Geoscientific Model Development
Reply to Anonymous Referee 2

"A Joint Reconstruction and Model Selection Approach for Large Scale Inverse Modeling"

by Malena Sabate Landman, Julianne Chung, Jiahua Jiang, Arvind K. Saibaba, and Scot M. Miller

September 25, 2024

We are grateful to the referee for their careful reading and thoughts on our manuscript. Below, we repeat the remarks and interleave our responses. All modifications in the revised manuscript are highlighted in **blue** and all references to equations, pages, lines, and citations correspond to the revised manuscript.

In this manuscript, the authors propose a modelling and algorithmic approach for performing model selection in large-scale inverse problems. The context is that a set of predictor variables (say, meteorological variables) are used in a linear model to predict an unknown quantity s (say, surface fluxes). The unknown quantity s is related to the observations through a forward model. Interest is on both estimating s and on identifying a small subset of the predictor variables that are most informative of s. To achieve this, the authors propose a sparsity-promoting prior on the coefficients of the predictors, and develop a hybrid iterative projection method based on flexible Krylov subspace methods for efficient optimization.

The paper is well written and the novelty and significance of the work is well justified. The paper is well organized and the figures are clear and informative. The authors perform extensive simulation studies to validate their method. My comments are mostly minor and relate to the (lack of) discussion of prior work in the statistical literature on model selection, as well as some minor points regarding the simulation studies.

My detailed comments are as follows:

1. The proposed model structure in (7) places a Laplace prior on β, the coefficients of the predictors. This is a common choice for promoting sparsity in the coefficients, which is often called the Bayesian LASSO (Park and Casella, 2008). It would be worthwhile for the authors to discuss this connection: indeed this present manuscript could be seen as the extension of the Bayesian LASSO into the inversion context. Park and Casella (2008) also address the problem of performing Markov chain Monte Carlo for this model so it may also be worth mentioning in the Conclusion where uncertainty quantification is discussed.

   We have clarified these connections and included the above reference in Section 3.1 and the conclusions.

2. Relatedly, there is literature discussing the implications of the choice of sparsity promoting prior. For example, Carvalho et al. (2010), in proposing an alternative sparsity promoting prior, discuss the limitations of the Laplace prior and other choices, and Piironen and Vehtari (2017) discuss ways to tune sparsity promoting priors. While an extensive discussion of these choices is beyond the scope of this paper, a brief mention of how these modelling choices can impact the results would be useful.

Thanks for these references. We have included these references in a discussion in Section 3.1.

3. I was a bit confused by the one-dimension simulation example in Section 4.1. Here are my questions, which I suggest the authors clarify in the manuscript:

   Thank you very much for pointing this out. There was indeed an error on the plots, which we understand made it confusing. This has been corrected now without affecting the message of the example, which we hope is clear now.

   (a) In Figures 2 and 3, what is the x-axis showing?

   For the 1D image deblurring example, the true signal and the observed blurred signal are defined on an equi-distant grid on $[0, 1]$ containing 100 points. A quadrature method is used to discretize the continuous linear inverse problem,

$$\int_0^1 h(s-t)f(t)dt = g(s), \quad 0 \leq s \leq 1.$$

   Thus, the x-axis denotes the domain where the signal and measurements are defined.

   We have now clarified this in the caption of Figure 2.

   (b) In Figure 2(b), can the authors show s and $X\beta$ as well as the blurred observation?

   We have now added a subfigure, Figure 2 (c), showing the true solution as well as its components.

   (c) In Figure 3(a), I was surprised to see that all methods noticeably overestimate the true function, particularly for the x-axis values between 0 and 0.5. Shouldn't the stochastic component, ζ, correct this? How does this error arise?

   Thank you very much for this comment. This was indeed an error on the plot, where we had plotted the mean rather than the full true solution. We have now updated all figures affected (Figure 3 and Figure 5).

   (d) What exactly is the "reconstruction error norm" and the "relative reconstruction error norm" shown in Figure 5 and mentioned in the text (also in the later simulation studies)?

   In Section 4.1.1, we have defined the relative reconstruction error norm and fixed all instances to be consistent.

4. In discussing the partial F-test in Section 2, it may be worth mentioning that the "smaller" model must nest within the "larger" model, a notable limitation of the F-test compared to competing methods.

   We have edited the corresponding paragraph of Sect. 2 to highlight this point.

5. Minor points:

   (a) Lines 419–420: the references Miller et al and Liu et al would be better parenthesised.

   Done.

(b) Lines 485–486: I'm surprised that the elements of Q have units (ppm)2—isn't this quantity a flux?

Thanks for catching this. It has been fixed.

References

- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. Biometrika, 97(2):465–480.
- Park, T. and Casella, G. (2008). The Bayesian Lasso. Journal of the American Statistical Association, 103(482):681–686.
- Piironen, J. and Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics, 11(2):5018–5051.