The Prediction of Leaf Area Index from Forest Polygons Decomposed Through the Integration of Remote Sensing, GIS, UNIX, and C

M. A. Wulder[1]

Pre-print of published version.

Reference:
Wulder, M., 1998; The prediction of leaf area index from forest polygons decomposed through the integration of remote sensing, GIS, UNIX, and C, Computers & Geosciences, Vol. 24, No. 2, pp. 151-157.

DOI.
http://dx.doi.org/10.1016/S0098-3004(97)00132-5

Current info

~~Address correspondence to:~~
~~M. A. Wulder,~~
~~Waterloo Laboratory for Earth Observations,~~
~~Department of Geography, University of Waterloo,~~
~~Waterloo, Ontario, N2L 3G1, Canada~~

~~Fax: (519) 888-6768~~
~~Email: mwulder@glenn.uwaterloo.ca~~

Natural Resources Canada / Ressources naturelles Canada

Mike Wulder, Ph.D.
Research Scientist / Chercheur scientifique

Canadian Forest Service
Pacific Forestry Centre
506 West Burnside Road
Victoria, BC V8Z 1M5

Service canadien des forêts
Centre de foresterie du Pacifique
506, chemin Burnside ouest
Victoria (C.-B.) V8Z 1M5

Tel: (250) 363-6090 Fax: (250) 363-0775
E-mail: Mike.Wulder@nrcan.gc.ca
http://cfs.nrcan.gc.ca/directory/mwulder

Canada

Abstract

Forest stand data is normally stored in a geographic information system (GIS) on the basis of areas of similar species combinations. Polygons are created based upon species assemblages and given labels relating the percentage of areal coverage by each significant species type within the specified area. As a result, estimation of leaf area index (LAI) from the digital numbers found within GIS stored polygons lack accuracy as the predictive equations for LAI are normally developed for individual species, not species assemblages. A Landsat TM image was acquired to enable a classification which allows for the decomposition of forest stand polygons into greater species detail. Knowledge of the actual internal composition of the stand polygons provides for computation of LAI values based upon the appropriate predictive equation resulting in higher accuracy of these estimates. To accomplish this goal it was necessary to extract, for each cover type in each polygon, descriptive values to represent the digital numbers located in that portion of the polygon. The classified image dictates the species composition of the various portions of the polygon and within these areas the raster pixel values are tabulated and averaged. Due to a lack of existing software tools to assess the raster values found within GIS polygons a combination of remote sensing, GIS, UNIX, and specifically coded C programs were necessary. Such tools are frequently used by the spatial analyst and indicate the complexity of what may appear to be a straight forward spatial analysis problem.

Key Words: Raster, Vector, Leaf Area Index, Net Primary Productivity, Normalized Difference Vegetation Index, Texture, Image Classification

Introduction

Leaf Area Index (LAI), the leaf area per unit ground area, a dimensionless index, is an important structural attribute of forest ecosystems because of its potential to be a measure of energy, gas, and water exchanges. Accordingly, estimation of LAI from remote sensing instruments is of wide interest and significance, prompting Running, *et al.*, (1986), to comment that LAI is "the single variable which may be derived from remote platforms that is of greatest importance for quantifying energy and mass exchange by plant canopies over landscapes." Physiological processes such as photosynthesis, transpiration, and evapotranspiration are related to LAI (Pierce and Running, 1988). Furthermore, understanding the relation between spectral response and LAI at the forest level may allow for the future determination of LAI by remote sensing at a global level, facilitating calculation and global modeling of canopy photosynthesis and evapotranspiration in a synoptic and repeatable fashion (Bonan, 1993; Gutman, 1991; Spanner, *et al.*, 1990).

The production of dry matter is related to the capacity of trees to synthesize carbohydrates; accordingly, the growth of individual trees is often related to crown size. It follows that dominant trees with large fully exposed crowns produce more dry matter than suppressed trees with small shade crowns. Leaf biomass, or LAI, is directly related to the productivity of forest stands (Running and Hunt, 1994). The LAI of temperate zone forests generally varies in the range from 1 to 20, as a function of stand type, age, and site conditions such as water supply and soil fertility. In general, deciduous species have an LAI from approximately 3 to 6, with coniferous trees reaching up to 20 (Waring and Schlesinger, 1985; Kozlowski, *et al.*, 1991). Accurate estimates of LAI are required from the remotely sensed imagery so that they may be applied to compute an actual NPP measurement that can be compared to the NPP value produced in a biogeochemical model, such as BIOME-BGC (Running and Hunt, 1994; Franklin, et al., 1997a; Wulder, et al., 1996a). If the modeled values are close to the values predicted from the remotely sensed imagery, the stand may be assumed to be in a normal condition, yet if the stand is found to be below the remotely sensed estimate of NPP there may be a need for some form of silvicultural treatment. These analyses are a component of forest productivity monitoring and assessment studies undertaken by the Canadian Forest Service, Maritime Region, based in Fredericton, New Brunswick.

Regression relationships are developed between field measured LAI and remotely sensed vegetation indices. The method applied in this study to obtain initial control estimates is based upon destructive samples and pipe model theory (Kaufmann and Troendle, 1981; Lavigne, et al., 1996). These predictive equations for LAI are species specific as the slope value relates the degree to which the vegetation index is increased to generate the LAI value. As a result, to achieve the highest precision in the estimation of LAI each major surface cover has an empirical LAI predictive equation associated with it. Forest stand data stored in a GIS is normally associated with a polygon based on a species assemblage, such as 60% white spruce and 40% aspen. If a vegetation index value is computed to represent the polygon it will not be representative of the species composition and there is also a question of which predictive equation to apply. Data storage in a *vector* GIS, such as ARC/INFO is accomplished through the creation of labeled polygons which represent a particular geographic area. The polygon label enables the relational data base properties of the GIS. Remotely sensed imagery is stored in a *raster* format, which is a grid of regularly sized and shaped areas, 30 metres squared in the case of Landsat TM (Lillesand, and Kiefer, 1994). Input to the predictive equations for LAI requires a normalized difference vegetation index (NDVI) specific to each cover type, which will entail extraction of a mean NDVI for each cover type found within each polygon. Texture has also been demonstrated as a valuable independent variable in the modeled prediction of LAI (Wulder, et al., 1996b).

Methods

The method utilized to extract cover type information from forest stand polygons has been demonstrated to result in varying estimates of LAI (Franklin, et al., 1997b). The extraction of intra-polygon cover type information may allow for the ability to increase the reliability of estimates of LAI from GIS polygon data through the addition of a remotely sensed image classification and image spectral and textural information. The project data flow and extracted variables are demonstrated in Figure 1. To maximize the potential utility of the data a variety of values will be extracted from the image within each class of each polygon, such as spectral band information, vegetation indices, textural information, pixel count per polygon, and pixel count per class per polygon. An analysis of spectral/forest relationships was undertaken to develop regression models between LAI, aerial remote sensing data, and the TM-derived NDVI values at 17 intensively sampled plot locations (Wulder, et al., 1997a). These regression equations will be utilized to test for an increase in precision of LAI estimates with the refined extraction process. Each cover type represented in a stand will have an LAI value computed which is then multiplied by the proportion of the polygon occupied by that cover type and summed for an LAI value. The predictive equations of LAI from NDVI applied in this study are,

Hardwood LAI = (2.99 * NDVI) + 5.34
Softwood LAI = (-5.05 * NDVI) + 11.44
Mixedwood LAI = (2.47 * NDVI) + 5.94

The LAI values computed for the intra-polygon data are then compared to LAI values derived from mean polygon values.

Study Area, Field Data, and GIS Data

The Fundy Model Forest (FMF) is a 420,000 hectare working forest in southeast New Brunswick, Canada. The model forest is located in the Acadian forest region and is composed of a variety of broadleaf deciduous and coniferous species and includes a wide range of forest conditions (Rowe, 1972), with stand structures ranging from regeneration to old growth. The Acadian forest region is characterized by a wide variety of forest species. Coniferous tree species are predominantly jack pine, white spruce, and balsam fir, and also include white pine and red spruce. The predominant deciduous species are red maple and white birch, with stands also including beech, striped maple, trembling aspen, long tooth aspen, sugar maple, yellow birch, and grey birch (Power and Matson, 1995). The main focus of the study area was centered near Sussex at 45.43.00˚ North latitude and 65.31.00˚ West longitude.

The field data available consisted of:

- a GIS database for the Fundy Model Forest,
- a general timber cruise consisting of data on species, density, age, height, and site class for a sample of 128 stands, and
- detailed plot-level information at 17 locations where destructive sampling was undertaken, to derive allometric equations between LAI and sapwood cross-sectional area, for commercially important species located in the study area (Lavigne, et al., 1996; Wulder, 1996).

The forest cover types were derived from airphoto interpretation, work which involved separation of some of the main cover types into species assemblages, such as stands dominated by tolerant hardwoods, intolerant hardwoods, spruce, pine, and fir, and a wide range of mixed wood combinations and structures.

Remotely Sensed Imagery and Supplemental Data

To enable the integration of the raster and vector data both data sets were geometrically transformed to real world coordinates using UTM projection and NAD 27 datum. The Landsat TM image, which was acquired on August 7, 1992, was geometrically corrected utilizing the EASI/PACE utility of GCP Works. Base maps (1:50,000) and GPS points were used for the ground control, resulting in a RMS accuracy of 0.125 pixels utilizing 25 ground control points and a first order nearest neighbor resampling technique.

Landsat TM spectral data selected for extraction were TM bands 3, 4, and 5. Prior to extraction the data was corrected for atmospheric effects using a dark pixel subtraction technique (Franklin and Giles, 1995). The dark pixel subtraction technique was appropriate due to the presence of two deep water bodies in the original Landsat TM imagery for the calibration of each band. Landsat TM band 3 (0.63 - 0.69 µm) was included for extraction because it senses a strong chlorophyll absorption region and a high reflectance region for most soils, hence it is appropriate for discerning between vegetation and soil. Landsat TM band 4 (0.76 - 0.90 µm) was selected for discriminating between differing vegetation and varieties and conditions. Landsat TM band 5 (1.55 - 1.75 µm) measures changes in leaf-tissue water content (as reflectance is found to decrease as water content increases) which may be related to differences between plant species or vigor (Avery and Berlin, 1992). The mid-infrared spectral region occupied by TM band 5 is also related to the degree of forest canopy closure.

Vegetation Indices and Image Texture

Vegetation indices have been developed to emphasize the difference between the absorption in the visible and reflectance in the infrared through the construction of ratios between multispectral bands (Curran, 1980). The normalized difference vegetation index is a commonly used metric, calculated from the red (R) portion of the visible and near infrared (NIR) radiation, in the form of, NDVI = (NIR-R) / (NIR+R). NDVI has also been demonstrated to assist in the compensation for changing illumination conditions, surface slopes, and viewing aspects (Avery and Berlin, 1992). NDVI captures information relating to the amount of radiation absorbed in the visible (Red) and reflected in the near infrared (NIR) wavelengths by vegetation. Vegetation indices, such as NDVI, may be viewed as a surrogate for scene vegetation content, and may be correlated with physical measures of vegetation, such as LAI.

Texture has been characterized as the spatial variation in tones. This property may be related to statistics which characterize the relationship between neighboring pixels. Such statistics are related to the image properties of an area, this is usually accomplished through a moving window. A user-defined moving window is typically passed over the imagery and statistics are calculated directly from the values reflectance in the window, which are known as first-order texture. Alternatively, second-order texture may be calculated from matrices which characterize the arrangement of the digital numbers within a window (Wulder, et al., 1997b). The variation in texture is related to changes in the spatial distribution of terrestrial vegetation, both in the vertical and horizontal direction. Texture values describe the spatial variation in image tones which are the result of the variation in the above ground organization of forest elements. Texture derivatives are supplemental to the image data, and, accordingly, provide an accessible, low cost, additional data source.

Multispectral Digital Image Classification

A maximum likelihood classifier was selected for the multispectral digital image classification (Jensen, 1996). The goal of the classification was to accurately assign pixels to a limited number of classes. The classes defined for discrimination are: conifer, deciduous, mixed forest, non-vegetation, sparse vegetation, water, and null. An overall accuracy of 97.23%, with an average accuracy of 93.27%, was achieved for the multispectral image classification, based on the correct assignment of classes to the training data. The classification was semi-supervised based upon high resolution *casi* data, field collected ground cover survey data, and forest development survey maps.

Intra-Polygon Species Information Extraction Procedure

To enable the automatic extraction of the raster digital numbers from within each class within each polygon a series of steps are necessary due to a current lack of commercially available software to complete the desired task. A solution was found for the pixel decomposition problem by exploiting what did exist in PCI's EASI/PACE and ESRI's ARC/INFO while creating patch routines in UNIX and C to fill the gaps where necessary. The methodology for extracting pixel image values from within a polygon and within a class within a polygon required the conversion of vector data to raster data. This transformation permits data to be read by the EASI/PACE image processing software where the two raster layers are geometrically overlaid so that the values within each class within each pixel may be extracted. The large number of polygon identification numbers required that 32 bit data be exported, which precluded most of the built-in Importing/Exporting utilities of both ARC/INFO (version 7.0.1) and EASI/PACE (version 6.0.1). For example, there is no command sequence in ARC/INFO GRID to undertake the polygon decomposition goals of this project, and there are too many polygons (77077) to export to EASI/PACE while keeping the vital polygon identification value. The maximum number of polygon labels that may be exported utilizing GRID in ARC/INFO is 10,000, and the maximum number of vector polygon labels that may be imported into EASI/PACE is 512, which are both short of the number of polygons present. The number of polygons, 77077, is beyond the 16 bit unsigned limitation of most commercial image processing and GIS software packages. Accordingly, the following steps outline the procedure undertaken to extract the image values from within polygons and classes within polygons. Short simple computer programs are presented to enable easy alteration based on the specific needs of a project.

1. In ARC/INFO:

- ensure that the map projection for the forest cover polygon data corresponds with the geocoding used in the EASI/PACE image database (ARC: PROJECT).
- POLYGRID, convert the polygons to raster format. Each pixel is given a value which corresponds to the ID of the polygon in which the pixel was found. A pixel size of 30 metres is specified to enable integration of the raster polygon coverage with the image data. (ARC: LIST cover_name.pat)
- in GRID, use the DESCRIBE command to view the dimensions of the newly created raster data matrix. These dimensions are necessary for importing the ASCII data to an image format in EASI/PACE.
- use GRIDASCII command to create a text file from the raster data.

2. In C:

- process the ACSII text file to convert the NO_DATA (-9999) values supplied by the GRIDASCII routine to zeros.
- process the large ASCII file to add line breaks (EOLs) as the exported data is of a line length which exceeds that allowed by the NUMREAD command in EASI/PACE.
- remove unreadable header information. (May not use a text editor as the editor creates line breaks as set by the editor and transforms the data by breaking individual numbers apart. (C program: fileprep.c)

3. In EASI/PACE:
- create a new image (CIM) with the dimensions provided in ARC/INFO GRID with a single 32 bit real channel to receive the information.
- combine the rasterized polygon data with the existing Landsat imagery, landcover classification, and textural information (III).
- use the imported forest stand data to create a mask to remove image areas outside of the FMF from the analysis (In C: binary.c) . Create a binary image to use a multiplier channel against the full image channels using the EASI/PACE task MODEL.
- create image channels (CIM) which correspond to the size of the FMF to move the image data (III) into.
- export the raster data to individual ASCII text files for each of the channels of interest (EASI/PACE: NUMWRIT).
- add a terminal value to the exported text files. Create a file containing only a 99999 value and use the UNIX cat command to append to the end of the exported files. (cat 99999_only_file >> text_file)

4. In C:
- convert the 32 bit real polygon raster data exported by NUMWRIT from an unreadable scientific notation format (ie. 0.13000000E+01) to non-decimal float values which may be read. (C program: inexpout.c)

5. In UNIX:
- check the exported ASCII text files for errors using the word count, command. The total number of words should equal the number of rows multiplied by the number of columns of the exported image channel. Recall to add one for the added terminator value. (UNIX: wc exported_file.asc)

The polygon ID numbers are not assigned to the grid cells in a sequential fashion. Accordingly this requires searching for each polygon which would entail a very large computational load. Starting at the beginning of the files each time to search for the next polygon ID to be assessed vastly increases the computation time. To avoid excess searching the data is sorted using the polygon ID as the key field. The following steps both in C and in UNIX outline the process to add the polygon ID to all text files as a new column of data, to sort the files in ascending order based on the polygon ID, and to remove the polygon ID value once the files have been sorted.

6. In C:
- combine the ASCII text files, the extracted image data, and the raster polygon data. Create files with column one as the polygon ID value and column two is the

image channel.   All files must be sorted in the same manner to retain the integrity of the data. (C program: joiner.c)

7. In UNIX:

- use the sort command to place polygon ID values in ascending order. The syntax to sort numeric values to a specified output files is as follows, the -n flag specifies to sort numeric values and the -o flag specifies to send the results to the named outfile:

```
$> sort -n -o outfile file_to_sort
```

8. In C:

- transform the data into a format suitable for input into the specifically written extraction code. The supplied C program parse.c processes the data into a single sorted column from which the within polygon and class species information can be extracted.
- sequentially process each of the lines of the input files for the polygon, class, and related image pixel values. The C program countmean.c extracts the following sets of values (see Figure 1 for the complete list of variables extracted):
- pixel counts within each polygon,
    - pixel counts within each class within each polygon,
    - mean digital number values within a polygon,
    - mean digital number values within each class within a polygon.
- provide the polygon file, class file, and image data file to run the countmean.c program. The procedure may be run in a sequential batch mode from an executable file, first joining the data to the polygon information, sorting, removing the polygon information, and then inputting the data to countmean.c.

Results and Conclusions

The final output of countmean.c provides tab delimited text files which are readily imported into statistical analysis packages. A demonstration of the increased amount of the original variance in LAI captured through the decomposed polygon estimates over a polygon based estimate is provided in Figure 2. A future article will outline the change in estimates of net primary productivity based upon the species specific computation of LAI from spectral and textural variables extracted from within the GIS polygons. The polygon decomposition routines have also proven to be valuable in change detection. Stands that have been harvested between field data collection and image acquisition are discovered through the presence of non-vegetated or shrub categories in the classification in place of the forest cover type classes.

The original field collected LAI values are appear to be approximately normally distributed with a median value of 8.6 (Figure 2). The refined LAI, from the estimates of polygon contents with an image classification, show a distribution that is near normal yet negatively skewed towards lower LAI values. The skew towards the lower range of LAI values may be due to incomplete polygon coverage by SW, HW, or MW stands from the presence of other categories, especially non-vegetation class pixels representing access roads. The results of the estimates of LAI from the dominant species in the polygon with a mean NDVI value shows little variation about the median of 7.75. The averaged NDVI values calculated to represent the polygons are from a constrained range and leave little opportunity to capture the variance associated with LAI within and between the polygons. The procedure to decompose GIS forest stand polygons based on species assemblages with a remotely sensed image classification has been demonstrated as a valuable methodology.

This methodology is also valuable for change detection and database modification. Imagery collected and classified after completion of the polygon coverage may be applied to update the polygon coverage following disturbances, such as, insect defoliation, fire, or harvesting activities.

Acknowledgments

The author would like to thank Dr. Mike Lavigne of the Canadian Forest Service of Natural Resources Canada, Dr. Steven Franklin of the University of Calgary, and Dr. Ellsworth LeDrew of the University of Waterloo, for data, support, and recommendations which improved the manuscript. Two anonymous reviewers are also thanked for the insightful and constructive criticisms which allowed for improvement upon the original draft.

References

Avery, T. and G. Berlin (1992) *Fundamentals of Remote Sensing and Airphoto Interpretation*, 5th edn., Maxwell Macmillan, Toronto, 472 pp.

Bonan, G. (1993) Importance of leaf area index and forest type when estimating photosynthesis in boreal forests. *Remote Sensing of Environment* **43**, 303-314.

Curran, P. (1980) Multispectral remote sensing of vegetation amount. Progress in Physical Geography **4**, 315-341.

Franklin, S., M. Lavigne, M. Deuling, M. Wulder and E. R. Hunt (1997a) Estimation of forest leaf area index using remote sensing and GIS data for modeling net primary productivity. *International Journal of Remote Sensing*, In press.

Franklin, S., M. Wulder, M . Lavigne, M. Deuling and E. R. Hunt (1997b) Classification of forest covertype for use in ecosystem models of net primary production. *Canadian Journal of Remote Sensing*, In review.

Franklin, S. and P. Giles (1995) Radiometric processing of aerial and satellite remote-sensing imagery. *Computers & Geosciences* **21**(3) 413-423.

Gutman, G. (1991) Vegetation indices from AVHRR: an update and future prospects. *Remote Sensing of Environment* **35** 121-136.

Jensen, J. (1996) *Introductory digital processing: A remote sensing perspective*, 2nd edn., Prentice-Hall, Upper Saddle River, NJ, 316 pp.

Kaufmann, M. and C. Troendle (1981) The relationship of leaf area and foliage biomass to sapwood conducting area in four subalpine forest tree species. *Forest Science* **27**(3) 477-482.

Kozlowski, T., P. Kramer and S. Pallardy (1991) *The physiological ecology of woody plants*. Academic Press Inc., Toronto, 657 pp.

Lavigne, M., J. Luther, S. Franklin and E. Hunt (1996) Comparing branch biomass prediction equations for *Abies balsamea*: Canadian Journal of Forest Research, **26** 611-616.

Lillesand, T. and R. Kiefer (1994) *Remote sensing and image interpretation*. John Wiley & Sons, Inc., Toronto, 750 pp.

Pierce, L. and S. Running (1988) Rapid estimation of coniferous forest leaf area index using a portable integrating radiometer. *Ecology* **69**(6) 1762-1767.

Power, R. and B. Matson (1995) *Ecological land classification of southeastern New Brunswick*. Fundy Model Forest Publication, Sussex, New Brunswick, 6 pp.

Rowe, J. (1972) *Forest regions of Canada*. Environment Canada, Canadian Forest Service Publication n. 1300, Ottawa, Ontario, 172 pp.

Running, S. and E. R. Hunt (1994) Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global scale models. In *Scaling physiological processes: Leaf to globe*, ed. Ehleringer, J. and C. Field, pp. 141-158. Academic Press, Inc., Toronto.

Running, S., D. Peterson, M. Spanner and K. Teuber (1986) Remote sensing of coniferous forest leaf area. *Ecology* **67**(1) 273-276.

Spanner, M., L. Pierce, D. Peterson and S. Running (1990) Remote sensing of temperate coniferous forest leaf area index: The influence of canopy closure, understory vegetation and background reflectance. International Journal of Remote Sensing. **11**(1) 95-111.

Waring, R. and W. Schlesinger (1985) *Forest Ecosystems: Concepts and Management*. Academic Press, Inc., Toronto, 340 pp.

Wulder, M. (1996) Airborne remote sensing of forest structure: Estimation of leaf area index. MES Thesis, University of Waterloo, Waterloo, Ontario, Canada, 101 pp.

Wulder, S. Franklin, M., Lavigne, M. Deuling and E. Hunt (1996a) Estimation of the net primary productivity of the Fundy Model Forest. *Proceedings of Eco-Informa '96*, Lake Buena Vista, Florida, November 4-7, **10** 327-332.

Wulder, M., Lavigne, M. and S. Franklin (1996b) High spatial resolution optical image texture for improved estimation of forest stand leaf area index. *Canadian Journal of Remote Sensing*. **22**(4) 441-449.

Wulder, M., E. LeDrew, M. Lavigne and S. Franklin (1997a) High resolution optical image texture information in the estimation of deciduous and mixed forest LAI. *Remote Sensing of Environment*, In review.

Wulder, M., E. LeDrew, M. Lavigne and S. Franklin (1997b) A comparison of texture measures in the estimation of LAI: Introduction of semivariance moment texture. *CD-ROM Proceedings GER'97, Canadian Remote Sensing Society Conference*, Ottawa, Ontario, May 24-30.

Figure 1, GIS polygon decomposition project flow and list of extracted data

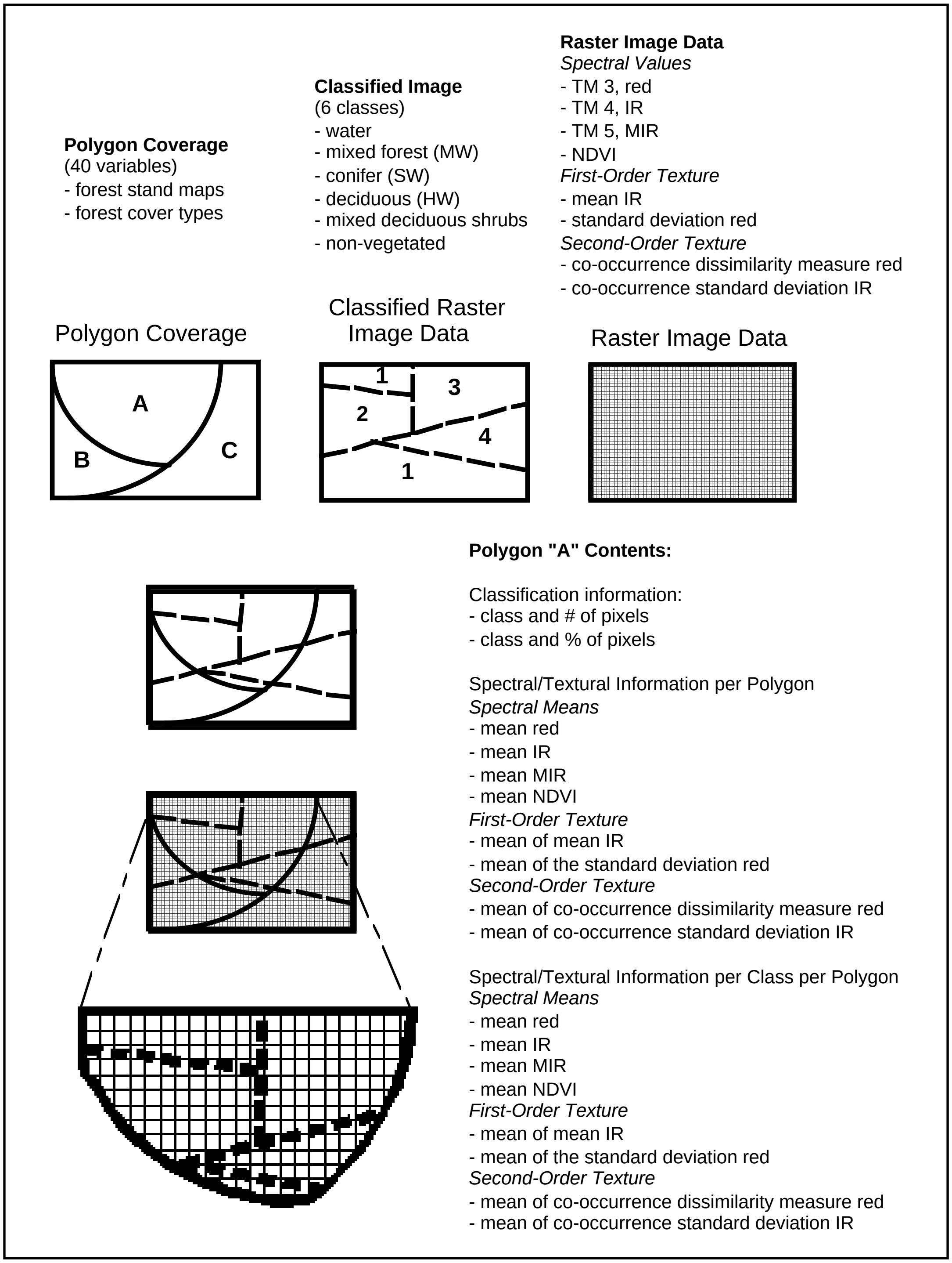

Figure 2, Box plots demonstrating the distribution of variance based upon LAI estimation technique. LAI is the field collected control, refined-LAI is the decomposed polygon estimate, and polygon-LAI is the estimate made from a single value to represent the polygon.