

## Review of the manuscript

“Interpreting the time variability of world-wide GPS and GOME/SCIAMACHY integrated water vapour retrievals, using reanalyses as auxiliary tools” by Roeland Van Malderen, Eric Pottiaux, Gintautas Stankunavicius, Steffen Beirle, Thomas Wagner, Hugues Brenot and Carine Bruyninx

The manuscript presents the results of a study focussing on the variability of integrated water vapour across the globe as provided by two different datasets (sub-daily GPS and monthly-mean GOMESCIA estimates).

I find the topic important and the results really quite interesting, and I liked the summary Figure 10. My main concern is about the way the scientific questions are addressed (too vague), the motivation behind the approaches proposed (not discussed), a lack of precision and explanations in numerous parts of the manuscript as well as too many speculations. As a result, this is sometimes confusing, so that I recommend at least a major review.

We appreciate that you find our topic important and the results interesting and we thank you for your exhaustive feedback! We agree with you and the third reviewer that the aim of the paper should be more clearly discussed in the introduction of the paper. It was provided only fragmentary at the beginning of the different sections. Also the motivation for the multiple linear regression approach (as opposed to using regional climate models which are validated by GPS retrievals, see our forthcoming study Berckmans et al. ACPD, 2018) to study the IWV time variability is mentioned at the beginning of the section, but not in the introduction.

We now clearly stated the two main research questions of the paper in the introduction:

“In this paper, we focus on answering the following questions.

1) How well are the spatial and temporal IWV variability represented by three different, independent, IWV datasets? As our primary dataset used here is a ground-based GPS IWV dataset covering a world-wide sample of 118 sites, we will assess the spatial variability only between those sites and the global IWV datasets will be sampled to the site locations. In this work, a first characterization of the spatial IWV variability is given by considering the geographical distribution of the IWV frequency distributions, an extension of the work by Foster et al. (2006). To assess the temporal IWV variability of the three datasets, we consider different time scales here: from the seasonal cycle to short inter-annual variability, and to trends over periods of around 15 years.

2) Can the spatial and temporal (inter-annual and trends) IWV variability be explained by changes of local meteorological variables (like e.g. surface temperature and pressure) and/or by low-frequency variability in atmosphere and ocean (on both global and regional scales), and if so, how? There are already a number of studies mentioning IWV patterns that result from interactions between the atmospheric circulation and the land and ocean surfaces (see e.g. Trenberth et al., 2005, Wagner et al., 2006, Shi et al., 2018, Wang et al., 2018). These studies are based on correlation studies between IWV and one such a teleconnection index (ENSO: all mentioned studies, Pacific Decadal Oscillation: see Shi et al., 2018). Here, we fit the monthly mean IWV time series by means of a stepwise multiple linear regression approach with proxies for the seasonal cycle and linear trend, and with local meteorological variables and teleconnection indices as explanatory variables. With this empirical approach, we aim at finding out which the most relevant variables are to explain the IWV variability for different regions, independent of the used IWV dataset. To our knowledge, it is the first time that such an analysis is done on the IWV time series of individual sites.”

For your general comment on a lack of precision and explanations in numerous parts of the manuscript, we went through the manuscript again and added explanations to the issues raised by you, but also identified by ourselves. About the high number of speculations: we know that especially Sect. 6.2.3 (with its summary graph Fig. 11) is speculative, because this is the first time that so many teleconnection indices are linked, all together, with the IWV in such an empirical approach, so we could not refer to existing literature.

Some examples of lack of precision:

1) the teleconnection indexes are presented with no logical link, for instance, you do not explain your choice, nor systematically provide the associated fluctuation time scales; as a result this is confusing;

The selection of the teleconnection indices is done by analyzing ourselves their correlation with IWV for different regions (see Sect. 6.2.1.). We made a reference to this subsection here.

The complete description of the explanatory variables is out of scope here, but we reorganized Sect. 2.4 and give examples how they differ (atmospheric vs. oceanic circulation, spatial scale of impact, time scale of their variability).

2) The section describing the GPS dataset is quite long but I could not find information on how the authors dealt with missing data;

As we consider the GPS dataset as our the primary IWV dataset here, and as this manuscript has been submitted to the ACP/AMT/ANGEOS Special Issue “Advanced Global Navigation Satellite Systems tropospheric products for monitoring severe weather events and climate (GNSS4SWEC)”, we think that a long description of the GPS IWV retrieval is needed here. We do not deal with the missing data: for the

analyses here (seasonal cycle, linear trend, multiple linear regression), they are not at all problematic, although they had an influence on the lower correlation coefficients for GPS when fitting harmonic functions to determine the phase and amplitude of the seasonal cycle. This has been added in the manuscript, as a response to a suggestion by reviewer 3. In the manuscript, after mentioning the percentage of gaps in the data, we added that we did not specifically deal with missing data here.

3) in the regression analysis, I am not sure surface pressure and atmospheric temperature are taken at the site location or over the region to which the site belongs;

The surface pressure and surface temperature are calculated from the respective values from the four grid points surrounding the site by horizontal interpolation, weighted with the inverse distance to the site. To account for the height difference between the site location and the ERA-Interim surface grid points, we assume a standard temperature lapse rate of  $-6.5 \text{ K km}^{-1}$  (typical for wet adiabatic conditions) for the surface temperature altitude correction, and the hydrostatic and ideal gas equations to adjust the surface pressure. A reference to this text, in Sect. 2.1 and Sect. 2.3.1 has been made when the meteorological parameters are first mentioned as candidate explanatory variables in the multiple linear regression.

4) sentences such as “for (sub)tropical sites and sites in East Asia two distinct lognormal distributions are needed, probably related to the monsoon and ENSO” why ENSO?

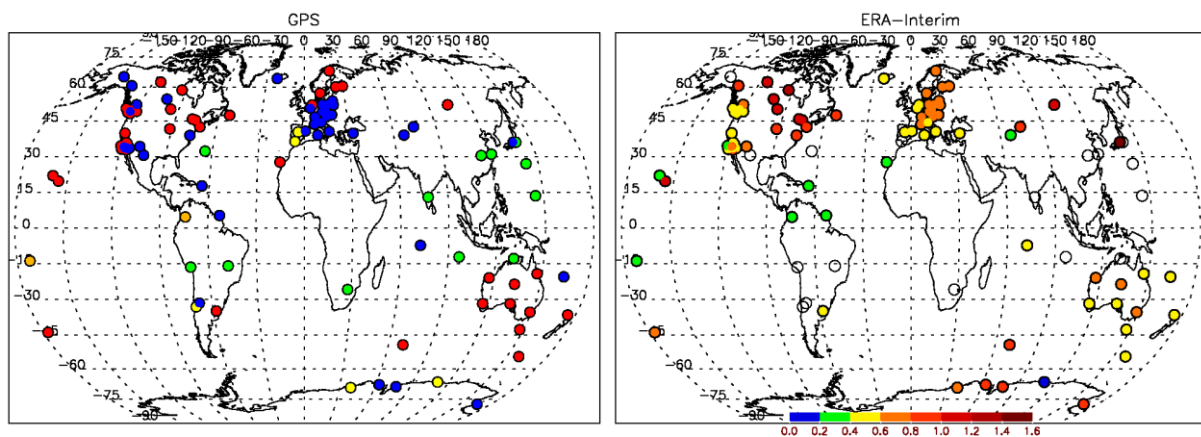
We added the reference to Foster et al. (2006) here, in which they showed the frequency distributions of some of those sites for different years. The authors could determine links between the shape of the frequency distribution of that year and the fact that the year was an El Niño, La Niña, or “normal” year. We also investigated this for our sample of sites and came to similar results. However, because of the interest of space, we decided not to include this in our manuscript.

5) in Figure 4, there are more tones of colours than indicated in the figure caption (e.g. pale green versus darker green, same for yellow to green).

This is done by purpose. The colour scale in the figure is continuous, while the colour bar is discrete. Here below, we add the same figure, but now with the colours used in the dots corresponding one to one to the colours of the colour bar. To our opinion, with this approach, we cannot represent enough the subtle gradients in GSD within a geographical region. Therefore, we propose our approach and added the following explanation to the figure caption: “Please note that the colour bar is a discrete indication of the colouring for the specified ranges. The colouring of the dots is done by a continuous scale, to better highlight the subtle GSD differences within a region”.

If you disagree with us, we can of course replace the figure, if requested.

The same representation by a continuous colour scale has been adopted in the (old) figures 7, 9, 10.



6) “Comparing our mean IWV trends with the 0.26 mm decade<sup>-1</sup> GPS IWV trend quoted by Wang et al. (2016a), we found slightly lower rates of 0.19, 0.08 and 0.11 mm decade<sup>-1</sup>”, 0.10 is NOT slightly lower than 0.26!

We removed “slightly”.

It may be a good idea to separate the manuscript in at least two parts, so that it may be easier to present proper presentations and discussions of the results (which are very interesting and numerous). Including more material in appendix may help too.

We thank you for this suggestion, but we are not in favour of splitting the manuscript in two parts. First of all, it is not a short paper, but also not excessively long. Secondly, the paper deals with the analysis of the

properties and spatial/time variability of the IWV measured at GPS sites, starting from its most simple statistical representation (the frequency distribution), over the seasonal cycle and linear trends to the inter-annual variability. Therefore, we consider this study as whole and prefer not to split it. We also do not see a solution how to do it. If you have some ideas about it, we would be happy to receive them.

Also, I think it would be excellent to take advantage of the global dataset (GOMESCIA) to add maps covering the whole globe rather than only presenting maps with results shown at GPS sites (data are sparse in large areas of the world and very clustered in others).

Here too, we do not completely agree with you. This paper is submitted to the ACP/AMT/ANGEOS Special Issue “Advanced Global Navigation Satellite Systems tropospheric products for monitoring severe weather events and climate (GNSS4SWEC)”, so the focus of this study are the GPS IWV retrievals at the sites. Therefore, we only considered the GOMESCIA and ERA-Interim IWV time series sampled at those sites. If we would show global maps based on GOMESCIA and ERA-Interim, and plot the GPS results on top of it, this would give the impression that GPS is used as a validation dataset for the other two datasets, which is exactly the opposite of our aim here. Moreover, some of the maps for ERA-Interim and GPS are already available in Parracho et al., ACP, 2018 and a similar study is under development for GOMESCIA, for which the first results have been presented at the EGU 2018: <https://meetingorganizer.copernicus.org/EGU2018/EGU2018-9217.pdf>. Also the second reviewer agreed that such an analysis is outside the scope of this paper. We nevertheless included the suggestion in the outlook.

On another subject, while I can understand your choice for a regression method where you test very numerous potentially explaining variables, I think that you should discuss more how your results compare or not with other studies and mechanism (I understand that it is difficult because there is a lot in the manuscript).

If the same explanatory variables were consistently retained for sites belonging to different regions, we tried to explain or discuss the relevant mechanisms behind this correlation with IWV. Unfortunately, as written in the introduction of our manuscript, at present, there are no other studies performing a multiple linear regression for IWV with those explanatory variables. The few that we mentioned performed a correlation analysis between one explanatory variable and IWV. Most of the past analyses try to link the explanatory variables with surface temperatures and precipitation, but not with IWV. We know that therefore, some of the arguments might sound hypothetical or too vague, but there is simply not much literature up to now.

Also, I did not understand your motivation for including the linear trend as an explanatory variable while it was already taken into account in Eqn (1) and finally did not explain much more.

We apologize for being not clearer on this point, also the second reviewer addressed the same issue. **All** the variables (describing the seasonal cycle, the linear trend and the explanatory variables) appearing in Eq. (1) are included only in the multiple linear regression if they significantly contribute to the regression coefficient. So, in most of the sites, the linear trend is not included in the final multiple linear regression. We hope that this has been clarified in the manuscript now.

Below I present some specific comments on the first part of the manuscript. This is not exhaustive, but I hope they can help the review process and help you to revise the following sections and conclusions accordingly for a second round of review.

#### Specific comments

Abstract: you need to be more precise about the time periods and time scales of analysis.

Page 1, line 14, “IWV variability”: please precise at which time scale and over which period, and the IWV sampling time step.

We added the period and for the different analyses we did, we mentioned the IWV sampling time step.

Page 1, line 18, “on average”: this is too vague, on average over what?

Averaged over all stations, this is added to the text.

Page 1, line 20-21, “the seasonal behaviour and the long-term variability are fitted together”: this is not exactly what I understood. Rather, you aim at reconstructing the time series of monthly-mean IWV from the mean annual cycle, linear trend and explanatory variables.

We changed the text to “Finally, we reconstruct the monthly mean IWV time series by means of a stepwise multiple linear regression from the mean annual cycle, the linear trend, and a selection of regionally dependent candidate explanatory variables.”

Page 1, line 25, “long term trend”: please precise, i.e. linear trend over the period [year1,year2]

Done.

Page 1, line 26: variableS

Changed.

## Introduction

In the last paragraph of the introduction, you present the work presented but I could not really find a clear presentation of the question(s) you want to address. In my opinion, an analysis of a new type is not a good enough motivation per se.

We provided a detailed description of the research questions and how we handled them in the manuscript (see above).

Page 2, line 4, 'on local scale': are you referring to mesoscale here? I would rather say that "At all scales" rather than local scale because precipitation for instance is not simply related to "local water vapour" alone; it typically involves larger, e.g. synoptic atmospheric circulations as well. In addition to the diabatic processes you mention, I would add radiative processes. I would remove "Of course".

Done.

Page 2, line 7, 1st sentence: a reference is needed there.

Done, we referred to Kämpfer, 2012.

Page 2, line 26, "have the potential to be used for climate change analysis, which is the subject of this paper": I do not agree about this statement, the results of this study are more focused on interannual variability (and trend over the 20-year-long period). The present study provides very valuable information about IWV variability in space and time. In time, it goes from the annual cycle to short inter-annual variability to trends over periods of 20 years or less. In my opinion, 20 years is too short to provide robust information on climate change. It is less than the time interval typically used to compute climate mean (30 years, e.g. see <http://www.metlink.org/climate/depthclimate-met-office/>).

We agree with you that for climatological means, time periods of 30 years are used (also at our meteorological institute). So, we are not posing that with our study, we can actually observe the climate change. But, as we know that the surface is warming, with this study, we would like to investigate the effect of e.g. this warming on the water vapour variability, at the time scales mentioned by you. We changed this statement in the text to. "In such a study (Van Malderen et al., 2014), we compared five different techniques and concluded that the techniques used here (GPS and GOMESCIA) are promising to study the inter-annual IWV variability within the context of a warming climate, which is one of the aims of this paper."

Page 2, line 31: add 'generally' before 'occur' as this is to my knowledge not strictly true for all geographic locations.

Done.

Page 3, line 2: remove 'can' as you precisely provides number who illustrate it.

Done.

Page 3, line 4: remove "of course" and replace "can be" by "are" as existing studies allow you to be more affirmative.

Done.

Page 3, line 6: I am not sure of what you mean by "autocorrelation" here.

Page 3, line 6, about ENSO: it seems to me that, more precisely, the relatively large magnitude of the signal induced by ENSO events at inter-annual scale affects trends computed on periods of 10-20 years. It would be good to reformulated a bit the sentence to be more informative.

We agree with you that this sentence does not add significantly new information with respect to the previous one. We therefore decided to remove it.

Page 3, lines 16-17: I think it is "on one hand", not "at one hand". More generally would be useful to check English throughout the manuscript. Avoid expressions such as "not surprisingly" or "of course" when you do not provide explanation nor references.

We went through the entire manuscript again and tried to improve the English and avoid using too suggestive language.

Page 4, lines 7-14, about homogenization: as the dataset in use here is not homogenized, I think this paragraph is unnecessary long. You could mention in the conclusion "for extension/improvement of this study, the use of a future homogenized dataset (as described in Van Malderen et al. (2017))".

We shortened this paragraph, but we still think it is necessary to point out that inhomogeneities can have a large impact on the calculated trends. We also made reference to this activity in our "Conclusions and outlook" section.

Page 4, presentation of GPS data processing: you may consider moving part of it in supplementary material.

As we already mentioned, as this manuscript has been submitted to a Special Issue with focus on GPS, we feel obliged to spend quite some text on the data processing, so that interested readers could compare this study with other published studies based on the same IGS repro 1 datasets (Wang et al., 2016, Parracho et al., 2018). Moreover, in this section, we also describe how the surface temperature and surface pressure time series (used in Sect. 6) were sampled and corrected to the GPS site locations and altitudes.

Page 6, about ERA5: this is not used, so the whole sentence is useless. It could be used in the future only, so potentially, you could mention it in the perspectives.

We removed this sentence and made a reference to ERA5 in the "Conclusions and outlook" section.

Page 6, lines 16-22: I am wondering why you mention this with so much details without linking this to your study.

We changed this paragraph to “The homogeneity of the extracted ERA-Interim IWV time series has been questioned recently by Schröder et al. (2016) and Ning et al. (2016) due to changes in the observing systems or changes of the input to assimilation schemes.”

Page 6, line 32: it seems to me that prior to Chen and Liu (2016), other studies such as by Bock et al. already extensively evaluated ECMWF and NCEP products.

Yes, but we only quoted the most recent one here.

Page 7, presentation of teleconnection indices: these indices are presented without much logic, you must re-write this section in a way that motivates your choice, and explain more over which time scales/regions they are relevant (and add somewhere the precise coordinates of the regions presented in the last figure).

The selection of the teleconnection indices is done by analyzing ourselves their correlation with IWV for different regions (see Sect. 6.2.1.). We made a reference to this subsection here.

The complete description of the explanatory variables is out of scope here, but we give examples how they differ (atmospheric vs. oceanic circulation, spatial scale of impact, time scale of their variability).

Page 8, section 3: 1) how did you deal with missing GPS data? 2) It is not well suited to use the word ‘bias’ as you do not have a reference dataset here.

We added “we also computed the statistical parameters mentioned here below only for the months for which the GPS monthly mean IWV dataset has actually values”

Strictly statistically speaking, you are right about the use of the word bias. But, this is a commonly used term when comparing datasets for a variable (including IWV, see <http://www.meteo.be/IWVintercomp>), without a clear reference dataset. For humidity/water vapour measurements, we speak about a dry bias (negative mean difference) and wet bias (positive mean difference) of one technique to another (not necessarily a reference).

Page 8, “We exclude the GOMESCIA dataset here, as only monthly means are available, which might be problematic to compute significant frequency distributions”: I do not understand what you mean. It seems you are mixing statistical robustness and time scale issues.

With only monthly means, you have at most 186 data points in your frequency distribution, which is not enough to be fitted well by lognormal or Gaussian statistical frequency distributions. We tried it, for the IWV monthly mean time series of the 3 datasets. You need either longer time series or capture the intra-monthly IWV variability (as we did with the 6h time sampling) to have observational frequency distributions that could be fitted well. So, the exclusion criterion is purely technical (or statistical robustness) and the motivation is not to study the intra-monthly IWV variability in particular. We clarified this in the text: “We exclude the GOMESCIA dataset here, as only monthly means are available, which puts constraints on the statistical robustness of the observational frequency distribution and its fit by a (log)normal distribution”.

Pages 8, 9, 10 and Figure 4: there are more tones of colours than indicated in the figure caption (e.g. pale green versus darker green, same for yellow to green). Please clarify.

See our answer on your general comment 5) on Figure 4.

Also, I would like to see the full maps obtained with ERA-I, NCEP as well as monthly-mean GOMESCIA IWV values. This would allow assessing the representativeness of the results obtained at the sites, and provide a clearer picture than emerging now at the end of this section.

We again refer to our response to your general comment about constructing those full maps based on the gridded datasets like GOMESCIA, ERA-Interim and NCEP/NCAR. We agree that this approach will provide an overall clearer picture, but we also agree with the second reviewer that this is out of the scope of this manuscript and the Special Issue it is submitted to. Moreover, this is also the subject of forthcoming work. We added this to the “Conclusions and outlook” section.

It is also very difficult to see the results over Europe because circles are overlapping each other.

Yes, we know. What do you suggest? Adding a snapshot over Europe for all figures to the Supplementary material?

Page 11, Figures 5 and 6: A discussion of the geographical patterns shown in Figure 5 is missing.

We added a small discussion on the geographical patterns shown: “This graph shows that the IWV seasonal cycle peaks in the summer months at both hemispheres. The amplitude of the seasonal cycle is largest for the Asian Northern Hemisphere sites, where there is a distinct difference between a dry and wet season. In Northern America, the central and eastern sites have larger amplitudes than the western coast sites, which is related to the fact that for most of these latter sites, the rainy season is in the winter, when the temperatures are low, while the summer are dry but warmer. In Europe, the Mediterranean sites have smaller amplitudes and the IWV seasonal cycle peaks one month later compared to the rest of the continent.”

We do not want to be too detailed here, as we focus here on how well the different datasets compare at representing the seasonal cycle at our sample of stations (see also next comment).

In Figure 6, I am wondering about how this graph was made: did you consider all the stations? I guess the uneven location of the stations is playing a large role in the shape of these histograms. I would like to see the same graph with all the global GOMESCIA dataset.

Yes, we considered all the stations. We specified this in the text. The uneven location of the stations is indeed playing a large role in the shape of the histograms of the amplitude and phase. However, the precise shape of those histograms does not matter here, as we focus on how well the different datasets compare in representing the seasonal cycle at our sample of stations (we refer here to the first scientific question that is addressed by this study), and are not interested in the global frequency distribution of the seasonal cycle phase and amplitude an sich. Therefore, adding the same graph making use of the global GOMESCIA (and ERA-Interim) datasets is not relevant here.

Page 11, section 6: Given the content of this section, I suggest that you modify the title (especially remove "long-term"). You could rather emphasize the idea of the trend over [year 1, year 2] and inter-annual variability in this time window. You could probably shorten your discussion of the statistics, and it would be clearer if you could add a few words about decadal and multi-decadal variability.

The title has been changed. Given the fact that only 1.5 decade is available for the datasets used here, we do not see which meaningful discussion could be added here about decadal and multi-decadal variability. Please share with us your ideas about this.

Page 11, lines 2-3: "As we have only 15 years available for most of the stations, our time series is too short to draw firm conclusions on the presence or magnitude of a trend." Then, just below you compute trends. The way it is written is very confusing. I suspect you mean an "expected climatic trend" in the first instance.

The point that we want to make here is that we will not discuss the magnitude or the presence of a trend (they cannot be statistically significant according to the formalism of Weatherhead), but we want to focus on the differences of the trends between the different datasets and on the interpretation of the inter-annual variability. So, we will not make firm statements about "IWV is increasing at this region", but merely "at this region, the three different datasets compare well and show an IWV increase". We tried to make this clearer in the manuscript.

Figure 8: It is important to add the residual obtained when removing the mean annual cycle and the linear trend to the time series, in order to be able compare it with its magnitude with the magnitude of the residual that you show.

The second reviewer also did not understand very well what was in this figure. For the two stations in this figure, the linear trend was not included in the multiple linear regression, as it did not contribute significantly enough to the correlation coefficient. The mean annual cycle was included for both sites. As requested by the second reviewer, we added some intermediate steps for the multiple linear regression procedure for those 2 sites as extra figures in the Supplementary material. We hope that this helps in the interpretation of this figure. We also added some extra clarifications in the figure caption.

I have more specific comments to come on a revised version of the manuscript where you would have taken into account my general and specific comments, as I think this manuscript should be published (perhaps as a two-part paper).