

General comments

This paper presents an operational multi-model forecasting system for air quality which has been developed to provide air quality services for urban areas of China (MarcoPolo-Panda). A companion paper describes the models involved in the forecasting platform. The prediction system provides a 72 hours forecast with a zoom over the largest Chinese cities. In the present paper the performances of each member and the ensemble are evaluated with available data for the main criteria pollutants.

This platform will be useful to develop services and improved in a near future by adding new members of the ensemble, use of satellite data for data assimilation and other MOS technique to improve the forecasting system. I am favorable to the publication of this work.

The author should add the sulfur dioxide in the analysis since this pollutant is in the target of the Chinese authorities in the new 3 years plan adopted to reduce air pollution in China.

Even if the model description is provided in the companion paper, a short description of models with the main features like resolutions, meteorological and emission data must be provided. As it is an analysis based on the first 24hours forecast, the ability of models to predict air pollutant concentrations could be due to the ability of the met driver to forecast the meteorological conditions or to the chemistry transport model itself.

Specific comments

As it is written and specified in the paper the main weakness of the exercise is to run models with outdated emissions. This must be improved in the next versions of the platform with certainly a yearly update.

To better understand the behavior on PM₂₅ concentrations a short analysis on sulfate, ammonium and nitrate species could be added to highlight the main pattern of these species.

The authors must justify why they did not include peri urban and rural stations in the vicinity of cities to better match with the resolution of models.

I suggest the authors to remind how POD, FAR and AQI indexes are calculated or they must add a reference.

L. 283: These low correlations for PM could be due to dust storms occurring during this period, Usually, for pollutants the occurrence of pollution events increase the time correlation that's why ozone time correlations are better in summer and PM correlation better in winter.

To understand the negative bias on PM₁₀, the authors should remind how dust are taken into account by models (boundary conditions, local emission parametrizations).

The author should comment the flat diurnal profile of PM_{2.5} in the observations while the model are very sensitive to the increase of the PBL during the afternoon, perhaps the effect of the secondary production of aerosols that is not well predicted by the models.

Boundary conditions are important drivers for several pollutant concentrations such as Ozone, PM10, PM2.5, dust, sulfates. This should be highlighted in this study. In this paper or in the companion paper an analysis of the behavior of the model used for the boundary conditions would be helpful to understand the performances of the models.

I suggest to cite this paper : *Bessagnet, B., Pirovano, G., Mircea, M., Cuvelier, C., Aulinger, A., Calori, G., Ciarelli, G., Manders, A., Stern, R., Tsyro, S., García Vivanco, M., Thunis, P., Pay, M.-T., Colette, A., Couvidat, F., Meleux, F., Rouil, L., Ung, A., Aksoyoglu, S., Baldasano, J. M., Bieser, J., Briganti, G., Cappelletti, A., D'Isidoro, M., Finardi, S., Kranenburg, R., Silibello, C., Carnevale, C., Aas, W., Dupont, J.-C., Fagerli, H., Gonzalez, L., Menut, L., Prévôt, A. S. H., Roberts, P., and White, L.: Presentation of the EURODELTA III intercomparison exercise – evaluation of the chemistry transport models' performance on criteria pollutants and joint analysis with meteorology, Atmos. Chem. Phys., 16, 12667-12701, <https://doi.org/10.5194/acp-16-12667-2016>, 2016. This paper provides an intercomparison of some of the models used in the MarcoPolo-Panda project.*

L. 106 : “We show that the application of bias correction to the models improves the forecasting skills of binary ozone predictions”. This sentence cannot be written in the introduction as it is a concluding remark.

L. 306 : For the overestimation of NO2 over cities where emissions are better documented, the missing urban parameterization could be one of the reason due to less vertical mixing in the model.

L. 647 Is it a correction based on analysis or forecast? The method is not well described.

L.756 “...predicts the occurrence of pollution events a few days before they occur.” Difficult to write this as the author only focus on the first 24 hours forecast.

L.780 “Furthermore, data assimilation of satellite and in situ observations should significantly improve the performance of the forecasting system.” This is very challenging but promising, the authors could add some references to support this initiative in terms of added value for such a system.

L782 What the authors mean by “a more advanced approach”?

I think that MOS (Model Output Statistics) techniques applied to the ENSEMBLE could be more useful to improve such a forecast system.

Technical comments

All along the paper put ensemble in capital letter ENSEMBLE

Table 3: Replace NaN by NA (not applicable) since NaN means Not a number

Figures 5,6,7 : Units?

Certainly the quality of figure can be improved and the figures harmonized.

L. 166 by relatively coarse resolution models

L.225 Should be ...“nitrogen” emissions

L.279 ...”exhibit small correlation”..., I would say low correlations

L313-314 “These cities also show an overestimation of NO₂ concentrations” would be better

L.612 The predictions of PM_{2.5} **concentrations**