Earth Syst. Sci. Data Discuss., 8, C320–C323, 2015
www.earth-syst-sci-data-discuss.net/8/C320/2015/

Open Access Earth System Science Data Discussions

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

Anonymous Referee #2

Received and published: 15 September 2015

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 inclusion 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.

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.

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?

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.

The time coordinates in seconds since 1970 is awkward. Would it be possible to provide a second variable that has the time in a more user-friendly, human readable version such as '1970, 1971, 1972. . .'?

P606 L13 But these aren't forecast models, consider replacing predictions with simula-

tions or projections although it's not clear why this dataset is useful for evaluating future ozone levels.

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

P611 Are these necessarily background conditions because urban areas are excluded, or is it more accurate to describe it as regional average conditions?

P617 L19-20 reference needed for the situation for carbon gases?

P618 L8 comforting → useful?

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.

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

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

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

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

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

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?

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

Figure 6. Enlarging panels would help readability.

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