APPLYING DIFFERENT METHODS FOR PREDICTION OF STAND VOLUME USING SPOT 5 DATA

NGUYEN, Thi Thanh Huong

Dr., Department of Forest resource and Environmental Management (Frem), Faculty of Agriculture and Forestry, University of Tay Nguyen
567 Le Duan, Buon Ma Thuot city, Daklak, Vietnam
Tel: 0084-500-3825553; Fax: 0084-500-3825553
Email: ngth_huong2003@yahoo.com

KEY WORDS: Remote sensing, regression, k-nearest neighbor, regression-kriging

ABSTRACT: This paper describes the potential of applying methods to estimate stand volume using remote sensing imagery for the disturbed forest stands in the Central Highlands of Vietnam. The equation of stand volume was defined in the first process based on field data. The methods of regression, k-Nearest neighbor, and regression-kriging were applied to estimate the stand volume using satellite image. The combination of 4 bands SPOT 5 along with normalized difference vegetation index (NDVI) and principal components (PCs) were used in these methods. Validation using independent data indicates that the regression-kriging is the best method for stand volume predictions, simultaneously, it is confirmed that the presence of all SPOT-5 bands improved the prediction result of stand volume. The results show that the more spectral bands included, the lower the root mean square error (RMSE) obtained for all these method.

1. INTRODUCTION

Quantitative assessment of forests is important at a variety of scales for forest planning and management. Wood volume is of fundamental importance to forestry, and consequently foresters have developed a variety of methods for estimating it. Traditionally, forest variables including stand volume are defined based on field plot sampling which is both ineffective cost and time. It is now relatively easy to estimate the forest variables by combining ground inventory with remote sensing data. They can be estimated by modeling relationship between digital number of image and volume inventoried on the ground. Nonspatial modeling and spatial modeling are common method to estimate the stand volume. A large quantity of literature has appeared to relate the estimation of stand volume using remote sensing data. Most of this literature has applied the regression method with linear or nonlinear regression (Trotter et al., 1997; Cohen et al., 2003; Awaya et al., 2004; Fransson et al., 2004; Rahman, 2004). Recently, kNN has found as a useful tool for forest mapping over a large geographic area using a fine spatial resolution (Tokola et al., 1996; Tomppo et al., 1999, 2001; Holmström and Fransson, 2003; Makela and Pekkarinen, 2004; Kurzer, 2008, Thesler et al., 2008). Geostatistics models have been used in some cases and gave promising results (Tuominen et al., 2003; Wallerman, 2003; Meng et al., 2009).

Although numerous studies of relationships between spectral responses and forest parameters have been conducted during the past several decades, conclusions about these relationships vary, depending on the characteristics of the study areas and the used data. Because of the complex forest stand structure and abundant vegetation species of tropical forests, the remote sensing spectral – forest attributes relation is poorly understood (Lu et al., 2004). However, relatively less attention has been devoted to the moist tropical regions such as Asian forests, this might be due to difficulty in field data collection and complex biophysical environments. Meanwhile, a better understanding of forest stand parameters and spectral relationships is a prerequisite for effectively using appropriate image bands for developing spectral response based estimation models (Lu et al., 2004).

This paper presents research that estimates stand volume in a site belonging to the Central Highlands of Vietnam. The study area is located between in $11^0$59' to $12^0$16' latitude North and $107^0$13' to $107^0$28' longitude East. The size of study area is about 500 square km (20 x 25km). The forest is dominated by evergreen broad-leaved tropical natural forest but disturbed by human over time at different levels. Many of valuable species trees have been selectively logged.

2. DATA SOURCES AND SOFWARE

Different data sources including satellite image, digital data, secondary forest data and sample plots were used under the study. The required satellite imagery product was SPOT 5 that is a multi-spectral optical data which was captured using the High Resolution Geometric (HRG) instrument onboard the satellite. The radiometric resolution is 256 digital levels, the spatial resolution is 10m x10m and the images cover 60km x 60km. The SPOT 5 image was rectified using GCPs and the elevation information was captured from a DEM created from a available 10m contour line GIS shapefile. The SPOT image was projected to UTM 48N, WGS84 to ensure compatibility between images and available digital data. A nearest neighbor resampling method was applied during this process with a pixel spacing of 10m x 10m in order to maintain the integrity of the pixel values.

Field data was used to construct relationship between stand volume and spectral responses. A set of 97 sample plots with approximately area of each plots being 0.1ha were conducted in the field during 2007-2008. Sample coordinates were recorded in the centroid position by GPS Garmin 72. 119 sample trees measured both diameter and height in sectional method was collected for constructing of standing volume equation. This equation was then applied for all trees in all sample plots. For each plot, volume of sample plots and the 9- pixel digital number means of SPOT 5 bands were calculated.

A variety of software was employed in this study following the different requirements of the work. The Erdas Imagine version 9.2 and ENVI software version 4.2 were used for image processing. Meanwhile, ArcGIS were employed for database development, spatial data analysis and producing thematic maps. This study used Microsoft excel and SPSS for statistical analysis. K nearest neighbor (kNN) software and the Gstat package were employed to extract stand volume from remote sensing data. kNN was developed by Stümer (2004), whereas the Gstat package which was designed by Pebesma (2005) played an important role in geostatistic models. Because R is open source software and has a user-friendly interface, the study performed the geostatistics procedure with the Gstat package running in the R environment.

3. METHOD

Stand volume was estimated using three modeling approaches: 1) regression; 2) k-nearest neighbor (kNN); and 3) geostatistics. In regression models, simple and multiple regression with both linear and non linear models were used to estimate the relationships between plot volume and the mean value of the digital number for each image. To evaluate the accuracy of models, the coefficient of determination (R^2) or adjusted coefficient of determination (R^2_{adj}) is usually used. The *kNN* software developed by Stümer (2004) was used in this study. For this

application, two input files are necessary, an 'image file' and a 'field sample file' in ASCII format. The required image data ('image file'), which are necessary for the *k*NN calculations, are converted from the corrected SPOT 5 bands and transformed bands (e.g., PCA, NDVI) into ASCII-files. For the purpose of mapping the whole interest area, the images were split into blocks, the predictions were run separately for each block, then merged into the total volume map of the whole area. Spatial autocorrelation is present when a particular value, or attribute, is found to exhibit some degree of dependence to location. Thus, the measurement of spatial autocorrelation involves the simultaneous consideration of both location and attribute information (Goodchild, 1986). This study performed the geostatistics process by the gstat package in R environment. Although there are several packages in R to do geostatistical analysis and mapping, R+gstat is recognized as the only complete and fully-operational package, especially when regression-kriging, multivariate analysis, geostatistical simulations and block predictions are desired (Hengl 2007; Rossiter 2007). In this case, univariate kriging and multivariable kriging are applied for volume prediction with predictors are four bands SPOT 5, three PCA band and a NDVI image.
For every trial, accuracy of the predicted volume was evaluated using the root mean square error (RMSE) and mean absolute error (MAE).

4. RESULT

4.1 Regression estimator

The correlation analysis result in table 5.1 provides a summary of the relationship between individual bands and stand volume. Most bands had a primarily negative correlation with wood volume except for the near infrared band.

Table 4.1 Pearson correlation matrix for the variables analyzed for stand volume estimation

	Volume(m^3)	Band 1	Band 2	Band 3	Band 4	NDVI	PC1	PC2	PC3
V(m^3)	1								
Band 1	-0.227*	1							
Band 2	-0.439**	0.867**	1						
Band 3	0.296**	0.250*	-0.099^{ns}	1					
Band 4	-0.354**	0.700**	0.612**	0.422**	1				
NDVI	0.504**	-0.372**	-0.692**	0.756**	-0.115^{ns}	1			
PC1	-0.559**	0.672**	0.888**	-0.449**	0.526**	-0.891**	1		
PC2	-0.118^{ns}	-0.483**	-0.169^{ns}	-0.938**	-0.682**	-0.546**	0.156^{ns}	1	
PC3	0.413**	-0.026^{ns}	-0.101^{ns}	0.070^{ns}	-0.629**	0.149^{ns}	-0.340**	0.103^{ns}	1

*, ** indicates 1% and 5% significant level, respectively and ns means insignificant.

The study examined linear as well as a number of non-linear models to find an optimal model for predicting stand volume from a satellite image. A combination of 4 band SPOT 5, along with normalized difference vegetation index (NDVI) and three band principle components (PCs) was employed in regression. The evaluation of the candidate regression models obtained from the analysis was based on the value of the adjusted coefficient of determination R^2_{adj} with statistical significance. In most cases, nonlinear regression exhibited the relative higher relationship compared to linear regression models. Of which the best result of the estimation of forest stand volumes by the multiple regression method was achieved using combination 4 SPOT bands. A comparatively high correlation coefficient (R^2 = 0.54) was found in this relationship. A map produced from this model shown in figure 4.2a

- $V = \exp(13.48 + 116139*(1/B12)/\log(1/B3) + 388820*(1/B24)/(1/B3) - 0.073*B4 - 3.3995E9*(1/B24)/B4)$ (Eq. 4.1)

  $R^2_{adj} = 0.54$ ($r = 0.73$) $P=0.000$

- $V = \exp(2.23 + 617.52*1/(PCA1)/\log(PCA3) - 16.24*\mathrm{sqrt}(PCA2)/\log(PCA3) + 3.36*\mathrm{sqrt}(PCA2))$ (Eq. 4.2)

  $R^2_{adj} = 0.48$ ($r = 0.69$) $P<0.000$

- $V = \exp(7.58 + 2.34*\log(NDVI))$ (Eq. 4.3)

  $R^2_{adj} = 0.31$ ($r = 0.56$) $P<0.000$

4.2 KNN method

The current study calculated the stand volume using the kNN method. The volume map made by using the kNN method is presented in figure 4.2b.

4.3 Geostatistical method with regression kriging

Based on an examination of the variogram cloud, experimental semivariograms were computed. Thereafter, different types of semivariogram models were used to fit the points including Spherical, Exponential, Gaussian, Circular, Bessen and Pentaspherical models. The current study found the Spherical model as the best fit. Hence, it was selected as the theoretical model applied for spatial predictions. The fit of the spherical model had a nugget of 0.30, a sill 0.075, and a range of 700 (figure 4.1). The volume map was created by regression kriging shown in Figure 4.2c.

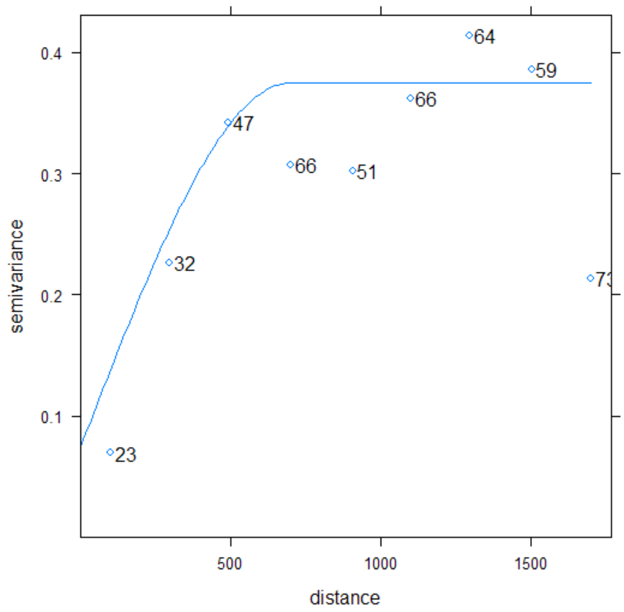

Figure 4.1 Experimental variogram with fitted model, log (volume)

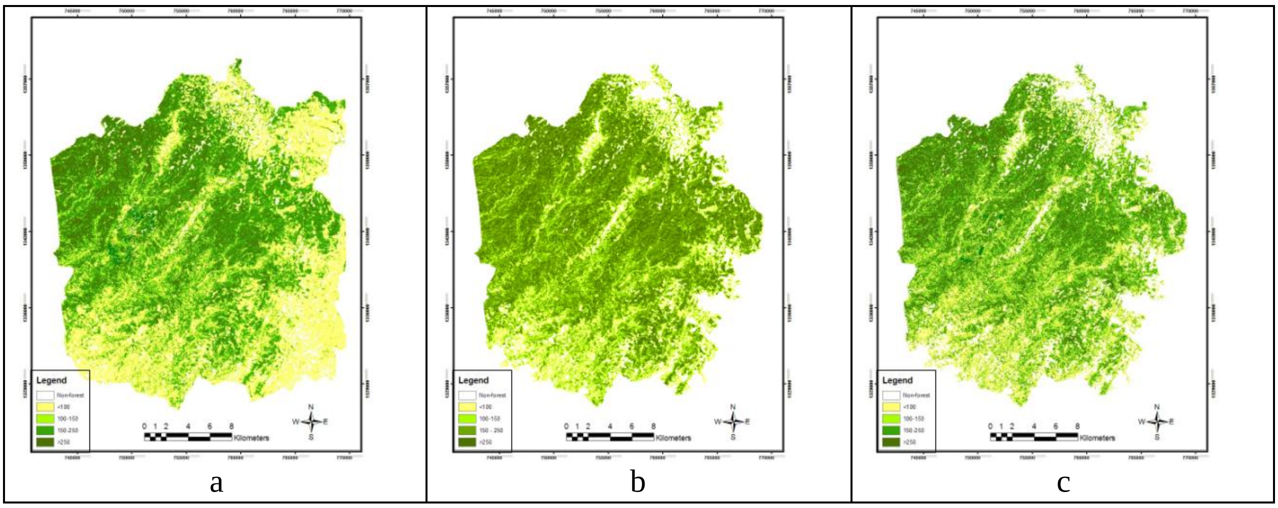

Figure 4.2 Volume map using SPOT 5 image and regression (a), kNN (b), and regression-kriging (c) method

4.4 Accuracy assessment of the models

The accuracy of the estimates at the sample plot level was expressed in terms of the RMSE, relative RMSE, mean bias and relative mean bias. Regression-kriging provided the smallest RMSE and mean bias. Multiple regression followed. The predicted volume using the NDVI resulted in the worst outcome for both regression – kriging and the kNN estimator. Within RK method, RK 4bands combination obtained the highest accuracy. Contrary to the RK 4 bands combination, the highest relative RMSE was observed when using RK- NDVI.

Table 4.2 Error of the volume estimates using three different methods

	RMSE (m3 ha -1)	RMSE (%)	MAE (m3 ha -1)	MAE (%)
Regression	77.54	40.38	47.2	23.61
Regression - Kriging 4bands	73.18	35.44	43.55	21.09
Regression – Kriging PCA	91.86	47.33	57.97	29.87
Regression – Kriging NDVI	143.23	90.09	95.86	60.3
kNN - 4bands	120.46	65.03	71.43	38.56
kNN – PCA	135.66	75.74	87.41	48.8
kNN- NDVI	148.4	77.28	97.25	50.64

5. DISCUSSION

Multiple regression models were used to assess the correlation of more than one spectral band (independent variables) with stand volume (dependent variable). The relationship between NDVI and stand volume was also tested in order to select the optimal model. The results indicated there was a significant relationship between stand volume and the groups of spectral bands, although only a combination of 4bands SPOT considerably improved the coefficient of determination. It can be explained that all SPOT bands which were used to predict stand volume appear to provide more information than individual bands or transformed bands. A positive relationship was found between NDVI and the stand volume (table 4.1) and coefficient of determination which was derived from the equation above (eq. 4.3) was low (R^2 =0.31). NDVI is widely used for many vegetation studies, but had the lowest performance for predicting stand volume in this case. This can be explained because the natural dynamic of forest conditions limits the ability to predict from a single information source, such as NDVI. Although according to Beusch et al. (2005), as the NDVI is derived from image spectral values, any forest structural variability will be related to the computed NDVI values. For example, factors such as differences in crown closure, shadows or stand density may result in markedly different stand structures, yet will be still represented by the same NDVI (Wulder et al., 1998).

Once again, it is confirmed that the presence of all SPOT-5 bands improved the prediction result of stand volume (see table 5.1). The more spectral bands included, the lower the RMSE obtained. Regardless the used method, it is clearly evident that the volume produced from RK using four bands SPOT had a lower RMSE than other RKs (PCs, NDVI). A similar outcome also occured with kNN estimators. In most cases, NDVI based on stand volume predictions gave poor results. Regression kriging was powerful compared to the others. This indicates that incorporation of spatial information in remote sensing based predictions improved the accuracy, when applied to stand volume.

6. CONCLUSION

Estimation of forest structure e.g. stand volume using remote sensing has been broadly investigated both in study and in practice with different methods. However, most work was focused on the boreal or temperate forest where the forest stand structure and species composition are comparatively simple. Relatively less attention has been devoted to moist tropical regions, and hardly has been made consideration of the forests in Vietnam. The findings of the relationship between stand volume and spectral reflectance in this study may be applicable potentially in complex structures such as Vietnamese natural forest stands. Although the regression model seems to be a potential method to extract stand volume using remote sensing, concern of spatial dependence within variables e.g. the regression-kriging algorithm improved the accuracy of the predicted result.

ACKNOWGEMENT

This research is part of my doctoral thesis. It was performed in Freiburg University, Germany. I would like to thank my supervisor I would like to express my deep appreciation to Professor Dr. Barbara Koch and Dr. Claus-Peter Gross of the Department of the Remote Sensing Landscape Information Systems (Felis) of the Albert Luwig University Freiburg for their kind supervision, guidance and support. I also wish to thank my colleagues from Felis for their technical support.

References

Awaya, Y., Tsuyuki, S., Kodani, E. and Takao, G., 2004. Potential of woody carbon stock estimation using high spatial resolution imagery: A case study of spruce stands. Pp. 425-440. In: Shiyomi, M. et al. (Eds). Global environmental change in the ocean and on land. TERRAPUB.

Beusch, M., Koetz, B., Kneubühler, M., and Itten, K., 2005. Forest canopy structure derived from spatial and spectral high resolution remote sensing data. www.geo.uZh.ch (accessed January 2008).

Cohen, W.B., Maiersperger, T.K., Gower, S.T., Turner, D.P., 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. Remote Sensing of Environment 84 (4), 561- 571.

Fransson J. E.S., Magnusson, M., Holmgren, J., 2004. Estimation of Forest Stem Volume uing optical SPOT-5 satellite and laser data in Combination. IEEE Transactions on Geoscience and Remote Sensing: 2318-2322.

Fransson J. E.S., Magnusson, M., Holmgren, J., 2004. Estimation of Forest Stem Volume uing optical SPOT-5 satellite and laser data in Combination. IEEE Transactions on Geoscience and Remote Sensing: 2318-2322.

Goodchild, M. 1986. Spatial autocorrelation. Concepts and Techniques in Modern Geography 47. Geo-Books. Norwich.

Hengl, T., 2007. A practical guide geostatistics mapping of environment variable. JRC European commission. 165p.

Holmström, H. and Fransson, J.E.S., 2001. Combining remotely sensed optical and radar data in kNN estimation of forest variables. Forest Science, 49(3):409–418.

Kutzer, C., 2008. Doctoral thesis. Potential of the kNN Method for estimation and Monitoring off-Reserve Forest Resources in Ghana. Faculty of Forest and Environmental Sciences, Albert-Ludwigs-Universität Freiburg im Breisgau, Germany. 130p.

Lu, D., Mausel, P., Brondízio, E., and Moran, E., 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. Forest Ecology and Management, 198: 149–167.

Mäkelä, H., Pekkarinen, A., 2004. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data, Forest ecology and management, Volume 196, issues 2-3: 245 – 255.

Meng, Q., Cieszewski, C., and Madden, M., 2009. Large area forest inventory using Landsat ETM+: A geostatistical approach. ISPRS Journal of Photogrammetry and Remote Sensing, 64: 27-36.
Pebesma, E.J., 2005. The Gstat Package. http://cran.r-project.org/doc/packages/gstat.pdf (accessed November 2008).
Rahman, M.M., 2004. Estimating carbon pool and carbon release due to tropical deforestation using high resolution Satellite Data. Doctoral thesis. Faculty of Forest, Geo and Hydro Sciences, Dresden University of Technology, Germany.191p.
Rossiter, D. G., 2007. Technical Note: Co-kriging with the gstat package of the R environment for statistical computing, 2nd edition. International Institute for Geo-information Science and Earth Observation (ITC), Enschede, Netherlands.
Stümer, W., 2004. Kombination von terrestrischen Aufnahmen und Fernerkundungsdaten mit Hilfe der kNN methode zur Klassifizierung und Kartierung von Wäldern. Dissertation, Fakultät für Forst- Geo- und Hydrowissenschaften, Technischen Universität Dresden, 153p.
Thessler, S., Sesnie, S., Bendaña, Z.S.R., Ruokolainen, K., Tomppo, E., and Finegan, B., 2008. Using k-nn and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica. Remote Sensing of Environment, 112: 2485–2494.
Tokola, T., Pitkanen, J., Partinen, S. and Muinonen, E., 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. International Journal of Remote Sensing, 17(12): 2333:2351.
Tomppo, E., Goulding, C., and Katila, M., 1999. Adapting Finnish multi-source forest inventory techniques to the New Zealand preharvest inventory. Scandinavian Journal of Forest Research, 14: 182−192.
Tomppo, E., Korhonen, K. T., Heikkinen, J., & Yli-Kojola, H., 2001. Multisource inventory of the forests of the Hebei Forestry Bureau, Heilongjiang, China. Silva Fennica, 35: 309−328.
Trotter, C.M., Dymond, J.R. and Goulding, C.J., 1997. Estimation of timber volume in a coniferous plantation forest using Landsat TM. International Journal of Remote Sensing, 18(10): 2209-2223.
Tuominen, S., Fish, S. and Poso S., 2003. Combining remote sensing, data from earlier inventories and geostatistical interpolation in multi source forest inventory. Canadian Journal of Forest Research, 33: 624-634.
Wallerman, J., 2003. Remote sensing aided spatial prediction of forest stem volume . Doctoral thesis. Swedish University of Agricultural Sciences, Umeå. 42p.
Wulder, M. A., Le Drew E. F., Franklin S. E., and Lavigne M. B., 1998. Aerial image texture information in the estimation of northern deciduous and mixed wood forest leaf area index (LAI). Remote Sensing of Environment, 64: pp. 64-76.