Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2018MS001403

Key Points:
- Availability of models for simulating surface albedo changes in climate impact assessment studies of forest management is limited
- A set of flexible models based on forest and climate variables can predict albedo changes from forest management and climate change
- The models capture the seasonal pattern of surface albedo and the interactive effects of forest structure and meteorological parameters

Supporting Information:
- Supporting Information S1

Correspondence to:
X. Hu,
xiangping.hu@ntnu.no

Citation:
Hu, X., Cherubini, F., Vezhapparambu, S., & Strømman, A. H. (2018). From remotely-sensed data of Norwegian boreal forests to fast and flexible models for estimating surface albedo. *Journal of Advances in Modeling Earth Systems*, *10*, 2495–2513. https://doi.org/10.1029/2018MS001403

Received 8 JUN 2018
Accepted 1 OCT 2018
Accepted article online 3 OCT 2018
Published online 21 OCT 2018

From Remotely-Sensed Data of Norwegian Boreal Forests to Fast and Flexible Models for Estimating Surface Albedo

Xiangping Hu[1], Francesco Cherubini[1], Sajith Vezhapparambu[1], and Anders Hammer Strømman[1]

[1]Industrial Ecology Program, Department of Energy and Process Engineering, Norwegian University of Science and Technology, Trondheim, Norway

Abstract The importance to consider changes in surface albedo and go beyond simple carbon accounting when assessing climate change impacts of forestry and land use activities is increasingly recognized. However, representation of albedo changes in climate models is complex and highly parameterized, thereby limiting their applications in climate impact studies. The availability of simple yet reliable albedo models can enhance consideration of albedo changes in land use studies. We propose a set of simplified models for estimating surface albedo in a boreal forest. We process and harmonize datasets of remotely-sensed albedo estimates, forest structure parameters, and meteorological records for different forest locations in Norway. By combining linear unmixing with nonlinear programming, we simultaneously produce albedo estimates at the same resolution of the land cover dataset (16 m, notably higher than satellite retrievals) and a variety of flexible models for albedo predictions. We test different combinations of functional forms, variables, and constraints, including variants specific for snow-free conditions. We find that models capture the seasonal pattern of surface albedo and the interactive effect of forest structures and meteorological parameters, and many of them show good statistical scores. The cross-validation exercise shows that the models derived from one area perform reasonably well when applied to other forested areas in Norway, regardless of the temporal and spatial scales. By incorporating changes in forest structure and climate conditions as explicit variables, these models are simple to be used in different applications aiming at estimating albedo changes from forest management and climate change.

Plain Language Summary Surface albedo is the fraction of solar radiation reflected back into the atmosphere by a surface, and it determines how much energy is re-distributed in the biosphere. It is one of the most important physical properties of land cover and a key mechanism for climate control. By combining satellite retrievals of Norwegian boreal forest with high resolution land cover data and meteorological records, our study produces a range of relatively simple models to estimate surface albedo using the forest structure parameters and climate data. The models require simple variables of forest structure information (age and/or volume) and temperature and/or snow water equivalents. These models are relatively easy and fast to be used for quantifying effects of forest management and climate change on surface albedo.

1. Introduction

Surface albedo is the percentage of solar radiation reflected back into the atmosphere by a surface, and it determines how much energy is re-distributed in the biosphere (Bonan, 2008; Juang et al., 2007). In northern latitudes, land surface albedo is one of the most important mechanisms for bio-geophysical climate control (Bala et al., 2007; Bathiany et al., 2010; Richard A. Betts, 2000; Bonan, 2008; Claussen et al., 2001; Francesco et al., 2018). Surface albedo depends on both local climate and vegetation structure (Bright et al., 2013; Kuusinen et al., 2014; Kuusinen et al., 2014; Kuusinen et al., 2016; Lukeš et al., 2014), and it has high seasonal and spatial variation, especially in areas affected by seasonal snow cover (Anderson et al., 2011). Albedo values of open land areas, such as croplands and grasses, are highly sensitive to snow cover, whereas the masking effect of coniferous forests results in a much smaller albedo increase in the presence of snow relative to open land or deciduous forests (*R. A.* Betts et al., 2007; Cherubini et al., 2017; Zhao & Jackson, 2014). Simulating the spatial and seasonal evolution of surface albedo and the dynamic effects of forest masking are critical in climate modeling and system analysis (Hibbard et al., 2010; Hollinger et al., 2010; Prestele et al., 2017). Surface albedo in the presence of snow still has a wide spread in simulations from CMIP5 models (Dickinson, 1983; X. Qu & Hall, 2014; Thackeray et al., 2015; Wang et al., 2016). Surface albedo changes can

exert a direct radiative forcing that is of the same order of magnitude of that from the associated carbon fluxes (Bala et al., 2007; Bathiany et al., 2010; Perugini et al., 2017), and the importance to go beyond a simple carbon accounting framework when assessing the impacts of land use activities on climate is increasingly highlighted in the scientific literature (Jackson et al., 2008; Mahmood et al., 2014). There are increasing efforts to estimate climate impacts from albedo together with those from carbon in studies that assess land-derived biomass products such as biomaterials and bioenergy (Caiazzo et al., 2014; Cherubini et al., 2012; Muñoz et al., 2010). However, the representation of changes in surface albedo in climate impact analysis, like Life-Cycle Assessment (LCA), is today mostly based on satellite-derived estimates for restricted areas (Bright et al., 2015). The availability of simplified yet reliable approaches with lower degrees of parameterization and complexity of the albedo models used in climate models, can improve climate impact assessment studies of land use changes (Cherubini et al., 2016; Perugini et al., 2017).

Surface albedo of forestland is commonly calculated using three major approaches (Essery, 2013; Xin Qu & Hall, 2007). The first approach (type 1) is rather complex, and it uses two-stream approximations for canopy radiative transfer (Dickinson, 1983). An example of this type of model is the CLM land surface scheme in the Community Earth System Model (Oleson et al., 2010). The second approach (type 2) treats the snow on the ground and snow intercepted in forest canopy differently, and it is used for instance by the climate model CLASS (Bartlett et al., 2006; Verseghy et al., 1993). The third approach (type 3), utilizes a weighted average of snow-covered and snow-free albedos depending on the vegetation type (Bright et al., 2013; Kuusinen, Lukes, et al., 2014; Volodin et al., 2010). A comparison of these three common methods shows similar albedo values of snow-covered forests in Northern Hemisphere, which is typically in line with observations, suggesting that the spread in winter albedo is due to varying vegetation structures (Essery, 2013). Type 1 and type 2 approaches are rather sophisticated and require complex parameterizations, which hinders their application in climate impact studies. On the other hand, type 3 models have simpler formulations and parameterizations that can facilitate the inclusion of changes in surface albedo in climate impact assessment of land use disturbances.

Studies are trying to constrain albedo values of land cover classes integrating estimates from satellite retrievals with high-resolution land cover maps from national forestry inventories and climatological records, thereby inferring sub-pixel land cover type albedos and semi-empirical predicting albedo models (Bright et al., 2013; Kuusinen et al., 2013; Lukes et al., 2016). Using observations to constrain modeled albedos of heavily vegetated landscapes with snow cover would also reduce uncertainty in predictions of regional climate change for physically consistent regions (*X*. Qu & Hall, 2014). Using datasets for the boreal forest in Norway, the aim of this paper is to further develop this approach and produce a range of relatively simple models for the prediction of surface albedo values to facilitate albedo change parameterization in climate impact studies. The models are in the form of nonlinear equations for each of the three most common boreal forest types (deciduous, spruce and pine) in Norway with different combinations of vegetation structure (biomass volumes and standing age) and climatological indicators like temperature and snow-water equivalent (SWE). We also test the performance of the full-year model relative to a model framework where a specific model is used for snow-free conditions. The model is flexible in the choice of the preferred functional form and input variables. We show the main characteristics of the best performing models with consideration of model complexity and test their transferability to other locations.

2. Methodology

2.1. Study Area

Norway lies in Northern Europe and it comprises the western part of Scandinavia. The western and southern parts have more precipitation but experience milder winters than the northern and eastern parts, and the western parts have heavier rain and snow totals than the areas to the east of the coastal mountains. Around 38%, corresponding to about 12 million ha, of the surface area in Norway is covered by forests, and the productive forests are about 58%. There are three most important tree species in Norway, namely spruce (*Picea abies*) (47%), Scots pine (*Pinus sylvestris*) (33%), and deciduous species (mostly *Betula pubescens* and *Betula pendula*) (18%). The southern and central region is dominated by spruce, whereas the south and coastal areas are dominated by pine. At northern latitude and higher elevation, deciduous forests are more abundant.

Table 1
Information on the Mean Elevation, Temporal Scale of Albedo Satellite Retrievals, and Forest Structure Parameters (Age, Volume, and Relative Abundancy) of the Datasets used in our Analysis

	Geographic coordinate (degree)	N (500 m)/(16 m)	Elev. (m)	Year	Relative abundancy, %			Age, years (mean/median)			Volume, m^3 (mean/median)		
					P.	S.	D.	P.	S.	D.	P.	S.	D.
Data 1	Lon: 11–12; Lat: 60–60.5	380257/9373644	258	2001–2005	14	85	1	100/106	55/51	16/16	76/79	148/145	56/63
Data 2	Lon: 9–10; Lat: 61–62	25461/848064	960	2001–2002	37	46	17	54/54	84/85	94/104	63/63	100/91	27/24

Note. N = number of data points with different resolutions; Elev. = elevation; P. = Pine forest; S. = Spruce forest; D. = deciduous forest.

2.2. Datasets

We harmonize datasets of forest structure, meteorological parameters and satellite-derived albedo estimates to be used as the input for our models. Forest abundance and structure information are from the National Forest Inventory (NFI) database 'Sat-Skog', which provides an overview of the forest resources of the country with information on relative abundance of tree species, standing age, and biomass volumes at 16-m resolution (Gjertsen & Nilsen, 2012; Tomter et al., 2010). The Sat-Skog product is created by multi-band classification of remote-sensing data and by correlating observed spectral reflectance with forest parameters observed at NFI sites, so that forest parameters are estimated on a nationwide scale with high spatial resolution (Tomppo et al., 2008). Daily observations of the meteorological parameters SWE and temperature at 2-m height are produced at 1-km resolution from interpolation of meteorological weather stations in Norway (Mohr, 2008). Albedo estimate is gathered from the Moderate Resolution Imaging Spectroradiometer (MODIS) MCD43A3 (collection 5) composites with a spatial resolution of about 500 meters. The black-sky shortwave albedo estimates are generated by clear-sky atmospherically corrected surface reflectance estimates at 8-day resolution, with 16-day acquisition periods (Schaaf et al., 2002). We retain only values with band quality 0 (best quality) or 1 (good quality). The daily temperature and SWE data are temporally averaged to match the retrieval intervals of the MODIS albedo time series (i.e. averaged into a new dataset of 8-day temporal resolution using a 16-day window), then intersected with the spatial extent of MODIS pixels. For each MODIS pixel, the intersecting land-cover polygons in the Stat-Skog data are identified, and the percentage area of the MODIS pixel represented by each land-cover type is determined. The same approach is undertaken for the meteorological parameters, which are combined and preprocessed so that they are attributed to each land cover polygon. All the missing values and/or invalid values in the land cover database, meteorological parameters, and satellite derived albedos (including the retrievals that failed the quality check) are filtered out. In addition, we reduce topographic effects on the accuracy of albedo estimates by filtering out pixels with average slope of the terrain higher than two percent (identified with elevation maps from the ETOPO1 Global Relief Model from the NOAA's National Geophysical Data Center (NGDC) (Amante & Eakins, 2009). More specific information about the datasets used and their harmonization are available elsewhere (Cherubini et al., 2017).

Given the large dimensions of these datasets and the complexity of the model formulations that are to be solved, we focus our analysis on two datasets of limited space and time from two different locations. These datasets refer to a forest area either in Southeast (Data 1) or central Norway (Data 2) and are representative of different vegetation classes, climatic conditions, spatial and temporal scales (see Table 1). Data 1 corresponds to a flat area characterized by intensive forestry, rich of spruce and pine forests but without deciduous forest, whereas Data 2 is an elevated area with sparser vegetation cover and more abundant in deciduous species. We consider different time frames for the two datasets. Data 1 is made of 6 years (2001–2005) of satellite retrievals, and Data 2 of 2 years (2001–2002). From these areas, we select all the sub-grid land cover pixels that are forests and connect them to the corresponding MODIS pixel grid. We summarize the information of the two datasets in Table 1. In this table, only single forest plots with the percentage of forests of at least 75% or more are considered in each dataset (mixed forests are thus filtered out). N stands for the number of pixels used in the model on 500 m resolution or 16 m resolution. More information about these three datasets can be found in Supplementary Figure S1, which shows the distributions of forest age and volume for the tree species in the different datasets.

Table 2
Summary of the Notations used in this Paper

Symbol	Description
m	MODIS pixel
s	Forest plot in MODIS pixel
t	Time
$\alpha_{m,\,t}$	Satellite-derived albedo estimates in each MODIS pixel over time
$A_{m,\,s,\,t}$	Forest standing age at forest plot in MODIS pixel over time
$T_{m,\,s,\,t}$	Temperature at forest plot in MODIS pixel over time
$P^1_{m,s,t}$	Area of Pine forest plots in MODIS pixel over time
$P^2_{m,s,t}$	Area of Spruce forest plots in MODIS pixel over time
$P^3_{m,s,t}$	Area of deciduous forest plots in MODIS pixel over time
$F_{m,\,s,\,t}$	Share of the area (in percentage) of forest plots of land cover within the MODIS pixel over time
$S_{m,\,s,\,t}$	SWE of forest plot in MODIS pixel over time
$V^1_{m,s,t}$	Volume of Pine forest of forest plot in MODIS pixel over time
$V^2_{m,s,t}$	Volume of Spruce forest of forest plot in MODIS pixel over time
$V^3_{m,s,t}$	Volume of deciduous forest of forest plot in MODIS pixel over time

2.3. Model Framework

Satellite-derived albedo estimates are gathered at MODIS resolution of 500 meters, and land cover information is available at 16-m resolution. This means that for each MODIS pixel there are many forest plots. The radiance recorded by the satellite sensor is a linear product of the single albedo of the individual land cover types within a MODIS pixel, as the MODIS algorithm assumes a linear scaling of surface reflectance (Lewis, 1995). It is thus possible to estimate albedo values at the land cover spatial resolution through the so-called linear unmixing method (Kuusinen et al., 2013; Kuusinen, Tomppo, et al., 2014; Lukeš et al., 2014), with which albedo estimates are produced at higher resolution (16-m) than the MODIS pixels (500-m) by combining remotely sensed data with forest inventory data.

To achieve this, we construct nonlinear models to predict surface albedo and use a nonlinear mathematical programming approach to estimate the unknown parameters by minimizing the squared distance between the modeled surface albedo and the satellite-derived albedo. The general form of our model can be written as follows:

$$\widehat{\alpha} = \sum_{p=1}^{3} P^p * F * \left\{ \begin{array}{l} i_c^p + \tau_1 f_1^p(A) + \tau_2 f_2^p(T) \\ +\tau_3 f_3^p(V) + \tau_4 f_4^p(S) + \tau_5 f_5^p(V,S) \end{array} \right\} \quad (1)$$

with indicator functions

$$\tau_i = \left\{ \begin{array}{ll} 1 & \text{when } f_i \text{ is present} \\ 0 & \text{otherwise} \end{array} \right. \quad \text{for } i = 1, 2, \ldots, 5 \quad (2)$$

where p indicates different species of forests, i.e., $p = 1$ for pine, $p = 2$ for spruce and $p = 3$ for deciduous forests, P^p is the area of the forest p of the individual forest plots s within a MODIS pixel m. F is the share of the area (in percentage) of the individual plot of land cover within the MODIS pixel. This equation applies for a given pixel, forest plot and time. Both P^p and F are time-invariant. The parameter i_c^p stands for the constant term for the forest p, $f_1^p(A), f_2^p(T), f_3^p(V), f_4^p(S)$ are the functional forms for describing the relationship between surface albedo and standing age (A), temperature (T) at 2-m height above ground from the interpolation of the station observations, biomass volume (V), and SWE (S) of the forest, respectively, and $f_5^p(V,S)$ is the functional form of the interaction between biomass volume and SWE. This term is used to capture possible interaction effects between forest structure and meteorological parameters. An alternative version of equation (1) has standing age$A$ instead of biomass volume V in the interaction term. For each forest p, the functional forms in the equation (1) are the same but with different unknown parameters, which need to be estimated from the data. $\widehat{\alpha}$ denotes the estimated surface albedo of each forest plot (16-m resolution) within each MODIS pixel at different time for different driven parameters, i.e., meteorological parameters and forest structure parameters. See Table 2 for a summary of the notations used in this paper.

In addition to the general form of our model, we investigated many other model frameworks and functional forms. For instance, another version of equation (1) uses temperature (T) instead of the snow-water equivalent (S) for the functional form. However, the version of the model in equation (1), which uses snow-water equivalent (S) as the meteorological parameter in the interaction term, is the option that shows better results, and it is therefore the only one further discussed in this paper.

2.4. Functional Forms

In the model given by equations (1) and its alternative version with age in the interaction term, a variety of different functional forms is used to capture the relationship between surface albedo and forest structure or metrological parameters.

2.4.1. Age

Regarding modeling the relationship between forest age and surface albedo, we consider the following three options:

$$\begin{aligned} f_{1,1}(A) &= i * \exp(-i_a * A) \\ f_{1,2}(A) &= i * \left(1 - \frac{1}{(1 + \exp(-i_a * (A - i_{ac})))}\right) \\ f_{1,3}(A) &= i * \left(1 - \frac{1}{(1 + \exp(-i_a * (A - i_{am})))}\right) \end{aligned} \tag{3}$$

where $f_{1,\,1}(A)$ is an exponential decay function with scaling parameter i and shape parameter i_a. The shape parameter i_a controls how fast the surface albedo decays with growing forests and scaling parameter i controls the overall scaling of the function. The exponential decay function $f_{1,\,1}(A)$is also used elsewhere to fit the relationship between surface albedo and forest age (Kuusinen, Tomppo, et al., 2014). In addition to it, we also test two more complicated functional forms that can more realistically represent the different growth rates of the forest at different life stages. The functional forms $f_{1,\,2}(A)$ and $f_{1,\,3}(A)$ are logistic functions with scaling parameter i, shape parameter i_a and sigmoid midpoint i_{ac} and i_{am}, respectively. In $f_{1,\,2}(A)$ the value of i_{ac} is fixed at predefined values, namely i_{ac} = 15 or i_{ac} = 30. On the other hand, in $f_{1,\,3}(A)$ the i_{am} is treated as an unknown parameter which needs to be estimated from the data. Testing two similar functional forms $f_{1,\,2}(A)$ and in addition to the exponential decay function will allow us to get information about the balancing of complexity with accuracy (in terms of model performance), because more complex functions and unknown parameters to be estimated increase model computational time and make the optimization harder. The same consideration holds true for the choices of functional forms for other parameters discussed below.

2.4.2. Temperature

The relationship between surface albedo and temperature can be modeled with logistic functions to capture the rapid change in albedo values around snow melting temperatures. This functional form is based on the data plots in Supplementary Figure S2(a), where MODIS surface albedo values are against mean temperature. We have chosen two functional forms for $f_2(T)$:

$$\begin{aligned} f_{2,1}(T) &= k * \left(1 - \frac{1}{(1 + \exp(-k_t * (T - i_{tc})))}\right) \\ f_{2,2}(T) &= k * \left(1 - \frac{1}{(1 + \exp(-k_t * (T - i_{tm})))}\right) \end{aligned} \tag{4}$$

with sigmoid midpoints i_{tc} and i_{tm}, respectively. The parameter k scales the function, and the parameter k_t controls the shape of the functions. The sigmoid midpoint i_{tc} and i_{tm} control the inflection points of the functions $f_{2,\,1}$ and $f_{2,\,2}$, respectively, and they are used to capture the dramatic change of surface albedo when temperature T becomes higher than 0 °C. As in the case for age, the difference between the two functional forms concerns the midpoint, which is assumed to be constant (i_{tc} = 270K) in $f_{2.1}$, whereas it is estimated from the data in $f_{2.2}$.

2.4.3. Volume

One functional form with an exponential decay function is chosen for describing the relationship between surface albedo and forest biomass volume:

$$f_3(V) = j * \exp(-j_v * V) \tag{5}$$

with scaling parameter j and shape parameter j_v. The parameter j_v dictates the speed of surface albedo decays with increasing volume, and parameter j scales the function. This functional form captures the decay of surface albedo with growing standing volume, which gradually stabilizes when the forest approaches maturity. We only use one functional form for the volume-albedo relationship because this trend is observed in several papers studying the link between (semi-)empirical albedo estimates and forest volume growth (Kuusinen et al., 2016; Lukes et al., 2013; Lukes et al., 2016; Lukeš et al., 2014).

2.4.4. Snow Water Equivalents (SWE)

Surface albedo is very sensitive to SWE. We simulate this relationship with three different functional forms, an exponential function and two logistic functions. The exponential function (similar to a second-half of a sigmoid) is chosen on the basis of the trend in the scatter plot in Figure S2(b). As for the case of age-albedo relationship, the other two sigmoidal functions are tested to investigate the tradeoffs between model complexity and accuracy:

$$\begin{aligned} f_{4,1}(S) &= q*(1 - \exp(-q_s*S)) \\ f_{4,2}(S) &= q*\left(\frac{1}{(1 + \exp(-q_s*(S - i_{sc})))}\right) \\ f_{4,3}(S) &= q*\left(\frac{1}{(1 + \exp(-q_s*(S - i_{sm})))}\right) \end{aligned} \tag{6}$$

with sigmoid midpoints i_{sc} and i_{sm} in $f_{4,\,2}$ and $f_{4,\,3}$, respectively. In all these three forms, the parameter q scales the whole function and the parameter q_s controls the shape of the function. The sigmoid midpoints i_{sc} and i_{sm} are supposed to control the dramatic change of albedo values with and without snow cover. The parameter i_{sc} is fixed at 15 or 30 to simplify the optimization, whereas i_{sm} is estimated from the data.

2.4.5. Vegetation-Climatology Interaction Term

We also introduce an interaction term f(V, S) in equation (1) to capture the joint effect from the meteorological parameter S and the forest structure parameterV,

$$f_5(V, S) = g_1(V)g_2(S) \tag{7}$$

g_1(V) concerns the volume and has the similar functional form as in equation (5), but with a different scaling parameter, and g_2(S) concerns SWE and is similar to the form for SWE discussed in equation (6). Equation (7) has a two individual forms that make easier to separately control and interpret the individual functions. We thus define three possible alternative forms for equation (7):

$$\begin{aligned} f_{5,1}(V, S) &= r*\exp(-j_v*V)*(1 - \omega*\exp(-r_s*S)) \\ f_{5,2}(V, S) &= r*\exp(-j_v*V)*\left(\frac{1}{(1 + \exp(-r_s*(S - i_{vsc})))}\right) \\ f_{5,3}(V, S) &= r*\exp(-j_v*V)*\left(\frac{1}{(1 + \exp(-r_s*(S - i_{vsm})))}\right) \end{aligned} \tag{8}$$

In $f_{5,\,1}$(V, S) we introduce a parameter ω, which is defined in the interval $0 \le \omega \le 1$, to scale the exponential function $g_2(S)$. In these forms, r is a scaling parameter, and r_s is a shape parameter that controls the shape of the function $g_2(S)$. The parameters i_{vsc}, i_{vsm} are the sigmoid midpoints where $i_{vsc}$is fixed at 15 or 30 and i_{vsm} is unknown, i.e. it is estimated from the data.

In order to explore the different model performances on the type of forest variable considered, we elaborate an alternative version of the interaction term by replacing biomass volume with forest age A:

$$f_5(A, S) = g_1(A)g_2(S) \tag{9}$$

Explicitly, it has the following three alternative forms

$$\begin{aligned} f_{5,1}(A, S) &= r*\exp(-j_v*A)*(1 - \omega*\exp(-r_s*S)) \\ f_{5,2}(A, S) &= r*\exp(-j_v*A)*\left(\frac{1}{(1 + \exp(-r_s*(S - i_{asc})))}\right) \\ f_{5,3}(A, S) &= r*\exp(-j_v*A)*\left(\frac{1}{(1 + \exp(-r_s*(S - i_{asm})))}\right) \end{aligned} \tag{10}$$

The interpretation of the parameters in equations (10) is similar to those in equations (8), with volume replaced by age.

2.5. Model Variant for Snow-Free Conditions

In addition to the general model described above, we also consider a variant that has a specific model to predict albedo values under snow-free conditions. To do this investigation, we first divide both the datasets Data 1 and Data 2 into two parts, with snow-free (S = 0) and snow-present (S > 0) conditions. The relationship between surface albedo and forest structure and meteorological parameters are shown in Supplementary Figure S3 for snow-free conditions. We see that in snow free conditions there is not a significant trend between albedo and temperature, and we use a linear function to model the relationship. A decaying exponential function is used to model the relationship between volume and albedo. This trend under snow-free conditions was observed in previous study as well (Rechid et al., 2009).

The following equation is adopted for snow-free conditions

$$\widehat{\alpha} = \sum_{p=1}^{3} P^p * F * \left\{ i_c^p + \tau_1 f_1^p(A) + \tau_2 f_2^p(T) + \tau_3 f_3^p(V) \right\}, \quad S = 0 \tag{11}$$

with

$$f_2(T) = k * T \tag{12}$$

which is a linear function with temperature T. We choose the same functional forms for $f_1(A)$ and $f_3(V)$ as described in the previous section. The additional term i_c^p is a constant terms which is different for different forest species.

We use the same model framework presented above for the snow-present conditions, and then combine the two independent estimates under snow and snow-free conditions to create a dataset that can be directly compared with the output of the full-year model.

2.6. Constraints in the Model

Some constraints are needed and introduced to confine the solution domain and make the optimization feasible. There are two groups of constraints. The first group includes constraints on the parameters, and they are used to constrain the shape and scaling parameter as well as the constant term i_c^p. Two parameters are introduced, and the constraints can be written as

$$\begin{aligned} \theta_l^{1,3} &= c_1 * \theta^2 \\ \theta_l^{1,3} &= c_2 * \theta^2 \\ \theta_u^{1,3} &= {}^{\theta^2}\Big/_{c_1} \\ \theta_u^{1,3} &= {}^{\theta^2}\Big/_{c_2} \end{aligned} \tag{13}$$

where, θ_l^p and θ_u^p indicate lower and upper bounds of parameters $\left\{ i_c^p, i, i_a^p, k, k_t^p, j, j_v^p, q, q_s^p, r, r_s^p \right\}$, with $p = 1$ for pine, $p = 2$ for spruce and $p = 3$ for deciduous forests. With this formulation, bigger values of c, which is defined in the interval $0 < c_i \leq 1$, $i = 1, 2$ introduce a harder constraint on the parameters. For instance, when $c_i = 1$, $i = 1, 2$, we have an idealized case where we do not distinguish among the different forest types but treat them identically. Since the coniferous forests, i.e., pine and spruce, are more similar than deciduous forest, we set a bigger value for c_1 and a smaller value for c_2 ($c_1 = 0.7$ and $c_2 = 0.25$).

In the case of open land without standing forest, forest age and volume is set to 0, and the predicted surface albedo becomes similar independently of the types of forest. This introduces the second group of constraints with the following form

$$\begin{aligned} \widehat{\alpha}_l^{1,3} &= c_3 * \widehat{\alpha}^2, \\ \widehat{\alpha}_u^{1,3} &= c_4 * \widehat{\alpha}^2 \end{aligned} \tag{14}$$

when volume and/or age are/is equal to 0. The parameters $\alpha_l^{1,3}$ and $\alpha_u^{1,3}$ stand for the lower and upper bounds of the predicted surface albedos for pine and deciduous forests, and $\widehat{\alpha}^2$ is the predicted surface albedo for the spruce forest. With $c_3 = c_4 = 1$, the predicted albedo for the three different forests with no vegetation (open land) is constrained to be the same.

2.7. Parameter Estimation

We use the nonlinear mathematical programming approach to estimate the unknown parameters in the models with the functional forms discussed above. The models are implemented in GAMS using the standard nonlinear optimization solver CONOPT (GAMS, 2017). The satellite–derived surface albedo α is available at each MODIS pixel m over time t. We compute the weighted average of the estimated albedo $\widehat{\alpha}_{m,s,t}$ of the individual forest plot s within each corresponding MODIS pixel m, and then compare the predicted albedo values with the satellite-derived estimates. Our question is then formulated as a nonlinear optimization problem by minimizing the squared distance between the predicted albedo and the satellite-derived albedo.

The mathematical optimization model can be summarized as:

$$\min \sum_{m,\ t}\left(\sum_{s}\widehat{\alpha}_{m,\ s,\ t}-\alpha_{m,\ t}\right)^2 \quad \forall m,\ t \tag{15}$$

subject to

$$\begin{aligned}\widehat{\alpha}_{m,\ s,\ t} &= \mathrm{P}^1_{m,\ s,\ t}{*}\mathrm{F}_{m,\ s,\ t}{}^*\left\{i_c^1+\tau_i\sum_j f_j^1\right\}\\ &+\mathrm{P}^2_{m,\ s,\ t}{*}\mathrm{F}_{m,\ s,\ t}{}^*\left\{i_c^2+\tau_i\sum_j f_j^2\right\} \quad j\in\{1,\ 2,\ \ldots,\ 5|\mathrm{Model}\}\\ &+\mathrm{P}^3_{m,\ s,\ t}{*}\mathrm{F}_{m,\ s,\ t}{}^*\left\{i_c^3+\tau_i\sum_j f_j^3\right\}\end{aligned} \tag{16}$$

and constraints discussed in Section 2.6. In equation (16), we use the notation $\sum_j f_j^p, j\in\{1,\ 2,\ \ldots,\ 5|\mathrm{Model}\}$ instead of the general form in equation (1) for each forest p, $p = 1,\ 2,\ 3$ since many of the models do not need to have all the items. We group the items in $\tau_j\sum_j f_j,\ j\in\{1,\ 2,\ \ldots,\ 5|\mathrm{Model}\}$ as a constant term, forest structure terms, meteorology terms, and the interaction term between the forest structure and meteorological parameters,

$$\begin{aligned}\text{Constant term} &= i_c^p,\\ \text{forest structure terms} &= \tau_1 f_1^p(\mathrm{A})+\tau_3 f_3^p(\mathrm{V}),\\ \text{meteorological terms} &= \tau_2 f_2^p(\mathrm{T})+\tau_4 f_4^p(\mathrm{S}),\\ \text{interaction term} &= \tau_5 f_5^p(\mathrm{V},\ \mathrm{S}) \text{ or } \tau_5\, f_5^p(\mathrm{A},\ \mathrm{S}), \quad p=1,\ 2,\ 3.\end{aligned} \tag{17}$$

With this grouping, we can investigate the contributions from forest structure parameters, meteorological parameters and the interaction between the forest structure and the meteorological terms. To make our model clear, one explicit example of equation (1) is given as follows

$$\begin{aligned}\widehat{\alpha} &= \sum_{p=1}^{3}\mathrm{P}^p{}^*\mathrm{F}^*\left\{\begin{array}{l} i_c^p+\tau_1 f_1^p(\mathrm{A})+\tau_2 f_2^p(\mathrm{T})\\ +\tau_3 f_3^p(\mathrm{V})+\tau_4 f_4^p(\mathrm{S})+\tau_5 f_5^p(\mathrm{V},\mathrm{S})\end{array}\right\}\\ &= \sum_{p=1}^{3}\mathrm{P}^p{}^*\mathrm{F}^*\left\{\begin{array}{l} i_c^p+\tau_1{}^* i{*}\exp(-i_a{*}A)\\ +\tau_2{}^* k^*\left(1-\dfrac{1}{(1+\exp(-k_t{}^*(\mathrm{T}-i_{tm})))}\right)\\ +\tau_3{}^* j^*\exp(-j_v{}^*\mathrm{V})\\ +\tau_4{}^* q^*\left(\dfrac{1}{(1+\exp(-q_s{}^*(\mathrm{S}-i_{sm})))}\right)\\ +\tau_5{}^* r^*\exp(-j_v{}^*\mathrm{V}){*}(1-\omega^*\exp(-r_s{}^*\mathrm{S}))\end{array}\right\}\end{aligned} \tag{18}$$

3. Results and Discussion

More than 300 models are investigated in our study, and we only present the most performing options. First, we provide an overview of the main results for some different models. Thereafter, we select two of the best

Table 3
Summary of Model Results for Data 1, Located in a Forestry Area in SE Norway

	$f_1(A)$			$f_2(T)$			$f_4(S)$			$f_5(V,S)$			
Nr.	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{2,1}$	$f_{2,2}$	$f_3(V)$	$f_{4,1}$	$f_{4,2}$	$f_{4,3}$	$f_{5,1}$	$f_{5,2}$	$f_{5,3}$	R^2
1	×			×		×	×						0.678
2				×						×			0.696
3					×					×			0.703
4					×		×			×			0.708
5					×			×		×			0.712
6					×				×			×	0.712
7					×				×	×			0.713
8			×		×			×			×		0.714
9			×		×	×		×			×		0.714
10	×				×				×	×			0.715
11	×				×	×			×	×			0.715
12			×		×			×		×			0.715
13			×		×	×		×		×			0.715
14		×			×				×	×			0.715
15			×		×				×	×			0.715
16			×		×				×	×			0.715
17	×				×	×			×			×	0.716
18		×			×	×			×			×	0.716
19			×		×	×			×			×	0.716
20			××		×				×			×	0.716

Note. This dataset is rich in coniferous forests, but few deciduous forests. The symbol × indicates that the model includes the corresponding functional form; otherwise, the corresponding item is missing. The constant term i_c is always included in all models, and hence not shown in the table.

models based on model complexity and performance, and then we depict their characteristics in detail. After this, we perform model inter-comparison to compare the behaviors of the selected models and test their spatial transferability. Finally, we conduct a sensitivity analysis on the constraints used in our study.

3.1. Summary of models' Performance

We summarize the results of the best performing models with volume as interaction term based on the regression coefficient R^2 (see Table 3 and Table 4). Table 3 shows the results for Data 1, which is rich in pine

Table 4
Summary of Model Results for Data 2, Located in a Aountainous Area in Central Norway

	$f_1(A)$			$f_2(T)$			$f_4(S)$			$f_5(V,S)$			
Nr.	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{2,1}$	$f_{2,2}$	$f_3(V)$	$f_{4,1}$	$f_{4,2}$	$f_{4,3}$	$f_{5,1}$	$f_{5,2}$	$f_{5,3}$	R^2
1										×			0.704
2	×			×						×			0.752
3				×						×			0.752
4									×	×			0.763
5					×					×			0.765
6					×		×			×			0.768
7				×						×			0.775
8				×				×		×			0.778
9				×								×	0.778
10					×							×	0.790
11				×					×	×			0.791
12					×				×	×			0.792
13	×				×				×	×			0.807
14	×				×	×			×	×			0.807

Note. This dataset contains both coniferous forest and deciduous forest. The symbol × indicates that the model includes the corresponding functional form; otherwise, the corresponding item is missing. The constant term i_c is always included in all models, and hence not shown in the table.

and spruce, whereas Table 4 shows the results for Data 2, which has a larger fraction of deciduous species. Models are numbered and for each of them we indicate the functional form used for age, temperature, volume, SWE, and for the interaction term. The explicit form given in equation (18) is the model Nr. 11 in Table 3 and Nr. 14 in Table 4.

Models applied in Data 2 achieve higher values of R^2 because the dataset is smaller in terms of temporal dimension (2 years against 6 years) and has lower entries because the forest has a lower density than Data 1 (and therefore there is a smaller number of forested MODIS pixels). In general, the regression coefficient R^2 increases with more complex models, i.e., with a higher number of more complicated functional forms. The results show that the inclusion of the interaction term between forest structure and meteorology is key for improving model performance. However, using a complex functional form such as $f_{5,\,2}(V,S)$ or $f_{5,\,3}(V,S)$ for SWE in the $f_5^p(\mathrm{V},\mathrm{S})$ does not significantly improve the results, and the simpler form $f_{5,\,1}(V,S)$ is sufficient to produce robust predictions at reduced complexity. In addition, when the interaction term $f_5^p(\mathrm{V},\mathrm{S})$ is included, extra additive terms of forest structure information like $f_1(\mathrm{A})$ and $f_3(V)$ are not usually necessary, as they do not improve the results. The inclusion of meteorological parameters, i.e., temperature and SWE, as additive terms in the models improve the overall performances. In general, $f_{2,\,1}(T)$and $f_{4,\,2}(S)$ produce lower scores than $f_{2,\,2}(T)$ and $f_{4,\,3}(T)$, meaning that estimating the inflection points of the functional forms for temperature and SWE with optimization provides better estimates than assuming fixed values.

There are many more models in our study, such as multiplicative models, but they do not produce better results. Furthermore, multiplicative models usually lead to very complex functional forms and increase nonlinearity, which makes estimating the unknown parameters more time-consuming and computationally demanding.

3.2. Model Applicability

Many models produce results with relatively high R^2 and can be chosen for practical applications. On the basis of the relationship between model complexity and accuracy, we select two models to test the applicability and explore model inter-comparison. The first model is number 7 in Table 3, or number 12 in Table 4. This model is based on volume as the variable for forest structure in the interaction term and has the following form:

$$\widehat{\alpha} = \sum_{p=1}^{3} P^p * F * \left\{ i_c^p + f_{2,2}^p(T) + f_{4,3}^p(S) + f_{5,1}^p(V,S) \right\} \tag{19}$$

with $f_{\cdot,\cdot}^p$ defined in Section 2.4. This model has $R^2 = 0.713$ in Data 1 and $R^2 = 0.792$ in Data 2. The estimates of the parameters for this type of model are available in the Supplementary Table S1, from which the simplicity of its application can be appreciated.

The second model is based on forest age, instead of volume, as the variable for forest structure in the interaction term and has the following form:

$$\widehat{\alpha} = \sum_{p=1}^{3} P^p \cdot F \times \left\{ i_c^p + f_{2,2}^p(T) + f_{4,3}^p(S) + f_{5,1}^p(A,S) \right\} \tag{20}$$

This model has $R^2 = 0.715$ over the first area and $R^2 = 0.791$ over the second area. See Supplementary Table S2 for the estimates of the parameters. These models are selected to reflect simplicity and model performance because only one forest structure parameter (either volume or age) is included in the interaction term but not as individual additive items in the model (which only has climatological parameters).

Predicted albedo values using the two model types are shown in Figure 1 for selected volume (Figure 1a and Figure 1b) and age (Figure 1c and Figure 1d) increments over months, assuming fixed monthly means of temperature and SWE. The Figure illustrates results for different forest species, pine, spruce and deciduous, and areas, Data 1 and Data 2. The fraction of deciduous species in Data 1 is very low (see Table 1 and Figure S1), and therefore we do not present the results for deciduous forests for this dataset.

As illustrated in Figure 1, the models capture the seasonal variation, with high surface albedo values in winter months and low values during summer months. Albedo decreases with growing forests and stabilizes when the forest gets mature, as the lines in the figures tend to overlap at high volumes and ages. Surface albedo differences between coniferous species pine and spruce are small, whereas the differences are more

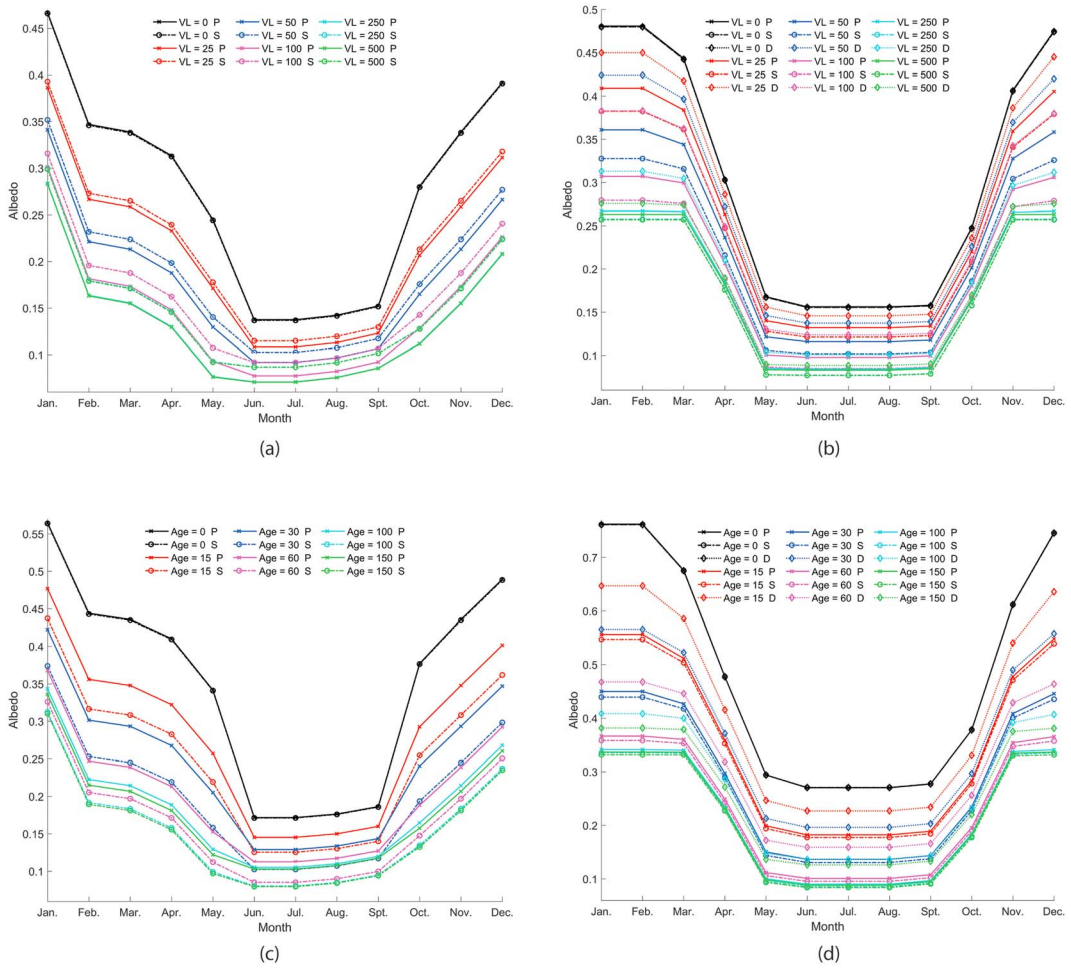

Figure 1. Predicted monthly-mean surface albedo values with different biomass volumes and forest ages. P stands for pine forest, S stands for spruce forest and D stands for deciduous forest. The results using volume in the interaction term are shown in a for data 1 and b for data 2, and the results using age in the interaction term are given in c for data 1 and (b) for data 2, respectively. The selected values for volume and age are V = {0; 30; 50; 100; 250; 500}m3/ha, and A = {0; 15; 30; 60; 100; 150}years, respectively, to represent the different states of the standing forests. The default fixed meteorological parameters are T = {260; 269; 270; 275; 280; 290; 290; 280; 275; 270; 270; 265}K for temperature and S = {300; 300; 150; 50; 10; 0; 0; 0; 0; 10; 80; 265} mm for SWE. Different colors are used to indicate different states of the forests (from young to mature forest), and different line styles to indicate different kinds of forests (pine forest with solid line, spruce forest with dashed line and deciduous forest with dotted line).

significant with deciduous forests. Deciduous forest has higher albedo values than coniferous forest, especially during winter, owing to the defoliation process and the reduced snow masking effect of snow on the ground. As expected, surface albedo values are higher in Data 2 because the average elevation is higher than in Data 1, where the snow appears earlier in the fall and melts later in springtime. The use of forest age for the vegetation structure parameter predicts higher albedo than using volume as forest structure parameter in the model when the forest is young. This is because the land cover datasets only contain information on forest but not open land, and therefore the model predicts the albedos for open land or very young forest by extrapolation. Since volume covers larger range comparing to age (see Supplementary Figure S1), the extrapolation is smoother and thus predict lower values.

3.3. Model Inter-Comparison

We investigate the performance of different models using volume as the forest structure parameter in the interaction term that differs in the SWE functional form. Selected models for this investigation are numbers 7, 4 and 3, from Table 3 for Data 1. The model number 7 has an additive sigmoid function for SWE-albedo

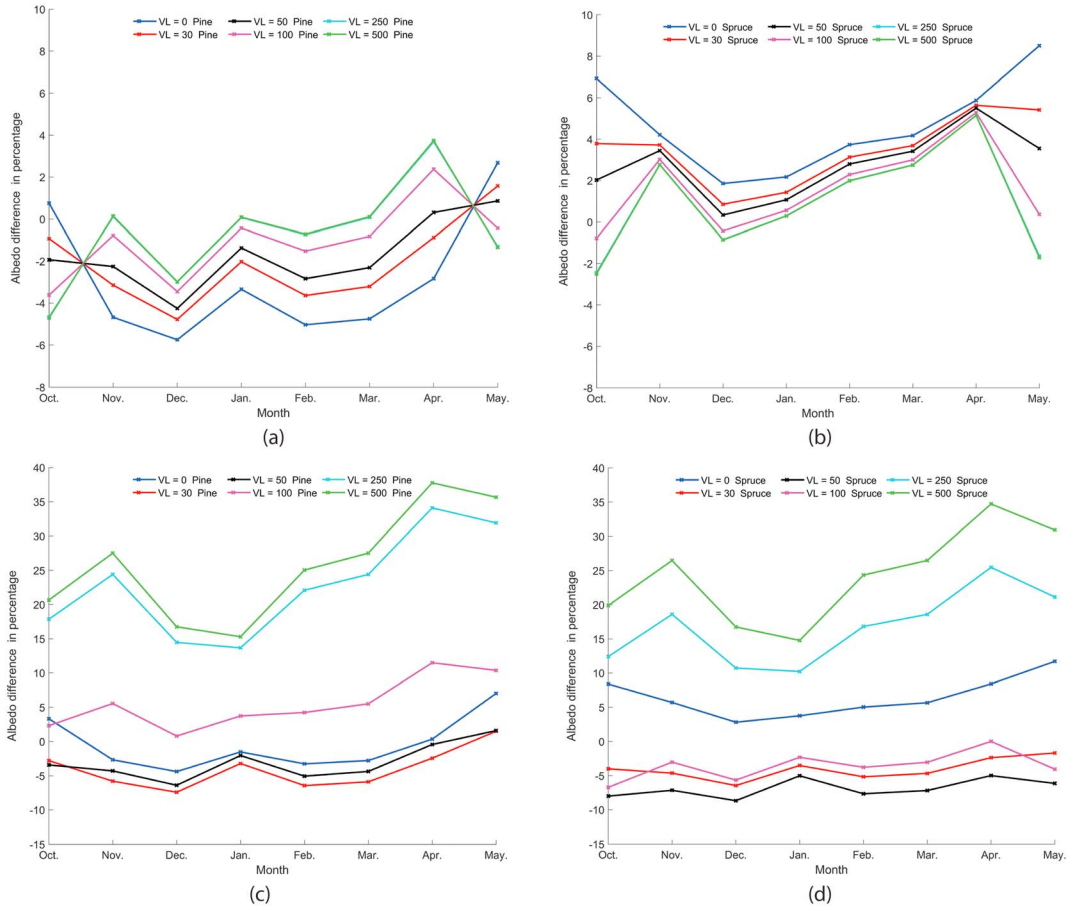

Figure 2. Results for model inter-comparison. The figures illustrate the relative differences among model outputs in terms of percentage with respect to model number 7 from Table 3 for data 1. Selected models for this investigation are numbers 4 and 3 from Table 3. Only the period with snow is chosen, and different colors represent different stages of the forests. The relative differences between model 4 and model 7 are given in a for pine forest and b for spruce forest, and corresponding results between model 3 and model 7 are shown in c for pine forest and (b) for the spruce forest, respectively.

relationship with a sigmoid midpoint estimated from the data. In model number 4, the additive sigmoid function term $f_4(S)$ is replaced by an exponential function, whereas model number 3 does not include the additive term, and the SWE is only present in the interaction term. The relative differences between model outputs are calculated in terms of relative percentage differences with respect to model number 7, and the results are shown in Figure 2. In this way, we can single out the importance of the additive term for SWE and the role of the functional form, either sigmoid or exponential, used for SWE-albedo relationship in the additive term.

With Data 1, the relative difference, with respect to model number 7, is generally smaller for model number 4 (Figures 2(a) and 2b) than for model number 3 (Figures 2c and 2d), for both pine and spruce forests over the whole year. This is reasonable since both forests are coniferous, and model number 7 and model number 4 are more similar to each other than model number 3. In model number 4, the relative differences are only within 6% and 10% for pine and spruce, respectively, and tend to decrease at higher volumes. The difference between model number 7 and model number 3 is much significant, especially for the forest with higher volumes or standing ages. The difference can up to 40% for mature forests for both pine and spruce in spring. The reason is that when the forests grow, the forest gets darker and the importance of the SWE is increasing, so the additive term becomes crucial in the model.

3.4. Transferability Analysis

One of the main purposes of our study is to produce models that can be used for easily predicting surface albedo using the forest structures and meteorological parameters in boreal areas. Therefore, we carry out

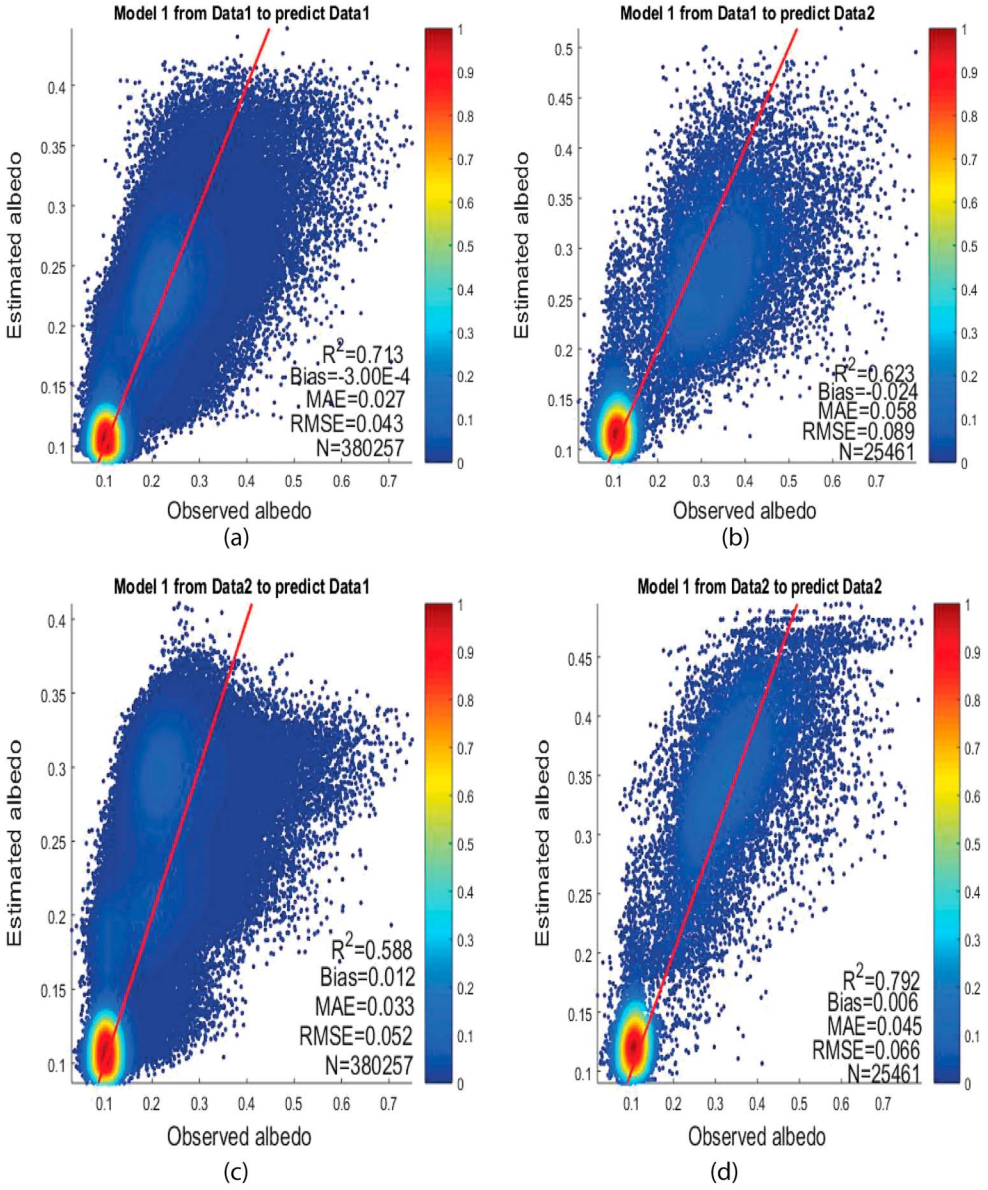

Figure 3. Results of the transferability analysis for model 1 (corresponding to number 7 in Table 3 for data 1 and number 12 in Table 4 for data 2) plotting the predicted surface albedo against the satellite-derived values (treated as observed values). Model 1 uses volume as the forest structure parameter in the interaction term. The figure also shows the respective statistical indicators, such as bias, mean absolute error (MAE), R^2, root mean square error (RMSE) and number of points (N). the red lines are $y = x$ and the color scale indicates the normalized densities of the points.

transferability analysis to validate the applicability of the selected models. In the first group of analyses, we use the model given in equation (19), named "Model 1" hereafter, which corresponds to number 7 in Table 3 for Data 1 and number 12 in Table 4 for Data 2. In this analysis, we first estimate the unknown parameters from the first area and then use the corresponding model to predict the surface albedo in the second area. Similarly, we estimate the parameters from the second area and then use the respective model to predict the surface albedo in the first area. The chosen characteristics, bias, mean absolute error (MAE), R^2 and root mean square error (RMSE) are calculated and reported. The results are shown in the off-diagonal sub-figures in Figure 3. As a benchmark, we include the results of predictive performance using estimates from the first and second areas to predict themselves, respectively. This is also informing about the model performances within the same region of the model estimates, thereby showing possible bias of the downscaling unmixing method from 500 m to 16 m.

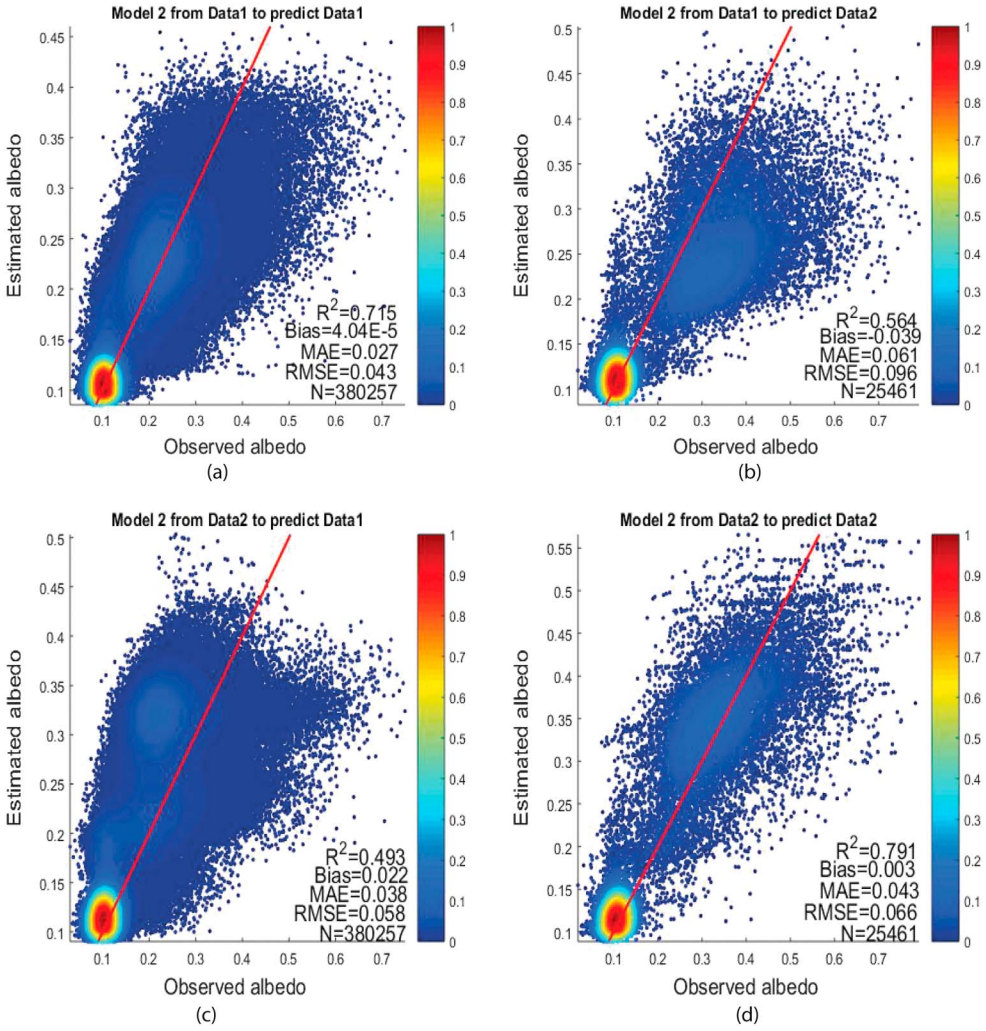

Figure 4. Results of transferability analysis for model 2 plotting the predicted surface albedo against the satellite-derived values (treated as observed values). Model 2 uses age as the forest structure parameter in the interaction term. The figure also shows the respective statistical indicators, such as bias, mean absolute error (MAE), R^2, root mean square error (RMSE) and number of points (N). the red lines are $y = x$ that are used as references and the color scale indicates the densities of the points.

The transferability of the selected model in equation (20), which uses age instead of volume as the forest structure in the interaction term, is also analyzed. The same procedure as above is applied, and the results are illustrated in Figure 4.

Figures 3(a), 3(d), 4(a), and 4d show that results are unbiased (the values of bias are close to 0) and robust, as the statistical indicators show good model performances. This means that each model produces robust estimates when applied to the same area of its original parameterizations. Figures 3(b), 3(c), 4b and 4c illustrate the model outputs and associated statistical indicators when the models are applied to different locations than those from their parameterization.

In general, after comparing the scores in Figure 3 and Figure 4, the model using volume in the interaction term $f_5(V, S)$ performs better than the model with age in the interaction term$f_5(A, S)$. The reason is that the size of the trees with same age can differ, and there are heterogeneities in the two different areas (flat managed forest vs. mountainous forest) that are better captured when volume is used instead of age as the forest structure parameter. The volume of trees tends to be more homogenous and a better indicator of forest growth and snow-masking effects. From Figure 3b and Figure 4b, a negative bias is observed when we use the estimates from the first area to predict the second area, meaning that surface albedo is likely to be

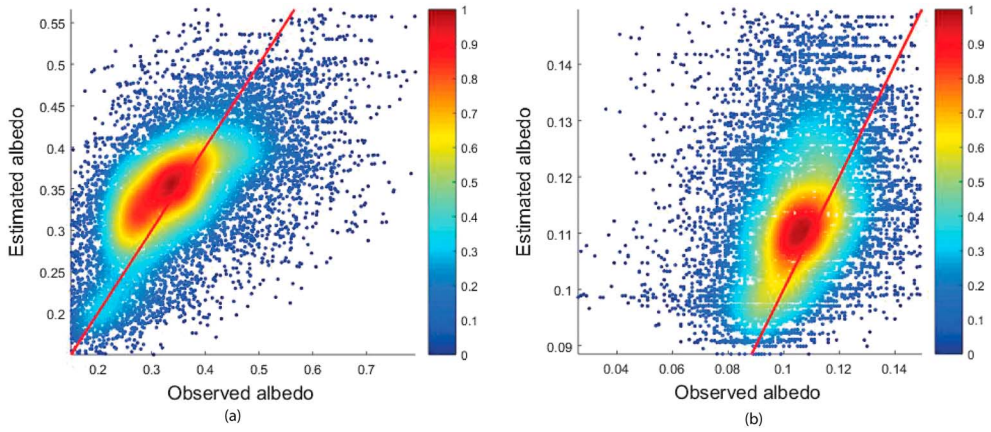

Figure 5. Split of Figure 4d with threshold value 0.15. A estimated albedo greater than 0.15 and b estimated albedo lower than 0.15 against the observed albedo.

underestimated. The reason is that the first area is rich in coniferous trees but has few deciduous trees, and therefore the model has limitations to represent the higher values of albedo in winter months of deciduous trees relative to coniferous trees, which are more abundant in the second area. For the same reason, there is the opposite effect when using the estimates from the second area to predict surface albedo in the first area, where we observe a positive bias. Results in Figure 3 and Figure 4 also show that our models do not capture the outliers in the datasets. This is expected, since we used a statistical approach and the models are unbiased in the mean. It is important to point out that it makes no sense to compare the results directly between Figure 3a and Figure 3d or between Figure 4a and Figure 4d, using the same standard statistical indicators. Data 1 and Data 2 have different densities of forests as well as temporal and spatial ranges, which leads to different number of points.

Figures 3 and 4 also indicate the possible presence of two major data clusters separated with a threshold around a value of 0.15. We therefore zoom into Figure 4d and explore possible hidden trends by splitting the figure into two parts, and show results in Figure 5. We find that both subplots generally exhibit a similar pattern, and model performances are relatively similar in both data clusters.

We further explore the transferability properties of the models under varying temporal and spatial scales. The temporal scale is investigated by introducing a third dataset, Data 3. Data 3 is a subset of Data 1 and it covers the same spatial domain, but for a shorter temporal period. Data 3 have the same temporal range of Data 2 (2 years of satellite-derived surface albedo estimates), but the numbers of the forest plots are different due to the lower forest density in the second area. We introduced this new dataset to test model transferability across datasets with different time ranges. The results are shown in Supplementary Figures S4 and S6, and the statistical scores indicate that model performances are not significantly affected by the temporal scale of the parameterization. Models results are thus relatively insensitive to variations in time scales. Comparing the corresponding results given in Figure 3a with Figure S4(d), and Figure 4a with Figure S5(d), the statistical scores are improved. This is because Data 3 is from the same area of Data 1, and it has a shorter temporal range and therefore lower scattered outputs (smaller number of N).

We test the transferability across different spatial scales of the sample domain introducing a fourth dataset, namely Data 4. Data 4 has the same time range of Data 1 and Data 3, but double size. We use the model estimates for Data 1 and Data 3 to predict albedo values in Data 4. Results show that the models do not decrease the performance when applied across different spatial ranges using either volume or age as the forest structure parameters (see Figure S6 in the supplementary material).

3.5. Model Variant for Snow-Free Conditions

The performance between the full-year model given in equation (19) is compared against a variant where a specific model for snow-free conditions only is used. We consider the option of equation with only volume as the forest structure parameter to be consistent with the version of the full-year model.

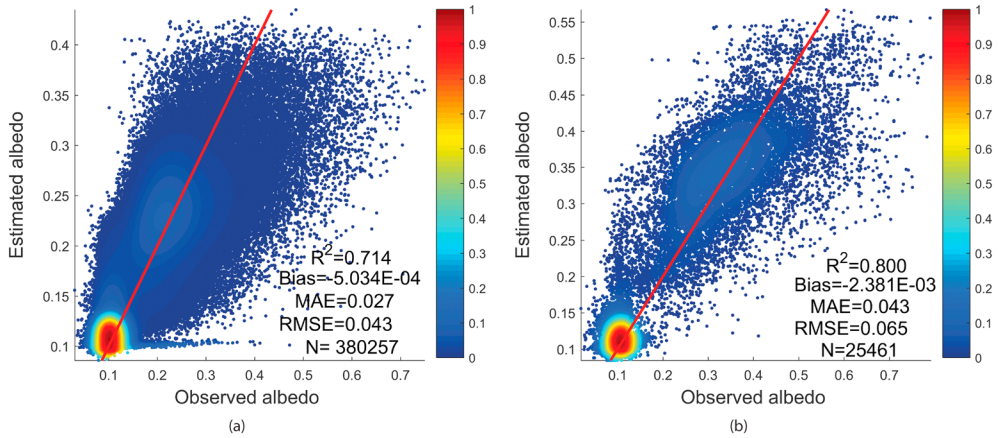

Figure 6. Results of snow-free and snow-present models with data 1 a and data 2 b plotting the predicted surface albedo against the satellite-derived values (treated as observed values). The approach refers to equations (20) and (21), where we use volume as the forest structure parameter in the interaction term. The figure also shows the respective statistical indicators, such as bias, mean absolute error (MAE), R^2, root mean square error (RMSE) and number of points (N). the red lines are $y = x$ that are used as references and the color scale indicates the densities of the points.

Figure 6 shows the statistical performance against observations of the combined snow-free and snow-present albedo estimates, so to make them comparable with the findings of the full-year model since the performance from these two variants is not directly comparable with the full-year model. The two coefficients of determination R^2 for Data 1 and Data 2 are 0.714 and 0.8, which are slightly better than the full year model (0.713 and 0.792, respectively). The other scores are slightly improved as well. From this analysis, we can conclude that the representation of the albedo-SWE relationship in the full-year model is sufficient to capture variations related to snow presence. The functional form used to model albedo response to SWE is highly sensitive to changes in SWE at low SWE values, so that the variability associated with the absence or presence of snow is duly represented in the full-year model. A model variant specific for snow free conditions does not improve the analysis of the two clusters either, because, as mentioned in the previous sections, low albedo values are also possible in presence of snow. The full-year model is thus a preferred option than the model variant with snow-free/present conditions, owing to its simpler applicability and not need for differentiated datasets and models.

3.6. Sensitivity Analysis Based on the Constraints

When constructing models, it is important to ensure that models are accurate and at the same time feasible to be solved. This is usually the main rational to introduce constraints in model frameworks. In order to simplify the analysis and harmonize the comparison, we introduce only one parameter to control the treatment of differences between forests and illustrate the results of the models given in the equation (19) and equation (20) with Data 1 and Data 2. The constraints in (13) can then be written as

$$\begin{aligned} \theta_l^{1,3} &= c^*\theta^2 \\ \theta_u^{1,3} &= \frac{\theta^2}{c} \end{aligned} \tag{21}$$

with $c = \{0.05, 0.1, 0.15, \ldots, 1\}$. With this setting, smaller values of c allow the bigger difference between the parameters for different forests, and bigger values of c introduce stricter constraints. The analysis is based on the investigation of the changes in the regression coefficient R^2 as a function of the changes in c (Figure 7).

Results show that the difference in the characteristics between the two types of coniferous forests (pine and spruce) is rather small since the values of the regression coefficient only change slightly when we tighten the constraints (see red lines in Figure 7a and Figure 7b for Data 1, which is rich in coniferous species). This means that the two different coniferous species can be treated as the same tree without significant losses in terms of overall model accuracy. On the other hand, results show the different characteristics between deciduous forests and coniferous forests. In this case, the regression coefficients change with tighter constraints (see blue

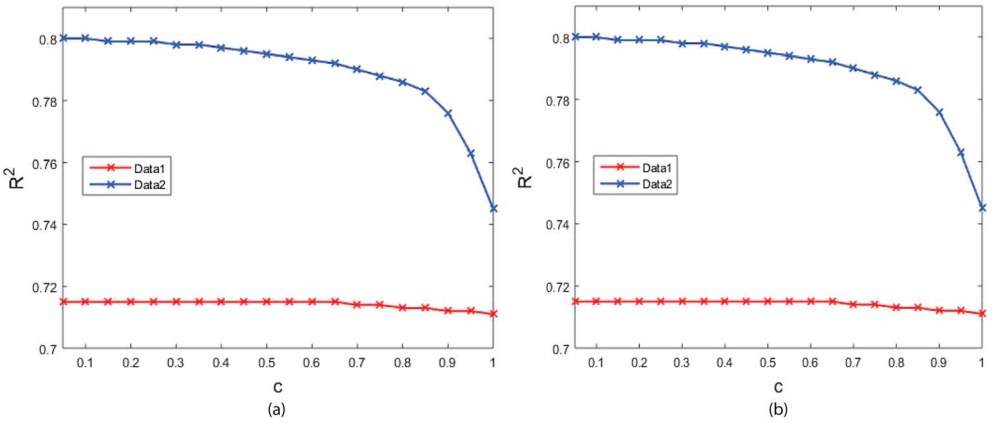

Figure 7. Results of sensitivity analysis based on the constraints. The results for model 1 and model 2 are shown in a and b, respectively. The red lines illustrate the results for data 1 and the blue line shows the results for data 2. The constraints parameter c has the value from 0.05 to 1 with step size 0.05, and bigger values of c introduce stricter constraints to the parameters for different forests.

lines in Figure 7a and Figure 7b for Data 2, which is rich in deciduous species), especially when c is increasing from 0.7 to 1. No significant change in R^2 occurs when c is smaller than about 0.6. This sensitivity analysis thus provides an indication for selecting the best vales of the constraints under the trade-offs between model robustness and solvability.

4. Conclusion

This study provides a simple and fast approach to simulate and predict surface albedo using forest structure and metrological parameters. By utilizing the high spatial resolution of the forest structure datasets, surface albedo can be estimated at high-resolution (from 500 m of satellite retrievals to 16 m) and predicted using statistical models. In this study, many models are investigated and the ones with the best statistical scores are presented and discussed in detail. We find that they are generally unbiased with high values of the regression coefficient R^2, low MAE, and low RMSE, and have good transferability properties. They can be used to predict surface albedo in another area with similar forest composition (deciduous vs. coniferous species) and meteorological parameters regardless of the spatial range and time scales.

This analysis offers many alternative models that can be used to study the effects of forest management and climate change on surface albedo. There is a large flexibility and choice for practitioners, depending on the particular case. Models can have different complexity in terms of functional forms and variables of forest structure (volume and/or age) and climate (temperature and snow water equivalents). The choice of the model can reflect the considerations highlighted in this paper and the different situations in terms of data availability of the particular study (for example, if one has only access to volume or age information, or both). In general, robust outcomes are obtained with climate variables included in model formulations as additive items and in the interaction term (especially SWE), whereas the representation of forest structure information is needed in the interaction term but not as additive terms. Models using volume instead of age as the forest structure parameter perform better, since volume is a more homogeneous indicator across different areas. The representation of the albedo-SWE interactions in the full-year models is sufficient to capture albedo sensitivity to snow, as benchmarked in the comparison with the snow-free/present model variant. The two models selected in the results section are an example of model options that reflect these considerations.

The computational time for different models depend on how complicated models are and how big the dataset is. It varies from several minutes to several hours on a desktop computer. However, after the model is trained and parameters estimated, it is fast to use these models in applications aiming at quantifying effects of changes in forest structure and climate conditions on surface albedo. The equations with estimated parameters can be used by non-expert by directly plugging in forest structure and meteorological parameters to obtain the surface albedo estimates.

Availability of simple parametric models for predicting surface albedo changes as a result of forest management or climate change can enhance the inclusion of biophysical effects in climate impact analysis. The approach presented in this paper provides a range of models that can fit a variety of applications. Currently, our models are tested and validated for Norwegian forests. The models are expected to perform in other boreal forest areas with similar tree species, and future work can specifically test the transferability in other boreal locations outside Norway. A possible extension of this work also includes the possibility to incorporate other information of forest structures and metrological parameters, and potentially integrate refined model versions (possibly re-estimated with the more recent MCD43A1 collection 6) with climate impact assessment frameworks and land surface models.

Acknowledgments

Authors acknowledge the support of the Norwegian Research Council through the project Bio4Clim (Nr. 244074). We acknowledge the National forest Inventory (NFI) database 'Stat-Skog' (http://www.skogoglandskap.no/kart/SAT-SKOG/map_view), which provides an overview of the forest resources of the country. We also acknowledge the Norwegian Meteorological Institute (https://www.met.no/en) for producing daily observations of meteorological information on surface air temperature and SWE in Norway (https://www.met.no/en/free-meteorological-data). Albedo estimate is gathered from the Moderate Resolution Imaging Spectroradiometer MCD43A3 (https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd43a3). The processed dataset and the model results shown in this paper are available on request to the corresponding author (xiangping.hu@ntnu.no).

References

Amante, C., & Eakins, B. W. (2009). ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis, edited, National Geophysical Data Center, NOAA, NOAA Technical Memorandum NESDIS NGDC-24. https://doi.org/10.7289/V5C8276M

Anderson, R. G., Canadell, J. G., Randerson, J. T., Jackson, R. B., Hungate, B. A., Baldocchi, D. D., et al. (2011). Biophysical considerations in forestry for climate protection. *Frontiers in Ecology and the Environment*, *9*(3), 174–182. https://doi.org/10.1890/090179

Bala, G., Caldeira, K., Wickett, M., Phillips, T. J., Lobell, D. B., Delire, C., & et al. (2007). Combined climate and carbon-cycle effects of large-scale deforestation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(16), 6550–6555. https://doi.org/10.1073/pnas.0608998104

Bartlett, P. A., MacKay, M. D., & Verseghy, D. L. (2006). Modified snow algorithms in the Canadian land surface scheme: Model runs and sensitivity analysis at three boreal forest stands. *Atmosphere-Ocean*, *44*(3), 207–222. https://doi.org/10.3137/ao.440301

Bathiany, S., Claussen, M., Brovkin, V., Raddatz, T., & Gayler, V. (2010). Combined biogeophysical and biogeochemical effects of large-scale forest cover changes in the MPI earth system model. *Biogeosciences*, *7*(5), 1383–1399. https://doi.org/10.5194/bg-7-1383-2010

Betts, R. A. (2000). Offset of the potential carbon sink from boreal forestation by decreases in surface albedo. *Nature*, *408*(6809), 187–190. https://doi.org/10.1038/35041545

Betts, R. A., Falloon, P. D., Goldewijk, K. K., & Ramankutty, N. (2007). Biogeophysical effects of land use on climate: Model simulations of radiative forcing and large-scale temperature change. *Agricultural and Forest Meteorology*, *142*(2–4), 216–233. https://doi.org/10.1016/j.agrformet.2006.08.021

Bonan, G. B. (2008). Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science*, *320*(5882), 1444–1449. https://doi.org/10.1126/science.1155121

Bright, R. M., Astrup, R., & Stromman, A. H. (2013). Empirical models of monthly and annual albedo in managed boreal forests of interior Norway. *Climatic Change*, *120*(1–2), 183–196. https://doi.org/10.1007/s10584-013-0789-1

Bright, R. M., Zhao, K., Jackson, R. B., & Cherubini, F. (2015). Quantifying surface albedo and other direct biogeophysical climate forcings of forestry activities. *Global Change Biology*, *21*(9), 3246–3266. https://doi.org/10.1111/gcb.12951

Caiazzo, F., Malina, R., Staples, M. D., Wolfe, P. J., Yim, S. H. L., & Barrett, S. R. H. (2014). Quantifying the climate impacts of albedo changes due to biofuel production: a comparison with biogeochemical effects. *Environmental Research Letters*, *9*(2), 024015.

Cherubini, F., Bright, R. M., & Stromman, A. H. (2012). Site-specific global warming potentials of biogenic CO2 for bioenergy: contributions from carbon fluxes and albedo dynamics. *Environmental Research Letters*, *7*(4), 045902.

Cherubini, F., Fuglestvedt, J., Gasser, T., Reisinger, A., Cavalett, O., Huijbregts, M. A. J., et al. (2016). Bridging the gap between impact assessment methods and climate science. *Environmental Science & Policy*, *64*, 129–140. https://doi.org/10.1016/j.envsci.2016.06.019

Cherubini, F., Vezhapparambu, S., Bogren, W., Astrup, R., & Stromman, A. H. (2017). Spatial, seasonal, and topographical patterns of surface albedo in Norwegian forests and cropland. *International Journal of Remote Sensing*, *38*(16), 4565–4586. https://doi.org/10.1080/01431161.2017.1320442

Claussen, M., Brovkin, V., & Ganopolski, A. (2001). Biogeophysical versus biogeochemical feedbacks of large-scale land cover change. *Geophysical Research Letters*, *28*(6), 1011–1014. https://doi.org/10.1029/2000gl012471

Dickinson, R. E. (1983). Land surface processes and climate surface albedos and energy-balance. *Advances in Geophysics*, *25*, 305–353. https://doi.org/10.1016/S0065-2687(08)60176-4

Essery, R. (2013). Large-scale simulations of snow albedo masking by forests. *Geophysical Research Letters*, *40*, 5521–5525. https://doi.org/10.1002/grl.51008

Francesco, C., Bo, H., Xiangping, H., Merja, T., & Anders Hammer, S. (2018). Quantifying the climate response to extreme land cover changes in Europe with a regional model. *Environmental Research Letters*, *13*(7). https://doi.org/10.1088/1748-9326/aac794

GAMS (2017). *Continuous nonlinear optimization for engineering applications in GAMS technology*. Washington, DC, USA: General Algebraic Modeling System (GAMS), Edited.

Gjertsen, A. K., & Nilsen, J.-E. (2012). SAT-SKOG. Et skogkart basert på tolking av satellittbilderRep., 54 pp, Rapport fra Skog og landskap 23/12.

Hibbard, K., Janetos, A., van Vuuren, D. P., Pongratz, J., Rose, S. K., Betts, R., et al. (2010). Research priorities in land use and land-cover change for the earth system and integrated assessment modelling. *International Journal of Climatology*, *30*(13), 2118–2128. https://doi.org/10.1002/joc.2150

Hollinger, D. Y., Ollinger, S. V., Richardson, A. D., Meyer, T. P., Dails, D. P., Martin, M. E., et al. (2010). Albedo estimates for land surface models and support for a new paradigm based on foliage nitrogen concentration. *Global Change Biology*, *16*(2), 696–710. https://doi.org/10.1111/j.1365-2486.2009.02028.x

Jackson, R. B., Randerson, J. T., Canadell, J. G., Anderson, R. G., Avissar, R., Baldocchi, D. D., et al. (2008). Protecting climate with forests. *Environmental Research Letters*, *3*(4), 044006. https://doi.org/10.1088/1748-9326/3/4/044006

Juang, J. Y., Katul, G., Siqueira, M., Stoy, P., & Novick, K. (2007). Separating the effects of albedo from eco-physiological changes on surface temperature along a successional chronosequence in the southeastern United States. *Geophysical Research Letters*, *34*, L21408. https://doi.org/10.1029/2007GL031296

Kuusinen, N., Lukes, P., Stenberg, P., Levula, J., Nikinmaa, E., & Berninger, F. (2014). Measured and modelled albedos in Finnish boreal forest stands of different species, structure and understory. *Ecological Modelling*, *284*, 10–18. https://doi.org/10.1016/j.ecolmodel.2014.04.007

Kuusinen, N., Stenberg, P., Korhonen, L., Rautiainen, M., & Tomppo, E. (2016). Structural factors driving boreal forest albedo in Finland. *Remote Sensing of Environment*, *175*, 43–51. https://doi.org/10.1016/j.rse.2015.12.035

Kuusinen, N., Tomppo, E., & Berninger, F. (2013). Linear unmixing of MODIS albedo composites to infer subpixel land cover type albedos. *International Journal of Applied Earth Observation and Geoinformation*, *23*, 324–333. https://doi.org/10.1016/j.jag.2012.10.005

Kuusinen, N., Tomppo, E., Shuai, Y., & Berninger, F. (2014). Effects of forest age on albedo in boreal forests estimated from MODIS and Landsat albedo retrievals. *Remote Sensing of Environment*, *145*, 145–153. https://doi.org/10.1016/j.rse.2014.02.005

Lewis, P. (1995). The utility of kernel-driven BRDF models in global BRDF and albedo studies, paper presented at Geoscience and Remote Sensing Symposium, 1995. IGARSS'95.'Quantitative Remote Sensing for Science and Applications', International, IEEE.

Lukeš, P., Rautiainen, M., Manninen, T., Stenberg, P., & Mõttus, M. (2014). Geographical gradients in boreal forest albedo and structure in Finland. *Remote Sensing of Environment*, *152*, 526–535. https://doi.org/10.1016/j.rse.2014.06.023

Lukes, P., Stenberg, P., Mottus, M., Manninen, T., & Rautiainen, M. (2016). Multidecadal analysis of forest growth and albedo in boreal Finland. *International Journal of Applied Earth Observation and Geoinformation*, *52*, 296–305. https://doi.org/10.1016/j.jag.2016.07.001

Lukes, P., Stenberg, P., & Rautiainen, M. (2013). Relationship between forest density and albedo in the boreal zone. *Ecological Modelling*, *261-262*, 74–79. https://doi.org/10.1016/j.ecolmodel.2013.04.009

Mahmood, R., Pielke, R. A., Hubbard, K. G., Niyogi, D., Dirmeyer, P. A., McAlpine, C., et al. (2014). Land cover changes and their biogeophysical effects on climate. *International Journal of Climatology*, *34*(4), 929–953. https://doi.org/10.1002/joc.3736

Mohr, M. (2008). New routines for gridding of temperature and precipitation observations for "seNorge. no", *Met. no Report*, *8*, 2008.

Muñoz, I., Campra, P., & Fernandez-Alba, A. R. (2010). Including CO2-emission equivalence of changes in land surface albedo in life cycle assessment. Methodology and case study on greenhouse agriculture. *The International Journal of Life Cycle Assessment*, *15*(7), 672–681. https://doi.org/10.1007/s11367-010-0202-5

Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., et al. (2010). Technical description of version 4.0 of the Community Land Model (CLM).

Perugini, L., Caporaso, L., Marconi, S., Cescatti, A., Quesada, B., de Noblet-Ducoudre, N., et al. (2017). Biophysical effects on temperature and precipitation due to land cover change. *Environmental Research Letters*, *12*(5). https://doi.org/10.1088/1748-9326/aa6b3f

Prestele, R., Arneth, A., Bondeau, A., de Noblet-Ducoudre, N., Pugh, T. A. M., Sitch, S., et al. (2017). Current challenges of implementing anthropogenic land-use and land-cover change in models contributing to climate change assessments. *Earth System Dynamics*, *8*(2), 369–386. https://doi.org/10.5194/esd-8-369-2017

Qu, X., & Hall, A. (2007). What controls the strength of snow-albedo feedback? *Journal of Climate*, *20*(15), 3971–3981. https://doi.org/10.1175/JCLI4186.1

Qu, X., & Hall, A. (2014). On the persistent spread in snow-albedo feedback. *Climate Dynamics*, *42*(1–2), 69–81. https://doi.org/10.1007/s00382-013-1774-0

Rechid, D., Raddatz, T. J., & Jacob, D. (2009). Parameterization of snow-free land surface albedo as a function of vegetation phenology based on MODIS data and applied in climate modelling. *Theoretical and Applied Climatology*, *95*(3–4), 245–255. https://doi.org/10.1007/s00704-008-0003-y

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., et al. (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, *83*(1–2), 135–148. https://doi.org/10.1016/s0034-4257(02)00091-3

Thackeray, C. W., Fletcher, C. G., & Derksen, C. (2015). Quantifying the skill of CMIP5 models in simulating seasonal albedo and snow cover evolution. *Journal of Geophysical Research: Atmospheres*, *120*, 5831–5849. https://doi.org/10.1002/2015jd023325

Tomppo, E., Olsson, H., Stahl, G., Nilsson, M., Hagner, O., & Katila, M. (2008). Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*, *112*(5), 1982–1999. https://doi.org/10.1016/j.rse.2007.03.032

Tomter, S., Hylen, G., & Nilsen, J.-E. Ø. (2010). Development of Norway's National Forest Inventory. National Forest InventoriesRep., 411-424 pp, National Forest Inventories. Pathways for common reporting.

Verseghy, D., McFarlane, N., & Lazare, M. (1993). CLASS—A Canadian land surface scheme for GCMs, II. Vegetation model and coupled runs. *International Journal of Climatology*, *13*(4), 347–370. https://doi.org/10.1002/joc.3370130402

Volodin, E. M., Dianskii, N. A., & Gusev, A. V. (2010). Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya Atmospheric and Oceanic Physics*, *46*(4), 414–431. https://doi.org/10.1134/S000143381004002x

Wang, L., Cole, J. N., Bartlett, P., Verseghy, D., Derksen, C., Brown, R., & et al. (2016). Investigating the spread in surface albedo for snow-covered forests in CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *121*, 1104–1119. https://doi.org/10.1002/2015JD023824

Zhao, K., & Jackson, R. B. (2014). Biophysical forcings of land-use changes from potential forestry activities in North America. *Ecological Monographs*, *84*(2), 329–353. https://doi.org/10.1890/12-1705.1