"Development and characterisation of a state-of-the-art GOME-2 formaldehyde air-mass factor algorithm" by Hewson et al. [2014]

Response to Reviewer 3

General Comments:

The paper by Hewson et al. presents an improved algorithm to compute AMFs for GOME-2 HCHO retrievals. The algorithm is not particularly innovative but it brings together a number of incremental improvements from the recent literature and evaluates their effects on the AMF computation in a consistent framework including error estimates. The paper is well written. This is in my opinion a valuable contribution worth publishing in AMT.

We thank the reviewer for their supportive and also insightful comments. We have provided detailed responses to comments raised about the AMF error analysis. We also have addressed all other specific issues and have adjusted the manuscript accordingly where appropriate.

I have one major comment that I think is critical to address. I am troubled by the error characterization presented here and the lack of comparison to the work of Millet et al. JGR2006 (cited in passing but not in that context). The authors find that the shape factor is the dominant source of AMF error but this is based on seemingly arbitrary +/- 25% perturbations to the HCHO profiles – without better characterization of the error on the HCHO profiles it seems to me that the inferred AMF error is of little value.

The study of Millet et al. [2006] is a valuable contribution in understanding AMF uncertainties, particularly as the in situ aircraft measurements of HCHO, aerosols and clouds, allowed key insight into their relative distributions and properties. However, there are two main limitations of that study: (1) it is only focused over US region and (2) it only focused over a small time period (1 July – 15 August 2004). Over North America, the GEOS-Chem simulation of tropospheric chemistry and HCHO performs particularly well for a number of reasons, one of the most important being the prevailing high NOx conditions which ensure HCHO production from isoprene oxidation is adequately described. This is demonstrated by the good agreement of the model and aircraft HCHO profiles shown in Figure 2 of Millet et al. [2006].

However, the results of Millet et al [2006] cannot (and should not) be readily extrapolated to cover all geographical regions. Over tropical regions, where the combination of high isoprene emissions, low NO_x conditions, and poorly parameterized VOC oxidation schemes (in low NO_x conditions), contribute to significantly more uncertainty in the simulation of HCHO distributions. For example, GEOS-Chem model comparisons to surface, tower, and aircraft observations over tropical South America, indicate the model struggles to reproduce isoprene and its oxidation products (MVK, MACR) and specifically HCHO; see Barkley et al. [2008, 2011]. In particular, Barkley et al. [2011] show a comparison of various GEOS-Chem simulations to GABRIEL aircraft observations over the Guyana's and surrounding area. Model biases in lowest 2 km of the atmosphere were of order 11-100% using a standard chemical mechanism and -12% to 49% using a more advanced Caltech chemistry scheme (see data presented in Table S2 of that work). Therefore our assumption of +/- 25% error in the HCHO profile ***is not arbitrary*** *but based on comparison to real data in a region where the model simulation is particularly poor. These high biases are a combination of very uncertain isoprene emissions and subsequent poorly described isoprene chemistry. Furthermore, they are not specific to GEOS-Chem alone, as most chemistry-transport models generally struggle with simulations of tropical tropospheric photochemistry.*

In summary, we assigned a default 25% error to the HCHO profile as it seemed more appropriate to use a higher uncertainty to cover likely errors in HCHO distributions found on a global scale. ***We have made it clearer in the text that this choice is not random but based on published work. We have also now examined the impact of 10% profile scaling on the AMF error (please read further).***

Even more of concern, the finding disagrees with the previous work of Millet et al. , which provided in my opinion a much better estimate of AMF error because it was based on a large collection of aircraft vertical profiles over the eastern US in summer including simultaneous HCHO, cloud, and aerosol

measurements. They reported a much smaller AMF error (15-24%, vs. 50-60% in this paper), and found that clouds were by far the largest source of error, which makes a lot of sense to me (and even more so for GOME-2 because of its large pixel size). By contrast, the authors here find clouds to be only a small source of error (smaller than aerosols, which makes no sense to me), but it's not clear to me how they estimated the cloud error. Cloud errors seem to get short shrift in this paper – they are discussed less than aerosol errors, which should be much smaller.

Our approach for estimating the AMF errors associated with a global trace gas retrieval is consistent with the standard methodology presented in Boersma et al. [2004], De Smedt [2011], and De Smedt et al. [2008, 2012]. Our GOME-2 AMF error estimates ***are consistent*** *with published estimates of De Smedt et al. [2012] indicating our presented results are in line with other leading retrieval groups.*

However, Barkley et al. [2012] also assessed SCIAMACHY and OMI HCHO AMF uncertainties over tropical South America, using a more 'brute force' approach by systematically varying input parameters and determining the corresponding change in AMFs and vertical columns. In that analysis, the cloud fraction was found the dominant individual error producing changes in the AMF over land of +/-20% for SCIAMACHY and 30% to -16% for OMI. We have therefore have now included extra discussion in section 6 on AMF cloud errors, using the approach of Barkley et al. [2012], to highlight their significance (see also our response to reviewer 1). However, even so, the overall effect of clouds on GOME-2 AMFs appears small.

I think that it is essential that this paper carefully compare their results to the experimentally-based AMF error analysis by Millet et al., and if they disagree to give a solid reason why. Millet et al. should be an important reference for this work. In particular there has to be a resolution of why this paper finds a 50-60% error, vs. 15-24% in Millet et al.

Whilst an AMF error of about 30% may be appropriate for the US (based on Millet's analysis), we firmly believe and have provided firm evidence that an AMF error of 50-60% is more applicable at the global scale, given the ***realistic uncertainties*** *in simulations of tropical chemistry.*

However, to reconcile the differences between our analysis and the Millet et al. [2006] study we have also conducted an extra sensitivity study in which the HCHO profile error has been recalculated by rescaling the profile by 10% instead of the default 25% (as previously done in Section 6). As a result the ***total AMF error drops to about 30%****, which is more consistent with the Millet et al. [2006] study.*

We have now included these extra profile estimates in the paper within section 6, and have indicated that over regions where HCHO chemistry is well-described (e.g., the US) AMF errors from the HCHO profile will likely be smaller, and clouds error more relevant. This has also been included in the abstract to emphasize this point.

Specific comments (page, line):

(1110, line 16) The authors advertise their computation of AMF errors for individual scenes. However, for individual scenes, isn't the spectral fitting error by far dominant? In my experience, this is why HCHO satellite data are reported only as monthly or seasonal averages. The spectral fitting error averages out but AMF errors tend to be more systematic. My point is that I'm not sure that an estimate of AMF error for individual scenes is valuable. Maybe the authors could give a better argument of why they think it is.

Given its weak spectral signature, spectral fitting errors are dominant especially over regions of low HCHO levels. However, over HCHO regions, the AMF error is non-negligible and can be comparable to the spectral fitting error (i.e. of order 20-100%). Furthermore, as shown in Barkley et al., [2012] AMF errors are not systematic but vary spatially and temporally, according to the accuracy of the key AMF input parameters. Individual AMF error estimates are valuable as they provided a more robust error characterization of the HCHO vertical columns, this in turn allows proper observational weighting when grid averaging and to also properly calculate the errors of top-down VOC emissions. This statement has been added to section 6 to emphasize to the reader the importance of AMF error estimates.

(1110, lines 17-18) I am skeptical of the 50-60% error and of the dominant contribution from the HCHO profile, for reasons stated above. If that statement is to remain in the abstract it needs to be better supported.

See our detailed responses above.

(1111, 4) It would be good idea to cite Palmer et al. JGR 2003 since they were the first to infer isoprene emissions from HCHO satellite observations**.**

Reference added.

(111, 21) I don't understand the 're-normalisation'.

Re-normalization is simply the adding back of the mean GEOS-Chem model vertical column background over the Pacific Ocean, once the slant column bias determined over the same Pacific region has been removed. This is step described in detail in our previous paper, Hewson et al. [2013]. This is a common technique applied in many HCHO retrievals (e.g., Barkley et al., [2008,2013], De Smedt [2011], De Smedt et al. [2008,2012], Marias et al. [2012], Fortems-Cheiney et al., [2012], Gonzalez Abad et al., [2015]). This is done to all GOME-2 observations on a daily and latitudinal basis. We have rephrased the text to make this point clearer.

(1112, 7) It makes no sense to me that aerosols would be a larger source of error than clouds. I believe that the 50% aerosol effect comes from an extreme case of an overhead biomass burning plume and is not really representative. Millet et al. 2006 should be prominently cited in these error estimates – it seems to me that they did a much better job at quantifying the AMF error than anyone else.

It is very likely that the greatest AMF error comes from high aerosol loading above the HCHO peak, typically associated with biomass burning plumes. However, these are indeed ***localized events*** *(but which can still cover large regional areas). We acknowledge this in section 6.2 to avoid confusion. Also Millet et al. [2006] is now thoroughly cited in the error discussion.*

(1112, 11) I would think that model PBL depth is the largest source of error in the HCHO profile. This should probably be mentioned.

The importance of the PBL depth as a significant source of error in the HCHO profile is already acknowledged in Section 6 (page 1127, line8).

(1114, 9) References given for surface albedo contribution to AMF error are for NO2. NO2 is more sensitive to surface albedo error than HCHO (longer wavelengths), as pointed out by Palmer et al. 2001. This should be acknowledged.

Now acknowledged.

(1114, 17) This section talks about surface albedo, HCHO profiles, aerosols, and LUTs as sources of AMF error but doesn't mention clouds.

Apologies. This is indeed missing. A short paragraph detailing likely AMF uncertainties due to cloud fraction and cloud top pressure are now included in this section, with particular references to the works of Millet et al. [2006], Boersma et al. [2004], and De Smedt et al. [2008].

(1120, 17) I would have expected the fine distribution of clouds to be a major factor of difference in the AMF calculation for individual scenes between model simulations at different resolutions. Maybe this gets washed out in temporal averaging, but if the goal is to have an error estimate for individual scenes…

The cloud fraction and cloud top pressure are the same for each set of AMF computations, even if the GEOS-Chem model resolution changes. It is not clear what the reviewer is implying here.

(1128, 17) The statement that biogenic emissions greatly affect the shape factor seems weird. Considering that almost all the HCHO is in the PBL in any case, it seems to me that the shape factor would be largely insensitive to the magnitude of biogenic emissions and instead respond most strongly to the PBL depth.

The analysis of SCIMACHY and OMI AMFs over South America, by Barkley et al., (2012) clearly

demonstrates surface isoprene emissions significantly affect HCHO AMFs by influencing the profile shape. The changes in HCHO VCDs due to differences in the HCHO profile shape can be up to +/- 40% locally but mostly are typically within +/-20%, i.e. still a non-negligible amount. Figure 6 of Barkley et al. (2011) and Figure S2 of Barkley et al. (2012) clearly show how the HCHO profile shape is affected by different emissions models (MEGAN vs LPJ-GUESS) and from different chemical mechanisms. In short, the magnitude and spatial-temporal variability of VOC emissions influences subsequent tropospheric oxidation chemistry and thus how HCHO is produced & distributed. ***Hence the underlying assumption that the shape factor is largely insensitive to BVOC emissions is incorrect.*** *However, we do acknowledge that the HCHO profile shape is affected both by the BL height and BL mixing. The latter effect was also assessed in Barkley et al., (2012), using full versus non-local BL mixing schemes. HCHO VCD differences resulting from the two schemes were about +/- 10% for SCIAMACHY and +/- 20% for OMI. Again, differences in the HCHO profile from the two BL mixing schemes are shown in Figure S2 of Barkley et al. (2012).*

(1130, 6) I question the punchline in the summary about the critical importance of the vertical profile; see my major comment above.

There is strong evidence that uncertainty in HCHO vertical distribution is one of the main AMF error sources; see e.g., De Smedt [2011], De Smedt et al [2008, 2012], and Barkley et al. [2012]. However, we However, we acknowledge that we have not stressed enough that that this large error is in part due to the relative HCHO vertical distribution to cloud height and aerosols. We also accept that we have not stressed enough the AMFs uncertainties associated with errors in cloud fraction and height. ***We hope that the revised analysis and refined main take home messages in the latest manuscript now properly highlight this error source and its impact.***