Geosci. Model Dev. Discuss., 6, C990–C995, 2013
www.geosci-model-dev-discuss.net/6/C990/2013/

Geoscientific Model Development Discussions Open Access

Interactive comment on "Evaluation of the United States National Air Quality Forecast Capability experimental real-time predictions in 2010 using Air Quality System ozone and NO_2 measurements" *by* T. Chai et al.

Anonymous Referee #2

Received and published: 10 July 2013

1 General comments

This paper provides a detailed summary of the forecast performance of the "US National Air Quality Forecast Capabillity" for the year 2010, compared to the routine surface measurements of ozone and NO2 over CONUS. The main aims of the paper is the documentation of the absolute performance, with little interpretation of the results. Despite this, the NAQFC is an important system and it is good that such statistics are available in the literature as a reference. As such I am in favour of publishing the paper

after my remarks, given below, have been incorporated.

The other aim of the paper is mentioned as "a view towards" the improvement of the model (see, e.g. the abstract). However, the discussion and interpretation of the results is very limited, and for this aspect the authors basically refer to the preprint of Stajner et al. 2013. To my opinion the line may be removed from the abstract (line 9), limiting the aim of the paper to the documentation of the performance of the NAQFC system against routine air quality observations.

The paper evaluates only the first 24 hours of the forecast. Why this limitation? It would be valuable to have at least one figure added which discusses the performance of the second forecast day compared to the first day. I suggest that such a result is added before publishing.

In the abstract I miss references to similar forecast systems worldwide (Europe, Asia). For instance the European model air quality forecast ensemble (gmes-atmosphere.eu) is an interesting comparable capability which should be referred to and the performance may be compared. References should be added, as well as some remarks on the comparisons.

To my opinion the text can be shortened in several places. The paper is written carefully and provides all the details needed to understand what is done. However, the text in section 4 has several long discussions listing the details in the figures. The figures and tables contain this information already and to my opinion the text should only highlight and summarise the main features (and not the details). I would suggest that the authors look where the text may be shortened.

I was a bit surprised the authors limit the statistics metrics to rather traditional bias, RMSE and exceedance scores. In particular, RMSE is well-known to be sensitive to outliers. Furthermore, when the bias is large, RMSE is not an independent measure. Please motivate the choice of metrics.

The paper of Stajner, 2013 is not available to me. This seems to contain a lot of the interpretation of the results presented here. The authors should make sure that the overlap is minimal.

Sensors with molybdenum converters suffer from an overestimate of NO2. From the paper it is not clear to me if corrections were applied to the NO2 data to account for this issue.

Why is especially the SouthEast difficult to model correctly ?

2 Detailed comments

2.1 Section 2

Please discuss also the boundary conditions ? Biases in ozone could partly be caused by long-range transport and influx from the stratosphere. What is known from previous work on the quality of these aspects of the model?

Wildfire emissions: these are based on a multi-year averaged emission. Wildfires are an important source of variability and there are several efforts worldwide to use space observations to improve estimates on a monthly/weekly/daily time scale, also in near-real time. Please comment.

What about the 100 hPa "zero flux assumption" (p2615): does this lead to a reasonable ozone concentration in the upper troposphere? Is there previous work studying the free troposphere ozone concentrations of the system?

2.2 Section 3

Near-real time data: Please expand the discussion on how the near-real time data compare to the quality controlled data sets which become available after some time. Are there many outliers?

p17 l15: "valid hourly observations": What does "valid" mean?

p17 l19: 10000 observations: please also mention how many stations are involved.

Fig 2: the term "overlapped" is not clear: is it the number of overlapping observations, or the total after removal of the overlap. If observations overlap, is one of them kept? What is "overlap*" ? This should be explained also in the caption of the figure.

p18 l18: in-line formula: why is there a division by 10 ?

Section 3 is a bit long and may be condensed. The overlap and time shift issues are details.

I was wondering about the NO2 observations. Sensors with molybdenum converters suffer from an overestimate of NO2, see e.g. Steinbacher et al, doi 10.1029/2006JD007971. Were corrections applied? Please discuss this.

2.3 Section 4

Why do authors use RMSE: this is sensitive to outliers. The mean-absolute difference would be a good alternative!? RMSE has large contribution coming from the bias, so it is not an independent statistical measure.

Why are there different panels for July and August ? If the conclusions are similar these may be combined. For me it would be more useful to see Summer JJA and winter DJF results.

p21, l4: It is not surprising the Rocky mountains have teh smallest bias because NOx is low in this region. Please quote also the relative errors compared to e.g. monthly and regional mean. In general, throughout the paper, it would be useful to have errors also as relative numbers (

p22: The exceedance scores are influenced considerably by the large bias. In Fig. 10 "d" is close to zero. Please discuss this point.

There is a lot of tables and a lot of numbers provided. This is in part useful (e.g. to document the results for the different regons) but sometimes similar results are shown in different tables. In particular, I would suggest to remove table 5 (Summer only is enough because that is when exceedances occur). In fig.6 it may be considered to remove one of the two limits (70 or 75) because they are close together and the results are similar. Alternative: for 70 keep only the CONUS row.

2.4 Section 5

p28 l14 " . . . in order to expose systematic model errors, which could be corrected in the future to improve NAQFC predictions " There is little interpretation of the results. Is it really the emissions which are to blame? Could it be a lifetime issue as well. What about free troposphere, influx from stratosphere ?

p29, l8 "monthly mean profiles from global model simulations for most chemical species" Which model is used? Has this been validated? Provide a reference please.

l9: "Dry deposition was modified based on the Monin-Obukhov similarity theory as well as by 10 including canopy height and density based on recent satellite observations" please provide a reference.

l11 "Planetary boundary layer height was constrained to be at least 50 m." How and when does this influence the results.

l14 "The emission data sets have been updated in June 2012, with a pronounced decrease in mobile NOx emissions." By how much?

Interactive comment on Geosci. Model Dev. Discuss., 6, 2609, 2013.