

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

Received and published: 22 December 2015

We thank Martin for his thoughtful and thorough feedback on our manuscript. His comments are below with our responses in bold. We wish to highlight that we have produced an updated dataset to reflect some of the suggestions provided by both of the reviewers. The new dataset can be found here: <http://catalogue.veda.ac.uk/uuid/c0455ab814224a05ac74642d3d44b73e>

M.G. Schultz:

The article by Sofen et al. describes a major undertaking which aims at pulling together all available surface ozone datasets from publically accessible network data centers across the world in order to make them available for (global) model evaluation in a

C410

harmonized, gridded format. This dataset constitutes a valuable asset to the global atmospheric chemistry modelling community and the authors need to be applauded for their rigorous efforts to collect the data, identify metadata and data issues, and harmonize the datasets into a single collection.

As is well described in the article, the processing of thousands of datasets from different networks constitutes a great challenge, and not all existing data centers have yet managed to provide all their datasets in fully quality-controlled and harmonized form. It can only be hoped that this article, alongside with other efforts within GAW, TOAR, and other communities, will lead to a better “global view” on these data which are often sampled in order to adhere to national or regional legal requirements.

The article is generally well written and concise, but I do have some issues with the Introduction section which are outlined in the detailed comments below. I also noticed one or two minor flaws with the applied methods or their description. Specifically, there is no explanation of any “data capture” criterion, which makes it difficult to understand section 3.6 on seasonal AQS sites. Also, I am somewhat concerned about the gridding procedure if high and low altitude sites are merged together in one grid box. Ozone mixing ratios can vary by up to 20 ppbv within 2000 m altitude, so the gridded product might be useless in mountain regions. At least some assessment of this effect should be given.

All-in-all this article describes a unique dataset which will be of great usefulness for model evaluation. Judging on the completeness of the data is difficult, because (i) as the authors themselves note it is limited to data from publicly accessible archives, and (ii) the content of the original data archives keeps changing, which means that this dataset represents a snapshot at a given (download) time. In any case this dataset provides a much more complete view on global surface ozone than what has been available in previous model evaluation exercises.

I recommend publication of this article after the following minor issues have been taken

C411

into account.

Detailed comments:

Abstract: spell out metrics SOMO35 and AOT40

Done. Added “sum of MDA8s over 35 ppb (SOMO35), accumulated ozone exposure above a threshold of 40 ppbv (AOT40)”

Introduction:

p. 606 line 5 – remove sentence “Overall ...” – this is an unnecessary motherhood statement

Done

Line 8 – I don’t fully agree with the statement “they have focused heavily . . .”; there are different communities who employ CTMs for other purposes such as aerosol modelling, mercury, POPs or other substances as well. You may say that there is a strong scientific community who have focused. . .

Done

Line 13 – “against”. Perhaps re-phrase that ozone is a central species in such evaluations?

Reworded.

Line 15 – “limited by the availability or type of observations”. This mixes two very different aspects in one rather general statement. Better to separate these two issues.

We have restructured the paragraph to address the issues in a different way.

Line 18 – The discussion of the issues with satellite data is too vague (“makes understanding the vertical distribution difficult”). It is simply physically impossible to obtain a good vertical resolution in the troposphere from spectral irradiances measured by the satellites. Even though there are now products emerging which claim to have at least

C412

some resolution in the troposphere, one can at most distinguish between upper and lower troposphere. Please state more explicitly that satellite retrievals of ozone do not allow for an adequate vertical resolution in the troposphere.

Revised to say “which can only provide vertical resolution that, at best, can separate the upper troposphere from the lower troposphere”

Line 24 – I don’t like the implication here that surface observations are only useful because there is nothing else available. There are good technical (and programmatic) reasons why people measure at ground-level (for example it is much easier to maintain stable instrument characteristics and calibrations on the ground compared to aircraft measurements). Furthermore, the boundary layer, where the stations reside, is often de-coupled from the free troposphere above, and aircraft usually can perform only very little sampling in the boundary layer. Hence, surface observations sample a different part of the atmosphere which is particularly relevant, because this is the air we (and the plants) breathe.

This is a good point. It was not our intention to imply that surface observations are not as good as other sources; only that from the perspective of evaluating a 3-D model, they only provide a (roughly) 2-D perspective. This section has been reworded to say “Much of the model-measurement comparison has relied upon comparisons with surface observations. Surface observations have the advantage of occurring at relatively high time resolution (typically hourly) and at fixed locations that can, for some sites, go back decades with very stable instruments. Surface observations also have the advantage of observing the portion of the atmosphere that is most relevant to human and plant life.”

p. 608 line 10 – “For convenience”; I don’t think that this is the major reason for (regional) model studies to have limited their evaluation to one region. Most of these models were built to support regional air pollution abatement policies, and so it is natural that they would choose data from their region as the only battlefield. I think it would

C413

be worthwhile if the authors included a somewhat more meaningful discussion about the differences between regional air quality and global model evaluation.

Our intention here was to point out that evaluation of global models has generally been done using regional datasets. However, the start of the paragraph has been rewritten to read: “Most previous model evaluations have emphasized one regional regulatory or background network for their comparisons. Regional models are evaluated over the domain for which they have been developed in support of air quality policy. In the evaluation of global models, the comparison with a single network, while expedient, has tended to ignore the other networks and datasets which obviously limits the veracity of the comparison.”

Section Datasets:

p. 610 line 3 – the authors should add a sentence here when these data were downloaded or last updated. The content of the archives changes frequently and it is important to have a reference date for the dataset described in this manuscript.

Data was last downloaded 29 June 2015. Added to manuscript.

Line 26 – maybe worthwhile mentioning that Airbase also contains data from some overseas representations (for example in the Carribean).

Done

p. 610 line 10 – “are removed.” Please add “in this analysis.”

Done

p. 611 line 24 – “unlikely to adhere to quality standards”. Strong statement. How would the authors know if they cannot access these data? At least replace “and” by “or” in line 25?

Changed “and” to “or”. During the development of the dataset, we downloaded some small datasets from other countries (generally downloaded one-at-a-time)

C414

such as South Africa and Chile and found that data and metadata quality was often quite poor based on visual inspection of time series, at which point efforts to assimilate these smaller datasets were abandoned.

p. 612 line 1 – for your information: at a recent meeting on GAW data management strategies the conclusion was almost opposite to your recommendation. Rather than one central archive one wants to work towards a federated, distributed archive. There are very strong programmatic and funding reasons why several data archives exist, and this will not change. Thus, rather than trying to bring everything together in one place the strategy is to harmonize data access and formats and build a central portal from where the individual archives can be accessed, ideally entirely transparent for the user.

Thank you for passing along this update from the GAW community. This is a good point. To the end user, a centralized archive or a portal to individual archives that can provide data in a harmonized format will hopefully serve the same purpose. We have added “ or a data portal and access interface that can provide data in a uniform manner,” to this sentence.

p. 614 line 2 – AQS sites do come with a station identifier “AQS site id” in the metadata file (format “NN-NNN-NNNN”)

AQS distributes their data in a couple of different formats. In the version we are working with (The Hourly_44201_YYYY.csv files), there is not a site ID given. However, there are separate “State Code”, “County Code”, and “Site Num” that are 2, 3, and 4 digits, respectively, that we use to construct a site ID that do follow the pattern you describe and is described in the AirData FAQ (http://www.epa.gov/airdata/ad_about_reports.html). We have clarified this description in the text.

p. 617 line 9 – Units correct? “ppmv ppbv-1”?

C415

Thank you for catching this. It appears to be an error introduced during typesetting. We meant that units are output as ppmv or ppbv by default.

p. 618/619 – there is no mentioning of data quality flags in the original datasets. In our experience, many outliers disappear if these flags are correctly taken into account. It is a problem, though, that in some networks the definition of the flags is not very clear or changes between stations. Specifically with respect to “missing values” we also found that there are cases when these are not flagged as invalid data (and the code used to mark a value as missing, i.e. the number reported in lieu of an ozone value, can also vary – even within the time series of one site.

We thank the reviewer for pointing this out. We now include statements in the text about how we have used the flags provided by the data providers. When data flags are defined and provided, we use them to screen out flagged data, such as in the AirBase dataset. Regarding filler values for bad ozone values (e.g. -999, 99999), we screen for multiple filler values. All negative values are set to NaN, so that handles the most common filler values.

p. 620 lines 1-14 – it is not clear to me what is the intention of this analysis? Do the authors want to exclude sites with incomplete annual series? Or do they want to detect these and include them only in statistics over the summer months, for example? I believe the problem here is that a “9 months minimum” data capture criterion is referred to without introducing this first. In general, it is a bit difficult to follow the text in section 3 where stations are removed in various steps of the analysis, without having said at the beginning what the criteria are for a “good” data record.

Some EPA sites only report during the summer months. Their inclusion could potentially lead to a bias in the seasonal signal of those gridboxes. To err on the side caution we have not included them in the average products. We have explained this thinking in the text.

p. 621 line 23 – grammar “reported hour”? **Changed to “reported in each hour”**

C416

p. 625 – this might be the only concern I have about the robustness of the methods applied: when combining information from mountain and “valley” stations in one grid box, the information in these grid boxes might be much less meaningful than in flatter terrain. It might be useful to create an additional dataset which includes only stations below or above a given threshold altitude. At least the impact of this “altitude merging” should be assessed.

We agree with Martin that the altitude variations are a challenging problem to address. In the updated version 2.7 of the dataset that we have created, we have restricted the analysis to sites with elevation of less than 1500 m to remove most high-alpine (mountain-top) sites while still preserving many elevated regions, although some plateau areas of the Mountain West will be removed.

p. 638 Table 3 – I am somewhat doubtful about the usefulness of the air quality metrics DaysMDA8_ge_60ppb_Gridded_Ozone and FourthHighestMDA8_Gridded_Ozone in a gridded product. These metrics analyze the high end of the distributions, and these are highly variable across sites from a single region. Have the authors demonstrated that a model can develop any skill with respect to these metrics?

These products were requested by modelling groups notably regional modeller. We have added a footnote to Table 3 to say that they may be of limited usefulness for some model evaluations.

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

C417