Author’s comment

Trends of atmospheric water vapour in Switzerland from ground-based radiometry, FTIR and GNSS data

Bernet et al., Atmos. Chem. Phys. Discuss., 2020, https://doi.org/10.5194/acp-2020-77

Dear referees,

we would like to thank you for the constructive and elaborated comments and suggestions to our manuscript. We have taken the remarks into account and hope that we have responded satisfactorily to your suggestions. The referee’s comments are given in blue italic typeface, our responses are given in black, and the corresponding changes in the manuscript in grey.

Kind regards,
Leonie Bernet (on behalf of all co-authors)

Contents

1 Author’s response to referee #1 2
- 1.1 General comments 2
- 1.2 Specific comments 5

2 Author’s response to referee #2 11
- 2.1 General comment 11
- 2.2 Minor comments 14

3 References 23

1 Author's response to referee #1

1.1 General comments

> *The paper is a detailed study of the IWV data (and trends) gathered by all available IWV measurement devices available in Switzerland, and extended with the most state-of-the-art reanalysis output available (ERA5 and MERRA-2). The use of a trend analysis model that account for possible biases (inhomogeneities) due to instrument changes is really a strong point of this study. At some places, the study does lack some interpretation of the findings (see my specific comments), and I identified three remaining caveats the study should deal with before publication:*

Thank you for your feedback, we have followed your suggestions to improve the manuscript.

> *In the analysis of equations (9), (10), and (11), it is assumed that the relative humidity is constant, as is quite often the case when assessing the impact of the Claussius-Clapeyron relationship. However, in the study of Wang et al. (2016), already referred to in the manuscript, but not in this section, it is shown that the relative humidity varies with temperature. As you might have relative humidity measurements available (in any case you might use the reanalyses output), it would be good to confront your approach with the one described in Wang et al. (2016). In the conclusions section (lines 505-509), you also highlight this weakness in your approach. I think that adding the variation of relative humidity with temperature as an extra piece of information to understand the variation of IWV with temperature in Switzerland should be considered.*

We thank the referee for this input and agree that analysing the variation in relative humidity can provide valuable information. We therefore added a figure showing RH trends from ERA5 (Fig. 15 in new manuscript). The text and interpretation were adapted accordingly.

Description of new Fig. 15:
ERA5 e_s is decreasing in winter, whereas ERA5 IWV winter trends are increasing. These conflicting results indicate that the assumption of constant relative humidity might not be valid in winter. This is confirmed by the ERA5 RH trends (Fig. 15), which are around zero for whole Switzerland in all seasons but slightly positive in winter. Even though these positive winter RH trends are not significantly different from zero, they raise the question whether it is justified to assume RH to be constant.

Conclusions:
Another reason for observed inconsistencies between temperature and IWV changes might be changes in relative humidity (RH). **Our temperature–IWV relation assumes that the relative humidity remains constant. However, we found positive RH trends in winter using lower tropospheric ERA5 data. Even though the RH trends are not significant, they might partly explain the disagreement between observed winter temperature and IWV changes. Wang et al. (2016) states that RH may not be constant because of limited moisture availability over land surfaces.** ~~However, this is not always the case, because of limited moisture availability over land surfaces (Wang et al. 2016)~~ Some studies found even a decrease of relative humidity with increasing temperature at midlatitudes (O'Gorman and Muller, 2010) or in the subtropics (Dessler et al., 2008). **Further analyses with additional data sets would be required to provide more insights into possible RH trends in Switzerland.**

> *Another major point I want to raise here, is the comparison of trends (especially in section 6.1) between GNSS sites that have a different length of time series. The length of the time series is only considered in the size of the markers in Fig. 8, but will interfere with other effects that have considered to compare the trends (e.g. the altitude of the station, in Fig. 9, to name only one). Due to the inter-annual variability of IWV, the starting/end date of the time period has an impact on the trend value. Therefore, I would urge for a detailed analysis of the impact of the length of the time series on the resulting IWV trends, by e.g. constructing an alternative Fig. 8 with trends calculated from a common time period and/or constructing a figure like Fig. 9, in which the trend is assessed against the starting year of the GNSS time series.*

The length of the time series can indeed affect the trend and its uncertainty in multiple ways.
1. The longer the time series is, the more information is available. However, the information content of the data is often overestimated because the data might not be truly independent. Any variability in the data which is not accounted for by the trend model can lead to autocorrelations in the fit residuals. These are, however, taken into account by the trend analysis tool used. Even if the length of our time series appears to be short, the impact of this shortness and the information loss via autocorrelations is included in the uncertainty estimates provided.
2. Also, systematic short-term variability can mimic trends which are not there in the true atmosphere. For example, an unaccounted annual cycle will give rise to an artificial trend if the length of the time series is different from an integer multiple of a full year. However, the trend model we use includes the annual cycle and three of its overtones. Thus, these effects should not cause any artefact.
3. Another possible cause of the dependence of the trend from the start and end date of the time series might be that the trend is not linear. Since our trend model assumes a linear trend, our resulting trends might indeed depend on the start and end date of the time series. We therefore indicate the length of the time series in Table 1 and show the different lengths by the marker size in Figure 8. To make these different trend lengths also visible in the altitude dependent Figure (Fig. 10 in new manuscript), we added trend length dependent marker sizes also in this figure. Further, we added GNSS trends to the reanalysis maps (Figures 11 and 12 in new manuscript), showing only GNSS trends with longest time series (18 and 19 years). Finally we would like to mention that Fig. 8 (Fig. 9 in new manuscript) shows for the majority of the Swiss GNSS stations positive trends of IWV in spite of different start and end times of the IWV series. This is also an important result of our study.

> *The section 4.1.1 is definitely not my favorite section. It leans on Fig. 5, in which the trend uncertainty bars should be added. This might be tricky, since, as far as I understand, those monthly trends were computed based on only one value a year, so based on about 20 points (20 years) only. Besides this small amount of points, the length of the time series is very short in climatological sense (where time periods of 30 years are the standard). As a matter of fact, there are some interesting studies that calculate the number of years that is needed to derive a statistically significant trend in IWV (e.g. Alshawaf, F., Zus, F., Balidakis, K., Deng, Z., Hoseini, M., Dick, G., and Wickert, J.: On the statistical significance of climatic trends estimated from GPS tropospheric time series. Journal of Geophysical Research: Atmospheres, 123.* `https://doi.org/10.1029/2018JD028703`*, 2018). So, the main question is: how significant are the trend value differences between the different months? The discussion in lines 309-316 also seems to indicate that the found "seasonal" trend differences (and the shape presented) is not a consistent feature.*

The uncertainties are indeed missing in Fig. 5 (Fig. 6 in new manuscript). We therefore added the maximum trend uncertainty to the figure. The error bar represents the month with largest uncertainty for each data set, even though some months have smaller uncertainties. Significant trends (usually with smaller uncertainties) are thus marked by filled dots. Please also note that the uncertainty of monthly means from TROWARA and GNSS has been adapted, because we assessed that the monthly uncertainty used is too small for a total uncertainty of GNSS or radiometer measurements. We therefore decided to add a systematic uncertainty based on results from Ning et al. (2016). The updated uncertainties result in slightly different trends, but the effect is small and the more realistic uncertainties justify the change. The manuscript has been changed as shown below, and the Figures have been adapted to the new trend values. We thank the referee for the constructive criticism, and we think that the reviewed uncertainty estimates and the changes in Fig. 5 (Fig. 6 in new manuscript) improved this section.

As monthly uncertainties σ_{mon}, we use for TROWARA and GNSS data

$$\sigma_{mon} = \sqrt{\sigma_{\bar{x}}^2 + \sigma_{sys}^2}, \quad (1)$$

where σ_{sys} is a systematic error and $\sigma_{\bar{x}}$ is the standard error of the monthly mean, given by

$$\sigma_{\bar{x}} = \sigma n^{-\frac{1}{2}}, \quad (2)$$

with the standard deviation of the monthly mean σ and the number of measurements per month n. The systematic error σ_{sys} is estimated to be 1 mm for TROWARA and $0.7\,\mathrm{mm}$ for GNSS data. These values are based on results from Ning et al. (2016a), who assessed IWV uncertainties from a radiometer and GNSS observations in Sweden.

Concerning the length of the time series, please note that the length of the time series is considered in the uncertainty of the trend estimate. A special characteristic of the trend tool we use is the consideration of the full data error covariance matrix instead of the mere error bars of the data. The trend uncertainty is estimated using generalized Gaussian error propagation in a sense that the dependence of the model parameters (trend, amplitudes of the components of the annual cycle) on the uncertainties of the ingoing data are calculated. The autocorrelated error components are estimated from the autocorrelation of the fit residuals. The necessary length of the time series depends on the included sources of variability. The more drivers of variability (e.g. overtones of annual cycle) are included in the trend model, the smaller the autocorrelated parts of the fit residual will be, and the better is the reliability of even a quite short time series. This is true as long as the number of data points is large enough to allow the fit of the related increasing number of model parameters, which is the case for our time series.

The significance of the trend differences between different months can be estimated as follows: The estimated error variances of the trends of both months under consideration are added to give the sum of the estimated variances. The square root of this is the estimated standard deviation of the difference. The trend difference can then be divided by this standard deviation. If this quotient is larger than 2, assuming Gaussian error statistics, the significance is approximately 5% (95% confidence limit). In our case, this test shows that only the difference between MERRA-2 trends in November and December is significantly different at 95% confidence interval. For the higher summer trends and the trend peak in October, we can be confident at 68% (1σ) that they differ from the trends of the other months (for TROWARA and reanalysis data). We therefore added a sentence to indicate this.

However, the differences between those monthly trends are significant only at 68% confidence level.

Beyond this, we would like to stress that we do not claim that our trend estimates have any validity outside the time period analysed. We do not endorse extrapolation. We restrict our analysis to descriptive statistics of the data available.

1.2 Specific comments

> *Page 1, lines 20-21: rephrase "it builds the link between temperature and precipitation" , building the link sounds awkward*

Modified.

[...] water vapour is involved in important tropospheric processes such as cloud formation~~, it builds the link between temperature~~ and precipitation [...]

> *Page 2, line 42: specify in which wavelength range the satellites that are restricted to oceans only operate.*

The text was adapted as follows.

Second, **visible and infrared satellite techniques are limited to clear-sky measurements. Further,** ~~many~~ satellite products from **passive mircowave sensors** are restricted to oceans only, because the well-known ocean surface emissivity makes retrievals generally easier over oceans than over land surfaces (Urban, 2013).

> *Page 2, line 47: drop "probably" in the sentence "Radiosondes probably provide the longest time series, ..."*

Modified.

> *Page 5, lines 147-148: it is a pity that you use the Bevis approximation to estimate Tm from Ts, while you could have used the ERA5 vertical profiles of temperature and specific humidity to calculate them. You might comment on the applicability of the Bevis approximation for Switzerland.*

We agree that alternatives to the Bevis approximation exist. However, we prefer not to use the ERA5 vertical profiles in order to keep the GNSS IWV data independent from ERA5. This is especially important because we compare the GNSS data with ERA5 data for validation.
Further, Alshawaf et al. (2017) showed that the use of reanalyses temperature and pressure data can lead to a bias in IWV compared to the use of surface measurements, especially in mountainous regions in Germany. We therefore follow their recommendation to use the Bevis approximation derived from surface temperature. We think that for the current study, the required additional effort by changing the retrieval would be larger than the benefit. Nevertheless, we will consider using a more advanced approach to determine the mean temperature at the GNSS stations in future.

> *Page 8, lines 237-238: explain what the impact of an antenna and receiver change might have on the data variability. I understand that these changes might cause a jump in the mean of a time series, but it is less clear to me how they might cause a higher/smaller data variability. Please comment on this.*

We agree that a receiver or antenna change is not directly linked to the data variability. The antenna was enhanced from GPS only to GPS and GLONASS, which might improve the tropospheric values. It is nevertheless questionable if this can effect the data variability. We therefore adapted the statement as followed.

The anomalies are less variable from 2007 to 2012, but it is not clear whether this is related to the antenna update in 2007.

> *Page 8, lines 239-240 and Fig. 2b: which differences are you referring here to? GNSS – ERA5 or ERA5 – GNSS. Please specify.*

As indicated in the caption of Fig. 2, we compute the relative difference with ERA5-GNSS/GNSS. We adapted the text for clarification.

Furthermore, the relative difference to ERA5 **((ERA5 - GNSS)**/**GNSS)** reveals a data jump [...]

> *Page 9, lines 264-and following: when comparing the IWV trends of TROWARA with the GNSS IWV trends, please mention if these trends are calculated for the same time periods, and if not, what the impact of the time period on the calculated trend is.*

The GNSS trends are estimated for a shorter time period than TROWARA and the reanalyis trends. We now mention this in the manuscript, referring to Table 1. As discussed in the general comment, the trend length might have an effect on the trend estimate, but is considered in the trend uncertainties. Note also, that for TROWARA, starting in 2000 instead of 1995 has quasi no effect on the trend estimate.

Reanalysis IWV at Bern increases significantly by 3.7 % per decade for MERRA-2 and by 2.3 % per decade for ERA5 data**, both for the period from 1995 to 2018**. [...]
The larger GNSS trend uncertainties compared to TROWARA and reanalysis trends are mainly due to the bias correction, which adds some uncertainty to the trend estimates. Further, all GNSS trends result from a shorter time period than TROWARA and reanalysis trends (see Table 1), which also increases the trend uncertainty and may lead to some trend differences.

> *Page 10, lines 291-292: to which GNSS stations with deviating trend values, due to instrumental issues are you referring here to? I guess not to Payerne, since the radiosonde measurements give trend values close to the GNSS IWV trends, no? Please specify the GNSS stations here.*

We were mainly refering to the EXWI trend, but we agree that the sentence is not clear in this summary and we removed it.

~~Some GNSS stations deviate from these trend values, probably due to instrumental issues.~~

> *Page 10, line 307-308: I don't understand this argument: strictly speaking, a (constant) bias between datasets will not have an impact on the trend difference between those datasets. Please explain.*

As mentioned in the last paragraph of section 4, the larger difference between MERRA-2 and TROWARA data in winter mainly occurs in the last decade, after 2008. This suggests, that TROWARA observed more IWV in winter than MERRA-2 in the last decade, whereas this feature was less visible in the years before. It is therefore not a constant bias and can be reflected in different winter trends. We adapted the text as follows.

This larger disagreement between TROWARA and reanalysis trends in December and January is consistent with the larger winter biases of TROWARA **starting in 2008** in Fig. 3c (Fig. 4c in new manuscript).

> *Page 12, line 345: this argument can be easily checked: do the same analysis for the reanalysis grids neighboring the used reanalysis grids at Bern. Please do so.*

As suggested, we investigated additionally monthly reanalysis trends for the used reanalysis grids. Indeed, they differ only slightly from the trends at the Bern grid, and are therefore not a main argument to explain differences between reanalysis and GNSS trends. We therefore removed these argument in the text and thank the referee for the remark.

p. 10, l. 304: ~~Most of the used GNSS stations are not located within the Bern reanalysis grid, which might explain some of the trend differences. The better agreement in summer suggests that the spatial difference is less critical in summer conditions.~~

p. 11, l. 345: ~~which might again partly be explained by the spatial difference of most GNSS stations to the used reanalysis grids at Bern.~~

> *Page 12, lines 346-following: as mentioned as a general comment: in this discussion, I was asking myself what the impact was on the constant RH approximation on the results of this analysis.*

We modified Fig. 5 (Fig. 6 in new manuscript) showing monthly trends around Bern, and added also RH trends from ERA5 to the figure. The text and interpretation have been adapted accordingly.

This discrepancy can be related to changes in RH, which was assumed to be constant (Eq. (9)). Indeed, our trends of ERA5 RH for the Bern grid (Fig. 6c) show that RH was not constant in those months, especially in winter but also in May and June. Even though the RH trends are not significantly different from zero, these results suggest that assuming RH to be constant may not be valid during all seasons, especially in winter. This makes the attribution of IWV trends to changes in temperature more challenging.

> *Page 12, line 365: here, but also at other places in the manuscript: please use the word "bias" only if you mention your reference. It is not clear from this sentence here if GNSS is biased low w.r.t. FTIR or if FTIR is biased low w.r.t. GNSS.*

Modified as followed.

~~A mean FTIR bias of $-0.26 \pm 0.3\,\mathrm{mm}$ is found when comparing monthly means of coincident GNSS and FTIR measurements. [...] Compared to the fully sampled GNSS time series, we found a bias of $1.05 \pm 0.61\,\mathrm{mm}$~~

~~(GNSS – FTIR), which corresponds to a bias of 34 % (using the mean of the fully sampled GNSS as reference).~~

Monthly means of coincident GNSS data have a mean dry bias of $-0.26 \pm 0.3\,\mathrm{mm}$ compared to FTIR ($\mathrm{GNSS_{coincident}} - \mathrm{FTIR}$) (Fig. 7c). Further, monthly means of fully sampled GNSS have a bias of $1.05 \pm 0.61\,\mathrm{mm}$ compared to FTIR (GNSS – FTIR), which corresponds to a bias of 34 % (using the mean of the fully sampled GNSS as reference).

> *Page 12, lines 371-377: as the IWV amounts at Jungfraujoch are very low, there might be another argument for the GNSS dry bias w.r.t. FTIR: as pointed out by Wang et al. (2007), under dry conditions, the GPS is less sensitive to low IWV values as other devices (reference: Wang, J., Zhang, L., Dai, A., Van Hove, T., and Van Baelen, J.: A near-global, 2-hourly data set of atmospheric precipitable water from ground-based GPS measurements, J. Geophys. Res., 112, D11107, doi:10.1029/2006JD007529, 2007). Please comment.*

We thank the referee for this indication. We checked the mentioned reference and also the COST report (Bock et al., 2020, p. 343) and added the comparison with Schneider et al. (2010) to the manuscript.

[...] the remaining bias of $-0.26mm$ when using coincident GNSS measurements indicates that GNSS measures slightly less IWV than FTIR. **This is consistent with results from Schneider et al. (2010), who report that GNSS at the high altitude Izaña Observatory (Tenerife) systematically underestimates IWV in dry conditions ($< 3.5\,\mathrm{mm}$).** ~~This result is consistent with previous studies that reported negative GNSS biases compared to~~ **Further, a dry bias has also been observed in previous studies that compared Jungfraujoch GNSS data with** Precision Filter Radiometer (PFR) data (Guerova et al., 2003; Haefele et al., 2004; Morland et al., 2006; Nyeki et al., 2005).

> *Page 13: the word "astonishingly" is not very scientific*

Modified.

From this point of view the achieved results are ~~astonishingly~~ good **and** a possible offset is not relevant for trend analyses as long as it is constant over the whole trend period.

> *Page 13, line 387: what is the trend estimate of the GNSS time series that coincides with the FTIR time series?*

The trend of GNSS data that coincides with FTIR measurements is 0.05mm per decade, which is similar to the FTIR trend (0.04mm per decade). We do not show this information in Fig. 7 to avoid overload, but we adapted the text accordingly.

The difference between both trends ~~might~~ **can** partly be explained by the dry sampling bias of the FTIR spectrometer, that measures only during clear-sky day conditions. **Indeed, the absolute GNSS trend is comparable with the FTIR trend when we use GNSS data coincident with FTIR measurements, with** $0.05\,\mathrm{mm}$ **per decade (not shown).**

> *Page 13, line 407: there's a typo in "autumn"*

Modified.

> *Page 14, lines 411-412: as already mentioned in the general comments: how many years do you need to obtain “stable” trends?*

As discussed in the general comment, it depends on the included sources of variability in the trend programme. We have shown that the trend uncertainty does not only depend on the length of the time series, but also on the number of change points in the time series. Our trend error estimates include the precision of the data, bias uncertainty between measurement periods after antenna updates, and atmospheric processes not accounted for by the trend model and thus showing up as auto-correlated fit residuals. They do not include nonlinear atmospheric variability on timescales longer than our analysis period. The latter are aliased into our linear trend. Thus our results should be understood as descriptive statistics only and should not be used to extrapolate beyond our analysis period.

> *Page 15, lines 454-455: the large spatial variability between the different grids in es changes in Merra-2 is surprising, especially when comparing with ERA5, which have a higher spatial resolution. What is the explanation for this?*

We assume that the large spatial variability in MERRA-2 e_s changes is due to the poor representation of the topography in the Alpine region, and thus lacking local dynamics, which results in underestimated temperature (compared to ERA5). Further, we found that this underestimation is reduced starting at around 2017, which leads to a jump in the MERRA-2 time series resulting in a very strong trend. We assume that this is related to a MERRA-2 model improvement in recent years (e.g. improved topography, changes in assimilated data) and adapted the manuscript accordingly.

In contrast to ERA5, The MERRA-2 changes in e_s show large spatial variability, with strong differences between different grids and generally larger values in the Alpine region. **These large values are caused by a jump in MERRA-2 tropospheric temperature data around the year 2017 in the Alpine grids, which results in strong temperature trends. This might be due to MERRA-2 model changes in topography or assimilated data. In the Alpine region** ~~On regional scales~~, the MERRA-2 changes in e_s (Fig. 14) thus differ strongly from the MERRA-2 IWV trends (Fig. 12).

To summarize, the reanalysis IWV trends follow on average the changes expected from temperature changes~~, but~~. **However** MERRA-2 shows large regional variability for changes in e_s ~~based on lower tropospheric temperature changes.~~ **due to biases in lower tropospheric temperature data in the Alpine region**.

Conclusions: The reanalysis grids probably miss regional variability in atmospheric stratification and convection, as it was also observed for zonal means by Wentz et al. (2000). **Further, a bias in MERRA-2 tropospheric temperature was observed for several Alpine grids after 2017.**

> *Page 16, lines 460-461: this statement can be easily verified: what is the winter temperature trend in the 1995-2018 interval?*

A simple linear trend analysis of MeteoSwiss surface temperature data (doi: `10.18751/Climate/Timeseries/CHTM/1.1`) confirms this statement. Whereas the temperature trend is positive for long-term winter temperature observations (e.g. starting in 1961), it is negative for the shorter study period from 1995 to 2018. We therefore adapted the text as follows.

This difference is ~~probably~~ due to our short study period. A few cold winters in the past 15 years ~~might~~ have hidden the overall positive temperature trend when looking only at the relatively short period from 1995 to 2018 (MeteoSwiss, 2019).

> *Page 28, Fig. 4: The datasets that have the smallest trend uncertainties (the reanalyses) are actually the ones of the course time evolution (monthly means), while GNSS data were available as hourly values, if I remember it correctly. Please comment on the trend uncertainties obtained here in this perspective.*

The GNSS data have indeed hourly resolution, but we use monthly means for the trend estimation. The initial resolution should therefore have little to no effect on the trend estimate. We think that the larger GNSS uncertainties are mainly due to the bias correction, which adds some uncertainty to the trend estimate, and due to the shorter time series. We added this explanation to the manuscript.

The larger GNSS trend uncertainties compared to TROWARA and reanalysis trends are mainly due to the bias correction, which adds some uncertainty to the trend estimates. Further, all GNSS trends result from a shorter time period (see Table 1) than TROWARA and reanalysis trends, which also increases the trend uncertainty and may lead to some trend differences.

2 Author's response to referee #2

2.1 General comment

> *This manuscript investigates the consistency of monthly means and linear trends of IWV from ground-based radiometry, FTIR, and GNSS data, and two modern reanalyses (ERA5 and MERRA-2) in Switzerland for the period 1995 to 2018. This study is of special interest to the climate community because it confronts several state of the art observational techniques and two of the currently best available reanalyses which assimilate a huge amount of observations (mainly from satellite remote sensors). Though the investigation is limited to the country of Switzerland, it shall shed some light on the uncertainties that generally impact both types of climate data and propose a methodology that may be replicated to other regions of the world. One specific source of uncertainty in trend analysis resides in the treatment of inhomogeneities in the data due to changes in the observing systems. In this respect the approach described by the authors with the GNSS data is original in the sense that it adjusts offsets (so-called jumps) in the time series that may be due to antenna and station changes. Though the proposed approach is interesting I think the way this issue is handled it is also the main weak point in the study.*

Thank you for this positive feedback and the constructive criticism, to which we respond in the following comments.

> *The approach of fitting simultaneously biases (or changes in the mean) and a linear trend is an ill- posed problem which requires special care and validation because of collinearity. Indeed, it is well known that these parameters are highly correlated, e.g. successive downward changes in the mean can be compensated by an upward trend hence over-estimating (underestimating) an existing positive (negative) trend in the series. Interpreting the trend parameter alone may thus lead to erroneous conclusions. Though the authors notice in several places that the trends after bias correction increase, they do not consider it as problematic, e.g. at NEUC the GNSS trend increases from 0.33 to 0.74 mm/decade when 3 jumps are adjusted and at PAYE the trend increases from 0.32 to 1.14 mm/decade after bias correction (when 3 jumps are fitted). It is noticeable that the uncorrected trends are in very good agreement with the ERA5 trend at Bern (0.34 mm/decade). Inspecting the station information table cited by the authors, reveals that 6 out of the 7 stations listed in Table 2 are subject to 3 or more jumps. The trends at all these stations are relatively high (above 0.58 mm/decade). One exception is station EXWI which did not undergo a bias correction and which has the smallest trend (0.02 mm/decade). It may be pure coincidence, but this point needs to be checked.*

It is true that most bias corrected trends are larger than the uncorrected trends. On the one hand, this can be due to positive biases in earlier years (or negative biases in later years), which leads to smaller trends if the bias is not corrected.
On the other hand, from the perspective of the trend model, this behavior tells us that the trends in each individual data block (after antenna updates) are larger than the overall (uncorrected) trend. The overall uncorrected trend is then assumed to be too small, due to unconsidered biases between the data blocks.
If a large bias is allowed (large bias uncertainty), the total trend tends to be the average of the individual trends of each data block. The algorithm then corrects the difference between the mean values of the blocks because they are biased to each other. Conversely, if a small bias correction is allowed (smaller bias uncertainty), the block differences contribute more to the trend estimate, and the intra-block trends have less weight. We chose a realistic bias uncertainty of 5% (see Sect. 3.1.1), which restricts the bias

correction to a realistic magnitude. The corrected trend is thus accounting for the biases between the data blocks, without giving the individual block trends too much weight. Therefore, the fact that the corrected trends are mostly larger than the uncorrected trends just means that the trends within the blocks are larger than the overall (uncorrected) trend. The latter, however, can be falsified (too small) because of the biases. The trend uncertainty of this corrected trend will be larger, which reflects the fact that the intra-block trends are less reliable than the overall trend would be.

> *[...] The large dispersion in the GNSS trend estimates over Switzerland (Fig. 8) compared to the relative flatness of the reanalysis trends (Fig. 10) is striking as well and may further support this problem.*

The newly added uncorrected trends in Table 2 show that the dispersion of uncorrected trends is not smaller than for the corrected trends. However, the trend values are often slightly larger, as discussed above.

> *I suggest that the authors further investigate the impact of the bias correction procedure on the GNSS trends and possibly reduce it.*

To better asses the bias correction, we added a section to the manuscript (Sect. 3.1) where the approach is illustrated with an artificial time series. We think, this better illustrates the effectiveness of the approach in correcting possible biases in a data set and justifies the use of the approach for the GNSS data.

The approach is illustrated with an exemplary case (Fig. 2). We used an artificial time series with a trend of $0.5\,\mathrm{mm}$ per decade, and added three change points with a constant bias for each subset. The change points represent for example an instrumental update, that leads to a constant bias in the following data. The biased time series has a trend of $1.19 \pm 0.06\,\mathrm{mm}$ per decade, which is too large compared to the true trend of $0.5 \pm 0.06\,\mathrm{mm}$ per decade. To improve the trend estimate, we add a fully correlated block in S_y for each biased subset, assuming a bias uncertainty of $5\,\%$. We obtain a corrected trend of $0.52 \pm 0.17\,\mathrm{mm}$, which corresponds within its uncertainties to the true trend of the unbiased time series. This demonstrates that the approach can reconstruct the true trend from a biased time series, with slightly increased trend uncertainties.

Further, we assessed the monthly uncertainties that were used for GNSS trends as being too small (only 1.5%), and added a systematic uncertainty of 0.7 mm to all GNSS monthly uncertainties, leading to monthly uncertainties of around 5% (see comment P7L200 below). This gives the algorithm some more liberty in the bias fit, which leads to a slightly smaller bias correction at some stations (e.g. trend of 7% in Payerne instead of 7.3% before). We think, this change is justified and the bias correction is improved.

> *One critical aspect here is with the position of the change points. A misplaced change point, or a change point inserted in a series where there is no jump, is very likely to bias the trend estimate. So it is of prime importance to detect and correct only the true change points. It is not clear from the manuscript how the jumps are detected. Is it from the station position time series? The authors should be careful that though some jumps in the position time series may have a coincident jump in the ZTD or IWV time series, this is not always the case. The reason is that, e.g. in the case of an antenna change, a station height offset can be due to both a PCO and a PCV change while a ZTD offset is mainly due to a PCV change (as long as*

> *the position is estimated simultaneously with ZTD). Only when change points are properly detected can the method of estimating biases and trends together provide accurate trend estimates.*

A changepoint was introduced as soon as an antenna has been replaced, even if it would not be visible in the coordinate time series. In the trend programme, the bias due to a changepoint in the data is fitted from the data itself. If we assume a jump even though the data do not show an offset (for example due to a PCO change that does not lead to a change in ZTD), the bias fit from the data would detect no jump and therefore not correct for it. A misplaced changepoint does thus not impact the trend itself. Only the uncertainty of the trend would be increased. This has been verified with an artificial time series, where we added a changepoint even though no jump in the data was observed. The trend estimate did not change, only the uncertainty increased slightly in the corrected trend. Thus, the bias correction method is a practical and objective tool for trend analysis of GNSS IWV series with possible jumps.

> *In order to provide proper insight into the impact of the correction, it would be useful to show the trends estimated from corrected and uncorrected time series.*

We agree that showing the uncorrected trends gives a better insight into the impact of the correction. Therefore, we added the uncorrected trends for the GNSS stations around Bern in Table 2 and added the statement shown below. However, the uncorrected trends may be falsified due to the missing consideration of jumps after antenna updates. We therefore prefer showing only the corrected trends in the figures.

For comparison, the GNSS trends without bias correction are also shown in Table 2. They are generally smaller than the bias corrected trends, suggesting that GNSS trends are mostly underestimated when biases are not accounted for. Further, their uncertainties are smaller, reflecting the additional uncertainty when biases are considered.

For your information, we provide the GNSS trend map with uncorrected trends here in the author's response (Fig. AR.1). We observe that most uncorrected GNSS trends are smaller, and have smaller uncertainties than the corrected trends. This is consistent with the statement that we added to the manuscript and the uncorrected trends added to Table 2. Uncorrected trends are significant at several stations. We think, however, that this is not reliable, because possible biases due to the instrumental updates are not considered, trends might thus be falsified and uncertainties underestimated. This is confirmed by our example with the artificial time series (Fig. 2), for which a false trend value with a small uncertainty is derived when the bias correction is not applied. We therefore show only the corrected trends in the figures of the manuscript. We think that the newly added uncorrected trends in Table 2 as well as the example cases (Figures 2 and 3 in the new manuscript) are sufficient to give an idea of the impact of the correction.

> *I recommend thus that the GNSS offset correction be improved or at least carefully assessed and the manuscript revised accordingly. The conclusions may change significantly.*

We think that our additional analysis with an artificial data set made a careful assessment of the bias correction approach possible. We think that our changes and this additional section improved the manuscript and thank the referee for this input.

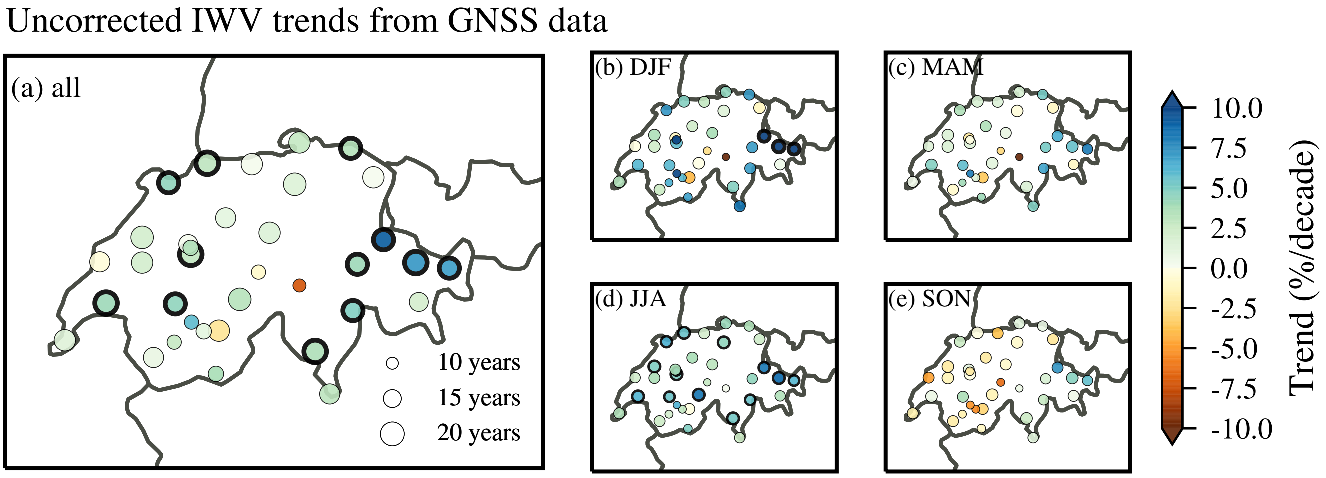

Figure AR.1: Uncorrected trends of IWV in Switzerland for the different GNSS stations for **(a)** the whole year and for the four seasons (**(b)** to **(e)**). The length of the GNSS time series is indicated by the size of the markers. Stations with trends that are significantly different from zero at 95% confidence interval are marked with a bold edge. Possible jumps in the time series due to GNSS antenna updates have not been considered in these trend estimates. The trends and their significance are therefore assumed to be not reliable. For the bias corrected trends, see Fig. 9 in the manuscript.

2.2 Minor comments

> *A number of additional specific comments are given below regarding other aspects of the work and the manuscript.*

> *P4L99: could there be an effect of the changes of the TROWARA instrumentation in 2002 and 2004 on the IWV data and IWV trend estimated with the radiometer? Is there a bias adjustment similar to that applied to the GNSS data or any kind of recalibration?*

Change points in TROWARA data due to instrumental changes before 2009 have been detected and corrected by a careful comparison of the TROWARA time series with a co-located weather station (Morland et al., 2009). No instrumental changes have been performed in recent years. We therefore presume that the data is well homogenized and use no bias correction in the TROWARA trend estimates.

> *Section 2.3: why did the author use the outdated Thayer constants and Bevis formula for the ZTD to IWV conversion? More elaborate and updated approaches are given in the COST GNSS4SWEC final report (see e.g. Chapter 6, Section 6.4.2). Note that the older refractivity constants may be responsible for a small bias in the GNSS IWV when comparing these data to other well calibrated observations (e.g. microwave radiometer and FTIR) and that ZHD and Bevis Tm computed from surface Ts observations and Ps may produce spurious diurnal and seasonal signals in the GNSS IWV estimates (due to spatial extrapolation errors and inaccuracies in the Bevis formula).*

We thank the referee for the suggestions of updated approaches for the ZTD to IWV conversation. We will consider using these newer approaches in future studies. We agree that the use of old refractivity constants may result in a bias in IWV, as shown in the mentioned GNSS4SWEC report (Bock et al., 2020). However, a possible bias which is constant in time would not affect the trend estimates. We therefore think that the approach used in our study is sufficient for the main purpose of our study, which is trend estimation.

Concerning surface pressure as well as surface and mean temperature data, we prefer to stay independent from reanalysis data, because we compare the GNSS data with reanalyses for validation. Also, the use of reanalysis data for the determination of T$_m$, T$_s$ and p$_s$ can lead to a bias, especially in mountainous areas, as shown by Alshawaf et al. (2017). We are aware of possible inaccuracies in the current approach and will consider using more advanced approaches in future studies. However, we think that for the current study, the required additional effort when reprocessing the whole data would be larger than the benefit.

> *P7L200: what does this standard deviation represent exactly? Is it the variability of the IWV data at the nominal time sampling (in this case what is the sampling of the GNSS IWV data? 1 hour?) or is it a formal error?*

The standard deviation used represents the variability around the monthly mean. We adapted the text to clarify this. The standard error depends indeed on the sampling. The time sampling of TROWARA is of several seconds (see Sect. 2.1), the GNSS sampling is 1 hour (see Sect. 2.3). For both datasets (TROWARA and GNSS), the standard error used for each monthly mean is around 1.5%. Within the review process, we assessed that this monthly uncertainty is too small for a total uncertainty of GNSS or radiometer measurements. We therefore decided to add a systematic uncertainty based on results from Ning et al. (2016). These new uncertainties are around 8% for TROWARA and 5% for a typical GNSS station. The updated uncertainties result in slightly different trends, but the effect is small and the more realistic uncertainties justify the change. The manuscript has been changed as shown below, and the Figures have been adapted to the new trend values.
For consistency, we also added the approximated value of FTIR uncertainties used. The FTIR uncertainties are much larger due to the low number of measurements per month and the model fit to obtain monthly means from the sparse data, as described in Sect. 2.2. We also add a sentence on the uncertainty used for the newly computed RH trends (10%) and for temperature trends (around 2.5K).

As monthly uncertainties σ_{mon}, we use for TROWARA and GNSS data

$$\sigma_{mon} = \sqrt{\sigma_{\bar{x}}^2 + \sigma_{sys}^2}, \tag{3}$$

where σ_{sys} is a systematic error and $\sigma_{\bar{x}}$ is the standard error of the monthly mean, given by

$$\sigma_{\bar{x}} = \sigma n^{-\frac{1}{2}}, \tag{4}$$

with the standard deviation of the monthly mean σ and the number of measurements per month n. The systematic error σ_{sys} is estimated to be 1 mm for TROWARA and $0.7\,\mathrm{mm}$ for GNSS data. These values are based on results from Ning et al. (2016a), who assessed IWV uncertainties from a radiometer and GNSS observations in Sweden. Our monthly uncertainties used for TROWARA and GNSS are on average 8 % for TROWARA and around 5 % for a typical GNSS station. FTIR uncertainties (around 25 %) are based on the model fit of daily means as described in Sect. 2.2. [...]
In addition to IWV trends, we determine reanalysis trends of RH (ERA5) and temperature (ERA5 and MERRA-2). We use monthly uncertainties of 10% to estimate RH trends, whereas the standard error of each averaged temperature profile (below 500 hPa) is used as monthly temperature uncertainties (around 2.5 K).

> *Why is the standard deviation of TROWARA 2 or 3 times smaller than the GNSS standard deviation? (is it because of the time sampling?)*

The larger trend uncertainty for the GNSS stations is not due to the assumed data uncertainty (which is similar to TROWARA, see comment before), but rather due to the shorter time series and the bias fitting which adds some uncertainty.

> *P8L221: explain how the jumps are detected (see also the general comments).*

A changepoint was introduced whenever a site antenna was replaced. A changepoint is also introduced when a jump would not be visible in the corresponding coordinate time series (especially station height). We adapted the manuscript to make this clearer. In rare cases, jumps were also set up when coordinate jumps were visible in the coordinate domain without an antenna change. Reasons for this are sometimes difficult to find out, but e.g. tree cuts or new buildings very close to the station may result in a different observation skyplot.

We introduced change points in the trend programme as soon as ~~antenna updates have led to~~ a **possible** jump in the GNSS data **has been** ~~as~~ recorded by swisstopo (available at `http://pnac.swisstopo.admin.ch/restxt/pnac_sta.txt`), **which is mostly due to antenna updates**.

> *Fig. 2b: how does the corrected GNSS time series compare to the ERA5 data? Are the biases reduced?*

In the corrected GNSS time series, the jump in the ERA5-difference is reduced. We added this corrected difference to ERA5 in Fig. 2b (Fig. 3b in new manuscript).

When the bias is considered in the trend model, **the jump in the difference to ERA5 is reduced (Fig. 3b). Further,** we obtain a larger bias corrected trend [...]

> *P9L269: compared to the TROWARA trends which are of similar magnitude, the GNSS trends are not deemed significant because the standard deviation is 2 or 3 times larger. This prompts again for a clarification on the reason of the difference in standard deviation.*

The larger uncertainties in GNSS trends are mainly due to the changepoints in the time series due to antenna updates, which adds some uncertainty to the trend estimates. Further, GNSS time series are shorter, which also slightly increases the uncertainty of the trend estimate. Both points explain the larger GNSS trend uncertainties. We added this explanation to the manuscript

The larger GNSS trend uncertainties compared to TROWARA and reanalysis trends are mainly due to the bias correction, which adds some uncertainty to the trend estimates. Further, all GNSS trends result from a shorter time period than TROWARA and reanalysis trends (see Table 1), which also increases the trend uncertainty and may lead to some trend differences.

> *P9L271: this sentence is very speculative. Revise or remove.*

We agree that this formulation is too vague and removed it.

~~However, the uncertainties of the monthly reanalysis data might be larger than the here used value of 10%, which would also lead to larger uncertainties of the reanalysis trends.~~

> *P9L274-275: The shorter time period of EXWI is suggested but WAB1 has a much shorter period. Revise or remove.*

We agree and removed this sentence.

The quasi zero trend at the EXWI station is less reliable than the other trends, because the EXWI station ~~provides data in a shorter time period (until 2016). Further, it~~ is not part of the AGNES network and therefore does not fulfil the same quality requirements.

> *P9L279-284: is the reliability of radiosonde data good enough to estimate trends? Similarly, the lidar trends are probably biased because they include clear sky measurements only. Moreover, the time period for the latter is much shorter. I think these results can hardly be inter-compared. Please remove.*

The reliability of radiosonde data for long-term trend analyses has been improved substantially with the homogenzitation efforts within the GCOS Reference Upper Air Network (GRUAN) (e.g. Dirksen et al., 2014). At Payerne, not all radiosonde profiles are GRUAN-certified, but Hicks-Jalali et al. (2020) showed that the difference between the homogenized GRUAN profiles and the operational radiosonde profiles lie within the GRUAN uncertainties, which justifies the use of the operational radiosonde data.
We agree that the short trend time period and a possible clear-sky bias of the lidar is a limitation of the study by Hicks-Jalali et al. (2020). However, the use of lidar data for water vapour retrieval is a new and promising technique, and we think that the comparison with the co-located GNSS data at Payerne is important. We therefore suggest to keep this short comparison, but added their limitations in the text as followed.

However, comparing their trend results with ours has to be done with care, because their trend time period is short and the lidar trends might contain a clear-sky bias.

> *P10L285-294: trend estimates computed for different periods must be compared with much care.*

We agree that comparing trends of different periods has to be done with care. However, we think it is important to set our results in the context of previous studies that used the same TROWARA data set. We added a sentence to point out to the reader that care has to be taken.

However, care has to be taken when comparing these TROWARA trends of different trend period lengths.

> *P10L295: why would trends be larger in summer? Do you have any clue for this?*

Absolute trends are largest in summer because more water vapour is available in summer. Relative trends might be larger in summer due to increased evaporation with increasing temperature. Hocke et al. (2019) showed that evaporation of surface water in Bern plays a major role in summer. Surface evaporation is temperature dependent, which might explain larger IWV summer trends due to a warming climate. IWV trends largest in summer have also been observed in China (Peng et al., 2017) or in Sweden (Nilsson et al., 2008). Further, we have shown that summer trends are consistent with the temperature increase in summer, as discussed in Sect. 4.2.

> *Fig. 5: what is the benefit of using the mean GNSS trend here? how is this mean computed?*

The arithmetic mean of trends from GNSS stations close to Bern is shown in Fig. 5 (Fig. 6 in new manuscript). We decided not to show all individual GNSS trends to avoid an overload of the figure. The monthly trends of the different GNSS stations close to Bern show a similar "shape" throughout the year. Therefore, the arithmetic mean is a good representation of those GNSS trends. The Figure caption has been adapted to clarify the mean computation.

Figure 5. Trends of IWV for different months for TROWARA in Bern, GNSS stations close to Bern **(arithmetic mean)**, and reanalysis grids (MERRA-2 and ERA5) at Bern.[...]
The mean trend (arithmetic mean) of the GNSS stations around Bern agrees with the other data sets in summer, [...]

> *Fig. 5: Add error bars on the monthly trend estimates to judge whether the variations are significant or not.*

Error bars have been added to Fig. 5 (Fig. 6 in new manuscript). The error bars represent the month with largest uncertainty for each data set. Some months have smaller uncertainties. Significant trends (usually with smaller uncertainties) are thus marked by filled dots. A sentence about the significance of the monthly trend differences has been added (see also general comment of referee 1).

However, the differences between those monthly trends are significant only at 68% confidence level.

> *Fig. 5: The GNSS trend for March is suspicious and needs to be checked.*

This strong negative averaged March trend is mainly due to strong negative trends at the EXWI station. However, also all the other stations around Bern show smaller March trends than reanalysis data. The reason for this negative GNSS March trend is not clear.

> *P10L304: The fact that the GNSS stations are not located in the same reanalysis grid (ERA5 or MERRA-2?) does not sound like a good explanation for the differences. Please revise or remove.*

We investigated monthly reanalysis trends for the used reanalysis grids. Indeed, they differ only slightly from the trends at the Bern grid point, and are therefore not a main argument to explain differences between reanalysis and GNSS trends. As suggested, we removed this argument in the text and thank the referee for the remark.

p. 10, l. 304: ~~Most of the used GNSS stations are not located within the Bern reanalysis grid, which might explain some of the trend differences. The better agreement in summer suggests that the spatial difference is less critical in summer conditions.~~
p. 11, l. 345: ~~which might again partly be explained by the spatial difference of most GNSS stations to the used reanalysis grids at Bern.~~

> *P11L315: mention that the temperature trends explain part of it and that they are further discussed in section 4.2.*

We thank for this adequate remark and adapted the manuscript as followed.

Nevertheless, we showed that the IWV trend peak is consistent with November temperature trends, suggesting that those trends are temperature driven (see Sect. 4.2 and Fig. 6).

> *P11L325-328: contains some inaccuracies. Suggested correction: "In case that the water vapour pressure e is smaller than es, the available water is in vapour phase, whereas for e > or = es it condenses. With increasing temperature, es increases, which leads for a given relative humidity (RH) to an increase of water vapour e. Changes in es can therefore directly be compared to changes in the amount of water vapour measured by e or IWV, assuming that the RH remains constant"*

Thank you for this clarification, we adapted the text as followed.

In case that the water vapour pressure e is smaller than e_s, the available water ~~evaporates~~ **is in vapour phase**, whereas for $e \geq e_s$ it condenses ~~until it reaches the equilibrium ($e = e_s$)~~. With increasing temperature, e_s increases, which leads for a given **relative humidity (RH)** to an increase of e ~~water vapour in the vapour phase~~. Changes in e_s can therefore directly be compared to changes in the amount of water vapour **assuming that RH remains constant [...]**

> *P11L325-328: Give a reference to other studies concluding that RH remains nearly constant and be more specific about the conditions for this to hold, namely is this result valid at regional scale?*

The assumption of fixed RH has been presented by Möller (1963), and Held et al. (2000) discusses the usefulness of this assumption to evaluate water vapour feedbacks. We added these references to the manuscript. Note that following the suggestion of referee 1, we added a full analysis of ERA5 RH trends, indicating that RH seems not to be constant in all seasons. The validity of the assumption is thus further discussed in Sect. 6.2 and 7.

[...] assuming that RH remains constant **(Möller, 1963; Held and Soden, 2000)**

> *P12L347-355: I am not convinced of the impact of temperature inversions. Do you have further evidence of this? Instead, the effect of moisture transport is a better explanation (e.g. also discussed by Parracho et al., 2018).*

We agree that this assumption would require further evidence, e.g. additional analyses. We have shown that inconstant relative humidity can partly explain the observed discrepancies in winter monhts. Further, we agree that moisture transport may be a more dominant factor. We think that those 2 aspects are sufficient to explain the observed differences and therefore removed the assumption of winter temperature inversions.

~~This discrepancy might be related to temperature inversions that are often present in winter. In these situations, the surface water vapour is trapped by the inversion and less propagating in the free troposphere (Fujita and Sato, 2017), which means that the water vapour content does not scale anymore to temperature above the inversion layer.~~

> *P13L392: could the snow and icing problems effect the GNSS data used in this study?*

As mentioned in Sect. 5, the old antenna (before 2016) at Jungfraujoch was never calibrated, because the special heated radome construction to avoid icing made the calibration complicated. The authors in the mentioned study (Nyeki et al., 2019) also refer to this problem, which is related with the snow and icing at Jungfraujoch. We reformulated the sentence to make their intention clearer. We account for this antenna change in the trend estimate by the bias correction.

They decided not to use GNSS IWV data from Jungfraujoch due to the high IWV variability and **the missing calibration of the antenna before the replacement in 2016.**

> *P13L403: is it 3 or 4 stations?*

Corrected, it is three stations.

Only ~~four~~ **three** stations (~~11~~$9\,\%$ of all stations) show negative IWV trends [...]

> *P14L410: trends computed for different periods and regions must be compared with much care.*

We agree that comparing GNSS IWV trends from different regions and periods can be problematic. However, we think that it is important to put our results in the context of other GNSS based trend results in Europe. We therefore keep this comparison, but add a sentence about the problematic of comparing trends of different lengths.

Note, however, that both studies use different trend periods lengths than in our study, which makes trend comparisons difficult.

> *Fig. 13: these plots are weird. It seems that there is glitch in the data computation here. Please check carefully. Otherwise, a proper explanation of the MERRA-2 results should be given.*

We investigated the MERRA-2 lower tropospheric temperature data in more detail. From Fig. 13 (Fig. 14 in new manuscript), it is evident that the unrealistic high MERRA-2 temperature trends occur mainly in grids with complicated topography (Alps). In those grids, the topography is poorly represented, which leads to lacking local dynamics and to biases in tropospheric temperature. We have seen that in all grids where the MERRA-2 temperature trend is large, the temperature is substantially underestimated compared to ERA5. However, this underestimation is reduced for all grids starting at around 2017, which leads to a jump in the MERRA-2 time series resulting in a very strong trend. We assume that this is related to a MERRA-2 model improvement in recent years (e.g. improved topography, changes in assimilated data). We adapted the manuscript accordingly.

In contrast to ERA5, The MERRA-2 changes in e_s show large spatial variability, with strong differences between different grids and generally larger values in the Alpine region. **These large values are caused by a jump in MERRA-2 tropospheric temperature data around the year 2017 in the Alpine grids, which results in strong temperature trends. This might be due to MERRA-2 model changes in topography or assimilated data. In the Alpine region** ~~On regional scales~~, the MERRA-2 changes in e_s (Fig. 14) thus differ strongly from the MERRA-2 IWV trends (Fig. 12).
To summarize, the reanalysis IWV trends follow on average the changes expected from temperature

changes~~, but~~. **However** MERRA-2 shows large regional variability for changes in e_s ~~based on lower tropospheric temperature changes.~~ **due to biases in lower tropospheric temperature data in the Alpine region**

Conclusions: The reanalysis grids probably miss regional variability in atmospheric stratification and convection, as it was also observed for zonal means by Wentz et al. (2000). **Further, a bias in MERRA-2 tropospheric temperature was observed for several Alpine grids after 2017. This reflects the problematic of using reanalysis data for trend estimates due to changes in observing systems or assimilated data.**

> *P15L444-445: I think it is not clear if it is the reanalyses results that are too smooth or the GNSS results that are too noisy. Can you clarify/discuss this point?*

The coarse resolution of the reanalysis grids is not capable to represent the topography correctly, especially in mountainous regions. As shown for example by Bock et al. (2005) or Alshawaf et al. (2017), reanalysis data can deviate substantially from GNSS data, especially in mountainous regions. If this bias is constant, it should not affect the trend estimates. However, spatial small-scale variability that is captured by GNSS stations and not by reanalyses might still lead to some differences in the trends. Further, it is evident that ungridded point measurements are more variable than gridded reanalysis data. To clarify this point, we adapted the text as followed.

However, the reanalyses do not resolve small scale variability, which can explain the differences to some GNSS station trends. **Further, the GNSS point measurements are generally more variable than the gridded reanalyses data.** Alshawaf et al. (2017) also observed larger differences in mountainous regions between GNSS derived IWV ~~trends~~ and reanalyses **data** in Germany.

> *P15L462-468: The uncertainty of the reanalyses is not directly addressed along in this work. So, drawing doubt on these data in the final discussion sound rather awkward. This being said, it is true that the use of reanalyses for trend analysis has been debated but the quoted references suggesting that ERA5 and MERRA-2 are more accurate at estimating trends are too general to support such an assertion. Please revise or remove this paragraph.*

We agree that the discussion of the reanalysis limitations and uncertainties is misplaced. We therefore moved it to the description of the reanalysis data sets in section 2.4. Further, we changed the formulation to respond to your suggestion.

When using reanalysis data for trend estimates, one has to keep in mind their limitations. Due to changes in observing systems of the assimilated data, the use of reanalyses for trend studies has been debated (e.g. Bengtsson et al., 2004; Dee et al., 2011; Parracho et al., 2018; Sherwood et al., 2010). ~~This problem has been addressed in the recent reanalysis products. Compared to the previous ECMWF reanalysis (ERA-Interim), the bias correction of assimilated data in ERA5 has been extended to more observation systems (Hersbach et al., 2018). Also MERRA-2 has been improved compared to the previous MERRA data, for example with reduced biases in the water cycle (Gelaro et al., 2017). Further analyses would be required to conclude if biases due to changes in observing systems still exist in the reanalysis data over Switzerland.~~ **The recent reanalysis products contain some improvements in handling possible steps in assimilated data. For example, the bias correction of assimilated data in ERA5 has been extended to more observation systems (Hersbach et al., 2018) and MERRA-2 reduced certain**

biases in water cycle data (Gelaro et al., 2017). Nevertheless, future studies have to assess whether these improvements affect the reliability of reanalysis data for trend estimates.

> *Fig. 10 and 11: to ease the comparison with GNSS results, add the GNSS results onto the reanalysis results. The plots may be resized.*

As proposed we added GNSS trends on the reanalysis IWV trend maps (Figs. 11 and 12 in new manuscript). We restricted the GNSS trends to stations with longest time series (18 or 19 years) to avoid the direct comparions with much shorter time series.

> *Table 1: add the number of jumps that are inserted in the trend analysis.*

The dates of change points have been added to Table 1.

3 References

Alshawaf, Fadwa, Kyriakos Balidakis, Galina Dick, Stefan Heise and Jens Wickert (2017). “Estimating trends in atmospheric water vapor and temperature time series over Germany”. In: *Atmos. Meas. Tech.* 10.9, pp. 3117–3132. DOI: `10.5194/amt-10-3117-2017`.

Bock, O., R. Pacione, F. Ahmed, A. Araszkiewicz, Z. Bałdysz, K. Balidakis, C. Barroso, S. Bastin, S. Beirle, J. Berckmans, J. Böhm, J. Bogusz, M. Bos, E. Brockmann, M. Cadeddu, B. Chimani, J. Douša, G. Elgered, M. Eliaš, R. Fernandes, M. Figurski, E. Fionda, M. Gruszczynska, G. Guerova, J. Guijarro, C. Hackman, R. Heinkelmann, J. Jones, S. Zengin Kazancı, A. Klos, D. Landskron, J. P. Martins, V. Mattioli, B. Mircheva, S. Nahmani, R. T. Nilsson, T. Ning, G. Nykiel, A. Parracho, E. Pottiaux, A. Ramos, P. Rebischung, A. Sá, W. Dorigo, H. Schuh, G. Stankunavicius, K. Stępniak, H. Valentim, R. Van Malderen, P. Viterbo, P. Willis and A. Xaver (2020). “Use of GNSS Tropospheric Products for Climate Monitoring (Working Group 3)”. In: *Advanced GNSS Tropospheric Products for Monitoring Severe Weather Events and Climate.* Ed. by J. Jones et al. Springer Nature Switzerland AG. DOI: `10.1007/978-3-030-13901-8_5`.

Bock, Olivier, Christian Keil, Evelyne Richard, Cyrille Flamant and Marie Noelle Bouin (2005). “Validation for precipitable water from ECMWF model analyses with GPS and radiosonde data during the MAP SOP”. In: *Q. J. R. Meteorol. Soc.* 131.612, pp. 3013–3036. DOI: `10.1256/qj.05.27`.

Dirksen, R. J., M. Sommer, F. J. Immler, D. F. Hurst, R. Kivi and H. Vömel (2014). “Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde”. In: *Atmos. Meas. Tech.* 7.12, pp. 4463–4490. DOI: `10.5194/amt-7-4463-2014`.

Held, Isaac M. and Brian J. Soden (2000). “Water Vapor Feedback and Global Warming”. In: *Annu. Rev. Energy Environ.* 25, pp. 441–475.

Hicks-Jalali, Shannon, Robert J. Sica, Alexander Haefele, Giovanni Martucci, Eliane Maillard Barras and Jordan Voirin (2020). “A Raman Lidar Tropospheric Water Vapour Climatology and Height-Resolved Trend Analysis over Payerne Switzerland”. In: *Atmos. Chem. Phys. Discuss.*, pp. 1–29. DOI: `https://doi.org/10.5194/acp-2019-1089`.

Hocke, Klemens, Leonie Bernet, Jonas Hagen, Axel Murk, Matthias Renker and Christian Mätzler (2019). “Diurnal cycle of short-term fluctuations of integrated water vapour above Switzerland”. In: *Atmos. Chem. Phys.* 19.19, pp. 12083–12090. DOI: `10.5194/acp-19-12083-2019`. URL: `https://www.atmos-chem-phys.net/19/12083/2019/`.

Möller, F. (1963). “ On the influence of changes in the CO 2 concentration in air on the radiation balance of the Earth’s surface and on the climate ”. In: *J. Geophys. Res.* 68.13, pp. 3877–3886. DOI: `10.1029/jz068i013p03877`.

Morland, J., Collaud M. Coen, Klemens Hocke, P. Jeannet and C. Mätzler (2009). “Tropospheric water vapour above Switzerland over the last 12 years”. In: *Atmos. Chem. Phys.* 9.16, pp. 5975–5988. DOI: `10.5194/acp-9-5975-2009`.

Nilsson, T. and G. Elgered (2008). “Long-term trends in the atmospheric water vapor content estimated from ground-based GPS data”. In: *J. Geophys. Res. Atmos.* 113.19. DOI: `10.1029/2008JD010110`.

Ning, T., J. Wang, G. Elgered, G. Dick, J. Wickert, M. Bradke, M. Sommer, R. Querel and D. Smale (2016). “The uncertainty of the atmospheric integrated water vapour estimated from GNSS observations”. In: *Atmos. Meas. Tech.* 9.1, pp. 79–92. DOI: `10.5194/amt-9-79-2016`.

Nyeki, Stephan, Stefan Wacker, Christine Aebi, Julian Gröbner, Giovanni Martucci and Laurent Vuilleumier (2019). “Trends in surface radiation and cloud radiative effect at four Swiss sites for the 1996–2015 period”. In: *Atmos. Chem. Phys.* 19.20, pp. 13227–13241. DOI: `10.5194/acp-19-13227-2019`.

Parracho, Ana C., Olivier Bock and Sophie Bastin (2018). “Global IWV trends and variability in atmospheric reanalyses and GPS observations”. In: *Atmos. Chem. Phys.* 18.22, pp. 16213–16237. DOI: `10.5194/acp-18-16213-2018`.

Peng, Wang, Xie Tongchuan, Dai Jiageng, Sun Jingmin, Wu Yanling, Song Qingli, Dai Xin, Yuan Hongliang, Song Dejun and Zhao Jinrong (2017). “Trends and Variability in Precipitable Water Vapor throughout North China from 1979 to 2015”. In: *Adv. Meteorol.* 2017, pp. 1–10. DOI: `10.1155/2017/7804823`.

Schneider, M., P. M. Romero, F. Hase, T. Blumenstock, E. Cuevas and R. Ramos (2010). “Continuous quality assessment of atmospheric water vapour measurement techniques: FTIR, Cimel, MFRSR, GPS, and Vaisala RS92”. In: *Atmos. Meas. Tech.* 3.2, pp. 323–338. DOI: `10.5194/amt-3-323-2010`.