

General comments:

The authors are presenting a paper that assesses the skills of an air quality model system forecasting air pollutants concentrations over Europe and Italy, as a nested domain, in the specific year of 2017.

Together with emissions, the weather conditions play also an important role in any air quality model system. Its analysis together with the model skills may help to prioritise the developments needed in the air quality systems and put in perspective the obtained results for this particular year. As example the Copernicus service has two small reports that also show the anomalies of the gridded observations of temperature and precipitation (E-OBS) over Europe, which can help to frame the year 2017 regarding the climatological normal of 1981-2010 (<https://climate.copernicus.eu/climate-2017-european-wet-and-dry-indicators>, <https://climate.copernicus.eu/climate-2017-focus-region-southwest-europe>).

As it was noticed during the specific comments, the CAMS reanalysis product can be used to better understand what air quality model system can do better for a similar domain. In principal yes, since it gathers the best knowledge of the atmospheric composition in those instants. However, it is worth to remember that this assessment procedure implies that the emission inventories have similar methodologies, the meteorological forcing come from the same model, and the observations used in validation are sampling the covered areas in the same way (i.e, in number of stations and radius of coverage). The authors give conclude that they have learned from the CAMS reanalysis related with the FORAIR_IT ozone skills. However, more lessons can be taken from this set of data both at the European and Italian level, namely on season surface patterns differences.

When discussing the skills of both systems this is often done through the confrontation between the FORAIR_IT realization and the CAMS_50 realization. However, the CAMS_50 realization is already a median of realizations. When reading the text one is induced on a single realization and starts to forget that fact, especially for those that are not too familiar with the products. Regarding the added value of the model forecast, it would be preferable to compare the same product, the CAMS forecast.

If the authors can afford that, it would be a valuable help if the text can be revised by someone that relates also the scientific topic.

As supplemental material is it possible to provide information about the mean concentration values related to every graph or table presenting the model skills?

Specific comments:

Figure 1: Colours green and blue are very difficult to distinguish from each other. It would help if a legend of the colours is added to the figure, increase the size of the figures text font, and also increase the size of the figure.

Figures 2-> 5, mention in the figures legend that the y-axis does not start at zero, in order to have a better description of the hourly fluctuation of the skills. To better see the curves when analysing the

different pollutants if the colour of the lines in panels a/b2, a/b3 and a/b4 can be defined in a way they contrast (especially when analysing the graphs in Figure 4 and 5)

Introduction

Pg1

L24: "...improve air quality by reducing anthropogenic emissions, through the improvement of industry, energy production, residential...". Suggestion to clarify this phrase: "...improve air quality by reducing anthropogenic emissions through the introduction of measures according to the activity sector as industry, ..."

Pg2

L4: "To tackle this issue...". Suggestion to clarify this phrase: "" To better manage and control air pollution, ..."

L3 and L5: Please explicitly designate the names of the directives, keeping the references.

L6: Making a paragraph after "In the last decades..." will help the reader.

L7: suggestion to clarify the text "...,, in the days (1 to 5) after the prediction, ...", try to change to something like "with valid forecasts varying between 1 and 5 days,"

L8: Can the authors be more specific with the term "diffusion" in this context?

L9:10: the text can be clarified specially after "...episodes allows to take actions..." . As example, "...take actions at different levels (starting at the policy-administration level, down to the private sector to the citizen level), reducing the risks for human health by limiting the population exposure to these episodes."

L11: "...to forecast the chemical..."

L12: instead "...physical and chemical dynamics..." I would suggest "physical and chemical processes..."

L16: Suggest starting a new paragraph with "Since 2009"

L23: Since DEHM and GEM-AQ are now operational the suggestion is to remove the text in parenthesis.

L24: " is providing an ensemble of air quality forecasts" instead of "e is providing air quality forecast ensembles".

L25: Suggestion is to change to a more general statement the text "...the national air quality forecasting systems are maintained or developed in order to..." to something similar to "...each member state is responsible to develop and maintain an air quality system in order..."

Pg 3

L1: "...and original it started..." -> "... original started..."

L2,3 and 4: "computing infrastructure of ENEA CRESCO. The new forecasting system, whose upgrades are now funded by the Italian Ministry of Environment, is maintained and operated by ENEA (Italian National Agency for New Technologies, Energy and Sustainable Economic Development)".

The suggestion is to change the order to:

...computing infrastructure of ENEA CRESCO (Italian National Agency for New Technologies, Energy and Sustainable Economic Development). The new forecasting system, whose upgrades are now funded by the Italian Ministry of Environment, is maintained and operated by ENEA.

L20: "...FORAIR_IT forecasts on daily basis the concentrations in the three following days." ,
Suggestion is to include the frequency of the output results:

"...FORAIR_IT forecasts on daily basis the hourly concentrations in the three following days."

L22:23 Suggestion is to remove the parenthesis enclosing the pollutants that are delivered in the forecast.

Pg5

L8: Have the authors any reference that can show the advantage of explicitly solving the convection in RAMS at 4 km grid resolution over Italy?

L36: There is an update on the Emission inventory over Europe. Is it possible to give an uncertainty of the emission considered over the European domain within the FORAIR_IT?

In the text, there is no mention to the emission inventory underlying the European interim reanalysis.

Pg6:

L16: Suggestion: "Over the Italian territory, only daily mean PM10 and PM2.5 datasets were available. As adequately representative of the FORAIR_IT spatial resolution the measurements at background stations were used in the model system evaluation."

Pg 7:

L11: The authors can remove "Anyway" from the text.

L12:14: This is one of the most sensitive parts of the work. The results of the model system FORAIR_IT over the European domain is compared with the Copernicus service regional reanalysis ensemble. As the authors pointed out, the error will be lower in the reanalysis product since it try to

represent the best knowledge of the pollutant surface fields, combining model results and observations. It would be fair to compare the FORAIR_IT model system also with the Copernicus the ensemble forecast product. Like this it would be possible to know how different are the skills of both forecasts, between them and the ensemble reanalysis.

L19:20: Suggestion " ...using the first day forecast of the previous forecasting cycle as the initial condition."

L21:22: Suggestion "The procedure is lopped for the 3 forecasting days".

L24: "The whole operational chain is automatically launched via **the** software utility cron."

L25: Suggestion "of them on" -> "of them at"

Section 3:

L2: Suggestion: "Main statistical..." -> "The main statistical..."

L5: Suggestion is to write the full name of the scores used and their acronyms.

Pg 8:

L24: Please remove the parenthesis enclosing the pollutants.

L25: Several questions arise in the way the results are presented in Tables 2 to 5, and Annex A2.

It seems not to be a fair to make a mean (my supposition) of the skill scores obtained in the European domain and the domain over Italy. The possible gain of having a high resolution forecast over Italy would be seen if the skills of both runs are compared over the same area using the same stations. Since this is not explored, the reader is induced to think on a general skill for both domains.

Pg 10:

L12: Suggestion "skills in the last three years,2015-2018 (Collin et al., 2018) This may be explained" -> "skills between 2015 and 2018 (Collin et al., 2018). This may be explained".

L14: Is it only one last long event occurred during January and February?

L21: The way it is done, the statement saying ". Overall RMSE CAMS_50 results are better than FORAIR_IT ones;" is an unfair comparison between the CAMS and FORAIR_IT, as commented **previously**.

L28:29: Suggestion, change "a 12-hour less accentuated cycle characterizes CAMS_50 results", to something as: "a 12-hour cycle with smaller amplitudes characterizes CAMS_50 results".

L36: Please, see comment L21.

L36-37: commenting the statement “: FORAIR_IT has a well-defined daily cycle for CORR but not for RMSE”. Looking to the graphs a2) and a49, a daily regular pattern can be seen for both.

Pg11:

L14:” lower are the values of the RMSED better is the performance of one realization respect to another”, is maybe better like: ”the lower the value of RMSED the better is the performance of one realization respect to another”.

L15:19: In spite of the fact that Italy does not have hourly PM10 data, the products that are being compared are not the same. PM10 is a pollutant that may be advected over long distances and the assimilation procedure will constrain the surface concentrations. That impact of this effect could be interesting to assess if the cams ensemble forecast and reanalysis were compared with the FORAIR_IT results over Italy (notably over its Northern Part).

L20: please see comment on table Table 2, pg 8, L25.

PM25: Higher spread of MME over Italy.

Pg 13: Figure 3, please see comments on Figure 2.

Pg14:

L11: days -> day

L11: Since only 2017 is being evaluated, and to not get it confusion, please rephrase “with higher values from November to March”. These are two periods of the same year that are apart.

Pg15

L15:18: “The better skills obtained over the Italian domain with respect to European one are probably due to the increase of spatial resolution. In addition, the Italian measurements, differently from the European ones, were not used for assimilation within CAMS_50 simulation; indeed on Italian domain, validation and assimilation dataset are independent for both FORAIR_IT and CAMS_50 realizations.” Probably this statement is more correct when analysing PM2.5. However, account the differences only to the increased resolution is possibly minimizing the role of the emission inventory used in the higher resolved domain.

Pg17

Figure 4. The figures in the panels are very small. Please consider also the general comment on Figures 2 -> 5.

L20:21 “at hours 5 and 18, MMB values are close to 0, whereas the highest underestimation (-0.6) is reported around midday”, It is never stated that the CAMS_50 ensemble results MMB values are most of the time underestimating the observations.

L22:23: “Again the reason of the different daily cycles of the two models performances may be related to the different temporal profiles used for emission disaggregation.” When comparing the FORAIR runs over the European domain with the CAMS ensemble, it is important to have in mind that the ensemble may reflect more than one temporal profile used for emission disaggregation (please take a look into the CAMS_50 URL: <https://www.regional.atmosphere.copernicus.eu/>, under CAMS Verification mean scores, especially for NO2 and PM2.5).

Pg19

L21:22: “may be due to either meteorological conditions, being more dispersive along with the forecast time, or to global boundary conditions”. The main reason is probably related to the degradation of both meteorological and chemical boundary conditions forecasts.

Pg21

Figure 5, please see general comments on figures 2->5

Pg22

L14: “Conversely the winter one is quite unusual; anyway it is present...” , please consider changing to “Conversely the winter one is quite unusual; This peak is present...”

L14:16: Can the authors add a reference supporting this sentence?

L27:28: Suggestion “...simultaneous bad RMSE and good MMB skills mean high CRMSE values...” change to ““ ...simultaneous high RMSE and near zero MMB skills mean high CRMSE values...”

L34: Can the authors, please, clarify the meaning of “large observed daily excursion since the 5% of the stations has skill variance less than 0.7.”?

Pg23

L7:8: “This outcome, meaning that O3 forecasted concentration values tend to increase passing from D0 to D2, is expected because, being O3 a secondary pollutant, its precursors decrease.” The way it is written, the phrase is difficult to read.

Pg 24

L11:12: Suggestion: Due to the causal relationship between high levels of PM and human health, it is therefore pivotal to have the possibility to predict with accuracy in space and time the expected PM concentration...

L16: "hearth" -> "heart" ? (hearth, the Cambridge dictionary defines it as: "the area around a fireplace or the area of floor in front of it").

L20: Figure 6, upper panel must be increased in size. As it is, there is a high risk of losing track of the different marks, dot and square black and no fill dot (white?). White square markers are not visible.

Pg25

Figure 6 caption will be clearer stating that the average PM10 concentrations are calculated with the model results.

GFAS or other biomass burning database, may help to justify the lack of model skill in modelling the nitrogen ions.