

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

Anonymous Referee #1

Received and published: 28 June 2018

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.

C1

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.

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.

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

C2

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

LI 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.

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

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

LI 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.

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

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

L 175: Do you mean 'quantitative' forecasting?

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

LI 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.

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

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

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

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

C3

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

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

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.

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

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

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

LI 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.

LI 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.

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

L 512: 'help experimenters think' is an interesting expression.

LI 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.

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

Figure 7: This figure is hard to understand and also the caption doesn't help much

C4

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.

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

Reference: T. Thum, N. MacBean, P. Peylin, C. Bacour, D. Santaren, B. Longdoz, D. Loustau, P. Ciais: The potential benefit of using forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites, *Agricultural and Forest Meteorology*, <https://doi.org/10.1016/j.agrformet.2016.12.004>.

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