DETECTING TRENDS OF ANNUAL VALUES OF ATMOSPHERIC POLLUTANTS BY THE MANN-KENDALL TEST AND SEN'S SLOPE ESTIMATES -THE EXCEL TEMPLATE APPLICATION MAKESENS

Timo Salmi
Anu Määttä
Pia Anttila
Tuija Ruoho-Airola
Toni Amnell

FINNISH METEOROLOGICAL INSTITUTE

Authors
Timo Salmi, Anu Määttä, Pia Anttila,
Tuija Ruoho-Airola and Toni Amnell

Name of project Air Quality Assessment in
the Baltic countries as a consequence of
local pollution and long range transport
- a co-operation between Nordic and
Baltic countries within the framework
of the EMEP's 20 years Assessment

Commissioned by Nordic Council of Ministers

Title

Detecting trends of annual values of atmospheric pollutants by the Mann-Kendall test and
Sen's slope estimates –the Excel template application MAKESENS

Abstract

An Excel template – MAKESENS – is developed for detecting and estimating trends in the
time series of annual values of atmospheric and precipitation concentrations. The procedure
is based on the nonparametric Mann-Kendall test for the trend and the nonparametric Sen's
method for the magnitude of the trend. The Mann-Kendall test is applicable to the detection
of a monotonic trend of a time series with no seasonal or other cycle. The Sen's method
uses a linear model for the trend. The theory of the calculation, the user's manual and the
macro code are presented. As an example the long term trends of precipitation and
atmospheric concentrations of some compounds at the Virolahti air quality monitoring
station of the Finnish Meteorological Institute are calculated and briefly discussed.

Table of Contents

1   Introduction

An Excel template MAKESENS (<u>Ma</u>nn-<u>Ke</u>ndall test for trend and <u>Sen's</u> slope estimates) is developed for detecting and estimating trends in the time series of the annual values of atmospheric and precipitation concentrations. The need for this kind of simple and easy- to-use tool came up in the research project "Air quality Assessment in the Baltic countries as a consequence of local pollution and long range transport - a co-operation between Nordic and Baltic countries within the framework of EMEP's 20-years Assessment" financed by the Nordic Council of Ministers. This subproject is a contribution to the EMEP's (Evaluation of the Long-Range Transmission of Air Pollutants in Europe) twenty years of assessment work initiated by the EMEP's Task Force on Measurements and Modelling in Vienna October 2000 (TFMM 2000). This European-wide effort will assess the outcome of the emission control measures in Europe between 1980-2000, focusing on the significance of changes in concentrations and deposition in relation to emission changes.

The calculation of the Mann-Kendall test and the nonparametric Sen's method are briefly described here. We also present a detailed user's manual of MAKESENS, its macro code and some examples. In addition to the statistical calculations, MAKESENS also provides a simple graphical interface to assist the visual inspection of the time series and the statistical results. The details of the theories can be found in Gilbert (1987). Also Sirois (1998) gives an educative summary of the application of these methods in atmospheric chemistry studies.

2   Calculation of the Mann-Kendall test and the magnitude of the trend with the Sen's method in MAKESENS

MAKESENS performs two types of statistical analyses. First the presence of a monotonic increasing or decreasing trend is tested with the nonparametric Mann-Kendall test and

secondly the slope of a linear trend is estimated with the nonparametric Sen's method (Gilbert 1987). These methods are here used in their basic forms; the Mann-Kendall test is suitable for cases where the trend may be assumed to be monotonic and thus no seasonal or other cycle is present in the data. The Sen's method uses a linear model to estimate the slope of the trend and the variance of the residuals should be constant in time. These methods offer many advantages that have made them useful in analysing atmospheric chemistry data. Missing values are allowed and the data need not conform to any particular distribution. Besides, the Sen's method is not greatly affected by single data errors or outliers.

2.1   Mann-Kendall test

The Mann-Kendall test is applicable in cases when the data values x_i of a time series can be assumed to obey the model

$$x_i = f(t_i) + \varepsilon_i \ ,$$
(1)

where $f(t)$ is a continuous monotonic increasing or decreasing function of time and the residuals ε_i can be assumed to be from the same distribution with zero mean. It is therefore assumed that the variance of the distribution is constant in time.

We want to test the null hypothesis of no trend, H_o, i.e. the observations x_i are randomly ordered in time, against the alternative hypothesis, H_1, where there is an increasing or decreasing monotonic trend. In the computation of this statistical test MAKESENS exploits both the so called S statistics given in Gilbert (1987) and the normal approximation (Z statistics). For time series with less than 10 data points the S test is used, and for time series with 10 or more data points the normal approximation is used.

2.1.1   Number of data values less than 10

The number of annual values in the studied data series is denoted by n. Missing values are allowed and n can thus be smaller than the number of years in the studied time series.

The Mann-Kendall test statistic S is calculated using the formula

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \mathrm{sgn}(x_j - x_k), \tag{2}$$

where x_j and x_k are the annual values in years j and k, $j > k$, respectively, and

$$\mathrm{sgn}(x_j - x_k) = \begin{cases} 1 & if\ x_j - x_k > 0 \\ 0 & if\ x_j - x_k = 0 \\ -1 & if\ x_j - x_k < 0 \end{cases}. \tag{3}$$

If n is 9 or less, the absolute value of S is compared directly to the theoretical distribution of S derived by Mann and Kendall (Gilbert, 1987). In MAKESENS the two-tailed test is used for four different significance levels α : 0.1, 0.05, 0.01 and 0.001. At certain probability level H_0 is rejected in favour of H_1 if the absolute value of S equals or exceeds a specified value $S_{\alpha/2}$, where $S_{\alpha/2}$ is the smallest S which has the probability less than $\alpha/2$ to appear in case of no trend. A positive (negative) value of S indicates an upward (downward) trend.

The minimum values of n with which these four significance levels can be reached are derived from the probability table for S as follows.

Significance level α	required n
0.1	≥ 4
0.05	≥ 5
0.01	≥ 6
0.001	≥ 7

The significance level 0.001 means that there is a 0.1% probability that the values x_i are from a random distribution and with that probability we make a mistake when rejecting H_0 of no trend. Thus the significance level 0.001 means that the existence of a monotonic trend is very probable. Respectively the significance level 0.1 means that there is a 10% probability that we make a mistake when rejecting H_0.

2.1.2 Number of data values 10 or more

If n is at least 10 the normal approximation test is used. However, if there are several tied values (i.e. equal values) in the time series, it may reduce the validity of the normal approximation when the number of data values is close to 10.

First the variance of S is computed by the following equation which takes into account that ties may be present:

$$VAR(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{q} t_p (t_p - 1)(2t_p + 5) \right]. \qquad (4)$$

Here q is the number of tied groups and t_p is the number of data values in the pth group.

The values of S and $VAR(S)$ are used to compute the test statistic Z as follows

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{VAR(S)}} & if \ S > 0 \\ 0 & if \ S = 0 \\ \dfrac{S+1}{\sqrt{VAR(S)}} & if \ S < 0 \end{cases} \qquad (5)$$

The presence of a statistically significant trend is evaluated using the Z value. A positive (negative) value of Z indicates an upward (downward) trend. The statistic Z has a normal distribution. To test for either an upward or downward monotone trend (a two-tailed test) at α level of significance, H_0 is rejected if the absolute value of Z is greater than $Z_{1-\alpha/2}$, where

10

$Z_{1-\alpha/2}$ is obtained from the standard normal cumulative distribution tables. In MAKESENS the tested significance levels α are 0.001, 0.01, 0.05 and 0.1.

2.2   Sen's method

To estimate the true slope of an existing trend (as change per year) the Sen's nonparametric method is used. The Sen's method can be used in cases where the trend can be assumed to be linear. This means that *f(t)* in equation (1) is equal to

$$f(t) = Qt + B \tag{6}$$

where Q is the slope and B is a constant.

To get the slope estimate Q in equation (6) we first calculate the slopes of all data value pairs

$$Q_i = \frac{x_j - x_k}{j - k} \ , \tag{7}$$

where $j > k$.

If there are *n* values x_j in the time series we get as many as $N = n(n-1)/2$ slope estimates Q_i. The Sen's estimator of slope is the median of these N values of Q_i. The N values of Q_i are ranked from the smallest to the largest and the Sen's estimator is

$$Q = Q_{[(N+1)/2]}, \text{ if } N \text{ is odd}$$

or $\tag{8}$

$$Q = \frac{1}{2}\left(Q_{[N/2]} + Q_{[(N+2)/2]}\right), \text{ if } N \text{ is even.}$$

A 100(1-α)% two-sided confidence interval about the slope estimate is obtained by the nonparametric technique based on the normal distribution. The method is valid for n as small as 10 unless there are many ties.

The procedure in MAKESENS computes the confidence interval at two different confidence levels; $\alpha = 0.01$ and $\alpha = 0.05$, resulting in two different confidence intervals.

At first we compute

$$C_\alpha = Z_{1-\alpha/2}\sqrt{VAR(S)}\,, \tag{9}$$

where $VAR(S)$ has been defined in equation (4), and $Z_{1-\alpha/2}$ is obtained from the standard normal distribution.

Next $M_1 = (\ N - C_\alpha\ )/2$ and $M_2 = (\ N + C_\alpha\ )/2$ are computed. The lower and upper limits of the confidence interval, Q_{min} and Q_{max}, are the M_1^{th} largest and the $(M_2 + 1)^{th}$ largest of the N ordered slope estimates Q_i. If M_1 is not a whole number the lower limit is interpolated. Correspondingly, if M_2 is not a whole number the upper limit is interpolated.

To obtain an estimate of B in equation (6) the n values of differences $x_i - Qt_i$ are calculated. The median of these values gives an estimate of B (Sirois 1998). The estimates for the constant B of lines of the 99% and 95% confidence intervals are calculated by a similar procedure.

3   User's manual for MAKESENS

The MAKESENS template was created using Microsoft Excel 97 and the macros were coded with Microsoft Visual Basic. The template consists of four worksheets: **About**, **Annual data, Trend Statistics** and **Figure**. The **About** worksheet gives general information about the template. The data of time series are entered into the **Annual data**

worksheet. The calculation macro can be activated by using the button **Calculate Trend Statistics** and the **Trend Statistics** worksheet contains the results. Finally, the original data and the statistics can be viewed numerically and visually in the **Figure** worksheet one time series at a time.

3.1   Entering the time series and other input values

All the input values are entered in the **Annual data** worksheet (see Figure 1). The years, the annual values and the names of the time series have to be typed or copied (**Paste Special/Values**) to the fixed places of the worksheet. The cells in the worksheet, in which the user may enter data have no shading. The other cells are protected and have grey shading.

- **A title to the whole data set** can be typed in cell **A3** . This title is also shown in cells **'Trend Statistics'!A3** and **'Figure'!B4.**

- **The names of the time series** are entered in cells **B13:Z13** starting from the column **B.** The maximum number of time series that can be entered is 25. The names are transposed into the cells **'Trend Statistics'!A6:A30.** The name of the time series is also shown in cell ´**Figure'!C10** and as the title of Y-axis in the figure.

- Cell **A13** is reserved for the column title "Year".

- From cell **A14** downwards the user must enter an **increasing and continuous series of years**. The range of the years shall cover all the time series but excess years in the beginning or at the end are allowed. The number of years is limited to 100 but can be extended by changing the value of the constant **MaxData** in the macro code (Appendix 1).

- The **annual values** of each time series must be entered below the name of the time series according to the years in the column **A**. An empty cell indicates a missing value. The values of a time series or the entire table of annual data can also be copied from another table.

	A	B	C	D	E	F	G	H	I	J	K
1	ANNUAL DATA										
2											
3	FI17 Virolahti 1987 - 2000			CALCULATE TREND STATISTICS							
4											
5											
6											
7											
8	Number of time series in the calculation:	9									
9	Number of annual values in the calculation:	14	14	14	14	12	11	14	12	12	0
10	Select the FIRST YEAR of the calculation:	1987	1987	1987	1987	1989	1990	1987	1989	1989	
11	Select the LAST YEAR of the calculation:	2000	2000	2000	2000	2000	2000	2000	2000	2000	
12											
13	Year	SO4 mg S	NH4 mg N	NO3 mg N	Ca mg/l	K mg/l	SO2 ug S/	SO4 ug S/	HNO3+NO	NH3+NH4 ug N/m3	
14	1987	1,46	0,73	0,48	0,69			1,68			
15	1988	1,22	0,49	0,49	0,75			1,59			
16	1989	1,19	0,66	0,47	0,56	0,23		1,22	0,42	1,07	
17	1990	1	0,45	0,44	0,54	0,19	2,01	1,27	0,38	0,95	
18	1991	1,11	0,52	0,43	0,64	0,3	2,17	1,5	0,41	0,95	
19	1992	1,16	0,72	0,5	0,66	0,31	1,31	1,2	0,35	0,93	
20	1993	0,78	0,38	0,34	0,44	0,15	1,84	1,18	0,33	0,82	
21	1994	0,84	0,41	0,36	0,43	0,22	1,61	1,04	0,34	0,75	
22	1995	0,79	0,45	0,38	0,42	0,14	1,05	0,89	0,28	0,69	
23	1996	0,85	0,43	0,43	0,41	0,12	1,51	1,09	0,35	0,8	
24	1997	0,54	0,33	0,33	0,26	0,13	0,79	0,75	0,26	0,96	
25	1998	0,49	0,34	0,3	0,27	0,2	1,02	0,81	0,29	0,85	
26	1999	0,61	0,41	0,41	0,28	0,12	0,96	0,85	0,32	1,16	
27	2000	0,66	0,43	0,39	0,27	0,13	0,66	0,67	0,28	0,95	
28											

Figure 1. The input data is entered to the **Annual data** worksheet of MAKESENS.

14

- **The first and the last year** to be handled in trend calculation must be entered in **rows 10 and 11**, above the names of the time series. Every time series must have its own starting and ending year. With these rows you can easily define different time spans for the trend statistics calculation of each time series. The starting and ending years are also shown in the cells **'Trend Statistics'!B6:C30**, and in **'Figure'!C11**.

- **The number of values** of each time series which are selected to trend calculation are shown in **row 9**. These values are also shown in the cells '**Trend Statistics'!D6:D30** and in **'Figure'!C12**. If the number of values for a time series is equal to or less than 9, the Mann-Kendall test is performed using the S statistics and the confidence interval for the Sen's slope estimate is not determined. If this number is at least 10, the Mann-Kendall test is performed using the Z statistic (normal approximation) and the 95% and 99% confidence intervals for the Sen's slope estimates are calculated.

- Cell **B8** shows **the number of the time series** from which the statistics will be calculated.

There is no input data checking in the trend-macro. If you get strange results or even errors, please check the input data you have entered. It is important that the values are entered in the appropriate cells. Some of the cells that are not allowed to be changed are protected.

The calculation of the trend statistics is started by clicking the '**Calculate Trend Statistics**' button in the **Annual data** worksheet. The Status bar near the bottom of the screen shows when the calculation process is ready. The summary table of the results will appear to the **Trend Statistics** worksheet (Figure 2) and the **Figure** worksheet (Figure 3) will be activated with the first time series.

3.2 Summary table of results

During the statistical calculations the summary table of the results in the **Trend Statistics** worksheet is updated. Results are given for each time series in rows starting from **row 6** (see Figure 2).

The values of the cell **A3** and of the area **A6:D30** are derived directly from the **Annual data** worksheet. The trend calculation procedure writes the results of the calculation to the area **E6:Q30** when you press the "**Calculate Trend Statistics"** button.

The columns in this worksheet have the following meanings:
- **Time series** (column **A**): the names of the time series are derived from the **Annual data** worksheet (from row **13**)
- **First year** (column **B**): starting year of each time series (from the **Annual data** worksheet)
- **Last year** (column **C**): ending year of each time series (from the **Annual data** worksheet)
- **n** (column **D**): the number of annual values in the calculation excluding missing values (from the **Annual data** worksheet).
- **Test S** (column **E**): If n is 9 or less, the test statistic S is displayed. The absolute value of S is compared to the probabilities of the Mann-Kendall nonparametric test for trend (Table A18 given in Gilbert 1987) to define if there is a monotonic trend or not at the level α of significance. A positive (negative) value of S indicates an upward (downward) trend. In n is larger than 9, this cell is empty.
- **Test Z** (column **F**): If n is at least 10, the test statistic Z is displayed. The absolute value of Z is compared to the standard normal cumulative distribution to define if there is a trend or not at the selected level α of significance. A positive (negative) value of Z indicates an upward (downward) trend. If n is 9 or less, this cell is empty.

Signific. (column **G**): the smallest significance level α with which the test shows that the null hypothesis of no trend should be rejected. If n is 9 or less, the test is based to

16

TREND STATISTICS

FI17 Virolahti 1987- 2000

Time series	First year	Last Year	n	Test S	Test Z	Signific.	Q	Qmin99	Qmax99	Qmin95	Qmax95	B	Bmin99	Bmax99	Bmin95	Bmax95
					Mann-Kendall trend				Sen's slope estimate							
SO4 mg S/l	1987	2000	14		-3,50	***	-0,063	-0,089	-0,034	-0,085	-0,044	1,30	1,46	1,10	1,44	1,19
NH4 mg N/l	1987	2000	14		-2,47	*	-0,016	-0,040	0,000	-0,035	-0,004	0,59	0,73	0,44	0,73	0,47
NO3 mg N/l	1987	2000	14		-2,47	*	-0,013	-0,020	0,002	-0,019	-0,004	0,48	0,52	0,40	0,51	0,45
Ca mg/l	1987	2000	14		-3,78	***	-0,034	-0,053	-0,021	-0,046	-0,025	0,69	0,82	0,60	0,79	0,61
K mg/l	1989	2000	12		-2,27	*	-0,010	-0,029	0,002	-0,023	-0,002	0,25	0,42	0,16	0,37	0,18
SO2 ug S/m3	1990	2000	11		-3,11	**	-0,135	-0,231	-0,048	-0,187	-0,060	2,51	3,22	1,55	2,92	1,68
SO4 ug S/m3	1987	2000	14		-4,05	***	-0,070	-0,096	-0,036	-0,092	-0,046	1,58	1,72	1,33	1,68	1,41
HNO3+NO3 ug N/m3	1989	2000	12		-2,82	**	-0,011	-0,023	-0,000	-0,020	-0,005	0,42	0,50	0,35	0,47	0,38
NH3+NH4 ug N/m3	1989	2000	12		-0,28		-0,005	-0,062	0,037	-0,047	0,025	0,96	1,22	0,60	1,17	0,69

About / Annual data / **Trend Statistics** / Figure /

Figure 2. The **Trend Statistics** worksheet of MAKESENS shows the results of the calculation.

the S statistic and if n is at least 10, the test is based to the Z statistic (normal approximation). For the four tested significance levels the following symbols are used in the template:

***	if trend at $\alpha = 0.001$ level of significance
**	if trend at $\alpha = 0.01$ level of significance
*	if trend at $\alpha = 0.05$ level of significance
+	if trend at $\alpha = 0.1$ level of significance

If the cell is blank, the significance level is greater than 0.1.

- **Sen's slope estimate Q** (column **H**): the Sen's estimator for the true slope of linear trend i.e. change per unit time period (in this case a year)

- **Qmin99** (column **I**): the lower limit of the 99 % confidence interval of Q ($\alpha= 0.1$)

- **Qmax99** (column **J**): the upper limit of the 99 % confidence interval of Q ($\alpha= 0.1$)

- **Qmin95** (column **K**): the lower limit of the 95 % confidence interval of Q ($\alpha= 0.05$)

- **Qmax95** (column **L**): the upper limit of the 95 % confidence interval of Q ($\alpha= 0.05$)

- **B** (column **M**): estimate of the constant B in equation (6) f(year)=Q*(year-firstYear)+B for a linear trend

- **Bmin99** (column **N**): estimate of the constant Bmin99 in equation (6) f(year)=Qmin99*(year-firstYear)+Bmin99 for 99% confidence level of linear trend

- **Bmax99** (column **O**): estimate of the constant Bmax99 in equation (6) f(year)=Qmax99*(year-firstYear)+Bmax99 for 99% confidence level of linear trend:

- **Bmin95** (column **P**): estimate of the constant Bmin95 in equation (6) f(year)=Qmin95*(year-firstYear)+Bmin95 for 95% confidence level of a linear trend:

- **Bmax95** (column **Q**): estimate of the constant Bmax95 in equation (6) f(year)=Qmax95*(year-firstYear)+Bmax95 for 95% confidence level of a linear trend

When calculating the constants B in MAKESENS the time is used in the form $t = year − firstYear$, where *firstYear* is the first year of all data in the **Annual data** worksheet.

The confidence intervals are valid only if n is at least 10 and there are not many ties (equal values). If n is less that 10, the constants Q and B for the confidence intervals are not shown in MAKESENS.

3.3 Visual viewing of data and results

After the statistics calculation the results can be viewed both numerically and visually on the **Figure** worksheet (see Figure 3). The results are shown one time series at a time. You can switch from one time series to another by pressing the button **Next Time Series** and **Previous Time Series**. The title in cell **A4** is derived from the cell **'Annual data'!A3**.
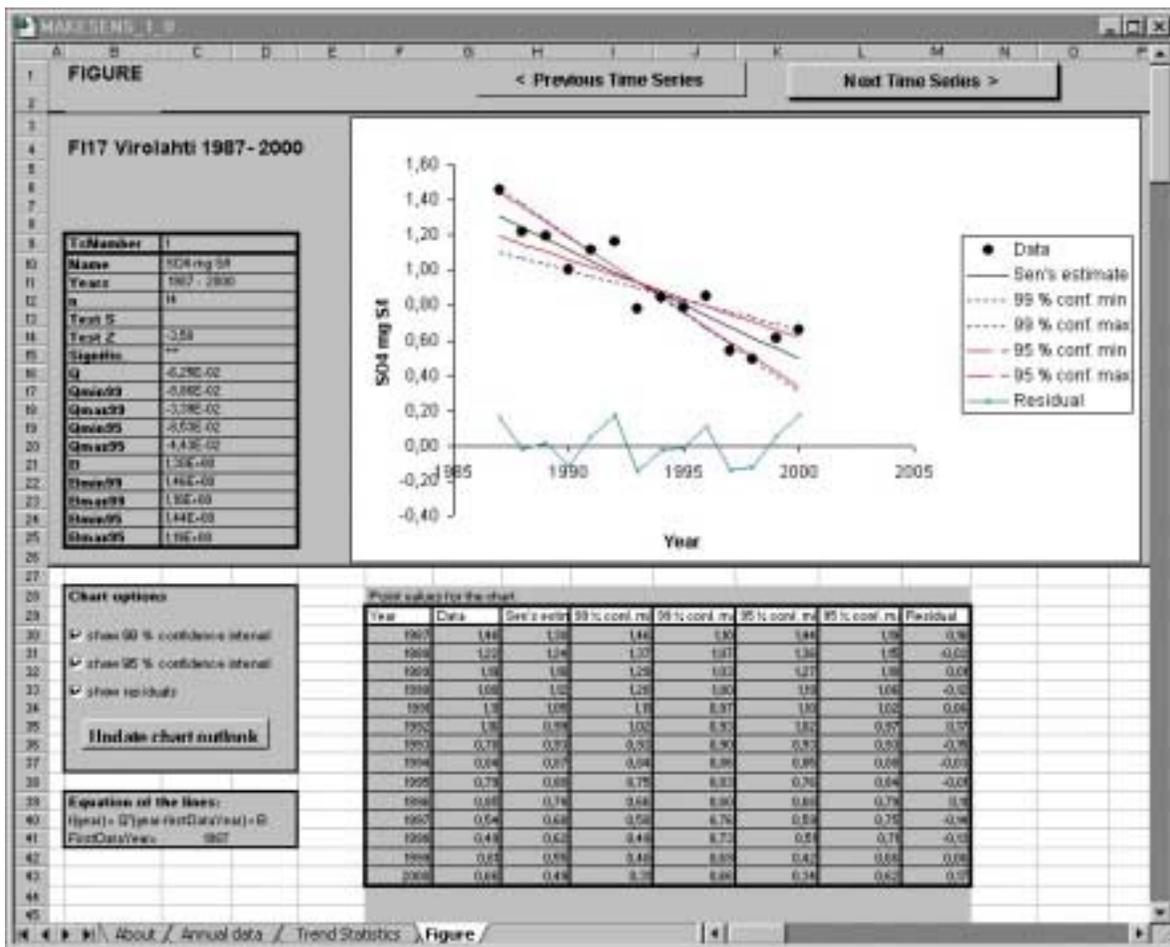

Figure 3. The **Figure** worksheet of MAKESENS.

The numerical results of the statistical calculation of one time series are shown in the area **B9:C25**. **'TsNumber'** is the ordinal number of the time series in the tables of the **Annual data** and **Trend Statistics** worksheets. The other values are derived from table **'Trend Statistics'!A5:Q30.** The first and the last year are combined to the same cell **C11**. All these values are updated when the button **Next Time Series** or **Previous Time Series** is pressed.

The table of point values for the chart is shown in columns **F:M,** from row 29 downwards. The years in column **F** and the data values in column **G** are derived from the **Annual data** worksheet. Column **H** contains the point values of the line for the Sen's estimator for a linear trend calculated with the equation $f(year) = Q*(year-firstYear) + B$. In columns **I:L** are the corresponding point values of the lines of 99% and 95% confidence intervals for the Sen's estimator. The calculated residuals (data minus trend) are shown in column **M.** All these values are updated when the **Next Time Series** or **Previous Time Series** button is pressed. The titles of the columns may be changed. The titles in the legend box of this chart are derived from these values.

In the chart the original data points of the time series, the Sen's estimator for a linear trend, the lines for 99% and 95% confidence intervals and the residuals (data minus trend) are shown.

In the **Chart options B28:D37** box you can quickly select whether the confidence intervals or residuals will be shown in the chart. The outlook of the chart can then be updated by pressing the **Update chart outlook** button. The macro recreates the chart and all the earlier chart settings are lost. The properties of the chart can also be modified directly with standard Excel tools but these modifications will be lost when the **Update chart outlook** button is pressed. The buttons **Next Time Series** or **Previous Time Series**, on the other hand, preserve all chart settings except the title of the value axis which will always be the name of the time series.

4 Examples

The graphs produced with MAKESENS serve as a visual help for the interpretation of the results given on the **Trend Statistics** worksheet. As an example the trend statistics of some time series from the Finnish EMEP station Virolahti (FI17) are shown in Figures 4 – 6 and briefly discussed here.

In Figure 4 there are two examples of annual time series: atmospheric SO_4^{2-} concentration and SO_4^{2-} concentration in precipitation which nicely fulfill the premises of the statistical methods used; The time series consist of annual averages with monotonously decreasing trends. The residuals seem to be from a random distribution indicating that a linear model may be applied. The statistical calculations give a high level of significance with narrow angles between the confidence lines.
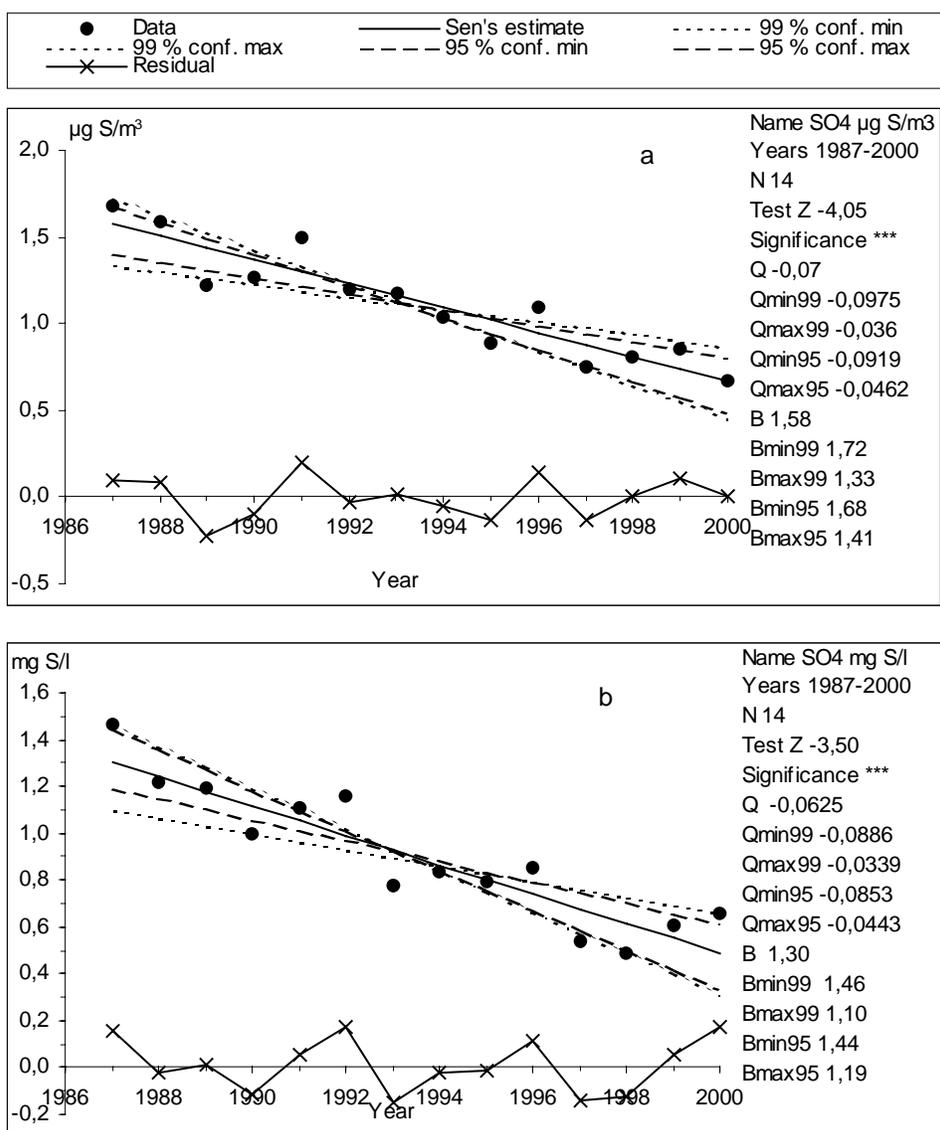
• Data	Sen's estimate	······ 99 % conf. min
···· 99 % conf. max	— — — 95 % conf. min	— — — 95 % conf. max
—×— Residual		

Chart a (μg S/m³):

Name SO4 µg S/m3
Years 1987-2000
N 14
Test Z -4,05
Significance ***
Q -0,07
Qmin99 -0,0975
Qmax99 -0,036
Qmin95 -0,0919
Qmax95 -0,0462
B 1,58
Bmin99 1,72
Bmax99 1,33
Bmin95 1,68
Bmax95 1,41

Chart b (mg S/l):

Name SO4 mg S/l
Years 1987-2000
N 14
Test Z -3,50
Significance ***
Q -0,0625
Qmin99 -0,0886
Qmax99 -0,0339
Qmin95 -0,0853
Qmax95 -0,0443
B 1,30
Bmin99 1,46
Bmax99 1,10
Bmin95 1,44
Bmax95 1,19

Figure 4. Annual time series and trend statistics of atmospheric SO_4^{2-} concentrations (a) and SO_4^{2-} concentrations in precipitation (b) at Virolahti FI17.

The case of NH_4^+ concentration in precipitation is a bit more problematic (Figure 5). The Mann-Kendall test indicates a decreasing trend at the 0.05 significance level but the Sen's slope gives non-positive slope even at the 1% confidence interval. Thus the decreasing trend seems to be more probable with the Sen's method than with the Mann-Kendall. The reason to this difference may be that the presumptions of the Sen's method are not totally

fulfilled. The Sen's method uses a linear model for the trend but here the slope seems to diminish at the end which also can be seen from the residuals which are not random. However the trend seems to be monotonic and thus the Mann-Kendall test is suitable. This time series contains also two tied groups which may reduce the accuracy of these methods with few data values as in this case.
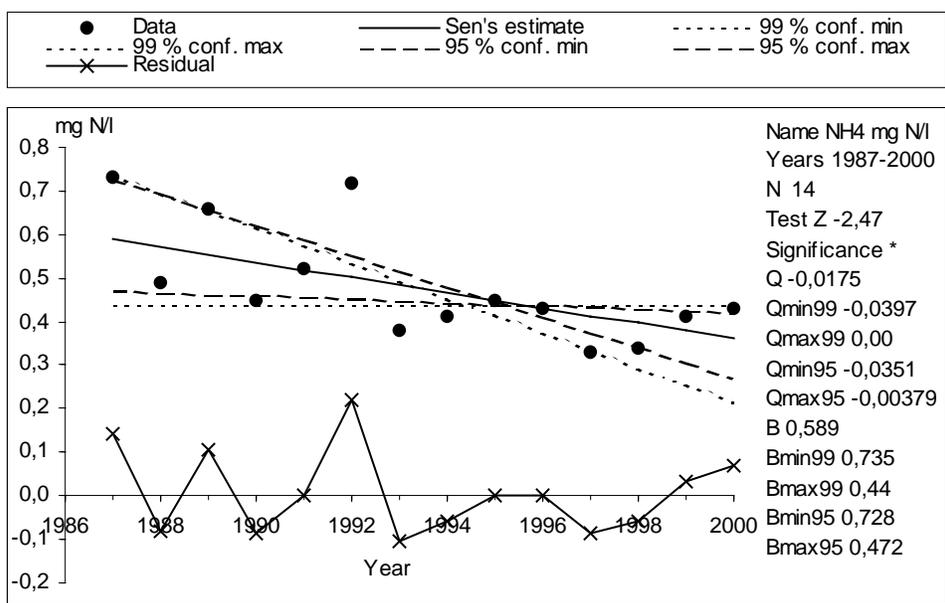

Figure 5. Annual time series and trend statistics of NH_4^+ in precipitation at Virolahti FI17.

The case of atmospheric $NH_3+NH_4^+$ concentration (Figure 6) is an example of time series to which neither of the used methods are statistically suitable. The trend is clearly first decreasing and then increasing and thus the trend is neither linear nor monotonic. This conclusion can also be made on the grounds of the residuals. However the statistical methods could be used separately to the decreasing and increasing parts of the time series but then the small number of data reduces the statistical significance of the results.
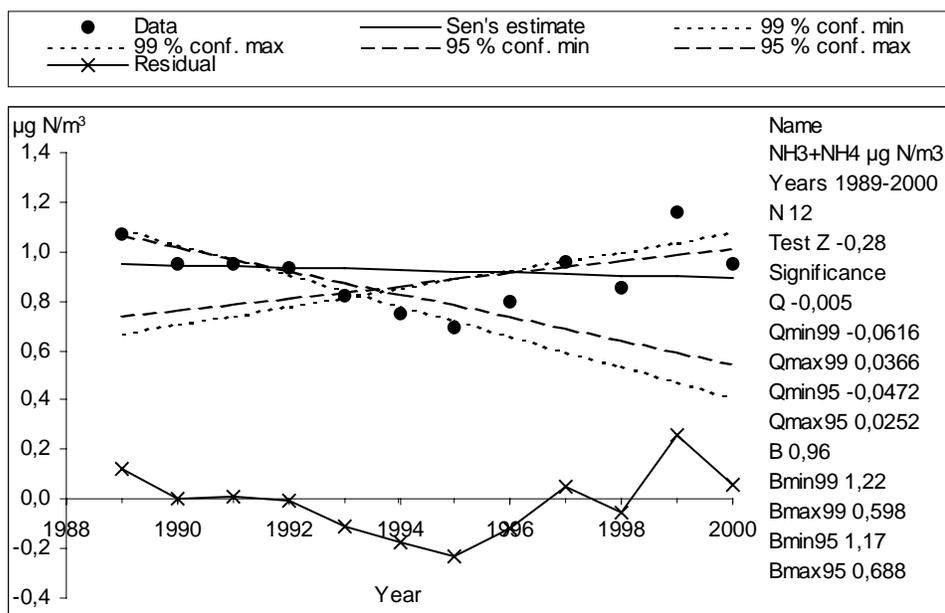

Figure 6. Annual time series and trend statistics of atmospheric $NH_3 + NH_4^+$ concentrations at Virolahti FI17.

5   Summary

An Excel template – MAKESENS – is developed for the calculation of the annual trend statistics. The procedure is based on the nonparametric Mann-Kendall test for the trend and the nonparametric Sen's method for the magnitude of the trend. The theory of calculation, the user's manual and the macro code are presented. As an example the annual trends of precipitation concentrations and atmospheric concentrations of some compounds measured at the Finnish air quality monitoring station Virolahti are calculated and briefly discussed.

Examination of the time series is very important before performing and interpreting the statistical calculations with MAKESENS. The Mann-Kendall method may be used in cases where the possible trend can be assumed to be monotonic. In the Sen's method it is assumed that the trend is linear and the residuals are from the same distribution with zero mean. The time series should fulfill these presumptions in order to produce correct statistical results with MAKESENS.

In the Mann-Kendall test missing values are allowed and the data need not conform to any particular distribution. The Sen's method is not greatly affected by gross data errors or outliers, and also it can be computed when data are missing.

Acknowledgements

6   References

Gilbert, R.O., 1987. Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold , New York.

Sirois, Allan, 1998. A Brief and Biased Overview of Time Series Analysis or How to Find that Evasive Trend. In WMO report No. 133: WMO/EMEP workshop on Advanced Statistical methods and their Application to Air Quality Data sets (Helsinki, 14-18 September 1998).

TFMM 2000. Minutes of the First Meeting of the Task Force (23 - 25 October 2000, Vienna, Austria). http://www.ubavie.gv.at/tfmm/pages/meet.htm

7 Appendix 1. The macro code of MAKESENS

```
'   Copyright 2002 Finnish Meteorological Institute
'   Timo Salmi & Anu Määttä & Toni Amnell
'   MAKESENS Version 1.0

'COMMON SETTINGS
Option Base 1    'Default indexing of arrays starts from 1
'Code for missing data in arrays:
Const MissingValue As Double = -999999#
'Maximum number of data in one time series:
Const MaxData As Integer = 100
'Codes for different significance levels:

'Minimum count of data to use normal approximation
' in Mann-Kendall test. Below this value the S statistics
' is used:
Const MinMannKendNorm As Integer = 10
'Minimum count of data to calculate confidence interval
'in Sen's method
Const MinSenConf As Integer = 10

Const S001 As String = "***"    'alpha = 0.001
Const S01 As String = "**"      'alpha = 0.01
Const S05 As String = "*"       'alpha = 0.05
Const S1 As String = "+"        'alpha = 0.1

'Arrays of critical values of Mann-Kendall statistic S
' for significance levels 0.001, 0.01, 0.05 and 0.1
' of two-sided test when n is between 4 and 10.
' The arrays will be filled by the subroutine fillS
Dim S_001(4 To 10) As Integer
Dim S_01(4 To 10) As Integer
Dim S_05(4 To 10) As Integer
Dim S_1(4 To 10) As Integer

Private Sub CB_CalculateStatistics_Click()
' The main program of calculation
' - Retrieves the data values from the sheet "Annual data" with
'   the subroutine GetData
' - Calculates the statistics with the subroutines MannKendall
'   and Sen and with the function calcB and saves the results into
'   the sheet "Trend statistics"
' - Finally calls the workbook level routines makeCollection and
'   drawFigure preparing the sheet Figure

Dim nofCol As Integer 'Number of columns i.e. time series
Dim colno As Integer  'Column number of a time series
Dim firstYear As Integer 'first year of a time series
Dim baseYear As Integer  'first year of all time series
Dim nYears As Integer    'number of years in a time series
```

```
Dim n As Integer           'true data values in a time series
                           'i.e. missing values are not considered
Dim x(MaxData) As Double 'Array for data values of a time series
Dim s As Integer           'Mann-Kendall test statistic for n=4..10
Dim Z As Double            'Mann-Kendall test statistic for n>10
Dim signif As String     'significance of trend
'Sen's slope estimator Q and its 99% and 95% confidence levels:
Dim Q As Double, Qmin99 As Double, Qmax99 As Double
Dim Qmin95 As Double, Qmax95 As Double
'Constants B for equation of lines of Sen's slope and conf. intervals:
Dim B As Double, Bmin99 As Double, Bmax99 As Double
Dim Bmin95 As Double, Bmax95 As Double


' The result cells are emptied before the calculation starts
Worksheets("Trend Statistics").Range("E6:Q30") = ""

nofCol = Worksheets("Annual data").Cells(8, 2).value
baseYear = Worksheets("Annual data").Cells(14, 1).value
Call fillS     'initializes arrays of  Mann-Kendall probabilities


'Calculation of trend statistics for each time series at a time
For colno = 2 To nofCol + 1
    If Not GetData(colno, baseYear, firstYear, nYears, n, x) Then
        Exit Sub
    End If
    If n >= 2 Then    'nothing can be computed, if n<2
    'First the existence of trend is tested using Mann-Kendall method.
        Call MannKendall(nYears, x, s, Z, signif)
        If n < MinMannKendNorm Then
            Worksheets("Trend Statistics").Cells(4 + colno, 5) = s
        Else
            Worksheets("Trend Statistics").Cells(4 + colno, 6) = Z
        End If
         Worksheets("Trend Statistics").Cells(4 + colno, 7) = signif
         'Evaluation of Sen's slope estimator and confidence intervals
        Call Sen(nYears, x, Q, Qmin99, Qmax99, Qmin95, Qmax95)
         Worksheets("Trend Statistics").Cells(4 + colno, 8) = Q
         B = calcB(nYears, x, firstYear, baseYear, Q)
         Worksheets("Trend Statistics").Cells(4 + colno, 13) = B

        If n >= MinSenConf Then
            Worksheets("Trend Statistics").Cells(4 + colno, 9) = Qmin99
            Worksheets("Trend Statistics").Cells(4 + colno, 10) = Qmax99
            Worksheets("Trend Statistics").Cells(4 + colno, 11) = Qmin95
            Worksheets("Trend Statistics").Cells(4 + colno, 12) = Qmax95
        'Coefficients B for equation of linear trend f(t)=Qt+B
            Bmin99 = calcB(nYears, x, firstYear, baseYear, Qmin99)
            Bmax99 = calcB(nYears, x, firstYear, baseYear, Qmax99)
            Bmin95 = calcB(nYears, x, firstYear, baseYear, Qmin95)
            Bmax95 = calcB(nYears, x, firstYear, baseYear, Qmax95)
            Worksheets("Trend Statistics").Cells(4 + colno, 14) = Bmin99
            Worksheets("Trend Statistics").Cells(4 + colno, 15) = Bmax99
            Worksheets("Trend Statistics").Cells(4 + colno, 16) = Bmin95
```

```
                Worksheets("Trend Statistics").Cells(4 + colno, 17) = Bmax95
            End If
        End If
Next colno

' Draw the figure of the first component
Sheets("Figure").Cells(9, 3).value = 1
Sheets("Figure").Cells(10, 3).value = Sheets("Annual data").Cells(13,
2).value
Application.Run "makeCollection"
Application.Run "DrawFigure"
End Sub 'CB_CalculateStatistics_Click

Private Function GetData(ByVal colno As Integer, ByVal baseYear As
Integer, firstYear As Integer, nYears As Integer, n As Integer, x() As
Double) As Boolean
' Retrieving of data of one time series into the array x()
' colno is the column of the worksheet "Annual data" where the
' values of the time series exist.
' The real number of annual values n in time series is calculated
' If the cell is empty it is understood as a missing value.

Dim rowno As Integer      'row of the data cell
Dim lastYear As Integer   'last year of the time series
Dim nVal As Integer       'counter for number of true data
Dim i As Integer          'counter for data loop
Dim Error As Integer
firstYear = Worksheets("Annual data").Cells(10, colno).value
lastYear = Worksheets("Annual data").Cells(11, colno).value
nYears = lastYear - firstYear + 1

nVal = 0
For i = 1 To nYears
    If firstYear < baseYear Then
        Error = MsgBox("For the time series """ + _
        Worksheets("Annual data").Cells(13, colno).value + _
        """ first year is too early!")
        GetData = False
        Exit Function
    End If
    rowno = 13 + i + firstYear - baseYear
    If IsEmpty(Worksheets("Annual data").Cells(rowno, colno)) Then
        x(i) = MissingValue
    Else
        nVal = nVal + 1
        x(i) = Worksheets("Annual data").Cells(rowno, colno)
    End If
Next i
n = nVal
GetData = True
End Function ' GetData

Private Sub MannKendall(ByVal nYears As Integer, x() As Double, s As
Integer, Z As Double, signif As String)
```

```
'Calculates the MannKendall test
'Calls the function tiedSum
'Uses the string constants S001, S01, S05 and S1

Dim absS As Integer        'value of absS
Dim varS As Double         'the variance of S
Dim absZ As Double          'value of abs(Z)
Dim k As Integer, j As Integer  'counters for slopes
Dim n As Long           'number of true values in x()

Z = MissingValue           ' returns MissingValue for Z
                           ' if they are not calculated
'Computing of the Mann-Kendall statistic S.
signif = ""
n = IIf(x(nYears) <> MissingValue, 1, 0)
s = 0
For k = 1 To nYears - 1
    If x(k) <> MissingValue Then
        n = n + 1
        For j = k + 1 To nYears
            If x(j) <> MissingValue Then
                s = s + Sgn(x(j) - x(k))
            End If
        Next j
    End If
Next k


If n < 4 Then
'If n is less than 4, the method can not be used at all
    Exit Sub
ElseIf n < MinMannKendNorm Then
'If n is between 4 and 10, S is compared directly to Mann-Kendall
statistics for S
    absS = Abs(s)
    signif = Switch(absS >= S_001(n), S001, absS >= S_01(n), S01, absS >=
S_05(n), S05, absS >= S_1(n), S1, True, "")
Else 'n>=MinMannKendNorm
'If n is at least 10, the normal distribution is used
'Firstly the variance VAR(S) is calculated
'The correction term for ties is calculated by the function tiedSum
    varS = (n * (n - 1) * (2 * n + 5) - tiedSum(nYears, x)) / 18#
'Calculation  of test statistic Z using S and its variance VAR(S)
    Z = Switch(s > 0, (s - 1) / Sqr(varS), s < 0, (s + 1) / Sqr(varS), s
= 0, 0#)
'The absolute value of Z is compared to critical value Z[1-alpha/2]
'which is obtained from the standard normal table. The presence and
'significance of the trend is evaluated by testing four different
'levels of significance: '0.001, 0.01, 0.05 and 0.1
    absZ = Abs(Z)
    signif = Switch(absZ > 3.292, S001, absZ > 2.576, S01, absZ > 1.96,
S05, absZ > 1.645, S1, True, "")
End If

End Sub 'MannKendall
```

```
Private Sub Sen(ByVal nYears As Integer, x() As Double, Q As Double,
Qmin99 As Double, Qmax99 As Double, Qmin95 As Double, Qmax95 As Double)
'Calculates Sen's slope estimator Q and its 99% (Qmax99,Qmin99)
' and 95 % (Qmax95, Qmin95)confidence levels
' Calls the function tiedSum and the subroutine
CalulateConfidenceInterval

Dim nofQ As Integer      'number of value pairs
Dim Qarray(MaxData * (MaxData - 1) / 2) As Double 'Array for the slopes
of value pairs
Dim k As Integer, j As Integer 'counters for loops
Dim n As Long            'number of true values in x()
Dim Calpha As Double     'C-alpha for calculation of conf.intervals of Q

'Computing of slopes of individual value pairs into Qarray
nofQ = 0    'used as counter for Qarray
n = IIf(x(nYears) = MissingValue, 0, 1)
For k = 1 To nYears - 1
    If x(k) <> MissingValue Then
        n = n + 1
        For j = k + 1 To nYears
            If x(j) <> MissingValue Then
                nofQ = nofQ + 1
                Qarray(nofQ) = (x(j) - x(k)) / (j - k)
            End If
        Next j
    End If
Next k

'The median of individual slopes in Qarray  is the Sen's
'slope estimator. The median is calculated by the function "median".
Q = median(nofQ, Qarray)

If n >= MinSenConf Then
'The confidence intervals are calculated only if n is at least 10.
'Computing of variance VAR(S) of Mann-Kendall statistics S.
'The correction term for ties is calculated by the function tiedSum
    varS = (n * (n - 1) * (2 * n + 5) - tiedSum(nYears, x)) / 18#

'The 100(1-alpha)% two-sided confidence intervals for the
'Sen's slope are computed with two values of alpha: 0.01 and 0.05
'which means 99% and 95% confidence intervals. The values of
'Z[1-alpha/2] are obtained from the standard normal table.
'Case alpha=0.01: Z[1-alpha/2]=Z[0.995]=2.576
    Calpha = 2.576 * Sqr(varS)
    Call CalculateConfidenceInterval(Calpha, nofQ, Qarray, Qmin99,
Qmax99)

'Case alpha=0.05: Z[1-alpha/2]=1.96
    Calpha = 1.96 * Sqr(varS)
    Call CalculateConfidenceInterval(Calpha, nofQ, Qarray, Qmin95,
Qmax95)
Else
```

```
    Qmin99 = MissingValue
    Qmax99 = MissingValue
    Qmin95 = MissingValue
    Qmax95 = MissingValue
End If
End Sub 'Sen

Private Function tiedSum(n As Integer, x() As Double) As Integer
'Calculates sum related to tied groups(= two or more equal values)
'   for the variance of Mann-Kendall statistics S
'n = number of values in the array x including missing values
'Function tiedSum is called by subroutines Sen and MannKendallNorm

Dim m As Integer           ' number of tied groups
Dim tval() As Double       ' data values of tied groups
ReDim tval(n)
Dim t() As Integer, nt As Integer ' number of data in tied groups
ReDim t(n)
Dim p, i As Integer        'indexes for the loops
Dim newValue As Boolean
Dim tSum As Integer

'Calculation of the number of tied groups m and the number of data
' in tied groups t()
m = 0
For i = 1 To n - 1
    If x(i) <> MissingValue Then
        newValue = True
        If m > 0 Then
            For p = 1 To m
                If x(i) = tval(p) Then
                    newValue = False 'this value is alredy managed
                    Exit For
                End If
            Next p
        End If

        If newValue Then
            nt = 1  'number of equal values x(i)
            For p = i + 1 To n
                If x(p) = x(i) Then
                    nt = nt + 1
                End If
            Next p

            If nt > 1 Then ' new group only if nt>1
                m = m + 1
                t(m) = nt
                tval(m) = x(i)
            End If
        End If
    End If
Next i
```

```
'Calculating the sum related to tied groups for variance
tSum = 0
If m > 0 Then
    For p = 1 To m
        tSum = tSum + t(p) * (t(p) - 1) * (2 * t(p) + 5)
    Next p
End If
tiedSum = tSum
End Function 'tiedSum


Sub CalculateConfidenceInterval(ByVal Calpha As Double, ByVal nofQ As
Integer, Qarray() As Double, lowerLimit As Double, upperLimit As Double)
'Computes confidence interval for Sen's slope estimate.
'Input parameters:  Calpha = Z[1-alpha/2],
'                    nofQ - number of slopes of all data pairs
'                    Qarray - array of slopes of all data pairs
'Subroutine returns the lowerLimit and upperLimit.
'Calls the subroutine SortArray
'Is called by the subroutine Sen
    Dim M1 As Double         'M1:th largest ordered slope
    Dim M2 As Double         'M2:th largest ordered slope
    Dim M1int As Integer     'integer part of M1 (>0)
    Dim M2int As Integer     'integer part of M2+1 (>0)
    Dim QarraySort() As Double
    ReDim QarraySort(nofQ)

    'The array Qarray is sorted to the array QarraySort
    Call SortArray(nofQ, Qarray, QarraySort)
    M1 = (nofQ - Calpha) / 2
    M2 = (nofQ + Calpha) / 2

    If M1 > 1 Then
    'to be sure that index does not point outside QarraySort
        M1int = Int(M1) 'find the integer part of M1
        'Interpolation of the lower limit
        lowerLimit = QarraySort(M1int) + (M1 - M1int) * (QarraySort(M1int
+ 1) - QarraySort(M1int))
    Else
        lowerLimit = QarraySort(1)
    End If

    If M2 < nofQ - 1 Then
    'to be sure that index does not point outside QarraySort
        M2int = Int(M2 + 1) 'because the indexing of QarraySort begins
from zero
        'Interpolation of the upper limit
        upperLimit = QarraySort(M2int) + (M2 + 1 - M2int) *
(QarraySort(M2int + 1) - QarraySort(M2int))
    Else
        upperLimit = QarraySort(nofQ)
    End If
End Sub 'CalculateConfidenceInterval
```

```
Public Function calcB(nYears As Integer, x() As Double, firstYear As
Integer, baseYear As Integer, Q As Double) As Double
' calculates the constant B for the equation of linear trend f(t)=Q*t+b.
' The zero point of time axis is the "baseYear"
' Calls the function median
Dim n As Integer     'the number of true values in time series
Dim year As Integer 'the true year of the data value
Dim i As Integer     'index for loop
Dim val() As Double 'array of differences
ReDim val(nYears)

n = 0
For i = 1 To nYears
    year = firstYear + i - 1
    If x(i) <> MissingValue Then
        n = n + 1
        val(n) = x(i) - Q * (year - baseYear)
    End If
Next i

' the estimate for B is median of the calculated differences
calcB = median(n, val)
End Function ' calcB

Private Function median(nofV As Integer, values() As Double) As Double
' calculates median of values in the array values(), indexed from 1 to
nofV
' calls the subroutine sortArray
' is called by the fuction calcB and by the subroutine Sen
    Dim i As Integer
    Dim sortedValues() As Double
    ReDim sortedValues(nofV)

    Call SortArray(nofV, values, sortedValues)
    If nofV Mod 2 = 0 Then       'nofv is even
        i = Int(nofV / 2)
        median = (sortedValues(i + 1) + sortedValues(i)) / 2
    Else                                'nOfvalues is odd
        median = sortedValues((nofV + 1) / 2)
    End If
End Function 'median

Sub SortArray(ByVal nofV As Integer, values() As Double, sortedValues()
As Double)
'This subroutine ranks the values of an array from smallest to largest.
'The sorting method is SELECTION SORT
'The ranked values are stored into the other array called sortedValues.
'Input parameters:  nofV - number of values in the array values
'                   values - values to be ranked, indexed from 1 to nofV
'Subroutine returns the sorted array at sortedValues.
'Is called by the function median and by the subroutine
'CalculateConfidence interval

Dim ind As Integer, i As Integer, j As Integer
```

```
Dim minV As Double, maxV As Double
Dim carray() As Double    'the data is first copied to this array
ReDim carray(nofV)
Dim ignoreV As Double     'value that is ignored in carray when sorting

For i = 1 To nofV         'Copy the original array to carray
    carray(i) = values(i)
Next i

'Find the smallest and largest value
ind = 1
minV = carray(1)  'initialize the smallest value
maxV = carray(1)  'initialize the largest value
For i = 2 To nofV
    If carray(i) < minV Then
        minV = carray(i)
        ind = i
    End If
    If carray(i) > maxV Then
        maxV = carray(i)
    End If
Next i

sortedValues(1) = minV 'the smallest data value
ignoreV = minV - 10    'smaller value than the smallest data value
carray(ind) = ignoreV  'this value is later ignored in sorting

'now sort the values
For j = 2 To nofV
    minV = maxV
    For i = 1 To nofV
        'find the minimum from the rest of the array
        If carray(i) <= minV And carray(i) > ignoreV Then
            minV = carray(i)
            ind = i
        End If
    Next i
    sortedValues(j) = minV
    carray(ind) = ignoreV  'from now on this element is ignored
Next j
End Sub 'SortArray

Private Sub fillS()
'Fills the arrays S_nnn of probabilities for two-tailed
' Mann-Kendall test
'The index of tables is the number of data if n=4...10
'Each array entry is an absolute value of the Mann-Kendall
' statistic S, with which the probability that there is no trend
' is less than the probability level p related to the array:
' S_001: p=0.001, S_01: p=0.01, S_05: p=0.05 and S_1: p=0.1.
' Source of values: Gilbert, 1987, Table A18
'Value 9999 indicates that the probability level can not be
' reached with given number of data
```

```
Dim n As Integer
For n = 4 To 10
    S_001(n) = 9999
    S_01(n) = 9999
    S_05(n) = 9999
    S_1(n) = 9999
Next n
S_001(7) = 21
S_001(8) = 26
S_001(9) = 30
S_001(10) = 35
S_01(6) = 15
S_01(7) = 19
S_01(8) = 22
S_01(9) = 26
S_01(10) = 29
S_05(5) = 10
S_05(6) = 11
S_05(7) = 15
S_05(8) = 18
S_05(9) = 20
S_05(10) = 23
S_1(4) = 6
S_1(5) = 8
S_1(6) = 11
S_1(7) = 13
S_1(8) = 16
S_1(9) = 18
S_1(10) = 21

End Sub
```