Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-96-RC1, 2018

Biogeosciences Discussions

Interactive comment on "Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation" *by* Istem Fer et al.

M. van Oijen (Referee)

mvano@ceh.ac.uk

Received and published: 19 March 2018

GENERAL COMMENTS

This paper shows how Bayesian calibration of computationally demanding ecosystem models can be sped up using various techniques. The authors test their methods on two ecosystem models, SIPNET and ED2, using synthetic calibration data and flux data from a forest site. A small fraction of the models' parameters (theta) is calibrated by means of Markov Chain Monte Carlo (MCMC) simulation. Instead of running the ecosystem models at each iteration of the MCMC to calculate the likelihood function, L(theta), the authors first derive emulators for the contribution to L(theta) of different

data sets (representing different measurement variables) at any point in parameter space. Building the emulators requires running the original models on a set of training points, but thereafter, in the MCMC, the fast emulators are used at every iteration instead of the slow models.

Contrary to statements made in the paper, the techniques used by the authors are for the most part not novel. There is in fact a substantial literature on replacing the likelihood function with more efficient calculation methods, and I shall give pointers to the literature below. Overall it seems that the literature is very poorly referenced in this paper.

However, in the field of ecosystem modelling, several techniques described by the authors have been used hardly at all, so the paper can be valuable in introducing the ideas to a new audience. Moreover, the tests carried out by the authors demonstrate the effectiveness of the techniques very well, and they present very clear figures and tables. But to introduce new methodological ideas to people, the language should be clear and consistent, and that is not the case here. There is a worrying lack of understanding of the difference between the concepts of 'error' and 'uncertainty'. The first refers to deviation from truth, the second to incomplete knowledge, but in this paper the terms are occasionally treated as synonyms, which makes the Introduction highly unclear. [Proper terminology for these concepts and others can, for example, be found in the review of Bayesian methods by Van Oijen (2017), where also additional references on MCMC, emulation and hierarchical modelling in ecosystem modelling can be found.]

Missing references to the literature include the following. Guttmann & Corander (2016) gave a useful overview of many different ways to replace the likelihood function with faster alternatives (their Table 1 is helpful). Oakley & Youngman (2017) showed many of the same methods as the present authors do. They even provided a six-step procedure that is almost identical to the steps outlined in Fig. 1 and section 2.1 of the present paper. For many examples of likelihood-emulation using Gaussian processes

etc. in cosmology, see Aslanyan et al. (2015) and references 7-24 therein (which also tend to focus on how much computations are made faster by likelihood-emulation). And just as in the present paper, Aslanyan et al.'s procedure alternates between posterior estimation and emulator improvement. Jandorov et al. (2014) used the same refinement employed in the present paper, of emulating sufficient statistics instead of the overall likelihood directly. In contrast to that, Kandasamy & Schneider show that instead of emulating the likelihood, it is also possible to emulate the product of prior and likelihood (i.e. the posterior up to a constant), an approach not mentioned by the present authors. On pages 6-7 of Yurko et al. (2015), some mathematical details are provided of using a 'GP emulator-modified likelihood function'. I further recommend that the authors inspect the literature on ABC and especially on Bayesian quadrature and Bayesian optimization to find ideas that may help refine their approach, and ground it in the wider literature. Further, as perhaps an unmentioned predecessor of calibrating scale parameters rather than the original parameters, see the ecosystem model Bayesian calibration approach of Van Oijen et al. (2011), where every separate data stream came with its own bias parameter.

The Introduction mentions that "Parameter error refers to the uncertainty about the true values of the model parameters", which is quite wrong. Parameter error means assigning a value to a parameter which differs from reality, e.g. stating that the light-use efficiency is 1 g MJ-1 when in reality it is 2 g MJ-1. Not knowing whether it is 1 or 2 or anything else is uncertainty. It is therefore also incorrect to state, as the authors do, that "parameter error asymptotically goes to zero with enough data". It is the conditional uncertainty that goes to zero, not the error. Every experimentalist knows that having any number of biased measurements makes no parameter converge to its correct value - and all measurements have their hidden or unhidden biases. There is no safe way to "estimate observation error from data".

The treatment of the subject matter in the Introduction is further hampered by poor terminology regarding parameters. Terms like "parameter", "parameter vector", "pa-

rameter set[s]" are used arbitrarily and inconsistently. [As an exercise for the reader: show that lines 98 and 147 cannot both apply.] Note that a set is unordered and a vector is ordered, so a point in parameter space can not be a "parameter set". And "covariances among parameters" are not real quantities but statistical quantities that capture part of our uncertainty and that change when more data come in. Therefore the covariances are in no way "accounted for". Please note that your subject matter of Bayesian calibration using MCMC is unfamiliar to many readers, so getting an idea of what is going on requires using precise language. Apologies for these pedantic remarks, but in my experience people stumble over the smallest inconsistency when learning Bayesian methods.

Can you elaborate on the limitations of your approach? What is the maximum number of parameters (p) that can be calibrated in general, and for your two models in particular? You set the number of model-runs at p3. Does that mean that calibrating 100 parameters is unfeasible because it would require 106 model evaluations just to build the emulator? And how exactly does PEcAn calculate the contributions of different parameters to overall uncertainty, i.e. what was the screening algorithm?

Published methods for Bayesian calibration increasingly take into account that models are imperfect. There is a discrepancy between model output and reality, even at the best possible setting of model parameter values. This discrepancy is often modeled as a Gaussian Process for which - in the Bayesian calibration - the hyperparameters are estimated together with the regular model parameters. Likelihood-emulation precludes including discrepancy-estimation because model outputs are not calculated during the MCMC. Please add a discussion of this limitation of your approach.

SPECIFIC COMMENTS

There are linguistic errors (plural subjects with singular verbs, missing definite articles etc.) on lines 54, 55, 92, 93, 100, 183, 201, 248, 294, 306 (twice), 309, 323, 351, 372, 418, 434, 436, 443, 454-455, 482 (twice), 483, 484, 485, 507, 511, 520, 539 (twice),

581. I will have missed many more errors - it might be good to involve a native speaker.

The last sentence of the Abstract (l. 34-36) can be deleted without loss of content.

How is the "Euclidean distance between confidence intervals" determined?

Why were 729 knots used for p=8 parameters of SIPNET, given that you state the need for p3 knots (729=93, not 83)?

Two of the references are not placed in their proper alphabetical position, and the reference to Hartig et al. (2012) is missing.

Can you explain the results shown in Tables A2 and A5? How can posterior distributions for parameters following MCMC neatly fall into parameterised probability distributions (which also are often of different type than their priors)? And what were the posterior covariances?

REFERENCES

Aslanyan, G. et al. (2015). Learn-as-You-Go Acceleration of Cosmological Parameter Estimates. J. Cosmology and Astroparticle Physics 09: 005. https://doi.org/10.1088/1475-7516/2015/09/005. [see also http://arxiv.org/abs/1506.01079]

Chowdhury, A. & Terejanu, G. (2016). An Enhanced Metropolis-Hastings Algorithm Based on Gaussian Processes. In: Model Validation and Uncertainty Quantification, Volume 3, 227-33. Conference Proceedings of the Society for Experimental Mechanics Series. Springer, Cham. https://doi.org/10.1007/978-3-319-29754-5_22.

Gutmann, M.U. & Corander, J. (2016). Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models. J. Machine Learning Res. 17: 1-47.

Jandarov, R. et al. (2014). Emulating a Gravity Model to Infer the Spatiotemporal Dynamics of an Infectious Disease. J. Roy. Stat. Soc., Series C, Appl. Statist. 63: 423-44. https://doi.org/10.1111/rssc.12042.

Kandasamy, K. & Schneider, J. (2015). Bayesian Active Learning for Posterior Estimation. Proc. 24th Int. Conf. Artifical Intelligence: 3605-3611.

Oakley, J.E. & Youngman, B.D. (2017). Calibration of Stochastic Computer Simulators Using Likelihood Emulation. Technometrics 59: 80-92. https://doi.org/10.1080/00401706.2015.1125391. [Preprint at http://arxiv.org/abs/1403.5196]

Van Oijen, M. et al. (2011). A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest, Agric. Forest Met. 151: 1609-1621. 10.1016/j.agrformet.2011.06.017.

Van Oijen, M. (2017). Bayesian Methods for Quantifying and Reducing Uncertainty and Error in Forest Models. Current Forestry Reports 3: 269-80. https://doi.org/10.1007/s40725-017-0069-9.

Yurko, J.P. et al. (2015). Demonstration of Emulator-Based Bayesian Calibration of Safety Analysis Codes: Theory and Formulation. Science and Technology of Nuclear Installations. https://doi.org/10.1155/2015/839249.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-96, 2018.