The paper presents a promising strategy for air quality forecast, which they have applied to large regions of three east-African countries including their capitals Kampala, Addis Ababa and Nairobi (with a special focus on Nairobi, capital of Kenya). Studies on air quality modelling are extremely scarce, so this work is, by itself novel, particularly in the fact that the authors present results for three large cities along with the corresponding data. The numerical methods used by the authors and their tools are exposed clearly and their results in terms of comparing model to observations are good (and sometimes extremely good).

However, I feel like there is more potential with this study if the authors would spend less time and space in presenting too many statistics (sometimes not relevant as I will discuss below for specific points). Statistical discussion should be refocused in model average, observation average, MFE and MFB which as the authors themselves recognize are better suited to their study than the metrics they use first. It is interesting that the model performs correctly relative to observations, but once this is done the reader feels he has the right not to see more statistics but to get insight on the physics and chemistry that are at play : if the model provides a relatively correct assessment of the PM time series, then it would be very interesting to know what the model tells us on the composition of PM (and therefore, possible, on its sources). Is it made of mineral dust, primary anthropogenic contaminants, SOA? If the model is correct on specific station, then we would like to see a map of contamination that it produces for the entire simulation domain (that of Fig. 7 for example): is Nairobi the maximum for the entire domain, are there other hot spots? How id pollution channelled – or not – between the mountain slopes? In its current shape, this article sometimes looks like a technical report on the feasibility of a particular forecast system for specific regions, which is not really what is expected from a research article. I think that with the additions above, this article could give many more indications on the specificities on Particulate matter composition in this region, and yield more interesting questions for future research. I feel this article will deserve publication because they obtain a great performance in reproducing pollution in areas where this has rarely been attempted; Once major changes are brought (making the statistical discussion more straightforward and give more scientific material from the model outputs), I feel that this may become a breakthrough article for air quality modelling in Africa.

A > The authors thank the reviewer for their interest in the manuscript and for the detailed review they made of its content and form. What the authors have tried to do in this work is to assess models for the simulations of meteorology and atmospheric chemistry over three urban domains within East Africa, for the first time at high resolution and using only available data from open-source provider and limited amount of field measurements. Hence, this paper wants to open the way for further and future scientific works focused on refining different aspect of the model configuration adopted here and to improve in this way the predictions made by the models. From this point of view the first novelty of this work is represented by the detailed analysis of the model performances simply because it is the first time that WRF and CHIMERE have been used to simulate PM$_{2.5}$ in this area of the world and at this resolution.

Forthcoming works will use this same configuration and will explore additional aspects of the atmospheric composition of East African conurbation focusing on the primary and secondary composition, transport effects and also fraction of elemental and black carbon in the PM. This type of analysis though requires an additional preparation and/or level of detail of the input data used for the simulations that are not in the aim of the present work. The presented work provides the necessarily starting point for these further studies by our and other groups. The authors thank the reviewer to suggest the possibility to investigate the formation of secondary aerosols, it is a hot topic in the analysis of the PM$_{2.5}$ levels within urban conurbations. The information we have at the moment for

the creation of primary emissions of PM$_{2.5}$ are though limited to the lumped species and few additional components of it (Organic and Elemental carbons) that would require many assumptions in the reliability of the secondary components modelled by CHIMERE. What we have done instead according to the reviewer suggestion is to substitute the original Figure 7 with a new Figure (9) providing the average concentrations of PM$_{2.5}$ in the whole domain KEN2K and giving the position of additional hotspots of concentrations outside Nairobi urban area.

Finally, our future efforts are oriented in refining the configuration used for the models and also increase the level of detail of the input used for the simulations. This means of course to obtain more information about local sources of anthropogenic emissions but also quantify with higher detail the biogenic emissions that have surely an impact on the levels of air pollutants in urban and rural locations.

MAJOR COMMENTS

A. Statistics on wind

l. 362: it is unclear what is meant by wind direction » and its unit. The vocabulary used i not appropriate for wind direction (« higher » wind direction has no meaning in my opinion). Mean Normalized bias has no clear meaning either in this sense (if as I think wind direction is in degree). Authors should explain how they build their indicators for wind. For calculating the RMSE and MNB, how do they account for the difference between a wind oriented at 1° and one at 359°? they are close but the difference between them is large. In general, MNB and RMSE are not adapted to deal with angles. Even the average does not make sense (average between 1° and 359° is 180° which does not make any sense…). I suggest that the authors remove the statistics they have done for wind direction (and possibly replace them by a more meaningful way to do these statistics, e.g., average and standard deviation of the geometric angle between observed and modelled wind speed). An alternative is to rely mostly on Fig. 5.

A > The wind direction shown in the manuscript is defined in degrees and the text has been modified according to a more appropriate description of the variation of this variable. The statistical parameters of MNB and RMSE are calculated as shown in the following equations:

$$MNB = \frac{\sum_{i=1}^{n}(M_i - O_i)}{\sum_{i=1}^{n}(O_i)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(M_i - O_i)^2}{n}}$$

We agree with the reviewer that this type of metric can be suitable for linear variables such as temperature and relative humidity, but they cannot be appropriate for the analysis of circular variables such as wind directions. The statistical evaluation of WRF has been therefore re-written using mean fractional bias and errors (MFB and MFE) for the weather variables and also using the calculation of the index of agreement (IOA):

$$IOA = 1 - \left[\frac{\sum_{i=1}^{n}(O-M)^2}{\sum_{i=1}^{n}(|M-\bar{O}|+|O-\bar{O}|)^2}\right]$$

The new statistics has been added to table 3 and a new discussion of these parameters for the 4 variables has been added in the manuscript.

B. Error on wind speed?

Wind speed seems largely and critically false for the Kenya domain (Table 3). Could the authors double-check their numbers?

A > The statistics of Wind Speed for the domain KEN2K have been modified due to the presence in the observations of data from a particular station that after further analysis, was found to be suspect. In absence of precise information on the possible cause for this (the mean wind speed in that particular station was 45 m s^{-1} with only few data available during the month) we have excluded that station from the statistical evaluation and performed the calculation again. The new statistics have been expressed in terms of MFB, MFE, R and IOA.

C. Use the same metrics throughout the paper

l. 528: here the authors argue (insightfully in my opinion) that MFB and MFE have less problems than MNB and NRMSE which they use above. Why not use MFB and MFE throughout the paper? MFB and MFE could be calculated in time-average foe all stations in the first place, given in table 4, and Table 4 could be used to discuss the results relative to the Boylan and Russel criteria. This would be less confusing than the current presentation and would avoid needing Fig. 6 which is an unusual figure style and, lacking the time dimension, does not bring much more understanding to the reader.

A > Thanks for the valuable comment on this aspect. The reviewer is right is highlighting the possibility to use the same metrics of MFB and MFE both for the evaluation of WRF and for the evaluation of CHIMERE. These two metrics are definitely more appropriate than the classic use of MNB and RMSE that can be indicative for the linear variables but misleading for circular variables like wind direction.

According to the comments of the reviewer great part of the statistical analysis WRF has been re-written using as new metrics MFB, MFE and also index of agreement calculated as follows:

$$\text{IOA} = 1 - \left[\frac{\sum_{i=1}^{n}(O-M)^2}{\sum_{i=1}^{n}(|M-\bar{O}|+|O-\bar{O}|)^2} \right]$$

For what concern Figure 6 and the representation of CHIMERE performance in terms of MFB and MFE we think that that type of figure showing the capability of the model to reproduce PM$_{2.5}$ concentrations both in low and in high peak of concentrations represent an important information at this stage of the numerical modelling of air quality in east Africa. One of the elements of uncertainties in this work is represented by the use of low spatial resolution anthropogenic emissions to simulate urban and rural concentrations compared with observation points. The reliability of the model needed to be evaluated in the representation of both the general hourly pattern of concentrations but also in the discrimination of period of low and high concentrations with the available data for simulations. In the vision of using WRF-CHIMERE as tool to assist policy making these are level of reliability needed and important to address.

D. What do the authors mean by « coupled models »?

Usually, coupled modelling means that the chemistry-transport model is able to give feedback to the meteorological model (both models running at the same time, similar to a general circulation

model with atmosphere + ocean). This is not the case here, so the expression « coupled models » is confusing and should be removed from the text.

A > The reviewer is right is pointing out the imprecise definition of coupled model. The authors reviewed this term along the manuscript and modify it accordingly. The WRF and CHIMERE models have been used independently one from the other. Output from the simulations of WRF have been used off-line as weather parameters for the CHIMERE simulations.

l. 75: lon-term → long-term

A > this typo (like all the other found in the manuscript) has been corrected.

l. 148: is it really two-way nesting (do the small domains retroact on the large one?)? Otherwise, this is one-way nesting.

A > Yes, it is. The two-way nesting is a proper option activable on in the WRF model that let the parent and nest domain communicate the variable values and give modification retroactively from the nested to the parent domain.

l. 343: why such a massive underestimation for temperature in Kenya? Such a difference would strongly affect photochemistry isn't it?

A > The value of 4.1 °C is a big difference in temperature but calculated on the average of all stations for Kenya. One station in particular, Narok shows a huge difference between model and observations (10.9 °C) and this big value influence the average bias on all the stations. The bias in temperature for the station of Nairobi is 1.3°C smaller. This difference between the value of the single station of Narok and its influence on the average of the stations has been added in the explanation of the statistics (Lines **391-400**).

l. 392 MNB is not really significant here, it would depend on if temperature is expressed in K or °C. In any case, 0.1 MNB is not small at all (relative to an average value of 20° this is a 2° bias which is not small). I see something strange in the numbers presented in Table 3. There is a negative bias of 4.1° relative to an observed mean of 23.2° so with the typical definition of the mean bias ((-) /) I would expect a MNB ~0.18 while the authors claim the MNB is 0.1 here. This looks underestimated. I am aware of the difference between NMB and MNB and I don't have the data to recalculate the MNB here (see the definitions in e.g., https://www.bnl.gov/envsci/schwartz/pres/metrics.pdf). Could the authors explain what exact definition they have retained for the mean normalized bias and how they deal with missing data? This all the more surprising as for UGA2K and ETH2K I find exactly the same number as in Table 3 by calculating -) / . This is also an argument to just drop MNB which behaves in a confusing way, as suggested in my major comment C.

A > The values of 4.1°C is the MNB calculated on the average of all the 5 stations for Kenya where the station of Narok in particular has a bias of 10.9 °C between model and observations. The absolute bias calculated for the station of Nairobi as previously explained is 1.3°C and it is related to the MNB calculated in the same station that is -0.06. In the same way the other number 0.1 is the MNB calculated for the station of Addis Ababa. The paragraph has been modified with and this aspect clarified.

The values of MNB and RMSE calculated for the statistics take in account the absence of the observation points in the calculation. The model-observation values are taken in account in calculation only when the observation is present for that particular hour.

l. 453: « negligible » is not the correct appreciation for biases up to 4°C in temperature. Table 4 : The figures don't always make sense. For Addis Ababa, there is a MNB 0.1 for daily data but 0.88 for hourly data, this is much likely affected by data points with an extremely small denominator driving the entire average. If the model yields 2.0 and 2.0, and if the observations yield 0.1 and 3.9 then the MNB would be 0.5 * (1.9 / 0.1 – 1.9 / 3.9) ~10 which hardly makes any sense. Again, see my major comment C.

A > The paragraphs of the WRF validation have been modified considering the new statistical parameters inserted in the analysis. Albeit there is a bias in the averaged station statistics to take in account the individual stations, the closest stations to the urban areas of interest shows acceptable biases and levels of MFB and MFE inside the range of acceptability of the model.

Fig. 6 : This figure is not really useful, just the statistics on MFB and MFE would be sufficient for the understanding. I do not think their distribution as a function of concentration really brings something to the discussion

A > As already explained in point C of the reviewer analysis, Figure 6 and the representation of CHIMERE performance in terms of MFB and MFE give an insight of the capability of the model to reproduce PM$_{2.5}$ concentrations both in low and in high peak of concentrations. The reliability of a numerical model for Air quality purposes needs to be evaluated in the representation of both the general hourly pattern of concentrations but also in the discrimination of period of low and high concentrations with the available data for simulations.

Fig. 8 : please use the same x-axis for both panels to ease reading the figure

A > The x-axis of Figure 8 has been modified and it is now in the same format of the other Figures.

Fig. 9 : It is not useful to compare modelled values in Nanyuki to observed values in Nyeri, 60km away in a mountain / plateau environment. No statistical link between the two timeseries can be expected a priori. I do not understand the point of the authors here, this should maybe be explained more.

A > The analysis of concentrations observed in Nanyuki takes in account that the location chosen by Pope et al. (2018) for the sampling of PM was a rural spot in a location of minimum local air pollution chosen to calculate the net urban increment subtracting the rural background concentrations of Nanyuki from the urban concentrations in Nairobi. The comparison that is proposed by Figure 8 it is only one of the options that can be taken in account considering the combined effect of meteorological parameters and location with higher contamination levels near Nanyuki that could influence the local level of PM. A first element to take in account to explain the peaks of contamination in Nanyuki could be the presence of local sources not accounted in the emission inventory used in CHIMERE. Despite this there is a clear change of trend in the concentration levels between February and March, in presence of local sources misrepresented we should see peaks at high concentration also in March but instead they are absent. A second element to take in account is the possible presence of precipitation during the period of March were the average concentrations of PM$_{2.5}$ doesn't exceed the 2 μg/m^3 but (Pope et al., 2018) affirm In their work that no rain was observed in that period and WRF model also doesn't model any in that particular period.

We are aware that to support the thesis of transport phenomena additional further analysis (e.g., trajectory analysis) are required as well as more observational point along the way between Nyeri

and Nanyuki. Further analyses are planned to go in that direction, what we argue in this paper is to give a possible explanation with the extent of the data available at the moment.

l. 643 : the authors claim that Nyeri is 0.43°N but n 10 on Fig. 7 seems to be at 0.43°S (or at least clearly nor 0.43°N)

A > The typo in the manuscript has been modified according to the suggestion of the reviewer.

References:

1. Pope, F. D., Gatari, M., Ng'ang'a, D., Poynter, A., and Blake, R.: Airborne particulate matter monitoring in Kenya using calibrated low-cost sensors, Atmospheric Chemistry and Physics, 18, 15403-15418, 10.5194/acp-18-15403-2018, 2018.