ELSEVIER

Agriculture, Ecosystems and Environment 85 (2001) 95–105

Agriculture Ecosystems & Environment

www.elsevier.com/locate/agee

Change detection, accuracy, and bias in a sequential analysis of Landsat imagery in the Pearl River Delta, China: econometric techniques

Robert K. Kaufmann [a,*], Karen C. Seto [a,b]

[a] *Department of Geography, Center for Energy and Environmental Studies, Boston University, Boston, MA 02215, USA*
[b] *Center for Environmental Science and Policy, Institute for International Studies, Stanford University, Stanford, CA 94305-6055, USA*

Abstract

Time series data from high resolution satellite imagery provide researchers with an opportunity to develop sophisticated statistical models of land-cover change. As inputs to statistical models, land-cover change data that are generated from satellite imagery must be both accurate and unbiased. This paper describes a new change detection method to determine the date of land-cover change in a sequential series of Landsat TM images of the Pearl River Delta, China. The method is a three-step change detection procedure that uses time series and panel econometric techniques. In the first step, regression equations are estimated for each of the six DN bands for each of seven stable land-cover classes. In the second step, the 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 (1989–1996). In the third step, the date of land-cover change is identified by comparing a pixel's DN values against the eight possible dates of change using tests for predictive accuracy. The accuracy and bias of the dates of change identified by the econometric technique compare favorably to a more conventional change detection technique. Furthermore, the econometric technique may reduce efforts required to assemble the training data and to correct the images for atmospheric effects. Together, these results indicate that is possible to generate land-use change estimates from a time series of satellite images that can be used in conjunction with socioeconomic data to estimate statistical models of land-use change.

Keywords: Land-cover change detection; Time series techniques; Landsat TM; Accuracy; Bias

1. Introduction

Understanding of the relation between human activity and land-use/cover can be enhanced by statistically coupling time series of satellite images with time series of socioeconomic data. The time series for many socioeconomic variables are available from a variety of sources (e.g. national yearbooks) at an annual frequency. To use these data in statistical analyses, land-cover estimates also must also be available at an annual frequency. This presents a methodological difficulty. Most change techniques are designed to analyze two images. Computationally, repeated applications of these techniques can introduce errors associated with post-classification comparison of images.

To couple land-cover information with economic activity, it is important to assess the date at which land-cover/use changes occur. If a change detection

* Corresponding author. Tel.: +1-617-353-3940; fax: +1-617-353-5986.
E-mail addresses: kaufmann@bu.edu (R.K. Kaufmann), kseto@stanford.edu (K.C. Seto).

0167-8809/01/$ – see front matter
PII: S0167-8809(01)00190-6

technique systematically identifies the date of change prior to actual date of change, the time series for changes in land cover/use will be biased. Such bias will undermine the interpretation of statistical results. For example, a regression equation that specifies land-use change as a function of current and lagged rates of economic activity will indicate no relation between these variables if the change detection technique consistently identifies a date of change prior to the actual date of land conversion.

This paper describes a method that uses econometric techniques to determine the date at which land-cover changes between seven classes in nine images (1988–1996) of the Pearl River Delta, China. Section 2 describes existing change detection methods. Section 3 describes the data sets that are used to assess the new technique. Section 4 describes the methodology. Results from comparison tests with a more conventional change detection technique are presented in Section 5 and a discussion of the results is presented in Section 6.

2. Existing change detection techniques

Numerous change detection techniques are available which achieve different levels of success in monitoring a variety of land-cover changes. These techniques include image regression, image subtraction, post-classification comparison, multidate principal components analysis, multidate Tasseled Cap transformation, change vector analysis, and neural networks (Singh, 1989; Fung, 1990; Lambin and Strahler, 1994; Collins and Woodcock, 1994; Gopal and Woodcock, 1996; Dai and Khorram, 1999). The method used depends largely on the landscape of the study area, the types of land-cover changes, and the temporal and spatial resolution of the data. As a result, there is no consensus regarding the 'best' technique.

Most change detection studies evaluate change between two periods (Howarth and Boasson, 1983; Green et al., 1994; Kwarteng and Chavez, 1998; Mas, 1999). While studies that use more than two dates of imagery exist (Jensen et al., 1995; Pax Lenney et al., 1996; Collins and Woodcock, 1996), most do not use consecutive dates, nor do they extract annual estimates of land-use or land-cover. Furthermore, analyses that use more than two images are largely limited to AVHRR data (Eastman and Fulk, 1993; Barbosa et al., 1999; Lambin, 1996).

It is well recognized that the accuracy of a change map made from two separate classifications will be at best the product of the two individual classifications (Stow et al., 1980). For example, if each individually classified map has an accuracy of 90%, the post-classification change map may only have an 81% accuracy. To evaluate land-cover changes in a series of images, repeated application of a change detection algorithm may also yield similar results. Therefore, it is important to develop a methodology that processes all the images simultaneously rather than sequentially or in a pair-wise fashion.

3. Data and data preprocessing

For this study, we acquire nine predominantly cloud-free Landsat TM images of the Pearl River Delta in Southern China (WRS2 path 122, row 44). Cloud-free images are not available throughout the year because the Delta is located in the tropics, between 21°N and 23°N, which is mostly cloudy during the rainy season. As a result, the images are obtained from October to March (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 m by 30 m pixels using a nearest neighbor resampling algorithm with a first-order polynomial. The number of ground control points (GCPs) 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). Other techniques are available (e.g. Schott et al., 1998; Hall et al., 1991) but they are not suitable for images of the Pearl River Delta because there is a dearth of well defined spectrally stable dark and bright ground features.

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 16 change land-cover classes

Table 1
Characteristics of TM images used in the study

Acquisition date	Landsat satellite	Number of GCPs[a] for registration	Geometric registration RMSE[b]
10 December 1988	4	25	±0.2893
13 December 1989	4	23	±0.2914
30 October 1990	5	18	±0.2364
2 February 1991	5	18	±0.2412
20 January 1992	5	16	Master image
24 December 1993	5	20	±0.2618
8 November 1994	5	17	±0.3162
30 December 1995	5	24	±0.2970
3 March 1996	5	25	±0.2776

[a] Ground control points.
[b] Root mean square error of x,y coordinate.

(Table 2). The stable classes are fairly 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. Ideally, the 4766 change pixels (469 sites) would be assigned a date of change through fieldwork and interviews with local residents. This is not possible because many residents recently migrated from other provinces in China and are not familiar with local land-use history. Because of this difficulty, date of change are assigned 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.

Table 2
Stable and change land-cover classes, and number of training sites for each class

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		

Data for the 469 change sites are separated randomly so that 80% of the sites are used for training while the remaining 20% are used for testing. This separation is repeated to create five data sets in which all pixels for each site appear in the training portion of four data sets and in the testing portion of one data set. Separating pixels from individual sites between training and testing allows us to perform a true out-of-sample evaluation of the change detection techniques.

4. Change detection methodology

The econometric change detection technique uses time series and panel techniques to identify the date of change for individual pixels in three-steps. In the first step, regression equations are estimated for each of the six DN 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 (Fig. 1). 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.

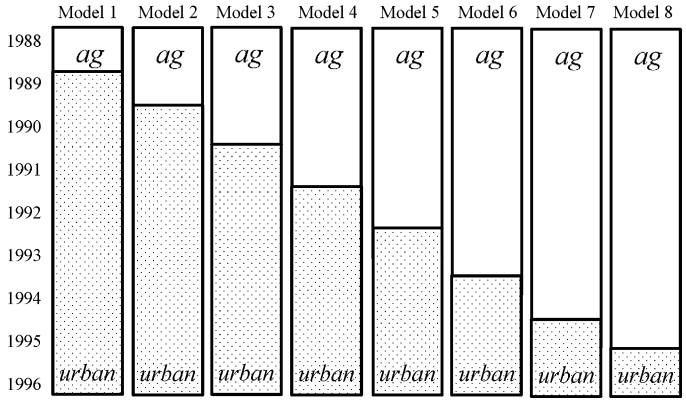

Fig. 1. Models of land-use change (agriculture to urban) for each of eight possible dates of change.

4.1. Modeling constant land cover classes using panel techniques

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

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

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

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

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

$$\mathrm{DN}^*_{5tj} = \beta X^*_t + \gamma Y^*_t + \mu_{5tj} \quad (5)$$

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

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 ($j = 1, \ldots, 7$), X is a vector (3×9) 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 (6×9) 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 0 for the other five dummy variables), β and γ are vectors of regression coefficients, and μ are a time series of error terms.

Eqs. (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 (Hsaio, 1986). For example, the transformed DN value for Band 1 is given by $(\mathrm{DN}_{1tj} - \overline{\mathrm{DN}}_{1j})$ where

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

in which N is the number of observations per pixel (9). The independent variables are transformed using a similar procedure, as indicated by $*$ in Eqs. (1)–(6). Because of these transformations, Eqs. (1)–(6) do not contain an intercept.

Eqs. (1)–(6) can be estimated individually using OLS, but the regression results may be inefficient because the errors (μ) 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, Eqs. (1)–(6) are estimated as a system of seemingly unrelated regressions (SURs). SUR is a generalized least squares estimator that allows for correlation among the off-diagonal elements of the variance–covariance matrix (Zellner, 1962). If there is no correlation

among error terms, the SUR estimate converges to the OLS estimate.

This procedure is repeated for each of the seven stable land-cover classes. The estimates for β and γ along with the values for X and Y variables are used to calculate the transformed DN value for each stable land cover class for each image. The transformed values are converted back to levels using Eq. (7).

4.2. Calculating models for all possible dates of change between stable land-cover classes

The DN values for each stable land cover class are combined to calculate DN values for each of the 16 change classes (Table 2). The DN values for the change class *agriculture to urban* are generated by combining the DN values generated by the regression equations for the stable *agriculture* class with the DN values generated by the regression equations for the stable *urban* class (Fig. 1).

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 change for the *agriculture to urban* class, the DN values generated by the *agriculture* regression equations for 1988 and 1989 are combined with the DN values generated by the *urban* regression equations for the years 1990–1996 (Fig. 1). This process is repeated to generate a model for the *agriculture to urban* 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 16 land-cover change classes.

4.3. Using tests of predictive accuracy to identify the date of change

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 *agriculture to urban* class, its DN values are compared to values generated by the stable *agriculture* and *urban* 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

$$d_{it} = |\mathrm{DN}_{it} - \widehat{\mathrm{DN}}_{itj}| - |\mathrm{DN}_{it} - \widehat{\mathrm{DN}}_{itj+1}| \tag{8}$$

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

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

$$I_+(d_t) = \begin{cases} 1, & \text{if } d_t > 0, \\ 0, & \text{otherwise} \end{cases}$$

to calculate the $S_{3\mathrm{a}}$ statistic which can be compared against a t distribution.

The $S_{3\mathrm{a}}$ test statistic weighs the value of d using a binary choice (zero or one), which 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 $S_{3\mathrm{a}}$ test statistic. The test statistic will be negative if the absolute forecast error associated the model that represents the change in land cover at date j is smaller than the absolute forecast error associated the model that represents the change in land cover 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 $S_{3\mathrm{a}}$ statistic that exceeds a threshold for statistical significance.

The threshold for statistical significance that is used to distinguish among competing models 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., $p < 0.05$) can differentiate between competing dates of change with a high degree of confidence. But for many pixels, the $S_{3\mathrm{a}}$ statistic may not be able to differentiate competing models with a high degree of confidence. Under these conditions, the methodology cannot 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. To find the optimal trade-off between classification coverage and accuracy, results using several threshold for statistical significance are compared.

4.4. Multidate Tasseled Cap technique

To assess the performance of the econometric technique, its results are compared to those generated by a more conventional remote sensing approach, a multidate Tasseled Cap technique. This technique requires two steps. In the first step, a two-date classification and change detection of the 1989–1996 images in comparison to the 1988 image is performed. In the second step, the classified maps are stacked to identify the date of change.

The Tasseled Cap transformation has been used to monitor changes in land-cover (Crist and Cicone, 1984; Fung, 1990; Collins and Woodcock, 1996). It rotates TM data to creates three planes: brightness (B), greenness (G), and wetness (W). For each of the five training sets described in Section 3, BGW values for the 1988 and 1996 images are calculated, and then the 1988 BGW values are subtracted from the 1996 BGW values to create change vectors. Next, the 1988 BGW values and the 1996–1988 change vectors are composited to create a six-band training set to calculate multivariate statistics for each of the stable and change classes. The training statistics generated from 1988 BGW and 1996–1988 BGW values are then used to classify the testing data with a Bayesian maximum likelihood classifier. Details of the multidate BGW technique are documented in Seto et al. (in press).

For each of the five testing data sets, we first identify training sites where land-cover changes between 1988 and 1996. The selection of these sites is based on a combination of visual interpretation of the sites by an analyst who is familiar with the study area and ground-based field assessments. For sites that change between 1988 and 1996, BGW values for the 1988 image and change vectors are calculated by subtracting the 1988 BGW values from BGW values for each subsequent image. For example, the 1990 testing data consists of a six-band composite of 1988 BGW and 1990–1988 BGW values for sites which change by 1996. These sites are not expected to have changed in 1990, but rather they were changed by 1996. Because the objective is to analyze the date of change in the series of images, we first identify sites that change by 1996, and evaluate the date at which the pixels within these sites change. For example, a pixel identified as agriculture to urban in the 1988–1996 analysis is evaluated in the intervening years with class membership limited to either agriculture or urban. Using this technique, two-date change maps are created for each of the eight pairs of dates from 1989 to 1996 compared to 1988.

After classification of all dates of imagery in the testing data, the date of change is determined by stacking the two-date change maps and identifying change in two consecutive years. That is, if in time t, a pixel is identified as having changed from its original class in 1988, but not changed in $t + 1$, then t is not the date of change. Only when both t and $t + 1$ are evaluated as change is t identified as the date of change for a pixel. The date of change is assigned only when a pixel has changed in two consecutive years rather than the first year in which a pixel has changed to reduce errors due to misclassification of a pixel as change when it actually has not changed.

5. 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. The regression coefficients associated with the independent variables generally are statistically different from zero ($p < 0.05$) 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^2 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.

Table 3
SUR regression results for the stable agriculture land cover class[a]

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
AOD	0.05 (*42.0*)	0.01 (*8.4*)	0.01 (*4.6*)	*−0.02* (*−2.2*)	*−1.98* (*−7.1*)	0.02 (*11.0*)
SZA	*359.17* (*36.4*)	*−54.24* (*−8.1*)	*−96.24* (*−7.2*)	*−142.46* (*−1.8*)	*1390.57* (*8.1*)	*39.88* (*2.5*)
Min DN	0.69 (*16.6*)	−0.07 (−1.6)	*0.39* (*6.2*)	*0.72* (*3.3*)	*350.75* (*7.1*)	
January	*−53.95* (*−37.4*)	*−9.47* (*−11.5*)	*−14.66* (*−8.5*)	*16.59* (*2.1*)	*1201.29* (*7.2*)	*−17.05* (*−8.6*)
March	*−35.28* (*−35.7*)	*7.14* (*11.6*)	*10.59* (*8.9*)	*3.32* (0.5)	*194.27* (*6.3*)	−1.23 (−0.8)
October	*−30.19* (*−34.0*)	*3.27* (*5.2*)	*6.29* (*4.9*)	*18.06* (*2.4*)	*219.64* (*6.4*)	*−5.75* (*−4.0*)
November	*−29.05* (*−34.6*)	*4.72* (*10.8*)	*8.34* (*9.3*)	*8.43* (*1.4*)	*397.49* (*6.8*)	*−2.58* (*−2.5*)
December	*17.50* (*26.8*)	*−1.91* (*−6.4*)	*−2.17* (*−4.3*)	*−7.35* (*−3.7*)	*404.39* (*7.2*)	*6.04* (*9.8*)
r^2	0.27	0.13	0.12	0.12	0.03	0.16

[a] *t*-Statistics in parenthesis. Values in italic exceed the $p < 0.05$ threshold.

Table 4
Results for the econometric time series technique[a]

Threshold	Percent classified	Exactly correct	Possible correct	Mean, $\bar{\mu} = 0$	Skew, Sk = 0	Kurtosis, Ku = 0
1.95 ($p < 0.1$)	54	57	84	*−0.39*	*−1.07*	*7.20*
1.75 ($p < 0.14$)	75	47	71	*−0.46*	*−0.51*	*5.22*
1.50 ($p < 0.19$)	82	50	75	*−0.29*	0.07	*6.11*
1.25 ($p < 0.26$)	86	50	74	*−0.22*	*0.77*	*6.99*
1.00 ($p < 0.36$)	90	50	73	*−0.13*	*0.94*	*6.31*
0.75 ($p < 0.49$)	96	50	72	*−0.08*	*0.67*	*4.83*
0.50 ($p < 0.64$)	97	51	71	*0.19*	*1.30*	*3.91*
0.25 ($p < 0.81$)	98	51	66	*0.52*	*1.21*	*2.01*
0.10 ($p < 0.92$)	99	51	64	*0.63*	*1.16*	*1.44*

[a] Values in italic exceed the $p < 0.05$ threshold.

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 ($t = 2.4$, $p < 0.05$). As the threshold is lowered to $p < 0.1$, the fraction classified rises, and exceeds 95% when the threshold is lowered to $p < 0.50$. There also is an negative relation between the fraction classified and the stringency of the threshold in the multidate BGW technique (Table 5).

Table 5
Results for the multidate BGW technique[a]

Threshold	Percent classified	Exactly correct	Possible correct	Mean, $\bar{\mu} = 0$	Skew, Sk = 0	Kurtosis, Ku = 0
5	52	10	47	*−2.14*	*−0.57*	*0.56*
6	63	12	48	*−2.00*	*−0.59*	*0.63*
7	71	13	49	*−1.84*	*−0.44*	*1.03*
8	76	15	50	*−1.65*	*−0.34*	*1.27*
9	81	16	49	*−1.51*	*−0.33*	*1.30*
10	85	19	51	*−1.38*	*−0.33*	*1.34*
20	96	30	56	*−0.61*	*0.13*	*1.64*
30	98	36	59	*−0.23*	*0.52*	*1.44*
40	99	38	59	*0.09*	*0.72*	*1.02*
∞	100	56	71	*0.64*	*1.06*	*1.39*

[a] Values in italic exceed the $p < 0.05$ threshold.

The econometric and BGW methods' 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 47–57% of the pixels classified. The multidate BGW method correctly identifies the date of change for 10–56% of the pixels classified. The BGW methods decline in accuracy as the threshold is made more stringent may be associated with how the Bayesian classifier classifies a pixel. Only when a pixel has a very high probability of belonging to a change class is it classified as change. Under these circumstances, it is likely that the analyst will 'see' change in the image before the algorithm detects change. In this case, with a stringent criteria, the algorithm only assigns a pixel as change when it is very certain it has changed, which is later than when the analyst detects change.

The accuracy of the two methods 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 methods 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% of the pixels classified, while the accuracy of the multidate BGW technique is 47–71%.

To evaluate bias in the errors generated by the two methodologies, a series (μ) is compiled that measures the size of the error which varies between 8 and -8. A value of zero is assigned to (μ) for pixels that the change detection technique correctly identifies the exact date of change. A value of -2 is assigned to (μ) for pixels that the change detection technique assigns a date of change 2 years after the actual change. A value of 3 is assigned to (μ) for pixels that the methodology assigns a date of change 3 years prior to the actual date of change. To eliminate the systematic bias associated with the staggered changes in land cover within sites, the series (μ) does not include errors from pixels located in sites where change is not simultaneous and the methodology assigns a date of change after the date associated with that site.

The series (μ) is used to investigate three components of bias: mean, skew, and coefficient of kurtosis. The mean value of the errors ($\bar{\mu}$) is calculated using Eq. (10)

$$\bar{\mu} = \frac{\sum_{i=1}^{N} \mu_i}{N} \tag{10}$$

in which N is the number of pixels classified. For all thresholds and both techniques, $\bar{\mu}$ is statistically different from zero (Table 4, $p < 0.05$). For both techniques, rigorous thresholds tend to identify a date of change after the actual date (negative mean error), while less rigorous thresholds tend to identify a date of change prior to the actual occurrence (positive mean error).

The decline in the absolute size of the errors around the mean (skew) is measured as follows:

$$\mathrm{Sk} = \frac{N \sum_{i=1}^{N} (\mu_i - \bar{\mu})^3}{(N-1)(N-2)\sigma^3} \tag{11}$$

in which σ is the standard error of the series μ. If the skew is consistent with a normal distribution, $\mathrm{Sk} = 0$. If the number of errors drops off more slowly on the positive side of mean, the errors are said to be skewed to the right ($\mathrm{Sk} > 0$). Alternatively, if the number of errors drops off more slowly on the negative side of the mean, the errors are said to be skewed to the left ($\mathrm{Sk} < 0$).

The skew is evaluated with a z-statistic that tests the null hypothesis $\mathrm{Sk} = 0$. The skew is affected by the choice of threshold. Using a threshold of 1.5 ($p < 0.19$), the econometric technique generates errors that have a skew consistent with a normal distribution (Table 4). More rigorous thresholds are associated with a positive skew while lower thresholds are associated with a negative skew. A similar pattern appears in the skew of the errors generated by the multidate BGW method (Table 5).

The degree to which the dates of change are near the actual date can be identified by calculating the coefficient of kurtosis. One measure for the coefficient of kurtosis is calculated as follows:

$$\mathrm{Ku} = \frac{(N+1)\sum_{i=1}^{N}(\mu_i - \bar{\mu})^4 - 3(n-1)(\sum_{i=1}^{N}(\mu_i - \bar{\mu})^2))^2}{(N-1)(N-2)(N-3)\sigma^4} \quad (12)$$

in which the variables are as defined above. If the coefficient of kurtosis is consistent with a normal distribution, Ku = 0. The null hypothesis (Ku = 0) can be evaluated with a z-statistic. For both the econometric and multidate BGW method, Ku > 0 for all thresholds. This indicates that the dates of change identified by the methodology are concentrated near the correct date of change relative to a normal distribution.

6. Discussion

The ability to generate statistically meaningful estimates for Eqs. (1)–(6) implies that the method used to calibrate the images fails to eliminate differences associated with SZA and AOD. 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 Min 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 use a linear transformation of the uncorrected images. 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 while the multidate BGW method is trained on both the stable and change land-cover classes. For many applications, there are more change classes than stable classes. In this study, there are seven stable classes and 16 change classes. Under these conditions, less effort is required to assemble the training data for the econometric methodology than the multidate BGW analysis.

The results indicate that analysts should chose the threshold used by the econometric or BGW techniques to identify the date of change based on the purpose of the analysis. Because the time series for changes in land-cover will be used as dependent variables in a statistical analysis of the socioeconomic drivers, the estimates must be relatively free of bias. Based on the results described above, a threshold between 0.75 ($\bar{\mu} = -0.08$) and 1.5 (Sk = 0.07) would generate the most satisfactory results (Fig. 2). Although these thresholds fail to classify 4–18% of the pixels, these omissions will have little effect on the results of the statistical analysis of the socioeconomic drivers if the date of change associated with the pixels not classified is distributed randomly among the eight possible dates of change. For the multidate BGW, a threshold of ∞, which classifies all pixels, would generate the most satisfactory results. The ∞ threshold generates a higher accuracy than a threshold of 40, but the bias associated with the ∞ threshold is more severe than the 40 threshold.

Although the two techniques are used to analyze the same data, it is difficult to compare their accuracy and bias because of differences in the 'out-of-sample' analysis. The multidate BGW analysis is trained on pixels from a subset of stable and change sites which are assembled from the 1988 and 1996 images. No information is included from intervening images (1989–1995). This constitutes the out-of-sample testing of the BGW technique. Changes in atmospheric and seasonal conditions in the intervening images are represented explicitly by the econometric methodology. For the econometric methodology, there are no equations for change classes. Applying results obtained from stable classes to change classes constitutes the out-of-sample testing of the econometric technique.

The differences in the out-of-sample testing may be partially responsible for the differences in the two methodology's predictive accuracy. One cause for the large fraction of pixels with multiple dates of change that is identified by the multidate BGW analysis may

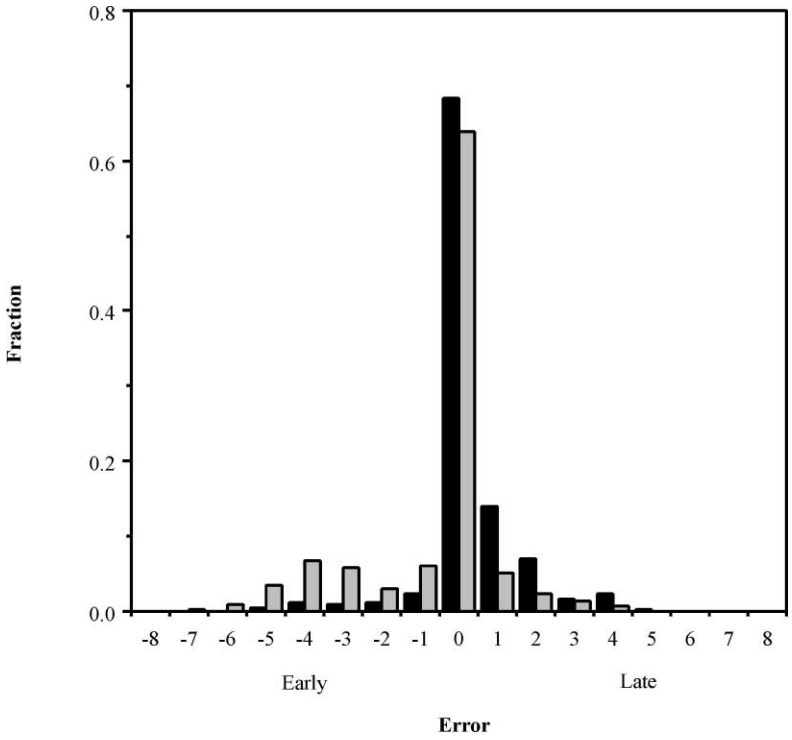

Fig. 2. Distribution of errors generated by the econometric technique using a threshold of 1.5 (darkened bars) and the errors generated by the multidate BGW technique using a threshold of ∞ (shaded bars). Correct values are assigned a value of zero. Fraction is the fraction of pixels classified.

be the lack of information about atmospheric or seasonal changes in the 1989–1995 images. This lack of information should be minimized by the techniques for atmospheric correction, but as described previously, the regression results for Eqs. (1)–(6) imply that the ridge correction technique does not eliminate these effects.

The results generated by the econometric technique generally seem to have less bias than the results generated by the multidate BGW technique. The mean value of the errors generated by the econometric technique generally is closer to zero and the coefficient of kurtosis generally is greater than the multidate BGW technique. The reduction in bias is consistent with the general difference between the two techniques. The econometric technique evaluates the probability of a change in land cover class for each of the eight possible dates and chooses the date at which the change is most likely to have occurred. This reduces the bias that is introduced by the multidate BGW technique, which individually compares the 1989–1996 images to the 1988 image and determines the date of change based on the first year in which two consecutive comparisons indicate a change in land cover.

7. 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 change from a time series of images after some other technique, such as a two date BGW technique, is used to identify pixels where land cover-changes between the first and last image of the time series.

Acknowledgements

The authors would like to thank Curtis Woodcock for comments on an earlier draft of the paper. This research was supported by NASA's Land Cover Land Use Change Program, grant NAG5-6214.

References

Barbosa, P.M., Gregoire, J.-M., Pereira, J.M.C., 1999. An algorithm for extracting burned areas from time series of AVHRR GAC data applied at a continental scale. Rem. Sen. Environ. 69, 253–263.

Collins, J.B., Woodcock, C.E., 1994. Change detection using the Gramm-Schmidt transformation applied to mapping forest mortality. Rem. Sen. Environ. 50, 267–279.

Collins, J.B., Woodcock, C.E., 1996. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. Rem. Sen. Environ. 56, 66–77.

Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data — the TM Tasseled Cap. IEEE Trans. Geosci. Rem. Sen. 22, 256–263.

Dai, X.L., Khorram, S., 1999. Remotely sensed change detection based on artificial neural networks. Photogramm. Eng. Rem. Sen. 65, 1187–1194.

Diebold, F.X., Mariano, R.S., 1995. Comparing Predictive Accuracy. J. Bus. Econ. Statist. 13, 253–263.

Eastman, R.J., Fulk, M., 1993. Long sequence time series evaluation using standardized principal components. Photogramm. Eng. Rem. Sen. 59, 1307–1312.

Fung, T., 1990. An assessment of TM imagery for land-cover change detection. IEEE Trans. Geosci. Rem. Sen. 28, 681–684.

Gopal, S., Woodcock, C.E., 1996. Remote sensing of forest change using artificial neural networks. IEEE Trans. Geosci. Rem. Sen. 34, 398–404.

Green, K., Kempka, D., Lackey, L., 1994. Using remote sensing to detect and monitor land-cover and land-use change. Photogramm. Eng. Rem. Sen. 60, 331–337.

Hall, F.G., Strebel, D.E., Nickeson, J.E., Goetz, S.J., 1991. Radiometric rectification: toward a common radiometric response among multidate, multisensor images. Rem. Sen. Environ. 35, 11–27.

Howarth, P.J., Boasson, E., 1983. Landsat digital enhancements for change detection in urban environments. Rem. Sen. Environ. 13, 149–160.

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

Jensen, J.R., Rutchey, K., Koch, M.S., Narumalani, S., 1995. Inland wetland change detection in the everglades water conservation area 2A using a time series of normalized remotely sensed data. Photogramm. Eng. Rem. Sen. 51, 199–209.

Kwarteng, A.Y., Chavez Jr., P.S., 1998. Change detection study of Kuwait City and environs using multi-temporal Landsat Thematic Mapper data. Int. J. Rem. Sen. 19, 1651–1662.

Lambin, E.F., 1996. Change detection at multiple temporal scales: seasonal and annual variations in landscape variables. Photogramm. Eng. Rem. Sen. 62, 931–938.

Lambin, E.F., Strahler, A.H., 1994. Change-vector analysis in multitemporal space: a tool to detect and categorize land-cover change processes using high temporal-resolution satellite data. Rem. Sen. Environ. 48, 231–244.

Mas, J.-F., 1999. Monitoring land-cover changes: a comparison of change detection techniques. Int. J. Rem. Sen. 20, 139–152.

Pax Lenney, M., Woodcock, C.E., Collins, J.B., 1996. The status of agricultural lands in Egypt: the use of multitemporal NDVI features derived from Landsat TM. Rem. Sen. Environ. 56, 8–20.

Schott, J.R., Salvaggio, C., Volchok, W.J., 1998. Radiometric scene normalization using psuedoinvariant features. Rem. Sen. Environ. 26, 1–16.

Seto, K.C., Woodcock, C.E., Song, C., Huang, X., Lu., J., Kaufmann, R.K., in press. Monitoring land-use change in the Pearl River Delta using Landsat TM. Int. J. Rem. Sen.

Singh, A., 1989. Digital change detection techniques using remotely-sensed data. Int. J. Rem. Sen. 10, 989–1003.

Song, C., Woodcock, C.E., Seto, K.C., Pax Lenney, M., Macomber, S.A., 2001. Classification and change detection using Landsat TM data: when and how to correct the atmospheric effects? Rem. Sen. Environ. 75, 230–244.

Stow, D.A., Tinney, L.R., Estes, J.E., 1980. Deriving land use/land cover change statistics from Landsat: a study of prime agricultural land. In: Proceedings of the 14th International Symposium on Remote Sensing of the Environment held in Ann Arbor in 1980, Ann Arbor, Environmental Research Institute of Michigan, Michigan, pp. 1227–1237.

Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J. Amer. Statist. Assoc. 57, 348–368.