Science of the Total Environment 912 (2024) 168772

Contents lists available at ScienceDirect

Science of the Total Environment

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

Producing context specific land cover and land use maps of human-modified tropical forest landscapes for infectious disease applications

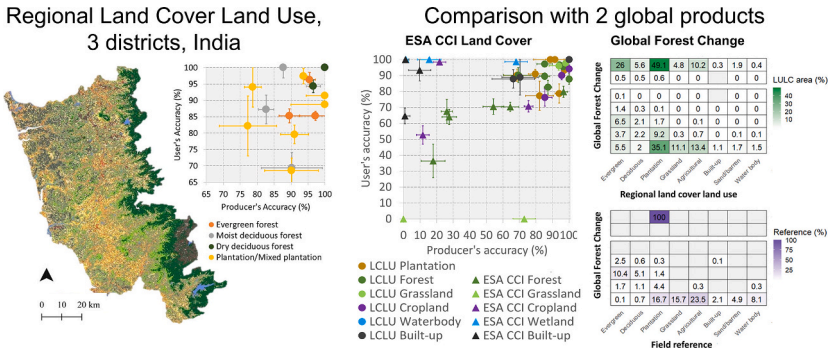
Abhishek Samrat [a,c,e], Bethan V. Purse [b], Abi Vanak [a], Anusha Chaudhary [a,d], Gowri Uday [a], Mujeeb Rahman [a], Richard Hassall [b], Charles George [b], France Gerard [b,*]

[a] Ashoka Trust for Research in Ecology and the Environment (ATREE), Srirampura, Jakkur post, 560064 Bengaluru, India
[b] UK Centre for Ecology and Hydrology (UKCEH), Maclean Building, Crowmarsh Gifford, Wallingford, Oxon OX10 8BB, UK
[c] Centre for Wildlife Studies (CWS), 37/5, Yellappa Chetty Layout, Ulsoor Road, 560064 Bengaluru, India
[d] Quantitative Disease Ecology and Conservation (QDEC) Lab Group, Department of Geography, University of Florida, Gainesville, FL, United States of America
[e] School of Engineering and Computing, University of Central Lancashire, Preston PR1 2HE, UK

HIGHLIGHTS

- Global land cover and use maps are unsuitable for zoonotic disease studies.
- We produced maps reflecting functional resource use by vectors, reservoirs and people.
- Local training and multiple map iterations ensured spatially stable and high accuracies.
- We compared our map with ESA CCI land cover and Hansen Global Forest Change.
- Regional maps represent the landscape components better than global products.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Martin Drews

Keywords:
Regional land cover
Zoonotic diseases
ESA CCI land cover
Hansen global forest change
Multiple map iterations
Western Ghats

ABSTRACT

Satellite-based land cover mapping plays an important role in understanding changes in ecosystems and biodiversity. There are global land cover products available, however for region specific studies of drivers of infectious disease patterns, these can lack the spatial and thematic detail or accuracy required to capture key ecological processes. To overcome this, we produced our own Landsat derived 30 m maps for three districts in India's Western Ghats (Wayanad, Shivamogga and Sindhudurg). The maps locate natural vegetation types, plantation types, agricultural areas, water bodies and settlements in the landscape, all relevant to functional resource use of species involved in infectious disease dynamics. The maps represent the mode of 50 classification iterations and include a spatial measure of class stability derived from these iterations. Overall accuracies for Wayanad, Shivamogga and Sindhudurg are 94.7 % (SE 1.2 %), 88.9 % (SE 1.2 %) and 88.8 % (SE 2 %) respectively. Class classification stability was high across all three districts and the individual classes that matter for defining key interfaces between human habitation, forests, crop, and plantation cultivation, were generally well separated. A

* Corresponding author.
E-mail address: ffg@ceh.ac.uk (F. Gerard).

https://doi.org/10.1016/j.scitotenv.2023.168772
Received 14 February 2023; Received in revised form 10 November 2023; Accepted 19 November 2023
Available online 24 November 2023

comparison with the 300 m global ESA CCI land cover map highlights lower ESA CCI class accuracies and the importance of increased spatial resolution when dealing with complex landscape mosaics. A comparison with the 30 m Global Forest Change product reveals an accurate mapping of forest loss and different dynamics between districts (i.e., Forests lost to Built-up versus Forests lost to Plantations), demonstrating an interesting complementarity between our maps and the % tree cover Global Forest Change product. When studying infectious disease responses to land use change in tropical forest ecosystems, we recommend using bespoke land cover/use classifications reflecting functional resource use by relevant vectors, reservoirs, and people. Alternatively, global products should be carefully validated with ground reference points representing locally relevant habitats.

1. Introduction

The association of infectious diseases with land use and landscape structure and change has long been known (Beck et al., 2000) but has been the subject of more intense focus since the late 20th century (Vanwambeke et al., 2010), particularly for ecologically complex diseases where multiple hosts and vectors are involved in transmission (Lambin et al., 2010). The mechanisms underpinning disease-landscape associations are often poorly described and involve a complex interplay between ecological dynamics of host, vectors and pathogens and social factors that determine contact rates between humans and pathogens within ecosystems (Reisen, 2010; Murray and Daszak, 2013; Gottdenker et al., 2014; Purse et al., 2020). Key tools in understanding disease-landscape associations have been remotely sensed Land Use Land Cover maps and Land Use Change products that have been widely integrated into epidemiological models of infectious disease patterns (Lambin et al., 2010; Vanwambeke et al., 2010, 2019).

The use of remote sensing imagery to map the land cover, use and forests in landscapes is well established and there are many examples of operational mapping that deliver land cover products. Their coverage varies from global (e.g., ESA CCI land cover (Buchhorn et al., 2020), GlobeLand30 (Jun et al., 2014), Global Forest Change (Hansen et al., 2013), NASA MODIS Land Cover (Friedl et al., 2002)) to national or regional land cover (e.g., for USA NLCD (Homer et al., 2012), China NLUD-C (Zhang et al., 2014) (Zhang et al., 2014), European CORINE (Büttner et al., 2004), Brazil LCLU Brazil (Instituto Brasileiro de Geografia e Estatística, 2022), India LCLU 2011–2012 (NRSC, 2012), and Australia DLCD (Lymburner et al., 2015)).

A wide range of studies have used these operationally produced land cover and land use (LCLU) maps for monitoring or predicting ecologically complex infectious disease patterns. For example, using the Harmonized Global Land Use product Chini et al. (2014) and Redding et al. (2016) related area cover of cultivated land, savanna, and conversion of savanna and grassland to croplands to the likelihood of haemorrhagic zoonotic disease (Lassa fever virus) outbreaks in Western Africa. Shah et al. (2018) used the intact forest landscapes and mangrove forests classes of the Global Forest Change product (Hansen et al., 2013) to predict the occurrence of emerging disease events linked to forests. The European CORINE LCLU was used to predict patterns in wide-ranging vector-borne diseases across Europe, including tick-borne encephalitis (Vanwambeke et al., 2010) and midge-borne diseases (Cuellar et al., 2020; Barceló et al., 2021) while global MODIS land cover was used for predicting the occurrence of wide ranging diseases including Rift Valley fever across Africa (Redding et al., 2017). However, these "off the shelf" LCLU products can prove unsuitable for several reasons which include coarse spatial resolution, low temporal resolution, unknown or low accuracy, and a lack of thematic detail (i.e., too few classes) (Kotchi et al., 2019). To model the distribution and dynamics of infectious diseases within ecosystems and to link future outbreak predictions to ecosystem management, underlying LCLU maps need to be tailored to the spatial scale at which people, wildlife, domestic animal reservoirs and vectors use the landscape. The maps also need to be tailored to the specific cover types and landscape features that act as functional resources for these species (Lambin et al., 2010; Hartemink et al., 2015).

In response, many eco-epidemiological studies have chosen to derive context specific LCLU maps by classifying finer scale imagery for a study region (Lambin et al., 2010; Kotchi et al., 2019; Vanwambeke et al., 2019). Ideally, the LCLU classes are defined a priori based on the functional resources or habitats used by the species in transmission (Hartemink et al., 2015). For example, Marston and Giraudoux (2018) derived land cover maps from Landsat OLI imagery and Sentinel-1 data to model spatial distributions of small mammal hosts and the related spread of a parasitic tapeworm among human habitation, agricultural and grassland habitats in Kyrgystan. Hardy et al. (2019) used Sentinel-1 to map vegetated and non-vegetated water body breeding habitats of the African malaria vector (*Anopheles gambiae* sensu *lato, Anopheles coustani, An. squamosus, An. ziemanni*) mosquito in Barotseland, Western Zambia. High spatial resolution SPOT imagery (2.5 m panchromatic and 10 m in multispectral mode) were used to produce LCLU maps and identify the environmental factors (distance from the forest, areal cover of forest, agriculture, human settlement, forest patch density) which drive leptospirosis incidences among humans and rodents in northern Thailand (Della Rossa et al., 2016). For the Thailand provinces of Buriram, Loei and Nan, SPOT imagery was used to produce high spatial resolution maps including the LCLU classes forest-converted agricultural land, reforestation areas and fallow area, potential habitats for *Orientia tsutsugamushi* (scrub typhus) transmission via rodents (Chaisiri et al., 2017). Common practice in such eco-epidemiological studies is to provide a single LCLU map, accompanied by accuracy metrics for the overall map and specific classes. However, as highlighted by Kotchi et al. (2019), accuracy within any given Earth observation image derivatives can spatially be highly variable and is generally not available. Providing an insight into cover type specific spatial variation in accuracy and covariance, would better inform the consequent uncertainty in disease-landscape relationships and disease pattern predictions. Lyons et al. (2018) proposed a spectral signature resampling framework for mapping spatial variation in accuracy, by generating an ensemble of LCLU classification solutions. This approach enables the production of a spatially explicit confidence map. However, such maps have not yet been widely implemented and then used in disease modelling.

Comparing the information content provided by context specific LULC maps with globally available operational LULC products, would help develop recommendations for the use of these global products for biodiversity and infectious disease applications. An ecosystem context in which this may be particularly important is tropical forest mosaics. Tropical forests play a critical role in supporting the livelihoods and economic development of large rural poor communities (Assessment, 2005) and provide a wide range of ecosystem services (Jenkins and Schaap, 2018). As the forest landscape changes, communities accessing forest services or living near forests may become exposed to infectious diseases (Campbell-Lendrum et al., 2005). Several studies have directly and widely linked spatial changes in forest cover and type, through deforestation and/or reforestation, to upsurges in human cases of infectious diseases (Fornace et al., 2019; Guégan et al., 2020; Purse et al., 2020; Morand and Lajaunie, 2021). Moreover, due to intense land use pressure and degradation, tropical agro-forestry landscapes present fine grained, spatially complex mosaics in which human habitation coincides with agriculture (including crops and plantations), primary natural forest, and mixed or regenerating secondary forests. This provides a

wide range of interface habitats in which interactions of people, wildlife and livestock hosts, vectors and pathogens can occur (Reisen, 2010; Loh et al., 2015; Das Neves, 2020; Purse et al., 2020). Thus, when investigating disease-landscape associations within rapidly changing tropical forest ecosystems using LCLU maps, it is important to understand the variable accuracy and information content of bespoke versus global LCLU products.

Here we develop context specific LCLU maps designed to capture the spatial distribution of the main land covers and uses in three Indian districts where the tick-borne viral infection Kyasanur Forest Disease (KFD) is present and in which plantation and paddy cultivation have degraded the Western Ghats tropical forests in recent decades. The work formed part of the OneHealth MonkeyFeverRisk project (www.monkeyfeverrisk.ceh.ac.uk), which aimed to understand links between human outbreaks of KFD and degraded forest landscapes and identify which landscape conditions increase the risk of spill-over to humans. Using our LCLU maps we highlighted that KFD outbreaks are linked to agro-forest habitat mosaics created as a result of forest degradation and that outbreaks were more likely in areas with high coverage of evergreen forest and plantations and lower coverage of dry deciduous forest (Purse et al., 2020). This required relevant context specific LULC maps, distinguishing land cover and use types found within these agro-forest mosaics. In particular, natural vegetation types, plantation types, agricultural areas, water bodies and settlements, which provide different resources used by humans and the small mammals, birds, primates, and ticks involved in KFD transmission (Pattnaik, 2006). We developed our LCLU class nomenclature (i.e., class names and their descriptions) with this purpose in mind. We used the spectral signature resampling framework approach of Lyons et al. (2018) to generate an ensemble of LCLU classification solutions and produce spatially explicit confidence maps. We then checked overall and class accuracy of our resulting map and carried out a class comparison with two global land cover products that are commonly used to predict disease distributions, vectors and reservoir hosts: the 300 m ESACCI land cover product (Buchhorn et al., 2020) and the 30 m Global Forest Change product (Hansen et al., 2013). The 300 m ESA CCI land cover has a spatial resolution that cannot differentiate smaller fragmented forest-agriculture-plantation classes while the 30 m Global Forest Change product is suitable in terms of its resolution but only separates land cover into forest and non-forest.

2. Methods

2.1. Focal areas: Shivamogga, Sindhudurg, and Wayanad

Our three focal districts, Shivamogga (Karnataka state), Sindhudurg (Maharashtra state) and Wayanad (Kerala state), are located in the central part of the Western Ghats, India (Fig. 1). Although there is a distinct difference in the climate and landscape between each of the districts, in all cases, the southwest monsoon drives their seasonal rainfall, with their western parts receiving more rainfall than their eastern parts. Shivamogga is the largest of the three districts (Table 1). It experiences a tropical climate. The landscape is generally flat in the east and has mountains in the west. The natural vegetation mainly consists of a gradient of Dry deciduous, Moist deciduous and Wet evergreen forest. Small patches of Shola grasslands are typically found at high elevations in the western part of the district. Sindhudurg experiences a moist and humid climate and has an undulating landscape that rises from the Arabian Sea in the west to the Western Ghats in the east (Table 1). The natural vegetation consists mainly of Semi-evergreen forests, Moist-deciduous forests, and Grasslands. Wayanad is the smallest of the three districts (Table 1). It has a tropical monsoon climate and is part of the mountainous plateau of the Western Ghats. The natural vegetation mainly consists of Dry deciduous, Moist deciduous, Wet evergreen forest and, at high elevation, Shola grassland.

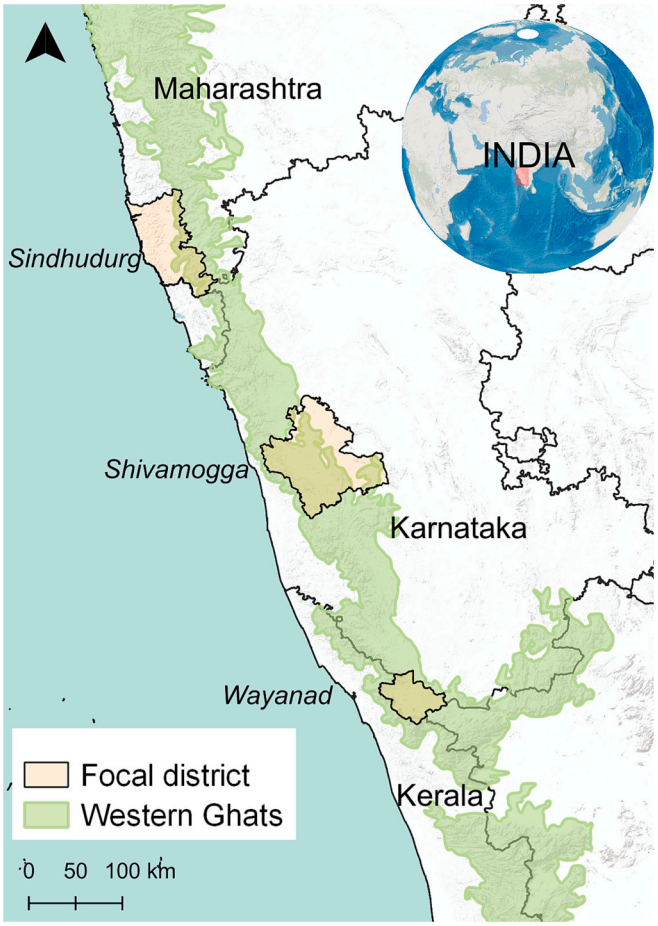


Fig. 1. Map showing the location of the three focal districts Sindhudurg, Shivamogga and Wayanad part of the Maharashtra, Karnataka, and Kerala state respectively. The administrative boundary dataset used in all figures is from SEDAC (Meiyappan et al., 2018). The background layer is from ESRI ocean and terrain.

2.2. Satellite imagery

We used Landsat 8 OLI Surface Reflectance Tier 1 imagery which are atmospherically corrected using LaSRC (Vermote et al., 2016). The correction includes a cloud, shadow, water and snow mask, produced using CFMASK (Foga et al., 2017), and a per-pixel saturation mask. We also applied an illumination correction (Tan et al., 2013) to compensate for shading caused by the topography.

Cloud cover is a major challenge for optical Earth observation in India and particularly in the Western Ghats, so the likelihood of finding images without some degree of cloud (and cloud shadow) is very low, even during the dry season. To deal with this, using an image compositing approach, we produced a median '2017' image based on layers of individual scenes with low cloud cover (< 2 % cloud), acquired during the dry season from November to February, between 2016 and 2018. Dry season imagery was used because of the lower cloud cover and also because during this period the spectral difference between deciduous and evergreen forests and other cover types was expected to be greater. Most importantly, this period coincides well with the peak season of human cases of KFD between December and March. We were particularly interested in how landscape structure affects the disease system in this part of the year. A widely accepted approach for compositing images is to apply the maximum NDVI rule (White et al., 2014). However, this approach favors observations acquired when a pixel contains a higher density of green vegetation and is therefore unsuitable for imagery taken in the dry season. Instead, we choose to use the median which produces

Table 1
Description of the three focal districts.

District	Area (km^2)	Elevation (m)	Rainfall (mm)	Temperature (°C)	Major crop
Shivamogga	8495	500–1340	1813	21–28	areca nut, paddy, cotton, ragi, jowar, maize, sugarcane, pulses, sunflower and small scale vegetables (Najeeb & Dhiman, 2012)
Sindhudurg	5207	0-700	3287	17–35	paddy, ragi, cashew nut, mango, and kokum (Kulkarni et al., 2013)
Wayanad	2132	700-2100	2322	18–29	coffee, tea, pepper, cardamom rubber, coconut, areca nut, pepper, paddy, vegetables, and tuber crops, (Personal communication, District Planning Officer, Wayanad, 2011)

imagery that better represents the time period of the original images (Flood, 2013).

Vegetation indices help discriminate between habitat types (Clerici et al., 2012; da Silva et al., 2019) and can improve image classification results, especially when combined with machine learning algorithms (Puletti et al., 2018). Therefore, EVI was calculated from the composite image and added as an extra information layer in the classification. Finally, because the natural cover types are generally associated with different elevations (i.e. Evergreen forests are found at the high elevation, Moist deciduous forest at middle elevation, and Dry deciduous forest at low elevation) and the majority of the cultivated land is found on flat land, we included the Shuttle Radar Topography Mission (SRTM) elevation data (Farr et al., 2007) as an ancillary physical band.

2.3. Land cover and use classes: nomenclature and reference polygons

The LCLU class nomenclature for the maps were developed with the purpose of capturing the spatial distribution of the main covers and uses in the landscape believed to be relevant to the spread of KFD Virus. The goal was to distinguish different natural forest types and plantation types, to map waterbodies and agricultural land and larger built-up areas such as towns. Table 2 lists the classes identified for each district, and their definitions.

We used the Open Data Kit (ODK), an android-based tool developed by Google (Hartung et al., 2010) to collect reference GPS points for the training and validation of our LCLU classification. Ideally, a spatially stratified random approach will ensure an unbiased collection of reference points, representing the relative abundance of each cover class (Stehman, 2009). However, in reality this was not achievable, mainly because of the lack of access, and difficulty in gaining permits to access nature reserves. Instead, similar to other large-scale mapping exercises (Morton et al., 2011), points were collected along roads or accessible paths (Fig. 2). For Shivamogga in particular, we were restricted to the main road arteries, because of unusual flooding during the field campaign (August 2018). In total, we collected 2588 reference points (i.e., 1038 for Shivamogga, 713 for Sindhudurg, and 837 for Wayanad). For each of these points, we checked the attributed cover class and digitized a polygon of homogeneous reflectance pixels manually (as seen on the 2017 composite image and matching google Earth imagery). Each polygon was treated as a single reference sample.

2.4. Image classification and accuracy assessment

We choose to use Support Vector Machine (SVM) (Burges, 1998) as our image classifier. The current consensus is that SVM-based image classifiers are very effective in achieving high class mapping accuracies in case of a smaller set of training sample size (Mountrakis et al., 2011). SVM uses a non-parametric machine learning technique to best fit the boundary/hyperplane, separating training samples into predefined classes, and as such only has one global optimum and a small number of tuning parameters (i.e., kernel, cost, gamma). We used visual evaluation of map results and overall mapping accuracies to optimize the SVM which involved tuning cost ($C = 10^3$) and gamma ($\gamma = 10^{-6}$) for the Radial Basis Function (RBF) kernel (Hsu et al., 2003).

In our case, because the occurrence and spread of KFD is thought to be linked to spatial landscape vegetation patterns, it was also important to produce, in addition to overall map and class accuracy, a measure of the spatial mapping accuracy (i.e., how accurate a class is likely to be mapped in a particular location). A solution proposed by (Lyons et al., 2018) is to create multiple classifications (i.e. iterations) by changing the training and validation sample set used for each classification and validation. The resulting majority of times a pixel is mapped into the same class can then be used to produce a land cover and use (LCLU) map that represents the mode solution (of the many iterations) for each pixel, with an associated measure of pixel specific classification stability.

We ran the classification 50 times, each iteration requiring training and testing. We summarized maps and accuracy metrics across sets of 10, 20, 30, 40, and 50 iterations to test the impact of iteration number on classification metrics and spatial patterns in accuracy (Table 3). The modal solution derived from each iteration set also required validation. So, our first step was to randomly allocate 20 % of the reference polygons (weighted by the proportion of reference polygons available per cover class) to an "independent" validation set used to evaluate the mode solution (Fig. 3). The remaining 80 % of reference polygons was then randomly split (again weighted by the proportion of reference polygons available per cover class) into 50 independent sample sets using a 70 % training and 30 % testing ratio. The sample split was done using the 'sample.split' method of caTools version 1.18.0 (Tuszynski, 2014) package in R (Team, 2018). For each iteration set, the mode solution was calculated by taking the per-pixel mode of the iteration's output LCLU maps. Finally, pixel specific class stability of the mode solution was calculated as the percentage of times the pixel was allocated the mode solution within an iteration set.

The classification accuracy of each mode solution was evaluated using overall accuracy (OA) and class-specific user's and producer's accuracy (Stehman, 1997) and involved a post-sampling area-weighted stratification (Stehman and Foody, 2019), building on the accuracy assessment design and analysis of Stehman and Czaplewski (1998). Following the advice of Foody (2011), Stehman and Foody (2019) and Pontius and Millones (2011), we did not include the widely used kappa coefficient. The accuracy measures were used to determine the optimum number of iterations required to achieve the highest classification accuracy for the mode solution. Image classifications and subsequent accuracy assessments were performed in Google Earth Engine (GEE) (Gorelick et al., 2017).

2.5. Comparison with ESA CCI and Global Forest Change

To evaluate how current global LCLU maps relate to our region-specific classification we compared our district maps with the 300 m spatial resolution global ESA CCI LCLU (Buchhorn et al., 2020) and the 30 m Global Forest Change v1.6 (Hansen et al., 2013) products. To simplify the comparison and improve the thematic match between classification systems, we reduced the 23 ESA CCI LCLU classes to 12 classes by combining similar classes (Supplement, Table 1) and we combined our LCLU map Evergreen forest and Semi-evergreen forest

Table 2
The districts' Land Cover and Land Use classes and their description.

Class no.	Class name	Class description	Shivamogga	Sindhudurg	Wayanad
1	Evergreen forest	*Wet Evergreen:* These forests grow in areas where the monsoon period lasts for several months and receive high annual rainfall (> 3000 mm). Evergreen trees are generally tall and straight, generally grow above 30 m height with a buttressed trunk or root on three sides like a tripod. The vertical structure is layered with shrubs in the understorey, short trees in the middle storey and tall canopy and emergent trees in the top storey. Prevalent species are *Areca catechu*, *Artocarpus hirsutus*, *Mangifera indica, Syzygium cumini*, *Vateria macrocarpa*, *Dipterocarpus bourdilloni*, *Dipterocarpus indicus*, *Myristica malabarica*, *Nageia wallichiana*, and *Palaquium ellipticum*.	X	–	X
		Evergreen & semi-evergreen: The transition zone between wet evergreen forest and moist deciduous forest. This forest is found in a region which receives annual rainfall between 2000 mm and 2500 mm. The canopy is less dense than wet evergreen forests and is predominated by evergreen tree species which are mainly in the middle storey. Prevalent species are *Antidesma menasu*, *Acrocarpus fraxinifolium*, *Artocarpus hirsutus*, *Carallia brachiata*, *Dalbergia sissoo*, *Dimocarpus longan*, *Ficus racemosa*, *Magnifera indica*, *Neolamarckia cadamba*, *Pterocarpus marsupium*, and *Terminalia paniculata.*	–	X	–
2	Moist deciduous forest	Forest dominated by deciduous tree species shedding leaves in summer but with evergreen tree species in the middle and understoreys. The top storey canopy is sparse and uneven with emergent trees exceeding 25 m that have broad and branching trunks and roots to hold them firmly to the ground. It is found in regions with annual rainfall between 1000 mm and 2000 mm. Prevalent species are H*orea robusta*, *Tectona grandis*, *Magnifera indica*, *Dalbergia sissoo*, and *Bambusa* sp.	X	X	X
3	Dry deciduous forest	Forest dominated by deciduous species. Deciduous	X	–	X

(*continued on next page*)

Table 2 (*continued*)

Class no.	Class name	Class description	Shivamogga	Sindhudurg	Wayanad
		forests are found in regions with a moderate amount of annual rainfall (< 1000 mm) that lasts for only a few months in a year. The deciduous trees shed their leaves during summer. Canopy trees do not normally exceed 20 m. Prevalent species are *Shorea robusta*, *Bambusa* sp., and *Acacia* sp.			
4	Grassland	Area dominated by high elevation grassland (Shola Forest) or low elevation grassland	X	X	X
5	Mixed plantation	A mixture of different types of plantations i.e., areca nut, coconut plantation, eucalyptus plantation, rubber plantation.	X	X	X
6	Cropland	Area with standing non woody seasonal crop	X	X	X
7	Fallow land	Area left fallow after crop harvesting	X	X	X
8	Waterbody	Open waterbody including reservoirs, lakes, ponds, rivers, and streams	X	X	X
9	Built-up	Built-up area including towns, villages, and hamlets.	X	X	X
10	Tea plantation	Area dominated by tea plantations	–	–	X
11	Teak plantation	Area dominated by teak plantations	–	–	X
12	Cashew plantation	Area dominated by cashew plantations	–	X	–
13	Mango plantation	Area dominated by mango plantations	–	X	–
14	Sand/ Barren	Sandy area or area without vegetation i.e., rocky area, sea shore.	–	X	–

X: Present in the district; –: Absent in the district.

into a single Evergreen forest class; Moist deciduous forest and Dry deciduous forest into a single Deciduous forest class; the different plantation classes into a single Plantation class; and Cropland and Fallow land into an Agricultural land class. After matching the products' projection system to our LCLU map, we used our map's pixel centroids to extract the corresponding ESA CCI cover class and GFC % tree cover values and produce correspondence matrices, showing the proportion of LCLU pixel centroids categorized as an ESA CCI class or as GFC % tree cover. In addition, we evaluated the ESA CCI map and the GFC using our complete set of LCLU in situ reference samples. We used the same combined ESA CCI and LCLU classes as for the ESA CCI/GFC – LCLU map comparisons. Here, we took the centroid location of the reference polygons to identify the corresponding ESA CCI cover class and GFC % tree cover. For ESA CCI, we calculated an area-weighted class specific accuracy (Stehman and Foody, 2019). For GFC, for both map and reference sample comparisons, we made a distinction between pixels where GFC shows tree cover loss any time prior to 2018, or tree cover gain any time up to 2015 (i.e., the most recent year for which gain is identified in GFC) and where GFC shows no tree cover loss or gain. The continuous % tree cover values (provided for 2000) were split into 0–20 %, 20–40 %, 40–60 %, 60–80 %, and 80–100 % bins.

3. Results

3.1. Accuracy

The overall classification accuracy of the LCLU maps is consistent and high (> 85 %, SE ≤ 2.1 %) for all districts and across all realization sets (Table 3). There is no significant difference in overall classification accuracy between realization sets for all three respective districts (tested using ANOVA on the normally distributed overall accuracies (Girden, 1992), see supplement Table 2a). However, the accuracy achieved for Wayanad is slightly higher than for Shivamogga and Sindhudurg (Table 3). Similarly, there is no significant difference in class specific user's accuracy and producer's accuracy between realization sets (tested using Kruskal and Wallis (1952) as user's and producer's accuracy distribution failed normality test, see Supplement Table 2b), but there is a clear spread in user's and producer's accuracy between classes, which differ between districts (supplement, Fig. 1). We decided to use 50 realizations for our final LCLU classification, as a higher number of realizations would give a more precise estimate of the mapping consistency across iterations. The resulting LCLU maps (mode of 50 realizations set) are shown in Fig. 4, their class specific user's and producer's accuracies in Fig. 5 and corresponding confusion matrices are provided in Supplement Tables 3 to 5. The map for Wayanad, the district with the least forest fragmentation, has the best results in terms of overall accuracy (94.7 %, SE 1.2 %) and class-specific accuracies (i.e., both user's accuracy and producer's accuracy are >85 % for all classes), showing minimal confusion between the forest, plantation, and cropland (Supplement, Table 5). The maps for Shivamogga and Sindhudurg have similar overall accuracies (i.e., 88.9 %, SE 1.2 %; 88.8 %, SE 2 % respectively). In Shivamogga, Cropland shows the lowest user's accuracy

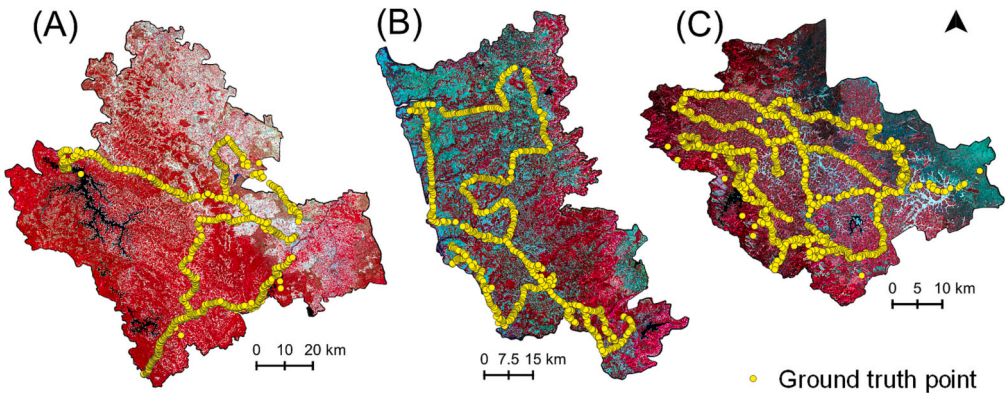


Fig. 2. Ground reference data location map for (A) Shivamogga, (B) Sindhudurg, and (C) Wayanad.

Table 3
For the resulting LCLU maps (mode of realizations), the area weighted mean (AWM) overall map accuracy (OA) and standard error (SE) for set of iterations.

Number of iterations	Shivamogga		Sindhudurg		Wayanad	
	OA (%)		OA (%)		OA (%)	
	AWM	SE	AWM	SE	AWM	SE
10	88.86	1.24	87.51	2.08	91.8	1.49
20	89.08	1.23	88.01	2.03	94.01	1.24
30	89.43	1.21	88.08	2.03	94.71	1.21
40	89.23	1.22	88.35	2.01	94.98	1.16
50	88.87	1.25	88.67	1.98	94.74	1.21

(76.2 %) and Built-up and Mixed plantation the lowest producer's accuracy (66.3 % and 79.6 % respectively) while in Sindhudurg, the least accurate classes are Mango plantation and Cashew plantation (user's accuracy:77.3 % and 78.7 % respectively) and Built-up, Mixed plantation and Moist deciduous forest (producer's accuracy: 70.4 %, 68.5 and 69.3 % respectively) (Supplement, Tables 3 and 4).

3.2. Land Cover Land Use patterns

The LCLU maps show variation in landscape patterns between the districts (Fig. 4, Table 4). Shivamogga is dominated by forest cover (Wet evergreen forest - 18 %, Moist deciduous forest - 16 %, Dry deciduous forest - 16 %) and arable land (Cropland - 11 % and Fallow land - 21 %). Across the three districts, Shivamogga also has the least proportion of plantations, which are predominantly Mixed plantations (10 %) and the highest proportion of Waterbody (4 %). Wayanad and Sindhudurg have similar high proportions of plantation cover (50 % and 46 % respectively), but the plantation types are different. Mixed plantations dominate in Wayanad while Cashew plantations dominate in Sindhudurg. The remaining land of the Wayanad is mostly forest (41 %). However, in Sindhudurg, it is divided between forests (24 %) and Grasslands (11 %). Sindhudurg and Wayanad have smaller proportions of arable land (14 % and 5 % respectively). In all three districts, human settlements, such as small towns, villages, and hamlets (i.e., Built-up) make up <2 % of the land area.

3.3. LCLU class stability

Generally, LCLU class classification stability was high across all three districts (Fig. 6). For Wayanad, 77 % of map pixels were allocated the same class at least 80 % of the time. For Shivamogga it was 74 % of pixels and for Sindhudurg it was 60 % (Table 5). Some of the spatial variations in stability reflect mapping accuracy of individual classes (see supplement Fig. 1 and supplement Tables 3 to 5). For example, across all

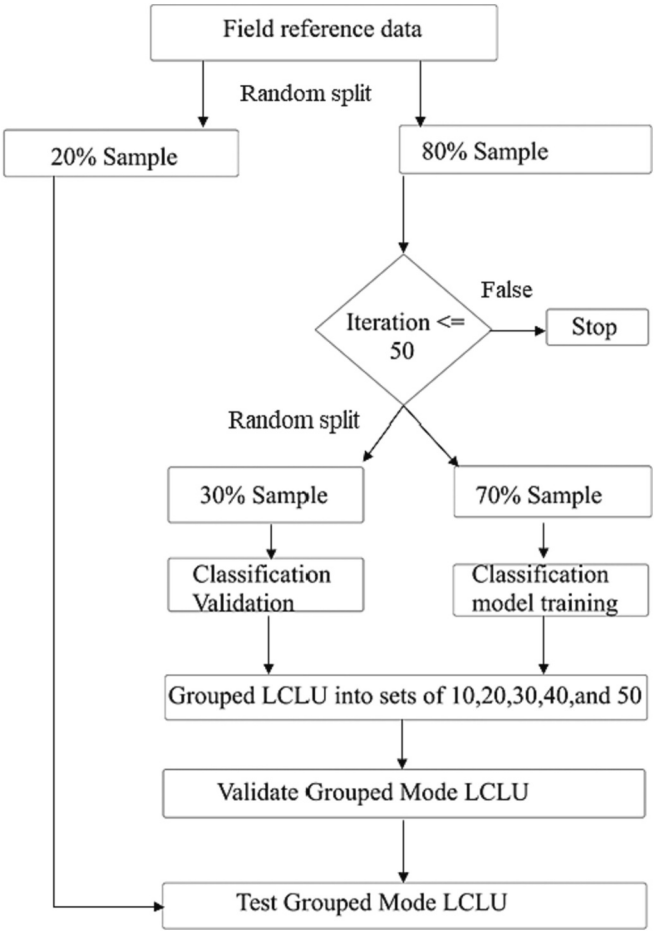


Fig. 3. Flow chart for field reference polygon processing.

districts, Waterbody shows as stable and has high user's and producer's accuracies. Also, stable areas in Sindhudurg coincide with the Evergreen & semi-evergreen forest class, which has high user's and producer's accuracies. In contrast, Cropland in Shivamogga and Cashew and Mango plantation areas in Sindhudurg show low stability and lower mapping accuracies. There are exceptions: for example, the Moist deciduous forest class, in Wayanad, has the highest commission errors (i.e., low user's accuracy), but shows high stability in the top left part of the map, suggesting that in that location the classification was likely to consistently overestimate Moist deciduous forest.

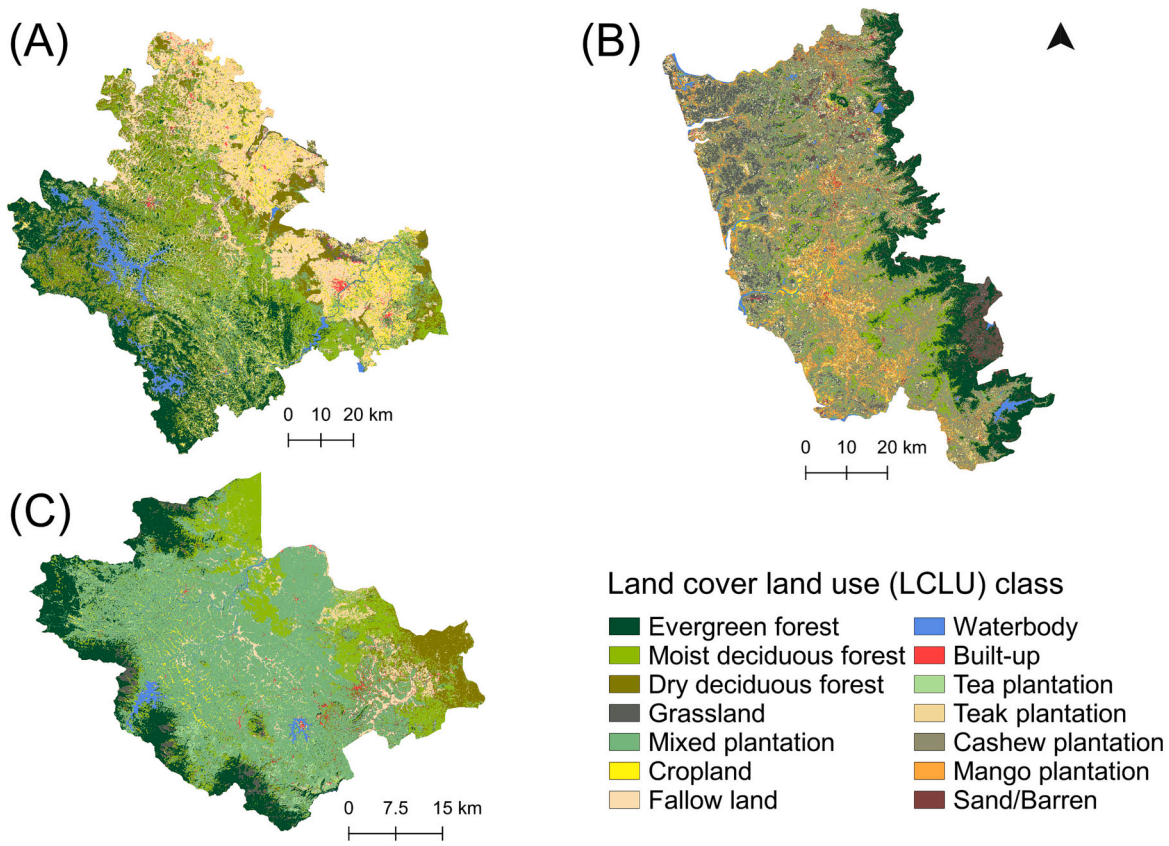


Fig. 4. The land cover land use map for (A) Shivamogga, (B) Sindhudurg, and (C) Wayanad districts, India.

3.4. LCLU classification vs ESA CCI land cover

ESA CCI has no regional specific forest and plantation classes, instead, designed to mainly cater for the global research communities, it has classes representing main plant functional types, separating woody from herbaceous, evergreen needle-leaved from deciduous broadleaved and dense from sparse woody covers. As a result, there are no clear one-to-one matches with LCLU (Fig. 7). Also, for LCLU Plantation and Grassland, there are no clear and consistent patterns of correspondence. The Plantation class was spread across multiple ESA CCI classes, but mainly the ESA CCI Crop classes (Crop/Nat Veg; Agricultural) and the Mixed tree & shrub with the remainder a mixture of Deciduous and Broadleaved Evergreen forest, and Shrubland. The LCLU forest classes and Agricultural land class show slightly clearer patterns. Evergreen forest corresponds mainly with a mixture of the ESA CCI Broadleaved Evergreen forest and Mixed tree & shrub, but also with Deciduous forest in Shivamogga and Sindhudurg, and Shrubland in Sindhudurg. Deciduous forest in Wayanad matches mainly with the ESA CCI Mixed tree & shrub class and Deciduous forest with some Evergreen forest. In Shivamogga the match is mainly with ESA CCI Mixed tree & shrub, Deciduous forest and crop classes (Crop/Nat Veg; Agricultural) and some match with Evergreen forest and Shrubland. Sindhudurg shows similar patterns as in Shivamogga but the spread between ESA CCI classes is more even. There is a good match between LCLU Agricultural land and the ESA CCI Crop classes (Crop/Nat Veg; Agricultural), for Shivamogga and Sindhudurg, while in Wayanad, where there is less Agricultural land, the class is mainly split between the ESA CCI Mixed tree & shrub and Crop classes (Crop/Nat Veg; Agricultural). In contrast, in Sindhudurg and Wayanad a large proportion of the ESA CCI Crop/Nat Veg corresponds with our LCLU Plantation class. There is a good match between the ESA CCI and LCLU Water class in Shivamogga but less so in Sindhudurg and Wayanad.

When evaluating ESA CCI with our reference data set, we only focused on ESA CCI classes for which there is a matching LCLU class (Supplement – Tables 6 to 8). We did not calculate overall accuracy because there was no overall one to one match. We found that, across the three districts, the user's accuracy of classes is consistently higher than the producer's accuracy (or class commission error is consistently lower than omission error), except for Agricultural in Shivamogga and Broadleaved Evergreen forest in Wayanad. Built-up and Wetland (which includes waterbodies) show the highest user's accuracies (> 93 %) and variable low producer's accuracies. Broadleaved Evergreen forest also shows relative high user's accuracies (i.e., above 70 %) but variable producer's accuracies (96.6 %, 64.5 %, 54.3 %). Class specific producer's accuracy varies between districts, with higher accuracies achieved by Broadleaved Evergreen forest (96.6 %) in Wayanad, Grassland (72.9 %) and Broadleaved Evergreen forest (54.3 %) in Sindhudurg, and Cropland (75.5 %), Wetland (68.0 %) and Broadleaved Evergreen forest (64.5 %) in Shivamogga. Overall, the class specific accuracies achieved by ESA CCI are lower than those achieved by our LCLU map (Fig. 8 and Supplement – Tables 6 to 8). Although ESA CCI does not map plantations separately, the reference-based correspondence matrix reveals a similar pattern as the map-based comparison: the Plantation class mostly matches with the Mixed tree & shrub class, except in Sindhudurg where it mainly matches with Crop/NatVeg. Where ESA CCI correctly identifies forest, the distinction between evergreen and deciduous is relatively good for Shivamogga and Wayanad.

3.5. LCLU classification vs Global Land Cover

Comparing GFC with our reference points shows interesting patterns (Fig. 8, panel B). In Shivamogga and Sindhudurg, a large proportion of pixels classified as forest or plantation have GFC % tree covers that are below 60 %. While in Wayanad most forest and plantation pixels have % tree covers above 60 %. Across all districts, all other (non-woody) LCLU classes are mainly within the 0–20 % GFC tree cover range. A similar

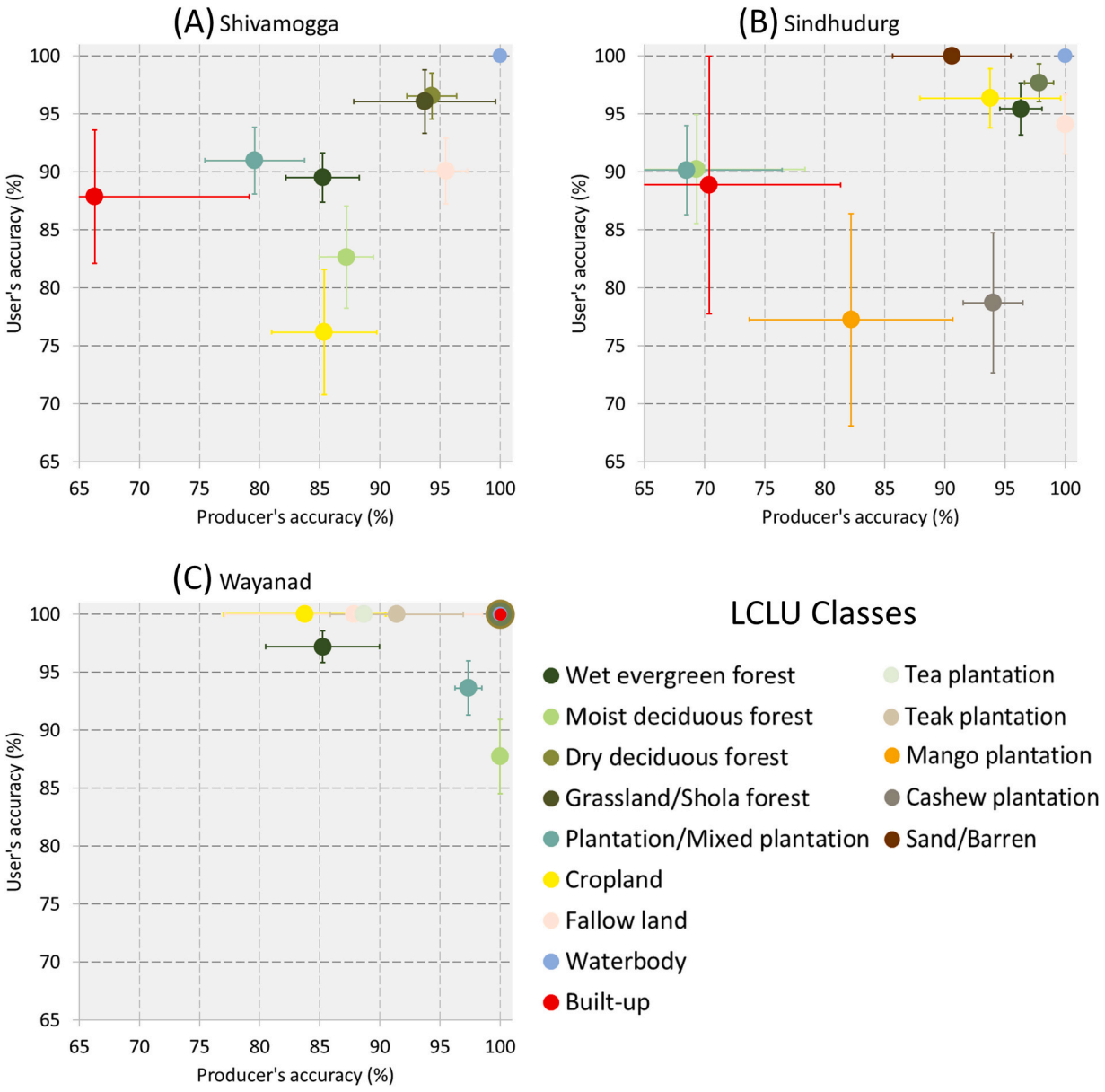


Fig. 5. Area-weighted user's and producer's LCLU class accuracies resulting from the mode of 50 realizations for (A) Shivamogga, (B) Sindhudurg, and (C) Wayanad district, India.

Table 4
Percentage of land cover land use (LCLU) class across the districts.

LCLU class	% LCLU cover		
	Shivamogga	Sindhudurg	Wayanad
Evergreen forest	18	16	15
Moist deciduous forest	16	6	15
Dry deciduous forest	16	0	7
Grassland	3	11	2
Mixed plantation	10	9	46
Cropland	11	7	2
Fallow land	21	7	3
Waterbody	4	1	1
Built-up	1	1	1
Tea plantation	0	0	3
Teak plantation	0	0	3
Cashew plantation	0	27	0
Mango plantation	0	10	0
Sand/Barren	0	4	0

pattern emerges when comparing GFC with our classification (Fig. 9, Panel A). There are some distinct differences when evaluating forest loss and forest gain pixels with our LCLU map (Fig. 9, panel A) and with our reference data points (Fig. 9, panel B). While the larger proportion of the forest loss and gain pixels are classified by the LCLU map as Plantation, Agricultural land, Built-up and Grassland a substantial number is also classified as Deciduous forest or Evergreen forest. In contrast, the forest loss and gain pixels that coincide with LCLU reference data points are nearly all labelled as non-forest LCLU classes. In Shivamogga and Sindhudurg most or all 'forest loss' pixels are identified as Plantation, while in Wayanad they are identified as Built-up.

4. Discussion

While global and national land cover products provide useful general land cover and use information, studying zoonotic diseases requires a product that provides spatially detailed information on LCLU classes that are unique to the local landscape and tailored to the functional

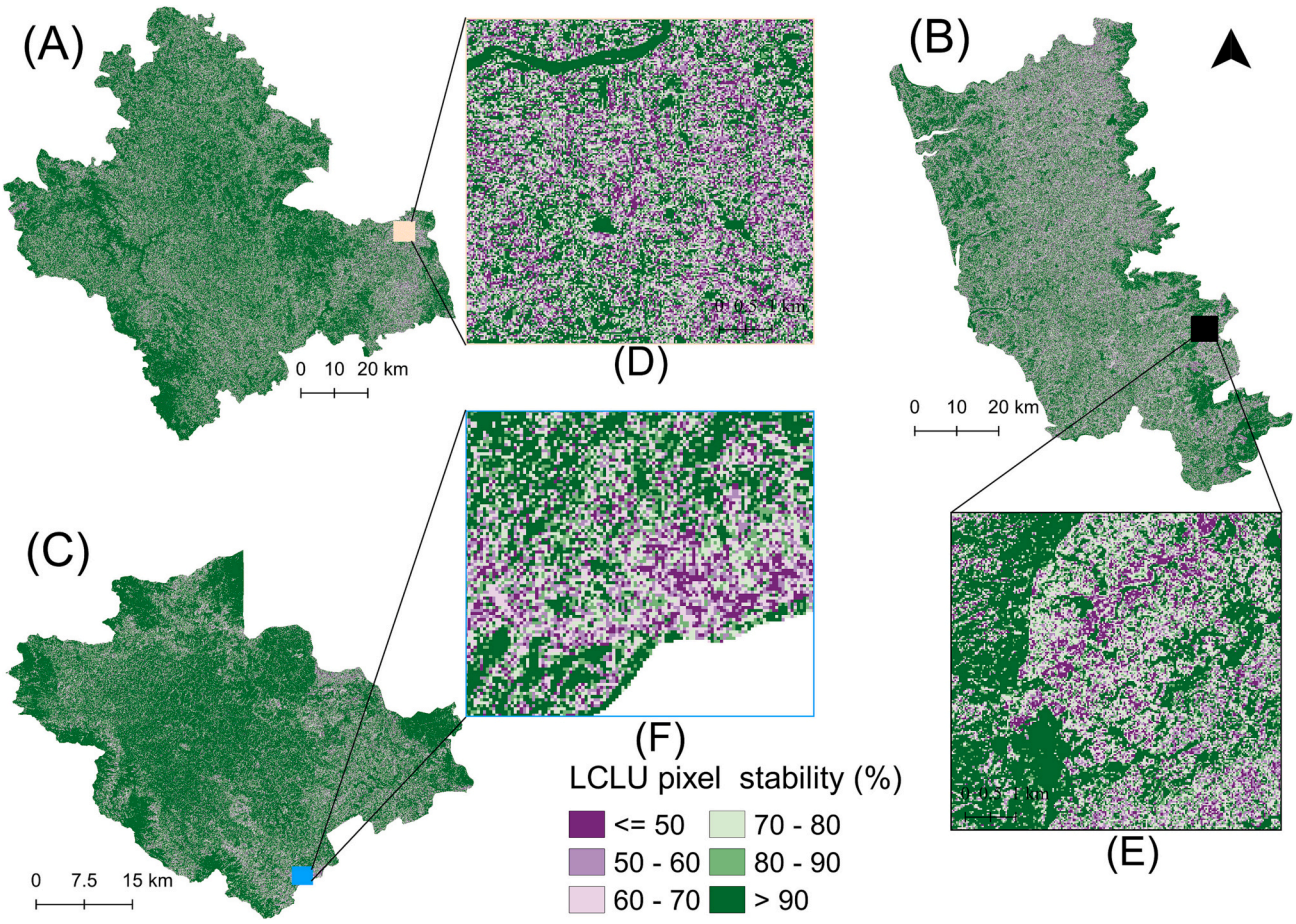


Fig. 6. Pixel specific class stability of the resulting LCLU maps (A) Shivamogga, (B) Sindhudurg, and (C) Wayanad (India). Here, stability is the percentage of times a pixel is allocated the pixel's mode solution within the iteration set of 50.

Table 5
For three districts, the % map pixels per stability range (for iteration set 50).

Stability (%)	LCLU map pixels (%)		
	Shivamogga	Sindhudurg	Wayanad
0–20	0	0	0
20–40	0.35	1	0.26
40–60	9.08	14.4	7.5
60–80	17	25.1	14.9
80–100	73.6	59.5	77.3

resource use of arthropod vectors and hosts involved in transmission (Chaisiri et al., 2017; Della Rossa et al., 2016; Figueiredo et al., 2020; Marston and Giraudoux, 2018; Hartemink et al., 2015; Vanwambeke et al., 2019). Our regionally focused mapping approach in India distinguished cover and use classes that were particularly relevant to the vertebrate hosts and tick vectors involved in KFD transmission in India (Purse et al., 2020), some of which were unique to the districts (i.e., Semi-evergreen forest, Cashew and Mango plantation in Sindhudurg, and Tea and Teak plantations in Wayanad). More importantly overall classification accuracy was high and the individual classes that matter for defining important interfaces between human habitation, forests, crop and plantation, were generally well separated (i.e., high user's accuracy / low commission error and high producer's accuracy / low omission error). This may be partly due to including elevation and NDVI in our approach which has also increased accuracies in other studies (White et al., 2014; Tan et al., 2013). However, some mapping confusion remained, particularly in districts (Shivamogga and Sindhudurg) where forest degradation has led to a more fine-grained forest-agriculture mosaic resulting in many mixed pixels. Also, broadleaved tree plantations, can spectrally be very similar to natural (primary and/or secondary) forest canopies (Fagan et al., 2015) and in our case this led to lower class accuracies. In Sindhudurg, broadleaved evergreen Cashew plantations were confused with Moist deciduous forest, because unlike the broadleaved deciduous Mango plantations, Cashew has a heterogeneous vertical canopy structure similar to Moist deciduous forest, particularly when its tree cover is low. Also, cashew, although planted in large fields, are mostly located within the Moist deciduous forest matrix. Cashew and Moist deciduous, both found at lower elevations, were easily separated from semi-evergreen forest by including elevation as a variable. In Shivamogga, the landscape is a mosaic of small farmland areas, producing seasonal crops, such as rice and vegetables, mixed among mixed tree plantations (mainly Areca nut, Coconut, Eucalyptus, and spices). This led to misclassifications between Mixed plantation and Cropland. Finally, in Wayanad, Moist deciduous forest (the least accurately mapped class) was confused with Wet evergreen forest, Mixed plantation, and Teak plantation. This is because, Moist deciduous forest is found as a transition between Wet evergreen forest and Dry deciduous forest, and Mixed plantations (i.e., a mixture of eucalyptus, rubber, black pepper, coffee, and other cash crops) are mostly found in the Wet evergreen - Moist deciduous transition zone, while Teak plantations are found in the Moist deciduous - Dry deciduous transition zone.

A successful separation between desired cover and use classes using single time period or "snapshot" remote sensing is not always guaranteed and this needs to be considered when embarking on a regional mapping exercise. To further improve mapping results, people often advocate the use of multi-temporal data to capture the seasonal variability in cover and use classes (e.g. Clerici et al., 2017; Morton et al.,

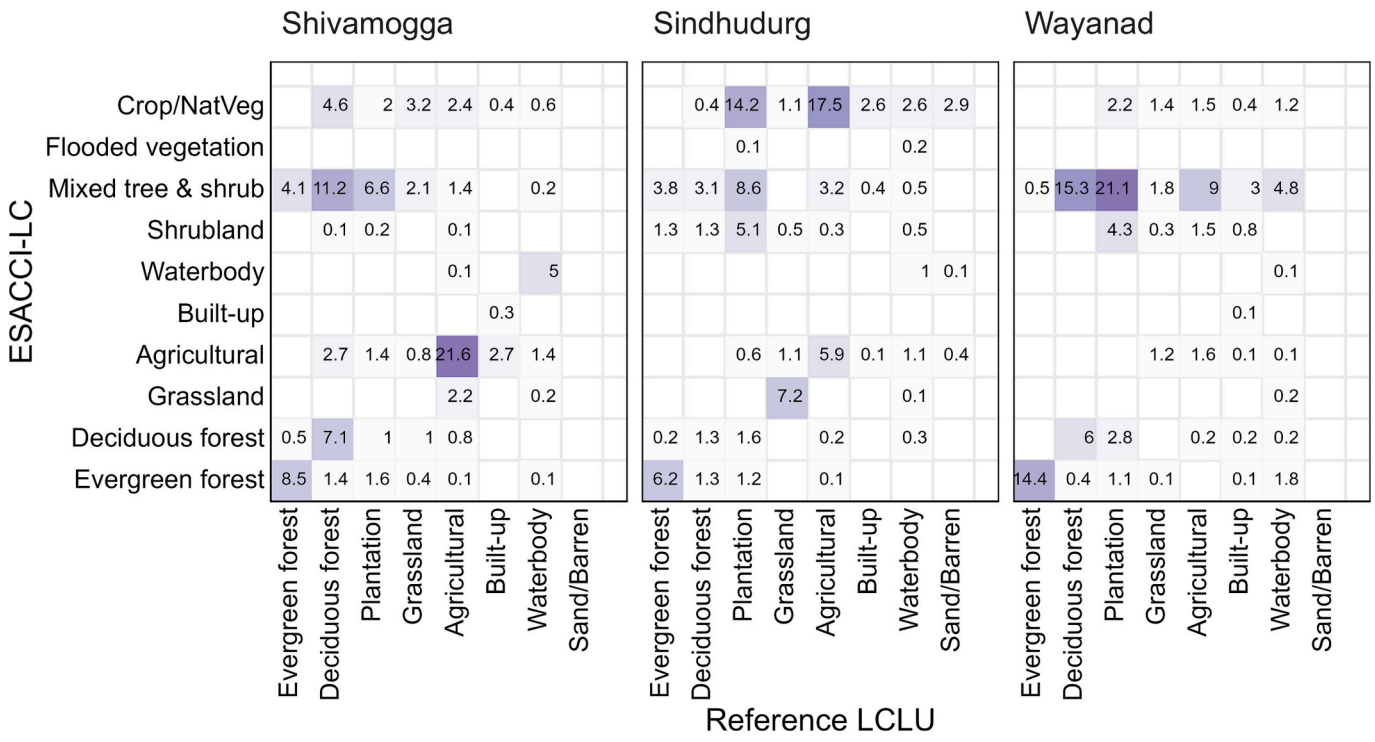


Fig. 7. Correspondence matrices for the Current LCLU class vs ESA CCI LC. The cell values were the percentage of current LCLU class pixels corresponding to ESA CCI LC classes. Cells with values >0.1 % are highlighted.

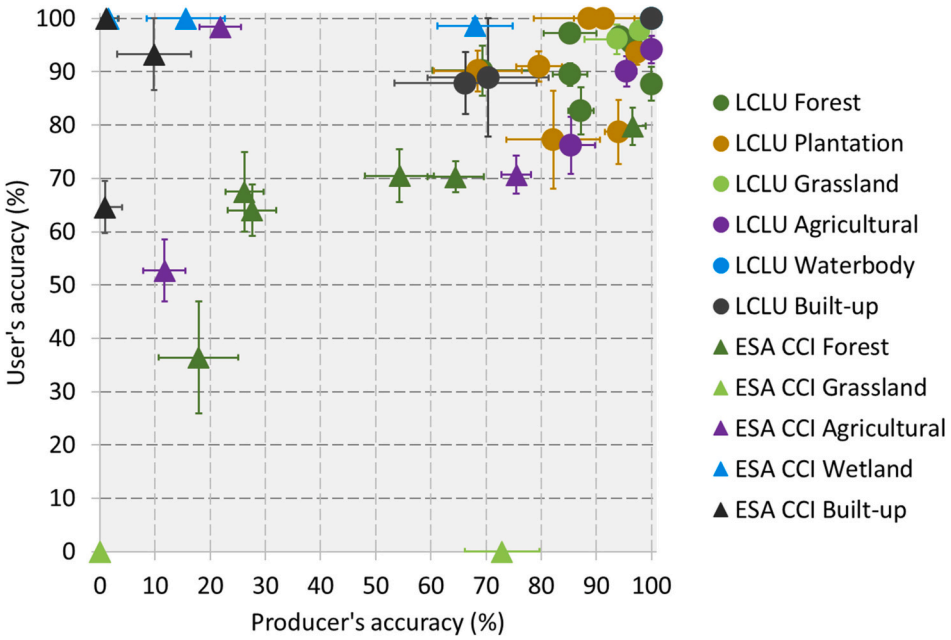


Fig. 8. User's and producer's accuracies of LCLU and ESA CCI classes (for which there is a matching LCLU class).

2011; Borges et al. 2020). This may be achieved with satellite radar since in many tropical or monsoon regions high and seasonal cloud cover limits the availability of cloud free optical data. In our case, the best option was to produce a cloud free mosaic from a multi-year post-monsoon period.

Off the shelf global land cover products do not make a distinction between woody plantations and natural forest canopies (Hansen et al., 2013; Buchhorn et al., 2020; Karra et al., 2021; Brown et al., 2022; Zanaga et al., 2022). These distinctions are most critical for the many infectious disease systems where the arthropod vectors or vertebrate reservoir hosts and/or the human activities that cause exposure are associated with specific plantation or forest types rather than forest overall. For example, human incidence of diverse vector-borne and zoonotic diseases has been linked to degradation of forests by oil palm and rubber plantations including malaria, hookworm, scrub typhus and rickettsial diseases (Shah et al., 2018). For the KFD system, outbreaks have been historically linked to replacement of moist evergreen forests by cashew plantations and landscape epidemiological analyses of recent

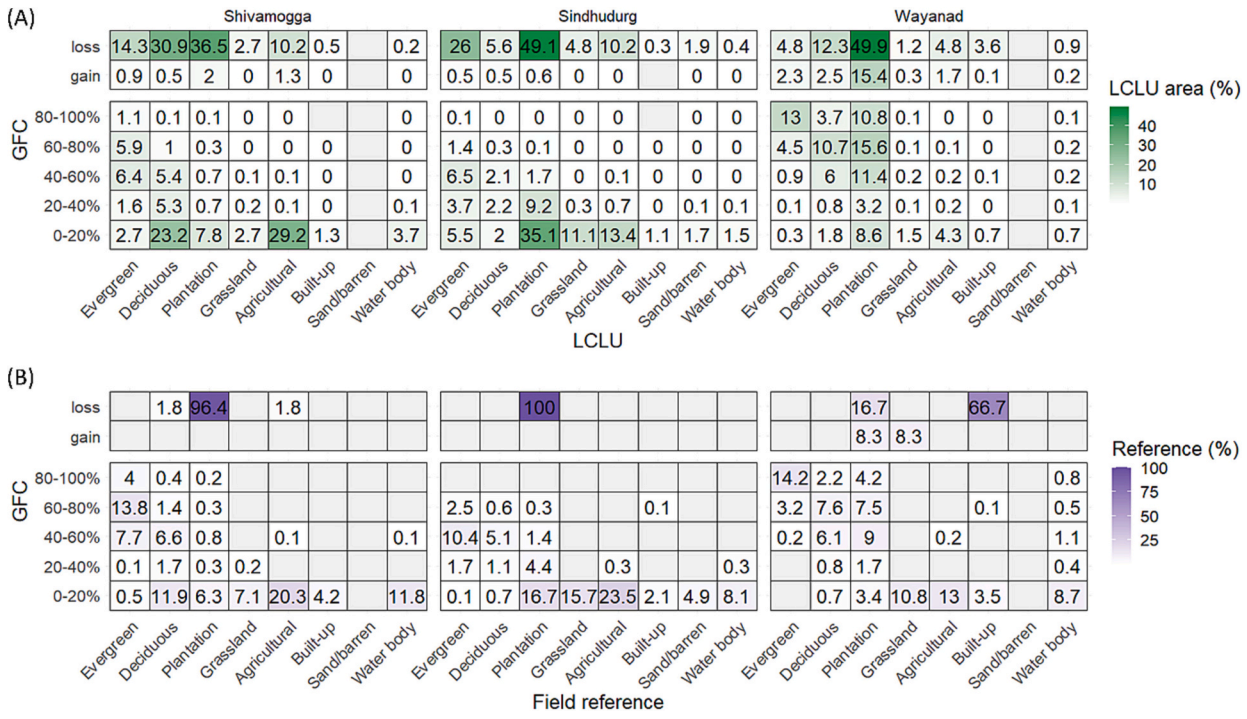


Fig. 9. (A) Correspondence matrices for GFC vs LULC pixels for each district. Top panel in A shows the comparison for pixels identified by GFC as having lost (in period up to 2017) or gained (in period up to 2015) forest cover. Bottom panel in A shows the comparison for pixels where GLC did not show a loss or gain. Both are expressed in % of total LCLU pixels. (B) Correspondence matrices for Hansen products vs LULC reference data points. Panels show the same as A but are expressed in % of reference points. Light grey cells show classes/categories with no data.

outbreak patterns have identified diverse mosaics of moist evergreen forest, plantation and paddy cultivation as being at highest risk for human outbreaks (Purse et al., 2020). In the case of the ESA CCI global land cover product (Buchhorn et al., 2020), we found that generally plantations and plantation types are classified across several broad classes representing cropland, a mosaic of tree and shrub, shrub land or forests, making it difficult to capture and map high risk forest-plantation mosaics for this disease. Designing a global landcover and use classification system that simultaneously is appropriate for many vector-borne diseases and EO-based mapping will always be a challenge, as globally, landscapes and their uses vary substantially with biogeographical region and there is a limit to the number of classes that can be distinguished using the visible and infra-red spectrum. Only by working regionally are we able to understand what is possible and so achieve the best possible outcome.

Defourny et al. (2016) validated the ESA CCI product (2010 epoch) with reference points acquired for the GlobCover 2009 product (Arino et al., 2012), showing an overall accuracy of 73.2 % with class specific user and producer accuracy ranging from 19 % to 97 % and 29 % to 95 % respectively. However, our LCLU, ESA CCI comparison reveals that its coarse spatial resolution (i.e., 300 m) does not suit the fragmented and complex mosaic landscapes found in parts of India. The lack of a clear pattern of correspondence between similar classes and in particular the higher omission errors (than commission) is partially caused by reference points falling within ESA CCI class pixels that represent a mixture of covers (e.g., tree and shrub mosaics; natural vegetation, and crop mosaics). Although, the 30 m pixel size of our LCLU map has helped resolve landscape patterns lost in the ESA CCI map, the degree of spatial detail may still be insufficient to resolve some forest type and land use classes that are key resources for hosts, vectors, and pathogens (e.g., small water bodies for mosquito breeding sites, or as sites of exposure to environmentally transmitted pathogens such as Leptospirosis). A potential solution is to use imagery with higher increasing the spatial resolution imagery such as, 10 m to 20 m of the imagery Sentinel-2 imagery (e.g. 20 m ESA CCI prototype: https://2016africalandcover20m.esrin.esa.int/; and 10 m global land cover maps of Karra et al. (2021), Brown et al. (2022) and Zanaga et al. (2022), or m resolution imagery from one of many commercial satellites (Marston and Giraudoux, 2018; Hardy et al., 2019). However, only visible and near infrared bands are available at these resolutions and a lack of shortwave infrared at higher spatial resolutions is likely to impact classification performance in terms of the thematic detail achievable (i.e., number of cover/use classes) or classification accuracy (Fagan et al., 2015). Establishing and contrasting the relative importance of thematic detail, spatial detail, and mapping accuracy, within and across infectious disease systems and different ecosystem contexts, will help in establishing guidelines for integrating future land cover mapping into eco-epidemiological studies.

Global Forest Cover provides canopy cover irrespective of the forest type and plantation types. The dense canopy Deciduous forest of Wayanad was in the range of 60–80 % canopy cover however in Shivamogga it was 0–20 % class. Most of the plantation area of the Shivamogga and Sindhudurg was in the 0–20 % canopy cover class. The comparison with GFC, using both our LCLU map and reference data points, suggests that in Wayanad the forests and plantations generally have a higher areal tree cover than in Shivamogga and Sindhudurg. Most of the Evergreen, semi-evergreen forests, and plantations of Shivamogga and Sindhudurg coincide with GCF tree covers that are <80 %, while in Wayanad they coincide with GCF tree cover that are >60 %. This makes sense as forests in Wayanad are generally denser than in Shivamogga and Sindhudurg, and plantations in Wayanad are dominated by old growth Teak and tea, both of which have high canopy covers. In Sindhudurg plantations are dominated by mixed Mango and Cashew plantation which generally are less high and sparse. The tree cover loss and gain pixel comparison with our reference data points reveals a high GCF mapping accuracy of loss and gain events. Focusing on the loss pixels, the comparison highlights how in Wayanad most of the lost forest was converted into built-up, while in the other two districts, forest was mainly replaced with cropland and plantation. The patterns are less clear when evaluating the map-to-map comparison, which is likely because of the classification

uncertainties of our LCLU map, but also because of the significantly larger number of pixels included in the comparison. Nevertheless, this LCLU - GFC map comparison highlights the complementarity of land cover, tree cover and forest loss and gain information and how their combined use can provide further insights into the infectious disease dynamics. In particular for the gain information, because no distinction is made between natural forest and plantation (Hansen et al., 2013), the provision of a land cover change product which includes plantation as a class is critical for understanding infectious disease changes associated with transitions from forest to plantation (Shah et al., 2018; Morand and Lajaunie, 2021). We avoided evaluating the forest mapping accuracy of GFC with our reference data through a confusion matrix. The GCF % cover data is only available for year 2000 with the following years showing loss and gain. Reconstructing an annual forest, non-forest layer using the 2000 % tree cover and the loss and gain data would require the choice of an area cover threshold to determine what is forest and what is not. Moreover, this type of reconstruction is not recommended by the product authors.

The class level stability (Table 5) map (Fig. 6) provides additional information about the landscape topography and land cover complexity. Regions with high landscape heterogeneity are more likely to result in low mapping accuracies (Herold et al., 2008). We found that landscape heterogeneity as well as hilly terrain also led to low local mapping consistency (particularly in Sindhudurg district).

5. Conclusions

Global EO derived LCLU products are readily and freely available and used for a wide variety of applications. However, for ecological complex infectious disease systems, such as Kyasanur Forest Disease, these products are not always suitable because of their limited spatial or thematic detail. By developing region specific LCLU maps (from Landsat imagery) with the right level of thematic detail, we were able to capture the heterogeneous and complex landscapes patterns of three districts in India (Shivamogga, Sindhudurg and Wayanad). Although we were not able to exploit multi-temporal data because of cloud cover, applying a Support Vector Machine classification on the median of post-monsoon Landsat images still delivered good results with overall accuracies exceeding 85 % and class specific accuracies varying between classes and districts but all exceeding 66 % in producer's accuracy and 76 % in user's accuracy. Deriving a map of land use class stability from the mode of many classification realizations not only delivered the most probable solution, but also highlighted where our maps were is less reliable. Comparisons with two global products (ESA CCI and GFC) highlighted that bespoke Land Use classifications can better capture local plantation, crop and forest classes and changes that may underpin changes in infectious disease systems. For other researchers who study infectious disease responses to land use change in tropical forest ecosystems we recommend that:

(1) global land use products are carefully validated with ground reference points representing locally relevant habitats;
(2) where possible a bespoke land use classification is developed, reflecting functional resource use by relevant vectors, reservoirs and people;
(3) classification stability is examined and integrated into landscape metrics used for epidemiological analyses, particularly where vectors, reservoirs and people are associated with particular forest and plantation types;
(4) experience of trade-offs in thematic detail, spatial detail and mapping accuracy are shared between Earth observation scientists, practitioners, and researchers, across infectious disease contexts.

CRediT authorship contribution statement

Abhishek Samrat: Software and method development and implementation, Field reference data collection, Data analysis and interpretation of results, Validation of maps, Data processing and curation, Writing - original draft and editing of further drafts

Bethan V. Purse: Project lead, Funding acquisition, Conceptualization, Interpretation of results, Writing - review & editing of final draft.

Abi Vanak: Funding acquisition, Conceptualization, Writing – review of drafts

Anusha Chaudhary: Field reference data collection, Writing – review of final draft

Gowri Uday: Field reference data collection, Writing – review of final draft

Mujeeb Rahman: Field reference data collection, Writing – review of final draft

Richard Hassall: Data analysis contributing to cover map comparisons, Writing – review of final draft

Charles George: Validation of maps, Writing – review of final draft

France Gerard: Conceptualization, Supervision, Field reference data collection, Interpretation of results, Writing – reviewing and editing of drafts

Declaration of competing interest

Bethan V. Purse reports financial support was provided by Natural Environment Research Council. Bethan V. Purse reports financial support was provided by UKRI Medical Research Council. Bethan V. Purse reports financial support was provided by Arts and Humanities Research Council. Bethan V. Purse reports financial support was provided by Biotechnology and Biological Sciences Research Council. Bethan V. Purse reports financial support was provided by Economic and Social Research Council. Bethan V. Purse reports financial support was provided by Global Challenges Research Fund.

Data availability

Data will be made available on request.

Acknowledgements

The MonkeyFeverRisk project that led to these results is supported by the Global Challenges Research Fund and funded by the MRC, AHRC, BBSRC, ESRC and NERC [grant numbers MR/P024335/1 and MR/P024335/2], awarded to BVP, AV and FG. Additional support was provided from the IndiaZooRisk Project, which is funded by UK Research and Innovation through the Global Challenges Research Fund [MR/T029846/1] and by the NERC SUNRISE project [grant number NE/R000131/1].

We thank government organizations for providing permission to conduct field work. Anuse, Ramesh Parab, and Sagar for helping during Sindhudurg field work. Dr. Darshan, Abhijit, M Mubashira for helping during Shivamogga field work, and M Mubashira for helping during Wayanad field work Anjan Katna, and Surya Narayanan for their comments and suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.168772.

References

Assessment ME, 2005. Ecosystems and Human Well-Being. Island press United States of America.

Arino, O., Ramos Perez, J.J., Kalogirou, V., Bontemps, S., Defourny, P., Van Bogaert, E., 2012. Global land cover map for 2009 (GlobCover 2009).

Barceló, C., Purse, B.V., Estrada, R., Lucientes, J., Miranda, M.Á., Searle, K.R., 2021 Jan 12. Environmental drivers of adult seasonality and abundance of biting midges culicoides (diptera: ceratopogonidae), bluetongue vector species in Spain. J. Med. Entomol. 58 (1), 350–364.
Beck, L.R., Lobitz, B.M., Wood, B.L., 2000. Remote sensing and human health: new sensors and new opportunities. Emerg. Infect. Dis. 6 (3), 217–227.
Borges, J., Higginbottom, T.P., Symeonakis, E., Jones, M., 2020. Sentinel-1 and sentinel-2 data for savannah land cover mapping: optimising the combination of sensors and seasons. Remote Sens. 12 (23), 3862.
Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., et al., 2022 Dec. Dynamic world, near real-time global 10 m land use land cover mapping. Sci. Data. 9 (1), 251.
Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N.-E., Herold, M., Fritz, S., 2020. Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe. https://doi.org/10.5281/zenodo.3939050.
Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. 2 (2), 121–167.
Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., Soukup, T., 2004. The CORINE land cover 2000 project. EARSeL eProceedings. 3 (3), 331–346.
Campbell-Lendrum, D., Molyneux, D., Amerasinghe, F., Davies, C., Fletcher, E., Schofield, C., et al., 2005. Ecosystems and vector-borne disease control. Ecosystems and Human well-being: PResponses 3, 353–372.
Chaisiri, K., Cosson, J.-F., Morand, S., 2017 Oct 6. Infection of rodents by Orientia tsutsugamushi, the agent of scrub typhus in relation to land use in Thailand. Trop Med. Infect. Dis. 2 (4).
Chini LP, Hurtt GC, Frolking S. LUH1: Harmonized Global Land Use for Years 1500–2100, V1. ORNL Distributed Active Archive Center. 2014.
Clerici, N., Weissteiner, C.J., Gerard, F., 2012 Jun 18. Exploring the use of MODIS NDVI-based phenology indicators for classifying Forest general habitat categories. Remote Sens. 4 (6), 1781–1803.
Clerici, N., Valbuena Calderón, C.A., Posada, J.M., 2017 Nov 30. Fusion of sentinel-1A and sentinel-2A data for land cover mapping: a case study in the lower Magdalena region. Colombia. J. Maps. 13 (2), 718–726.
Cuéllar, A.C., Kjær, L.J., Baum, A., Stockmarr, A., Skovgard, H., Nielsen, S.A., Andersson, M.G., Lindström, A., Chirico, J., Lühken, R., Steinke, S., 2020. Modelling the monthly abundance of Culicoides biting midges in nine European countries using Random Forests machine learning. Parasites & vectors 13 (1), 1–18.
da Silva, V.S., Marília, Salami G., da Silva, Isabelle Oliveira, Silva, Emanuel Araújo, Junior, José Jorge Monteiro, Alba, Elisiane, 2019. Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. Geol. Ecol. Landsc. 4, 159–169.
Das Neves CG. IPBES (2020) Workshop Report on Biodiversity and Pandemics of the Intergovernmental Platform on Biodiversity and Ecosystem Services. 2020.
Defourny, P., Achard, F., Boettcher, M., Bontemps, S., Eberenz, J., Herold, M., 2016. Global and Regional Land Cover Mapping and Characterization for Climate Modelling: Current Achievements of the Land Cover Component of the ESA Climate Change Initiative.
Della Rossa, P., Tantrakarnapa, K., Sutdan, D., Kasetsinsombat, K., Cosson, J.F., Supputamongkol, Y., et al., 2016 May. Environmental factors and public health policy associated with human and rodent infection by leptospirosis: a land cover-based study in Nan province. Thailand. Epidemiol. Infect. 144 (7), 1550–1562.
District Planning Officer, Wayanad KSPB. Kerala State Planning Board [Internet]. Kerala State Planning Board. 2011 [cited 2022 Sep 6]. Available from: https://spb.kerala.gov.in/en/wayanad.
Fagan, M., DeFries, R., Sesnie, S., Arroyo-Mora, J., Soto, C., Singh, A., et al., 2015 May 5. Mapping species composition of forests and tree plantations in northeastern Costa Rica with an integration of hyperspectral and multitemporal Landsat imagery. Remote Sens. 7 (5), 5660–5696.
Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al., 2007 May 19. The shuttle radar topography mission. Rev. Geophys. 45 (2).
Figueiredo, A.M., Valente, A.M., Fonseca, C., de Carvalho, L.M., Torres, R.T., 2020. Endoparasite diversity of the main wild ungulates in Portugal. Wildl. Biol. 2020 (1), 1–7.
Flood, N., 2013 Dec 2. Seasonal composite Landsat TM/ETM+ images using the Medoid (a multi-dimensional median). Remote Sens. 5 (12), 6481–6500.
Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., et al., 2017 Jun. Cloud detection algorithm comparison and validation for operational Landsat data products. Remote Sens. Environ. 194, 379–390.
Foody, G.M., 2011 May. Impacts of imperfect reference data on the apparent accuracy of species presence-absence models and their predictions. Glob. Ecol. Biogeogr. 20 (3), 498–508.
Fornace, K.M., Alexander, N., Abidin, T.R., Brock, P.M., Chua, T.H., Vythilingam, I., et al., 2019 Oct. Local human movement patterns and land use impact exposure to zoonotic malaria in Malaysian Borneo. eLife 22, 8.
Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., et al., 2002 Nov. Global land cover mapping from MODIS: algorithms and early results. Remote Sens. Environ. 83 (1–2), 287–302.
Girden, E., 1992. ANOVA. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America, p. 10.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017 Jul. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.
Gottdenker, N.L., Streicker, D.G., Faust, C.L., Carroll, C.R., 2014 Dec. Anthropogenic land use change and infectious diseases: a review of the evidence. Ecohealth 11 (4), 619–632.
Guégan, J.-F., Ayouba, A., Cappelle, J., de Thoisy, B., 2020. Forests and emerging infectious diseases: unleashing the beast within. Environ. Res. Lett. 15 (8), 083007.
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., et al., 2013 Nov 15. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853.
Hardy, A., Ettritch, G., Cross, D., Bunting, P., Liywalii, F., Sakala, J., et al., 2019 Mar 12. Automatic detection of open and vegetated water bodies using sentinel 1 to map african malaria vector mosquito breeding habitats. Remote Sens. 11 (5), 593.
Hartemink, N., Vanwambeke, S.O., Purse, B.V., Gilbert, M., Van Dyck, H., 2015 Nov. Towards a resource-based habitat approach for spatial modelling of vector-borne disease risks. Biol. Rev. Camb. Philos. Soc. 90 (4), 1151–1162.
Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., Borriello, G., 2010. Open data kit: Tools to build information services for developing regions. Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development - ICTD '10. New York, New York, USA: ACM Press, pp. 1–12.
Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., Schmullius, C., 2008 May. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. Remote Sens. Environ. 112 (5), 2538–2556.
Homer, C.H., Fry, J.A., Barnes, C.A., 2012. The national land cover database. US Geological Survey Fact Sheet 3020 (4), 1–4.
Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2003. A Practical Guide to Support Vector Classification.
Instituto Brasileiro de Geografia e Estatística, 2022. Monitoramento da cobertura e uso da terra do Brasil : 2018/2020. IBGE, Coordenação de Meio Ambiente, Rio de Janeiro, IBGE, p. 39.
Jenkins, M., Schaap, B., 2018. Forest Ecosystem Services.
Jun, C., Ban, Y., Li, S., 2014 Oct 23. China: open access to Earth land-cover map. Nature 514 (7523), 434.
Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021. Global land use / land cover with Sentinel 2 and deep learning. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, pp. 4704–4707.
Kotchi, S.O., Bouchard, C., Ludwig, A., Rees, E.E., Brazeau, S., 2019 May 2. Using earth observation images to inform risk assessment and mapping of climate change-related infectious diseases. Can. Commun. Dis. Rep. 45 (5), 133–142.
Kruskal, W.H., Wallis, W.A., 1952 Dec. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47 (260), 583–621.
Kulkarni J, Mehta P, Kamble C. A Study of Status, Distribution and Dynamics of Private and Community Forests in Sahyadri-Konkan Corridor of Maharashtra Western Ghats [Internet]. Wildlife Research and Conservation Society, Pune; 2013 Feb. Available from: https://indiabiodiversity.org/biodiv/content/projects/project-3222640e-6fc9-4f59-897f-ef0ad20fa300/94.pdf.
Lambin, E.F., Tran, A., Vanwambeke, S.O., Linard, C., Soti, V., 2010 Oct. Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. Int. J. Health Geogr. 27 (9), 54.
Loh, E.H., Zambrana-Torrelio, C., Olival, K.J., Bogich, T.L., Johnson, C.K., Mazet, J.A.K., et al., 2015 Jul. Targeting transmission pathways for emerging zoonotic disease surveillance and control. Vector Borne Zoonotic Dis. 15 (7), 432–437.
Lymburner, L., Tan, P., McIntyre, A., Thankappan, M., Sixsmith, J., 2015. Dynamic Land Cover Dataset Version 2.1.
Lyons, M.B., Keith, D.A., Phinn, S.R., Mason, T.J., Elith, J., 2018 Apr. A comparison of resampling methods for remote sensing classification and accuracy assessment. Remote Sens. Environ. 208, 145–153.
Marston, C., Giraudoux, P., 2018 Dec 27. On the synergistic use of optical and SAR time-series satellite data for small mammal disease host mapping. Remote Sens. 11 (1), 39.
Meiyappan, P., R, P.S., Soliman, A., Li, T., Mondal, P., Wang, S., Jain, A.K., 2018. India Village-Level Geospatial Socio-Economic Data Set: 1991, 2001. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Accessed 15-06-2021.
Morand, S., Lajaunie, C., 2021 Jul 23. Biodiversity and COVID-19: a report and a long road ahead to avoid another pandemic. One Earth 4 (7), 920–923.
Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., et al., 2011. Final report for LCM2007-the new UK land cover map. Countryside survey technical report no 11/07.
Mountrakis, G., Im, J., Ogole, C., 2011 May. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66 (3), 247–259.
Murray, K.A., Daszak, P., 2013 Feb 13. Human ecology in pathogenic landscapes: two hypotheses on how land use change drives viral emergence. Curr. Opin. Virol. 3 (1), 79–83.
Najeeb, K.M.D., Dhiman, S.C., 2012 Aug.. Ground Water Information Booklet Shimoga District, Karnataka [Internet]. Available from: http://cgwb.gov.in/District_Profile/karnataka/2012/SHIMOGA-BROCHURE%202012.pdf.
NRSC I. Manual of National Land Use/Land Cover Mapping (Second Cycle) Using Multi Temporal Satellite Data. Department of Space, Hyderabad. 2012.
Pattnaik, P., 2006 Jun. Kyasanur forest disease: an epidemiological view in India. Rev. Med. Virol. 16 (3), 151–165.
Pontius, R.G., Millones, M., 2011 Aug 10. Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int. J. Remote Sens. 32 (15), 4407–4429.
Puletti, N., Chianucci, F., Castaldi, C., 2018. Use of Sentinel-2 for forest classification in Mediterranean environments. Ann. Silvic Res. 42, 32–38.
Purse, B.V., Darshan, N., Kasabi, G.S., Gerard, F., Samrat, A., George, C., et al., 2020 Apr 7. Predicting disease risk areas through co-production of spatial models: the example of Kyasanur Forest Disease in India's forest landscapes. PLoS Negl. Trop. Dis. 14 (4), e0008179.

Redding, D.W., Moses, L.M., Cunningham, A.A., Wood, J., Jones, K.E., 2016 Jun. Environmental-mechanistic modelling of the impact of global change on human zoonotic disease emergence: a case study of Lassa fever. Methods Ecol. Evol. 7 (6), 646–655.

Redding, D.W., Tiedt, S., Lo Iacono, G., Bett, B., Jones, K.E., 2017 Jul 19. Spatial, seasonal and climatic predictive models of Rift Valley fever disease across Africa. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 372 (1725).

Reisen, W.K., 2010. Landscape epidemiology of vector-borne diseases. Annu. Rev. Entomol. 55, 461–483.

Shah, V., Shah, A., Joshi, V., 2018 May 9. Predicting the origins of next forest-based emerging infectious disease. Environ. Monit. Assess. 190 (6), 337.

Stehman, S.V., 1997 Oct. Selecting and interpreting measures of thematic classification accuracy. Remote Sens. Environ. 62 (1), 77–89.

Stehman, S.V., 2009 Sep 23. Sampling designs for accuracy assessment of land cover. Int. J. Remote Sens. 30 (20), 5243–5272.

Stehman, SV, Czaplewski, RL, Design and analysis for thematic map accuracy assessment: fundamental principles, Remote Sens. Environ., Volume 64, Issue 3, 1998, Pages 331–344, ISSN 0034-4257, https://doi.org/https://doi.org/10.1016/S0034-4257(98)00010-8.

Stehman, S.V., Foody, G.M., 2019 Sep. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199.

Tan, B., Masek, J.G., Wolfe, R., Gao, F., Huang, C., Vermote, E.F., et al., 2013 Sep. Improved forest change detection with terrain illumination corrected Landsat images. Remote Sens. Environ. 136, 469–483.

Team RC, 2018. R: A Language and Environment for Statistical Computing. Vienna. http s. www R-proje ct org, R Foundation for Statistical Computing.

Tuszynski J. caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc. R package version 2014 Apr 21;1.17.

Vanwambeke, S.O., Sumilo, D., Bormane, A., Lambin, E.F., Randolph, S.E., 2010 Jun. Landscape predictors of tick-borne encephalitis in Latvia: land cover, land use, and land ownership. Vector Borne Zoonotic Dis. 10 (5), 497–506.

Vanwambeke, S.O., Linard, C., Gilbert, M., 2019 Jun. Emerging challenges of infectious diseases as a feature of land systems. Curr. Opin. Environ. Sustain. 38, 31–36.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016 Apr 28. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185 (2), 46–56.

White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., et al., 2014 May 4. Pixel-based image compositing for large-area dense time series applications and science. Can. J. Remote. Sens. 40 (3), 192–212.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., et al., 2022. ESA WorldCover 10 m 2021 v200.

Zhang, Z., Wang, X., Zhao, X., Liu, B., Yi, L., Zuo, L., et al., 2014 Jun. A 2010 update of National Land Use/Cover Database of China at 1:100000 scale using medium spatial resolution satellite images. Remote Sens. Environ. 149, 142–154.