Extending the CAR Model to Account for General Temporal Neighbourhood Structures

Eric Kalendra

NCSU

2009-12-08

SAMSI: Spatial Epidemiology

Introduction

- Preterm birth refers to the birth of a baby of less than 37 weeks gestational age.
- Because it is by far the most common cause of prematurity, preterm birth is the major cause of neonatal mortality in developed countries.
- Premature infants are at greater risk for short and long term complications, including disabilities and impediments in growth and mental development.

- Understanding what places certain mothers at higher risk for having a poor pregnancy outcome is important in trying to prevent preterm and low weight births from occurring.

The Question of Interest

- From previous work on birth outcome data from New York City, we know there is spatial variation even after accounting for race.
- With multiple years of data it is reasonable that the demographics may change overtime.
- We would like a method that allows smoothing over space and time after accounting for other factors.

- The temporal structure needs to be flexible.

- Is there any evidence that the spatial pattern of preterm birth changes over time?

Description of the Data

We have data for

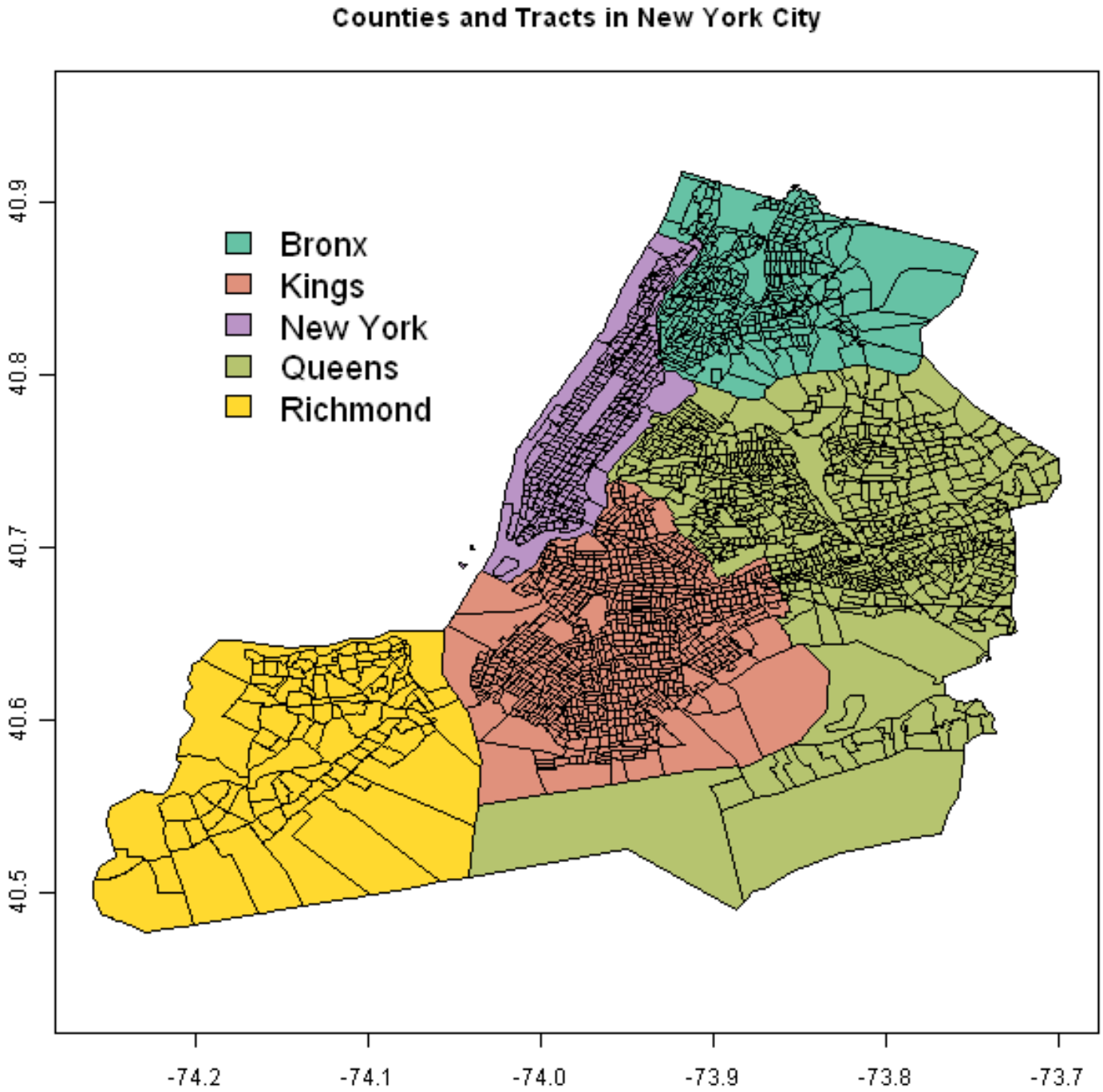
- White mothers
- who live in Bronx or New York counties in New York
- for pregnancies which resulted in a single birth
- from 1995 to 2003

Data for each mother:

- Location geocoded to census tract level,
- Age,
- Education, and
- Country of birth.

Data for each baby:

- Gestational age in weeks (preterm is less than 37 weeks),
- Sex of the baby, and
- Month of delivery.


Counties and Tracts in New York City
Bronx
Kings
New York
Queens
Richmond
40.9
40.8
40.7
40.6
40.5
-74.2
-74.1
-74.0
-73.9
-73.8
-73.7

The Model

To study the questions of interest, we fit a probit regression model expanded to include a spatial-temporal intercept:

$$\Phi^{-1}(E(y_i)) = Z_i\gamma + \beta(s_i, t_i), \tag{1}$$

- y_i is indicator of whether mother i gives birth prematurely,
- Z_i is the vector of confounders (individual level characteristics),
- γ are the corresponding fixed effects,
- β is the spatial-temporal intercept having a STCAR prior,
- s_i is the census tract where mother i lives, and
- t_i is the time period when mother i gives birth (month, year, ...).

Spatial-temporal Component: β has a STCAR Prior

β follows a spatial-temporal Conditional Autoregressive (STCAR) model. This allows the model to:

- Vary the effect of each racial group smoothly across space.
- Vary the effect of each racial group smoothly across time.
- Allow the model to have a flexible temporal neighbourhood structure
- Impose the condition that neighbouring census tracts will have similar risks.
- Allows variation in risk which is not explained by covariates to be explained in the spatial component.

This is different from the typical *spatiotemporal autoregressive* (STAR) models.

Defining the STCAR

The STCAR model is defied as

$$\beta \sim MVN\left(0, \left[\tau\left(\alpha(M_s - \rho_s D_s) + (1-\alpha)(M_t - \rho_t D_t)\right)\right]^{-1}\right)$$

- τ is a precision component that controls the variance
- M_s and M_t are diagonal matrices listing the number of neighbours for space and time respectively
- ρ_s and ρ_t control the smoothing
- D_s and D_t are the neighbourhood matrices for space and time respectively
- α controls the amount of smoothing coming from spaces versus time
  - $\alpha -> 0$ implies smoothing across time only
  - $\alpha -> 1$ implies smoothing across space only
- This is not the dynamic linear model
- It is not clear if there is a corresponding SAR equivalent

STCAR Full conditionals

The STCAR model is defined through the conditional

$$\beta(s,t)|\beta(-(s,t)) \sim N\left(\mu(s,t), per = \tau\left(\alpha M_{s=s} + (1-\alpha)M_{t=t}\right)\right) \quad (2)$$

$$\mu(s,t) = \frac{\alpha\rho_s M_{s=s}\bar{\beta}_s + (1-\alpha)\rho_t M_{t=t}\bar{\beta}_t}{\alpha M_{s=s} + (1-\alpha)M_{t=t}} \quad (3)$$

- $\mu(s,t)$ is between the average of the spatial neighbours and temporal neighbours
- The precision is between τ times the number of spatial neighbours and τ times the number of temporal neighbours
- We assume the weights to be equal to 1 if two tracts are neighbours and 0 if the two tracts are not neighbours.

Example analysis

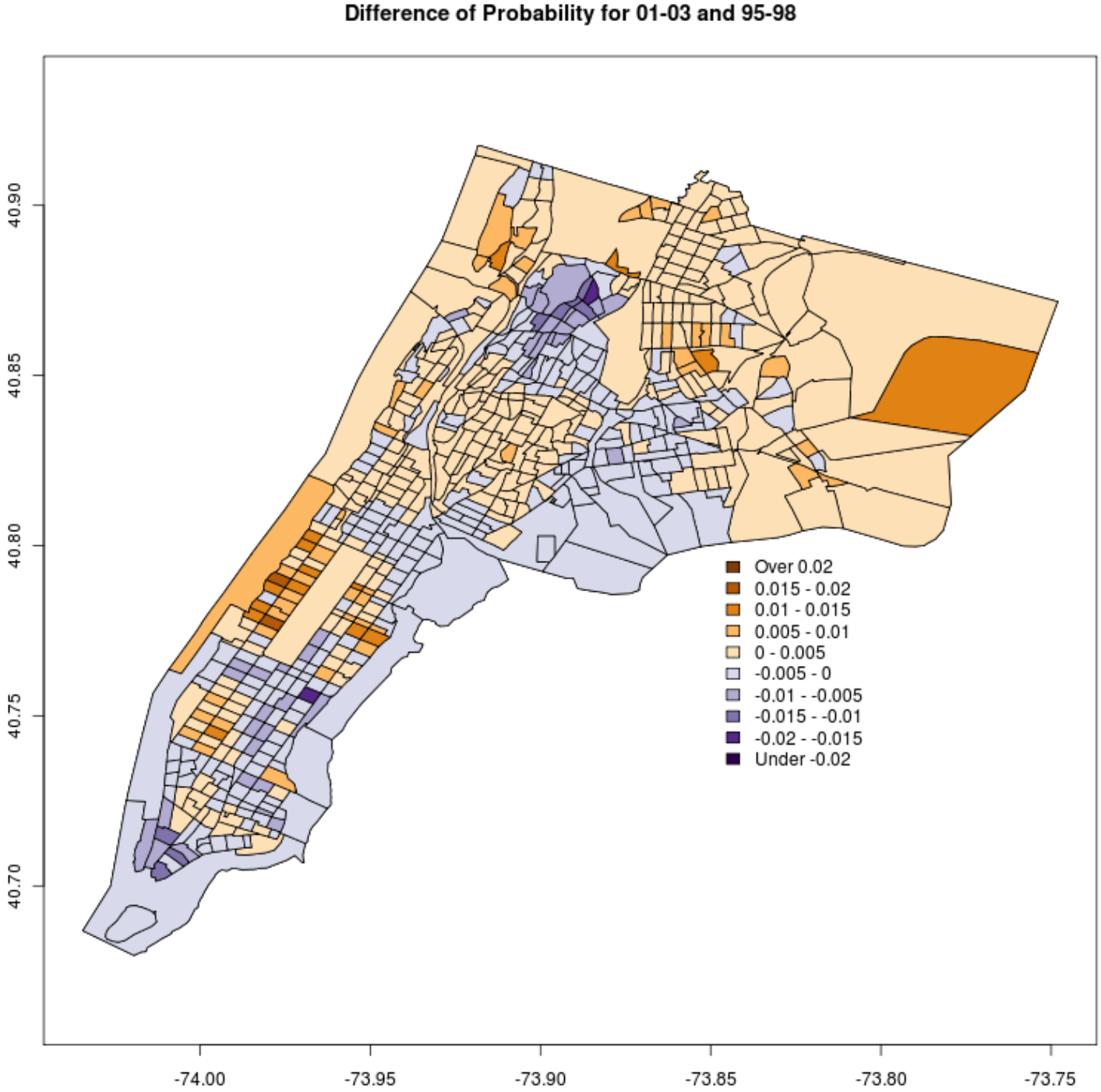
A subset of the birth outcome data is used to demonstrate the method.

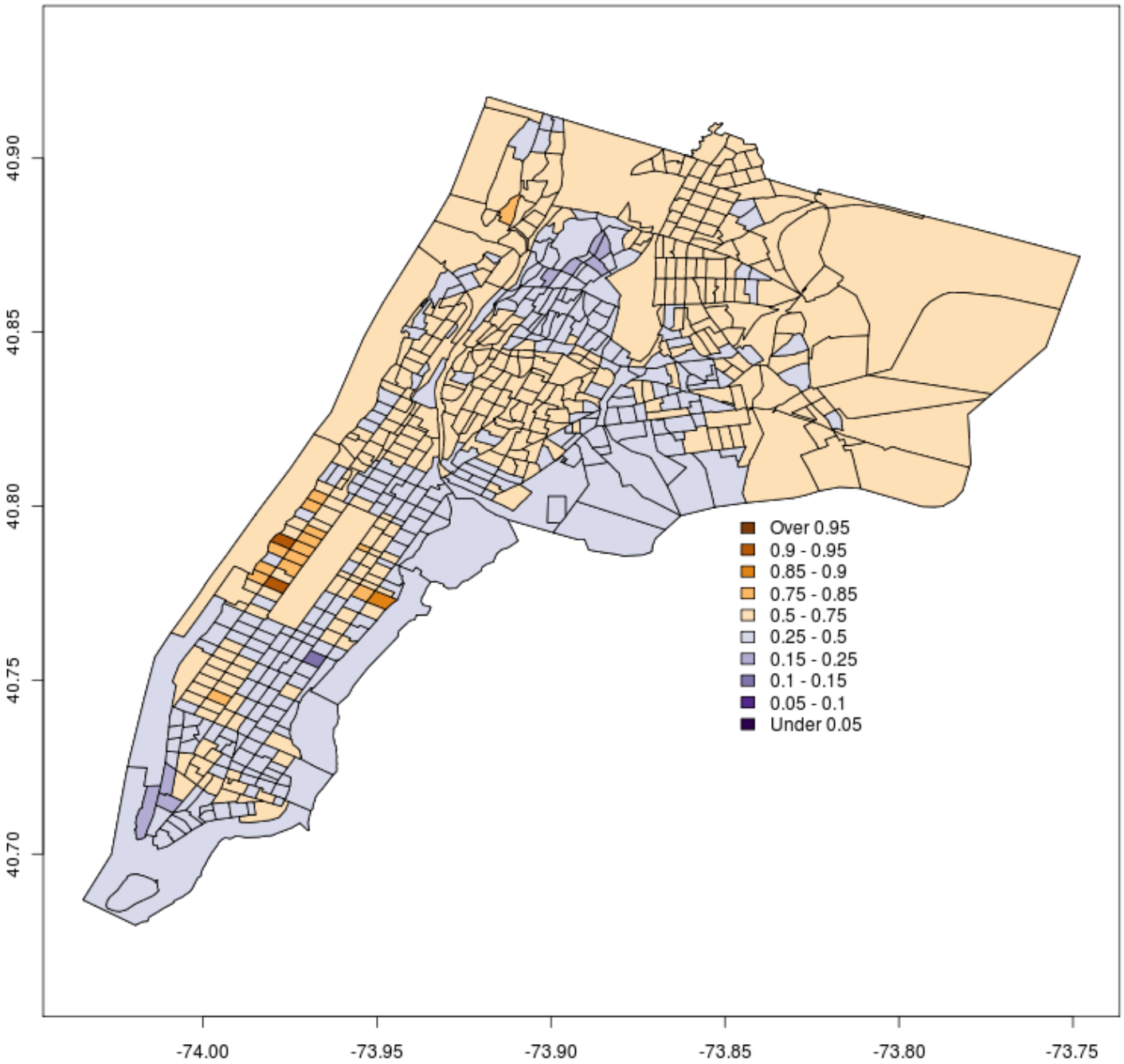
- Spatial structure is first order neighbours
- Temporal structure is previous and next year (if available)
  - The time domain is 1995 to 2003
  - Time is broken into 3 categories
    1995 - 1997, 1998 - 2000, and 2001 - 2003
- The spatial domain is New York and Bronx counties
  - In total, 650 census tracts
- 65,628 mothers used in the analysis
  - $\approx$ 30 mothers per census tract per time period
- A total of 3,442 preterm births

Estimates of the Fixed Effects

	$\hat{\gamma}$	SE	P>0
Intercept	6.32	0.60	1.00
$<$ 18 years old	0.31	1.41	0.55
18 - 25	-0.05	0.40	0.44
26 - 30			
31 - 35	0.28	0.30	0.84
36 - 40	0.85	0.33	1.00
$>$ 40	2.53	0.53	1.00
Elementary	-0.59	0.75	0.21
High School			
College	-1.81	0.33	0.00
Grad School	-1.89	0.35	0.00
Smoking	6.98	1.13	1.00
Male Child	0.91	0.23	1.00
Mforeign	-1.00	0.23	0.00
WIC	0.80	0.45	0.96

These results are similar to previously obtained results for the fixed effects


Difference of Probability for 01-03 and 95-98
40.90
40.85
40.80
40.75
40.70
-74.00
-73.95
-73.90
-73.85
-73.80
-73.75
Over 0.02
0.015 - 0.02
0.01 - 0.015
0.005 - 0.01
0 - 0.005
-0.005 - 0
-0.01 - -0.005
-0.015 - -0.01
-0.02 - -0.015
Under -0.02


Signifiance of Difference of Probability for 01-03 and 95-98
40.90
40.85
40.80
40.75
40.70
-74.00
-73.95
-73.90
-73.85
-73.80
-73.75
Over 0.95
0.9 - 0.95
0.85 - 0.9
0.75 - 0.85
0.5 - 0.75
0.25 - 0.5
0.15 - 0.25
0.1 - 0.15
0.05 - 0.1
Under 0.05

Posterior of α

α controls the dependence between spatial smoothing and temporal smoothing.

The Following table characterizes the posterior of α:

	α
Min.	:0.30
1st Qu.	:0.46
Median	:0.61
Mean	:0.61
3rd Qu.	:0.75
Max.	:0.99

The model is favouring more spatial smoothing over temporal smoothing.

Conclusions

- There does appear to be a limited amount of spatial variation in the temporal pattern
- There are not any regions which show up to be statistically significant in a change in spatial structure over time
  - Although, the data that are used is very limited
  - Using a larger spatial region may provide more diverse results
- Even with a very small number of preterm births (outcomes) the model does a good job smoothing across space
- Extending the analysis to cover all of New York city for a more complete analysis would be the next step in evaluating the model

Further Work

We plan to do the following things:

- compare the STCAR model to the more typically used STAR models (is there a SAR equivalent to the STCAR model?),
- compare the STCAR model to the dynamic linear model,
- incorporate the use of a model comparison method such as DIC to evaluate the the fit of the model instead of looking at statistical significance of the change in individual census tracts over time,
- allow for non-stationary models for α, as there is no reason to believe the smoothness across time and space is constant, and
- compare the performance of other covariance structures (directionality, increased temporal neighbourhood, more temporal categories, ...).