Dear Referees,

On behalf of all the co-authors I thank you for the insightful and constructive comments directed to the manuscript "Improving the representation of high-latitude vegetation in Dynamic Global Vegetation Models". We have prepared point-by-point responses to each of the comments and amended the manuscript in line with these comments. For convenience and reference, we have numbered the Referee comments with "RC-x.x", where the first "x" corresponds to the referee number and the second "x" to the respective comment. Each of our responses is offered below the respective comment emphasized in blue italics. Please note that the line numbers point to the "marked-up" version of the manuscript attached below the responses.

Kind Regards,

Peter Horvath

Contents

1 Anonymous Referee #1

Received and published: 30 June 2020

General comments

The overall objective of this paper was to identify biases in a dynamic global vegetation model (DGVM) and, if possible, to find ways of reducing the biases. The analysis focused primarily on relatively undisturbed landscapes in Norway. The target model output was the within-gridcell plant functional type (PFT) distribution. One unique and valuable aspect of the manuscript was that the PFT distributions predicted by the DGVM were compared to multiple products, including field surveys, satellite products, and the output of species distribution models. Field surveys were much more similar to the satellite products and distribution models than to the DGVM. Improvement to the DGVM was realized by incorporation of a precipitation seasonality index, although it was clear that this improvement would not be the end of the story.

Given that PFT distribution is an important quantity that is still challenging for DGVMs to predict, I think that the manuscript covers a topic that will be interesting and useful to readers of Biogeosciences. I also appreciated how the DGVM was compared to multiple products and how the distribution model was leveraged. However, I think that the value of the manuscript could be increased by being more thorough with the methods (see below). Also, I think that more could be done to make the manuscript interesting to readers who use models other than CLM.

We are thankful to Referee #1 for his/her positive response and constructive comments.

Specific comments

RC-1.1 - The title should be modified. It mentions "Dynamic Global Vegetation Models" in the plural, but only one model is discussed. I also think the title is too general. I would suggestion "high-latitude vegetation distributions" rather than simply "high-latitude vegetation".

This is a good suggestion. We have adjusted the title to specify that high-latitude vegetation distributions are considered. With regard to the plural mention of DGVMs, we believe that even though we tested this particular exercise only on one DGVM (namely CLM4.5BGCDV), the procedures/methods of implementing variables from DM as new parameters in DGVM can be used in several DGVMs not just the tested one (thus the plural form).

RC-1.2 - Lines 83-84: This point is overstated. There are publications that have evaluated PFT distributions from dynamic vegetation models against field-based datasets, at least on regional and national scales.

In line with response to Referee #3 on this same point (see also comment RC-3.6), we have adjusted the formulation of the sentence and added a reference (line 95).

RC-1.3 - Methods: I am puzzled by the limitation of the study to only 20 plots. Certainly these 20 plots span the range of mean annual temperature and precipitation, but other factors are also commonly perceived to be important. Indeed, the distribution model seems to take 100+ inputs. Some questions that come to mind is whether the plots span the range of observed precipitation seasonality (identified by this study as an important factor!), soil texture, and soil nutrients.

We agree that a higher number of plots would have been beneficial. Ideally, we would want 1000+ plots or perhaps a regional/global simulation. However, labor-demanding preparation of all data layers for each *plot was one of the critical factors for this study and we had to find a compromise between what was practically possible and what was considered robust in terms of the aim of the study. From a methodological perspective, our opinion is clearly that a representative sample of 20 plots is sufficient to demonstrate the differences between the three methods of representing the vegetation distribution.*

The gradients of precipitation and temperature are known to be among the most influential for vegetation distribution (e.g., Ahti et al. 1968; Bakkestuen et al. 2008), thus we have chosen to include these particular two variables when selecting the 20 plots. However, we also agree with the Referee #1 in the argument that the 20 plots' representativity across the range of precipitation seasonality should be tested (since this is identified as an important factor). We have clarified this more thoroughly, included a comparative test and added a third diagram to the Supplementary Figure S2 (lines 145-161 -chapter 2.3 and Fig S2 lines 20-30 in the supplement). Please also see the response RC-2.6 to Referee #2 with a similar request.

RC-1.4 - Line 157: Why not assign the observed soil texture to the 20 plots?

The observed data on the 20 plots unfortunately do not include information about soil texture. The plots were mapped using wall-to-wall vegetation mapping, where only data about the type of vegetation cover are available.

RC-1.5 - Section 2.4.3: I am concerned that the DGVM and the DM uses different driver data to represent the same phenomenon. For example, does one use SeNorge2 and the other reanalysis to represent precipitation? Does one use observed soil texture and the other "default" soil texture? If so, might differences in inputs account for differences in the DGVM and DM predictions?

Absolutely. Ideally, we would use the same climate input data for both DM and DGVM. However, there are technical obstacles: DM uses multi-year monthly averaged climate data as input, while DGVM requires 3-hourly meteorological data as the input. SeNorge2 dataset, which is used in DM, has only daily data available, therefore can only be used for DM but not for driving DGVM. For DGVM, we had to use available reanalysis or regional climate model data for present day climate (CORDEX data in this manuscript). To compare the differences between the driving data for DGVM and DM, we have listed mean annual temperature and precipitation for both datasets in the Table S1 and Figure S4 of the supplement (lines 5-8 and 50-55). There are indeed some minor differences between the two sets of driving data, however it is beyond this study to quantify the effect of these differences. We have devoted a paragraph to clarify the potential bias this may imply in the discussion (lines 542 - 550).

Soil texture does not come in as an explanatory variable in the DM, whereas DGVM is using soil texture as an important parameter affecting various processes in soil, such as soil temperature, moisture and organic matter decomposition. We have added a comment on the differences between the input data in the paper and discuss its potential implications (lines 542 - 550).

RC-1.6 - Line 183: Was the DM model previously tuned to these 20 plots? To Norway?

The DM was not tuned specifically to these 20 plots. The training data for DM included the whole set of 1081 plots (across Norway) at a different thematic resolution (detailed vegetation types instead of PFTs) and at a scale of one point per polygon. Although the 20 plots were included as a subset of the total 1081 plots, we believe the influence is minimal, since they have gone through a spatial and thematic conversion. Moreover, the DM was evaluated with a completely independent dataset.

RC-1.7 - Line 414: Might phenology also be an issue? Further, what is the light compensation point of the PFTs? Perhaps the authors can use the light compensation point to directly evaluate the relative shade tolerance of the different PFTs.

Please also see comments to Referee #2 (RC-2.14) and Referee #3 (RC-3.26) regarding this paragraph in the discussion. Phenology is likely to be an issue, as evergreen plants seem to have advantage in competing with deciduous plants in general in the high-latitude region in the model. It is therefore suggested that stress for evergreen plants in winter and spring may not be well represented in the model to limit the growth of boreal NET in some regions. However, we admit that this issue is not well documented through our results and therefore have decided to remove this paragraph from the discussion.

RC-1.8 - Discussion: Are there lessons for people who use other models? The more the authors can draw out such lessons, the broader the audience this paper would appeal to. The TEM model, which has a more detailed representation of boreal PFT diversity than CLM, immediately comes to mind as one example.

Thanks for the suggestions. The present-day vegetation distribution outputs from dynamical vegetation models could more often be evaluated by use of multiple products complementing the RS, such as by including DM and AR as presented in this study. We also believe that the procedure of identifying new parameter values from DM, running a set of sensitivity tests and implementing the sensible new parameters into a DGVM is not limited to CLM4.5BGCDV (the DGVM tested here) but transferrable also to other DGVMs, such as the TEM model. We have clarified this and included more thorough discussion with regard to applicability to other models in the revised manuscript (lines 575-579, 614-619).

Technical corrections

RC-1.9 - The manuscript is very readable, but it should still be reviewed for grammar.

We have carefully searched the manuscript for grammatical errors and corrected where applicable.

RC-1.10 - Page 3, Lines 43-45: There is a problem with word choice in this sentence. Vegetation distributions are not implemented in ESMs, but rather are predicted by ESMs. The ESM predictions can then be evaluated with satellite products (as done in the present analysis).

We have rewritten the sentence according to the referee's comment (lines 50-51).

RC-1.11 - Section 2.4.1: It would be useful for the authors to briefly describe how the DGVM determines the amount of area to each PFT.

We have added a brief description on how the area of each PFT (i.e. percentage cover fraction %) is determined by DGVM in the revised manuscript (lines 212-217). The percentage cover fraction of each PFT is equal to the average individual's fraction projective cover (FPC_{ind}) multiplied by the number of individuals (N_{ind}) and average individual's crown area ($CROWN_{ind}$). FPC_{ind} is a function of the maximum leaf carbon achieved in a year, while $CROWN_{ind}$ is related to dead stem carbon simulated by the model. N_{ind} is mainly determined by establishment and survival rate controlled by establishment and survival threshold conditions.

RC-1.12 - Data availability: Note that the GitHub link not up yet. I understand if the authors do not want to release the link prior to manuscript acceptance, but it is still important not to forget to release the link.

This is an important point. We have made all the available data accessible on the following repositories (link to DGVM scripts: https://github.com/huitang-earth/Horvath_etal_BG2020*; link to script for analysis:* https://github.com/geco-nhm/DGVM_RS_DM_Norway*; and link to larger spatial data outputs from RS and DM on DRYAD:* https://doi.org/10.5061/dryad.dfn2z34xn*).*

2 Anonymous Referee #2

Received and published: 19 August 2020

General Comments

This study evaluates estimates of PFT distributions from a DGVM in comparison to those of remote sensing and empirical models, and against a field-based dataset, for 20 plots of high-latitude vegetation types across Norway. The topic investigated, approach taken, and results reported will be of interest to the modeling community. The paper could benefit from more or better explanation of the methods, especially the CLM simulations. For example, it is unclear whether or not this is intended to be any kind of 'temporally-explicit' analysis; this seems a sort of model estimation of some 'average' PFT distribution from the spin-up results that was compared to field plots and remote sensing data, both of which presumably represent a specific point in time (that is not specified in either case in the methods here).

Thank you for this to-the-point comment. We agree that more careful explanation of some aspects of the methods is necessary. We have adjusted the manuscript with respect to the specific comments you provided here.

This study represents a temporally explicit analysis of the 'present-day' vegetation distribution. We agree and have emphasized this more clearly. In line with further replies to RC-2.10, the temporal context has been specified for each of the three modelling methods as well as for the AR in the respective sub-chapter 2.4 (lines 102, 170, 206-207 and 226).

RC-2.1 - To properly interpret the results, the sensitivity tests need more explanation and clarification to justify and understand what was done here in this study (vs. previous work).

We have added a much more detailed explanation of the sensitivity tests in the revised manuscript (chapter 4 - lines 348-397). Also, we shall review the formulations of what was done in this study vs previous work.

RC-2.2 - The "RS method" as one of the three methods compared here seems kind of out of place in this analysis since it is not a method for predicting future PFT distributions as with the DGVM and DM methods. What is the reasoning / purpose behind including RS in this comparison? Or could / should it be used in this study more as a 'reference' data set, like the AR data?

We understand the concern of Referee #2 on this point. We also agree that RS is often being used as a verification/reference dataset in land surface modelling. However, the emphasis of this work is on improving the DGVM for the 'present-day', based on the premise that the better DGVM are able to predict the present-day distribution of vegetation (based on the processes/parameters driving the DGVM), the more reliable vegetation predictions for the future will the model be able to produce. Moreover, RS is also of interest from the perspective that products derived from RS data may also be burdened with uncertainties, needing evaluation - just as DM and DGVM - against a ground-truth/reference data set, which in this case is AR (see also our response to RC-3.5). We have devoted lines 79-84 to making this clearer in the revised version of the manuscript.

Specific Comments

RC-2.3 - 25-26. please consider this statement carefully; numerous authors could claim that this is untrue

Thank you for pointing this out. This comment accords with a comment of Referee #3 (RC-3.6) and we have modified this statement in the abstract of the revised manuscript (lines 25) as well as the introduction (lines 95-96) where the amended sentence is now supported by references (e.g., Druel et al 2017).

RC-2.4 - 34. can these three thresholds be named here, or at least hint at what they are (e.g. ". . . based on . . .)?

Yes, we agree that the thresholds should be mentioned here. Also, in line with another of your comments (RC-2.15), we have adjusted the text to clarify that only precipitation seasonality (bioclim_15) is influential (lines 36-41).

RC-2.5 - 115-116. this is not quite clear and perhaps needs to be specified or qualified; i.e. don't many "countries" have national-scale inventory programs?

This has been re-worded (lines 130-131). What is meant here is that wall-to-wall vegetation surveys on national scale are rarely made. AR (the reference dataset) is an example of an area-representative survey.

RC-2.6 - 126-131. Selecting only 20 plots seems limited, even if deemed acceptable for bioclimatic variation. There needs to be better explanation / justification for this choice, how "acceptable" was determined, and whether a kriging of temperature and precipitation really captures "bioclimatic" variation across the country.

We agree that a set of 20 plots is a rather limited number. Referee #1 raises the same issue (RC-1.3), and our response (and justification for the choice) is given in comments to Referee #1. We have amended the text to explain our choice better (section 2.3 - lines 145-161).

The representativeness was tested for and explained in supplements S2 and S3 (see also Fig.S2 and Table S3 – lines 17-49). By acceptable representativeness we mean that the selection of 20 plots does capture the variation across the whole range of temperature and precipitation (in the revised version we have also added "precipitation seasonality" - Fig.S2 – following comment RC-1.3) compared to the full set of 1081 AR plots. The representativeness of the 20 plots was also tested against the full dataset of 1081 AR plots with regard to PFT coverage, where a Chi-square test showed that the two datasets are much more similar than expected by chance (Supplement S3 – lines 35-49).

We have reformulated the sentence on line 131. Also, in line with the comment RC-2.1, we have clarified what was done in this study vs. previous studies. Kriging was used in a previous study to interpolate the original SeNorge2 dataset from 1km down to 100m for the purpose of distribution modelling (a procedure which was done and described in Horvath et al. 2019). We agree that this information is not relevant for the representativeness comparison, and it is more important to include a specific description of how the representativeness test was done in this study (in addition to the existing description in supplement S3). We have reformulated this paragraph and revised manuscript accordingly (section 2.3 - lines 143-161).

RC-2.7 - 150. curious decision to give a new acronym to CLM. why not just refer to it as "CLM"? and actually, you do, somewhat, as it seems to switch back-and-forth between "DGVM" and "CLM4.5" for the rest of the manuscript. I see the idea to associate the results from CLM as representative of the "DGVM" approach, but when describing or referring to the specifics of CLM then just call it "CLM" (or "CLM4.5")

We understand the confusion here. This has been clarified and we have explained the terms further in lines 180-182. CLM has an option to run will full vegetation dynamics (CLM4.5BGCDV), this option is further referred to as DGVM. The abbreviation of DGVM is used throughout the manuscript to refer to this particular setup of CLM. Consistency in the use of terms have be carefully checked.

RC-2.8 - 154. it may be useful here to point out what these simple assumptions are, and how different (or not) they are from those for which the DM method is based on.

We have added more details about the assumptions used in DGVM in describing establishment, survival, mortality and light competitions (lines 185-186). Compared to DM which uses statistical relationships (line 231-232) to predict the probability of VTs/PFTs from environmental variables, DGVM assume a simple environmental threshold for establishment, survival and mortality of a PFT to occur (see supplement S7).

RC-2.9 - 171. was soil C initialized somehow, or was it a separate (longer) spin-up? are these mostly undisturbed sites or was that taken into consideration for the vegetation spin-up at each site? was the CORDEX climate used for the spin-up? average or de-trended?

Thanks for pointing this out. In our experiments, soil C and N were firstly initialized using the restart file from an existing global present-day spin-up simulation with prescribed vegetation. Then, they were spun-up together with vegetation for 400 years. All the selected sites are mostly undisturbed. The 30-year CORDEX data were cycled during the spin-up. A 30-year period is consistent with WMO climatological normals based on the rational that 30 year is short enough to avoid large long-term trends while long enough to include the range of variability. Thus, the data were not de-trended or averaged. We noticed that vegetation distribution was insensitive to interannual variation or decadal variation of the climate forcing when it reached equilibrium state in most of our study sites (see supplement S10). This has now been specified in more detail in the manuscript (section 2.4.1 - lines 195-207).

RC-2.10 - 174. what year / era does this RS map represent? Table 2. I don't think all of this detail is necessary in the main text.

A very good point, which should be clarified indeed. The RS product used in this study is created from satellite images covering the period of 1999-2006 (Johansen, 2009). This has been clarified in the manuscript (line 226).

We agree that Table 2 might be too detailed for the main text. We have moved Table 2 into the supplement S5 (lines 60-64 in the supplement).

RC-2.11 - 278, 279 & 305 are confusing uses of sub-headings

We agree that further splitting the chapter 4 (Sensitivity experiments and model improvement) into methods and results might seem untraditional. We suppose that it has not been made clear that the paper falls into two parts: an analysis of data, and a sensitivity analysis which is based upon the results of the analysis. We have added a motivation sentence at the end of the introduction (line 106), clarifying that the sensitivity experiments are a separate chapter, which builds upon the results of the analyses. Chapter 4 describing the sensitivity experiments has remained, but the sub-headings have been removed and the text into separate paragraphs (lines 342-402) (see also reply to RC-3.23).

RC-2.12 - 287. swe_10 and tmin_5 make sense as described but can "precipitation seasonality" be explained? "bioclim_15" is not as obvious as the other two parameters

A very good point. We have now included a description and a reference to how "precipitation seasonality" is calculated (O'Donnell & Ignizio, 2012) on lines 357-359. "Precipitation seasonality" is defined as the ratio of the standard deviation of the monthly total precipitation to the mean monthly total precipitation (also known as the coefficient of variation) and is expressed as a percentage.

RC-2.13 - 293-299. there just seems like so much of the justification and explanation of decisions and approaches for the sensitivity test are glossed over here. For example, why are these particular parameters chosen, how was bioclim calculated, is the stepwise order important, what does it mean "three PFTs at the same time", how were the thresholds determined, etc etc. Perhaps a little more explanation than just "see Horvath et al 2019" (line 286) would be helpful.

We agree with the Referee #2. Since a lot of the sensitivity experiments are based on the results from the previous study by Horvath et al. 2019, referring to this article is necessary. However, we agree that explicitly describing the sensitivity experiments is important. We have now added more detailed explanation on the reasoning behind the set-up of the sensitivity experiments, including the specific topics that Referee #2 is pointing to in this comment (lines 342-402).

RC-2.14 - 414-415. this seems like a bit of a leap without a more direct connection to the results of this study.

We agree that the arguments in this paragraph are not supported by the results of this study. In line with the comments from Referee #3 (RC-3.26) and request from Referee #1 (RC-1.7) we have removed this argument from the revised version of the manuscript (lines 508-513).

RC-2.15 - 468. but in line 312 it was stated that two of those three "had little effect"

Yes, this must be a remnant of a previous formulation. We have removed the two parameters that did not improve the DGVM performance from this sentence (line 587). We have also amended lines 36-41 the abstract with regard to this (see also reply to a comment for RC-2.4 and RC-2.17).

RC-2.16 - 498-499. when are high-quality RS products ever not available anymore in this day-and-age?

We agree that this needs to be reformulated to explain the challenges clearly. It is not the "high-quality" of RS products in terms of resolution or coverage that we are concerned about, but rather in terms of being able to supply proxies of other properties (such as deriving parameter improvements, traits or in some cases vegetation distribution in high enough thematic resolution). In particular, at high latitudes low sun-angle results in large shadow effects. Furthermore, our results show that analyses of high spatial resolution RS images have limitations when it comes to thematic precision and resolution. We have now reformulated this sentence (lines 628-629).

RC-2.17 - 503. Just to be clear, it seems that these parameters were identified in a previous study, not this one, correct? And actually in this study only one of them (bioclim_15) was found to be useful, no? This same claim is made in the abstract, as well, and should be used with care.

Yes, we agree, and we have carefully re-formulated the sentences with this regard both in the conclusion and abstract. Please see also related comment RC-2.4 and RC-2.15.

Technical Corrections

RC-2.18 - - please review the grammar, wording and sentence structure throughout

All the technical and wording amendments suggested below have been implemented in the revised version of the manuscript. The text has been carefully searched and corrected for erroneous grammar.

42. please re-word and fix the grammar of this sentence one way or the other

55. remove “the” before DGVMs

60. latitudes

150. replace “further” with “hereafter”

170. “recalculated”

Table 2. “AR” is missing from the caption

292. change “NEB” to “NET”, I think

341. “spectre” should be “spectrum”?

412. “overprediction of Boreal NET”?

3 Anonymous Referee #3

Received and published: 25 August 2020

The manuscript "Improving the representation of high-latitude vegetation in Dynamic Global Vegetation Models" by Horvath et al analyses the performance of three different vegetation modeling approaches with regard to the spatial distribution and relative abundance of plant functional types (PFT) in Norway. The modeling approaches include a dynamic global vegetation model (DGVM), remote sensing (RM), and a statistical distribution model (DM), which relates occurrences of vegetation types to multiple environmental variables. The authors found that both RM and DM showed a better performance than the DGVM when compared to observational data from a range of field sites. They then tested if it was possible to use the DM to improve the predictions of the DGVM with regard to PFT composition and distribution. It was found that, through inclusion of three further bioclimatic constraints based on the analysis of the DM, the performance of the DGVM could be improved. The authors recommend DM as a complementary tool for the assessment and improvement of DGVMs.

RC-3.1 - The manuscript is well written and easy to understand in general. The research topic (assessing and improving DGVMs at high latitudes) is certainly relevant, and the chosen approach is original and seems useful to me. However, the description of the methods needs to be improved, with regard to the chosen statistical approaches, and also the motivation to carry out certain analyses. It often becomes clear only later in the manuscript why a certain method was applied. I therefore recommend minor revisions before a new version of the manuscript may be submitted.

We thank Referee#3 for a set of thorough comments. We have improved the sections of the manuscript in line with these comments.

Comments:

RC-3.2 - L 28 While the term 'DGVM' is explained at the beginning of the abstract, the term 'distribution model (DM)' is used in this sentence without previous explanation. Please explain shortly in the abstract what a DM is and how it differs from a DGVM, since some readers may not be familiar with the concept.

Good point. We have added a sentence about the difference between process based (DGVM) and correlative (DM) models (lines 29-31).

RC-3.3 - L 58 Please define or explain in more detail what you mean by 'thematic resolution'. Furthermore, it should be mentioned that recently, specific high-latitude PFTs, such as mosses, for instance, have been added to a number of DGVMs, e.g. Jules (Chadburn et al, 2015, The Cryosphere), JSBACH (Porada et al 2016, The Cryosphere), or ORCHIDEE (Druel et al 2017, Geoscientific Model Development) and several more.

The term thematic resolution is meant to refer to number of classes (ex. PFTs) in a model. This has now been explained in line 66. Thank you for pointing to these references, we have included them as examples in this paragraph (line 64).

RC-3.4 - L 60 Three examples are given for the difficulties of DGVMs to simulate extents of high-latitude PFTs correctly. However, I do not see how the underestimation of forest carbon storage by DGVMs relates to this, since this is rather a consequence, and not a reason for the incorrectly predicted extent. Please explain in more detail.

Good point. The sentence about carbon storage underestimation has been reformulated (line 68) to clarify that discrepancies in the DGVM have implications on different systems (e.g. carbon storage).

RC-3.5 - L 71 Please add a short statement to describe in which regard the RS products are not consistent.

The study by Myers-Smith et al. (2011) reports a mismatch in the spatial resolution between satellite observations and the spatial heterogeneity of vegetation patches in tundra ecosystems. This will be clarified in the introduction. Also, different satellite products produce varying results with regard to vegetation classification (Majasalmi, T. et al. 2018). We have devoted lines 80-81 to describing these inconsistencies in the manuscript (please, also see RC-2.2).

RC-3.6 - L 83 At least one study (Druel et al 2017, Geoscientific Model Development), uses site data to assess the DGVM's performance with regard to plant traits. Please be more specific in this regard, and explain what exactly is new in the validation method.

Yes, we have reformulated this sentence to make clear that our study focuses on evaluation of vegetation distributions between different models/methods (lines 94-99). Also, we mention the study by Druel et al. (2017) as an example of evaluation with field data.

RC-3.7 - L 121 I do not understand this sentence: If one plot is 0.9 km2 large, then 1081 plots are around 1000 km2, but 18x18 km are only 324 km2. Also, the plots are distributed throughout Norway, so the 18x18 km area has to mean something else. Is it the distance between the plots on a grid which covers Norway? Please explain.

Thank you for pointing this out! For us, having worked with these data for so long time, it is easy to forget that it is not obvious how they are structured! There is a regular grid covering the whole land area of

Norway on which the plots (in total 1081 plots), each with a size of 0.9km2, is placed every 18 km (in latitude) by 18 km (in longitude). This has now been explained in more details in lines 136-139.

RC-3.8 - L 129 To me it seems that low values of temperature and precipitation are underrepresented in the 20 selected plots compared to the full data set. This should be mentioned here briefly and then considered later in the Discussion section.

We agree that there is a slight underrepresentation in the frequency of plots with the lower values for temperature and precipitation. However, the most important factor was to include plots covering the range of the temperature and precipitation values experienced, which we have succeeded in (Fig S3). We have added a brief description in lines 155-156.

RC-3.9 - L 156ff By using the default surface parameter values for CLM, the DGVM may miss some relevant information to correctly predict PFT distribution, compared to RS and DM. Furthermore, by using climate forcing from 1980-2010 and running the DGVM into a steady state with regard to this period, historical climatic effects, which may influence today's PFT distribution are not considered. These points should be mentioned in the Discussion section of the manuscript.

We understand the concern of the Referee #3 regarding this aspect. In line with replies to the RC-1.5 we have added a more detailed discussion on the issues raised in this comment (lines 542-550). As to the concern on the usage of the climate forcing data, we indeed overlooked the historical climate effects on vegetation distribution, which usually lag several years or decades behind climate changes. However, this is considered to have minor impacts on the large biases observed in DGVM (e.g., too much boreal NET and too few shrubs), even though historical climate effects (such as cooler temperature in the past) might favor more boreal shrub than boreal NET (please, also see our reasoning to comment RC-2.9). We have devoted a paragraph to clarify this in the Discussion (lines 536-542).

RC-3.10 - L 162 Why was the CORDEX data not also used for the DM method? This should be briefly mentioned here.

In a previous study (Horvath et al. 2019) the authors have created distribution models for vegetation types with a range of predictors (including SeNorgre2 data), where the statistically important predictors were selected in a forward selection procedure. At that point the SeNorge2 was the most reliable climate dataset available for the whole study area. We have now added a comment on the choice of climate data sets in DM in the section 2.4.3 (lines 235-237). Also see the paragraph in discussion on lines 536-642.

RC-3.11 - L 175 Please explain 'supervised' and 'unsupervised' in more detail.

While in supervised classification, training data are based on well labeled data from part of the study area, unsupervised classification is only supplied with the number of output classes. 'Supervised' and 'unsupervised' classification methods are now shortly explained on lines 226-228.

RC-3.12 - L 182 the number of explanatory variables (116) is rather high. It should be shortly explained what these are, and why such a large number is necessary for the regression. Even if this information is provided in Horvath et al (2019), it should be summarized here.

We have added a short description of the explanatory variables (grouped into categories) on lines 232-233. Also, a sentence about forward variable selection procedure has been added, to make clear that only a few of the 116 variables were actually included in each final DM (lines 233-237).

RC-3.13 - L 183 It would be good to add a short summary of the evaluation method for the DM here, so the reader can assess the DM better.

We have now added a short summary of the evaluation procedure on lines 237-242. Evaluation of each model was carried out using an independent evaluation data set and by calculating the area under the receiver operator curve (AUC), a threshold-independent measure of model performance commonly used in Distribution modelling. AUC can be interpreted as the probability that the model predicts a higher suitability value for a random presence grid cell than for a random absence grid cell (Fielding & Bell, 1997).

RC-3.14 - L 186 I wonder if, by discarding all other VT except the most probable one, biases in the distribution of the VTs are introduced. Let us assume the logistic regression predicts a certain VT always with a slightly higher probability than a second one; according to the description, only the first VT would occur in the predicted map at all pixels, and all observations of the second one would be discarded, although this VT occurs quite frequently in reality. Please explain this in more detail.

This is an interesting and intriguing topic. As the Referee #3 rightfully points out, there is a possibility of slight biases in certain regions, for the reason outlined. However, as far as we are aware, this has not yet been closely investigated. We are preparing a manuscript covering this topic in more detail - The results so far suggest that the approach for compiling the wall-to-wall map from 31 DMs, which we also use here, is performing the best out of the tested approaches (Horvath et al., manuscript in prep.). Additionally, as the probability of presence for each VT is predicted separately for each grid-cell, the probability values for every VT varies independently of the probabilities for the other VTs, throughout the study area. Thus, we regard the chance that one VT consistently outperforms another VT over all the grid cells to be negligible. We have now explained this more carefully in the discussion (lines 477-482).

RC-3.15 - L 200 I don't understand why an aggregated PFT profile is needed, I thought that the comparison of the 3 modeling approaches and the AR data is done for each of the 20 plots?

Indeed, the main comparison is between the 3 modelling approaches and AR on each of the 20 plots (this can be found in figure 2 and 3). But besides, it was also worth investigating the overall performance of the three methods across the study area. In order to do that, we needed the aggregated PFT profiles. We have now clarified this in the sentence (line 259).

RC-3.16 - L 208ff This sounds like one comparison was done with the aggregated profiles (one for each method, aggregated over all 20 plots), using the chi-square test. Then, for each of the 20 plots the profiles were compared regarding their dissimilarity. It is not clear to me, why two different statistical methods were used to compare the models (DM, RS, DGVM) to AR.

The point here is that we wanted to compare the three models (DM, RS, DGVM) to AR both with respect to the overall pattern (represented by the aggregated profiles) and with respect to their performance on each plot; the latter in order to identify the circumstances under which some of the models deviated strongly from the reference. Accordingly, the chi-square test was used to formally test if the models overall deviated from the reference, while the proportional dissimilarity index (which does not come with a statistical test) was calculated to address the purpose of identifying strongly deviating modelling results at plot scale. This is now clarified in lines 267-271.

RC-3.17 - L 222 I thought the dissimilarity index was used to assess the similarity between the 3 modeling approaches and the AR data. Why is it then necessary to do a pairwise Wilcoxon-Mann-Whitney test in addition? Please explain the reasons for the chosen statistical approach in a more detailed way.

Our statistical analyses serve several purposes of which one is to assess the goodness-of-fit of the modeling results to the reference (I.e., to assess their performance); another (which is addressed by the Wilcoxon-Mann-Whitney tests) is to assess the degree to which the models produce pairwise similar differences. We have added a sentence to explain this in the paragraph (lines 283-284).

RC-3.18 - L 230ff As mentioned above (L200), by aggregating the PFT profiles of the 20 plots, differences in profiles between plots are lost. Hence, it is not possible to evaluate the 3 models with respect to the correct prediction of differences in profiles between individual plots. Also, while the AR data (for each plot) can be interpreted as a random sample, it is not clear to me how the model approaches can be consistently included in this Chi-square test. Moreover, the number of elements (6 PFTs) is actually too small for a Chi-square test. The authors need to justify this better, or change their testing approach.

The mere purpose of analyzing the aggregated profiles is to assess the models' ability to produce overall predictions of PFTs that accord with the PFTs' overall frequency (as given by the reference). We do not see any reason why the chi-square test should not be useful for a contingency table of 6 classes.

RC-3.19 - L 249 If I understand Fig. 2 correctly, the lines which connect the dots denote the individual plots, which means that for one method (e.g. DGVM), the dissimilarity can be high (1.0), while for another method (e.g. RS) it can be much lower. The result that the goodness of the fit between a given method and AR data depends on the set of chosen plots may point to some underlying systematic deficiencies of each method and should be discussed later.

Exactly as you describe, the values of dissimilarity index portrayed as dots connected by lines in Fig.2 represent the similarity of each plot between a particular method and the reference dataset AR for that plot. While the individual dissimilarities may be high, we have good reasons to believe that the selection of 20 plots is sufficiently representative for the study area that the major patterns emerging from the analyses reflect real major patterns. Furthermore, you are right that systematic deficiencies in some of the methods are reflected in the single-plot patterns shown in Fig. 2. Some of these topics were discussed in the previous version of our manuscript and we have now expanded on some of the aspects (sections 5.1.1., 5.1.2., 5.1.4., and section 5.1.5 of the discussion chapter).

RC-3.20 - L 252 The statement in this sentence is not evident to me in Fig. 3, because this figure simply shows the profiles for each plot (which is a good way of illustrating the results, in my opinion). Wrong reference?

Absolutely. This typo has been corrected to Fig.2.

RC-3.21 - L 254 Please see also my comment to L 222; I assume that the authors use the Wilcoxon test to assess if the median values of the dissimilarity indices for the 3 models are significantly different from each other. However, I think it is more relevant how the models differ to each other with regard to the AR data. This information is contained in the values of the dissimilarity index, and it should be reported more clearly here. The pairwise comparison of the 3 models seems to me of secondary importance to assess the goodness of the fit to AR data.

This is correct. The core result we report in this paragraph is the dissimilarity between the methods and the reference dataset. This is reported on lines 311-315 "While RS had the lowest median proportional dissimilarity with the AR reference (0.19, compared to 0.26 for DM and 0.41 for DGVM), …".

The pairwise comparison results of the Wilcoxon rank-sum tests are mentioned only after the core findings to support the similarity between RS and DM at most plots (lines 315-319).

RC-3.22 - L 262ff The visual comparison of the 3 models in Fig.3 and the associated description is more helpful to assess the modeling approaches than the statistical methods described before.

The paragraph on lines 325-336 summarizes the visual inspection of Fig. 3 in terms of performance of the three methods and describes the regional deviations of DGVM and DM from the reference. The issues are further discussed on lines 446-471 and in paragraph 5.1.5.

RC-3.23 - L 279ff This belongs into the Methods section. Explaining the sensitivity analysis earlier also makes it much easier to understand the goal of the overall approach.

We agree. Please see also our reply to RC-2.11. We have added a clarification in the introduction (line 106), that the sensitivity experiments are described in a separate chapter, which builds upon the results of the analyses. We have deleted the subheadings 4.1 Methods and 4.2. Results to avoid confusion.

RC-3.24 - L 287 The term 'precipitation seasonality' should be better described, in particular since it is found later that it is important to improve DGVM parameterization.

Please see also our reply to RC-2.12. "Precipitation seasonality" is defined as the ratio of the standard deviation of the monthly total precipitation to the mean monthly total precipitation (also known as the coefficient of variation) and is expressed as a percentage. This has now been explained on lines 357 -359.

RC-3.25 - L 379ff The point about 'good' and 'poor' DMs is not clear to me. Why should poor DMs be used at all? Please explain, and also consider my comment above (L 186).

The terms 'good' and 'poor' refer to the predictive performance of the individual DMs (i.e. AUC - see also reply to comment RC-3.13). The study by Horvath et al. (2019) provides predictions of the distribution of a total of 31 vegetation types across the study area of Norway (with AUC values ranging from 0.671 to 0.989). Reasons for the low predictive performance of some DM may vary, but in this case is most likely caused by missing important predictors. The set of predictor variables used in the study (n=116) might seem excessive, but nevertheless the authors conclude that several important factors are not represented among these 116 (soil nutrients, NDVI, LiDAR etc.). The reason for this is that variables representing these factors were not available in the required formats/resolution/coverage at the time-point the study was carried out; a general problem in distribution modelling. By using the chosen set of predictor variables, statistical approach and settings, the authors obtained the best possible distribution models, even though with regard to the AUC values, some might be considered weak/poor. The direct answer to the comment is that the DM method requires estimates for the probabilities of occurrence for (almost) all vegetation types to create a seamless vegetation map, which in turn is required for making estimates for the PFT profiles as robust as possible. Thus, in this context, 'poor' models are better than no model. We have devoted the paragraph on (lines 455-465) to making this (important) point more clear.

RC-3.26 - L 411 It may not be clear to readers why the lack of a shade-intolerant birch-PFT in the DGVM leads to the over-representation of NET in plots 17 and 18. The birch-PFT should rather have an advantage in mountainous regions compared to NET, which is currently lacking in DGVMs. Please clarify.

Please see also our reply to RC-1.7 and RC-2.14. We agree with the Referee #3 that this argument is not clear and without a clear support from our results. We have decided to remove the argument from the revised manuscript.

RC-3.27 - L 450 Please check the literature for the recent progress in including high-latitude vegetation types into the PFT scheme of DGMVs, and add this to the discussion.

We have added recent studies about this topic in the discussion (lines 572-575). See also our reply to RC-1.8.

RC-3.28 - L 467 This sentence is hard to understand, please reformulate.

Yes, this has been reformulated.

RC-3.29 - L 475 It should be mentioned if increased seasonality promotes or impedes growth of NET.

Thanks for pointing this out. By applying the new threshold, the growth of NET is impeded if the value for precipitation seasonality is larger than 50 (Table 4, Supplement S6 and S11). This is now mentioned in the lines 597-598.

Supplement:

RC-3.30 - L 40 missing reference L 51 missing reference L 52 missing reference

Thanks for pointing this out. This is a remnant of splitting the document into manuscript and supplement. All the references are now fixed.

RC-3.31 - L 55 The PFTs for this study are not in bold font, but shaded grey, please make this consistent.

This has been fixed.

RC-3.32 - L 56 The caption of Tab. S6 should be a bit more detailed: Is zbot the bottom height of the canopy (11.5 m above ground)? How is the coefficient of variation in precipitation seasonality computed?

We have adjusted the caption to clarify all the mentioned abbreviations.

RC-3.33 - L 90 The cover fractions in plots 801,2108,4268 are clearly not in a steady state. Please check if this significantly affects the results (e.g. by extrapolating the trends in cover), and repeat the DGVM runs, if necessary.

Thanks for pointing this out. We have now extended the running time of our simulations for these three plots by 400, 200 and 200 years respectively to check the vegetation distribution at the equilibrium state. We found that the average of the last 20 years at the end of each simulation does not deviate substantially from our previous results at the end of 400 years (see the new plots added in the supplement S11.2). We have therefore decided to consistently use the original 400 year spin-up data for the analysis for all 20 plots. This is also clarified in lines 204-207).

RC-3.34 - L 122 missing reference

Comments on style:

All the following comments on style have been implemented in the revised version of the manuscript.

L 42 I think ’an’ is not needed here.

L 55 ’DGVMs’ instead of ’the DGVMs’

L 60 ’at high latitudes’ instead of ’in the high latitude’

L 66 ’in’ not necessary

L 138 the second "of the" is not necessary

L 373 add ’the’ before ’reason’

L 401 ’differ’ instead of ’differs’

4 REFERENCES:

Ahti, T., Hämet-Ahti, L. & Jalas, J. 1968. Vegetation zones and their sections in northwestern Europe. – Annls bot. fenn. 5: 169-211.

Bakkestuen, V., Erikstad, L. & Halvorsen, R. 2008. Step-less models for regional environmental variation in Norway. – J. Biogeogr. 35: 1906-1922.

Fielding, A. H., & Bell, J. F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation, 24(1), 38–49.

Horvath, P., Halvorsen, R., Stordal, F., Tallaksen, L. M., Tang, H., and Bryn, A.: Distribution modelling of vegetation types based on area frame survey data, Applied Vegetation Science, 22, 547-560,

Johansen, B. E.: Satellittbasert vegetasjonskartlegging for Norge, Direktoratet for Naturforvaltning, Norsk Romsenter, 2009.

Majasalmi, T., Eisner, S., Astrup, R., Fridman, J., and Bright, R. M.: An enhanced forest classification scheme for modeling vegetation–climate interactions based on national forest inventory data, Biogeosciences, 15, 399–412,

O’Donnell, M.S., and Ignizio, D.A., 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691, 10 p

Improving the representation of high-latitude vegetation distribution in Dynamic Global Vegetation Models

Peter Horvath[1, 4], Hui Tang[1,2, 4], Rune Halvorsen[1], Frode Stordal[2, 4], Lena Merete Tallaksen[4, 5], Terje Koren Berntsen[2, 4], Anders Bryn[1, 3, 4]

[1] Geo-Ecology Research Group, Natural History Museum, University of Oslo, P.O. Box 1172, Blindern NO-0318 Oslo, Norway
[2] Section of Meteorology and Oceanography, Department of Geosciences, University of Oslo, Norway
[3] Division of Survey and Statistics, Norwegian Institute of Bioeconomy Research, P.O. Box 115, NO-1431 Ås, Norway
[4] LATICE Research Group, Department of Geosciences, University of Oslo, Norway
[5] Section of Physical geography and Hydrology, Department of Geosciences, University of Oslo, Norway

Correspondence to: Horvath, P. (peter.horvath@nhm.uio.no)

Keywords: Area frame survey, Community Land Model, CLM4.5BGCDV, Distribution model, Earth System Model, Plant functional types, Remote sensing, Vegetation types,

Abstract. Vegetation is an important component in global ecosystems, affecting the physical, hydrological and biogeochemical properties of the land surface. Accordingly, the way vegetation is parameterised strongly influences predictions of future climate by Earth system models. To capture future spatial and temporal changes in vegetation cover and its feedbacks to the climate system, dynamic global vegetation models (DGVM) are included as important components of land surface models. Variation in the predicted vegetation cover from DGVMs therefore has large impacts on modelled radiative and non-radiative properties, especially over high-latitude regions. DGVMs are mostly evaluated by remotely sensed products, ~~but rarely~~less often by other vegetation products or by in-situ field observations. In this study, we evaluate the performance of three methods for spatial representation of present-day vegetation cover with respect to prediction of plant functional type (PFT) profiles – one based upon distribution models (DM), one that uses a remote sensing (RS) dataset and a DGVM (CLM4.5BGCDV). While DGVMs predicts PFT profiles based on ~~are~~ physiological and ecological processes, DM relies on ~~a~~ statistical correlations between a set of predictors and the modelled target, and the RS dataset is based on classification of spectral reflectance patterns of satellite images. PFT profiles obtained from an independently collected field-based vegetation data~~-~~set from Norway were used for the evaluation. We found that RS-based PFT profiles matched the reference dataset best, closely followed by DM, whereas predictions from DGVM often deviated strongly from the reference. DGVM predictions overestimated the area covered by boreal needleleaf evergreen trees and bare ground at the expense of boreal broadleaf deciduous trees and shrubs. Based on environmental predictors identified by DM as important, three new environmental variables (e.g. minimum temperature in May, snow water equivalent in October and precipitation seasonality) were selected as the threshold for the establishment of these high-latitude PFTs. We performed a series of sensitivity experiments to ~~demonstrate~~investigate ~~whether~~if ~~that~~these thresholds improve the performance of the DGVM. Based on our results, we suggest implementation of one of these~~three~~ novel PFT-specific thresholds (i.e., precipitation seasonality) ~~for establishment~~ in the DGVM. ~~We performed a series of sensitivity experiments to demonstrate that these thresholds improve the performance of the DGVM.~~ The results highlight the potential of using PFT-specific thresholds obtained by DM in development ~~and benchmarking~~ ~~of also other~~ of DGVMs ~~for~~in broader regions. Also, we emphasize the potential of establishing DM as a reliable method for providing PFT distributions for evaluation of DGVMs alongside RS.

1 Introduction

Vegetation plays an important role in the climate system, as changes in the vegetation cover alter the biogeophysical and biogeochemical properties of the land surface (Davin and de Noblet-Ducoudré, 2010; Duveiller et al., 2018). Therefore ~~an~~ accurate descriptions of the vegetation distribution hold a key role in Earth system models (ESM) (Bonan, 2016; Poulter et al., 2015). Historical and present vegetation distributions ~~are~~ can be ~~implemented~~ prescribed in ESMs by means of datasets prepared from ~~satellite~~ observations (Lawrence and Chase, 2007; Li et al., 2018; Lawrence et al., 2011). However, in order to predict the future temporal and spatial changes in natural vegetation cover and subsequently the processes, dynamics and feedbacks to the climate system, dynamic global vegetation models (DGVMs) are needed.

DGVMs have been implemented as components of ESMs (Bonan et al., 2003) to represent long-term vegetation changes by a set of parameterizations describing general physiological principles, including ecological disturbances, successions (Seo and Kim, 2019) and species interactions (Scheiter et al., 2013). DGVMs represent the heterogeneity of land surface processes and interactions with other components of the Earth system by characterising land areas by their composition of type units defined by plant functional types (PFTs) (Bonan et al., 2003; Oleson et al., 2013). PFTs are groupings of plant species with similar eco-physiological properties – which express differences in growth form (woody vs herbaceous), leaf longevity (deciduous vs evergreen) and photosynthetic pathway (C3 and C4) (Wullschleger et al., 2014). Even though ~~the~~ DGVMs are being constantly developed and improved to incorporate more complex plant processes (Fisher et al., 2010), and more PFTs (Chadburn et al., 2015; Porada et al., 2016; Druel et al., 2017), there are still fundamental challenges for DGVMs to correctly simulate the extents of ~~the~~ PFTs that characterise boreal and Arctic ecoregions (Gotangco Castillo et al., 2012). For instance, the thematic resolution (i.e. the number of classes or PFTs in a model) of high-latitude PFTs is still limited (Wullschleger et al., 2014), important interactions between vegetation and fire ~~in~~ at high latitudes are still missing (Seo and Kim, 2019) which in turn has implications on the ~~, and~~ forest carbon storage in ~~the~~ high latitudes ~~is~~ still being underestimated by most DGVMs (Song et al., 2013). The large uncertainties in simulating high-latitude PFT distributions may also lead to discrepancies between modelled and observed energy fluxes and hydrology (Hartley et al., 2017), ~~or~~carbon cycles (Sitch et al., 2008) or surface albedo (Shi et al., 2018). Accordingly, systematic evaluation of PFT distributions modelled by DGVMs is required to improve the DGVMs and, subsequently, to reduce uncertainties in estimates of climate sensitivity and in predictions by ESMs.

Remote sensing (RS) is often used for evaluation, benchmarking and improvement of parameters ~~in~~ of DGVMs (Zhu et al., 2018). RS products are commonly used to describe vegetation cover using vegetation classes derived from multispectral images based on vegetation indices, such as the normalized difference vegetation index (NDVI) (Xie et al., 2008; Franklin and Wulder, 2002). For evaluation, RS products are translated into distributions of the PFT classes used in the DGVMs (Lawrence and Chase, 2007; Poulter et al., 2011). However, inconsistencies between various available RS-based land cover or vegetation products ~~have been reported~~(Majasalmi et al., 2018) as well as mismatch between the spatial resolution in RS observations and the spatial heterogeneity of vegetation patches (Myers-Smith et al., 2011; Lantz et al., 2010)~~–~~have been reported. ~~Therefore~~The fact that~~–and~~ benchmarking DGVMs only to these RS-based products may ~~therefore~~ lead to different conclusions in ESMs (Poulter et al., 2015), motivates for exploring ~~and~~ other vegetation products ~~are worth exploring~~ as a ~~to~~ supplement to RS.

Among the less explored methods to generate wall-to-wall vegetation cover predictions is distribution modelling. Distribution models (DMs) are most often used to predict the distribution of a target, by establishment of statistical relationship between the target (response) and the environment (predictors) (e.g. Halvorsen, 2012). The most common use of DM in ecology is for prediction of species distributions (Henderson et al., 2014), but DM methods have proved valuable also for prediction of targets at higher levels of bio-, geo- or eco-diversity (i.e. vegetation types and land-cover types) (Ullerud et al., 2016; Horvath et al., 2019; Simensen et al., 2020). DM methods are inherently static, in contrast to the dynamic DGVMs (Snell et al., 2014). Nevertheless, they may be a useful corrective to DGVMs by providing insights into important environmental factors driving the distribution of individual targets, which may, in turn, improve PFT parameter~~ settings~~ization in DGVMs.

Comparative studies that evaluate the present-day PFT distributions of DGVMs in a systematic manner, with reference to a field-based evaluation dataset~~, are~~are, with ~~apart ~~some exceptions ~~so far lacking~~ (Druel et al., 2017), ~~scarce~~few. In this study, we evaluate ~~representations of ~~vegetation distribution, translated to PFT profiles, obtained by ~~the ~~three different methods (DGVM, RS, DM) and ~~. We ~~use an independently collected field-based dataset of vegetation distribution, ~~(~~AR~~; –~~, (the Norwegian National map series for Area Resources) for the evaluation. Furthermore, we explore if environmental correlates of vegetation-type distributions identified by DM can be used to improve DGVMs by adjusting parameter settings for high-latitude PFTs.

To approach these aims, we constructed a conversion scheme to harmonize the classification schemes of RS, DM and AR into the PFTs used by the DGVM. We represent the present-day vegetation coverage by using plant functional type profiles (PFT profiles), vectors of relative abundances of PFTs within an area, e.g. a given study plot, summing to ~~1~~one. We then compare the PFT profiles obtained by DGVM, RS and DM with the AR reference on 20 selected study plots across the Norwegian mainland. Finally, we conduct a series of sensitivity experiments (ref. chapter 4) which build~~s~~ upon the results ~~from ~~of the analyse~~i~~s performed ~~of~~in this study to explore if the DGVM performance can be improved by adjusting DGVM parameters for selected environmental drivers ~~sourced~~identified by ~~ from ~~DM.

2 Methods

2.1 Study area – Norway

The study area covers mainland Norway, spanning latitudes from 57°57'N to 71°11'N and longitudes from 4°29'E to 31°10'E. Norway is characterized by a gradient from a rugged terrain with deep valleys and fjords in the western, oceanic parts to gently undulating hills and shallow valleys in the central and eastern, more continental parts. Temperature and precipitation show considerable variation with latitude, distance from the coast and altitude (Førland, 1979). While the mean annual precipitation ranges from 278 mm in the central inland of S Norway to more than 5000 mm in mid-fjord regions along the western coast, the yearly mean temperature ranges from 7°C in the southwestern lowlands to –4°C in the high mountains (Hanssen-Bauer et al., 2017).

The vegetation of Norway is structured along two main bioclimatic gradients (Fig. 1); one related to temperature/growing-season length and one to humidity/oceanity (Bakkestuen et al., 2008). Broadleaf deciduous forests, regularly found in the southern and southwestern parts (the boreonemoral bioclimatic zone), are further west and north (in the southern boreal zone) restricted to locally warm sites (Moen, 1999). With declining temperatures northwards and towards higher altitudes, ~~i.e. in the southern and middle boreal zones, ~~evergreen coniferous boreal forests dominate in the southern and middle boreal zones~~,~~. In the northern boreal zone the coniferous boreal forests ~~they~~ pass gradually into subalpine birch forests, which form the tree line in Norway. A total of about 38% of mainland Norway is covered by forests, and about 37% of the land is situated above the forest line (of which two thirds is covered by alpine mountain heaths). Wetlands cover approximately 9% and broadleaf deciduous forests about 0.4% of the land area (Bryn et al., 2018).

2.2 The AR reference dataset

Data obtained by in-situ field mapping, which is considered among the most reliable sources of land-cover information (Alexander and Millington, 2000), is practically and economically impossible to obtain in a wall-to-wall format for large land areas such as countries (Ullerud et al., 2020). As an alternative, ~~A~~area-frame surveys based upon stratified statistical sampling may~~, however,~~ provide accurate, area-representative, homogeneous and unbiased land-cover and land-use data for large areas. To evaluate the three methods for representing vegetation addressed in this study, we used the 'Norwegian land cover and land resource survey of the outfields' (*Arealregnskap for utmark*) dataset (Strand, 2013), a Norwegian implementation of the mapping program LUCAS (Eurostat, 2003). Data were collected in the period between 2004–2014 in a systematic regular grid covering the whole land area of Norway on which the plots (in total 1081 plots, each 0.6×1.5 km, i.e. 0.9 km^2) were placed every 18 km (in latitude) by 18 km (in longitude)~~18×18 km grid of 1081 rectangular plots (each 0.6×1.5 km, i.e. 0.9 km~~^{2}~~)~~ (Bryn et al., 2018; Strand, 2013). In each plot, expert field surveyors performed land-cover mapping by use of a system with 57 land-cover and vegetation-type classes (Bryn et al., 2018), mapped at a scale of 1:25 000. The data were provided in vector format with vegetation-type attributes assigned to each mapped polygon.

2.3 Study plots

Twenty out of the 1081 rectangular AR plots were selected to make up our reference dataset, AR (Fig. 1; center coordinates in Table S1). The AR plots spanned elevations from 88 to 1670 m a.s.l., with mean annual temperatures between –~~-~~4.0°C and 7.1°C and mean annual precipitation between 466 and 2661 mm (Table S1). ~~A test showed that the selectionwere acceptable representative for bioclimatic variation in Norway (see Fig. S3 and Fig. S4).~~ The gradients of precipitation and temperature are known to be among the most influential for vegetation distribution (e.g., Ahti et al. 1968; Bakkestuen et al. 2008). ~~The~~A series of Kolmogorov-Smirnov tests for ~~(~~comparison of sample mean and variance for these two variables~~,~~ using ~~gridded temperature and precipitation~~ data from seNorge2 (Lussana et al., 2018a; Lussana et al., 2018b)~~for these two variables~~ were obtained to investigate ~~whether~~ if the ~~selection of~~ 20 selected plots ~~does~~ capture the variation across in ~~the whole range of~~ temperature and precipitation in Norway-(Fig. S2), acceptable well -compared to the full set of 1081 AR plots~~, interpolated for each plot by kriging in accordance with Horvath et al. (2019)~~. Additionally, we ~~have~~ tested the representativeness across the range of variation for a third variable (precipitation seasonality) which was later selected for sensitivity experiments (see further section 4). While low values of temperature and precipitation are slightly underrepresented in the 20 plots, the total range of v~~lues is~~ariation was well covered. ~~All~~None of the ~~the~~ tests for temperature, precipitation and ~~for~~ the additional variable ~~of~~ (precipitation seasonality) ~~do not~~ indicate that ~~significant difference of~~ the sample of the 20 plots deviate from the full set of 1081 plots. The representativeness of the 20 plots was also tested against the full dataset of 1081 AR plots with regard to PFT coverage (Supplement

S3, Table S3), using ~~, where a~~ a Chi-square test. This test showed that the two datasets are not more dissimilar ~~much more similar~~ than expected by chance.

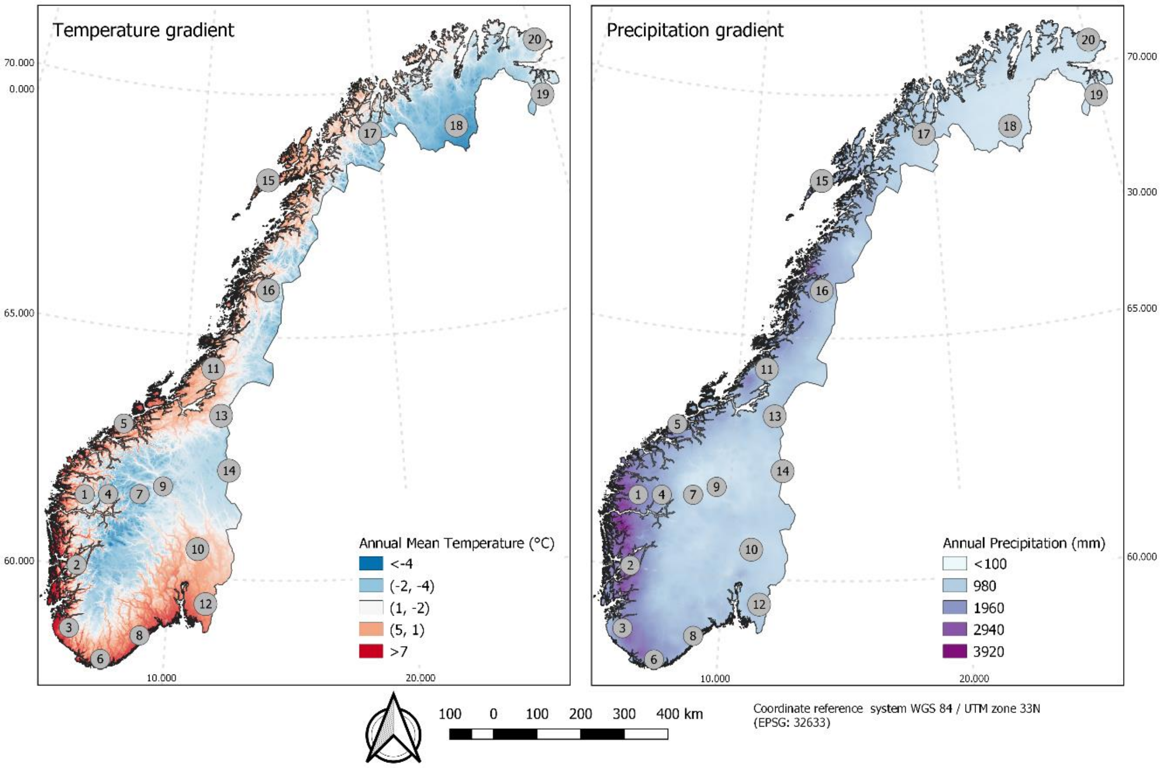


Figure 1 - Locations of the 20 plots across the two main bioclimatic gradients in the study area: temperature (left) and precipitation (right). The plots are numbered by longitude from west to east. Exact values of temperature, precipitation and altitude for each plot are given in Table S1.

2.4 Methods for representing vegetation

In this study, we use 'plot' as a collective term for two partly overlapping spatial units: (i) the 0.9-km^2 rectangles of the AR ~~of the~~ reference dataset; and (ii) the 1-km^2 quadrats with the same centerpoint as, and edges parallel to those of, the AR rectangles. The latter were used for the three methods of DGVM, RS and DM (Fig. S2).

Representations of the present-day vegetation ~~of~~ for each of these 20 plots were obtained by three different methods: (i) as the result of single-cell DGVM simulations for each plot; (ii) inferred from an RS vegetation map of the study area; and (iii) from vegetation-type DM models (Table 1). In order to make the three methods comparable, vegetation was represented by plant functional type profiles (PFT profiles), obtained by a conversion scheme (Table S5~~2~~ and Sect. 2.5). We define a PFT profile as a thematic representation of the land surface in a given plot or a group of plots, described as a vector of relative PFT abundances, i.e. values that sum up to 1.

Table 1 – Details of each of the ~~presented~~ methods for representing vegetation. DGVM – dynamic global vegetation model, RS – remote sensing, DM – distribution model. PFT – plant functional type, VT – vegetation type.

	DGVM	RS	DM
Model type	Process-based mechanistic model	Supervised and unsupervised classification	Statistical model
Software / model name and version	Community Land Model 4.5 – CLM4.5-BGCDV	ENVI (image analysis) and ArcGIS (classification)	R version 3.6.2, generalized linear model
Reference	Oleson et al., 2013	Johansen, 2009	Horvath et al., 2019
Thematic resolution	14 PFTs	25 VTs	31 VTs
Spatial resolution (grid cell)	1 km	30 m	100 m

2.4.1 The DGVM method

The DGVM employed in this study was the CLM4.5BGCDV (~~further~~ hereafter referred to as DGVM), ~~embedded~~ an option provided in NCAR's Community Land Model version 4.5 (CLM4.5) with vegetation dynamics, plant-soil carbon/nitrogen cycle, and multi-layer vertical soil enabled (Oleson et al., 2013). In DGVM, plant photosynthesis, stomatal conductance, carbon/nitrogen allocation, plant phenology and multi-layer soil biogeochemistry are described in accordance with default CLM4.5, while vegetation dynamics (establishment, survival, mortality and light competition) are handled separately based upon ~~relatively~~ simple assumptions of environmental thresholds for establishment, survival and mortality of each PFT (see supplement S6) (Oleson et al., 2013). We used DGVM in the form of single-cell simulations for the 20 plots with grid-cell size set to 1×1 km (Table 1) to simulate the fractional cover of each PFT. All models were run with default CLM4.5 values for surface parameters (e.g. soil texture and depth), with prescribed atmospheric forcing derived from the 3-hourly hindcast of the regional model (SMHI-RCA4) driven by ERA-interim reanalysis for the European Domain of the Coordinated Downscaling Experiment – CORDEX for 1980–2010 (Dyrrdal et al., 2018). The CORDEX model simulation was used because it has a higher spatial resolution than the default atmospheric forcing used in CLM4.5 (0.11°×0.11° and 0.5°×0.5°, respectively). An inspection of the choice of atmospheric forcing, by which the CORDEX data were compared with the SeNorge data used for DM, showed minimal differences (Fig. S5). Only results obtained using CORDEX data are therefore shown in this paper. The 30-year CORDEX data was cycled during the spin-up. A 30-year period is consistent with WMO climatological normal based on the rationale that a 30 year-period is short enough to avoid large long-term trends while long enough to include the range of variability. Thus, the data ~~is~~are not de-trended or averaged.

The model was run with default PFT parameters (Table ~~S6~~S7). All the selected sites are mostly undisturbed. In our experiments, soil C and N were firstly initialized using a restart file from an existing global present-day spin-up simulation with prescribed vegetation. Each model simulation was spun-up for 400 years to establish a vegetation in equilibrium with the current climate after initialization from bare ground. In three plots where the equilibrium of vegetation was questionable (plot 6, 12 and 17), we extended the spin-up by 400, 200 and 200 years respectively to check if any effect on PFT profile could be seen. No significant changes in the PFT profile was noted in these three instances (Fig. S11.1 and Fig. S11.2) and therefore we behold the initial 400 year spin up for all the sites. A 20-year average at the end of the spin-up was used as input for calculation of PFT profiles (representing years 1990–2010), which corresponds with the data-collection timeframe of DM, RS and AR.

Among the 15 PFTs used in CLM4.5 to represent vegetated surfaces globally(Lawrence and Chase, 2007), only six (plus bare ground) were relevant for our study area (Table S5~~2~~). Bare ground was predicted to occur where plant productivity was below a threshold value (Dallmeyer et al., 2019). The DGVM simulates the vegetated land unit only (non-grey boxes in Fig. S~~7~~8), while other land units within the 20 plots, including glaciers, wetlands, lakes, cultivated land and urban areas, make up the "EXCL" PFT category (Table S5~~2~~). The percentage cover fraction of each PFT is equal to the average individual's fraction projective cover (*FPCind*) multiplied by the number of individuals (*Nind*) and average individual's crown area (*CROWNind*). *FPCind* is a function of the maximum leaf carbon achieved in one year, while *CROWNind* is related to dead stem carbon simulated by the model. *Nind* is mainly determined by establishment and survival rate controlled by establishment and survival threshold conditions (Levis et al., 2004)~~.~~. We obtained PFT profiles for each plot by excluding the EXCL category and ~~recalculating~~ recalculated fractions of the vegetated land unit covered by each PFT to sum up to one. ~~Each~~

~~model simulation was spun-up for 400 years to establish a vegetation in equilibrium with the current climate after initialization from bare ground. A 20-year average at the end of the spin-up was used as input for calculation of PFT profiles.~~

2.4.2 The RS method

As RS product we used SatVeg (Johansen, 2009), a vegetation map for Norway with 25 land-cover classes and a spatial resolution (~~pixel~~ grid cell size) of 30 m (Table 1). SatVeg is obtained by a combination of unsupervised and supervised classification methods, applied to Landsat 5/TM and Landsat 7/ETM+ images within the near-infrared and mid-infrared spectrum covering the period ~~of~~ 1999~~–~~-2006. While with the supervised classification, training data is based on well-labelled data from the study area, during the unsupervised classification the algorithm is only supplied with the number of output classes without further interference of the user. Only ~~pixel~~grid cells that were within each 1-km^2 plot with a majority of their area were taken into consideration for further calculations.

2.4.3 The DM method

The distribution models (DMs) for 31 vegetation types (VT) obtained by Horvath et al. (2019) using generalized linear models (GLMs, with logit link and binomial errors, i.e. logistic regression), were used for this study. The VT data were collected during years 2004~~–~~-2014. The DMs were obtained by using wall-to-wall data for 116 environmental ~~variables~~predictors from six groups (topographic, geological, proximity, climatic, snow and land cover), gridded to a spatial resolution of 100×100 m (Table 1) as predictors. Important predictors were selected by an automated stepwise forward-selection procedure for each of the 31 VTs individually, thus each final model is built upon only a narrow selection of important predictors (Horvath et al., 2019 supplement S7). All DMs were evaluated using an independent evaluation data set and by calculating the area under the receiver operator curve (AUC), a threshold-independent measure of model performance commonly used in DM. ~~by use of an independently collected data set~~ (see Horvath et al., 2019 for details). -AUC can be interpreted as the probability that the model predicts a higher suitability value for a random presence grid cell than for a random absence grid cell (Fielding and Bell, 1997). A seamless vegetation map (i.e. with one predicted VT for each ~~pixel~~ grid cell with no overlap and no gaps) was obtained from the stack of 31 probability surfaces by assigning to each grid cell the VT with the highest predicted probability of occurrence within that cell (Ferrier et al., 2002). ~~Pixels~~ Grid cells with the majority of their area ~~that were~~ within ~~each~~a 1-km^2 plot ~~with majority of their area~~ were used for further calculations (Fig. ~~S2~~S6).

2.5 Conversion to PFT profiles

Harmonisation of the various vegetation classification systems was accomplished by a conversion scheme that represented each grid cell (RS and DM) or polygon (AR) in each of the 20 plots with one out of the six PFTs recognised by DGVM (Table S5~~2~~ and Fig. ~~S2~~S6). The scheme was obtained by expert judgements and solicited by a consensus process which involved ecologists participating in the AR18x18 survey as well as scientists working with RS and DGVMs.

We used the conversion scheme of Table S5~~2~~ to generate wall-to-wall PFT maps from the original RS, DM and AR datasets (Table 1) by assigning one PFT to each 30×30 m grid cell, 100×100 m grid cell or VT polygon, respectively. PFT profiles for each plot, at the same thematic resolution as for DGVM, were obtained as the vector with fractions of grid cells or polygons assigned to each of the six PFTs. 'EXCL' classes not represented in DGVM (cf. Table S5~~2~~) were left out ~~in order~~ to minimise the effect~~s~~ of land use, which could otherwise have brought about differences in PFT profiles among the compared methods. PFT profiles were obtained for each combination of method and plot. To test for deviations in PFT coverage between the methods across the whole study area, a~~A~~ggregated PFT profiles were obtained by averaging the 20 PFT profiles obtained for each method.

~~Table 2 Conversion scheme for harmonizing vegetation and land cover types across methods (RS, DM and AR) into plant functional types (PFTs). DGVM – dynamic global vegetation model, RS – remote sensing, DM – distribution model. PFT – plant functional type, VT – vegetation type.~~

~~DGVM~~		~~RS~~	~~DM~~	~~AR~~
~~PFT code~~	~~plant functional type~~	~~vegetation / land cover type – remote sensing~~	~~vegetation type – distribution model~~	~~vegetation type – area frame survey~~
~~BG~~	~~Bare ground~~	~~Exposed alpine ridges, scree and rock complex~~	~~Frozen ground, leeward~~	~~Frozen ground, leeward~~
		-	~~Frozen ground, ridge~~	~~Frozen ground, ridge~~
		-	~~Boulder field~~	~~Sand dunes and gravel beaches~~
		-	~~Exposed bedrock~~	~~Pioneer – alluvial vegetation~~
		-		~~Barren land~~
		-		~~Boulder field~~
		-	-	~~Exposed bedrock~~
~~Boreal NET~~	~~Boreal needleleaf evergreen tree~~	~~Coniferous forest – dense canopy layer~~	~~Lichen and heather pine forest~~	~~Lichen and heather pine forest~~
		~~Coniferous forest and mixed forest – open canopy~~	~~Bilberry pine forest~~	~~Bilberry pine forest~~
		~~Lichen rich pine forest~~	~~Lichen & heather spruce forest~~	~~Meadow pine forest~~
		-	~~Bilberry spruce forest~~	~~Pine forest on lime soils~~
		-	~~Meadow spruce forest~~	~~Lichen & heather spruce forest~~
		-	~~Damp forest~~	~~Bilberry spruce forest~~
		-	~~Bog forest~~	~~Meadow spruce forest~~
		-		~~Damp forest~~
		-	-	~~Bog forest~~
~~Temperate BDT~~	~~Temperate broadleaf deciduous tree~~	~~Low herb forest and broadleaved deciduous forest~~	~~Poor / Rich broadleaf deciduous forest~~	~~Poor broadleaf deciduous forest~~
		-	-	~~Rich broadleaf deciduous forest~~
~~Boreal BDT~~		~~Tall herb – tall fern deciduous forest~~	~~Lichen and heather birch forest~~	~~Lichen and heather birch forest~~

	~~Boreal broadleaf deciduous tree~~	~~Bilberry- low fern birch forest~~	~~Bilberry birch forest~~	~~Bilberry birch forest~~
		~~Crowberry birch forest~~	~~Meadow birch forest~~	~~Meadow birch forest~~
		~~Lichen-rich birch forest~~	~~Alder forest~~	~~Birch forest on lime soils~~
		-	~~Pasture land forest~~	~~Alder forest~~
		-	~~Poor / rich swamp forest~~	~~Pasture land forest~~
		-		~~Poor swamp forest~~
		-	-	~~Rich swamp forest~~
~~Boreal BDS~~	~~Boreal broadleaf deciduous shrub~~	~~Heather-rich alpine ridge vegetation~~	~~Lichen heath~~	~~Lichen heath~~
		~~Lichen-rich heathland~~	~~Mountain avens heath~~	~~Mountain avens heath~~
		~~Heather- and grass-rich early snow patch communities~~	~~Dwarf shrub / Alpine calluna heath~~	~~Dwarf shrub heath~~
		~~Fresh heather and dwarf-shrub communities (u/l)~~	~~Alpine damp heath~~	~~Alpine calluna heath~~
		-	~~Coastal heath / Coastal calluna heath~~	~~Alpine damp heath~~
		-	~~Damp heath~~	~~Flood-plain shrubs~~
		-		~~Coastal heath~~
		-		~~Coastal calluna heath~~
		-		~~Damp heath~~
		-	-	~~Crags and thicket~~
~~C3~~	~~C3 grass~~	~~Graminoid alpine ridge vegetation~~	~~Moss snowbed / Sedge and grass snowbed~~	~~Moss snowbed~~
		~~Herb-rich meadows (up-/lowland)~~	~~Dry grass heath~~	~~Sedge and grass snowbed~~
		~~Grass and dwarf willow snow-patch vegetation~~	~~Low herb / forb meadow~~	~~Dry grass heath~~
		-		~~Low herb meadow~~
		-		~~Low forb meadow~~
		-	-	~~Moist and shore meadows~~
~~EXCL~~	~~Excluded~~	~~Ombrotrophic bog and low-grown swamp vegetation~~	~~Bog / Mud-bottom fen and bog~~	~~Bog~~
		~~Tall-grown swamp vegetation~~	~~Deer-grass fen / fen~~	~~Deer-grass fen~~
		~~Wet mires, sedge swamps and reed beds~~	~~Sedge marsh~~	~~Fen~~
		~~Glacier, snow and wet snow-patch vegetation~~	~~Pastures~~	~~Mud-bottom fen and bog~~
		~~Water~~		~~Sedge marsh~~

		~~Agricultural areas~~		~~Cultivated land~~
		~~Cities and built-up areas~~		~~Pastures~~
		~~Unclassified and shadow affected areas,~~		~~Built-up areas~~
		-		~~Scattered housing~~
		-		~~Artificial impediment~~
		-		~~Glaciers and perpetual snow~~
		-		~~Sea and ocean~~
		-	-	~~Water bodies (fresh)~~

2.6 Comparison of PFT profiles

To examine the overall pattern across the study area and to assess the models' ability to produce overall predictions of PFTs that accord with the PFTs' overall frequency (as given by the reference) ~~A~~aggregated PFT profiles obtained by each of the DGVM, RS and DM methods were compared with the aggregated PFT profile of the AR reference dataset by a chi-square test (Zuur et al., 2007). ~~.~~

-To identify strongly deviating modelling results at a plot scale, ~~For each plot,~~ the dissimilarity between PFTs profiles obtained by each of the DGVM, RS and DM methods and the PFT profile of the AR dataset for each plot was calculated by using proportional dissimilarity (Czekanowski, 1909):

$$d_{hj}=\sum|y_{hji}-y_{0ji}|/\sum(y_{hji}+y_{0ji})=1-2\sum min(y_{hji},y_{0ji})/\sum(y_{hji}+y_{0ji})$$

where y_{hji} refers to the specific element in a PFT profile vector (the fraction occupied by the PFT in question) given by method h (DGVM, RS or DM; h = 1, ..., 3; the value h = 0 refers to the AR reference dataset), j refers to sampling unit (j = 1, ..., 20) and i refers to PFT (i = 1, ..., 6). Proportional dissimilarity is the Manhattan measure standardized by division by the sum of the pairwise sums of variable values (here PFTs). Since the values of each PFT profile sums to one, the index reduces to

$$d_{hj}=1-\sum min(y_{hji},y_{0ji})$$

The proportional dissimilarity index is appropriate for incidence data like PFT abundances, i.e. variables that take zero or positive values. The index reaches a maximum value of 1 when two objects have no common presences (here, PFTs present in both compared objects) and ignore joint absences (zeros). To assess the degree to which the models produce pairwise similar differences, ~~W~~we compared the pairwise differences between the proportional dissimilarity values among methods, using a Wilcoxon-Mann-Whitney paired samples test.

All raster and vector operations related to DM, RS and AR were carried out in R (version 3.4.3) (R Core Team, 2019) using packages "rgdal" (Rowlingson, 2019), "raster" (Hijmans, 2019) and "sp" (Pebesma and Bivand, 2005), while graphics are produced using the "ggplot2" package (Wickham, 2016). Statistical analyses were carried out in R (version 3.4.3), using the "vegan" package (Oksanen et al., 2019). All maps were produced in QGIS (QGIS Development Team, 2019).

3 Results

The aggregated PFT profiles for the RS and DM datasets did not differ significantly from those of the reference AR dataset according to the chi-square test, while a significant difference was found for the DGVM profiles (Table 2). While the proportion of ~~pixels~~ grid cells attributed to the PFT 'boreal NET' by the RS and DM methods underestimated AR values by 3.0 and 2.8 percentage points, respectively, DGVM overestimated the proportion of boreal NET by 20.4 percentage points compared to the AR reference. Also, unproductive areas (BG) were overestimated~~represented~~ by DGVM (by 16.6 percentage points), less so by RS (4.0 percentage points), while this PFT was slightly underrepresented by DM (by 5.0 percentage points). Discrepancies were also observed for the cover of the C3 PFT, which was overestimated by RS and DM (by 7.2 and 2.9 percentage points, respectively) and underestimated by DGVM (by 3.0 percentage points)~~ by DGVM~~. Furthermore, DGVM overestimated BG and temperate BDT cover on the expense of boreal BDT and boreal BDS.

Table 2 - PFT profiles (columns) aggregated across all 20 plots for the three methods compared in this study and the AR reference dataset. Results of comparisons of aggregated PFT profiles for each of the three methods with the reference are also given. DGVM – dynamic global vegetation model, RS – remote sensing, DM – distribution model, AR – reference dataset. BG – bare ground, boreal NET – boreal needleleaf evergreen trees, temperate BDT – temperate broadleaf deciduous trees, boreal BDT – boreal broadleaf deciduous trees; boreal BDS - boreal broadleaf deciduous shrubs, C3 – C3 grasses.

PFT	Compared methods			Reference
	DGVM (%)	RS (%)	DM (%)	AR (%)
BG	29.5	17.0	7.9	12.9
Boreal NET	57.2	34.0	33.8	36.8
Temperate BDT	5.6	2.0	0.2	0.5
Boreal BDT	3.1	12.5	17.2	15.5
Boreal BDS	4.1	23.8	34.5	30.8
C3	0.5	10.7	6.4	3.5
Chi-square test	$\chi^2 = 45.98$, df = 5, $p < 0.05$	$\chi^2 = 6.36$, df = 5, p = 0.27	$\chi^2 = 2.61$, df = 5, p = 0.75	

In accordance with results from comparisons between aggregated PFT profiles obtained by the three methods and those obtained for the reference dataset, DGVM profiles for individual plots were significantly more dissimilar to the AR reference than ~~were~~ RS and DM profiles (Fig. 2). While RS had the lowest median proportional dissimilarity with the AR reference (0.19, compared to 0.26 for DM and 0.41 for DGVM), DM had the lowest spread of dissimilarity values, measured as interquartile difference (0.12, compared to 0.19 for RS and 0.72 for DGVM), among the three methods (Fig. 2~~3~~). While no dissimilarity value for RS was above 0.50, two ~~sampling unit~~plots (4, 19) acted as strong outliers in the distribution of DM values (cf. Fig. 2 ~~and Fig. 3~~). Additionally, ~~A~~a comparison of proportional dissimilarity between pairs of methods revealed significant differences between DGVM profiles and those obtained by RS and DM (Wilcoxon rank-sum tests: W = 111, p = 0.0167; and W = 88, p = 0.0026, respectively), while RS and DM profiles were not significantly different from each other (Wilcoxon rank-sum test: W = 161, p = 0.3013).

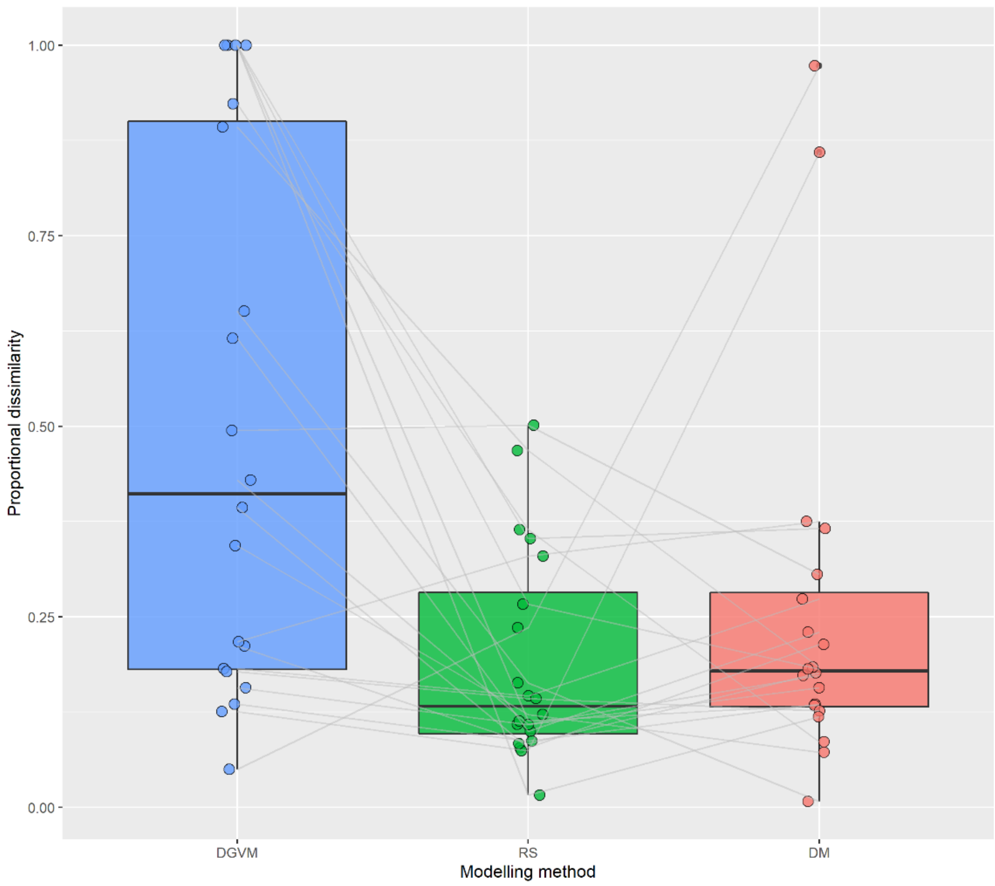


Figure 2 - Proportional dissimilarity values between PFT profiles for each combination of 20 plots and ~~one~~ each of the three methods ~~compared~~ evaluated in this study, and the corresponding plot in the AR reference dataset. The thick horizontal line, the box and the whiskers represent the median, the interquartile difference and the range of values for each method.

Visual inspection of spatial patterns of PFT profile characteristics across the 20 plots suggests that the best agreement among the methods was obtained for the south-eastern part of the study area, dominated by the boreal NET (Fig. 3 and Table S10). Compared to the AR reference dataset, PFT profiles obtained by DGVM were strongly biased: in the north (plots 17 and 18) towards boreal NET on the cost of boreal BDT, near the west coast (plots 1, 2, 5 and 15) towards boreal NET on the cost of boreal BDS, and in southern coastal areas (plots 3, 6 and 12) towards temperate BDT instead of boreal NET. In ~~sampling unit~~plots 13 and 16 DGVM failed to establish vegetation (predicting bare ground) where AR reported boreal BDS. RS represented the PFT profiles of the AR reference well in most cases, but tended to overestimate the frequency of dominance by C3 grasses at several locations (plots 3, 16 and 20). While DM showed no general spatial pattern of PFT profile deviations from the reference dataset, PFT profiles of plots 4 and 19 obtained by DM had almost no similarity to the corresponding profiles of the AR reference dataset: C3 grasses and boreal BDT were predicted instead of bare ground and boreal NET, respectively.

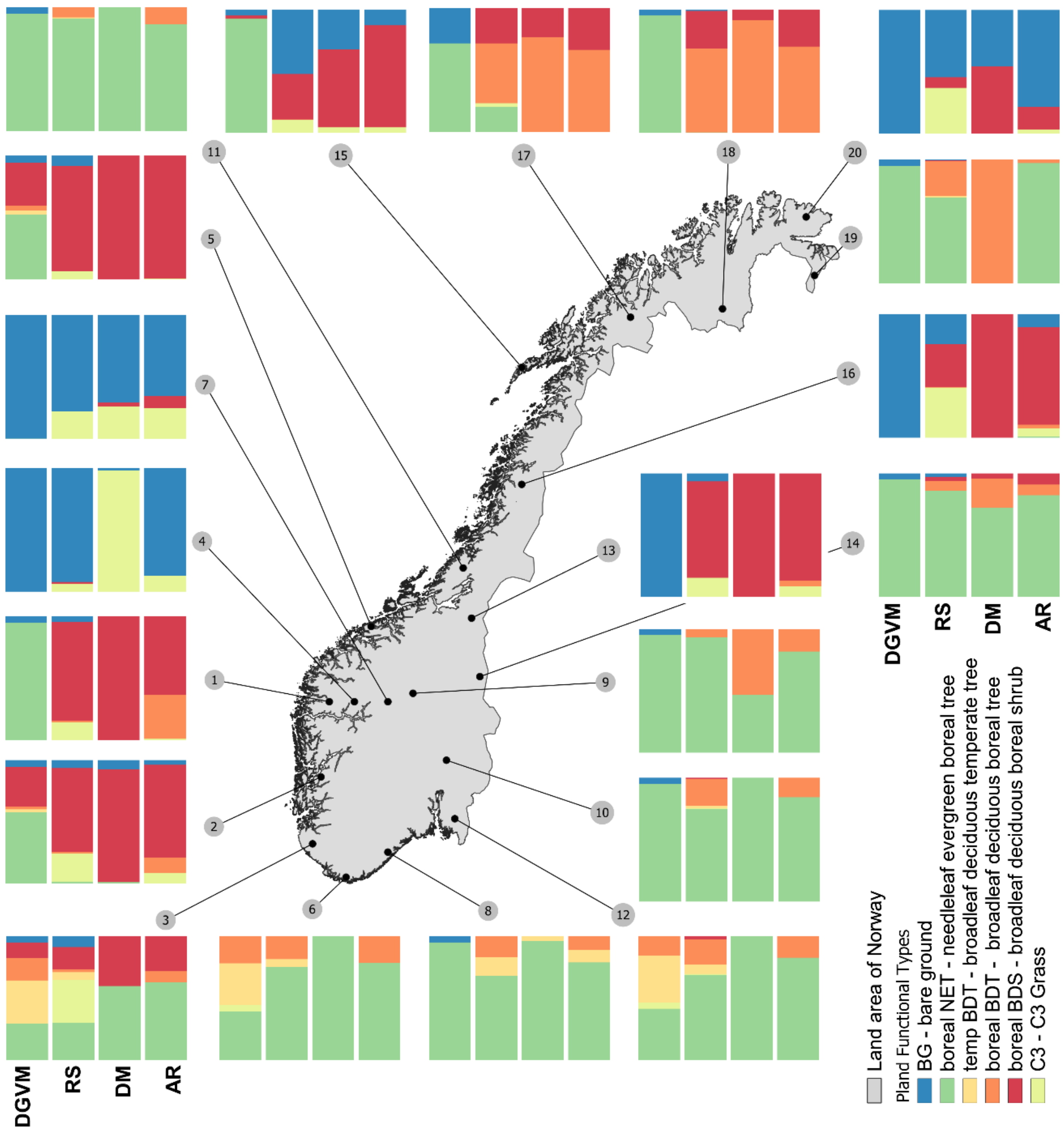


Figure 3 – PFT profiles for each of the 20 plots for the three methods compared in this study and the AR reference dataset. The columns in each cluster of four bar-charts represent, from left to right, the methods dynamic global vegetation model (DGVM), remote sensing (RS) and distribution model (DM), with the AR reference dataset to the right.

4 Sensitivity experiments and model improvement

~~4.1 Methods~~

We used the results of PFT profile comparisons between DGVM and the AR reference (Fig. 3) and the results obtained for the DM dataset as a starting point for the exploring possible ~~relationships~~causes of ~~between~~ the poor performance of DGVM ~~and DGVM parameter settings~~. We first identified the three most abundant PFTs (i.e. boreal NET, boreal BDT and boreal BDS) in our set of plots (Table ~~S4~~S3). Thereafter, we identified the major VTs predicted by DM in those plots that were translated into these PFTs using the conversion scheme (Table S5) ~~(to be~~ pine forest, birch forest and dwarf shrub heath, respectively ~~(~~Table 3). Based on the results from Horvath et al. (2019), the corresponding final models for these three VTs were examined to identify important environmental variables that were driving the distribution of the VTs but not represented in DGVM. We ~~selected~~ recognized ~~the~~ three ~~of the most important~~ environmental predictors that are critical for ~~for~~ the distribution of each of these VTs~~,~~ and exhibit ~~with~~a clear threshold signature in the frequency-of-presence plots (i.e. graphs showing variation in the abundance of the VT as a function of an environmental predictors, also see Fig. S12):~~ as identified by DMs (see Horvath et al. 2019) – for further sensitivity experiments of DGVM parameter settings (~~Table 3~~):~~ snow water equivalent in October (swe_10), minimum temperature in May (tmin_5) and precipitation seasonality (bioclim_15). ~~–~~Precipitation seasonality is defined as the ratio of the standard deviation of the monthly total precipitation to the mean monthly total precipitation (i.e. the coefficient of variation), expressed as percentage (O'Donnell and Ignizio, 2012). ~~We used~~Based on visual inspection of the frequency-of-presence plots~~, (i.e. graphs showing variation in the abundance of the VT as a function of an environmental variable) to~~ we identif~~y~~ied specific threshold values for ~~presence~~ each of the three VTs (see Fig. S12 for details) and implemented these threshold values into DGVM as new limits for establishment of the three PFTs as shown in Table 3~~ (also see Fig S11)~~. ~~(F~~For example, in line with Fig. S12, VT 2ef and its respective PFT - boreal BDS ~~–~~can only~~ is not~~ establish~~present along the~~ when variable swe_10 is less than ~~above the value of~~ 380mm.~~ – thus the threshold was decided appropriately; Fig S12)~~.

We explored the extent to which these additional thresholds ~~revised parameter settings~~ improved the performance of DGVM on the subset of six plots (i.e. ~~numbers~~ plot 1, 2, 5, 15, 17 and 18) in which the PFT profiles are most biased ~~boreal NEB was most strongly overrepresented~~ compared to the AR reference dataset due to the overrepresentation of the boreal NEB. In total, three s~~S~~ensitivity experiments were carried out by a stepwise process, in each step ~~adding one~~a new threshold was added ~~compared~~cumulatively to the previous experiments (Table 3).~~specific for each of the three PFTs (see Table S13 for details on the stepwise process of DGVM parameter adjustments), specific for the three PFTs at the same time. Parameters were cumulatively added in the following order:~~ Namely, in the first sensitivity experiment (i), we added ~~swe_10 as~~ the swe_10 threshold. In the second experiment (ii), we added both swe_10 and~~,~~ tmin_5~~ and bioclim_15~~ as the threshold. In the last experiments (iii), we added all the three novel thresholds.~~(the last only relevant for the boreal NET)~~ Only the results of the third sensitivity experiment~~DGVMs run (iii)~~ with all the three thresholds added ~~parameters changed~~ are reported here. Re~~(re~~sults of the other two experiments are summarised in Table S13~~2~~). ~~For example, in the three sensitivity model runs (i–iii), (i) the requirement for establishment of boreal NET was set to swe_10 > 150 mm; in (ii) and (iii) the additional demands tmin_5 > – 5 °C and bioclim_15 < 50, respectively, were enforced.~~

Table 3 – New ~~parameter~~ thresholds for establishment of the three PFTs explored in DGVM sensitivity experiments. ~~Variables for which parameter settings~~The variables ~~were~~ explored were: swe_10 – snow water equivalent in October given in mm; tmin_5 – minimum temperature in May (°C); bioclim_15 – precipitation seasonality (unitless index representing annual trends in precipitation).

		Sensitivity model run		
		(i)	(ii)	(iii)
VT	PFT	~~SWE~~swe_10 (mm)	t~~T~~min_5 (°C)	~~B~~bioclim_15
2ef – Dwarf shrub heath / Alpine calluna heath	Boreal broadleaf deciduous shrub	<~~>~~ 380	> -10	–
4a – Lichen and heather birch forest	Boreal broadleaf deciduous tree	<~~>~~ 180	> -7.5	–
6a – Lichen and heather pine forest	Boreal needleleaf evergreen tree	<~~>~~ 150	> -5	< 50

4.2 Results

The results show that w~~W~~hile the added thresholds for swe_10 and tmin_5 had little impact on the results (Table S13), the addition of the threshold for bioclim_15 (i.e., the third sensitivity experiment) largely ~~the~~ improved the performance of the DGVM on the experimental plots explored (Fig. 4). ~~After the third run (iii) ofby which A~~adding ~~a new parameter thresholds was added in accordance with~~ Table 3~~, made~~ PFT profiles simulated~~identified~~ by ~~DGVM~~this experiment were more similar to those of the AR reference dataset for four out of the six plots in the experimental subset (plot 1, 2, 5 and 15): in plots 1 and 15, Boreal NET was correctly replaced by boreal BDS; in plots 2 and 5 boreal NET was replaced by boreal BDT, BDS and temperate BDT. Addition of new ~~parameter~~ threshold (bioclim_15)~~s~~ also reduced the modelled abundance of boreal NET in plots 17 and 18, but DGVM still failed to populate these plots with another PFT (Fig. 4).~~The improved performance of DGVM on the experimental sampling units was mainly due to the implementation of the threshold for bioclim_15, while the changes made for swe_10 and tmin_5 had little impact on the results (Table S12)~~.

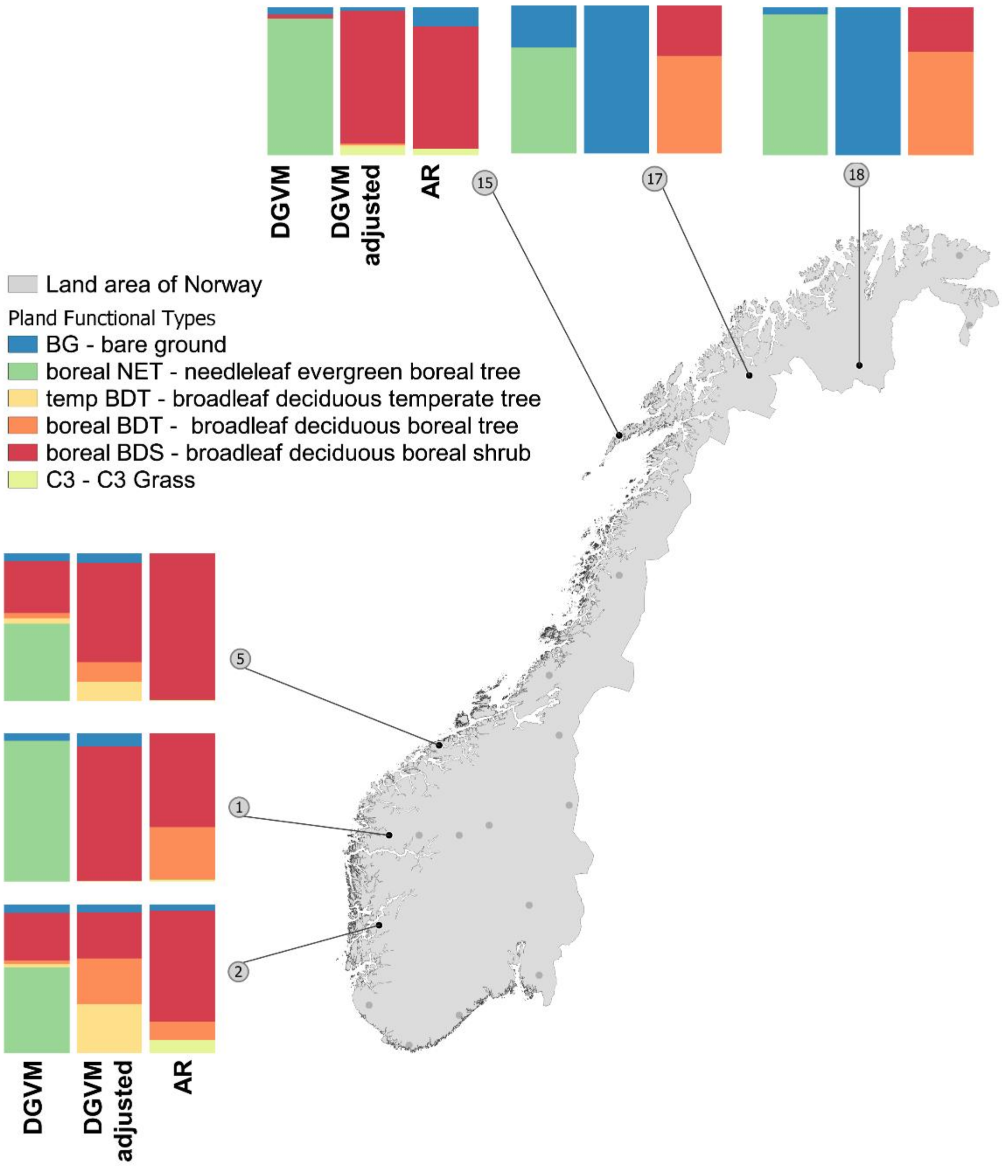


Figure 4 – PFT profiles for the subset of six plots subjected to sensitivity experiments with new DGVM establishment thresholds. The columns in each cluster of three bar-charts represent, from left to right, dynamic global vegetation model (DGVM) with original (default) parameter settings, DGVM with revised parameter settings, and the AR reference dataset. For further details, see Table S13~~2~~.

5 Discussion

5.1 Comparison of PFT profiles

The maps of PFT distributions generated by DM and RS are generally similar (Fig. ~~S8~~S9) across most of our study area. This indicates that output from DM, which is rarely used for evaluating PFT distributions from DGVMs, can be used for this purpose in addition to the commonly used RS-based datasets. There are, however, some differences between results obtained by the two methods near the northern Norwegian coast and in the mountain areas of western Norway, which will be discussed below in more details.

We recognise six possible explanations for the differences in PFT profiles obtained by DGVM, RS and DM for the 20 plots (see Table 5), related to the following issues: (i) the conversion scheme (ref. Table S5~~2~~); (ii) what is actually modelled by DGVM, RS and DM, e.g. in terms of potential vs actual vegetation; (iii) the performance of individual DM models; (iv) transforming predictions from single DMs into a seamless vegetation map, i.e. that assigns one VT to each ~~pixel~~grid cell; (v) DGVM performance; and (vi) missing PFTs in DGVM.

5.1.1 The conversion scheme

The conversion schemes used to reclassify vegetation and land cover classes into PFTs have been reported as a possible attributor to erroneous PFT distributions (Hartley et al., 2017). While we use a simple conversion scheme ~~which~~ that assigns each land cover type/vegetation type to one and only one PFT (Dallmeyer et al., 2019), more complex conversion schemes exist, by which each land cover class is translated into a multi-PFT composition that co-occur within a grid cell (Bonan et al., 2002; Li et al., 2016; Poulter et al., 2011; Poulter et al., 2015). Our approach may be advantageous when the classes to be converted are homogeneous, in the sense that one PFT is clearly dominating in the type, and in the sense that the range of variation within the class in PFTs is negligible, such as is the case for 90% of the DM- and RS-classes in our study. Our simple scheme may, on the other hand, be a source of ~~bias~~ uncertainty when quantitatively important VTs are ambiguous in one way or the other, or, more commonly, in both ways at the same time. The set of VTs used in our study includes several relevant examples: VTs that may include a wide spectrum~~e~~ of tree-dominant types; the VT '*1a/1b - Moss snowbed / Sedge and grass snowbed*', which covers a range of variation in the relative abundance of graminoids and, hence, shows affinity to C3 as well as to BG; and the VT '*8a - Damp forest*', which is usually dominated by the evergreen Scots pine and converted into boreal NET, but that in some instances (e.g. after clear-cutting) is dominated by deciduous trees like *Betula* spp. and should then be converted into boreal BDT (Bryn et al., 2018). However, a close inspection of DM shows that our method reproduced~~s~~ similar PFT profiles as the reference dataset for all plots, except two out of 20 plots (the two outliers on Fig. 2, ~~represented by~~ plots 4 and 19 in Fig. 3).

In our case, a more complicated conversion scheme is likely to be compensated for by the sub-grid complexity introduced in the process by which PFT profiles are obtained. Rather than estimating a PFT profile for the 1-km^2 plot directly, i.e. in one operation as in DGVM, the RS-based classes and VTs are first converted into PFTs in their original resolution, and then subsequently subjected to aggregation to obtain the PFT profiles. This results in a sub-grid PFT heterogeneity that could otherwise be implemented by using a more complex conversion scheme.

5.1.2 What is modelled by DGVM, RS and DM

The methods used in this study produce different representations of the vegetated land surface in terms of actual or potential natural vegetation (Table 4). In order to model future vegetation changes and feedbacks, functional type-based models like DGVM implicitly address the processes that control the distribution of vegetation (Bonan et al., 2003; Song et al., 2013). Simulating natural vegetation processes under a given climatic equilibrium scenario (at any given time), DGVM produces a model of potential natural vegetation (ex. Bohn et al., 2000, Hengl et al. 2018). RS-based classifications, on the other hand, describe the land surface at a specific time-point or changes through time (e.g. Arctic greening and browning) (Myers-Smith et al., 2020) and, accordingly, portrays actual vegetation as influenced by previous and ongoing land use (Bryn et al., 2013). Depending on the modelling setup, DM may pragmatically describe the current ecological envelope of a target or aim at revealing the proximate causes for its distribution (Ferrier and Guisan, 2006), thus modelling either actual or potential natural vegetation, depending on the input data used for modelling (Hemsing and Bryn, 2012; Hengl et al., 2018).

In this study, we carefully restricted our attention to PFTs that represent natural vegetation, excluding VTs with strong anthropogenic influences. This was done for all methods and the AR reference. Nevertheless, differences with respect to what is actually modelled by the different methods, potential vegetation by DGVM and actual vegetation by RS and DM, may have contributed to the observed among-model differences in PFT profiles.

5.1.3 DM performance

While the performance of the DM method is overall good, distribution models of individual VTs vary in performance (with AUC values ranging from 0.671 to 0.989) according to the study by Horvath et al. (2019). Several reasons for the low predictive performance of some DM are identified, of which the most important is considered to be important predictors missing in the training data. This might seem counter-intuitive, given the large number of predictor variables used in the study ~~(~~n=116~~) ~~. However, the authors conclude that several important factors for the distribution of vegetation are not at all represented in the data set (e.g. NDVI, LiDAR etc.), amo~~u~~ng others because they are almost impossible to obtain data for with required spatial resolution (e.g. soil nutrients). The DM method requires estimates for the probabilities of occurrence for (almost) all individual vegetation types to create a seamless vegetation map, which in turn is required for making estimates for the PFT profiles as robust as possible. Thus, in this context, ‘poor’ models are better than no model.

Individual models’ performance might be the reason for the two plots ~~stand out by~~ whose PFT profiles ~~that~~ deviate strongly from the AR reference (Fig. 2 and Fig. 3). For plot 4, the discrepancy is due to VT “*1a/1b - Moss snowbed / Sedge and grass snowbed*”, which is represented by one of the best performing among the 31 DMs. For this VT, conversion scheme bias is a more likely reason for the deviant PFT profile. For plot 19, boreal BDT is modelled because the VT predicted by DM is *“4a – Lichen and heather birch forest”*. The fact that the DM for this VT is among the inferior DMs (see the ranking of individual models presented in Horvath et al. (2019)) makes this explanation more likely in this case.

5.1.4 Transformation of single-DM predictions into a vegetation map

The performance of DM on the particular plots may also be influenced by the method chosen for transforming predictions from one DM for each VT into a seamless vegetation map. Assigning to each grid cell the VT with the highest predicted probability of presence in that cell, which is a commonly used method for this purpose (Ferrier and Guisan, 2006), favours VTs represented by good DMs. This is brought about by good DMs having a distribution of predictions that is more spread out (with larger predictions for the ~~pixels~~ grid cells identified as the most favourable cells) than poor DMs (Halvorsen, 2012). However, since the probability of presence for each VT was predicted separately for each grid-cell, the probability values for every VT vary independently of the probabilities for the other VTs, throughout the study area. Thus, we regard the chance that one VT consistently outperforms another VT over all the grid cells to be negligible. Alternative methods for this purpose should be tested in the context of DGVM evaluation~~ng~~.

To avoid uncertainties associated with conversion between type systems and perhaps even further improve the performance of DM, we recommend exploring the option of using PFTs directly as targets in DM. Direct modelling of PFTs rather than taking the detour via VT models may reduce the number of environment predictors required (116 layers used in Horvath et al. (2019)) in addition to circumventing the complicated process of modelling thematically narrow vegetation types (VTs). Another potential advantage of modelling PFT targets directly is that the model parameters will then be PFT specific, and not in need of being converted (from VT into PFT).

To further reduce the biases and uncertainties of DM-based PFT profiles, we recommend exploring the use of variables derived from RS directly as predictors in DM. Previous studies have shown that RS -based predictors may enhance DM performance on different scales: on vegetation-type level (Álvarez-Martínez et al., 2018); on the habitat-type level (Mücher et al., 2009); and on the PFT level (Assal et al., 2015). Further suggestions for improvement of the methods used in this study are found in Table 4.

Table 4 – A summary of the key properties of the three methods compared in this study. DGVM – dynamic global vegetation model, RS – remote sensing and~~,~~ DM – distribution model~~, AR – reference dataset~~.

Key property	Method		
	DGVM	RS	DM
Modelled property	Process-based vegetation model – using ~~on~~ *a priori* parameterizations	Classification based on satellite imagery (spectral reflectance)	Statistically based model of a target (response) and the environment (predictors)
Main purpose	Feeding vegetation changes into ESM for further quantification of feedbacks between land surface and the atmosphere	Mapping of land cover or land use for descriptive purposes, management or monitoring	Predicting the spatial distribution of a target and/or to summarise its relationship with the environment
Material	Climate forcing, PFT parameters, host model	Satellite imagery in different bands	Presence-absence training data, environmental predictors
Spatial extent	Global to regional (Single-cell tests)	Global to local	Regional to local
Modelling outcome	Potential vegetation	Actual vegetation	Potential or actual vegetation, depending on the training data
Advantages	– Addresses the processes – Feedback loops with other Earth system components can be included – Continuous temporal scale of prediction into the future	– Observation-based – High spatial resolution – Good temporal coverage	– Opens for use of proxies for important predictors – May provide insight into drivers of distributions
Disadvantages	– Low performance (e.g. compared with RS and DM) as long as the underlying processes are not fully understood and properly parameterised – Parameter intensive – Resource demanding	– Data are sensitive to cloud cover and shaded areas – Atmospheric correction needed – Provides limited insight to the processes that regulate the distributions of land cover types – No feedback included	– Provides limited insight to the processes that regulate the distributions of targets – Temporally static (one time-point addressed by each model) - No feedback included
Possible interactions with	– May improve DM by pointing at relevant predictor variables	– May improve DGVM by improved parameterization (based on RS indices)	– May improve parameterization and envelope discrimination of DGVM

the other methods	– May improve RS by identifying threshold values	– May improve DM by providing predictor variables, directly or as indices (NDVI, ~~PAR~~ etc)	– May improve RS by targeting specific PFTs that have similar reflectance, but different ecology

5.1.5 DGVM performance

Our results show that, for many plots, the PFT profiles simulated by DGVM differ~~s~~ from those of the AR reference dataset. According to our results, DGVM overestimates~~predicts~~ the coverage of bare ground and boreal NET and underpredicts the cover of C3 grasses, boreal BDT and boreal BDS. While the AR reference dataset shows that the northern plots (specifically plots 17 and 18) are covered by mountain birch forest and shrubs (boreal BDT and boreal BDS), DGVM predicts dominance of boreal NET in these plots. Overestimation of boreal NET has also been reported by Hickler et al. (2012) for large parts of Scandinavia, who attributed this to the lacking representation of shade tolerance classes in DGVM models. A similar pattern is seen in our results: the PFT profiles obtained by DGVM during the 400-year spin-up (Fig. S11~~0~~) show no sign of boreal BDT in the early phases of model prediction, as would be expected of an early successional forest in Norway.

~~The western parts of Scandinavia are dominated by shade intolerant birch forests (Bryn et al., 2018) which gradually give way to coniferous forests along the oceanity-continentality gradient towards east (Wielgolaski, 2005). The overprediction of DGVM in the west indicates that the DGVM does not only lack shade-intolerant PFTs, but also that improved representation of winter-time respiration loss and soil frost-induced drought stress of boreal NET in spring in regions with higher temperature fluctuations around 0°C during winter time compared to the more continental regions (see e.g. Oksanen, 1995; Sevanto et al., 2006) are needed.~~

Our results further suggest that the DGVM underrepresents grasses and shrubs compared to the reference dataset. This may be explained by the built-in constraints in the light competition scheme of DGVM. ~~For example Oleson et al. mention that~~ The model assumes that r~~r~~egardless of grass and shrub productivity, trees will cover up to 95% of the land unit when their productivity permits (Oleson et al., 2013). The priority given to a PFT in DGVM decreases with the stature of the organisms in question because of the increasing probability that a lower layer is covered by another layer. The degree of underrepresentation is therefore expected to increase from shrubs to grasses. Accordingly, DGVM predicts dominance by trees in the most productive regions, by grasses in less productive regions, and by shrubs in the least productive non-desert regions (Zeng et al., 2008). The underrepresentation of C3 grasses by DGVM across the 20 ~~study~~ plots ~~in our study~~ accords with the results of Zhu et al. (2018), who found that C3 grasses are underpredicted on a global level in an earlier version of DGVM.

Inappropriate parameterisation of shrubs may be a reason why the DGVM underestimates boreal BDS in many of the coastal plots (1, 2, 5, 15) (Table S6). The implementation of shrubs as a new PFT in an earlier version of DGVM (CLM3-DGVM) by Zeng et al. (2008), ~~which is~~ parameterised for representation of taller shrubs with heights between 0.1 and 0.5 m, may not suit the majority of dwarf shrubs (of genera *Calluna, Betula, Empetrum*) that abundantly occurs in Norwegian ecosystems. To this, Castillo et al. (2012) add that the sparse shrub and grass vegetation cover simulated by DGVM in the tundra regions may be caused by the soil moisture bias inherited from the host land model CLM4 (Lawrence et al., 2011). Another reason for DGVM's underestimation of boreal BDS in coastal areas could be the 4000-yr tradition of coastal heath management in Norway (Bryn et al., 2010) which causes a large discrepancy between the actual vegetation modelled by RS, DM and AR and the potential natural vegetation simulated by DGVM under present-day climatic conditions (e.g. Bohn et al., 2000, Hengl et al. 2018).

We therefore argue that more sensitivity studies of PFT-specific parameters for height, survival, establishment etc., across all PFTs, are needed.

Some discrepancies in the DGVM output might be caused by the climate forcing used in the simulations, looped for the period 1980~~—~~–2010. Long-term historical climate effects on vegetation distribution were not included in our model simulation. However, we noticed that vegetation distribution was insensitive to interannual variation or decadal variation of the climate forcing when it reached equilibrium state in most of our study sites. Even though long-term historical climate effects (such as cooler temperature in the early 20^{th} century) may favour boreal BDS rather than boreal NET, we consider such historical effects to have only minor impact on the already large biases observed in DGVM (e.g., too much boreal NET and too few BDS). We also note that DGVM used a spatially coarser CORDEX reanalysis (11x11 km) to supply high temporal resolution (6-hourly) atmospheric forcing data, while the climate predictors used in DM was derived from observation-based SeNorge v2 dataset with 1x1 km spatial resolution and daily temporal resolution. The larger biases in CORDEX reanalysis data may also contribute to the large mismatch between DGVM and the reference dataset. We have compared the average annual temperature and annual precipitation of the two input datasets used in DGVM and DM to look for differences (see Fig. S4). It appears that precipitation estimates by CORDEX for the 20 plots were slightly higher than SeNorge estimates, the converse (but less strongly) was true for temperature. The consequences of these differences in the input data might be investigated in follow-up studies.

Despite the shortcomings discussed above, DGVM performs reasonably well for some PFTs. One example is the temperate BDT, which is correctly predicted by the model to be restricted to the southern coastal plots (Bohn et al., 2000; Moen, 1999). This finding suggests that some climatically driven PFTs (i.e. temperate BDT) are well implemented by the existing parameters in the ~~current~~ DGVM used in this study.

5.1.6 Missing PFTs

DGVM coerces the World's immense variation in plant species composition (vegetation) into a very limited number of predefined PFTs, compared to classification schemes used by the other methods in this study (RS, DM and AR; see Table S5~~2~~) and by other approaches to systematisation of ecodiversity (e.g. ~~(~~Dinerstein et al., 2017; Keith et al., 2020~~)~~). In particular, the number of high-latitude specific PFTs is insufficient to realistically represent the biodiversity of these ecoregions, as pointed out by Bjordal (2018) and Vowles & Björk (2017). Comparisons between PFT profiles obtained by DGVM and profiles obtained by DM ~~may~~ suggest specific vegetation types that need to be better represented in DGVMs, either by improving an existing PFT or by adding a new PFT (e.g. dwarf shrubs vs. tall shrubs; moss dominated snow-beds, wetlands, lichens). In our study, the PFT profile of DGVM is represented by the six boreal PFTs, whereas the original data for RS, DM and AR include an average of 17% (ref. Table ~~S4~~S3) of the total area ~~which cannot be~~that are not represented by these six PFTs (classes for "Excluded" PFT category ref. Table S5~~2~~). This ~~reminds~~ points to ~~us of~~ the missing PFTs in the classification scheme of the DGVM, but ~~it~~ also ~~points~~ to the ~~problem~~ challenge that certain ecosystems in our study area do not have a ~~real~~ representation in the PFT schemes of DGVM. This is exemplified by wetlands; important ecosystems that are still not represented in many of the current DGVMs. This is not only problematic from the perspective of land surface energy balance (Wullschleger et al., 2014), but has also ~~brings issues~~ implications for modelling of carbon storage and cycling, and other interactions between the land surface and the atmosphere (Bjordal, 2018).

Some recent examples with improvements to the thematic resolution of PFTs in DGVMs are available in the literature (Druel et al., 2019; Coppell et al., 2019; Chadburn et al., 2015; Porada et al., 2016; Druel et al., 2017), and further examples of DGVMs with a larger number of high-latitude PFTs also exist (Euskirchen et al., 2009). In line with these studies, ~~O~~our results demonstrate a great potential for increasing the thematic resolution of DGVMs in general and not limited to the DGVM tested here in terms of developing and parameterizing new specific PFTs to be representative of the high-latitude and high-altitude habitats, ~~as exemplified by Druel et al. (2017)~~ and also deriving parameters from observations, DMs or RS products (Bjordal, 2018; Wullschleger et al., 2014), specific for the high latitudes (Druel et al., 2017).

5.2 Sensitivity ~~tests~~experiments

Adjusting DGVM parameters so that they correspond better with environmental drivers known to be functional in the high-latitude PFTs has been suggested as a measure to improve the performance of DGVM ~~in these parts of the World~~ (Wullschleger et al., 2014). Our ~~simple~~ sensitivity experiments demonstrate that DM results can inform ~~parameterisation, in~~ DGVM ~~of new~~ parameterisations~~,~~ based upon ~~of the~~ suitability ranges of the environmental predictors recognized by DM in determining the distribution of a PFT. ~~where a PFT occurs along variables predictors used in DM where a PFT occurs.~~ Most notably, we recognize~~d~~ that the implementation of ~~three important environmental drivers~~ precipitation seasonality (bioclim_15 < 50) as a threshold for the establishment of NET, which has not yet been used in the DGVM~~,~~ ~~for~~ improves~~d~~ the distribution of high-latitude PFTs simulated by the DGVM ~~not yet represented well in DGVM~~. This adds to the environmental thresholds for establishment ~~, survival or mortality~~ of a PFT previously used in DGVMs to restrict the predicted distribution of PFTs to realistic geographic regions (Miller and Smith, 2012). Even though our sensitivity experiments focus on a limited number of additional thresholds across three PFTs, this approach shows promising results and is worth ~~should be tested~~ exploring~~ed~~ more extensively in ~~the~~ future studies.

~~Adjustment of the climatic thresholds for the establishment of the high-latitude PFTs (i.e. boreal NET, BDT, BDS) seemingly bring the PFT profiles of DGVM closer to those of the reference data (Fig. 4). In particular, the sensitivity experiments with DGVM highlight~~ T~~t~~he importance of precipitation seasonality (i.e. bioclim_15) as a critical limiting factor for the establishment of boreal NET indicates~~, and show~~ that the increased seasonality impedes growth of boreal NET. While some studies have emphasized the importance of seasonal distribution of rainfall on vegetation in the semi-arid areas (Zhang et al., 2018), the importance of this factor for high-altitude areas is less well studied (Oksanen, 1995; Sevanto et al., 2006). Better representation of the processes related to the response of boreal NET to water availability, especially spring-drought in DGVM, also warrants further investigation. From our results for ~~Site~~plots 17 and 18, we notice that adjusting the climatic thresholds for the establishment~~growth~~ of boreal NET does not ~~automatically make~~necessarily lead to other PFTs grow. Boreal BDT and BDS can establish at both ~~sites~~plots, but their growth rates are too slow to make them occupy a large area at these ~~sites~~plots. This ~~prevents development of similarity with the PFT profiles of AR reference dataset (Fig. 4) and~~ implies that other environmental conditions, e.g., nitrogen availability, might play a more important role in limiting the growth of BDT and BDS in the tested DGVM ~~(CLM4.5BGCDV)~~. The biases of the DGVM~~s~~ in simulating ~~boreal~~ ~~broadleaf deciduous tree~~BDT and ~~shrub~~BDS has been widely noticed in previous~~other~~ studies (Castillo et al., 2012), and~~-~~ remains a challenge requiring more ~~should be~~ investigation in the future~~ed further~~.

While going into further details of which additional PFTs should be included in DGVMs and how these and other PFTs should be parameterised is beyond the scope of the present paper, we emphasize the potential of using DM for improving the parameters of DGVMs. More specifically, we propose more intensive exploration of DM as a tool for identification of potential environmental drivers for the high-latitude PFTs, which may enhance the performance of DGVMs in high-latitude ecoregions. The specific focus of our study is the boreal region~~s~~, both because of the importance of these ecosystems in the climate system and because of the data availability of vegetation-type DM and the field-based reference dataset (AR). However, we believe that the improved DGVM parameters resulting from our sensitivity experiments may be applicable to other DGVMs such as TEM and LPJ-GUESS (Euskirchen et al., 2009; Miller and Smith, 2012). Also, the results from this study are likely to be transferable to other high-latitude areas in the circumboreal region.

6 Conclusions

This study ~~emphasizes~~ demonstrates the potential of using distribution models (DM) for representing present-day vegetation in evaluations of plant functional type (PFT) distributions simulated by dynamic global vegetation models (DGVMs) and for improvement of specific PFT parameters within DGVMs. By identification of the main differences among PFT profiles obtained by three methods (DGVM, RS and DM) in selected high-latitude plots distributed across climatic gradients in Norway, we show that PFT profiles derived from DM and RS are in the same range of reliability, judged by resemblance to a reference dataset (AR). Hence, we suggest that DM results can be used as a complementary evaluation dataset to benchmark the present-day DGVMs. This approach is recommended when high-quality RS products are not available in desired thematic resolution or when they are not ~~.~~able to supply proxies of other properties (such as deriving parameter improvements or PFT-specific traits).

Comparing the twenty PFT profiles obtained by DGVM with those obtained by AR shows a large overestimation by DGVM of boreal needleleaf evergreen trees (boreal NET) and bare ground at the expense of boreal broadleaf deciduous trees and shrubs. This is attributed to missing processes and PFT parameterizations of high-latitude PFTs in DGVM. We use DM results to identify ~~three~~ a new PFT-specific environmental parameter – precipitation seasonality – ~~s~~which, in a series of sensitivity experiments, improves the distribution of boreal NET predicted by DGVM. Thi~~s~~e new PFT-specific threshold~~s~~ for establishment decreases the bias of boreal NET in DGVM across four out of six plots and as a result, the distribution of other high-latitude PFTs is also better represented. We argue that this~~ese~~ new threshold~~s~~ should be transferable to other DGVMs simulating high-latitude PFTs, and that our DM-based approach can be well applied~~transferre~~ to other ecosystems.

Further development of DGVM, such as refining parameters for existing boreal PFTs and increasing the thematic resolution of PFTs for boreal areas, should be strongly encouraged to achieve a more realistic simulation of the distribution of ~~actual~~ vegetation by DGVM, to increase the reliability of future predictions, and the reliability of predicted vegetation feedbacks in the climate system.

7 Acknowledgements

NIBIO is acknowledged for providing access to the area-frame survey AR18X18 dataset. UNINET Sigma2 is acknowledged for providing computing facilities. Geir-Harald Strand is acknowledged for providing scientific assistance and Michal Torma for providing technical assistance.

8 Data availability.

The model scripts for running the DGVM are available in the GitHub repository https://github.com/huitang-earth/Horvath_etal_BG2020, while -the script used to carry out the analysis of~~in~~ this study ~~are~~ is available in the GitHub repository https://github.com/geco-nhm/DGVM_RS_DM_Norway. High-resolution DM-based and RS-based PFT maps are available for download at the Dryad Digital Repository https://doi.org/10.5061/dryad.dfn2z34xn ~~)from the authors on request~~ (Fig. ~~S8~~S9). DGVM outputs are provided in the Table S10~~9~~, Table S13~~2~~ and Fig. S1~~0~~1.

9 Author contributions.

All authors have contributed to conceptualizing the research idea. PH curated the data and was responsible for the distribution modelling and for compiling and analysing the data from all methods. HT carried out the modelling and sensitivity tests using the DGVM (CLM4.5-BGCDV). PH together with AB, RH and HT were responsible for writing, with all authors contributing to reviewing and editing the paper. FS, AB, TKB and LMT acquired funding for this research.

10 Competing interests.

The authors declare that they have no conflict of interest.

11 Financial support.

This work forms a contribution to LATICE (https://www.mn.uio.no/latice), which is a Strategic Research Initiative funded by the Faculty of Mathematics and Natural Sciences at the University of Oslo (UiO/GEO103920). It is also part of the EMERALD project (294948) funded by the Research Council of Norway.

12 References:

Alexander, R., and Millington, A. C.: Vegetation mapping: From Patch to Planet, Vegetation Mapping, John Wiley & Sons, LTD, Chichester, England, 321-331 pp., 2000.

Álvarez-Martínez, J. M., Jiménez-Alfaro, B., Barquín, J., Ondiviela, B., Recio, M., Silió-Calzada, A., and Juanes, J. A.: Modelling the area of occupancy of habitat types with remote sensing, Methods in Ecology and Evolution, 9, 580-593, https://doi.org/10.1111/2041-210X.12925, 2018.

Assal, T. J., Anderson, P. J., and Sibold, J.: Mapping forest functional type in a forest-shrubland ecotone using SPOT imagery and predictive habitat distribution modelling, Remote Sensing Letters, 6, 755-764, https://doi.org/10.1080/2150704x.2015.1072289, 2015.

Bakkestuen, V., Erikstad, L., and Halvorsen, R.: Step-less models for regional environmental variation in Norway, J. Biogeogr., 35, 1906-1922, https://doi.org/10.1111/j.1365-2699.2008.01941.x, 2008.

Bjordal, J.: Potential Implications of Lichen Cover for the Surface Energy Balance: Implementing Lichen as a new Plant Functional Type in the Community Land Model (CLM4.5), Master Thesis, Department of Geosciences, University of Oslo, Oslo, 99 pp., 2018.

Bohn, U., Gollub, G., Hettwer, C., Neuhäuslova, Z., Raus, T., Schlüter, H., and Weber, H.: Map of the Natural Vegetation of Europe. Scale 1 : 2 500 000., Federal Agency for Nature Conservation, Münster, 2000.

Bonan, G. B., Levis, S., Kergoat, L., and Oleson, K. W.: Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models, Global Biogeochemical Cycles, 16, 5-1-5-23, https://doi.org/10.1029/2000gb001360, 2002.

Bonan, G. B., Levis, S., Sitch, S., Vertenstein, M., and Oleson, K. W.: A dynamic global vegetation model for use with climate models: concepts and description of simulated vegetation dynamics, Global Change Biol., 9, 1543-1566, https://doi.org/10.1046/j.1365-2486.2003.00681.x, 2003.

Bonan, G. B.: Forests, Climate, and Public Policy: A 500-Year Interdisciplinary Odyssey, Annual Review of Ecology, Evolution, and Systematics, 47, 97-121, https://doi.org/10.1146/annurev-ecolsys-121415-032359, 2016.

Bryn, A., Dramstad, W., Fjellstad, W., and Hofmeister, F.: Rule-based GIS-modelling for management purposes: A case study from the islands of Froan, Sør-Trøndelag, mid-western Norway, Norsk Geografisk Tidsskrift, 64, 175-184, https://doi.org/10.1080/00291951.2010.528224, 2010.

Bryn, A., Dourojeanni, P., Hemsing, L. Ø., and O'Donnell, S.: A high-resolution GIS null model of potential forest expansion following land use changes in Norway, Scand. J. For. Res., 28, 81-98, https://doi.org/10.1080/02827581.2012.689005, 2013.

Bryn, A., Strand, G.-H., Angeloff, M., and Rekdal, Y.: Land cover in Norway based on an area frame survey of vegetation types, Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 72, 1-15, https://doi.org/10.1080/00291951.2018.1468356, 2018.

Chadburn, S. E., Burke, E. J., Essery, R. L. H., Boike, J., Langer, M., Heikenfeld, M., Cox, P. M., and Friedlingstein, P.: Impact of model developments on present and future simulations of permafrost in a global land-surface model, The Cryosphere, 9, 1505-1521, 10.5194/tc-9-1505-2015, 2015.

Coppell, R., Gloor, E., and Holden, J.: A process-based Sphagnum plant-functional-type model for implementation in the TRIFFID Dynamic Global Vegetation Model, Geosci. Model Dev. Discuss., 2019, 1-44, 10.5194/gmd-2019-51, 2019.

Czekanowski, J.: Zur differentialdiagnose der Neandertalgruppe, Friedr. Vieweg & Sohn, 1909.

Dallmeyer, A., Claussen, M., and Brovkin, V.: Harmonising plant funtional type distributions for evaluating Earth System Models, Climate of the Past, 15, 335-366, https://doi.org/10.5194/cp-15-335-2019, 2019.

Davin, E. L., and de Noblet-Ducoudré, N.: Climatic Impact of Global-Scale Deforestation: Radiative versus Nonradiative Processes, J. Clim., 23, 97-112, https://doi.org/10.1175/2009jcli3102.1, 2010.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., Price, L., Baillie, J. E. M., Weeden, D., Suckling, K., Davis, C., Sizer, N., Moore, R., Thau, D., Birch, T., Potapov, P., Turubanova, S., Tyukavina, A., de Souza, N., Pintea, L., Brito, J. C., Llewellyn, O. A., Miller, A. G., Patzelt, A., Ghazanfar, S. A., Timberlake, J., Kloser, H., Shennan-Farpon, Y., Kindt, R., Lilleso, J. B., van Breugel, P., Graudal, L., Voge, M., Al-Shammari, K. F., and Saleem, M.: An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm, Bioscience, 67, 534-545, https://doi.org/10.1093/biosci/bix014, 2017.

Druel, A., Peylin, P., Krinner, G., Ciais, P., Viovy, N., Peregon, A., Bastrikov, V., Kosykh, N., and Mironycheva-Tokareva, N.: Towards a more detailed representation of high-latitude vegetation in the global land surface model ORCHIDEE (ORC-HL-VEGv1.0), Geoscientific Model Development, 10, 4693-4722, https://doi.org/10.5194/gmd-10-4693-2017, 2017.

Druel, A., Ciais, P., Krinner, G., and Peylin, P.: Modeling the Vegetation Dynamics of Northern Shrubs and Mosses in the ORCHIDEE Land Surface Model, Journal of Advances in Modeling Earth Systems, 11, 2020-2035, 10.1029/2018ms001531, 2019.

Duveiller, G., Hooker, J., and Cescatti, A.: The mark of vegetation change on Earth's surface energy balance, Nature Communications, 9, 679, https://doi.org/10.1038/s41467-017-02810-8, 2018.

Dyrrdal, A. V., Stordal, F., and Lussana, C.: Evaluation of summer precipitation from EURO-CORDEX fine-scale RCM simulations over Norway, International Journal of Climatology, 38, 1661-1677, https://doi.org/10.1002/joc.5287, 2018.

Eurostat: The Lucas Survey: European Statisticians Monitor Territory, Office for Official Publications of the European Communities, Luxembourg, 2003.

Euskirchen, E. S., McGuire, A. D., Chapin III, F. S., Yi, S., and Thompson, C. C.: Changes in vegetation in northern Alaska under scenarios of climate change, 2003–2100: implications for climate feedbacks, Ecol. Appl., 19, 1022-1043, 10.1890/08-0806.1, 2009.

Ferrier, S., Watson, G., Pearce, J., and Drielsma, M.: Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling, Conserv. Biol., 11, 2275-2307, https://doi.org/10.1023/a:1021302930424, 2002.

Ferrier, S., and Guisan, A.: Spatial modelling of biodiversity at the community level, J. Appl. Ecol., 43, 393-404, https://doi.org/10.1111/j.1365-2664.2006.01149.x, 2006.

Fielding, A. H., and Bell, J. F.: A review of methods for the assessment of prediction errors in conservation presence/absence models, Environ. Conserv., 24, 38-49, 1997.

Fisher, R., McDowell, N., Purves, D., Moorcroft, P., Sitch, S., Cox, P., Huntingford, C., Meir, P., and Ian Woodward, F.: Assessing uncertainties in a second-generation dynamic vegetation model caused by ecological scale limitations, New Phytol., 187, 666-681, https://doi.org/10.1111/j.1469-8137.2010.03340.x, 2010.

Franklin, S. E., and Wulder, M. A.: Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas, Progress in Physical Geography, 26, 173-205, https://doi.org/10.1191/0309133302pp332ra, 2002.

Førland, E.: Precipitation and topography [in Norwegian with English summary], Klima, 79, 23–24, 1979.

Gotangco Castillo, C. K., Levis, S., and Thornton, P.: Evaluation of the New CNDV Option of the Community Land Model: Effects of Dynamic Vegetation and Interactive Nitrogen on CLM4 Means and Variability, J. Clim., 25, 3702-3714, https://doi.org/10.1175/jcli-d-11-00372.1, 2012.

Halvorsen, R.: A gradient analytic perspective on distribution modelling, Sommerfeltia, 35, 1-165, https://doi.org/10.2478/v10208-011-0015-3, 2012.

Hanssen-Bauer, I., Førland, E., Haddeland, I., Hisdal, H., Lawrence, D., Mayer, S., Nesje, A., Nilsen, J., Sandven, S., and Sandø, A.: Climate in Norway 2100–A knowledge base for climate adaptation, The Norwegian Centre for Climate Services, The Norwegian Centre for Climate Services, 2017.

Hartley, A. J., MacBean, N., Georgievski, G., and Bontemps, S.: Uncertainty in plant functional type distributions and its impact on land surface models, Remote Sens. Environ., 203, 71-89, https://doi.org/10.1016/j.rse.2017.07.037, 2017.

Hemsing, L. Ø., and Bryn, A.: Three methods for modelling potential natural vegetation (PNV) compared: A methodological case study from south-central Norway, Norsk Geografisk Tidsskrift - Norwegian Journal of Geography, 66, 11-29, https://doi.org/10.1080/00291951.2011.644321, 2012.

Henderson, E. B., Ohmann, J. L., Gregory, M. J., Roberts, H. M., and Zald, H.: Species distribution modelling for plant communities: stacked single species or multivariate modelling approaches?, Applied Vegetation Science, 17, 516-527, https://doi.org/10.1111/avsc.12085 2014.

Hengl, T., Walsh, M. G., Sanderman, J., Wheeler, I., Harrison, S. P., and Prentice, I. C.: Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential, PeerJ, 6, e5457, https://doi.org/10.7717/peerj.5457, 2018.

Hickler, T., Vohland, K., Feehan, J., Miller, P. A., Smith, B., Costa, L., Giesecke, T., Fronzek, S., Carter, T. R., Cramer, W., Kuhn, I., and Sykes, M. T.: Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model, Global Ecol. Biogeogr., 21, 50-63, https://doi.org/10.1111/j.1466-8238.2010.00613.x, 2012.

Horvath, P., Halvorsen, R., Stordal, F., Tallaksen, L. M., Tang, H., and Bryn, A.: Distribution modelling of vegetation types based on area frame survey data, Applied Vegetation Science, 22, 547-560, https://doi.org/10.1111/avsc.12451, 2019.

Johansen, B. E.: Satellittbasert vegetasjonskartlegging for Norge, Direktoratet for Naturforvaltning, Norsk Romsenter, 2009.

Keith, D. A., Ferrer, J. R., Nicholson, E., Bishop, M. J., Polidoro, B. A., Llodra, E. R., Tozer, M. G., Nel, J. L., Nally, R. M., Gregr, E. J., Watermeyer, K. E., Essl, F., Faber-Langendoen, D., Franklin, J., Lehmann, C. E. R., Etter, A., Roux, D. J., Stark, J. S., Rowland, J. A., Brummitt, N. A., Fernandez-Arcaya, U. C., Suthers, I. M., Wiser, S. K., Donohue, I., Jackson, L. J., Pennington, R. T., Pettorelli, N., Andrade, A., Kontula, T., Lindgaard, A., Tahvanainan, T., Terauds, A., Venter, O., Watson, J. E. M., Chadwick, M. A., Murray, N. J., Moat, J., Pliscoff, P., Zager, I., and Kingsford, R. T.: The IUCN Global Ecosystem Typology v1.01: Descriptive profiles for Biomes and Ecosystem Functional Groups, IUCN, CEM, New York, 172, 2020.

Lantz, T. C., Gergel, S. E., and Kokelj, S. V.: Spatial Heterogeneity in the Shrub Tundra Ecotone in the Mackenzie Delta Region, Northwest Territories: Implications for Arctic Environmental Change, Ecosystems, 13, 194-204, 10.1007/s10021-009-9310-0, 2010.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z. L., Levis, S., and Sakaguchi, K.: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, Journal of Advances in Modeling Earth Systems, 3, https://doi.org/10.1029/2011MS00045, 2011.

Lawrence, P. J., and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), Journal of Geophysical Research, 112, n/a-n/a, https://doi.org/10.1029/2006jg000168, 2007.

Levis, S., Bonan, B., Vertenstein, M., and Oleson, K.: The community land model's dynamic global vegetation model (CLM-DGVM): technical description and user's guide, National Center for Atmospheric Research, Boulder, Colorado, 2004.

Li, W., Ciais, P., MacBean, N., Peng, S., Defourny, P., and Bontemps, S.: Major forest changes and land cover transitions based on plant functional types derived from the ESA CCI Land Cover product, International Journal of Applied Earth Observation and Geoinformation, 47, 30-39, https://doi.org/10.1016/j.jag.2015.12.006, 2016.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R. A., and Peng, S.: Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015), Earth Syst. Sci. Data, 10, 219-234, https://doi.org/10.5194/essd-10-219-2018, 2018.

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J.: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, Earth System Science Data, 10, 235-249, https://doi.org/10.5194/essd-10-235-2018, 2018a.

Lussana, C., Tveito, O., and Uboldi, F.: Three-dimensional spatial interpolation of 2 m temperature over Norway, Quarterly Journal of the Royal Meteorological Society, 144, 344-364, https://doi.org/10.1002/qj.3208, 2018b.

Majasalmi, T., Eisner, S., Astrup, R., Fridman, J., and Bright, R. M.: An enhanced forest classification scheme for modeling vegetation–climate interactions based on national forest inventory data, Biogeosciences, 15, 399-412, 10.5194/bg-15-399-2018, 2018.

Miller, P. A., and Smith, B.: Modelling Tundra Vegetation Response to Recent Arctic Warming, Ambio, 41, 281-291, https://doi.org/10.1007/s13280-012-0306-1, 2012.

Moen, A.: Vegetation, Norwegian Mapping Authority, Hønefoss, 200 s. ill. 234 cm pp., 1999.

Mücher, C. A., Hennekens, S. M., Bunce, R. G. H., Schaminée, J. H. J., and Schaepman, M. E.: Modelling the spatial distribution of Natura 2000 habitats across Europe, Landscape Urban Plann., 92, 148-159, https://doi.org/10.1016/j.landurbplan.2009.04.003, 2009.

Myers-Smith, I. H., Forbes, B. C., Wilmking, M., Hallinger, M., Lantz, T., Blok, D., Tape, K. D., Macias-Fauria, M., Sass-Klaassen, U., Lévesque, E., Boudreau, S., Ropars, P., Hermanutz, L., Trant, A., Collier, L. S., Weijers, S., Rozema, J., Rayback, S. A., Schmidt, N. M., Schaepman-Strub, G., Wipf, S., Rixen, C., Ménard, C. B., Venn, S., Goetz, S., Andreu-Hayles, L., Elmendorf, S., Ravolainen, V., Welker, J., Grogan, P., Epstein, H. E., and Hik, D. S.: Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities, Environmental Research Letters, 6, 045509, https://doi.org/10.1088/1748-9326/6/4/045509, 2011.

Myers-Smith, I. H., Kerby, J. T., Phoenix, G. K., Bjerke, J. W., Epstein, H. E., Assmann, J. J., John, C., Andreu-Hayles, L., Angers-Blondin, S., Beck, P. S. A., Berner, L. T., Bhatt, U. S., Bjorkman, A. D., Blok, D., Bryn, A., Christiansen, C. T., Cornelissen, J. H. C., Cunliffe, A. M., Elmendorf, S. C., Forbes, B. C., Goetz, S. J., Hollister, R. D., de Jong, R., Loranty, M. M., Macias-Fauria, M., Maseyk, K., Normand, S., Olofsson, J., Parker, T. C., Parmentier, F.-J. W., Post, E., Schaepman-Strub, G., Stordal, F., Sullivan, P. F., Thomas, H. J. D., Tømmervik, H., Treharne, R., Tweedie, C. E., Walker, D. A., Wilmking, M., and Wipf, S.: Complexity revealed in the greening of the Arctic, Nature Climate Change, 10, 106-117, https://doi.org/10.1038/s41558-019-0688-1, 2020.

O'Donnell, M. S., and Ignizio, D. A.: Bioclimatic predictors for supporting ecological applications in the conterminous United States, 2012.

Oksanen, L.: Isolated occurrences of spruce, Picea abies, in northernmost Fennoscandia in relation to the enigma of continental mountain birch forests, Acta Bot. Fenn., 81-92, 1995.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., and Thornton, P. E.: Technical Description of version 4.5 of the Community Land Model (CLM), NCAR Earth System Laboratory Climate and Global Dynamics Division, BOULDER, COLORADO, USA, 2013.

Pebesma, E. J., and Bivand, R. S.: Classes and methods for spatial data in {R}, R News, 5, 9-13, 2005.

Porada, P., Ekici, A., and Beer, C.: Effects of bryophyte and lichen cover on permafrost soil temperature at large scale, The Cryosphere, 10, 2291-2315, 10.5194/tc-10-2291-2016, 2016.

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., and Zimmermann, N. E.: Plant functional type mapping for earth system models, Geoscientific Model Development, 4, 993-1010, https://doi.org/10.5194/gmd-4-993-2011, 2011.

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., Bontemps, S., Boettcher, M., Brockmann, C., Defourny, P., Hagemann, S., Herold, M., Kirches, G., Lamarche, C., Lederer, D., Ottlé, C., Peters, M., and Peylin, P.: 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, https://doi.org/10.5194/gmd-8-2315-2015, 2015.

Scheiter, S., Langan, L., and Higgins, S. I.: Next-generation dynamic global vegetation models: learning from community ecology, New Phytol., 198, 957-969, https://doi.org/10.1111/nph.12210, 2013.

Seo, H., and Kim, Y.: Interactive impacts of fire and vegetation dynamics on global carbon and water budget using Community Land Model version 4.5, Geoscientific Model Development, 12, 457-472, https://doi.org/10.5194/gmd-12-457-2019, 2019.

Sevanto, S., Suni, T., Pumpanen, J., Grönholm, T., Kolari, P., Nikinmaa, E., Hari, P., and Vesala, T.: Wintertime photosynthesis and water uptake in a boreal forest, Tree Physiology, 26, 749-757, https://doi.org/10.1093/treephys/26.6.749, 2006.

Shi, Y., Yu, M., Erfanian, A., and Wang, G.: Modeling the Dynamic Vegetation–Climate System over China Using a Coupled Regional Model, J. Clim., 31, 6027-6049, 10.1175/jcli-d-17-0191.1, 2018.

Simensen, T., Horvath, P., Erikstad, L., Bryn, A., Vollering, J., and Halvorsen, R.: Composite landscape predictors improve distribution models of ecosystem types, Divers. Distrib., https://doi.org/10.1111/ddi.13060, 2020.

Sitch, S., Huntingford, C., Gedney, N., Levy, P. E., Lomas, M., Piao, S. L., Betts, R., Ciais, P., Cox, P., Friedlingstein, P., Jones, C. D., Prentice, I. C., and Woodward, F. I.: Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs), Global Change Biol., 14, 2015-2039, https://doi.org/10.1111/j.1365-2486.2008.01626.x, 2008.

Snell, R. S., Huth, A., Nabel, J. E. M. S., Bocedi, G., Travis, J. M. J., Gravel, D., Bugmann, H., Gutiérrez, A. G., Hickler, T., Higgins, S. I., Reineking, B., Scherstjanoi, M., Zurbriggen, N., and Lischke, H.: Using dynamic vegetation models to simulate plant range shifts, Ecography, 37, 1184-1197, https://doi.org/10.1111/ecog.00580, 2014.

Song, X., Zeng, X., and Zhu, J.: Evaluating the tree population density and its impacts in CLM-DGVM, Advances in Atmospheric Sciences, 30, 116-124, https://doi.org/10.1007/s00376-012-1271-0, 2013.

Strand, G.-H.: The Norwegian area frame survey of land cover and outfield land resources, Norsk Geografisk Tidsskrift, 67, 24-35, https://doi.org/10.1080/00291951.2012.760001, 2013.

Ullerud, H. A., Bryn, A., and Klanderud, K.: Distribution modelling of vegetation types in the boreal–alpine ecotone, Applied Vegetation Science, 19, 528-540, https://doi.org/10.1111/avsc.12236, 2016.

Ullerud, H. A., Bryn, A., and Skånes, H.: Bridging theory and implementation – Testing an abstract classification system for practical mapping by field survey and 3D aerial photographic interpretation, Norsk Geografisk Tidsskrift, 73, 301-317, https://doi.org/10.1080/00291951.2020.1717595, 2020.

Vowles, T., Gunnarsson, B., Molau, U., Hickler, T., Klemedtsson, L., and Björk, R. G.: Expansion of deciduous tall shrubs but not evergreen dwarf shrubs inhibited by reindeer in Scandes mountain range, J. Ecol., 105, 1547-1561, https://doi.org/10.1111/1365-2745.12753, 2017.

Wielgolaski, F. E.: History and Environment of the Nordic Mountain Birch, in: Plant Ecology, Herbivory, and Human Impact in Nordic Mountain Birch Forests, edited by: Caldwell, M. M., Heldmaier, G., Jackson, R. B., Lange, O. L., Mooney, H. A., Schulze, E. D., Sommer, U., Wielgolaski, F. E., Karlsson, P. S., Neuvonen, S., and Thannheiser, D., Ecological Studies, Springer Berlin Heidelberg, Berlin, Heidelberg, 3-18, 2005.

Wullschleger, S. D., Epstein, H. E., Box, E. O., Euskirchen, E. S., Goswami, S., Iversen, C. M., Kattge, J., Norby, R. J., van Bodegom, P. 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., 114, 1-16, https://doi.org/10.1093/aob/mcu077, 2014.

Xie, Y., Sha, Z., and Yu, M.: Remote sensing imagery in vegetation mapping: a review, Journal of Plant Ecology, 1, 9-23, https://doi.org/10.1093/jpe/rtm005, 2008.

Zeng, X., Zeng, X., and Barlage, M.: Growing temperate shrubs over arid and semiarid regions in the Community Land Model–Dynamic Global Vegetation Model, Global Biogeochemical Cycles, 22, n/a-n/a, https://doi.org/10.1029/2007gb003014, 2008.

Zhang, W., Brandt, M., Tong, X., Tian, Q., and Fensholt, R.: Impacts of the seasonal distribution of rainfall on vegetation productivity across the Sahel, Biogeosciences, 15, 319-330, https://doi.org/10.5194/bg-15-319-2018, 2018.

Zhu, J., Zeng, X., Zhang, M., Dai, Y., Ji, D., Li, F., Zhang, Q., Zhang, H., and Song, X.: Evaluation of the New Dynamic Global Vegetation Model in CAS-ESM, Advances in Atmospheric Sciences, 35, 659-670, https://doi.org/10.1007/s00376-017-7154-7, 2018.

Zuur, A. F., Ieno, E. N., and Smith, G. M.: Analysing ecological data, in, Springer, New York, 163-178, 2007.

Supplementary information to the paper:

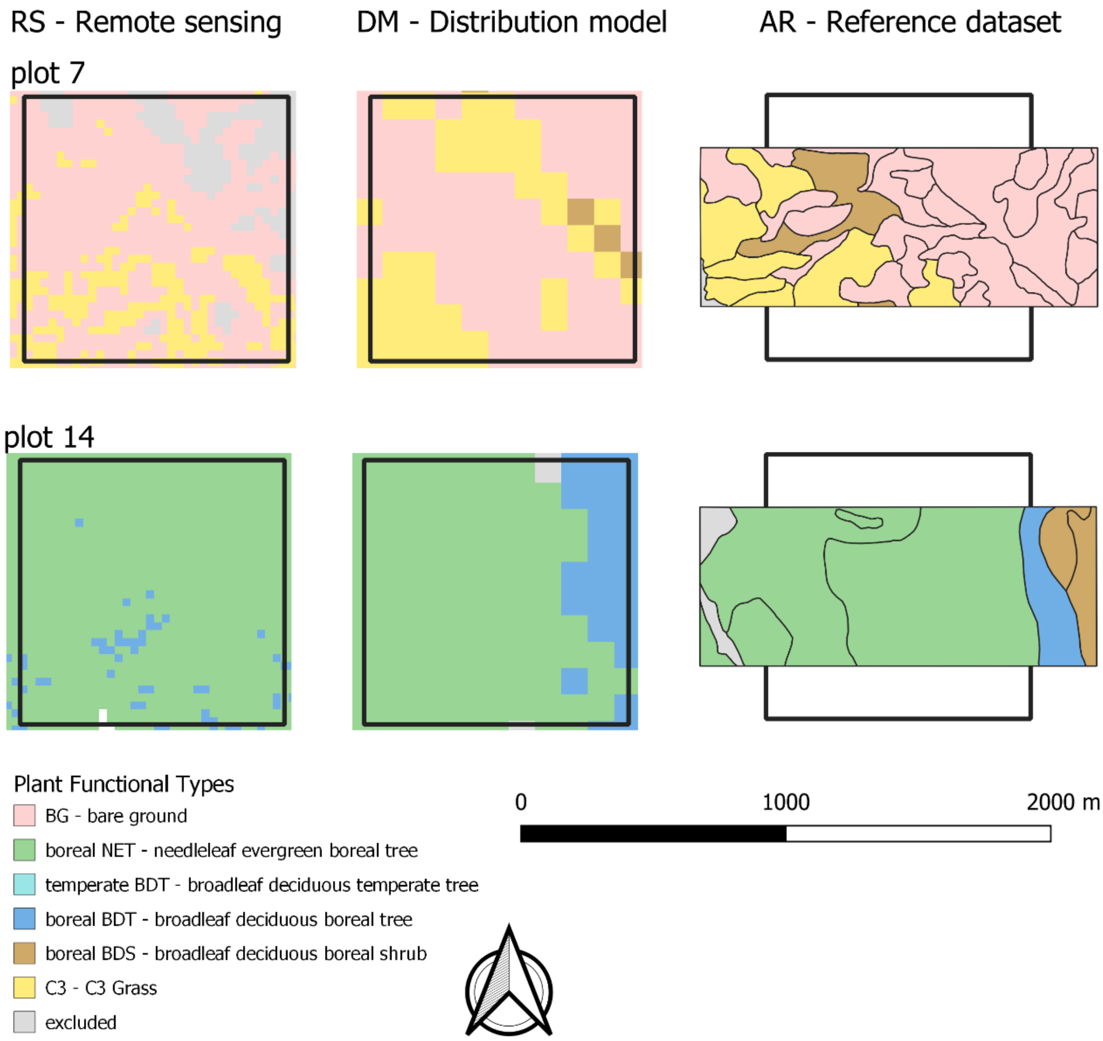
Horvath et al. Improving the representation of high-latitude vegetation distribution in Dynamic Global Vegetation Models

Supplement S1 – Locations of 20 study plots

Table S1 – Centre coordinates (latitude and longitude) and climatic data for the 20 plots used in this study. Estimates of mean annual precipitation and mean annual temperature are obtained from two sources; data from SeNorge (C. Lussana et al., 2018; Lussana, Tveito, & Uboldi, 2018) interpolated to each centrepoint and from CORDEX (the forcing climate dataset in DGVM).

					SeNorge v2 data (used in DM)		CORDEX climate data (used in DGVM)	
ID	Plot # from (AR18x18)	LAT	LONG	Elevation (m a.s.l) at centre	Mean Annual Precipitation (mm)	Mean Annual Temperature (°C)	Mean Annual Precipitation (mm)	Mean Annual Temperature (°C)
3	405	6.061	58.635	200	2662	6.3	2916	4.7
2	513	6.035	59.934	710	2628	1.0	3530	2.9
1	622	5.956	61.392	596	2520	2.0	2606	2.0
6	801	7.429	58.074	184	1542	6.7	2055	5.9
4	922	6.957	61.456	1437	1799	-3.6	2958	-2.9
5	1131	7.264	62.935	454	1976	4.0	1716	4.8
8	1304	8.862	58.638	88	1395	7.1	1640	4.9
7	1322	8.298	61.529	1670	827	-3.1	2418	-6.1
9	1623	9.278	61.735	852	555	-0.1	808	-3.9
10	2015	10.812	60.496	606	804	1.9	1517	0.5
12	2108	11.268	59.377	130	1072	5.5	1223	4.4
11	2238	11.000	64.223	222	1349	4.3	1542	2.1
13	2332	11.492	63.266	721	1029	0.3	2001	-0.2
14	2425	11.968	62.145	744	715	-1.2	1013	-2.0
16	2948	13.508	65.886	529	1513	1.1	1819	-0.3
15	2962	13.363	68.146	393	1339	5.8	1075	4.4
17	4268	19.167	69.072	354	715	0.7	1122	-1.8
18	5369	24.147	69.040	395	466	-4.0	695	-3.1
19	6473	29.382	69.334	69	503	-1.1	640	-2.5
20	6380	29.703	70.465	387	552	0.2	1132	-2.5

~~Supplement S2 – Sampling design – RS, DM and AR~~



~~**Figure S2 – Sampling design used by the remote sensing (RS) and distribution modelling (DM) methods and to obtain the AR reference dataset. Like DGVM plots (see Fig. S7), the RS and DM plots are 1×1 km, while the AR plots are 1.5×0.6 km. Plots 7 and plot 14 (AR18x18 plot #1322 and plot #2425) are used as examples.**~~

~~Supplement S3~~Supplement S2 – Assessment of climatic representativeness of selected plots

We assessed the representativeness of the 20 plots, selected from the original AR18×18 dataset which consists of 1081 plots, by comparing frequency distributions with respect to the two main bioclimatic gradients in Norway, expressed as annual mean temperature and annual precipitation. We also included a comparison of precipitation seasonality, as the only one of the three tested new parameters that improved the DGVM in the sensitivity tests. For each of temperature, ~~and~~ precipitation and precipitation seasonality, we obtained ~~interpolated~~ values for the centrepoint of each AR18×18 plot (cf. Fig. S1) and compared the frequency distributions of the selected plots with those of all plots (Fig. S3). A series of Kolmogorov-Smirnov tests for these three variables (comparison of sample mean and variance) indicate that the subsample does not deviate from the full dataset substantially. The 20 selected plots span elevations from 88 to 1670 m a.s.l., covers an annual temperature range from -4°C to 7.1°C, and an annual precipitation range from 466 to 2661 mm (Fig. S1), which accords well with the variation in the AR18×18 dataset (Fig. ~~S3~~S2).

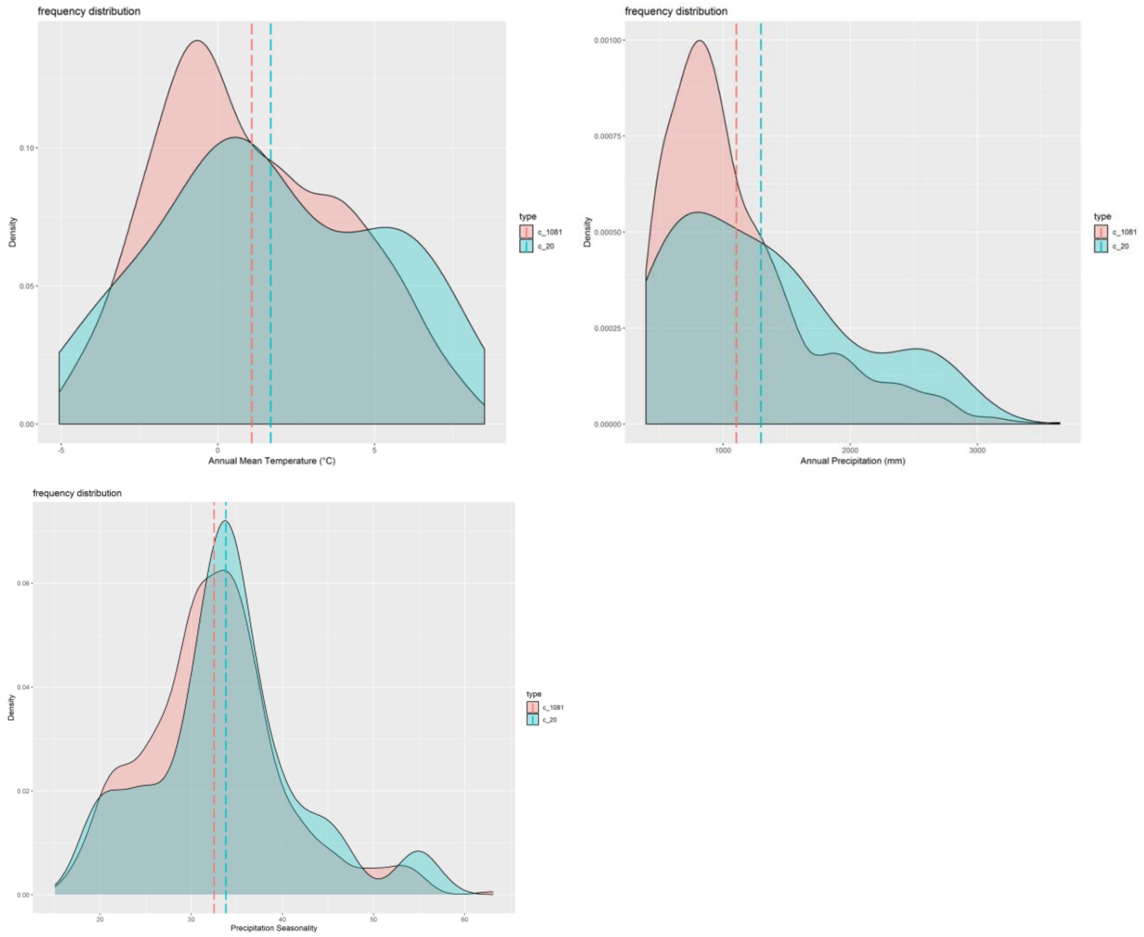


Figure ~~S3~~S2– Frequency distributions of plots in the original AR18×18 dataset (n=1081; in red) and in the set of 20 plots selected for this study (in blue), with respect to annual mean temperature (top left), ~~and~~ annual precipitation (top right) and precipitation seasonality (bottom left). Dashed lines indicate means for the respective datasets.

~~Supplement S4~~Supplement S3 – Assessment of the representativeness of PFT profiles

We also assessed the representativeness of the 20 study plots, selected from the original AR18×18 dataset which consists of 1081 plots, by comparing the aggregated PFT profiles for the two datasets given in Table S4. PFT profiles were first obtained for each plot by the conversion scheme in Table 2, thereafter aggregated to dataset level by calculation of mean frequencies for each of the six PFTs (and 'EXCL'; land not assigned to any PFT type).

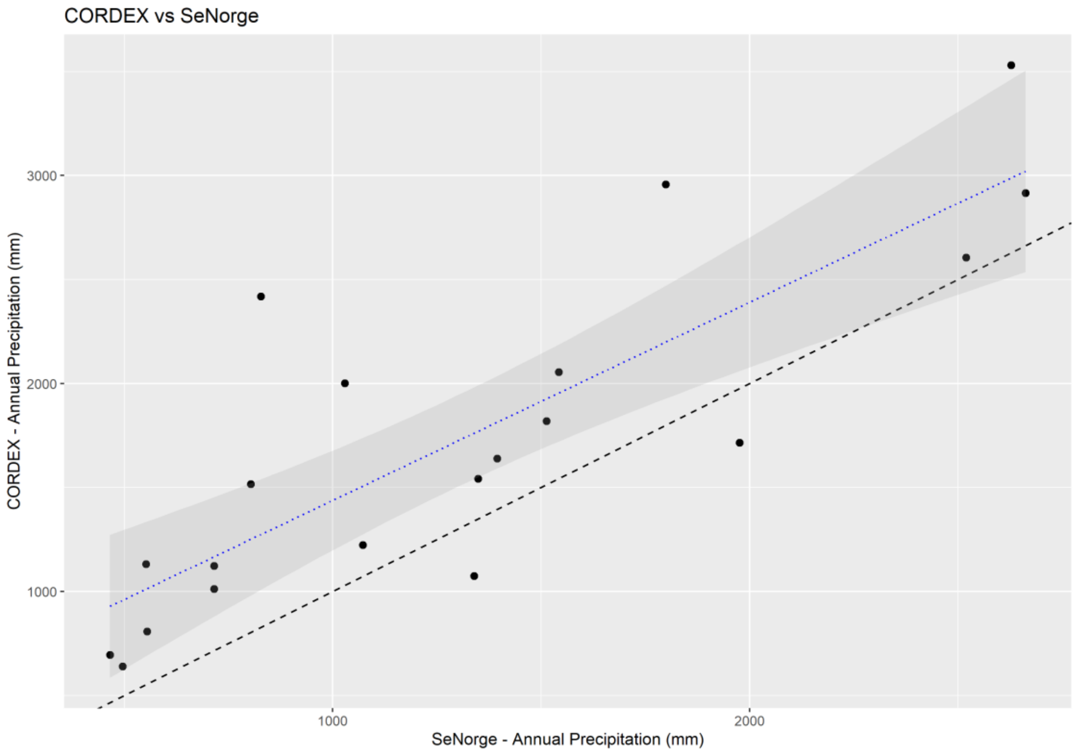
The comparison between the aggregated PFT profiles in Table S4 by use of the chi-square test (see section 2.6 for method) shows that the two datasets are much more similar than expected by chance (χ^2=1.991, df = 6, p = 0.079). Despite slight overrepresentation of the boreal NET PFT and underrepresentation of boreal BDT and C3 grasses, we conclude that the selected plots are sufficiently representative for the conclusions drawn from the sample of 20 plots to be acceptably representative for Norway. Note that percentage for EXCL category has been proportionally re-distributed through relevant PFTs in the study as shown on the Table 3 (so that the six PFTs cover 100%).

Table ~~S4~~ S3 – PFT profiles of the full AR18x18 dataset (n = 1081) and the 20 plots selected for this study.

PFT code	PFT name	Fraction of PFT in 1081 plots (%)	Fraction of PFT in 20 plots (%)
BG	Bare Ground	10.37	10.95
Boreal NET	needleleaf evergreen tree - boreal	21.50	31.18
Temp BDT	broadleaf deciduous tree - temperate	0.46	0.40
Boreal BDT	broadleaf deciduous tree - boreal	16.02	12.55
Boreal BDS	broadleaf deciduous shrub - boreal	25.11	24.35
C3	C3 grass	7.27	3.00
EXCL	excluded	19.27	17.57

~~Supplement S5~~Supplement S4 – Assessment of the representativeness of climate forcing data

The comparison of SeNorge and CORDEX estimates of temperature and precipitation in Fig. S5.1 shows that precipitation estimates by CORDEX for the 20 plots were generally higher than SeNorge estimates while the converse (but less strongly) was true for temperature.



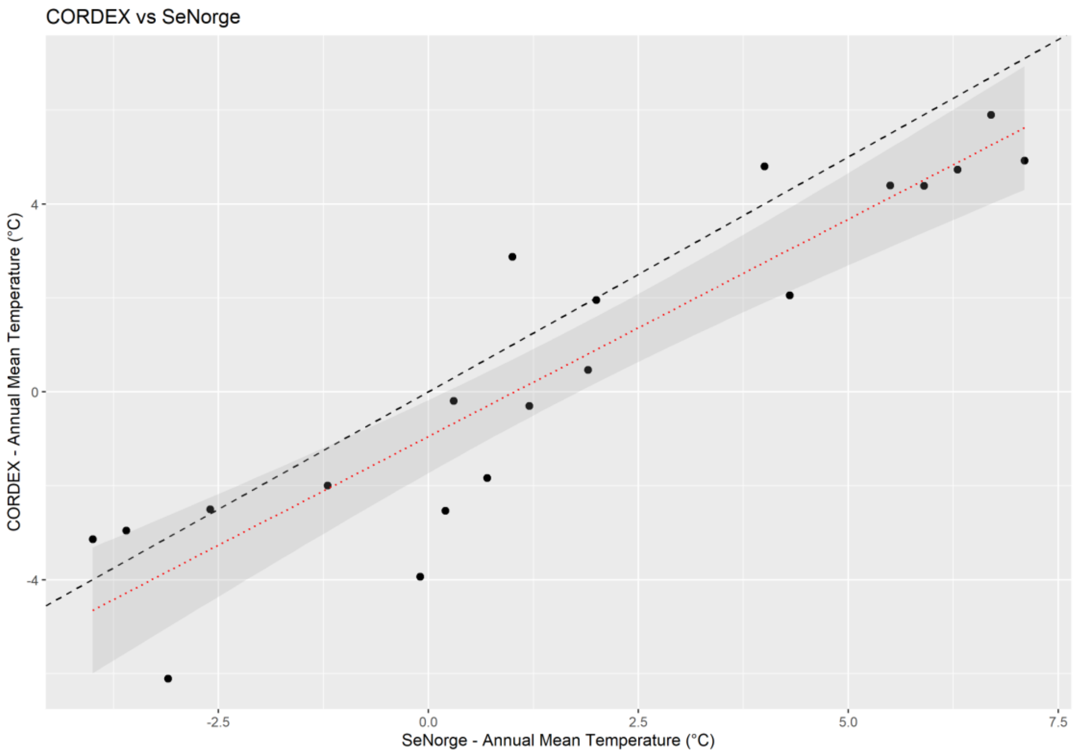


Figure ~~S5~~ S4 – Scatterplots showing the relationship between temperature and precipitation estimates obtained by the two data sources used in this study; SeNorge for DM (see Sect. 2.4.3) on the horizontal axes and CORDEX for climate forcing data used in DGVM (see Sect. 2.4.1) on the vertical axis. The dashed black line represents the 1:1 relationship, while the dotted red line represents a linear model of y~x.

Supplement S5 – PFT Conversion scheme

Table S5– Conversion scheme for harmonizing vegetation and land cover types across methods (RS, DM and AR) into plant functional types (PFTs). DGVM – dynamic global vegetation model, RS – remote sensing, DM – distribution model, AR – reference dataset. PFT – plant functional type and VT – vegetation type.

DGVM		RS	DM	AR
PFT	plant functional type	vegetation / land cover type – remote sensing	vegetation type – distribution model	vegetation type – area frame survey
BG	Bare ground	Exposed alpine ridges, scree and rock complex	Frozen ground, leeward	Frozen ground, leeward
			Frozen ground, ridge	Frozen ground, ridge
			Boulder field	Sand dunes and gravel beaches
			Exposed bedrock	Pioneer alluvial vegetation
				Barren land
				Boulder field
				Exposed bedrock
Boreal NET	Boreal needleleaf evergreen tree	Coniferous forest – dense canopy layer	Lichen and heather pine forest	Lichen and heather pine forest
		Coniferous forest and mixed forest - open canopy	Bilberry pine forest	Bilberry pine forest
		Lichen rich pine forest	Lichen & heather spruce forest	Meadow pine forest
			Bilberry spruce forest	Pine forest on lime soils
			Meadow spruce forest	Lichen & heather spruce forest
			Damp forest	Bilberry spruce forest
			Bog forest	Meadow spruce forest
				Damp forest
				Bog forest
Temperate BDT	Temperate broadleaf deciduous tree	Low herb forest and broadleaved deciduous forest	Poor / Rich broadleaf deciduous forest	Poor broadleaf deciduous forest
				Rich broadleaf deciduous forest
Boreal BDT	Boreal broadleaf deciduous tree	Tall herb - tall fern deciduous forest	Lichen and heather birch forest	Lichen and heather birch forest
		Bilberry- low fern birch forest	Bilberry birch forest	Bilberry birch forest
		Crowberry birch forest	Meadow birch forest	Meadow birch forest
		Lichen-rich birch forest	Alder forest	Birch forest on lime soils
			Pasture land forest	Alder forest
			Poor / rich swamp forest	Pasture land forest
				Poor swamp forest
				Rich swamp forest
Boreal BDS	Boreal broadleaf deciduous shrub	Heather-rich alpine ridge vegetation	Lichen heath	Lichen heath
		Lichen-rich heathland	Mountain avens heath	Mountain avens heath
		Heather- and grass-rich early snow patch communities	Dwarf shrub / Alpine calluna heath	Dwarf shrub heath
		Fresh heather and dwarf-shrub communities (u/l)	Alpine damp heath	Alpine calluna heath
			Coastal heath / Coastal calluna heath	Alpine damp heath
			Damp heath	Flood-plain shrubs
				Coastal heath
				Coastal calluna heath
				Damp heath
				Crags and thicket
C3	C3 grass	Graminoid alpine ridge vegetation	Moss snowbed / Sedge and grass snowbed	Moss snowbed
		Herb-rich meadows (up-/lowland)	Dry grass heath	Sedge and grass snowbed
		Grass and dwarf willow snow-patch vegetation	Low herb / forb meadow	Dry grass heath
				Low herb meadow

		_		Low forb meadow
		_	_	Moist and shore meadows
EXCL	Excluded	Ombrotrophic bog and low-grown swamp vegetation	Bog / Mud-bottom fen and bog	Bog
		Tall-grown swamp vegetation	Deer-grass fen / fen	Deer-grass fen
		Wet mires, sedge swamps and reed beds	Sedge marsh	Fen
		Glacier, snow and wet snow-patch vegetation	Pastures	Mud-bottom fen and bog
		Water		Sedge marsh
		Agricultural areas		Cultivated land
		Cities and built-up areas		Pastures
		Unclassified and shadow affected areas,		Built-up areas
		_		Scattered housing
		_		Artificial impediment
		_		Glaciers and perpetual snow
		_		Sea and ocean
		_	_	Water bodies (fresh)

Supplement S6 – Sampling design – RS, DM and AR

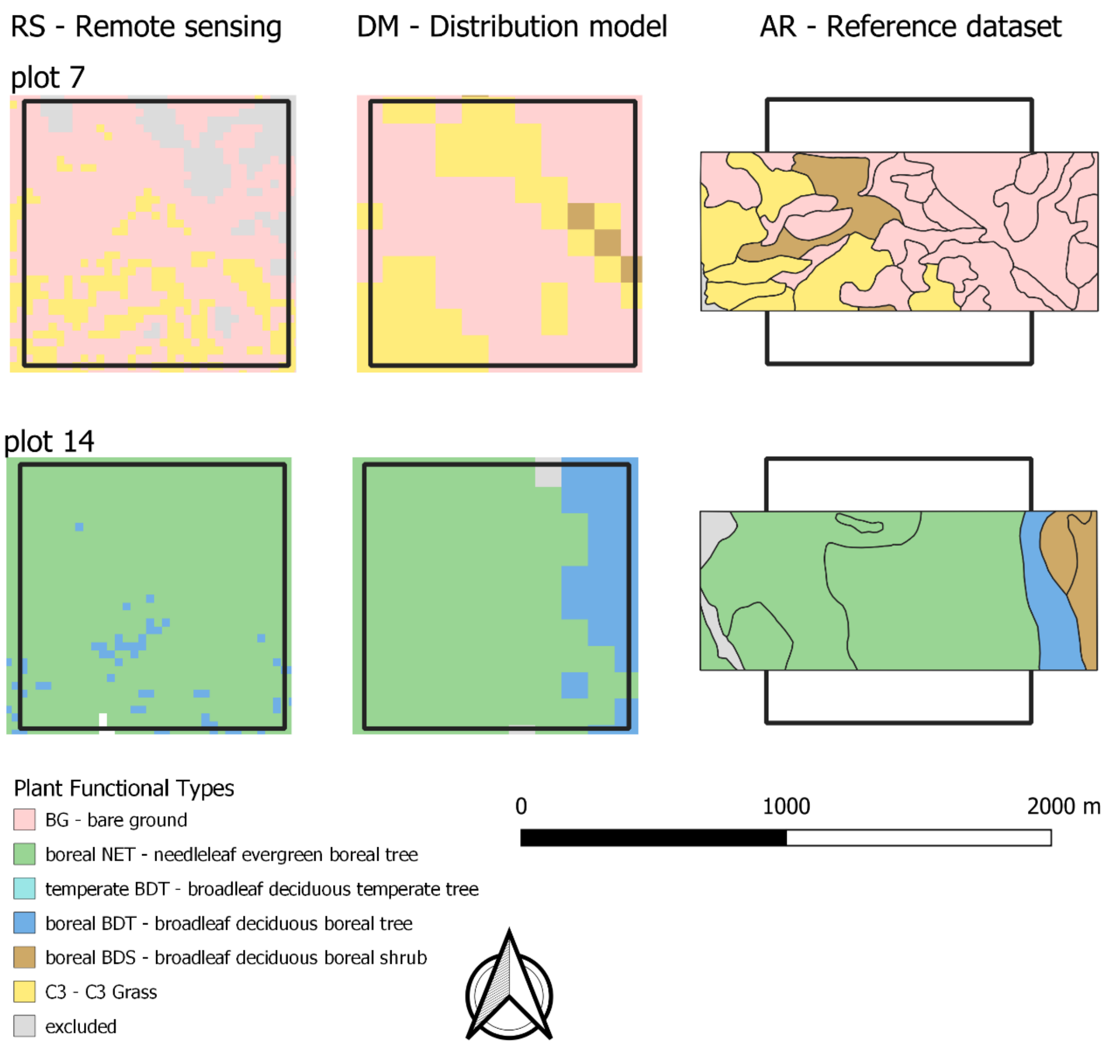


Figure S6 – Sampling design used by the remote sensing (RS) and distribution modelling (DM) methods and to obtain the AR reference dataset. Like DGVM plots (see Fig. S7), the RS and DM plots are 1×1 km, while the AR plots are 1.5×0.6 km. Plots 7 and plot 14 (AR18x18 plot #1322 and plot #2425) are used as examples.

~~Supplement S6~~Supplement S7 – DGVM parameters for PFTs (CLM4.5-BGCDV)

Table ~~S6~~ S7 – Some important PFT parameter settings for DGVM (CLM4.5-DV). PFTs relevant for the study area (Norway) are shaded grey~~in bold font~~. Prescribed heights for the canopy are indicated by the upper and lower limits in columns “ztop” and “zbot” respectively. Limiting temperatures for survival and establishment are mentioned in columns “Tc,min” and “Tc,max” respectively. Minimum growing degree days for establishment are contained for relevant PFTs in column “GDDmin”. The last three columns contain the adjusted parameters thresholds used in the sensitivity experiment. ~~Bioclim_15 – Precipitation Seasonality (Coefficient of Variation); SWE~~swe_10 – ~~S~~snow water equivalent in October (mm); ~~TMIN~~tmin_5 – ~~M~~minimum ~~T~~temperature in May (°C) bioclim_15 – precipitation seasonality (coefficient of variation);

		Prescribed heights		Survival	Establishment		Sensitivity tests		
Plant functional type (PFT)	Acronym	ztop (m)	zbot (m)	*T*c,min (°C)	*T*c,max (°C)	GDDmin	swe~~SWE~~_10 (mm)	~~TMIN~~tmin_5 (°C)	bioclim_15
Needleleaf evergreen tree – temperate	Temp NET	17	8.5	–2	22	900			
Needleleaf evergreen tree – boreal	Boreal NET	17	8.5	–32.5	–2	600	150	–5	50
Needleleaf deciduous tree – boreal	Boreal NDT	14	7						
Broadleaf evergreen tree – tropical	Trop BET	35	1	15.5	No limit	0			
Broadleaf evergreen tree – temperate	Temp BET	35	1	3	18.8	1200			
Broadleaf deciduous tree – tropical	Trop BDT	18	10	15.5	No limit	0			
Broadleaf deciduous tree – temperate	Temp BDT	20	11.5	–17	15.5	1200			
Broadleaf deciduous tree – boreal	Boreal BDT	20	11.5	No limit	–2	350	180	–7.5	
Broadleaf evergreen shrub – temperate	Temp BES	0.5	0.1						
Broadleaf deciduous shrub – temperate	Temp BDS	0.5	0.1	–17	No limit	1200			
Broadleaf deciduous shrub – boreal	Boreal BDS	0.5	0.1	No limit	–2	350	380	–10	
C3 arctic grass	C3 A	0.5	0.01	No limit	–17	0			
C3 grass	C3	0.5	0.01	–17	15.5	0			
C4 grass	C4	0.5	0.01	15.5	No limit	0			
Non vegetated/bare ground	BG								

~~Supplement S7~~Supplement S8 – Representation of grid-cells in the CLM 4.5 model

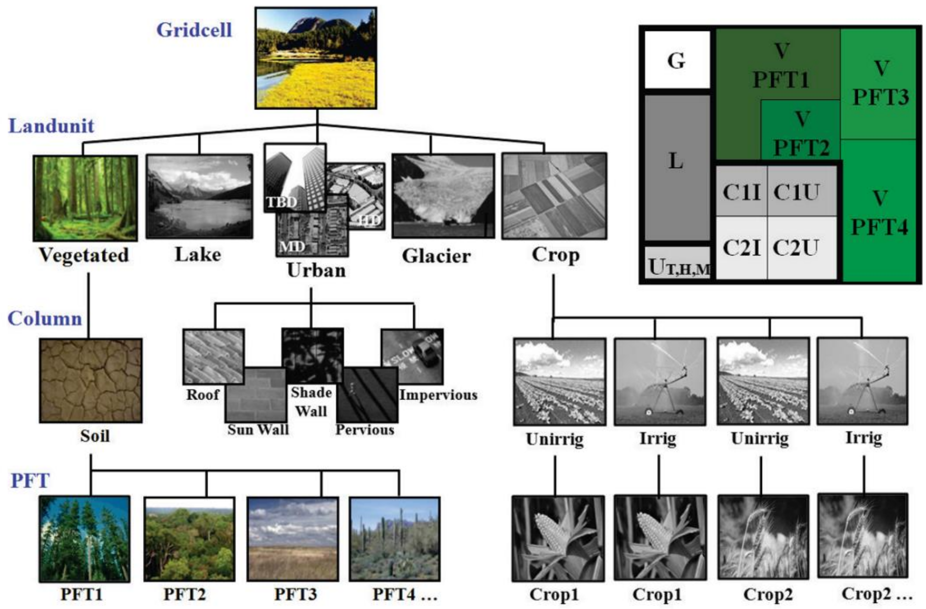


Figure ~~S7~~ S8 – Representation of a grid-cell in the DGVM model (obtained by CLM4.5-BGCDV method); figure adapted from Oleson et al. (2013). Land units in grey (lake, urban, glacier and crop) were excluded from this study.

~~Supplement S8~~Supplement S9 – DM- and RS-units reclassified to PFT units

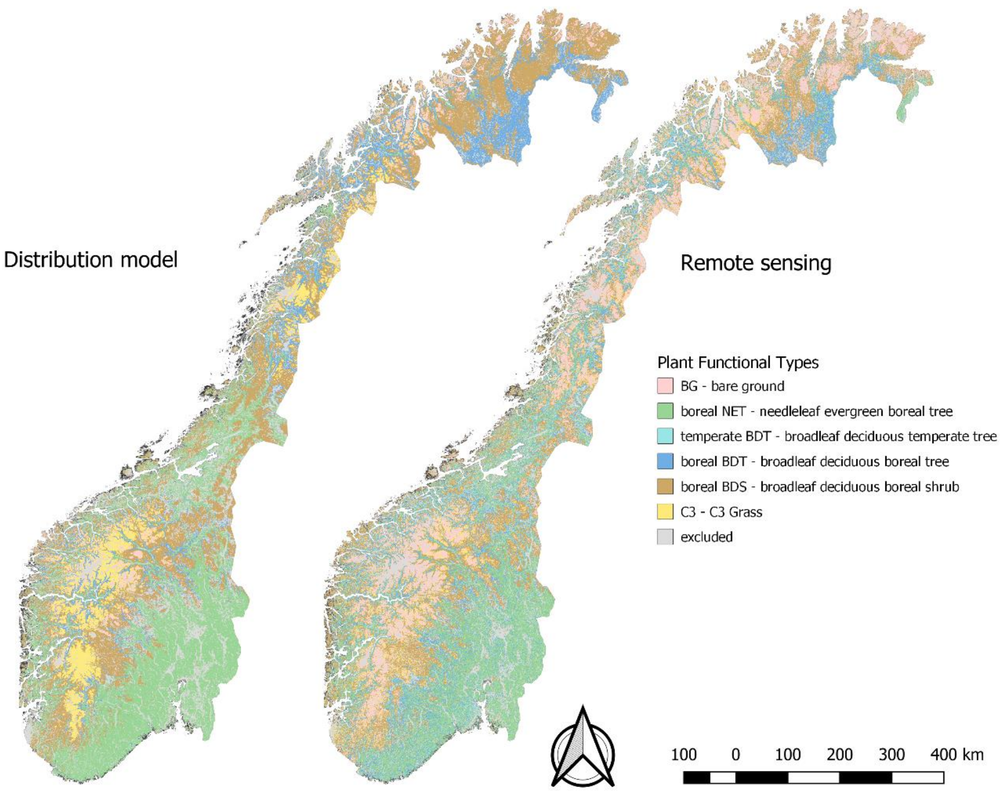


Figure ~~S8~~S9– The distribution in Norway of vegetation types (used in distribution modelling – DM) and units obtained by remote sensing (RS), after reclassification to PFT units (see Table 2 for conversion scheme and explanation of PFT codes). The dominating PFT in each grid cell (of 100×100 m for DM and 30×30 m for RS) is shown.

The distributions in Norway of PFTs obtained by conversion of DM- and RS-units using the conversion scheme in Table 2 exhibit considerable similarities (Fig. S8). Both methods show dominance of boreal needleleaf evergreen forest (boreal NET) in southeastern Norway, while most of the western and northern Norway is covered by boreal broadleaf deciduous shrub (boreal BDS) and boreal broadleaf deciduous forest (boreal BDT). Slight differences between the two methods can be seen in the western mountainous part of Norway, where DM predicts dominance by C3 grasses where RS suggests bare ground, and in North Norway where DM predicts boreal BDS where RS predicts bare ground. Accordingly, the fractional area classified to PFTs that are converted to bare ground is three times higher with RS than with DM (Table S8). Full resolution raster images are available at the Dryad repository (https://doi.org/10.5061/dryad.dfn2z34xn).

Table ~~S8~~ S9 – Area statistics for Norway for vegetation types (used in distribution modelling – DM) and units obtained by remote sensing (RS), after reclassification to PFT units (see Table 2 for conversion scheme and explanation of PFT codes).

	RS (%)	DM (%)
BG	17.1	5.6
Boreal NET	25.3	31.4
Temperate BDT	5.2	0.1
Boreal BDT	16.9	15.0
Boreal BDS	27.9	39.0
C3	7.5	8.9

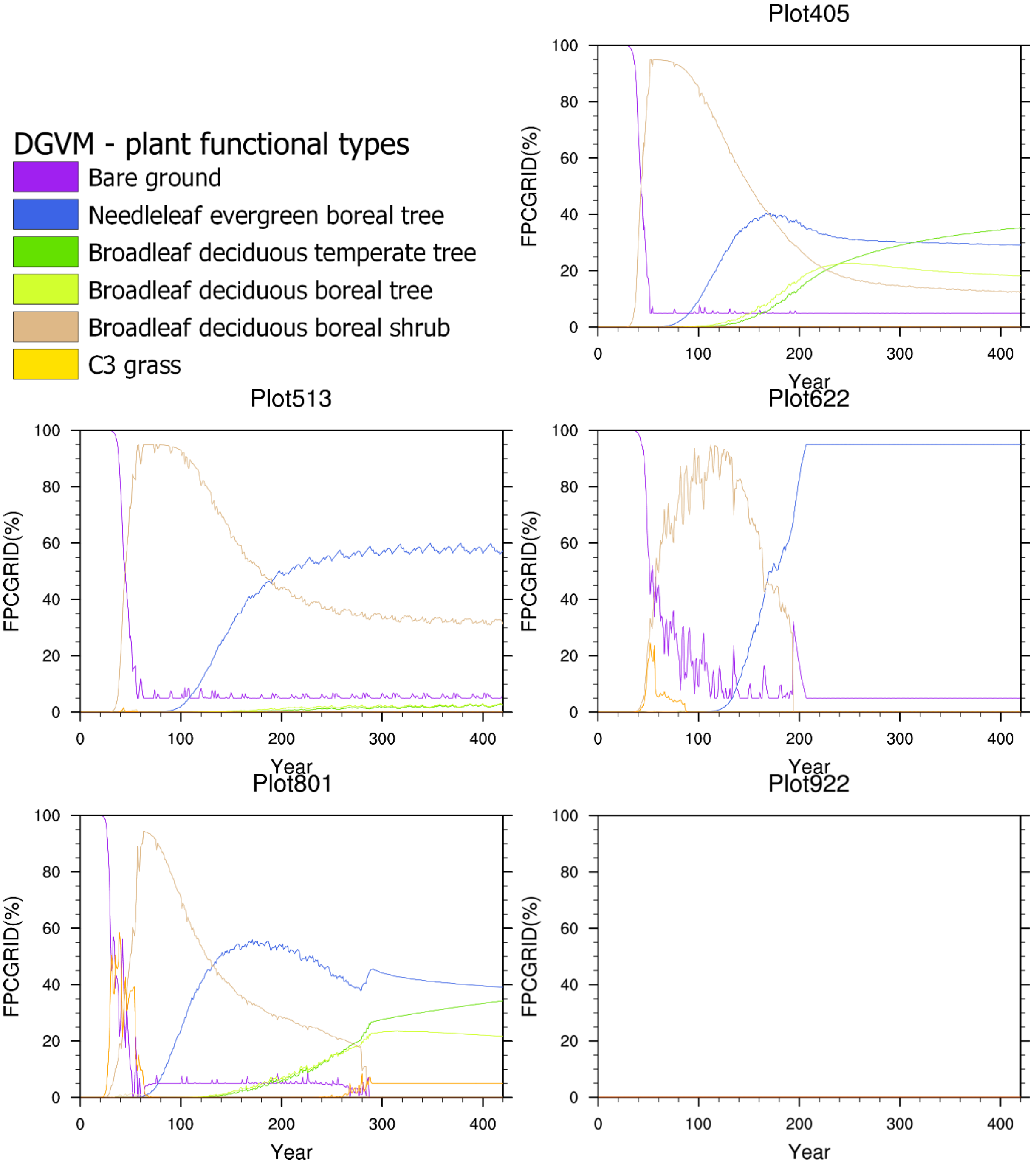
~~Supplement S9~~Supplement S10 – PFT profiles for each of the 20 plots

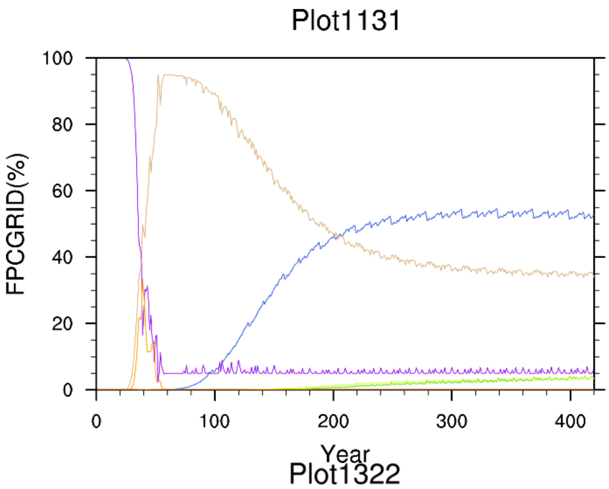
Table S10~~9~~– PFT profiles (percentage of vegetated land assigned to each of six PFTs) for each of the 20 plots in this study, obtained by remote sensing (RS) and distribution modelling (DM) methods and for the AR reference dataset. Original units (vegetation types, etc.) are converted to PFTs by use of the scheme in Table 2.

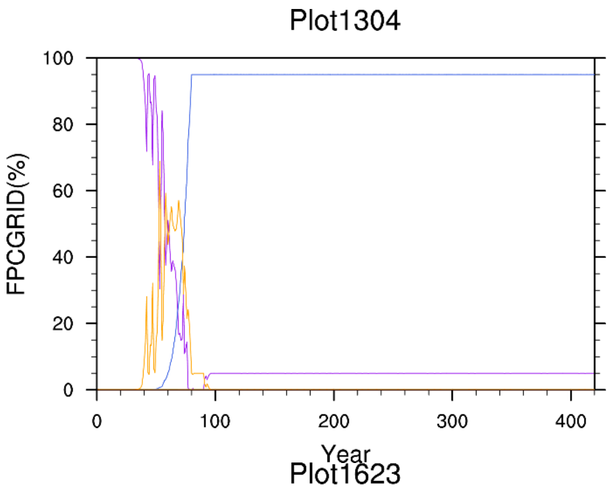
Method	PFT_shortcut	plot 3	plot 2	plot 1	plot 6	plot 4	plot 5	plot 8	plot 7	plot 9	plot 10	plot 12	plot 11	plot 13	plot 14	plot 16	plot 15	plot 17	plot 18	plot 19	plot 20
DGVM	BG	5	6	5	0	100	6	5	100	5	5	0	5	100	5	100	5	28	5	100	5
DGVM	boreal NET	29	58	95	39	0	52	95	0	95	95	41	95	0	95	0	92	72	95	0	95
DGVM	temp. BDT	35	2	0	34	0	4	0	0	0	0	38	0	0	0	0	0	0	0	0	0
DGVM	boreal BDT	18	2	0	22	0	4	0	0	0	0	16	0	0	0	0	0	0	0	0	0
DGVM	boreal BDS	13	32	0	0	0	35	0	0	0	0	0	0	0	0	0	3	0	0	0	0
DGVM	C3	0	0	0	5	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
RS	BG	9	7	4	0	92	8	0	78	0	0	0	0	7	3	24	52	0	1	54	1
RS	boreal NET	30	2	0	75	0	0	68	0	93	75	69	91	0	86	0	0	20	0	0	70
RS	temp. BDT	6	0	0	6	0	0	15	0	0	2	7	1	0	0	0	0	0	0	0	1
RS	boreal BDT	2	1	1	19	0	0	17	0	7	22	20	8	0	8	0	0	48	68	0	28
RS	boreal BDS	18	68	80	0	1	85	0	0	0	1	3	0	78	3	35	37	28	30	9	1
RS	C3	35	23	14	0	7	7	0	22	0	0	1	0	16	0	41	11	3	0	37	0
DM	BG	0	8	0	0	2	0	0	70	0	0	0	0	0	0	0	33	0	0	46	0
DM	boreal NET	60	1	0	100	0	0	96	0	47	100	100	100	0	72	0	0	0	0	0	0
DM	temp. BDT	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
DM	boreal BDT	0	0	0	0	0	0	0	0	53	0	0	0	0	23	0	0	77	91	0	100
DM	boreal BDS	40	91	100	0	0	100	0	3	0	0	0	0	100	4	100	63	23	9	54	0
DM	C3	0	0	0	0	98	0	0	26	0	0	0	0	0	0	0	4	0	0	0	0
AR	BG	0	4	0	0	87	0	0	66	0	0	0	0	0	0	11	13	0	0	78	0
AR	boreal NET	63	0	0	79	0	0	79	0	82	84	83	86	0	82	1	0	0	0	0	97
AR	temp. BDT	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
AR	boreal BDT	9	12	35	21	0	0	11	0	18	16	17	14	5	9	3	0	66	70	0	3
AR	boreal BDS	28	75	63	0	0	99	0	10	0	0	0	0	87	9	79	83	34	30	18	0
AR	C3	0	9	1	0	13	1	0	25	0	0	0	0	8	0	6	5	0	0	3	0

~~Supplement S10~~Supplement S11 – DGVM spin-up and simulation of PFT profiles for each plot

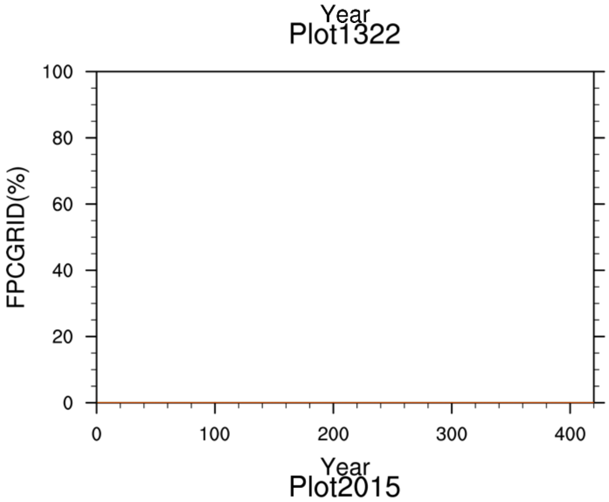
DGVM spin-up for 400 years and 20 years of simulation of PFT profiles for each of the 20 plots used in this study. For plots #801, #2108 and #4268, the spin-up was extended by additional 400, 200 and 200 years respectively.

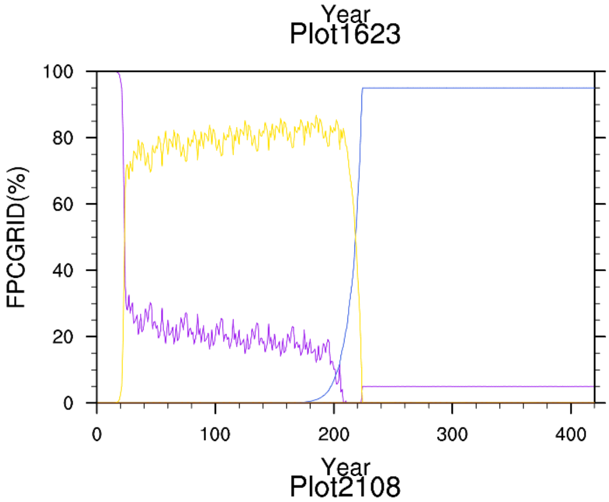




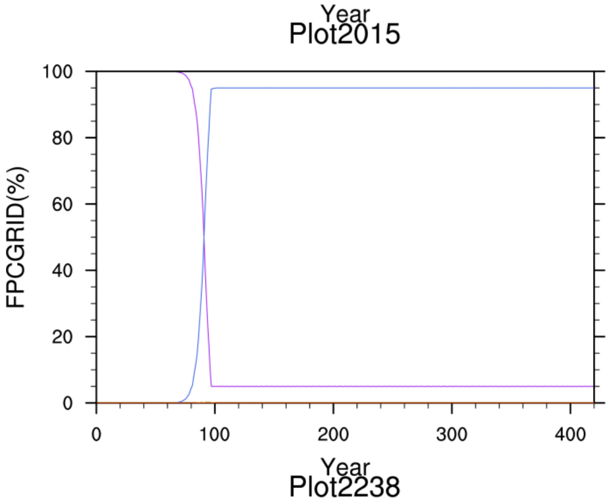


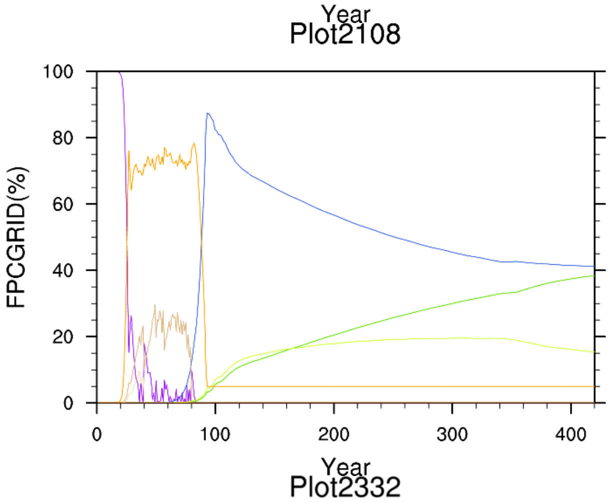
114



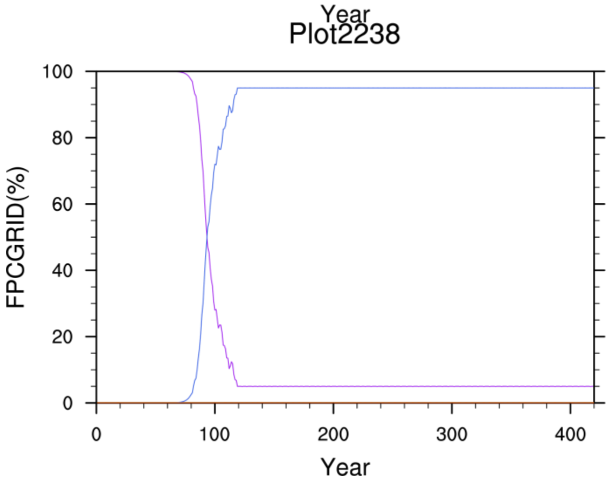


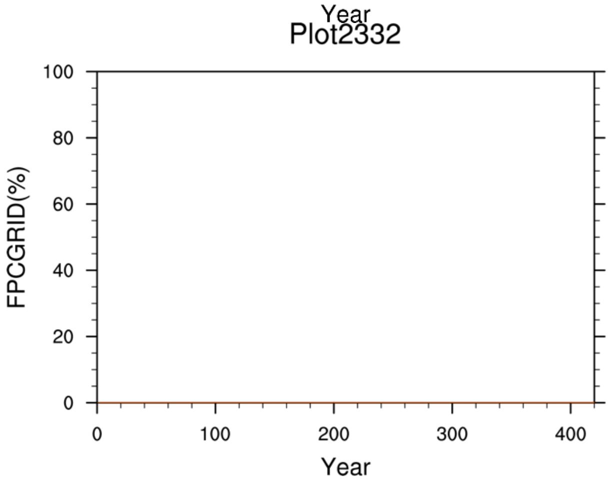
115



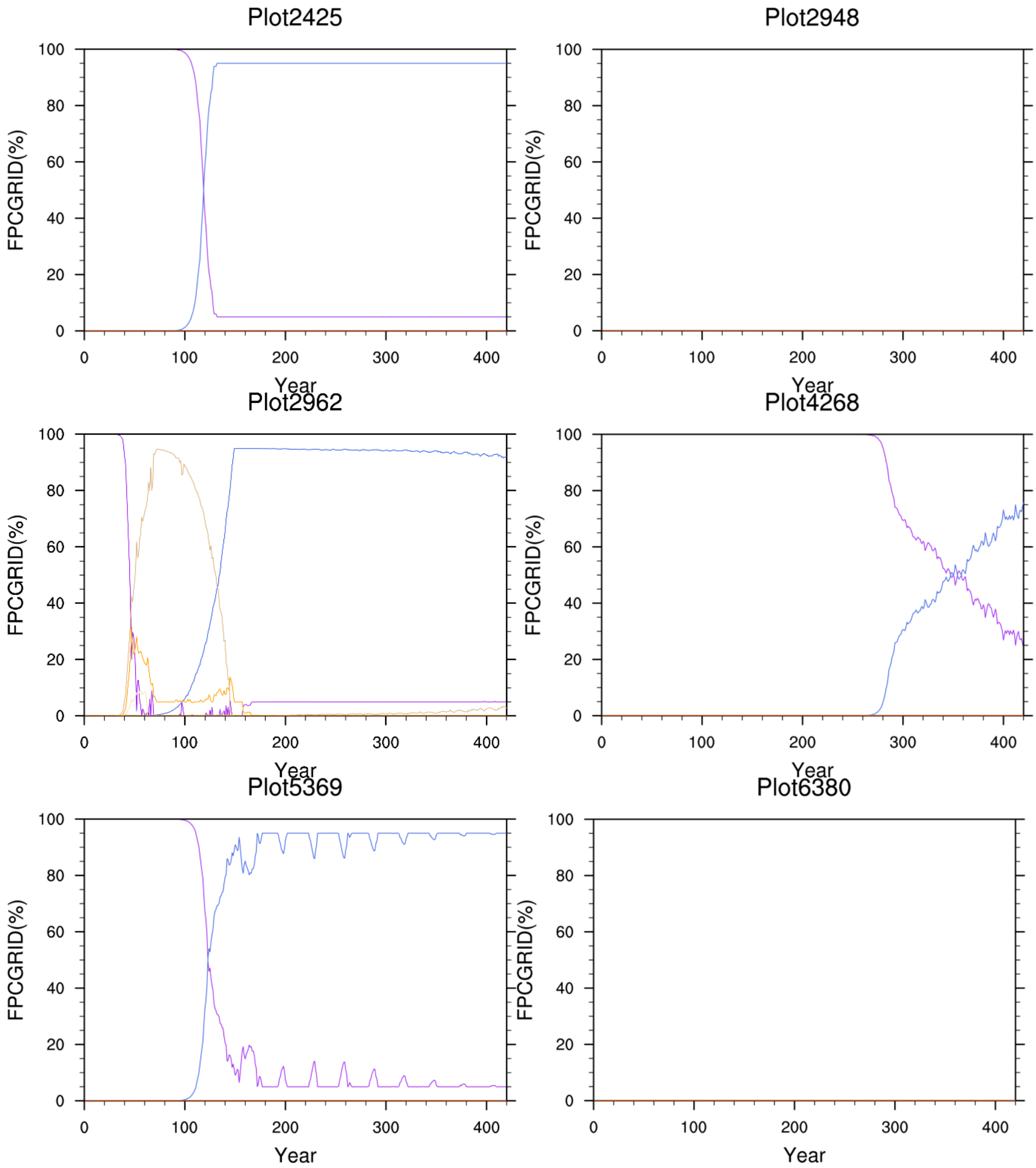


116





117


Plot2425
Plot2948
Plot2962
Plot4268
Plot5369
Plot6380
FPCGRID(%)
Year
0
20
40
60
80
100
0
100
200
300
400

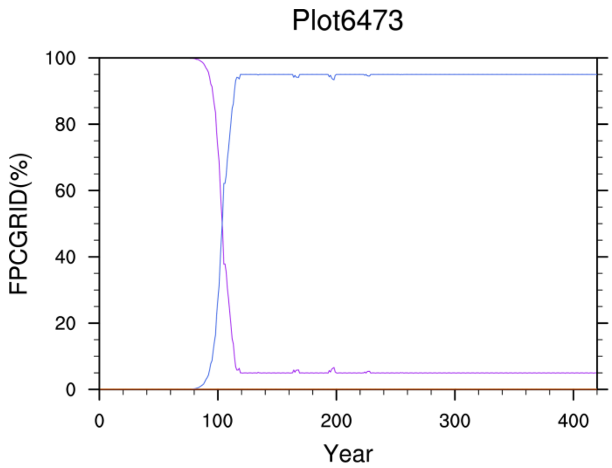


Figure ~~S10~~ S11.1 – DGVM spin-up for 400 years and simulation of PFT profiles for each of the 20 plots used in this study. FPCGRID – estimated percentage per PFT per grid cell. Reference number of plots accords with the AR18x18 dataset, and plot numbers can be found in Table S1.

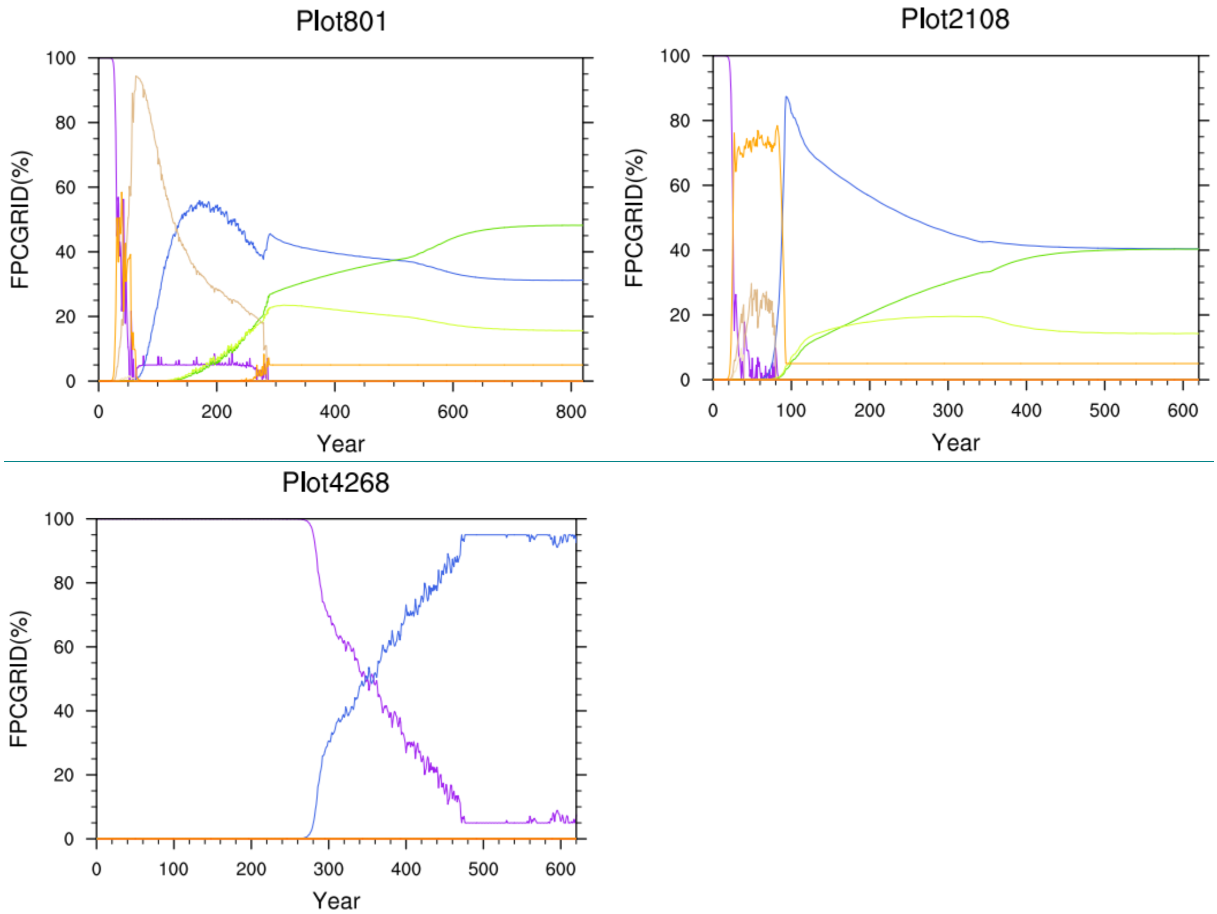


Figure S11.2 – Three plots (number 6, 12, 17) where DGVM spin-up was prolonged beyond 400 years and simulation of PFTs was extended by 400, 200 and 200 years respectively in order to check for equilibrium. FPCGRID – estimated percentage per PFT per grid cell. Reference number of plots accords with the AR18x18 dataset, and plot numbers can be found in Table S1

~~Supplement S11~~Supplement S12 – Sensitivity experiments: frequency-of-presence (FoP) plots

Frequency-of-presence (FOP) plots based upon output from distribution models (DM) for the nine combinations of three environmental variables and three vegetation types modelled, used to indicate threshold values that were explored in the sensitivity experiments, are shown in Fig. S11. Thresholds for new variables in DGVM models were chosen based upon visual inspection of the FoP plots. For example, while boreal BDS are abundant below swe_10 value of 380mm, boreal BDT and boreal NET are abundant at values of swe_10 below 180mm and 150mm respectively. Also, while we identified no clear threshold of variable bioclim_15 for boreal BDS and BDT (frequency of presence is never zero along the variable x-axis - lower left and middle panel of Fig S12), threshold for boreal NET was set to 50 (a value above which no presences occur - lower right panel of Fig S12).

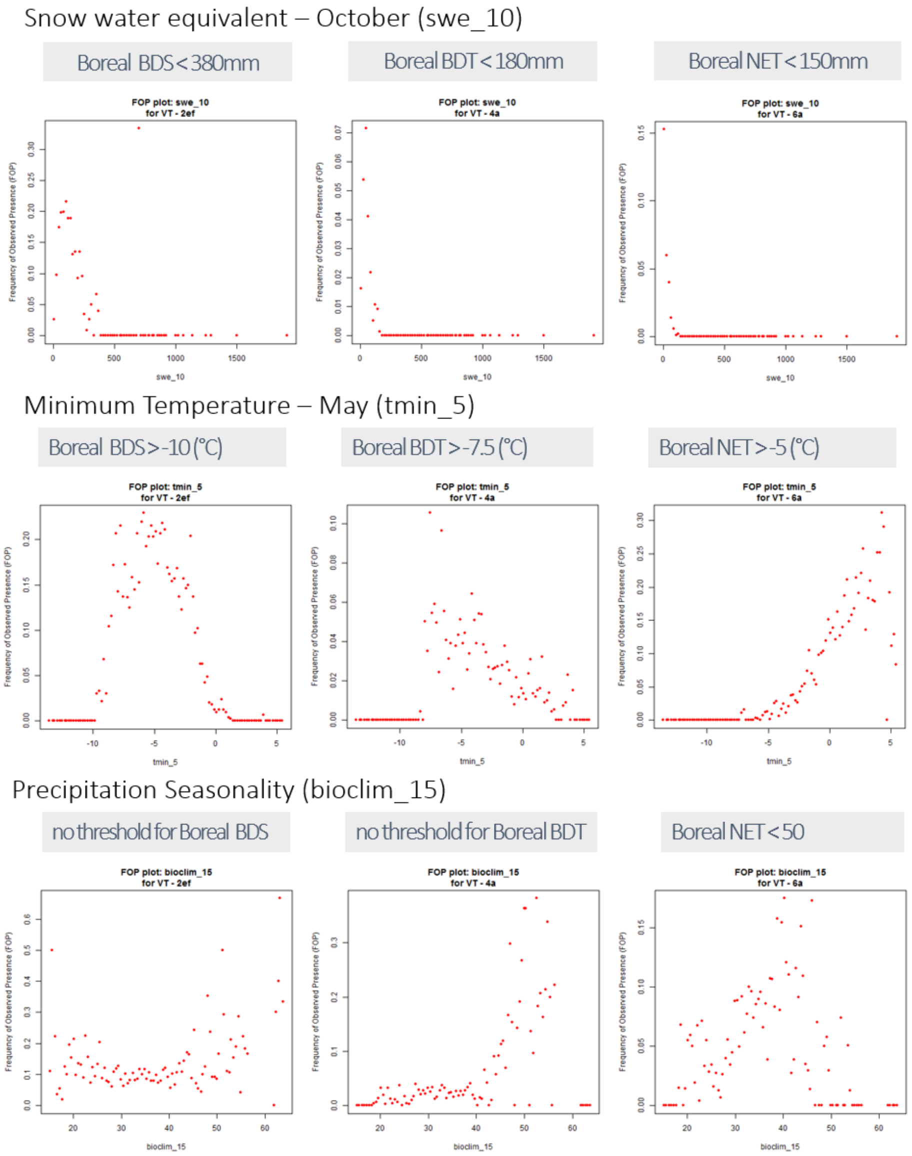


Figure ~~S11~~ S12 – Frequency-of-presence plots from the distribution modelling (DM) study by Horvath et al. (2019) for the combinations of environmental predictors and vegetation types (VTs) used in the sensitivity experiments with DGVM. FOP is the frequency of 100×100 m pixels in the AR18×18 dataset in which the VT in question is present, expressed as a fraction of all pixels in that interval along the environmental variable. All environmental variables were a priori divided into 100 intervals with the same number of pixels. The environmental gradients were: swe_10 – snow water equivalent in October (mm); tmin_5 –- minimum temperature in May (°C); bioclim_15 – precipitation seasonality (unitless index). Boreal BDS – boreal broadleaf deciduous shrubs, Boreal BDT - boreal broadleaf deciduous trees, Boreal NET - boreal needleleaf evergreen shrubs.

~~Supplement S12~~Supplement S13 – Sensitivity experiments: results

Table ~~S12~~ S13 – PFT profiles for the six out of the 20 plots (plot numbers 1, 2, 5, 15, 17, 18) which were included in the sensitivity experiments, for four ‘generations’ of DGVM parameter settings and the AR reference dataset. From left to right the column represent: DGVM before adjustment of parameters thresholds; DGVM_adj1 after ~~adjustment~~ first adding parameter threshold of swe_10; DGVM_adj2 after ~~adjustment~~ also adding parameter threshold of tmin_5; DGVM_adj3 after finally adding parameter threshold ~~adjustment~~ of bioclim_15; and the PFT profile of the reference dataset AR. All parameter thresholds were added cumulatively. Full names for the PFTs are given in Table ~~S6~~ S7 and names of parameters and their values in Table 3.

	DGVM	DGVM adj1	DGVM adj2	DGVM adj3	AR	DGVM	DGVM adj1	DGVM adj2	DGVM adj3	AR	DGVM	DGVM adj1	DGVM adj2	DGVM adj3	AR	DGVM	DGVM adj1	DGVM adj2	DGVM adj3	AR	DGVM	DGVM adj1	DGVM adj2	DGVM adj3	AR	DGVM	DGVM adj1	DGVM adj2	DGVM adj3	AR
	plot 1					plot 2					plot 5					plot 15					plot 17					plot 18				
BG	5	5	5	9	0	6	5	5	5	4	6	6	6	7	0	5	5	5	3	13	28	100	100	100	0	5	100	100	100	0
boreal NET	95	95	95	0	0	58	58	58	0	0	52	52	52	0	0	92	92	92	0	0	72	0	0	0	0	95	0	0	0	0
temp. BDT	0	0	0	0	0	2	2	2	33	0	4	4	4	13	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
boreal BDT	0	0	0	0	35	2	2	2	31	12	4	4	4	13	0	0	0	0	2	0	0	0	0	0	66	0	0	0	0	70
boreal BDS	0	0	0	91	63	32	32	32	31	75	35	35	35	67	99	3	3	3	89	83	0	0	0	0	34	0	0	0	0	30
C3	0	0	0	0	1	0	0	0	0	9	0	0	0	0	1	0	0	0	6	5	0	0	0	0	0	0	0	0	0	0