Article

Toward a More Robust Estimation of Forest Biomass Carbon Stock and Carbon Sink in Mountainous Region: A Case Study in Tibet, China

Guanting Lyu 1,2, Xiaoyi Wang 1,*, Xieqin Huang 3, Jinfeng Xu 1,4, Siyu Li 5, Guishan Cui 5,6 and Huabing Huang 6

1 State Key Laboratory of Tibetan Plateau Earth System, Resources and Environment (TPESRE), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China; guantinglv@itpcas.ac.cn (G.L.); jinfengxu@itpcas.ac.cn (J.X.)
2 University of Chinese Academy of Sciences, Beijing 100101, China
3 Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; hxq23@mails.tsinghua.edu.cn
4 College of Ecology, Lanzhou University, Lanzhou 730000, China
5 College of Geography and Ocean Sciences, Yanbian University, Yanji 133002, China; honertjecodin@gmail.com (S.L.); cuiguishan@ybu.edu.cn (G.C.)
6 School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai 519082, China; huanghb55@mail.sysu.edu.cn
* Correspondence: xiaoyiwang@itpcas.ac.cn; Tel.: +86-010-8409-7037

Abstract: Mountainous forests are pivotal in the global carbon cycle, serving as substantial reservoirs and sinks of carbon. However, generating a reliable estimate remains a considerable challenge, primarily due to the lack of representative in situ measurements and proper methods capable of addressing their complex spatial variation. Here, we proposed a deep learning-based method that combines Residual convolutional neural networks (ResNet) with in situ measurements, microwave (Sentinel-1 and VOD), and optical data (Sentinel-2 and Landsat) to estimate forest biomass and track its change over the mountainous regions. Our approach, integrating in situ measurements across representative elevations with multi-source remote sensing images, significantly improves the accuracy of biomass estimation in Tibet’s complex mountainous forests ($R^2 = 0.80$, root mean squared error = 15.8 MgC ha$^{-1}$). Moreover, ResNet, which addresses the vanishing gradient problem in deep neural networks by introducing skip connections, enables the extraction of complex spatial patterns from limited datasets, outperforming traditional optical-based or pixel-based methods. The mean value of forest biomass was estimated as 162.8 $±$ 21.3 MgC ha$^{-1}$, notably higher than that of forests at comparable latitudes or flat regions in China. Additionally, our findings revealed a substantial forest biomass carbon sink of 3.35 TgC year$^{-1}$ during 2015–2020, which is largely underestimated by previous estimates, mainly due to the underestimation of mountainous carbon stock. The significant carbon density, combined with the underestimated carbon sink in mountainous regions, emphasizes the urgent need to reassess mountain forests to better approximate the global carbon budget.

Keywords: mountain forest; Sentinel-1; VOD; landsat; forest biomass; carbon sink; deep learning

1. Introduction

Mountainous forests, spanning extensive territories, are pivotal in the sequestration of atmospheric carbon [1]. Refining our estimates of carbon reservoirs within these biomes is crucial for climate change mitigation strategies and to inform sustainable forestry [2–5]. Nonetheless, the diverse forest types, challenging topography, and heterogeneous climate of mountain regions impose significant difficulties in calculating forest biomass and carbon sequestration rates [6,7]. Across the elevational gradient, forest types transition from broadleaf to mixed, and then to coniferous forests and eventually an alpine tree line, which...
occurred over short horizontal spans, and paralleled by varied biomass. Marked biomass
differences also persist between sunlit and shaded slopes in mountain regions. Traditional
assessment methods could not capture the intricate spatiotemporal dynamics of mountain
forests, hindered by limited in situ measurements and methods suitable for their complex
spatial variations [8]. Consequently, accurate biomass and carbon sequestration estimates
for mountainous terrain remained challenging [9].

First, in situ measurements in mountainous areas were limited, as monitoring sites
often were unevenly distributed and too few to accurately reflect the complex spatial
variations, especially across large elevational gradients. Such insufficiency compromised
the accuracy of forest carbon stock assessments and their variation within these complex
landscapes. For example, a widely used forest biomass inventory dataset for Eurasia [10]
included only 30.6% of measurements from mountain regions as defined by the Global
Mountain Biodiversity Assessment (GMBA) datasets [11], and included only 19.9% of
measurements which were located in rugged regions with slopes greater than 10°. This
deficiency was more pronounced across China, where only 6.45% of measurements were
located in mountain regions with an elevation higher than 1000 m. Therefore, there was
an emergent need to conduct widespread measurements over mountain regions [12,13].
However, the remoteness and access difficulties in mountainous regions made extensive
field surveys challenging, often precluding the collection of widespread observational
data [14,15]. Thus, determining whether representative sampling along elevational gradi-
ents could enable efficient monitoring and enhance forest biomass estimations emerged as
a primary issue to address.

Second, forest biomass carbon stock and sink estimations had increasingly relied
on remote sensing data, ranging from optical images (e.g., Landsat, Sentinel-2, IKONOS,
QuickBird) to SAR (e.g., Sentinel-1, SSDD, ALOS-2) and VOD (e.g., L-VOD, X-VOD, P-VOD)
datasets, each offering unique insights into vegetation structure [16–18]. However, how to
effectively combine multi-source remote sensing data remained to be solved. Specifically,
optical remote sensing imagery provided spectral reflectance of the forest canopy, enabling
the acquisition of features related to the forest’s top canopy, notably leaves [19,20]. However,
it fell short in detecting the structural components of trunks and branches, resulting in
potential bias in forest carbon estimations [21]. Advanced remote sensing techniques such as
LiDAR [22,23] and microwave remote sensing [24,25] could effectively penetrate the forest
canopy, provide forest structure for biomass analysis, and were less susceptible to weather
conditions [26]. For example, C-band SAR data provided by Sentinel-1A/B satellites aboard
the ESA (European Space Agency) had been consistently providing valuable data on forest
vertical structure through multiple polarization modes since 2014 [27,28]. There have
been studies showing the proficiency of SAR data in regional biomass studies [29,30], yet
the penetration power of SAR has limitations in dense or high-biomass forests where it
might underestimate biomass [31]. The L-band vegetation optical depth (L-VOD), which
 correlates with aboveground biomass carbon and water content within vegetation, had
been extensively utilized as a proxy for vegetation biomass to monitor changes in carbon
stocks [32,33]. VOD data, notably sensitive to woody biomass and non-saturating at high
levels due to low-frequency microwaves [34], offered an effective biomass proxy. For
instance, Fan et al. [35] detected changes in terrestrial tropical carbon stocks using L-VOD
and found that the tropics were roughly carbon neutral during 2010–2017. Wigneron
et al. [36] further showed that tropical forest biomass was a carbon source during the strong
2015–2016 El Niño event and did not recover to the pre-event level. While methods utilizing
L-VOD were constrained by the assumption that the plant water-holding capacity remained
constant. This assumption could potentially be challenged by climate variations and the
physiological responses of vegetation [37–39]. Therefore, how to effectively combine the
multi-source remote sensing data that reflect the multi-dimensional forest structure and
attributes to accurately quantify forest biomass remained an unresolved issue.

Third, traditional methods that upcaled in situ measurements to regional wall-to-
twall maps failed to capture the detailed spatial variations present in mountainous terrain.
Such methods often employed pixel-based modeling, which neglected neighboring pixel information, including textural patterns crucial for depicting spatial variation [40,41]. However, with advancements in deep learning, particularly Convolutional Neural Networks (CNNs), these methods showed promise in overcoming various geospatial challenges, as evidenced by numerous researchers [42–44]. Unlike traditional pixel-based approaches that formed predictions from isolated pixels, CNNs leveraged the spatial context by considering information from neighboring pixels, thus offering a more complex representation of forest biomass distribution [45]. Integrating CNNs with multi-source remote sensing data, which provided insights into different forest components such as leaves, branches, and trunks, held the potential to significantly improve the mapping of regional biomass and carbon sinks.

Forests in Tibet, characterized by intact and natural state [45–47], potentially hold considerable carbon stock and serve as substantial carbon sinks. The elevation of forest over Tibet ranges from 8 to 4900 m. The rugged and complex terrain poses challenges for traditional methods of estimating biomass in mountainous forests, making it an ideal case study for our research. Hence, our objective is to devise a robust methodology for estimating forest biomass and biomass carbon sink, with a particular focus on mountainous forests, exemplified by those in the mountain regions of Tibet. Our research aimed to answer the following key questions: (1) whether representative in situ measurements along elevation could improve mountain forest biomass. (2) Do multi-source remote sensing data outperform single source results? (3) Whether the patch-based deep learning method outperforms the traditional pixel-based method in mountain forest biomass and carbon sink estimation. (4) The size of forest biomass carbon stock and carbon sequestration rate over Tibet.

2. Data Collection

2.1. Study Area

The Tibet Autonomous, situated in southwestern China, boasts a significant altitude range. This considerable variation in altitude harbors diverse climate types across Tibet [48], facilitating the growth of plants ranging from subtropical to boreal species. Consequently, Tibet harbors a forest ecosystem of exceptional biodiversity, earning recognition as one of China’s three major forested regions [46]. Forest distribution in Tibet is predominantly concentrated in the mountains and valleys of the Himalayas and the Hengduan Mountains in the southeast. Noteworthy is the Nyingchi Prefecture, home to China’s largest primeval forest area and revered as one of the most representative regions for global biodiversity.

2.2. Forest Inventory Data

We conducted in situ measurements \( (N = 67, \text{red dots in Figure 1}) \) during 2019 and 2023 with a special focus on the elevational gradient. Here, we set the elevation intervals as 100 m, ensuring that at least one sample was collected every 100 m from 33 m (a.s.l.) to 4500 m (a.s.l.). Each plot was set as a square of \( 10 \times 10 \text{ m} \). We then measured the diameter at breast height (DBH), and height of each tree (DBH > 5 cm) and also recorded tree types. The forest aboveground biomass was obtained using the allometric model of each tree type. The average biomass of these sample points is 135.4 MgC ha\(^{-1}\), with an average elevation of 4673.4 m. In addition, we compiled data documented in the literature \( (N = 98, \text{blue dots in Figure 1}) \). Data should be measured near or after the year 2015, and satisfy the following two criteria (1) the inventory should be conducted at the plot scale rather than at the individual tree level [49]; (2) data should be collected from stable sites with no recent fires or human disturbances, verified through high-resolution imagery on Google Earth (interpreted manually from Google Earth high-resolution images). Data recorded only forest aboveground biomass was transformed into biomass by the shoot-to-root ratio of 0.35, following Liu et al. [8]. The distribution of inventory data which met those criteria is shown in Figure 1. However, the observation data were obtained between
2015 and 2023, while our analysis focused on the period from 2015 to 2020. This temporal mismatch could potentially introduce uncertainties in our estimation.

![Figure 1. Distribution of forests and in situ measurements over Tibet.](image)

2.3. Microwave Remote Sensing Images

The Sentinel-1 data offer backscatter information in the Interferometric Wide Swath (IW) mode, capturing both VV (Vertical transmit–Vertical receive) and VH (Vertical transmit–Horizontal receive) polarizations at a spatial resolution of 10 m. We acquired median value data spanning the growing seasons (May to September) in 2015 and 2020 using the Google Earth Engine (GEE) platform [50] (https://earthengine.google.org/, accessed on 10 January 2024). These datasets underwent meticulous pre-processing, including noise elimination, terrain correction, and radiometric calibration, to ensure accuracy and reliability.

The vegetation optical depth (VOD) data were from the SMOS-ICV2-RE06 L-VOD product (L-VOD), which provides monthly VOD data at different microwave band frequencies since 2010. The VOD measured at microwave frequencies is related to the vegetation water content. And the vegetation water content can be related to the biomass and to its moisture status, which can be parameterized by the gravimetric moisture content Mg (%). The VOD data we utilized also originate from both the 2015 and 2020 periods, and we downscaled L-VOD data to 10 m resolution by linear spatial interpolation.

2.4. Optical Remote Sensing Image

We also obtained top-of-atmosphere reflectance data from Sentinel-2 level 2A and Landsat 8 images during the growing seasons (June to September) of 2015 and 2020 via the GEE platform. We selected five bands, including green, red, near-infrared, and two shortwave infrared bands (Table 1). To ensure compatibility with the Sentinel-1 data, we applied the nearest-neighbor resampling method to adjust the Landsat data to a 10-m resolution, thereby minimizing uncertainties due to disparities of spatial scale between the two datasets.
Table 1. Details of the multi-source remote sensing imagery used in this study.

<table>
<thead>
<tr>
<th>Multi-Source Remote Sensing Dataset</th>
<th>Band</th>
<th>Description</th>
<th>Spatial Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentinel 1</td>
<td>VV</td>
<td>Vertical transmit–Vertical receive</td>
<td>10 m</td>
</tr>
<tr>
<td>VH</td>
<td>VH</td>
<td>Vertical transmit–Horizontal receive</td>
<td>10 m</td>
</tr>
<tr>
<td>VOD SMOS-ICV2-RE06</td>
<td>L-VOD</td>
<td>VOD</td>
<td>0.25°</td>
</tr>
<tr>
<td>Sentinel-2A</td>
<td>Band 3</td>
<td>Green</td>
<td>10 m</td>
</tr>
<tr>
<td>Band 4</td>
<td>Red</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Band 8</td>
<td>Near-infrared</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Band 11</td>
<td>Shortwave infrared</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Band 12</td>
<td>Shortwave infrared</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landsat 8 OLI</td>
<td>Band 3</td>
<td>Green</td>
<td>30 m</td>
</tr>
<tr>
<td>Band 4</td>
<td>Red</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Band 5</td>
<td>Near-infrared</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Band 6</td>
<td>Shortwave infrared</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Band 7</td>
<td>Shortwave infrared</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Method

To provide a benchmark of mountainous forest biomass and estimated changes in biomass, that is the biomass carbon sink, we proposed a deep learning method that integrated a Residual Convolutional Neural Networks based model, ResNet, with a relatively large number of in situ measurement data, microwave remote sensing (Sentinel-1 and VOD), and optical imagery (Sentinel-2A and Landsat 8). Initially, we compiled multi-source remote sensing data covering the years 2015 to 2020. We then aligned in situ measurements with the remote sensing data to ensure spatial consistency, incorporating neighboring information as additional input for ResNet regression (Step 1 in Figure 2). Subsequently, we trained and optimized the ResNet model to minimize the loss value (Step 2 in Figure 2). The well-optimized ResNet model was then deployed to predict regional forest biomass at the pixel level. Changes in forest biomass were subsequently translated into forest biomass carbon sink (Steps 3 and 4 in Figure 2).

Figure 2. Schematic flow of the mountain forest biomass and biomass carbon sink mapping proposed in this study.
3.1. ResNet CNN Model

The principle of Convolutional Neural Networks (CNN) was that the convolutional structure was added to the traditional Deep Neural Network (DNN) model. The ResNet model introduced a novel architectural innovation known as residual learning to the traditional CNN model, primarily through the use of residual blocks [51]. Each block contained a shortcut connection that bypassed one or more layers with an identity mapping. This approach allowed the model to learn the additive residuals of the underlying mapping, rather than the entire mapping at once, which avoided the vanishing gradients and the degradation problem, and enabled much deeper networks to be trained effectively. Furthermore, the core of the algorithm are convolutional layers, which apply convolution operations to the input images using learnable filters. These filters act as feature detectors, capturing patterns and spatial relationships within the images. Following convolution, pooling layers are employed to down-sample the feature maps, reducing computational complexity and preventing overfitting by retaining only the most relevant information. The extracted features are then passed through fully connected layers, where they are transformed and aggregated to produce the final regression output. Each layer is activated using rectified linear unit (ReLU) activation functions, facilitating non-linear transformations and enhancing model expressiveness. In our case, we constructed the ResNet model under the Google TensorFlow framework [52]. Our ResNet architecture comprises a series of convolution layers, repeated four times in this instance, followed by a pooling layer and six fully connected layers. The parameters employed in this architecture include Convolution layers: these utilize a kernel size of $3 \times 3$, a stride size of $1 \times 1$, and a filter size initially set to 64, with an increment by a factor of 2 as the model depth increases. The filter size regulates the number of features extracted by convolution. Pooling layers: utilizing max pooling, these layers extract the maximum value of each feature with a stride size of $1 \times 1$. Residual blocks: these blocks consist of multiple convolutional layers followed by identity shortcut connections, which skip one or more layers. Fully connected layers: the architecture includes six fully connected layers. These parameters are carefully chosen to optimize feature extraction and model performance across varying depths. To optimize the algorithm's performance, we employed techniques such as grid search to fine-tune hyperparameters, weight regularization (L2 regularization) to prevent overfitting, and dropout layers to randomly deactivate neurons during training, thereby promoting model generalization.

3.2. Other Machine Learning

To demonstrate the proficiency of the ResNet model in estimating forest biomass carbon, we conducted a comparative analysis with traditional machine learning approaches, including Random Forest (RF), Support Vector Machine (SVM), eXtreme Gradient Boost (XGBoost) and a Deep Neural Network (DNN) that had no convolutional layers to extract neighboring information, and two advanced deep learning methods, AlexNet and VGG. We optimized and fine-tuned all the models to ensure that the superior performance of ResNet was not solely due to model tuning. The RF model consisting of 3000 decision trees (ntree = 3000), an SVM model with a “rbfdot” kernel and a penalty coefficient of $C = 5$ and a XGBoost with 1000 base-classifiers and 0.01 learning rate was ultimately selected (Figure 2). Furthermore, to determine whether the convolutional structure was the main reason the ResNet model effectively captured forest biomass in mountainous areas, we also established a DNN model without a convolutional structure. Parameters were set to be consistent with the fully connected structure in a CNN. AlexNet and VGG share the same convolutional layer parameters ($3 \times 3$ kernel size; 64 filter size) and fully connected structure (6 layers with L2 regularization and Dropout) with ResNet. However, they lack residual modules, and AlexNet has shallower convolutional layers (3 layers) compared to VGG and ResNet (8 layers).
3.3. Validation and Accuracy Assessment

We employed a 10-fold cross-validation method to randomly reserve 20% of the training data as independent validation data to assess the uncertainty of biomass mapping at field measurement scales. Specifically, the cross-validation process was repeated 10 times, with each validation dataset being non-repeated to reduce the particularity of validation results. In addition, we extracted biomass values corresponding to the training sample locations from the forest biomass dataset used for comparison, and evaluated the dataset accuracy by contrasting their differences with the true values at sample points. We also randomly selected 80% of the training samples for model training, with the remaining 20% used for model validation. For accuracy quantification, we utilized the coefficient of determination ($R^2$) and root mean square error (RMSE) (Equations (1) and (2)).

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2}$$  \hspace{1cm} (1)

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}$$  \hspace{1cm} (2)

where $m$ is the number of observations, $y_i$ and $\hat{y}_i$ is the $i$th observed dependent variable value and the predicted $y$ value for the $i$th observation from the regression model, respectively; and $\bar{y}_i$ is the mean of all $y$ values.

4. Result and Discussion

4.1. Validation of Mountain Forest Biomass Mapping

To validate the robustness of our proposed deep learning-based algorithm, we randomly reserved 20% of in situ measurements as independent validation samples. Figure 3 illustrates the scatter plot presenting this comparison. Each point on the plot represented a measured biomass value along with the corresponding prediction generated by our ResNet model. The validation showed that the proposed model could capture well the variation in mountain forest biomass at the plot scale, revealing a high coefficient of determination ($R^2$) of 0.80. Furthermore, the root mean square error (RMSE) was found to be 15.8 MgC ha$^{-1}$, indicating a low level of discrepancy between the measured and predicted forest biomass values. Notably, the RMSE constituted only 10.1% of the average forest biomass estimated by the ResNet model, highlighting the model’s robust performance in accurately estimating forest biomass in mountainous regions.

![Figure 3](image-url)  \hspace{1cm} Figure 3. Relationship between the independent in situ measurement and the predicted forest biomass. The solid blue line and the dashed black line represent the regression line and 1:1 line, respectively. Here, independent samples ($N = 62$) were only used for verification. $R^2$ represents the coefficient of determination, and RMSE is the root mean squared error.
4.2. Comparative Studies

4.2.1. Comparison in Different Field Sampling Strategies

To reveal the importance of in situ observations across elevational gradients in estimating mountainous forest biomass, we compared results using elevation gradient sampling ($N = 67$), and with traditional random sampling from the literature review ($N = 98$). The comparison revealed that elevational gradient sampling strategy, though small in numbers (67 vs. 98), surpassed traditional random sampling strategy in $R^2$ by 15.1% (Table 2) in terms of estimating mountain forest biomass. We further assessed the impact of elevational interval settings of in situ measurements on the results, considering intervals of 100 m ($N = 67$), 200 m ($N = 32$), 300 m ($N = 19$), and 500 m ($N = 12$). Here, one measurement was randomly selected at each interval, and the Monte Carlo method [53] was used to augment the sample size, ensuring a consistent sample size after sampling. The results showed that as the elevational interval increased, the accuracy of the results decreased rapidly, with $R^2$ dropping from 0.63 to 0.51 and RMSE increasing from 22.4 to 24.8 MgC ha$^{-1}$. These results highlighted the crucial role of field observation data spanning different elevational gradients, and emphasizes the need for comprehensive sampling strategies that capture the variability in forest characteristics along elevation gradients to improve the accuracy of biomass estimates in complex mountain ecosystems.

Table 2. Estimations using different field sampling strategies.

<table>
<thead>
<tr>
<th>Sampling Strategy</th>
<th>$R^2$</th>
<th>RMSE (MgC ha$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation gradient sampling (Total, $N = 67$)</td>
<td>0.63</td>
<td>22.4</td>
</tr>
<tr>
<td>Traditional random sampling ($N = 98$)</td>
<td>0.53</td>
<td>24.6</td>
</tr>
<tr>
<td>Elevation gradient sampling (interval = 200 m, $N = 32$)</td>
<td>0.59</td>
<td>23.9</td>
</tr>
<tr>
<td>Elevation gradient sampling (interval = 300 m, $N = 19$)</td>
<td>0.56</td>
<td>24.7</td>
</tr>
<tr>
<td>Elevation gradient sampling (interval = 500 m, $N = 12$)</td>
<td>0.51</td>
<td>24.8</td>
</tr>
</tbody>
</table>

4.2.2. Comparison of Different Sources of Remote Sensing Data

Multi-source remote sensing data offer a comprehensive view of vegetation structure, with Sentinel-2 and Landsat providing information on leaf characteristics; Sentinel-1 C band backscatter reflecting trunk and branch features, and L-VOD indicating vegetation water content. We first revealed a relationship between each remote sensing data and forest biomass at the field measurement scale. Scatter plots (Figure 4) indicated that the Sentinel-1 VV and VH bands (Figure 4a,b) outperform L-VOD (Figure 4c) and optical data from Sentinel-2 (Figure 4d–g) and Landsat (Figure 4h,i) in predicting field-measured forest biomass. However, Sentinel-1 data still saturated at high forest biomass above 150 MgC ha$^{-1}$, potentially underestimating high-biomass forests. Here, it should be emphasized that the weaker performance of L-VOD for biomass may be due to the loss of information and blurring caused by its lower resolution data. Furthermore, we conducted a comparative analysis using different sources of remote sensing data. Our analysis revealed that relying solely on single-source remote sensing data yielded limited accuracy in estimating forest biomass (Table 3). For instance, using Sentinel-2 and Landsat data alone to capture leaf characteristics resulted in an incomplete representation of forest structure and subsequently led to suboptimal biomass estimates ($R^2 = 0.65$, RMSE = 23.3 MgC ha$^{-1}$). It was noted that using optical data alone could significantly bias mountain forest biomass, as the underestimation could reach ~50% when compared with multi-source data results (107.0 ± 11.0 MgC ha$^{-1}$ vs. 162.8 ± 21.9 MgC ha$^{-1}$). Similarly, employing Sentinel-1 data, which primarily reflects trunk and branch characteristics, provided only a partial
insight into forest biomass dynamics ($R^2 = 0.57$, $RMSE = 28.7 \text{ MgC ha}^{-1}$). In contrast, when integrating multi-source remote sensing data, including Sentinel-2, Landsat, Sentinel-1, and L-VOD, our model achieved significantly improved performance in estimating forest biomass ($R^2 = 0.80$, $RMSE = 15.8 \text{ MgC ha}^{-1}$). By leveraging the complementary information from each data source, our model was able to capture a more comprehensive range of vegetation structure features, leading to enhanced accuracy in biomass estimation.

![Figure 4. Relationship between the forest biomass and the multi-source remote sensing data at field measurement scale. The subplots reveal data using multi-source remote sensing data, including (a) VV, (b) VH bands from Sentinel-1, (c) Vegetation Optical Depth (VOD), (d–g) reflectance data from Sentinel-2, and (h–l) Landsat 8. All remote sensing-based data have been normalized.](image)

**Table 3.** Comparison in forest biomass accuracy using different sources of remote sensing data.

<table>
<thead>
<tr>
<th>Features</th>
<th>$R^2$</th>
<th>RMSE (MgC ha$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentinel-2A + Landsat OLI</td>
<td>0.65</td>
<td>23.3</td>
</tr>
<tr>
<td>Sentinel-1</td>
<td>0.57</td>
<td>28.7</td>
</tr>
<tr>
<td>Sentinel-2A + Landsat OLI + Sentinel-1</td>
<td>0.78</td>
<td>18.1</td>
</tr>
<tr>
<td>Sentinel-2A + Landsat OLI + Sentinel-1 + VOD</td>
<td>0.80</td>
<td>15.8</td>
</tr>
</tbody>
</table>
4.2.3. Comparison with Pixel-Based Method

To reveal the effectiveness of incorporating spatial context into the ResNet model for enhancing forest biomass estimation, we conducted a comparative analysis with traditional pixel-based machine learning methods. Here, training and validation samples were kept the same. The ResNet model, equipped with convolutional layers capable of capturing spatial features within the forest canopy, outperformed traditional pixel-based methods such as RF ($R^2 = 0.59$, $RMSE = 29.3$ MgC ha$^{-1}$) and SVM ($R^2 = 0.56$, $RMSE = 30.4$ MgC ha$^{-1}$). This superiority was evident in the evaluation results presented in Table 4, where ResNet exhibited higher accuracy levels, particularly during validation at the plot scale. Additionally, XGBoost outperformed other machine learning methods with an $R^2$ value of approximately 0.66. However, its performance remains weaker than that of the ResNet model. This is primarily because XGBoost, despite algorithmic improvements [21,54], operates on individual pixels and lacks a convolutional structure, which prevents it from fully exploring spatial distributions like ResNet does. To ascertain the impact of the convolutional structure, that could consider the spatial context in ResNet, we constructed a pixel-based DNN model devoid of convolutional layers. This model relied solely on single-pixel information for forest biomass estimation and has other structures same to ResNet. Our analysis revealed a marked discrepancy in performance, with the DNN model trailing behind CNN notably ($R^2 = 0.68$ vs. 0.81, $RMSE = 24.1$ vs. 15.6 MgC ha$^{-1}$). This disparity highlighted the critical role played by the convolutional structure. Incorporating spatial context allowed a model like ResNet to “understand” and process such multi-dimensional data, enabling it to recognize and analyze complex patterns and relationships within the data that are essential for accurate forest biomass estimation, especially in challenging and varied landscapes like mountainous regions.

Table 4. Comparison with the pixel-based method for forest biomass estimation.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$R^2$</th>
<th>RMSE (MgC ha$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet</td>
<td>0.80</td>
<td>15.8</td>
</tr>
<tr>
<td>DNN</td>
<td>0.68</td>
<td>24.1</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.66</td>
<td>23.3</td>
</tr>
<tr>
<td>RF</td>
<td>0.59</td>
<td>29.3</td>
</tr>
<tr>
<td>SVM</td>
<td>0.56</td>
<td>30.4</td>
</tr>
</tbody>
</table>

4.2.4. Comparison with Other Advanced Deep Learning Methods

To reveal the effectiveness of including residual learning and deep neural network structures in the ResNet model for enhancing forest biomass estimation, we compared results with two other advanced convolution-based deep learning methods: AlexNet and VGG. AlexNet, with its simple three-layered architecture, offered a baseline with shallow depth, while VGG presented a more intricate setup with eight layers, yet without the residual learning integral to ResNet. Our comparative analysis indicated that the shallow depth model AlexNet performed poorly, achieving a model accuracy of only 87.5% (Table 5). Although increasing the depth of the model appeared to improve the performance of VGG, it still did not match that of ResNet. VGG’s accuracy lagged by 7.5%, potentially owing to the absence of residual connections, which are vital for the efficient training of deep networks. Residual connections streamline in ResNet introduced “shortcuts” that skip layers, simplifying gradient backpropagation and preventing diminishing gradients. This approach speeded up training while maintaining data integrity, enabling deeper network structures without added complexity. In this way, ResNet could reach high accuracy with complex deep networks in forest biomass mapping.
Table 5. Comparison in forest biomass accuracy using different CNN models.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$R^2$</th>
<th>RMSE (MgC ha$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet</td>
<td>0.80</td>
<td>15.8</td>
</tr>
<tr>
<td>VGG</td>
<td>0.74</td>
<td>18.2</td>
</tr>
<tr>
<td>AlexNet</td>
<td>0.70</td>
<td>21.1</td>
</tr>
</tbody>
</table>

4.3. Benchmark Mapping of Mountainous Forest Biomass over Tibet

After obtaining satisfactory validation results at the plot scale, we proceeded to generate a comprehensive wall-to-wall forest biomass map for the year 2020 (Figure 5). Our analysis revealed that regional forest biomass ranges from 95.2 to 275.3 MgC ha$^{-1}$, with an average of 162.8 ± 21.3 MgC ha$^{-1}$. This equals a total carbon storage of approximately 2.05 PgC, constituting 40.2% of the entire forest carbon storage 5.10 PgC on the Tibetan Plateau [55]. Notably, the mountainous forests harbor high carbon density, with nearly 90% of the forest biomass exceeding 135 MgC ha$^{-1}$. This is significantly higher than forests in eastern regions at similar latitudes in China (86.0 ± 10.6 MgC ha$^{-1}$) [56], as well as flat regions (slopes lower than 10 degrees) in China (61.3 ± 22.4 MgC ha$^{-1}$). Forest biomass reveals large spatial variation, regions with higher forest biomass are predominantly concentrated in Shannan City, situated in the southeastern part of Tibet, where the average biomass is approximately 1.24 times higher than that of the entire Tibetan region (202.4 ± 24.3 MgC ha$^{-1}$). Conversely, regions such as Nagqu and Chamdo in northeastern Tibet exhibit relatively lower forest biomass, attributed to the constraints of low temperature and limited water availability. Furthermore, forest biomass exhibits significant variation with the elevation. With increasing elevation, the forest biomass accumulation began to be inhibited.

Figure 5. Spatial pattern of mountain forest biomass over Tibet in 2020. (a) Distribution of forest biomass estimated by the newly proposed deep learning method, that integrates CNN with Sentinel-1 and Landsat data. The inset displays the frequency of forest biomass. (b) Mean elevational value of forest biomass over Tibet.

4.4. Forest Biomass Carbon Sink over Tibet

We then estimate changes in forest biomass during 2015–2020 (Figure 6). The mean forest biomass increased from 162.3 to 168.2 MgC ha$^{-1}$, resulting in regional carbon storage increasing from 0.85 to 1.07 PgC. This indicates an annual forest biomass carbon sink of
3.35 ± 0.08 TgC year⁻¹. Warming and wetting climate, along with CO₂ fertilization contributed to the forest biomass carbon sink [57,58]. The size of the carbon sink varied significantly across the elevation, with lower elevation forests (below 2500 m) generally experiencing a larger carbon sink. Local forest afforestation and natural forest protection projects likely contributed to this observed increase in carbon sink size [59]. High-elevation forests (above 4000 m) constitute only 32.9% of the total forest area in Tibet, yet their carbon sink capacity accounts for 41.8% (1.40 TgC year⁻¹) of the total carbon sink. Notably, forests near the treeline (elevation > 4500 m) also make a substantial contribution to carbon sequestration, with an annual carbon sink of 0.97 ± 0.03 TgC year⁻¹. The warming climate would facilitate trees upslope expansion [60], which also contributes to the regional carbon sink.

![Figure 6](image_url)  
**Figure 6.** Spatial pattern of the forest biomass carbon sink between 2015 and 2020. (a) Distribution of forest biomass carbon sink. The inset displays the frequency of forest biomass carbon sink. (b) Mean elevational value of forest biomass carbon sink over Tibet. The red region indicates a decrease in biomass and the green region is an increase.

We further compared our results with existing forest biomass carbon sink estimation (Table 6). Xu et al. [24] estimated global forest aboveground biomass and biomass carbon sink using a random forest method combined with ICESat/GLAS lidar data, MODIS, SRTM digital elevation model, and climate variables. We transformed aboveground biomass into total biomass, considering the root-to-shoot ratio. Liu et al. [8] provide changes in forest biomass based on VOD datasets. However, compared to our results, the previous estimation of mountainous forest biomass carbon sink (0.0068–1.19 TgC year⁻¹) [8,24] is less than half of our results. The underestimation of forest biomass carbon sink is primarily attributed to the underestimation of mountain forest carbon stocks (0.01–0.7 PgC) in Tibet due to the lack of forest structure information and valid in situ measurement data. Specifically, the mean forest biomass over mountainous Tibet reaches only 40.0 and 102.9 MgC ha⁻¹, which is 4 times less than our estimation. The validation results using our independent samples (N = 62) indicate that neither of these datasets effectively estimates the biomass of forests in the Tibetan region (Figure 7). This comparison shows models lacking representative field observation and spatial context can notably undervalue the biomass of mountain forests.
Table 6. Comparison in existing mountain forest biomass carbon sink estimation in Tibet.

<table>
<thead>
<tr>
<th>Method</th>
<th>Data</th>
<th>Biomass (MgC ha(^{-1}))</th>
<th>Carbon Stocks (PgC)</th>
<th>Carbon Sink (TgC year(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>This study</td>
<td>ResNet</td>
<td>162.6</td>
<td>5.10</td>
<td>3.35</td>
</tr>
<tr>
<td>Liu et al. [8]</td>
<td>Sentinel-2A</td>
<td>40.0</td>
<td>0.01</td>
<td>0.0068</td>
</tr>
<tr>
<td></td>
<td>Landsat OLI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sentinel-1 VOD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xu et al. [24]</td>
<td>Random forest</td>
<td>102.9</td>
<td>0.7</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>GLAS MODIS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QuikSCAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WorldClim SRTM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7. Validation of existing forest biomass maps in mountainous areas of Tibet based on measured samples. (a,b) are the estimated results of Liu et al. [8] and Xu et al. [24], respectively.

4.5. Limitation and Perspective

Here, we revealed significant elevational distribution of mountain forest biomass carbon stock and sink. However, the temporal mismatch between observation data (2015–2023) and our analysis period (2015–2020) could introduce uncertainties. Field observations were scarce at high elevations, especially near the treeline, and we lack repeated sampling to quantify biomass changes at the field scale. Future studies should include repeated sampling, particularly at high elevations. Furthermore, forest biomass varied greatly across sunny and shaded slopes, but we lack detailed comparative data on terrain aspect-related differences along the elevation gradient. Finer sampling could refine our understanding of microtopography’s impact on forest biomass and the carbon cycle in mountainous regions. Regarding remote sensing, we utilized optical, SAR, and VOD data for mapping forest biomass but omitted satellite LiDAR data from sources like ICESat/GLAS, ICESat-2 and GEDI. Integrating canopy height and direct vertical structure data from these sources might enhance our biomass estimates and provide better insights into the carbon cycle of mountain forests.

5. Conclusions

In summary, our study advances the benchmark mapping of forest biomass and biomass carbon sinks in mountainous regions. We introduce a deep learning-based method
that integrates microwave remote sensing (C band data from Sentinel-1, and L-VOD data) with optical images (Sentinel-2A and Landsat 8). The Regset model used in our study could automatically learn information from complex spatial features and distribution characteristics unique to mountainous forests, evidenced by comparison with the pixel-based method, advanced deep learning methods, and single-source remote sensing data. Our research also presents high-resolution (10 m) mapping of biomass and biomass carbon sink dynamics across mountainous forests during 2015–2020, thereby aiding in our understanding of how forests respond to climate change, especially in rugged mountainous regions. The magnitude of the forest carbon sink across high elevations of mountains can serve as a basis for formulating regional strategies to enhance carbon sequestration in response to climate change. However, it should be noted that our estimation still lacks observation data near the upper range limit of trees, especially near the treeline position due to the difficulty of access. The use of UAV-based or satellite-based LiDAR may be promising in those high-elevation forest carbon accounting.

Author Contributions: Conceptualization, X.W.; methodology, X.W. and G.L.; writing—original draft preparation, G.L., X.W. and J.X.; writing—review and editing, X.W., X.H., S.L., G.C. and H.H.; funding acquisition, X.W. and H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (42371118, 41977401) and the Science and Technology Plan Project of Tibet Autonomous Region (XZ202201ZD0005G01).

Data Availability Statement: All computer codes used in this study can be provided by the corresponding author upon reasonable request.

Acknowledgments: We thank the reviewers for their valuable comments and suggestions, and gratefully acknowledge the INRAE BORDEAUX remote sensing products for providing the SMOS-ICV2-RE06 L-VOD product (L-VOD) data.

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

References
13. Xie, X.; Li, A. An adjusted two-leaf light use efficiency model for improving GPP simulations over mountainous areas. J. Geophys. Res. Atmos. 2020, 125, e2019JD031702. [CrossRef]

15. Xie, X.; Li, A. Development of a topographic-corrected temperature and greeness model (TG) for improving GPP estimation over mountainous areas. Agric. For. Meteorol. 2020, 295, 108193. [CrossRef]


43. Liang, P.; Wang, X.; Sun, H.; Fan, Y.; Wu, Y.; Lin, X.; Chang, J. Forest type and height are important in shaping the altitudinal change of radial growth response to climate change. *Sci. Rep.* **2019**, *9*, 1336. [CrossRef] [PubMed]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.