

COVID-19, Climate and Socioeconomic Status in the United States: A Bayesian Analysis of County-Level Case Data

Jordan Bramble, Charlie Frazier, Liam Hanson, and Aaron Marshall



NSA and NSF for support through:
NSA:
H98230-20-1-0259
and
H98230-21-1-0014
NSF: DMS#2051010

1 Bayesian SIR Model

We model COVID-19 spread in county j during week t via an SIR model:



$\beta_j(t)$ is the county-dependent, time-varying transmission rate and γ_j is the county-dependent removal rate. Transmission rates are assumed to evolve temporally according to a stationary log AR(1) process:

$$\begin{cases} b_j(t) = \log(\beta_j(t)) \\ b_j(1) \sim \text{Normal}(0, \sigma_j^2) \\ b_j(t) \sim \text{Normal}(\rho_j b_j(t-1), (1-\rho_j^2) \sigma_j^2), t > 1 \end{cases}$$

The process is governed by variance σ_j^2 and autocorrelation coefficient ρ_j .

Let $C_{j,true}(t) = |S_j(t-1) - S_j(t)|$ and $C_{j,obs}(t)$ be the true and observed numbers of new cases in county j during week t . We assume the following data model:

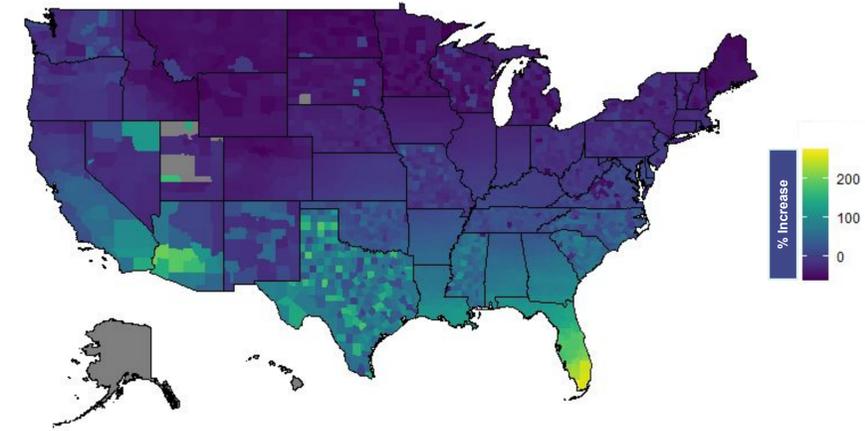
$$C_{j,true}(t) | C_{j,obs}(t) \sim \text{Poisson}(C_{j,true}(t - \ell)), \ell = \text{Reporting Lag}$$

Finally, we assume gamma priors for $I_j(0)$, and $R_j(0)$; beta priors for γ_j and ρ_j ; and a half-normal prior for σ_j .

Abstract

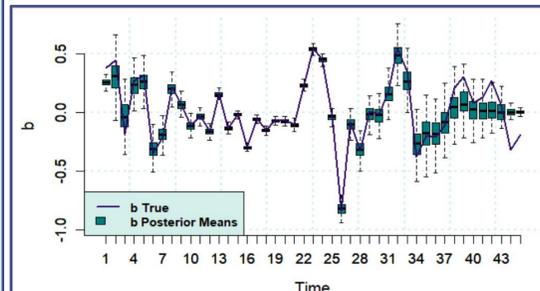
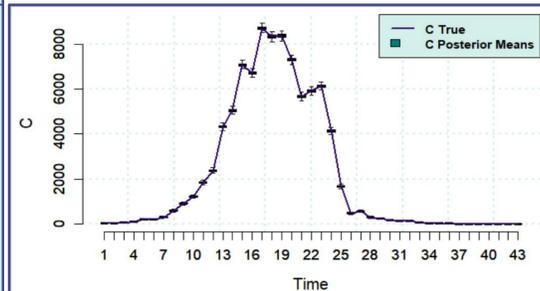
SARS-CoV-2 has infected over 80 million people in the United States. Its spread is heterogeneous, and differences may be partially explained by temporal and spatial variations in climate and sociological variables. Literature suggests that temperature and humidity may impact COVID-19 spread, but their influence remains unclear. Similar is true of population density, poverty, race, ethnicity, and age. Meanwhile, literature suggests that air pollution may increase spread. We utilize SIR models with county-dependent parameters to simulate SARS-CoV-2 spread across the U.S. Transmission rates are assumed to evolve temporally according to a stationary AR(1) process. These models are fit to county-level COVID-19 case data, and parameter estimates for all U.S. counties are obtained via Markov chain Monte Carlo sampling. We analyze the above variables' relative influence on SARS-CoV-2 transmission using a Bayesian linear regression model with the estimated means of county-level transmission rates as the response and climate and sociological variables as covariates. Results indicate that climate and sociological variables significantly impact SARS-CoV-2 transmission.

U.S. Map: % Increase in Transmission Rate from Temperature Increase



2 Simulation Study

- 6 parameter set types, 5 randomly generated versions each
 - Parameter variability increased going from parameter set i to parameter set $i+1$
 - $b(t)$ generated via draws from stationary log AR(1) process
- 30 parameter sets total
 - Simulated 250 observed case sets per parameter set
 - Performed 3 fittings per observed case set
- Obtained posterior means and standard deviations for $C_{true}(t)$ and $b(t)$; calculated coverage, bias, and MSE
 - Results from fittings pooled to obtain means and S.D.s
 - Coverage generally high, bias and MSE generally low
- True $C_{true}(t)$ generally recovered across-the-board; true $b(t)$ generally recovered when:
 - Case curves could not be produced by simpler model
 - Case counts sufficiently high

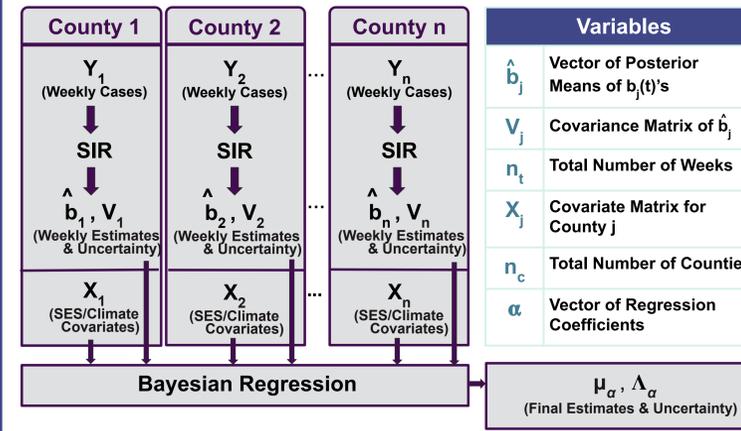


4 Bayesian Regression Model

The posterior means $\hat{b}_j = \langle b_j(1), \dots, b_j(n_j) \rangle$ and covariance matrices V_j are used in a Bayesian Linear Regression Model:

$$\hat{b}_j \sim \text{Normal}(X_j \alpha, V_j), \alpha \sim \text{Normal}(0, 100 \cdot I_p)$$

Candidate covariate matrices X_j included climate and sociological variables; their interactions; and state and month indicator variables. The latter account for large-scale spatiotemporal trends. Included covariates were chosen via forward selection, and models were evaluated via BICs. The posterior means, standard deviations, and 95% credible intervals for α were obtained for the best model.

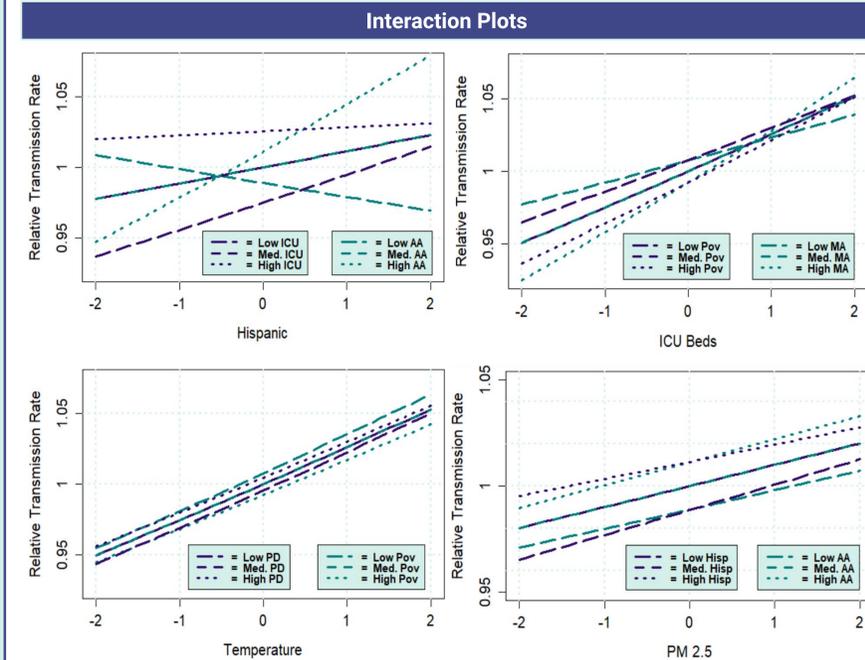


5 Bayesian Regression Results

Scaled coefficient estimates for all first-order terms (excluding indicators) and the 9 most significant interactions are given below. Covariates were scaled to allow for comparison.

Coeff. Estimates: First-Order Terms			Top 9 Second-Order Terms		
Covariate	Mean	95% Interval	Interaction	Mean	95% Interval
Temperature (Temp)	258.1	(254,262.2)	AA*Hisp	213.5	(199.5,227.5)
ICU Beds (ICU)	252.9	(234.0,271.8)	MA*ICU	99.3	(90.7,107.9)
Humidity (Humid)	135.0	(131.6,138.5)	Hispanic*ICU	-86.1	(-90.8,-81.4)
Hispanic (Hisp)	113.7	(103.3,124.1)	MA*PD	-60.6	(-67.8,-53.5)
African American (AA)	111.3	(99.2,123.3)	Humid*PM	-49.4	(-51.9,-46.9)
PM 2.5 (PM)	99.7	(96.3,103.1)	AA*ICU	-45.8	(-50.9,-40.7)
Poverty (Pov)	-77.2	(-86.3,-68.1)	Pov*PM	-41.1	(-43.4,-38.8)
Median Age (MA)	-76.0	(-84.7,-67.3)	Pov*ICU	36.0	(27.8,44.2)
Population Density (PD)	44.9	(34.0,55.7)	Hispanic*Humid	-34.5	(-36.7,-32.2)

Interaction plots for Hisp*ICU & Hisp*AA, ICU*Pov & ICU*MA, Temp*PD & Temp*Pov, and PM*Hisp & PM*AA are given below.



3 Case Data & Model Fitting

Case Data Overview

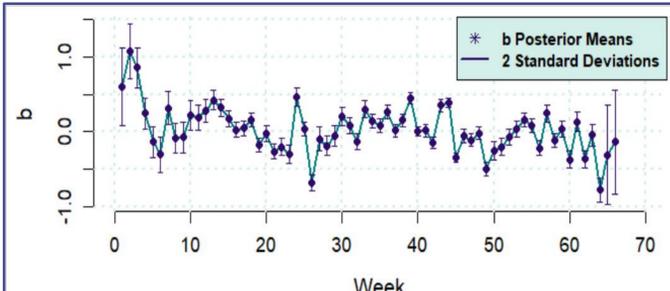
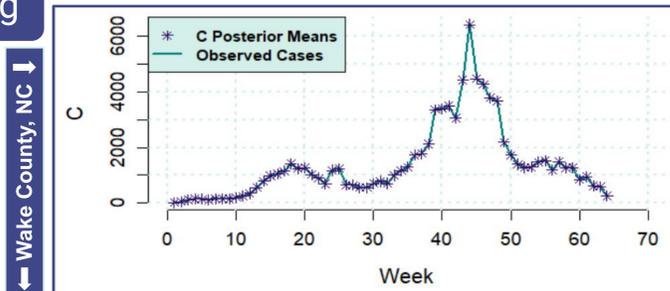
- Aggregated by week to address case-count anomalies
- Trimmed to exclude leading weeks with 0 cases
- Spans ~30-70 weeks

Model Fitting Overview

- Model was fit to each county's case data 3 times to obtain county-level parameter estimates
 - Multiple fittings addressed potential convergence issues
 - Pooled results from fittings to obtain summary statistics
- Estimates for $C_{j,true}(t)$ and $b_j(t)$ were obtained via JAGS in R on NCSU's Henry2 cluster

Results

The model's predicted case counts are generally close to the observed case counts. Slight deviations are observed when case counts are low (< approx. 100). Low case counts are observed for many small counties and at the beginnings and ends of many counties' outbreaks.



Conclusions & Looking Forward

- All climate and sociological variables significantly impacted SARS-CoV-2 spread. In isolation:
 - Temperature and ICU beds most significantly impacted spread
 - Increases in temperature, humidity, PM 2.5, ICU beds, population density, African American proportion, and Hispanic proportion were associated with increases in transmission rates
 - Increases in poverty and median age were associated with decreases in transmission rates
- All climate and sociological variables (save for poverty & proportion African American and temperature & Hispanic) significantly modulated each others' impact on spread
 - African American proportion & Hispanic proportion and ICU beds & median age most significantly modulated each others' impact
 - Most variables interacted negatively (21 negative interactions)
 - Some variables interacted positively (13 positive interactions)
- Results need to be understood and explained in terms of previously established knowledge regarding climate and sociological variables