In this paper an analysis of integrated water vapour (IWV) data from three different sources (ground based GPS data, GOME/SCIAMACHY/GOME-2 satellite data and ERA Interim model data) is presented. Beside intercomparisons of the three data sets major topics of the paper are the analysis of time series and the association of variabilities in IWV with 'explanatory variables', e.g. physical / dynamical processes described by teleconnection indices.

This is an interesting approach which is well presented, but some aspects need further clarifications. Especially, the method of stepwise linear regression needs to be explained in more detail. Which terms of eq. (1) are fitted in the individual steps (and in which sequence), which corrections are applied to the data sets used in each step? Especially, it should be clarified if/how e.g. seasonal variability and trends of the 'explanatory variables' is considered. Have these data been de-seasonalised before the fit? If not, how are correlations with the linear trend and seasonal terms in eq. (1) handled? It seems that e.g. linear trends are sometimes not fitted at all because they are covered by explanatory variables variations - how is this decided? Does the sequence of fitted variables matter?
Thank you very much for your positive feedback. We elaborated more on the description of the method and the practical issues that you mentioned here in the manuscript. You will also find our answers to these questions here below, when they were repeated in your specific comments!

Maybe it would help to add e.g. for one example a plot of the different fit steps with the different time series used in each step (not necessarily for all explanatory variables, but for the major ones); this could possibly be an extension of Fig. 8. Another aspect is the selection of relevant explanatory variables for the fit. I would assume that many of the variables (e.g. temperature, pressure, precipitation) are strongly correlated - it needs to be explained in more detail how this is considered in the fitting procedure as well as in the interpretation of the results.
We did not change Fig. 8 (now Fig. 9), but instead added two extension figures of Fig. 8 in the Supplementary Material: in one figure (S2), we present the time series of the 5 most important explanatory variables that have been included in the multiple linear regression fit of both the good and bad example. In a subsequent figure S3, we then show how the fit evolves when adding each of those explanatory variables, step by step, to the multiple linear regression.
The other part of your comment (about the correlation of the explanatory variables) is treated in our response to your specific comment 7 and is also elucidated in the manuscript.

Specific Comments:
1. p. 6, l. 28–29:
'The database has been enhanced with many observations not available in real time for operational use.'
What is meant with this? Which database/observations do you refer to? Did you use/produce a dedicated version of NCEP/NCAR data? Please clarify.
It is meant here that the data assimilation database is enhanced with many observations that have not been used in the operational version of January 1995 that was taken to perform the reanalysis. We adapted this sentence to "The assimilation database has been enhanced with many observations not available in real time for operational use at that time."

2. p. 9, 1st paragraph:
As I understand, the difference between lognormal and reversed lognormal distribution is that there is either a (non-zero) lower or upper limit for the IWV. Is there a physical reason why a lognormal distribution occurs for subtropical and temperate climate and a reversed lognormal form in tropical oceanic environments?
This is indeed a very interesting question. Foster et al. (2006) also reflected on this issue. The absolute lower limit (threshold parameter t=0) is simply a completely dehydrated atmosphere, while the most obvious interpretation for the upper limit in the reversed lognormal situation is that it corresponds to some maximum carrying capacity of the atmosphere: complete saturation of the atmospheric column up to some characteristic level for example. So, reverse lognormal distributions appear to be connected with areas that experience almost total saturation, typically oceanic equatorial zones, while for subtropical and temperate climate zones, the lower bound is more decisive. A possible interpretation of a simple lognormal form for a IWV time series is that the source region(s) for the observed moisture is(are) effectively mixed over the considered time period, as this ensures that one mean and variance can be used to describe the region(s). This explanation is shortly added to the text.

3. p. 9, l. 10ff. and Fig. 3:
Since the 'shouldered' lognormal distribution is a new category I suggest to add also an example plot for this in Fig. 3.
Fig 3.c already shows an example of this category. This is also clarified in the text.

4. p. 11, l. 5–6:
'the IWV seasonal cycle for about 15 sites in the Northern Hemisphere peaks one month later in the GOMESCIA dataset with respect to the GPS and ERA-Interim datasets' Is there an explanation for this?
The GOMESCIA IWV peaks for these sites more frequently in August than in July, compared to GPS and ERA-Interim, but the difference between the mean July and August values are often very small at those sites. We refined this in the text. Part of the difference can be ascribed to the diurnal variation of the IWV and the difference between the fixed satellite overpass times for GOMESCIA and the 6h time sampling for the GPS and ERA-Interim. For about 10% of the sites, we find a difference of one month in the IWV peak between the monthly means calculated e.g. at 0h and 12h UTC for GPS and ERA-Interim. The clear sky observation bias for GOMESCIA might contribute as well.

5. p. 12, l. 7–8:
'We calculated linear trends as the slope of the linear regression line that was fitted (by minimising the least squares) through the monthly anomaly IWV time series.'
Please explain how these anomalies are calculated (e.g. by subtraction of harmonics or long-term monthly means).
We added the sentence "obtained by subtracting the long-term monthly means from the monthly averages"

6. p. 12, l. 16–18:
Is here an explanation for the differences in GPS trends from this study and Wang et al. (2016a)? If the same data set is used (as stated in section 2.1), why are there differences?
There are different explanations. First of all, we used the same raw data set (IGS network of 117 (Wang et al. 2016a)/118 sites with data extending from 1995/1996 until 2010/2011), but the ZTD processing might be different (IGS repro 1 in our case, not clear which processing has been used for the different periods by Wang et al. (2016a). The conversion from ZTD to IWV has been done in our study from the ERA-Interim surface pressure and weighted mean temperature at the 6h temporal grid. For constructing the 2-hourly IWV dataset in Wang et al. (2016a), Ps is derived from global, 3-hourly surface synoptic observations with temporal, vertical and horizontal adjustments, and Tm is calculated from NCEP/NCAR reanalysis with temporal, vertical and horizontal interpolations. And finally, the trends calculated in our study range from 1996-2010, while in Wang et al. (2016a), the time period 1995-2011 is considered. In our manuscript, we included "The raw GPS dataset used in their study and ours is identical, but the trend difference can be explained by different factors: a different data processing, the use of different meteorological data sources for the ZTD to IWV conversion (see the Appendix for e.g. an assessment of this factor on the trends), a different time resolution (and temporal interpolation in Wang et al., 2016a), and the use of a different time period for which the trends were calculated (see above)".

7. p. 14, 1st and 2nd paragraph:
Are seasonal variations considered in the checks for independency of explanatory variables? Where does the limit of 0.90 for the linear correlation coefficients come from (in the section on linear trends a correlation of R_2=0.66 is considered as large).
Yes, we calculated the correlation between the entire time series of the (global) explanatory variables (including the seasonality) like the teleconnection patterns. But additionally, for every site separately, we calculated the correlations between the local meteorological parameters (surface pressure, surface temperature, tropopause pressure, precipitation), and between these local meteorological parameters and the teleconnection patterns. The 0.90 limit for the linear correlation coefficient is chosen so high to exclude only strongly correlated explanatory variables at this stage. This has also been clarified in the manuscript. We are confident enough in our statistical t-test in the stepwise multiple linear regression technique not to retain an explanatory variable if, in the previous step, another explanatory variable with which it is strongly correlated, was already selected. See also immediately here below for more explanations.

In this context, and also for the regression tests used to determine the significance of the different explanatory variables: Have the data sets been deseasonalised before the comparisons? Is the preparation of variables for this test consistent with their later use in the fit (see also general comment)? Is the fact that 'Variables with a significance level lower than 5% are discarded' the reason that in the fits instead of 100-200 variables only 6-8 (see end of this page) are considered?
Please explain.
We did not make any corrections or did not deseasonalise the time series of the explanatory variables for the linear regression. The time step of the linear regression is one month, so we used the monthly mean time series for both the IWV and all explanatory variables. Some of the teleconnection patterns are provided as monthly means on the web pages mentioned in Table S2, for the other explanatory variables, we calculated monthly means based on daily values or on the 6h time grid (surface temperature, surface pressure, tropopause pressure).

8. p. 14, l. 17–19:
'a significant positive trend is still present in the residual time series (although the annual trend was not retained as a significant explanatory variable in the multiple linear regression).' This is unclear. If (as in eq. (1)) the linear trend is fitted, why is there a remaining trend in the residual?
The linear trend is considered as an explanatory variable as well in the stepwise multiple linear regression, and is included only if it significantly contributes to the regression coefficient (by means of a t-test, at the 95% significance level). We clarified this in the text.

9. p. 14, l. 23–25:
'part of the seasonal behaviour present in the time series still has to be explained by other variables, especially by the surface temperature and precipitation time series'
As mentioned in the general comments, it has to be clarified how seasonal variations are considered in each of the fitted time series and how this is in line with the seasonal terms in eq. (1).
The term(s) describing the seasonal behavior in Eq. (1) are treated as explanatory variable(s) as well, just like the linear term, and they are included only in the multiple linear regression if they significantly contribute to the regression coefficient. We hope this is now also clarified in the manuscript itself.

10. p. 17, l. 23–25:
'the NAO index is present in only one third of the sites as explanatory variable, although its relationship with precipitation is well established in Europe' If NAO index and precipitation are closely related, why can these be considered as independent explanatory variables?
As shown in the referenced paper, the NAO index and mean/extreme precipitation are spatially correlated, especially in summer and winter, and with different signs in different parts of Europe. This does not necessarily mean that their time series are also strongly correlated. For every European site we considered here, the linear correlation between the precipitation and the NAO index never exceeds 0.5 (in absolute terms).

11. p. 18, l. 26:
'we do not expect an effect of IWV on precipitation' Why not? Please explain.
We removed this sentence.

12. p. 19, l. 10–13:
'Moreover, whereas the majority of the sites have positive trends in their IWV time series, especially for the GPS and ERA-Interim datasets (see Sect. 6.1), the residual time series after applying the multiple linear regression show an equal amount of positive and negative trends (GPS and ERA-Interim) or even a higher amount of sites with a negative trend (GOMESCIA).'
Obviously, there are different ways to determine trends. As I understand, trends from section 6.1 originate from a simple linear regression (and probably a to be defined seasonal correction, see above). It seems that the multiple linear regression trends mentioned here are those remaining after subtracting the effects of explanatory variables without a linear trend fit (although a linear term is given in eq. (1)). This should be clarified.
As should be explained now here before and in the manuscript, the residual time series is obtained by subtracting from the original IWV monthly mean time series the multiple linear regression fit in eq. (1), which might contain the long-term monthly means, a linear function and explanatory variables, depending on their significant contribution to the regression coefficient. We also clarified this in the text. The main message that we want to give here is that some of the explanatory variables, besides the linear trend term, are able to explain the linear trends discussed in Sect. 6.1.

13. p. 19, l. 25–27:
'the linear trend sign of the explanatory variable's term (coefficient multiplied with its time series) is in agreement with the linear trend sign of the IWV time series of the same site'
It is not clear what is meant with 'trend sign of the explanatory variable's term' and how this is derived. Do you fit a linear trend to the explanatory variable's term? Please explain.
We give an example here. Suppose that an IWV time series shows a positive trend (as calculated in Sect. 6.1). For this time series, among others, e.g. the surface pressure was kept as an explanatory variable. This is time series $X_j(t)$ (with e.g. j=2) in Eq. (1). At this station, the surface pressure decreased over the time period considered. So, for this station, the linear trend in IWV could not be explained by the linear trend in the surface pressure, unless the coefficient B_j (with j=2) is negative. Let us assume it is indeed negative. So, for this example, the linear trend sign of the surface pressure term $B_jX_j(t)$ (with j=2) (positive) is in agreement with the linear trend sign of the IWV time series (positive). We are aware of the complexity of the formulation used here and we propose to change into: "As the time series of the local meteorological variables Xj(t) appear with coefficients Bj in the multiple linear regression equation (1), a positive trend in Xj(t) might be compensated by a negative coefficient Bj, so that the negative trend in BjXj(t) could still be linked to an IWV decrease at that site, for instance. As a matter of fact, more or less independent of the

dataset used, we found that, on average, for about 70% of the cases for which the mentioned (local) explanatory variables Xj(t) are present in the multiple linear regression of the sites, the (linear) trend sign of the explanatory variable's term BjXj(t) is in agreement with the (linear) trend sign of the IWV time series (see Sect. 6.1) of the same site."

14. p. 20, l. 25–27:
'the precise identification of the main contributor to the IWV trend is almost impossible'
This sounds rather pessimistic. In most cases there will not be one single contributor to the IWV trend. The analysis presented in this paper at least shows for certain regions the main contributors, and this is a very useful result which could possibly even be the basis for further investigations (see suggestion below).
Thanks for this more optimistic formulation. We changed it in the text as well.

15. p. 22, l. 14:
The 'meteorological station' is only introduced in the next paragraph, should be explained before.
Solved by using the more general term "data source".

16. p. 22, l. 14:
'the slope of the linear regression (with correlation coefficient 0.84) between the Ps and IWV biases between the different corresponding datasets for the 40 IGS stations is equal to the -0.34, confirming the acceptable data quality of the pressure observations at the retained stations.'
It is not fully clear to me what has been done here and how the derived numbers are to be interpreted. Do you refer here to the begin of the appendix 'a 1 hPa change in Ps gives an IWV change of 0.36 mm'? Table A1/A2 list three different Ps sources - to which do you refer here?
Yes, we refer here to the 'a 1 hPa change in Ps gives an IWV change of 0.36 mm'. As is made clearer in the text now, the correlation is made between "the individual values of the first column of b), not the mean".

17. Table A1:
(a) Please define "abs bias" and "abs trend". Probably these are the bias and trend of the absolute differences? Why are these absolute numbers relevant in relation to the interpretation of the multilinear fit?
We changed this to "In Table A1, we present the means – weighted by the number of observations for each station – of (from left to right) the IWV differences, the absolute value of the IWV differences, the standard deviation of the IWV differences, the linear correlation coefficients between the IWVs, the IWV trend differences, and finally the absolute IWV trend differences between two GPS IWV datasets that disagree only by one or more of these auxiliary meteorological parameters."
This appendix is not directly relevant for the interpretation of the multilinear fit, but should give an idea how sensitive the linear IWV trends calculated in Sect. 6.1 are on the used meteorological parameters needed to convert the GPS ZTD to IWV.
(b) What exactly is meant with case a)? Is this a comparison between the IGS and the ERA Interim IVW data?
Indeed! So, this is an addition to the Sect. 3 "Dataset comparison" for ERA-Interim and IGS. We added this reference to this section in the text.

(c) What is meant with 'the two different databases of the meteorological variable whose impact on the IWV is studied'? For example, I interpret case b) as a comparison between IWV results based on Ps from ERA Interim and Ps from synoptic stations (this should be the non-italic numbers). What are the italic numbers in this case?
You are right about your interpretation of the plain numbers for case b). We wrote in the manuscript, immediately after the cited text to answer your issue 17 (a): "The numbers in italic are more informative and denote these same statistical means of (trend) differences, but now for the meteorological parameter that differs between the two datasets (resp. Ps, Ts, and Tm for cases [b], [c], and [d])."

18. Figure 4:
Why do you only show Classification for GPS (a) and GSD for ERA Interim data (b)? I assume these plots should be available for both data sets.
These plots are similar for the other datasets. We nevertheless included them in the Supplementary Material.

19. Figure 8: It should be clarified that no linear trends are fitted, only proxies (if this is the case here).
We hope this is clear now.

Technical Corrections:

1. p. 2, l. 16:
differntial → differential
Corrected.
2. p. 2, l. 20:
An inventory many of → An inventory of many of
Corrected.
3. p. 7, l. 18:
Antarctic (AO) → Antarctic (AAO)
Corrected.
4. p. 11, l. 21:
statistical significant → statistically significant
Corrected.
5. p. 16, l. 30:
our sample our located → our sample are located
Corrected.
6. Caption Figure 3:
with its contribution lognormal distributions → with its contributing lognormal distributions
Corrected.

Suggestion:

The current work is limited to the geographical sampling of the GPS stations. I suggest to perform (outside the scope of this paper) a similar multiple linear regression analysis for global time series (e.g. from ERA Interim and/or GOMESCIA). This could help to identify reasons for IWV variations on a regional scale.

Thank you for this suggestion, we included it in the conclusions. As a matter of fact, such a study is under development and the first results have been presented at the EGU 2018: https://meetingorganizer.copernicus.org/EGU2018/EGU2018-9217.pdf