NEURAL NETWORK-BASED ANALYTICAL MODEL FOR BIOMASS ESTIMATION IN POYANG LAKE WETLAND USING ENVISAT ASAR DATA

Dong Lei, Liao Jingjuan, Shen Guozhuang

State Key Laboratory of Remote Sensing Science
Institute of Remote Sensing Applications, Chinese Academy of Sciences，Beijing 100101 - ldong@irsa.ac.cn

KEY WORDS: biomass sar wetland monitoring

ABSTRACT:

Poyang Lake is the largest freshwater lake in China with an area of about 3,000 km2. Its wetland ecosystem has a significant impact on China’ s environment change. Traditional way to monitor biomass change on this area is to use linear or nonlinear model from TM/ETM data. In this paper we discuss the neural network algorithms(NNA) to retrieve wetland biomass from multi-polarization(HH and VV) backscattering values using Envisat ASAR data. Two field measurements were carried out concomitant to the acquisition of ASAR images in this area through the hydrological cycle from Apr and Nov. Training data of the network are generated by MIMICS model which is often used in the forest. We simplify the model to make it available on the wetland system. The model input parameters are defined according to the real wetland circumstance. NNA retrieval results are validated with experimental data. The inversion results show that the NNA is capable of performing the retrieval with good accuracy. Finally, the trained neural network is used to estimate the overall biomass of the Poyang Lake. The wetland biomass reaches a level of 1.06x109 kg,1.72x108 kg, 1.0x109 kg in April, July and November 2007.

1. INTRODUCTION

Traditional optical remote sensing using Landsat TM data to estimate biomass in this region has been conducted in the last few years. Using vegetation index obtaining from TM satellite image and wetland vegetation data from the corresponding field spots, non-linear regression analysis model has been proved more accurate than the linear model.(Jian et al., 2005). Overall biomass estimations have been reported in this area using ETM data (Rendong and Jiyuan, 2001).

In recent years, research in remote sensing has led to the development of methods for retrieving wetland biomass from backscattering values collected by Radar system. The feasibility of inverting ERS data into biomass shows the relevance of using C-band VV polarized data (Moreau and Toan, 2003). A combination of Radarsat and JERS-1 images were used to understand the saturation point in the logarithmic relationship between backscattering coefficients and biomass in the Amazon floodplain. (Costa et al.,2002). There exists no studies using SAR images to estimate the biomass in this area before, especially at C-band. A major problem in wetland biomass inversion from SAR data is that influence caused by a number of other environmental variables such as water content, vegetation height. (Frate, 2004). To reduce to effect, multiple-frequency and multiple datasets have been mainly considered. (Dobson et al., 1992).

Most of the study in retrieving biomass has been focused on the implement of linear and non-liner regression model (Kasischke et al., 1995), (Polatin et al., 1994), (Rignot et al., 1994). But the mapping between SAR image and ground surface parameters are always very complex because of the strong nonlinearities (Frate and Wang, 2001). The semi-empirical regression model based on real measurement can not

express the relationship sufficiently. In this paper, a neural network algorithms is combined with a canopy scattering model to perform the inversion. Neural network is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve complex problems. This structure makes NN inherently suitable for solving nonlinear problems.

Poyang Lake is the largest freshwater lake in China with an area of about 3,000 km2. (Rendong and Jiyuan, 2001). The predominant vegetation is large homogeneous stands of grass-like plants and some aquatic plants such as bulrush. Its wetland ecosystem has a significant impact on China’ s environment change, so it is of great importance to monitor the change of biomass in this region. In this paper, hh and vv polarization data have been used to estimate biomass around the lake, and we focuses on the use of canopy scattering model and Neural Network Algorithms (NNA) to establish the relationship between the backscatter values and the biomass instead of using linear or non-linear model. Although data-driven regression models are easier, they have no exact meaning, so we can not determine the influence caused by each variable. While the theoretical model can give us a clear structure and approach to understand the physical mechanism during the interaction between the microwave and the vegetation.

Comparison between the Neural Network algorithms and both linear and nonlinear regression algorithms points out the overall superior performance of the neural algorithm using SAR images at both P and L band (Frate, 2004). In this paper we use the algorithms to invert total biomass in Poyang Lake in three temporal stage from April to November.

2. TEST SITE AND DATASET

2.1 Test Site

Poyang Lake is located in Jiangxi Province. It is connected to the Yangtze (Chang Jiang) River through a narrow channel, exiting at the top left corner of the image (Figure 1).

Figure1. Study area: Poyang Lake

In summer it is the largest freshwater body in China, by the end of the rainy season the Lake can extend up to 3500 km2 in area, but during the dry season it may shrink to under 1000 km2. The receding waters leave behind a system of wetlands and mudflats which attract up to half a million migratory waterfowl. The predominate vegetation in the Poyang Lake is mostly carex and miscanthus floridulus.(Figure2). The height, water content and plant density of vegetation vary in different hydrological stage from April to November. In April, the carex is starting to grow with green leaves at about 50 - 70 cm. While in July, the plants are submerged by the water. In November, it turns yellow and the plant water content is becoming lower.

Figure 2. Large homogeneous stands of carex

2.2 SAR Data

Three Envisat ASAR scenes were collected over this area in three period of 2007: April, July and November. Table1 shows the ASAR data details. To radiometrically correct the ASAR imagery, we used the BEST toolbox provided by the European Space Agency. Then the different images were georeferenced and mean backscatter values were determined by averaging 7x7 pixels surrounding each study site. ASAR is sensitive to water or waterlogged surfaces, the images has a pixel resolution of 30m.

Date	Polarization	Incident Angle
2007-04-01	HH/HV	30-36°
2007-04-04	HH/VV	39-43°
2007-04-06	HH/VV	39-43°
2007-07-28	HH/VV	18-26°
2007-11-27	HH/VV	18-26°

Table 1. ASAR data for the study area

2.3 Field Measurement

To validate the accuracy of NNA, two field measurements were carried out concomitant to the acquisition of SAR imagery inApril and November. Figure3 shows a dynamic change of water level of the lake in 2007.

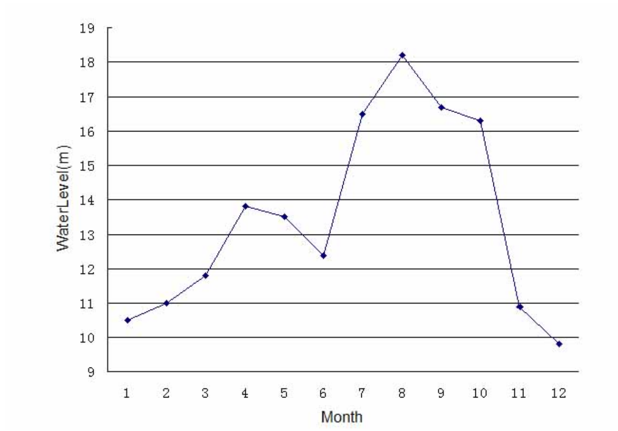

Figure 3. Poyang Lake water level in 2007

In April, when the carex is in growing season, the water is rising with a high soil moisture, at this time the Poyang Lake is typical wetland system. It reaches the peak level in August. By the end of November when the carex tends to wilt, due to the serious shortage of rainfall in Jiangxi province and the impact of the low water level in Yangtze River, most of the sample sites have low soil moisture, basically in drought conditions. (Table2)

Date	Samples	Vegetation	Soil Moisture
2007-04	52	Growing	High
2007-11	57	withering	low

Table 2 Field collection summary in April and November

In both stage, ground data: including plant water content, above ground biomass, plant height, were collected in more than 50 study sites. The biomass was measured using a square quadrat of 0.5m x 0.5m. Aboveground vegetation samples were harvested and the heights of the vegetation were measured. Each sample was weighted and oven-dried at 120oC to a constant weight and reweighed. According to the ground data, biomass is correlated with plant height, with a correlation coefficient of 0.61 in a 90% confidence interval.

3. METHODS

In the previous study of biomass estimation based on neural network, real ground measurement are mostly used as input unit and the backscattering coefficients extracted from SAR image are output unit. However, the training data are completely rely on the ground truth, and it becomes unavailable when the study area are hard to access or ground data are hard to get. The inversion accuracy is mostly determined by the measurement accuracy.

In this paper, we use MIMICS model to generate training data for the neural network. The purpose of using theoretical model is mainly to know how the major input parameters influence the backscattering and form a certain calculation pattern for the biomass estimation in the unaccessible field.

The inversion is organized with the following steps:
First, MIMICS model is used to simulate training inputs (biomass, plant height, water content)and output(HH and VV backscattering).
Then train the neural network with the data generated in the previous step and invert ASAR image backscattering values to biomass.
Third, validate the results with the real ground measurement.
Finally, calculate the overall biomass of the lake.

3.1 Generating training data

MIMICS(Michigan Microwave Canopy Scattering) is a canopy scattering model which has been widely used for the tree canopy comprising a crown layer, a trunk layer and a rough-surface ground boundary. (Ulaby et al., 1990). The canopy scattering model supposes that the backscatter value is governed by the following two
factors: the direct scattering from the vegetation canopy and the backscatter from the ground surface (including the ground vegetation scattering and the multiple-path scattering between the surface and vegetation canopy.)

For the vegetation in our study area, this model should be changed because there is no trunk layer in the grass-like carex. So we made the flowing two hypothesizes:
(1) Simplify the three-layer (canopy-trunk-surface) model to two-layer (canopy-surface) because of the grass-like plants.
(2) Most of the study area are wetland system with high soil moisture in April and July 2007. So the ground surface is considered as water surface.

Training data, as prior knowledge of the network, is very important and it mainly determine the accuracy of the training results.
It needs to fit the following two conditions:
(1) The training data should vary in a wide range to cover different stages of plant growth.
(2) In the training data, the changes of backscattering coefficient with biomass, height, polarization, moisture content, incident angle and other parameters should be consistent with the ground measurements.
(3) Choose less correlated data as training input to reduce redundancy and improve simulation speed and accuracy.

In fact, real ground data can rarely meet the conditions for such training data. It's possible that the data don't meet the actual backscattering model because of the random error in measurement, and also may not cover all range of biomass in study area, which may lead to poor accuracy of the simulation. But theoretical model can simulate backscattering values including HH and
VV polarizations under all kinds of conditions, given the different range of ground parameters.

Table3 shows the input parameters of MIMICS model. Plant height, plant water content and above-ground biomass have been collected in the filed campaign. Leaf density and dimensions and others parameters are calculated according to the samples of study area. Soil moisture is estimated according to the water level. With this input parameters, a set of 50 network input-output pairs were generated.

Major Parameters	Other Parameters
Plant Height	Soil Moisture
Plant Water Content	Roughness of Soil
Leaf Density	Density of plants
Leaf Dimensions	Incident Angle

Table 3 Input parameters for the MIMICS model

3.2 Neural Network Algorithm

The topology of the NN that we used is a one-hidden-layer back-propagation (BP) neural network with three inputs elements and two outputs (Figure4).

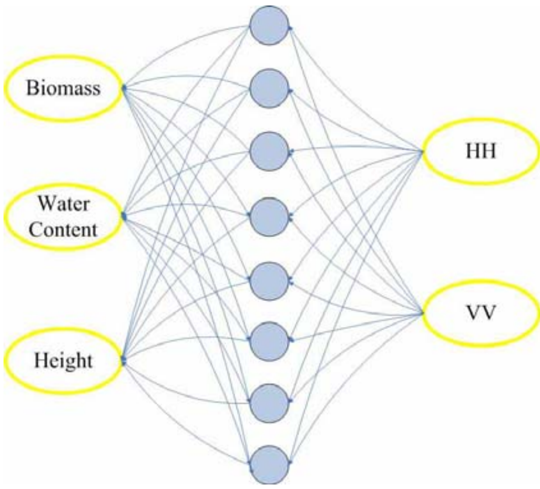

Figure 4. Neural Network training model

The activation function of each element in the hidden layer is the sine function, and logistic function in the output element defined by Equation(1)

Figure 5 shows that the training process quickly meet a performance goal of 0.001 after 257 epoches.

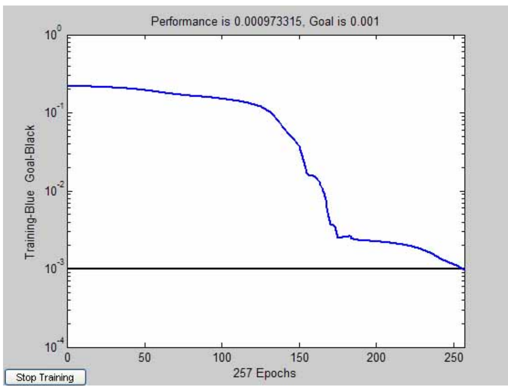

Figure 5. Neural Network training result

4. VALIDATION WITH GROUND DATA

In April, 52 samples were collected and 42 of them were used to validate the inversion accuracy of NNA.

Figure6 and Figure7 shows the error between ground truth and estimated biomass in April and November. The intermediate level of biomass between 400 - 800 g/km2 is well simulated, while the high and low biomass inversion results is not satisfactory. The high biomass estimation error is mostly due to the observed saturation of the backscattering coefficients versus biomass (Costa etal., 2002). After reaching the saturation point, high biomass level has similar backscattering values with the intermediate one, resulting in low biomass estimation in the results. The low biomass estimation error is mostly due to the soil surface. Low biomass always has low plant density, and the microwave can directly interact with the ground surface. So most of the backscattering distribution are due to soil surface instead of vegetation above them. This result in error while low biomass level vegetation is being inverted.

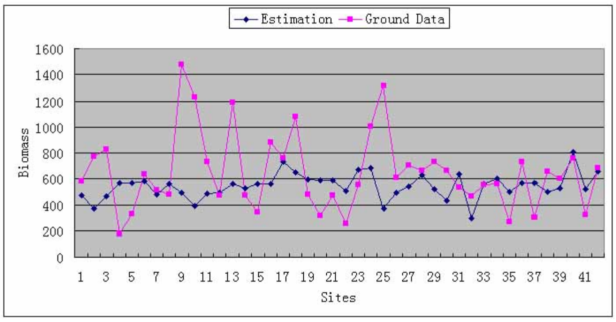

Figure 6. Estimation Biomass validation with ground data in April, 2007

The Radar backscattering values is determined by soil surface, water content, biomass level, vegetation height and many other factors. The inversion from backscattering values to biomass may be difficult because it is influenced by so many other factors. So with a high backscattering value in the image it is difficult to tell it's due to the high biomass or other factors. It is something like in the optical remote sensing that different ground objects may result in similar spectral response.

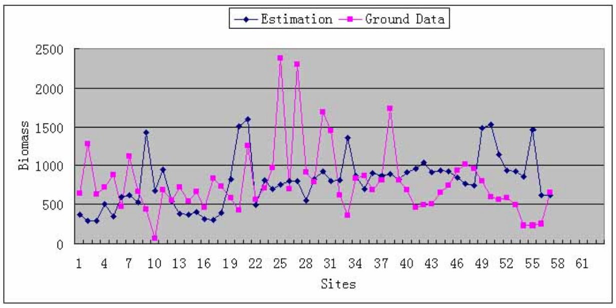

Figure 7. Estimation Biomass validation with ground data in November, 2007

The overall results of ASAR data inversion into biomass have a root mean square error (RMSE) of 0.3kg/m2 in April, and 0.4kg/m2 in November. While the intermediate biomass level estimation have a RMS of 0.17kg/m2 and 0.3kg/m2, this shows good inversion accuracy using NNA combined with MIMICS model.

5. RESULTS ANALYSIS

Figure8 gives the biomass mapping results using the above trained neural network. These maps of dry biomass on three dates, from April 2007 to November 2007, show temporal variations of biomass values. They clearly depict the development phases of carex, as
well as the changes of water level. In April, water level is rising and most of the wetlands are visible. The vegetation is mainly distributed in the southwest, south and southeast of the lake. In July, the water level is approximately 17m, near to the peak of the year, so most of lake is flooded, except the shoaliness in the higher ground in the south of the lake. In November, the lowest water level among the three months results in the smallest area of water body, we can see the most area of wetlands in the image.

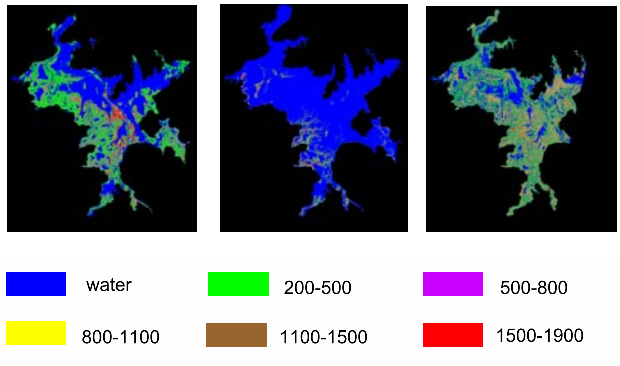

Figure 8: Above water dry biomass maps of Poyang Lake derived from Envisat ASAR data

Table4 shows the biomass level distribution in the three periods. In April, the overall dry biomass is 1.06x109 kg. It is the growing season of carex, nearly half of the biomass is in 250-500 g/km2. While in July, the percent of high biomass level is rising because the carex grows very fast at this time of the year, but with the rising of the water level, most of the wetlands are covered with water, so the overall biomass decrease to 1.72x108 kg2. During November as the carex wither and leaf turn to yellow the biomass level start to decrease, about 61.73% of the area are in range of 250-500(g/km2).The total biomass in November is nearly the same with in April. The average biomass is nearly consistent in the three images, because in April and November the area of vegetation is vast but most of them are very low, while in July, the vegetation grow very good and have a high biomass but the area of wetland are less than other two months.

Biomass Level(g/m^2)	April	July	November
250-500	46.76	51.04	61.73
500-800	31.0	21.43	21.1
800-1100	11.98	12.43	11.18
1100-1500	6.54	10.81	5.54
1500-2000	3.66	4.28	0.43
Average Biomass(g/m^2)	608	609	627
Total Biomass(kg)	$1.06x10^9$	$1.72x10^8$	$1.01x10^9$

Table 4: Biomass distribution comparison in three period.

6. CONCLUSIONS

This study focus on the application of neural network algorithms to retrieve wetland biomass from co-polarization backscattering coefficients using Envisat ASAR image. The training data of neural network is simulated by MIMICS model according to the certain circumstances in different time period or places. Trained neural network is used to estimate the overall above-ground dry biomass of the lake. The intermediate biomass level are well inverted but the inversion accuracy in high and low level need to be improved. The wetland biomass reaches a level of 1.06x109 kg,1.72x108 kg, 1.0x109 kg in April, July and November 2007. The retrieval result shows that the inversion model is capable of performing the estimation with a good level of precision with error of 0.3kg/m2 and 0.4kg/m2 in April and November. General problem of neural network is its ability in global-scale usage. In the future study, efforts for better understanding the model to make it globally used is of great worth.

ACKNOWLEDGEMENTS

The work is supported by The National High Technology Research and Development Program of China (2006AA12Z122).

REFERENCES

Costa, M., O. Niemann, Novo, E. and Ahern, F., 2002. Biophysical properties and mapping of aquatic vegetation during the hydrological cycle of the Amazon foodplain using Jers-1 and Radarsat. *International Journal of Remote Sensing* 23, pp. 1401–1426.

Dobson, M., Ulaby, F. T., Toan, T. L., A.Beaudoin, kasischke, E. and N.L.Christensen, 1992. Dependence of radar backscatter on coniferous forest biomass. IEEE Trans. *Geosci. Remost Sensing* 30, pp. 412–415.

Frate, F. D., 2004. On neural network algorithms for retrieving forest biomass from sar data. IEEE Trans. on *Geosci. Remote Sensing* 42, pp. 24–34.

Frate, F. D. and Wang, L., 2001. Sunflower biomass estimation using a scattering model and a neural network algorithm. *International Journal of Remote Sensing* 22, pp. 1235–1244.

Kasischke, E. S., Christensen, N. L. and Bourgeau-Chavez, L. L., 1995. Correlating radar backscatter with components of biomass in loblolly pine forest. IEEE Trans. *Geosci. Remost Sensing* 32, pp. 643–659.

Li, J., Shu, X. and Chen, S., 2005. Establishment of wetland vegetation biomass and remote sensing observation in poyang lake area. *Acta Geographica Sinica* 4, pp. 494–498.

Li, R. and Liu, J., 2001. An estimation of wetland vegetation biomass in the poyang lake using landsat etm data. *Acta Geographica Sinica* 56, pp. 532–540.

Moreau, S. and Toan, T. L., 2003. Biomass quantification of andean wetland forages using ers satellite sar data for optimizing livestock management. *International Journal of Remote Sensing* 84, pp. 477–492.

Polatin, P. F., Sarabandi, K. and Ulaby, F. T., 1994. An iterative inversion algorithm with application to the polarimetric radar response of vegetation canopies. IEEE Trans. *Geosci. Remost Sensing* 32, pp. 62–71.

Rignot, E., Way, J. B., Williams, C. and Viereck, L., 1994. Radar estimates of aboveground biomass in boreal forests of interior alaska. IEEE Trans. *Geosci. Remost Sensing* 32, pp. 1117–1124.

Ulaby, F. T., Sarabandi, K., McDonald, K., Whitt, M. and Dobson, M., 1990. Michigan microwave canopy scattering model. *International Journal of Remote Sensing* 11, pp. 1223–1253.