Geostatistical Analysis in Space and Time

Logan Stundal, Benjamin Bagozzi, and John Freeman

August 9, 2021

Spatial temporal processes are at the heart of many important political questions. How do ideologies of mass publics evolve within and across state boundaries. How does electronic media affect the attitudes and behavior of citizens over time within counties in different states, attitudes and behaviors which through political communication diffuse to and from neighboring counties? How do candidates decide where and when to to raise funds ("search for gold") in their constituencies? How do their efforts spawn new contributions over time from locations outside their constituencies? Why and how do civil wars break out in and then spread across governmental jurisdictions dying out in some places but escalating in others?

Political methodology lags behind other disciplines like criminology, meteorology, environmental science, and epidemeology in its analysis of spatial temporal processes. We made real advances in spatial measurement (Monogan and Gill 2015; Cho 2003), causal inference based on geographical designs (Keele and Titiunik 2015) and the study of conflict dynamics (Brandt, et al 2008, 2014). But, many of our spatial analysis ignore or downplay temporal factors; they assess treatment effects at a single point in time. Or, we study spatial processes but in separate or highly aggregated slices of time (Cho and Gimpel 2007). Leading textbooks on time series in political science and economics (Box Steffensmeier et al. 2014, Enders 2010) say nothing about spatial-temporal processes. Difference in difference causal analysis use two time points to make grand assumptions about time trends in complex political process without any provision for medium and long term dynamics, let alone diffusion across (omitted) units of analysis (Keele and Minozzi 2013). Many conflict studies are macro level investigations in which spatial factors amount to country dyadic exchanges or exchanges between belligerants (Brandt et al. 2008). Subnational conflict studies make a "localization" assumptions–that conflict occurs over time within but not across units of analysis (Silverman 2021).

If they allow for spatial temporal factors in their analysis, political scientists most often place two way fixed effects in their models and, concomitantly,used cluster errors to account for any remaining unit heteroscedasticity. Occasionally, a simple (static) spatial model is used to assess the robustness of the investigator's results. As we explain below, conceptually, two way fixed effects models do not allow for diffusion of the effects of the variables of interest across units and they treat the effect of "time" as identical on all units in each time slice; two way fixed effects set-ups ignore the possibility of individual space-time unit differences in the effects of causal mechanisms.

Some of us have applied and made advances in neighborhood type spatial temporal modeling. These modelers carve up space into discrete units and include in regression models right hand side variables with both spatial lags with *prespecified* connectivity matrices of the dependent variable and also temporal lags of the values of the own unit's past values of the dependent variable. The development of such models for binary dependent variables along with techniques to assess spatial and temporal marginal effects is illustrative (Ward and Gleditsch 2002, Franzese et al. 2015). An innovative application of spatial vector autoregression in the study of subnational conflict also has been applied in international relations (Linke et al. 2012).

With one or two exceptions, what has not been applied and developed in political science, is the geostatistical approach to spatial temporal analysis. This approach treats the data generating process as continuous in space; it employs a set of techniques for "point process" modeling. Rather prespecifying a connectivity matrix, the geoestatistical approach *estimates* the extent of spatial interdependence between units and allows for unit variant temporal evolution of causal mechanisms like diffusion. Out of the geostatistical approach emerges methods for tracking *the distinct effects of causal mechanisms for every individual unit at each point in time.* Its is this methodology that has proven so useful recently in other disciplines.[1]

This research note explains a popular geostatistical approach and compares it to the more familiar two way fixed effects and neighborhood methods of modeling spatial temporal processes. We begin by reviewing the ways two familiar methods conceive of spatial temporal processes. The key idea of separability is introduced in this context. Geostatistical modeling based on a stochastic partial differential equation then is introduced. A formal definition of separability and the additional concepts of space-time stationarity and isotropy are discussed briefly here. We then the compare the three approaches. We use the modeling of insurgent attacks in Iraq–in particular, Silverman's recent (2021) two way fixed effects based investigation–as our benchmark. As an added payoff of our analysis, we compare the modeling results for ground truth (army reported SIGACTS), machine coded (ICEWS) and human coded (GED) databases; the rationale for our comparison of the databases is explained elsewhere (Bagozzi et al. 2018; Stundal et al. forthcoming).

We find....

1 Geostatistical Modeling Spatial Temporal Processes

Spatial temporal modeling is motivated by theoretical expectations predict and empirical observations which suggest the existence of certain causal mechanisms: as regards space, the possibility of strategic interaction, diffusion, spillover, relocation, and advection and as

[1] Geostatistical modeling is mentioned as a possible approach in one paragraph at the end of the second edition of Ward and Gleditsch (2019). And it has been applied in the form of kriging by by Monogan and Gill 2015, Gill 2020, Cho and Gimpel 2007. But these studies are static in character. The spatial-temporal versions of kriging, what is called, co-kriging, to our knowledge, have not been applied by American political methodologists. But see Pavia et al. 2008.

regards time, persistence, cumulation, and memory. Spatial temporal models aim to capture the *combination* of such mechanisms. The notion that civil conflict in a particular location in one part of country not only spills over to another part of the same or neighboring countries contemporaneously and historically at the same time the conflict persists (is remembered) within the unit in which it originated is illustrative. Schutte and Weidman (2011) are among those who explain why spatial-temporal interdependence should be expected in civil conflicts. They detect evidence of "escalation diffusion" in the Bosnian, Burundi, and Rwandan cases.

Two way fixed effects set ups are among the most common approaches political scientists use to study spatial temporal political processes. Assume we observe $i = 1 \ldots N$ units at $t = 1 \ldots T$ points in time. And we are interested in the effect of a set of covariates, X_{kit} on variable Y_{it}. This most simple model is of the form

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \epsilon_t \tag{1}$$

where Y_{it} is a Nx1 vector of dependent variables, X_{kit} is a Nxk matrix of covariates, α_i is the fixed (time invariant) affect of factors at each location i, γ_t is the fixed (unit invariant) affect of unknown factors at each respective time point, β is a coefficient to be estimated and ϵ_{it} is a Nx1 error vector which may be heteroscedastic. The fixed effects are separated and added together. There is no provision for interdependence between the values of Y_{it} of any units, there is no persistence or other kinds of temporal mechanisms governing the evolution of Y_{it}, and there is no variation in the impact of historical factors across units at each point in time.[2]

A second, familiar model is the so called Spatial Temporal Autoregressive (STAR) Neighborhood model. These are usually expressed as

[2]Other simple models are used in political science such as the unit fixed effects model with a lagged outcome variable, $Y_{it} = \alpha_i + \beta X_{kit} + \rho Y_{i,t-1} + \epsilon_{it}$. And more complete representations of the effect of history are often employed such as time polynomials. But, again, these alternative set ups ignore the possibility of spatial interdependence and treat the effect of history as unit invariant. On the latter point see the discussion Franzese et al. 2015 esp. pps. 152 and 165.

$$Y_{it} = \rho W Y_t + \phi M Y_{i,t-1} + \beta X_{kit} + \epsilon_{it} \tag{2}$$

As Franzese et al. (2007: 159) explain, W is the Kronecket product of a TxT identity matrix and an NxN prespecified spatial weights matrix ($I_t \bigotimes W_N$), M is an NTxNT matrix of zeros except for ones on the minor diagonal at coordinates (N+1,1), (N+2,2) ...(NT,NT-N), and ϕ is the unit invariant temporal autoregressive coefficient. Anselin (2006: Section 26.2.1)) argues that this set-up captures diffusion, persistence and the other causal mechansms enumerated above. The first term on the right represents spatial interdependence while the second term captures unit persistence and related temporal factors. Again, these factors are assumed to be separable and additive. Methods [3]

Neighborhood models may suffer from "inappropriate discretization" (Lindgren and Rue 2015: 3). The prespecified connectivity matrices used in neighborhood models treats spatial dependence as a step function—the same for some subset of units and nonexistant for another subset of units. The same is true of methods used to detect spatial-temporal interdependence such as that developed by Schutte and Weidman (2011) who use a "Moore neighborhood" to assess escalation diffusion and relocation. In reality the spatial dependence may vary *continuously* in space. The alternative—geostatistical spatial error models—produce *estimates* of both the range of spatial error dependence and of the site specific impact of unobserved factors.

[3]There is a related model that according to Anselin and others is best suited to analyzing spatial connections in the errors.Stundal et al. (2021) argue that conceptually, the spatial error model (SEM)for answering our questions about the validity of human and machine coded event data. SEMs capture model errors for neighboring units that cluster together—"smaller(larger) errors for observation $i \ldots$ go together with smaller [larger] errors for [neighbor] j" (Ward and Gleditsch 2019: 76). Errors also may be correlated because of the mismatch between the spatial scale of a process and the discrete spatial units of observations (Anselin 2006: 907). These error patterns correspond respectively to what researchers call the remoteness problem. For example, remoteness means that a model's underestimates of violence in a unit distant from a capital city correlate with underestimates of violence in a neighboring unit which is also distant from the same city. An example, is the spatial probit error model, SPEM. A technique based on the conditional log likelihood and variance-covariance matrix of the model can be used to estimate it (Martinelli and Geniaux 2017). The model provides estimates of a λ parameter which, with a row-standardized connectivity matrix, indicates the average dependence in the errors of a prespecified set of neighbors on the estimation error in a unit of interest.

Geostatistical models analyze point referenced data. These models are based on the idea of a continuous spatial domain. For example, even though terrorist events are observed at specific locations and therefore "inherently discrete" these events are interpreted as realizations of a continuously indexed space-time process of violence perpetrated against civilians and government officials (Python et al. 2017, 2018). Formally, the data are defined by a process indexed in time and space: $Y(s,t) \equiv [y(s,t),(s,t) \in D \subset R^2 xR)$. The spatial-temporal covariance function for the process is written as $Cov(y(s_i,t)y(s_j,u)) = C(y_{it}, y_{ju})$. Under an assumption of stationarity in space and time (see below) the covariance function can be expressed in terms of a combination of spatial distance, $\Delta_{ij} = ||s_i - s_j||$, and temporal lag $\Lambda_{tu} = |t-u|$. Thus $Cov(y_i, y_{ju}) = C(\Delta_{ij}, \Lambda_{tu})$. [4]

As in neighborhood modeling, several concepts are especially important in geostatistical analysis. First, as in the fixed effects models and STAR setups, spatial and temporal factors often are assumed to be *separable*. This means that covariance function can be written as a sum or product of its spatial and temporal parts, for instance, $Cov(y_{it}, y_{ju}) = C_1(\Delta_{ij})C_2(\Lambda_{tu})$. Gneiting (2002) proposes a test for separability for modeling spatial-temporal data.[5] *Stationarity* is a second key concept. Stationarity in space implies that mean function of the process is constant in space and that the spatial covariance function depends only on the distance vector $s_i - s_j \in R$. A closely related concept is *isotropy*; this is the idea that the covariance function depends only on Euclidian distance, $||s_i - s_j|| \in R$ (rotational invariance). For instance, Bandyopadhyay and Rao (2017) provide a test for stationarity for irregularly spaced data. In time series analysis stationarity means, roughly, that the mean and all its autocovariances are unaffected by a change in its time origin. The Dickey Fuller test is an example of a test for nonstationarity. Because time is ordered and space is not, isotropy has no meaning in the spatial-temporal context (Harvill 2010: 375).[6]

[4]As regards the spatial location of the data, s, can vary continuously in a fixed domain; s is a two-dimensional vector with latitudes and longitudes (three-dimensional if altitudes are considered) This definition of the process is taken from Blangiardo and Cameletti 2015: 235-236.

[5]Harvill (2010: 375) explains that "covariance functions imply that small changes in the locations of observations can lead to large changes in the correlations between certain linear combinations of observations."

[6]Technically this definition of spatial stationarity is a definition of second order stationarity. "Formally,

A popular geostatistical approach is Continuous Domain Bayesian Modeling with Integrated, Nested Laplacian Approximation (INLA).[7] Briefly, this approach "does not build models solely for *discretely observed data* but for approximations of *entire processes* defined over continuous domains" (Lindgren and Rue 2015:3, emphasis in the original). It assumes that the data generating process is a Gaussian field, $\xi(s)$, where s denotes a finite set of locations, $(s_1 \ldots s_m)$. As such it suffers from a "big n problem;" analyzing the Gaussian field is costly computationally (Lindgren et al. 2011). Therefore, a particular linear, stochastic partial differential equation is assumed to apply to the Gaussian field:

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}}(\tau\xi(s)) = W(s), \qquad s \in D \tag{6}$$

where Δ is a Laplacian, α is a smoothness parameter such that $\alpha = \lambda + 1$ (for two dimensional processes), $\kappa > 0$ is a scale parameter, τ is a precision parameter, the domain is denoted by D, and $W(s)$ is Gaussian spatial white noise. The stationary solution to this equation is a Gaussian field with the Matérn covariance function:

$$Cov(\xi(s_i), \xi(s_j)) = \sigma^2_{\xi_i} \frac{1}{\Gamma(\lambda)2^{\lambda-1}} (\kappa \mid\mid s_i - s_j \mid\mid)^{\lambda} K_{\lambda}(\kappa \mid\mid s_i - s_j \mid\mid) \tag{7}$$

where $\mid\mid s_i - s_j \mid\mid$ denotes the Euclidean distance between locations s_i and s_j, $\sigma^2_{\xi_i}$ is the

the spatial process is a Gaussian field if for any $n \geq 1$ and for each set of locations $(s_1, \ldots, s_n)$ the vector $(y(s_1), \ldots, y(s_n)$ follows a multivariate Normal distribution with mean $\mu = (\mu(s_1) \ldots, \mu(s_n)$ and spatially structured covariance matrix Σ. [Then in terms of the notation in the text above] the spatial process is second order stationary if the mean function is constant in space, i.e., $\mu(s_i) = \mu$ for each i and the spatial covariance function depends only on the distance vector $(s_i - s_j) \in R^{2}$" (Blangiardo and Cameletti 2015: 193. In atmospheric studies second order stationarity might be violated due to togological factors. Jun and Cook (2020) propose to model terrorism and related phenomenon as nonstationary processes. As regards the temporal part of the process, let y_t be a time series. Then this series is *covariance stationary* if for all t and t-s,

$$E(y_t) = E(y_{t-s}) = \mu \tag{3}$$

$$E(y_t - \mu)^2 = E[(y_{t-s} - \mu)^2] = \sigma^2_y, \qquad var(y_t) = var(y_{t-s}) = \sigma^2_y \tag{4}$$

$$E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s \tag{5}$$

where μ, σ^2_y, and γ_s are all constants (Enders 2010: 64). Financial time series and, according to some scholars, approval of the President, are nonstationary.

[7] The following description draws from Blangiardo and Cameletti (2015: Chapter 6) and especially the passage on pps. 234-5 of Python et al. (2017). Extensions that allow for modeling nonstationary processes are reviewed Ingebritsen et al. (2015).

marginal variance, $\Gamma(\lambda) = \lambda!$, K_λ is the modified Bessel function of the second kind and order $\lambda > 0$. The distance at which the spatial correlation becomes negligible (for $\lambda > .05$) is the range, r. The solution to the SPDE implies that the formula for the marginal variance is $\sigma^2 = \frac{\Gamma(\lambda)}{\Gamma(\alpha)(4\pi)^{\frac{d}{2}}\kappa^{2\lambda}\tau^2}$ where $d = 2(\alpha-\lambda)$. And the formula for the range is $r = \frac{\sqrt{8\lambda}}{\kappa}$. In this way, the Gaussian field can be represented (approximated) by a GMRF. A finite element method using basis functions defined on a Constrained Refined Delaunay Triangularization (mesh) over a corresponding shapefile of latitude-longitude event data is used for this purpose.

A hierarchical Bayesian framework then can be used to model one's data. For the variable of interest in our application below—insurgent violence incidents (V) normalized by unit population—we employ:

$$V(s_i,t)|\theta_V, \zeta_V(s_i,t) \sim Gaussian[\mu_V(s_i,t)] \tag{8}$$

$$\mu_V(s_i,t) = \beta_{V0} + x_V(s_i,t)\beta_V + \zeta_V(s_i,t) + \epsilon_V(s_i,t) \tag{9}$$

$$\theta_V \sim p(\theta_V) \tag{10}$$

where $V(s_i,t)$ is the observation at point s_i at time t, x_V is a vector of covariates, $x_V(s_i,t) = (x_{V1}(s_i,t)\ldots, x_{Vq}(s_i,t))$, β_V is the corresponding vector of coefficients, $\beta_V = (\beta_{V1}\ \ldots \beta_{Vq})'$, $\zeta_V(s_i,t)$ is the Gaussian Markov Random Field, and $\epsilon_V(s_i,t) \sim N(0, \sigma^2_{\epsilon_V})$ is measurement error with variance $\sigma^2_{\epsilon_V}$. Cameletti et al. (2010) refer to the covariates as (separable) large scale factors and the ζ_V as small scale processes. The latter capture diffusion and related, *unit and time variant* spatial-temporal causal mechanisms.[8] The Gaussian Markov Random field aometimes is assumed to follow a first or second order random walk (Python et. al. 2018). For the first order random walk we have $\zeta_V(s_i,t)|\zeta_V(s_{i,t-1}) \sim N(\zeta_V(s_{i,t-1}, \sigma^2_{rw}))$ so that $cov(\zeta_V(s_i,t), \zeta_V(s_j,u)) = 0$ if $t \neq u$ and the spatial interdependence is $cov(\zeta_V(s_i), \zeta_V(s_j))$

[8]Python et al. (2018: 17) say the space time structure of the GMRF captures, in the form of their "aggregate effects," "the marginal effects of unobserved individual factors." Cameletti et al. (2010), Harvill (2010) and others interpret the GMRF as small scale factors manifesting diffusion, spillover, advection and other causal processes.

if t=u. Modeled in this way, the temporal dependence amounts to simple, *unit variant*, one lag persistence in the mean of the GMRF (not temporal dependence on more distant lags of $\zeta_V(s_i,t)$ nor on lagged values of the $\zeta_V(s_j,t)$ at other locations). In other cases (Python et al. 2016), the the GMRF is assumed to follow a first order autoregressive process: $\zeta(s_i,t) = \rho\zeta(s_i,t-1) + \psi(s_i,t)$ where $\psi(s_i,t)$ is a time independent zero mean Gaussian field with $Cov(\zeta(s_i,t),\zeta(s_j,u)) = 0$ if $t \neq u$ and $Cov(\zeta(s_i)\zeta(s_j)$ if t = u. We use the latter specification in our analysis below.[9][10]

Besides estimates of the effects of the covariates on the rate of violent events, this geo-statistical model produces useful estimates of the parameters in the GMRF—in particular, the mean and standard deviation of the latent field at each point in space and at each poin in time. These estimates thus tell us not just that spatial-temporal interdependence exists as in the Schutte and Weidman (2011) method, but *where and when* such interdependence exists.

Table 1 summarizes these three (separable) spatial temporal models along with some others that have been used to study civil conflict.

2 Three Approaches to Modeling Spatial-temporal Processes Compared

Much progress has been made in recent years in analyzing the subnational, microdynamics of civil war. Compensation studies are illustrative. They show that insurgent violence tends to be lower if counterinsurgents compensate civilians for various kinds of harm the counterinsurgents create. Informational mechanisms—expressed through "civilian agency"—account for the relationship between post-harm compensation and violence reduction; the

[9] This notation follows that used by Python et al. (2018):7-9.

[10] A more complex, dynamic set up in which the weights on the bases functions of the spatial representation are governed by a linear dynamic system is developed by Cseke et al. (2016) and used to analyze data from the Afghan War Diary.

Neighborhood, Discrete Space Time	Geostatistical, Continuous Space Time
-General Linear Models with Two Way Fixed Effects and Error Clustering Condra and Shapiro 2015 Berman et al. 2015 Weidman and Shapiro 2015 Blair, 2021; Silverman 2021	-Stochastic Integral Differential Equation Models (square lattice) Zammit-Mangion 2012
-Spatial Temporal Autoregressive Models Weidman and Ward 2010 Franzese et al. 2015	-Stochastic Partial Differential Equation Models (triangular mesh) Python et al. 2016,2018
-Spatial Vector Autoregressions Linke et al. 2012	-State Space Models Cseke et al. 2016
-Spatial Hierachical Linear Models Hagan et al. 2016	

Table 1: Types of Separable Space Time Models Used in the Study of Subnational Civil Conflict and Terrorism with Selected Citations

compensated civilians share key information to counterinsurgents , information which enable counterinsurgents to thwart subsequent violence. In this way, compensation produces *de-escalation diffusion* and(or) *relocation de-escalation.* As the citations Table 1 indicates, (single equation) fixed effects models usually are employed to establish this relationship. Iraq often is the testbed for the analysis.[11]

An important, recent study in this genre is Silverman's article in the 2021 volume of *International Organization.* Silverman's main model is:

$$Y_{i,j} - Y_{i,t-1} = \alpha(c_{it} - c_{i,t-1}) + \beta(d_{it} - d_{i,t-1}) + \gamma(e_{i,j} - e_{i,j-1}) + G_y + H_i + I_{g,y} + \mu_{i,j} \quad (11)$$

where Y_{it} is the level of insurgent attacks in Iraq district i at time t, c_{it} is the spending

[11]Authors like Condra and Shapiro (2012), Silverman (2021) and others go to great lengths to establish the exogeneity of compensatory variables. We do not address their efforts in this regard but rather focus on the way in which spatial-temporal mechanisms are treated in the compensation genre.

on condolence payments in Iraq district i at time t, d_{it} is the spending on supplemental form of condolence payments, Ruzicka parments, in Iraq district i at time t, e_{it} is a vector of time varying controls including the amount of collateral damage created by Coalition forces and insurgents, G_y is a set of half year fixed effects, H_i is a set of district fixed effects, and I_{gy} is a set of interactions between governorate level vote shares for Sunni parties in the 2005 parliamentary elections and half years intended to pick up "broad sectarian shifts such as the Sunni Awakening." Silverman uses the Army's SIGACTS-III database to measure violence; he estimates the model using OLS for the period February 2004-December 2019 using clustered errors and with and without population weights. He assumes conflict is "localized" and that it evolves strictly as a function of changes in the covariates; he makes no provision for unit specific changes in the rate of violence due diffusion nor any unit specific persistence of violence in time.[12]

The first column of Table 2 replicates Silverman's main results. Most important it shows that compensation payments reduce the rate of violence in Iraq in his study period. *Should we replicate Silverman's analysis using ICEWS and GED data for his dependent variable?*

—————————————-

STAR Models. Emphasize Anselin's (2006) argument about causal mechanisms captured by START model. Note Silverman doesnt analyze data with STAR setup Logan's new results for effects of compensation when allowing for unit invariant persistence in form of lagged dependent variable. Does this hold up with ICEWS and GED? Footnote Blair (2021) use of STAR as robustness checks and flaws in his STAR application ——————

Geostatistical reanalysis of Silverman. Use of district centroids as points. Specifications for triangularization and settings for prior distributions, etc. Admit to temporal discretization ala Cseke et al. 1747 top col 2. The effects of the localization assumption in Silverman

[12]The sources of Silverman's condolence payments in the DoD's Commanders Emergency Response Program (CERP) and the USAID's Maria Ruzicka Iraq War Victims Fund. Population data are taken from the World Food Programme database. His controls include coalition troop strength, population density, per cent urban and other variables. See his article for the sources of these data (2021: 9ff).

ought to be illuminated in a way not revealed by the STAR model.

Excerpt for this section: INLA was implemented in the programming language of R. R-INLA performs numerical calculation of posterior densities and in this regard it is more efficient than Markov Chain Monte Carlo methods (Blangiardo and Cameletti, 2015). We use the default priors in R-INLA for the covariate coefficients and measurement errors, i.e., we assume iid zero-mean normal distributions with precision equal to .001 for the vector β_V and ϵ_V. An improper uniform prior between $-\infty$ and ∞ is used for the intercept, β_{V0}. As regards the settings for the Gaussian Markov Random Field, ...For its random walk structure we assume $P(\sigma_V^2 > 1 = 0.01$.

3 Statistical Analyses of the External Validity of Machine and Human Coded Data for Iraq

4 References

Anselin, Luke 2006. "Spatial Econometrics" Chapter 26 in *Palgrave Handbook of Econometrics* Volume 1. Econometrics. Basingstone, Palgrave MacMillon, pps. 901-969

Bandyopadhyay, Soutir, and S.S. Rao. 2017. "A Test for Stationarity in Irregularly Spaced Data." *Journal of the Royal Statistical Society* 79(1): 95-123.

Bagozzi, Benjamin E., Patrick T Brandt, John R. Freeman, A, Kim, A. Palao, and Carly Potz-Nielsen. 2018 " The Prevalence and Severity of Underreporting Bias in Machine and Human Coded Data" *Political Science Research and Methods* 7(3): 641-649.

Berman Eli, Michael Callen, Joseph H. Felter and Jacob Shapiro. 2011. "Do Working Men Rebel? Insurgency in Afghanistan, Iraq, and the Philippines." *Journal of Conflict Resolution* 55(4): 496-528//

Biddle, Stephen, Jeffrey A. Friedman, and Jacob N. Shapiro. 2020. "Testing the Surge: Why Did Violence Decline in Iraq in 2007? *International Security* 37(1): 7-40.//

Blair, Christopher W. 2021 "Restitution or Retribution? Detainee Payments and Insurgent Violence" Unpublished Manuscript. State College, PA.

Blangiardo, Marta and Michela Cameletti. 2015. *Spatial and Spatial-temporal Bayesian Models with R-INLA* New York, John Wiley and Sons.

Box-Steffensmeier, Janet, John R. Freeman, Matthew P. Hitt, and Jon C. Pevehouse. 2014. *Time Series Analysis for the Social Sciences* New York. Cambridge U

Braithwaite, Alex and Shane D. Johnson. 2012 "Space-Time Modeling of Insurgency and Counterinsurgency in Iraq" *Journal of Quantitative Criminology* 28: 31-48.

Brandt, Patrick T., John R. Freeman and Philip Schrodt. 2014. "Evaluating Forecasts of Political Conflict Dynamics." *International Journal of Forecasting* 30(4): 944-962.

Brandt, Patrick T., Michael Colaresi, and John R. Freeman 2008. "The Dynamics of Reciprocity, Accountability, and Credibility." *Journal of Conflict Resolution* 52(3): 343-374.

Cameletti, Michela, Rosaria Ignaccolo, and Stefano Bande. 2011 "Comparing Spatio-temporal models for Particulate Matter in the Piemonte." *Environmentrics* 22: 985-996.

Center for Army Analysis. Deployed Analyst History Report–Volume 1. Analytic Support to Combat Operations in Iraq 2002-2011. 2012 Fort Belvoir, Virginia.

Cseke, Botond, Andrew Zammit-Mangion, Tom Heskes, and Guido Sanguinetti. 2016. "Sparse Approximate Inference for Spatio-Temporal Point Process Models." *Journal of the American Statistical Association*" 111(516)" 1746-1763.

Cho, Wemdy K. "Contagion Effects and Ethnic Contribution Networks" *American Journal of Political Science* 47(2): 368-387.

Cho, Wendy K. and James G. Gimpel. 2003 "Prospecting for Gold" *American Journal of Political Science* 51(2): 255-268.

Condra, Luke N. and Jacob N. Shapiro. 2012 "Who Takes the Blame? The Strategic Effects of Collateral Damage" *American Journal of Political Science* 56(1): 167-187.

Donnay, Karsten and Vladimir Filmanov. 2014. "Views to a War: Systematic Differences in Media and Military Reporting of the War in Iraq" *EPJ Data Science* 3 XXX

Enders, Walter. 2010. *Applied Econometric Time Series* Third Edition. New York, Wiley.

Franzese, Robert Jr. , Jude C. Hays, and Scott J. Cook. 2015. "Spatial and Spatial-temporal-Autoregressive Probit Models of Interdependent Binary Outcomes." *Political Science Research and Methods* 4(1): 151-173.

Franzese, Robert Jr. and Jude C. Hays. 2007. "Spatial-Econometric Models of Cross-Section Interdependence in Political Science Panel and Time Series Cross Section Data." *Political Analysis* 15(2): 140-164.

Gill 2020.XXXX

Gneiting, Thomas. 2002 "Nonseparable, Stationary Covariance Functions for Space-Time Data" *Journal of the American Statistical Association* 97(458): 590-600

Hagan, John, Joshua Kaiser, and Anna Hanson. 2016. *American Sociological Review* 81(2): 316-346.

Hammond, Jesse and Nils B. Weidman. 2014. "Using Machine Coded Even Data for the Micro-level Study of Political Violence" *Reaearch and Politics* July-September: 1-6.

Imai,Kosuke, and InSong Kim 2020. "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis* 29: 405-415

Imai, Kosuke, and InSong Kim 2019 "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63(2): 467-490.

Ingebrigtsen, R., F. Lindgren, I. Steinsland, and S. Martino. 2015. “Estimation of a Nonstationary Model of for Annual Participation in Southern Norway Using Replicates of the Spatial Field.” *Spatial Statistics*C 14: 338-364.

Harvill, Jane L. 2010 "Spatio-temporal Processes." wires.wiley. com/compstats. vol 2: 376-382

Jun, and Scott Cook 2020.

Keele, Luke and Rocio Titiunik. 2015. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23: 127-155.

Keele, Luke and William Minozzi. 2013. "How Much is Minnesota Like Wisconain? Assumptions and Inference in Observational Data." *Political Analysis* 21(2): 193-216

Linke, Andrew M., Frank W. Witmer, and John O‘Loughlin. 2012. "Space Time Granger Analysis of the War in Iraq: A Study of Coalition and Insurgent Action-Reaction" *International Interactions* 38: 402-425.

Lindgren, F. and H. Rue. 2015. "Bayesian Spatial and Spatial-Temporal Modeling with R-INLA." *Journal of Statistical Software* 63(19).

Lindgren, F., H. Rue, and J. Lindstrom. 2011. "An Explicit Link Between Gaussian fields and Gaussian Markov Random Fields. The Stochastic Differential Equation Approach (with Discussion)." *Journal of the Royal Statistical Society* Series B 73(4): 423-498

Martinelli, D. and G. Geniaux. 2017 "Approximate Likelihood Estimation of Spatial Probit Models" *Regional Science and Urban Economics* 64: 30-45.

Monogan III, J.E. and Jeff Gill 2015. "Measuring State and District Level Ideology With Spatial Realignment." *Political Science Research and Methods* 4(1): 97-121.

Pavia, Jose Manuel, Beatriz Larraz, and Jose Maria Montero. 2008. "Election Forecasts Using Spatiotemporal Models" *Journal of the American Statistical Association* 103(483):

1050-1059

Python, A., J.B. Illian, C.M. Jones-Todd, and M.Blangiardo. 2018. "A Bayesian Approach to Modeling Subnational Spatial Dynamics of World Wide Non-state Terrorism, 2010-2016." *Journal of the Royal Statistical Society* Series A pps. 1-22.

Python, A., J. Illian, C. Jones-Todd, and M. Blangiardo. 2017. "Explaining the Lethality of Boko Haram's Terrorist Attacks in Nigeria, 2009-2014." In *Bayesian Statistics in Action* R.Argento et al. eds. Springer.

Schutte, Sebastian and Karsten Donnay. 2014. "Matched Wake Analysis: Finding Causal Relationships in Spatiotemporal Data" *Political Geography* 41: 1-10.

Schutte, Sebastian and Nils Weidman. 2011. "Diffusion Patterns of Violence in Civil Wars." *Political Geography* 30: 143-152.

Shapiro, Jacob N. and Nils B. Weidman 2015. "Is the Phone Mightier Than the Sword? Cellphones and Insurgent Violence in Iraq" *International Organization* 69: 247-274

Silverman, Daniel 2021. "Too Late to Apologize? Collateral Damage, Post-Harm Compensation, and Insurgent Violence in Iraq." *International Organization* XX

Stundal, Logan, Benjamin E. Bagozzi, John R. Freeman, and Jennifer S. Holmes. forthcoming. "Human Rights in Space: Statistical Modelss of Machine Coded vs. Human Coded Data." *Political Analysis*

vonBorzykowski, Inkenand Michael Wahman. 2021 "Systematic Measurement Error in Election Violence Data: Causes and Consequences." *British Journal of Political Science* XXX

Ward, Michael D. and Kristian Skrede Gleditsch. 2019 *Spatial Regression Models* Second Edition. Los Angeles California. Sage Publications.

Ward, Michael D. and Kristian Skrede Gleditsch. 2002. "Location, Location, Location. A MCMC Approach to Modeling the Spatial Context of War and Peace." *Political Analysis* 10(3): 244-260.

Weidman, Nils B. and Michael Ward. 2010. "Predicting Conflict in Space and Time" 2010. *Journal of Conflict Resolution* 54(6): 883-901

Zammit-Mangion, Andrew, Michael Dewar, Visakan Kadirkamanathan, and Guido Sanguinetti. 2012. "Point Process Modeling of the Afgan War Diary". *PNAS* 109(31): 12414-12419.