2019-03-26

Review of

Consistency and representativeness of integrated water vapour from ground-based GPS observations and ERA-Interim reanalysis

Olivier Bock and Ana C. Parracho

Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-28

General Comments

The work presented focus on comparing atmospheric integrated water vapour (IWV) estimated from ground-based GPS observations with the corresponding IWV values in four nearby grid points of the ERA Interim reanalysis. The structure of the manuscript is straightforward and reasonably easy to follow, although I needed quite some time until I was familiar with the nomenclature and the symbols.

The stated motivation for the work was to identify GPS stations where ERA Interim is not recommended to be used when searching for inhomogeneities in the GPS time series of IWV.

A question that is not answered after reading the manuscript is an approximate quantitative relation between representativeness statistics and the size of the break in the GPS IWV time series. I think like this: if representativeness errors (at a specific GPS site) are stable with time, it should still be possible to detect a break in the GPS time series if it is above a certain size? It would be interesting to have the authors ideas about how large, or small, breaks that could be detected, given some example values of the representativeness statistics.

Figures 2–6 are presented and discussed in Section 3. Some of them have red dotted lines defining limits in order to identify outlying results/stations. However, it is only in Section 4 that these limits are explained. I think it would help the reader if they were introduced already in Section 3. Related to this it is clearly stated that the method is subjective. Nevertheless, if the method is to be applied by others, it would be informative to also document the reasoning behind the choices. For

example, why did you choose non-symmetric limits for the mean differences in Figures 2 and 3?

Specific comments

P1,L19: It is not surprising that the comparison results are significantly improved when the worst 15 sites (of 120 sites) are removed. It would be informative to quantify the improvement.

P2,L25-26: Is that not obvious? I mean it is stronger than "a tendency".

P2,L32-34: Are representativeness errors never to be referred to model errors? I interpret the definition of a representativeness error as that the only cause is the limited model resolution? If this is correct it can be stated explicitly, because I can also argue that the limited resolution of a model can be the cause of "model errors".

P3,L15: Explain/give examples, what is meant by "atmospheric environment" already here? Although it is clarified later when presenting Figure 9, my reading stopped here wondering what atmospheric environment could be out of many different things?

P4,L9-11: I think you should mention that the GPS time series used have passed some kind of quality check, because a very large break should have an impact on the overall standard deviation of the differences GPS – ERA Interim.

P6,L5: It cannot be taken for granted that the discrepancy is not due to GPS errors just because the formal errors do not increase. For example, a nearby installation of say a metallic structure may introduce significant multipath errors without affecting the formal errors.

P6,L30-33: An additional explanation could be that you only required 15 days of data for a specific month in order to be included. That would also affect the reduction of the standard deviation, unless it is very rare that so much data are missing from a month?

P7,L2-5: Can you compare this standard deviation of 0.81 kg/m^2 to what is obtained for stations located in the same area of the present study, in order to quantify the improvement obtained for the higher resolution model?

P7,L13-15: Perhaps the GPS sites that do not show an improvement using bi-linear interpolation are located close to one of the four grid points that is more representative compared to the others?

P11,L11: delete "strong" and just give the value? It should be up to the reader to decide what is a strong and a weak correlation.

P12,L23: delete "good", or state your definition for "good". Which parameter values do you typically see for Antarctica that is not seen globally?

P12,L25-26: This last sentence is not clear. It is the word "also" that raise questions. Because isn't that what you have done in the study? And what is meant by "other observation types"

Fig. 3: The figure caption refers to Figure 2. Are really the black dashed lines in Figure 3 of order 5 to 9?

Technical Corrections

P1,L7: IWV is not defined

P1,L17: don't $\Rightarrow$ do not (style ? + a couple of additional ones in the manuscript)

P1,L18: topography and coast-lines, strong $\Rightarrow$ topography, coast-lines, and a strong

P3,L7: delete "here"

P3,L24: sites is $\Rightarrow$ sites are

P6,L20: past study and led the $\Rightarrow$ past studies and led to the

P6,L25: GPS errors $\Rightarrow$ GPS formal errors?

P8,L25: strongly varying $\Rightarrow$ (very) different (altitudes of specific sites do not vary?)

P8,L30: worst $\Rightarrow$ worse

P9,L21: excessing $\Rightarrow$ excessive ?

P10,L11: thought $\Rightarrow$ though (or although?)

P10,L32: realists $\Rightarrow$ realistic ?

P12,L24-25: such trend $\Rightarrow$ such as trend

All figures: the red dotted lines would benefit from having larger dots

Fig. 2: (a,b) Mean and (c,d) $\Rightarrow$ (a,c) Mean and (b,d)

Fig.2: The units specified in the caption for graphs (b) and (c) do not agree with the labels.

Fig. 2: It is difficult to see the difference between the blue and the black dashed lines. Black and green may be better? Or make one of them dash-dotted?

Fig. 5: It is hard to read "dav1"(?) in graphs (b) and (d), plot labels after symbols?