

Response to Referee 1:

We would like to thank Referee 1 for the detailed review of our manuscript and the highly valuable feedback. Changes of the manuscript are highlighted in blue and the referee's comments in black.

In “Formaldehyde and hydroperoxide distribution around the Arabian Peninsula – evaluation of EMAC model results with ship-based measurements”, the authors describe observations of these species from the AQABA campaign. They find that, for formaldehyde, agreement between observations and EMAC is acceptable, although there is a notable low bias in the model in polluted regions. Agreement for hydroperoxides is worse and is driven in large part to errors in OH and HO₂ abundance. They also compare deposition velocities determined from observations to model output, finding reasonable agreement for both HCHO and H₂O₂ over the Arabian Sea but large underestimates by the model in the Mediterranean. In general, this is a well-written, thorough paper and is suitable for publication in ACP once the following minor comments are addressed.

Line 59: Need a space between “collision” and “partner”

Now corrected in the manuscript.

Line 115: “Distinguish” is misspelled

Now corrected in the manuscript.

Line 141: The statement “Tropospheric HCHO, which is not removed heterogeneously via deposition” is incorrect and contradicted in the next paragraph where you discuss heterogenous loss of HCHO (Lines 158 – 159). Please remove or clarify your point.

Now corrected in the manuscript:

Line 145: ‘Photochemical losses of HCHO are the reaction with OH and its photolysis (R15 – R17) (e.g. Heikes et al., 2001).’

Lines 161 – 163: ‘Wet and dry deposition are major loss processes of HCHO, even though it is less soluble than H₂O₂. Highest mixing ratios of HCHO are typically found in the boundary layer and decrease with altitude in the free troposphere (Zhu et al., 2020; Anderson et al., 2017; Stickler et al., 2007).’

Line 191: Later on, you mention that you correct for line losses of HCHO. Is there any concern, given the “sticky” nature of HCHO, that you will also get HCHO sticking to your inlet tubing and then later desorbing? This could artificially bias your results high, although it would likely only matter in clean background air with low concentrations. Does the relatively large flow rate prevent this?

Theoretically, these desorbing effects could bias the results in clean conditions with rapidly changing HCHO mixing ratios, e.g. when detecting ship emission plumes during rather clean MBL conditions of the Arabian Sea. Desorption should lead to elevated / variable zero air measurements, which was not observed.

Additionally, the determination of sampling artifacts was performed based on a HCHO permeation source, with gaseous HCHO standard injections at the start of the sampling line compared to injections

right in front of the AL4021 instrument. No significant losses were detected and the HCHO mixing ratio declined fast to stable ambient air levels. We did not observe any significant desorbing effects or longer decay intervals of HCHO mixing ratios after these HCHO standard injections.

Line 206: Does that mean Sect. 2.5? Check the ACP formatting guidelines, but I think you need say Section explicitly. As it's written, it's slightly confusing.

Now corrected in the manuscript.

Section 2.6: Given the importance of the OH and HO₂ measurements to your analysis, I think more discussion about these observations is warranted, particularly as they relate to the instrumental uncertainty/accuracy. Measurement uncertainty of these species can be relatively large and could have an impact on your comparisons.

Added more information about the OH and HO₂ measurements and the data quality:

Lines 270 – 275: 'The instrument utilizes LIF of the OH radical at 308 nm, which is created by a pulsed tunable laser system (Nd:YAG) operated at a pulse frequency of 3 kHz. OH radicals are excited in a low pressure detection cell (White cell setup) with a flow rate of 10 L min⁻¹. The detection of HO₂ is achieved by chemical conversion by the addition of NO downstream of the OH detection. The resulting sum of ambient plus chemically converted OH is measured in a second detection axis (Hens et al., 2014; Marno et al., 2020). HORUS achieved an instrumental LOD of 0.03 – 0.15 pptv for OH and 0.22 – 2.01 pptv for HO₂ with a TMU of 17 % (OH) and 20 % (HO₂), respectively.'

Line 289: How does this coarse grid resolution affect your results? The land/water interface is a difficult region to capture accurately, even at high resolution. Why was this resolution chosen (I assume you were using a model run for a different purpose than just for this paper)?

The AQABA dataset is ideal with its complex photochemistry in the MBL to examine the accuracy of EMAC's complex chemistry mechanism (MOM) and the involved submodels (Pozzer et al., 2022; <https://doi.org/10.5194/gmd-15-2673-2022>). Please keep in mind that EMAC simulates the atmosphere globally, and thus its resolution is still limited. Even though I agree that a higher resolution would be a benefit for the study, e.g. if the model would be able to resolve the air pollution over the Arabian more accurately with a higher resolution. A short comparison was also done with a higher resolution based on the regional model WRF-Chem, but the increased resolution did not improve the results (e.g. over the Arabian Gulf) and was thus not included in the manuscript.

Line 293: What emissions inventory are you using? Have there been evaluations of these inventories before, particularly in the middle east?

The EMAC results were based on EDGAR (v4.3.2) see also Pozzer et al. (2022). The AQABA campaign demonstrates the first ship-based measurements over the Arabian Gulf, as far as I am aware of. To the best of my knowledge I do not know of any evaluations of emission inventories within the region.

Line 299 – 301: 'The anthropogenic emissions are based on the Emissions Database for Global Atmospheric Research (EDGARv4.3.2). Further details are presented in Pozzer et al. (2022).'

Line 314: The technique you are using to determine the deposition velocity is highly dependent on your assumption that there is very little spatial heterogeneity in the concentrations of either HCHO or H₂O₂. I think some further analysis is needed to show that this is the case, either here or in Section 3.4. Do you, for example, remove any plumes that you might encounter at nighttime? What distance did the ship cover in a night? Is it a large enough distance that differences in surface conditions (winds, waves, etc.) could affect the deposition velocity? How can you be sure that you're not sampling air that has recently been advected from a region with a different background HCHO value?

Plumes of HCHO (e.g. due to ship emissions) have been removed with the use of NO_x, CO, SO₂ and wind direction data (Lines 344 – 346 and caption of Fig. 2). No significant plumes were detected for H₂O₂. This analysis was only performed over the rather clean Arabian Sea and over the Mediterranean Sea, where we detected well aged air masses.

HYSPLIT trajectories (Section 2.8) showed that we measured air masses with a similar origin over the Arabian Sea, while winds over the Mediterranean Sea originated in Europe (Fig. S17). The *Kommandor Iona* moved with an average speed of $3.4 \pm 1.8 \text{ m s}^{-1}$, while we were sailing slowest during the high wave conditions over the Arabian Sea and the Mediterranean Sea ($\leq 10 \text{ km h}^{-1}$). We cannot exclude the possibility that high wave conditions affected our results e.g. due to high amounts of sea spray.

Figure 1: Was there any particular rationale as to how you divided the regions up, beyond names, particularly for the RN and RS? Were the chemical environments significantly different or was this just an arbitrary decision?

The regions were divided according to changes in chemical conditions and changes of the air mass origin based on HYSPLIT trajectories. We added 'according to different chemical regimes' to the caption of Fig. 2. Additionally, air mass origin plots based on HYSPLIT trajectories were added to the supplement (Fig. S17, S18), HYSPLIT is described in section 2.8.

Figure 2 caption: Should be "Contaminated HCHO data were removed"

Now corrected in the manuscript. Additionally, the readability of Fig. 2 was improved by increasing the size of the captions and its resolution.

Line 389: If MS and RS were relatively clean regions, why do you think you saw EHP there?

Thanks for this input, we agree that air masses over the MS were not clean compared to background MBL conditions. We changed the description of the chemical regime over the MS, which reflected mainly well-aged air pollution transported from Europe.

Line 407 – 409: Significantly enhanced amounts of EHP were only detected over the Arabian Gulf, although small amounts of EHP were also detected in the enriched samples of MS (Fig. S10), where we detected aged air masses originating from Europe (Fig. S17).

Line 405: How were the EMAC data adapted? Linear/bilinear interpolation? Did you interpolate in space and time? What was the time resolution of the model output?

The model results were interpolated bilinearly along the ship track (GPS data) with the S4D submodel (Jöckel et al., 2010). No interpolation is present on the time axis.

Line 294 – 300: ‘The model simulations were carried out in the T106L31 resolution, which correspond to a grid of $1.1^\circ \cdot 1.1^\circ$ (~110 km) with 31 vertical pressure layers and a time resolution of 10 minutes. The EMAC data was interpolated bi-linearly along the GPS track of the ship with the S4D submodel (Jöckel et al., 2010). The model was initialized from a previous evaluated simulation (Pozzer et al., 2022) and started on the 1st of June 2017 covering the entire campaign. The dynamics have been weakly nudged (Jeuken et al., 1996; Jöckel et al., 2006) towards the ERA-interim data (Berrisford et al., 2011) of the European Centre for Medium-Range Weather Forecasts (ECMWF) to reproduce the actual day-to-day meteorology in the troposphere.’

Line 412 (Fig. S2): It would be helpful to color code figure S2 by region to emphasize your point.

Thanks for the idea, we included updated versions of the scatter plots in the SI.

Figure 3: Could the persistent low bias in HCHO in the polluted regions from EMAC also result from the coarse model resolution?

It seems unlikely that the elevated HCHO mixing ratios over the AG and SU were only due to the coarse model resolution and the resulting continental influence in some grid cells. There is a distinct trend for higher amounts of air pollution during the first leg over the AG and SU for both the HCHO measurements and the EMAC results, while the second leg was less polluted in both datasets (Fig. S1). Even though we agree that a higher resolution of EMAC would be helpful to resolve more details.

Line 428: Should be “Even though”.

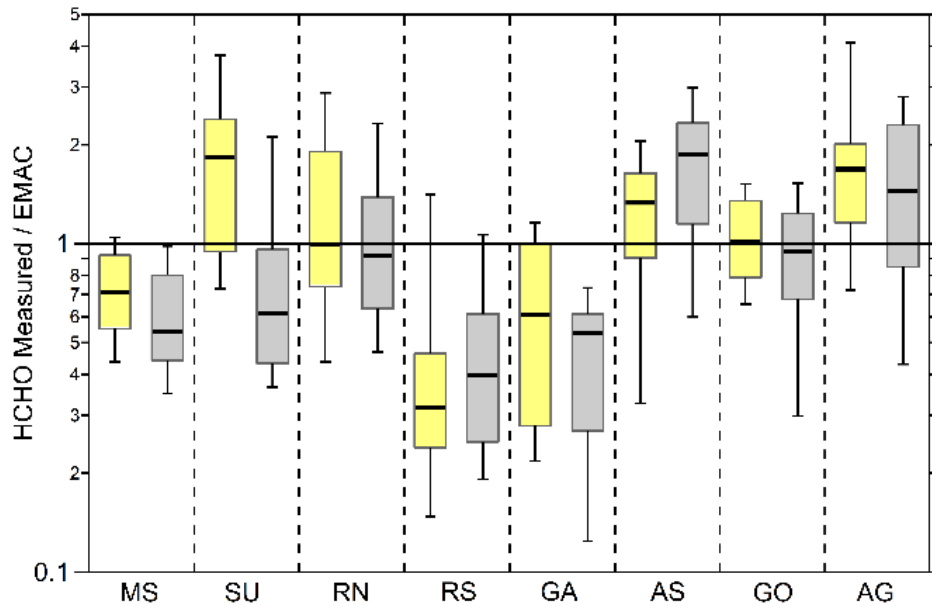
Now corrected in the manuscript.

Line 436: How well does the model capture NO_x in the more polluted regions? Could errors in NO_x abundance also affect the model HCHO accuracy there? If modeled NO is too high in polluted regions, would production from the CH₃O₂ + NO reaction also be too high, in which case you would be producing HCHO from the wrong source?

According to Tadic et al. (2020) median NO_x values over the Arabian Gulf matched within a factor of less than 2 with 1.26 ppbv for the measurements and 1.61 ppbv for the EMAC simulation. The averages show that we observed a higher variation of NO_x than the model simulated, with 3.65 ppbv measured and 1.91 ppbv simulated. EMAC was of course not able to resolve fine details, as is the case for O₃ and HCHO due to its coarse resolution, still it simulated satisfactory results for NO_x and matched the observations within a factor of 2.

Line 437: If you just look at daytime values, does the measurement/model agreement change at all, since the model does have a pronounced diurnal cycle?

In a previous version of the manuscript these plots were separated into photochemistry (yellow) and nighttime (grey). As there is no general trend that the accuracy is higher during night or day, we decided to remove these plots. The enhanced background concentrations of HCHO over the Arabian Sea were likely caused by ship emissions since we were sailing on a major shipping route (e.g. Line 378).



Line 461: I might consider breaking Section 3.2 into parts, one for HCHO and one for the other species. It's very long otherwise. Also, I feel like you contradict yourself in the sentence where you say "The model-measurement comparison for H₂O₂ is even worse...". You say in the previous sentence that HCHO "agreement is quite good." I would reword one of those sentences. (I feel like "quite good" might be an overstatement given the regressions you show in the supplement, but, given the limitations, the model is at least satisfactory).

Thanks for the idea, but we would like to keep the major structure as it simplifies comparisons between the HCHO and H₂O₂ observations within the discussion.

Line 476 – 477: 'Given the multitude of potential HCHO sources both from direct emissions and a large variety of photochemical precursors, and the limited resolution of EMAC the agreement within a factor of 2 is satisfactory.'

Line 485 – 486: 'The model-measurement comparison for H₂O₂ reveals that EMAC systematically overestimates H₂O₂ mixing ratios by up to an order of magnitude.'

Line 465: Couldn't this also be due to an underestimate in the magnitude of the sinks?

Line 488 – 489: 'This consistent overestimation by the model indicates either a significant overestimation of H₂O₂ sources or missing sinks in the model, or a combination of both.'

Figure 6: How accurate are the meteorological variables in EMAC? Do you find general agreement between observed temperature and pressure (the variables that affect reaction rate) and the model? If there are significant land/water differences, given the grid size, this could impact the accuracy of the loss rate calculations.

We agree, and additionally checked for the accuracy of water, which influences the production of H₂O₂ through recombination of HO₂. EMAC shows a slight offset of the temperature when cruising close to the coastline of Egypt (RN) and Oman (GO) probably due to the land / water difference. These only marginally affected the reaction rates, but the underestimation of H₂O led to an underestimation of

$k_{\text{HO}_2+\text{HO}_2}$. Main cause of the overestimated H_2O_2 remains the overestimated HO_2 , underestimated dry deposition during the day may be possible, but remains unclarified.

