

* Reviewer 2

** Comment 1

Increasing the time resolution after staggered sampling of long-time samples using mathematical methods is a well-known but experimentally not practicable method for highest analytical challenges of environmental samples over long field experiment durations. The selected example of 220 hours is a “no go” for staggered sampling because the high number of samplers needed and the enormous costs for analyses of the high number of samples. Often measurement campaigns have a much longer duration so staggered sampling can be used only in very few cases. Therefore I recommend to shorten the discussion of the staggered sampling in favor of sequential sampling. Sequential sampling is a much better possibility in view of the experiment costs how the authors have shown in Fig. 9 and discussed at some different places elsewhere.

- *Response:*

We thank the reviewer for his/her constructive comments, which have helped to improve the manuscript. Our response to each comment is detailed below.

The mathematics of deconvolution are certainly well-known, as recognised by appropriate citations throughout our manuscript. In the field of image processing, a number of studies have used deconvolution techniques to combine low-resolution measurements into higher-resolution products, and in the aerosol community deconvolution is used to invert overlapping transfer functions in size distribution measurements. However, to our knowledge the techniques have not yet been extended to target higher time resolution in environmental sampling. While we have searched for prior literature, we have not found any and would be happy to be directed to the appropriate publications. To our knowledge, the application of deconvolution to staggered measurement samples is a novel contribution.

We also note that the number of samples required by staggered sampling is not actually impractical, and would be determined by the objectives of a study. For the example considered, 223 staggered, 4-hr samples would be required to cover a 220-hour measurement period, which is not a large number of samples considered within the context of past work. For example, several hundreds of filter samples are analysed by FTIR in a typical intensive field campaign (e.g. Russell et al., 2011). Thousands of samples are analysed by FTIR each year for the major US air quality monitoring networks (e.g. Reggente et al., AMT 2016). Therefore, staggering FTIR samples is not impractical, particularly in relation to the cost of expensive online methods for measuring organic aerosol concentrations at high time resolution (e.g. aerosol mass spectrometry), which can easily become prohibitively expensive if sampling at multiple sites is required. These calculations and considerations will obviously change depending on the analysis technique in question, which is why we included Fig. 9 in the manuscript with the statement on P9, L27 that "it is up to the experimenter to decide if the extra practical costs of staggered sampling (e.g. Fig. 9) are worth" the benefits it provides.

Given these arguments, we have retained the discussion of staggered sampling alongside sequential sampling.

**** Comment 2**

After my opinion, the applicability of this method is limited to such regimes like the selected one with clear and distinct daily variations. For a remote sampling site with low differences of concentrations between day and night or a sampling strongly influenced by meteorological changes and rapid changes in emission strength of specific sources this method seems to be unsuitable.

- Response:

We agree with the reviewer that the method likely performs best when daily variations are clear and distinct. However, we note that if pronounced variations are not present, we would have much less interest in investing the effort to resolve them. Although the occurrence of daily concentration peaks in the test data set are consistent, the magnitude and shapes of the peaks vary substantially on a day-to-day basis, and between the two aerosol species. Fig. 1 indicates that the test data includes sharp (high gradients), broad (low gradients), large magnitude, and relatively flat regions. Therefore, we believe it is more likely than not that the methods will be suitable also for other organic aerosol sampling sites. Nevertheless, as explained in response to Reviewer 1's 2nd comment, we have not tested the generality of the results, as such an aim is beyond the scope of a single paper. To further highlight the limitations of the study we have added the following text to the Conclusion:

- Additions to the manuscript:

“These conclusions are based on the two time series we have investigated, which included sharp (high gradients), broad (low gradients), large magnitude, and relatively flat regions (Fig. 1). However, further work is required to test the generality of the conclusions by applying these sampling strategies and post-processing methods to different time series types (e.g. cooking organic aerosols, which may display even sharper peaks in concentrations). The theoretical and modeling frameworks provided in Sections 3 and 4 do not depend on the specific test case in question and can be applied to time series of any variable.”

**** Comment 3**

The here discussed and properly described method is for my eyes as an experimentally working field chemist not more than an interesting mathematical play! When the analytical method doesn't allow to take hourly samples than you have to accept it or you must search for an improvement of the analytical method or apply another one.

- Response:

We thank the reviewer for this opinion and take it very seriously. The primary conclusion of this study is that for the cost of only a relatively small increase in absolute error (quantified as the upsampling error in Section 9), post-processing methods can be used to increase the time resolution of OA measurements requiring long sample collection times. It is not necessary to simply accept a low-resolution measurement, even if one cannot find, obtain or construct an instrument capable of measuring at higher resolution. One of our primary goals is to communicate this result well to experimental scientists who could benefit directly from the result. Indeed, the first-author also identifies himself primarily as a working, experimental scientist.

With this goal in mind, we point out that many modern measurement techniques are in fact a combination of an underlying measurement and a data processing algorithm. For example, submicrometer size distributions measured by a Scanning Mobility Particle Sizer (SMPS) are today considered a standard measurement. Although the modern SMPS user need not know it, SMPS particle size distributions are actually a combination of particle electrical mobility measurements and a deconvolution algorithm. The combination of long measurement samples and post-processing algorithms that we have investigated can be thought of in exactly the same way. Taking this even further, one can characterize the bias and error in the concentration outputs of the post-processing methods in the same way as the bias and error of any new instrument is characterized. We have added these arguments and an example of bias and error characterization to the revised manuscript and supplementary information.

- Additions to the manuscript:

This text has been added to Section 9:

“One way to frame these errors is to consider each combination of noisy 4 – 8-hour measurement samples and post-processing method as a self-contained measurement technique or instrument that measures OA concentrations at hourly resolution. For example, submicrometer size distributions measured with a Scanning Mobility Particle Sizer (SMPS) are typically considered as a standard, self-contained measurement. In fact, SMPS measurements are a combination of particle electrical mobility measurements and an inversion algorithm. SMPS inversion algorithms are analogous to the post-processing methods we have tested here, and are even based on the same underlying mathematics of deconvolution (e.g., Pfeifer et al., 2014), although it is not necessary for the modern SMPS user to know this fact. In this framing, the total error of each hourly resolved OA concentration measurement (*RE*) can be considered as a combination of random error in the underlying measurement (*ME*) and error introduced by the processing algorithm (*UE*). *UE* is the error cost of increasing the measurement time resolution.

Taking this interpretation further, one can also use estimated concentrations to characterize the equivalent bias and error of the hourly-resolution measurements as a whole, analogously to the way bias and error would be characterized for any new instrument. An example of equivalent bias and error characterization is provided in Supplementary Section S5 for the sequential high and low, smeared, and recovered cases considered in Section 7. We have not quantitatively characterized equivalent errors for these cases because Supplementary Fig. S7 indicates that the post-processing methods alter the structure of the errors in the estimated concentrations, and the linear error model described by Eq. 6 is no longer applicable. Therefore, further work would be required to find a more suitable error model and to quantify equivalent error. However, the example still demonstrates how the hourly-resolved outputs of the post-processing methods that we have tested can be treated in the same manner as the output of any given instrument or measurement technique.”

The following example has been added to the Supplementary information:

“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

$$\tilde{c} = \beta_0 + \beta_1 c \quad (\text{S1})$$

where \tilde{c} is concentration estimated by the instrument (a series of values of which comprises an estimated time series $\hat{\mathbf{f}}$), c is true concentration (a series of values of which comprises a true time series \mathbf{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 $\kappa_m = 20\%$, 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 ($< 5 \mu\text{g m}^{-3}$), 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 $\kappa_m = 0.2$ 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.

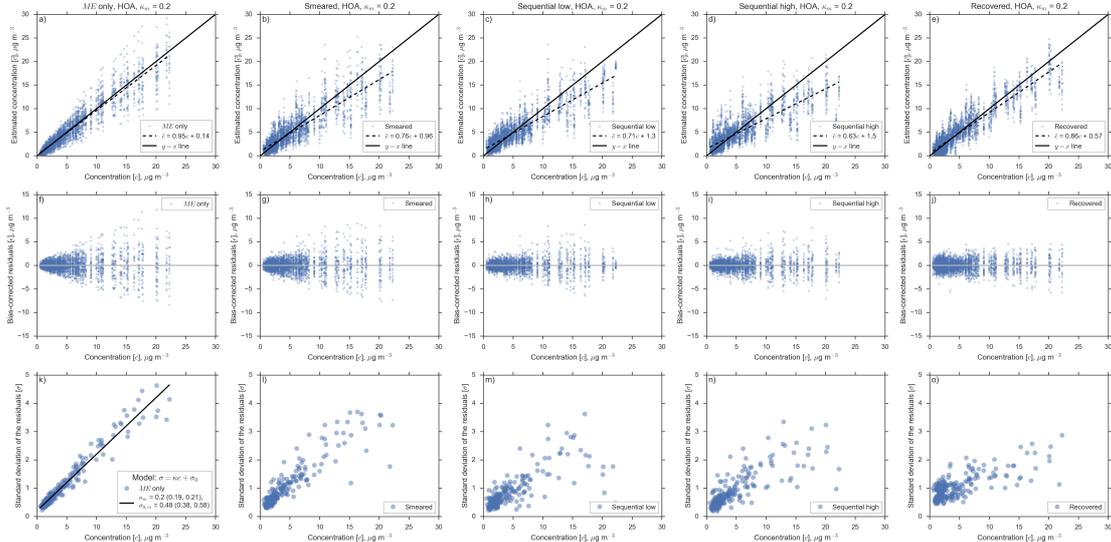


Figure S7. Equivalent bias and error characterization for the Measurement Error (*ME*) only, sequential high and low, smeared and recovered cases. Estimated concentrations \tilde{c} and corresponding true concentration c are taken from HOA time series with $T = 57$ hours, $\Delta\tau = 4$ hours and $\kappa_m = 20\%$. **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. ”

**** Comment 5**

Nevertheless, the calculations have shown that staggered or sequential sampling and the applied model can improve the time resolution of an experimental data set under specific conditions with acceptable low errors but I'm not sure if the method is applicable for another place with more complex influences.

- Response:

We discuss the generality of the results in response to the reviewer's 3rd comment. We limit the scope of this manuscript to suggest possibilities for reducing the time resolution of conventional measurements through new sampling strategies and post-processing algorithms, and anticipate that this work can serve as a starting point for further investigation.

**** Specific and technical comments:**

p.3 l.26: Please be careful with such statements – I don't know one work which has done a full (100%) characterization of the OA.

- Response: We have changed full to "A detailed OA picture..."

p.9 l.1: Only a Fig. 3 is existing, please rewrite this sentence! The x-axis at Fig.3 should be renamed because there is no date.

- Response: Sentence has been rewritten and the x-axis label of Fig. 3 changed to 'Time'.

p.14 l.1: "Filter sample length" you cannot use instead of sampling time, sampling period or sampling duration.

- Response: This has been changed to "filter sampling periods"

p.14 l.17-21: These are two very important sentences! For a real experiment you don't have the possibility to recalculate an optimal starting time you have to decide it before the collection starts. Based on experiences you can try to optimize the starting point but here I have to recapitulate that staggered sampling is in most cases experimentally not practicable and too expensive as well.

- **Response:** We agree with that this is an important point, which is why we investigated and quantified the effect of filter start time on sequential sampling error (Fig. 6g).

p.15 l.5: You discuss here OOA and HOA but not aerosol mass – these are different things.

- **Response:** Indeed it is not possible to measure HOA and OOA concentrations directly with FTIR. Our model could represent, for example, the measurement of organic functional groups by FTIR that represent HOA and OOA. This point has already been discussed in detail on P7 L24 to P8 L9.

p.15 l. 29 and p.18 l.23 Please, use only one language for dependance!

- **Response:** We have changed all instances of dependance to dependance.

p.25 l.17. In the figure caption Fig. 11 it is correctly described that 4 h samples were discussed but here in the text it should be added, too.

- **Response:** Added as suggested.

References:

Reggente, M., Dillner, A. M. and Takahama, S.: Predicting ambient aerosol thermal–optical reflectance (TOR) measurements from infrared spectra: extending the predictions to different years and different sites, *Atmos. Meas. Tech.*, 9(2), 441–454, doi:10.5194/amt-9-441-2016, 2016.

Russell, L. M., Bahadur, R. and Ziemann, P. J.: Identifying organic aerosol sources by comparing functional group composition in chamber and atmospheric particles, *PNAS*, 108(9), 3516–3521, doi:10.1073/pnas.1006461108, 2011.