

Response to Reviewer 2 Comments

Thanks for reviewing our paper “Estimation of Forest Stock Volume based on CNN-LSTM-Attention from Landsat8 and SRTM: A Case Study of Baishanzu Forest Park, China ”. The Authors really appreciate your valuable comments for improving the work.

We have carefully revised our paper according to the reviewers’ comments. This includes the following: 1) we improved English language through the entire manuscript; 2) we improved the figures as suggested by the reviewers; 3) we strengthened the conclusion to emphasize our contributions; 4) we addressed each of the reviewers’ suggestions and comments.

Below we give a point-to-point response to the reviews’ comments. The questions are displayed in red font, and the answers are displayed in black font. All of the revisions in the revised manuscript are marked in yellow and the all the page and line numbers are referring to the revised manuscript.

Comments 1: Authors only suggest about the passive data, not includes about the other format of data in line 43-48.

Response: Thank you for your valuable comment. In the Introduction section, remote sensing data used for FSV prediction is divided into passive remote sensing and active remote sensing. In line 45-67, we introduce the application methods and research status of passive remote sensing in the forestry field. And in line 68-87, we introduce the application methods and research status of active remote sensing in the forestry field. Both types of remote sensing data are supplemented with detailed explanations and presentations.

Comments 2: Line no 106, 109 the text R^2 should be replaced with ‘Coefficient of Determination (R^2)’ and other places in the MS.

Response: Thank you for your valuable comment. We checked the description of “ R^2 ” in the whole manuscript, and revised all the errors to “the Coefficient of Determination (R^2)”, with yellow texts.

Comments 3: Author should measure the Goodness of fit not only with Coefficient of Determination and MAE, they also used the Chi square, RMSE, AUC and other methods. The coefficient of determination doesnot provide the complete over view of the correlation.

Response: Thank you for your valuable comment. In addition to the coefficient of determination, we also used mean square error (MSE), mean absolute error (MAE) and root mean square error (RMSE) totally four evaluation indexes to compare the prediction accuracy of different models.

Among them, The value of the coefficient of determination (R^2) is between 0 and 1, which can reflect the correlation between the survey data of forest stock volume and the predicted value

of the model. The closer to 1, the higher the inversion accuracy. The average absolute error (MAE) is obtained by taking the mean of the absolute error, and its value is a natural number greater than or equal to 0, which can reflect the overall prediction error. The root mean square error (MSE) is opened to obtain the root mean square error (rMSE). The root mean square error is consistent with the dimension of the predicted value, which can reflect the degree of deviation between the predicted value and the true value. The smaller the value, the better the quality of the model and the higher the prediction accuracy. Through the above four evaluation indexes, we can compare and evaluate the prediction accuracy and robustness of different models more comprehensively, as highlighted in line 406-414.

Comments 4: The summary explained in the introduction section (lno 111-118), should be changed with the inference of literatures or hypothesis, the research on FSV used the mentioned datasets. Any summary and conclusions should be moved to the conclusion or discussion section of the MS.

Response: Thank you for your valuable comment. We have deleted the data and methods mentioned in summary and only retained the research direction and objectives of this paper as highlighted in line 112-114. The summary and results are moved to the Section 4 “Results and Analysis” section of manuscript.

Comments 5: Write about the modelling factors (lno:113-118), what all the factors is it some kind of black box or is it a language error?

Response: Thank you for your valuable comment. Based on Landsat 8 remote sensing data, SRTM global digital elevation data and survey data in the study area, we extracted six kinds of factors: spectrum, vegetation index, texture, principal component, topography and soil. Among them, we have 7 spectrum factors (Landsat B1~B7), 9 vegetation index factors (including RVI, NDVI, DVI, EVI, PVI, TVI, BVI, GVI and WVI), 56 texture factors (including contrast, correlation, dissimilarity, entropy, homogeneity, mean, second moment and variance from B1 to B7), 3 principal component factors, 2 topography factors (DEM and Slope), 4 soil factors(soil texture, soil layer thickness, humus layer thickness and soil category), totally 81 factors.

These factors are stored in a csv file in the form of image grayscale values or specific numeric codes to form the dataset. All factors are not black boxes, they are selected according to some references or designed based on experience.

Comments 6: Table 1 the Landsat8 band B10-TIR, spatial resolution should be check, is it of 30/100m. if it is for 100m, then explain about how the thermal band of the landsat8 image is used with other 30m bands for PCA and texture based product derivation, Pearson correlation in between the variables and how it is correlated for the goodness of fit?

Response: Thank you for your valuable comment. After double checking the homepage of Landsat, we found that the spatial resolution of Landsat 8 band B10-TIR was 100m. Therefore, when re-designing the input variables of the experiment, we abandoned Band10 and only

retained the band B1~B7 as spectrum factors.

In the experiment of this study, we extracted reflectance of each band of pixels as input. Therefore, when extracting reflectance of B10-TIR band, due to its low resolution, adjacent pixels would extract the same reflectance, and we ignored the impact of this problem on the experimental results.

Comments 7: In PCA factor (3.4.2), After three components have great noise??, who is it so, and kindly take look in the process of PCA, how the eigen value was calculated between the components. Author should correctly describe about why they are using PCA in place of lansat8 image.

Author suggest the MS to read more about the PCA

A machine learning-based classification of LANDSAT images to map land use and land cover of India. Remote Sensing Applications:Society and Environment.

Response: Thank you for your valuable comment. We consulted the relevant knowledge about principal component analysis in remote sensing field and read the articles recommended by reviewers. Principal Component Analysis is a statistical method that filters out important variables by reducing the dimension of multiple variables after linear transformation. The transformed variables are called principal components. For remote sensing images, the single-band image of each band corresponds to an input variable of PCA. For multi-spectral data, principal component analysis is very useful for extracting effective information. In our study, the principal component was screened by the characteristic contribution rate. The calculation formula is as follows:

$$R_i = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i}$$

And we redraw the principal component characteristic quantity graph of each variable after principal component analysis.

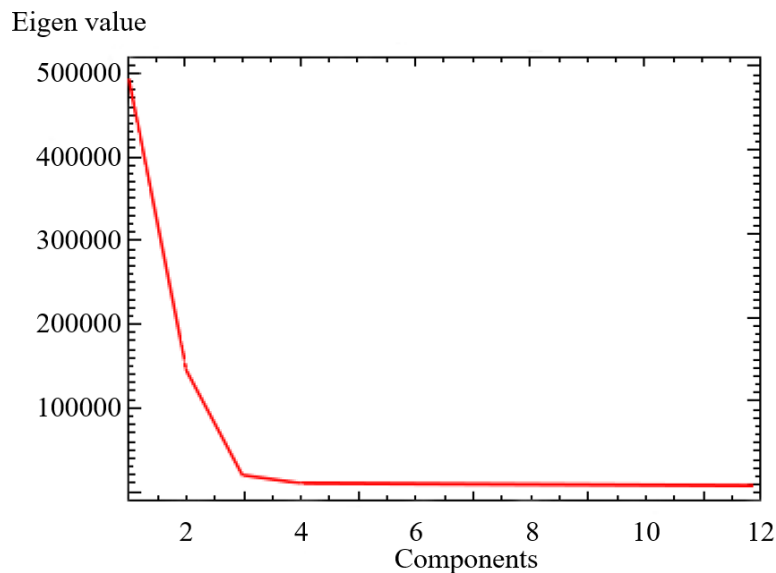


Figure 3. Eigenvalues of each principal component.

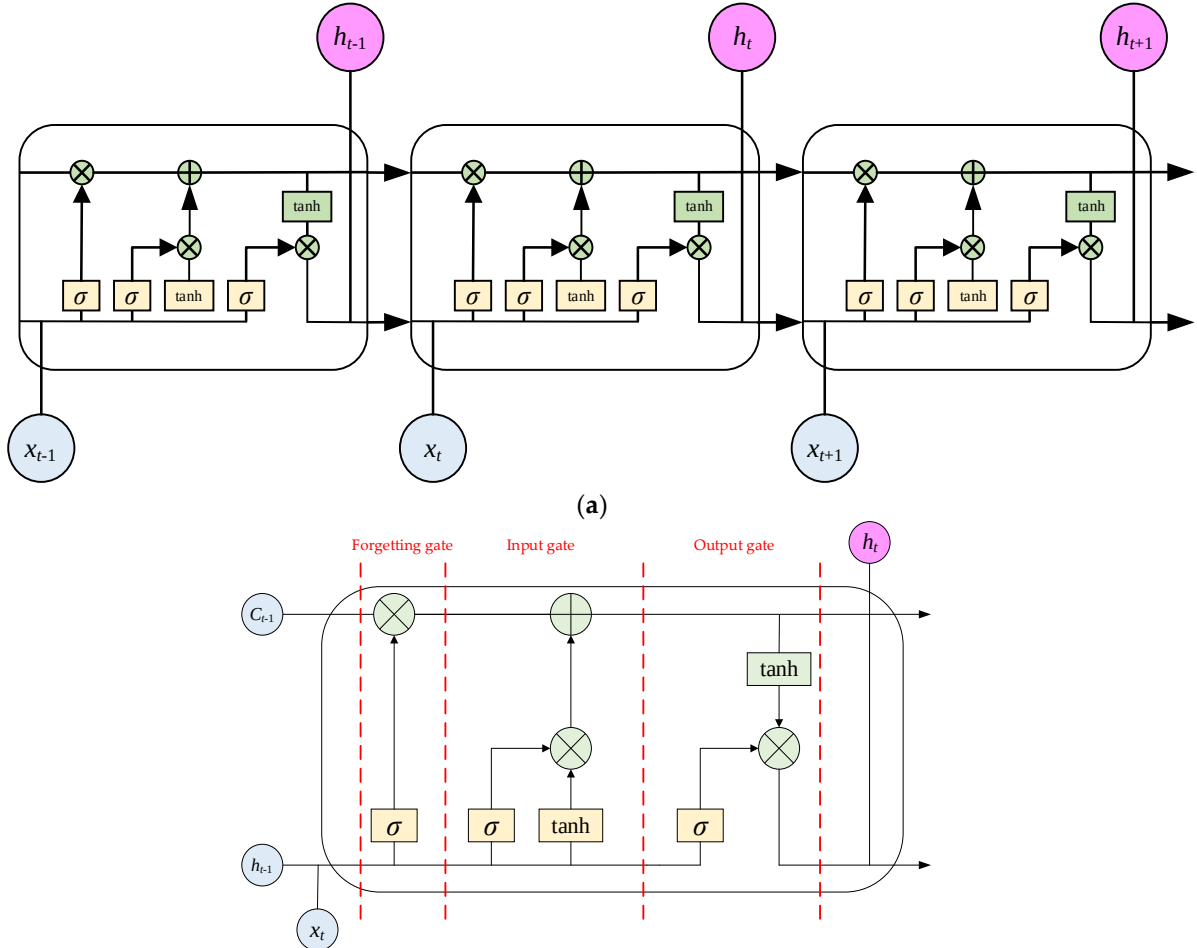
It can be seen from the figure that the image feature information is mainly distributed in the first, second and third components. Therefore, this study used the first, second and third principal component components to construct the forest stock volume estimation model of Baishanzu Forest Park.

Comments 8: PCA graph (Figure 3:) I hope it should be corrected as graph between ‘components’ (X axis) and ‘eigen value’ (Y axis). It caption should also be corrected accordingly.

Response: Thank you for your valuable comment. Please kindly refer to our response to your comment #7 (your previous comment).

Comments 9: Author should correct the flow diagram (Figure 6), direction of flow at every point and explain it with basic English what is input, what is intermediate(hidden) and what is predicted outcome. Clearly show the flow direction at every point function.

Response: Thank you for your valuable comment. In Figure 6, we use different colors to distinguish the input, output and neural network layer, and expand the internal structure of the network, and supplement the explanation of the network output, output and internal information flow in section 3.6. The process introduction of this model is supplemented in the manuscript, in line 308-329, with yellow texts.



(b)

Figure 6. Expansion diagram of LSTM structure: (a) Chain structure of LSTM; (b) The vector direction inside the LSTM's three gate structure.

LSTM has a chain structure, which stores the state of neurons through the gate structure. The chain structure is shown in the figure 6. Each yellow box represents a neural network layer, which is composed of weight, bias and activation function; Each green circle represents a pointwise operation; Arrows indicate the direction of the vector; Intersecting arrows represent concatenation of vectors; The bifurcated arrows represent the copy of the vector. LSTM has three input: cell state C_{t-1} (blue circle), hidden layer state of last moment h_{t-1} (purple circle), t time input vector x_t (blue circle), and the output has two: cell state at t time C_t , hidden layer state h_t . The information of the cell state C_{t-1} is always transmitted on the line above. The hidden layer state h_t at time t and the input x_t will modify C_t appropriately and then pass it to the next moment. C_{t-1} will participate in the calculation of the output h_t at time t . The information of the hidden layer state h_{t-1} modifies the cell state through the gate structure of LSTM and participates in the calculation of the output. In general, the information of the cell state has been transmitted on the upper line, and the hidden layer state has been transmitted on the lower line. They interact with each other through the gate structure. In the three gate structures, the results of 0~1 are calculated by the activation function σ to affect the proportion of information access and abandonment of the previous neuron.

The gate structure increases the number of network iterations, and the error of the activation function can still be transmitted in reverse to avoid long-term dependence. At the same time, the output of the upper layer neurons is accepted by the three gate structures of forgetting gate, input gate and output gate, and the effective information of the historical moment is selectively retained.

Comments 10: Author should try to use the PCA in Data 5/6 data set and tried to predict the model outcome. It reduces the noise and data dimensionality between all kind of texture, vegetation index products, srtm data.

Response: Thank you for your valuable comment. Based on your valuable suggestions, we found that the original experimental dataset design was unreasonable. Therefore, we reextract the factors. All the corrected factors are as follows: Based on Landsat 8 remote sensing data, SRTM global digital elevation data and survey data in the study area, we extracted six kinds of factors: spectrum, vegetation index, texture, principal component, topography and soil. Among them, we have 7 spectrum factors (Landsat B1~B7), 9 vegetation index factors (including RVI, NDVI, DVI, EVI, EVI, PVI, TVI, BVI, GVI and WVI), 56 texture factors (including contrast, correlation, dissimilarity, entropy, homogeneity, mean, second moment and variance from B1 to B7), 3 principal component factors, 2 topography factors (DEM and Slope), 4 soil factors(soil texture, soil layer thickness, humus layer thickness and soil category), totally 81 factors.

After correlation analysis, among all 81 characteristic variables, 69 characteristic variables were significantly correlated with the stock volume when the confidence level (bilateral) was 0.01, and 5 feature variables were significantly correlated with FSV when the confidence level (bilateral) was 0.05. They are the B1 band of spectral features, the B5 difference feature of texture features, the homogeneity feature of B6 and B7 bands, and the third component of principal

component features. Seven feature variables were not significantly correlated, namely, B6 and B7 bands of spectral features, the correlation features of B1, B5, B6 and B7 bands of texture features, and the difference features of B2 bands.

In order to analyze the contribution of different feature variable types to stock estimation among the 69 retained feature variables, 6 different species feature variables were combined into 7 different data sets, The new dataset design is shown in the table below:

Variable Sets	Types of characteristic variables
DataSet1	Vegetation index, Texture, PCA, Topography, Soil
DataSet2	Spectrum, Texture, PCA, Topography, Soil
DataSet3	Spectrum, Vegetation index, PCA, Topography, Soil
DataSet4	Spectrum, Vegetation index, Texture, Topography, Soil
DataSet5	Spectrum, Vegetation index, Texture, PCA, Soil
DataSet6	Spectrum, Vegetation index, Texture, PCA, topography
DataSet7	Spectrum, Vegetation index, Texture, PCA, Topography, Soil

As shown in the table, the 7 dataset are: (1) Datasets including vegetation index feature, texture feature, principal component feature, topographic feature and soil feature as independent variables; (2) Datasets including spectrum features, texture features, principal component features, topographic features and soil features as independent variables; (3) Datasets including spectrum features, vegetation index features, principal component features, topographic features and soil features as independent variables; (4) Datasets including spectrum features, vegetation index features, texture features, topographic features and soil features as independent variables; (5) Datasets including spectrum features, vegetation index features, texture features, principal component features and soil features as independent variables; (6) Datasets including spectrum features, vegetation index features, texture features, principal component features and topographic features as independent variables; (7) Datasets that retain all feature variables as independent variables.

Through the above grouping method, we compared the prediction accuracy of the model without various features, and then evaluated the contribution degree of various variables to FSV prediction.

Comments 11: In Dataset 3/4/5/6, why image band and its PCA together used for the analysis. PCA or image band should be used analysis. It also affected the correction between the products and goodness of fitness model.

Response: Thank you for your valuable comment. Please kindly refer to our response to your comment #10 (your previous comment).

Comments 12: Figure 10 representing absolute error between real and predicted values from different models, kindly check the point value representation for all the blue color points (difference value between 50 and 100 unit value)

Response: Thank you for your valuable comment. In the experimental results, we have checked all the sample prediction results. In the prediction results of each model, the maximum absolute

error between real and predicted values of the sample is within 100, so we set two intervals of 50 m³/ha and 100 m³/ha when drawing. So the points in the graph contain all the test samples.

Comments 13: The result section should add the discussion about the data, methods selection and Spatial map which are representing the predicted FSV outcome from the CNN-LSTM model.

Response: Thank you for your valuable comment. In the result section, we added section 4.2 “Evaluation of contribution degree of characteristic variables” to compare the FSV prediction accuracy of the seven experimentally modified data sets on the CNN-LSTM-Attention model, The evaluation indicators are shown in the table below:

Variable Sets	R ²	MAE	MSE	RMSE
Dataset1	0.6481	25.2271	1625.2519	40.3144
Dataset2	0.6208	25.5053	1751.4848	41.8507
Dataset3	0.7311	22.0368	1241.9801	35.2417
Dataset4	0.8299	17.4181	785.5955	28.0284
Dataset5	0.7633	21.5353	1093.2510	33.0643
Dataset6	0.8519	16.7629	683.8263	26.1501
Dataset7	0.8544	15.0753	672.3985	25.9306

According to the test results of different kinds of feature factor combinations shown in figure 10, it can be seen that the highest coefficient of determination(R²) of the dataset7 with the most types of characteristic variables is 0.8544 ; the second is the dataset6 that lacks the information of soil characteristics. The coefficient of determination is 0.8519, and the root mean square error between the two data sets is only within 0.3 m³/ha, which may be due to the fact that the soil in the study area is generally not very different, which weakens the influence of soil environment on vegetation growth. Among all kinds of characteristic variables, the original spectral characteristics of remote sensing images and the vegetation index characteristics that can directly reflect the vegetation growth status have the greatest impact on the estimation of volume. In the data set lacking spectral characteristics and vegetation index characteristics, the root mean square error of the model is larger than that of other data sets, reaching 40.3144 m³/ha and 41.8507 m³/ha, respectively. The coefficient of determination(R²) decreased to 0.6481 and 0.6208. It can be seen that spectral data is the characteristic factor that can directly reflect the growth status of ground vegetation, and has a high contribution in the characteristic variables of volume estimation.

Besides, we added the FSV prediction results of the three models used in the comparison experiment in the whole study area, and compared the estimation accuracy of the four models, as shown in the following table:

Models	Minimum FSV(m ³ /ha)	Maximum FSV(m ³ /ha)	Total FSV(m ³)	Accurate(%)
MLR	9.76	396.85	6331850.47	66.74
RF	4.13	359.17	3017532.16	71.40
BP	1.27	724.68	5798483.77	72.88
CNN-LSTM-Attention	4.83	402.38	5066562.23	83.41
Survey data	6.95	414.6	4226019.56	1

We compared the minimum and maximum FSV per unit area predicted by each model, and predicted the FSV of the whole study area based on the trained model, and calculated the prediction accuracy with the FSV of the survey data. In addition, we briefly analyzed the spatial distribution of FSV with altitude change in the study area.

Response to Comments on the Quality of English Language:

Point 1: Check English in line 33 to 36.

Response: Thank you for your valuable comment. We checked the English of the entire manuscript to make it easier to read and understand.

Point 2: Heading 3.6 what is 'FSV Rrediction Model', kindly correct it.

Response: Thank you for your valuable comment. We have corrected this spelling error to 'FSV Prediction Model' in the manuscript.

Point 3: Figure 8 the label fonts should be readable. Other wise use it notation in the graph label. It caption should also be corrected.

Response: Thank you for your valuable comment. Considering that there are a large number of feature variables involved in correlation calculation in this study, the correlation heat map will affect the reading experience of the manuscript, and it is impossible to extract effective information from the graph, so we delete the correlation graph.

Point 4: Font size of all the diagram label should be corrected.

Response: Thank you for your valuable comment. Now, the meaning of these graphs and the axes is clearer.