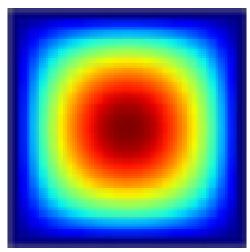
MGWR 1.0 User Manual



MGWR 1.0 Development Team

Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu and Wei Luo

Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA

Source code is available at: https://github.com/pysal/mgwr

1. Introduction

What is MGWR 1.0?

MGWR 1.0 is a new release of an application software (compatible with both Microsoft Windows and MacOS) for calibrating geographically weighted regression (GWR) and multi-scale geographically weighted regression (MGWR) models, which can be used to explore geographically varying relationships between dependent/response variables and independent/explanatory variables.

GWR

Traditional or global regression assumes that the relationships being examined through the model's parameters are constant over space. This assumption is relaxed in GWR by allowing the parameters to vary spatially. The GWR model formulation can be described as follows. Assuming that there are n observations, for the observation $i \in \{1, 2, . . . , n\}$ at location (u_i, v_i), the linear regression model is

$$y_i = \sum_j \beta_j (u_i, v_i) x_{ij} + \varepsilon_i ,$$

where x_{ij} is the jth predictor variable, $\beta_j (u_i, v_i)$ is the jth coefficient, ε_i is the error term, and y_i is the response variable.

Using the framework of geographically weighted generalised linear modelling, Binomial and Poisson regression models with geographically varying coefficients can also be used in MGWR 1.0 to explore geographically varying relationships for binary or count data.

MGWR

Whereas GWR constrains the local relationships within each model to vary at the same spatial scale, MGWR allows the conditional relationships between the response variable and the different predictor variables to vary at different spatial scales. That is, the bandwidths indicating the data-borrowing range can vary across parameter surfaces. It can be described as follows:

$$y_i = \sum_j \beta_{bwj} (u_i, v_i) x_{ij} + \varepsilon_i ,$$

Where *bwj* in β_{bwj} indicates the bandwidth used for calibration of the jth conditional relationship.

MGWR 1.0 enables the fitting of such GWR (Gaussian, Binomial and Poisson) and MGWR (currently supports Gaussian only) models with their associated statistical tests and model selections by user-defined data and model settings.

Main features

(1) Interface (GUI)

A user-friendly interface has been introduced to enable modelling sessions to intuitively proceed by selecting user-specified settings to the model. Results from the model are displayed once it completes and are saved in a text and table format. Several popular file types can be used as input data files (space, comma, tab separated text, and dbase IV formats). In addition, Areal key field (a unique ID field) can be integrated into the output of GWR and MGWR modelling, enabling you to join your output file to a GIS attribute table via the key field for mapping the result in a GIS environment.

(2) Requirements

MGWR 1.0 runs on Windows (Vista, 7, 8 and 10) and MacOS environments. The maximum size of data is dependent on your local machine environment. MGWR 1.0 dynamically allocates memory for large matrices (n by n, where n is the number of regression points) even for conventional GWR models. Thus it is recommended to use a PC having relatively large memory size (equal to or larger than 4GB) for running the software. If the system has a multi-core processor, MGWR 1.0 automatically uses multithreading routines to speed-up the computation.

Notes for use of MGWR 1.0

(i) MGWR 1.0 is copyrighted by the development team.
(ii) MGWR 1.0 can be freely distributed and used for academic and non-profit purposes.
(iii) When any results using the software are published, the author(s) must clearly state that MGWR 1.0 was used and give reference to SPARC website. Recommended citations for the theoretical background of MGWR modelling may be found in the *References* section of this manual.
(iv) Please note that the images used as graphic aids in the manual are produced on a system with Windows OS but all the same functionalities are also available for a MacOS based system.

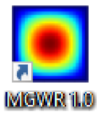
2. Installation / Uninstallation

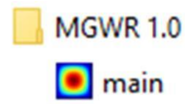
How to install MGWR 1.0

Download MGWR 1.0 for Windows (mgwr_v1_pc) or MacOS (mgwr_v1_macos) and double-click the installer (MGWR 1.0 for PC installer/MGWR 1.0.app) within the compressed downloaded package.

When the installer starts

Follow the instructions to select the MGWR 1.0 installation folder and users. On successful installation, a shortcut to the program will appear on your desktop and in the MGWR 1.0 program group.



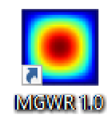


To uninstall

To uninstall MGWR 1.0 from your local environment, you may use "uninstall" option in the MGWR 1.0 program group. Alternatively, you can use the "Uninstall Programs" option in the Control Panel in a Windows system or simply drag the program's icon to "Trash" for MacOS.

3. Starting the program, Exiting the program, and GWR Modes

Starting the program



To start the program, double click the MGWR 1.0 shortcut icon on the desktop, or select it from the MGWR 1.0 program group.

GUI Introduction

On opening the program you will see the interface as below. There are two main modes to run your model - GWR and MGWR.

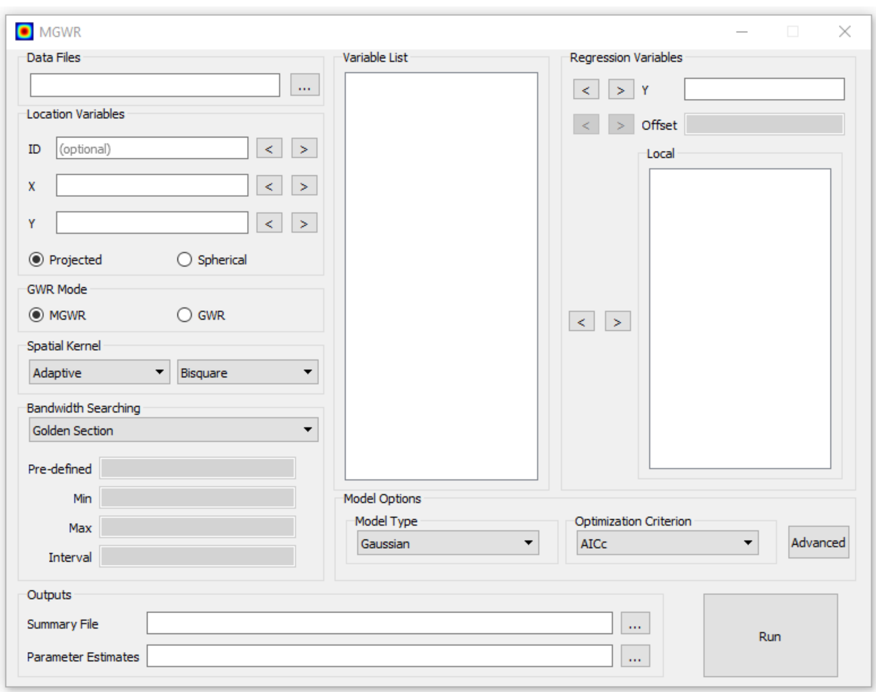
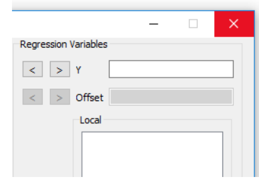


Figure 3.1: MGWR 1.0 main interface startup screen

Exiting the program

To exit the program you can press the Esc key or click the close button in the top-right corner of the window.



GWR Modes

As mentioned briefly above, there are two main modes for running the model in the program - GWR and MGWR. This program incorporates the widely used approach to modeling process heterogeneity - Geographically Weighted Regression (GWR) as well as the more recent and advanced approach - Multiscale GWR (MGWR) which relaxes the assumption that all of the processes being modeled operate at the same spatial scale. Inferences for both these models can be made within the software. We now describe (i) the data preparation step for running the models, (ii) the typical operation of a GWR model calibration and (iii) the typical operation of an MGWR model calibration.

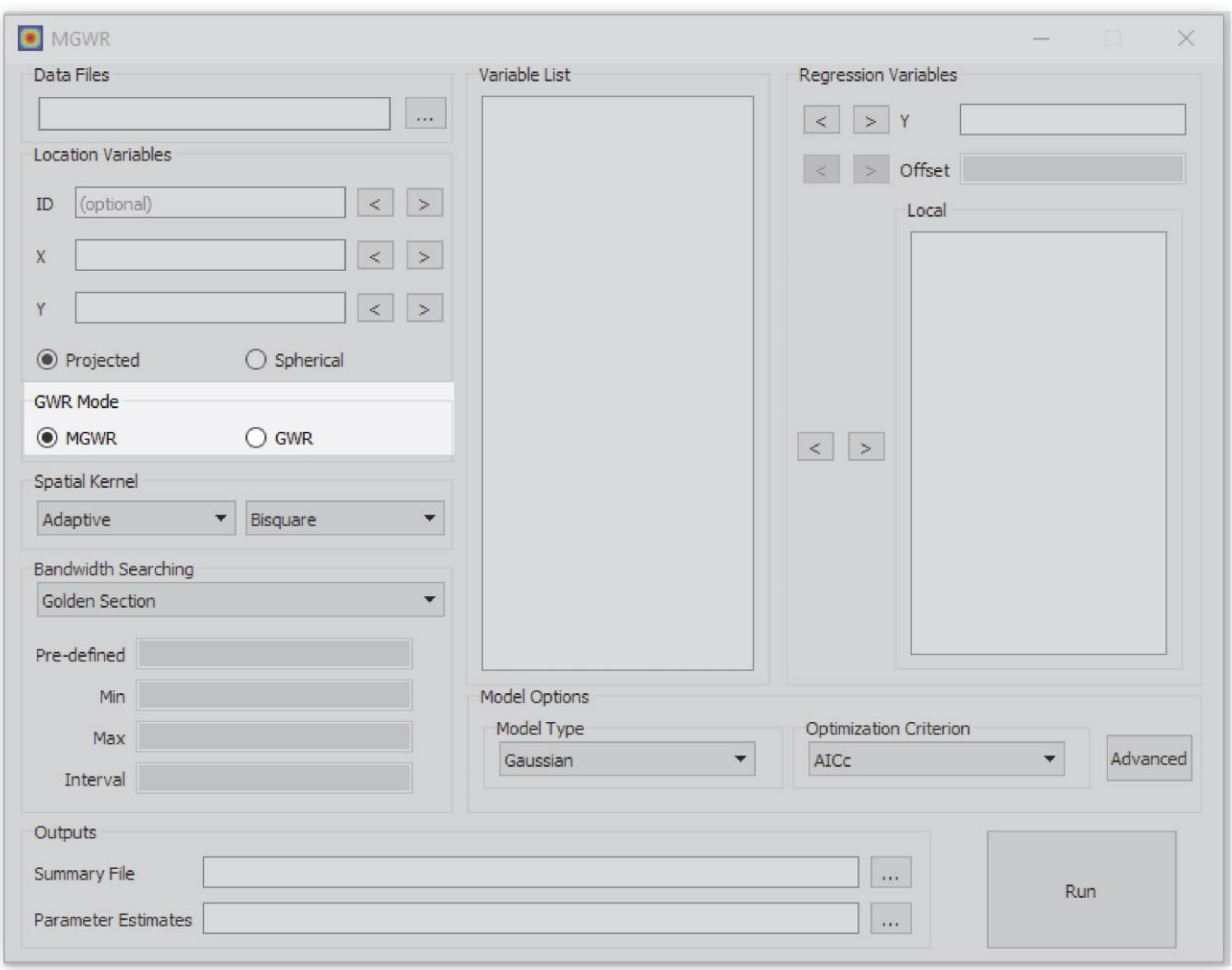


Figure 3.2: The two modes for operating MGWR 1.0

4. Data preparation and basic operation definitions

4.1 Preparing your data

What fields do I have to prepare in my dataset?

To calibrate a GWR model, you must prepare a tabular dataset that contains fields of dependent and independent variables, and x-y coordinates. Every variable must contain numeric values, with an exception of the ID column which may contain string characters. The ID column is treated as a string field in the program. No 'strange' characters (like $, &, *, @, # etc.) are permitted in any of the field values in the table.

Coordinates

Both 'Projected' and 'Spherical' values for locations are supported in the program and can be used as x and y coordinates in MGWR 1.0. 'Projected' is typically used for coordinates projected onto an orthogonal two-dimensional space, such as UTM coordinates while 'Spherical' is used for x-y coordinates stored in decimal degrees such as latitude longitude. These can be selected as options on the interface as shown in the image below. Also shown are examples of the two types of values.

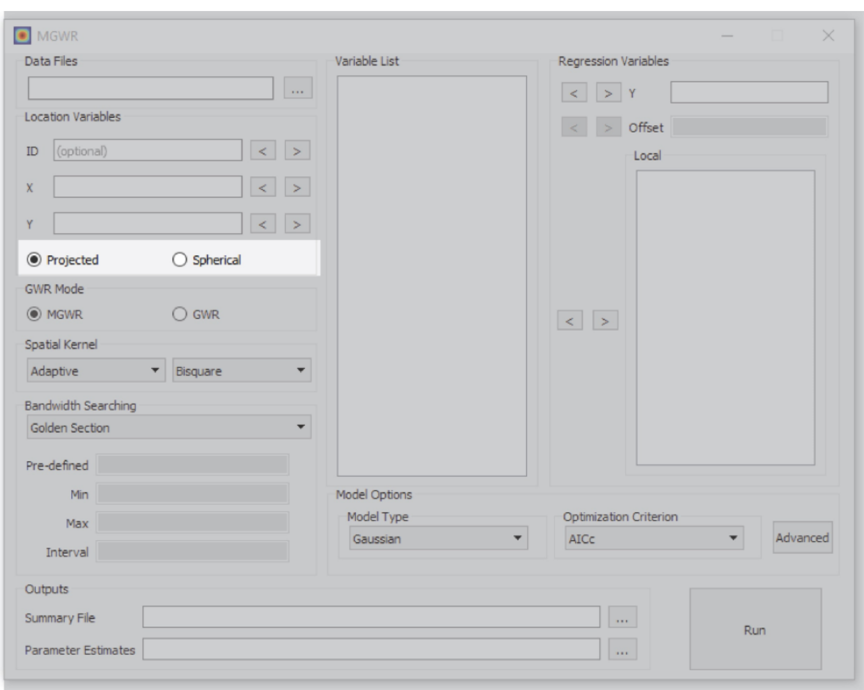


Figure 4.1: Coordinate value options (Projected and Spherical)

X	Y
941396.6	3521764
895553	3471916
930946.4	3502787
745398.6	3474765
849431.3	3665553
819317.3	3807616
803747.1	3769623
699011.5	3793408
863020.8	3520432

Latitude	Longitude
31.75339	-82.28558
31.29486	-82.87474
31.55678	-82.45115
31.33084	-84.45401
33.07193	-83.25085
34.3527	-83.50054
33.99347	-83.71181
34.2384	-84.83918
31.7594	-83.21976

Example of Projected (top) and Spherical (bottom) latitude/longitude values

ID field

If you have a unique locational ID field (such as place names or regional codes), this option lets you include that in your model for later use. Using this ID key, you can join your resulting parameter estimates table to other tables for eg. a shapefile with the same ID field to map the results.

Possible data formats

MGWR 1.0 supports excel data file formats (*.xls, *.xlsx), comma delimited text files (*.csv) and dbase IV file format (*.dbf). Below are examples of some of these file formats.

AreaKey	Latitude	Longitude	TotPop90	PctRural	PctBach	PctEld	PctFB	PctPov	PctBlack	ID	X	Y
13001	31.75339	-82.28558	15744	75.6	8.2	11.43	0.64	19.9	20.76	133	941396.6	3521764
13003	31.29486	-82.87474	6213	100	6.4	11.77	1.58	26	26.86	158	895553	3471916
13005	31.55678	-82.45115	9566	61.7	6.6	11.11	0.27	24.1	15.42	146	930946.4	3502787
13007	31.33084	-84.45401	3615	100	9.4	13.17	0.11	24.8	51.67	155	745398.6	3474765
13009	33.07193	-83.25085	39530	42.7	13.3	8.64	1.43	17.5	42.39	79	849431.3	3665553
13011	34.3527	-83.50054	10308	100	6.4	11.37	0.34	15.1	3.49	23	819317.3	3807616
13013	33.99347	-83.71181	29721	64.6	9.2	10.63	0.92	14.7	11.44	33	803747.1	3769623
13015	34.2384	-84.83918	55911	75.2	9	9.66	0.82	10.7	9.21	24	699011.5	3793408
13017	31.7594	-83.21976	16245	47	7.6	12.81	0.33	22	31.33	138	863020.8	3520432
13019	31.27424	-83.23179	14153	66.2	7.5	11.98	1.19	19.3	11.62	153	859915.8	3466377
13021	32.80451	-83.69915	149967	16.1	17	12.23	1.06	19.2	41.68	85	809736.9	3636468
13023	32.43552	-83.33121	10430	57.9	10.3	12.6	0.64	18.3	22.36	100	844270.1	3595691
13025	31.19702	-81.98323	11077	100	5.8	9.02	0.33	18.2	4.58	159	979288.9	3463849
13027	30.84653	-83.57726	15398	65.6	9.1	13.68	1.76	25.9	41.47	169	827822	3421638
13029	32.02037	-81.43763	15438	80.6	11.8	7.22	0.45	13.2	14.85	118	1023145	3554982
13031	32.39071	-81.74391	43125	63.2	19.9	9.56	1.16	27.5	25.95	97	994903.4	3600493
13033	33.05837	-81.99939	20579	72.3	9.6	10.6	0.43	30.3	52.19	71	971593.8	3671394
13035	33.28834	-83.95713	15326	73.4	7.2	10.41	0.72	15.6	35.48	65	782448.2	3684504
13037	31.52793	-84.61891	5013	100	10.1	15.94	0.1	31.8	58.89	149	724741.2	3492653

Figure 4.2: Georgia sample data file (available on software website) in .xlsx format

```
AreaKey,Latitude,Longitude,TotPop90,PctRural,PctBach,PctEld,PctFB,PctPov,PctBlack,ID,X,Y
13001,31.75339,-82.28558,15744,75.6,8.2,11.43,0.64,19.9,20.76,133,941396.6,3521764
13003,31.29486,-82.87474,6213,100,6.4,11.77,1.58,26,26.86,158,895553,3471916
13005,31.55678,-82.45115,9566,61.7,6.6,11.11,0.27,24.1,15.42,146,930946.4,3502787
13007,31.33084,-84.45401,3615,100,9.4,13.17,0.11,24.8,51.67,155,745398.6,3474765
13009,33.07193,-83.25085,39530,42.7,13.3,8.64,1.43,17.5,42.39,79,849431.3,3665553
13011,34.3527,-83.50054,10308,100,6.4,11.37,0.34,15.1,3.49,23,819317.3,3807616
13013,33.99347,-83.71181,29721,64.6,9.2,10.63,0.92,14.7,11.44,33,803747.1,3769623
13015,34.2384,-84.83918,55911,75.2,9,9.66,0.82,10.7,9.21,24,699011.5,3793408
13017,31.7594,-83.21976,16245,47,7.6,12.81,0.33,22,31.33,138,863020.8,3520432
13019,31.27424,-83.23179,14153,66.2,7.5,11.98,1.19,19.3,11.62,153,859915.8,3466377
13021,32.80451,-83.69915,149967,16.1,17,12.23,1.06,19.2,41.68,85,809736.9,3636468
13023,32.43552,-83.33121,10430,57.9,10.3,12.6,0.64,18.3,22.36,100,844270.1,3595691
13025,31.19702,-81.98323,11077,100,5.8,9.02,0.33,18.2,4.58,159,979288.9,3463849
13027,30.84653,-83.57726,15398,65.6,9.1,13.68,1.76,25.9,41.47,169,827822,3421638
13029,32.02037,-81.43763,15438,80.6,11.8,7.22,0.45,13.2,14.85,118,1023145,3554982
13031,32.39071,-81.74391,43125,63.2,19.9,9.56,1.16,27.5,25.95,97,994903.4,3600493
13033,33.05837,-81.99939,20579,72.3,9.6,10.6,0.43,30.3,52.19,71,971593.8,3671394
13035,33.28834,-83.95713,15326,73.4,7.2,10.41,0.72,15.6,35.48,65,782448.2,3684504
13037,31.52793,-84.61891,5013,100,10.1,15.94,0.1,31.8,58.89,149,724741.2,3492653
```

Figure 4.3: Georgia sample data file (available on software website) in .csv format

Field names

The first row in any of the file formats listed above must be the list of field names. (A dbase IV file automatically defines the first line as field names)

Missing values

MGWR 1.0 does not have functions for handling missing values representations. Records/rows having blank items in any variable field/column are skipped completely in the model fitting. In case of any missing values in your dataset, please leave those spaces blank instead of having NULL or NAN or any equivalent value representations to keep the GWR model from failing.

4.2 Basic operation definitions on the MGWR 1.0 interface

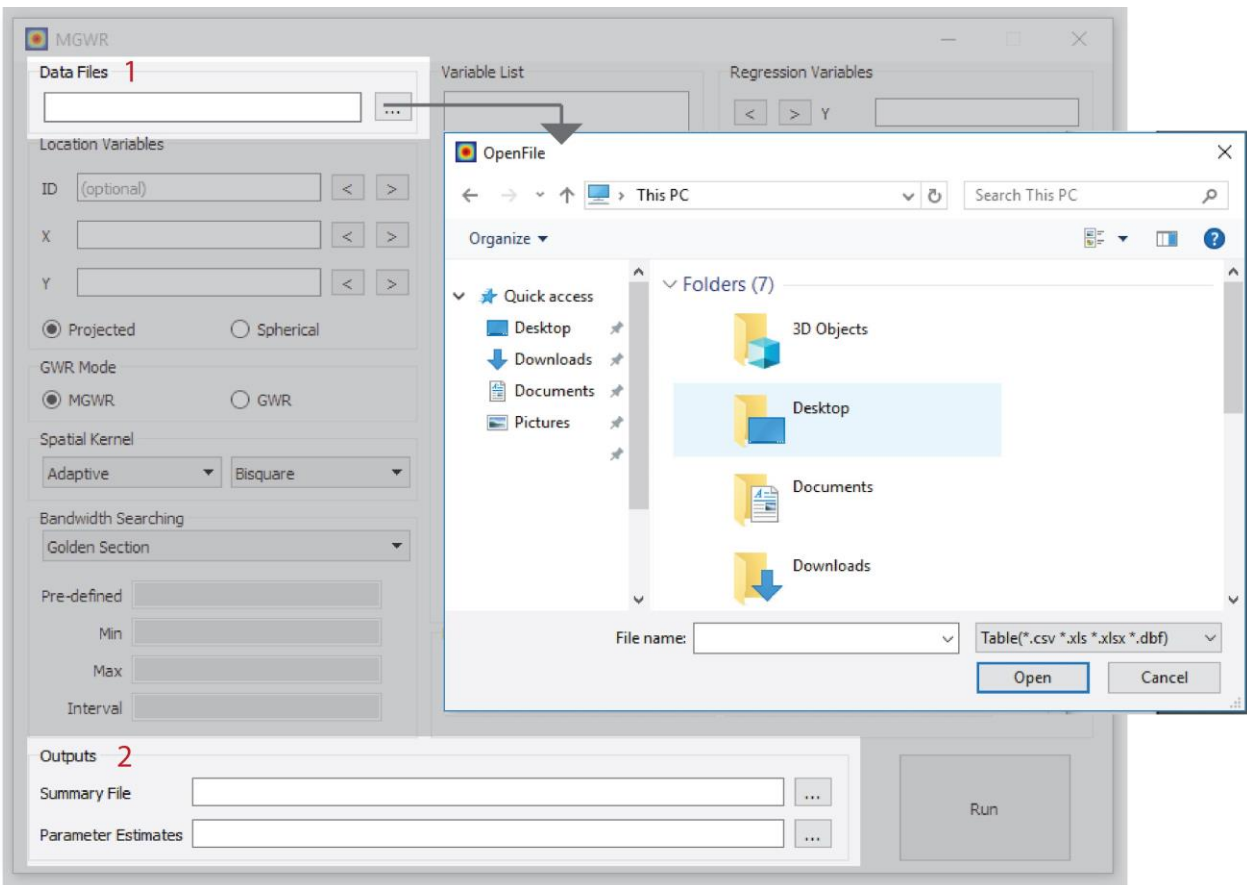


Figure 4.4: Data input and output tabs on the MGWR 1.0 interface

- The ‘Data Files’ tab (1) lets you input a local data file through the selection window from your computer

- The ‘Outputs’ tab (2) opens a similar selection window to select a spot on your computer to save the output file

  Two outputs are generated by the MGWR 1.0 program:
  - Summary file (MGWR_session_summary.text)- This is a text file that contains the Global regression results, Geographically weighted regression results and summary statistics for the GWR model
  - Parameter Estimates (MGWR_session_betas.csv) - This table contains all the parameter estimates for each location in the subject area and its t-values. This table can be easily plotted in any standard mapping package (like ArcMap, QGIS etc) or with mapping packages in Python (like Geopandas etc.)

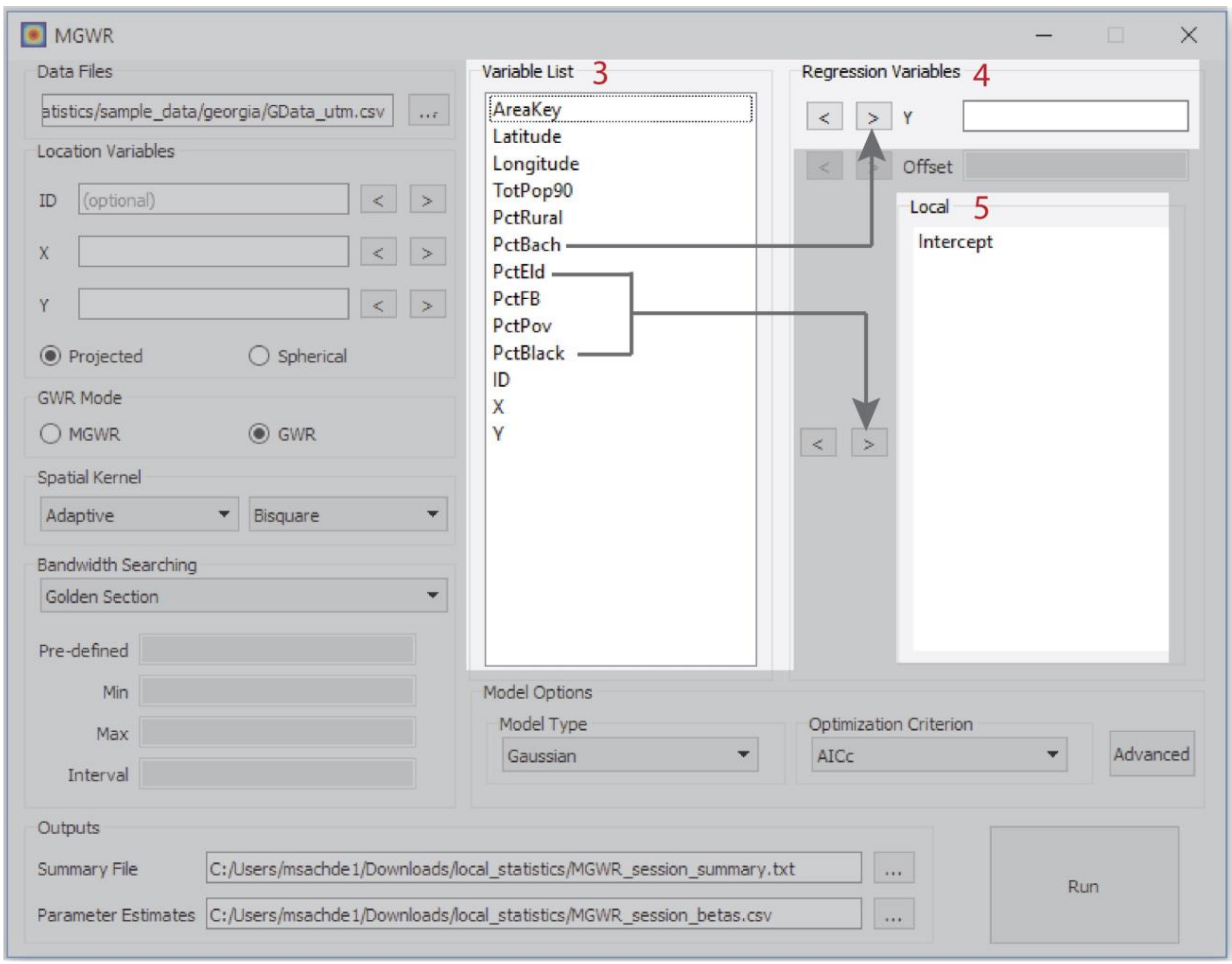


Figure 4.5: Variable list, independent and dependent variable specification options

- Once you upload the data file into the program, the variables of your table appear in the ‘Variable List’ tab (3).

- From there you can select a variable and use the arrow key (“>”) to move it to the ‘Regression Variable’ tab (4) to specify it as your dependent variable

- You can then select multiple variables at once and use the arrow key (“>”) to move it to the ‘Local’ tab (5) to specify the independent variables. You can also undo the action by selecting the “<” key.

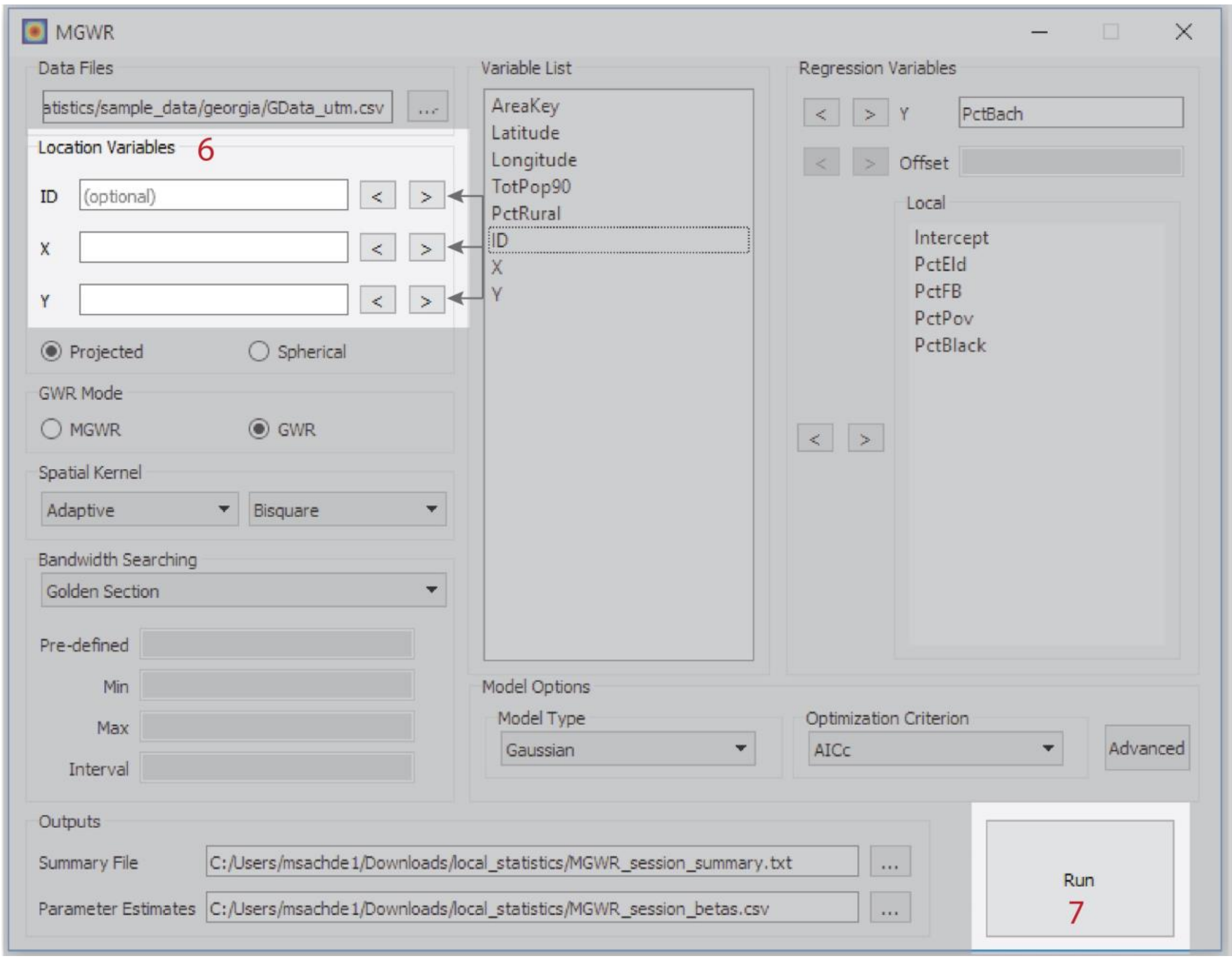


Figure 4.6: Location Variables tab to specify the ID and location variables

- Once the dependent and independent variables have been selected and specified in the model you can move the ID variable and coordinate variables (projected or spherical as discussed above) in the ‘Location Variables’ tab (6)

- After going through these basic steps and the model options (elaborated in the next section) you can run the model by pressing the ‘Run’ tab (7)

4.3 Model options

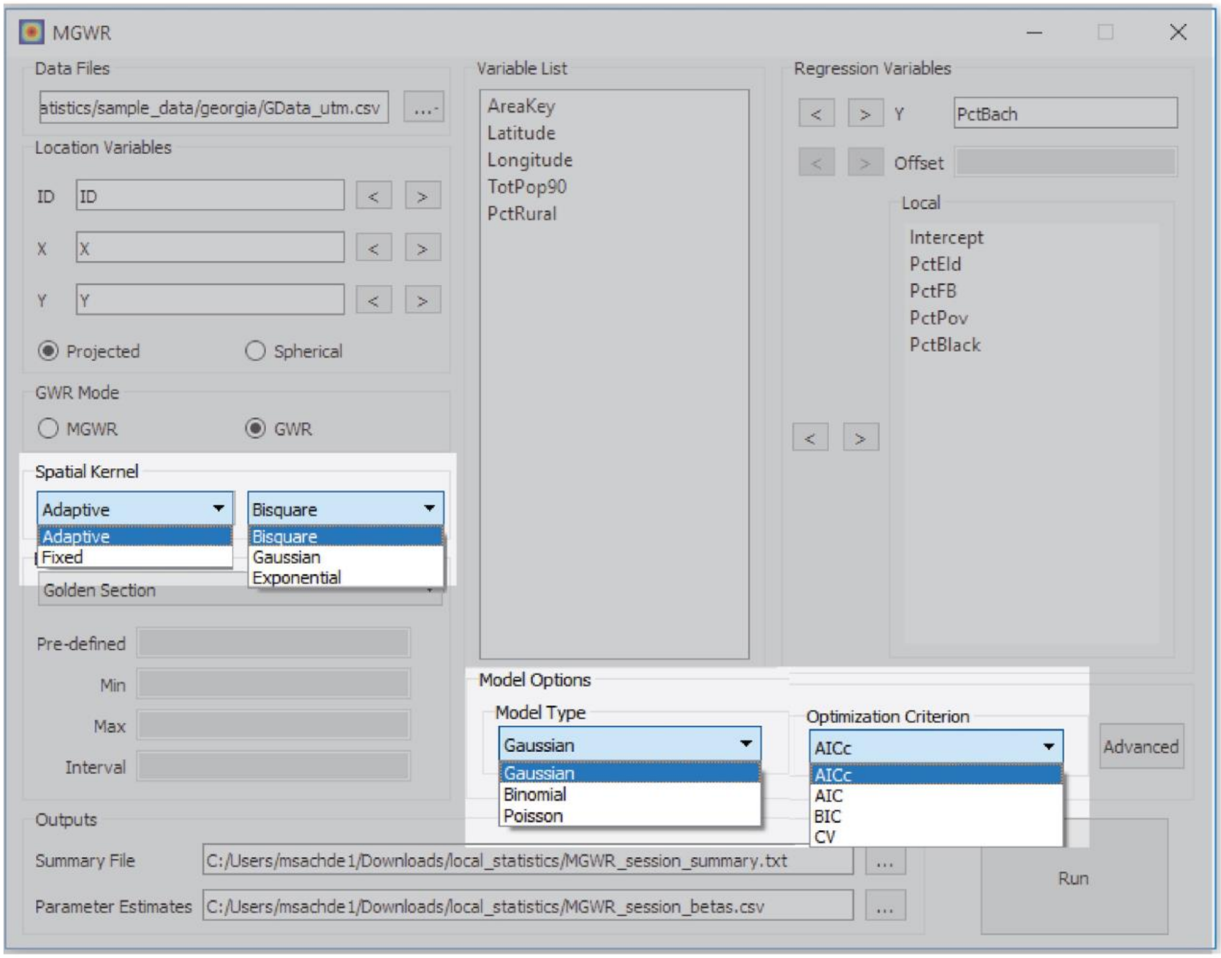


Figure 4.7: Model specification options

There are several detailed specification options for your model such as 'Spatial Kernel', 'Bandwidth Searching' methods, 'Model Type' and 'Optimization Criterion'. Let's explore these one by one.

1. **Spatial Kernel:** This lets you define the weighting scheme for your GWR/MGWR model. You can choose the function of the weighting scheme to be either Bisquare, Gaussian or Exponential. The default setting is the Bisquare function. The second option lets you specify whether the kernel would be 'Adaptive' or 'Fixed'. An adaptive kernel controls for an optimal number of *k* neighbors to be included in the model fitting whereas a fixed kernel controls for an optimal bandwidth which is assumed constant over space for each point. Adaptive is the default option in the program though you can change it to fixed if you know your dataset to have evenly distributed points. (See below for diagrams explaining the concept)

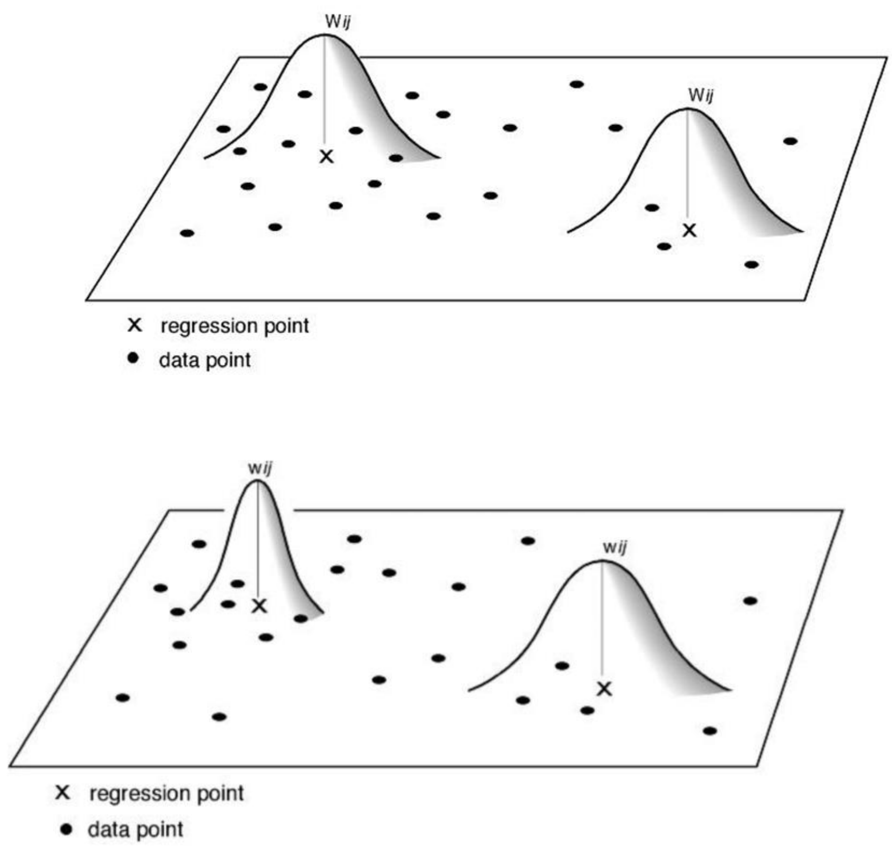


Figure 4.8: Conceptual diagrams explaining fixed (top) and adaptive (bottom) weighting schemes

2. **Model Options:** Research in GWR has found the result of a model to be relatively agnostic to the choice of weighting function (bisquare, gaussian or exponential) as long as it is a continuous distance-based function. It is however sensitive to the degree of distance-decay which can be optimized in a number of ways. MGWR 1.0 program lets you choose this Optimization Criterion from AICc (Corrected Akaike Information Criterion), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and CV (Cross Validation Score).
   Standard GWR uses Ordinary Least Squares which assumes a Gaussian error term. This is however not the best option for all kinds of data. In addition to Gaussian, depending on your data type you can choose Binomial (for model forms that predict the probability for binary outcomes like yes/no or 1/0) or Poisson (for count data like number of crimes, cases of illnesses etc.). Model selection option is only available for the GWR mode currently. The current version of the software allows only Gaussian implementation for MGWR model calibration and implementation of Poisson and Binomial models within MGWR is an area of ongoing research.

3. **Bandwidth Searching:** There are options that you can select from to define the algorithm the model will follow to select an optimal bandwidth. The default is the Golden Section search which finds the optimal value for the bandwidth by successively narrowing the range of values inside which the optimal value exists and comparing the optimization score of the model for each - returning the value which has the lowest score. Users can change that to an Interval search which allows you to define a limited range (Min and Max) within which the model will look and a step value (Interval) by which it will successively increment, to search for the optimal bandwidth. There is a third option which allows you to input a pre-defined bandwidth for your model, which is particularly useful as an exploratory method to understand the behavior of your data and hence the model better.

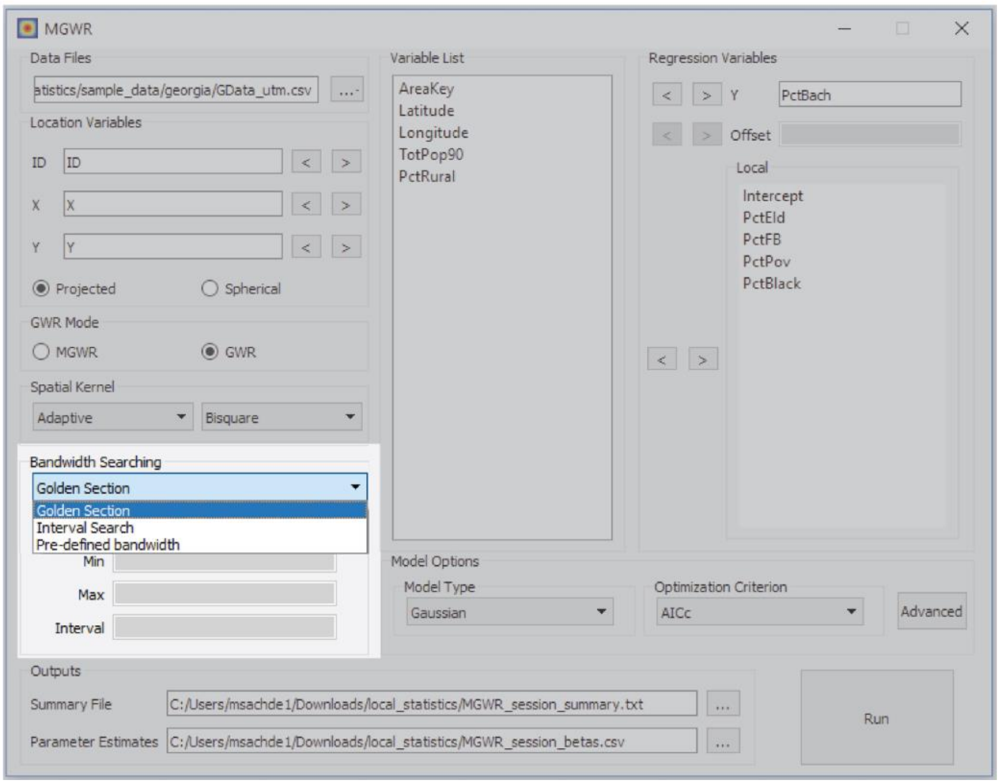


Figure 4.9: Option to define the bandwidth-search algorithm for the model

4.4 Advanced options tab

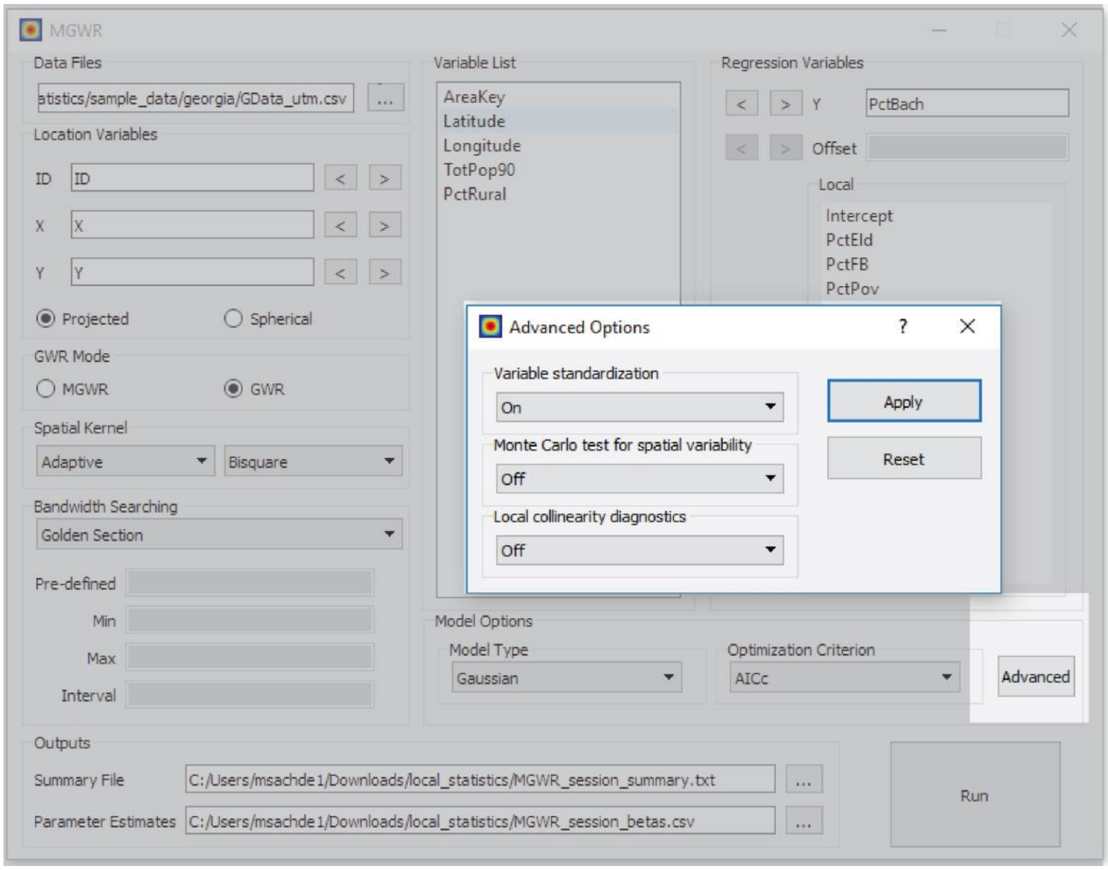


Figure 4.10: Advanced options in the MGWR 1.0 program

You can also choose from advanced options in the program to specify other things for the model. Below, we discuss the three common options that can be selected for both, GWR and MGWR modes.

1. **Variable standardization:** This option performs a z-transformation on dependent and independent variables so that each variable has a mean of 0 and standard deviation = 1. The default option for this option is 'On' as in most cases standardization makes iterative computation of model fitting faster. However, you can turn the feature 'Off' for better interpretability of parameter estimates.
2. **Monte Carlo test for spatial variability:** To test whether the spatial variability of the local estimates is attributable to sampling variation or a result of other inherent processes, the program provides the option to run a Monte Carlo test which basically runs once to derive the local parameter estimates with the given distribution and then many times after by randomly rearranging points - to measure that the variability of each parameter surface could have arisen by chance. Please note that this test takes significantly more time for the model to run.
3. **Local collinearity diagnostics:** To identify collinearity issues locally, you can turn the local collinearity diagnostics option 'On'. This returns a local condition index called 'local_CN' which

identifies the number of near dependencies among the columns of the design matrix. In addition, this test also provides a variance decomposition proportion ('local_VDP') and a local variance inflation factor ('local_VIF') for each covariate which in conjunction with the condition index provides a measure of the degree to which the corresponding regression estimate has been degraded by the presence of collinearity.

4.5 Running the model

Once you have selected all the options in the model, click the run button to the right corner. Once it runs successfully, you will see the following screen and the output files will download automatically to your selected local folder.

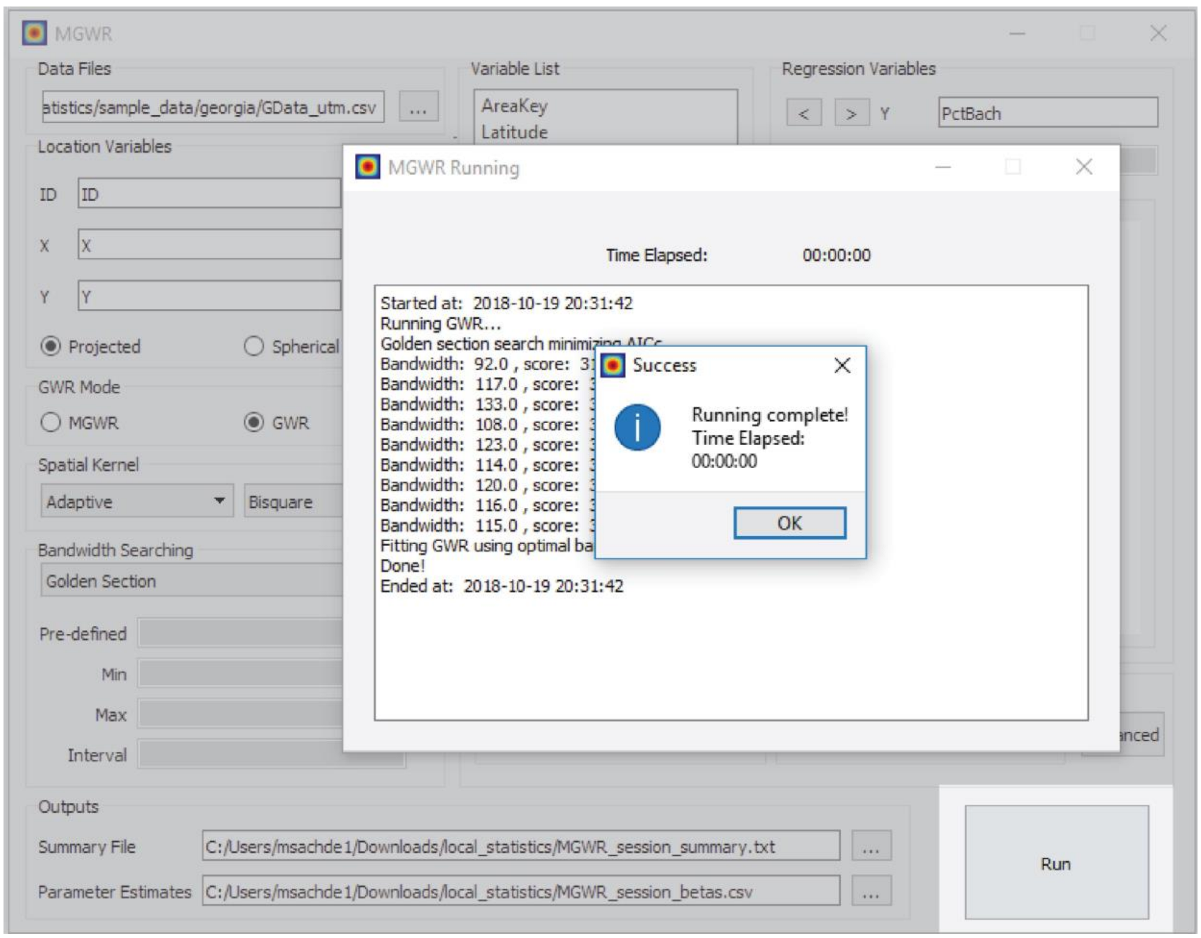


Figure 4.11: Screen when the model runs successfully

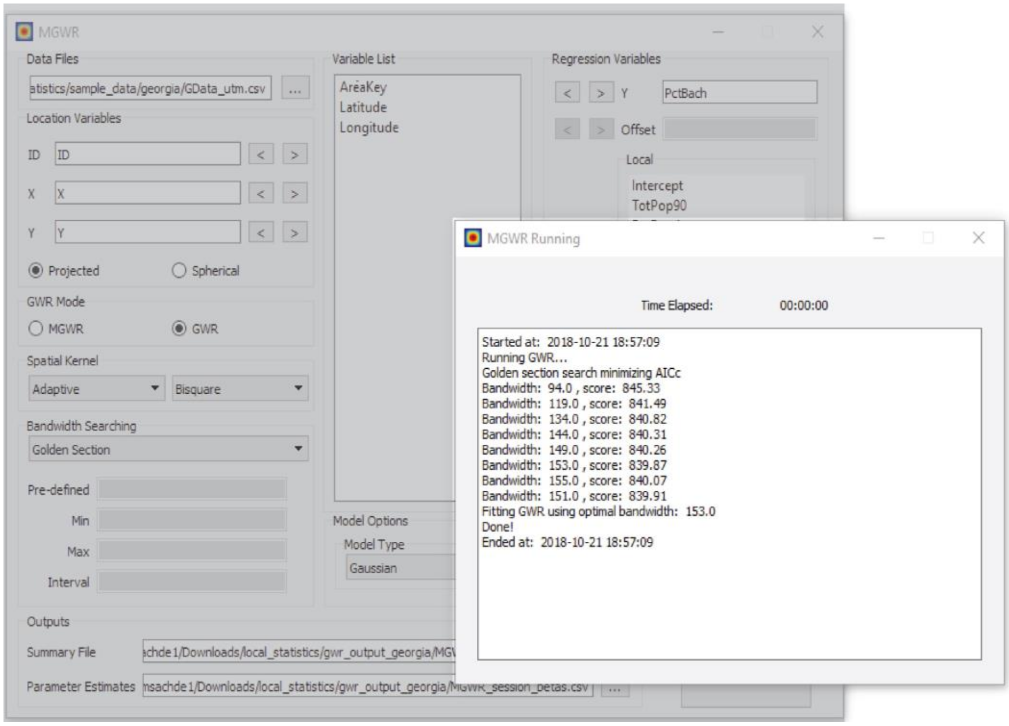


Figure 4.12: Window showing iterations searching for optimal bandwidth

The summary file from the output (see below) opens automatically on successful completion of the program.

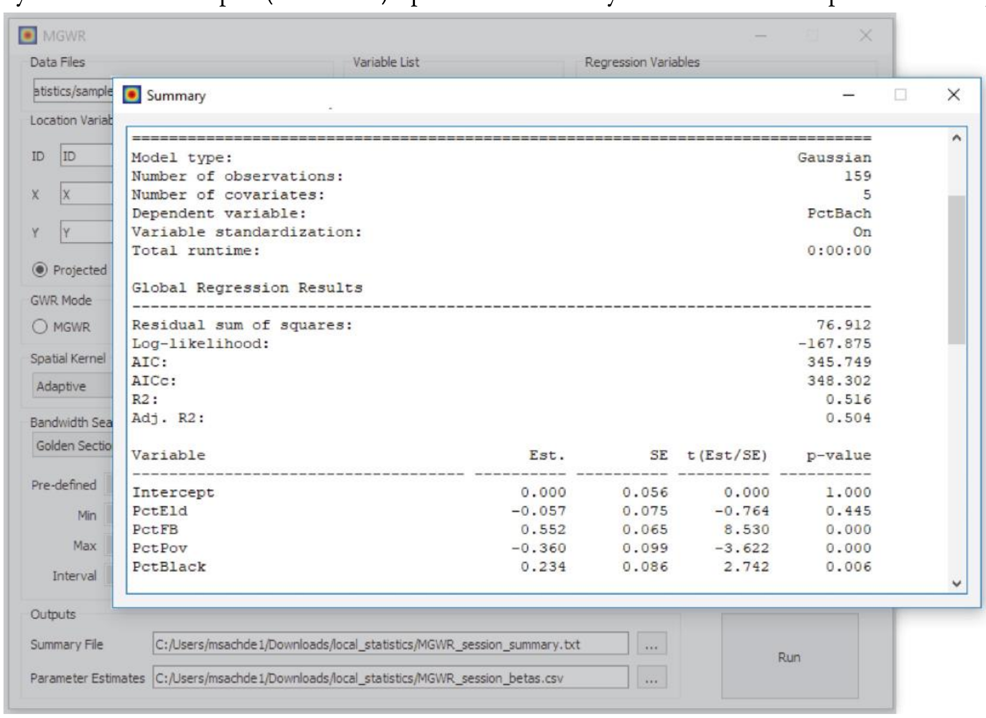


Figure 4.13: Output summary file

Let's run an example model for each model type next. To follow along, download the sample data (Tokyo, Georgia and Clearwater) from the program installation page.

5. Running different GWR model types with sample data

5.1 Gaussian GWR model

A conventional Gaussian GWR model is described as:

$$y_i = \sum_j \beta_k (u_i, v_i) x_{ij} + \varepsilon_i ,$$

where y_i, $x_{k,i}$ and ε_i are, respectively, dependent variable, *kth* independent variable, and the Gaussian error at location *i*; (u_i,v_i) is the x-y coordinate of the *ith* location; and coefficients β_k (u_i,v_i) are varying conditionals on the location. Usually, the first variable is constant by setting $x_{0,i} = 1$, after which β_0 (u_i,v_i) becomes a geographically varying "intercept" term.

Let's explore running a Gaussian GWR model using the Georgia sample data which has the following variables:

PctBach - percentage of inhabitant with at least a bachelor degree
TotPop90 - total population in 1990
PctRural - percentage of rural population
PctEld - percentage of elderly
PctFB - percentage of foreign-born inhabitants
PctPov - percentage of inhabitants living below the poverty level
PctBlack - percentage of African-Americans

and the following specifications:

$$\begin{aligned}\mathbf{PctBatch}_i = \; & \beta_0(X_i, Y_i) + \beta_1(X_i, Y_i)\,\mathbf{PctRural}_i \\ & + \beta_2(X_i, Y_i)\,\mathbf{PctPov}_i + \beta_3(X_i, Y_i)\,\mathbf{PctBlack}_i \\ & + \beta_4(X_i, Y_i)\,\mathbf{PctFB}_i + \beta_5(X_i, Y_i)\,\mathbf{PctEld}_i \\ & + \beta_7(X_i, Y_i)\,\mathbf{PctTotPop90}_i + \varepsilon_i\end{aligned}$$

where X_i and Y_i are projected x-y coordinates in this example.

For the input file we enter the 'GData_utm.csv' file and set the model features in the following way.

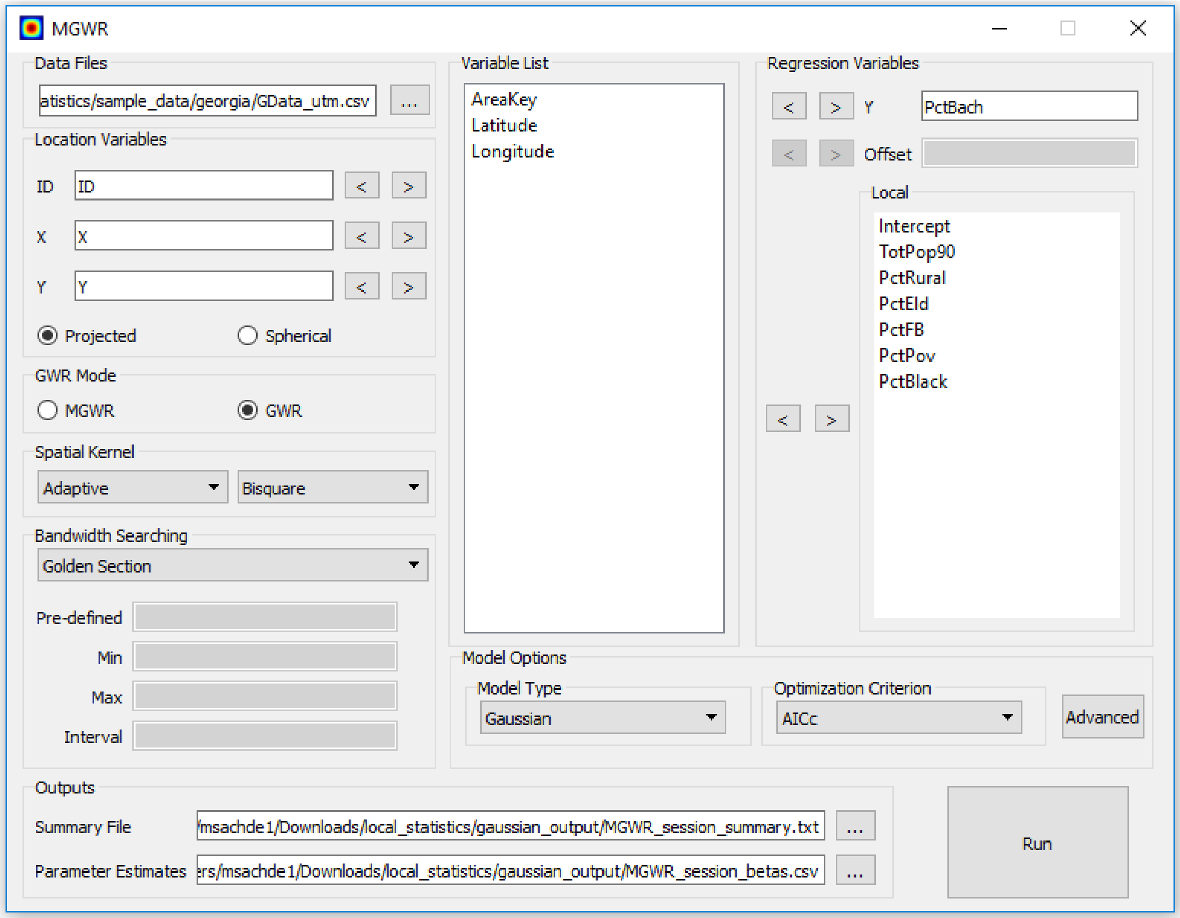
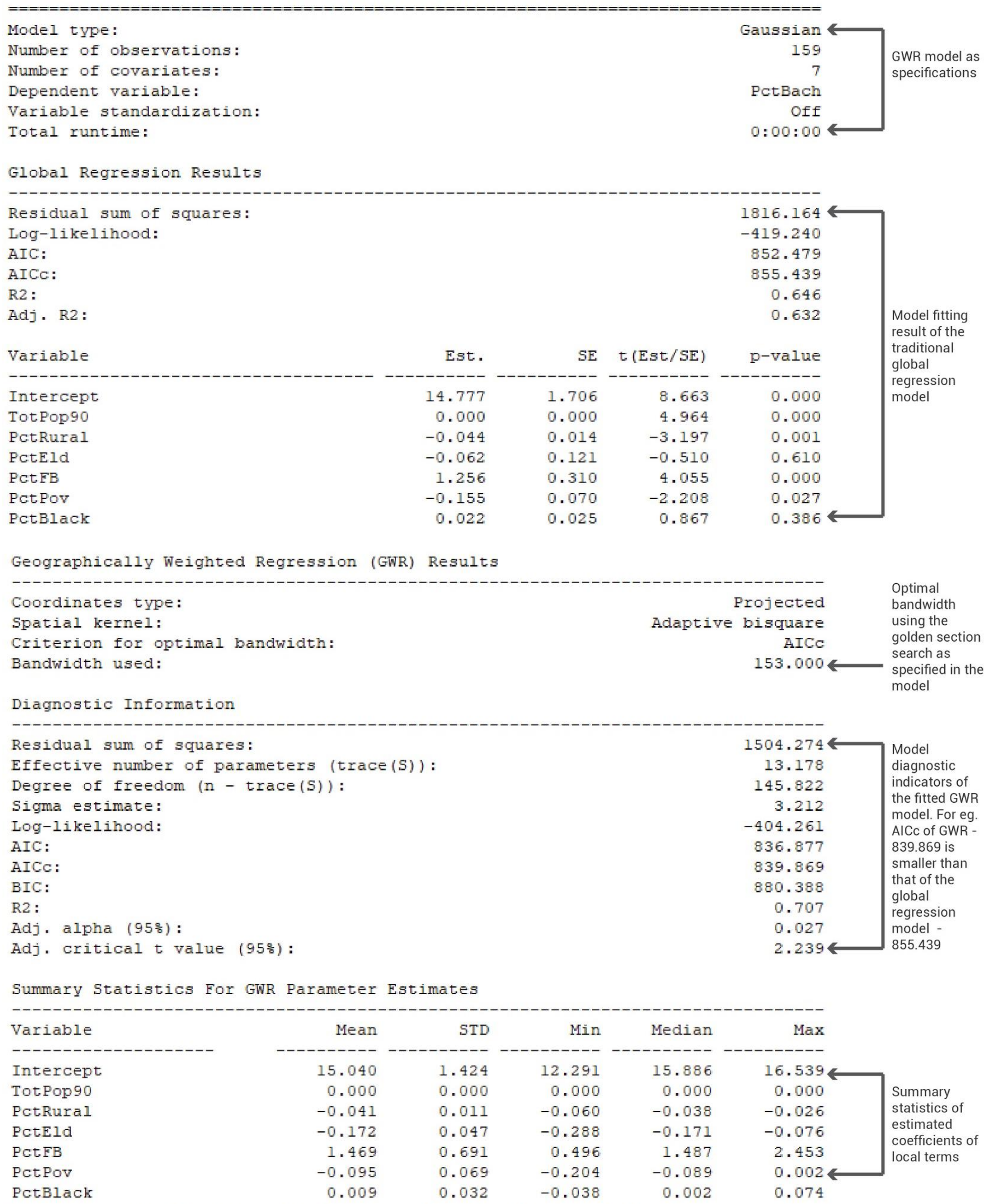


Figure 5.1: GWR Gaussian model specification in MGWR 1.0

Once you run the model, a results-summary appears in a text box. The content includes your modeling settings, global model result, best bandwidth, model diagnostic information of the GWR model. These are also saved in the output file as entered in the Summary File. The other GWR result file of parameter estimates (default name is MGWR_session_betas.csv) can be joined with the shapefile (G_utm.shp) based on the field 'ID'. Using AcrGIS, QGIS, R or Python , these local results can be mapped.



```
===========================================================================
Model type:                                                       Gaussian
Number of observations:                                                159
Number of covariates:                                                    7
Dependent variable:                                                PctBach
Variable standardization:                                              Off
Total runtime:                                                     0:00:00

Global Regression Results
---------------------------------------------------------------------------
Residual sum of squares:                                          1816.164
Log-likelihood:                                                   -419.240
AIC:                                                               852.479
AICc:                                                              855.439
R2:                                                                  0.646
Adj. R2:                                                             0.632

Variable                                   Est.         SE  t(Est/SE)    p-value
----------------------------------- ---------- ---------- ---------- ----------
Intercept                               14.777      1.706      8.663      0.000
TotPop90                                 0.000      0.000      4.964      0.000
PctRural                                -0.044      0.014     -3.197      0.001
PctEld                                  -0.062      0.121     -0.510      0.610
PctFB                                    1.256      0.310      4.055      0.000
PctPov                                  -0.155      0.070     -2.208      0.027
PctBlack                                 0.022      0.025      0.867      0.386

Geographically Weighted Regression (GWR) Results
---------------------------------------------------------------------------
Coordinates type:                                                Projected
Spatial kernel:                                          Adaptive bisquare
Criterion for optimal bandwidth:                                      AICc
Bandwidth used:                                                    153.000

Diagnostic Information
---------------------------------------------------------------------------
Residual sum of squares:                                          1504.274
Effective number of parameters (trace(S)):                          13.178
Degree of freedom (n - trace(S)):                                  145.822
Sigma estimate:                                                      3.212
Log-likelihood:                                                   -404.261
AIC:                                                               836.877
AICc:                                                              839.869
BIC:                                                               880.388
R2:                                                                  0.707
Adj. alpha (95%):                                                    0.027
Adj. critical t value (95%):                                         2.239

Summary Statistics For GWR Parameter Estimates
---------------------------------------------------------------------------
Variable                   Mean        STD        Min     Median        Max
--------------------      ---------- ---------- ---------- ---------- ----------
Intercept                15.040      1.424     12.291     15.886     16.539
TotPop90                  0.000      0.000      0.000      0.000      0.000
PctRural                 -0.041      0.011     -0.060     -0.038     -0.026
PctEld                   -0.172      0.047     -0.288     -0.171     -0.076
PctFB                     1.469      0.691      0.496      1.487      2.453
PctPov                   -0.095      0.069     -0.204     -0.089      0.002
PctBlack                  0.009      0.032     -0.038      0.002      0.074
---------------------------------------------------------------------------
```


Figure 5.2: Summary description for the model

5.2 Poisson GWR model

A Gaussian error term is suitable for modeling numerical responses. However, in the case of modeling count or binary (dichotomous) responses, other model types of generalized linear modeling, particularly Logistic and Poisson regression, are quite popular. Let's explore an example of using the geographically weighted poisson regression for modeling count data output.
A Poisson GWR model and can be shown as:

$$y_i \sim \text{Poisson}\left[N_i \, exp\left(\textstyle\sum_k \boldsymbol{\beta_k} \, (\boldsymbol{u_i}, \boldsymbol{v_i}) \, x_{ki}\right)\right],$$

The dependent variable should be an integer that is greater than or equal to zero. N_i is the offset variable at the *ith* location. This term is often the size of the population at risk or the expected size of the outcome in spatial epidemiology. In cases where the "offset variable" box is left blank, N_i becomes 1.0 for all locations.
For example, in the Tokyo sample dataset for modeling premature mortality in Tokyo (dependent variable db2564 - premature mortality count) we set the offset variable to eb2564 - total mortality count and independent variables as POP65 (population over 65), OCC_TEC (population with professional occupation), OWNH (home ownership) and UNEMP (unemployment).

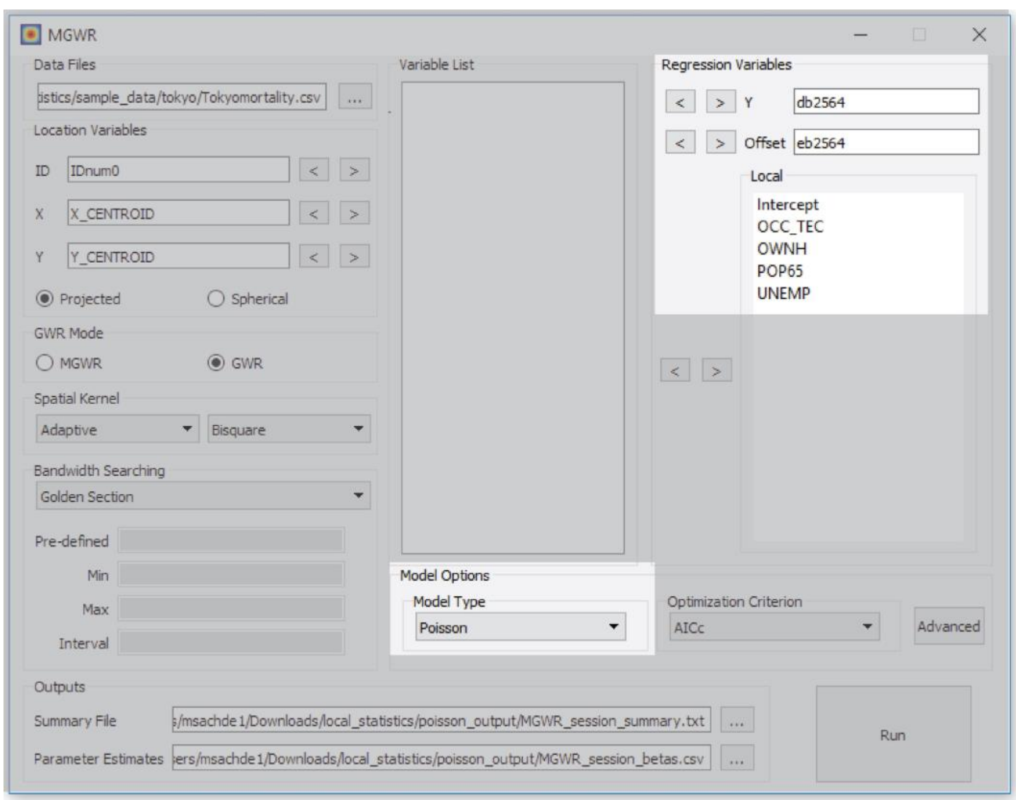


Figure 5.3: GWR Poisson model specification in MGWR 1.0

5.3 Binomial GWR model

A Binomial GWR model is shown as

$$y_i \sim \text{Bernoulli}\,[p_i]$$
$$\text{logit}\,(p_i) = \sum_k \beta_k\,(u_i, v_i)\, x_{ki}$$

The dependent variable must be 0 or 1. p_i is the modelled probability that the dependent variable becomes one.

From the Clearwater dataset, we use Binomial GWR to model the dependent variable Landslid (probability of landslide occurrence in Clearwater National Park) using independent variables Elev(elevation) and Slope(slope in percentage).

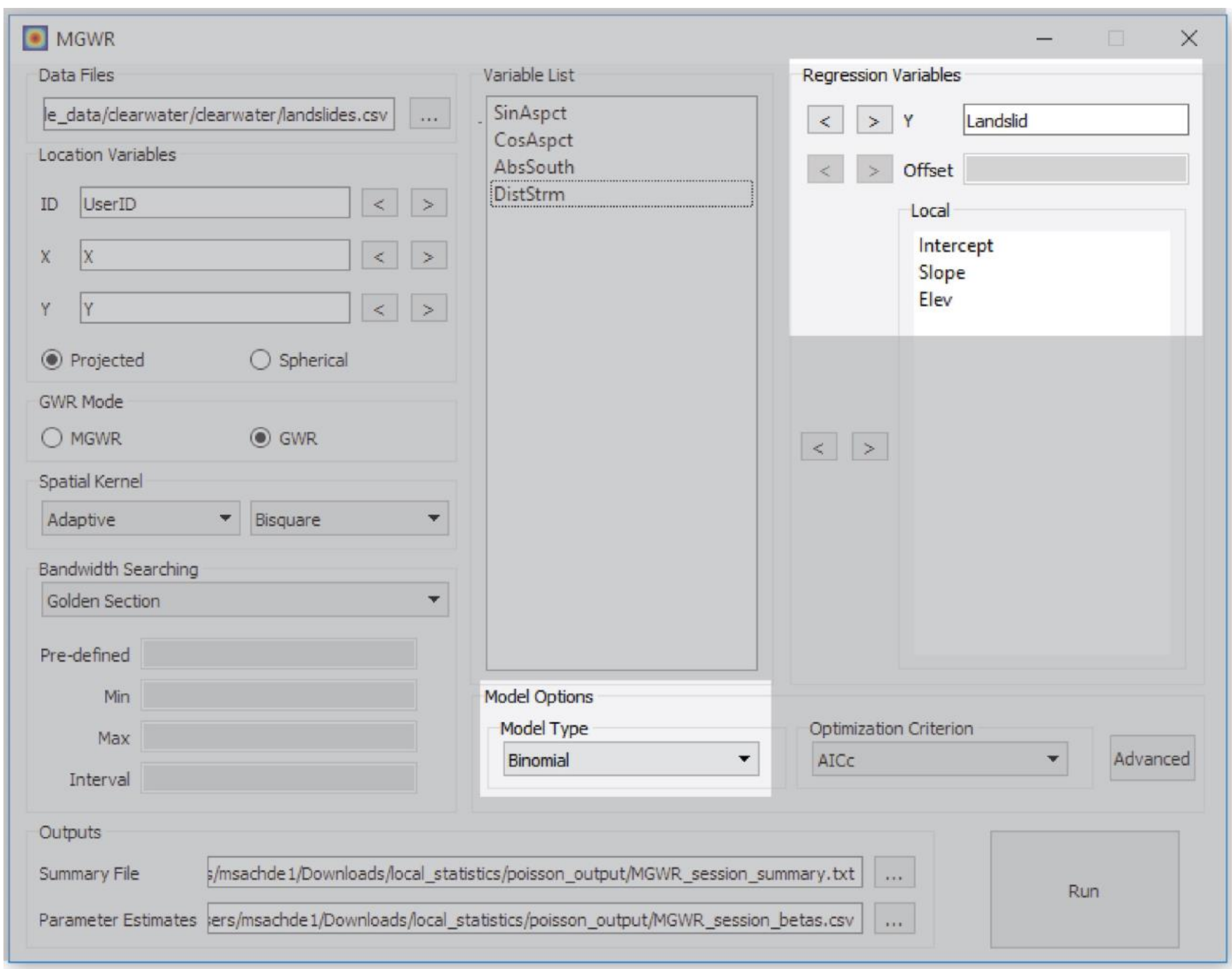


Figure 5.4: GWR Binomial model specification in MGWR 1.0

6. Running an MGWR model

Let's work through an example of the MGWR model with the sample Georgia dataset.

PctBach - percentage of inhabitants with at least a bachelor degree
TotPop90 - total population in 1990
PctRural - percentage of rural population
PctEld - percentage of elderly
PctFB - percentage of foreign-born inhabitants
PctPov - percentage of inhabitants living below the poverty level
PctBlack - percentage of African-Americans

To examine the spatial variation in percentage of inhabitants with at least a bachelor degree (**PctBach**), we use data of four other predictors percentage of African-Americans, foreign-born inhabitants, elderly and total population for the year 1990 (**PctBlack, PctFB, PctEld, TotPop90**). For the main interface in the program, most settings remain the same as for specifying a GWR model.

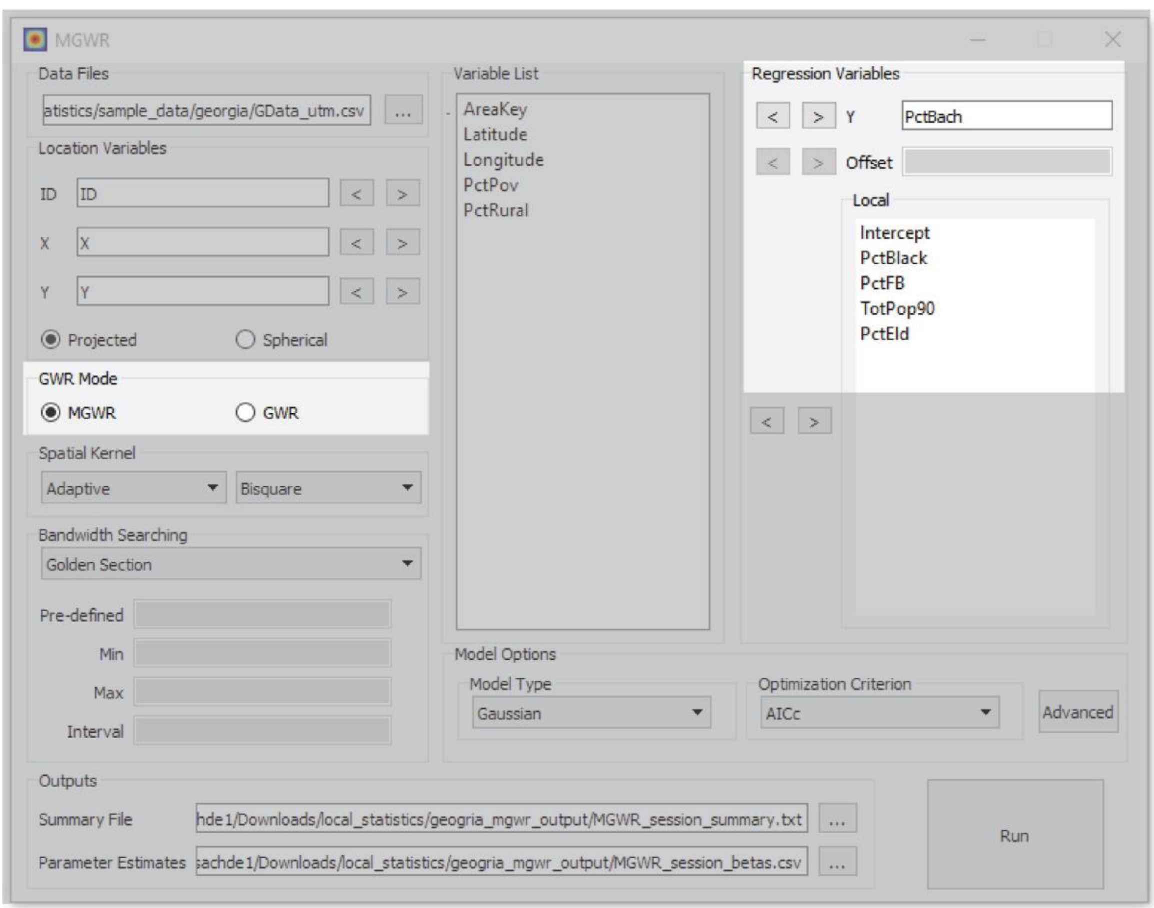


Figure 6.1: MGWR model specification in MGWR 1.0

In an MGWR model, as discussed in the introduction, optimal bandwidths vary across parameter surfaces. These different bandwidths imply that each relationship at the same location will have a different spatial weighting matrix. The estimator for GWR, therefore, is not applicable here and we use a smoother function in a back-fitting algorithm for the calibration of an MGWR model. The basic idea of back-fitting is to calibrate each term in the model with a smoother assuming that all the other terms are known. It is an iterative process where additive terms are first initialized by assigning initial estimates to each local coefficient. Through these initial estimates, each variable is then regressed with the initial estimate, producing an optimal bandwidth for each as they go and updating the initial estimates with new local estimates. These iterations continue until the changes of all the terms on successive iterations are sufficiently small to declare convergence (for more information see Fotheringham, Yang and Kang, 2017).
For the calibration, three main inputs from the user are involved in the algorithm. In addition to the options described in section 4.4 of the manual, these calibration inputs can be defined in the 'Advanced' tab in the software.

1. **Initialization:** First is the **choice of the initial estimates** for which the available options in the software are OLS and GWR as shown below. This choice might affect the number of iterations needed to reach convergence while not influencing the selected optimal bandwidth vector.

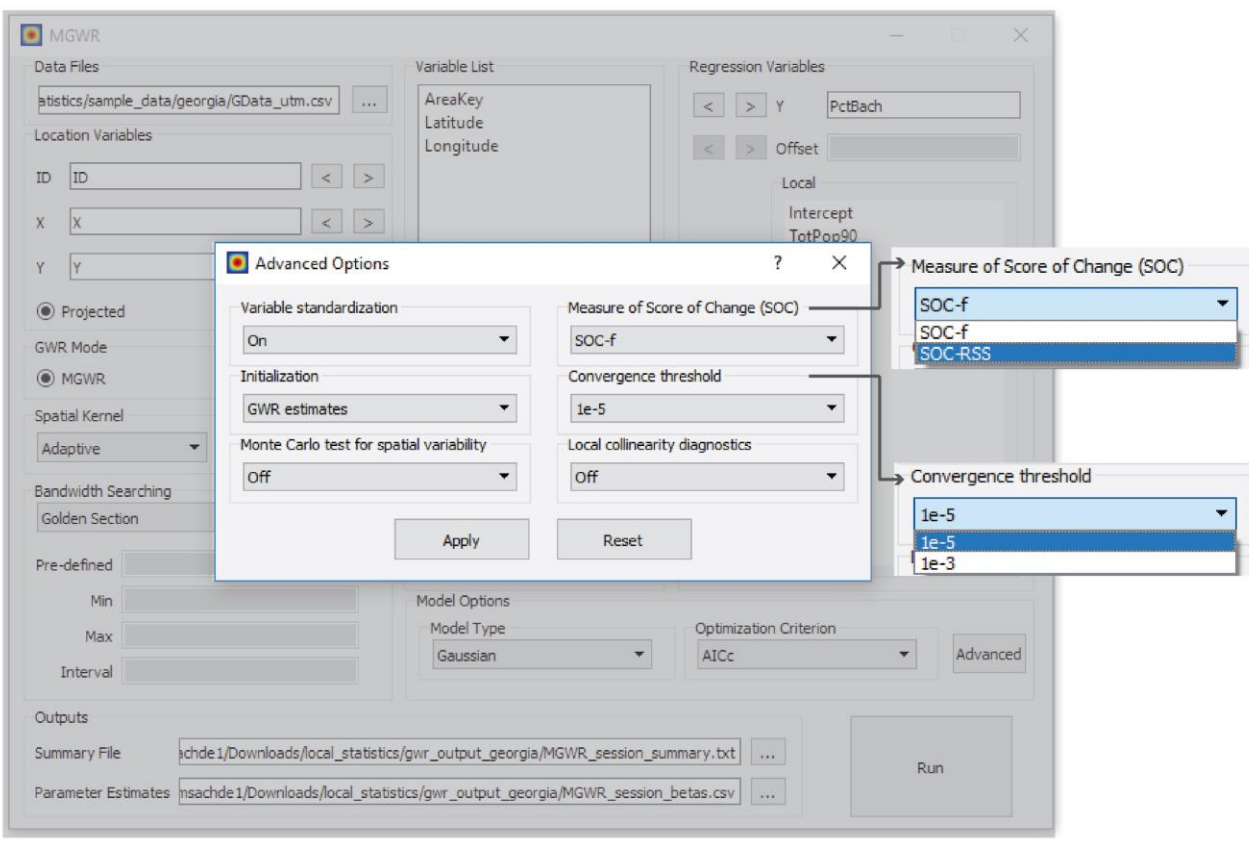


Figure 6.2: Advanced options for MGWR mode in MGWR 1.0

2. **Measure of Score of Change (SOC):** Second is the choice of termination criterion for the iterations to be deemed to have converged which is decided by the value of the differential between successive iterations (**score of change or SOC**). Two types of SOC can be used - SOC-RSS which is the proportional change in the residual sum of squares (RSS) and SOC-f which is the change in the GWR smoother. Both these are scale-free but SOC-f has the advantage of being focused on the relative changes in the additive terms rather than on the overall model fit, though it also may take longer to converge in some cases.
3. **Convergence Threshold:** Third is the choice of the convergence threshold where options are 10^{-5} (1e-5) and 10^{-3} (1e-3). This is simply the threshold of change below which the model is declared to have converged.

To run the MGWR model it is suggested to always **turn the 'Variable Standardization' option 'On'.** This follows from the fact that in the operation of MGWR models, the interpretation and comparison of the individual bandwidths is facilitated by standardizing all of the variables in the model (to have mean = 0 and standard deviation = 1). This allows the bandwidths to be direct indicators of the spatial scale at which the conditional relationship between y and the predictor variable varies. Without standardization the bandwidths will also reflect the scale and variation in each predictor variable.
For running the model we use Score of Change to be SOC-f, set the convergence threshold at 1e-5 and initialize our estimates with GWR estimates. The MGWR model calibration converges after 15 iterations.

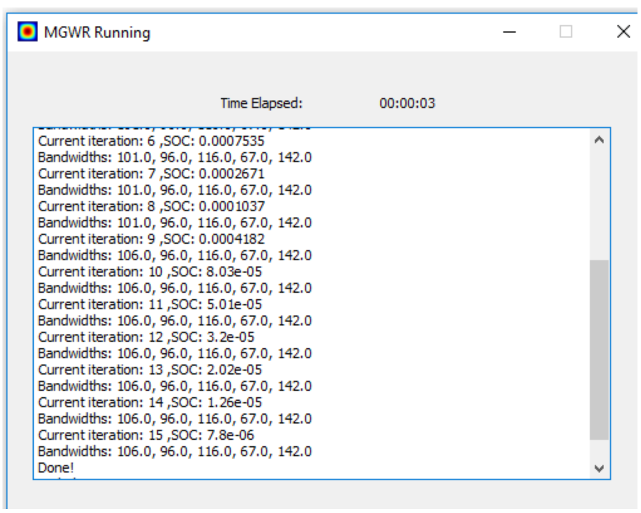


Figure 6.3: Window showing iterations for optimal bandwidths search in MGWR mode

After the model runs, a window showing summary results opens similar to the one in the GWR model and the results files are stored in the specified local folder automatically. The individual optimal bandwidths specific to each parameter in the MGWR model are found as below.

MGWR bandwidths

Variable	Bandwidth	ENP_j	Adj t-val(95%)	Adj alpha(95%)
Intercept	106.000	2.939	2.412	0.017
PctBlack	96.000	3.458	2.473	0.014
PctFB	116.000	2.706	2.381	0.018
TotPop90	67.000	4.456	2.566	0.011
PctEld	142.000	2.247	2.309	0.022

Figure 6.4: Bandwidths for each covariate on running an MGWR model

On running the same in GWR mode, the GWR calibration yielded an optimal bandwidth of 117 nearest neighbors implying only broad regional spatial variation in processes given 159 observations. In MGWR it can be seen that the processes vary at different spatial scales with the parameter estimates associated with the variable TotPop90 having an impact that varies over relatively short distances with the optimal bandwidth being 67 and the parameter estimates associated with the variable PctEld being global with an optimal bandwidth of 142. Apart from providing summary results for a global regression model and optimal bandwidths as discussed above, MGWR 1.0 also provides diagnostic and summary statistics for the MGWR model.

Diagnostic Information

Residual sum of squares:	45.299
Effective number of parameters (trace(S)):	15.806
Degree of freedom (n - trace(S)):	143.194
Sigma estimate:	0.562
Log-likelihood:	-125.790
AIC:	285.193
AICc:	289.432
BIC:	336.770
R2:	0.715

Model diagnostic indicators of the fitted MGWR model. For eg. AICc of MGWR =289.432 is smaller than that of the global regression model = 313.199

Summary Statistics For MGWR Parameter Estimates

Variable	Mean	STD	Min	Median	Max
Intercept	0.090	0.055	0.021	0.074	0.219
PctBlack	-0.030	0.082	-0.163	-0.029	0.126
PctFB	0.386	0.161	0.135	0.438	0.587
TotPop90	0.593	0.373	0.262	0.397	1.700
PctEld	-0.145	0.028	-0.213	-0.142	-0.094

Figure 6.5: Summary screen showing diagnostic and summary statistics

References:

Information on GWR:

Fotheringham, AS, Brunsdon, C, Charlton, M (2002): *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Wiley, 282 pages. ISBN: 978-0-471-49616-8

Information on MGWR:

A. Stewart Fotheringham, Wenbai Yang & Wei Kang (2017) Multiscale Geographically Weighted Regression (MGWR), Annals of the American Association of Geographers, 107:6, 1247-1265, DOI: 10.1080/24694452.2017.1352480