This study, "Realized ecological forecast through interactive Ecological Platform for Assimilating Data into model (EcoPAD)," by Huang et al. introduces a web-based platform for data-model integration framework that "can" be used for ecological forecasting. The manuscript introduces the conceptual components of the framework relatively well, albeit with too much generic details on databases, web-based workflow, metadata, data API, which are not the focus of GMDD. Even though I feel that the platform presented in the manuscript has a huge potential, I find the current state of the platform, and the example cases presented are not mature yet (with only one model, one data assimilation, one site). There is a little scientific advance from the study with results based on previous studies. The results and evidences presented, therefore, do not back the claim of 'have-it-all' platform that can be used by the scientists and 'citizens' alike. I think the manuscript should focus on specifically "what has been done" with thorough scientific discussion, and not "what can potentially be done." This would help judge if the platform is truly flexible and interactive.

The web-based workflow can be viewed as a specific technological advancement in the field of ecological forecasting, but web-based frameworks have been around for a while in the field of geoscience, e.g., PEcAn (as cited in the manuscript in line 138) and PALS. Therefore, I do not agree that it is already the first flexible framework as the manuscript claims. In fact, such claims are not always necessary, but that might just be my personal opinion.

The quality of the figures should be improved, and the redundant information in the schematics should be eliminated. Also, the sources of the images used in the figures are not shown in the respective figures or captions. In general, the schematics can be more technical to suit the expertise of the reader-base of the GMDD.

The manuscript has several Grammatical errors and typos. At times, it feels like even a simple "spelling checker" has not be run through the whole manuscript once. At the same time, some paragraphs are excellently written without a blip.

Major comments:

- The manuscript does not clarify what the "ecological forecasting" means. In the first paragraph, there are several examples of "ecological forecasting" from previous studies. In the end, the

EcoPAD seems to be simulating the carbon stocks and fluxes, which is just an aspect of ecological forecasting. The title should be revised to be more specific to the scope and capabilities of EcoPAD.

- There is no specific section on the benchmarking of the performance of the EcoPAD simulations. This is a critical step to have a reliable platform that can be used for forecasting. Therefore, evaluation of model performance should be presented in detail in this manuscript.

- The examples presented here are for SPRUCE sites. It is not clear whether EcoPAD can be used easily in other sites, even though manuscript ends with statements on how the framework can easily be implemented for different FLUXNET sites and at continental scales. If such implementations are simple as the manuscript claims, they should be the main focus of the manuscript because the results presented here have been, at least partly, used in previous studies.

- Once again, the results presented here just seem like a summary and discussion of previously published manuscripts from the main author and/or co-authors of the manuscript. In fact, I found the results presented in the Appendix A2 to be far more interesting than the results presented in the main text. There should be discussion on why most of the parameters are not well-constrained (Figure A5, right panel), or why Q10 parameters for CH4 is not as well constrained as those for r and why they differ for different temperature treatments (Figure 6). I understand that there may be counter argument on this issue being out of the scope of the current paper, but, it is necessary to discuss how these potentially unconstrained parameters affect the forecast skills of EcoPAD. After all, general public, who do not understand the technical and scientific details, may easily be misinformed with the uncertain forecast of EcoPAD.

- It is not clear from the manuscript what models or assimilation methods are currently available. There are several instances of "ensembles" and "structural uncertainty" but as far as I could decipher, EcoPAD only has one model and one data assimilation method until now. This is also relevant to explaining how EcoPAD can be used to quantify uncertainty from different sources. Such quantification of uncertainty would require factorial experiments with multiple model structures, process/mechanism formulations, cost functions, optimization/assimilation schemes with multiple observational constraints, and so on. It is not clear if EcoPAD already has such

functionalities or if it is yet another potential use. If so, an explanation of how "ecologists" can add such functionalities would be useful. For example, is the interested developer responsible for creating a separate docker that satisfy all the system requirements for his/her own system? I could not test "adding functionality" because it requires registration to the system.

Specific comments:

Line 53-55: The manuscript does not have any results or discussion on this, and thus this sentence should be removed from the abstract and the whole manuscript.

Line 61: one science → a science?

Line 62: Isn't forecasting always for future?

Line 87-88: what are the "relevant mechanisms" that the previous systems are lacking and how does EcoPAD, and TECO therein, address these shortcomings?

Line 110: one-directionary → unidirectional

Line 114-128: I think the CARDAMOM model-data fusion system (Bloom et al., 2016) deserves a mention in this paragraph (http://www.pnas.org/content/113/5/1285) and in further discussions.

Line 132, 141, 146, 147, 252: spelling errors. I am not mentioning all the places here. Please check the whole manuscript carefully.

Line 151-153: It's not clear what this sentence means.

Line 176: qualitative means better quality or is it quantitative?

Line 210: Should clarify what 'big data' means in this context. Diverse data?

Line 215: cite FLUXNET

Line 305: MongDB → MongoDB

Line 338: May be better to define what IT stands for, just for the sake of completeness.

Line 345: Does docker have a website or citation?

Line 348-350: Isn't it redundant (unnecessary duplication) to include input data in the docker?

Line 381: I think the web-based platform is for job submission and not web-based simulation.

Line 404: clarify what 'scientific values' means.

Line 415-422: Bayesian statistics has been used in previous ecosystems studies. Please find and cite these previous studies.

Line 433-438: I wonder if experimental/observational setup can be carried out in such short notice.

Line 473: Is +0 the same as control experiment?

Line 479: Is there any difference between data-model integration and data-model communication? If so, this should be clearly stated at the beginning. Both have been used frequently in the manuscript.

Section 3.3: Is this process done systematically or through personal communication? If systematically, how are the updates (both of models and experiments) carried out theoretically and technically?

Line 548-552: In figure 6, it seems like the parameter 'r' is not well constrained for higher treatments of temperature. Discuss the reasons for this.

Line 580: EcoPAD only includes one model, but the sentence says that it relies on ensembles of ecosystem models. This statement is misleading.

Line 586-611: Summary of Jiang et al., 2018 and Huang et al., 2017. So, the results presented in current study are specific for the models and tools used in those previous studies? If there is any additional scientific advancement in EcoPAD, this should be highlighted here.

Line 620-630: To what extent does the parameter ranges depend on the uncertainty in the forcing? Is there a particular reason why the parameter values are randomly sampled from the posterior? Doesn't it make sense to use the posterior distributions to get the parameter ranges within certain confidence intervals?

Line 631-642: It is not clear if GPP has an effect on carbon stocks in the TECO model.

Line 668: It is not clear how these 'scientific' information is directly useful for general public.

Line 680-681: I am not convinced that all 7 characteristics of EcoPAD have been backed by evidences presented in this manuscript. At least, this has not been clearly presented in the manuscript.

Line 688-705: The discussion here should be divided into the users (those who run the model) and developers (those who add processes and methods to EcoPAD). Since the developers need to carry out a lot of set-up using the GitHub repository, the web-based platform seems more suited to the users. This limits the options of the users to only the ones already available in EcoPAD, which is, as of now, only one model and one data-assimilation system for one site. As such, the potential applications of the model are not applicable to the web-based system. This should be clearly mentioned in the abstract, main text, and the conclusions.

Line 722: 'model structure' → In this use, does it mean different formulations of one process as in Jiang et al., 2018?

Line 744-745: What about the interactions between fluxes and pools?

Line 787-788: Assuming this statement is based on Table 1. But, it is not clear if the table is just a hypothetical example or based on the actual experience.

Line 790-791: I just wonder if it is too risky for experimenters to invest resources on carrying out experiments recommended by modelers who used one-single model?

Line 804-817: I think these tasks of including several sites or using EcoPAD at continental studies should be a part of this manuscript. As I have mentioned previously, the results presented here have been published in previous studies. Using it in different ecosystems will validate the scientific soundness of EcoPAD and it will provide sufficient evidence of its potential wide-scale applications.