

Atmos. Meas. Tech. Discuss., referee comment RC1
<https://doi.org/10.5194/amt-2021-51-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2021-51

Anonymous Referee #1

Referee comment on "Revisiting matrix-based inversion of scanning mobility particle sizer (SMPS) and humidified tandem differential mobility analyzer (HTDMA) data" by Markus D. Petters, Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2021-51-RC1>, 2021

This manuscript presents a software package to invert aerosol size distributions from measurements, in particular from scanning mobility particle sizers (SMPSs), using the Tikhonov regularization approach.

This manuscript sits at the intersection of producing open-source, scientific code and presenting new scientific ideas. The reviewer admits this is an awkward position, as existing dissemination methods are not amendable to publishing well-maintained software, which is a critical component to modern, reproducible analysis. That being said, the scientific contributions of the underlying code are not hugely significant. Inversion of aerosol distributions using Tikhonov regularization is well-established. In fact, the author already notes one other instance of open-source software designed - at least in part - for this task (Hansen). (Other codes undoubtedly exist, though, admittedly most of these codes are closed source or not immediately available to the user, with very few exceptions, as the author notes.) Otherwise, this code does little to innovate on existing methods and is somewhat behind in terms of state-of-the-art, such as not presenting any form of uncertainty quantification - see Kandlikar and Ramachandran (1999); Voutilainen et al., (2001); and Voutilainen, Kolehmainen, Stratmann, & Kaipio (2001). The use of a GSVD to speed computation is insightful but is still based on existing literature. The code does also extend existing analysis tools to the Julia programming language, but it is this reviewer's opinion that this contributes little in terms of a novel scientific contribution.

Of note, the author could focus on the less-investigated HTDMA problem and the specific challenges that arise for that application (e.g., present the underlying integral equation for that scenario), which the authors note in the conclusion is one of the more novel aspects of this manuscript.

Altogether, this reviewer thinks the manuscript could be reoriented more towards novel scientific components, including more of a focus on HTDMA. As such, the author SHOULD be given the chance to respond to comments and refine the manuscript. MAJOR REVISIONS are recommended.

SPECIFIC COMMENTS

The focus on the programming aspects also often distracts from the underlying science. For example, presenting the underlying mathematics in a programming language- and program-specific representation without the more standard mathematical forms (e.g., the underlying integral equations) makes the manuscript harder to follow. It seems that in an attempt to tread a line between a scientific manuscript and code documentation, the manuscript does not accomplish either task particularly well. In this respect, the manuscript may be better structured by clearly presenting the underlying scientific principles in a more standard mathematical notation, moving coding references out of the body of the manuscript. The alternative - presenting the manuscript as a form of documentation for a program - is better structured with specific coding examples in the text. However, this latter route is less amendable to a research article in AMT. In this case, another platform (a technical note in a journal or an article in a computational journal) may be more suitable. As a hybrid, the SI could be formally formed into documentation for the code that refers to the scientific principles in the base article without cluttering the body with code snippets and representations. Regardless, clarifications should be made before further review.

In the abstract, the authors note that the inversion speed is improved by ~200 times down to 2-5 ms. Is the implication to work towards online inversion of the measurements? If not, there is a fair degree of flexibility in terms of inversion speed, such that speed may not be the only or the best metric gauging improvement. Can the authors comment? If the hope is for online inversion, can the authors comment on the interface with the instrument, which would be a substantial component of the overall process.

Related to the above, this code is 200 times faster compared to what? A previous version of this code? It is worth noting that Tikhonov regularization for these distributions is a relatively straightforward problem, solving a simple linear system. As such, the speed improvements are likely linked to the external libraries that solve the linear system, something which the authors do imply later in the work. However, this does limit the novelty of using those methods for a different problem.

Line 93: What is the dimension/size of the different quantities defined here? Based on the

subsequent discussion, it seems that A is assumed to be square (same reconstruction and measurement discretization). The A matrix is not required to be square, but this reviewer thinks it does make the GSVD simpler to compute (a non-square matrix may require special treatment) and should be stated clearly.

Line 108: Consider explicitly noting that automating the L-curve method, while feasible, is often more challenging than other automated methods and can be affected by noise and type of solver (which the authors admittedly imply later when they state that the L-curve algorithm used previously occasionally failed).

Line 112: Clarify "standard form". What is the standard form? How would one compute this standard form? Under what conditions does one not use the standard form?

Line 208: Is Petters (2018) the best reference for this? The underlying equations for the discrete transfer function of the SMPS have been stated more formally many times before this work. If there is something specific in Petters (2018) about which the authors can be more explicit? There are also multiple ways to discretize the problem, which could be a route to a more specific representation from Petters. Further, why not present this in a more standard form, such as the transfer function given by Stolzenburg (2018), rather than a programming language-specific representation?

Through Section 2.3: Similar to above, why not present all of the physics in terms of its underlying integration equations rather than language-specific concatenation and mapping operators or convolution "*" operators? For the discrete version, why not state these in terms of matrices instead? Interestingly, there are multiple ways to discretize the problem (e.g., finite element bases), which is also not detailed here. The HTDMA problem is based on a double convolution with three components to the underlying integral equation/kernel: 1) the transfer function of the first DMA, 2) a kernel describing the humidification process, and 3) the transfer function of the second DMA. This feature is not clear from the current reading.

Line 268+: The current manuscript structure makes it challenging to ascertain the role of the 30 (or other) bins for the growth factor in the overall procedure. This reviewer would expect that the growth factor would contribute to another matrix that bridges the mobility distribution output by the first DMA to the mobility distribution input to the second DMA. In this respect, since the other components of the problem have a constant number of bins (at least this reviewer gathered as such), would it not make more sense to have a matrix with the same dimension/number of bins as the larger problem? Further, depending on whether one is inferring these quantities or not, this matrix could be combined with one or more of the DMA transfer function kernels and thus be pre-computed, with little effect on the overall computational effort. If one is inferring the growth factor, the structure of the problem deviates somewhat from the more general aerosol inversion problem, a fact that should be clarified. Namely, there will be at least two integrations (over the mobility distributions for each DMA) with an intermediary quantity that is being inferred. Then, there is also the question of the uncertainties in the input size distribution, which is measured independently, also inferred, or assumed.

Overall, these definitions could be clarified.

Line 276: The use of Poisson noise could be used to appropriately weight the data. Why was this not considered (i.e., use weighted least-squares instead of naïve least-squares)? One limitation is that measurements that span multiple orders of magnitude will result in numerical instabilities, such that a baseline amount of background noise may be required. Can the authors comment?

Line 280: How often would this a priori information be known? In the experimental section to follow, there is a short phrase about this being computed using the inverse of the S matrix. Would it be worth noting this here? Also, what is S? This information does not seem to be immediately available.

Line 280: Continuing from above, do the choices for x_0 make sense given the chosen Tikhonov matrix? For example, a first-differences Tikhonov matrix encodes information about the expected slope of the solution. Using $(x - x_0)$ implies regularization of the slope of the residual with respect to an a priori estimate. Can the author comment?

With respect to, "Higher resolution grids generally lead to poor performance even for method LOD1e-3B[0,1]." This is *slightly* surprising. Given the way Tikhonov prior operates, one may expect the extra grid points to be filled using the prior (a little like interpolating between lower resolution points, but not quite the same). Could this be an indication of limitations in the error metric used (there are most points at which the error is being calculated such that one is not comparing the same quantity)? Alternatively, the regularization parameter would change depending on the reconstruction grid. Was the regularization parameter re-optimized each time?

Paragraph around Line 315: Is the unweighted residual really the best metric? How about measurement noise (one may have more confidence in some measurements than others)? Should this be accommodated in terms of calculating this residual?

Line 355: Small values in the A matrix do not matter as much as where they are located. Small diagonals or nearly all-zero rows/columns are the real issues. Consider clarifying. There is the issue of numerical noise (scattered small values) in the kernel, which does little but slow down the inversion. Was this dealt with?

Figure 5: The real-world noise in Fig. 5 does not seem to match noise in other number concentrations reported in the theoretical components of the work. Can the authors comment on this difference and or update the earlier scenarios to be more representative?

For the temporally-evolving measurements, recent work by Ozon et al. (<https://acp.copernicus.org/preprints/acp-2021-99/>) presents an improvement to this existing technique and is closer to state-of-the-art. Can the authors comment and cite appropriately?

Line 460: Code would never involve writing out the Fredholm integral equations, making this statement a bit confusing. Further, scientific manuscripts supporting such code probably should state the underlying Fredholm integral equations. As before, program-specific language makes the scientific components of the article harder to follow.