

---

# Manual for the multilinear parametric trend estimation model

Lorena Moreira, Klemens Hocke

---



---

<sup>b</sup>  
**UNIVERSITÄT  
BERN**

Research Report No. 2017-05-MW  
October 2017

Institute of Applied Physics

Microwave Physics Division

---

Sidlerstr. 5  
CH-3012 Bern  
Switzerland

Tel. : +41 31 631 89 11  
Fax : +41 31 631 37 65  
E-mail : [iapemail@iap.unibe.ch](mailto:iapemail@iap.unibe.ch)

## Abstract

This report describes the multilinear parametric regression model used to calculate trends in middle atmospheric ozone from GROMOS ([Moreira et al., 2015](#)). The program was also utilised to analyse the effects of certain natural oscillation in stratospheric ozone ([Moreira et al., 2016](#)). Here you will find the answer to questions such as *How the model works?*, *What are the advantages compare to other regression methods?*, *How to run the program?*, *What do you need as input?* and *What will you get as output?*.

This trend estimation method was developed by Thomas von Clarmann at the Karlsruhe Institute of Technology (Germany) ([von Clarmann et al., 2010](#)). I strongly recommend you to contact him ([thomas.clarmann@kit.edu](mailto:thomas.clarmann@kit.edu)) to ask for the most recent version of the model. And in case you use his method, please include him as co-author of your work.

# Contents

1 Method	4
2 Regression function	5
3 Program	6
References	8

# 1 Method

The estimation of a trend of an atmospheric state variable is usually performed by fitting a linear regression line to a set of data of this variable (von Clarmann et al., 2010).

Assuming a linear trend, we can approximate the temporal development of an atmospheric state variable  $y$  as a straight line.

$$\hat{y} = a + bx \quad (1)$$

The  $\hat{y}$  indicates a model atmospheric state variable,  $x$  is the time of the measurement,  $a$  is the constant term and  $b$  the slope of the regression line. This straight line is the optimal regression line for which the cost function is minimum.

$$\chi^2 = (\mathbf{y} - (a\mathbf{e} + b\mathbf{x}))^T \mathbf{S}_y^{-1} (\mathbf{y} - (a\mathbf{e} + b\mathbf{x})) \quad (2)$$

where  $\mathbf{y} = (y_1, \dots, y_N)^T$ ,  $\mathbf{x} = (x_1, \dots, x_N)^T$  and  $\mathbf{e} = (1, \dots, 1)^T$ , and  $^T$  denotes the transpose of a matrix.

Therefore, the program minimizes the cost function in order to estimate the linear trend of a time series. The cost function consists of the quadratic norm of the residual between a regression model and the measured time series, weighted by the inverse covariance matrix of the data errors. The usage of the cost function along with the error covariance matrix is the main difference of this method compared to other multiple linear regression methods.

The diagonal elements of the error covariance matrix are the uncertainties of the ozone monthly mean profiles. The off-diagonal elements are initially set zero. In a second iteration, the correlation coefficients between each data point and its  $n$ th neighbour are estimated from the fit residuals, and an additional error term is built, based on these covariances, describing the deficiency of the multi-parametric model chosen. This error term (in terms of covariance matrix) is scaled according to chi-square statistics and added to the initial measurement error covariance matrix.

This generic trend method takes the entire effect of correlated residuals into account but it does not correct them. The method is particularly useful when a time series is constructed of multiple datasets where different calibration standards and other errors with systematic components have to be considered. Further, this method takes into account autocorrelated residuals which are caused by atmospheric variability on time scales larger than the sampling distance of the data but not described properly by the multilinear trend model chosen. Since the related additional autocorrelated error term can change the weight of the data points, its inclusion does not only lead to a larger, more realistic, error estimate but can slightly change the trend estimate, too.

The advantages of this scheme, that supports the use of the full covariance matrix, is that can solve problems having to do with covariances such as model deficiencies which go along with autocorrelated residuals, correlated measurement errors and biases between data subsets.

This method was used in several publications for example [Moreira et al. \(2016, 2015\)](#); [Eckert et al. \(2014\)](#); [Stiller et al. \(2012\)](#).

## 2 Regression function

The regression model consists of an axis intercept, a linear trend, sine waves, and several proxies, being the regression function:

$$\hat{y}(t) = a + b \cdot t + c_1 \cdot qbo_1(t) + d_1 \cdot qbo_2(t) + e \cdot F10.7(t) + f \cdot MEI(t) + \sum_{n=2}^m (c_n \cdot \sin(\frac{2\pi \cdot t}{l_n}) + d_n \cdot \cos(\frac{2\pi \cdot t}{l_n})) \quad (3)$$

where  $t$  is time,  $a$  and  $b$  represent the constant term and the slope of the fit, respectively. The terms  $qbo_1$  and  $qbo_2$  are the normalised Singapore zonal winds at 30 and 50 hPa as provided by the Free University of Berlin via <http://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/>. These wind series are approximately orthogonal such that their combination can emulate any Quasi-Biennial Oscillation (QBO) phase shift (Kyrölä et al., 2010). The F10.7 term stands for fitting against the solar 10.7 cm flux, which is commonly used as a measure of solar activity. MEI is the Multivariate ENSO index (MEI), which monitors the El Niño - Southern Oscillation (ENSO) phenomenon with six variables (sea-level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature, and total cloudiness fraction of the sky). Both indices are available from [www.esrl.noaa.gov](http://www.esrl.noaa.gov). The sum term comprises a number of sines and cosines functions with period length  $l_n$ , including for instance the annual and semi-annual oscillation as well as several harmonics. The selection of these harmonics can be done by calculating the power spectra of your time series in order to infer which periods affect your atmospheric state variable.

With the aim to assess the linear variation of the time series, the coefficients  $a$ ,  $b$ ,  $c_1, \dots, c_m$ ,  $d_1, \dots, d_m$ ,  $e$  and  $f$  are fitted to the ozone monthly means using the method of von Clarmann et al. (2010), where the full error covariance matrix of mixing ratios is considered. The diagonal elements of the error covariance matrix are the uncertainties of the ozone monthly mean profiles.

Therefore, the inputs required by the trend estimation program are the ozone monthly mean profiles and their uncertainties, the proxies of QBO, solar activity cycle and ENSO updated and the period lengths of the harmonics that affect your time series. The uncertainties are calculated by the standard error of the monthly mean values. Keep in mind that this uncertainty of the natural variability is divided by the number of included data points and in case you have a large number of data points within a month this uncertainty would lead to an underestimated uncertainty of the trend estimate. In Moreira et al. (2015) is presented a solution to avoid this problem.

### 3 Program

In this section I am going to explain how to use the trend program according to the latest version (**Version 1.4.3**) of the program I have, dated September 2016.

The folder **KIT-trend** includes:

- The program: **trend\_1\_4\_3.f90**
- Control file: **control\_trend.dat**
- Input files folder: **Input\_trendf90** (make sure the proxies (**f107monthly.dat**, **MEI.txt**, **singapore.dat**) are up to date)
- Short manual on how to use trend.f90, provided by Ellen Eckert from KIT:  
**Trend\_ShortManual\_Version1\_4\_3.txt**
- An extended documentation for the control file, provided by Ellen Eckert from KIT:  
**doku\_control\_trend.txt**
- **daten.mod** and **library.mod**
- Monthly means input file: **monatsmittel.dat**
- Python script: **reformat\_input.py** needed to convert **monatsmittel.dat** into something readable for the program: **new.dat**
- Output folder: **trendoutput2** which after the trend calculation will include: **ausgabe.dat**, **zeitreihe\_ausgabe.dat**, **ausgabe\_mit\_autokorr.dat**, **zeitreihe\_ausgabe\_autokorr.dat**

First of all, you need a HDF file containing your monthly means and their uncertainties (**monatsmittel.dat**). It has to be built as follows:

1. line: minimum latitude [degree] (e.g. -50), maximum latitude (e.g. -40), number of altitudes (e.g. 60)
2. line: altitude grid (consisting of 51 altitude grid points in this case)
3. line: number of data points (e.g. number of months included in the data set)

Afterwards the data-blocks are added, which consist of 4 lines each:

1. line: name of the input file for that block (could also be 'xxx', i.e. is not used in trend estimation routine, just should not be left blank)
2. line: date-time → YYYYMMDD'T'hmmss'Z' (Y = year, M = month, D = day, h = hour, m = minute, s = second; 'T' and 'Z' are fixed), we just use months in this example, which is why hhmss = 000000

3. line: monthly mean values at each altitude grid point (-100 is the missing value)
4. line: standard error of the mean values at each altitude grid points (that  $\text{std\_err}=\text{std\_dev}/\sqrt{n}$ ), where  $\text{std\_dev}$  is the standard deviation of the sample at that altitude and  $n$  is the number of included data points at that altitude, -100 is the missing values here as well)

This pattern continues until the end of the file.

Secondly, it is necessary to convert **monatsmittel.dat** into something readable for the program **new.dat**. For example, the Python script, **reformat\_input.py**, will convert it.

To use the Python script on whichever original input file you like you might have to make it executable: `chmod u+x reformat_input.py`

Then you simply run it by typing: `./reformat_input.py`

1. When you run the script it first asks your input (**monatsmittel.dat**) and output (**new.dat**) file name.
2. Then it will ask you for:
  - integer and float missing values (these will be filled in output of the trend.f90 where there is not result)
  - unit of your data
  - altitude unit

OR you can just hit ENTER for all of them and use the default values

Then your new output file should be created, with the name you picked in the beginning. I simply chose **new.dat** but name it whatever you like. This input file is used instead of **monatsmittel.dat** to run the **trend\_1\_4\_3.f90** with using the **control\_trend.dat**. This step can be avoided if you just simple create a file similar to **new.dat** and therefore readable for the program. I made this step since I previously used older versions of the trend estimation program.

Afterwards, you have to prepare the control file (**control\_trend.dat**) according to your needs. In case any doubt arises and no answer is found in the documentation you can ask Leonie Bernet (IAP-UniBe), Klemens Hocke (IAP-UniBe), Thomas von Clarmann or someone in his group (KIT).

To run the program you just need to compile it using FORTRAM, typing:

```
gfortran -o trend trend_1_4_3.f90 -llapack -lablas
./trend
```

Finally, in the output folder (**trendoutput2**) you will find several files with your trend results ready to plot. Good luck!

## References

- E. Eckert, T. von Clarmann, M. Kiefer, G. P. Stiller, S. Lossow, N. Glatthor, D. A. Degenstein, L. Froidevaux, S. Godin-Beekmann, T. Leblanc, S. McDermid, M. Pastel, W. Steinbrecht, D. P. J. Swart, K. A. Walker, and P. F. Bernath. Drift-corrected trends and periodic variations in mipas imk/iaa ozone measurements. *Atmospheric Chemistry and Physics*, 14(5):2571–2589, 2014. doi: 10.5194/acp-14-2571-2014.
- E. Kyrölä, J. Tamminen, V. Sofieva, J. L. Bertaux, A. Hauchecorne, F. Dalaudier, D. Fussen, F. Vanhellemont, O. Fanton d’Andon, G. Barrot, M. Guirlet, T. Fehr, and L. Saavedra de Miguel. Gomos  $\text{o}_3$ ,  $\text{no}_2$ , and  $\text{no}_3$  observations in 2002–2008. *Atmospheric Chemistry and Physics*, 10(16):7723–7738, 2010. doi: 10.5194/acp-10-7723-2010.
- L. Moreira, K. Hocke, E. Eckert, T. von Clarmann, and N. Kämpfer. Trend analysis of the 20-year time series of stratospheric ozone profiles observed by the gromos microwave radiometer at bern. *Atmospheric Chemistry and Physics*, 15(19):10999–11009, 2015. doi: 10.5194/acp-15-10999-2015. URL <http://www.atmos-chem-phys.net/15/10999/2015/>.
- L. Moreira, K. Hocke, F. Navas-Guzmán, E. Eckert, T. von Clarmann, and N. Kämpfer. The natural oscillations in stratospheric ozone observed by the gromos microwave radiometer at the ndacc station bern. *Atmospheric Chemistry and Physics*, 16(16):10455–10467, 2016. doi: 10.5194/acp-16-10455-2016. URL <http://www.atmos-chem-phys.net/16/10455/2016/>.
- G. Stiller, T. von Clarmann, F. Haenel, B. Funke, N. Glatthor, U. Grabowski, S. Kellmann, M. Kiefer, A. Linden, S. Lossow, and M. López-Puertas. Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period. *Atmospheric Chemistry and Physics*, 12(7):3311–3331, 2012.
- T. von Clarmann, G. Stiller, U. Grabowski, E. Eckert, and J. Orphal. Technical note: Trend estimation from irregularly sampled, correlated data. *Atmospheric Chemistry and Physics*, 10(14):6737–6747, 2010. doi: 10.5194/acp-10-6737-2010.