Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-1170-RC3, 2019

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Interpreting the time variability of world-wide GPS and GOME/SCIAMACHY integrated water vapour retrievals, using reanalyses as auxiliary tools" *by* Roeland Van Malderen et al.

Anonymous Referee #3

Received and published: 6 February 2019

Overall, this is an interesting topic, but I found it hard to figure out what the paper is trying to accomplish. The purpose of the paper needs to be more clearly discussed in the first few pages. I find that the comparison between the GPS, GOMESCIA and reanalysis IPW data valuable and fairly well described in the paper. I would like to see a clearer discussion of the effect of clear-sky bias on the GOMESCIA data, and whether this could be an explanation for its larger discrepancy from the other data sources. The sorting of histograms of IPW at different sites is also valuable and clearly extends the earlier work in this area. I would like to see a more organized discussion of the impact

of seasonal behavior on these histograms. The paper asserts that the seasonal behavior is important, but does not explicitly show that if the seasonal cycle is removed, the resulting distributions of the residuals are simpler (e.g gaussian or lof-normal). I am less convinced by the "step-wise multiple linear regression." First, I don't understand the name – what is step-wise about it? Second, so many potential explanatory variables are used (which are said to be at least partly independent), that at least some of them are likely to have explanatory "power" by chance for the relatively short times series studied. I would be happier if the authors could clearly state a hypothesis, and then test it with a more limited set of explanatory variables consistent with the hypothesis. I am not a fan of a "throw everything at it and see what sticks" approach. There are numerous cases of strange English usage/wording, some of which I mention below, but there are far more than I can explicitly call out. I recommend that the paper be edited by a native English speaker. Some more detailed comments below: Page 3, line 20-22. Strange Wording ("vastly"). Also, what does neutral mean in this context? Page 3, line 30. Strange wording ("disposed of"). Maybe change to "This process results in a world-wide. . .." Page 4, line 29. "downsized"? How was the conversion from 5-minute observations to 6 hourly observations performed? And why is this needed/appropriate considering that the reanalysis can be considered a snapshot at the synoptic times? (I'm not saying that the downsizing is wrong, I just want it explained better. Page 8. I can't follow the discussion of the clear sky/cloudy biases. Is the GOMESCIA only available in clear-sky conditions? If so, wouldn't it be expected that the GOMESCIA is biased low compared to measurements available in all-sky conditions? Page 9 and 10. Discussion of different distribution classes at different locations. It is unclear how the sorting into classes was performed. Are all types of fits tried, and then some criterion applied? Please explain more clearly. Also, if the various distributions mostly occur because of seasonal cycles, maybe it would be good to show analysis after the seasonal cycle is removed?? Page 11, lines 19-20. This could be tested by subsetting the reanalysis data so that it too has gaps at the same time as the GPS data. Page 16. I wonder if the reason for the poor fitting for the west coast of North America sites it

due to the fact that for much of this region, the rainy season is in the winter, where the temperatures are low. In the summer, it is dry but warmer, so there is not that much of a change in IPW. Other regions, such as eastern NA or Europe, there is still significant rainfall in the summer season, leading to a much larger correlation with surface temperature.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-1170, 2018.