Supplement of Atmos. Meas. Tech., 9, 3337–3354, 2016
http://www.atmos-meas-tech.net/9/3337/2016/
doi:10.5194/amt-9-3337-2016-supplement

Atmospheric Measurement Techniques
Open Access

Supplement of

Sampling strategies and post-processing methods for increasing the time resolution of organic aerosol measurements requiring long sample-collection times

Rob L. Modini and Satoshi Takahama

Correspondence to: Satoshi Takahama (satoshi.takahama@epfl.ch)

Section S1: Regularization parameter choice

In a real life experiment, the true time series that one seeks to measure cannot be known *a priori*. Therefore, when performing TSVD or Tikhonov regularization, it is not possible to choose the optimal regularization parameter corresponding to the aims of a given experiment (e.g. the regularization parameter that minimizes the RMSE error between the deconvolved and true time series as utilized in this work, or the regularization parameter that minimizes the difference between the peak concentrations in the deconvolved and true time series if one was mainly interested in maximum concentrations). One must select a parameter by calibration to a training set similar to the data set of interest, or employ a parameter choice method based only on available measurement data.

A number of such parameter choice methods have been devised (Hansen, 1992) and implemented in software packages for inverse modeling (e.g. Regularization Tools Version 4.1 for MATLAB; Hansen, 2007). Perhaps the most convenient and intuitive of these methods is the L-curve criterion, which seeks to balance minimization of the solution ($\|\boldsymbol{g}\|$) and residual ($\|\boldsymbol{H}\hat{\boldsymbol{f}} - \boldsymbol{g}\|$) norms. A plot of the solution norm versus the residual norm for all valid regularization parameters often yields an L-curve on a log-log scale. This indicates that beyond a certain point, less filtering of singular values produces only minimal reductions in the residual norm, but very strong increases in the solution norm (hence the vertical stroke of the L-curve). Examples of such solutions are the curve corresponding to $k = 53$ in Fig. 5d) and the curve corresponding to $\lambda = 0.1$ in Fig. 5e). The L-curve criterion chooses the regularization parameter corresponding to the corner of the L-curve. In other words, the method chooses the smoothest solution that produces an acceptably low residual norm.

Cross validation can also be used to estimate the regularization parameter. This method amounts to successively leaving out a single element of the measurement vector $\boldsymbol{g}$, and choosing the regularization parameter that best predicts the left-out observations.

Section S2: Detection limits and sampling times

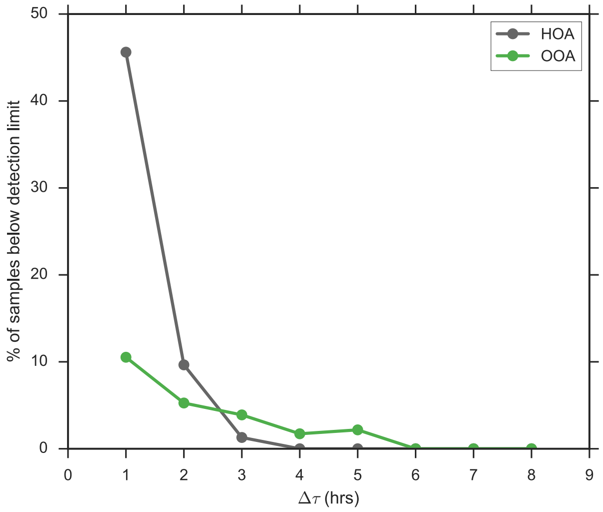

Figure S1. The percentage of HOA and OOA measurement samples below detection limit as a function of the sampling interval *Δτ*. Detection limits are defined as 3 times the constant error term expressed as a concentration ($\sigma_{0,c}$).

Section S3: Additional sequential sampling results

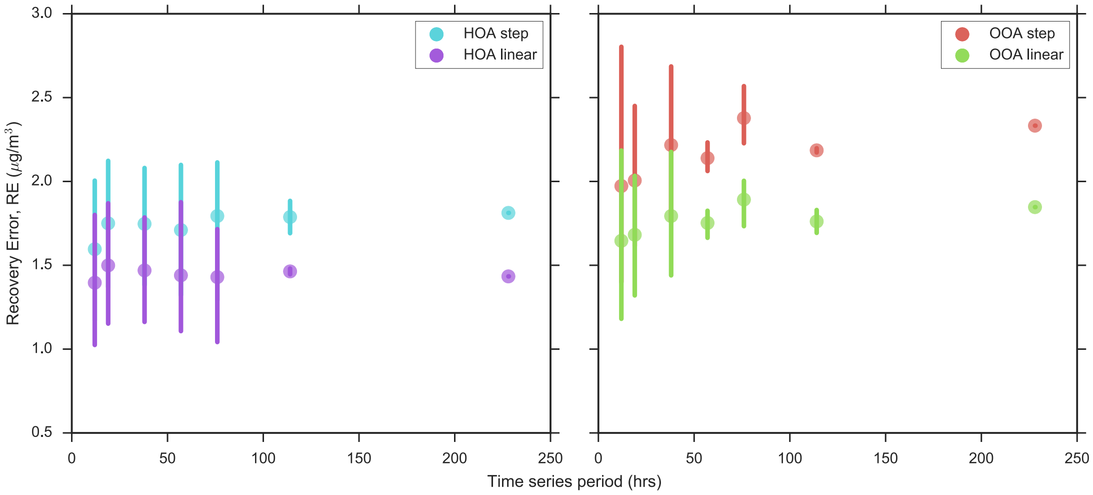

Figure S2. Mean Recovery error (*RE*) as a function of the time series period *T* for HOA and OOA time series constructed by step and linear interpolation between sequential measurements of length (*Δτ*) 4 hours. κ_m = 20%. The vertical bars represent 95% confidence intervals determined by bootstrapping the mean estimates.

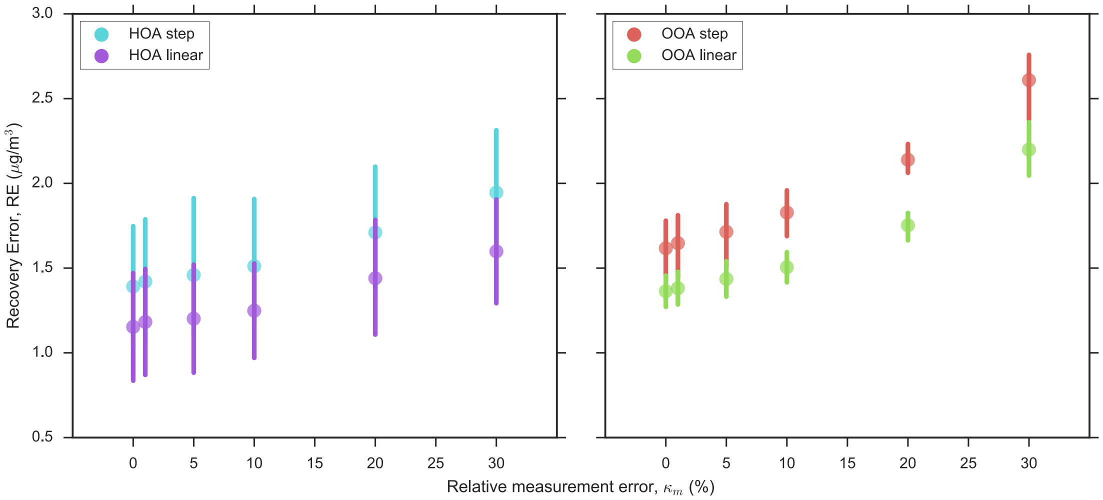

Figure S3. Mean Recovery error (*RE*) as a function of the relative measurement error κ_m for HOA and OOA time series constructed by step and linear interpolation between sequential measurements of length (*Δτ*) 4 hours. *T* = 57 hours, meaning each data point is an average over 4 (=228/57) time series segments. The vertical bars represent 95% confidence intervals determined by bootstrapping the mean estimates.

Section S4: Additional deconvolution results

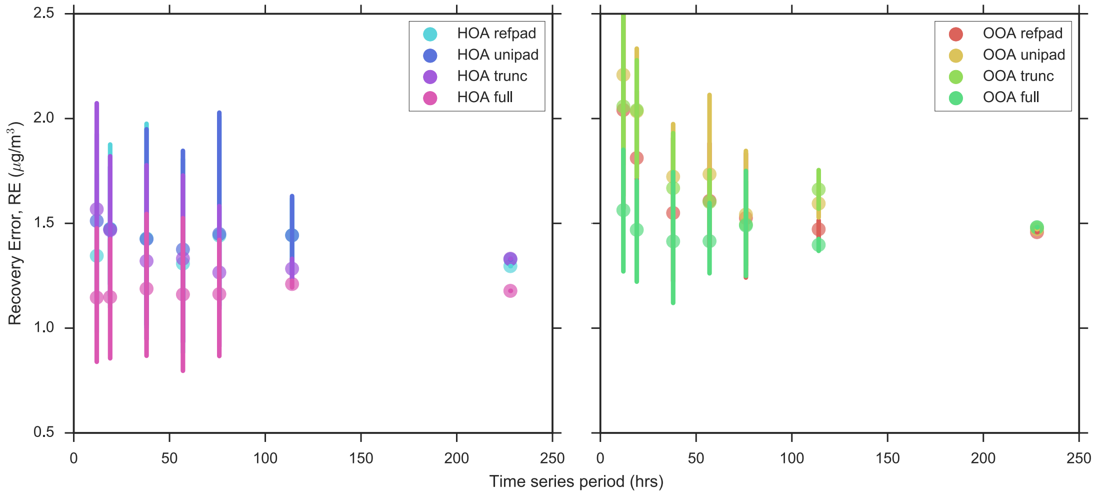

Figure S4. Mean Recovery error (*RE*) as a function of the time series period *T* for HOA and OOA time series constructed by deconvolution with TSVD regularization of staggered measurements of length ($\Delta\tau$) 4 hours. κ_m = 20%. The boundary value methods are full; trunc, truncated; unipad, uniformly padded; and refpad, reflectively padded. The vertical bars represent 95% confidence intervals determined by bootstrapping the mean estimates.

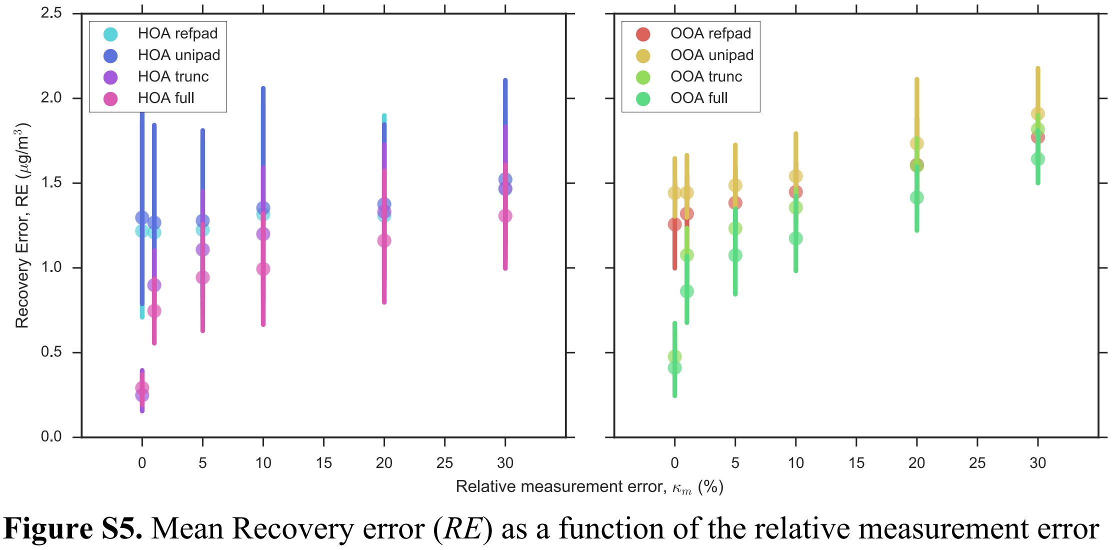

Figure S5. Mean Recovery error (*RE*) as a function of the relative measurement error κ_m for HOA and OOA time series constructed by deconvolution with TSVD regularization of staggered measurements of length ($\Delta\tau$) 4 hours. *T* = 57 hours, meaning each data point is an average over 4 (=228/57) time series segments. The boundary value methods are full; trunc, truncated; unipad, uniformly padded; and refpad, reflectively padded. The vertical bars represent 95% confidence intervals determined by bootstrapping the mean estimates.

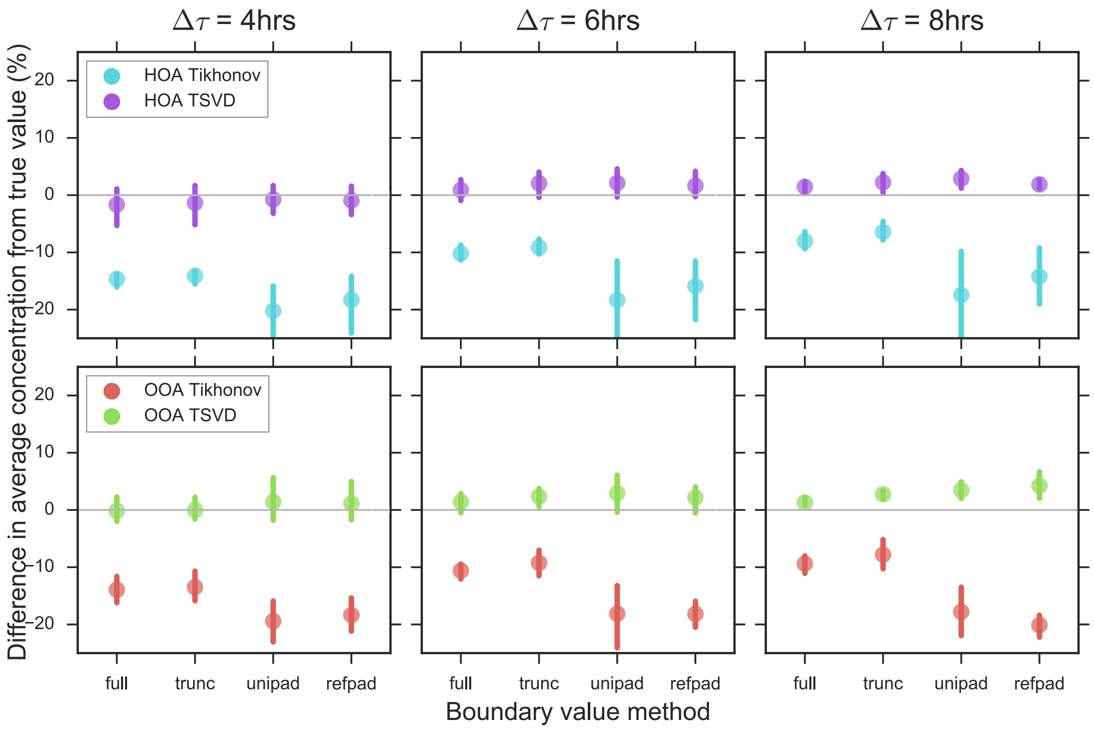

Figure S6. Percentage deviations of average concentrations over the full time series from the true values for different boundary value methods applied to HOA and OOA time series constructed by deconvolution with TSVD and Tikhonov regularization of staggered measurements of length ($\Delta\tau$) 4, 6, and 8 hours. κ_m = 20% and T = 57 hours, meaning each data point is an average over 4 (=228/57) time series segments. The boundary value methods are full; trunc, truncated; unipad, uniformly padded; and refpad, reflectively padded. The vertical bars represent 95% confidence intervals determined by bootstrapping the mean estimates.

Section S5: Equivalent bias and error characterization

Each combination of 4 – 8-hour measurements and post-processing algorithm can be considered as a separate, self-contained instrument. The estimated concentrations can be used to characterize the equivalent bias and error of that instrument.

For example, equivalent bias can be characterized with the following linear model

$$\tilde{c} = \beta_0 + \beta_1 c \tag{S1}$$

where $\tilde{c}$ is concentration estimated by the instrument (a series of values of which comprises an estimated time series $\hat{\boldsymbol{f}}$), c is true concentration (a series of values of which comprises a true time series $\boldsymbol{f}$), and β_0 and β_1 are the fitted parameters of the linear model. For the case T = 57 hours, $\Delta\tau$ = 4 hours and κ_m = 20%, HOA $\tilde{c}$ versus c plots are shown in Figs. S7a – e) for the sequential high and low, smeared, and recovered cases considered in Section 7 of the main text. For comparison, a baseline *ME* only case is also shown. The *ME* only case is obtained by processing the true, hourly concentrations through the linear error model defined by Eq. 6 in the main text. As expected, minimal bias is observed for the *ME* only case. In contrast, the concentrations estimated by the post-processing methods are biased high at low concentrations (< 5 µg m^{-3}), and biased low at higher concentrations.

Bias-corrected residuals are displayed in Figs. S7f – j). The standard deviations σ of these residuals are plotted against true concentration c in Figs. S7k – o). For the *ME* only case, the standard deviations of the residuals are a linear function of concentration (Fig. S7k). As expected, the slope and intercept of the line obtained by an ordinary least squares fit to the data are statistically equivalent (at the 95% confidence level) to the input parameters of the linear error model used to generate the data (Eq. 6 with κ_m= 0.2 and $\sigma_{0,m}$ = 0.5).

However for the other estimation methods, a simple linear model does not adequately capture the dependence of σ on c (Figs. S7l – o), indicating that the post-processing methods have altered the structure of the errors in the estimated concentrations (particular for the smeared and recovered cases). We may still observe that for each of the methods there is a greater fixed error component (leading to large σ values even at low concentrations), and a weaker dependence of σ on concentration compared to the *ME* only case. Appropriate error models would need to be found to fully quantify these differences. Such an effort is beyond the scope of this paper.

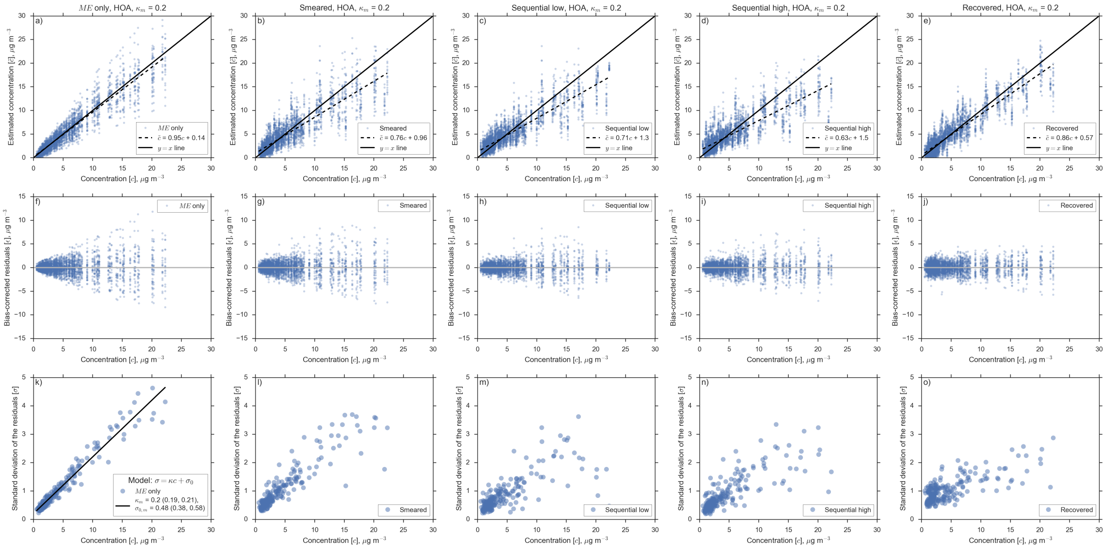

Figure S7. Equivalent bias and error characterization for the Measurement Error (*ME*) only, sequential high and low, smeared and recovered cases. Estimated concentrations $\tilde{c}$ and corresponding true concentration c are taken from HOA time series with T = 57 hours, $\Delta\tau$ = 4 hours and κ_m = 20%. **a – e)** Estimated versus true concentrations. **f – j)** Bias-corrected residuals versus true concentrations. The bias-corrected residuals were calculated by subtracting the means of the estimated concentrations from the individual estimated concentrations at each concentration value (recall from Section 4 that 20 realisations of each measurements signal were generated, creating 20 values of $\tilde{c}$ for each value of c). **k – o)** Standard deviations of the bias-corrected residuals σ versus true concentration. A linear error model is only considered appropriate for the *ME* only case, so linear functions have not been plotted for each of the other cases.

References:

Hansen, P. C.: Analysis of discrete ill-posed problems by means of the L-curve, SIAM Review, 34, 561–580, 1992.

Hansen, P. C.: Regularization tools version 4.0 for Matlab 7.3, Numer. Algorithms, 46, 189–194, doi:10.1007/s11075-007-9136-9, 2007. 11, 12, 16, 17, 22.