**Geoscientific
Model Development
Discussions**

Interactive comment on "Evaluation of a Dynamic Global Vegetation Model using time series of satellite vegetation indices" *by* F. Maignan et al.

F. Maignan et al.

maignan@lsce.ipsl.fr

The first of objective of the research, which is to develop a quantitative method for evaluating DGVMs based on satellite data, is both very relevant and timely given the plethora of models and versions of models, and the need to benchmark and compare performances. That being said, however, I think that looking at FPAR by itself would not be a sufficient test (not that the authors are claiming it to be). The bigger significance perhaps would be possibly in adapting the methodology for other parameters of interest, based on specific research objectives, and for other models. In this sense, the paper can potentially have broad impact. The second objective, which is to evaluate versions of ORCHIDEE, is specifically relevant to ORCHIDEE users, but it still informative and useful.

We do agree with these general comments. It is clear that one parameter alone is not sufficient for a full assessment of a model. However, one also needs a reliable "truth" and the FPAR is probably the parameter, while representative of vegetation presence and activity, which is the most reliable as fairly directly related to the measurement.

In general, the authors do a good job of explaining and accounting for their results, particularly with regards to the model deficiencies and unexpected outcomes. However, editing and minor additions may help improve clarity, flow and reproducibility.

Although we had asked the service of a native speaker for this manuscript version, we have applied a new set of editorial corrections based on another one's suggestions.

*Specific Questions:

1. Why have the authors chosen to use a static vegetation distribution rather than a dynamic vegetation distribution (section 2.1, line 17)? As the authors point out in the model description, vegetation responds to climate not just in terms of changes in LAI but also in terms of fractional coverage of their respective grid cells. Using a fixed PFT distribution restricts the model response, and the LAI may act to compensate for the inability to change fractional coverage. Intuitively, one would think that having vegetation respond to climate and competition would be a more "realistic" representation of land surface processes and we would therefore want to test this to correspond to satellite observations. Admittedly, since the objective is to evaluate this particular version of the model using the new technique described and since many modelers do use fixed vegetation anyway, the choice of static vs. dynamic vegetation may not be a crucial issue. However, the text may benefit from a sentence or two just clarifying the rationale behind this choice.

We added the following sentences and references: "As we are evaluating the model over a short recent period, we chose to use a fixed vegetation distribution, derived from recent observations, rather than a dynamic one. The state of the art models of vegetation dynamics are still rather crude and lead to biases in simulated vegetation types as compared to the observations (Krinner et al., 2005; Bonan and Levis, 2006)."

Bonan, G. B., and Levis, S.: Evaluating aspects of the community land and atmosphere models (clm3 and cam3) using a dynamic global vegetation model, Journal of Climate, 19, 2290-2301, 2006.

2. The authors cite Hansen et al. (2000) and Heymann et al. (1993) as the basis for deriving the fixed PFT distribution map used in the simulations. It would be more helpful to actually present a figure showing the PFT distribution in terms of fractional coverage, especially since these references are dated earlier than the MODIS data from which NDVI is derived for the evaluation.

As suggested by the reviewer we added such a figure, showing the fractional coverage for each PFT, in a new Supplementary Material section.

Over what time period was the input PFT distribution determined? If the NDVI time series covers years 2000-2008, and the vegetation maps used as input to the model are based on earlier references, how sure are we that the PFT distribution used in ORCHIDEE reasonably matches actual PFT distributions over the period of interest for the research? Couldn't this potentially

affect the correlations, depending on land cover changes between the period over which the input vegetation maps were determined and during 2000-2008? The authors already state in Section 5 that there were no significant trends in satellite data over 2000-2008 so maybe just a comment on the input PFT distribution and how well it represents 2000-2008 coverage is required.

Indeed the land cover maps used to generate the ERA-I PFT distribution are derived from observations acquired between 1995 and 2000. However, global land cover maps are not so numerous and exhibit around 70% accuracy (Bontemps et al., 2011). Deriving a PFT distribution from land cover maps introduces again some uncertainties, especially for C4 and tropical PFTs (Poulter et al., 2011). Hence over such a short comparison period, we assume that, at global scale, the land cover change lays within the uncertainty of the PFT distribution map. For example, Hansen et al. (2010) estimate the global gross forest cover loss at 0.6% per year. We also argue that it's a second order improvement, not noticeable as long as the phenology modeling is not improved, and has been noted as such in Jung et al. (2007b).

Bontemps, S. , Herold, M., Kooistra, L., van Groenestijn, A. , Hartley, A. , Arino, O. , Moreau, I., and Defourny, P.: Revisiting land cover observations to address the needs of the climate modelling community, Biogeosciences Discuss., 8, 7713–7740, 2011, doi:10.5194/bgd-8-7713-2011.

Hansen, M. C., Stehman, S. V., and Potapov, P. V.: Quantification of global gross forest cover loss, Proceedings of the National Academy of Sciences of the United States of America, 107, 8650-8655, DOI 10.1073/pnas.0912668107, 2010.

none**Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S. & Zimmerman, N. E. (2011). Plant functional type mapping for earth system models. Geosci. Model Dev. Discuss., 4, 2081–2121, 2011, doi:10.5194/gmdd-4-2081-2011.**

3. In the Discussions section, line 17, the authors state that they "observe no significant trends over the nine year-period of the study, neither on the satellite data nor on the modeled FPAR." What about the meteorological forcing data? If there was no significant trend in the meteorological forcing data that would cause changes in PFT fractional cover in a version with dynamic vegetation, then this would be additional justification for just using static vegetation.

By "non significant", we mean that any potential trend is small in regard to the interannual variability and cannot be identified on the data because of the rather short period. We make use of nine years of data which is clearly insufficient as the trend (difference between the "typical" value of two successive years), if any, is small in comparison to the year-to-year variability.

4. Not being an ORCHIDEE user myself, I am just curious as to why NEP is used as an indicator of steady-state equilibrium (Section 2.1.2, line 25 onwards) rather than Net

Ecosystem Exchange (=-(NEP-fire flux)) as is done for LPJ, Community Land Model

DGVM, CNDV. Does ORCHIDEE not calculate fire flux? In addition, what forcing is used during the spin-up to reduce the decadal mean value of NEP?

The debate about taking the fires into account or not during the spinup phase is still open among ORCHIDEE users. The standard configuration has the fire module active. This indeed implies to have some interannual climatic variability during the spinup phase. The steady state equilibrium is thus reached by cycling over the available ERA-Interim years, for around 1300 years.

The following references mention using fires during the spinup phase:

Krinner et al., 2005

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, Global Change Biology, 9, 161-185, 2003.

Thornton, P. E., and Rosenbloom, N. A.: Ecosystem model spin-up: Estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model, Ecological Modelling, 189, 25-48, DOI 10.1016/j.ecolmodel.2005.04.008, 2005.

5. Could the authors speak a little more about the significance and purpose of comparing the results of two different meteorological forcings? I am not sure that the full added value of this section is clear to me – it just seems that modelers can use a variety of forcings depending on their purpose. For example, I might use meteorological forcings describing a paleoclimate or future SRES or RCP. So why make a comparison between CRU-NCEP and ERA-Interim forcings, in particular? Aside from the impact of spatial resolution, how can the findings from this comparison be useful for researchers utilizing different meteorological input? Is the purpose to test the sensitivity or stability of the model between different forcings with the new phenology scheme?

For people doing simulations over the last decades, studying for example carbon fluxes and budgets, having the best possible meteorological forcings is of crucial importance: a specific process may be perfectly modeled, if the meteorological input is wrong, the outputs may not be of the expected quality. The meteorological driver is pointed by Jung et al. (2007b) as the second major source of uncertainty (after the model itself). They find a relative difference of 20% for GPP over Europe between two different forcings. Zhao et al. (2006) report similar differences on MODIS global GPP and conclude that ECMWF has the most accurate temperature and VPD.

Zhao, M., Running, S. W., and Nemani, R. R.: Sensitivity of moderate resolution imaging spectroradiometer (modis) terrestrial primary production to the accuracy of meteorological reanalyses, Journal of Geophysical Research-Biogeosciences, 111, G01002, doi 10.1029/2004jg000004, 2006.

6. In Section 4.2, Evaluation of meteorological forcings, the paper states in line 13 "as the simulations have two different vegetation maps. . ." This part slightly is confusing to me. Why do ERA-I and CRU-NCEP have different vegetation maps – is it because of interpolation to their respective spatial resolution despite being based on the same PFT distribution? The authors cite Jung (2007) as justification for having different vegetation maps but perhaps in supplemental material, these maps can be presented with some discussion of their differences.

We added a new paragraph "2.1.3 PFT spatial distribution" to warn that the global simulations were done independently within different contexts, using different PFT maps. We agree with the reviewer that this is not an "ideal" situation, but one that was made necessary due to disk space and CPU time restrictions. We added the images of the two fractional coverages in the supplementary material.

7. In the section "Scoring of PFT" line 7, the paper says "we have used a high-resolution vegetation map to identify the dominant PFT. ." Which map is being used here – is it the same one as was used to develop the PFT distribution? Perhaps a reference is needed here, and maybe just something to state or show that this map reasonably corresponds with actual vegetation distribution from 2000-2008.

In fact we have a tool that calculates PFT fractional coverage at any desired spatial resolution, based on the same high resolution entry maps (CORINE and UMd). We thus modified the text as follows: "We have used the same high-resolution vegetation maps that were used to derive the PFT distribution for the ERA-Interim simulations, to identify the dominant PFT at the resolution of the satellite dataset (5 km)."

8. One rationale for doing this study as stated in the Introduction is because DGVMs are used to address the question of the impact of increased CO_2. However, the rest of the manuscript makes no mention of CO_2. Is it kept at a fixed value? At what ppmv? Can the correlation scores be partially explained by any effect of increasing CO_2 over the time period 2000-2008 not captured in the model but reflected in satellite data?

The increase of the CO_2 atmospheric concentration is taken into account through a mean annual global value. As we mentioned earlier we see no trend in the data or in the model over this rather short period, so that the increase of the CO_2 atmospheric concentration cannot explain the correlation.

*Technical Comments:

1. I appreciate that the authors try to make the methodology clearer by illustrating it as they do in Section 3, but in some parts I find that it detracts from the flow of the paper. For example, section 3.3 already presents important results from the ERA-Interim simulation regarding areas where the correlations are high and where they are low, and discusses the deficiencies in the model to account for the latter. However, these may better be presented and discussed more systematically in Sections 4 and 5, including also the correlation map for the other meteorological forcing. The authors might consider restructuring the paper slightly to improve organization and flow.

We followed the reviewer's suggestion. We have more clearly divided the method and result sections so that the figures are now discussed outside of the methodology section.

2. Figure 1 label reads "a mean annual cycle for both the NDVI and FPAR signals is derived (middle right plot)" but you are referring to the plot on the left of the middle panel right?

The reviewer is right, we changed the text to "middle left plot".

Interactive comment on Geosci. Model Dev. Discuss., 4, 907, 2011.