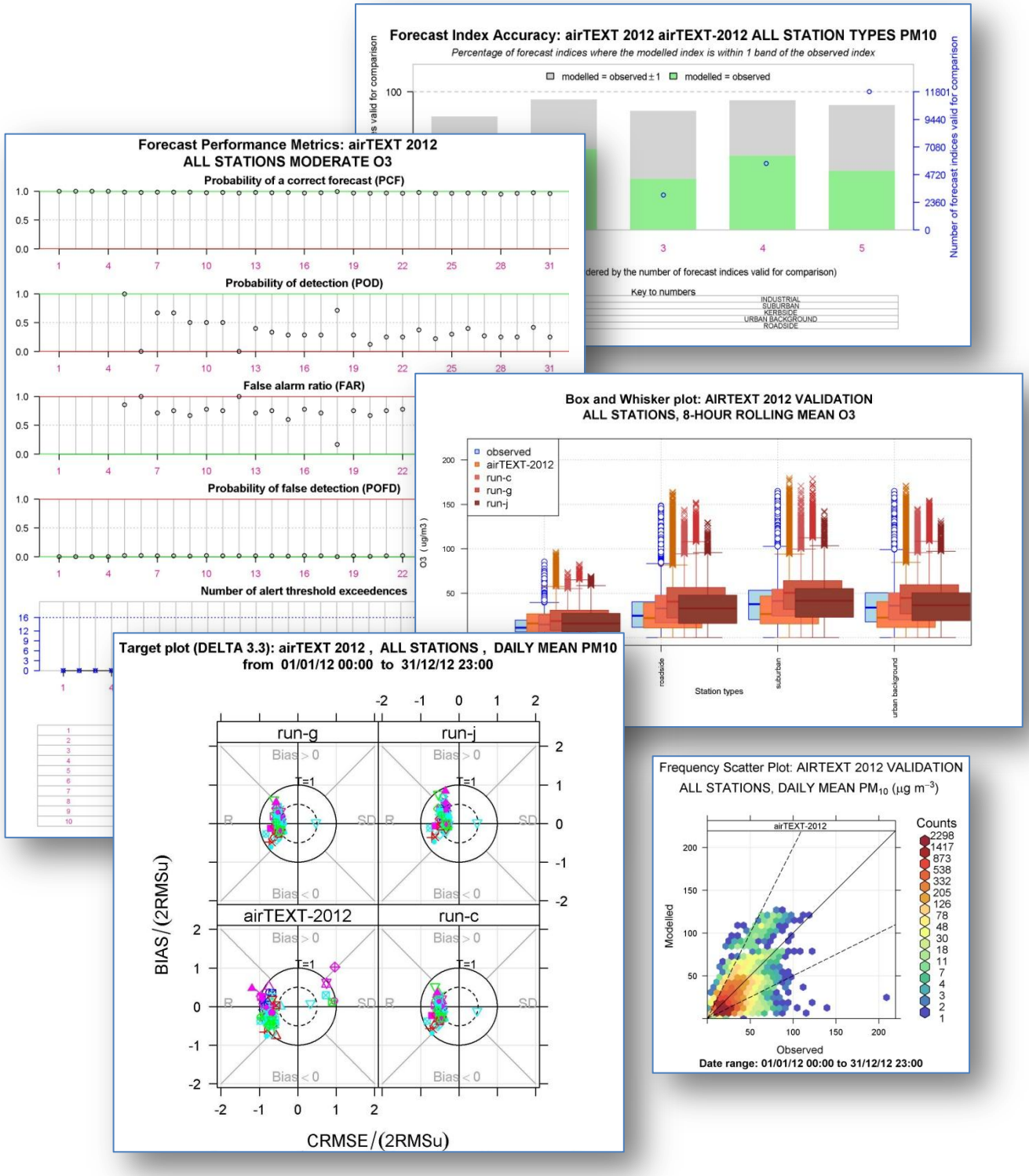
Model Evaluation Toolkit
User Guide
Version 4.0, January 2016



Support
The website for the Model Evaluation Toolkit is www.cerc.co.uk/ModelEvaluationToolkit. Please follow the links to the User Forum to report issues or suggest improvements; alternatively you can send an email to help@cerc.co.uk.

Contents

1 Introduction

Regional and municipal governments are increasingly interested in providing services to assess and forecast local- and city-level air quality. Air quality forecasts on these scales can be disseminated to health services and the public in terms of air quality alerts, to inform and warn 'at-risk' groups about impending pollution episodes and provide advice. Local air quality modelling is critical in assessment of air quality against the EC air quality directive as it can provide high resolution maps of concentration where the population is most dense and allows the investigation of proposed mitigation measures on short or long time scales. Understanding the benefits, limitations and performance of individual models, the input data required of them as well the extent of the options available to them is often lacking. Setting standard evaluation criteria and comparing model capabilities in a structured way is therefore a crucial task.

The Model Evaluation Toolkit has been developed under the local forecast model evaluation support work package of the EU's 7th Framework, PASODOBLE project. It draws on existing best practice such as the EU Joint Research Council's (JRC) FAIRMODE initiative on model evaluation [1] and the openair project tools [2,3].

The Toolkit is a simple-to-install, user-friendly environment that guides the user through the process of evaluating model predictions of local air quality and investigating the model performance. It runs on Windows operating systems.

The Toolkit can take modelled data from regional or local scale models as input. Observed data are *in situ* time series data. Missing data are handled if they are indicated by a standard value. As output, the toolkit creates plots of the model performance in predicting concentrations and predicting alerts with respect to defined thresholds, for single or multiple sites, single or multiple pollutants and single or multiple modelled datasets. Results can be classified by the type of monitoring site and the pollutant for each modelled dataset. The diagnosis of model performance for individual sites and individual pollutants produces time series plots, scatter plots and analyses with respect to month, day of the week and hour of the day. All the plotted data are also exported to data files to provide an audit trail and make the data available for further analysis and visualisation.

2 Getting started

The Toolkit can be used on Windows operating systems and does not require any software to be purchased. Before using the Toolkit you will need to install R and the Toolkit itself; this will just take a few minutes. Detailed installation instructions are given in Sections 2.1 and 2.2.

2.1 Install R

The Model Evaluation Toolkit version 4.0 is compatible with R version 3.2.3.

Follow these step-by-step instructions to download and install R from the internet:

1. Go to http://www.r-project.org/
2. Select CRAN from the links on the left-hand side of the page
3. Choose a CRAN mirror for your locality (in the UK, choose the mirror for the University of Bristol) and click on the link
4. Under 'Download and install R' click on the link for your operating system
5. Click on 'base'
6. Under 'Other builds' click on 'Previous releases'
7. Click on 'R 3.2.3 (December, 2015)'
8. Click on 'Download R for...' to download the install program
9. Run the install program, taking care to install R in a directory where you have write privileges.

NOTE: If you do not have direct access to the internet from your computer, for example you access the internet through a university network, then when you install R, instead of accepting all the defaults, at the Setup screen choose not to accept all the defaults and when offered, choose 'Internet2' as the internet option. This will force R to use the same proxy settings used by Internet Explorer. The defaults for all other options can be accepted.

2.2 Install the Model Evaluation Toolkit

Follow these step-by-step instructions to install the Model Evaluation Toolkit:

1. Log on as Local Administrator for the PC.
2. Unzip the downloaded *.zip* file to a local directory. In Explorer, browse to this directory and double-click on the file *setup.exe*. The **Welcome** window in Figure 2.1 will be launched.

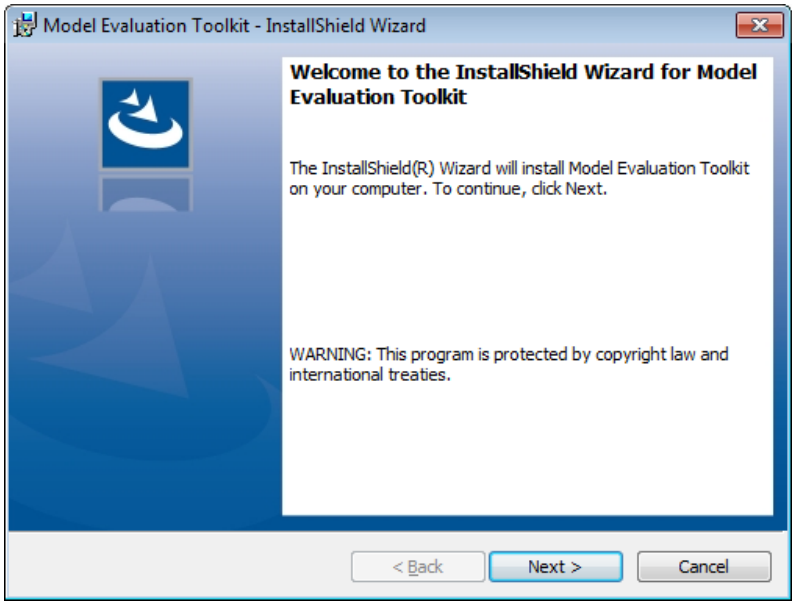


Figure 2.1 The Model Evaluation Toolkit *Welcome* screen.

3. Click 'Next' on the **Welcome** screen and then enter you Customer Information as shown in Figure 2.2. Click 'Next' to proceed to the **Destination Folder** screen, as shown in Figure 2.3.

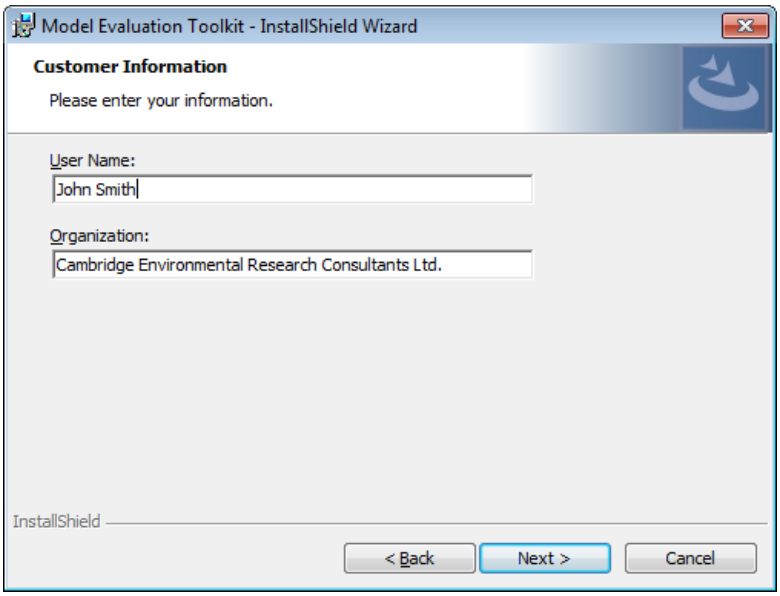


Figure 2.2 The Model Evaluation Toolkit Customer Information screen.

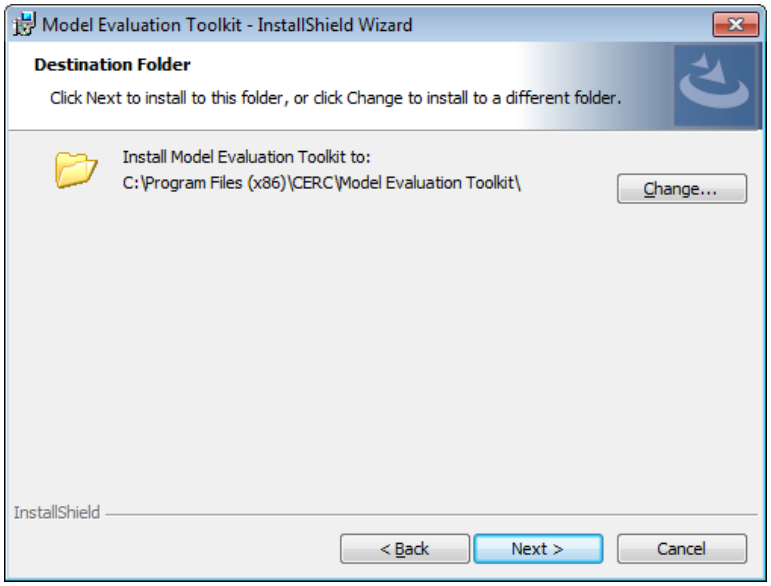


Figure 2.3 The Model Evaluation Toolkit Destination Folder screen.

4. The default installation directory is *C:\Program Files (x86)\CERC\Model Evaluation Toolkit*. If required, use the **Change...** button to select another destination directory. Click **OK** to return to the 'Destination Folder' screen.
5. Click **Next >** to choose who should be able to use the Model Evaluation Toolkit, as shown in Figure 2.4.

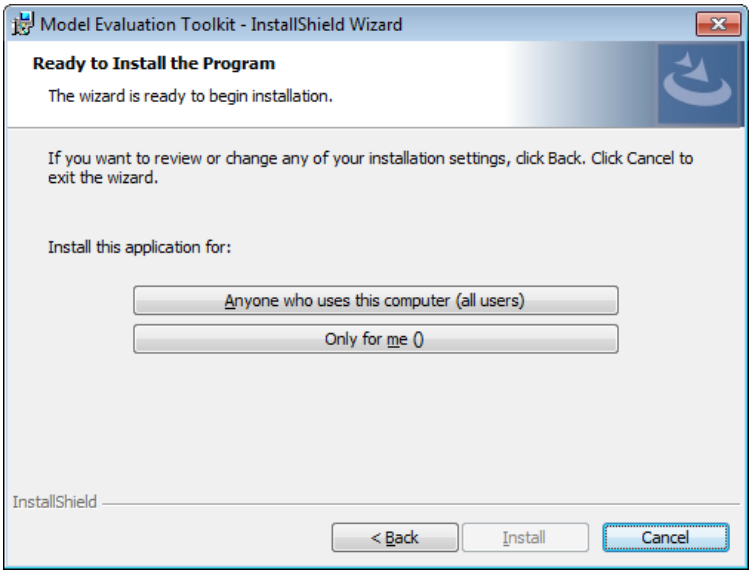


Figure 2.4 The Model Evaluation Toolkit *Ready to Install* screen.

6. Choose whether to install for all users of just the current user.
7. If you wish to amend any details, press the **< Back** and **Next >** buttons as appropriate. Once the **Install** button has been pressed, and the Model Evaluation Toolkit files have been successfully installed, the installer may prompt a restart, as shown in Figure 2.5. The restart can be delayed by clicking **No** but it is required for proper use of the Model Evaluation Toolkit. A final screen will now appear, as shown in Figure 2.6.

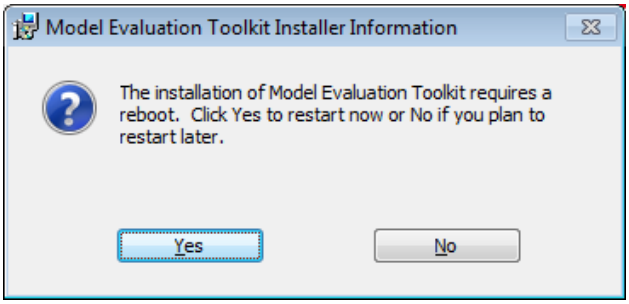


Figure 2.5 The installer may prompt you for a restart.

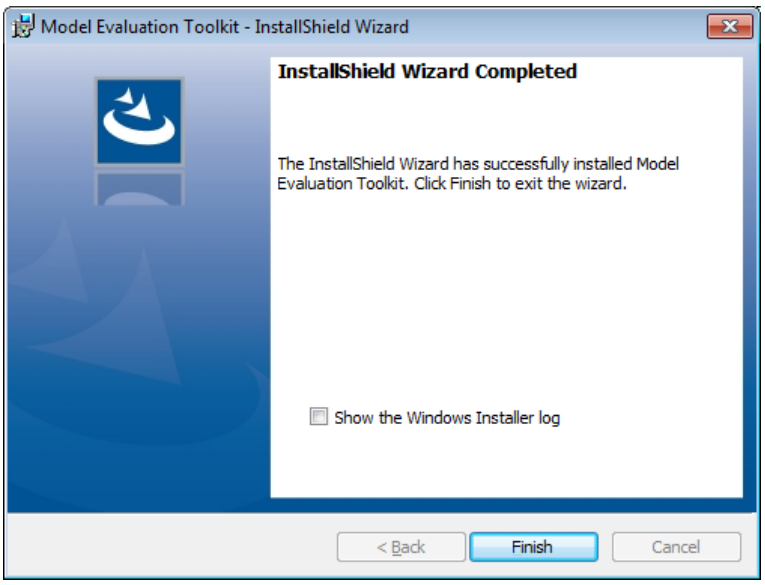


Figure 2.6 The Model Evaluation Toolkit *InstallShield Wizard Completed* screen.

8. Click **Finish** to complete the installation. The install wizard automatically creates three shortcuts in the **Programs** menu under **Model Evaluation Toolkit**; one each for the **Data Input Tool**, **Model Diagnostics Tool** and **Model Evaluation Tool**.

3 Using the toolkit

The Model Evaluation Toolkit consists of three tools:

1. Data Input

   This tool processes your modelled and observed concentration data, saving the processed data in an R workspace and (optionally) a CSV file.

2. Model Evaluation

   This tool takes in the workspace created by the Data Input tool and evaluates all or some of the data, producing graphs and (optionally) CSV files.

3. Model Diagnostics

   This tool takes in the workspace created by the Data Input tool and produces diagnostic graphs for one station and one pollutant at a time.

This section of the User Guide gives step-by-step instructions for using each tool in the Toolkit.

3.1 Data Input tool

The data input tool processes modelled and observed data, saving it in an R workspace file, which later can be imported into the Model Evaluation and Model Diagnostics tools.

The tool supports both gridded and point modelled data. Gridded data are interpolated to the monitoring station locations.

In situ observed data can either be automatically downloaded from the internet and imported (UK only) or input using simple-format CSV files.

The Toolkit supports the evaluation of multiple modelled datasets, either from multiple models or from multiple runs of the same model, over the same time period. Initially, a new R workspace is created with one modelled dataset and an associated observed dataset. Further modelled datasets can then be added to the same R workspace. There is no limit to the number of models that the Data Input tool will load into the same R workspace, but memory issues may be encountered if the number of models is large.

Tip: In the Model Evaluation Toolkit installation directory you will find a 'DataSamples' sub-directory. Here you will find sample files in the formats recognised by the Toolkit, and a ReadMe.txt file describing each file.

Step 1: Open the Data Input Tool

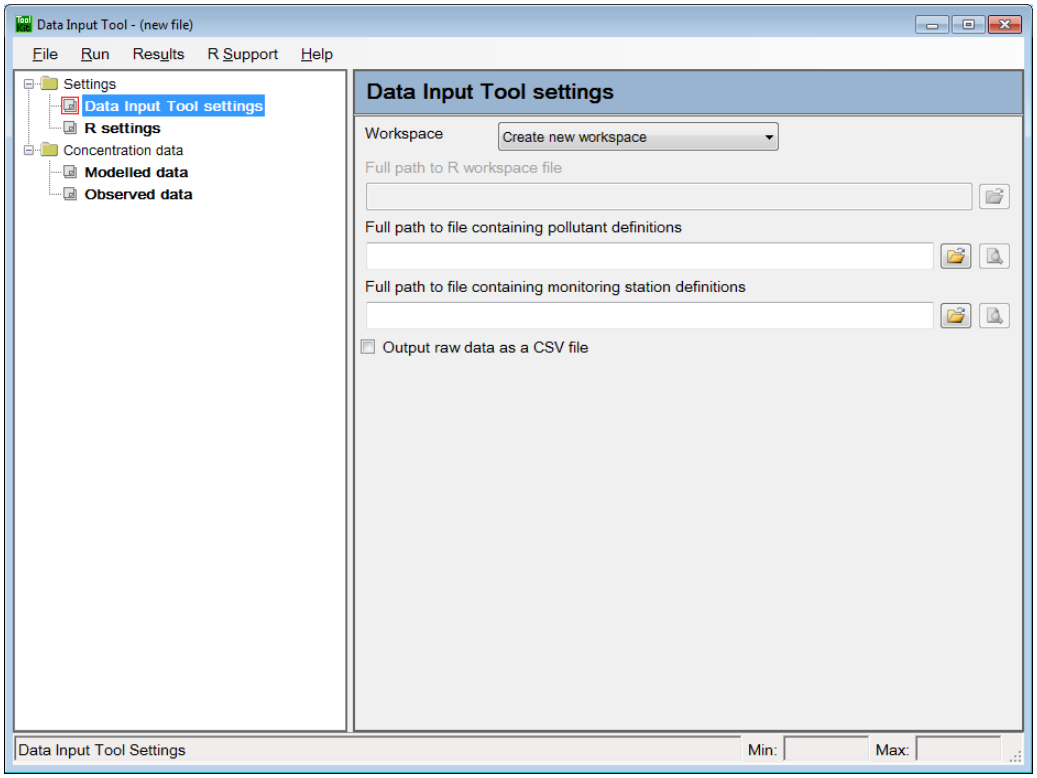


Figure 3.1 The Data Input Tool "Settings" screen

Step 2: Select whether you wish to create a new workspace or add modelled data

You can choose to either create a new R workspace or to add a modelled dataset to an existing R workspace created with the Data Input Tool.

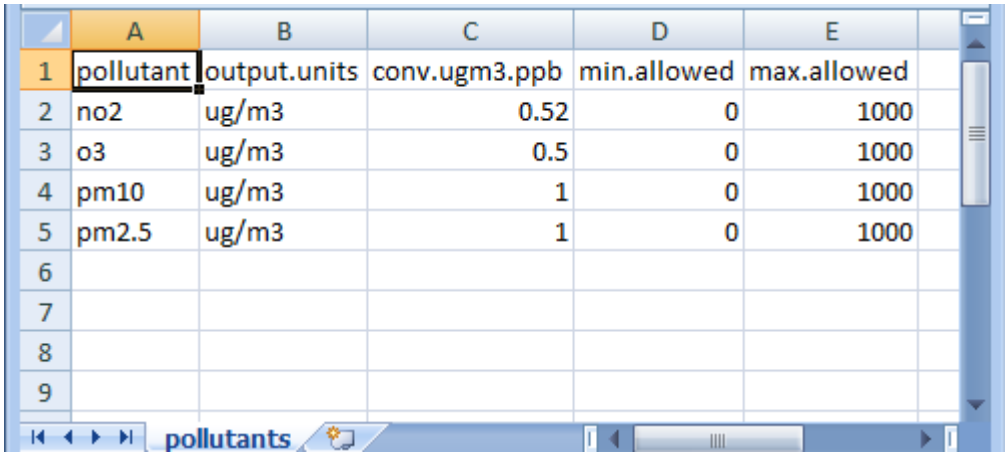
Note that, when creating a new workspace, the data range of the modelled dataset should cover the start and end time you wish to evaluate.

When adding a modelled dataset to an existing R workspace, the R workspace must have been created by version 4.0 of the Data Input tool.

When adding multiple modelled datasets to a workspace, the modelled data averaging time and statistic must be the same for all modelled datasets. Refer to Table 3.1 for details.

Step 3: If creating a new workspace, browse to select a pollutant definitions file

Your pollutant definitions file should be a CSV file with a comma separator, containing a list of all the pollutants for which you wish to process data. For each pollutant a set of parameters must be set. **It is very important these parameters are set correctly for your data.** Refer to Table 3.1 for details of the information required.



	A	B	C	D	E
1	pollutant	output.units	conv.ugm3.ppb	min.allowed	max.allowed
2	no2	ug/m3	0.52	0	1000
3	o3	ug/m3	0.5	0	1000
4	pm10	ug/m3	1	0	1000
5	pm2.5	ug/m3	1	0	1000
6					
7					
8					
9					

Figure 3.2 Example of a pollutant definitions file

Column header	Description	Allowed values
pollutant	Name of the pollutant to be used in all output	*n/a*
output.units	Concentration units to be used in all output	'mol/mol', 'ppb', 'ppm', 'ng/m3', 'ug/m3', 'mg/m3', 'g/m3' or 'kg/m3'
conv.ugm3.ppb	Conversion factor from µg/m^3 to ppb (used for unit conversions)	Any numeric value
min.allowed	Minimum allowed concentration value	Any numeric value
max.allowed	Maximum allowed concentration value	Any numeric value

Table 3.1 Details of the pollutant definitions CSV file columns.

Step 4: If creating a new workspace, browse to select a stations definitions file

Your stations definitions file should be a CSV file with a comma separator, containing a list of all the stations for which you wish to process data. Each station should have a station type. Also, if you have selected one of the gridded netCDF modelled data formats you will also need to include the station longitude and latitude.

	A	B	C	D	E
1	name	type	latitude	longitude	
2	BG1	Suburban	51.56375	0.177928	
3	BG2	Suburban	51.52939	0.132889	
4	BG3	Roadside	51.54043	0.074451	
5	BL0	Urban Bac	51.52206	-0.12578	
6	BN1	Roadside	51.61467	-0.17658	
7	BN2	Urban Bac	51.59189	-0.20596	
8	BQ5	Industrial	51.47769	0.190833	
9	BQ6	Industrial	51.47849	0.185298	
10	BQ7	Urban Bac	51.49476	0.134384	
11	BQ8	Urban Bac	51.49476	0.134384	
12	BT1	Suburban	51.58962	-0.27549	
13	BT4	Roadside	51.55247	-0.25806	
14	BT5	Industrial	51.55265	-0.24875	
15	BT6	Roadside	51.53779	-0.24776	

stations_KCL_latlong

Figure 3.3 Example of a station data file

If you are using either the London KCL or UK AURN observed data options then the station names must match the station codes used by those networks.

Step 5: Select the option to output raw data to a CSV file if required

The processed data generated by the Data Input Tool will always be output as an R workspace for analysis in the Model Evaluation Tool and/or Model Diagnostics Tool.

If you select this option then this processed data will also be output as a CSV file, which can be useful for further analysis in other programs such as Excel.

Step 6: Select your modelled data.

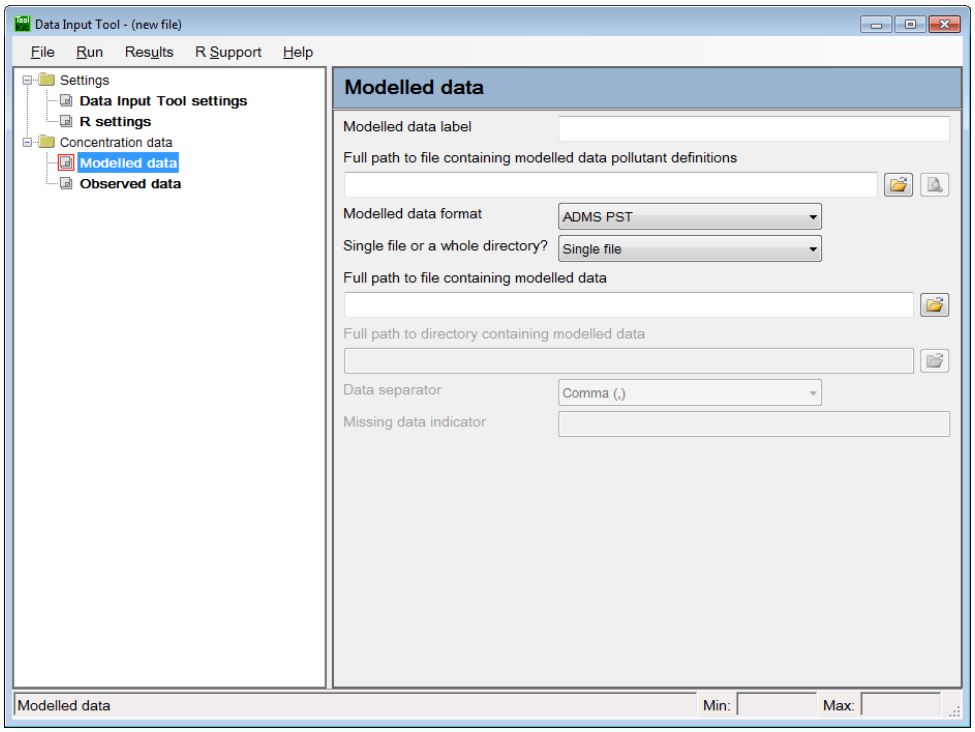


Figure 3.4 The Data Input Tool "Modelled data" screen

You need to label the modelled dataset being read in to the tool.

> *This label will be used in the subsequent plots to identify the modelled dataset. The label should be alpha-numeric and not include any special characters such as '&', '!' or commas.*

You can select either a single file or a whole directory of data files.

The Data Input tool supports the following formats for modelled data:

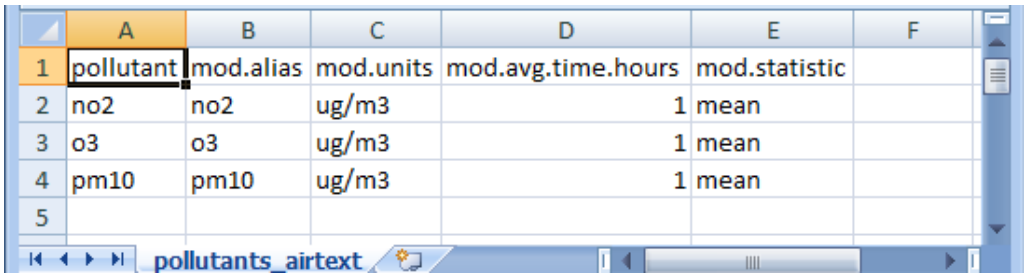
- ADMS PST: point receptor output format from the ADMS suite of atmospheric dispersion models
- AIRSHEDS netCDF: gridded modelled data from the PASODOBLE IC-AIRSHEDS netCDF work package

  > *This option supports output from the PASODOBLE Web Coverage Service (WCS) and output from individual IC-AIRSHEDS partners*

- CAMS Ensemble netCDF: gridded data from the CAMS regional ensemble air quality product
- CMAQ netCDF: gridded modelled data from the CMAQ model
- Generic CSV: standard text file
  - Refer to Section 5.1 for details of the required format
  - Select the separator used in your CSV files – either comma or semicolon
  - Enter the missing data indicator used in your CSV files (e.g. -999 or NA)

Step 7: **If creating a new workspace, select the modelled data pollutant definitions file**

Your modelled data pollutant definitions file should be a CSV file with a comma separator, containing a list of all the pollutants for which you wish to process data. For each pollutant a set of parameters must be set. **It is very important these parameters are set correctly for your data.** Refer to Figure 3.5 and Table 3.2 for details of the information required.



	A	B	C	D	E	F
1	pollutant	mod.alias	mod.units	mod.avg.time.hours	mod.statistic	
2	no2	no2	ug/m3	1	mean	
3	o3	o3	ug/m3	1	mean	
4	pm10	pm10	ug/m3	1	mean	
5						

pollutants_airtext

Figure 3.5 Example of a modelled data pollutant definitions file

pollutant	Name of the pollutant to be used in all output	*n/a*
mod.alias	Pollutant name as it appears in the modelled data	*n/a*
mod.units	Units that apply to the modelled data	'mol/mol', 'ppb', 'ppm', 'ng/m3', 'ug/m3', 'mg/m3', 'g/m3' or 'kg/m3'
mod.avg.time.hours	Averaging time in hours of the modelled data (minimum 1 hour)	An integer value, minimum 1
mod.statistic	Statistic that applies to the modelled data	'max', 'mean' or 'rolling mean'

Table 3.2 Details of the modelled data pollutant definitions file columns

Notes:

- The 'mod.alias' settings will be the same as 'pollutant' unless the pollutant names within the modelled data are different to 'pollutant', in which case use the aliases to make sure the correct fields are extracted from the data.
- When adding multiple modelled datasets to a workspace, the modelled data averaging time and statistic for a particular pollutant must be the same in all modelled datasets.

Step 8: **If creating a new workspace, leave the 'Observed data' option selected and continue; if adding data to an existing workspace, de-select 'Observed data' and skip to Step 12**

Step 9: **Select the observed data pollutant definitions file**

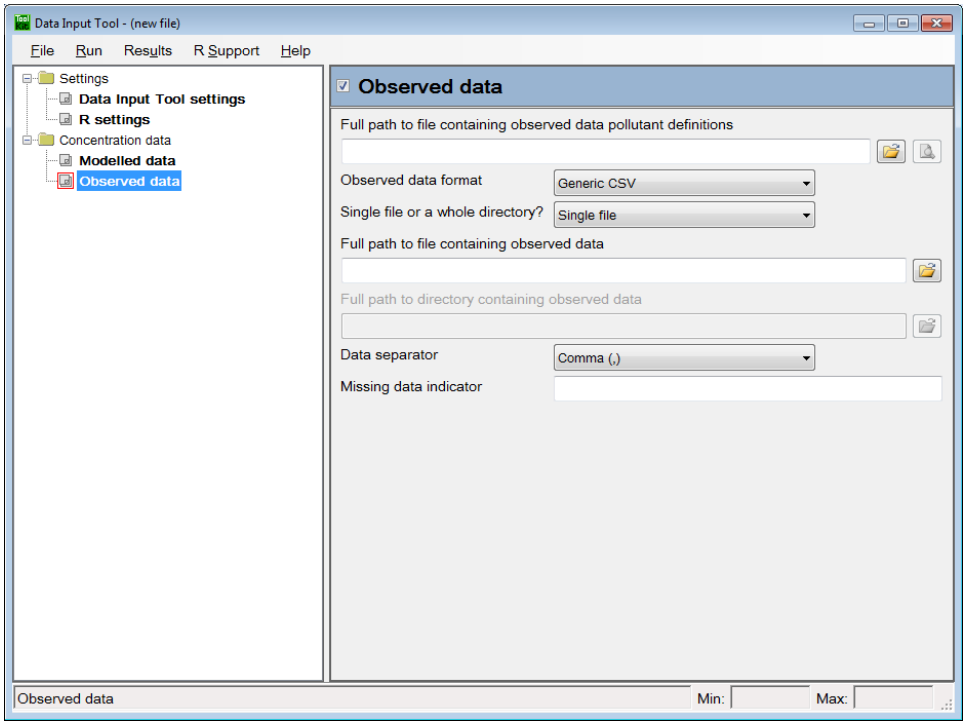
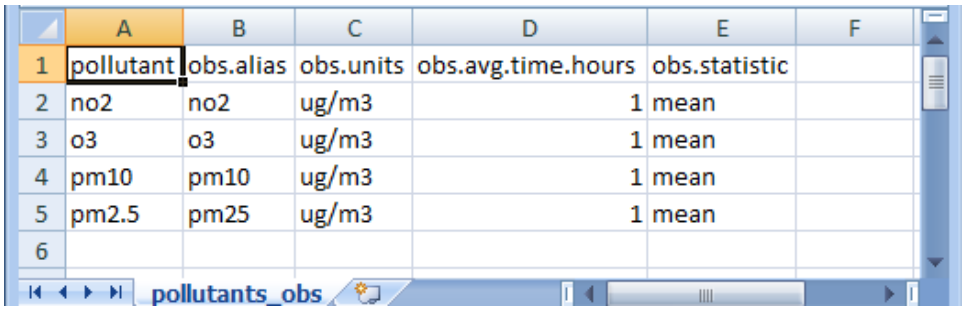


Figure 3.6 The Data Input Tool "Observed Data" screen

Your observed data pollutant definitions file should be a CSV file with a comma separator, containing a list of all the pollutants for which you wish to process data. For each pollutant a set of parameters must be set. **It is very important these parameters are set correctly for your data.** Refer to Figure 3.7 and Table 3.3 for details of the information required.



	A	B	C	D	E	F
1	pollutant	obs.alias	obs.units	obs.avg.time.hours	obs.statistic	
2	no2	no2	ug/m3	1	mean	
3	o3	o3	ug/m3	1	mean	
4	pm10	pm10	ug/m3	1	mean	
5	pm2.5	pm25	ug/m3	1	mean	
6						

pollutants_obs

Figure 3.7 Example of an observed data pollutant definitions file

pollutant	Name of the pollutant to be used in all output	*n/a*
obs.alias	Pollutant name as it appears in the observed data	*n/a*
obs.units	Units that apply to the observed data	'mol/mol', 'ppb', 'ppm', 'ng/m3', 'ug/m3', 'mg/m3', 'g/m3' or 'kg/m3'
obs.avg.time.hours	Averaging time in hours of the observed data (minimum 1 hour)	An integer value, minimum 1
obs.statistic	Statistic that applies to the observed data	'max', 'mean' or 'rolling mean'

Table 3.3 Details of the observed data pollutant definitions file columns.

Notes:

- The 'obs.alias' setting will be the same as 'pollutant' unless the pollutant name within the observed data is different to 'pollutant', in which case use the alias to make sure the correct fields are extracted from the data.
- The observed data averaging time and statistic must either be the same as for the modelled data, or 1 hour mean (then it will be recalculated to the same averaging time and statistic as the modelled data).

Step 10: Select your required observed data option.

The Data Input tool supports the following formats for observed data:

- Generic CSV: standard text file
  - Refer to Section 5.1 for details of the required format
  - Select the separator used in your CSV files – either comma or semicolon
  - Enter the missing data indicator used in your CSV files (e.g. -999 or NA)
  - Select either a single CSV file or a whole directory of CSV files

- London KCL

  This option requires internet access

  - King's College London (KCL) maintain a network of 123 monitoring sites in Greater London
  - Data for all required stations and pollutants will be downloaded and imported for a time period to match the modelled data

- UK Automatic and Rural Network (AURN)

  This option requires internet access

  - The UK AURN is the national network of automatic monitoring stations around the UK
  - Data for all required stations and pollutants will be downloaded and imported for a time period to match the modelled data

Step 11: If 'Generic CSV' is selected as the observed data type, do the following:

- Browse to select the observed data file or directory
- Choose the data separator used in the CSV file; the available options are comma (default) and semi-colon
- Specify the missing data indicator, for example -999 or NA

Step 12: Save the settings file and Run the Tool

To save the settings file go to "File" on the toolbar and select "Save" or "Save As...". To run the tool with the current settings select "Run" from the toolbar and then "Data Input Tool".

Note: If Run is selected without saving a .tki file, the Tool will prompt the user to save.

Step 13: Check the log file

The Data Input Tool generates a log file when it runs; this records all the processing messages that are shown in the run screen as well as any error and warning messages issued. It is good practice to check the log file for any problems that may have occurred.

To see the log file go to "Results" on the toolbar and select "View log". This option will open the log file in your preferred text editor (Notepad by default).

If "View log" is not available to select, this is because no log file exists for the currently loaded .tki file.

Step 14: View results folder

To open the results folder for the currently loaded .tki file, go to "Results" on the toolbar and select "Open results folder". This will open the results folder in Explorer.

If "Open results folder" is not available to select, this is because the .tki file has not yet been run and therefore no results are available to view.

Refer to Section 4.1 for details of the output from the Data Input Tool.

3.2 Model Evaluation tool

This section gives a step-by-step guide to using the Model Evaluation tool. For details of each graph type, please refer to Section 4.2.

Tip: In the Model Evaluation Toolkit installation directory you will find a 'DataSamples' sub-directory. Here you will find sample files in the formats recognised by the Toolkit, and a ReadMe.txt file describing each file.

Step 1: Open the Model Evaluation tool

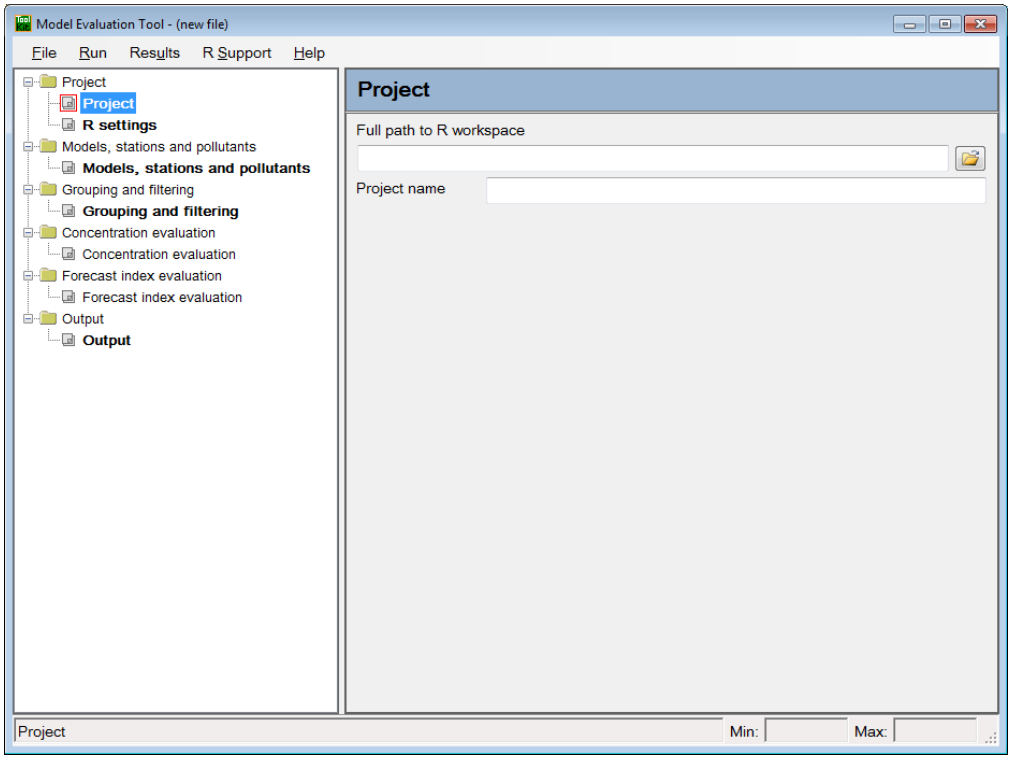


Figure 3.8 The Model Evaluation Tool "Project" screen

Step 2: Browse to select the workspace previously created by the Data Input tool

Step 3: Enter a project name. This will appear on the titles of graphs.

Step 4: Select which modelled datasets to evaluate

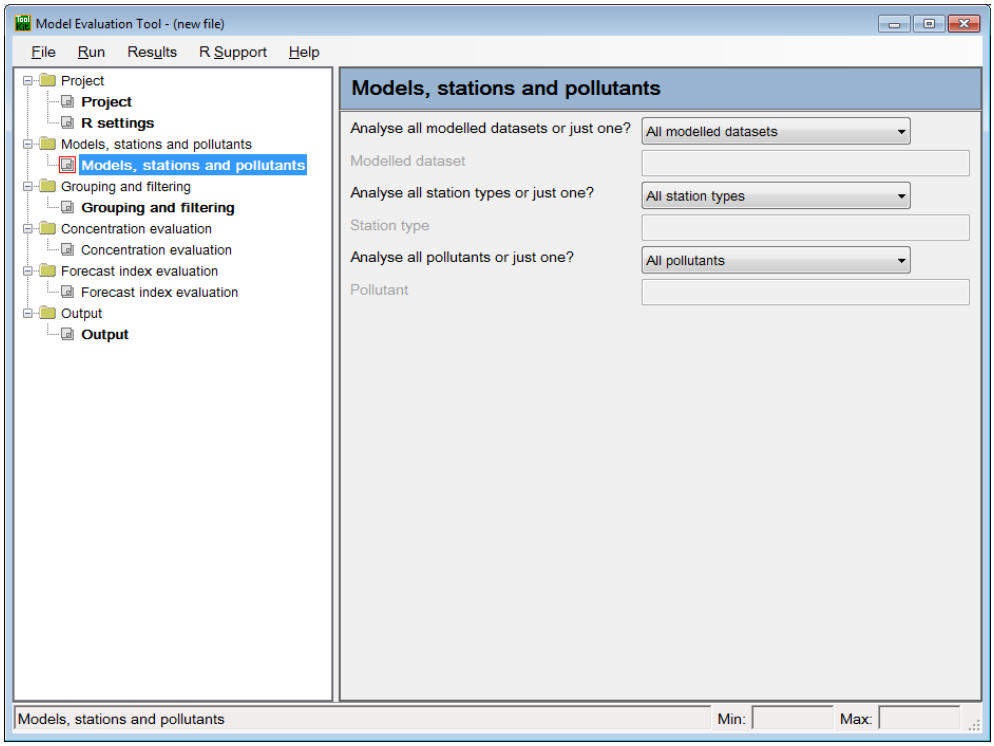


Figure 3.9 The Model Evaluation Tool "Models, stations and pollutants" screen

You can either evaluate all modelled datasets or just one modelled dataset. If you only want to evaluate one modelled dataset, enter its label (as specified in the Data Input tool) in the box provided.

Tip: If you don't know the labels of the modelled datasets that are available, type any text into the box. When you run the tool it will declare that no data are available for that modelled dataset and give you a list of the available labels.

Step 5: Select which station types to evaluate

You can either evaluate all station types or just stations of one type. If you only want to evaluate stations of one type, enter the name of this station type in the box provided.

Tip: If you don't know which station types are available, type any text into the box. When you run the tool it will declare that no data are available for that station type and give you a list of the available station types.

Step 6: Select which pollutants to evaluate

You can either evaluate all pollutants or just one pollutant. If you only want to evaluate one pollutant, enter the name of this pollutant in the box provided.

Tip: If you don't know which pollutants are available, type any text into the box. When you run the tool it will declare that no data are available for that pollutant and give you a list of the available pollutants.

Step 7: Choose how to group and filter the data

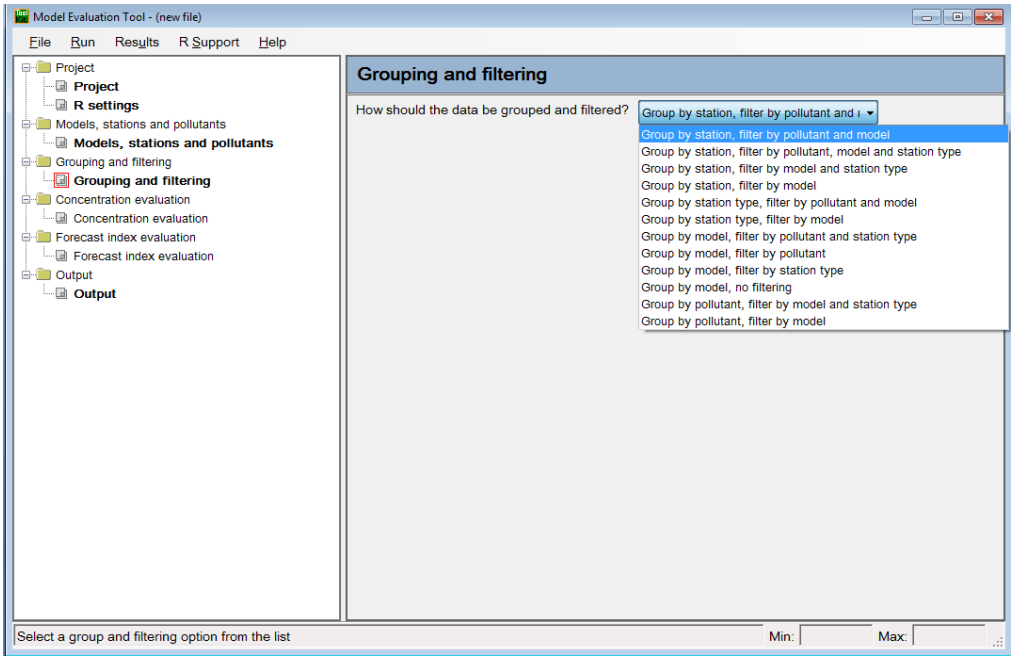


Figure 3.10 The Model Evaluation Tool "Grouping and filtering" screen

'Grouping' describes what each data point on a plot represents. The Model Evaluation tool supports grouping the data by:

- Station[1]
- Station type
- Pollutant
- Model (or modelled dataset)

'Filtering' describes how the data is to be split across plots; for example if the option to filter by station type is selected then data for different station types will be plotted on separate plots.

The default option is to group by station and filter by pollutant and model. Data must always be either grouped by or filtered by model (or modelled dataset) because the datasets must always be evaluated separately against the observed data.

Note that there are limitations on the grouping and filtering variables available for each plot option, appropriate to what each plot presents. For example, the forecast index evaluation plots present normalised variables and so all options are available with the only constraint being that the data must be either grouped or filtered by model. Whereas for concentration evaluation plots, the data must be grouped or filtered by both pollutant and model. Table 3.4 presents the available paired options for grouping and filtering for each plot. Refer to Sections 4.2.1 and 4.2.2 for details about these graphs.

[1] It is noted that, for the frequency scatter plot, conventional scatter plot and QQ plot options, the option to group by station should be chosen. In these specific cases, the data are plotted for all stations together on the same graph. If a graph for an individual station is required, refer to Section 3.3 for the Model Diagnostics tool.

Plot	Allowed Group	Allowed Filter
Target plot	Station	Model and Pollutant
	Station	Model and Pollutant and Station type
	Station type	Model and Pollutant
	Pollutant	Model
	Pollutant	Model and Station type
	Model	Pollutant
	Model	Pollutant and Station type
Box and whisker plot	Station	Model and Pollutant
	Station	Model and Pollutant and Station type
	Station type	Model and Pollutant
	Model	Pollutant and Station type
Scatter plot	Station	Model and Pollutant
	Station	Model and Pollutant and Station type
Quantile-quantile plot	Station	Model and Pollutant
	Station	Model and Pollutant and Station type
Forecast index plots	Station	Model
	Station	Model and Pollutant
	Station	Model and Station type
	Station	Model and Pollutant and Station type
	Station type	Model
	Station type	Model and Pollutant
	Pollutant	Model
	Pollutant	Model and Station type
	Model	-
	Model	Pollutant
	Model	Station type
	Model	Pollutant and Station type
Time plot	Station	Model and Pollutant

Table 3.4 Grouping and filtering options for the Model Evaluation tool

Step 8: Concentration Evaluation

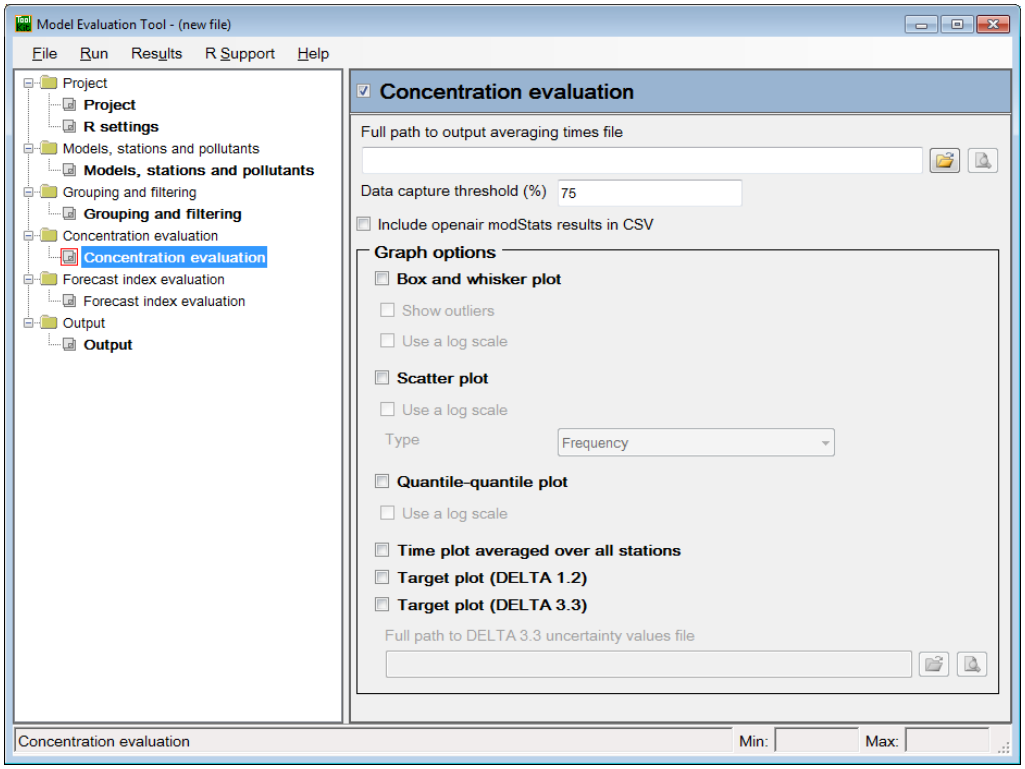


Figure 3.11 The Model Evaluation Tool "Concentration evaluation" screen

This part of the tool compares the modelled concentrations to observed concentrations.

- Check the box to perform the concentration evaluation.
- Select your output averaging times file.

  Your output averaging times file should be a comma-separated CSV file containing one row per pollutant. For each pollutant, the required averaging time and statistic must be set. Table 3.5 below describes these parameters.

	A	B	C	D	E
1	pollutant	output.avg.time.hours	output.statistic	output.daily.max	
2	no2	24	max	no	
3	pm10	24	mean	no	
4	o3	8	rolling mean	yes	
5	pm2.5	24	mean	no	

output_averaging_times

Figure 3.12 Example of an output averaging times file

Column header	Description	Allowed values
pollutant	Name of the pollutant as defined in the pollutant data file used in the Data Input tool	*n/a*
output.avg.time.hours	Averaging time in hours to be applied to the concentration evaluation output (minimum 1 hour)	An integer value, minimum 1
output.statistic	Statistic that applies to the concentration evaluation output	'max', 'mean' or 'rolling mean'
output.daily.max	Should output be based on the daily maximum of the calculated values?	'yes' or 'no'

Table 3.5 Details of the output averaging times CSV file columns. n/a indicates that values are unrestricted.

- Enter a data capture threshold to apply to the output averaging process. For example, for an 8-hour rolling mean, a 75% data capture threshold means that at least 6 hours of data must be valid in each period for the averaged data for this period to be considered valid and used in the evaluation.

- Six graph options are available: a box and whisker plot, a scatter plot as either a frequency (binned) or conventional scatter plot, a quantile-quantile plot, a time series plot averaged over all stations, a target plot from version 1.2 of the FAIRMODE DELTA Tool and a target plot from version 3.3 of the FAIRMODE DELTA Tool. Check the boxes of the graphs you require. Refer to Section 4.2.1 for details about these graphs.

- If you have selected the box plot option, check the 'Show outliers' option if you want outliers to be plotted on the graph

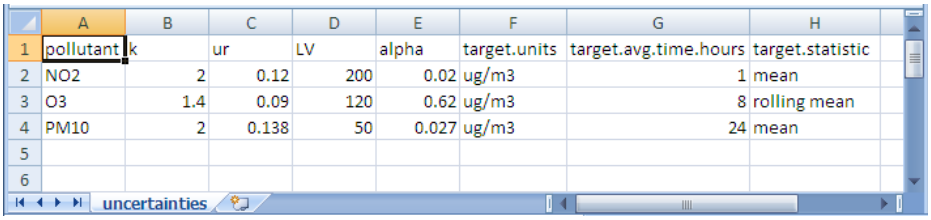
- If you have selected the scatter plot option, select which type of plot you require; a frequency scatter plot or conventional scatter plot.

  Tip: The frequency scatter plot is better suited to larger sets of data, the conventional scatter plot to smaller sets.

- If you have selected any of the box plot, scatter plot or quantile-quantile plot options, check the appropriate 'Use log scale' option if you wish the numerical axes of the plot to be scaled to $\log_{10}$.

- If you have selected the target plot from DELTA v3.3, select your uncertainties file.

  Your uncertainties file should be a comma-separated CSV file containing one row per pollutant. For each pollutant, the coefficients of the measurement uncertainty given by the FAIRMODE DELTA v3.3 methodology [1] must be specified. Table 3.6 describes these parameters.



	A	B	C	D	E	F	G	H
1	pollutant	k	ur	LV	alpha	target.units	target.avg.time.hours	target.statistic
2	NO2	2	0.12	200	0.02	ug/m3	1	mean
3	O3	1.4	0.09	120	0.62	ug/m3	8	rolling mean
4	PM10	2	0.138	50	0.027	ug/m3	24	mean
5								
6								

uncertainties

Figure 3.13 Example of an uncertainties file

Column header	Description	Allowed values
pollutant	Name of the pollutant as defined in the pollutant data file used in the Data Input tool	*n/a*
k	Coverage factor, *k*.	Refer to the DELTA v3.3 methodology [1]
ur	Measurement uncertainty, u_r^{LV}	Refer to the DELTA v3.3 methodology [1]
LV	Limit value (or reference value)	Refer to the DELTA v3.3 methodology [1]
alpha	Proportion of the measurement uncertainty that is independent of the limit value.	Refer to the DELTA v3.3 methodology [1]
target.units	Units that apply to the measurement uncertainty	'mol/mol', 'ppb', 'ppm', 'ng/m3', 'ug/m3', 'mg/m3', 'g/m3' or 'kg/m3'
target.avg.time.hours	Averaging time in hours that applies to the measurement uncertainty, which should be equal to that applied to the concentration evaluation output (minimum 1 hour)	An integer value, minimum 1
target.statistic	Statistic that applies to the measurement uncertainty, which should be equal to the concentration evaluation output	'max', 'mean' or 'rolling mean'

Table 3.6 Details of the uncertainties CSV file columns. n/a indicates that values are unrestricted.

Step 9: Forecast Index Evaluation

This part of the tool converts both the observed and modelled concentrations to 'forecast indices' and performs an evaluation based on these forecast indices.

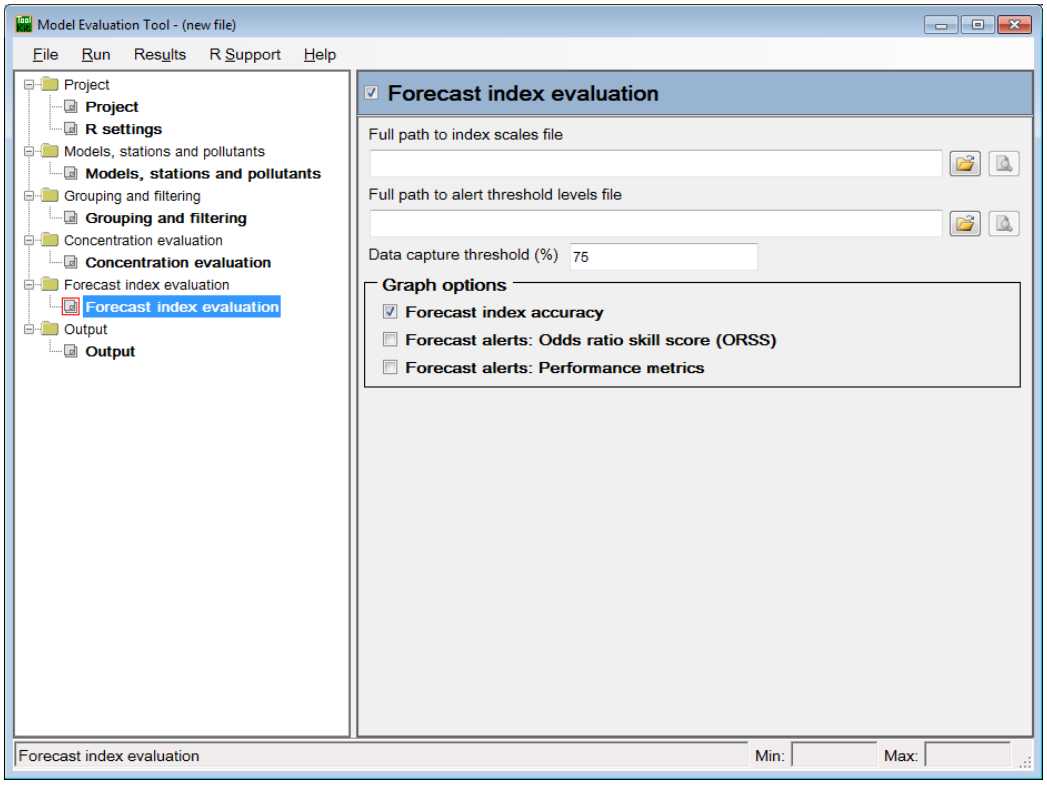


Figure 3.14 The Model Evaluation Tool "Forecast index evaluation" screen

- Check the box to perform the forecast index evaluation.
- Select your index scales file.

  Your index scales file should be a comma-separated CSV file containing one row per pollutant. For each pollutant, index-related parameters must be set. Table 3.7 describes these parameters. At least one index threshold concentration must be defined.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	pollutant	index.units	index.avg.time.hours	index.statistic	index.daily.max	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	
2	no2	ug/m3	1	mean	yes	0	68	135	201	268	335	401	468	535	601	
3	o3	ug/m3	8	rolling mean	yes	0	34	67	101	121	141	161	188	214	241	
4	pm10	ug/m3	24	mean	yes	0	17	34	51	59	67	76	84	92	101	
5	pm2.5	ug/m3	24	mean	yes	0	12	24	36	42	48	54	59	65	71	
6																

index_scales

Figure 3.15 Example of an index scales file

Column header	Description	Allowed values
pollutant	Name of the pollutant as defined in the pollutant data file used in the Data Input tool	*n/a*
index.units	Units in which the index threshold concentrations are given	'mol/mol', 'ppb', 'ppm', 'ng/m3', 'ug/m3', 'mg/m3', 'g/m3' or 'kg/m3'
index.avg.time.hours	Averaging time in hours that applies to the index threshold concentrations (minimum 1 hour)	An integer value, minimum 1
index.statistic	Statistic that applies to the index threshold concentrations	'max', 'mean' or 'rolling mean'
index.daily.max	Is the forecast given as the daily maximum of the calculated indices?	'yes' or 'no'
i1	Threshold concentration for index level 1	Any numeric value
i2	Threshold concentration for index level 2	Any numeric value
i{n}	Threshold concentration for index level n	Any numeric value

Table 3.7 Details of the index scales CSV file columns. n/a indicates that values are unrestricted.

- Select your alert thresholds file.

  Your alert thresholds file should be a comma-separated CSV file containing one row per alert. Each alert should have a name and an alert threshold defined in terms of the user-defined index scales. Table 3.8 describes these parameters.

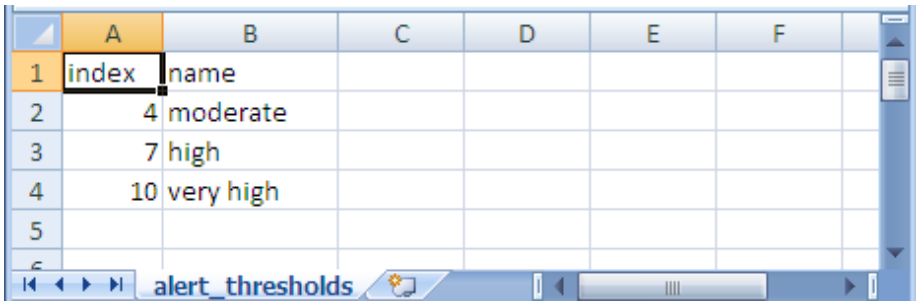


Figure 3.16 Example of an alert thresholds file

Column header	Description	Allowed values
index	Threshold index for this alert	Any integer value
name	The name to give to this alert in all output	*n/a*

Table 3.8 Details of the alert thresholds CSV file columns. n/a indicates that values are unrestricted.

- Enter a data capture threshold to apply to the forecast index averaging process. For example, for an 8-hour rolling mean, a 75% data capture threshold means that at least 6 hours of data must be valid in each period for the averaged data to be valid and used in the evaluation. The threshold will also be applied over each day if the daily maximum is set to "yes" in your index scales file.

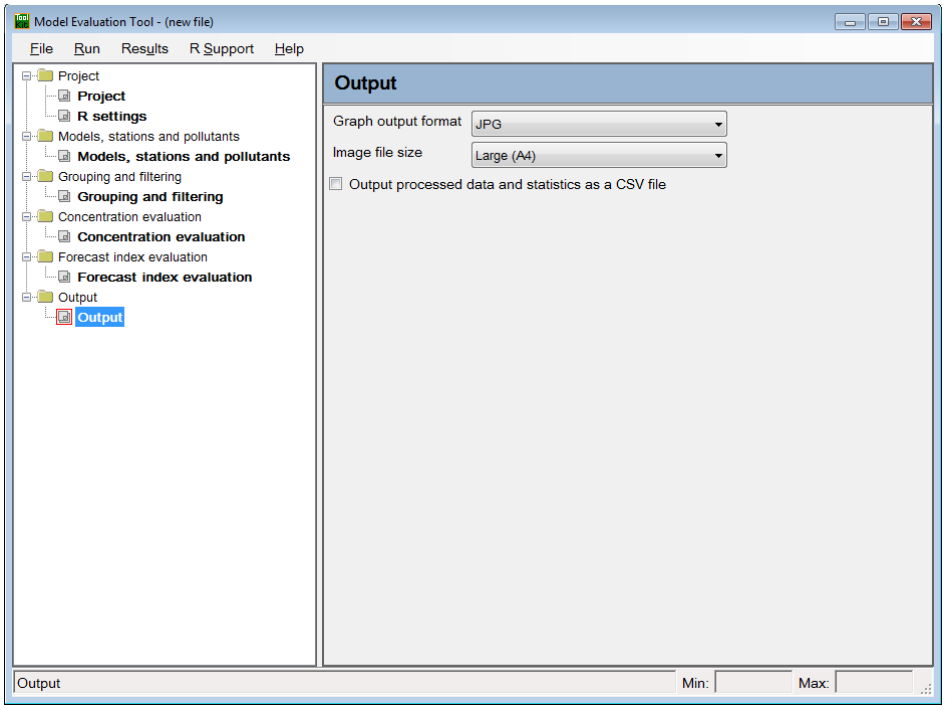
- Select the graphs you require

  There are three graph options (for details about these graphs, refer to Section 4.3):

  1. Forecast index accuracy
  2. Forecast alerts: Odds ratio skill score (ORSS)
  3. Forecast alerts: Performance metrics

  Check the boxes for the graph output you require.

Step 10: Select graph output format



There are three graph output options: JPG, PNG and PDF.

The first two options produce image files that can be imported into other documents. One image file is produced for each graph. The PDF option produces PDF files (size A4), with one PDF file per graph type.

Step 11: Select image file size

There are three options: Large (A4), Medium and Small.

These options only apply to the image output formats JPG or PNG. PDF output is always produced on A4. Reducing the overall image size will increase the proportional size of text. These options provide flexibility to produce graphs for reports or presentations.

Step 12: Choose whether to output processed data and statistics as CSV

This is a very useful option that produces CSV files containing all of the numerical data used to create the graphs, also some statistics not shown on the graphs. Refer to Section 4.2 for details of the contents of these files.

Step 13: Save the settings file and Run the Tool

To save the settings file go to "File" on the toolbar and select "Save" or "Save As...". To run the tool with the current settings select "Run" from the toolbar and then "Model Evaluation Tool".

Note: If Run is selected without saving a .tkm file, the Tool will prompt the user to save.

Step 14: Check the log file

The Model Evaluation Tool generates a log file when it runs; this records all the processing messages that are shown in the run screen as well as any error and warning messages issued. It is good practice to check the log file for any problems that may have occurred.

To see the log file go to "Results" on the toolbar and select "View log". This option will open the log file in your preferred text editor (Notepad by default).

If "View log" is not available to select, this is because no log file exists for the currently loaded .tkm file.

Step 15: View results folder

To open the results folder for the currently loaded .tkm file, go to "Results" on the toolbar and select "Open results folder". This will open the results folder in Explorer.

If "Open results folder" is not available to select, this is because the .tkm file has not yet been run and therefore no results are available to view.

Refer to Section 4.2 for details of the output from the Model Evaluation tool.

3.3 Model Diagnostics Tool

The aim of the model diagnostics tool is to enable further investigation of the performance of a model at one particular monitoring station for one particular pollutant.

Tip: In the Model Evaluation Toolkit installation directory you will find a 'DataSamples' sub-directory. Here you will find sample files in the formats recognised by the Toolkit, and a ReadMe.txt file describing each file.

Step 1: Open the Model Diagnostics Tool

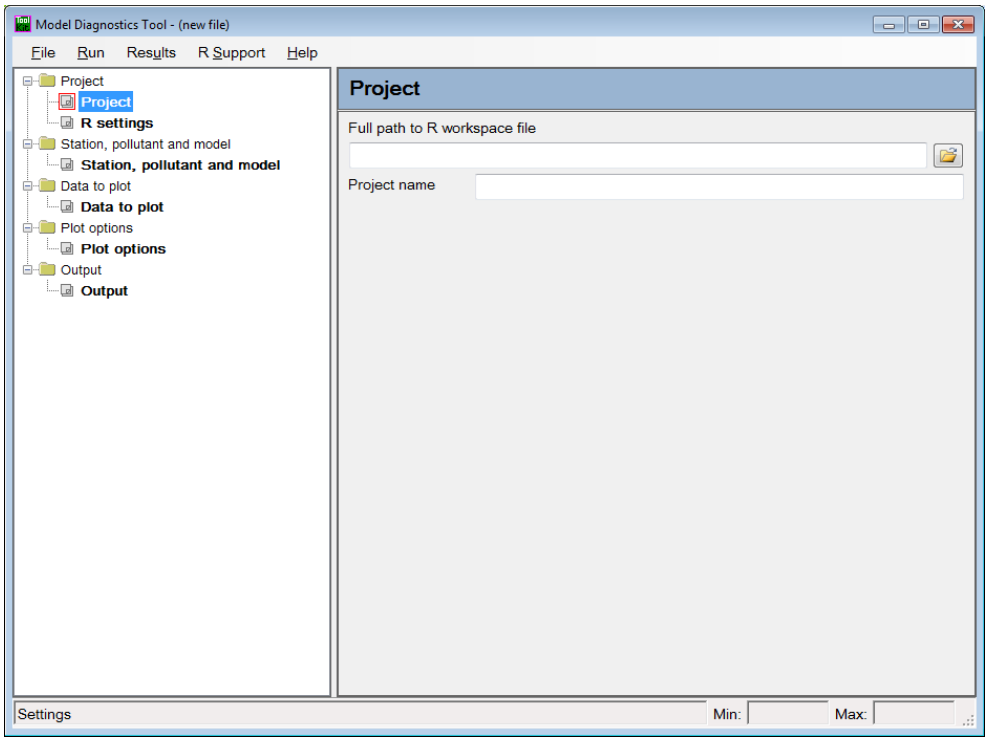


Figure 3.17 The Model Diagnostics Tool "Project" screen

Step 2: Browse to select the workspace previously created by the Data Input tool

Step 3: Enter a project name; this will appear on the titles of graphs

Step 4: Enter station, pollutant and modelled dataset label

Here you should type in the appropriate boxes the name of the station, pollutant and modelled dataset that you wish to produce plots for.

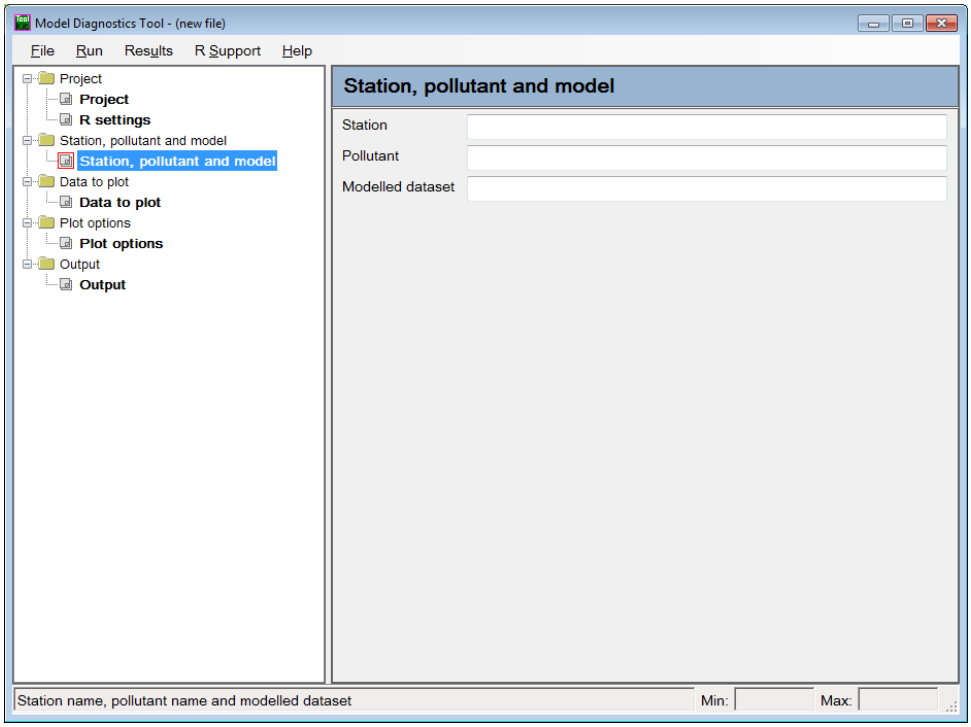


Figure 3.18 The Model Diagnostics Tool "Station, pollutant and model" screen

Step 5: Choose which type of data to plot

There are three options for the type of data to plot (see Figure 3.19):

1. Plot raw concentration data

   This plots the data as it was entered, in the averaging time of the modelled data.

2. Plot calculated statistics

   This plots the selected statistic, from a choice of the following: 8-hour rolling mean, 8-hour mean, daily mean, daily maximum and the daily maximum of the 8-hour rolling mean.

3. Plot calculated forecast indices

   This converts both the modelled and observed data to indices (as defined in the index scales file described in Section 3.2, step 9, and detailed in Table 3.7).

If 'Raw concentration data' or 'Calculated statistics' is selected, then one to ten reference values may also be plotted on the graph. To enter more than one reference value, separate the values with a comma, e.g. '50, 100'. If 'Forecast indices' is selected, then the alert thresholds defined in the selected alert thresholds file (described in Section 3.2, step 9) can be plotted on the graphs as reference lines.

If either calculated statistics or forecast indices are selected, you also need to enter a data capture threshold to apply to the averaging process. For example, for an 8-hour rolling mean, a 75% data capture threshold means that at least 6 hours of data must be valid in each period for the averaged data for that period to be valid and used in the graph.

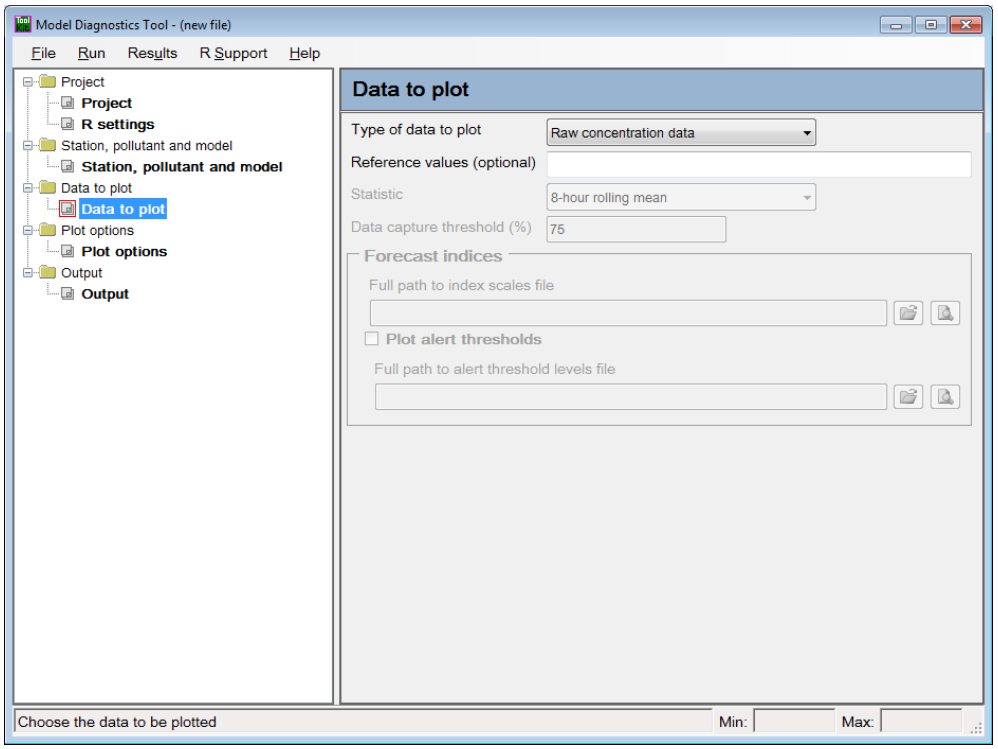


Figure 3.19 The Model Diagnostics Tool "Data to plot" screen

Step 6: Select which plot options you require

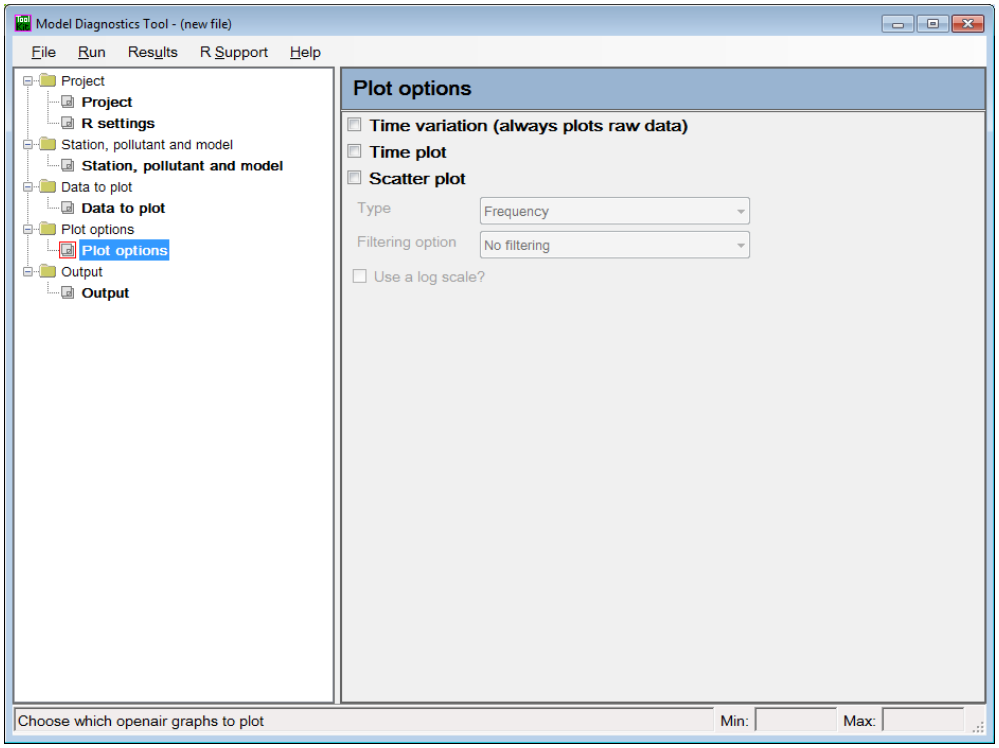


Figure 3.20 The Model Diagnostics Tool "Plot options" screen

There are three plot options (for full details on the graphs refer to Section 4.3):

1. Time variation

   This openair graph plots averages by weekday, month and hour. The settings in the 'Data to plot' section do not apply to this graph; it uses the raw data as it was entered into the Data Input tool.

2. Time plot

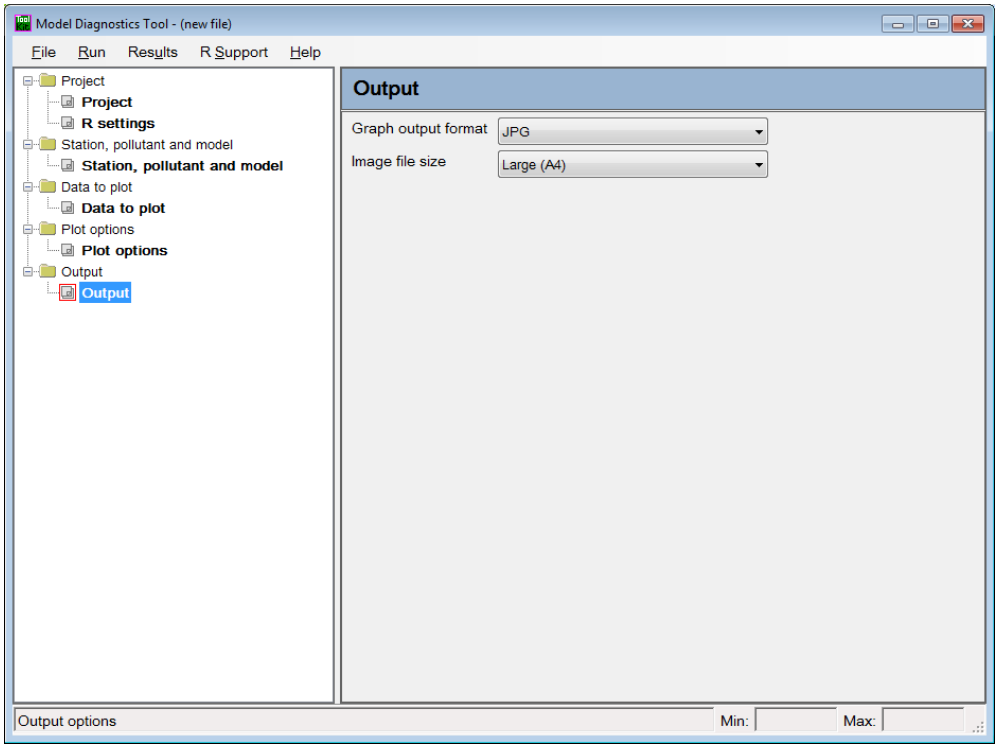
   This openair graph is a time series plot of both modelled and observed data.

3. Scatter plot

   This openair graph plots modelled against observed data. Select which type of plot you require; a frequency scatter plot shows the frequency of occurrence of each point whereas a conventional scatter plot shows one point per pairwise modelled-observed data point. The scatter plot can be scaled to $\log_{10}$ and filtered by weekday, month or season.

   Tip: The frequency scatter plot is better suited to larger sets of data, the conventional scatter plot to smaller sets.

Step 7: **Select graph output format and image file size**



There are three **graph output format** options: JPG, PNG and PDF. The first two options produce image files that can be imported into other documents. The PDF option produces PDF files.

There are three **image file size** options: Large (A4), Medium and Small. These options only apply to the image output formats JPG or PNG. PDF output is always produced on A4. Reducing the overall image size will increase the proportional size of text. These options provide flexibility to produce graphs for reports or presentations.

Step 8: Save the settings file and Run the Tool

To save the settings file go to “File” on the toolbar and select “Save” or “Save As...”. To run the tool with the current settings select “Run” from the toolbar and then “Model Diagnostics Tool”.

Note: If Run is selected without saving a .tkd file, the Tool will prompt the user to save.

Step 9: Check the log file

The Model Diagnostics Tool generates a log file when it runs; this records all the processing messages that are shown in the run screen as well as any error and warning messages issued. It is good practice to check the log file for any problems that may have occurred.

To see the log file go to “Results” on the toolbar and select “View log”. This option will open the log file in your preferred text editor (Notepad by default).

If “View log” is not available to select, this is because no log file exists for the currently loaded .tkd file.

Step 10: View results folder

To open the results folder for the currently loaded .tkd file, go to “Results” on the toolbar and select “Open results folder”. This will open the results folder in Explorer.

If “Open results folder” is not available to select, this is because the .tkd file has not yet been run and therefore no results are available to view.

Refer to Section 4.3 for details of the output from the Model Diagnostics tool.

4 Output

All three tools generate output files that are in the same directory as the input *.tki*, *.tkm* or *.tkd* file, with the name of the input file as a prefix, for easy identification. In addition to tool-specific output, all three tools produce a *.log* file that records messages written to the output screen, including any error or warning messages.

4.1 Data Input Tool Output

4.1.1 R Workspace

This is an R workspace containing all the data imported by the Data Input tool, processed and ready for import into the Model Evaluation and Model Diagnostic tools.

This workspace can also be loaded into R to explore the data further.

4.1.2 CSV file

This contains the time series of concentrations (in output units as defined in the pollutants information file) for every station and pollutant for which both monitored and modelled data are available, with the same averaging time and statistic as the modelled data.

4.2 Model Evaluation Tool Output

4.2.1 Concentration Evaluation Output

4.2.1.1 Target plot (DELTA 1.2)

The 'Target' plots produced by the concentration evaluation part of the Model Evaluation tool are similar to the plot produced by the FAIRMODE DELTA tool [1]. These plots can be output as defined by either version 1.2 or version 3.3 of the DELTA tool. This section describes output in line with version 1.2 of the DELTA tool.

For DELTA version 1.2, the metrics calculated by the tool for each monitoring station and pollutant and shown on the target plot are:

Centralised root-mean-square error (CRMSE):

$$\mathrm{CRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left((M_i - \bar{M}) - (O_i - \bar{O})\right)^2}$$

$\bar{M}$ is the mean modelled concentration, $\bar{O}$ is the mean observed concentration.

Mean bias (BIAS):

$$\mathrm{BIAS} = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)$$

Standard deviation of the observations (σ_{obs}):

$$\sigma_{obs} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - \bar{O})^2}$$

Standard deviation of the modelled data (σ_{mod}):

$$\sigma_{mod} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - \bar{M})^2}$$

Target (T):

$$T = \sqrt{\left(\frac{\text{BIAS}}{\sigma_{obs}}\right)^2 + \left(\frac{\text{CRMSE}}{\sigma_{obs}}\right)^2}$$

The target plot shows $\frac{\text{BIAS}}{\sigma_{obs}}$ against $\left[\frac{\text{CRMSE}}{\sigma_{obs}}\right]_{\sigma}$ where

$$\left[\frac{\text{CRMSE}}{\sigma_{obs}}\right]_{\sigma} = \begin{cases} -\frac{\text{CRMSE}}{\sigma_{obs}} & \text{if } \sigma_{mod} > \sigma_{obs} \\ \frac{\text{CRMSE}}{\sigma_{obs}} & \text{if } \sigma_{obs} > \sigma_{mod} \end{cases}$$

The radial distance to a data point on the target plot is equal to T for that station. The smaller the value of T, the better the concentration prediction for that station.

The black circle represents T=1, the dark green circle represents T=0.8, the light green circle represents T=0.65 and the dotted line represents T=0.3. In DELTA v1.2, 0.8 is the 'criteria' value, 0.65 is the 'goal' value and 0.3 is the uncertainty.

Figure 4.1 shows an example of target plot (DELTA 1.2) output from the concentration evaluation part of the model evaluation tool (all pollutants, all station types).

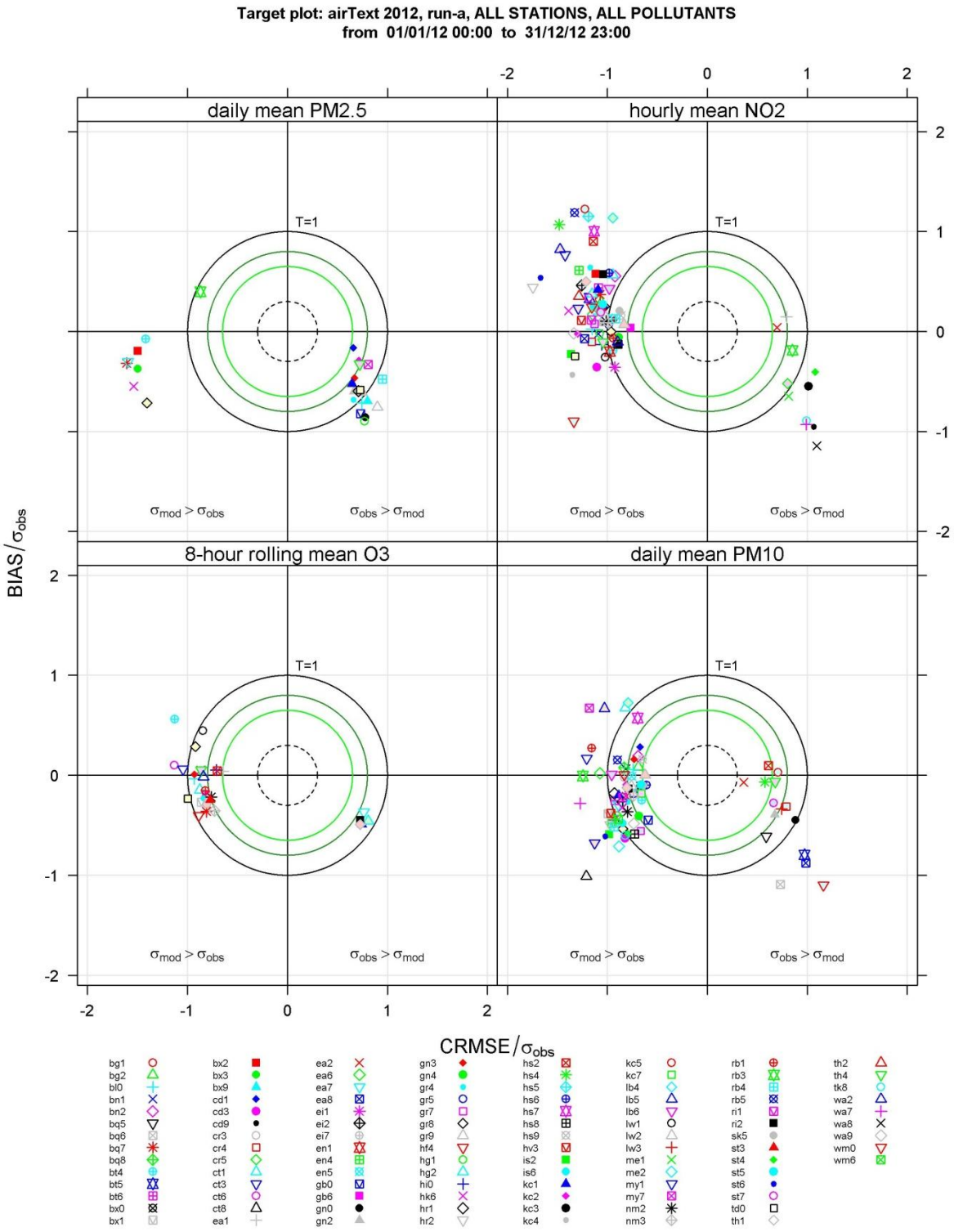


Figure 4.1 Example of a target plot (DELTA 1.2) from the concentration evaluation part of the Model Evaluation tool (grouped by station and filtered by model and pollutant for a single model 'run-a')

4.2.1.2 Target plot (DELTA 3.3)

The 'Target' plots produced by the concentration evaluation part of the Model Evaluation tool are similar to the plot produced by the FAIRMODE DELTA tool [1]. This section describes output in line with version 3.3 of the DELTA tool.

For DELTA version 3.3, the metrics calculated by the tool for each monitoring station and pollutant and shown on the target plot are:

Centralised root-mean-square error (CRMSE):

$$\mathrm{CRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left((M_i - \bar{M}) - (O_i - \bar{O})\right)^2}$$

$\bar{M}$ is the mean modelled concentration, $\bar{O}$ is the mean observed concentration.

Mean bias (BIAS):

$$\mathrm{BIAS} = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)$$

Root-mean-square of the observation measurement uncertainty (RMSu):

$$\mathrm{RMS_u} = k u_r^{LV}\sqrt{\left((1-\alpha)(\bar{O}^2 + \sigma_o^2)\right) + \alpha LV^2}$$

where σ_o^2 is the standard deviation of the observed data, *LV* is the limit value of interest and the other coefficients, k, u_r^{LV} and α, are derived for a specific pollutant from measurement data, as described in the FAIRMODE DELTA v3.3 methodology [1] and associated papers [4,5]. An example uncertainties file is provided in the data examples, using the values reported by DELTA v3.3:

	k	u_r^{LV}	α	*LV* (ug/m3)
NO2	2	0.120	0.020	200
O3	1.4	0.090	0.620	120
PM10	2	0.138	0.027	50

Target (*T*):

$$T = \sqrt{\left(\frac{\mathrm{BIAS}}{2\mathrm{RMS_u}}\right)^2 + \left(\frac{\mathrm{CRMSE}}{2\mathrm{RMS_u}}\right)^2}$$

The target plot shows $\frac{\mathrm{BIAS}}{2\mathrm{RMS_u}}$ against $\left[\frac{\mathrm{CRMSE}}{2\mathrm{RMS_u}}\right]_{\mathrm{NMSD}}$ where

$$\left[\frac{\mathrm{CRMSE}}{2\mathrm{RMS_u}}\right]_{\mathrm{NMSD}} = \begin{cases} -\dfrac{\mathrm{CRMSE}}{2\mathrm{RMS_u}} & \text{if } \dfrac{\mathrm{NMSD}}{\sqrt{2(1-R)}} < 1 \\ \dfrac{\mathrm{CRMSE}}{2\mathrm{RMS_u}} & \text{if } \dfrac{\mathrm{NMSD}}{\sqrt{2(1-R)}} > 1 \end{cases}$$

R is the Pearson's correlation coefficient and NMSD is the normalised mean standard deviation $(\sigma_m - \sigma_o)/\sigma_o$.

The radial distance to a data point on the target plot is equal to T for that station. The smaller the value of T, the better the concentration prediction for that station.

The black circle represents T=1, which represents the performance criteria and should be fulfilled by at least 90% of stations. The dashed black circle represents T=0.5. The grey lines separate quadrants representative of the dominant error in the modelled data, from positive or negative BIAS, or from standard deviation (SD) or correlation (R) errors in the CRMSE.

Figure 4.2 and Figure 4.3 show examples of output from the concentration evaluation part of the model evaluation tool with different grouping and filtering options.

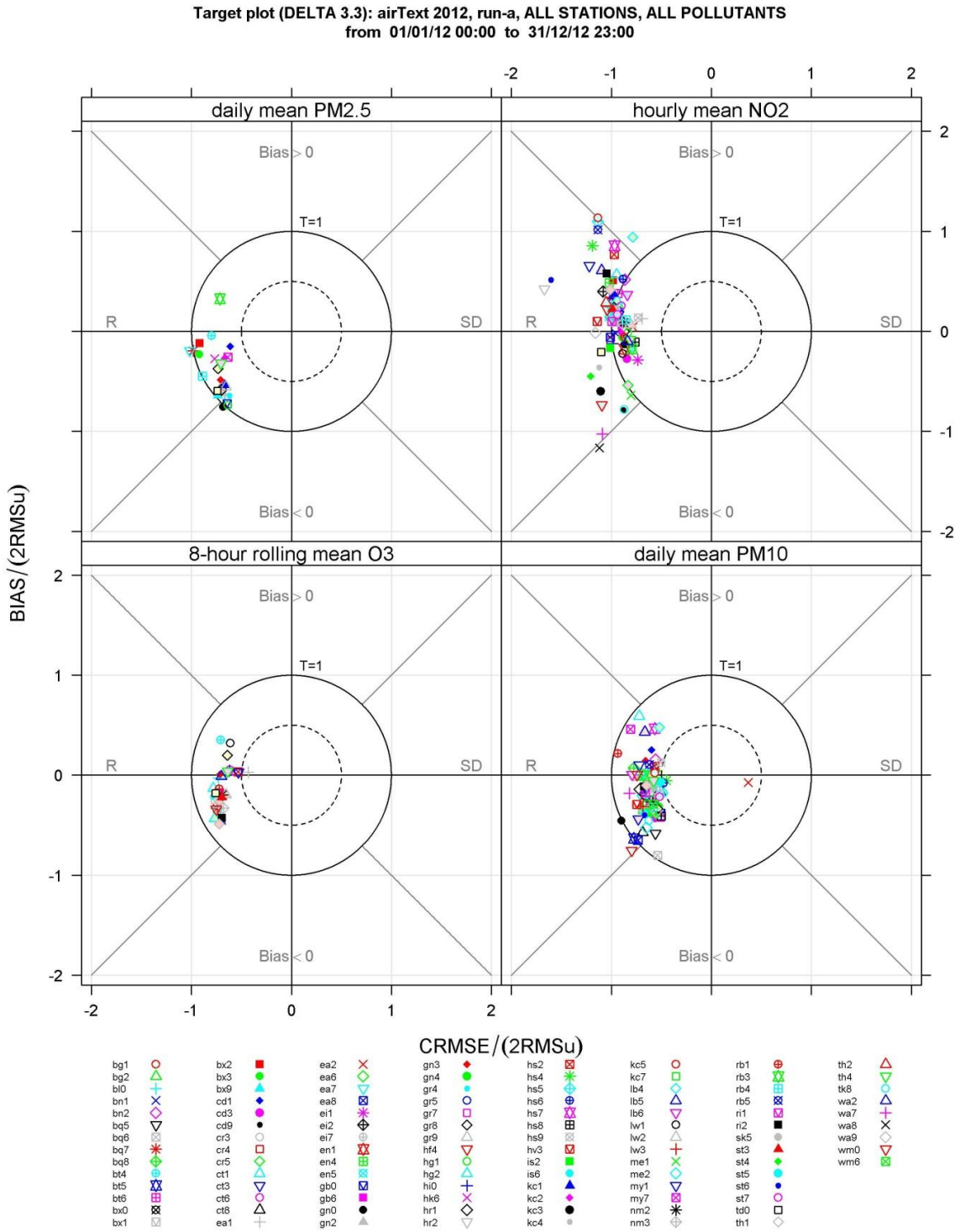


Figure 4.2 Example of a target plot (DELTA 3.3) from the concentration evaluation part of the Model Evaluation tool (grouped by station and filtered by model and pollutant for a single model 'run-a')

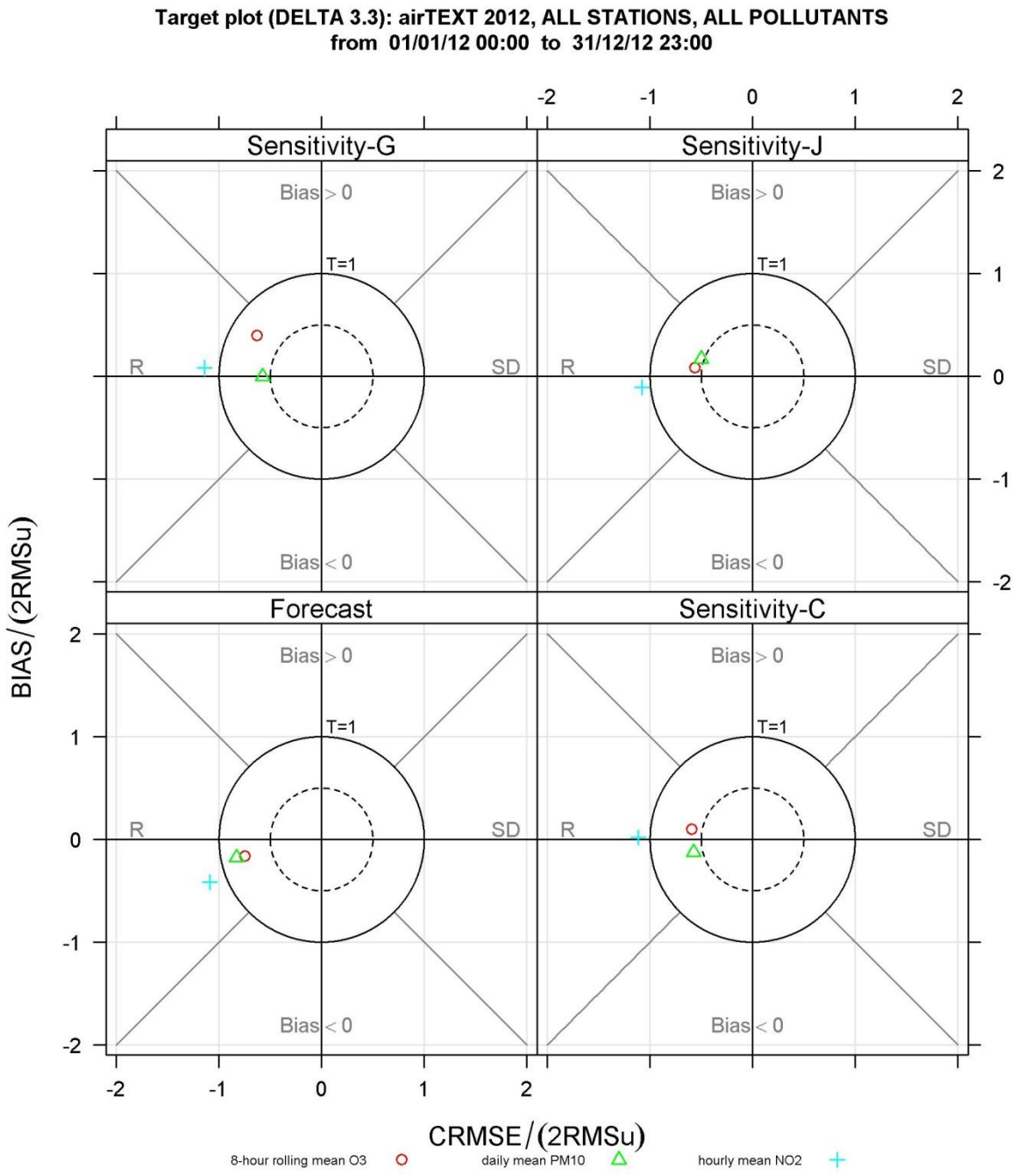


Figure 4.3 Example of a target plot (DELTA 3.3) from the concentration evaluation part of the Model Evaluation tool (grouped by pollutant and filtered by model)

4.2.1.3 Box and Whisker plot

The Box and Whisker plot shows 5 pieces of information for each station, for each of the modelled and observed datasets:

1. The lower whisker
2. The 25th percentile (the lower quartile – the lower end of the box)
3. The 50th percentile (the median – the horizontal line inside the box)
4. The 75th percentile (the upper quartile – the upper end of the box)
5. The upper whisker

The inter-quartile range (IQR) is defined as the 75th percentile minus the 25th percentile, i.e. the length of the box. The lower whisker is defined as the lowest concentration value still within 1.5xIQR of the lower quartile. The upper whisker is defined as the highest concentration value still within 1.5xIQR of the upper quartile. Optionally, the outliers lying outside of the upper and lower whisker can also be plotted, and the plot can be displayed on a $\log_{10}$ scale. An example box and whisker plot is shown in Figure 4.4.

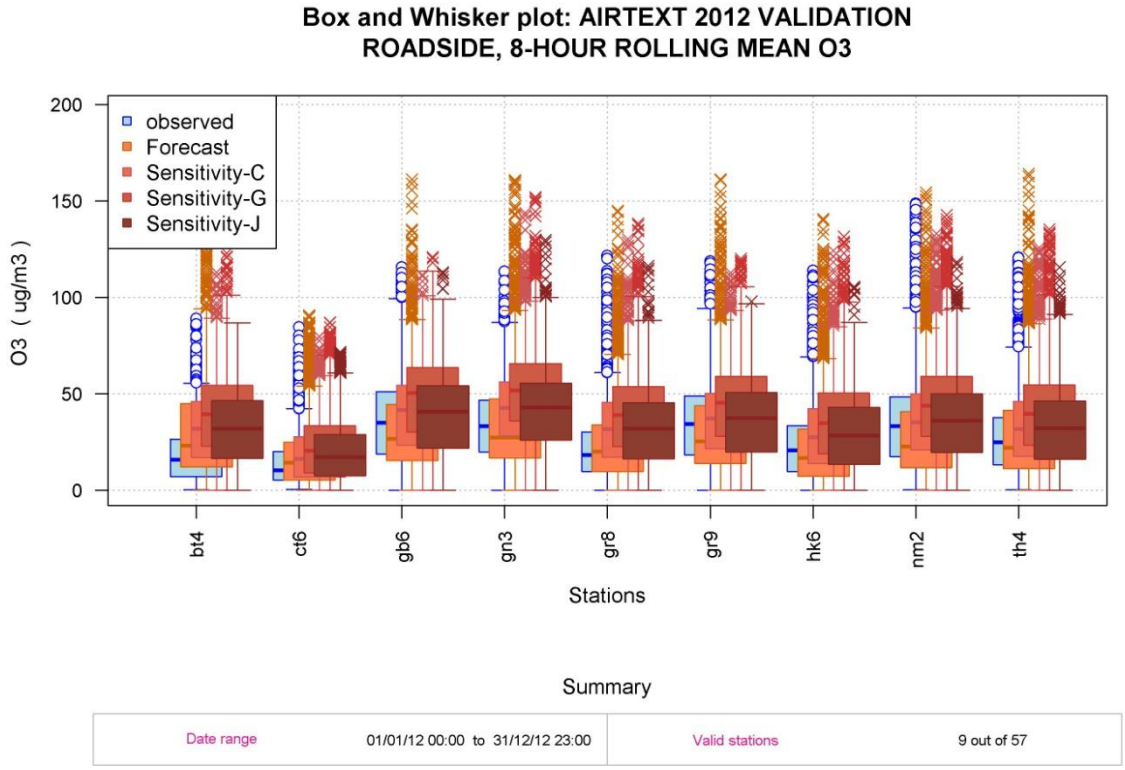


Figure 4.4 Example of a box and whisker plot from the concentration evaluation part of the Model Evaluation tool (grouped by station and filtered by model and pollutant; displaying single station type 'Roadside' for pollutant O_3)

4.2.1.4 Scatter Plot

This plot compares the modelled and observed concentrations on a scatter plot. The frequency scatter plot shows the frequency of occurrence of each point whereas the conventional scatter plot shows one point per pairwise modelled-observed data point. The data can optionally be plotted on a $\log_{10}$ scale. The black solid line is the 1:1 line. The dotted lines are the factor of 2 lines. Figure 4.5 presents an example frequency scatter plot and Figure 4.6 presents an example conventional scatter plot.

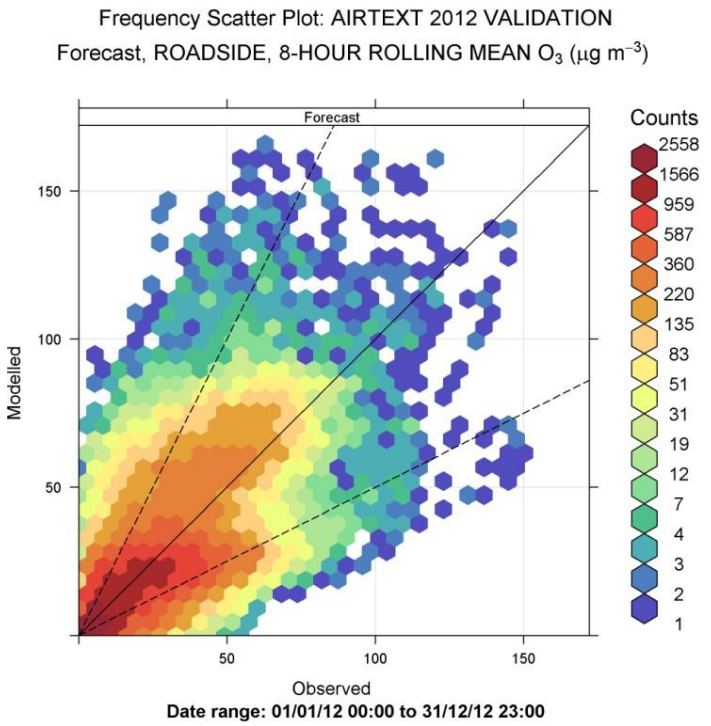


Figure 4.5 Example of a frequency scatter plot from the concentration evaluation part of the Model Evaluation tool (grouped by station and filtered by model and pollutant; displaying single model 'Forecast' and single station type 'Roadside' for pollutant O_3)

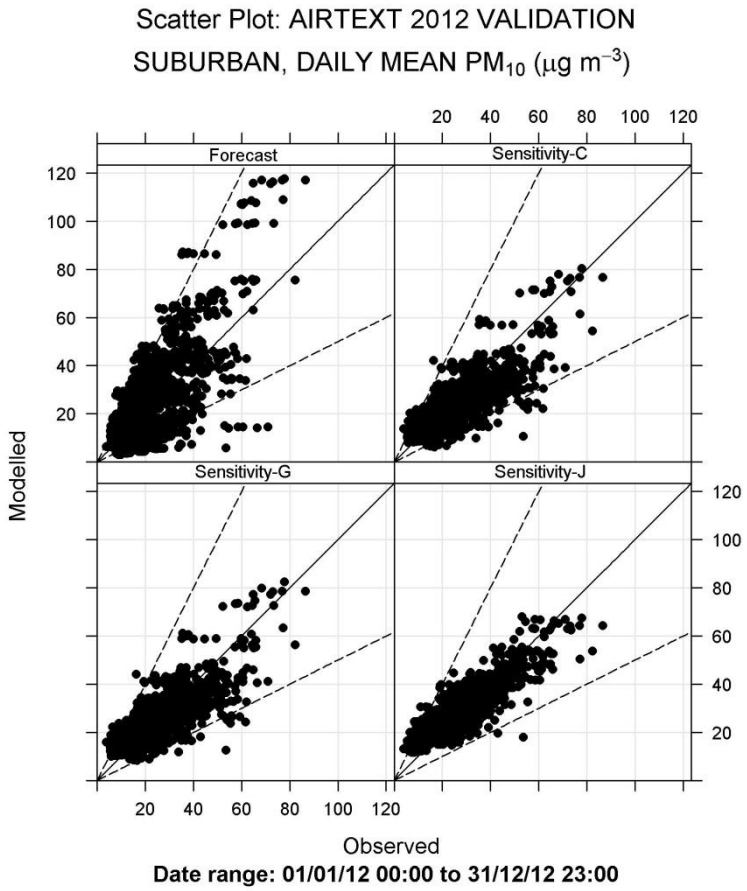


Figure 4.6 Example of a conventional scatter plot from the concentration evaluation part of the Model Evaluation tool (grouped by station and filtered by model and pollutant; displaying single station type 'Industrial' and pollutant PM_{10})

4.2.1.5 *Time plot averaged over all stations*

This plot compares the modelled to the observed concentrations for each pollutant in a number of time series plots, averaged over all stations. If a forecast assessment is also being carried out then the plot includes alert thresholds. This plot will only be produced by the Toolkit if the "Group by station, filter by model and pollutant" option is selected. Figure 4.7 shows an example time plot averaged over all stations.

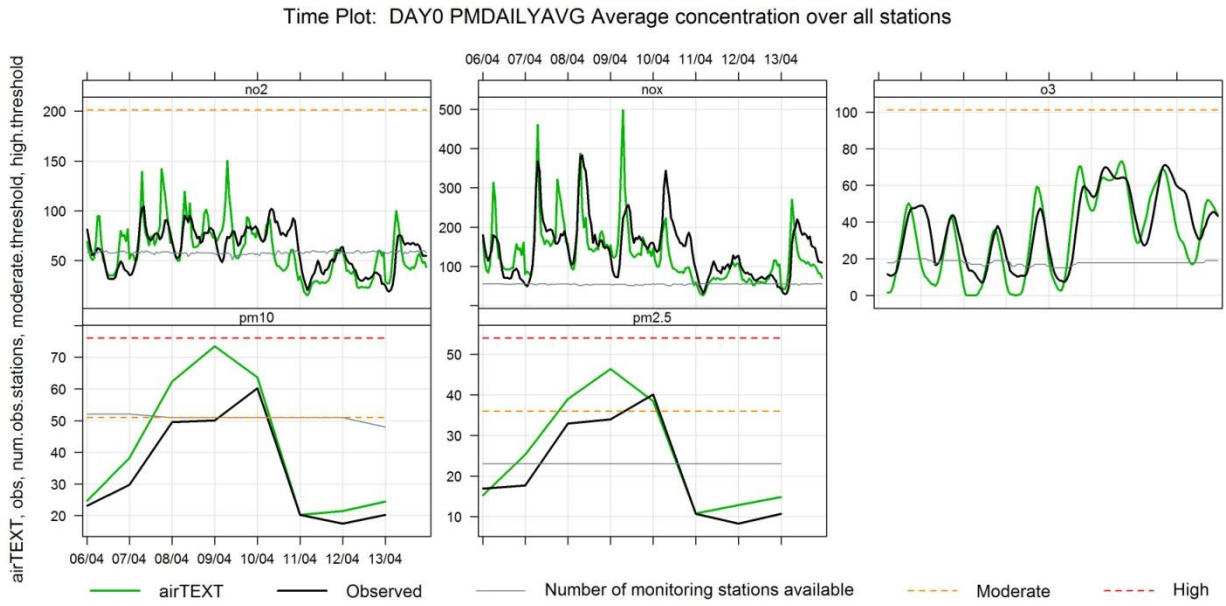


Figure 4.7 Example of a time plot averaged over all stations from the concentration evaluation part of the Model Evaluation tool (grouped by model and filtered by pollutant)

4.2.1.6 *Quantile-Quantile Plot*

This plot compares the modelled and observed concentrations ordered independently from lowest to highest concentration, as a quantile-quantile plot. The data can optionally be plotted on $\log_{10}$ scales. The black solid line is the 1:1 line. The dotted lines are the factor of 2 lines. Figure 4.8 presents an example quantile-quantile plot.

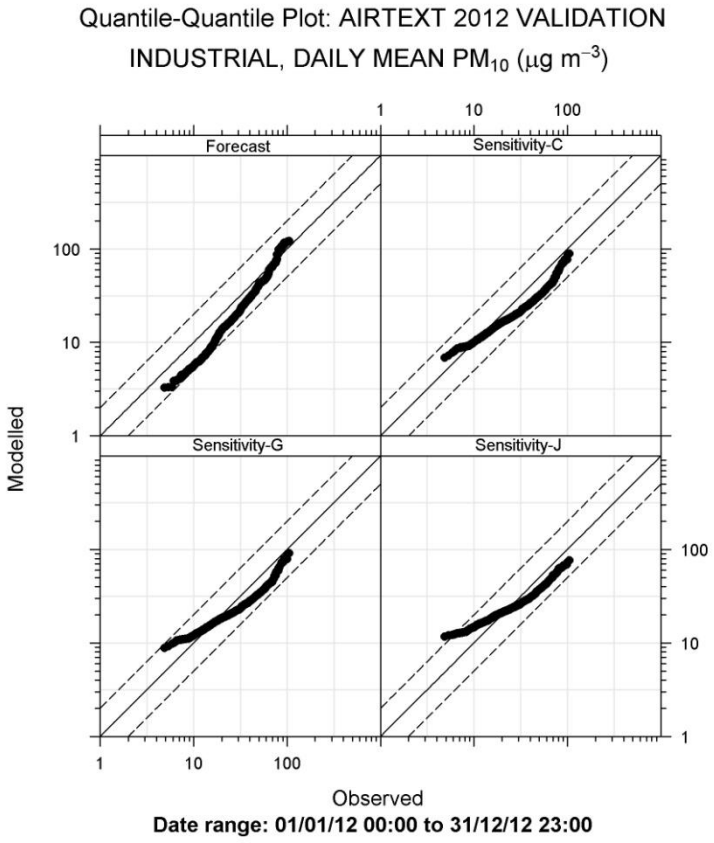


Figure 4.8 Example of a quantile-quantile plot from the concentration evaluation part of the Model Evaluation tool (grouped by station and filtered by model and pollutant; displaying pollutant PM_{10} on a $\log_{10}$ scale)

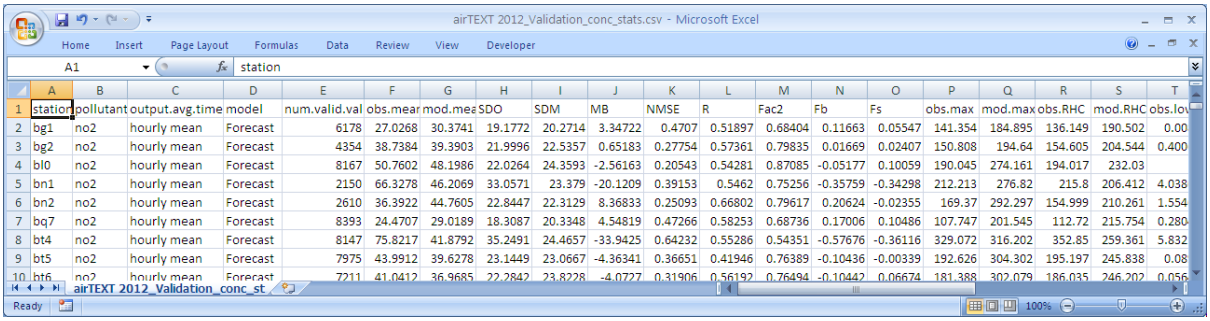
4.2.1.7 CSV output files

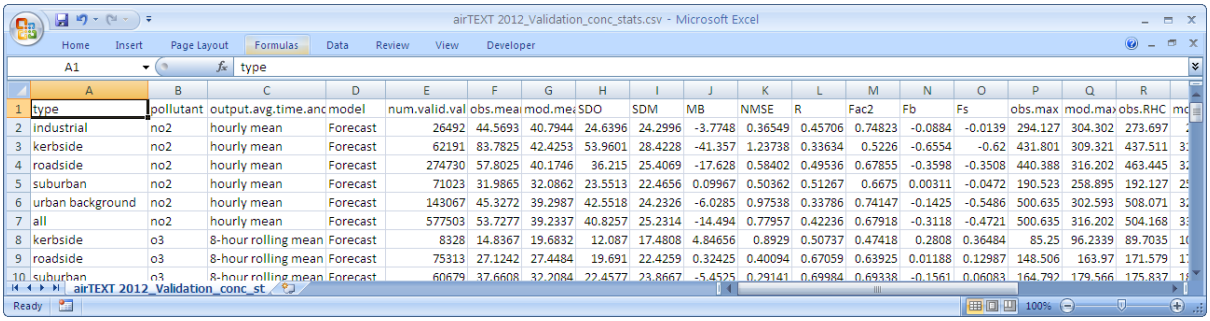
When the "output processed data and statistics as CSV" option is checked, statistics are calculated for each pollutant (in output units) and are output in a CSV file, *<*_conc_stats.csv>*. The statistics are calculated for the variable by which the data is grouped and additionally over all data, filtered by the filters selected. For example, if data is grouped by station and filtered by model and pollutant, the statistics are output for each station and for all valid stations in each model and pollutant.

The statistics that are output in are given in Table 4.1. Where 'obs' or 'mod' occur in variable names, these indicate observed or modelled values respectively.

Name	Description	Equation
num.valid.values	Number of values	
obs.mean mod.mean	Mean	$1/n \sum C$
SDO SDM	Standard Deviation	$\sqrt{1/n \sum (C - \overline{C})^2}$
MB	Mean Bias	$\overline{(C_p - C_o)}$
NMSE	Normalised Mean-Square-Error	$\overline{(C_p - C_o)^2} / \overline{C_o C_p}$
R	Pearson's Correlation Coefficient	$\mathrm{cov}(C_p, C_o)/\sigma_{C_p}\sigma_{C_o}$
Fac2	Factor of 2	Fraction of data where $0.5 \leq C_p/C_o \leq 2$ (when $C_o = 0$, $C_p/C_o \rightarrow \infty$ and the data pair is not counted)
Fb	Fractional Bias	$(\overline{C_p} - \overline{C_o})/0.5(\overline{C_o} + \overline{C_p})$
Fs	Fractional Standard Deviation	$(\sigma_{C_p} - \sigma_{C_o})/0.5\,(\sigma_{C_o} + \sigma_{C_p})$
obs.max mod.max	Maximum	$\max C$
obs.RHC mod.RHC	Robust Highest Concentration	$\chi(n) + (\chi - \chi(n))\ln\left(\frac{3n-1}{2}\right)$, where n is the number of values used to characterise the upper end of the concentration distribution, χ is the average of the $n-1$ largest values, and $\chi(n)$ is the n^{th} largest value; n is taken to be 26.

Table 4.1 Details of the statistics output by the concentration evaluation part of the Model Evaluation tool





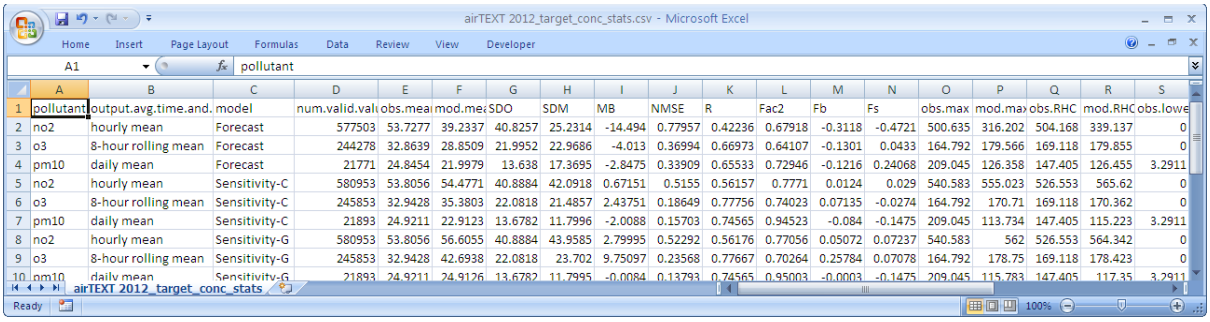


Figure 4.9 Examples of concentration evaluation output CSV file (grouped by station, station type and pollutant and filtered by model and pollutant)

In addition, some statistics for each station that are presented on the target plots and the box plots are also output in the CSV file for each station:

1. Lower whisker (observed)
2. Lower quartile (observed)
3. Median (observed)
4. Upper quartile (observed)
5. Upper whisker (observed)
6. Lower whisker (modelled)
7. Lower quartile (modelled)
8. Median (modelled)
9. Upper quartile (modelled)
10. Upper whisker (modelled)
11. centralised root-mean-square-error (CRMSE)
12. root-mean-square measurement uncertainty (RMSu)
13. normalised mean standard deviation (NMSD)
14. CRMSE.sign.DELTA.1.2 (CRMSE multiplied by +1 if SDO>SDM and multiplied by -1 if SDM>SDO)
15. Target parameter (T.DELTA.1.2)
16. CRMSE.sign.DELTA.3.3. (CRMSE multiplied by +1 if $\mathrm{NMSD} > \sqrt{2(1-R)}$ and multiplied by -1 if $\mathrm{NMSD} < \sqrt{2(1-R)}$)
17. Target parameter (T.DELTA.3.3)

In addition, if the “Include openair modStats results in CSV file” option is selected, then a number of further additional columns of data are added to the CSV file; these are described briefly in Table 4.2. For more information about the statistics generated by the modStats function, please refer to openair documentation [3].

Statistic	Description
N	The number of complete pairs of data
FAC2	Fraction of predictions within a factor of two
MB	The mean bias
MGE	The mean gross error
NMB	The normalised mean bias
NMGE	The normalised mean gross error
RMSE	The root mean squared error
R	The Pearson correlation coefficient.
COE	The *Coefficient of Efficiency* based on Legates and McCabe (1999, 2012) which spans from -1 to +1 with values approaching +1 representing better model performance
IOA	The Index of Agreement based on Willmott et al. (2011), which spans between -1 and +1 with values approaching +1 representing better model performance

Table 4.2 Details of the statistics used by the openair modStats package

If the 'Time plot averaged over all stations' option is selected, then an additional CSV file, *<*_conc_agg.csv>* is created containing the aggregated values plotted on the graph. More details are given in Table 4.3.

Name	Description
date	The date and time of each data point
pollutant	Name of the pollutant measured
obs	The measured concentration of a pollutant averaged over all relevant observed stations and averaging time determined by the *output averaging times* file. The units are determined by the *pollutants* file
airTEXT	The measured concentration of a pollutant averaged over all relevant modelled stations and averaging time determined by the *output averaging times* file. The units are determined by the *pollutants* file
num.obs.stations	Number of observation stations that measure each pollutant. The *obs* column values are averaged over these stations
num.airTEXT.stations	Number of modelled stations that measure each pollutant. The *airTEXT* column values are averaged over these stations
moderate.threshold	The moderated concentration threshold as defined in the *index scales* and *alert threshold* files. This column will contain *NA*s when Forecast Evaluation is not performed
high.threshold	The high concentration threshold as defined in the *index scales* and *alert threshold* files. This column will contain *NA*s when Forecast Evaluation is not performed

Table 4.3 Details of the headers in the CSV file produced when the 'Time plot averaged over all stations' option is selected

4.2.2 Forecast Index Evaluation Output

The forecast index evaluation part of the model evaluation tool produces three types of output; these are described in this section.

All three types of graph are based on a 'forecast index', which is calculated from the modelled and observed data according to the index threshold definitions in the index scales file (described in Section 3.2, step 9).

4.2.2.1 Forecast index accuracy

This part of the tool assesses the performance of the model's forecast index predictions against forecast indices calculated from observed concentrations.

The graph is a stacked bar chart that shows, for each station, the percentage of calculated forecast indices valid for comparison where the modelled index was equal to the observed index (green) and where the modelled index was equal to the observed index plus or minus one band (grey). Only forecast periods for which both modelled and observed indices can be calculated are included in the assessment. The stations are sorted by the number of indices valid for comparison, which is also shown on the chart by the blue circles and the right-hand y-axis. A key to the stations is given below the graph. Refer to Figure 4.10 for an example.

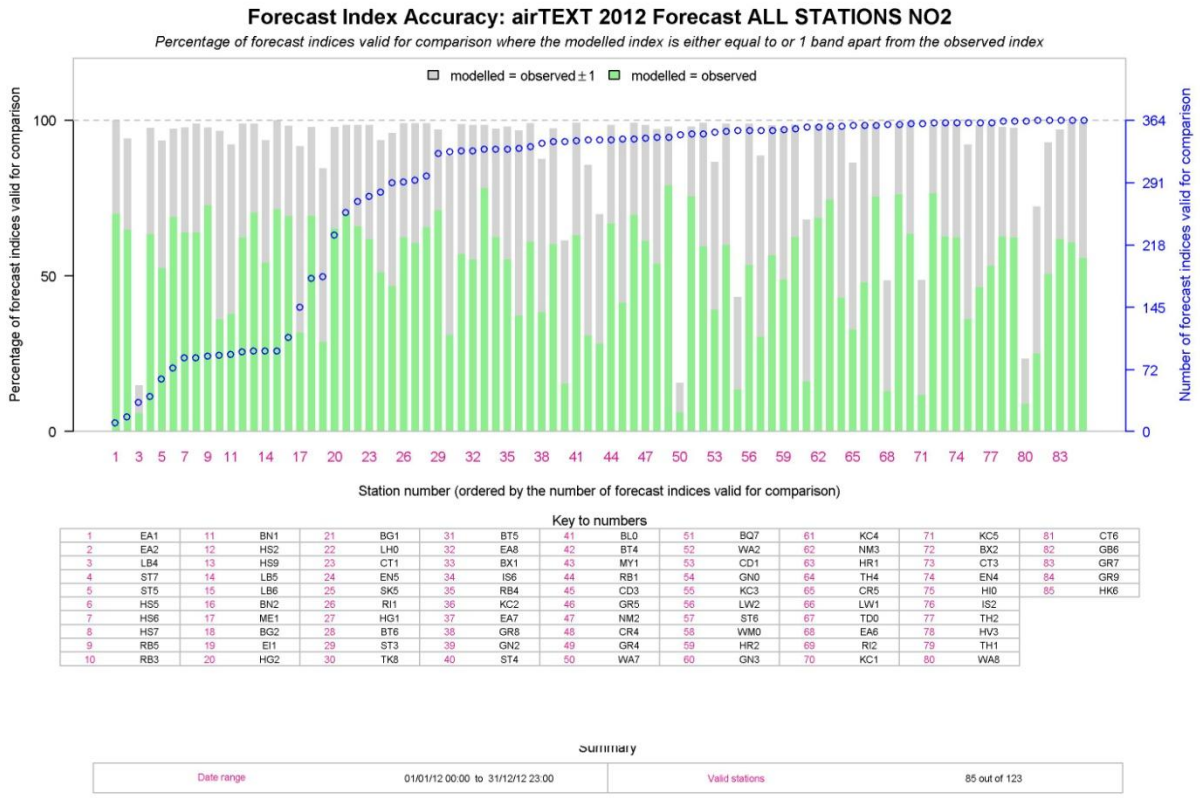


Figure 4.10 Example of the forecast index accuracy graph (grouped by station and filtered by model and pollutant; displaying model 'Forecast' and pollutant NO_2)

4.2.2.2 *Forecast alert accuracy*

It is usual in operational pollution forecasting to use pollution bandings to help communicate pollution levels to the public. For example, a common set of bandings for a 1 to 10 forecast index scale is shown in Table 4.4.

Band	Forecast index range	Alert threshold
LOW	1 to 3	n/a
MODERATE	4 to 6	4
HIGH	7 to 9	7
VERY HIGH	10	10

Table 4.4 Example set of bandings, with associated alert thresholds

Depending on the system, a forecast index in the MODERATE, HIGH or VERY HIGH range may trigger an alert to the public; it is therefore important for system operators to understand whether the system issues these alerts correctly.

The assessment of forecast alerts is carried out by calculating metrics for each monitoring station based on considering the exceedence of an alert threshold as an 'event'. The number of events observed and modelled, modelled but not observed, observed but not modelled, not modelled and not observed are summed to get the parameters *a*, *b*, *c* and *d* respectively. This is summarised in Table 4.5.

		Event observed	
		Yes	No
Event modelled	Yes	*a*	*b*
	No	*c*	*d*

Table 4.5 Definition of the forecast alert parameters

The forecast index evaluation includes two sets of graphical output for the assessment of the accuracy of forecast alerts:

1. **Odds ratio skill score (ORSS)**

   The odds ratio skill score (ORSS) is calculated from the alert metrics as follows:

$$\text{Odds ratio (OR)} = \frac{ad}{bc}$$

$$\text{ORSS} = \frac{OR - 1}{OR + 1}$$

   A perfect system will have *b* and *c* equal to zero, which means $\text{OR} \to \infty$, which means $\text{ORSS} \to 1$.

   A poor system will have *a* and *d* equal to zero, which means $\text{OR} = 0$, which means $\text{ORSS} = -1$.

   The odds ratio is a good metric for determining if a model is good at correctly issuing and not issuing alerts. It gives equal weighting to the correct prediction of an alert and to the correct non-prediction of a non-alert. If no alerts are observed or no alerts are forecast then ORSS is invalid. The graph shows ORSS for each station, where the stations are ordered by the number of observed alert threshold exceedences. The number of observed and modelled alerts for each station is also plotted in blue (right-hand y-axis). A key to the stations is shown below the graph. Figure 4.11 shows an example of an ORSS graph.

Figure 4.11 Example of a ORSS graph produced by the forecast index evaluation (grouped by pollutant and filtered by model; displaying model 'Forecast')

2. Performance metrics

In an operational pollution forecasting system, it is important not to issue alerts when there should not be an alert, but it is arguably *more* important to accurately issue an alert when an alert should be issued.

The following performance metrics give information about the skill of a model in terms of its ability to issue accurate alerts, and any tendency towards 'false alarms'.

$$\text{Probability of a correct forecast (PCF)} = \frac{a+d}{a+b+c+d}$$

$$\text{Probability of detection (POD)} = \frac{a}{a+c}$$

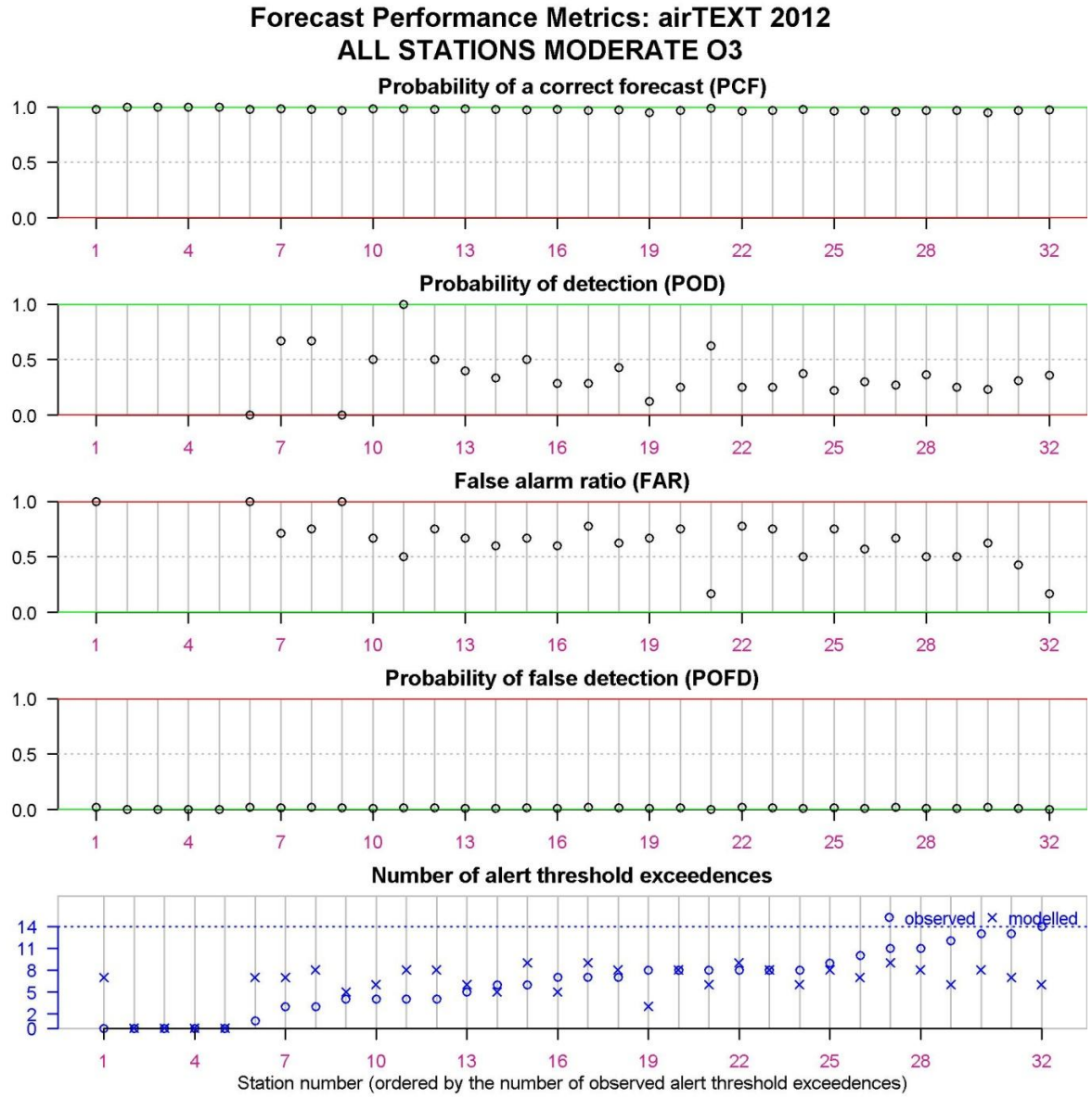
$$\text{False alarm ratio (FAR)} = \frac{b}{a+b}$$

$$\text{Probability of false detection (POFD)} = \frac{b}{b+d}$$

Each of these metrics lies in the range 0 to 1. A good score for PCF and POD is 1; a good score for FAR and POFD is 0.

The graphical output from the performance metrics option is a series of five graphs; the four metrics described above and a graph showing the number of observed and modelled alert threshold exceedences for each station. Again, the stations are sorted by the number

of observed alert threshold exceedences and a key to the stations is shown below the graphs. Figure 4.12 and Figure 4.13 show examples of this output grouped and filtered by different variables.



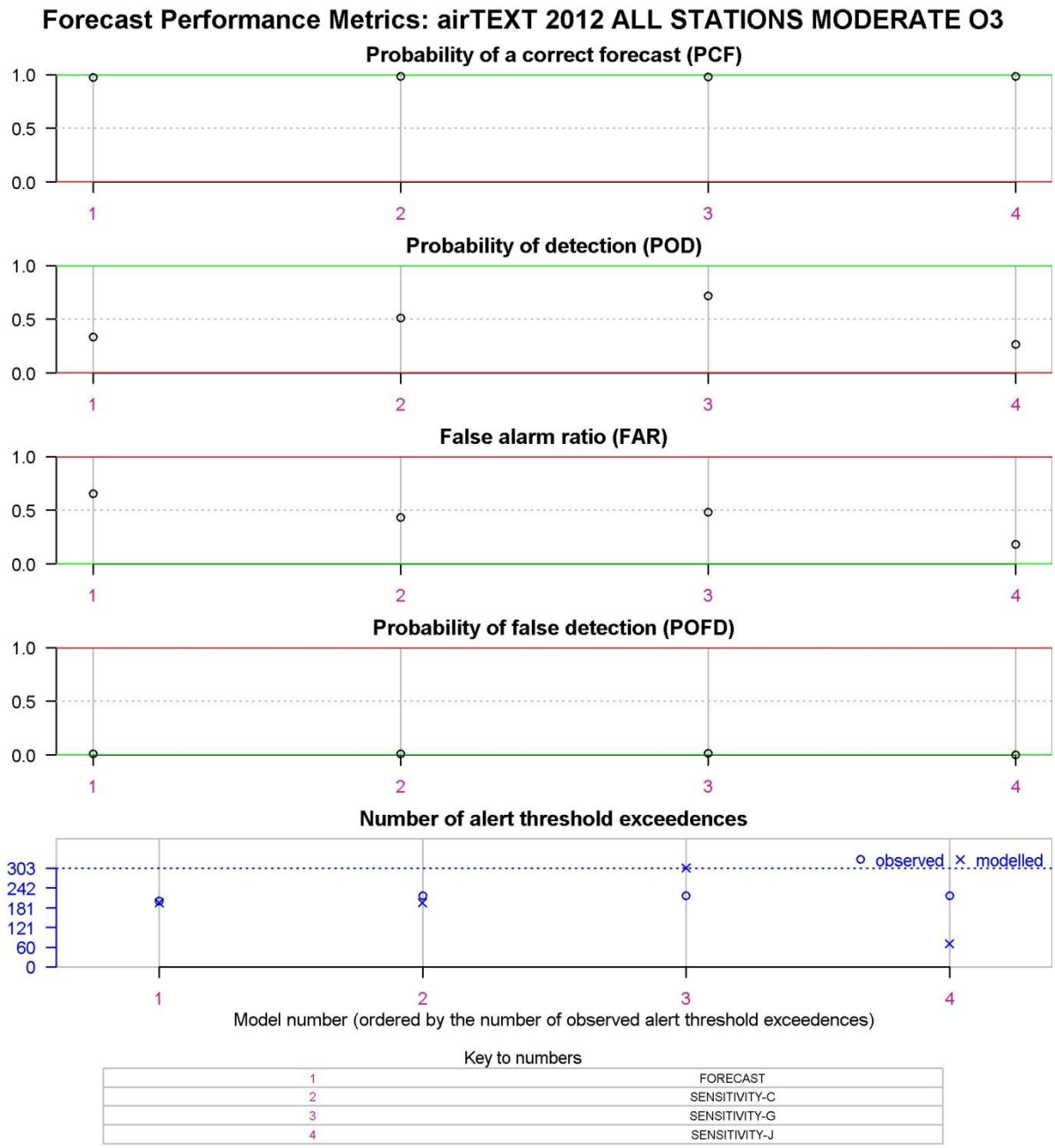
Key to numbers

1	BT4	11	LH0	21	HI0	31	TH1
2	CT6	12	TH4	22	RB1	32	RI2
3	EA1	13	GR9	23	ST3		
4	HS2	14	EA7	24	WM0		
5	MY1	15	TD0	25	GR4		
6	GN3	16	BL0	26	NM3		
7	HK6	17	BX1	27	BQ7		
8	LW1	18	WA2	28	NM2		
9	CR3	19	CT1	29	KC1		
10	GR8	20	GB6	30	HG2		

Summary

Date range	01/01/12 00:00 to 31/12/12 23:00	Valid stations	32 out of 123

Figure 4.12 Example of a forecast performance metrics graph produced by the forecast index evaluation (grouped by station and filtered by model and pollutant; displaying pollutant O3)



Key to numbers

1	FORECAST
2	SENSITIVITY-C
3	SENSITIVITY-G
4	SENSITIVITY-J

Summary

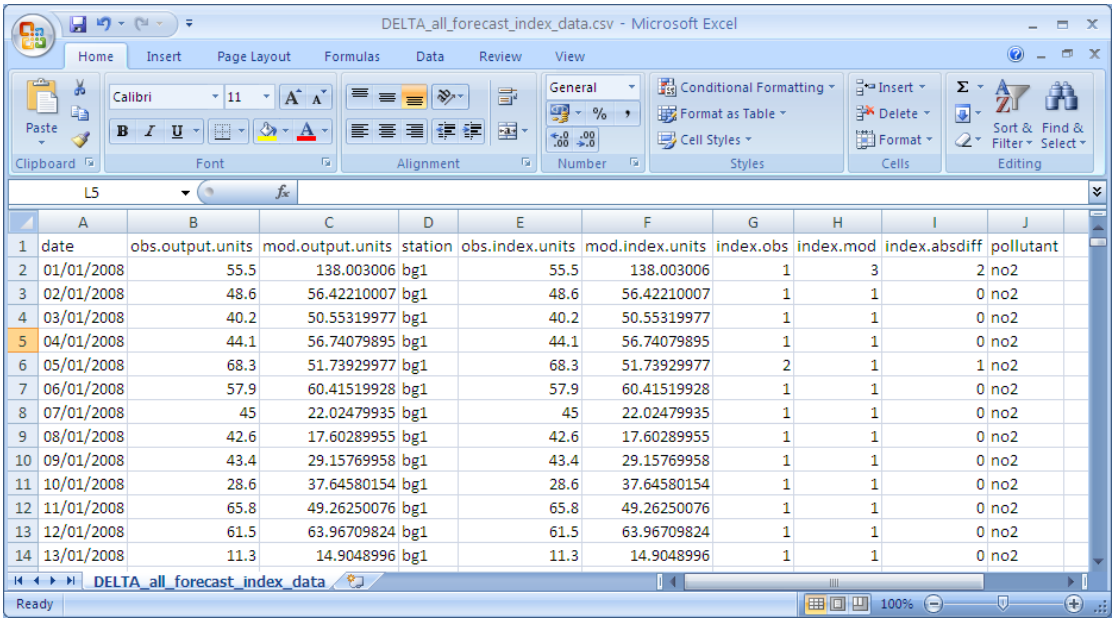
Date range	01/01/12 00:00 to 31/12/12 23:00	Valid models	4 out of 4

Figure 4.13 Example of a forecast performance metrics graph produced by the forecast index evaluation (grouped by model and filtered by pollutant; displaying pollutant O3)

4.2.2.3 CSV output files

*_forecast_index_data.csv:

This contains the modelled and observed concentrations using the index averaging time and statistic, in both output units and index units, also the observed and modelled index value for each averaging period and the absolute difference between the observed and modelled indices. For an example refer to Figure 4.14.

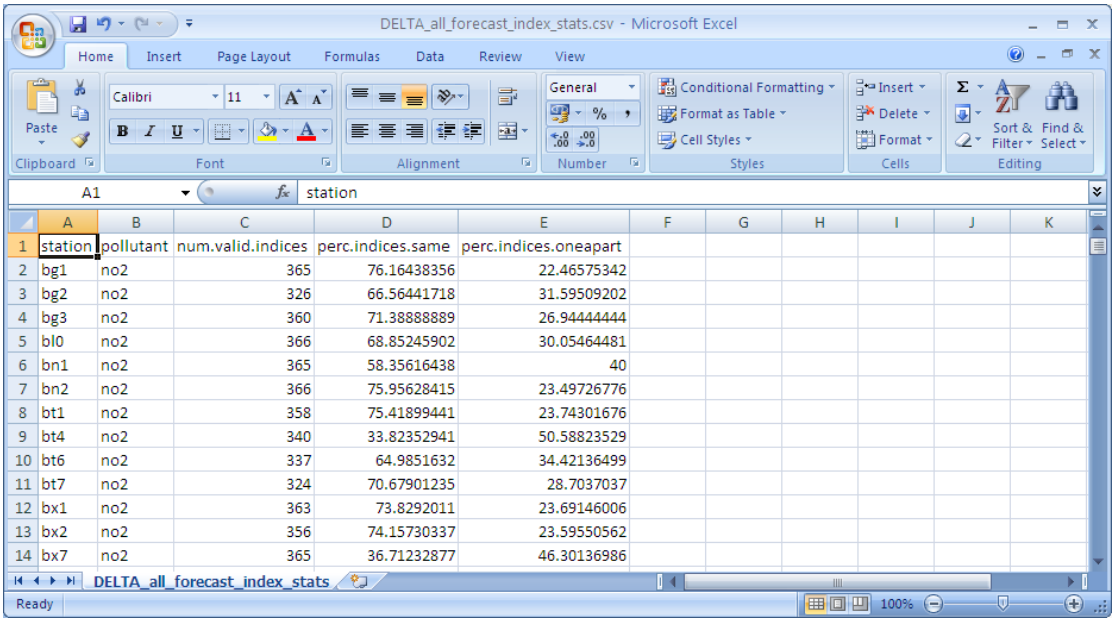


	A	B	C	D	E	F	G	H	I	J
1	date	obs.output.units	mod.output.units	station	obs.index.units	mod.index.units	index.obs	index.mod	index.absdiff	pollutant
2	01/01/2008	55.5	138.003006	bg1	55.5	138.003006	1	3	2	no2
3	02/01/2008	48.6	56.42210007	bg1	48.6	56.42210007	1	1	0	no2
4	03/01/2008	40.2	50.55319977	bg1	40.2	50.55319977	1	1	0	no2
5	04/01/2008	44.1	56.74079895	bg1	44.1	56.74079895	1	1	0	no2
6	05/01/2008	68.3	51.73929977	bg1	68.3	51.73929977	2	1	1	no2
7	06/01/2008	57.9	60.41519928	bg1	57.9	60.41519928	1	1	0	no2
8	07/01/2008	45	22.02479935	bg1	45	22.02479935	1	1	0	no2
9	08/01/2008	42.6	17.60289955	bg1	42.6	17.60289955	1	1	0	no2
10	09/01/2008	43.4	29.15769958	bg1	43.4	29.15769958	1	1	0	no2
11	10/01/2008	28.6	37.64580154	bg1	28.6	37.64580154	1	1	0	no2
12	11/01/2008	65.8	49.26250076	bg1	65.8	49.26250076	1	1	0	no2
13	12/01/2008	61.5	63.96709824	bg1	61.5	63.96709824	1	1	0	no2
14	13/01/2008	11.3	14.9048996	bg1	11.3	14.9048996	1	1	0	no2

Figure 4.14 Example of forecast index evaluation output *_forecast_index_data.csv file

*_forecast_index_stats.csv:

This contains, for each pollutant and station, the number of valid calculated index values, the percentage of indices where the modelled index was correct and the percentage of indices where the modelled index was one band either above or below the correct index. For an example refer to Figure 4.15.

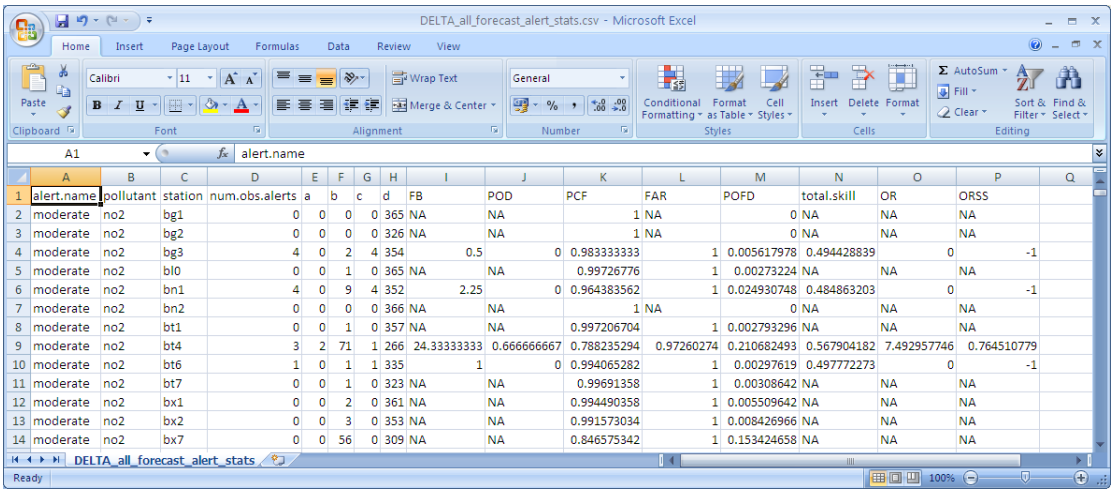


	A	B	C	D	E	F	G	H	I	J	K
1	station	pollutant	num.valid.indices	perc.indices.same	perc.indices.oneapart						
2	bg1	no2	365	76.16438356	22.46575342						
3	bg2	no2	326	66.56441718	31.59509202						
4	bg3	no2	360	71.38888889	26.94444444						
5	bl0	no2	366	68.85245902	30.05464481						
6	bn1	no2	365	58.35616438	40						
7	bn2	no2	366	75.95628415	23.49726776						
8	bt1	no2	358	75.41899441	23.74301676						
9	bt4	no2	340	33.82352941	50.58823529						
10	bt6	no2	337	64.9851632	34.42136499						
11	bt7	no2	324	70.67901235	28.7037037						
12	bx1	no2	363	73.8292011	23.69146006						
13	bx2	no2	356	74.15730337	23.59550562						
14	bx7	no2	365	36.71232877	46.30136986						

Figure 4.15 Example of forecast index evaluation output *_forecast_index_stats.csv file

*_forecast_alert_stats.csv:

This contains, for each station, for each pollutant, and for each alert threshold, the number of observed alerts, the 4 event parameters a, b, c and d, the performance metrics PCF, POD, FAR, POFD and also the odds ratio OR and the odds ratio skill score ORSS. Refer to Figure 4.16 for an example.



DELTA_all_forecast_alert_stats.csv - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	alert.name	pollutant	station	num.obs.alerts	a	b	c	d	FB	POD	PCF	FAR	POFD	total.skill	OR	ORSS	
2	moderate	no2	bg1	0	0	0	0	365	NA	NA	1	NA	0	NA	NA	NA	
3	moderate	no2	bg2	0	0	0	0	326	NA	NA	1	NA	0	NA	NA	NA	
4	moderate	no2	bg3	4	0	2	4	354	0.5	0	0.983333333	1	0.005617978	0.494428839	0	-1	
5	moderate	no2	bl0	0	0	1	0	365	NA	NA	0.99726776	1	0.00273224	NA	NA	NA	
6	moderate	no2	bn1	4	0	9	4	352	2.25	0	0.964383562	1	0.024930748	0.484863203	0	-1	
7	moderate	no2	bn2	0	0	0	0	366	NA	NA	1	NA	0	NA	NA	NA	
8	moderate	no2	bt1	0	0	1	0	357	NA	NA	0.997206704	1	0.002793296	NA	NA	NA	
9	moderate	no2	bt4	3	2	71	1	266	24.33333333	0.666666667	0.788235294	0.97260274	0.210682493	0.567904182	7.492957746	0.764510779	
10	moderate	no2	bt6	1	0	1	1	335	1	0	0.994065282	1	0.00297619	0.497772273	0	-1	
11	moderate	no2	bt7	0	0	1	0	323	NA	NA	0.99691358	1	0.00308642	NA	NA	NA	
12	moderate	no2	bx1	0	0	2	0	361	NA	NA	0.994490358	1	0.005509642	NA	NA	NA	
13	moderate	no2	bx2	0	0	3	0	353	NA	NA	0.991573034	1	0.008426966	NA	NA	NA	
14	moderate	no2	bx7	0	0	56	0	309	NA	NA	0.846575342	1	0.153424658	NA	NA	NA	

Figure 4.16 Example of forecast index evaluation output *_forecast_alert_stats.csv file

4.3 Model Diagnostics Tool Output

4.3.1 Time Variation Plot

This compares modelled and observed concentrations by (clockwise from top) hour and day of the week, day of the week, month of the year and hour of the day. The shaded area indicates the 95% confidence interval in the mean. For an example, refer to Figure 4.17.

4.3.2 Scatter Plot

This compares the modelled and observed concentrations on a scatter plot, optionally filtered by weekday, month or season. The frequency scatter plot shows the frequency of occurrence of each point whereas the conventional scatter plot shows one point per pairwise modelled-observed data point. When the calculated forecast indices are chosen in the data to plot, the output defaults to a frequency scatter plot. The black solid line is the 1:1 line. The dotted lines are the factor of 2 lines. Figure 4.18 to Figure 4.21 show examples of frequency and conventional scatter plots for different 'data to plot' and different filtering options.

4.3.3 Time Plot

This plots a time series of modelled and observed concentrations. Figure 4.22 and Figure 4.23 show examples of a time plot for different 'data to plot'.

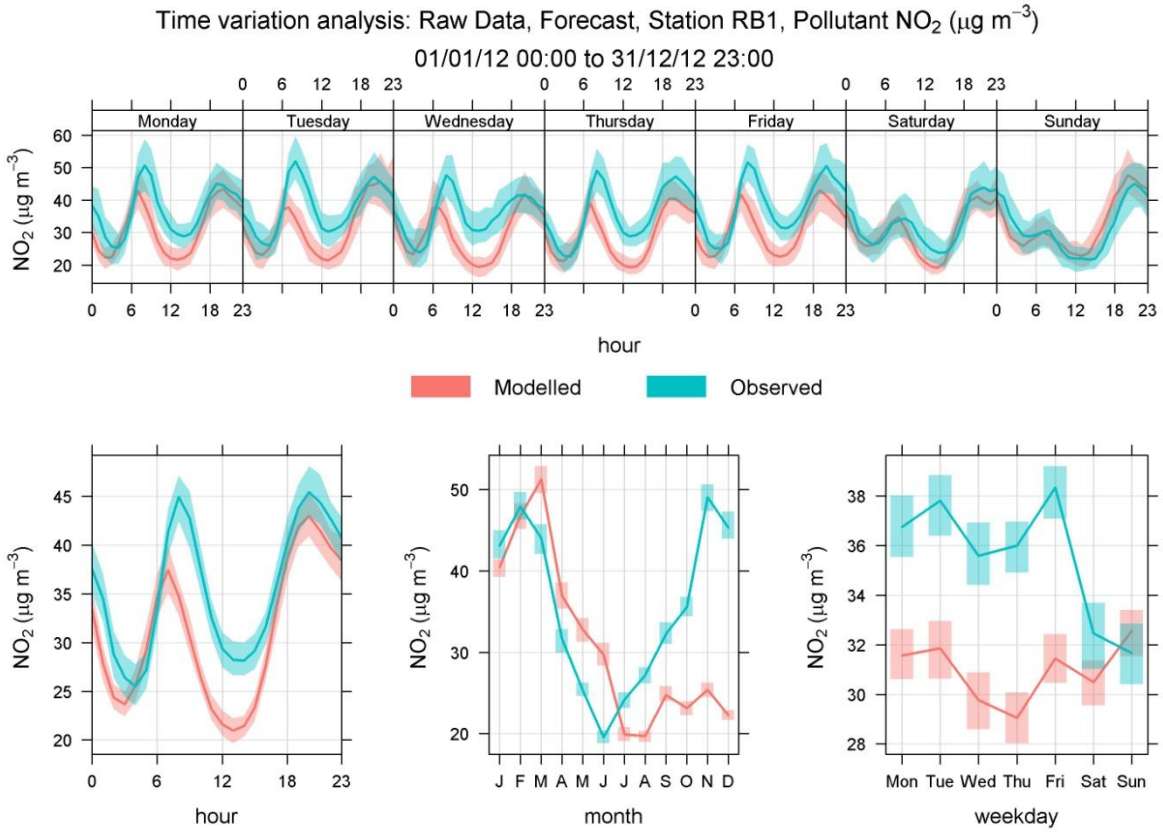


Figure 4.17 Example of a time variation plot from the Model Diagnostics tool.

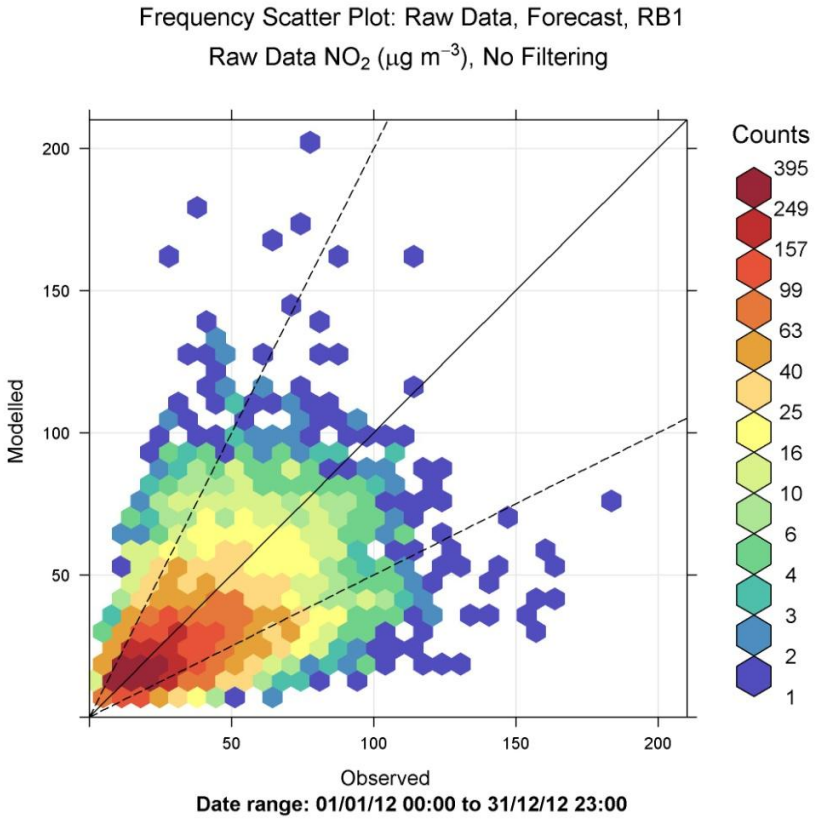


Figure 4.18 Example of a frequency scatter plot from the Model Diagnostics tool; raw data with no filtering

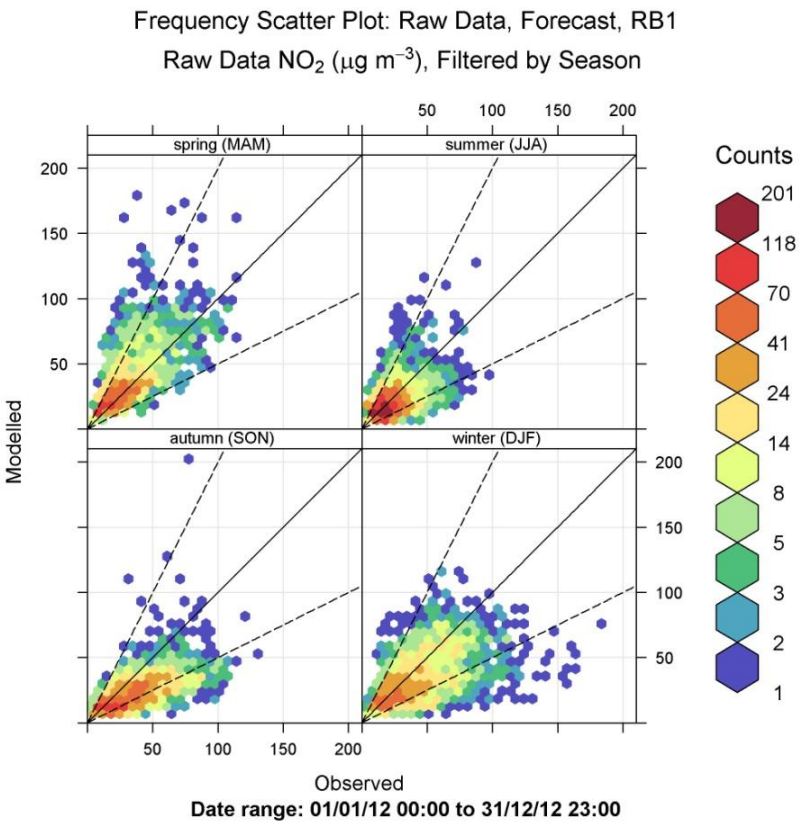


Figure 4.19 Example of a frequency scatter plot from the Model Diagnostics tool; raw data filtered by season

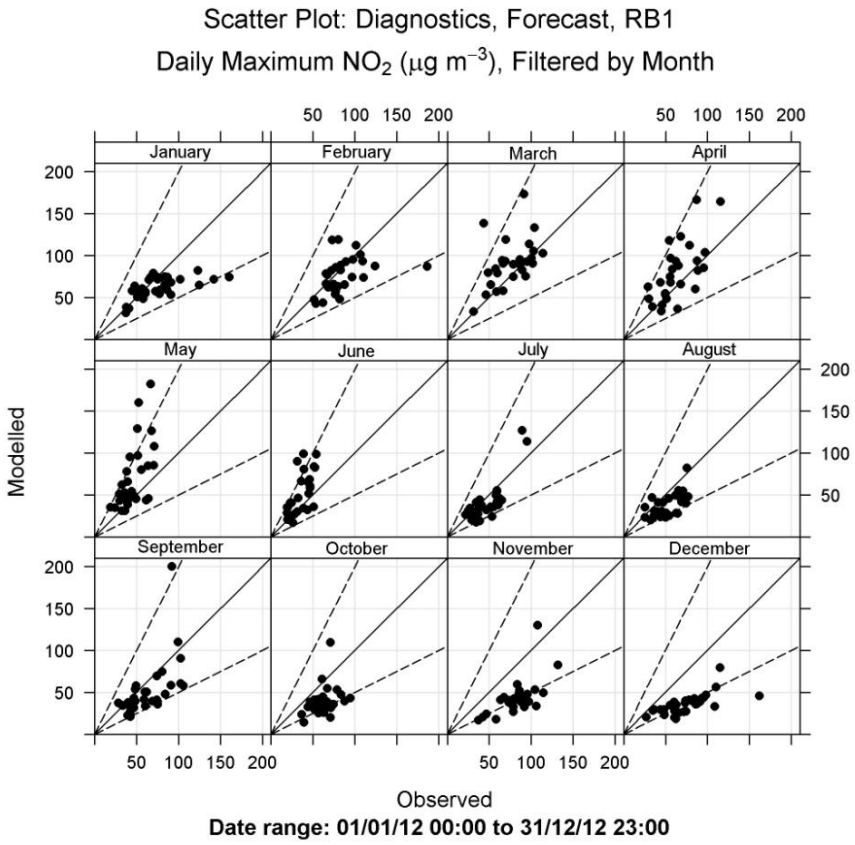


Figure 4.20 Example of a conventional scatter plot from the Model Diagnostics tool; calculated statistics (daily maximum NO_2) filtered by month

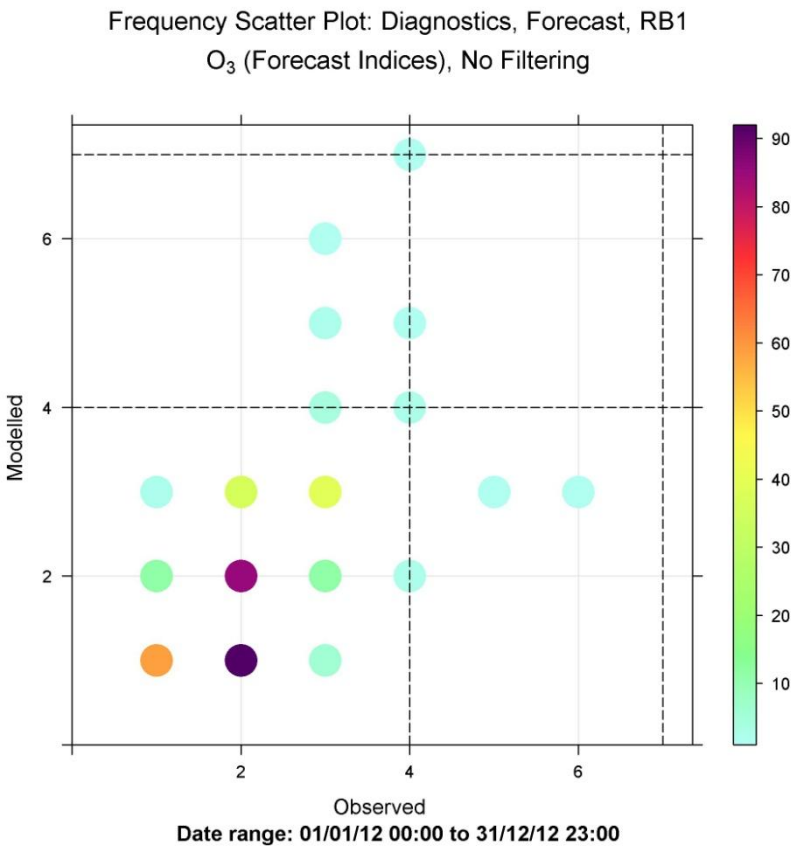


Figure 4.21 Example of a frequency scatter plot from the Model Diagnostics tool; forecast indices with no filtering

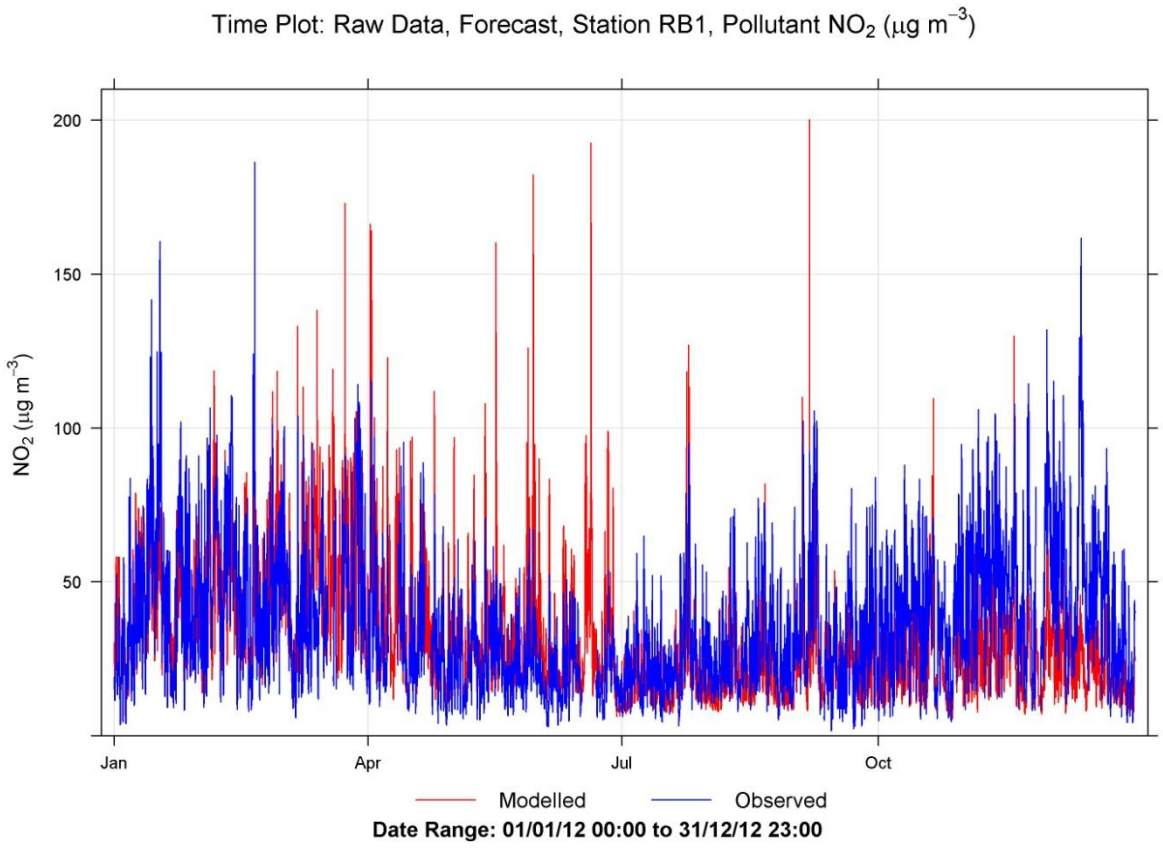


Figure 4.22 Example of a time plot from the Model Diagnostics tool; raw data

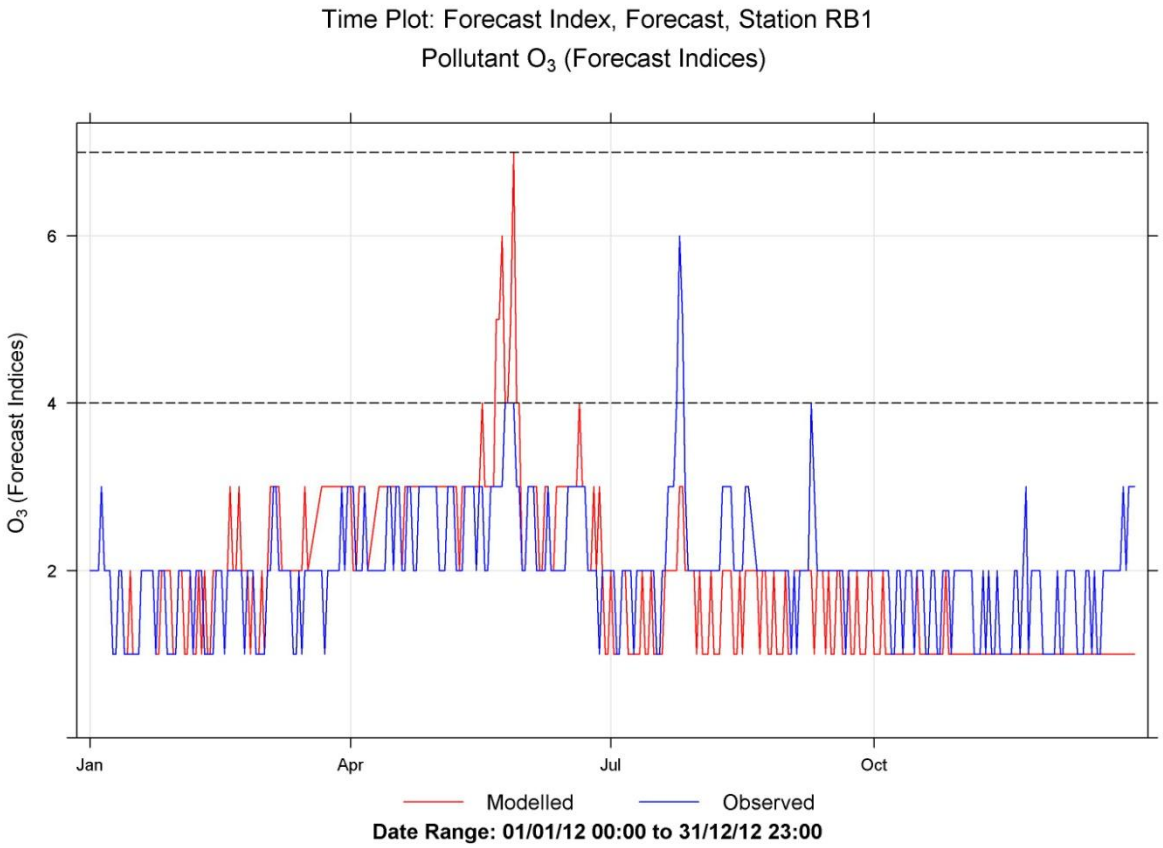
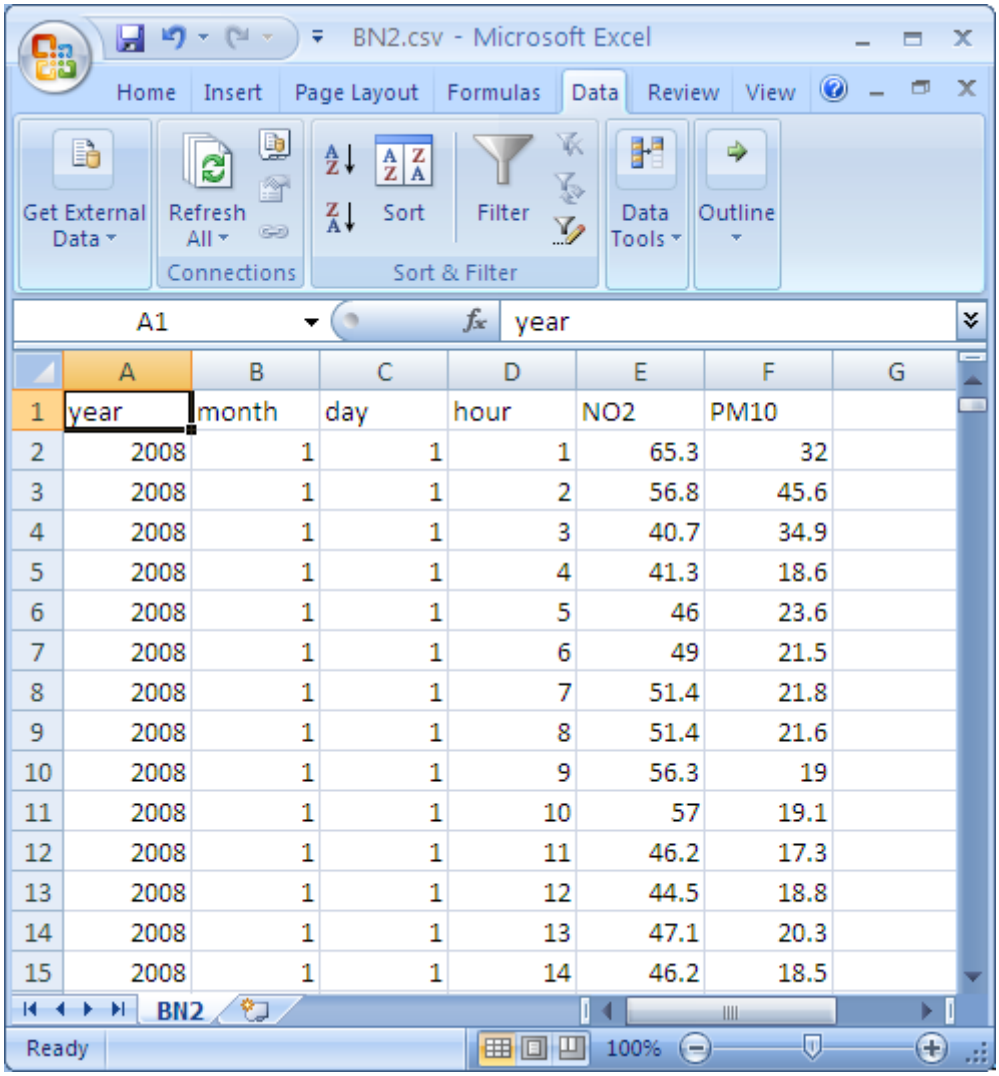


Figure 4.23 Example of a time plot from the Model Diagnostics tool; forecast indices

5 File formats

5.1 CSV modelled and observed data



	A	B	C	D	E	F	G
1	year	month	day	hour	NO2	PM10	
2	2008	1	1	1	65.3	32	
3	2008	1	1	2	56.8	45.6	
4	2008	1	1	3	40.7	34.9	
5	2008	1	1	4	41.3	18.6	
6	2008	1	1	5	46	23.6	
7	2008	1	1	6	49	21.5	
8	2008	1	1	7	51.4	21.8	
9	2008	1	1	8	51.4	21.6	
10	2008	1	1	9	56.3	19	
11	2008	1	1	10	57	19.1	
12	2008	1	1	11	46.2	17.3	
13	2008	1	1	12	44.5	18.8	
14	2008	1	1	13	47.1	20.3	
15	2008	1	1	14	46.2	18.5	

Figure 5.1 Example of an observed dataset in CSV file format

Restrictions:

- Include no more than one column per pollutant to be analysed
- The date must be given by the year, month, day and hour.
- The station column is optional. If it is present, the header should read “station”; if it is missing (as in the screenshot above), then all the data in that file is assigned to a ‘station’ with the name of the filename, e.g. BG1.csv would be assigned to station ‘BG1’.
- Modelled data must contain data for each pollutant to be analysed. Note that if multiple files are used, that means each pollutant to be analysed must be present in at least one file. Observed data does not have to include data for all required pollutants.

6 R Support

The three Model Evaluation Toolkit tools each contain a number of utilities for managing the interaction of the Toolkit with R.

6.1 R Settings screen

Each tool has an 'R Settings' screen. The Data Input Tool R Settings screen with the default settings is shown in Figure 6.1.

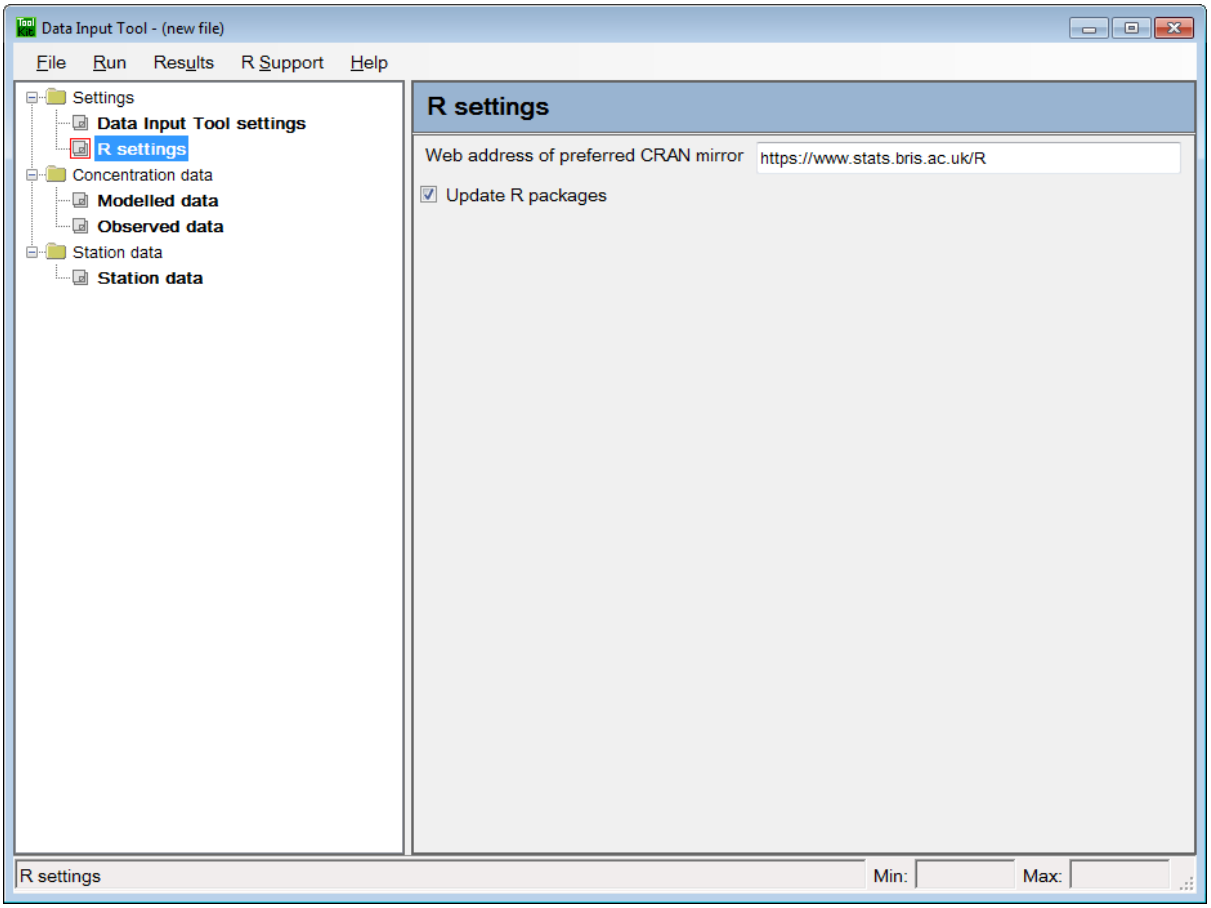


Figure 6.1 Data Input Tool R Settings screen, showing the default settings

6.1.1 Web address of preferred CRAN mirror

The CRAN mirror must be specified and should ideally be the mirror most relevant for the user's geographical location. The default mirror web address is for the UK; for other locations go to the 'R Support' menu and select 'See a full list of CRAN mirrors...'. This will bring up the R project website that lists the web addresses of all CRAN mirrors; copy the link that applies to your location and paste into the Toolkit. To use this CRAN mirror as the default mirror in future, see Section 7.

6.1.2 Update R packages

Each time a Model Evaluation Toolkit Tool is run, by default any installed R packages will be updated to the latest version, according to the CRAN mirror specified. If you do not wish the Toolkit to update packages in this way for any reason, untick this option.

6.2 R Support menu

Each tool has an 'R Support' menu, as shown in Figure 6.2.

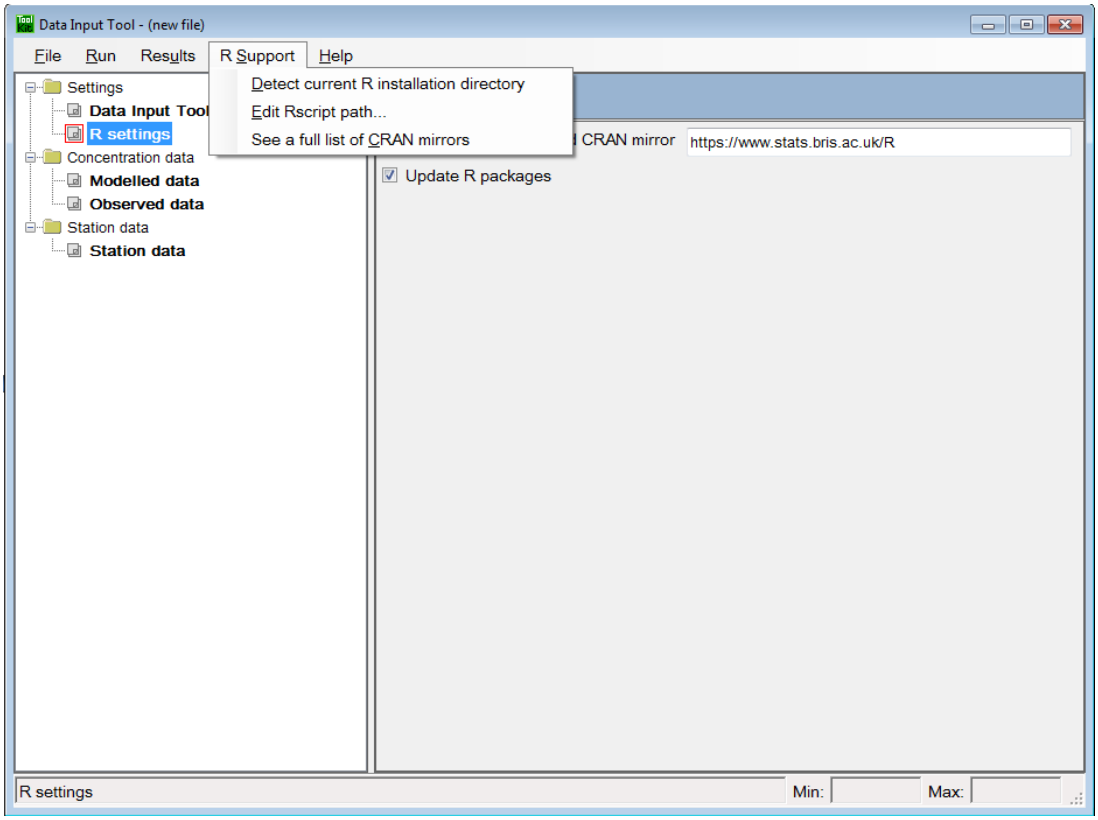


Figure 6.2 Data Input Tool R Support menu

6.2.1 Detect current R installation directory

The first time you run any Model Evaluation Toolkit tool, the Toolkit will automatically set itself to use the *version of R that was most recently installed*. If you install a different version of R and wish to update the Toolkit to use this newly-installed version of R, select 'Detect current R installation directory' to re-set the installation used by the Toolkit to the most recently installed version.

Note: The .log file for any tool run records the version of R used.

6.2.2 Edit Rscript path...

As described above, by default the Toolkit will use the most recently installed version of R. If you want the Toolkit to use a different installed version of R, click 'Edit Rscript path...' to open the file shown below in Figure 6.3. Edit the path as required and save the file.

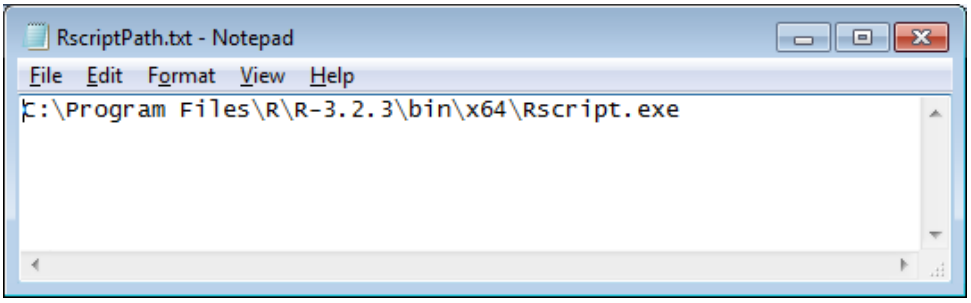


Figure 6.3 The RscriptPath.txt file

6.2.3 See a full list of CRAN mirrors…

Select this option to bring up the R project website that lists the web addresses of all CRAN mirrors, as shown in Figure 6.4.

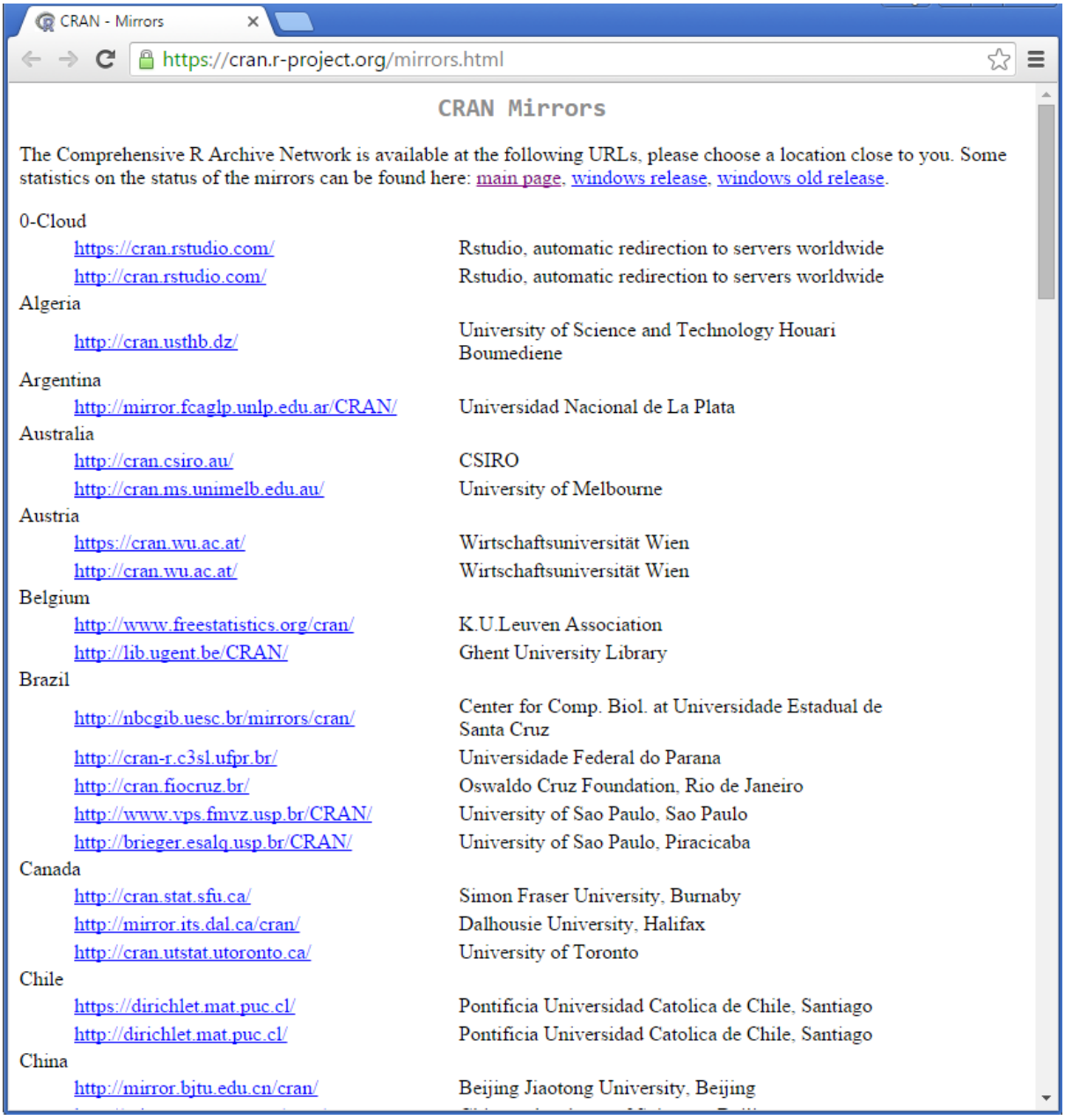


Figure 6.4 Screenshot of the R project website that lists all CRAN mirrors

7 Using a template file

If you wish to change any of the default settings used by any of the Toolkit tools to defaults that are more appropriate for you, for example if you are outside the UK and need to use a different CRAN mirror, you can do this using a 'template' file.

Step 1: **Open the tool concerned**

Step 2: **Change the settings concerned to your new default values**

Step 3: **Save the file**

Step 4: **From the 'File' menu, select 'Preferences...' and then select the 'Template' tab**

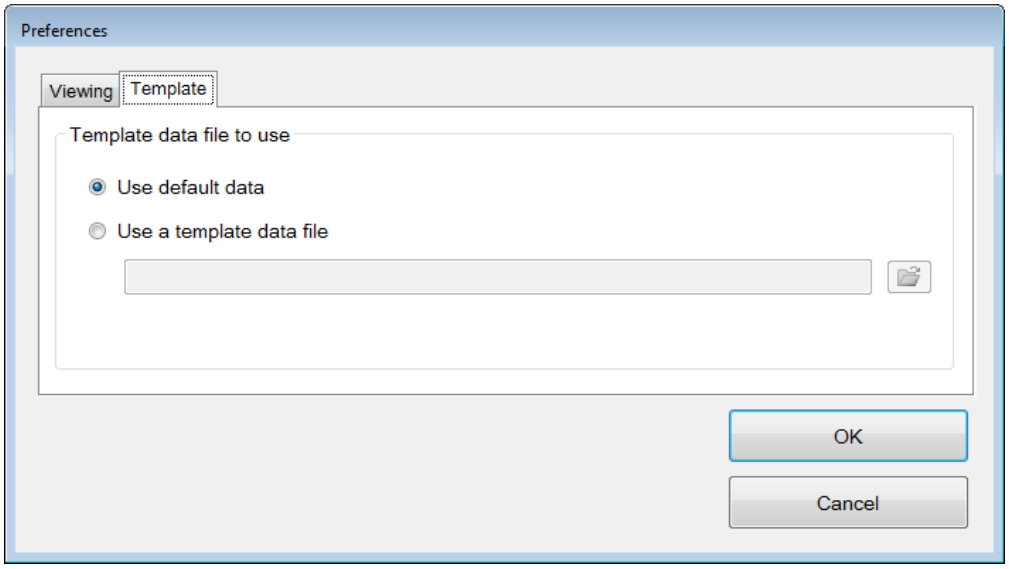


Figure 7.1 Preferences screen showing the Template Tab

Step 5: **Select 'Use a template data file' and browse to select the file you saved in Step 3.**

Step 6: **Click OK.**

Step 7: **If necessary (for instance to change the default CRAN mirror) repeat Steps 1 to 6 for the other tools**

In future, any new input files will use the settings in your saved template files as the default settings instead of the standard default settings. If at any time you wish to revert to the standard default settings, simply change the Preferences screen setting back to 'Use default data'.

8 Batch mode facility

The Data Input, Model Evaluation and Model Diagnostics tools can be run from the command-line, making it possible to automate tasks, for example in a batch file.

The procedure is the same for each of the above tools:

1. Open the tool, enter your settings and save the settings file
2. On the command-line (or in a batch file) type

   {R's Rscript.exe path} {Path of the tool's .r file} {settings file path}

More information is given below about the three components of the instruction:

{R's Rscript.exe path}: This is the pathname of the Rscript.exe file installed with R. This must be used when running R from the command line. A typical path on a Windows PC (where R was installed in "C:\Program Files\R\R-3.2.3") would be

"C:\Program Files\R\R-3.2.3\bin\x64\Rscript.exe"

{Path of the tool's .r file}: This is the pathname of the R-script for the tool you want to run. For example, for the Data Input tool, if the Toolkit is installed at "C:\ModelEvaluationToolkit" then the path would be

"C:\ModelEvaluationToolkit\DataInput.r"

{settings file path}: This is the path of the settings file that you saved. For example, if you saved a Data Input tool settings file to the directory "C:\Toolkit_output" with the filename "toolkit" then the settings file path would be

"C:\Toolkit_output\toolkit.tki"

9 References

1. Thunis, P., A. Pederzoli, E. Giorgieva, C. Cuvelier and D. Pernigotti (2013). **The DELTA tool and Benchmarking Report template: Concepts and User guide version 3.2.** http://aqm.jrc.ec.europa.eu/index.aspx
2. Carslaw, D.C. and K. Ropkins, (2012). **openair — an R package for air quality data analysis**. Environmental Modelling & Software. Volume 27-28, pp. 52–61.
3. Carslaw, D.C. (2015). **The openair manual — open-source tools for analysing air pollution data**. Manual for version 1.1-4, King's College London. http://www.openair-project.org/PDF/OpenAir_Manual.pdf
4. Thunis, P., D. Pernigotti and M. Gerboles (2013). **Model quality objectives based on measurement uncertainty. Part I: Ozone.** Atmospheric Environment, In Press. http://dx.doi.org/10.1016/j.atmosenv.2013.05.018
5. Thunis, P., D. Pernigotti, M. Gerboles and C. Belis (2013). **Model quality objectives based on measurement uncertainty. Part II: PM10 and NO2.** Atmospheric Environment, Submitted.
6. R Development Core Team (2010). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.