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”

Line 115: “Distinguish” is misspelled

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.

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?
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.

Section 2.6: Given the importance of the OH and HO\textsubscript{2} 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.

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)?

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

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?

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?

Figure 2 caption: Should be “Contaminated HCHO data were removed”

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

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

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

Line 428: Should be “Even though”.

Line 436: How well does the model capture NOx in the more polluted regions? Could errors in NOx 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?

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?

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).

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

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.