

Response to reviewer #2 comments to manuscript:

Curci, G., Palmer, P. I., Kurosu, T. P., Chance, K., and Visconti, G.: Estimating European volatile organic compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument, *Atmos. Chem. Phys. Discuss.*, 10, 19697-19736, doi:10.5194/acpd-10-19697-2010, 2010.

R = Reviewer's comment

A = Authors' reply

The authors thank reviewer for his/her careful revision and thoughtful comments which we believe improved our manuscript. In particular, we improved our "sensitivity tests" section on the robustness of our inversion analysis. In the related revision process we also found a bug in our routines related to calculation of continental totals of averages: emission monthly totals in Table 2 and Figures 5 and S2, and column averages in Table 1 are corrected in the current version of the manuscript. In view of new analysis developed in section 4.2 and 4.3, we also revised to a more conservative statement the final comments on the robustness of Bayesian inversion. Detailed response to your comments are enclosed below.

R2.1) 19702, L27. You should discuss the potential effects of using α -pinene as a surrogate for all monoterpenes. In effect you're assuming that all monoterpenes have a HCHO yield equivalent to that of α -pinene.

A2.1) Right, we are assuming that all monoterpenes have a HCHO yield equivalent to that of α -pinene. Atkinson and Arey (2003) reviewed HCHO yield estimates for several monoterpenes. α -pinene have a HCHO yield from reaction with OH of about 0.21 per C reacted. Other major monoterpenes with estimated HCHO yields from reaction with OH: β -pinene 0.50, myrcene 0.3, 3-carene 0.21, terpinolene 0.29. Limonene has an estimated HCHO yield from reaction with ozone of about 0.14.

Among monoterpenes with appreciable amount of HCHO as oxidation product, α -pinene is thus at lower end of yield range. Other monoterpenes don't have an estimate of HCHO yield, or they only have appreciable HCHO formation from ozonolysis, which is generally slower than reaction with OH and thus having more spatial smearing of their "HCHO signal".

The effect of the oxidation of monoterpenes may thus be underestimated, but, as can be seen in Figure 3 and as explained in section 3.3 their oxidation contribute generally very little to HCHO column over Europe.

We added comments on this issue in section 2 (after P19702-L27) and further work on chemical mechanisms is suggested for future developments in the conclusive section.

R2.2) 19703, L10-20. This is a nice discussion of the uncertainty in the HCHO yield, but it seems that this uncertainty needs to be propagated through the error analysis of section 4.3. Section 4.3 claims to investigate uncertainties in the Eqn 2 terms, and the Jacobian $K = dy/dx$ is strongly dependent on the HCHO yield, right?

A2.2) Right, the uncertainty on HCHO yield, as discussed here, needs to be propagated into section 4.3. The uncertainty on estimation of K is already addressed in the paper, but its link with HCHO yield is indeed not emphasized in the text. We modified the related paragraph as follows. In the first version a perturbation of 10% to K is introduced to evaluate the effect of its uncertainty. This was a conservative estimate related to a <5% uncertainty due to non-linearity of chemistry. Considering that 10% is the uncertainty related to isoprene yield only, we increased the perturbation of K to 15% and repeated the sensitivity test. Text changes are found in section 4.3, after P19718-L10.

R2.3) Figure 1 seems to show much more spatial smearing for OMI versus the Chimere output. For instance, the high HCHO values north and east of Valencia. Why is this?

A2.3) This is related to a still unexplained model underestimation of HCHO column over Mediterranean sea with respect to satellite observations. A clarification of this point needs more work which is beyond the aim of the present work, which focuses on observations over land. Speculative suggestions on possible reasons of disagreement to be explored in the future are given on P19707-L3-16. See also related comment A2.5.

R2.4) 19705, L15-16, "the mean uncertainty of HCHO columns is 9.1E15". Not clear what you mean by this, please clarify. Is this the averaged SE for the domain based on aggregating the single-retrieval errors that you described earlier (which range from 50-105%)? Or..?

A2.4) Yes. We meant the average single-retrieval error over the domain. We rephrased the sentence. Moreover, the number is changed because of erroneous calculation of domain average as pointed out at the beginning of this response.

R2.5) 19707, L3, "Observed HCHO concentrations over the SE Mediterranean.." Based on Figure 1, it seems that this discrepancy applies across the entire Mediterranean, not just the SE. Also, clarify, observed by whom? Not clear if you're talking about OMI here (i.e. Fig 1) or some other observations.

A2.5) The paragraph is not properly introduced. It is devoted to a brief discussion on the model bias (which is a general feature of current CTMs) over Mediterranean sea with respect to satellite observations. The observations over SE Mediterranean refer to the MINOS campaign, as reported by Kormann et al. (2003). The paragraph is now moved to section 3.2 where we present comparison of our model with OMI and the scope of the paragraph is clarified at the beginning.

R2.6) 19708, L17-18, "The model bias over land is always within the estimated OMI uncertainty". Clarify what you mean here . . . standard error for aggregated scenes? The uncertainty on an individual scene? Also, as phrased this statement seems to call into question any analysis that follows . . .

A2.6) The sentence is now removed.

R2.7) 19709, L1-2. "The model over the Iberian Peninsula under (over) predicts. . ." This is not clear to me from Figure 1 since the difference plot for July seems to be blue or white everywhere over the Iberian Peninsula. Are you referring to a different month? In any case clarify timeframe.

A2.7) The sentence describing model bias over Iberian Peninsula was confused. Now the sentence is rephrased and clarified.

R2.8) 19709, L9-13. "production . . . largely controlled by photochemistry . . . main loss . . . reaction with OH and photolysis" A bit of an odd statement since isn't this the case for HCHO no matter where or when you look?

A2.8) Here probably the statement was confused by the specification of "... production at EMEP rural sites...". We point out here again that the HCHO budget is controlled by photochemistry and not by direct emission or deposition processes. We changed "production" with "budget" for clarity.

R2.9) 19710, L7-9, ". . . also suggested by OMI yearly cycle shown in Fig. 1". State what in Figure 1 suggests this.

A2.9) We removed the statement, because too speculative.

R2.10) 19714, L16 and L22. How valid is this assumption that the observational error covariance and the error covariance matrices are diagonal? To the extent it is not likely to be valid, need to discuss what effect this will have on your inversion results.

A2.10) Off-diagonal elements of both matrices are not known. According to Rodgers (2000), off-diagonal elements in covariance matrices may speed up convergence toward the "true state", i.e. we would obtain the same MAP solution solved with diagonal covariance matrices with less observations. However, if one introduces wrong estimates in off-diagonal elements of covariance matrices the MAP solution may be biased. This is now pointed out on P19714-L16.

R2.11) 19714, L17-22. The specification of the error covariance has a strong impact on inversion outcome. Need to elaborate on why this comparison of two biogenic emission models is a good approach for estimating it. E.g., are the Derognat and MEGAN models derived from independent (versus overlapping) sets of field observations? Why is it not a better approach to estimate this error based on a typical level of agreement between MEGAN model predictions and actual field-measured fluxes (e.g. eddy covariance)? It's not clear from the manuscript, but are you assuming the error is equal to the straight difference between the two inventories? I.e., in places where the two happen to agree (perhaps b/c they used the same field observations for that ecosystem type) then the a-priori error is very small? Please clarify this.

A2.11) The error on a-priori is estimated as the straight difference between the two inventories, and the difference is always non null (e.g. see Figure 4 for July and new Table 3 for country scale emissions). This allows a gridded estimate of the uncertainty on our current knowledge on biogenic emissions. It is the same approach followed in recent literature by e.g. Steinbrecher et al. (2009) and Poupkou et al. (2010). A quite strong sensitivity test on the robustness of this approach was given in section 4.3, where we calculated the MAP solution with error doubled and found small variations. However, we further tested the method assuming a flat uncertainty on isoprene emissions of a factor of 3, as recommended by Simpson et al. (1999). Those points are now developed in sections 4.1.1 and 4.3.

R2.12) 19716, L22. The fact that R with respect to OMI improves seems like a foregone conclusion since you have adjusted the emissions based on OMI.

A2.12) The entire part concerning results reported in Table 1 is rephrased in order to make clear that this is an expected result.

R2.13) 19717, L17-18, "since it is strictly related to intrinsic instrument characteristics." Not a true statement. The AMF depends on prior information including the HCHO profile shape and aerosol distribution, and radiative transfer through partially cloudy scenes, as you already pointed out. I question the claim that the uncertainty in estimating epsilon is negligible.

A2.13) We agree, the statement is too strong. We discussed errors related to AMF and other model input also in section 3.1, and we only meant to point out that the bulk of the observational error comes from the

fitting error. Thus the uncertainty on the knowledge of this error can be regarded as small compared to other terms in MAP solution, e.g. error on a-priori or calculation of K, as discussed soon after. The sentence is rephrased in order to smooth the statement.

R2.14) 19718, L22-23. "We conclude that our Bayesian top-down estimate . . . is robust against choice of the a-priori." I'm not sure this is a fair statement. Comparing the a-posteriori isoprene panels in Figs 5 and S2 I can see some regions that appear to change strongly with the a-priori. For instance, Greece, parts of the UK, and the northeast portion of the domain (Poland/Baltics). The domain total stays fairly similar but that doesn't mean you're not sensitive to the a-priori.

A2.14) We agree, the conclusion is not fully justifiable. We modified the final (and abstract) statements also in view of further analysis carried out in A2.11.

R2.15) 19701, L1: change to "are estimated to contribute"

A2.15) Text changed.

R2.16) 19704, L20, change to "compute normalized HCHO distributions for the AMF using the . . ."

A2.16) Text changed.

R2.17) 19706, L6 "Lagrangian model calculations . . ." this sentence doesn't really seem relevant.

A2.17) The sentence is now deleted.

R2.18) Fig 2, HCHO in ug/m3? Change to ppb.

A2.18) Units in figure are now changed to ppb.

R2.19) 19709, L19, "higher values over the sea", I think you must mean higher relative contribution over the sea?

A2.19) Right. The text is now corrected.

R2.20) 19709, L20-24, re-write this paragraph to read " Variability was found to be driven", "Isoprene oxidation was estimated to contribute", etc, etc. Need to distinguish between model findings and statements of fact.

A2.20) Right. The paragraph is rewritten according to the suggested spirit.

R2.21) 19709, L25 "Model calculations agree with description . . ." Not clear what you mean by this.

A2.21) The sentence is removed.

R2.22) 19712, L9, why two elements in the state column vector?

A2.22) The two elements of x are emission flux of isoprene and emission flux of RAVOC. Now explicitly stated in the text.

R2.23) 19712, L23, "the sensitivity of the retrieval". The term "retrieval" shouldn't be used here to avoid confusion with the actual satellite retrieval of HCHO.

A2.23) Right. "retrieval" is changed with "state vector".

R2.24) 19713, L20, "would only be not ill-conditioned" awkward, suggest rephrasing

A2.24) The sentence is now removed.

R2.25) 19718, L4, should be Figure 3.

A2.25) Text changed.