We would like to thank both Referees for their valuable comments and suggestions for improving our manuscript. Following Referees' comments, we carefully revised our manuscript. Please find below the point-to-point responses (in black) to referee comments (in blue). For your convenience, changes in the revised manuscript are highlighted with dark red.

Reviewer #2

Comment

Authors attempt to improve and test the dynamic vegetation module of the ORCHIDEE model to primarily show that inclusion of new bioclimatic constraints that induce mortality lead to better simulation of fractional coverage of PFTs in mid- to high-latitude regions.

The manuscript is reasonably written but as a reader I have some concerns, which if addressed will strengthen the manuscript significantly. In addition, I am attaching the scanned version of the annotated manuscript, as a supplement, on which I have made several comments. These are primarily minor comments.

Major comments

1. My first major concern is that there is no equation in the manuscript that will allow a reader to see how competition between PFTs is modelled. Scanning through the Krinner et al. (2005) GBC paper, I am unable to find an equation like the following ...
df/dt = establishment + encroachment into inferior PFTs - mortality - take over by superior PFTs
where f is the fractional coverage of a PFT and I assume is the primary variable of interest.

Response

The following equations were added after P2219,L25 accordingly: "…which simulates the dynamic area covered by each PFT as functions of bioclimatic limitation, competition, mortality and establishment. The basic equations to calculate fractional cover of each PFT are listed below:

$$V = CA \times P$$

$$\frac{dP}{dt} = E - M \times P$$

where V is fractional vegetation cover (dimensionless); CA is crown area of individual plant (m^2); P is population density (m^{-2}); E is establishment rate ($m^{-2}\ d^{-1}$); M is mortality rate (100% d^{-1}), including components described in Sect. 2.2.1."

---

2. Second, the paper fails to acknowledge that by including more and more bioclimatic constraints we are essentially turning DGVMs into biogeography models. We all realize that the current generation DGVMs use phenomenological approaches. If the physiological processes in the model were sufficiently process-based we would never need bioclimatic constraints to include mortality. Yet, as modellers, we keep digging empirical evidence to find

more and more bioclimatic constraints. Consider the three additional constraints used in this manuscript - tree mortality during extremely cold days, broadleaf tree mortality caused by spring frost and growing-season temperature limits to tree extension - all of which are temperature related in one form or another.

In absence of a df/dt equation, and an overall large stress on mortality due to bioclimatic constraints, I am inclined to ask to what extent has ORCHIDEE become a biogeography model, in which the spatial distribution of PFTs is determined primarily by their bioclimatic constraints and not by the explicit competition between them.

Response

This is a very good remark, touching some general and fundamental discussions on the current issues in DGVMs. We agree that an ideal DGVM should contain sufficient physiological processes that enable the model to realistically simulate vegetation distribution, with the least empirical bioclimatic constraints. However, for now, many well-established DGVMs like LPJ, Sheffield-DGVM and ORCHIDEE still contain empirical extreme-temperature constraints that work on vegetation dynamics (Sitch et al., 2008).

On one hand, forest mortality is a complex process, involving interactions between management, disturbances and direct climatic effects on tree physiology. The lack of fundamental understanding of mortality prevents mechanistic parameterization of mortality in DGVMs (Steinkamp et al., 2015; Wang et al., 2012). Thus, modelers have to choose among various logical yet unconfirmed algorithms to calculate mortality, including growth efficiency related mortality (as Eq. 1) and climate constraints (McDowell et al., 2011; Steinkamp et al., 2015).

On the other hand, temperature-related constraints indeed appear to be the most reasonable explanation of treeline locations at high latitudes and high elevations (Richardson et al., 2009; Körner et al., 2004). The physiological pathways of cold temperature remaining unresolved, we think it acceptable to adopt temperature constraints derived from large scale measurements (Körner et al., 2004).

References:

Körner, C. and Paulsen, J.: A world-wide study of high altitude treeline temperatures, J. Biogeogr., 31, 713–732, 2004.

McDowell, N. G., Beerling, D. J., Breshears, D. D., Fisher, R. a, Raffa, K. F. and Stitt, M.: The interdependence of mechanisms underlying climate-driven vegetation mortality., Trends Ecol. Evol., 26, 523–32, 2011.

Richardson, A. D. and Friedland, A. J.: A review of the theories to explain arctic and alpine treelines around the world, J. Sustain. For., 28, 218–242, 2009.

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), Glob. Chang. Biol., 14, 2015–2039, 2008.

Steinkamp, J. and Hickler, T.: Is drought-induced forest dieback globally increasing?, J. Ecol., 103, 31–43, 2015.

Wang, W., Peng, C., Kneeshaw, D. D., Larocque, G. R. and Luo, Z.: Drought-induced tree mortality: ecological consequences, causes, and modeling, Environ. Rev., 20, 109–121, 2012.

---

3. As a reader, I found several of the new metrics difficult to appreciate. The beta metric used in equation (7) and (8) is essentially the square root of sum of square of difference between model and observations over all PFTs. I am unable to understand why is this limited between 0 and square root of 2. If there is only one PFT in a grid cell covering 100% of the grid cell and model simulates its fractional coverage to be zero, maximum value of beta is obtained equal to 1. If there are two PFTs covering the grid cell say 50% each, and say the model again simulates zero fractional coverage then beta = sqrt((0.5-0)ˆ2 + (0.5-0)ˆ2) = 0.70.

Why not use the already established root mean square error (RMSE). Beta in essence is very similar to RMSE. Why unnecessarily confuse your reader?

The S_V metric used in equation (9) is okay, but would make more sense if it were based on RMSE rather than the beta metric.

Finally, another metric D (absolute difference) is introduced when comparing PFT groups and although an argument is made at the bottom of page 2231 why beta is not used, I am unable to follow this argument.

Note that, with all these new metrics, the manuscript still does not compare the good old mean fractional coverages of PFTs with observations. What is instead shown is the composite color map, which if I am not wrong shows relative abundances and not the absolute values. I realize that a composite map can show more PFTs but relative abundances is a derived quantity and that's not what the model simulates. In my humble opinion, composite maps should be complementary to the usual maps of absolute fractional coverages, not something that replaces them.

Response

Beta diversity (β) was firstly proposed as a metric to estimate the variation in species composition among different sites (Legendre et al., 2005; Legendre et al., 2013). Poulter et al. (2011) use the β metric to assess the reclassification similarity of different PFT maps derived from remotely-sensed land-cover datasets. The β metric was calculated as the root of the sum of square error over all PFTs (Eq. 7 and 8). It is larger than or equal to zero, and can be $\sqrt{2}$ at maximum, in the limit case a grid cell has 100% of one single PFT in one dataset and has 100% of another PFT in the other dataset.

The β metric is similar to root mean square error (RMSE) which is widely used in many fields. But if we use RMSE with the following equation, the value will be dependent on the number of PFTs in the model or dataset. Unlike the β metric which has a fixed range ($[0, \sqrt{2}]$), RMSE will have smaller maximum value as the total number of PFTs increase, making it incomparable between different models.

$$RMSE_{c,M\_Oi} = \sqrt[2]{\frac{\sum_{k=1}^{n}(V_{k,c,M} - V_{k,c,Oi})^2}{n}}$$

where $V_{k,c,M}$ is fractional abundance for PFT *k* and for grid cell *c*, simulated by model; $V_{k,c,O}$ is fractional abundance for PFT *k* and for grid cell *c*, from observational dataset *i*; and *n* is the number of PFTs.

Another way to calculate RMSE is to use the following equation:

$$RMSE_{k,c,M\_O} = \sqrt[2]{\frac{\sum_{i=1}^{S}(V_{k,c,M} - V_{k,c,Oi})^2}{S}}$$

where *S* is the number of datasets.

This method also has a shortcoming: it gives one value for each PFT, and taking the mean RMSE over all PFTs is not appropriate because the redundant PFTs in a grid cell may lead to too optimistic results, blurring the information about the major PFTs in this grid cell.

Considering the shortcomings of RMSE and use of the β metric in assessment of dissimilarity in PFT maps (Poulter et al., 2011; Ottlé et al., 2013), we think it appropriate to adopt β rather than RMSE to evaluate the model results in vegetation distribution.

As for PFT groups, β could be calculated for each group using Eq. 7 and 8, saying that there are only two PFTs in the equation. But we used dissimilarity index (*D*) instead of β because, take needleleaf deciduous trees (PFT9) as an example: they are mainly distributed in eastern Siberia; outside this region, models and observational datasets have ~0 of needleleaf deciduous and ~1 of non-needleleaf deciduous; thus, the Northern Hemisphere average of $\beta_{\text{needleleaf-deciduous}}$ will be very small due to "high agreement" outside Siberia. Unlike *D*, in β calculation, we cannot simply exclude the grid cells where the corresponding group does not exist, since β, by definition, takes into account the case when both maps give "absence" of the corresponding group in the grid cell. Therefore, we chose *D* for PFT groups rather than β. The last sentence on P2231 was revised as: "…because in that case the average $\beta_{\text{group,M\_O}}$ (or $\beta_{\text{group,O\_O}}$) for Northern Hemisphere (20-90°N) would be too optimistic, considering that many of the pixels will be equal to zero, due to the limited distribution range of the corresponding *group*."

Compared to the usual maps of fractional coverage for each PFT, we believe that a composite color map is more concise and captures the main information. Nevertheless, following the comment, we added a figure in the Supplement (Fig. S3), showing fractional coverage for each PFT simulated by both OLD and NEW to allow a comparison of the distribution of all PFTs.

Reference:

Legendre, P., Borcard, D. and Peres-Neto, P. R.: Analyzing beta diversity: partitioning the spatial variation of community composition data, Ecol. Monogr., 75, 435–450, 2005.

Legendre, P. and De Cáceres, M.: Beta diversity as the variance of community data: dissimilarity coefficients and partitioning, Ecol. Lett., 16, 951–63, 2013.

Ottlé, C., Lescure, J., Maignan, F., Poulter, B., Wang, T. and Delbart, N.: Use of various remote sensing land cover products for plant functional type mapping over siberia, Earth Syst. Sci. Data, 5, 331–348, 2013.

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S. and Zimmermann, N. E.: Plant functional type mapping for earth system models, Geosci. Model Dev., 4, 993–1010, 2011.

Minor comments

P2215,L7: Reword. Because it hasn't been updated doesn't necessarily imply it yields unrealistic results.

Response

This sentence was revised as: "The vegetation dynamics module (ORC-VD) within the process-based ecosystem model ORCHIDEE (Organizing Carbon and Hydrology in Dynamic Ecosystems) has not been updated and evaluated since many years and is known to produce unrealistic results."

P2215,L24: "…and the return frequency…" unclear, please reword.

Response

This sentence was revised as: "…and the effects of frequency and severity of extreme cold events during the spin-up phase of the model."

P2216,L6-8: "To simulate…" reword.
P2216,L10-17: can be better written, somewhat weak at present.

Response

The sentences at P2216,L6-17 were revised as: "To simulate past and future changes on long time scales, Earth system models must represent how the distribution and structure of ecosystems respond to changes in climate, CO_2 and land use. This need provides the motivation for the development of dynamic global vegetation models (DGVM). In DGVMs, vegetation distribution, carbon stocks and fluxes exchanged with the atmosphere are simulated through fast processes (canopy exchange, soil heat and moisture dynamics, photosynthesis), intermediate processes (vegetation phenology, carbon allocation and growth, soil carbon decomposition) and slow processes (vegetation dynamics, recovery from disturbances) (Sitch et al., 2003; Krinner et al., 2005). DGVMs have been used to study the response of ecosystems to recent climate change (e.g., Piao et al., 2006) and to project the evolution of the coupled carbon-climate system (e.g., Cox et al., 2000). The coupling of vegetation dynamics with a climate model allows for the inclusion of vegetation-atmosphere interactions related to ecosystem migration in global climate simulations (Quillet et al., 2010)."

P2216,L21-22: Are you sure about CO2 and soil? My understanding is that biogeography models just use climate info.

Response

The early biogeography models simulated the natural potential distribution of ecosystems as a function of climate and soil properties (Prentice et al., 1992; Neilson et al., 1992). Then the new generation of process-based, equilibrium biogeographic models (Neilson, 1995; Haxeltine and Prentice, 1996) incorporated physiological CO_2 effect, allowing direct CO_2 effects on both productivity and water use efficiency. So we think it appropriate to write the sentence like this.

Reference:

Haxeltine, A. and Prentice, I. C.: Biome3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types, Global Biogeochem. Cy., 10, 693–709, 1996.

Neilson, R., King, G. and Koerper, G.: Toward a rule-based biome model, Landsc. Ecol., 7, 27–43, 1992.

Neilson, R. P.: A model for predicting continental-scale vegetation distribution and water balance, Ecol. Appl., 5, 362–385, 1995.

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., and Solomon, A. M.: A global biome model based on plant physiology and dominance, soil properties and climate, J. Biogeogr., 19, 117–134, 1992.

P2217,L9: "at the end of each time step", at this point the reader doesn't know the time step of your model.

Response

We deleted "at the end of each time step" in this sentence accordingly.

P2217,L11-19: This discussion seems unnecessary and confusing without introducing the reader to "variable-trait" approach.

Response

We included this discussion about newly developed "variable trait" approach to provide readers with some recent developments in DGVMs. To avoid confusion, the sentences at P2217,L10-15 were revised as: "The competence of any PFT is dependent on the underlying plant traits that define this PFT. The traits for a given PFT are fixed in most DGVMs, but can also be variable within PFTs based on trait-climate relationships derived from trait database. For example, Verheijen et al. (2013) conducted a variable trait simulation with the JSBACH DGVM for three leaf traits (Specific Leaf Area, and the constants defining the maximum rate of photosynthesis, v_{cmax}, j_{max}), showing significant difference in predicted dominant PFTs compared with fixed trait simulation. Higgins et al. (2014) however, pointed out…"

P2217,L26: "…evaluated for static runs…" -> "for runs in which geographical distribution of PFTs is specified"

Response

This sentence was revised accordingly as: "These new parameterizations have been evaluated for static runs in which geographical distribution of PFTs is specified based on observed satellite land-cover information."

P2218,L2-4: without any reference to "updates" you can simply say ORC-VD produces unrealistic results.

Response

We mentioned "updates" because the dynamic vegetation module had reasonable results in its first version in Krinner et al. (2005), but did not work well after the later developments in ORCHIDEE physical and biogeochemical processes. This sentence was revised as follows to avoid the improper logic that "not updated" necessarily leads to "unrealistic results": "ORC-VD has not been updated and evaluated since the Krinner et al. (2005) description, and it produces unrealistic results in dynamic runs."

P2218,L17: "Vcmax/Jmax" just say photosynthesis parameters. At this point reader doesn't know what these mean.

Response

Revised accordingly.

P2218,L18-20: already mentioned in previous para.

Response

We deleted the repetitious part in this sentence as: "The results of the original module (ORC-HL-OVD) and of the new parameterization (ORC-HL-NVD) are evaluated (Sects. 4 and 5)."

P2219,L21: "…in which different processes…" -> "…in which processes different…"

Response

Revised accordingly.

P2220,L2: How is population density related to fractional cover? What is the state variable in the model, is it fractional coverage or something else.

Response

Fractional cover equals to the product of population density (unit: m^{-2}) multiplied by crown area of individual plant (unit: m^2) (Krinner et al., 2005, Eq.1).

The main state variable in ORCHIDEE is fractional coverage, and all the carbon variables are defined on fractional coverage.

P2220,L2 and L12: "Mortality is defined as the percentage reduction…" "…(d^{-1})" these are not same units.

Response

The sentence at P2220,L2 was revised as: "Mortality is defined as the reduction in population density during each time step (daily)."

P2220,L19: mortality by itself doesn't determine competition.

Response

This sentence was revised as: "The dynamic mortality formulation M_{BG} takes into account the influence of growth efficiency on tree mortality, and thus can simulate the competitiveness of tree PFTs under various climates…".

P2221: Isn't the new mortality also instantaneous, just that the rate increases as T_{min} becomes greater than $T_{min,crit}$

Response

The major difference here between ORC-HL-OVD and NVD is that, in the old version, the tree PFTs will be completely eliminated once the minimum temperature in a day drops below the PFT-dependent threshold, while in new version, we defined an extreme coldness-induced mortality as a function of daily minimum temperature. To clarify it, the first sentence on P2221 was revised as: "…the corresponding tree PFT was completely eliminated."

P2221,L19: Is this competition? No, this is biogeographic limitation

Response

We agree that this is an empirical biogeographic limitation, but mortality can also be regarded as part of the PFTs' competiveness. The boreal needleleaf deciduous trees have higher tolerance to extreme cold climate than other trees, thus in the model they have smaller mortality in face of coldness, and win against other tree PFTs in eastern Siberia through indirect competition.

P2223,L3: "warm season air temperature (T_{WS})" Define this. Is this the average of the temperature in a year?

Response

In fact, Eq.4 is the definition of T_{WS}. It is not the annual average, but a (similarly) running mean of daily mean temperature (see Krinner et al., 2005, Eq.3).

P2223,L3: "…to exclude trees…" which tree PFTs

Response

This sentences was revised as: "…to exclude all tree PFTs…".

P2223,L27: So do you replace T_{WS} by T_{GS} in eq(4)

Response

Eq.4 is the calculation method for T_{WS}. The T_{WS} criterion existed in Krinner et al. (2005) but not in ORC-HL-OVD. We re-introduced a criterion (T_{GS}) to constrain tree expansion to Arctic regions, based on more recent literature results (Körner et al., 2004; Randin et al., 2013).

P2224: "Code availability" seems more suitable info for an Appendix.

Response

This Sect. 2.3 Code availability was moved to the end of the manuscript.

P2225,L20: "…from bare ground" -> "where fractional coverage of all PFTs are zero"
P2225,L22: "…cycling CRU-NCEP…" -> "repeated using"

Response

Revised accordingly.

P2226,L14: STAT1 & STAT2 are not listed in Table 2.

Response

The original "STAT" in Table 2 was separated into "STAT1" and "STAT2" accordingly.

P2226,L24-29: Not clear what is the purpose of this comparison.

Response

Since fire is an important vegetation succession process in boreal regions, we did a test similar to NEW but deactivated the fire module. In current ORC-HL, the fire module is still the relatively simple one as described in Krinner et al. (2005), rather than the recently developed SPITFIRE that has been implemented in ORCHIDEE standard version. In order to justify the use of the old fire module, we conducted this comparison of burned area simulated by ORC-HL (old fire) and ORCHIDEE standard (SPITFIRE).

P2228,L3: "In order to account for uncertainties of observations…" -> "In order to account for uncertainties in observation-based estimates"

Response

Revised accordingly.

P2230: "beta diversity" Isn't this similar to RMSE? Why the fancy name?

"β is bound to the interval $[0, \sqrt{2}]$" Not obvious why? Is β forced to be in this interval?

Response

Please refer to the previous response to "Major comments 3"

P2230,L16: "In order to derive a bounded score" What does bounded implies here.

Response

The metric for model skill at simulating vegetation distribution (S_V) is defined as the mean β of data vs. data divided by the mean β of model vs. data (Eq.9). If S_V for a grid cell is larger than 1 for both models, indicating that the uncertainties in the observation-based estimates are too large to be qualified for model evaluation, this grid cell is excluded in the calculation of regional average S_V. Thus, S_V ranges from 0 to 1 (i.e., bounded range). If we inverse the numerator and denominator, S_V will range from 0 to infinity, with lower values representing better performance, which is counter-intuitive. To clarify it, the following sentence was added after P2230,L21: "If $S_{V,c}>1$ for both models, indicating that the observation-based estimates have too large uncertainties to be qualified for model evaluation, then this grid cell *c* is left out."

P2232,L14: "discrepancies" -> "uncertainty"

Response

Revised accordingly.

P2233,L1-4: Does this still allow to compare means over a given period?

Response

Yes, the skill score for GPP (S_G) is intended for evaluation against mean values over years. In this study, we used 10-year average (1999-2008) of data-driven MTE GPP for evaluation. On the contrary, for the evaluation of time series of GPP, other metrics like IOA (index of agreement, Willmott et al., 2012) may be more suitable.

Reference:

Willmott, C. J., Robeson, S. M. and Matsuura, K.: A refined index of model performance, Int. J. Climatol., 32, 2088–2094, 2012.

P2233,L26: "…observed land-cover uncertainty…" -> "uncertainty in observation-based estimates of land cover"

Response

Revised accordingly.

P2237,L13: "or 25%" 25% of what? 25% seems large.

Response

In this sentence, “or 25%” was revised as “or 25% of their mean”. 25% itself may seem large, but after multiplying by delta fraction of corresponding PFTs, the relative difference of total GPP in the grids will usually be less than 10%.

P2238,L13-14: This is not exactly true because so many processes in the model affect turnover. I believe you can’t just take the turnover number in years from the model and multiply it with NPP.

Response

We agree that turnover time in the model is affected by many processes, and biomass is not directly derived by multiplying turnover with NPP. The sentence in P2238,L13-14 was revised accordingly as: “Biomass at equilibrium is positively correlated with both NPP and turnover time of carbon in biomass pools.”

P2238,L24: “This bias may be caused by non-modeled forest management in this region.” No need to speculate.

Response

This sentence was deleted accordingly.

P2239,L22: “Given the large…climate-carbon feedbacks.” Seems redundant.

Response

This sentence was deleted accordingly.

P2241,L1-9: What’s the overall message?

Response

This discussion is to explain why decrease in water availability when soil freezing is activated (Fig. 12b) leads to inconsistent changes in tree fractional cover (Fig. 12a). It is because fractional cover equals to population density multiplied by individual crown area, and decrease in WA affects these two variables contrarily. To clarify it, the following sentence was added at P2241,L9: “Therefore, reductions in WA may lead to inconsistent changes in tree fraction, depending on their relative effects on crown area and population density.”

P2243,L4-8: Not essentially. It depends how models implement bioclimatic constraints.

Response

The sentence at P2243,L4-7 was revised as: “…it is notable that this may bias DGVMs to produce unrealistic or unstable results, if vegetation distribution is sensitive to extreme temperatures in the model.”

P2244,L3-6: How would this help?

Response

The plant traits that describe the characteristics of each PFT define the behavior of PFTs in terms of distribution and vegetation carbon cycle. Using fixed traits, the terrestrial vegetation is represented by a limited number of PFTs; while using variable traits allows more variation in vegetation responses in the model. Verheijen et al. (2013) showed in their trait-variation simulation an improvement in resulted dominant vegetation types compared to fixed trait simulation, as well as enhanced climate-vegetation feedbacks when the DGVM was coupled to atmosphere model. Therefore, we think trait-variation might be an interesting direction in future development, especially to simulate vegetation acclimation to paleo or future climates.

P2252: Why start at 2?

Response

PFT1 in ORCHIDEE represents bare land. To clarify it, PFT1 was added in Table 1.

P2256: Why not include biomass densities in this table as well? Table 4 can be merged with this table as well.

Response

Forest biomass density can be readily calculated as total biomass in Table 5 divided by forest area in Table 4, so we did not include it.

P2267: Bad choice of colors. Please use better color scale.

Response

The color scale of Fig. 11 was changed.

P2268,2269: Tell your reader what + and - values mean rather than having them interpret it themselves.

Response

Variable names were added in each sub figure of Fig. 12 and 13.