The authors acknowledge the referee for spending time and providing valuable and punctilious comments. Although most of the suggestions were applied in the revised paper, here are a few points to be mentioned:

The study assesses the modeled surface ozone variability over two parts, north and northwestern, of Iran. Two CAMS products, namely reanalysis and daily forecast, were evaluated against surface-based measurements at the city of Tehran (a megacity) and Tabriz (an industrialized city).

Please note that in this document, the tables and figures are referred based on their numbers in the revised article.

- Although most studies (prefer to) evaluate global models versus observations at the background (rural) sites, neglecting urban stations leads to a biased evaluation, in particular where a large fraction of the grid box is associated with urban conditions, e.g., megacities. Assuming a single grid box over the cities, that can give some relevant information about ozone variability in those regions. The regional nature of ozone, with a space scale up to 500 km, has been shown over different parts of the globe. Apart from it, using several stations distributed in a grid box provide more reliable results than the case of using one station per grid box; because the results can be affected depending on the position of the station, i.e., close to a grid point, or away from it. Most of the air quality monitoring stations in Iran are installed in the cities, as they are aimed for the public health report. There is no information about stations' type or availability of the data at background sites. The Geophysics station is located at the Geophysics institute, University of Tehran, Tehran. This station measures surface ozone, along with several other variables such as air temperature, nitrogen oxides, wind, total ozone column, etc. These data are often used for research studies.

- Global models with coarse resolutions might have deficiencies to describe the process at local scales, but the current generation of reanalysis datasets or global models with fine resolutions are expected to capture some of those processes (at least to some extent). The altitudes of the stations are larger than 1000 m, which might not be resolved by the models. So, the results of the reanalysis datasets are based on the surface model level (= 60). For forecast datasets, the optimum vertical model levels are shown in Table A1. In addition, the results at the surface model (= 137) of forecast datasets are discussed in Sect. 5.

- That was a great idea and described in the (revised) paper. The initial thought of the authors was to refer the reader to the other publications, but since each study uses a given version of the models, that was best to describe these aspects.

- Data series of ozone and several proxies are taken for all datasets, i.e., observations and both models. The predictors were chosen based on data availabilities and literature, in which these quantities are often used as proxies in explaining ozone variabilities. Nevertheless, there are studies in which other quantities such as radiation and cloud cover are used as ozone predictors. The first point to be mentioned is that there is a significant correlation

between temperature and radiations, so using both quantities may not add (or explain) more information. That can be tested using AIC, BIC, and lasso. Second, the point of this analysis is to compare the relationship between these quantities and ozone in reality and simulations, so the data needs to be available for three datasets. For easier comparisons of the results (both components of three datasets), they are shown in one figure (Fig. 4) instead of listing in a table.

The suggestions were to the point. The mentioned plots were meant to show the components over a given time period, but the suggestion is right. Therefore, the plot (Fig. 2) was created upon the referee's opinion. The naming conventions had arisen from the programming style of the authors, where 'df' refers to the data frame. Table 1 shows the coefficients for standardized datasets to show the relative importance of the variables. They only depend on the standard deviations, which are given in table A2 and table A3. The coefficients for non-standardized data are shown in table A4. To have a co-located data series of ozone and predictors, we selected datasets of station 8. Although model observation comparisons based on aggregated data and associative metrics provide information about the model performance; they often describe major features, and do not target the sources of error. Here we present an approach to compare observation and models that focus on given timescales. Scale analysis helps to identify the possible sources of errors, and underlying physical or chemical processes in each model. Metric used in the study helps to find the nature of the errors (bias, model-related, or observation-related error), which can be a diagnostic tool helping developers to improve the models. We used a regression model to infer the relevant processes which could have generated the model errors. Furthermore, it can be used to understand how the relationship between ozone and its proxies changes on various timescales.