WETDEPNORM 1.0

- a Visual Basic program for computing wet deposition of substances and extracting anthropogenic signals from time series of deposition data

User's Manual

2005-04-14

Claudia Libiseller
Department of Mathematics
Linköping University

Introduction

en sammanfattning., kanske från artikeln

The Visual Basic programme WETDEPNORM has been developed to facilitate:

- **calculation of substance deposition;**
- **extraction of anthropogenic signals from time series of deposition data.**

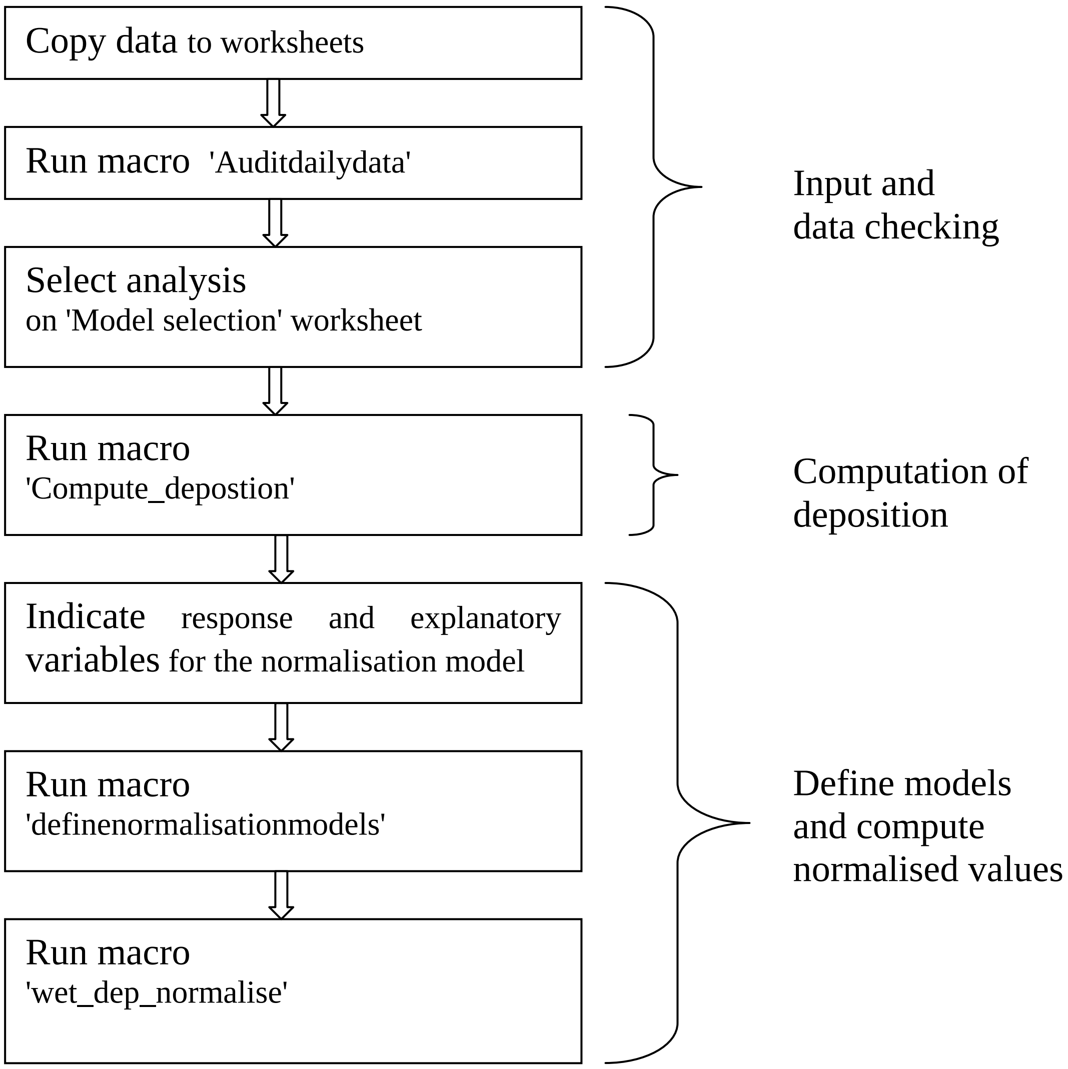
WETDEPNORM, Version 1.0, consists of five Visual Basic macros having the names and functions listed below. Two Excel worksheets containing concentration and precipitation data for the sampling site form the starting point for an analysis. It is also possible to include sector information on these worksheets. The final result consists of time series of **normalised loads**, i.e. deposition values that have been adjusted to remove natural fluctuations and clarify anthropogenic impacts. **Theoretical details** can be found in the end of this description.

Macro	Function
auditdailydata	Identify illegal entries in raw data
matchprecipandconc	Define pairs of precipitation/sector and concentration data for load calculations
compute_deposition	Compute monthly and annual deposition from precipitation and concentration data ordered by date
definenormalisationmodels	Define response variables and normalisation models
wet_dep_normalise	Select normalisation models by cross-validation and compute monthly and annual normalised deposition

Each macro operates on predefined worksheets for inputs and outputs. The table below shows which worksheets that are used for the different macros. Further details are given in the documentation of each macro.

Macro	Input worksheets	Output worksheets
auditdailydata	"Concentration by date" or "Precip and sector by date"	The same as input, and "Concentration data summary" or "Precipitation data summary"
matchflowandcons	"Concentration by date" and "Precip and sector by date"	"Matched pairs"
compute_deposition	"Matched pairs", "Precip and sector by date", and "Concentration by date"	"Annual totals" and "Seasonal or sector totals"
definenormalisationmodels	"Seasonal or sector totals"	"Normalisation models"
wet_dep_normalise	"Normalisation models"	"Normalisation models", "Normalised seasonal totals", and "Normalised annual totals"

Running the WETDEPNORM programme



Input and data checking

Copy data to worksheets

The concentration and precipitation data (and sector data if necessary) are copied to the worksheets 'Concentration by date'' and 'Precip and sector by date', respectively. Figure 1 shows how the input data shall be organised for the worksheet 'Concentration by date'. The date column shall be followed by an arbitrary number of columns containing the values of the monitored variables. Strings found in the row immediately above the first observations (row 2 in Figure 1) are interpreted as variable names. Sampling site names are extracted from the cell two lines above the first date. Missing values shall be entered as empty cells. 'Less-than-values' of the form <0.05 are permitted. Figure 2 illustrates the 'Precip and sector by date' worksheet. The format of data inserted is the same as for concentration data, but notice that precipitation data should be given in column 2 and sector in column 3. Sector data should be given as values 1-8 for classified sectors and 9 for remaining (missing or unclassified).

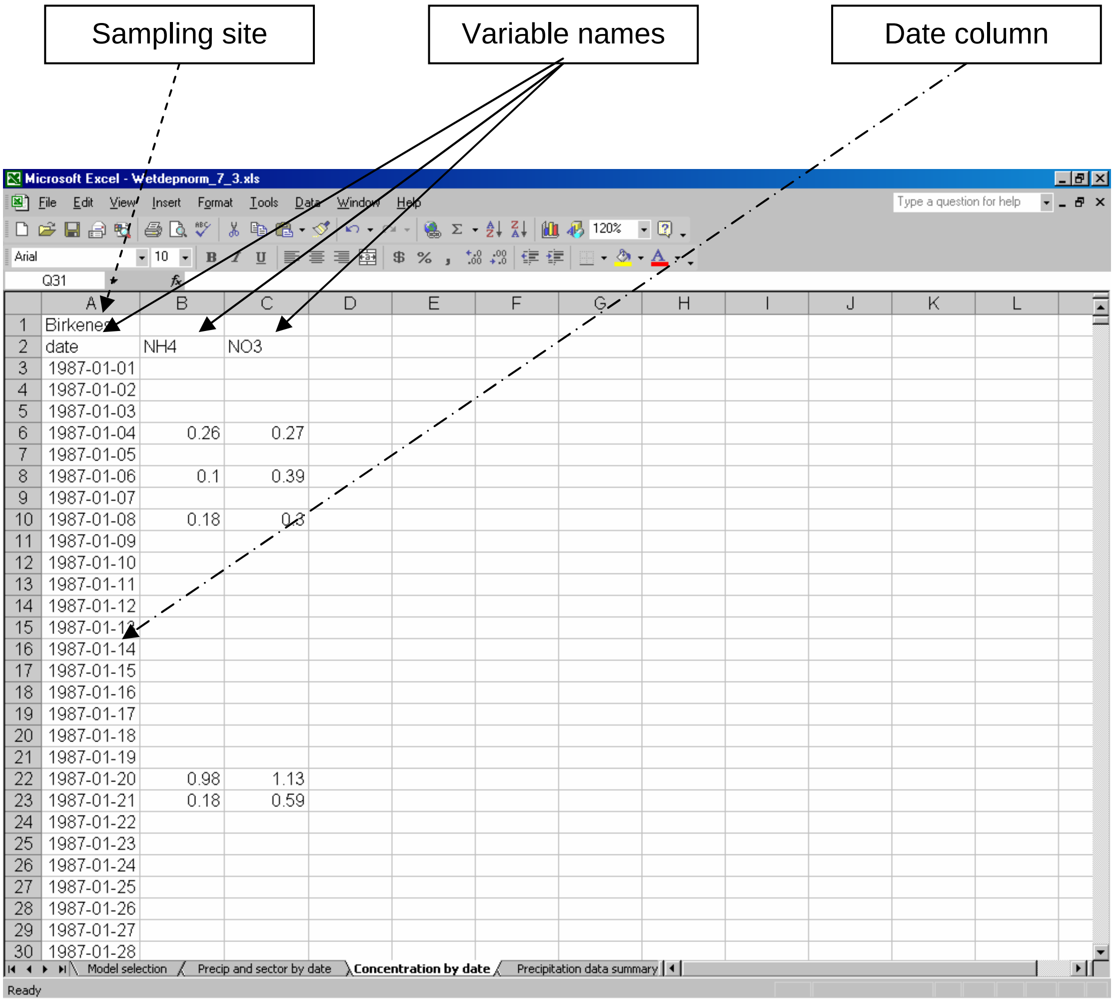


Figure 1: The worksheet ' Concentration by date'. Insert you concentration data.

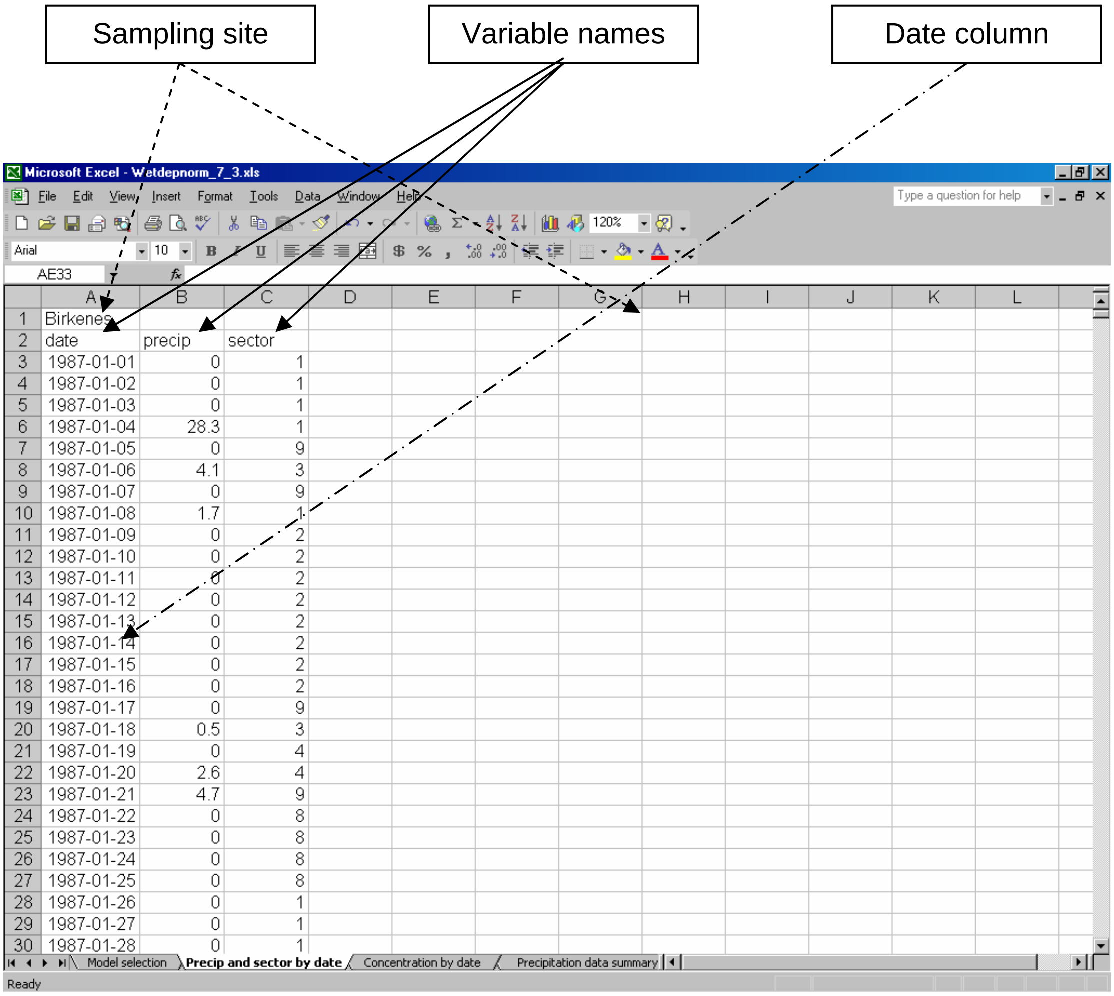


Figure 2: The 'Precip and sector by date' worksheet. Insert precipitation and sector data here. Sector data should be given in the third column, and can be omitted if the analysis is done on seasonal data only.

Run macro 'Auditdailydata'

This macro checks that the data pasted on the worksheets 'Concentration by date' and 'Precip and sector by date' are of correct type and properly organised. When 'Auditdailydata' is run, the macro first identifies the name of active worksheet ('Concentration by date' or 'Precip and sector by date'), and then the first cell containing date values.

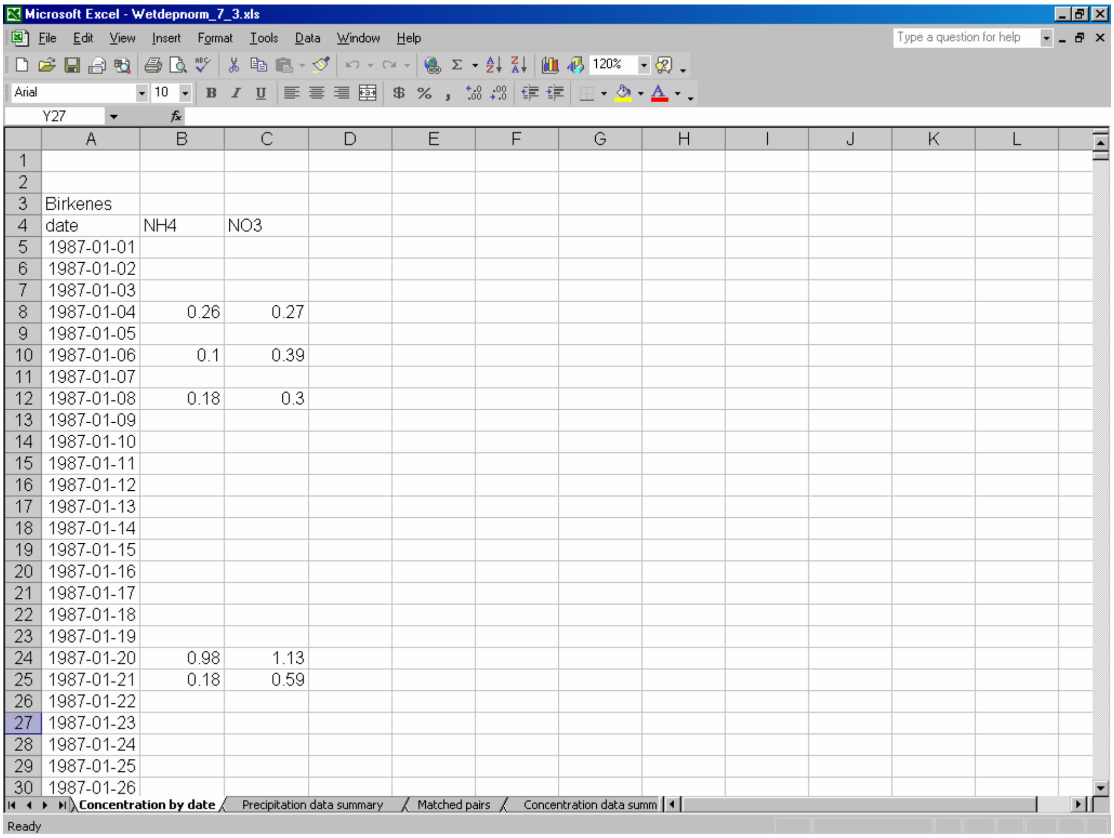


Figure 3: 'Concentration by date' worksheet after the macro 'Auditdailydata' was run. The first data row is moved to row five.

Run macro 'matchprecipandconc'

This macro facilitates the matching of precipitation/sector values and concentration data for deposition calculations. Based on the identified names of the sampling sites for flow and concentration the macro prints a preliminary list of matched pairs on the worksheet ‘Matched pairs’. This list can then be edited prior to the load calculations.

Figure 4 shows the result of the 'matchprecipandconc' macro on the 'Matched pairs' worksheet. In our dataset both series (precipitation and concentration) were labelled correctly with the site name 'Birkenes'. The macro matches the two series. If the series were labelled using different site names, but should be analysed together, the combinations on the 'Matched Pairs' worksheet can be edited now.

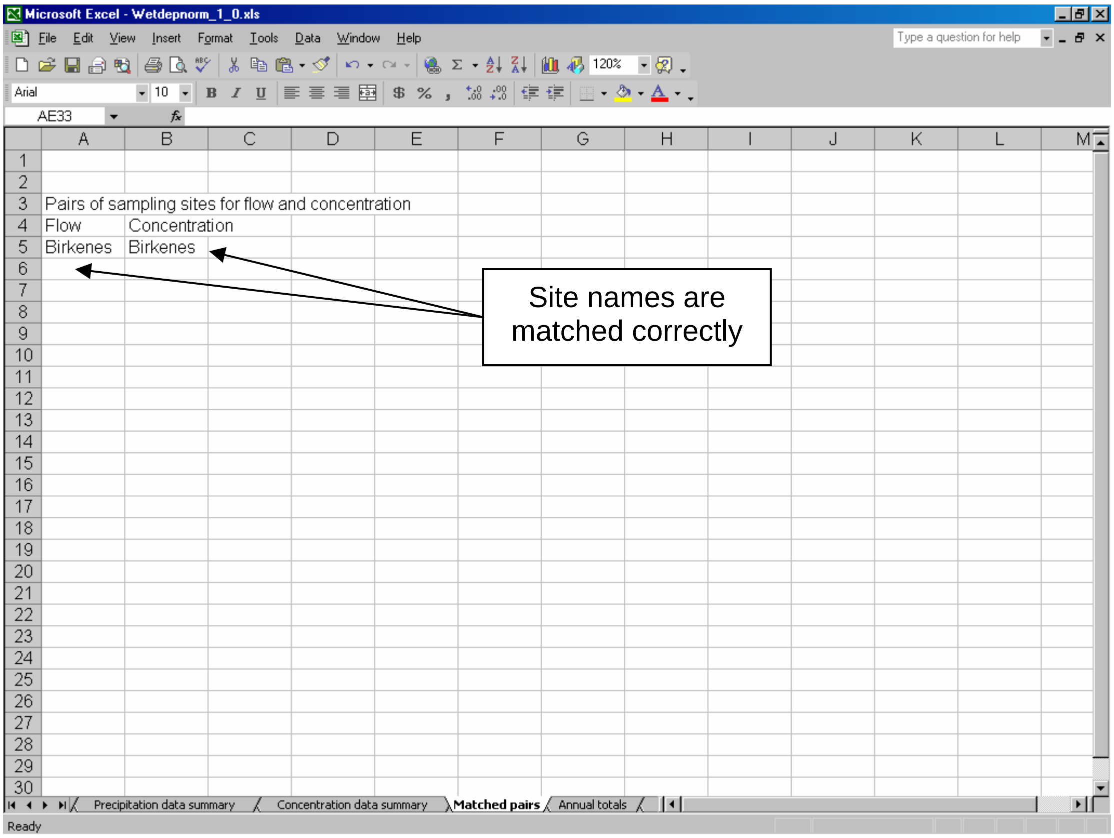


Figure 4: The matched pairs as result of the macro 'matchprecipandconc'.

Select Analysis

On the worksheet 'Model selection' you have to choose, if the computation of depositions and the normalisation of depositions should be done on annual summaries by sector or on seasonal summaries. One of the methods has to chosen and it is not possible to choose both at the same time, this implies that one of the cells must be empty, while the other contains a 'yes' or 'y'. When one of the classification methods is chosen for the computation of deposition, also the normalisation of deposition must be conducted using the same classification.

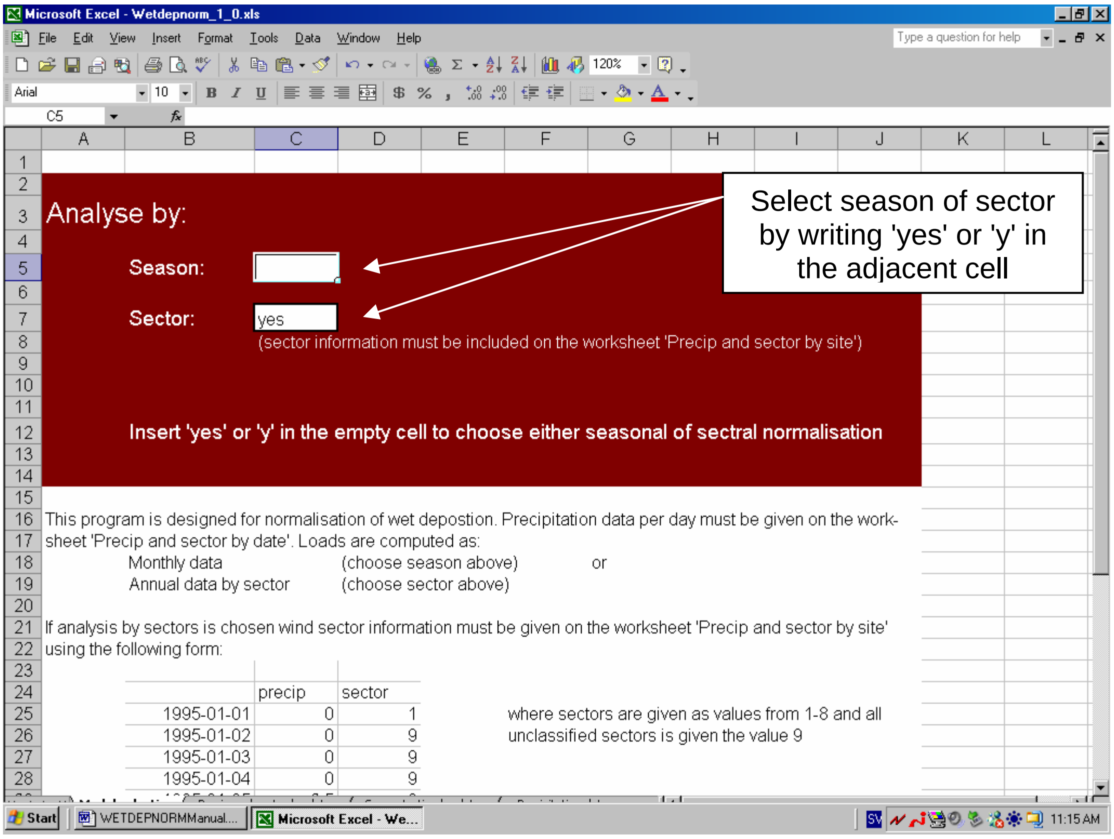


Figure 5: Select which class to use (season or sector) for computation of deposition and normalisation of deposition.

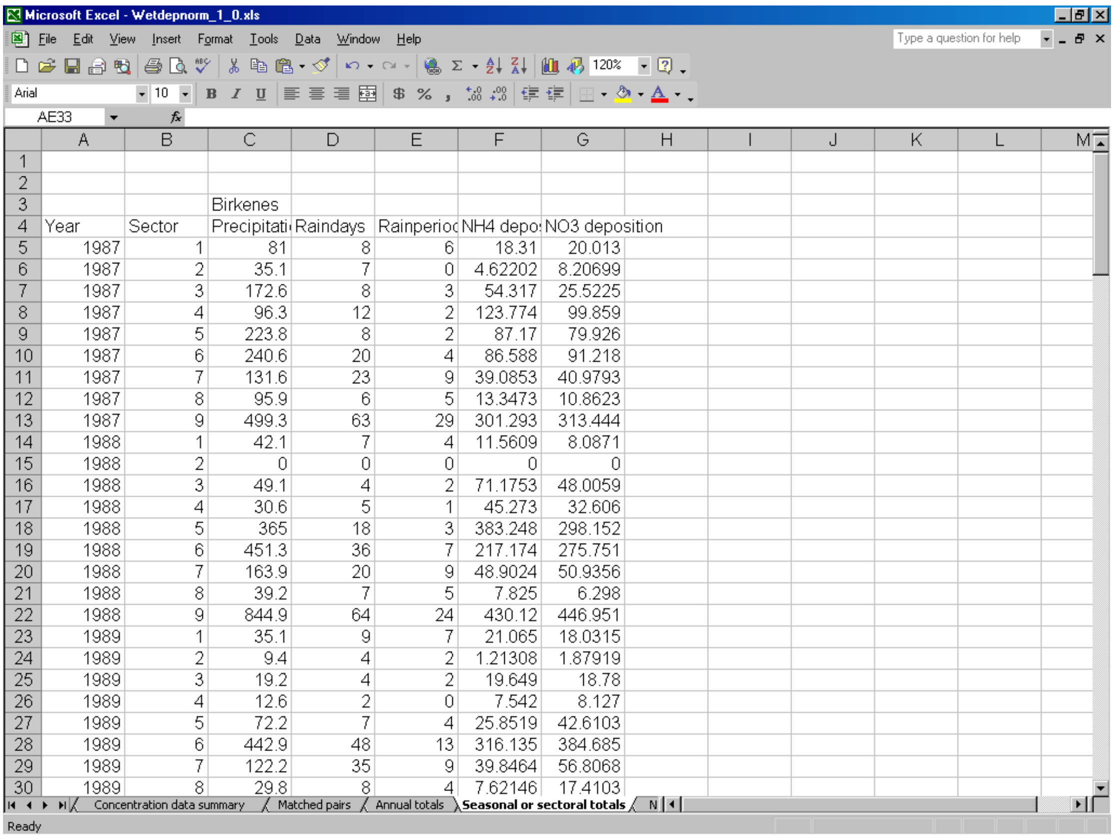
Computation of depositions

Run macro 'Compute_deposition'

This macro operates on the worksheets 'Model selection', 'Precip and sector by date', 'Concentration by date' and 'Matched pairs'. The output worksheets are 'Annual totals' and 'Monthly or sector totals'. 'Less-than-values' are replaced by a fixed percentage of the detection limit. The user is asked to enter the desired percentage when the macro is run.

In Figure 6 the computation of deposition is made by sector, giving 9 outputs per year. The response variables NH4 and NO3 concentrations and precipitation are used to compute annual deposition by sector. The explanatory variable precipitation forms three new variables: annual precipitation amount by sector, number of precipitation days per sector, number of precipitation periods (=1st day of several consecutive precipitation days) per sector.

Computation of depositions by seasons works in the same way, then 12 values are given for each year, the variables computed are the same.



	A	B	C	D	E	F	G
1							
2							
3			Birkenes				
4	Year	Sector	Precipitati	Raindays	Rainperioc	NH4 depo	NO3 deposition
5	1987	1	81	8	6	18.31	20.013
6	1987	2	35.1	7	0	4.62202	8.20699
7	1987	3	172.6	8	3	54.317	25.5225
8	1987	4	96.3	12	2	123.774	99.859
9	1987	5	223.8	8	2	87.17	79.926
10	1987	6	240.6	20	4	86.588	91.218
11	1987	7	131.6	23	9	39.0853	40.9793
12	1987	8	95.9	6	5	13.3473	10.8623
13	1987	9	499.3	63	29	301.293	313.444
14	1988	1	42.1	7	4	11.5609	8.0871
15	1988	2	0	0	0	0	0
16	1988	3	49.1	4	2	71.1753	48.0059
17	1988	4	30.6	5	1	45.273	32.606
18	1988	5	365	18	3	383.248	298.152
19	1988	6	451.3	36	7	217.174	275.751
20	1988	7	163.9	20	9	48.9024	50.9356
21	1988	8	39.2	7	5	7.825	6.298
22	1988	9	844.9	64	24	430.12	446.951
23	1989	1	35.1	9	7	21.065	18.0315
24	1989	2	9.4	4	2	1.21308	1.87919
25	1989	3	19.2	4	2	19.649	18.78
26	1989	4	12.6	2	0	7.542	8.127
27	1989	5	72.2	7	4	25.8519	42.6103
28	1989	6	442.9	48	13	316.135	384.685
29	1989	7	122.2	35	9	39.8464	56.8068
30	1989	8	29.8	8	4	7.62146	17.4103

Figure 6: Annual deposition by sector. Besides NH4 and NO3 deposition and precipitation amount by sector two new variables are added: raindays (number of precipitation days per sector) and rainperiods (number of 1st days of longer precipitation periods) per sector.

Define normalisation models and compute normalised values

Indicate response and explanatory variables

In this step the user must indicate which variables should be interpreted as explanatory and response variables, respectively. This is done by writing 'x' for explanatory variables and 'y' for response variables in the second row just above the data columns that are chosen.

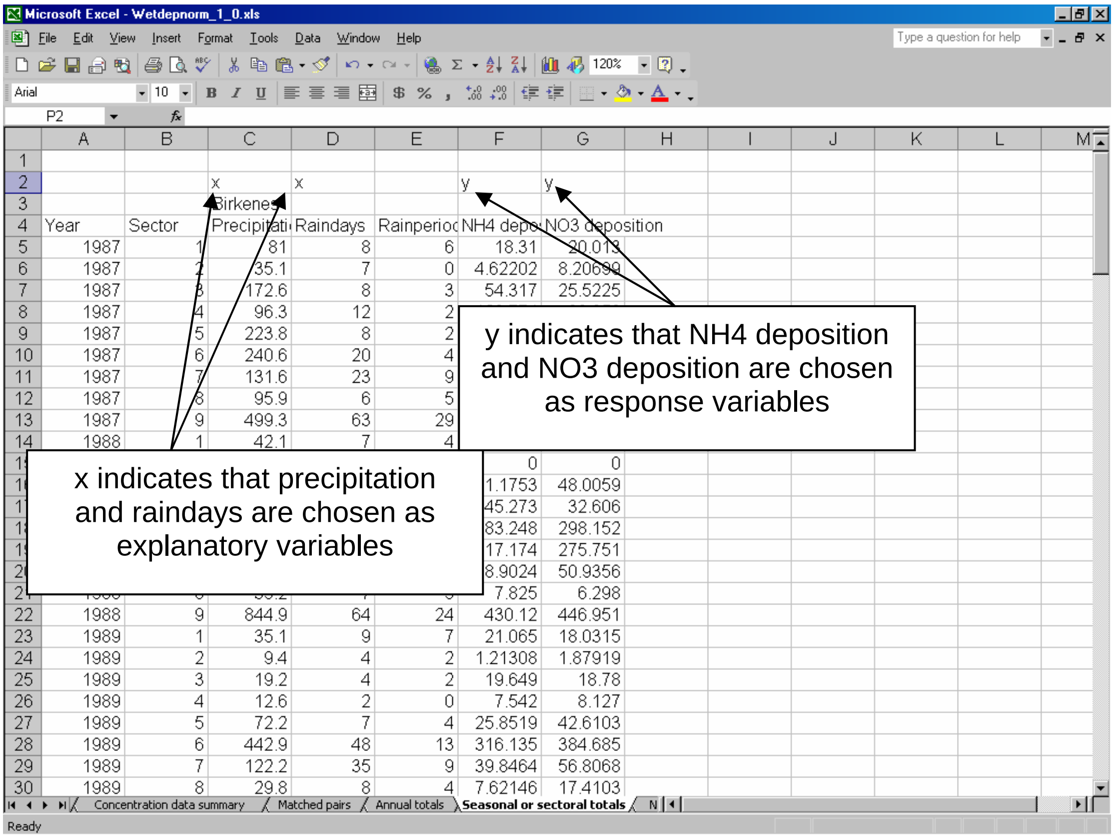


Figure 7: Indicate explanatory variables by 'x' and response variables by 'y'.

Run macro 'definenormalisationmodels'

After indicating explanatory and response variables run the macro 'definenormalisationmodels', which combines the variables in all possible ways (See Figure 8 shows the output of this macro on the worksheet 'Normalisation models'. Also this sheet can be edited prior to further analysis (Figure 9).

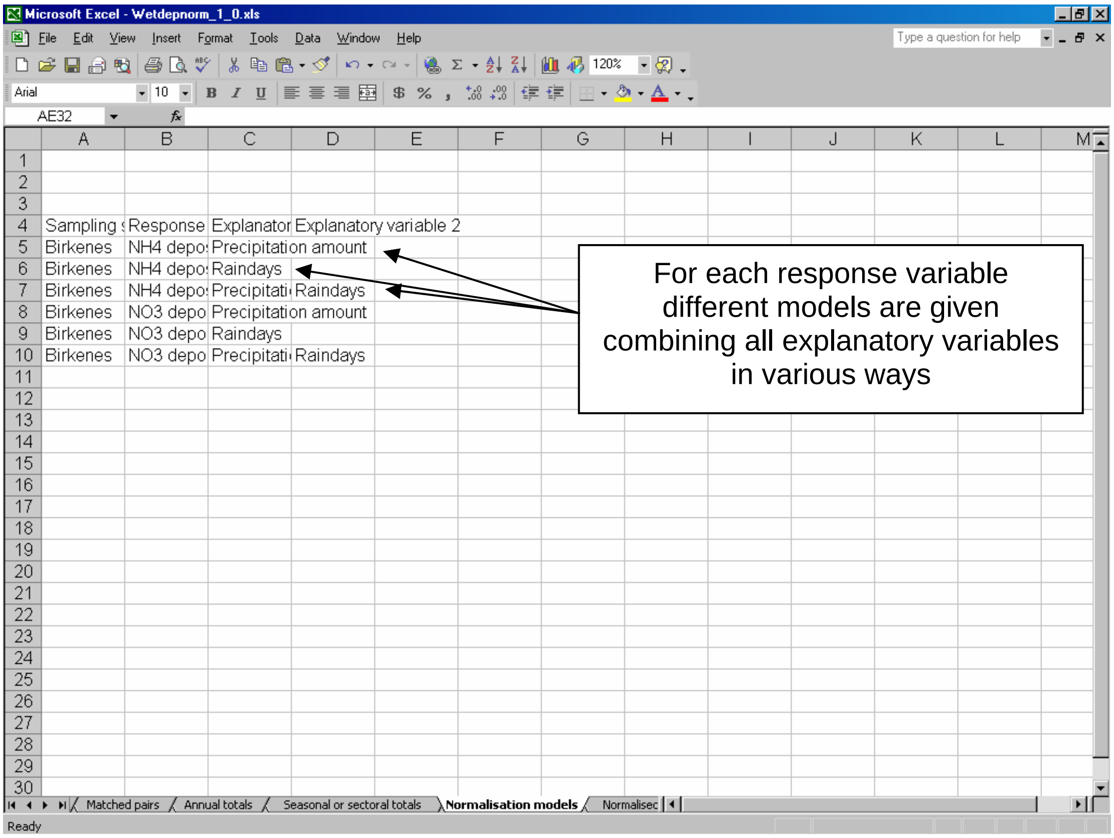


Figure 8: Output of the macro ' definenormalisationmodels' on the worksheet 'Normalisation Models'.

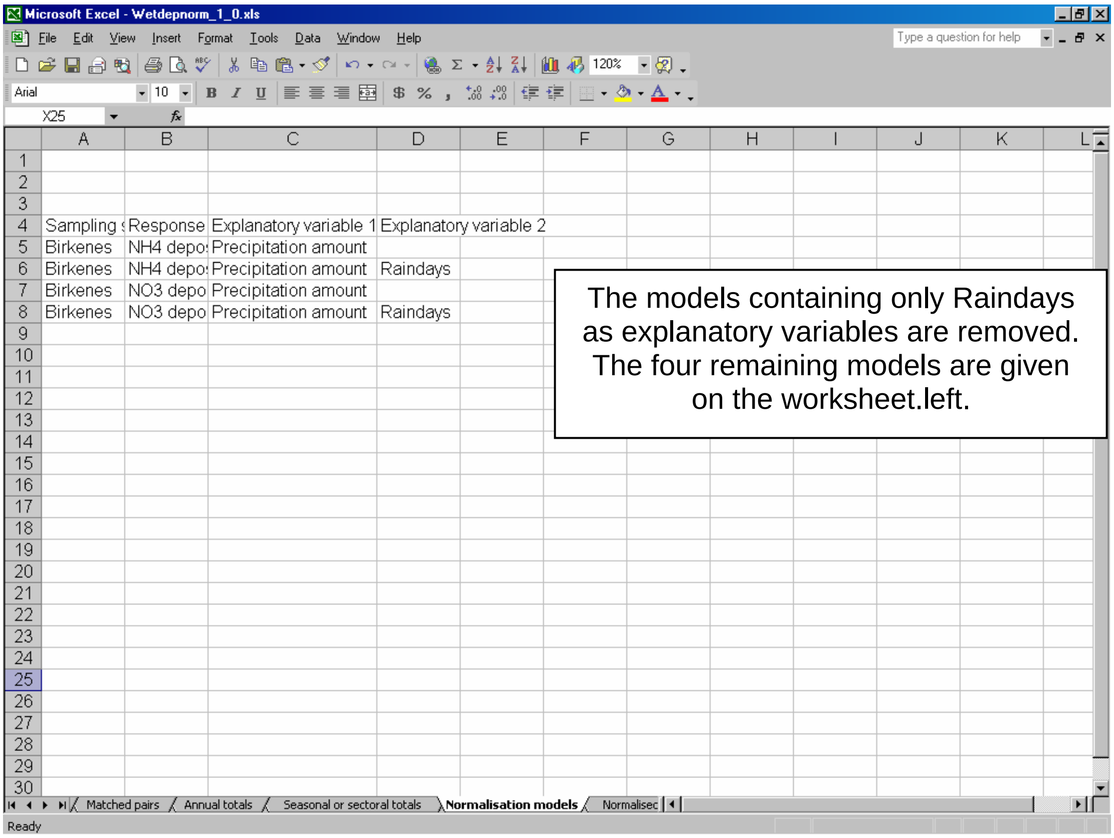


Figure 9 The worksheet 'Normalisation Models' after editing.

Run macro ' wet_dep_normalise'

This macro aims to remove or suppress the natural variation in deposition data (either monthly or annual by sector). The different normalisation models to be computed are read from the worksheet ‘Normalisation models’. Input data are read from the worksheet ‘Seasonal and sectoral totals’, and the outputs of the macro are printed on the worksheets ‘Normalised annual totals’ (see Figure 11) and ‘Normalised seasonal totals’ (Figure10).

When running the macro 'wet_dep_normalise, there are two possibilities:

- use cross-validation to determine smoothing parameters (recommended)
- insert smoothing parameters manually.

After you start the program you will see a message box asking if cross-validation should be used. Insert 'yes' for cross-validation and 'no' for manual input. If you choose cross-validation the macro will start running, if you choose manual you get a new message box asking for the first smoothing parameter (determining the smoothness between years) and after that a last message box asking for the second smoothing parameter (determining the smoothness between classes, i.e. sectors or seasons).

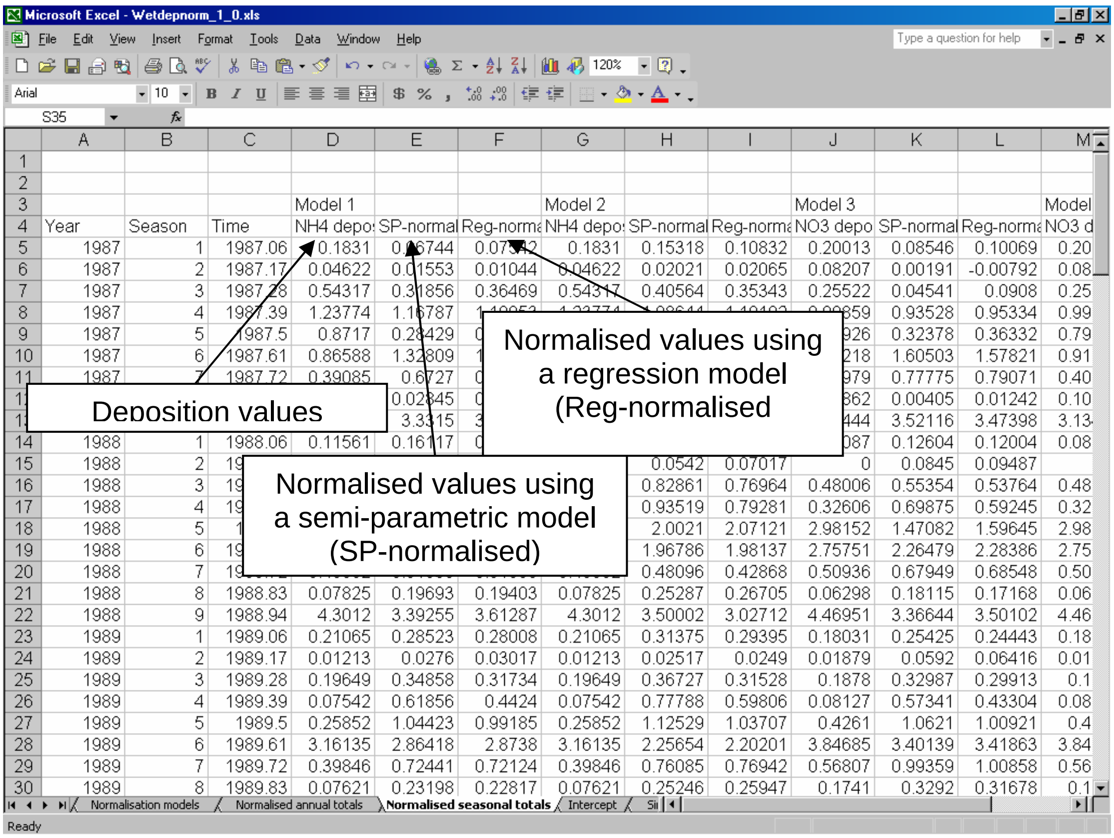


Figure 10: Output on the worksheet 'Normalised seasonal totals'. For each model defined on the worksheet 'Normalisation models' three outputs are given: deposition values, SP-normalised and Reg-normalised values. The same output can be found on the worksheet 'Normalised annual totals', but there aggregated for each year

On the worksheet 'Normalisation Models' (Figure 11) some information on the fitted models can be found. In the first columns after the chosen explanatory variables the number of observations is given. Then the smoothing parameters are listed as Lambda1 and Lambda2. Furthermore Mean squared prediction error (MSPE) and mean squared residuals (MSRes) are given for all combinations of explanatory and response variables and for both the semi-parametric and the regression approach.

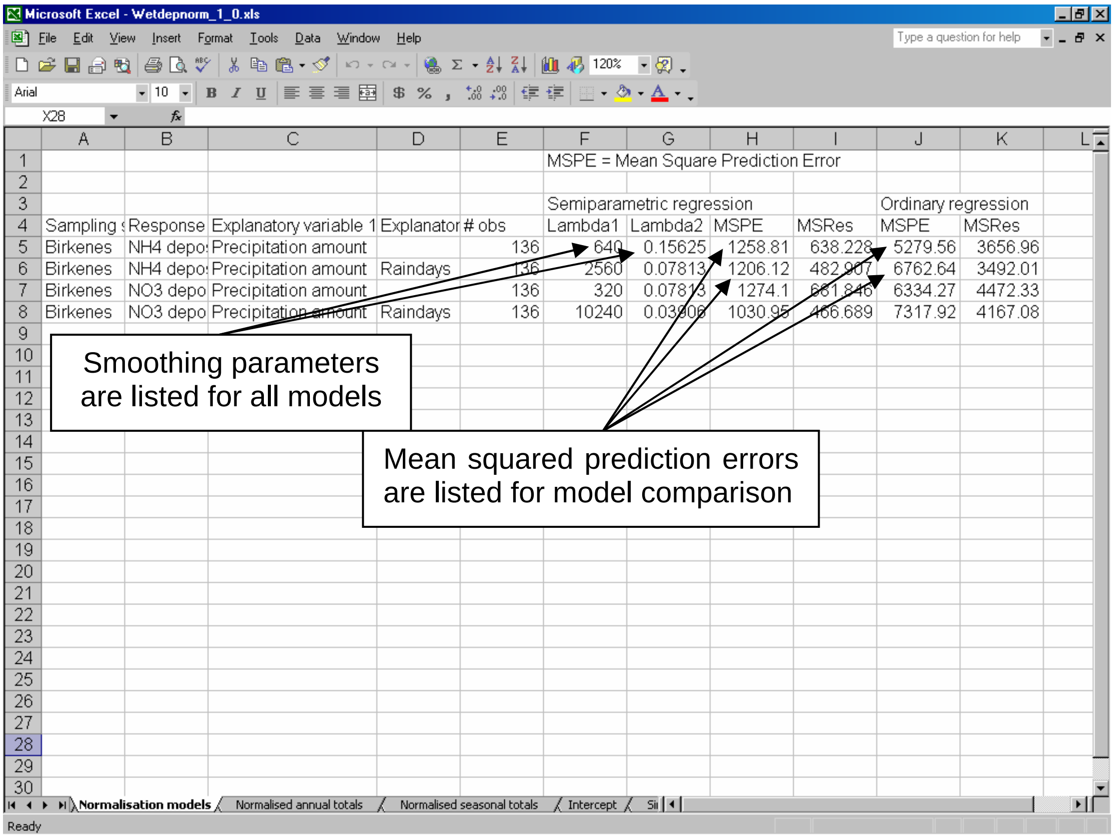
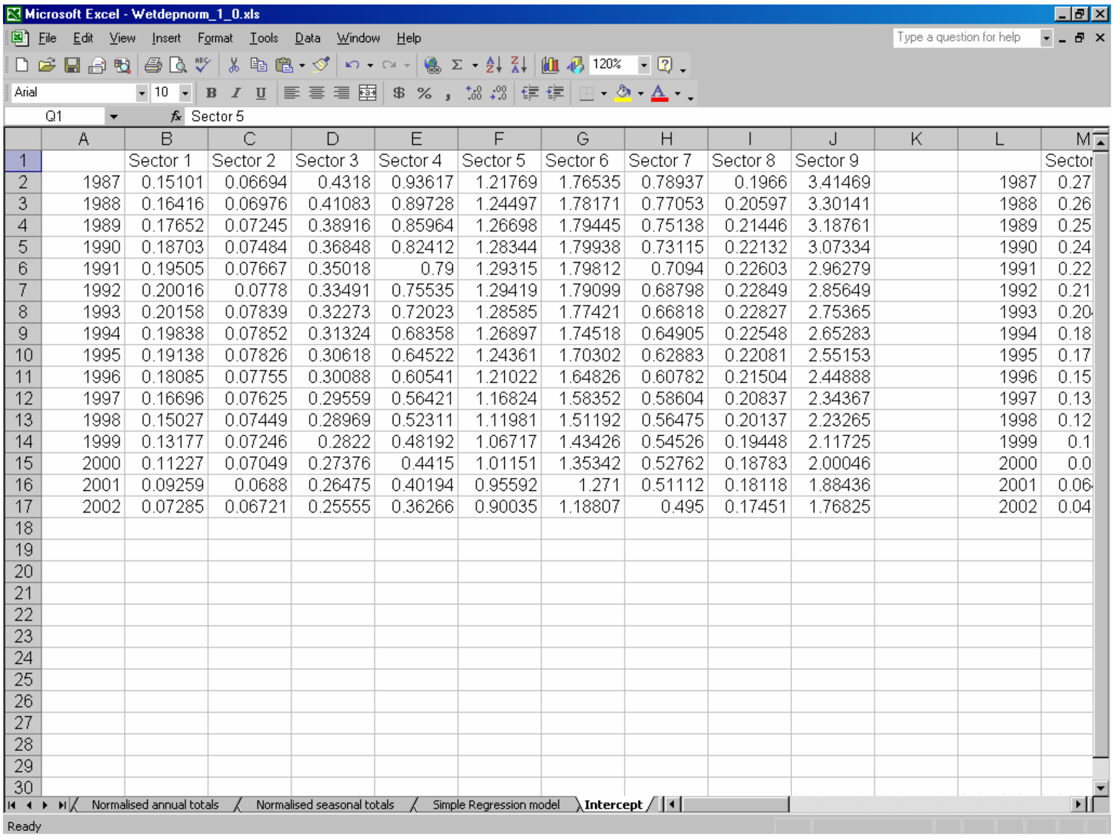


Figure 11: The output on the worksheet 'Normalisation models'.

The estimated intercepts of the semi-parametric models are given on the worksheet 'Intercept' (Figure 12). In this case nine columns are given for each pre-specified model providing a time-varying intercept for each sector.
Normalised values from a simple regression model, only using annual deposition and annual precipitation amount is also provided. This can be found on the worksheet 'Simple Regression Model'. No picture is given here.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6	Sector 7	Sector 8	Sector 9			Sector
2	1987	0.15101	0.06694	0.4318	0.93617	1.21769	1.76535	0.78937	0.1966	3.41469		1987	0.27
3	1988	0.16416	0.06976	0.41083	0.89728	1.24497	1.78171	0.77053	0.20597	3.30141		1988	0.26
4	1989	0.17652	0.07245	0.38916	0.85964	1.26698	1.79445	0.75138	0.21446	3.18761		1989	0.25
5	1990	0.18703	0.07484	0.36848	0.82412	1.28344	1.79938	0.73115	0.22132	3.07334		1990	0.24
6	1991	0.19505	0.07667	0.35018	0.79	1.29315	1.79812	0.7094	0.22603	2.96279		1991	0.22
7	1992	0.20016	0.0778	0.33491	0.75535	1.29419	1.79099	0.68798	0.22849	2.85649		1992	0.21
8	1993	0.20158	0.07839	0.32273	0.72023	1.28585	1.77421	0.66818	0.22827	2.75365		1993	0.20
9	1994	0.19838	0.07852	0.31324	0.68358	1.26897	1.74518	0.64905	0.22548	2.65283		1994	0.18
10	1995	0.19138	0.07826	0.30618	0.64522	1.24361	1.70302	0.62883	0.22081	2.55153		1995	0.17
11	1996	0.18085	0.07755	0.30088	0.60541	1.21022	1.64826	0.60782	0.21504	2.44888		1996	0.15
12	1997	0.16696	0.07625	0.29559	0.56421	1.16824	1.58352	0.58604	0.20837	2.34367		1997	0.13
13	1998	0.15027	0.07449	0.28969	0.52311	1.11981	1.51192	0.56475	0.20137	2.23265		1998	0.12
14	1999	0.13177	0.07246	0.2822	0.48192	1.06717	1.43426	0.54526	0.19448	2.11725		1999	0.1
15	2000	0.11227	0.07049	0.27376	0.4415	1.01151	1.35342	0.52762	0.18783	2.00046		2000	0.0
16	2001	0.09259	0.0688	0.26475	0.40194	0.95592	1.271	0.51112	0.18118	1.88436		2001	0.06
17	2002	0.07285	0.06721	0.25555	0.36266	0.90035	1.18807	0.495	0.17451	1.76825		2002	0.04

Figure 12: On the worksheet 'Intercept' the estimated time-varying intercept for each class is given.

Theoretical background

Estimation of monthly and annual deposition and annual deposition by sector

Monthly and annual deposition data are calculated by first replacing all missing precipitation data by zeros. Missing sector values are replaced by sector value 9, and missing concentrations are imputed by a simple regression model using precipitation amount as explanatory variable. Subsequently daily concentration is multiplied with daily precipitation values and then summed up to form monthly or annual data by sector.

Normalisation of deposition data

Deposition data is normalised by employing parametric and semiparametric regression models to remove or suppress the temporal variation that can be attributed to fluctuations in precipitation.

The normalisation can be conducted on two different kinds of input data, either monthly deposition or annual deposition by sector. We therefore use the term class, with means either month or sector. The semiparametric normalisation model has then the general form:

$$y_{ij} = \alpha_{ij} + \beta_{1,j} x_{1,ij} + ... + \beta_{p,j} x_{p,ij} + \varepsilon_{ij}, \quad i = 1, ... , n \quad j = 1, ... , m$$

where y_{ij} is the observed response for the jth class of the ith year, $x_{k,ij}$, k=1, … ,p represent precipitation values, number of precipitation days and/or number or precipitation periods, and ε_{ij} is a random error term with mean zero. The slope parameters ($\beta_{k,j}$, k=1, … ,p) are permitted to vary with the class (j) under consideration, and the intercept (α_{ij}) is permitted to vary with both class (j) and year (i). However, rapid changes in the intercept are controlled by so-called roughness penalty factors (λ_1 and λ_2), and the intercept and slope parameters are estimated by minimising the expressions:

for seasonal data:

$$S(\alpha, \beta) = \sum_{i,j} (y_{ij} - \alpha_{ij} - \beta_{1,j} x_{1,ij} - ... - \beta_{p,j} x_{p,ij})^2 + \lambda_1 \sum_{i,j} (\alpha_{ij} - \frac{\alpha_{i+1,j} + \alpha_{i-1,j}}{2})^2$$
$$+ \lambda_2 \sum_{i,j} (\alpha_{ij} - \frac{\alpha_{i,j-1} + \alpha_{i,j+1}}{2})^2,$$

where

$$\alpha_{i,0} = \alpha_{i-1,m}, \quad i = 2, \dots, n; \quad \alpha_{i,m+1} = \alpha_{i+1,1}, \quad i = 1, \dots, n-1$$

to ensure that smoothness between december one year and january the next year.

for data by sector:

$$S(\alpha,\beta)=\sum_{i,j}(y_{ij}-\alpha_{ij}-\beta_{1,j}x_{1,ij}-...-\beta_{p,j}x_{p,ij})^2+\lambda_1\sum_{j=1}^{r}\sum_{i=2}^{n-1}(\alpha_{ij}-\frac{\alpha_{i+1,j}+\alpha_{i-1,j}}{2})^2$$
$$+\lambda_2\left\{\sum_{i=1}^{n}\left((\alpha_{i1}-\frac{\alpha_{i,r}+\alpha_{i,2}}{2})^2+\sum_{j=2}^{r-1}(\alpha_{ij}-\frac{\alpha_{i,j-1}+\alpha_{i,j+1}}{2})^2+(\alpha_{ir}-\frac{\alpha_{i,r-1}+\alpha_{i,1}}{2})^2\right)\right\}$$
$$+\lambda_1\sum_{i=2}^{n-1}(\alpha_{i,r+1}-\frac{\alpha_{i+1,r+1}+\alpha_{i-1,r+1}}{2})^2$$

Detailed information about algorithms for parameter estimation of the seasonal model has been published by Stålnacke and Grimvall (2001). Details for the model using sectoral data can be found in Libiseller et al.(2005)

The parametric normalisation model is an ordinary multiple regression model of the general form

$$y_{ij}=\alpha+\beta_{1,j}\,x_{1,ij}+...+\beta_{p,j}\,x_{p,ij}+\varepsilon_{ij},\quad i=1,...,n\quad j=1,...,m$$

where y_{ij}, $x_{k,ij}$, and ε_{ij} have the same meaning as above. The slope parameters ($\beta_{k,j}$, k=1, … ,p) are permitted to vary with the class (j) under consideration, whereas the intercept is assumed to be constant.

Selection of roughness penalty factors and assessment of the predictive ability of the tested normalisation models

If the penalty factors λ_1 and λ_2 are determined by cross-validation, the entire data set is separated into an estimation set (or training set) and a test set. The model is first fitted to the estimation set and is subsequently used to predict the observations in the test set, that is, the values that have been left out of the estimation step. If the observation period covers m years, we define m estimation sets M_i, i=1, … ,m by leaving out one-year-long blocks of observations, and then we compute a so-called PRESS-value (i.e., a sum of squared prediction errors):

$$S(\lambda_1,\lambda_2)=\sum_i\sum_{(i,j)\notin M_i}(y_{ij}-\hat{\alpha}_{ij}-\hat{\beta}_{1,j}x_{1,ij}-...-\hat{\beta}_{p,j}x_{p,ij})^2.$$

Finally, the factors λ_1 and λ_2 are selected in such a way that $S(\lambda_1,\lambda_2)$ is minimised, and the corresponding Root Mean *PRESS* value

$$min\left\{\sqrt{\frac{1}{N}S(\lambda_1,\lambda_2)};\lambda_1>0,\lambda_2>0\right\}$$

is used as a measure of the predictive ability of the normalisation model under consideration.

Literature references

Libiseller C., Grimvall A. and Hallberg L. (2005) Meteorological normalisation of time series of wet deposition (manuscript).

Stålnacke, P., and Grimvall, A.: 2001, 'Semiparametric approaches to flow-normalisation and source apportionment of substance transport in rivers', *Environmetrics* **12**, 233-250.