Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-41-AC1, 2016

Geoscientific Model Development Discussions

Interactive comment on "Estimation of trace gas fluxes with objectively determined basis functions using reversible jump Markov chain Monte Carlo" *by* Mark F. Lunt et al.

Mark F. Lunt et al.

mark.lunt@bristol.ac.uk

Received and published: 20 July 2016

We thank the anonymous reviewer for their comments on the manuscript. We have replied to each of the specific comments in turn below, with the author's response following each comment. Page and line numbers given in the author response refer to the marked up version of the manuscript, provided as a supplement to this comment.

1. The authors have restricted the dimensionality problem to the 2D space, and seem to have forgotten that the problem has a temporal dimension as well. This question should be addressed somehow right from the start of the paper and in the pseudo-data example.

Author's response: We agree with the reviewer that the aggregation of the temporal dimension is an important component of inverse problems, and indeed, it is something we considered carefully. However, in the manuscript we intentionally focused only on the spatial domain for the following reasons:

- There are a limited number of examples of three-dimensional transdimensional inversion in the geoscientific literature, although to our knowledge none of these involve a temporal dimension explicitly. However, one study (Piana Agostinetti et al. 2015) that used three dimensional Voronoi cells reported a computation time for the reversible jump algorithm of approximately one month, given 9700 data points and 10^6 iterations of the Markov chain. Therefore, were we to explore the aggregation across three dimensions we would anticipate that a vastly more efficient procedure might be required, perhaps involving an alternative approach to Voronoi cells (e.g. Hawkins and Sambridge 2015). This would be a substantial undertaking which we believe could form an entirely new work of itself.

- The aggregation of the spatial basis functions into Voronoi cells relies on calculating the Euclidian distance between each grid cell and Voronoi nucleus. However, if this was extended to a space-time domain, it is not immediately obvious how one would calculate equivalent "distances" in space and time. One solution might be to normalise these distances, thus allowing the same Voronoi tessellation to be used across three dimensions. However, we decided that such an extension would lead to further complication and again, would be better tackled in a future paper.

In light of this comment we have made the following additions to the main text, which highlights that we have focused only on the spatial part of the problem, and we have included a discussion of how the reversible jump algorithm could be applied to the temporal aggregation of emissions in the final section.

Page 1, Line 4: "Here, we present an objective method for reducing the spatial dimension of the parameters space. . ."

Page 2, Line 10: “In addition to the spatial partitioning, some form of temporal aggregation must also be performed, over which the parameters are assumed constant. Each basis function then represents some 3-D aggregation of the underlying fluxes. In this work we choose to focus only on the 2D spatial component of emissions, making the assumption that the fluxes are constant over a fixed period of time.”

Page 23, Line 11: “In this work, we intentionally chose to focus only on the 2-D spatial aggregation of the fluxes and ignored the temporal dimension in this work due, primarily, to concerns about the computational demands of extending this particular implementation to 3-D (Piana Agostinetti et al. 2015). However, there is no inherent reason that the transdimensional approach could not be further extended to the 3-D problem. Such an extension would inevitably incur higher computational expense, particularly with the frequent need to recalculate 3-D Voronoi cells. It may be possible to ameliorate these demands by prescribing an alternative form of basis function such as a tree structure similar to Bocquet et al. (2011), which may be both faster to calculate and more efficient at exploring the 3-D parameter space (e.g. Hawkins and Sambridge 2015).”

2. P. 2, l. 10: the correlation is on emission errors, not on emissions.

Author’s response: We thank the reviewer for pointing out this error and have corrected this in the text.

3. P. 2, l. 15: “that do not exist in the true field” actually applies to any flux estimate, since it remains uncertain.

Author’s response: We agree with this point and for the removal of doubt have removed this comment.

4. P. 4, l. 22: to be fair, the authors should also cite earlier publications like Michalak et al. (2005, doi : 10.1029/2005JD005970), Berchet et al. (2013, doi:10.5194/acp-13-7115-2013) or Wu et al. (2013, doi :10.3402/tellusb.v65i0.20894).

Author’s response: We acknowledge that our overview of previous studies in this field came across as unfairly brief. The point that we wished to make was that, in a similar vein to this work, Ganesan et al. (2014) treated the solving of hyperparameters as a single-step problem alongside the estimation of emissions in an MCMC framework. The work previous to this, as mentioned by the reviewer, considered the problem as a two-step process, that we believe leads to some difficulties in accurately apportioning uncertainties in a Bayesian framework. We accept that the way this was written made it appear as if only Ganesan et al. (2014) had addressed this problem which was not our intention. We have rewritten this paragraph from page 4 line 31 and extended it to encompass a review of previous work as follows:

“In addition to being dependent on the partitioning of basis functions, Bayesian inversions are also dependent on the form of the PDFs used to describe the prior and likelihood. The terms that describe these PDFs such as the mean, standard deviation and correlation length are commonly referred to as hyperparameters. The dependence of the posterior parameters on these hyperparameters, and a lack of objective determination of their values have been previously identified as a limitation of Bayesian inverse methods (e.g. Rayner et al.,1999). There have since been a number of studies that have proposed methods for determining hyperparameter values using the data (e.g. Michalak et al., 2005, Berchet et al., 2013, Wu et al., 2013). In general, these methods rely on Gaussian assumptions and are performed in a two-step process whereby the hyperparameters are first optimised, and then parameter inference is performed based on these optimal values. Winiarak et al. (2012) also extended this to a semi-Gaussian prior PDF, such that the source term was constrained to be positive. However, as noted by Berchet et al. (2015), one issue is that the uncertainty in the specification of the hyperparameters in step one is not included in the second step. Ganesan et al. (2014) presented an alternative method, where the hyperparameters and parameters were estimated simultaneously using an MCMC algorithm. This framework explored the “uncertainties in the uncertainty", resulting in a more complete characterization of the uncertainty in the posterior parameters. The framework also has the advantage

that the data is used only once, thus remaining strictly Bayesian, and PDFs are able to take forms other than Gaussian. In the transdimensional case, the posterior distribution of the number of unknowns can be heavily dependent on the prescribed uncertainties (Bodin et al. 2012). As such, it is important to incorporate data driven hyperparameters into the transdimensional inversion, if the derived number of unknowns is to be truly dependent on the data."

5. P. 5, l. 25: "its".

Author's response: We thank the reviewer for pointing out this spelling mistake.

6. P. 6, l. 7: "unintelligent" sounds harsh and a softer word would be more appropriate

Author's response: We have replaced "unintelligent" with . . . "unrefined".

7. P. 7, l. 7: the authors should also refer to earlier studies.

Author's response: On reflection this sentence seems redundant, and we have significantly edited this passage in response to the comment below and those of reviewer 2 (see below).

8. P. 7, l. 10: the validity of this assumption should be discussed. At first glance, it looks poor. For instance a large dependency in the case of natural CO2 fluxes over land was shown by Chevallier et al. (2012, doi:10.1029/2010GB003974, their Fig. 5). More generally for instance, it is very likely that hyperparameters are not the same at city-scale and at country-scale.

Author's response: There was an omission in the original manuscript that neglected to mention that this assumption only applies to the dimension of the hyperparameters describing the data, θy. The hyperparameters describing the prior parameters PDF, θx, are also dependent on the dimension of the basis functions (i.e. each basis function is described by its own hyperparameters). However, the hyperparameters describing the model-measurement covariance structure are still independent of the number of basis functions, since the prior and data must maintain independence to be strictly

Bayesian. We have rewritten Eq. (8) to account for the fact that there are two distinct sets of hyperparameters, those describing the prior emissions error structure and those describing the model-measurement error structure. We have edited the text in this passage from page 8, line 8, so that it now reads:

"In addition to m and k we also wish to solve for the set of hyperparameters that describe the prior parameters PDF, θx, and the likelihood PDF, θy. The dimension of the latter can be assumed independent of k since it is a property of the data. However, we prescribe the dimension of the emissions hyperparameters θx, to be dependent on k, alongside the parameters. The full form of the transdimensional, hierarchical Bayesian equation then becomes:"

9. P. 9, l. 14: how can the prior location and emissions variables be independent of each other?

Author's response: The prior probability of a Voronoi nucleus occupying any particular position on the grid is dependent only on the total number of Voronoi nuclei, since each nucleus cannot occupy a grid cell that is already occupied. A priori, we assume that there is an equal probability of choosing each grid cell as a nucleus location, and hence this is independent of the emissions. While the magnitude of emissions within each Voronoi cell will be dependent on its location, the prior scaling of this magnitude is independent of the value within it, and thus the location. A priori the scaling of the prior is the same throughout the spatial domain and so the location and emission variables are independent. We note that this condition is only met a priori, and thus significant correlations might be expected on the introduction of the data, and therefore in the posterior distribution. We believe that a reordering of statements may help explain this independence, and have changed the text accordingly. We first stipulate that a uniform distribution is assumed for the location of each nucleus, and that we are solving for a scaling of the underlying emissions distribution. In response to this comment, and a similar one from reviewer 2 we have included the following on page 10, line 21:

"If the emissions value is taken to be some scaling of a prior distribution of emissions then the a priori scaling of the prior emissions field should be one everywhere, and hence this is not dependent on location. In this work we assume a uniform distribution for the location of the Voronoi nuclei, meaning that the prior distribution is independent of the emissions. Given this independence of the variables, the term p(m|k) can be decomposed into two terms expressed as:"

10. P. 13, l. 7 and l. 9: why is there a notion of convergence (l. 9; like if we were looking for just the most-likely state) while the algorithm explores the space of the posterior pdf (l. 7 and 24)?

Author's response: The notion of convergence refers to the convergence in our exploration of the posterior PDF, rather than convergence to a point. The chain starts from one distribution (the prior) and on the introduction of the data moves to another distribution (the posterior). While the individual iterations of the chain will continue to explore the parameter space, the posterior distribution itself should be stationary in order for convergence to be said to have occurred. In light of this comment we have attempted to be clearer about our definition of convergence in the text. Page 16, Line 12 now reads:

"The chain must be run for a sufficient number of iterations in order for convergence of the posterior distribution to occur. The convergence refers to the stability of the distribution across the sampled iterations of the Markov chain."

11. P. 13, l. 19: what does "typically" mean here?

Author's response: Typically here is possibly a misnomer since it will be somewhat dependent on the acceptance rate of the dimension changing proposals in particular. Previous examples of transdimensional inversions in the geosciences (e.g. Bodin and Sambridge 2009, Ray and Key, 2012) have reported running for around one million iterations, although the acceptance rates were fairly low in these studies. Our acceptance rates for the dimension changes are around 30%, although, as mentioned in the

manuscript, this may be due to limited constraint over areas of low emissions such as the sea. The key is that it is important to run the chain for a sufficient number of iterations such that it returns a meaningful stationary estimate of the posterior distribution. Our own tests have shown that O(10^5) iterations are required in order to achieve a robust estimate of the posterior distribution in the particular problems we have attempted. We have altered page 16, line 23 to be more explicit about our meaning:

"In order to achieve a stationary posterior distribution for the parameters, the number of iterations for which the chain. . ."

12. P. 13, l. 27: why should the solution of the problem (independent of the resolution method) be smooth? In other words, is it an advantage or an inconvenient to generate a smooth solution?

Author's response: The word "smooth" was used to mean that the mean of the posterior distribution can provide a spatial distribution that is at a higher resolution than the coarser basis function partitioning at each individual iteration. In this work, since we limited the shape of the Voronoi cells to follow the underlying NAME output grid, the smooth solution is at the resolution of this grid, and therefore still discretized. For inference on national scale fluxes a smooth solution such as this may not be necessary. However, we believe that for the regional or spatial attribution of emissions then a smooth solution is an advantage, since the derived spatial patterns are not dependent on a single partitioning of the basis functions. In light of this comment we have attempted to be clearer about what naturally smoothed means on page 17, line 1:

"a naturally smoothed solution, (i.e. at the resolution of the underlying finite grid) without the need to specify. . ."

13. P. 15, l. 10: is "twice as small" significant here?

Author's response: We do not think that the approximate factor of two is significant per se. It is simply the fact that the RMSE is smaller that is significant. We have edited this

sentence to reflect this point: “The RMSE value of 1.0 ppb was smaller (approximately a half) than that of the subjectively determined grid, for this particular pseudo-data example.”

14. P. 17, l. 8: how stable is the estimate with respect to the number of iterations?

Author’s response: The UK total is stable with respect to the number of iterations after the 100,000 iteration burn-in period, showing how this distribution has converged. Given the thinning of the chain, for each 100,000 iterations 1000 samples are stored. 500,000 iterations were chosen to allow for the 5 different proposal types at each iteration and after thinning of the chain, the posterior distribution is then estimated by 5000 samples. The first 1250 return a mean UK estimate of 2.27 (2.04-2.48) Tg/yr, the second quarter 2.25 (2.07-2.47) Tg/yr, the third 2.30 (2.04-2.57) Tg/yr and the final 1250 iterations have a mean of 2.27 (2.02-2.53) Tg/yr. This shows how relatively stationary the distribution is with respect to the number of iterations. If this were not the case then either a longer burn-in period may be required, or a change to the proposal jump sizes to allow for more efficient exploration of the chain. In light of this comment we have added the following line to page 21, line 21:

“The UK and Ireland estimates were found to be stable with respect to the number of iterations from which the posterior distribution was sampled. This shows that the burn-in period was sufficient for convergence of these national scale emission totals to occur.”

15. P. 18: the first paragraph on the page reminds of the discussion by Berchet et al. (2013, doi:10.5194/acp-13-7115-2013, their sections 3.1, 3.2 and 3.3) on the same topic. This may be acknowledged.

Author’s response: In response to a point from reviewer 2 we have added the following section, which also makes reference to Berchet et al. (2013) on page 22, line 9:

“No significant difference was found between the uncertainties derived for times when

local influence was high and those when it was not. By contrast, Berchet et al. (2013) reported CH4 observation uncertainties that were on average 23-31% smaller during the day than at night for a number of sites across Europe using three different hyperparameter optimization schemes. There are known errors in boundary layer modelling that are likely to be greater at night, although these may be more systematic than random. A better understanding of modelling uncertainties, and how they can be accounted for in the hierarchical framework would be necessary to include this potential bias.”

16. P. 18, l. 11: “To avoid this, the. . .”

Author’s response: We thank the reviewer for pointing this out and have changed it accordingly in the text.

17. P. 19, l. 6: the sentence is too trivial to be the last one.

Author’s response: We have removed this sentence and replaced it with the following: “The framework provides an alternative approach to using a single partitioning of basis functions when performing dimension reduction.”

References

Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Conil, S., Geever, M., Laurila, T., Lavric, J., Lopez, M., Moncrieff, J., Necki, J., Ramonet, M., Schmidt, M., Steinbacher, M., and Tarniewicz, J.: Towards better error statistics for atmospheric inversions of methane surface fluxes, Atmos Chem Phys, 13, 7115-7132, 2013.

Bodin, T. and Sambridge, M.: Seismic tomography with the reversible jump algorithm, Geophys J Int, 178, 1411-1436, 2009.

Ganesan, A. L., Rigby, M., Zammit-Mangion, A., Manning, A. J., Prinn, R. G., Fraser, P. J., Harth, C. M., Kim, K. R., Krummel, P. B., Li, S., Mühle, J., O’Doherty, S. J., Park, S., Salameh, P. K., Steele, L. P., and Weiss, R. F.: Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods, Atmos. Chem.

Phys., 14, 3855-3864, 2014.

Hawkins, R. and Sambridge, M.: Geophysical imaging using trans-dimensional trees, Geophys J Int, 203, 972-1000, 2015.

Piana Agostinetti, N., Giacomuzzi, G., and Malinverno, A.: Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, Geophys J Int, 201, 1598-1617, 2015. Ray, A. and Key, K.: Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm, Geophys J Int, 191, 1135-1151, 2012.

Please also note the supplement to this comment:
http://www.geosci-model-dev-discuss.net/gmd-2016-41/gmd-2016-41-AC1-supplement.pdf

---

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-41, 2016.