Geosci. Model Dev. Discuss., 6, C3068–C3072, 2014
www.geosci-model-dev-discuss.net/6/C3068/2014/

Geoscientific Model Development Discussions
Open Access

GMDD
6, C3068–C3072, 2014

Interactive Comment

Interactive comment on "Influences of calibration data length and data period on model parameterization and quantification of terrestrial ecosystem carbon dynamics" *by* Q. Zhu and Q. Zhuang

Q. Zhu and Q. Zhuang

zhuq@purdue.edu

Received and published: 12 June 2014

Comment#1.Methods for identifying critical time periods in the observations, which contain most of the information for parameter identification, can provide useful guidance when selecting the evaluation data series. However this issue is not new and has been explored in some detail in previous studies including recent work (Bárdossy and Singh, 2008; Bennett et al., 2013; Singh and Bárdossy, 2012). And there are recent and more quantitative treatments of this problem in the recent literature that the authors missed in their review, and that should serve as a starting point for this work (Bárdossy and Singh, 2008; Bennett et al., 2013; Singh and Bárdossy, 2012). Without this, in the opinion of this reviewer the novelty (scientific significance) of the work in the context of general model development is in question.

Response: In the revised manuscript, we have added those missed references regarding the recent advances in the related topic. However, we noticed that all these studies focused on hydrology modeling. To our knowledge, there is no related study in terrestrial ecosystem and biogeochemistry modeling. We believe our study is an important step forward in terrestrial ecosystem and biogeochemistry modeling field. The measurements of ecosystem scale carbon fluxes are relatively short compared with hydrology sciences (e.g., precipitation data), which makes the model calibration and evaluation difficult. One of the central questions is that which subset of observed carbon flux should be used to constrain our models. Thus, exploring the sensitivity of model parameterization to calibration data length and data period becomes extremely important for terrestrial ecosystem modeling community. We believe this study is among the very first efforts towards such topic and will be beneficial for our modeling community.

Comment#2 The paper aptly illustrated these expected results and their limited site-specific value. However, it does not seem to significantly advance the core issue of "when is enough enough?"

Response: As is mentioned by the reviewer, our conclusions are limited at specific sites. And our findings are also conditioned on the specific model we used. However, it will be impossible to draw a universal conclusion for all ecosystem models or for all ecosystem types at various sites in a single study. We find that two-year data is enough for calibration of TEM's carbon dynamics at our study sites. Although our finding may not necessarily be true for other ecosystem models, we demonstrate that there exists a threshold for model calibration data length. Beyond the threshold, more calibration data may only marginally improve the model performance. Further studies of this threshold for different ecosystem models are still needed to address the question of "how long ecosystem carbon flux data is sufficient to constrain an ecosystem model".

Comment #3.I was hoping to find a more general analysis of observed variance thresholds that could serve as the basis to identify the data length needed for a specific site (e.g. the variance contained in each period length for a specific site would be compared to the required variance threshold to identify the minimum length required for each site).

Response: It is challenging to provide a threshold of climate variability that can help the ecosystem modeling community to identify the minimum calibration data length. That is because the intrinsic variability of climate variables changes from site to site even though the vegetation types are the same. As is pointed out in Introduction, our objective is to test "if calibrations using the data with high climate variability are superior to the calibrations using data with low climate variability" rather than to identify a threshold of climate variability. And through calibration experiments, we successfully show that model calibration could benefit from actively selecting data with relative high climate variability. Our study didn't aim to provide a threshold of "climate variability", which could be used to identify minimum calibration data length. The key message we trying to convey is that "climate variability" is an effective indicator of information content within the data. To obtain such a threshold, further researches are needed.

Comment#4. Another important issue of this study is the dissimilarity in length of experimental records available for the different ecosystem monitoring sites analyzed. These differences will likely hinder the representativity of the results and the comparisons across these ecosystem types. Again, the observed variance captured at each site can grossly vary with the length of the monitoring record, and in turn this will depend on the intrinsic characteristics of each ecosystem type (some ecosystems exhibit larger variability in canopy and productivity than others). If so the comparisons presented among the ecosystem types are questionable, and possibly misleading.

Response: The reviewer concerns that the differences in total length of observations at the five selected sites may greatly affect the results. In order to address this data dissimilarity issue, we selected AmeriFlux sites that have the longest published mea-

GMDD
6, C3068–C3072, 2014

surements for each vegetation type in this revision. So that, the differences in total length of observation among different sites are largely reduced. We: (1) replaced the Lost Creek shrub land site (5 years data) with Kennedy Space Center Scrub site (7 years data); (2) replaced the UCI boreal forest site (4 years data) with Wind River Field Station site (8 years data). In the revised manuscript, we re-did the calibration experiments for shrubland and boreal forest. The results at the new sites are consistent with previous findings at the old sites. We agree that the "absolute climate variance captured at each site can grossly vary with the length of the monitoring record, and in turn this will depend on the intrinsic characteristics of each ecosystem type (some ecosystems exhibit larger variability in canopy and productivity than others)". However, we did not use the "absolute climate variance". Instead, we use "relative climate variance". For example, throughout the manuscript, we grouped the calibration experiments into two categories: (1) above mean climate variability (2) below mean climate variability; and compared the model performances. We show that model calibration using the data with relative higher climate variability is better than that using relative lower climate variability.

Comment #5. I also find the selection of the RMSE as a calibration performance statistic of limited value without the use of a complementary relative value statistic. While RMSE provides a magnitude based error between the observed and predicted quantity, its minimization it does not ensure that the end point of the calibration is satisfactory. Many studies suggest that this magnitude-based indicator (RMSE) be complemented by a relative statistic that compares the model error to the variance of the observed data, or against a desirable/accepted benchmark (see for example recent discussions of Bennett et al., 2013 ; Ritter and Muñoz-Carpena, 2014).

Response: Thanks for providing the references. All evaluation methods may suffer certain biases. For example, the RMSE will underestimate the bias if the predicted carbon flux and observed carbon flux are small (during non-growing season). Following the reviewer's suggestion, in this revision, we selected two other metrics as complementary criteria to assess the model performances: (1) Mean Absolute Percentage Error (MAPE); (2) Nash-Sutcliffe efficiency coefficient (NSE or R2).The MAPE evenly weights small and large predicted carbon fluxes, therefore, provides us a more appropriate metric (compared with RMSE) to assess the model performance during non-growing season. And the NSE tells how well the calibrated model reproduced the variation (seasonality) of observations. In the revised manuscript, we added two tables (Table 3 and Table 4) to summarize the model performance of calibration experiments in terms of MAPE and NSE.

Minor comments. Table 2, figures and text. For consistency refer to the sites by the ecosystem type as in Fig. 6-7, rather than by site type (rest of figs. and parts of text). Table 2 contains the correspondence between them. Response: According to reviewer's suggestions, we have updated the figures 1-5 accordingly. We have also carefully checked the entire manuscript for spelling, grammar issues, and sentence structure.

References: Bennett, N.D., B.F.W. Croke, G. Guariso, J.H.A. Guillaume, S.H. Hamilton, A.J. Jakeman, S. Marsili-Libelli, L.T.H. Newham, J.P. Norton, C. Perrin, S.A. Pierce, B. Robson,R. Seppelt, A.A. Voinov, B.D. Fath. 2013. Characterising performance of environmental models. Environ. Modell. Soft., 40:1–20 Bárdossy, A., Singh, S.K., 2008. Robust estimation of hydrological model parameters.Hydrol. Earth Syst. Sci. 12, 1273–1283. Ritter, A. and R. Muñoz-Carpena. 2013. Predictive ability of hydrological models: objective assessment of goodness-of-fit with statistical significance. J. of Hydrology 480(1):33-45. doi:10.1016/j.jhydrol.2012.12.004 Singh, S.K., Bárdossy, A., 2012. Calibration of hydrological models on hydrologicallyunusual events. Adv. Water Resour. 38, 81–91

Interactive comment on Geosci. Model Dev. Discuss., 6, 6835, 2013.

GMDD

6, C3068–C3072, 2014

Interactive Comment