Review Article

Digital change detection methods in ecosystem monitoring: a review

P. COPPIN*, I. JONCKHEERE, K. NACKAERTS, B. MUYS
Geomatics and Forest Engineering Group, Department of Land Management, Katholieke Universiteit Leuven, Vital Decosterstraat 102, B-3000 Leuven, Belgium

and E. LAMBIN
Department of Geography, Université Catholique de Louvain, 3 Place Pasteur, B-1348 Louvain-la-Neuve, Belgium

Abstract. Techniques based on multi-temporal, multi-spectral, satellite-sensor-acquired data have demonstrated potential as a means to detect, identify, map and monitor ecosystem changes, irrespective of their causal agents. This review paper, which summarizes the methods and the results of digital change detection in the optical/infrared domain, has as its primary objective a synthesis of the state of the art today. It approaches digital change detection from three angles. First, the different perspectives from which the variability in ecosystems and the change events have been dealt with are summarized. Change detection between pairs of images (bi-temporal) as well as between time profiles of imagery derived indicators (temporal trajectories), and, where relevant, the appropriate choices for digital imagery acquisition timing and change interval length definition, are discussed. Second, pre-processing routines either to establish a more direct linkage between remote sensing data and biophysical phenomena, or to temporally mosaic imagery and extract time profiles, are reviewed. Third, the actual change detection methods themselves are categorized in an analytical framework and critically evaluated. Ultimately, the paper highlights how some of these methodological aspects are being fine-tuned as this review is being written, and we summarize the new developments that can be expected in the near future. The review highlights the high complementarity between different change detection methods.

1. Introduction

Ecosystems are in a state of permanent flux at a variety of spatial and temporal scales all around the world. Causes of these fluxes can be natural as well as anthropogenic, or may be a combination of the two. Moreover, scientific evidence clearly points to the fact that impacts of, for example, global change on land surface attributes are not uniformly distributed on the face of the Earth. The fact that sustainability has become a primary objective in present-day ecosystem management has as one of its consequences the continuous need for accurate and
up-to-date resource data. What is more, where it concerns large-area processes, any in-depth understanding of the changes has to be based on an accurate monitoring of land surface attributes over at least a few decades. At such regional scales, monitoring poses a number of methodological challenges. The lack of quantitative, spatially explicit and statistically representative data on ecosystem change has left the door open to simplistic representations, e.g. the advance of deserts (Lamprey 1975). While it is possible to find local examples of such extreme changes, empirical studies in grassland, savannas and open forest ecosystems generally revealed the predominance of inter-annual climatic variability, ecosystem resilience and complex land-cover change trajectories over secular (irregularly spaced at long intervals in time) land-cover conversions.

Digital change detection encompasses the quantification of temporal phenomena from multi-date imagery that is most commonly acquired by satellite-based multi-spectral sensors. The scientific literature, however, reveals that digital change detection is a difficult task to perform. An interpreter analysing aerial photography will almost always produce more accurate results with a higher degree of precision (Edwards 1990). Nevertheless, visual change detection is difficult to replicate because different interpreters produce different results. Furthermore, visual detection incurs substantial data acquisition costs. Apart from offering consistent and repeatable procedures, digital methods can also more efficiently incorporate features from the non-optical parts of the electromagnetic spectrum.

This paper reviews the state of the art of digital change detection in the optical/infrared domain with emphasis on ecosystems. It is a thorough revision of an eight-year-old previous effort (Coppin and Bauer 1996) that was largely confined to bi-temporal change detection methodologies (comparing the same area at two points in time). The present review has been expanded now to encompass also the emerging field of change detection based on temporal trajectory analysis (comparing the same area over longer time intervals with multiple imagery, e.g. over the length of a growing season).

2. Ecosystem change and multi-temporal imagery
2.1. Change

Ecosystems are continuously changing, where change is defined as ‘an alteration in the surface components of the vegetation cover’ (Milne 1988) or as ‘a spectral/spatial movement of a vegetation entity over time’ (Lund 1983). The rate of change can either be dramatic and/or abrupt, as exemplified by fire; or subtle and/or gradual, such as biomass accumulation. Change can therefore be seen as a categorical variable (class) or in a continuum. Authors generally distinguish between land-cover conversion, i.e. the complete replacement of one cover type by another, and land-cover modification, i.e. more subtle changes that affect the character of the land cover without changing its overall classification. Land-cover modifications are generally more prevalent than land-cover conversions. Some ecosystem modifications are human-induced, for example tree removal for agricultural expansion. Others have natural origins resulting from, for example, flooding and disease epidemics.

Various classifications of change in ecosystems have been proposed. A problem with many of these categorical schemes is that the change classes are often not mutually exclusive. Colwell et al. (1980) therefore suggested a more hierarchical framework, whereby the classification is mutually exclusive and totally exhaustive. However, it still falls short by not providing a link between change event and causal
agent. Hobbs (1990) focused more on ecological aspects and differentiated between seasonal vegetation responses, inter-annual variability and directional change. The last may be caused by intrinsic vegetation processes (e.g. succession), land-use conversion, other human-induced changes (e.g. pollution stress) and alterations in global climate patterns (e.g. global warming). Khorram et al. (1999) concentrated on the spatial environment in which the change occurs: ‘Some changes may affect entire areas uniformly and instantaneously, while others may take the form of slow advances or retreats of boundaries between classes, and still other changes may have very complex spatial textures.’ In the spatial context, they proposed four types of change whereby spatial entities either (1) become a different category, (2) expand, shrink or alter shape, (3) shift position, or (4) fragment or coalesce.

The ability of any system to detect and monitor change in ecosystems depends not only on its capability to deal adequately with the initial static situation, but also on its capacity to account for variability at one scale, e.g. seasonal, while interpreting changes at another, e.g. directional (Hobbs 1990). Moreover, the ability to detect is a function of the ‘from’ and ‘to’ classes, the spatial extent and the context of the change (Khorram et al. 1999).

Not all detectable changes, however, are equally important. It is also probable that some changes of interest will not be acquired very well, or at all, by any given system. Of particular interest to the ecosystem scientist and/or manager are first and foremost vegetation disturbances caused by short-term natural phenomena such as insect infestation, fire and flooding, and changes resulting from human activities, e.g. resource exploitation and land-use conversion. While the natural phenomena are likely to be temporary and may in some cases even be self-correcting, evidence of anthropogenic activities generally remains much longer. Equally important are the ecosystem changes that ensue from alterations in larger-scale processes, such as global warming etc. In this case, changes in trends are analysed, rather than actual change events. The proper understanding of the nature of the change and the principles that enable its detection and categorization usually encompass more sophistication than the simple detection of the change event itself.

In summary, the main challenges facing ecosystem change monitoring from space come from the requirements to: (i) detect modifications in addition to conversions (e.g. quantify forest cover degradation due to selective logging or fires); (ii) monitor rapid and abrupt changes in addition to the progressive and incremental changes (e.g. assess the impact of a flood, drought or fire, versus a progressive expansion of agriculture); (iii) separate inter-annual variability from secular trends, given the shortness of the available time series—20 to 30 years (e.g. assess dry land degradation); (iv) understand and correct for the scale dependence of statistical estimates of change derived from remote sensing data at different spatial resolutions; and (v) match the temporal sampling rates of observations of processes to the intrinsic scales of these processes (e.g. monitor rapidly evolving processes such as floods or biomass burning).

2.2. Imagery acquisition

Change detection for ecosystem monitoring generally assumes overall phenological conditions to be comparable, be it on a two-point timescale (bi-temporal change detection; e.g. change detection between two peak-green summer images), or on a continuous timescale (temporal trajectory analysis; e.g. change detection between growing seasons with similar climatic conditions). However, in the latter
case, it is also possible that these conditions themselves are the objects that are being monitored, e.g. in natural variability monitoring.

2.2.1. Bi-temporal change detection

The appropriate selection of imagery acquisition dates is as crucial to a bi-temporal change detection method as is the choice of the sensor(s), change categories and change detection algorithms. The problem has two dimensions: the calendar acquisition dates and the change interval length (temporal resolution).

Anniversary dates or anniversary windows (annual cycles or multiples thereof) are often used because they minimize discrepancies in reflectance caused by seasonal vegetation fluxes and Sun angle differences. However, even at anniversary dates, or within anniversary windows, phenological disparities due to local precipitation and temperature variations may present real problems. Forest ecosystem dynamics in the temperate regions illustrate this clearly. All else being equal, the reflectance of tree leaves in the visible part of the electromagnetic radiation (EMR) spectrum is higher in spring and autumn than in the middle of the growing season. Changes in the near-infrared (NIR) part of the spectrum are less distinct. The stage of change at a particular time in spring and autumn depends on the site, tree species and varying genotypes of the same species. Local seasonal effects are especially confusing during leaf-out and autumn coloration. Hame (1988) therefore concluded that for bi-temporal change detection, summer and winter are the best seasons because of their phenological stability. What is more, selecting the summer, or the driest period of the year for the locale, will enhance spectral separability, yet minimize spectral similarity, because of excessive surface wetness prevailing during other periods of the year (Burns and Joyce 1981). However, the optimal selection of the season for bi-temporal forest cover change detection data acquisition remains a topic of contention in the pertinent literature.

For most documented studies, the periodicity of the data acquisition seems to have been determined according to the availability of satellite sensor data of acceptable quality. While visually analysing Landsat-1 multi-spectral scanning system (MSS) data for forest cover regeneration assessment, Aldrich (1975) concluded that in most cases a minimum time interval of three years was required to detect non-forest to forest changes. Colwell et al. (1980) judged a two-year separation between MSS scenes insufficient to map reliably re-vegetating areas in South Carolina. They advised a periodicity of at least five years. Gregory et al. (1981) reported that in southern Oklahoma three separate classes of clear cuts have been identified from processed Landsat MSS data: those created less than six years before image acquisition, those that dated from six to 15 years before data capture, and those older than 15 years. Park et al. (1983), again using Landsat MSS data, suggested a one-year interval to detect forest to non-forest (successional herbs, urban or agricultural development) changes, a three- to five-year interval to monitor non-forest to successional shrubs stage, and another five to 10 years to detect the consecutive establishment of a forest cover. Coppin and Bauer (1995) tested two-, four- and six-year intervals for canopy change detection and found that a two-year cycle was optimal to study aspen establishment and storm damage in the Great Lakes region with Landsat Thematic Mapper (TM) imagery. However, their four- and six-year cycles gave better results in the case of human-induced and natural canopy disturbances such as thinning, cutting and dieback.
2.2.2. Temporal trajectory analysis

To circumvent the problem of the selection of optimal imagery acquisition dates, some investigators have approached ecosystem monitoring by comparing seasonal development curves or profiles. These require time series of remotely sensed indicators of relevant land surface attributes which, in turn, are constructed from daily imagery acquisitions provided by such sensors as the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR), Système Probatoire de l’Observation de la Terre (SPOT) VEGETATION, Sea-Viewing Wide Field-of-View Sensor (SeaWIFS), Moderate-Resolution Imaging Spectrometer (MODIS), Along Track Scanning Radiometer (ATSR), etc. Change detection based on such profiles has been proven appropriate for regional studies of largely climate-driven land surface attribute changes (Lambin and Ehrlich 1997, Kawabata et al. 2001), of phenology modifications (Myneni et al. 1997), and of inter-annual net primary production variability (Behrenfeld et al. 2001).

One of the advantages of profile-based techniques lies in the fact that the issue of influence of phenology on change detection performance is resolved, because data are collected throughout the growing season. As such, changes inherently linked to seasonality can be separated from other changes. A serious disadvantage, however, is that at present only coarse (AVHRR, VEGETATION) to moderate (MODIS) spatial-resolution sensors provide the high temporal frequency of observations that is necessary to establish time profiles. This drastically limits the change categories that can be detected and monitored. Lawrence and Ripple (1999) have nevertheless been able to monitor the ecosystem disturbances caused by a volcanic eruption by calculating change curves from a time series of Landsat TM data acquired at high temporal frequency.

3. Data pre-processing for change detection

The primary challenge in deriving accurate natural ecosystem change information is representative of the standard remote sensing problem: maximization of the signal-to-noise ratio. Inherent noise will affect the change detection capabilities of a system or even create unreal change phenomena. Causes of such unreal changes can be, among others, differences in atmospheric absorption and scattering due to variations in water vapour and aerosol concentrations of the atmosphere at disparate moments in time, temporal variations in the solar zenith and/or azimuth angles, and sensor calibration inconsistencies for separate images. Pre-processing of satellite sensor images prior to actual change detection is essential and has as its unique goals the establishment of a more direct linkage between the data and biophysical phenomena, the removal of data acquisition errors and image noise, and the masking of contaminated (e.g. clouds) and/or irrelevant (e.g. waterbodies when looking at changes in vegetation) scene fragments.

The pre-processing of multi-date sensor imagery, when absolute comparisons between different dates or periods are to be carried out, is much more demanding than in the single-date case. It commonly comprises a series of sequential operations, including (but not necessarily in this order) calibration to radiance or at-satellite reflectance, atmospheric correction or normalization, image registration, geometric correction, mosaicking, sub-setting and masking (e.g. for clouds, water, irrelevant features). Often these procedures are accompanied by a data transformation to vegetation indices that are known to exhibit a strong positive
The relationship between upwelling radiance and the vegetative cover of natural ecosystems. The principal advantages of vegetation indices over single-band radiometric responses are their ability to reduce considerably the data volume for processing and analysis, and their inherent capability to provide information not available in any single band. However, no single vegetation index can be expected to summarize totally the information in multidimensional spectral data space. Wallace and Campbell (1989) aptly stated that adequate indices can be found for different purposes and that indices derived for one analysis may be inappropriate in another context.

3.1. Bi-temporal change detection

Of the various aspects of pre-processing for bi-temporal change detection, there are two outstanding requirements: multi-date image registration and radiometric rectification. It should be evident that accurate spatial registration of the multi-date imagery is absolutely essential to digital change detection. In a study on the impact of mis-registration on change detection simulations of MODIS data run with spatially degraded Landsat MSS images, Townshend et al. (1992) followed by Dai and Khorram (1998) clearly demonstrated that to achieve errors of only 10% in vegetation index values, registration accuracies of 0.2 pixels or less were required. However, change detection capabilities are intrinsically limited by the spatial resolution of the digital imagery. Further, residual mis-registration at the below-pixel level commonly degrades the area assessment of the change events somewhat, specifically at the change/no-change boundaries. This within-pixel shift is inherent to any digital change detection technique (Coppin and Bauer 1994).

The second critical requirement for successful change detection is that a common radiometric response is required for quantitative analysis of one or more image pairs acquired on different dates. For many change detection applications, absolute radiometric correction is unnecessary, but variations in solar illumination conditions, in atmospheric scattering and absorption, and in detector performance need to be normalized or, in other words, radiometric properties of a subject image need to be adjusted to those of a reference image. Duggin and Robinove (1990) strongly insisted on the calibration of raw sensor data to meaningful physical units prior to any multi-temporal analysis. Only through reliable radiometric calibration can a researcher be confident that observed spatial or temporal changes are real differences and not artefacts introduced by differences in the sensor calibration, atmosphere and/or Sun angle (Robinove 1982).

Collins and Woodcock (1996) examined three levels of radiometric pre-processing: no pre-processing, an intermediate level (matched or normalized digital counts) and full radiometric correction (matched satellite/ground reflectances). They considered some form of radiometric correction essential, but the matched digital counts, in which several invariant features were used to calibrate a regression equation to predict time-2 digital numbers from those at time-1 (relative routine), was found to be as accurate as and easier to apply than the matched reflectances (absolute routine). Songh et al. (2001) demonstrated that one relative and seven absolute atmospheric correction algorithms all improved final data analysis when tested for change detection purposes. The more complicated algorithms did not necessarily lead to greater accuracy; however, with respect to atmospheric correction, the authors recommended either a simple dark object subtraction with or
without a Rayleigh atmospheric correction (absolute routine), or atmospheric normalization (relative routine).

Atmospheric normalization can be full-scene-based and is then called simple regression normalization (SR) (Jensen 1983) or pseudo-invariant-feature-based (PIF). Hall et al. (1991a) have developed such a PIF radiometric normalization technique that corrects or rectifies images from common scenes through use of sets of landscape elements whose reflectance is nearly constant over time. The technique provides a relative calibration and does not require sensor calibration or atmospheric turbidity data, although correction to absolute surface reflectance can be accomplished if sensor calibration coefficients, an atmospheric correction algorithm (model) and atmospheric turbidity data are available. The technique resulted in data accurate to within 1% absolute reflectance in the visible and NIR bands. Similar procedures have been used by Caselles and Lopez Garcia (1989), Conel (1990) and Coppin et al. (2001). Elvidge et al. (1995) developed an automatic scattergram-controlled regression (ASCR) normalization technique. It uses pixels from no-change regions identified from scattergrams, instead of from whole images as in the SR method or from spectral/spatial features as in the PIF procedure. More sophisticated approaches, which use atmospheric models for correction, but derive the atmospheric parameters from the satellite imagery, include Ahern et al. (1988), Chavez (1989) and Hill and Sturm (1991). All of these studies give clear evidence of the fact that only when all sources of variation but the surface cover can be adjusted for (absolute calibration) or normalized to a common standard (relative calibration), will it be possible to detect and identify changes in natural ecosystems from multi-date imagery.

The debate around the use of filtering as a pre-processing technique for change detection has not been settled and remains very application-dependent. Riordan (1981) applied a modification of Nagao’s and Matsuyama’s edge-preserving image-smoothing algorithm to reduce minor variations in radiance values and to allow the comparison of relatively homogeneous groups of pixels. He judged the result ineffective. For tropical deforestation assessment, Singh (1989) tried both image smoothing and edge enhancements, but did not detect any increase in the final change detection accuracy with either method. Baraldi and Parmiggiani (1989), however, suggested the application of edge-preserving image-smoothing filters previous to image analysis in order to enhance the homogeneity of the spectral response of a thematic class and at the same time to eliminate noise effects. Bruzzone and Prieto (2000) reported increased change detection accuracies when exploiting the spatial–contextual information contained in the neighbourhood of each pixel to reduce the effects of noise. In addition, the definition of an adaptive pixel neighbourhood or parcel allowed for a more precise location of the borders of changed areas. To overcome low frequency/high contrast problems in Landsat TM imagery covering sand sheet desert environments, Kwarteng and Chavez (1998) advocated the use of high pass spatial filters with relatively large kernel sizes, with a 50% add-back option and followed by edge enhancement, before submitting the data to actual change detection.

Some researchers advocate the use of texture features for change detection. Most consulted literature sources explicitly state, however, that texture information must be used only in conjunction with spectral data, and that the two sources are complementary to each other (e.g. He and Wang 1990). Reed (1988) cautioned the user community about the integration of texture measures in data analysis procedures. Classification accuracy is increased only under certain conditions.
In his research, the addition of textural features, derived from a transformed vegetation index map of TM data, degraded the classification accuracy in distinguishing spectrally similar vegetation covers in all but one case. He conceded, however, that the choice of texture index (angular second moment and contrast) and window size \((7 \times 7 \text{ pixels})\) might have influenced the outcome significantly. Smits and Annoni (2000) computed texture distance measures based on the texture indices' contrast and homogeneity. They report mis-detection and false alarm rates of only 15% in urban expansion change detection and demonstrated their method robust for mis-registration.

3.2. Temporal trajectory analysis

As time-profile-based change detection methods work with data acquired on a large number of observation dates, adequate pre-processing is of utmost importance. Just as with bi-temporal approaches, pre-processing includes the geometric registration of successive images at the sub-pixel level. For example, Roy (2000) showed that high contrast boundaries on images from wide field-of-view sensors might be shifted when mis-registered data are composited. This may exaggerate reverse or obscure change phenomena. As one is dealing with a series of images, data artefacts that are inherent to the type of imagery used and that render the comparison of data observations or measurements at different times difficult, must be removed. Moreover, to remove cloud and other atmospheric effects (i.e. water vapour content, aerosols), a process of temporal compositing must be executed. Finally, many present-day high-temporal-frequency sensors have a wide field of view, which necessitates a correction for directionality (bidirectional reflectance distribution function (BRDF)) effects. For example, changes in illumination conditions due to a drift in equator crossing time of the AVHRR sensors result in a viewing angle difference for off-nadir observations (Gutman 1999). And alterations in AVHRR sensor responsivity over time constitute another major source of image noise (Cihlar et al. 1998, Gutman 1999).

The fact that some controversy arose on whether previous time series analyses detected real trends in the Earth’s climate system (e.g. Myneni et al. 1997), or resulted at least in part from a combination of calibration residuals and satellite-orbit drift (Gutman 1999), illustrates how critical the artefact issue still is. In addition, natural phenomena can render temporal trajectory analysis much more difficult: the eruptions of Mount Pinatubo in mid 1991 caused significant noise in the global AVHRR time series due to the varying presence of aerosols in the atmosphere.

With respect to temporal compositing, a range of procedures has been suggested for wide-angle sensors. Several authors have proposed other criteria for temporal compositing than the widely used maximum vegetation index value. Actually, the latter criterion preferentially selects off-nadir pixels from the fore scatter region, this effect varying with land-cover type (Cihlar et al. 1994). Other single-step compositing criteria (e.g. maximum temperature) (Roy 1997) or two-step procedures (e.g. maximum vegetation index value followed by minimum scan angle) (Cihlar et al. 1994) were proven to yield good composite images. These studies also highlight the impact of the compositing procedure on results derived from time series analysis for specific applications. Therefore, some authors have applied compositing criteria designed for a specific aim, such as monitoring burned areas at the regional scale (Barbosa et al. 1998).
Other researchers have attempted to resolve problems that contaminate time series of wide field of view sensors by addressing each issue separately rather than attempting to solve all the issues through a single compositing procedure. Concerning residual clouds, various threshold-based methods have been proposed for cloud screening, e.g. an automated series of five tests applied to each pixel (Saunders and Kriebel 1988) or thresholding of the AVHRR channel 4 (centred around 11 μm) brightness temperature (Gutman et al. 1994). Threshold-based methods suffer, however, from the heterogeneity of the land surface and the seasonal variability of surface radiance. As an alternative, a Fourier series approximation to the seasonal trajectory of vegetation index data has been developed and then modified to closely represent seasonal trends for different land-cover types (Cihlar et al. 2001). It is used to determine time- and pixel-specific cloud contamination thresholds and then to produce contamination masks. Concerning other atmospheric effects, atmospheric corrections of AVHRR data using seasonal averages of atmospheric water vapour and aerosols optical depths were shown to result in corrections that were similar to the full correction using daily values based on in situ data, for biophysical studies in the African Sahel using time series of vegetation index data (Hanan et al. 1995).

The correction for BRDF effects in wide field of view satellite sensors is probably the area where progress has been more explicit recently. Hu et al. (2000) have produced surface albedo data corrected for angular effects from a time series of daily multi-angular AVHRR data. They used a kernel-driven semi-empirical land surface BRDF model where kernels are derived from approximations to physical BRDF models. The angle-corrected vegetation index data displayed more consistent surface properties than monthly maximum value composites and were used to investigate seasonal and inter-annual changes in albedo and vegetation index value. Schaaf et al. (2002) applied this approach to global MODIS data to generate nadir BRDF-adjusted reflectances. This product, computed for 16-day periods, is free from view angle effects, as well as from cloud and aerosol contamination. It is thus ideal for applications that traditionally depended on compositing methods, and in particular for monitoring land-cover change and inter-annual variability in land surface conditions at broad spatial scales.

In another work also aimed at the data from the new generation of satellite sensors, van Leeuwen et al. (1999) proposed a compositing algorithm for MODIS vegetation index data based on a view angle standardization approach. Nadir equivalent reflectance values were produced using the simple Walthall BRDF model. In a different approach, Csiszar et al. (2001) corrected daily AVHRR data for angular effects using coincident multiangle polarization and directionality of the Earth’s reflectance (POLDER) land surface data products. The POLDER data were used to derive BRDF functions for 6 km² AVHRR grid boxes, and the anisotropic factors were then applied to the individual AVHRR grid cell reflectances.

4. Change detection algorithms
4.1. Bi-temporal algorithms
4.1.1. Background

All digital change detection is affected by spatial, spectral, temporal and thematic constraints. The type of method implemented can profoundly affect the qualitative and quantitative estimates of the disturbance (Colwell and Weber 1981). Even in the same environment, different approaches may yield different
change maps. The selection of the appropriate method therefore takes on considerable significance. Most documented digital change detection methods are based on per-pixel classifiers and pixel-based change information contained in the spectral–radiometric domain of the images. They combine both the procedures for change extraction (change detection algorithm) and those for change separation/labelling (change classification routine). Change extraction and change separation/labelling are preliminaries to any change modelling. In other words, predicting with a statistical or ecological model using independent variables where, when, and/or why change occurs, requires prior detection, measurement and categorization of ecosystem change patterns.

Statistical and/or spatial decision rules that are derived from a heuristic understanding of the change event often constitute the backbone of the separation/labelling exercise. These rules, however, are not change-detection-algorithm-specific. In other words, they can be applied irrespective of the algorithm that generated the change data. They comprise the complete range of pattern recognition procedures available to the image analyst, from edge enhancement and simple thresholding (visual or statistically-based), to supervised and unsupervised classification, to segmentation and spatial analysis rules.

A wide variety of digital change detection algorithms have been developed over the last two decades. They basically can be summarized in two broad categories to which different reviewers have attached definitions that vary in complexity and, to a certain extent, in coverage. Malila (1980) recognized the categories as change measurement (stratification) methods versus classification approaches. Pilon et al. (1987) amplified the description of the first category to ‘enhancement approaches involving mathematical combinations of multi-date imagery which, when displayed as a composite image, show changes in unique colours’. Singh (1989) changed the focus slightly by centering the definitions more on a temporal scale: simultaneous analysis of multi-temporal data versus comparative analysis of independently produced classifications for different dates. Other scientists (Nelson 1983, Milne 1988) have employed multi-class schemes.

Whatever combination of change detection algorithm and classification routine is applied, it is obvious that a wide assortment of alternatives exist and that all have varying degrees of flexibility and availability. As already stated, change classification routines are not specific to change detection. The overview that follows is therefore restricted to change detection algorithms only. All change detection algorithms that have been found documented in the literature can be grouped into nine distinctly different categories plus one heterogeneous group encompassing hybrid methods and less frequently implemented or more esoteric algorithms. The first six are those most frequently used for monitoring vegetative canopies, while the other algorithms are less common and/or remain in an experimental stage. However, the order in which the algorithms are presented does not imply any ranking or qualitative judgement. Moreover, because the algorithms are not necessarily independent of the data sources for which their implementation has been documented, examples typical for natural ecosystem monitoring are given.

4.1.2. Post-classification comparison

Post-classification comparison is sometimes referred to as ‘delta classification’. It involves independently produced spectral classification results from each end of
the time interval of interest, followed by a pixel-by-pixel or segment-by-segment comparison to detect changes in cover type. By adequately coding the classification results, a complete matrix of change is obtained, and change classes can be defined by the analyst.

The principal advantage of delta classification lies in the fact that the two dates of imagery are separately classified, thereby minimizing the problem of radiometric calibration between dates. By choosing the appropriate classification scheme, the method can also be made insensitive to a variety of types of transient changes in selected terrain features that are of no interest (Colwell et al. 1980). However, the accuracy of the post-classification comparison is totally dependent on the accuracy of the initial classifications. The final accuracy very closely resembles that resulting from the multiplication of the accuracies of each individual classification and may be considered intrinsically low. The difficulty thus lies in securing completely consistent, analogous and highly accurate target identifications for each iteration. Mis-classification and mis-registration errors that may be present in the original images are compounded and results obtained using post-classification comparison are therefore frequently judged unsatisfactory (Howarth and Wickware 1981).

A significant example of the use of the post-classification comparison approach is the work of Hall et al. (1991b) in which Landsat images acquired in 1973 and 1983 were classified into five forest successional classes (clearings, regeneration, broadleaf, conifer and mixed). Application of a PIF normalization (see §3.1) enabled classification of the 1973 image using 1983 ground data for classifier training. Following the two classifications a matrix of class changes over the 10-year interval was constructed and the transition rates between classes were calculated.

Xu and Young (1990) preceded their post-classification comparison by a manual segmentation of the images according to ground features and characteristics of the scene. They then classified all segments separately for each date via a supervised maximum likelihood pattern recognition routine. They concluded that this approach, sometimes referred to as ‘pre-stratified delta classification’, enabled them to avoid some obvious errors in classification (e.g. pixels classified as built-up areas on areas known to be moorlands in south-east Scotland).

4.1.3. Composite analysis

By using combined registered datasets, or corresponding subsets of bands, collected under similar conditions with respect to ecosystem phenology but from different years, classes where vegetative canopy change is occurring would be expected to have statistics significantly different from those where no change has occurred, and could be identified as such. The method can incorporate multistage decision logic and is sometimes referred to as ‘spectral/temporal change classification’, ‘multi-date clustering’ or ‘spectral change pattern analysis’. While this technique necessitates only a single classification, it is a very complex one, in part because of the added dimensionality of two dates of data. In numerous cases it requires many classes and many, often redundant features when no discriminant analysis has preceded the process. It furthermore demands prior knowledge of the logical interrelationships of the classes and should be used only when the researcher is intimately familiar with the study area (Jensen 1983). Burns and Joyce (1981) found the method to produce only change in forest cover per se without providing accurate information on the character of the change. Schowengerdt (1983)
remarked that, since spectral and temporal features have equal status in the combined dataset, they could not be easily separated in the pattern recognition process. As a consequence, class labelling may be difficult. Researchers who have used this technique for natural ecosystem change detection include Colwell et al. (1980), Hall et al. (1984) and Sader (1988).

4.1.4. Univariate image differencing

From the analysis of the relevant scientific literature, univariate image differencing is the most widely applied change detection algorithm. It involves subtracting one date of original or transformed (e.g. vegetation indices, albedo, etc.) imagery from a second date that has been precisely registered to the first. With ‘perfect’ data, this would result in a dataset in which positive and negative values represent areas of change and zero values represent no change.

Banner and Lynham (1981) used a multi-temporal vegetation index difference based on calculated Normalized Difference Vegetation Indices (NDVIs) for MSS datasets. They then density-sliced the difference vegetation index image. They found the method impractical for forest cutover delineation owing to the sensitivity of the NDVI to grass growth and the development of other vegetation in the clear-cuts, but useful for monitoring vegetative competition within the cutovers. Analogously, Lyon et al. (1998) implemented NDVI differencing and found it the better vegetation change detection technique for monitoring deforestation and loss of vegetation. Nelson (1983) delineated forest canopy changes due to gypsy moth defoliation in Pennsylvania more accurately with vegetation index differencing than with any other single band difference or band ratiing (see §4.1.5). Mukai et al. (1987) computed normalized difference channels for MSS5 and MSS6 to detect areas infested by pine bark beetle in Japan. They were able to distinguish three classes of infestation from light, over moderate, to heavy, with increasing band-5 and decreasing band-7 pixel values. Thresholds for each class were set separately and interactively using multiples of a ‘residual’ standard deviation.

Hame (1986) suggested histogram matching and Yasuoka (1988) band-to-band normalization before differencing TM data so as to yield bands with comparable means and standard deviations and to reduce scene-dependent effects. In comparative analyses of the six reflective TM difference channels, Hame (1986) as well as Fung (1990) found that the TM3 difference channel contained the highest information content for vegetative cover monitoring.

Coppin and Bauer (1994) suggested a standardization of the differencing algorithm (difference divided by the sum) to minimize the occurrence of identical change values depicting different change events.

Serneels et al. (2001) applied double univariate image differencing in a spatial–contextual approach to separate anthropogenic changes from climate variability in a savanna environment. The technique computed and combined changes at pixel and landscape scales. First, univariate image differencing was applied to pairs of smoothed (101 × 101 pixels low-pass filter) vegetation index images to delineate landscape-scale changes. Second, local-scale patterns were detected via pixel-level image differencing between the two original (unsmoothed) full-resolution images. Finally, the latter change image was subtracted from the former, resulting in a change image wherein all pixels that behaved differently over time at both scales had the highest change values.
Cohen et al. (1998) expanded bi-temporal image differencing in a multi-temporal context. They comparatively assessed two approaches: merged versus simultaneous image differencing. The first was based on an unsupervised classification, repeated five times, using five sequential date-pairs of different Landsat images between 1972 and 1993, and requiring the merging of the results from five separate time intervals into a single change map. The second involved a single unsupervised classification of the full sequential difference image set. Both methods gave consistent and comparable results, with simultaneous image differencing being considerably more cost-effective to implement.

4.1.5. Image ratioing

Though not as quick and simple as image differencing, image ratioing is also one of the conceptually easier to understand change detection methods. Data are ratioed on a pixel-by-pixel basis. A pixel that has not changed will yield a ratio value of one. Areas of change will have values either higher or lower than one.

To represent the range of the ratio function in a linear fashion and to encode the ratio values in a standard eight or 16 bit code common to many PC-based image analysis software packages, a two-level (for values smaller than or greater than zero) normalizing function can be applied to all ratio values not equal to zero (Jensen 1983). Because of the non-Gaussian bimodal distribution of ratioed multiple-date images, Riordan (1981) criticized the ratio change detection algorithm in combination with an empirical threshold definition as being statistically invalid. The areas delineated on either side of the distribution mode are not equal. Consequently the standard deviation cannot be used for threshold definition.

Howarth and Wickware (1981) combined MSS5 and MSS7 ratios in a single colour composite. They found that, while the band-5 ratio emphasized changes in water level due to flooding, the band-7 ratio assigned the brightest pixel values to areas where changes in vegetation cover were dominant. They were not, however, able to make a quantitative assessment of the changes.

4.1.6. Bi-temporal linear data transformation

Various linear data transformation techniques can be applied to two-date imagery that has been stacked in \(2n\)-dimensional space (where \(n\) is the number of input bands per image). They concentrate information pertaining to statistically minor modifications in the state of the natural ecosystem (minor as contrasted to the entirety of the image scene) in orthogonal components, producing uncorrelated differences. The most important linear transformation is the one based on principal component analysis (PCA), with tasseled cap (Crist and Cicone 1984) and multivariate alteration detection (MAD) (Nielsen et al. 1998) being less frequently implemented. While linear transformation techniques often showed remarkable results, more complicated nonlinear transformations like the curve-theorem based approach of Yue et al. (2002) could not yet show their utility.

The exact nature of the principal components derived from multi-temporal datasets is difficult to ascertain without a thorough examination of the eigenstructure of the data and a visual inspection of the combined images (eventually via multidimensional temporal feature space analysis, see §4.1.10). To avoid drawing
faulty conclusions, the analysis should not be applied as a change detection method without a thorough understanding of the study area (Fung and LeDrew 1987). The link between vegetative canopy change and tasselled cap transforms, on the other hand, appears to be more solid, an observation supported by Collins and Woodcock (1994).

Richards (1984) applied a normal PCA procedure to two-date MSS imagery to monitor brush-fire damage and vegetation regrowth over extensive areas in Australia. Provided the major portion of the variance in the multi-temporal sequence was associated with correlated (constant, unchanged) land cover, areas of localized change were enhanced in some of the lower components, particularly principal components three and four. Ingebritsen and Lyon (1985) did exactly the same thing to detect and monitor vegetation changes around a large open-pit uranium mine in Washington and a wetland area in Nevada. Under the assumptions that the two original images both had an intrinsic dimensionality of two (first two principal components, rest primarily noise), that these dimensions were related to soil brightness and vegetation greenness, and that the change in land cover and/or vegetation condition exceeded some threshold value, four meaningful principal components resulted. They were stable brightness, stable greenness, change in brightness (somewhat analogous to an albedo difference signature) and change in greenness (somewhat similar to a NIR/red ratio difference image). The latter component proved to be well related to change in vegetative cover and insensitive to variations in slope and aspect.

According to Fung and LeDrew (1987), standardized PCA (as opposed to the normal PCA procedure) gave more accurate results because its PCs were found to be better aligned along the object of interest: change. Singh and Harrison (1985) concurred, citing a substantive improvement in signal-to-noise ratio and image enhancement by using standardized variables in the PCA. On the subject of subsetting the image to derive multi-temporal PCs, the authors stated that, although even lower principal components can detect some land cover changes, the statistics extracted from data subsets are not recommended for change detection due to the great variability and uncertainty of the unextracted part of the data. They furthermore found the multi-temporal PCA too scene-dependent and suffering from the serious drawback that no prior information was available regarding the nature of the components before actual processing.

Fung (1990) reported on a comparative analysis of a multi-date standardized PCA and a multi-date tasselled cap transformation of a multi-temporal Landsat TM dataset. The PCA rotation was based on the merged 12-band TM image. Three components were found associated with change: PC3 with changes in soil brightness, PC5 with changes in NIR reflectance and thus vegetative vigour, and PC6 with changes in the contrast between the middle infrared bands and the photosynthetically active radiation (PAR) bands, thus changes in wetness. The usefulness of these lower-order multi-date PCs to highlight localized change was confirmed by Lee et al. (1989). They ran a principal factor analysis on the transformed image and found the lower-order PCs to contain significantly more unique variance. The tasselled cap transformation was carried out as follows. The spectral bands of the two dates were assigned the tasselled cap coefficients as derived by Crist and Cicone (1984) with positive coefficients for the first date and negative coefficients for the second. The derived vectors were not orthogonal and consequently were subjected to a Gramm–Schmidt transformation. The process generated output vectors that were effectively orthogonal to each other. Three
change tasselled cap images were produced detailing differences in greenness, brightness and wetness. The greenness change image gave the highest classification accuracy between all resulting PCA and tasselled cap change-related images. Fung (1990) clearly advocated the use of the tasselled cap transformation. Under this algorithm, the inherent data structure could be clearly depicted and the derived variables were physically based and independent of scene content. Twelve TM input bands could moreover be reduced to two, maximum three, significant change bands. While Hame (1988) concluded that the original TM bands accomplished more than the transformed features (bi-temporal PCs or differences between paired PCs based on either the covariance matrix or the correlation matrix) to separate change classes in Finland, Coppin and Bauer (1994) found the second principal component of vegetation index band pairs to be an excellent indicator of change in the temperate forest cover in the north-central USA. Kwarteng and Chavez (1998) used a similar selective PCA approach to successfully detect and map surface changes dealing with urban development, vegetation growth, and coastal wetland and sand sheet surface differences. Collins and Woodcock (1994) have developed a multi-temporal generalization of the tasselled cap transformation. Application of the technique produced multi-temporal analogues of the brightness, greenness and wetness—the three primary dimensions of the tasselled cap transformation—and a component measuring change. Nielsen et al. (1998) proposed the multivariate alteration detection or MAD, which is an extension of the traditional canonical correlations analysis and is invariant to linear scaling of the input data. As such MAD is insensitive to, for example, differences in sensor gain settings, or to linear radiometric and atmospheric correction schemes. MAD, in combination with a posterior maximum autocorrelation factor (MAF) transformation, gave significantly better results than PCA of simple differenced data in the detection of coherent patterns of spatial change in urbanization zones in Australia. Moreover, the MAD transformation provides a way of combining different data types that may be useful in historical change detection studies.

4.1.7. Change vector analysis

Change vector analysis (CVA) is a multivariate change detection technique that processes the full dimensionality (spectral + temporal) of the image data and produces two outputs: change magnitude and change direction. A major advantage is its capability to analyse change concurrently in all data layers as opposed to selected bands.

The first automated CVA change detection algorithm that furthermore took account of spatial scene characteristics was developed at the Environmental Research Institute of Michigan in the late 1970s. It consolidated a tasselled cap transformation to greenness–brightness, an image-segmentation to spatially contiguous pixel groups or ‘blobs’, and a characterization of the movement of the individual segments in spectral space in terms of magnitude and direction (Malila 1980).

Colwell et al. (1980) applied the algorithm in Kershaw County, South Carolina, and found that, to be effective, it required absolutely precise image registration and normalization (changes in brightness needed to be scaled to be approximately equal to changes in greenness to avoid elliptical change thresholds) and a considerable
amount of operator interaction. A forest mask had to be created, and parameters had to be set to control the formation of blobs and to threshold the change vectors for change, with respect to their magnitude (defining change from no change) as well as their direction (labelling the type of change). While the relative utility of the technique to assess the type of change was not clear, CVA performed well with respect to automated change detection.

Prior to computing change as a vector or distance in multi-spectral space, Yokota and Matsumoto (1988) applied a preliminary transformation to all raw digital numbers (new value = old value minus the average brightness over the six TM bands) to ‘accentuate’ the disturbances. They then calculated a multi-band Euclidean distance measure as the spectral differentiation norm. This measure was computed as the square root of the sum of the squares of the pixel value differences between the two dates over the six bands.

Lambin and Strahler (1994) applied similar principles of segmentation and vector movement characterization to multi-temporal vegetation indices, surface temperatures and spatial texture data. Combined with PCA of the change vectors, the technique proved to be effective in detecting and categorizing different land-cover changes operating at different timescales in West Africa. Johnson and Kasischke (1998) amply illustrated the capability of CVA in general to be an effective technique to capture all changes and not just a priori defined change events.

4.1.8. Image regression

A mathematical model that describes the fit between two multi-date images of the same area can be developed through stepwise regression. The algorithm assumes that a pixel at time-2 is linearly related to the same pixel at time-1 in all bands of the EMR spectrum acquired by the sensor. This implies that the spectral properties of a large majority of the pixels have not changed significantly during the time interval (Vogelmann 1988). The dimension of the residuals is an indicator of where change occurred. The regression technique accounts for differences in mean and variance between pixel values for different dates. Simultaneously, the adverse effects from divergences in atmospheric conditions and/or Sun angles are reduced.

The critical part of the method is the definition of threshold values or limiting dimensions for the no-change pixel residuals. When Burns and Joyce (1981) applied the technique to each pair of spectral MSS bands for land-cover change detection via a third-degree polynomial linear equation, they found the green band (MSS4) to perform better than the other band pairs; however, still with relatively low accuracy. Singh (1986), on the other hand, reported the highest change detection accuracy for tropical forest change detection with the regression method and the MSS5 band. A couple of years later (Singh 1989), he reviewed that statement and concluded that the regression method performed only marginally better than univariate image differencing techniques in detecting tropical forest cover changes. This conclusion was confirmed in a later study in urban environment (Ridd and Liu 1998).

4.1.9. Multi-temporal spectral mixture analysis

The increased data dimensionality associated with high-spectral resolution (HSR) data gave rise to spectral mixture analysis (SMA), based on the premise that
HSR image elements are composed of multiple pure spectral signatures or end-members. If a linear mixing model is assumed, the overall reflectance of the image element may be computed from the reflectances of the composing end-members, weighted by their respective surface proportions. As a consequence, image element changes over time are directly mirrored in modifications of the end-member proportions, especially where more subtle natural ecosystem changes are concerned. Because end-members represent the spectra of known natural ecosystem components (e.g. canopy, soil, shadow, etc.), they have the advantage of providing physically based standardized measures of fractional abundance.

Multi-temporal SMA (MSMA) algorithms were implemented with good results by Adams et al. (1995) and Roberts et al. (1998) to monitor the physical nature of land cover changes in the Brazilian Amazon with Landsat TM imagery. The first group used reference end-members (derived from field or laboratory spectra of known components), while the second implemented an iterative process whereby reference library spectra were manipulated using image end-members in order to derive candidate reference end-members for green vegetation, soil, shade and non-photosynthetic vegetation. The major advantage of MSMA when compared with other bi-temporal change detection algorithms lies in its capability to recover natural ecosystem change signals at much finer event scales, e.g. thinning in forest ecosystems.

4.1.10. Multidimensional temporal feature space analysis

This change detection algorithm uses image overlay as a digital enhancement technique for on-screen change delineation. It comprises the one-step combination of a maximum of three individual bands (more steps are possible) which, when displayed as a composite image via the blue, green and red colour guns of a cathode ray tube (CRT), portray changes in unique colours. The multidimensional temporal feature space analysis method seems to be most appropriate in natural environments where changes are relatively subtle. However, it provides the analyst with little information regarding the nature of the change (Pilon et al. 1988). Its use is mostly restricted to the creation of binary change masks to eliminate no-change areas before further analysis with any of the other algorithms, or to the visual definition of training areas related to particular change phenomena. Banner and Lynham (1981) obtained a better forest clear-cut identification and boundary delineation displaying MSS5 of time-1 via the blue CRT colour gun and MSS5 of time-2 via the red colour gun, than with multi-temporal vegetation index differencing (see §4.1.4). Areas of change were highlighted in one of the two gun colours, while areas of no change appeared grey after adjustment of the colour balance for the composite image. Hall et al. (1984) applied the same technique with MSS7 data, however assigning the red colour gun to the MSS7 of time-1 and the green one to the MSS7 of time-2. They were able to connect the variations in red hues to qualitative variations in aspen tree defoliation in Alberta, Canada. Areas without defoliation appeared yellow on the composite image. Renicz (1985), however, remarked that the fast vegetative regeneration in cutover areas precluded the use of any of the MSS reflective infrared bands for clear-cut monitoring with this approach. Because of its relative insensitivity to this phenomenon, the use of red-band overlays was advised.

Werle et al. (1986) created composite multi-temporal images visually to monitor clear-cut and regeneration areas on Vancouver Island, Canada, assigning blue to
TM3 for time-1, green to TM3 for time-2, and red to TM4 for time-2 on the CRT. In a multi-seasonal assessment of regeneration cover in burned forest in Alberta, Canada, Knepp and Ahern (1989) applied another colour combination: blue to TM4 for summer, green to TM4 for autumn, and red to TM3 for summer. Alwashe and Bokhari (1993) merged TM bands 2, 4 and 5 of two different acquisition dates via an intensity–hue–saturation (IHS) transform to represent directly the bi-temporal variations within one single image product. Vegetation differences showed up in distinctly different colours.

Wilson and Sader (2002) recently developed an image overlay technique based on band ratios like NDVI and normalized difference moisture index (NDMI) rather than spectral bands and applied it to temperate forests in Maine. They intensively discussed the accuracy using both indices and found NDMI superior to the use of NDVI. High accuracies in classification were reached especially on the shorter time intervals (two or three years) and even in application of the same method to tropical rainforests.

4.1.11. Hybrid and less frequently implemented algorithms

Only a few authors have documented the combined use of different change detection algorithms (hybrid schemes) in an attempt to minimize commission errors. Pilon et al. (1988) applied such a hybrid scheme to change detection in semi-arid north-western Nigeria, performing post-classification comparison in areas where change was detected by other algorithms. Zhan et al. (2000) applied five change detection methods in parallel (red–NIR space partitioning, red–NIR space change vector, modified delta space thresholding, texture and linear feature) and then integrated the measures of change through a voting method; a change was confirmed where three out of the five methods did flag a land-cover conversion.

At about the same time CVA was developed, change detection procedures called ‘inner product analysis’ and ‘correlation analysis’ (Yasuoka 1988) were implemented in single instances. For both, the difference between the multi-spectral vectors of a pixel at two points in time was expressed as the cosine of the angle between them, but the correlation analysis also took into account the means of the multi-spectral vectors. While the inner product method was found to be sensitive to mis-registration, mixed pixels, sensor gain and offset fluctuations, and changes in absolute radiance, the correlation method substantially reduced these effects.

In 1985, Rencz proposed a multi-temporal biomass index for forest clear-cut monitoring, using the ratio of an MSS7 and MSS5 difference at time-2 over an MSS7 and MSS5 sum at time-1. The results had a very limited usefulness. The error was concentrated in the omission of forest cutovers in which there remained a relatively large number of unfelled residual hardwoods.

In theory, no-change areas can be treated as having slowly varying background grey levels. These variations can be approximated by a background image, for example, a low-pass filtered variant of the original. A subtraction of such an approximation from the image potentially could be used to create a new dataset that accentuates change phenomena. Singh (1986, 1989) used the technique in tropical deforestation monitoring with only modest results.

Ridd and Liu (1998) explored a chi square transformation encompassing the six reflective Landsat TM bands to create a single change image. The method is
applicable only if the scenes are relatively unchanged. A disadvantage is that change related to specific spectral directions might not be readily identified.

4.1.12. Comparison and evaluation of methods

The literature indicates that ecosystem changes can be monitored via a variety of detection methods, with most providing positive results. With the exception of Singh’s 1989 paper, it has taken until the mid 1990s, however, to see comparisons and evaluations of alternative approaches documented.

Sunar (1998) studied the equivalence of multidimensional temporal feature space analysis, univariate image differencing, PCA and composite analysis, in order to detect and delineate land-cover changes in ecosystems under intense development pressure. He could define no optimal method, as each algorithm had its own merits with respect to production ease, information content and interpretability. Similarly, Coppin and Bauer (1994), Ridd and Liu (1998) and Cohen and Fiorella (1998) could make no conclusive statements regarding the superiority or inferiority of univariate image differencing versus in the first case selective PCA, in the second case image regression, and in the third case composite analysis. Hayes and Sader (2001) found a composite analysis of NDVI bands as resulting in more accurate change detection results in tropical forest environments when compared to univariate differencing and PCA approaches.

When comparative conclusions must be made, often univariate image differencing is cited as the, or one of the, preferred change detection algorithms. Singh (1986) found univariate image differencing more effective than bi-temporal PCA for tropical deforestation monitoring. Muchoney and Haack (1994) tested four methods on multi-temporal SPOT sensor data: merged PCA, image differencing, spectral–temporal change classification (composite analysis) and post-classification comparison, in order to identify changes in hardwood defoliation caused by gypsy moth advances. Defoliation was most accurately detected by the image differencing and PCA approaches. Michener and Houhoulis (1997) cross-referenced composite analysis, PCA and univariate image differencing for monitoring vegetation responses to extensive flooding in south-west Georgia. Univariate NDVI differencing most accurately identified vegetation changes in their multi-temporal SPOT HRV dataset. Comparable conclusions were reached by Macleod and Congalton (1998) when comparing post-classification comparison to univariate image differencing and PCA algorithms: image differencing performed significantly better in monitoring alterations in aquatic vegetation ecosystems encompassing submerged eelgrass. In a pilot study on North American landscape characterization (NALC) and land-cover change detection, Yuan and Elvidge (1998) systematically tested and evaluated 75 change detection methods using both visual and statistical procedures. Initial results again suggested image differencing resulting in better final accuracy results when compared to the other algorithms. While Sohl (1999) found specific quantitative values of change most accurately and efficiently provided by an enhanced image differencing algorithm, he cited the CVA as excelling at providing rich qualitative detail about the nature of the change. Mas (1999), on the other hand, obtained the highest accuracy in land-cover change detection in a coastal zone of Mexico using post-classification comparison. In single band analysis, selective PCA even outperformed image differencing due to its capacity to more efficiently remove inter-image variability left over after radiometric normalization.
Li and Yeh (1998) indicated a clear final-accuracy-based supremacy of PCA over post-classification comparison in urban expansion change detection.

Collins and Woodcock (1996) have compared three multi-temporal linear data transformation algorithms: PCA, tasselled cap and Gramm–Schmidt orthogonalization. Tasselled cap and PCA both gave better results than the Gramm–Schmidt technique. However, the authors recommended the Kauth–Thomas approach because it identified change in a more consistent and interpretable manner. Rogan et al. (2002), on the other hand, found lower classification accuracy results for the Kauth–Thomas approach when compared with multi-temporal spectral mixture analysis using decision tree classification.

4.2. Temporal trajectory analysis

Temporal trajectory analysis requires the comparison of temporal development curves, also called time trajectories or time profiles, of different relevant indicators, and this for successive growing seasons or years. The inherent high temporal frequency in data acquisition not only expedites the detection of ecosystem modifications, but also greatly facilitates the characterization of phenological variations in ecosystem status. When the time trajectory of one or several remotely sensed indicators for a particular pixel departs from the normal (or average, or optimal, depending on the objectives of the study), a seasonal or inter-annual change event or process is detected (Lambin and Strahler 1994). In order to bring this about, several investigators computed simple anomalies in time profiles of vegetation indices. For example, Myneni et al. (1997) and Plisnier et al. (2000) calculated a change parameter as the vegetation index value for a given month minus the average value for the time series, divided by the standard deviation. The use of signal processing techniques such as Fourier analysis to examine frequency distributions of high temporal resolution AVHRR data has also been documented. The algorithm showed definite capabilities in determining seasonal and sub-seasonal variability in Brazilian Amazon forest cover (Andres et al. 1994).

Standardized PCA of regional to continental scale time series from wide-angle sensors also proved a powerful technique to separate changes that were taking place at different time frequencies, e.g. decadal time-scale changes in productivity, seasonal changes, sensor-related value drifts, and vegetation index value variations related to the El Niño Southern Oscillation (ENSO) phenomenon (Eastman and Fulk 1993, Young and Wang 2001). Finally, the change vector analysis method has been adapted to temporal trajectory analysis by computing a change vector in a multi-temporal feature space rather than in a multi-spectral feature space (Lambin and Strahler 1994). This allows detecting the magnitude of change as well as the nature of land-cover change, through an analysis of the directions of the change vectors (Lambin and Ehrlich 1997).

Temporal trajectory analysis methods have been implemented with different wide field-of-view, high temporal resolution sensors (AVHRR, SeaWIFS, VEGETATION) and different indicators (vegetation indices, surface temperature, spatial structure data). With the availability of still relative short time series (10 to a couple of years), mostly climate-driven fluctuations in surface conditions of natural ecosystems and natural disaster (drought, flood, fire) impacts have been detected (Behrenfeld et al. 2001, Lupo et al. 2001). In addition to allowing an unambiguous detection of many abrupt changes in different ecosystems, temporal trajectory analysis has proven sensitive to more subtle alterations in primary productivity,
vegetation phenology, ecosystem dynamics and seasonality, often more so than classical bi-temporal approaches. Even in a multi-stage context, the latter techniques may suffer from an obvious under-sampling at the time-scale. This phenomenon has proven especially problematic when dealing with abrupt and often brief ecological events such as fire, flooding, dry spells and vegetation stress. However, given the typical coarse spatial resolution and the large area coverage of imagery from wide field-of-view sensors, validation with independent datasets is a major challenge for ecosystem monitoring.

When cross-referencing temporal trajectory analysis and bi-temporal change detection methods, Borak et al. (2000) found statistically significant relationships between AVHRR time-profile-based change metrics and classical change indicators computed from Landsat and SPOT imagery for several African study sites.

5. New developments

The focus of the above review has been on operational data enhancement and data analysis procedures for ecosystem change detection with digital imagery, especially satellite acquired. There are, however, a number of recent and expected advancements that will increase the accuracy and effectiveness of change detection with digital satellite sensor data. Further improvements in hard- (sensor and computer) and software systems, in ecosystem management models, and in change detection algorithms and models, can all be expected to improve the capability of satellite remote sensing for ecosystem monitoring. Since the mid 1990s we have witnessed the development and launch of an unprecedented number of satellite sensing systems. It now appears that by 2005 there will be 15–20 commercial and government satellites acquiring data at spatial resolutions of 1–30 m. In addition to the technical improvements, the 1992 US Land Remote Sensing Act now also provides for the distribution of the data at the cost of reproduction. Access to space-based digital information has thus become much better and continues to do so.

Although geographic information systems (GIS) are not a new development, it is only in the last decennium that they have gained widespread acceptance as a practical tool for ecosystem management. The incorporation of GIS technology in digital change detection methods enables the delivery of change maps, derived from any descriptive change model, in a timely fashion at scales that are consistent with ecosystem management objectives. Today, most image processing systems are integrated with, or at least compatible with, GIS systems, and classifications of remotely sensed data are commonly viewed as inputs to GIS. At the same time increasing attention is being given to using GIS data layers as ancillary inputs to classification of remote sensing data. Further developments in image analysis and display systems, and in the integration of graphical user interfaces, database management systems, statistical analysis (including spatial statistics) and process modelling subroutines, along with advanced GIS and image analysis functions, are to be expected.

Artificial intelligence or knowledge-based expert systems offer further opportunities. They provide a way to integrate other features of vegetative cover categories besides spectral change information, thereby overcoming some of the limitations of the traditional statistical classifiers. Such change category recognition methods make use of existing or prior knowledge of the scene content (e.g. original ecosystem status, location, size, relationship with other cover types, shape, socio-economic data, etc.) to guide and assist the classification, which follows spatial reasoning lines. As such, they parallel the interpretative procedures employed by a
photo-interpreter much more closely. The methods assume that there are similar and identifiable characteristics that each cover type possesses. Although artificial intelligence approaches to natural ecosystem change detection have largely remained in a conceptual design stage (McRoberts et al. 1991), researchers developing ecological models have started incorporating inputs from remote sensing and GIS techniques to analyse spatial patterns and processes (Ustin et al. 1993, Mladenoff and Host 1994). Also the incorporation of econometric techniques has been documented (Kaufmann and Seto 2001). Land cover and change determined from remotely sensed imagery are integral components of such models.

In recent years, the use of machine-learning algorithms, among which artificial neural networks and decision tree classifiers, has gained considerable attention as an alternative to conventional approaches such as the maximum likelihood classification (Benediktsson et al. 1990, Bischof et al. 1992). Increased classification accuracy is often cited as the primary reason for developing and applying these techniques. However, machine-learning algorithms can also be computationally very complex and require a considerable number of training samples. Documented implementation of machine-learning algorithms in the change detection environment is rather scarce and of recent dates, including the use of multi-layer perceptrons (Gopal and Woodcock 1996, Dai Long and Khorram 1999), of learning vector quantization (Chang Cheung-Wai et al. 2001), and of decision tree classifiers such as, for example, Quinlan’s C5.0 (Chang Cheung-Wai et al. 2001). The first study also provided evidence that the neural network approach made use of the same spectral signals and scene characteristics as bi-temporal linear data transformation algorithms, e.g. the Gramm–Schmidt orthogonalization. This suggests that the nonlinearities in the relationship between the spectral inputs and, in this case, forest mortality patterns as the change event, accounted for the improved results using the neural network (with back-propagation training). While perceptron procedures were reported as the most difficult to replicate, tree classifiers were the easiest to use, and learning vector quantization the best performer where it concerned change detection accuracy (Chang Cheung-Wai et al. 2001). Although the implicit use of the time dependency of the datasets represents a major advantage of machine-learning algorithms in bi-temporal change analysis, it is not clear yet if the benefits of the higher accuracy outweigh the cost of the additional training data.

Lastly, one can look forward to the development of improved and new change detection algorithms and models. Bruzzone and Serpico (1997a) conceptualized the unsupervised SMI or “selective use of multi-spectral information” algorithm to reduce the effect of bi-temporal registration noise, and a supervised non-parametric iterative technique based on the compound classification rule for minimum error (Bruzzone and Serpico 1997b) to improve change detection accuracy. To provide a richer information base on class membership and its dynamics, Foody and Boyd (1999) suggested running the post-classification comparison algorithm with fuzzy classifications instead of with conventional hard ones. The effect was especially significant when the ecosystem changes were operating at a scale finer than the spatial resolution of the sensor. Morisette et al. (1999) explored the use of generalized linear models (GLM) for enhancing standard methods of satellite-based land-cover change detection. GLMs were shown to be helpful in examining different change metrics and useful by applying the resulting model throughout the image to get a probability of change estimate as well as pixel-specific estimates of the variability of change estimate. Smits and Myers (2000) investigated a method that can be used to characterize and understand the spatial behaviour of change by decomposing the change intensity image into a tree of entities
called echelons. They indicated that such a tree could be extremely helpful in discovering connections between changes. In 2001, Hazel advocated the use of object-level change detection (OLCD) instead of the traditional pixel-based algorithms, and Yamamoto and Hanaizumi (2001) proposed three-dimensional segmentation for temporal change detection.

6. Conclusions

The data-gathering capabilities of space-borne remote sensors have generated great enthusiasm over the prospect of establishing remote sensing based systems for the continuous monitoring of ecosystems. Although Aldrich’s prediction in 1975 of the accuracy of satellite remote sensing for monitoring forest change was not quickly or easily achieved, today it is well established that remote sensing imagery, particularly digital data, can be used to monitor and map changes in ecosystems. It has been demonstrated (e.g. Collins and Woodcock 1994, Coppin and Bauer 1995) that it is feasible to develop automated forest cover monitoring methodologies. When the inherent limitations of digital approaches are appropriately dealt with, pre-processing is adequately incorporated and optimal change detection algorithms are selected, then Aldrich’s prediction can be met.

The in-depth analysis of a very substantive number of change detection studies has demonstrated that the choice of change detection algorithm was in almost all cases pragmatic rather than scientifically based, and that most authors report success rather than the extent of the shortcomings of their approach(es). In other words, the selection was driven more by the application itself, than by the main issues of change monitoring in general (see §2.1). This makes it very difficult to draw summarizing comparative statements. Table 1 is nevertheless an attempt to cross-evaluate the bi-temporal change detection algorithms, as categorized under §4.1, against these change monitoring issues. The qualitative cross-evaluations represent majority use as reported in the literature, not a judgement by the authors of this manuscript.

As many change detections are application- or context-specific, they should be viewed as complementary to each other. A parallel implementation of several change detection methods followed by an integration of results is thus the most effective way to detect change in a wide range of environments (Zhang et al. 2002).

It is clear that temporal trajectory analysis offers the greatest opportunities to meet all challenges described in §2.1, but its major drawback is nowadays inherently tied to the coarse spatial resolution of the imagery, and the limitations on the available time series length. Therefore, the implementation of the technique is still largely limited to the study of dynamic processes and their effect on vegetation on regional to global scales (e.g. for ENSO; Young and Wang 2001). Here also, temporal trajectory methods applied at broad spatial scales to identify areas of rapid change are highly complementary to bi-temporal change detection methods than can be used to zoom in at finer resolutions over areas where change has been identified at coarser resolutions. Any effective regional to global scale monitoring system of ecosystem change should be based on such a multi-scale, nested approach, with different change detection methods applied at each scale.

Although all of the possible change detection methods have not been applied to the same data for cross-evaluation, it is evident from this review that:

1. Vegetation indices are more strongly related to changes in the scene than the responses of single bands.
Table 1. Qualitative cross-evaluation of bi-temporal change detection algorithms against main ecosystem change monitoring issues (see §2.1).

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Modifications vs conversions</th>
<th>Abrupt changes vs progressive changes</th>
<th>Annual variability vs secular events</th>
<th>Sampling rates vs process scales</th>
<th>Scale dependence</th>
<th>Typical example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-classification comparison composite analysis</td>
<td>conversions</td>
<td>abrupt changes</td>
<td>secular events</td>
<td>simply bi-temporal sampling</td>
<td></td>
<td>forest succession; Hall et al. 1991b</td>
</tr>
<tr>
<td>Univariate image differencing</td>
<td>modifications</td>
<td>progressive changes</td>
<td>inter-annual variability</td>
<td>simply bi-temporal sampling</td>
<td></td>
<td>defoliation; Hall et al. 1984</td>
</tr>
<tr>
<td>Image ratioing</td>
<td>conversions</td>
<td>abrupt changes</td>
<td>secular events</td>
<td>simply bi-temporal sampling</td>
<td></td>
<td>savanna monitoring; Serneels et al. 2001</td>
</tr>
<tr>
<td>Bi-temporal linear data transformation</td>
<td>conversions and modifications</td>
<td>abrupt and progressive changes</td>
<td>inter-annual variability and secular events</td>
<td>bi-temporal sampling at different process-related intervals</td>
<td>up-scaling from pixel to landscape entity</td>
<td>land cover change; Howarth and Wickware 1981</td>
</tr>
<tr>
<td>Change vector analysis</td>
<td>conversions and modifications</td>
<td>abrupt and progressive changes</td>
<td>inter-annual variability and secular events</td>
<td>bi-temporal sampling at different process-related intervals</td>
<td></td>
<td>forest stand monitoring; Coppin and Bauer 1994</td>
</tr>
<tr>
<td>Image regression</td>
<td>conversions</td>
<td>abrupt changes</td>
<td>secular trends</td>
<td>simply bi-temporal sampling</td>
<td></td>
<td>land cover change; Lambin and Strahler 1994</td>
</tr>
<tr>
<td>Multi-temporal spectral mixture analysis</td>
<td>modifications</td>
<td>progressive changes</td>
<td>inter-annual variability</td>
<td>simply bi-temporal sampling</td>
<td></td>
<td>tropical deforestation; Singh 1989</td>
</tr>
<tr>
<td>Multidimensional temporal feature space analysis</td>
<td>modifications</td>
<td>progressive changes</td>
<td>inter-annual variability</td>
<td>simply bi-temporal sampling</td>
<td></td>
<td>tropical forest monitoring; Roberts et al. 1998</td>
</tr>
</tbody>
</table>

P. R. Coppin et al.
2. Precise registration of multi-date imagery is a critical prerequisite of accurate change detection. However, residual mis-registration at the below-pixel level somewhat degrades area assessment of change events at the change/no-change boundaries.

3. Some form of image matching or radiometric calibration is recommended to eliminate exogenous differences, for example due to differing atmospheric conditions, between image acquisitions. The goal should be that following image rectification, all images should appear as if they were acquired with the same sensor, while observing through the atmospheric and illumination conditions of the reference image.

4. Image differencing and linear transformations appear to perform generally better than other bi-temporal change detection methods. Differences among the different change maps and their accuracies are undoubtedly related to the complexity and variability in the spatial patterns and spectral–radiometric responses of ecosystems, as well as to the specific attributes of the methods used.

5. Patterns of seasonal and inter-annual variations in land surface attributes, which can be driven by climatic variability (e.g. ENSO), natural disasters (e.g. fires, floods), land-use changes (e.g. deforestation) or global climate change (e.g. climate warming), can be detected using high temporal frequency data from wide field of view sensors, provided that great care has been taken to remove sensor-related artefacts in time series and that an appropriate profile-based change detection method is applied. This research area is still in its infancy compared to the more classic bi-temporal change detection techniques used with medium to fine spatial resolution remote sensing data.

6. There is a high complementarity between different change detection methods. This is certainly true when one seeks to detect a wide range of ecosystem changes at one given scale. It also applies to the design of multi-scale monitoring systems that combine methods adapted to detect changes at regional to global scales with methods better suited for landscape-scale temporal analyses. While the former can be implemented continuously over large territories, the latter could only be applied where and when a change has been detected at a broader scale.

7. The capability of using remote sensing imagery for change detection will be enhanced by improvements in satellite sensor data that will become available over the next several years, and by the integration of remote sensing and GIS techniques, along with the use of supporting methods such as expert systems and ecosystem simulation models.

Analysis of the literature provides ample evidence to support the conclusion that multi-date satellite imagery can be effectively used to detect and monitor changes in ecosystems. At the same time we agree with the observation of Collins and Woodcock (1996) that one of the challenges confronting the remote sensing research community is to develop an improved understanding of the change detection process on which to build an understanding of how to match applications and change detection methods.

References


Chilar, J., Du, Y., and Latifovic, R., 2001, Land cover dependence in the detection of


COLWELL, J. E., DAVIS, G., and THOMSON, F., 1980, Detection and measurement of changes in the production and quality of renewable resources. USDA Forest Service Final Report 145300-4-F, ERIM, Ann Arbor, MI, USA.


HAME, T. H., 1986, Satellite image aided change detection. In Remote sensing-aided forest inventory, Research Notes No.19, Department of Forest Mensuration and Management, University of Helsinki, Helsinki, Finland, pp. 47–60.

HAME, T. H., 1988, Interpretation of forest changes from satellite scanner imagery. In Satellite imageries for forest inventory and monitoring: experiences, methods, perspectives, Research Notes No.21, Department of Forest Mensuration and Management, University of Helsinki, Helsinki, Finland, pp. 31–42.


LUND, H. G., 1983, Change: now you see it—now you don’t!. *Proceedings of the International Conference on Renewable Resource Inventories for Monitoring Changes and Trends, Oregon State University, Corvallis, OR, USA* (Corvallis, OR: Oregon State University), pp. 211–213.


SADER, S. A., 1988, Remote sensing investigations of forest biomass and change detection in tropical regions. In Satellite imageries for forest inventory and monitoring: experiences, methods, perspectives, Research Notes No.21, Department of Forest Mensuration and Management, University of Helsinki, Helsinki, Finland, pp. 31–42.


van LEEUWEN, W. J. D., HUETE, A. R., and LAING, T. W., 1999, MODIS vegetation index


