Geoscientific
Model Development
Discussions

Interactive comment on "Surrogate-assisted Bayesian inversion for landscape and basin evolution models" *by* Rohitash Chandra et al.

Anonymous Referee #1

Received and published: 19 June 2019

Description

This paper attempts to present Bayesian inference for an inverse problem in landscape and sedimentary basin evolution models, with acceleration by parallel tempering and use of a surrogate likelihood in the form of a neural net that is incrementally trained along the chain using evaluations of the true model. The paper attempts optimization to perform inversion, and the chain adapts the surrogate.

Comments on content

I will restrict my comments to the computational Bayesian aspects as I am no expert in models for landscape and sedimentary basin evolution. However, the paper makes virtually no mention of the evolution model used, and the majority concerns Bayesian computation, so my comments actually account for the brunt of the content of this paper.

Parallel tempering is an arcane algorithm that probably gives no advantage

One can see from the references that the computational Bayesian method used in this paper predates 1997, which is when the first serious attempts at Bayesian computation for inverse problems were made. There has been tremendous development in algorithms since then, particularly in terms of efficiency with respect to inverse problems. This manuscript seems entirely unaware of these developments. (The more recent statistical references in this manuscript are not to sampling methods, and frequently to the authors' own papers.)

It was recognized soon after 1997 that while parallel tempering is a nice idea, it is not very effective for inverse problems, and suffers from fundamental difficulties. One such major difficulty is the need for tuning "psuedo priors" to allow the parallel chains to mix – this is mentioned in the original paper [MP92] (cited in the present manuscript) and studied in subsequent papers, but no mention of this issue is made in the present manuscript. It also became known that parallel tempering's use of parallel resources is essentially trivial; one can see from the distribution for the ensemble (Eqn. on p 4 line 3) that the parallel chains are statistically independent, so there is no efficiency gain over simply running a parallel instance of single chains that move up and down tempering levels. That multilevel algorithm has been implemented in multiple guises, most effectively in the delayed acceptance algorithm [CF05] and multilevel versions,

and the variance-reduction methods of multi-level Monte Carlo [DKST15].

These algorithms significantly outperform parallel tempering as used in this manuscript; see [HRMVF11] for a review of some of these topics. On the other hand, there is no evidence presented in the present manuscript that parallel tempering actually leads to improved computational efficiency (beyond running separate parallel chains), and my impression is that it does not

A further note is that the present manuscript makes no mention of the random-walk proposal used, though this is VERY critical to the efficiency of the method. There are modern methods that use *correlated* parallel chains and use the information across chains to adapt to an optimal random-walk proposal [SOLHTJ12] that would be much more efficient than the method used in this manuscript.

0.0.1 The proposed method is not ergodic for the target distribution

One can see in Alg. 1 (p11) that the surrogate is trained, or *adapted* in the language of MCMC, as the algorithm proceeds. Hence this MCMC is not stationary and does not satisfy the conditions for standard MCMC to be ergodic for the desired target distribution. Indeed, it is easy to see from the structure of Alg. 1 that it will not target the desired posterior distribution. That is to say, Alg. 1 converges (if it converges at all) to some distribution that is not the target. So it is somehow useless for performing a quantitative solution of this inverse problem.

The present literature contains potential fixes to these problems, yet the present manuscript makes no mention of them. In particular, the surrogate transition method [L01] would allow Alg. 1 to correctly use a fixed surrogate, *at no increase in computational cost*, while the adaptive algorithm in [CFO11] gives a framework for provably ergodic methods that accommodate the adaptation of the surrogate, again *at no increase in computational cost.*

0.0.2 The algorithm in the manuscript is not run to convergence

This is evident from Fig. 8, and elsewhere.

0.0.3 This manuscript does not implement Bayesian methods

Bayesian analysis produces a posterior distribution over possible solutions to the inverse problem. It is then necessary to summarize the posterior distribution, in a way that is appropriate for the problem at hand. Without specific evidence, there is no reason that optimizing the posterior distribution to give the MAP estimate, as attempted in this manuscript, is a good summary statistic. Indeed, it is known that in even moderate dimension inverse problems the MAP estimate can be arbitrarily far from the bulk of feasible solutions, and can be sensitive to noise realizations – that is to say that it is a hopeless summary statistic. The computation attempted in this manuscript is not a sensible summary statistic for a Bayesian analysis.

Also, the use of Bayesian modelling in this manuscript is entirely bogus. For example, the sum-of-squares log likelihood (p10 l 6, and elsewhere) has no physical meaning for sediment transport, while plenty of more physically-realistic measures are available. Thus, the likelihood function used is not a sensible Bayesian model. Prior models are implicitly uniform, which makes no sense in terms of Bayesian probabilistic modelling of the model parameters. Uncertainty in distributions, via hyperpriors, is not even considered.

This manuscript widely advertises its Bayesian credentials and the use of Bayesian inference (in the title, abstract, introduction), yet does not implement any sensible Bayesian methods. It certainly does not achieve "a rigorous approach to uncertainty quantification" (p2, l1).

0.0.4 The technical writing is poor in places

For example, lines 5, 6, 7 on p10 notation is inconsistent: "The likelihood function $L_e(\theta)$ is given by $L_l(\theta) = ...$ where the subscript e, in $L_e(\theta)$," then again $L_l(\theta)$ used in l 19 p10, and so on. Line 9-10 in Alg. 1, L_{local} is not defined. Line 5 in Alg. 1, proposal density is not defined. Line 22 on p1 "deterministic geophysical forward model can be seen as a probabilistic model via Bayesian inference" is nonsense.

0.0.5 There are some positive aspects about this manuscript that could be publishable

Some of the computed results in Fig. 2, 5, 6 look interesting, to my untrained eye. The idea of using a neural net surrogate to accelerate computation is interesting, though a somewhat obvious one given the current hype around neural nets. Nevertheless, the attempt to use neural nets in this way is interesting. I must reemphasize, as stated above, that the use of the surrogate ought to be performed within one of the well-established algorithms for correct use of a surrogate, while the algorithm in this manuscript is ad-hoc and incorrect. The analysis in this manuscript is actually a maximum likelihood calculation – though these estimates are known to suffer from quantitative problems, as outlined above, and notwithstanding the issue with the unphysical likelihood modelling, mentioned above. The authors might consider re-presenting this work with an accurate description of the calculation as a maximum likelihood (and dropping all mention of Bayesian methods, which appears to be far from the authors' skill set).

C5

Conclusions

This manuscript makes no contemporary contribution to Bayesian methods or computation. Indeed, in terms of Bayesian computation, the methods presented in this manuscript constitute a step backwards by some decades, while attention to convergence is completely absent and hence a major deficiency. The probabilistic modelling is crude, to the point of being worthless. The use of parallel chains and a neural net surrogate seem more an exercise in programming than an efficient, quantified solution to a scientific problem. Publishing this paper would be a disservice to the community as it grossly misrepresents the current literature on Bayesian modelling and computation for this problem. The current literature contains better and more computationally efficient solutions.

If this manuscript were submitted to a journal that deals with Bayesian computation, I doubt the editor would even send it out for review.

I recommend this manuscript be rejected.

References

Marinari, E. and Parisi, G.: Simulated tempering: a new Monte Carlo scheme, EPL (Europhysics Letters), 19, 451, 1992.

Christen, J.A. and Fox, C., 2005. Markov chain Monte Carlo using an approximation. Journal of Computational and Graphical statistics, 14(4), pp.795-810.

Dodwell, T.J., Ketelsen, C., Scheichl, R., Teckentrup, A.L.: A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. SIAM/ASA Journal on Uncertainty Quantification 3(1), 1075–1108 (2015)

Higdon, D., Reese, C.S., Moulton, J.D., Vrugt, J.A. and Fox, C., 2011. Posterior exploration for computationally intensive forward models. Handbook of Markov Chain Monte Carlo, pp.401-418.

Solonen, A., Ollinaho, P., Laine, M., Haario, H., Tamminen, J. and Järvinen, H., 2012. Efficient MCMC for climate model parameter estimation: Parallel adaptive chains and early rejection. Bayesian Analysis, 7(3), pp.715-736.

Liu, J.S., 2001. Monte Carlo strategies in scientific computing. Springer Series in Statistics.

Cui, T., Fox, C. and O'Sullivan, M.J., 2011. Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. Water Resources Research, 47(10).

C7