Referee Comments for Manuscript AMT-2021-51 Ver. 3 "Revisiting Matrix-Based Inversion of SMPS and HTDMA Data" Markus D Petters

This manuscript addresses the important issue of automating the processing of SMPS and tandem DMA data. The idea of inverting data with regularization is sound. However, there is still a functional problem with the development of the matrix-based forward model of calculating system response from a known input distribution. Furthermore, some of the descriptions in that section could be much clearer. If these issues can be properly addressed, the resulting software package should prove of great utility.

Major Comments

As before, the main focus of this review is the development of the forward matrix model in the manuscript. The author has generally addressed my previous comments satisfactorily. In particular, the changes to properly account for multiply-charged particles in the second DMA of a tandem DMA setup are greatly appreciated. However, there appears to be a remaining problem in this situation with tracking the true mobility diameter of the particles after growth, D_{wet}, to the same diameter in DMA2.

When changing thermodynamic state, Λ or (T,p), true mobility diameter (*i.e.* physical diameter) of the particles is the only "size" parameter that remains constant regardless of charge state. For multiply-charged particles, apparent mobility diameter does not equal true mobility diameter and it depends on (T,p). Therefore, for multiply-charged particles transitioning from DMA1 conditions to DMA2 conditions, the relationship between apparent and true mobility diameters changes. If apparent mobility diameter is held constant across the transition, then it is necessarily true that mobility diameter is not. The somewhat garbled explanation of the function of $\Pi_k^{\Lambda,\delta}$ (lines 299-301) seems to indicate that the particle size sent to the convolution matrix, $\mathbf{O}_k$, for transport through DMA2 is the apparent mobility diameter at DMA1 conditions. Though the actual end effect may be small, this is technically incorrect for multiply-charged particles.

In equation form, this may be seen as follows. Let the function h represent the forward calculation of mobility from diameter as $z_1/k = h_1(D_m)$ where the calculation is done at DMA1 conditions. Then the inverse calculation is represented as $D_m = h_1^{-1}(z_1/k)$. Using this, the apparent diameter at DMA1 conditions is given as $D_{a1} = h_1^{-1}(z_1) = h_1^{-1}(k \cdot h_1(D_m))$. If D_{a2} is equated to D_{a1}, then the corresponding true mobility diameter in DMA2 is given as $D_{m2} = h_2^{-1}(h_2(D_{a2})/k)$. For $|k|=1$, all these equations simply collapse such that $D_{m2}=D_{a2}=D_{a1}=D_m$. But for $|k|>1$, $D_{m2}\neq D_m$, meaning that the true mobility (physical) diameter is not preserved across the transition. Furthermore, if the output of $\Pi_k^{\Lambda,\delta}$ is apparent mobility diameter at

DMA1 conditions, then in order to correctly compute the diffusion effects associated with DMA2, this conversion to apparent mobility diameter would need to be directly reversed (at DMA1 conditions) just to get back to true mobility diameter, a computational waste. As far as I can see, the apparent mobility diameter is of no use in the context of these TDMA equations.

Eliminating that concept entirely from the development would also eliminate the need for such strange terminology as an "apparent +1 mobility diameter". At least to me, that equates to a "false true mobility diameter". The term "mobility diameter" alone is well-defined in terms of the particle dynamic mobility (z/ke where e is the unsigned electronic charge). The addition of "+1" is then, in fact, somewhat redundant. There is no need to retain the "+1" when referring to an "apparent mobility diameter".

Since I have not even looked at the code for this, I admittedly know little of the details of the programming of these equations for the calculation of the forward model. However, with the implementation of the corrections discussed above, it seems likely that there remains no compelling reason to treat the input parameters of DMA1, $\Omega^{\Lambda,\delta}(Z, z^s/k, k)$ (Eq. 10), and DMA2, $\Omega^{\Lambda,\delta}(Z, z^s, k)$ (Eq. 15), differently in these equations. Assuming that the Ω algorithm has no hidden switches for different input structures, these variations in input can be quite confusing and seemingly contradictory.

In addition, there is an interpolation step from the input distribution diameter grid to the DMA1 mobility grid and a very similar interpolation step from the grown particle diameter grid, D_{wet}, to the DMA2 mobility grid. Each of these interpolations takes as input the incoming diameter grid and the outgoing mobility grid with its associated DMA Λ parameter. From this, a relatively sparse matrix is generated to operate on the incoming concentration vector resulting in an outgoing concentration vector interpolated to the new grid. In these equations, the two interpolations could be similarly codified as $\Pi^{\Lambda_1}(Z_{s,1}, D_{in})$ and $\Pi^{\Lambda_2}(Z_{s,2}, d_1^s \cdot G)$ where D_{in} is the input diameter grid, $d_1^s = D_p(z_1^s)$ is the mobility diameter associated with z_1^s, G is the vector of discretized g_0 values, and $d_1^s \cdot G$ is the grid of grown diameter bins.

Notation such as used above can greatly improve the clarity of these equations with significantly less reliance on explanations in the surrounding text. In general, there is much room to improve the consistency and logical arrangement of the notation. For instance, there is no obvious reason why the Z dependence, which makes the result a vector, is omitted from the expression $T_{size}^{\Lambda,\delta}(k, z^s)$ (Eq. 10) and the subscript *size* seems to add no information at all. The use of a subscript 1 (to indicate singly-charged?) in the expression $D_{p,1}=D_p(z,k=1)$ (line 259) becomes confusing when used in Eq. (15) for DMA2 and especially so in $g=D_{p,1}/D_d$ (line 345) where it represents the diameter associated with an element of $Z_{s,2}$. It would be far better to reserve subscripts 1 and 2 for their most frequent usage in associating the parameter with a

particular DMA. To distinguish particular values of diameter from the function that generated them, lowercase 'd's could be used as above. And though the superscript s of z^s was apparently originally representative of "star" in "z-star" or z*, at this distant removal from that original notation and in this context, it simply looks out of place. Even the author changed it to a subscript when the Z_s parameter was introduced. Given the number of flaws uncovered at this point in the formulation of the forward model in Petters (2018), there would seem to be little value in clinging to its confusing notation. At least in terms of the forward model, the current work should be seen as superseding the previous work.

Minor Comments and Corrections

Line 31: "For ~~examples~~ example, …". No matter how many examples are proffered in the following text, standard usage is "For example".

Lines 87-88: "Capital bold-roman letters denote matrices (**A**), lower case roman letters denote vectors (x) and lowercase italic symbols denote scalars (*n*)." However, this list is far from exhaustive. For instance, uppercase italic Ts are used as vectors in the later development of the forward model.

Line 88 or line 94: Include the definition of the pseudo inverse, $\mathbf{A}^{+} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}$, as this is perhaps less well known to many. It also helpful to have that available to compare to the form in Eq. (3).

Lines 100-101: "Common choices are the first ~~and~~ or second derivative operator defined as the upper bidiagonal(-1; 1) and the ~~upper~~ tridiagonal(1;-2; 1) matrix, respectively." This same reversal is also found in line 145. There is no such thing as an "upper tridiagonal"; the matrix is symmetrically populated about the main diagonal.

Line 106: "… the derivative of the right hand side of Eq. (2) with respect to x, …"

Lines 108-109: "The L-curve method~~s~~ involves a plot of $\log\left\|\mathbf{A}\,\mathrm{x}_\lambda - \mathrm{b}\right\|_2^2$ vs. $\log\left\|\mathbf{L}\left(\mathrm{x}_\lambda - \mathrm{x}_0\right)\right\|_2^2$." Note that x has been changed to x_λ in these expressions.

Line 109: "The optimal λ occurs at the corner of the resulting L-shaped curve, …"

Lines 111 and following: The description up this point has been very helpful allowing for at least some intuitive grasp of what is going on in these equations and their optimization. Then the GCV estimator is introduced (Eq. 4) with absolutely no physical context for what it means. I realize that the full definition for this is somewhat involved and beyond the scope of this paper, but is there any sort of simplified view of what this parameter physically represents? At least for me, the form of the equation provides no clue at all. Perhaps that is in part due to my ignorance of the standard form and what that physically represents.

Line 113: $\mathbf{A}^T$ should be $\mathbf{A}^{\mathrm{T}}$ (no italics on the superscript), two places.

Lines150-151: Doesn't $\mathrm{D}_{\hat{x}} = diag\left(\left|\hat{x}_1\right|,...,\left|\hat{x}_n\right|\right)$ imply $\left(\mathrm{D}_{\hat{x}}\right)_{ii} = \left|\hat{x}_i\right|$? But then that together with $\left(\mathrm{D}_{\hat{x}}\right)_{ii} = \varepsilon$ for all $\left|\hat{x}_i\right| < \varepsilon$ creates a contradiction if the set $\left|\hat{x}_i\right| < \varepsilon$ is not empty.

Line 179: "*RegularizationTools.jl* also provides an abstract generic interface …"

Line 196: "… the raw response function defined as the integrated response …"

Line 201: "… requires ~~solving~~ evaluating integrals …"

Lines 203-206: "For the forward calculation, the objective is to find a design matrix that maps the growth factor frequency distribution to the raw TDMA response function. ~~The objective is to find a design matrix that maps the growth factor frequency distribution to the raw TDMA response function.~~"

Line 208: "… provide a domain ~~specific~~ independent language …" This change would seem to make more sense in this context.

Line 223: "… +1 (singly charged) mobility diameter bin edges …" If you wish to continue to use the somewhat superfluous "+1" notation, it should be explained such as this on first use.

Line 226: "For ~~examples~~ example, …"

Line 227: "…$f * \mathrm{n}$ is the uniform scaling of the concentration fields by factor *f*." Here and in the subsequent example of $f \cdot \mathrm{n}$, *f* is treated as a simple scalar. But in the very next paragraph *f* and *g* are used as functions. Rather than tempt confusion, use some other letter such as *a* for scalar multiplication, *e.g.* $a * \mathrm{n}$.

Line 233: "… and sums the results."

Lines 234 and following: The description of the $\mathrm{foldl}()$ function is clearer than that of the previous splatting function. However, you have stated the generic rule for the first element while giving examples directly from Julia documentation which use a different rule. The generic rule is that for the first element, *a* is the neutral value. In the Julia implementation of foldl, the first element acts as *a* for the processing of the second element. For the example foldl(–; [1; 2; 3]), the neutral value is 0 and the operation is – (*a*, *x*) = *a* – *x*. The generic implementation of this would be $0 - 1 - 2 - 3 = -6$ while the Julia implementation is $1 - 2 - 3 = -4$. The way foldl is used in the development of the forward model would appear to conform to the Julia implementation. So the description of the first element rule should be changed.

Lines 265-266: "… which corresponds to the centroid mobility setting for the DMA to transmit particles ~~with the size of particles~~ with k charges under the assumption that they carry only a single charge." Even with this change, I find it very difficult to understand what is going on

in this section. The only way I can make sense of this is that the Z vector at this point contains values of z/k, not simply z. If this is the case, it needs to be stated as so somewhere. Also, if this is the case, then I believe the DMA transfer function dependence in Eq. (10) given as $\Omega(Z, z^s/k, k)$ should actually be $\Omega(Z, z^s/k, 1)$, assuming $\Omega(z, z^s, k)$ is the "normal" arrangement of inputs.

Line 268-270: "The function $T_{size}^{\Lambda,\delta}(1, z^s)$ evaluates to a vector of the same length as Z. Performing an elementwise sum over all $T_{size}^{\Lambda,\delta}(k, z^s)$ produces the net mobility distribution transmitted by the DMA." Perhaps it is clear enough from context, but noting that the summation is over all k could be helpful. More importantly, the result of the summation is only the net mobility distribution if $\sum_k T_{size}^{\Lambda,\delta}(k, z^s)$ is multiplied by the input distribution, n. Otherwise, the bare summation is simply the net transmission probability function.

Line 270: "Examples for … ~~is~~ <u>are</u> shown in Figure 2, …"

Lines 276-277: "… and <u>Z_s</u> is a vector of centroid mobilities scanned by the DMA." The vector Z_s is missing from this description.

Line 277: "The matrix is square if $Z_s = Z$ in <u>Eq. 10</u>." I think you mean Eq. 11. As noted earlier, this would be a lot clearer if the explicit Z dependence were actually shown in $T_{size}^{\Lambda,\delta}(k, z^s)$ as $T_{size}^{\Lambda,\delta}(Z, z^s, k)$.

Lines 289-290: "The latter is the raw response function defined as integrated response downstream of the DMA …" The latter refers to r in Eq. 12. r is similar to a size distribution, it is NOT integrated over all sizes as would be the case for the response of a CPC detector.

Line 295: "The mobility distribution ~~exiting~~ <u>entering</u> DMA 2 …"

Line 297: "… Eq. (13) … evaluates to … that exit the <u>humidity conditioner</u> ~~DMA1~~ …"

Lines 298-299: "Subscripts are used to differentiate DMA 1 and 2 which possibly have different geometries, flow rates, and grids, …" This list should include thermodynamic states, that is, different temperature and/or pressure. DMA2 is definitely at a lower pressure than DMA1 and this is reflected in the relationship between mobility and mobility diameter.

Line 311: "The total humidified mobility distribution $\mathrm{m}_t^{\delta_2}$ exiting DMA 2 is given by (Eq. 14) …" The total humified mobility distribution exiting DMA2 would be that observed at fixed $z_{s,2}$ without integrating over particle size. $\mathrm{m}_t^{\delta_2}$ is the total integrated concentration versus $z_{s,2}$. It is not a true particle distribution.

Line 322: "… the humidified distribution function exiting DMA 2 …" Same as previous comment. This might be considered to be a pseudo-distribution if what is meant by that is explained up front.

Line 324: In reference to Eq. 16, "…the diameters in $\mathrm{M}_k^{\delta_1}$ are normalized by D_{dry}." This does not seem correct. The diameters, or at least the associated mobilities, cannot be altered until after passage through DMA2. I can't figure out where in Eq. 16 you could safely normalize diameters.

Line 327: "… the DMA ~~setup~~ setups Λ_1, Λ_2, δ_1, δ_2 …"

Lines 332-333: "Transmission through DMA 1 is computed for a specified z^s (the dry mobility) and g_0 (the growth factor) via Eq. (13)." The humidity conditioner as represented by g_0 is NOT part of DMA1.

Line 339: "… the matrix is non square." This is only true if $j \neq i$. Perhaps j is typically much smaller than i, but you haven't said anything about that.

Line 342: "The advantage of interpolation is that the ~~the~~ matrices $\mathbf{O}_k$ are smaller." Here again it is implied that j is significantly smaller than i but you have not explained why that is.

Line 347: "… the humidified mobility pseudo-distribution function …" or some such modification.

Lines 350-351: "If the grids for P_g and that of DMA 2 do not align, interpolation is used to map the P_g grid onto the DMA 2 grid." The P_g grid would refer to the discretization of g_0, or G, using the notation from above. What is being interpolated onto the DMA2 grid is $d_1^s \cdot G$.

Lines 351-352: "The choice of i, j, n, the ranges of mobility grids for DMA 1, DMA 2, and the range of P_g, is only constrained …" Insert comma after P_g. The range of P_g is simply [0,1]; it is the range of g_0 that is needed here or perhaps "the range of the growth grid for P_g".

Line 359: A new paragraph should start with "Figure 1 …".

Line 390: Concerning bounds for P_g, if $P_g = dF/dg$ from Figure 1 caption and as would be appropriate in Eq.16, then there is no upper bound for Pg. For if one introduced a perfectly monodisperse as the input aerosol, P_g would range from zero to infinity. However, in discretizing the integral of Eq. 16 it is possible to assign the bin width factor, Δg, to the dF/dg vector to get a ΔP_g vector. This would be properly bounded by [0,1]. This should be explained more clearly or the notation in the equations appropriately modified to show this.

Lines 396-397: "Visual evaluation … ~~suggest~~ suggests …"

The balance of the paper was not reviewed.