Dynamic Estimation of Forest Volume Based on Multi-Source Data and Neural Network Model

Dasheng Wu¹,² & Yongquan Ji¹,²

¹ Key Laboratory of Zhejiang Province about Forestry Intelligent Monitoring and Information Technology Research, Zhejiang A & F University, Lin’an, Zhejiang, China
² School of Information Engineering, Zhejiang A & F University, Lin’an, Zhejiang, China

Correspondence: Dasheng Wu, Key Laboratory of Zhejiang Province about Forestry Intelligent Monitoring and Information Technology Research, Zhejiang A & F University, Lin’an 311300, Zhejiang, China. Tel: 1-396-802-3105. E-mail: 458752249@qq.com

Received: November 17, 2014   Accepted: December 25, 2014   Online Published: February 15, 2015
doi:10.5539/jas.v7n3p18          URL: http://dx.doi.org/10.5539/jas.v7n3p18

Abstract

It is quite necessary to explore some more efficient and reliable estimation models which could integrate or, in some cases, substitute the traditional and expensive measuring techniques in forest resources management owing to the rising investigation costs. Thanks to their flexibility and adaptability, artificial neural networks (ANN) constitute a valid approach for modelling complex long-lived dynamic forest ecosystems.

The evaluation indexes set was established, including 17 factors: elevation, slope, aspect, surface curvature, solar radiation index, topographic humidity index, tree ages, the soil depth, the A-layer depth of soil, canopy density, Normalized Difference Vegetation Index (NDVI), and the spectral characteristics of the bands from Enhanced Thematic Mapper (ETM+) or Thematic Mapper (TM), Band 1 to Band 5, and Band 7 from Landsat. Then, integrating the remote sensing images of ETM+ or TM, Digital Elevation Model (DEM), and forest resource planning investigation data of fir of the key forestry city of Longquan, Zhejiang Province, China, the membership of each factor was empirically fitted by polynomials, and the forest volumes were estimated via an improved back propagation (BP) neural network (NN) model with Levenberg-Marquardt (LM) optimization algorithm (LM-BP). The results showed that the average individual relative errors (IARE) were from 26.38% to 34.41%; the group relative errors (GRE) were from 2.04% to 6.69%, and all of the group estimation precisions were more than 90% which is the highest standard of overall sampling accuracy about volume of forest resource inventory in China.

Keywords: LM-BP, forest volume, estimation

1. Introduction

Forest inventories provide objective and scientifically reliable information on key forest ecosystem processes, and constitute an effective tool for forest management and forest resource monitoring. Forest inventory data define the extent, size distribution, and species composition of forested and non-forested lands and through periodical updating, they track the changes that occur in natural resources over time (Gianfranco et al., 2007).

In China, the traditional large-scale survey of forest resources include forest inventory and forest resource planning investigation, where, forest inventory is repeated per 5 years, and the forest resource planning investigation is conducted per 10 years. However, the traditional long cycle of forest resources survey has been unable to meet the actual demand both in the ecological monitoring and the perspective of the use of forest products (Liu, 2006).

Forest managers are continuously searching for more efficient and reliable estimation models that could integrate or, in some cases, substitute the traditional and expensive measuring techniques. Many simulation models have been built to be used to predict forest growth and yield response to treatments (Robinson et al., 2003). Traditional statistical methods are not always suited to solve unstructured problems occurring in natural resource assessment (Gimblett et al., 1995) mainly because statistical methods are based on some assumptions on the data distribution. Moreover, they have shown to have several limitations when variables that are involved interact in a complex manner and have difficulties in handling poor and noisy data. Such conditions are very frequent in
forest data where classes may display a range of distributions, relationships between variables may be non-linear, and outliers and noise may exist in the data (Liu et al., 2003; Gianfranco et al., 2007).

Owing to their adaptability and flexibility, artificial neural networks (ANNs) constitute an alternative and valid approach for modelling non-linear and complex long-lived dynamic biological ecosystems such as forests. ANNs have become very popular because they can learn complex patterns and trends in the data, they are slightly affected by data quality problems and bias, and they are robust to data structures with highly interrelated relationships (Gianfranco et al., 2007). During the last 2 decades, ANNs have received a great deal of attention as a valid alternative to traditional statistical methods to predict the behaviours of non-linear systems (Gianfranco et al., 2007), and have been showing potential for solving some difficult problems in forest resources management (Shataee, 2011; Wang et al., 2011; Castaño-Santamaría et al., 2013).

Multi-source data aggregation has been used to solve complementarity and collaboration about different information on forest resources which provides a possibility to further improve the prediction accuracy (Shataee, 2011; Ying et al., 2011; Mäkelä et al., 2011; Han et al., 2013).

To facilitate the use of the implemented previsional model, dominant tree species of fir were chosen as the research object, and the forest volumes of the key forestry city, Longquan in Zhejiang province of China, were predicted dynamically. Firstly, the evaluated factor set with lower cost was established, including topography, climate, soil, forest structure, and spectral characteristics of forest etc. Secondly, research data which include the satellite images, Digital Elevation Model (DEM), forest resource planning investigation data, permanent sample plot survey data and other data sources, were integrated. Finally, the membership of each variable was empirically fitted by polynomials, and the forest volume was estimated via an improved back propagation neural network (BPNN) with Levenberg-Marquardt (LM) algorithm (BP-LM).

2. Study Area

The key forestry city of Longquan, 3,059 km² in extent, is a largely mountainous area located in the southwestern part of Zhejiang province in China, where the longitude is from 118°42′E to 119°25′E, and latitude is between 27°42′N to 28°20′N. The administrative map of Longquan city was showed as figure 1.

There are abundant forest resources with 3,985,000 mu (1/15 hectare) of areas, forest volume reached 14.56 million cubic meters, and the forest coverage rate up to 84.2% (Hong, 2012).

3. Materials and Methods

3.1 Evaluation Indexes Set

Forest development must be carried out under certain site conditions which commonly evaluated by environmental factors, forestry vegetation factors and human activity factors (Shen, 2001).

Typically, in the natural state, the development of forest resources affected by the environmental factors which mainly include three classes:

- Climate, mainly includes solar radiation and precipitation.
- Topography, directly related to water potential and soil conditions, including elevation, aspect, slope, slope position, slope-type, and small terrain, etc.
- Soil, including soil type, soil depth, soil texture, soil structure, soil nutrients, soil humus, soil PH, soil erosion degrees, all levels of gravel reserves in the soil, soil salinity, soil-forming rock and parent material type, etc.

But, they are not always suited to estimate the forest growth increment because the variety of environmental factors greatly increased the costs for data acquisition and the complexity of research (Zhao, 2007), which led to many experts and scholars try to select part of them to involve in their experiment and have showed some good results (Hong et al., 1998; Deng & Li, 2002; Xie, 2004; Xu, 2011).

Thanks to their large-scale, information-rich and relatively low-cost advantages, remote sensing images have been constituted a valid addition for improving prediction accuracy to forest volume or forest biomass (Guo et al., 2002; Wang & Xing, 2008; Xu, 2008).

Responding to the specific precision of 90% which is the highest standard of overall sampling accuracy about volume of forest resource inventory in China (Liu, 2005), one of the aims of this study was to establish an evaluation indexes set which was, first and foremost, the best to meet the low-cost requirements to monitor forest volume, and included the environmental factors and remote sensing image factors as much as possible.

Thus, a comprehensive evaluated index set including 17 factors: elevation, slope, aspect, surface curvature, solar radiation index, topographic humidity index, tree ages, the soil depth, the A-layer depth of soil, canopy density,
Normalized Difference Vegetation Index (NDVI), and the spectral characteristics of the bands from Landsat Thematic Mapper-TM (Band 1 to Band 5, and Band 7), was established.

3.2 Data Sources

The development of ANN models requires sufficient research data included the satellite images, Digital Elevation Model (DEM), forest resource planning investigation data, and permanent sample plot survey data. The research data should be divided into modeling sample set and estimating sample set. ANN models can learn the input/output relationship from the modeling samples and estimate the forest volume from estimating samples.

3.2.1 Research Data

(1) Administrative map of Longquan city.

(2) DEM with 30 meter resolution, data type of IMG and projection of UTM/WGS84, provided by Geospatial Data Cloud Platform, Computer Network Information Center, CAS (http://www.gscloud.cn), is generated by the first version data of the instrument Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)-Global Digital Elevation Model (GDEM) in 2009.

(3) The forest resource planning investigation data in 2007, consisting of 83,078 subplots which composed 39,377 forest subcompartments. In order to eliminate the erroneous and incorrect data, those samples located in non-forested land and non-volume land were removed by preliminary processing. Thus, the forest resource planning investigation data were remained with 28,707 subcompartments and 40,249 subplots, and there were 17,250 subcompartments and 20,296 subplots where the dominant tree species are fir.

(4) 207 valid investigation data from permanent sample plots in 2004, and 76 valid investigation data from permanent sample plots in 2010, in which, the dominant tree species are all fir.

(5) Satellite images include the Enhanced Thematic Mapper (ETM+) from Landsat on March 26, 2003,
Thematic Mapper (TM) from Landsat on January 8, 2007, and the Thematic Mapper (TM) from Landsat on October 31, 2010.

3.2.2 Modeling Sample Set and Estimating Samples Set

The 20296 subplots in 2004, whose dominant tree species are fir, were independently divided into 2 sets: a modeling sample set with 18,000 samples and an estimating samples set with other 2,296 samples. In addition, 207 valid data from permanent sample plots in 2004, and 76 valid data from permanent sample plots in 2010, have been as estimating samples set.

3.3 Improved BP Neural Network Mode Based on LM Algorithm

Despite the large number of possible network models available, multilayer feed-forward neural networks trained by back propagation algorithm (BPNN) represent the most prominent and well researched class of ANNs in classification and pattern recognition (Lek & Guégan, 1999). Usually, a back propagation system comprises three types of successive layers: input layer, hidden layer and output layer. The three layers consist of simple computational units called nodes. During training, the input signal propagates through the network in a forward direction, from left to right on a layer-by-layer basis, generating a set of values on the output units and fixing all networks synaptic weights. Then, difference between the actual and desired output values is measured, and the network model connection strengths are changed so that the outputs produced by the network become closer to the desired outputs. This is achieved by a backward pass during which connection changes are propagated back through the network starting with the connections to the output layer and ending with those to the input layer (Gianfranco et al., 2007).

However, in traditional BPNNs, there are some shortcomings, such as slow convergence speed and easy to fall into local minimum. Fortunately, LM algorithm which is actually a combination of gradient descent algorithm and Newton algorithm, compare to the traditional BPNNs, significantly reduce the number of iterations, accelerate the convergence speed, and get a higher accuracy. Especially, whose convergence speed is the fastest of all traditional and other improved BPNNs for medium-sized networks. In recent years, the improved BPNNs by LM algorithm have been widely used in the fields of evaluation and forecasting and showed a lot of good effects (Hua et al., 2008; Zheng & Jiang, 2010; Miao et al., 2011; Jian et al., 2012; Wang, 2013).

In order to obtain a better result for the experiment, we chose improved BP neural network model based on LM algorithm to estimate the volume of forest resources.

3.4 Data Preprocessing

3.4.1 Data Integration

In 2007, the average volume per unit (m³/mu) of forest resources was the only estimated factor, whose data were stored in the database of forest resource planning investigation. Moreover, the data of soil depth, the A-layer depth of soil, tree ages, and the canopy density were stored in the same database also. But in 2004 and 2007, all data came from permanent samples and were stored in EXCEL files.

In addition, the data about elevation, slope, aspect, surface curvature, solar radiation index, topographic humidity index were derived from DEM. Correspondingly, the NDVI, and the spectral characteristics of the bands (Band 1 to Band 5, and Band 7) were from satellite images of TM.

To facilitate the data storage and analysis, all the data have been integrated into the same database of forest resource planning investigation.

3.4.2 Membership about Evaluation Indexes

Generally, membership was calculated by following steps:

Step 1: To group each evaluation index data according to the experience;

Step 2: To statistics their average volume per unit of forest resources according to each group evaluation index, and to obtain their polynomial fitting curves and fitting equations;

Step 3: To get the fitted values of each evaluation index according the fitting equations, and to get their membership by normalization through equation as shown in Equation 1.

\[ zi = \left| \frac{yi}{max(yi)} \right| \]  \hspace{1cm} (1)

Where, \( yi \) was the fitted value of each index of every monitoring unit, \( max(yi) \) was the maximum of all \( yi \), and \( zi \) was the membership of each index.

Exceptionally, in this paper, the indexes of aspect, the soil depth, and the A-layer depth of soil, their membership
had special rules. Specifically, we come to solve the membership for each evaluation index:

(1) Aspect: firstly, according to their degree range, to divide aspect into 9 classes: flat, north, northeast, east, southeast, south, southwest, west, northwest, north; secondly, to statistics their average volume per unit of forest resources grouping by the 9 classes; finally, to get the membership of aspect according to Equation 1. The classification and membership about aspect showed as Table 1.

Table 1. Classification and membership of aspect

<table>
<thead>
<tr>
<th>Aspect classification</th>
<th>Degrees range (º)</th>
<th>Actual volume of forest (m³/mu)</th>
<th>Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td>&lt;=0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>North</td>
<td>(&gt;0 and &lt;=22.5) or (&gt;337.5 and &lt;=360)</td>
<td>5.62143108917871</td>
<td>0.940494698</td>
</tr>
<tr>
<td>Northeast</td>
<td>&gt;22.5 and &lt;=67.5</td>
<td>5.97710024243702</td>
<td>1</td>
</tr>
<tr>
<td>East</td>
<td>&gt;67.5 and &lt;=112.5</td>
<td>5.52788764320642</td>
<td>0.924844393</td>
</tr>
<tr>
<td>Southeast</td>
<td>&gt;112.5 and &lt;=157.5</td>
<td>5.6463109082021</td>
<td>0.944657222</td>
</tr>
<tr>
<td>South</td>
<td>&gt;157.5 and &lt;=202.5</td>
<td>5.6286008322792</td>
<td>0.941694235</td>
</tr>
<tr>
<td>Southwest</td>
<td>&gt;202.5 and &lt;=247.5</td>
<td>5.54310961835899</td>
<td>0.927391108</td>
</tr>
<tr>
<td>West</td>
<td>&gt;247.5 and &lt;=292.5</td>
<td>5.62936390352901</td>
<td>0.941821899</td>
</tr>
<tr>
<td>Northwest</td>
<td>&gt;292.5 and &lt;=337.5</td>
<td>5.69270408541559</td>
<td>0.952419042</td>
</tr>
</tbody>
</table>

(2) Soil depth: a positive correlation between soil depth and plant height has been presented (Li, 2012). Similarly, in this paper, the experimental data also reflected a generally positive linear correlation between the soil depth and the volume of forest resources. So, the membership of soil depth was calculated by Equation 1 directly.

(3) The A-layer depth of soil: according to the data of forest resource planning investigation, the A-layer depth of soil qualitatively recorded as thick, medium, thin or null (State Forestry Administration of China, 2003). Accordance with experts’ experience, the membership values of the A-layer depth of soil were quantified as: thick to 1; medium to 0.7; thin to 0.4 and null to 0.

![Figure 2. Polynomial fitting curves of elevation](image)

![Figure 3. Polynomial fitting curves of slope](image)
Figure 4. Polynomial fitting curves of surface curvature

Figure 5. Polynomial fitting curves of solar radiation index

Figure 6. Polynomial fitting curves of topographic humidity index

Figure 7. Polynomial fitting curves of tree ages

Figure 8. Polynomial fitting curves of canopy density

Figure 9. Polynomial fitting curves of NDVI
(4) Other indexes: the membership of other evaluation indexes was calculated by the step 1 to step 3. Where, their polynomial fitting curves were showed as Figure 2 to Figure 15, and their polynomial fitting equations were showed as Table 2.
### 3.5 Dynamic Estimation for Forest Volume Based on BP-LM Neural Network Model

Step 1: Setting Model Parameters

The L-M BP neural network was comprised three successive layers: input layer, hidden layer and output layer. In which, the nodes of hidden layer were calculated by Equation 2.

\[
\text{Hidden}\_\text{Num} = 2 \times \text{Input}\_\text{Num} + \text{Output}\_\text{Num} \quad (2)
\]

Where, *Hidden\_Num* was the number of nodes about the hidden layer, *Input\_Num* was the number of nodes of the input layer, and *Output\_Num* was the number of nodes of the output layer.

Specifically, those model parameters were set as follows:

- **Epochs = 1000;** %the maximum of epochs.
- **Input\_Num = 17;** %the number of nodes of input layer.
- **Output\_Num = 1;** %the number of nodes of output layer.
- **Hidden\_Num = 2*Input\_Num+Output\_Num;** %the number of nodes of the hidden layer

TransferFcn = '\text{tansig}' '\text{purelin}'; %\text{tansig} was the transfer function transferring values from the input layer to the hidden layer, and \text{purelin} was the transfer function transferring values from the hidden layer to the output layer.

TrainFcn = 'trainlm'; %training function was trainlm corresponding to LM algorithm.

LearnFcn = 'learngdm'; %learning function was learngdm.

PerformFcn = 'mse'; %performing function was mse (mean square error).

Step 2: Creating Net

\[
\text{Net} = \text{newff} (P, T, \text{Hidden}\_\text{Num});
\]

Where, *P* was the input vector, *T* was the output vector, and *Hidden\_Num* is the number of nodes of the hidden layer.

Step 3: Training Net

\[
[\text{Net TR}] = \text{train} (\text{Net}, P, T); \quad \%\text{net training}
\]

Step 4: Simulation

\[
Y = \text{sim} (\text{Net}, P\_\text{test});
\]
Where, Net was the trained net, P_test was the input vector of simulating samples, and y was the estimation result.

3.6 Error Representation

IARE was the individual average relative error which calculated by Equation 3, and GRE was the group relative error which calculated by Equation 4.

\[
IARE = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{t_i - y_i}{t_i} \right)
\]

\[
GRE = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\sum_{i=1}^{n} (t_i - y_i)}{\sum_{i=1}^{n} t_i} \right)
\]

Where, n was the number of simulating samples, ti was the observed value of the i-th sample, and yi was the calculated value of the i-th sample.

4. Results and Discussion

Estimation results of forest volume based on improved BPNN with LM algorithm were showed as Table 3.

<table>
<thead>
<tr>
<th>The year</th>
<th>Total samples</th>
<th>Modelling samples</th>
<th>Estimating samples</th>
<th>Observed value (m^3/mu)</th>
<th>Calculated value (m^3/mu)</th>
<th>IARE (%)</th>
<th>GRE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>20,296</td>
<td>18,000</td>
<td>2,296</td>
<td>6.7717</td>
<td>6.3189</td>
<td>26.57</td>
<td>6.69</td>
</tr>
<tr>
<td>2004</td>
<td>207</td>
<td>-</td>
<td>207</td>
<td>4.1524</td>
<td>4.1608</td>
<td>34.41</td>
<td>2.04</td>
</tr>
<tr>
<td>2010</td>
<td>76</td>
<td>-</td>
<td>76</td>
<td>6.1652</td>
<td>5.8193</td>
<td>26.38</td>
<td>5.61</td>
</tr>
</tbody>
</table>

As shown in Table 3, GRE were from 2.04% to 6.69%, but IARE were from 26.38% to 34.41%. The calculated and observed values scatter plots (Figures 16, 17, and 18) showed that the model could well fit the data when the observed values were approximately between 2 and 8, but when observed values were less than 2, it is a clear tendency that observed values were overestimated, and when observed values were over than 8, the values were obviously underestimated. Similarly, residuals display more homogeneous distributions when the observed values were between 2 and 8, but a clear tendency of heteroscedasticity to overestimate the observed values which were less than 2, and to underestimate the observed values which were over than 8 (Figures 19, 20, and 21).

Analytic trees’ data (Kan, 2010; Duan, 2010), temporary standard data or remote sensing data were often chosen as modeling data in most of the current studies about forest growth (Zeng, 2010; Wang, 2008). The narrower distribution range of analytic trees leads to a lack of representation about the experimental results and limits the popularization and application of their model. Temporary standard method to estimate forest growth is only suitable for the cases with short growth period and assume that the number of trees is constant, but in practice, medium and long term forecast is required, and forest stand density exists natural sparse phenomenon also (Che, 2012). Thanking to the integrating data of the remote sensing images of ETM+ or TM, Digital Elevation Model (DEM), forest resource planning investigation data, and permanent sample plot survey data of fir, the research data reflects the features with a wide distribution range and lower costs of data acquisition in this paper. When the improved BPNN model with LM algorithm will be used to forecast the forest volume in the the same research area, canopy density is the only index, which need to be re-measured, can be measured by a simple visual observation or a diagonal measurement about the sample plots, so the measuring costs are far less than those conventional estimation models which require to measure the dbh or tree height for each tree (Wei, 2012; Wu, 2004; Lin, 2000). And the model has solved the problems which are difficult to adapt to the uncertainty, time-varying environment in forest growth forecast with the traditional methods based on an accurate model (Zhu, 2010), and is more suitable for predicting the forest growth (Huang, 2006; Liu, 2007; Diamantopoulou, 2010; Shen, 2009; Chen, 2006, 2009; Zhao, 2003; Huang, 2005).
Figure 16. The calculated and observed values scatter plot for 2007 simulating samples

Figure 17. The calculated and observed values scatter plot for 2004 samples

Figure 18. The calculated and observed values scatter plot for 2010 samples
5. Conclusions
In this study, the forest volumes of fir in Longquan, of Zhejiang Province (China), were estimated dynamically. First, the evaluation indexes set was established, which includes 17 factors: elevation, slope, aspect, surface curvature, solar radiation index, topographic humidity index, tree ages, soil depth, A-layer depth of soil, canopy
density, Normalized Difference Vegetation Index (NDVI), and the spectral characteristics of the bands from ETM+ or TM (Band 1 to Band 5, and Band 7). Then, the membership of each evaluation index was empirically fitted by polynomials, and the forest volume was estimated via an improved BPNN model with LM optimization algorithm. The results showed that the average individual relative errors (IARE) were from 26.38% to 34.41%; the group relative errors (GRE) were from 2.04% to 6.69%, this indicated that group estimation precisions were all more than 90% which is the highest standard of overall sampling accuracy about volume of forest resource inventory in China.

Since the model showed that when observed values were less than 2, it is a clear tendency that observed values were overestimated, and when observed values were over than 8, the values were obviously underestimated, a further improvement of the model predicting capability could be achieved by including, removing the abnormal modeling sample data by the physical or statistical identification methods, optimizing the BPNN model with genetic algorithm. So the result of this paper could be further used to forecast the growth of dbh, tree height, and biomass of fir, even it could be used to predict the growth of other trees.

Acknowledgements
This work was financially supported by Zhejiang provincial major science and technology project (2011C12047), Zhejiang provincial natural science foundation (Y5110145), and Zhejiang A & F University professionals' surporting project (2014FR078).

References


State Forestry Administration of China. (2003). *Technical regulations of forest resources planning and design*.


**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).