

Response to Reviewer #3

We thank Reviewer#3 for his/her helpful comments on our AMT discussion paper. We carefully addressed all comments and accounted for them in our paper as stated below. The original comments of the reviewer are cited in *italic font*, our response is put below each comment in standard font. The changes stated below are also highlighted in the Revised Manuscript (RM).

1. In my understanding, model analyses data should not be directly involved in the statistical optimization in principle. In this study, the model analyses are used to estimate the “unbiased” background bending angle and the observation error correlation matrix. In addition, the same model analyses data are used as references to evaluate the performance of proposed SO method statistically.

We agree with this understanding that model analysis data should not be directly involved in the statistical optimization in principle. This is the very reason why we use co-located short-range forecast data as the primary ingredient of background profiles and why we are very careful to use analysis data only in an indirect way and in a form of certain statistical quantities. We do this in order to keep their influence on the statistical optimization results arguable beneficial in a statistical sense (bias-calibration step, Eq. 5 in the paper; involving a large ensemble of analysis profiles, leaving negligible dependence on the co-located analysis profile) or otherwise negligible (as is the case for the purely technical use of analysis data for enabling subtraction of smooth/noise-free “zero order profiles” in order to construct correlation matrices; section 2.2 in the paper).

For example, using forecast profiles instead of analysis profiles for subtracting a suitable smooth state in correlation matrix construction does not induce any relevant change to the resulting correlation matrix (which contains correlation functions normalized to unity in their diagonal). Also, the construction might include an additional step towards delta-difference profiles (according to Eq. 8 in the paper) before these are then used for correlation matrix generation. This makes the difference profiles essentially independent from the choice of the original smooth state (analysis, forecast, or any similar profile that provides a reasonable smooth state for being subtracted), but again the changes in the resulting correlation matrix are in practice negligible.

In the performance evaluation, the co-located analysis profiles just serve as a convenient joint reference in this paper, and we therefore do not “over-interpret” the results from the real observed data in the sense that we would claim independence from this joint reference; in on-going work for a next paper (including climatological use over many months) we additionally employ independent reference data over the stratosphere and mesosphere, which makes the evaluation more rigorous and we can derive even stronger conclusions that consolidate and complete the introduction of this new SO approach.

2. Obviously the proposed SO method is not applicable for the (near) real-time RO data processing. It needs to be clarified in the paper, especially when you compare your SO method with those methods designed for real-time RO data processing.

It is applicable for near real time processing. But it is true that we did not mention that for this application we suggest to build the empirical covariance matrices from the previous week (i.e., using the seven-day-history data $d_{oi}-1d$ to $d_{oi}-7d$ instead of the centered-time-window data $d_{oi}+/-3d$, where d_{oi} is the day of interest that contains the occultation events to be processed). Use of the seven-day-history data is justified, with very slight degradation in performance if any, since the dynamical change of the matrices over time is practically very slow (as indicated by Figs 2 and 3 in the paper and discussed in the related text).

In order to point to the applicability also in near-real-time and fast-track processing applications we now added after the last paragraph of section 2.2: “These slow dynamics of the observation error covariance matrices, and of the background error covariance matrices as discussed in Section 2.1, enable reliable use also in near-real-time or fast-track processing (i.e., processing within 3 hours or within follow-on day of observations). Instead of using seven days centered about the day being processed (including three days before and after the center day) seven-day-history data (from the previous day to seven days ago) may be used in these cases, with insignificant degradation in performance.”

3. It would be important to understand what information comes from the pure observation, and what information is affected by the background in optimized RO bending angle. So could you please present the histograms of the median weighting height where the background and observation bending angle uncertainty are same for different SO methods.

We agree that this is an important aspect and we carefully considered how to best account for it in this paper without unduly expanding it (also based on a comment received in the quick review towards the AMTD paper). We decided for a twofold improvement in the revised manuscript:

1) we included explicit illustration of the observation-to-background weighting ratio (r_{obw}) for the three representative occultation events shown in Fig. 6 (left column). These cover typical simMetOp, CHAMP, and COSMIC r_{obw} ratios and provide the readers with a good impression of how the information is typically weighted between observations and background for settings as used in this paper. The $r_{obw} = 100 \cdot u_b^2 / (u_b^2 + u_o^2)$ [%] is a simply and convenient approximate variable for the purpose, since it expresses how the weighting of observation-to-background information would be if the covariance matrices in the statistical optimization equation (Eq. 2 in the paper) were diagonal.

The related improved text of Section 3.1, describing Fig. 6, now reads as follows:

“Figure 6 illustrates the effects of statistical optimization on individual bending angle profiles by a few representative RO events. The left panels show the background and observation uncertainties as well as the observation-to-background (Obs-to-Bgr) weighting ratio $r_{obw} = 100 \cdot u_b^2 / (u_b^2 + u_o^2)$, which expresses on a percentage scale how the

information is weighted between observations and background. r_{obw} is a convenient approximate variable for the purpose, which exactly applies if the covariance matrices in the statistical optimization equation (Eq. 2) were diagonal. Observation uncertainty is smallest for the simMetOp event (top), largest for the CHAMP event (middle), and in between for the COSMIC event (bottom). At 60 km these observation uncertainties are roughly 0.4 μrad , 3 μrad , and 1.5 μrad for simMetOp, CHAMP, and COSMIC, respectively. These differences in observation uncertainty yield the largest r_{obw} for the simMetOp event and the smallest for CHAMP. Related to this, the altitude where both observations and background receive equal weight ($r_{\text{obw}} = 50\%$), which may be considered the transition altitude below which the observation information dominates the retrieval, is highest for simMetOp (>60 km), medium for COSMIC (around 55 km), and smallest for CHAMP (near 45 km). Ongoing follow-on work on large RO datasets of many months analyzes the r_{obw} statistically and confirms that these few example events are typical for the respective data sources.”

On histograms of r_{obw} , we decided not to add yet another figure to this paper (on top of illustrating r_{obw} profiles now in Fig. 6 as noted above) but leave this to the follow-on paper with bigger and longer-term ensembles of data. From the updated Fig. 6 we can see that the altitude where r_{obw} equals 50% (i.e., the altitude where the background and observation uncertainty are the same so that both background and observation receive equal weight) is over 60 km for the simMetOp event, near 45 km for CHAMP event, and around 55 km for the COSMIC event. We also have performed some internal sensitivity test with preliminary histograms of r_{obw} , which indicate similar typical $r_{\text{obw}}=50\%$ altitudes for the new SO method (around 75 km for simMetOp, around 45 km for CHAMP, around 55 km for COSMIC, with the linear $f_{\text{bc}} = 1$ to 15 as applied here); existing SO methods have “less well behaved” results, due to a higher degree of ad-hoc construction of the optimization, or the r_{obw} is unknown.

2) we made further sensitivity tests for the bias coverage factor f_{bc} —that is available to penalize the background bias uncertainty according to how users of the algorithm decide is best serving their application (Eq. 3 in the paper)—and modified its use and improved the related description in the paper. In the AMTD manuscript, we had introduced $f_{\text{bc}} = 5$ as a reasonable choice, applied over the whole altitude range from 30 km (28 km) to 80 km, pointing to the option that this may be considered somewhat large for background uncertainty at high altitudes, and somewhat small for background uncertainties at lower altitudes. We now therefore cross-checked cases with linearly increasing f_{bc} with altitude and decide to show the results for a case with linear variation of f_{bc} , with $f_{\text{bc}} = 15$ at 30 km (3 times more than $f_{\text{bc}} = 5$) to $f_{\text{bc}} = 1$ at 80 km (no bias penalty at the top boundary of the altitude domain). Based on this modification, all the related description and result figures were updated accordingly. As a reference we also computed a case without any penalty (i.e., $f_{\text{bc}} = 1$ at all altitudes), reflecting use of the algorithm as an optimal estimation between observations and background without any user intervention to down-weight the background for the sake of avoiding potential background biases. In order to illustrate the difference between the three settings $f_{\text{bc}} = 1$, $f_{\text{bc}} = 5$, and $f_{\text{bc}} = 1$ to 15 (from 80 km to 30 km) a new Fig. 9 was included, showing the differences. Altogether these

improvements make the role of the bias coverage factor, as the one parameter of the algorithm available to the user for influencing the degree of background information flowing into the optimization, more clear and transparent.

The modified text in Section 2.1, introducing f_{bc} below Eq. 3, now reads as follows:

“Herein the bias coverage factor f_{bc} is introduced as a user-defined parameter that can be employed to penalize the estimated bias-type uncertainty u_b^k relative to the estimated random uncertainty s_{f-a}^k . This enables to minimize the influence from potential residual background biases, relative to observation uncertainty, on the resulting optimized profile α_s^k . In this study the bias coverage factor was chosen to linearly decrease with altitude, setting $f_{bc} = 15$ at 30 km (strong penalty in lower stratosphere) and $f_{bc} = 1$ at 80 km (no penalty at top boundary). This choice was found to be useful for climate applications (more discussion of f_{bc} , including for comparison also example cases with constant $f_{bc} = 1$ and $f_{bc} = 5$, is given in Section 3.2).”

Subsequently, the related new text of Section 3.2, describing Fig. 9, reads as follows:

“In order to discuss the effects of different choices of f_{bc} on the resulting optimized bending angles, we compared the choice of this study for linear altitude dependence (see Section 2.1; termed $f_{bc} = 1\text{To}15$ here) with the choice in the algorithmic introduction by Li et al. (2013) ($f_{bc} = 5$) and with a reference case intentionally making no use of the bias penalty option ($f_{bc} = 1$).

Figure 9 shows the comparative results for these three f_{bc} choices in the same statistical result format as used for Fig. 8; the results shown for “Dynamic $f_{bc} = 1\text{To}15$ ” and OPSv5.6 replicate the ones of Dynamic and OPSv5.6 of Fig. 8 for context. It can be seen that for simMetOp the $f_{bc} = 5$ and $f_{bc} = 1\text{To}15$ results are rather similar, both for systematic differences and standard deviations, whilst the $f_{bc} = 1$ choice indicates how a strong un-penalized weight of background profiles forces the optimized solution towards the background at high altitudes (here visible above about 45 km). As is to be expected, any residual biases in the background will therefore best survive in the optimized solution in case of an $f_{bc} = 1$ choice. This is the very reason why the background is intentionally penalized if the dynamic scheme is employed for climate-quality retrievals that strive to minimize background influence.

For the CHAMP and COSMIC data, standard deviations are largest for the $f_{bc} = 1\text{To}15$ case, medium for the $f_{bc} = 5$ case, and smallest for the $f_{bc} = 1$ case. This is in line with the expectation of how the r_{obw} changes under these different choices, since the observational noise is more and more mitigated the more relative weight the background receives. The effect on the systematic differences (for CHAMP and COSMIC relative to the co-located ECMWF analysis profiles) is relatively small in these global-scale statistics; but also here the direction is that no bias penalty forces the results towards the background mean state at high altitudes.

Overall Figure 9 demonstrates that the sensitivity to the detailed quantitative choice of f_{bc} is fairly weak, as can be seen from the moderate differences between these three cases with very different f_{bc} choices. This is favorable since it implies that not any detailed quantitative tuning of f_{bc} is needed in practice; the f_{bc} as the only free user-defined variable in the new dynamic algorithm is rather a clear and transparent option to

predefine the influence of background information according to what users deem suitable for their application.”

4. As mentioned in section 2.1, the authors updated the calculation of the mean variable in each grid cell by averaging the variable over a longer temporal period (7 days) and a larger geographic region (1000 km x 3000 km) compared to the b-dynamic scheme. The choice of average domain is a trade-off between in capturing the mean field variations and in removing short-term/random variations. So could the authors please say few more words about what kind of improvement in performance get from this update? If more RO data available per day in the future, would it impact the choice of the average?

After applying this update, we found that the mean background and observed variables in each grid cell were somewhat smoother than those obtained using the previous settings, while the magnitudes of the variables stayed essentially the same. This means random variability was a bit more suppressed (ok) while the mean field variation was still similarly captured (ok). If more RO profiles become available per day in future this is favorable, since random errors are averaged out even better. Combined with improved residual ionosphere bias correction in future (as discussed in the response to reviewer #1) this enables more weight on the averaged-observed bending angle in the mean background profile, leading to better background-bias calibration (Eq. 5 of the paper). In addition, the empirical construction of the observation error covariance matrix can gain further robustness.

We improved the relevant text in section 2.1 as follows: “...this update allows more data to be used for more reliable statistical estimates (especially important for mean observed bending angles) at each 10° latitude $\times$ 20° longitude grid point, while still capturing the slow variations of the mean field well.”

5. As mentioned in section 3.2, the authors applied different quality checks for RO data retrieved with different SO schemes. I would like to suggest the authors to make sure the same dataset used for comparison.

We confirm that we were careful in this respect. In the quality control, the quality of the input RO bending angles (ionosphere-corrected bending angles), optimized bending angles, refractivity, and dry temperature is checked in various ways. If for the same RO event, any of the quality flags of any SO scheme showed a bad quality, we rejected this event from the statistical inter-comparison. Therefore we can confirm that we used the same datasets (same ensemble of occultation events) for all SO schemes involved for the statistics.

Many thanks again to Reviewer #3 for his/her valuable comments that helped us to further improve our manuscript.