Brief note on using geomatics to study land-cover change in the Tarai since the 1950s

Notes sur l'utilisation de la géomatique pour étudier les changements de l'occupation des sols dans le Téraï depuis les années 1950

Jérôme Picard
Brief note on using geomatics to study land-cover change in the Tarai since the 1950s

Notes sur l'utilisation de la géomatique pour étudier les changements de l'occupation des sols dans le Téraï depuis les années 1950

Jérôme Picard

1 Geomatics, a catch-all term for geography and information science, is the computer-assisted analysis and reprocessing of spatial data, usually combined with the use of a geographic information system (GIS). This tool is capable of highlighting and visualising changes in land-cover, that is the biophysical nature of land surface. Geomatics makes use of many different documents (topographic maps, aerial photographs, satellite images) and therefore allows comparisons over time but also on different scales. It requires field observations and GPS point surveys to support image interpretations.

2 The purpose of this note is to explain the main principles of geomatics (especially for social scientists) and to show how certain geomatic methods contribute to our studies on land-cover change in the Tarai. I first present the sources of data and the methodology, and then highlight some general results prior to outlining the limitations of the methods used.

3 The cartographic sources for this region are, on the whole, relatively varied and numerous but not always easy to use for studying changes in land-cover. There are three main types. The first type concerns topographic maps that have been in existence since 1925 (Boulnois 1967), the oldest ones I had access to being colour topographic maps from the Survey of India to a scale of 1:63,360 (one inch to one mile), and those from the US army to a scale of 1:250,000, both sets dating back to the 1950s–60s. Those from the 1990s, produced by Nepal’s Survey Department, are to a scale of 1:25,000 for the Tarai (1:50,000 for several mountain regions) and have the advantage of being vectorised and georeferenced in eight main information layers: transportation (mainly
roads), buildings, topography, hydrography, utility (power lines), administration, designated areas (parks and nature reserves) and land-cover. Three types of land cover are identified: cultivation, vegetation (forest, orchard, plantation, nursery, grass, bush, bamboo, scattered trees, tree), and others (rock outcrop, swamp, sand, barren land). The second type of map I used in my work is the vectorised thematic map from the Land Resource Mapping Project (LRMP 1986), to a scale of 1:50,000, and the land-use map from the Land Use National Project (2007, 1:50,000). The latter has the advantage over the LRMP map of showing the succession of crops and of delimiting categories more accurately. Although the scale of these maps is less accurate than the scale of topographic maps for the same period, they nevertheless contribute to the interpretation of certain satellite images by spatially delimiting the various categories they provide, and are especially suitable for studies at district level. When dated after 1957, these first two types of maps are derived from aerial photographs. The third type of source that can be used in geomatics is the satellite image. Landsat images, which are free of charge, have existed since 1975 with a resolution of 80m per pixel (ie a scale of approximately 1:200,000), and since 1984 with a resolution of 30m (approximately 1:75,000). The images produced are 180km x 180km views that offer visible and infrared channels. The Spot satellite, with roughly similar channels but a finer resolution over time of up to 1.5cm provides archives dating back to 1986.

The research project purchased SPOT 4 and SPOT 5 images at resolutions of 20m and 2.5m respectively, each covering 60 x 60km for the years 2006, 2014 and 2015 for Sunsari and Saptari, and for 2001 and 2011 for Rupandehi. For 2016, it acquired 6m-resolution SPOT 6 colour images and 1.5m-resolution SPOT 7 black and white images. The SPOT 5 images available since 2006 with a 2.5m resolution give a much better view of the different types of land-cover (forest, bamboo, crops at different stages, buildings) than Landsat images and offer wide coverage, which generally makes it possible to circumscribe an entire district.

Google Earth images however, are pseudo-satellite images because they are assemblages and overlays of various images. They have the dual advantage of having been archived online since the early 2000s and of being extremely precise for certain sectors (up to 60cm for one pixel), but have the disadvantage of not providing information on the radiometry of objects. We have not used Pleiades images – which have been available since 2013 and are of a very high resolution (50cm) – because of their very limited coverage (20 x 20km view) and because they require complex processing of the ‘object-oriented classification’ type due to their precision.

The methodology used here consisted of integrating and analysing the layers of information using a geographic information system (GIS). This involved geo-referencing sources if this had not already been done (ie applying geographic coordinates to them using the same projection to allow a perfect layering of the maps), analysing the results obtained and retrieving them in mapped form. The analysis of satellite images is based on the remote identification of objects that either have the same spectral signatures (remote-sensing analysis) (fig 1b, fig 2) or are recognisable by a specific shape and colour (photo-interpretation analysis) (fig 10). Some of these objects are easily identifiable because they are relatively homogeneous. This is the case for forests (which can be distinguished from cultivated areas next to them) (fig 1), sand, watercourses and certain roads. Homogeneous objects therefore belong to the same category or ‘class’ of land-cover. The processing of satellite images by supervised or
unsupervised classifications (both included in a ‘pixel classification’) using software (such as Arcgis, QGIS, ENVI) is used to produce land-cover maps (fig 4). ‘Unsupervised’ means classifications based solely on pixel values, with no prior human interpretation. They give a first (radiometric) overview of the land-cover classes (fig 4a), with a thematic interpretation of the resulting spectral classes being made afterwards. ‘Supervised’ classification refers to the use of sampled and observed areas in the field for running classification algorithms (fig 4b). Supervised classification has to be verified by field studies and then adjusted by applying an iterative process that alternates between data analysis and fieldwork. Google Earth images are also used in supervised classification to complete field data and to identify classes because of their very high resolution that often leaves room for an interpretation of the data (fig 3c). ‘Field checks’ were carried out through several observations and surveys conducted among farmers in Sunsari district in October 2013, March 2015 and November 2016. During these periods, cropping and harvesting rapidly followed one another; agricultural land-use therefore varied, enabling the landscape to be studied in all its diversity. The season in which satellite images are captured is important because at the end of the dry season (March to May) it can be difficult, for example, to distinguish almost bare and uncultivated land (especially next to houses) from harvested fields. It is logical to privilege supervised classification if one knows the terrain, though even in this case both classification methods are used.

Fig 1. Example of an area according to a Google Earth image extract (dated 8 Dec 2014) with pixel resolution less than 1m (1a) and a Landsat image extract (dated 30 Nov 2014) with pixel resolution 30m at infrared false-colour (1b)
For example, figure 1 shows an area of Sunsari where the forest appears very clearly in red (chlorophyllous vegetation has a high near-infrared reflectance, the channel of which has been colour-coded red here, see fig 1b). The rest of the area, which generally shows colours ranging from green to bright white, is made up of cultivated or harvested plots and of rural settlements. Theoretically, the whiter the appearance, the barer the ground, but here a halo of mist amplifies this phenomenon in places. The image underwent an unsupervised classification for which the user specified seven classes (fig 2) showing the main land-cover categories. A Landsat resolution of 30m is sufficient here to isolate the forest, but the level of detail is not fine enough to individualise houses properly. Figure 3 illustrates the difference in visuals provided by Landsat, SPOT and Google Earth images for the centre of a village (Pipara).
The different land-cover categories stand out more or less in the two classifications made on a portion of the Spot 5 image where the centre of the village is located (fig 3b, fig 4a and 4b).

Fig 4. Unsupervised classification (k-means) into 8 classes on SPOT 5 image extract dated 8 Nov 2014 using QGIS software (SCP plug-in) (4a) and supervised classification, using the k-nearest neighbours algorithm into 8 classes on the same image extract with QGIS software (SCP plug-in) (4b)

Sources: ©CNES, 2014, Spot Image SA distribution.

Shadows, bamboo and trees, with homogeneous spectral signatures, are recognisable and often well circumscribed. The length of the shadows depends on the position of the sun and on the height of the objects. At this scale, and with field knowledge, it is easier
to identify land-use. The plots of agricultural land not yet harvested and showing reduced chlorophyll activity (in light blue) are fully visible, as is rice stubble (ie harvested plots in orange). However, individual buildings (in bright red) are not always properly demarcated. In an unsupervised classification (fig 4a), the number of buildings is underestimated because the algorithm tries to create the most homogeneous classes possible without considering their significance. Moreover, classes with a small number of pixels cannot be easily mapped. However, we know that houses are very diverse, combining bare surfaces (flat roofs, tin roofs), vegetated surfaces (thatched roofs, roofs made of leaves) with very different spectral signatures. That is why in a supervised classification (fig 4b), a process in which the user themselves selects representative samples, the built environment is better represented, though there is some obvious ‘confusion’ here and there with plots of agricultural land. In order to analyse the SPOT image dated 8 November 2014, I decided to create a classification of eight categories, based on the information visible on the December 2014 Google Earth image and on our knowledge of the field. Indeed, I spent six days in the field, interviewed farmers about the history of some plots, took pictures (figs 6 and 7) and GPS coordinates of the different types of land-use, observed the different stages in plant development and the colour of plants before harvesting. Together with this, I drew 20 homogeneous samples or polygons (regions of interest or rois) to represent these eight classes (see fig 5). Then I ran the k-nearest neighbours algorithm that produced the supervised classification (fig 4b).

Fig 5. The 20 samples (regions of interest or rois) used to define the eight land-use classes

The roofs made of different materials (sheet metal, tiles) present a high level of surface reflectance but to varying degrees, as is the case of the farmyard and the country lane, hence possible confusion with bare plots, whether harvested or ploughed, near the village.

Supervised and unsupervised classifications based on the grouping of pixels only according to their radiometric criterion and the manual vectorisation of objects represent the main traditional image-processing techniques. We would like to emphasise that other more recent methods exist, they are sometimes more effective in the sense that they are better at individualising homogeneous objects and provide a more accurate mapping. For example, it is worth mentioning the ‘object-oriented’ classification that carries out a preliminary grouping of pixels into ‘segments’ with the aim of preparing and improving the classification. A segment is a homogeneous grouping of pixels according to the latter’s radiometric value and to other criteria such as shape and texture (texture is a spatial repetition of a basic pattern in different directions; it is determined by the arrangement of the groupings of pixels of different shades of grey). Some authors working in Nepal (Uddin et al 2015) have used this type of classification on Landsat images. The implementation of this method is relatively complex and time consuming. We tested it merely because its implementation using Open Source software had not proved successful. In fact, our classifications could have been refined by using more sophisticated methods such as multi-temporal analysis which is based on seasonal radiometric changes. This method, which requires several images, provides several reflectances throughout the year for the same object and the classification is therefore improved. We have limited ourselves to presenting examples based solely on the number given for each spectral band. However, there are other methods based on spectral indices: by combining several spectral bands, we obtain additional information, such as the normalised difference vegetation index (NDVI).
which combines near infrared and red and provides information on the state of the vegetation.

The methodology used here, which compiles maps or images from different dates, also makes it possible to produce diachronic maps. With the help of students, I studied the changes in land-cover and land-use (Butler 2015, Phay 2015, Dumont 2015) and then focused on the changes in the built environment (Havet et al 2017, Cusset et al 2018). The case of forest cover regression is a good example. For Rupandehi (fig 8), I had to vectorise the data from the 1958 US map. However, for the years 1975, 1989, 1999 and 2020, the result comes directly from a supervised classification using Landsat images. In Sunsari, the forest and brushwood represented 60,514ha (50.8% of the district) in 1962 and spread over the whole district (fig 9). It decreased to 54,168ha in 1977 (43%) and then to 33,744ha (27%) in 2014 (Dumont 2015), the loss of tree cover having benefited cultivated areas. Concerning the progression of built-up areas in the cultivated space, this has also been analysed for Tarai districts by Rimal (2018) based on supervised classifications of Landsat images from 1989 to 2016 at 30m. It has also been modelled to assess the urban footprint over the coming decades (Rimal 2019).

Fig 8. Changes in forest cover in Rupandehi from 1958 to 2020

Some limitations to our study became apparent due to the sources available and the specificities of the field. Firstly, few digital maps of Nepal exist apart from 1:25,000 topographic maps dating back to 1992–96. This scale is interesting because it provides access to numerous details on maps. Satellite imagery has made it possible to partially overcome the relative scarcity of maps, even if there are problems concerning the temporal availability of images and their resolution. Secondly, Landsat has been providing images since 1975, but these are often only available for the months of October to December when there is less cloud cover. The average resolution of 30m is suitable for the recognition of forest areas (fig 1b), but not for houses, which require greater spatial resolution (fig 3b, fig 3c). An isolated object measuring only a dozen metres in length is not always identifiable because its size is close to the basic pixel resolution. This leads to confusion between classes because our use of the algorithms does not always allow us to differentiate between houses and small fields, especially when the latter have no vegetation cover, even at a resolution of 2.5m (fig 3b, fig 4). Agricultural space and dwellings are often intermingled. Moreover, buildings are heterogeneous, with various sorts of roofing (tiled, cement, sheet metal, thatched) with different spectral signatures or reflectances, some of which can be mistaken for courtyards, trees or even fields embedded in the inhabited space. Plot size is very small and the various crops are at different stages of growth, which implies numerous variations in the spectral signatures. Generally speaking, it is difficult to precisely identify isolated or ungrouped settlements (those outside previously settled villages) prior to 2005 because the resolutions of the two main providers, SPOT and Landsat, used to be decametric. Due to the difficulties the students and myself encountered in
isolating houses in pixel classification on the images at a resolution of 2.5m – difficulties linked to confusion between previously mentioned classes –, we chose to conduct a photo-interpretation study (SPOT 5 images) and a vectorisation of buildings (with the disadvantage of being very time-consuming) at different dates (years 2006 and 2014), superimposed on the map of dwellings from the 1992 topographic map (fig 10). We especially see the densification phenomenon and then the sprawl of houses from the road but also from the town of Jhumka Bazaar.

Fig 10. Spatial changes in settlements in Jhumka Bazaar from 1992 to 2014

To conclude, the software used, QGIS-extension SCP and Arcgis, made it possible to obtain land-cover maps on a district scale with Landsat (30m) or SPOT (20 to 2.5m) images and thus to carry out the first identifications. However, this work needs to be followed up by more detailed analyses using specialised remote sensing software (ENVI, eCognition) on judiciously chosen and very localised sectors and on a fine scale, if one wishes to obtain a precise analysis of the housing sprawl phenomenon. However, the 2.5m resolution seems to limit extraction of the different elements of the very complex and intertwined dwellings of the rural Tarai using algorithmic remote sensing and ‘pixel classification’. Plantations and large fields are easy to identify. For buildings, Pleiades images would theoretically be more suitable but for very small sectors (villages, neighbourhoods) analysed in an object-oriented classification, which multiplies the number of analyses and images for a result on the scale of a district and limits it to a recent period (post-2011). Photo-interpretation followed by manual vectorisation on very detailed images remains an alternative – though time-consuming – to the problems of image classification using an algorithm.
BIBLIOGRAPHY


NOTES

1. The satellite is equipped with sensors that record the solar radiation reflected by objects on the Earth's surface. This radiation consists of waves of the electromagnetic spectrum. The spectral bands or channels of the spectrum are categorised according to their wavelength. The channels of the visible range of the spectrum correspond to the short waves visible to the naked eye (colours of the rainbow). The channels of the infrared range, made up of longer waves, are not visible to the naked eye.

2. The project ‘Terre Eau/Of lands and waters’ – Projet-ANR-12-AGRO-0002 (see introduction to the section).

3. I would like to thank the ISIS (Incentives for the Scientific Use of Spot Images) programme for providing several SPOT 4 and 5 images. Two images were also purchased from Airbus DS Geo SA.

4. These images were provided free of charge by the EQUIPEX GEOSUD project, which was recently supplanted by the DINAMIS system (https://dinamis.data-terra.org).
ABSTRACTS

This note addresses the use of geomatics for studying the changes in land-cover in the Tarai since the 1950s. The author explains the main principles of geomatics (for social scientists) and shows how certain geomatic methods contribute to addressing this theme. After a presentation of the inventory of sources – digitized maps and Landsat and Spot satellite images for the most part – the methodology is explained. The latter is based on the integration into a small geographic information system (GIS), of various finely reworked, georeferenced maps, both raster and vector, that ultimately show land-cover in the Tarai on various scales and at different periods.

These maps can be the result of satellite-image classifications using various remote-sensing techniques, and in particular pixel-supervised classifications used here and which identify spatial objects based on their known spectral signatures. However, while our land-cover classifications are fairly accurate at district level, they are less accurate at local level. Indeed, it is difficult to individualise and map small objects by pixel classification, such as rural dwellings in the Tarai that can be mistaken for bare or harvested fields, even on images of a very high spatial resolution. The use of manual vectorisation of these small space objects partly compensates for these inaccuracies.

Cette note concerne l’utilisation de la géomatique pour étudier les changements dans l’occupation du sol dans le Téraï depuis les années 1950. L’auteur explique les principes fondamentaux de la géomatique (à destination de chercheurs en sciences sociales) et montre l’apport de certaines méthodes géomatiques utilisées pour ce thème. Après l’inventaire des sources, cartes digitalisées et images satellite Landsat et Spot pour la plupart, la méthodologie est exposée. Cette dernière repose sur l’intégration dans un petit système d’information géographique (SIG) de cartes géoréférencées diverses, raster et vecteur, retravaillées finement qui montrent au final l’occupation du sol dans le Téraï à différentes échelles et périodes. Ces cartes peuvent être le résultat de classifications d’images satellite avec diverses techniques de télédétection, en particulier les classifications supervisées pixel utilisées ici et qui identifient les objets spatiaux grâce à leurs signatures spectrales connues. Cependant, si nos classifications montrant l’occupation du sol sont assez précises à l’échelle du district, elles le sont moins à l’échelle locale. En effet, il est difficile d’individualiser et de cartographier des objets de petite taille par classification pixel, comme les habitations rurales du Téraï qui peuvent se confondre avec des champs nus ou moissonnés, même sur des images à très haute résolution spatiale. Le recours à la vectorisation manuelle de ces petits objets spatiaux pallie en partie ces imprécisions.

INDEX

Mots-clés: géomatique, SIG, classification supervisée, Landsat, SPOT 5, Téraï

Keywords: Geomatics, GIS, supervised classification, Landsat, SPOT 5, Tarai

AUTHOR

JÉRÔME PICARD

Jérôme Picard is a cartographer trained in geographic information science and a geographer at the Center for South Asian and Himalayan Studies (CNRS/EHESS-CESAH). He analyses cultivated
and urban areas in the intertropical zone using remote sensing and geographic information systems (GIS).