I. Editor comments related to the response to the previous referee comments

Referee #3:

1) Q. Not entirely clear how statistical measures are averaged (lines 336-338). Are measures calculated for each site and then those values for each domain are averaged (e.g., the 5 Relative Humidity NRMSE values for the 5 KEN2K weather stations are averaged to produce the KEN2K NRMSE value?) Or are the observed and modelled data for all the sites within a domain used together to calculate the average measure?

A > The statistical analysis both for WRF and for CHIMERE has been done calculating the statistics for each station individually and the averaging all station together so that e.g., the 5 values of the individual relative humidity NRMSE are averaged to produce the final NRMSE value for the domain. The calculation has been done on the original hourly values from observations and model outputs and consider hourly values from the model only if the corresponding hourly observation is present. According to comments made by reviewer 4 and 5, MNB and RMSE have been substituted by MFB and MFE in the validation of WRF and CHIMERE.

Editor: Is this information included in the text? If not, please do so.

2) Q. It would be helpful to specify how wind direction statistics were calculated. Since wind direction is a circular variable, calculating means, RMSE, etc. is different than for linear variables. Also, I'm not sure that normalized measures, MNB, NRMSE make sense for wind direction.

A > The statistics presented originally in the manuscript has been calculated as follows:

$MNB = \sum (Mi - Oi)\ n\ i{=}1 \sum (Oi)\ n\ i{=}1$

$RMSE = \sqrt{\sum (Mi - Oi)\ n\ 2\ i{=}1\ n}$

As the review suggests these operators can be used for linear variables such as temperature and relative humidity but they haven't the same meaning for what concern circular variables like in the case of the wind direction. Moreover, they rely also on the number of observations point included in the denominator and the final value can be misleading. For this reason, the statistical analysis in the new manuscript has been changed and the MNB and RMSE values substituted with mean fractional bias and error (MFB and MFE) originally used only for the validation of CHIMERE. Moreover, for WRF we also use the Index of Agreement calculated as follows:

$IOA = 1 - [\sum (O-M)\ n\ 2\ i{=}1 \sum n\ (|M-\bar{O}|+|O-\bar{O}|)2\ i{=}1\ ]$

Editor: Please add these equations to the manuscript, equivalently to Eqs. 2 and 3

3) Q. In the discussion of statistical evaluation of meteorological parameters it would be helpful to include criteria for what constitutes "good agreement" (line 361), "acceptable agreement" (line 443), etc.

A > These qualitative terms have been deleted and the paragraphs modified to include quantitative statements.

Editor: In lines 522 and 525, you still use 'acceptable' , in l. 410 'reasonable' without quantifying it. Please either avoid such statements or define them properly.

In particular, you may want to consider whether the paragraph at the end of section 3.1.2 is completely needed. – What do you want to say here – specify 'acceptable' for what.

4) Q. In Figure 8 the data for Nanyuki show what appears to be a nearly constant baseline PM2.5 concentration of around 2 to 2.5 µg m-3 . Why would this be occurring?

A > The observations used to validate CHIMERE performance for Kenya comes from previous work by Pope et al., 2018 [1]. In that work the site of Nanyuki was chosen as rural spot in a location of minimum local air pollution influence. The data from Nanyuki has been used for the calculation of the net urban

increment subtracting the rural background concentrations of Nanyuki from the urban concentrations in Nairobi. The average concentrations around 2 µg m-3 in the period between the 4th and the 11th are the levels of the rural background in absence of any external influence from meteorological parameters and in absence of local sources. The peak of concentrations visible is the other days are between 4 and 15 µg/m3 that is in any case a low value in comparison with the concentrations from the urban area. The difference in the baseline concentrations is given by the big difference between the days with possible transport of pollutants from days where this phenomenon is not visible, but it is exaggerated by the low scale of the concentrations (0-16 µg m-3)

Editor: Please add the relevant information on baseline PM to the manuscript.

5) Q. In presenting data table results, the text is often mainly just stating the values that are already shown in the tables. (e.g., sections 3.1.2, 3.2.1, 3.2.2) These sections could be condensed and/or modified to include additional description and discussion of what the data values mean.

Editor: In the revised manuscript, there are still instances of such descriptive text only listing values that are reported in the table without any discussion. Condense such texts (e.g. l. 418 -443) and add more interpretation as it has been done around l. 400.

6) Technical corrections: Throughout the manuscript the authors mention "low air quality index". This could be interpreted as a low numerical value of the air quality index, indicating good air quality, but from the context it seems the authors are instead describing poor, or low, air quality. It would be better to use a different word than "low".

Editor: This comment was not addressed in the previous response. It is a fair concern as 'low air quality index' may be interpreted either as 'low index for air quality' or 'index for low air quality'. Please address it and replace 'low air quality (index)' by a less ambiguous expression.

Referee #4:

1) Introductory comment:

.... 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.

Editor: I agree with the referee that your article raises many important questions regarding air quality in Africa. I respect your response that your study is focused on presenting the model performance for a few locations but may be a considered a starting point for future analyses and additional aspects. Your paper would indeed benefit if you added a few sentences along those lines towards the end of the paper as an outlook on further research question that should be explored in forthcoming studies. This could be part of the conclusion section.

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 PM2.5 doesn't exceed the 2 µg/m3 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.
Editor: I am not convinced that your response fully addresses the referee's concern. Please explain in the manuscript why the comparison as performed in Figure 8 is justified.

II. Additional Editor comments:

A. Comments regarding content and structure

l. 15: Add the model resolution here.

l. 219: (1) which conversion factor from organic carbon to aerosol mass was applied? (2) replace 'for' by 'with' (multiplied with...)

l. 220: Why is it assumed that PM2.5 is only composed of carbon-containing components? How about other compounds, such as sulfate etc?

l. 320 – 325: This paragraph is neither a result nor a discussion of your results. Therefore, either connect it better to the results or remove it, as it seems out of place and redundant here.

l. 327 - 375: This text is still a description of the methodology and therefore should be a subsection of Section 2. Lines 327 – 349 could be included in a subsection 'Statistical parameters'; l. 350 – 375 describes 'Model resolution and simulations'.

l. 585: ' CHIMERE 'better reproduces' than what?

l. 682 - 693: The model-observation comparison in Fig 7a shows clearly that the model tends to overestimate the PM2.5 concentration. If the emissions in the model were correct, one would expect the opposite trend – as you correctly describe, i.e. lower predicted values as compared to observations since the latter represent point measurements whereas the former are gird-averaged values.
However, in Figure 7a, there seems to be a period where model/observation agreement is particularly poor (~ 28/02 – 05/03) that shows a very distinct trend, opposite to the expected one. What was

different during this period? If indeed this discrepancy is due to an incomplete/inappropriate emission inventory in the model, can the characteristics of the air mass give a hint on the missing/wrong emissions as a function of air mass type/history?

Table 7: Is this table necessary? It provides the same information as in l. 794 – 798 and in some of the following lines. I suggest removing it as it is neither a result nor part of their discussion.

l. 847: The simulation of 'weather patterns' were not the main goal of this study but simulation of trends of air pollution.

B. Technical comments (language, journal standards etc)

l. 18: replace 'tool' by 'model'

l. 37: define 'WWP' and add it as database to reference list as detailed on the journal website https://www.atmospheric-chemistry-and-physics.net/submission.html#manuscriptcomposition

l. 43: Can you give a reference to the data base? – Add to reference list.

l. 54: add 'UN Habitat, 2017' to reference list

l. 98/99: A verb seems to be missing in this sentence.

l. 123: (1) Figures should be numbered according to their reference in the text. Since here Fig. 3 is cited before Figure 2, please change them accordingly. (2) remove 'a, b, c' here and in the remainder if the manuscript – see my comment below regarding 'panel 3d'.

l. 247 – 249: This sentence is quite convoluted given its rather simple message. How about "The emissions used in this work might not reflect the true values due to missing emission sources and the mismatch of the simulated time period and the date of the emission inventories. "

l. 276 – 278: I do not understand this sentence.

Table 2: (1) Spell out Latitude, Longitude, Elevation. (2) Use consistent terminology for latitude. Here you use – whereas later in the text, you specify S, N.

Figure 3: Remove 'd)' from the last panel. It is a legend and therefore does not need a label. In the caption, replace 'in table d' by 'in the legend'.

Table 3: Do not use random abbreviations in the table and caption. Spell out all words (obs., rel., ...) or define them in the caption (e.g. relative humidity (RH) which then can be used as RH in the table).

l. 321: (1) remove 'from the real world'. (2) replace 'systems' by 'simulations'

l. 345 & 347: See my comment above regarding referring to Figures in the correct order. For simplicity, I suggest removing the text in the parentheses here. You can refer to it later.

l. 603: A subject is missing in this sentence (that starts with 'Is therefore...')

l. 721 and 729: Correct the units (m s^{-1})

l. 756: ‘and’ seems wrong here (‘and large hotspots...’) – should it read ‘a’?

l. 786: why ‘e.g.’? Is 25 μg m^{-3} a limit set by the WHO for comparable areas?
Table 6: (1) Replace ‘WHO exceeding limit’ by ‘Exceedances of WHO limit’; (2) The last two columns do not include essential information: ‘Ratio’ is unclear and not very meaningful; ‘model false positive’ is described in the text and therefore does not need to be repeated here.

Data availability: Please provide at a minimum the model input and output data in a public repository https://www.atmospheric-chemistry-and-physics.net/policies/data_policy.html

1127 – 1129: Provide complete information for these references.