Review of manuscript amt-2016-378: “GPS-PWV jumps before intense rain events” by Luiz F. Sapucci and co-authors.

This manuscript presents an analysis of time series of PWV from a GPS station and of precipitation from a nearby X-band radar for a period of 56 days in Brazil. The characteristics of PWV variations before and after the peak in precipitation rate are described with wavelet analysis and more classical tools (spearman correlation coefficient and histograms). The discussion is focused on the rapid variation of PWV preceding the peak in precipitation (increase followed by decrease). Though this feature is quite well known, both for the Amazon and for other regions of the world, the main novelty I see from this work is the fact that the increasing phase in PWV can be built upon successive pulses as suggested by Fig. 7. This subtle modulation of PWV was detected in this study thanks to the high temporal resolution of the PWV time series (1-minute). The potential for predicting the occurrence of intense precipitation with the help of real-time GPS PWV data is also suggested but not demonstrated. The topic of this study is interesting and might contribute to a better understanding of the moist processes involved in deep convection over land in the tropics (though such physical interpretation is beyond the scope of this paper). However, in its present form, this manuscript cannot be accepted for publication. A major rework of the data analysis and presentation is required. Major and specific comments are given below.

A) Major comments

1) Wavelet analysis:

- The wavelet correlation method and results are not properly used and interpreted. Several times the authors interpret the period bands for which the wavelet correlation values are displayed as lag times (P2L29, P620, P6L29, P9L19). For example (P6L29): “an interrelation analysis can be performed with the WCC of GPS-PWV and precipitation time series to evaluate the correlation at different lags in time, scale by scale”. The results provided in figure 5 and 6 are correlation values for a range of time periods indicated in the x-axis and wrongly labelled “wavelet scale”. This kind of plot is typically given for a zero-lag (see e.g., Whitcher et al., 2000). Where are the wavelet correlations for different lags? The analysis could be complemented with wavelet cross-correlations as a function of lag for different scales such as in Whitcher et al., 2000. Proper use of wavelet cross-correlations for lead/lag analysis would probably not require using the classical Spearman correlation (Fig. 8).

- The significance of wavelet spectra (Fig. 3 and 4) and correlations (Fig. 5 and 6) is not addressed, though mention to “significant” frequencies is made in several places in the text. Most tools used for wavelet analysis provide also results of significance tests or confidence intervals. These should be used to highlight objectively the significant frequencies.

- It is not said what the error bars in Fig. 5 and 6 represent.

- The description of wavelet analysis methodology in section 3 is not sound. Only fragmented information is given on the used methods, without justification of the choices (e.g. of mother wavelets, continuous vs. discrete wavelet transforms). On the other hand very highly-technical information is given (e.g. use of non-decimated discrete wavelet transform, use of pyramidal algorithm…) which won’t be explicit to the general readers of AMT journal. So I recommend that this section explains a bit more about the general principles and choices that were made, and the implication of the choices that are made. For example, one might question about the robustness of the results regarding the choice of mother wavelets. For the more technical aspects, refer to the proper literature where appropriate.

2) Analysis of data and interpretation of results

The GPS data processing options should be more detailed (see specific comments below). Some fundamental aspects are missing, such as the constraints of the stochastic model of the ZTD parameters. They impact directly the magnitude of variability in the retrieved PWV time series.

The dimensions of the area around the GPS station used to study the PWV-precipitation relationship is said to be a "key factor" (P5L5). In consequence, some preliminary tests should be presented which give insight into the sensitivity of results to this parameter. It is not clear how the representativeness (P5L11) is estimated.

Fig. 7 shows one case of heavy precipitation where PWV shows a strong peak achieved in several steps (so-called "pulses" in the text and inconsistently labelled "jumps" in the Figure). This kind of analysis should be made for other cases to establish if the pulses are a robust feature of the heavy precipitation events. It is not clear if the authors associate the 32-64 min periodicities detected by the WCC with these pulses? It would also be interesting to add a composite of time series to highlight the most prominent features of the PWV jumps.

The authors refer extensively to Adams et al. (2013) regarding a maximum of PWV 1-h before the maximum of precipitation. Inspecting this reference carefully reveals that on average the peak in PWV is rather coincident with the peak in precipitation (Fig. 2 in Adams et al., 2013) or slightly ahead (Fig. 3 and 4), in accordance with other results (e.g. Holloway and Neelin, 2010). The case illustrated in Fig. 7 shows indeed a lead time of 1 hour but about 50% of the cases show lead times between 0 and 30 min (Fig. 8). So the 1-h lead time should not be considered as a general rule.

In section 4.1, show an example of correlation function to help explaining what is meant by the "positive and negative correlations" (in fact maximum positive/minimum negative correlations) and specify the time window of analysis (+/- 1 hour). Again, the histograms in Fig. 8 suggest that the case in Fig. 7 might be a specific case because the minimum correlation is probably reached for a lag time around 0. Illustrating other cases and providing composite plots would help to better catch the general situation. Moreover, Fig. 7 is a case where the pulses preceding the peak in PWV are outside of the correlation window. The link between the pulses and the peak in PWV and in precipitation should be investigated in more detail as this is to my opinion the real innovation of this work.

In section 4.2, the PWV derivatives are computed in a 1-hour window preceding the peak in precipitation. This window is too small, namely for the case of Fig. 7, as it does not include the part of time series with the pulses. Extend the time window to analyse this feature in more detail and also to not give too much weight to the decreasing phase after the peak in the statistics computed in Table 2. Some sensitivity tests should be made to choose the best window size.

I suggest to add also a figure with the time series of PWV derivatives for all 6 cases of the upper tercile (e.g. superposed on one graph showing also the average PWV derivative and precipitation).

The sentence P2L29 announces a "study of correlation and lags with rainfall events to form a conceptual model with predictive capacity that can hence be used in developing a nowcasting tool for strong precipitation events". However, section 4.3 doesn't provide any proof of concept of this model. I suggest a short hindcast study is performed, based on the data from this campaign, to evaluate the skill of the proposed detection method.

Some interpretation is made about the underlying physical processes leading to the observed PWV jumps:

P8L25-29: "The presence of low-level water vapor convergence can be attributed to mechanisms such as gravity wave forcing (Raymond 1987) or other larger-scale forcing mechanisms. The increased water vapor convergence (i.e., positive PWV derivatives) may also simply be a reflection of the unstable surface parcels accelerating upwards, thereby vertically advecting a larger surface specific humidity to higher levels in the atmosphere without necessarily any larger-scale dynamical forcing"

P10L31: "This increase in the stronger negative derivative frequency before more intense events (upper and middle terciles) is associated with the conversion process from water vapor to liquid water…"

P11L5: 'the processes responsible for the maintenance of precipitable water suspended in the atmosphere are very complex and highly nonlinear"

P11L8: "The increase in moisture convergence appears to be due to an increase in the frequency of convergence pulses."

These assertions relate either to very general atmospheric processes or are highly speculative as recognized by the authors, P8L29, "Given the limitations of the observations, our interpretation of the physical mechanisms responsible for the jumps remains speculative."

I suggest to remove sentences which cannot be supported by other studies of similar phenomena or provide the necessary proofs, e.g. by analysing additional data.

3) Inappropriate and vague terminology

The PWV jumps which are the main topic of this work (as advertised in the title) are defined inconsistently in several places in the manuscript: P2L28 ("sharp increase in PWV"), P8L22 (a "pattern" of maximal PWV), P9L14 ("oscillations"), and Fig. 7 (pulses defined P8L25 are labelled "jumps"). Please be consistent.

Precipitation intensity is measured as the 95th percentile of all precipitation data in the area (P5L22). However, later, precipitation fractions are referred to as precipitation intensity (Table 1 and 2). Intensity terciles are referred to in section 4.2 and 5. Please be consistent.

It is not clear if the precipitation fractions are computed from the 95th percentile of precipitation or from all data and grid points?

Expressions such as "expressive", "expressive changes", "evident", "more evident" are not adequate. Please use conventional scientific terminology.

What is a positive oscillation? (P8L18, P12L3…) or a positive increase?

B) Specific comments

P4L7: specify which "possible noise sources" are taken into account and how.

I understand that you used the regional model for Tm of Sapucci 2014. So it is not the Bevis 1992 relationship that is used (P4L13). Please correct.

This Tm model requires RH. It is not said what RH data are used to compute Tm.

P4L17: “3.1% of data are missing in 2 days”. This represents about 1.7 days for the total period of 56 days. I suggest that you completely remove these 2 days from the analysis because the spline interpolation won’t give correct results.

It is important to specify the tropospheric model used in the GPS data processing: is it a random walk? Are both ZWD and gradients estimated? What are the constraints of temporal evolution of these two parameters? Did you make some tests with different constraints? (the analysis of 6-min time differences in section 4.2 might be strongly impacted if the constraints parameters are too small or too big).

P6L8-10 and in many other places, results from Adams et al., 2013, should be rather described in the introduction.

P7L17: what do you mean by “amplification”?

P7L31: why is quasi-symmetry (“a good property for the mother wavelet”) important to this study?

P8L19:20: “maximum peak” replace with “maximum”

P10L12: “the lowest threshold” complete with “among the 3 that were tested”.

P10L16: what are rainfall events with periods without precipitation?

P11L11-19: description of results from Fig. 9 should go to section 4.2

P11L21-25: results from past studies about nowcasting should go to the introduction. Note that these references can probably be updated.