International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Patterns and trends in land-use land-cover change research explored using self-organizing map

Yi-Chen Wang \textsuperscript{a} & Chen-Chieh Feng \textsuperscript{a}
\textsuperscript{a} Department of Geography, National University of Singapore, Singapore, 117570

Available online: 28 Jun 2011

To cite this article: Yi-Chen Wang & Chen-Chieh Feng (2011): Patterns and trends in land-use land-cover change research explored using self-organizing map, International Journal of Remote Sensing, 32:13, 3765-3790

To link to this article: http://dx.doi.org/10.1080/01431161.2010.540590

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Research on land change has a long history, has generated numerous publications and continues to receive international research attention. To facilitate the understanding of the patterns and trends of land-change research, this article uses a content-based text-retrieval approach and self-organizing map to analyse more than 700 peer-reviewed remote-sensing and natural-science papers on land-use/cover change (LUCC) from the past two decades. We present the results in map-like displays and discuss papers within the identified clusters to examine the research activities. A new cluster of research, which has emerged in the last 5 years of analysis, has focused on mixed-pixel issues for land-use/cover mapping, particularly in the context of forest catchments. Studies of LUCC consequences after 2000 have been concerned with the effects of forest conversion on soil-nutrient pools and nitrate cycling. Incorporating information on resolutions and extents into the representations reveals a dominant scale of analysis for some research activities. Analysing time frames of examination in the papers suggests that research on long-term LUCC consequences started to use presettlement land survey records. Few attempts, however, have been made to investigate the uncertainties in the historical sources of information for LUCC research, thereby presenting a future research topic.

1. Introduction

Several decades of research have revealed that land-use/cover change (LUCC) is one of the most essential interacting components of global change affecting the Earth’s system (Foley et al. 2005). The importance of monitoring LUCC has been recognized by scientists and practitioners, and the capacity to monitor this has been greatly facilitated by the number of and improvement in remote-sensing sensors over the past two decades (Gutman et al. 2004). Burgeoning research programs and projects on utilizing datasets derived from Earth observation have generated diverse publications, including, but not limited to, monitoring LUCC (Townshend and Justice 2002), simulating its change (Parker et al. 2003), linking social with physical processes (Walsh and Crews-Meyer 2002) and investigating the consequences of LUCC (DeFries et al. 2002). Keeping track of the latest scientific advances of this subject is, however, challenging because the increasing volume of the literature and the interdisciplinary nature of LUCC studies could easily obscure the patterns, trends and relationships in research publications. A visualization approach that can reveal the pattern and structure of the established LUCC research is therefore desirable.
Research results are primarily disseminated using text written in natural language. Text analysis of research documents hence allows the recognition of structures and relationships in and among the text sources. Exploratory data-analysis techniques, such as self-organizing map (SOM) and information visualization tools, have been useful in various knowledge-discovery and document-visualization tasks (Lin 1992, Kohonen 2001, Wang and Feng 2008). Skupin (2004) employed SOM with cartographic techniques to successfully create map-like knowledge-domain visualizations for the geography discipline based on the abstracts of the 1999 Annual Meeting of the Association of American Geographers. Koua et al. (2006) found the effectiveness of SOM-based representations in exploring the correlations of the attributes of the dataset.

Using SOM, this article analyses peer-reviewed journal papers related to LUCC and its environmental consequences, and portrays the subject in a visual form to help further our understanding of the LUCC research activities. We focus on LUCC and its environmental consequences because these issues have been the major research challenges for environmental sciences (NRC 2001), and remain of concern to the emergent land-change research community (Turner et al. 2007). We recognize that LUCC research, in its most comprehensive form, is interdisciplinary, linking human, environmental and geographical information remote-sensing sciences (Turner et al. 2007). However, a complete description of all the materials concerning LUCC would almost be impossible because, just in 2005, there were more than 2000 papers on the Web of Science with ‘land use’ as either a keyword or in the abstract (Aspinall 2006). Instead, this article serves as a beginning to employ information-visualization techniques to explore the patterns and relationships of research publications of LUCC and its environmental consequences from selected journals in remote sensing and natural sciences over the last two decades. Specifically, we address four questions. First, what are the patterns of research on LUCC in individual domains? Do publications from remote-sensing and natural-science journals form their own clusters? Second, have the research activities changed over time? Third, are LUCC studies scale dependent in terms of data resolution employed and spatial extent examined? Fourth, do different research topics consider different time frames?

2. Methods

2.1 Data preprocessing

The data for analysis were research publications relevant to LUCC and its environmental consequences. We used peer-reviewed articles published between 1987 and 2007 from seven journals, including International Journal of Remote Sensing, Remote Sensing of Environment, Landscape Ecology, Ecosystems, Journal of Biogeography, Hydrological Processes and Journal of Land Use Science, for three reasons. First, these journals represent some of the main publication outlets for studies on the techniques commonly used to monitor and quantify LUCC, for which research investment is still needed, particularly for land-change research (Turner et al. 2007), and on the consequences of LUCC in ecosystem and hydrology, which are of concern to global-change research (NRC 2001, GLP 2005). Second, LUCC research requires interdisciplinary effort. Information visualizations can ease collaboration in interdisciplinary research settings (Skupin 2002). Thus, instead of only analysing papers from remote-sensing journals, visualization of papers from different subjects allows remote-sensing scientists to see their own work in the broader LUCC context, thereby enabling research synergies. Third, the time period between 1987 and 2007 is about one decade
before and after the joint LUCC project by the International Geosphere-Biosphere Programme (IGBP) and the International Human Dimensions Programme (IHDP) started in the late 1990s (i.e. IGBP-IHDP 1995, 1999). The two decades of analysis may reveal changes in research trends. Additionally, we had access to the electronic versions of most papers in these journals in Adobe Portable Document Format (PDF) for this time period, facilitating computational text analysis for information visualization. Although *Ecosystems* and *Journal of Land Use Science* are relatively new, we included them to investigate if recently established journals would contribute essentially to certain types of research activities.

We searched articles with ‘land use’ or ‘land cover’ in their titles or keywords. We did not use ‘land change’ or ‘land system change’ in the search because LUCC did not give way to these two terms until recently (Turner *et al.* 2007). For each identified paper, we extracted the title, author information, abstract and introduction from its original PDF file into a plain text file. Unlike prior work, which used abstracts (Skupin 2002, 2004), we included both the abstract and the introduction because the use of only abstracts in our preliminary analysis did not reveal much variation in the patterns and structures of the text documents. Since an introduction of a paper provides background on the topic and purposes of the research, we argue that the inclusion of the introduction section allows much more differentiation of research activities.

### 2.2 Content-based text retrieval

Computational text analysis retrieves and transforms text documents (i.e. journal papers here) from their unstructured, plain form into a structured form amenable to computation, comparison of document-to-document contents and visualization of structures and relationships in and among literature. Instead of using a citation-based approach, we followed Skupin (2004) to use a content-based text-retrieval approach because the content of a document offers richer descriptions of the document than citations. The content-based approach incorporated the vector-space model and the SOM method, which analysed document relationships based on shared contents.

#### 2.2.1 Vector-space modelling

In the vector-space model, each paper was represented as a vector with dimensions corresponding to index terms and numerical weights expressing the importance of each index term in the document (Salton and McGill 1983). The index terms that we used were author-chosen keywords in the journal papers so as to capture the main themes of the text documents. The use of author-chosen keywords also ensured an automated creation of an index, without prior knowledge of the particular domain, while restricting the keyword set to relatively meaningful terms (Skupin 2002). For the 722 papers that we searched, we obtained a total of 876 keywords, representing different LUCC research topics. This set of index terms already excluded ‘land use’, ‘land cover’ and ‘land use/cover change’ because we used them in the search. For papers that did not list keywords, we carefully checked their contents to ensure that the main themes of these papers were captured by the terms in the compiled index-term pool.

The numerical weights, which measured the importance of an individual term in the papers, were the frequencies of occurrence for the term. The higher the frequency of an index term in a paper, the more relevant the paper was to that term. Papers with similar sets of index terms and frequencies should be similar in content. We wrote a program in Perl to calculate the frequencies of the index terms in the abstract and introduction of each paper formatted in plain text (figure 1). The extraction of the
frequencies of occurrence was straightforward for terms without synonyms in our index-term pool. We further combined the frequencies of index terms for synonyms. For example, common and Latin species names, such as Pine and *Pinus* spp., were combined because they referred to the same tree genus. We also excluded terms with zero frequencies in all the analysed texts from further analysis, such as chemical mixing, spring flow and mixed mesophytic. This resulted in 832 final index terms.

### 2.2.2 The SOM method.

The SOM method is one of the most widely used forms of artificial neural networks. It uses an unsupervised learning strategy where the similarity relationships between the data and the clusters are used to categorize the data for knowledge discovery, as opposed to a supervised method that learns to associate a set of inputs with a set of outputs using a training dataset, for which both input and output are known. We used Kohonen’s vector-quantization algorithm, which takes a set of *N*-dimensional observations as input and maps the high-dimensional data onto a low-dimensional space. The output layer consists of a network of neurons or nodes in a certain manner, typically a two-dimensional (2D) lattice (Kohonen 2001).

To apply the SOM method, we first developed a document-term matrix. Each row of the matrix was a document vector that contained 832 integers, representing the frequencies of the 832 index terms in the paper. We used the document-term matrix as the input to train a SOM that consisted of 384 neurons because 24 rows by 16 columns produced the results in a reasonable computing time and allowed the differentiation of patterns of research activities. Next, we carried out the training with SOM_PAK Version 3.1 (Kohonen *et al.* 1996), resulting in a vector of 832 elements for each of the 384 neurons. Last, the document vector of each paper was compared to all 384 neuron
vectors in order to identify the similarity of the document vector to each of the 384 neuron vectors. Once the most similar neuron vector was identified, the paper associated with the document vector is assigned to the location of the most similar neuron. This step was applied to all 722 papers to generate the final SOM (figure 1).

2.3 Knowledge-domain visualization

The 2D neuron lattice of SOM allows map-like displays using geographic information systems (GIS) and cartographic techniques, as geographic phenomena are often represented using maps, mostly 2D planar representations. The SOM method assigned individual documents to the most similar neuron by comparing document vectors to neuron vectors. A single neuron may thus be associated with multiple documents. Following the visualization approach of Skupin (2004), we derived the 2D coordinates for each document and represented the documents, which are the 722 LUCC papers, as points in GIS. This enabled direct access to the papers and the attributes associated with them, such as year of publication and spatial extent of analysis, for later exploration of patterns in LUCC research. We then used Thiessen Polygons in ArcGIS (ESRI, Redland, CA, USA) to create a geometric based map of 722 polygons, each representing an analysed paper (figure 1).

Portraying LUCC research activities in a visual form involved two main steps: clustering and labelling. First, to facilitate the understanding of the range of LUCC research activities, instead of showing all the 722 polygons, we merged neighbouring document polygons if they were part of a statistically determined cluster. The purpose of clustering was not to provide the ‘best and true’ partition, but to serve as a stepping-stone in the support of visual exploration towards understanding the knowledge domain (Skupin 2004). We used hierarchical clustering because it presented multiple levels of clusters simultaneously; clusters at one level were joined as clusters at the next higher level, thereby revealing different levels of research activities. We constructed a dendrogram to indicate the clustering of papers at different levels and determined the similarity using the Pearson correlation coefficient. We tested different similarity levels to explore the patterns and eventually employed the similarity levels of 80% and 90%. A 90% cutoff denoted that papers within a cluster were 90% similar in content based on the frequencies of the index terms.

Second, knowledge-domain visualization involved careful labelling because it would be too complex to label all the index terms for all the derived clusters of different similarities on a map. To illustrate the contents of each cluster, we derived labels by extracting the three index terms of the highest frequencies in descending order for labelling clusters at the 80% similarity. Because the 90% clusters were nested within the 80% clusters, the three terms of the highest frequencies of the 80% clusters would likely appear in the terms of high frequencies for the 90% clusters. Therefore, for labelling at the 90% similarity, we used six index terms of the highest frequencies so as to provide more information on the research activities at the 90% level.

2.4 Exploring patterns of LUCC research

We conducted the following analyses using clusters at the 90% similarity to help further understand LUCC research activities. To investigate if papers from similar types of journals would occupy a cluster, we labelled papers as points based on journal types, and overlaid them with the polygons of the 90% clusters. To comprehend whether certain research activities dominated a specific time period, we divided the
papers into four time periods, approximately 5 years as an interval, based on their year of publication, and represented the trend of change on a map. To examine if LUCC studies were scale related, we extracted information on spatial extent and resolution by going through each of the 722 papers. We excluded studies that only mentioned place names of their sites or types of data used without explicitly providing scale-related information in numbers. Spatial extent denoted the size of the study area. When a study consisted of multiple sites of various sizes, we used the average. Resolution here referred to the size of a pixel or grid cell. For studies with multi-resolution datasets, we recorded the finest resolution or the most dominant one if specified. To understand whether different LUCC research activities were concerned with different time frames, we recorded the start and end dates of the imageries or maps analysed or of the in situ data collected to derive the duration of consideration. We then discussed the contents of the papers in clusters that revealed certain patterns or trends. We further reviewed LUCC papers published after 2007 to investigate if the trend of research activities identified by the SOM method remained unchanged.

3. Results and discussion

3.1 Clusters of research activities from the SOM method

Using similarity cutoffs at the 80% and 90% levels, we obtained a total of 8 and 30 clusters, respectively, illustrating research activities of remote sensing and natural science from 1987 to 2007 on LUCC and its environmental consequences (figure 2(a)). Because ‘data’ and ‘area’ are among the highest frequency terms in the derived 80% clusters, to allow a clearer distinction of research activities in the map, we excluded these two terms from labelling and used the next three terms of the highest frequencies. For easy referencing in this article, we labelled each cluster at the 90% similarity with numerical numbers (figure 2(b)), so cluster 1.1 corresponded to the ‘spatial-resolution-pixel’ cluster located at the top-left corner of the map (figure 2(a)). All the 1.X clusters (X = 1, 2, …) were within the 80% cluster labelled as ‘forest-model-catchment’, and we referred to this 80% cluster as cluster 1.X. Note that these numerical labels are unique identifiers, rather than implying that cluster 2.X has newer or more important research topics than cluster 1.X. Clusters nearer in distance are more similar in content than those farther apart; thus, clusters adjacent to each other are, in general, of similar labels. The map portrays the content of LUCC research activities approximately in three parts: forests on the left, soils on the middle and images on the right (figure 2). Research activities distributed from the middle to the right of the map are relevant to image and classification, such as 6.X ‘landscape-classification-image’ and 7.X ‘image-classification-spatial’.

Each journal has its aims and foci of research areas. Publications from a specific type of journal may therefore form a cluster on the map. Nevertheless, the publications of the seven analysed journals, which were indicated as different point symbols (figure 2(a)) dispersed among each other. Papers from relatively new journals, Ecosystems and Journal of Land Use Science, did not just occur in one single cluster. At the coarse 80% level, there was no clear aggregation for remote-sensing and natural-science types of journals; at the fine 90% level, most of the clusters still consisted of papers from these two fields, showing LUCC research collaboration from different fields. The exception was cluster 4.1, with an aggregation of remote-sensing papers. The lack of papers from natural-science journals suggested that its research activities probably focused on methodologies or techniques for land-use/cover observation or classification. As
indicated by its labelling ‘validation-video-resolution’, research within this cluster used airborne digital video to collect validation data for independent accuracy assessment over large areas in British Columbia, Canada (Wulder et al. 2007). Studies related to resolution in this cluster consisted of two types. The first type focused on applying an artificial neural network to improve the classification accuracy when integrating multi-resolution or multi-type data, such as textural and terrain information, particularly in areas with cloud-cover issues or limited ground information (Bagan et al. 2005, Peng et al. 2005). The second type of studies developed techniques to extract land-cover information from coarse-resolution data (Cihlar et al. 1996,
It is interesting that the technical content of cluster 4.1 did not result in a position at the right portion of the map where clusters were associated with images and classifications (figure 2(a)). Instead, the SOM training placed cluster 4.1 at the top-left of the map near to forest-relevant clusters. Close examination of the papers indicated that these studies were indeed forest relevant because the land-cover type of their interests was mostly forests (Cihlar et al. 1996, Grandell et al. 1998, Wulder et al. 2007).

3.2 Changes in research activities over time

The volume of papers on LUCC has been growing (figure 3). To understand whether the research activities have changed over time, we divide the papers into four time periods based on the year of publication and illustrate the trend of change. A steady increase in the number of papers over time is evident for most clusters, with a much higher number of publications in the latest (2003–2007) than in the earliest period (1987–1992) (figure 4). This pattern probably results from the emerging research agendas and activities under the auspices of the joint IGBP-IHDP LUCC project since the 1990s.

3.2.1 Emerging themes. Most research clusters consist of papers from all four time periods, suggesting that these LUCC research activities are still vibrant. The SOM derived map also reveals some possible new topics. The most notable one is cluster 1.1, which includes only papers published in the latest time period (2003–2007) (figure 4). The labelled terms indicate that its research activities are related to ‘spatial-resolution-pixel’ (90% level labelling) in the context of forest catchments (80% level labelling) (figure 2(a)). Spatial resolution is one of the primary factors that determines the level of details of LUCC patterns observable to researchers. Assessing sensitivities of pattern detection and subsequent inferable processes to change in resolution of remotely sensed imageries remains an important research agenda for remote-sensing specialists (Rindfuss et al. 2004, Aplin 2006). Papers in cluster 1.1 underscore this issue by proposing new methods and techniques to cope with the spatial heterogeneity in the resolution of analysis in
Luco research. Atkinson and Aplin (2004) used geostatistics, and Kasetkasem et al. (2005) employed Markov random-field models to represent the spatial dependence within and between pixels for super-resolution land-cover mapping. Pontius et al. (2006) presented new techniques to convert the predictions of forest-cover types for each pixel into conditional probabilities for land-change models. These studies echo the promising super-resolution mapping technique (e.g. Tatem et al. 2002).

At the 90% similarity level, clusters 5.2, 2.2, 3.1 and 3.5 represent relatively new LUCC research activities because these four clusters consist of papers published in the latter two periods of analysis (i.e. 1998–2002 and 2003–2007) (figure 4). Cluster 5.2 ‘forest-conversion-pool’ is the only cluster without remote-sensing journal papers, suggesting its association with LUCC consequences. Indeed, studies within cluster 5.2 have focused on the conversion of forests into other land-use/cover types, and the subsequent effects on soil-nutrient pool dynamics (McGrath et al. 2001, Powers 2004). When assessing long-term LUCC impacts on soil nutrients, recent biogeographical research has estimated forest-stand age from historical aerial photographs so as to understand how the edaphic environment changes through forest successional time (Matlack 2009). During the past decade, there has been considerable work on soil carbon dynamics of forest ecosystems, but data for non-forest ecosystems are still under-represented. Accordingly, Cernusca et al. (2008) contribute to the understanding of LUCC consequences on carbon cycling in mountainous grasslands.
Cluster 2.2 ‘water-spatial-catchment’, within 2.X ‘forest-model-river’, signifies that the physical systems under consideration are forest and water. A review of its large number of papers reveals three types of research. First, indicated by the three other index terms (i.e. forest, river, agriculture) of cluster 2.2 (figure 2(a)), studies examined the relationship between LUCC, in particular the agricultural activities in originally forested catchments and water quality, especially nitrate cycling, in rivers (van Herpe and Troch 2000), riparian zones (Schilling 2007) and groundwater (Ritter et al. 2007). Efforts have continued to investigate factors controlling nitrate concentration in groundwater of agricultural areas (Koh et al. 2009) and quantify the relationship between in-stream nitrogen concentration and the proximity to the streams of agricultural land (Mourad and van der Perk 2009). The second focus is on boundary-layer dynamics. De Beurs and Henebry (2004) used land-surface phenology as a means to detect changes in agricultural land cover, and Biftu and Gan (2000) applied various models to estimate evapotranspiration and its association with different vegetated land-cover types. The third type of research takes a land-use history perspective. In conjunction with a GIS, Hall et al. (2002) employed presettlement land-survey records (PLSRs), while Pascarella et al. (2000) used aerial photographs, to document changes in forest-community distribution and composition.

Conversely, clusters 3.1 and 3.5 have relatively small amounts of research activities. Cluster 3.1 ‘albedo-estimate-component’ was aimed at estimating energy and water balance (Wei and Fu 1998, Davidson and Wang 2004). Cluster 3.5 ‘soil-crop-rate’ included studies highlighting the importance of understanding biogeochemical cycling, a research priority identified in NRC (2001) and an issue still of concern (GLP 2005). As opposed to some of the studies in 2.2 ‘water-spatial-catchment’, which focused on nitrate cycling in water systems, cluster 3.5, seen in its derived labels (figure 2(a)), investigated nutrient (Zhu et al. 2006) and soil methane (Verchot et al. 2000) fluxes in terrestrial cropland or pasture. Estimating methane emissions in the context of LUCC is challenging due to the high spatial and temporal variability of environmental drivers. To combat this, high spatial-resolution information on the distribution and extent of the ecosystem environment is essential. Schneider et al. (2009) therefore generated new and high-resolution land-cover data to quantify methane emissions from the largest Arctic delta ecosystem. Dias et al. (2010) used plant-species composition as a proxy to predict methane fluxes in peatland ecosystems in the context of LUCC.

### 3.2.2 Re-emerging topics
Apart from revealing new research topics, analysis of year of publication exhibited that 5.4 ‘classification-network-forest’ and 7.2 ‘map-sediment-accuracy’ lacked research activities from 1993 to 1997 (figure 4). Detailed examination of these papers indicated that the temporal gap of research activities in cluster 5.4 was almost 10 years. The adjacency of these two clusters on the SOM derived map (figure 2(a)) suggested some degree of similarity in research contents. Indeed, the labelling term ‘classification’ in both clusters stressed their research themes on land-use/cover classification. Three examples from the two clusters showed that advances in methods or images might have enabled researchers to revisit certain topics on land-use/cover classification.

First, recent endeavours in both clusters addressed issues related to classification accuracy. Cluster 5.4 ‘classification-network-forest’ incorporated artificial neural networks into the accuracy improvement for land-use/cover classification (Kavzoglu and Mather 2003, Verbeke et al. 2004). Cluster 7.2 ‘map-sediment-accuracy’ was concerned with various components of the accurate mapping process, such
as assessment of accuracy (Foody 2002) and improvement of LUCC estimates (van Oort 2005). The accuracy of land-use/cover mapping and change detection is still a major concern in LUCC studies. Consequently, Stehman (2009) provided guidance for choosing a sampling design to assess land-cover classification accuracy. Stehman et al. (2009) further recommended that rare land-cover classes require large sample sizes to ensure that the accuracy estimators have a small bias. Foody (2010) directed the attention to the impacts of ground reference-data error on the accuracy of land-cover change detection. The re-emerging work in cluster 7.2 also noted that semantic differences in land-cover class definitions between two maps of different dates could contribute to uncertainties in LUCC estimation (van Oort 2005). The differences in land-cover nomenclature not only occur over time, but also across space, when multiple data sources are required. Ahlqvist (2008) thus proposed methods to address the semantic issues, and Herold et al. (2008) harmonized the thematic legends of four global land-cover maps to address the lack of interoperability of different datasets.

Second, attempts were found in cluster 7.2 to use new or high-resolution imageries to update previous land-use/cover mapping. For instance, Wikantika et al. (2007) assessed the use of IKONOS and QuickBird to identify land-use changes in a post-tsunami disaster. The launch of satellites of increased spatial resolution (e.g. WorldView-1 in 2007 and WorldView-2 in 2009) offers the potential for more detailed and accurate land-use mapping using textural features. Arising from this, Pacifi ci et al. (2009) employed very high resolution panchromatic images from QuickBird and WorldView-1 to classify the land use of four different urban environments. The availability of the MODerate Resolution Imaging Spectroradiometer (MODIS) Terra imageries might also account for the re-emergence of research activities in cluster 7.2 after 1999 on characterizing land-cover types (Thomlinson et al. 1999) and vegetation parameters such as leaf area index (LAI) (Cohen et al. 2003). Over time, there has been substantial work based on MODIS imageries and products. Maselli et al. (2009), for example, enriched the existing land-cover map of the Food and Agriculture Organization of the United Nations with information on vegetation productivity and phenology derived from the MODIS Normalized Difference Vegetation Index (NDVI) data. Coops et al. (2009) explored the relative predictive power of environmental descriptors, such as land cover and productivity derived from MODIS, for estimating breeding-bird species richness. Time-series MODIS products on vegetation, such as NDVI and LAI, add another dimension of information for species-distribution modelling (Bradley and Fleishman 2008). Such analysis may provide insights into habitat quality, thereby enabling creative approaches in the examination of LUCC consequences.

Third, efforts from a landscape ecological perspective were made in cluster 5.4 to link patterns of forest patches with processes in agricultural landscapes. For instance, Turner and Ruscher (1988) and Pan et al. (2001), two studies that published more than 10 years apart, both integrated historical aerial photographs with GISs and described land-cover change using landscape metrics. The latter study (Pan et al. 2001) argued that they enhanced the research topic by linking the spatial patterns of the forest patches to the processes, particularly the land-use dynamics that generated them. Synergies between remote-sensing and landscape ecological analyses are necessary because measurements of the spatial configuration of landscape patterns derived from remote-sensing imageries provide insights into processes (Walsh et al. 2008). Effects of landscape heterogeneity and patch size on classification accuracy were previously evaluated in Smith et al. (2003) of cluster 7.2, while Lechner et al. (2009) recently contributed to the accurate mapping of small remnant and linear vegetation
patches because of their ecological values in providing landscape connectivity and changing the degree of fragmentation. Small and linear features and subtle changes in the landscape may be under-represented when mapped using remote sensing. Accordingly, Houet et al. (2010) stressed the need for fine-scale analyses to better understand the LUCC processes.

3.2.3 Fading activities. A decreasing number of publications in a cluster over time may indicate a vanishing topic. Most of the clusters consist of studies from the four time periods of analysis with a growing number of publications, except for clusters 6.5, which exhibit no research activities in the latest time period (i.e. 2003–2007) (figure 4). The six highest frequencies of index terms of cluster 6.5 are forest, tropical, image, radar, synthetic aperture radar (SAR) and system (figure 2), suggesting its research activities on radar images and tropical forests. Indeed, imaging radars have the advantage of monitoring tropical LUCC in the presence of cloud cover. Detailed examinations show that papers of this cluster were published before 2000 and were mostly concerned with using radar images in deforestation monitoring of tropical rainforests (Stone and Woodwell 1988, Kuntz and Siegert 1999).

The lack of publications after 2000 in 6.5 ‘forest-tropical-radar’ suggests that researchers might have taken the advantages of new optical satellite imageries, such as MODIS, to examine LUCC in tropical forests, as seen in Hayes and Cohen (2007), which was classified into 3.3 ‘forest-landscape-surface’. Although the humid tropics were especially cloudy at the time of the Landsat overpass (Ju and Roy 2008), the low cost of Landsat data continuously facilitates the applications of the imageries in tropical deforestation issues (Muñoz-Villers and López-Blanco 2008). Alternatively, the SOM approach may have classified LUCC studies published after 2000 using radar imageries in tropical forests into other clusters, such as Saatchi et al. (2000) and Roberts et al. (2003) in the adjacent cluster 6.2. Although both studies employed radar imageries, Roberts et al. (2003) used a diversity of satellite sensors and Saatchi et al. (2000) compared the results with maps from other data sources and sensors. Many studies have combined optical and radar data for tropical land-cover mapping, as also seen in the recent work by Erasmi and Twele (2009), which combined Landsat Enhanced Thematic Mapper Plus and Envisat–Advanced SAR (ASAR) data. The frequency of occurrence of ‘radar’ might have thus been diluted, resulting in the SOM approach excluding these studies from cluster 6.5 ‘forest-tropical-radar’. Additionally, these papers used the word ‘Amazon’ more often than ‘tropical’, even though the Amazon region is mostly tropical forests. This presents an issue on index-term meanings that deserves further research for information retrieval and knowledge-domain visualization using SOM. Depending on the purpose of visualization, studies may consider to further combine some index terms, such as Amazon and tropical.

3.3 Spatial scale of analysis

To explore if the SOM-derived LUCC research clusters are scale relevant, we divide the spatial resolutions and extents noted in the papers into several classes. Because observation and monitoring of LUCC often involve satellite-based remote sensing, we adopt the classification of satellite sensors based on resolutions (Jensen 2000) to divide the resolutions of LUCC analysis into four classes: coarse resolution (>1 km); moderate resolution (250 m–1 km); high resolution (20–250 m) and very high resolution (<20 m). A study with a very high resolution dataset means that it uses an image with a pixel size or a dataset with a grid cell smaller than 20 m × 20 m (i.e. 400 m$^2$). For
spatial extent, we consider the classification of Zielinski (2002) to derive four size classes for the study area: small (<250 km$^2$); intermediate (250–2500 km$^2$); large (2500–25 000 km$^2$) and very large (>25 000 km$^2$). We exclude studies that do not provide information on resolutions or extents from mapping.

Most of the LUCC research activities involved various resolutions and extents (figures 5 and 6). Although some degree of correlation is considered to exist between resolution and extent such as that a small extent often requires a high resolution to depict the patterns and processes, extent can vary independently of resolution (Turner et al. 2001). There is probably no single appropriate scale at which researchers may expect to examine a specific LUCC topic, but the SOM derived maps reveal some interesting patterns. The most noticeable clusters are within 5.X ‘model-landscape-soil’, which used data of no resolutions (i.e. cluster 5.2) and of high resolutions at small extents (i.e. cluster 5.4). Cluster 5.2 ‘soil-forest-conversion’ recorded no information on resolution (figure 5) and lacked remote-sensing journal papers (also see section 3.2.1), suggesting that remotely sensed data were not their main datasets. Indeed, to assess the impacts of LUCC, research within the cluster used field measurements at a single

![Figure 5](image)

**Figure 5.** (a) Data resolution used in the clusters of land-change research activities. Each square symbol represents a paper that provided information on data resolution. The locations of the symbols correspond to those in figure 2(a). (b) Numerical labels are provided for the clusters at the 90% similarity level for easy referencing in this article.
location, rather than a fixed spatial resolution, to represent a whole catchment or a specific land-use/cover type. For example, Powers (2004) collected soils from 50 sites of different land covers in northeastern Costa Rica to evaluate the effects of LUCC on soil carbon and nitrogen pools. McGrath et al. (2001) used published data from 39 studies on nutrient dynamics in natural forests and forest-derived land uses to test the hypotheses concerning the effects of LUCC on soil properties.

Adjacent to the no-resolution cluster 5.2 is cluster 5.3 ‘spatial-climate-region’ (figure 2(b)), which includes some studies using coarse resolutions. Indicated by its labelling ‘region’, cluster 5.3 also includes studies at very large extents (figure 6). Zaehle et al. (2007), for example, projected changes in terrestrial carbon storage under climate and land-use change on a 10' × 10' coarse resolution grid for Europe. The adjacency of clusters 5.2 and 5.3 on the SOM derived map, the relevance of their topics on soil carbon and yet the differences in the employed resolutions and extents suggest potential research opportunities in linking their analyses and results across scales.

Figure 6. (a) Spatial extent of analysis in the clusters of land-change research activities. Each square symbol represents a paper that mentioned its study-area size. The locations of the symbols correspond to those in figure 2(a). (b) Numerical labels are provided for the clusters at the 90% similarity level for easy referencing in this article.
On the other hand, cluster 5.4 ‘classification-network-forest’ used high and very high resolutions (figure 5). Studies within the cluster either did not mention study-area size or were only concerned with small extents (figure 6). Research without mentioning study-area size focused on methodologies, such as testing land-cover classification methods in mountainous terrains (Blesius and Weirich 2005) and improving agricultural land-use mapping (El-Magd and Tanton 2003). Research of small extents mostly analysed aerial photographs, rather than satellite imageries. In conjunction with a GIS, Turner and Ruscher (1988) used aerial surveys to estimate herded-livestock distributions, which reflected local land-use patterns and other environmental factors; Pan et al. (2001) employed statistical methods to detect LUCC patterns from aerial photographs and related these patterns to the underlying physical structures of landscape elements.

Investigations of consequences of agricultural land use seem to focus on catchments of intermediate extents, as seen in cluster 8.3 ‘catchment-agriculture-river’ (figures 2 and 6), with the high-resolution data used (figure 5). Fisher et al. (2006) examined the intensity of land use, such as fertilizer application and human-population density, on producing eutrophication. Kiage et al. (2007) linked land-cover changes to the rising human and livestock population in a catchment in Kenya and found that deforestation and subsequent land degradation increased turbidity and sediment yield in the lake. A number of studies of LUCC consequences have now focused on catchments of very large extents with the high-resolution Landsat data analysed. Ruelland et al. (2008) evaluated the potential of Landsat imageries to monitor vegetation cover for the improvement of hydrological modelling of a catchment of 120 000 km$^2$ in West Africa. Rodriguez et al. (2010) analysed streamflow records with deforestation maps derived mainly from Landsat to detect hydrological response signals potentially related to LUCC for an Amazonian drainage basin of 33 000 km$^2$.

Satellite sensors appear as the high-frequency terms for labelling for cluster 7.3, with the Advanced Very High Resolution Radiometer (AVHRR) as the second term and for cluster 4.3, with Landsat as the fifth term (figure 2). Most studies within these clusters probably used images from the respective sensors. For example, Amundson et al. (2003) in cluster 4.3 determined soil disturbance using the US National Land Cover Data, a dataset with a 30-m resolution interpreted from the Landsat Thematic Mapper data. For cluster 7.3, ‘global’ and ‘region’ also appear as the labelling terms, signifying research activities at large extents. In fact, studies within cluster 7.3 conducted at large extents all involved the use of AVHRR. Lambin and Ehrlich (1995) analysed multi-temporal AVHRR data for the integration of thermal information on surface temperature with a vegetation index to increase biome discrimination at a continental level; Giri et al. (2003) analysed multi-temporal AVHRR data to prepare historical land-cover maps and to identify areas undergoing major land-cover transformations of continental southeast Asia. Global datasets derived from the AVHRR in the 1990s have assisted in many LUCC studies of large extents. The availability of the newer remote-sensing data sources (e.g. MODIS and SPOT VEGETATION) have caused the methodologies and products for land-cover analysis of large extents to evolve rapidly, e.g. the MODIS Collection 5 Land Cover Type product recently described in Friedl et al. (2010).

3.4 Temporal dimension of examination

We obtained the temporal dimension of examination using the start and end dates of the maps, imageries, or in situ data analysed in the studies, and divided it into four time periods. The first time period was the shortest. We referred to it as ‘single year’,
including studies that used a map or an image from one single year, or whose data start and end dates were within 1 year. The second time period, referred to as ‘short term’ here, was between 2 and 35 years. We used 35 years as the cutoff because satellite imageries have been widely used in LUCC studies and Landsat-1, the first unmanned satellite designed to acquire data on Earth resources (Lillesand et al. 2004), was launched in 1972, 35 years before the latest year of the papers (i.e. 2007) analysed in this study. The third time period, referred to as ‘intermediate term’, was from 36 to 100 years, covering studies using data in the 20th century. The fourth time period, referred to as ‘long term’, included studies using data of more than 100 years. Apart from representing the four time periods of examination on the SOM derived map (figure 7), we summarized the number of papers in different time periods by clusters at the 80% level, including papers with no indication of temporal dimension analysed (table 1).

The following clusters at the 90% level mostly consisted of studies of single year and short term: clusters 4.1 ‘validation-video-resolution’, 2.1 ‘global-estimate-vegetation’, 3.1

![Figure 7](image)

Figure 7. (a) Temporal dimension of consideration in the clusters of land-change research activities. Each square symbol represents a paper that noted the duration of study. The locations of the symbols correspond to those in figure 2(a). (b) Numerical labels are provided for the clusters at the 90% similarity level for easy referencing in this article.
albedo-estimate-component’, 3.5 ‘soil-crop-rate’, 7.4 ‘NDVI-spectral-technique’, 7.3 ‘global-AVHRR-region’ and 6.5 ‘forest-tropical-radar’ (figure 7). These clusters had a high proportion of papers from remote-sensing journals (figure 2(a)) and their labels, such as resolution, NDVI, spectral, AVHRR and radar, were associated with satellite imagery. Thus, if the focus of an LUCC study was on the use of satellite imageries, such as developing techniques to extract land-cover information (Cihlar et al. 1996), its temporal dimension of analysis was likely to be single year or short term.

Of the 722 analysed papers, 63 were of intermediate term. The two adjacent clusters, 2.X ‘forest-model-river’ and 5.X ‘model-landscape-soil’ (figure 7), together accounted for one-third of the intermediate term LUCC research (table 1). In terms of journal type, 29 (46%) and 18 (29%) of the intermediate term studies were from Landscape Ecology and Hydrological Processes, respectively (result not shown). The temporal dimension of examination in landscape ecological studies seemed to depend on the time span for which aerial photography was available, although this might subsequently limit the scale of analysis to a small extent (e.g. Turner and Ruscher (1988), Pan et al. (2001) in section 3.3). The availability of data from meteorological stations and stream gauges established during the 20th century also facilitated hydrological studies on LUCC consequences in the intermediate term (e.g. Sato et al. (2007) in 2.3 ‘urban-image-surface’ and Hejazi and Moglen (2007) in 5.2 ‘forest-conversion-pool’).

Of the 24 long-term LUCC studies, approximately half of them (46%) were found in cluster 6.X (table 1). Their distributions, indicated as large black squares in figure 7, corresponded to the aggregations of the papers from natural-science journals.
especially on the right-hand side of clusters 6.2 and 6.4 (figure 2(a)). Because a centurial time frame is needed to investigate environmental processes related to LUCC, such as human-induced disturbances and their resulting biophysical responses in the landscapes (Delcourt et al. 1983), long-term LUCC studies rely on the availability of historical references. For example, in cluster 6.2, based on the topographic maps of ca. 1898, Korytny et al. (2003) were able to analyse changes in the river network structure and show the intensity of the erosion process as a result of agricultural activities during the 20th century. Systematic flood recording at York in the UK since 1878 enabled Longfield and Macklin (1999) to assess the influence of large-scale land-use changes on flood magnitude and sediment availability. With a creditable 1883 land-cover map and climate data back to 1915, Cuo et al. (2009) found that the effects of the changing land cover and climate, primarily temperature, on streamflow occurred differently at high and low elevations of the Puget Sound basin, USA. Another valuable historical source of information is the PLSRs that record vegetation and landscape conditions preceding widespread European settlement in the US (Wang 2005). Foster et al. (1998) in cluster 1.2 ‘soil-landscape-region’ analysed a range of archival data, including PLSRs, to assess historical changes in forest vegetation and land use in New England over the past three centuries. Schulte et al. (2007) in 6.3 ‘forest-pixel-sediment’ used a combination of PLSRs and current forest-inventory and land-cover data to quantify the regional level consequences of one century of Euro-American land use in the US Great Lakes region. The inclusion of PLSRs in these LUCC studies revealed that environmental change did not occur only in the last several decades. Instead, their results raised the possibility that the effects of historical land-use practices could be longer, might have altered vegetation–environment relationships across broad geographic regions and hence should be considered as major components of global change.

The incorporation of historical maps and surveys and GIS into remote-sensing studies has greatly facilitated the reconstruction of long-term LUCC and the assessment of its environmental impacts. Historical maps and surveys extend the time frame of consideration from decades to centuries, beyond what satellite imagery and aerial photography can provide. GIS enables the integration of land-use/cover data of different time periods and facilitates the modelling of additional datasets, e.g. soils and human populations, to estimate the impacts of LUCC. Gomarasca et al. (1993), in cluster 3.3 ‘forest-landscape-surface’, digitized thematic maps of 1888–1990 to compare with present land-use/cover maps derived from satellite imageries to study LUCC in areas affected by rapid agriculture transformation and residential/industrial development in Milan, Italy. Fisher et al. (2006), in 8.3 ‘catchment-agriculture-river’, reconstructed a 350-year LUCC history in the mid-Atlantic region of the US using remote sensing, aerial photographs, historical maps and socio-economic records. To estimate the effects of LUCC, they assembled historical data between 1660 and 1850 on human populations, agricultural crop yields, labour statistics and crop exports, analysed water-chemistry data available from the mid-20th century, and then incorporated these datasets into modelling of biogeochemical impacts.

Alternatively, as Cousins (2001) in 7.2 ‘map-sediment-accuracy’ commented, correctly rectified historical maps are crucial for addressing questions related to LUCC in historical time. Although many LUCC studies have developed methods to improve the accuracy of recent land-use/cover maps, particularly those derived from satellite imageries, few studies have attempted to investigate the uncertainty in the historical maps for LUCC research and the possible integration issues of historical maps and
present remotely sensed imageries. Recently, Leyk and Zimmermann (2007) of 2.2 ‘water-spatial-catchment’ presented a method for correcting inherent classification bias in historical maps for which subsequent LUCC analysis could be improved. Because historical sources of information often contain considerable inherent uncertainty, comparing historical maps with contemporary maps and imageries for long-term LUCC studies requires an extensive evaluation of data quality.

4. Conclusions and future work

To facilitate the understanding of the patterns and trends of land-change research, this article used a content-based text-retrieval approach, SOM and cartographic techniques to visualize research activities on LUCC and its environmental consequences from 722 journal papers over the past two decades. Using hierarchical clustering, we obtained a total of 8 and 30 clusters, respectively, at the 80% and 90% similarity levels based on research contents measured using frequencies of index terms. We presented the results in map-like displays and discussed patterns of research activities in the identified clusters based on journal types, year of publication, spatial scale of analysis and temporal dimension of examination. The results are not intended to be all inclusive, but draw examples from analyses of a large number of remote-sensing and natural-science literature.

The results showed that publications from remote-sensing and natural-science journals did not form clear aggregations at the 90% similarity level, suggesting research collaboration from different fields. The number of papers in each of the derived clusters has been growing, showing continuous research investment. Analysis of the publication years of the papers revealed emerging themes, re-emerging topics and possible fading activities. The most noticeable cluster of emerging themes was studies that developed new methods and techniques in remote sensing, such as the super-resolution mapping techniques, for mixed-pixel issues in land-cover mapping in forest catchments. Another new focus of the past decade was on the LUCC consequences on the dynamics of soil pools and nitrate cycling. Re-emerging topics centred on issues of land-use/cover classification because advances in satellite imageries and incorporation of different approaches might have facilitated researchers to revisit topics such as accuracy improvement for classification. Suggested by the lack of publications after 2000 in cluster 6.5 ‘forest-tropical-radar’, a possible fading activity in the seven selected journals was monitoring tropical deforestation with radar images. This finding, however, presents two issues for further investigation. First, in terms of LUCC study, researchers need to scrutinize if research using radar images in monitoring tropical deforestation has also diminished in other publication outlets during the last decade. Because tropical forests have undergone dramatic LUCC in the last decade (Achard et al. 2002), it is important to understand whether this finding presents a potential concern that more LUCC studies in tropical forests are needed, or in fact, whether studies have employed imageries from optical sensors or multisensors for such tasks. Second, in terms of the SOM approach, the issue on the meanings of the index terms deserves further research for information visualization. Some papers might have used Amazon more often than tropical, even though the Amazon region is mostly tropical forests. Depending on the purpose of visualization, researchers may consider to further group some index terms.

For the spatial scale of analysis, most clusters involved resolutions and extents of various sizes. Additionally, the derived labelling terms, such as Landsat, AVHRR and
global, provided insights into the dominant resolution and extent of analysis for certain research activities. For example, cluster 7.3 ‘global-AVHRR-region’ indeed consisted of studies using coarsely resolved AVHRR images for land-cover mapping over large areas. The advantage of coarse-resolution imageries is that they often have a better temporal resolution than high-resolution imageries, thereby providing a consistent land-cover map over a large area at a fixed point in time. This is important for LUCC research because many of the environmental processes related to LUCC are time dependent. The challenge of using coarse-resolution imageries, however, is that at a resolution such as 1 km × 1 km, most pixels have mixed land-cover types. Studies that took the advantage of coarse-resolution imageries for LUCC research have formed a cluster (i.e. 7.3 ‘global-AVHRR-region’) in the LUCC knowledge domain (figures 2(a) and 5), while studies that proposed new methods to deal with the challenge of the mixed-pixel issue have appeared to be an emerging theme as cluster 1.1 ‘spatial-resolution-pixel’ (figures 2(a) and 4). That the SOM-based representations reveal research activities on both taking the advantage and dealing with the challenge of coarse-resolution imageries warrants continuous research investment on this topic.

In terms of temporal dimension of examination, the duration of consideration for intermediate term studies seems to depend on the availability of aerial photography. Historical sources of information, such as presettlement land-survey records and archival flooding records, facilitated assessment of long-term consequences of LUCC. Unlike the large amount of efforts on accuracy improvement for contemporary land-use/cover maps derived from satellite images, few attempts have been made to examine the uncertainties in the historical maps for LUCC research. The integration issues of historical maps and present remotely sensed imageries, such as positional errors and semantic interoperability of land-cover classes between two datasets, hence present important topics for further research.

Acknowledgements
The National University of Singapore provided funding (Grant no. R-109-000-070-101) for this research. We thank André Skupin for his valuable suggestions on the analysis; Seeleng Quek, Cheryl Chen and Huan VuDuc for their help in preprocessing the text files; Yikang Feng for his assistance in mapping and two anonymous reviewers for their helpful and encouraging comments.

References


