Responses to Reviewer Comments on "Air Quality Predictions with an Analog Ensemble"

We thank the editor for coordinating the review of our manuscript, and the two anonymous reviewers for a thorough review of our manuscript. Please find our point by point responses to all the comments below. All of the reviewer's comments appear in regular font and our responses appear in the bold font.

Response to Reviewer #1

Comment: "The paper is yet another application of the technique already presented by the authors in several other instances, there is no new scientific value in the current manuscript, and the paper barely suites GMD. Many concepts are given for granted; several are the example of the imprecise use of language. The literature selection seems incomplete considering only weather forecast and air quality as examples, US papers mainly, and the author works. All literature on atmospheric dispersion that has preceded the air quality is neglected as if there is a specificity in the application case of techniques. The classification of multi-model ensemble is awkward (multi-physics, multi-model?) while there are classifications accepted that could be instrumental to the authors. The paper has a potential but not for ACP in my view." This was my assessment at the "short report" level, but I am afraid by point of view has not changed. The paper is poor, it is just an application of a method used for weather prediction, solar power and now air quality forecasting. The authors clearly state it stems from Delle Monache et al. (2013, DM13), improved by the method of Alesandrinni(2015) and so much so that they do not even care about explaining the methods and present the novelty that relates to this paper. Then there is a publication in AE from almost the same authors referred to as "PM2.5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model" that also refers to Delle Monache 2013 where the analog definition is also used and that already dealt with AQ. I see nothing new here compared to those publications and original that is worth publication in ACP. ACP aims at scientific novelty and originality since model developments and applications have long been confined to GMD. The editorial style is that of an internal report, with reference to other publications for the details and rushing to the results. Sure the results are good but what is the surprise here? that a method that works when applied to dense networks of data (in space and time) works also when the variable is not called temperature but solar radiation or ozone or PM2.5? This is not serious in my view and this paper does not fit ACP at all. It did not at the beginning, I pointed it out, I gave an opportunity to intervene and make clear what is the scientific relevance, nothing has changed, I am sorry but I must reject it. Accepting this publication not only would give a false sense of value to the authors, which was indeed in DM13, but it is not present here, but also would take away value from a large number of scientifically original and valuable publications that are present in the journal to date. To which very demanding reviews were presented that challenge the scientific standpoints at their very essence. This is an application of a method that has the value of yet another application. The value of demonstrating that something that was proven in the past works also for this case, as if one were to publish papers on the laws of gravity showing that they work to a couch, a lorry, a cow, a planet, an asteroid or the pen that sits in front of me on my desk. The authors are strongly encouraged to rewrite it and submit it to GMD, which is as prestigious and rigorous as ACP and fitting much better the content of this paper. Do re-write it however since the English is a bit strange at times and many concepts are rushed over like for example: The disputable idea that operational AQ forecast prevents deaths and societal costs. In my view, the planning does more than the operational forecasting in that respect. Ensembles of many different kinds are discussed generically as ensembles but never presented for what they are and their differences. Many omissions in terms of ensemble applications are present thus giving a false sense of completeness to the paper content These are minor issues compared to the lack of originality and scientific relevance but yet they will become important if the authors opt of GMD.

Reply: We strongly disagree with the reviewer's assessment that the paper has no new scientific value. The reviewer has raised five concerns, i.e., a simple replication of the technique, novelty of the work, citing relevant literature, paper writing, and relevance to ACP. Here, we respond to these concerns one by one.

<u>Simple replication of the analog ensemble technique</u>: Yes, we have employed the analog ensemble technique in a variety of applications but, contrary to the reviewer's belief, the previous applications do not guarantee that the technique will work for air quality application – a statement that is true for any method, which is the reason why in science the generality of a new algorithm is never guaranteed and needs to be tested. The predictive skill of numerical models for weather parameters (e.g., wind and temperature) may vary significantly when compared to the predictive skill of models for quality variables (e.g., ozone and $PM_{2.5}$), and that can significantly affect the performance of any given postprocessing method. For example, the extension of analog ensemble application from one area to another (e.g., weather to air quality) requires careful selection of the predictors that best identify the similar (analogous) atmospheric conditions in the past.

For air quality, the predictors have to be selected in such a way that they are able to (1) identify air pollution episodes of similar magnitude in the past, and (2) identify the meteorological and chemical conditions leading to similar past air pollution episodes. Following these two criteria, we selected O_3, $PM_{2.5}$, 10-m wind speed and direction, 2-m air temperature, 2-m specific humidity, and cloud cover as the predictor variables in our implementation of analog ensemble for air quality. The rationale for selecting these variables as predictors is now described in the revised manuscript and reproduced here for reference:

"The rational for selecting the aforementioned air quality and meteorological variables as predictor variables is as follows. O_3 and $PM_{2.5}$ allow us to identify pollution episodes of similar magnitude in the past. Temperature plays a vital role in several processes relevant to air quality including atmospheric chemical kinetics, biogenic emissions, and mixing. The wind speed and wind direction allow us to ensure that similar transport pathways contributed to the analogous air pollution episodes in the past. Humidity is selected for its key role in the formation and destruction of both O_3 and $PM_{2.5}$. Water vapor (H_2O) in conjunction with O_3 photolysis is the main source of hydroxyl (OH) radical, which in turn initiates photochemical production of O_3 through oxidation of different volatile organic compounds (VOCS). In the case of $PM_{2.5}$, humidity determines the aerosol water content, which is important for secondary aerosol formation. Cloud cover determines the amount of solar radiation available for atmospheric photochemical reactions that produce both O_3 and $PM_{2.5}$. In summary, the predictors are strategically selected in such a way that they are not only able to identify the pollution episodes of similar magnitude in the past but also identify the meteorological and chemical conditions leading to similar air pollution episodes."

Because of the research required to identify the suitability of a method for a given application, we believe it is naive to compare different applications of analog ensemble with

the application of universal law of gravity that simply requires replacement of masses of the bodies involved. Furthermore, research also allows us to challenge the universally accepted laws. For example, even following the line of thought of the reviewer, the Newton's law of gravity and motion were found to be not accurate enough to deal with very strong gravitational fields or to describe with extreme precision the orbit of Mercury around the Sun. Scientists discovered that when trying to apply the Newton's laws to objects other than a couch or a cow.

<u>Novelty of the work</u>: To the best of our knowledge, we are proposing for the first time a novel approach to generate probabilistic predictions for air quality, which is based on a significant shift in paradigm with respect to traditional ensemble methods: i.e., rather than running a numerical model with several different configurations to create the ensemble members, we run the air quality model in real time only once, and then generate the necessary uncertainty quantification by inference from the training data set. Additionally, a strategic selection of the predictors is required for using the analog ensemble method in air quality applications (see the response above for details). The selection of these predictors with their corresponding weights (as explained in Section 2.3 and 3.2) contributes further to the novelty of this work. It is worth noticing that Djalalova et al. (2015) did not touched the subject of probabilistic predictions as in this proposed work, it was focused only on $PM_{2.5}$ (here we analyze the performance of the proposed method on both ozone and $PM_{2.5}$), and it involved the combination of analog-based deterministic methods with the Kalman filter (the latter is not part of the current manuscript).

<u>Citing relevant literature</u>: We did cite several papers on transport and dispersion modeling as requested by this Reviewer in his preliminary comment (i.e., we cite among others Galmarini et al., 2001; Galmarini et al. 2004; Kioutsioukis and Galmarini et al. 2014; Potempski et al. 2008; Potempski and Galmarini, 2009; Solazzo et al. 2012). Here, we focus on reducing biases and quantifying uncertainty in the air quality forecasts produced by the three-dimensional Eulerian chemistry transport model (CTM). Therefore, we discuss potential sources of uncertainties in CMAQ and the relevant literature attempting to reduce biases and errors in CTM forecasts. We believe that citing our previous work is quite relevant and fully consistent with ACP guidelines, to provide a reader with the background information on where and how the analog ensemble has been used previously.

<u>Paper Writing</u>: We apologize for the imprecise use of language. The manuscript has been revised throughout to improve clarity and proof-read by a native English speaker. We agree with the reviewer that planning is important to mitigate health impacts of air pollution but operational air quality forecasts are as important because air quality managers cannot plan anything until they know about the forthcoming air pollution episodes. Several sections of the manuscript including Introduction, Prediction System and Data, and results have been revised to reflect this sentiment.

<u>Relevance to ACP</u>: Here, we have neither developed nor described a numerical model and/or model component, which is a key requirement for the GMD. Rather, we provide robust evidence that the analog ensemble technique is capable of reducing errors and biases in air quality predictions, and to generate reliable and calibrated probabilistic predictions. ACP

has previously published articles that focused on reducing errors and biases in air quality predictions either using chemical data assimilation over both the US and Europe, or a dynamical ensemble in China (e.g., Saide et al., 2013; Flemming et al., 2017; Kioutsioukis et al., 2016; Potempski and Galmarini, 2009; Hu et al., 2017). Therefore, we strongly believe that our paper is suitable for publication in ACP rather than GMD.

References

Flemming, J., Benedetti, A., Inness, A., Engelen, R. J., Jones, L., Huijnen, V., Remy, S., Parrington, M., Suttie, M., Bozzo, A., Peuch, V.-H., Akritidis, D., and Katragkou, E.: The CAMS interim Reanalysis of Carbon Monoxide, Ozone and Aerosol for 2003–2015, Atmos. Chem. Phys., 17, 1945-1983, https://doi.org/10.5194/acp-17-1945-2017, 2017.

Saide, P. E., Carmichael, G. R., Liu, Z., Schwartz, C. S., Lin, H. C., da Silva, A. M., and Hyer, E.: Aerosol optical depth assimilation for a size-resolved sectional model: impacts of observationally constrained, multi-wavelength and fine mode retrievals on regional scale analyses and forecasts, Atmos. Chem. Phys., 13, 10425-10444, https://doi.org/10.5194/acp-13-10425-2013, 2013.

Hu, J., Li, X., Huang, L., Ying, Q., Zhang, Q., Zhao, B., Wang, S., and Zhang, H.: Ensemble prediction of air quality using the WRF/CMAQ model system for health effect studies in China, Atmos. Chem. Phys., 17, 13103-13118, https://doi.org/10.5194/acp-17-13103-2017, 2017.