FUSION OF AIRSAR AND TM DATA FOR VARIABLE CLASSIFICATION AND ESTIMATION IN DENSE AND HILLY FORESTS

Mahta Moghaddam[1], Jennifer Dungan[2], and Joseph Coughlan[3]

[1] Jet Propulsion Laboratory, MS 300-227
4800 Oak Grove Drive, Pasadena, CA 91109, USA
Phone: (+1) 818-354-1591, Email: mahta@radar-sci.jpl.nasa.gov
[2] CSUMB, NASA Ames Research Center, MS 242-4
Moffett Field, CA 94035-1000 USA
Phone: (+1) 650-604-3618, Email: jdungan@mail.arc.nasa.gov
[3] NASA Ames Research Center, MS 242-4
Moffett Field, CA 94035-1000 USA
Phone: (+1) 650-604-5689, Email: jcoughlan@mail.arc.nasa.gov

ABSTRACT

Radar and optical remote sensing data are used to develop a classification algorithm based on nonlinear estimation theory. The study site is the H. J. Andrews forest in Oregon, USA, which has significant topography and several old-growth conifer stands with biomass values sometimes exceeding 1000 tons/hectare. Polarimetric C-band, L-band, and P-band AIRSAR data, interferometric C-band TOPSAR data, and six channels of Landsat TM data are used in a regression analysis that relates them to several measurements of one or more forest variables. Parametric expressions are derived and used to estimate the same variable(s) at other locations from the combination of AIRSAR and TM data. Statistical characteristics of the variable estimates are derived and used to define variable classes. These classes are the basis for future analytic estimation algorithms.

1. INTRODUCTION

The expanded remotely sensed data space consisting of coincident radar backscatter and optical reflectance data provides for a more complete description of the Earth surface. This is especially useful where many variables are needed to describe a certain scene, such as in the presence of dense and complex-structured vegetation or where there is considerable underlying topography. The goal of this paper is to use a combination of radar and optical data to develop a methodology for forest variable classification for dense and hilly forests, and further, class-specific variable estimation.

To achieve classification, first a number of variables are defined which are of interest to ecologists for forest process modeling (Coughlan and Dungan, 1996). These include leaf biomass, LAI, and tree height. The remote sensing data from radar and TM are used to formulate a multivariate regression analysis problem given the ground measurements of the variables. Each class of each variable is defined by a probability density function (pdf), the spread of which defines the range of that class. Classification accuracies are defined using the degree of overlap between the variable estimate pdfs: the smaller the overlap region, the better the separation of classes. Using the regression relations derived using ground measurements, the variables can be estimated and classified for other areas in the scene. Validation is carried out using ground measurements at points not already used in deriving the parametric regression relations. Classification results provide the basis for the future work of class-specific variable estimation using radar and optical data.

2. SITE DESCRIPTION

The area to be used in this study is the H. J. Andrews Forest in Oregon, one of the Long-Term Ecological Research (LTER) sites in the US. This area consists of various dense old-growth conifer stands, with biomass values ranging from less than 100 tons/hectare to over 1000 tons/hectare. The average altitude is about 950 m, with the lowest and highest points at about 600m and 1700m, respectively. The Andrews forest has been the subject of many ecological studies over several decades, resulting in an abundance of ground measurements. There are several permanent GPS monuments and meteorological stations. The forest stand characteristics have also been extensively documented. In particular, foliage biomass and leaf-area index (LAI) values for approximately 30 reference stands have recently become available (Means et al, 1999). We will use half of these measurements in deriving the regression relations and the other half for validation.

3. REMOTE SENSING DATA

The remote sensing data types to be used are the C-, L-, and P-band polarimetric radar data from the JPL airborne SAR (AIRSAR), the C-band single-polarization data from the JPL topographic SAR (TOPSAR), and the Thematic Mapper (TM) data from Landsat, all acquired in late April 1998. The total number of useful separate data channels from the AIRSAR is 15 (three frequencies, each with three unique polarizations and amplitude and phase of the like-polarized correlation), from the TOPSAR is 2 (amplitude and phase of the interferometric correlation), and from the TM is 6 (the thermal band is not used).

These data types are of varying pixel resolutions, and hence must be resampled to a common resolution. The range pixel spacing of the AIRSAR is 3.3m for C- and L-bands and 6.6m for P-band. The TOPSAR pixel spacing is 10m, and the TM pixel size is 30m.

Radiometric and polarimetric calibrations have been carried out on the AIRSAR data. Due to pronounced topography, the radiometric calibration involves an added step to remove the effect of local slopes. Furthermore, for the areas where the backscattering process involves ground or double-bounce scattering, a correction factor has to be applied to take the local slopes into account. The Landsat TM data were acquired under almost cloud-free conditions. All radar data are coregistered to the TM data, since the latter are already geocoded. Ground control points are chosen manually to tie the various datasets.

4. REGRESSION MODEL

For demonstration purposes, the analysis will be performed for only one stand variable, e.g., foliage biomass. We shall denote it by X. The various data channels will be denoted as D_i, where $i = 1,\ldots,N$ with N being the total number of available data channels. As a result of regression analysis, the measured data can be fit into polynomials in the measured variable X. This can be written as

$$D_i = \Sigma_{j=0}^{M} a_{ij} X^j \pm err_i,$$

where M is the order of regression and err_i is the fitting error for each D_i. Our analysis will not further carry errors resulting from regression fitting. However, errors on ground measurements X are used to study their effect in the derived regression relations, i.e., on a_{ij}. If the fitting errors are ignored, we can simplify the above equation as

$$\mathbf{D} = \mathbf{f}(X),$$

where **D** is the vector containing all data sets and **f** simply denotes a vector function. An example is shown in Figure 1, where the polynomial fits to two radar and two TM channels are derived, with foliage mass as the independent variable.

5. ESTIMATION AND CLASSIFICATION

The derived relationships between the data and the selected ground measurements of the variable can be used to estimate that variable elsewhere in the scene covered by the remote sensing instruments. To do so, an estimation algorithm can be used that minimizes the quantity

$$L = |\mathbf{D} - \mathbf{f}(X)|^2 + |X - X_{apriori}|^2.$$

Here, $X_{apriori}$ represents our previous knowledge about the variable to be estimated. The L_2 norms are both defined through the respective covariances, and hence allow the inclusion of the statistical characteristics of both the data and the variable measurements (Moghaddam and Saatchi, 1999).

To minimize L, an iterative algorithm is used as follows:

1. Assume an initial arbitrary estimate for X.
2. Calculate **f**(X) from the previously established regression model.
3. Calculate L and decide if it is small enough. If so, X is the solution.
4. If L is not small enough, calculate the gradient or other appropriate direction of change for L, followed by the calculation of a step length in that direction. This step involves calculation of derivatives of **f**(X).
5. Update X and go back to the second step above.

The statistical estimation errors can be found by superimposing simulated errors on the derived regression coefficients of **f**(X), including measured covariances of the data channels and the stand variable, and repeating the estimation procedure for each added value of noise. It will be assumed that the noise has a uniform Gaussian distribution. Its mean value can be determined from the regression analysis if the error in ground measurements of X are known.

6. RESULTS

The above algorithm was applied to the radar and optical data described in Section 3 to classify foliage biomass within the H.J. Andrews forest. The advantage of fusing optical and radar data over using only one of the two data types to estimate foliage mass is shown in Figure 2. The figure shows that the overall estimation accuracy for the entire range of variable values is higher when both data tyes are used.

To show the statistical accuracy of the algorithm, a uniformly Gaussian error was superimposed on the coefficients. The noise amplitude was 5% of the coefficient valus. The error analysis described in Section

5 was carried out. The results are shown in Figure 3, where it is observed that the estimates fall in "ranges," as opposed to being deterministically and uniquely calculated. In other words, depending on the spread of the estimated values, they may not be all statistically independent. Rather, they may be more uniquely described by a "class" with a spread given by the width of a defined dependence range.

As such, the estimated variable values and their statistical distribution suggest a classification scheme. Although classification of the variable X might connote a reduced information content, it in fact contains all the "useful" and unique information derived from the regression-estimation algorithm. For the same number of classes, the classification accuracy, i.e., the statistical independence of classes, is expected to become higher as the number of data channels used is increased. Conversely, for similar classification accuracies, fewer classes are expected to be identified if fewer channels are used. This method brings optical and radar into a unified statistical framework, allowing "fusion.'"

ACKNOWLEDGMENT

This work was performed in part by the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, and in part by the NASA Ames Research Center, Moffett Field, CA, both under contract from the National Aeronautics and Space Administration. The authors would like to thank Dr. Steve Acker of Oregon State University for providing ground measurements.

REFERENCES

Coughlan, J., C., and J. L. Dungan, 1996. Combining remote sensing and forest ecosystem modeling: An example using the Regional HydroEcological Simulation System (RHESSys). The Use of Remote Sensing in the Modeling of Forest Productivity at Scales from the Stand to the Globe, H.L. Gholz, K. Nakane and H. Shinoda (eds.). Kluwer Academic Publishers: Dordrecht. pp. 139-158.

Means, J., S.A. Acker, D.J. Harding, J.B. Blair, M.A. Lefsky, W.B. Cohen, M.E. Harmon, and W.A. McKee, 1999. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western Cascades of Oregon, *Remote Sens. Environ.*, 67:298-308.

Moghaddam, M., and S. Saatchi, 1999. Monitoring tree moisture using an estimation algorithm applied to SAR data from BOREAS. *IEEE Trans. Geosci. Remote Sensing.*, vol. 37, no. 2, pp. 901-916.

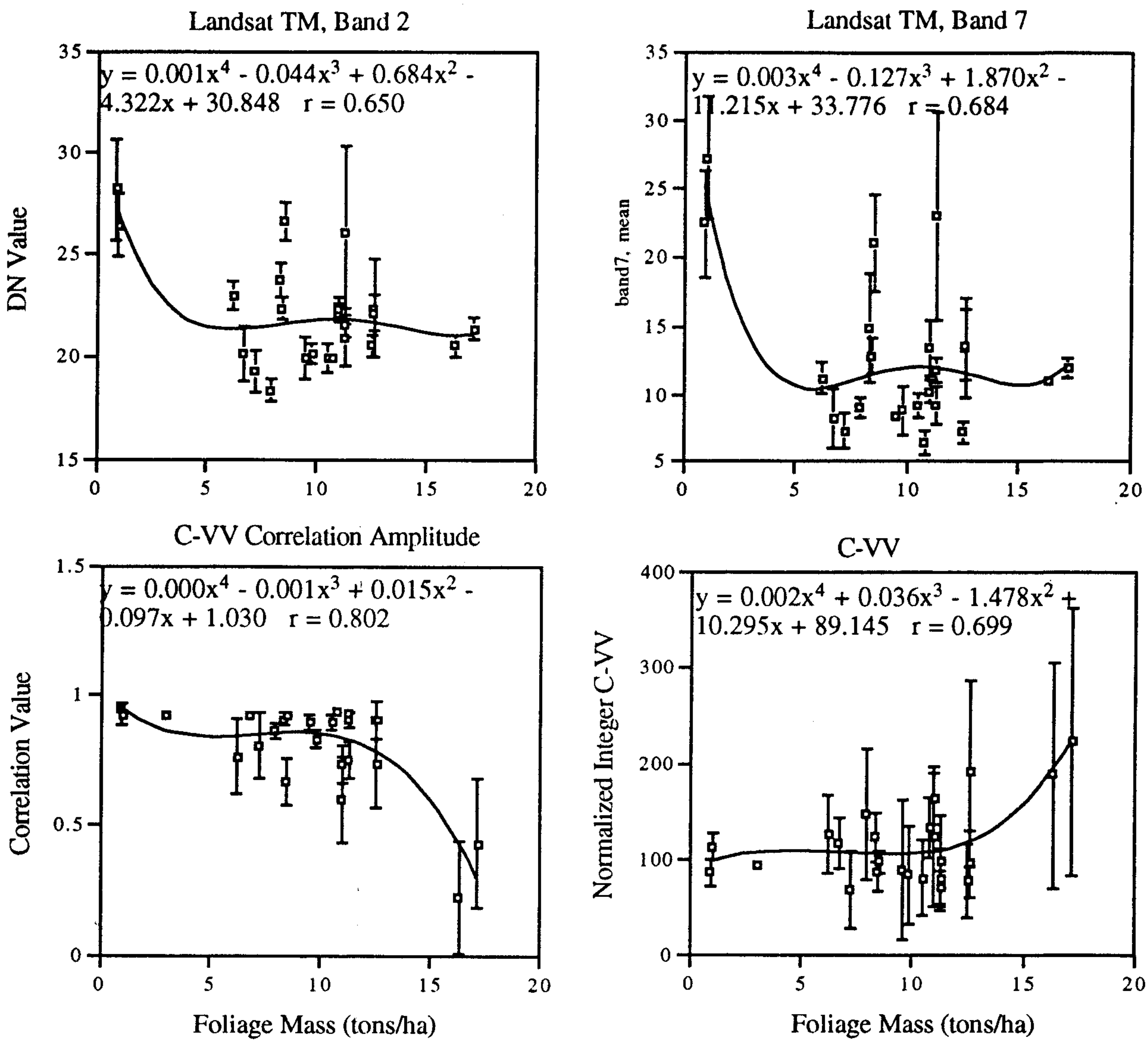

Figure 1. Polynomial regression relations between remote sensing data and measured foliage mass

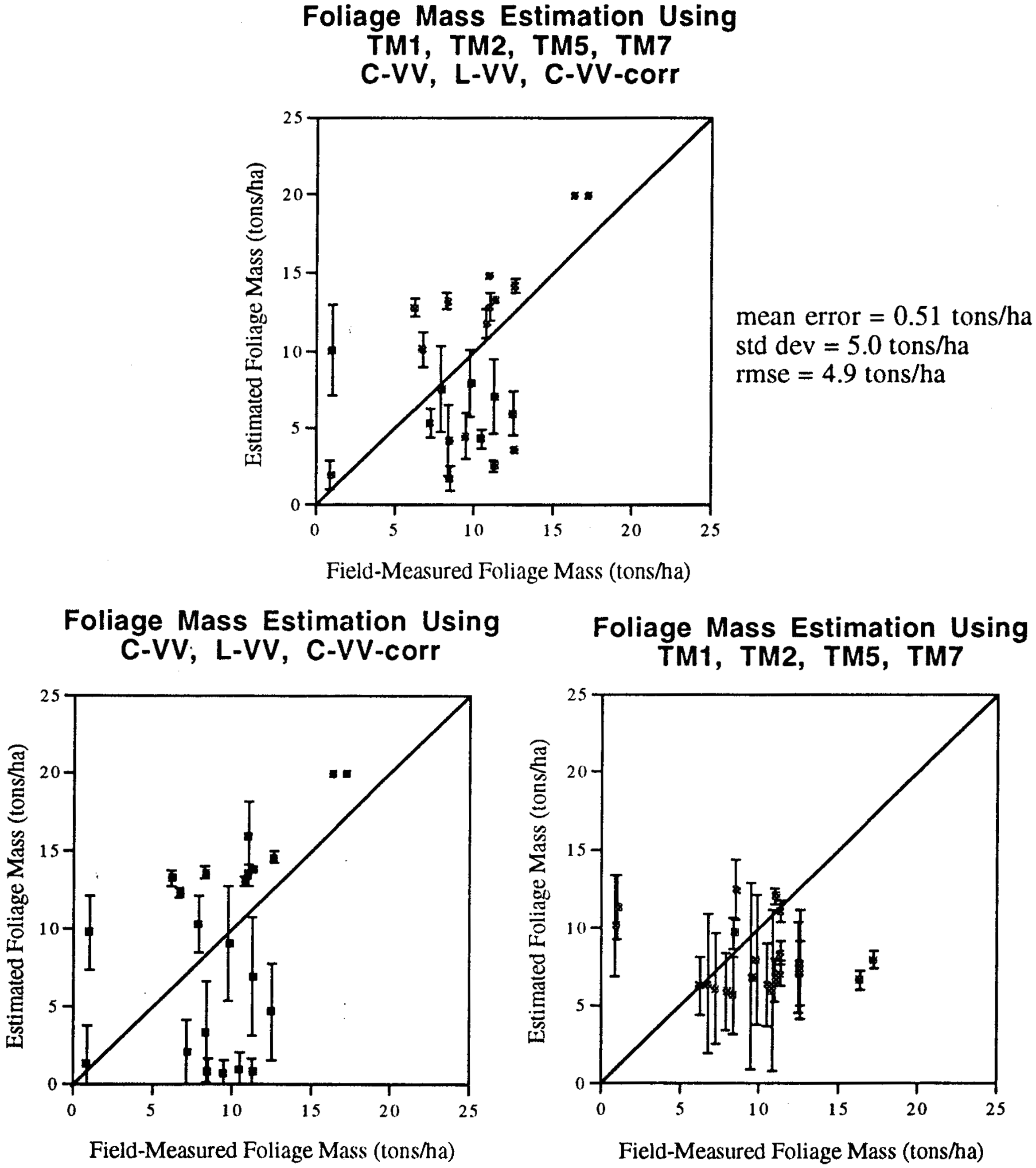

Figure 2. Simultaneous use of both optical and radar data improves the accuracy of estimation compared to using only one data type.

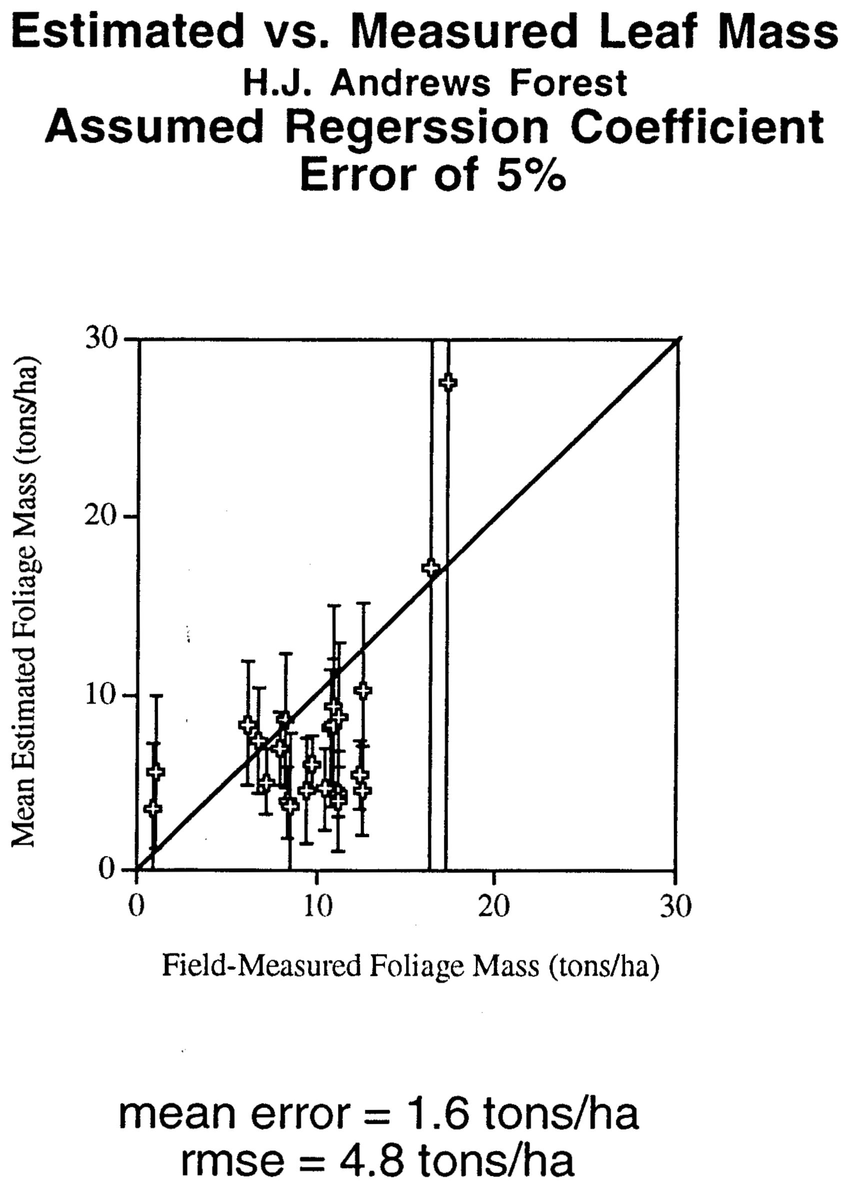

Figure 3. Statistical accuracy of estimated values assuming that the regression coefficients have a 5% Gaussian error.