

# Analysis of global water vapour trends from satellite measurements in the visible spectral range

S. Mieruch, S. Noël, H. Bovensmann, and J. P. Burrows

Institute of Environmental Physics, University of Bremen, FB 1, P. O. Box 330440, 28334 Bremen, Germany

Received: 15 June 2007 – Published in Atmos. Chem. Phys. Discuss.: 10 August 2007

Revised: 25 October 2007 – Accepted: 3 January 2008 – Published: 6 February 2008

**Abstract.** Global water vapour total column amounts have been retrieved from spectral data provided by the Global Ozone Monitoring Experiment (GOME) flying on ERS-2, which was launched in April 1995, and the Scanning Imaging Absorption spectrometer for Atmospheric CHartographY (SCIAMACHY) onboard ENVISAT launched in March 2002. For this purpose the Air Mass Corrected Differential Optical Absorption Spectroscopy (AMC-DOAS) approach has been used. The combination of the data from both instruments provides us with a long-term global data set spanning more than 11 years with the potential of extension up to 2020 by GOME-2 data on MetOp.

Using linear and non-linear methods from time series analysis and standard statistics the trends of H<sub>2</sub>O columns and their errors have been calculated. In this study, factors affecting the trend such as the length of the time series, the magnitude of the variability of the noise, and the autocorrelation of the noise are investigated. Special emphasis has been placed on the calculation of the statistical significance of the observed trends, which reveal significant local changes from –5% per year to +5% per year. These significant trends are distributed over the whole globe. Increasing trends have been calculated for Greenland, East Europe, Siberia and Oceania, whereas decreasing trends have been observed for the north-west USA, Central America, Amazonia, Central Africa and the Arabian Peninsular.

## 1 Introduction

Water vapour is the most important natural greenhouse gas in the atmosphere and plays a crucial role in the context of climate change, because of strong feedback mechanisms (Held and Soden, 2000). H<sub>2</sub>O plays an essential role in atmospheric

chemistry, e.g. the rapid conversion of sulfur trioxide to sulfuric acid, it is a source of the OH radical, and is also important for the ozone chemistry (Stenke and Grewe, 2005). Thus the knowledge of the global distribution of H<sub>2</sub>O and its evolution in time is of utmost importance for climate system studies.

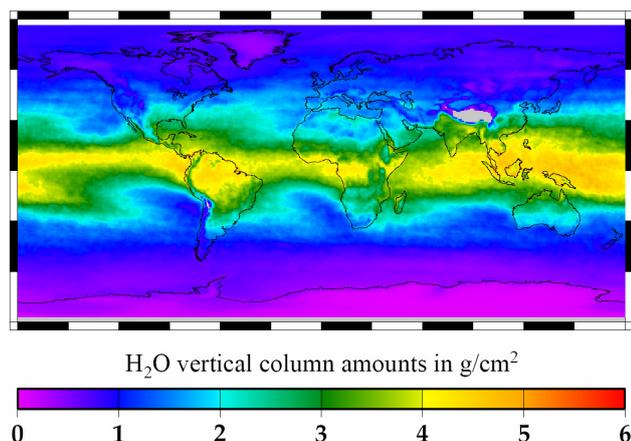
The strong infrared radiation absorbing character of H<sub>2</sub>O generates the natural greenhouse effect. Without H<sub>2</sub>O the global mean temperature at the surface would be 20°C lower than today (Häckel, 1999). In this context the transport of H<sub>2</sub>O plays an important role for the climate system. Atmospheric H<sub>2</sub>O represents the movement of energy in the form of latent heat. By condensation this latent heat can be released and could yield to a warming of the atmosphere, which affects global circulation systems associated with weather and climate.

The Earth's surface temperature results from an equilibrium state of the incoming solar radiation and the outgoing terrestrial radiation. Changes in the atmospheric composition, especially those of greenhouse gases such as H<sub>2</sub>O, carbon dioxide and methane, can alter the outgoing terrestrial radiation which leads to a new equilibrium state between the incoming and outgoing radiation fluxes, thus resulting in a changing Earth surface temperature (IPCC, 2007). CO<sub>2</sub> and CH<sub>4</sub>, which are also measured with the SCIAMACHY instrument (Buchwitz et al., 2006), are particularly important in the discussion of the anthropogenic greenhouse effect.

In the debates about climate change and the greenhouse effect H<sub>2</sub>O plays an extremely important role. For instance, climate models predict a global increase of H<sub>2</sub>O contents due to the global warming caused by increasing CO<sub>2</sub> and other greenhouse gases (Dai et al., 2001). This increased H<sub>2</sub>O reduces the outgoing long-wave radiation, which yields to an additional warming of the troposphere (IPCC, 2007). Together with these indirect effects on the atmospheric H<sub>2</sub>O contents, direct influences of anthropogenic interventions such as irrigation (Boucher et al., 2004) and deforestation

---

Correspondence to: S. Mieruch  
(sebastian.mieruch@iup.physik.uni-bremen.de)



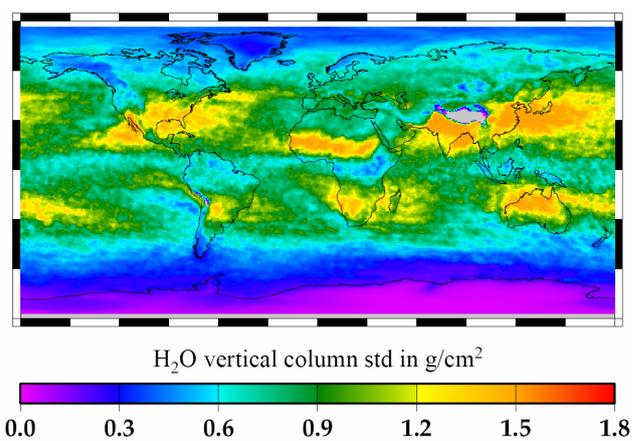
**Fig. 1.** Annual mean of H<sub>2</sub>O column amounts for the year 2006 derived from SCIAMACHY measurements.

(Gordon et al., 2005) alter the water vapour cycle and thereby the concentrations on local as well as on global scale.

For clarification, we have to note that several synonyms are used for the H<sub>2</sub>O total column amounts in the related literature, e.g. IWV (Integrated Water Vapour), TWV (Total Water Vapour), precipitable water, etc. In this paper we talk about H<sub>2</sub>O columns and H<sub>2</sub>O column amounts and denote therewith the complete amount of water vapour in grams per atmospheric column on a 1 cm<sup>2</sup> base (unit: g/cm<sup>2</sup>).

The H<sub>2</sub>O column of the atmosphere can be seen as a proxy for the climate state of a region, whether it is, for instance, humid or dry. Moreover, it is strongly linked to the surface temperature and to the lower altitude temperature of air. This strong correlation is shown by Wagner et al. (2006) for H<sub>2</sub>O columns retrieved from GOME. The H<sub>2</sub>O column amounts are high in the tropics, low over the poles and medium over the temperate zone. Figure 1 shows as an example the global annual mean and Fig. 2 depicts the respective variability (standard deviation) of the H<sub>2</sub>O column amounts for the year 2006 retrieved by the AMC-DOAS method (cf. Sect. 2) from SCIAMACHY data.

The H<sub>2</sub>O trends can be seen as tracers following the climate state of a specific region. A decreasing trend, for example, could be a change from a humid state to a dry state of a specific region. An infinite decreasing trend is impossible, so the trend has to stagnate at a certain point. If the H<sub>2</sub>O columns have significantly changed, dramatic consequences for the flora (major vegetation types, savanna, tundra etc. as reported by Melillo (1999)), fauna, agriculture and therefore inescapably for men cannot be ruled out. Moreover this new state could be stable and a way back is perhaps not easy, or, connected with a strong hysteresis (Scheffer and Carpenter, 2003). The same arguments are valid for increasing trends vice versa.



**Fig. 2.** Respective standard deviation of the annual means of H<sub>2</sub>O column amounts for the year 2006 derived from SCIAMACHY measurements.

The H<sub>2</sub>O columns and their changes are strongly linked to the climate state and the vegetation type of a region. Plants, animals and humans are adapted to their environmental conditions. Changes or trends of the atmospheric H<sub>2</sub>O columns, e.g. to dryer or more wet situations, can have critical consequences for life. Moreover, H<sub>2</sub>O trend calculations are important to prove model results and increase our knowledge of the hydrological cycle on global and local scale. The understanding of H<sub>2</sub>O correlated atmospheric processes, (e.g. evaporation, precipitation and cloud distribution) are supported by our study.

The importance and usefulness of H<sub>2</sub>O trends is enormous, but the detection of such trends is difficult. The trends can be influenced by several kinds of effects, such as instrumental changes or natural phenomena like autocorrelation. Another important natural phenomenon influencing the H<sub>2</sub>O columns is the ENSO (El Niño Southern Oscillation). El Niño is a natural recurring (without a constant period) climate phenomenon mostly (but not solely) impacting the tropics. With respect to atmospheric H<sub>2</sub>O the connection is performed through increasing and decreasing (depending on geolocation) surface temperatures, which cause increase and decrease of evaporation. The influence of the large El Niño event in 1997/1998 on the H<sub>2</sub>O columns is shown in Wagner et al. (2005). Also sea surface temperature is influenced by El Niño, but trend studies by Good et al. (2007) showed, that El Niño is not influencing the trends significantly for a 20 years data record, which is a great advantage of long data sets. Our data comprise 11 years, thus we have to investigate the impact of El Niño on the calculated trends.

After the strong 1997/1998 El Niño, two small El Niño events took place in 2002 and 2006. Figure 3 shows the sea surface temperature (SST) anomalies (red) and the GOME/SCIAMACHY H<sub>2</sub>O total column anomalies (blue)

for the area from 4° N to 4° S and 150° W to 90° W, which are both smoothed by a 5 months running mean filter. The El Niño event in 1997/1998 exceeds the other events by a factor of about 3. This strong coupling of the near-surface temperature anomalies with the H<sub>2</sub>O total column anomalies is also shown in Wagner et al. (2006) for GOME measurements.

Our H<sub>2</sub>O trend study comprises the years 1996 to 2006, i.e. 11 years of global satellite data. This length of data cannot resolve long-term oscillation. However, it is enough to show significant H<sub>2</sub>O changes in several regions on Earth.

An overview on the H<sub>2</sub>O retrieval method and validation efforts is given in the following Sect. 2.

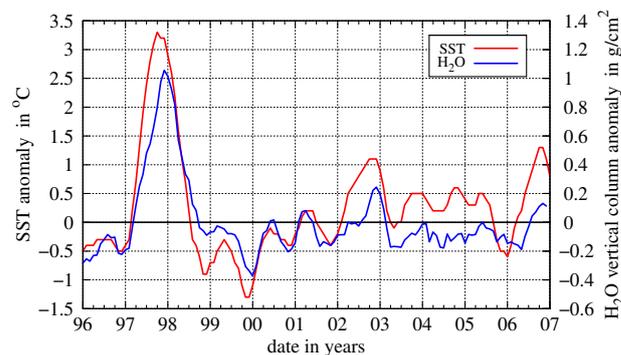
In Sect. 3 requirements for the combination of the two data sets are discussed, which are implemented in Sect. 4, where we describe the trend estimation including the statistical modelling of the time series.

Section 5 shows the results from the global trend analysis for the combined data set, and the influence of the 1997/1998 El Niño event on the trends is investigated.

## 2 Data analysis

The global H<sub>2</sub>O total column amounts used in the present study have been retrieved by the Air Mass Corrected Differential Optical Absorption Spectroscopy approach (AMC-DOAS) (Noël et al., 2004) from spectral data measured by the Global Ozone Monitoring Experiment (GOME) flying on ERS-2 which was launched in April 1995 and the SCanning Imaging Absorption spectrometer for Atmospheric CHartography (SCIAMACHY) onboard ENVISAT launched in March 2002. The basic principle of the method is to calculate the difference between the measured Earthshine radiance and the solar irradiance at wavelengths where H<sub>2</sub>O absorbs radiation (here we use the wavelength band from 688nm to 700 nm) and relate this absorption-depth to the H<sub>2</sub>O column concentration.

Within the AMC-DOAS retrieval certain surface and atmospheric conditions are assumed, namely no surface elevation, a surface albedo of 0.05, a tropical atmosphere and especially the absence of clouds. Usually these conditions differ from the real ones, which is accounted for by the so called Air Mass Correction Factor (AMCF) derived from O<sub>2</sub> absorption. Via the AMCF the H<sub>2</sub>O columns are scaled such that the correct O<sub>2</sub> optical depth is achieved (see Noël et al. (2004) for details). Deviations of the AMCF from unity indicate discrepancies between the assumed and the real conditions and if these deviations are too large (AMCF < 0.8), the H<sub>2</sub>O measurements are discarded. One of the main reasons for AMCF's differing from unity is the presence of clouds in the observed scene. Therefore the AMCF limit efficiently sorts out too cloudy scenes, but it is possible to derive H<sub>2</sub>O columns also from partly cloudy scenes, as long as the cloud fraction is low (AMCF ≥ 0.8). In this sense the AMC-DOAS products provide a cloud-cleared climatology.



**Fig. 3.** Monthly mean sea surface temperature (SST) anomalies (red) and GOME/SCIAMACHY H<sub>2</sub>O total column anomalies (blue) averaged for the area 4° N to 4° S and 150° W to 90° W and both smoothed by a 5 months running mean filter. SST Data taken from <http://coaps.fsu.edu/jma.shtml>

In the presence of clouds, the AMC-DOAS method is most probably slightly underestimating the H<sub>2</sub>O columns, because in contrast to the well mixed O<sub>2</sub>, the H<sub>2</sub>O volume mixing ratio increases towards the surface. Thus the derived AMCF should be typically lower in this case. However, this second order effect should not influence the H<sub>2</sub>O trends other than via trends in the cloud cover. For example, on the one hand an increase of the cloud cover with time would result in an increase of underestimated measurements of H<sub>2</sub>O and thus would decrease the water vapour trend. On the other hand a decrease of the cloud cover with time would yield to a decrease of underestimated measurements and thus to slightly increased trends. It has to be noted, that this is a second order effect, because the climatology is cloud-cleared, but it cannot be fully excluded.

The AMC-DOAS method provides a completely independent data set, because it does not rely on any additional external information. The retrieval of H<sub>2</sub>O data from the GOME instrument is described in Noël et al. (1999), where also validation results of the data with SSM/I (Special Sensor Microwave Imager) data are shown. Likewise, SCIAMACHY H<sub>2</sub>O data have been validated with SSM/I and ECMWF (European Centre for Medium-Range Weather Forecasts) data (Noël et al., 2005). An intercomparison and a preliminary connection of both, the GOME and the SCIAMACHY data sets, is shown in Noël et al. (2006). The high quality of the two H<sub>2</sub>O data sets is demonstrated from validation and comparison results, which shows that they can be merged well together. Thus, the trend analysis presented in this paper is built on a solid fundament. A good overview of other H<sub>2</sub>O measuring instruments from space can be found in Brocard (2006). Previous investigations of other H<sub>2</sub>O retrievals from GOME are described e.g. in Maurellis et al. (2000) and Lang et al. (2003). A similar H<sub>2</sub>O trend study to ours is presented by Wagner et al. (2006) for the GOME data, based on

a different retrieval method described in Wagner et al. (2003). In this study we extend the data set with the SCIAMACHY measurements and concentrate on the definition and calculation of statistically significant trends.

### 3 The combination of GOME and SCIAMACHY data

GOME on ERS-2 has been measuring since June 1995 up to the present, but since June 2003 no global coverage is provided as a result of a breakdown of the on-board tape recorders. SCIAMACHY data are available since August 2002, but the SCIAMACHY instrument did not achieve final flight conditions until January 2003. The quality of the SCIAMACHY H<sub>2</sub>O data is furthermore slightly reduced in 2002, because of the non-availability of one of the diffuser plates for solar observation prior to December 2002. Overall the most appropriate time for the change from GOME to SCIAMACHY data results in January 2003.

When combining the data sets possible level shifts have to be accounted between GOME and SCIAMACHY measurements. Therefore the period of near simultaneous global measurements of GOME and SCIAMACHY, August 2002 to June 2003, has been studied explicitly. The global agreement results in an average deviation of  $-0.01 \text{ g/cm}^2$  with a scatter of  $\pm 0.25 \text{ g/cm}^2$  (Noël et al., 2007). This means, that on a mean, there is strictly speaking no difference between the results of both instruments. This is anticipated, because we use the same retrieval method (AMC-DOAS) for both instruments and the method is quite insensitive to existing calibration differences between the GOME and SCIAMACHY instruments. The scatter of the H<sub>2</sub>O differences between the two instruments results from local (single grid pixel) time series, which show deviations. Although these differences on a local scale are small ( $\pm 0.25 \text{ g/cm}^2$ ) compared to the total H<sub>2</sub>O column, they can influence the trend and have to be considered. Two ways of determining the level shifts are possible. Firstly we can estimate (in the sense of arithmetic means) the local level shifts on the basis of the 5 months overlapping data. Secondly the level shifts can be estimated from the complete (merged) data set comprising 132 months using a least square method. With respect to the size of the underlying data base we will use the least square method (described in Sect. 4) to estimate the level shifts from the combined data sets.

Since the calibration between the instruments as a cause for the level shifts on local scale can be ruled out, two main aspects are responsible for the differences:

1. Different equator crossing time. GOME on ERS-2 and SCIAMACHY onboard ENVISAT, respectively, cross the equator at 10:30 and 10:00 local time. That means SCIAMACHY and GOME measure at different times slightly different states of atmospheric composition. This arises on the one hand from the diurnal cycle of the

H<sub>2</sub>O column, but with quite small impact during the 30 minutes. On the other hand and more important are fluctuations in the H<sub>2</sub>O column on fast time scales caused by e.g. winds and clouds, which are most likely responsible for the level shifts. Here we refer to our statistical analysis of the observed level shifts in Sect. 5.3. It follows that a possible mean level shift between both data sets has to be allowed for the combination of the data on a local scale.

2. Differing spatial resolutions. The spatial resolution of the GOME data is  $40 \text{ km} \times 320 \text{ km}$ , whereas it is (typically)  $30 \text{ km} \times 60 \text{ km}$  for SCIAMACHY data. When combining both data sets, different (higher) seasonal amplitudes have to be accounted for in the SCIAMACHY data with respect to GOME. Because of the higher resolution of SCIAMACHY, higher peaks (negative as well as positive) of H<sub>2</sub>O can be detected. Furthermore the resolution together with the cloud cover contributes to the level shift. Due to the higher resolution, SCIAMACHY “sees” more cloud free pixels than GOME which introduces a potentially positive bias for the SCIAMACHY data. However, this bias is observed on local scale, it is not visible on average. (cf. Sect. 5.3). As mentioned in Sect. 2 we expect a negative bias for the AMC-DOAS data, due to remaining clouds. Because of the different spatial resolutions, partly cloudy scenes are more probably for GOME; therefore a more negative bias for the GOME data compared to the SCIAMACHY data is expected. This is in line with the findings of higher SCIAMACHY columns and thus positive level shifts around the equator regions, where high cloudiness is more probable (see Sect. 5.3).

The local level shifts result from a complex interaction of atmospheric processes (clouds, winds, small scale fluctuations, diurnal cycle) within the 30 min time delay of GOME and SCIAMACHY and instrumental differences (resolution).

The daily H<sub>2</sub>O columns are gridded on a  $0.5^\circ \times 0.5^\circ$  lattice and averaged to yield monthly mean H<sub>2</sub>O columns.

The H<sub>2</sub>O columns are retrieved on a daily basis, but it has to be noted that ERS-2 and ENVISAT fly on a sun-fixed orbit, i.e. passing each point on Earth at constant local time. Thus measurements from GOME and SCIAMACHY are snap-shots of the actual atmospheric conditions at specific locations at specific times.

A global coverage is achieved for GOME data within 3 days and for SCIAMACHY nadir measurements within 6 days. Thus, in principle monthly mean data provide a data set without gaps. However, few gaps are observed even in the monthly mean data, because high cloudiness and high mountain area (e.g. the Himalayas) measurements are removed from the data by the AMC-DOAS algorithm. Moreover, since GOME and SCIAMACHY are spectrometers using the sunlight, measurements are only possible during daylight, and therefore no data is available at night, which results

in a lack of measurements at high latitudes during the polar night. Since GOME and SCIAMACHY are measuring in the nadir viewing geometry no profile information of H<sub>2</sub>O can be retrieved in this mode.

The derivation of H<sub>2</sub>O columns from GOME-type instruments has also some unique advantages: The retrieval is possible over land and ocean and no external calibration sources like radiosondes are required. Although the resulting H<sub>2</sub>O time series is quite short compared to other instruments like SSM/I which are looking forward to a 40 years series, it will be extended by other SCIAMACHY measurements and especially by the series of GOME-2 instruments on MetOp. The series of GOME-type instruments can therefore provide independent and consistent H<sub>2</sub>O data sets on both land and ocean for at least 25 years in 2020.

## 4 Methods

### 4.1 Trend estimation

The detection of trends is difficult and depends on the length of the time series, the magnitude of variability and autocorrelation of the data (Weatherhead et al., 1998). The trends can be influenced by level shifts inside the time series from instrument changes or new instrumental calibration etc.. Short time series as well as high variability, autocorrelation and level shifts in the data increase the uncertainty of trend detection. Statistical methods are used to reveal trends and explore their uncertainties. The methods used here are based on the approach of Weatherhead et al. (1998) and Tiao et al. (1990) and have been adapted to our requirements.

### 4.2 Statistical modelling

The time series of the data at one geolocation (i.e. a single grid point) can be described by the following trend model:

$$Y_t = \mu C_t + S_t + \omega X_t + \delta U_t + N_t, \quad t = 0, \dots, T, \quad (1)$$

where  $Y_t$  contains the monthly mean H<sub>2</sub>O measurements.  $\mu$  is the mean water vapour column of the time series at time  $t=0$  and  $C_t$  is a constant, which is unity for all  $t$ .  $\omega$  represents the trend and  $X_t$  contains the time (in our case from January 1996 until December 2006 or from month 0 to 131, respectively), which is not necessarily equidistant as there may be missing data.  $\delta$  is the magnitude of a mean level shift at time  $t=T_0$  ( $0 < T_0 < T$ ), where  $T_0=84$  represents the intersection of GOME and SCIAMACHY data on January 2003.  $U_t$  describes a step function:

$$U_t = \begin{cases} 0, & t < T_0 \\ 1, & t \geq T_0 \end{cases}, \quad (2)$$

and the seasonal component  $S_t$  is modelled by a Fourier series

$$S_t = \eta \sum_{j=1}^4 [\beta_{1,j} \cdot \sin(2\pi j t / 12) + \beta_{2,j} \cdot \cos(2\pi j t / 12)]. \quad (3)$$

Expanding the methods used by Weatherhead et al. (1998) an additional term  $\eta=1+(\gamma-1)U_t$  is used and describes an amplitude change of magnitude  $\gamma$  at time  $t \geq T_0$ .

The last term  $N_t$  in Eq. (1) contains the unexplained portion of the data, i.e. the noise. The noise  $N_t$  is assumed to be an autoregressive process of the order of one (AR(1)) (Schlittgen and Streitberg, 1997), i.e.

$$N_t = \phi N_{t-1} + \epsilon_t, \quad (4)$$

where  $\epsilon_t$  are independent random variables with zero-mean and variance  $\sigma_\epsilon^2$ . This assumption is used because environmental data is often autocorrelated, e.g. if the temperature is high at one day, a high temperature is likely on the next day. The magnitude or the memory of the autocorrelation is presented by  $\phi$ , which is restricted to  $-1 < \phi < 1$ , so the noise process  $N_t$  is stationary. The memory of the data at lag one can be calculated using the autocorrelation function  $\phi = \text{Corr}_{N_t N_{t-1}}$ , which is directly linked to the well known correlation coefficient. More sophisticated approaches of analysing long-term correlations in environmental data are the Detrended Fluctuation Analysis (DFA) (Rybski et al., 2006) and the modelling of time series with Markov-Chains (Freund et al., 2006). On a mean, the autocorrelation functions of our H<sub>2</sub>O noise time series converge to zero at lag two, thus the autocorrelation at lag one is adequate for our purposes.

Generally the autocorrelation function is restricted to continuous, statistically stationary stochastic functions, or in the discrete case equidistantly sampled data. Since there are gaps in our time series the discrete correlation function for analysing unevenly sampled data which was originally developed by Edelson and Krolik (1988) for astronomical problems, was applied.

To calculate the autocorrelation of the noise, the noise itself has to be determined by applying the model (Eq. (1)) to the data and subtract the fit from the data. For minimising the model in a least square sense we used the well known Levenberg-Marquardt algorithm for non-linear least square regression. The noise  $N_t$  is then given by the remaining residuals:

$$N_t = Y_t - (\hat{\mu} C_t + \hat{S}_t + \hat{\omega} X_t + \hat{\delta} U_t) \quad (5)$$

where  $\hat{\mu}$ ,  $\hat{\omega}$ ,  $\hat{\delta}$  are the least square estimators and  $\hat{S}_t$  stands for the seasonal component resulting from the fitted parameters  $\hat{\beta}_{i,j}$  and  $\hat{\gamma}$ . The  $N_t$  are used to calculate first the set of unbinned discrete correlations

$$\theta_t = \frac{N_t \cdot N_{t-1}}{\sigma_N^2}, \quad t = 1, \dots, T, \quad (6)$$

where the  $N_t$  have zero-mean and variance  $\sigma_N^2$ . Following the  $\theta_t$  have to be assigned to their lags,  $\tau_t$  with

$$\tau_t = X_t - X_{t-1}, \quad t = 1, \dots, T. \quad (7)$$

Now, the magnitude  $\phi$  of autocorrelation at lag  $\tau=1$  can be determined by averaging over the number  $M$  of  $\theta_t$  with corresponding  $\tau_t=1$ :

$$\phi = \frac{1}{M} \sum_{i=1}^M \theta_i(\tau_i = 1). \quad (8)$$

The aim of the above calculations concerning autocorrelations is to account for them during the fitting procedure. For this purpose the estimated seasonal component  $\hat{S}_t$  is subtracted from the data, because it has negligible effect on the estimation of the other parameters (Weatherhead et al., 1998). The model then becomes

$$A_t = Y_t - \hat{S}_t = \mu C_t + \omega X_t + \delta U_t + N_t, \quad t = 0, \dots, T. \quad (9)$$

After the non-linear terms have been removed from the model a linear matrix transformation to consider the autocorrelations is possible. Making the connection to the autoregressive process of Eq. (4), the model has absorbed the autocorrelations of  $N_t$  into the transformed data  $A_t^*$ ,  $C_t^*$  (which is no more constant), the time  $X_t^*$  and the step function  $U_t^*$ , whereas the  $N_t$  have lost their autocorrelations and have become white noise  $\epsilon_t$ :

$$A_t^* = \mu C_t^* + \omega X_t^* + \delta U_t^* + \epsilon_t, \quad t = 0, \dots, T. \quad (10)$$

Now a linear regression is applied, which can be solved analytically for the least square estimators  $\hat{\mu}$ ,  $\hat{\omega}$ ,  $\hat{\delta}$  and their errors  $\sigma_{\hat{\mu}}$ ,  $\sigma_{\hat{\omega}}$ ,  $\sigma_{\hat{\delta}}$ . Details of the transformation and regression are given in the Appendix and in Weatherhead et al. (1998).

#### 4.2.1 Trend fitting and estimation of the uncertainty

After the implementation of the autocorrelations into the model and solving the linear least square equations (where we denote the least square estimator of the trend with  $\hat{\omega}$ ) a good approximation of the error of the trend  $\sigma_{\hat{\omega}}$  is given by (Weatherhead et al., 1998):

$$\sigma_{\hat{\omega}} \approx \frac{\sqrt{12} \sigma_N}{\ell^{\frac{3}{2}}} \cdot \sqrt{\frac{1+\phi}{1-\phi}} \cdot \frac{1}{[1-3\vartheta(1-\vartheta)]^{\frac{1}{2}}}. \quad (11)$$

$\sigma_{\hat{\omega}}$  depends on the standard deviation  $\sigma_N$  of the noise, the length of the time series  $\ell$ , the autocorrelation  $\phi$  of  $N_t$  and the fraction  $\vartheta=T_0/\ell$  of the data before the level shift occurs.

#### 4.2.2 Significance of the trend

One main question concerning trends is whether the trend is significant or not. The answer to this question can only be given in a probabilistic sense. Based on the null hypothesis that the observed trend is equal to zero  $H_0 : \hat{\omega}=0$  the alternative hypothesis is the observation of a nonzero trend  $H_1 : \hat{\omega} \neq 0$ . The least square method assumes Gaussian distributed data around the fitted function. Using standard rules of random variables it can be shown that the trend  $\omega$  is a linear function of the data  $Y_t$  and therefore also Gaussian distributed (Fahrmeir et al., 2004). The probability of measuring a trend with magnitude greater than two times its error becomes  $P_{H_0}\{|\hat{\omega}|>2\sigma_{\hat{\omega}}\}=0.05$  and the chance of making an error in rejecting the null hypothesis is 5%. Accordingly, the likelihood to be correct in confiding the alternative hypothesis is 95%. Therefore we will adopt the rule that a trend  $\hat{\omega}$  is statistically significant when a probability of 95% is achieved with  $|\hat{\omega}|>2\sigma_{\hat{\omega}}$ .

## 5 Results

### 5.1 Global trend patterns

The global trend patterns are determined from the long-term time series from January 1996 to December 2006 including GOME and SCIAMACHY globally gridded monthly mean data on a  $0.5^\circ \times 0.5^\circ$  grid. Two ways of investigating the trends are informative; on the one hand displaying the absolute trends  $\hat{\omega}$  in  $\text{g/cm}^2$  per year (Fig. 4) and on the other hand displaying the relative trends  $\hat{\omega}/\hat{\mu}$  in % per year (Fig. 5), where  $\hat{\mu}$  represents the deseasonalised H<sub>2</sub>O columns at the beginning of the time series.

The absolute trends shown in Fig. 4 are stronger near the equator and smaller near the poles. Bluish as well as yellowish and reddish patches are seen, thus there are negative as well as positive trends observed, however most trends are small and distributed around zero.

Now the question arises if these observed trends are significant in a statistical way. Here it has to be noted, that a non-significant trend does not mean that the results are wrong, but that the magnitude of the observed trend has a higher uncertainty. We will use the significance definition from Sect. 4.2.2 that a trend is significant if it is greater than two times its error (Weatherhead et al., 1998). In addition we will extend the significance criterion by the claim, that the time series has to contain at minimum 2/3 of the maximum data points and denote this additional criterion with  $\ell \geq 2/3L$ , where  $\ell$  is the number of data points of a specific time series and  $L$  is the number of maximum data points. In our case we have 11 years of monthly data, yielding  $L=132$  and  $\ell \geq 88$ . Figures 6 and 7 show the absolute and relative significant trends, respectively, which are distributed over the whole globe. The significant trends agree with either strong

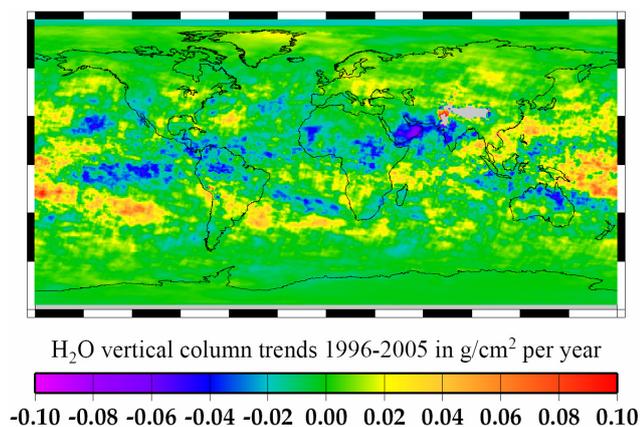


Fig. 4. Global absolute H<sub>2</sub>O trends (1996 to 2006).

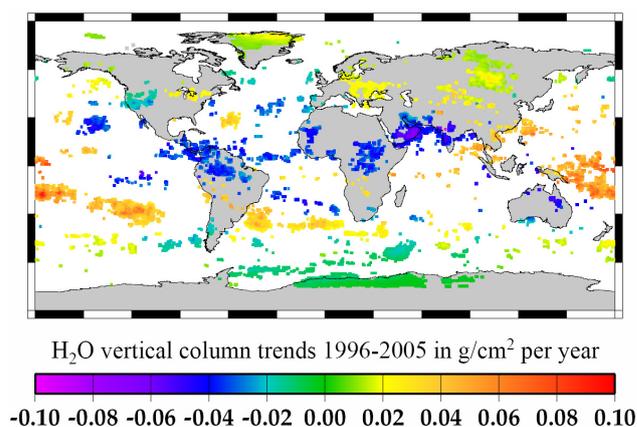


Fig. 6. Global absolute significant H<sub>2</sub>O trends (1996 to 2006).

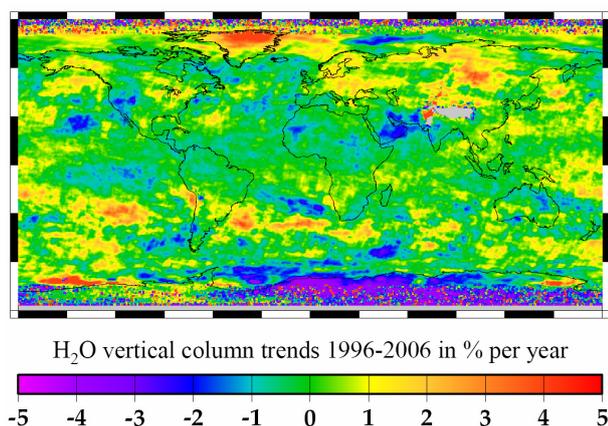


Fig. 5. Global relative H<sub>2</sub>O trends (1996 to 2006).

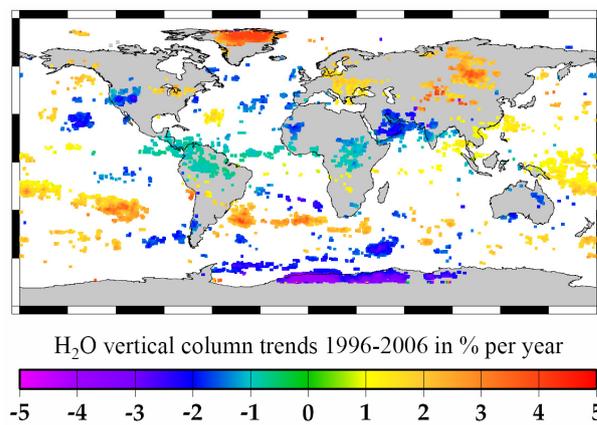


Fig. 7. Global relative significant H<sub>2</sub>O trends (1996 to 2006).

absolute or strong relative trends. However, it is interesting that also small absolute (e.g. Antarctica) or small relative trends (e.g. Amazonia) can be significant.

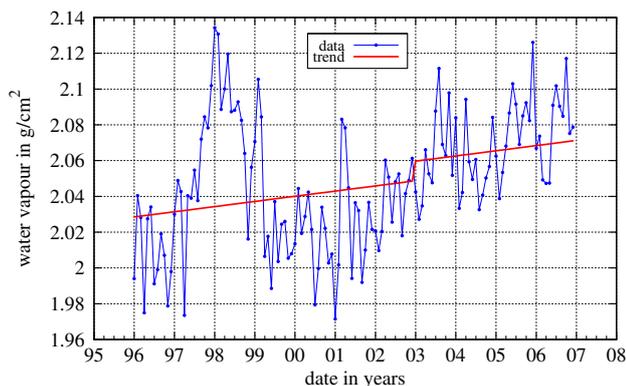
Figures 6 and 7 reveal several local regions with significant trends, for instance increasing H<sub>2</sub>O columns in Greenland, East Europe, Siberia and Oceania.

Water vapour decrease is observed in the northwest USA, Central America, Amazonia, Central Africa and the Arabian Peninsular.

Trenberth et al. (2005) calculated H<sub>2</sub>O total column trends from SSM/I data (over ocean only) for the time span from 1988 to 2003. Although Trenberth et al. (2005) analysed a different time interval, several similar patterns to ours are observed on the global maps, e.g. negative trends at the east coast of Australia, positive trends in the south west Pacific, positive trends covering a band from the east coast of India over Southeast Asia until the open Pacific. Of course there are differences, for instance positive trends at the west coast

of Peru spread far into the ocean seen by Trenberth et al. (2005), whereas our trends are also positive at the coast, but zero and negative in the ocean. Furthermore, the Trenberth et al. (2005) SSM/I trends are about one magnitude smaller than our trends, which is most probably due to the difference in the used time interval. A first inspection of SSM/I H<sub>2</sub>O total column trends, for the time span from 1995 to 2005 provided by the HOAPS data base (Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data, Andersson et al., 2007), reveal trends of the same magnitude as our trends and we are looking forward to a more detailed intercomparison.

Since H<sub>2</sub>O trends are usually quite small, the consideration of both, the level shift and the amplitude change during the fit routines, is highly important, especially the level shift. Considering or not considering the level shift  $\delta$  has crucial consequences for the trends. This is revealed from a trend calculation only for the GOME data, where we observed quite similar results to the combined data using the level shift. Disregarding the level shift yields rather different



**Fig. 8.** Time series of deseasonalised spatially averaged monthly means of the entire globe with the trend (red line) regarding auto-correlations.

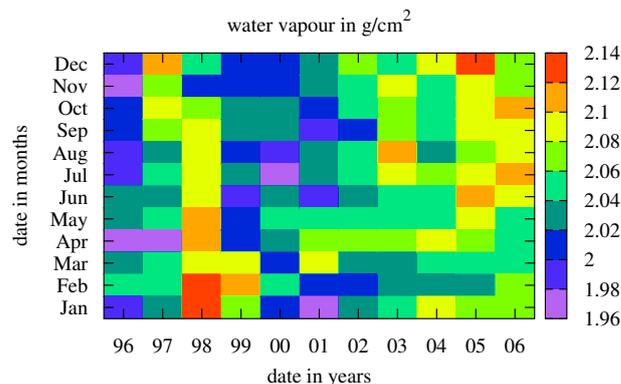
findings. Concluding, the integration of the level shift term (which is investigated statistically in Sect. 5.3) in our model is absolutely necessary for our trend calculation.

Neglecting the change in the amplitude yields on the one hand a higher noise signal in the deseasonalised data  $A_T$ . On the other hand a remaining seasonal component is left in the  $N_T$ , which results in changing autocorrelations. Both aspects are not critical for the trends, but crucial for the estimation of the errors of the trends (Eq. 11) and therefore the significance of the trends.

## 5.2 Globally averaged monthly mean H<sub>2</sub>O trend

Our trend analysis is applied to a time series of deseasonalised globally averaged monthly mean H<sub>2</sub>O columns shown in Fig. 8 including the strong 1997/1998 El Niño data. As can be seen from Fig. 3 the two El Niño events in 2002 and 2006 are small compared to the El Niño in 1997/1998. Here, we can benefit from the consideration of the autocorrelation, because the possible change in H<sub>2</sub>O caused by an El Niño event changes the autocorrelation of the data. For instance increasing H<sub>2</sub>O columns over a limited time yield to systematics in the noise and therefore to increasing autocorrelation which yields to a higher error  $\sigma_{\hat{\omega}}$  of the trend, because autocorrelations are considered in Eqs. (10) and (11). Hence it is not necessary to remove small events such as 2002 and 2006.

It has to be noted that a weighted mean is used when accumulating spatial measurements on a regular latitude/longitude grid, where the weights are given by the cosine of the latitude of each grid point, to account for the different surface areas. The red line in Fig. 8, corresponding to the fit parameter  $\hat{\omega}$ , shows an increase of  $0.0029 \text{ g/cm}^2 \pm 0.0028 \text{ g/cm}^2$  per year, i.e. 0.14% per year related to the fitted parameter  $\hat{\mu} = 2.03 \text{ g/cm}^2$ . This trend is non-significant, because the high autocorrelation of  $\hat{\phi} = 0.6$



**Fig. 9.** Time series of months plotted against years, while the deseasonalised globally averaged H<sub>2</sub>O column amounts are coded with colours.

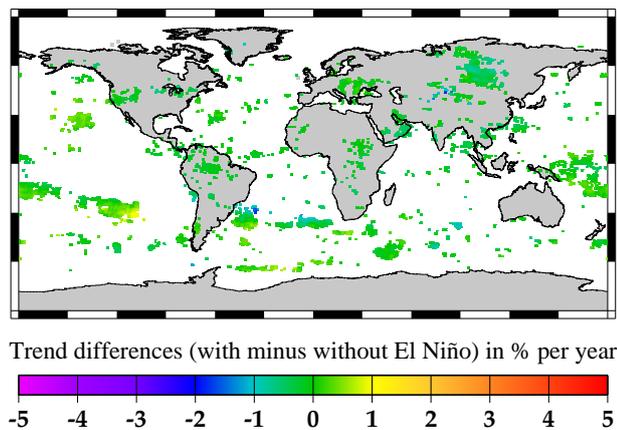
increases the error of the trend  $\sigma_{\hat{\omega}}$  as can be seen from Eq. (11).

One reason for the high autocorrelation is the presence of high H<sub>2</sub>O column amounts around the year 1998, which are most likely caused by the El Niño event. These higher columns are also reported by Wagner et al. (2005) for H<sub>2</sub>O retrieved from GOME data by a different algorithm.

As stated above the 1997/1998 El Niño event is most likely influencing the trend in Fig. 8, and probably data obtained during the El Niño time have to be removed as a kind of recurring phenomenon. Otherwise it is not clear if El Niño can be totally separated from the trend, because it cannot be excluded that for instance due to an increasing H<sub>2</sub>O trend the magnitude of the El Niño is increased. Nevertheless, we identify the strong 1997/1998 El Niño in the time series and remove the corresponding data to quantify the effect on the trends, especially on the significance of the trends. The influence of the 1997/1998 El Niño is shown in Fig. 9, where the months are plotted against the years and the globally averaged deseasonalised H<sub>2</sub>O column amounts are coded with colours.

As can be seen from Fig. 9, high H<sub>2</sub>O columns are observed from September 1997 until March 1999. Accordingly, the global trend analysis is performed again with the data set where we removed these potentially El Niño influenced data. The differences between the collocated significant trends on the global map based on the complete data and the data where we removed El Niño is quite small with a mean of  $-0.08\% \pm 0.34\%$  and shown in Fig. 10.

From this finding it reveals, that removing the potentially El Niño influenced data is not really needed for the data set, which is most satisfiable, because removal of data is often critical. However, for single time series, such as the globally averaged data, the El Niño influence can be crucial and removing of data points may be required.



**Fig. 10.** Differences between the collocated significant trends estimated from the complete time series and the data, where we removed potential El Niño influenced measurements (September 1997–March 1999).

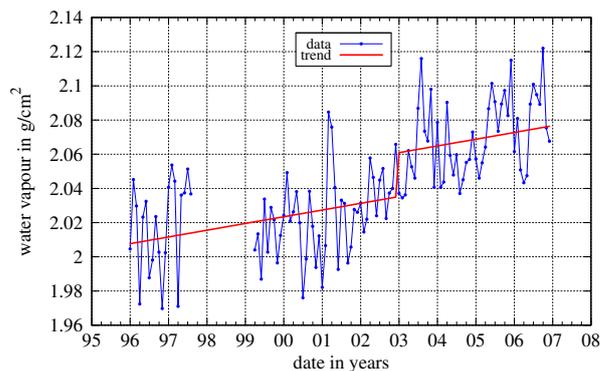
Figure 11 depicts the deseasonalised spatially averaged monthly mean column amounts of the data with the El Niño event removed.

The trend (red line) yields  $0.0039 \text{ g/cm}^2 \pm 0.0015 \text{ g/cm}^2$  per year or 0.19% per year thus the trend is significant with  $\hat{\omega} > 2.6\sigma_{\hat{\omega}}$ .

The magnitude of the level shift in Fig. 8 is increased with respect to the level shift in Fig. 11 due to the removed data. The error of the trend without El Niño data is decreased, because the autocorrelation of the time series is decreased to  $\hat{\phi} = 0.2$  and affects the errors of the fit parameters much less. Hence, there is a strong contribution of the 1997/1998 El Niño event to the autocorrelations of the time series.

To demonstrate the sensitivity of the calculated trends  $\omega$  and their errors  $\sigma_{\omega}$  to the used regression model and data set, several fitting procedures are performed for the single time series of deseasonalised globally averaged monthly mean data. The trend estimation (shown in Table 1) is applied, with and without the 1997/1998 El Niño data, to GOME (January 1996 to December 2002) measurements only and to the complete data set based on GOME and SCIAMACHY measurements (January 1996 to December 2006). The first column of Table 1 indicates the fitting method used, i.e. all permutations of considering and neglecting (denoted as cancelled parameter) the level shift  $\delta$  and the autocorrelation  $\phi$  of the noise  $N_t$ .

As can be seen from Table 1, fitting a single H<sub>2</sub>O time series (in this case the global monthly mean data) is quite sensitive to the regression model used. Applying the full model used in this paper (considering  $\delta$  and  $\phi$ ) delivers the most reliable results, which is explained below:



**Fig. 11.** Time series (with removed El Niño measurements) of deseasonalised spatially averaged monthly means of the entire globe with the trend (red line) regarding autocorrelations.

$[\delta, \phi]$ : Both parameters, the level shift and the autocorrelation, are fitted to the complete data set. Including El Niño yields the trend from Fig. 8, which is relatively high, because the consideration of  $\phi$  attenuates the high H<sub>2</sub>O measurements in 1997/1998. Otherwise a high error is observed, because fitting  $\delta$  and  $\phi$  introduces additional uncertainties and therefore increases the error (cf. Eq. 11). Neglecting El Niño increases the trend and decreases the error, because the resulting data contains less autocorrelations and less noise.

$[\delta, \phi]$ : The trend calculation is performed without regarding autocorrelations with the consequence that the high H<sub>2</sub>O columns in 1997/98 lift up the trend curve at that time and the magnitude of the trend is nearly zero. Furthermore the error is decreased, because no autocorrelation is considered. However, the relative error is still about 100%. If the El Niño data is masked the trend and its error have hardly changed compared to the “full fit”, because  $\phi$  is quite small without El Niño data.

$[\delta, \phi]$ : Fitting only the GOME time series from January 1996 to December 2002 and including autocorrelations yields an almost zero trend, because of the strong influence of El Niño in 1997/98. On the contrary, if the El Niño data is removed, the trend and its error is quite similar to those calculated for the complete data considering both parameters. If the regression without  $\delta$  is performed for the complete data set, the trend is increased, which is clear, because a positive level shift (see Fig. 8), which exists but is not fitted, increases the trend. When the El Niño data is removed the level shift is even larger, which can be seen in Fig. 11. In this case neglecting the level shift strongly increases the trend.

$[\delta, \phi]$ : Neglecting autocorrelations results for the GOME data with El Niño in a nearly zero slightly

**Table 1.** Results from several fitting procedures show the sensitivity of the trends  $\omega$  and their errors  $\sigma_\omega$  (in g/cm<sup>2</sup> per year) to the consideration of the level shift  $\delta$  and the autocorrelations  $\phi$ .

	including El Niño data		neglecting El Niño data	
	GOME	GOME & SCIA	GOME	GOME & SCIA
$\delta, \phi$	–	0.0029±0.0028	–	0.0039±0.0015
$\delta, \phi$	–	0.0002±0.0002	–	0.0037±0.0012
$\delta, \phi$	0.0006±0.0040	0.0043±0.0019	0.0034±0.0015	0.0072±0.0010
$\delta, \phi$	–0.0010±0.0019	0.0041±0.0009	0.0033±0.0012	0.0073±0.0007

negative trend, whereas without El Niño the trend is quite high as in the case above. For the complete GOME and SCIAMACHY data the situation is similar to the above scenario. The negligence of  $\delta$  extremely influences the trend.

From the results presented above we come to the following conclusions:

1. Only significant trends can be trusted. No conclusions should be drawn from non-significant trends.
2. For the global time series all significant trends (with or without data during El Niño times) agree within their errors as long as the level shift  $\delta$  is considered. Not considering  $\delta$  has a large impact on the absolute values of the trend and especially results in large deviations between the trends derived from the combined GOME/SCIAMACHY time series and the GOME data set alone. If the assumption of a constant linear trend (as in our model) is correct, the derived trend should not depend on the length of the time series; only the error of the derived trend should be affected. Therefore we conclude that the level shift has to be considered in the trend determination.
3. The global trends derived with and without autocorrelation agree within their errors. Nevertheless, a correct estimation of the error (and thus the significance of a trend) requires autocorrelations to be taken into account. This is especially important, if the El Niño data are included because these show a high autocorrelation.
4. The influence of El Niño on the derived trends depends strongly on the region for which trends shall be estimated. Depending on the region the inclusion of data measured during El Niño times may result in a non-significant trend (as in the case of the global time series). On the other hand, Fig. 10 shows that local significant trends derived from the full time series are hardly changed and remain significant if the El Niño times are excluded. Therefore, we suggest to use as baseline for all trend calculations the full fit applied to the complete data set without removing any measurements (what we

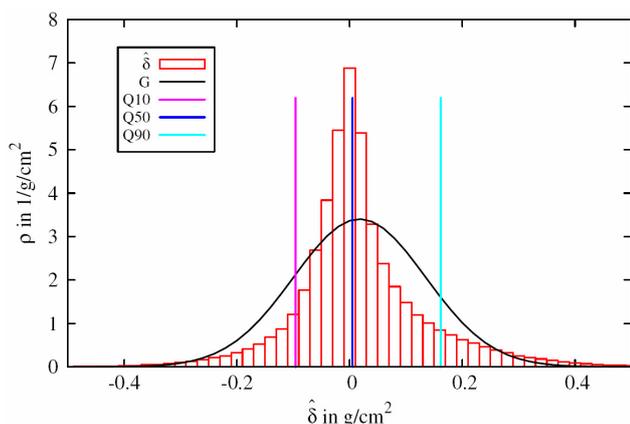
have done). A removal of data should only be done after a careful inspection of an individual time series, be it for a single grid pixel or for spatially averaged data (like the whole globe).

For the globally averaged time series from Fig. 8, where a strong El Niño signal is observed, the removal of the potential El Niño data is needed (Fig. 11). In this case the most reliable method is the full fit applied to El Niño adjusted combined GOME/SCIAMACHY data, which results in a globally averaged water vapour trend of 0.0039 g/cm<sup>2</sup>±0.0015 g/cm<sup>2</sup> or 0.19%±0.07 % per year. This value is in line with the corresponding result for the GOME time series of 0.0034 g/cm<sup>2</sup>±0.0015 g/cm<sup>2</sup> or 0.17%±0.07% per year. For comparison, Wagner et al. (2006) calculated the GOME trend from annual averaged water vapour columns for the time from 1996 to 2002 with a magnitude of about 0.4% per year, which is about a factor of two higher.

There are various possible explanations for this discrepancy: First, Wagner et al. (2006) use a different retrieval algorithm, hence it is a comparison of two different data sets. Furthermore it is most likely that different data preprocessings have been used in the generation of temporal or spatial means, which comprise gridding and averaging. Since the trend analysis of GOME annual means is based on only 7 data points, slightly differing annual means between both data sets can have crucial impact on the linear regression. However, further investigation of these differences is needed.

### 5.3 Influence of the level shift and the amplitude change

The above analysis showed that the derived trends are significantly affected by the level shift. Therefore we present the results of a statistical investigation of the fitted level shift  $\delta$  and also the amplitude change  $\gamma$  and reveal their distributions. A statistical analysis of the least square estimator  $\hat{\delta}$  of the mean level shift is shown in Fig. 12 where the density distribution (red bars) with binsize 0.02 g/cm<sup>2</sup> of the  $\hat{\delta}$  for the whole globe is plotted. The black line denoted with G describes a Gaussian which is fitted to the distribution of the  $\hat{\delta}$ .

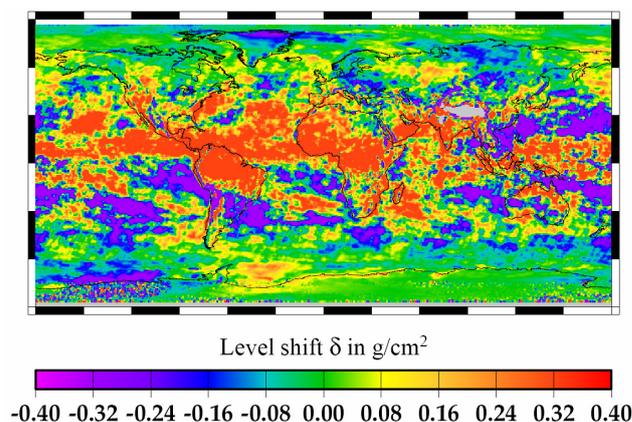


**Fig. 12.** Density distribution of the mean level shift ( $\hat{\delta}$ ) together with a fitted Gaussian (G), the median (Q50), and the 10% (Q10) as well as 90% (Q90) quantiles.

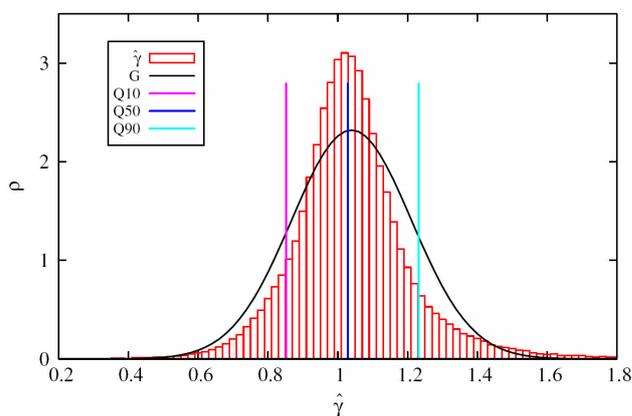
The level shifts between GOME and SCIAMACHY data reveal a slightly narrower distributed offset than a Gaussian normal distribution between both instruments. A description in terms of quantiles is quite suitable. The 10% quantile, denoted as Q10 (magenta line) in Fig. 12 lies at  $-0.096 \text{ g/cm}^2$ , the 50% quantile (the median blue line) has a magnitude of  $Q50=0.005 \text{ g/cm}^2$  and the 90% quantile (cyan line) is observed at  $Q90=0.162 \text{ g/cm}^2$ . Strictly speaking, in most cases the H<sub>2</sub>O columns do not change, which is reflected by the nearly zero median, but with a probability of 10% it changes less than  $-0.096 \text{ g/cm}^2$  and more than  $0.162 \text{ g/cm}^2$  (also with a probability of 10%), which corresponds to  $-4.7\%$  and  $8.0\%$ , respectively, related to the global mean H<sub>2</sub>O column at the beginning of the time series (fitted parameter  $\mu$ ) of about  $2.03 \text{ g/cm}^2$ . This scatter of the level shift is attributed to the time delay between the two instruments as well as to the high variability of atmospheric H<sub>2</sub>O and the cloud distribution in connection to the different instrumental resolutions (cf. Sect. 3).

In Fig. 13 we present the spatial distribution of the level shifts  $\delta$ , which reveals a patchy structure of positive and negative level shifts. Larger level shifts are observed near the equator (where also the H<sub>2</sub>O columns are large) and smaller level shifts are located at higher latitudes. Over the equator and over rain forest regions (e.g. Amazonia) mostly positive level shifts are detected. This is most likely caused by the different resolution between GOME and SCIAMACHY as mentioned in Sect. 3. This can be understood by assuming that a cloudy scene contains on average more water vapour than a cloud free scene. Partly cloudy scenes are better resolved by SCIAMACHY, thus in these situations SCIAMACHY columns tend to be higher than GOME columns.

The statistical analysis of the level shifts  $\delta$  confirms the assumptions from Sect. 3. The reason for the level shift is a complex superposition of instrumental effects (different resolution) in combination with local cloud covers, different



**Fig. 13.** Spatial distribution of the level shifts.



**Fig. 14.** Density distribution of the amplitude change between GOME and SCIAMACHY H<sub>2</sub>O columns ( $\hat{\gamma}$ ) together with a fitted Gaussian (G), the median (Q50), and the 10% (Q10) as well as 90% (Q90) quantiles.

measurement times (30 min time delay between ERS-2 and ENVISAT) and natural aspects like the high variability of atmospheric H<sub>2</sub>O.

Figure 14 depicts the global density distribution (binsize 0.02) of the amplitude changes  $\hat{\gamma}$  together with a fitted Gaussian (black curve) denoted by G. As expected, as a result of the higher resolution of the SCIAMACHY instrument, the amplitude factors  $\hat{\gamma}$  are higher than one with a median of  $Q50=1.029$  (blue line), but with a scatter of  $Q10=0.852$  (magenta line) and  $Q90=1.232$  (cyan line). However, in the same manner as for the level shifts, also the reason for the amplitude change is a complex interaction of instrumental and natural phenomena.

Concluding, both the mean level shift and the amplitude change at the intersection of the GOME and SCIAMACHY data are small compared to the H<sub>2</sub>O column amounts.

However they need to be considered in the trend analysis, because the observed trends are also small.

## 6 Conclusions and discussion

The trend analysis (of global monthly mean H<sub>2</sub>O data from 1996 to 2006) focussed on the estimation of the statistical significance of the observed trends. First the trends were calculated from monthly mean H<sub>2</sub>O column amounts where we removed the seasonal component. Special emphasis was placed on the consideration of autocorrelations in the data. The trend calculation, which is based on the well known least square linear regression, provides an error for the trend. This error is influenced by the length of the time series, the noise, the autocorrelation of the noise, and the level shift between GOME and SCIAMACHY data.

Two criteria for a significant trend are proposed:

- A trend has to be greater than 2 times its error, which arises from standard statistics.
- The time series has to comprise minimum 2/3 of the maximum data, which is required as a quality criterion.

For the period of January 1996 to December 2006 we found significant increase in the H<sub>2</sub>O columns (cf. Figs. 6 and 7) in Greenland, East Europe, Siberia and Oceania, and we have significant decrease in the northwest USA, Central America, Amazonia, Central Africa, and the Arabian Peninsular. The significant trends can be interpreted as tracers of the climate state, hence these regions could change their states, e.g. from dry to humid or from moist to dry. However long-term oscillations cannot be excluded.

For the whole globe the increasing trend is non-significant when taking into account the 1997/1998 El Niño event, which is seen in the globally averaged data as strong positive H<sub>2</sub>O columns from September 1997 to March 1999. Masking the El Niño time span – which should be done in this case – we find a significant H<sub>2</sub>O trend of  $0.0039 \text{ g/cm}^2 \pm 0.0015 \text{ g/cm}^2$  per year or 0.19% per year. Even stronger trends up to 5% per year are observed on local scales.

The H<sub>2</sub>O column is changing and the human impact on this is not clear. Though the anthropogenic intervention in nature is beyond all question, on the one hand humans irrigate fields (which has a direct effect on the atmospheric H<sub>2</sub>O columns reported by Boucher et al. (2004)) for agriculture, on the other hand they drain swamps. Woods are deforested and grassland is concreted. Diamond (2005) refers to drastic anthropogenic interventions such as deforestation and high consumption of groundwater in the northwest USA (especially in Montana), where we detect significant H<sub>2</sub>O decrease. For instance Gordon et al. (2005) attribute a decrease in H<sub>2</sub>O flow of the Brazilian Amazon region to 15% deforested rainforest, which is in line with our observed decreasing trends.

At present the contribution of natural and anthropogenic induced changes remains unclear.

Also the influence of the El Niño event on the trends, which can be seen in our data, is an interesting point and needs further investigation.

One can imagine that at a certain length of the time series a linear regression is not suitable, i.e. when there is a non-linear trend in the time series. Dose and Menzel (2006) describe how a changing trend over time can be estimated for long-term time series, which could be a useful method for the analysis of the extended data set comprising GOME, SCIAMACHY and GOME-2 (on MetOp) measurements.

## Appendix A

### Trend fitting and estimation of the uncertainty

The following steps show the calculation of the trend  $\omega$  and the uncertainty of the trend  $\sigma_\omega$  regarding autocorrelations. Equation (9) can be cast into compact matrix notation

$$A = \mathbf{X}\beta + N \quad (\text{A1})$$

where  $A$  is the  $\ell \times 1$  vector of observation,  $\mathbf{X}$  is a  $\ell \times 3$  matrix consisting of the constant  $C_t$ , time  $X_t$  and step function  $U_t$ .  $\beta = (\mu, \omega, \delta)$  represents the vector of unknown regression coefficients and  $N$  is the noise vector afflicted with autocorrelations.

The  $N_t$  are directly calculated from the time series (cf. Eq. (5)) and with the connection to the  $\epsilon_t$  from Eq. (4) only the  $\epsilon_t$  for  $t=1, \dots, T$  can be calculated via

$$\epsilon_t = N_t - \phi N_{t-1} \quad (\text{A2})$$

because no  $N_{-1}$  exists. Therefore the  $\epsilon_0$  has to be estimated by  $\epsilon_0 = \sqrt{1 - \phi^2} N_0$  which is motivated by the assumption

$$\frac{\sigma_\epsilon}{\sigma_N} \approx \frac{\epsilon_t}{N_t} \quad (\text{A3})$$

A matrix  $\mathbf{P}'$  is constructed which satisfies:

$$\mathbf{P}'N = \epsilon \quad (\text{A4})$$

which is in detail:

$$\begin{pmatrix} \sqrt{1-\phi^2} & 0 & 0 & \dots \\ -\phi & 1 & 0 & \dots \\ 0 & -\phi & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \cdot \begin{pmatrix} N_0 \\ N_1 \\ N_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \vdots \end{pmatrix} \quad (\text{A5})$$

so that  $N = \mathbf{P}'^{-1}\epsilon$ .

The model Eq. (A1) becomes:

$$A = \mathbf{X}\beta + \mathbf{P}'^{-1}\epsilon \quad (\text{A6})$$

Using matrix algebra, the model can be written as

$$\mathbf{P}'\mathbf{A}=\mathbf{P}'\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon} \quad (\text{A7})$$

or using the transformed variables  $\mathbf{A}^*=\mathbf{P}'\mathbf{A}$  and  $\mathbf{X}^*=\mathbf{P}'\mathbf{X}$  we have

$$\mathbf{A}^*=\mathbf{X}^*\boldsymbol{\beta}+\boldsymbol{\epsilon}. \quad (\text{A8})$$

Now we have absorbed the autocorrelations in the transformed variables  $\mathbf{A}^*$  and  $\mathbf{X}^*$  of model Eq. (A8) and we can apply a least square fit. The least square estimator can be calculated by:

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}^*\mathbf{X}^*)^{-1}\mathbf{X}^*\mathbf{A}^* \quad (\text{A9})$$

Denoting the diagonal elements of  $(\mathbf{X}^*\mathbf{X}^*)^{-1}$  with  $v_j$  the variance of  $\hat{\boldsymbol{\beta}}$  becomes:

$$\text{Var}(\hat{\beta}_j)=\sigma_\epsilon^2 v_j, \quad j=1,2,3, \quad (\text{A10})$$

where  $\sigma_\epsilon^2$  stands for the variance of the  $\epsilon_t$ . Therefore the variance of the trend estimator  $\hat{\omega}$  is

$$\sigma_{\hat{\omega}}^2=\text{Var}(\hat{\omega})=\sigma_\epsilon^2 v_2 \quad (\text{A11})$$

The variance  $\sigma_{\hat{\omega}}^2$  or the standard deviation  $\sigma_{\hat{\omega}}$ , respectively, of the trend estimator considers the length of the data ( $\ell$ ), the contained noise ( $\sigma_\epsilon$ ), the autocorrelation of the noise ( $\phi$ ) and additionally the position of the level shift ( $\vartheta$ ), but not its magnitude. Thus  $\sigma_{\hat{\omega}}$  can be written as

$$\sigma_{\hat{\omega}}=\frac{\sqrt{12}\sigma_\epsilon}{(1-\phi)\cdot[\ell(\ell^2-1)]^{\frac{1}{2}}}\cdot\frac{1}{[1-3\vartheta(1-\vartheta)]^{\frac{1}{2}}} \quad (\text{A12})$$

where  $\vartheta=T_0/\ell$  is the fraction of the data before the level shift occurs. With the assumption  $\ell(\ell^2-1)\approx\ell^3$  Eq. (A12) can be written as:

$$\sigma_{\hat{\omega}}\approx\frac{\sqrt{12}\sigma_\epsilon}{(1-\phi)\cdot\ell^{\frac{3}{2}}}\cdot\frac{1}{[1-3\vartheta(1-\vartheta)]^{\frac{1}{2}}} \quad (\text{A13})$$

The variance  $\sigma_N$  of the autocorrelated noise  $N_t$  is directly related to the variance  $\sigma_\epsilon$  of the white noise  $\epsilon_t$  by  $\sigma_N^2=\sigma_\epsilon^2/(1-\phi^2)$ , thus a good approximation is found with

$$\sigma_{\hat{\omega}}\approx\frac{\sqrt{12}\sigma_N}{\ell^{\frac{3}{2}}}\cdot\sqrt{\frac{1+\phi}{1-\phi}}\cdot\frac{1}{[1-3\vartheta(1-\vartheta)]^{\frac{1}{2}}} \quad (\text{A14})$$

More details on the estimation of the trend uncertainty can be found in Tiao et al. (1990) and Weatherhead et al. (1998).

**Acknowledgements.** SCIAMACHY is a national contribution to the ESA ENVISAT project, funded by Germany, the Netherlands and Belgium. SCIAMACHY data have been provided by ESA. This work has been funded by DLR-Bonn and by the University of Bremen. This work is part of and supported by the EU ACCENT Network of Excellence.

Edited by: I. Aben

## References

- Andersson, A., Bakan, S., Fennig, K., Grassl, H., Klepp, C. P., Schulz, J.: Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data – HOAPS – 3 monthly mean, doi:10.1594/WDC/HOAPS3.MONTHLY, 2007.
- Boucher, O., Myhre, G., and Myhre, A.: Direct human influence of irrigation on atmospheric H<sub>2</sub>O and climate, *Climate Dynam.*, 22, 597–603, 2004.
- Brocard, E.: Overview on satellite experiments which measure atmospheric water vapor, IAP Research Report, Institut für angewandte Physik, Universität Bern, Bern, Switzerland, 2006.
- Buchwitz, M., de Beek, R., Noël, S., Burrows, J. P., Bovensmann, H., Schneising, O., Khlystova, I., Bruns, M., Bremer, H., Bergamaschi, P., Körner, S., and Heimann, M.: Atmospheric carbon gases retrieved from SCIAMACHY by WFM-DOAS: version 0.5 CO and CH<sub>4</sub> and impact of calibration improvements on CO<sub>2</sub> retrieval, *Atmos. Chem. Phys.*, 6, 2727–2751, 2006, <http://www.atmos-chem-phys.net/6/2727/2006/>.
- Dai, A., Meehl, G. A., Washington, W. M., Wigley, T. M. L., and Arblaster, J. A.: Ensemble simulation of twenty-first century climate changes: Business-as-usual versus CO<sub>2</sub> stabilization, *Bull. Amer. Meteor. Soc.*, 82, 2377–2388, 2001.
- Diamond, J.: Collaps, How Societies Choose to Fail or Succeed, Viking, Penguin Group, New York, 2005.
- Dose, V. and Menzel, A.: Bayesian correlation between temperature and blossom onset data, *Glob. Change Biol.*, 12, 1451–1459, 2006.
- Edelson, R. A. and Krolik, J. H.: The discrete correlation function: A new method for analyzing unevenly sampled variability data, *The Astrophysical Journal*, 333, 646–659, 1988.
- Fahrmeir, L., Pigeot, I., and Tutz, G.: Statistik, Springer, Berlin, Germany, 5. edn., ISBN 3-540-21232-9, 2004.
- Freund, J. A., Pöschel, T., and Wiltshire, K. H.: Markovsche Analyse nasser Gemeinschaften, Logos, Berlin, ISBN 978-3-8325-1350-4, 99–110, 2006.
- Good, S. A., Corlett, G. K., Remedios, J. J., Noyes, E. J., and Llewellyn-Jones, D. T.: The Global Trend in Sea Surface Temperature from 20 Years of Advanced Very High Resolution Radiometer Data, *J. Climate*, 20, 1255–1264, 2007.
- Gordon, L. J., Steffen, W., Jörnsen, B. F., Folke, C., Falkenmark, M., and Johannessen, A. A.: Human modification of global water vapor flows from the land surfaces, *PNAS*, 102, 7612–7617, 2005.
- Häkel, H.: Meteorologie, Ulmer, Stuttgart, Germany, 4.edn., ISBN 3-8252-1338-2, 1999.
- Held, I. M. and Soden, B. J.: Water Vapor Feedback And Global Warming, *Annu. Rev. Energy Environ.*, 25, 441–75, 2000.
- Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Winden, P., and Dai, X.: Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report, Cambridge Univ. Press, 423pp., 2001.
- IPCC: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp., 2007.
- Lang, R., Williams, J. E., van der Zande, W. J., and Maurellis, A.

- N.: Application of the Spectral Structure Parameterization technique: retrieval of total water vapor columns from GOME, *Atmos. Chem. Phys.*, 3, 145–160, 2003, <http://www.atmos-chem-phys.net/3/145/2003/>.
- Maurellis, A. N., Lang, R., van der Zande, W. J., Aben, I., and Ubachs, W.: Precipitable water vapor column retrieval from GOME data, *Geophys. Res. Lett.*, 27(6), 903–906, 2000.
- Melillo, J.M.: Climate Change: Warm, Warm on the Range, *Science*, 8, 183–184, 1999.
- Noël, S., Buchwitz, M., Bovensmann, H., Hoogen, R., and Burrows, J. P.: Atmospheric Water Vapor Amounts Retrieved from GOME Satellite Data, *Geophys. Res. Lett.*, 26(13), 1841–1844, 1999.
- Noël, S., Buchwitz, M., and Burrows, J. P.: First retrieval of global water vapour column amounts from SCIAMACHY measurements, *Atmos. Chem. Phys.*, 4, 111–125, 2004, <http://www.atmos-chem-phys.net/4/111/2004/>.
- Noël, S., Buchwitz, M., Bovensmann, H., and Burrows, J. P.: Validation of SCIAMACHY AMC-DOAS H<sub>2</sub>O columns, *Atmos. Chem. Phys.*, 5, 1835–1841, 2005, <http://www.atmos-chem-phys.net/5/1835/2005/>.
- Noël, S., Mieruch, S., Buchwitz, M., Bovensmann, H., and Burrows, J. P.: GOME and SCIAMACHY global H<sub>2</sub>O columns, in: Proceedings of the First Atmospheric Science Conference, ESA Publications Division, Noordwijk, The Netherlands, <http://earth.esa.int/cgi-bin/confatmos06.pl?abstract=16%6>, 2006.
- Noël, S., Mieruch, S., Bovensmann, H., and Burrows, J. P.: A combined GOME and SCIAMACHY global H<sub>2</sub>O data set, in: ENVISAT Symposium 2007, SP\_636\_ENVISAT, ESA Publications Division, Noordwijk, The Netherlands, <http://envisat.esa.int/envisatsymposium/proceedings/contents.html>, 2007.
- Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in climate and the detection problem, *Geophys. Res. Lett.*, 33, L03718, doi:10.1029/2005GL025591, 2006.
- Scheffer, M. and Carpenter, S. R.: Catastrophic regime shifts in ecosystems: linking theory to observation, *Trends Ecol. Evol.*, 18, 648–656, 2003.
- Schlittgen, R., and Streitberg, B. H. J.: *Zeitreihenanalyse*, Oldenbourg, München, 9.edn., ISBN 3-486-25725-0, 1997.
- Stenke, A. and Grewe, V.: Simulation of stratospheric water vapor trends: impact on stratospheric ozone chemistry, *Atmos. Chem. Phys.*, 5, 1257–1272, 2005, <http://www.atmos-chem-phys.net/5/1257/2005/>.
- Tiao, G. C., Reinsel, G. C., Xu, D., Pedrick, J. H., Zhu, X., Miller, A. J., DeLuisi, J. J., Mateer, C. L., and Wuebbles, D. J.: Effects of autocorrelation and temporal sampling schemes on estimation of trend and spatial correlation, *J. Geophys. Res.*, 95, 20507–20517, 1990.
- Trenberth, K. E., Fasullo, J., and Smith, L.: Trends and variability in column-integrated atmospheric water vapour, *Climate Dynam.*, 24, 741–758, 2005.
- Wagner, T., Heland, J., Zöger, M., and Platt, U.: A fast H<sub>2</sub>O total column density product from GOME – Validation with in-situ aircraft measurements, *Atmos. Chem. Phys.*, 3, 651–663, 2003, <http://www.atmos-chem-phys.net/3/651/2003/>.
- Wagner, T., Beirle, S., Grzegorski, M., Sanghavi, S., and Platt, U.: El niño induced anomalies in global data sets of total column precipitable water and cloud cover derived from GOME on ERS-2, *J. Geophys. Res.*, 110, D15104, doi:10.1029/2005JD005972, 2005.
- Wagner, T., Beirle, S., Grzegorski, M., and Platt, U.: Global trends (1996–2003) of total column precipitable water observed by Global Ozone Monitoring Experiment (GOME on ERS-2) and their relation to near-surface temperature, *J. Geophys. Res.*, 111, D12102, doi:10.1029/2006JD006523, 2006.
- Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E.: Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res.*, 103, 17149–17161, 1998.