GEOGRAPHICALLY WEIGHTED REGRESSION

Lecture 3

GWR 3.0 Software
Geographically Weighted Regression

stewart.fotheringham@nuim.ie
http://ncg.nuim.ie/GWR/
Outline

- Running the GWR software
- Understanding the GWR listing
Part I
Running a model in the GWR software
First, create a data file...

- File type xxxxx.dat (xxxx.csv)
- First line is a comma separated list of variable names (<= 8 characters)
- Data lines have **numeric items only** terminated by a carriage return
- One line of data per location
- Space or comma delimited (easily created in Excel)
Example of a data file  Georgia.dat

The first 10 lines of the file...

<table>
<thead>
<tr>
<th>ID</th>
<th>Lat</th>
<th>Lon</th>
<th>TotPop90</th>
<th>PctRural</th>
<th>PctBach</th>
<th>PctEld</th>
<th>PctFB</th>
<th>PctPov</th>
<th>PctBlack</th>
</tr>
</thead>
<tbody>
<tr>
<td>13001</td>
<td>31.75339</td>
<td>-82.28558</td>
<td>15744</td>
<td>75.6</td>
<td>8.2</td>
<td>11.43</td>
<td>0.635</td>
<td>19.9</td>
<td>20.76</td>
</tr>
<tr>
<td>13003</td>
<td>31.29486</td>
<td>-82.87474</td>
<td>6213</td>
<td>100.0</td>
<td>6.4</td>
<td>11.77</td>
<td>1.577</td>
<td>26.0</td>
<td>26.86</td>
</tr>
<tr>
<td>13005</td>
<td>31.55678</td>
<td>-82.45115</td>
<td>9566</td>
<td>61.7</td>
<td>6.6</td>
<td>11.11</td>
<td>0.272</td>
<td>24.1</td>
<td>15.42</td>
</tr>
<tr>
<td>13007</td>
<td>31.33084</td>
<td>-84.45401</td>
<td>3615</td>
<td>100.0</td>
<td>9.4</td>
<td>13.17</td>
<td>0.111</td>
<td>24.8</td>
<td>51.67</td>
</tr>
<tr>
<td>13009</td>
<td>33.07193</td>
<td>-83.25085</td>
<td>39530</td>
<td>42.7</td>
<td>13.3</td>
<td>8.64</td>
<td>1.432</td>
<td>17.5</td>
<td>42.39</td>
</tr>
<tr>
<td>13011</td>
<td>34.35269</td>
<td>-83.50053</td>
<td>10308</td>
<td>100.0</td>
<td>6.4</td>
<td>11.37</td>
<td>0.340</td>
<td>15.1</td>
<td>3.49</td>
</tr>
<tr>
<td>13013</td>
<td>33.99347</td>
<td>-83.71181</td>
<td>29721</td>
<td>64.6</td>
<td>9.2</td>
<td>10.63</td>
<td>0.922</td>
<td>14.7</td>
<td>11.44</td>
</tr>
<tr>
<td>13015</td>
<td>34.23840</td>
<td>-84.83918</td>
<td>55911</td>
<td>75.2</td>
<td>9.0</td>
<td>9.66</td>
<td>0.816</td>
<td>10.7</td>
<td>9.21</td>
</tr>
<tr>
<td>13017</td>
<td>31.75939</td>
<td>-83.21975</td>
<td>16245</td>
<td>47.0</td>
<td>7.6</td>
<td>12.81</td>
<td>0.332</td>
<td>22.0</td>
<td>31.33</td>
</tr>
</tbody>
</table>
Starting GWR

• The software is usually stored in the C:\GWR3 folder and the program is called GWR30.exe

  - Start/Programs/Geographically Weighted Regression
  - Desktop icon
  - Explorer
This brings up the GWR Wizard

• You have a number of options to choose from in creating and running a GWR model

• The job of the Wizard is to provide suitable guidance in making the right choices
If you want to create and run a new GWR model, click on the option ‘Create a new model’.

If you created and saved a GWR model in a previous session and you want to access this, click on ‘Open an existing model using the GWR model editor’.
Inputting data – click and drag .dat or .csv file from appropriate folder
Regression Points

- Do you want to run GWR at the data point locations or some other set of locations?
Name the output file...

either click and drag an existing file name or give a new file name.

Three types of output file are possible:

- .e00 ArcInfo export file
- .mif MapInfo file <---- there is a problem with this
- .csv comma separated variable file

For the GMAC workshop use .csv

Remember to choose a name that is different from the data file name.
The Model Editor

To specify a dependent variable, highlight it in the list on the left and click on the [->] symbol.
The Model Editor

To specify independent variables, highlight them individually in the list on the left and click on the [- ->] symbol.
The Model Editor

To specify location variables, highlight them individually in the list on the left and click on the corresponding [- ->] symbols.
The Model Editor

Next you specify the type of kernel: this can be **fixed** (Gaussian) or **adaptive** (bisquare)
The Model Editor

You can either preset the bandwidth in the units that the location variables are measured in (for example, *metres*)
Or if you want the program to determine the optimal bandwidth, specify *either* crossvalidation or AIC minimisation.

For large files, there is a sampling option to speed the process.
The Model Editor

Select the type of coordinate system you are using for your location variables – choice is either Cartesian or spherical.
The type of output in the printed listing can also be controlled by clicking on Model Options.
# Bandwidth selection

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.043532255000</td>
<td>913.159190588348</td>
</tr>
<tr>
<td>84.50000000000000</td>
<td>885.119969660068</td>
</tr>
<tr>
<td>112.956467745000</td>
<td>872.910381423844</td>
</tr>
<tr>
<td>130.543532046749</td>
<td>868.887720190066</td>
</tr>
<tr>
<td>141.412935569545</td>
<td>869.149708997055</td>
</tr>
<tr>
<td>123.825871267796</td>
<td>870.450868861077</td>
</tr>
<tr>
<td>134.695274741431</td>
<td>869.114420384913</td>
</tr>
<tr>
<td>127.977613962479</td>
<td>869.551269557617</td>
</tr>
</tbody>
</table>

** Convergence after 8 function calls
** Convergence: Local Sample Size= 131

Useful if you want to plot the relationship to see how steep or flat it is
If you have requested an output file, this information and the diagnostics are also written to this file.

<table>
<thead>
<tr>
<th>Obs</th>
<th>Y(i)</th>
<th>Yhat(i)</th>
<th>Res(i)</th>
<th>X(i)</th>
<th>Y(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.200</td>
<td>9.006</td>
<td>-0.806</td>
<td>-82.286</td>
<td>31.753</td>
</tr>
<tr>
<td>2</td>
<td>6.400</td>
<td>6.958</td>
<td>-0.558</td>
<td>-82.875</td>
<td>31.295</td>
</tr>
<tr>
<td>3</td>
<td>6.600</td>
<td>8.524</td>
<td>-1.924</td>
<td>-82.451</td>
<td>31.557</td>
</tr>
<tr>
<td>4</td>
<td>9.400</td>
<td>8.308</td>
<td>1.092</td>
<td>-84.454</td>
<td>31.331</td>
</tr>
<tr>
<td>5</td>
<td>13.300</td>
<td>13.835</td>
<td>-0.535</td>
<td>-83.251</td>
<td>33.072</td>
</tr>
<tr>
<td>6</td>
<td>6.400</td>
<td>8.910</td>
<td>-2.510</td>
<td>-83.501</td>
<td>34.353</td>
</tr>
<tr>
<td>7</td>
<td>9.200</td>
<td>11.760</td>
<td>-2.560</td>
<td>-83.712</td>
<td>33.993</td>
</tr>
<tr>
<td>8</td>
<td>9.000</td>
<td>11.446</td>
<td>-2.446</td>
<td>-84.839</td>
<td>34.238</td>
</tr>
<tr>
<td>9</td>
<td>7.600</td>
<td>10.231</td>
<td>-2.631</td>
<td>-83.220</td>
<td>31.759</td>
</tr>
<tr>
<td>10</td>
<td>7.500</td>
<td>9.104</td>
<td>-1.604</td>
<td>-83.232</td>
<td>31.274</td>
</tr>
</tbody>
</table>
**List Pointwise Diagnostics**

<table>
<thead>
<tr>
<th>Obs</th>
<th>Observed</th>
<th>Predicted</th>
<th>Residual</th>
<th>Std Resid</th>
<th>R-Square</th>
<th>Influence</th>
<th>Cook's D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.20000</td>
<td>8.84819</td>
<td>-0.64819</td>
<td>-0.182251</td>
<td>0.784156</td>
<td>0.032346</td>
<td>0.000073</td>
</tr>
<tr>
<td>2</td>
<td>6.40000</td>
<td>6.39738</td>
<td>0.00262</td>
<td>0.000759</td>
<td>0.775286</td>
<td>0.085459</td>
<td>0.000000</td>
</tr>
<tr>
<td>3</td>
<td>6.60000</td>
<td>8.48954</td>
<td>-1.88954</td>
<td>-0.549871</td>
<td>0.782090</td>
<td>0.096687</td>
<td>0.002126</td>
</tr>
<tr>
<td>4</td>
<td>9.40000</td>
<td>8.35258</td>
<td>1.04742</td>
<td>0.302776</td>
<td>0.808351</td>
<td>0.084519</td>
<td>0.000556</td>
</tr>
<tr>
<td>5</td>
<td>13.30000</td>
<td>14.60358</td>
<td>-1.30358</td>
<td>-0.377493</td>
<td>0.834522</td>
<td>0.087768</td>
<td>0.000901</td>
</tr>
<tr>
<td>6</td>
<td>6.40000</td>
<td>8.29036</td>
<td>-1.89036</td>
<td>-0.538081</td>
<td>0.839070</td>
<td>0.055846</td>
<td>0.001125</td>
</tr>
<tr>
<td>7</td>
<td>9.20000</td>
<td>12.02529</td>
<td>-2.82529</td>
<td>-0.794934</td>
<td>0.841344</td>
<td>0.033697</td>
<td>0.001448</td>
</tr>
<tr>
<td>8</td>
<td>9.00000</td>
<td>10.97210</td>
<td>-1.97210</td>
<td>-0.554246</td>
<td>0.846250</td>
<td>0.031492</td>
<td>0.000656</td>
</tr>
<tr>
<td>9</td>
<td>7.60000</td>
<td>10.73917</td>
<td>-3.13917</td>
<td>-0.892715</td>
<td>0.778960</td>
<td>0.054083</td>
<td>0.002994</td>
</tr>
<tr>
<td>10</td>
<td>7.50000</td>
<td>9.04908</td>
<td>-1.54908</td>
<td>-0.444086</td>
<td>0.778295</td>
<td>0.069182</td>
<td>0.000963</td>
</tr>
</tbody>
</table>
Once the model specification is completed, give the file a title and save it before you run it. The saved file can be used in later sessions.
- the listing of the output will be saved here

and then click on ‘Run’
Run Completed

View Listing File?

Yes
No
Geographically Weighted Regression
Release 3.0.1
Dated: 06-vii-2003
Martin Charlton, Chris Brunsdon
Stewart Fotheringham
(c) University of Newcastle upon Tyne

** Program limits:
** Maximum number of variables..... 35
** Maximum number of observations.. 12000
** Maximum number of fit locations. 12000

Georgia
** Observed data file: C:\GWR3\GeorgiaData.dat
** Prediction location file: Estimation at sample point locations
** Result output file: C:\GWR3\georgia.e00

** Variables in the data file...
ID    Latitude Longitud TotPop90 PctRural PctBach PctEld PctFB
PctPov  PctBlack

** Dependent (y) variable.........PctBach
** Easting (x-coord) variable.....Longitud
** Northing (y-coord) variable.....Latitude
** No weight variable specified
** Independent variables in your model...
PctRural PctEld PctFB PctPov PctBlack
• The local parameter estimates will be saved in your named output file e.g. `georgia.csv`

This can be used for subsequent mapping in MapInfo...

Before we do any mapping we need to check the listing file
Part II
Understanding the Listing File
The Listing File

• An important part of the GWR output is the listing file.
• This is an ASCII file which you view in the GWR software
• There are a number of helpful diagnostics to guide you
• Let’s look at its organisation and content
Viewing the output

• Click on **Yes** to view the file in a scrollable window
Part 1: The Header

**************************************************************
* Geographically Weighted Regression                      *
* Release 3.0.1                                           *
* Dated: 06-vii-2003                                       *
*                                                        *
* Martin Charlton, Chris Brunsdon                         *
* Stewart Fotheringham                                     *
* (c) National University of Ireland                      *
**************************************************************

** Program limits:
** Maximum number of variables..... 52
** Maximum number of observations.. 80000
** Maximum number of fit locations. 80000

Check the program limits and your problem limits before mailing us for help!!
Part 2: File information

Educational Attainment Model 1
** Observed data file:        C:\GWR3\SampleData\GeorgiaData.dat
** Prediction location file:  Estimation at sample point locations
** Result output file:        C:\Users\Martin\GISRUK\gmodel1.csv

** Variables in the data file...
ID       Latitude Longitud TotPop90 PctRural PctBach PctEld   PctFB   PctPov   PctBlack

We have here a list of the files to be used, and the list of variables from the first line of the data file
Part 3: Model options

** Dependent (y) variable..........PctBach
** Easting (x-coord) variable.....Longitud
** Northing (y-coord) variable.....Latitude
** No weight variable specified
** Independent variables in your model...
  PctEld   PctFB    PctPov
** Kernel type: Adaptive
** Kernel shape: Bi-Square
** Bandwidth selection by AICc minimisation
** Use all regression points
** Calibration history requested
** Prediction report requested
** Output estimates to be written to .txt file
** No significance test for spatial variation
** Casewise diagnostics requested

These are the choices you made in the Model Editor. Check to see that these are correct
Part 4: Model Header

*** Analysis method ***
*** Geographically weighted multiple regression
** Cartesian coordinates: Euclidean Distance
******************************************************************************
*
* GEOGRAPHICALLY WEIGHTED GAUSSIAN REGRESSION *
*
******************************************************************************
Number of data cases read: 159
Observation points read...

Dependent mean= 10.9471693
Number of observations, nobs= 159
Number of predictors, nvar= 3
Observation Easting extent: 4.41947222
Observation Northing extent: 4.20193577

Here we’ve chosen a Gaussian regression. GWR3 prints some summary information concerning the data
Part 5: Calibration

*Finding bandwidth...
... using all regression points
This can take some time...
*Calibration will be based on 159 cases
*Adaptive kernel sample size limits: 10 159
*AICc minimisation begins...

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.043532255000</td>
<td>910.503643223293</td>
</tr>
<tr>
<td>84.500000000000</td>
<td>894.591466346276</td>
</tr>
<tr>
<td>112.956467745000</td>
<td>891.242236922868</td>
</tr>
<tr>
<td>130.543532046749</td>
<td>892.205443937283</td>
</tr>
<tr>
<td>102.087064301749</td>
<td>891.191941160401</td>
</tr>
<tr>
<td>95.369403522796</td>
<td>892.011818377660</td>
</tr>
<tr>
<td>106.238806966048</td>
<td>891.303511556728</td>
</tr>
<tr>
<td>99.521146217479</td>
<td>891.260634243019</td>
</tr>
<tr>
<td>103.67288881778</td>
<td>891.193359720619</td>
</tr>
</tbody>
</table>

** Convergence after 9 function calls
** Convergence: Local Sample Size= 102

The results of the calibration step are printed next: again check these, and if necessary, graph the function
Minimisation

![Graph showing Minimisation and Calibration with AICc and Bandwidth (objects)]
Part 6: Global Model Estimates

**********************************************************
*           GLOBAL REGRESSION PARAMETERS                 *
**********************************************************
Diagnostic information...
Residual sum of squares........ 2602.920095
Effective number of parameters.. 4.000000
Sigma........................... 4.097930
Akaike Information Criterion.... 906.096726
Coefficient of Determination.... 0.492418

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Err</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.332323298897</td>
<td>1.666228106775</td>
<td>7.401341438293</td>
</tr>
<tr>
<td>PctEld</td>
<td>-0.163871467223</td>
<td>0.139054993298</td>
<td>-1.178465127945</td>
</tr>
<tr>
<td>PctFB</td>
<td>2.653034523214</td>
<td>0.301957476064</td>
<td>8.786119461060</td>
</tr>
<tr>
<td>PctPov</td>
<td>-0.127244149993</td>
<td>0.054776894921</td>
<td>-2.322952985764</td>
</tr>
</tbody>
</table>

These are the parameter estimates and diagnostics for a global model – they’re similar to the output from a program like SPSS or R.
Part 7: Local Model Summary

**********************************************************
*                GWR ESTIMATION                          *
**********************************************************
Fitting Geographically Weighted Regression Model...
Number of observations............ 159
Number of independent variables... 4
  (Intercept is variable 1)
Number of nearest neighbours...... 102
Number of locations to fit model.. 159

Diagnostic information...
Residual sum of squares............ 1854.384815
Effective number of parameters.. 13.997842
Sigma........................... 3.576125
Akaike Information Criterion.... 875.142043
Coefficient of Determination.... 0.638386
** Results written to .e00 file

This section provides a summary of the local model fit
### Part 8: Casewise Diagnostic Information

<table>
<thead>
<tr>
<th>Obs</th>
<th>Observed</th>
<th>Predicted</th>
<th>Residual</th>
<th>Std Resid</th>
<th>R-Square</th>
<th>Influence</th>
<th>Cook's D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.20000</td>
<td>9.38396</td>
<td>-1.18396</td>
<td>-0.251849</td>
<td>0.599310</td>
<td>0.034376</td>
<td>0.000161</td>
</tr>
<tr>
<td>2</td>
<td>6.40000</td>
<td>9.65340</td>
<td>-3.25340</td>
<td>-0.703125</td>
<td>0.520499</td>
<td>0.064544</td>
<td>0.002437</td>
</tr>
<tr>
<td>3</td>
<td>6.60000</td>
<td>8.12451</td>
<td>-1.52451</td>
<td>-0.327863</td>
<td>0.566709</td>
<td>0.055310</td>
<td>0.000450</td>
</tr>
<tr>
<td>4</td>
<td>9.40000</td>
<td>8.99214</td>
<td>0.40786</td>
<td>0.086889</td>
<td>0.679146</td>
<td>0.037269</td>
<td>0.00021</td>
</tr>
<tr>
<td>5</td>
<td>13.30000</td>
<td>14.33952</td>
<td>-1.03952</td>
<td>-0.226632</td>
<td>0.822233</td>
<td>0.080736</td>
<td>0.000322</td>
</tr>
<tr>
<td>6</td>
<td>6.40000</td>
<td>9.03684</td>
<td>-2.63684</td>
<td>-0.561450</td>
<td>0.818496</td>
<td>0.036263</td>
<td>0.000847</td>
</tr>
</tbody>
</table>

If you request a casewise listing of the diagnostics, you obtain line of information for each regression point. Here the regression and sample points are the same, so residuals and other associated statistics can be computed.
The ANOVA test gives you a brief guide to the improvement in model fit when you compare the local and global models. You either need a table of critical values of $F$, or use the FINV function in Excel $[\text{FINV}(0.05, 74.8697, 12.7887) \text{ is 2.368, so the ANOVA is significant: } p<0.001]$
### Part 10: Parameter Estimate Summaries

<table>
<thead>
<tr>
<th>Label</th>
<th>Minimum</th>
<th>Lwr Quartile</th>
<th>Median</th>
<th>Upr Quartile</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7.895511</td>
<td>10.407367</td>
<td>11.925733</td>
<td>13.882247</td>
<td>17.442397</td>
</tr>
<tr>
<td>PctEld</td>
<td>-0.562128</td>
<td>-0.406957</td>
<td>-0.238977</td>
<td>-0.065320</td>
<td>0.266729</td>
</tr>
<tr>
<td>PctFB</td>
<td>0.506413</td>
<td>1.447798</td>
<td>3.023736</td>
<td>3.754073</td>
<td>3.898406</td>
</tr>
<tr>
<td>PctPov</td>
<td>-0.230242</td>
<td>-0.136669</td>
<td>-0.070377</td>
<td>-0.032474</td>
<td>0.001953</td>
</tr>
</tbody>
</table>

**Run terminates normally**

Because the output from GWR3 would be voluminous if printed, the program provides a 5-number summary of the distribution of each parameter. If you don’t have the last line, something’s wrong!!
• If we’re satisfied that we’ve run the model that we intended to run...

• … and that the results appear to make sense...

• … we can proceed to map the output
End