Review of essd-2021-208 PAPILA dataset: a regional emission inventory of reactive gases for South America based on the combination of local and global information by Paula Castesana et al.

The paper describes a dataset intended on providing air quality and climate modellers with a complete dataset for South America (SA) by creating a so-called mosaic inventory. This implies using a complete but rather generic global dataset with less granularity than local or national data as a starting point and merging it with more detailed national scale inventories. The result of such a mosaic inventory still provides a complete dataset but with higher granularity and includes more local knowledge thereby providing modellers with a better starting point for their model exercises which can lead to more accurate (scientific) results and analysis. The merged dataset needs to be carefully evaluated and discrepancies explained and documented as a risk is that "apples and oranges" are treated in the same way and the end result can also be confusion. The advantage of the approach as also stressed by the authors is that the dataset can improve over time incorporating more data as they become available and the current PAPILA dataset is not intended nor expected to be the final dataset but a well-documented starting point that can improve over the years.

As said such datasets can be an asset for modellers and support / improve regional air quality analysis. The topic is fitting for ESSD and the paper could be published after several clarifications and improvements are made. .

My main concerns are on the discussion of the results and clarity of the associated messages.

- The idea of the mosaic approach is not new and has been successfully applied in the framework of theTask Force on Hemispheric Transport of Air Pollution (TF HTAP). In the introduction it would be good to refer to the HTAP v2.2 dataset as an example (Maenhout et al., 2015) including the fact that this dataset has been widely used also outside of HTAP. This could be added in the paragraph L68-73
- The discussion on the impact of the incorporation of the national/local data is unintentionally misleading. For example in L339 "resulting in a difference for SA of only -5%". Strictly speaking this is true BUT you have only replaced 3 countries of total SA and not even for all pollutants (Figure 1). When looking at Table 2 it can be deduced that these 3 countries are only responsible for 20-25% of the total SA emissions (in the case of CO). Thus over 75% of the emissions are unchanged and hence the impact on total SA is very limited. It should be explained in this way and then followed by what the impact is on the three countries that were adjusted. (For CO this is ~-20% which is substantially more than the -5% for all SA, and in fact is still a bit misleading because the countries compensate for each other but this is properly stated in L338.) This applies to all species (e.g. L365 NOx – same story, this 3% is caused by the bulk of the emission being unchanged because a.o. Brazil and Venezuela are kept constant)
- Modelling is done for MABA (Buenos Aires) only. It should be more clearly stated in the conclusions that these results do not give any insight in the performance for e.g. Colombia, Chile etc. SA is such a big area that a better performance over MABA cannot be taken as sign that also in other countries local data result in better AQ modelling results. I agree it is what we may expect / hope for but it is no prove whatsoever. As an outlook it could be mentioned that an evaluation of the PAPILA dataset over other regions where local data are integrated is not only highly recommended but truly needed to prove the added value of the approach. (see your conclusion section)

- L116 – the shipping might need a bit more discussion. Inland navigation is clear but the domestic coastal (and what do you mean with deep-sea ?) cannot be so easily separated from SHP-INT – how do you do this? Is that not leading to double counting? SHP-INT will be based on AIS signals but no separation is made between flag states, so how will e.g. Chilean coastal shipping be distinguished from SHP-INT? .
- L398. The main point is that CAMS treats countries uniform w/o correcting for the climatic zones in the country. However, SA countries are large and the climatic / temperature zones vary widely from tropics to antartica. In the discussion this could be mentioned and suggested as an improvement for global inventories as the temperature zones are well known and available on grids. Theis could be used to redistribute within a country.
- L504 – here and throughout the MS – where you use "driver" you mean "proxy". (Economic growth is a driver for growing emissions; population density is a proxy to distribute such emissions). Please check the whole section 3.3. and further for this terminology.
- The discussion in L505-511 is unclear, please rewrite and check for using the right words. (it can be a low-resolution map but not low-resolution density)
- L512 as should be or? ; population as a driver should be "population density as a proxy"
- L562 remove "which is reflected in the uneven levels of development of local inventories" as such it does not add information.
- Final paragraph – I think you can add that especially Brazil and 1 or 2 others should be added because the 3 countries added now are responsible for less than 25% of the emissions. How would this change if Brazil is addd?
- L571 sentence "in addition etc. can be removed, the sentence before is a stronger ending with a clear message.

Table A2 – is that really a summary of NOx ? or do you mean NH3? I'm surprised to see NOx emissions come from dairy cattle.

Minor issues

L5 place "derived" after "data"

L13 "obtained" should be "found"

L15 ....lower **levels**....

L20 replace: PAPILA-based modelling results had a lower bias for CO and NOx concentrations in winter while CAMS-based results for the same period tended to deliver an underestimation of these concentrations.

L28 From the energy standpoint ->   Regarding energy use,

L39 it's a bit awkward to use a reference of 20 years ago here. Much would / may have hanged since then. I think this is also discussed in Huneeus et al. 2020; might be a added here?

L43 change larger to increasing

L77 change bibliography to literature

L91 choose either LAC or SA and use only one. Please check full MS and SI for this

L98 2 times have

L99 Explain a bit better why you need complete country inventories (because this is the common / shared administrative entity that can be exchanged with the global inventory)

L140 – yes that is a solution but it is not really necessary, the combination of coordinates AND a country code could also generate unique values and the cell can than be shared by 2 or even 3 countries. The difficulty is probably that the global inventory does not have these country codes

Figure 1 – I would call that a Table and not a figure. Replace SA on the bottom with "Rest of SA" or "Other SA"

L156 explain how you distributed the rest as area sources? With what proxy?

L179 taken to -> converted to

L183 Such => This

L191 considered => assumed

L215 check word missing (we to report?)

L228 applied => apply

L378 minority => minor

Table 2. it would be good to indicate that the first 3 subregions are in Argentina and the last 3 in Chile. This is well explained in the text but for international readers it might still be difficult to remember this. Moreover it should be considered (but only a suggestion) if it is not better to e.g. in L250 where MABA is introduced to simply state that in this study Buenos Aires includes the entire Metropolitan area of BA. In the paper you can then discuss simply Buenos Aires and do not have to use this MABA which for many people will be something they are not familiar with. Same applies to MRS. Everybody knows Santiago is the capital of Chile, MRS will not be known to many. Also in the Table 2 acronyms like MABA and MRS are not very helpful. You could add a table footnote that the entire metropolitan region is included and use the city names in the table.

Table 2 – I suggest to only show a digit if the value is smaller than 100. E.g. 31225.2 is not very helpful. PAPILA NOx column has 2 times 16.8 – pls check if correct

Fig 3 . Pls consider if it is not better to change the order and show Argentina – followed by its 3 sub regions (same for Chile). This may help to see patterns. For example when looking at SO2 the pattern for ARG and sub region is exactly the same (so it's a national consistent difference) for Chile it is different with a redistribution between Antofagasta and MRS.

L415. How relevant are these emissions? NH3 is dominated by AGRI; 80% sounds big but if 80% of almost nothing is still nothing. Is it worth the attention?

L430 non => not

L435 see earlier remark on comparison for the entire region (which is not informative as bulk of the emissions are taken from CAMS and thus constant)

Table 3 – explain acronyms in table footnote (NMB, NMGE, FAC2)

L457 null => no

L460 goodness of the => better

L467 high => height

Fig 6 change top legend : remove "simul" (it is not a known word), make it bigger and have e.g. CAMS simulation; PAPILA simulation

L477 goodness of => quality

L491 than => as

L495 what is intended with hydraulic availability??? Change and explain better, different word?

L497 the incorporation of diesel-based motor generators => the use of diesel-fuelled generators

L504-505 Why does this imply a difficulty?

L520 tools => monitoring

L520-527 – Please rewrite this paragraph and check with co-authors; it is rather cryptic. E.g. "information generated from a local perspective" can simply be "local information"

L529 Why are aviation emissions not included? Are they not important? Not provided by EDGAR or CEDS?

L534 item v) – Please explain this better and be a bit more informative on its relevance – are these emissions missing or possibly reported elsewhere? If it is CO, NOX, SO2 it must be from waste burning not landfills or waste water. Is this not present for all SA counties? Or only for Chile and Columbia? When you look at the local information where you have these emissions included, how important is it? What is the share for the country or city? If it is missing but only adds e.g. 0.1% to the national total is it relevant?

L541 replicated in => addressed in the

L541 remove "for the case of"

Section 4 – I did not check the dataset and the sums with the tables. Please ask one of your 10 co-authors who was not involved in uploading the data to download and do a check. I assume it will be good but advise to do this check.

L552 made for a limited number of countries – change to "only for Argentina, Chile and Colombia" (be as specific as possible) and add that this is a living dataset and that in the future other countries can be added.

Refs

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmüller, R., Denier van der Gon, H., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411–11432, https://doi.org/10.5194/acp-15-11411-2015, 2015.