* Reviewer 1

** Comment 1

This paper deals with post-processing methods for mimicking the determination of time-resolved concentrations of atmospheric organic aerosols (OAs) starting from time-averaged measurements. This is an important issue because, even if it has been acknowledged that the existing spectroscopic methods for OA measurement provide complementary information, the actual applicability of some offline methods (e.g., FTIR, NMR) is limited for not providing the desired time resolution. The paper presents a comprehensive theoretical discussion of a case study, an AMS field campaign, where a hypothetical time-averaged measurement series was produced based on a set of sampling strategies and post-processing methods which were then compared and evaluated.

- Response:
We thank the reviewer for his/her succinct summary of the work, and for the constructive comments that follow. These comments have helped to improve the manuscript. Our responses to each comment are listed below.

** Comment 2

The authors conclude that the simple linear interpolation of sequential samples of 4 h duration provides the best approximation of a time-resolved (hourly) timeline of observations for many practical purposes, although some uncertainty still relies in the actual sampling strategy (how much start/stop times overlap with peak concentrations). This is a useful recommendation, even if, I believe, it is not straightforward to generalize it to all possible real sampling conditions. The Mexico City campaign was characterized by very pronounced, consistent diurnal variations in the concentrations of HOA and OOA (Figure 2), and simply most of this variability can be captured by a 4 h sampling strategy. In areas where diurnal cycles are more complicated (e.g., lunch-time peak of cooking aerosols) or meteorological conditions are more variable (marine/coastal sites), I am not sure the Authors would have found the same conclusions.

- Response:
We thank the reviewer for recognizing that this work results in useful recommendations. We agree that we would not necessarily come to the same conclusions with other types of time series. As suggested, generalization is not straightforward, and therefore such an aim is beyond the scope of a single paper. This is why we have chosen a test case. On P7, L14 we discuss the rationale for choosing this particular HOA and OOA case. The daily peaks of the two time series are separated by only a few hours, so it is important to measure the time series at high time resolution, to help resolve the primary and secondary origins of these two aerosol species. We would also like to add that although the occurrence of the daily peaks is 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 set includes sharp (high gradients), broad (low gradients), large magnitude, and relatively flat regions. Therefore, we think that this is a well-chosen example. Furthermore, we have provided a consistent theoretical and modeling framework that

can be easily extended to time series of other variables. We have added the following text to the Conclusion to make these points more clearly:

- Additions to the manuscript:
The following text has been added to the Conclusion.
"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
I suggest to include the average concentration calculated throughout the full measurement period as an additional diagnostic, because some post-processing methods seem to qualitatively capture the time trend but not the absolute amounts respect to the true time series (Figure 5d,e).

- Response:
We calculated average concentrations over the full measurement periods corresponding to the results presented in Sections 5, 6, and 7. In almost all of the cases, except the one that we discuss below, average concentrations were reproduced to within 10% (and very often to within only a few percent).

The sole exception is the deconvolution result presented in Section 7. Concentrations recovered by Tikhonov regularization were found to be systematically lower than the true concentrations, resulting in average concentrations 10 - 20% below the corresponding true values. We have included a new Figure in the supporting information to highlight this result. This calculation further strengthens our recommendation of TSVD over Tikhonov regularization, and we now point out this fact in Section 6 and the Conclusion.

- Additions to the manuscript:
The following Figure has been added to the Supporting Information:
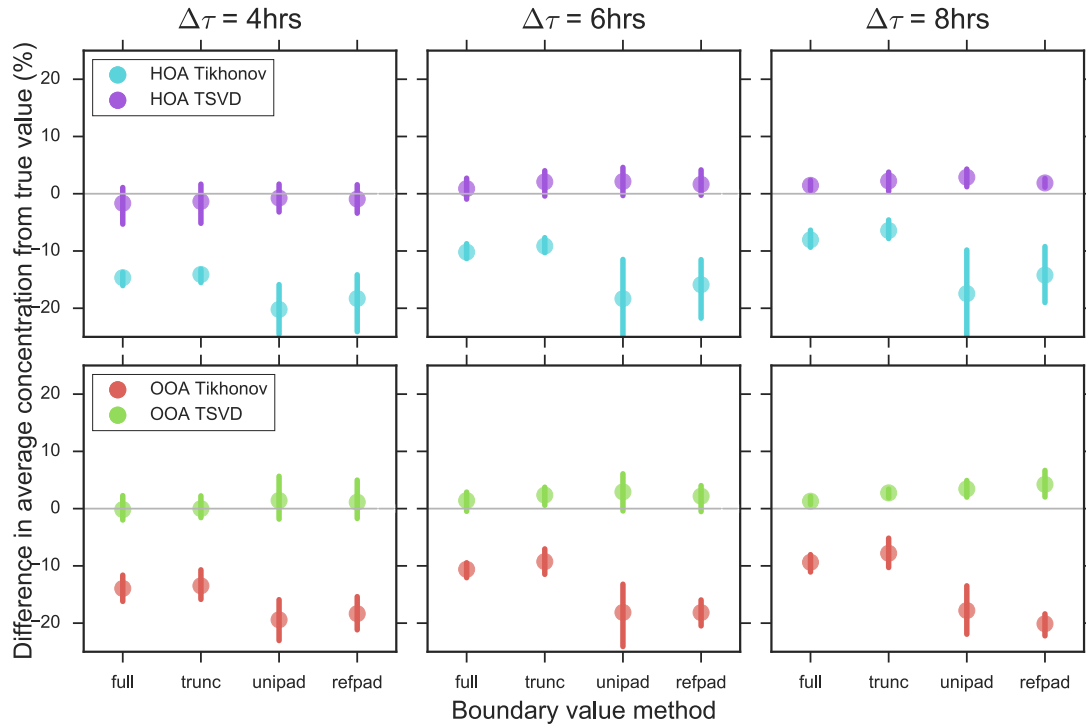
Figure S6. Percentage deviations of average concentrations over the full time series from the true values for different boundary value methods applied to HOA and OOA time series constructed by deconvolution with TSVD and Tikhonov regularization of staggered measurements of length ($\Delta\tau$) 4, 6, and 8 hours. $\kappa_m = 20\%$ 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.

The following text has been added to Section 6:
"Although the *RE*s are similar, concentrations recovered with Tikhonov regularization are generally lower than the true concentrations. As a result, the overall average concentrations of time series recovered with Tikhonov regularization are 10 - 20% below the corresponding averages of the original time series. The average concentrations of the time series recovered with TSVD regularization are very similar to the true values (Supplementary Fig. S6)."

And the result has also been referenced in the Conclusion:
"If a deconvolution algorithm is applied, we recommend using TSVD regularization because it resulted in more accurate average concentrations over full sampling periods, and marginally better peak capture and *RE*s than Tikhonov regularization."

** Comment 4
The regularization parameters remain a source of error that cannot simply constrained in a real case scenario, when the true time series is (obviously) not known. Are there any recommendations for the selection of k and l in the TSVD and Tikhonov methods beside looking at when the solution provides a "reasonable" time trend?

- Response:
Yes, we have added the following discussion to the Supplementary section, which references previous work on this topic and describes two common parameter choice methods. This Supplementary section is now referred to in the discussion of regularization parameter choice in Section 4.

- Additions to the manuscript:
The following text has been added to the Supplementary information
"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 ($\|g\|$) and residual $\|H\hat{f} - g\|$ 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 g, and choosing the regularization parameter that best predicts the left-out observations."

** Comment 5
Figure 6 shows that the error associated with sequential sampling increases with increasing sampling duration. Here, it is assumed that measurement errors are invariant with sampling duration except for the effect of mass loading approaching the detection limit. In the real world, this is complicated by sampling artifact effects. A number of studies have shown that extending the sampling time allows for a better equilibration of vapors absorbing into the filter matrix (e.g., Kirchstetter and Novakov, Atmos. Environ. 2001, 1663-1671), therefore, ME would decrease with Dt faster than it is supposed in the present study for many practical purposes (but depending on the technique and on the substrate, actually).

- Response:
This is an important point and we thank the reviewer for raising it. Firstly it is important to clarify the difference between systematic measurement artefacts and random measurement error, and how we account for these two concepts in our numerical model. We assumed that the random measurement error (σ) has both a fixed component (σ_0) and a relative component depending on concentration ($\kappa*c$). The random perturbations due to measurement error were assumed to be normally distributed about 0 with a standard deviation of σ (Eq. 6). The centering of the normal distributions about a mean of 0 amounts to assuming that there were no systematic measurement artefacts. We made this assumption because as suggested, measurement artefacts depend strongly on the measurement technique in question and even the specific materials used (e.g. batch lots of filters) and are therefore difficult to model. If known, measurement artefacts for a specific technique could be addressed by assuming a non-zero mean of the normal distribution in Eq.6. The time dependance discussed by the reviewer could be addressed by making the non-zero mean sampling time dependant.

- Additions to the manuscript:
We have added the following discussion to Section 4 to make our assumption of no measurement artefacts explicit, and to discuss how non-zero, time dependant sampling artefacts could be addressed in the model:

"By setting the means of the ε distributions to 0 we have assumed that the simulated measurements are not affected by systematic measurement artefacts. Systematic measurement artefacts depend strongly on the measurement technique in question and even the specific batch of materials used (e.g. filter lot). They can be positive or negative, and can depend on sampling time (e.g., Kirchstetter et al., 2001; Subramanian et al., 2004). If known, measurement artefacts could be addressed in this modeling framework by the setting the means of the ε distributions to non-zero, time-dependant values."

** Comment 6
Page 10, line 6: "The temporal resolution of f is dt, the temporal resolution of g." There is something missing in this phrase. (Dt?)

- Response:
We have modified the phrase to "The temporal resolution of f is the same as that of g (i.e. δτ)."