AGE ESTIMATION TECHNIQUE OF INDUSTRIALIZED TIMBER PLANTATION USING VARIOUS REMOTE SENSING DATA

Jong-Hong Kim, Joon Heo, Ji-Sang Park

School of Civil and Environmental Engineering, Yonsei University, South Korea
kkdevil@hanmail.net, jheo@yonsei.ac.kr, utheman05@gmail.com

ABSTRACT Timber stand age information of timber in industrialized plantation forest is generally collected by field surveying which is labor-intensive, time-consuming, and very costly. It is also inconsistent in analyses perspective. As an alternative, The objective of this research is to present a practical solution for estimating timber age of loblolly pine plantation using Landsat thematic mapper (TM) images, shuttle radar topography mission (SRTM), and national elevation dataset (NED). A multivariate regression model was developed based upon satellite image-based information (i.e.normalized difference vegetation index (NDVI), tasseled cap (TC) transformation, and derived tree heights). A residual studentized technique was applied to remove potential outliers. After that, a refined age estimation model with a correlation coefficient R-square of 84.6% was obtained. Finally, the feasibility test of estimated model was performed by comparing estimated and measured stand ages of timber plantations using test datasets of plantation stands (2,032 stands). The result shows that the proposed method of this study can estimate loblolly pine stand age within an error of 2~3 years in an effective and consistent way in terms of time and cost.

KEY WORDS: Stand Age Estimation, Tasseled Cap Transformation, Linear Regression, SRTM, Studentized Residual,

1. INTRODUCTION

Major companies which are related to trees usually own their industrialized tree plantation site for mass production and keep their eyes on monitoring the overall production of timber, by-products, site information and management. The importance of acquiring and analyzing information of stands in plantation site is significantly increasing. For applying appropriate and in-time management activities, a set of forest stand information (e.g., height, age, stand density, canopy closure, leaf area index, etc) needs to be collected and updated periodically based upon forest growths, which means that the management and update of the information is the key-issue recently (Dodge and Bryant 1976; Sivanpillai et al. 2006). Traditionally, forest information like stand ages has often been collected by interpreting aerial photographs or field surveying (Niklasson 2002; Worbes et al. 2003). However, it wastes lots of time and it has a high cost. Therefore, various research has been conducted to find alternative solutions for acquiring forest stand information using different data sources. Among them, satellite imagery has emerged as a useful resource for characterizing key forest stand characteristics (Sivanpillai et al. 2006). A lot of research has been done about extracting information from satellite imagery.(Franklin, 1986; Gemmell, 1995; Steininger 2000; Tian et al. 2002; Wulder et al. 2004; Berterretche et al. 2005, Sivanpillai et al. 2006) However, it is hard to generate an exact mathematical model of information about trees from satellite imagery and the application of satellite images have not been tested for retrieving forest stand information in industrialized plantations (Sivanpillai et al. 2006). The objective of this research is to derive multi-derivative regression model to estimate the stand ages using various kinds of remote sensing data. The initial regression model is refined by removing outliers of estimated stand age using residual studentization technique. Finally, the feasibility of the model is evaluated by comparing estimated and measured stand ages.

2. DATASET AND PROJECT AREA

2.1 Dataset

Figure 1 illustrates the project area which includes industrialized timber plantations of loblolly pine.

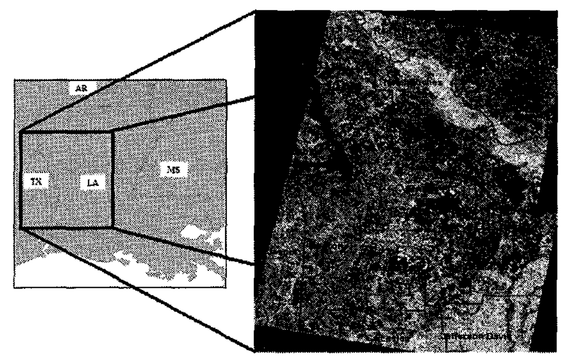

Figure 1. Project Area

The area of this area is 2,000km^2 and covers western Louisiana and eastern Texas. Four datasets (i.e., Landsat TM images, forest inventory data which is used for ground truth, SRTM, NED) were used in this study. Three Landsat TM images were acquired on different years (1986, 2000, 2004) and used for creating vegetation index images (NDVI, Tasselcap Transformation). Forest inventory data which is composed of polygon provides a variety of information associated with timber management (e.g., stand location, area, perimeter, site index, stand age, timber type, and chronicle of silvicultural activity). Forest inventory data are managed by timber companies for asset management of site. Theses attributes and spatial data should be updated by asset managers and field foresters regularly since latest information of forest is very critical factor for management operations. In this study, loblolly pine plantation stands established between the years of 1977 and 2004 were used.

SRTM "Finished" dataset is composed of points of 1 arc second with respect to latitudes and longitudes (i.e., approximately 30 meters), and represents heights of grounds and features.

NED represents the elevation of the "bare earth" at 30m resolution. NED is generated by merging elevation data across the Unites States into a seamless raster format (USGS 2005). In other words, NED is the 1:24,000-scale digital elevation model (DEM) data for the conterminous U.S. provided by USGS.

2.2 Data Processing

The NDVI is the typical vegetation index and the well-known equation is;

$$NDVI = \frac{IR - R}{IR + R} \tag{1}$$

where, IR : infra red band
R: red band

TC is a radiometric-rotational transformation which emphasizes soils, vegetations, and wet areas in several layers. Among three layers, layers of greenness index and wetness index were included for regression modeling due to their proven capacity for estimating ages of trees (Cohen et al. 1995). And the difference image between SRTM and NED contains the information of estimated interferometric tree height. Lastly, from the timber inventory database, loblolly pine stands were selected to reduce the effects of uneven structural and environmental conditions. Among the pine stands, only stands whose area is greater than 15*ha* are selected to improve tree height stimation accuracy during the model development. Figure 2 represents the flow chart of this study.

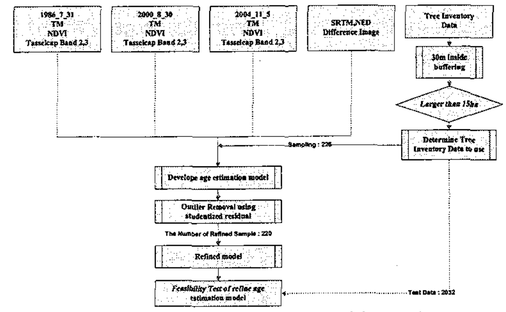

Figure 2. The flow chart of this study.

2.3 Model Development and Refinement

A multivariate regression model for estimating stand ages of loblolly pine plantations was developed. The *Year* information was used as a dependent variable, and NDVI, TC, and estimated interferometric heights from difference image were used as independent variables. One percent of total stands data (i.e., 226 stands) were selected among 2,258 stands data randomly, and an age estimation model was developed to determine inter-relationships among variables. The value of adjusted R_2 is 76.0%. The equation used for multi-derivative linear regression is

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \cdot & \cdot & x_{1,10} & x_{1,11} \\ x_{2,1} & x_{2,2} & \cdot & \cdot & x_{2,10} & x_{2,11} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{n-1,1} & x_{n-1,2} & \cdot & \cdot & x_{n-1,10} & x_{n-1,11} \\ x_{n,1} & x_{n,2} & \cdot & \cdot & x_{n,10} & x_{n,11} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_{10} \\ a_{11} \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_{n-1} \\ Y_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_{n-1} \\ v_n \end{bmatrix}$$

$$\mathrm{AX} = \mathrm{Y} + \mathrm{v} \tag{2}$$

A matrix represents for the pixel values that are from NDVI, Tasselcap Transform, and the difference image. X matrix stands for the coefficients of each independent variables and y-intersection. Y is the stand ages from the tree inventory as ground truth. V matrix represents residuals. Using least square adjustment, the coefficient matrix X is estimated as follows.

$$\hat{X} = (A^T A)^{-1} A^T Y \tag{3}$$

And residual v is as below.

$$\hat{v} = Y - \hat{Y} = Y - A\hat{X} = (I - A(A^T A)^{-1} A^T)Y = (I - H)Y \tag{4}$$

Where, I is an identify matrix, and H is a hat matrix. The variance of i-th residual is,

$$\mathrm{var}(\hat{v}) = (1 - h_{ii})\sigma^2 \tag{5}$$

where, h_{ii} is the i-th diagonal element of the hat matrix. The variances of residuals are estimates of errors based upon estimations. This explains that simple comparison of magnitudes of residuals is not appropriate for detecting outliers, since residuals cannot be independent and do not all have the same variance (Park 1992). Therefore, residuals need to be adjusted using their estimated standard deviations. The standardized *i*-th residual is,

$$r_i = \frac{v_i}{s\sqrt{1-h_{ii}}} \tag{5}$$

where, $s = \sqrt{\frac{ee^T}{n-k}}$

k = number of unknowns

There is another criterion which is the approximate distribution of the standardized residual by Lund (1975) and studentized residual by Weisberg(1980). Any observations, which have displacement of three times of RMSE, were considered as outliers to be removed. This process was iteratively continued until no outlier is detected. After three iterations, six of 226 observations were removed as outliers.

3. EXPERIMENT AND RESULTS

3.1 Development of Age Estimation Model.

A refined estimation model with 84.6% adjusted-R^2 value using remained 220 observations. Table 1 represents the coefficients of each variables and R^2 value of refined model.

Y_intersection	2000.52909	R^2	0.8534
SRTM	-0.4574261	**Adjusted** R^2	0.8464
NDVI_2000	-28.295652		
NDVI_2004	-12.601424		
NDVI_1986	1.56775996		
TC3_1986	0.08104753		
TC3_2000	-0.0833383		
TC3_2004	-0.5735998		
TC2_1986	-0.162229		
TC2_2000	0.33725868		
TC2_2004	0.93075851		

Table 1. The properties of refined model.

3.2 Evaluate the feasibility of estimation model.

The RMSE of the residual of each observation is used as the criteria. The residuals of stand age estimation are calculated as the difference between stand ages estimated using the regression model and true stand ages in forest inventory databases. The developed estimation model is applied to 2,032 observations for evaluation. Table 2 shows that the result of the experiment.

Case	Time Period	RMSE
1	1977 ~ 2004	3.034
2	1982 ~ 2002	2.416
3	1982 ~ 2004	2.647

Table 2. RMSE and time period of each case

There are 3 cases to evaluate the feasibility. According to the Table 2, the estimated regression model is able to provide stand age estimations with 3.034 year time difference for the whole time period (case 1). Moreover, the age estimations accuracy is increased by using shorter time periods, such as the RMSE is 2.416 years in 1982 ~ 2002(Case 2) And the case 3 has RMSE which is 2.647 years in 1982 ~ 2004, lastly. This result happens due to absence of available imagery before the year of 1986. The oldest imagery available for this study was taken in 1986 which means that there's no information about stands before 1986. If the Landsat MSS imagery is available (i.e, before 1984), then the estimation accuracy will be significantly improved further even in early years (i.e., before the year of 1982) with a further comprehensive regression model.

Table 3 illustrates the residual distribution of total 2,032 residuals.

Year	Average	RMSE	No. of data	σ~-σ	2σ~-2σ	3σ~-3σ	out of 3σ
1977	9.184	9.630	16	0	0	0	16
1978	6.963	7.281	28	0	1	2	26
1979	5.879	6.268	37	0	3	22	15
1980	3.727	4.328	19	4	11	16	3
1981	3.121	3.837	79	26	59	75	4
1982	0.875	2.174	83	57	74	82	1
1983	0.925	2.534	105	67	100	103	2
1984	0.530	2.204	117	84	112	115	2
1985	-0.063	2.048	139	92	134	138	1
1986	-0.662	2.019	128	84	123	127	1
1987	-0.029	2.296	113	79	106	113	0
1988	-0.058	1.933	135	92	129	135	0
1989	-0.208	1.811	104	72	100	104	0
1990	-0.511	1.690	107	68	103	107	0
1991	-0.593	2.053	85	61	80	84	1
1992	-1.989	2.948	65	26	56	64	1
1993	-1.304	2.547	51	30	45	50	1
1994	-2.084	2.865	54	23	46	52	2
1995	-1.984	2.608	45	22	37	43	2
1996	-2.208	2.939	51	19	40	50	1
1997	-1.846	2.705	41	21	34	41	0
1998	-1.688	2.844	32	17	30	32	0
1999	-2.502	3.196	28	14	23	26	2
2000	-0.887	2.864	27	18	27	27	0
2001	-0.536	2.897	59	44	54	59	0
2002	-2.811	4.147	70	32	62	70	0
2003	-3.142	3.845	106	33	78	100	6
2004	-3.345	4.172	108	43	80	102	6

Most residual of total observations are within the range of $-3\delta \sim 3\delta$ except the observations from 1977~1980. As a result, the age estimation model created from 220 observations could say that the stand age at least within 3 years.

4. CONCLUSION AND FUTURE WORKS

4.1 Conclusion

In this study, the stand age estimation model is created by linear regression using various kinds of remote sensing data (SRTM, NED, Landsat images) and tree inventory data. Initial estimation model is developed from 226 sample data and 6 of them is removed as outliers. This refined model is applied to 2,032 data to evaluate the feasibility. The RMSE value of residual which is the difference between ground truth and estimated year is 3.034 (1977 ~ 2004), 2.416 (1982 ~ 2002), 2.647 (1982 ~ 2004). In short, this model can estimate the stand age around 3 years or better.

4.2 Future Works

Firstly, the selection of variables should be done. Instead of using all 10 variables, we must select the variables which are highly correlated with the developed model and in addition try to add landsat images of 70's during model development. At last, the effect caused by terrain features should be considered and result has to be compared and analyzed.

REFERENCE

Berterretche, M., HUdak, A.T., Cohen, W.B., Maiersperger, T.K., Gower, S.T., and Dungan, J. ,2005, "Comparison of Regression and Geostatistical Methods for Mapping Leaf Area Index (LAI) with Landsat ETM+ Data Over a Boreal Forest," *Remote Sensing of Environment*, 96:49-61.

Cohen, W.B., Spies, T.A., and Fiorella, M, 1995, "Estimating the Age and Structure of Forests in Multi-Ownership Landscape of Western Oregon, U.S.A.," *International Journal of Remote Sensing*, 16:721-746.

Dodge, A.G. and Bryant, E.S. ,1976, "Forest Type Mapping with Satellite Data," *Journal of Forestry*, 74:526-531.

Franklin, J. , 1986 , "Thematic Mapper Analysis of Coniferous Forest Structure and Composition," *International Journal of Remote Sensing*, 10:1287-1301.

Gemmell, F.M. ,1995, "Effects of Forest Cover, Terrain, and Scale on Timber Volume Estimation with Thematic Mapper Data in the Rocky Mountain Site," *Remote Sensing of Environment,* 51:291-305.

Lund, H.G. and Thomas, C.E., 1989, "A Primer on Stand and Forest Inventory Designs," General Technical Report WO-54, USDA Forest Service, Washington D.C.

Niklasson, M., 2002, "A Comparison of Three Age Determination Methods for Suppressed Norway Spruce: Implications for Age Structure Analysis," *Forest Ecology and Management*, 161:279-288.

Park, S.H. ,1992, *Regression Analysis*, Min-Young-Sa, Seoul, Korea, 654p.

Sivanpillai, R., Smith, C.T., Srinivasan, R., Messina, M.G., and Wu, X.B. ,2006, "Estimation of Managed Loblolly Pine Stand Age and Density with Landsat ETM+ Data," *Forest Ecology and Management*, 223:247-254.

Steininger, M.K. ,2000, "Satellite Estimation of Tropical Secondary Forest Above-Ground Biomass: Data from Brazil and Bolivia," *International Journal of Remote Sensing*, 21:1139-1157.

Tian, T., Wang, Y., Zhang, Y., Knyazikhinm Y., Bogaert, J., and Myneni, R.B. ,2002, "Radiative Transfer Based Scaling of LAI/FPAR Retrievals from Reflectance Data of Different Resolutions," *Remote Sensing of Environment*, 84:143-159.

U.S. Geological Survey (USGS), 2005, "The USGS National Elevation Dataset (NED)," URL: http://ned.usgs.gov.

Worbes, M. Staschel, R., Roloff, A., and Junk, W.J. , 2003, "Tree Ring Analysis Reveals Age Structure, Dynamics and Wood Production of a Natural Forest Stand in Cameron," *Forest Ecology and Management*, 173:105-123.

Wulder, M.A., Skakun, R.S., Kurz, W.A., White, J.C. , 2004, "Estimating Time Since Forest Harvest Using Segmented Landsat ETM+ Imagery," *Remote Sensing of Environment*, 93:179-187.