

Response to interactive comment on
“Reconciling differences in stratospheric ozone composites”
by William T. Ball et al.,
Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2017-142-RC2, 2017

General response

We thank both referees for their thorough consideration and constructive feedback. As a result of the review process, we have made a significant effort to improve semantics regarding methods, models and algorithms. In the revised version of the manuscript we have replaced the particle filtering method by a Hamiltonian Monte Carlo (HMC) approach to sample the full posterior distribution, conditioned on the full data vector as is required, rather than just the data up to time t , as in the particle filter. HMC is well documented in the literature, and as such, the length section on the particle filter has been significantly reduced. We re-did the sampling algorithm from scratch, re-ran everything and re-made all relevant plots. In practice, we found that the new results are broadly similar to the particle filtering results and none of the key findings are changed. We now refer to the composite constructed using Gaussian-mixture likelihood and transition prior, with SVD uncertainty estimates, as the BAYesian Integrated and Consolidated (BASIC) composite.

We note that there are some differences that you should be made aware of compared to the previous version. These include:

- the time-dependent error bars are much tighter, and much closer to Gaussian than before; this is good because the DLM analysis will better represent the data with Gaussian errors included;
- the problem we found (in only a few limited regions) following Pinatubo has gone and BASIC performs well during this period, given the data supplied to the process;
- there is a longer section in the appendix that encompasses requests for information on the Gaussian-mixture likelihood construction (in BASIC), and the parameter estimation (in the DLM);
- Northern and southern hemispheres in the profiles were actually the wrong way around; given the symmetry between hemispheres, the conclusions do not change.

We reply to all comments below, with referee comments in black, and our responses in blue.

Marko Laine (Referee)

Received and published on 20 April 2017

The PF method is presented like a model, but in fact it is a numerical algorithm

For example

line 17: "Particle filtering and DLM",

line 20: "The particle filter results",

line 779: "using a particle filter",

line 804 "the particle filter as a method".

In my opinion, the distinction between a model for data and a numerical algorithm should be made more clear. You should first describe the model (your dynamical mixture-Gaussian model as a Bayesian hierarchical model) behind the data merge and then the numerical Monte Carlo filtering algorithm (PF/SIR) for actually estimating the merged data set.

PF (or SIR) is a numerical method of computing a certain Monte Carlo estimate of a posterior (predictive) distribution in a dynamical model. You propagate an ensemble ensemble of possible

model states (time series realizations) by a model (here the assumed month-to-month seasonal change and known deficiencies) to produce prior ensemble for the next state, which is then weighted by the likelihood function defined by the observed satellite composites. This will form a sample of the posterior uncertainty of the merged series given the observation up to the current time point. In effect this is a non-linear, non-Gaussian generalization of a Kalman filter.

You could contrast this to DLM or MLR "methods". DLM (and MLR, too) is a model for the processes and the system generating the observations (see below for a general state space description. DLM is a structural state space model that constructs a time series from basic building blocks, like trend, seasonality and proxies. For DLM one can use Kalman filter and smoother as an estimation algorithms. For MLR you can use the least squares algorithm for estimation, but other algorithms are available, as well.

We agree that the semantics regarding methods, models and algorithms needed cleaning up. In the updated version we refer to the composite constructed using Gaussian-mixture likelihood and transition prior, with SVD uncertainty estimates, as the BAYesian Integrated and Consolidated (BASIC) composite, and refer elsewhere to specific methods and algorithms appropriately.

SVD for uncertainty estimates

A similar comment is valid for the SVD "method" for construction of uncertainty estimates for the individual composites. SVD is an algorithm for a certain matrix decomposition. For the uncertainty analysis, you will have a some kind of model based on principle components and then you use the SVD algorithm for estimating the components. Is there any references the "SVD" approach used? I think the approach would need more motivation. You could write a model for the sources of uncertainties for each composite, having a common source and other sources that might be instrument specific. Then you could estimate these by principle components. As an example, a model for composite d_i would be $d_i = p_i T = p_{1i} T_1 + p_{2i} T_2 + p_{3i} T_3 + p_{4i} T_4$, where T are the principle components and p are the corresponding loadings. Then use it to build a model for variance components of a composite d_i , as $\text{var}(d_i) = \dots$, that would include the composite uncertainty as one of the components.

We have updated the discussion of the uncertainty estimation in Section 3.1 to give more clarity about the role of the SVD (essentially as a numerical method to implement PCA) and also to more fully explain our heuristic error estimation. There are references to use of SVD, but none we are aware of in the form we have put forward here. There are no references for this method itself, but we have done empirical tests to show that it produces sensible results for reasonably discrepant data-sets such as those being analysed here. That said, we explicitly state that more work is needed on this aspect of the overall data analysis problem and we fully expect to attempt this in future papers.

Filter vs. smoother

You should motivate why "filtering" is adequate for the data merge and no "smoothing" is needed. A filter calculates $p(y_t | \{d_t\})$ for each $t = 1 : T$, but not $p(y_t | \{d_{1:T}\})$ nor the joint distribution $p(y_{1:T} | \{d_{1:T}\})$. The latter are what are estimated by Kalman smoother in DLM calculations for a linear state space model.

Additional question: could PF be replaced by suitable weighted average of the composites, that just takes into account the prior information about problems in the individual series? In DLM and MLR

you will need to assume Gaussian uncertainty, so the PF results need to be summarized as mean and standard deviation.

What are the benefits of PF over some simpler (non-Monte Carlo) averaging method?

Thanks for pointing this out. You are correct that the particle filtering algorithm samples from the posterior distribution of the true time series conditioned on the data "up to that point", rather than the full data vector, and conditioning on the full data vector would require a subsequent smoothing step. The smoothing step comes with some considerable technical difficulties - to get around this whole issue, in the new version we have abandoned the particle filtering method completely and resorted to Hamiltonian Monte Carlo (HMC) sampling to sample the full posterior distribution, conditioned on the full data vector as is required. HMC is well suited to ultra-high dimensional sampling problems and is well documented in the literature. We re-coded the sampling algorithm from scratch, re-ran everything and re-made all relevant plots. In practice, we found that the new results (now correctly conditioned on the full data vector) are broadly similar to the particle filtering results and none of the key findings are changed. Nonetheless we thank you again for pointing this out and the new approach is now correct and more robust.

Regarding to what extent our method is akin to performing a weighted-average of the composites: for sure, some of the data-artefacts will be reduced/removed by taking an inverse-variance weighted mean, and for a lot less effort. However, use of the (fat-tailed) Gaussian-mixture likelihood combined with the month-to-month transition prior allows our approach to identify where certain data are corrupted without a priori knowledge of specific issues — these (many) cases cannot be captured by simply averaging. We also provide non-Gaussian uncertainties; it's true that DLM/MLR assumes Gaussian errors, but we encourage extension of these tools to allow for non-Gaussian uncertainties and/or marginalization over a full systematics model - this is a first step on a long road to a more principled approach to trend analysis from ozone data.

About MCMC

I would like to see some MCMC results for the DLM analysis. You are using uniform priors for the variance parameters (line 689). Do these parameter identify, especially, if you assume unconstrained smoothness for the trend?

How do the AR parameters identify?

You use uniform $[-1, 1]$ for the AR parameter, but do you consider negative autoregression as a realistic model for an ozone observation time series?

You could include some plots of the posterior distributions.

Thanks for raising these issues. Over the course of the work we experimented with different prior assumptions for the DLM. We found that in some cases leaving the "smoothness of the trend" parameter σ_{trend} unconstrained leads to a wiggly "trend" that captures all of the variability (with enough burn-in), and in the most recent version we use a half-Gaussian prior on σ_{trend} with variance $5e-4$. The other parameters are left with improper uniform priors, and the AR correlation coefficient prior is updated to being uniform on $[0, 1]$ rather than $[-1, 1]$ — we agree that negative AR correlations are difficult to justify physically (although the strictly positive prior made little/no different in practice). In tests on simulated data we find that all hyper-parameters identify well under these priors - we have included new plots of the parameter posteriors in the appendix.

General state space model approach

I suggest that you describe the merge and trend analyses as a general hierarchical state space model. In both merging the data and in the DLM analysis you are dealing with a dynamical state space model. A general framework to describe the statistical model is by a hierarchical description, with a process model for the model state dynamics, a parameter model for model (nuisance) parameters and a data model for the likelihood. The Bayes formula would provide the posterior estimate from the individual conditional components as (see [1,2,3]):

$$[\text{process, parameters} | \text{data}] \propto [\text{data} | \text{process, parameters}] [\text{process} | \text{parameters}] [\text{parameters}]$$

Filtering and smoothing algorithms can be used to estimate various marginal and conditional posterior distributions. The nuisance parameter could be integrated out by MCMC, for example.

For ozone data merge the process model includes the month-to-month variability and external events like volcanos, trends etc. The observation processes could describe the instrument effects. Lastly, there is the prior distributions for model parameters. The whole will in effect be a hierarchical Bayesian model to describe and estimate the state together with the parameters. This could provide a common framework for both merging and analysing.

We agree that the merge and trend analyses should really be done simultaneously in a single Bayesian hierarchical model (BHM). We have an on-going project where we are developing a sophisticated BHM for analyzing ozone data from scratch, going back to the original instrument records and modeling systematics explicitly rather than attempting to merge already-merged composites. However, this is well beyond the scope of the current paper, although it is a first step that resolves some of the key data-issues and is a coarse approximation to the full BHM approach.

A related issue that some readers have raised is concern over “using the data twice” — once to construct the transition prior (and uncertainty estimates) and once again in the main analysis (i.e. posterior sampling). Estimating the prior hyper-parameters and uncertainties a priori and fixing them can really be seen as approximation to the full BHM solution.

To cover these issues, we’ve added a section titled “BASIC as an approximation to a Bayesian hierarchical state-space model” where we briefly describe the full BHM approach and make explicit the fact that pre-computing the transition prior and uncertainties is an approximation to the BHM approach, which is good in the fortuitous case where those pre-computed quantities are strongly constrained by the data and do not strongly co-vary with the parameters of interest.

[1] L. M. Berliner. Physical-statistical modeling in geophysics. Journal of Geophysical Research: Atmospheres, 108(D24):8776, 2003. doi: 10.1029/2002JD002865.

[2] C. K. Wikle and M. B. Hooten. A general science-based framework for dynamical spatio-temporal models. TEST, 19(3):417–451, 2010. doi: 10.1007/s11749-010-0209-z.

[3] N. Cressie and C. K. Wikle. Statistics for Spatio-Temporal Data. Wiley, 2011.

Other comments

line 385, equation (2): I do not see how the parameters γ and β give rise to bimodality for an individual composite as the mean is the same $d^{\wedge}c_t$ for both modes. It probably will make the tails of the likelihood heavier than for a standard Gaussian likelihood.

The heavier tails of the likelihoods for individual data points leads to enhanced bi-modality when these likelihoods are multiplied together. See the new Figure. A3 and comment/response to reviewer 1 where the figure has been included there too.

line 460: The PF distribution is said not to be Gaussian but in DLM and MLR you need Gaussian uncertainty. Is this a problem for the trend analysis?

A “most principled” and optimal trend analysis will consider full non-Gaussian uncertainties. We are in the process of developing extensions to DLM that can deal with non-linear models and non-Gaussian likelihoods - however, this is well beyond the scope of this work. It’s difficult to assess quantitatively to what extent the Gaussian assumption biases the trend analysis without knowing the “right answer” accounting for non-Gaussian errors. However, we feel that the impact of non-Gaussian errors is one of a large number of remaining deficiencies in trend analyses performed in the community, such as the linear-model assumption, fixed regressor phases etc. It is very likely not the biggest evil in this basket of remaining issues.

line 801: "using the same instrument dataset more than once". The transition prior is inferred from the same observations that are used in the model, so the data is used twice. Also, the uncertainty is inferred from the same data by SVD. Maybe this is ok here, but it violates the Bayesian assumptions.

See discussion above under “General state-space model approach” — pre-estimation of the transition prior and uncertainties can be thought of as an approximation to the full Bayesian hierarchical model. We leave the full hierarchical treatment to future work.

Can you elaborate more the claim that PF method can resolve the problems in data merging? Do you claim that PF is capable to extract the background truth behind different biased estimates. Or does it just make the "error bars" larger, so that the trend analysis is not affected by instrument artefacts?

The heavy-tailed Gaussian-mixture likelihood combined with the transition prior is able to identify where one or more datasets are biased, and result in a posterior whose mean is un/less-biased without necessarily ballooning the error bar. This can be seen from the fact that the product of Gaussian-mixture likelihoods can result in a multi-model joint-likelihood where the widths of the individual modes are not expanded as much as for a product of normal Gaussians. If the multiplication of the transition prior then excludes one of these modes, the resulting posterior effectively rejects the data in the excluded mode and what is left does not necessarily have an inflated uncertainty.

I agree that construction of a merged data set is of interest in itself. For trend analysis one could start from individual observations. You could discuss the possibility of a general data fusion approach that assimilates all the different composites or individual retrievals to a common time series model. You might still be able to use linear model, but with carefully designed (linear) observation operator, that would account for the instrument artefacts. Or use some non-linear generalization of DLM.

As discussed earlier, we completely agree that this is the way forward and have an exciting on-going project concerned with exactly this problem, but we feel it’s beyond the scope of the current paper.

Conclusion

I can recommend the article to be published, if the author formulate the modelling approach for merging and uncertainty estimation a little more consistently, motivate the adequacy of the filter in the data merge and the use of SVD for the uncertainty variance components, and describe the MCMC results for DLM.