

Exploring local variability in statistical relationships with

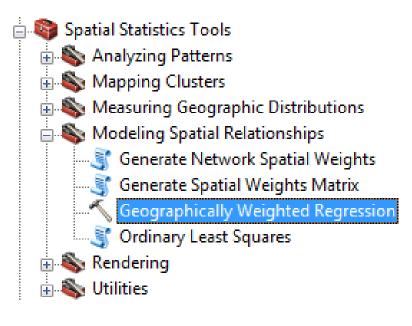
### GEOGRAPHICALLY WEIGHTED REGRESSION

Betsy Breyer Department of Geography Portland State University breyer@pdx.edu

# GWR

# OLS

#### Geographically Weighted Regression



#### Ordinary Least Squares (regression)

🜍 Spatial Statistics Tools
🛓 🗞 Analyzing Patterns
🗄 🦠 Mapping Clusters
🛓 🗞 Measuring Geographic Distributions
🚔 🗞 Modeling Spatial Relationships
🧃 Generate Spatial Weights Matrix
Geographically Weighted Regression
Ordinary Least Squares
🗄 🦠 Rendering
🗄 🦠 Utilities

# **OVERVIEW**

- 1. ORIGINS OF GWR
- 2. CRITIQUES OF GWR
- 3. USING GWR IN ARCGIS
  - a) Exploring data
  - b) Developing an OLS model
  - c) Running GWR, interpreting results
  - d) Model comparison
- 4. REFERENCES & RESOURCES
- 5. DISCUSSION

Section 1

### **ORIGINS OF GWR**

Chris Brunsdon, A. Stewart Fotheringham and Martin E. Charlton

#### Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity

Spatial nonstationarity is a condition in which a simple "global" model cannot explain the relationships between some sets of variables. The nature of the model must alter over space to reflect the structure within the data. In this paper, a technique is developed, termed geographically weighted regression, which attempts to capture this variation by calibrating a multiple regression model which allows different relationships to exist at different points in space. This technique is loosely based on kernel regression. The method itself is introduced and related issues such as the choice of a spatial weighting function are discussed. Following this, a series of related statistical tests are considered which can be described generally as tests for spatial nonstationarity. Using Monte Carlo methods, techniques are proposed for investigating the null hypothesis that the data may be described by a global model rather than a non-stationary one and also for testing whether individual regression coefficients are stable over geographic space. These techniques are demonstrated on a data set from the 1991 U.K. census relating car ownership rates to social class and male unemployment. The paper concludes by discussing ways in which the technique can be extended.

Brunsdon, C., A.S. Fotheringham, and M.E. Charlton. 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis* 28 (4): 281-298.

Chris Brunsdon, A. Stewart Fotheringham and Martin E. Charlton

### Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity

Spatial nonstationarity is a condition in which a simple "global" model cannot explain the relationships between some sets of variables. The nature of the model must alter over space to reflect the structure within the data. In this paper, a technique is developed, termed geographically weighted regression, which attempts to capture this variation by calibrating a multiple regression model which allows different relationships to exist at different points in space. This technique is loosely based on kernel regression. The method itself is introduced and related issues such as the choice of a spatial weighting function are discussed. Following this, a series of related statistical tests are considered which can be described generally as tests for spatial nonstationarity. Using Monte Carlo methods, techniques are proposed for investigating the null hypothesis that the data may be described by a global model rather than a non-stationary one and also for testing whether individual regression coefficients are stable over geographic space. These techniques are demonstrated on a data set from the 1991 U.K. census relating car ownership rates to social class and male unemployment. The paper concludes by discussing ways in which the technique can be extended.







Brunsdon, C., A.S. Fotheringham, and M.E. Charlton. 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis* 28 (4): 281-298.

# **GWR: A local regression model**

- GWR estimates an OLS-like regression for each feature
  - Analogous to other local/global statistics
    - Morans I vs Local Morans
    - Getis-Ord General G vs Getis-Ord Gi\*
- Designed to explore spatial nonstationarity in parameters
  - Nonstationary model, heterogeneous pattern
  - Spatial heterogeneity vs. spatial dependence
- Key to the analysis:
  - Model development
  - Define the neighborhood

# The nature of local variation in statistical relationships

- 1. Primacy of place
  - a) Nonstationarity arises from intrinsic local differences
  - b) Global statements of spatial behavior are not possible
  - c) Raises additional questions: what defines a neighborhood?
- 2. Model misspecification
  - a) Apparent nonstationarity simply results from omitted variables
  - b) When all sources of variability are included, the effect disappears
  - c) Sampling bias?

See also

Fotheringham, A.S. and C. Brunsdon. 1999. Local forms of spatial analysis. *Geographical Analysis* 31 (4): 340-358.

Fotheringham, A.S., C. Brunsdon, and M.E. Charlton. 2000. *Quantitative geography: Perspectives* on spatial data analysis. London: SAGE Publications

# **CRITIQUES OF GWR**

Section 2

### Simulation results

- GWR has not consistently differentiated between stationary and nonstationary data-generating processes
- Multicollinearity in estimated coefficients may bias results
- Unclear what tests can reliably diagnose model problems

#### Simulation studies:

Wheeler D, Tiefelsdorf M, 2005, "Multicollinearity and correlation among local regression coefficients in geographically weighted regression" Journal of Geographical Systems 7 (2) 161 - 187.

Páez, A., S. Farber, and D. Wheeler. 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A* 43 (12): 2992 – 3010.

# "notoriously unreliable"

http://r-sig-geo.2731867.n2.nabble.com/A-question-about-gwr-morantest-pvalue-td7292670.html

There is no "correct" here, as ?pchisq shows that if lower.tail= is not set it is taken by default to be TRUE. You will need to check this yourself. Note that GWR is a notoriously unreliable technique, and simulation studies indicate that it finds pattern in coefficients even when there is none. So any tests are doubtful anyway - it should only be used for exploring the data for possible missing variables or inappropriate functional forms.

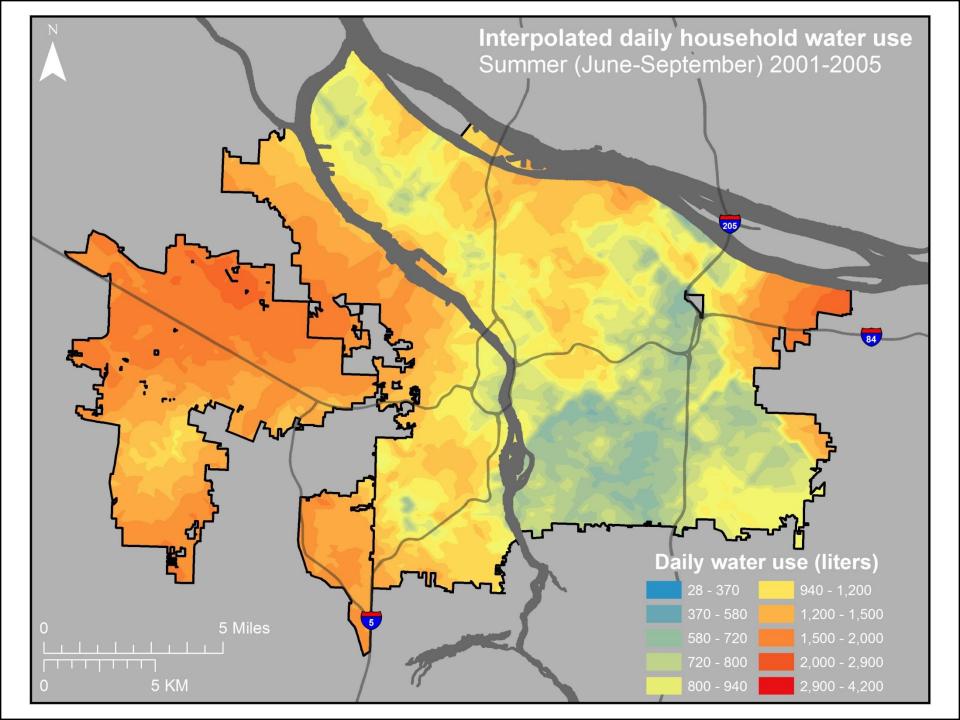
Hope this clarifies,

Roger --Roger Bivand Department of Economics, NHH Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway. voice: +47 55 95 93 55; fax +47 55 95 95 43 e-mail: Roger.Bivand at nhh.no

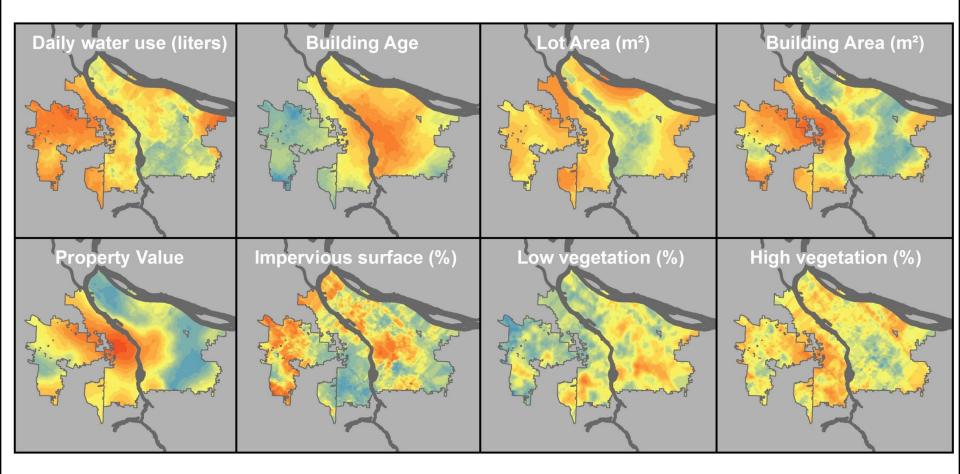
R package spgwr: No standard errors and t-statistcs ArcGIS: includes standard errors, does not calculate t-statistics GWR4: includes both standard errors and t-statistics

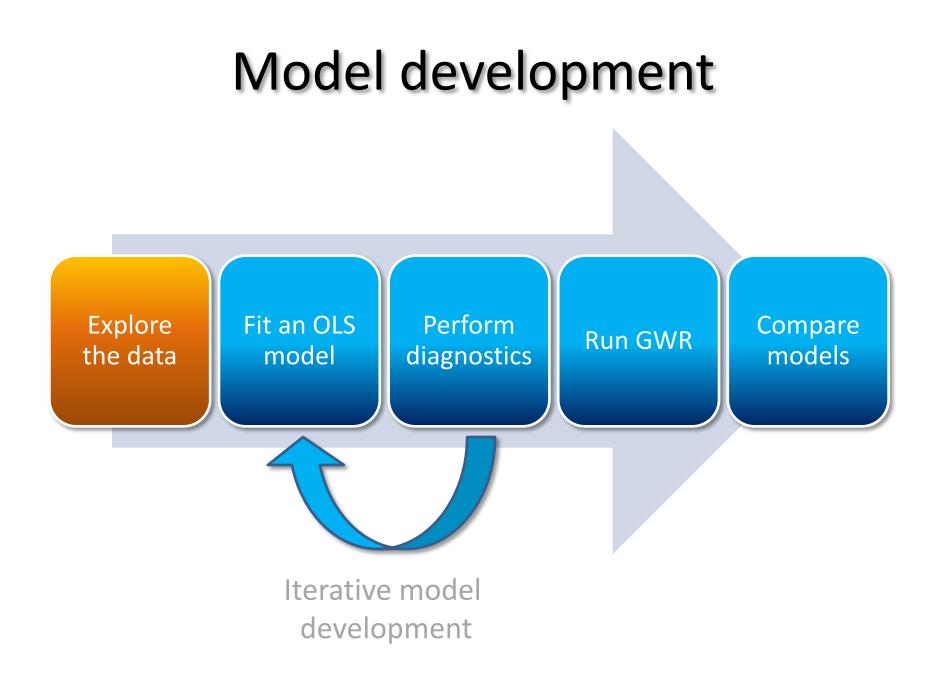
# RESIDENTIAL LAND USE AND DAILY HOUSEHOLD WATER USE

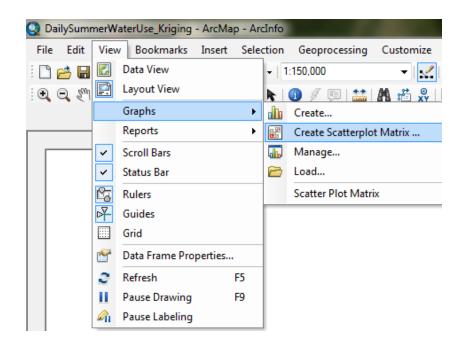
Section 3: An example in ArcGIS 10



# Candidate explanatory variables



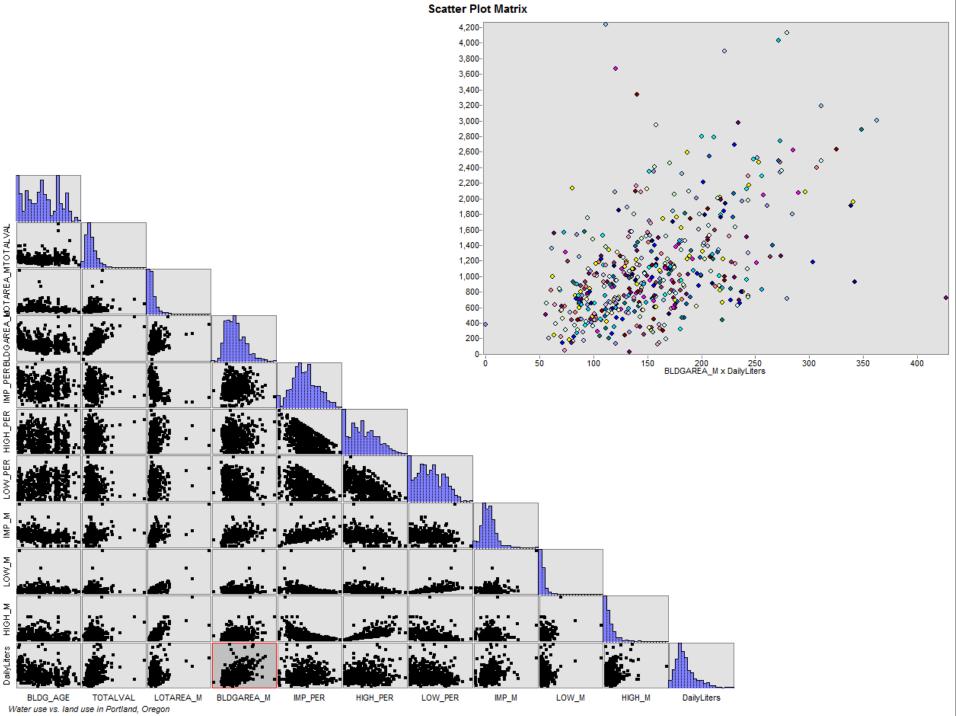




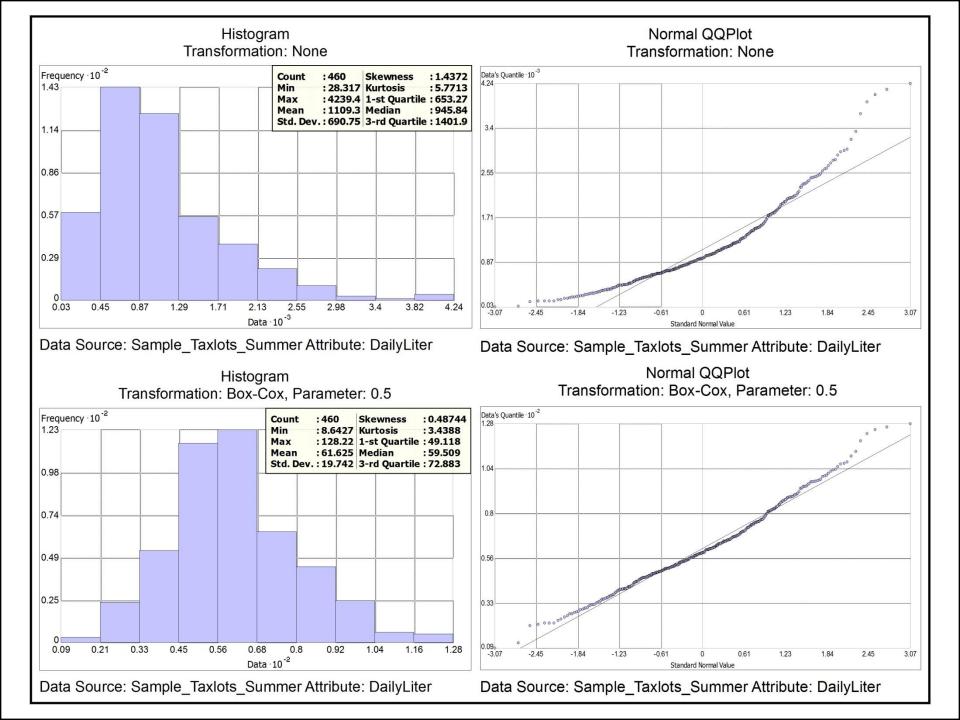
### Scatterplot matrix: USEFUL

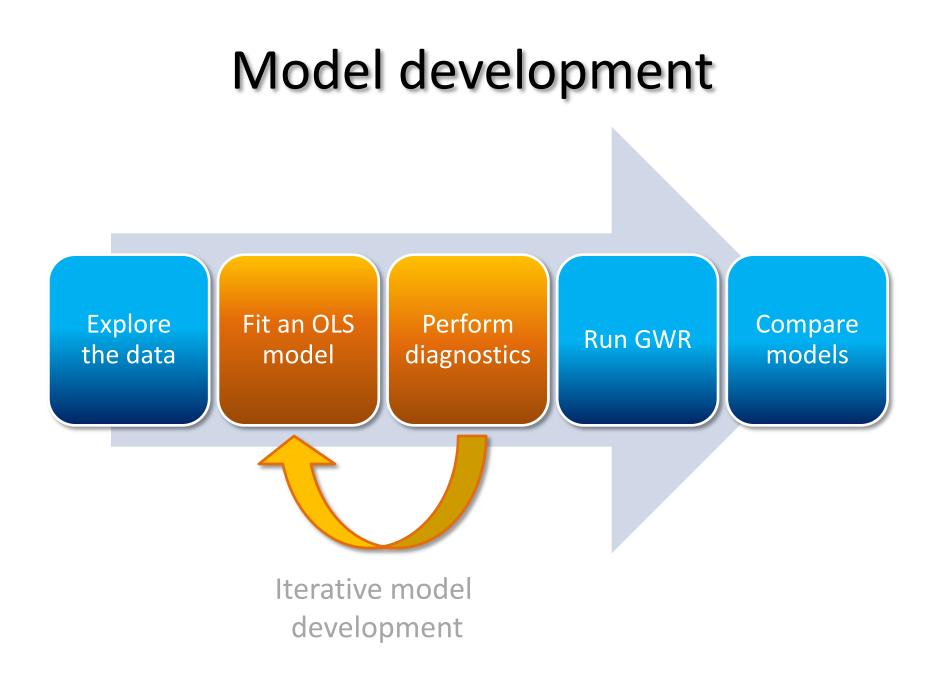
Q Sqr	tTransformatio	on - ArcMap -	ArcInf	fo		
File	Edit View	Bookmarks	Inse	rt S	election Geoprocessing	Customize
: 🗅 🛛	🐴 日 🖨 🛛	). 🖻 🛍 🗙	5	C	- I:159,682	▼   <u>⊀</u>
i 🗨 🔍 🕙 i 💥 23 i 🦛 🔶 i 🔯 - 🖾 i 💽 🚺 🖉 💷 🔛 👪 🚜 🖧 🔗 i 🛙						
Geo	statistical Analy	yst 🕶 🐠 💂		_		
	Explore Data		►		Histogram	
1	Geostatistical	Wizard		2	Normal QQPlot	
- 53	Subset Featu	res		۰	Trend Analysis	
?	Tutorial			<b>8</b>	Voronoi Map	
?	Geostatistical	l Analyst Help		<i>6</i>	Semivariogram/Covarianc	e Cloud
				2	General QQPlot	
				9	Crosscovariance Cloud	

### Geostatistical Analyst Toolbar: AWESOME



LOW\_PER M\_ MI M\_WOU HIGH\_M DailyLiters





### Fit an OLS model

🛐 Ordinary Least Squares		x
Input Feature Class	Ordinary Least Squares	*
Sample_Taxlots_Summer_xycoord		
Unique ID Field	Performs global Ordinary Least Squares	
ID 🗸	(OLS) linear regression to generate	
Output Feature Class	predictions or to model a dependent	
I:\Students\Instructors\changh\Betsy\Thesis\GWR\OLS\OLS_bldg.shp	variable in terms of its relationships to a	
Dependent Variable	set of explanatory variables. Results are accessible from the Results window.	
SqrtLiters 🗸	accessible noni the Results window.	
Explanatory Variables		
BLDG_AGE		
LANDVAL		
E BLDGVAL	100 -	
TOTALVAL	100-	
LOTAREA_M	80 - 9	
BLDGAREA_M	60 -	
IMP_PER		
HIGH_PER	40-	
LOW PER	20- Observed Values (y)	
	Predicted Values (ŷ)	
Select All Unselect All Add Field	0 20 40 60 80 100	
Output Options     Coefficient Output Table (optional)		
I:\Students\Instructors\changh\Betsy\Thesis\GWR\OLS\OLS_bldg_coefficients		
Diagnostic Output Table (optional)		
I:\Students\Instructors\changh\Betsy\Thesis\GWR\OLS\OLS_bldg_diagnostics		
	Ŧ	Ŧ
OK Cancel Environments << Hide Help	p Tool Help	

Results	
🖃 🔂 Curr	rent Session
ė- <u>A</u> (	Ordinary Least Squares [191942_05232013]
	🖸 Output Feature Class: OLS_bldg.shp
	Coefficient Output Table: OLS_bldg_coefficients
	Diagnostic Output Table: OLS_bldg_diagnostics
÷	Inputs
Đ-P	F Environments
	1 Messages
	Executing: OrdinaryLeastSquares Sample_Taxlots_Summer_xycoord ID I:\Students\Instructors\changh\Betsy\Thesis\GWR
	Start Time: Thu May 23 19:19:36 2013
	Running script OrdinaryLeastSquares
	Summary of OLS Results
	Uariable Coefficient StdError t-Statistic Probability Robust_SE Robust_t Robust_Pr
	Intercept 19.415272 1.100009 17.650108 0.000000* 1.485714 13.067976 0.000000*
	BLDGAREA_M 0.078753 0.006507 12.103116 0.000000* 0.009660 8.152434 0.000000*
	OLS Diagnostics
	Number of Observations: 460 Number of Variables: 2
	Degrees of Freedom: 458 Akaike's Information Criterion (AIC) [2]: 3287.210420
	Multiple R-Squared [2]: 0.242331 Adjusted R-Squared [2]: 0.240676
	Joint F-Statistic [3]: 146.485414 Prob(>F), (1,458) degrees of freedom: 0.000000*
	Joint Wald Statistic [4]: 66.462176 Prob(>chi-squared), (1) degrees of freedom: 0.000000*
	Koenker (BP) Statistic [5]: 13.687422 Prob(>chi-squared), (1) degrees of freedom: 0.000216*
	Jarque-Bera Statistic [6]: 49.307420 Prob(>chi-squared), (2) degrees of freedom: 0.000000*
	Notes on Interpretation
	Notes on Interpretation
	statistically significant at the 0.05 level.
	[1] Large VIF (> 7.5, for example) indicates explanatory variable redundancy.
	[2] Measure of model fit/performance.
	[1] Significant p-value indicates overall model significance.
	[4] Significant p-value indicates robust overall model significance.
	[5] Significant p-value indicates biased standard errors; use robust estimates.
	[6] Significant p-value indicates residuals deviate from a normal distribution.
	WARNING 000851: Use the Spatial Autocorrelation (Moran's I) Tool to ensure residuals are not spatially autocorrelated.
	Writing Coefficient Output Table
	I:\Students\Instructors\changh\Betsy\Thesis\GWR\OLS\OLS_bldg_coefficients.dbf
	Writing Diagnostic Output Table
	I:\Students\Instructors\changh\Betsy\Thesis\GWR\OLS\OLS_bldg_diagnostics.dbf
	Completed script OrdinaryLeastSquares Succeeded at Thu May 23 19:19:42 2013 (Elapsed Time: 6.00 seconds)
SI 18	Sourcected of the May 20 19:19:42 2010 (Elabsed Time: 0:00 Seconds)

# A properly specified OLS model

#### Coefficients

- Expected sign
- Significant, p-value < 0.05</li>
- No multicollinearity (VIF)

#### • Goodness of fit

- Model significance F-statistic
- Goodness of fit: R<sup>2</sup>, AICc

#### Summary of OLS Results

 Variable
 Coefficient StdError t-Statistic Probability Robust\_SE Robust\_t Robust\_Pr

 Intercept 19.415272
 1.100009 17.650108
 0.000000\*
 1.485714
 13.067976
 0.000000\*

 BLDGAREA\_M 0.078753
 0.006507 12.103116
 0.000000\*
 0.009660
 8.152434
 0.000000\*

	Number of Observations:	460	Number of Variables: 2	2
	Degrees of Freedom:	458	Akaike's Information Criterion (AIC) [2	]: 3287.210420
[i]	Multiple R-Squared [2]:	0.242331	Adjusted R-Squared [2]:	0.240676
	Joint F-Statistic [3]: 14	6.485414	Prob(>F), (1,458) degrees of freedom:	0.000000*
(i)	Joint Wald Statistic [4]: (	56.462176	Prob(>chi-squared), (1) degrees of fr	eedom: 0.000000*

#### • Residuals ~ N (0, σ)

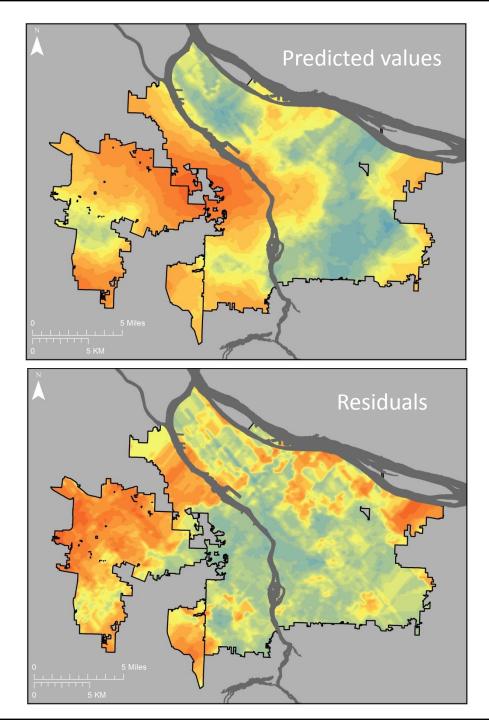
- N: normally distributed
  - Jarque-Bera p-value > 0.05
- 0: mean = zero
- σ: constant variance
  - Breusch-Pagen p-value > 0.05

#### Residuals independent

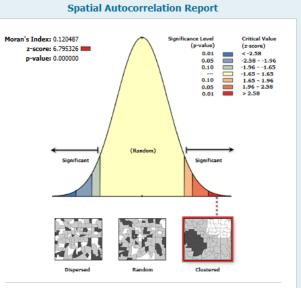
Moran's I p-value > 0.05

 Image: Statistic [5]: 13.687422
 Prob(>chi-squared), (1) degrees of freedom: 0.000216\*

 Image: Statistic [6]: 49.307420
 Prob(>chi-squared), (2) degrees of freedom: 0.000000\*



#### Residuals independent? No.



Given the z-score of 6.80, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

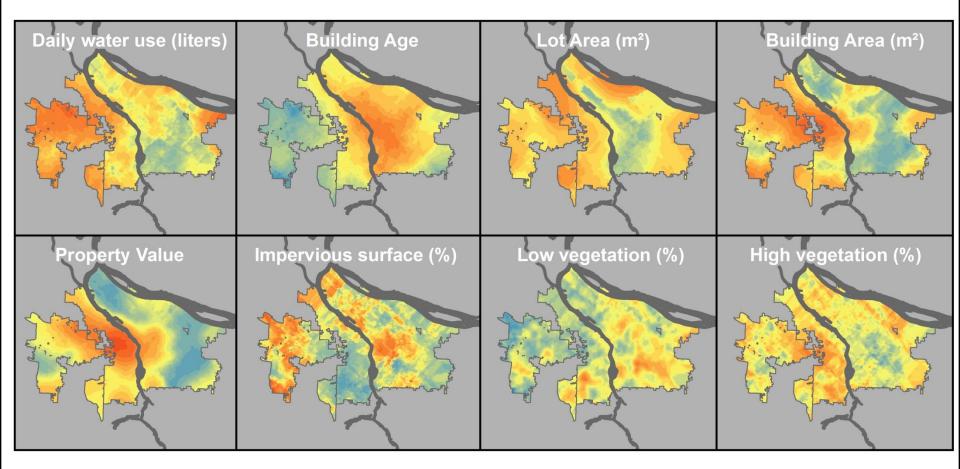
#### **Global Moran's I Summary**

Moran's Index:	0.120487
Expected Index:	-0.002179
Variance:	0.000326
z-score:	6.795326
p-value:	0.000000

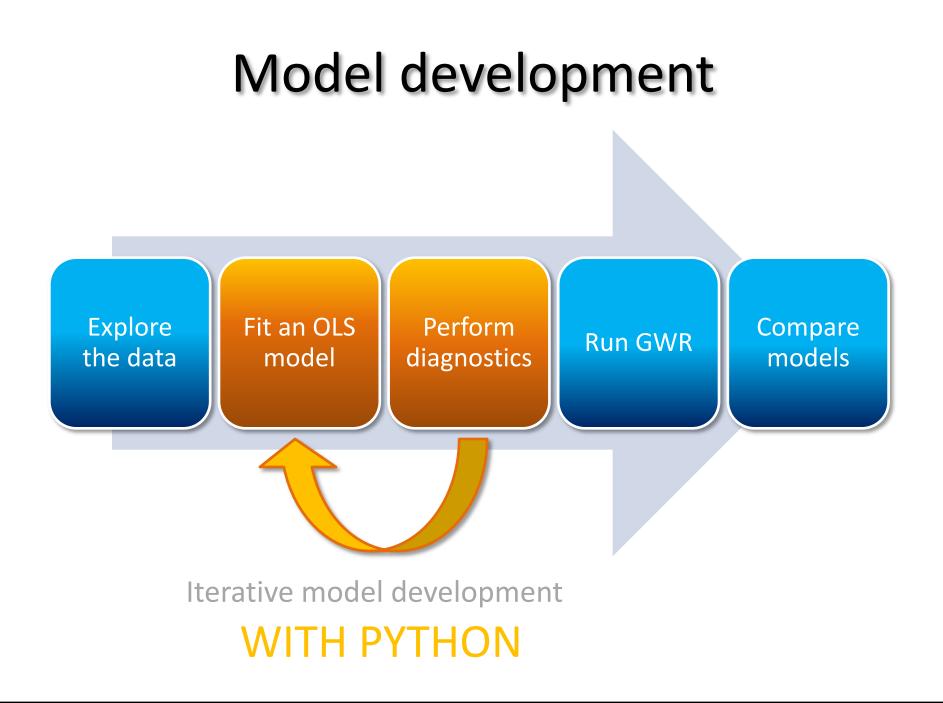
#### Dataset Information

Input Feature Class:	OLS_bldg
Input Field:	RESIDUAL
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN
Row Standardization:	False
Distance Threshold.:	11948.6849698
Weights Matrix File:	None

### What combination of variables makes a properly specified OLS model?



7! = 5040 possible models



#### Exploratory Regression

Input Feature Class	
Sample_Taxlots_Summer_thiessenClip	- 2
Dependent Variable	
SgrtLiters	-
Independent Variables	
BLDG_AGE	
V LANDVAL	
BLDGVAL	=
V TOTALVAL	
V LOTAREA_M	
BLDGAREA M	
₩ IMP_PER	
V HIGH_PER	
V LOW PER	-
< III	•
Select All Unselect All	Add Field
Input Spatial Weights Matrix	
I: \Students \Instructors \changh \Betsy \Thesis \GWR \SWM \InverseDistance.swm	<b>2</b>
Output Report File	
I:\Students\Instructors\changh\Betsy\Thesis\GWR\InverseDistance.txt	
Output Table Workspace (optional)	
Tables to Create	
MAX_NUMBER_OF_EXPLANATORY_VARIABLES_ONLY	· · · · · · · · · · · · · · · · · · ·
* Search Criteria	
Max Number of Explanatory Variables	
5	
1	20
Min Number of Explanatory Variables	
1	
1	20
Min Adj. R-Squared	
	0.5
Max Coefficient p-value	
	0.05
Max VIF Value	
	7.5
Min Jarque-Bera p-value	
	0.1
Min Spatial Autocorrelation p-value	
	0.1
OK Cancel Environments	<< Hide Help

#### Exploratory Regression

Tool Help

The Exploratory Regression tool evaluates all of the possible combinations of the input candidate explanatory variables, looking for models that best explain the dependent variable within the context of user-specified criteria. It produces an output report file and optional tables. A full explanation of each output is provided in the Interpreting Exploratory Regression Results document, found in the Documentation folder of the Supplementary Spatial Statistics folder. This tool uses Ordinary Least Squares (OLS) and Spatial Autocorrelation (Global Moran's I), both from the Spatial Statistics toolbox.

- -

X

# **Exploratory regression**

- Tests all possible combinations of explanatory variables
- All candidate models tested against criteria for a properly specified OLS model
  - Includes Moran's I on residuals
    - Spatial weights matrix
- Summary statistics on candidate models:
  - Frequency of variable significance
  - Collinear explanatory variables
  - Models with highest goodness of fit
  - Models with normally distributed residuals
  - Models with spatially uncorrelated residuals

# **Exploratory regression**

- Built into ArcGIS 10.1
  - Also new: Incremental Spatial Autocorrelation
- Both tools available for ArcGIS 10.0
  - Supplementary Spatial Statistics toolbox
  - <u>http://esriurl.com/spatialstats</u>
  - Python is available, too.

### Using Exploratory Regression in ArcGIS: Possible workflow

- 1. Calculate Distance Band from Neighbor Count
- 2. Incremental Spatial Autocorrelation
  - Bandwidth potentially useful for mapping clusters
- 3. Generate Spatial Weights Matrix
- 4. Exploratory Regression
  - Ordinary least squares regression
  - Geographically weighted regression

Source: ESRI Spatial Statistics Best Practices

### Calculate distance band from Neighbor Count

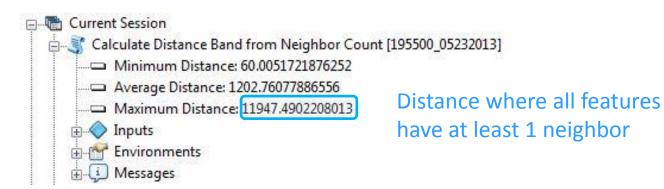
St Calculate Distance Band from Neighbor Count	
Input Features	Calculate Distance Band from
Sample_Taxlots_Summer_xycoord	Neighbor Count
Neighbors       5       1       9999       Distance Method       EUCLIDEAN_DISTANCE	Returns the minimum, the maximum, and the average distance to the specified Nth nearest neighbor (N is an input parameter) for a set of features. Results are accessible from the Results window.
	1 Neighbor Distance Band - 5.2 miles
	2 Neighbors Distance Band - 9.5 miles
OK Cancel Environments << Hide Help	Tool Help

# Current Session Calculate Distance Band from Neighbor Count [195500\_05232013] Minimum Distance: 60.0051721876252 Average Distance: 1202.76077886556 Maximum Distance: 11947.4902208013 Inputs Environments

🗄 🤳 Messages

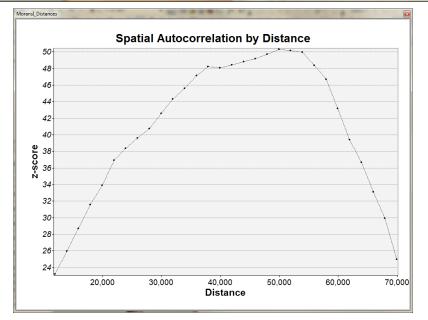
### Calculate distance band from Neighbor Count

Calculate Distance Band from Neighbor Count	Et cont		
Input Features		Calculate Distance Band from	
Sample_Taxlots_Summer_xycoord Neighbors	- 🖻	Neighbor Count	
I Distance Method EUCLIDEAN_DISTANCE	9999	Returns the minimum, the maximum, and the average distance to the specified Nth nearest neighbor (N is an input parameter) for a set of features. Results are accessible from the Results window.	
		1 Neighbor Distance Band - 5.2 miles	
	*	2 Neighbors Distance Band - 9.5 miles	
OK Cancel Environments.	. << Hide Help	Tool Help	



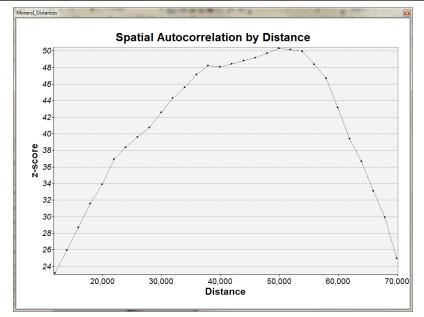
### **Incremental Spatial Autocorrelation**

Incremental Spatial Autocorrelation	
Input Feature Class Sample_Taxlots_Summer_xycoord	Incremental Spatial
	Autoconclution
SqrtLiters 🗸	Measures spatial autocorrelation at
Number of Distance Bands	incremental distances and creates a graph
30 /	of those distances and their corresponding
2 30	z-scores. The graph can be used to choose an appropriate scale of analysis
Beginning Distance (optional) 11947.4902208013	(distance band) to use for further analysis
Distance Increment (optional)	for instance in a Hot Spot Analysis (Getis-
2000	Ord Gi*). Peaks in the output graph
Distance Method (optional)	indicate distances at where clustering is
EUCLIDEAN	most pronounced. When more than one peak is present, clustering is pronounced
Row Standardization (optional)	at each of those distances. Select the
Output Table (optional)	distance that best corresponds to the
I:\Students\Instructors\changh\Betsy\Thesis\GWR\SWM\IncrementalMoransI.dbf	scale of analysis you are interested in;
	often this is the first peak encountered.
☑ Display Results Graphically (optional)	
v	Ŧ
OK Cancel Environments << Hide Help	Tool Help



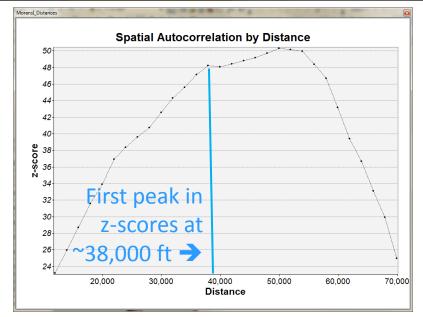
#### **STEP 2** Incremental Spatial Autocorrelation

Input Feature Class		Incremental Spatial
Sample_Taxlots_Summer_xycoord		Autocorrelation
Input Field		
SqrtLiters	-	Measures spatial autocorrelation at
Number of Distance Bands		incremental distances and creates a gra
30	0	of those distances and their corresponding
2	30	z-scores. The graph can be used to
Beginning Distance (optional)		choose an appropriate scale of analysis
	11947.4902208013	(distance band) to use for further analysi
Distance Increment (optional)		for instance in a Hot Spot Analysis (Geti
	2000	Ord Gi*). Peaks in the output graph
Distance Method (optional)		indicate distances at where clustering is most pronounced. When more than one
EUCLIDEAN	-	peak is present, clustering is pronounced
Row Standardization (optional)		at each of those distances. Select the
		distance that best corresponds to the
Output Table (optional)		scale of analysis you are interested in;
I:\Students\Instructors\changh\Betsy\Thesis\GWR\SWM\Increment	alMoransI.dbf 🛛 🔁	often this is the first peak encountered.
Display Bas the Graphically (antional)	_	· · · ·
Display Results Graphically (optional)		
		r



#### **STEP 2** Incremental Spatial Autocorrelation

Input Feature Class		Incremental Spatial
Sample_Taxlots_Summer_xycoord	I 🖻	Autocorrelation
Input Field		
SqrtLiters	-	Measures spatial autocorrelation at
Number of Distance Bands		incremental distances and creates a gra
30	0	of those distances and their correspondi
2	30	z-scores. The graph can be used to
Beginning Distance (optional)		choose an appropriate scale of analysis
	11947.4902208013	(distance band) to use for further analysi
Distance Increment (optional)		for instance in a Hot Spot Analysis (Geti
	2000	Ord Gi*). Peaks in the output graph
Distance Method (optional)		indicate distances at where clustering is
EUCLIDEAN	-	most pronounced. When more than one peak is present, clustering is pronounce
Row Standardization (optional)		at each of those distances. Select the
,		distance that best corresponds to the
Output Table (optional)		scale of analysis you are interested in;
I:\Students\Instructors\changh\Betsy\Thesis\GWR\SWM\Incremental	MoransI.dbf 🔂 🔁	often this is the first peak encountered.
Disalas Davita Carabiantha (artisanth	_	
Display Results Graphically (optional)		
		-



### Generate spatial weights matrix

Generate Spatial Weights Matrix	0.00.00	
Input Feature Class I:\Students\Instructors\changh\Betsy\Thesis\GWR\Sample_Taxlots_Summer_xycoord.shp		Conceptualization of Spatial Relationships
Unique ID Field ID		Specifies how spatial relationships among
Output Spatial Weights Matrix File		features are conceptualized.
I:\Students\Instructors\changh\Betsy\Thesis\GWR\SWM\Inverse_Distance.swm Conceptualization of Spatial Relationships	Ċ	INVERSE_DISTANCE—The impact
INVERSE_DISTANCE Distance Method (optional)		of one feature on another feature decreases with distance.
EUCLIDEAN Exponent (optional)	-	<ul> <li>FIXED_DISTANCE—Everything within a specified critical distance of</li> </ul>
Threshold Distance (optional)	1	each feature is included in the analysis; everything outside the
Number of Neighbors (optional)	38000	critical distance is excluded. • K NEAREST NEIGHBORS—The
		closest "k" features are included in the analysis; k is a specified
Row Standardization (optional) Input Table (optional)		numeric parameter. • CONTIGUITY EDGES ONLY—
		Polygon features that share a boundary are neighbors.
		CONTIGUITY_EDGES_CORNERS—     Polygon features that share a
		boundary and/or share a node are neighbors.
		DELAUNAY_TRIANGULATION—A
		mesh of nonoverlapping triangles is created from feature centroids;
		features associated with triangle nodes that share edges are
		<ul> <li>neighbors.</li> <li>CONVERT_TABLE—Spatial relationships are defined in a table.</li> </ul>
		Note: Polygon Contiguity methods are only available with an ArcInfo license.
	Ŧ	۰
OK Cancel Environments	<< Hide Help	Tool Help

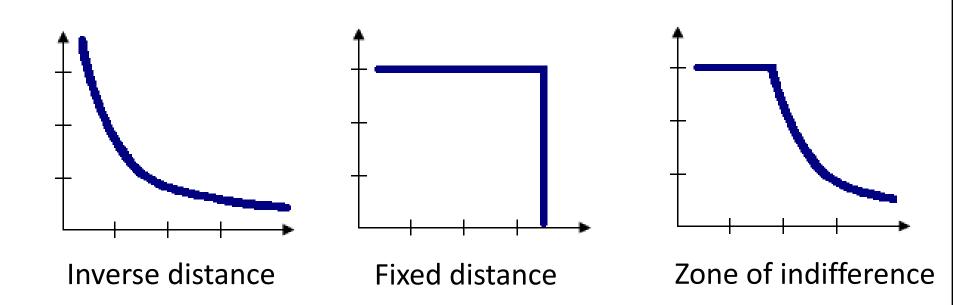
### Generate spatial weights matrix

Generate Spatial Weights Matrix	
Input Feature Class I:\Students\Instructors\changh\Betsy\Thesis\GWR\Sample_Taxlots_Summer_xycoord.shp Unique ID Field	Conceptualization of Spatial Relationships
ID U U U U U U U U U U U U U U U U U U U	Specifies how spatial relationships among features are conceptualized.  INVERSE_DISTANCE—The impact of one feature on another feature decreases with distance.  FIXED DISTANCE—Everything
Exponent (optional)  Threshold Distance (optional)  Number of Neighbors (optional)  Row Standardization (optional)  from Stand 2	<ul> <li>within a specified critical distance of each feature is included in the analysis; everything outside the critical distance is excluded.</li> <li>K_NEAREST_NEIGHBORS—The closest "k" features are included in the analysis; k is a specified numeric parameter.</li> </ul>
Input Table (optional)	<ul> <li>CONTIGUITY_EDGES_ONLY— Polygon features that share a boundary are neighbors.</li> <li>CONTIGUITY_EDGES_CORNERS— Polygon features that share a boundary and/or share a node are neighbors.</li> <li>DELAUNAY_TRIANGULATION—A mesh of nonoverlapping triangles is created from feature centroids; features associated with triangle nodes that share edges are neighbors.</li> <li>CONVERT_TABLE—Spatial relationships are defined in a table.</li> </ul>
OK     Cancel     Environments     << Hide Help	Note: Polygon Contiguity methods are only available with an ArcInfo license.

### Generate spatial weights matrix

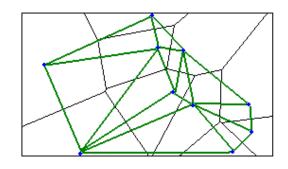
Generate Spatial Weights Matrix	
Input Feature Class	Conceptualization of Spatial
I:\Students\Instructors\changh\Betsy\Thesis\GWR\Sample_Taxlots_Summer_xycoord.shp	Relationships
Unique ID Field	
ID 🗸	Opecilies now spatial relationships among
Output Spatial Weights Matrix File	features are conceptualized.
I:\Students\Instructors\changh\Betsy\Thesis\GWR\SWM\Inverse_Distance.swm	
Conceptualization of Spatial Relationships	INVERSE_DISTANCE—The impact
INVERSE_DISTANCE	
Distance Method (optional)	decreases with distance.
EUCLIDEAN	
Exponent (optional)	within a specified critical distance of each feature is included in the
	analysis; everything outside the
Threshold Distance (optional) 38000	
Number of Neighbors (optional)	K NEAREST NEIGHBORS—The
Input is the bandwidth 🛧	
	the analysis; k is a specified
Row Standardization (optional)	numeric parameter.
Input Table (optional)	CONTIGUITY_EDGES_ONLY—
Input Table (optional)	Polygon features that share a
	boundary are neighbors.
	CONTIGUITY_EDGES_CORNERS—     Polygon features that share a
	boundary and/or share a node are
	neighbors.
	DELAUNAY TRIANGULATION—A
	mesh of nonoverlapping triangles is
	created from feature centroids;
	features associated with triangle
	nodes that share edges are
	neighbors.
	CONVERT_TABLE—Spatial
	relationships are defined in a table.
	Note: Debuger Continuity with 1
	Note: Polygon Contiguity methods are only available with an ArcInfo license.
	available with an Arcinio license.
	▼ <b>(</b> Ⅲ <b>)</b>
OK Cancel Environments << Hide He	Tool Help

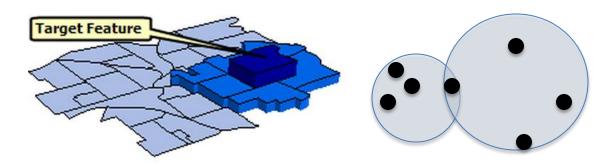
# Conceptualization of spatial relationships



http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000005000000

# Conceptualization of spatial relationships





Delaunay Triangulation

Polygon Contiguity

K nearest neighbors

See also: Space-time clustering, Grouping Analysis tool

Exploratory Regression

- 0 X

	A REAL PROPERTY AND	
Input Feature Class	^	Exploratory Regression
Sample_Taxlots_Summer_thiessenClip	- 2	
Dependent Variable		The Exploratory Regression tool evaluates all of the
SqrtLiters	-	possible combinations of the input candidate
Independent Variables		explanatory variables, looking for models that best
BLDG_AGE	*	explain the dependent variable within the context of
V LANDVAL		user-specified criteria. It produces an output report file
BLDGVAL	=	and optional tables. A full explanation of each output is provided in the Interpreting Exploratory Regression
V TOTALVAL		Results document, found in the Documentation folder
V LOTAREA_M		of the Supplementary Spatial Statistics folder. This
BLDGAREA_M		tool uses Ordinary Least Squares (OLS) and Spatial
IMP_PER		Autocorrelation (Global Moran's I), both from the
V HIGH_PER		Spatial Statistics toolbox.
V LOW PER	· · ·	
Select All Unselect All	Add Field	
Input Spatial Weights Matrix		
I:\Students\Instructors\changh\Betsy\Thesis\GWR\SWM\InverseDistance.swm		
Output Report File		
I:\Students\Instructors\changh\Betsy\Thesis\GWR\InverseDistance.txt	2	
Output Table Workspace (optional)		
	<b>1</b>	
MAX_NUMBER_OF_EXPLANATORY_VARIABLES_ONLY	× ]	
* Search Criteria		
Max Number of Explanatory Variables		
5		
1	20	
Min Number of Explanatory Variables		
1		
1	20	
Min Adj. R-Squared	0.5	
New Coefficient e velue	0.5	
Max Coefficient p-value	0.05	
Max VIF Value	0.05	
	7.5	
Min Jarque-Bera p-value	7.5	
rim burgae bera prvalae	0.1	
Min Spatial Autocorrelation p-value	011	
	0.1	
OK Cancel Environments	<< Hide Help	Tool Help

Sample_Taxlots_Summer_thiessenClip		Exploratory Regression
Dependent Variable		The Exploratory Regression too
SqrtLiters	_	possible combinations of the in
Independent Variables	•	explanatory variables, looking for
BLDG AGE		explain the dependent variable v
V LANDVAL	- All	user-specified criteria. It produc
BLDGVAL	=	and optional tables. A full expla
V TOTALVAL		is provided in the Interpreting Ex
V TOTALVAL		Results document, found in the
BLDGAREA_M		of the Supplementary Spatial S
		tool uses Ordinary Least Squar
		Autocorrelation (Global Moran's
V HIGH_PER	-	Spatial Statistics toolbox.
V LOW PER		
	,	
Select All Unselect All	Add Field	
Input Spatial Weights Matrix		
I: \Students \Instructors \changh \Betsy \Thesis \GWR \SWM \InverseDistance.swm		
I: \Students\Instructors\changh\Betsy\Thesis\GWR\InverseDistance.txt		
Output Table Workspace (optional)		
Tables to Create		
MAX_NUMBER_OF_EXPLANATORY_VARIABLES_ONLY		
	1	
Search Criteria Max Number of Explanatory Variables		
5		
1	20	
Min Number of Explanatory Variables		
1		
1	20	
Min Adj. R-Squared		
	0.5	
Max Coefficient p-value		
	0.05	
Max VIF Value		
	7.5	
Min Jarque-Bera p-value		
	0.1	
Min Spatial Autocorrelation p-value		

aluates all of the candidate odels that best in the context of an output report file ion of each output ratory Regression cumentation folder stics folder. This OLS) and Spatial both from the

- • ×

#### .swm file **→**

St Exploratory Regression

Exploratory Regression

- 0 X

	Input Feature Class	^	Exploratory Regression
	Sample_Taxlots_Summer_thiessenClip	- 🖻	
	Dependent Variable		The Exploratory Regression tool evaluates all of the
	SqrtLiters	-	possible combinations of the input candidate
	Independent Variables		explanatory variables, looking for models that best
Candidate	BLDG_AGE	<u> </u>	explain the dependent variable within the context of
	I LANDVAL		user-specified criteria. It produces an output report file
explanatory	BLDGVAL	=	and optional tables. A full explanation of each output is provided in the Interpreting Exploratory Regression
capitalitatory	TOTALVAL		Results document, found in the Documentation folder
variables	LOTAREA_M		of the Supplementary Spatial Statistics folder. This
Valiables	BLDGAREA_M		tool uses Ordinary Least Squares (OLS) and Spatial
	IMP_PER		Autocorrelation (Global Moran's I), both from the
	HIGH_PER	_	Spatial Statistics toolbox.
	LOW PER		
	Select All Unselect All	Add Field	
	Input Spatial Weights Matrix		
.swm file 🗲	I: \Students\Instructors\changh\Betsy\Thesis\GWR\SWM\InverseDistance.swm		
	I:\Students\Instructors\changh\Betsy\Thesis\GWR\InverseDistance.txt		
	Output Table Workspace (optional)		
		<b>2</b>	
	Tables to Create		
	MAX_NUMBER_OF_EXPLANATORY_VARIABLES_ONLY	<b>T</b>	
	* Search Criteria		
	Max Number of Explanatory Variables		
	5		
	1	20	
	Min Number of Explanatory Variables		
	1		
	1	20	
	Min Adj. R-Squared	0.5	
	May Caefficient a unive	0.5	
	Max Coefficient p-value	0.05	
	Max VIF Value	0.05	
	יומג ענו למוטב	7.5	
	Min Jarque-Bera p-value	7.5	
		0.1	
	Min Spatial Autocorrelation p-value		
		0.1	
			Ψ
	OK Cancel Environments	<< Hide Help	Tool Help

Exploratory Regression

		The set of
	Input Feature Class	Exploratory Regression
	Sample_Taxlots_Summer_thiessenClip	
	Dependent Variable	<ul> <li>The Exploratory Regression tool evaluates all of the possible combinations of the input candidate</li> </ul>
	SqrtLiters Independent Variables	explanatory variables, looking for models that best
Candidate	BLDG_AGE	explain the dependent variable within the context of
Curranduce	LANDVAL	user-specified criteria. It produces an output report file
explanatory	✓ BLDGVAL	and optional tables. A full explanation of each output is provided in the Interpreting Exploratory Regression
	V TOTALVAL	Results document, found in the Documentation folder
variables	BLDGAREA_M	of the Supplementary Spatial Statistics folder. This
	₩ IMP_PER	tool uses Ordinary Least Squares (OLS) and Spatial Autocorrelation (Global Moran's I), both from the
	HIGH_PER	_ Spatial Statistics toolbox.
	LOW PER	
	Select All Unselect All Add Field	
.swm file 🗲	Input Spatial Weights Matrix I: \Students \Instructors \changh \Betsy \Thesis \GWR \SWM \InverseDistance.swm	
	1: Students this actors (changingersy (mess (swik (swik (swik (swik)))))	
	I:\Students\Instructors\changh\Betsy\Thesis\GWR\InverseDistance.txt	
	Output Table Workspace (optional)	
	Tables to Create	
Min/max	MAX_NUMBER_OF_EXPLANATORY_VARIABLES_ONLY	<b>•</b>
iviiii/iiiax	* Search Criteria	
number of	Max Number of Explanatory Variables	
number of	1 20	
variables	Min Number of Explanatory Variables	
	1	
per model	1 20 Min Adj. R-Squared	
		0.5
	Max Coefficient p-value	
		0.05
Significance	Max VIF Value	7.5
-	Min Jarque-Bera p-value	
test p-values		0.1
•	Min Spatial Autocorrelation p-value	0.1
		v
	OK Cancel Environments	de Help Tool Help
		loon loop

**Exploratory regression** 

Running script ExploratoryRegression... Choose 1 of 12 Summary Highest Adjusted R-Squared Results R2 AICc JB BP VIF MI Model 0.24 3289.26 0.00 0.00 1.00 0.00 + BLDGAREA M\*\*\* 0.17 3332.10 0.00 0.00 1.00 0.00 + LANDVAL\*\*\* 0.16 3338.06 0.00 0.14 1.00 0.00 -BLDG\_AGE\*\*\* Passing Models R2 AICc JB BP VIF MI Model Passing models would be here i Choose 2 of 12 Summary (i) Highest Adjusted R-Squared Results R2 AICc JB BP VIF MI Model 0.31 3248.04 0.00 0.05 1.10 0.00 -BLDG\_AGE\*\*\* +BLDGAREA\_M\*\*\* 0.28 3264.27 0.00 0.00 1.01 0.00 -BLDG\_AGE\*\*\* + TOTALVAL\*\*\* 0.28 3268.99 0.00 0.00 1.22 0.00 + BLDGAREA\_M\*\*\* + IMP\_M\*\*\* Passing Models R2 AICc JB BP VIF MI Model I Choose 3 of 12 Summary Highest Adjusted R-Squared Results R2 AICc JB BP VIF MI Model 0.33 3236.40 0.00 0.00 1.50 0.00 -BLDG\_AGE\*\*\* +LANDVAL\*\* +BLDGAREA\_M\*\*\* 0.32 3240.67 0.00 0.09 2.15 0.00 -BLDG\_AGE\*\*\* +TOTALVAL\*\* +BLDGAREA\_M\*\*\* 0.32 3241.65 0.00 0.05 1.36 0.00 -BLDG\_AGE\*\*\* +BLDGAREA\_M\*\*\* +IMP\_M\*\*\* (i) Passing Models R2 AICc JB BP VIF MI Model i T 1 Choose 4 of 12 Summary (i) Highest Adjusted R-Squared Results R2 AICc JB BP VIF MI Model 0.33 3231.96 0.00 0.00 1.60 0.00 -BLDG\_AGE\*\*\* +LANDVAL\*\* +BLDGAREA\_M\*\*\* +IMP\_M\*\* 0.33 3233.74 0.00 0.01 1.54 0.00 -BLDG AGE\*\*\* +LANDVAL\*\* +LOTAREA M\*\* +BLDGAREA M\*\*\* 0.33 3234.62 0.00 0.01 6.72 0.00 -BLDG AGE\*\*\* -BLDGVAL\*\* +TOTALVAL\*\* +BLDGAREA M\*\*\* 1 Passing Models I R2 AICc JB BP VIF MI Model

### Which diagnostic tests are not being passed?

\*\*\*\*\*\*\*\*\*\*\*\*\* Exploratory Regression Global Summary (SQRTLITERS) \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Percentage of Search Criteria Passed
 Search Criterion Cutoff Trials # Passed % Passed
 Min Adjusted R-Squared > 0.50 1585 0 0.00
 Max Coefficient p-value < 0.05 1585 283 17.85</li>
 Max VIF Value < 7.50 1585 1351 85.24</li>
 Min Jarque-Bera p-value > 0.10 1585 3 0.19
 Min Moran's I p-value > 0.10 18 6 33.33

### Which variables are

#### consistently significant?

- 🛄 Summary of V	ariable Significance
🔄 🕖 Variable 🕺 Si	gnificant
BLDG_AGE	100.00
LANDVAL	80.60
BLDGVAL	65.12
JOTALVAL	77.76
LOTAREA_M	72.95
BLDGAREA_M	100.00
IMP_PER	42.70
🔄 🕕 HIGH_PER	30.60
LOW_PER	33.63
IMP_M	97.51
LOW_M	48.22
HIGH_M	43.24

## **Exploratory regression**

#### Which models have normal residuals?

- Ó			Sum	mary o	f Resid	lual Norma	lity				
	JB	R2	AICc	BP	VIF	MI Mod	el				
	0.124366	0.1732	28 3331	.519326	0.0000	03 7.90975	L 0.000000	+BLDGVAL***	+LOTAREA_M*	** -LOW_PER	-LOW_M
	0.110102	0.1727	88 3330	720833	0.0000	00 2.976807	7 0.000000	+BLDGVAL***	+LOTAREA_M*	** -LOW_M***	*
	0.102163	0.1719	89 3331	164556	0.0000	23 1.046833	3 0.000000	+BLDGVAL***	+LOTAREA_M*	** -LOW_PER*	***
									10.		
	V	Vhi	ch n	nod	els	have s	spatia	ally unc	orrelated	l residu	als?
	V	Vhi	ch n	10d Sumn	els nary of	have s Residual A	spatia utocorrela	ally unc	orrelated	l residu	als?
	MI	R2	AICc	Sumn JB	nary of BP	Residual A VIF Mod	utocorrela del	ation			
	MI 0.223620	R2 0.3302	AICc 70 3234	Sumn JB .618079	BP 0.0000	Residual A VIF Mo 000 0.014348	utocorrela del 3 <mark>6.72409</mark>	ation 0 -BLDG_AGE*	** -BLDGVAL**	+TOTALVAL**	+BLDGAREA_M**
	MI 0.223620 0.320787	R2 0.3302 0.3341	AICc 70 3234 35 3231	Sumn JB .618079 .956051	0.0000	Residual A VIF Mo 000 0.014348 000 0.001452	utocorrela del 3 6.72409 2 1.59690	ation 0 -BLDG_AGE* 1 -BLDG_AGE*	** -BLDGVAL** ** +LANDVAL**	+TOTALVAL** +BLDGAREA_	

# **OLS regression: Estimated coefficients**

nput Feature Class		Ordinary Least Squares
Sample_Taxlots_Summer_thiessenClip	I 🔁	
nique ID Field		Performs global Ordinary Least Squares (OLS) linea
ID	-	regression to generate predictions or to model a
lutput Feature Class		dependent variable in terms of its relationships to a
I:\Students\Instructors\changh\Betsy\Thesis\GWR\OLS\OLS_blgdval_lotarea_lowm_thiessen.shp		set of explanatory variables. Results are accessible from the Results window.
ependent Variable		from the Results window.
SqrtLiters	-	
xplanatory Variables		
BLDG_AGE	*	<b>†</b>
LANDVAL .		19
V BLDGVAL	E	100-
TOTALVAL		80 - 9
V LOTAREA_M		
BLDGAREA_M		60-
IMP_PER		40-
HIGH_PER		20- Observed Values (y)
LOW PER	*	Predicted Values (ŷ)
	,	0 20 40 60 80 100
Select All Unselect All	Add Field	0 20 40 60 80 100
lutput Options		
Coefficient Output Table (optional)		
	e 1	
Diagnostic Output Table (optional)		
	e	
	-	

 Summary of OLS Results

 Variable Coefficient StdError t-Statistic Probability Robust\_SE Robust\_t Robust\_Pr VIF [1]

 Intercept 22.341917
 1.078627 20.713298
 0.000000\*
 1.344188
 16.621123
 0.000000\*
 ------ 

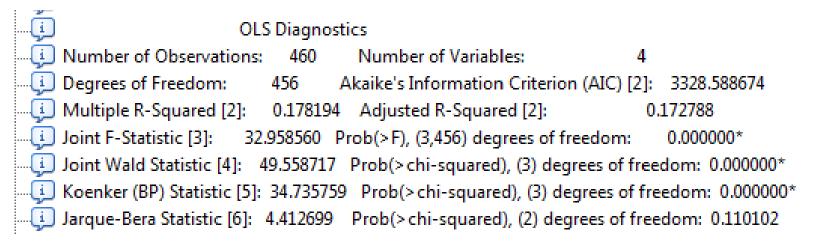
 BLDGVAL
 0.000025
 0.000005
 5.087609
 0.000001\*
 0.000006
 4.471951
 0.000012\*
 1.051748

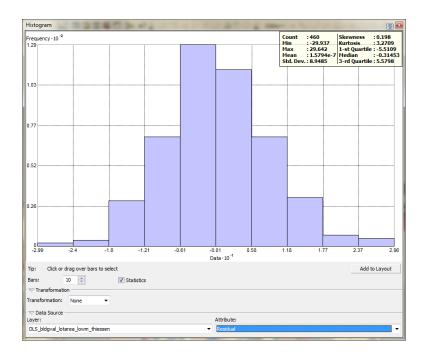
 LOTAREA\_M
 0.009847
 0.001469
 6.702138
 0.000000\*
 0.001925
 5.115796
 0.000001\*
 2.976807

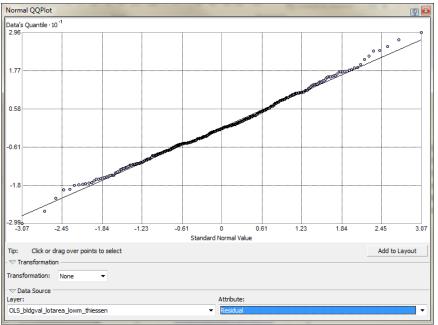
 LOW\_M
 -0.008885
 0.002490
 -3.567676
 0.000411\*
 0.002627
 -3.382256
 0.000795\*
 2.951698

#### Y = 22.34 + 0.00025(BLDGVAL) + 0.00985(LOTAREA\_M) - 0.008885(LOW\_M) + error

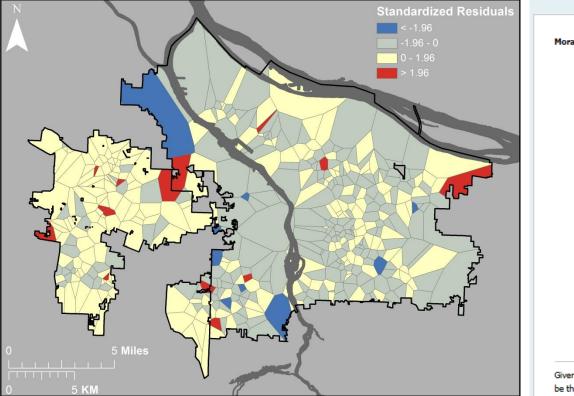
### **OLS regression: Residual normality**

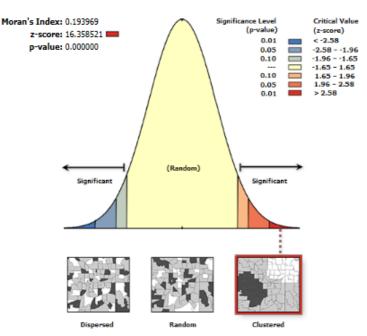






# **OLS regression: Residual dependence**



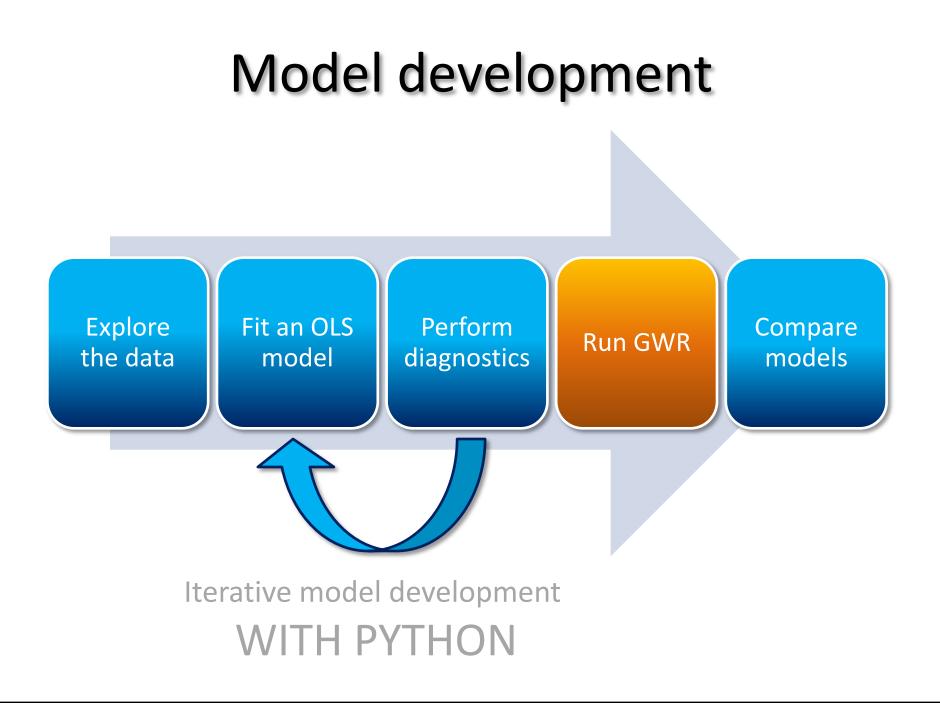


Spatial Autocorrelation Report

Given the z-score of 16.36, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Moran's Index:	0.193969
Expected Index:	-0.002179
Variance:	0.000144
z-score:	16.358521
p-value:	0.000000

#### Global Moran's I Summary



# **GWR: Setting kernel and bandwidth**

🤇 Geographically Weighted Regression			
Input features Sample_Taxlots_Summer_thiessenClip			Bandwidth method
Dependent variable			Specifies how the extent of the kernel should be determined. When AICc or
Explanatory variable(s)	•		CV are selected, the tool will find the optimal distance/neighbor parameter for you.
BLDGVAL LOTAREA_M LOW_M	<ul> <li>+</li> <li>×</li> <li>↑</li> <li>↓</li> </ul>	III	<ul> <li>AICc—The extent of the kernel is determined using the Akaike Information Criterion (AICc).</li> <li>CV—The extent of the kernel is determined using Cross Validation.</li> <li>BANDWIDTH PARAMETER—The extent of the kernel is determined by a fixed distance or</li> </ul>
Output feature class I:\Students\Instructors\changh\Betsy\Thesis\GWR\GWR_bldgval_lotarea_lowm_thiessen_aic.shp Kernel type ADAPTIVE	-		a fixed number of neighbors.
Bandwidth method BANDWIDTH PARAMETER AICc CV BANDWIDTH PARAMETER BANDWIDTH PARAMETER	30	Ţ	
OK Cancel Environments << H	lide Help		Tool Help

# **GWR: Output**

E messages	
Executing: GeographicallyWeightedRegression	
Start Time: Tue May 28 16:37:36 2013	
EffectiveNumber : 33.793550676073622	
🧓 Sigma : 7.8059865944561873	
AICc : 3218.5804496988549	

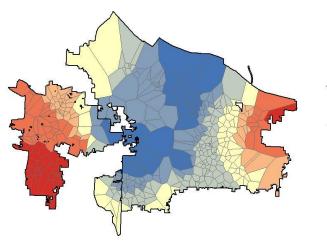
in Messages

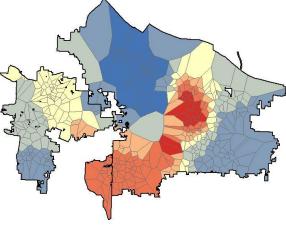
□ -	
GWR_bldgval_lotarea_lowm_thiessen_aic	×
Observed Cond LocalR2 Predicted Intercept C1_BLDG C2_LOTA C3_LO Residual StdError StdErr_Int StdErrC1 StdErrC2 StdErrC3 Std	Resi
▶ 53.739534 10.220 0.208487 53.572759 19.87490 0.000034 0.010397 -0.0094 0.166775 2.862581 2.337176 0.000011 0.002891 0.00416 0.0	5826
23.890516 7.2591 0.215906 27.938022 16.87395 0.000033 0.008712 -0.0011 -4.04750 7.569381 2.547647 0.000011 0.003393 0.006247 -0.	5347
33.052823 8.8667 0.168227 33.010089 17.64097 0.000054 0.003033 0.00435 0.042734 7.127231 3.175386 0.000016 0.003929 0.006976 0.0	0599
63.504146 9.7458 0.263538 41.802754 13.37000 0.00008 0.004307 0.00920 21.70139 6.518928 3.464891 0.000017 0.004406 0.007025 3.3	2898
38.798366 8.5593 0.111953 32.279123 22.30576 0.000017 0.009133 -0.0087 6.519243 7.269473 2.14511 0.000009 0.004395 0.007334 0.8	9679
44.229499 6.3124 0.16125 44.521959 27.86461 0.000022 0.00218 0.00594 -0.29246 5.747762 1.692899 0.000007 0.002721 0.00676 -0.	0508
20.397856 6.9361 0.241764 29.772157 25.08383 0.000036 0.002785 0.00316 -9.37430 7.594773 1.78285 0.000008 0.002709 0.005121 -1.	2343
45.690303 9.6165 0.229426 26.559803 14.67353 0.000067 0.005141 0.00815 19.1305 7.523589 3.292605 0.000014 0.004415 0.007004 2.5	4273
13.583328 8.9639 0.115431 25.634288 19.63844 0.000033 0.006154 -0.0001 -12.0509 7.558961 2.690557 0.000009 0.004108 0.006693 -1.	5942
24.759926 8.9757 0.090382 26.398506 22.65538 0.000017 0.008834 -0.0055 -1.63857 7.487637 2.061032 0.000009 0.004441 0.006892 -0.	2188 -
	•
$1 \leftarrow 1 \rightarrow 1$ $\square$	
GWR_bldgval_lotarea_lowm_thiessen_aic	

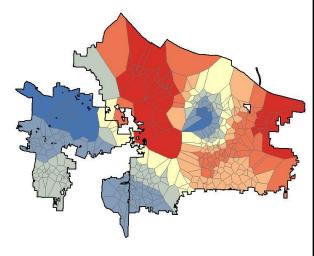
#### Full explanation of statistical output:

http://resources.esri.com/help/9.3/arcgisdesktop/com/gp\_toolref/spatial\_statistics\_tools/interpreting\_gwr\_results.htm

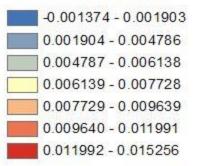
# **GWR: Coefficient estimates**



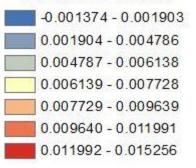




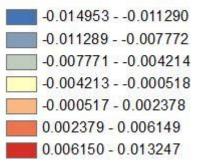
#### Building value Estimated coefficient



#### Lot area Estimated coefficient



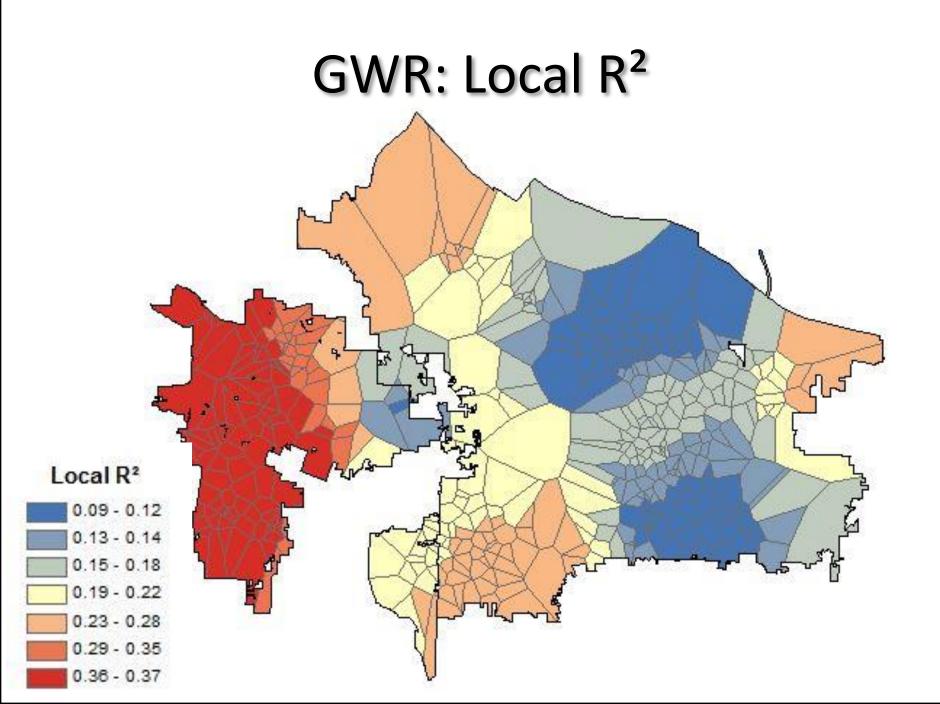
#### Low vegetation (m<sup>2</sup>) Estimated coefficient

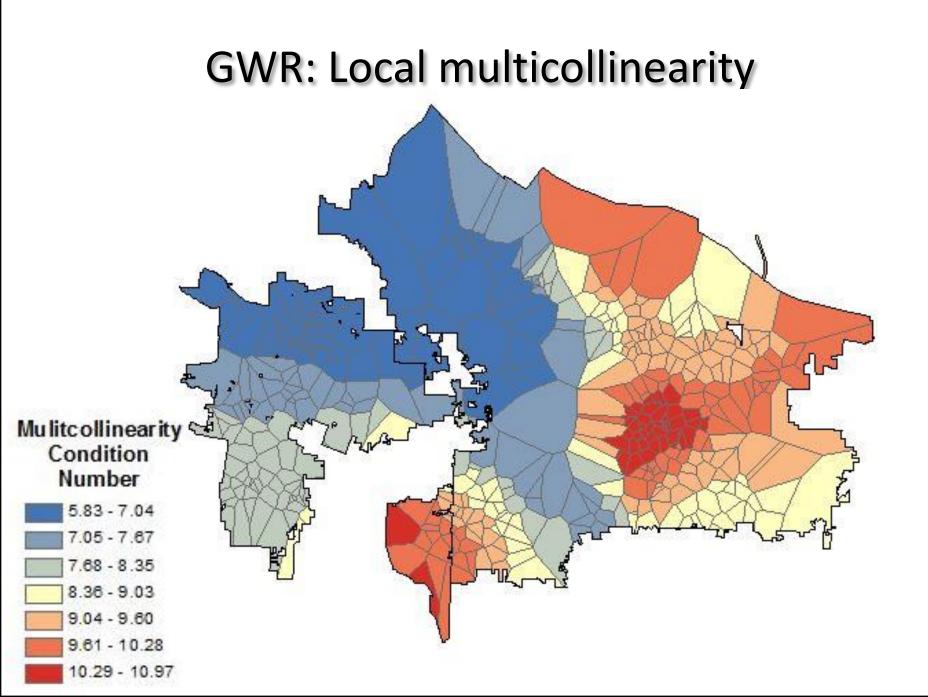


Global: 0.000025

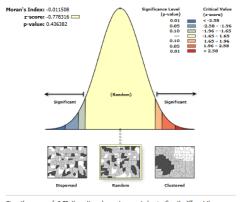
Global: 0.009847

Global: -0.008885

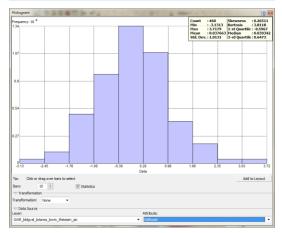


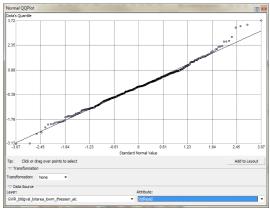


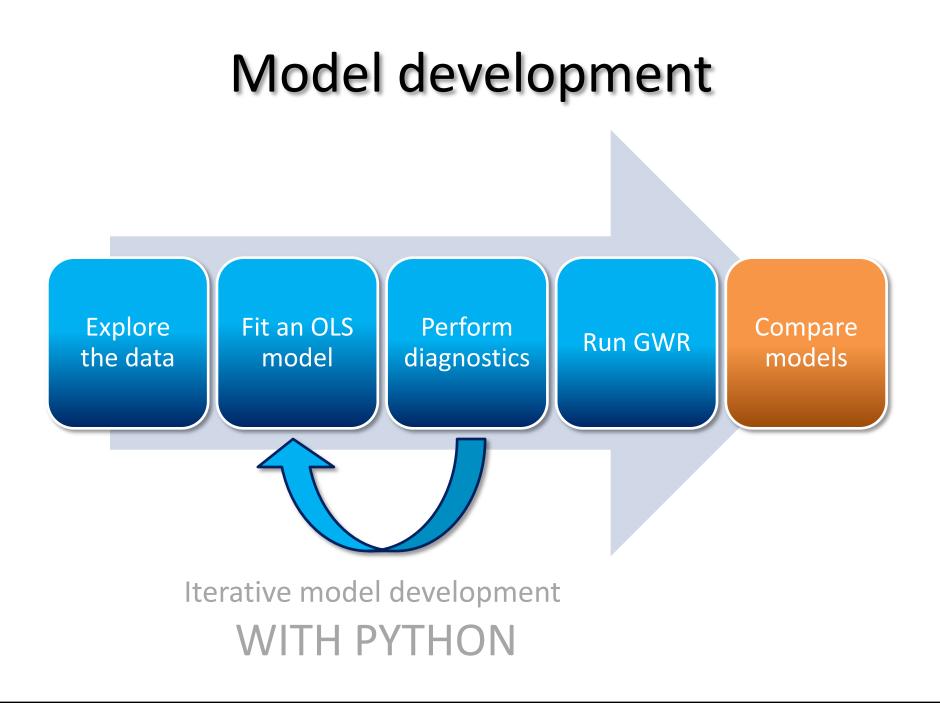
# **GWR: Standardized residuals** Standardized residuals -3.13 - -1.85 -1.84 - -0.91 -0.90 - -0.24 -0.23 - 0.40 0.41 - 1.15 1.16 - 2.13 2.14 - 3.72



Given the z-score of -0.78, the pattern does not appear to be significantly different than random.







# How to compare models?

- Compare with tables
  - Global OLS coefficient against distribution of GWR coefficients
  - Improvement in goodness of fit (R<sup>2</sup>, AICc)
  - Improvement in residual diagnostics
- ANOVA table to compare models
   GWR4 outputs this automatically
- Model validation?
  - Divide data into calibration and validation subsets.
  - Use prediction locations (under additional parameters) to compare predicted to observed

# **Calibration and validation**

- Create random points
  - Contraining feature = input data
    - Selects a random subset of points from input data
    - Save as your validation dataset
- Join random points back to input data
  - Select points that join, export points as validation dataset with attributes
  - Switch selection, export as calibration dataset
- Run GWR on calibration dataset
  - Include validation dataset as Prediction Locations
  - Add explanatory variables in same order

# A few tips for GWR

- Use only on a large dataset ( > 160 features)
- Continuous variables with no NULL values
- SEVERE MODEL DESIGN PROBLEMS
  - Local multicollinearity > 30
    - Center variables (subtracting out the mean)
- Residuals still autocorrelated?
  - Try including a distance-based variable
    - E.g. Distance to urban center, major arterial
    - Geostatistical Analyst, Explore data, trend analysis may assist in selection

# New directions in GWR research

- Flexible bandwidth
  - Why would we have a global bandwidth in a local model?
  - Allow each explanatory variable to have its own bandwidth
- Nonlinear models
- Non-Euclidean distance metrics
  - Manhattan distance, Network distance
  - Social, cost, aspatial 'distance'
- Check out GWR4 for the following functionality:
  - Semiparametric model some coefficients estimated globally, others locally
  - Poisson (for count data) or logistic (binary response)
  - Bisquare in addition to Gaussian kernel
  - Automatically compares OLS with GWR using ANOVA

Section 4

# **REFERENCES & RESOURCES**

# Software

### GWR4 software (Forthingham, University of St Andrews)

http://www.st-andrews.ac.uk/geoinformatics/gwr/gwr-software/

- GWR4 (free download)
- Link to various R packages

### ESRI software

http://esriurl.com/spatialstats

- Spatial statistics toolbox (for < 10.1) with python code
- Also, videos, tutorials, presentations, documentation

#### ARTICLES

- Brunsdon, C., A.S. Fotheringham, and M.E. Charlton. 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis* 28 (4): 281-298.
- Chaix, B., J. Merlo, and P. Chauvin. 2005. Comparison of a spatial approach with a multilevel approach for investigating place effects on health: The example of healthcare utilization in France. *Journal of Epidemiology and Community Health* 59 (6): 517-526.

Fotheringham, A.S. and C. Brunsdon. 1999. Local forms of spatial analysis. *Geographical Analysis* 31 (4): 340-358.

- Páez, A., S. Farber, and D. Wheeler. 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A* 43 (12): 2992 3010.
- Wheeler D and M. Tiefelsdorf. 2005, Multicollinearity and correlation among local regression coefficients in geographically weighted regression *Journal of Geographical Systems* 7 (2): 161 187.

#### BOOKS

- Fotheringham, A.S., C. Brunsdon, and M.E. Charlton. 2000. *Quantitative geography: Perspectives on spatial data analysis*. London: SAGE Publications.
- Fotheringham, Stewart A., Brunsdon, Chris, and Charlton, Martin.. *Geographically Weighted Regression: The analysis of spatially varying relationships.* John Wiley & Sons, 2002.
- Mitchell, Andy. The ESRI Guide to GIS Analysis, Volume 2. Redlands: ESRI Press, 2005.

#### **RECOMMENDED VIDEO**

ESRI Users Conference 2011. Spatial statistics: Best practices.

http://video.esri.com/watch/903/spatial-statistics-best-practices

Charlton, M. 2011. Geographically weighted regression: Modeling spatial heterogeneity. Lecture at Hamilton Institute.

http://www.podcasts.com/hamilton institute seminars hd large/episode

geographically weighted regression modelling spatial heterogeneity

Section 5

# DISCUSSION