Response to Referee #3

We thank R3 for this helpful review. Enclosed please find a detailed explanation of the revisions we made based on R3's comments. For your convenience, comments are in bold and our response is in italic. Revisions we made in the manuscript are presented in italic with grey background.

This manuscript describes a revision to the ORCHIDEE land surface model to improve the way in which tundra and subarctic vegetation are simulated by the model. The authors achieve this update by implementing three new plant functional types (PFTs) – these are a boreal shrub type, an arctic graminoid type, and a non-vascular plant type – into the model framework. Implementing new PFTs in ORCHIDEE has two steps, 1) changing process representations where necessary, and 2) defining the set of parameters that characterizes each PFT. The new shrub and grass PFTs needed few changes to process representation to implement, while on the other hand, simulating the non-vascular plant PFT required a different way of dealing with plant water uptake, gross productivity, and mortality. Parameter sets for each of the new PFTs were estimated using a Bayesian estimation process. The authors use the result of the new PFTs, updated process representations, and parameter sets and run the new version of the (ORC16), and compare the result to field-based observations, to satellite remote sensing products, and to the previous version of the model (ORC13) to highlight the effects of the update.

In general, this manuscript is valuable and should be published. It describes a valuable update to ORCHIDEE, which will undoubtedly be used in a number of forthcoming and future studies. The changes to the model lead to improvements in the comparison with observations, and thus represent progress over ORC13. However, the manuscript presentation is not particularly good: the text requires a thorough copyediting to clarify grammar and usage style, some of the figures are too small, and there a few small issues concerning the presentation of units and values which are elaborated below. Aside from these presentation issues, my major concern of this study was the choice of data used to inform the parameter optimization, and the appropriateness of comparing site-level measurements with model simulations performed on a 2-degree grid.

The entire manuscript was re-read by a native English speaker. Possible further improvements may be done upon request, for a next stage of revision or upon acceptance. However the grammar and usage style changes are not reported in this response. Furthermore, the size of all figures has been increased. Concerning your other comments, please find some answers below.

The largest concern I have with the current study is the authors' apparent inability to assemble a larger, more representative dataset of high-latitude plant characteristics with which to parameterize the model. Their Bayesian optimization relies exclusively on the Peregon et al., 2008 biomass and NPP dataset. These data were specifically collected on *wetland* vegetation, while ORCHIDEE, in this paper, is intended to simulate *upland* vegetation. This mismatch between what the data represent and what the model is trying to simulate is a very serious limitation and calls into question the appropriate- ness and quality of the model parameterization. Use of such a limited and specialized dataset to parameterize a global model might be acceptable in regions of the world for which there are very few ecological and ecophysiological data, e.g., in parts of the tropics, but for the Arctic, it is practically inexcusable because of enormous amount of field research that has been performed over the last 50 years. Data from iconic arctic research sites such as Toolik Lake in North America, Abisko in Europe, and Zacken- berg in Greenland were ignored in development of the testing dataset. Large amounts of data on key characteristics such as aboveground biomass were collected in the en- tire circumarctic region as part of, e.g., the ITEX experiment. Data from all of these locations outside of west Siberia, while perhaps more difficult to assemble, could have provided valuable information on the status of upland tundra and subarctic vegetation that would have been more appropriate for performing the model parameterization. If the authors prefer to not improve their parameterization using more widespread and representative field data, at very least they should explain and justify their choice for using the wetland dataset of limited spatial extent more clearly in the manuscript.

We are aware of your concern about the spatial representativeness of the dataset used for the Bayesian optimisation. However part of our choice is justified by specific needs for the calibration and by the accessibility to the data. We needed total living biomass and productivity at different sites, with multi-annual observations, and for the three new PFTs. The published and unpublished data provided by Peregon et al. satisfied these criteria, while we did not find easily other data sets satisfying all criteria. We agree that there is

a large amount of recent campaigns in the Artic with numerous in situ measurements especially at specific highly instrumented sites; however these data are not assembled into a freely available and comprehensive database. Note also that the western Siberian data are acquired mainly on lowlands but not exclusively on very humid sites. As you suggested, in this case it is important to clarify our approach and we have thus added in the text: p.14 l.36-38: "Note finally that using a single dataset in Western Siberia (mainly lowlands) for the model calibration may introduce some biases, which will have to be evaluated."

However, to account for your very relevant suggestion, we have searched for additional data for the model evaluation, especially from the sites you recommended. We did not find any complete data set, in the mass of published literature, which could be used easily for the optimization step along with the Siberian data. Nevertheless, we now use two North-South Arctic transects (with biomass data in lowlands and uplands): one in Eurasia (Walker et al, 2011a) and one in North America (Walker et al, 2011b; and previous reports since 2007). While these data are not sufficient for the optimisation (the productivity is missing), we propose to use them to evaluate the model. We do not claim that a larger set of data could not have been gathered but given the focus of the paper, i.e. on the new process description, we believe the two sets of data that are now used (from Western Russian and from two transects) are sufficient. We added a new figure (Fig. 9) for the model evaluation with associated comments reported below. Note finally that we discuss in the text the potential shortcomings due to the use of mainly lowland data for the calibration of a global model.

P.17 l.14-19: "We further compare the simulated biomasses with two other Arctic transects. The first one is the North America Arctic Transect (NAAT). It is situated in a continental area, and includes eight field locations (70°N 149°W to 79°N 100°W) sampled from 2002 to 2006 (Walker et al., 2011b) chosen as representative of zonal conditions. The second, located in a marine-influenced area, is the Eurasian Arctic Transect (EAT). It includes six field locations (58 to 73°N, between 67 to 81°E) sampled from 2007 to 2010 (Walker et al., 2008, 2009a, 2009b, 2011a)."

P.19 l.1-15: "Carbon stock with two Arctic transect

To evaluate the modelled biomass in other Arctic sites (not used in the calibration step), including uplands and lowlands, Fig. 9 shows scatter plots of observed and simulated biomass along two transects: the NAAT (North America) and the EAT (Eurasia) Artic Transect. The NVPs and shrub

biomasses are relatively well reproduced by the model (i.e. within the error bars). For both PFTs, the standard deviation of the observations includes the 1:1 line, but the observed biomasses are on average higher than the simulated biomasses. Simulated shrub biomasses are biased low for the NAAT transect but not for the EAT transect.

In contrast, the mean value of observed biomass for boreal C3 grasses (Fig. 9.c) is low compared to the simulated biomasses for both cases. For half of the sites the simulated low biomass is in accordance with the observations, but for the other half the values are much larger (> 300 gC.m^2 whereas the observation do not exceed 54 gC.m^2). Despite the optimization with observations from western Siberia (Fig. 7; leading to a decrease of biomass compared to temperate C3 grasses) there is likely an overestimation of the biomass for boreal C3 grasses, probably associated with an overestimated productivity."

Walker et al, 2011a: Vegetation of zonal patterned-ground ecosystemsalong the North America Arctic bioclimate gradient. Applied Vegetation Science 14, 440–463. Doi: 10.1111/j.1654-109X.2011.01149.x

Walker et al, 2011. 2010 Expedition to Krenkel Station, Hayes Island, Franz Josef Land, Russia, Data Report, Alaska Geobotany Center, Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK. 63 pp.

Specific comments

Page 2, line 3

The last glacial inception began around 126.5-120 ka; correct this error

Done

Page 2 line 28

The model described is called BIOME4; please correct the model name

Done

Page 8 line 16

Anoxic conditions affect the activity of all types of soil microorganisms, not only bacteria, e.g., fungi, archaea, and multi-celled microorganisms. Please be more inclusive instead of using the word "bacteria"

We thank the reviewer for this relevant comment. We changed "bacteria" for "soil microorganism" (p9. l.1)

Page 13 line 14-18

Why not make the root profile shape parameter a function of the mean active-layer thickness? The model simulates active layer thickness, and presumably most plants would optimize their rooting profile to be compatible with this value

We agree that this is an important suggestion. Using a dynamical root profile could be appropriate to take into account the active layer thickness or the water table or the plant growth status. However, to keep model consistency, it should be applied to all PFTs of the model, and not only the PFTs developed in this study. Given the requested work especially for the calibration issues, we chose not to change the general equation of the current version of ORCHIDEE. Note that this is currently under investigation for all PFTs.

Page 13 lines 21-23

This sentence is confusing. Please revise for clarity by explaining how this version of ORCHIDEE uses prescribed vegetation cover and therefore survival and establishment limits are not relevant.

We changed the sentence to: "Note that we did not add any bioclimatic limits, such as i) survival or establishment temperature thresholds as proposed by Bonan et al., (2003) and Oleson et al. (2013) or ii) a cumulated degree-day threshold (above the zero degree criteria) for the plant growth (Miller and Smith, 2012). In this study we use ORCHIDEE without the dynamic vegetation module, but with a prescribed vegetation cover preventing vegetation development in unfavourable areas" (p.14 l.5-9).

Page 13 line 31

Explain why using observational data collected in "boreal wetlands" is appropriate for a parameterizing a global model that simulates predominantly upland systems, indeed, there is no representation of wetlands at all in this version of ORCHIDEE (as far as I could understand).

We have already partially answered this comment above. The first reason is that it was the most appropriate dataset that was available to us, even though it concerns mainly lowlands. Secondly, although it is considered as lowland on average, such data set comprises some sites that are not so-called wetlands. Finally, although we have kept this data set for the model calibration, in order to evaluate results at a global scale, we now use an additional set of observations for the model evaluation. These new data include both upland and wetland observations (Fig. 9 and associated comment, p.17 l.14-19, p.19 and l.1-15).

Page 14 line 14-16

If the model was run on a 2-degree grid, why were the site-level data aggregated only to half-degree? Wouldn't it have made more sense to aggregate the data at the same spatial scale as the model simulations? Also, the choice of dataset (from wetlands) clearly limits the amount of data coming from non-vascular plants, shrubs, and grasses; wouldn't an effort to assemble a more spatially global and upland-representative dataset have helped here?

For the first part of your comment, indeed this point was not clear enough. In fact, the optimization is also done at 0.5° resolution. We have now added a new sentence to clarify this in the Section about the optimization (2.6.1): "The simulation for the optimisation was done with CRU-NCEP meteorological forcing (Wei et al., 2014; Viovy, 2015), at 0.5° resolution" (p. 16, l.36-37 and p.17 l.1).

The second part of your comment was already answered above.

Page 15 line 4-5

The phrase starting ". . .in CAVM Mapping Team. . ." is awkward and hard to understand. Rephrase.

We rephrased as follows: "In the map from Loveland et al. (2000), we noticed that the tundra biome corresponds to the "sparse vegetation" or to the "lichens and mosses" LCCs distribution. In CAVM Mapping Team (2003), the tundra biome is described as containing ~30 to 60% NVPs." (p.15, l.39-31)

Page 16 line 20-21

As we know multi-annual and decadal climate cycles exist, e.g., ENSO, and that there was a clear trend on climate during the 1st half of the 20th Century, is it appropriate to select individual years randomly over this period for the model spinup? I realize that many other vegetation modeling protocols prescribe the same thing, but that doesn't mean that it is correct. Using a detrended climate timeseries would be a minimum first step towards improving the quality of the model spinup.

Thank you for this remark. Indeed this is probably a better solution. However, for this study, this would lead to re-running all simulations, which was not possible at this stage. Moreover, the impact on above-ground boreal vegetation after a century of stable climate would probably be minor.

Page 17 line 24

If the 2-degree resolution used to run the model presents problems in terms of comparison with observations, why wasn't the model run at finer resolution, or in an "individual point" model with local forcing. This version of ORCHIDEE does not simulate any 2D spatial processes that would be impossible to implement in a point mode.

The aim of this study was to improve boreal representation on a global scale. However, at such scale, fine-resolution (e.g. 0.5°, used for the optimization) simulations would be too computationally demanding. Moreover, local (point) meteorological forcing data, including precipitation, temperature, downward longwave and shortwave radiation, relative humidity and wind, were not available. This is why we chose to run the model at 2° resolution with a global climate forcing based on a merge of climate reanalysis and in situ observations. Else we agree that if the local forcing data would have been available, we should have used them.

Page 17 line 32-34

Making an effort to assemble a larger calibration-evaluation dataset would have helped here. If these data really do not exist, this has to be clearly explained in the manuscript.

This comment was already answered above. Data we found for other sites are not complete enough to be used for the calibration. However, they were used to improve the evaluation of our results (Fig. 9 and associated comment, p.17 l.14-19, p.19 and l.1-15).

Page 18 line 1-2

Again, having more, and more widespread observations might have helped here.

Same as above

Page 18 line 24-26

I would be very helpful for the reader if the meteorological variables were provided in terms of more ecologically relevant units. For example, provide precipitation in terms of annual totals, and temperature in terms of summertime (JJA) or growing season means (instead of annual? – it's not clear what is provided here).

We have changed most units to more ecologically relevant ones (in "$mm.y^{-1}.m^{-2}$" in p.21 l.21,29-31,34, p.50 l.12-13, Figs. 12 and S5). Moreover, we have now indicated more clearly on which period temperature is considered ("growing

season (AMJ) mean air temperature" p.50 l.13) and we have updated the values (p.50 l.16)

Page 19 line 16-17

Again, what are these temperature anomalies referring to – seasonal, annual, individual months? A +10 anomaly in winter temperature in an place where the mean winter temperature is -40 C may not really be ecological relevant.

It was annual temperature anomalies, but which are present all along the year (winter, growing season or summer). In order to be clearer and consistent with precedent changes, we changed by the growing season: p.51 l.11-12: "growing season temperatures" and "+6°C and + 10°C compared to America and Asia respectively".

Page 20 line 3-4

The phrase with ". . .too low LAI. . ." is awkward. Revise.

We rephrased as follows: "However, the model underestimates LAI in the central-west of Siberia" (p.19 l.21).

Page 21 line 1-2

Again, provide ecologically relevant units, e.g., total transpiration per month.

Page 21 line 11-14

Again, adjust units of evapotranspiration, runoff, etc. to monthly, seasonal, or annual sums. Annual is probably best here.

We change all values by day (in $mm.d^{-1}.m^{-2}$) by values in "$mm.y^{-1}.m^{-2}$" (p.21 l.21,29-31,34, p.50 l.12-13), including the Fig. 12 and S5.

Page 22 line 35

In the boreal regions and Arctic, the shrub vegetation is composed of both evergreen and deciduous (summergreen) broadleaved plants (angiosperms), and evergreen needleleaf plants (gymnosperms). Thus, there are at least three types of shrubs.

Indeed, we forgot one type here. This sentence was changed to: "Concerning shrubs, we selected a boreal broad-leaved deciduous phenology, although in reality there is a mix of deciduous and evergreen broadleaf shrubs and evergreen needled-leaf shrubs." (p. 23, l. 26-27)

Figure 5

The maps should be reproduced in a larger size

Figure 6

The plots should be reproduced in larger size, or at least the points should be plotted a bit larger. It is hard to see some of the points, especially the cyan colored dots

Figure 11

The maps should be reproduced in a larger size

Figure 12

The maps should be reproduced in a larger size

All of the designed figures have been enlarged. In particular, the figure 6 has been improved to be more readable.