Earth System
Science
Data

Open Access

Discussions

Interactive comment on "Gridded global surface ozone metrics for atmospheric chemistry model evaluation" *by* E. D. Sofen et al.

E. D. Sofen et al.

esofen@gmail.com

We thank the reviewer for their feedback. We provide their comments with our responses in bold below. As noted in our response to M.G. Schultz, drawing on both of their comments, we have produced an updated version of the dataset that is available from the BADC here: http://catalogue.ceda.ac.uk/uuid/ c0455ab814224a05ac74642d3d44b73e

REFEREE 2:

This dataset represents a tremendous effort to meld disparate datasets, overcoming numerous barriers that generally prevent models from being evaluated with a full suite of available data. It will undoubtedly become a highly referenced paper, and the inclu-

sion of non-traditional statistics including those relevant for meeting policy goals may lead to their becoming routinely used in model evaluations of surface ozone distributions. The paper is well-written with appropriate detail describing methodology, with only minor suggestions for improving clarity described below. It will be important to the atmospheric chemistry and air pollution modeling communities to keep this dataset current as new data are brought online, particularly in areas of the world that are not well observed. The recommendations section is a useful summary for those making the measurements, and also for the modelers using the dataset to be mindful of the various uncertainties involved in using the measurements. The inclusion of the number of sites, and variability across sites and time in the dataset is useful.

Detailed comments:

The U.S. EPA AQS dataset should go back to 1980. Was the earlier data all excluded because of the year-round requirement? While discarding the summer-only and urban sites is understandable, it may limit the utility of the dataset for those interested in summertime air quality, and perhaps there could be value in including the urban values for a spatial average especially if the classification isn't always accurate.

The EPA AQS data for 1980–1989 was not directly available from the EPA website, but we have since requested it and added it to the new version (2.7) of our dataset. We have applied the same summer-only and urban screens to the early EPA data.

A range of other data were potentially available to us but we did not want to end up in a position where we would have to negotiate individual agreements with a range of organizations to be able to release the data. We note that the IGAC sponsored TOAR project is undertaking this kind of negotiation.

How is the spatial gridding done? Is any effort made to grid at a finer scale first, and then to a larger grid or are all sites given equal weight? Gridding at a finer scale first might allow for inclusion of urban sites without unduly influencing the regional grid-cell

average.

We agree with the reviewer that there are a range of challenges in how the averages are calculated. There have been some recent efforts to use more complex approaches [Schell, 2014] but for the sake of simplicity we have used a simple averaging excluding sites that we do not feel are representative of the regional concentration. We have added comments in this regard to section 6 of the text.

On a related note, could a flag be included to alert users to when sites measuring free tropospheric air have been merged with ground-level sites? Modelers will often sample above the surface layer when comparing with mountain sites, and perhaps in this case it would be more appropriate to average multiple model layers together for a clean comparison?

As discussed above, version 2.7 of the dataset includes only those sites below 1500 m which should provide a better representation of boundary layer ozone without mountaintop sites.

For metrics that depend on daylight hours or the local time of day (like MDA8), are all of the calculations done initially in local time? Seeing as only overall statistics are reported, it probably doesn't matter for this dataset, but it would be nice as a quality check if the authors could confirm that their statistics, at least for one site, match those provided, for example by U.S. EPA.

MDA8 is calculated in UTC for the sake of simplicity and for comparison to global models which often output in UTC. This is described in the text in Section 6. AOT40 does need a local time calculation as it uses hours of daylight. Our local time calculation for AOT40 is based on longitude as described in Section 6 and Table 3. We have evaluated by hand the numbers for a range of sites to ensure algorithmic accuracy.

The time coordinates in seconds since 1970 is awkward. Would it be possible to pro-

vide a second variable that has the time in a more user-friendly, human readable ver sion such as '1970, 1971, 1972. . .'?

Time in seconds since 00:00:00 1 Jan 1970 is equivalent to seconds since the "Unix epoch" and is a standard computer time format. Most modern programming languages and netCDF reading utilities can automatically convert from this time convention to a calendar date-time. Furthermore, it is the only datetime format in netCDF that complies with the CF specifications.

P606 L13 But these aren't forecast models, consider replacing predictions with simulations or projections although it's not clear why this dataset is useful for evaluating future ozone levels. **Changed from "predictions" to "simulations" and reworded the sentence to read "Our confidence in the unbiased nature of these models and hence the value of these models in developing mitigation policies is assessed by comparing model simulations against observations."**

P607 L10-15 might also note the evaluation of multiple global models with MDA8 in Reidmiller et al., 2009 See http://www.atmos-chem-phys.net/9/5027/2009/acp-9-5027-2009.pdf

Thank you for this suggestion. It has been added to the list of relevant papers on ozone model evaluation.

P611 Are these necessarily background conditions because urban areas are excluded, or is it more accurate to describe it as regional average conditions? **We agree that is is probably better to refer to a "regional average" and have adopted this description throughout the manuscript.**

P617 L19-20 reference needed for the situation for carbon gases? **Added reference to Dlugokencky et al., 2005 that describes the intercalibration of CH4 standards.**

P618 L8 comforting → useful?

Done

P618 L15-17 It's unclear here if the routine was in error or if this discovery led to further cleaning up of the original datasets.

Yes. There were cases where where data/metadata inconsistencies led to large errors (large overlaps between timeseries), but once caught by the overlap-checker, additional development of the data parsing routines led to improved data quality. Modified the sentence to read "Discovery of long overlaps (up to a year) has led to improvements in the date parsing routines, so that long over-laps are eliminate in the final version of the dataset."

P622 L5 Why are summertime-only sites not screened out here? Do they have more than 9 months of data?

The algorithm to remove those sites that are 'summertime-only' (or have an ozone season less than 9 months) is not perfect, and a small number of sites slipped through the summer time removal process and the signal seen reflects this.

P623 L22. Are the standard deviations here reflecting larger amplitude seasonal cycles?

Yes. We have a paper in preparation which is providing a systematic analysis of the dataset and will show that these regions have both a large seasonal and diurnal cycle driven by photochemical activity.

P628 L27 Consider cutting "available for model evaluation" as it seems this expansion of measurements would be more broadly useful. **Done**

Table 3 clarify if MDA8 is calculated based on local time 24-hour periods **Done**

Table A1. Please provide the country name as in A2. **Done**

Figure 1. The white text is hard to make out, consider lighter shading and black text? Why is metadata listed multiple times in the brown boxes for the individual datasets?

We have revised this figure to use lighter boxes and black text (See attached). The multiple metadata boxes for individual datasets are meant to illustrate that some datasets have multiple sources of metadata. For example, EMEP provides most metadata within the data file, but also requires some auxiliary metadata files for the site latitude/longitude. We now describe this in the caption.

Figure 5. Note in caption the R greater than 0.707 criterion used here. **Done**

Figure 6. Enlarging panels would help readability. **We ask the editor to convert the figure to a 2-panel figure.**

Interactive comment on Earth Syst. Sci. Data Discuss., 8, 603, 2015.

Fig. 1.
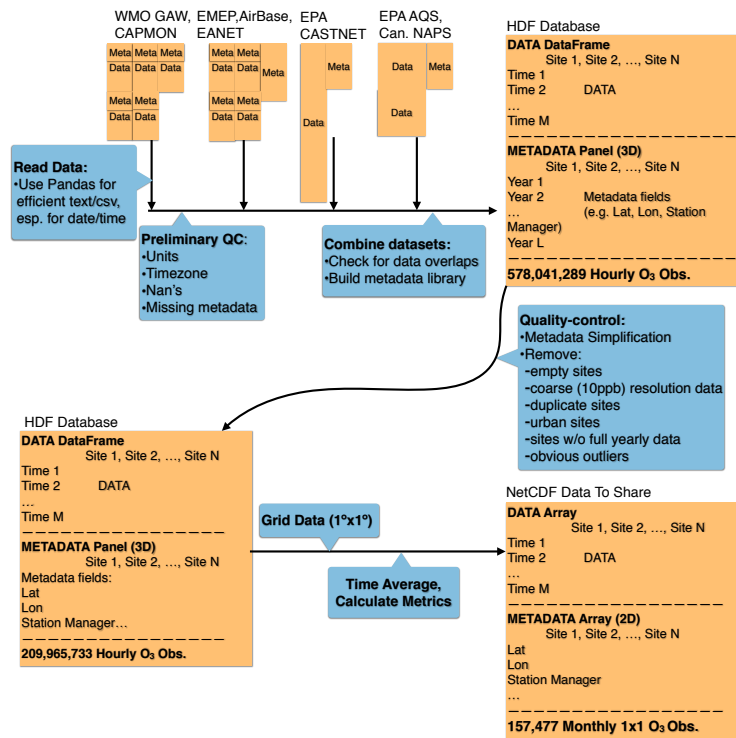
C424