The author's response file is structured as follow:

Comment(s) of the reviewers in bold

Authors' responses

Authors' changes to the manuscript

Editor in chief

Dear authors,

Unfortunately, after checking your manuscript, it has come to our attention that it does not comply with our "Code and Data Policy".

https://www.geoscientific-model-development.net/policies/code_and_data_policy.html

You have archived your code on GitHub and pesthome.org. However, neither of them are suitable repositories for scientific publication. GitHub itself instructs authors to use other alternatives for long-term archival and publishing, such as Zenodo.

In this way, if you do not fix this problem, we will have to reject your manuscript for publication in our journal. I should note that, actually, your manuscript should not have been accepted in Discussions, given this lack of compliance with our policy. Therefore, the current situation with your manuscript is irregular.

Therefore, please, publish your code in one of the appropriate repositories, and reply to this comment with the relevant information (link and DOI) as soon as possible, as it should be available for the Discussions stage.
Also, you must include in a potentially reviewed version of your manuscript the modified 'Code and Data Availability' section, the DOI of the code.

Finally, I have to note that the pesthome.org page says that the software is free. However, it does not seem to be in the sense of FLOSS "free-libre-open-source", which is what we request in our journal. In the GitHub repository and web page, there is no license listed for the software. If you do not include a license, despite what you state on the web page, the code is not "open-source"; it continues to be your property, and nobody can use it. Therefore, when uploading the model's code to one of the repositories that we can accept, you could want to choose a FLOSS license. We recommend the GPLv3. You only need to include the file 'https://www.gnu.org/licenses/gpl-3.0.txt' as LICENSE.txt with your code. Also, you can choose other options that the repositories provide: GPLv2, Apache License, MIT License, etc.

Please, fix the issues mentioned, and reply to this comment with the information requested.

Juan A. Añel
Geosci. Model Dev. Exec. Editor

Dear Editor,

Thank you for considering our paper. We have added the source codes for PEST and especially DSI in a Zenodo repository (https://doi.org/10.5281/zenodo.7913402) with a license file in it. Modification has been made accordingly concerning the Code Availability section in the revised manuscript.

We have also made revisions throughout the document to improve overall clarity and consistency of expression to demonstrate the DSI methodology and how it can (and should) be used in decision-based modelling contexts. We hope that this will facilitate the quantification of uncertainty of complex, highly parametrized models.

Best,

H. Delottier, P. Brunner and J. Doherty

Reviewer #1

This paper represents an interesting approach to greatly reducing the computational costs of estimating prediction uncertainties for decision-relevant variables arising from complex spatial fields of parameters using the Data Space Inversion (DSI) surrogate model approach for both model calibration and the evaluation of data worth. I cannot find fault with the methodology, only with the assumptions on which it is based.

In effect the modelling problem is shoehorned into an aleatory Gaussian framework (with transformation of variables being suggested if necessary), when in general in real problems of this type we are dealing with epistemic uncertainties rather than purely aleatory uncertainties. In the example given, this is partly avoided because the same model is used to create the reality, as is run to create the realisations that provide the information for the DSI methodology. There is some nonlinearity in the model (as is illustrated by the difference between the DSI and ensemble smoothing results) but there is absolutely no doubt that the Gaussian assumptions about "measurement error" and the structural features of the field are correct. The reasoning is therefore to a degree circular.

I would be much happier if all the demonstrations that the method works pretty well for the synthetic case were consigned to an electronic supplement and the authors applied the methodology to a real case where, given the likely epistemic uncertainties, the assumptions might be more difficult to justify.

We welcome your comment and for your time and consideration.

Epistemic uncertainty is difficult to explore as, by definition, it is not part of the prior parameter probability distribution. Hence all uncertainty methods have difficulties with this issue, and it is a given that the outcomes of any method of uncertainty analysis will be compromised.

Sometimes prior-data conflict is apparent from the failure of model outputs to span field observations when the prior is sampled. Software from both the PEST and PEST++ suites that work in conjunction with the approach described in the paper are able to detect whether this is the case, based on samples of the prior that are required to build the DSI model.

Another means through which a defective prior parameter probability distribution can be detected is through the estimation of parameters that violate the prior. Hence if, in history-matching the DSI surrogate model, a large number of parameters are forced to adopt values that are of low probability for independent Gaussian parameters, this can be construed as prior data conflict. A modeller would be well advised to take note of this. Both of the above methodologies for detecting epistemic uncertainty are easily implemented in conjunction with workflows that are discussed in our paper.

We also note that it is uncommon for papers on uncertainty analysis to draw attention to epistemic uncertainty in general, and prior-data conflict in particular, because it is obvious that uncertainty analysis is only as good as the prior – regardless of the method that is used to undertake uncertainty analysis. On the other hand, papers that are dedicated to methodologies that explore and expose prior-data conflict (rightly) focus on this aspect of uncertainty analysis. In short, we do not see that it is incumbent on anyone who wants to discuss an uncertainty analysis methodology to also discuss epistemic uncertainty, for everybody knows that (a) it is an issue and (b) it is unquantifiable but important.

We note that there have been some recent papers in which hyperparameters of highly-parameterized prior probability distributions are estimated at the same time as the parameters themselves. These methodologies do address, to some extent, inadequacy of the prior parameter distribution by treating these inadequacies as uncertainties. This is a worthy undertaking, but is also accompanied by a considerable amount of numerical difficulty. Hence they are worthy topics of publication in their own right.

See, for example:

Oliver, D.S., 2022. Hybrid iterative ensemble smoother for history matching of hierarchical models. Math. Geosci, 54:1289-1313,

Chada, N.K., Iglesias, M.A., Roininen, L. and Stuart, A.M. (2018). Parameterisations for ensemble Kalman inversion. Inverse problems 34 https://doi.org/10.1088/1361-6420/aab6d9

It is also worthy of note that both of these high-quality papers demonstrate their methodologies using entirely synthetic cases – with less of a resemblance to hydrogeological reality than our own example.

It is, therefore, our opinion that our paper covers the topic to which it is dedicated to the extent required to demonstrate the veracity of the method that it employs, in such a way as to allow other modellers to use this method with little difficulty. We make no claims that the methodology that we present is exempt from problems which beset all uncertainty analysis methods. We note this in our paper, but would rather leave the discussion of these important (but ancillary) issues to other papers. We have added the following paragraph in the conclusion section of our paper to make this clear.

« *Realisations that compose the initial ensemble from which model outputs are calculated do not have to be multiGaussian. The multiGaussian assumption used to link measurements of past system behaviour to predictions of future system behaviour is independent of any assumptions about the prior realisations. Because a direct link is made between measurements and predictions (thereby bypassing parameters), an assumption of multiGaussianality is likely to have a weaker effect on the results of the predictive uncertainty analysis process than highly parameterised methods that rely implicitly or explicitly on parameter adjustment (such as linear Bayesian methods, randomised maximum likelihood methods and iterative ensemble smoother methods). Thus, prior realisations used to start the DSI process can accommodate both aleatory and epistemic uncertainties. Uncertainties in prior parameter distributions can also be readily accommodated.* »

In addition to these changes, we've also added a paragraph at the beginning of Section 3 to make it clear that our application is based on a synthetic model rather than a real-world model, and that we, therefore, do not consider the potential effects of epistemic uncertainty.

« The objective of this section is to demonstrate the performance and utility of DSI in quantifying posterior uncertainties of predictions made by a model whose run time is long and whose parameter field is complex. As is common in the literature, where the performance of a new method is tested and documented, we base our analyses on a synthetic model rather than on a real-world model. This allows us to assess, and document the performance of the method. It also dispenses with the need to account for epistemic uncertainties which accompany real-world modelling. »

Just one further comment. Not sure I understand how the lack of persistence of connectivity of the posterior parameter fields can lead to faster travel times than in the prior sample of simulations (Line 532) – is it not the fast pathways in the connected high conductivity alluvial channels that would give the fastest times? If that connectivity is lost how can it speed things up? Perhaps a bit more explanation needed.

We agree with the reviewer and would like to emphasise that this is exactly what we wanted to mention in the text. In response to this comment, it is true that this paragraph needed some clarification. To this end, we have added some further explanation and clarification to this paragraph (see below) (see between line 514 and line 527 of the revised manuscript).

«An interesting feature of Figure 3a is that some posterior parameter fields have fast travel times that exceed those calculated using prior parameter fields. This can be explained by the fact that some posterior parameter fields have lost some of the connectivity exhibited by the prior parameter fields. (see Appendix B). This is an outcome of the PESTPP-IES parameter adjustment process which is only truly Bayesian where prior parameter distributions are Gaussian on a cell-by cell basis (if cell-by-cell parameterisation is employed). However, the prior realisations that compose the initial ensemble from which the IES inversion process has started is not multiGaussian (see Appendix A). Neither the theory on which IES is based, nor numerical implementation of that theory in its history-matching algorithm, can guarantee the maintenance of long-distance hydraulic property connectedness which cannot be characterised by a multiGaussian distribution. Indeed, history-match-constrained adjustment of connected and categorical parameter fields is still an area of active research (Khambhammettu et al., 2020). Note that while uncertainty analysis methods such as rejection sampling or Markov Chain Monte Carlo approaches do not require a Gaussian prior nor a Gaussian likelihood function, these methods are impractical in contexts where the number of parameters is high and model run times are long, which is the case for the many hydrogeological applications and for the example used in this paper. »

<u>Reviewer #2</u>

The paper is well written and addresses an important topic in hydrologic systems analysis, namely robust model evaluation. The authors draw inspiration from the Kalman Filter and bring linear algebra, Tikhonov-regularized inversion and surrogate modeling to bear to decompose and/or approximate the forward model and quantify its prediction uncertainty using relatively few model simulations. This is yet another addition to the methods of the PEST toolbox designed to enable as thoroughly as possible the uncertainty quantification of highly parameterized and CPU-demanding surface-subsurface hydrologic models for which existing Monte Carlo simulation methods are too demanding and/or cumbersome to implement. One can only applaud the efforts of the authors, particularly the 2nd author, John Doherty, to provide workable solutions (with sufficient theoretical rigor) to practical, real-world, problems.

I do not have comments on the methodology. The assumptions are almost always listed and/or defined, and the mathematics (linear algebra) articulates the implementation. I only have one comment, which I think could help to further strengthen this paper.

Thank you very much for your kind words on our approach and our work in general. As reviewer #1, no methodological issues were identified.

The present case study is well chosen to illustrate the DSI methodology. But this case study is not easy to immediately repeat. I think the authors should consider including a relatively simple hydrologic modeling study which (a) is easy to reproduce and (b) most readers are familiar with. I believe that this may help articulate the detailed workings of the presented DSI methodology.

This case study does not have to be a distributed and/or computationally demanding modeling problem.

One may interpret this as a moderate revision but at the same time, I also understand if authors wish to publish their work as is.

Thank you for your comment, your time and consideration. While reviewer #1 suggests implementing a real case study you suggest to rather implement a much simpler model, so the suggestions are contrary. For the following reasons we feel that our model is at the sweetspot to effectively demonstrate our approach:

- The model is actually easy to reproduce and straightforward to implement. All boundary conditions are clearly described, the grid is easy to reproduce with the information provided in the paper. In any case, the input files are all provided, and can readily be used.
- We believe that the setting we are simulating is probably one of the most common settings in hydrogeology: A well next to a river. We do not see the benefit of having a simpler model. The model has to be tailored to provide a solid basis for verification and to demonstrate the usefulness of the approach.

One of the most important features of our approach is that it is capable of dealing with slow, and computationally very demanding models. We do not think that using a computationally non-demanding model is useful to demonstrate this feature. We have added a paragraph at the beginning of the application section to make this clear by discussing the veracity of our modelling example (see also our response to reviewer #1).

« The objective of this section is to demonstrate the performance and utility of DSI in quantifying posterior uncertainties of predictions made by a model whose run time is long and whose parameter field is complex. As is common in the literature, where the performance of a new method is tested and documented, we base our analyses on a synthetic model rather than on a real-world model. This allows us to assess, and document the performance of the method. It also dispenses with the need to account for epistemic uncertainties which accompany real-world modelling. »

An additional advantage of such a simple study is that the uncertainty of the DSI methodology can be benchmarked against Bayesian methods using a full exploration of the model's parameter space using MCMC simulation with/without the use of advanced distribution-free likelihood functions. This will readers to better understand the strengths and weaknesses of the presented methodology.

We do not see MCMC simulation directly comparable to DSI. DSI is independent of the number of model parameters, while the efficiency of MCMC is dependent on the number of parameters employed. We understand the point of verifying the methods against a more traditional method. This work has already been done by Sun and Durlofsky (2017), who compare the results of DSI with the rejection sampling procedure. We preferred to compare our DSI procedure against the Iterative Ensemble Smoother, which is a well established method that can be used with complex models and highly parameterised environments. This is the closest approach to DSI, and therefore suitable for a benchmark comparison.

As mentioned above, to some extent our work continues that of Sun and Durlofsky (2017). This continuation embodies use of the DSI surrogate model in conjunction with linear analysis tools that support data worth analysis at very little cost. The worth of data is judged by its ability to reduce the uncertainties of model predictions of interest. To be sure, linear analysis is approximate. Its strength, however, is that it does not require that values be assigned to posited observations, nor to the parameters that they may inform. Furthermore, it can be undertaken extremely quickly once a sensitivity matrix is available. We see a demonstration of this methodology using the DSI model as an important component of our paper. However, we see a comparison of the results of this analysis with MCMC (which is unable to handle enough parameters to characterise the heterogeneity of aquifers, and for which data worth assessment would need to be nonlinear) as well beyond the scope of our paper. This is especially the case where uncertainty is dominated by parameter nonuniqueness – a context in which the numerical cost of MCMC can be very high indeed. (As we point out in our paper, comparison with IES – a method that IS able to accommodate parameter nonuniqueness - was a numerically costly exercise.)

To address the reviewer comment, we have added clarifications in the introduction and complements one paragraph between lines 524 and 527 of the revised manuscript. (see below).

« Note that while uncertainty analysis methods such as rejection sampling or Markov Chain Monte Carlo approaches do not require a Gaussian prior nor a Gaussian likelihood function, these methods are impractical in contexts where the number of parameters is high and model run times are long, which is the case for the many hydrogeological applications and for the example used in this paper. »

One may interpret this as a moderate revision but at the same time, I also understand if authors wish to publish their work as is.

We thank the reviewer that he is not objecting to publication as it is. While reviewer #1 suggests a more complex model for a real case, reviewer #2 suggests a simpler model. For the reasons we outlined above, we believe that our choice of model is the most appropriate one in terms of demonstrating the capability of our approach. It is easy to reproduce, provides all the information required to assess the robustness of our approach, corresponds to a very common hydrogeological setting and allows us to demonstrate the very high performance of the proposed methodology.