


SAAR: Spatial Analysis using ArcGIS engine and R

Introduction

This lab demonstrates the functionalities of SAAR using an empirical dataset. It follows a general spatial data analysis process, ranging from simple mapping, exploratory spatial data analysis, and confirmatory spatial data analysis, to an uncertainty exploration in the analyses and data.

The empirical data are obtained from the 2010-2014 five-year American Community Survey (ACS) data for Texas counties. Specifically, this lab mainly presents a series of spatial data analyses using the estimates of median house value (*hvalue*) for the 254 counties in Texas. Two additional variables, the unemployment rate (*unemploy*) and median year in which a structure was built (*year*), are used as covariates in the context of regression.



Setting up the lab

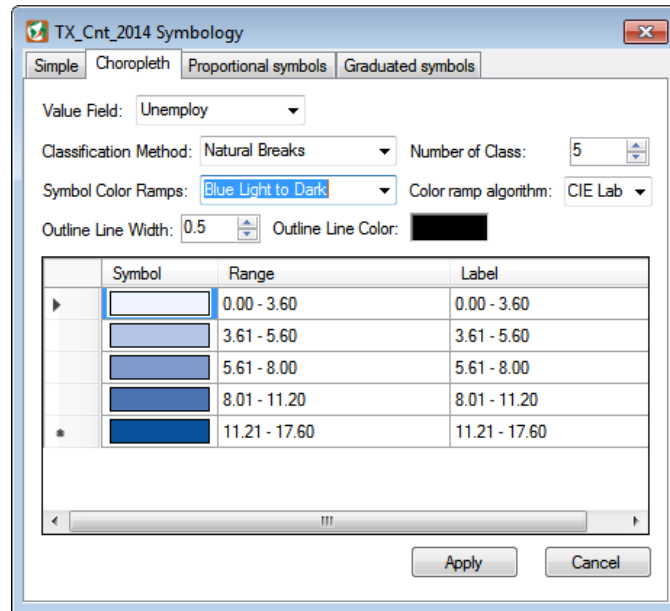
- Lab data and installation file for the software are available on T:\ywchun\SAAR.
- Make a working folder and get data files in the folder
- The software is already installed on your local machine. Start SAAR by clicking the icon .

Task 1. Simple mapping and creating a layout

1-1. Drawing a choropleth map

► We will draw a simple choropleth map of unemployment.

1. Select File → Open Document . Browse to your own folder and open “**Value of Home.mxd**”.
2. Click  to open the "Add shapefile" dialog. Browse to your own folder and add “**TX_Cnt_2014.shp**”.
3. Right-click on *TX_Cnt_2014* and choose "Symbology". You should see a “Symbology” dialog window.
4. Click on the tab called “Choropleth”.
5. We need to specify the field (i.e., attribute) we want to display. Pull down the box right to "Value Field" and select ***Unemploy***.



→ The classification was automatically done. We can also define the number of classes and a classification method. In addition, we can determine the break point for each class manually by clicking the range list box.





6. Change “Symbol Color Ramps” to **Blue Light to Dark** located at the top of the color scheme list. Then, click Apply to draw a choropleth map. Close the “Symbology” dialog window.

1-2. Export the map layout as an image file

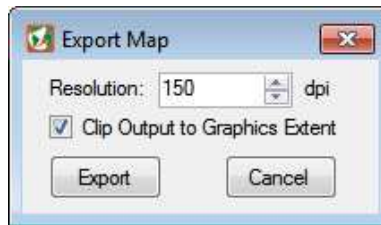
► Now, we will export the map as an image file.

1. From the menu bar, select Layout → Open Layout View.
2. We will work on only the *TX_Cnt_2014* layer. Turn off the display of *Hvalue* by checking off the boxes before their layer names.
3. Select Insert → Legend and North Arrow to add a legend and a north arrow on the map.
4. Choose Insert → Scale Bar and a window will pop up. Select “Division Units” as **Miles** and set the “Number of Division” as **2**.




5. Using navigation tools (e.g., zoom in  and out , full extent ), change the location and scale of the map. Also, using the select element tool , change the location and size of the map elements (i.e., legend, scale bar, and north arrow).

6. Select File → Export to open an Export window. Check “Clip Output to Graphics Extent”, and then, click Export.



7. Change “Save as type” to “Tiff files (*.tif)”, and type a proper name of the map.

8. Close the layout view by clicking the close button .

Task 2. Exploratory data analysis (EDA)

■ SAAR contains EDA tools with the support of dynamic linking and brushing techniques among statistical graphs and a layer map in a map view. EDA tools support not only generic graphics such as histogram, boxplot, and scatter plot, but also violin and quantile-comparison plots.

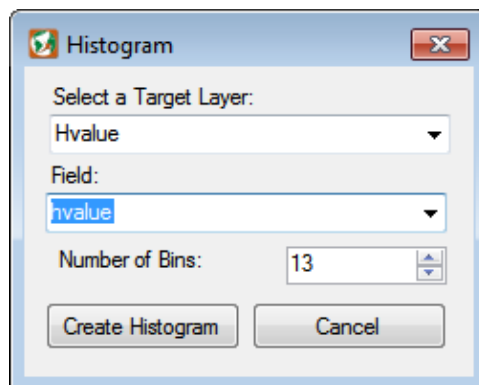
2-1. Drawing a histogram and scatter plots

► In here, we will explore the distribution of *hvalue*, and a relationship between *hvalue* and the selected covariates; *unemploy* and *year*.

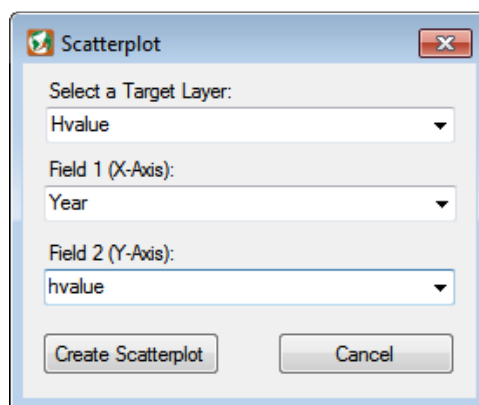
1. Now, we will mainly work on the *Hvalue* layer. Drag the *Hvalue* layer to the top of the Table of Content. Then, turn off the display of *TX_Cnt_2014* by checking off the boxes.

2. Click Explore → Histogram. You can see a “Histogram” window.

3. Set a target layer to the *Hvalue* layer, and a field to *hvalue*. Then, click the “Create Histogram” button.



4. Explore the numerical distribution of *hvalue*, and do **NOT** close the histogram.
5. Select Explore → Scatterplot. Choose a target layer as the *Hvalue* layer.
6. Set the “Field 1 (X-Axis)” to *year* and “Field 2 (Y-Axis)” to *hvalue*. Then, click the “Create Scatterplot” button.

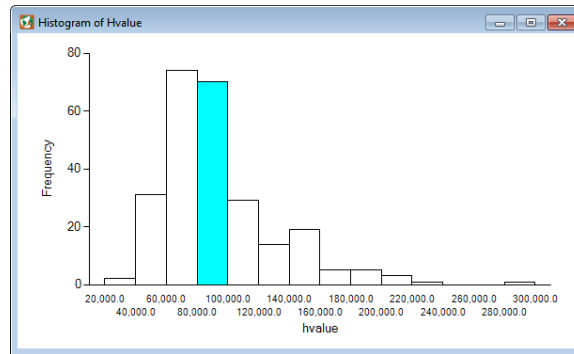


7. In the same way, create a scatter plot of *hvalue* and *unemploy* to see a relationship between them.
8. Explore the relationship between *hvalue* and its covariates. Do **NOT** close the scatter plots.

2-2. Dynamic linking and brushing

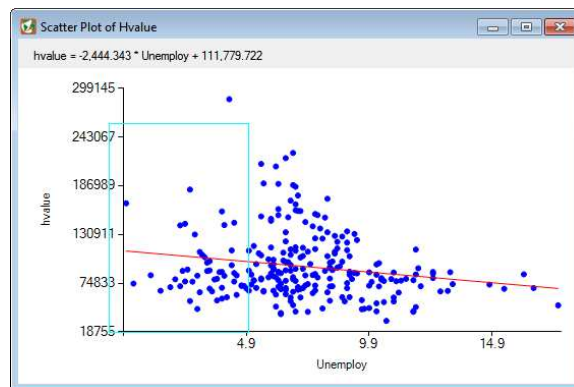
► Now, the relationship among the three variables can be explored further with dynamic linking and brushing.


1. Arrange the size and position of three statistical plots and the map view to see all windows simultaneously.
2. Click an individual bin in the histogram. Then, you will see the corresponding counties in the scatter plots and the map view.




→ In this dynamic linking and brushing, different color symbols are used for an input (in cyan) and target sources (in red) to distinguish them easily from each other.

3. Now, on the scatter plot of *hvalue* and *unemploy*, click a left mouse button and drag to choose points. Then, explore the corresponding counties in different windows.



4. A map view can also be used as an input of dynamic linking and brushing. Click “Brushing and Linking” tool  on the toolbar in the main window. Then, select regions that you want to see in the histogram and the scatter plots.

→ We will use dynamic linking and brushing with different graphs in the next task.

5. Close all graphs. If you have selected counties in the map view, click the “Clear Selected Features” tool  on the toolbar.

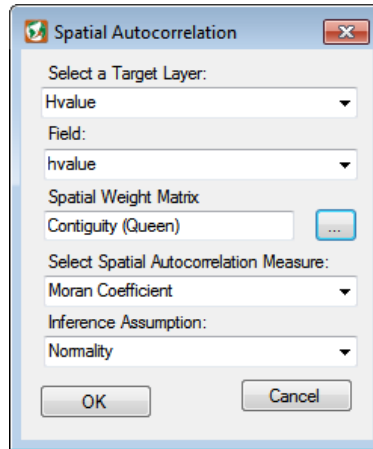
Task 3. Exploratory spatial data analysis (ESDA) and spatial autocorrelation

■ SAAR provides a set of global and local spatial autocorrelation measures, and ESDA tools, including Conditioned Choropleth maps, a connectivity histogram and map, a Moran scatter plot, and spatial correlogram.


3-1. Spatial autocorrelation

► In here, we will explore both global and local spatial autocorrelation for *hvalue*.

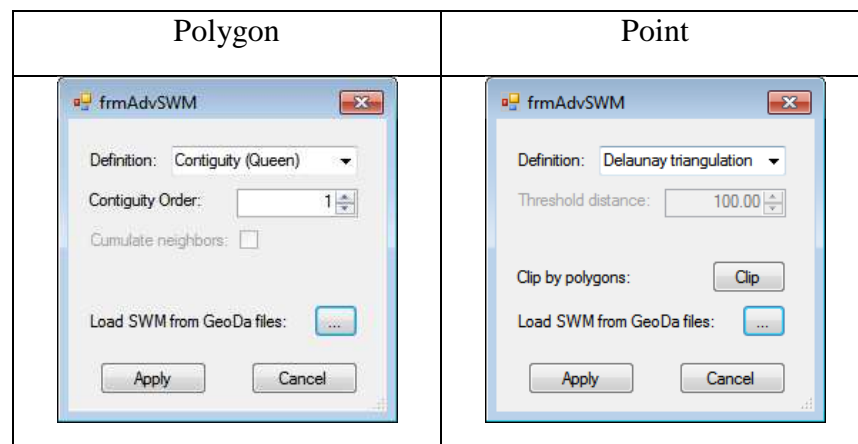
1. Select Analysis → Global Spatial Autocorrelation.
2. Set a target layer to the *Hvalue* layer, and a field to *hvalue*.



→ The default for the definition of a spatial weights matrix is a queen definition for polygon contiguities.


When you click the button , You can also use a rook definition and higher order spatial weights. In addition, SAAR can load a spatial weights matrix from GeoDa files (i.e., *.gal and *.gwt).

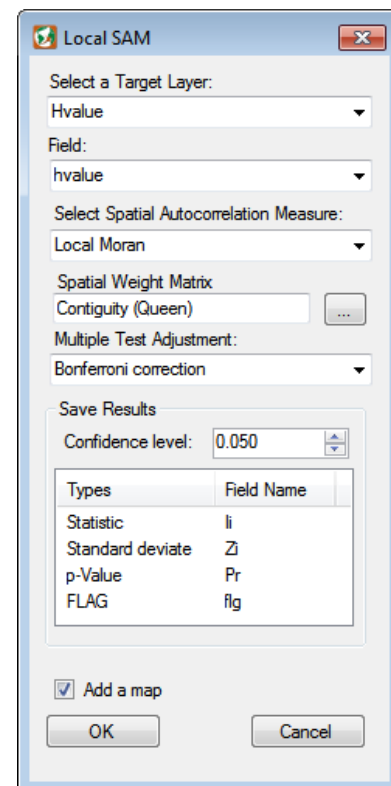
→ For point contiguities, the default for the definition is a Delaunay triangle, which can be clipped by a polygon. You can also use a fixed distance and k-nearest neighbors definitions in the advanced spatial weights matrix window.



3. Select a spatial autocorrelation measure as **Moran Coefficient**, and an inference assumption as **Normality**.

→ This tool also supports a Geary ratio and a randomization assumption.

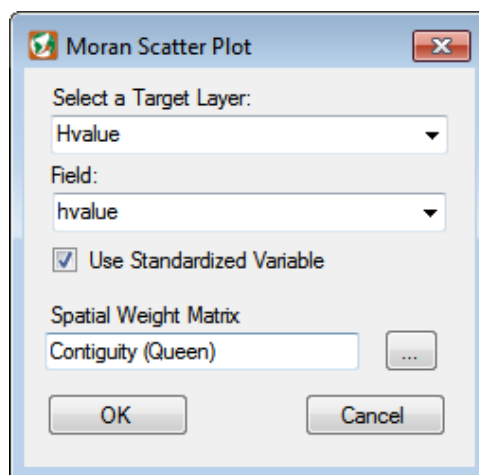
4. Click OK, then you can see a “Summary” window.
5. Now, we will explore local spatial autocorrelation. Select Analysis → Local Spatial Autocorrelation (univariate). Then, a “Local SAM” window will pop up.
6. Set a target layer to the *Hvalue* layer, and a field to *hvalue*.
7. Select a spatial autocorrelation measure as **Local Moran**.
8. Set a multiple test adjustment to **Bonferroni correction**.
9. Check on the “Add a map” box, and then Click OK.
10. The results of local spatial autocorrelation are save in the target shape file. Based on the results, a local Moran Coefficient map will be drawn.
11. Right-click on a *Local Moran of Hvalue* layer and choose "Open Attribute Table". You can see an attribute of the layer from the window. In there, you can also find the results of the local spatial autocorrelation.
12. Close the attribute table by clicking the close button .



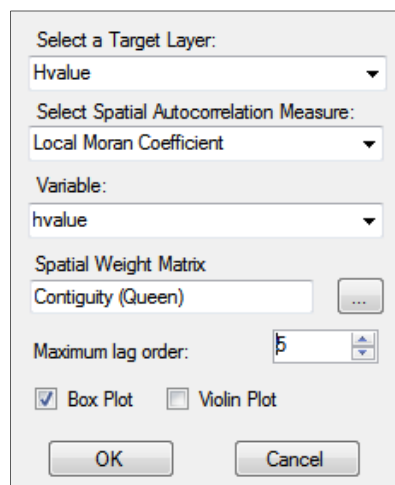
3-2. Moran scatter plot and spatial correlogram

► Now, we will explore spatial autocorrelation further using a Moran scatter plot and spatial correlogram with dynamic linking and brushing.


2. Click Explore → Moran Scatterplot. You can see a “Moran Scatter plot” window.



3. Set a target layer to the *Hvalue* layer, and a field to *hvalue*. Check on the box of “Use Standardized Variable”.
4. Click OK, then you will see the result of the Moran scatter plot.
 - The slope of a regression line represents a global Moran Coefficient value when using standardized variables.
5. Click a left mouse button and drag to select points. Then, you can see a corresponding county and its neighbors. Do **NOT** close the Moran scatter plot.
6. Now, select Analysis → Spatial Correlogram.
7. Again, set a target layer to the *Hvalue* layer, and a field to *hvalue*.
8. Select a spatial autocorrelation measure as Local Moran, and the maximum lag order as **5**. Then, click OK. The spatial correlogram will appear in a left panel.



→ A spatial correlogram tool can help to explore spatial autocorrelation at a particular spatial lag and its trend across spatial lags.

9. In the same way to the previous Moran scatter plot, choose points in the spatial correlogram to see corresponding counties and their neighbor on the map view and the Moran scatter plot.
10. Explore the spatial autocorrelation in different spatial lags using dynamic linking and brushing.
11. Close both plots by clicking the close button .
12. Right-click on a *Local Moran of Hvalue* layer and choose "Remove layer" to remove this layer from the table of contents.

Task 4. Data transformations and regression analysis

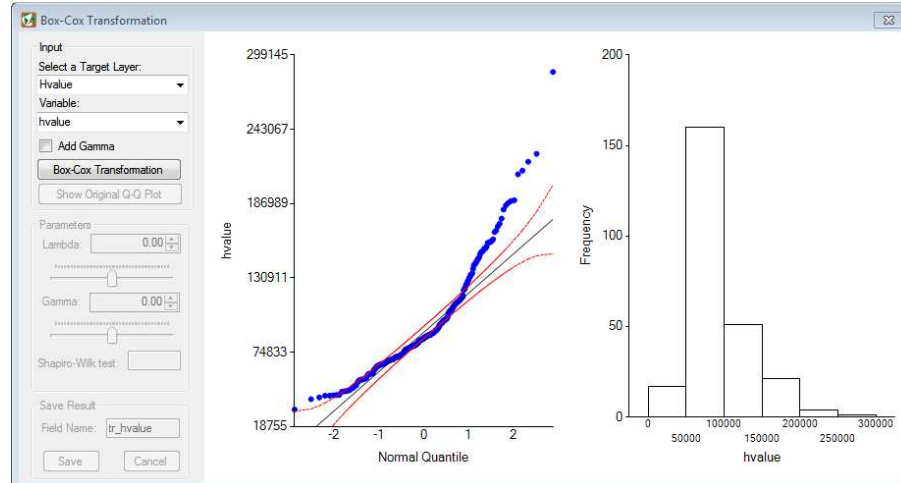
■ SAAR supports various regression analysis tools, including simple linear regression, generalized linear model (GLM), various spatial regressions, and Moran eigenvector spatial filtering (MESF) for linear and GLM specifications. In addition, SAAR furnishes a Box-Cox transformation tool to convert an input variable to one that better mimics a normal distribution (i.e., bell-shaped curve).

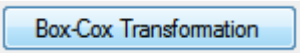
4-1. Box-Cox transformation

1. Select Data → Data Transformation, and then a “Box-Cox Transformation” window will pop up.
2. Set a target layer to the *Hvalue* layer, and a field to *hvalue*. A normal quantile-quantile (Q-Q) plot and a histogram will be drawn based on the select field.

→ This tool applies Box-Cox transformation for a selected variable using the following equation. If the selected variable contains 0 (zero) values, an “Add Gamma” box will be automatically checked on.



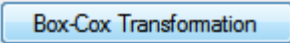
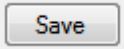
$$y_i^{(\lambda)} = \begin{cases} \frac{(y_i + \gamma)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln(y_i + \gamma) & \text{if } \lambda = 0 \end{cases}$$




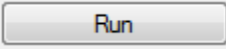

3. Click the “Box-Cox Transformation” button . Then, an estimated parameter of “Lambda” (λ) will be shown. Also, the normal Q-Q plot and the histogram will be updated based on the transformed variable.

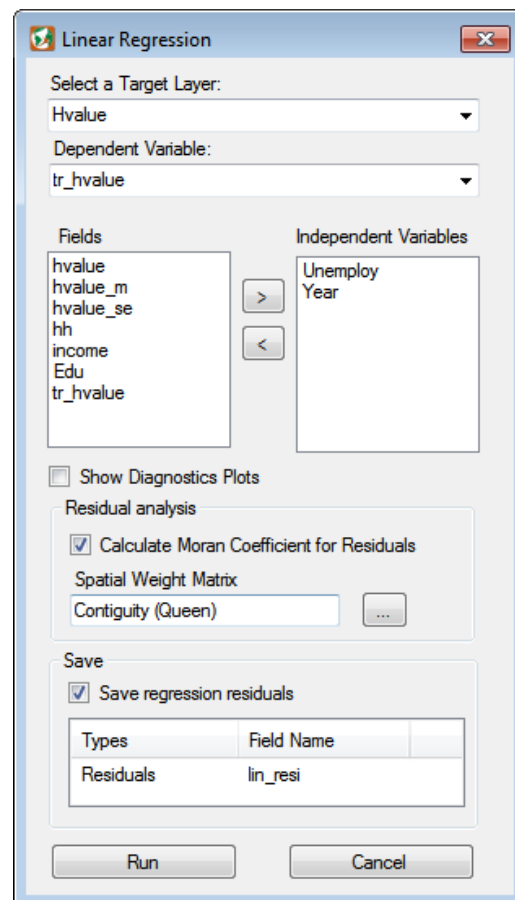
→ The parameters for the transformation are estimated by the maximum likelihood estimation.


4. See the normal Q-Q plot, the histogram, and the result of Shapiro-Wilk test to evaluate the transformation result.

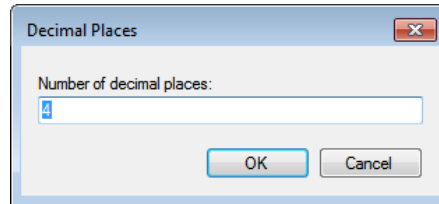
5. Click the “Show Original Q-Q plot button”  to compare the distribution of the transformed variable with that of the original variable.
6. Now, move the slide bar  under the Lambda parameter to interactively see the impact of the parameter on the transformation result.
7. Click the “Box-Cox Transformation” button  again to get the optimal parameters.
8. In the “Save Result” group box, set the “Field Name” to *tr_hvalue*. Then, click the “Save” button  to store the transformed variable as an additional field. We will use this field for regression analyses.

4-2. Linear regression


1. Click Regression → Linear Regression. You can see the “Linear Regression” window.
2. Choose a target layer as the *Hvalue* layer, and a dependent variable as *tr_hvalue*, which is derived from the previous step.
3. Set independent variables to *unemploy* and *year* by click the button  or double-click on a field name.
→ If you check on the “Show Diagnostics plots”, five diagnostic plots (e.g., plots of residuals against fitted values, square root of absolute residuals against fitted values, a Normal Q-Q plot, and residuals against leverages) will be appear.
4. In the “Residual analysis” group box, check on the “Calculate Moran Coefficient for Residuals” box. Set the spatial weight matrix to **Contiguity (Queen)**.
5. Also, in the “Save” box, check on the “Save regression residuals” box to draw the map of the linear regression residuals.
6. Click the “Run” button , and then, the “Linear Regression Summary” window will pop up. Evaluate the linear regression result.
7. Close the “Linear Regression Summary” and “Linear Regression” windows by clicking the close button .



8. Right-click on the *Hvalue* layer and choose "Symbology". Choose the value field as *lin_resi*.
9. Change the number of classes to **6** and symbol color ramps to "**Blue to Red**".
10. Click the head  of label lists. Then, you can the "Decimal Places" window. Change the value to **4**.

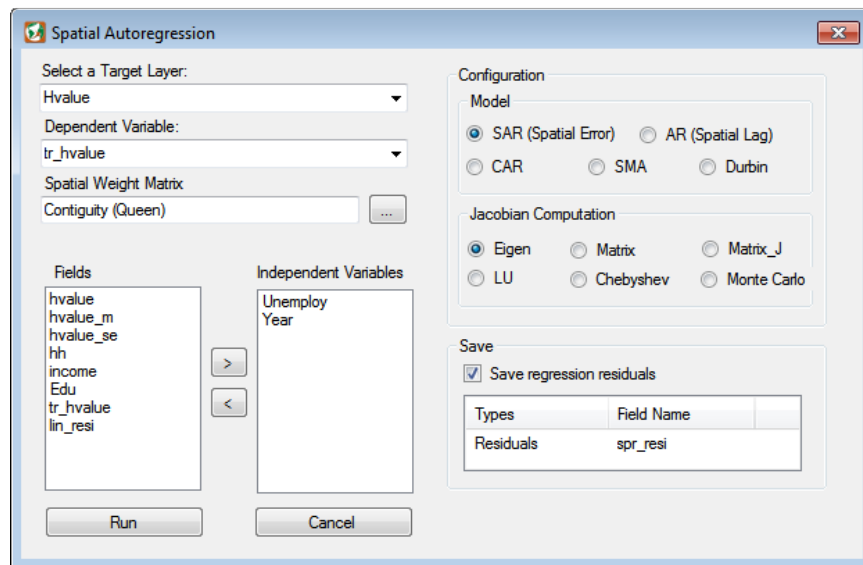


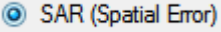

→ For further adjustment, You can manually change the range of classes by clicking the range list box.

11. Click the "Apply" button , and close the "Symbology" window. Explore the spatial distribution of the linear regression residuals.

4-3. Spatial regression

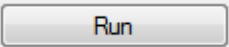
1. Click Regression → Spatial Autoregression. Then, the "Spatial Autoregression" window.
2. Choose a target layer as the *Hvalue* layer, and a dependent variable as *tr_hvalue*.
3. In the same way to the previous step, assign *unemploy* and *year* to independent variables.



4. In the “Configuration” group box, we will use simultaneous autoregressive (SAR) model  with the exact eigenvalue method  for Jacobian computation.

→ For further information, refer to https://www.rdocumentation.org/packages/spdep/versions/0.7-4/topics/do_1det.

5. In the “Save” box, check on the “Save regression residuals” box to explore the spatial distribution of spatial regression residuals.

6. Click the “Run” button , and then, the “Spatial Autoregressive Model Summary” window will pop up. Evaluate the result.

→ You might see a warning message. Exact Jacobian computation using eigenvalue might require relatively long time for the calculation depending on a performance of your computer.

7. Close the summary and spatial autoregression windows.

8. Right-click on the *Hvalue* layer and select "Symbology".

9. Now, using the *spr_resi* field, draw a residual map. You can refer the previous step (8-11 in 4.2 Linear regression).

10. Explore the spatial distribution of the spatial regression residuals.

4-4. Moran eigenvector spatial filtering (ESF)

1. Select Regression → Eigenvector Spatial Filtering. Then, the “Eigenvector Spatial Filtering” window will appear.

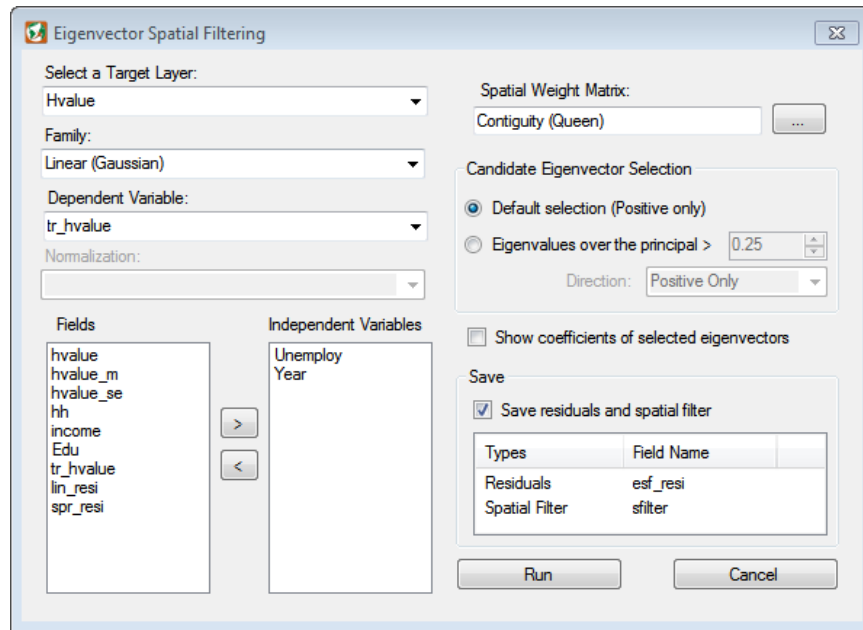
2. Choose a target layer as the *Hvalue* layer, a dependent variable as *tr_hvalue*, and dependent variables as *unemploy* and *year*.


3. We will use a linear regression specification for MESF. Choose the “Family” as **Linear (Gaussian)**.

4. In the “Candidate Eigenvector Selection”, select “**Default selection (positive only)**.”

→ For further information for the candidate selection method, you can refer Chun Y, Griffith D A, Lee M, and Sinha P 2016 Eigenvector selection with stepwise regression techniques to construct eigenvector spatial filters. *Journal of Geographical Systems* 18: 67–85.

6. In the “Save” box, check on the “Save residuals and spatial filter” box to explore a residual distribution and selected spatial filter.



7. Click the “Run” button , and then, the “ESF Linear Regression Summary” window will pop up. Evaluate the result.

→ Again, MESF need to calculate eigenvalues and eigenvectors. If the number of observation become large, the time needed for the estimation increases.

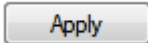
8. Close the summary and spatial autoregression windows.

8. Right-click on the *Hvalue* layer and select "Symbology".

9. Again, using the *esf_resi* field, draw a residual map, and explore the map.

10. Now, we will draw the map of the estimated spatial filter. Open the “Symbology” window from the *Hvalue* layer.

11. Choose the value field as *lin_resi*, Then, change the number of classes to **6** and symbol color ramps to “**Green to Red**”.

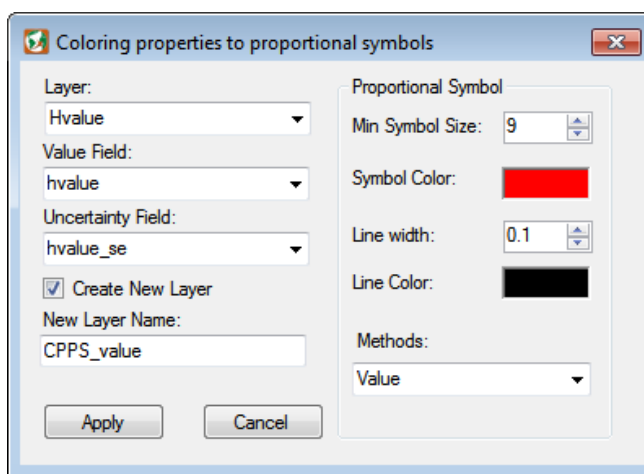
12. Click the “Apply” button , and close the “Symbology” window. Explore the spatially structured random effect from the regression model.

Task 5. Geovisualization for spatial data uncertainty

■ SAAR provides toolsets to explore uncertainty in spatial data and SDA output in two different contexts: geovisualization, and map classification. In this lab, we will use uncertainty geovisualization tools developed based on bivariate mapping technique.

5-1. The coloring properties to proportional symbols

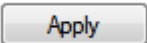
1. Click Uncertainty → Visualization (Static) → Proportional Symbol with Color Combination. Then, the “Coloring properties to proportional symbols” window will pop up.



2. Choose a target layer as the **Hvalue** layer. Set a value and uncertainty fields as **hvalue** and **hvalue_se**, respectively.

3. Check on the “Create New Layer” box, and type a new layer name to “**CPPS_Value**”.

4. Set the “Min Symbol Size” to **9**, and the “Methods” to **Value**.

5. Click the “Apply” button , and close the window.

6. Explore the median house value map, and examine the reliability of the value.

- The End-