

The GEWEX Water Vapor Assessment: Results from Intercomparison, Trend, and Homogeneity Analysis of Total Column Water Vapor

MARC SCHRÖDER AND MAARIT LOCKHOFF

Deutscher Wetterdienst, Offenbach, Germany

JOHN M. FORSYTHE AND HEATHER Q. CRONK

Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado

THOMAS H. VONDER HAAR

Colorado State University, Fort Collins, Colorado

RALF BENNARTZ

Vanderbilt University, Nashville, Tennessee, and University of Wisconsin–Madison, Madison, Wisconsin

(Manuscript received 16 October 2015, in final form 12 April 2016)

ABSTRACT

The Global Energy and Water Cycle Exchanges project (GEWEX) water vapor assessment's (G-VAP) main objective is to analyze and explain strengths and weaknesses of satellite-based data records of water vapor through intercomparisons and comparisons with ground-based data. G-VAP results from the intercomparison of six total column water vapor (TCWV) data records are presented. Prior to the intercomparison, the data records were regridded to a common regular grid of $2^\circ \times 2^\circ$ longitude–latitude. All data records cover a common period from 1988 to 2008. The intercomparison is complemented by an analysis of trend estimates, which was applied as a tool to identify issues in the data records. It was observed that the trends over global ice-free oceans are generally different among the different data records. Most of these differences are statistically significant. Distinct spatial features are evident in maps of differences in trend estimates, which largely coincide with maxima in standard deviations from the ensemble mean. The penalized maximal F test has been applied to global ice-free ocean and selected land regional anomaly time series, revealing differences in trends to be largely caused by breakpoints in the different data records. The time, magnitude, and number of breakpoints typically differ from region to region and between data records. These breakpoints often coincide with changes in observing systems used for the different data records. The TCWV data records have also been compared with data from a radiosonde archive. For example, at Lindenberg, Germany, and at Yichang, China, such breakpoints are not observed, providing further evidence for the regional imprint of changes in the observing system.

1. Introduction

Water vapor plays a central role in Earth's energy and water cycle, making it a key variable also for climate analysis. The strong absorption properties of water vapor in the infrared part of the radiation spectrum make it the most important natural greenhouse gas. Fast-acting water vapor feedback mechanisms are critical

for understanding Earth's climate (Lacis et al. 2010). In addition, it is one constituent in the large and complex nonlinear climate system, and its interactions with other components of the climate system such as clouds and precipitation are still not fully understood. Analyzing recent decades of global water vapor distribution and variability is expected to help extend our understanding of how the climate system responds to increasing greenhouse gas concentrations.

To date, a large variety of water vapor data records is available (see, e.g., http://gewex-vap.org/?page_id=309 or <http://ecv-inventory.com>). Without proper background

Corresponding author address: Marc Schröder, Deutscher Wetterdienst, Frankfurter Str. 135, 63067 Offenbach, Germany.
E-mail: marc.schroeder@dwd.de

information and understanding of the limitations of available data records, these data may be incorrectly utilized or misinterpreted. Thus, GDAP has initiated G-VAP, the primary purpose of which is to quantify the current state of the art in water vapor products being constructed for climate applications. G-VAP intends to answer, among others, the following questions:

- 1) How large are the differences in observed temporal changes in long-term satellite data records of water vapor on global and regional scales?
- 2) Are the differences in observed temporal changes within uncertainty limits?
- 3) What is the degree of homogeneity of each long-term satellite data record? Are there any discontinuities (breakpoints)?

G-VAP considers upper-tropospheric humidity, specific humidity, and temperature profiles as well as TCWV. More details on G-VAP can be found online (<http://www.gewex-vap.org>).

In this study, we focus on G-VAP results related to TCWV. We use the acronym TCWV here to describe the mass of water vapor per unit area from Earth's surface to space. TCWV is a widely used variable in atmospheric science, but the nomenclature in the field is not standardized. TCWV is sometimes referred to as precipitable water or integrated water vapor, and by acronyms such as TPW, IPW, PWAT, and IWV. TCWV is measured in kilograms per meter squared.

Several studies have been conducted to characterize the quality of individual records or of a selection of TCWV data records by intercomparison (e.g., Divakarla et al. 2014; Schröder et al. 2013; Sohn and Smith 2003), by comparison with ground-based data (e.g., Bedka et al. 2010; Lindstrot et al. 2014; Rienecker et al. 2011), or both (e.g., Reale et al. 2012). Also, various studies have inter-compared and analyzed trend estimates using a subset of available long-term data records of TCWV (e.g., Trenberth et al. 2005; Mears et al. 2007; Mieruch et al. 2014, partly summarized in Sherwood et al. 2010). These studies focus on the ice-free ocean and have shown, among others, an increasing overall mean trend in TCWV. Results from these and other studies are often difficult to compare and interpret because different reference data records might have been used, the data records might have been subjected to different types of preprocessing, different metrics might have been considered, or different periods and/or regions were analyzed. Thus, the comparability of the results from the quality analysis and trend estimation is limited and challenging to conduct through a literature review. To our knowledge, a consistent characterization of the quality and stability of TCWV data records considering all mature and freely available data records with at least a

10-yr record has not been completed to date and is currently the focus of G-VAP as well as of this paper.

This paper presents results from the G-VAP analysis and attempts to provide answers to the questions given above. Six of the longest (longer than 25 yr) TCWV data records that were readily available form the basis for this study. These are 1) HOAPS from the EUMETSAT's CM SAF, 2) SSM/I-based TCWV merged 1° monthly water vapor product provided by REMSS, 3) NVAP-M from the NASA MEaSUREs program, and reanalysis products from 4) ECMWF, 5) NASA, and 6) NCEP: ERA-Interim, MERRA, and CFSR.

Our approach here is to use the analysis of trend estimates to identify issues in the data records. Our results should not be interpreted as supporting or otherwise ruling out any possible physical changes of TCWV associated with climate change. A side aspect of this paper is to increase the awareness that caution is needed when using these data for trend analysis and to foster improvements toward more homogeneous long-term datasets.

An overview of the various data records is given in section 2. Section 3 introduces the methodologies used for the intercomparison, the trend estimation, and the homogeneity analysis. Results are presented in section 4, first on a global ice-free ocean scale, then on selected land regions, and finally on local scale at Lindenberg, Germany, and Yichang, China. A summary and conclusions are given in section 5. Acronyms are defined in the appendix.

2. Data

A brief description of the satellite-based, reanalysis, and in situ products utilized in this study follows. An overview of the considered data records is provided in Table 1. TCWV data records are readily available as specified in Table 1.

a. Satellite-based and merged products

1) HOAPS

The TCWV product is part of the HOAPS product suite (Andersson et al. 2010) and is purely based on satellite information. TCWV from HOAPS covers the months from July 1987 to December 2008 and is defined over the global ice-free oceans. Here, monthly means on a regular $0.5^\circ \times 0.5^\circ$ longitude–latitude grid are utilized. TCWV is derived from SSM/I passive microwave radiometers on board the polar-orbiting DMSP platforms *F-08*, *F-10*, *F-11*, *F-13*, *F-14*, and *F-15*. Information on the satellites utilized per month is contained in the HOAPS products. The statistical retrieval for TCWV is described in Schlüssel and Emery (1990) and is applicable

TABLE 1. Data record overview. Resolutions are given for the data as downloaded from the institution's homepage and are not necessarily the highest possible resolution available from the record owner.

Data record name	Owner	Model scheme, version	Temporal coverage	Spatial resolution	Temporal resolution
NVAP-M	NASA	Climate	1988–2009	1°	Daily
HOAPS	EUMETSAT	Version 3.2	Jul 1987–Dec 2008	0.5°	Monthly
REMSS	Remote Sensing Systems	Version 7	1988–present	1°	Monthly
ERA-Interim	ECMWF	ECMWF IFS, version Cy31r2, 4D-Var	1979–present	1°	Monthly
MERRA	NASA	GEOS, version 5.2, GSI + IAU (3D-Var)	1979–present	0.5° × 0.667°	Monthly
CFSR	NCEP	NCEP Coupled Seasonal Forecast System, GSI	1979–2010	0.5°	Monthly

under all sky conditions, except in presence of strong scattering such as during heavy precipitation. The mean fields are computed from the swath-based data by aggregating all SSM/I pixels that have their center of the field of view falling in the respective grid box and arithmetic averaging over the month. A monthly mean is computed only if two-thirds of the days within a month contain valid observations. Further details are given in Andersson et al. (2010) and Fennig et al. (2012). HOAPS was jointly developed by the University of Hamburg and the MPI-M and has meanwhile successfully been transferred to CM SAF. The most recent release, version 3.2, is based on homogenized SSM/I observations (Fennig et al. 2013).

2) REMSS

The TCWV data record from REMSS is available over the global ice-free ocean on a regular 1° × 1° longitude–latitude grid and contains monthly means from January 1988 to the month of most recent processing. The TCWV values come from SSM/I (*F-08*, *F-10*, *F-11*, *F-13*, *F-14*, and *F-15*), the *F-16* and *F-17* SSMIS, AMSR-E (*Aqua*), and WindSat (*Coriolis*) instruments. These microwave radiometers have been carefully intercalibrated at the brightness temperature level and the version 7 (V7) ocean products have been produced using a consistent processing methodology for all sensors. Information on the satellites utilized per month is contained in the V7 products. The TCWV retrieval is described in Wentz (1997) and the data record from REMSS is made using a two-step construction process. First, monthly 1° maps from the individual satellite TCWV values are generated based on 0.25°-grid satellite products from REMSS. In a second step, quality control measures and post hoc adjustments are applied. This includes for example, the application of small constant corrections to AMSR-E (−0.2 kg m^{−2}) and WindSat (−0.05 kg m^{−2}) data prior to merging. This way the observed bias between AMSR-E/WindSat and SSM/I as well as SSMIS is removed. The combined TCWV values from

all instruments are calculated using simple averaging. TCWV values for a specific 1° × 1° grid cell are calculated only if the cell contains more than 160 observations during one month, if ice is present for only ≤30 of the cases, and if the calculated mean day of the month, derived by averaging the time of the data falling within the cell, is within 6 days of the center day of the month. The V7 products are described in Hilburn et al. (2010) and results from comparisons to TMI are shown in Wentz (2015). The TCWV data record was obtained online (<http://www.remss.com>; accessed 11 April 2013). Note that, in contrast to HOAPS and NVAP-M, the REMSS TCWV data record also includes data from SSMIS, AMSR-E, and WindSat.

3) NVAP-M

NVAP-M was created under the NASA MEaSUREs program and consists of three parallel data records for climate, weather, and ocean uses (Vonder Haar et al. 2012). The NVAP-M climate component is considered here and contains daily TCWV on a regular 1° × 1° longitude–latitude grid from January 1988 to December 2009. NVAP-M is a global, merged atmospheric water vapor data record that includes TCWV from SSM/I (*F-08*, *F-10*, *F-11*, *F-13*, *F-14*, and *F-15*) over ocean, HIRS (*NOAA-9–NOAA-12*, *NOAA-14–NOAA-17*) and AIRS (*Aqua*) water vapor products over land, and ocean and data from the IGRA. Details on the utilized retrieval schemes are given in Vonder Haar et al. (2012). The merging process performs an error-weighted average on all SSM/I, HIRS, and AIRS retrievals within a grid box for each day. Each sensor platform is assigned a variance based on comparisons with other observations or from published values, which determines the sensor's weight [see Vonder Haar et al. (2012) for more details]. Daily averages were obtained online (https://eosweb.larc.nasa.gov/project/nvap/nvap-m_table; accessed 21 November 2012). Note that, in contrast to HOAPS and REMSS, the NVAP-M TCWV data record also includes data from HIRS and AIRS over ocean.

b. Reanalysis products

Reanalyses products are generated in offline runs of advanced operational atmospheric general circulation models, including their data assimilation systems. These systems were operated with a fixed version during the production period, while data input to the system have changed over the years. Three reanalyses products that cover the modern satellite era from 1979 to the present are considered. In the following, a short overview is given for each of the reanalysis products used.

1) ERA-INTERIM

ERA-Interim (Dee et al. 2011) is a third-generation reanalysis and improves on previous versions, for example, by using a four-dimensional variational data assimilation (4D-Var) scheme for atmospheric analysis and variational bias correction. ERA-Interim is operated in 12-hourly analysis cycles using the ECMWF IFS version Cy31r2. ERA-Interim output has a native horizontal resolution of about 0.75° (T255), 60 vertical levels, and a temporal resolution of 6 h (3D fields) and 3 h (2D fields). It assimilates observations from a large variety of instruments, among them radiances from several satellites, among them ATOVS and SSM/I [see Dee et al. (2011) for more details]. The monthly means of TCWV were obtained online (<http://apps.ecmwf.int/data-records/>; accessed 9 November 2012) with a spatial resolution of 1° longitude–latitude.

2) MERRA

MERRA (Rienecker et al. 2011) was generated with version 5.2.0 of the GEOS atmospheric model and a 3D-Var assimilation scheme that is based on the GSI scheme, with a 6-hourly update cycle. Also, the GEOS-5 implements IAU to slowly adjust the model states toward the observed state. MERRA assimilates a large variety of ground-based, in situ, and satellite data, among them radiances from ATOVS and SSM/I, and applies a variational bias correction scheme. An overview of assimilated data is given in Rienecker et al. (2011) (and at <http://gmao.gsfc.nasa.gov/research/merra/input.php>). MERRA output has a native model resolution of $0.5^\circ \times 0.667^\circ$ with 72 vertical levels and a temporal resolution of up to 1 h (2D fields). The monthly means of TCWV were obtained online (<http://disc.sci.gsfc.nasa.gov/>; accessed 10 November 2012) in native spatial resolution.

3) CFSR

The CFSR (Saha et al. 2010) is a third-generation global reanalysis using the operational GSI version from 2007, which includes a 3D-Var and first-order time interpolation to the observations and variational bias correction. The atmospheric analysis system, the GFS

of 2003, is operated at a 6-hourly cycle. As with the other reanalyses, it assimilates a large variety of measurements and is the first NCEP reanalysis that assimilates satellites radiances, among them radiances from ATOVS [see Saha et al. (2010) for details on assimilated data]. The CFSR global atmosphere resolution is ~ 38 km (T382) with 64 vertical levels and is available on a 0.5° regular grid with hourly temporal resolution. Here, the monthly means of TCWV were obtained online from UCAR (<http://rda.ucar.edu/pub/cfsr.html>; accessed 21 November 2012) in 0.5° spatial resolution. Note that SSM/I-based wind speed products, but no SSM/I radiances, are assimilated.

All considered satellite and reanalysis products utilize SSM/I observations from *F-08*, *F-10*, *F-11*, *F-13*, *F-14*, and *F-15*. While the use of the data from these spacecraft is a common denominator, the data are not identical. There are differences in the sensor intercalibration, precipitation and sea ice masking, and the retrieval algorithms–assimilation systems. The common denominator data are not that common.

c. In situ observations

The HomoRS92 data record is a multistation long-term radiosonde archive and is based on IGRA (Durre et al. 2006). This data record consists of quality controlled radiosonde and pilot balloon observations at more than 1500 globally distributed stations with varying periods of record. The archive has been further improved by additional quality control, data gap filling at existing stations, and additional radiosonde data. The homogenization method described in Dai et al. (2011) has been applied and the solar radiation dry bias for Vaisala RS92 radiosonde data from 63 stations has also been corrected (Wang et al. 2013). HomoRS92 covers the period from January 1945 to December 2010 (L. Zhang 2012, unpublished data; available on request).

The data records considered in this study are a subset of the currently available data records. A decent overview of available satellite, reanalysis, ground-based, and in situ data records is provided at the G-VAP web page (http://gewex-vap.org/?page_id=309).

3. Methodology

a. Data preprocessing

All subsequent analysis is carried out on the basis of monthly means. The CFSR, ERA-Interim, HOAPS, MERRA, and REMSS data records are available as monthly means. NVAP-M contains daily averages, and the daily values within a month are arithmetically averaged using all valid observations to compute monthly

means. Prior to further processing fill values, missing values and values that are outside the validity range are assigned a unique undefined value.

To allow a straightforward intercomparison, the analysis is carried out on a common grid and time period. The common period is defined as the minimum (maximum) start (stop) time common in all data records. This leads to a common time range covering the full years 1988–2008. The common grid is defined as the minimum integer multiple applicable to all data record grids that leads to a grid resolution of $2^\circ \times 2^\circ$. To remove a shift in the spatial grid between the satellite-based products and the reanalysis products, the reanalysis grids are shifted by half a grid box. Therefore, the CFSR, ERA-Interim, and MERRA monthly means are linearly interpolated to a grid with unchanged resolution but changed center positions. Then, all six data records are arithmetically averaged onto the common grid by considering all valid observations within a grid cell.

HOAPS and REMSS are only available over the ice-free oceans, whereas the reanalysis products and NVAP-M provide global coverage. Thus, common land–sea and sea ice masks need to be applied to allow fair comparisons. The land–sea mask is computed from the GTOPO30 of the USGS (available at <https://lta.cr.usgs.gov/GTOPO30>; accessed 12 August 2008). If a single value of the high spatial resolution GTOPO30 data within a $2^\circ \times 2^\circ$ grid cell is classified as land, this grid cell is classified as land. The sea ice mask is based on the HOAPS sea ice mask (Andersson et al. 2010). Again, we apply a conservative sea ice mask, that is, the grid is classified as ice contaminated if one pixel within the grid box at any time during the common period was classified as ice covered.

Only those stations from the HomoRS92 data record were taken into account where monthly data series were available without gaps over the entire common time period, that is, from 1988 to 2008. Monthly means were calculated, if at least two profiles reaching 300 hPa were available per day on 20 days per month. These filtering criteria are fulfilled by 55 stations with the majority being located in China (see Fig. 4). Data from these stations are used for comparison.

b. (Inter)comparison and trend analysis

The intercomparison is carried out relative to the ensemble mean. The ensemble mean is calculated as arithmetic average using all valid data and all six data records. Area means are calculated using latitude-weighted averaging. The intercomparison analysis considers the absolute and relative standard deviation (relative to the ensemble mean).

The trend estimation largely follows the work of Weatherhead et al. (1998) and Mieruch et al. (2014). A

linear trend model was applied that fits the SST-based index for determining El Niño strength (available at <http://coaps.fsu.edu/jma>; accessed 6 June 2014) and four frequencies to allow an asymmetric fitting of the annual cycle, each of the four terms normalized by its standard deviation. These steps are carried out simultaneous with the trend estimation such that the results comprise the anomaly, the trend estimate, and the mean amplitude associated with the SST index and the four frequencies. The estimation of the uncertainty of the trend again follows Weatherhead et al. (1998) and considers the noise as computed from the anomaly and autocorrelation. The ratio of the trend and uncertainty estimates is subjected to a two-sided *t* test to compute the coverage probability. If the coverage probability exceeds 95%, the trend estimate is considered to be significantly different from the null hypothesis of a trend of $0 \text{ kg m}^{-2} \text{ decade}^{-1}$; that is, the level of significance is 0.05.

To put the trend estimates into a physical perspective we also computed the regression of TCWV in percent and SST in kelvins using the NOAA Optimal Interpolation SST, version 2 (Reynolds et al. 2002). Prior to the regression, the SST and TCWV anomaly time series are smoothed with a 12-month low-pass filter as in Mears et al. (2007). Following, for example, Hyland and Wexler (1983), the change in saturation vapor pressure as function of air temperature and of air temperature change can be computed and, assuming constant relative humidity and pressure, be transferred into a change in mixing ratio. For a temperature change of 1 K the expected change in mixing ratio is between 6% at 300 K and 7.5% at 275 K. We consider SST and TCWV instead of near-surface air temperature and mixing ratio. Both might lead to an amplification of the regression (Trenberth et al. 2005). Here we consider the regression using global ice-free ocean values within 60°N – 60°S . Advection associated with low pressure systems in the extratropics can lead to observed regression values larger than the maximum given above. Note that the regression will also be affected by the quality of the utilized SST data record (see the Group for High Resolution SST for a discussion of the quality of SST data; www.ghrsst.org). Other factors that can affect the correlation between TCWV and SST are described in, for example, Mieruch et al. (2014).

c. Homogeneity analysis

The aim of the homogeneity analysis is to detect abrupt (artificial) shifts in time series, that is, a discontinuity in the time series of, in our case, TCWV not due to geophysical causes. Here these are called breakpoints, which might arise because of a variety of causes including sensor channelization and calibration, algorithm changes or

changes in their ancillary data, and changes in spatial sampling coverage. Here the penalized maximal F test (Wang 2008a,b) is utilized to detect breakpoints because it can be applied to time series of (deseasonalized) anomalies and of anomaly differences and because it does not require supervision. The PMF test assumes that the time series is not affected by any sudden change in the linear trend (Wang 2008a). Wang (2008a) assessed the accuracy of the PMF test relative to the accuracy of a two-phase regression model and found superior quality in particular also for the rate of detecting the breakpoint position correctly. Note that the uncertainty in the estimation of the time of a break increases when the internal variance is larger than the variance introduced by the magnitude of the breakpoint (Lindau and Venema 2016). Here breakpoints are considered if the associated level of significance is 0.05 or smaller. Then, the null hypothesis of a break-free time series needs to be rejected. Note that the PMF test is applied to identify potential breakpoints in the data records and not to homogenize the data records. An overview of available homogenization schemes, among them the PMF test, and a discussion of the quality of the homogenization is given in Venema et al. (2012).

For each breakpoint detected, the PMF test returns the step size that corresponds to the size of the shift in the model fitted to the time series at the breakpoint. We also calculate the step size relative to the variability, further called break size. The variability used to get the break size is calculated as the mean standard deviation based on 6-month time segments before and after the detected breakpoint (i.e., in total 12 months). In some cases, two breakpoints occur so close to each other that the time span for calculating the variability is less than 6 months. Then the variability was still calculated, if at least one of the two time segments had the minimum length of 6 months.

For assessing the homogeneity of the six datasets, the following time series are computed and analyzed as outlined above: 1) anomaly differences relative to HOAPS (ocean) and ERA-Interim (land) and 2) anomaly differences relative to data at selected stations of HomoRS92. ERA-Interim has been chosen as reference because it exhibits the least suspicious features in terms of seeming breakpoints and intensity of maxima and minima in Hovmöller diagrams of TCWV anomalies (not shown). Similar conclusions can be drawn for HOAPS and REMSS, which both have the additional advantage of relying on SSM/I observations only and which both utilize homogenized SSM/I radiance data. Here, HOAPS has been chosen. Note that seeming breakpoints in ERA-Interim or HOAPS will also be visible in the anomaly differences and that the chosen reference will affect the time, the number, and the step size.

Last, we compare the breakpoint positions with changes in the observing systems and changes in the assimilation scheme of the considered reanalyses. Temporal coincidence is defined here if the difference between both events is within ± 3 months, as in the accuracy analysis in Wang (2008a).

4. Results

We first show results from the analysis of global maps of trend estimates and homogeneity tests using anomaly differences over global ice-free oceans for the six data records. This way the degree of consistency and the reasons for observed differences in trends are analyzed and explained (section 4a). Then, differences in absolute trend estimates are compared to the standard deviation of differences to the ensemble mean to find regions of maxima common and uncommon in both analyses (section 4b). Anomaly differences and heat maps are presented based on data from three distinct regions evident in results from both analyses. Last, results from the comparisons with HomoRS92 and associated homogeneity tests are shown (section 4c).

a. Global analysis

Figure 1 shows global maps of TCWV trend estimates ($\text{kg m}^{-2} \text{yr}^{-1}$) for all six data records. Overall the trend patterns are quite similar among the data records, in particular over the ocean, with generally large positive trends in the west Pacific and smaller areas of negative trends in the east Pacific within two bands north and south of the equator. Here the imprint of El Niño–Southern Oscillation is seen. In Fig. 1 it seems that the dominating factor for significance is the magnitude of the trend, with lowest number of grids with significant trends in ERA-Interim and maximum number of significant trends in CFSR.

Although the trend patterns are similar, the associated regional averages of trends differ in magnitude. The two passive microwave only products, HOAPS and REMSS, agree reasonably well in terms of trend magnitudes. Both records heavily rely on SSM/I observations and can thus be expected to be similar. In contrast to these SSM/I-based data records, the reanalysis datasets show larger differences with ERA-Interim (CFSR) showing smallest (largest) average trends (over ocean) among the reanalysis records and in general. When looking closer at Fig. 1 the impression of a general bias in trend estimates is actually dominated by contributions from the oceans, in particular in the tropics. Thus, for each data record the trend has been estimated for the global ice-free ocean within 60°N – 60°S , and the results are given in Table 2 together with associated uncertainties.

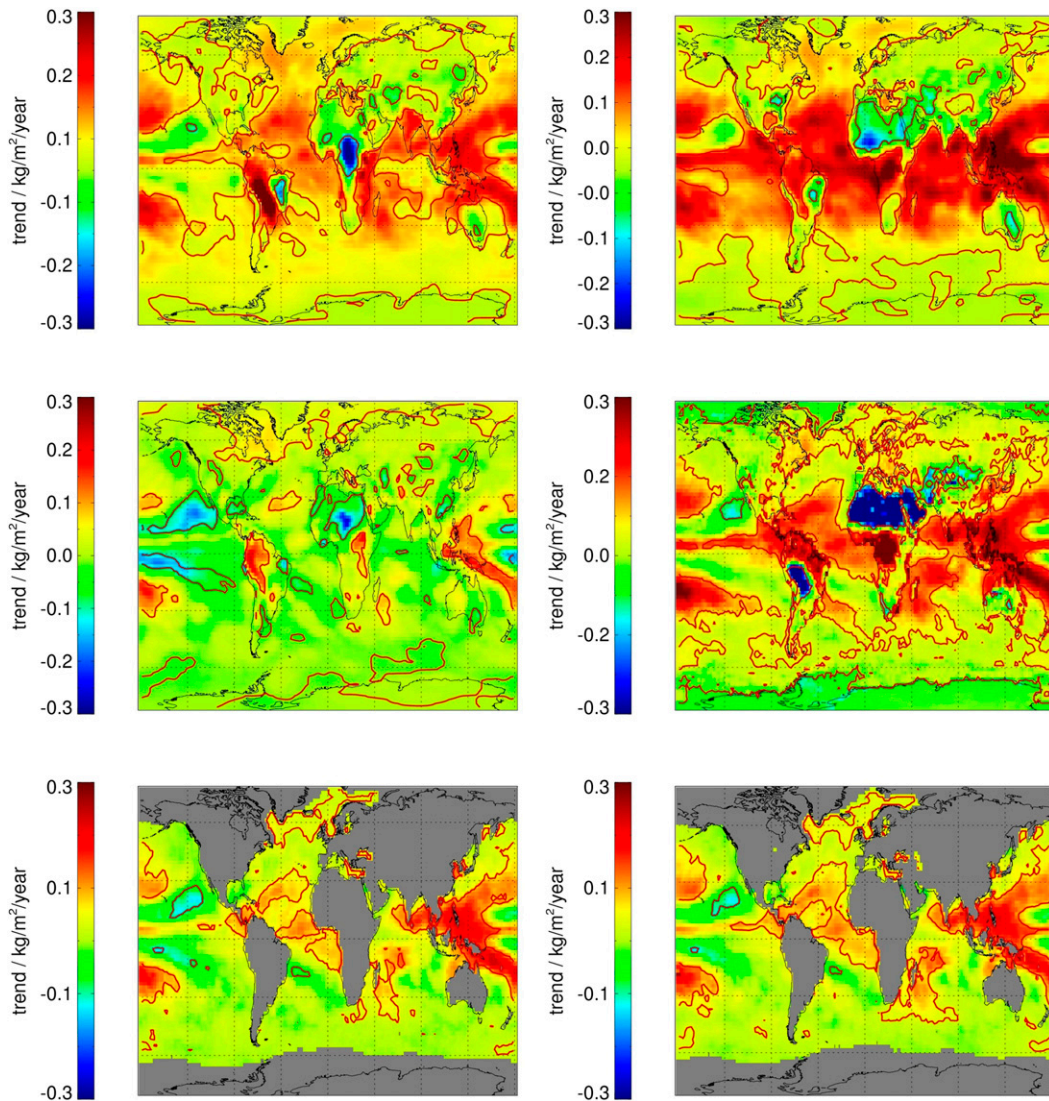


FIG. 1. Trend estimates in TCWV ($\text{kg m}^{-2} \text{yr}^{-1}$) for the six different data records: (top left) MERRA, (top right) CFSR, (middle left) ERA-Interim, (middle right) NVAP-M, (bottom left) HOAPS, and (bottom right) REMSS. Red contours indicate the 95% coverage probability. Data from HOAPS and REMSS are defined over ice-free oceans.

The trend estimates in Table 2 are sorted in ascending order and confirm the impression based on results shown in Fig. 1 that there are large differences in the trends over the ocean. The values range from $-0.11 \text{ kg m}^{-2} \text{decade}^{-1}$ (ERA-Interim) to $1.21 \text{ kg m}^{-2} \text{decade}^{-1}$ (CFSR). Associated uncertainties are typically $\geq 0.06 \text{ kg m}^{-2} \text{decade}^{-1}$ (REMSS) and $\leq 0.20 \text{ kg m}^{-2} \text{decade}^{-1}$ (NVAP-M). Following Mieruch et al. (2014) and ignoring the covariance between trend estimates, the trends given in Table 2 are significantly different, except for HOAPS and REMSS as well as for NVAP-M and MERRA. Sherwood et al. (2010) provide a table of trend estimates based on literature values on data of various coverage and temporal

length. Here, we confirm their conclusion that the “precise climate dependence of water vapor” cannot yet be established “from observed trends” by using data records on common grid and common period. The regression coefficients are also provided in Table 2 to comment on the physical soundness. The order and the (dis)agreement are similar to the results for the trends. It seems that, except for HOAPS and REMSS, the regression values are outside the expected range of $6\% \text{ K}^{-1}$ at 300 K and $7.5\% \text{ K}^{-1}$ at 275 K (see section 3b). When using the global ice-free average SST, the theoretical expected value is $6.2\% \text{ K}^{-1}$, which is significantly different from all estimates given in Table 2. The latter conclusion is still

TABLE 2. Results from the estimation of trends over the global ice-free ocean within 60°N–60°S. The values are sorted in ascending trend order. Recall that the use of the trend analysis is to locate breakpoints and it is not claimed that these are the trends occurring in nature.

Data record	Trend and uncertainty [$\text{kg m}^{-2} (10 \text{ yr})^{-1}$]	Regression and uncertainty (% K^{-1})
ERA-Interim	-0.11 ± 0.09	2.9 ± 0.5
HOAPS	0.25 ± 0.07	7.2 ± 0.3
REMSS	0.34 ± 0.06	7.6 ± 0.3
NVAP-M	0.68 ± 0.20	8.4 ± 0.7
MERRA	0.75 ± 0.09	15.8 ± 0.3
CFSR	1.21 ± 0.16	24.9 ± 0.5

valid when the analysis is carried out for the tropical ocean (not shown). Note that the observed regression can be larger than expected because of the impact of, for example, advection and tropospheric amplification of surface warming (e.g., [Santer et al. 2005](#)). We conclude that the analysis of regression is a supportive and valuable approach for the interpretation of trend estimates. Results from an analysis of regression alone need to be considered with caution because of the assumptions that enter the expectation.

The large differences in trends bring up the question of how these can be explained. To study this in more detail, anomaly differences for the global ice-free ocean are analyzed using the PMF test. The results are shown in [Fig. 2](#). Note the difference in the range at the y axis among the panels and keep in mind that HOAPS also can be affected by breakpoints. All five anomaly differences exhibit statistically significant breakpoints of various step sizes. The largest (smallest) step sizes are found for NVAP-M and CFSR (REMSS). The moistening in the mid-1990s in NVAP-M is also observed over the tropical ocean and tropical land areas as shown in [Vonder Haar et al. \(2012\)](#). ERA-Interim exhibits the largest number of breakpoints. This can partly be explained by the small variance, even minimal variance among the considered data records in the early 2000s, and its impact on significance estimation. Obviously the breakpoints largely explain the observed differences in trends, in particular the large negative step size for ERA-Interim in December 1991 and the large positive step size for CFSR in October 1998.

We analyzed the temporal occurrence of changes in the observing system and changes of the input to assimilation schemes with observed breakpoints by screening the data record references given in [section 2](#) and the launch dates of relevant satellites. The results are summarized in [Table 3](#). It seems that all breakpoints can be explained with changes in the observing system. The maximum step size in CFSR coincides with a start of

assimilation of data from *NOAA-15* [see [Saha et al. \(2010\)](#) for an overview on assimilated radiances]. CFSR and MERRA both assimilate the low-frequency channels of AMSU-A but ERA-Interim does not ([Rienecker et al. 2011](#)). The breakpoint in ERA-Interim in late 1991 coincides with a change of assimilating *F-10* instead of *F-08* data and has also been observed in [Schröder et al. \(2013\)](#) using HOAPS, version 3.1, as reference. This breakpoint is also evident in ERA-Interim precipitation anomalies, and even the series of downward and upward breaks in [Fig. 2](#) exhibits similarities to the precipitation anomalies shown in [Dee et al. \(2011\)](#). The breakpoint in July 2006 is observed in the anomaly differences of ERA-Interim and REMSS but not for NVAP-M, MERRA, and CFSR. The noise level of the anomaly differences of MERRA, CFSR, and NVAP-M might obscure the presence of a breakpoint in July 2006. However, the breakpoint in July 2006 coincides with the activation of a radar calibration beacon on *F-15*. HOAPS and ERA-Interim do not use SSM/I data from *F-15* after July 2006, while REMSS includes beacon-corrected data from *F-15* after July 2006 ([Hilburn and Wentz 2008](#)).

One breakpoint cannot be explained with a change in the observing system: a breakpoint in REMSS in April 1993. Here an increase in TCWV anomaly difference between approximately 1992 and 1995 might be misinterpreted as breakpoint. During this period the HOAPS and REMSS data records use data from *F-10* and *F-11*. Observations from both satellites are affected by orbital drift. In addition, *F-10* exhibits strong annual Earth incidence angle variations ([Berg et al. 2013](#); [Fennig et al. 2013](#)). Different treatment in calibration and intercalibration between processing at RSS and CM SAF might explain the observed feature.

The information on the various events is partly taken from figures published in the literature. It would be helpful to support data record users and assessments if data providers publish a list of input data that finds its way into the final product together with main technical specifications such as start and stop dates and number of observations per instrument and month.

b. Regional analysis

We start this analysis by showing results from the intercomparison and comparison of trend estimates to identify regions of pronounced maxima in differences among the data records. In [Fig. 3](#) the standard deviation and the relative standard deviation of differences relative to the ensemble mean are shown. Both figures show a contrast between land and ocean, with ocean values on average being smaller than land values. Regional exceptions are the relatively large values at the

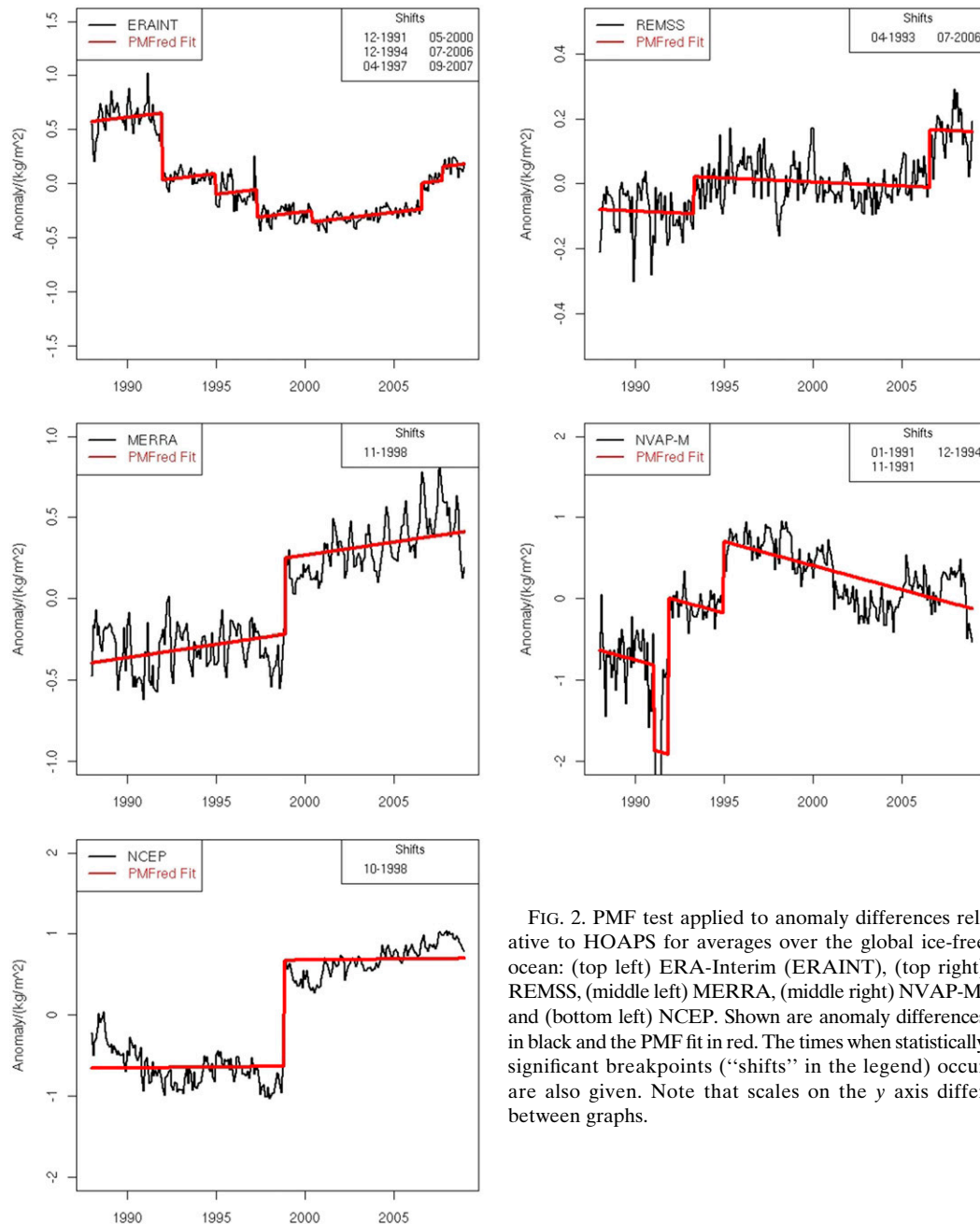


FIG. 2. PMF test applied to anomaly differences relative to HOAPS for averages over the global ice-free ocean: (top left) ERA-Interim (ERAINT), (top right) REMSS, (middle left) MERRA, (middle right) NVAP-M, and (bottom left) NCEP. Shown are anomaly differences in black and the PMF fit in red. The times when statistically significant breakpoints (“shifts” in the legend) occur are also given. Note that scales on the y axis differ between graphs.

ITCZ. Relative values are typically larger at higher latitudes and in elevated terrain, which correlates with generally small TCWV values. The following regions exhibit maxima in standard deviation: ITZC, tropical rain forest over South America and central Africa, deserts like the Sahara, and mountainous regions in combination with a general wet environment like the Andes. Additionally, the following regional maxima in relative standard deviation are found: Arctic and Antarctic and mountainous regions in the high latitudes.

The local maxima at mountainous regions can to some extent be explained by the various native resolutions of the individual data records as well as the by interpolation and averaging processes. It is further likely that sampling differences—for example, by cloud or rain screening (Sohn et al. 2006; Schröder et al. 2013) and different handlings of surface emissivities—contribute to the observed differences, in particular over tropical land surfaces. Such aspects are not further discussed here and are partly subject to future efforts within G-VAP.

TABLE 3. Dates of observed breakpoints and coincident changes in the observing system or changes of the input to the assimilation schemes based on the analysis of anomaly differences relative to HOAPS for the global ice-free ocean. When launch dates are provided, the dates match their consideration in the assimilation schemes within ± 3 months.

Date	Break size (kg m^{-2})	Dataset	Event
Jan 1991	-1.05	NVAP-M	Launch <i>F-10</i> : Dec 1990
Nov 1991	1.92	NVAP-M	Launch <i>F-11</i> : Dec 1991 Stop date <i>F-08</i> : Dec 1991
Dec 1991	-0.62	ERA-Interim	See Nov 1991
Apr 1993	0.11	REMSS	See text
Dec 1994	-0.19	ERA-Interim	Launch of <i>NOAA-14</i> : Dec 1994, approx stop of assimilation of <i>NOAA-11</i> data (see Dee et al. 2011)
Apr 1997	0.88	NVAP-M	
	-0.26	ERA-Interim	Approx change from assimilation of data from <i>NOAA-12</i> to <i>NOAA-11</i> (see Dee et al. 2011)
Oct 1998	1.31	CFSR	Begin of assimilation of <i>NOAA-15</i> data in Oct 1998 (Chelliah et al. 2011); approx end of assimilation of <i>NOAA-11</i> and <i>NOAA-14</i> data; change from assimilation of data from <i>GOES-9</i> to <i>GOES-10</i> (Saha et al. 2010)
Nov 1998	0.47	MERRA	Start of assimilation of <i>NOAA-15</i> data (Rienecker et al. 2011)
May 2000	-0.10	ERA-Interim	Approx start of assimilation of <i>F-15</i> data and end of <i>NOAA-11</i> and <i>NOAA-15</i> data (see Dee et al. 2011)
Jul 2006	0.24	ERA-Interim	Close to end of assimilation of <i>F-15</i> data, close to change from <i>GOES-10</i> to <i>GOES-11</i> , start of <i>Meteosat-5</i> and <i>Meteosat-8</i> , approx end of assimilation of <i>NOAA-14</i> data (see Dee et al. 2011)
	0.18	REMSS	See text
Sep 2007	0.13	ERA-Interim	Approx end of assimilation of <i>NOAA-16</i> data (see Dee et al. 2011)

In addition, the mean absolute difference in trend estimates and the number of valid observations are shown in [Fig. 4](#). In general, the differences are small in high latitudes and over the ocean and larger over land and in the ITZC. Occasionally, a few grid values stick out. This can be explained with the small absolute number of cases in combination with minima in the number of valid observations. Most striking are the pronounced maxima over the following land areas: central Africa, the Sahara, and (central) South America. Looking at [Fig. 1](#) reveals that over central Africa trend patterns are generally different and even exhibit opposite trends; over the Sahara all data records show negative trends, with NVAP-M having the maximum negative trend covering the entire Sahara region and Saudi Arabia. The trends from re-analyses data records agree in sign over South America, with a dipole of positive and negative values, with strongest maxima and minima in MERRA. NVAP-M exhibits the opposite pattern with a similar absolute magnitude as MERRA. These three distinct regions of maxima largely overlap in [Figs. 3](#) and [4](#).

In the following, we try to identify explanations for the observed maxima. To study this in more detail, we looked at regional anomaly differences for the areas with the most pronounced differences: central Africa, the Sahara, and South America, which are marked by black/red boxes in [Fig. 4](#). Regional TCWV anomaly differences relative to ERA-Interim are shown as black

lines in [Fig. 5](#) for data records with valid data over land, that is, CFSR, MERRA, and NVAP-M. Also shown is the model fitted by the PMF test (red line) and the time of significant breakpoints. The range of the y axis is variable. Note that ERA-Interim is likely affected by breakpoints as well, which would appear as breakpoints in all three comparisons. The time, the number, and the step size of the breakpoints are a function of data record and also differ from region to region. The maximum number of breakpoints is found for NVAP-M over central Africa and the largest step size for NVAP-M over the Sahara. Comparing the step direction and the sign of the trends in the regions of central Africa and the Sahara (see [Fig. 1](#)), it can be concluded that breakpoints are responsible for the observed trend patterns and thus also for maxima in mean absolute trend differences. An exception is found for CFSR over the Sahara, where the step size has opposite sign to the trend. Results for South America are specific because the step sizes in CFSR and in NVAP-M have opposite sign to the trend estimates and MERRA does not exhibit a statistically significant breakpoint but has the strongest trend in this region. In view of these results, the difference in sign of the trends, and the dipole patterns of the trends, a more in-depth analysis of long-time-series water vapor data records over central South America is needed.

The time of occurrence and the break size of breakpoints as well as changes in the observing systems are

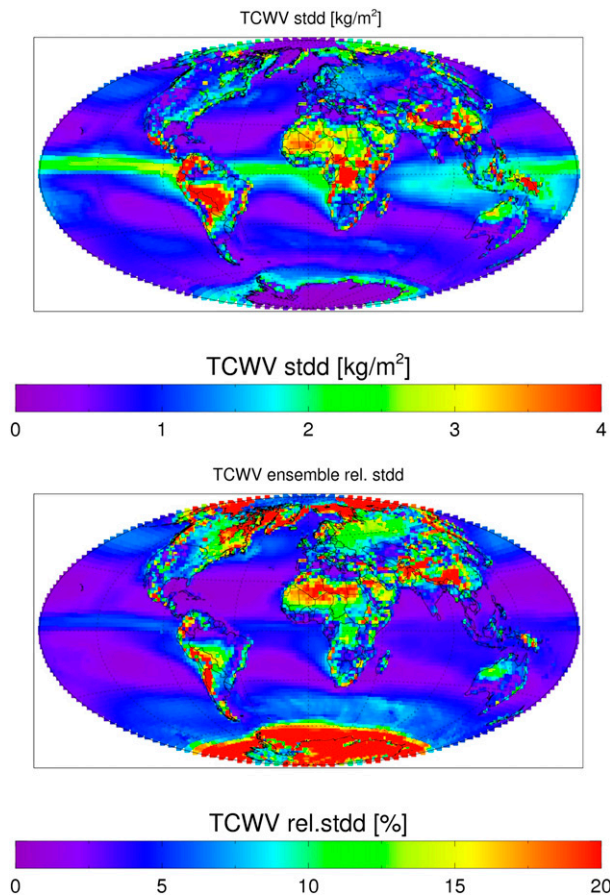


FIG. 3. (top) Absolute and (bottom) relative standard deviation between the six long-term data records and the ensemble mean.

shown as a heat map in Fig. 6. Results are based on anomalies for ERA-Interim and for anomaly differences relative to ERA-Interim for CFSR, MERRA, and NVAP-M for each of the three regions. Most of the detected breakpoints can be matched to changes in the observing systems or changes of the input to assimilation schemes. The following breakpoints remain unexplained: 1) central Africa: September 1992/NVAP-M, December 2001/MERRA, February 2006/NVAP-M; and 2) the Sahara: November 1993/NVAP-M. Changes in the assimilation of ERA-Interim are likely not causing these breakpoints because they would need to be evident in all three anomaly differences—this is not the case here. The supplemental material of Vonder Haar et al. (2012) includes a figure of the area covered by instrument that enters NVAP-M. The results for HIRS exhibit a small decrease in number and a relatively large change in variability between 1993 and 1994. It seems that this break is associated with the processing of HIRS data in NVAP-M. The channel central wavelength for

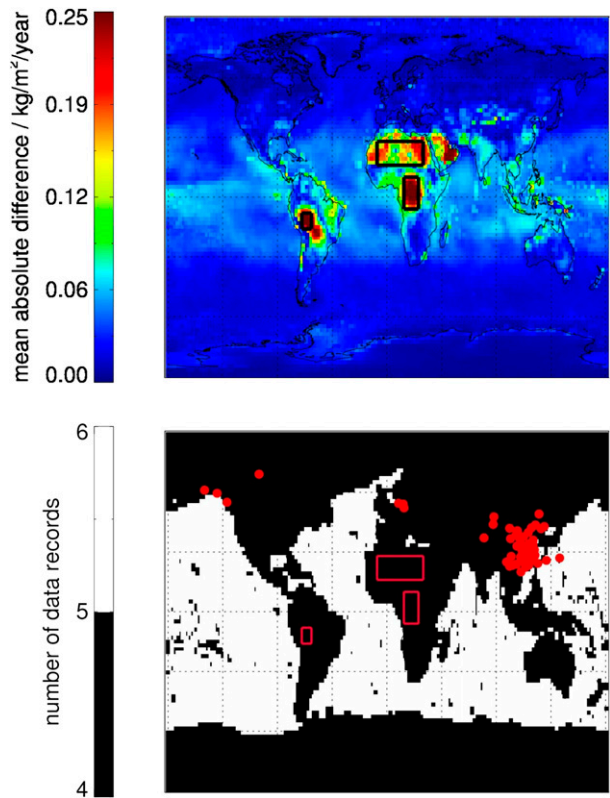


FIG. 4. (top) Mean absolute difference of trend estimates using the six data records. (bottom) The number of data records, either 4 or 6. Areas marked in black/red rectangles exhibit distinct differences and overlap with areas of distinct differences shown in Fig. 3 (central Africa, the Sahara, and South America). Red circles mark the position of stations from the HomoRS92 (see text for details).

one of the relevant HIRS water vapor channels was changed from $8.6\ \mu\text{m}$ on NOAA-12 to $12.5\ \mu\text{m}$ on NOAA-11 and NOAA-14. Also, in NVAP-M NOAA-12 sampling actually ended in December 1993. The mix of these two HIRS versions in 1993–94 implies these are years of increased uncertainty for NVAP-M. The number of observations entering the reanalyses is dominated by satellite observations over the period 1988–2008. Even though the spatial density is significantly smaller in ground-based and in situ measurement relative to satellite data, single stations can have a larger-scale impact on the product. Bosilovich et al. (2011) observed a significant hydrological anomaly in central Africa in late 1995 that could be explained by the impact of a single station. In particular, the impact was also evident in specific humidity at 850 hPa. As in Fig. 5 the noise level is relatively large (see Fig. 15b in Bosilovich et al. 2011). There results seem to be consistent with the decrease of TCWV between late 1995 and the early 2000s in the MERRA anomaly difference time series (Fig. 5). The

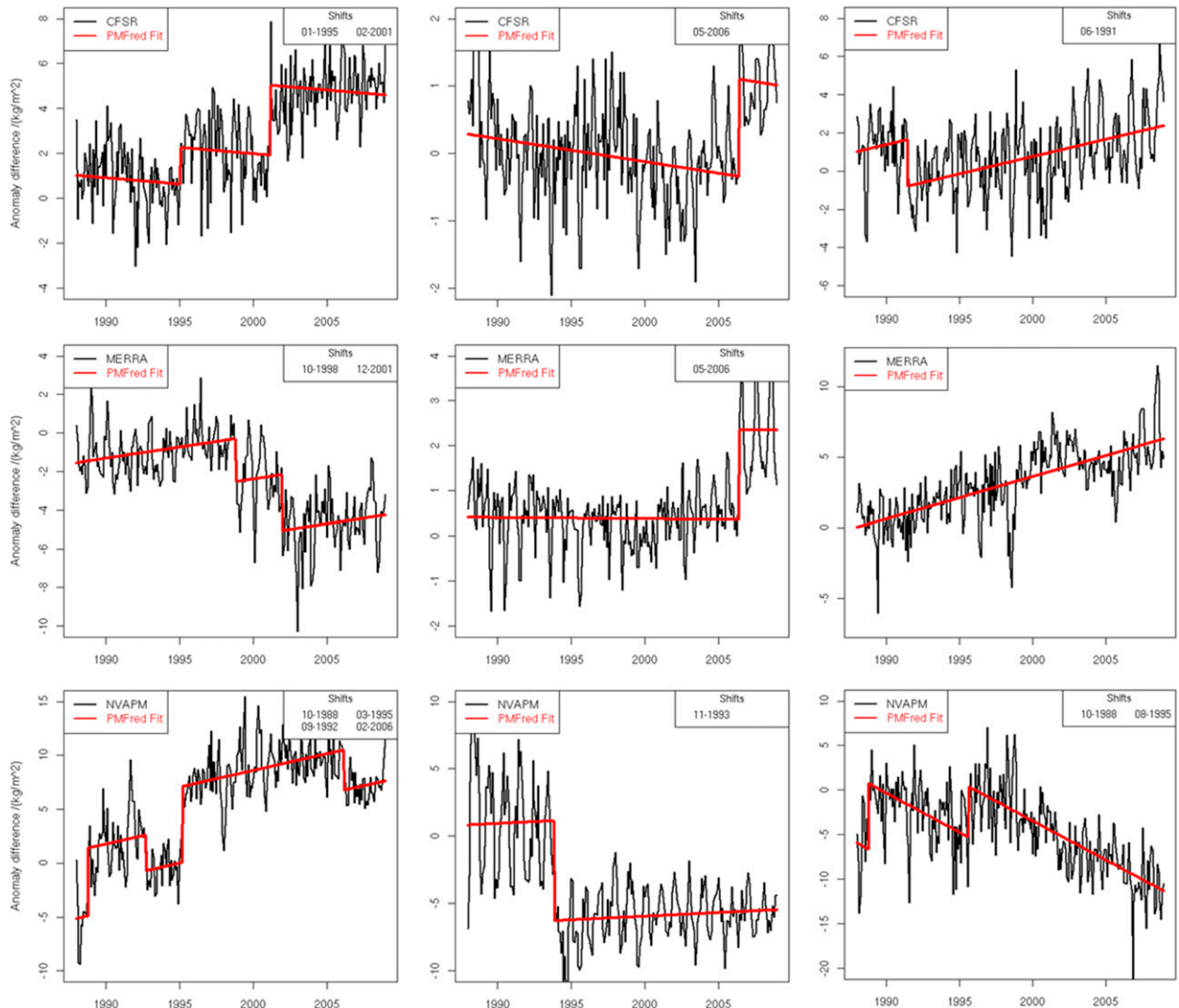


FIG. 5. As in Fig. 2, but based on anomaly differences for the regions (left) central Africa, (center) the Sahara, and (right) South America using ERA-Interim as reference, for (top) CFSR, (middle) MERRA, and (bottom) NVAP-M. Note that scales on the y axis differ between graphs.

relatively large noise level in combination with a seeming decrease in TCWV might obscure the series of two breakpoints in late 1998 and late 2001. More efforts are needed to find reasons for the unexplained breakpoints. In a first step, other homogeneity tests will be applied to confirm or disprove these observed breaks. Note that there are unexplored linkages to global or regional oscillations that may possibly manifest themselves as breakpoints.

It seems that the breakpoints explain not only the differences in trend estimates but to some extent and for the considered regions also the maxima in standard deviation relative to the ensemble mean. To give a rough estimate based on the step size given in Fig. 6 for the Sahara and

NVAP-M, we get an artificial standard deviation of more than 3 kg m^{-2} when assuming no trend, white noise everywhere except at breakpoint, and the difference to the mean of the anomaly difference, that is, 0 kg m^{-2} .

We summarize that the impact on the product quality of a change in the observing system seems to be a function of space and data record. In certain areas the quality of the retrievals and assimilation schemes are not significantly affected, while in other regions (e.g., through the handling of surface properties) the change in the observing systems manifests itself in the product quality. We conclude that breakpoints caused by changes in the observing system largely explain differences in TCWV and associated trends.

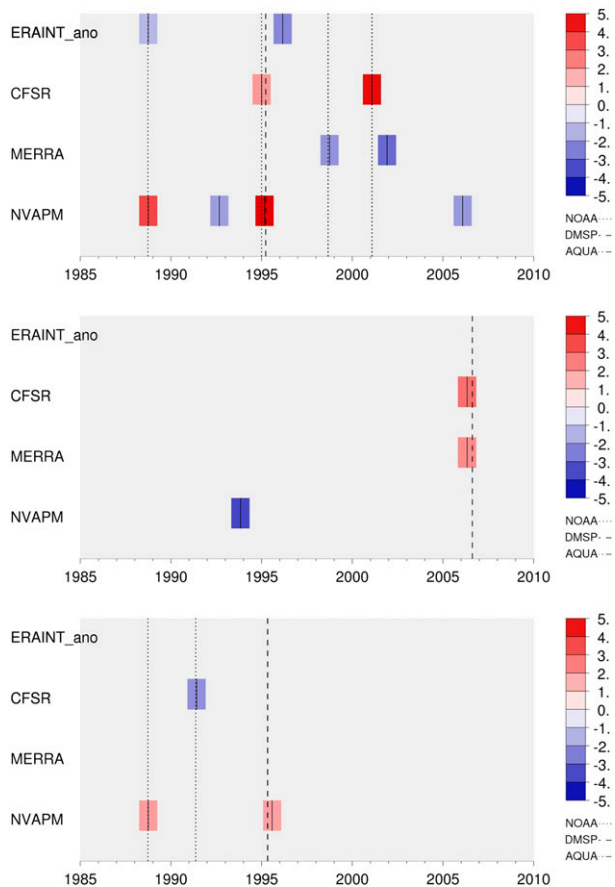


FIG. 6. Occurrences and break size of breakpoints detected by the PMF test for (top) central Africa, (middle) the Sahara, and (bottom) South America. Vertical lines indicate a match between changes in observing systems and detected breakpoints. Anomaly differences are considered using ERA-Interim as reference. Results for ERA-Interim are based on the ERA-Interim anomaly time series (labeled ERAINT_ano).

c. Comparison with HomoRS92

The above results are based on intercomparison results of data records, which largely rely on satellite observations, in particular observations from SSM/I. We extend this intercomparison through a comparison with data from HomoRS92. For the comparison with HomoRS92, TCWV values were extracted from the gridded data records for the grid box closest to the respective station. Exemplary comparisons of TCWV anomalies and anomaly differences relative to HomoRS92 for two GUAN sites, Lindenberg and Yichang, are shown in Figs. 7 and 8, respectively. For Lindenberg, all data records show a small bias close to 0 kg m^{-2} after 1994. The consistency in anomalies is generally large, except for generally increased variances prior to 1994 and a relatively large anomaly in 1992. At Yichang the discrepancies are larger and show a clear seasonal dependency, with the largest absolute values in boreal

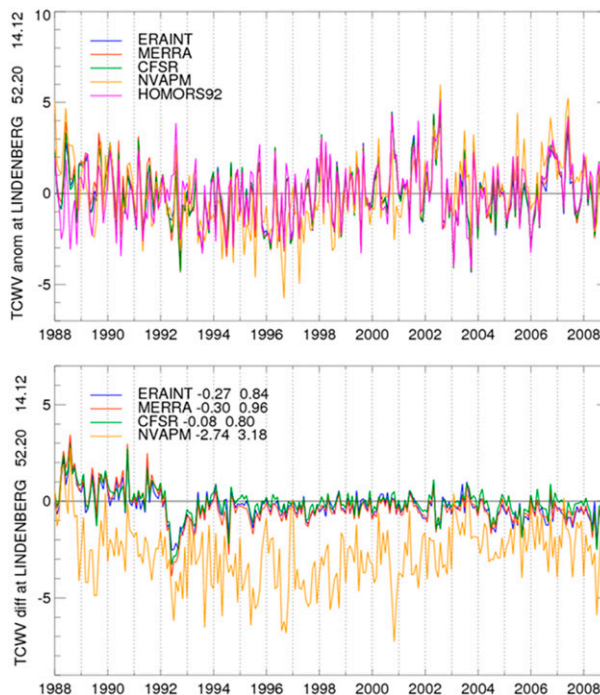


FIG. 7. (top) TCWV anomaly plots (kg m^{-2}) for ERA-Interim, MERRA, CFSR, NVAP-M, and HomoRS92 for Lindenberg. (bottom) Anomaly differences using HomoRS92 as a reference (data record minus HomoRS92); the numbers next to the dataset names indicate bias and RMS, respectively.

summer and mean bias values up to 3.7 kg m^{-2} for the re-analysis datasets. An exception in both cases is NVAP-M, which exhibits larger variance and a larger bias, occasionally exceeding 5 kg m^{-2} for Lindenberg and 20 kg m^{-2} for Yichang. The breakpoints in October/November 1998 in MERRA and CFSR anomaly differences over the global ice-free ocean (Fig. 5) are not evident in Figs. 7 and 8. Even the strong breakpoints in NVAP-M over the Sahara and central Africa in November 1993 and March 1995, respectively, are not evident here. These results again support the conclusion that the imprint of changes in the observing systems is a function of region.

For Lindenberg, NVAP-M and all other data records show a distinct anomaly in summer 1992. This might be linked to the volcanic eruption of Pinatubo in June 1991. Robock (2002) shows that the Pinatubo eruption caused temperature anomalies on global scale: a decrease in low-tropospheric temperatures in boreal summer in 1992 and an increase in boreal winter temperatures in 1991/92 and 1992/93. On a global scale, the associated phase shifts in mean temperature and in TCWV agree, as shown by Soden et al. (2002). On a more regional scale, the negative anomaly relative to data from Lindenberg coincides with a regional maximum in temperature anomalies over central Europe in summer 1992 (Robock and Oppenheimer 2003).

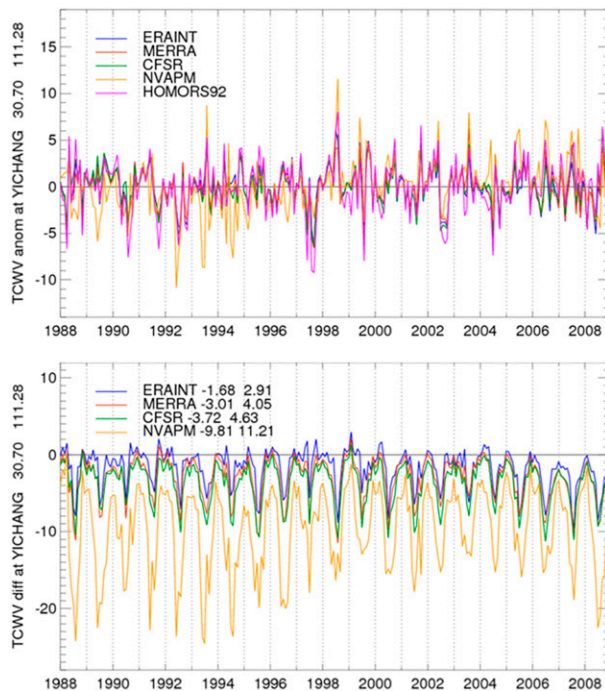


FIG. 8. As in Fig. 7, but for Yichang.

The imprint of the eruption on the anomaly differences appears to explain the observed features in Fig. 7. This feature requires further analysis, and a first step can be an analysis of the correlation between these anomaly differences and a climate index of volcanic activity.

Distinct anomaly features for the site of Yichang (Fig. 8, top) include the pronounced negative anomalies during boreal summer in the years 1997 and 2002 visible in all datasets. As both years are El Niño years, these anomalies might be linked to the weakening of the Asian summer monsoon observed during El Niño events (e.g., Ju and Slingo 1995).

We also applied the homogeneity test to the time series of anomaly differences between HomoRS92 and NVAP-M for a series of stations and analyzed heat maps similar to Fig. 6 but with stations on the y axis (not shown). Assuming a consistent degree of homogeneity among time series from the different stations of HomoRS92, one would expect that in the presence of a breakpoint in NVAP-M this will be evident in a series of consistently detected breakpoints. We found that the breakpoints are generally scattered in time and among stations and, for example, the strong breakpoint over the Sahara in 1995 is not observed at any of the considered stations. This might be explained by breakpoints being a function of region and/or by a lack of temporal and spatial homogeneity among the data from the different stations in HomoRS92.

When comparing HomoRS92 station data that fulfill the sampling criteria (not shown), most regional issues cannot be observed because the regions of maxima in standard deviation (see Fig. 3) and maxima in differences of trend estimates (see Fig. 4) are not covered by a station from HomoRS92. Establishing a station dedicated to climate monitoring and satellite evaluation in such regions can be highly beneficial for future evaluations of satellite observations.

5. Conclusions

To our knowledge, a consistent characterization of the quality and stability of TCWV data records considering all mature and freely available data records with at least a 10-yr record has not been completed to date and is currently the focus of G-VAP. In this paper, three satellite-based TCWV data records and TCWV output from three reanalyses were transferred to a common $2^\circ \times 2^\circ$ longitude–latitude grid and a common period ranging from January 1988 to December 2008. The corresponding monthly means were inter-compared to characterize the differences among these data records on global and regional scales. The comparison of trend estimates was used as a tool to identify issues in the data and to complement the inter-comparison study and was not meant as an attempt to characterize climate change. Output from the inter-comparison and from the trend estimation were analyzed to identify regions of distinct differences and differences in breakpoints. Several regions of maxima in standard deviation and in differences in trend estimates were identified. Homogeneity tests were applied to anomalies and anomaly differences based on data from such regions to find explanations for the observed distinct regional features. For the same reason, comparisons with a long-term multistation radiosonde archive, here HomoRS92, were carried out.

On the basis of consistently applied tools, major differences in state-of-the-art CDRs were identified, documented, and to a large extent explained. This will allow data record providers in future updates to easily assess their data record's improved stability given the results presented here. Also, the science questions given in the introduction were largely answered. The results and the answers are summarized as follows:

- On a global ice-free ocean scale, the trend estimates among the different data records were found to be significantly different between the different data records, with the exception of just two pairs: HOAPS/REMSS and NVAP-M/MERRA.
- Except for HOAPS and REMSS, all data records exhibit regression values outside the theoretically

expected range. This is an indication of issues in long-term stability.

- Regions with distinct differences in standard deviation from the ensemble mean largely coincide with the mean absolute difference of the trend estimates. The regions with the most pronounced differences are central Africa, the Sahara, and South America.
- The differences in trend estimates in these regions and on global ice-free ocean scale were found to be caused by breakpoints or series of breakpoints. In most cases these breakpoints coincide temporally with changes in the observing system.
- The time, sign, and step size of breakpoints are typically a function of region and data record. In particular, the breakpoint characteristics are different between time series from the regional and the global ice-free ocean scale. The imprint of changes in the observing systems is a function of region.
- The majority of these breakpoints are not evident when comparisons with the HomoRS92 data record were carried out. One reason is that areas with distinct differences in trend estimates are not covered with stations.

Data record users and assessments would benefit from a list of input data that enters the final product together with main technical specifications such as start and stop dates and number of observations per instrument and month.

Lessons learned about regional changes have provided guidance for future improvements of data records. One of the major advantages of an effort like G-VAP is to suggest and encourage improvements to data records included in the G-VAP analysis. Discussions between G-VAP participants over the last two years have allowed participants to receive new perspective on their work. Analysis of data records by outside, independent scientists willing to provide critical feedback is of great benefit. For example, the discovery of the regional breakpoints in NVAP-M over the Sahara by G-VAP members has prompted further investigation by the NVAP-M team into the challenges of using infrared data over a surface with variable emissivity and a variable atmosphere that is often impacted by dust storms. These factors will be addressed in future production efforts. Our results generally confirm the conclusions in [Rienecker et al. \(2011, p. 3643\)](#) that the differentiation between the impacts from changes of the observing system and climate variations “pose perhaps the greatest challenge for the next generation of reanalyses.” Here this conclusion is extended from reanalyses to state-of-the-art satellite-based data records. A careful recalibration and intercalibration of raw data records, retrieval harmonization/improvements, and refined assimilation schemes are key elements to increase the level of homogeneity and stability. A sound uncertainty estimation is required as well, and such efforts

should be carried out on a regular basis in conjunction with a reassessment of the achieved change in quality.

We emphasize the regional aspect of the impact of changes of the observing system and its relevance for the emerging need of regional climate analysis. It is important to verify the stability of a data record on global and all regional scales. The latter is a challenge because of missing reference observations with sufficient global and temporal coverage that at the same time are not affected by changes in the observing system.

Acknowledgments. M. Lockhoff and M. Schröder acknowledge the financial support by the EUMETSAT member states through CM SAF. J. Forsythe, H. Cronk, and T. Vonder Haar acknowledge the support of the NOAA NEAT program under Fuzhong Weng. The authors are grateful to ECMWF, NASA, NCEP, and REMSS for producing and making available the various data records. J. Schulz, L. Shi, B. Bojkov, F. Fell, C. Kummerow, and R. Roca are acknowledged for initiating G-VAP and for fruitful discussions. Comments from Michael Bosilovich and three anonymous reviewers are acknowledged. J. Trentmann supported the implementation of the homogeneity tool developed by X. Wang. We acknowledge the participants of the 3rd G-VAP workshop at CSU, in particular B. Weatherhead, for initiating the publication of this work.

APPENDIX

Acronyms

This appendix gives definitions for the acronyms that are used in this paper.

AIRS	Atmospheric Infrared Sounder
AVHRR	Advanced Very High Resolution Radiometer
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
CDR	Climate data record
CFSR	Climate Forecast System Reanalysis
CM SAF	Satellite Application Facility on Climate Monitoring
DMSP	Defense Meteorological Satellite Program
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA-Interim	ECMWF interim reanalysis
ESDR	Earth system data record
EUMETSAT	European Organisation for Exploitation of Meteorological Satellites
GCOS	Global Climate Observing System
GDAP	GEWEX Data and Assessments Panel

GEOS	Goddard Earth Observing System
GFS	Global Forecast System
GUAN	GCOS Upper-Air Network
GSI	Gridpoint Statistical Interpolation analysis system
GTOPO30	Global 30 arc s elevation dataset
G-VAP	GEWEX water vapor assessment
HIRS	High Resolution Infrared Radiation Sounder
HOAPS	Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data
IAU	Incremental analysis updates
IFS	Integrated Forecast System
IGRA	Integrated Global Radiosonde Archive
IPW	Integrated precipitable water
ITCZ	Intertropical convergence zone
IWV	Integrated water vapor
MEaSUREs	Making Earth Science Data Records for Use in Research Environments
MPI-M	Max Planck Institute for Meteorology
MERRA	Modern-Era Retrospective Analysis for Research and Applications
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
NVAP	NASA's Water Vapor Project
NVAP-M	NASA's Water Vapor Project-MEaSUREs
PMF	Penalized maximal F (test)
PWAT	Precipitable water
REMSS	Remote Sensing Systems
SSMIS	Special Sensor Microwave Imager/Sounder
SSM/I	Special Sensor Microwave Imager
SST	Sea surface temperature
TCWV	Total column water vapor
TPW	Total precipitable water vapor
UCAR	University Corporation for Atmospheric Research
USGS	U.S. Geological Survey

REFERENCES

- Andersson, A., K. Fennig, C. Klepp, S. Bakan, H. Graßl, and J. Schulz, 2010: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data—HOAPS-3. *Earth Syst. Sci. Data*, **2**, 215–234, doi:10.5194/essd-2-215-2010.
- Bedka, S., R. Knuteson, H. Revercomb, D. Tobin, and D. Turner, 2010: An assessment of the absolute accuracy of the Atmospheric Infrared Sounder v5 precipitable water vapor product at tropical, midlatitude, and arctic ground-truth sites: September 2002 through August 2008. *J. Geophys. Res.*, **115**, D17310, doi:10.1029/2009JD013139.
- Berg, W., M. R. P. Sapiano, J. Horsman, and C. Kummerow, 2013: Improved geolocation and Earth incidence angle information for a fundamental climate data record of the SSM/I sensors. *IEEE Trans. Geosci. Remote Sens.*, **51**, 1504–1513, doi:10.1109/TGRS.2012.2199761.
- Bosilovich, M. G., F. R. Robertson, and J. Chen, 2011: Global energy and water budgets in MERRA. *J. Climate*, **24**, 5721–5739, doi:10.1175/2011JCLI4175.1.
- Chelliah, M., W. Ebisuzaki, S. Weaver, and A. Kumar, 2011: Evaluating the tropospheric variability in National Centers for Environmental Prediction's climate forecast system reanalysis. *J. Geophys. Res.*, **116**, D17107, doi:10.1029/2011JD015707.
- Dai, A., J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang, 2011: A new approach to homogenize daily radiosonde humidity data. *J. Climate*, **24**, 965–991, doi:10.1175/2010JCLI3816.1.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Divakarla, M., and Coauthors, 2014: The CrIMSS EDR algorithm: Characterization, optimization, and validation. *J. Geophys. Res. Atmos.*, **119**, 4953–4977, doi:10.1002/2013JD020438.
- Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the Integrated Global Radiosonde Archive. *J. Climate*, **19**, 53–68, doi:10.1175/JCLI3594.1.
- Fennig, K., A. Andersson, S. Bakan, C.-P. Klepp, and M. Schröder, 2012: Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data—HOAPS 3.2—Monthly means/6-hourly composites. Satellite Application Facility on Climate Monitoring, accessed 22 March 2012, doi:10.5676/EUM_SAF_CM/HOAPS/V001.
- , —, and M. Schröder, 2013: Fundamental climate data record of SSM/I brightness temperatures. Satellite Application Facility on Climate Monitoring, doi:10.5676/EUM_SAF_CM/FCDR_SSMI/V001.
- Hilburn, K. A., and F. J. Wentz, 2008: Mitigating the impact of RADCAL beacon contamination on *FI5* SSM/I ocean retrievals. *Geophys. Res. Lett.*, **35**, L18806, doi:10.1029/2008GL034914.
- , —, C. A. Mears, T. Meissner, and D. K. Smith, 2010: Description of Remote Sensing Systems version-7 geophysical retrievals. *17th Conf. on Satellite Meteorology and Oceanography/17th Conf. on Air–Sea Interaction*, Annapolis, MD, Amer. Meteor. Soc., JP1.5. [Available online at <https://ams.confex.com/ams/17Air17Sat9Coas/webprogram/Paper174339.html>]
- Hyland, R. W., and A. Wexler, 1983: Formulations for the thermodynamic properties of the saturated phases of H₂O from 173.15 K to 473.15 K. *ASHRAE Trans.*, **89**, 500–519.
- Ju, J., and J. Slingo, 1995: The Asian summer monsoon and ENSO. *Quart. J. Roy. Meteor. Soc.*, **121**, 1133–1168, doi:10.1002/qj.49712152509.
- Lacis, A. A., G. A. Schmidt, D. Rind, and R. A. Ruedy, 2010: Atmospheric CO₂: Principal control knob governing Earth's temperature. *Science*, **330**, 356–359, doi:10.1126/science.1190653.
- Lindau, R., and V. Venema, 2016: The uncertainty of break positions detected by homogenization algorithms in climate records. *Int. J. Climatol.*, **36**, 576–589, doi:10.1002/joc.4366.
- Lindstrot, R., M. Stengel, M. Schröder, J. Fischer, R. Preusker, N. Schneider, T. Steenbergen, and B. Bojkov, 2014: A global climatology of total columnar water vapour from SSM/I and

- MERIS. *Earth Syst. Sci. Data*, **6**, 221–233, 2014, doi:[10.5194/essd-6-221-2014](https://doi.org/10.5194/essd-6-221-2014).
- Mears, C. A., B. D. Santer, F. J. Wentz, K. E. Taylor, and M. F. Wehner, 2007: Relationship between temperature and precipitable water changes over tropical oceans. *Geophys. Res. Lett.*, **34**, L24709, doi:[10.1029/2007GL031936](https://doi.org/10.1029/2007GL031936).
- Mieruch, S., M. Schröder, S. Noel, and J. Schulz, 2014: Comparison of decadal global water vapor changes derived from independent satellite time series. *J. Geophys. Res. Atmos.*, **119**, 12 489–12 499, doi:[10.1002/2014JD021588](https://doi.org/10.1002/2014JD021588).
- Reale, T., B. Sun, F. H. Tilley, and M. Petey, 2012: The NOAA Products Validation System (NPROVS). *J. Atmos. Oceanic Technol.*, **29**, 629–645, doi:[10.1175/JTECH-D-11-00072.1](https://doi.org/10.1175/JTECH-D-11-00072.1).
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, doi:[10.1175/1520-0442\(2002\)015<1609:AIIASAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIIASAS>2.0.CO;2).
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:[10.1175/JCLI-D-11-00015.1](https://doi.org/10.1175/JCLI-D-11-00015.1).
- Robock, A., 2002: Pinatubo eruption: The climatic aftermath. *Science*, **295**, 1242–1244, doi:[10.1126/science.1069903](https://doi.org/10.1126/science.1069903).
- , and C. Oppenheimer, Eds., 2003: *Volcanism and the Earth's Atmosphere*. *Geophys. Monogr.*, Vol. 139, Amer. Geophys. Union, 360 pp.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:[10.1175/2010BAMS3001.1](https://doi.org/10.1175/2010BAMS3001.1).
- Santer, B. D., and Coauthors, 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, **309**, 1551–1556, doi:[10.1126/science.1114867](https://doi.org/10.1126/science.1114867).
- Schlüssel, P., and W. J. Emery, 1990: Atmospheric water vapor over oceans from SSM/I measurements. *Int. J. Remote Sens.*, **11**, 753–766, doi:[10.1080/01431169008955055](https://doi.org/10.1080/01431169008955055).
- Schröder, M., M. Jonas, R. Lindau, J. Schulz, and K. Fennig, 2013: The CM SAF SSM/I-based total column water vapor climate data record: Methods and evaluation against re-analyses and satellite. *Atmos. Meas. Tech.*, **6**, 765–775, doi:[10.5194/amt-6-765-2013](https://doi.org/10.5194/amt-6-765-2013).
- Sherwood, S. C., R. Roca, T. M. Weckwerth, and N. G. Andronova, 2010: Tropospheric water vapor, convection, and climate. *Rev. Geophys.*, **48**, RG2001, doi:[10.1029/2009RG000301](https://doi.org/10.1029/2009RG000301).
- Soden, B. J., R. T. Wetherald, G. L. Stenchikov, and A. Robock, 2002: Global cooling following the eruption of Mt. Pinatubo: A test of climate feedback by water vapor. *Science*, **296**, 727–730, doi:[10.1126/science.296.5568.727](https://doi.org/10.1126/science.296.5568.727).
- Sohn, B. J., and E. A. Smith, 2003: Explaining sources of discrepancy in SSM/I water vapor algorithms. *J. Climate*, **16**, 3229–3255, doi:[10.1175/1520-0442\(2003\)016<3229:ESODII>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3229:ESODII>2.0.CO;2).
- , J. Schmetz, R. Stuhlmann, and J.-Y. Lee, 2006: Dry bias in satellite-derived clear-sky water vapor and its contribution to longwave cloud radiative forcing. *J. Climate*, **19**, 5570–5580, doi:[10.1175/JCLI3948.1](https://doi.org/10.1175/JCLI3948.1).
- Trenberth, K. E., J. Fasullo, and L. Smith, 2005: Trends and variability in column-integrated atmospheric water vapour. *Climate Dyn.*, **24**, 741–758, doi:[10.1007/s00382-005-0017-4](https://doi.org/10.1007/s00382-005-0017-4).
- Venema, V., and Coauthors, 2012: Benchmarking homogenization algorithms for monthly data. *Climate Past*, **8**, 89–115, doi:[10.5194/cp-8-89-2012](https://doi.org/10.5194/cp-8-89-2012).
- 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**, L16802, doi:[10.1029/2012GL052094](https://doi.org/10.1029/2012GL052094).
- Wang, J., L. Zhang, A. Dai, F. Immler, M. Sommer, and H. Vömel, 2013: Radiation dry bias correction of Vaisala RS92 humidity data and its impacts on historical radiosonde data. *J. Atmos. Oceanic Technol.*, **30**, 197–214, doi:[10.1175/JTECH-D-12-00113.1](https://doi.org/10.1175/JTECH-D-12-00113.1).
- Wang, X. L., 2008a: Penalized maximal F test for detecting undocumented mean shift without trend change. *J. Atmos. Oceanic Technol.*, **25**, 368–384, doi:[10.1175/2007JTECHA982.1](https://doi.org/10.1175/2007JTECHA982.1).
- , 2008b: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. *J. App. Meteor. Climatol.*, **47**, 2423–2444, doi:[10.1175/2008JAMC1741.1](https://doi.org/10.1175/2008JAMC1741.1).
- Weatherhead, E. C., and Coauthors, 1998: Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.*, **103**, 17 149–17 161, doi:[10.1029/98JD00995](https://doi.org/10.1029/98JD00995).
- Wentz, F. J., 1997: A well-calibrated ocean algorithm for SSM/I. *J. Geophys. Res.*, **102**, 8703–8718, doi:[10.1029/96JC01751](https://doi.org/10.1029/96JC01751).
- , 2015: A 17-yr climate record of environmental parameters derived from the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager. *J. Climate*, **28**, 6882–6902, doi:[10.1175/JCLI-D-15-0155.1](https://doi.org/10.1175/JCLI-D-15-0155.1).