

# Geoscientific Model Development

## Reply to Referee 1 (Ian Enting)

“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 Professor Enting for his careful reading and thoughts on our manuscript. Below, we repeat his 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.

**Overview** This paper describes an inversion technique where the regularisation is based by prior predictions from models, with the new feature that selection of an appropriate subset of models is being selected as part of the inversion. It defines an appropriate objective function, and a new algorithm for minimising this objective function.

The technique is then illustrated by three synthetic data examples, firstly a signal processing example and then two inversions of simulated OCO-2 data.

The paper is generally well written, although there are a few points at which a little more detail might make it easier to read. It might benefit from a table of notation — although what appears below was produced for my own benefit while writing this report.

To conclude, this paper is suitable for publication in Geoscientific Model Development. My various comments should be regarded as suggestions for the authors to consider and the editor to take into account, rather than being prescriptive.

### General comments

- Probably it should be explicitly stated that the formalism as presented here is restricted to linear (or linearised) models.

We have included “for linear inverse modeling” in the abstract and “Linear” in the title.

- As the authors note, model selection (effectively choosing a set of regression variable from a set of  $p$  candidate variables) would strictly involve  $2^p$  comparisons, and iterative approaches, e.g. successive rejection, still require large numbers of comparisons (and  $\binom{p}{q}$  comparisons if  $q$ , the size of the target subset is pre-determined). The approach here is analogous to the use of L1 norms in regression, fitting a subset of items closely at the expense of downplaying the lack of fit of other items (and thus being less sensitive to outliers).

Yes, that is correct. We have added a remark to this effect in Section 3.1.

- Choosing one ‘best’ model from a set of models describing the same process seems an appropriate approach. If one has two or more candidate models of different processes that give similar contributions

to  $z$ , then rather than choose one model, the appropriate conclusion is that the inversion can estimate a linear combination of these models but not distinguish between them. This could be particularly relevant for the third example, where 'lumping' of poorly distinguished weak source regions might be more appropriate than selecting a subset. The example is a good illustration of the power of the method but the method may not be the best way of inverting actual OCO-2 data.

Thank you for this comment. We agree that the third example provides a nice illustration of the msHyBR method's ability to perform model selection in a one-step approach. In Section 4.3, we emphasize that this example highlights the applicability of our approach to very large datasets, although other approaches or additional analysis may be needed for actual data.

### Line by line

83 Suggest using  $q$ , as above, rather than  $s$ , as number of target models to avoid confusion with  $s$ .

Great suggestion. We have changed  $s$  to  $q$ .

various Suggest upright  $\mathbf{T}$  to indicate matrix-vector-transpose, since the  $\mathbf{T}$  is not a mathematical variable. There are various online discussions on best form. My own preference would be  $\mathbf{X}^T$

Done. All transposes have been fixed.

various The importance of this will depend on the fonts used for final production, but in the discussion preprint, the use of bold font to distinguish vectors works poorly for the Greek beta and particularly poorly for gamma. While the intended usage can be deduced from the context, it makes the article a little bit harder to read, since beta does not always appear as a vector (see eqn 7) and both forms are subscripted, once by component index within the vector, and otherwise with subscript to the vector indicating the iteration number.

We are using the Copernicus Publications Manuscript Preparation Template for LaTeX Submissions, so it is likely that the fonts will remain. To help with clarity, we changed all  $\boldsymbol{\gamma}$  to  $\mathbf{g}$ . We also noted that  $\gamma$  was being used to refer to the noise level.

Regarding  $\boldsymbol{\beta}$ , we decided to keep this notation to match the model selection literature. However, we clarify that  $\beta_j$  corresponds to the components of the vector  $\boldsymbol{\beta}$  on pages 7 and 10.

259 Strictly,  $t_{k+1,k}$  and  $m_{k+1,k}$  seem not to be defined at this point. Needs forward reference to Algorithm 1.

We have clarified that these are normalization scalars and include a forward reference to the Algorithm. We also clarify that these are the elements of the upper Hessenberg matrix  $\mathbf{M}$  and upper triangular matrix  $\mathbf{T}$  respectively.

### Terms where a few words of description might help

- Krylov subspace methods - Methods for working with subspaces of large problems
- Laplace distribution - Also known as double exponential.

In Section 3.2 (the first instance of the word Krylov except for the abstract), we added the statement, “Solving such optimization problems can be challenging, and Krylov subspace methods, which are iterative approaches based on Krylov subspace projections, are ideal for working in subspaces for large-scale problems.”

In Section 3.1, we clarified that the Laplace distribution is also known as the double exponential.

### Notation

- $k$  Iteration count in inversion algorithm. Dimension of the Krylov subspace at that step??  $m$  Number of observations (dimensionality of  $z$ ).
- $n$  Number of quantities to be estimated (dimensionality of  $s$ ).
- $p$  Number of candidate predictors (models).
- $z$  Vector of observations that are to be fitted by the inversion.
- $s$  Vector of quantities to be estimated by the inversion.
- $\zeta$  ‘Random’ component of  $s$ , assumed to be distributed as zero mean multivariate Gaussian described by  $Q$ .
- $\beta$  Vector of contributions from each of the candidate models, where model selection corresponds to taking zero as the estimate of particular components. The selection is ‘controlled’ by the regularisation parameter  $\alpha$ .
- $\epsilon$  Error component of  $z$ . Assumed to be zero mean multivariate Gaussian described by  $R$ .
- $R$  Covariance matrix for  $\epsilon$ , the error component of observations,  $z$ .  $H$  Mapping from  $s$  to observations  $z$ .
- $\lambda$  Regularisation parameter, estimated as part of the inversion, applied as scale factor of  $Q$ .
- $\alpha$  Regularisation parameter, estimated in the inversion, applied  $Y$  as defining the scale of the Laplace distribution of  $\beta$ .
- $\gamma$  Transformation such that minimisation over  $s$ ,  $\beta$  becomes minimisation over  $\gamma$ ,  $\beta$ , avoiding the need to invert  $Q$ .
- $\mathbf{u}_k, \mathbf{v}_k$  Additions to the solution subspace after step  $k$ .
- $X$  Mapping from candidate models,  $\beta$ , to  $s$ .
- $j$  Component index within a vector.

Thank you for this suggestion and for your list of terms. We agree that such a table can be very useful and may help with any confusion. Thus, in the revision we have included a list in Appendix A that builds on your above list. Thanks!