Primitives as building blocks for constructing land cover maps

David Saah[a,b,w], Karis Tenneson[a], Ate Poortinga[a,b,*], Quyen Nguyen[b,c], Farrukh Chishtie[b,c], Khun San Aung[b,c], Kel N. Markert[d,e], Nicholas Clinton[f], Eric R. Anderson[d,e], Peter Cutter[a,b], Joshua Goldstein[g], Ian W. Housman[g], Biplov Bhandari[b,c], Peter V. Potapov[h], Mir Matin[i], Kabir Uddin[i], Hai N. Pham[j], Nishanta Khanal[i], Sajana Maharjan[i], Walter L. Ellenberg[d,e], Birendra Bajracharya[i], Radhika Bhargava[w], Paul Maus[k], Matthew Patterson[g], Africa Ixmucane Flores-Anderson[d,e], Jeffrey Silverman[l], Chansopheaktra Sovann[m], Phuong M. Do[n], Giang V. Nguyen[o], Soukanh Bounthabandit[p], Raja Ram Aryal[v], Su Mon Myat[q], Kei Sato[r], Erik Lindquist[s], Marija Kono[t], Jeremy Broadhead[x], Peeranan Towashiraporn[b,c], David Ganz[u]

[a] Spatial Informatics Group, LLC, 2529 Yolanda Ct., Pleasanton, CA 94566, USA
[b] SERVIR-Mekong, SM Tower, 24th Floor, 979/69 Paholyothin Road, Samsen Nai Phayathai, Bangkok 10400, Thailand
[c] Asian Disaster Preparedness Center, SM Tower, 24th Floor, 979/69 Paholyothin Road, Samsen Nai Phayathai, Bangkok 10400, Thailand
[d] Earth System Science Center, The University of Alabama in Huntsville, 320 Sparkman Drive, Huntsville, AL 35805, USA
[e] SERVIR Science Coordination Office, NASA Marshall Space Flight Center, 320 Sparkman Drive, Huntsville, AL 35805, USA
[f] Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
[g] RedCastle Resources, Inc., Contractor to: USDA Forest Service Geospatial Technology and Applications Center (GTAC), Salt Lake City, UT 84119, USA
[h] Department of Geographical Sciences, University of Maryland, College Park, MD, USA
[i] SERVIR-HKH, International Centre for Integrated Mountain Development, GPO Box 3226, Kathmandu, Nepal
[j] Forest Inventory and Planning Institute (FIPI), Training and International cooperation Division, Vinh Quynh, Thanh Tri, Ha Noi, Viet Nam
[k] US Forest Service – Remote Sensing Applications Center (RSAC), Salt Lake City, UT, USA
[l] Wildlife Conservation Society, #21, Street 21 Sangkat Tonle Bassac, Phnom Penh 12000, Cambodia
[m] Royal University of Phnom Penh, Russian Federation Blvd. (110), Phnom Penh, Cambodia
[n] National Institute of Agriculture Planning and Projection, 61 Hang Chuoi, Hoan Kiem, Ha Noi, Viet Nam
[o] Space Technology Institute (Viet Nam Academy of Science and Technology), 18 Hoang Quoc Viet, Ha Noi, Viet Nam Vietnam
[p] Department of Forestry, Ministry of Agriculture and Forestry, P.O. Box 2932, Vientiane, Laos Freetown, Sierra Leone
[q] Planning and Statistics Division, Forest Department, Ministry of Natural Resources and Environmental Conservation, Nay Pyi Taw, Myanmar
[r] PASCO Thailand, 18 Khlong Toei, Bangkok, Thailand
[s] Food and Agriculture Organization of the United Nations, Rome, Italy
[t] SilvaCarbon, USA
[u] RECOFTC – The Center for People and Forests, P.O. Box 1111, Kasetsart Post Office, Bangkok 10903, Thailand
[v] Forest Research and Training Centre, P.O.Box 3339, Babarmahal, Kathmandu, Nepal
[w] Geospatial Analysis Lab, University of San Francisco, 2130 Fulton Street, San Francisco, CA 94117, USA
[x] European Forest Institute (EFI), c/o Embassy of Finland to Malaysia, Wisma Chinese Chamber, 258 Jalan Ampang, 50450 Kuala Lumpur, Malaysia

ARTICLE INFO

ABSTRACT

Land cover maps play an integral role in environmental management. However, countries and institutes encounter many challenges with producing timely, efficient, and temporally harmonized updates to their land cover maps. To address these issues we present a modular Regional Land Cover Monitoring System (RLCMS) architecture that is easily customized to create land cover products using primitive map layers. Primitive map layers are a suite of biophysical and end member maps, with land cover primitives representing the raw information needed to make decisions in a dichotomous key for land cover classification. We present best practices to create and assemble primitives from optical satellite using computing technologies, decision tree logic and Monte Carlo simulations to integrate their uncertainties. The concept is presented in the context of a regional land cover map based on a shared regional typology with 18 land cover classes agreed on by stakeholders from Cambodia, Laos PDR, Myanmar, Thailand, and Vietnam. We created annual map and uncertainty layers for the period 2000–2017. We found an overall accuracy of 94% when taking uncertainties into account. RLCMS

* Corresponding author at: Spatial Informatics Group, LLC, 2529 Yolanda Ct., Pleasanton, CA 94566, USA.
E-mail address: apoortinga@sig-gis.com (A. Poortinga).

produces consistent time series products using free long term historical Landsat and MODIS data. The customizable architecture can include a variety of sensors and machine learning algorithms to create primitives and the best suited smoothing can be applied on a primitive level. The system is transferable to all regions around the globe because of its use of publicly available global data (Landsat and MODIS) and easily adaptable architecture that allows for the incorporation of a customizable assembly logic to map different land cover typologies based on the user's landscape monitoring objectives

1. Introduction

Land cover and global land cover change are some of the most important variables affecting all aspects of society (GEOSS, 2005; Herold et al., 2008b). Timely, comprehensive, accurate information about land cover and change dynamics play an integral role in informing decision making in all sectors including policy evaluation, sustainable development, food production, water and energy security, natural hazards, ecosystem services and natural resource management (Turner et al., 2007; Bui et al., 2014; Running, 2008; Pettorelli et al., 2016; Stürck et al., 2014). Land cover is also required as a boundary condition for many biogeochemical cycling and hydrologic models (Poortinga et al., 2017). The capacity to monitor and understand changes in land cover helps countries and regional organizations identify and understand risk in order to make informed decisions on a range of issues, from land use planning, water resources management, carbon accounting, disaster risk reduction, and climate change action (Simons et al., 2017; Tolentino et al., 2016).

To address these needs, a number of research initiatives have developed global land cover products to provide information about land cover dynamics, such as the 1 km spatial resolution GLOBCOVER products (Bontemps et al., 2011), the MODIS global land cover product (Friedl et al., 2002), University of Maryland global land cover (Hansen et al., 2000), IGBP DISCover (Loveland et al., 2000), and Global Land Cover 2000 (Bartholomé and Belward, 2005). Scientists have harmonized the legends and compared these data sets and report limited compatibility between products due to differences in methodologies, sensors, and product development requirements (e.g., Giri et al., 2005; Herold et al., 2008a; Fritz et al., 2010). The 1 km spatial resolution of these products is too coarse to represent heterogeneous landscapes for some decision making objectives. An assessment of the four global products revealed poor performance in heterogeneous landscapes, specifically with mixed classes characterized by a mosaic of trees, shrubs, and herbaceous vegetation (Herold et al., 2008a). Even with information on differences between these products, there is little information on the relative utility of these different land cover products to meet specific land management objectives or evaluate performance at specific locations. Accuracy assessments have been completed at the global scale on most products (Scepan, 1999; Mayaux et al., 2006; MODIS Land Cover Team, 2003), however due to small sample sizes and associated estimates are not easily down-scaled to reflect local or regional levels.

There has been a paradigm shift in land cover science, supported by changes to data access policies that opened the archive of Landsat satellite images, new moderate resolution data are available with the launch of new satellites (e.g., Sentinel 2), improvements in time series analysis algorithms, and accessibility of free cloud computing infrastructure (Wulder et al., 2018, 2012; Poortinga et al., 2018; Yang et al., 2017b, 2013; Markert et al., 2018; Azzari and Lobell, 2017). As a result finer resolution global products are currently being produced from the Landsat archive, such as the annual global forest loss maps (Hansen et al., 2013). These provide valuable information about global trends at a finer spatial resolution. Yet because the land cover classes and associated definitions are not tailored to inform policies at national and local levels, their relative utility is still limited (Herold et al., 2006b; McCallum et al., 2006). The land cover classes vary between products and often do not align with national policy objectives, and when classes are similar the definitions and cover thresholds can still vary considerably (Giri et al., 2005; Jung et al., 2006; McCallum et al., 2006; Fritz et al., 2010).

Agencies are not inclined to adopt or harmonize the available global land cover products since their reports and work-flows are built around specific definitions which cannot easily be changed. Global land cover mapping initiatives are collaborating across a broad spectrum of thematic areas and with a variety of countries to attempt to make global data more relevant to policy and management applications (Herold et al., 2008b). Even so countries are more likely to trust and adopt data that they are generating themselves. At the same time, government and civil society organizations face challenges and roadblocks that prevent the effective use of appropriate land cover data for policy making, planning and other decision contexts (Saah et al., 2019b). A 2015 geospatial needs assessment with decision makers and technical experts from these organizations within five Lower Mekong countries, namely Cambodia, Laos PDR, Myanmar, Thailand, and Viet Nam revealed key institutional and technical impediments, including insufficient commitment to the provision of resources (staff time, computing infrastructure, etc), building of capacity, and setting up of unrealistically tight time lines for work for geospatial applications (Saah et al., 2019b).

The concept of primitives

We respond to the need for timely and accurate land cover information by making the land cover mapping and monitoring process more accessible and easily adapted to each users' specific information needs and land management objectives. Hence, we present both an approach and a system based on 'primitives' to develop and update land cover products which takes advantage of recent technological advances including publicly accessible archives of satellite data, free and cloud computing infrastructure. Our system also adopts the harmonization for data sets approach developed in the Land Cover Classification System (LCCS) (Herold and Schmullius, 2004; Herold et al., 2006a) and supports global capacity building efforts. The RLCMS method is both flexible and scalable, using biophysical definitions (e.g. forest canopy cover, tree height, percent impervious area) that are directly mappable to create land cover assessments. The biophysical information is combined with user defined definitions of those land cover types making it easily flexible for different agencies and applications, which is difficult to achieve with traditional land cover mapping methods.

Primitive map layers are a suite of biophysical and end member maps, such as canopy height and percent canopy cover. Land cover primitives represent raw information needed to make decisions in a dichotomous key for land cover typing. For example, to classify a location as forest, one may need to know the percent canopy closure. Percent canopy closure, expressed as a geographic raster dataset, is a land cover primitive (after land cover "endmembers" suggested by Gong et al. (2013)). We suggest that this approach to not only be highly suitable for remotely sensed raster data sets, but also highly flexible in terms of being able to meet needs of multiple stakeholders. As such, primitives are conceptualized as key building blocks of our system.

These primitive layers are reassembled to create a final land cover map product according to a decision logic that results in land cover classes corresponding to the desired land cover typology. The accuracy of the primitive maps and the final land cover maps will be assessed using independent validation data. Finally, we developed accuracy

assessment techniques appropriate for the novel primitive-based RLCMS assemblage structure using Monte Carlo simulation methods. In this study, we adopted best practices of satellite data processing (Young et al., 2017a) and leveraged free cloud computing technologies (Gorelick et al., 2017). We present how we constructed those primitives from our processing pipeline and how primitives are assembled into a final land cover and uncertainty map. Overall, we present primitives and the integration of their uncertainties into an agile land cover mapping system as a key innovation in this study. They serve as building blocks in the system which makes it modular and also enable explicit pixel based error quantification in a systematic manner.

To substantiate our claims, we use the Mekong region as a case study to test the system architecture and performance. We present the results of the land cover maps created from the shared regional typology agreed to during the first SERVIR regional land cover monitoring workshop with country stakeholders from Cambodia, Laos, Myanmar, Thailand, and Viet Nam. SERVIR is a unique partnership between the U.S. Agency for International Development (USAID) and the U.S. National Aeronautics and Space Agency (NASA) focused on bringing space-based technologies to environmental decision makers in developing regions. Though the RLCMS was developed for the Lower Mekong countries, the system has been modified for use in nearly any region and time period.

In the following sections we document our methods, including a description of the Mekong study area, the computational platform, system architecture, imagery and image processing, methods to create the biophysical primitive layers, the assemblage process, uncertainty propagation, and validation. Then we present the results, main findings and conclude with a discussion about the implications of this research.

2. Materials and methods

2.1. Study region

The greater Mekong region covers a 1.9 million km^2 area, has a population of 240 million people, and includes the five continental Southeast Asian countries: Cambodia, Laos, Myanmar, Thailand, and Vietnam (Fig. 1). In Vietnam, rapid urbanization has put increased pressure on land and water resource management, with implications for communities and the ecosystems upon which they rely. Cambodia has seen one of the most rapid declines in forest cover of any country since 2000. Since the end of military rule, Myanmar has seen swift transformations and development opportunities, creating challenges for land use and resource management. Lao PDR has experienced forest cover reductions due to shifting cultivation practices, commercial logging, and agriculture. In Thailand and across the region, communities live in flood- and landslide-prone areas that put people and development gains at risk. The common thread in each of these cases is the land use choices of countries in the Lower Mekong Region and their implications for sustainable development.

2.2. Methods overview

There are four phases in the creation of the regional monitoring system: defining a land cover classification typology, a supervised classification to create the primitive (or biophysical) layers, the assemblage of biophysical layers into a customized land cover map, and an accuracy assessment. These steps are outlined in Fig. 2.

An annual time series of primitives and their associated probabilities are created from optical satellite data applying machine learning supervised classification. Primitives were created based on the user defined typology. Primitives were then post processed to ensure temporal consistency. The primitives were assembled into a land cover map with uncertainty estimates with a user defined decision tree and Monte Carlo simulations. A logical transition table was used to post-process the data. The final product was validated using a stratified random sample. We describe the full process in more detail in the following sections.

2.3. Land cover typology

Defining the land cover classes is the first step in producing a land cover map. The criteria used to develop the land cover classes depend on user objectives and available monitoring resources. A classification system, or typology, should be clear, precise, and based upon objective criteria. It involves defining class boundaries even though transitions across landscapes are often gradual (Faber-Langendoen et al., 2009; FGDC, 1977). Robust land cover typologies and definitions are required for the selection of primitives and to define the assembly logic rules.

Stakeholders in the Mekong region collaboratively defined a shared land cover typology with 18 land cover classes. They used the FAO LCCS philosophy and framework (Di Gregorio et al., 2016; Di Gregorio, 2005; Di Gregorio and Jansen, 1998) and the definitions from the International GeosphereBiosphere Programme (Loveland and Belward, 1997) to systematically define land cover categories from specific observable land cover characteristics or attributes, along with their spatial and temporal relationships. The final typology includes definitions built with biophysical elements that can be mapped and assembled into a final land cover map. The classes and definitions include:

- *Aquaculture* is the farming of aquatic organisms, including fish, molluscs, crustaceans and aquatic plants. It includes man-made pond systems within fresh and salt water bodies or temporarily flooded regions.
- *Barren* areas are natural and semi-natural lands comprised of exposed soil, sand, and rocks.
- *Cropland* includes lands with herbaceous and shrubby crops followed by harvest and a bare soil period (Loveland and Belward, 1997). This category includes single, mixed, multiple, and seasonal cropping systems. Examples include cereals, oils seeds, vegetables, root crops and forages. Tea and coffee plantations are included in this layer; but orchards, forest croplands, and forest plantations are not. It also excludes irrigated or flooded rice fields and low land

Fig. 1. The greater Mekong region includes five countries: Cambodia, Lao People's Democratic Republic, Myanmar, Thailand, and Vietnam.
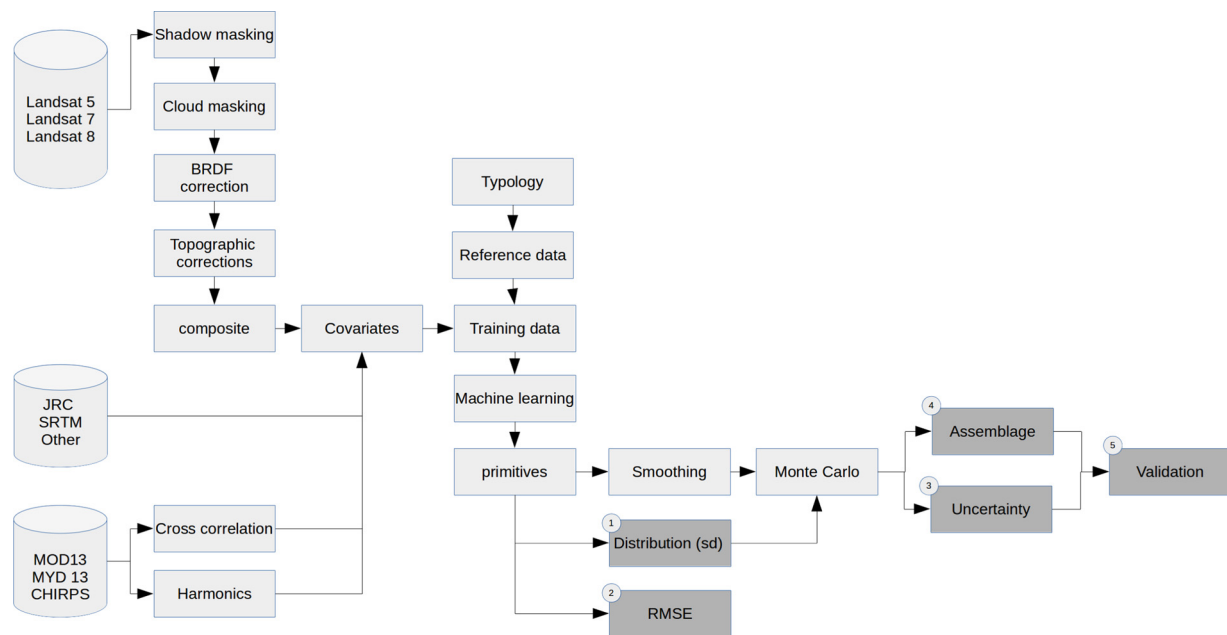
Fig. 2. Schematic overview of the overall workflow. Training data were collected based on the typology and combined with covariates in the machine learning algorithm to calculate primitives. The primitives were post processed and used in an accuracy assessment using data obtained from Collect Earth. The accuracy of the primitives was used in a Monte-Carlo simulation using a decision tree for the classification. The assemblage was post processed. The accuracy of the final product was assessed using the field data.

paddy fields where rice is intensively planted for more than 1 cycle per year.

- *Deciduous forest* lands are dominated by trees > 60% canopy cover with a tree height above 5 m. Deciduous tree species make up > 60% of the total tree cover.
- *Evergreen Broadleaf* lands are dominated by trees > 60% canopy cover with a tree above 5 m. Dominant tree species are evergreen broadleaf, and make up > 60% of the total tree cover.
- *Flooded forests* have fresh water inland habitats with > 10% tree canopy cover, a tree height above 2 m, and seasonal or permanent flooding.
- *Grassland* areas are lands with herbaceous cover, where wetland obligate species are scarce.
- *Shrublands* are lands where the majority of woody vegetation cover is less than 5 m in height and greater than 10% canopy cover. Shrub species can be evergreen or deciduous.
- *Mangroves* are coastal sediment habitats with more than 10% woody vegetation canopy cover and the majority of cover is higher than 2 m.
- *Mining lands* are comprised mostly of exposed soil, sand, or rocks originating from mining, gravel production, or other human activity.
- *Mixed forests* are have > 60% tree canopy cover, tree height is greater than 5 m, and the forest composition is mixed such that no single forest type makes up > 60% of the total tree cover.
- *Orchard and plantation forests* include lands cultivated with perennial crops that reach heights above 5 m and occupy the land for long periods (Blanchez, 1997). Commercial tree crops in the region are mainly rubber, palm oil, cashew nut, and coconut plantations.
- *Rice paddies* include irrigated or flooded rice fields and low land paddy fields where rice is intensively planted for more than 1 cycle per year (can be 2 or 3 cycles). Rice makes up the majority of vegetation cover.
- *Snow and ice* lands are under snow and/or ice cover throughout the year.
- *Surface Water* was defined as open water larger than 30 m by 30 m that is open to the sky, including fresh and saltwater (Pekel et al., 2016).

- *Urban and built-up areas* were defined as cultural lands covered by buildings, roads, and other built structures.
- *Wetlands* are seasonally flooded regions dominated by herbaceous or shrub vegetation. Wetland obligates are common.

2.4. Methods for creating primitive layers

The land cover maps are developed by combining primitives. Primitives can be extracted from the definition of each class, which are provided in the previous section. In the following sections we describe the feature space and reference data sets used to create the annual primitive layers for the Mekong product.

2.5. Landsat archive and image processing

We used the USGS Landsat 4, 5, 7, and 8 surface reflectance products in order to have a consistent time-series. The atmospherically corrected orthorectified surface reflectance data products are hosted in the Earth Engine data archive. Images from Landsat missions 4, 5 and 7 have been atmospherically corrected using LEDAPS (Masek et al., 2006; Schmidt et al., 2013; Vermote et al., 1997; Ju et al., 2012), and comes with band with a cloud, shadow, water and snow mask produced using CFMASK (Zhu and Woodcock, 2012), as well as a per-pixel saturation mask. Landsat 8 data have been atmospherically corrected using the Landsat Surface Reflectance Code (LaSRC) (Vermote et al., 2016; Holden and Woodcock, 2016; Roy et al., 2016a) and also contains the data produces by CFMASK. Landsat 7 ETM+ for after the 2003 Scan Line Corrector failure were not included in the analysis as scan line effects were found to propagate through the data analysis into the final product. Discarding Landat 7 left a gap for 2012, as Landsat 5 was decommissioned in May 2012, whereas Landsat 8 was launched in 2013. The 2012 primitives were created by temporal interpolation using the Whittaker smoothing algorithm.

Additional image pre-processing was applied since these images are subject to distortion as a result of sensor, solar, atmospheric, and topographic effects (Young et al., 2017a). In order to produce reliable and consistent time series it is important to account for these effects. We applied shadow and cloud removal, a bidirectional reflectance

distribution function (BRDF) and topographic correction. We describe each in the following sections. Then we include the calculations and derivative list we used to create the full stack of Landsat derived covariates in the supervised classification model.

2.5.1. Cloud masking

We mask clouds using the pixel-qa band and a cloudScore algorithm. The cloudScore algorithm uses the spectral and thermal properties of clouds to identify and remove pixels with cloud cover from the imagery. The algorithm identifies pixels that are bright and cold, then compares to the spectral properties of snow. The snowscore was also calculated using the Normalized Difference Snow Index (NDSI) to prevent snow from being masked. The algorithm calculates scaled cloud scores for the blue, all visible, near-infrared and short-wave infrared bands and then takes the minimum. The algorithm was described by Chastain et al. (2019).

2.5.2. Cloud shadow masking

To remove cloud shadows, we used the Temporal Dark Outlier Mask (TDOM) algorithm (Housman et al., 2018) which identifies pixels that are dark in the infrared bands but are found to not always be dark in past and/or future observations. This is done by finding statistical outliers with respect to the sum of the infrared bands. Next, dark pixels were identified by using the sum of the infrared bands (NIR, SWIR1, and SWIR2). The pixel quality attributes generated from the CFMASK algorithm (pixel-qa band) was also used for shadow masking.

2.5.3. BRDF correction

The nadir view angles of the Landsat satellites cause directional reflection on the surface which can be described by the bidirectional reflectance distribution function (BRDF) (Roy et al., 2008, 2017; Gao et al., 2014; Lucht et al., 2000). We applied the method of Roy et al. (2016b) to all images in the image collection.

2.5.4. Topographic correction

Topographic correction is a radiometric process to account for illumination effects from slope, aspect, and elevation that can cause variations in reflectance values for similar features with different terrain positions (Colby, 1991; Riano et al., 2003; Shepherd and Dymond, 2003). We applied the Modified Sun-Canopy-Sensor Topographic Correction method as described by Soenen et al. (2005). The algorithm combines the sun-canopy-sensor (SCS) (Gu and Gillespie, 1998) with a semiempirical moderator (C) to account for diffuse radiation (Justice et al., 1981; Smith et al., 1980; Teillet et al., 1982). The model contains physically based corrections that preserve the sun-canopy sensor geometry (SCS, SCS + C) while adjusting for terrain.

2.5.5. Landsat derivatives and annual statistics

The medoid (Flood, 2013) and standard deviation were calculated for the blue, green, red, nir, swir1, and swir2 bands. We also calculated the standard deviation of the normalized difference (Eq. (1); Angiuli and Trianni, 2014) for the nir and swir2, green and swir1, and nir and red. To represent yearly variability, the medoid of the 20th and 80th percentile were also included in the yearly composite.

The bands in the composites were used to calculate a series of covariates. Table Table A.1 provides an overview of covariates used in the study including the spatial resolution. The letters ND indicate that the normalized difference between the first and second bands were calculated, p20 and p80 refer to the 20th and 80th percentile respectively. For some combinations there are more common names such as Normalized Difference Water Index (NDWI, McFeeters, 1996), Normalized Burn Ratio (NBR, Key and Benson, 1999), Normalized Difference Snow Index (NDSI, Salomonson and Appel, 2004) and Normalized Difference Vegetation Index (NDVI, Rouse et al., 1974). The source or calculation for the other covariates are described in the next paragraphs.

$$ND(band1, band2) = \frac{band1 - band2}{band1 + band2} \quad (1)$$

Two ratio (R) bands were included; these are calculated by the division of two bands. This was done for the swir1 and nir and red and swir1. We also included the Enhanced Vegetation Index (EVI, Eq. (2)) (Jiang et al., 2008) and the Soil-adjusted vegetation index (SAVI, Eq. (3) using L = 0.5) (Huete, 1988). The Index-based Built-Up Index (IBI) (Xu, 2008) was calculated using Eq. (4).

$$EVI = 2.5*\frac{NIR - red}{NIR+6*red - 7.5*blue + 1} \quad (2)$$

$$SAVI = \frac{(1 + L)(NIR - red)}{NIR+red + L} \quad (3)$$

$$IBI = \frac{NIR}{(NIR + red)} + \frac{green}{(green + swir1)} \quad (4)$$

The composites were used to calculated the Tasseled Cap transformation derivatives, using coefficients from Crist and Cicone (1984), resulting in indices representing "brightness", "greenness", "wetness", "fourth", "fifth", and "sixth". The Tasseled Cap angles (tcAngle) and distances (tcDist) were calculated for all pairs of brightness, greenness, and wetness using Eqs. (5 and 6) (Powell et al., 2010).

$$tcAngle(band_1, band_2) = atan2(\frac{band_1}{band_2}) \quad (5)$$

$$tcDist(band_1, band_2) = \sqrt{band_1^2 + band_2^2} \quad (6)$$

2.6. Tree canopy cover and height

Yearly tailor made products mapping fractional tree canopy cover (TCC) and tree canopy height (TCH) derived from summary statistics of annual Landsat surface reflectance products and global sub-pixel training data (Hansen et al., 2011, 2016) were included as covariates to map the primitive layers. The processing for the creation of these annual products includes temporal smoothing using linear regression and median filters for inter-annual variation.

2.7. Water persistence metrics

The JRC Global Surface Water Mapping dataset contains maps of the location and temporal distribution of surface water from 1984 to 2015 and provides statistics on the extent and change of those water surfaces (Pekel et al., 2016). The mapping layers product consists of 1 image containing 6 bands (Table 1). It maps different facets of the spatial and temporal distribution of surface water over the last 32 years.

2.8. Cross correlation between rainfall and vegetation greenness using MODIS and CHIRPS data

In contrast with natural areas, agricultural fields are often supplemented with irrigation water to ensure crop growth and optimize

Table 1
The JRC Global Surface Water Mapping dataset contains 6 layers with information on different facets of the spatial and temporal distribution of surface water over the last 32 years.

Band	Description
occurrence	The frequency with which water was present
change_abs	Absolute change in occurrence
change_norm	Normalized change in occurrence
seasonality	Number of months water is present
transition	Categorical classification of change between first and last year
max_extent	Binary image containing 1 anywhere water has ever been detected

yields. Irrigation practices occur throughout the year, but is most important during the dry season. Intensive agricultural systems – that produce two or three crops a year – can be found throughout the Mekong region depending on water availability. Areas that remain wet in the dry season are generally greener than the surrounding natural areas. For natural areas, a high cross-correlation can be expected between rainfall and greenness. For irrigated agricultural areas this correlation is expected to be less pronounced as these areas remain green with no rainfall. As such the cross-correlation of EVI MODIS products (MOD13Q1 and MYD13Q1) (Huete et al., 2002) and precipitation was calculated. We used the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) rainfall product, which is a 30 + year quasi-global rainfall dataset (Funk et al., 2015).

The cross-correlation was calculated in three steps. First, linear regression was applied to the EVI timeseries. Regression coefficients were used to create a de-trended time-series. This time-series was combined with CHIRPs. A lag of 30 days was used to account for the delayed response of vegetation to rainfall. Cross-correlation was calculated using eq. (7), with the EVI pixel value at time, t, and the rainfall of time $t − 1$. It can be noted that the numerator of Eq. (7) is an expression of the cross-covariance. The data was added as a band to the image stack of covariates.

$$\text{crossCorrelation} = \frac{E\left[(\text{EVI}_t − \text{EVI}[p_t])(p_{tl} − E[p_{t−l}]))\right]}{\sigma(\text{EVI})\sigma(p_{t−1})} \tag{7}$$

2.9. Representing vegetation phenology using harmonic trend analysis on MODIS data

Phenology is another important metric that can be calculated and used to distinguish different types of vegetation. Different phenology stages of croplands and plant communities can be linked with trends in EVI values. For example, low EVI values are associated with harvest and planting when the soil is bare. High EVI values represent peak growing season.

Harmonic trend analysis was applied to the MODIS EVI time series products (MOD13Q1 and MYD13Q1) (Shumway and Stoffer, 2011). We applied one, two, and three harmonic terms. The latter is shown in eq. (8). Eq. (8) has a linear component (β_0 and β_1) and six harmonic coefficients (with β_2–β_6). For each harmonic, the coefficient of determination (R^2) was calculated. Higher R^2 values indicate a better fit between measured and modeled and help to identify areas with 1, 2 and

3 cropping cycles. The coefficient of determination for the different harmonics were added as bands to the image.

$$\text{EVI} = \beta_0 + \beta_1 t + \beta_2 \cos(2\pi t) + \beta_3 \sin(2\pi t)$$
$$+ \beta_4 \cos(4\pi t) + \beta_5 \sin(4\pi t)$$
$$+ \beta_5 \cos(6\pi t) + \beta_6 \sin(6\pi t) \tag{8}$$

2.10. Terrain indices

We computed terrain properties to be used as inputs in creating primitives. The five terrain properties include elevation, slope, aspect, and two aspect derivatives. These include a measure of the deviation from east (the sine of aspect) and deviation from north (the cosine of the aspect) (Trimble and Weitzman, 1956; Beers et al., 1966). All terrain indices were derived from the digital elevation data from the 1-arc-second (approximately 30 m ground resolution) SRTM dataset (Farr et al., 2007).

2.11. Ancillary data sources

Numerous studies have demonstrated the improvement of land cover mapping results with the use of auxiliary information in the classification process (Zhu et al., 2016; Khatami et al., 2016; Franklin and Wulder, 2002). Therefore we included a suite of ancillary data set in addition to the information derived from optical and elevation imagery. These included distance to coast, distance to roads, distance to buildings and ecoregions (Olson et al., 2001). The road and building maps were created from open street map data (OpenStreetMap Contributors, 2017).

3. Reference data

We trained the primitive models with reference data that was provided by our colleagues. Reference data included a combination of observations that were recorded in the field and using high resolution imagery and time series information. In total we used over 50,000 data points. The temporal distribution of the reference datapoints in each class is specified in Fig. 3. The spatial distribution of points is shown in Fig. 4.

The reference data of the RLCMS system was collated from our national partners and additional data was generated from their land cover products and high resolution imagery. We compiled over 3000
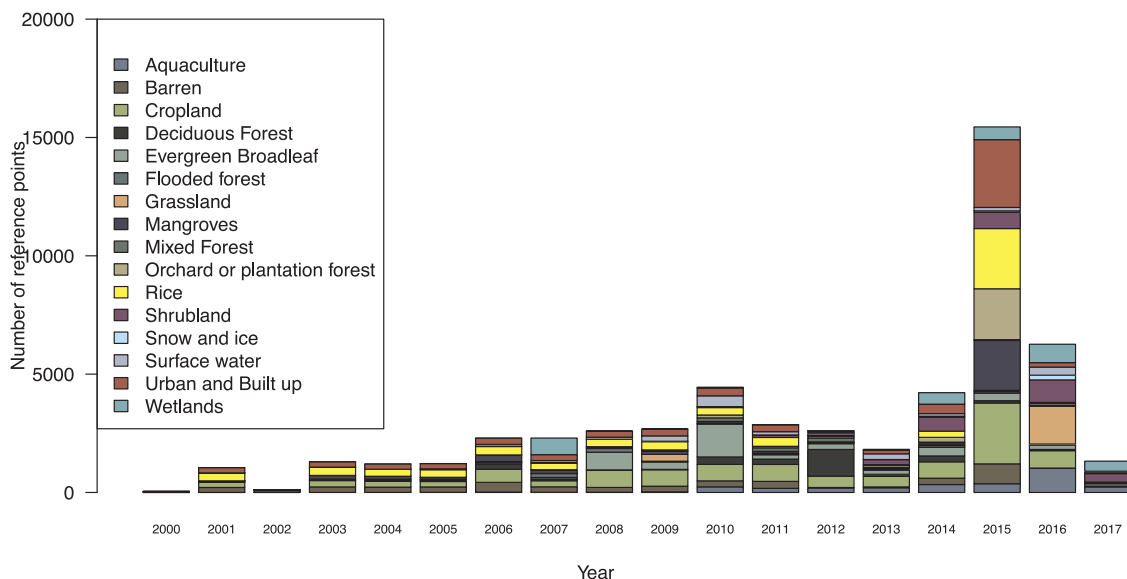

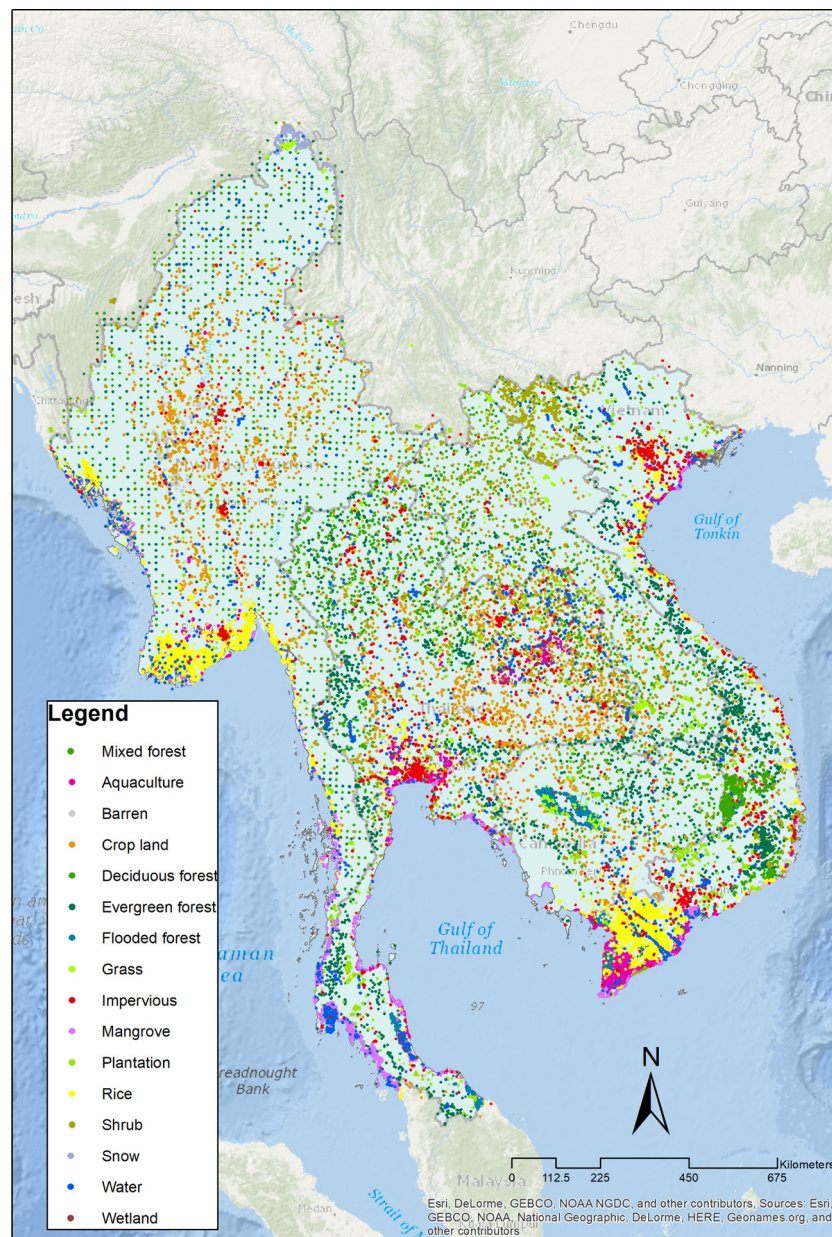**Fig. 3.** Number of reference data points per year and class.

Fig. 4. Spatial distribution of reference data points per land cover class.

points from our partners data collection efforts for the Global Forest Resources Assessment (GFRA). This conducted by the Food and Agriculture Organization (FAO) over period of 2010–2016. We also used data collected by the Forest Department of Myanmar and Forest Inventory Planning Division of Lao PDR; these were photo-interpreted plots that were labeled using VHR image in period 2000–2016 in the Collect Earth desktop application. Space Technology Institute, Viet Nam, and WCS, Cambodia, colleagues shared their field data collected at different sites.

Finally, we augmented these data sets by collecting additional data with our stakeholder colleagues using the following methodology. We placed random points within land cover map products (period of 2008-2012) from the five countries in the Mekong region. All the random points were placed within 200 m from the edge of each land cover patch. The land cover type of each point were verified using Google Earth with date associated. We also added supplemental points in land covers with low coverage and in areas of high model uncertainty using imagery from Google Earth.

3.1. Machine learning

Each supervised classification was set up to predict the primitive class of interest. The other land cover classes were aggregated into a absence class. For example, for the barren primitive we had reference data that was assigned a label as either barren or not-barren. Another random training sample of about equal size was created from the other reference data.

At each point all the potential covariates, the full imagery stack, was sampled and the coincident values were evaluated in R (R Core Team, 2018; Liaw and Wiener, 2002; Breiman, 2001). The list of possible covariates is quite large and in some cases provides redundant information. Selecting a smaller subset reduces the computational expense of applying the model across the landscape in GEE. Therefore we selected a subset of model inputs from the full imagery stack by eliminating covariates that provided little gains in accuracy. We ran a principal components analysis on two measures of variable importance, Mean Decrease Accuracy and Mean Decrease Gini. Mean decrease in

accuracy is the number or proportion of observations that are incorrectly classified by removing the feature (or values from the feature) in question from the random forest model. The mean decrease in Gini is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest. Variables that result in nodes with higher purity have a higher decrease in Gini coefficient. Neither of these metrics provide information about the model performance, but rather indicate the predictive power, or importance, of each metric included in the feature space. Finally the covariates with a primary component analysis coefficient above zero were selected.

Once the subset of covariates was finalized we applied the random forest classifier in GEE using the selected image bands. The model was run in probability mode with 100 trees. The random forest model was applied to each class (class versus other) for every year. The output of the random forest model (primitive) is the probability map for the class.

3.2. Temporal smoothing of primitive layers

Temporal smoothing was applied to the primitives for the period 2000–2017 to ensure a consistent time-series. We applied the Whittaker smoothing algorithm (Whittaker, 1922; Eilers, 2003). The method is based on penalized least squares and has been successfully applied in smoothing remote sensing time series (Atzberger and Eilers, 2011a,b; Zuliana and Perperoglou, 2017; Atkinson et al., 2012). Outputs of the smoothing algorithm include a time series of smoothed probabilities and a map of the Root Mean Square error (RMSE) of the fitted curve versus the original data.

3.3. Assemblage

Finally, we combined all the primitive data layers into a final land cover map using an assemblage logic. The assemblage is a method for combining different primitive layers into a land cover map, while incorporating users' land cover definitions and priorities – while simultaneously preserving uncertainty information. We use a hierarchical, decision tree structure (dichotomous key) to prioritize the integration from the primitive layers. We propagate uncertainty along the decision tree by running a series of Monte Carlo simulations. During each iteration of the simulation primitive layers are randomly generated according to the accuracy of the primitive layer. The simulated values are then passed through the user-defined decision tree to generate a series of land cover predictions. These predictions are aggregated to produce a final land cover map. We elaborate on these details in the sections below.

In comparison with alternative probabilistic classification methodologies, such as Bayesian networks or fuzzy logic, Monte Carlo sampling over deterministic decision trees enables end-users to construct bifurcating decision trees by posing yes/no land cover-related questions. Because the primitives are probabilistic, sampling from them independently within the group-designed decision trees and collecting aggregate statistics enables retention of as much of uncertainty in the input data in the final products as possible. Although a Bayesian network approach may provide a clearer uncertainty propagation scheme to those familiar with probabilistic mathematics, the Monte Carlo methodology is considered more easily and widely understood and therefore more appropriate.

3.4. User defined rule set

The assemblage is constructed from the user defined rule set. We constructed a decision tree to generate the final 17 class typology. Placement near the top of the tree are more likely to be classified, while primitives lower in the decision tree are more likely to be masked. The tree was designed based on the priority of the land cover types but also the accuracy was taken into consideration.

3.5. Uncertainty propagation with Monte Carlo simulations

We propagate uncertainty through the assemblage process using information on the standard deviation of each primitive. Each iteration of the simulated primitive values are then passed through the user-defined decision tree to generate a series of land cover predictions. These predictions are aggregated to produce a final land cover map with associated probabilities for each land cover category specified in the assemblage.

We start with a random raster with values between -1 and 1. Each random pixel value is multiplied by the standard deviation of its respective primitive layer to simulate a value that takes into account a random error value, the error term associated with any modeling process. This value is then added to the respective probability primitive layer. Once we have a simulated value for each primitive, we run these through the assembly logic rule set. The resulting land cover value is saved at each iteration. This is used to generate a probability layer for each land cover end type by summing the final count of each land cover class and dividing it by the number of iterations. The map land cover is the mode of the simulated values, and the probability layers represent the assemblage uncertainty information.

3.6. Post-processing

A temporal categorical smoothing algorithm was applied that incorporated rules restricting land cover transitions based on knowledge

Table 2

Transitions between classes. A 0 indicates an illogical transition whereas transitions with a 1 are permitted.

| Class	A	B	C	D	E	F	G	Ma	Mi	MF	O	R	SL	SI	SW	U	W
Aquaculture	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Barren	1	1	1	0	0	0	1	0	1	0	0	1	1	1	1	1	1
Cropland	1	1	1	0	0	0	1	0	0	0	1	1	1	0	1	1	0
Deciduous Forest	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	0
Evergreen Broadleaf	0	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	0
Flooded forest	1	1	1	0	0	1	1	1	0	1	1	1	1	0	1	1	1
Grassland	1	1	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1
Mangroves	1	1	1	0	0	0	1	1	0	0	0	1	1	0	1	1	1
Mining	0	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0
Mixed Forest	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Orchard or plantation	1	1	1	0	0	0	1	0	0	0	1	1	1	1	1	1	1
Rice	1	1	1	0	0	0	1	0	0	0	1	1	1	1	1	1	1
Shrubland	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Snow and Ice	0	1	0	1	1	1	1	0	0	1	1	0	1	1	1	0	1
Surface Water	1	1	1	0	0	0	1	1	1	0	0	1	1	1	1	1	1
Urban and built up	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Wetlands	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	1
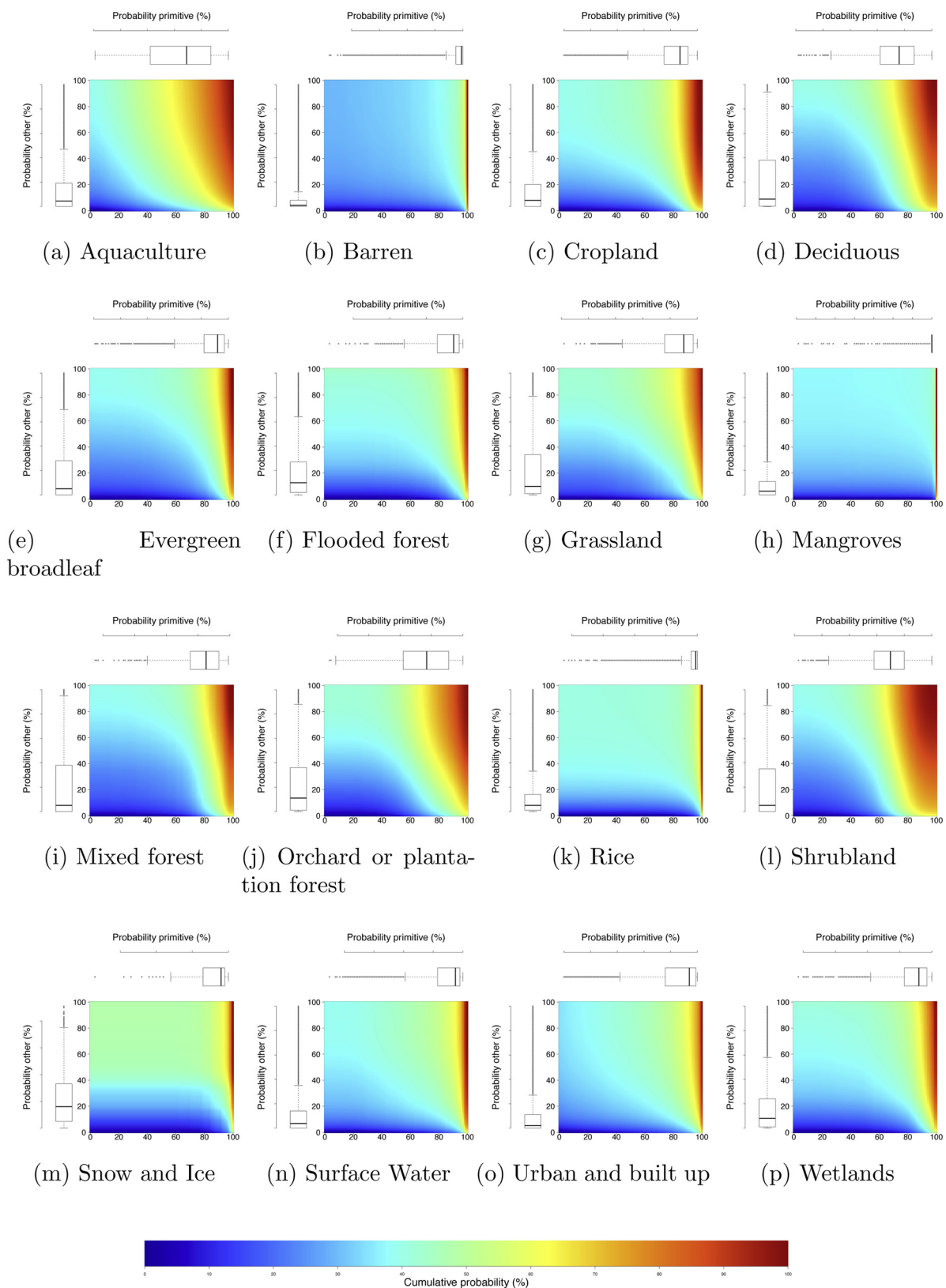
Fig. 5. The box-plots show the distribution of primitives sampled with the reference data. The top image shows the distribution of the primitive, the image on the left the distribution of all other points. The overlap between the primitive and other categories is shown by the cumulative distribution plots.

of regional land cover dynamics (Liu and Zhou, 2004; Gómez et al., 2016; Yang et al., 2016; Hermosilla et al., 2018). Illogical transition are defined as changes that are unlikely to occur as they violate ecological rules within a given time period (Cai et al., 2014; Liang and Gong, 2010; Townsend et al., 2009). The temporal frequency of our map production is annual, therefore we are defining illogical transitions as those that are unlikely to happen within one year. Illogical changes are indicated with a 0 in Table 2 whereas logical changes are indicated with a 1. An iterative approach was used to ensure transitions were propagated through the time-series. Data unmasking was applied to the time-series in ascending order.
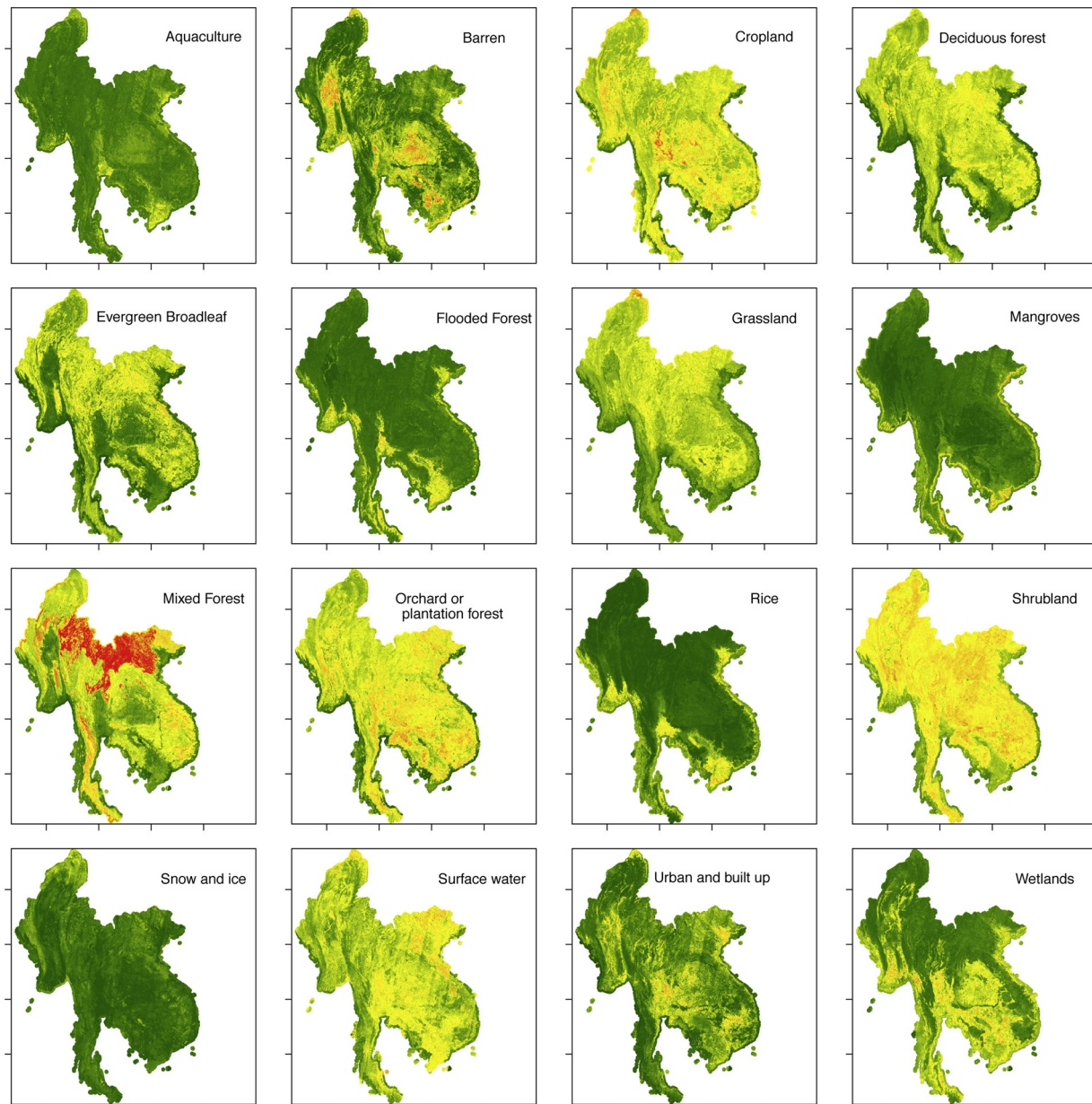
(a) maps
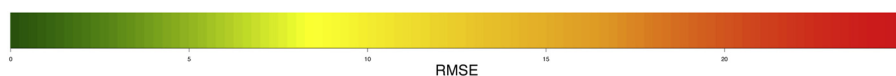
(b) legend

Fig. 6. Root mean square error from the Whittaker smoothing algorithm. The RMSE shows the difference between the original probability and the smoothed one over the timeseries. Green indicates a high agreement whereas red means more disagreement. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

shrubland, deciduous forest and orchards or plantation forest show more mixing.

The Whittaker smoothing was applied to all primitive time-series products. Fig. 6 shows the per-pixel level of agreement between the original timeseries and the smoothed one indicated by the RMSE. Primitives of Snow and Ice, urban and build-up, rice and flood forest have generally low RMSE while shrubland, barren and mixed forest show higher values. Large variability in the temporal signal of the primitive can indicate rapid changes in land cover or inconsistent performance in the method. It can be seen that areas with low probabilities, i.e. features that look very different from the class under consideration have low error rates and thus perform consistent over the time-series.

The decision tree contains two main branches (Fig. 7). The branch on the right contains all forest classes, the one on the left all other classes. The deciduous and evergreen forest classes have a sub category of mixed forest. Water also has two sub-classes, aquaculture and wetlands. Aquaculture, wetlands, and mangroves are included at multiple locations in the decision tree as they exists as sub-class of an overarching class, but also as a separate entity. The numbers indicate the thresholds that were applied.

The forest layer is created as an intermediate primitive because it is a combination of two other biophysical primitives. It is defined based on tree canopy cover, tree canopy height, and a minimum mapping unit. To create it we applied a threshold of 10% on the canopy cover primitive. Then patches that met a minimum patch size requirement of 5 connected pixels (roughly 0.5 ha) were selected. These remaining patches were combined with tree height information which only

patches with a canopy height greater than 5 m were classified as forest lands.

The land cover map and uncertainty layer are shown in Fig. 8, with the accuracy per land cover class in Fig. 9. From both images it is clear that snow, rice, barren, mangroves and cropland have high probabilities, whereas aquaculture, shrubland, plantations and the different forest classes have higher uncertainties. Maps for all years are available on the SERVIR-Mekong website (https://rlcms-servir.adpc.net/en/landcover/).

The results of the validation are shown in Fig. 3 . A total of 560 points were included in the analysis and 11 were discarded because no high resolution imagery was available or the quality was not sufficient to objectively classify the points. We found an overall accuracy of 0.76 with the highest accuracy for water and the lowest for barren, which was often confused with cropland. Also flooded forest and shrubland have a low accuracies. To evaluate if the information in the uncertainty layer is reliable, we evaluated the accuracy and Kappa using different thresholds. Fig. 10 shows the accuracy and number of points using a threshold between 0 and 100. It can be seen that there is an almost linear increase from 50 until 100. Values increase from 0.76 to 0.94 for the accuracy, discarding 324 from the analysis. This demonstrates that the uncertainty layer adds valuable information to the analysis as validation points with high certainties produce a high accuracy in the analysis.
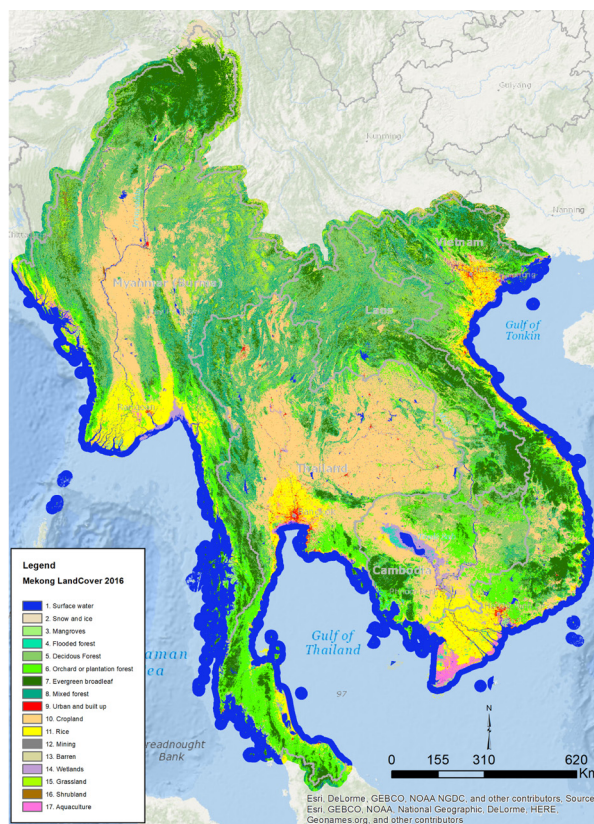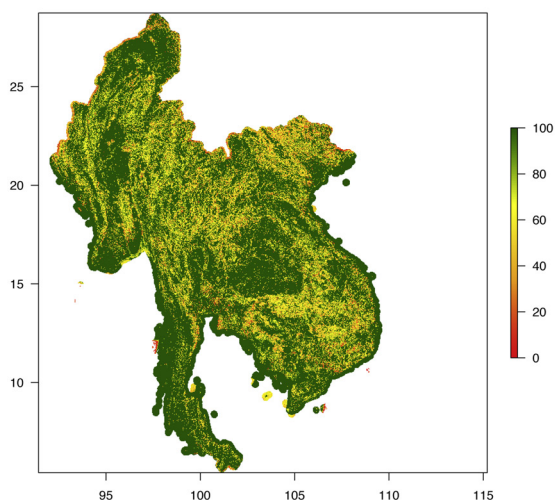
(a) Probability calculated from the monte-carlo simulation. The probability is the count of the assembly mode divided by the total number of model runs.

(b) Land cover map of 2016.

Fig. 8. Probability and landcover map.
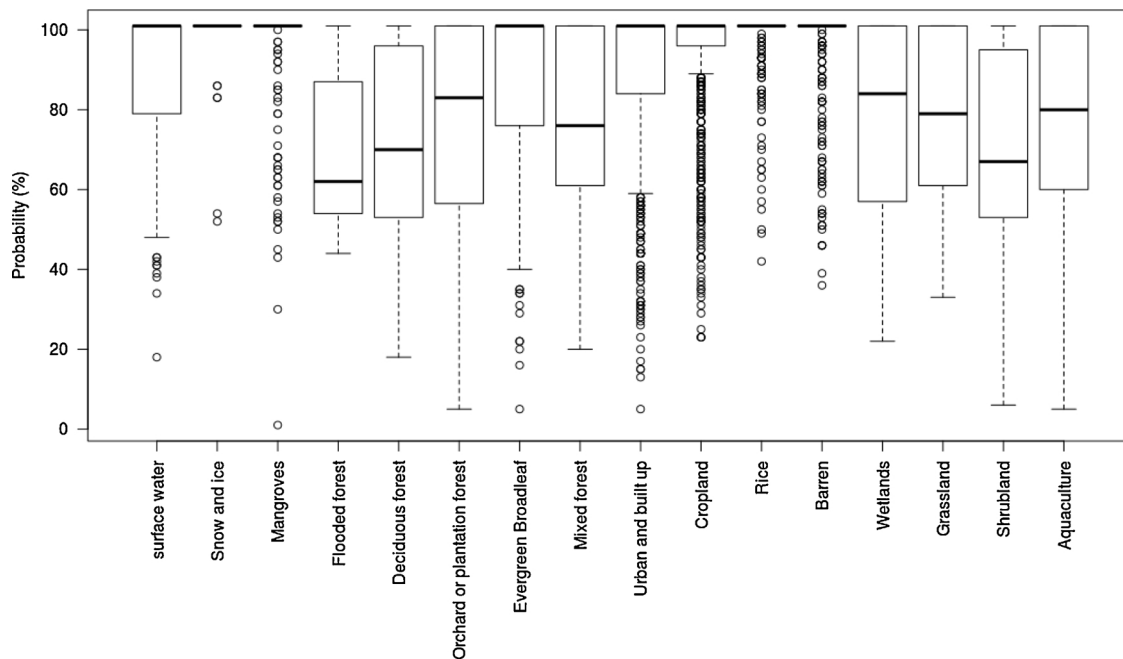
Fig. 9. Distribution of the probability for the different classes sampled using the validation data. The image shows the mean for all years.

Table 3
Confusion matrix for land cover validation of the year 2016 using high resolution satellite imagery. Accuracy of all classes is 0.76.

	SW	SI	M	FF	F	OP	UB	C	R	B	W	G	S	A
Surface water	48	1	0	0	0	0	0	0	0	0	0	0	0	1
Snow and Ice	0	21	0	0	1	0	0	0	0	0	0	1	0	0
Mangrove	0	0	27	3	0	3	0	0	1	3	1	0	0	1
Flooded Forest	0	0	2	10	1	1	0	0	0	0	1	0	0	0
Forest	0	1	0	0	111	6	0	0	0	0	0	5	5	0
Orchard or plantation forest	0	0	0	0	21	30	1	9	3	0	0	1	1	0
Urban and built up	0	0	0	0	0	0	22	0	0	1	0	0	0	0
Cropland	0	0	0	0	3	1	5	60	1	7	2	1	1	0
Rice	0	0	0	1	0	0	0	2	14	2	0	0	0	0
Barren	0	0	0	0	0	0	0	0	0	11	0	0	0	0
Wetlands	1	0	0	3	0	0	1	3	1	1	21	0	0	1
Grassland	0	1	0	0	5	1	0	1	0	0	0	18	4	0
Shrubland	0	0	0	0	6	0	0	0	0	0	0	0	13	0
Aquaculture	1	0	4	1	0	0	0	0	0	0	0	0	0	18
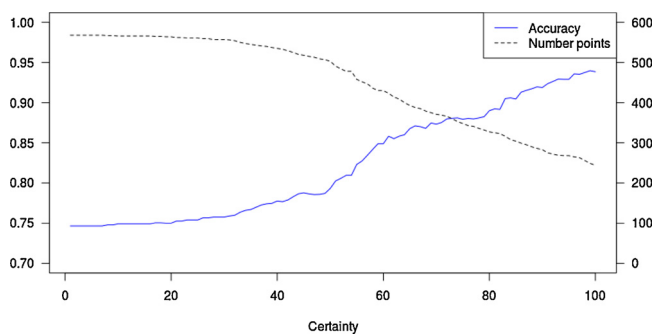
Fig. 10. The accuracy and Kappa was calculated using different thresholds for the probability. The accuracy and Kappa are shown on the left vertical axis, the number of points included in the analysis on the right vertical axis.

5. Discussion

In this work, we have detailed both the architecture and the outputs of an innovative land cover mapping approach termed as the RLCMS, which was developed based on key user requirements. This system produces reliable and accurate land cover maps and is based on primitives which are key building blocks of this system. Yearly composites were created applying all critical steps of cloud and shadow removal, BDRF and topographic correction, as suggested by Young et al. (2017a). These steps were applied on the USGS surface reflectance products to produce consistent timeseries. We found it critical to use surface reflectance products rather than top-of-atmosphere as atmospheric distortions often times led to mis-classification. We also found that it was better to exclude after the scan line correction error, as the striping effect also led to miss-classifications which were also clearly visible in the final land cover maps. However, discarding Landsat-7 led to a data-gap in 2012. Topographic correction was also found to be critical to account for shadow effects. Other issues we found were high percentages of cloud cover throughout the region, but specifically for northern Vietnam. A limitation of the method is that we did not account for specific sensors' characteristics. Spectral reflectance characteristics of Landsat-5, 7 and 8 were all considered consistent. Furthermore, data with different spatial resolutions were combined as bands in the random forest classifier. Multi-sensor data fusion or spatial downscaling technologies might further improve the quality of the composites (Gevaert and García-Haro, 2015; Gao et al., 2006).

The Monte-Carlo simulation shows that the highest uncertainties can be found in the mixed forest classes, orchard and plantations,

shrubland and aquaculture (Fig. 9). This is also represented in the confusion matrix (Table 10) which shows that plantations and shrub are often mixed with forest. This is in agreement with findings of Herold et al. (2008a). They compared four global coarse resolution land cover products and found that mixed tree types and other mixed classes characterized by a mosaic of trees, shrubs, and herbaceous vegetation have commission errors. Accuracies reported in this study are higher than global products (Yang et al., 2017c) which are generally made from coarser resolution imagery and landscape heterogeneity possibly lowers performance. Class accuracies are comparable with other studies that use Landsat as a primary data source (e.g. Huang et al., 2017; Stibig et al., 2007). It is notable, however, that our study includes explicit and systematic error quantification.

Reference data collection was found to be quite challenging as the regional product covers multiple administrative regions. As such, reference data was collected from different sources with variations in typologies, quality, collection methods and temporal density. For example, data from countries were often collected in the field and specifically focused around the country's typologies. The more generic reference data collected by international programs and institutions such as the Food and Agricultural Organization were often collected using high resolution satellite imagery using data collection software such as collect earth (Bey et al., 2016; Saah et al., 2019a). For some classes, some extra reference data was collected applying opportunistic sampling on high resolution on GEE. Applying this method introduces new uncertainties as points were not randomly selected and often times cover areas with a rather homogeneous land cover. A large portion of the reference data used in this study were from 2015 and there was a bias towards the more recent past due to availability of high resolution satellite imagery. The reference data also showed an unequal spatial distribution with higher densities in parts of Vietnam and Thailand. Other potential sources of error in reference data collection include subjectivity in reference data labeling (Powell et al., 2004).

The LMR covers a large area with different ecological regions. Covering these different ecological regions presents a challenge in terms of reference data collection and training the model. For example, the central part of Myanmar has a climate vastly different from the region with mostly dry forests and dry agriculture covering the area. The phenology and spectral signature of these dryland classes are very different from the rest of the area whereas reference data was scarce. Additional reference data collection was needed as dryland agriculture was often confused with barren and urban whereas dryland forest were classified as shrub. However, additional training data often introduced a new error in other geographic regions of the region. We therefore advise governments, institutes and other users of the RLCMS to apply the system on smaller more digestible and ecological homogeneous units.

The smoothing algorithm was applied to remove noise from the primitives while creating a more consistent time-series. Consistency in time-series is important when the final purpose of the map is related to long-term analysis such as the FAO forest resource assessment and IPCC reporting. However, applying temporal smoothing removes short term changes from the time-series, which makes it less suitable for purposes such as change detection. Algorithms such as landtrendr (Kennedy et al., 2010) and BFAST (Verbesselt et al., 2010) are more suited for change detection as they are applied directly on the source data. Information from the RLCMS regional product can add valuable information to these types of analysis. The RMSE, as an output of the smoothing algorithm provides valuable information on the consistency of the timeseries and can be used to identify primitives that show a lot of temporal variability or areas that need more reference data. More research is needed as different primitives might require specific smoothing approaches.

It was demonstrated that the Monte-Carlo simulation included in the assemblage method provides valuable information on the uncertainty of the final land cover map. It was shown that explicit

quantification of the uncertainty can improve the final accuracy up to 94%. This accuracy is higher than comparable studies of the region (Spruce et al., 1910) while including more classes with a higher spatiotemporal resolution. Moreover, in our approach, the assemblage process allows a user to tailor the decision tree structure and thresholds to match a map product's area for a land cover of interest which align with estimates derived from a statistical inventory. Future work may include refinements via application of pixel-based approaches while including distributions functions to the thresholds rather than random numbers on a single standard deviation for a map. These methods might improve the quality of the uncertainty maps which can then be used to collect more reference data in areas with high uncertainty. Repeating this cycle a number of times is expected to lead to a cumulative improvement of the final product.

The strength of the system is that it is modular and that each component can easily be replaced by another one. Thematic primitives can be combined in a customizable assembly protocol with the users' specific target land cover classes. In this study we used a single composite for each year, but composites that contain temporally more dense information could also be used in order to include information on phenology. The data density could also be increased along the spectral axes by including more covariates or using sensor fusion techniques. The random forest model is one of many machine learning methods that be can easily be replaced by newer methods such as U-Nets. Similarly, there are many methods for temporal smoothing. Performance might differ based on the primitive under consideration. All methods and data are open and free. Collaborative and iterative development of the system will improve the final results and help policy makers in developing countries making more informed decision regarding natural resource management.

In regards to future work, new satellites from the Sentinel constellation provide open data with higher spectral, spatial and temporal resolution. Moreover, they include synthetic aperture radar which can be very useful for tropical countries with persistent cloud cover. Data fusion techniques and approaches are being increasingly developed and utilized for land cover mapping (Poortinga et al., 2019).

6. Conclusions

In this study, we described the architecture and demonstrated the performance of a modular regional land cover mapping system using primitives as building blocks for constructing land cover maps. We successfully demonstrated this in the context of the Lower Mekong Region, and found that our decision tree logic and Monte Carlo simulations lead to an explicit error quantification resulting in a overall 94% accuracy when taking these errors into account. The architecture is fully customizable and can include a variety of sensors and machine learning algorithms to create primitives. Most suitable smoothing approaches can be selected and applied on a primitive level which are then combined in a fully customizable assembly protocol. Given these findings we propose that using primitives and a measure of their uncertainties can potentially serve as key ingredients in developing flexible land cover monitoring systems which can serve multiple stakeholders and geographic regions in an internally consistent fashion.

Conflict of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Covariates

Table A.1 and Table A.2

Table A.1
Covariates used to generate the primitives: (a) Urban and built up, (b) Cropland, (c) Rice, (d) Surface water, (e) Shrubland, (f) Mangrove, (g) Barren, (h) Aquaculture, (i) Wetlands, (j) Orchard or plantation forest, (k) Grassland, (l) Flooded Forest, (m) Evergreen Broadleaf, (n) Deciduous forest, (o) Mixed forest and (p) Snow and ice.

		Resolution	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
1	EVI	30						x						x		x	x	
2	ND_blue_nir	30	x				x					x						
3	ND_blue_red	30	x															
4	ND_blue_swir1	30	x			x	x					x	x					
5	ND_blue_swir2	30				x							x					x
6	ND_green_nir	30				x	x	x				x			x		x	
7	ND_green_swir1	30				x	x			x		x	x			x	x	x
8	ND_green_swir2	30		x		x	x						x			x	x	
9	ND_nir_red	30						x					x		x			
10	ND_nir_swir1	30						x										x
11	ND_nir_swir2	30						x			x							
12	ND_red_swir1	30				x	x			x		x	x			x		
13	ND_red_swir2	30		x		x	x					x	x					
14	ND_swir1_swir2	30	x		x			x						x	x	x	x	
15	R2_cycle1	500					x						x		x	x		
16	R2_cycle2	500						x				x	x		x	x		
17	R2_cycle3	500			x			x		x		x	x		x			
18	R_red_swir1	30				x	x			x		x	x			x	x	x
19	R_swir1_nir	30						x										x
20	SAVI	30							x					x	x	x		
21	aspect	30								x								
22	auto	500			x	x	x				x		x	x		x	x	
23	blue	30				x			x			x						
24	brightness	30		x					x	x							x	
25	change_abs	30								x								
26	change_norm	30								x	x							
27	distBuildings	30		x	x		x				x			x		x		x
28	distCoast	30		x	x			x		x	x		x	x				
29	distRoad	30	x		x		x			x	x		x	x	x	x		x
30	distStream	30				x												
31	eco	30					x	x		x	x			x	x	x		
32	ecoForest	30						x		x					x	x		x
33	elevation	30		x	x	x	x	x		x	x	x	x	x	x	x	x	x
34	fifth	30	x	x		x						x			x	x	x	
35	fourth	30	x															x
36	green	30			x		x		x				x	x	x		x	
37	greenness	30										x						
38	hand	30				x												
39	max_extent	30									x							
40	nir	30	x			x				x								
41	occurrence	30								x	x							
42	p20_EVI	30						x				x				x	x	
43	p20_IBI	30						x										
44	p20_ND_blue_nir	30				x	x					x						
45	p20_ND_blue_red	30	x											x				
46	p20_ND_blue_swir1	30	x		x	x				x		x	x					x
47	p20_ND_blue_swir2	30		x		x							x					x
48	p20_ND_green_nir	30					x					x			x		x	
49	p20_ND_green_swir1	30	x	x						x								
50	p20_ND_green_swir2	30		x														
51	p20_ND_nir_red	30						x				x			x			
52	p20_ND_nir_swir1	30						x										
53	p20_ND_nir_swir2	30						x										
54	p20_ND_red_swir1	30		x		x				x		x	x					x
55	p20_ND_red_swir2	30		x		x	x			x		x	x					x
56	p20_ND_swir1_swir2	30	x		x			x			x			x				
57	p20_R_red_swir1	30		x		x	x			x		x	x					x
58	p20_R_swir1_nir	30						x							x			
59	p20_SAVI	30						x				x			x	x		
60	p20_blue	30						x	x									x
61	p20_brightness	30	x	x		x			x		x			x			x	
62	p20_fifth	30	x								x			x		x	x	
63	p20_fourth	30	x															x

Table A.2

Covariates used to generate the primitives: (a) Urban, (b) Cropland, (c) Rice, (d) Water, (e) Shrub, (f) Mangrove, (g) Barren, (h) Aquaculture, (i) Wetlands, (j) Plantations, (k) Grass, (l) Flooded Forest, (m) evergreen, (n) Deciduous, (o) Mixed forest and (p) Snow.

		Resolution	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
64	p20_green	30		x	x		x		x						x		x	
65	p20_greenness	30						x				x						x
66	p20_nir	30	x			x				x								
67	p20_red	30							x		x						x	x
68	p20_swir1	30	x	x	x				x	x				x	x	x		
69	p20_swir2	30		x	x				x	x				x	x			
70	p20_tcAngleBG	30				x		x				x			x	x		
71	p20_tcAngleBW	30		x			x	x										x
72	p20_tcAngleGW	30													x			
73	p20_tcDistBG	30		x					x	x	x							
74	p20_tcDistBW	30	x	x					x	x	x			x			x	
75	p20_tcDistGW	30	x	x			x			x		x						x
76	p20_wetness	30	x	x			x	x								x		x
77	p80_EVI	30												x		x		
78	p80_ND_blue_nir	30	x									x						
79	p80_ND_blue_red	30								x								
80	p80_ND_blue_swir1	30	x			x	x					x	x					
81	p80_ND_blue_swir2	30				x							x					x
82	p80_ND_green_nir	30			x						x	x			x			x
83	p80_ND_green_swir1	30				x	x					x						x
84	p80_ND_green_swir2	30	x			x	x											x
85	p80_ND_nir_red	30						x										
86	p80_ND_nir_swir1	30	x															
87	p80_ND_nir_swir2	30	x															
88	p80_ND_red_swir1	30		x		x	x					x		x		x	x	
89	p80_ND_red_swir2	30			x	x	x					x					x	
90	p80_ND_swir1_swir2	30	x		x								x			x	x	
91	p80_R_red_swir1	30				x						x	x			x	x	
92	p80_R_swir1_nir	30	x															
93	p80_SAVI	30										x	x	x	x		x	
94	p80_blue	30	x															
95	p80_brightness	30				x			x	x								
96	p80_fifth	30	x								x							
97	p80_fourth	30	x															
98	p80_green	30			x			x	x		x		x			x		x
99	p80_greenness	30			x							x						
100	p80_nir	30	x							x								x
101	p80_red	30		x					x				x	x	x			
102	p80_sixth	30																x
103	p80_swir1	30	x			x	x		x									
104	p80_swir2	30	x						x	x								
105	p80_tcAngleBG	30			x			x							x	x		
106	p80_tcAngleBW	30	x			x	x											
107	p80_tcAngleGW	30				x	x		x									
108	p80_tcDistBG	30				x			x	x								
109	p80_tcDistBW	30			x				x	x								
110	p80_tcDistGW	30	x															x
111	p80_wetness	30					x		x							x		x
112	protected	30											x					
113	red	30							x				x	x		x		
114	sixth	30			x						x				x			x
115	slope	30		x	x		x				x	x	x	x	x	x	x	x
116	swir1	30	x	x		x			x									
117	swir2	30		x	x			x	x		x			x	x		x	
118	tcAngleBG	30			x			x			x				x			
119	tcAngleBW	30		x												x		
120	tcAngleGW	30	x			x			x							x		
121	tcDistBG	30				x			x	x								
122	tcDistBW	30		x					x	x								
123	tcDistGW	30	x				x		x									
124	tcc	30			x		x				x	x	x		x	x	x	
125	transition	30								x	x							
126	treeheight	30			x	x		x			x	x	x	x	x	x	x	
127	water	30				x												
128	wetness	30					x									x		x

References

Angiuli, E., Trianni, G., 2014. Urban mapping in Landsat images based on normalized difference spectral vector. IEEE Geosci. Remote Sens. Lett. 11, 661–665.

Atkinson, P.M., Jeganathan, C., Dash, J., Atzberger, C., 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. Remote Sens. Environ. 123, 400–417.

Atzberger, C., Eilers, P.H., 2011a. A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of South America. Int. J. Digit. Earth 4, 365–386.

Atzberger, C., Eilers, P.H., 2011b. Evaluating the effectiveness of smoothing algorithms in

the absence of ground reference measurements. Int. J. Remote Sens. 32, 3689–3709.

Azzari, G., Lobell, D., 2017. Landsat-based classification in the cloud: an opportunity for a paradigm shift in land cover monitoring. Remote Sens. Environ. 202, 64–74.

Bartholomé, E., Belward, A.S., 2005. Glc2000: a new approach to global land cover mapping from earth observation data. Int. J. Remote Sens. 26, 1959–1977.

Beers, T.W., Dress, P.E., Wensel, L.C., 1966. Notes and observations: aspect transformation in site productivity research. J. Forest. 64, 691–692.

Bey, A., Sánchez-Paus Díaz, A., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., Bastin, J.-F., Moore, R., Federici, S., Rezende, M., et al., 2016. Collect earth: land use and land cover assessment through augmented visual interpretation. Remote Sens. 8, 807.

Blanchez, J., 1997. Forest resources and roundwood supply in the Asia pacific countries: situation and outlook to the year 2010. Asia-Pacific Forestry Towards 2010. Asia-Pacific Forestry Sector Outlook Study Working Paper Series (FAO).

Bontemps, S., Defourny, P., Bogaert, E.V., Arino, O., Kalogirou, V., Perez, J.R., 2011. GlobCover 2009 – products description and validation report.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bui, Y.T., Orange, D., Visser, S., Hoanh, C.T., Laissus, M., Poortinga, A., Tran, D.T., Stroosnijder, L., 2014. Lumped surface and sub-surface runoff for erosion modeling within a small hilly watershed in northern Vietnam. Hydrol. Process. 28, 2961–2974.

Cai, S., Liu, D., Sulla-Menashe, D., Friedl, M.A., 2014. Enhancing MODIS land cover product with a spatial-temporal modeling algorithm. Remote Sens. Environ. 147, 243–255.

Chastain, R., Housman, I., Goldstein, J., Finco, M., 2019. Empirical cross sensor comparison of Sentinel-2a and 2b MSI, Landsat-8 OLI, and Landsat-7 ETM+ top of atmosphere spectral characteristics over the conterminous united states. Remote Sens. Environ. 221, 274–285.

Colby, J.D., 1991. Topographic normalization in rugged terrain. Photogramm. Eng. Remote Sens. 57, 531–537.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46.

Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data – the TM Tasseled Cap. IEEE Trans. Geosci. Remote Sens. 256–263.

Di Gregorio, A., Jansen, L.J., 1998. Land Cover Classification System (LCCS): Classification Concepts and User Manual. FAO, Rome.

Di Gregorio, A., 2005. Land Cover Classification System: Classification Concepts and User Manual: LCCS, Volume 8. Food Agriculture Org.

Di Gregorio, A., et al., 2016. Land Cover Classification System: Classification Concepts, Software Version 3.

Eilers, P.H., 2003. A perfect smoother. Anal. Chem. 75, 3631–3636.

Faber-Langendoen, D., Tart, D.L., Crawford, R.H., 2009. Contours of the Revised US National Vegetation Classification Standard. Bull. Ecol. Soc. Am. 90, 87–93.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al., 2007. The shuttle radar topography mission. Rev. Geophys. 45.

F.G.D. Committee, 1977. National Vegetation Classification Standard. FGDC-STD-005-2008. Vegetation Subcommittee, Federal Geographic Data Committee, US Geological Survey, Reston, Virginia, USA. Available online: http://www.fgdc.gov/standards/documents/standards/vegetation/vegclass.pdf.

Flood, N., 2013. Seasonal composite Landsat TM/ETM+ images using the medoid (a multi-dimensional median). Remote Sens. 5, 6481–6500.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201.

Franklin, S., Wulder, M., 2002. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. Prog. Phys. Geogr. 26, 173–205.

Friedl, M.A., McIver, D.K., Hodges, J.C., Zhang X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., et al., 2002. Global land cover mapping from MODIS: algorithms and early results. Remote Sens. Environ. 83, 287–302.

Fritz, S., See, L., Rembold, F., 2010. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. Int. J. Remote Sens. 31, 2237–2256.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al., 2015. The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes. Sci. Data 2, 150066.

Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: A review. ISPRS J. Photogramm. Remote Sens. 116, 55–72.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44, 2207–2218.

Gao, F., He, T., Masek, J.G., Shuai, Y., Schaaf, C.B., Wang, Z., 2014. Angular effects and correction for medium resolution sensors to support crop monitoring. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 7, 4480–4489.

GEOSS, 2005. The Global Earth Observation System of Systems (GEOSS): 10-Year Implementation Plan.

Gevaert, C.M., García-Haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sens. Environ. 156, 34–44.

Giri, C., Zhu, Z., Reed, B., 2005. A comparative analysis of the global land cover 2000 and MODIS land cover data sets. Remote Sens. Environ. 94, 123–132.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., et al., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens. 34, 2607–2654.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Gu, D., Gillespie, A., 1998. Topographic normalization of Landsat TM images of forest based on subpixel sun-canopy-sensor geometry. Remote Sens. Environ. 64, 166–175.

Hansen, M.C., DeFries, R.S., Townshend, J.R., Sohlberg, R., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens. 21, 1331–1364.

Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., Loveland, T.R., 2011. Continuous fields of land cover for the conterminous united states using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project. Remote Sens. Lett. 2, 279–288.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T., et al., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853.

Hansen, M.C., Potapov, P.V., Goetz, S.J., Turubanova, S., Tyukavina, A., Krylov, A., Kommareddy, A., Egorov, A., 2016. Mapping tree height distributions in sub-Saharan Africa using Landsat 7 and 8 data. Remote Sens. Environ. 185, 221–232.

Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2018. Disturbance-informed annual land cover classification maps of Canada's forested ecosystems for a 29-year Landsat time series. Can. J. Remote Sens. 44, 67–87.

Herold, M., Schmullius, C., 2004. Report on the harmonization of global and regional land cover products. Workshop report at FAO, Rome, Italy 14–16.

Herold, M., Latham, J., Di Gregorio, A., Schmullius, C., 2006a. Evolving standards in land cover characterization. J. Land Use Sci. 1, 157–168.

Herold, M., Woodcock, C.E., Di Gregorio, A., Mayaux, P., Belward, A.S., Latham, J., Schmullius, C.C., 2006b. A joint initiative for harmonization and validation of land cover datasets. IEEE Trans. Geosci. Remote Sens. 44, 1719–1727.

Herold, M., Mayaux, P., Woodcock, C., Baccini, A., Schmullius, C., 2008a. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. Remote Sens. Environ. 112, 2538–2556.

Herold, M., Woodcock, C.E., Loveland, T.R., Townshend, J., Brady, M., Steenmans, C., Schmullius, C.C., 2008b. Land-cover observations as part of a global earth observation system of systems (GEOSS): progress, activities, and prospects. IEEE Syst. J. 2, 414–423.

Holden, C.E., Woodcock, C.E., 2016. An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. Remote Sens. Environ. 185, 16–36.

Housman, I., Chastain, R., Finco, M., 2018. An evaluation of forest health insect and disease survey data and satellite-based remote sensing forest change detection methods: case studies in the United States. Remote Sens. 10, 1184.

Huang, H., Chen, Y., Clinton, N., Wang, J., Wang, X., Liu, C., Gong, P., Yang, J., Bai, Y., Zheng, Y., et al., 2017. Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. Remote Sens. Environ. 202, 166–176.

Huete, A.R., 1988. A Soil-Adjusted Vegetation Index (SAVI). Remote Sens. Environ. 25, 295–309.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.

Jiang, Z., Huete, A.R., Didan, K., Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. Remote Sens. Environ. 112, 3833–3845.

Ju, J., Roy, D.P., Vermote, E., Masek, J., Kovalskyy, V., 2012. Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. Remote Sens. Environ. 122, 175–184.

Jung, M., Henkel, K., Herold, M., Churkina, G., 2006. Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sens. Environ. 101, 534–553.

Justice, C.O., Wharton, S.W., Holben, B., 1981. Application of digital terrain data to quantify and reduce the topographic effect on Landsat data. Int. J. Remote Sens. 2, 213–230.

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr – temporal segmentation algorithms. Remote Sens. Environ. 114, 2897–2910.

Key, C.H., Benson, N.C., 1999. The Normalized Burn Ratio (NBR): a Landsat TM radiometric measure of burn severity. United States Geological Survey, Northern Rocky Mountain Science Center, Bozeman, MT.

Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens. Environ. 177, 89–100.

Liang, L., Gong, P., 2010. An assessment of MODIS collection 5 global land cover product for biological conservation studies. IEEE 18th International Conference on Geoinformatics 1–6.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.

Liu, H., Zhou, Q., 2004. Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison. Int. J. Remote Sens. 25, 1037–1050.

Loveland, T.R., Belward, A., 1997. The IGBP-DIS global 1km land cover data set, DISCover: first results. Int. J. Remote Sens. 18, 3289–3295.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP discover from 1 km AVHRR data. Int. J. Remote Sens. 21, 1303–1330.

Lucht, W., Schaaf, C.B., Strahler, A.H., 2000. An algorithm for the retrieval of Albedo from space using semiempirical BRDF models. IEEE Trans. Geosci. Remote Sens. 38, 977–998.

Markert, K., Schmidt, C., Griffin, R., Flores, A., Poortinga, A., Saah, D., Muench, R., Clinton, N., Chishtie, F., Kityuttachai, K., et al., 2018. Historical and operational monitoring of surface sediments in the Lower Mekong Basin using Landsat and Google Earth Engine Cloud Computing. Remote Sens. 10, 909.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F.,

Kutler, J., Lim, T.-K., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. IEEE Geosci. Remote Sens. Lett. 3, 68–72.

Mayaux, P., Eva, H., Gallego, J., Strahler, A.H., Herold, M., Agrawal, S., Naumov, S., De Miranda, E.E., Di Bella, C.M., Ordoyne, C., et al., 2006. Validation of the global land cover 2000 map. IEEE Trans. Geosci. Remote Sens. 44, 1728–1739.

McCallum, I., Obersteiner, M., Nilsson, S., Shvidenko, A., 2006. A spatial comparison of four satellite derived 1 km global land cover datasets. Int. J. Appl. Earth Observ. Geoinform. 8, 246–255.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.

MODIS Land Cover Team, 2003. Validation of the Consistent Year 2003. V003. MODIS Land Cover Product.

Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sens. Environ. 129, 122–131.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., et al., 2001. Terrestrial ecoregions of the world: a new map of life on earth. BioScience 51, 933–938.

OpenStreetMap Contributors, 2017. Planet Dump. Retrieved from: https://www.planet.osm.org, https://www.openstreetmap.org.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418.

Pettorelli, N., Wegmann, M., Skidmore, A., Mücher, S., Dawson, T.P., Fernandez, M., Lucas, R., Schaepman, M.E., Wang, T., O'Connor, B., et al., 2016. Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. Remote Sens. Ecol. Conserv. 2, 122–131.

Poortinga, A., Bastiaanssen, W., Simons, G., Saah, D., Senay, G., Fenn, M., Bean, B., Kadyszewski, J., 2017. A self-calibrating runoff and streamflow remote sensing model for ungauged basins using open-access earth observation data. Remote Sens. 9, 86.

Poortinga, A., Clinton, N., Saah, D., Cutter, P., Chishtie, F., Markert, K.N., Anderson, E.R., Troy, A., Fenn, M., Tran, L.H., et al., 2018. An operational before-after-control-impact (BACI) designed platform for vegetation monitoring at planetary scale. Remote Sens. 10, 760.

Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., San Aung, K., Chishtie, F., Saah, D., 2019. Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification. Remote Sens. 11, 831.

Powell, R., Matzke, N., de Souza Jr., C., Clark, M., Numata, I., Hess, L., Roberts, D., 2004. Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. Remote Sens. Environ. 90, 221–234.

Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., Ohmann, J.L., 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. Remote Sens. Environ. 114, 1053–1068.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Riano, D., Chuvieco, E., Salas, J., Aguado, I., 2003. Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types (2003). IEEE Trans. Geosci. Remote Sens. 41, 1056–1061.

Rouse Jr., J., Haas, R., Schell, J., Deering, D., 1974. Monitoring Vegetation Systems in the Great Plains with ERTS.

Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization. gap filling, and prediction of Landsat data. Remote Sens. Environ. 112, 3112–3130.

Roy, D.P., Kovalskyy, V., Zhang, H., Vermote, E.F., Yan, L., Kumar, S., Egorov, A., 2016a. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sens. Environ. 185, 57–70.

Roy, D.P., Zhang, H., Ju, J., Gomez-Dans, J.L., Lewis, P.E., Schaaf, C., Sun, Q., Li, J., Huang, H., Kovalskyy, V., 2016b. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. Remote Sens. Environ. 176, 255–271.

Roy, D.P., Li, J., Zhang, H.K., Yan, L., Huang, H., Li, Z., 2017. Examination of Sentinel-2a multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. Remote Sens. Environ. 199, 25–38.

Running, S.W., 2008. Ecosystem disturbance, carbon, and climate. Science 321, 652–653.

Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., Poortinga, A., Markert, K., Quyen, N.H., San Aung, K., et al., 2019a. Collect earth: an online tool for systematic reference data collection in land cover and use applications. Environ. Model. Softw.

Saah, D., Tenneson, K., Matin, M., Uddin, K., Cutter, P., Poortinga, A., Nguyen, Q.H., Patterson, M., Johnson, G., Markert, K., Flores, A., Anderson, E., Weigel, A., Ellenberg, W.L., Bhargava, R., Aekakkararungroj, A., Bhandari, B., Khanal, N., Housman, I.W., Potapov, P., Tyukavina, A., Maus, P., Ganz, D., Clinton, N., Chishtie, F., 2019b. Land cover mapping in data scarce environments: challenges and opportunities. Front. Environ. Sci. 7.

Salomonson, V.V., Appel, I., 2004. Estimating fractional snow cover from MODIS using the normalized difference snow index. Remote Sens. Environ. 89, 351–360.

Scepan, J., 1999. Thematic validation of high-resolution global land-cover data sets. Photogramm. Eng. Remote Sens. 65, 1051–1060.

Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., Gao, F., 2013. Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. Technical Report, US Geological Survey.

Shepherd, J., Dymond, J., 2003. Correcting satellite imagery for the variance of reflectance and illumination with topography. Int. J. Remote Sens. 24, 3503–3514.

Shumway, R.H., Stoffer, D.S., 2011. Time series regression and exploratory data analysis. Time Series Analysis and its Applications. Springer, pp. 47–82.

Simons, G., Poortinga, A., Bastiaanssen, W.G., Saah, D., Troy, D., Hunink, J., Klerk, M.d., Rutten, M., Cutter, P., Rebelo, L.-M., et al., 2017. On spatially distributed hydrological ecosystem services: bridging the quantitative information gap using remote sensing and hydrological models.

Smith, J., Lin, T.L., Ranson, K., et al., 1980. The Lambertian assumption and Landsat data. Photogramm. Eng. Remote Sens. 46, 1183–1189.

Soenen, S.A., Peddle, D.R., Coburn, C.A., 2005. SCS + c: a modified sun-canopy-sensor topographic correction in forested terrain. IEEE Trans. Geosci. Remote Sens. 43, 2148–2159.

Spruce, J., Bolten, J., Srinivasan, R., Lakshmi, V., 1910. Developing land use land cover maps for the Lower Mekong Basin to aid hydrologic modeling and basin planning. Remote Sens. 10 (2018).

Stürck, J., Poortinga, A., Verburg, P.H., 2014. Mapping ecosystem services: the supply and demand of flood regulation services in Europe. Ecol. Indicat. 38, 198–211.

Stehman, S.V., Czaplewski, R.L., 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sens. Environ. 64, 331–344.

Stibig, H.-J., Belward, A., Roy, P., Rosalina-Wasrin, U., Agrawal, S., Joshi, P., Hildanus, R., Beuchle, S., Fritz, S., Mubareka, et al., 2007. A land-cover map for South and southeast Asia derived from SPOT-VEGETATION data. J. Biogeogr. 34, 625–637.

Teillet, P., Guindon, B., Goodenough, D., 1982. On the slope-aspect correction of multi-spectral scanner data. Can. J. Remote Sens. 8, 84–106.

Tolentino, P.L.M., Poortinga, A., Kanamaru, H., Keesstra, S., Maroulis, J., David, C.P.C., Ritsema, C.J., 2016. Projected impact of climate change on hydrological regimes in the Philippines. PLOS ONE 11, e0163941.

Townsend, P.A., Helmers, D., Kingdon, C.C., McNeil, B.E., de Beurs, K.M., Eshleman, K.N., 2009. Changes in the extent of surface mining and reclamation in the central Appalachians detected using a 1976–2006 Landsat time series. Remote Sens. Environ. 113, 62–72.

Trimble Jr., G.R., Weitzman, S., 1956. Site index studies of upland oaks in the northern Appalachians. For. Sci. 2, 162–173.

Turner, B.L., Lambin, E.F., Reenberg, A., 2007. The emergence of land change science for global environmental change and sustainability. Proc. Natl. Acad. Sci. U.S.A. 104, 20666–20671.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114, 106–115.

Vermote, E.F., Tanré, D., Deuze, J.L., Herman, M., Morcette, J.-J., 1997. Second simulation of the satellite signal in the solar spectrum, 6 S: an overview. IEEE Trans. Geosci. Remote Sens. 35, 675–686.

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

Whittaker, E.T., 1922. On a new method of graduation. Proc. Edinb. Math. Soc. 41, 63–75.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 122, 2–10.

Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T., 2018. LandCover 2.0. Int. J. Remote Sens. 39, 4254–4284.

Xu, H., 2008. A new index for delineating built-up land features in satellite imagery. Int. J. Remote Sens. 29, 4269–4276.

Yang, C., Xu, Y., Nebert, D., 2013. Redefining the possibility of digital earth and geosciences with spatial cloud computing. Int. J. Digit. Earth 6, 297–312.

Yang, G., Fang, S., Dian, Y., Bi, C., 2016. Improving seasonal land cover maps of Poyang Lake Area in China by taking into account logical transitions. ISPRS Int. J. Geoinf. 5, 165.

Young, N.E., Anderson, R.S., Chignell, S.M., Vorster, A.G., Lawrence, R., Evangelista, P.H., 2017a. A survival guide to Landsat preprocessing. Ecology 98, 920–932.

Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F., 2017b. Big data and cloud computing: innovation opportunities and challenges. Int. J. Digit. Earth 10, 13–53.

Yang, Y., Xiao, P., Feng, X., Li, H., 2017c. Accuracy assessment of seven global land cover datasets over china. ISPRS J. Photogramm. Remote Sens. 125, 156–173.

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94.

Zhu, Z., Gallant, A.L., Woodcock, C.E., Pengra, B., Olofsson, P., Loveland, T.R., Jin, S., Dahal, D., Yang, L., Auch, R.F., 2016. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. ISPRS J. Photogramm. Remote Sens. 122, 206–221.

Zuliana, S.U., Perperoglou, A., 2017. Two dimensional smoothing via an optimised Whittaker smoother. Big Data Analyt. 2, 6.