Applied Geography 74 (2016) 136–150

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Mapping long-term land use and land cover change in the central Himalayan region using a tree-based ensemble classification approach

Anusheema Chakraborty [a], Kamna Sachdeva [a], P.K. Joshi [b,*]

[a] Department of Natural Resources, TERI University, New Delhi, 110070, India
[b] School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, 110067, India

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form
25 June 2016
Accepted 21 July 2016

Keywords:
Landsat
Multi-temporal LULC maps
Random forests
Decision level fusion
Himalaya

ABSTRACT

Forest cover and its change analysis along with nexus between other land cover types are often seen as insufficient data quality for operational applications in the Himalayan region. Despite extensive documentation reporting rapid demographic, socio-economic and environmental changes in this region, we lack comprehensive detailed assessments of spatial distribution of land use/land cover (LULC) change over an extended period of time. In this study, we overcame this limitation by producing annual maps of change among forests and other LULC classes in the Kumaon division in the state of Uttarakhand, India. This is the first attempt to develop a database for this region using public domain Landsat images and replicable mapping techniques. To deal with high spatial and temporal variability as well as complex multi-signature classes, this study uses a tree-based ensemble classification approach. The central premise of the approach is to exploit multi-seasonal information using characteristic temporal signatures in several spectral regions along with various environmental variables to identify twenty (20) LULC classes spanning three decades, focussing on distinguishing geographically dominant forest types. The maps were combined into seven LULC classes with reference to global databases. Random forest (RF) classifier was used to create seasonal maps, and knowledge-based decision level fusion was used to produce annual composite maps. Overall accuracies were equal to 82% (kappa = 0.75), 87% (kappa = 0.81), 87% (kappa = 0.82), and 88% (kappa = 0.83) for 1990, 1999, 2009 and 2014, respectively, while detailed maps had moderately high (~70%) overall accuracies. As forests in the Himalayan region represent the most widespread vegetation structure, development of such time series analysis in this region can be potentially used for national and regional resource management efforts. This study, therefore, gives an insight on the potential of using a tree-based ensemble classification approach to provide a baseline database, which can aid in developing practical field inventories and forest conservation policies.

1. Introduction

Mapping and monitoring land use/land cover (LULC) change is fundamental information required in a number of fields such as climate modelling (Brovkin et al., 2013; Pielke et al., 2011; Verburg, Neumann, & Nol, 2011), hydrological modelling (Bulygina, Ballard, McIntyre, O'Donnell, & Wheater, 2012; Nosetto, Jobbágy, Brizuela, & Jackson, 2012), biogeochemical cycling studies (Sohl et al., 2012), environmental conservation (Giri et al., 2011; Xiuwan, 2002), forestry studies (Hansen et al., 2013), biodiversity (Tuanmu & Jetz, 2014; Turner et al., 2015), and natural resource management (Lambin et al., 2001; Turner, Lambin, & Reenberg, 2007). Over the last few decades, our understanding of the interdependencies between LULC change, socio-ecological system and environmental system has increased (Adam, Mutanga, Odindi, & Abdel-Rahman, 2014; Eisavi, Homayouni, Yazdi, & Alimohammadi, 2015). This necessitates the need to develop timely accurate information in the field of land change science (LCS) to understand spatial and temporal dynamism inherent in any landscape (Cassidy, Southworth, Gibbes, & Binford, 2013).

With the continuous availability of satellite-based remotely sensed data, the efforts to accurately classify forest types have increased over the years in order to correctly capture the true dynamism of any socio-ecological landscape. The ambiguity in

* Corresponding author.
E-mail address: pkjoshi27@hotmail.com (P.K. Joshi).

http://dx.doi.org/10.1016/j.apgeog.2016.07.008

management approaches for forest conservation raises the need to develop newer approaches to facilitate higher classification accuracies in mapping different forest types and other LULC patterns (Paneque-Gálvez et al., 2013). In practice, the trade-offs between the choice of spatial, temporal and spectral resolution, in addition to availability and costs of acquiring datasets determine the quality of LULC mapping products (Eisavi et al., 2015). Very high-resolution satellite images, unlike moderate- or medium-resolution satellite images, provide a far better advantage for differentiating species and mapping individual trees (Pu & Landry, 2012). However, unavailability of free and open data policies restricts the adequate use of these datasets to efficiently process and quantify forest cover change (Hansen et al., 2013). Although, a number of reliable historical and current LULC information are available at the local, regional, and global level, with spatial resolution ranging from 300 m to 1 km, the demand for new LULC products with improved accuracy has been increasingly recognized by the remote sensing community (Giri, Pengra, Long, & Loveland, 2013).

The advent of the free U.S. Landsat data policy, the longest-running archive, has often helped to capture most significant human activities on the Earth's land surface for the last 40 years (Kennedy et al., 2014). Despite facing difficulties in consolidation of the global archive, especially in the mid-1980s (Loveland & Dwyer, 2012), Landsat data still remains the best Earth-observation data coverage, as only few other sensors are available over such a long period of time (Wulder, Masek, Cohen, Loveland, & Woodcock, 2012). One of the key concerns over recent years include reducing deforestation and forest degradation, where historical assessments through Landsat images have proven to be very useful (Ankersen et al., 2015; DeVries, Verbesselt, Kooistra, & Herold, 2015; Kim et al., 2014). At the global level, attempts to generate LULC information with 30 m Landsat products manage to map ten (10) land cover classes and have overall accuracies between 60% (Gong et al., 2013) and 80% (Chen et al., 2015). Mapping global LULC changes with medium-resolution satellite data (for example, 30 m Landsat data) is very complex, particularly if based on single and simplistic algorithms that could only provide limited insights into real forest dynamics at local scales (Tropek et al., 2014). Nonetheless, given the long-term regularly sampled historical archive, both in terms of frequency and coverage, Landsat remains an optimum solution to effectively capture and quantify where and when important natural and human-caused changes have occurred, by proving preliminary identification of on-going changes (Kennedy et al., 2014; Wulder et al., 2015).

Since the launch of the first land-observation satellite back in 1972 (Landgrebe, 2005), field survey techniques have been commonly used to classify LULC types (Adam et al., 2014). Subsequently, large numbers of new algorithms were developed to generate reliable LULC maps from satellite images (Lu & Weng, 2007), in attempt to improve classification accuracies (Foody, 2002). Different image classifiers generate different LULC outputs, with each classifier having its unique capabilities and the likely chance of high performance depending on the field of application, image characteristics, and ultimate goal of the study (Kennedy et al., 2014; Liu, Skidmore, & Van Oosten, 2002; Lu & Weng, 2007; Srivastava, Han, Rico-Ramirez, Bray, & Islam, 2012; Szuster, Chen, & Borger, 2011).

Mapping complex, heterogeneous landscapes where vegetation types lack spectral discernibility and are poorly represented by large pixels can be highly challenging. The use of traditional classifiers (such as conventional parametric statistical techniques) are often unable to detect finer-scale changes including disturbances, long-term trends, and seasonal variations (Cassidy et al., 2013). For higher accuracies, various machine learning algorithms have emerged as more efficient classification approaches for dealing with large dimensional and complex data spaces to detect high spectral and spatial variability in mountain landscapes (Cingolani, Renison, Zak, & Cabido, 2004; Fan, 2013; Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012). For classifying multiple sources of input data, random forest (RF) classifiers perform better than parametric approaches, like maximum likelihood, or non-parametric classification approaches, such as support vector machines (SVMs), decision trees (DT), and neural networks (Mellor, Boukir, Haywood, & Jones, 2015; Stefanski, Chaskovskyy, & Waske, 2014; Tatsumi, Yamashiki, Canales Torres, & Taipe, 2015). LULC classifications are convenient abstractions that can be better characterized by considering other significant variables. For instance, continuous surfaces, such as the Principal Component Analysis (PCA) (Deng, Wang, Deng, & Qi, 2008), Normalized Difference Vegetation Index (NDVI) (Cassidy et al., 2013), or the Digital Elevation Model (DEM) (Ghosh, Sharma, & Joshi, 2014), reflect range of variability within the LULC classification schemes. These layers can be associated and quantified with actual biologically and/or physically meaningful units (Southworth, Munroe, & Nagendra, 2004). In addition to this, multi-seasonal images can often significantly improve classification accuracies by discriminating various LULC categories, especially for accuracies of classes with rapid temporal change behaviour such as agriculture, grasslands and forest scrub (Ghosh et al., 2014; Rodriguez-Galiano & Chica-Olmo, 2012; Schneider, 2012; Senf, Leitão, Pflugmacher, van der Linden, & Hostert, 2015).

The Himalayan region, despite being rich in terms of biodiversity, have limited number of key studies (Gairola, Procheş, & Rocchini, 2013; Sharma & Chettri, 2005), making it difficult to understand and accordingly plan resource usage, developmental priorities, and conservation efforts. This region is often considered data insufficient with prior mapping efforts being sporadic, inconsistent, and also inaccessible (Uddin et al., 2014). Although, there has been an increase in LULC databases using different remote sensing techniques, there exists great diversity in the area and scale of study, purpose of mapping, classification schemes, methods of mapping, and accuracy assessments (Table 1).

Long-term broad-scale spatial and seasonal variation in vegetation phenology has been reported in the Himalayan region (Mishra & Chaudhuri, 2015; Panday & Ghimire, 2012; Shrestha, Gautam, & Bawa, 2012). However, to be able to detect fine-scale vegetation trends where field-based research is extremely difficult due to topographic relief, high altitudinal range, and spatial heterogeneity, it requires exhaustive training sets and multiple field visits. Moreover, local communities residing in this region are highly dependent on forest resources, both for subsistence and commercial purposes (Rayamajhi, Smith-Hall, & Helles, 2012). Therefore, considering the vast extent of this region and also the ecological and economic importance of forests, it becomes essential to analyse seasonal vegetation trends in this highly sensitive and fragile region. Most of the documented mapping efforts lack detailed classification schemes and species-level distribution maps, predominantly in the higher altitudes (Gairola et al., 2013). Therefore, to address this paucity of LULC baseline database, the aim of the study is to develop annual maps of change among forests and other LULC classes spanning three decades in the Kumaon division in the state of Uttarakhand, India. Furthermore, we focus on distinguishing geographically dominant forest types (pine, oak, mixed and deciduous forest). The objective is to produce annual composite maps for each year using a tree-based ensemble classification approach with information from multi-seasonal Landsat data (pre-monsoon from February to March and post-monsoon from October to November) to account for patterns of LULC change over a period of time.