I will start by apologizing for the delay in providing this review but will also note that this paper is VERY long. Much too long in my estimation. Overview papers are different than typical journal articles in that they are not expected to provide great scientific depth. Rather, they are intended to convey details that summarize a field study and can serve as a reference for the many related manuscripts, absolving them of repeating the many details summarized in the overview. As an overarching comment, I would suggest that the salient details of EMeRGe could be encapsulated in a much more condensed manuscript, but I will not attempt to specify how this might be done. Rather, I will comment on the material presented in its present form while also noting that the length of the paper will frustrate many readers and likely reduce its impact. I will also say that the length does not translate into scientific depth as myriad topics are touched upon but never satisfactorily developed to be useful. I would prefer to wait for the papers.

The most important omission in the current manuscript is the lack of any statement or information on the availability of the EMeRGe data. I apologize if I have missed it somehow, but a check of the EMeRGe website at http://www.iup.uni-bremen.de/emeerge/home/home.html is also devoid of any instructions or indication of how the data might be obtained. I would emphasize that this is a data set that is expected to have value to broader community beyond the EMeRGe team, and should be available at this point, ideally through a public portal, but minimally through a request that can be submitted via the website. With the campaign completed four years ago and manuscripts being written, the time is long past for any reasons to withhold data for quality control or proprietary purposes.

Specific Comments:

**Lines 112-133** “secondary pollutants such as ozone or secondary organic aerosols (SOA)”

It is not only happening in this paper, but sometimes it feels like the community has
forgotten that secondary aerosol production includes more than just organics. Why not say “secondary aerosol (organic and inorganic)”?

Lines 124-132 “important unresolved issues” “inaccurate modelling”, “insufficient sub-grid parameterization”, Inadequate characterization”, Inaccurate prediction”

The phrasing of these bullets emphasizes the negative, leaving the impression that we are a community without capability. I think they could be stated more positively in terms of advancing capability. In that sense, EMeRGe contributes to the long history of providing observations to challenge models and investigate processes that contribute to continuous and incremental progress in the capability to forecast and simulate atmospheric composition and chemistry. In that sense, EMeRGe is unlikely to fully “resolve” any issues but still provides a clear step forward.

Lines 139-140 “The predicted changes in these patterns indicate that future air quality in MPCs will generally be less influenced by local emission sources than by the mixing of anthropogenic and natural emissions outside the MPC (Butler et al., 2012).”

I both disagree strongly with this statement and think that it also misrepresents the referenced paper. The paper refers to MPC impacts on the global atmosphere, not local conditions. Even the abstract states that under one of the scenarios “the local influence of megacities in increased”. That said, I think the real problem is that localities around the world continue to point a finger upwind when local emissions will always dominate the local problem in an MPC. I would remove this sentence.

Lines 143-154 “Medium and long-term effects of anthropogenic emissions and their interaction with natural and biogenic emissions in the local and regional surroundings of individual MPCs are poorly understood and imprecisely quantified.”

Again, I think this language is too strong. While I agree that quantification is a problem, I do not agree that it is poorly understood. We actually know quite a bit about how this all works, but properly capturing the combined influence of emissions, chemistry, and dynamics for specific locations is still a big challenge.

Lines 149-150 “The current knowledge on all these aspects is still insufficient.”

While this statement is not necessarily incorrect, I think it would be more constructive to say that even as we make progress in understanding MPC emissions and their local and downwind impacts, the landscape of emissions and conditions in each location continues to evolve and requires ongoing attention.

Line 176 “might” should be changed to “is expected to”

Lines 286-287 and Figure 1 “The differences observed are most likely related to the special weather situation in 2017, as described in 286 Sect. 3.1.”

While I appreciate the detailed offered by the four panels, there are no striking or noteworthy differences. All the same hotspots of NO2 are there in each image. There is also no further discussion of these images in the text. Why not cut this down to a single satellite image with the hotspots for the MPC selections highlighted. After all, “Selection of MPC targets” is the purpose of this short section.

Table 1a FLEXPART is listed in the table, but never discussed. Does this reflect that it was not used as much?

Table 1b This list seems quite limited. Were products from any other satellites (e.g.,
The word “shuttle” is used here in a nonconventional manner and seems to have been coined to describe a specific pattern of stacked flight legs. Since the word is used throughout, I think it would help to formally define it. It would help to say “…incorporated vertical shuttles. Shuttles are defined here as...”

Section 3.1 This meteorological discussion is rather cursory and does little to help place later discussions in context.

Section 3.2 The discussion here is again overly simplistic. The statistics in the figure and text are rounded to one significant figure, which leads to strange numbers with all standard deviations being either 0.01 or 0.02. When I look at Figure 3, Rome has the least interannual variability of all, so I would consider it to be most representative of long-term conditions, but it is held out as the anomaly.

Lines 496-498 “Overall, 60% of the HALO measurements during EMeRGe in Europe were performed below 3000 m to probe fresh and transported outflows of selected MPCs (see Fig. 5 for the distribution of HALO flight altitudes during the EMeRGe IOP).”

It would also be interesting to know how much of the data falls into the PBL, which varies with altitude. Also, is the altitude in figure 5 a pressure altitude?

Table 3: The use of Flight Route (FR) could be confusing to the reader as it implies a specific flight path, but all of the paths are unique. You could keep “FR” but changing the term to “Flight Region” seems more appropriate.

Line 599: Further details on all the flight tracks and shuttles are given in the supplement (S9).

The tables in S9 read like shorthand for the science team and are not easily deciphered. If the authors feel it is necessary to describe every flight in detail, it should be easier to follow.

Figure 8 These composite images of CO are somewhat overwhelming as they contain details that are not specific to any particular flight day. Is this much detail necessary? How is it helping the reader?

Figures 10 and 11 In the captions for each figure, readers should be alerted that the domain of the upper right panel is different, reaching much further to the east. The upper right panel also appears to be mislabelled as 925 hPa in both figures. Isn’t it showing 500 hPa?

Line 660-661 “The chemical composition and the extent of photochemical activity of the air masses probed during the EMeRGe IOP were different for the different flight routes and tracks.”

This sentence doesn’t say anything useful. I would remove it.

Table 4 I don’t find the values in this table or the supplement to be terribly useful. I would much rather see contrasts between MPC values and differences in pollutant mixtures between MPCs and the surrounding atmosphere. If you keep the table, it is important to specify what was measured in situ versus remotely sensed. For instance, it is important to acknowledge that if NO2 and HONO were remotely sensed, those values are not entirely local to the aircraft.
**Figure 12 (and Figure 13)** As expressed in other comments, I don’t find these campaign average statistics to be terribly useful or insightful. There is absolutely nothing surprising in them. This is somewhat consistent with the very cursory discussion they receive in the text. Why is PAN only shown below 3000 m? I would be much more interested in seeing contrasts between MPC plumes against the rest of the atmosphere.

**Lines 729-731** “The HCHO mixing ratios observed in the PBL and middle troposphere during EMeRGe are somewhat lower than the North American mixing ratios (see Fig. 14). This might be related to the fact that several EMeRGe flight tracks were carried out far from emission sources over the North and the Mediterranean Seas.”

This is precisely the problem with aggregating all campaign data into a single figure without any discriminating information. It skews the ability to compare with HCHO over North America. That said, I think the North American numbers look a little on the low end. What was the source of information? The reference to Kluge et al. 2020 is a paper about Amazonia. This deserves some scrutiny.

**Lines 745-747** “A detailed analysis of the complexity of the air masses measured and the variations encountered in individual flights is beyond the scope of the present work and will be presented in dedicated publications.”

I agree with this statement, but on the opposite end of the spectrum, lumping all of the campaign data together for presentation purposes does not solve the problem.

**Figure 15** The caption states “Note that mixing ratios measured at different altitudes in the shuttle areas are not distinguishable in the figure.”

The lack of ability to distinguish altitude is only one of the problems with figures like this. These large-scale views are just confusing, and you aren’t pointing to any particularly interesting features in the text. Am I supposed to study them and see things for myself?

**Lines 784-786** “Oxygenated VOC (OVOC) result from the oxidation of VOC emissions (e.g. CH3COCH3 or HCHO) and are strong sources of HO2 and CH3O2.”

While acetone is an OVOC, it is most certainly not a particularly strong source of HO2 and CH3O2 in the lower atmosphere. HCHO is dominant and there are a slew of other reactive VOCs that are making much larger contributions than acetone, which is not worth mention in this sentence.

**Figure 17** The whole discussion of this figure is rather wandering. I am not sure what is trying to be said and why CO and CCN are being used together. Is it just that that there are north-south gradients? Specific layers are mentioned, but I don’t know what I am to take away from this figure. When doesn’t CCN ever show a strong vertical gradient?

**Lines 835-836** “Hence, C6H6 enhancements in the absence of CH3CN can be used to identify relatively “pure” anthropogenic pollution.”

When first introducing the idea of using enhancements to filter the data, it is important to be more quantitative in your language. Enhancement is too vague. What constitutes an enhanced value?

**Lines 898-901** “These maxima are not apparent in the profiles of particle larger than 0.25 μm. This is consistent with the attribution of LRT of air masses from North America, where they had contact with BB emissions. New particle formation events cannot be excluded but are considered unlikely.”
Why would aged air only show a signal in the smaller particles? There is no expectation of gravitational settling for particles in these size ranges. I do not follow the logic in these sentences.

**Lines 905-907** “As a result of the time required by the emitted precursor VOCs to be converted into secondary organic aerosol, the anthropogenic organic aerosol concentration increases above 2000 m altitude.”

The reasoning for this statement is insufficient. PBL mixing would easily disrupt such a gradient. Isn’t it also possible that there is a temperature effect on the volatility of SOA?

**Figure 20** Based on the figure and text, I frankly don’t know what it is you want me to see in this very busy data.

**Figure 21** Why is the peak in the size distribution for AP&BB greater than either of them separately?

**Line 1213** “Fig. 26” should be “Fig 33”

**Figure 37** From the figure, I have a hard time convincing myself that the various aging metrics provide any nuanced information beyond a simple distinction between background versus polluted conditions.

**Line 1400** “remanence” do you mean “remnant”?

**Lines 1434-1435** “The results obtained from EMeRGe provide new insights into the transport and transformation of pollution plumes over Europe during the IOP in July 2017.”

This paper does not cover results, nor does it provide any new insights. That will come from the other publications. I would suggest changing this sentence to say, “The ongoing analysis and publication of EMeRGe results is expected to provide new insights into the transport and transformation of pollution plumes over Europe during the IOP in July 2017.”

**Lines 1441-1442** “The selected MPCs are confirmed as pollution hot-spots by analysis using the aircraft measurements, backward and forward trajectories, dispersion models, CAMS tracer simulations and satellite observations.”

The word “confirmed” seems rather strange. Did the concept of cities as major pollution sources need confirmation? Why not say “The downwind impact of pollution from MPCs was identified and explored using…”

**Lines 1543-1544** “Prospective deployments of similar characteristics are desirable to consolidate and contextualise the EMeRGe results in Europe.”

I am going to disagree with this statement. While it is not my job as a reviewer to judge the approach taken in EMeRGe, I must confess that I am not convinced that this was the best approach to the problem. Thus, suggesting more flights of the same type is something I cannot fully endorse. Instead, I would prefer to hear something like the following. “Continued scrutiny of the EMeRGe observations and the development of lessons learned will be needed to build upon and improve airborne measurement strategies for future deployments focusing on pollution in Europe.”