Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

An enhanced forest classification scheme for modeling vegetation-climate interactions based on national forest inventory data

Titta Majasalmi[1], Stephanie Eisner[1], Rasmus Astrup[1], Jonas Fridman[2], Ryan M. Bright[1]

[1]Norwegian Institute of Bioeconomy Research (NIBIO), 1431 Ås, Norway
[2]Swedish University of Agricultural Sciences (SLU), 901 83 Umeå, Sweden

Correspondence to: Titta Majasalmi (titta.majasalmi@nibio.no)

Abstract. Forest management greatly affects the distribution of tree species and the age class composition of a forest, shaping its overall structure and functioning, and in turn, surface-atmosphere exchanges of mass, energy, and momentum. In order to attribute climate effects to anthropogenic activities like forest management, good accounts of forest structure are necessary. Here, using Fennoscandia as a case study, we make use of regional National Forest Inventory (NFI) data to systematically classify forest cover into groups of similar aboveground forest structure. An enhanced forest classification scheme and related Look-Up-Table (LUT) of key forest structural parameter values was developed, and the classification was applied for NFI maps from Norway, Sweden and Finland. To provide a complete surface representation, our product was integrated with the European Space Agency Climate Change Intiative's Land Cover (ESA CCI LC) map of present land cover (v.1.6.1) (http://maps.elie.ucl.ac.be/CCI/). An enhanced grouping by aboveground structure can improve climate predictions in intensively managed forested regions and is consistent with climate model routines that simulate the effects of land transitions through area-based changes in vegetation cover. Further, such a classification scheme is congruent with existing forestry tools employed to predict the evolution of forest structure over interannual time scales, and as such, may be viewed as a tool that links the climate and forest modeling communities.

Keywords: Land use; land use change; forest management; LULCC; NFI; PFT; SVAT; CCI LC-product

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

1 Introduction

The structural properties of a forest largely determine the amount of mass, energy, and momentum exchanged with the atmosphere contributing to weather and climate at multiple scales (Bonan 2008). Given their controls on photosynthesis, albedo, evapotranspiration, structural attributes like canopy leaf area and heights are crucial variables in modeling of carbon, water, and energy budgets. Many land models employed in climate research characterize forests and other vegetated surfaces according to their biophysical properties, grouping them into what is often termed Plant Functional Types (PFTs). At the heart of many land model parameterizations are structural variables like Leaf Area Index (LAI) and canopy heights, whose values are often PFT-dependent and sourced from Look-Up Tables (LUTs). LAI quantifies the areal interface between the land surface and the atmosphere and belongs to a group of Essential Climate Variables (ECVs) (GCOS, 2012). Canopy top and bottom heights z_{top} and z_{bottom} are required for calculating roughness length and displacement height (Bonan et al., 2002) that determine resistances to heat, moisture, and momentum transfer.

Differences in forest structure within a given forest cover type (henceforth PFT) can differ substantially depending on whether the forest is natural or intensively managed, and capturing this difference in climate simulations of anthropogenic land use activities remains a large challenge. Because structure is PFT-dependent, climate impacts from anthropogenic disturbances in forests that do not involve changes in PFT area but which do modify structure – like a clear-cut harvest, for example -- go undetected. One way to overcome this is by expanding the number of forest PFTs with sufficient differentiation in key structural attributes. In a natural or primary forest (i.e., unmanaged) of a given species or phenology grouping, the structural attributes like LAI and canopy heights are more likely to be normally distributed in space, and thus a single PFT classification with these and other structural terms calibrated to the spatial mean may lead to reasonable climate predictions. However, in intensively managed forests of the same species or phenology grouping, key structural attributes are less likely to be normally distributed, thus use of the same PFT-dependent parameters will lead to biased predictions. In addition to grouping forests according to their shared phenological characteristics, further grouping according to their shared structural characteristics would strengthen prediction confidence in intensively managed regions. Further, an enhanced grouping by structure would make it easier to attribute climate effects to forest management activities as the relationships between forest structure and various intervention types (i.e., harvest, thinnings, fertilization, etc.) are relatively well-known and understood.

Climate modelers, however, have no real practical way of handling and assimilating the vast amounts of information stemming from the forestry science community surrounding the relationships between forest management and forest structure. Many regions, however, have well established National Forest Inventory (NFI) systems that continuously monitor the physical state of forests. In intensively managed regions, forest inventories will reflect the human influence on forest structure, and as such, they have often been used in research aiming to attribute climate effects to management activities

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

(Bright et al., 2014; Naudts et al., 2015; 2016). Classifying forests based on the structural properties they share at various successional stages under similar management conditions may be one way to link models of forestry with the land models employed in climate research. Important transient effects could then be included, for example, through changes in area under a given successional stage, with forestry models providing the link to the time dimension. Alternatively, distinct rule sets for successional dynamics following management disturbances could be developed analogous to those which are used to govern growth and competition in dynamic vegetation models (or land models run in dynamic vegetation mode).

Here, we exploit NFI data to develop a forest classification scheme that better reflects the diversity in forest structure under managed conditions, and which facilitates the modeling of transient behavior connected to forest succession. From the perspective of climate modeling, NFI data is well-suited for enhancing the structural description of forests in global land cover datasets, because similar data is available for most developed countries and new data is collected annually. The scheme is based on K-medoids clustering and Mahalanobis distance analysis of NFI data grouped by forest composition. We focus on the Fennoscandian region (Norway, Sweden, and Finland) to develop our concept as it represents one of the most intensively managed forested regions of the world. We illustrate how an NFI-based classification scheme can be applied to enhance the representation of forests in global LC datasets like the European Space Agency (ESA) Climate Change Initiative's (CCI) Land Cover (LC) -product (CCI LC, 2016).

2 Materials and Methods

2.1 Field data

Norwegian NFI data (Tomter et al., 2010) from 2007 to 2015 and Swedish NFI data (Fridman et al., 2014) from 2011 to 2015 were used in this study. NFI employs a network of field plots from which trees are measured and the growth monitored systematically. NFI data are systematically collected and processed by forest authorities and are used to quantify the amount and extent of forest at national level. The large diversity in forest structure throughout the Fennoscandian region (**Fig. 3**) as shaped by climate, management, and topography are well-represented in the Swedish and Norwegian NFI data. The Norwegian NFI contained data from 10,813 circular 8.92 m radius sample plots (250 m^2), while the Swedish data contained data from 14,032 circular 10 m radius sample plots (314 m^2). Plots which were divided (i.e. not completely circular) or which did not have trees were excluded from the data prior to the analysis. The main tree species of the area are Norway spruce (*Picea abies* (L.) H. Karst.), Scots pine (*Pinus sylvestris*, L.), and Silver and Downy birches (*Betula pendula* Roth and *pubescens* Ehrh.). Monocultural plots of birch are rare, but birches are common in plots with different species mixtures. Field data were classified as spruce, pine or deciduous (contains also other tree species) dominated forests based on species with the largest share of total stem volume (m^3/ha) on the sample plot (**Table 1.**).

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

2.2 Forest classification scheme

The process flow of developing and applying the forest classification scheme is illustrated in **Fig. 1,** and summarized briefly here: NFI field data were first used to develop the forest classification scheme based on four key forest structural attributes: total stem volume (V), height (H), crown length (CL) and LAI. H is Lorey's height (the basal area-weighted height) which corresponds well with the aerodynamic height (Nakai et al., 2010), and LAI is the maximum growing season LAI. Models to calculate the plot total maximum LAI and CL are described in supplementary file (S1 and S2). A clustering analysis was subsequently employed to define clusters and solve class memberships based on species, V, and H. Then, the classification was applied to Multi-Source-NFI maps (MS-NFI) from Norway, Sweden, and Finland at high resolution. MS-NFI maps extrapolate forest characteristics for areas between NFI field plots using a non-parametric k-Nearest Neighbor (kNN) estimation method (e.g. Tomppo et al., 2014). This extrapolation step is called "multi-source" because it employs data from different remote sensing systems (i.e. satellite and aerial platforms) and field plots. MS-NFI applies high resolution satellite images to separate forested areas from other LC-categories and digital terrain models to correct topographical distortions. Finally, the classified maps were reprojected, aggregated, and resampled to the ESA CCI LC-product resolution (~300m). For each forested CCI LC-pixel, forest class and within-pixel tree cover fraction (%) were obtained based on classified high resolution maps. Two exceptions occurred: 1) If the original CCI LC-pixel was not classified as forest, but the MS-NFI maps indicated the presence of forest, and 2) If the original CCI LC-pixel was classified as a forest but MS-NFI maps indicate non-forest. In the former case, the pixel was classified as forest, whereas in the latter case, a gapfilling method was employed. In addition, for each forest CCI LC-pixel, coverage fractions ('percentage layers') for each of the twelve forest classes were calculated (i.e. twelve layers with values ranging between 0 and 100 based on subgroup abundance within the CCI LC-pixel). These layers were calculated to allow more flexibility in terms of number of input land cover classes in different land models (i.e. modelers may use e.g. three most abundant forest classes instead of keeping to a single class).

2.2.1 K-medoids clustering

The Norwegian and Swedish NFI data were merged prior to the classification exercise. The plots were grouped based on predefined vegetation classes (i.e. spruce, pine, and deciduous dominated to account for differences in forest structural properties between different species groups (**Table 1.**). A K-medoids algorithm was used for clustering because only the number of clusters is required as an input, and also because it is robust against outliers. K-medoids clustering was used to define the 'centroids' within V-H-CL-LAI – space. CL of deciduous species was modeled based on birch models (described in S2). The cluster median values of V-H-CL and LAI were used to create LUT for the three species groups (i.e. spruce, pine and deciduous). The K-medoids algorithm selects a random set of *n* medoids and computes distances of all data points to cluster medoids. Points are classified belonging to the cluster they are most similar to according to the sum of minimized squared Euclidean distances. All variables were normalized prior to analysis. Centers of the medoids are adjusted iteratively until medoids do not change. The analysis was run using the 'cluster' package in R. The optimal number of clusters was

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

assessed as the decrease in total within sum of squares as the number of clusters increased (i.e. 'elbow method' (Ketchen & Shook, 1996)) using R package 'factoextra'. The optimal number of clusters was determined to lie between three and five, thus the number of clusters was set to four.

2.2.2 Mahalanobis distance

A method to assess cluster boundaries was needed, because many plots were located near the edges of the four-dimensional clusters. We chose to determine class memberships using two variables V and H since these are often available for large geographical areas from NFIs. Mahalanobis Distance (MD) was used to quantify the within-cluster variation within V- and H–space (i.e. VH-space), because it corresponds to the Euclidean distance after V and H have been normalized. MD is a multidimensional method to determine how many standard deviations a data point is away from the class mean. MD values were calculated for each species group and the respective subgroups. The binning (i.e. grid of 14 × 14) interval for V-space was set subjectively to add resolution on younger forest structures. For each grid cell, and for each subgroup, a median MD value was calculated. To represent results using a grid surface, the cell was assigned to a group with the smallest mean MD. Analysis was conducted using the 'stats' package in R.

2.3 Application of the classification to MS-NFI maps

2.3.1 Description of the MS-NFI maps

We applied our forest classification scheme to MS-NFI maps of 2013 from Finland (LUKE, 2016) and of 2010 from Sweden (SLU, 2016) (background information provided in S4). Our classification was applied to high resolution maps of V by species and H (Sweden: 0.025 km^2, Finland: 0.016 km^2). For Norway, MS-NFI data (compiled during the first decade of the twenty-first century) called 'SAT-SKOG' (Gjertsen, 2010) and a forest resource map called 'AR5' (Ahlstrøm et al., 2014) were used to obtain all required inputs and cover the northernmost forest areas (i.e. Finnmark county). SAT-SKOG is provided as geospatial vector data and was rasterized to 0.025 km^2 resolution. SAT-SKOG does not contain tree height information which was modeled based on tree species, tree age at breast-height (i.e. 1.3 m above ground surface), and site index using equations by Tveite and Braastad (1981). Separate equations were used for pine, spruce, and birch (birch model was applied to all deciduous species). However, as our forest classification requires H (i.e. Lorey's height), and not the mean tree height calculated using site index, a separate model was developed based on Norwegian NFI data to scale the plot mean height into H (described in S3).

2.3.2 Processing the MS-NFI maps

MS-NFI data were classified as spruce, pine or deciduous dominated based on species with the largest share of pixel total stem volume (m^3/ha). The share of other tree species than pine or spruce was assigned to deciduous group. After the species group was assigned, a separate LUT was used to determine pixel subgroup in VH-space (see **section 2.2.2**). Possible VH-

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

combinations without MD value (i.e. falling outside the VH-space) were assigned to the closest subgroup based on V. After classifying all data, the forest classes were recoded as integers between one and twelve (i.e. three species groups × four subgroups). Finnmark county, the northernmost forested area in Fennoscandia, is currently not covered by SATSKOG but contains NFI-field plots and forest extent information from AR5. The NFI-field plots were classified based on species, V and H. AR5-based forest mask (resolution 0.02 km^2) was used to identify forested areas, and for all pixels within a forest mask, forest class was assigned based on nearest neighbor.

The classified high resolution maps were reprojected, aggregated, and resampled to complement the new ESA CCI LC-product from the 2008-2012 time period (v.1.6.1) (CCI LC, 2016), which, temporally, approximately corresponds to the MS-NFI data used in this study. The LC-product has three PFTs to describe boreal forests in Fennoscandia: broadleaved deciduous (60-62), needleleaved evergreen (70-72), and mixed broadleaved and needleleaved (90) (original LC-label values in parenthesis, shown in **Fig. 4**). Second digit from the left is designed to indicate the forest fraction within a LC-pixel. The canopy is 'closed' when forest pixel cover fraction is >40% (class labels 61 and 71) or 'open' when forest pixel cover fraction is between 15-40% (class labels 62 and 72). Labels 60 and 70 are used to indicate that the forest pixel fraction within that pixel is more than 15%, but it is not known whether that pixel is closed or open. The CCI LC-product used in this study contained only classes 60, 70 and 90 i.e. no pixels were assigned to subclasses (61, 62, 71 or 72). However, we add this subclass information to the CCI LC-product based on the high resolution MS-NFI data. Two types of aggregation routines were used for upscaling: forest class was assigned based on mode (among the twelve forest classes), and forest cover based on mean (for this purpose, forested pixels in high resolution data were recoded as 100 and other pixels as 0).

We imported non-forest LC-classes from the CCI LC-product to supplement our forest classification. For pixels which were classified as forest by the CCI LC-product, but forest cover fraction within that pixel did not exceed the 15% threshold (i.e. definition used by the original CCI LC-product) according to the MS-NFI data, forest class was assigned based on nearest neighbor. Gapfilling was necessary because land (climate) models require completeness in LC to resolve computations of mass, energy, and momentum fluxes. In order to assess changes between the LC-classes, the percentage layers of different forest subgroups were classified to correspond with the original CCI LC-classes: If >=70% the pixels in high resolution data, within the CCL LC-pixel, were classified into conifer or deciduous groups, the pixel was considered as 'needleleaved' (class 70) or 'broadleaved' (class 60), but otherwise 'mixed' (class 90). Raster analyses were performed using 'rgdal' and 'raster' packages and confusion matrix (i.e. error matrix between original CCI LC-product and 'back-classified' enhanced CCI CL-map) calculated using 'caret' package in R.

2.3.3 Recoded class labels for the Enhanced CCI LC-product

The forest classes (labels: 1-12) were recoded by adding 300 (i.e. forest class 12 would be coded as 312). In addition, two digits were added after recoded forest class number to indicate forest cover within the CCI LC-pixel, and whether a given

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

pixel is a 'true' forest pixel based on MS-NFI data or whether it is gapfilled. The fourth digit is used to indicate the fraction of forested pixels within an LC-pixel: A value of '1' indicates that the fraction of forested pixels within an LC-pixel is > 40%, and value '2' denotes that the fraction of forested pixels within an LC-pixel is between 15-40%. For 'true' forested pixels the last digit is '0', whereas for gapfilled pixels the last digit is '1'.

3 Results

3.1 Forest classification scheme

As a result of our classification scheme, a LUT of the key structural variables (i.e. V, H, CL, and LAI) was created (**Table 2.**). The boundaries of the subgroups were determined based on MD, which can be visualized using a gridded representation of vegetation subgroups within the VH-space (**Fig. 2**) to select the right values from the LUT. The classified grid area and subgroup membership patterns reflect the variability of V and H in NFI data, which was used to define the classes. For example, the spruce dominated plots may have V up to 1500 m^3/ha. In pine dominated plots the V did not exceed 900 m^3/ha, and in deciduous plots the highest V was 1100 m^3/ha. In spruce dominated plots, the H exceeded 30 m with many different Vs, whereas for pine the 30 m was exceeded either when the respective V was less than 50 m^3/ha (i.e. tree is left for seed production during harvesting which is a common forest regeneration strategy in Fennoscandia) or large (more than 500 m^3/ha). In plots dominated by deciduous species the 30 m exceeded after V was more than 150 m^3/ha. The location and size (i.e. patterns) of different subgroups in VH-space cannot be directly compared between different species groups, as Euclidean distances were used for their classification.

3.2 Enhanced CCI LC-product

The majority (59%) of the forest pixels in Fennoscandia were classified as pine dominated which was also the largest species group in Sweden (58%) and in Finland (72%) (**Table 3.**). However, in Norway the largest species group was deciduous (41%). Finland had slightly higher percentage of deciduous forests than Sweden. Spruce dominated forest was the smallest species group in Finland (14%) and in Norway (26%). Visual assessment of spatial distribution of different species groups and their subgroups showed that low-land areas in Finland and in Sweden were mainly dominated by pines and spruces, whereas deciduous species were most abundant in the northernmost, mountainous and coastal areas (**Fig. 3**). In Fennoscandia the most abundant subgroup within spruce dominated forest was 'Spruce 3' with class median values of V= 201 m^3/ha and H= 17m (see **Table 2.**). Within the pine dominated forest the most abundant subgroup was 'Pine 2' with class median V= 80 m^3/ha and H= 12m. For deciduous species group the median values of the largest subgroup 'Deciduous 1' were V= 7 m^3/ha and H=5m.

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

In order to assess agreement between different LC-classes, the enhanced CCI LC-product was back-classified into original CCI LC-classes using the percentage layers of different forest subgroups. Kappa coefficient (measure of agreement which takes into account possible agreement occurring by chance) for classification was 0.55 and classification accuracy was 0.64 (calculated based on **Table 5.**).

Confusion matrix between the original CCI LC-product and the enhanced back-classified CCI LC-map showed that the highest agreement (30.4%) between the two classifications schemes occurred for forest class 70 (i.e. needleleaved evergreen trees) (**Table 5.**). The biggest discrepancy between LC-classes occurred between different forest types, as expected. Results showed that 14.5% of original CCI LC-class 70 was classified as 90 (i.e. mixed broadleaved and needleleaved trees) and 1.4% of class 70 was classified into class 60 (i.e. broadleaved deciduous trees). The largest fraction of forested pixels in the back-classified enhanced CCI LC-map were classified as conifer dominated (=36.2%) (**Fig. 4**). The share of forest class 'mixed' (i.e. class 90) was also high (=30.3%). However, the share of pixels classified as deciduous was relatively low (=5.3%). Overall, the enhanced CCI LC-product contained 16% more forest pixels than the original CCI LC-product. For example, the forest area increased as 5.1% of original CCI LC-class 180 (i.e. shrub or herbaceous cover), 3.6% of class 100 (i.e. mosaic tree and shrub (>50%)), and 2.1% of crop LC-classes were classified as forest in the enhanced CCI LC-product. The fraction of gapfilled forest pixels was 4.2% of all classified pixels in the enhanced CCI LC-product. The classified land area increased by 0.5% as areas classified as no-data in the original CCI LC-product were classified as forest in the enhanced CCI LC-product. The spatial distribution of different LC-classes and their frequencies in the back-classified enhanced CCI LC-map are shown in **Fig. 4**, which shows both LC-class labels and descriptions.

4 Discussion

We developed a method for adding forest PFT classes based on K-medoids clustering of four key structural attributes extracted from NFI field data in which the differentiation is based on the sum of minimized Euclidean distances to cluster centroids. The LUT values obtained from clustering analysis are medians, and thus provide conservative estimates (i.e. cannot represent extreme values) of V, H, CL and LAI. While the clustering analysis was performed using Swedish and Norwegian NFI data, the LUT may be assumed applicable in Finland because the biomass functions by Marklund (1988) are applicable in Finland (Kärkkäinen 2005) and given the similarities in commercial species and forest management practices in Fennoscandia.

Recently, other approaches have been developed for incorporating forest management into existing land surface (climate) models. For example, the radiative transfer based land-surface model ORCHIDEE was parameterized to simulate the effects of forest management for biogeochemical and biophysical variables (Naudts et al., 2015). The model was parameterized using diameter-at-breast-height (dbh) data from different European NFIs (French, Spanish, Swedish and German) (i.e. the key input values were modeled based on dbh using allometric models), and twelve parameter sets for specific tree species (instead of presenting groups of species such as PFTs) were presented. However, a major drawback of individual tree-based approaches is that existing global LC maps are not designated to distinguish between individual species, which limits the spatial domain where such approaches can be applied. In addition, the need for residual groups remains because individual tree based approaches are not suited for areas where the forests are essentially mixtures of different tree species. The benefit of defining 'broader' PFT classes, such as those developed in this study, is that the broad functional types may be separated from optical satellite data based on differences in optical and structural characteristics of the forests. In the future, as both the spectral, spatial, and temporal resolution of the optical satellite data improves, definition of narrower forest classes may be justified. Alternatively, Functional Traits (FT) may be used for modeling vegetation-climate interactions (Wullschleger et al., 2014; Verheijen et al., 2013). Commonly, the community-weighted-mean trait value (i.e. based on relative abundances of species and their trait values) is used in models which apply the FTs concept. While FTs are highly scalable (i.e. from organism to ecosystem scale), well assembled (i.e. leaf, stem and root traits), and measurable, the downside of FTs is that the applicability of FTs is in its infancy and the lack of standards hinders its practical application.

At present, some countries, such as Finland and Sweden, have national Airborne Laser Scanning (ALS) datasets, which could be used to obtain more accurate forest height estimates or to develop forest classification schemes for different land models. However, the drawback of these ALS datasets is that they cannot be used to separate between different tree species, which is one of the most important forest structural attributes. In addition, as few countries have national ALS datasets the geographical extent which could be covered using ALS based forest classification schemes would remain limited. The use of optical satellite data to classify the forests is unquestionable due to its superior spatial and temporal resolution, and thus will sustain its role as the most valuable tool for environmental monitoring and mapping. However, in the future, approaches combining both optical and ALS data may be expected to become more common, and thus allow development of more sophisticated forest classification schemes to increase the accuracy of the climate predictions in managed forested regions.

This paper is response to the 'call to action' raised in the review by Ellison et al., (2017) which highlighted an urgent need to incorporate forest management impacts into existing modeling schemes in order to reach better political decisions regarding climate change adaptation, mitigation, and land use and water management. One of goals of this paper is to foster interdisciplinary discussions on alternative information sources, such as the existing NFI datasets, in Fennoscandia spanning back to the 1920's, to enhance representation of forests in different land modeling frameworks. We developed a simple clustering and classification scheme to allow reiteration of our approach to NFI data from other countries. The enhanced LC maps and temporal descriptions of forest key structural attributes are needed for forecasting and back-casting the impacts of forest management on energy, water and carbon cycling.

The presented forest classification scheme has many levels to serve the needs of different users. For example, for climate and hydrological modeling requiring full spatial coverage, the gapfilled pixels and non-forest LC-classes are provided.

Researchers that are able to run their models with no-data may select to remove the gapfilled pixels prior to analysis. Remote sensing scientists may wish to use only 'true' forest pixels and extract areas belonging to different species groups or subgroups, or select areas where the fraction of forests is lower or higher (i.e. 'open' or 'closed'). In addition, the sub-pixel fraction – or the relative abundance of different forest subgroups within each CCI LC-pixel -- provides land modelers more control and flexibility in terms of the number of input LC-classes in different land models: The percentage layers for different forest subgroups may be used to obtain complete PFT-distributions (retrieved from high resolution data) for Fennoscandia or alternatively, modeler may choose to use e.g. three most abundant forest classes instead of holding on one class. The sub-pixel fractions also provide greater "cross-walking" (Poulter et al. 2015) ability at finer scale resolution. Our forest classification scheme and the related map products (i.e. Enhanced CCI LC-product and the respective percentage layers) allow customized model 'inputs' to fit the needs (or requirements) of various land models.

Simple clustering analysis of the NFI data was used to optimize grouping of forest cover based on four key structural attributes that are strong controls of surface-atmosphere exchanges of mass, energy, and momentum. Our new forest cover classification was subsequently applied to MS-NFI maps, which were then aggregated and fused with the ESA CCI LC-product. The resulting enhanced CCI LC-product and related LUT of the key structural variables, can now be used to better quantify surface fluxes linked to present-day forest cover. To our knowledge, this is the first study to use NFI field data together with MS-NFI maps to enhance the characterization of forest structure in a format that is compatible with many land surface (climate) models (i.e. in modeling frameworks) where changes in vegetation structure are captured by area-based changes in PFTs. The methods used for creating the LUT were carefully explained to allow other researchers to replicate the same procedures using NFI data from other countries. The benefit of the classification scheme described in this study is that the required data (i.e. NFI data and MS-NFI maps of species, V and H) are readily available for many countries. Future research is needed to evaluate the sensitivity of present-day predictions in carbon, moisture, and energy fluxes to the different PFT grouping levels, and to develop recommendations and guidelines for prescribing future forest transitions under changing climate and management regimes.

Data availability

The MS-NFI Forest resource maps for Finland are available through Natural Resources Institute Finland (LUKE) portal: http://kartta.luke.fi/opendata/valinta.html. For Sweden the forest maps may be obtained through Swedish University of Agricultural Sciences (SLU) portal: http://www.slu.se/en/Collaborative-Centres-and-Projects/the-swedish-national-forest-inventory/forest-statistics/slu-forest-map/. For Norway the MS-NFI data are available by request.
The enhanced CCI LC-product for Fennoscandia, including the percentage layers, can be downloaded from: (DOI will be added here)

Acknowledgements

The research was funded by the Research Council of Norway, grant number 250113/F20.

We acknowledge the work done by ESA CCI Land Cover project.

References

Ahlstrøm, A. P., Bjørkelo, K. and Frydenlund, J. AR5 KLASSIFIKASJONSSYSTEM: Klassifikasjon av arealressurser. Rapport fra Skog og landskap 06/14: III, 38s. http://www.skogoglandskap.no/filearchive/rapport_06-2014.pdf , 2014.

Bonan, G.B.Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. Science, 320, 1444-1449, 2008.

Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., ... and Yang, Z. L. The land surface climatology of the community land model coupled to the NCAR community climate model. J. Clim., 15, 3123-3149, 2002.

Bright R.M., Antón-Fernández C, Astrup R, Cherubini F, Kvalevåg MM, and Strømman A.H. Climate change implications of shifting forest management strategy in a boreal forest ecosystem of Norway. Glob. Chang. Biol., 20, 607-621, 2014.

CCI LC, 2016 European Space Agency (ESA) Climate Change Initiative Land Cover (CCI LC) -product http://maps.elie.ucl.ac.be/CCI/ [downloaded: 7 Oct 2016]

Ellison, D., Morris, C. E., Locatelli, B., Sheil, D., Cohen, J., Murdiyarso, D., ... and Gaveau, D. Trees, forests and water: Cool insights for a hot world. Glob. Environ. Chang., 43, 51-61, 2017.

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., Ringvall, A. H., and Ståhl, G. Adapting National Forest Inventories to changing requirements – the case of the Swedish National Forest Inventory at the turn of the 20th century. Silva Fenn. 48, 1-29, 2014. http://dx.doi.org/10.14214/sf.1095.

Gjertsen, A.K. SAT-SKOG kom på Internett i 2009. Årsmelding fra Skog og landskap 2009: 27. http://www.skogoglandskap.no/filearchive/sat_skog_kom_pa_internett_i_2009.pdf , 2010.

Global Climate Observing Systems -network (GCOS). Essential Climate Variables. http://www.wmo.int/pages/prog/gcos/index.php?name=EssentialClimateVariables, 2012.

Ketchen Jr, D. J., and Shook, C. L. The application of cluster analysis in strategic management research: an analysis and critique. Strategic Manage J., 441-458, 1996.

Kärkkäinen, L. Evaluation of performance of tree-level biomass models for forestry modeling and analyses. Finnish Forest Research Institute, Research papers, 940: p123, 2005.

Marklund, L. G. Biomassafunktioner för tall, gran och björk i Sverige. Sveriges lantbruksuniversitet, Institutionen för skogstaxering, 1988.

Nakai, T., Sumida, A., Kodama, Y., Hara, T., and Ohta, T. A comparison between various definitions of forest stand height and aerodynamic canopy height. Agric. For. Meteorol., 150, 1225-1233, 2010.

Natural Resources Institute Finland (LUKE), Forest resource maps. http://kartta.luke.fi/opendata/valinta.html [downloaded: Oct 15 2016]

Naudts, K., Ryder, J., McGrath, M. J., Otto, J., Chen, Y., Valade, A., ... and Ghattas, J. A vertically discretised canopy description for ORCHIDEE (SVN r2290) and the modifications to the energy, water and carbon fluxes. Geosci. Model Dev., 8, 2035-2065, 2015.

Naudts, K., Chen, Y., McGrath, M. J., Ryder, J., Valade, A., Otto, J., and Luyssaert, S. Europe's forest management did not mitigate climate warming. Science, 351, 597-600, 2016.

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., ... and Hagemann, S. Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative. Geosci Model Dev, 8, 2315-2328, 2015.

Swedish University of Agricultural Sciences (SLU) Forest maps. http://www.slu.se/en/Collaborative-Centres-and-Projects/the-swedish-national-forest-inventory/forest-statistics/slu-forest-map/ [downloaded: Oct 29 2016]

Tomppo, E., Katila, M., Mäkisara, K. and Peräsaari, J. The Multi-source National Forest Inventory of Finland - methods and results 2011. Metlan työraportteja / Working Papers of the Finnish Forest Research Institute 319. 224 p. http://www.metla.fi/julkaisut/workingpapers/2014/mwp319.htm, 2014.

Tomter S, Hylen G, Nilsen J-E, Norways national forest inventory. In: Tomppo E, Gschwantner T, Lawrence M, McRoberts RE (Eds.): National forest inventories: pathways for common reporting. Springer, Berlin. 2010

Tveite, B. and Braastad H. Bonitering for gran, furu og bjørk. Norsk Skogbruk 27, 17-22, 1981.

Verheijen, L. M., Brovkin, V., Aerts, R., Bönish, G., Cornelissen, J. H. C., Kattge, J., ... and Van Bodegom, P. M. Impacts of trait variation through observed trait-climate relationships on performance of a representative Earth System model: a conceptual analysis. Biogeosciences, 10, 5497-5515, 2013.

Wullschleger, S. D., Epstein, H. E., Box, E. O., Euskirchen, E. S., Goswami, S., Iversen, C. M., ... and Xu, X. Plant functional types in Earth system models: past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. Ann. Bot., mcu077, 2014.

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

Figures and tables

Table 1. Descriptive statistics for the National Forest Inventory (NFI) data. Abbreviations: n=number of sample plots, dbh= diameter-at-breast-height, H=basal area weighted mean tree height (i.e. Lorey's height) and V=total stem volume.

			dbh (cm)		H (m)		V (m^3/ha)	
Area	Species	n	mean	range	mean	range	mean	range
Norway	Spruce	3364	12.6	5.0 - 49.0	13.0	2.9 - 32.3	152.0	0.2 - 1492.4
	Pine	3650	14.1	5.1 - 48.9	11.5	2.4 - 28.6	97.8	0.2 - 656.6
	Deciduous	3799	9.4	5.0 - 99.9	8.3	2.4 - 24.8	53.0	0.2 - 592.9
Sweden	Spruce	4552	16.2	1.0 - 52.0	15.3	1.4 - 40.2	177.8	0.5 - 1010.2
	Pine	7028	16.2	1.0 - 64.6	13.6	1.4 - 32.1	120.0	0.6 - 752.3
	Deciduous	2452	12.8	1.0 - 81.2	12.8	1.5 - 32.6	101.9	0.4 - 1001.5

Table 2. A forest classification scheme Look-Up-Table (LUT). Abbreviations: V=total stem volume (m^3/ha), H=Lorey's height (m), CL=Crown length (m), and LAI=Leaf Area Index (Maximum during growing season; m^2/m^2). Recoded label -column is a key to be used with the Enhanced CCI CL-product.

			LUT values			
Species group	Subgroup	Recoded label	V	H	CL	LAI
Spruce	1	301	22.0	7.5	6.3	1.4
	2	302	92.2	12.3	10.1	4.3
	3	303	201.3	16.8	13.2	6.7
	4	304	373.9	22.0	15.8	9.1
Pine	1	305	20.8	7.5	4.6	0.9
	2	306	80.0	11.6	6.7	2.4
	3	307	129.5	17.0	9.4	2.3
	4	308	236.4	17.2	8.4	4.4
Deciduous	1	309	7.2	4.9	3.2	0.5
	2	310	36.1	8.4	5.5	1.8
	3	311	97.6	12.2	7.9	3.9
	4	312	227.0	18.3	10.3	7.0

Table 3. The percentage (%) of forest pixels (i.e. excluding gapfilled pixels) belonging to different species groups in the Enhanced CCI LC-product (referred as 'Fennoscandia') and separately for each country (The spatial distribution of different forest subgroups and their frequency distributions are shown in **Fig. 3)**. Values are based on MS-NFI data.

	Fennoscandia(%)	Norway (%)	Sweden (%)	Finland (%)
Spruce	22.4	26.2	29.0	13.7
Pine	58.8	32.7	58.3	71.7
Deciduous	18.8	41.1	12.6	14.6

Table 4. The percentage (%) of different forest classes in the Enhanced CCI LC-product (i.e. contains gapfilled forest pixels). The grey bars inside the cells are used to visualize distributions of different forest classes in Fennoscandia and separately for each country.

Group	Subgroup	Recoded label	Fennoscandia (%)	Norway (%)	Sweden (%)	Finland (%)
Spruce	1	301	2.5	0.2	1.9	0.4
	2	302	4.3	1.2	2.5	0.6
	3	303	11.7	2.5	6.0	3.1
	4	304	3.9	1.0	1.7	1.2
Pine	1	305	15.3	1.1	6.5	7.7
	2	306	20.4	1.5	8.8	10.1
	3	307	11.2	3.0	1.7	6.5
	4	308	11.9	0.6	7.3	4.0
Deciduous	1	309	8.2	2.7	3.0	2.6
	2	310	4.9	2.9	0.8	1.2
	3	311	3.4	1.4	0.5	1.4
	4	312	2.3	0.7	0.9	0.7

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

Table 5. Confusion matrix in percentage (%) between the original CCI LC-product and the back-classified enhanced CCI LC-map. Grey background is used to indicate the ten classes with the highest percentage of pixels, and small percentages are shown as 0.00. Kappa coefficient was 0.55. Label definitions shown in **Fig 4**.

Original CCI Land Cover Classification

Enhanced back-classified CCI Land Cover Classification	0	10	11	12	30	40	60	70	90	100	110	120	130	140	150	152	180	190	200	201	202	210	220
0																							
10		2.78																					
11			0.06																				
12																							
30					0.12																		
40						0.04																	
60	0.01	0.03	0.01		0.03	0.01	2.70	1.38	0.25	0.22	0.07				0.24	0.01	0.30	0.01	0.00			0.01	
70	0.09	0.14	0.01		0.03	0.03	0.67	30.38	1.68	1.21	0.04				0.19	0.04	1.45	0.08	0.00			0.17	
90	0.40	1.98	0.06		0.36	0.13	1.91	14.53	2.30	2.18	0.45		0.01		0.70	0.07	3.35	0.60	0.01	0.00		1.22	
100										0.72													
110											3.91												
120																							
130													0.04										
140														0.02									
150															9.09								
152																0.08							
180																	3.16						
190																		0.62					
200																			1.71				
201																				0.06			
202																					0.00		
210																						5.40	
220																							0.4

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

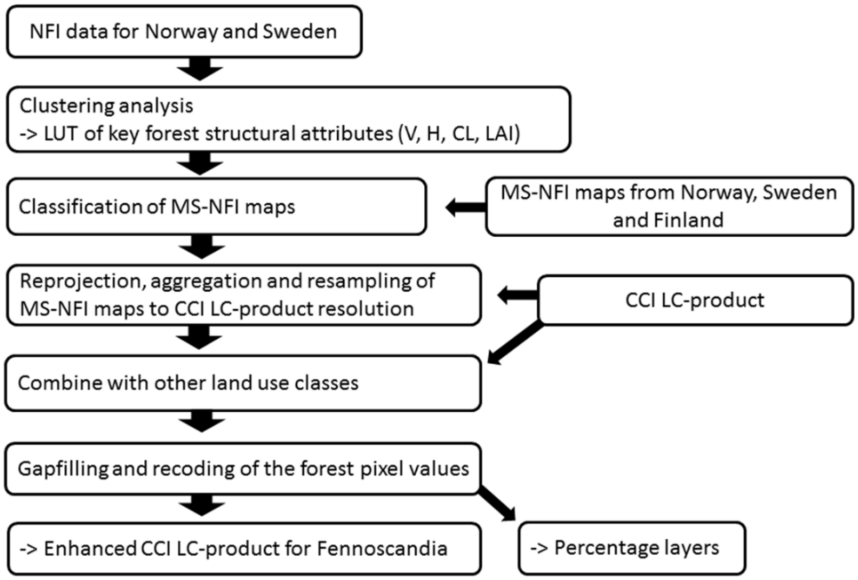

Figure 1. Flowchart for developing and applying the forest classification scheme. Abbreviations: National Forest Inventory (NFI) data, Look-Up Table (LUT), total stem volume (m^3/ha) (V), Lorey's height (H), Crown Length (CL), Leaf Area Index (LAI), Multi-Source NFI (MS-NFI, i.e. products provided by forest authorities), Climate Change Initiative Land Cover (CCI LC) product.

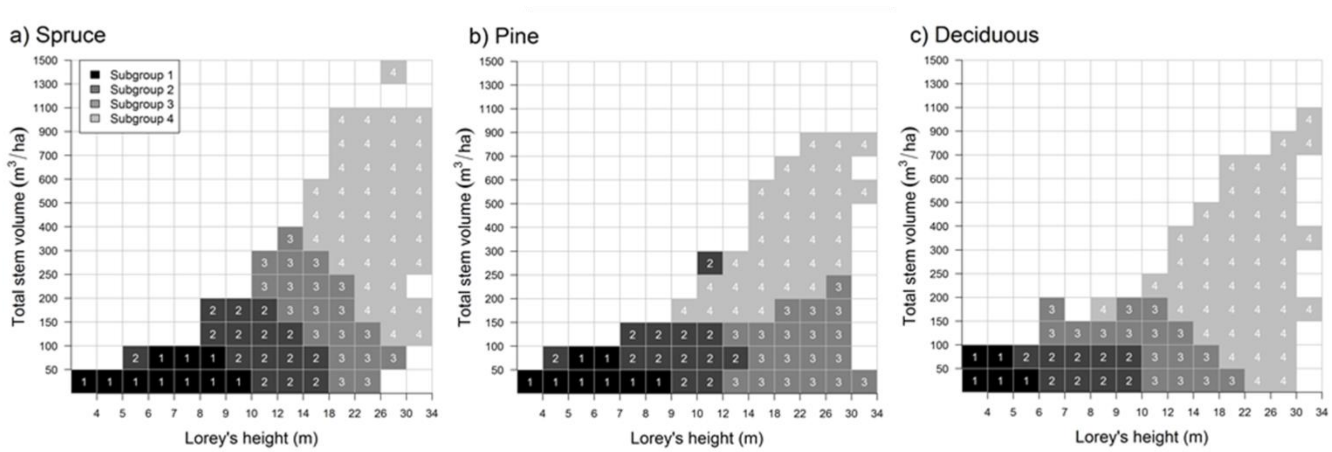

Figure 2. Gridded representation of vegetation subgroups, i.e. a) spruce, b) pine and c) deciduous, within the total stem volume (V) and Lorey's height (H) –space (referred as VH-space) based on NFI data. Visualization is required to map their distribution in VH -space and used to apply the classification to the MS-NFI maps.

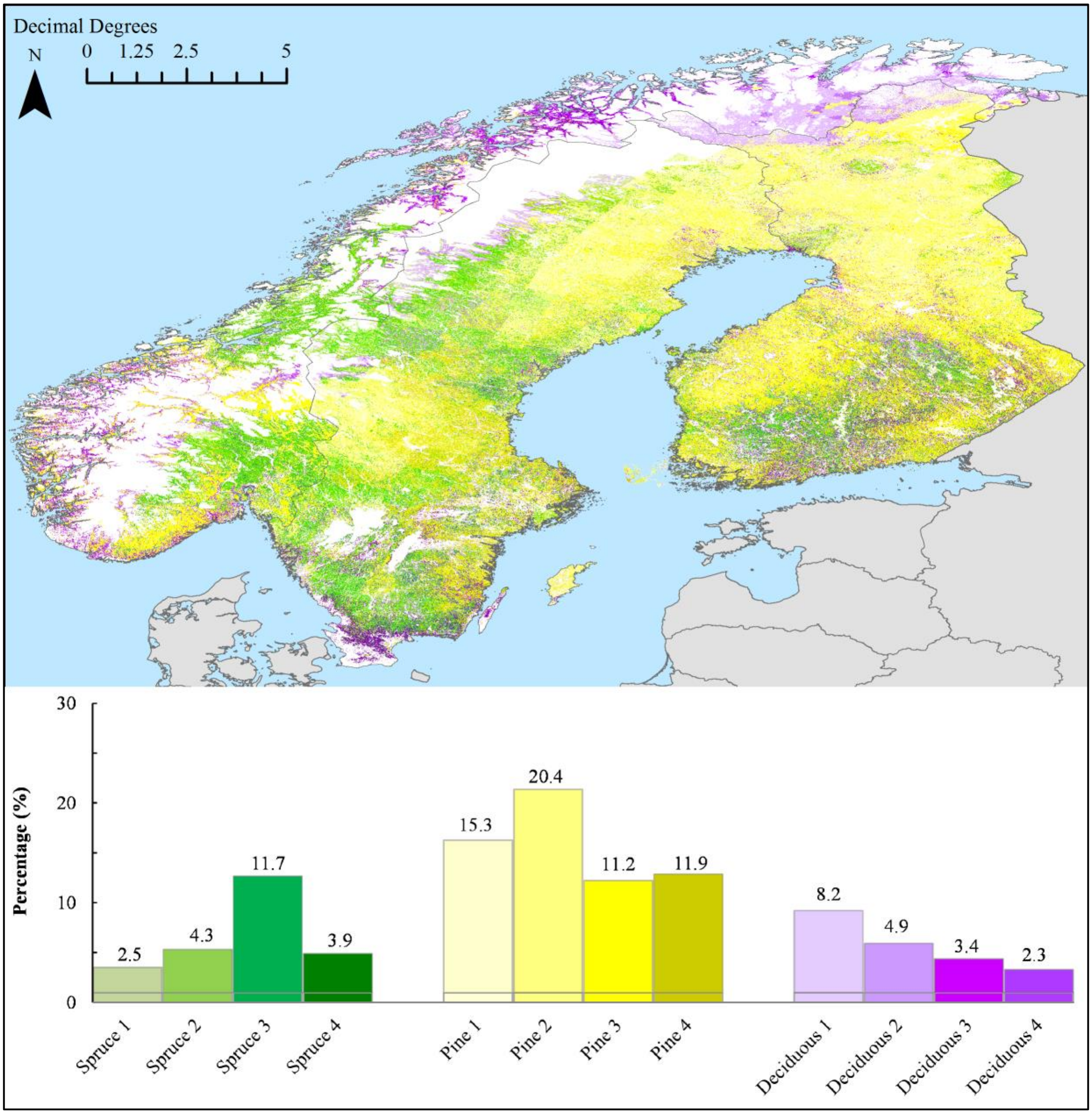

Figure 3. Spatial distribution of MS-NFI forest classes (i.e. without gapfilled forest pixels) in Fennoscandia. The forest subgroup was assigned based on most abundant forest class within the CCI LC-pixel. For colors, see online version of the article.

Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-301
Manuscript under review for journal Biogeosciences
Discussion started: 9 August 2017

Biogeosciences Discussions
Open Access

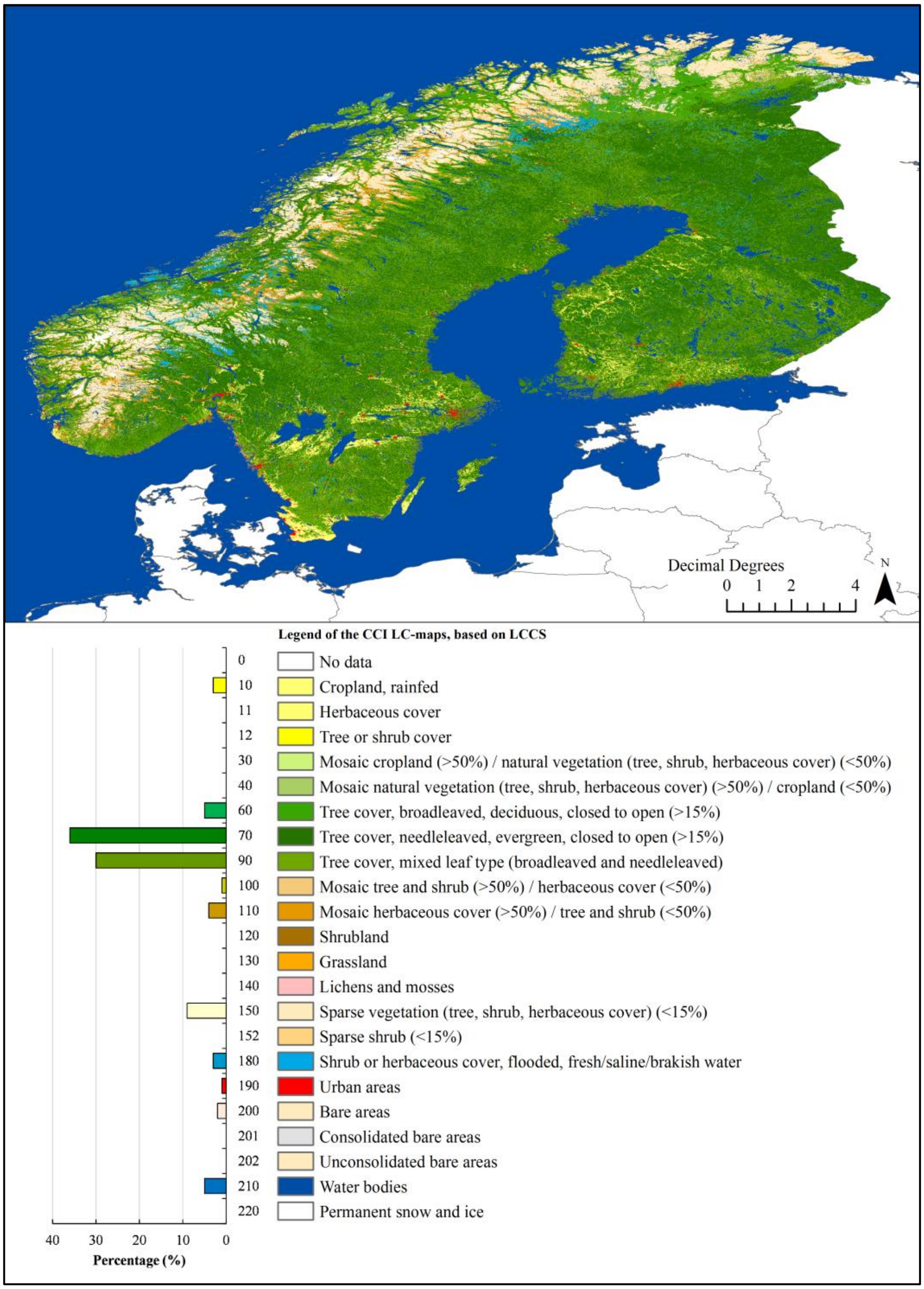

Figure 4. Back-classified enhanced CCI LC-map with non-forests LC-classes from the original CCI LC-product for Fennoscandia. The percentage layers of each forest subgroup were used for back-classifying our data into original CCI LC-forest classes (see section **2.3.2.**). Histogram shows LC-class percentages of the back-classified enhanced CCI LC-map. For colors, see online version of the article.