Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-76-AC2, 2018

Geoscientific Model Development Discussions

Interactive comment on "Realized ecological forecast through interactive Ecological Platform for Assimilating Data into model (EcoPAD)" *by* Yuanyuan Huang et al.

Yuanyuan Huang et al.

yyhuang@ufl.edu

Received and published: 18 September 2018

Anonymous Referee 1 RC1: The manuscript "Realized ecological forecast through interactive Ecological Platform for Assimilating Data into model (EcoPAD)" by Y. Huang et al. presents the development of a web-based software system for quantitative ecological forecasting. The system is based on the availability of observational data, a process-oriented model, an algorithm for assimilating the observations into the model and a web-based workflow. Furthermore the paper describes the application of EcoPAD to the Spruce and Peatland Responses Under Climatic and Environmental change (SPRUCE) experiment in North Minnesota using the Terrestrial ECOsystem (TECO)

model and a Markov Chain Monte Carlo assimilation technique in forecasting carbon fluxes and pools. The manuscript is mostly well written, however, at times (sections 1 and 2) it reads more like a 'sales pitch' for EcoPAD with quite a few repetitive elements (e.g. the list of elements included in the workflow appears at multiple places) and at other times (section 3) it reads as a review on the previous applications of EcoPAD. So in essence, my major concern is that there is little new science in the current version of the manuscript except for the technical engineering of the web-based software system, which in itself is not described in great detail. My recommendation is to focus the manuscript on these technical developments and provide a more in-depth description of the technical details of this system, however, I am not sure if this then still fits to GMD because the web-based software system development is very much focussed on information technology developments.

Response: We appreciate the Reviewer's valuable suggestions. The manuscript is organized to express the motivation of building EcoPAD, or why we need a platform like EcoPAD in ecological forecasting (Introduction, section 1), the technical support (section 2) and what we can benefit from EcoPAD (or its application and scientific values, sections 2, 3). The technical engineering is an important part of EcoPAD and the manuscript. The functionality of EcoPAD or the role of EcoPAD in advancing ecological forcasting is built upon the technical elements. But the manuscript is not only about technical details. Equally important is what we can benefit from such a platform for ecological forecasting. And the goal of the technical advances is to improve ecological forecasting. We emphasize that iterative interactions between model and data, as well as between modellers and experimenters, are valuable for ecological forecasting.

We do not agree with the Reviewer that there is little new science in the manuscript. The near real time ecological forecasting itself is a new scientific advance in ecology. In addition, we integrated different case studies to illustrate how different components contribute to improve ecological forecasting. Case 3 and Case 4 comes from previous studies. Case 3 is about uncertainty and Case 4 is related to biophysical estimation.

Cases 1, 2, and 5 are new case studies from this manuscript. Case 1 focuses on the communication between modellers and experimenters. We believe that good ecological forecasting is built upon efforts from both modellers and experimenters. Even though this case is not direct technical advance, the techniques embedded in EcoPAD allow near- and real-time interactions between modellers and experimenters. This itself represents an important advance for scientific research that is enabled by modelling. Case 2 is related to acclimation or shift in parameter values. Case 5 compares realised vs. unrealised forecasting. The focus of this study is ecological forecasting. The practice of ecological forecasting is still at its early stage and good forecasting needs to integrate resources from different aspects. Each case study provides valuable information from different perspectives. But none of these cases alone guarantees good ecological forecasting. We keep Case 3 and Case 4 as they reflect important aspects, i.e. uncertainty and boundary conditions, that lead to good ecological forecasting. We have a section discuss the implications of these case studies for better ecological forecasting (section 4.2). And please also refer to our responses to Reviewer 2.

RC1: Another concern relates to the use of the tool by the 'general public' or even experimentalists lacking the background knowledge on data assimilation as promoted by the authors of the manuscript. The concern is that with such a level of automation (essentially only clicking a button on a webpage to get the results of a complex data assimilation experiments) of a very complex system involving experts' concepts from multiple disciplines the user could easily lose the connection to the underlying tools, such as the capability of the ecological model and the data assimilation algorithm. Both components may not be fit for the user's purpose, so a misuse (even and especially unconsciously) of the system can easily happen without the user being able to notice because the user is not an expert of either the ecological model nor the data assimilation algorithm. An erroneous result (which can easily happen if e.g. some observations used in the assimilation are outliers or the assimilation algorithm produces parameter values outside of physical meaningful values etc) of such an automated system could be taken as real and thus be misused. In that sense there should be some caution in

promoting this system to non-specialist users.

Response: We agree with the Reviewer that there are risks of misuses. Tool itself does not necessary equal to misuse. It depends on the people who use it and how it is used. Misuse is not unique to web-based simulation and can also occur to non-web-based model simulation and data assimilation. For example, sometimes people who run complex process-based models, such as these embedded in big Earth system models, may not necessary know how different components of the model work. Or an experienced modeler of carbon cycling may not know much about how hydrology in the model works. In these situations, there are also risks of misuse. This is why we emphasize effective communication between different experts. Experimenters may not know the technical details of how to build a model or how to code the data assimilation algorism, but it is not to say they do not need to know how the system works. The communication between modelers and experimenters help the experimenters to understand what works in the background, what is the meaning of a parameter or process, what they can, or cannot do with the platform. The platform is carefully designed to avoid potential errors. For example, the experimenter is asked to prescribe the minimal and maximum values of the parameter they are interested in, avoiding the situation of non-meaningful parameter values. When it comes to outliers in observations or physical/biological boundaries of a parameter, actually, experimenters are more experienced than modellers in making judgements. And normally modellers consult experimenters on the quality and to which degree we can trust and use observation data. The observational data we used in EcoPAD-SPURCE went through the quality control from experimenters. We promote the hands-on experience for the 'general public' with prescribed examples to connect the 'general public' and ecological research. It is not to say we expect the 'general public' to understand the result displayed from the webpage without any guidance or consultancy with a specialist. We still need the modellers to support these activities and play an important role.

Nevertheless, we do not rule out the possibility of potential errors, it is good to be cautious. EcoPAD archives relevant model parameters, boundary conditions, model structure and observational data for each modelling activity. If there are erroneous results, they can be traced through the archives. It does not provide a mechanism to detect unaware erroneous results, but it helps in the situation when people suspect that there are errors.

Detailed comments: Ll 31-33: This sentence is hard to understand, what are updated data?

Response: We changed"updated data" into "new data".

L40: What is your definition of near real-time?

Response: In the SPRUCE study, EcoPAD is setup to automatically update forecasting every week and is adaptable to different updating frequency depending on the research goal. In this specific case, we refer to "weekly" as near real-time.

Ll 67-73: Maybe put a 'e.g.' in front of the mentioned references because these are only examples and there are many more possible references to cite as examples.

Response: Good suggestion. We add 'e.g.'

Ll 92-94: Unrepresented processes and unknown parameter values are two different reasons for large uncertainties in simulating ecological systems.

Response: We agree that unrepresented processes and unknown parameter values can be two different reasons for large uncertainties in ecological modelling. But uncertainty of parameters sometimes also contains information about unrepresented processes. The separation between processes and parameters are context and scale dependent. For example, the decomposition of soil organic matter or litter can be represented through the parameter decomposition rate. The uncertainty of decomposition rate partly reflects unrepresented processes such as microbial dynamics.

Ll 98/99: 'to communicate model with data' seems to be a weird expression.

Response: We change this expression to "to combine model with data".

Ll 122/123 Model improvements do not necessarily happen after the end of an field experiment, other ways of improving a model rely on literature or new theoretical understanding.

Response: We agree that there are other ways to improve model. We add " Data-informed" at the beginning of the sentence.

L 128: Interactive ecological forecasting does not require web-based technology.

Response: We modify the sentence to "The web-based technology facilitates interactions". There are different levels of "interactive", in this manuscript "The interactive feature of EcoPAD (v1.0) is reflected in the iterative model updating and forecasting through dynamically integrating models with new observations, bidirectional feedbacks between experimenters and modellers, and flexible user-model communication through web-based simulation, data assimilation and forecasting." (Lines 191-194, tracked manuscript)

Ll 148/149: This sentence is hard to understand, please clarify what you mean here.

Response: We rewrite this part as "The iterative model-data integration provides an approach to constantly improve ecological forecasting and is an important step especially for realizing near real-time ecological forecasting." And we explained that "Instead of projecting into future through assimilating observations only once, the iterative forecasting constantly updates forecasting along with ongoing new data streams or/and improved models."

L 175: Do you mean 'quantitative' forecasting?

Response: Yes.

L 220: Please specify in the manuscript how this is done.

Response: We add "Each project has a separate folder where data are stored. Data are generally separated into two categories. One is used as boundary conditions for modelling and the other category is related to observations which are used for data assimilation. Scheduled sensor data are appended to existing data files with prescribed frequency." (Lines 254-258, tracked Manuscript)

Ll 226/227: It would be interesting to see more details on how the data assimilation system can be independent on the specific ecological model. Usually, in a data assimilation system the underlying model and the applied data assimilation algorithm are closely connected on a code level.

Response: We agree that there are connections between different components. We added "Linkages among the workflow, data assimilation system and ecological model are based on messaging. For example, the data assimilation system generates parameters that are passed to ecological models. The state variables simulated from ecological models are passed back to the data assimilation system. Models may have different formulations. As long as they take in the same parameters and generate the same state variables, they are functionally identical from the "eye" of the data assimilation system." (Lines 256-262, tracked manuscript)

Ll 241-246: Hard to understand, maybe split in two sentences.

Response: We rewrite this part as "SOM decomposition modelling follows the general form of the Century model [Parton et al., 1988] as in most earth system models. SOM is divided into pools with different turnover times (the inverse of decomposition rates) which are modified by environmental factors such as the soil temperature and moisture."

Ll 249-252: Again, hard to understand, maybe split in two sentences.

Response: We rewrite the sentence as "Data assimilation is growing in importance as the process-based ecological models, despite largely simplifying the real systems, are in great need to be complex enough to address sophisticate ecological issues. These

ecological issues are composed of an enormous number of biotic and abiotic factors interacting with each other."

Ll 257-259: The underlying principle of Bayesian modelling is that the ingredients are specified by probability density functions.

Response: It is not clear to us what information the Reviewer intended to add here.

Ll 264/265: Complicated formulation, essentially what you want to say is that the posterior uncertainty is smaller than the prior after assimilating observations.

Response: We agree that what we want to express is that the posterior uncertainty is likely to be smaller than the prior after assimilating observations. We elaborate on this part because some readers of the manuscript might be ecologist/experimentalist with limited background in modelling and Bayesian statistics.

Ll 267-269: Please specify in the manuscript how you choose between DA techniques and what are the criteria for the selection.

Response: Please refer to our response to L 401. EcoPAD is open to different DA techniques.

Ll 271-273: Again, hard to understand, maybe split in two sentences.

Response: We delete "which makes Bayesian inference, especially these with multi-dimensional integrals, workable".

L 275: What are the various uncertainty sources and why do other methods do not take all these sources into account? Please specify in the manuscript.

Response: We remove the statement "is advantageous for better ecological forecasting as it" as it is not the objective of this manuscript to compare different data assimilation techniques.

Ll 296/297: What is a good management in the sense here?

Response: Good management is a subjective term. Nowadays Ecologists are working with large and heterogeneous ecological datasets routinely. Good management can broadly refer to management that improves the efficiency of activities that involve these large and heterogeneous ecological datasets.

Ll 394/395: What are youngster? And why should they study ecological dynamics through their phones and tables opposed to seniors or others?

Response: Youngster is a random example, instead of all-inclusive listing. We use youngster to delegate people who are not experts in ecology. We do not think we intend to have implicit meaning that says seniors or others should not do it. We apologize if we made readers feel in such way. To reduce over interpretation, we replaced youngster with "Non-ecologists, such as youngsters".

L 401: Doesn't that contradict your earlier statement that you need to choose a DA technique that is fit for purpose (Ll 267-269)?

Response: Ll 267-269 states "EcoPAD is open to different data assimilation techniques depending on the ecological questions under study since the scientific workflow of EcoPAD is independent on the specific data assimilation algorithm. For demonstration, the Markov chain Monte Carlo (MCMC) [Xu et al., 2006] is described in this study." We choose a DA technique for demonstration purposes and we do not state that only the chosen DA technique fits. Instead, we think our system is open to different DA techniques and L401 is not in contradiction with our previous statement.

Ll 428-430: How is the automated forecast done? And who is analysing the results of the automated forecast? I suppose if something goes wrong in the automated processing and forecasting an experimentalist won't be able a) notice that something went wrong and b) would be able to fix the bug/problem in the modelling chain.

Response: EcoPAD-SPRUCE is built upon the teamwork. There are both modellers and experimenters. We emphasize the interaction between experimenters and mod-

ellers, as illustrated through the section 3.4.1. Modellers built the automated forecasting algorism/code and experimenters also played an important role, such as, in preparing observations and interpretation of the modelling results. Experimenters are not good at finding out software bugs, but they might be more experienced in telling whether the modelling results make sense in reality. Details about how the automated forecast is done can be find in Section 3.3.

Ll 443-446: It seems that there is a misconception between parameters and parameterisations: parameters should be invariant in time otherwise they are not parameters but a result of a parameterisation that depends on independent inputs. Could you please clarify this point in the manuscript.

Response: We think the statement that whether a parameter should be time-independent is context dependent. People commonly link a parameter to a constant that does not change with time. But parameter does not equal to constant. The wiki takes parameter as "A parameter, generally, is any characteristic that can help in defining or classifying a particular system (meaning an event, project, object, situation, etc.). That is, a parameter is an element of a system that is useful, or critical, when identifying the system, or when evaluating its performance, status, condition, etc." (https://en.wikipedia.org/wiki/Parameter). And it is not uncommon to find "time-varying parameters" or "time-variant parameters" in literature, e.g., Tucci 1995, Lauzon and Bates, 1991; Zellner et al., 1991; Zeng et al., 1998; Jiang et al., 2015.

L 500: What are the SPRUCE communities doing with the results?

Response: The results are used mostly for research. From the modelling part, Case 5 (section 3.4.5) is based on this part and ongoing studies are using these archived near-time forecasting to track the time-shift in acclimation and to track model elements that contribute to reducing forecasting uncertainty. The experimenters may adjust their sampling scheme, e.g., the sampling frequency or additional variables to be measured to reduce the forecasting uncertainty.

L 512: ‘help experimenters think’ is an interesting expression.

Response: We do not understand what the Reviewer intended to express here.

Ll 712-714 Could you please clarify this statement. I don’t think this is true, complex models can of course assimilate pool-related data, see e.g. Thum et al., 2017.

Response: The sentence is “In the past, complex models could not assimilate pool-related data to constrain their parameter estimation due to insurmountable computational demand in large scale studies.” The context is “large scale studies”. Thum et al., 2017 is about site level studies, not large-scale studies. For example, Bloom et al., 2016 assimilated large-scale pool-based observations. So we deleted this paragraph.

Ll 729-732: Again, hard to understand, please clarify what you want to say here.

Response: We replace it with “Parameter values derived under the ambient condition was not applicable to the warming treatment in our methane case due to acclimation”.

Figure 7: This figure is hard to understand and also the caption doesn’t help much to understand the panels. What exactly has been changed between S1-S3? What is realised and unrealised forecasting? And there seems to be no difference in time-scale among the panels.

Response: The differences between S1-S3 are weather forcings and are indicated by “The upper panels show 3 series of forecasting with updated vs. stochastically generated weather forcing (Lines 1352-1353, tracked manuscript)”. We changed “realised” and “unrealised” to “updated” and “un-updated” respectively to reduce confusion. S1 is “un-updated” forecasting and the forecasting is generated with stochastically generated weather forcings over our whole forecasting period (2015-2024). S2 and S3 are updated forecasting. S2 is updated through replacing the stochastically generated weather forcings by measured real weather forcings from January 2015 to July 2016. And S2 then forecasts the period from August 2016 to 2024 with updated states. S3 is updated with measured forcings from January 2015 to December 2016 and forecast

after the end of the real measured forcing. The timing of updating is randomly chosen for demonstration purposes. We added specific time-periods hopefully to make it clear about when measured vs. stochastically generated forcings are used. We also cleared it in the description with “ red corresponds to updated forecasting with two stages, that is, updating with measured weather forcing from January 2015 to July 2016 followed by forecasting with 100 stochastically generated weather forcing from August 2016 to December 2024 (S2)” (Lines 1355-1357, tracked manuscript).

Typos: Ll 126, 140, 154, 160, 187, 324, 456, 566

Response: We correct typos throughout the manuscript.

References: Bloom, A. A., J. F. Exbrayat, I. R. van der Velde, L. Feng, and M. Williams (2016), The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times, Proceedings of the National Academy of Sciences of the United States of America, 113(5), 1285-1290, doi:10.1073/pnas.1515160113

Jiang, C., L. H. Xiong, D. B. Wang, P. Liu, S. L. Guo, and C. Y. Xu (2015), Separating the impacts of climate change and human activities on runoff using the Budyko-type equations with time-varying parameters, Journal of Hydrology, 522, 326-338, doi:10.1016/j.jhydrol.2014.12.060

Lauzon, A. M., and J. H. T. Bates (1991), ESTIMATION OF TIME-VARYING RESPIRATORY MECHANICAL PARAMETERS BY RECURSIVE LEAST-SQUARES, Journal of Applied Physiology, 71(3), 1159-1165.

Zellner, A., C. Hong, and C. K. Min (1991), FORECASTING TURNING-POINTS IN INTERNATIONAL OUTPUT GROWTH-RATES USING BAYESIAN EXPONENTIALLY WEIGHTED AUTOREGRESSION, TIME-VARYING PARAMETER, AND POOLING TECHNIQUES, Journal of Econometrics, 49(1-2), 275-304, doi:10.1016/0304-4076(91)90016-7

Zeng, Z., R. M. Nowierski, M. L. Taper, B. Dennis, and W. P. Kemp (1998), Complex population dynamics in the real world: Modeling the influence of time-varying parameters and time lags, Ecology, 79(6), 2193-2209, doi:10.2307/176721 Tucci, Marco P. 1995, Time-varying parameters: a critical introduction. Structural Change and Economic Dynamics, Volume 6, Issue 2, June 1995, Pages 237-260.

Please also note the supplement to this comment:
https://www.geosci-model-dev-discuss.net/gmd-2018-76/gmd-2018-76-AC2-supplement.pdf

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-76, 2018.