*Paper presented at the 22nd Asian Conference on Remote Sensing, 5 - 9 November 2001, Singapore.*


# MONITORING LAND-USE CHANGE IN A TIME SERIES OF SEQUENTIAL LANDSAT IMAGERY FOR THE PEARL RIVER DELTA, CHINA: AN ECONOMETRIC TECHNIQUE

Karen C. SETO
Fellow, Institute for International Studies
Center for Environmental Science and Policy
Stanford University
Stanford, CA 94305-6055
Tel: (1)-650-736-1295 Fax: (1)-650-725-1992
Email: kseto@stanford.edu
USA

Robert K. KAUFMANN
Associate Professor, Department of Geography
Boston University
Boston, MA 02215
Tel: (1)-617-353-3940 Fax: (1)-617-353-5986
Email: kaufmann@bu.edu
USA

**KEY WORDS:** Land-use change, econometric techniques, Pearl River Delta

**ABSTRACT:** In this paper, we describe a new change detection method that determines the date at which land-use changes occur in a sequential series of Landsat TM images. We use this approach to extract annual rates of land conversion in the Pearl River Delta, China, for the period 1988-1996. The technique is a three-step change detection procedure that uses time series and panel econometric methods. First, regression equations are estimated for each of the six TM bands for unchanged, stable land-cover classes. Second, the regression equations for these stable land-cover classes are used to calculate DN values for change classes for each of the eight possible dates of change (1989-1996). Lastly, the year of land-use change is identified by comparing a pixel's DN values against the eight possible dates of land conversion using tests for predictive accuracy. The accuracy of the dates of land-use change identified by the econometric technique is upwards of 85 percent. Additionally, the econometric technique may reduce efforts required to assemble the training data and to correct the images for atmospheric effects.

## 1 INTRODUCTION

Statistically coupling time series of satellite images with time series of economic and demographic data can provide insight on the relation between economic activity and land-use change. Time series data for many socioeconomic variables are available from a variety of sources (e.g. statistical yearbooks) at annual frequencies. To use these data in statistical analyses, estimates of rates of land-use change must also be available at annual frequencies. This presents a methodological difficulty. Most change detection techniques are designed to analyze two or three images. Computationally, repeated applications of these techniques can introduce errors associated with post-classification comparison of images. This paper describes a method that uses econometric techniques to determine the date at which land-use changes occur in nine images (1988-1996) of the Pearl River Delta China.

## 2 DATA AND DATA PREPROCESSING

Our remote sensing data consists of nine predominantly cloud-free Landsat TM images of the Pearl River Delta in Southern China (Table 1). Eight images are georeferenced to the 1992 master image with a Universal Transverse Mercator (UTM) map projection provided by the Institute for Remote Sensing Application in Beijing. The images are resampled to 30- by 30-m pixels using a nearest neighbor resampling algorithm with a first order polynomial. The number of ground control points (gcp) used for the registration varies by image, and in all but one case, the root mean square error (RMSE) of the registration process is less than a third of a pixel (Table 1). To correct for changes in atmospheric conditions, illumination angles, and seasonal variation across the images, a relative radiometric normalization technique is used (Song *et al*., 2001). Based on fieldwork and visual interpretation of the images, we identify 809 sites with 7807 pixels for training and testing. The sites are distributed among seven stable and sixteen change land-cover classes (Table 2). Of the sixteen change land-cover classes, ten of them are land-use change classes (marked in bold).

| *Acquisition date* | *Number of gcp's for registration* | *Geometric Registration RMSE* |
|---|---|---|
| 10 December 1988 | 25 | ± 0.2893 |
| 13 December 1989 | 23 | ± 0.2914 |
| 30 October 1990 | 18 | ± 0.2364 |
| 02 February 1991 | 18 | ± 0.2412 |
| 20 January 1992 | 16 | master image |
| 24 December 1993 | 20 | ± 0.2618 |
| 08 November 1994 | 17 | ± 0.3162 |
| 30 December 1995 | 24 | ± 0.2970 |
| 03 March 1996 | 25 | ± 0.2776 |

Table 1. Characteristics of TM images used in the study

| *Stable Classes* | *Sites* | *Change Classes* | *Sites* |
|---|---|---|---|
| water | 34 | water to fish pond | 22 |
| | | water to agriculture | 48 |
| | | **water to transition** | 36 |
| | | **water to urban** | 26 |
| forest | 31 | forest to water | 21 |
| | | **forest to transition** | 36 |
| | | **forest to urban** | 18 |
| shrub | 35 | shrub to water | 14 |
| | | **shrub to transition** | 33 |
| | | **shrub to urban** | 23 |
| fish pond | 22 | **fish pond to transition** | 12 |
| agriculture | 117 | agriculture to water | 26 |
| | | agriculture to fish pond | 34 |
| | | **agriculture to transition** | 83 |
| | | **agriculture to urban** | 47 |
| transition | 28 | **transition to urban** | 24 |
| urban | 39 | | |

Table 2. Stable and Change Land-Use Classes, and Number of Training Sites for Each Class

The stable classes are self-explanatory except for the transition class, which represents land where the previous land-cover has been removed, but the structures associated with the new use have not been put in place. Each of the 469 change sites is assigned a date at which land-use changes occur based on a visual interpretation of the images. An analyst who visited the study area twice and is familiar with the region evaluated the nine images to identify the year (hereafter termed date of change) in which the first pixel within the site changed. A second label is attached to 39% of the sites in which all of the pixels do not change simultaneously.

## 3 CHANGE DETECTION METHODOLOGY

The econometric change detection technique uses time series and panel techniques to identify the date of land conversion for individual pixels in three-steps. In the first step, regression equations are estimated for each of the six TM bands for each of the seven stable land-cover classes. In the second step, the estimated regression equations for each class are used to calculate DN values for change land-cover classes for each of the eight possible dates of change. In the third step, the date of change is identified by comparing a pixel's DN values against the eight possible dates of change using tests for predictive accuracy.

### 3.1 Models of Stable Land-Use

For each of the seven stable land-use classes, six regression equations are estimated that specify DN values for TM Bands 1-5 and 7 as follows:

$$DN^*_{1tj} = \beta X^*_t + \gamma Y^*_t + \mu_{1tj} \qquad (1)$$

$$DN^*_{2tj} = \beta X^*_t + \gamma Y^*_t + \mu_{2tj} \qquad (2)$$

$$DN^*_{3tj} = \beta X^*_t + \gamma Y^*_t + \mu_{3tj} \qquad (3)$$

$$DN^*_{4tj} = \beta X^*_t + \gamma Y^*_t + \mu_{4tj} \qquad (4)$$

$$DN^*_{5tj} = \beta X^*_t + \gamma Y^*_t + \mu^{5tj} \qquad (5)$$

$$DN^{*}_{7tj} = \beta X^{*}_{t} + \gamma Y^{*}_{t} + \mu_{7tj} \qquad (6)$$

in which DN is a time series for the DN value for band i (i = 1-5, 7) at time t (t= 1988-1996) for stable land class j (j=1,…7), X is a vector (3 x 9) of time series for physical variables thought to effect reflectivity (solar zenith angle [SZA], aerosol optical depth [AOD], and minimum DN [Min DN]) value for that band), Y is a vector (6 x 9) of six dummy variables (January, February, March, October, November, December) for the month the image is obtained (the dummy variable for the month when the image is obtained has a value of 1 and a value of zero for the other five dummy variables), $\beta$ and $\gamma$ are vectors of regression coefficients, and $\mu$ are a time series of error terms.

Equations 1-6 are estimated with a fixed effect estimator to account for the spectral heterogeneity of land-cover classes and unobservable variables. A random effects estimator also can be used, but the results would be nearly identical to the fixed effects estimator because the elements of the X vector do not vary among pixels within individual images. To estimate the fixed effect estimator, the data are transformed such that the mean value of a variable for a pixel is removed from the nine annual observations of the variable for that pixel (Hsiao, 1986). For example, the transformed DN value for Band 1 is given by ($DN_{1tj} - \overline{DN}_{1j}$) where:

$$\overline{DN}_{1j} = \frac{1}{N}\left(\sum_{t=1}^{N} DN_{1jt}\right) \qquad (7)$$

in which N is the number of observations per pixel (9). The independent variables are transformed using a similar procedure, as indicated by the * in equations 1-6. Because of these transformations, equations 1-6 do not contain an intercept. Equations 1-6 can be estimated individually using OLS, but the regression results may be inefficient because the errors ($\mu$) for individual equations within a given land-cover class may be correlated due to the correlations that exist among bands 1-5, and 7. To avoid this potential source of inefficiency, equations 1-6 are estimated as a system of seemingly unrelated regressions (SUR). This procedure is repeated for each of the seven stable land-cover classes. The estimates for $\beta$ and $\gamma$ along with the values for X and Y variables are used to calculate the transformed DN value for each stable land-use class for each image. The transformed values are converted back to levels using equation 7.

### 3.2 Models Of All Possible Dates Of Land-Use Change

The DN values for each stable land-use class are combined to calculate DN values for each of the sixteen land-cover change classes (Table 2). To represent the eight possible dates of change, the DN values for the stable land classes are spliced together at each of eight points. To represent the DN values associated with a 1990 date of land-use change for the <u>agriculture to urban</u> class, the DN values generated by the <u>agriculture</u> regression equations for 1988 and 1989 are combined with the DN values generated by the <u>urban</u> regression equations for the years 1990 through 1996. This process is repeated to generate a model for the <u>agriculture to urban</u> class for each of the eight possible dates of change between 1989 and 1996. This process is repeated to generate a model for each date of change for each of the sixteen land-cover change classes, including the ten land-use change classes.

### 3.3 Identifying The Date Of Land-Use Change With Tests Of Predictive Accuracy

These models serve as ideotypes against which each pixel can be compared to determine the date of change. For each pixel in a change class, its DN values are compared against the eight models for that change land-cover class, each of which represents one of eight possible dates of change. For a pixel in the <u>agriculture to urban</u> class, its DN values are compared to values generated by the stable <u>agriculture</u> and <u>urban</u> regression equations that are spliced together in 1989, or 1990,… ,1996. The model that best describes a pixel's date of change is chosen using a test of predictive accuracy. The test compares the difference (d) in the absolute value of the forecast errors, which is given by:

$$d_{it} = \left|DN_{it} - D\hat{N}_{itj}\right| - \left|DN_{it} - D\hat{N}_{itj+1}\right| \qquad (8)$$

in which DN is the DN value for pixel i at time t, $D\hat{N}_{itj}$ is the DN value predicted for pixel i at time t that changes between stable land-use classes at time j and $D\hat{N}_{itj+1}$ is the value predicted for pixel i at time t that changes between stable classes at time j plus 1. Following Diebold and Mariano (1995), the values of d are weighted and summed as follows:

$$S_{3a} = \frac{\sum_{t=1}^{N} I_{+}(d_t)\,rank(|d_t|) - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}) \qquad (9)$$

$$I_+(d_t) = 1 \qquad if\ d_t > o$$
$$= 0\ otherwise$$

to calculate the s3a statistic which can be compared against a t distribution. The s3a test statistic weighs the value of d using a binary choice (zero or one), that is determined by the model for the date of change that has the smaller absolute forecast error. This binary choice is weighted by the rank order of the errors. The model with the smaller forecast error is indicated by the sign on the s3a test statistic. The test statistic will be negative if the absolute forecast error associated the model that represents the change in land-use at date j is smaller than the absolute forecast error associated the model that represents the change in land-use at date j+1. The date of change is chosen based on the model that 'best' describes a pixel's DN values as indicated by the value of the s3a statistic that exceeds a threshold for statistical significance.

The threshold for statistical significance that is used to distinguish among competing mo dels is chosen empirically based on the trade-off between the fraction of pixels for which a date of change can be identified and the accuracy of that date of change. A rigorous threshold (e.g., $p < 0.05$) allows can differentiate between competing dates of change with a high degree of confidence. However, for many pixels, the s3a statistic may not be able to differentiate competing models with a high degree of confidence. Under these conditions, the methodology can not assign a date of change and the pixel's date of change is unclassified. Alternatively, most of the pixels can be classified by using a relatively low threshold (e.g., $p < 0.5$) to distinguish between competing models. However, a low threshold may increase the probability of falsely detecting change.

## 4 RESULTS

The regression results for the six regression equations for each of the seven stable land-cover classes are statistically meaningful. Results for the stable agriculture class are shown in Table 3 and are similar to those obtained for the other six classes.

| | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 7 |
|---|---|---|---|---|---|---|
| AOD | 0.05<br>**(42.0)** | 0.01<br>**(8.4)** | 0.01<br>**(4.6)** | **-0.02**<br>**(-2.2)** | **-1.98**<br>**(-7.1)** | 0.02<br>**(11.0)** |
| SZA | **359.17**<br>**(36.4)** | **-54.24**<br>**(-8.1)** | **-96.24**<br>**(-7.2)** | **-142.46**<br>**(-1.8)** | **1390.57**<br>**(8.1)** | **39.88**<br>**(2.5)** |
| MINDN | 0.69<br>**(16.6)** | -0.07<br>(-1.6) | **0.39**<br>**(6.2)** | **0.72**<br>**(3.3)** | **350.75**<br>**(7.1)** | |
| JAN | **-53.95**<br>**(-37.4)** | **-9.47**<br>**(-11.5)** | **-14.66**<br>**(-8.5)** | **16.59**<br>**(2.1)** | **1201.29**<br>**(7.2)** | **-17.05**<br>**(-8.6)** |
| MARCH | **-35.28**<br>**(-35.7)** | **7.14**<br>**(11.6)** | **10.59**<br>**(8.9)** | **3.32**<br>(0.5) | **194.27**<br>**(6.3)** | -1.23<br>(-0.8) |
| OCT | **-30.19**<br>**(-34.0)** | **3.27**<br>**(5.2)** | **6.29**<br>**(4.9)** | **18.06**<br>**(2.4)** | **219.64**<br>**(6.4)** | **-5.75**<br>**(-4.0)** |
| NOV | **-29.05**<br>**(-34.6)** | **4.72**<br>**(10.8)** | **8.34**<br>**(9.3)** | **8.43**<br>**(1.4)** | **397.49**<br>**(6.8)** | **-2.58**<br>**(-2.5)** |
| DEC | **17.50**<br>**(26.8)** | **-1.91**<br>**(-6.4)** | **-2.17**<br>**(-4.3)** | **-7.35**<br>**(-3.7)** | **404.39**<br>**(7.2)** | **6.04**<br>**(9.8)** |
| $r^2$ | 0.27 | 0.13 | 0.12 | 0.12 | 0.03 | 0.16 |
| t statistics in parenthesis; values in bold exceed the p<.05 threshold | | | | | | |

Table 3. SUR Regression Results for the Stable Agriculture Land-Cover Class

The regression coefficients associated with the independent variables generally are statistically different from zero ($p < .05$) as indicated by a t test. The sign on the regression coefficients cannot be used to evaluate their consistency with theory. The images are corrected for atmospheric effects therefore, the regression coefficients represent the effect of errors in these corrections and the effects that are not corrected fully by the techniques for atmospheric correction. The r-squared for all equations ranges from 0.03 to 0.64. The lower range of values is consistent with the transformation used to calculate the fixed effects estimator.

The fraction of change pixels for which the econometric technique identifies a date of change, which is termed the percent classified, is negatively related to the threshold used to determine the statistical significance of differences between competing models (Table 4). No pixels are classified when a rigorous threshold is used ($t=2.4$, $p < .05$). As the threshold is lowered to $p < .1$, the fraction classified rises, and exceeds 95 percent when the threshold is lowered to $p < .50$.

| Threshold | Percent Classified | Exactly Correct | Possibly Correct |
|---|---|---|---|
| 1.95 (p < 0.1) | 54 | 57 | 84 |
| 1.75 (p < 0.14) | 75 | 47 | 71 |
| 1.50 (p < 0.19) | 82 | 50 | 75 |
| 1.25 (p < 0.26) | 86 | 50 | 74 |
| 1.00 (p < 0.36) | 90 | 50 | 73 |
| 0.75 (p < 0.49) | 96 | 50 | 72 |
| 0.50 (p < 0.64) | 97 | 51 | 71 |
| 0.25 (p < 0.81) | 98 | 51 | 66 |
| 0.10 (p <0.92) | 99 | 51 | 64 |
| Values in bold exceed the p < 0.05 threshold | | | |

Table 4. Results for the Econometric Time Series Technique

The econometric method's accuracy is evaluated two ways. One measures the fraction of pixels classified for which the methodology identifies the date of change correctly. Depending on the threshold used to determine statistical significance, the econometric technique correctly identifies the date of change for 47 to 57 percent of the pixels classified.

The accuracy also can be measured by the fraction of pixels that *may* be classified correctly. As described previously, all pixels in a site are assigned the same date of change, although the pixels in 39% of the sites do not change simultaneously. This staggered change implies that a pixel's date of change may be later than the date assigned to all pixels in the site. Under these conditions, the focus on a single date of change may systematically understate each method's accuracy. To account for this bias, a second measure of accuracy is calculated. If a pixel is assigned a date of change that occurs after the date assigned to the entire site, and if our visual examination of that site indicates that the pixels do not change simultaneously, the pixel's date of change is considered to be identified correctly. Defined this way, this measure gives an upper bound on each method's accuracy. Using this measure of accuracy, the econometric technique identifies the correct date of change for 64 – 84 percent of the pixels classified.

## 5 DISCUSSION

The ability to generate statistically meaningful estimates for Equations 1-6 implies that the method used to calibrate the images fails to eliminate differences associated with solar zenith angle and aerosol optical depth. The regression coefficients for these variables would be statistically insignificant if the 'ridge' correction technique eliminated systematic differences in the reflectivity among images. This failure also is indicated by the statistically significant regression coefficients associated with the minimum DN values and the dummy variables for the month of acquisition.

This failure may be obviated by the econometric technique. Many techniques for atmospheric correction to calibrate images use a linear transformation of the uncorrected image. This linear transformation can be viewed as a shift in the mean value for each band for each pixel. Such a shift is similar to the transformation used to calculate the fixed effect estimator. This similarity implies that the regression coefficients estimated by the fixed effect estimator are unaffected by atmospheric correction. Under these conditions, the econometric technique may be able to identify the dates of change from a series of Landsat images without 'correcting' them for atmospheric effects.

The econometric technique also may reduce the effort required to assemble the data that are needed for supervised classification. The econometric technique is trained on the stable land-cover classes only. For many applications, there are more change classes than stable classes. In this study, there are seven stable classes and sixteen change classes. Under these conditions, less effort is required to assemble the training data for the econometric methodology than more conventional change detection techniques that require training data for both stable and change classes.

## 6 CONCLUSION

Econometric techniques rarely are used to process satellite images because many of their underlying assumptions are not be consistent with the data collected by the sensor. For this application, these inconsistencies appear to have relatively little effect on accuracy or bias relative a more traditional methodology. Indeed, the time series technique may alleviate bias, which may be more important than accuracy when the data generated from remotely sensed images are used in statistical analyses that seek to identify the causes of change in land cover/use. Nonetheless, the econometric technique is not designed to replace existing techniques—it cannot be used to classify an image or a time series of images. Rather, the time series technique is designed to identify the date of land-use change, or another form of land conversion, from a time series of images after another technique is used to identify pixels where land-use changes between the first and last image of the time series.

## REFERENCES

Diebold, F. X., Mariano, R. S., 1995. Comparing Predictive Accuracy. Journal of Business and Economics Statistics. 13, 253-263.

Hsiao, C., 1986. Analysis of Panel Data. Cambridge University Press. Cambridge.

Song, C., Woodcock, C. E., Seto, K. C., Pax Lenney, M., and Macomber, S. A., 2001. Classification and Change Detection Using Landsat TM Data - When and How to Correct Atmospheric Effects? *Remote Sensing of Environment* **75**(2): 230-244.