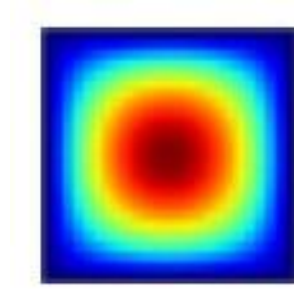# MGWR 2.2 User Manual

MGWR 2.2 Development Team

Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin

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

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

Acknowledgements

Suggested citation:

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, *8*(6), 269.

Table of Contents

1.   Introduction

MGWR 2.2 is the latest version of the MGWR application software, which can be used to calibrate geographically weighted regression (GWR) and multi-scale geographically weighted regression (MGWR) models. MGWR 2.2 offers a user-friendly, graphical interface for exploring geographically varying relationships between dependent and independent variables of interest.

MGWR 2.2 is compatible with Windows and MacOS environments. Because GWR and MGWR modeling is computationally intensive, it is recommended that the software be installed on machines with a minimum of 4GB of memory. The software will dynamically allocate memory for large matrices and automatically uses multi-processing routines to speed-up computation.

This document serves as a reference for using the MGWR 2.2 interface and offers a short tutorial for conducting GWR and MGWR analyses. Some familiarity with the basics of geographic information systems (GIS) and regression modeling is assumed. For those interested in additional functionality, the open-source python package is available via https://github.com/pysal/mgwr.

MGWR was first released in October 2018. The current version of MGWR covered in this user's manual is version 2.2 and was released in March 2020. A summary of changes can be found at the end of this document. This software is copyrighted by the development team, however, it may be freely distributed and used for academic and noncommercial purposes. Publications and research reports developed using MGWR 2.2 must clearly cite the software using the citation below. Additional information regarding geographically weighted regression modeling can be found in the References section of this manual.

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, *8*(6), 269.

2.   Getting Started

2.1   Installing MGWR 2.2

To install MGWR 2.2, navigate to https://sgsup.asu.edu/sparc/mgwr and click the "DOWNLOAD" button for the appropriate computing environment. The download will begin automatically. Once the download is complete, a zipped folder called "mgwr_2.2._for_pc.zip" for MGWR 2.2 for Windows or a zipped folder called "mgwr_2.2_macos.zip" for MGWR 2.2 for MacOS will appear in your downloads folder. For MacOS, double click on the mgwr_2.2_macos.zip to complete the installation. For Windows, double click on mgwr_2.2_for_pc.zip and then double-click on the installer

"MGWR 2.0 for Windows installer" to complete the installation. Upon successful installation, a shortcut to the program will appear on your desktop and in the MGWR 2.2 program group.

If you later decide to uninstall MGWR 2.2 from your local environment, you may use the "uninstall" option in the MGWR 2.2 program group or the "Uninstall Programs" option in the Control Panel for Windows environments. To uninstall MGWR 2.2 from a Mac environment, simply drag the program's icon to the "Trash" folder.

2.2 Launching the software

Once MGWR 2.2 is installed (see 2.1) you can open the program by double clicking the MGWR 2.2 shortcut icon on the desktop or by selecting MGWR 2.2 from the applications folder. Once opened, the MGWR 2.2 interface window will appear (Figure 1 below).
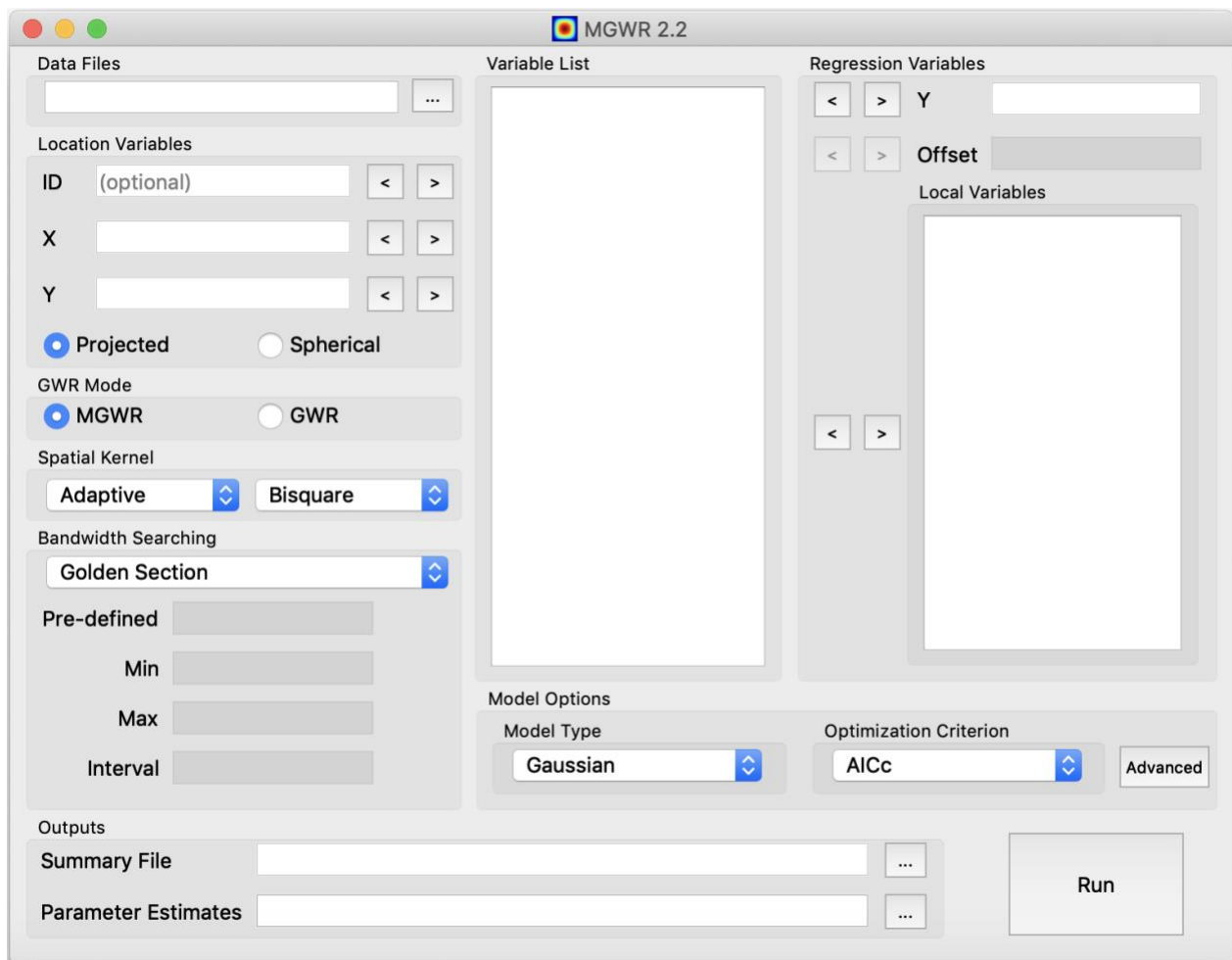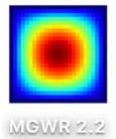
Figure 2.1: MGWR 2.2 main interface startup screen

4

3.   Interface Overview

The GUI interface is divided into 10 main components, as depicted in Figure 3.1. Components 1 through 10 are described in greater detail in the subsequent subsections.
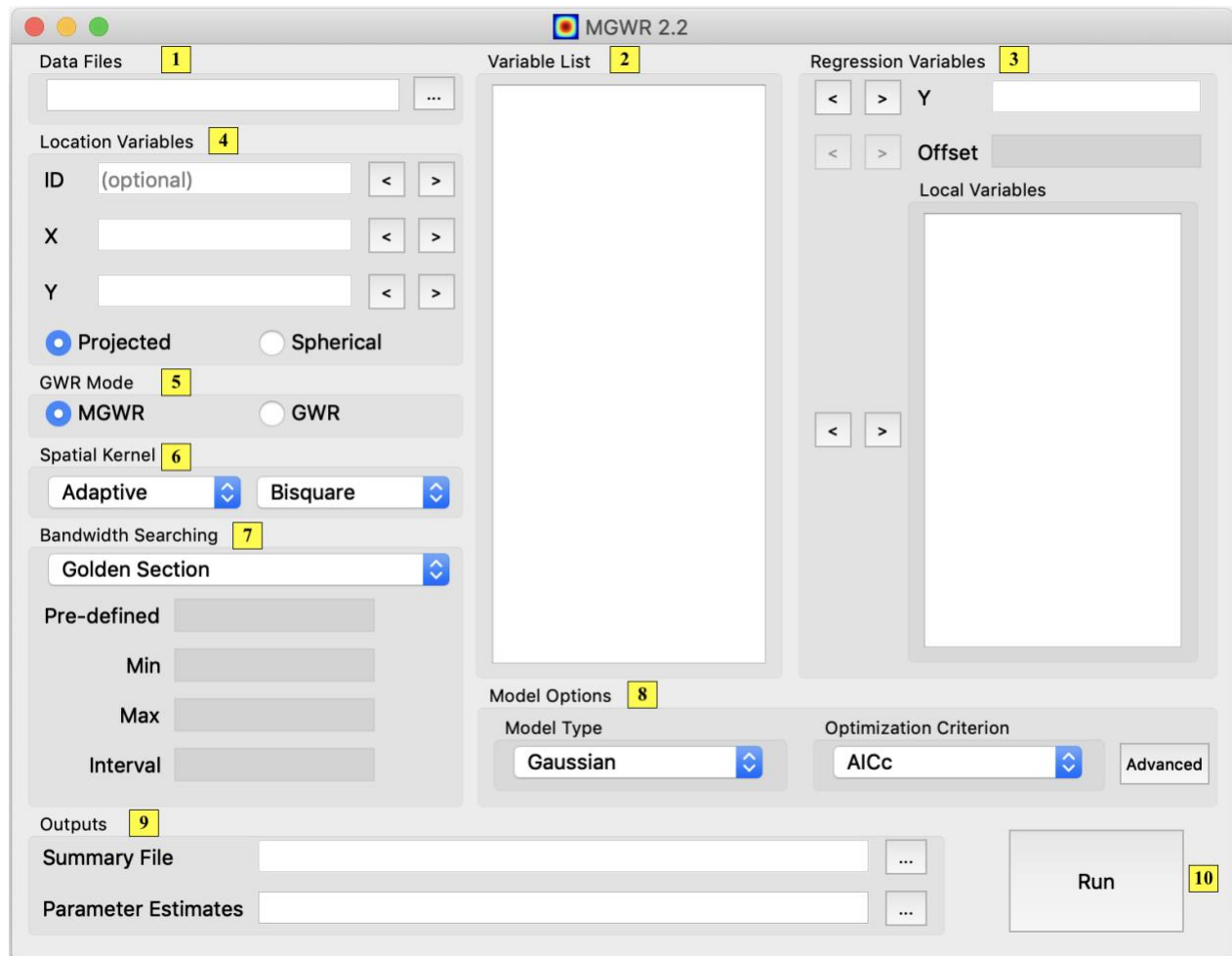
Figure 3.1: Components of MGWR 2.2 main interface

3.1   Data Files

To load a data set into the MGWR 2.2 software, click on the white text box below the "Data Files" header and type in the pathway of your data file. Alternatively, click on the icon to the right of the white text box to open a file explorer window. Navigate to the location of your data file and click "Open."
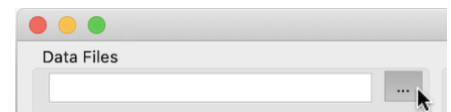
MGWR 2.2 supports the following file formats: excel (*.xls, *.xlsx), comma delimited (*.csv), and dbase IV file format (*.dbf). By default, MGWR 2.2 will treat the first row of any data file as variable names. To calibrate a GWR or MGWR model, a data file must contain the minimum following fields:

1) One field corresponding to a dependent variable
2) At least one field corresponding to an independent variable

3) One field containing the x-coordinate of each observation
4) One field containing the y-coordinate of each observation.

An optional "ID" field containing a unique identifier for each observation may also be specified. Every variable must contain numeric values except for the ID variable which may contain alphanumeric values. Relatedly, missing values must be stored as blank values rather than "NULL" or "NAN". The x- and y-coordinate fields may be in projected or spherical units; additional information on x- and y-coordinates is provided in section 3.4 below.

Note: MGWR 2.2 does not have the capability to handle missing values. Observations containing blank values in any variable will be dropped from the model fitting process. MGWR 2.2 also does not currently support space-time analysis. It is recommended to have a minimum of 300 observations for performing a GWR or MGWR analysis.

3.2   Variable List

After successfully loading a data file into MGWR 2.2, the "Variable List" menu will populate with all variables contained in the data file (Figure 3.2.1).
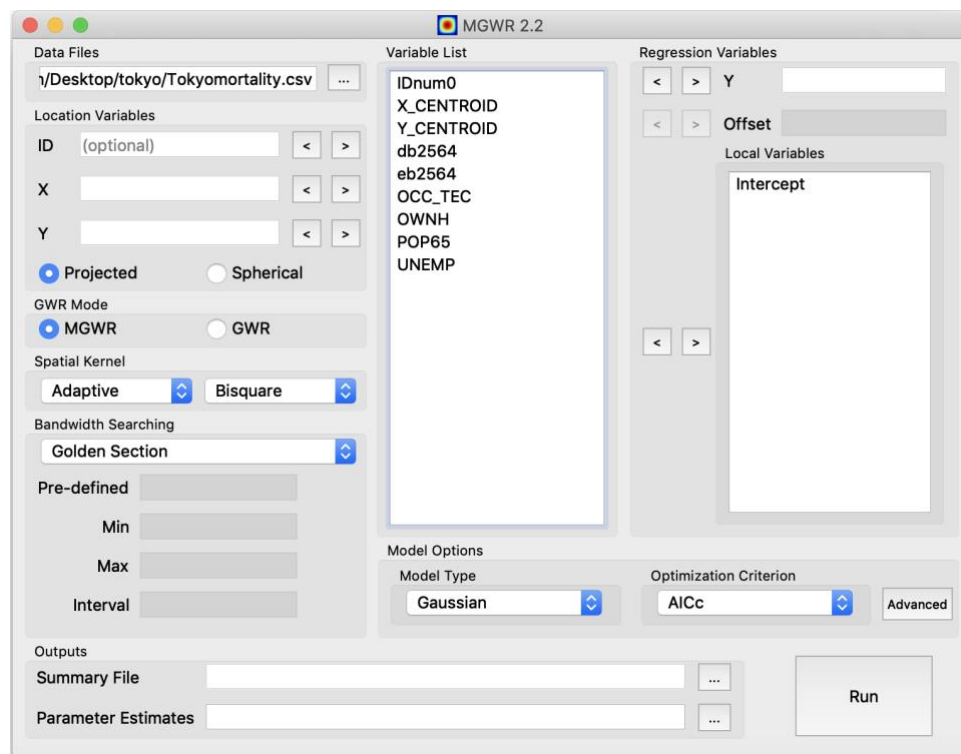
Figure 3.2.1: Variable list for Tokyo mortality data set loaded into MGWR 2.2

3.3   Regression Variables

The "Regression Variables" menu contains three options: (1) Y, (2) Offset, and (3) Local Variables. Using the forward arrow icons > , you can select a variable from the Variable List for each of these options. Each of these options is described more fully in the subsections below.

3.3.1 Selecting Y Variable

To select the Y variable, also known as the outcome or dependent variable, click on the variable name within the Variable List.
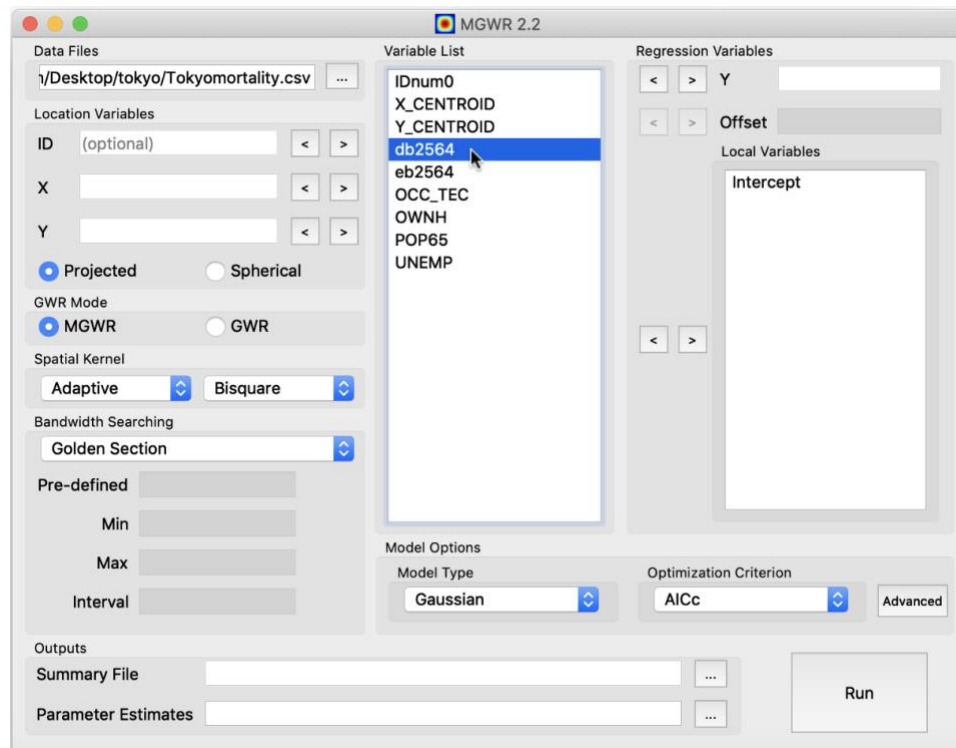
Figure 3.3.1a: Selecting variable from Variable List

Then, click on the forward arrow [>] next to the Y component.
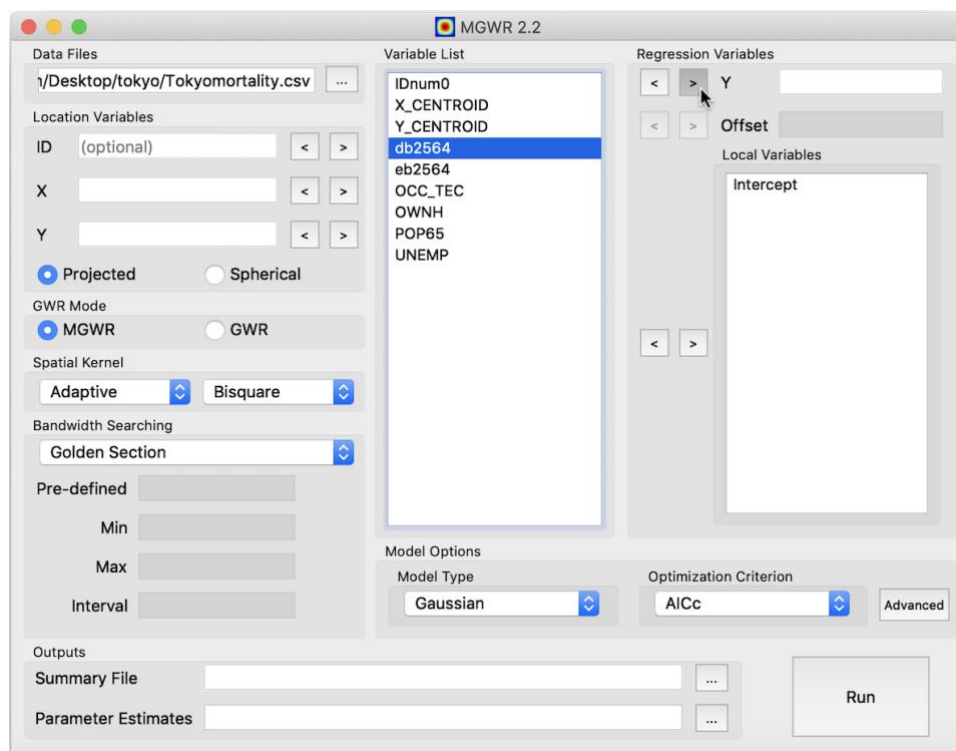
Figure 3.3.1b: Adding selected variable to Y component of the Regression Variables

The variable should now appear in the white box to the right of the Y component.
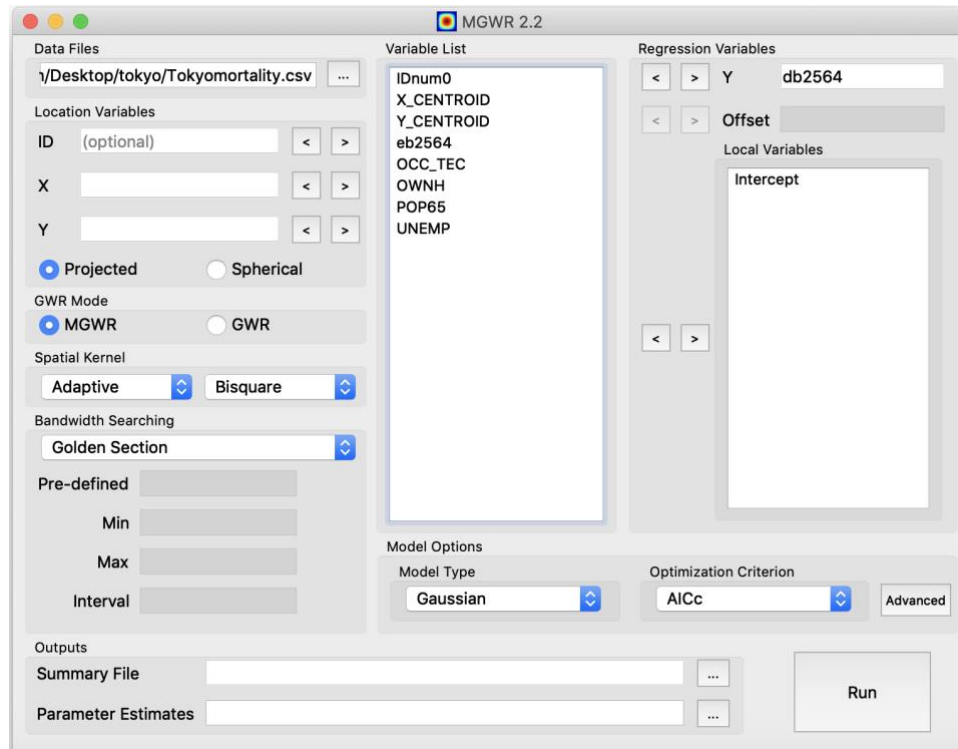
Figure 3.3.1c: Variable db2564 successfully added to Y component of Regression Variables

3.3.2 Selecting Offset

The Offset option is only available for Poisson GWR models. See section 3.8 for how to specify an offset when performing a Poisson GWR model.

3.3.3 Selecting Local Variable(s)

To select the Local variable(s), also known as the covariate(s) or independent variable(s), click on the variable name(s) within the Variable List. You can select multiple variables at once by holding down the Control key on a Windows systems or the Command key on a MacOS as you click on variable names. Then, click on the forward arrow ▶ next to the Local Variables component. The variable(s) should now appear in the white box below the Local Variables heading.

3.3.4 Removing Regression Variables

To remove a variable from the Regression Variables menu, click on the variable name and then click on the back arrow ◀. Once removed, the variable name will reappear in the Variable List menu.

3.4   Location Variables

Similar to the Regression Variables component described in Section 3.3, the "Location Variables" component contains three options: (1) ID, (2) X, and (3) Y. Each of these options can be selected using the backward arrow icon ◀. In addition to these options, there is the

8

"Projected" and "Spherical" coordinate system option. Each of these options is described more fully in the subsections below.

3.4.1 Selecting ID Variable

As mentioned earlier, a unique ID variable is not required to calibrate a GWR or MGWR model in MGWR 2.2. However, if your data file contains a unique ID field, such as a place name or Census FIPS code, specifying the variable in the ID component will retain the ID field in the model output. This will allow you to easily join your output with other information (e.g. a shapefile, other data sets, etc.) using the ID field.

To specify an ID variable, click on the variable name within the Variable List. Then, click on the back arrow < next to the ID component. The variable should now appear in the white box to the right of the ID component.

3.4.2 Selecting X and Y Coordinate Variables

In order to calibrate a GWR or MGWR, x- and y-coordinates must be specified correctly. To specify the x-coordinate variable, click on the variable name within the Variable List. Then, click on the back arrow < next to the X component. The variable should now appear in the white box to the right of the X component. Repeat the steps above for the y-coordinate to specify the Y component.

To specify the type of coordinate system your x- and y-coordinates are stored as, click on "Projected" or "Spherical" as appropriate.
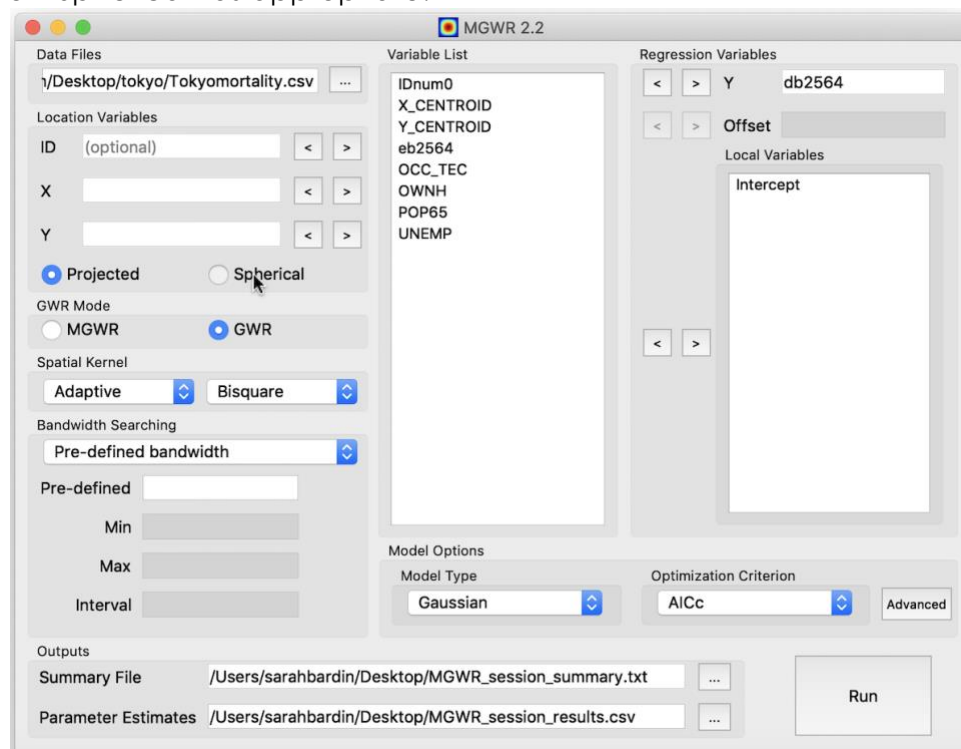
Figure 3.4.2: Selecting a spherical coordinate system within the Location Variables menu

Note: The "Projected" option should be used for coordinates that have been projected onto an orthogonal two-dimensional space. Common projected coordinate systems include UTM, State Plane, and Albers Equal Area. The "Spherical" option should be used for coordinates stored in decimal degrees, such as latitude and longitude.

3.5   GWR Mode

MGWR 2.2 allows users to calibrate both MGWR and GWR models. Unlike traditional, global regression modeling techniques, which assume that the relationships examined through the model's parameters are constant over space, both MGWR and GWR allow for the parameters to vary across space. Although both MGWR and GWR allow users to model process heterogeneity, they are based on different model formulations.

3.5.1 GWR Model Equation

For a GWR model, the linear regression model is as follows:

Assuming that there are *n* observations, for observation $i \in \{1,2,...,n\}$ at location (u_i, v_i),

$$y_i = \beta_0(u_i, v_i) + \sum_j \beta_j(u_i, v_i) x_{ij} + \varepsilon_i \quad ,$$

where $\beta_0(u_i, v_i)$ is the intercept, 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.

3.5.2 MGWR Model Equation

For a MGWR model, the linear regression model is as follows:

Assuming that there are *n* observations, for observation $i \in \{1,2,...,n\}$ at location (u_i, v_i),

$$y_i = \beta_0(u_i, v_i) + \sum_j \beta_{bwj}(u_i, v_i) x_{ij} + \varepsilon_i \quad ,$$

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

As shown in the equations above, unlike GWR which assumes that the local relationships within each model 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 that indicate the range over which data is borrowed can vary by parameter surface.

3.5.3 Specifying GWR Mode

To specify the modeling approach, click on "MGWR" or "GWR" within the GWR Mode menu, as appropriate.
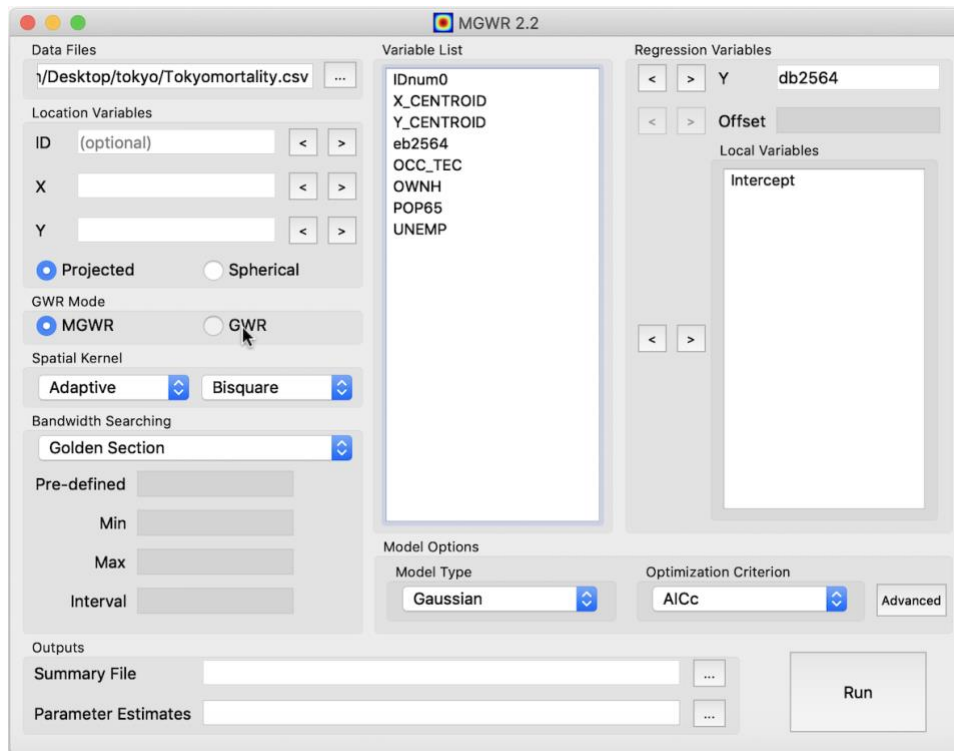
Figure 3.5.3: Selecting GWR model

NOTE: MGWR 2.2 currently supports Gaussian, Binomial, and Poisson GWR modeling and Gaussian MGWR modeling. More information regarding these model types is provided in Section 3.8 below.

3.6   Spatial Kernel

The spatial kernel menu allows users to define the weighting scheme used for calibrating a GWR or MGWR model. Users have the option of specifying an adaptive bisquare or a fixed gaussian spatial kernel. 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 to be constant over space for each point. By default, an adaptive bisquare spatial kernel will be applied. However, a fixed gaussian spatial kernel may be appropriate if your dataset contains evenly distributed points. Figure 3.6 below provides a visual illustration of the difference between a fixed gaussian spatial kernel (top panel) and an adaptive bisquare spatial kernel (bottom panel).

11
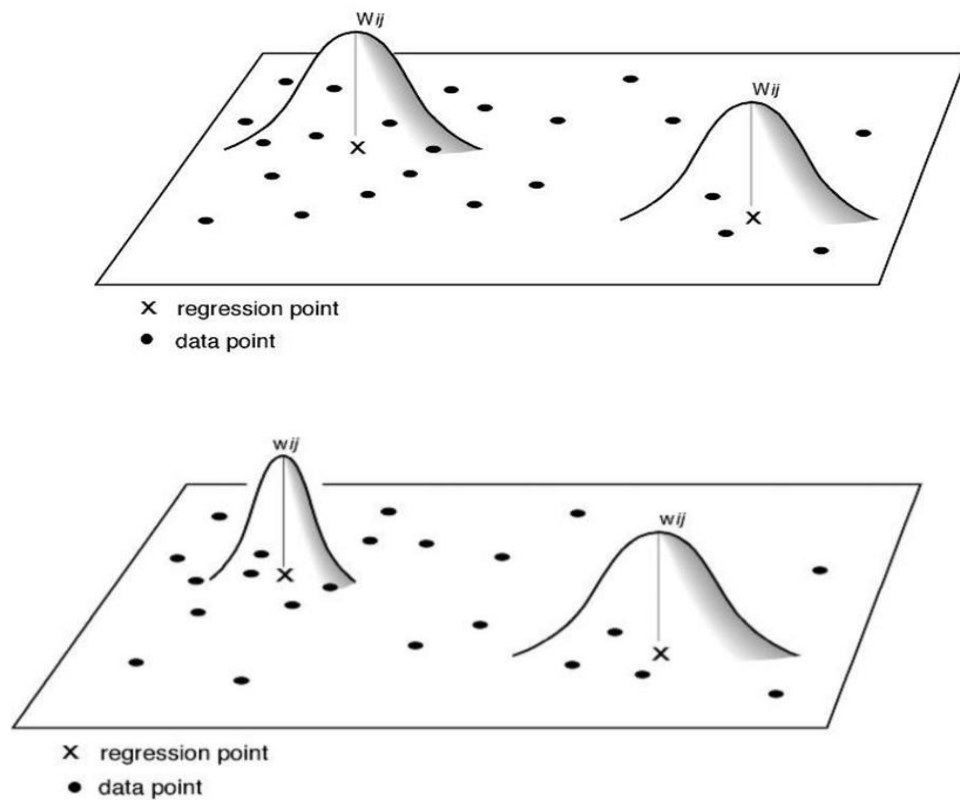
Figure 3.6a: Conceptual diagrams explaining fixed (top) and adaptive (bottom) weighting schemes

To specify a fixed gaussian spatial kernel, click on the "Adaptive" menu and select "Fixed".
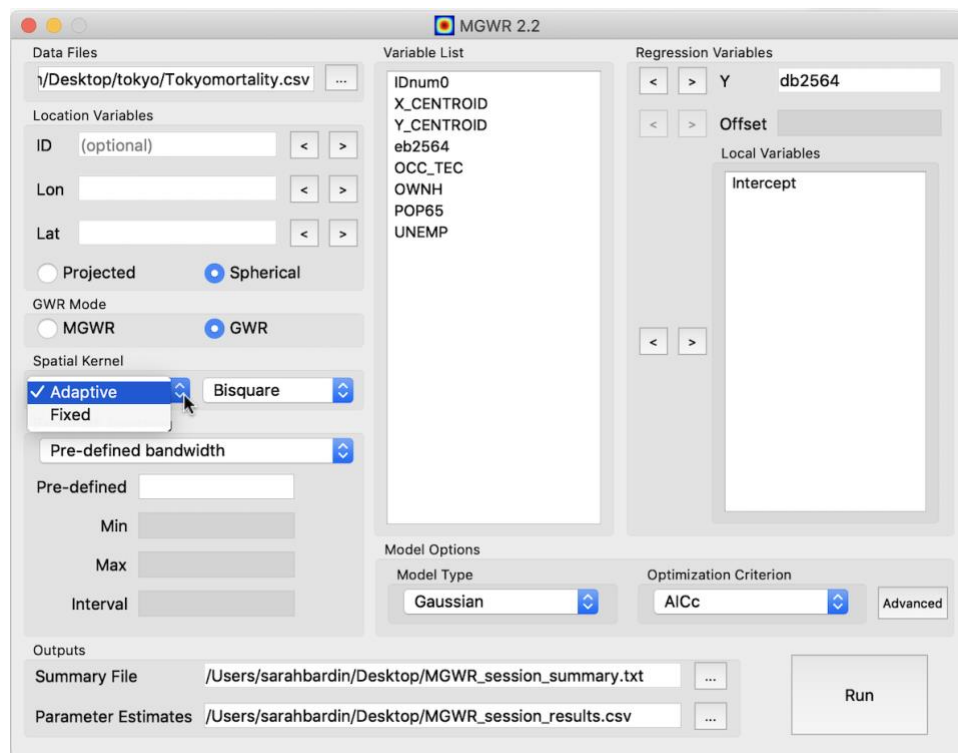
Figure 3.6b: Selecting a new spatial kernel

3.7   Bandwidth Searching

The bandwidth searching option allows users to define the algorithm used to select an optimal bandwidth when calibrating a GWR or MGWR model. By default, a Golden Section search will be used. The Golden Section search 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. Alternatively, users may specify an Interval search, which requires the user to define a limited range (min and max) and step value (interval) within which the model will search and successively increment over until an optimal bandwidth is found. For GWR models, users may also specify a pre-defined bandwidth, which may be useful for exploratory modeling.

To specify an Interval Search or Pre-defined bandwidth, click on the "Golden Section" menu and select the appropriate option.
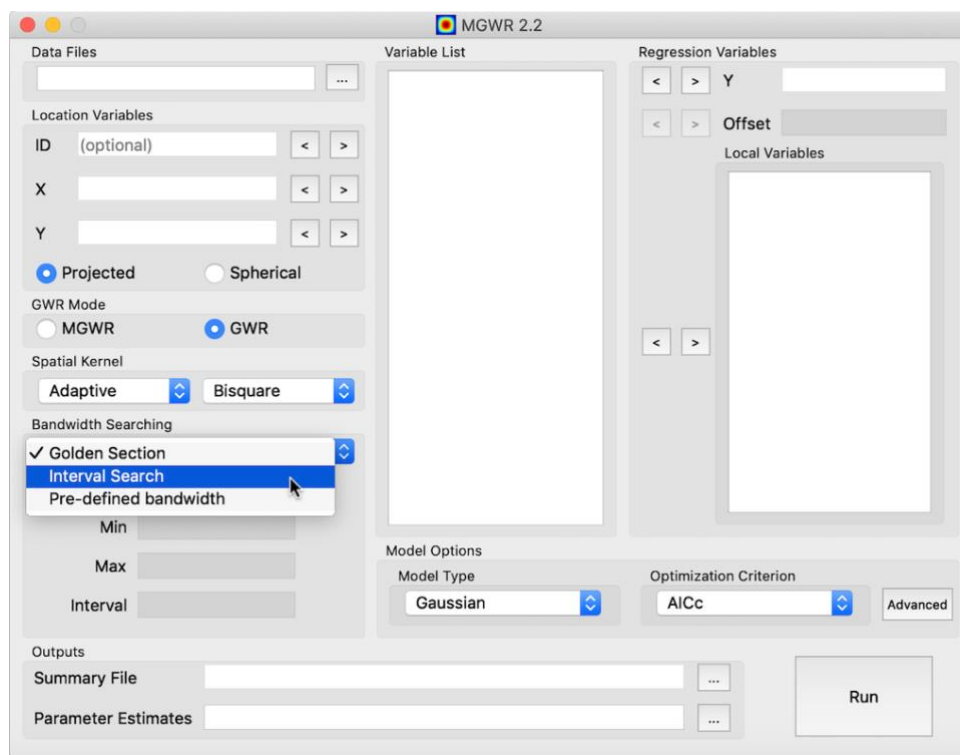
Figure 3.7a: Selecting the Bandwidth Searching Procedure

If Interval Search is selected, click on the white text box next to "Min", "Max", and "Interval" to define the range and step value for the search criteria.
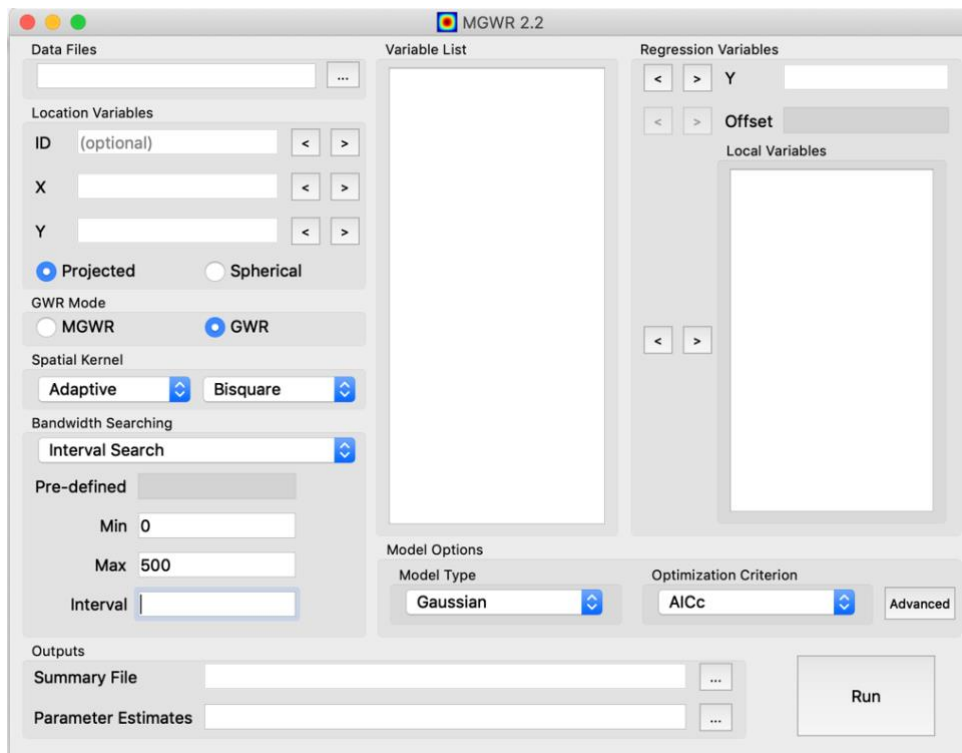
Figure 3.7b: Specifying Min, Max, and Interval for Interval Search Bandwidth Searching Procedure

If Pre-defined bandwidth is selected, click on the white text box next to "Pre-defined" to define the bandwidth value for the model.
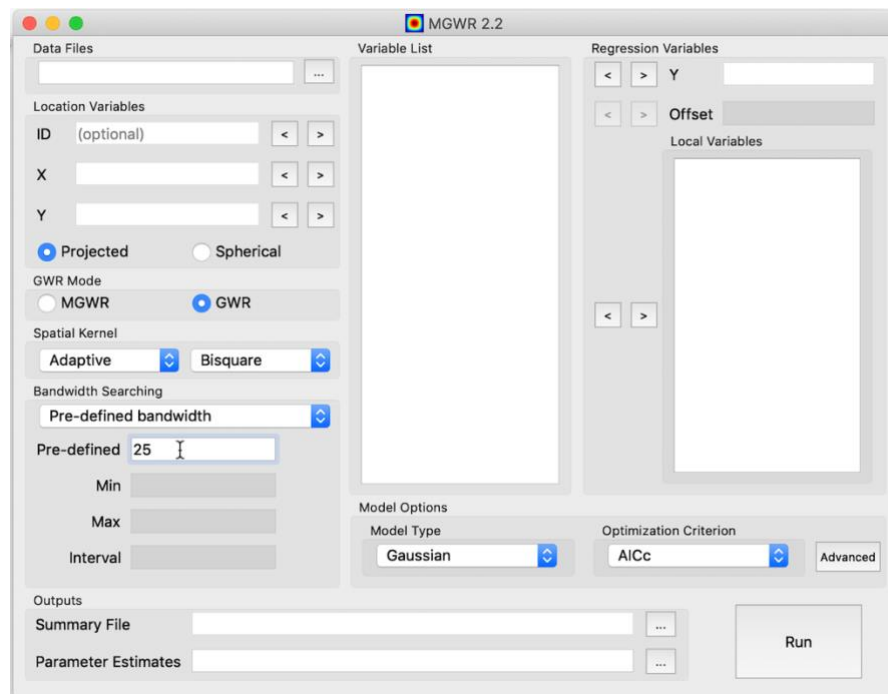
Figure 3.7c: Specifying Pre-defined search area for Pre-defined Bandwidth Searching Procedure

3.8   Model Options

The Model Options component allows users to define the Model Type used for calibrating the model, along with the Optimization Criterion, and also offers Advanced options. The following subsections describe each of these options more fully.

3.8.1 Model Type

MGWR 2.2 allows users to specify one of three model types for calibrating GWR models. These model types are (1) Gaussian, (2) Binomial, and (3) Poisson. Only a Gaussian model type is currently available for MGWR modeling. By default, a Gaussian model type will be used. However, a Binomial model type may be more appropriate if the outcome of interest is a binary variable (e.g. 1/0). Similarly, a Poisson model type may be more appropriate if the outcome of interest is a count variable (e.g. number of crimes, number of illnesses, etc).

To change the Model Type for GWR modeling, click on the "Gaussian" menu and select the appropriate option.
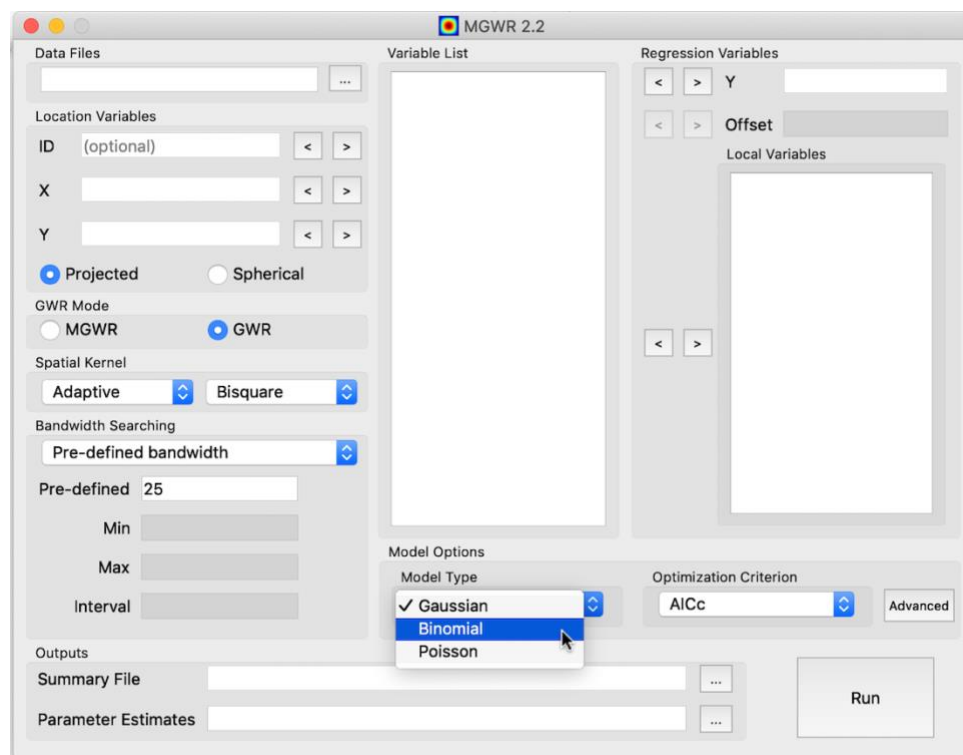
Figure 3.8.1: Specifying Model Type

For Poisson GWR models, MGWR 2.2 allows users to specify an offset variable. In spatial epidemiology, the offset variable typically corresponds to the expected size of the outcome or the size of the population at risk. In Poisson GWR models, if no offset variable is selected, the value of 1 will be used for all locations.

To specify an offset, click on the variable name within the Variable List. Then, click on the forward arrow ⟩ next to the Offset component. The variable should now appear in the white box to the right of the Offset component.

3.8.2 Optimization Criterion

In MGWR 2.2 users may specify the optimization criterion used for calibrating both MGWR and GWR models. By default, a Corrected Akaike Information Criterion (AICc) will be used. However, an Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or Cross Validation Score (CV) may be used instead.

To specify the Optimization Criterion, click on the "AICc" menu and select the appropriate option.
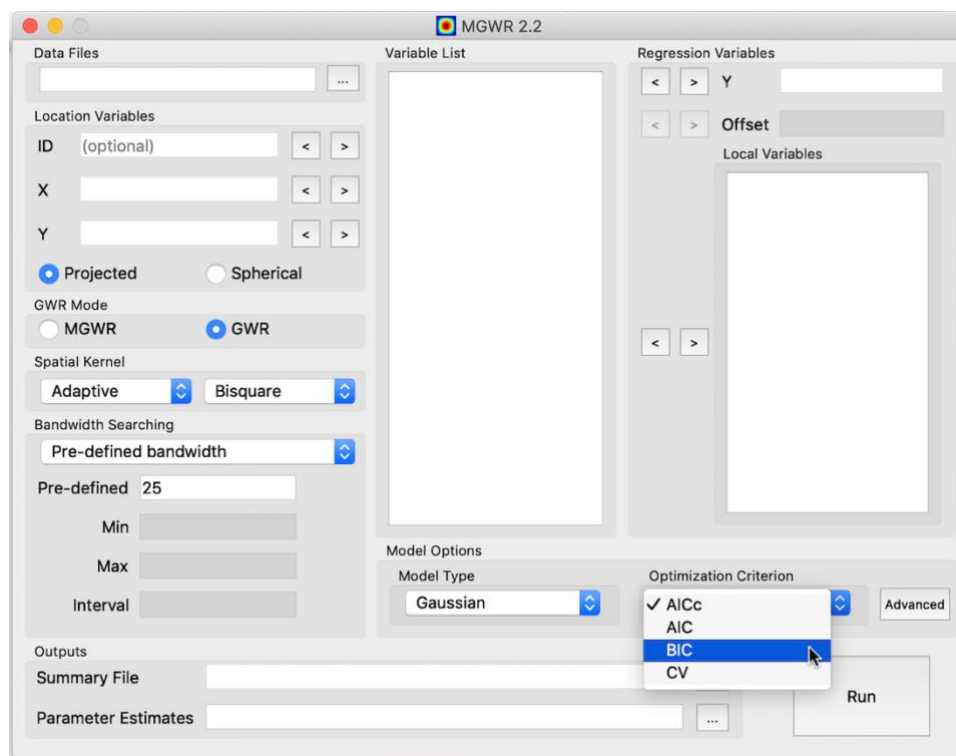
Figure 3.8.2: Specifying Optimization Criterion

3.8.3 Advanced

MGWR 2.2 also provides advanced menu options related to variable standardization, Monte Carlo testing for spatial variability, local collinearity diagnostics, and bandwidth confidence intervals. For MGWR models, users may also change the measure of score change and convergence threshold. By default, only the Variable Standardization advanced option will be turned on. The following sections provide greater detail regarding each of these advanced options.

16

1. **Variable standardization:** This option performs a z-transformation on the Y (dependent) and local (independent) variables so that each variable has a mean of 0 and standard deviation of 1. By default, this option is 'On' as standardization makes iterative computation of model fitting faster in most cases. For MGWR modeling, it is recommended to always turn the variable standardization on, as doing so, will allow for the interpretation and comparison of the individual bandwidths. With standardization, the bandwidths serve as direct indicators of the spatial scale at which the conditional relationship between the outcome (Y) variable and the independent (local) variable varies. Without standardization, the bandwidths also reflect the scale and variation within each predictor variable.

2. **Monte Carlo test for spatial variability:** To test whether the spatial variability of the local estimates is attributable to sampling variation versus a result of other inherent processes, MGWR 2.2 provides the option to run a Monte Carlo test. The Monte Carlo test runs once to derive the local parameter estimates and then repeatedly derives new local parameter estimates after randomly rearranging the data points to measure whether the variability of each parameter surface could have arisen by chance. **NOTE:** Monte Carlo tests are computationally intensive and will significantly increase the model run time. Additional information can be found in Li and Fotheringham (2020).

3. **Local collinearity diagnostics:** When turned on, the Local collinearity diagnostics will return a local condition index called a 'local_CN,' which identifies the number of near dependencies among the columns of the design matrix. In addition, the Local collinearity diagnostics 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. Additional information can be found in Oshan et al. (2019)

4. **Bandwidth confidence interval:** When turned on, the Bandwidth confidence interval will provide the lower and upper bounds corresponding to the 95th confidence interval for each bandwidth in a MGWR and GWR model. Additional information can be found in Li et al. (2020).

5. **Measure of Score of Change (SOC):** Only available for MGWR models, users may change the measure of score of change, which is the termination criterion for the iterations to be deemed to have converged. Two types of SOC can be used: (1) SOC-RSS which is the proportional change in the residual sum of squares (RSS) and (2) SOC-f which is the change in the GWR smoother. Both of these options are scale-free but the SOC-f option has the advantage of focusing on relative

changes in the additive terms rather than on the overall model fit, however it also may take longer to converge in some cases. By default, SOC-f is chosen.

6. **Convergence Threshold:** Only available for MGWR models, users may change the convergence threshold, which determines the threshold of change below which the model is deemed to have converged. By default, 10_{-5} (1e-5) is selected.

To specify any advanced options, click on the "Advanced" icon.
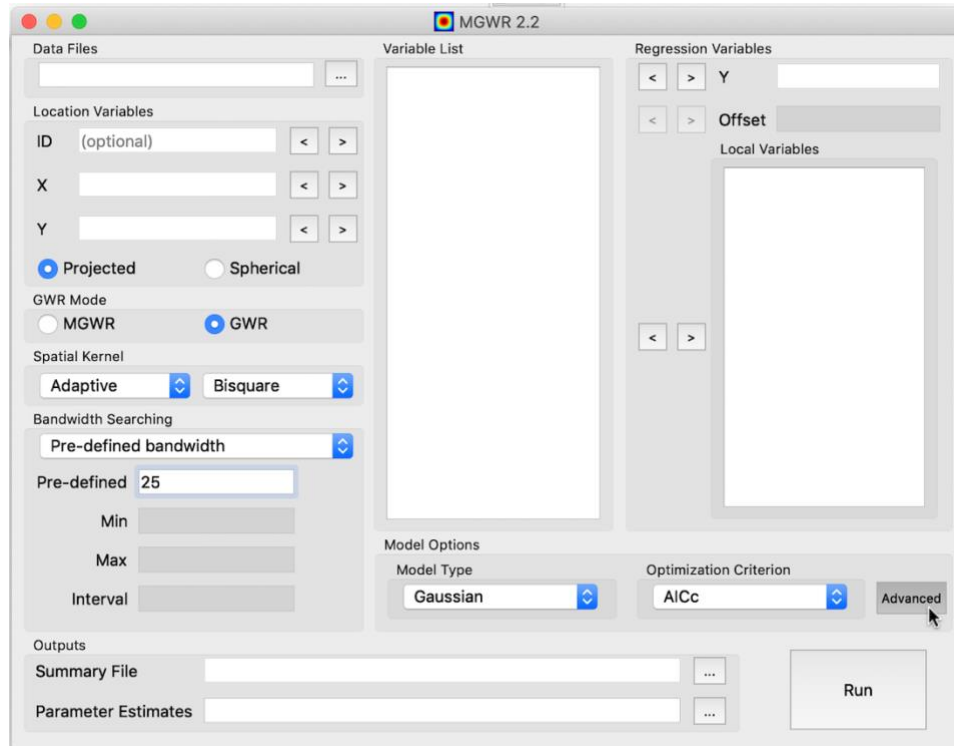
Figure 3.8.3a: Specifying Advanced options

An Advanced option menu will appear in a new window. Click on the option of interest and select "On" to turn the option on or select "Off" to turn off the option.
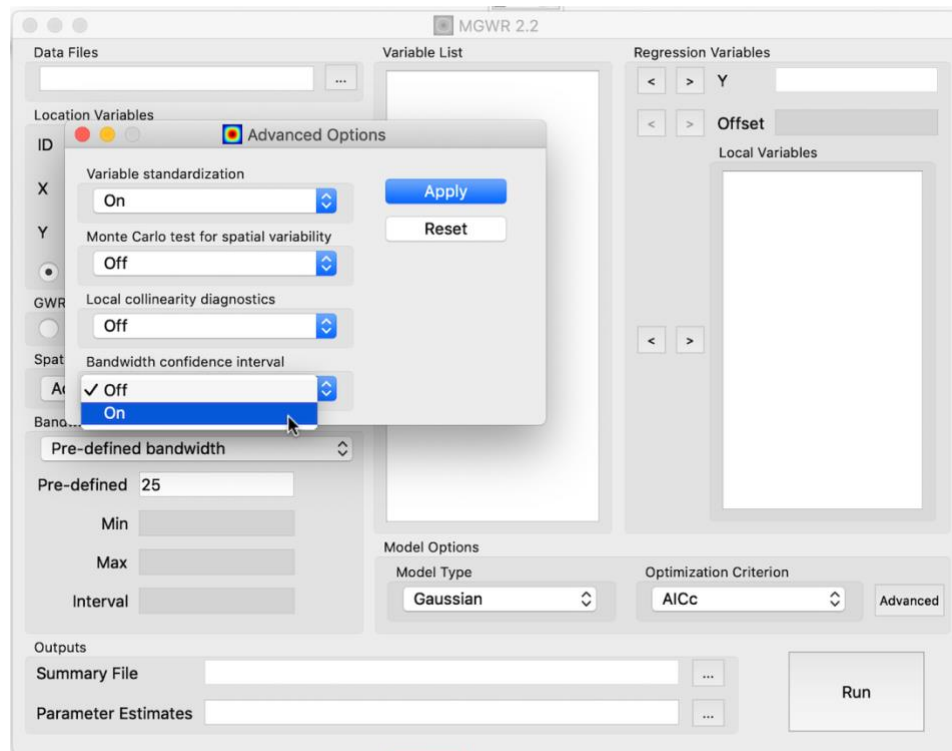
Figure 3.8.3b: Modifying Advanced options for GWR model

When done, click "Apply" to apply the options to your model.
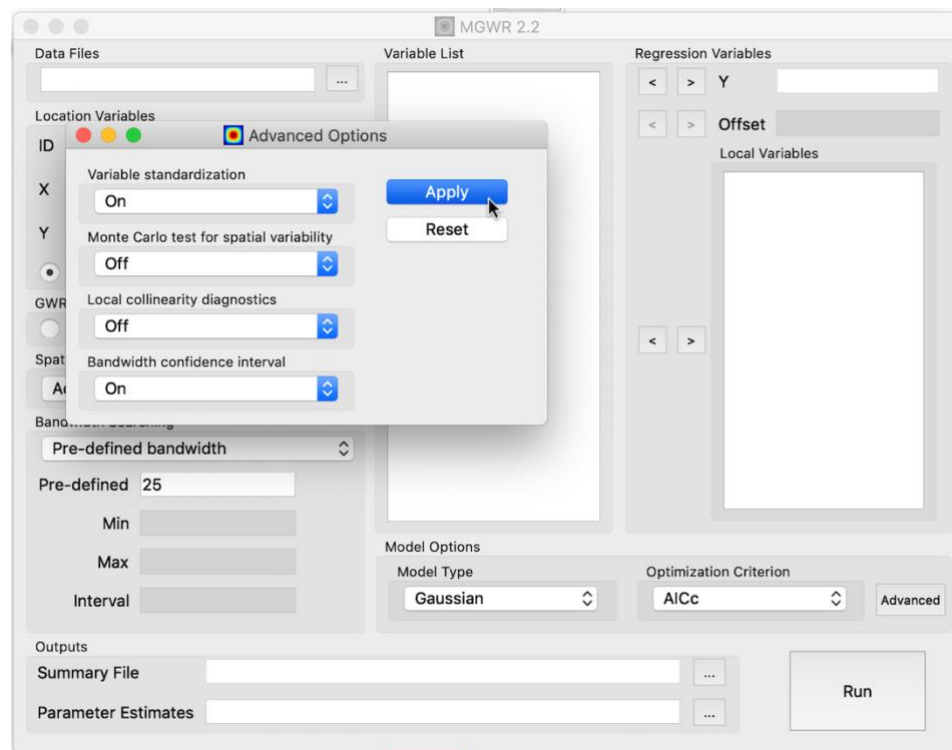
Figure 3.8.3c: Applying Advanced options for GWR model

3.9   Outputs

The Outputs menu specifies where the model results should be saved. MGWR 2.2 generates two output files: (1) a summary file and (2) a parameter estimates file. The summary file (MGWR_session_summary.txt) contains the results from the global regression model, the MGWR or GWR model, and summary statistics associated with the MGWR or GWR model. The parameter estimates file (MGWR_session_betas.csv) contains the parameter estimates for each location and the corresponding t-values. The parameter estimates can be easily plotted using any standard mapping package (e.g. ArcMAP, QGIS, etc) or with mapping packages within Python and R (e.g. geopandas, sf, etc).

To specify where the summary and parameter estimate output should be saved, click on the ⋯ icon next to Summary File option, navigate to the folder location of choice, and click save.

3.10   Run

To run the model, simply click the "Run" button in the bottom, righthand corner of the MGWR 2.2 interface. Once the model is finished running, a screen will appear indicating that the model has successfully completed running (see Figure 3.10 below). Click ok. The output files will automatically save to the location specified by the user.
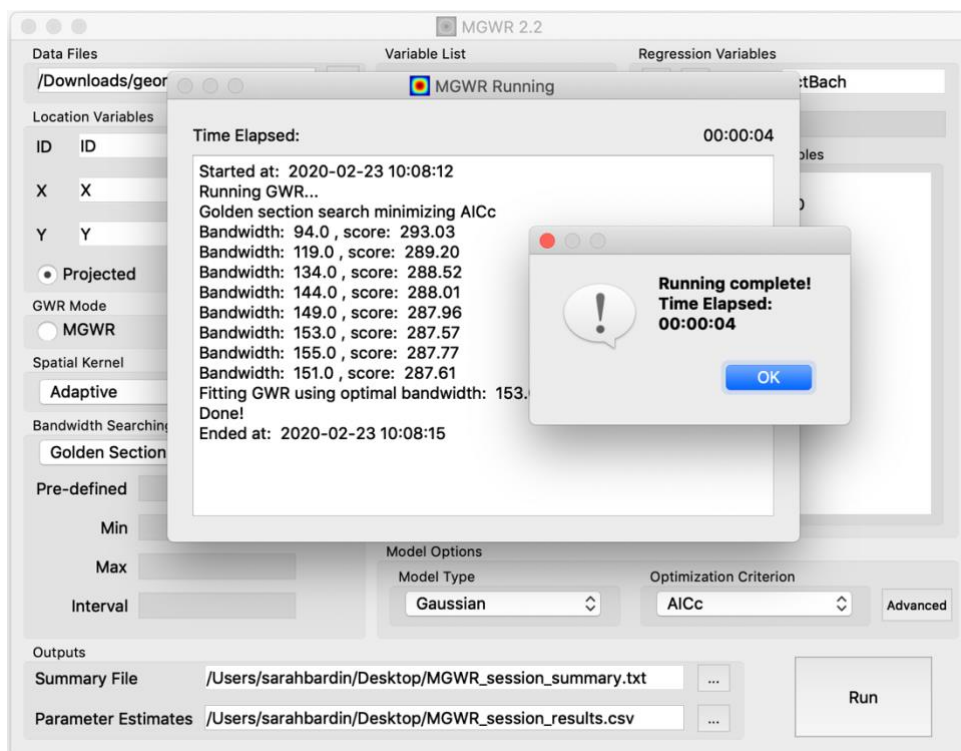
Figure 3.10a: Running of MGWR 2.2

The summary file from the output (see below) opens automatically upon successful completion of the program. Run times will vary depending on the number of observations

20

in the data set and number of predictors included in the model (see Li and Fotheringham, 2020 for more information).
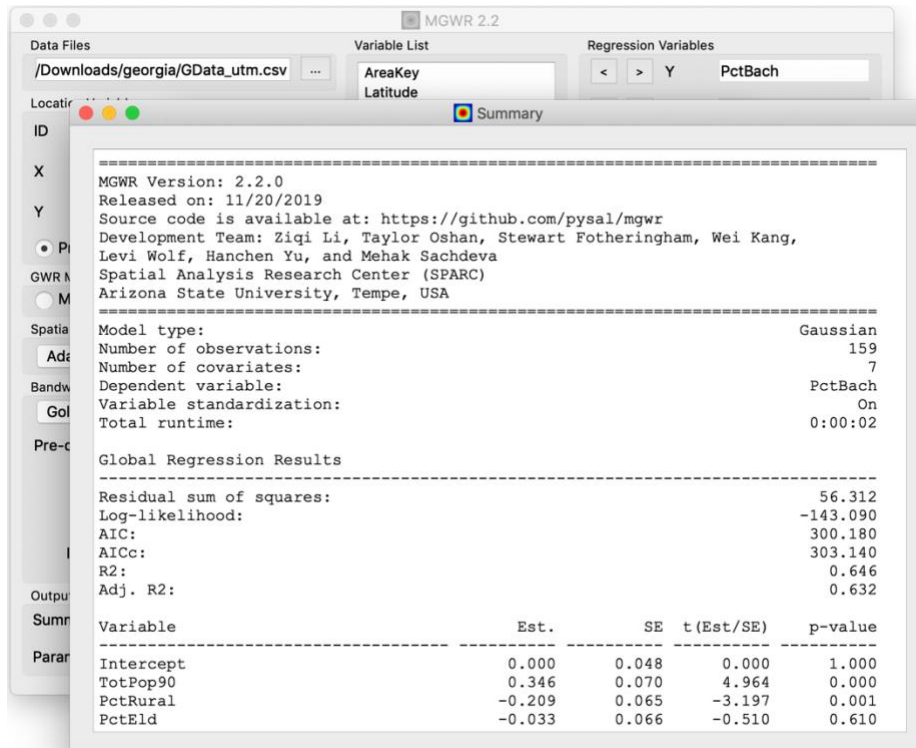
Figure 3.10b: Output summary file

4. GWR Tutorial with Sample Data Sets

The following GWR tutorials provide guidance on running Gaussian GWR, Poisson GWR, and Binomial GWR models. For each of these tutorials, we will use the sample data sets available from the [program installation page](#). After downloading the data files, unzip the files to reveal their contents.

Disclaimer: All sample data sets are for tutorial purposes only. To perform MGWR or GWR analysis it is generally recommended to have a minimum of 300 observations.

4.1 Gaussian GWR model

For this tutorial, we will use the Georgia sample data (GData_utm.csv) to estimate the relationship between the percentage of residents with at least a bachelor's degree and various county-level characteristics. The Georgia sample data contains the following variables:

AreaKey   – Census FIPS code
Latitude   – spherical y-coordinate
Longitude – spherical x-coordinate
TotPop90 – total population in 1990

PctRural   – percentage of rural population

PctBach   – percentage of residents with at least a bachelor's degree

PctEld   – percentage of elderly

PctFB   – percentage of foreign-born inhabitants

PctPov   – percentage of inhabitants living below the poverty level

PctBlack   – percentage of African-Americans

ID   – identifier

X   – projected x-coordinate

Y   – projected y-coordinate

For our model, we will specify a Gaussian GWR model, which takes the following general form:

$$y_i = \beta_0(u_i, v_i) + \sum_j \beta_k (u_i, v_i)\, x_{ij} + \varepsilon_i,$$

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 $\beta_0 (u_i, v_i)$ becomes a geographically varying "intercept" term.

For our analysis, we will fit the following Gaussian GWR model using the Georgia sample data:

$$PctBatch_i \;=\; \beta_0\,(X_i, Y_i) + \beta_1\,(X_i, Y_i)\,PctRural_i + \beta_2\,(X_i, Y_i)\,PctPov_i + \beta_3\,(X_i, Y_i)\,PctBlack_i$$
$$+\, \beta_4\,(X_i, Y_i)\,PctFB_i + \beta_5\,(X_i, Y_i)\,PctEld_i + \beta_7\,(X_i, Y_i)\,PctTotPop90_i + \varepsilon_i$$

where X_i and Y_i are the projected x-y coordinates.

After loading the GData_utm.csv file into MGWR 2.2, we will specify the following options:
1. Regression Variables
   a. Y = PctBach
   b. Local = Intercept, TotPop90, PctRural, PctEld, PctFB, PctPov, PctBlack
2. Location Variables
   a. ID = ID
   b. X = X
   c. Y = Y
   d. Projected
3. GWR Mode = GWR

We will keep all other model defaults, such that the spatial kernel is an adaptive bisquare, the bandwidth searching uses Golden Section, the model type is Gaussian, and the Optimization Criterion is AICc. After making these selections, the MGWR 2.2 interface should look like Figure 4.1a.

Note: Variable standardization by default is turned on, however, users may want to turn off this option to improve interpretability of coefficients.

Figure 4.1a: GWR Gaussian model specification in MGWR 2.2

Once the model is finished running, a "Summary" text box will appear. The summary includes the model settings, global model results, best bandwidth, and the model diagnostic information of the GWR model. Figure 4.1b describes the summary output in more detail. These results are also saved to the "MGWR_session_summary.txt" file, which is saved to the location specified in the "Summary File" section in the Outputs menu.

The parameter estimates for each location generated by the GWR model are saved to the "MGWR_session_beta.csv" file, which is saved to the location specified in the "Parameter Estimates" section in the Outputs menu. Figure 4.1c describes the parameter estimate output in more detail. The GWR parameter estimates can be joined with shapefile "G_utm.shp" using the ID field. The G_utm.shp can be found in the Georgia data folder, where the GData_utm.csv is located. After joining the parameter

23

results to the shapefile, using a mapping software, such as ArcGIS, QGIS, R, or python, the parameter estimates can be mapped to visualize the relationships found.

```
===============================================================================
Model type:                                                         Gaussian
Number of observations:                                                  159
Number of covariates:                                                      7
Dependent variable:                                                  PctBach
Variable standardization:                                                 On
Total runtime:                                                       0:00:02

Global Regression Results
-------------------------------------------------------------------------------
Residual sum of squares:                                              56.312
Log-likelihood:                                                     -143.090
AIC:                                                                 300.180
AICc:                                                                303.140
R2:                                                                    0.646
Adj. R2:                                                               0.632

Variable                            Est.         SE   t(Est/SE)     p-value
---------------------------------- ---------- ---------- ---------- ----------
Intercept                          0.000      0.048      0.000      1.000
TotPop90                           0.346      0.070      4.964      0.000
PctRural                          -0.209      0.065     -3.197      0.001
PctEld                            -0.033      0.066     -0.510      0.610
PctFB                              0.273      0.067      4.055      0.000
PctPov                            -0.198      0.090     -2.208      0.027
PctBlack                           0.067      0.077      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:                                             46.641
Effective number of parameters (trace(S)):                           13.178
Degree of freedom (n - trace(S)):                                   145.822
Sigma estimate:                                                       0.566
Log-likelihood:                                                    -128.111
Degree of Dependency (DoD):                                          0.875
AIC:                                                                284.578
AICc:                                                               287.570
BIC:                                                                328.088
R2:                                                                   0.707
Adj. R2:                                                              0.680
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              0.025      0.056     -0.062      0.035      0.114
TotPop90               0.316      0.062      0.212      0.320      0.400
PctRural              -0.194      0.052     -0.287     -0.183     -0.122
PctEld                -0.093      0.026     -0.156     -0.093     -0.041
PctFB                  0.319      0.150      0.108      0.323      0.532
PctPov                -0.121      0.087     -0.260     -0.113      0.003
PctBlack               0.027      0.097     -0.116      0.006      0.227
```

Annotations at right:
- GWR model specifications
- Model fitting results for the traditional, global regression model
- Optimal bandwidth based on Golden Section search
- Model diagnostic indicators for the fitted GWR model.
- Summary statistics of estimated coefficients of local terms

Figure 4.1b: Summary description for the Gaussian GWR model

	A	B	C	D	E	F	G	H	I	J
1	ID	x_coor	y_coor	y	ols_residual	gwr_yhat	gwr_residual	localR2	influ	CooksD
2	133	941396.6	3521764	-0.483734	-0.192034	-0.311214	-0.17252	0.6719628	0.0232564	0.0001721
3	158	895553	3471916	-0.800686	-0.341841	-0.657209	-0.143477	0.6743307	0.0669405	0.0003755
4	146	930946.4	3502787	-0.76547	-0.341591	-0.409146	-0.356324	0.672913	0.0760371	0.0026829
5	155	745398.6	3474765	-0.272433	0.4090737	-0.522752	0.2503188	0.6992132	0.0645606	0.0010968
6	79	849431.3	3665553	0.4142971	-0.000991	0.4524657	-0.038169	0.7062185	0.0761202	3.08E-05
7	23	819317.3	3807616	-0.800686	-0.297062	-0.385685	-0.415002	0.7117908	0.0525301	0.002391
8	33	803747.1	3769623	-0.30765	-0.335884	0.1137518	-0.421401	0.7155012	0.0260067	0.001155

	K	L	M	N	O	P	Q
1	beta_Interce	beta_TotPop	beta_PctRur	beta_PctEld	beta_PctFB	beta_PctPov	beta_PctBlac
2	-0.030366	0.3704693	-0.26906	-0.062979	0.1730151	-0.207812	0.0970875
3	-0.044993	0.3879884	-0.26178	-0.067327	0.1381983	-0.225837	0.1398587
4	-0.035393	0.3772778	-0.267567	-0.063676	0.1596365	-0.214675	0.1113441
5	-0.050935	0.3947135	-0.21344	-0.094465	0.1238974	-0.239929	0.1907754
6	0.1047703	0.2561338	-0.175011	-0.155606	0.4245419	-0.018631	-0.105606
7	0.08977	0.2305019	-0.158606	-0.098096	0.5205631	-0.006573	-0.074103
8	0.0891896	0.2396111	-0.154122	-0.102589	0.5059619	-0.008926	-0.074128

	R	S	T	U	V	W	X
1	se_Intercept	se_TotPop90	se_PctRural	se_PctEld	se_PctFB	se_PctPov	se_PctBlack
2	0.0553191	0.0752916	0.0692626	0.0834347	0.0753976	0.0919666	0.0860727
3	0.0560422	0.073245	0.0689967	0.0851285	0.0738496	0.0925595	0.0867434
4	0.0553953	0.0744664	0.069027	0.0837037	0.0743725	0.091981	0.0860879
5	0.0573902	0.0711572	0.0726918	0.0942966	0.0763779	0.0984034	0.0909665
6	0.0547478	0.0713126	0.0767649	0.0852318	0.0844975	0.0983172	0.0865856
7	0.0552747	0.0720641	0.0753709	0.0742281	0.0839356	0.1041738	0.0846029
8	0.05457	0.0713826	0.0745255	0.0739118	0.0825325	0.1032321	0.0839805

	Y	Z	AA	AB	AC	AD	AE
1	t_Intercept	t_TotPop90	t_PctRural	t_PctEld	t_PctFB	t_PctPov	t_PctBlack
2	-0.548918	4.9204631	-3.884632	-0.754826	2.2947035	-2.259649	1.1279709
3	-0.80285	5.2971306	-3.794102	-0.790889	1.8713481	-2.439916	1.6123274
4	-0.638916	5.0664192	-3.876269	-0.760732	2.1464471	-2.333904	1.2933777
5	-0.887529	5.5470629	-2.936229	-1.001791	1.6221623	-2.43822	2.0972048
6	1.9136898	3.5917032	-2.279837	-1.825685	5.0243135	-0.189499	-1.21967
7	1.6240694	3.1985702	-2.104346	-1.321544	6.2019323	-0.0631	-0.875889
8	1.6344058	3.3567163	-2.068039	-1.387992	6.1304533	-0.086466	-0.882683

	AF	AG	AH	AI	AJ	AK	AL
1	p_Intercept	p_TotPop90	p_PctRural	p_PctEld	p_PctFB	p_PctPov	p_PctBlack
2	0.5838364	2.15E-06	0.0001504	0.4514771	0.0230664	0.0252099	0.2610421
3	0.4232673	3.89E-07	0.0002107	0.4301941	0.0631457	0.0157967	0.1088864
4	0.5238028	1.12E-06	0.0001552	0.4479511	0.0333635	0.0208601	0.1977682
5	0.3761434	1.20E-07	0.0038188	0.3179761	0.1067623	0.0158682	0.0375677
6	0.0574679	0.0004382	0.0239551	0.0697858	1.35E-06	0.8499452	0.2244077
7	0.1063543	0.0016689	0.0369311	0.1882304	4.68E-09	0.949767	0.3824199
8	0.1041646	0.0009881	0.0402665	0.1670936	6.73E-09	0.9312057	0.3787486

Figure 4.1c: Parameter estimate description for the Gaussian GWR model

4.2   Poisson GWR model

For this tutorial, we will use the Tokyo sample data (Tokyomortality.csv) to estimate the relationship between the premature mortality and various characteristics, including the population over 65 years of age, the population with a professional occupation, home ownership, and unemployment. The Tokyo sample data contains the following variables:

IDnum0	– Census FIPS code
X_CENTROID	– projected x-coordinate
Y_CENTROID	– projected y-coordinate
db2564	premature mortality count
eb2564	total mortality county
OCC_TEC	percentage of population with professional occupation
OWNH	percentage of home ownership
POP65	percentage of population over 65 years of age
UNEMP	unemployment rate

For our model, we will specify a Poisson GWR model. Because the outcome variable of interest, premature mortality count, is a count variable, a Poisson regression is better suited for this analysis. The general form for a Poisson GWR model is as follows:

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

For a Poisson GWR model, 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. In spatial epidemiology, the offset term is often the size of the population at risk or the expected size of the outcome. In cases where the "offset variable" option is left blank, N_i becomes 1.0 for all locations.

After loading the Tokyomortality.csv file into MGWR 2.2, we will specify the following options:
1. Regression Variables
   a. Y = db2564
   b. Offset = eb2564
   c. Local = Intercept, POP65, OCC_TEC, OWNH, UNEMP
2. Location Variables
   a. ID = IDnum0
   b. X = X_CENTROID
   c. Y = Y_CENTROID
   d. Projected
3. GWR Mode = GWR
4. Model Type = Poisson

We will keep all other model defaults, such that the spatial kernel is an adaptive bisquare, the bandwidth searching uses Golden Section, and the Optimization Criterion is AICc. After making these selections, the MGWR 2.2 interface should look like Figure 4.2a:

Note: Variable standardization by default is turned on, however, users may want to turn off this option to improve interpretability of coefficients.

Figure 4.2a: GWR Poisson model specification in MGWR 2.2

Once the model is finished running, a "Summary" text box will appear. The summary includes the model settings, global model results, best bandwidth, and the model diagnostic information of the GWR model. Figure 4.2b describes the summary output in more detail. These results are also saved to the "MGWR_session_summary.txt" file, which is saved to the location specified in the "Summary File" section in the Outputs menu.

27

```
===============================================================================
Model type:                                                            Poisson
Number of observations:                                                    262
Number of covariates:                                                        5
Dependent variable:                                                     db2564
Offset variable:                                                        eb2564
Variable standardization:                                                   On
Total runtime:                                                          0:00:00

Global Regression Results
-------------------------------------------------------------------------------
Deviance:                                                            12029.202
Log-likelihood:                                                        101.600
AIC:                                                                  -193.200
AICc:                                                                12039.437
Percent deviance explained:                                             -0.016
Adj. percent deviance explained:                                        -0.032

Variable                          Est.          SE    t(Est/SE)      p-value
------------------------------ ---------- ---------- ---------- ----------
Intercept                      -35.044      0.470      -74.603        0.000
OCC_TEC                          0.000      0.215        0.000        1.000
OWNH                            -0.000      0.407       -0.000        1.000
POP65                            0.000      0.301        0.000        1.000
UNEMP                           -0.000      0.197       -0.000        1.000

Geographically Weighted Regression (GWR) Results
-------------------------------------------------------------------------------
Coordinates type:                                                    Projected
Spatial kernel:                                               Adaptive bisquare
Criterion for optimal bandwidth:                                          AICc
Bandwidth used:                                                        261.000

Diagnostic Information
-------------------------------------------------------------------------------
Effective number of parameters (trace(S)):                               6.376
Degree of freedom (n - trace(S)):                                      255.624
Deviance:                                                            12029.202
AIC:                                                                 12041.955
AICc:                                                                12042.325
BIC:                                                                 12064.709
Percent deviance explained:                                             -0.016
Adj. percent deviance explained:                                        -0.042
Adj. alpha (95%):                                                        0.039
Adj. critical t value (95%):                                            2.072

Summary Statistics For GWR Parameter Estimates
-------------------------------------------------------------------------------
Variable                   Mean        STD        Min     Median        Max
------------------------ ---------- ---------- ---------- ---------- ----------
Intercept                -35.044      0.000    -35.044    -35.044    -35.044
OCC_TEC                    0.000      0.000     -0.000      0.000      0.000
OWNH                      -0.000      0.000     -0.000     -0.000      0.000
POP65                     -0.000      0.000     -0.000     -0.000      0.000
UNEMP                      0.000      0.000     -0.000      0.000      0.000
```

GWR model specifications

Model fitting results for the traditional, global regression model

Optimal bandwidth based on Golden Section search

Model diagnostic indicators for the fitted GWR model

Summary statistics of estimated coefficients of local terms

Figure 4.2b: Summary description for the Poisson GWR model

28

4.3   Binomial GWR model

For this tutorial, we will use the Clearwater sample data (landslides.csv) to estimate the probability of a landslide occurrence in Clearwater National Park using elevation and slope. The Clearwater sample data contains the following variables:

UserID   – Census FIPS code
X   – projected x-coordinate
Y   – projected y-coordinate
Elev   – elevation
Slope   – slope (in percentage)
Landslid   – Landslide occurrence

For our model, we will specify a Binomial GWR model. Because the outcome variable of interest, occurrence of a landslide, is a binary variable, a Binomial regression is better suited for this analysis. The general form for a Binomial GWR model is as follows:

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

For a Binomial GWR model, the dependent variable must be 0 or 1, where p_i is the modelled probability that the dependent variable becomes one.

After loading the Tokyomortality.csv file into MGWR 2.2, we will specify the following options:
1. Regression Variables
    a. Y = Landslid
    b. Local = Intercept, Slope, Elev
2. Location Variables
    a. ID = UserID
    b. X = X
    c. Y = Y
    d. Projected
3. GWR Mode = GWR
4. Model Type = Binomial

We will keep all other model defaults, such that the spatial kernel is an adaptive bisquare, the bandwidth searching uses Golden Section, and the Optimization Criterion is AICc. After making these selections, the MGWR 2.2 interface should look like Figure 4.3a:

Note: Variable standardization by default is turned on, however, users may want to turn off this option to improve interpretability of coefficients.
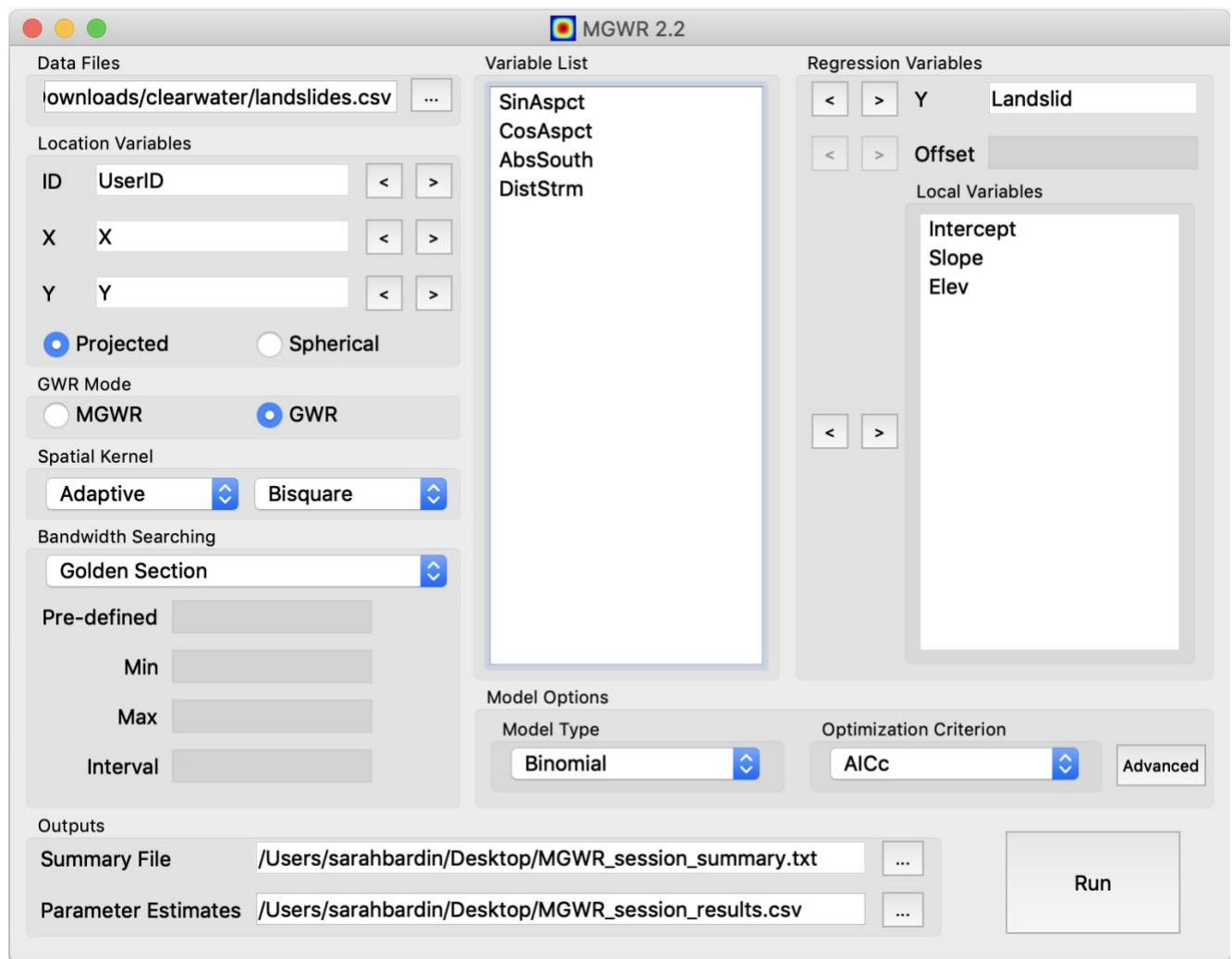
Figure 4.3a: GWR Binomial model specification in MGWR 2.2

Once the model is finished running, a "Summary" text box will appear. The summary includes the model settings, global model results, best bandwidth, and the model diagnostic information of the GWR model. Figure 4.3b describes the summary output in more detail. These results are also saved to the "MGWR_session_summary.txt" file, which is saved to the location specified in the "Summary File" section in the Outputs menu.

```
================================================================
Model type:                                            Binomial
Number of observations:                                     239
Number of covariates:                                         3
Dependent variable:                                    Landslid
Variable standardization:                                    On
Total runtime:                                          0:00:03


Global Regression Results
----------------------------------------------------------------
Deviance:                                               366.281
Log-likelihood:                                         -88.206
AIC:                                                    182.411
AICc:                                                   372.383
Percent deviance explained:                              -0.125
Adj. percent deviance explained:                         -0.134


Variable                         Est.        SE  t(Est/SE)    p-value
-------------------------------- ---------- ---------- ---------- ----------
Intercept                       -0.034     0.140     -0.244      0.807
Slope                            0.597     0.150      3.970      0.000
Elev                            -0.598     0.149     -4.007      0.000


Geographically Weighted Regression (GWR) Results
----------------------------------------------------------------
Coordinates type:                                     Projected
Spatial kernel:                                Adaptive bisquare
Criterion for optimal bandwidth:                           AICc
Bandwidth used:                                         237.000

Diagnostic Information
----------------------------------------------------------------
Effective number of parameters (trace(S)):               5.696
Degree of freedom (n - trace(S)):                      233.304
Deviance:                                              363.588
AIC:                                                   374.981
AICc:                                                  375.309
BIC:                                                   394.783
Percent deviance explained:                             -0.117
Adj. percent deviance explained:                        -0.144
Adj. alpha (95%):                                        0.026
Adj. critical t value (95%):                            2.235

Summary Statistics For GWR Parameter Estimates
----------------------------------------------------------------
Variable                  Mean        STD        Min     Median        Max
------------------- ---------- ---------- ---------- ---------- ----------
Intercept              -0.002      0.123     -0.221     -0.013      0.198
Slope                   0.531      0.119      0.343      0.548      0.750
Elev                   -0.550      0.053     -0.670     -0.546     -0.426
```

Figure 4.3b: Summary description for the Binomial GWR model

5.   MGWR Tutorial with Sample Data Set

For this tutorial, we will use the Georgia sample data (GData_utm.csv) to explore the spatial variation in the relationship between the percentage of residents with at least a bachelor's degree and various county-level characteristics. The Georgia sample data contains the following variables:

    AreaKey   – Census FIPS code
    Latitude   – spherical y-coordinate
    Longitude – spherical x-coordinate
    TotPop90 – total population in 1990

PctRural – percentage of rural population

PctBach – percentage of residents with at least a bachelor's degree

PctEld – percentage of elderly

PctFB – percentage of foreign-born inhabitants

PctPov – percentage of inhabitants living below the poverty level

PctBlack – percentage of African-Americans

ID – identifier

X – projected x-coordinate

Y – projected y-coordinate

For our model, we will specify a Gaussian MGWR model, which takes the following general form:

$$y_i \quad = \beta_{i0} + \sum_j \beta_{bwj}(u_i, v_i) \, x_{ij} \quad + \varepsilon_i$$

As discussed in Section 3.5.2, in an MGWR model, an optimal bandwidth is computed for each parameter surface, allowing the bandwidths to vary across parameter surfaces. As a result, these different bandwidths imply that every relationship at each location will have a different spatial weighting matrix. In order to estimate an MGWR model, a different back-fitting algorithm is used for calibration than is used in a traditional GWR model (for more information see Fotheringham, Yang and Kang, 2017).

After loading the GData_utm.csv file into MGWR 2.2, we will specify the following options:
1. Regression Variables
    a. Y = PctBach
    b. Local = Intercept, TotPop90, PctEld, PctFB, PctBlack
2. Location Variables
    a. ID = ID
    b. X = X
    c. Y = Y
    d. Projected
3. GWR Mode = MGWR
4. Advanced options
    a. Turn on Monte Carlo test for spatial variability
    b. Turn on Bandwidths confidence intervals
    c. Turn on Local collinearity diagnostics

We will keep all other model defaults, such that the spatial kernel is an adaptive bisquare, the bandwidth searching uses Golden Section, and the Optimization Criterion is AICc. In the advanced settings, we will keep the default options such that variable standardization will be turned on, the measure of score of change will be SOC-f, and

the convergence threshold will be 1e-5. After making these selections, the MGWR 2.2 interface should look like Figure 4.3:

Figure 5.1: MGWR Gaussian model specification in MGWR 2.2

Once the model is finished running, a "Summary" text box will appear. The summary includes the model settings, global model results, best bandwidth, and the model diagnostic information of the MGWR model. Figure 5.2 describes the summary output in more details. These results are also saved to the "MGWR_session_summary.txt" file, which is saved to the location specified in the "Summary File" section in the Outputs menu.

The parameter estimates for each location generated by the MGWR model are saved to the "MGWR_session_beta.csv" file, which is saved to the location specified in the "Parameter Estimates" section in the Outputs menu. The MGWR parameter estimates can be joined with shapefile "G_utm.shp" using the ID field. The G_utm.shp can be found in the Georgia data folder, where the GData_utm.csv is located. After joining the parameter results to the shapefile, using a mapping software, such as ArcGIS, QGIS, R, or python, the parameter estimates can be mapped to visualize the relationships found.

```
================================================================================
Model type:                                                    Gaussian          ┐
Number of observations:                                             159          │  MGWR model
Number of covariates:                                                 5          │  specification
Dependent variable:                                             PctBach          │
Variable standardization:                                            On          │
Total runtime:                                                 0:00:08          ┘

Global Regression Results
--------------------------------------------------------------------------------
Residual sum of squares:                                         61.675          ┐
Log-likelihood:                                                -150.323          │
AIC:                                                            310.646          │
AICc:                                                           313.199          │  Model fitting results
R2:                                                               0.612          │  for the traditional,
Adj. R2:                                                          0.602          │  global regression
                                                                                 │  model
Variable                         Est.          SE  t(Est/SE)     p-value         │
---------------------------- ---------- ---------- ---------- ----------         │
Intercept                       0.000       0.050      0.000       1.000         │
PctBlack                       -0.028       0.053     -0.523       0.601         │
PctFB                           0.315       0.068      4.627       0.000         │
TotPop90                        0.471       0.064      7.376       0.000         │
PctEld                         -0.135       0.060     -2.258       0.024         ┘

Multiscale Geographically Weighted Regression (MGWR) Results
--------------------------------------------------------------------------------
Coordinates type:                                             Projected          ┐
Spatial kernel:                                        Adaptive bisquare         │  MGWR advanced
Criterion for optimal bandwidth:                                   AICc          │  model
Score of change (SOC) type:                                  Smoothing f         │  specifications
Termination criterion for MGWR:                                1.0e-05          │
Number of iterations used:                                           15          ┘

MGWR bandwidths
--------------------------------------------------------------------------------
Variable              Bandwidth      ENP_j  Adj t-val(95%)         DoD_j         ┐
Intercept               106.000      2.939           2.412         0.787         │  Optimal
PctBlack                 96.000      3.458           2.473         0.755         │  bandwidths for
PctFB                   116.000      2.706           2.381         0.804         │  each model
TotPop90                 67.000      4.456           2.566         0.705         │  predictor based
PctEld                  142.000      2.247           2.309         0.840         ┘  on Golden
                                                                                    Section search

Bandwidth Confidence Intervals (95%)
--------------------------------------------------------------------------------
Intercept                                                  (87.0, 131.0)         ┐  Confidence
PctBlack                                                   (70.0, 114.0)         │  intervals for
PctFB                                                      (87.0, 131.0)         │  optimal
TotPop90                                                   (59.0, 114.0)         │  bandwidths for
PctEld                                                     (87.0, 148.0)         ┘  each model
                                                                                    predictor

Monte Carlo Test for Spatial Variability
--------------------------------------------------------------------------------
Variable                                                         p-value         ┐
Intercept                                                         0.326          │  P-values
PctBlack                                                          0.071          │  corresponding
PctFB                                                             0.605          │  to Monte Carlo
TotPop90                                                          0.019          │  test for each
PctEld                                                            0.511          ┘  model predictor
```

34

```
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
Degree of Dependency (DoD):                                       0.773
AIC:                                                            285.193
AICc:                                                           289.432
BIC:                                                            336.770
R2:                                                               0.715
Adj. R2:                                                          0.683


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
```

Model diagnostic indicators for the fitted MGWR model

Summary statistics of estimated coefficients of local terms

Figure 5.2: Summary description for the Gaussian GWR model

6.   References

1. 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

2. Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). Annals of the American Association of Geographers, 107(6), 1247-1265. DOI: 10.1080/24694452.2017.1352480

3. Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS International Journal of Geo-Information, 8(6), 269.

4. Li, Z., & Fotheringham, A. S. (2020). Computational improvements to multi-scale geographically weighted regression. International Journal of Geographical Information Science, 1-20. DOI: 10.1080/13658816.2020.1720692

5. Li, Z., Fotheringham, A. S., Oshan, T. M., & Wolf, L. J. (2020). Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights. Annals of the American Association of Geographers, 1-21.

# 7.   MGWR Software Change Log

v2.2(02/21/2020)

1. Add bandwidth confidence interval.

v2.1 (04/01/2019)

1. Kernel option: only allowing Fixed Gaussian + Adaptive Bi-square kernel combination.
2. For Gaussian kernel, report new bandwidth = old bandwidth times 2.45 (when the weight drops below 0.05). This can be considered as an effective bandwidth in the Gaussian setting.
3. Add sum of local weights for GWR and MGWR in the output csv file. The sum of weights is a measure for effective number of data points used at each location.
4. Add local_R2 for MGWR This is calculated based on optimal GWR bandwidth.
5. Add degree of dependency for GWR and MGWR as $\frac{\log(mk)-\log(ENP)}{\log(mk)-\log(k)}$. For MGWR, there is also covariate-specific DoD_j calculated as $\frac{\log(m)-\log(ENP_j)}{\log(m)-\log(1)}$. This measure is ranging from 0 to 1, and 1 indicates total dependency (global process).
6. Add OLS residual in the output.
7. Add interval search for MGWR.

v2.0 (04/01/2019)

1. This update contains computational improvements to MGWR (see Li and Fotheringham, 2020 for details). For general guideline of time taken, please refer to the figure below.
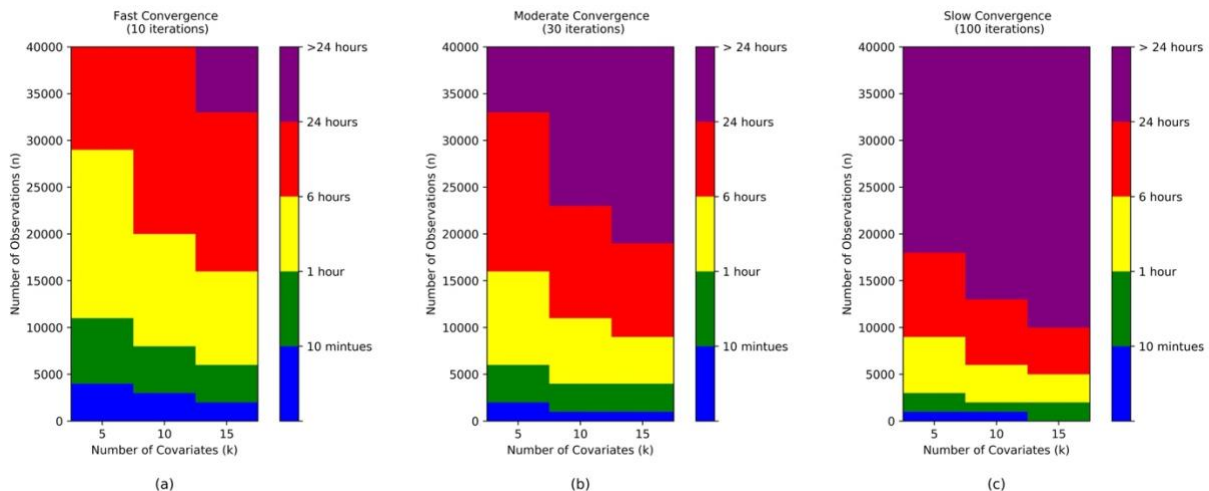
Li, Z., Fotheringham, A. S. (2020). Computational improvements to multi-scale geographically weighted regression. *International Journal of Geographical Information Science. doi:* 10.1080/13658816.2020.1720692

2. Local VIF and local VDP are available for both GWR and MGWR as measures of collinearity.

v1.0 (10/01/2018)

Initial release.