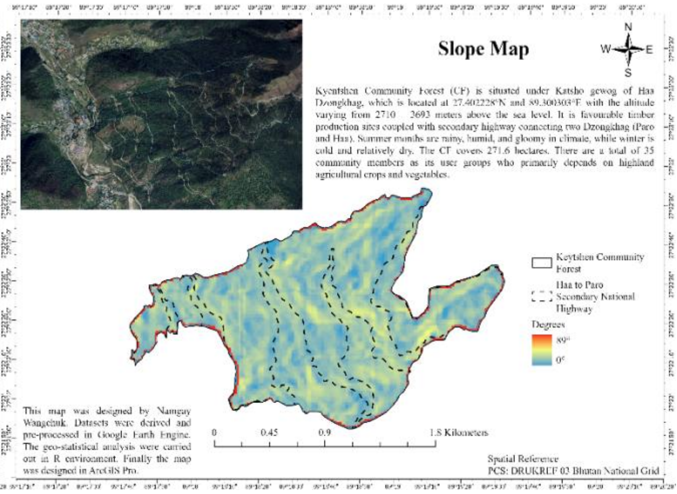

Sentinel Band 9 Map
N
W
E
S
Kyentshen Community Forest (CF) is situated under Katsho gewog of Haa Dzongkhag, which is located at 27.402228°N and 89.300303°E with the altitude varying from 2710 – 3693 meters above the sea level. It is favourable timber production sites coupled with secondary highway connecting two Dzongkhag (Paro and Haa). Summer months are rainy, humid, and gloomy in climate, while winter is cold and relatively dry. The CF covers 271.6 hectares. There are a total of 35 community members as its user groups who primarily depends on highland agricultural crops and vegetables.
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.2813
0.1094
This map was designed by Namgay Wangchuk. Datasets were derived and pre-processed in Google Earth Engine. The geo-statistical analysis were carried out in R environment. Finally the map was designed in ArcGIS Pro.
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

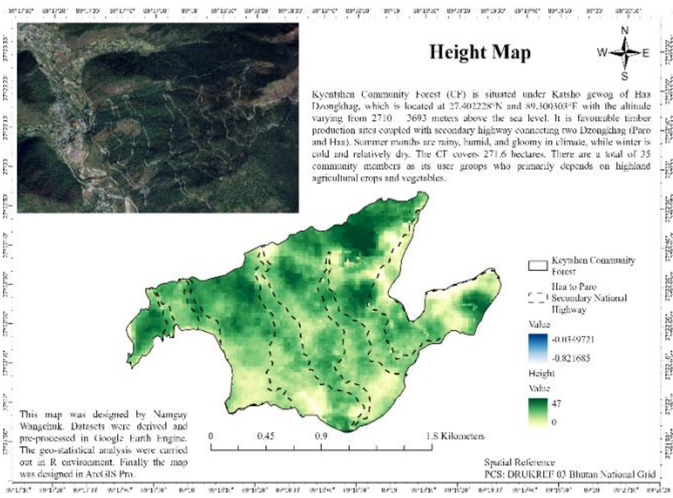

Sentinel Band 11 Map
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.3905
0.0419
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

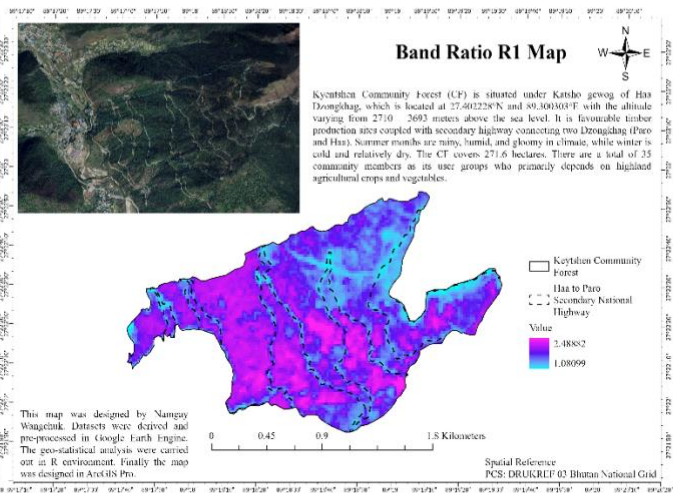

Sentinel Band 12 Map
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
0.3518
0.0242
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

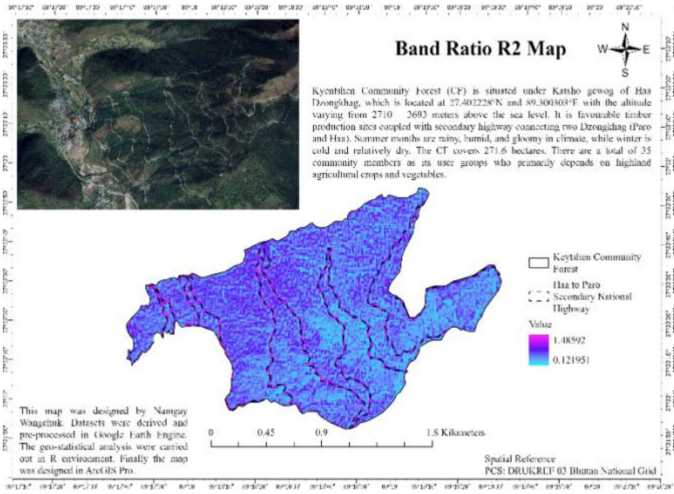

Aspect Map
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
360°
0°
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

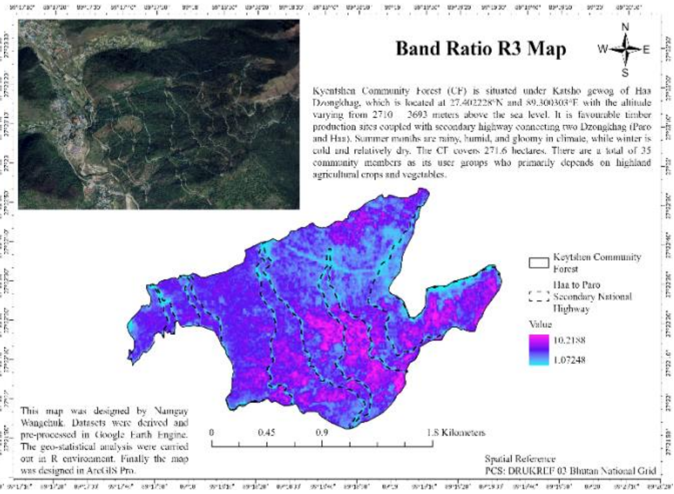

Elevation Map
Keytshen Community Forest
Haa to Paro Secondary National Highway
masl
3631
2700
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid

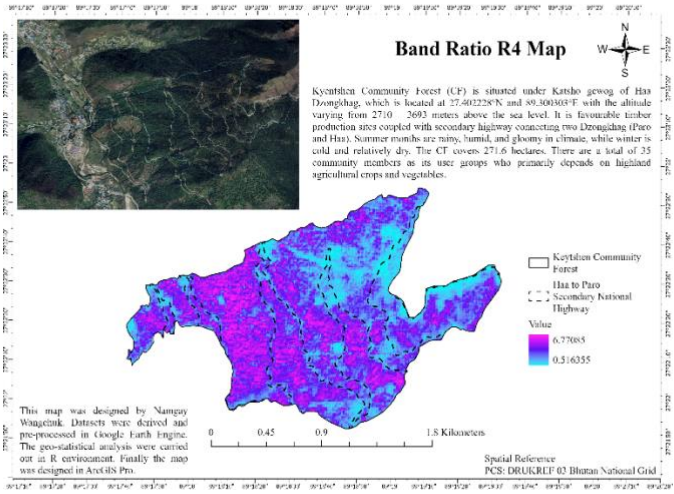

Slope Map
Keytshen Community Forest
Haa to Paro Secondary National Highway
Degrees
89°
0°
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid


Height Map
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
-0.0349721
-0.821685
Height
Value
47
0
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid


Band Ratio R1 Map
Keytshen Community Forest
Haa to Paro Secondary National Highway
Value
2.48882
1.08099
0
0.45
0.9
1.8 Kilometers
Spatial Reference
PCS: DRUKREF 03 Bhutan National Grid







ANNEX B: Predictor Variable Specifications and Descriptive Statistics

Contains: Table 2 (predictor catalog) + Table 3 (descriptive statistics)

Table 2: List of predictors that were used to compute forest AGB

Predictors	Name	Specifications	References
Topographic Variables (3)			
	Elevation	Metre	(NLCS)
	Slope	Degrees	(NLCS)
	Aspect	Cardinal direction	(NLCS)
Geophysical Variables (4)			
	Canopy cover	Percentage	
	DBH	Centimetre	
	Tree height	Metre	(Lang et al., 2023)

Figure 10: Maps of the predictor variables used in the study

	Geo-coordinates of individual tree	Degrees-minutes-seconds (DMS)	
Sentinel-2A band reflectance (10)			
	Band 2 (Blue)	10m, 490 nm	(Kirches et al., 2024)
	Band 3 (Green)	10m, 560 nm	
	Band 4 (Red)	10m, 665 nm	
	Band 5 (Red edge 1)	20m, 705 nm	
	Band 6 (Red edge 2)	20m, 749 nm	
	Band 7 (Red edge 3)	20m, 783 nm	
	Band 8 (Near Infrared)	10m, 842 nm	
	Band 8A (Red edge 4)	20m, 865 nm	
	Band 11 (Shortwave Infrared)	20m, 1610 nm	
	Band 12 (Shortwave Infrared)	20m, 2190 nm	
Sentinel-2A derived Spectral Indices' (9)			
	Normalised Differenced Vegetation Index (NDVI)	(NIR - Red) / (NIR + Red)	(Rouse *et al.*, 1973)
	Green Normalised Difference Vegetation Index (GNDVI)	(NIR–Green)/ (NIR + Green)	(Gitelson *et al.*, 1996)
	Enhanced Vegetation Index (EVI)	2.5 × (NIR-Red)/(NIR + 6 × Red 7.5xBlue) + 1	(Huete *et al.*, 2002)
	Advanced Vegetation Index (AVI)	2.5 * ((B8 – B4) / (B8 + 6 * B4 – 7.5 * B2 + 1))	(Gitelson, 2004)
	Soil Adjusted Vegetation Index (SAVI)	(1 + L)((NIR–Red)/ (NIR + Red + L))	(Hute, 1988)
	Specific Leaf Area Vegetation Index (SLAVI)	(NIR/Red) + SWIR	(Lymburner *et al.*, 2000)
	Normalised Difference Water Index (NDWI)	(Green–NIR)/ (Green + NIR)	(Gao, 1996)

	Visible Atmospherically Resistant Vegetation Index (VARI)	(Green–Red)/ (Green + Red–Blue)	Gitelson (2002)
	Wide Dynamic Range Vegetation Index (WDRVI)	((0.1 × NIR) − Red/ (0.1 × NIR) + Red)	Gitelson (2004)
Sentinel-2A derived Band Ratios (4)			
	r1	SWIR1/SWIR2	(Purohit *et al.*, 2020)
	r2	Red/NIR	
	r3	NIR/Green	
	r4	NIR/SWIR	

Table 3: Descriptive statistics of all the predictor variables

Predictors	**Min**	**Max**	**Mean**	**Median**	**StdDev**
Aspect (°)	0.000	359.862	250.256	253.520	47.285
AVI	0.082	0.489	0.265	0.263	0.057
B11	0.041	0.394	0.113	0.103	0.041
B12	0.022	0.352	0.063	0.052	0.032
B2	0.003	0.423	0.019	0.017	0.011
B3	0.006	0.452	0.032	0.030	0.015
B4	0.005	0.430	0.030	0.024	0.020
B5	0.014	0.244	0.061	0.057	0.021
B6	0.033	0.302	0.130	0.126	0.035
B7	0.042	0.309	0.150	0.145	0.038
B8	0.028	0.467	0.158	0.153	0.047
B8A	0.048	0.332	0.167	0.163	0.040
B9	0.109	0.287	0.172	0.168	0.031
Elevation (masl)	2712.000	3634.000	3217.731	3271.577	564.088
EVI	-0.050	0.621	0.265	0.260	0.086
GNDVI	0.032	0.827	0.663	0.682	0.086
Height (m)	5.000	47.000	32.430	32.000	4.753
NDVI	-0.014	0.890	0.687	0.736	0.144
NDWI	-0.827	-0.032	-0.663	-0.682	0.086
r1	1.101	2.497	1.851	1.848	0.239

r2	0.115	1.502	0.445	0.433	0.115
r3	1.067	10.556	5.270	5.280	1.420
r4	0.513	6.924	2.800	2.733	1.139
SAVI	-0.006	0.276	0.143	0.144	0.042
SLAVI	1.393	17.252	6.660	6.657	2.805
Slope (°)	0.000	89.717	24.218	23.820	13.165
VARI	-0.359	0.432	0.091	0.123	0.173
WDRVI	-0.823	0.263	-0.245	-0.207	0.216

ANNEX C: Recursive Feature Elimination Model Performance
Contains: Table 4 (RFE optimization results for 5-28 variables)

Table 4: Model summary of the RFE for selecting the most important predictors for AGB estimation

No. of Variables	***RMSE* (t/ha)**	R^2	***MAE* (t/ha)**
5	76.11	0.26	59.64
6	73.4	0.28	57.13
7	73.73	0.26	57.79
8	72.99	0.27	57.44
9	72.98	0.27	57.21
10*	**72.63***	**0.28***	**56.72***
11	74.69	0.25	58.24
12	73.65	0.26	57.67
13	74.02	0.27	58.29
14	74.12	0.26	58.39
15	74.21	0.26	58.4
28	75.36	0.27	60.31

*Selected

ANNEX D: Random Forest Hyperparameter Tuning Results
Contains: Table 5 (RF optimization with different Mtry, splitrule, min.node.size)

5: Model summary of RF based on the permutation importance

M_{try}	**splitrule**	**min.node.size**	***RMSE* (t/ha)**	R^2	***MAE* (t/ha)**
3	variance	5	68.17	0.34	52.81
3*	**variance***	**10***	**67.98***	**0.34***	**52.31***
3	extratrees	5	70.60	0.34	55.19
3	extratrees	10	70.66	0.33	55.92
10	variance	5	70.86	0.32	57.06
10	variance	10	69.13	0.33	54.15
10	extratrees	5	69.27	0.35	53.73
10	extratrees	10	69.48	0.35	53.87

*Selected

ANNEX E: Tree Species and Wood Density Reference Database
Contains: Table 6 (KCF tree species with wood density values)

Table 6: Tree species found in KCF along with their wood densities

Species	**Wood Density**
Abies densa	0.50
Euyria sp	0.62
Juniperus recurva	0.30
Larix griffithii	0.46
Lyonia ovalifolia	0.50
Picea spinulosa	0.39
Pinus wallichiana	0.48
Poplus sp	0.40
Quercus semecarpifolia	0.86
Rhododendron aroboreum	0.64