

Supporting information for:

Correcting Bias in Log-Linear Instrument Calibrations in the Context of Chemical Ionization Mass Spectrometry

Chenyang Bi¹, Jordan E. Krechmer², Manjula R. Canagaratna², Gabriel Isaacman-VanWertz¹

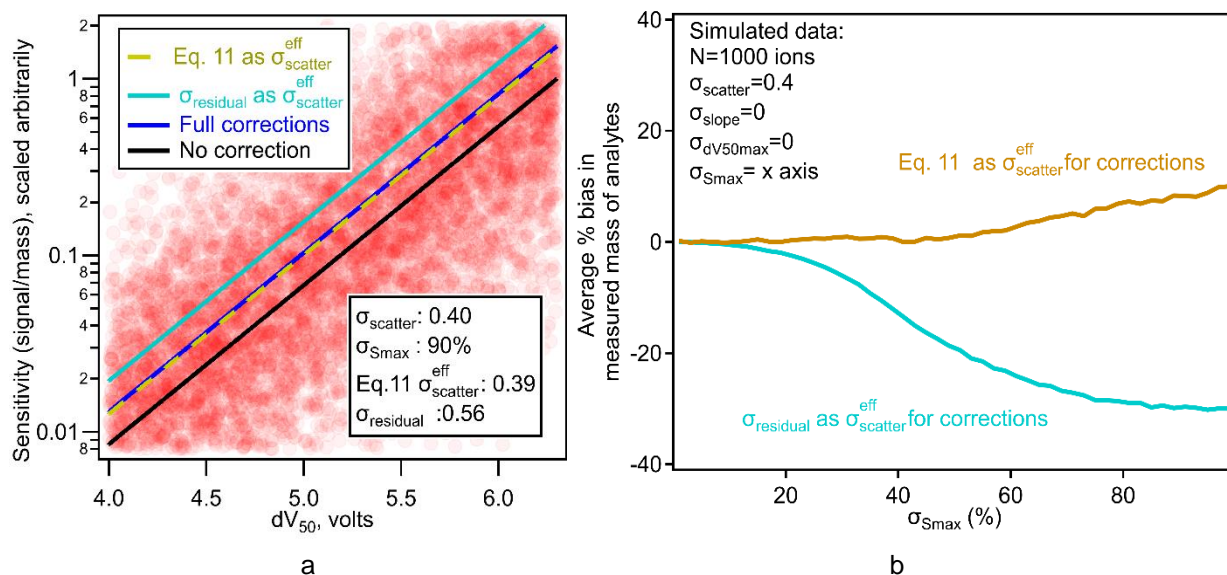
5 ¹Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24060, USA

²Aerodyne Research Inc., Billerica, Massachusetts 01821, USA

Correspondence to: Gabriel Isaacman-VanWertz (iwv@vt.edu)

Determination of $\sigma_{scatter}^{eff}$

10 A critical issue in the accurate implementation of a simplified, non-parameter-explicit bias correction is the selection of a reasonable $\sigma_{scatter}^{eff}$. As described in the main text, a reasonable estimate of $\sigma_{scatter}^{eff}$ must be represented by the standard deviation of the residuals, $\sigma_{residual}$, between measured and fitted sensitivity in log units, $\log(S_{measured,i}) - \log(S_{fitted,i})$. However, the effect of σ_{Smax} on $\sigma_{residual}$ must be considered and removed as described by Eq. 11 because, in cases of high σ_{Smax} (e.g., 90%, Figure S1), using $\sigma_{residual}$ as $\sigma_{scatter}^{eff}$ overcorrects for the bias by up to 30% (Figure S1b)



15 **Figure S1.** Simulated data set with $\sigma_{scatter}$ dominating the source of uncertainty (N=1000 ions, $\sigma_{scatter}=0.4$, $\sigma_{slope}=0$, $\sigma_{dV50,max}=0$, $\sigma_{Smax}=90\%$). (a) The nominal (black line), parameter-explicit bias corrected (dark blue line), and simplified $\sigma_{scatter}^{eff}$ corrected relationship with $\sigma_{residual}$ as $\sigma_{scatter}^{eff}$ (light blue line) and with $\sigma_{scatter}^{eff}$ estimated based on Eq. 11 (yellow dashed line). (b) The influence of σ_{Smax} on the average percent bias of analytes using $\sigma_{residual}$ or Eq. 11 as $\sigma_{scatter}^{eff}$ for the simplified bias correction.

20 For use in Eq. 11, Eq.12 provides a conversion of linear uncertainty, $\sigma_{S_{max}}$, to log-equivalent uncertainty, $\sigma_{S_{max}(log)}$, for cases where $\sigma_{S_{max}}$ is below $\sim 50\%$. The the analytical relationship between $\sigma_{S_{max}(log)}$ and $\sigma_{S_{max}}$ is complex, but can be calculated numerically using a simulated set of S_{max} data with normally distributed error. The relationship between $\sigma_{S_{max}(log)}$ from $\sigma_{S_{max}}$ is described by Eq. S1:

$$\sigma_{S_{max}(log)} = -0.0635 + \frac{0.476}{1 + \exp\left(\frac{0.325 - \sigma_{S_{max}}}{0.179}\right)} \quad (S1)$$

25 For values of $\sigma_{S_{max}}$ greater than $\sim 50\%$, this equation provides a reasonable estimate of $\sigma_{S_{max}(log)}$, while for values below 50% it is in approximate agreement with Eq. 12.

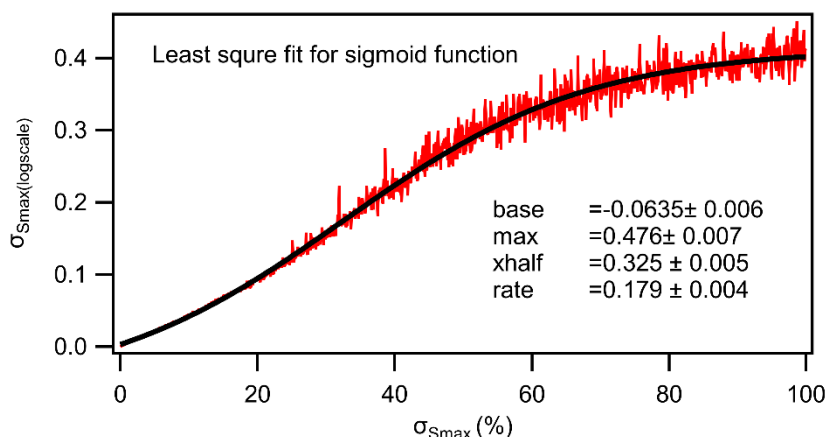


Figure S2. The simulated and fitted relationship between $\sigma_{S_{max}(logscale)}$ and $\sigma_{S_{max}}$.

The obtained $\sigma_{scatter}^{eff}$ using Eq 11 is very close to the theoretical correction factor, $\sigma_{scatter} = 0.4$, regardless of $\sigma_{S_{max}}$ values and the bias-corrected mass approximates the true mass of analytes (dark yellow line in Figure S1b). Notice that the simplified

30 $\sigma_{scatter}^{eff}$ correction approach still overestimates analyte mass by 5-10% when $\sigma_{S_{max}}$ is greater than 70%, probably due to the uncertainties in transforming $\sigma_{S_{max}}$ to $\sigma_{S_{max}(log)}$. Although such bias correction method does not fully remove the bias in some circumstances, it is relatively close to the full parameter-explicit bias correction relationship, as shown in Figure S1a, and values of $\sigma_{S_{max}}$ this large should probably be considered with some caution.

Alternatives for $\sigma_{scatter}^{eff}$

35 While Eq. 11 is effective to determine $\sigma_{scatter}^{eff}$ for the simplified approach, it relies on $\sigma_{S_{max}}$. Typically, $\sigma_{S_{max}}$ can be determined by multi-point measurements of analytes that are known to be maximumly sensitive in the chemical ionization mass spectrometer (CIMS) (Lopez-Hilfiker et al., 2016), but there may be cases where $\sigma_{S_{max}}$ is not well known. To address

the lack of $\sigma_{S_{max}}$ information, we examine the possibility of other statistical parameters in the log-linear fit to correct the bias using a real-world case reported by Isaacman-Vanwertz et al. (2018)

40

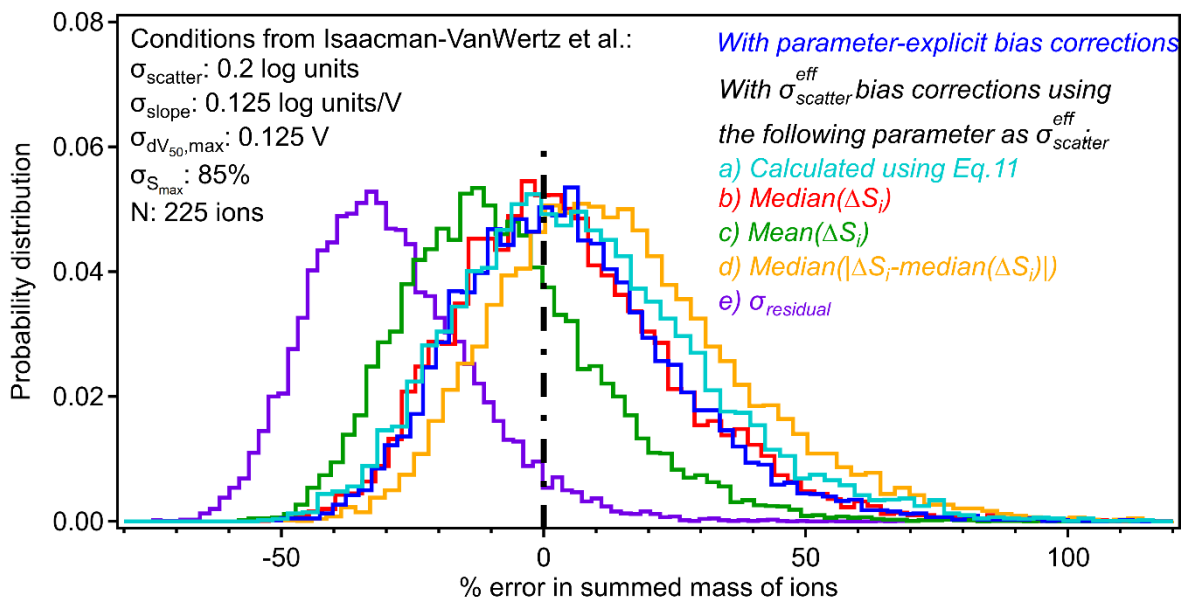
As shown in Figure S3, we compare the error for the prediction of summed mass using different statistical parameters as $\sigma_{scatter}^{eff}$ including (a) Eq 11; (b) the median of absolute residuals; (c) the mean of absolute residuals; (d) the median deviation of absolute residuals, and (e) $\sigma_{residual}$. Absolute residuals are defined here as the magnitude of the difference between measured and fitted log sensitivity, $\Delta S_i = |\log(S_{measured,i}) - \log(S_{fitted,i})|$. The detailed definitions of each

45

- a) Eq. 11: calculate $\sigma_{scatter}^{eff}$, with $\sigma_{S_{max}(log)}$ estimated using Eq. S1
- b) Median of absolute residuals, $median(\Delta S_i)$
- c) Mean of absolute residuals, $mean(\Delta S_i)$
- d) Median deviation of absolute residuals, $median(|\Delta S_i - median(\Delta S_i)|)$
- e) $\sigma_{residual}$: the standard deviation of the residuals

50

Figure S3 suggests that using the median of absolute residual (red line) as $\sigma_{scatter}^{eff}$ in the simplified approach has similar skill as Eq. 11 to match the parameter-explicit bias corrections (dark blue lines). However, Figure S3 only presents a case study and a broader investigation of cases in Figure S4 demonstrates the superiority of Eq. 11 over the median of absolute residuals.



55

Figure S3. Errors in summed mass of ions with median absolute residual, mean absolute residual, median absolute deviation, and gaussian width of residuals, as $\sigma_{scatter}^{eff}$. The residuals are the differences between measured and fitted log(sensitivities) of voltage-scanned calibrants.

Figure S4 shows the influence of σ_{scatter} , σ_{slope} , and $\sigma_{\text{dV50,max}}$ on the average percent bias in summed mass of analytes when (a) Eq. 11 is used as $\sigma_{\text{scatter}}^{\text{eff}}$ for the simplified bias correction; (b) median of absolute residuals is used as $\sigma_{\text{scatter}}^{\text{eff}}$ for the simplified bias correction; and (c) no bias correction. We found that despite subject to some bias at high values of σ , using Eq. 11 is able to remove most forms of bias, except in cases of high σ_{slope} . Conversely, the median of absolute residuals is not able to capture most forms of bias. The positive results shown in the case study of Figure S3 are likely attributable to the large role of σ_{Smax} in the uncertainty, which does not generally introduce bias. In cases where σ_{Smax} is not known, the median of absolute residuals therefore represents a worst-case option, but Eq. 11 or a parameter-explicit bias correction (Eq. 9) is strongly preferred.

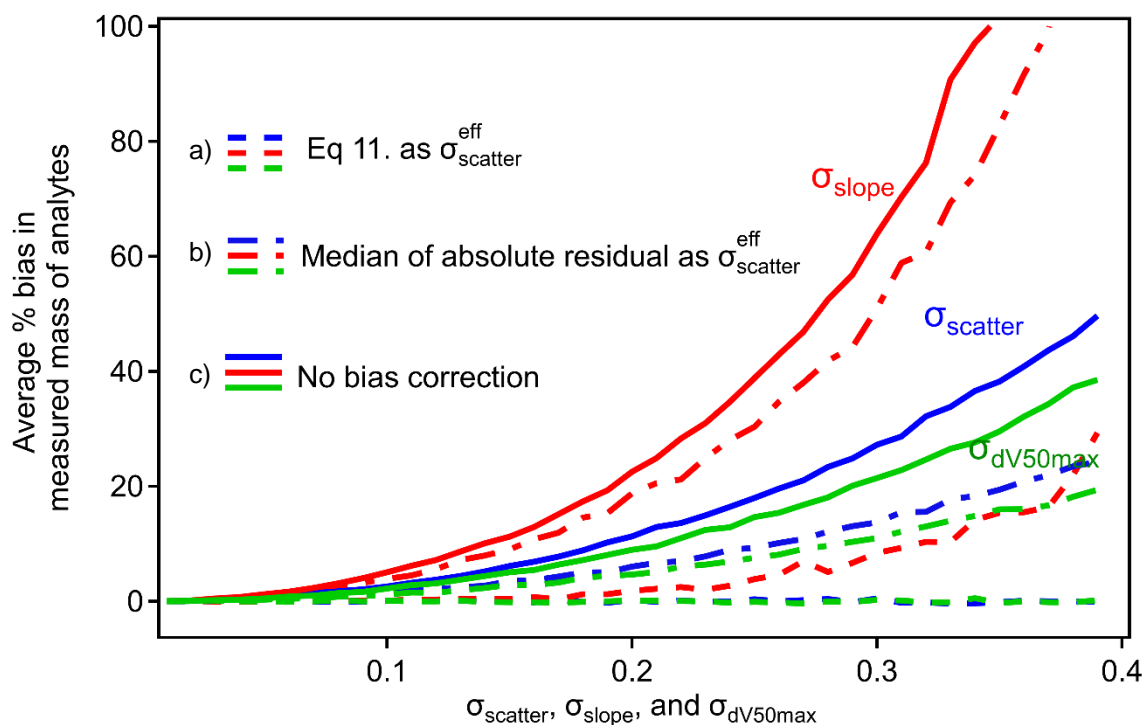


Figure S4. The influence of σ_{scatter} , σ_{slope} , and $\sigma_{\text{dV50,max}}$ on the average percent bias in summed mass of analytes when (a) Eq. 11 is used as $\sigma_{\text{scatter}}^{\text{eff}}$ for the simplified bias correction; (b) the median of absolute residuals is used as $\sigma_{\text{scatter}}^{\text{eff}}$ for the simplified bias correction; and (c) no bias correction is applied.

70 **Comparison of $\sigma_{scatter}^{eff}$ to parameter-explicit correction**

While $\sigma_{scatter}^{eff}$ is shown in the Figures 6 and S3 to capture average bias with low average error. However, the skill with which the simplified approach is able to capture bias in each individual analyte is function of the heteroscedasticity of the error. This, in turn, is dependent on the relative importance of each source of uncertainty. In Figure S5, the impacts of varying $\sigma_{scatter}$ and σ_{slope} is explored. As an illustration of the limitations of the simplified approach, consider panels (c) and
75 (e), for which $\sigma_{scatter}^{eff}$ is similar, but the simplified average approach more closely matches the parameter-explicit correction when $\sigma_{scatter}$ is substantially larger than σ_{slope} .

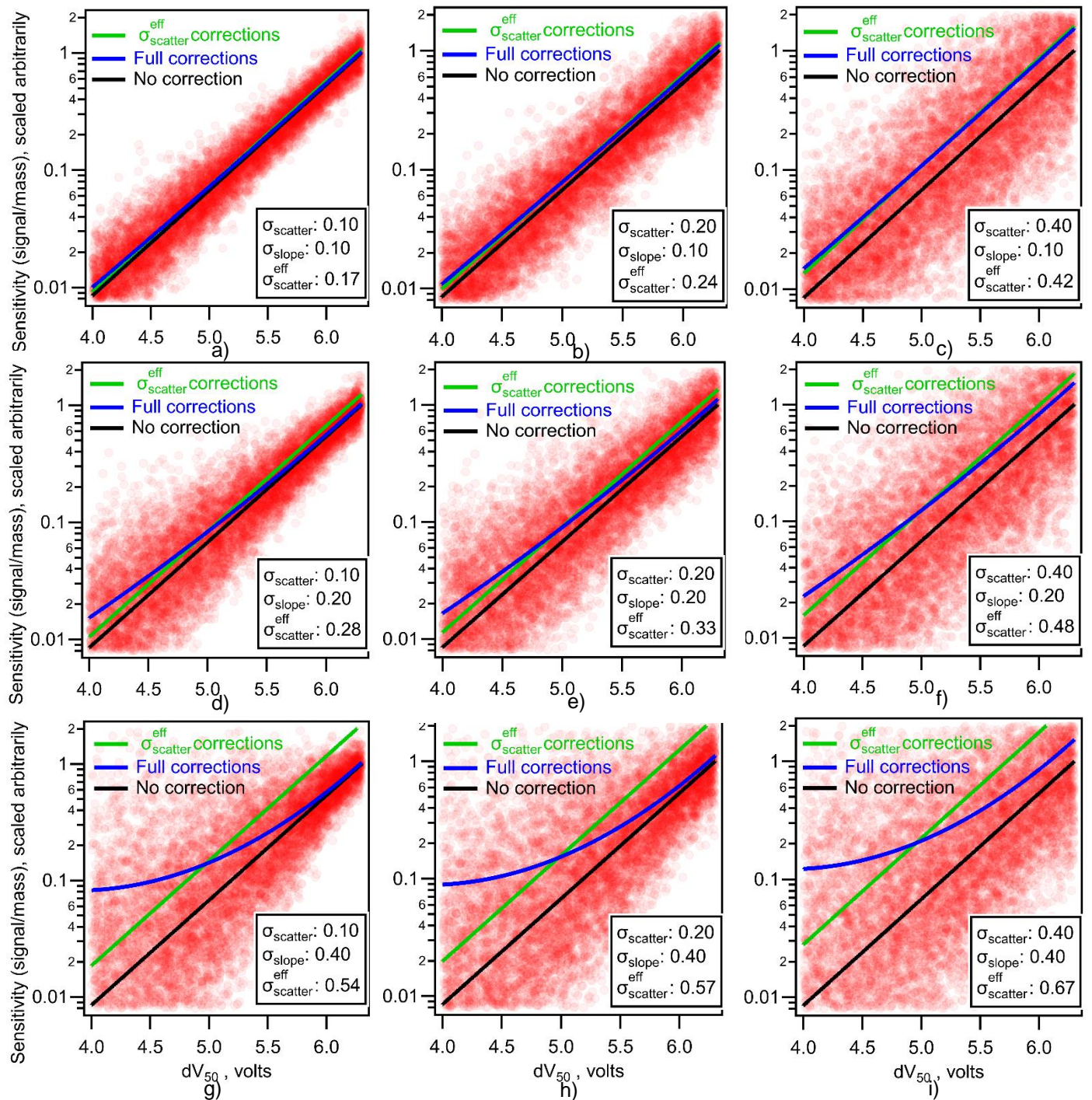


Figure S5. The influence of σ_{scatter} (log units) and σ_{slope} (log units/volt) of the analyte sensitivity distribution on the fitted relationship with parameter-explicit bias corrections and simplified effective σ_{scatter} corrections (Eq. 11). Each circle represents one ion with the distribution of σ_{scatter} and σ_{slope} uncertainties listed. The number of simulated ions, $\sigma_{dV_{50}, \text{max}}$, and $\sigma_{S_{\text{max}}}$ in the dataset are 10000, 0 V, and 0%, respectively.

References:

- 85 Isaacman-Vanwertz, G., Massoli, P., O'Brien, R., Lim, C., Franklin, J. P., Moss, J. A., Hunter, J. F., Nowak, J. B., Canagaratna, M. R., Misztal, P. K., Arata, C., Roscioli, J. R., Herndon, S. T., Onasch, T. B., Lambe, A. T., Jayne, J. T., Su, L., Knopf, D. A., Goldstein, A. H., Worsnop, D. R., and Kroll, J. H.: Chemical evolution of atmospheric organic carbon over multiple generations of oxidation, *Nature Chemistry*, 10, 462-468, 2018.
- Lopez-Hilfiker, F. D., Iyer, S., Mohr, C., Lee, B. H., D'Ambro, E. L., Kurtén, T., and Thornton, J. A.: Constraining the sensitivity of iodide adduct chemical ionization mass spectrometry to multifunctional organic molecules using the collision limit and thermodynamic stability of iodide ion adducts, *Atmospheric Measurement Techniques*, 9, 1505-1512, 2016.