

General comment by authors:

We appreciate the generally positive and encouraging feedback from the two reviewers. We also acknowledge the reviewer comments which led to improvements of the manuscript.

Below comments from the reviewers are recalled in black while our response is given in **blue** and *italic* and changes to the manuscript in **green**.

Anonymous Referee #1

Received and published: 19 March 2018

This well written paper introduces the archive of water vapor related data records that have been used within the phase 1 of the GEWEX Water vapor Assessment. All the data have been regridded on a common 2x2degrees grid allowing easy comparisons. A large number of products are considered (22 for TCWV). The products are quickly presented individually and the references to the original papers are given making the present paper a very useful resource for climate scientists. I think this database is very unique in the sense that it simplifies the collection of a wide suite of individual products and facilitates their use though common gridding. I think the contextualisation (the sections 2 and 3), which is broader than the actual list of products of the archive, also contributes to making this paper an excellent entry point in the field. The illustrations given at the end of the paper are very relevant science-wise and clearly show the benefit of using this archive for further analysis.

Thanks for the positive and encouraging feedback.

Changes to the manuscript: None.

I recommend publication of the paper but I have couple of comments that could possibly help in making the paper even more clear, if accounted for.

1) mention of scientific results: I would propose to move the references from Schroder et al page 25 line 18 to the introduction of the paper to very quickly provide the readers with the information that the archive is part of a large project that is already producing refereed literatures.

We will update the manuscript as proposed.

Changes to the manuscript: We propose to include the following sentences: “. This archive forms the basis for scientific analysis in the characterisation effort of satellite and reanalysis data records and results from such analysis have already been published (Schröder et al. 2016, 2017a).” (page 2, line 32, after “...data archive”).

2) some references are a bit dated. For instance: page 5 section 2.2 line 23 reference is Shi et al., 2008. Don't you have any more recent efforts on the intercalibration of the HIRS instruments to list here? I suggest the authors try to update as much as they can their reference list.

We will go through the manuscript and update the references. Please note that the references should fit to the version of the elements of the G-VAP data archive.

Changes to the manuscript: A couple of references have been updated.

3) Mixing between algorithms and products. Section 4.1 is listing the products that compose the database but for some of these it is indeed algorithms that are mentioned. While I fully understand there are not yet any formal way to report these information, I wish a better distinction between the algorithms and products is made in this section. Similarly I have hard time with the nicknames used. For instance "TMI". This is no algorithm nor product but an instrument and the data record should be REMSS TMI only TCWV version 7. I think it will benefit the readers and the climate scientists to have a more homogeneous way of naming the products.

We will carefully go through section 4.1 and separate between algorithm and data record. It is our intention that these paragraphs mention the algorithms and describe the data record's technical specification. Note that frequently the retrieval, the data and the validation results are described in a single reference.

Whenever a data record name is available or has been provided by the data record's PI (i.e., merged microwave REMSS) we will use it, even when it is also used for the algorithm. If not available our original intention was to

use short and unique names. Thus, we were speaking of the TMI data record. We agree that a more appropriate approach should be chosen. When no data record name is available we propose to combine sensor and data record owner to form the data record name and only mention the version and the parameter if required by context. We think that this way a reasonable balance between information and readability is reached. Also, the overview table provides full details.

Changes to the manuscript: various minor, editorial changes. The following introduction was included in section 4.1: “The elements of the G-VAP data archive are given in Table 2 and are marked with asterisks. In this section abstracts are provided for each element of the archive. These abstracts briefly introduce the input data, the retrieval scheme, technical specifications, provide basic references to the algorithm and/or the data record and give access details.” In order to have the abstracts in alphabetic order we slightly changed the order – this is NOT in track change in order to enhance readability. We have also changed the date record names given in figures 3+4 and adapted the figure captions accordingly.

4) The future of the database. It is only alluded to in the conclusion, and I think it deserves more discussions. Do the authors plan to update the record (with the same regridding)? Some records are stalled in 2008 that is already 10 years ago. Do the authors have a plan to update some of the active records? What about new versions? Changes of the algorithms? Given the uniqueness of the database, I think it is important to better document its much expected evolution.

This is a valid comment. Actually we were thinking in the same direction: we plan to update the archive by including new data records or new data record versions. Reasons for the availability of new data record versions can be manifold and may include retrieval changes or extended temporal coverage. It is in particular planned to include data records on original temporal coverage into the G-VAP data archive. We are confident that this will enhance the usability of the archive. Note that G-VAP is a generally unfunded and volunteer effort, which limits how much can be done.

Changes to the manuscript: The following paragraph on the future of the G-VAP data archive has been added to the manuscript (section 4.3, after last paragraph): “As G-VAP will be continued the G-VAP data archive will be updated over the next years and released in approximately 2020. The updated release will include newly generated data records, data records which have not been considered yet and new versions of elements of the current G-VAP data archive. It is planned to keep the overall format and technical specifications of the G-VAP data archive, with one important change: In order to keep the gap between last year covered by elements of the archive and release date at minimum and in order to enhance applicability, the data records will not be provided on common period but on individual temporal coverage.”.

Misc. page 7 line 23. "aata" should be "data"

Changed as proposed.

Changes to the manuscript: modified accordingly.

Anonymous Referee #2

Received and published: 5 April 2018

This impressive paper provides a great deal of detail on observations and data assimilation products from which total column water vapor as well as vertical profiles can be obtained. It will be an outstanding and very useful reference for researchers interested in evaluating and intercomparing estimates of water vapor in the troposphere. I couldn't tell whether the stratospheric water vapor estimates were treated as comprehensively; they are less familiar to me and I may have failed to fully grasp some of the details. The paper is well written for the most part, but is a bit uneven in some spots; for example, the data record descriptions in section 4.1 appear to have been pasted together from individual contributions without any linking overview. In a few other spots, there are typos and language errors that should be fixed, for example in the first paragraph of section 4 (bottom of page 8). I think that the manuscript will benefit from a careful reading and editing, but these are small points that do not impact the general merit of the paper.

Section 4.1 would indeed benefit from a short introductory paragraph. We will carefully read the manuscript and check for typos and language errors. In addition we will also go through section 4.1 and harmonise the wording in order to enhance readability.

Changes to the manuscript: various minor changes and newly introduced paragraph at the beginning of section 4.1: "The elements of the G-VAP data archive are given in Table 2 and are marked with asterisks. In this section abstracts are provided for each element of the archive. These abstracts briefly introduce the input data, the retrieval scheme, technical specifications, provide basic references to the algorithm and/or the data record and give access details." In order have the abstracts in alphabetic order we slightly changed the order – this is NOT in track change in order to enhance readability.

I applaud the opportunity for the authors to publish such a paper, which will make the wide range of water vapor information products much more accessible than would otherwise be the case. I have had the impression that the most usable such products for studies of climate variability, and to some extent change, were those derived from data assimilation systems, since all of the satellite observations and estimates are significantly biased in their sampling - IR products can't see into or below clouds, microwave products are challenged over land and ice, near IR and visible products tend to be daytime only - and therefore the integrated analyses formed from them have relatively poorly known biases. Of course, since the reanalyses attempt to use these estimates and are themselves challenged by biased input, my inclination to favor them may be faulty as well.

It is indeed the case that data from reanalysis are widely used. We would argue that the dominant reasons for this are that reanalysis data is gap-free and includes a wide range of variables. In terms of biases we would argue that both satellites and reanalysis may be affected by systematic uncertainties which are in both cases typically not known. Various recently initiated activities focus on the estimation of sampling uncertainties. First results have been published in Diedrich et al. (2016), Hersbach (2017), Höschen and Schröder (2016) and Mears et al. (2018).

References:

Diedrich, H., F. Wittchen, R. Preusker, and J. Fischer, 2016: Representativeness of total column water vapour retrievals from instruments on polar orbiting satellites. *Atmos. Chem. Phys.*, 16, 8331–8339, <https://doi.org/10.5194/acp-16-8331-2016>.

Hersbach, H., 2017: News from C3S: ERA5. Available at <https://www.ecmwf.int/sites/default/files/elibrary/2017/17312-news-c3s-era5.pdf>

Höschen, H., and M. Schröder, 2016: An analysis of the diurnal sampling bias using GNSS data, G-VAP report, available at http://gewex-vap.org/?page_id=19.

Mears, Carl A., Deborah K. Smith, Lucrezia Ricciardulli, Junhong Wang, Hannah Huelsing, Frank J. Wentz, 2018: Construction and Uncertainty Estimation of a Satellite-Derived Total Precipitable Water Data Record over the World's Oceans. *Earth and Space Science*, accepted article, <https://doi.org/10.1002/2018EA000363>

Changes to the manuscript: None.

One great benefit of this paper will be to provide scientists with the raw material that will enable them to do a much better job of untangling the errors and uncertainties in the estimation and analysis of water vapor. The great significance of water vapor to tracking and understanding climate change, as well as climate variability on shorter time scales, makes this a very important task greatly facilitated by this paper and the availability of the datasets described. I compliment the authors on a challenging task well done, and I recommend publication.

Thanks for the encouraging and positive feedback.

Changes to the manuscript: None.