

Supplement of **A Software Package to Simplify Tikhonov Regularization with Examples for Matrix-Based Multi-Charge Inversion of SMPS and HTDMA Data**

Markus D Petters¹

¹NC State University, Department of Marine, Earth, and Atmospheric Sciences, Raleigh, NC, 27695-8208

Correspondence: Markus D Petters (mdpetter@ncsu.edu)

Monte-Carlo analysis was used to test the fidelity of the inversion of Eq. (15). Each simulation follows the example presented in Figure 2 of the main text. Starting from a selected test case, the noise-perturbed apparent growth factor distribution function is simulated and then inverted it using one of the 12 bounded methods. The root mean square error of the reconstructed solution and the input is evaluated. Test cases include the *Bimodal*, *Uniform*, *Truncated Normal*, and *Population* examples shown in Figure 2 of the main test. The *Population* examples here are denoted *Single Channel* and *Two Channel*, corresponding to one or two populations, respectively. The simulations are a combination of the following test matrix: $D_d \in [20, 50, 100, 200, 300]$ nm, $N_t \in [500, 1000, 5000, 50000]$ cm⁻³, $n \in [20, 30, 40, 50, 60]$, and $p \in$ [“Single Channel”, “Two Channel”, “Uniform”, “Bimodal”, “Truncated Normal”], where D_d is the dry diameter, N_t is the total number concentration, and n the number of bins of the growth factor frequency distribution, and p denotes the test case. The 12 inversion methods are $[L_0B, L_1B, L_2B, L_0D_\epsilon B, L_1D_\epsilon B, L_2D_\epsilon B, L_0x_0B, L_1x_0B, L_2x_0B, L_0x_0D_\epsilon B, L_1x_0D_\epsilon B, L_2x_0D_\epsilon B]$. For all methods, the *a-priori* estimate x_0 is taken to be the normalized apparent growth factor distribution, where the normalization ensures that the sum over all bins equals to unity. All D_ϵ are taken to be D_{1e-3} and the lower and upper bounds are [0,1] for all simulations. Each combination of input parameters is simulated for 10 random seeds, resulting in $5 \times 4 \times 5 \times 5 \times 12 \times 10 = 60000$ inversions.

The input aerosol size distribution is fixed and bimodal with mode diameters of 50 and 130 nm, geometric standard deviations of 1.4 and 1.6 and number concentration of $0.66N_t$ and $0.33N_t$ in modes 1 and 2, respectively. The truncated probability density function is a Gaussian with a mean of 1.2, standard deviation of 0.2, and is truncated for $gf < 1$. The uniform distribution has equal probability density between $1.3 < gf < 1.8$. The bimodal probability density function are Gaussian with means and standard deviations of 1.3, 1.7, and 0.07, 0.02 for modes 1 and 2 respectively and 70% of particles reside in the less hygroscopic mode. The single channel example has a growth factor of 1.6. The two channel example has growth factors separated by three size channels and frequencies of 0.7 and 0.3. The Poisson noise is evaluated for a particle counter with 1 L min⁻¹ flow rate and 2 s bin integration time. However, the total noise scales with the total number concentration of the size distribution.

Figure S1 summarizes the simulations, stratified by method, test-case and number of bins. The figure shows box-and-whisker plots of the log10 transformed root mean square error, with 250 values contributing to each box-and-whisker entry. For the uniform, bimodal, and truncated normal cases, typical average log10 root mean square errors are close to -2, i.e. root mean

25 square errors near 0.01. All of the methods perform approximately equally well. In contrast, most methods perform poorly for the single channel and two channel model, with root mean square errors exceeding 0.1. The exceptions are the $L_0D_\epsilon B$ methods for $n = 20$ and $n = 30$ bins. Since $L_0D_\epsilon B$ does not perform worse for uniform, bimodal, and truncated normal, but better for the single channel, and some extent the two channel test cases, $L_0D_\epsilon B$ is selected as the main regularization method for inversion involving tandem DMA data.

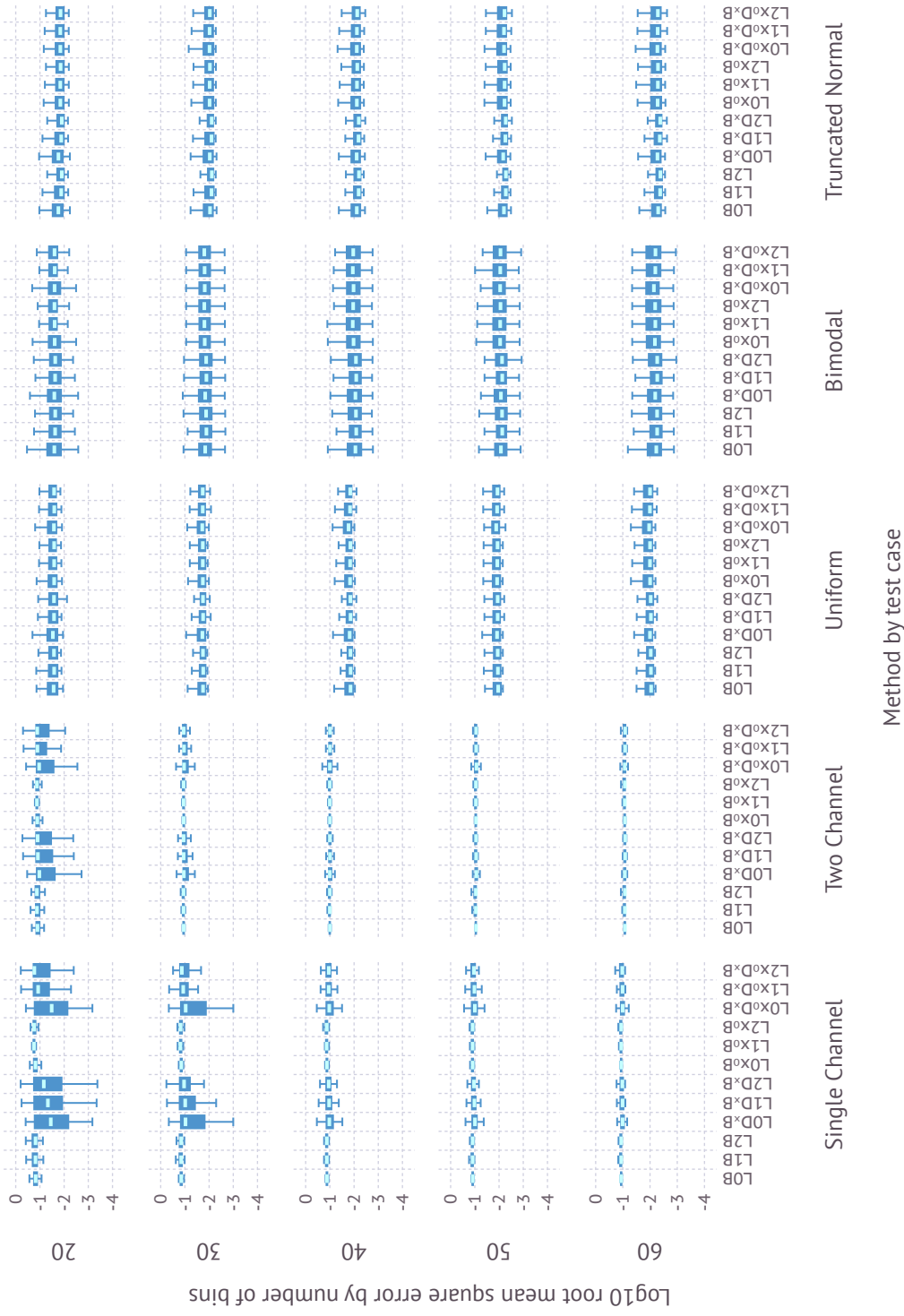


Figure S1. Results from simulations grouped by method and number of bins. Note that the legend of the figure denotes the method for example as L0DxB. The reason for this is the automatic generation of the label, which did not allow for splicing Unicode subscripts from variable names (hence L0 instead of L_0). Unicode superscript ϵ is not defined, hence D_x is used instead for the axis labels.