Quantitative analysis of the links between forest structure and land surface albedo on a global scale

Sara Alibakhshi[a,*], Babak Naimi[b], Aarne Hovi[a], Thomas W. Crowther[c], Miina Rautiainen[a,d]

[a] Department of Built Environment, School of Engineering, Aalto University, P.O. Box 14100, 00076 Aalto, Finland
[b] Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, 00014 Helsinki, Finland
[c] Institute of Integrative Biology, ETH Zürich, Universitätstrasse 16, P.O. Box 8092, Zürich, Switzerland
[d] Department of Electronics and Nanoengineering, School of Electrical Engineering, Aalto University, P.O. Box 15500, 00076 Aalto, Finland

ABSTRACT

Forests are critical in regulating climate by altering the Earth's surface albedo. Therefore, there is an urgent need to enhance our knowledge about the effects of forest structure on albedo. Here, we present a global assessment of the links between forest structure and albedo at a 1-km spatial resolution using generalized additive models (GAMs). We used remotely sensed data to obtain variables representing forest structure, including forest density, leaf area index, and tree cover, during the peak growing season in 2005 with pure forest pixels that cover ~7% of the Earth's surface. Furthermore, we estimated black-sky albedo at a solar zenith angle of 38° using the most recent collection of the moderate resolution imaging spectroradiometer (MODIS; version 6) at shortwave, near-infrared, and visible spectral regions. In addition, for the first time, we mapped the magnitude of the relationship between forest structure and albedo at each pixel with a 0.5-degree spatial resolution. Our results suggested that forest structure may modulate albedo in most of the sub-biomes. The response of shortwave albedo was always positive to the leaf area index and negative to the tree cover (except for deciduous broadleaf forests in mediterranean and temperate regions), while the response to forest density varied across space in 2005. The spatial map affirmed that the links between forest structure and albedo vary over geographical locations. In sum, our study emphasized the importance of forest structure in the surface albedo regulation. This paper provides the first spatially explicit evidence of the magnitude of relationships between forest structure and albedo on a global scale.

1. Introduction

Forests are critical for regulating climate by altering the amount of radiation that Earth reflects back to the atmosphere, known as surface albedo. Although albedo is considered an essential variable in climate studies (GCOS, 2003, 2006), it is still among the most uncertain components of the radiation budget in the climate modeling (Liang, 2007). The role of forests in regulating climate through albedo has been highlighted in recent studies (Bright et al., 2017; Luyssaert et al., 2018; Doughty et al., 2018). This suggests an urgent necessity for a better understanding of the factors that control the surface albedo variations in forests on a global scale.

Theoretically, land surface albedo may decrease if changes in a forest result in Earth's surface absorbing more incoming solar radiation (e.g., due to vegetation cover development) (Betts, 2000). However, the surface albedo may increase if the changes in a forest result in reflecting more incoming solar radiation (e.g., due to wider canopy gaps where

more underlying forest floor is exposed). Previous studies showed that forest structure can physically determine the albedo of forests (Hovi et al., 2019; Kuusinen et al., 2016; Lukeš et al., 2014; Manninen and Stenberg, 2009).

Despite numerous studies to investigate the links between forest structure and albedo (Abera et al., 2019; Bright et al., 2015, 2018; Culf et al., 1995; Gao et al., 2005; Roberts et al., 2004), inconsistent results in previous research suggest that the relationship between albedo and forest structure has not been understood sufficiently (Dore et al., 2012; Lukeš et al., 2013a; Sun et al., 2010). In some studies, measurements showed a negative relationship between forest structure (i.e., forest density and tree height) and albedo affected by management activities, such as logging and thinning (Sun et al., 2010), while other studies showed slightly positive relationships between forest structure (i.e., forest density, tree height, leaf area index, diameter at breast height) and albedo (Dore et al., 2012; Lukeš et al., 2013a).

Part of the challenge in producing consistent results is that forest

---

structure, along with the forest floor that usually has different optical properties than the tree canopy layer, can considerably influence albedo values (Hovi et al., 2016). Thereby, the degree to which forest structure can explain albedo may vary due to the confounding effects of understory and overstory in a forest (Alibakhshi et al., 2019). Another challenge is that the structural variables and albedo datasets used in the past studies were obtained from different satellite images with various spatial resolutions and/or used different methods that lead to difficulties in precisely understanding the mechanisms behind the links. In addition, the majority of the previous studies have been conducted over limited geographic locations, and therefore they may not represent different forest structures with varying proportions of different understory species or compositions that might be the reason behind inconsistent results of previous studies. Thus, the clear lack of representation of the links between forest structure and albedo at different biomes highlights the need for a study on a global scale. Until such contradictions can be resolved and the influence of forest structure on albedo is better understood, the research question of how forests modulate albedo remains unanswered.

This study aims to enhance our knowledge of how variations in forest structure affect albedo and to quantify the degree to which forest structure and forest types can explain albedo changes on a global scale. The global spatial coverage from satellite images can offer new insights into this study. We hypothesized that the links between forest structure (represented in this study by forest density, tree cover, and leaf area index) and albedo are dependent on geographical locations. We tested this hypothesis by generating the first global map representing the magnitude of the relationship between forest structure and albedo locally at the pixel-level.

Since solar zenith angle and snow can considerably influence albedo at different geographical locations (Ni and Woodcock, 2000), we removed pixels containing snow and estimated albedo at a fixed solar zenith angle. Our analyses were strengthened by (1) identifying pure forest type pixels including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savannah forests (WS) at a 1-km spatial resolution over major biome zones, including boreal, mediterranean, temperate, and tropical regions, (2) assessing the quality of the images using water and slope masks at a 30-m spatial resolution, and (3) using the most recent version of moderate resolution imaging spectroradiometer (MODIS) satellite images, i.e., version 6 (Wang et al., 2018).

2. Materials and methods

2.1. Overview

In this study, we used a large variety of freely available remotely sensed data (Table 1). Although no comprehensive definition exists for forest structure (Delang and Li, 2012), we used a set of variables describing different aspects of forest structure by presenting information on the distribution of forest layers horizontally, including forest density, tree cover, and leaf area index (LAI).

First, we performed a set of pre-processing steps to keep only high-quality pixels in the datasets used in this study (Section 2.2). Next, we described the procedure of identifying sub-biomes based on the ecoregions and land cover maps (Section 2.3), to later identify the timing of the peak growing season (Section 2.4). Then, we prepared a separate dataset for each sub-biome to test the hypothesis (Section 2.5). The information related to the multicollinearity test, fitting of generalized additive models (GAMs), quantifying R^2, variable importance, and response curve estimations are described in Sections 2.6 and 2.7. We explored further by quantifying the magnitude of changes in albedo in response to forest structure at each sub-biome (Section 2.8) and mapping the R^2 of the link between forest structure and albedo (Section 2.9).

All statistical analyses and visualizations were performed in R statistical software (R Core Team, 2013), QGIS (QGIS Development Team, 1991), and Google Earth Engine (Gorelick et al., 2017). We used different R packages, including "raster" (for basic raster analysis) (Hijmans and van Etten, 2012), "rts" (for time series analysis) (Naimi, 2016), "mgcv" (for fitting GAMs) (Wood, 2011), "phenopix" (for detecting peak growing season) (Filippa et al., 2016), "usdm" (for multicollinearity test) (Naimi et al., 2014), "ggplot2" (Wickham, 2016), and "rasterVis" (for visualizations) (Lamigueiro and Hijmans, 2018).

2.2. Datasets and quality

2.2.1. Major biome zone

We used the ecoregions map (Olson et al., 2001) to identify four major biome zones. Ecoregions were defined as extensive lands containing distinct ecosystems and species with boundaries that approximate their original extent. This map was calculated from biogeographical studies that synthesized information about the distributions of plants and animals, the world's biotic province maps, and different vegetation types (Olson et al., 2001). To identify the major biome zones, we selected all ecoregions representing forests, except mangrove forests (Table A1, Fig. 1). Mangrove forests are usually mixed with water, which can profoundly influence their reflectance, causing uncertainties in albedo values over MODIS pixels (see Section 2.2.7).

2.2.2. Forest type

To detect the boundary of forest types, we used the classes provided by the International Geosphere-Biosphere Program (IGBP, Channan et al., 2014), obtained from the Land Cover Type 1 of the MODIS product (MCD12Q1 version 6: Friedl and Sulla-Menashe, 2019) with a 500-km spatial resolution for the year 2005. This product identified forests based on the canopy cover percentage and tree heights (e.g., tree cover > 10%, and height > 2-m), the leaf type (broadleaf, needleleaf), and their phenology (evergreen, deciduous), using a supervised classification method. We selected the MCD12Q1 product over other land cover maps, because its spatiotemporal resolution is well-matched with other data chosen in this study (Table 1) and because it had an acceptable overall accuracy of 73.6% (Sulla-Menashe et al., 2019).

According to the Food and Agriculture Organization of the United Nations (Carle and Holmgren, 2003), land with more than 10% tree cover and an area of more than 0.5 ha is considered a forest. However, we only selected forest types with tree cover greater than 30%, including EN, EB, DN, DB, Mixed, and WS forest classes (Table A2), because of low classification accuracy (i.e., 25%) reported for forest types with tree cover less than 30% in the MCD12Q1 product (Friedl et al., 2010). Furthermore, we used the quality assessment flags of MCD12Q1 to exclude low-quality pixels. As an additional step, we also excluded the pixels with tree cover less than 30% using the vegetation continuous fields (VCF) dataset (Section 2.2.5).

2.2.3. Leaf area index (LAI)

LAI is defined as a one-sided green leaf area per unit ground area in broadleaf canopies and as one-half of the total needle surface area per unit ground area in coniferous canopies (Myneni et al., 1999). LAI describes the number of equivalent layers of leaves relative to a unit ground area (Myneni et al., 1999). In this study, we used LAI for two purposes: first, to detect the peak growing season in each sub-biome (a forest type in a biome), and second, as one of the forest structure variables that may explain albedo. We obtained LAI for the year 2005 from the MODIS product (MCD15A3H version 6, 4-day composite, 500-m x 500-m spatial resolution) (Myneni et al., 2015). In the MCD15A3H, LAI is estimated from atmospherically corrected bi-directional reflectance factors in the red and NIR spectral bands of MODIS images using a radiative transfer model designed for vegetation (Knyazikhin et al., 1998). We used the quality tags in the ancillary data of the MCD15A3H product (Myneni et al., 1999) to exclude low-quality LAI

Table 1

List of data products used in this study.

Name	Unit	Product	Spatial resolution and sensor	Year	Projection system	Data source
			Sub-biomes			
Ecoregions	–	–	–	1973–2000	Geographic	Olson et al., 2001
Forest type	–	MCD12Q1	500-m (MODIS)	2005	Sinusoidal	Friedl and Sulla-Menashe, 2019
			Forest structure components			
Forest density	number of trees/km^2	–	1-km	2005	Geographic	Crowther et al., 2015
Tree cover	percentage, %	VCF[a]	30-m (Landsat)	2003-2008	Geographic	Sexton et al., 2013
LAI	m^2/m^2	MCD15A3H	500-m (MODIS)	2005	Sinusoidal	Myneni et al., 2005
			Albedo and ancillary data[b]			
BRDF parameters	-	MCD43A1	500-m (MODIS)	2005	Sinusoidal	Schaaf and Wang, 2015
BRDF-albedo quality	–	MCD43A2	500-m (MODIS)	2005	Sinusoidal	Schaaf and Wang, 2015
Solar zenith angle	degree, °	MCD43A2	500-m (MODIS)	2005	Sinusoidal	Schaaf and Wang, 2015
Snow	–	MCD43A2	500-m (MODIS)	2005	Sinusoidal	Schaaf and Wang, 2015
Digital elevation model	m	–	90-m (SRTM[d])	–	Geographic	Jarvis et al., 2008
Water area	–	JRC[c]	30-m (Landsat)	2005	Geographic	Pekel et al., 2016

[a] Vegetation continuous fields (VCF).
[b] Google Earth engine (GEE).
[c] Joint Research Center (JRC).
[d] Shuttle radar topography mission (SRTM).
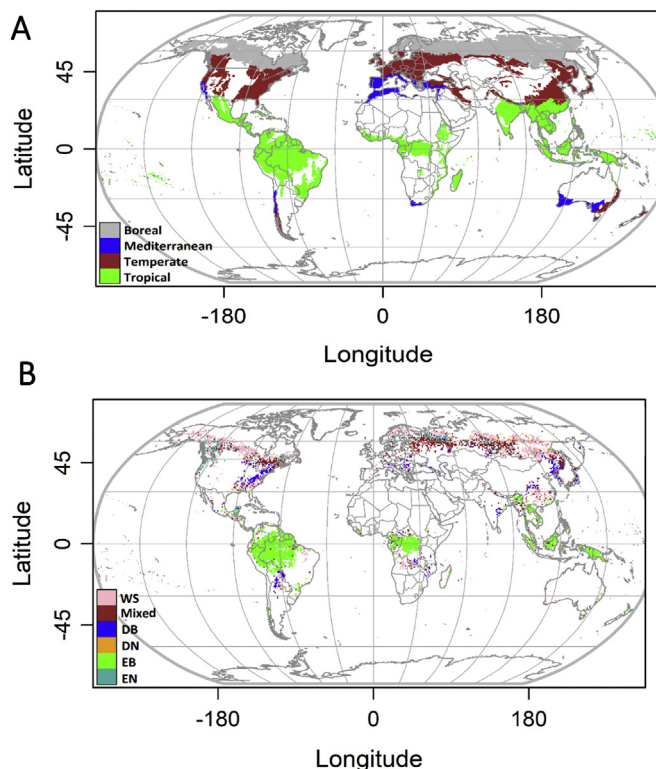
Fig. 1. Forest characterization based on forest type and major biome zone. A: major biome zones according to the ecoregions map, B: forest types, including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savanna forests (WS), according to the MODIS land cover map for the year 2005.

pixels as well as pixels with snow based on the main algorithm retrievals (both with and without saturation).

2.2.4. Forest density

To obtain the forest density, we used the only available forest density map on a global scale with a spatial resolution of 1-km, produced by Crowther et al. (2015), for the year 2005. This map was originally calculated from a predictive model that used 19 variables as predictors, including topographic (i.e., latitude, slope, elevation, aspect, and roughness), climatic (i.e., temperature, precipitation, aridity, and evapotranspiration), and vegetation variables (i.e., enhanced vegetation index and LAI, obtained from MODIS products), and over 420,000 forest inventory field plots as training data. The R^2 of the forest density was reported as 0.97 for country-wise total values (Crowther et al., 2015). From this variable, we excluded the pixels with a value of zero, which discarded 0.5% of all the pixels in the dataset.

2.2.5. Tree cover

Tree cover is generally defined as the proportional, vertically projected area of vegetation (including leaves, stems, branches, etc.) of woody plants above a given height (Sexton et al., 2013). We obtained the high-quality tree cover map from the vegetation continuous fields (VCF) dataset with a 30-m spatial resolution for the year 2003–2008 (Sexton et al., 2013). This product was derived from seven bands of Landsat-5 Thematic Mapper (TM) and/or Landsat-7 enhanced thematic mapper plus (ETM+) satellite images. VCF was estimated using a piecewise linear function of surface reflectance and temperature. The training data to estimate the tree cover were derived from the MODIS tree cover layer at 250-m.

In this study, all the forest structure variables and albedo were obtained for the year 2005, except tree cover that was calculated from data obtained between 2003 and 2008, since it was the only available product that was matched with other datasets. We assumed the VCF

product could sufficiently represent tree cover for the year 2005, because land cover changes between 2003 and 2008 were removed in this product by calculating the standard deviation of annual tree cover changes over 10% (Sexton et al., 2013). In addition, the comparison between this product and the light detection and ranging (lidar)-based measurements for the growing season of 2005–2006 demonstrated reasonable agreement with root mean square error (RMSE) of 16.8% (Sexton et al., 2013). In this study, we removed the uncertain tree cover pixels with RMSE greater than 30%, using pixel-wise reports of uncertainty in the product.

2.2.6. Slope and water masks

We used a digital elevation model (DEM) dataset, obtained from the shuttle radar topography mission (SRTM) with a 90-m spatial resolution, to calculate slope values (Jarvis et al., 2008) that were used for pre-processing of albedo datasets (Section 2.2.7). We generated the slope values using the R "raster" package (the terrain function). We then aggregated the slope values spatially from 90-m to 1-km using the arithmetic mean function.

Furthermore, we obtained the JRC yearly water classification history dataset (v1.0) with a 30-m spatial resolution in 2005 (Pekel et al., 2016), and used it for pre-processing of albedo (Section 2.2.7). This product is derived from top-of-atmosphere reflectance and brightness temperature images of Landsat. This dataset contains the locations and distribution of permanent surface water, with a classification accuracy of 98.56% (Pekel et al., 2016).

2.2.7. Albedo dataset

We estimated albedo for three different spectral regions, i.e., shortwave (SW, 300–5000 nm), near-infrared (NIR, 700–5000 nm), and visible (VIS, 300–700 nm) (Wang et al., 2018). We used these three spectral regions because they represent essential information on forest and climate. Also, a combination of NIR and VIS spectral regions can provide a wealth of information about vegetation dynamics (Treitz and Howarth, 1999).

We used MODIS products of MCD43A1 and MCD43A2 with a 500-m spatial resolution to obtain the parameters of bidirectional reflectance distribution function (BRDF) model, solar zenith angle (SZA), incremental quality values and snow information. We took advantage of numerous improvements in algorithms of the latest version of the MODIS product (i.e., v006) (Wang et al., 2018). To eliminate the effects of SZA variations between biomes, we computed black-sky albedo at fixed SZA using the following equation (Strahler et al., 1999):

$$\alpha_{bs}(\theta, \lambda)$$
$$= f_{iso}(\lambda) + f_{vol}(\lambda)(-0.007574 - 0.070987 \times \theta^2 + 0.307588 \times \theta^3) + f_{geo}$$
$$(\lambda)(-1.284909 - 0.166314 \times \theta^2 + 0.041840 \times \theta^3) \qquad (1)$$

where $\alpha_{bs}(\theta, \lambda)$ is black sky albedo, which describes the albedo under direct illumination conditions (i.e., the sun as a point source of illumination), at band λ, and θ is the SZA. In this study, θ of 38° was used, because it is a realistic value for all the sub-biomes. We selected this value by computing the minimum, maximum and mean values of SZA in each sub-biome (Table A3). Based on this analysis, we observed that the mean of SZA for all the forest pixels included in this study was 28°.

However, 28° would be an unrealistic value for the boreal region, because SZAs in the boreal region are usually larger. Therefore, we decided to use the mean SZA of the boreal region (i.e., 38°) to calculate albedo for all forest pixels (Table A3). The $f_{iso}(\lambda)$, $f_{vol}(\lambda)$, and $f_{geo}(\lambda)$ refer to the three weighting parameters in the model (isotropic, volumetric, and geometric, respectively) for λ (i.e., SW, NIR and VIS). We selected the black-sky albedo in this study because it is free from atmospheric effects and thus reveals the effects of forest structure without disturbing factors. Note that white- and blue-sky albedos were highly correlated with black sky albedo (slope: ~1 and intercept: ~ −0.0002), and therefore, results for white- and blue-sky albedos would likely be similar to black sky albedo (see Section A1.1 and Fig. A1).

To avoid the error in forests located in rugged terrain due to topography-related issues, we excluded pixels with a slope greater than 10°. The number of pixels excluded due to high slope values (greater than 10°) was equal to 3%, 6%, 9%, and 5% of the total pixels in the boreal, mediterranean, temperate, and tropical regions, respectively (Fig. A2).

Furthermore, the very low reflectance of water may affect albedo values significantly in pixels that are partly water-covered. Although the MCD12Q1 product has a separate class for water, it is possible that pixels partially covered by water have been classified as forest. Thus, we used a water mask with 30-m resolution (data in Section 2.2.6) to exclude forest pixels that had water cover greater than 5% of a pixel area this step resulted in excluding low-quality pixels equal to 15% of all the pixels in the boreal region, 5% in the mediterranean region, 6% in the temperate region, and 4% in the tropical region (Fig. A3).

2.3. Sub-biomes identification

We used the major biome zones (Section 2.2.1) and the forest types (Section 2.2.2) to identify sub-biomes. We considered different forest types within each biome as a sub-biome. For example, boreal DB forest is a sub-biome of the boreal region. To identify pure forest pixels, we aggregated the 500-m forest type pixels to 1-km spatial resolution using a modal function. Then, we set a criterion that if all the fine resolution pixels have the same forest type within a coarse resolution pixel, it is then considered as a pure pixel. This procedure resulted in removing 7,113,859 out of 42,069,856 high-quality forest pixels at a 1-km resolution (Table 2). The remaining pixels were classified either as mixed forests (tree cover more than 60%) or woody savanna (tree cover between 30%–60%), based on tree cover (land cover map definitions: Table A2 and Fig. 1).

2.4. Peak growing season

We calculated the peak growing seasons in different sub-biomes to use their corresponding times later for obtaining LAI and albedo (Fig. A4). In this study, the peak growing season was defined as the period over a year when environmental conditions are suitable for plants to reach their maximum leaf area index in a forest. We limited our analyses to the peak growing season to exclude the influence of snow and varying phenological stages of vegetation on the relationship between forest structure and albedo. Thereby, our results represent a similar phenological stage in different sub-biomes.

To estimate the timing of the peak growing season, we first

Table 2

Area [km^2] of pure forest pixels over the growing season in 2005. The columns refer to total area as well as area per each forest type, including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savannah forests (WS).

	EN	EB	DN	DB	Mixed	WS	Total
Boreal	1,579,456	0	362,540	114,201	3,965,227	6,370,820	12,392,244
Mediterranean	22,612	32,573	0	11,997	41,924	155,178	264,284
Temperate	1,045,315	389,547	24,328	1,561,611	3,046,696	2,794,644	8,862,141
Tropical	14,792	10,020,388	0	282,731	977,587	2,141,830	13,437,328
Total	2,662,175	10,442,508	386,868	1,970,540	8,031,434	11,462,472	34,955,997

extracted high-quality LAI time series (Section 2.2.3) using an arithmetic mean function for each sub-biome between the 1st of January and the 31st of December 2005. Then, we identified the time window of the stable phase around the peak growing season. We noted that the length of the window should not be long enough to include leaf area changes, for example, due to seasonality or timing of flushing that may change the albedo since the newly emerged leaves usually have higher albedos (Hovi et al., 2017), and also not so short that it would substantially reduce the number of high-quality observations.

We explored the methods available in the "phenopix" R package to detect the time window of the stable phase (Filippa et al., 2016). We found that only the Gu method produced consistent results with the least likelihood of failure compared with the other methods. The same reasoning for selecting the Gu method is reported by Snyder et al. (2016). The Gu method quantifies the photosynthetic cycle of plant communities and breaks down the dynamics of plant community photosynthesis into five distinctive phases in sequence. One of these five phases is a stable phase that refers to the length of the period between the stabilization day (i.e., the day on which the peaked canopy photosynthetic capacity is predicted to occur), and the downturn day (i.e., the day on which the peak growing season starts to decrease sharply). We detected the time window of peak growing season in LAI time series for the year 2005 by looking at the stable phase derived from the Gu method. As an example, we showed two plots of the stable phase estimated by the Gu method, one from the boreal region and another from the tropical region (Fig. A5).

We found that in most of the biomes, selecting one month as the stable phase around the peak growing season would meet our criteria, i.e., ± 15 days around the max value of LAI in one year. We assigned a month (i.e., ± 15 days) around the maximum LAI in each sub-biome, except for EB forests in the tropical region where we considered two months (i.e., ± 30 days), as it experienced a lengthier stable phase than the other sub-biomes (Fig. A5). The EB forests in the tropics rarely had good-quality pixels due to the cloud conditions. The selection of a two-month window allowed us to have a greater number of high-quality pixels (around three million more pixels), that ultimately resulted in a better representativity of the tropical region in our analysis. Next, we used the high-quality LAI product to identify when the LAI value reached its maximum (Fig. A4), and then we assigned the length of the window obtained from the Gu method.

2.5. Data extraction and quality assessment

To have a consistent spatial database, all the data were aggregated to a 1-km spatial resolution. We used a mean function for all the datasets, except for the land cover map that was aggregated with a modal function. In addition, all the data were analyzed with the geographic coordinate system, and therefore, the MODIS products were transformed from a sinusoidal to a geographic coordinate system. We performed all the analyses using geographic coordinates, since only the albedo, LAI, and land cover products had a sinusoidal coordinate system among all the eight datasets.

Based on the quality assessments (Section 2.2), we kept not only high-quality records of all the variables but also ensured that if a pixel was missing from one variable, it was excluded from all the other variables, even if it had high-quality values for the other variables. Thereby, for example, the mean LAI was taken from the pixels considered to be "high quality and not missed" for all other variables (see Table 3). This step was essential in exploring the relationships correctly (e.g., comparison between mean LAI and mean albedo at each sub-biome), as many albedo pixels were removed due to the quality issues, though they were not necessarily removed due to low quality in the other products. Finally, we calculated the mean of each variable (albedo, LAI, forest density and tree cover) extracted from the pure forest pixels in each sub-biome.

2.6. Test of multicollinearity

We performed a multicollinearity test as a final step in the data preparation procedure. Multicollinearity refers to a situation in which at least two strongly correlated variables are used as predictors in a predictive model. Multicollinearity may lead to a misleading inference, therefore, performing a multicollinearity test is a necessary step before fitting the model. Given that we used a predictive model (Section 2.7) to explore the links between forest structure and albedo, we tested whether the predictor variables (i.e., forest density, tree cover, and LAI) were subject to the multicollinearity issue. We used two standard methods to test multicollinearity, called variance inflation factor (VIF) and pairwise correlation (PC), which uses the Pearson correlation coefficient (Naimi et al., 2014). As a rule of thumb, a VIF greater than 10 or a PC greater than 0.7 could be a signal that the predictors have a collinearity problem (Dormann et al., 2013). The results of multicollinearity test can be seen in Section 3.2.

2.7. Generalized additive model (GAM)

In order to investigate the relationship between forest structure and albedo, we used generalized additive models (GAMs) (Hastie and Tibshirani, 1990) implemented in the "mgcv" library in R (R Core Team, 2013; Wood et al., 2016). This modeling method is a flexible non-parametric approach for exploring linear or non-linear relationships (Eqs. (2)–(3)), that can characterize patterns in a complex system such as a forest ecosystem. Unlike parametric linear models, the shape of the relationship in GAMs is data-driven (rather than model-driven as specified by assuming a form of parametric relationship), which allowed for the use of smoothing functions to deal with highly non-linear and non-monotone relationships (Hastie and Tibshirani, 1990). In this study, GAMs were fitted using the restricted maximum likelihood estimator (REML), which has been shown to be more robust compared with the other estimators (Wood, 2017). We fitted the following models:

$$\alpha_{bs}(\lambda) = \gamma_0 + s_1(\text{Density}) + s_2(\text{Tree cover}) + s_3(\text{LAI}) + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2) \tag{2}$$

$$\alpha_{bs}(\lambda) = \gamma_0 + s_1(\text{Density}) + s_2(\text{Tree cover}) + s_3(\text{LAI}) + t_{i1}(\text{Density, Tree cover}) + t_2(\text{Tree cover, LAI}) + t_3(\text{Density, LAI}) + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2) \tag{3}$$

where α_{bs} for λ (i.e., SW, NIR and VIS) is the black-sky albedo at 38° of SZA, γ_0 refers to the model intercept, s_1, s_2, s_3 are the smooth functions fitted using splines, t_1, t_2, t_3 are the tensor product interactions smooth functions for the interaction terms, and ε refers to the residuals of the model assumed to follow a normal distribution as $N(0, \sigma^2)$. We then used a variance partitioning method to test the proportion of albedo explained by each forest structure variable. A common problem in using flexible nonlinear methods, like GAMs, is the overfitting issue (Hawkins, 2004), which makes the model more complex than is needed. We used the penalized iteratively re-weighted least squares (P-IRLS) method (Wood, 2000) implemented in the mgcv library to avoid and control the overfitting issue.

We used the deviance explained and adjusted R^2 statistics to measure the performance (goodness-of-fit) of the GAMs. For each model, the p-value was extracted that shows whether the model fit, i.e., the links between the forest structure and albedo, was significant (Hastie and Tibshirani, 1990).

To understand whether there is any interaction between the predictor variables that explain the albedo variation, we also fitted a new model, to which the interactions between different pairs of the forest structure variables were added as new terms to the original form of the model (Eq. 3).

To fit the models, we first drew 10,000 pixels randomly from each

Table 3
Mean values of variables used in this study during the peak growing season in 2005 in pure forest pixels at a 1-km spatial resolution. The columns refer to forest types, including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savannah forests (WS). The density values are rescaled (divided by 10,000) representing the number of 10,000 trees (e.g., 6.09 = 60,900 trees).

	SW albedo						NIR albedo						VIS albedo					
	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS
Boreal	0.09	NA	0.11	0.12	0.11	0.1	0.15	NA	0.18	0.21	0.19	0.17	0.02	NA	0.02	0.03	0.02	0.03
Mediterranean	0.1	0.11	NA	0.15	0.12	0.11	0.16	0.18	NA	0.25	0.19	0.17	0.03	0.03	NA	0.02	0.03	0.03
Temperate	0.09	0.11	0.11	0.14	0.12	0.12	0.15	0.18	0.19	0.25	0.21	0.2	0.02	0.02	0.02	0.02	0.02	0.03
Tropical	0.09	0.12	NA	0.11	0.12	0.13	0.15	0.2	NA	0.18	0.2	0.21	0.02	0.02	NA	0.03	0.03	0.03

	Density [number of trees/km^2]						Tree cover [%]						Leaf area index [m^2/m^2]					
	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS
Boreal	6.09	NA	4.47	7.18	6.95	5.22	60	NA	46	48	57	41	2.7	NA	3.96	3.11	3.68	2.45
Mediterranean	7.31	8.57	NA	16.5	9.27	6.74	46	54	NA	53	47	37	2.14	4.13	NA	4.93	3.4	1.86
Temperate	6.06	5.9	5.49	5.27	5.94	4.48	60	63	55	56	57	42	2.87	4.64	3.76	4.77	4.12	3.4
Tropical	7.04	4.25	NA	5.08	4.54	3.39	60	73	NA	52	50	40	3.39	5.71	NA	2.42	4.14	3.5

sub-biome and extracted the high-quality data from the forest structure and albedo. The dataset (with 10,000 records) was then used to fit two GAMs (with and without interaction terms). We repeated this procedure 100 times, resulting in 200 GAMs (100 for each modeling form) for each sub-biome. We used these models to explore the links between forest structure and albedo and quantify the confidence interval for each parameter extracted (e.g., R^2, response curve).

In addition, we measured the importance of each forest structure variable (hereafter "variable importance") in explaining albedo in the models using a permutation test. This method measures Pearson's correlation coefficient (COR) between the predicted values of albedo (by the model) and predictions where a predictor variable is randomly permuted. If the contribution of a predictor variable to the model is high, it is expected that the predictions are more affected by a permutation, and therefore, the correlation is lower. Therefore, "1 − COR" can be considered as the measure of variable importance (Naimi et al., 2014).

Furthermore, we defined the marginal effects (Williams, 2012) as the expected changes of albedo in response to changes in a forest structure variable, while other forest structure variables are kept constant at their mean values. The results were visualized as a plot for each variable that clarifies the form and shape of the relationship. The plot is also named the model's response curve (Elith et al., 2005; Heegaard, 2002).

2.8. Quantifying the trend of changes

To quantify the magnitude of the albedo trend change (%) due to the forest structure change, we used the Theil-Sen's slope (hereafter called Sen's slope) test (Sen, 1968). Sen's slope is a non-parametric test that is known as an alternative to parametric linear regression since it is robust to outliers. This method estimates the slope between all the possible pairs of records (Eq. (4)) and then takes the median of all the slope values. This test computes the slope (i.e., the linear rate of change) as:

$$d_{ij} = median \frac{y_i - y_j}{x_i - x_j} \tag{4}$$

where y_i and y_j are data values at times x_i and x_j (or two consecutive values), respectively with $i > j$. The median of N values of d_{ij} is Sen's estimator of the slope. Then, we estimated the change magnitude of the slope (Yue and Hashino, 2003) as a percentage of mean expressed as a percentage (% change), i.e. % change of albedo due to change in a forest structure variable as:

$$\%change = \frac{d_{ij} \times (max(x) - min(x))}{mean(y)} \tag{5}$$

We calculated the % change of albedo due to change in each forest structure variable in each sub-biome, using the drawn sample data used for the GAMs.

2.9. Spatial analyses

To examine the spatial dependencies of the links between forest structure and albedo, we fitted a new set of GAMs over a globally extended coarse resolution (coarse grid cell: 50-km × 50-km) regular grid map. Instead of sub-biomes, each coarse grid cell was used as a spatial unit over which we fitted a GAM (using Eq. 2 following the same procedure used in Section 2.7) and then assigned the model's R^2 to the coarse grid cell. We repeated this procedure to also test the links between each forest structure variable and albedo separately (i.e., each model used an individual forest structure variable as the predictor). These tests allowed for spatially explicit explorations of the relationship between forest structure and albedo, in order to understand the magnitude of the relationships, and to test whether and how the relationships vary over space.

3. Results

The following sub-sections focus on the main results of the links between forest structure and albedo. We reported the results of the quality assessment of observations in Section A1.2 of the supplement and, based on these analyses, we summarized the number of available pixels at each sub-biome (Table A4). We also reported the results for the detection of peak growing season (Fig. A4, Section A1.3) and variable importance of GAMs (Section A1.4) in the supplement.

3.1. The mean and standard deviation of forest structure components and albedo

The mean and standard deviation values of albedo showed obvious variations in SW, NIR, and VIS regions in each sub-biome (Table 3, Table A5). The mean and standard deviation of VIS albedo showed notably lower values than NIR albedo, while SW albedo exhibited mid-range values between NIR and VIS albedos (Table 3, Table A5). In addition, the mean albedo varied significantly over major biome zones; otherwise, the variation of albedos was reasonably low over forest types (Table 3). The lowest values of SW albedo were evident for EN forests (mean albedo was 0.09) and the highest for DB forests (mean albedo was 0.13).

3.2. Multicollinearity

Multicollinearity test among the predictors showed that VIF values were always between 1.02 and 1.89 in all the cases, meaning that the

Table 4

Mean R^2 values derived from 100 models (GAMs), with and without interaction terms, that were fitted between albedo (response variable) and forest structure predictor variables including forest density [number of trees/ km^2], tree cover [%], and leaf area index [m^2/m^2] at the peak growing season in 2005 in each sub-biome. The columns refer to forest types, including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savannah forests (WS).

	GAMs with interactions																	
	SW albedo						NIR albedo						VIS albedo					
	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS
Boreal	0.5	NA	0.29	0.39	0.39	0.4	0.53	NA	0.28	0.42	0.4	0.46	0.24	NA	0.19	0.29	0.19	0.17
Mediterranean	0.16	0.28	NA	0.14	0.43	0.7	0.18	0.33	NA	0.18	0.54	0.77	0.47	0.22	NA	0.33	0.36	0.32
Temperate	0.4	0.33	0.26	0.07	0.46	0.63	0.42	0.37	0.27	0.08	0.47	0.67	0.21	0.39	0.1	0.16	0.16	0.27
Tropical	0.57	0.34	NA	0.58	0.34	0.33	0.59	0.35	NA	0.69	0.4	0.42	0.57	0.25	NA	0.71	0.4	0.24

	GAMs without interactions																	
	SW albedo						NIR albedo						VIS albedo					
	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS
Boreal	0.47	NA	0.27	0.36	0.37	0.37	0.5	NA	0.26	0.4	0.38	0.43	0.22	NA	0.17	0.27	0.18	0.15
Mediterranean	0.14	0.26	NA	0.07	0.42	0.69	0.15	0.31	NA	0.1	0.53	0.76	0.46	0.21	NA	0.28	0.35	0.3
Temperate	0.38	0.29	0.23	0.06	0.43	0.58	0.4	0.34	0.24	0.06	0.44	0.63	0.19	0.36	0.08	0.13	0.13	0.23
Tropical	0.49	0.3	NA	0.5	0.3	0.32	0.53	0.31	NA	0.6	0.37	0.41	0.48	0.22	NA	0.62	0.37	0.23

predictors were independent, and therefore, there was no sign of multicollinearity (Table A6). In addition, the results of the pairwise correlation (Table A7) were consistent with the VIF test and showed no sign of the issue.

3.3. Performance of GAMs in each sub-biome

The GAM models that used the interaction terms between the forest structure components (in addition to the original variables) did not significantly improve the goodness of fit. Therefore, we used the models without the interaction terms in this study (Table 4).

The relationships between the forest structure and SW, NIR, and VIS albedos during the peak growing seasons, as estimated by GAMs, were significant, with R^2 varying from 0.07 to 0.77 over different sub-biomes (Table 4). In addition, we observed that the R^2 of the SW albedo models were more similar to the R^2 of the NIR albedo models than to the R^2 of the VIS albedo models. Further, our results showed a varying range for relative importance of the forest structure variables in explaining albedo over different sub-biomes (Fig. A6).

3.4. The response curve of albedo to forest structure

The responses of SW, NIR, and VIS albedos to forest density were weak (Fig. 2). The response of SW albedo to the wide range of forest density variation (between 330 and 833,735 trees per km^2) was within a rather narrow range (between 0.07 and 0.16) with large variations in sign and magnitude of response between sub-biomes. NIR albedo variations were slightly higher than SW albedo variations (between 0.10 and 0.27) across the wide range of forest density values (Fig. A7). VIS albedo varied only marginally (between 0.01 and 0.04) in response to the forest density variations (Fig. A8). In addition, the confidence intervals of albedo response to forest density (Fig. 2) were wider than those of albedo response to tree cover (Fig. 3) and LAI (Fig. 4) (see also the response curves in Fig. A7-A12 and scatter density plots in Fig. A13-A21).

Regarding the response of albedo to tree cover, negative relationships were obtained in all the sub-biomes, except for DB forests in the mediterranean and temperate regions (Fig. 3). By looking at the response curves, it is apparent that SW, NIR, and VIS albedos responded to the tree cover variations (from 30% to 83%) within a range between 0.08 and 0.16, 0.12 and 0.28, and 0.01 and 0.05, respectively (Fig. 3, Fig. A9-A10). Furthermore, for denser canopies with tree cover of more than ~70% in particular, the albedo values showed a sudden increase

in some sub-biomes, that might be related to the GAM's adapting (overfitting) to the sparse observations in this range (Fig. A16).

We observed positive relationships between SW, NIR and VIS albedos and LAI (varying from ~0.1 to ~6.9) in all the sub-biomes (Fig. 4). The range for SW albedo was between 0.06 and 0.17, for NIR albedo between 0.08 and 0.27, and for VIS albedo between 0.01 and 0.06 in different sub-biomes (Fig. 4, Fig. A11–12).

3.5. The % change of the albedo in response to forest structure variables

While increases in NIR and SW albedos with increasing LAI were clear in all the sub-biomes, the relationships between tree cover and NIR and SW albedos in all the sub-biomes were negative, except for DB forests in the mediterranean and tropical regions (Table 5). The % change of the albedo in response to forest density varied diversely among sub-biomes (Table 5). VIS albedo decreased in response to increasing forest density, tree cover and LAI in the majority of the sub-biomes.

3.6. Spatial analyses

We mapped the spatial distribution of albedo and forest structure (Fig. A22-A27), as well as the spatial distribution of R^2 between forest structure and albedo (Fig. 5, A28-A30). The results showed no clear pattern in the R^2 values across latitudinal and longitudinal gradients (grey shades surrounded the images in Fig. 5, A28-A30). The mean R^2 substantially varied pixel-wise on a global scale. The forest structure showed a slightly weaker relationship with VIS albedo (mean R^2 = 0.40) than with SW and NIR albedos (mean R^2 = 0.43 for both SW and NIR) (Fig. 5).

4. Discussion

4.1. Relationships between forest structure and albedo

In this study, we explored the links between forest structure and albedo on a global scale. We observed varying R^2 of the relationships between forest structure and SW albedo within the same forest types (Table 4). For example, our results showed that among all the forest types, the lowest R^2 (i.e., 0.06) was observed in DB forests in the temperate region (Table 4), which is in line with the marginal response of albedo to the increasing tree cover and LAI in this sub-biome (Figs. 2-4). Observing the lowest R^2 of relationship may be related to a varying
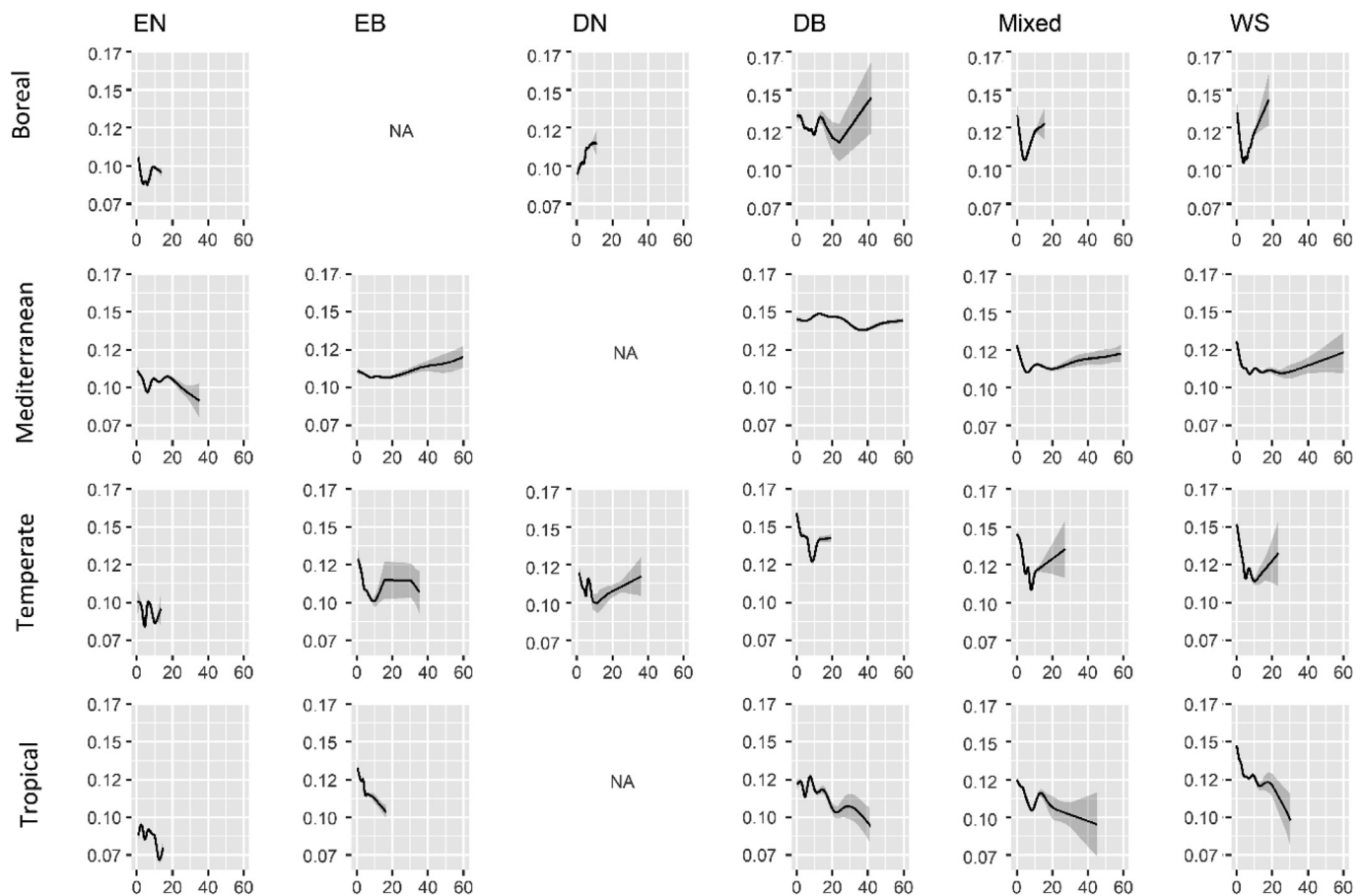
Fig. 2. The response of SW albedo to forest density [number of trees/km^2] during the peak growing season in 2005. The rows refer to the biomes, and the columns refer to forest types, including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savannah forests (WS). In each graph, the x-axis refers to the value of forest density, and y-axis refers to albedo values. The density values are rescaled (divided by 10,000, e.g., 6.09 = 60,900 trees). The grey areas around the response curves represent the confidence intervals.

range of forest structure variables among different sub-biomes (Table 3, Table A5), or the confounding effects of optical properties of forest structure and forest floor in this specific sub-biome (DB forests in temperate region), compared with the other sub-biomes.

Regarding the relationships between forest density and SW albedo, the results showed diverse responses in different sub-biomes (Fig. 2 and Table 5). This may imply that forest density alone cannot be considered as a key explanatory factor for SW albedo in different sub-biomes during the peak growing season. In general, forest density is one of the explanatory variables used in physical modeling of albedo (e.g., Wang, 2005). However, forest density together with other parameters (e.g., the biomass of foliage) explains the forest albedo variation. In addition, forest density may have no strong link to the other relevant variables, e.g., LAI, as a certain value of LAI may represent a few big trees (low density), or several small trees (high density).

Our results showed a negative relationship between the tree cover and albedo in all forests (except for DB in the mediterranean and temperate regions) (Fig. 3, Table 5). The tree cover usually has a lower reflectance compared with soil and understory (Bonan, 2008; Zhao and Jackson, 2014). This is why an increase in tree cover can lead to a decrease in albedo through the higher absorption (Planque et al., 2017).

Previous studies have reported that LAI is negatively connected with SW albedo through a year (Alibakhshi et al., 2019; Tian et al., 2018). This might be because, in some forest environments, developing tree-layer vegetation can decrease surface albedo by eliminating the effects of background such as the forest floor, which may have higher surface albedo than overstory vegetation. In this study, however, we showed positive relationships between LAI and SW albedo during the peak

growing season across the space (Fig. 4 and Table 5), which is in line with the results of other studies during peak growing seasons (Abera et al., 2019; Hollinger et al., 2010).

4.2. Spatially explicit assessment of the links between forest structure and albedo

The importance of geographic locations has been frequently emphasized in previous studies (Hovi et al., 2016, 2019; Lukeš et al., 2014), while no spatially-explicit map exists representing the links between forest structure and albedo on a global scale. In previous studies, geographic dependencies have been examined by, e.g., exploring patterns in R^2 values across latitudinal and longitudinal gradients (e.g., Lukeš et al., 2014). In this paper, we also visualized the overall changes of R^2 during the peak growing season across latitudes and longitudes (grey shades surrounding the images in Fig. 5) that revealed no significant pattern. The reason is related to having both high and low values of R^2 along a latitude (or a longitude), which may compensate each other through arithmetic mean. This argument can be confirmed by the observed R^2 of the relationships with substantially varying values over different geographic locations (Fig. 5). Therefore, the spatial map of the links between forest structure and albedo implies that the R^2 values of the relationships are very much dependent on the location, as it has also been previously shown by Hovi et al. (2019) for the boreal region.

Some contradictory results for the relationships between forest structure and albedo have been shown in previous studies (Dore et al., 2012; Lukeš et al., 2013a; Sun et al., 2010). Despite the weak latitudinal
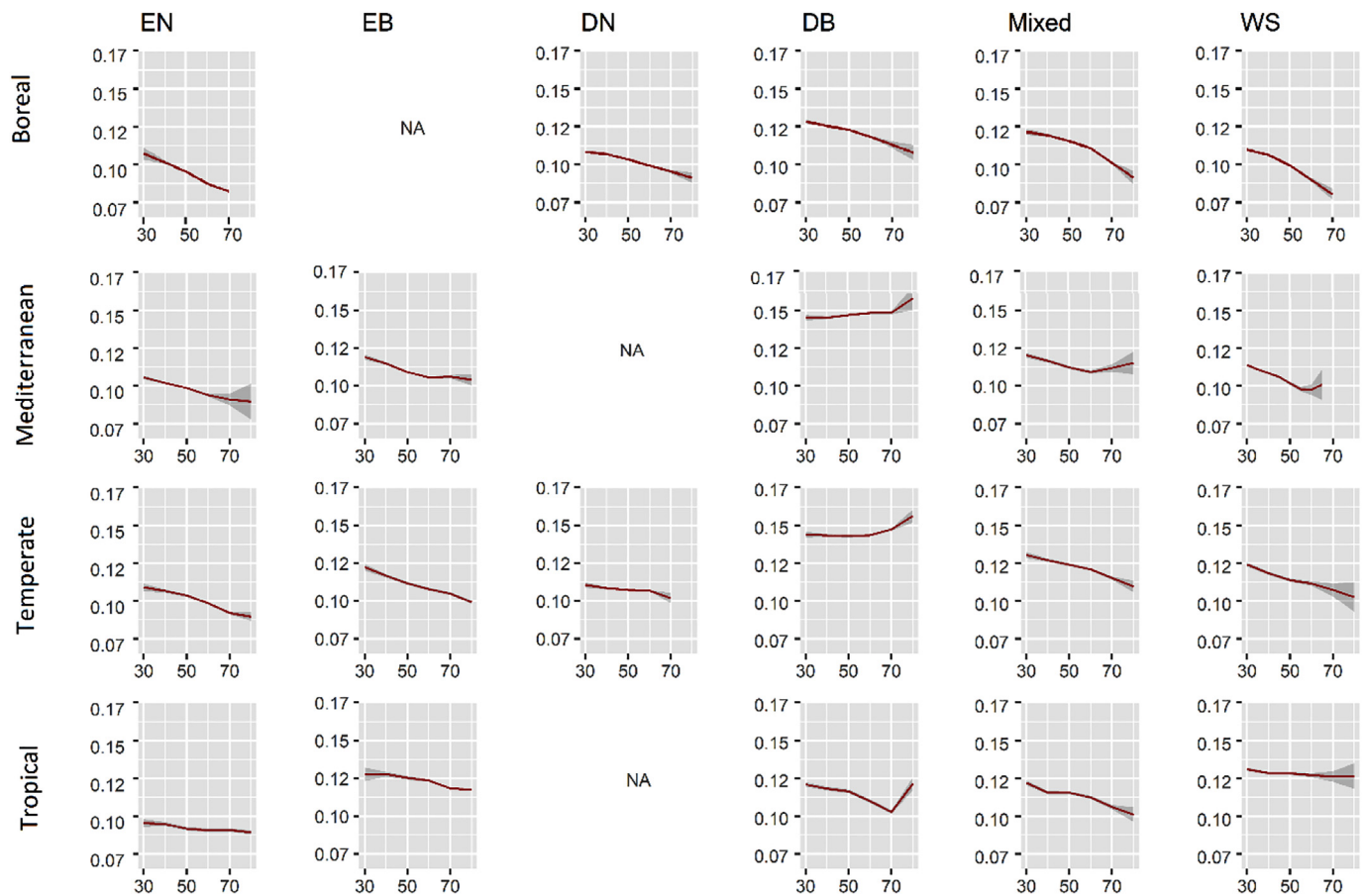
Fig. 3. The response of SW albedo to tree cover [%] during the peak growing season in 2005. The rows refer to the biomes, and the columns refer to the forest types, including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savannah forests (WS). In each graph, the x-axis refers to the value of tree cover, and the y-axis refers to albedo values. The grey areas around the response curves represent the confidence intervals.

and longitudinal gradients of SW, NIR, or VIS albedos, there is a sharp diversity in R^2 values globally (Fig. 5). Such varying dependencies can be related to diverse values in forest structure variables (Fig. A22–24) with varying proportions of forest floor, such as different species composition and visible understory in each pixel. The large geographical diversity in the response of albedo to forest structure (Fig. 5) may explain the contradictory results in previous studies.

The spatial maps showed that the R^2 values were higher in the northern parts of the boreal region (e.g., North America and central Eurasia) compared with the other regions (Fig. 5). The R^2 values of the links between albedo and individual forest structure variables suggest that all the variables contributed to explaining albedo (Fig. A28–30). As expected, R^2 of the models that used all the forest structure variables as predictors were greater than those of the models that used each individual variable separately.

We also quantified the spatial distribution of SW, NIR and VIS albedos (Fig. A24-A27). We showed that VIS albedo was considerably smaller than SW and NIR albedos, and NIR albedo always had higher values than SW and VIS albedos (Fig. 2). We observed obvious dependencies of mean SW, NIR, and VIS albedo on forest type, as was also reported by Gao et al. (2005). The mean values of SW albedo of needleleaf forests (either evergreen or deciduous) were lower than those of broadleaf forests (either evergreen or deciduous). In general, the albedo of a needle is usually smaller than that of leaves (e.g., Lukeš et al., 2013b).

4.3. Uncertainty assessment

Several factors contributed to uncertainty in this study. First, considering a single time period as the peak growing season for all areas within one sub-biome causes a potential uncertainty. Some biomes, such as mediterranean and temperate, are distributed partly in the northern hemisphere and partly in the southern hemisphere, which results in different peak growing season timing. However, these potential inconsistencies have a negligible impact on our analysis for two reasons: (1) the number of high-quality forest pixels located in mediterranean and temperate regions in the southern hemisphere was significantly lower than in the northern hemisphere (Fig. 1); (2) we compared the values of R^2 over forest types separately in the northern and southern hemispheres, reaffirming tight similarity in the values, except for EN forests in the mediterranean region (Fig. 5). The second source of uncertainty is related to the varying quality over pixels in the data products we used in this study. To offset the potential uncertainty among the products, we performed the analysis with only high-quality data through careful preprocessing actions (Section 2.2). It is noteworthy to mention that we showed a weak relationship between albedo and forest density (Fig. 2). However, it should be noted that although the overall accuracy of the forest density dataset on a global scale is quite acceptable (Crowther et al., 2015), there is no report of confidence in forest density prediction on a local scale. Third, the spatial representativeness of the data may involve some uncertainties. Although the number of high-quality pixels was quite substantial in each sub-biome, we also had to remove a considerable number of low-quality pixels from some sub-biomes, such as the tropical region. The
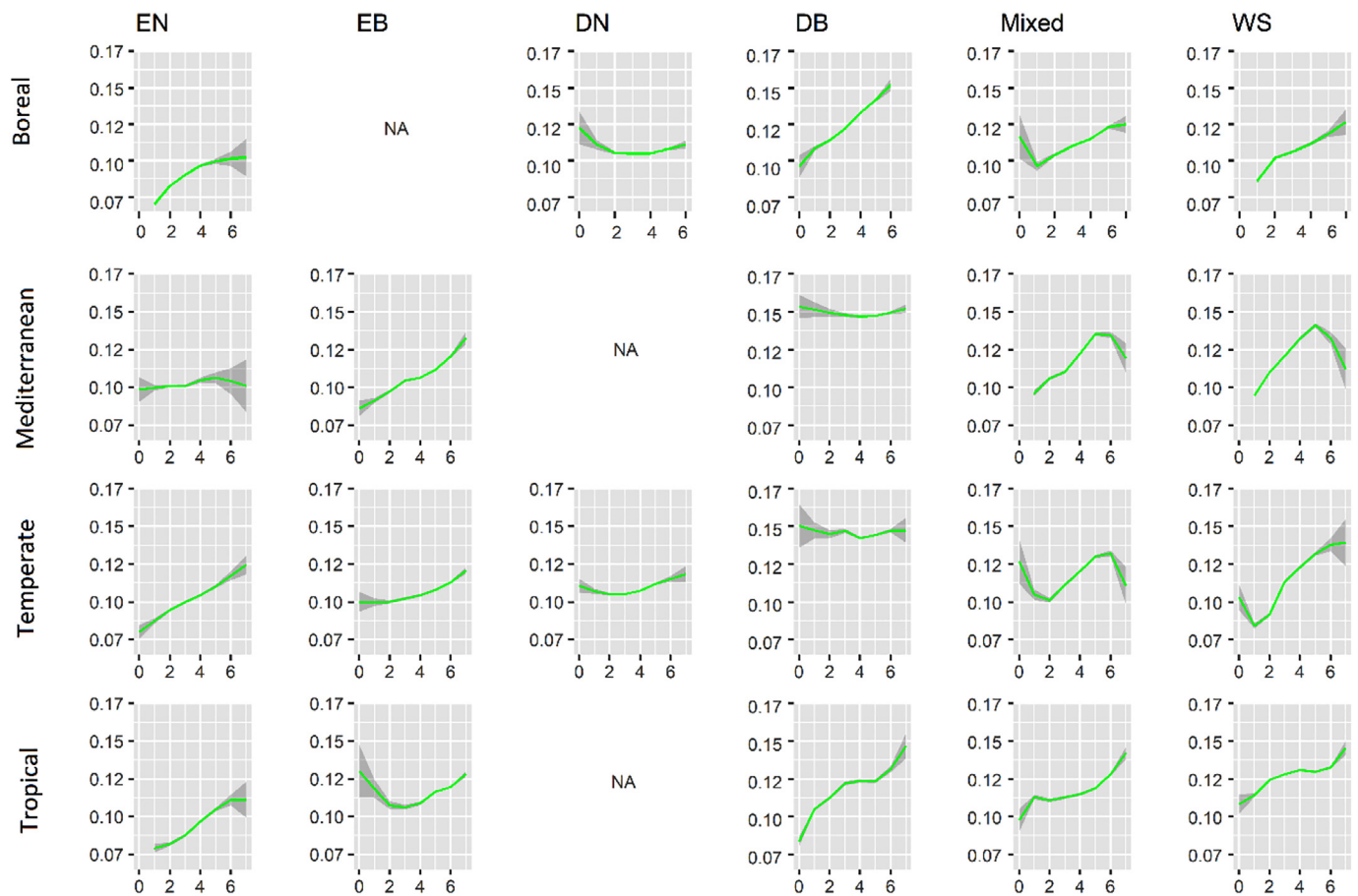
Fig. 4. The response of SW albedo to leaf area index (LAI) [m^2/m^2] during the peak growing season in 2005. The rows refer to the biomes, and columns refer to the forest types, including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savannah forests (WS). In each graph, the x-axis refers to the value of LAI, and the y-axis refers to albedo values. The grey areas around the response curves represent the confidence intervals.

Table 5
The % change of SW, NIR, and VIS albedos due to the forest structure variation. The columns refer to forest types, including evergreen needleleaf (EN), evergreen broadleaf (EB), deciduous needleleaf (DN), deciduous broadleaf (DB), mixed forests (Mixed), and woody savannah forests (WS). All the trends have been significant at the level of 90%.

								SW albedo										
			Density						Tree cover						LAI			
	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS
Boreal	17	NA	19	−18	16	9.3	−34	NA	−24	−19	−24	−6	57	NA	2	50	41	55
Mediterranean	−13	16	NA	−4	8	4.7	−20	−15	NA	4	−9	−25	12	50	NA	6.2	57	74
Temperate	8.1	−8	−9	11.3	6	−10	−26	−29	−6	5	−19	−13	41	31	10	3	25	50
Tropical	−22	−20	NA	−29	−16	−14	−12	−14	NA	−3	−19	−6	59	10	NA	39	29	27
								NIR albedo										
	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS
Boreal	24	NA	−2	−3	−13	34	−43	NA	−25	−15	−35	−43	65	NA	4	78	54	67
Mediterranean	−14	34	NA	−5	13	14	−25	−11	NA	9	0	−24	31	45	NA	−7	62	94
Temperate	−2	30	6	2	−6	−9	−25	−33	−2	7	−10	−21	52	44	14	3	30	67
Tropical	−25	−20	NA	−41	3	−32	0	−15	NA	−1	−24	−7	70	29	NA	59	57	46
								VIS albedo										
	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS	EN	EB	DN	DB	Mix	WS
Boreal	−17	NA	0	0	0	12	−64	NA	−20	−32	−33	−46	−8	NA	1	−48	−13	−26
Mediterranean	−26	2	NA	−3	−5	−9	−21	−82	NA	−57	−50	−40	−77	37.7	NA	−17	−6	−46
Temperate	0	0	−10	−3	7	11	−47	−72	−17	−32	−40	−24	1	−46	2.1	−4	−9	−21
Tropical	−65	−28	NA	−49	5	−32	−99	−44	NA	41	−4	−11	−11	−18	NA	−105	−73	−58
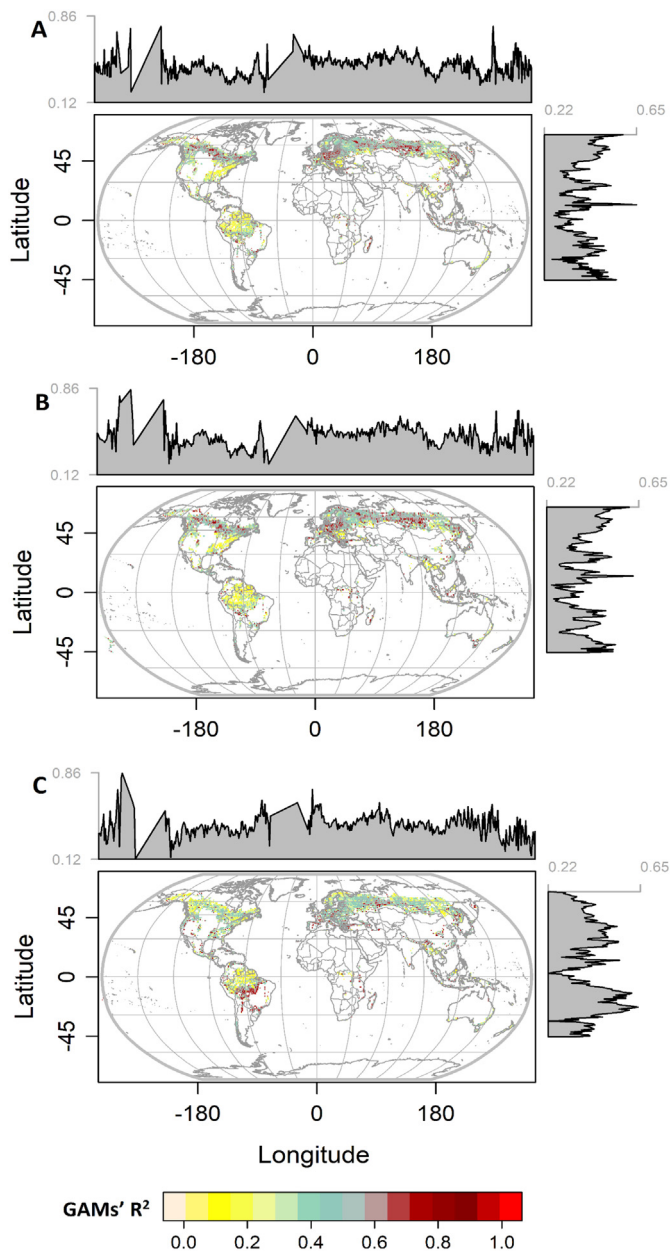
Fig. 5. Spatial distribution of the R^2 of the GAMs representing the links between forest structure and SW, NIR and VIS albedos during the peak growing season in 2005 at a 0.5-degree spatial resolution. A: SW albedo; B: NIR albedo; C: VIS albedo. The grey shades represent the mean R^2 values over longitudes (top) and latitudes (right), respectively.

unbalanced distribution of pixels may influence the results in the tropical region (Fig. 5). We tested if our results are biased towards the varying number of available pixels over space by estimating a Pearson correlation between the number of forest pixels and the obtained R^2 from GAMs over spatial locations. The low correlation coefficient (0.21) confirmed no strong bias (the results are not illustrated). Finally, another source of uncertainty could be a narrow SW, NIR, or VIS albedo range within a sub-biome (Fig. A13-A21), which allows for sudden variation due to outliers. To reduce the influence of outliers, we used high-resolution satellite images (slope and water masks) to perform additional quality assessment on top of the standard preprocessing assessments (Section 2.2.7). In addition, it should be noted that we limited our study to peak growing season, as the results of this study would probably have been considerably different throughout a year, due to seasonality of vegetation and environmental conditions (see Section

2.4).

5. Conclusion

We used spatial analysis of remotely sensed images to explore the links between forest structure and albedo, quantifying the degree to which forest structure can explain albedo in different sub-biomes on a global scale. We showed that forest structure changes could significantly affect SW, NIR, and VIS albedos during the peak growing season ($R^2 = 0.43$). However, we observed a considerably high variation in R^2 in different sub-biomes. Our results demonstrated that LAI had a positive relationship with SW albedo, while the response of SW albedo to tree cover was negative (except for DB forests in mediterranean and temperate regions). The relationships between albedo and forest density were not consistent, and varied from negative to positive over different sub-biomes. The first global map of links between forest structure and SW, NIR, and VIS albedo reaffirmed that the relationships between forest structure and albedo were extremely dependent on geographic location. All the data and maps generated in this study are freely available on our web application: https://albedo.shinyapps.io/shiny_apps/.

Declaration of Competing Interest

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.

Acknowledgments

Author contribution

Sara Alibakhshi was the main author of the paper, and responsible for the corresponding results and implementation. All authors participated in discussing the studies and commenting on the manuscripts.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2020.111854.

References

Abera, T.A., Heiskanen, J., Pellikka, P., Rautiainen, M., Maeda, E.E., 2019. Clarifying the role of radiative mechanisms in the spatio-temporal changes of land surface temperature across the horn of Africa. Remote Sens. Environ. 221, 210–224.

Alibakhshi, S., Hovi, A., Rautiainen, M., 2019. Temporal dynamics of albedo and climate in the sparse forests of Zagros. Sci. Total Environ. 663, 596–609.

Betts, R.A., 2000. Offset of the potential carbon sink from boreal forestation by decreases in surface albedo. Nature 408, 187.

Bonan, G.B., 2008. Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. Science 320 (5882), 1444–1449.

Bright, R.M., Zhao, K., Jackson, R.B., Cherubini, F., 2015. Quantifying surface albedo and other direct biogeophysical climate forcings of forestry activities. Glob. Chang. Biol. 21, 3246–3266.

Bright, R.M., Davin, E., O'Halloran, T., Pongratz, J., Zhao, K., Cescatti, A., 2017. Local temperature response to land cover and management change driven by non-radiative processes. Nat. Clim. Chang. 7, 296.

Bright, R.M., Eisner, S., Lund, M.T., Majasalmi, T., Myhre, G., Astrup, R., 2018. Inferring surface albedo prediction error linked to forest structure at high latitudes. J. Geophys. Res. Atmos. 123, 4910–4925.

Carle, J., Holmgren, P., 2003. Definitions related to planted forests. In: UNFF Intercessional Expert Meeting on the Role of Planted Forests in Sustainable Forest Management, Wellington, New Zealand, pp. 329–343.

Channan, S., Collins, K., Emanuel, W.R., 2014. Global Mosaics of the Standard MODIS Land Cover Type Data. University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA, pp. 30. [Data Set]. URL: https://lpdaac.

usgs.gov/products/mcd12q1v006/ (Date of access: 20.04.2020).

Crowther, T.W., Glick, H.B., Covey, K.R., Bettigole, C., Maynard, D.S., Thomas, S.M., Smith, J.R., Hintler, G., Duguid, M.C., Amatulli, G., 2015. Mapping tree density at a global scale. Nature 525, 201.

Culf, A.D., Fisch, G., Hodnett, M.G., 1995. The albedo of Amazonian forest and ranch land. J. Clim. 8, 1544–1554.

Delang, C.O., Li, W.M., 2012. Ecological Succession on Fallowed Shifting Cultivation Fields: A Review of the Literature. Springer Science & Business Media [eBook]. URL: https://books.google.ch/books?hl=en&lr=&id=vugb914WvxsC&oi=fnd&pg=PP3 &dq=Ecological+succession+on+fallowed+shifting+cultivation+fields: +A+review+of+the+literature.+Springer+Science+%26+Business+Media& ots=xDzloFKTnG&sig=zoUxtX4RxLHn5QZ7aXdmPAv8ovk#v=onepage&q= Ecological%20succession%20on%20fallowed%20shifting%20cultivation%20fields %3A%20A%20review%20of%20the%20literature.%20Springer%20Science%20% 26%20Business%20Media&f=false (Date of access: 20.04.2020).

Dore, S., Montes-Helu, M., Hart, S.C., Hungate, B.A., Koch, G.W., Moon, J.B., Finkral, A.J., Kolb, T.E., 2012. Recovery of ponderosa pine ecosystem carbon and water fluxes from thinning and stand-replacing fire. Glob. Chang. Biol. 18, 3171–3185.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.

Doughty, C.E., Santos-Andrade, P.E., Shenkin, A., Goldsmith, G.R., Bentley, L.P., Blonder, B., Díaz, S., Salinas, N., Enquist, B.J., Martin, R.E., 2018. Tropical forest leaves may darken in response to climate change. Nat. Ecol. Evol. 2, 1918.

Elith, J., Ferrier, S., Huettmann, F., Leathwick, J., 2005. The evaluation strip: a new and robust method for plotting predicted responses from species distribution models. Ecol. Model. 186, 280–289.

Filippa, G., Cremonese, E., Migliavacca, M., Galvagno, M., Forkel, M., Wingate, L., Tomelleri, E., Di Cella, U.M., Richardson, A.D., 2016. Phenopix: AR package for image-based vegetation phenology. Agric. For. Meteorol. 220, 141–150.

Friedl, M., Sulla-Menashe, D., 2019. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006, NASA EOSDIS Land Processes DAAC. [Data Set]. URL: https://lpdaac.usgs.gov/products/mcd12q1v006/ (Date of access: 20.04.2020).

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182.

Gao, F., Schaaf, C.B., Strahler, A.H., Roesch, A., Lucht, W., Dickinson, R., 2005. MODIS bidirectional reflectance distribution function and albedo climate Modeling grid products and the variability of albedo for major global vegetation types. J. Geophys. Res. Atmos. 110.

GCOS, 2003. Second Report on the Adequacy of the Global Observing Systems for Climate Technical Report GCOS-82 (WMO/TD No 1143). [Report]. URL: https://library. wmo.int/doc_num.php?explnum_id=3931 (Date of access: 20.04.2020).

GCOS, 2006. Systematic Observation Requirements for Satellite-Based Products for Climate Technical Report GCOS-107, WMO/TD. 1338/. [Report]. URL: https:// library.wmo.int/doc_num.php?explnum_id=3813 (Date of access: 20.04.2020).

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.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models, Volume 43 of. Monographs on Statistics and Applied Probability. 15.

Hawkins, D.M., 2004. The problem of overfitting. J. Chem. Inf. Comput. Sci. 44, 1–12.

Heegaard, E., 2002. The outer border and central border for species–environmental relationships estimated by non-parametric generalised additive models. Ecol. Model. 157, 131–139.

Hijmans, R.J., van Etten, J., 2012. Raster: Geographic Analysis and Modeling with Raster Data. (R Package Version 2.0–12).

Hollinger, D.Y., Ollinger, S.V., Richardson, A.D., Meyers, T.P., Dail, D.B., Martin, M.E., Scott, N.A., Arkebauer, T.J., Baldocchi, D.D., Clark, K.L., 2010. Albedo estimates for land surface models and support for a new paradigm based on foliage nitrogen concentration. Glob. Chang. Biol. 16, 696–710.

Hovi, A., Liang, J., Korhonen, L., Kobayashi, H., Rautiainen, M., 2016. Quantifying the missing link between forest albedo and productivity in the boreal zone. Biogeosciences 13, 6015.

Hovi, A., Raitio, P., Rautiainen, M., 2017. A spectral analysis of 25 boreal tree species. Silva Fenn 51, 51(4).

Hovi, A., Lindberg, E., Lang, M., Arumäe, T., Peuhkurinen, J., Sirparanta, S., Pyankov, S., Rautiainen, M., 2019. Seasonal dynamics of albedo across European boreal forests: analysis of MODIS albedo and structural metrics from airborne LiDAR. Remote Sens. Environ. 224, 365–381.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-Filled SRTM for the Globe Version 4. Available from CGIAR-CSI SRTM 90m. [Data Set]. URL: https:// developers.google.com/earth-engine/datasets/catalog/AU_GA_DEM_1SEC_v10_ DEM-H.

Knyazikhin, Y., Martonchik, J.V., Myneni, R.B., Diner, D.J., Running, S.W., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. J. Geophys. Res. Atmos. 103, 32257–32275.

Kuusinen, N., Stenberg, P., Korhonen, L., Rautiainen, M., Tomppo, E., 2016. Structural factors driving boreal forest albedo in Finland. Remote Sens. Environ. 175, 43–51.

Lamigueiro, O., Hijmans, R., 2018. rasterVis. (R package version 0.45).

Liang, S., 2007. Recent developments in estimating land surface biogeophysical variables from optical remote sensing. Prog. Phys. Geogr. 31, 501–516.

Lukeš, P., Stenberg, P., Rautiainen, M., 2013a. Relationship between forest density and albedo in the boreal zone. Ecol. Model. 261–262, 74–79.

Lukeš, P., Stenberg, P., Rautiainen, M., Mottus, M., Vanhatalo, K.M., 2013b. Optical properties of leaves and needles for boreal tree species in Europe. Remote Sens. Lett. 4, 667–676.

Lukeš, P., Rautiainen, M., Manninen, T., Stenberg, P., Mottus, M., 2014. Geographical gradients in boreal forest albedo and structure in Finland. Remote Sens. Environ. 152, 526–535.

Luyssaert, S., Marie, G., Valade, A., Chen, Y.-Y., Djomo, S.N., Ryder, J., Otto, J., Naudts, K., Lansø, A.S., Ghattas, J., 2018. Trade-offs in using European forests to meet climate objectives. Nature 562, 259.

Manninen, T., Stenberg, P., 2009. Simulation of the effect of snow covered forest floor on the total forest albedo. Agric. For. Meteorol. 149, 303–319.

Myneni, R.B., Knyazikhin, Y., Privette, J.L., Running, S.W., Nemani, R., Zhang, Y., Tian, Y., Wang, Y., Morissette, J.T., Glassy, J., Votava, P., 1999. MODIS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product. Modis Atbd Version 4, 130.

Myneni, R., Knyazikhin, Y., Park, T., 2005. MCD15A3H MODIS/Terra+ Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006. [Data Set]. URL: https:// lpdaac.usgs.gov/products/mcd15a3hv006/ (Date of access: 20.04.2020).

Naimi, B., 2016. rts: Raster Time Series Analysis. (R Packag. version 0–1).

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., 2014. Where is positional uncertainty a problem for species distribution modelling? Ecography. 37, 191–203.

Ni, W., Woodcock, C.E., 2000. Effect of canopy structure and the presence of snow on the albedo of boreal conifer forests. J. Geophys. Res. Atmos. 105, 11879–11888.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., 2001. Terrestrial ecoregions of the world: a new map of life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. Bioscience 51, 933–938.

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.

Planque, C., Carrer, D., Roujean, J.-L., 2017. Analysis of MODIS albedo changes over steady woody covers in France during the period of 2001–2013. Remote Sens. Environ. 191, 13–29.

R Core Team, 2013. R: A Language and Environment for Statistical Computing.

Roberts, D.A., Ustin, S.L., Ogunjemiyo, S., Greenberg, J., Dobrowski, S.Z., Chen, J., Hinckley, T.M., 2004. Spectral and structural measures of northwest forest vegetation at leaf to landscape scales. Ecosystems 7, 545–562.

Schaaf, C., Wang, Z., 2015. MCD43A1 MODIS/Terra+ Aqua BRDF/Albedo Daily L3 Global—500 m V006. LP DAAC. [Data Set]. URL: https://lpdaac.usgs.gov/products/ mcd43a1v006/ (Date of access: 20.04.2020).

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. 63, 1379–1389.

Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. Int. J. Digit. Earth 6, 427–448.

Snyder, K., Wehan, B., Filippa, G., Huntington, J., Stringham, T., Snyder, D., 2016. Extracting plant phenology metrics in a great basin watershed: methods and considerations for quantifying phenophases in a cold desert. Sensors 16, 1948.

Strahler, A.H., Muller, J.P., Lucht, W., Schaaf, C., Tsang, T., Gao, F., Li, X., Lewis, P., Barnsley, M.J., 1999. MODIS BRDF/albedo product: algorithm theoretical basis document version 5.0. MODIS Doc. 23, 42–47.

Sulla-Menashe, D., Gray, J.M., Abercrombie, S.P., Friedl, M.A., 2019. Hierarchical mapping of annual global land cover 2001 to present: the MODIS collection 6 land cover product. Remote Sens. Environ. 222, 183–194.

Sun, G., Noormets, A., Gavazzi, M.J., McNulty, S., Chen, J., Domec, J.-C., King, J.S., Amatya, D.M., Skaggs, R.W., 2010. Energy and water balance of two contrasting loblolly pine plantations on the lower coastal plain of North Carolina, USA. For. Ecol. Manage. 259, 1299–1310.

Team, Q.D., 1991. QGIS Geographic Information System. Open Source Geospatial Foundation Project. URL: http://qgis.osgeo.org (Date of access: 20.04.2020).

Tian, L., Chen, J., Shao, C., 2018. Interdependent dynamics of LAI-albedo across the roofing landscapes: Mongolian and Tibetan plateaus. Remote Sens. 10, 1159.

Treitz, P.M., Howarth, P.J., 1999. Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystems. Prog. Phys. Geogr. 23, 359–390.

Wang, S., 2005. Dynamics of surface albedo of a boreal forest and its simulation. Ecol. Model. 183, 477–494.

Wang, Z., Schaaf, C.B., Sun, Q., Shuai, Y., Román, M.O., 2018. Capturing rapid land surface dynamics with collection V006 MODIS BRDF/NBAR/Albedo (MCD43) products. Remote Sens. Environ. 207, 50–64.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. R package. Springer.

Williams, R., 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. Stata J. 12, 308–331.

Wood, S.N., 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. J. R. Stat. Soc. Ser. B. Stat. Methodol. 62, 413–428.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B. Stat. Methodol. 73, 3–36.

Wood, S.N., 2017. Generalized Additive Models: An Introduction With R. Chapman and Hall/CRC. [eBook] URL: https://books.google.ch/books?hl=en&lr=&id=HL-PDwAAQBAJ&oi=fnd&pg=PP1&dq=Generalized+additive+models:+an +introduction+with+R&ots=NpZM9jxAp3&sig= 1uOEneEvY8p1aSArZXmczqc1rc4#v=onepage&q=Generalized%20additive %20models%3A%20an%20introduction%20with%20R&f=false (Date of access: 20.04.2020).

Wood, S.N., Pya, N., Säfken, B., 2016. Smoothing parameter and model selection for general smooth models. J. Am. Stat. Assoc. 111, 1548–1563.

Yue, S., Hashino, M., 2003. Long term trends of annual and monthly precipitation in Japan 1. JAWRA J. Am. Water Resour. Assoc. 39, 587–596.

Zhao, K., Jackson, R.B., 2014. Biophysical forcings of land-use changes from potential forestry activities in North America. Ecol. Monogr. 84, 329–353.