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 all referee comments and short comment (in blue). For your convenience, changes in the revised manuscript are highlighted with dark red.

Referee #1

General Comments

This paper provides a description of revisions to the ORCHIDEE-HL (high latitude) land surface model intended to improve the simulation of Northern Hemisphere vegetation cover. The results are evaluated against several fractional land cover datasets and gridded observations of GPP, biomass and soil carbon. The authors claim "significant improvements" in simulated tree distributions and this appears to be justified. A particularly strength of the paper is that simulated PFT fractions are compared with multiple observational estimates, which takes into account the combined uncertainty in the source data and in the mapping from land cover classes to model PFTs. This allows the authors to place an informed emphasis on model errors and improvements in different regions.

The manuscript is well-written throughout and the figures are clear and understandable. With only a couple of exceptions, details of the model description that were not provided explicitly in the manuscript were found easily in the references provided (e.g., Krinner et al, Gouttevin et al).

Specific Comments

1. Despite being well used, it's not clear to me whether the 6-hourly CRU-NCEP forcing resolves the diurnal cycle adequately. In particular, the simulation of photosynthesis will depend strongly on the sub-daily representation of surface insolation. How are the forcing data downscaled from 6 hours to the 30 minute model time step? If these forcing fields are valid at the same UTC time rather than the same local solar time, is there any significant longitudinal variation in how well the diurnal cycles of insolation and GPP are represented?

Response

In ORCHIDEE, the meteorological fields of climate forcing are interpolated from their original time step to the half-hourly model time step. For fields other than downward solar radiation and precipitation, the 6-hourly data in CRU-NCEP are linearly interpolated to half-hourly resolution. For the short-wave radiation in particular, it is distributed as a function of solar angle, calculated based on longitude/latitude, the day of the year and the hour, according to the method used by GSWP (Dirmeyer, 2011; ORCHIDEE code see http://dods.ipsl.jussieu.fr/orchidee/DOXYGEN/webdoc/d1/db6/solar_8f90_source.html). The forcing fields and model outputs are valid at the same UTC time, for example at each time step, only half of the earth surface has solar radiation. The diurnal cycles of insolation and

GPP at different longitudes are thus corresponding to UTC time rather than their local time.

Reference:

Dirmeyer, P. A.: A history and review of the global soil wetness project (GSWP), J. Hydrometeorol., 12, 729–749, 2011.

---

2. The β diversity metric shows well the improvement in the high latitude tundra (Fig 5), but it doesn't highlight the greatly improved tree PFT fractions in northern Europe and eastern Canada. I would have expected this improvement between simulations to be more apparent in the metric, especially in the mean given the agreement between the observational datasets in these regions (Fig 3). It is more visible in the skill score (Fig 6) so, are there model errors and improvements that we should not expect to be able to evaluate through the use of this metric?

Response

It is true that in Fig. 5, the most highlighted regions are arctic tundra, with large β values between OLD and observational datasets and substantial improvement (β reduction) in NEW. But it can also be seen from Fig. 5 that the β metric is reduced in eastern Canada (from ~0.7 to ~0.3), and northern Europe and European Russia (from ~0.7 to ~0.4).

Tundra regions are more apparent in Fig. 5, because the OLD simulation produced very high fraction (>0.9) of needleaf deciduous trees in these regions that in reality have high fraction of bare land (PFT1); according to the definition of β diversity, this "extreme" bias of 2 PFTs, compared with evenly distributed bias among all PFTs, will more enlarge the value of β diversity. By contrast, in eastern Canada and northern Europe, besides the dominant needleleaf evergreen trees, other PFTs including broadleaf trees, grass and bare land can account for ~0.3. This relatively evenly distribution in vegetation (compared to that in tundra regions) avoids very large values of β diversity, even though the OLD simulation highly overestimated broadleaf trees in these regions. Therefore, the significant improvement in eastern Canada and northern Europe as shown in Fig. 4 did not turn into very obvious decrease of β Fig. 5.

The skill score (S_V) in Fig. 6, however, presents more visible improvement in northern Europe and eastern Canada. This is consistent with the high agreement among observational datasets (i.e., small β for data vs. data) in these regions shown in Fig. 3, because S_V is defined as β (data vs. data) divided by β (model vs. data) (Eq.9), and the small S_V value for OLD, due to a small numerator, makes the difference between OLD and NEW more visible. S_V highlights the improvement that is intuitionally shown in Fig. 3; in this sense, S_V is a good metric to evaluate model performance in simulating vegetation distribution.

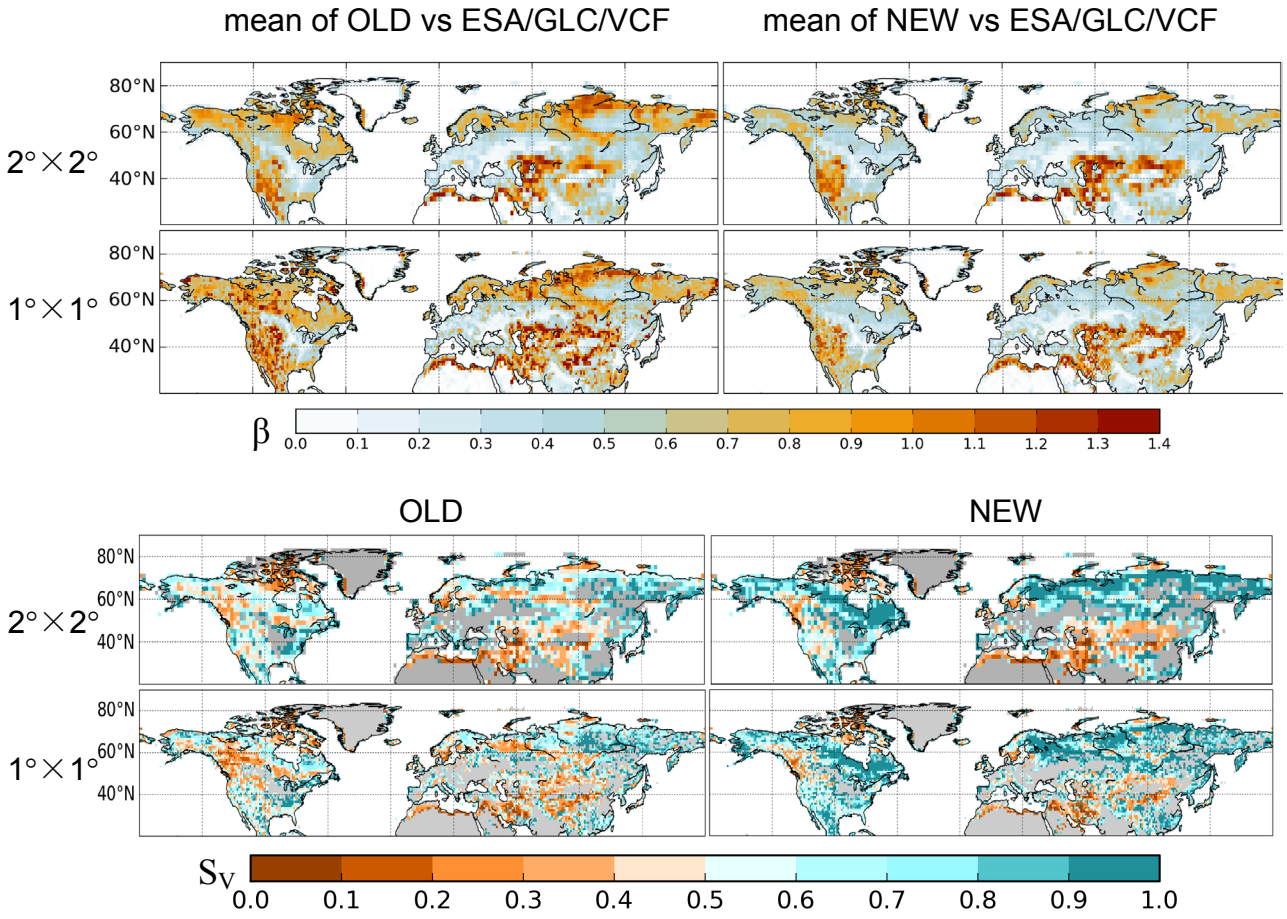
---

3. The authors highlight that these metrics (β, D and S) provide a framework that could be used by other models, and this type of multi-dataset analysis should undoubtedly be done in other studies. But how resolution dependent are these metrics likely to be? This would be a

tradeoff between the smoothing of coarser grids making it easier for a model to match observations, but also easier for observations to match each other. So would it be reasonable to compare models using significantly different grids? Could I calculate values for another model and compare them fairly with those in Table 3?

Response

Following this comment, we conducted two additional runs similar to OLD and NEW except for a 1°×1° resolution, in order to test the resolution dependency of these metrics. The figure below displays the new β value and skill score (S_V), compared with Fig. 5 (bottom panel) and Fig. 6:



As the figure shows, both β and S_V have similar spatial pattern in 1°×1° runs as in previous 2°×2° runs.

The β metric (Northern Hemisphere (20-90°N) mean) between models and observational datasets, and average S_V over different countries/regions are listed in the following tables:

β	2°×2°				1°×1°			
	OLD	NEW	ESA	GLC	OLD	NEW	ESA	GLC
ESA	0.58	0.56			0.70	0.62		
GLC	0.56	0.48	0.25		0.68	0.54	0.29	

VCF	0.65	0.47	0.37	0.35	0.77	0.52	0.43	0.41

S_V		Asian Russia	European Russia	Canada	USA	Europe	China	Northern Hemisphere (20°N-90°N)
2°×2°	OLD	0.68	0.63	0.53	0.66	0.62	0.57	0.60
	NEW	0.89	0.89	0.70	0.69	0.65	0.61	0.72
1°×1°	OLD	0.69	0.57	0.52	0.63	0.58	0.53	0.59
	NEW	0.87	0.91	0.71	0.73	0.67	0.66	0.74

In the coarser 2°×2° runs, due to smoothing effect, the β values for both model vs. data and data vs. data are decreased by 9~18% compared with 1°×1° runs. For S_V however, there is little difference between the two resolutions (relative differences are mostly within 5%), since the smoothing effect on both numerator and denominator partly offset each other. It indicates that the resolution at which the model runs has minor influence on the S_V metrics, and is not supposed to change the ranking of different models. Therefore, it is reasonable to calculate the skill score for other DGVMs with different grids, and compare them with the results in this study, if they adopt the same simulation protocol.

4. In the sensitivity experiments one piece of information that I couldn't glean was how do variations in the "1850" forest cover and GPP owing to spin up methodology (e.g., 1901 vs 1914, Fig 14) compare with the magnitude of 20th century change in the NEW and OLD simulations? Section 6.2 quotes 11.5% and 4.8% 20N-90N forest fraction increases with and without CO2 fertilisation, but it is difficult to compare these aggregate figures with the maps in Fig 14. This would provide some context for the warnings about spin up methodology. Also in section 6.3, the apparent motivation for the individual year simulations ("...recycled one-year climatic data are sometime used...") appears near the end after the results. It would be clearer if this was mentioned earlier in the section.

Response

Sect. 6.3 focused on the spin up methodology in terms of climate forcing, and the vegetation distribution results shown in this sector corresponded to the last year of spin up, i.e., the initial state (1850) of the transient simulation. The EXP3 experiment used the 20-year average climatology as forcing file in spin up; compared to NEW, total forest area in EXP3 increase by 5.1 Mkm^2 (22%), among which temperate trees (PFT4-6) increase by 2.7 Mkm^2, boreal needleleaf evergreen (PFT7) and broadleaf deciduous (PFT8) trees increase by 6.3 Mkm^2, and needleleaf deciduous tree (PFT9) decrease by 3.9 Mkm^2. This large variation in the initial forest cover owing to different climate forcings in spin up brings a warning on spin up methodology. Accordingly, the following sentences were added at P2242,L26: "In EXP3, temperate trees (PFT4-6) can extend northward, taking up the boreal tree positions, while the distribution of boreal needleleaf evergreen (PFT7) and broadleaf deciduous (PFT8) trees is squeezed to the climatic range of needleleaf deciduous tree (PFT9). Compared with the initial state after spin up in NEW, total forest area in the studied region (20-90°N) in EXP3 increase by 5.1 Mkm^2 (22%), among which PFT4-6 increase by 2.7 Mkm^2, PFT 7 and 8 increase by

6.3 Mkm^2, and PFT9 decrease by 3.9 Mkm^2."

To explain more clearly the motivation for the spin-up tests forced by individual year climate, the sentence "The large variance…" in P2242,L26 was replaced by "Apart from average climatology, recycled one single year climate is occasionally used in spin-up phase, which can also lead to large variance in initial vegetation distribution after spin-up due to interannual climate variability."

Technical Comments

Title: "...northern..." is a bit too vague. "...Northern Hemisphere high latitude..." would be more informative (and would reflect the model version).

Response

The title was revised accordingly: "Improving the dynamics of Northern Hemisphere high latitude vegetation in the ORCHIDEE ecosystem model".

P2219,L20: The repository that "rev1322" corresponds to isn't mentioned until Section 2.3.

Response

Since the original "Sect. 2.3 Code availability" was moved to the end (after "Sect. 7 Conclusions"), the two sentences in P2219,L20 were revised as "The basic structure of ORC-HL used in this study is shown in Fig. S1 in the Supplement, in which different processes from Krinner et al. (2005) are highlighted with red."

P2220,L13-15: It's not clear if V can be negative, e.g., though net biomass loss, which makes the range of possible MBG values unclear.

Response

Here V cannot be negative. To clarify it, the following sentence was added in the end of P2220,L15: "V equals to 0 in case of net annual biomass loss."

P2222,L21: "M_{SF} (t)" should be "M_{SF} (t, T_{min})", if I've interpreted the model correctly.

Response

$M_{SF}(t)$ was revised as $M_{SF}(t,T_{min})$ accordingly.

P2224,L8: Kuppel et al (2012) references a PhD thesis; can the same information be gleaned from Kuppel et al (2012), Biogeosciences, doi:10.5194/bg-9-3757-2012 ? If so, the latter reference is preferable.

Response

Kuppel et al (2012, Biogeosciences) presented a data assimilation system to optimize some ORCHIDEE parameters using measurements from temperate deciduous broadleaf forest sites, thus their results were only applied to PFT6 in ORCHIDEE; while in Kuppel's PhD thesis (Kuppel, 2012), parameters of other PFTs were optimized using the same method. So we cited

the PhD thesis (accessible from Internet) rather than the paper in Biogeosciences.

P2224,L19-21: It's not clear whether the leaf age dependency was switched off entirely or whether just very long time constant (a_{crit}) was used. The values in Table 1 for evergreen needleleaf are unchanged from Krinner et al, so is a_{crit} used elsewhere in the model? If not, why quote unused acrit values at all?

Response

Apart from the v_{cmax} (or j_{max}) dependency on leaf age discussed in Sect. 2.2.3, a_{crit} is also used to calculate leaf senescence in the turnover module in ORCHIDEE, so we still listed the a_{crit} values for evergreen needleleaf (PFTs 4 and 7) in Table 1.

The leaf age dependency of v_{cmax} (or j_{max}) for PFTs 4 and 7 was switched off. This v_{cmax} (or j_{max})–leaf age relationship was introduced in Krinner et al. (2005) to account for the influence of seasonal variation in leaf age on photosynthetic activity for trees; and we removed this rule for needleleaf evergreen trees since they do not have such significant seasonal variation in leaf age as deciduous trees do. To clarify it, we added a sentence at the end of Table 1 notes: "a_{crit}: critical leaf age for leaf senescence (days); the dependence of v_{cmax} and j_{max} on leaf age for PFTs 4 and 7 was eliminated as described in Sect. 2.2.3."

P2230,L3 L9: (Equation pedantry) The sum should be from "k = 1" rather than just "k".
P2230,L19: Similarly, the sums are missing upper limits.

Response

Equation (7), (8) and (9) were revised accordingly.

P2232,L26: Should be σ_O rather than σO.
P2233,L1: Are there missing modulus symbols, i.e., $|X_{c,M} - X_{c,O}| < \sigma_O$?
P2237,L7: "SG" should be "S_G".

Response

Revised accordingly.

Fig 2: "Brighter colors..." is ambiguous wording, "Deeper colors..." would be better. Should "...relative fraction..." be just "...fraction. . .", else it's not clear what it's relative to?

Response

Fig 2 caption was revised as: "…Color indicates the fraction of three PFT groups…Deeper colors represent higher fractional covers." Similarly, the "relative" in Fig 4 caption was deleted.

Fig 2 4: I find it difficult to determine how deep or pale these maps are relative to each other (e.g., OSIB vs IIASA). A limited scale (e.g., 25%,50%,75%,100%) for the pure RGB hues would be useful.

Response

A color scale was added in Fig. 2 and 4 accordingly.

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)^2 + (0.5-0)^2) = 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.

Short Comment by D. Lunt

Comment

Dear authors,

In my role as Executive editor of GMD, I would like to bring to your attention our Editorial:

http://www.geoscientific-model-development.net/gmd_journal_white_paper.pdf

http://www.geosci-model-dev.net/6/1233/2013/gmd-6-1233-2013.html

This highlights some requirements of papers published in GMD, which is also available on the GMD website in the 'Manuscript Types' section:

http://www.geoscientific-model-development.net/submission/manuscript_types.html

In particular, please note that for your paper, the following requirements have not been met in the Discussions paper – please correct this in your revised submission to GMD.

"– The paper must be accompanied by the code, or means of accessing the code, for the

purpose of peer-review. If the code is normally distributed in a way which could compromise the anonymity of the referees, then the code must be made available to the editor. The referee/editor is not required to review the code in any way, but they may do so if they so wish. “

“– All papers must include a section at the end of the paper entitled "Code availability". In this section, instructions for obtaining the code (e.g. from a supplement, or from a website) should be included; alternatively, contact information should be given where the code can be obtained on request, or the reasons why the code is not available should be clearly stated. ”

Yours,

Dan Lunt

Response

Dear Dan Lunt,

Thank you for the comment.

Following your comment, we carefully read these documents, and moved the original “Sect. 2.3 Code availability” to the end of the paper, with minor changes (in red): “The ORCHIDEE model used as a starting point in this study is ORCHIDEE-MICT rev1322. The source code can be obtained at http://forge.ipsl.jussieu.fr/orchidee/browser/branches/ORCHIDEE-MICT/ORCHIDEE?rev=1322. A detailed documentation and the forcing data needed to drive ORCHIDEE can be found at http://forge.ipsl.jussieu.fr/orchidee/wiki/Documentation and http://forge.ipsl.jussieu.fr/orchidee/wiki/Forcings. ORC-HL-NVD is derived from rev1322 with the modifications presented in Sect. 2.2, the source code of which can be obtained upon request (http://labex.ipsl.fr/orchidee/index.php/contact). The modifications of ORC-HL-NVD from rev1322 are also implemented in ORCHIDEE standard version (trunk), recorded as the difference between rev2672 (source code: http://forge.ipsl.jussieu.fr/orchidee/browser/trunk/ORCHIDEE?rev=2672) and rev2658 (source code: http://forge.ipsl.jussieu.fr/orchidee/browser/trunk/ORCHIDEE?rev=2658)”

Improving the dynamics of Northern Hemisphere high latitude ~~northern~~ vegetation in the ORCHIDEE ecosystem model

D. Zhu[1], S. S. Peng[1,2], P. Ciais[1], N. Viovy[1], A. Druel[2], M. Kageyama[1], G. Krinner[2], P. Peylin[1], C. Ottlé[1], S. L. Piao[3], B. Poulter[4], D. Schepaschenko[5], A. Shvidenko[5]

[1] Laboratoire des Sciences du Climat et de l'Environnement, LSCE CEA CNRS UVSQ, 91191 Gif Sur Yvette, France

[2] UJF Grenoble 1, Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE, UMR5183), Grenoble, France

[3] Department of Ecology, College of Urban and Environmental Sciences, Peking University, Beijing 100871, People's Republic of China

[4] Institute on Ecosystems and Department of Ecology, Montana State University, Bozeman, MT 59717, USA

[5] International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

Correspondence to: D. Zhu (dan.zhu@lsce.ipsl.fr)

Abstract

Processes that describe the distribution of vegetation and ecosystem succession after disturbance are an important component of dynamic global vegetation models (DGVMs). 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 ~~does not match the progress in modeling the rest of the physical and biogeochemical processes. Therefore, ORC-VD~~ is known to produce unrealistic results. This study presents a new parameterization of ORC-VD for mid-to-high latitude regions in the Northern Hemisphere, including processes that influence the existence, mortality and competition between tree functional types. A new set of metrics is also proposed to quantify the performance of ORC-VD, using up to five different datasets of

satellite land cover, forest biomass from remote sensing and inventories, a data-driven estimate of gross primary productivity (GPP) and two gridded datasets of soil organic carbon content. The scoring of ORC-VD derived from these metrics integrates uncertainties in the observational datasets. This multi-dataset evaluation framework is a generic method that could be applied to the evaluation of other DGVM models. The results of the original ORC-VD published in 2005 for mid-to-high latitudes and of the new parameterization are evaluated against the above-described datasets. Significant improvements were found in the modeling of the distribution of tree functional types north of 40°N. Three additional sensitivity runs were carried out to separate the impact of different processes or drivers on simulated vegetation distribution, including soil freezing which limits net primary production through soil moisture availability in the root zone, elevated CO_2 concentration since 1850, and the effects of frequency and severity of extreme cold events during the spin-up phase of the model.~~the return frequency of cold climate extremes causing tree mortality during the spin-up phase of the model.~~

1 Introduction

The terrestrial biosphere plays an important role in the carbon (Schimel, 1995; Ciais et al., 2013), water (Oki and Kanae, 2006) and energy balances of the Earth (Trenberth et al., 2009). Interactions between vegetation and the atmosphere involve complex biophysical and biogeochemical processes and feedbacks (Heimann and Reichstein, 2008; Foley et al., 2003). 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). To simulate past and future changes on long time scales, Earth system models must represent how the distribution of terrestrial ecosystems adjusts in response to changes in climate, CO_2 and land use. This need provides the motivation for the development of dynamic global vegetation models (DGVM), which couple 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) to simulate the distribution of vegetation, its carbon stocks and the fluxes exchanged with the atmosphere (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 representation of vegetation structural dynamics in DGVMs builds on principles previously applied in biogeography models and "gap models" (Sitch et al., 2003). Biogeography models define the patterns of vegetation physiognomy based on plant functional types (PFT) driven by temperature, precipitation, CO_2, climate-related disturbances, and soil properties (Prentice et al., 1992; Haxeltine and Prentice, 1996). Gap models on the other hand simulate forest dynamics at patch scale, including demographic processes (recruitment, growth, death), competition, and disturbance (Prentice and Leemans, 1990; Bugmann, 2001). 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).

Vegetation distribution largely depends on bioclimatic limits and competition between species, which are regrouped into PFTs in most DGVMs (Woodward, 1987; Sitch et al., 2003; Krinner et al., 2005). Bioclimatic limits consist of direct limiting factors (e.g., minimum temperature for survival) and indirect limitations that control primary productivity and in turn the competitive ability of a PFT (e.g., optimal temperature for photosynthesis, various temperature and moisture phenological controls of leaf-out and senescence). PFTs with a better tolerance to extreme climate conditions and higher growth efficiency during the growing season are more competitive than others, and their distribution will therefore expand at the end of each time step. 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. ~~In most DGVMs, the traits for a given PFT are fixed, but Verheijen et al. (2013) conducted a variable trait simulation with the JSBACH model for three leaf traits (Specific Leaf Area, and the constants defining the maximum rate of photosynthesis, v_{cmax}, j_{max}) based on observed trait-climate relationships, emphasizing the need for climate-dependent and regional trait variation modeling.~~ Higgins et al. (2014) however, pointed out the inherent limitations in Verheijen et al. (2013) using a statistical method to parameterize plant trait diversity, and proposed that the focus should not be on trait values, but rather on the trade-offs between traits (Scheiter et al., 2013). In this study, we will use a fixed trait approach to describe the characteristics of each PFT in ORCHIDEE (the PFTs are listed in Table 1).

ORCHIDEE is the terrestrial surface component of the Institut Pierre Simon Laplace (IPSL) Earth system model. Since the first model description by Krinner et al. (2005), the representation of existing processes has been improved and new processes have been implemented, such as a physically-based multi-layer soil hydrology scheme (de Rosnay et al., 2002), and a scheme describing soil freezing and its effects on root-zone soil moisture and soil thermodynamics (Gouttevin et al., 2012). These new parameterizations have been evaluated for static runs in which geographical distribution of PFTs is specified based on observed satellite land-cover information~~PFT maps were prescribed based on observed satellite land-cover information~~. Yet, their influence on the simulated PFT distribution when the vegetation dynamics module is activated has not been addressed. The original vegetation dynamics module in ORCHIDEE (hereafter "ORC-VD") described by Krinner et al. (2005) was adapted from the LPJ model (Sitch et al., 2003) with minor modifications. Unlike the rest of the model, ORC-VD has not been updated since the Krinner et al. (2005) description, and it produces unrealistic results in dynamic runs.~~which produces unrealistic results in dynamic runs.~~ For example, Woillez et al. (2011) have shown that the boreal forest area is largely modeled as broadleaf deciduous, whereas in reality it is mainly comprised of needleleaf trees.

The work described here improves ORC-VD, with a focus on Northern Hemisphere vegetation dynamics. Different sets of recent observations have been used to evaluate model performance using quantitative metrics, either related directly to the spatial distribution of

vegetation (satellite-observed land-cover and tree fraction) or resulting from it (data-driven spatial distribution of gross primary production (GPP), biomass and soil carbon stocks). The evaluation methodology developed here could be used for other DGVMs as well, and is thus of general interest for the DGVM modeling community.

We present a new parameterization of vegetation dynamics in the ORCHIDEE High Latitude version (ORC-HL) described by Gouttevin et al. (2012), with modifications to the equations and parameters describing tree mortality, thermal constraints and a calibration of photosynthesis parameters (v_{cmax}/j_{max}) (Sect. 2.2). The results of the original module (ORC-HL-OVD) and of the new parameterization (ORC-HL-NVD) are evaluated ~~against different satellite land-cover products, forest inventory data for forest area and biomass, and data-driven GPP and soil carbon products~~ (Sect. 4 and 5). Because the biogeochemical and physical processes that characterize high latitudes interact in a complex way with the processes that control vegetation structure, in Sect. 6 we performed and analyzed factorial model simulations changing one process or driver at a time, to isolate their impacts on vegetation distribution. In addition, because the initial distribution of the vegetation in 1850 is sensitive to pre-industrial climate conditions, we also tested the effect of the return frequency of cold extremes relating to tree mortality during the spin-up phase of the model and discussed its implications.

2 Model description

2.1 ORCHIDEE High Latitude

ORCHIDEE consists of two main modules: SECHIBA (the surface-vegetation-atmosphere transfer scheme) which simulates energy and water exchanges between the atmosphere and land surface at a half-hourly time-step, as well as photosynthesis based on enzyme kinetics (Ducoudré et al., 1993; de Rosnay and Polcher, 1998), and STOMATE (Saclay Toulouse Orsay Model for the Analysis of Terrestrial Ecosystems, Viovy, 1997) which simulates carbon dynamics at a daily time-step, including carbon allocation, biomass accumulation, litter and soil carbon decomposition, and phenology. STOMATE includes a dynamic vegetation module with equations adapted from the LPJ model (Sitch et al., 2003) as described by Krinner et al. (2005).

ORCHIDEE High Latitude version (ORC-HL) is an evolution of ORCHIDEE including additional high latitude processes, described by Gouttevin et al. (2012). In particular, the simple 2-layer soil hydrology (Ducoudré et al., 1993) was replaced by an 11-layer diffusion scheme (de Rosnay et al., 2002), which describes water infiltration and diffusion through soil in a physically-based way. A soil-freezing scheme is implemented in the 11-layer model to calculate liquid and ice water fractions in each soil layer. This scheme has been shown to improve the representation of pan-Arctic river discharge and soil thermal regimes in permafrost regions (Gouttevin et al., 2012). The basic structure of ORC-HL used in this study is shown in Fig. S1 in the Supplement,~~The version number of the ORC-HL used in this study is ORCHIDEE rev1322. Its basic structure is shown in Fig. S1 in the Supplement,~~ in which ~~different~~ processes different from Krinner et al. (2005) are marked red.

2.2 Modifications to ORCHIDEE vegetation dynamics

Figure 1 is a schematic of ORC-VD, 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:

$$V = CA \times P$$
$$\frac{dP}{dt} = E - M \times P \quad (1)$$

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 (d^{-1}), including components described in Sect. 2.2.1. The modifications made in this study are described in the following, shown red in Fig. 1.

2.2.1 Tree mortality

Mortality is defined as the ~~percentage~~ reduction in population density ~~at the end of each day~~during each time step (daily). The overall tree mortality rate (maximum 1) is the summation of each component including background mortality (M_{BG}), extreme coldness (M_{EC}) and spring frost (M_{SF}) related mortalities, fire-induced mortality, and light competition-induced mortality.

Background mortality

In ORC-HL-OVD, the default calculation of mortality rate for tree PFTs was the inverse of a PFT-specific longevity parameter (30 years for tropical trees, 40 years for temperate trees, 80

years for boreal trees). An alternative calculation in ORC-HL-OVD was a dynamic mortality related to growth efficiency, inherited from LPJ (Sitch et al., 2003):

$$M_{BG} = \left(\frac{k_{BG}}{1 + 0.035V} \right) / 365 \quad (21)$$

where M_{BG} is the dynamic background mortality for tree PFTs (d^{-1}); k_{BG} is maximum background mortality rate (yr^{-1}), set to 0.1 for all tree PFTs in ORC-HL-OVD; and V is vigor or growth efficiency, defined as the ratio of the net annual biomass increment to maximum LAI of the preceding year. V equals to 0 in case of net annual biomass loss.

The default calculation defines a constant mortality for each PFT in all grid cells, without considering the variations in mortality of that PFT caused by adaptation to different climate conditions. 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 ~~competition between tree PFTs~~ under various climates, but it does not consider longevity differences between PFTs. In the new version, ORC-HL-NVD, the dynamic M_{BG} formulation, Eq. (12), is again adopted, but k_{BG} is set to different values for tropical (0.14), temperate (0.1) and boreal (0.05) tree PFTs, proportional to the inverse of their respective longevities in the original ORC-HL-OVD model code.

Tree mortality during extremely cold days

In ORC-HL-OVD, when instantaneous minimum temperature on each day (T_{min}) drops below a PFT-dependent threshold ($T_{min,crit}$, Table 1), the corresponding tree PFT was completely eliminated ~~instantly~~. This assumption makes the vegetation distribution highly sensitive to the minimum temperature during a few extremely cold days, which varies from year to year. In reality, trees within a grid cell are unlikely to all die during a single extremely cold event, and moreover, at the resolution at which global models usually run (0.5° or coarser), a single minimum temperature cannot depict the heterogeneity within each grid cell. Therefore, we replaced the original threshold-based LPJ equation by a linearly increasing mortality rate as a function of daily minimum temperature, such that when $T_{min} < T_{min,crit}$

$$M_{EC} = k_{EC}(T_{min,crit} - T_{min}) \quad (23)$$

where M_{EC} is mortality caused by extreme coldness in winter (d^{-1}); k_{EC} =0.04, estimated by trial and error based on the return frequency of below-threshold T_{min} both within and between years according to the CRU-NCEP climate forcing.

The PFT-specific $T_{min,crit}$ (Table 1) confines the distribution of each tree PFT to their adaptable temperature zones. Boreal needleleaf deciduous trees (PFT9) have no $T_{min,crit}$ value, meaning that they are insensitive to extreme coldness, and thus can prevail over other boreal tree PFTs in the model in regions with extreme winters such as eastern Siberia.

Broadleaf tree mortality caused by spring frost

Broadleaf species have the specific property of being vulnerable to freezing events that occur after the spring leaf-out. Spring frost can cause damage to leaf buds, developing shoots and flowers, leading to reproductive failure and reduced peak growing-season leaf area index. These effects may result in a natural selection of species with a higher frost resistance, and affect species distribution in the long term (Augspurger, 2009). Kollas et al. (2013) found that minimum temperature during bud-break was a better predictor of the climate space of seven broadleaf tree species in Europe than winter temperature or mean growing-season temperature.

The change of temperature variability projected by climate models (Cohen et al., 2012; Screen, 2014) may increase or alleviate the risk of spring frost damage. Warmer winters and springs and earlier leaf presence may lead to a greater exposure of mid-latitude broadleaf species to spring frost events (Bokhorst et al., 2009; Gu et al., 2007), while the severity of individual cold spells may also decrease because of a faster warming of the Arctic compared to mid-latitudes (Screen, 2014). DGVMs must therefore represent spring frost induced mortality if they are to account for the response of broadleaf trees to altered climate variability.

We added a frost damage limitation to the distribution of the two broadleaf deciduous tree PFTs (PFT6 and PFT8). After leaf-out in the model, if daily minimum temperature drops below a threshold of –3°C (Kollas et al., 2013), tree mortality is assumed to increase with decreasing temperature. This frost-induced mortality is multiplied by the period elapsed since leaf-out, because the more time that has elapsed, the larger the mass of vulnerable foliage. Thus, during the consecutive 40 days after leaf-out when,

$T_{min} < T_{SF,crit}$ and $t - t_{leaf\text{-}out} < 40$ days

$M_{SF}(t, T_{min})$, the spring frost induced mortality for broadleaf deciduous trees in PFT6 and PFT8 (d^{-1}), is given by:

$$M_{SF}(t, T_{\min}) = 0.01(T_{SF,crit} - T_{min})\left\{\frac{t - t_{leaf-out}}{40}\right\} \quad (43)$$

where $T_{SF,crit}$ =–3°C; and $t_{leaf\text{-}out}$ is the day of the year when leaf-out was simulated in the model.

2.2.2 Growing-season temperature limits to tree extension

In the version of ORCHIDEE described by Krinner et al. (2005), a warm season air temperature (T_{ws}) limit was set to exclude all trees PFTs from cold Arctic regions, with T_{ws} being required to exceed 7°C for trees to become established or be able to stay at a grid point. T_{ws} was calculated using a linear relaxation method (a substitute for the running mean method to reduce computer memory requirement) given by:

$$T_{ws,t} = \frac{(\tau - \Delta t)T_{ws,(t-\Delta t)} + \Delta t T_{daily}}{\tau} \quad (54)$$

where Δt = time-step, 1day; τ = relaxation time of 60 day; and T_{daily} = daily mean air temperature.

In ORC-HL-OVD, used as a starting point for this study, this T_{ws} criterion had been removed. In ORC-HL-NVD, we re-introduced a growing-season temperature criterion to constrain tree extension to Arctic regions, but modified the original formulation using recent results. In their global study of temperature controls on high altitude treelines, Körner et al. (2004) found a growing-season mean soil temperature of 6.7±0.8°C to be the most consistent criterion to predict treelines across different climate zones. Other predictors tested (growing-season length, thermal sums and thermal extremes) were shown to have too large amplitudes and therefore be less suitable indicators of the altitudinal treeline position (Körner et al., 2004). We assumed that the cold limits of trees at both high altitude and high latitude are similar, which is supported by the recent study of Randin et al. (2013), and thus used the Körner et al. (2004) empirical results to re-define the thermal constraint on the existence of trees (treeline) in ORCHIDEE.

Combining the same definition of growing season as Körner et al. (2004), i.e., the period during which 10 cm depth soil temperature exceeds 3.2°C, with their linear relationship between soil temperature in the root zone and canopy air temperature, we prescribe the large-scale thermal limitation of trees in ORC-HL-NVD as follows: mean weekly air temperature during the growing season (T_{GS}) must exceed 7°C, corresponding to $T_{GS,root}$ larger than 6.7°C;

the growing season is calculated as the period when weekly air temperature is greater than 0°C, which corresponds closely to T_{root} above 3.2°C. The new T_{GS} criterion shows more consistency with the current treeline positions than the earlier T_{ws} criterion described by Krinner et al. (2005) (Fig. S2).

2.2.3 Modifying v_{cmax} and j_{max}

The values of the maximum rate of Rubisco carboxylase ($v_{cmax,opt}$) and maximum rate of photosynthetic electron transport ($j_{max,opt}$) for each PFT were revised using the results of the ORCHIDEE parameter optimization against flux tower measurements from Kuppel ~~et al.~~ (2012). Corresponding values are given in Table 1. In ORC-HL-OVD, v_{cmax} (or j_{max}) is the product of $v_{cmax,opt}$ (or $j_{max,opt}$) and a leaf efficiency factor (e_{rel}), itself determined by relative leaf age (a_{rel}). a_{rel} is defined as the ratio of the calculated leaf age since leaf-out considering four leaf cohorts to a PFT-dependent leaf longevity (a_{crit} in Table 1) (Krinner et al., 2005). As the value of a_{rel} increases with time since $t_{leaf\text{-}out}$, e_{rel} increases from 0 to 1 quickly at the beginning of the growing season, and then gradually decreases if $a_{rel} > 0.5$ when leaves become senescent near the end of the growing season. This rule was originally implemented to simulate the influence of seasonal variation in leaf age on photosynthetic activity for all tree PFTs. However, unlike deciduous trees, temperate and boreal evergreen needleleaf trees can keep their needles for 4-6 consecutive years, or even longer for some species (Richardson et al., 2000), resulting in a rather constant leaf age. Thus, we removed the dependence of v_{cmax} and j_{max} on leaf age for temperate and boreal evergreen needleleaf trees (PFTs 4 and 7) in ORC-HL-NVD.

~~2.3 Code availability~~

~~The ORCHIDEE model used as a starting point in this study is ORCHIDEE rev1322. The source code can be obtained at http://forge.ipsl.jussieu.fr/orchidee/browser/branches/ORCHIDEE-MICT/ORCHIDEE?rev=1322. A detailed documentation and the forcing data needed to drive ORCHIDEE can be found at http://forge.ipsl.jussieu.fr/orchidee/wiki/Documentation and http://forge.ipsl.jussieu.fr/orchidee/wiki/Forcings. ORC-HL-NVD is derived from rev1322 with the modifications presented in the Sect. 2.2, the source code of which can be obtained upon request.~~

3 Datasets and methods

3.1 Simulation protocol

Six different runs with ORC-HL (Table 2) were performed to test the impact of the new dynamic vegetation parameterizations and parameter calibrations. Since the modifications in vegetation dynamics module were mainly for temperate and boreal PFTs, the simulation domain is Northern Hemisphere from 20°N to 90°N. All runs were conducted at 2° resolution. The climate forcing files were from the 6-hourly CRU-NCEP dataset (http://dods.extra.cea.fr/store/p529viov/cruncep/V4_1901_2012/readme.htm), resampled from their original 0.5° data. CRU-NCEP is widely used as standard climate forcing in current offline terrestrial models, such as MsTMIP (Multi-scale synthesis and Terrestrial Model Intercomparison Project, Huntzinger et al., 2013) and TRENDY (Trends in net land-atmosphere carbon exchange over the period 1980-2010). Tests with different resolutions were carried out, showing quite similar results in the simulated vegetation distribution and carbon fluxes and pools (results not shown), indicating that the results presented below do not depend significantly on the spatial resolution of input climate and soil property data within the tested resolution range [0.5°, 2°].

Each simulation was preceded by a spin-up from bare ground (i.e., fractional cover of PFT1 equals to 1 everywhere). For the standard run with the new vegetation dynamics parameterizations (NEW), in spin-up, ORC-HL-NVD was forced by ~~cycling~~ repeatedly using CRU-NCEP 1901–1920 climate data and constant pre-industrial CO_2 concentration (285 ppm) for 250 years. Then the soil carbon sub-model was driven by the previous outputs for 1000 years for the soil carbon pools to reach equilibrium; this was followed by another 50 years of ORC-HL-NVD to complete the spin-up. Each transient simulation from 1850 to 2010 was started from the last year of the spin-up, forced by historical CRU-NCEP climate and rising CO_2 concentration. No climate data were available before 1901, so for that period, randomly selected years between 1901 and 1920 were used. The OLD run used the original vegetation dynamics equations from Krinner et al. (2005) in the ORC-HL version so that comparing NEW and OLD allows us to evaluate the improvements listed. The other four runs (EXP1–3, STAT) were similar to NEW except for one different setting for each run (Table 2). In EXP1, we deactivated soil freezing to test its impact on vegetation distribution. In EXP2, we used fixed CO_2 concentration at 285 ppm to test the sensitivity of vegetation distribution to rising CO_2. In EXP3, the model spin-up was forced by the CRU-NCEP 1901–1920 averaged

climatology instead of the 20-year cycle, in order to examine the impact of interannual climate variability on the initial PFT distribution after spin-up. In STAT runs, dynamic vegetation was deactivated and a fixed land-cover map was prescribed, in order to separate the effect of simulated versus observed PFT fractions on GPP, biomass and soil carbon. In STAT1 and STAT2, the PFT map was prescribed from ESA CCI land cover v1.1 (ESA, Bontemps et al., 2013, http://maps.elie.ucl.ac.be/CCI/viewer/index.php) and a synergetic land-cover product (SYNMAP, Jung et al., 2006), respectively.

Fires play an important role in determining vegetation patterns by preventing trees from achieving their climate potentials of height, biomass and fractional cover (Bond et al., 2005). Fire occurrence in ORC-HL is formulated using the fire model of Thonicke et al. (2001), based on litter quantity and moisture (Krinner et al., 2005). In this study, the fire module was activated in all the runs. But in a separate test, ORC-HL-NVD was run without the fire module. Compared to NEW, this simulation showed a small increase (5%) in the total temperate and boreal forest area in Northern Hemisphere (20°N–90°N) without fire. In this study, we used the relatively simple Thonicke et al. (2001) fire module, but compared the results with those obtained with SPITFIRE (Thonicke et al., 2010), a more sophisticated fire model, which explicitly simulates natural and human ignition, fire propagation and fuel combustion (Yue et al., 2014). The average annual burned area during 1981–2010 simulated by the Thonicke et al. (2001) fire module (as implemented in ORC-HL) is 2.7 Mkm^2 in Northern Hemisphere forests, similar to that simulated by SPITFIRE (2.1 Mkm^2, implemented in ORCHIDEE standard version).

In this study, agriculture is excluded from all the dynamic runs in order to simulate the potential vegetation distribution without croplands and pasture. The results were post-processed for comparison with observed vegetation cover or carbon stocks. For vegetation cover, this is done by subtracting the observed cropland fraction from the simulated natural PFT fraction in each grid:

$$V_{k,c} = V_{k,c,orig} \times (1 - V_{crop,c}) \quad (5\underline{6})$$

where $V_{k,c,orig}$ is the model simulated fractional vegetation cover for PFT k (except C3 and C4 crops) and for grid cell c; $V_{k,c}$ is the fraction of PFT k for grid cell c, after post-processing; and $V_{crop,c}$ is observed fraction of cropland for grid cell c, in this study we use croplands estimated from the ESA land-cover map.

For total GPP and soil carbon stocks, since ORCHIDEE outputs the values per unit PFT, which are multiplied by PFT fractions and summed up to derive the total amount, the results from dynamic runs were post-processed using the following equation (taking GPP as an example), to compare with observational data:

$$GPP_c = \sum_k^n (GPP_{k,c} \times V_{k,c}) + GPP_{crop,c} \times V_{crop,c} \quad (76)$$

where $GPP_{k,c}$ is GPP for natural PFT k and for grid cell c (g C m^{-2} yr^{-1} PFT^{-1}), simulated by dynamic runs; $GPP_{crop,c}$ is GPP for crops (including C3 and C4) for grid cell c (g C m^{-2} yr^{-1} PFT^{-1}), simulated by STAT1 (prescribed from the ESA map); GPP_c is total GPP for grid cell c (g C m^{-2} yr^{-1}), after post-processing; and n=11, the number of natural PFTs.

3.2 Evaluation datasets

We use satellite observations of land cover translated into the PFTs of ORCHIDEE to evaluate the simulated vegetation distribution. In order to account for uncertainties ~~of observations~~in observation-based estimates, we used three different land-cover maps: the ESA CCI land cover v1.1 for year 2010, GLC2000 (JRC, 2003) and ISLSCP II vegetation continuous field for 1992-1993 (Defries and Hansen, 2009). The first two land-cover products (hereafter "ESA" and "GLC") were converted from their original classifications (22 categories based on LCCS system) into PFT maps, using the cross-walking method of Poulter et al. (2011). The third product (hereafter "VCF") provides the fractional cover of bare ground, herbaceous vegetation and forest (further split into evergreen or deciduous, and broadleaf or needleleaf), and was merged with climate zones of the Köppen-Geiger classification system to resolve to PFT classes, based on Poulter et al. (2011). For Siberia, two additional regional land-cover maps were used, the PFT map of Siberia at 1km scale from Ottlé et al. (2013) based on the GlobCover2005 product (Bicheron et al., 2006), hereafter "OSIB", and the Russian land-cover dataset produced by International Institute for Applied Systems Analysis (Schepaschenko et al., 2011), hereafter "IIASA", which was converted into PFT map using the cross-walking method of Poulter et al. (2011). Along with ESA, GLC and VCF, the five land-cover products were used to evaluate the model skill at simulating the vegetation distribution across Siberia. The PFT maps were aggregated at 2°×2°, matching the resolution run by ORCHIDEE in this study. Figure 2 displays an RGB composite-color map of the vegetation fractional cover partitioned between broadleaf (including evergreen and deciduous,

red), needleleaf evergreen (green), and needleleaf deciduous (blue) trees, from the five PFT maps.

Simulated GPP was evaluated using the data-derived field obtained from FLUXNET data, satellite fAPAR and gridded climate and land-cover data using a model tree ensemble (Jung et al., 2011), hereafter "MTE". A recent forest carbon density map (Thurner et al., 2013) for Northern Hemisphere boreal and temperate forests (30°N–80°N), derived from radar remote sensing of growing-stock volume (GSV), was used to evaluate modeled forest biomass. For soil carbon stocks, the simulated soil carbon density was compared with the Harmonized World Soil Database (HWSD, 0–1m depth, FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) and the Northern Circumpolar Soil Carbon Database (NCSCD, Hugelius et al., 2013). Since the model results for soil carbon are not fully comparable to NCSCD due to lack of peatland carbon accumulation and cryoturbation processes in ORC-HL, metrics were not applied to soil carbon for establishing a model score. All gridded observation-derived data were aggregated at 2°×2°.

Apart from gridded data products based on satellite observations, independent forest inventory data at country/region level as compiled by Pan et al. (2011), including forest area and biomass, were also compared with model results.

3.3 Metrics for model evaluation

Different metrics can be used to quantify the agreement between model results and observations, including Pearson correlation, model-to-data deviation, mean error, root mean square error (see Kelley et al., 2013; Cadule et al., 2010). However, most of these metrics do not consider observational uncertainty. When there are multiple observations available and no particular dataset can be proved to be more accurate than others, which is the case for land cover, the choice of an observational dataset for model evaluation may have a large influence on the model performance score. In order to quantify the agreement between simulated and observed fields, as well as to integrate the uncertainty of observations, a metric normalized by observational uncertainty (Skill, *S*) was defined to evaluate model performances in terms of PFT fractional cover, GPP and forest biomass. For the following equations, M refers to the model results and O to observational data.

3.3.1 Metrics for PFT fractional abundance evaluation

For PFT fractions, a beta diversity metric (β) was used to calculate the disagreement between two different PFT maps, defined as the Euclidian distance of PFT classes (Poulter et al., 2011; Ottlé et al., 2013). For every grid cell c, beta diversity between model and observational dataset i (β_{c,M_Oi}) was calculated as:

$$\beta_{c,M\_Oi} = \sqrt[2]{\sum_{k=1}^{n}(V_{k,c,M} - V_{k,c,Oi})^2} \quad (87)$$

where $V_{k,c,M}$ is fractional abundance for PFT k and for grid cell c, simulated by model; $V_{k,c,Oi}$ is fractional abundance for PFT k and for grid cell c, from observational dataset i; and n=11, the number of natural PFTs.

Similarly, the disagreement between two observations was quantified using β_{c,Oi_Oj} , as:

$$\beta_{c,Oi\_Oj} = \sqrt[2]{\sum_{k=1}^{n}(V_{k,c,Oi} - V_{k,c,Oj})^2} \quad (98)$$

where $V_{k,c,Oi}$ and $V_{k,c,Oj}$ are fractional abundances from different observations i and j separately.

β is bound to the interval $[0, \sqrt{2}]$, with higher values representing larger discrepancies between two PFT maps. To take into consideration uncertainties of the different satellite land-cover products (Sect. 3.2), we use the mean β_{c,M_O} of the model versus all datasets normalized by the mean β_{c,O_O} of all combinations between different datasets. In order to derive a bounded score, with higher values representing better model performance, the metric for the model skill at simulating vegetation distribution in every grid cell ($S_{V,c}$) was defined as:

$$S_{V,c} = \left(\frac{1}{P}\sum_{i\neq j}^{P}\beta_{c,Oi\_Oj}\right)/\left(\frac{1}{Q}\sum_{i=1}^{Q}\beta_{c,M\_Oi}\right) \quad (910)$$

where P is the number of all combinations between different datasets; Q is the number of datasets; If $S_{V,c}>1$ for both models, indicating that the observation-based estimates have too large uncertainties to be qualified for model evaluation, this grid cell c is left out.

The $S_{V,c}$ of each grid cell was averaged over the Northern Hemisphere (20°N–90°N) to get an overall score (S_V). In the calculation of S_V, grid cells where mean β_{c,O_O} is higher than mean β_{c,M_O} for both models ($S_{V,c}>1$) were excluded, because in these pixels the uncertainties in the observational data are too large to qualify them for model evaluation – the choice of dataset might significantly alter the model evaluation result. Grid cells where both model and datasets have 100% bare ground (Sahara Desert and Greenland), and grid cells with a crop fraction higher than 0.5, were masked out (18% of the total number of land points in that part of the

Northern Hemisphere included in the study). The same rules were also applied to the calculation of regional average β_{c,M_O} and β_{c,O_O}.

To analyze the improvement of NEW over OLD for different PFTs, a dissimilarity index (D) was also calculated for groups of PFTs: broadleaf evergreen (PFT 2 and 5), broadleaf deciduous (PFT 3, 6 and 8), needleleaf evergreen (PFT 4 and 7), needleleaf deciduous (PFT 9), total tree, and grass (PFT 10 and 11). For each PFT *group* and grid cell c, $D_{group,c}$ was defined as the absolute bias in fractional cover between two maps:

$$D_{group,c,M\_Oi} = \left|V_{group,c,M} - V_{group,c,Oi}\right|$$
$$D_{group,c,Oi\_Oj} = \left|V_{group,c,Oi} - V_{group,c,Oj}\right| \quad (1\underline{1}\sout{0})$$

where $V_{group,c,M}$ is fractional abundance for PFT *group* and for grid cell c, simulated by the model; and $V_{group,c,Oi}$ and $V_{group,c,Oj}$ are fractional abundances from different observations i and j separately.

The average D_{group,M_O} and D_{group,O_O} were calculated over the studied region, in which the grid cells where the corresponding *group* does not exist in any of the models or observations, were excluded. In practice, we set a threshold of 0.01 to determine the existence of each *group*. We did not use the β equation here after re-grouping PFTs (e.g., needleleaf deciduous versus non-needleleaf deciduous, so that there are only two PFTs in the β equation), because in that case the average $\underline{\beta_{group,M\_O}}$ $\underline{(or\ \beta_{group,O\_O})\ for\ Northern\ Hemisphere\ (20\text{-}90°N)}$ $\sout{D_{group,M\_O}\ (or\ D_{group,O\_O})\ for\ the\ studied\ region}$ would be too optimistic, considering that many of the pixels will be equal to zero, due to the limited distribution range of the $\underline{corresponding}$ *group*.

Figure 3 shows the spatial pattern of β between the three observational datasets (ESA, GLC and VCF), and mean D among them for different PFT groups. The β between different datasets show a higher agreement for ESA versus GLC (an average β of 0.25) and lower agreement for VCF versus ESA or GLC (average β of 0.37 and 0.35 respectively). ESA and GLC legends are based on the FAO Land Cover Classification System (LCCS); while in VCF, the original 1 km continuous field data (DeFries et al., 2000), in which the forest fractional area is given for each grid cell instead of a discrete classification scheme, was aggregated to 0.5 degree resolution for ISLSCP II under the guidance of IGBP (International Geosphere-Biosphere Programme), by DeFries and Hansen (2009). LCCS uses a low threshold (15%) of tree cover for forest definition, whereas IGBP uses a threshold of 60% (Poulter et al., 2011), resulting in relatively lower tree cover in VCF than in either ESA or GLC land-cover maps.

For the PFT groups, higher D values were found for grassland, indicating significant ~~discrepancies~~ uncertainty in observed grassland fractions. The difference may come from uncertainties in the remotely sensed land-cover products, as well as from uncertainty in the reclassification of land-cover classes into PFT categories. The overlap of broadly defined arid-land classifications (i.e., grassland, shrubland, barren) of land-cover products can introduce errors in partitioning between trees, grass and bare land, in deserts and tundra regions (Poulter et al., 2011).

3.3.2 Metrics for GPP and forest biomass evaluation

GPP and forest biomass were evaluated using gridded observational data containing uncertainty estimates. The metric for model performances was defined as:

$$S_{G,c} \text{ or } S_{B,c} = \frac{\sigma_O}{|X_{c,M} - X_{c,O}|} \quad (12\text{~~1~~})$$

where $S_{G,c}$/$S_{B,c}$ is model skill at simulating GPP or forest biomass for grid cell c; $X_{c,M}$ is GPP or forest biomass for grid cell c, simulated by model; $X_{c,O}$ is GPP or forest biomass for grid cell c, from observation; and σ_O is the standard deviation of the observation.

In grid cells where $|X_{c,M} - X_{c,O}| < \sigma_O$, indicating a model-data difference within the uncertainty of the observational data, $S_{G,c}$ or $S_{B,c}$ is set to 1. The $S_{G,c}$ or $S_{B,c}$ of each grid cell were averaged over the Northern Hemisphere to get an overall score (S_G or S_B).

4 Modeled and observed vegetation distribution

4.1 Northern Hemisphere vegetation distribution

The present-day vegetation distributions simulated by OLD and NEW are shown in Fig. 4 as RGB composite-color maps the same as Fig. 2. Fractional covers for each PFT are shown in Fig. S3. Compared with OLD, NEW introduces two major improvements to the results. First, the tree distribution in cold subarctic regions has a northern boundary consistent with observations, mostly due to the introduction of a growing season temperature constraint (Sect. 2.2.2). Second, the observed dominance of needleleaf evergreen trees over broadleaf deciduous trees in northern Europe and North America is reproduced by NEW and not by OLD, an improvement mainly due to the introduction of the spring frost limitation for broadleaf deciduous trees (Eq. 4~~3~~) and the removal of the v_{cmax} (and j_{max}) leaf-age dependency for evergreen needleleaf trees (Sect. 2.2.3).

Figure 5 displays the spatial pattern of β index for OLD, NEW and different satellite land-cover products. Compared with OLD, the NEW results significantly reduce β in the boreal forests of Canada, western Siberia and northern Europe, consistent with results shown in Fig. 4. The disagreement is also reduced in pan-arctic tundra regions, after correction of the unrealistically high fraction of trees in these regions originally present in OLD. The average β over the Northern Hemisphere land surface (20°N–90°N, excluding bare ground and agricultural grid cells) for NEW versus ESA, GLC and VCF are 0.56, 0.48 and 0.47 respectively, equivalent to a 3.5%, 13% and 28% reduction (i.e., improvement) compared with OLD. The large variation of β for different observations shows the importance of accounting for uncertainty in observation-based estimates of land cover~~observed land-cover uncertainty~~ in DGVM evaluations, because the arbitrary choice of a specific land-cover product may result in quite different scores.

Accounting for uncertainty in observed PFT ~~maps~~distributions, the model skill at simulating the vegetation distribution (S_V) is shown in Fig. 6 for OLD and NEW. The average S_V for the major Northern Hemisphere forested countries or regions are listed in Table 3, showing improvement in all countries/regions. Larger improvements of NEW over OLD are found in European Russia (42%), Asian Russia (29%) and Canada (33%). The overall S_V for the Northern Hemisphere is 0.72 in NEW compared to 0.61 in OLD, equivalent to 18% improvement. In OLD, 13% of the land grid cells have a β_{c,M_O} value less than the uncertainty between different satellite products (β_{c,O_O}); in NEW, this fraction increases to 27%.

The forest areas simulated by the dynamic simulations and estimated from the land-cover products were aggregated to country level and compared with independent forest area from national forest inventories (Pan et al., 2011) (Table 4). In OLD, forest areas are systematically overestimated, especially for Asian Russia and Canada. The bias is decreased in NEW, for which most of the differences are less than 30% except for an overestimation in Canada (50%). This overestimation is, however, within the differences between the three land-cover products and the forest inventory data at country scale (Table 4). Forest areas estimated by VCF are systematically lower than inventory data, due to the difference in forest definition mentioned previously. The largest underestimation of VCF occurs in Asian Russia, where the vast taiga-tundra transition zones with relatively sparse trees make the definition-related biases more prominent.

4.2 Distribution of specific groups of Plant Functional Types

For the different PFT groups described in Sect. 3.3.1, the Northern Hemisphere average dissimilarity index (D) is plotted in Fig. 7 for OLD and NEW versus observations, as well as between different observational datasets. For the broadleaf evergreen group, D is small for both OLD and NEW, and similar to the uncertainty in the data, because the broadleaf evergreen fraction is smaller than other tree PFT groups in temperate and cold zones. For the broadleaf deciduous, needleleaf evergreen and needleleaf deciduous groups, the average D for NEW versus the three datasets is reduced (i.e., improved) by 53%, 13% and 67% respectively, compared with OLD. The OLD overestimation of broadleaf deciduous area in Canada, Scandinavia and European Russia is corrected in NEW (Fig. 8b). The large underestimation of needleleaf evergreen in OLD is partly corrected in NEW, but a significant underestimation of the needleleaf evergreen coverage still exists in southern Siberia and western Canada (Fig. 8c). For needleleaf deciduous, the unrealistically high fractions in subarctic regions in OLD are corrected in NEW, but needleleaf deciduous fractions in southern Siberia and Canada are still higher than observations, at the cost of needleleaf evergreen (Fig. 8d).

A strong disagreement between simulated and observed grassland fractions persists in NEW (average D of 0.35), but the data-data comparison also shows significant discrepancy (average D of 0.19) (Figs. 7 and 3). Since there are no specific shrubland and tundra PFTs in ORCHIDEE, the NEW simulation has high fractions of C3 grass (PFT10) in both arid and cold areas, including subarctic regions, the western USA and the middle of Eurasia (Fig. 8e). The average D for the grass fraction between OLD and observed land-cover maps is 0.27, lower than NEW, because the overestimations of tree cover in OLD decrease the distribution ranges of grassland, leading to a relatively higher agreement with observations for grassland cover than NEW.

4.3 Case study for Siberia, using regional land-cover datasets

For Siberia, the same metrics were calculated based on five observational datasets (ESA, GLC, VCF, OSIB and IIASA). As shown in Fig. 9a, the average β for NEW versus all datasets is significantly reduced compared to OLD along all longitudes, with a larger reduction (improvement) in central Siberia and the most eastern part of Russia. The average values of β over Siberia for NEW versus ESA, GLC, VCF, OSIB and IIASA are 0.59, 0.46, 0.38, 0.35 and 0.41 respectively, equivalent to 0%, 10%, 51%, 45% and 26% reduction

compared with OLD, respectively. The average β between different datasets is 0.37, with larger β between ESA and VCF (0.50) and between GLC and VCF (0.47), and smaller β for GLC and IIASA (0.23), VCF and OSIB (0.28), ESA and GLC (0.29). OSIB and VCF both have lower fractions of tree PFTs than the other three maps. In particular, the needleleaf deciduous fractions in OSIB and VCF for the densest forest areas are less than 0.65, while other maps can reach 0.85.

The model skill (S_V) that integrates observational uncertainty for Siberia is shown in Fig. 9b (OLD) and 9C (NEW). The average S_V for Siberia is 0.87 in NEW compared to 0.65 in OLD, equivalent to 32% improvement. In OLD, 11% of the Siberian grid cells have a β_{c,M_O} value less than the uncertainty between different satellite products (β_{c,O_O}); in NEW, this fraction increases to 40%.

5 Modeled and observed carbon stocks and GPP

5.1 Gross primary productivity

The latitudinal pattern of annual gross primary productivity (GPP) averaged for 1999–2008 from OLD and NEW is shown in Fig. 10, compared with STAT1 and STAT2 (prescribing ESA and SYNMAP land cover) and from the data-driven MTE GPP (Jung et al., 2011). For total GPP in the Northern Hemisphere (20°N–90°N), the 10-year average annual GPP simulated by NEW is 45.4 P g yr^{-1}, close to OLD (42.6 P g yr^{-1}) and MTE (42.2±2.4 P g yr^{-1}). As for the static runs, total GPP in STAT1 is 35.9 P g yr^{-1}, smaller than MTE. Since MTE by Jung et al. (2011) was based on SYNMAP land-cover data (Jung et al., 2006) to describe the vegetation at FLUXNET sites, STAT2 has a GPP (42.3 P g yr^{-1}) closer to MTE. The difference between STAT1 and STAT2 shows that the choice of land-cover map makes a strong impact on modeled GPP. Compared with ESA, SYNMAP has a larger forest area (29 versus 22 Mkm^2) and similar grassland area (~11 Mkm^2) for the northern hemisphere, explaining its larger GPP.

The spatial patterns of GPP simulated by OLD and NEW are similar (Figs. 10 and 11a). Compared with MTE, both NEW and OLD overestimates GPP in eastern USA, western Europe and southern Asia, and underestimates GPP in middle and eastern Siberia (Fig. 11a), indicating that the similarity in total Northern Hemisphere GPP between NEW and MTE masks compensating regional biases. The STAT1 and STAT2 runs produce very similar

patterns of GPP to those from NEW (not shown), suggesting that the regional bias of GPP in ORCHIDEE is not related to the modeled PFT distribution, but to other non-modeled factors such as nitrogen interactions.

The model skill at simulating annual GPP (S_G) averaged over different countries is given in Table 3. The average S_G for the Northern Hemisphere in OLD and NEW are similar (~0.6). The improvement in vegetation distribution in NEW does not lead to a significant improvement of GPP, probably because simulated GPP in the same grid cells for high latitudes has only a weak dependence on the modeled PFT. For example, in Canada and northern Europe needleleaf evergreen trees (PFT7) are dominant in NEW, but broadleaf deciduous trees (PFT8) are dominant in OLD, the GPP differences between these two PFTs are less than 1.5 g C m^{-2} yr^{-1} per PFT (or 25%), explaining why different modeled PFT fractions in this region do not result into large differences in GPP. This result means that GPP is not a discriminant variable for evaluating the performance of a vegetation dynamics module at high latitudes.

5.2 Forest biomass

The country-level forest biomass (above- and belowground) simulated by OLD, NEW and the two static runs with prescribed PFT maps were compared with forest inventory data from Pan et al. (2011) (Table 5). The satellite-based spatially explicit forest biomass estimates from Thurner et al. (2013) over temperate and boreal forests in 30°N–80°N were also aggregated to country level, showing generally good agreement with the data from Pan et al. The results in NEW are lower than the inventory for all countries, with the largest underestimation by 61% in Asian Russia. OLD gives a higher total forest biomass in Asian Russia, but the biomass density of OLD and NEW are similar (~2.4 kg C m^{-2} forest) and both lower than Pan et al. (4.1 kg C m^{-2} forest). The large overestimation of biomass in Canada by OLD is reduced in NEW, due to both reductions in forest area (Table 4, from 6.0 to 3.4 Mkm^2) and in biomass density (from 5.6 to 3.8 kg C m^{-2} forest). Considering the 50% overestimation of forest area in Canada by NEW compared to the inventory data from Pan et al. (Table 4), the small underestimation (6%) in total biomass results from a negative bias in biomass density simulation in the model. It is notable, however, that the biomass density in Canada estimated by Thurner et al. (3.7 kg C m^{-2} forest) is also significantly lower than that given by Pan et al. (6.1 kg C m^{-2} forest).

In order to separate the bias of simulated biomass density from the bias of modeled tree cover, the spatial distributions of forest biomass per unit forest area (kg C m^{-2} forest) simulated by OLD and NEW are shown in Fig. 11b and compared with the satellite-based estimates by Thurner et al. (2013). The original overestimation in eastern Canada, northern Europe and European Russia by OLD is improved in NEW, although underestimation in western Canada and Siberia still exists in NEW. Biomass at equilibrium is positively correlated with both NPP and turnover time of carbon in biomass pools.~~defined by the product of woody NPP multiplying the turnover time of carbon in biomass pools~~. Natural disturbances and forest management can thus lower biomass by reducing the turnover time (Jandl et al., 2007; Litton et al., 2004). Since older forests store more biomass carbon than younger forests (Wei et al., 2013; Luyssaert et al., 2008), managed and frequently burned forests may not be able to reach their climate-dependent maximum biomass.

In order to diagnose the possible causes of the biomass deviation from data, the ratio of forest biomass from NEW to that from Thurner et al., as well as the ratio of forest NPP (average during 2001–2010) from NEW to MODIS-NPP (NTSG), is plotted in Fig. S5~~4~~. In eastern Canada, forest biomass is overestimated by NEW, while NPP is close to MODIS NPP, indicating an overestimation of biomass carbon turnover time in ORCHIDEE compared to reality. ~~This bias may be caused by non-modeled forest management in this region.~~ In western Canada and southern Siberia, the underestimation of biomass is attributable to underestimation of NPP.

The model skill at simulating forest biomass (S_B) averaged over different countries is given in Table 3. S_B is improved in NEW for all countries compared to OLD, with the largest improvement found in Canada (66%). The overall S_B for 30°N–80°N is 0.59 in NEW, compared to 0.46 in OLD, equivalent to 28% improvement.

5.3 Soil carbon

The spatial patterns of soil carbon density simulated by OLD and NEW (0–2m depth) are shown in Fig. 11c, compared with that from HWSD (0–1m depth, FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) and NCSCD (0–1m depth, Hugelius et al., 2013). Over the grid cells present in NCSCD, the total soil carbon is 285 Pg in HWSD, markedly lower than that in NCSCD (460 Pg C for the upper meter of soil), indicating large uncertainties in the empirical soil carbon data. Since the ORC-HL in this study does not

include processes of peatland and wetland carbon accumulation, whereas in NCSCD the peat deposits contain about 30% of the total soil organic carbon mass in the upper meter (Tarnocai et al., 2009), and wetland carbon stock is estimated to account for 20% of the total 1-m-deep soil organic carbon pool in Russia (Schepaschenko et al., 2013), the model results are not fully comparable to NCSCD. The spatial patterns of soil carbon from OLD and NEW are similar (Fig. 11c). Over the grid cells present in NCSCD, the total soil carbon simulated by OLD and NEW is 263 and 283 Pg C, respectively.

A comparison of soil carbon simulations from several land surface models coupled with climate models in CMIP5 (Todd-Brown et al, 2013) suggested that most models cannot reproduce grid-scale variation in soil carbon; and that the substantial disagreement between the HWSD and NCSCD datasets and their lack of quantitative uncertainty estimates limit their ability for benchmarking land carbon models. ~~Given the large carbon storage in northern high latitude soils, the ability to accurately simulate high latitude processes such as permafrost, wetland and peatland carbon accumulation, is a prerequisite for realistic projections of future climate-carbon feedbacks.~~

6 Critical model processes influencing vegetation distribution

6.1 Soil freezing

The area of seasonally frozen ground covers 50% of the Northern Hemisphere land, or 48 Mkm^2 (Zhang et al., 2003). Soil freezing limits plant access to soil moisture, and thus impacts the simulated PFT distribution through a set of complex interactions between productivity, tree-grass competition, and soil water limitations. In permafrost regions, the limitation of growing-season water availablity due to soil freeze-thaw processes was shown to substantially contribute to the low vegetation carbon densities (Beer et al., 2007). In ORCHIDEE, a soil heat diffusion equation with latent heat (Gouttevin et al., 2012) is solved for each soil layer that impacts soil temperature and liquid water content. In this study, we tested the effects of soil freezing on the vegetation distribution by comparing NEW and EXP1 in which soil freezing processes were not activated (all other parameters being the same). In EXP1, soil temperature can drop below 0℃, but liquid water continues to be available in the root zone irrespective of soil temperature conditions. Figure 12 shows the difference in tree fraction and in water availability (WA) during the growing season (May-September) between NEW and

EXP1. In the model, soil moisture available to plants is defined by WA, the relative soil moisture in the root zone, weighted by PFT-specific root profiles. A value of WA = 0 defines the wilting point, and WA = 1 the field capacity. A stress factor is applied to stomatal conductance and canopy photosynthesis if WA drops below a critical value of 0.4, and this stress factor increases linearly for $0 < WA \leq 0.4$ (Krinner et al., 2005).

When soil freezes in autumn and winter, the amount of liquid water in the root zone is reduced as water is immobilized as ice in soil pores. In the growing season, WA in NEW is also lower than that in EXP1 (Fig. 12b). This is consistent with previous results of model validation at site scale (Gouttevin et al., 2012), in which the upper layer (0–20 cm) soil moisture in summer was found to be more depleted if the soil freezing module was activated. In regions underlain by permafrost, there is a spring peak in runoff originating from meltwater which does not infiltrate into frozen soils (Gouttevin et al., 2012). If soil freezing is not modeled as in EXP1, meltwater will infiltrate into soil, leading to overestimated soil water content in the growing season. The reduction of tree fraction in the presence of freezing occurs where there is significant reduction of WA (Fig. 12a). In areas with a small reduction (less than 0.1) in WA, however, there is a slight increase in tree fraction. The tree fraction in the model ~~is related to~~equals to population density ~~and~~ multiplied by individual crown area. On the one hand, as WA decreases, GPP, LAI and crown area are smaller; yet on the other hand, reduced LAI leads to increased available space for establishment, resulting in a subsequent increase in population density, compensating for the loss of crown area. Therefore, reductions in WA may lead to inconsistent changes in tree fraction, depending on their relative effects on crown area and population density.

6.2 Changing CO_2 since 1850

Terrestrial plants respond to elevated atmospheric CO_2 concentration by increasing assimilation rate and reducing diffusive stomatal conductance (Lammertsma et al., 2012), both processes are included in ORCHIDEE (Krinner et al., 2005). Under elevated CO_2 concentration, the enhanced photosynthetic capacity and thus increased NPP of forest (Norby et al., 2005; Hickler et al., 2008) leads to higher growth efficiency of trees and thus higher tree fractional coverage. In the model, tree PFTs are superior to grass PFTs in terms of light competition, i.e., when trees expand, grass PFTs will give way to trees. Therefore, tree cover is expected to increase at the cost of grasslands under elevated CO_2. Here we conducted a sensitivity test (EXP2) with fixed pre-industrial CO_2 concentration (285 ppm). Compared

with EXP2, the simulation NEW forced by historical CO_2 concentration produces higher tree fractions (Fig. 13a) by 2010, the spatial pattern of which mirrors the pattern of tree NPP increase (Fig. 13b) in the model. In NEW, total temperate and boreal forest area in the studied region (20°N–90°N) are modeled to increase by 2.6 Mkm^2 (11.5%) from 1850 to 2010. In EXP2 the increase is only 1.1 Mkm^2 (4.8%) indicating that about 58% of the increase in forest area is attributable to the historical increase of CO_2, the rest being attributable to climate warming (longer growing seasons) and changes in rainfall.

Since the processes of CO_2 uptake by photosynthesis and water loss by transpiration are tightly coupled, increasing CO_2 concentration results in increased water use efficiency (Lammertsma et al., 2012; O'ishi et al., 2009). Figure 13c displays the difference of WA for trees between NEW and EXP2. Compared to the fixed-CO_2 simulation, NEW produces higher WA by ~5% in mesic regions such as Europe, western Siberia and the eastern part of North America, and similar WA in drier regions such as middle and eastern Siberia and the western part of North America.

6.3 Effects of the return frequency and severity of extreme cold events during the spin-up

As mentioned in Sect. 2.2.1, the distribution range of tree PFTs in ORCHIDEE is influenced by extremely cold days in winter that varies from year to year. When the PFT-dependent threshold $T_{min,crit}$ (Table 1) is applied (Eq. 32), this mechanism results in a considerable difference in modeled tree fraction between the results of a spin-up forced by cycling multi-year climatic data versus an average climatology. In EXP3, the model spin-up used the average climatology of the period 1901–1920 from CRU-NCEP, and was compared with NEW where interanually variable climate from years 1901–1920 was repeated in a loop. The minimum temperature in winter (T_{min}) derived from the climatology is significantly higher than T_{min} considering the 20 individual years (Fig. 14a). Since the intra-annual variations among different years are not synchronous, a low temperature of a day in one year is offset by a higher temperature of the same day during another year; this leads to a milder climate in the climatology.

The vegetation distributions after spin-up are very different between NEW and EXP3 as shown in Fig. 14. In EXP3, temperate trees (PFT4-6)~~s~~ can extend northward, taking up the boreal tree positions, while the distribution of boreal needleleaf evergreen (PFT7) and

broadleaf deciduous (PFT8) trees is squeezed to the climatic range of needleleaf deciduous tree (PFT9). Compared with the initial state after spin up in NEW, total forest area in the studied region (20-90°N) in EXP3 increase by 5.1 Mkm^2 (22%), among which PFT4-6 increase by 2.7 Mkm^2, PFT7 and 8 increase by 6.3 Mkm^2, and PFT9 decrease by 3.9 Mkm^2. Apart from average climatology, recycled one single year climate is occasionally used in spin-up phase, which can also lead to large variance in initial vegetation distribution after spin-up due to interannual climate variability.~~The large variance induced by interannual climate variability during the spin-up also holds if one single year is used instead of the multi-year average.~~ Figure 14d shows the considerable difference in the fraction of PFT 7 and 8 between two spin-ups forced by two different single years arbitrarily chosen (1914 and 1901). Similar results were obtained when three sets of forcings (one-year, climatological mean, and cycling of the whole period 1960–1999) were used in the spin-up process of CLM-DGVM (Li et al., 2011). Since climatology or recycled one-year climatic data are sometimes used in the spin-up of land surface models, 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.~~considering the sensitivity of vegetation distribution to extreme temperatures.~~ Thus, it is more appropriate to cycle multi-year climatic data to force DGVMs in a spin-up.

7 Conclusions

This study has presented an improved parameterization and a calibration of Northern Hemisphere vegetation dynamics in the ORCHIDEE process-based ecosystem model, based on a version that includes frozen soil moisture and its impacts on plant productivity. Keeping the original model's concept of plant functional types, we modified the processes that influence tree existence, mortality and competition. A new performance metric applicable for DGVM evaluation in terms of vegetation fractional cover was used to evaluate ORCHIDEE, which integrates uncertainties in different land-cover maps. The new version of the ORCHIDEE vegetation dynamics module shows marked improvement in the simulated PFT distribution compared to the previous version. A more realistic simulation of the northern tree limit is obtained, as well as of the distribution of evergreen and deciduous conifers in the boreal zone. The model still overestimates grass fraction in dry regions of central Asia and western North America, possibly because of the lack of a specific shrubland PFT. Grass fraction was also overestimated in the Arctic tundra. Considering the large coverage of

shrubland and tundra in northern middle and high latitudes, a proper representation of shrub and tundra plant functional types in DGVMs, as well as their biophysical and biogeochemical processes, should be a priority for future development. The better PFT distribution results in improvements in simulated forest biomass, while significant regional biases still remain for GPP, forest biomass and soil carbon distributions, indicating other structural biases in the carbon cycle parameterizations in the model. Incorporating PFT trait variation into DGVMs, which allows the functional properties to vary within PFTs based on trait-climate relationships, might be a promising method to simulate vegetation acclimation that impacts both vegetation competition and the carbon cycle, and be an interesting future development.

Code availability

The ORCHIDEE model used as a starting point in this study is ORCHIDEE rev1322. The source code can be obtained at http://forge.ipsl.jussieu.fr/orchidee/browser/branches/ORCHIDEE-MICT/ORCHIDEE?rev=1322. A detailed documentation and the forcing data needed to drive ORCHIDEE can be found at http://forge.ipsl.jussieu.fr/orchidee/wiki/Documentation and http://forge.ipsl.jussieu.fr/orchidee/wiki/Forcings. ORC-HL-NVD is derived from rev1322 with the modifications presented in the Sect. 2.2, the source code of which can be obtained upon request (http://labex.ipsl.fr/orchidee/index.php/contact). The modifications of ORC-HL-NVD from rev1322 are also implemented in ORCHIDEE standard version (trunk), recorded as the difference between rev2672 (source code: http://forge.ipsl.jussieu.fr/orchidee/browser/trunk/ORCHIDEE?rev=2672) and rev2658 (source code: http://forge.ipsl.jussieu.fr/orchidee/browser/trunk/ORCHIDEE?rev=2658).

Acknowledgements

Dan Zhu received a PhD grant funded through the GAP Swedish-French project. P. Ciais acknowledges support of the ERC-Syg Grant P-Imbalance.

References

Augspurger, C. K.: Spring 2007 warmth and frost: phenology, damage and refoliation in a temperate deciduous forest, Funct. Ecol., 23, 1031–1039, 2009.

Beer, C., Lucht, W., Gerten, D., Thonicke, K. and Schmullius, C.: Effects of soil freezing and thawing on vegetation carbon density in Siberia: a modeling analysis with the Lund-Postdam-Jena dynamic global vegetation model (lpj-dgvm), Global Biogeochem. Cy., 21, GB1012, doi:10.1029/2006GB002760, 2007.

Bicheron, P., Leroy, M., Brockmann, C., Krämer, U., Miras, B., Huc, M., Ninô, F., Defourny, P., Vancutsem, C., Arino, O., Ranera, F., Petit, D., Amberg, V., Berthelot, B., and Gross, D.: GLOBCOVER: A 300m global land cover product for 2005 using ENVISAT/MERIS time series, Proceedings of the Second Recent Advances in Quantitative Remote Sensing Symposium, 538–543, 2006.

Bokhorst, S. F., Bjerke, J. W., Tømmervik, H., Callaghan, T. V. and Phoenix, G. K.: Winter warming events damage sub-arctic vegetation: consistent evidence from an experimental manipulation and a natural event, J. Ecol., 97, 1408–1415, 2009.

Bond, W. J., Woodward, F. I. and Midgley, G. F.: The global distribution of ecosystems in a world without fire, New Phytol., 165, 525–538, 2005.

Bontemps S., Defourny P., Radoux J., Van Bogaert E., Lamarche C., Achard F., Mayaux P., Boettcher M., Brockmann C., Kirches G., Zülkhe M., Kalogirou V., Seifert F.M., Arino O.: Consistent global land cover maps for climate modelling communities: current achievements of the ESA's land cover CCI, Proceedings of the ESA Living Planet Symposium, Edimburgh, 9-13 September, 2013

Bugmann, H.: A review of forest gap models, Climatic Change 51, 259–305, 2001.

Cadule, P., Friedlingstein, P., Bopp, L., Sitch, S., Jones, C. D., Ciais, P., Piao, S. L. and Peylin, P.: Benchmarking coupled climate-carbon models against long-term atmospheric CO_2 measurements, Global Biogeochem. Cycles, 24, 1–30, 2010.

Ciais, P., C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, C. Jones, C. Le Quéré, R.B. Myneni, S. Piao and P. Thornton: Carbon and Other Biogeochemical Cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the

Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Cohen, J. L., Furtado, J. C., Barlow, M. A., Alexeev, V. A., and Cherry, J. E.: Arctic warming, increasing snow cover and widespread boreal winter cooling, Environ. Res. Lett., 7, 014007, 20 doi:10.1088/1748-9326/7/1/014007, 2012.

Cox, P. M., Betts, R. A., Jones, C. D. and Spall, S. A.: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, Nature, 408, 184–187, 2000.

de Rosnay, P. and Polcher, J.: Modelling root water uptake in a complex land surface scheme coupled to a GCM, Hydrol. Earth Syst. Sci., 2, 239–255, doi:10.5194/hess-2-239-1998, 1998.

de Rosnay, P., Polcher, J., Bruen, M., and Laval, K.: Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes, J. Geophys. Res.-Atmos., 107, 4118, doi:10.1029/2001JD000634, 2002.

DeFries, R. and Hansen, M. C.: ISLSCP II continuous fields of vegetation cover, 1992–1993: in: ISLSCP Initiative II Collection, Data set, edited by: Forrest, G., Collatz, G., Meeson, B., Los, S., Brown de Colstoun, E., and Landis, D., available at: http://daac.ornl.gov/ (last access: November 2014), Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, 2009.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., and Loveland, T. R.: A new global 1 km dataset of percentage tree cover derived from remote sensing, Glob. Change Biol., 6, 247–254, 2000.

Ducoudré, N. I., Laval, K., Perrier, A.: SECHIBA, a new set of parameterizations of the hydrologic exchanges at the land-atmosphere interface within the LMD atmospheric general circulation model, J. Climate, 6, 248–273, 1993

FAO/IIASA/ISRIC/ISSCAS/JRC: Harmonized World Soil Database (version 1.10), FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2012.

Foley, J. A., Costa, M. H., Delire, C., Ramankutty, N., Costaz, M. H. and Snyder, P.: How terrestrial ecosystems could affect earth's climate, Front. Ecol. Environ., 1, 38–44, 2003.

Gouttevin, I., Krinner, G., Ciais, P., Polcher, J. and Legout, C.: Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology, Cryosphere, 6, 407–430, 2012.

Gu, L., Hanson, P. J., Post, W. Mac, Kaiser, D. P., Yang, B., Nemani, R., Pallardy, S. G., and Meyers, T.: The 2007 eastern US spring freeze: increased cold damage in a warming world, Bioscience, 58, 253–262, doi:10.1641/B580311, 2008.

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.

Heimann, M. and Reichstein, M.: Terrestrial ecosystem carbon dynamics and climate feedbacks, Nature, 451, 289–292, 2008.

Hickler, T., Smith, B., Prentice, I. C., Mjöfors, K., Miller, P., Arneth, A. and Sykes, M. T.: CO_2 fertilization in temperate FACE experiments not representative of boreal and tropical forests, Glob. Chang. Biol., 14, 1531–1542, 2008.

Higgins, S. I., Langan, L., and Scheiter, S.: Progress in DGVMs: a comment on "Impacts of trait variation through observed trait–climate relationships on performance of an Earth system model: a conceptual analysis" by Verheijen et al. (2013), Biogeosciences, 11, 4357–4360, doi:10.5194/bg-11-4357-2014, 2014.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P. and Swanson, D. K.: The northern circumpolar soil carbon database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst. Sci. Data, 5, 3–13, 2013.

Huntzinger, D. N., Schwalm, C., Michalak, a. M., Schaefer, K., King, a. W., Wei, Y., Jacobson, a., Liu, S., Cook, R. B., Post, W. M., Berthier, G., Hayes, D., Huang, M., Ito, a., Lei, H., Lu, C., Mao, J., Peng, C. H., Peng, S., et al.: The North American carbon program multi-scale synthesis and terrestrial model intercomparison project -part 1: overview and experimental design, Geosci. Model Dev., 6, 2121–2133, 2013.

Jandl, R., Lindner, M., Vesterdal, L., Bauwens, B., Baritz, R., Hagedorn, F., Johnson, D. W., Minkkinen, K. and Byrne, K. A: How strongly can forest management influence soil carbon sequestration? Geoderma, 137, 253–268, 2007.

JRC: Global Land Cover 2000 database, European Commission, Joint Research Centre, available at: http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php/ (last access: November 2014), 2003.

Jung, M., Henkel, K., Herold, M. and Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling, Remote Sens. Environ., 101, 534–553, 2006.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., Williams, C.: Global patterns of land–atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteoro logical observations, J. Geophys. Res.-Biogeo., 116, G00J07, doi:10.1029/2010JG001566, 2011.

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B. and Willis, K. O.: A comprehensive benchmarking system for evaluating global vegetation models, Biogeosciences, 10, 3313–3340, 2013.

Kollas, C., Körner, C. and Randin, C. F.: Spring frost and growing season length co-control the cold range limits of broad-leaved trees, J. Biogeogr., 41, 773–783, 2014.

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

Koven, C., Friedlingstein, P., Ciais, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: On the formation of high-latitude soil carbon stocks: e□ects of cryoturbation and insulation by organic matter in a land surface model, Geophys. Res. Lett., 36, L21501, doi:10.1029/2009GL040150, 2009.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere–biosphere system, Global Biogeochem. Cy., 19, GB1015, doi:10.1029/2003GB002199, 2005.

Kuppel, S.: Assimilation de mesures de flux turbulents d'eau et de carbone dans un modèle de la biosphère continentale, Ph.D. thesis, Le Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Université de Versailles Saint-Quentin-en-Yvelines, France, 2012.

Lammertsma, E. I., de Boer, H. J., Dekker, S. C., Dilcher, D. L., Lotter, A. F. and Wagner-Cremer, F.: Global CO_2 rise leads to reduced maximum stomatal conductance in florida vegetation., Proc. Natl. Acad. Sci. U. S. A., 108, 4035–4040, 2011.

Li, F., Zeng, X., Song, X., Tian, D., Shao, P., and Zhang, D.: Impact of spin-up forcing on vegetation states simulated by a dynamic global vegetation model coupled with a land surface model, Adv. Atmos. Sci., 28, 775–788, doi:10.1007/s00376-010-0009-0, 2011.

Litton, C. M., Ryan, M. G. and Knight, D. H.: Effects of tree density and stand age on carbon allocation patterns in postfire lodgepole pine, Ecol. Appl., 14, 460–475, 2004.

Luyssaert, S., Schulze, E.-D., Börner, A., Knohl, A., Hessenmöller, D., Law, B. E., Ciais, P. and Grace, J.: Old-growth forests as global carbon sinks., Nature, 455, 213–5, 2008.

Norby, R. J., Delucia, E. H., Gielen, B., Calfapietra, C., Giardina, C. P., King, J. S., Ledford, J., McCarthy, H. R., Moore, D. J. P., Ceulemans, R., De Angelis, P., Finzi, A. C., Karnosky, D. F., Kubiske, M. E., Lukac, M., Pregitzer, K. S., Scarascia-Mugnozza, G. E., Schlesinger, W. H. and Oren, R.: Forest response to elevated CO_2 is conserved across a broad range of productivity, Proc. Natl. Acad. Sci. U. S. A., 102, 18052–6, 2005.

NTSG (Numerical Terradynamic Simulation Group): MODIS GPP/NPP Project (MOD17A3), available at: http://www.ntsg.umt.edu/project/mod17 (last access: November 2014), 2004.

O'ishi, R., Abe-Ouchi, A., Prentice, I. C., and Sitch, S.: Vegetation dynamics and plant CO_2 responses as positive feedbacks in a greenhouse world, Geophys. Res. Lett., 36, L11706, doi:10.1029/2009GL038217, 2009.

Oki, T. and Kanae, S.: Global hydrological cycles and world water resources., Science, 313, 1068–1072, 2006.

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.

Pan, Y., Birdsey, R. a, Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. a, Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire, a D., Piao, S., Rautiainen, A., Sitch, S. and Hayes, D.: A large and persistent carbon sink in the world's forests., Science, 333, 988–93, 2011.

Piao, S., Friedlingstein, P., Ciais, P., Zhou, L., and Chen, A.: Effect of climate and CO_2 changes on the greening of the Northern Hemisphere over the past two decades, Geophys. Res. Lett., 33, L23402, doi:10.1029/2006GL028205, 2006.

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.

Prentice, I. C. and Leemans, R.: Pattern and process and the dynamics of forest structure: a simulation approach, J. Ecol., 78, 340–355, 1990.

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.

Quillet, A., Peng, C. and Garneau, M.: Toward dynamic global vegetation models for simulating vegetation–climate interactions and feedbacks: recent developments, limitations, and future challenges, Environ. Rev., 18, 333–353, 2010.

Randin, C. F., Paulsen, J., Vitasse, Y., Kollas, C., Wohlgemuth, T., Zimmermann, N. E. and Körner, C.: Do the elevational limits of deciduous tree species match their thermal latitudinal limits?, Glob. Ecol. Biogeogr., 22, 913–923, 2013.

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

Schepaschenko, D., McCallum, I., Shvidenko, A., Fritz, S., Kraxner, F. & Obersteiner, M.: A new hybrid land cover dataset for Russia: a methodology for integrating statistics, remote sensing and in situ information. Journal of Land Use Science, 6, 245–259, 2011.

Schepaschenko, D. G., Mukhortova, L. V., Shvidenko, a. Z. and Vedrova, E. F.: The pool of organic carbon in the soils of Russia, Eurasian Soil Sci., 46, 107–116, 2013.

Schimel, D. S.: Terrestrial ecosystems and the carbon-cycle, Glob. Chang. Biol., 1, 77–91, 1995.

Screen, J. A.: Arctic amplification decreases temperature variance in northern mid- to high latitudes, Nat. Clim. Chang., 4, 577–582, doi:10.1038/NCLIMATE2268, 2014.

Sitch, S., Smith, B., Prentice, I. C., Arneth, a., Bondeau, a., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K. and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Glob. Chang. Biol., 9, 161–185, 2003.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, Global Biogeochem. Cy., 23, GB2023, doi:10.1029/2008GB003327, 2009.

Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L. and Carmona-Moreno, C.: The influence of vegetation, fire spread and fire behaviour on biomass burning and trace gas emissions: results from a process-based model, Biogeosciences, 7, 1991–2011, 2010.

Thonicke, K., Venevsky, S. and Sitch, S.: The role of fire disturbance for global vegetation dynamics : coupling fire into a dynamic global vegetation model, Global Ecol. Biogeogr., 10, 661–677, 2001.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R. and Schmullius, C.: Carbon stock and density of northern boreal and temperate forests, Glob. Ecol. Biogeogr., 23, 297–310, 2014.

Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. a. G. and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 earth system models and comparison with observations, Biogeosciences, 10, 1717 – 1736, 2013.

Trenberth, K. E., Fasullo, J. T. and Kiehl, J.: Earth's global energy budget, Bull. Am. Meteorol. Soc., 90, 311–323, 2009.

Verheijen, L. M., Brovkin, V., Aerts, R., Bönisch, G., Cornelissen, J. H. C., Kattge, J., Reich, P. B., Wright, I. J. and Van Bodegom, P. M.: Impacts of trait variation through observed trait-climate relationships on performance of an earth system model: a conceptual analysis, Biogeosciences, 10, 5497–5515, 2013.

Viovy, N.: Interannuality and CO_2 sensitivity of the SECHIBA-BGC coupled SVAT-BGC model, Phys. Chem. Earth, 21, 489–497, 1997.

Wang, T., Ottlé, C., Boone, A., Ciais, P., Brun, E., Morin, S., Krinner, G., Piao, S. and Peng, S.: Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model, J. Geophys. Res. Atmos., 118, 6064–6079, 2013.

Wei, Y., Li, M., Chen, H., Lewis, B. J., and Yu, D.: Variation in Carbon storage and its distribution by stand age and forest type in boreal and temperate forests in Northeastern China, PLoS ONE, 8, e72201, doi:10.1371/journal.pone.0072201, 2013.

Woillez, M.-N., Kageyama, M., Krinner, G., de Noblet-Ducoudré, N., Viovy, N. and Mancip, M.: Impact of CO_2 and climate on the last glacial maximum vegetation: results from the ORCHIDEE/IPSL models, Clim. Past, 7, 557–577, 2011.

Woodward, F. I. and Williams, B. G.: Climate and plant distribution at global and local scales, Vegetatio, 69, 189-197, 1987.

Yue, C., Ciais, P., Cadule, P., Thonicke, K., Archibald, S., Poulter, B., Hao, W. M., Hantson, S., Mouillot, F., Friedlingstein, P., Maignan, F., and Viovy, N.: Modelling the role of fires in the terrestrial carbon balance by incorporating SPITFIRE into the global vegetation model ORCHIDEE – Part 1: simulating historical global burned area and fire regimes, Geosci. Model Dev., 7, 2747–2767, doi:10.5194/gmd-7-2747-2014, 2014.

Zhang, T., Barry, R., Knowles, K., Ling, F., and Armstrong, R.: Distribution of seasonally and perennially frozen ground in the Northern Hemisphere, in: Proceedings of the 8th International Conference on Permafrost, 21-25 July 2003, Zurich, Switzerland, 1289–1294, 2003.

Table 1. PFT-specific parameters in ORC-HL-NVD.

PFT	$T_{min,crit}$	k_{BG}	$v_{cmax,opt}$	$j_{max,opt}$	a_{crit}
1: bare ground	/	/	/	/	/
2: tropical broadleaf evergreen trees	0	0.14	65	130	730
3: tropical broadleaf dry-season deciduous trees	0	0.14	65	130	180
4: temperate needleleaf evergreen trees	-30	0.1	35	70	910
5: temperate broadleaf evergreen trees	-14	0.1	45	90	730
6: temperate broadleaf summergreen trees	-30	0.1	55	110	180
7: boreal needleleaf evergreen trees	-45	0.05	33	66	910
8: boreal broadleaf summergreen trees	-45	0.05	30	60	180
9: boreal needleleaf summergreen trees	/	0.05	35	70	180
10: natural C3 grass	/	/	70	140	120
11: natural C4 grass	/	/	70	140	120
12: agricultural C3 grass	/	/	100	200	90
13: agricultural C4 grass	/	/	100	200	90

$T_{min,crit}$: minimum temperature limitation (°C), below which the mortality rate will increase as Eq.(~~2~~3). k_{BG} : maximum background mortality rate (yr^{-1}) for tree PFTs. $v_{cmax,opt}$: optimal maximum rubisco-limited potential photosynthetic capacity (μmol m^{-2} s^{-1}). $j_{max,opt}$: maximum rate of photosynthetic electron transport (μmol m^{-2} s^{-1}). a_{crit} : critical leaf age for leaf senescence (days); the dependence of v_{cmax} and j_{max} on leaf age for PFTs 4 and 7 was eliminated as described in Sect. 2.2.3.~~-~~

Table 2. Characteristics of each ORC-HL off-line runs. OLD follows the same simulation protocol as NEW. EXP1–3 and STAT is similar to NEW except for one different setting for each run.

Name	Model	Module	Spin-up		Simulation (1850-2010)	
			Climate forcing	CO_2 level	Climate forcing	CO_2 level
NEW	ORC-HL-NVD	Activate ORC-VD, soil freezing and fire schemes	CRU-NCEP 1901-1920 cycle	285ppm	CRU-NCEP 1901-2010 (for 1850-1900: randomly select from 1901-1920)	rising
OLD	ORC-HL-OVD	/	/	/	/	/
EXP1	ORC-HLNVD	Deactivate soil freezing				
EXP2	ORC-HL-NVD	/	/	/	/	fixed at 285ppm
EXP3	ORC-HL-NVD	/	CRU-NCEP 1901-1920 average climatology	/	/	
STAT1	ORC-HL-NVD	Deactivate ORC-VD (PFT map prescribed from ESA)	/	/	/	/
STAT2	ORC-HL-NVD	Deactivate ORC-VD (PFT map prescribed from SYNMAP)	/	/	/	/

Table 3. Model skills at simulating vegetation distribution (S_V), GPP (S_G) and forest biomass (S_B), averaged over different countries/regions. STAT1 and STAT2 are static runs prescribing different PFT maps, ESA and SYNMAP.

		Asian Russia	European Russia	Canada	USA	Europe	China	Northern Hemisphere (20°N-90°N)
Vegetation distribution	OLD	0.69	0.63	0.53	0.66	0.62	0.57	0.61
	NEW	0.89	0.89	0.70	0.69	0.65	0.61	0.72
GPP	OLD	0.53	0.70	0.59	0.63	0.60	0.57	0.63
	NEW	0.58	0.68	0.50	0.65	0.60	0.56	0.62
	STAT1	0.52	0.63	0.63	0.54	0.55	0.53	0.60
	STAT2	0.50	0.68	0.65	0.65	0.53	0.50	0.63
Forest biomass	OLD	0.52	0.54	0.37	0.49	0.49	0.56	0.46
	NEW	0.62	0.73	0.62	0.57	0.55	0.56	0.59
	STAT1	0.58	0.55	0.50	0.53	0.54	0.52	0.56
	STAT2	0.57	0.47	0.46	0.47	0.54	0.53	0.54

Table 4. Forest areas (Mkm^2) for different countries/regions simulated by models (OLD and NEW) and estimated from land cover products (ESA, GLC, VCF), in comparison with that from Pan et al. (2011). The relative differences compared to Pan et al. (2011) are given in parentheses.

	Asian Russia	European Russia	Canada	USA	Europe	China
Pan et al. (2011)	6.77	1.69	2.30	2.57	2.05	1.56
OLD	10.0 (48%)	1.96 (16%)	6.00 (160%)	3.33 (30%)	2.14 (5%)	2.80 (80%)
NEW	5.00 (-26%)	1.80 (7%)	3.44 (50%)	2.61 (2%)	1.56 (-24%)	1.23 (-21%)
ESA	6.54 (-3%)	1.58 (-6%)	3.64 (58%)	3.00 (17%)	1.81 (-12%)	2.19 (41%)
GLC	8.42 (25%)	2.02 (20%)	4.50 (96%)	4.73 (84%)	2.40 (17%)	2.23 (43%)
VCF	3.43 (-49%)	1.18 (-30%)	2.54 (10%)	2.00 (-22%)	1.19 (-42%)	1.10 (-30%)

Table 5. Forest biomass (Pg C) for different countries/regions simulated by models (OLD, NEW and two static runs) and estimated from Thurner et al. (2013), in comparison with that from Pan et al. (2011). STAT1 and STAT2 prescribe different PFT maps, ESA and SYNMAP. The relative differences compared to Pan et al. (2011) are given in parentheses.

	Asian Russia	European Russia	Canada	USA	Europe	China
Pan et al. (2011)	27.9	9.6	14.0	19.4	13.0	6.5
Thurner et al. (2013)	25.2 (-10%)	9.0 (-6%)	15.9 (14%)	-	10.6 (-18%)	-
OLD	24.3 (-13%)	14.7 (53%)	33.4 (138%)	17.7 (-9%)	16.2 (27%)	8.1 (25%)
NEW	11.0 (-61%)	7.6 (-21%)	13.2 (-6%)	12.0 (-38%)	8.3 (-36%)	3.5 (-47%)
STAT1	6.9 (-75%)	10.8 (12%)	21.2 (52%)	8.7 (-55%)	8.0 (-39%)	3.6 (-44%)
STAT2	13.7 (-51%)	15.5 (62%)	36.1 (158%)	17.9 (-8%)	13.8 (6%)	4.5 (-31%)

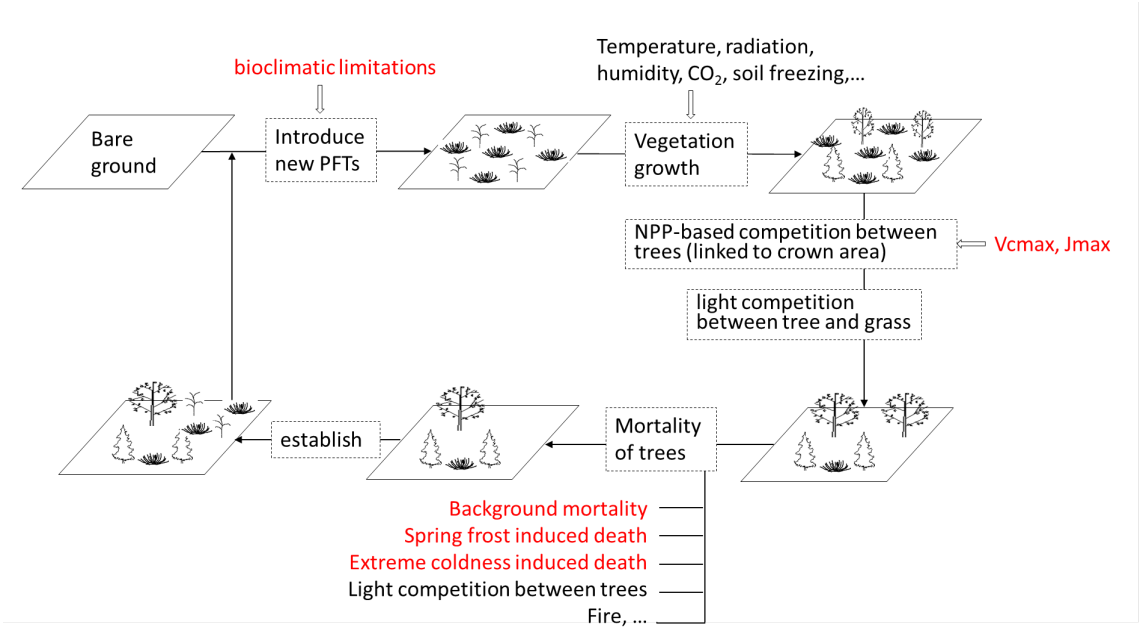


Figure 1. Schematic of ORCHIDEE vegetation dynamics module (ORC-VD). The modifications in this study are marked red.

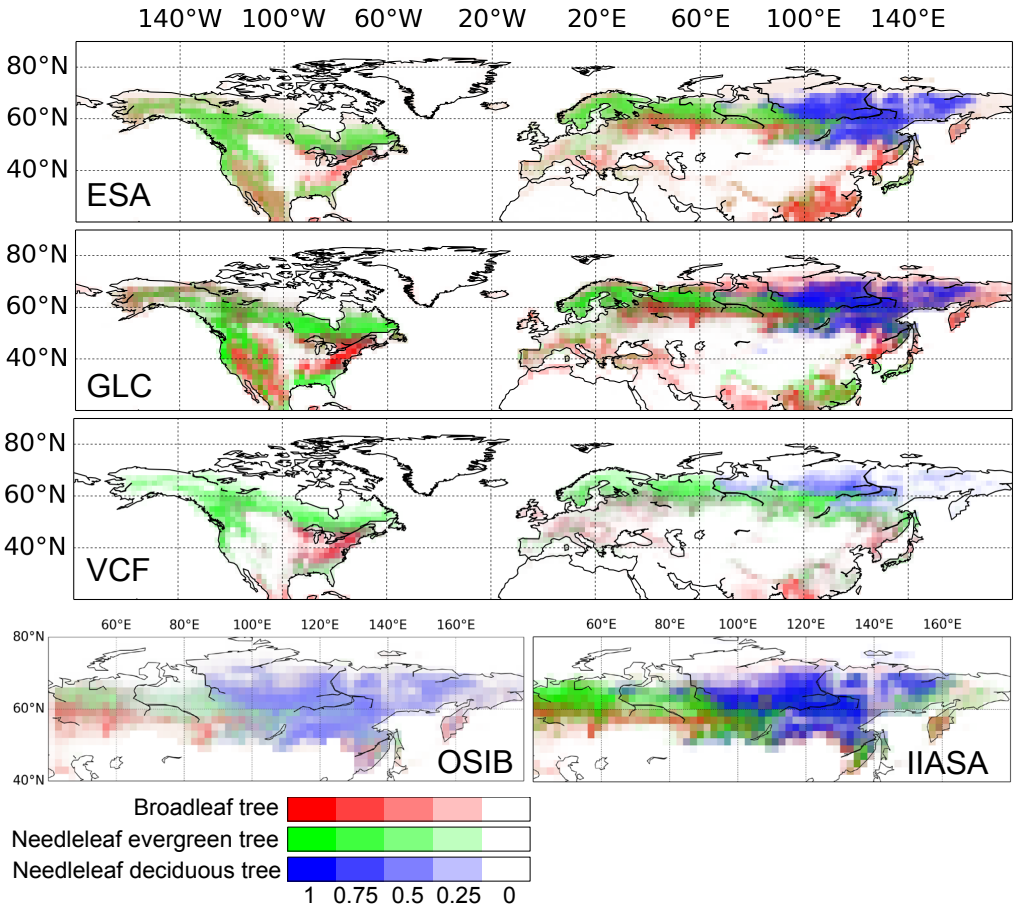


Figure 2. Composite-color map of the fractional vegetation cover in PFT maps converted from the five land-cover products based on Poulter et al. (2011). Color indicates the ~~relative~~ fraction of three PFT groups: broadleaf (including evergreen and deciduous, red), needleleaf evergreen (green), and needleleaf deciduous (blue) trees. Deeper ~~Brighter~~ colors represent higher fractional covers.

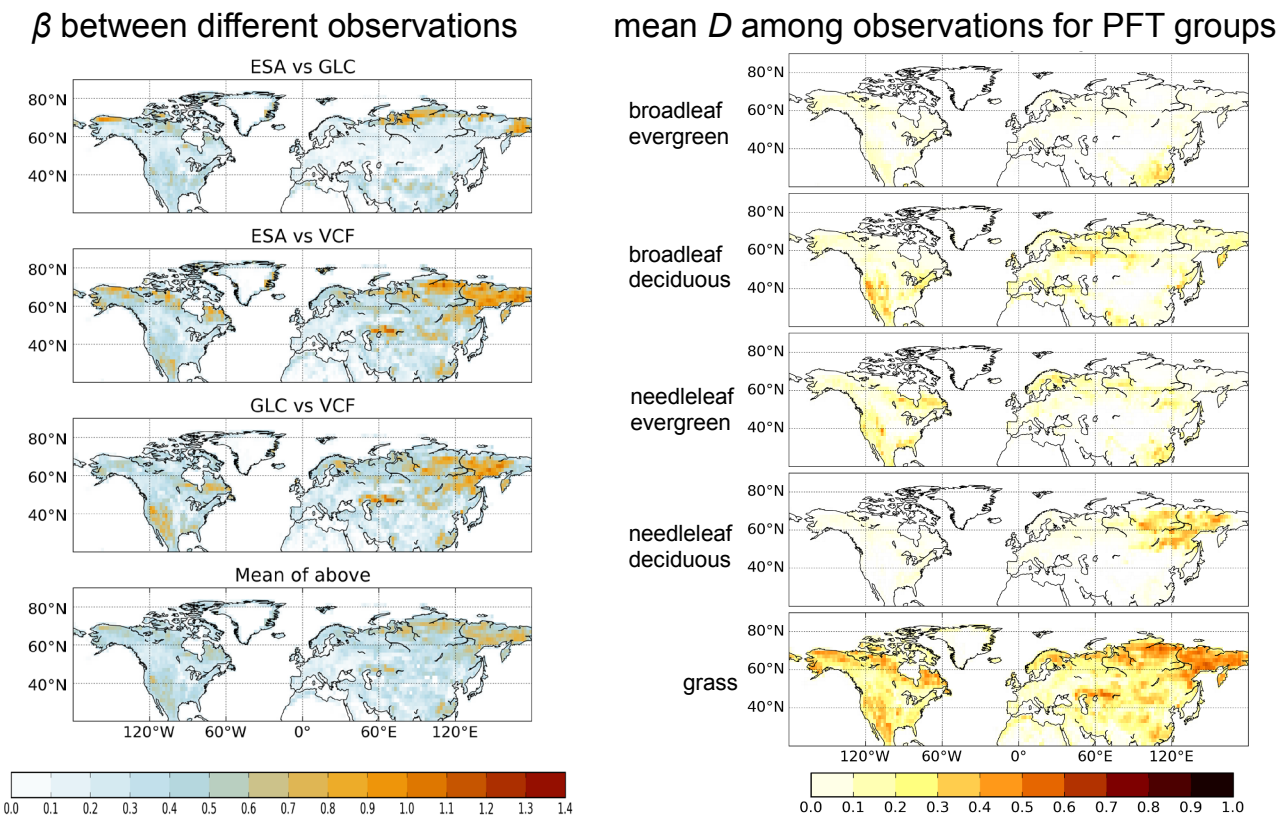


Figure 3. Beta diversity (β) between the three observational datasets (ESA, GLC and VCF) (left panel), and mean dissimilarity index (D) among them for different PFT groups (right panel). β ranges from 0 to $\sqrt{2}$, and D ranges from 0 to 1, both with higher values representing larger disagreement.

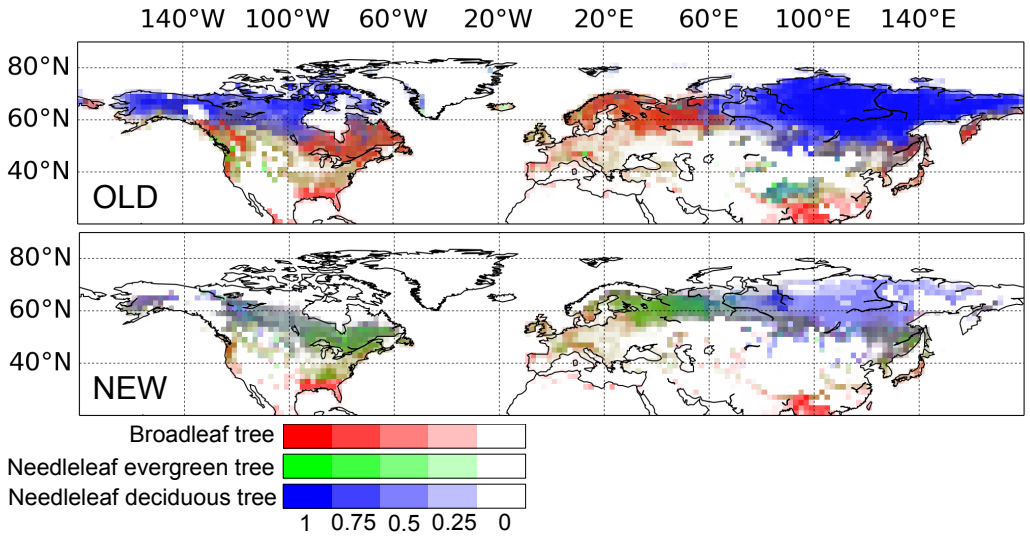


Figure 4. Composite-color map of the fractional vegetation cover in OLD and NEW. Color indicates the ~~relative~~ fraction of three PFT groups: broadleaf (including evergreen and deciduous, red), needleleaf evergreen (green), and needleleaf deciduous (blue) trees. Deeper colors represent higher fractional covers.

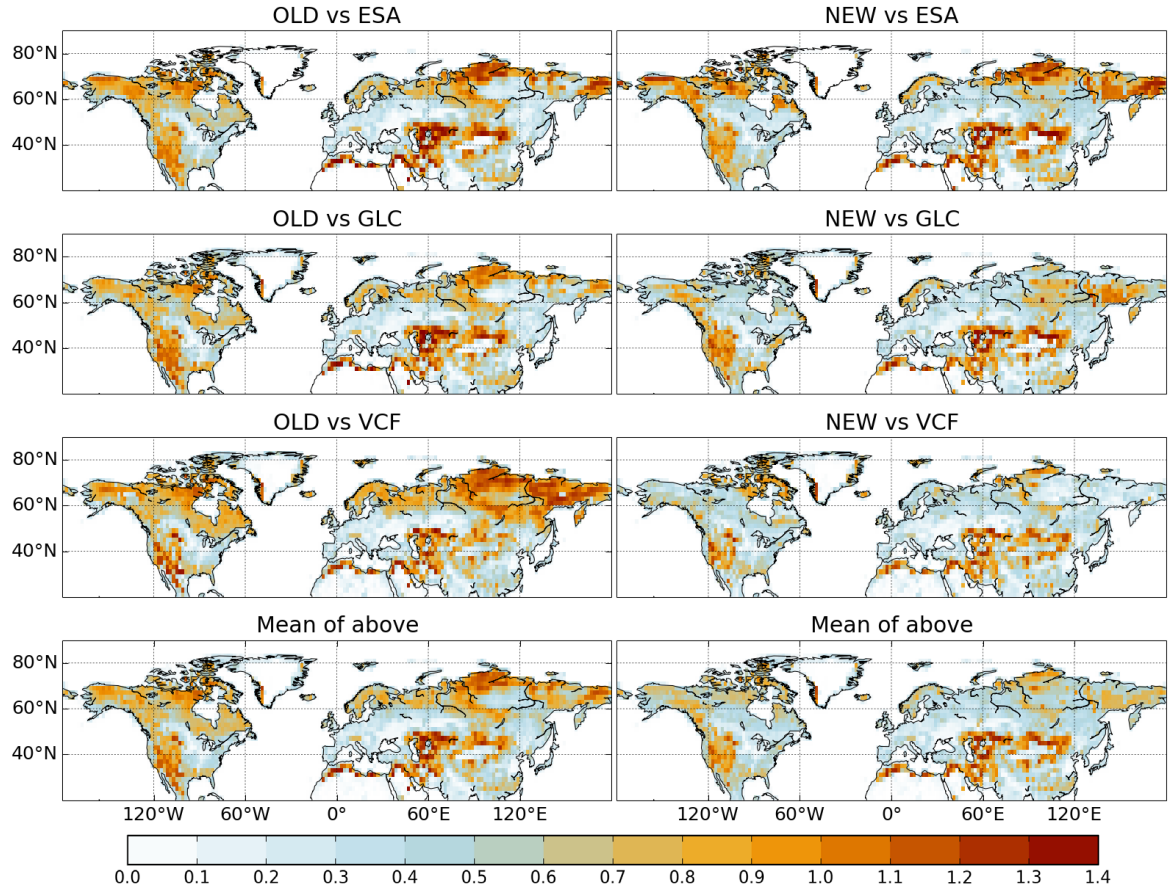


Figure 5. Beta diversity (β) to quantify the disagreement in vegetation distribution between model and observational datasets. β ranges from 0 to $\sqrt{2}$, with higher values representing larger disagreement.

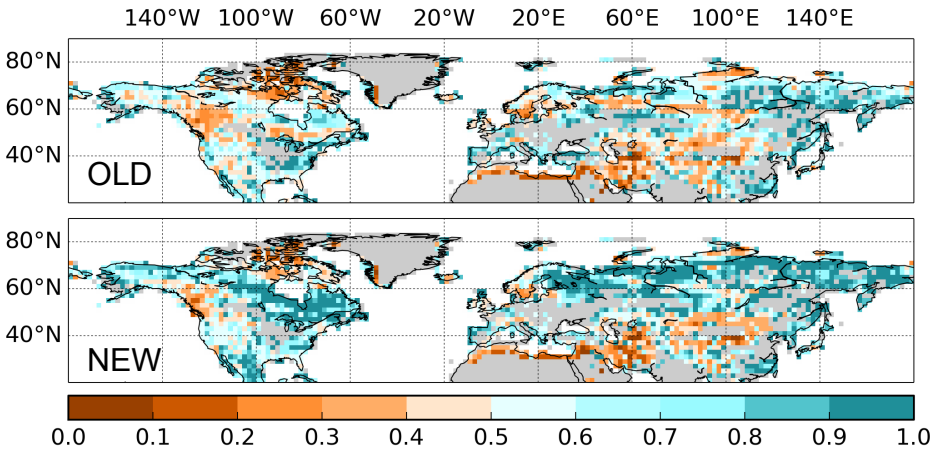


Figure 6. Model skill at simulating vegetation distribution (S_V, Eq. ~~9~~10) for OLD and NEW. S_V ranges from 0 to 1, with higher values representing better model performances, integrating observational uncertainty. Three kinds of grid cells are masked out (in gray): 1) the grid cells where S_V >1 for both models, indicating that the observational data have too large uncertainties to be qualified for model evaluation (13% of the total land points for the studied region); 2) the grid cells where all models and datasets have 100% bare ground in Sahara Desert and Greenland (10%); and 3) the grid cells where crop fraction is higher than 0.5 (8%).

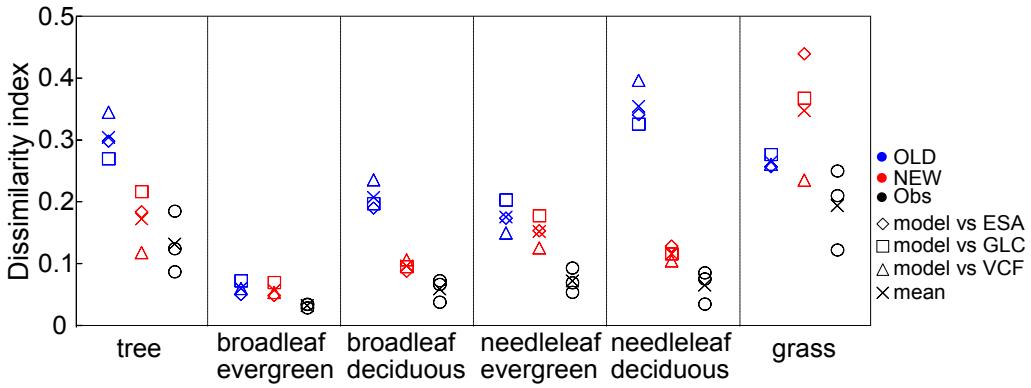


Figure 7. Dissimilarity index (D, Eq. 1~~0~~1) for fractional cover of PFT groups including total tree, grass, and four tree subtypes between model (OLD, blue, and NEW, red) and observations, and between different observations (black), averaged over Northern Hemisphere (20°N-90°N). D ranges from 0 to1, with higher values representing larger disagreement.

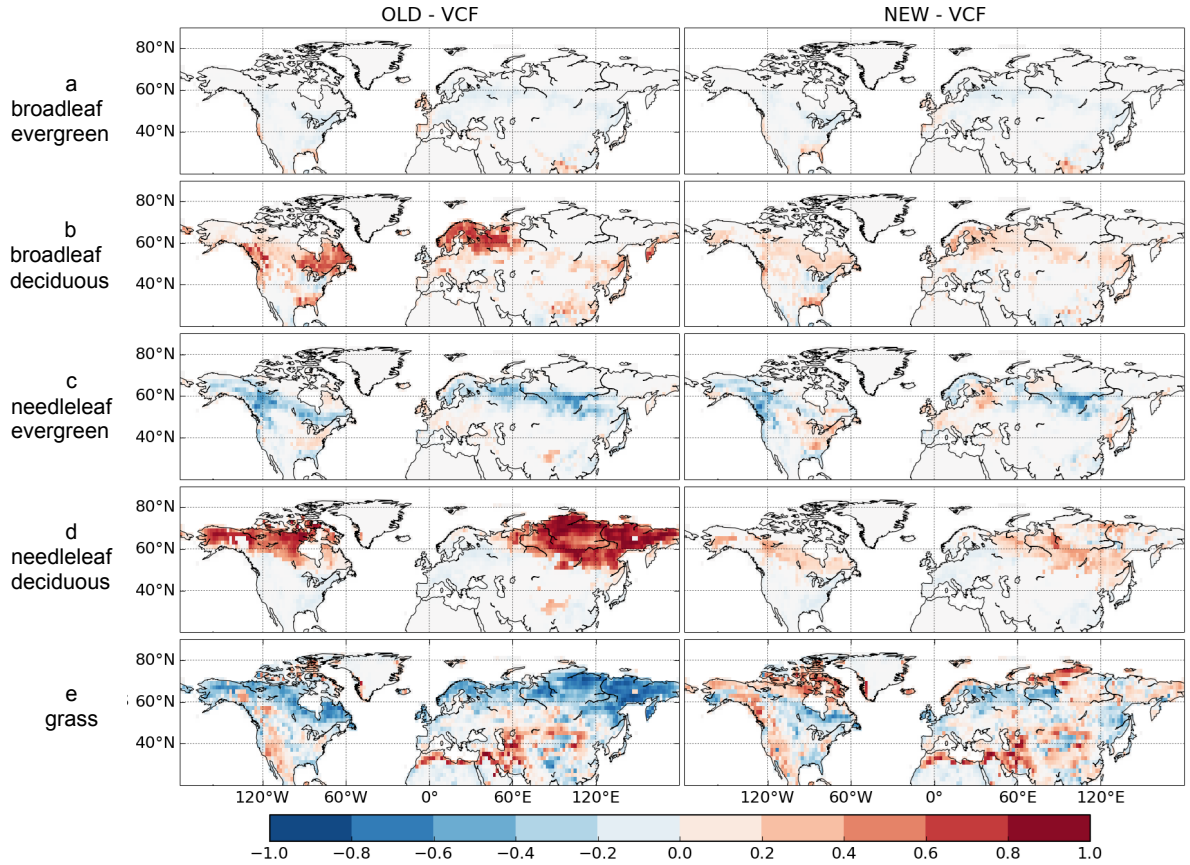


Figure 8. Difference in fractional cover of PFT groups between model (OLD and NEW) and observation-derived PFT map (VCF). Similar map for the difference between model and ESA/GLC is shown in Fig. S4.

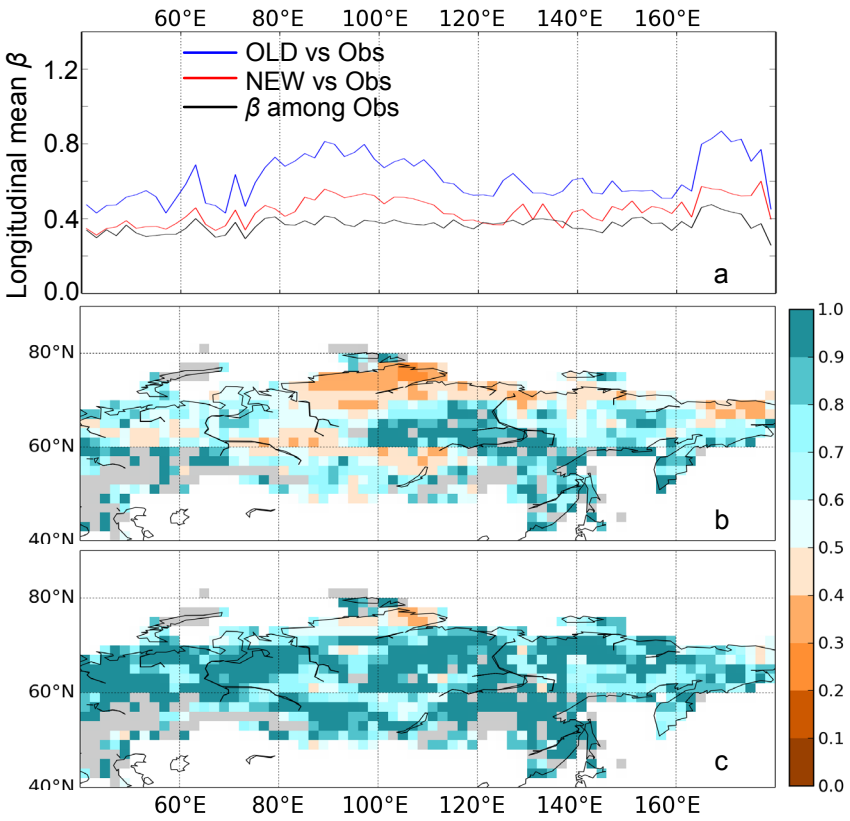


Figure 9. **(a)** Longitudinal average beta diversity (β) between model (OLD, blue and NEW, red) and observational datasets (including ESA, GLC, VCF, OSIB and IIASA) and between different observations (black) in Siberia. β ranges from 0 to $\sqrt{2}$, with higher values representing larger disagreement. **(b)** and **(c)**: Model skill at simulating vegetation distribution (S_V) for OLD and NEW in Sibera. S_V ranges from 0 to 1, with higher values representing better model performances, integrating observational uncertainty. The pixels where S_V >1 for both models, indicating that the observational data have too large uncertainties to be qualified for model evaluation (12% of the total land points in Siberia), were masked out (in gray).

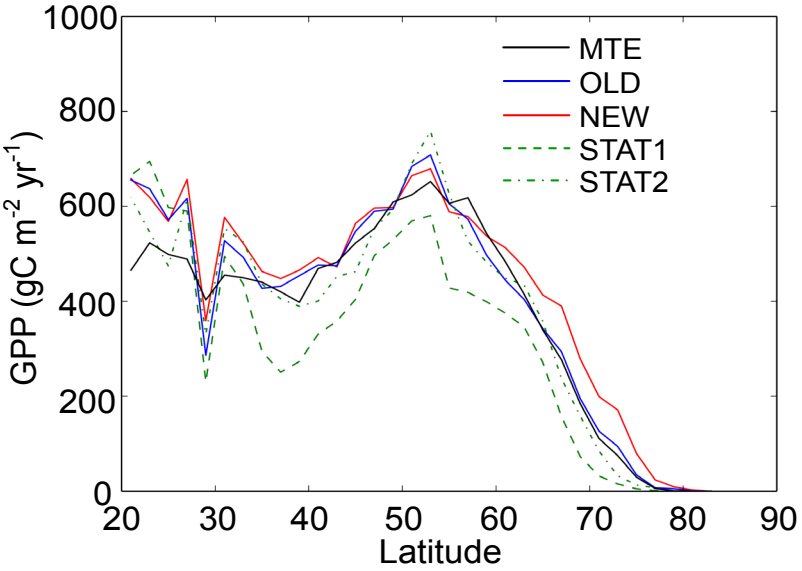
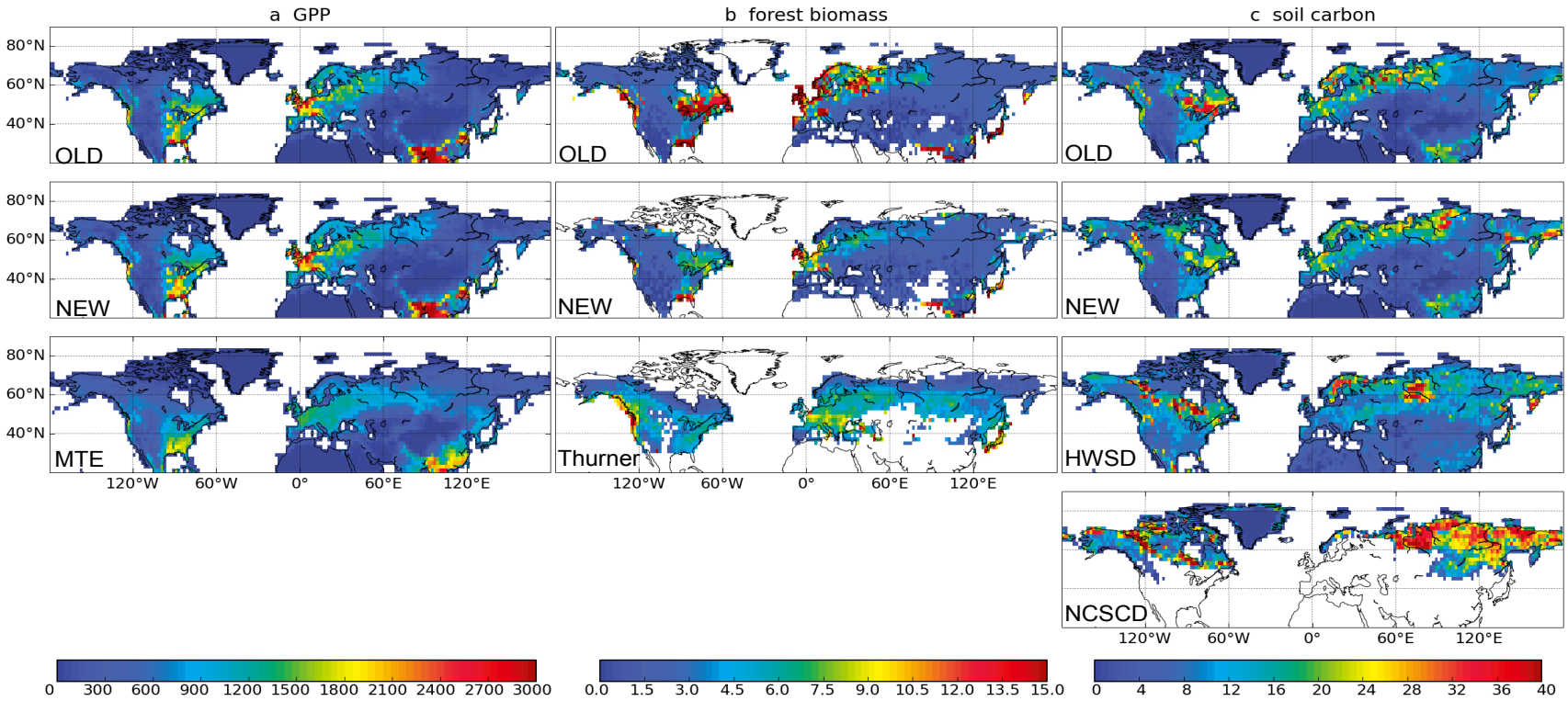


Figure 10. Latitudinal mean annual GPP (2° bands) during 1999~2008 from OLD (blue) and NEW (red), compared with that from STAT (static run in which ORC-VD is deactivated, green dashed lines) and MTE (Jung et al., 2011, black). In STAT1 and STAT2, PFT map is prescribed from ESA and SYNMAP respectively.

1



2

3 Figure 11. Spatial pattern of **(a)** mean annual GPP (g C m^{-2} yr^{-1}) during 1999~2008 from OLD, NEW, and MTE (Jung et al., 2011); **(b)** forest
4 biomass density (per forest area, kg C m^{-2} forest) from OLD, NEW and Thurner et al. (2013); and **(c)** total soil carbon density (kg C m^{-2})
5 simulated by OLD and NEW (0–2 m depth), and from HWSD (0–1 m depth) and NCSCD (0–1m depth).

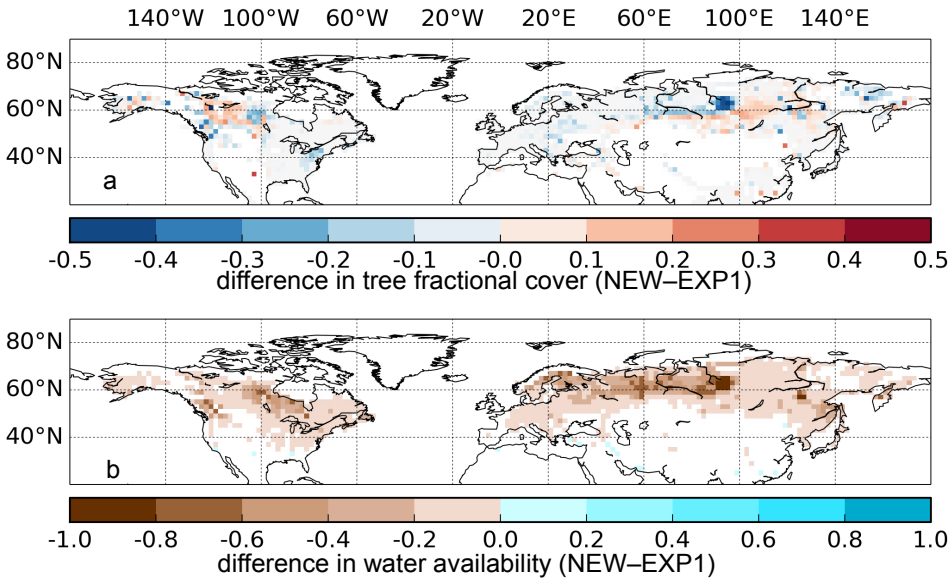


Figure 12. Difference of tree fractional cover **(a)** and water availability (WA, **b**) between with and without soil freezing (NEW–EXP1). WA is averaged over the growing season (May–September) and over tree PFTs (PFT 2–9) weighted by their fractions.

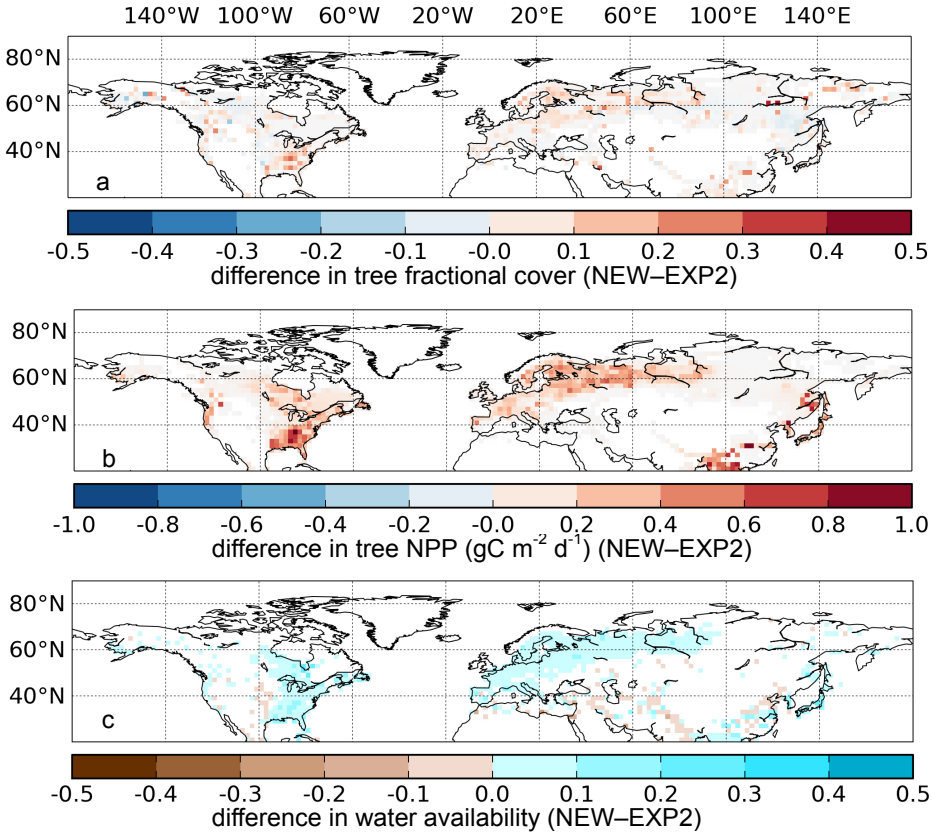


Figure 13. Difference of tree fractional cover **(a)**, tree NPP (g C m^{-2} d~~yr~~$^{-1}$, **b**) and water availability (WA, **c**) between with and without CO_2 rising (NEW–EXP2). NPP is averaged over tree PFTs (PFT 2–9) weighted by their fractions. WA is averaged over the growing season (May–September) and over tree PFTs (PFT 2–9) weighted by their fractions.

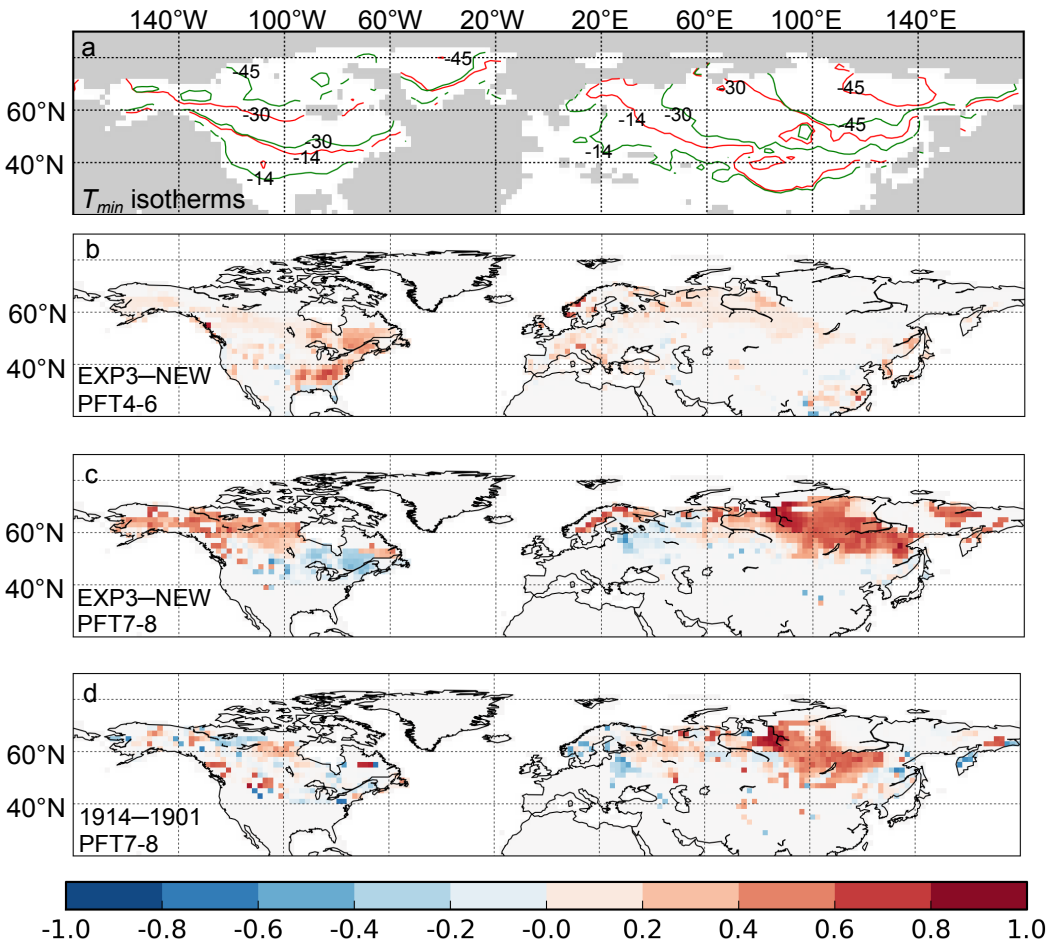


Figure 14. **(a)** minimum temperature (T_{min}) isotherms calculated from the 20-year average climatology (red lines) and the mean of the twenty T_{min} for each year (green lines). The T_{min} values are labeled on the lines, corresponding to the PFT-dependent $T_{min,crit}$ for temperate and boreal trees (see Table1). **(b, c)** difference of ~~the~~ vegetation fractional cover for the last year of spin-up between EXP3 (using 20-year climatology as forcing file in spinup) and NEW for temperate trees (PFT4–6, **b**) and boreal broadleaf deciduous / needleleaf evergreen trees (PFT7–8, **c**). **(d)** difference in fraction of PFT7–8 between spinup results forced by climatic data of two different single years (1914 and 1901).