Biogeosciences
Discussions

Interactive comment on "Assessment of model estimates of land–atmosphere CO_2 exchange across Northern Eurasia" *by* M. A. Rawlins et al.

A. J. Dolman (Referee)

han.dolman@vu.nl

This is a useful and timely paper, in general well written that compares the output of 9 different models for GPP, Respiration and NEP over a large regions in Norther Eurasia. The model output is benchmarked against eddy covariance data from 4 sites and against a satellite remote sensing product of GPP (MOD17). While the general line is good, the paper misses several opportunities to present a clearer analysis and hence at times the paper ends in a somewhat lacklustre description of numbers, rather than that it tries to identify which process description in models produces what behaviour. As an example in the discussion the impact of nitrogen is mentioned, but nowhere in the paper is an attempt made to use the fact that two of the 9 models incorporate a nitrogen cycle. The authors also conclude that the model's treatment of respiration needs to be

improved, similar to previous authors (e.g. Dolman et al., 2012), but again do not use differences between the models to shed more light on how they see this improvement. I would therefore suggest that the authors consider the differences between the models (Table 2) more in their explanation of the results.

p 2260 line 6-12. This is an complicated sentence to read, certainly for the introduction. Please reformulate. The reference to Cox et al., 2000, is a bit strange. This paper, while pointing to the importance of a carbon-climate feedback, mainly identifies the tropics as the key region.

p 2260 line 15-17. It is important to stress that, yes there may be increase in GPP at the beginning of the season, but also an extended period for respiration at the autumn (e.g. Parmentier et al., 2011 show that this may lead to no change in the net flux).

page 2060. It is important to nam the different period for which the numbers quoted in this section are obtained. This is relevant to the discussion of trends later on in the paper.

page 2262 line 6 efficacy means according to Wikipedia the capacity for beneficial change (or therapeutic effect) of a given intervention. I am sure the authors mean efficiency. That being said, the error analysis is a bit poor, basically the error is defined as modelled minus means. With these comparisons I would suggest that more advanced metrics, such as Nash-Sutcliffe efficiencies or other allow a clear interpretation of the results, give more depth to the analysis and might also help with meeting my earlier comment on the character of the analysis).

Page 2263. It would help if the authors could say something about the modelling protocol that was followed: what was the spinup procedure, which forcing data (or model) sets were used etc. In fact, the big question here is do all model start from the same initial condition in 1960, or are there already big differences after spinup. A big question is also why some models are apparently run as DGVM, while it is also possible to fix the land use. Some of the differences may thus be due to different land

use types, rather than process description.

Page 2264. Is the Zotino site used, the fir, or pine forest. The fir is rather a strong sink (see Dolman et al., 2012, figure 4. I am also rather surprised that there are only four sites used. There is more data available. Why was that not used?

Page 2065. It would help to cite the references agains which the MOD17 products appears to have been "extensively evaluated"

Page 2266. Line 4. assessment, not assessments.

Page 2066. While accepting whether this is to some extent subjective, I do not quite agree with the description of the model at match. Please be as objective as possible. A model like ISBA is clearly way off of properly simulation both GPP and R, and only comes back at NEP because they cancel by and large. Also the timing is not always consistent, some models are a month of. That is in a growing season of 2-2.5 month a substantial mistake, even if the use of monthly means would exaggerate such a mismatch. See also my earlier remark about using more objective metrics.

Page 2267. line 7-10. I do understand what you are trying to say here. Please reformulate.

Page 2267 line 25-29. A score of less than 50% would suggest you might as well use climate (T,P) and empirical relation between these two variables and GPP or even a random number generator. Please realise these results are really bad, and question the use of these models. So try to reformulate this and emphasise that the models are doing an extremely poor job here.

Page 2268 line 10-12. Most models have hardwired a relation between GPP and R, so this high correlation does not say much.

2270. While I appreciate the discussion on residence times, I do find it questionable to calculate this, given the discrepancies in GPP, R, let alone soil carbon stocks (see also remark on spin up). I wonder whether this should not be tuned down a little and

mentioned that as a results of wrong fluxes, a very large variability in residence time is obtained, with opposing trends, and signs.

Discussion. 1) I really miss a remark on thermokarst, changing hydrology or other cryogenic processes This is crucially important in this part of Eurasia. 2) A discussion where use is made of the differences in model processes and parametrizations would bring more depth to the analysis, han the current line which basically states that respiration, soil dynamics and productivity need to be improved. These three processes are of course the core purpose of the models, so either one concludes that they fail for that purpose, or one should make an effort to go a little bit more in depth.

Parmentier, F. J. W., van der Molen, M. K., van Huissteden, J., Karsanaev, S. J., Kononov, A. A., Suzdalov, D. A., Maximov, T. C., and Dolman, A. J.: Longer growing seasons do not increase net carbon uptake in the northeastern Siberian tundra, J. Geo- phys. Res., 116, G04013, doi:10.1029/2011JG001653, 2011.