Predicting Elections and Examining Gerrymandering in North Carolina Using a Spatiotemporal CAR Model

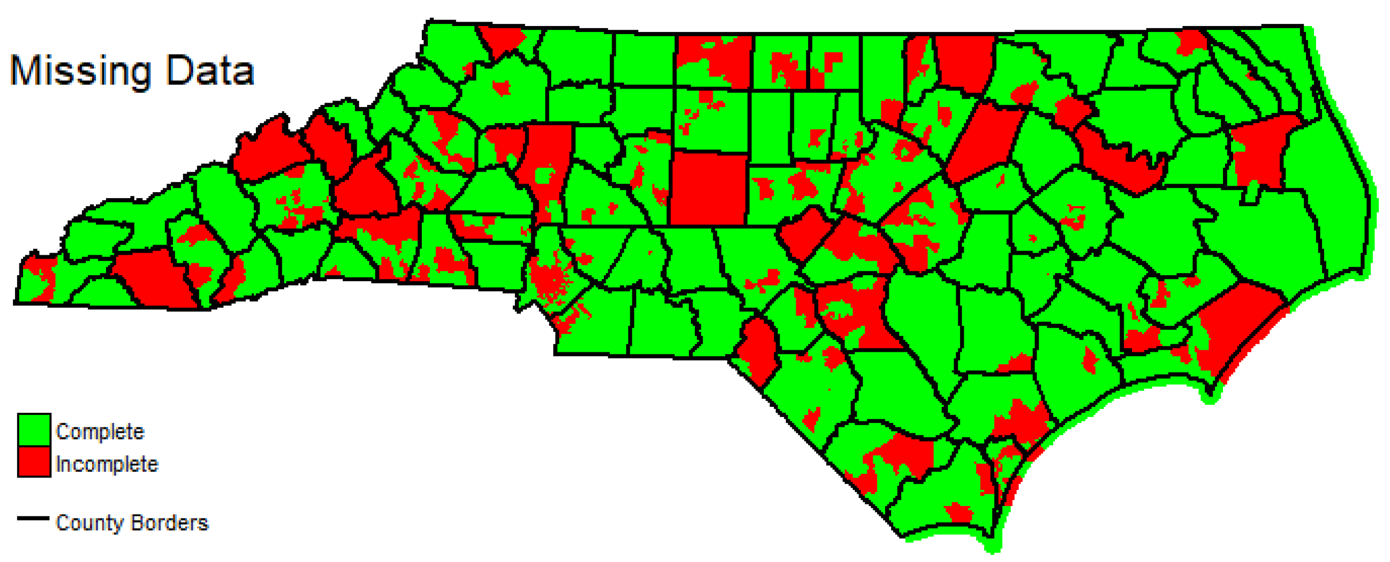
Marschall Furman, Andrew Giffin, Matt Miller

North Carolina State University

Overview

- The 13 US House districts in NC are unconstitutionally racially gerrymandered according to courts
- Motivation from work done by Mattingly's team at Duke [1]
- Electorate about half Democrats and half Republicans
- Republicans have 10-3 advantage in House seats
- House Rules Chairman said "it is only 10-3 because they could not make it 11-2" [2]
- Fit a spatiotemporal CAR model to predict House election results
- Aggregate precinct-level predictions to get district-level outcomes
- Substantial missing data and no polling information in model
- Closely match true outcomes for 2018 House elections

Data



- NC voting data for 9 years of US House races from 2002-2018 [3]
- Currently 2,704 precincts. Only model 2,045 (76%) because of name changes and missing data
- Use precinct-level vote counts and registration data from public record [2]
- Cleaning the data took a large amount of time and effort (19 public data sets)
- Caveats:
  - Only complete precincts included
  - Absentee and early voting not reported at precinct-level, so these were excluded ($\approx 3\%$ of votes)
  - covariates formed using relative proportions of self-reported age and gender (ignored unreported)
  - 4 out of the 117 races in the 9 years had a candidate running without major party opposition (e.g. Libertarian candidate or unopposed), creating outliers

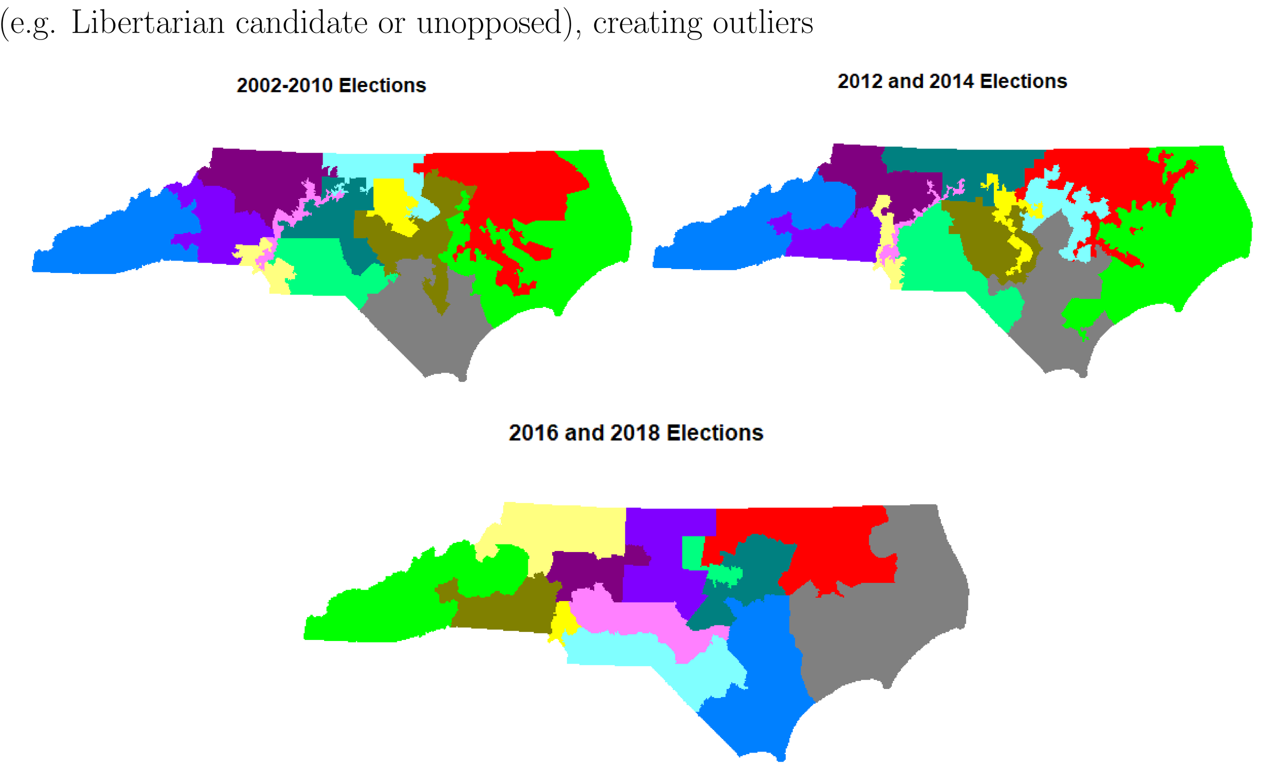


Figure: District Boundaries for each election of interest. The lines are redrawn after each census and, for NC, when they are ruled an unconstitutional gerrymander.

Model Statement

Variable Definitions

- Y_{kt}: Number of Democratic votes in precinct k for year t
- n_{kt}: Sum of Democratic and Republican votes in precinct k for year t
- θ_{kt}: Relative proportion of Democratic votes in precinct k for year t
- $\mathbf{x_{kt}}$: Vector of covariates in precinct k for year t
- ψ_{kt}: Latent component incorporating spatiotemporal effects in precinct k for year t

Model form

$$Y_{kt} \sim \text{Binomial}(n_{kt}, \theta_{kt})$$

$$\log\left(\theta_{kt}/(1-\theta_{kt})\right) = \mathbf{x}_{kt}^{\top}\beta + \psi_{kt}$$

$$\psi_{kt} = \beta_1 + \phi_k + (\alpha + \delta_k)\frac{t-\bar{t}}{N}$$

$$\phi_k | \boldsymbol{\phi}_{-k}, \mathbf{W} \sim \text{Normal}\left(\frac{\rho_{int}\sum_{j=1}^{K} w_{kj}\phi_j}{\rho_{int}\sum_{j=1}^{K} w_{kj} + 1 - \rho_{int}}, \frac{\tau^2_{int}}{\rho_{int}\sum_{j=1}^{K} w_{kj} + 1 - \rho_{int}}\right)$$

$$\delta_k | \boldsymbol{\delta}_{-k}, \mathbf{W} \sim \text{Normal}\left(\frac{\rho_{slo}\sum_{j=1}^{K} w_{kj}\delta_j}{\rho_{slo}\sum_{j=1}^{K} w_{kj} + 1 - \rho_{slo}}, \frac{\tau^2_{int}}{\rho_{slo}\sum_{j=1}^{K} w_{kj} + 1 - \rho_{slo}}\right)$$

$$\beta \sim \text{Normal}(\mathbf{0}_{p\times 1}, 100000\mathbf{I}_p)$$

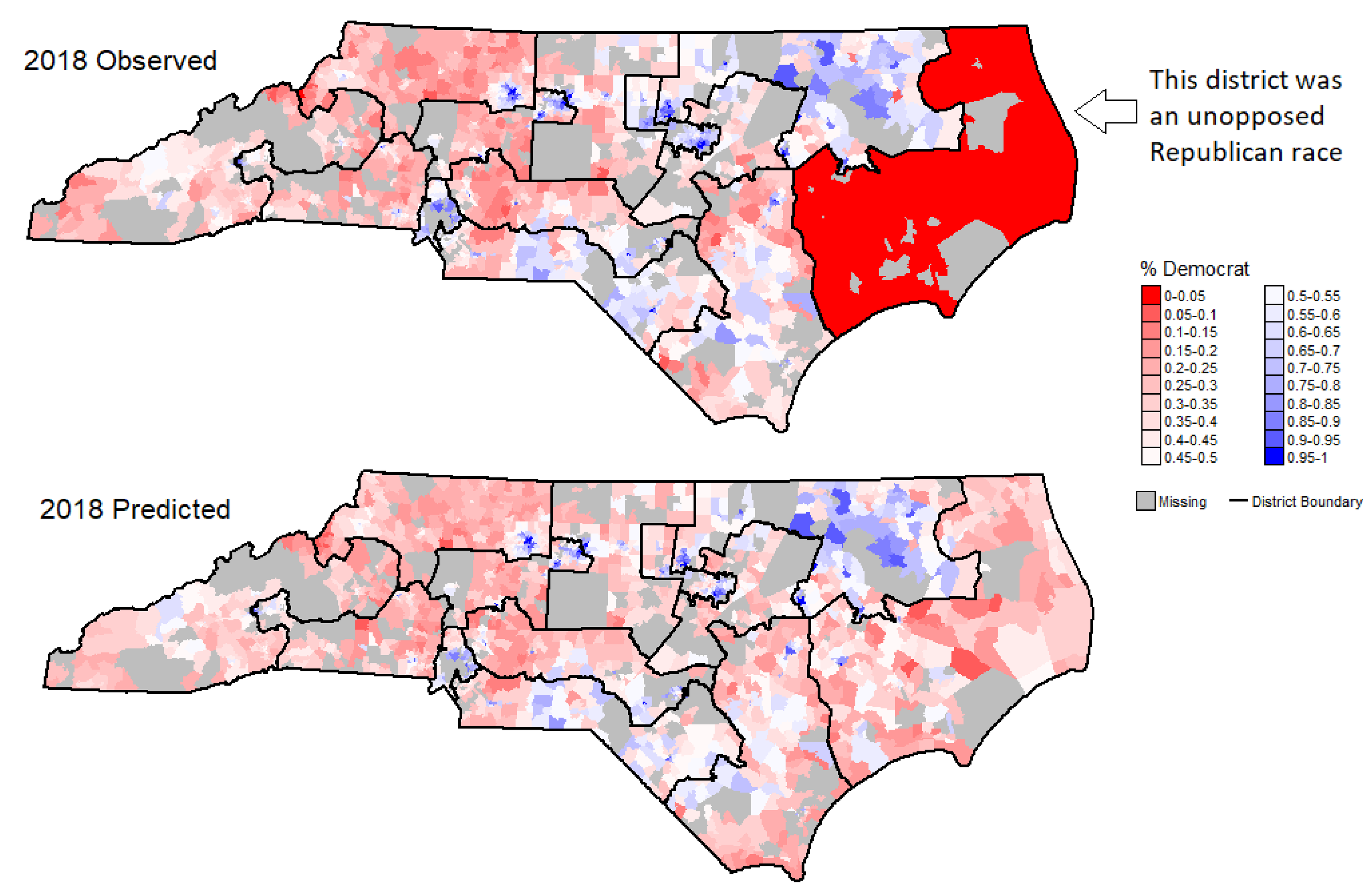
$$\tau^2_{int}, \tau^2_{slo} \sim \text{Inverse Gamma}(1, .01)$$

$$\rho_{int}, \rho_{slo} \sim \text{Uniform}(0, 1)$$

$$\alpha \sim \text{Normal}(0, 1000)$$

Model Prediction

Figure: Actual and Predicted 2018 precinct-level outcomes, by % of Democratic votes



Model Explanation

- Many spatiotemporal CAR models in literature
- Used `CARBayesST` package [4], with 8 possible model options
- Decided on this model to capture temporal trends in each precinct
- The random effect ψ_{kt} incorporates both the spatial effect ϕ_k and linear spatiotemporal trend $(\alpha + \delta_k\frac{t-\bar{t}}{N})$ for that individual location k. (int and slo refer to "intercept" and "slope".)
- ρ_{int} and ρ_{slo} are analogous to spatial dependence parameter in the original CAR model
- The terms ϕ_k and δ_k are modeled using their full conditional distributions, as determined by their neighbors (specified in adjacency matrix $\mathbf{W}$.)
- $\bar{t} = N^{-1}\sum_{t=1}^{N} t$, so $(t-\bar{t})/N \in [-\frac{1}{2}, \frac{1}{2}]$
- $t = 1, \ldots, T = 9$ is the number of years and $k = 1, \ldots K = 2045$ is the number of precincts
- $\sum_{j=1}^{K}\phi_j = \sum_{j=1}^{K}\delta_j = 0.$

Results and Conclusions

- Trained model on data from 2002-2016 elections
- Predicted on 2018
- Burn-in of 5,000 and sample of 50,000, thinning every 10, making posterior samples 5,000.
- Visual inspection of trace plots shows parameters have converged, though effect sample size for some is small. Posterior SDs are small enough not to worry.

	Median	2.5%	97.5%
(Intercept)	**3.86**	3.79	3.95
I(Pres. Elec. Year)	**0.10**	0.10	0.10
% Reg. Male	**-2.43**	-2.66	-2.28
% Reg. White	**-2.61**	-2.63	-2.59
% Reg. Age 26-40	**-0.45**	-0.49	-0.39
% Reg. Age 41-65	**-1.77**	-1.82	-1.73
% Reg. Age 66+	0.04	-0.01	0.09
τ^2_{int}	**0.55**	0.48	0.63
τ^2_{slo}	**1.08**	0.92	1.25
ρ_{int}	**0.19**	0.14	0.25
ρ_{slo}	**0.41**	0.33	0.51

Year	Fitted	Data	Truth
2002	6	6	6
2004	6	6	6
2006	6	7	7
2008	7	8	8
2010	4	8	7
2012	3	3	4
2014	3	3	3
2016	4	3	3
2018	3	4	2 or 3*

* Currently in dispute due to alleged ballot tampering.

Table: *Left.* Parameter posterior summary. Medians from credible intervals not containing 0 in bold. *Right.* Total Democratic district winners calculated by summing precinct results from the fitted counts and 2018 prediction and the data with missing precincts.

- Parameter posterior summaries reflect common knowledge about voters and voter turnout (Dems turn out more in presidential election years, males, whites and older generations vote more Republican than their counterparts)
- Model is close to correct number of winners for almost every year, despite missing data
- Impressive that this model did so well despite not using polling data. All popular models rely heavily on polling data

References

[1] Herschlag et al. (2018). "Quantifying Gerrymandering in North Carolina". arXiv.

[2] Fain, Travis. "Partisan gerrymandering case could force another congressional primary in NC". August 27, 2018. wral.com

[3] North Carolina State Board of Elections, ncsbe.gov

[4] Lee D, Rushworth A, Napier G (2018). "Spatio-Temporal Areal Unit Modeling in R with Conditional Autoregressive Priors Using the `CARBayesST` Package." *Journal of Statistical Software, Articles*, 84(9), 139.