Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas

S.E. Franklin\textsuperscript{a,*} and M.A. Wulder\textsuperscript{b}

\textsuperscript{a}Department of Geography, University of Calgary, Calgary, Alberta, Canada T2N 1N4
\textsuperscript{b}Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Canada V8Z 1M5

Abstract: Numerous large-area, multiple image-based, multiple sensor land cover mapping programs exist or have been proposed, often within the context of national forest monitoring, mapping and modelling initiatives, worldwide. Common methodological steps have been identified that include data acquisition and preprocessing, map legend development, classification approach, stratification, incorporation of ancillary data and accuracy assessment. In general, procedures used in any large-area land cover classification must be robust and repeatable; because of data acquisition parameters, it is likely that compilation of the maps based on the classification will occur with original image acquisitions of different seasonality and perhaps acquired in different years and by different sensors. This situation poses some new challenges beyond those encountered in large-area single image classifications. The objective of this paper is to review and assess general medium spatial resolution satellite remote sensing land cover classification approaches with the goal of identifying the outstanding issues that must be overcome in order to implement a large-area, land cover classification protocol.

Key words: land cover mapping, large-area classification, remote sensing.

1 Introduction

Satellite remote sensing has long been considered an ideal technology and data source for large-area land cover classifications (Gregory, 1971; Saint, 1980; Iverson \textit{et al}., 1989). This has resulted in numerous national, regional, continental and global land cover mapping efforts, including many nationally based forest inventory and land cover...
mapping efforts by governments and independent organizations to support regional landscape planning and integrated resource management (Table 1). The emergence of such large-area mapping programs has been accompanied by comprehensive political activities reflecting continuing concern over deforestation, environmental degradation, conversion or changes in land cover and land use, and the impact on the global carbon budget and linkages between national economies, development, resources and complex ecological dynamics (Bryant, 1997; Ahern et al., 1998; Franklin, S.E., 2001).

The wide availability of spatially extensive Advanced Very High Resolution Radiometer (AVHRR) 4 km and 1 km data sets possessing high temporal resolution beginning in the 1980s helped spur the research community to develop methods to characterize the land cover of large regions (Tucker et al., 1985; Cihlar et al., 1994) and, eventually, to create global land cover data sets (Eidenshink and Faundeen, 1994; Justice and Townshend, 1994; Sellers et al., 1994; Thomlinson et al., 1999). Loveland et al. (1991) developed a land cover map of the conterminous USA with AVHRR data. They also used available ancillary digital elevation model (DEM) and climate data, ecoregions and digital land-use and land cover maps available through the process of aerial photointerpretation over a 20-year period since ‘... ancillary data provide information crucial to the success of large-area land-cover characterization based on NOAA AVHRR data’ (Brown et al., 1993: 984). In Canada, a land cover classification covering the entire country was obtained using multitemporal AVHRR 1 km data (see Table 2; Cihlar et al., 1997a,b, 1998, 1999, 2000). The resulting map showed good separability in more than 34 land cover classes (14 forest classes, 10 open land classes, more than 10 developed land classes).

These AVHRR mapping efforts are complementary to those resulting from finer spatial resolution products such as those generated from Landsat and free-flying satellite synthetic aperture radar (SAR) sensors, and image mosaics now becoming available from MODIS and other Earth Observing System (EOS) sensors. Typically, these are medium spatial resolution data that capture land cover characteristics with details smaller than 1 ha (Table 3). Their use is also accompanied by the tendency to include data from more than one sensor in the mapping procedure. For example, Mayaux et al. (2000) prepared a regional forest cover map of $20 \times 10^6$ ha of the Congo Basin by the combination of ‘best sensors available’ (the Central Africa Mosaic Project (CAMP) 470 image ERS-1 SAR mosaic, available data from the ERS-1 ATSR, and a 10-year time series of AVHRR data). Accuracy assessments, conducted by a visual comparison and a cross-tabulation procedure of the new multisensor map with two independently generated remote sensing maps suggested the thematic content and spatial detail of land cover map produced by the best available sensor data were superior to single-source products.

Numerous methodological problems are associated with using data from sensors with higher spatial resolution than AVHRR, and when using different spatial and spectral resolutions in a single mapping procedure (see Table 4; Hall et al., 1991a; Fuller et al., 1998; Muller et al., 1999). Critical decisions within the procedures used to develop such maps range from issues related to image acquisition and cost, availability of image processing systems and expertise, and the full range of radiometric and geometric pre-processing that must be implemented to ensure compatibility across space and time. Other issues include cloud removal and masking of ‘unchanging’ features, task or process automation and the possible role and integration of other geospatial
<table>
<thead>
<tr>
<th>Long Form Name</th>
<th>Name</th>
<th>Scale</th>
<th>Resolution</th>
<th>Study Area</th>
<th>Sensor(s)</th>
<th>URL</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-ordination of</td>
<td>CORINE</td>
<td>Continental/ National/ Regional</td>
<td>Composite</td>
<td>Europe</td>
<td>MSS/TM</td>
<td><a href="http://etc.satellus.se/the_data/overview.htm">http://etc.satellus.se/the_data/overview.htm</a></td>
<td><a href="#">Fuller et al., 1998</a></td>
</tr>
<tr>
<td>Information on the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gap Analysis Program</td>
<td>GAP</td>
<td>Continental/ National/ Regional</td>
<td>Fine (1–30 m)</td>
<td>USA (some other areas)</td>
<td>TM</td>
<td><a href="http://www.gap.uidaho.edu/">http://www.gap.uidaho.edu/</a></td>
<td>See state reports</td>
</tr>
<tr>
<td>Trees Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remote Sensing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Present Land Use Mapping at 1:250 000, 1995 Province of British Columbia Ministry of Environment, Lands and Parks Surveys and Resource Mapping Branch
<table>
<thead>
<tr>
<th>Long Form Name</th>
<th>Name</th>
<th>Scale</th>
<th>Resolution</th>
<th>Study Area</th>
<th>Sensor(s)</th>
<th>URL</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeric Terradynamic Simulation Group</td>
<td>NTSG</td>
<td>Global</td>
<td>Coarse (250 m+)</td>
<td>Global</td>
<td>AVHRR</td>
<td><a href="http://www.forestry.umt.edu/nts/Projects/landcover/">http://www.forestry.umt.edu/nts/Projects/landcover/</a></td>
<td>Nemani and Running, 1996</td>
</tr>
<tr>
<td>International Geosphere–Biosphere Programme</td>
<td>IGBP</td>
<td>Global</td>
<td>Coarse (250 m+)</td>
<td>Global</td>
<td>MODIS (AVHRR)</td>
<td><a href="http://geography.bu.edu/landcover/">http://geography.bu.edu/landcover/</a></td>
<td>Friedl et al., 2000</td>
</tr>
<tr>
<td>Geography</td>
<td>Date of Collection</td>
<td>Resolution</td>
<td>Region(s)</td>
<td>Source(s)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>--------------------</td>
<td>------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tropical Ecosystems Environmental Observation by Satellite</td>
<td></td>
<td>Composite</td>
<td>SE Asia: Continental and Insular West and Central Africa, Madagascar Latin America (Central and South)</td>
<td>AVHRR/ATSR/V GT/TM5 POT/ERS TRESS.HTML Malingreau et al., 1993</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td></td>
<td>Coarse (250 m+)</td>
<td>USA</td>
<td>AVHRR <a href="http://www.ciesin.org/docs/005-347/005-347.html">http://www.ciesin.org/docs/005-347/005-347.html</a> Loveland et al., 1991</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring and Assessment of Resources In Europe – Forest</td>
<td></td>
<td>Fine (1–30 m)</td>
<td>Europe (5 test sites)</td>
<td>TM <a href="http://www.ears.nl/MARIEF/index.htm">http://www.ears.nl/MARIEF/index.htm</a> MARIE-F final report, December 1999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land Cover Map 1990</td>
<td></td>
<td>Fine (1–30 m)</td>
<td>UK</td>
<td>TM <a href="http://www.ceh.ac.uk/products_services/data/lcm.html">http://www.ceh.ac.uk/products_services/data/lcm.html</a> 1994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humid Tropical Forest Deforestation Project</td>
<td></td>
<td>Composite</td>
<td>Congo Basin/Pan Amazon (UMD); Amazon/SE Asia (MSU)</td>
<td>MSS/TM <a href="http://glcf.umd.edu/documents/pfinder.html">http://glcf.umd.edu/documents/pfinder.html</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long Form Name</td>
<td>Name</td>
<td>Scale</td>
<td>Resolution</td>
<td>Study Area</td>
<td>Sensor(s)</td>
<td>URL</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>---------------</td>
<td>--------------------------------</td>
<td>------------------</td>
<td>-----------------------------</td>
<td>-----------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Land Use in Temperate East Asia</td>
<td>LUTEA</td>
<td>Continental/National/Regional</td>
<td>Coarse (250 m+)</td>
<td>China, Japan, N &amp; S Korea, Mongolia, Asian Russia</td>
<td>AVHRR</td>
<td><a href="http://nrel.colostate.edu/projects/lutea/lutea.htm">http://nrel.colostate.edu/projects/lutea/lutea.htm</a></td>
<td></td>
</tr>
<tr>
<td>Multi-Resolution Land Characteristics</td>
<td>MRLC</td>
<td>Continental/National/Regional</td>
<td>Fine (1–30 m)</td>
<td>USA</td>
<td>TM</td>
<td><a href="http://www.epa.gov/mrlc/">http://www.epa.gov/mrlc/</a></td>
<td>Vogelmann et al., 1998b</td>
</tr>
</tbody>
</table>
information. Approaches to general land cover classification of large areas with diverse data sets suggested in the literature range from multispectral classifiers in an unsupervised protocol (e.g., hyperclustering and labeling) (Homer et al., 1997) to advanced classification models that rely on training areas developed using various supervised methods (Carpenter et al., 1997, 1999). Not yet clear is the role of context classifiers and ‘reclassifiers’ (Groom et al., 1996), which could assume various positions within a decision-tree (Hansen et al., 1996) that might also include neural networks or fuzzy logic algorithms (Foody, 1999).

Table 2  Processing steps used in the creation of the landcover map of Canada from multitemporal, single-season AVHRR data (after Cihlar et al., 1999)

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Calibration of daily AVHRR images for post-launch degradation;</td>
</tr>
<tr>
<td>2</td>
<td>Georeferencing and resampling;</td>
</tr>
<tr>
<td>3</td>
<td>10-day cloud-free compositing;</td>
</tr>
<tr>
<td>4</td>
<td>Conversion of top-of-atmosphere radiances to surface reflectance with a standard viewing geometry (atmospheric correction; detection of snow, subpixel clouds, smoke, other significant atmospheric contamination; bidirectional corrections; NDVI corrections for solar zenith angle effects; replacement of contaminated pixels through temporal interpolation);</td>
</tr>
<tr>
<td>5</td>
<td>Removal of nongrowing season image pixels;</td>
</tr>
<tr>
<td>6</td>
<td>Application of a water mask;</td>
</tr>
<tr>
<td>7</td>
<td>Hyperclustering and labelling (details of the algorithm dubbed ‘Classification by Progressive Generalization’ (CPG) provided by Cihlar et al. (1998));</td>
</tr>
<tr>
<td>8</td>
<td>Development of a final classification legend (for the map product);</td>
</tr>
<tr>
<td>9</td>
<td>Postclassification refinement;</td>
</tr>
<tr>
<td>10</td>
<td>Accuracy assessment (comparisons with existing analogue and digital TM classifications).</td>
</tr>
</tbody>
</table>

Table 3  Relationship between scale and spatial resolution in satellite-based land cover mapping programs (after Franklin, S.E., 2001)

<table>
<thead>
<tr>
<th>Resolution Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low spatial resolution imagery</td>
<td>Optimal applications are in study of phenomena that can vary over 100s or 1000s of metres (small scale) and could be supported with GOES, NOAA AVHRR, EOS MODIS, SPOT VEGETATION sensor data</td>
</tr>
<tr>
<td>Medium spatial resolution imagery</td>
<td>Optimal applications are in study of phenomena that can vary over 10s or 100s of metres (medium scale) typically with imagery from sensors on the Landsat, SPOT, IRS, JERS, ERS, Radarsat and Shuttle platforms</td>
</tr>
<tr>
<td>High spatial resolution imagery</td>
<td>Optimal applications are in study of phenomena that can vary over scales of centimeters to metres (large scale), are currently supported by aerial remote sensing platforms, IKONOS and very specific applications of coarse resolution satellite imagery (e.g., coarse pixel resolution unmixing studies)</td>
</tr>
</tbody>
</table>
The full range of possible methods and the rationale for each step in a classification protocol for land cover mapping of large areas has never been fully documented such that the advantages and disadvantages can be understood in the context of emerging national land cover mapping and change detection programs (e.g., Goodenough et al., 1998; Lunetta et al., 1998; Lunetta and Elvidge, 1999). In this paper we provide a review of the issues in large-area, medium spatial resolution satellite image land cover mapping. First, we discuss several existing projects that have generated initial land cover maps of large portions of Africa, South America, North America and Europe. Then, the methodological issues associated with these projects are reviewed, including the development of classes and map legends, image selection, georadiometric processing of imagery, and selection of mapping variables, classification approach, decision rules and validation approach. We conclude this review with a summary of the issues in establishing a protocol for conducting large-area land cover classifications with satellite remote sensing data.

II High spatial detail, large-area, land cover mapping applications

An obvious new requirement in multiple image applications is the need for image mosaicing to create a seamless, large-area database that would cover whole regions of similar ecology and physiography (Homer et al., 1997) or political units (e.g., states, countries) (Gaston et al., 1997). Homer and Gallant (2001) introduced the concept of ‘mapping zones’ in the USA; the idea is that within structurally similar ecoregions or physiographic zones there may be image mosaics created with similar data acquisition characteristics, which could then be handled together in a precategorization stratification approach (Pettinger, 1982) to maximize spectral differentiation and provide a means to facilitate partitioning the workload into logical, feasible units. This approach represented a reasonable compromise between the need to reduce individual workload demands and the need to work within physiographically defined areas (Table 5). Image-by-image processing results in difficult edge-class matching and tedious manual editing steps, even in relatively small image mosaics. For example, when using a mosaic
of eight Thematic Mapper (TM) images in Maine, Hepinstall et al. (1999) suggested that the disadvantages of edge matching greatly exceeded the advantages of speed of processing and consistency in class definition.

1 SAR mosaics

All-weather sensing capability was noted by De Grandi et al. (2000a) as a requirement to generate a regional, medium spatial resolution, land cover map of tropical Africa as part of the Global Rain Forest Mapping project; 3900 ERS-1 SAR scenes were used in the mosaic. Internal calibration was a semi-automated procedure; the goal was to calibrate radar reflectivity as input to the generation of thematic products either by visual interpretation or automatic supervised image classification. Artifacts introduced by sensor attitude errors, a range amplitude pattern, striping and between-scene gain imbalance were empirically corrected. Not corrected were radiometric problems associated with topography or variations in imagery resulting from changing single-season vegetation phenology and ‘different eco-climatic states of vegetation classes as a function of latitude’ (De Grandi et al., 2000b: 1246).

Image data and texture data were used in the supervised classification of local areas scaled up to the entire mosaic (stratified and locally filtered). The map was validated in a two-stage process with TM imagery and other available maps; accuracies in the 68-70% range were reported for classes including water, swamp forest, lowland rain forest and a single class of ‘other land cover types’ (De Grandi et al., 2000b).

A similar project for the Amazon was described by Siqueira et al. (2000); more than 1500 JERS-1 scenes were used to create a single-season (September to December 1995) mosaic. JERS-1 data were selected for this application partly because of ‘difficulties in interpreting the spectral information of Landsat data acquired at different years or seasons’ (Saatchi et al., 2000: 1202). The image data were resampled to create a 1 km spatial resolution land cover map with 14 land cover and vegetation classes within five general land cover types driven by a supervised classification method that incorporated texture decision rules. The supervised classification step was used to generate five general categories that corresponded well with a regional physiographic/climate-based vegetation map; then, a hierarchical decision rule based on texture measures was imposed on these classes to separate vegetation types based on taxonomy and woody biomass levels. Validation consisted of comparisons with existing (local) vegetation maps, AVHRR-based land cover maps, and selected field sites; 78% accuracy in 14 vegetation classes was reported.

The difficulties in interpreting radar backscatter data, compared with optical spectral

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Advantages of physiographical stratification for definition of mapping zones (after Homer and Gallant, 2001)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Obtain economics of size by considering scale issues</td>
</tr>
<tr>
<td>2</td>
<td>Apply boundary definition through physiography</td>
</tr>
<tr>
<td>3</td>
<td>Reduce total number of classes by considering land cover distribution</td>
</tr>
<tr>
<td>4</td>
<td>Maximize spectral uniformity</td>
</tr>
<tr>
<td>5</td>
<td>Minimize edge-matching</td>
</tr>
</tbody>
</table>
data, are not inconsiderable (Gerstl, 1990); the influence of topography, for example, is an incompletely understood phenomenon in both optical and radar remote sensing, but clearly exerts a more debilitating effect in digital classifications of SAR imagery (Bayer et al., 1991). However, based on the large-area classification work in Congo and Amazon forests (De Grandi et al., 2000a,b; Saatchi et al., 2000), it seems likely that general land cover classes can be readily mapped (perhaps manually rather than digitally) with satellite SAR data. Earlier experiences in working with individual satellite radar images in a wide variety of the world’s forest environments would appear to support this view (Ranson and Sun, 1994; Thompson and Macdonald, 1995; Michelson et al., 2000). A new challenge is the combined use of SAR and optical information (Haack et al., 2000); for example, by using available SAR data to ‘fill-in’ holes in a Landsat multiscope mosaic such that the classification process does not need to be significantly modified to take into account the different characteristics of the data (Cihlar, 2000).

2 Landsat mosaics

The classification of 24 Landsat TM scenes mosaiced together and radiometrically normalized was described for a state-wide mapping application in Utah by Homer et al. (1997). Fourteen base scenes were used to construct the mosaic; ten scenes were used as secondary cloud patch scenes. All scenes used were acquired in June, July or August during full leaf-out conditions. The base scenes were acquired in 1988 or 1989 and the cloud patch scenes were acquired over a span of nine years. The imagery were atmospherically adjusted using a histogram-matching process in overlap areas to a single master scene that appeared to contain a wide range of environments to be mapped. Ecological stratification was introduced at this stage of the analysis to ensure the complete sampling of scene variability; this approach was considered superior to individual scene clustering because within-scene clustering differences could be avoided and edge-matching was not required.

The classification process was initiated with unsupervised clustering (100–125 clusters targeted) separately in three large ecoregions. The next step was to develop the relationship between the clusters and field classes by photointerpretation and field visits. Collecting training data independent of the classification optimized field time, but was thought to be inadequate in resolving several spectral clustering difficulties. Subsequently, training polygons and each class were modelled using ecological rules that included topographic information from a DEM or information on land-use obtained from ancillary maps; significant post-classification ancillary modeling or cluster-busting (iterative masking) was required. This post-classification modelling was the key to analysing and framing local spectral class segments into the large-area final map classes. This final map of the state yielded 36 covetypes and was found to be approximately 75% correct.

The Utah project was one of a series of Landsat TM-based land cover mapping projects that adopted the US Gap Analysis protocol (Scott et al., 1996). The GAP program accommodated a wide range of techniques depending on the size of the area (State-based) and the particular status of land cover mapping and data resources in each project. A general goal was to achieve 85% classification accuracy across all mapping classes with Landsat TM data. To achieve this goal, the Maine Gap project
provided several recommendations, including the application of stratification procedures before image classification (rather than after), a more integrated use of the aerial videography and satellite data (particularly when more than one image date is to be classified) and more detailed use of ancillary data (e.g., climate and geomorphic data to predict vegetation types within the broad classes) (Table 6).

In Alaska, Muller et al. (1999) created two large-area Landsat Multispectral Scanner (MSS) mosaics: (1) a 13-scene mosaic of the Kuparuk River watershed in the Alaska interior, and (2) a 26-scene mosaic of a larger area of Alaska covering almost the entire North Slope. Images were acquired over a 10-year period from July to September. The images were processed for noise and sensor striping, and radiometrically matched to an early August 1985 scene. Eight classes were of interest (barrens, three types of tundra, shrublands, water, clouds and ice, and shadows). Two bands were used in an ISODATA K-means clustering in which a total of 69 clusters were specified and interpreted for ecological relevance. Several clustering iterations were needed in which land cover classes within clusters were ‘broken out’ in ways that conformed to detailed field knowledge and expertise in this environment. Some cloud and mountain shadows were misclassified as water or wet tundra and were fixed manually by digitizing on the screen. One class (tundra on sand dunes in one small area) was manually delineated from available surficial geology maps, in essence creating a new class that was not present in the clustering process. The maps were filtered and resampled to 100 m from 50 m to facilitate accuracy assessment from field and air photographs; map accuracy was determined to be 87%. This Landsat-based map was found to be about 55% in agreement with an AVHRR map product.

Table 6  GAP mapping procedures in development and assessment of a large-area land cover classification map of Maine using nine full and partial Landsat TM scenes, the collection of aerial videography to develop training areas, and ancillary GIS data (after Hepinstall et al., 1999)

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Clouds, cloud shadows and water were masked out of each date using a combination of thresholds and on-screen digitizing</td>
<td></td>
</tr>
<tr>
<td>2. Dark-target subtraction was used to reduce atmospheric effects in each scene; a series of ratios was developed for input to the classifier</td>
<td></td>
</tr>
<tr>
<td>3. Unsupervised clustering was used to develop preliminary cluster statistics then subjected to distance-based (Bhattacharyya) separability analysis</td>
<td></td>
</tr>
<tr>
<td>4. The aerial videography data were randomly sampled to develop training sites for use in the Landsat mosaic; these were subsequently evaluated with the spectral clusters and either retained, deleted, renamed, or merged. This created the ‘hybrid’ structure for the final classification runs</td>
<td></td>
</tr>
<tr>
<td>5. Stratification was used to identify several ‘special’ classes, known to be undifferentiated in the TM data (e.g., abandoned farmfields, certain types of wetlands)</td>
<td></td>
</tr>
<tr>
<td>6. Several postclassification steps were implemented to reduce errors; these steps included additional wetland stratification with National Wetland Inventory (NWI) data, spatial filtering and edge-matching (each TM scene was classified separately, which necessitated this final step)</td>
<td></td>
</tr>
<tr>
<td>7. Accuracy assessment was conducted using the aerial videography; the authors noted that what was actually reported was an ‘agreement between two methods’ rather than accuracy per se. Overall map accuracy was 88% with a kappa coefficient of 0.71</td>
<td></td>
</tr>
</tbody>
</table>
Hyperclustering appears to be employed less in areas with a well-developed classification structure; in such areas, another use of hyperclustering has been to generate initial map data to be used in the development of final (often supervised) map products. Such second-generation large-area maps are often the recipient of more plentiful resources in the form of field data and methodological developments. In California, Franklin, J., et al. (2000) mapped vegetation and land cover over $10 \times 10^6$ ha of Forest Service lands using a hierarchical vegetation classification system with vegetation series-level information nested within broad life-form/land cover classes from a single summer image (high sun angle, no snow) of each National Forest. The procedure was based on the seamless integration of remotely sensed data, advanced image processing techniques, use of collateral spatial data and the compilation of georeferenced field data within a GIS (Table 7). Generally, lifeform class accuracy exceeded 85% in most of the forests tested using these methods. Some remaining problems were noted in open forests and in detailed vegetation types. Cost of the mapping effort, including acquisition, processing and validation, was estimated at US$0.30–0.40 ha$^{-1}$. These costs were considered extremely favourable when compared with the 10– to 100-fold increase in cost of aerial photographic methods or field-based surveys, but were accompanied by a significant reduction in map detail and accuracy.

The land cover map of Great Britain was produced by semi-automated mapping of a 1988–1990 Landsat TM mosaic in which summer and winter imagery were combined to exploit seasonal differences within cover types (Fuller et al., 1998). These summer–winter composites were used to develop a supervised classification training data set that was also processed using knowledge-based context reclassification

<table>
<thead>
<tr>
<th>Table 7</th>
<th>Steps involved in the integration of remotely sensed data, advanced image processing techniques, collateral spatial data, and georeferenced field data within a GIS to provide landcover maps of US National Forests in California (after Franklin et al., 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure</td>
<td>Key references</td>
</tr>
<tr>
<td>Image segmentation to create homogeneous forest stands or stand-like map units</td>
<td>Woodcock and Harward, 1992</td>
</tr>
<tr>
<td>Iterative unsupervised classification of spectral and texture data within these units</td>
<td>Ryherd and Woodcock, 1997</td>
</tr>
<tr>
<td>Vegetation gradient models based on DEM data to predict vegetation labels</td>
<td>Davis and Dozier, 1990; Franklin, J. 1995</td>
</tr>
<tr>
<td>Canopy model and spectral mixture analysis for forest cover assessment and crown size</td>
<td>Woodcock et al., 1997</td>
</tr>
<tr>
<td>Editing on screen manually with reference to aerial photography</td>
<td></td>
</tr>
<tr>
<td>Accuracy assessment using fuzzy sets on a stratified random sample of mapped stands and permanent sample plots</td>
<td>Woodcock and Gopal, 2000</td>
</tr>
</tbody>
</table>
procedures (Groom et al., 1996). Results suggested that the 17 cover classes mapped – which included two forest classes, four grassland classes, shrubs and several types of crop land – were 79–84% accurate in comparison with field surveys, themselves determined to be approximately 90% accurate. The authors noted that information on the state of the countryside across the whole of Great Britain could conceivably be provided by ‘the emergence of operational remote sensing systems and procedures for automated interpretation of imagery’ (Fuller et al., 1998: 102), but then cautioned that ‘enormously complex issues concerning scale and generalization profoundly affect the level and detail of information that can be presented, either as maps or statistical summaries . . . and] . . . there are method-dependent factors which determine the representation of the landscape to a greater or lesser extent’ (Fuller et al., 1998: 122). Isolating these ‘method-dependent factors’ requires the development of a thorough understanding of both the rationale and implications of decision-points used in the process of generating a large-area, Landsat-based, land cover classification map.

This brief assessment of recent Landsat and satellite SAR large-area mosaic land cover mapping applications suggests a minimum number of tasks that must be accomplished satisfactorily for success in any large-area remote sensing land cover classification. The tasks are similar to those used in single-image classifications, but are more complex because of the large area that necessitates acquisition and processing of multiple images, from potentially different sensors and at different times. Typically, this situation also would be characterized by a dearth of field data.

III Methodological issues in large-area, multiple image, land cover classification

The major tasks required in any large-area, multiple image, land cover classification are listed in Table 8 and discussed in the following sections.

1 Hierarchical class structures (map legend)

Differences have existed in perception of land cover and the practices of land mapping as resource management needs have matured (Sauer, 1921; Christian, 1958; Mabbutt, 1968; Bailey et al., 1978; Townshend, 1981). Practically speaking, land cover is almost always used in the sense of the dominant physiographic attribute for a given parcel of land – vegetation, rock, water or soil. Such a statement originates in the climatic and physiognomic classifications of the previous century that were generally broad mapping systems that covered large areas, such as continents, usually with little spatial detail. At larger and larger spatial scales, finer and finer class divisions can be introduced, and these divisions are based on a refinement in the level of physiographic, genetic (e.g., soil types) or ecological detail used (Whittaker, 1975; Zonneveld, 1989; Kimmins, 1997). Divisions within broad physiographic land cover classes such as vegetation require the flexible definition of covertypes.

Land cover and individual vegetation covertypes are usually considered within the context of a classification hierarchy invoked by a conceptual model of vegetation as a geographic phenomenon (gradients or patches, mapped as fields or entities, on the basis of vegetation attributes alone, or vegetation and environmental attributes)
Remote sensing methods

(Franklin, J. and Woodcock, 1997). An example of a hierarchical land cover classification system is the Anderson et al. (1976) Land Use and Land Cover Classification System comprised of four Levels (I, II, III, IV) designed for use with a variety of remotely

Table 8 The major tasks required in any large area, multiple image, land cover classification

<table>
<thead>
<tr>
<th>Task</th>
<th>Key references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection of land cover classification map legend suitable for use with satellite remote sensing data</td>
<td>Anderson et al., 1976; Robinove, 1981; Running et al., 1995; Franklin, J. and Woodcock, 1997</td>
</tr>
<tr>
<td>Acquisition of the image data (different sensors and spatial resolutions, and multitemporal data)</td>
<td>Ahern et al., 1993; Schriever and Congalton, 1995; Lauer et al., 1997; King et al., 1999; Trichon et al., 1999</td>
</tr>
<tr>
<td>Geometric preparation of the image data (geolocation and mosaicing)</td>
<td>Steiner, 1974; Hayes and Cracknell, 1987; Itten and Meyer, 1993; Ehlers, 1997; Cheng et al., 2000</td>
</tr>
<tr>
<td>Radiometric preparation of the image data (calibration, correction, standardization)</td>
<td>Holben and Justice, 1980; Teillet, 1986; Chavez, 1988; Foody, 1988; Richter, 1990; Bayer et al., 1991; Hall et al., 1991b; Sandmeier and Itten, 1997; Dymond and Shepherd, 1999; Ouaidrari and Vermote, 1999</td>
</tr>
<tr>
<td>Selection of the mapping variables</td>
<td>Hutchinson, 1982; Haralick, 1986; Ahearn, 1988; Blaszcynski, 1997; Carr and Pelon de Miranda, 1998; Mickelson et al., 1998; Bruzzone and Fernandez Prieto, 2000;</td>
</tr>
<tr>
<td>Selection of the classification approach</td>
<td>Fleming and Hoffer, 1975; Skidmore, 1989; Lillesand, 1996; Hansen et al., 1996; Cihlar et al., 1998; Beaubien et al., 1999</td>
</tr>
<tr>
<td>Implementation of any preclassification processing steps (e.g., semi-automated training site selection, purification of spectral clusters, etc.)</td>
<td>Chuvieco and Congalton, 1988; Buchheim and Lillesand, 1989; Warren et al., 1990; Bolstad and Lillesand, 1991, 1992; McCaffrey and Franklin, 1993; Bauer et al., 1994</td>
</tr>
<tr>
<td>Selection of the decision rule</td>
<td>Swain and Davis, 1978; Bezdek et al., 1984; Lee et al., 1987; Benediktsson et al., 1990; Peddle, 1995; Adams et al., 1995; Foody, 1996; Friedl et al., 1999</td>
</tr>
<tr>
<td>Validation of process and map accuracy assessment strategy</td>
<td>Kloditz et al., 1998; Scepan et al., 1999; Zhu et al., 2000</td>
</tr>
</tbody>
</table>
sensed data. The system assumes that no one ideal classification of land use and land cover can be developed, but flexible classes and an open-ended structure can be used to accommodate many of the different uses that such classification maps are intended to serve (Talbot and Markon, 1988; Gonzalez-Rebeles et al., 1998; Vogelmann et al., 1998a,b; Raytheon STX, 1999). This classification logic can be applied almost anywhere.

A key implementation characteristic of this approach lies in the use of vegetation structural features (Graetz, 1990), rather than taxonomic or floristic criteria, to distinguish classes at various levels in the hierarchy. Thus, in forested areas, classes of conifer, deciduous and mixedwood stands – distinct by virtue of individual plant and spatially explicit assemblages with structural uniqueness – could be further divided by density or crown closure differences; but species differences would be largely ignored, unless such differences were accompanied by measurable structural distinctiveness. A logical extension of this idea has been provided at the global scale by Running et al. (1995). They suggested a classification system, based on classes distinguishable in the coarsest resolution satellite imagery for which global converage was available (e.g., AVHRR, SPOT VEGETATION or MODIS data). Fundamental vegetation classes differ in permanence of aboveground biomass, leaf longevity and leaf type (Running et al., 1994). The key issue in developing such global or regional lists of classes is to ensure class consistency over the large area.

2 Acquisition of imagery

A significant feature of the large-area Landsat TM coverage problem is the use of imagery acquired at different times of the year or even in different years. Neighbouring images will differ when obtaining multitemporal data over a large area (Hall et al., 1991b). Understory conditions, bud break, chlorophyll absorption rates, water moisture levels and leaf biomass levels were identified by Schriever and Congalton (1995) as the principle influences causing seasonal differences in classification accuracy. In certain circumstances (e.g., areas of persistent cloud, gaps in the mosaic) there will be a need to acquire SAR mosaics, or to augment Landsat TM mosaics with data from different sensors, thus entailing more complex data integration methods (e.g., cloud and cloud shadow patching). The creation of seamless (i.e., sensor invariant) maps is the goal at the classification level.

3 Geometric processing of imagery

Spatial or locational accuracy is required to satisfy mapping needs for large areas (Hayes and Cracknell, 1987; Burkholder, 1999). The basic need is to tie the image coordinate system to earth coordinates through a map projection. Image geometric distortions are related not only to the sensor and imaging geometry, but also to the topography (Itten and Meyer, 1993); corrections, then, are applied to account for known geometric distortions based on the topography and sensor/platform characteristics. Particularly in mountainous terrain, image points may be shifted because of scan line perspective displacement, a random characteristic of the orbital parameters and the terrain. This effect is not normally dealt with during polynomial transformations, even if higher order polynomials are defined (Cheng et al., 2000). Users concerned with the
relief displacement and geometric distortions caused by topographic shifting of pixels must consider orthorectification procedures. The ready availability of high quality DEMs – or the ability to derive these DEMs directly from stereocorrelated digital imagery (e.g., Chen and Rau, 1993) – has provided a foundation for the orthorectification of digital imagery.

In any geometric processing of imagery, a decision must be made on the type of resampling algorithm to use; little has been reported in the literature to guide users in this choice (Hyde and Vesper, 1983). A general preference for the nearest neighbour resampling algorithm exists. This algorithm is thought to minimize the radiometric modification to the original image data that are introduced by area (mean) operators, such as the cubic convolution or bilinear interpolation algorithms that use small windows in the vicinity of the pixel location to determine the actual pixel values. However, even nearest neighbour resampled data can differ from original imagery, since some pixels and scan lines may be duplicated and individual pixels can be skipped, depending on the resolution of the output grid.

Geocorrection could be based directly on a comparison of image detail rather than relying solely on ground control points (GCPs) (Steiner, 1974); such a procedure would be based on feature or area comparisons (Dai and Khorram, 1999). This type of registration implies that distinct entities, such as roads and drainage networks can be automatically extracted and used to match imagery over time. An area-based technique works on the correlation of image data within small windows of image data. Correlation is effective with multiple images from the same sensor with only small geometric misalignment (Shljen, 1979). The processing software is not widely available (Fogel and Tinney, 1996), but manual identification of GCPs will likely soon be augmented by automated methods of georeferencing (Ehlers, 1997).

4 Radiometric processing of imagery

The accurate and quantitative interpretation of remotely sensed data, over large areas or in single image applications, requires that digital images be corrected radiometrically prior to analysis (Robinove, 1982; Teillet, 1997; Teillet et al., 1997). Some sensor-induced distortions, including variations in the sensor point-spread response function, cannot be removed without complete recalibration of the sensor. Some environmentally based distortions, such as variations in atmospheric transmittance across a scene, or over time, during the acquisition of imagery, cannot be removed easily (Duggin, 1985; Woodham and Lee, 1985). Often, in a single-image classification exercise, it is likely that such effects are small relative to the first-order differences caused by the atmosphere and topographic effects.

Image processing systems often contain algorithms designed to remove or reduce influences attributable to atmospheric, topographic and view-angle illumination differences. Some available systems deal simultaneously with geometric, topographic and atmospheric corrections (Itten and Meyer, 1993). Typically, the idea is to develop corrections to remove first-order sensor-based (e.g., view angle variations) and environmental-based (e.g., illumination differences owing to topographic effects, atmospheric absorption and scattering) errors. The goal is to acquire image data over different seasons and years, but to radiometrically process these data such that
differences not resulting from actual land cover changes are removed; hopefully, satisfactory image classification or estimation of biophysical variables (such as LAI) can be conducted independently of the sensor or conditions (Crist, 1985; Gallo and Daughtry, 1987; Cohen et al., 1998).

The simplest atmospheric correction is to relate image information to pseudo-invariant reflectors, such as lakes or asphalt/rooftops (Campbell and Ran, 1993; Teillet and Fedosejevs, 1995). Known as the ‘dark-object subtraction’ procedure, the analyst checks the visible band radiances over the lakes or other dark objects, and then correspondingly adjusts the observed values to more closely match the expected reflectance (which would be very low, close to zero). The difference between the observed value and the expected value would be attributed to the atmospheric influences at the time of image acquisitions; the other bands would be adjusted accordingly. This procedure removes only the additive component of the effect of the atmosphere. An enhancement of this approach uses measurements over lakes with radiative transfer models to correct for both path radiance and atmospheric attenuation, by deriving the optical depth internally (Teillet and Fedosejevs, 1995). A rare alternative to such scene-based corrections relies on ancillary data, such as measurements from incident light sensors and field-deployed calibration targets.

A relatively simple model-based atmospheric correction can be used in which the analyst selects a representative standard atmosphere and estimates reflectances for common targets (Richter, 1990). Model-based atmospheric corrections are now much more commonly available in commercial image processing systems (Franklin, S.E. and Giles, 1995; Richter, 1997). However, it is important to be aware of the assumptions in the standard atmosphere model. Thin (invisible) clouds, smoke or haze, for example, will confound the algorithm because these atmospheric influences are not modelled in the standard atmosphere approach.

The topographic effect is defined as the variation in radiance from inclined surfaces, compared with radiance from a horizontal surface, as a function of the orientation surface relative to the light source and sensor position (Holben and Justice, 1980). In relatively flat terrain, no topographic corrections may be needed (Allen, 2000). However, the physics involved in radiant transfers in mountainous areas, where topographic effects can be significant and must be acknowledged or explicitly accounted for, are incompletely understood (Smith et al., 1980; Kimes and Kirchner, 1981). The complexity of atmospheric and topographic effects is increased by the non-Lambertian reflectance behaviour of many surfaces depending on the view and illumination geometry (Burgess et al., 1995). Generally, flat surfaces are assumed to be equally bright from all viewing directions. But since vegetated surfaces are ‘rough’ it is clear that there will be strong directional reflectances; forests, even on flat ground, are ‘brighter’ when viewed from certain positions. When slopes are introduced, the effects can be enormously complex. Unfortunately, while the various geodetic factors are all inter-related to some extent (Teillet, 1986), it is clear that the effects of topography and bi-directional reflectance properties of vegetation cover are inextricably linked (Hugli and Frei, 1983) and the Lambertian assumption is still widely used (Sandmeier and Itten, 1997).

Although there have been attempts to provide internally referenced topographic corrections (i.e., relying solely on the image data to separate topographically induced variations from target spectral differences) (Eliason et al., 1981; Pouch and Compagna,
Remote sensing methods

190

190, most corrections use a digital elevation model to calculate the illumination difference between sloped and flat surfaces (Civco, 1989; Colby, 1991). These techniques typically assume that the illumination effects depend mainly on the solar incident angle cosine of each pixel (i.e., angle between the surface normal and the solar beam) (Leprieur et al., 1988); but this assumption is not valid for all cover types, and not just because of the non-Lambertian nature of most forested surfaces. In particular, forests contain trees which are geotropic (Gu and Gillespie, 1998). In forests, the main illumination difference between trees growing on slopes and on flat surfaces is in the amount of sunlit tree crown and shadows that are visible to the sensor, rather than the differences in illumination predicted by the underlying slopes.

One of the more powerful methods to deal with the topographic effect has been to use the DEM data together with the spectral data in the analysis (Frank, 1988); for example, Carlotto (1998: 905) built a ‘multispectral shape classifier instead of correcting for terrain and atmospheric effects’. The shape classifier relied on the fact that the differences in reflectance in different bands would be similar regardless of the terrain conditions. In considering only the shape of the reflectance curve in known terrain types, the interband variations resulting from nonspectral/target interactions are assumed to be insignificant within classes (Franklin, S.E., 1991; Franklin, S.E. and Wilson, 1992). This idea is examined more fully in the following section.

New multiangle sensors (e.g., ASTR, POLDER, MISR) have been deployed with the express purpose of characterizing land cover Bidirectional Reflectance Distribution Function (BRDF) (Diner et al., 1999; Sandmeier and Deering, 1999; Gemmell and McDonald, 2000). The variations because of BRDF are sometimes apparent in wide-angle or pointable satellite systems, such as the SPOT, AVHRR, SPOT VEGETATION or EOS MODIS sensors (Donoghue, 1999; Running et al., 2000). The view angle effect can ‘mask’ or hinder the extraction of information as is often the case with single-pass wide-view-angle airborne data (Kriebel, 1978; Irons et al., 1991). Wide-angle and off-nadir views will introduce variable atmospheric path lengths in an image scene, thereby introducing different atmospheric thicknesses that need to be corrected during the atmospheric processing. The view angle will also determine the projected area of each pixel and introduce a more complex geometric correction (Barnsley, 1984; Barnsley and Kay, 1990). However, such multiangle variations in SPOT and Landsat imagery had no significant effect on land cover classification accuracy in earlier tests (Muller, 1993) and therefore are thought to be safely neglected (Gerstl and Simmer, 1986; Guyot et al., 1989). This situation may change if the predominately narrow viewing angles used in current satellite remote sensing were to be altered or if greater use is made of ‘overlap’ areas and other wide-angle viewing sensors.

In the microwave portion of the spectrum, radiometric corrections are needed to derive backscatter coefficients from slant-range power density, typically provided through calibration target deployment. By far the strongest georadiometric effects on SAR imagery are caused by azimuth (flight orientation) and incidence angles (defined as the angle between the radar beam and the local surface normal (Domik et al., 1988; Waring et al., 1995). The influence of local topography can be dramatic, as high elevations are displaced toward the sensor, and the backscattering on slopes is either brightened or foreshortened. Simple image corrections using DEM-derived slopes and aspects do not completely restore the thematic information content of the imagery. The wavelength-dependent energy interactions are too complex to be well-represented by
simple cosine models (van Zyl, 1993); however, cosine-corrected imagery will likely be more useful (Wu, 1990).

5 Selection of mapping variables (spectral, ancillary)

Issues that must be considered when selecting mapping variables relate to any tradeoffs that can be identified when using a subset of bands compared with the additional effort and cost in using all available spectral data, transformations of the data (including band-to-band manipulations and spatial variables such as texture) and the role of supplemental data (e.g., from other satellites). Generally, there are few guidelines that exist for the potential differences in land cover classification accuracy, over large areas and with the types of environments that will be characterized in any large-area mapping effort, that might result in using the full complement of Landsat bands compared with a reduced, statistically manipulated, data set (e.g., Tasseled Cap Transformations or vegetation indices). There is compelling evidence that context (Gurney, 1981), texture (He and Wang, 1992) and DEM (Hutchinson, 1982) information are critical data resources that can be used to increase land cover classification accuracy. The use of geospatial data with imagery has increased considerably with greater understanding of their characteristics and increased availability across wide regions (White et al., 1995).

In a Wisconsin forest landscape, Bolstad and Lillesand (1992) reported that a rule-based approach using Landsat TM data, soil texture information, terrain position and soil–plant relationships could separate 13 land cover classes corresponding to Anderson Levels II and III with 89% accuracy, a 16% improvement over a standard spectral classification method. Boresjö Bronge (1999) provided a recent example of an integrated TM/topographic map approach in Sweden. Using Landsat TM imagery and a series of masks derived from topographic maps, spectral confusion in four conifer, three deciduous, eight mire and eight other classes was eliminated in successive stages based on topography, land-use and other salient features expressed on the base map.

Another similar strategy has been to employ existing GIS-based forest cover data in some fashion during a classification procedure (Goodenough, 1988; Chalifoux et al., 1998). Clearly, the existence of a previous satellite remote sensing classification could be an excellent source of information in any new classification.

Increased classification accuracy can be obtained if it is possible to avoid relying exclusively on the per-pixel spectral response pattern and instead employ different image data. The two most obvious types of image data that are underutilized in current classification procedures are image context and image texture. Context relies on a summary description of the relationships among pixels or, more frequently, among classes. The premise is that a pixel’s most probable classification, when viewed in isolation, may change when viewed in some context (Haralick and Joo, 1986; Khazenie and Crawford, 1990; Binaghi et al., 1997; Chen, 1999; Solberg, 1999). Image texture has a long history in land cover classifications (e.g., Weszka et al., 1976); texture is a quantification of the spatial variation of image tone values that defies precise definition because of its perceptual character (Hay et al., 1996). Insight into how texture might be analysed by computer has focused on the structural and statistical properties of textures (Haralick et al., 1973; Haralick, 1986); subsequently, much effort has been expended on optimizing satellite image texture measures for the land cover mapping application (Shih and
6 Classification approach

An ideal classification approach does not yet exist (Cihlar et al., 1998); all classifiers at some point must handle a three-way compromise between the information classes that are desired, the spectral information content of the imagery and the method of making class decisions. The information classes and the spectral classes are never in complete agreement, and every decision rule yet devised will make less than optimal choices in the presence of real noise or various levels of ambiguity (Swain and Davis, 1978). These problems have led many to consider the classification process – not just the decision rules, but the entire set of procedures that comprise the classification protocol – as a special circumstance of fuzzy logic. Fuzzy classification methods have recently attracted a great deal of attention because of their ability to ‘soften’ the decision for each pixel; instead of only a single class assignment, each pixel would belong to every class but with different degrees of membership (Wang, 1990; Foody, 1999).

Typically, a classification can be considered as principally a supervised, unsupervised or modified approach (Fleming and Hoffer, 1975). Supervised classification is not typically the process of choice in large-area land cover classifications because of the enormous requirement for training data (Bauer et al., 1994); ‘thus, where spatial distribution information is not available, e.g., when mapping a large area previously not well known, unsupervised classification is arguably the better strategy’ (Cihlar, 2000: 1102). This situation has not prevented significant efforts based on supervised classification methods from proceeding (e.g., Muchoney et al., 2000). One option has been to develop semi-automated or unsupervised methods of finding training areas or homogeneous clusters for input to supervised classification (e.g., Buchheim and Lillesand, 1989; Bolstad and Lillesand, 1991; McCaffrey and Franklin, 1993); another has been to use fine resolution data to develop training data for use in the larger area, coarser resolution information (Friedl et al., 2000), and to concentrate on those observations that are more difficult to classify – so-called ‘boosting’ classification results (Friedl et al., 1999); still another has been to rely on the development of classifier decision trees (Hansen et al., 1996; Hansen and Reed, 2000) and neural networks (Carpenter et al., 1999; Friedl et al., 2000). The human mind is perhaps the finest available tool for synthesis and analysis of image patterns (Buiten, 1993); an appropriate strategy is to utilize, as much as is possible, the expertise of the interpreter (Beaubien, 1994). Modified supervised methods attempt to accomplish this goal (Beaubien et al., 1999).

Unsupervised classification usually refers to a classification in which the classes are not predetermined; the statistical or clustering properties of the image data are used to find a set of ‘natural’ classes (Thomas et al., 1987). This has led to widespread use of unsupervised classification procedures in large-area land cover mapping projects in which the classes are reasonably straightforward. In regional or small area studies, spectral classes generated in an unsupervised way can be labelled in a systematic process. For example, Debinski et al. (1999: 3284) generated 50 ISODATA clusters from a single Landsat TM image to discriminate the gross land cover types (forest versus meadow) and among a gradient of meadow types (xeric to hydric). There were five
forest classes mapped in their Greater Yellowstone Ecosystem study area; two pure conifer and three mixed conifer classes based on density (sparse, medium and dense stands). The purpose of the initial clustering was to provide remote sensing input to habitat class characterization that was used to structure a biodiversity sampling scheme. Clustering before collecting training data helped avoid a later step in which classes were spectrally combined because they are not separable with the necessary degree of accuracy.

A layered classification approach (Jensen, 1978) or decision tree approach (Hansen et al., 1996), can outperform a classification that uses the same decision rule for all data and all classes. Using different classifiers at different points in the classification process holds the promise of achieving maximum classification accuracy through ‘selective decision-making’; in other words, the algorithm is designed to invoke a particular classifier decision rule only for a decision to which it is optimally suited (Wilkinson, 1996). An early example of this approach was provided by Franklin, S.E. and Wilson (1992). They developed a three-stage classifier that used spectral data, DEM data and spectral-DEM data in a mountain environment. In valley reaches, where DEM data were not contributing to discrimination of vegetation classes, the classification was based solely on the spectral response data and a minimum distance to means rule. In more complex terrain, in which different slopes and landforms were of interest, the DEM data were more powerful in separating land cover classes. A maximum likelihood classifier was used in those situations. In other areas, both spectral and DEM data were needed. The layering occurred with the data (i.e., spectral, DEM or combined), the methods (minimum distance or maximum likelihood) or both.

These decision-tree classifiers (Hansen et al., 1996), can provide the same type of increases in accuracy that have been reported in more complex decision rule processes provided the data are not too complex (i.e., they are restricted to one or two data types). The EOS MODIS classifier decision tree, for example, was constructed ‘to estimate classifications based on training data by recursively splitting the data into successively more homogeneous data sets’ by measuring the reduction in entropy of the training data (Friedl et al., 2000: 977). The improvement in accuracy over a straightforward maximum likelihood supervised classification procedure was significant.

7 Selection of classifier decision rule

Clustering methods ‘were developed to detect inherent or natural structure in data’ (Swain and Davis, 1978: 178); such classifiers rely on measures of distance (e.g., Euclidean, Mahalanobis), cluster ‘characteristics’ (e.g., means, divergence) and clustering criteria (e.g., a measure of the quality of assignment of points to clusters). They are widely available and their behaviour under a wide range of conditions can be readily predicted. The most powerful decision rule available in the statistical realm is the maximum likelihood algorithm which relies on the development of class covariance-matrices from training samples (Kershaw and Fuller, 1992). Decision rules for assigning clusters to classes – i.e., the labelling process – tend to be dominated by human analyst rather than computer-based algorithms. In a large-area exercise this step could conceivably be augmented in several ways for specific strata or over the entire mosaic; such labelling tools require some development. Initially the clusters developed
by hyperclustering receive limited ‘labelling’ and are simply documented through differing statistical properties. An improvement in this process would be to consider a wide range of measures from the clusters, and to develop a decision rule between clusters and classes based on access to ancillary data (i.e., data not used to define or ‘detect’ the cluster).

Traditional statistical classifiers used in remote sensing, such as the maximum likelihood algorithm, the discriminant function, the minimum distance to means algorithm, the parallelepiped classifier, all operate in Euclidean space. These classifiers, and their assumptions, are well-described in the pattern recognition, statistical and remote sensing literature. Under many conditions, these common statistical classifiers are robust and well-behaved; they provide optimal or near-optimal decisions on the covertype class based on simple statistics such as the mean, standard deviation, and covariance of spectral response in classes, and a straightforward formulation of the probabilities of class membership. These algorithms have the advantages of ease-of-use and widespread availability; however, statistical classifiers do not always perform well when the data display non-normal distributions (common with DEM and texture data). But much of the new data that may be available to a large-area classification project might not be ratio-level data, but rather data including a mix of ordinal (e.g., ranked soil fertility classes), interval (e.g., dominant species) or nominal (e.g., class 1 of 12) data. A classifier that is nonparametric and thus can use a range of data types, including ratio-level data, would be needed in these situations (Bezdek et al., 1984; Cannon et al., 1986; Lee et al., 1987; Benediktsson et al., 1990; Peddle, 1995; Jensen et al., 1999; Trichon et al., 1999; Hall et al., 2000).

These new decision rules are increasingly available in commercial image processing systems or as add-on packages. Decision are made, not on probabilistic rules, but using different mathematical theory and logic. A few of these classifiers operate in the same way that conventional classification algorithms, such as maximum likelihood, operate. For example, they require the same preparation and iterative steps; only the actual decision rule is different. For others, the entire process of classification must be adjusted to take into account the demands of the classifier. The complexity of the nonstatistical classification decision rule and ultimately of the entire classification process (including interpreting the results) may be formidable. For example, Pinz et al. (1996) described ‘active fusion’, a computer procedure to combine information from multiple sources on the basis of three different mathematical theories: (1) probability theory, (2) the Dempster–Shafer theory of evidence and (3) fuzzy set logic. Experimentally, they showed a significant reduction in the number of information sources required for a reliable decision on classification of Landsat data for agricultural crops. In another example, Desachy et al. (1996) wrote 11 expert system rules for a southern India tropical forest vegetation classification using Landsat TM data and a DEM. The improvements in classification accuracy ranged up to 14% when compared with the average result for supervised maximum likelihood classification, with the final map determined to be 83% in agreement with field observations.
8 Class and map validation

Land cover maps must be validated to be useful (Thomlinson et al., 1999), and derived products must meet claimed specifications (Cihlar et al., 1997b). Validation, which differs from calibration, is ‘the process of assessing by independent means the accuracy of the data products derived from the system outputs’ (Justice et al., 2000: 3383). Initial product validation is the process of establishing the quality of an algorithm by assessing the product generated by the algorithm, and continuing (process) validation is the process of establishing how well the algorithm performs if the area of interest, time and data are changed (e.g., new satellite sensor in a different year in a different forest type) (Cihlar et al., 1997b).

Intercomparisons of data products and model outputs provides an initial indication of product accuracy and uncertainty. Another image classification product, generated using different data, could be used (Moody and Woodcock, 1995; Kloditz et al., 1998). The validation of maps and models based on an AVHRR classification with a higher spatial detail classification derived by Landsat TM was successful – in boreal forests (Fazakas and Nilsson, 1996), in tropical areas (Mayaux and Lambin, 1997) and now globally through the use of image interpretation keys (Kelly et al., 1999). This is the principle underlying the validation of the IGBP DISCover 1.0 product; ‘a stratified random sample design and a methodology that relies on testing the DISCover thematic classes against an independent data source, in this instance, higher spatial resolution satellite imagery’ (Scepan, 1999: 1051).

Maps must be validated using a comparison to some independent assessment – preferably comprised of ground validation data collected expressly for such purposes (Muller et al., 1998). Such data are – almost by definition – rare and expensive. Independent test areas, not used in development of the maps (e.g., through training the classifier), should be examined to report on classification accuracy, usually in the form of a contingency table or confusion matrix (Congalton and Green, 1999). A convenient way of considering the typical land cover accuracy assessment involves examination of the response design (how reference data are collected), the sampling design (choosing a sampling plan that is appropriate for the study) and the analysis and estimation (calculations of accuracy and estimating standard errors) (Stehman and Czaplewski, 1998). In large-area classifications from spatially coarse resolution data, the response design may be difficult to execute in the field; how does one observe classes over many has corresponding to individual 1 km pixels?

For large-area land cover classifications by remote sensing, an operational, standard, validation protocol does not yet exist (Justice et al., 2000). Outstanding issues include the design of a statistically valid and logistically feasible field sampling strategy, the assessment of the accuracy of reference (including field) data, establishing and presenting accuracy metrics that are stable and informative, and dealing with the host of registration and correlation problems that arise when using coarse and high spatial resolution satellite imagery together and with other geospatial data sets. By far the most frequent method of validating land cover classifications has been through aerial photointerpretation. Zhu et al. (2000) described the validation of the USGS Landsat TM conterminous land cover map in the New York and New Jersey region. Of the 23 land cover classes in the land cover map legend, 15 were found in the region and seven of these were ‘rare’ classes. A two-stage cluster sampling design was implemented with consid-
eration of known inclusion probabilities, good spatial distribution and representation of all classes, and cost. Only the data acquired in the National High Altitude Aerial Photography (NHAP) program had the required response design characteristic of full area coverage. Sample sites were located visually on the NHAP photography, and the photointerpreted and digital class estimates of land cover class assembled in contingency tables. Producer’s and user’s accuracies and associated standard errors were generated overall and using various stratification formulae.

IV Summary

Recently, several large-area land cover mapping applications based on Landsat TM and free-flying satellite SAR mosaics have been completed with success in terms of map accuracy and utility. Many more such mapping efforts are underway or are to be initiated in the near future. Radiometric processing of multiple scenes and multiple sensor inputs are a continuing problem. Corrections for complex atmospheric effects, view-angle effects, topographic effects and multiple sensor differences cannot yet be handled through radiative transfer models. None of these corrections are operational using empirical methods or models, and while often simple to implement, are thus rarely recommended for widespread use. Geoprocessing of large-area image mosaics continues to rely heavily on manual identification of GCPs and relatively simple resampling methods. It is thought that such techniques will likely be augmented in the near future with greater automation and feature- or area-based geocorrections, but the lack of these tools are not considered major impediments to large-area classifications with multiple scenes.

A more significant issue is the fact that an ideal classification approach to the large-area, multiple image, land cover mapping application has not yet been developed. Supervised and unsupervised classifications each contain significant weaknesses when applied to the enormous accumulations of diverse data typically assembled in mapping efforts over large areas. Modified classification approaches, implemented iteratively within submosaics of large areas, perhaps stratified by ecoregion and data characteristics, can be designed to address the weaknesses of the supervised and unsupervised methods without compromising the ability of the system to deliver repeatable and stable results. Decision-tree classifiers are one approach to utilize ancillary data, such as DEMs and texture, in classes and areas for which they are ideally suited to contribute to correct decisions. Statistical techniques continue to dominate the classification problem. Alternative classification algorithms remain difficult to access and use; in future, it is expected that fuzzy classifiers, rule-based and neural network algorithms, and nonstatistical formulations will assume a more prominent role in the development and implementation of the large-area land cover classification protocol.

Acknowledgements

The authors would like to acknowledge the financial support of the Earth Observation for Sustainable Development of Forests (EOSD) program of the Canadian Space Plan of the Canadian Space Agency and the Canadian Forest Service.
References


Buchheim, M.P. and Lillesand, T.M. 1989: Semi-automated training field extraction and


imagery. *Photogrammetric Engineering and Remote Sensing* 64, 293–300.


**Fleming, M.D. and Hoffer, R.M.** 1975: Computer-aided analysis of Landsat-1 MSS data: a comparison of three approaches, including a ‘modified clustering’ approach. Laboratory for Applications of Remote Sensing Information Note 072475. West Lafayette IN: Purdue University.


Gu, D. and Gillespie, A. 1998: Topographic nor-


**Jensen, J.R., Qiu, F. and Ji, M.** 1999: Predictive


