

Focal Sum Model User Guide

Version: focalsum_04082010.mxd

Date 4 August 2010

By

Kayoung Lee, Danielle Vienneau, Kees de Hoogh and David Briggs

Department of Epidemiology and Biostatistics
Imperial College London

Introduction

The Focal Sum Model is a GIS-based model designed to estimate air pollution exposures from source emissions. It is based on the principle that proximity to the source (e.g. roads or industry) is one of the main determinants of pollutant concentrations.

It uses the ArcGIS *focalsum* function with a weighted kernel to simulate the distance decay from sources. The weighted emissions are then calibrated with monitored air pollution data, using linear regression, to produce a concentration surface. An optional validation using a reserved set of independent monitoring is also included (Figure 1).

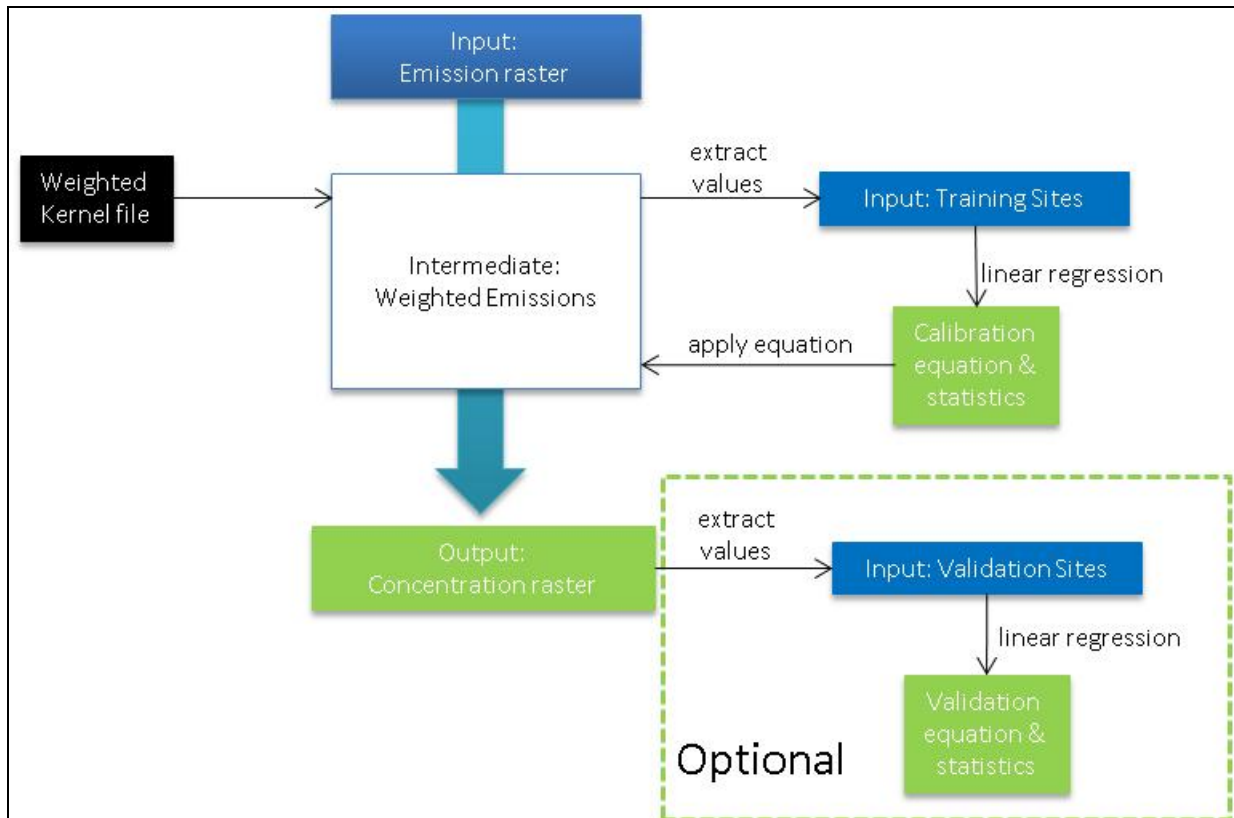


Figure 1. The modelling approach

Focal functions are applied to raster data, and compute an output raster where the cell value at each location (target cell) is some function of the input cells in a specified window (defined by the kernel file) around the location (Figure 2).

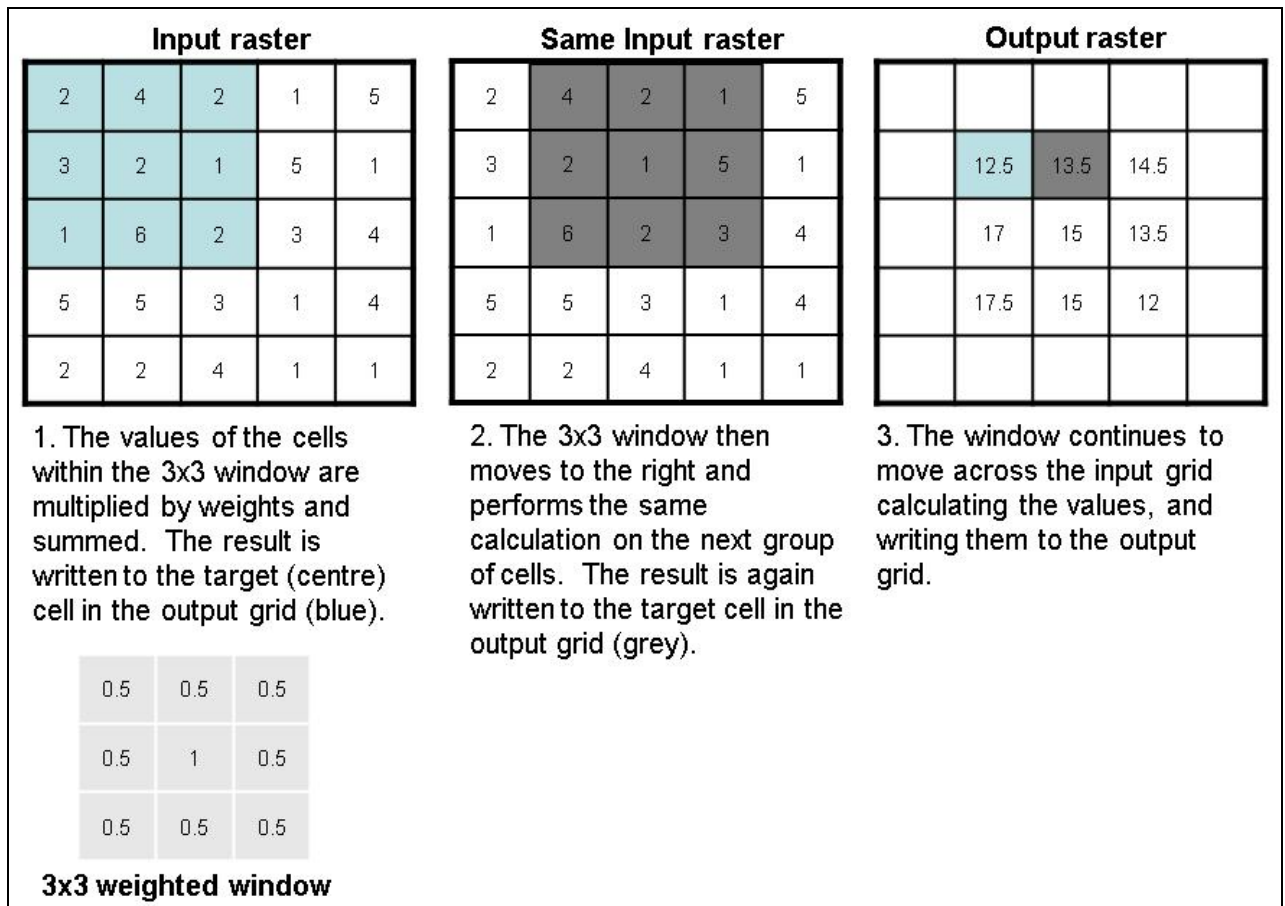


Figure 2. Focal sum example with 3x3 weighted kernel

In the Focal Sum Model, the window is moved across an input emissions raster to collect (i.e. sum) the source contributions and assign them to the target cell. With a weighted kernel one can assign greater contributions from nearby sources. Cells directly downwind of an industrial stack, for example, are expected to receive a greater source contribution than cells adjacent to but upwind of the cell in which the source is located.

In the Focal Sum Model the size, basic shape (square, circle) and weights ($1/d$, $1/d^2$ where d = distance in cells) for the kernel file can be specified by the user. Users may also wish to use existing kernel files either manually created or derived, for example, from dispersion models.

Software

The Focal Sum Model runs in ESRI's ArcGIS v9.3 ArcMap, and requires Spatial Analyst. The code is written in VBA and is embedded and saved in an ArcMap document.

To avoid overwriting the original, the user should make a copy of this document before running the model.

Data needs

Data format: ESRI's ArcGIS shape file and raster

Two main data inputs are required:

(1) Monitored air pollution at a sample of sites (point shape file).

One shape file is required for the training set. To include model validation, an optional second shape file containing an independent subset of sites is required. Both the training and validation shape files require a unique ID of 'long' integer type (i.e. in addition to the default FID). If it does not exist, the tool will prompt the user to create a new unique ID field.

(2) Emissions, or a proxy for them, to provide measures of source intensity (raster).

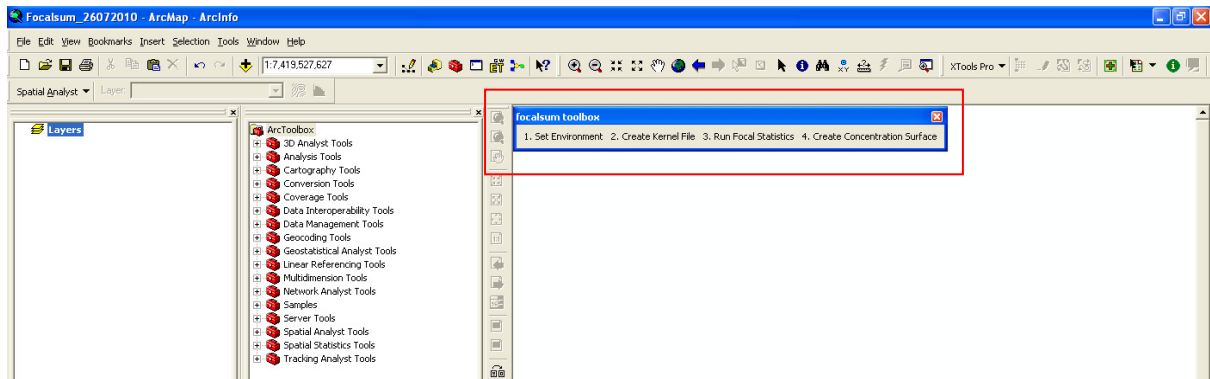
Test dataset has been provided by Daniela Fecht who developed urban Simulation for ENvironment health Analysis (SIENA). This includes:

- 250m grids for
 - PM₁₀ and NO_x (t/km²/year)
 - Emissions from main roads, minor roads and total road emissions
- Monitoring network shape files
 - PM₁₀ and NO₂
 - Representing two-week average concentrations measured in summer
 - Routine monitoring network of 10 sites
 - Study area network of 45 sites

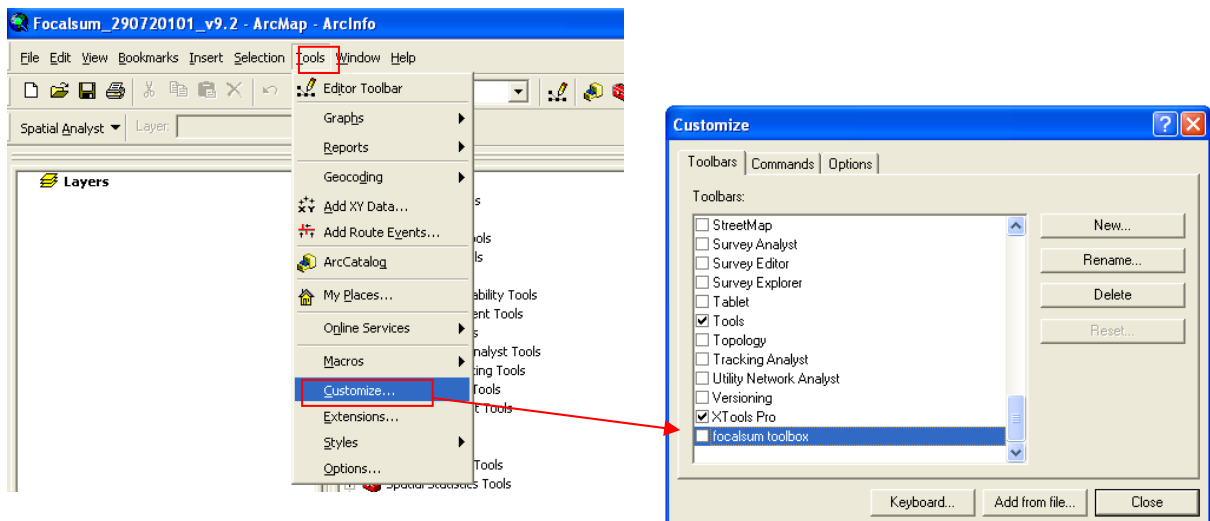
When saving output files (shape files, tables or rasters) please specify file names not already used. Suggested output names are provided, but the user should customise if running the model more than once.

Running the model

When opened the user will see the following toolbar:



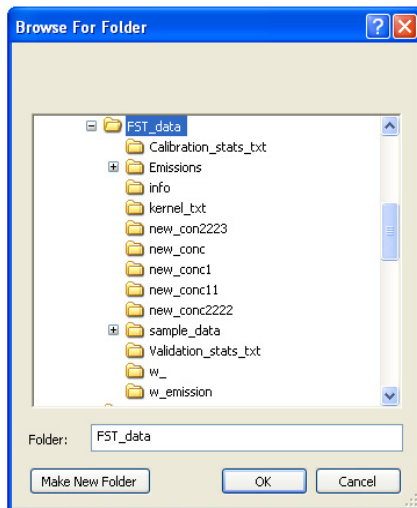
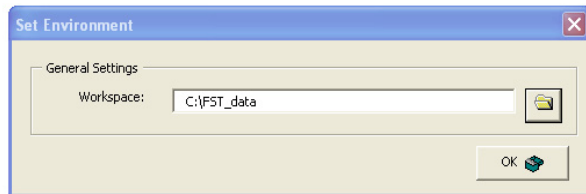
If the 'focalsum toolbox' toolbar does not appear: go to Tools – Customize – scroll to the bottom and click 'focalsum toolbox'. The Model tool will then be displayed onto the ArcMap window.



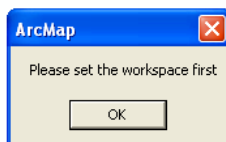
The steps to run this tool are outlined below:

1. Set environment

The set environment window prompts the user to define the folder in which all the output data from the model run will be stored.

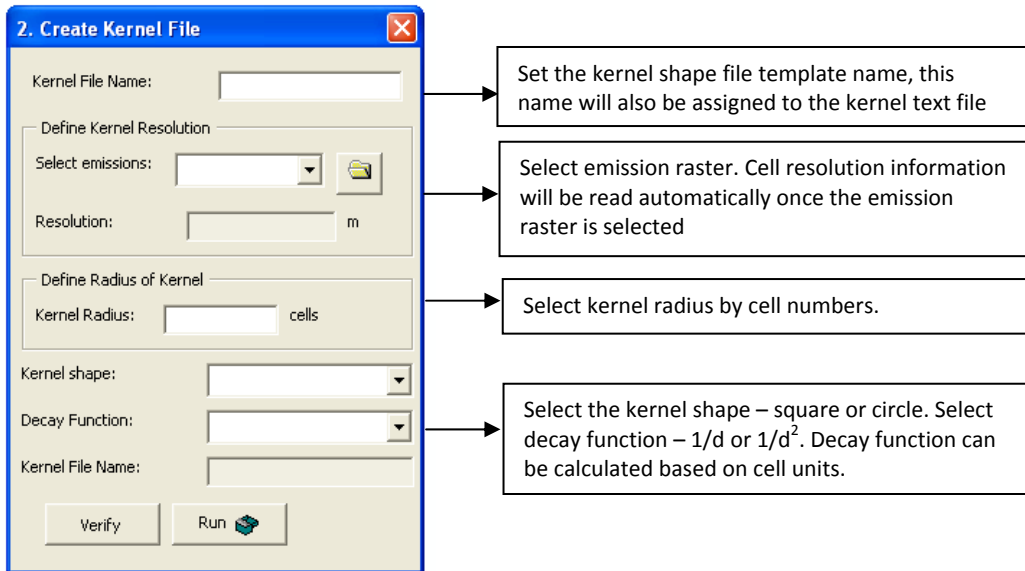


If this step is skipped the following message will appear to prompt the user to set the workspace:

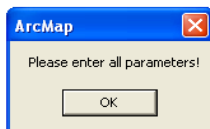


2. Create Kernel File

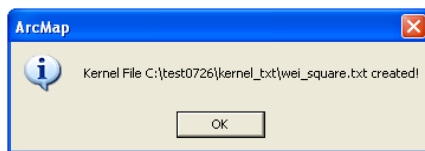
In this step the user specifies the parameters for creating the weighted kernel file. The tool creates a kernel with an odd number of rows and columns (e.g. 3x3 or 7x7 kernel) so that a true middle is defined. Weights are then computed using linear distance in cell units from the centroid of the 'middle kernel cell' to all other 'kernel cell' centroids.

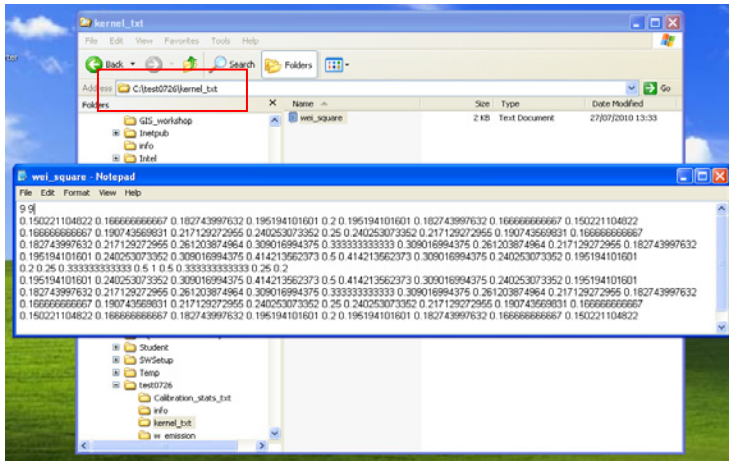


If the user does not select all parameters, the following message will be shown. Once the missing parameters are entered the model can be run.



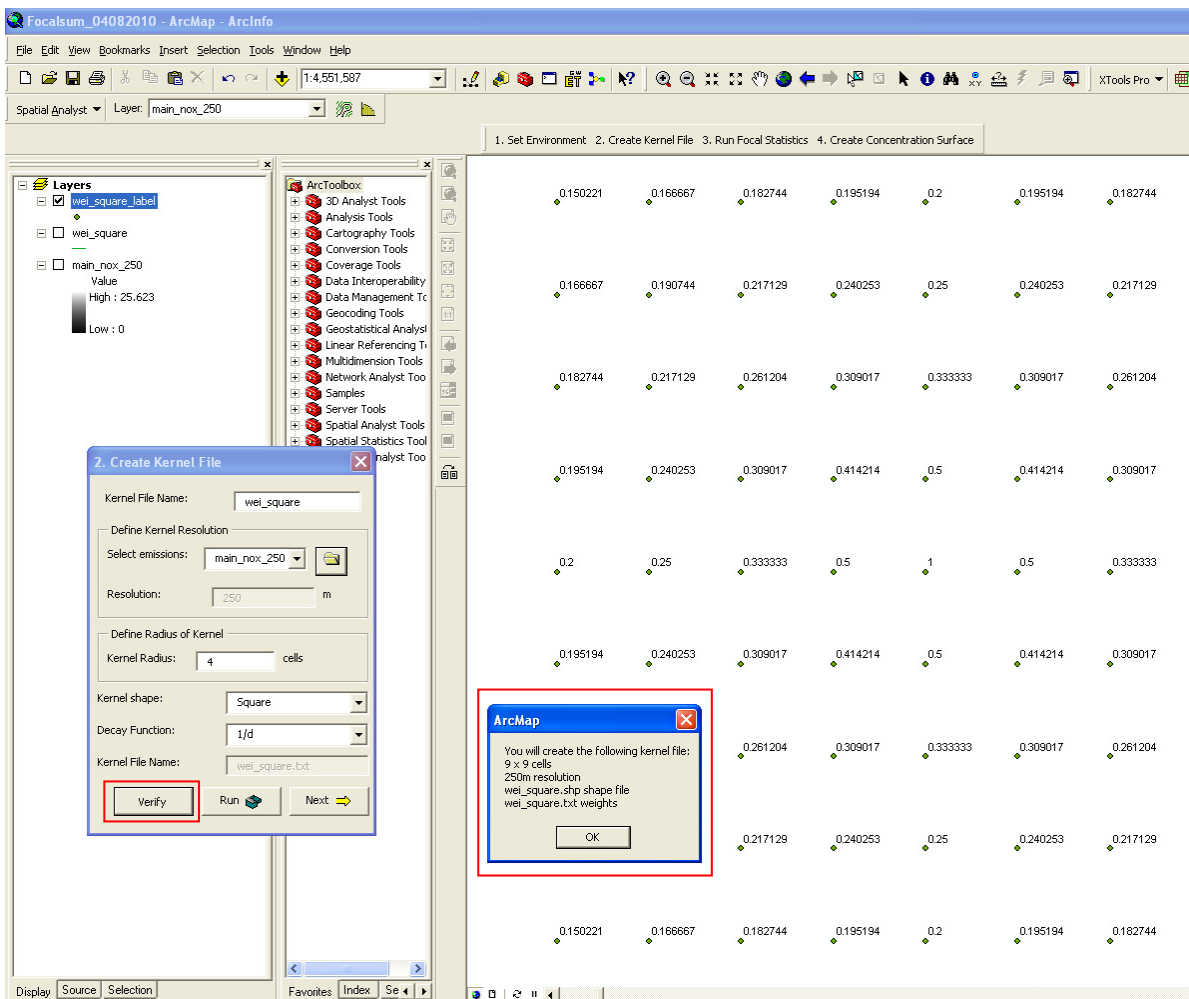
When Step 2 finishes running, a message box (shown below) tells the user that the kernel file is stored in: '*workspace which is set from step 1*\kernel_txt*kernel file name*'. This kernel file will be used as an input in step 3.






When the user selects the 'Verify' button, the summary information for the parameters used to create the kernel file will be shown in a message box.

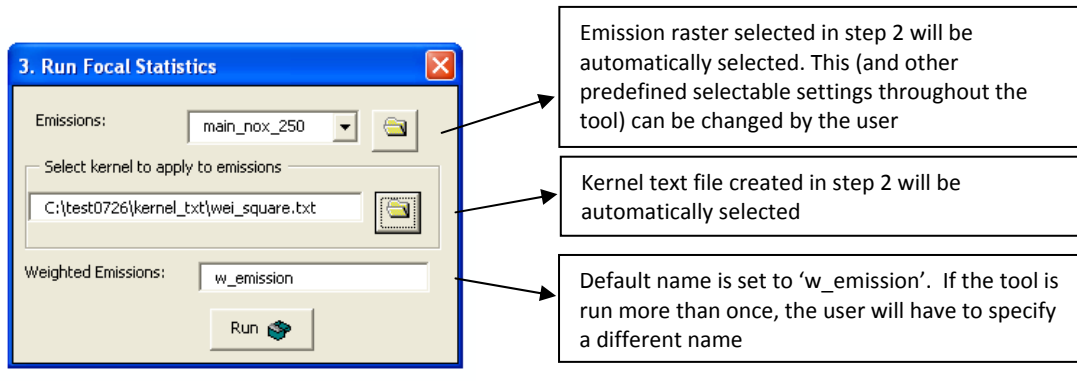
After clicking 'Run', if a kernel radius of less than 10 cells was defined, the calculated weights for the distance decay function will automatically be visualised in the map view. Otherwise, just the points will display, but the user can add labels to visualise.



Once step 2 finishes,  button will be shown.

3. Run Focal Statistics

In this step, the kernel file is used with a *focalsum* function to weigh and sum the emissions to create a weighted emission raster. This process is illustrated in Figure 2 above, and an example raster input and output are shown below.

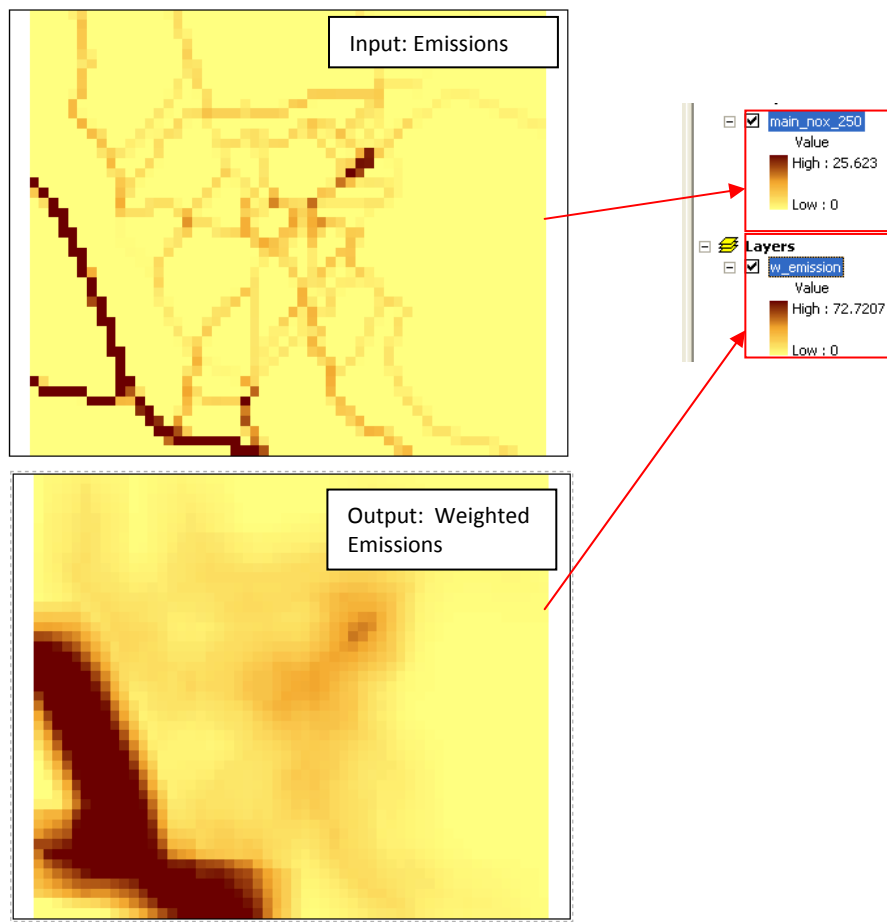


The screenshot shows the '3. Run Focal Statistics' dialog box with the following settings:

- Emissions:** main_nox_250
- Select kernel to apply to emissions:** C:\test0726\kernel_txt\wei_square.txt
- Weighted Emissions:** w_emission

Callout boxes provide additional information:

- Arrow 1: Emission raster selected in step 2 will be automatically selected. This (and other predefined selectable settings throughout the tool) can be changed by the user.
- Arrow 2: Kernel text file created in step 2 will be automatically selected.
- Arrow 3: Default name is set to 'w_emission'. If the tool is run more than once, the user will have to specify a different name.



4. Create Concentration Surface

In these steps, the weighted emissions are calibrated with monitored air pollution data, using linear regression. In step 4.1 the weighted emission value for each monitoring site is extracted and added to the calibration shape file. The parameters used to develop the regression model are defined in step 4.2. If a suitable calibration model is obtained, the user can apply the model to the weighted emission raster in the final step to produce a concentration surface.

4.1. Extract weighted emissions to training sites

Select training sites – input shape file should have 'Long Integer' field type. If not, tool will prompt user to create a new field called 'uniqueID'

Weighted emission raster from step 3 will be automatically added

Define calibration shape file name. Tool gives 'calib_' as a default name

If the user selects an input file without a Long integer unique id field, a warning message will appear. If the user clicks 'Yes', the uniqueID field will be created automatically. If the user clicks 'No' then the tool will terminate.

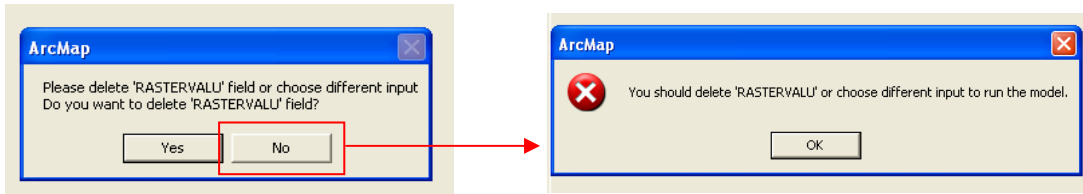
Please create a 'LONG' Integer Unique ID in your input point file
If not, you cannot run this function!
Do you want to create 'uniqueID' field?

Yes No

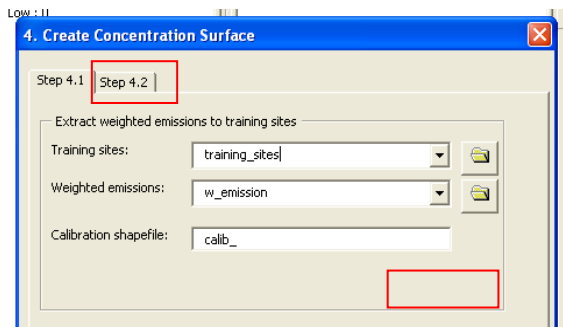
You should create a 'LONG' Integer unique ID to run the model.

OK

The 'Extract Values to Points' tool is used in this step. If this has previously been run with the selected training sites the shape file may already contain a 'RASTERVALU' field. The tool will warn the user if a 'RASTERVALU' field already exists and ask if it should be deleted. If the user does not want to delete the existing 'RASTERVALU' field, the program will terminate.



Once run by the user, the 'Run' button on 4.1 will be hidden and 4.2 tab will appear so that the user is directed to the next step. This functionality also applies to other steps with tabs.



4.2. Run simple regression to calibrate model

Shape file of monitoring sites containing both measurement data (dependent) and weighted emissions (explanatory) variables for analysis

Long integer Unique ID field

Numeric field containing measured concentrations

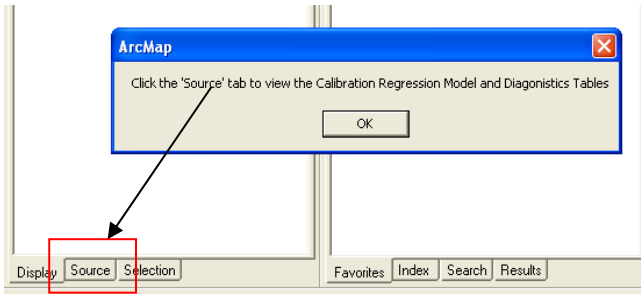
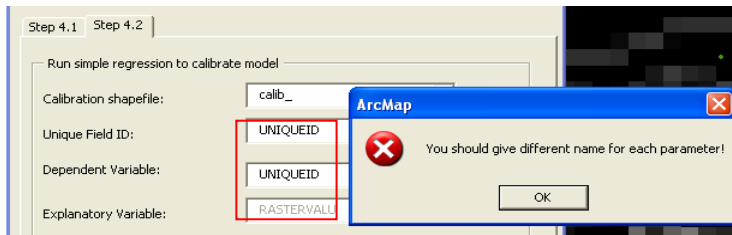
Explanatory variable is always 'RASTERVALU' representing weighted emissions(output from step 4.1)

Name of output shape file containing estimated value and residuals

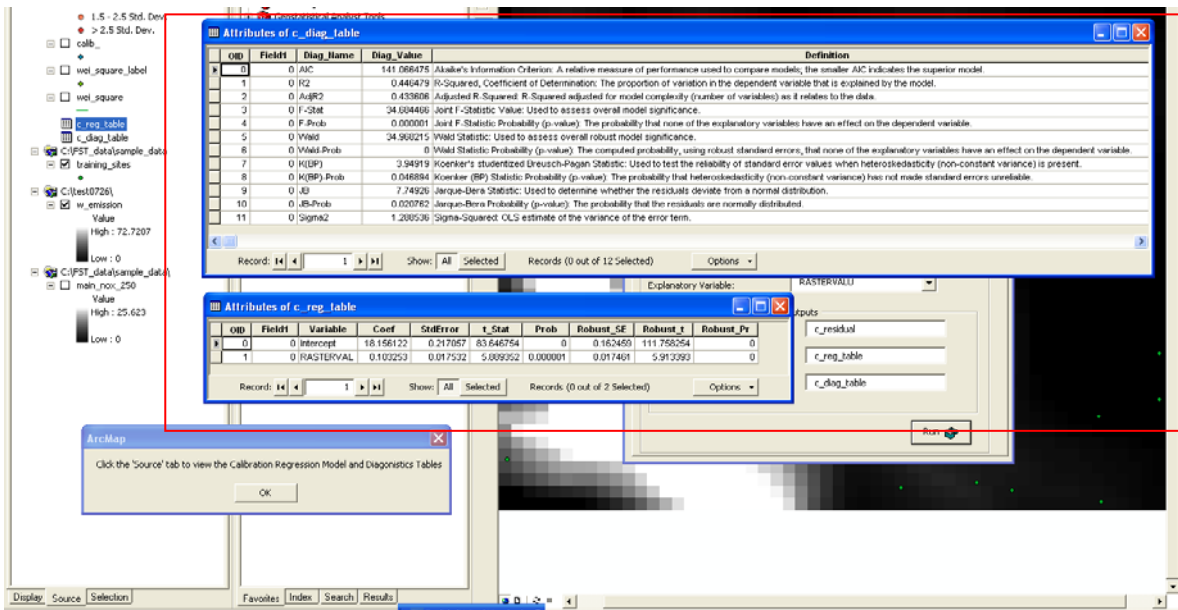
Name of output table containing model coefficients and standard errors

Name of output table summarising model building statistics, including coefficient of determination (R^2)

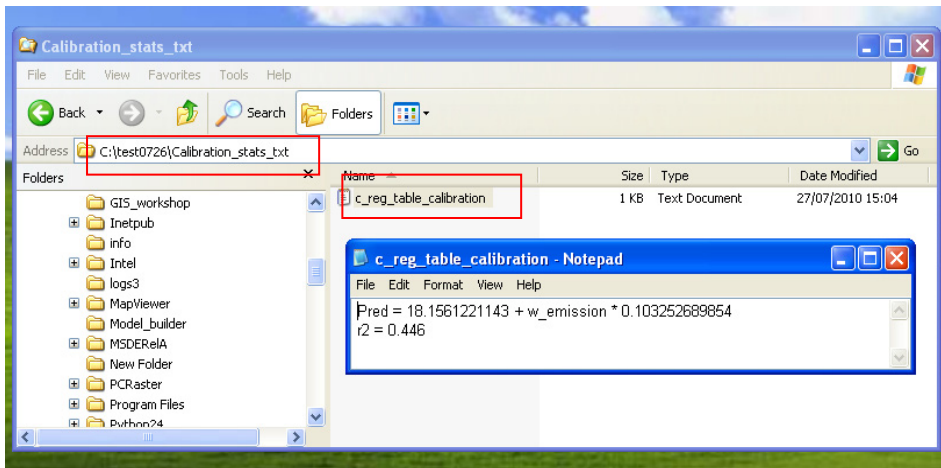
Please note that when setting parameters in Step 4.2, different inputs are required for each of Unique Field ID, Dependent variable and Explanatory variable. Otherwise the tool will show an error message and terminate.



After simple regression is run, a message appears indicating that the detailed regression output (e.g. R^2 , regression coefficients, etc) can be viewed by going to the 'source' tab as shown below.

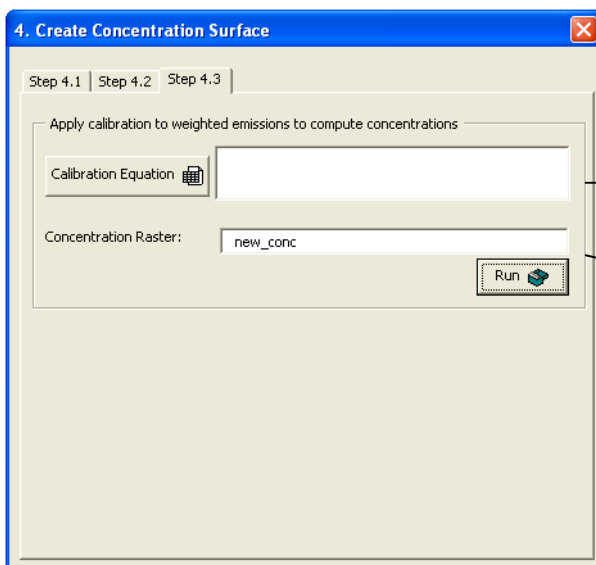


The regression equation is also written to a text file found under: 'workspace which is set from step 1\Calibration_stats_txt\Regression Model Table name_calibration.txt'



This equation is used to calculate new concentration surface in step 4.3.

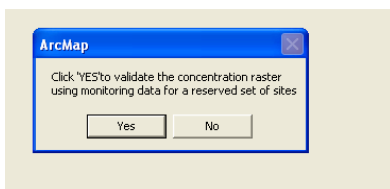
4.3. Apply calibration to weighted emissions to compute concentrations



Click 'Calibration equation' to automatically load the equation derived in step 4.2. This will be applied to the weighted emissions to compute an output concentration raster

Define a name for the new output concentration raster. The default name is set as 'new_conc'

After Step 4.3 is run and the new concentration raster created, a message will appear asking whether the user will proceed the next step for validation.

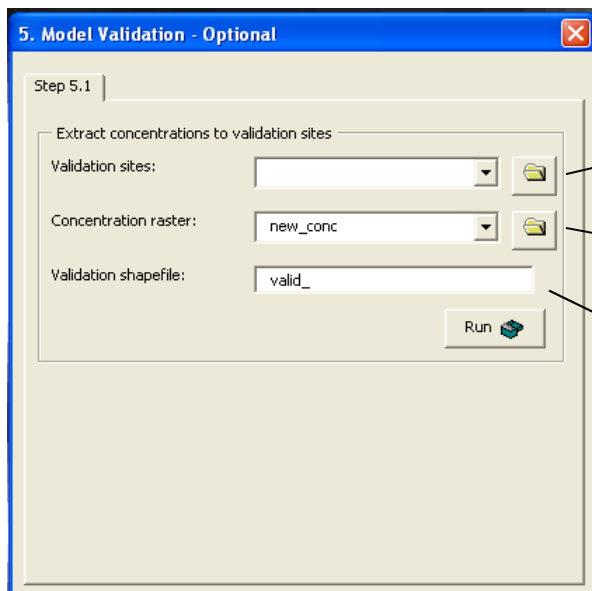


If the user clicks 'Yes' the optional Model Validation window will open. Step 5.1 and 5.2 use the same process as in step 4.1 and 4.2, respectively, though should be applied to the shape file containing the monitoring sites reserved for validation. If the user clicks 'No' then the tool will exit. The new concentration raster and regression outputs are automatically saved to the set workspace.

5. Model Validation – Optional

In these steps, the new concentration raster is validated with independent monitored air pollution data, also using linear regression. The final output includes regression statistics indicating model performance.

5.1. Extract concentrations to validation sites



Select validation sites – input shape file should have 'Long Integer' field type. If not, tool will prompt user to create a new field called 'uniqueID'

New concentration raster from step 4.3 will be automatically added

Define validation shape file name. Tool gives 'valid_' as a default name

As previous, the 'Run' button on 5.1 will be hidden and the 5.2 tab will appear so that the user is directed to the next step.

5.2. Extract concentrations to validation sites

5. Model Validation - Optional

Step 5.1 | Step 5.2

Run Simple Regression to validate model

Validation shapefile: valid_

Unique Field ID: ID|

Dependent Variable: NO2

Explanatory Variable: RASTERVALU

Enter names for the validation outputs

Residuals Shapefile: v_residual

Regression Model Table: v_reg_table

Diagnostics table: v_diag_table

Run

Shape file of monitoring sites containing both measurement data (dependent) and weighted emissions (explanatory) variables for analysis

Long integer Unique ID field

Numeric field containing measured concentrations

Explanatory variable is always 'RASTERVALU' representing modelled concentrations (output from step 5.1)

Name of output shape file containing estimated value and residuals

Name of output table containing model coefficients and standard errors

Name of output table summarising model validation statistics, including coefficient of determination (R^2)

ArcMap

Validation Statistics: $r^2 = 0.51$
Click source tab to see detailed table information
Also see, 'C:\test0726\validation_stats_txt\v_reg_table_validation.txt'

OK

As per step 4.2, when successfully run a message appears showing the validation R^2 and indicating that the detailed regression output (e.g. R^2 , regression coefficients, etc) can be viewed by going to the 'source' tab.