Social Determinants of Low Birth Weight: Lessons Learned in Spatial Analysis of Sub-County Health Outcomes

Tabassum Insaf, MBBS, MPH, Phd (Epidemiology)

Bureau Of Occupational And Environmental Epidemiology

New York State Department Of Health

Tabassum.Insaf@health.ny.gov

Analyzing Social Determinants

- EPHT funded project
- NYS goals: Develop robust analysis methods to describe the socio-demographic characteristics of NYS populations that are associated with higher rates of low birth weight at census tract level.
- Develop smoothed fine-resolution maps to identify community-level patterns of low birth weight rates and socio-demographic risk factors

What we will discuss today

- Methodological challenges with sparse data at sub-county level-Tom
- Methodological challenges with regression/modeling of Spatially dependent data
- Compare approaches for analysis of Spatially dependent data
- Case-Study -NYS analysis of Social Determinants of Low Birth Weight-Preliminary findings

Problems with using Linear Regression

- A basic assumption for linear regression is that the observations and regression residuals should be independent of each other.
- Ignoring lack of independence (existence of correlated observations) can result in the estimated regression coefficients being biased, inconsistent, or inefficient
- **Tobler's First Law of Geography**: Things closer in space tend to be more similar than things further apart

Issues with Spatially Specified Data

- Spatial dependence -what happens at one point in space influences what happens elsewhere.
  - Positive Spatial Autocorrelation/Negative Spatial Correlation
- Spatial heterogeneity- variation in relationships over space.
  - regionally-specific circumstances influence structural relationships.

How do I know if I have a problem?

- Apply tests for spatial autocorrelation to residuals of “conventional” regression model.
- This may be done with various software packages SAS, R, ArcGIS, GeoDa
- A significant ($p<0.05$) Moran’s I test suggests that there is spatial autocorrelation in the data.
- When residual spatial autocorrelation is present, several approaches may be taken to adjust for it.

How to fix spatial autocorrelation

Step 1:

Regression assumes *all* relevant variables influencing the dependent variable are included

Missing (omitted) variables may cause spatial autocorrelation

Try to identify omitted variables and include them in a multiple regression.

Step 2:

If additional variables cannot be identified and SA still exists, use a *spatial regression model*

Regression Approaches to Spatially specified data

- Spatial Autocorrelation
  - Fixed effect-location dummy variable
  - Mixed effects model-location random effect
  - Spatial lag model-weighted spatial lag of outcome variable
  - Spatial error model
- Spatial Heterogeneity
  - Geographic weighted regression

Geographically Weighted Regression

- Designed to explore **spatial heterogeneity** in parameters-assumes heterogeneity arises from intrinsic local differences, global statements of spatial behavior are not possible
- estimates an OLS-like regression for each census tract by applying the spatial weights matrix to the standard formulae for regression
- Each regression is based on few observations , the estimates of the regression parameters (b) are unreliable
- Need strong theory to explain why the regression parameters are different at different places

Spatial Error model

$$y_i = \boldsymbol{x}_i\beta + \lambda w_i \xi_{i.} + \varepsilon_i.$$

- decomposes overall error into two components, spatially uncorrelated error term ε_i that satisfied the normal regression assumption, and $\xi_{i.}$ which is a term indicating the spatial component of error term.
- It controls autocorrelation in both the dependent and the independent variables
- Spatial error model suggests that there is some spatially clustered feature that influences the value of y for i and its neighbors but is omitted from the specification.

Case-Study

Example variables

Social Determinants

- Education
- WIC participation
- Race
- Housing status

Behavioral Factors

- Prenatal care
- Smoking
- Drug use
- Maternal age

Low Birth Weight

Steps in Analysis

- Initial Data-cleaning and management
- Obtain counts at census tract level from birth certificate data
- Merge census tract level birth data with demographic data from Census and ACS
- Aggregate data and explore spatial correlation
- Apply Spatial Regression models
- Assess model fit
- Obtain predicted counts/rates and residuals from selected model
- Map predicted rates to identify areas of concern
- Map residuals to identify unique patterns

All data-management and analysis completed in ArcGIS and SAS 9.3

Low Birth Weight in New York State

- The Study population consisted of all births in upstate New York (excluding the five counties in New York City) for the years 2008-2012
- The study sample consisted of 562,586 births for the study period.
- Birth outcomes were reported in 2713 tracts.
- Tract level birth data was merged with socio-demographic data from Census and ACS.
- Geographic Aggregation tool was used to aggregate census tract such that each unit had atleast 250 births, aggregated areas were restricted to not cross county boundaries.
- This resulted in 1268 unique tract-areas in upstate NewYork

Creating count data

```
proc sql;
create table lbwrates as  select mi_tract10, count(*)"births",count (case
when lbw=1 then "count me"
end)"lbwc",
count (case when shortrace=1 then "count me" end)"Hispanics",
count (case when shortrace=2 then "count me" end)"Whites",
count (case when shortrace=3 then "count me" end)"Blacks", ……………………………
from tractdata
group by mi_tract10 */census tract variable*/
order by mi_tract10;
quit;
```

Patient ID	Lbw	NHW	WIC	Census tract
123	0	0	0	10001
124	1	0	0	10001
125	0	1	1	10002

Census tract	LBW	NHW	WIC	Births
10001	5	3	3	10
10002	12	3	8	30

Merging tract level data sets

```
proc sql;
create table gat_out as select *
        from lbwrates as A
                left join acs1 as b on a.mi_tract10=b.mi_tract10
                left join sf1 as c on a.mi_tract10=c.geoid11
                quit;
```

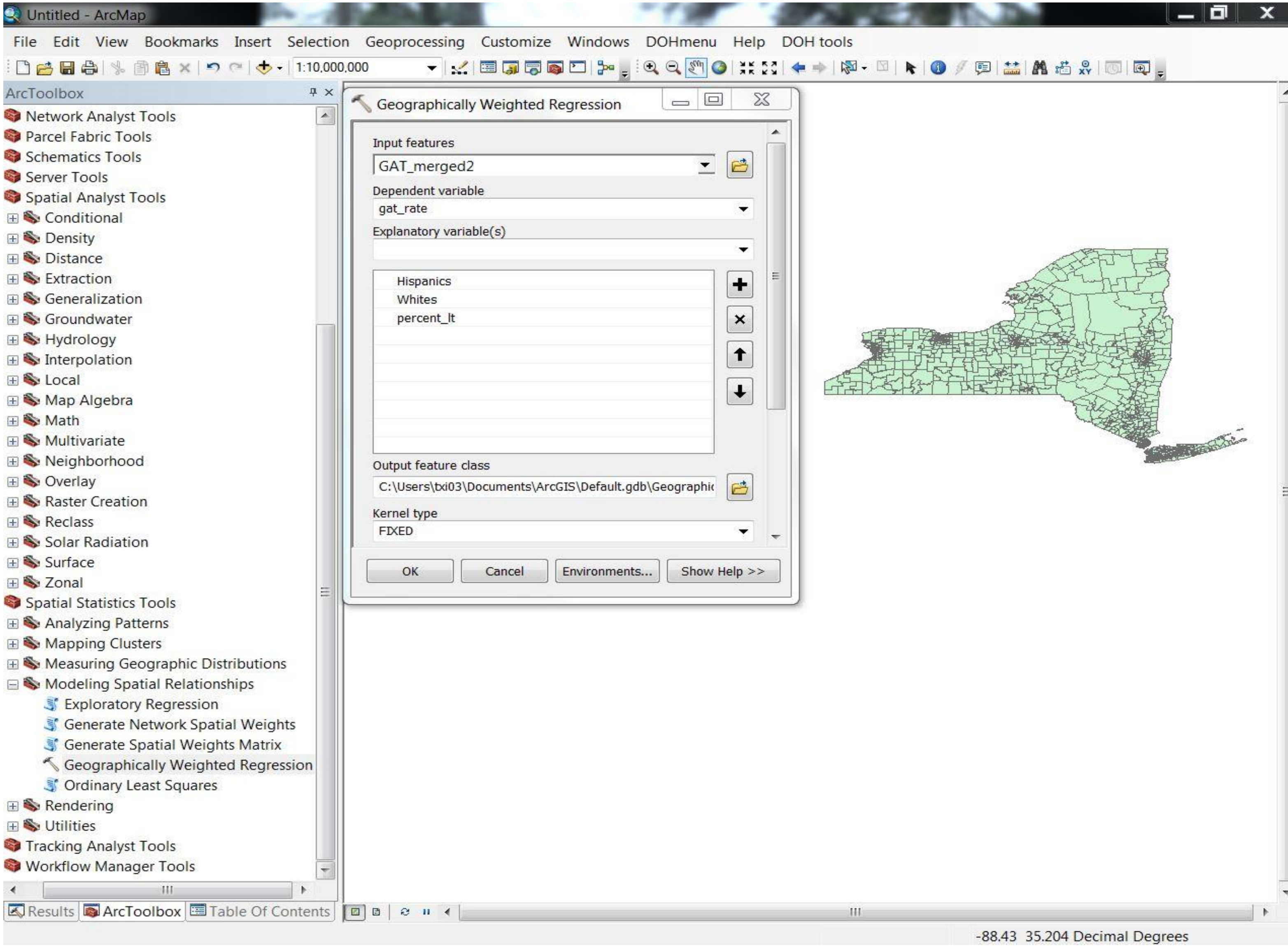

Rochester Area
Albany-Capital District Area
Buffalo Area
Long Island/Westchester area
Predicted LBW Rate
1.76 - 2.50
2.51 - 4.50
4.51 - 6.50
6.51 - 8.50
8.51 - 15.44
N
W
E
S
0
85
170
340 Miles

Approach 1-GWR

Geographic Weighted Regression using Arc GIS

- Geographic weighted OLS model with a fixed distance band and row standardization.
- Problems:
  - assumes that LBW is a continuous variable with normal distribution- GWR is only available as a linear regression (OLS) model in ArcGIS
  - Provides different estimates for each census tract (may not be useful for Public Health Action)
  - Poor fit- Spatial autocorrelation of residuals persists

ArcGIS –Geographic Weighted Regression

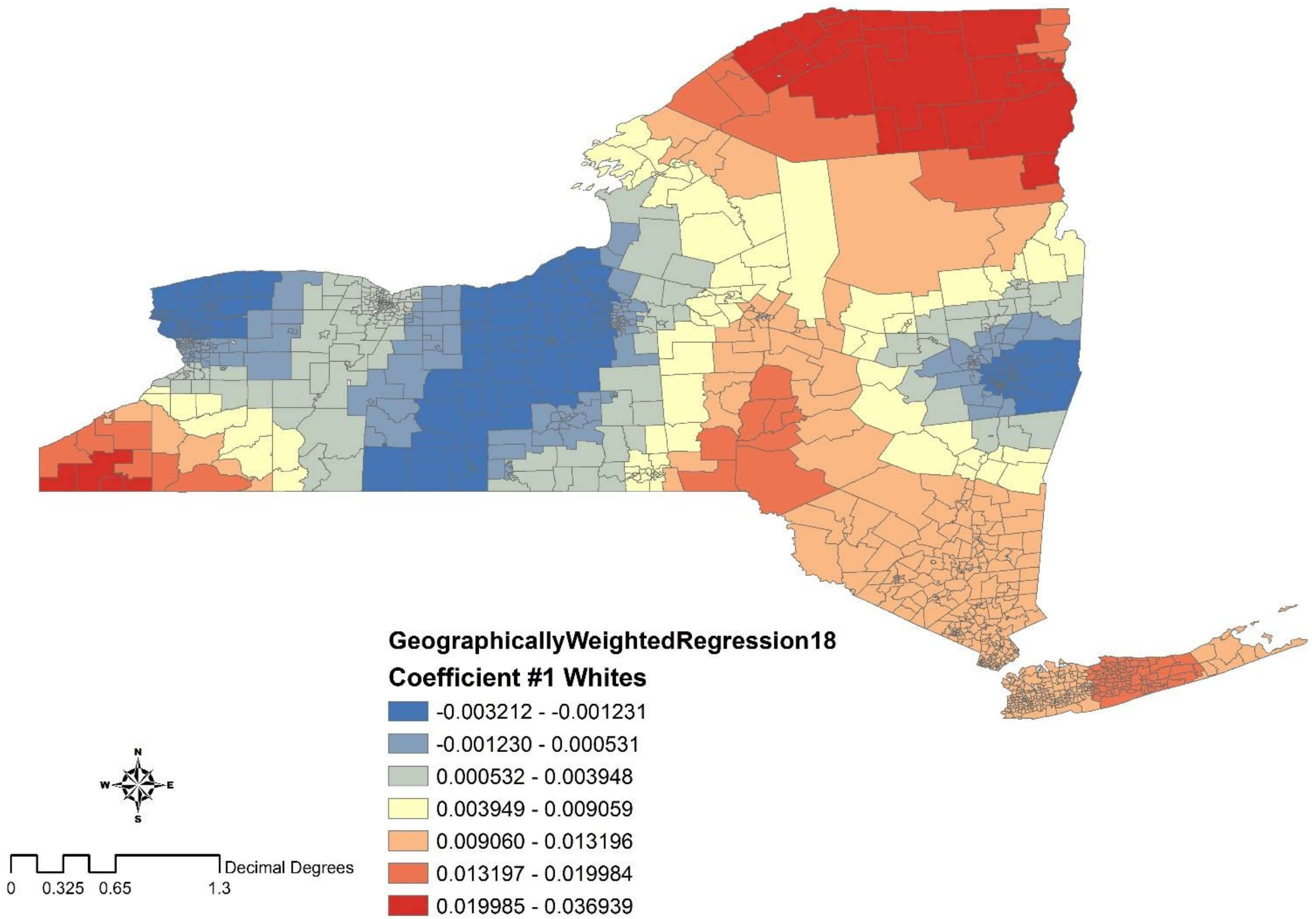


Table

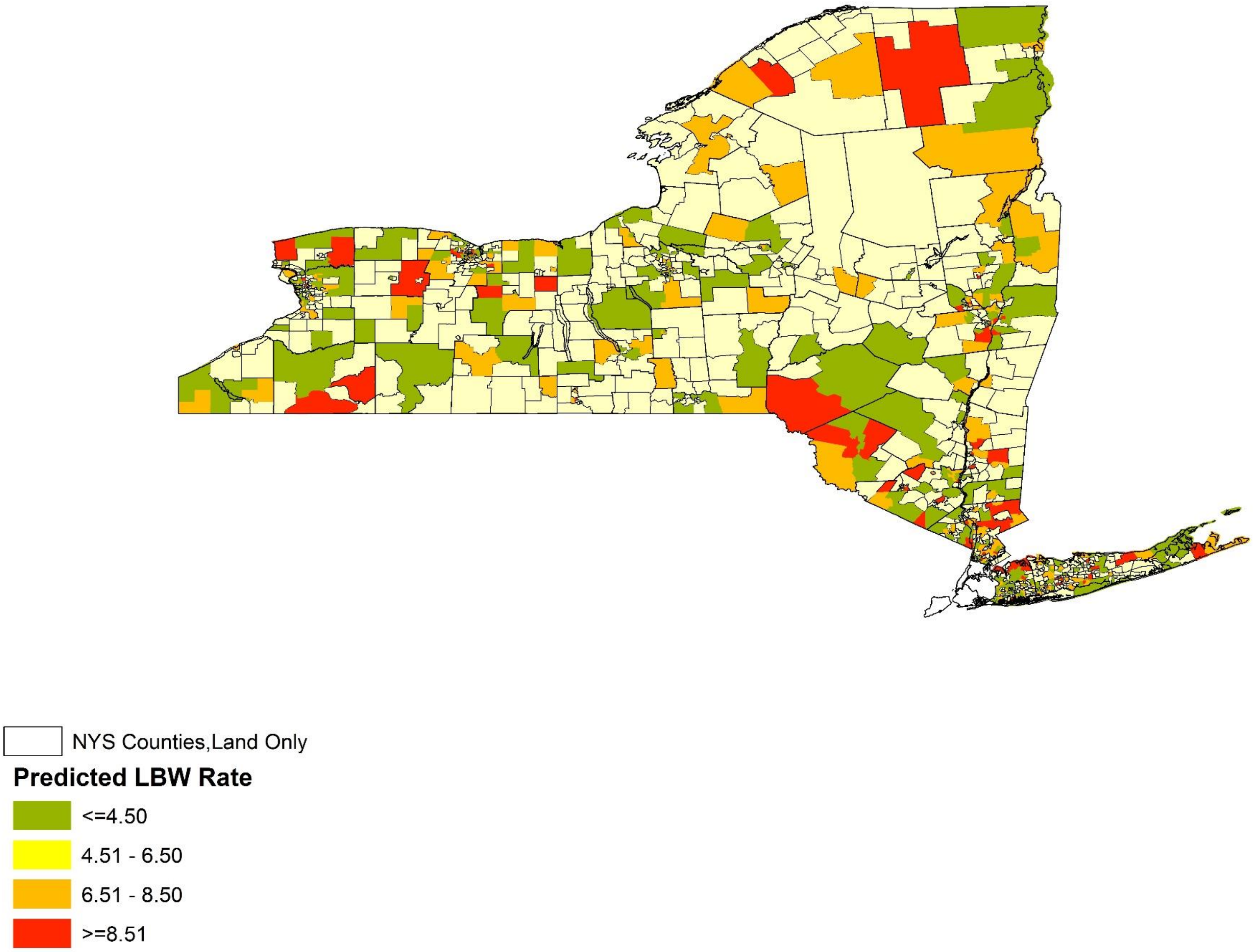
GeographicallyWeightedRegression18

OBJECTID *	Observed	Local R2	Predicted	Coefficient Intercept	Coefficient #1 Whites	Residual	Standard Error	SE_Intercept	SE_Whites
1	73	0.854564	78.426595	0.955952	-0.001823	-5.426595	3.477087	1.290254	0.004687
2	27	0.853754	32.356764	0.924926	-0.00175	-5.356764	4.115015	1.292666	0.0047
3	21	0.854284	24.073872	0.937548	-0.001735	-3.073872	4.216836	1.284711	0.004667
4	42	0.85425	52.858463	0.942288	-0.001788	-10.858463	3.991237	1.290146	0.004689
5	40	0.855413	44.777066	0.94443	-0.00162	-4.777066	4.04029	1.264581	0.004585
6	13	0.850516	13.929074	0.7857	-0.001304	-0.929074	4.274682	1.307406	0.004765
7	19	0.876889	15.167153	0.850058	0.002129	3.832847	4.23008	0.972602	0.004025
8	17	0.849999	12.226278	0.200021	0.007497	4.773722	4.058985	1.359727	0.005129
9	6	0.873104	5.486609	-0.848083	0.004472	0.513391	4.154725	1.543659	0.006975
10	14	0.887445	19.691117	-3.15231	0.018349	-5.691117	4.123058	1.298527	0.005342
11	17	0.848251	16.637805	2.983488	-0.002727	0.362195	4.307156	1.057351	0.004745
12	9	0.845706	13.640232	3.004572	-0.002705	-4.640232	4.27472	1.068771	0.004789
13	24	0.848386	19.122466	2.954769	-0.002708	4.877534	4.251469	1.056456	0.004742
14	13	0.916426	12.565736	-2.871529	0.022945	0.434264	3.681237	1.950243	0.007436
15	14	0.884273	19.490068	-0.417427	0.000907	-5.490068	4.149919	1.570435	0.007371
16	18	0.880074	12.160311	0.348448	-0.000542	5.839689	4.241748	1.335526	0.006283
17	14	0.886871	17.205076	0.065465	-0.00046	-3.205076	4.107707	1.481906	0.007053

Geographically Weighted Regression- Example Results



Geographic Weighted Regression- Predicted Rates of Low Birth Weight



Approach 2- Spatial Error Model

Spatial error model

- Poisson mixed effects model using a spherical semi-variogram model to model spatial auto-correlation at the tract level.
- Advantages
  - Better model fit
  - Spatial autocorrelation accounted for
  - One set of estimates for NYS

Spatial Error Model

Spatial-error model- solved using SAS Proc Glimmix

```
proc glimmix data=&data ;
class mi_tract10  ;
model lbwc=&varlist /solution dist=poisson link=log offset=lnbirth ddfm=residual  chisq ;
random mi_tract10/ type=sp(exp)(lng lat) ;
output out=scored_glim_cen_rnd_all pred(blup ilink)=predicted resid=resid;
ods output ParameterEstimates=estimates3;
run;
```

Other alternatives:

STATA-spatreg
R- spgwr, spdep

Poisson-Spatial Error Models Draft Results

Variables	Estimate	Standard Error	Pr > |t|	Relative Risk	95% Confidence Interval	
Intercept	-3.0897	0.07069	<.0001	0.04552	0.03963	0.05228
pncrate	-0.3095	0.09053	0.0006	0.73381	0.6145	0.87629
agelt18	0.3029	0.06204	<.0001	1.35379	1.19878	1.52885
wicrate	0.5892	0.05723	<.0001	1.8026	1.61133	2.01658
drugrate	1.2839	0.324	<.0001	3.61079	1.91338	6.81402

Spatial Autocorrelation Analysis

Autocorrelation Statistics for LBW (unadjusted)

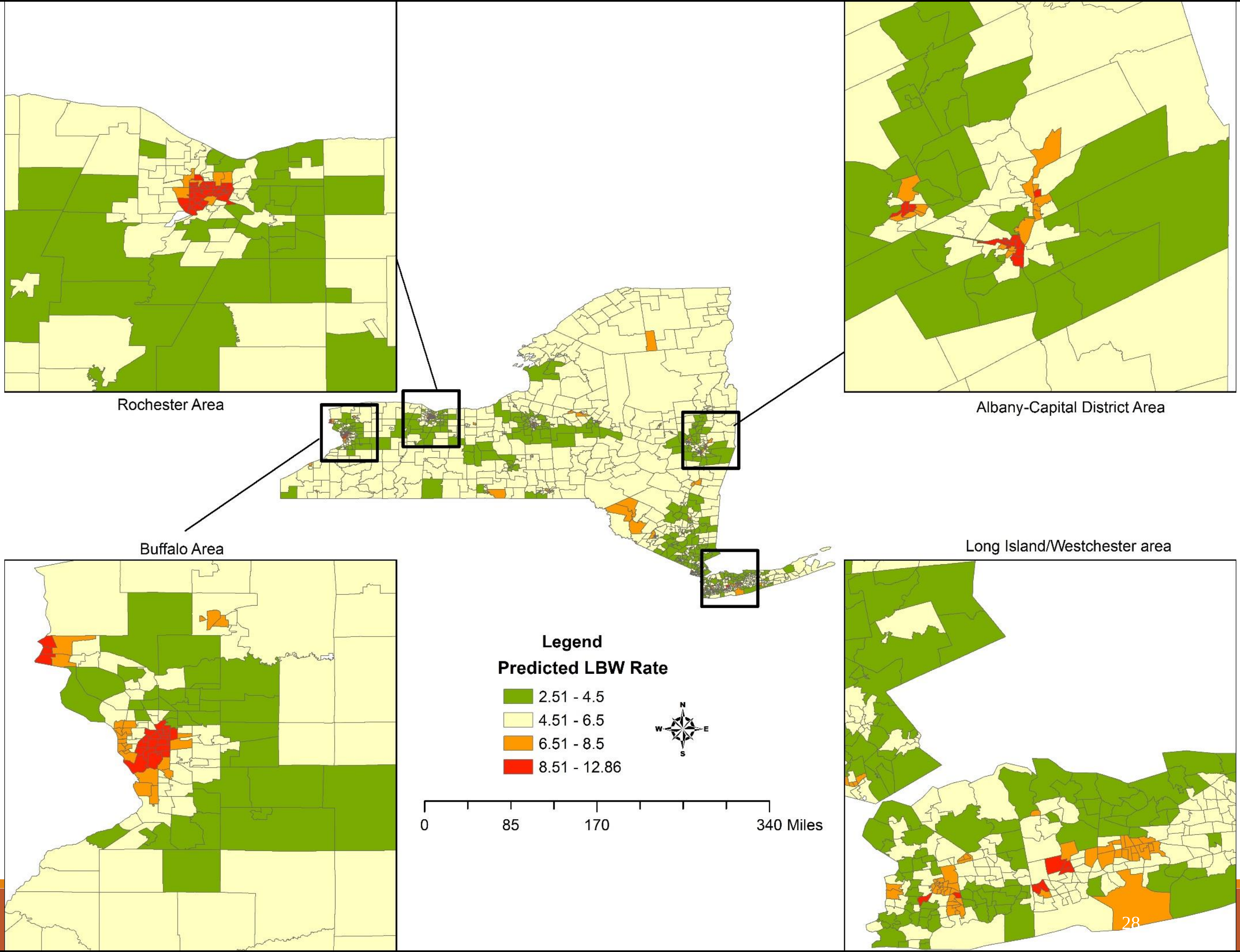
Assumption	Coefficient	Observed	Expected	Std Dev	Z	Pr > |Z|
Normality	Moran's I	0.0199	-0.000789	0.00559	3.69	0.0002

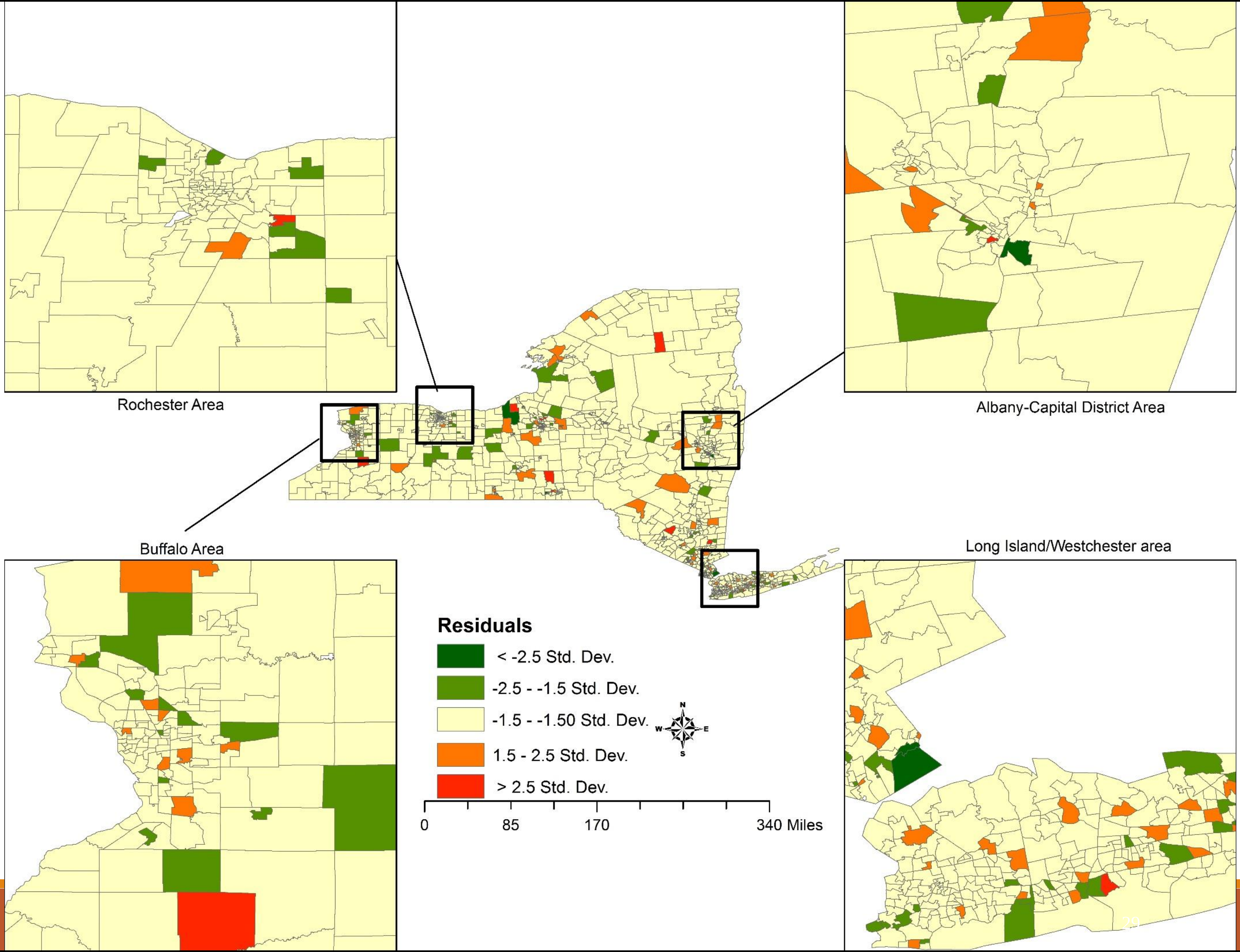
Autocorrelation Statistics for Student's Residuals for GWR

Assumption	Coefficient	Observed	Expected	Std Dev	Z	Pr > |Z|
Normality	Moran's I	0.0146	-0.000789	0.00559	3.07	0.0021

Autocorrelation Statistics-Residuals for Spatial Regression

Assumption	Coefficient	Observed	Expected	Std Dev	Z	Pr > |Z|
Normality	Moran's I	-0.00969	-0.000789	0.00559	-1.59	0.1112


Rochester Area
Albany-Capital District Area
Buffalo Area
Long Island/Westchester area
Legend
Predicted LBW Rate
2.51 - 4.5
4.51 - 6.5
6.51 - 8.5
8.51 - 12.86
N
S
W
E
0
85
170
340 Miles



Rochester Area
Albany-Capital District Area
Buffalo Area
Long Island/Westchester area
Residuals
< -2.5 Std. Dev.
-2.5 - -1.5 Std. Dev.
-1.5 - -1.50 Std. Dev.
1.5 - 2.5 Std. Dev.
> 2.5 Std. Dev.
N
W
E
S
0
85
170
340 Miles

Conclusion

- Spatial data can cause problems with standard correlation and regression
- The problems are caused by spatial autocorrelation/spatial heterogeneity
- Using appropriate spatial regression techniques will account for spatial autocorrelation and provide robust results
- Spatial Regression results can be used to develop Smoothed(Predicted) maps and Residual maps which can be useful for Program Planning.