Remote Sensing in Ecology and Conservation
Open Access
ZSL LET'S WORK FOR WILDLIFE

ORIGINAL RESEARCH

A Landsat composite covering all Amazonia for applications in ecology and conservation

Jasper Van doninck & Hanna Tuomisto

Amazon Research Team, Department of Biology, University of Turku, FI-20014 Turku, Finland

Keywords
Amazonia, BRDF, image compositing, preprocessing, satellite imagery, soils

Correspondence
Jasper Van Doninck, Amazon Research Team, Department of Biology, University of Turku, FI-20014 Turku, Finland. Tel: +358 29 450 4235; E-mail: jasper.vandoninck@utu.fi

Funding Information
The present research was funded by a grant from the Academy of Finland to Hanna Tuomisto.

Editor: Harini Nagendra
Associate Editor: Duccio Rocchini

Received: 13 November 2017; Revised: 1 February 2018; Accepted: 19 February 2018

doi: 10.1002/rse2.77

Remote Sensing in Ecology and Conservation 2018;**4** (3):197–210

Abstract

Studies at small spatial extents have shown that local floristic and edaphic patterns within the hyper-diverse Amazonian forests can be identified at a high thematic resolution using Landsat imagery. This suggests that Landsat images have the potential to indicate ecologically relevant environmental and biotic variation in the forests also at the extent of the entire basin. However, the full potential of Landsat data for these purposes has not yet been exploited in ecological and biodiversity research or in conservation applications. This is largely because the artifactual noise that is introduced by atmospheric and directional effects into multi-scene composite images can swamp the subtle spectral differences between different types of primary forest. Here, we present a new Landsat TM/ETM+ image composite for the entire Amazon biome that largely overcomes these problems. It is based on more than 16 000 individual image acquisitions from the 10-year period 2000–2009. The images were individually processed to directionally and topographically normalized surface reflectance and combined into 2.5 degree tiles using the medoid compositing criterion. Visual inspection showed that the resulting image composite is radiometrically clearly more consistent than other currently available Landsat composites. We tested the ecological relevance of the new Landsat composite by comparing its reflectance values with edaphic properties measured in more than 300 field sampling localities spread across 2000 km of Amazonia. We found a strong correlation between observed and predicted concentration of exchangeable base cations in the surface soil, which indicates that the compositing approach has succeeded in removing most of the artifactual noise. The Landsat composite image should be of great value for a multitude of applications in ecology, biodiversity research and conservation planning that require environmental data layers combining detailed spatial resolution, basin-wide coverage and high radiometric accuracy.

Introduction

Remote sensing data products have become crucial in many applications in ecology and biodiversity research, as well as in conservation planning (Gillespie et al. 2008; He et al. 2015). This is especially the case in regions like Amazonia, which combine a large size, difficult access, high biological diversity and a significant threat of deforestation and other anthropological pressures. Under these conditions, obtaining direct field information about biodiversity patterns is very slow and laborious, which has led to the situation that observational data are strongly biased towards easily accessible sites, and general understanding of broad-scale patterns is virtually lacking.

Satellite images hold great promise as a source of uniform data with global coverage for applications that require conclusions to be drawn beyond the few localities for which accurate field data exist. Different satellite sensors provide data that are appropriate for different purposes, depending on their spatial, temporal and radiometric properties (Kerr and Ostrovsky 2003; Wang et al. 2010). Landsat TM/ETM+ imagery has a global coverage at a relatively high 30 m resolution and a 16-day revisit time, and the Landsat archive is unique in

spanning several decades. For these reasons, Landsat images have been extensively used in Amazonia for monitoring such processes as fluvial dynamics (Kalliola et al. 1991, 1992), deforestation and forest degradation (Skole and Tucker 1993; Souza et al. 2013; Potapov et al. 2014; Hansen et al. 2016), forest fires (Alencar et al. 2011; Morton et al. 2011; Cardozo et al. 2014), and secondary forest succession (DeVries et al. 2015; Müller et al. 2016).

At the landscape scale, several studies carried out in Amazonia have documented that both gradual and abrupt spectral patterns in Landsat TM/ETM+ imagery can be related to changes in floristic composition or soil properties (Tuomisto et al. 2003a,c; Salovaara et al. 2005; Thessler et al. 2005; Anderson et al. 2009; Higgins et al. 2012; Sirén et al. 2013; Ceddia et al. 2017). This suggests that Landsat images could be used as an environmental layer in species distribution modelling and other applications that require *in situ* data to be interpolated across large areas. However, studies that have combined field inventories with quantitative spectral data within primary Amazonian forests have typically been limited to a single Landsat scene. Those studies that have aimed to detect vegetation heterogeneity or other biodiversity patterns at broader spatial scales and across multiple Landsat scenes have been limited to visual interpretation of the most obvious patterns (Tuomisto et al. 1995; Higgins et al. 2011). Therefore, Landsat data have remained relatively underexploited in broad-scale studies that require high thematic detail over the Amazon region, and their potential usefulness has not been realized.

One obvious reason for the limited use of Landsat imagery in large-scale ecological and conservation studies is that the data are available as individual scenes of *c.* 180 km by 180 km, and larger study areas can only be covered by combining several such images. Scenes in adjacent flight paths are acquired on different days, so they differ in atmospheric conditions. This is a big problem for studies that require high thematic resolution, because the atmospheric effects can easily be larger than the subtle spectral differences between the terrain features of interest. Even within a single Landsat scene, spectral artefacts caused by the viewing geometry can be larger than the actual spectral differences between different forest types (Toivonen et al. 2006; Higgins et al. 2015).

To minimize these problems for the end-users, considerable effort has been invested in automated cloud screening (Zhu and Woodcock 2012) and atmospheric correction (Masek et al. 2006) of Landsat data. In addition, composite imagery has been made available by the Web-enabled Landsat Data (WELD) project (Roy et al. 2010) which used all globally available Landsat TM and ETM+ images to generate composite images that are provided to the scientific community through an online portal.

Many existing Landsat compositing approaches generate image composites for repeated monthly or yearly time steps, or incorporate the temporal aspect in the compositing procedure itself (Roy et al. 2010; Griffiths et al. 2013; White et al. 2014; Hermosilla et al. 2015). These are particularly useful for studies with a temporal dimension. In the wet tropics, however, cloud cover is so persistent that a non-negligible fraction of pixels may not have any cloud-free observations within such short compositing periods, which leaves gaps in the image composite. In many ecological and conservation applications, the end users' needs would be best served by a single, gap-free and radiometrically consistent image, even if the compositing period had to be extended to achieve this.

Here, we present a new Landsat image composite for the entire Amazon rain forest biome that uses a compositing period of a full decade in order to minimize the occurrence of data gaps. It has been generated specifically to serve the needs of such applications that aim to understand relatively stable features of forest heterogeneity, since these can be expected to be most important for spatial ecology, biogeography and conservation planning. In the following, we explain the image processing and compositing method in some detail and provide a visual assessment of the end product. In order to assess its utility for continental-scale applications in ecology, biodiversity research and conservation planning, we tested how well Landsat reflectance can predict the concentration of exchangeable base cations in the surface soil in central and western Amazonia. The availability of these plant nutrients has direct influence on plant physiology and growth. Earlier ecological studies have found that variation in soil cation concentration is an excellent overall predictor of plant species turnover in Amazonian forests, and that these patterns are highly correlated with Landsat reflectance patterns at the landscape scale (Tuomisto et al. 2003a,b,c, 2016; Higgins et al. 2011). A radiometrically consistent image composite should, therefore, be able to predict variation in soil cation concentration also at broader spatial scales. If the image composite is not radiometrically consistent, artefacts related to directional effects and atmospheric disturbances would swamp the signal and lead to poor predictive power.

Materials and Methods

Landsat data availability over the Amazon biome

We decided to cover the entire Amazon biome as defined by WWF.[1] This includes tropical rain forests both in the

[1]http://wwf.panda.org/what_we_do/where_we_work/amazon/about_the_amazon/.

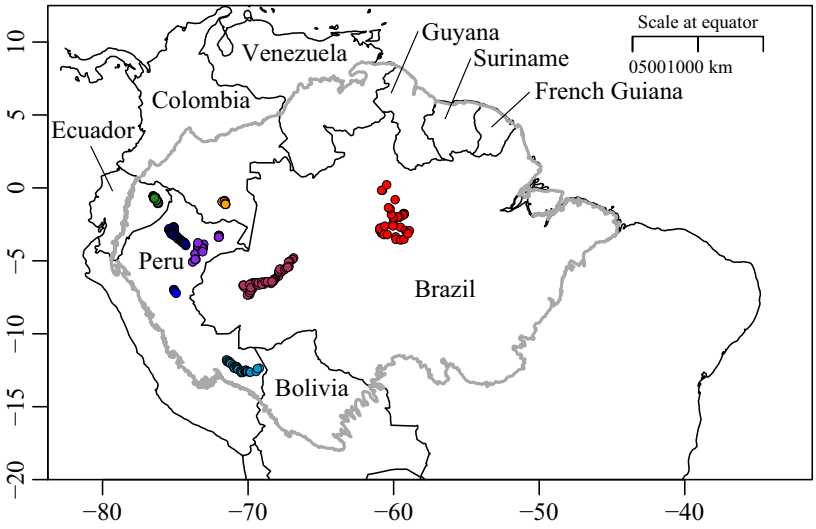

Figure 1. The Amazon biome (thick grey line), with locations of the field study sites described in Section 2.4 indicated. Sites from different regions are shown in different colours to facilitate assessing Figure 10.

Amazon river basin and in some adjacent forested ecoregions, and covers *c.* 6.7 million square kilometres (Fig. 1). The time period over which to carry out the image compositing is a compromise between the desire to avoid temporal and phenological changes within the image and the need to ensure sufficient data availability in spite of the high incidence of clouds. We settled for the 10-year period 2000–2009, because both Landsat 5 and Landsat 7 were operating at that time, which maximized potential data availability. We then searched the U.S. Geological Survey's (USGS) Earth Resources Observation and Science (EROS) center EarthExplorer archive[2] for all Level 1T processed TM/ETM+ images that had <60% cloud cover and had been acquired within the defined spatial and temporal limits.

Overall data availability was clearly higher over Brazilian Amazonia, especially over its south-eastern part, than over northern and western Amazonia and the Guianas (Fig. 2). This is partly due to more persistent cloud cover in the north-west, but historic image acquisition and storage strategies have also played an important role. In particular, the ground station in Cuiabá, Brazil, operated by the Brazilian National Institute for Space Research (INPE), holds significant amounts of Landsat data over the Brazilian, Peruvian and Bolivian territories, and these have recently been incorporated into the USGS archive (Wulder et al. 2016). In contrast, the ground receiving stations in Cotopaxi, Ecuador, and Mayaguez, Puerto Rico, which cover the northern and western parts of the Amazon biome, ceased receiving Landsat data in 2001 and 2003 respectively.[3]

[2]http://earthexplorer.usgs.gov.

[3]https://landsat.usgs.gov/historical-international-ground-stations.

The monthly availability of relatively cloud-free Landsat imagery shows a clear seasonal pattern (Fig. 3) that is related to the position of the Intertropical Convergence Zone and the alternation of dry and rainy seasons caused by it. Consequently, overall data availability is much lower from December to May than from July to October.

We decided to use for the compositing only images from the 3 months with highest data availability: July, August and September. Focusing on just one season avoids the undesired introduction of confounding factors that could be introduced to the compositing process by changes in vegetation phenology. A 10-year compositing period was deemed adequate to guarantee sufficient data availability without risking excessive temporal change.

Scene-based processing

We downloaded more than 16 000 Landsat TM and ETM+ Level 1T images over the Amazon biome for the selected time period through the USGS/EROS Inventory Service, totalling more than 2.5 TB of compressed data. At the time of image processing, the migration of the USGS Landsat archive to Collection 1 Tiers was not yet completed, so we downloaded the pre-collection products. Each scene was individually processed to directionally and topographically normalized surface reflectance through the processing chain summarized in Figure 4.

Atmospheric correction

The Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Masek et al. 2006, 2013) was used to generate surface reflectance images. LEDAPS

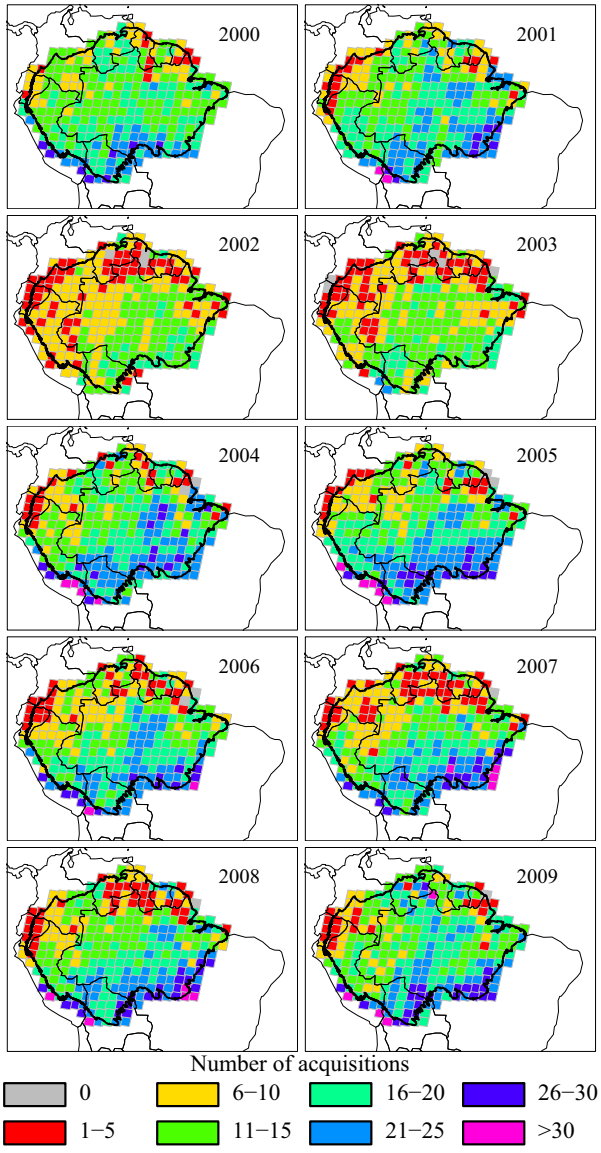

Figure 2. Number of Landsat TM/ETM+ acquisitions with <60% cloud cover available in the USGS EarthExplorer archive for each WRS-2 scene per year in the period 2000–2009. USGS, U.S. Geological Survey's.

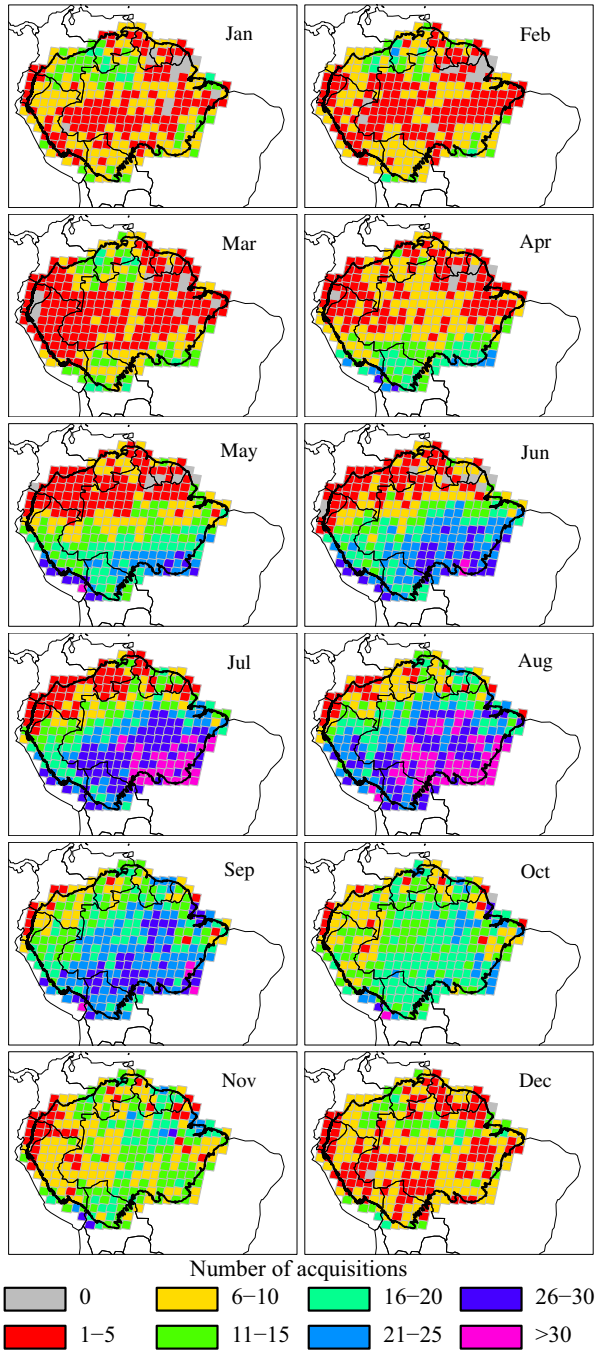

Figure 3. Number of Landsat TM/ETM+ acquisitions with <60% cloud cover available in the USGS EarthExplorer archive for each WRS-2 scene per month in the period 2000–2009. USGS, U.S. Geological Survey's.

applies the Second Simulation of a Satellite Signal in the Solar Spectrum (6S) radiative transfer model (Vermote et al. 1997) on Level 1 data products and estimates of water vapor, ozone, geopotential height, aerosol optical thickness and elevation. We used LEDAPS Version 2.5.0[4] (March 2016), in combination with the ESPA product formatter Version 1.6.0[5] (March 2016).

In addition to surface reflectance layers for each spectral band, LEDAPS software generates quality assessment (QA) layers for cloud cover, pixels adjacent to cloud cover and cloud shadow, as well as the CFMASK (Zhu and Woodcock 2012). We combined the LEDAPS QA layers and the CFMASK to create a single mask for clouds

[4]https://github.com/USGS-EROS/espa-surface-reflectance.
[5]https://github.com/USGS-EROS/espa-product-formatter.

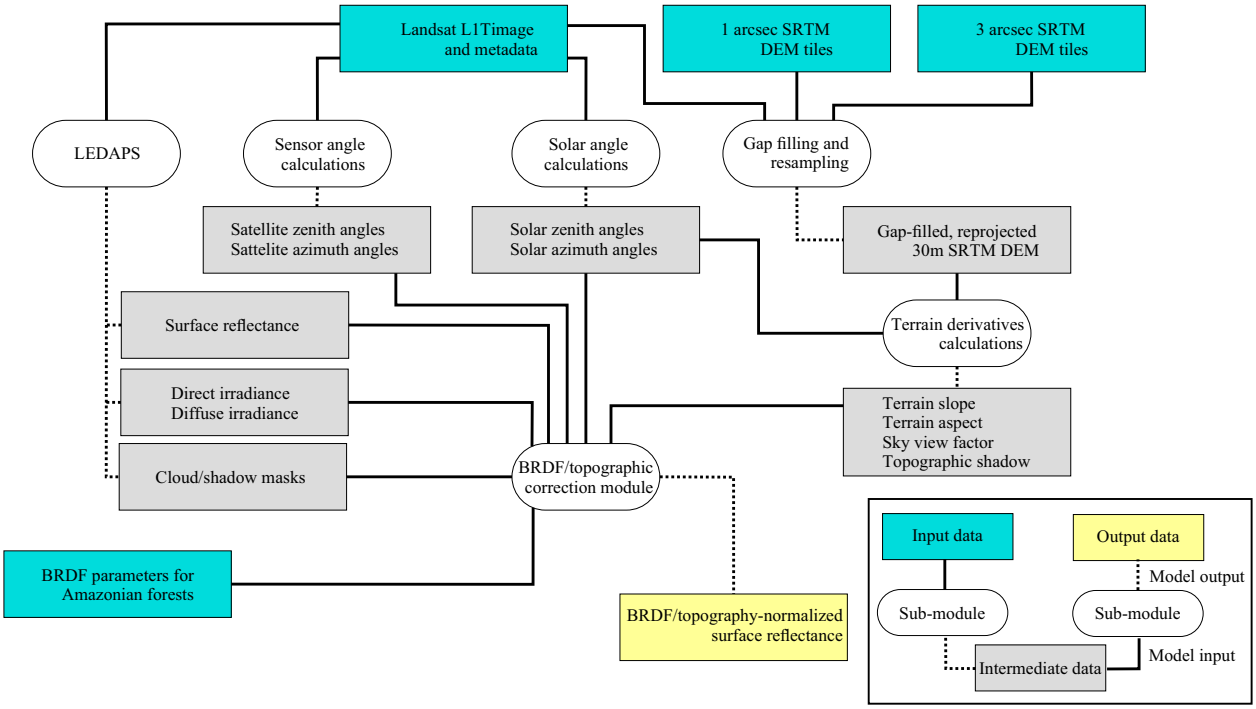

Figure 4. Flow chart for scene-based image processing.

and shadows. This is a rather conservative way of masking, but minimizes influence of possible clouds or shadows on the image composite.

The LEDAPS software derives a large number of intermediate parameters, including aerosol optical thickness and atmospheric transmittance for each band. We modified the LEDAPS source code to provide these parameters as outputs, and combined them with estimates of solar exoat-mospheric irradiance for each spectral band,[6] sun-earth distance for the day of image acquisition and solar zenith angle at the scene centre, to obtain estimates of direct and diffuse irradiance.

Angular calculations

Solar zenith and azimuth were calculated for each pixel in a scene with an astronomical model (Reda and Andreas 2004, 2007) that uses the pixel's latitude and longitude, the acquisition date and the time at the scene centre, which are supplied in the image metadata. Landsat ephemeris information, required for the calculation of the sensor viewing vector, is not provided in the Pre-Collection products. To obtain the sensor zenith and azimuth angle, we therefore used an approximation based on the coordinates of the scene corners in the near infrared band (Van doninck and Tuomisto 2017a).

Terrain derivatives

We used the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) to calculate terrain derivatives. The SRTM DEM is provided through USGS EarthExplorer at 1 arcsec spatial resolution, in tiles of 1 degree by 1 degree. The global 1 arcsec SRTM DEM may still contain void pixels. We therefore used the void-filled 3 arcsec SRTM DEM to replace these missing values. The gap-filled DEM was then resampled to the 30 m spatial resolution Universal Transverse Mercator (UTM) coordinate system and extent of the processed Landsat scene, and terrain slope and aspect were calculated for each pixel using the eight surrounding pixels. We also derived the sky view factor for each pixel (Dozier and Frew 1990), and masked self shadow and cast shadow pixels based on the set of solar angles obtained earlier.

Directional normalization

Landsat images over the tropics are strongly influenced by effects of the bidirectional reflectance distribution function (BRDF). The BRDF describes the scattering of light as a function of the relative orientation of sun,

[6]landsat.usgs.gov/esun.

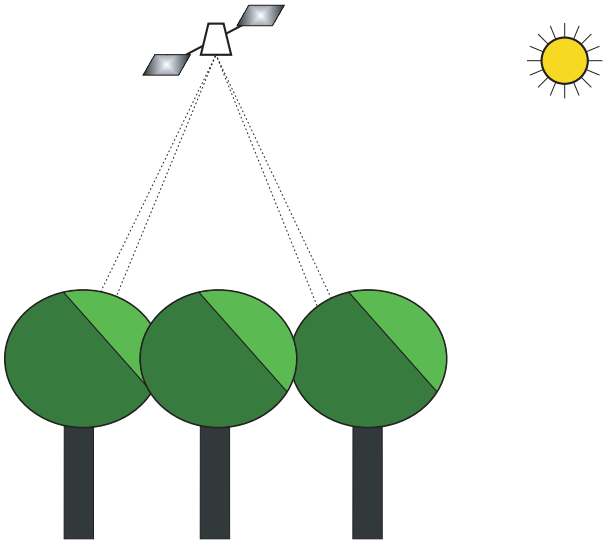

Figure 5. Origin of bidirectional effects in forests: the sensor's view angle changes along the scan line from one side of each scene to the other. Depending on the relative position of the view angle and the solar incidence angle, the sensor will register different fractions of radiation from the sunlit and shaded parts of tree canopies and other surface features.

sensor and terrain (Fig. 5), and over tropical forests it introduces an artificial gradient whereby reflectance increases from the eastern to the western edge of each Landsat scene (Toivonen et al. 2006; Nagol et al. 2015; Muro et al. 2016). Since directional effects on surface reflectance can be larger than the differences in reflectance between floristically different vegetation types (Higgins et al. 2012; Muro et al. 2016; Van doninck and Tuomisto 2017a), they have to be removed from each scene individually prior to image compositing.

We combined the corrections for the effects of sun-sensor geometry and topography into a single BRDF/topographic model, which was adopted from the methodology presented by Flood et al. (2013). This model normalizes surface reflectance using as inputs the direct and diffuse irradiance for each Landsat band, the sky view factor, incidence and exitance angles for each pixel, and a set of BRDF parameters. We calculated incidence and exitance angles from the solar and sensor angles and the terrain slope and aspect. The BRDF model parameters describe the change in surface reflectance with changing sun-terrain-sensor geometry, and we developed a method to derive them specifically for Amazonian forests (Van doninck and Tuomisto 2017a).

The BRDF/topographic model was used to normalize directional surface reflectance to nadir viewing geometry, horizontal terrain and a solar zenith angle of 30°. The latest WELD version uses a spatially varying modelled zenith angle (Zhang et al. 2016), but we chose to normalize to a fixed solar zenith angle to ensure that identical land surfaces have similar surface reflectance irrespective of their geographical location.

Tile-based compositing

After individually processing each Landsat Level 1 scene to directionally and topographically normalized surface reflectance, a second step combined all these scenes into a composite image covering the entire Amazon biome. Because the file size of the entire composite image would be much too large to handle on most personal computers, we adopted a tile-based approach. The Amazon biome was divided into 2.5 degree by 2.5 degree tiles with an 1 arcsec (*c.* 30 m) spatial resolution. Each tile has 9001 rows and 9001 columns, so that adjacent tiles always have either one row or one column of overlapping pixels.

All the processed Landsat image acquisitions that at least partially overlapped a given tile were reprojected to the tile's geometry using bilinear interpolation, thus creating a stack of multitemporal observations over the tile. Duplicate acquisitions in the overlap area of successive scenes in the same satellite orbit were first removed. The best available observation to be stored in the composite image for each pixel was then selected from this stack. If only a single un-masked observation was available for a target pixel, that acquisition was stored in the composite image. If there were two observations to select from, the reflectance of the six spectral bands corresponding to the observation with the higher NDVI was selected. For three or more available observations, the medoid criterion (Flood 2013) was applied. This method selects the observation with the smallest sum of euclidean distances (in six-dimensional spectral space) to all other observations. When comparing different pixel-based compositing methods, we found that the medoid criterion gave radiometrically more consistent composite images than other proposed selection methods (Van doninck and Tuomisto 2017b). If all observations for a pixel were masked due to cloud cover or cloud shadow, a 'no data' value was assigned.

Our tile-based compositing algorithm provides, in addition to the multispectral composite image, a layer reporting how many unmasked (i.e. cloud- and shadow-free) observations were available to select from for each pixel, and another layer giving the acquisition year and date (in the format YYYYDDD) of the selected observation.

Image composite evaluation with edaphic field data

The Landsat TM/ETM+ image composite presented here was developed to serve as an environmental data layer for applications in ecology, biodiversity research and

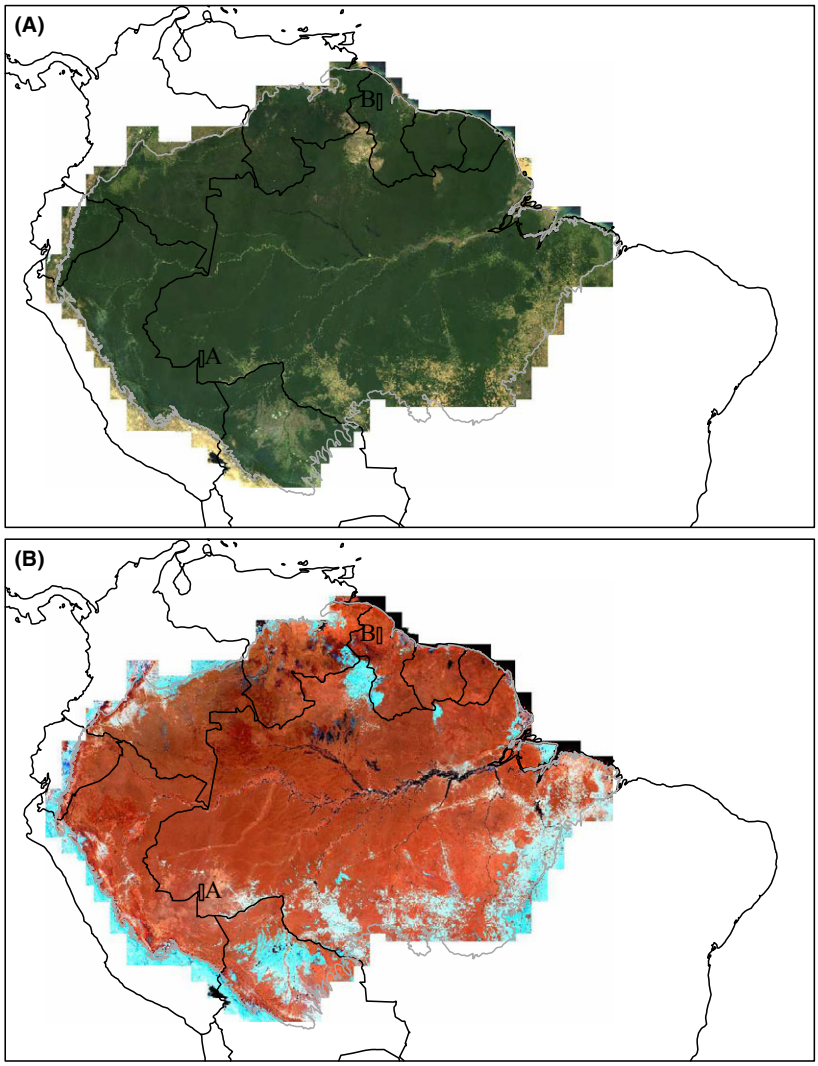

Figure 6. Landsat TM/ETM+ composite image over the Amazon biome (limits in thick grey line) based on image acquisitions from July–September in the years 2000–2009. (A) True colour composite. (B) False colour composite (red: band 4, green: band 5, blue: band 7). The boxes labelled A and B indicate the locations of composite image details shown in Figures 8 and 9 respectively.

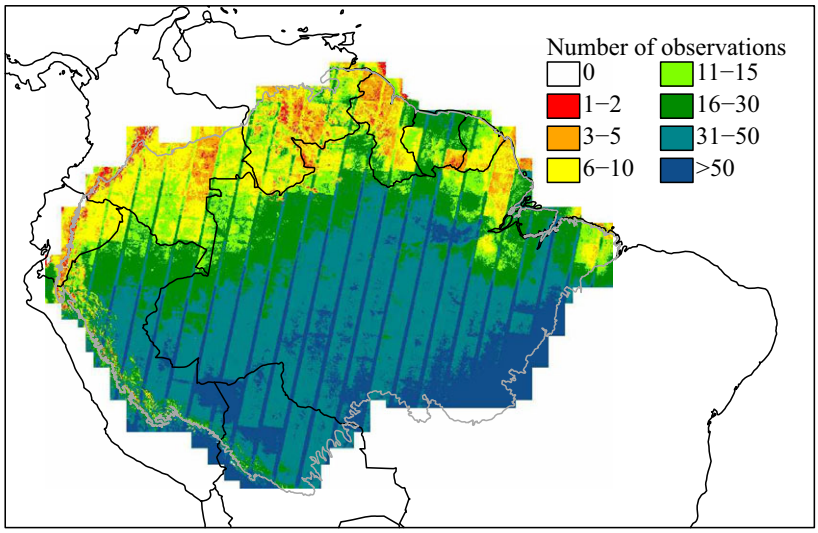

Figure 7. Number of unmasked observations (i.e. free of clouds and cloud shadows) in the Landsat TM/ETM+ acquisitions that were available for the image compositing period from July to September in the years 2000–2009 over the Amazon biome (limits in grey line).

conservation planning that are primarily interested in relatively stable aspects of landscape and forest heterogeneity, and require simultaneously a large spatial extent and high thematic and spatial resolution. To assess the utility of the composite for such purposes, we tested how well it is able to predict the concentration of exchangeable base cations (Ca, Mg, K and Na) in the surface soil (here after referred to as soil cation concentration). As ground truth data, we used data from 327 line transects established in central and western Amazonia (Fig. 1). The field data were collected during several field campaigns carried out between 1994 and 2012 in Brazil, Colombia, Ecuador and Peru, and they span an area of *c.* 2000 km East to West, and 1500 km North to South. Each field inventory consisted of a 500 m line transect that was georeferenced in the field using a hand-held GPS receiver. The transects were aimed to be representative of the local landscape, so they were placed such that they ran more or less perpendicularly across the local topography and thereby sampled both the drier hill tops and the moister valley bottoms. Surface soil samples (top 5 cm of the mineral soil) were collected at three points along each transect: near the beginning, midpoint and end, making sure that they represented both the higher and the lower parts of the local topography. Each soil sample consisted of five subsamples collected within an area of *c.* 5 m by 5 m and pooled. The cations Ca, Mg, K and Na were extracted with 1 mol L^{-1} ammonium acetate at pH 7. Their sum (in cmol(+) kg^{-1}) was first calculated for each soil sample separately and then averaged per transect. More details on the methodology of data collection can be found in previous publications (Tuomisto et al. 2003a,c, 2014, 2016; Higgins et al. 2011).

To extract reflectance data from the image composite, a window of 15 by 15 pixels (*c.* 450 m by 450 m) was centred on the midpoint coordinates of each transect. The average reflectance value within that window was then extracted for each spectral band separately. A minimum NDVI threshold of 0.6 was used to exclude non-forested pixels within the window before calculating the average.

Correspondence between Landsat reflectance and logarithmically transformed soil cation concentration was assessed using a simple linear model. The six Landsat surface reflectance bands were used as independent variables and soil cation concentration as the dependent variable. Performance of the model was assessed using leave-one-out cross-validation.

Results and Discussion

Visual assessment

Two versions of the Amazon-wide composite image are shown in Figure 6, a true colour composite based on the visible bands and a false colour composite based on the infrared bands. Both images show a clear contrast between the forested areas within the Amazon biome and the areas that are not forest-covered either naturally or due to deforestation. The channels and floodplains of the largest rivers also stand out.

Otherwise, the true colour composite shows little spectral variability within the primary forests, and the next most prominent visual feature becomes the alternation of brighter and darker stripes that correspond to Landsat flight paths. This indicates that the pre-processing of the images has not entirely succeeded in eliminating the atmospheric effects from the visible bands, especially in the northern areas with relatively poor availability of could-free data. Pre-processing of the infrared bands was more successful. The seams between scenes are hardly noticeable in the false colour composite, and regional patterns in surface features are more easily identifiable. Earlier vegetation studies have found infrared bands to be more informative than visible bands, even within a single scene (Tuomisto et al. 2003a,b; Salovaara et al. 2005).

The number of observations available for compositing varied greatly (Fig. 7), making it difficult to select a single compositing interval for the entire Amazon biome. With the medoid compositing criterion, each additional observation per pixel improves the radiometric consistency of the composite image (Van doninck and Tuomisto 2017b), but this effect becomes marginal after about 15 multitemporal observations. Over most of the Brazilian, Peruvian and Bolivian Amazonia, data abundance can be expected to result in a highly consistent composite, but this comes at the cost of unnecessarily heavy data processing. On the other hand, for large parts of Colombia, Venezuela and the Guianas, the composite is based on <5 unmasked observations per pixel, so additional data would have improved the result. Data availability was poorest in the mountainous parts of Ecuador, Colombia and Venezuela, where some areas had no cloud-free observations at all during the compositing period.

In spite of these problems, our composite image allows a better discrimination of land surface characteristics than the product generated by the WELD project (Roy et al. 2010). Our image composite and the WELD composite for the year 2010 (version 3.0) can be compared over an area with good data availability in Figure 8 and over an area with poor data availability in Figure 9 (data availability is shown in Fig. 7). From a user's perspective, what matters is the quality of the available end product, and differences in this respect are quite large between the WELD mosaic and our mosaic. Of course, the two composites were generated for different time periods and using different numbers of images, so from the producer's perspective they are not strictly comparable.

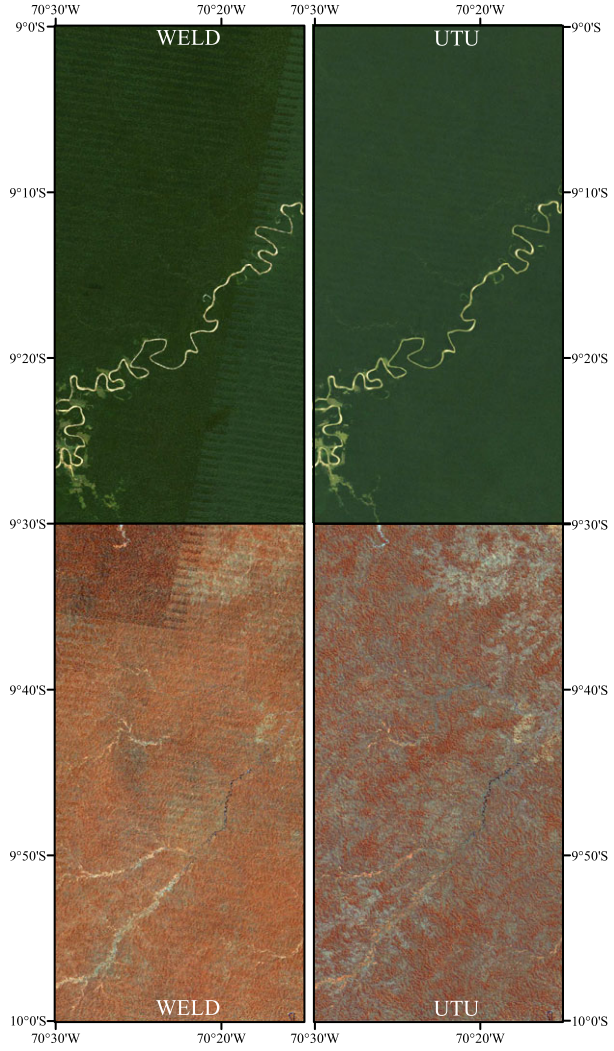

Figure 8. True colour composite (top) and false colour composite (bottom; red: band 4, green: band 5, blue: band 7) image details from the Web-Enabled Landsat Data composite (left) and the new Amazon-wide image composite presented here (right), for a region with good availability of Landsat acquisitions (rectangle labelled 'A' in Fig. 6).

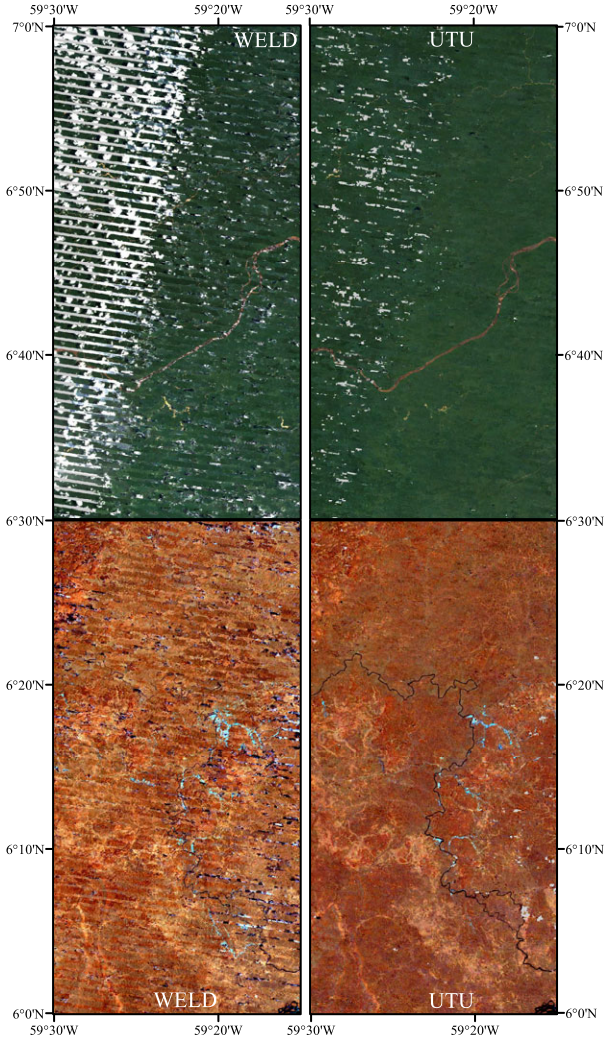

Figure 9. True colour composite (top) and false colour composite (bottom; red: band 4, green: band 5, blue: band 7) image details from the Web-Enabled Landsat Data composite (left) and the new Amazon-wide image composite presented here (right), for a region with poor availability of Landsat acquisitions (rectangle labelled 'B' in Fig. 6). In the WELD composite, grey pixels indicate that no observations were available and white pixels correspond to unmasked cloudy observations. In our composite, gray pixels indicate that no unmasked (i.e. cloud- and shadow-free) observations were available. WELD, Web-enabled Landsat Data.

Therefore, Figure S1 in Supporting Information also shows a window composited for the same time period as the WELD product but using the methodology described here.

Although the latest version of WELD includes a correction for directional effects (Roy et al. 2016), a clear discontinuity in surface reflectance is still visible at the seam between adjacent Worldwide Reference System (WRS-2) scenes in the data-rich example area (Fig. 8). Aside from an imperfect directional normalization, such artefacts may also originate from an inadequate atmospheric correction in combination with a suboptimal compositing algorithm. Because the magnitude of the artefacts is often visibly larger than the magnitude of the spectral differences in interest, the WELD image composites are not useful for practical applications that focus on vegetation heterogeneity. In contrast, our composite image is practically free of these artefacts in areas of high data abundance (Fig. 8) thanks to the use of a large number of images combined with the medoid criterion. Some artifactual striping is still apparent in the visible bands, but in the infrared bands

all the observable features correspond to natural vegetation patterns. Artefacts are also mostly eliminated when using our approach over a shorter compositing period (Fig. S1).

Artifactual disturbances are more prominent in regions with poor data availability. The area covered in Figure 9 had <10 observations per pixel during our compositing period, and for several pixels there was not a single cloud-free observation. The quality, and therefore also usefulness, of our image composite is clearly lower in data-poor than in data-rich areas. However, the land cover patterns can still be identified more readily in our composite than in the WELD image over the same area.

The radiometric consistency could be increased, and the number of 'no data' pixels reduced, by extending the compositing period. However, this would have the unwanted consequence that redundant image processing would be performed for data-rich regions. One possible solution for this problem would be to use a dynamic compositing length, with the compositing period defined as a function of the overall availability of cloud-free observations. For example, one could select a target year, and increase the time window around it for each WRS-2 scene individually until a predefined number of images with a selected maximum cloud threshold has been reached. Alternatively, our methodology can be used to generate a new composite image for any sub-area of the Amazon biome such that the temporal constraints are defined on the basis of local data availability.

Validation with edaphic data

Reflectance values extracted from our basin-wide Landsat image composite were able to predict the log-transformed soil cation concentration relatively well with a simple linear model: the coefficient of determination between observed and predicted values was 0.59 (Fig. 10). This can be considered a surprisingly high overall accuracy, given that the field sites span an area of 1500 km by 2000 km in central and western Amazonia.

The Landsat TM/ETM+ bands with the highest contributions to the predictive model were the red band (3) and the two shortwave infrared bands (5 and 7), whereas the blue band (1) had no predictive power at all (Fig. 11). This is in line with the fact that atmospheric disturbance is strongest in the short wavelengths (see Figs. 6, 8, and 9). More detailed studies and more field data will be needed to assess to what degree more flexible models could further improve the predictive power of the spectral data.

Testing the usefulness of the Landsat image composite with soil cation concentration can be considered a stringent test, because Landsat registers reflectance of the top surface rather than soil properties. For any relationship between

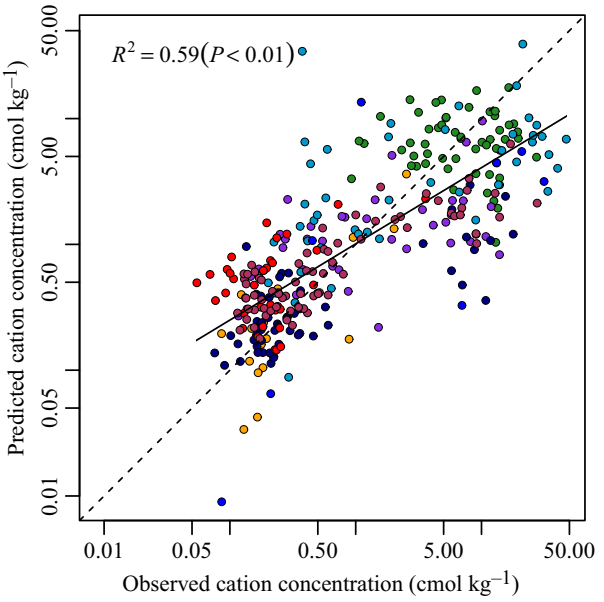

Figure 10. The relationship between predicted and observed concentration of exchangeable base cations in the surface soil for field sampling sites across Amazonia. Explanatory variables in the linear regression model were Landsat TM/ETM+ spectral values from the Amazon-wide image composite and leave-one-out cross validation was used. Sites from different regions are shown in different colours as specified in Figure 1.

soils and surface reflectance to exist at the basin-wide extent, there has to be an intermediary that is causally related to both. The obvious candidate intermediaries are plant (especially tree) species composition and forest structure. Earlier ecological studies have indeed found that these strongly covary with soil cation concentration in Amazonia (Phillips et al. 2003; Tuomisto et al. 2003a,b,c, 2016; Ruokolainen et al. 2007; Higgins et al. 2011, 2015; Baldeck et al. 2016; Cámara-Leret et al. 2017). The earlier studies were carried out at relatively restricted spatial extents. Our results suggest both that similar soil-vegetation relationships apply across the basin and that our method has succeeded in producing a Landsat image composite that captures the environmental information in a relevant way.

Although tree species distribution models (SDMs) in Amazonia have used soil cation exchange capacity and soil pH as predictive environmental data layers (Levis et al. 2017), both of these are poor surrogates of nutrient availability in Amazonian soils (Quesada et al. 2010; Moulatlet et al. 2017). Our results suggest that it may be possible to combine Landsat imagery with *in situ* observations of soil cation concentration to obtain spatially extensive soil nutrient concentration proxies for entire Amazonia at medium spatial resolution. Such a layer could be more informative as an abiotic predictor variable in SDMs than the currently available variables.

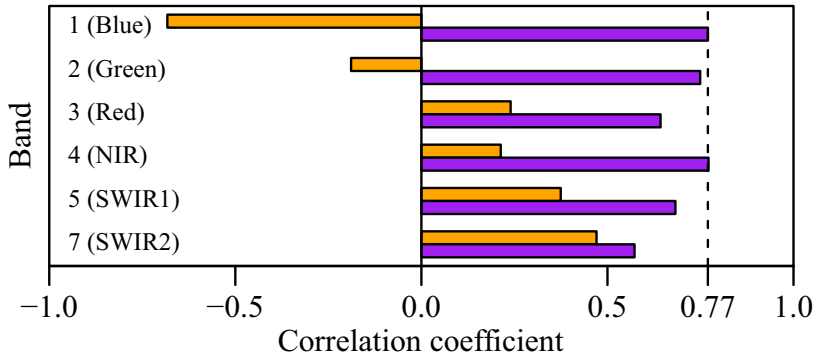

Figure 11. Assessment of variable importance of the different Landsat TM/ETM+ bands as predictors of the concentration of exchangeable base cations in the surface soil for field sampling sites across Amazonia. The correlation coefficient between the predicted and observed soil cation concentration (both log-transformed) is shown both when each band is used in a linear model on its own (orange) and when that band is omitted from a model including the other bands (purple).

Furthermore, data from the radiometrically highly consistent Landsat composite can be used in SDMs directly, either as original reflectance values or in the form of derived indices (He et al. 2015). The 30 m spatial resolution of Landsat means that the gap between the scales of *in situ* data acquisition and of satellite data is relatively small. This is in contrast to lower resolution imagery, such as AVHRR and MODIS, whose pixel size is much larger than the typical field data site. On the other hand, higher resolution data mostly lack basin-wide coverage. The spatial resolution of Landsat corresponds to the size of a single or a few tree crowns, which also makes it possible to derive spectral heterogeneity metrics that can then be related to biodiversity estimates (Rocchini et al. 2016).

The Amazon-wide Landsat composite presented here only combines data for 3 months of the year, namely July–September. This restriction was meant to focus the compositing to the driest part of the year and to avoid phenological changes within the compositing period. No doubt there are applications for which information on seasonal phenological changes would be useful. Such information can be obtained simply by applying the same compositing method to different seasons. This would allow using information on yearly averages together with seasonal variability in ecological and biodiversity studies.

Summary and Conclusions

The Landsat TM/ETM+ image composites that have previously been available over the Amazon region generally suffer from a significant fraction of missing observations and a poor radiometric consistency. Although they can be useful for some applications, they are of limited utility for studies that require a high thematic resolution. Here, we have presented a new Landsat composite for applications in ecology, biodiversity research and conservation planning over the Amazon biome.

Compromising on temporal resolution, our Landsat composite image represents the months with maximum data availability over the entire Amazon region – July, August and September – during the years 2000 through 2009. First, we individually processed all Landsat TM and ETM+ images acquired within this time period to directionally and topographically normalized surface reflectance. Then, we applied a pixel-based compositing approach to combine all available data into 2.5 degree tiles. Use of the medoid compositing method ensured that the effects of artefacts, such as residual atmospheric contamination, were minimized.

Despite the long compositing period, some parts of the composite image still have data gaps or poor radiometric consistency. This is especially observable in the northern part of the Amazon biome, where historic data acquisition and archiving strategies, together with persistent cloud cover, limited data availability. For studies specifically targeting these areas, the compositing algorithm could be run again for a different time period in order to increase data availability and improve the radiometric quality of the final product. However, even in data-poor areas, our image composite is radiometrically more consistent than the WELD composite, and the difference is especially large in regions with good data availability. Our compositing method seems to have succeeded in eliminating directional effects and atmospheric contamination rather well, since the reflectance values from the image composite were able to predict field-measured soil cation concentrations in field sites spanning 2000 km in the east-west direction. Our image composite provides an unprecedented combination of high spatial resolution, large spatial extent and high thematic accuracy over the Amazon biome. Such data should be useful for many studies focusing on spatial ecology, biogeography, species distribution modelling and conservation planning in Amazonia.

Acknowledgments

Image analyses were carried out using the computing facilities provided by CSC-IT Center For Science. Field

data on soils were accumulated during many previous projects, and we are grateful to numerous collaborators, especially Kalle Ruokolainen, for their participation in those projects. The present research was funded by a grant from the Academy of Finland to Hanna Tuomisto.

Data Accessibility

Original Landsat TM/ETM+ data are available from the U.S. Geological Survey. The total file size limits the possibilities to make the image composite over the Amazon biome available through a public data repository, but it can be accessed through agreement with the authors (details at www.utu.fi/amazon).

References

Alencar, A., G. P. Asner, D. Knapp, and D. Zarin. 2011. Temporal variability of forest fires in eastern Amazonia. *Ecol. Appl.* **21**, 2397–2412.

Anderson, L. O., Y. Malhi, R. J. Ladle, L. E. O. C. Aragão, Y. Shimabukuro, O. L. Phillips, et al. 2009. Influence of landscape heterogeneity on spatial patterns of wood productivity, wood specific density and above ground biomass in Amazonia. *Biogeosciences* **6**, 1883–1902.

Baldeck, C. A., R. Tupayachi, F. Sinca, N. Jaramillo, and G. P. Asner. 2016. Environmental drivers of tree community turnover in western Amazonian forests. *Ecography* **39**, 1089–1099.

Cámara-Leret, R., H. Tuomisto, K. Ruokolainen, H. Balslev, and S. Munch Kristiansen. 2017. Modelling responses of western Amazonian palms to soil nutrients. *J. Ecol.* **105**, 367–381.

Cardozo, F. D. S., G. Pereira, Y. E. Shimabukuro, and E. C. Moraes. 2014. Analysis and assessment of the spatial and temporal distribution of burned areas in the Amazon forest. *Remote Sens.* **6**, 8002–8025.

Ceddia, M. B., A. S. Gomes, G. M. Vasques, and E. F. M. Pinheiro. 2017. Soil carbon stock and particle size fractions in the central amazon predicted from remotely sensed relief, multispectral and radar data. *Remote Sens.* **9**, 124.

DeVries, B., M. Decuyper, J. Verbesselt, A. Zeileis, M. Herold, and S. Joseph. 2015. Tracking disturbance-regrowth dynamics in tropical forests using structural change detection and Landsat time series. *Remote Sens. Environ.* **169**, 320–334.

Dozier, J., and J. Frew. 1990. Rapid calculation of terrain parameters for radiation modeling from digital elevation data. *IEEE Trans. Geosci. Remote Sens.* **28**, 963–969.

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

Flood, N., T. Danaher, T. Gill, and S. Gillingham. 2013. An operational scheme for deriving standardised surface reflectance from Landsat TM/ETM+ and SPOT HRG imagery for Eastern Australia. *Remote Sens.* **5**, 83–109.

Gillespie, T. W., G. M. Foody, D. Rocchini, A. P. Giorgi, and S. Saatchi. 2008. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* **32**, 203–221.

Griffiths, P., S. van der Linden, T. Kuemmerle, and P. Hostert. 2013. A pixel-based landsat compositing algorithm for large area land cover mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**, 2088–2101.

Hansen, M. C., A. Krylov, A. Tyukavina, P. V. Potapov, S. Turubanova, B. Zutta, et al. 2016. Humid tropical forest disturbance alerts using Landsat data. *Environ. Res. Lett.* **11**, 034008.

He, K. S., B. A. Bradley, A. F. Cord, D. Rocchini, M. N. Tuanmu, S. Schmidtlein, et al. 2015. Will remote sensing shape the next generation of species distribution models? *Remote Sens. Ecol. Conserv.* **1**, 4–18.

Hermosilla, T., M. A. Wulder, J. C. White, N. C. Coops, and G. W. Hobart. 2015. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sens. Environ.* **158**, 220–234.

Higgins, M. A., K. Ruokolainen, H. Tuomisto, N. Llerena, G. Cardenas, O. L. Phillips, et al. 2011. Geological control of floristic composition in Amazonian forests. *J. Biogeogr.* **38**, 2136–2149.

Higgins, M. A., G. P. Asner, E. Perez, N. Elespuru, H. Tuomisto, K. Ruokolainen, et al. 2012. Use of Landsat and SRTM data to detect broad-scale biodiversity patterns in Northwestern Amazonia. *Remote Sens.* **4**, 2401–2418.

Higgins, M. A., G. P. Asner, C. B. Anderson, R. E. Martin, D. E. Knapp, R. Tupayachi, et al. 2015. Regional-scale drivers of forest structure and function in Northwestern Amazonia. *PLoS One* **10**, e0119887.

Kalliola, R., M. Puhakka, J. Salo, H. Tuomisto, and K. Ruokolainen. 1991. The dynamics, distribution and classification of swamp vegetation in Peruvian Amazonia. *Ann. Bot. Fenn.* **28**, 225–239.

Kalliola, R., J. Salo, M. Puhakka, M. Rajasilta, T. Hme, R. J. Neller, et al. 1992. Upper amazon channel migration. *Naturwissenschaften* **79**, 75–79.

Kerr, J. T., and M. Ostrovsky. 2003. From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* **18**, 299–305.

Levis, C., F. Costa, F. Bongers, M. Peña Claros, C. Clement, A. Junqueira, et al. 2017. Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science* **355**, 925–931.

Masek, J. G. M., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, et al. 2006. A Landsat surface

reflectance dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* **3**, 68–72.

Masek, J., E. Vermote, N. Vermote, R. Wolfe, F. Hall, K. Huemmrich, et al. 2013. LEDAPS calibration, reflectance, atmospheric correction preprocessing code, version 2. https://doi.org/10.3334/ornldaac/1146.

Morton, D. C., R. S. DeFries, J. Nagol, C. M. Souza Jr, E. S. Kasischke, G. C. Hurtt, et al. 2011. Mapping canopy damage from understory fires in Amazon forests using annual time series of Landsat and MODIS data. *Remote Sens. Environ.* **115**, 1706–1720.

Moulatlet, G. M., G. Zuquim, F. O. G. Figueiredo, S. Lehtonen, T. Emilio, K. Ruokolainen, et al. 2017. Using digital soil maps to infer edaphic affinities of plant species in Amazonia: problems and prospects. *Ecol. Evol.* **7**, 8463–8477.

Müller, H., P. Rufin, P. Griffiths, L. de Barros Viana Hissa, and P. Hostert. 2016. Beyond deforestation: differences in long-term regrowth dynamics across land use regimes in southern Amazonia. *Remote Sens. Environ.* **186**, 652–662.

Muro, J., J. Van doninck, H. Tuomisto, M. A. Higgins, G. M. Moulatlet, and K. Ruokolainen. 2016. Floristic composition and across-track reflectance gradient in Landsat images over Amazonian forests. *ISPRS J. Photogramm. Remote Sens.* **119**, 361–372.

Nagol, J. R., J. O. Sexton, D. H. Kim, A. Anand, D. Morton, E. Vermote, et al. 2015. Bidirectional effects in Landsat reflectance estimates: is there a problem to solve? *ISPRS J. Photogramm. Remote Sens.* **103**, 129–135.

Phillips, O. L., P. Núñez Vargas, A. L. Monteagudo, A. P. Cruz, M. E. C. Zans, W. G. Sánchez, et al. 2003. Habitat association among Amazonian tree species: a landscape-scale approach. *J. Ecol.* **91**, 757–775.

Potapov, P. V., J. Dempewolf, Y. Talero, M. C. Hansen, S. V. Stehman, C. Vargas, et al. 2014. National satellite-based humid tropical forest change assessment in Peru in support of REDD+ implementation. *Environ. Res. Lett.* **9**, 124012.

Quesada, C. A., J. Lloyd, M. Schwarz, S. Patiño, T. R. Baker, C. Czimczik, et al. 2010. Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences* **7**, 1515–1541.

Reda, I., and A. Andreas. 2004. Solar position algorithm for solar radiation applications. *Sol. Energy* **76**, 577–589.

Reda, I., and A. Andreas. 2007. Corrigendum to solar position algorithm for solar radiation applications [Solar Energy 76 (2004) 577589]. *Sol. Energy* **81**, 838.

Rocchini, D., D. S. Boyd, J. B. Féret, G. M. Foody, K. S. He, A. Lausch, et al. 2016. Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sens. Ecol. Conserv.* **2**, 25–36.

Roy, D. P., J. Ju, K. Kline, P. L. Scaramuzza, V. Kovalskyy, M. Hansen, et al. 2010. Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sens. Environ.* **114**, 35–49.

Roy, D. P., H. K. Zhang, J. Ju, J. L. Gomez-Dans, P. E. Lewis, C. B. Schaaf, et al. 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sens. Environ.* **176**, 255–271.

Ruokolainen, K., H. Tuomisto, M. J. Mac′ıa, M. A. Higgins, and M. Yli-Halla. 2007. Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? *J. Trop. Ecol.* **23**, 13–25.

Salovaara, K. J., S. Thessler, R. N. Malik, and H. Tuomisto. 2005. Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sens. Environ.* **97**, 39–51.

Sirén, A., H. Tuomisto, and H. Navarrete. 2013. Mapping environmental variation in lowland Amazonian rainforests using remote sensing and floristic data. *Int. J. Remote Sens.* **34**, 1561–1575.

Skole, D., and C. Tucker. 1993. Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988. *Science* **260**, 1905–1910.

Souza, J., J. A. V. Siqueira, M. H. Sales, A. V. Fonseca, J. G. Ribeiro, I. Numata, et al. 2013. Ten-year Landsat classification of deforestation and forest degradation in the Brazilian Amazon. *Remote Sens.* **5**, 5493–5513.

Thessler, S., K. Ruokolainen, H. Tuomisto, and E. Tomppo. 2005. Mapping gradual landscape-scale floristic changes in Amazonian primary rain forests by combining ordination and remote sensing. *Glob. Ecol. Biogeogr.* **14**, 315–325.

Toivonen, T., R. Kalliola, K. Ruokolainen, and R. N. Malik. 2006. Across-path DN gradient in Landsat TM imagery of Amazonian forests: a challenge for image interpretation and mosaicing. *Remote Sens. Environ.* **100**, 550–562.

Tuomisto, H., K. Ruokolainen, R. Kalliola, A. Linna, W. Danjoy, and Z. Rodriguez. 1995. Dissecting Amazonian biodiversity. *Science* **269**, 63–66.

Tuomisto, H., A. D. Poulsen, K. Ruokolainen, R. C. Moran, C. Quintana, J. Celi, et al. 2003a. Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. *Ecol. Appl.* **13**, 352–371.

Tuomisto, H., K. Ruokolainen, M. Aguilar, and A. Sarmiento. 2003b. Floristic patterns along a 43-km long transect in an Amazonian rain forest. *J. Ecol.* **91**, 743–756.

Tuomisto, H., K. Ruokolainen, and M. Yli-Halla. 2003c. Dispersal, environment, and floristic variation of western Amazonian forests. *Science* **299**, 241–244.

Tuomisto, H., G. Zuquim, and G. Cárdenas. 2014. Species richness and diversity along edaphic and climatic gradients in Amazonia. *Ecography* **37**, 1034–1046.

Tuomisto, H., G. M. Moulatlet, H. Balslev, T. Emilio, F. O. G. Figueiredo, D. Pedersen, et al. 2016. A compositional turnover zone of biogeographical magnitude within lowland Amazonia. *J. Biogeogr.* **43**, 2400–2411.

Van doninck, J., and H. Tuomisto. 2017a. Evaluation of directional normalization methods for Landsat TM/ETM+ over primary Amazonian lowland forests. *Int. J. Appl. Earth Obs. Geoinf.* **58**, 249–263.

Van doninck, J., and H. Tuomisto. 2017b. Influence of compositing criterion and data availability on pixel-based Landsat TM/ETM+ image compositing over Amazonian forests. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**, 857–867.

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

Wang, K., S. E. Franklin, X. Guo, and M. Cattet. 2010. Remote sensing of ecology, biodiversity and conservation: a review from the perspective of remote sensing specialists. *Sensors (Basel)* **10**, 9647–9667.

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

Wulder, M. A., J. C. White, T. R. Loveland, C. E. Woodcock, A. S. Belward, W. B. Cohen, et al. 2016. The global Landsat archive: status, consolidation, and direction. *Remote Sens. Environ.* **185**, 271–283.

Zhang, H. K., D. P. Roy, and V. Kovalskyy. 2016. Optimal solar geometry definition for global long-term Landsat time-series bidirectional reflectance normalization. *IEEE Trans. Geosci. Remote Sens.* **54**, 1410–1418.

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

Supporting Information

Additional supporting information may be found online in the supporting information tab for this article.

Figure S1. True colour composite (top) and false colour composite (bottom; red: band 4, green: band 5, blue: band 7) image details from the Web-Enabled Landsat Data composite for 2010 (left), from the image composite for 2010 derived using the methodology presented (middle), and from the new Amazon-wide image composite for July–September 2000–2009 presented (right).

