MesosphEO WP 3.3
Development, software set up, and documentation of merging algorithms

Stefan Lossow, KIT
Michael Kiefer, KIT

FMI-TN-MesosphEO-WP3-003
Version 1.4
Wednesday 13th December, 2017 14:20 CET

1 Preamble

Based on the considerations and discussions in WP 3.2 we focus on the time series merging of de-seasonalised water vapour and temperature data, to avoid the irreconcilable differences in the annual variation among the individual data sets. As input monthly zonal mean (using 10° bins) time series are used, preferably those produced in WP 4.3. In the following we describe the de-seasonalisation and the merging which uses a weighted approach. At last, examples of merged time series are shown.

2 De-seasonalisation

De-seasonalisation can be achieved in multiple ways. A common approach is to calculate for a given calendar month the average over several years. Subsequently this average is subtracted from the individual months contributing to the average. This approach requires that a data set covers every calendar month at least twice. The MIPAS data sets obtained with full spectral resolution from July 2002 to March 2004 do not fulfil this criterion. To keep these data sets as a part of the merging we use instead a regression approach for the de-seasonalisation. Every data set is regressed with following regression model:

$$\begin{aligned} f(t,\phi,z) =& C_{\text{offset}}(\phi,z)+ \\ & C_{\text{AO}_1}(\phi,z)\cdot\sin(2\cdot\pi\cdot t/p_{\text{AO}}) + C_{\text{AO}_2}(\phi,z)\cdot\cos(2\cdot\pi\cdot t/p_{\text{AO}})+ \\ & C_{\text{SAO}_1}(\phi,z)\cdot\sin(2\cdot\pi\cdot t/p_{\text{SAO}}) + C_{\text{SAO}_2}(\phi,z)\cdot\cos(2\cdot\pi\cdot t/p_{\text{SAO}}). \end{aligned} \tag{1}$$

The model contains an offset, the annual (AO) and the semi-annual (SAO) variation. The AO and SAO are parametrised by orthogonal sine and cosine functions. In the equation t denotes the time in years, $f(t,\phi,z)$ the fit of the time series for a given latitude band ϕ and altitude z and C are the regression coefficients of the individual model components. p_{AO}=1 year is the period of the annual variation, likewise p_{SAO}=0.5 year is the period of the semi-annual variation. To calculate the regression coefficients we follow the method outlined by *von Clarmann et al.* [2010] using the standard errors $\epsilon(t,\phi,z)$ (their inverse squared) of the monthly zonal means as statistical weights. Autocorrelation effects and empirical errors [*Stiller et al.*, 2012] were not considered in the regression. The de-seasonalised time series $y_d(t,\phi,z)$ are then given as:

$$y_d(t,\phi,z) = y(t,\phi,z) - f(t,\phi,z), \tag{2}$$

with $y(t,\phi,z)$ denoting the non-de-seasonalised (or absolute) data. For the sake of simplicity we do not assign any error to the regression fit, so that $\epsilon_d(t,\phi,z) = \epsilon(t,\phi,z)$.

3 Merging

To merge the individual data sets at a given time t (i.e. monthly mean) we calculate a weighted average:

$$y_m(t,\phi,z) = \frac{\sum_{i=1}^{n(t,\phi,z)} w_i(t,\phi,z) \cdot y_{d_i}(t,\phi,z)}{\sum_{i=1}^{n(t,\phi,z)} w_i(t,\phi,z)} \tag{3}$$

where $w_i(t,\phi,z)$ is the weighting of the individual data sets and $n(t,\phi,z)$ describes the total number of data sets to be merged. Implicitly, all parameters depend on latitude band and altitude. The weighting of the individual data sets is calculated as follows:

$$w_i(t,\phi,z) = \frac{1}{\epsilon_{d_i}(t,\phi,z)^2 + s(z)^2} \tag{4}$$

The weighting is based on two terms, i.e. the standard error of the monthly zonal means and a typical spread s among the data sets. The former term characterises a combination of noise error and natural variability. The spread term is intended to characterise systematic uncertainties of the individual data sets. In an ideal situation dedicated analyses of the systematic uncertainties would exist for all data sets. Since this is, however, not the case we choose a different approach to characterise the systematic uncertainties. For that we combine the data of all data sets. At a given time and latitude band we derive the spread (i.e. the maximum difference) among all de-seasonalised data sets. Prior to this determination we screen the data to avoid outliers and consequently overly large and unrealistic spread estimates. The screening is based on the median and the median absolute difference (MAD) [*Jones et al.*, 2012]. Data points outside the range $\{\text{median}[y_d(t,\phi,z)] \pm 7.5 \cdot \text{MAD}[y_d(t,\phi,z)]\}$ are discarded. For the merging we have tested weightings with the described latitude- and time-dependent spreads. However, any such separation does not matter much for the final result. In correspondence, we use a single spread profile, derived as the average over the spreads derived for all monthly means and latitude bands. In Fig. 1 the spread profiles $s(z)$ are shown for water vapour and temperature.

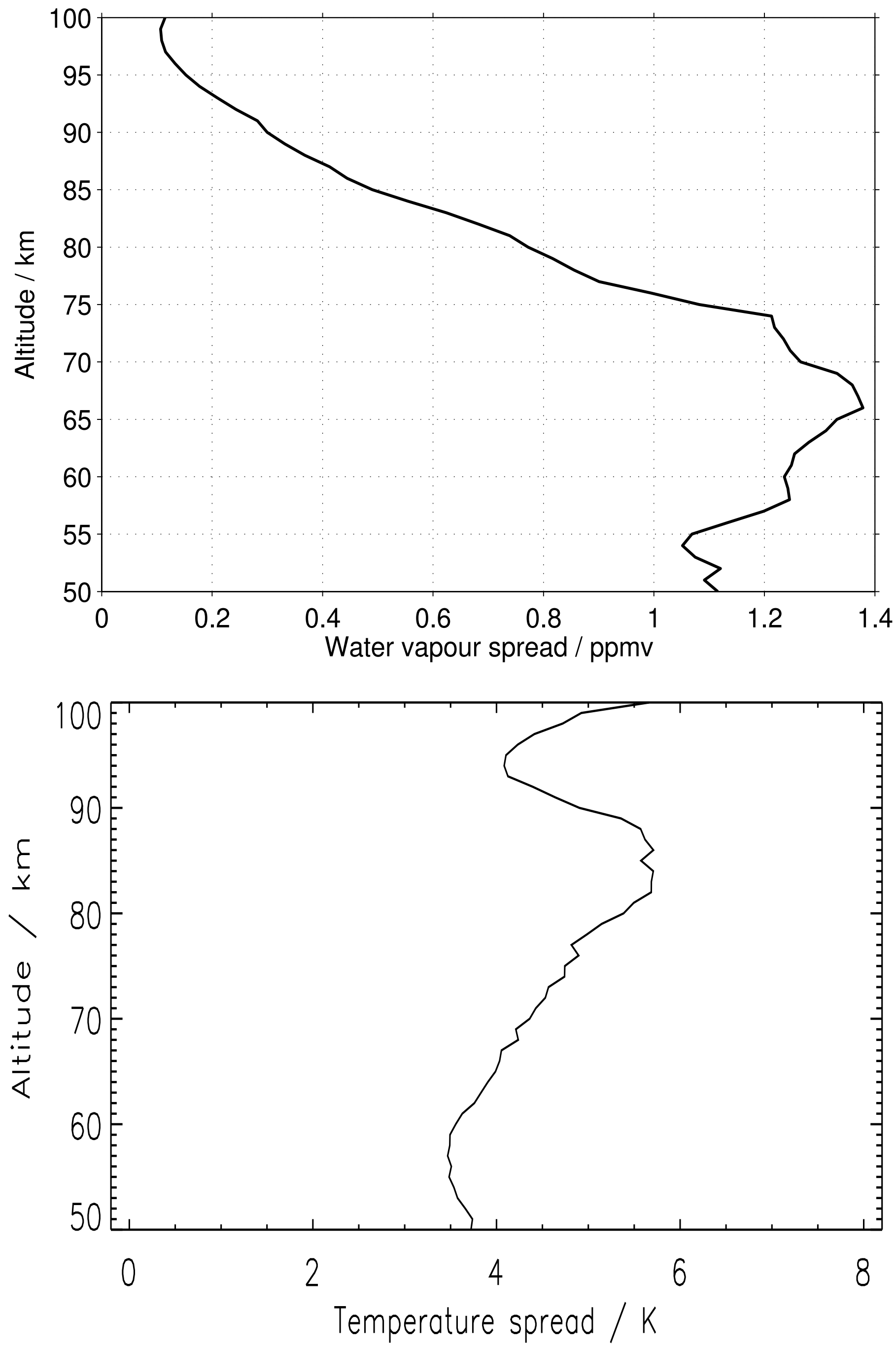

Figure 1: Spread profiles $\mathbf{s}(z)$ used to describe the systematic uncertainties of the data sets in the weighted merging for water vapour (top panel) and temperature (lower panel).

The combined error of the merged data set is given as follows:

$$\epsilon_m(t,\phi,z) = \frac{1}{W(t,\phi,z)} \cdot \sqrt{\sum_{i=1}^{n(t,\phi,z)} w_i(t,\phi,z)^2 \cdot \epsilon_{d_i}(t,\phi,z)^2} \quad \text{with} \tag{5}$$

$$W(t,\phi,z) = \sum_{i=1}^{n(t,\phi,z)} w_i(t,\phi,z).$$

This equation assumes that there no covariances between the individual data sets.

We merge the time series derived from the complete data sets. In addition there are also attempts to merge the time series separated according to daytime and nighttime. This, however, limits the number of available data sets, as the solar occultation measurements by ACE-FTS or SCIAMACHY are performed in twilight conditions and thus do not allow a meaningful separation according to the time of day. Likewise the solar scatter observations by OSIRIS and SCIAMACHY cover primarily daytime.

4 Examples

In this section some examples of merged time series are shown, focusing on water vapour and temperature. Figure 2 shows de-seasonalised water vapour time series at 55 km for the latitude band between 50°N and 60°N. Here all water vapour data sets provide data and they are given by the coloured lines. In black the merged de-seasonalised time series is overlaid, spanning the time period from 2001 to 2017. Figure 3 shows for the same latitude band the merged data as function of altitude. White areas with x-marks indicate that no data are available. While there certainly some inter-annual and longer-term variations are visible, the figure highlights also some of the issues that are encountered. One one hand, there are strong variations from month to month that appear unrealistic. On the other hand, transition regions in the employed data sets are observed, in particular the upper limit of the MIPAS observations in nominal mode (i.e. V5H and V5R NOM) close to 70 km. A similar picture arises from Fig. 4. This figure shows the de-seasonalised temperature data of five contributing datasets (see titles of the respective panels) and the merged dataset (lower right panel), again for the latitude band between 50°N and 60°N.

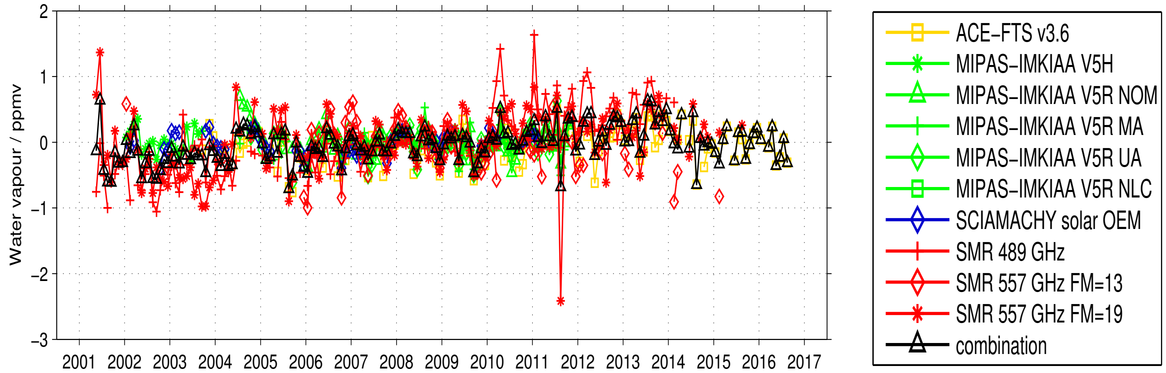

Figure 2: De-seasonalised water vapour time series for the latitude band between 50°N and 60° considering the altitude of 55 km. The coloured lines show the time series from the individual data sets while in black the merged time series is given.

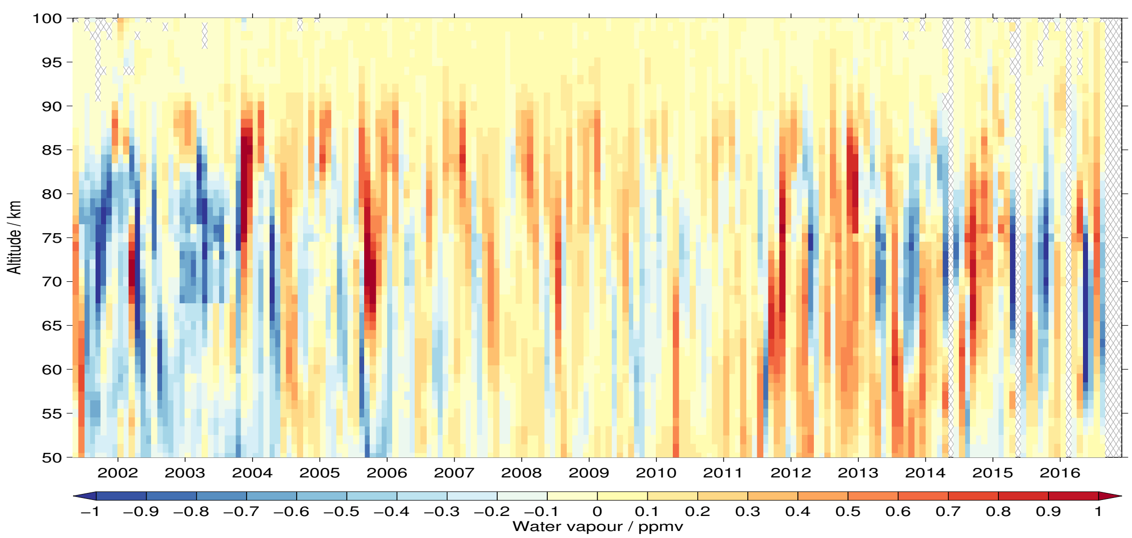

Figure 3: Contour of the merged de-seasonalised water vapour time series for the latitude band between 50°N and 60°.

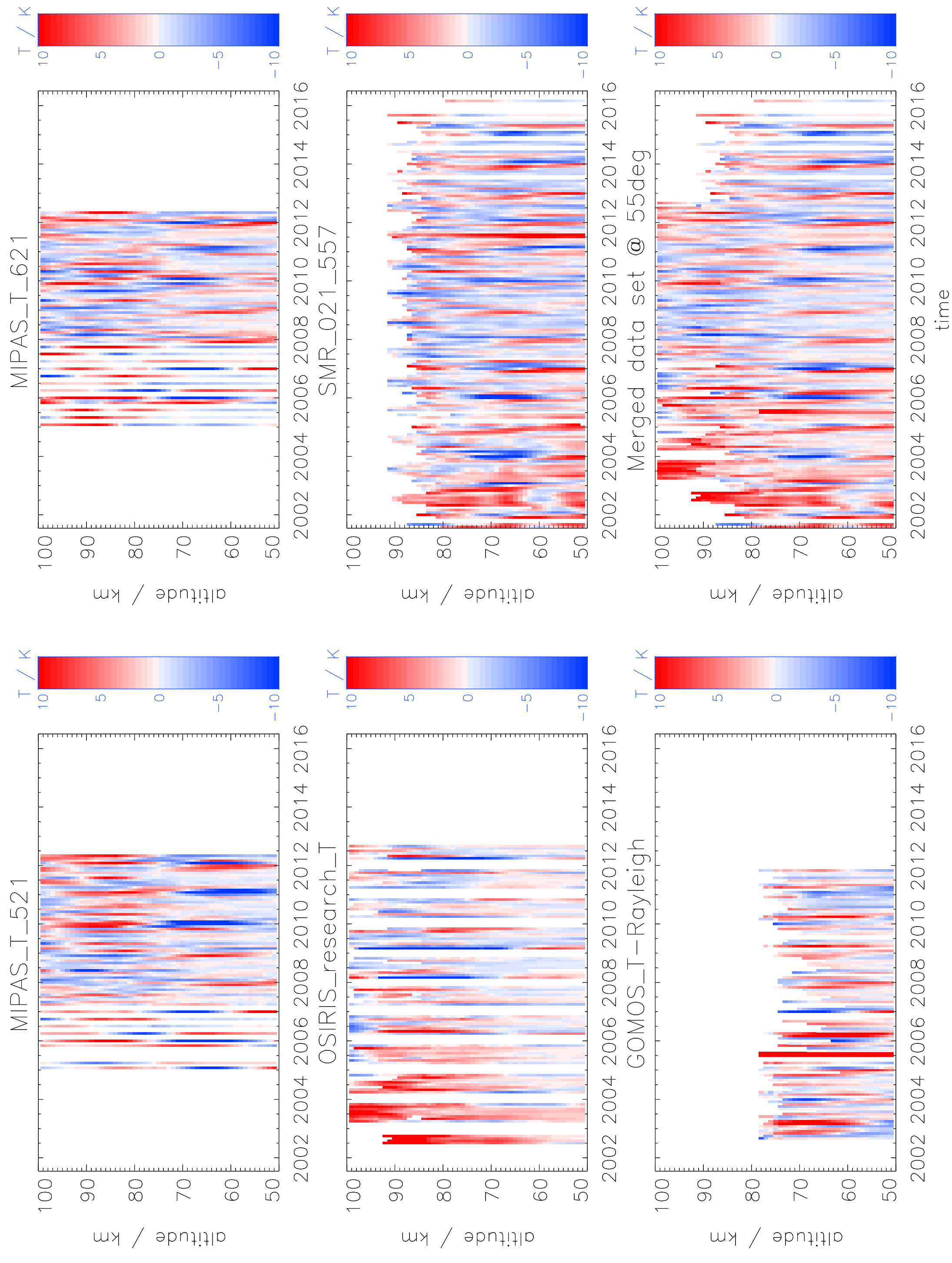

Figure 4: Contours of de-seasonalised time series of temperature for five contributing data sets (see panel titles) and the resulting merged de-seasonalised time series (all data for latitude band between 50°N and 60°).

5 Version history

version 1.0 / 14 May 2017

- the initial version

version 1.1 / 17 August 2017

- add term representing the systematic uncertainties of the data sets to the weighting used in the merging
- update figures
- more detailed equations

version 1.2 / 30 September 2017

- consideration of the comments by Alexei Rozanov, University of Bremen

version 1.3 / 27 October 2017

- add text and figures relating to temperature

version 1.4 / 13 December 2017

- small bugfixes

References

Jones, A., et al., Technical Note: A trace gas climatology derived from the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) data set, *Atmospheric Chemistry & Physics*, *12*, 5207 – 5220, doi:10.5194/acp-12-5207-2012, 2012.

Stiller, G. P., et al., Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, *Atmospheric Chemistry & Physics*, *12*, 3311 – 3331, doi:10.5194/acp-12-3311-2012, 2012.

von Clarmann, T., G. Stiller, U. Grabowski, E. Eckert, and J. Orphal, Technical note: Trend estimation from irregularly sampled, correlated data, *Atmospheric Chemistry & Physics*, *10*, 6737 – 6747, 2010.