International Journal of u- and e- Service, Science and Technology
Vol.9, No. 1 (2016), pp.295-306
http://dx.doi.org/10.14257/ijunesst.2016.9.1.31

A Regional Forest Tree Layer Biomass Estimation Method Based on Clustering Analysis

Wang Nihong[1,*], Gao Meng[1], Liu Lichen[2] and Gao Lewen[3]

[1]*Information and Computer Engineering College, Northeast Forestry University, Harbin, China*
[2] *College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin, China*
[3]*China Mobile Communications Corporation Design Institute Co., Ltd. of Heilongjiang Branch, Harbin, China*
wnh@mail.nefu.edu.cn

Abstract

As an area always contains varies of tree spices or forest types, therefore, when using biomass estimation model based on single tree or forest stand to estimate regional biomass, the modeling workload is big, and the existing models do not adequately reflect the factors that influence the biomass. Aiming at the problems above, this paper proposes a regional forest tree layer biomass estimation method based on clustering analysis, using the forest resources survey data of the study area as the research object, using principal component analysis to extract characteristic factors from 17 indexes, using the improved K-means algorithm to clustering the forest subcompartment, and using support vector regression algorithm to separately build the biomass estimation model based on clusters. The results show that 8 principal components can reflect over 80% information of the original data; the subcompartment of the study area can be divided into 6 classes, the coefficients of each model are ranging from 0.7 to 0.92, the average relative error absolute values of each model are ranging from 11.173% to 23.583%, this method has got a satisfactory accuracy, which can provide a new way for regional biomass estimation.

Keywords: *Biomass estimation; Principal components analysis; K-means; SVR algorithm*

1. Introduction

The global greenhouse effects are becoming more and more severe, and the role that forest plays in absorbing CO_2 and releasing O_2 is becoming more and more important, therefore, research on forest biomass has become a hotspot in recent years, under the background of Kyoto Protocol, quantify the forest biomass can make foundation for promoting the carbon sink marketization [1,2].

At present, single tree biomass models based on species are often been used to estimate biomass, Feng Zongwei [3], Li Fengri [4], Wang Hongyan [5] and other scholars have established many single tree biomass models of different species by using the allometric growth equation. Some scholars have established forest stand biomass models based on forest types, Jia Weiwei established planted Larix forest biomass model [4], Wang Lihai [6] established coniferous and broad-leaved forest biomass model. Researches in above have obtained reliable precision on biomass estimation, however, as an area always exists varies of tree species or forest types, therefore, workload of using the above models to

*Corresponding author. Tel.: +86 0451 82191401.
E-mail address: wnh@mail.nefu.edu.cn (N.H. Wang)

ISSN: 2005-4246 IJUNESST

estimate regional biomass is too big. In addition, in order to simplify the modeling complexity, the existing single tree or forest stand biomass models do not adequately reflect the factors that influence the biomass.

Therefore, this paper studies a regional forest tree layer biomass estimation method based on clustering analysis, based on forest subcompartment, use principal component analysis to extract characteristic factors from the forest resources survey data, and calculate the component score of each subcompartment; based on the component score, use cluster analysis to clustering the subcompartment, and obtain clusters with far cluster distance and close internal distance; based on the clusters, take component score as input and the corresponding biomass as output, and build the biomass estimation model based on subcompartment clusters. This method can reduce the workload in estimating regional biomass, and the model evaluation results show that the established models can get high estimation precision.

2. Data Resource and Pre-processing

2.1 Research Data and Pre-processing

Research area is Mengjiagang forest farm, which is located in Jiamusi City, Heilongjiang Province of China, and the research data is obtained from second type forest resources inventory data surveyed in 2012, it includes 1370 subcomparment records, and each record includes 79 survey factors, such as forest category, forest origin, area, volume, age group, forest type, forest density, soil and so on.

Delete the record which belongs to non-forest land and whose forest age is under 10 years, after this process, 829 records are remaining, in order to cleanup the data, use methods such as average value and formula derivation to fill the missing values, and amend the obviously abnormal properties.

2.2 Calculate the Actually Measured Biomass of Subcompartment

Due to the limited manpower and material resources, and the biomass of a subcompartment is hard to actually measured, therefore, collect biomass estimation models whose study area is similar with Mengjiagang forest farm in climate, geomorphic and other environment conditions, the collected models include single tree biomass models and forest stand biomass models, use the estimated biomass as the actually measured biomass as the real biomass of the sample subcomparment. Models used in this paper are shown in Tab.1.

Table 1. Single Tree and Forest Stand Biomass Estimation Models

Tree species /Forest types	Model	MRE (%)	P (%)	Source
Pinus koraiensis	$W = 0.0761627D^{2.5326936}$	6.9221	91.3471	Wang Hezhi, Li Fengri [7]
Picea asperata	$W = 0.0983775D^{2.4851609}$	0.5136	92.3674	
Larix gmelinii	$W = 0.2692528D^{2.1905198}$	-6.0022	91.919	
Tilia tuan	$W = 0.3024688D^{2.111704}$	12.7024	82.5733	
Quercus palustris	$W = 0.1892824D^{2.3519584}$	20.6167	80.9469	
Tree species /Forest types	Model	MRE (%)	P (%)	Source
Populus L.	$W = 0.2229582D^{2.0825374}$	-2.7421	95.1379	Wang Hezhi, Li Fengri [7]
Betula platyphylla Suk	$W = 0.312379D^{2.1983203}$	-9.6588	89.4095	
Populus davidiana	$W = 0.0762433D^{2.5669417}$	-6.4665	93.2052	
Pinus sylvestnis var	$W = 0.14457142D^{2.3016098}$	-1.3806	92.9659	Xing Yanqiu,
Broad-leaved forest	$W = 3.328 + 0.521V + 0.446D - 0.004H$	5.82	98.83	

Mixed forest	$W = -0.763 + 0.503V + 2.778H$	-5.66	97.02	Wang Lihai [6]

Note: D is the average diameter at breast height (cm); H is the average tree height (m); C is the forest canopy density ; V is the forest volume of per hectare (m^3/hm^2); A is the forest average age.

3. Characteristic Factors Extraction and Clustering

3.1 Extract the Characteristic Factors that Influence the Forest Biomass

There are 79 survey factors in the original data, it needs to find critical factors first, and they are strongly related to forest biomass. Considering that too many factors may cause collinearity problem, use principal component analysis method to extract characteristic factors, they can reflect the most information of the original data, and the extracted factors are independent mutually [8], they can be used to do clustering analysis and build the biomass estimation models. Assume that it has n subcompartment records in the original data, each record has m properties, $contributeLine$ is threshold of the cumulative contribution rate, the characteristic factors extraction algorithm can be described as follows:

(1) $x_{ij}(1 \le i \le n, 1 \le j \le m)$ is the value of property j in subcomparment i, quantize x_{ij} and build the orginal data matrix X_{ij}.

(2) normalize X_{ij}, eliminate dimensional influence, and get the standard matrix X_{ij}'.

(3) under the principle that the characteristic value is greater than a specific value, extract k mutually independent principal components, record each component's intial feature root $\lambda_p, 1 \le p \le k$ and its contribution rate $contributeRate_p$ on summarizing original information, sequencely calculate each component's accumulated contribution rate $totalContributeRate_p = \sum_{q=1}^{k} contributeRate_q$.

(4) when $totalContributeRate_p \ge contributeLine$, the contained p principal components are the extracted new characteristic factors.

(5) calculate the factor loading matrix M_{jp}, m_{jp} is the loading that property j works on pricipal component p.

(6) on the basis of M_{jp} and λ_p, according to $a_{jp} = \frac{m_{jp}}{\sqrt{\lambda_p}}$, calculate characteristic vector matrix A_{jp}.

(7) according to A_{jp}, seperately bulid p components expressions, calculate the component score of n subcomparment records, and get the subcomparment component score matrix F_{np}.

3.2 The Improved K-means Algorithm

Traditional K-means algorithm is based on the setting of target clusters number k and the selection of initial clustering centers, however, if lacking of enough domain knowledge, it is difficult to determine the number k, if k is not accurate, it will directly affect the clustering effect. In addition, traditional K-means algorithm is random in

selecting initial clustering centers [9,10], it can cause initial clustering centers are varies under different number k, therefore, the comparability of clustering effect under different k becomes poor. For this reason, use hierarchical clustering, BWP clustering validity index and maximum minimum distance algorithm to improve the traditional K-means algorithm, the improved algorithm can be described as follows:

(1) Use hierarchical clustering analysis to determine the searching scope of k, namely $[k_{\min}, k_{\max}]$.

(2) Set target clustering number k, at initial time, $k = k_{\min}$, if $k \le k_{\max}$, turn to step (3)；else, turn to step (10).

(3) Use average value to calculate the center c_0 of dataset $D_n = \{x_1, x_2, ..., x_n\}$, regard the sample which is closet to c_0 as the first initial clustering center c_1.

(4) Within $D_n = \{x_1, x_2, ..., x_n\}$, separately calculate the distance between each remaining sample x_i and c_1, regard the sample which is farthest from c_1 as the second initial clustering center c_2.

(5) Within $D_n = \{x_1, x_2, ..., x_n\}$, separately calculate the distance between each remaining sample x_i and c_1 as well as x_i and c_2, calculate the minimum value d_i of them, regard the sample x_t corresponding with $\max(d_i)$ as the third initial clustering center c_3.

(6) If it has already existed $k-1$ centers, within $D_n = \{x_1, x_2, ..., x_n\}$, separately calculate the distance d_{ij} between each remaining sample x_i and c_1, x_i and c_2, …, x_i and c_{k-1}, calculate the minimum value $d_i = \min(d_{i1}, d_{i2}, ..., d_{i(k-1)})$, regard the sample x_r corresponding with $\max(d_i)$ as the kth initial clustering center c_k.

(7) Calculate the distance d_{ij} between x_i and c_1, x_i and c_2, …, x_i and c_k.

(8) Divide x_i into cluster C_j corresponding with $\min(d_{ij})$, and recalculate the center of C_j.

(9) Judge whether each $C_1, C_2, \ldots, C_k$ is stable, if each cluster is not stable, turn to step (7); else, turn to step (3).

(10) Compare the average BWP index value $avg(BWP)$ under different k, and regard the k corresponding with the maximum $avg(BWP)$ as the best clustering number k_{opt}.

4. Experiment Results and Analysis

4.1 Characteristic Factor Extraction Results and Analysis

4.1.1 Biomass Characteristic Factor Extraction: Based on the second type forest resources inventory data and the subcomparment biomass calculated in the above, use SPSS19.0 to do partial correlation analysis between biomass and 79 initial factors, set significance level $\alpha = 0.05$. Through analyzing, 17 factors related with biomass pass the significance testing, they are forest category, forest origin, forest type, forest canopy density, site type, mean dominant tree height, underwood coverage, ground cover coverage, slope aspect, slope position, slope gradient, elevation, DBH, mean tree height, volume per hectare, growth rate, forest age.

Before principal analyzing, quantize the descriptive properties of the 17 factors, in order to eliminate the dimensional influence, use SPSS19.0 to do Z-standardization processing on original data, then conduct the KMO and Bartlett sphericity testing on the standarded data, the result shows the KMO value is 0.771 and the Sig value of Bartlett sphericity testing is 0.0, which means that the 17 factors has strong correlation mutually, namely, they are suitable for principal analyzing.

According to the principle that the each factor's initial characteristic value is greater than 0.7, use SPSS19.0 to extract principal components, and get the total variance each factor has explained as Tab.2.

Table 2. Total Variance Each Factor has Explained

Factor	Initial characteristic value			Extracted square sum loading		
	Total	Explained variance (%)	Accumulated variance (%)	Total	Explained variance (%)	Accumulated variance (%)
1	4.658	27.399	27.399	4.658	27.399	27.399
2	2.286	13.445	40.844	2.286	13.445	40.844
3	1.660	9.764	50.608	1.660	9.764	50.608
4	1.503	8.842	59.450	1.503	8.842	59.450
5	1.117	6.568	66.018	1.117	6.568	66.018
6	0.935	5.498	71.516	0.935	5.498	71.516
7	0.793	4.667	76.183	0.793	4.667	76.183
8	0.734	4.318	80.501	0.734	4.318	80.501
9	0.679	3.995	84.496			
10	0.601	3.538	88.034			
11	0.500	2.940	90.974			
12	0.454	2.668	93.642			
13	0.375	2.208	95.850			
14	0.327	1.925	97.775			
15	0.246	1.446	99.221			
16	0.113	0.668	99.888			
17	0.019	0.112	100.000			

It can be seen that the former 8 principal components can explain over 80% accumulated variance of the total variance, then calculate the coefficient between 8 components and 17 factors, and get the initial factor loading matrix, in order to explain the components better, rotate the initial factor loading matrix, and the rotated initial factor loading matrix is shown as Tab.3.

Table 3. The Rotated Initial Factor Loading Matrix

Factor	component							
	1	2	3	4	5	6	7	8
Forest category	0.145	0.140	-0.112	0.111	0.869	0.013	0.057	-0.086
Forest origin	-0.072	0.692	-0.085	0.119	0.395	0.325	-0.038	0.025
Site type	0.033	0.125	0.031	0.861	0.037	-0.050	-0.139	0.046
Forest type	-0.007	0.929	-0.039	0.032	0.001	-0.141	0.067	-0.054
Mean dominant tree height	0.921	0.006	-0.027	0.048	-0.027	0.261	-0.001	-0.018
Underwood coverage	-0.040	0.101	0.828	-0.045	-0.041	-0.078	-0.094	-0.110
Ground cover coverage	-0.014	-0.176	0.619	-0.054	-0.361	-0.069	0.299	-0.011
Slope aspect	-0.017	0.041	-0.018	0.048	0.027	0.032	0.949	-0.003
Slope position	-0.034	-0.027	-0.064	-0.095	-0.068	-0.025	-0.003	0.980
Slope gradient	0.010	0.064	0.055	-0.817	-0.060	-0.072	-0.195	-0.055

Elevation	-0.029	-0.441	-0.316	0.204	-0.406	-0.328	0.076	-0.041
DBH	0.894	0.177	0.057	0.013	0.099	0.125	-0.032	-0.036
Mean tree height	0.912	0.014	-0.003	0.066	0.011	0.279	-0.010	-0.018
Volume per hectare	0.489	-0.004	-0.068	0.072	0.111	0.714	0.045	-0.028
Growth rate	-0.843	0.154	0.063	0.037	-0.038	0.070	-0.027	-0.031
Forest age	0.878	-0.067	-0.174	-0.058	0.061	-0.092	-0.014	-0.031

4.1.2 Biomass Characteristic Factor Explanation: In the first component, mean dominant tree height (0.921), mean tree height (0.912), DBH (0.894) and forest age (0.878) have played the important role, which means this component reflects the forest growth situation.

In the second component, forest type (0.929) and forest origin (0.692) have played the important role, which means this component reflects the forest natural attribute.

In the third component, underwood coverage (0.828), forest canopy density (0.630) and ground cover coverage (0.619) have played the important role, which means this component reflects the vegetation covered situation.

In the fourth component, site type (0.861) and slope gradient (-0.817) have played the important role, which means this component reflects the forest site conditions.

In the fifth component, forest category (0.869) has played the important role, which means this component reflects the forest expected utility.

In the sixth component, volume per hectare (0.714) has played the important role, which means this component reflects the condition and quality of the forest resources.

In the seventh component, slope aspect (0.949) has played the important role, which means this component reflects the slope orientation of forest.

In the seventh component, slope aspect (0.949) has played the important role, which means this component reflects the slope orientation of the forest.

In the eighth component, slope position (0.980) has played the important role, which means this component reflects the slope height of the forest.

Finally, calculate 8 factors' expression coefficients on 17 variables, multiply each subcompartment's 17 attributes' value, we can get each subcomparment's 8 components' score.

4.2 Clustering Results and Analysis

4.2.1 Clustering Results: Based on the calculated 829 records' score in the above, under MATLAB R2011a, separately use traditional K-means and the improved K-means algorithm to do clustering analysis on the 829 records. Use SPSS19.0 to do hierarchical clustering analysis first, clustering method uses Ward and between-class distance measurement uses square Euclidean distance, combine with multiple experiment results, the target clusters number k can be set between 4 and 9, namely $k_{\min} = 4$ and $k_{\min} = 9$.

As traditional K-means algorithm is random in selecting initial clustering centers, in order to reduce this defect's influence on clustering results as far as possible, running the algorithm 30 times, and take the mean value of the 30 times as the final results. The changes of mean BWP value and the running efficiency along with clusters number k are shown as Fig.1 and Fig.2.

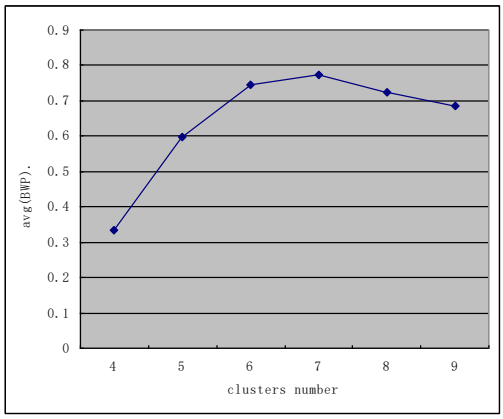

Figure 1. Average BWP Value Changing with k

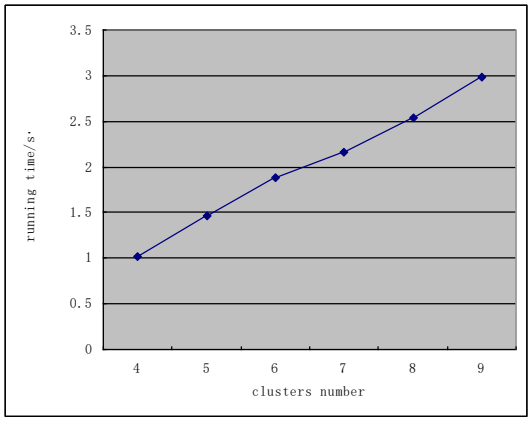

Figure 2. Running Time Changing with k

When using the improved K-means algorithm to select the initial clustering centers, as long as the sample dataset is the same, the final clustering results will be the same, compared with traditional K-means algorithm, the improved K-means algorithm is stable. Therefore, it is not necessary to running the algorithm repeatedly. Based on the same dataset used in traditional algorithm, the changes of mean BWP value and the running efficiency along with clusters number k are shown as Fig.3 and Fig.4.

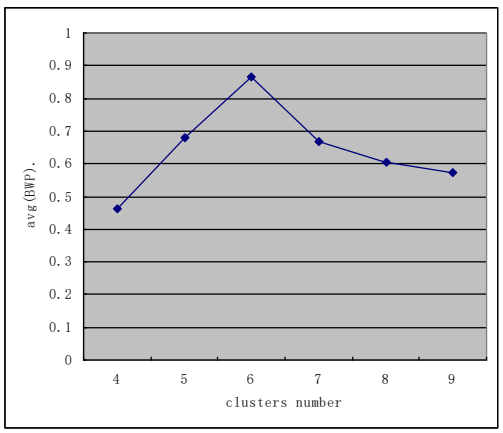

Figure 3. Average BWP Value Changing with k

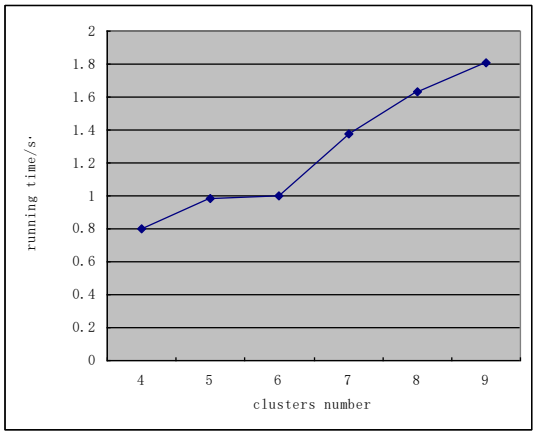

Figure 4. Running Time Changing with k

Through experiments in the above, the optimum cluster numbers, the maximum BWP value and the algorithm running time of traditional and the improved K-means algorithms are shown in Tab.4.

Table 4. Comparision Result of Two Algorithms

index / algorithm	$k_{\min}$	$k_{\max}$	k_{opt}	Maximum $avg(BWP)$	Running time/s
Traditional K-means	4	9	7	0.7736	2.163
The improved K-means	4	9	6	0.8664	1.002

It can be seen that the maximum $avg(BWP)$ of the improved K-means algorithm (0.8664) is higher than traditional K-means algorithm (0.7736) based on the same dataset,

namely, clusters number is 6 can get the best clustering effect; the algorithm running time of the improved K-means algorithm (1.002s) is shorter than traditional K-means algorithm (2.163s), namely, the improved algorithm has the better running efficiency.

In conclusion, the subcompartment of the study area can be divided into 6 classes, records number of each class are respectively 133, 150, 119, 163, 188 and 76.

4.2.2 Clustering Results Explanation: In the first class, the records all belong to natural water conservation forest, the forest ages are ranging from 30a to 50a, and forest types include soft broad-leaved mixed forest and oak forest. It means that the second component has played an important role in the first class.

In the second class, the records all belong to planted general timber forest, the forest ages are ranging from 35a to 55a, and the canopy densities are over 0.5, most are between 0.7 and 0.9, slope positions are focused on middle and up location, slope gradients are focused on flat and slow degree, slope aspects are focused on east, west, northeast and northwest. It means that the fifth, sixth, seventh and eighth component have played an important role in the second class.

In the third class, most records belong to the planted forest, the natural forest's rate is 1.32%, the forest ages are ranging from 20a to 40a, and the canopy densities are over 0.5, most are between 0.6 and 0.8, slope positions are focused on middle and down location, slope aspects are focused on east, west, north, site types are focused on shade slide slope and sunny slide slope thick-layer dark brown earth. It means that the seventh, eighth, third and fourth component have played an important role in the third class.

In the fourth class, all records belong to the planted forest, the canopy densities are ranging from 0.5 to 0.9, slope positions are focused on middle and up location, and slope gradients are focused on slow and oblique degree. It means that the fourth, seventh, eighth, second and third component have played an important role in the fourth class.

In the fifth class, all records belong to planted forest, the forest ages are ranging from 35a to 50a, forest types include *Larix gmelinii*, *Pinus koraiensis* and *Pinus sylvestris*, forest canopy densities are over 0.5, most are between 0.6 and 0.9, slope gradients are focused on flat and slow degree, site type are focused on shade slide slope and sunny slide slope thick-layer dark brown earth. It means that the third and fourth component have played an important role in the fifth class.

In the sixth class, forest ages are ranging from 10a to 30a, canopy densities are below 0.5, most are between 0.2 and 0.5, slope gradients are focused on flat and slow degree, and site types are focused on shade slide slope and sunny slide slope thick-layer dark brown earth. It means that the first, third, sixth and fourth component have played an important role in the sixth class.

4.3 Biomass Estimation Modeling Results and Analysis

4.3.1 Biomass Estimation Modeling Results: 829 records can be divided into 6 classes, records number of each class are respectively 133, 150, 119, 163, 188 and 76, randomly choose typical records from each class, and 75% of them are the modeling samples and the other 25% records are the validating samples. Sample and factor numbers of each class are shown in Tab.5.

Table 5. Modeling, Validating Sample and Modeling Factor Numbers of Each Class

Class	Total samples	Modeling samples	Validating samples	Modeling factors
1	106	80	26	8
2	120	90	30	8
3	95	72	23	8

4	130	100	30	8
5	152	114	38	8
6	63	48	15	8

Based on support vector regression method [11], use libsvm to find optimal parameters of each class, separately modeling based on 4 kinds of kernel functions (linear, polynomial, RBF, Sigmoid) and 2 kinds of models ($\varepsilon - SVR$, $v - SVR$), the parameter optimization results of each class are shown in Tab.6.

Table 6. Parameter Optimization Results of Each Class

Class	Model type	Kernel function	parameter			MSE	R^2
			c	g	p/v		
1	$\varepsilon - SVR$	RBF	1024.0	0.0625	1.0	47.0075	0.922873
2	$\varepsilon - SVR$	RBF	1024.0	0.125	64.0	788.65	0.870854
3	$v - SVR$	RBF	512.0	0.0625	1.0	611.915	0.868646
4	$v - SVR$	Sigmoid	512.0	0.125	1.0	1097.02	0.758751
5	$v - SVR$	Sigmoid	64.0	0.25	1.0	1071.25	0.705027
6	$\varepsilon - SVR$	RBF	1024.0	0.125	8.0	618.4	0.796332

4.3.2 Models Precision Evaluation: In order to validate the modeling effect, based on the same modeling and testing samples, separately use linear regression and BP neural network to build the model of each class, and use determination coefficient R^2 and absolute mean relative error(AMRE) to evaluate the 3models of each class. The evaluation results of 3 models are shown in Fig.7.

It can be seen that support vector regression models of all classes are optimal in R^2 and AMRE, BP neural network models are inferior to support vector regression models, and multiple linear regression models are the worst, namely, biomass has a strong and complex non-linear relationship with the impact factors, and the results have improved that support vector regression has the superiority on generalization ability.

Table 7. Biomass Estimation Model Precision Evaluation Results of Each Class

Class	Testing samples	Multiple linear regression model		BP neural network model		Support vector regression model	
		R^2	AMRE（%）	R^2	AMRE（%）	R^2	AMRE（%）
1	26	0.576	41.07	0.747	17.577	0.922873	10.433
2	30	0.569	27.73	0.713	12.167	0.870854	11.173
3	23	0.645	30.3478	0.664	20.3478	0.868646	17.387
4	30	0.535	44.967	0.725	26	0.758751	22.658
5	38	0.463	45.5263	0.698	21.0526	0.758751	17.487
6	15	0.641	42.749	0.657	31.022	0.705027	23.583

5. Conclusions

(1) 8 biomass characteristic factors have been extracted, the calculation results show that 8 components can reflect over 80% original information, each component has an obvious meaning, compared with the parameters of traditional biomass models, and the components extracted in this paper can fully reflect the factors that influence the biomass level.

(2) A biomass estimation modeling method based on clustering analysis has been proposed, use the improved K-means algorithm to clustering the forest subcomparment, and the clustering results show that the subcomparment of the study area can be divided

into 6 classes, each class has the clear characteristics, and the differences among different classes are obvious. Compared with traditional biomass modeling method based on tree species or forest types, this method has advantages in regional biomass estimation.

(3) 6 biomass estimation models based on support vector regression method have been established, compared with multiple linear regression model and BP neural network model, the SVR models have the highest fitting precision, the R^2 are ranging from 0.7 to 0.92, and the AMRE are ranging from 11.173% to 23.583%.

Acknowledgements

This work is supported by The Foundation of National 863 Plan of China (2012AA102003-2) and The Fundamental Research Funds for the Central Universities (2572014AB22).

References

[1] V. Uri, M. Varik and J. Aosaar, "Biomass production and carbon sequestration in a fertile silver birch (Betula pendula Roth) forest chronosequence", Forest Ecology and Management, vol. 267, (**2013**), pp. 117-126.
[2] C. Liu and X. Li, "Carbon storage and sequestration by urban forests in Shenyang, China", Urban Forestry & Urban Greening, vol. 11, **(2012)**, pp. 121-128.
[3] X. Wang, Z. Feng and O. Zhiyun, "Vegetation carbon storage and density of forest ecosystems in China", Chinese Journal of Applied Ecology, vol. 1, **(2001)**, pp. 13-16.
[4] W. Jia, F. Li, L. Dong and X. Zhao, "Carbon density and storage for Pinus sylvestris var. mongolica plantation based on compatible biomass models", Journal of Beijing Forestry University, vol. 1, **(2012)**, pp. 6-13.
[5] H. Wang, W. Wang, L. Qiu , D. Su, J. An, G. Zheng and Y. Zu, "Differences in biomass, litter layer mass and SOC storage changing with tree growth in Larix gmelinii plantations in Northeast China", Acta Ecologica Sinica, vol. 3, **(2012)**, pp. 833-843.
[6] Y. Xing and L. Wang, "Compatible biomass estimation models of natural forests in Changbai Mountains based on forest inventory", Chinese Journal of Applied Ecology, vol. 1, **(2007)**, pp. 1-8.
[7] H. Wang and F. Li, "Dynamic simulating system for stand growth of forests in Northeast China", Northeast Forestry University, vol. 6, **(2012)**, pp. 86-90.
[8] B. Yan, Z. Hu and B. Yu, "Application of principal component analysis and cluster analysis to evaluating ecological safety of land use", Transactions of the CSAE, vol. 8, **(2006)**, pp. 88-90.
[9] M. K. Pakhira, "A fast K-means algorithm using cluster shifting to produce compact and separate clusters", International Journal of Engineering, vol. 1, **(2015)**, pp. 36-45.
[10] G. Chen, M. Li, D. Wei and M. Xin, "Applied research of combinatorial algorithm of clustering, rough set and decision tree method in productivity evaluation", Scientia Agricultura Sinica, vol. 7, **(2011)**, pp. 4833-4840.
[11] J. Gu, J. Liu and M. Chen, "A modified regression prediction algorithm of large scale data based on SVM", Computer Engineering, vol. 1, **(2014)**, pp. 161-166.

Authors

Wang Ni-hong, Female, born in 1952, Ph.D. supervisor, working in Information and Computer Engineering College of Northeast Forestry University, mainly engaged in forestry informatization, Internet of Things.

Gao Meng, Female, born in 1989, Ph.D., studying in Information and Computer Engineering College of Northeast Forestry University, mainly engaged in data mining and data analysis.

Liu Li-chen, Male, born in 1983, Ph.D., studying in College of Mechanical and Electrical Engineering of Northeast Forestry University, working in Harbin Kejia Universal Electric Mechanical Corporation Co., Ltd., mainly engaged in electromechanical integration technology , vehicle control.

Gao Lewen, Male, born in 1981, Master's Degree, studied in Information and Computer Engineering College of Northeast Forestry University, working in China Mobile Communications Corporation Co., Ltd. of Heilongjiang Branch, mainly engaged in Engineering consulting and design.

