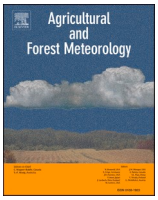
Agricultural and Forest Meteorology 332 (2023) 109362

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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



A simplified multi-model statistical approach for predicting the effects of forest management on land surface temperature in Fennoscandia

Bo Huang [a,*], Yan Li [b,*], Yi Liu [c], Xiangping Hu [a], Wenwu Zhao [d], Francesco Cherubini [a]

[a] Industrial Ecology Programme, Department of Energy and Process Engineering, Norwegian University of Science and Technology (NTNU), N-7491, Trondheim, Norway
[b] Institute of Meteorology, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany
[c] EmTech Group, Group of Research and Development (GRD), Det Norske Veritas (DNV), Oslo, Norway
[d] State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China

ARTICLE INFO

Keywords:
Forest management
Climate change
Surface temperature
Machine learning

ABSTRACT

Forests interact with the local climate through a variety of biophysical mechanisms. Observational and modelling studies have investigated the effects of forested vs. non-forested areas, but the influence of forest management on surface temperature has received far less attention owing to the inherent challenges to adapt climate models to cope with forest dynamics. Further, climate models are complex and highly parameterized, and the time and resource intensity of their use limit applications. The availability of simple yet reliable statistical models based on high resolution maps of forest attributes representative of different development stages can link individual forest management practices to local temperature changes, and ultimately support the design of improved strategies. In this study, we investigate how forest management influences local surface temperature (LSTs) in Fennoscandia through a set of machine learning algorithms. We find that more developed forests are typically associated with higher LST than young or undeveloped forests. The mean multi-model estimates from our statistical system can accurately reproduce the observed LST. Relative to the present state of Fennoscandian forests, fully develop forests are found to induce an annual mean warming of 0.26 °C (0.03/0.69 °C as 5th/95th percentile), and an average cooling effect in the summer daytime from -0.85 to -0.23 °C (depending on the model). On the contrary, a scenario with undeveloped forests induces an annual average cooling of -0.29 °C (-0.61/-0.01 °C), but daytime warming in the summer that can be higher than 1 °C. A weak annual mean cooling of -0.01 °C is attributed to forest harvest from 2015 to 2018, with an increased daytime temperature in summer of about 0.04 °C. Overall, this approach is a flexible option to study effects of forest management on LST that can be applied at various scales and for alternative management scenarios, thereby helping to improve local management strategies with consideration of effects on local climate.

1. Introduction

Forests cover around 30% of the global ice-free land surface and are widely distributed from tropical to boreal regions (Crowther et al., 2015; Hansen et al., 2013). They are a primary component of many national economies and interact with the global carbon cycle as they can be both sources and sinks of CO_2 (Luyssaert et al., 2018; Pan et al., 2011). Forests also play a vital role to support ecosystem services, including local climate regulation via water and heat exchanges with the atmosphere (called biophysical effects) (Anderson et al., 2011; Bonan, 2008). Depending on locations, tree species, and forest management type, these biophysical effects can either reinforce or counteract the carbon benefits of forests. Generally, presence of forests instead of open land induces regional land surface cooling in the tropics due to large amounts of evapotranspiration, while the effects are more uncertain and spatially heterogenous at mid-latitudes (Alkama and Cescatti, 2016; Li et al., 2016, 2015; Perugini et al., 2017). Outside tropical climates, contributions from radiative (i.e., changes in surface albedo and radiation fluxes) and non-radiative (i.e., evapotranspiration) effects of forests greatly vary depending on background conditions (climate, soil moisture), forest type (deciduous or coniferous), type of open land used for the benchmark (grassland or cropland), or method used for the analysis (observations vs. climate models) (Ge et al., 2019; Li et al., 2016; Pitman et al., 2011; Tian et al., 2022). For example, the nonradiative

* Corresponding authors.
E-mail addresses: bo.huang@ntnu.no (B. Huang), yan.li@met.fu-berlin.de (Y. Li).

https://doi.org/10.1016/j.agrformet.2023.109362
Received 19 October 2022; Received in revised form 9 January 2023; Accepted 5 February 2023
Available online 9 February 2023

cooling effect of forests is typically stronger using satellite retrievals (Ge et al., 2019), but for the same region the use of a coupled land-atmosphere model instead shows a dominant radiative effect on local temperature changes (i.e., a net warming effect) (Tian et al., 2022). In the boreal region, where the snow masking effect in late winter and early spring is stronger, forests typically have a low surface albedo that tends to warm the surface, with net effects that are of a similar order of magnitude to the carbon benefits (Mykleby et al., 2017; Perugini et al., 2017). However, forests usually have stronger latent heat fluxes in the summer, with seasonal cooling relative to open lands (Alkama and Cescatti, 2016; Naudts et al., 2016).

Although less studied due to inherent complexities, the net response of local temperature is also sensitive to the forest structure and management practices that influence its development (Kellomaki et al., 2021; Kumkar et al., 2020; Luyssaert et al., 2014; Naudts et al., 2016). Forest management is widespread in Fennoscandia (i.e., Norway, Sweden, and Finland) (McGrath et al., 2015; Zhou et al., 2021). It affects forest structure attributes such as tree density, standing volume (V), leaf area index (LAI), crown length (CL), and canopy height (CH), which in turn control the surface energy, moisture, and momentum fluxes via surface albedo, evapotranspiration, and surface roughness (Anderson et al., 2011; Jackson et al., 2008). Changes in forest structure thus have the potential to influence the local climate. However, there is a weak representation of various forest development stages within existing land surface models, where the same parameterization of LAI or CH is typically used for the different vegetation classes (e.g., coniferous or deciduous forests have their own set of fixed parameters, irrespective of forest development stage or age) (Lawrence et al., 2019). There are increasing efforts in embedding different forest stage parameterizations within gridded land use datasets for various forest types (Majasalmi et al., 2018; McGrath et al., 2015), but applications remain limited and largely idealized due to the complexity to link forest dynamics with climate models (Kumkar et al., 2020; Luyssaert et al., 2018; Naudts et al., 2016). Global climate models are not suitable to study the local impacts of forests on climate due to their coarse spatial resolutions and uncertainties in the physical processes. Regional climate models can achieve a finer resolution, but they still have challenges to simulate the complex spatial and temporal patterns of the effects of forest dynamics, and models often show contradictory results to land cover changes (Davin et al., 2020; de Noblet-Ducoudré et al., 2012).

Preliminary studies have investigated the effects of forest management on European climates. The application of a land-atmosphere model to the reconstructed history of forest management in Europe shows an increase of 0.12 °C in summertime atmospheric temperature due to species conversion from broadleaf to coniferous (Naudts et al., 2016). A similar study found that options that locally maximize the carbon sink through carbon sequestration are preferable over management practices to meet climate objectives in Europe (Luyssaert et al., 2018). More specific to the Fennoscandian context, prescribed idealized scenarios of structural changes were used to explore their influence on LST via offline simulations with a land surface model (Kumkar et al., 2020). Results show that older forests induce a light cooling effect (unclear if statistically significant), and younger forests tend to warm the surface because of lower evapotranspiration. These studies are based on numerical models, either fully coupled climate models (Luyssaert et al., 2018; Naudts et al., 2016) or the land surface model only (Kumkar et al., 2020), which are highly demanding in terms of time and costs. The availability of simplified yet reliable approaches that directly link forest attributes to local temperature can favour the understanding of the effects on local climate of alternative forest management practices. Machine learning methods have the potential for mining data in linear or non-linear systems (Doan and Kalita, 2015; Kolevatova et al., 2021; Sharma et al., 2013), and can offer simplified solutions to study climate-forestry interactions.

In this work, we integrate advanced forest maps with class- and stage-specific parameterization with remotely-sensed land surface temperature (LST) data to explore if different forest management stages can influence LST. We then apply a set of machine learning algorithms to estimate the effects of forest management scenarios on surface temperature. As the advanced map represents the current management-induced forest structure in Fennoscandia, a simplified statistical model is built to link changes in LST with changes in forest structure parameters (such as LAI and CH), under a set of defined variables (elevation, radiation, snow cover). This approach secures certain advantages over numerical models, as it is computationally more efficient (and cheaper), relatively easier to be used by non-experts, and open to flexible applications (e.g., able to investigate alternative forest management scenarios). It also directly links temperature changes with forest structure parameters, so avoiding the need to decompose the surface energy budget into different components to estimate the temperature response. The analysis is also performed at a horizontal resolution (about 5 km) that is higher than what is typically allowed by coupled numerical models, and it captures gradients in topography, forest structure and tree composition. The model is applied to three case studies: two idealized cases of temperature response to forests at either low or high development stages, and one case that quantifies the temperature response to historical forest harvest.

2. Methods

2.1. Land surface temperature and surface attributes

LST is the radiative skin temperature of the land, and it is a fundamental aspect of climate and biology, affecting near surface air temperature, organisms and ecosystems from local to global scales. Land surface cover (e.g., snow cover, vegetation cover, bare soil, etc.) and elevation are key determinants of local LST.

Daytime and night-time LST are retrieved from the MODIS MYD11C3 version 6 product (Wan et al., 2015), a monthly dataset available at a 0.05-degree resolution that retrieves data from clear-sky conditions over each 8-day period (Wan, 2014). The satellite passes over the region at approximately 13:30 (daytime LST) and 01:30 (night-time LST), which can be interpreted as the daily maximum and minimum temperatures. In order to compare potential differences under all-sky conditions, we also use a recently produced spatiotemporally continuous dataset that reconstructed daytime and night-time LST by integrating clear-sky retrievals from MODIS (MOD11C1 and MYD11C1) with all-sky conditions from the ERA5-Land climate reanalysis dataset (Yu et al., 2022). We calculated monthly all-sky LST using the daily average of the reconstructed LST data. We use a digital elevation model at 1 km resolution for the elevation topographic variable and geographical coordinates (Amatulli et al., 2018). Daytime and night-time net shortwave and longwave radiation at the surface are from the Global Land Data Assimilation Systems (GLDAS) (Rodell et al., 2004) and the snow cover data are from the MODIS/Terra monthly snow-cover product (Hall and Riggs, 2021). In order to reduce the effects of interannual variability, we averaged monthly mean LST data, snow cover, and radiation from 2013 to 2017 (5 years mean), as the land forest map is representative of 2015. The resulting monthly means are used in the regression model. LST retrievals for the grids where forests are less than 50% of the grid area are excluded from the analysis. Totally, 56 134 grids are used in the analysis. Supplementary Figure S1 shows a simplified flowchart with the main steps of the analysis (pre-processing the datasets, training the models and their applications). We separately apply the clear-sky and all-sky LST to train different models and obtain parameters for each regression model, which is then applied to either clear-sky or all-sky LST. Supplementary Table S1 summarizes the datasets used in our study. All datasets are remapped to a 0.05° resolution with bilinear interpolation.

2.2. Forest structure and composition

Forest structure parameters rely on an enhanced land cover dataset produced to represent the current forest status in Fennoscandia in 2015 (Majasalmi et al., 2018). This land cover data is integrated within the ESA CCI LC map (ESA, 2017), where forest classes are updated with the attributes from National Forest Inventory (NFI) data. This dataset distinguishes the three major tree species in the region (two coniferous species, pine and spruce, and one deciduous, birch), and for each of them it specifies different structural attributes indicating four different forest development stages. These stages range from Development Class 1 (DC1, a forest with predominance of young or poorly developed trees, which are typical in post-harvested sites) to DC4 (a forest with highly developed tree structure and canopy, index of low management intensity or well regenerated secondary forest) (see Supplementary Figure S2). For each tree species and DC, the dataset has attributes like total stem volume (V, in m^3), maximum growing season leaf area index (LAI_{max}, in m^2/m^2), tree crown length (CL, in m), and mean tree Lorey's height (H, in m). The resulting map of structural attributes of forests in Fennoscandia is shown in Supplementary Figure S3, and the individual parameters for forest type and development class in Supplementary Table S2.

For our analysis, we further improve this dataset by considering monthly variations in LAI (Table 1), as this attribute can considerably vary over the year (especially for birch, a deciduous specie). We used the given LAI_{max} to proportionally re-scale the monthly LAI variations for each class of forest and tree type using the look-up table available from the coupled community land model (Lawrence et al., 2019). This look-up table includes monthly LAI of needleleaf (spruce and pine) and broadleaf forest (birch). The resulting enhanced forest cover dataset with monthly LAI values is aggregated at a 0.05-degree resolution to match LST data. In the statistical model, the fraction of forest types and attributes per grid cell is considered by computing the weighted mean LAI, CL, H, and V for each tree species. We then calculate the monthly values of the different DC parameters by doing the average across the different tree species.

2.3. Multivariable regression model

Multivariable regression models are widely used to estimate the relationship between two or more explanatory variables and a response variable by fitting a linear model to observed data. Every value of the independent variable x is associated with a value of the dependant variable y. Formally, the model for multiple linear regression, given p explanatory variables $x_1, x_2, \ldots, x_p$ and n observations, is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_p x_{ip} + \varepsilon_i \; for \; i = 1, 2, \ldots n \quad (1)$$

Where, for each month, y_i is LST for a grid i, β_0 is the model intercept, $\beta_{1\ldots}\ \beta_p$ are the regression coefficients, and $x_{1\ldots p}$ is the full list of the variables, that is, both site-specific (latitude, altitude, net shortwave radiation, net longwave radiation, snow cover) and forest structure parameters (volume, canopy height, crown length, LAI). We note that variables of forest structure are not completely independent, as fully developed forests typically show a high canopy height and a large volume. Multicollinearity is known to affect the coefficients, but it does not influence the predictions and the goodness-of-fit statistics (Kutner et al., 2004). As our primary goal is to make LST predictions, we only assess model performances in reproducing observed LSTs and we don't individually explore the role of each independent variable (which might be influenced by the multicollinearity).

The population regression line for p explanatory variables is defined to be $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_p x_p$. This line describes how the mean response μ_y changes with the explanatory variables. The observed values for y vary about their means μ_y and are assumed to have the same standard deviation σ. The fitted values $b_0, b_1, \ldots, b_p$ estimate the parameters $\beta_0, \beta_1, \ldots, \beta_p$ of the population regression line. Generally, we can employ the regression method to obtain the fitted values $b_0, b_1, \ldots, b_p$, and then acquire the fitted population regression line.

2.4. Multiple-regressor system (MRS)

Several regression models are available to estimate the regression parameters (Fernandez-Delgado et al., 2019). In our analysis, we build a multiple-regressor system using seven regression models to capture the relationship between land surface temperature and forest structure attributes in Fennoscandia. The seven regression models, selected through an iterative process based on best fitting models, are bagEarth, bMachine, cubist, earth, extraTrees (ET), gbm, and svr (Table 2). They are all included amongst the best 20 regression models to predict

Table 2
Regression models used in this study.

Model	Regression family	Reference	Code repository
bagEarth	Bagging	Kuhn (2019)	http://topepo.github.io/caret/train-models-by-tag.html#bagging
bMachine	Bayesian models	Kapelner and Bleich (2016)	https://cran.r-project.org/web/packages/bartMachine/index.html
cubist	Regression rules	Quinlan (1992)	https://pypi.org/project/cubist/
earth	Additive models	Friedman (1991)	https://contrib.scikit-learn.org/py-earth/
extraTrees	Random forests	Geurts et al. (2006)	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesRegressor.html
gbm	Boosting	Greenwell et al. (2022)	https://cran.r-project.org/web/packages/gbm/index.html
svr	Support vector regression	Chang and Lin (2011)	https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html

Table 1
Re-scaled monthly LAI phenology from tree- and stage-specific LAI_{max}. DC = development class of the forest.

Species	DC	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Spruce	1	5.2	5.3	5.8	6.0	6.2	6.3	6.0	5.9	5.8	5.3	5.0	5.0
	2	8.3	8.5	9.3	9.7	9.9	10.1	9.7	9.5	9.3	8.5	8.1	8.1
	3	10.8	11.1	12.1	12.7	12.9	13.2	12.7	12.4	12.1	11.1	10.6	10.6
	4	13.0	13.3	14.5	15.2	15.5	15.8	15.2	14.9	14.5	13.3	12.6	12.6
Pine	1	3.8	3.9	4.2	4.4	4.5	4.6	4.4	4.3	4.2	3.9	3.7	3.7
	2	5.5	5.6	6.2	6.4	6.6	6.7	6.4	6.3	6.2	5.6	5.4	5.4
	3	7.7	7.9	8.6	9.0	9.2	9.4	9.0	8.8	8.6	7.9	7.5	7.5
	4	6.9	7.1	7.7	8.1	8.2	8.4	8.1	7.9	7.7	7.1	6.7	6.7
Birch	1	0.0	0.0	0.2	0.8	2.0	3.2	3.1	2.3	0.8	0.2	0.0	0.0
	2	0.0	0.0	0.4	1.4	3.5	5.5	5.3	4.0	1.4	0.4	0.0	0.0
	3	0.0	0.0	0.5	2.0	5.0	7.9	7.6	5.7	2.0	0.5	0.0	0.0
	4	0.0	0.0	0.7	2.6	6.6	10.3	9.9	7.5	2.6	0.7	0.0	0.0

small-to-large, easy-to-difficult datasets (Fernandez-Delgado et al., 2019). The seven regression models are from seven different regression families, as models from the same regression family have been excluded. We randomly used 80% of our grid cells to train the model, and then used the trained model to predict the entire domain. This procedure is repeated for both clear-sky and all-sky LST data.

The model bagEarth is a member of bagging regression family, which uses bagging wrapper for multivariate adaptive regression splines (MARS) via the earth function (Kuhn, 2019). bagEarth is a form of non-parametric regression analysis technique which automatically models non-linearities and interactions between features. We do not provide an estimate of the coefficients for Eq. (1), as solutions for each term are based on multiple splines.

Using the Bayesian additive regression tree, bMachine model consists of a sum of regression trees and a regularization process developed on the parameters of the tree set (Kapelner and Bleich, 2016). We use the default number of trees and tunable hyperparameters to train the bMachine model.

The model cubist learns a M5 rule-based model with corrections based on nearest neighbours in the training set (Quinlan, 1993), implemented by the Cubist package. A tree structure is created and translated to a collection of rules, which are pruned and combined, and each rule gives a regression model, applied to the patterns which accomplish that rule. The nearest neighbour based to correct the rule-based prediction is used in cubist model.

Earth belongs to the additive model regression family. This method is a hybrid of generalized additive model (GAM) and regression trees which uses an expansion of product spline functions to model non-linear data and interactions amongst inputs (Friedman, 1991). The spline number and parameters are automatically determined from the data using recursive partitioning, and distinguishing between additive contributions of each input and interactions amongst them. The functions are added iteratively to reduce maximally the residual, until its change is too small or a number of iterations is reached. Same as bagEarth, we do not provide an estimate of the coefficients for Eq. (1) in earth model.

Extremely Randomized Trees (extraTrees, ET) is an ensemble learning method which creates extra trees randomly in sub-samples of datasets to improve the predictivity of the model and control over-fitting (Geurts et al., 2006). It randomizes the input and cut-point of each split (or node in the tree), using a parameter that tunes the randomization strength. The full training set is used instead of a bootstrap replica.

The generalized boosting regression model (gbm) is the stochastic gradient boosting in the caret model list (Kuhn, 2019). The hyperparameters are the maximum depth of input interactions, with integer values from 1 to 5, and number of trees for prediction, with values from 50 to 250 with step 50. We use a Gaussian distribution and shrinkage=0.1 (default values).

Support vector regression (svr) is a supervised learning algorithm that uses the same principle as the support vector machines (Chang and Lin, 2011). It is used to predict discrete values. The basic idea behind svr is to find the best fit line which has the maximum number of points in a hyperplane.

2.5. Forest management scenarios and historical harvest data

The enhanced observationally-based dataset represents the current (i.e., 2015) structure and compositional state of Fennoscandian forests, and it is used to prescribe the present state of the forest, or control forest management scenario (CTRL). A key advantage of this dataset is its flexibility in terms of specification of tree species and development class of the forest, which allows to design management scenarios by creating new datasets with modified forest structure attributes. Two idealized forest management scenarios are defined to investigate effects of two opposite structural changes: undeveloped forests (UDF) and fully developed forests (FDF). In UDF, structural attributes of CTRL are changed to DC1 (the least developed class) for each tree species in all forested areas of the domain, so to mimic the effects of an idealized extensive management. In FDF, the structural attributes of CTRL are changed to DC4, the most developed forest class, to exemplify a forest state without management interventions. In addition to these extreme cases, we define a forest harvest scenario (HARV) to explore the effects on LST of historical forest management. In this case, the forest areas in Nordic countries where harvest occurred from 2015 to 2018 are changed from highly developed to low-developed classes. A new database of spatially explicit forest harvest data is used (Zhou et al., 2021). It was produced by integrating national roundwood production statistics from FAOSAT with high resolution maps of forest losses and gains (Hansen et al., 2013) attributed to forest management (Ceccherini et al., 2020). This dataset was then integrated with the European Space Agency Climate Change Initiative Land Cover (ESA CCI LC) maps (v.2.0.7) (ESA, 2017) to obtain temporally and spatially homogenous maps at 300 m resolution of annual (2003–2018) forest harvest areas in Fennoscandia (Zhou et al., 2021). This map is resampled at 0.05° resolution as the other datasets, preserving information on the fraction of forest harvested per tree species (see Supplementary Figure S4). In our study, this dataset is used to identify the grids where harvest occurred between 2015 and 2018. In these grids, the corresponding fraction of a grid affected by harvest is converted from highly developed classes (DC4 and DC3, in order of priority in the conversion) to DC1. The resulting dataset is averaged again for the different parameters in forest structure, so to get a new map with modified monthly mean values.

For all the case studies (UDF, FDF, and HARV), the modified forest maps are applied to Eq. (1) to calculate the resulting LST, and the difference with the reference state of the forest (e.g., HARV – CTRL) shows the local temperature impacts of different forest structure. In all management scenarios, the implemented changes in forest structure are constrained by observations and the approach taken allows to interpret the temperature effects as the result of alternative forest management strategies: widespread harvesting (UDF), reduced harvest and frequency of management interventions to enhance stand volume and tree density (FDF), and historical harvest (HARV).

3. Results

3.1. Forest structure distribution and its relationship to LST

More than 30% of the land area in Norway, 68% in Sweden and 73% in Finland is covered by forest. In total, the area covered by forests in Fennoscandia is more than 60 Mha (57% of the land area). The most abundant tree species is pine (which dominates in Sweden and Finland), followed by spruce (predominant in Norway), and birch (Fig. 1). The most common development class is Pine DC2, followed by Pine DC1 (Fig. 1a). This is the result of intensive forest management activities that historically took place in Fennoscandia. Spruce DC1 has the lowest presence in the domain, since Norway, the country where spruce dominates, has historically lower management interventions than Sweden and Finland (Iordan et al., 2018). Deciduous species (mostly birch) are more widespread and mostly connected to natural forest succession or early-stage tree encroachment. DC1 and DC2 are common in mountainous areas and at high latitudes, where low temperature is a limiting factor for tree growth.

Forest development stage is a key factor in shaping local surface temperature. We find that the annual mean clear-sky LST is generally increasing with the forest development class. Across an elevation gradient of an area experiencing similar climatic conditions (black box in Fig. 1b), the average annual mean LST of a pine-dominated forest tends to be warmer for DC4 (data points: $n = 303$) than DC2 ($n = 340$) and DC1 ($n = 161$), especially at low elevation (Fig. 1c). Values for DC3 are affected by a low number of points ($n = 26$). The trends are confirmed when projecting all temperature values for the different tree species and development classes in the domain (Supplementary Figure S5). For all tree species, and for nearly all elevation ranges,

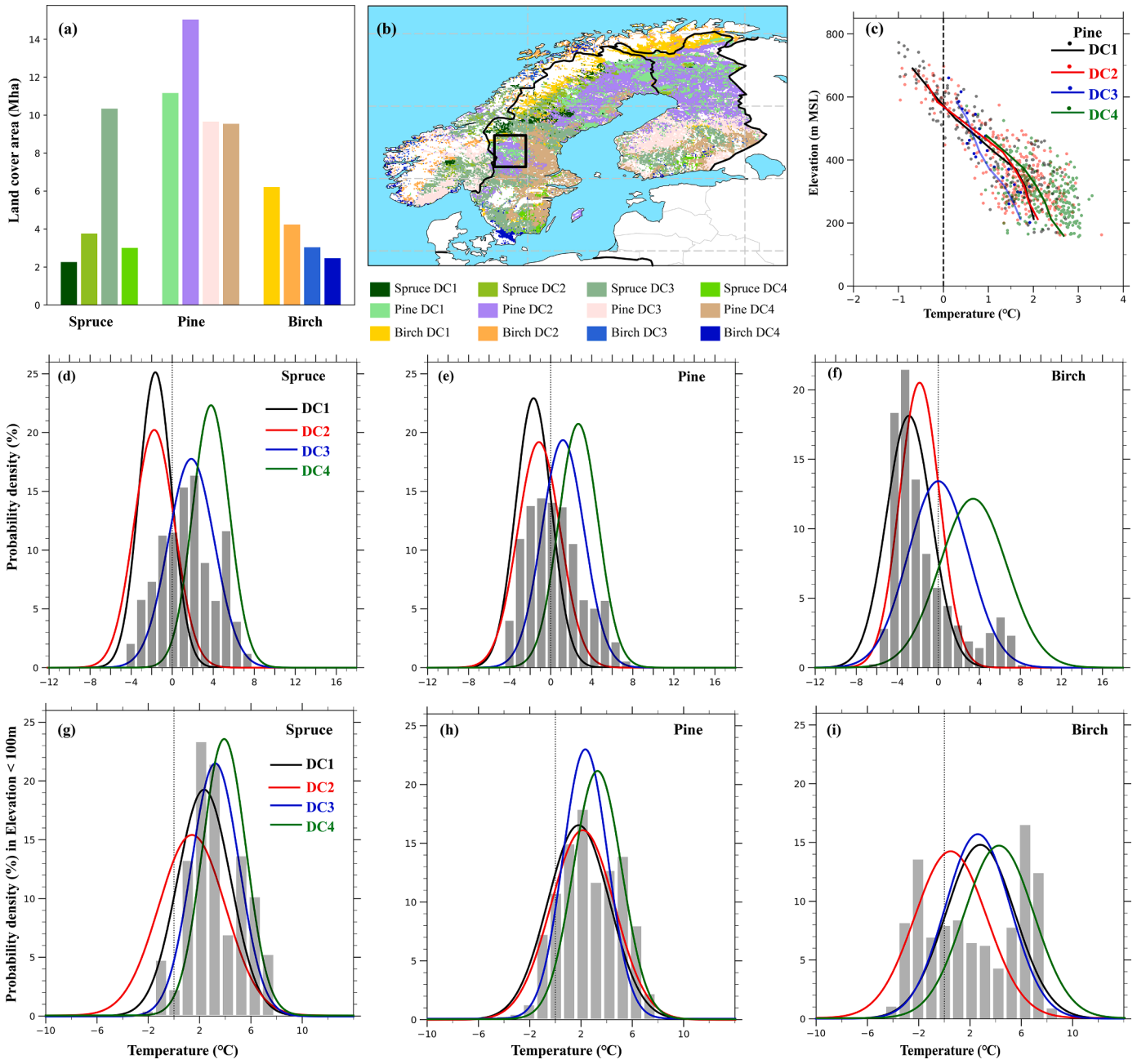


Fig. 1. Forest distribution in Fennoscandia, its breakdown into development classes (DC), and associated annual mean clear-sky LST. (a) Land cover area (in Mha) of four development classes for spruce, pine, and birch. (b) Dominant forest class distribution in Fennoscandia at 0.05° resolution (aggregated for visualization purposes only; weighted averages of each DC and tree per grid cells dominated by forests are used in the analysis). (c) Annual mean LST over an elevation profile from the area within the black box in (b). Probability density distribution (in%) of annual average land surface temperature (clear-sky) for four development classes of spruce (d), pine (e), and birch (f) for all the domain and for forest areas located at an elevation lower than 100 m (g, h, i). See Figure S6 for a similar analysis for all-sky LST. Colour lines show different forest classes and grey bars represent the probability density distribution of all the grids. LST values are multi-annual averages (2013–2017). .

higher LST values are found for higher development classes.

Similar results emerge when assessing the probability distribution of LST across the different DCs for each of the tree species for all the domain (Fig. 1d-f) or for forest areas located at an elevation lower than 100 m (Fig. 1g-i). On average, more developed forests (DC3 and DC4) are associated with higher annual mean LST than poorly structured forests (DC1 and DC2). The same trend is observed for all-sky LST data, whose probability distributions of LST values across DCs and tree species only show small differences relative to those under clear-sky conditions (Supplementary Figure S6). Although not specifically linked to forest development stages, a warming trend associated with forest cover at high latitudes relative to open land is typically found in other observational studies as well (Alkama and Cescatti, 2016; Li et al., 2015; Perugini et al., 2017). Fig. 1 further shows the role potentially played by forest management and development stage of the forest, thereby suggesting that a multivariable regression model using forest structure parameters and local surface geographic information as predictor variables can infer the variability in LST. The resulting parameters of regression models can be used to predict the effects of the management-induced changes in forest structure on LST.

3.2. Regression model performance

The average of the predictions from the seven individual machine learning algorithms in the multiple-regressor system (MRS) show a good performance in reproducing observed LST (Fig. 2). When compared to observations, both daytime (Fig. 2a) and night-time (Fig. 2c) LSTs show nearly symmetrical bias scattered all over the domain without any relevant cluster, with some exceptions (such as a cold bias in southern Norway and warm bias in central eastern Finland). The 5th and 95th percentiles of the differences of estimated LST vs. observations range from −0.75 to 0.78 °C for daytime temperature, and between −1.02 and 0.83 °C for night-time temperature.

For the seasonal and annual mean daytime and night-time LST, the temperature differences between simulated and observed LST shows a gaussian distribution with a near zero mean. The R-squared (R^2) of the predictions from each model of land surface temperature vs. observed LST per grid cell mostly fall between 0.81 and 0.98 for all the seasons, except for summer daytime, where the R^2 is smaller (0.51 - 0.81) (Fig. 2b). On annual average, the R^2 for daytime temperature from all the models is high, ranging from 0.95 (svr) to 0.99 (cubist), resulting in a multi-model mean of 0.97. Values are relatively lower for night-time, as some models have $R^2 = 0.91$ (bMachine and svr), with a multi-model mean of 0.95. In daytime summer, the R^2 of model predictions ranges from 0.51 (bMachine and svr) to 0.81 (cubist), with multi-model mean of 0.71.

A similar trend can be appreciated by the root-mean-square error (RMSE) of the predicted LSTs from the individual models (Fig. 2d). The average RMSE of annual mean daytime and night-time LST is 0.49 °C and 0.58 °C, respectively. Nearly all regressors in the MRS produce constrained RMSE to less than 0.8 °C for annual mean daytime and night-time LST. On the seasonal scale, the range of multi-model mean RMSE is from 0.46 (autumn) to 0.90 °C (winter) for daytime LST, and from 0.57 (spring) to 1.04 °C (winter) for night-time LST. Winter thus has the highest seasonal values of RMSE, and might be due to the presence of snow, which can largely affect the surface energy budget. In general, svr and bMachine shows the highest RMSEs, and cubist and ET the lowest.

The average annual variance of the simulated daily LST from the MRS ranges from 0.10 (cubist) to 0.42 (svr), but in some cases it can reach above two (daytime LST for winter from bMachine and svr method). In general, all machine learning algorithms show a large variance in daytime LST in winter, while it is usually smaller than 0.85 for the other seasons.

The full range of performance statistics (R^2, RMSE, variance, and bias) of the individual models is available in Supplementary Table S3 for daytime, night-time, and daily averaged temperature. Generally, svr and bMachine show poorer performances than the other regression models, as their simulated LSTs always have the highest RMSE, lowest R^2, largest variance, and highest model bias. On the other hand, cubist has the best statistical scores: the seasonal mean RMSE is from 0.33 to 0.66 °C for daytime LST and from 0.40 to 0.58 °C for night-time LST. Model bias are relatively low for all simulations, as the regression models are based on the least-square-error. The estimated coefficients for the monthly-mean LST from the different machine learning algorithms are available in Supplementary Tables S4-S7, and are applied to study the effects of alternative forest structure variables on LST.

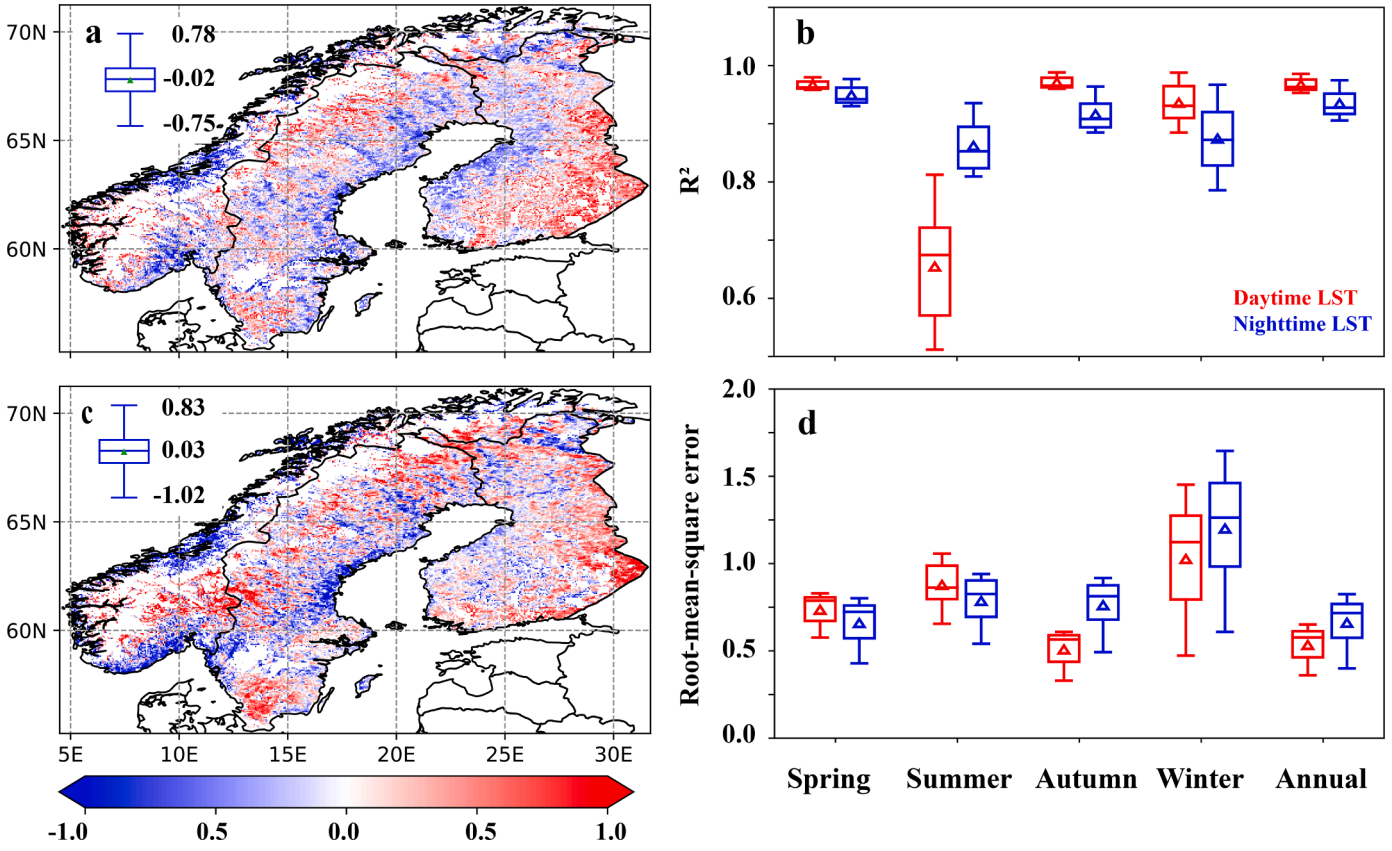


Fig. 2. Performance of the multiple regressor system (MRS) in reproducing observed clear-sky LST. Difference in the annual mean LST (unit: °C) between multi-model means of the MRS and MODIS products for daytime (a) and night-time LST (c). The blue range in the top left corner refers to the 5th and 95th percentile of the spatial variability in the map (the box indicates the standard deviation, the line in the box is the median, and the dot the average). The R-squared (R^2) of the seasonal mean LST between each model in the MRS and MODIS products for daytime and night-time LST (b). The root-mean-square error of the differences in seasonal mean LST (unit: °C) between each model in the MRS and MODIS products for daytime LST and night-time LST (d). Triangles indicate the average R^2 and root-mean-square errors of seasonal or annual mean LST from the 7 models, the range is the maximum and minimum, and the line is the median. .

Overall, the MRS captures the mean status and spatial distribution of LST. It shows a relatively low level of bias and RMSE, especially when compared to other modelling tools. Its statistical performance can be compared to that of regional climate models (RCM), which are widely used in simulating local-to-regional climate (Kotlarski et al., 2014) and for exploring land-climate interactions (Cherubini et al., 2018; Davin et al., 2020; Hu et al., 2019; Huang et al., 2020). In most RCMs simulations, the regional average annual mean temperature differences in northern Europe goes from −1.98 to −0.02 °C (Kotlarski et al., 2014), but in some grids it can reach up to −2.90 °C. The RCM error ranges of annual mean and seasonal mean temperature are thus remarkably larger than those from our MRS, despite the RCM being much more demanding in terms of resources and computational power/time.

3.3. Effects of idealized forest management on LST

The MRS is first applied to evaluate the LST changes from two idealized forest management scenarios, FDF and UDF (Fig. 3). If all forests in Fennoscandia would be at a highly developed stage (DC4), representative of no or limited management interventions, the effects on daily land surface temperature are an annual average increased warming of 0.26 °C (0.03–0.69 °C as 5th-95th percentile) (Fig. 3a). This warming is widespread across the domain, and it has higher intensity at high latitudes and at higher elevation, when seasonal snow cover is more pronounced. This suggests that changes in surface albedo can be a main driver of the annual mean changes in LST. However, the temperature changes vary dramatically between seasons and between day and night (Fig. 3b). Night-time temperatures are positive all year round, meaning that a fully developed forest causes warmer nights. During the daytime, temperature changes are positive (i.e., warming) in winter and negative (i.e., cooling) in summer. In particular, the average summer cooling is estimated by the different models to range from −0.23 (ET) to −0.85 °C (earth). The statistics for the FDF scenarios produced from the application of the individual models are shown in Supplementary Table S8.

Similar differences in the seasonal and diurnal LST response are found in other observational studies assessing the local temperature of nearby forested vs. non-forested sites (Alkama and Cescatti, 2016; Peng et al., 2014; Tang et al., 2018). Overall, the climate impacts are sensitive to the incoming radiation and, as a consequence, the largest response occurs during the summer solstice. In line with other studies (Alkama and Cescatti, 2016; Naudts et al., 2016; Zhang and Liang, 2018), the summer cooling of forests is the result of increased evapotranspiration and surface roughness under higher radiation load. Forest development generally reduces surface albedo owing to the snow masking effect in winter and to the darker colour of canopies than short vegetation in summer, thereby increasing both net radiation at the surface and latent

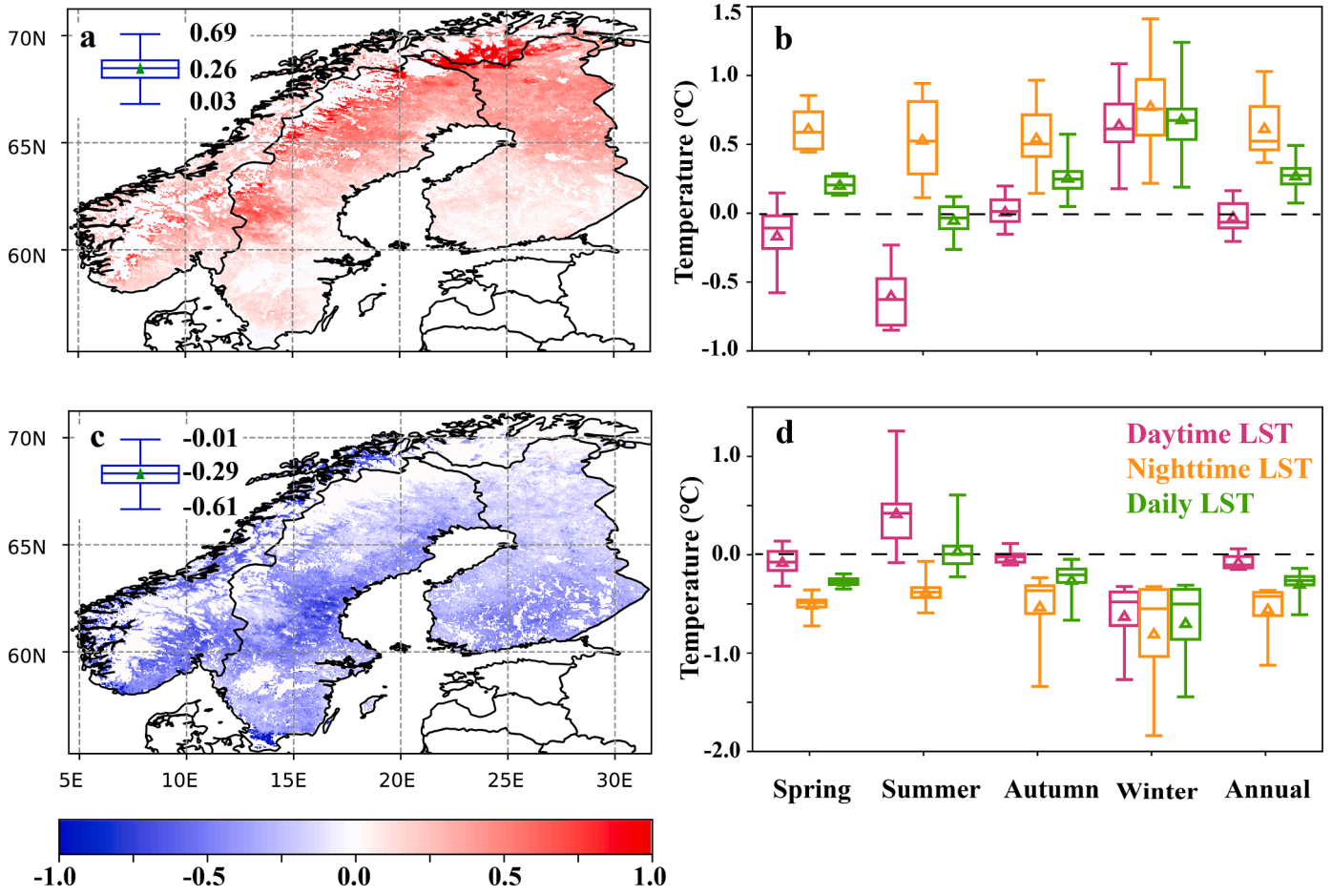


Fig. 3. Land surface temperature changes (clear-sky) induced from idealized forest management scenarios. The same results under all-sky conditions are shown in Figure S7. Annual mean daily temperature differences (unit: °C) between a scenario with a full development class (DC4) of all forested areas and present-day forest structure (FDF – CTRL) (a), and between all un-developed forest (DC1) and present-day forest structure (UDF - CTRL) (c). Both (a) and (c) show the multi-model mean LST from the regression coefficients of the individual machine learning models. Boxplots in the top-left corner in (a) and (c) show the spatial variability across the forested areas (the range indicates the 5th and 95th percentile, the box the standard deviation, the line in the box the median, and the dot the average). The seasonality of the LST changes of FDF (b) and UDF (d) relative to CTRL is shown for daytime, night-time, and daily LST. The boxplots are based on the average regional LST change for each individual model used (the range indicates the minimum and maximum values, the box the standard deviation, the triangle the average and the line the median). .

and sensible heat fluxes. The vegetation cover largely determines how this extra energy is partitioned, and trees typically increase latent heat flux instead of sensible heat, thereby inducing cooling effects. This leads to an increase of the diurnal variation, that is, the difference between the daily maximum (daytime LST) and minimum (night-time LST) temperature. The net effects of periodically changed albedo, surface roughness and ET under varying radiation loads shape the seasonal changes in the response of the local climate to forest dynamics. These aspects are discussed in more details in the discussion section.

A predominant cooling of −0.29 °C (−0.61/−0.01 °C as 5th/95th percentile) is found in the UDF scenario (Fig. 3c). A more limited forest structure in Fennoscandia is thus associated with an annual mean reduction in land surface temperature. As in the FDF scenario, there is a strong seasonality in the induced temperature effect. Depending on the machine-learning method used, daytime LST in summer can increase up to about 1.2 °C (with bagEarth) or be almost unchanged (weak reduction with cubist). Average reductions in night-time temperatures are around −0.4 °C, ranging from about 0 (bagEarth) to 0.6 °C (earth). In winter, changes in both day- and night-time temperatures are negative (i.e., cooling), and in one model (earth) they can be up to −2 °C. The statistics for the UDF scenarios produced from the application of the individual models are shown in Supplementary Table S9.

The results for both FDF and UDF are highly consistent to those obtained when the all-sky LST data (and corresponding MRS) are used

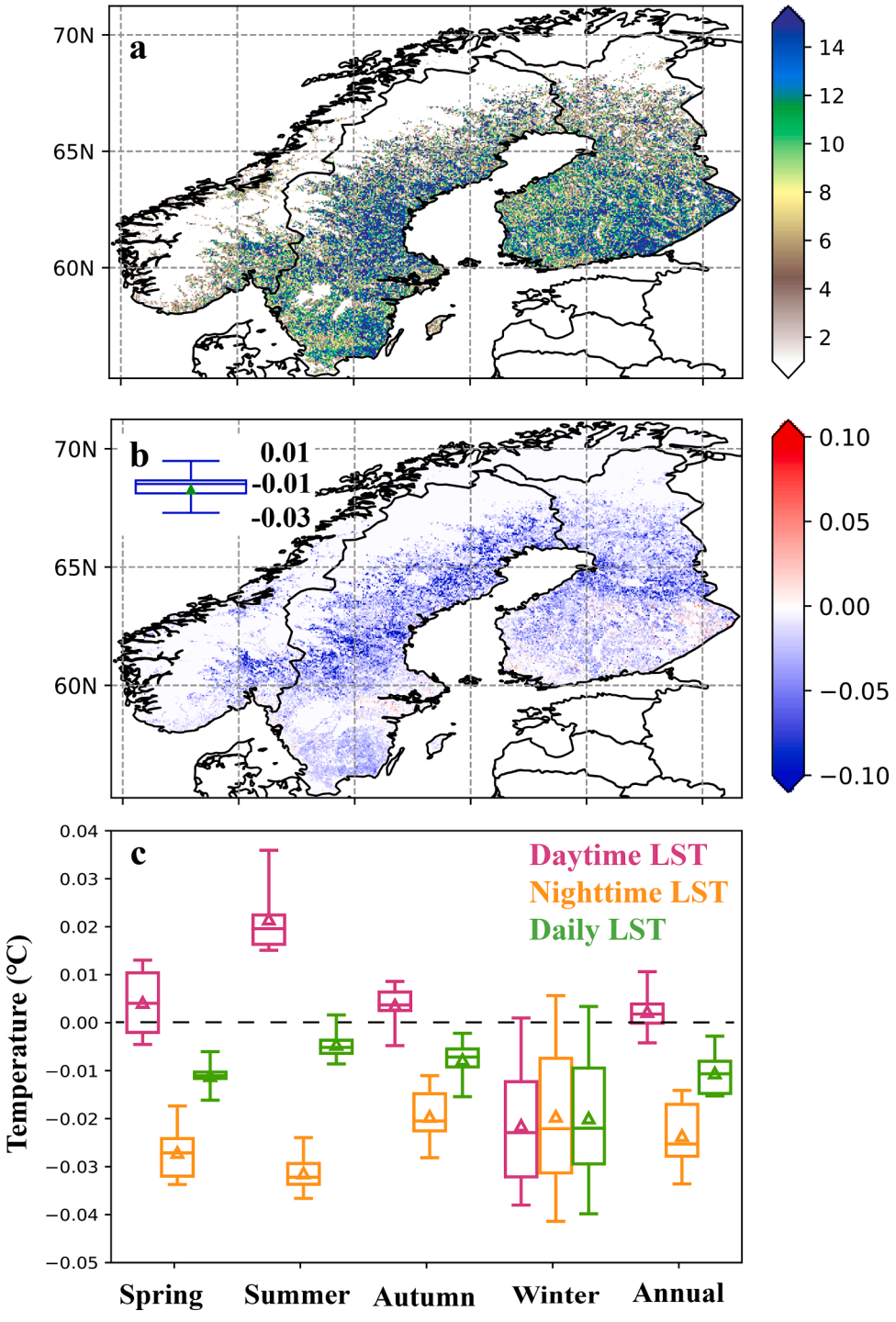


Fig. 4. Total harvested areas from 2015 to 2018 in Fennoscandia and effects on land surface temperature (clear-sky). Volume of forest harvest in m^3 per ha (a). Annual mean land surface temperature changes caused (HARV – CTRL) by forest harvest in °C (b). Boxplots of regional average seasonal and annual mean temperature changes (unit: °C) in the harvested grids caused by forest harvest from 2015 to 2018 (c). The boxplots show the average regional LST change for each individual model in the MRS (the range indicates the minimum and maximum values, the box is the standard deviation, the triangle is the average and the line is the median).

(Supplementary Figure S7). The same seasonality and magnitude of the temperature changes is observed, with the main difference being the spring night-time response in UDF (for which one model estimates an all-sky LST reduction of about −1.5 °C, against a range of values between −0.4 °C and −0.75 °C from the other models). In general, these findings are in line with the outcomes of observational studies performed in boreal forests (Alkama and Cescatti, 2016; Lee et al., 2011; Zhang et al., 2014), which show that forest clearing increases the daily temperature during summer months, while afforestation reduces it. They also agree that changes in temperature are smaller in the other seasons. Similar trends in diurnal LST asymmetries are found in other observational studies of both afforestation and deforestation. In China, forest plantations cooled daytime temperatures and warmed nights (Ma et al., 2017; Peng et al., 2014), and the same is found for forests relative to nearby open lands in Europe (Tang et al., 2018). In particular, the latter found a mean annual daytime cooling effect of forests of −1.06 ± 0.03 °C, a mean annual night-time warming of 0.58 ± 0.01 °C, and a mean annual daily cooling of −0.24 ± 0.01 °C. This is highly consistent with our estimates, which are lower in magnitude due to the different vegetation changes considered. Our scenarios investigate effects of forest management caused by smaller differences in forest structure, e.g., from a mixed developed forest to a fully (FDF) or under-developed (UDF) forest. The other studies compare temperature differences between forests and open land, either between nearby pair sites or after afforestation/deforestation, for which changes in vegetation structure are more pronounced. In summary, letting boreal forests to develop to more mature stages increase annual mean surface temperatures, although it brings cooling benefits in the summer, with a potentially important contribution in mitigating effects of summer heat. An opposite effect is found when forest structures are simplified.

3.4. Effects of historical forest management on LST

From 2015 to 2018, about 252 million m^3 of roundwood was harvested in Finland, 49 million m^3 in Norway and 296 million m^3 in Sweden (see Supplementary Table S10 and Figure S4). Forest harvest has mostly taken place in south-east Norway, central Sweden and southern Finland (Fig. 4a). The harvest reduces the local total volume of trees and its overall structural properties (i.e., LAI, CL, and H). The effects of historical harvest on LST (HARV scenario) were simulated by applying the MRS to a newly created forest map where, in the areas where harvest occurred, high development classes (DC3 and DC4) are converted to the lowest development class (DC1).

Results show a largely homogeneous multi-model annual mean cooling of −0.01 °C associated with the historical harvest, with confidence intervals (5th and 95th percentiles) ranging from −0.03 °C to +0.01 °C (Fig. 4b and Table S11). A few warming signals are detected in the southern coastal areas. The seasonality shows that the day- and night-time LST response have opposite sign but similar magnitude in summer, while in winter they are both negative (e.g., cooling), except for one model (earth) (Fig. 4c). In particular, a cooling effect at night is observed throughout the year, while daytime temperature increases in all seasons but winter. These results are highly consistent with those observed in the UDF scenario, as they are associated with a (less extreme) simplification of the forest structure.

4. Discussion

In this study, we use a multi-model regression system and satellite observations to investigate the effects of forest management on local surface temperatures. The multi-model mean has a reasonable accuracy in reproducing observed LST and models show a large consistency in estimating the temperature response to different forest management scenarios. We find that local biophysical processes triggered by enhanced forest structures can effectively mitigate summer temperatures. On the other hand, a simplification of the forest structure reduces absorbed radiation, especially during winter months in areas affected by seasonal snow cover owing to the well-known snow-albedo effect (Anderson et al., 2011; Betts et al., 2007), which can be dominating throughout the year and results in annual net cooling (but with summer daytime warming).

The predicted effects of forest dynamics from our multi-model statistical approach are consistent with other studies that investigated the relationship between forest cover and LST. A study that quantified the seasonal effects of forests on LST in Europe found that, relative to nearby open land, forests cool daytime clear-sky LSTs during the warm season, but the opposite occurs in winter, and that the forest night-time warming effect occurs year-round (Tang et al., 2018). Our analysis of the FDF scenario achieved the same conclusions, although numerical values are smaller because in our case we are assessing effects from marginal increases in forest development, rather than an alternative land cover. Another study located in China reached similar conclusions, i.e., large-scale afforestation decreased daytime LST and increased night-time LST, especially during summer (Peng et al., 2014).

The comparison with previous numerical models is challenging because of the large diversity in climate system response from different climate models. For example, there is no agreement across models on the sign of temperature changes in summer from afforestation in Europe, as some regional climate models predict a widespread cooling (well below −2 °C), others a widespread warming (around +2 °C or above) or even a mixed response (Davin et al., 2020). More specifically to the Scandinavian domain, a study based on an individual climate model found that an expansion of forests leads to additional warming of surface temperatures in winter and spring (between 1.0 °C and 1.5 °C) and cooling in summer (between −1.6 °C and −1.3 °C) (Mooney et al., 2021). Despite it affects a limited number of grid cells, this response is relatively stronger than what we found in our forest management scenarios, but the seasonal trend is similar. More specifically on the sensitivity of LST to forest structure, an off-line land surface model found that more developed forests have a light annual cooling of 0.04 °C, and undeveloped forests an annual mean warming of 0.14 °C (Kumkar et al., 2020). In general, modelling studies provide amplified responses relative to observationally-constrained estimates (Perugini et al., 2017). Differences can be connected to the accuracy in the representation of land cover classes and physical processes, as modelling studies do not usually distinguish amongst forest stages and directly assess larger scale changes between two alternative land covers with default parameterization schemes.

The observed diurnal asymmetry in day- and night-time temperature is due to different energy balance processes (Hain and Anderson, 2017; Peng et al., 2014). Daytime temperatures are shaped by incoming solar radiation, land surface properties (e.g., albedo and emissivity), latent and sensible heat fluxes, and near-surface atmospheric boundary layer conditions. The amount of absorbed radiation is determined by surface albedo, and this energy is then partitioned into latent and sensible heat fluxes depending on vegetation cover and soil moisture (Mu et al., 2011). The higher LAI of highly developed forests makes them more efficient in dissipating energy as latent heat into the atmospheric boundary layer through turbulent diffusion (Rotenberg and Yakir, 2010), so resulting in larger daytime cooling than less structured forests (which dissipate a higher fraction of energy as sensible heat). In the night, the evapotranspiration from vegetation is negligible, and the LST is mainly influenced by energy stored during the day and the status of the near-surface atmospheric boundary layer. Forests thus tend to be warmer at night as they get developed, because taller vegetation enhances turbulence and draws heat from the air towards the surface (Lee et al., 2011; Tang et al., 2018). There are two other relevant factors reported in the scientific literature as possible drivers of higher temperature at night of forests. The typical increased soil moisture of forests tends to increase the surface heat capacity, and thus the daytime heat storage and night-time heating (dos Santos et al., 2021; Schultz et al., 2017). Further, the higher evapotranspiration from mature forests

increases air humidity and boundary layer clouds, thereby increasing the downward longwave radiation from the atmosphere and decreasing the one upward (longwave) from the surface, which increases surface temperatures at night (Li et al., 2016; Peng et al., 2014). This longwave radiative imbalance has a stronger effect during night-time, when the boundary layer is thinner and more stable (Dai et al., 1999; Peng et al., 2014; Zhou et al., 2007). From summer to winter, the importance of these effects and that of evapotranspiration fluxes declines, while the albedo effect becomes dominating (especially in presence of snow), so that more developed forests result in a net warming effect in winter (at both day and night).

Our analysis focuses on land surface temperature, which may differ from air temperature. Changes of LST induced by forest cover dynamics can be up to 50% larger than that of air temperature (especially in the maximum values, although differences are smaller in mean values), but the patterns and directions of a change are the same (Alkama and Cescatti, 2016). This numerical difference is probably driven by satellite retrievals, as estimates of LST occur under clear sky conditions. Our analysis compared the findings under both clear-sky and all-sky conditions, showing that results are highly consistent. The presence of clouds decreases the incoming shortwave radiation from the sun and increases the downward longwave radiation during the night, potentially causing variability between cloud-covered LST and a cloud-free LST. Relative to all-sky conditions, LST values under clear sky conditions are usually higher, especially in the summer (Gallo and Krishnan, 2022). As no significant differences are found in our analysis, we can argue that vegetation dynamics induce proportional changes in LST that are the same under clear-sky and all-sky conditions. However, limitations can lie on the dataset used for all-sky LST, which relies on interpolating empirical orthogonal functions of climate reanalysis data for cloud-contaminated LST reconstruction. Given the lower resolution of the reanalysis data, the capabilities of this new dataset to robustly detect all-sky LST variability induced by forest development stages that occur at a much finer scale should be further explored, and as such its feasibility to estimate effects of forest management. In addition, our approach estimates local impacts on temperature resulting from small-scale variations in forest structure, without including possible large scale feedbacks due to land-atmosphere interactions and teleconnections (Portmann et al., 2022).

Our analysis considers average climatic conditions to better identify the effects of forest management on LST and avoid biases from inter-annual climate variability. However, the biophysical processes at play may be sensitive to variable background climate from year to year (Li et al., 2016; Pitman et al., 2011). For example, in the case of a warmer year, the amount of snow cover duration and snow depth is reduced, and thus the importance of the albedo mechanism. A better representation of snow phenology in the MRS can increase the accuracy of the predictions in winter and early spring, because snow phenology is highly sensitive to temperature (Ma et al., 2020; Peng et al., 2013), with potential repercussions on the surface energy balance. For example, earlier snow cover termination is typically correlated on a year-to-year basis with a positive temperature anomaly during the snowmelt month (Peng et al., 2013). More complex snow-related metrics beyond snow cover could be considered, such as daily (diurnal) freeze-thaw dynamics, mean melt onset date, snow cover depletion date, and snowmelt duration in spring. However, availability of high-quality data is typically a constraint for Nordic regions, where there is a sparse regional network of weather stations relative to the diversity of the landscape. Further, the long season of polar darkness, which largely overlap with the period of snow cover, and persistent cloud cover in the shoulder seasons, strongly reduce monitoring possibilities from satellite optical sensors. Reanalysis datasets of snow phenology exist, but they usually have coarse resolutions and poor representation of surface properties (especially of different forest development stages), making their suitability for applications relying on high horizontal resolution and high diversity of forest structures limited. Adding more variables to the MRS would also make it more complex and data intensive, and optimal solutions in terms of marginal additions and relative gains in prediction estimates should be pursued. Similarly, the number of models required to produce robust estimates could be optimized. We tested the influence of the number of selected models by repeating the analysis with five models instead of seven (reiteratively excluding two models from the regression system). The results are similar and did not show changes in trends. Our analysis retained one model per regression family to show their performances, but a reduction in the number of regression models would not affect our conclusions.

The evapotranspiration effect is highly sensitive to soil moisture, and periods of droughts can influence it. Modelling studies generally show that the cooling effect of forest losses (which we can interpret as a proxy of forest harvest) decreases with a warmer background climate, although with a latitudinal gradient (lower decreases at high latitudes) (Armstrong et al., 2016; Li et al., 2016; Pitman et al., 2011; Winckler et al., 2017). For the Fennoscandian region, the risks of warmer winters under progressing climate change can become a key factor in reducing the snow-albedo cooling effects of harvested sites, and thereby reduce the biophysical temperature benefits induced by forest harvest. At the same time, summer droughts can induce soil moisture benefits and reduce the latent heat fluxes, and thus the cooling effects, of developed forests. More specific model simulations combined with ground observations are necessary to robustly estimate the resulting effects on annual mean temperature of these two contrasting effects caused by climate change.

5. Conclusions

This study offers a simplified statistical approach to assess the effects of forest management on land surface temperature in Fennoscandia. We found a consistent pattern in the domain of higher mean surface temperature in presence of more developed forests relative to poorly structured forests, indicating that changes induced by forest management contribute to shape the local climate. The multi-model mean estimates from the regression system made of seven machine-learning models can reproduce the temperature values observed in the forest at reasonable accuracy, and thanks to their integration with high resolution maps of forest attributes can assess the influence of forest management on local temperature.

Compared to other numerical models, our approach is much cheaper and easier to use (a desktop computer is sufficient). It is also more flexible to be applied at different regional scales and locations, provided the number of grids is large enough to allow a robust estimate of the regression coefficients. The regression coefficients from our study are made available in the supplementary materials, and they can be used by non-experts to evaluate LST response to different forest management scenarios, either for the entire domain or for a subset of it, by using as inputs previous and post-harvest forest structure parameters. Finer scales different than the one used in this study can be explored, and they can deliver more specific predictions thanks to reduced geographical and climatic variability. An optimal solution in terms of size of the domain and accuracy of multi-model estimates of observed LST is to be identified. Possible extensions of this work include the possibility to ingest more land surface information such as aspect of the terrain, soil moisture, snow phenology, or forest root parameters, and capture the local LST response to extreme events such as drought periods or heat waves. At the same time, there should be a preference to limit the number of variables used in the model for simplification and reduce the volume of data needed to run the statistical model. Future work could explore an optimal solution for the minimum number of input variables that are needed to arrive at robust predictions. This optimal would probably differ per geographical region, meaning that the best combination of variables can be a compromise between available data and accuracy of the estimates.

Availability of simple parametric models for predicting surface

temperature changes from forest management can enhance the inclusion of biophysical effects in climate impact analysis. Currently, our models are tested and validated for Fennoscandian forests and cannot be directly used in other studies with different climate and species composition. Although future work should specifically test its transferability, the same approach is expected to perform in other forest areas. As LST data are available from all the globe, a major limitation can be the availability of tree- and stage-specific maps of forest attributes, which are essential to model the effects of management-induced changes in forest structure. In such cases, a statistical approach can be unpracticable, and calibration of physical/numerical modelling systems can be the only option to gain insights on forest-climate interactions. Overall, this work is a step forward to facilitate the consideration of the effects of forest management on land surface temperature and contributes to develop strategies that bridge the gap between forestry and climate sciences.

Author Contributions

B.H., Y.Li, and F.C. designed this study. B.H., Y. Li., Y. Liu and X.H. created and developed the statistical model and applied the machine learning approach. B.H. made the figures. B.H., Y.Li, X.H., W.Z. and F.C. analysed and interpreted the results. All authors wrote the paper.

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.

Data availability

Data will be made available on request.

Acknowledgements

B.H., X.H., and F.C. acknowledge the support of the Norwegian Research Council (project no. 286773 and 294534). Y. Li acknowledges the support from the China Scholarship Council (CSC). W.Z. acknowledges support from the National Natural Science Foundation of China (project no. 41861134038) and State Key Laboratory of Earth Surface Processes and Resource Ecology (2022-ZD-08). Simulations were performed on the resources provided by UNINETT Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2023.109362.

References

Alkama, R., Cescatti, A., 2016. Biophysical climate impacts of recent changes in global forest cover. Science 351 (6273), 600–604.
Amatulli, G., et al., 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Sci. Data 5, 180040.
Anderson, R.G., et al., 2011. Biophysical considerations in forestry for climate protection. Front. Ecol. Environ. 9 (3), 174–182.
Armstrong, E., Valdes, P., House, J., Singarayer, J., 2016. The role of CO2 and dynamic vegetation on the impact of temperate land-use change in the HadCM3 coupled climate model. Earth Interact. 20 (10), 1–20.
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. 25th Conf. Agric. For. Meteorol., 12th Jt. Conf. Appl. Air Pollut. Meteorol. A&WMA Fourth Symp. Urban Environ. 142 (2–4), 216–233.
Bonan, G.B., 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science 320 (5882), 1444–1449.
Ceccherini, G., et al., 2020. Abrupt increase in harvested forest area over Europe after 2015. Nature 583 (7814), 72–77.
Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intel. Syst. Tech. 2 (3).
Cherubini, F., Huang, B., Hu, X.P., Tolle, M.H., Stromman, A.H., 2018. Quantifying the climate response to extreme land cover changes in Europe with a regional model. Environ. Res. Lett. 13 (7), 074002.
Crowther, T.W., et al., 2015. Mapping tree density at a global scale. Nature 525 (7568), 201–205.
Dai, A., Trenberth, K.E., Karl, T.R., 1999. Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range. J. Climate 12 (8), 2451–2473.
Davin, E.L., et al., 2020. Biogeophysical impacts of forestation in Europe: first results from the LUCAS (Land Use and Climate Across Scales) regional climate model intercomparison. Earth Syst. Dynam. 11 (1), 183–200.
de Noblet-Ducoudré, N., et al., 2012. Determining robust impacts of land-use-induced land cover changes on surface climate over North America and Eurasia: results from the first set of LUCID experiments. J. Climate 25 (9), 3261–3281.
Doan, T., Kalita, J., 2015. Selecting Machine learning algorithms using regression models. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 1498–1505.
dos Santos, T., Keppel-Aleks, G., De Roo, R., Steiner, A.L., 2021. Can land surface models capture the observed soil moisture control of water and carbon fluxes in temperate-to-boreal forests? J. Geophys. Res. 126 (4).
ESA, 2017. Land Cover CCI Product User Guide Version 2. https://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf. Accessed 27 May 2022.
Fernandez-Delgado, M., et al., 2019. An extensive experimental survey of regression methods. Neural Netw. 111, 11–34.
Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19 (1), 1–67.
Gallo, K., Krishnan, P., 2022. Evaluation of the bias in the use of clear-sky compared with all-sky observations of monthly and annual daytime land surface temperature. J. Appl. Meteorol. Clim. 61 (10), 1485–1495.
Ge, J., et al., 2019. The nonradiative effect dominates local surface temperature change caused by afforestation in China. J. Climate 32 (14), 4445–4471.
Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Int. J. Mach. Learn. Cybern. 63 (1), 3–42.
Greenwell, B., Boehmke, B., Cunningham, J., 2022. Generalized Boosted Regression Models. URL. https://cran.r-project.org/package=gbm.
Hain, C.R., Anderson, M.C., 2017. Estimating morning change in land surface temperature from MODIS day/night observations: applications for surface energy balance modeling. Geophys. Res. Lett. 44 (19), 9723–9733.
Hall, D.K., Riggs, G.A., 2021. MODIS/Terra Snow Cover Monthly L3 Global 0.05Deg CMG, Version 61 [Data Set]. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA. https://doi.org/10.5067/MODIS/MOD10CM.061. Accessed 27 May 2022.
Hansen, M.C., et al., 2013. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853.
Hu, X.P., Huang, B., Cherubini, F., 2019. Impacts of idealized land cover changes on climate extremes in Europe. Ecol. Indic. 104, 626–635.
Huang, B., et al., 2020. Predominant regional biophysical cooling from recent land cover changes in Europe. Nat. Commun. 11 (1), 1066.
Iordan, C.M., Hu, X., Arvesen, A., Kauppi, P., Cherubini, F., 2018. Contribution of forest wood products to negative emissions: historical comparative analysis from 1960 to 2015 in Norway, Sweden and Finland. Carbon Balance Manage. 13 (1), 12.
Jackson, R.B., et al., 2008. Protecting climate with forests. Environ. Res. Lett. 3 (4).
Kapelner, A., Bleich, J., 2016. bartMachine: machine learning with bayesian additive regression trees. J. Stat. Softw. 70 (4), 1–40.
Kellomaki, S., Vaisanen, H., Kirschbaum, M.U.F., Kirsikka-Aho, S., Peltola, H., 2021. Effects of different management options of Norway spruce on radiative forcing through changes in carbon stocks and albedo. Forestry 94 (4), 588–597.
Kolevatova, A., Riegler, M.A., Cherubini, F., Hu, X.P., Hammer, H.L., 2021. Unraveling the impact of land cover changes on climate using machine learning and explainable artificial intelligence. Big Data Cogn. Comput. 5 (4).
Kotlarski, S., et al., 2014. Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci. Model. Dev. 7 (4), 1297–1333.
Kuhn, M., 2019. Caret: Classification and Regression training, R package. URL. http://topepo.github.io/caret/train-models-by-tag.html.
Kumkar, Y., Astrup, R., Stordal, F., Bright, R.M., 2020. Quantifying regional surface energy responses to forest structural change in nordic Fennoscandia. J. Geophys. Res. Atmos. 125 (15).
Kutner, H., Nachtsheim, J., Neter, J., Li, W., 2004. Applied Linear Statistical Models. McGraw-Hill/Irwin, New York, p. 701.
Lawrence, D.M., et al., 2019. The community land model version 5: description of new features, benchmarking, and impact of forcing uncertainty. J. Adv. Model. Earth Syst. 11 (12), 4245–4287.
Lee, X., et al., 2011. Observed increase in local cooling effect of deforestation at higher latitudes. Nature 479 (7373), 384–387.
Li, Y., et al., 2016. The role of spatial scale and background climate in the latitudinal temperature response to deforestation. Earth Syst. Dynam. 7 (1), 167–181.
Li, Y., et al., 2015. Local cooling and warming effects of forests based on satellite observations. Nat. Commun. 6, 6603.
Luyssaert, S., et al., 2014. Land management and land-cover change have impacts of similar magnitude on surface temperature. Nat. Clim. Change 4 (5), 389–393.
Luyssaert, S., et al., 2018. Trade-offs in using European forests to meet climate objectives. Nature 562 (7726), 259–262.

Ma, N., et al., 2020. Ground observed climatology and trend in snow cover phenology across China with consideration of snow-free breaks. Clim. Dyn. 55 (9–10), 2867–2887.
Ma, W., Jia, G.S., Zhang, A.Z., 2017. Multiple satellite-based analysis reveals complex climate effects of temperate forests and related energy budget. J. Geophys. Res. Atmos. 122 (7), 3806–3820.
Majasalmi, T., Eisner, S., Astrup, R., Fridman, J., Bright, R.M., 2018. An enhanced forest classification scheme for modeling vegetation-climate interactions based on national forest inventory data. Biogeosciences 15 (2), 399–412.
McGrath, M.J., et al., 2015. Reconstructing European forest management from 1600 to 2010. Biogeosciences 12 (14), 4291–4316.
Mooney, P.A., Lee, H., Sobolowski, S., 2021. Impact of quasi-idealized future land cover scenarios at high latitudes in complex terrain. Earths Future 9 (2).
Mu, Q.Z., Zhao, M.S., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115 (8), 1781–1800.
Mykleby, P.M., Snyder, P.K., Twine, T.E., 2017. Quantifying the trade-off between carbon sequestration and albedo in midlatitude and high-latitude North American forests. Geophys. Res. Lett. 44 (5), 2493–2501.
Naudts, K., et al., 2016. Europe's forest management did not mitigate climate warming. Science 351 (6273), 597–600.
Pan, Y., et al., 2011. A large and persistent carbon sink in the world's forests. Science 333 (6045), 988–993.
Peng, S.S., et al., 2014. Afforestation in China cools local land surface temperature. P Natl. Acad. Sci. USA, 111 (8), 2915–2919.
Peng, S.S., et al., 2013. Change in snow phenology and its potential feedback to temperature in the Northern Hemisphere over the last three decades. Environ. Res. Lett. 8 (1).
Perugini, L., et al., 2017. Biophysical effects on temperature and precipitation due to land cover change. Environ. Res. Lett. 12 (5), 053002.
Pitman, A.J., et al., 2011. Importance of background climate in determining impact of land-cover change on regional climate. Nat. Clim. Change 1 (9), 472–475.
Portmann, R., et al., 2022. Global forestation and deforestation affect remote climate via adjusted atmosphere and ocean circulation. Nat. Commun. 13 (1), 5569.
Quinlan, J.R., 1992. Learning with continuous classes. In: 5th Australasian Joint Conference on Artificial Intelligence, pp. 343–348.
Quinlan, J.R., 1993. Combining instance-based and model-based learning. In: 10th International Conference on International Conference on Machine Learning, pp. 236–243.
Rodell, M., et al., 2004. The global land data assimilation system. Bull. Am. Meteorol. Soc. 85 (3), 381. -+.
Rotenberg, E., Yakir, D., 2010. Contribution of semi-arid forests to the climate system. Science 327 (5964), 451–454.
Schultz, N.M., Lawrence, P.J., Lee, X.H., 2017. Global satellite data highlights the diurnal asymmetry of the surface temperature response to deforestation. J. Geophys. Res.-Biogeo 122 (4), 903–917.
Sharma, A., Agrawal, J., Agarwal, S., Sharma, S., 2013. Machine learning techniques for data mining: a survey. In: 2013 IEEE International Conference on Computational Intelligence and Computing Research, pp. 1–6.
Tang, B.J., Zhao, X., Zhao, W.Q., 2018. Local effects of forests on temperatures across Europe. Remote Sens.-Basel 10 (4).
Tian, L., Zhang, B.Q., Wang, X.J., Chen, S.Y., Pan, B.T., 2022. Large-scale afforestation over the loess plateau in China contributes to the local warming trend. J. Geophys. Res. Atmos. 127 (1).
Wan, Z., Hook, S., Hulley, G., 2015. MYD11C3 MODIS/aqua land surface temperature/emissivity monthly L3 Global 0.05Deg CMG V006. In: 2015, distributed by NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MYD11C3.006. Accessed 27 May 2022.
Wan, Z.M., 2014. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. Remote Sens. Environ. 140, 36–45.
Winckler, J., Reick, C.H., Pongratz, J., 2017. Why does the locally induced temperature response to land cover change differ across scenarios? Geophys. Res. Lett. 44 (8), 3833–3840.
Yu, P., et al., 2022. Global spatiotemporally continuous MODIS land surface temperature dataset. Sci. Data 9 (1), 143.
Zhang, M., et al., 2014. Response of surface air temperature to small-scale land clearing across latitudes. Environ. Res. Lett. 9 (3).
Zhang, Y.Z., Liang, S.L., 2018. Impacts of land cover transitions on surface temperature in China based on satellite observations. Environ. Res. Lett. 13 (2).
Zhou, L., Dickinson, R.E., Tian, Y., Vose, R.S., Dai, Y., 2007. Impact of vegetation removal and soil aridation on diurnal temperature range in a semiarid region: application to the Sahel. Proc. Natl. Acad. Sci. USA, 104 (46), 17937–17942.
Zhou, N., et al., 2021. Overview of recent land cover changes, forest harvest areas, and soil erosion trends in Nordic countries. Geogr. Sustain. 2 (3), 163–174.