
Instruction Manual for the OSPM Validation Tool

Created: Thor-Bjørn Ottosen *tbot@dmu.dk*
Checked: Matthias Ketzel *mke@dmu.dk*

November 25, 2014

Version 1.20

The OSPM Validation Tool is used for exploratory data analysis of the performance of the OSPM® Model in comparison to measurements. It is meant for generating a first set of quantitative model performance metrics. Additionally, it is possible to analyse the dependency of model performance on various parameters (e.g. wind direction, hour of day etc.).

The following guide will cover how to obtain the tool as described in section 1, an introductory description of the elements of the tool is given in section 2, the everyday usage of the OSPM Validation Tool as described in section 3 and inserting the user's own data is described in section 4.

1 Obtaining the Validation Tool

The OSPM validation tool can be downloaded in a fully functional version with a set of sample data as a zip-file from the webpage <http://envs.au.dk/en/knowledge/air/models/background/ospmtool>. The tool consists of an Excel-workbook containing both the example data, the user-interface, and the output.

2 Overview of the Worksheets

When the workbook opens, the worksheet shown on fig. 1 appears. The active worksheet is *SummaryYear* which contains the long time series of the modelled and measured concentrations. The worksheets *Pivot_wdir*, *Pivot_hour*, and *Pivot_month* are the worksheets used for visualising the data with respect to wind direction, hour of day, and seasonality.

Pivot_wdir, Pivot_hour, and Pivot_month:	Most of the graphs displayed on the worksheets <i>Pivot_wdir</i> , <i>Pivot_hour</i> , and <i>Pivot_month</i> contain three curves: A blue representing observations, a pink representing model results, and a green representing background concentrations. This can also be seen from the variable names, where the variable names ending on " <i>_obs</i> " represent observations, the variable names ending on " <i>_mod</i> " represent model results, and the variable names ending on " <i>_b</i> " represent background values. Two graphs do not follow this pattern:
--	---

- The first graph, as shown on fig. 2, shows the number of data points as a function of a variable, in this case, wind direction. A similar graph for the number of data points as a function of respectively hour of day and month of year exists on the worksheets *Pivot_hour* and *Pivot_month*. As with the other graphs, the blue color represent observations and the pink color represent model results. The function *nXY* is a system variable and should be ignored by the user.

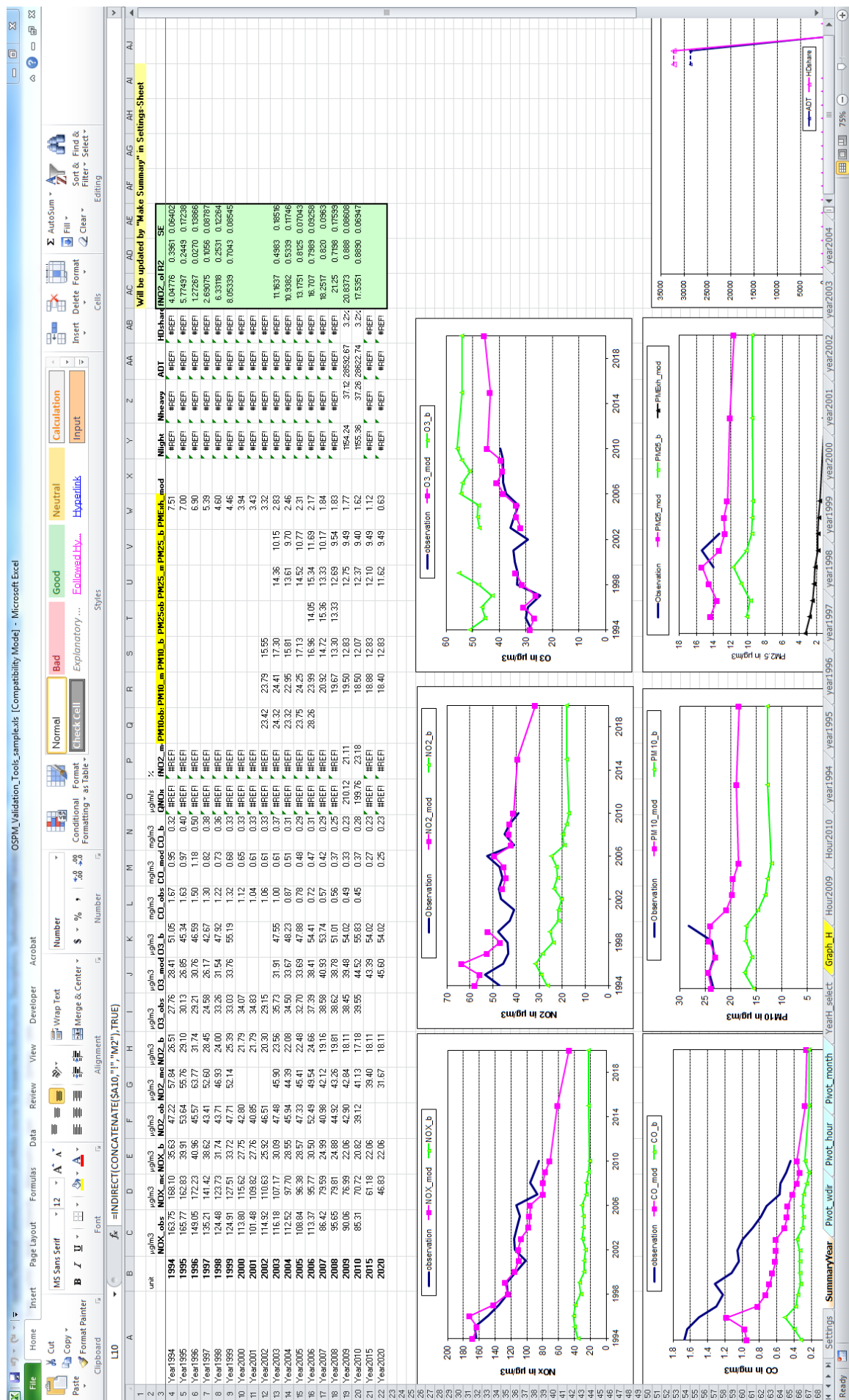


Figure 1: Screenshot of the opening window of the OSPM Validation Tool

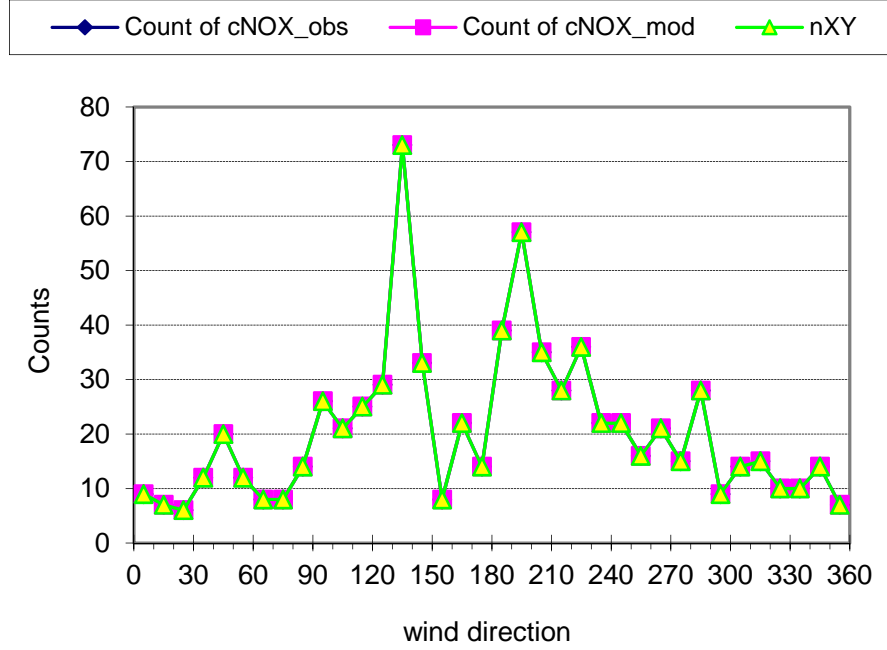


Figure 2: Illustration of the graph displaying the number of data points as a function of wind direction

- The second graph, as shown on fig. 3, shows the percentage of directly emitted NO_2 in respectively the model and the observations (see the following section for the details of the calculation). Again, similar graphs with the percentage depicted as functions of respectively hour of day and month of year exist on the worksheets *Pivot_hour* and *Pivot_month*.

O_x : To understand fig. 4 it is necessary to introduce the conception of O_x :

$$[\text{O}_x] = [\text{O}_3] + [\text{NO}_2]$$

Where $[X]$ is the measured or modelled concentration of specie X in ppb. If no NO_2 were directly emitted from the vehicles the concentration of O_x would be constant from the urban background to the street canyon, since NO is converted to NO_2 in the presence of O_3 . However, since there is a direct emission of NO_2 from the vehicles the concentration of O_x is larger in the street canyon than in the urban background:

$$[\text{O}_x]_{\text{st}} - [\text{O}_x]_{\text{bg}} = p \left([\text{NO}_x]_{\text{st}} - [\text{NO}_x]_{\text{bg}} \right) \quad (1)$$

Where X_{st} is the street canyon concentration of compound X and X_{bg} is the urban background concentration of compound X . Thus dividing the increase in O_x from the urban background to the street canyon by the increased NO_x concentration in the street canyon yields the fraction of directly emitted NO_2 represented by p in eq. (1).

Graph_H: The worksheet *Graph_H* contains a series of scatter plots, where various measurements and model data are compared. A large group of the scatter plots on the worksheet *Graph_H* are plots of the measured against the modelled data for various chemical species. Another group of plots show scatter plots of both the measured and modelled data points as a function of e.g. wind speed. These are used to illuminate model-measurement discrepancy dependencies on various variables.

SummaryYear: The worksheet *SummaryYear* contains the time series spanning several years. This is used to illuminate the long time temporal trend in concentrations and model performance. Again, most of the graphs contain a blue curve representing measurements, a

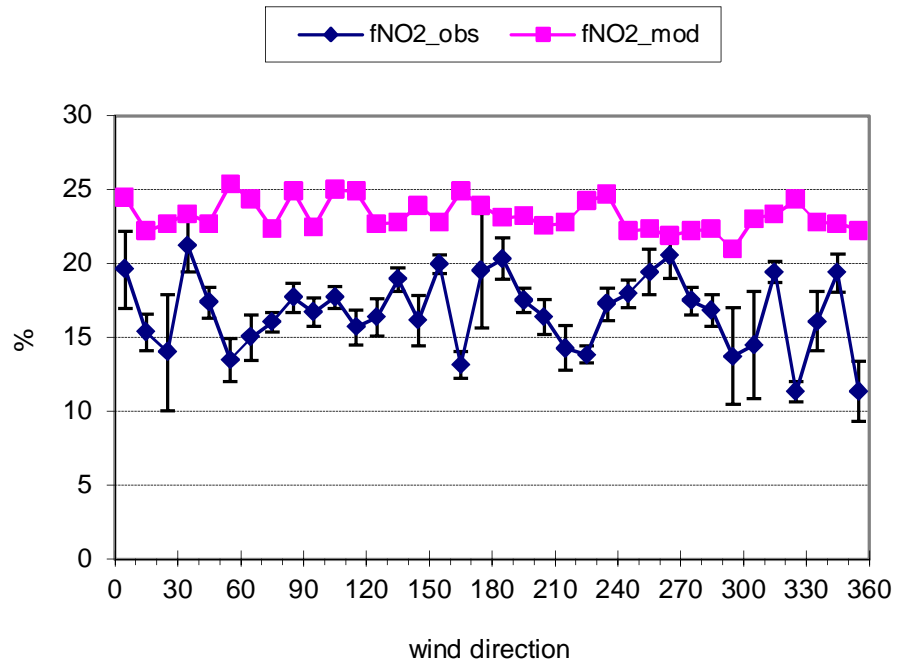


Figure 3: Percentage of directly emitted NO_2 as a function of wind direction.

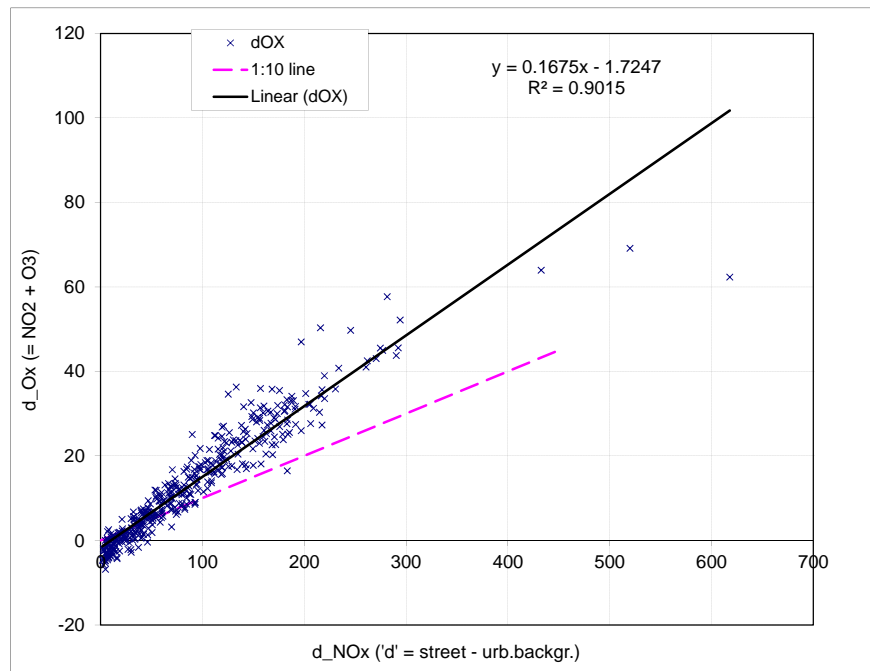


Figure 4: Scatter plot of hourly $\Delta[\text{O}_x]$ versus $\Delta[\text{NO}_x]$. The slope of the straight line fit is an estimate of the fraction directly emitted NO_2 .

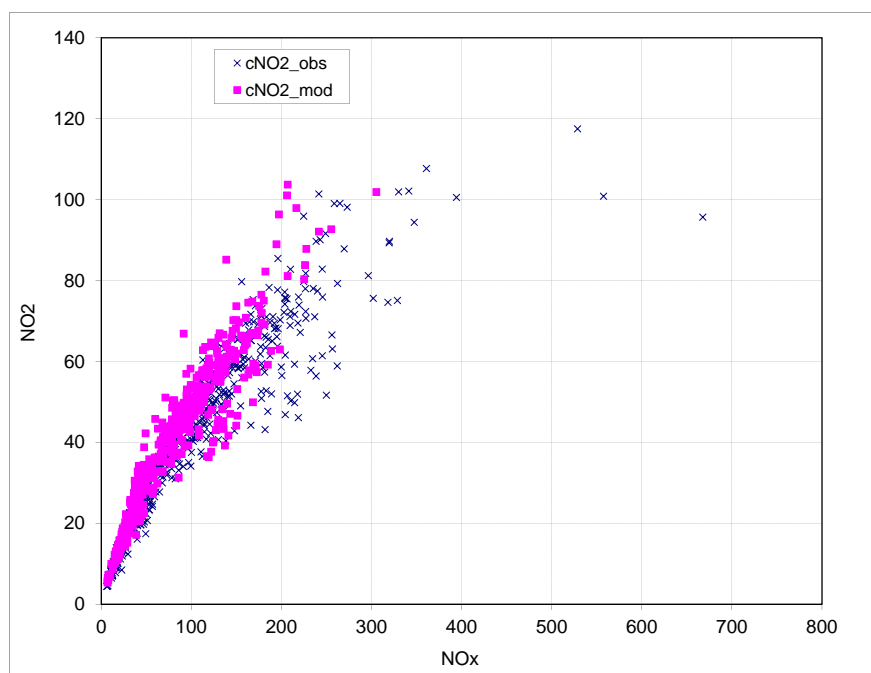


Figure 5: Scatter plot of the hourly NO_2 -concentration versus hourly NO_x concentration in respectively the model and the measurements. This plot is used to evaluate whether the model reproduces the correlation between NO_x concentrations and NO_2 concentrations from the measurements.

pink curve representing model results, and a green curve representing the development in the background concentrations. The model results and background results are also projected to the year 2015 and the year 2020 for assessment of future temporal development. Two figures are distinguishable from the rest on the worksheet *SummaryYear*:

ADT = Average
Daily Traffic
HD = Heavy
Duty

- The development in ADT and HD vehicles is shown as illustrated on fig. 6.
- Figure 7 shows the temporal development in modelled NO_x emissions plus the temporal development in the modelled and measured ratio between NO_2 emissions and NO_x emissions. This is used for assessing whether the model reproduces measured NO_2 emissions as a function of time plus the general trend in modelled NO_x emissions over time.

YearH_select, Settings
and other worksheets:

The worksheet *YearH_select* is a system worksheet and should **not** be modified by the user. The worksheet *Settings* is used when inserting new data in the worksheet as described in section 4. The rest of the worksheets are data-worksheets containing the sample data, which consist of in total two years of hourly data for model and measurements and 15 years of annual data.

3 Daily Use of the Validation Tool

The following section uses the sample data provided with the validation tool as an example of the everyday use of the validation tool.

Only the worksheets *Pivot_wdir*, *Pivot_hour*, and *Pivot_month* are used in the everyday exploratory data analysis. Each of these worksheets contain a number of graphs, some data, and a number of dropdown boxes as illustrated on fig. 8 for the worksheet *Pivot_wdir*. As can be seen, all the graphs show the concentration of a specific compound as a function of wind direction. Likewise with the other two

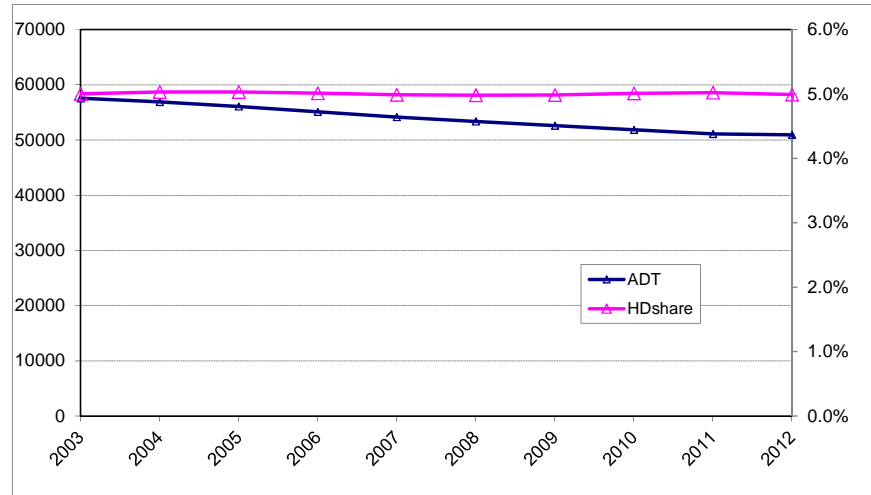


Figure 6: Annual trends in ADT and percentage of HD vehicles.

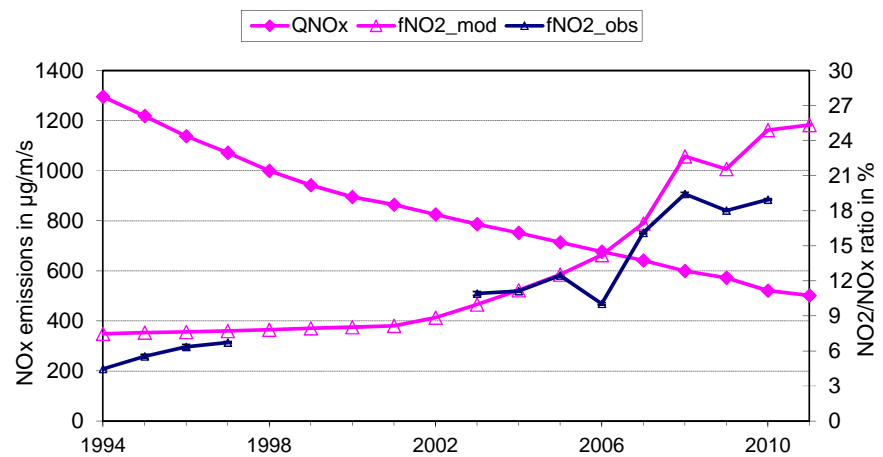


Figure 7: Temporal development in modelled NO_x emission density plotted on the left y-axis and the ratio of NO_2 to NO_x emissions for respectively the model and the measurements plotted on the right y-axis.

worksheets where the same species are plotted as a function of respectively the hour of the day and the month of the year. The compounds featured in all the worksheets are: NO_x , NO_2 , CO and O_3 . The worksheets *Pivot_hour* and *Pivot_month* furthermore contain one graph for PM_{10} and $\text{PM}_{2.5}$.

Choice of Year: The three worksheets operate on a one year dataset at a time. The year of interest is chosen in the dropdown box marked with the number "1" in fig. 8. This dropdown box is located differently on the worksheets *Pivot_hour* and *Pivot_month*, but the functionality is essentially the same – displaying the data from the respective year on the graphs. Choosing another year on one of the three worksheets also influences the scatterplots on *Graph_H*.

Choice of a subset of data: When the year is chosen, it is possible to choose to visualise the average values based on all the data for a full year or only to visualise a subset of the data. The latter is done using the dropdown boxes marked with the number "2" on fig. 8. All worksheets contain four dropdown boxes:

- **Month** Data from one or more months to be displayed in the graphs.
- **DayCase** Data from one or more types of day to be displayed in the graphs.
- **DayOfWeek** Data from one or more weekdays (numbered one to seven) to be displayed in the graphs.
- **U_GT_3** Data points where the wind speed is respectively larger or smaller than 3 m/s to be displayed in the graphs.
- **Hour** (Only on worksheet *Pivot_month*) Data from one or more hours a day to be visualised in the graphs.

The settings are linked between the different worksheets meaning that settings made on one worksheet are automatically transferred to the other worksheets. In this way, it is possible to quickly generate a series of standard graphs representing different subsets of large datasets. Moreover, it is possible to easily browse the dataset and thus examine various dependencies in the data.

The type of day (**DayCase**) is represented by a number between one and six: One being weekdays, two being Saturdays, and three being Sundays. The numbers four, five, and six are defined similarly, but only for the month of July, where traffic is lower due to holiday. This is similarly to the structure in the traffic input file to OSPM. For more info see Chapter 7 (*Traffic Data*) in the WinOSPM manual.

The only interaction the user has with these worksheets is through the dropdown boxes. DO NOT modify the content of the actual cells on the worksheet otherwise the workbook might break. This is also the case for the worksheet *SummaryYear*.



Figure 8: Screenshot of the worksheet *Pivot_wdir* in the OSPM Validation Tool

4 Inserting own data into the validation tool

To use the validation tool with the user's own data, the following procedure is used:

1. Generate a new set of model- and measurement data using the OSPM[®] model and save the output in an Excel-file. Make sure to name the hourly output "HourXXXX" and the yearly statistics "yearXXXX" with XXXX being the year in consideration. For info see User's Guide to WinOSPM[®] chapter 15 (*List of Input and Output Variables*).
 - (a) Enter the following string in the variable list of the hourly output file:
Date, hour, cNOX_r, cNOX_b, cNOX_mod_1, cNO2_r, cNO2_b, cNO2_mod_1, cO3_r, cO3_b, cO3_mod_1, cCO_r, cCO_b, cCO_mod_1, DailyTraffic, Nlight, Nheavy, cNOX_obs_1, cNO2_obs_1, cO3_obs_1, cCO_obs_1, QNOX, fractionNO2, cPM10_b, cPM10_mod_1, cPM2.5_b, cPM2.5_mod_1, cPMExh_mod_1, cPM10_obs_1, cPM2.5_obs_1, cNOX_str_mod_1, cNOX_str_obs_1, DayCase, DayOfWeek, Wind_dir, u_mast, cNOX_obs_2, cNO2_obs_2, cO3_obs_2, cCO_obs_2, cPM10_obs_2, cPM2.5_obs_2, Ntotal, Nheavy, Nlight, Speed_heavy, Speed_light, QNOX, QPMExh
 - (b) Enter the following string in the variable list columns for the file containing annual statistics:
"19xx", NOX_b, NOX_mod_1, NO2_b, NO2_mod_1, O3_b, O3_mod_1, CO_b, CO_mod_1, NOX_obs_1, NO2_obs_1, O3_obs_1, CO_obs_1, PM10_b, PM10_mod_1, PM2.5_b, PM2.5_mod_1, PMExh_mod_1, PM10_obs_1, PM2.5_obs_1
 - (c) Enter the following string in the variable list rows for the file containing statistics:
Average, MaxHourly, nHours, MaxDaily, 98_Perc
2. Copy the data sheets into the validation tool and delete the sample data sheets.
3. Activate the worksheet called *Settings*.
4. In cell C2 write "1" or "2" depending on which side of the street the monitoring station is located. This corresponds to the use of measurement station "1" or "2" in OSPM.
5. In cell A2 write the first year of the dataset. Continue with the next year in cell A3 and so on.