An Introduction to Using Generalized Linear Models to Enhance Satellite-based Change Detection

Jeffrey T. Morisette and Siamak Khorram
Computer Graphic Center
North Carolina State University
Box 7106
Raleigh NC 27695 - 7106
Phone: (919)515-3434 / Fax: (919)515-3439/ e-mail: jtmorise@unity.ncsu.edu

Abstract -- A popular satellite-based land cover change detection technique it to use the spectral information to set up a binary "change/no-change" mask. For each pixel, if there is a big enough difference between the reflectance values for two images acquired at different times, the area represented by that pixel is considered to have changed. The different change detection methods are different in how they determine a "big enough difference". The analyst is left to choose which function of the reflectance values to use and where to set the "change" threshold. This choice is often subjective and effects the accuracy of the change detection. In our paper we explore the use of Generalized Linear Models (GLM) as a way to enhance satellite based change detection by helping determine the most appropriate function of the reflectance values to use then apply the modeling to select the threshold. The main idea is that reflectance values from satellite imagery can be incorporated into a GLM to predict the *probability of change* from one land cover to another and that this method of change detection will provide more information than current change detection methods. A major benefit of using generalized linear models is to determining the significance and most useful combinations of the spectral data for predicting changed areas.

INTRODUCTION

A standard change detection technique for large areas is to use the spectral information from two satellite images acquired at different dates. This method considers individual pixels in an image and compares them to their corresponding pixels from the other image. For each pixel, if the ***difference*** between the reflectance values from the two images is ***big enough***, the area represented by that pixel is considered to have changed. Change detection methods vary in how they determine if the "difference" is "big enough". First there is the question of what is meant by "difference". Is it changes in infrared reflectance? Is it changes is some vegetation index? In general, the "difference" is some function of the reflectance values. For example, the difference might be a simple subtraction of the infrared reflectance for a pixel at one time from the infrared reflectance for the corresponding pixel at another time. The next question is how big is "big enough". This is the issue of choosing a threshold. If the difference is beyond the threshold then the area represented by the pixel is believed to have experienced a change in land cover sometime between the two images' acquisition. A binary change images is created by labeling those pixels whose functional values are beyond the threshold as changed and those pixels whose functional values are less than the threshold to be unchanged. Choung [1] and Fung and LeDrew [2] have investigated the determination of optimal threshold levels. Both studies reveal an intuitive balance of errors for various threshold levels. For example, with image differencing the lower the threshold, the more non-changed pixels were classified as changed (commission errors); the higher the threshold, the better the chance of changed pixels classified as non-changed (omission errors).

In all studies that create a binary change image, the value at which the threshold is set is somewhat arbitrary. With respect to image differencing, Jensen states that "most analysts prefer to experiment empirically...thus the amount of change...is often subjective and must be based on familiarity with the study area" [3, p.269]. With respect to the application of a binary change mask, the final outcome is dependent on the quality of the mask [3, p.270] and the quality of the mask will depend on the appropriate threshold level for the mask [4]. In order to avoid the subjectivity and arbitrary nature of setting a threshold, we have worked to develop a procedure that can be used as a more analytical method for investigating which function of radiance values is most effective for determining change areas.

In this paper we introduce the use of generalized linear models (GLMs) as a way to enhance satellite based change detection. Generalized Linear Models are a broad class of regression-type models. The use of GLMs has continued to gain attention since they were introduced by Nelder and Wedderburn in 1972 [5]. Unlike standard linear regression, GLMs can be used to model binary or discrete response variables. In this paper, the main idea is that GLMs can be used to regress the binary response of "change/no-change" on the reflectance values from satellite imagery. The binary response is determined from the interpretation of air photos. The radiance values are extracted from the satellite images. The object is to describe how GLMs can be used to enhance satellite-based change detection. This will be done by describing an example change detection utilizing GLMs.

0-7803-3836-7/97/$10.00

EXAMPLE

First, we collected a sample of "ground truth" or reference data. The sample was based on a set of particular ground locations. For each location we determined if the land cover had changed by interpreting the sample areas on the air photos. We do this from the digitized air photos for 1994 and from the hardcopy air photos for 1988. For each point in the sample we therefore have the land cover type for each time. For each point, if there is a different land cover type for the two time periods then the response for that point is a 1. If there is no change, the sample point will be assigned a response of 0. The next step is to relate the 0 and 1 responses from the sample points to the radiance values from the satellite data. So, for each point in the sample we obtain the satellite radiance values from Landsat TM data from 1988 and 1994 for each sample area. From this, for each sample point, we have a paired set:

{Change determination from reference data : Satellite radiance values from corresponding area}

Similar to a standard regression procedure, for every point in the sample we have one response (the "Y" value) and a set of possible explanatory variable (the "X" values). Using the LOGISTIC procedure in SAS [6] we can determine which function of the radiance values does the best at discriminating between the 0's and 1's – the unchanged and changed sample areas. The objective is to find a function of radiance values that can be used to predict where changes have occurred. Once an appropriate function is found, this function will serve as a model from which to predict the probability of change for the entire satellite image. That is, we use the sample points, for which we have air photo reference data coupled with the radiance values, to develop a model to estimate the probability of change for every pixel in the image.

Fig 1 shows the empirical data in the two dimensions corresponding to the two variables that the GLMs indicated as the most significant indicators of change out of a set of 27 different functions of the radiance values. By using the empirical data and the statistical model we were able to sort through the different possible functions of radiance values and select those which proved to be the most significant for discriminating change areas. The most significant variables were the difference in tassel cap band 5 and a Euclidean distance measure of all tassel-cap bands (expressed in the equation below as "vector-tas"). The GLM procedure produces a model in this two-dimensional space. The fitted logistic model follows the equation:

$$\mathit{Prob\ of\ change} = \frac{e^{-4.2954 - 0.9294 \times dif\text{-}tas\text{-}band5 + 0.1551 \times vector\text{-}tas}}{1 + e^{-4.2954 - 0.9294 \times dif\text{-}tas\text{-}band5 + 0.1551 \times vector\text{-}tas}}$$

Fig 2 shows the surface produced by this model. The model can be incorporated with the image data to produce an image in which the pixel values represent the probability of change (POC). The POC image shows one of the unique qualities of using GLM for a change detection product: there is no need to set a threshold. Instead the change detection product gives a probabilistic measure for change where values near zero represent a less likely chance of change and value near one represent a higher likelihood of change. Currently, we are research methods to utilize this continuous change product for display and modeling purposes. Fig 3 shows an image that uses the logistic model to produce a POC image. Darker areas represent a low POC with the chances increasing through the lighter colors, which represent the highest POC. Another unique quality of using GLMs is that the logistic model can also be used to estimate the variability of the estimated POC. As with any regression, the logistic model can be used to put a confidence interval on the estimates derived from the model [6]. Fig 4 shows an image representing the same ground area as in Fig 3. In Fig 4 the pixel values represent the width of the confidence interval for the estimated POC. Larger confidence intervals have brighter values. In Fig. 4 we see how the boundaries around change areas are more variable. This makes sense intuitively in that we are less certain of changes in these transitional zones. Fig 4 shows how the GLM can be used to indicate the spatial nature of the variability in the change detection estimates.

CONCLUSIONS

The GLM help to select, from a group of possible functions of radiance values, the most significant function of the radiance values. The resulting model can be used to produce a POC image as well as image which represents the variability of the estimates.

ACKNOWLEDGMENTS

Thanks to NASA's Earth System Science Graduate Student Fellowship Program for support of this research. Thanks to the U.S. National Oceanic and Atmospheric Administration Coastal Change Analysis Program for the project which supplied the data for this research.

REFERENCES

[1] Choung, Song Hak. 1992. "Wetland Change Detection Using Landsat-5 Thematic Mapper Data in Jackson Hole, Wyoming." Ph.D. Dissertation, University of Idaho, Ulliman, Joseph J.

[2] Fung, Tung and Ellsworth LeDrew. 1987. "Application of Principal Components Analysis to Change Detection." *Photogrammetric Engineering and Remote Sensing* 53(12):1649-58.

[3] Jensen, John R. 1996. "Introductory Digital Image Processing: A Remote Sensing Perspective" . Second ed. Upper Saddle River, NJ: Prentice Hall.

[4] Jensen, John R., David J. Cowen, John D. Althausen, Sunil Narumalani, and Oliver Weatherbee. 1993. "An Evaluation of the CoastWatch Change Detection Protocol

in South Carolina." *Photogrammetric Engineering and Remote Sensing* 59(6):1039-46.

[5] Agresti, Alan. 1990. "Categorical Data Analysis." New York: John Wiley and Sons.

[6] SAS. 1990. *SAS/Stat User's Guide, Version 6, Forth Edition*. Cary NC: SAS Institute Inc.

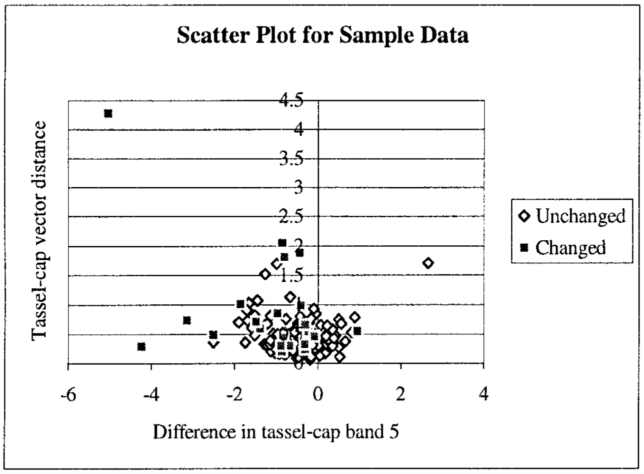

Figure 1: Distribution of empirical data

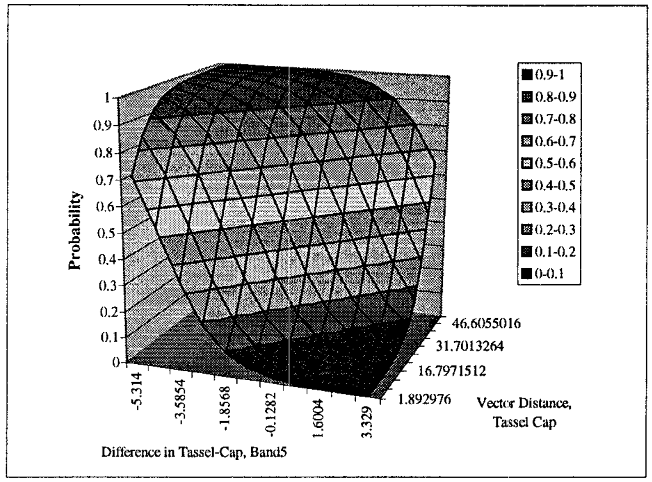

Figure 2: Model surface

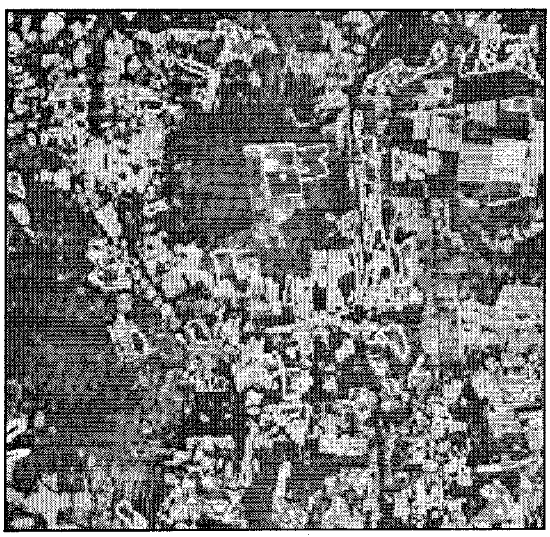

Figure 3: Probability of change image

Figure 4: Variability image