Interactive comment on "The CM SAF SSM/I-based total column water vapour climate data record: methods and evaluation against re-analyses and satellite" by M. Schröder et al.

Anonymous Referee #3

Received and published: 21 January 2013

This paper provides information on how the long term daily and monthly column integrated water vapor data were produced through homogenization of DMSP SSM/I brightness temperature measurements and by kriging objective analysis method to fill the gap, and validation results against other SSM/I product and reanalysis products. Considering that long-term water vapor data have been valuable and important for climate analysis; for example water vapor trend analysis has been a core of water vapor feedback and circulation changes under the global warming condition. It is true that there is no absolute standard or true value of water vapor amongst data sets used in this study. However, it is always valuable to add independent data set, in addition to existing data sets, not only because of quality assurance but also benefits to understand retrieval characteristics and to have different views on the derived scientific implications. I am sure that this type of data set will bring in much benefit to the science community, especially in climate science area. For the given importance and impact, I recommend the paper to be published in Atmos. Meas. Tech. Nonetheless I request to improve manuscript by clarifying some places which the reviewer thought are somewhat confusing. My minor comments are following:

Title: against satellite does not sound all right. It would be other SSM/I product.
As suggested by reviewer #1 we now compare also to WVPA from TMI and therefore prefer not to change the title.

6424.5: WVPA – I am not sure about what it is. It doesn't seem to be linked to total column water vapor.
For historical reasons the term WVPA is used. Also, the product is well documented by CM SAF (see www.cmsaf.eu). Throughout the documentation the term WVPA is used.

6424.23: Since there is no standard value, the absolute bias misleads the meaning. How about relative bias?
No – here, we refer to the absolute value of the bias, that is, independently of its sign. We changed to "absolute value of the bias".

6427.1-7: This paragraph appears not to be finished. Make the paragraph more complete.
We changed the last sentence to: "Considering…observed which is slightly larger than the theoretically expected change (Trenberth et al., 2005)."

6430 Eq. (1): The equation is expressed with perturbations because variables were then treated to produce variance and covariance. Please say so.
No – the equation is expressed with errors as outlined in line 5. Neighbouring values with smaller errors will get larger weights when similar redundancies are observed. Variance and covariance are introduced after differentiation of Eq. 2 and are determined using x_i. The error covariance appears on the diagonal only – see Eq. (3). We also changed 6430, line 12 into: "…as a matrix equation (Eq. 3)."

6430 Eq. (2) In the parenthesis, sign should be positive, consistent with in Eq. (1).

Modified accordingly.

6430. 16: _xi/_xj ; i should not be same as in (2) because i=1, ..., n. You cannot simply use i, and j here. Better find another notation, and i = dummy, j = dummy too. Please clarify this.
When considered the error covariance would appear as off-diagonal terms in Eq. (3), similar to the covariance terms. Therefore, i, j=1,…,n with i unequal j – the i=j case is the error variance on the diagonal.
We changed to: "Futhermore, all error covariances [delta_x_i, delta_x_j] (with i unequal j) are zero,…"

6430. 18: Is the second assumption spatial correlation?
The (spatial) error covariance (after normalisation "correlation") is meant. We deleted this sentence in response to reviewer #2 and because the main message is in line 19.

6432. 8: What is 'there intermediate means'? Is it gridded data in different times? Better to describe it.
The main point here is that the averaging process occurs in two steps. The result of the first averaging leads to "intermediate means" which are based on synchronous observations at different spatial positions. We changed to "…are averaged to intermediate means at time of overpass on a…" (line 7).

6433. 2: How did you obtain the climatological averages here?
The climatological averages are arithmetic averages with contributions from all defined data of the full data record.

6433. 6: Provide some reasoning of high standard deviation near the edge of high humidity area. It may be related to seasonal progress of the humid area.
We added "The maxima in standard deviation to the north and south of the maxima in WVPA are associated with the annual North-South migration of the inner tropical convergence zone (ITCZ)."

6433. 10: I cannot understand what the time series average is? It needs to be clarified.
We changed line 2 to "The arithmetic average per grid of the defined data values of the full time series are …" and line 10 to "…as in the top left panel but …"

Typos
6434. 4: A comparison is made to the WVPA....
6434. 22: radio –>radiosonde
6435. 2: precursor –>previous
6441. 15: where –>when
6442. 4: a another–>another
All: Modified accordingly.

Anonymous Referee #2
Received and published: 14 January 2013

General vote: Minor revision

General comments:
The manuscript presents a climatological assessment of global total column water vapor (WVPA) data obtained by SSM/I satellite observations. Different existing retrieval approaches are compared, in order to assess the accuracy and the stability of these observations. In a further step global analysis (ECMWF) and reanalysis (ERA, JMA) data are

used to confirm the results. It can be shown that the stability of the retrievals is sufficient to observe long-term WVPA trends in the atmosphere.

In general, the manuscript is well structured; however some specific comments and technical corrections which request minor revision are listed below.

Specific comments:
In section 2.1 ("Retrieval"), the accuracy of the statistical retrieval by Schüssel and Emery (1990) is given as 0.15 kg m-2. However, in the original publication this is stated as 0.15 g cm-2 which translates to 1.5 kg m-2, thus 10 times less!! It has to be seen throughout the manuscript if this has any implications on the discussion of results. In addition, the phrase: "More than 98% of the variance within the training data set could be explained by the regression formula" is not correct, the variance can be explained by the 22 GHz channel only.

The accuracy is indeed 1.5 kg/m2. This will be corrected in the manuscript.

We propose to rephrase and re-structure the two paragraphs:

"Schlüssel and Emery (1990) developed four retrieval schemes for WVPA utilising different combinations of SSM/I channels. These schemes are of semi-physical nature, i.e., they are based on forward radiative transfer calculations on a set of atmospheric profiles followed by a statistical inversion using linear regression. All four variants of the total column water vapour retrieval purely depend on SSM/I brightness temperatures and regression coefficients derived from the initial set of atmospheric profiles. The retrievals do not depend on any additional information or ancillary data source.
For long-term applications the availability of the channels used by the retrieval is a crucial limitation. In view of the failure of the 85 GHz channel on F-08, the Schlüssel and Emery (1990) retrieval variant employing the 22 and 37 GHz channels has been implemented for long-term application, although the use of the 85 GHz channel data might improve accuracies slightly when present. The continuous availability and homogeneity of the input data source is clearly an advantage of the two channel retrieval scheme. With the two-channel variant 98.9% of the variance within the training data set could be explained by the regression formula. The remaining uncertainty, which may be interpreted as the accuracy of the statistical retrieval, is given by the authors as 1.5 kgm^{-2}. A detailed description of the retrieval schemes and the statistical inversion technique can be found in Schlüssel and Emery (1990)."

Section 2.2 which describes the kriging interpolation is rather detailed for the purpose of the paper and should be shortened.

We will shorten section 2.2 by removing p.6429, l. 23-27; two lines in the last paragraph of p.6430 and the second last paragraph on page 6431; three lines on p. 6432 and Eq. (5).
We described the kriging method in some detail because this paper is the first publication in peer-reviewed literature of the kriging method as implemented at CM SAF and because section 3.3 relies on details presented in section 2.2.

Technical corrections:

p.6425, l.25 ff.: Please rewrite this phrase for better understanding!
The phrase will be changed into: "Also, the analysis of synoptic scale water vapour transports can yield valuable insights into the dynamics of the atmosphere and its evolution".

p.6426, l.4-6: Please mention the values for accuracy and stability here!
Lines 4-6 will be changed into: "Requirements for accuracy (stated as bias) and stability, which is the temporal variation of the bias, for WVPA are provided by GCOS and published in GCOS-107 (2011): 1% (bias) and 0.3% per decade (stability)."

p.6434, l.1 -7: please rewrite this passage to more fluent sentences
Passage will be rewritten as follows: "Furthermore the CM SAF HOAPS products are compared to the previous HOAPS version as provided by MPI/UHH, to WVPA products from SSM/I derived by RSS (V6) and to various re-analysis data sets. The first comparison aims at the analysis of the effect of different mapping algorithms onto the products while the comparison to the RSS product and the reanalyses considers bias, RMSE and differences in linear trends (RSS product only)."

p.6436, l.6: afflicted does not seem to be the right word here
We propose to change into: "…and relative differences are mainly observed in storm track regions in which frequent rain events occur".

p.6424, l.11: operation capabilities (not operationS)
p.6424, l.13: An objective analysis for interpolation, NAMELY kriging, has been APPLIED (…)
p.6424, l.24: reanalyses (not reanalysis)
p.6425, l.4: better: (…) Earth's energy budget and water cycle (…)
p.6426, l.14: better: (…) was initiated by WMO in order to establish (…)
p.6427, l.16: (…) have been positively evaluated (…), not HAS!
p.6428, l.10-12: rewrite to: "Using the overlap between the DMSPs, probability density functions (PDFs) based on ten days of brightness temperatures have been calculated and statistically matched in each channel."
p.6428, l.18: rewrite to: The data are available (…)
p.6428, l.18-19: cancel "of the data set"
p.6428, l.27: cancel "or ancillary data source"
p.6431, l.2: rewrite to: The result is a linear set of equations, containing (…)
p.6434, l. 21-23: rewrite to: "Nevertheless, the physical retrievals show a significant positive bias and greater RMSE compared to the statistical retrievals which are based on a temporally and geographically limited selection of radio soundings.
p.6435, l.16: must not -> do not need
p.6435, l.26: rewrite to: (…) due to rain are filled (…)
p.6436, l.3: rewrite as: "These considerations will likely lead to patterns of absolute and relative difference as in Fig. 2 because (…)"
p.6436, l.9: deriving -> to derive
p.6436, l.15: add after CM SAF something like: "and will be explained below"
p.6436, l.20: cancel "assessed"
p.6437, l.2: cancel "before"
p.6437, l.17: introduced -> caused
p.6437, l.21: rewrite to: "(…) CM SAF data set that is not using sea surface temperature data."
p.6437, l.28: add "only" between "source" and "allows"
p.6438, l.3: better "removing" instead of "taking out"
p.6438, l.20-21: rewrite to: "(…) within the constraints of the model physics"
p.6438, l.22: use plural ("versions", "systems", "reanalyses")
p.6439, l.3: "In order to assess (…)
p.6439, l.4: add "additionally"
p.6439, l.11-12: rewrite to: "In all analysis data sets SSM/I information is assimilated either by using retrievals of total column water vapour (ERA40 and JCDAS-25) or by taking directly the SSM/I radiances (ERA-Interim and ECMWF operational analysis).
p.6439, l.24: cancel: "which is"
p.6440, l.3-5: rewrite as: "This is due to the number of satellites used in the CM SAF data set (see Fig. 3 for SSM/I temporal coverage)."
p.6440, l.26: change to: "bias" (singular) and "analyses" (plural)
p.6440, l.27: "bias" (singular)
p.6440, l.28: "values" instead of "biases"

p.6440, l.29: rewrite to: "In the years after 2002 (…)"
p.6441, l.1: "as bias is concerned" -> "in terms of bias"
p.6441, l.1: but -> and
p.6441, l.3: "temporally less stable" -> "more fluctuating"
p.6441, l.15: "started to be used" instead of "is used"
p.6442, l.19: kg m-2
p.6448, rewrite figure caption to: "Fig. 2. Absolute (in kgm-2, left panel) and relative (in percent, right panel) difference of WVPA between the CM SAF HOAPS v3.1 and the MPI/UHH HOAPS v3.0 averaged over the period 1992–2005."
p.6450: rewrite figure caption to: "Regional trends of monthly mean water vapour derived from deseasonalized CM SAF (left) and RSS (right) data sets for 1991–2006. Shown is the monthly trend of the regression analysis in kgm−2."
All: Modified accordingly.

Anonymous Referee #1
Received and published: 6 November 2012

This paper presents an improved record of daily and monthly SSM/I TPW over ocean, processed with homogenized brightness temperatures and an improved HOAPS algorithm. It represents an improvement over previous versions of the SSM/I TPW dataset.
As water vapor is a principal greenhouse gas, it is important to have high-quality observational datasets to monitor any significant increases in TPW. The presented dataset compares well to other well-known TPW datasets from both observations and re-analyses.

Specific Comments:

Page 6426 Line 24 :A paper was recently published in Geophysical Research Letters describing the NVAP-M dataset and indicates the temporal coverage will be 1988-2009.
The temporal coverage was changed as suggested and a new reference is included (Vonder Haar, T. H., J. L. Bytheway, and J. M. Forsythe (2012), Weather and climate analyses using improved global water vapor observations, Geophys. Res. Lett., 39, L15802, doi:10.1029/2012GL052094).

Page 6428 Line 2: SSM/I was not included on DMSP F9, and did fail on F12
The sentence has been changed to clarify which platforms are used:
"The WVPA climatology is based on data from six SSM/I instruments on DMSP F-08, F-10, F11, F13, F14, and F15 platforms."

Page 6428-29 Section 2.1: This section is somewhat confusing as written. You first state that the retrieval algorithm uses only the 19 and 22 GHz channels, then in the last paragraph mention that the algorithm was modified to not use the 85 GHz channel due to its failure on F8. It would be more clear to state up front what channels the algorithm uses, then that it was modified.

Schlüssel and Emery (1990) developed four retrieval versions for WVPA. For the reasons provided in the manuscript the version that utilises 22 and 37 (!) GHz observations is used. The section will be changed as follows:

Schlüssel and Emery (1990) developed four retrieval schemes for WVPA utilising different combinations of SSM/I channels. These schemes are of semi-physical nature, i.e., they are based on forward radiative transfer calculations on a set of atmospheric profiles followed by a statistical inversion using linear regression. All four variants of the total column water vapour retrieval purely depend on SSM/I brightness temperatures and regression coefficients

derived from the initial set of atmospheric profiles. The retrievals do not depend on any additional information or ancillary data source.
For long-term applications the availability of the channels used by the retrieval is a crucial limitation. In view of the failure of the 85 GHz channel on F-08, the Schlüssel and Emery (1990) retrieval variant employing the 22 and 37 GHz channels has been implemented for long-term application, although the use of the 85 GHz channel data might improve accuracies slightly when present. The continuous availability and homogeneity of the input data source is clearly an advantage of the two channel retrieval scheme. With the two-channel variant 98.9% of the variance within the training data set could be explained by the regression formula. The remaining uncertainty, which may be interpreted as the accuracy of the statistical retrieval, is given by the authors as 1.5 kgm^{-2}. A detailed description of the retrieval schemes and the statistical inversion technique can be found in Schlüssel and Emery (1990).

Page 6435 Paragraph 2: Was homogenized SSM/I used in HOAPS V3? Could there be differences in the two products due to the use of a homogenized Tb dataset in one and not the other?

No, both datasets use the same level-2 (swath-based) input data from HOAPS-V3.

Page 6436 Section 3.4: What version of RSS SSM/I TPW was used? V7 had higher TPW at higher values than V6 and could impact your results somewhat.

The comparison is based on RSS SSM/I TPW V6. This was the most recent version at the time of writing. V7 was released around time of submission.We will include the data set version in the updated manuscript.
RSS provides the following estimates on the difference between V6 and V7: Vapour values between 50-60 mm and above 60 mm increased by 1% and 2-3%, respectively. Note that, the majority of grid points contain values less than 50 mm as can be seen in Figure 1 of this manuscript. When considering V7 an impact can be expected but with the above arguments we are certain that general statements remain valid and that average bias values are not strongly changed.

Section 3: All comparison data include SSM/I in some way. A comparison to a completely independent dataset, such as TOPEX or TRMM-TMI from 1997-2006 would strengthen your argument for successful monitoring of oceanic water vapor trends.

We have added the TMI water vapour from Remote Sensing Systems (TMI monthly products version 04) to the satellite product evaluation (section 3.4).

Section 3: Was there any accounting for the kriging error in the comparisons? That is, were areas with high uncertainty values due to interpolation included in the comparisons with other data sets as well as in the trend analysis? If so, how was this high uncertainty accounted for with respect to RMSE, bias, and trend calculations?

The kriging error was not accounted for in the inter-comparisons, that is, all valid values contribute to the uncertainties independently of the kriging error. It is not expected that accounting for the kriging error would change the results very much as data gaps are very small for daily estimates and not existing for monthly estimates. Thus, high uncertainties do not arise due to filling of data voids with interpolated data.
The trend uncertainty estimation utilises the standard deviation of the anomalies.

Figures 3 and 7: Are somewhat difficult to read without being very close up and could stand to be improved.
Both Figures were changed to improve the readability.

Technical Comments

P 6434 L 4: First sentence is not complete. Assuming something like "is performed" is missing from the end.
"is performed" was inserted.

P 6441 L 8: "than" should be "as". Overall this sentence is not clear.
We changed to: "ERA40 exhibits similar RMSE values as the operational analysis from ECMWF for the overlapping period for reasons outlined above."

Page 6425 Line 26: No comma needed after "transports".
P 6429 L 21: No comma needed after "Kriging provides both"
P 6436 L 26: "are" is needed after "RMSE"
P 6439 L 11: A comma is needed after "information"
P 6440 L 13: "inner" should be "inter"
P 6442 L 4: "a" after "establish" should be removed.
All: Modified accordingly.