Supplemental Appendix For: Human Rights in Space: Statistical Models of Machine Coded Vs. Human Coded Data[1]

Logan Stundal, Benjamin E. Bagozzi,
John R. Freeman, and Jennifer S. Holmes

August 3, 2021

Contents

[1]This research is supported by NSF Grant Number SBE-SMA-1539302. We thank Scott Cook, Jude Hays, Phil Schrodt, Clayton Webb and Nils Weidmann for comments.

1 Overview

In this supplemental appendix, we first provide a review of two past event data validation designs. This is followed by a general discussion of currently employed techniques for human- and machine-based event data geolocation. We next provide a detailed overview of the event coding processes and steps employed by our two primary—and widely used—event datasets of interest. We then discuss our Colombia event data subsetting, aggregation, and formatting decisions in detail. Finally, we present a series of (tabular and graphical) auxiliary model results for our discrete and continuous spatial models.

2 Past Event Data Validation Designs

Seminal works on validation of event data either ignore external validation or they do not offer a methodological framework to perform external validation. King and Lowe (2003: 619, 624) stress the importance of "independent evaluation" of event data which, in their case, is defined as validation by humans who did not develop the machine coding software.

They have the VRA Reader code 45,000 articles. King and Lowe assign the events in those articles to 157 bins, bins corresponding to the IDEA ontology. The bins then are divided into three groups: (i) bins containing at least five events, (ii) bins containing one to four events, and (iii) bins which are empty. To construct their internal validation test bed, King and Lowe randomly choose five events from each bin in the first group and all of the events in the bins in the second group. They also include in their test bed, twenty five randomly chosen events from among those for which the VRA Reader assigned a source and/or target but could not assign an IDEA category (King and Lowe 2003: 626-627). King and Lowe compare the codings by a handful of experts with those produced by one software routine and three undergraduates.[2] In their comparisons, King and Lowe ignore source-

[2]Based upon this approach, King and Lowe have 12 bins with no machine coded events/leads, and thus

target information in the leads, focusing instead on the assignment of events or leads to the final event-categories.

Schrodt and Gerner (1994) evaluate the validity of machine and human coded events data. Their analysis of internal validity is based on the correlations between their machine coded events and those produced by human coders independently at the U.S. Naval Academy.[3] As for external validity, Schrodt and Gerner (1994) compare their machine coded data to death counts tabulated by the Palestinian Human Rights Information Center and to "critical shifts" in historical narratives of Israeli-Lebanese and Israeli-Syrian relations. King and Lowe (2003) and Schrodt and Gerner (1994) both recognize that aggregation (temporal, event class, etc.) affects validation. King and Lowe (2003) point out that higher aggregations across event class and time leads to better validation results. They internally validate with the IDEA and WEIS event code levels aggregated to the more general cue categories; they do this before examining how these comparisons vary across the conflict-cooperation dimension. Both King and Lowe's and Schrodt and Gerner's studies are of limited dimensionality. They compare human coding of a single body of text with the coding produced by a single piece of software.

The validation exercises in King and Lowe (2003) and in Schrodt and Gerner (1994) are illustrated in Figures A.1 and A.2. As these schematics show, internal and external validation are distinct concepts. And the texts that are used to validate event data can be several levels removed from ground truth.[4]

---

cannot provide any leads of these event types to their coders, although they do include a sample of leads that were not classified by the machine.

[3]Schrodt reports in personal communication that he does not know the design that was used to produce the human coded events for this comparison.

[4]Recent important articles on the production of political text by news sources include Cook and Weidmann (2019) and Hellmeier et al. (2018). For a still more complex schematic of the sources behind newspaper reports of human rights violations see Davenport and Ball (2002).

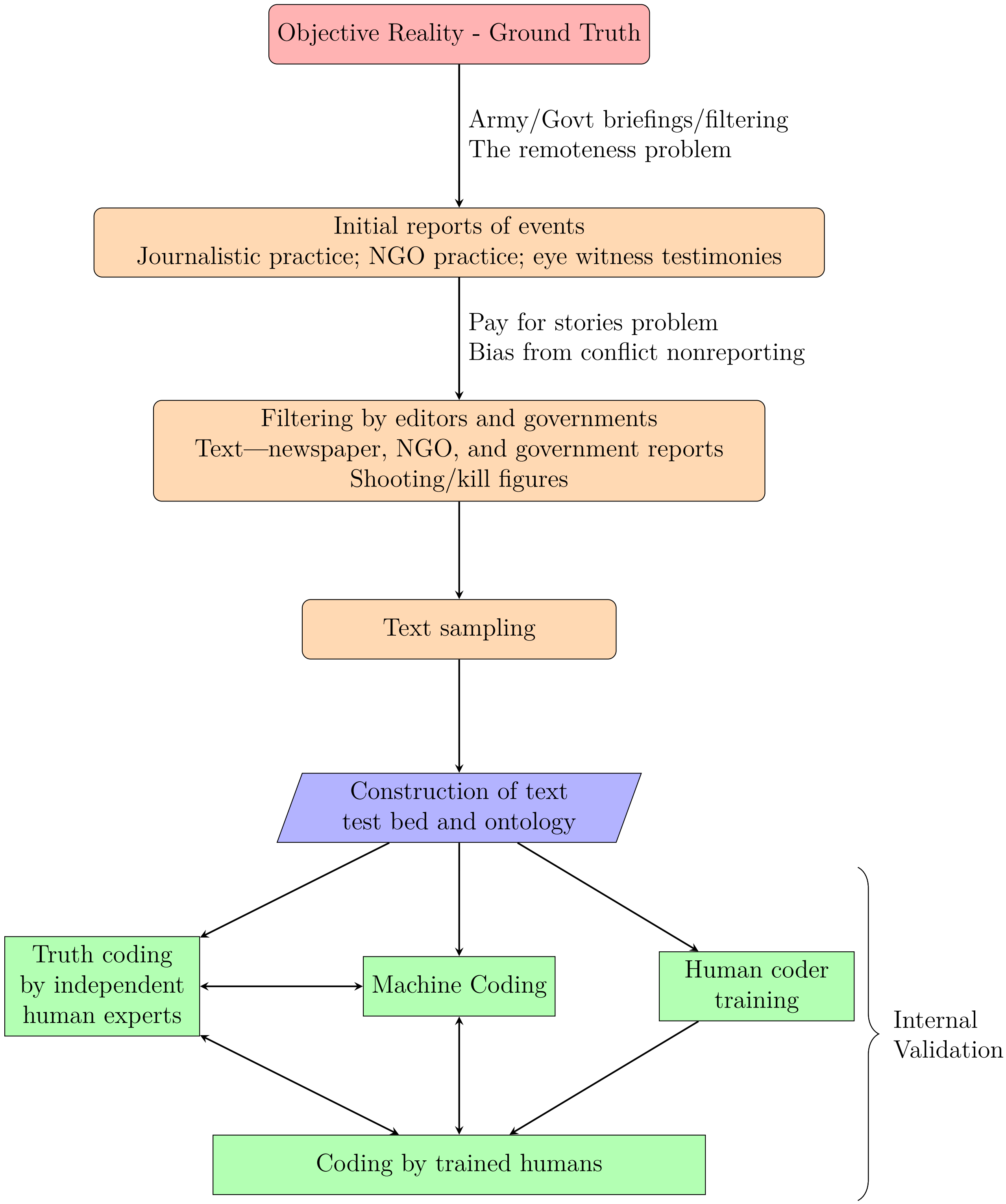


Figure A.1: **Validation in King and Lowe (2003)**

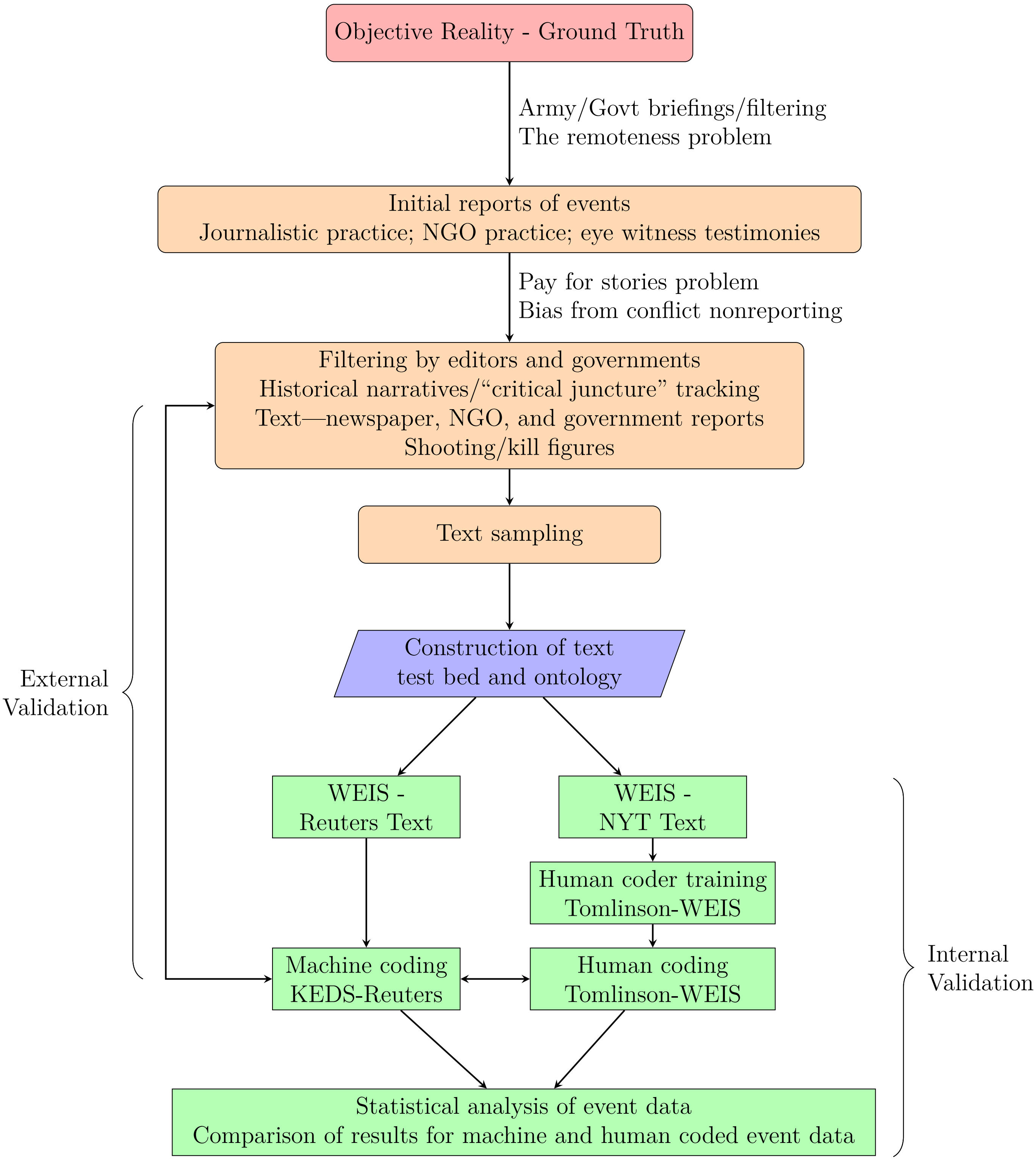


Figure A.2: **Validation in Schrodt and Gerner (1994)**

3 Spatial coding of events

Both human and machine coded event datasets primarily code events (in terms of who did what to whom, and where/when) from international news(wire) reports. Based upon the location mentions in these reports, human coders use a wide range of supplemental sources when assigning locations to events. GED coders, for example, draw upon National Geospatial Intelligence Agency databases, Google Earth, maps produced by aid agencies, and field atlases. Using Global ISO 31662 standards for assigning administrative divisions, the GED then reports a seven point precision scale for identified geolocations (Sundberg and Melander 2014: 526; UCDP Codebook pps. 14, 22-24). Another example is the Political Instability Task Force's Worldwide Atrocities Dataset (PITF). The PITF primarily relies on human coder lookups of identified place names in news articles via the GeoNames geographical database, alongside additional resources when necessary such as Google Search (PITF 2009). Location is assigned with village- or city-level precision unless otherwise noted. A third example is the Cline Center's Social Political Economic Event Dataset (SPEED). SPEED similarly relies on the GeoNames database for geolocation (Nardulli et al. 2019). It proceeds first in an automated fashion by identifying place names in relevant articles with natural language processing and then passes these place names to GeoNames to obtain confidence scores associated with potential locations for each event. Human coders are presented with the latter information via drop down menus for actual event geolocation.

Machines fully automate the geolocation process described for SPEED above. Lautenschlager, Starz, and Warfield (2017) and Lee, Liu, and Ward (2018) each characterize the automation of the political event geolocation process as following three sequential steps. First, named entity recognition (NER) is used to identify the words in a given news article that correspond to location names. Second, each location name that is identified within a given news article is disambiguated to establish that location name's most likely true geographic location. GeoNames or similar geographic databases (i.e., gazatteers) are typically employed during this second step, alongside additional contextual information from the original news text. Third, the (potentially multiple) disambiguated location name(s) recovered in step two are then evaluated to identify an optimal geographic location for a given news article's (separately machine coded) political event.

Several automated routines have been proposed to implement the three steps described above. One common approach employs the CLIFF/CLAVIN geolocation software (D'Ignazio et al. 2014), oftentimes within additional automated event extraction software such as PETRARCH-2 (Norris, Schrodt, and Beiler, 2012). CLIFF/CLAVIN is currently implemented in both the historical and real-time Phoenix event data projects (OEDA 2016, Althaus et al. 2017). Lee, Liu, and Ward (2018: 5) characterize the CLIFF/CLAVIN approach as one that leverages the frequency of location mentions to associate the proper disambiguated geolocation with a particular event. Halterman (2019: 3) criticizes this CLIFF/CLAVIN geolocation process for Phoenix[5] because it identifies a single "top" geolocation for an event in a manner that does not leverage any additional available information pertaining to the event itself. Lee, Liu, and Ward (2018: 5) likewise point out that ICEWS' comparable geolocation routine only selects a single most likely geolocation for each event based upon a statistical ranking of all NER-identified locations in a news article. Given the potential inaccuracies in this single-shot geolocation assignment approach, recent research has proposed improved machine-based geolocation steps that leverage additional natural language processing (NLP) and machine learning to better disambiguate and assign geolocations to events (Halterman 2017, 2019; Lee, Liu, and Ward 2018). At the time of writing, these innovations have not yet been widely integrated into existing event data projects.

While extant machine-based geolocation approaches identify a single latitude-longitude 2-tuple for relevant events, these coordinates do not always correspond to the city or village level. They can instead correspond to a municipality, department, or country centroid, depending on the location identified by NER. As such, several machine coded event datasets report the name of the most spatially accurate geographic unit associated with

[5] As well as the geolocation steps implemented within the Global Database of Events, Language and Tone (GDELT) dataset (Leetaru and Schrodt 2013).

each event, separately from the latitude-longitude 2-tuple recorded. For example, ICEWS includes latitude-longitude 2-tuples for all geo-located events, but additionally reports separate variables for "city," "district," "province," and "country." For these variables, location names only appear for an event's relevant level(s) of geographic accuracy, and are missing otherwise. This allows one to recover geo-coding precision for each event to produce the kind of precision estimate that is provided by (e.g.) GED, albeit with less granularity. To evaluate the precision and accuracy of ICEWS' subnational geolocations, Lautenschlager, Starz, and Warfield (2017) assess multiple random samples from the full ICEWS event dataset, in some cases with trained human evaluators. Based on their assessments, ICEWS was found to geolocate 85% of all events to the subnational (i.e., below "country") level, and to exhibit an accuracy rate in its subnational geolocations of 78% (pgs. 341-342). We provide a more detailed explanation of ICEWS coding in the next section.

The above points notwithstanding, many past evaluations of the validity of geo-located event data find that machine coded data are less accurate than human coded data. For example, Althaus et al. (2018:20ff) report that, in comparison to their SPEED (human) coders, PETRARCH-2's CLIFF/CLAVIN machine coder missed country level geolocation information in roughly 30%-70% of event reports pertaining to recent protests, imposed curfews, and suicide bombings for Nigeria. Notably, they found that CLIFF/CLAVIN's performance was even poorer at the state and province levels in this context.

In a parallel stream of research, several researchers have sought to *internally validate* the geolocation of events when proposing new automated geolocation algorithms. Lee, Liu, and Ward (2018), compared the performance of their new two stage supervised machine learning algorithm for geolocation with the performances of the machine geolocation coders used by the creators of the Phoenix and ICEWS data sets relative to human-coded geolocations for the same collections of news stories regarding China, Syria, the Democratic Republic of the Congo and Colombia. Lee, Liu, and Ward show that their new geolocation algorithm is more accurate than the internal geolocation routines currently used by ICEWS and Phoenix, though none of these machine coders is able to accurately classify their human-labeled geolocations at levels of accuracy greater than 90% (2018, Table 3). Halterman (2019) likewise develops a novel machine learning approach for the geolocation of events at the sentence level. This approach leverages the verbs and place names included in event sentences within a neural network to appropriately associate geolocations with event verbs of interest. Halterman validates this approach via a collection of 8,000 hand-labeled sentences, and via descriptive comparisons of geolocated Syrian military offensives.

The studies described above are of tremendous value in gauging the *internal validity* of machine coded data and in advancing automation software. However, as we argued in the main article's introduction, all of them treat the expert (training set) coding as ground truth. Uncertainty in expert coding is not incorporated into the comparisons of human and machine coding in a systematic way. Confusion matrices and related tools are frequently employed in these internal assessments. Inferences about where errors are prevalent—such as the remoteness problem—depend the characteristics of the handful of sites used in the comparison of machine and human coding.

4 ICEWS and GED Data Generating Processes

This section provides an overview of the event coding processes utilized by ICEWS and GED. Attention is given to ICEWS' and GED's (i) news(wire) repositories and news(wire) sources, (ii) news(wire) article identification and article selection steps, (iii) actor and event coding processes, and (iv) geolocation procedures. Extant validity assessments pertaining to these various coding processes are highlighted for ICEWS and GED where appropriate. Further details on ICEWS' and GED's respective coding pipelines can also be found in the sources cited below, and especially in Raytheon BBN Technologies (2015), Lautenschlager, Starz, and Warfield (2017), Croicu and Sundberg (2015) and Sundberg and Melander (2013).

4.1 News Source Repositories

ICEWS and GED each primarily code events from news(wire) reports. In both cases, these news(wire) reports are international and domestic focus. GED additionally codes events from a selection of nongovernmental organization (NGO), inter-governmental organization (IGO) reports, and similar sources. For coding purposes, both event data projects primarily identify candidate news(wire) articles from existing news(wire) repositories, rather than by scraping or obtaining these candidate news(wire) articles individually from their original publishers. In the case of the W-ICEWS dataset to the ICEWS project (i.e., the ICEWS data component that is the focus of our analysis), candidate news(wire) articles are drawn from the Factiva Global News Monitoring Resource (Lockheed Martin 2021). The set of Factiva news(wire) sources that ICEWS considers for our time period of analysis rests at approximately 300 sources (Raytheon BBN Technologies, 2015). These news(wire) sources are multilingual in the sense that ICEWS considers candidate news(wire) articles written in English, Spanish, French and Portuguese; which as noted above are both international (e.g., Reuters) and regional/national (e.g., the Times of India) in their focus.

GED likewise uses Factiva Global News Monitoring as the primary source for its global newswire corpus. However, its source focus in this respect is on Factiva's newswire article records for five specific English-language newswire sources: Reuters News, Agence France Presse (AFP), Associated Press (AP), Xinhua and BBC Monitoring (Croicu and Sundberg 2015). These five sources together represent a significantly smaller news(wire) testbed in relation to ICEWS' 300+ sources mentioned above. However, GED then compliments its global newswire sample (to which it attributes 60% of all GED events) with a set of independently collected secondary sources known to have unique coverage of political violence—including local news sources and monitoring organizations (e.g., Radio Okapi or SATP), international NGOs (e.g., Human Rights Watch), IGO reports (from, e.g., the United Nations), governmental publications (such as Truth and Reconciliation Commissions), and academic publications (Croicu and Sundberg 2015). While these details for GED's and ICEWS' sources leave the relative number of sources used by each dataset to be ambiguous, our investigations into the Colombia FARC-to-civilian event samples (for 2002-2009) discussed further below suggest that ICEWS' corresponding events are drawn from roughly[6] 40 distinct sources, whereas GED's events are drawn from roughly 20 distinct sources.[7]

4.2 Article Identification & Retrieval

Using the Factiva Global News Monitoring Resource described above, ICEWS retrieves candidate news(wire) articles for coding via a set of automated queries of this full Factiva corpus that are implemented separately for English, Spanish, French, and Portuguese language news(wire) stories. For each language-specific query, date ranges are specified alongside a set of key search query terms and discard terms (specified in each case relative to a target language) that together ensure the broad retrieval of socio-political news stories whilst largely excluding news stories related to tertiary topics such as sports or entertainment (Raytheon BBN Technologies, 2015). This fully automated article retrieval process employs a mild form of deduplication during retrieval in an effort to minimize duplicate stories from the same publisher, same headline, and same date (ICEWS 2015). Partially at this stage and partially at the coding stage outlined below, ICEWS also omits purely domestic US events from its event coding routines and event data (ICEWS 2015). Prior to event coding, all retrieved news stories based on the steps described above are next translated to English in a fully automated fashion. Validation of the latter translation step for an ICEWS sample of French and Spanish language stories reported event coding precision levels of 61.7% (French) and 63.4% (Spanish) relative to a precision level of 74.3% for a comparable, English language news story sample (Raytheon BBN Technologies, 2015).

GED instead uses a two step process identify and retrieve relevant news(wire) articles

[6]The number of sources reported for both ICEWS and GED is approximate, given both event datasets' recorded source information is denoted via strings for source name, which at times include variants of a news(wire) sources' name as distinct entries.

[7]That being said, a substantial share of events across both datasets for the Colombia sample described below correspond to a smaller set of overlapping newswire sources, namely AFP, AP, and EFE News Service.

and reports for human coding. First, the global newswire sources mentioned above are queried in Factiva for selected date and country ranges with a set of English-language search terms designed to identify news(wire) reports pertaining specifically to political violence. This can be seen to be a more narrow subset of the news(wire) articles obtained via the ICEWS query process described in the paragraph above. Based upon the degree of country and/or conflict coverage obtained from GED's automated Factiva query step, a second pass is then implemented to identify and add additional local coverage of a relevant conflict and/or country. Per GED (Crociu and Sundberg 2015), the decision on whether—and which—secondary sources to add for event coding at this second stage depends upon project leaders' and area experts' reviews of the sufficiency of materials retrieved in the initial Factiva search query, with the aim of providing comparable textual coverage of the relevant conflicts and countries that are to be coded for events. Herein, articles and reports are retrieved for coding in correspondence with the UCDP's more broadly identified conflicts across the world, which require a threshold of at least 25 battle-related deaths for consideration (Sundberg and Melander 2013). However, proximate conflict-country years (ie., non-active years) that fall below this aggregate threshold are also coded for events by GED (Crociu and Sundberg 2015).

4.3 Coding Processes

For the news sources that are retrieved and then machine-translated by ICEWS via the steps described above, the ICEWS project then applies automated event and actor coding routines to each retrieved news(wire) story in an effort to code relevant socio-political events. These coding steps entail the application of shallow parsing routines—and related dictionary-based event and actor codings—to the first six (i.e., lede) sentences of each news(wire) story considered. The former, dependency parsing-based natural language processing (NLP) steps enable ICEWS' proprietary BBN-ACCENT event coding software to identify any potential source and target actors, and any event actions arising between these source and target actors, from each of the relevant lede sentences considered. Potential event actions, when identified, are next compared against dictionaries for action terms (and synonyms) for each of the CAMEO project's 300-plus event type taxonomy (Schrodt, Gerner, and Yilmaz 2009). Event matches are then mapped onto an appropriate CAMEO two-digit category (and three- or four-digit category, when information is available). Identified source and target actors are likewise evaluated against ICEWS' own specialized entity dictionaries, which encompass over 50,000 named and time-indexed entities and over 700 generic agent names (e.g., "police," or "protestor") for matches to relevant source/target actors and country-assignments (Lockheed Martin 2021). This latter actor information is then included in separate variable fields for source and target actors and sectors, when identified.

The quality of ICEWS' event coding precision (i.e., the proportion of ICEWS-coded events that were in fact events) was evaluated by ICEWS' current automated coding system (BBN-ACCENT).[8] Two human evaluators assessed random samples of 500 ICEWS-coded events for each two-digit CAMEO action category coded, as drawn from events coded during the 2011-2013 time period. Across CAMEO's 20 two-digit categories, ICEWS' BBN-ACCENT system had on average 75.6% precision—with a minimum precision of 58.7% (10: Demand) and a maximum precision of 88.1% (17: Coerce)—which was notably superior to past ICEWS event-coding natural language processing (NLP) routines (Raytheon BBN Technologies, 2015). Event coding recall (i.e., the proportion of true events that were in fact coded by ICEWS) was similarly evaluated via a random sample of 1,000 doubly-annotated news stories alongside a sample of 1,100 singly-annotated news stories—where the latter were designed to over-sample less frequent two-digit CAMEO action categories. For the former sample, human annotators generally disagreed in their own event category codings approximately half the time. In comparing ICEWS event codings to these human annotations, ICEWS' current event coding software obtained an average relative recall of 34%. This average relative recall declines to 24% if actor/agent codes are also considered alongside

[8]Also see Wang et al. (2016) for an assessment of ICEWS' protest event precision, which found superior precision to GDELT, but evidence of duplicate events in ICEWS.

an event's CAMEO action category; but increases to 49% if one considers only those events where both human annotators agreed upon an event's actor/agent codes (Raytheon BBN Technologies, 2015).

For GED, only events with identifiable source and target actors are coded. This is distinct from the coding rules of ICEWS, wherein ICEWS codes events arising from unknown or ambiguous source and/or target actors. For events that then meet the more stringent GED coding criteria mentioned above, at least two separate human coders code each event at distinct points in time, and with the aid of distinct coding procedures, so as to avoid cross-coder influence (Crociu and Sundberg 2015). Strict rules and protocols are utilized during this event coding to maximize consistency in coding attributes. Subsequent checks are then applied to verify that event attributes have been properly coded. The GED codebook notes that these checks encompass both (i) a set of over 50 automated tests and (ii) manual checks by a project leader (Crociu and Sundberg 2015). Sundberg and Melander (2013) describe these quality checks in further detail in noting that "[d]ata quality is at least triplechecked, where the coder first runs through a checklist of consistency and streamlining tests. Secondly, a project manager performs similar tests, as well as controls of the geocoding through a set routine of visualization. Thirdly, PHP and Python scripts are run on the data to check consistency across IDs, coordinates, fatality counts, and more" (525-526).

Eck (2012) compared GED's resultant subnational geo-located events to those of a second prominent geolocated event dataset (ACLED) for the conflict cases of Algeria (1997) and Burundi (2000). While this comparison was primarily focused on geolocation accuracy, it was also determined through these validation assessments that (i) GED generally had fewer duplicate codings of events (corresponding to 1% of GED's coded events for Algeria and 0% of GED's coded events for Burundi) relative to ACLED (comprising 7% and 12% of ACLED's events, respectively) and that (ii) GED exhibited a comparably low rate of missing (i.e., non-coded) events in relation to ACLED (of 3% and 2% for Algeria, and 0% and 0% for Burundi). On the other hand, Otto's (2013) assessment of GED's codings of one-sided violence for Afghanistan found that GED's events potentially misattribute unknown perpetrators to the Taliban in 47% of all cases reviewed. These misattributions were at least partly attributable to the ambiguities of news(wire) reports' discussions of perpetrators and intentionality when reporting on one-sided violence (Otto 2013).

4.4 Geolocation

ICEWS employs automated methods to geolocate the events that were identified and coded via the ICEWS coding steps described in the subsection above. A hybrid named entity-based approach—leveraging both (i) a detailed dictionary of location name[9] and (ii) fast-string matching algorithms—is first applied to identify candidate location names for geo-coding (Lautenschlager, Starz, and Warfield 2017). Named entity resolution (NER) is then performed to match each retrieved event or story's candidate location name to the geolocation names and coordinates contained within a modified version of the GeoNames gazetteer, along with assignments of location specificity (i.e., precision) levels. Where multiple matches are obtained, ICEWS selects only the single most likely geolocation based upon a statistical ranking of all NER-identified matches (Lee, Liu, and Ward 2018). Several participants in the ICEWS project evaluated ICEWS' subnational geolocation accuracy via multiple random samples of ICEWS, in some cases with trained human evaluators. ICEWS was determined to geolocate roughly 85% of all events to the subnational level, and to exhibit an accuracy rate in subnational geolocation to an appropriate country of 78% (Lautenschlager, Starz, and Warfield 2017).

GED solely uses human coders to identify relevant strings in news(wire) articles and reports for geo-coding. Once these strings are identified, GED then implements NER in either a human-directed or semi-automated fashion. For these tasks, and like ICEWS, GED relies on a hybrid set of gazetteers as a reference set for georeferencing identified events, which

[9]Which extend those of the GeoNames gazetteer, and encompass first and second order administrative districts, mid-to-large-sized cities, and additional geographical features while excluding names that commonly pertain to individuals or organizations (Lautenschlager, Starz, and Warfield 2017, 336).

it then more uniquely complements with local and historical maps (Sundberg and Melander, 2013; Crociu and Sundberg 2015). These hybrid sources represent a more expansive set of reference materials than that which is reported for ICEWS above. GED's geolocation routines then either rely on human coders to geolocate events from this hybrid gazatteer and map-based reference set, or on the semi-automated geolocation of events. In the latter case, GED utilizes supervised "semi-automatic geocoding is employed in a number of cases (mainly in Europe and the former Soviet Union), using Google Geocoding API, Yandex, and Bing" with subsequent human and automatic vetting procedures to verify geolocation accuracy (Crociu and Sundberg 2015, 23). Geolocations are then attributed based upon standardized geo-coordinates for the single most precisely mentioned location within a given event's corresponding news(wire) story or non-media based report. Related information on administrative divisions is then added (where relevant) alongside a designation for a given event's geolocation precision based upon a seven point scale. Together, these geolocation steps likely ensure more precise geolocation accuracy in relation to ICEWS. Past validation assessments are suggestive of this conclusion, in finding (e.g.) that the quality of GED's event geolocation information is far superior to that of other prominent global event datsets such as GDELT (Hammond and Wedmann 2014) and ACLED (Eck 2012).[10]

5 Formatting Decisions for Colombia Data

We aggregate our GED, ICEWS, and CINEP data on FARC-directed human rights violations at the Colombian municipality-level for the years 2002-2009. Due to distinct processes of geolocation and event coding, these datasets each exhibit different levels of spatio-temporal precision, have unique definitions of what ultimately comprises a human right violation event, and contain varying levels of specificity regarding the identities of violence perpetrators and

[10]For example, in Eck's detailed comparison of ACLED's and GED's geolocation accuracy for two relevant conflicts (Algeria in 1997 and Burundi in 2000), 25%-50% of ACLED's events were incorrectly coded on at least one of the coding dimensions considered, whereas such instances of incorrect coding arose in only 2%-5% of GED's events.

victims. These differences necessitate several important decisions when spatially aggregating and combining these datasets for comparison. What follows is a detailed discussion of our efforts to format and combine each of these event datasets in a manner that ensures that our retained events are as comparable as possible across all three sources.

We first formatted our (machine coded) ICEWS data (Boshee et al. 2016) to correspond as closely as possible to FARC perpetrated instances of material violence against civilians. As an initial step, this required that we identify and retain only those ICEWS events that were perpetrated by FARC in Colombia against civilian targets. To filter our events according to the latter (target actor) designations, we subset ICEWS to include only those events with target actors containing mention of "general population," "civilian," and/or "social." To then subset our events to correspond only to FARC source actors, we used the sourcename identifiers contained within the ICEWS data to retain any events that were perpetrated by any actor that was definitively associated with the FARC. This accordingly excludes events that were perpetrated by Colombia's main other rebel groups—the ELN and EPL. This also excludes an HRV events in ICEWS that were possibly FARC-directed, but were nevertheless only recorded in ICEWS as being perpetrated by "guerillas," "communist guerillas," "communist rebels," "rebels," or similar designations. We identified all FARC-based ICEWS events with the aid of ICEWS' sourcename variable. This variable includes the non-standardized, source-actor entity name identified within an event's corresponding ICEWS-coded news article. We selected only those events that had "FARC" or similar variants of the FARC rebel group as a sourcename[11] or that recorded a sourcename

[11] Including "Armed Rebel (Revolutionary Armed Forces of Colombia)," "Revolutionary Armed Forces of Colombia," "FARC Secretariat," "Activist (Revolutionary Armed Forces of Colombia)," "Armed Services Deserter (Revolutionary Armed Forces of Colombia)," "Armed Band (Revolutionary Armed Forces of Colombia)," "Armed Force (Revolutionary Armed Forces of Colombia)," "Rebel Commander (Revolutionary Armed Forces of Colombia)," "Rebel Group (Revolutionary Armed Forces of Colombia)," "Armed Gang (Revolutionary Armed Forces of Colombia)," "Armed Insurgent (Revolutionary Armed Forces of Colombia)," "Armed Opposition (Revolutionary Armed Forces of Colombia)," "Secretariat (Revolutionary Armed Forces of Colombia)," "Armed Separatist (Revolutionary Armed Forces of Colombia)," "Criminal (Revolutionary Armed Forces of Colombia)," "Death Squad (Revolutionary Armed Forces of Colombia)," "Guerilla (Revolutionary Armed Forces of Colombia)," "Guerrilla Leader (Revolutionary Armed Forces of Colombia)" "Insurgent (Revolutionary Armed Forces of Colombia)" "Insurgency (Revolutionary Armed Forces of Colombia)," "Kidnapper (Revolutionary Armed Forces of Colombia)," "Militia (Revolutionary Armed Forces of

as a particular individual that we identified as a member or leader of the FARC.[12] We then retained only those events with a FARC-associated sourcename.

For each FARC→civilian event identified via the approach described above, we next filter all retained events to only include events (1) occurring in Colombia during the years 2002-2009 that (2) were geo-located to the city/town level(s) of geographic precision. This was achieved by dropping any events that contained no textual information within ICEWS' "city" variable. With these city-specific (2002-2009) FARC→civilian events in hand, we next retained only ICEWS' CAMEO category 18 (ASSAULT) and CAMEO category 20 (USE UNCONVENTIONAL MASS VIOLENCE) events with the following three or four digit CAMEO codes:

180: Use unconventional violence, not specified below
181: Abduct, hijack, or take hostage
182: Physically assault, not specified below
1821: Sexually assault
1822: Torture
1823: Kill by physical assault
183: Conduct suicide, car, or other non-military bombing, not specified below
1831: Carry out suicide bombing
1832: Carry out car bombing
1833: Carry out roadside bombing
184: Use as human shield
185: Attempt to assassinate
186: Assassinate
200: Use unconventional mass violence, not specified below
201: Engage in mass explusion
202: Engage in mass killings
203: Engage in ethnic cleansing

The above steps generated the ICEWS-set of all Colombian human rights violation events

Colombia)," "People Associated with the Opposition (Revolutionary Armed Forces of Colombia)," "Militant (Revolutionary Armed Forces of Colombia)," "Terrorist (Revolutionary Armed Forces of Colombia)," "Guerilla Faction (Revolutionary Armed Forces of Colombia)," or "Combatant (Revolutionary Armed Forces of Colombia."

[12] In each case, we reviewed all individuals coded for Colombian events with source actor designations assigned as 'rebel," "separatist," or "insurgent" to determine whether an individual was in fact a FARC-based individual as opposed to an individual associated with the ELN or EPL.

involving FARC source actors and civilian targets that were coded to a city level of geographic precision for the years 2002-2009. We then applied an initial de-duplication criterion to ensure that only one event(-type) was recorded per day, source, and latitude-longitude coordinate. This step was necessary because ICEWS only does very mild de-duplication at the coding stage—effectively eliminating duplicate stories bearing the same publisher, headline, and date—while still allowing for some duplicate stories given (e.g.,) variation in headlines (ICEWS, 2015; Schrodt, 2015: 14). We next turned to further aggregating these retained events for our anticipated analyses. More specifically, we first aggregated our deduplicated ICEWS FARC perpetrated events to the municipality level by matching these events to shapefiles of Colombia's municipalities via latitude longitude coordinates. After this join, we obtain a mean (median) FARC event count for our 2002-2009 municipality sample of 1.1 (0.0). For the three municipality-subperiods of 2002-2004, 2005-2007, and 2008-2009 that we consider further below, the corresponding mean (median) FARC event counts are 0.475, 0.406, and 0.220 (0.00, 0.00, and 0.00), respectively. As a final step, we then dichotomized our aggregated spatial event counts to binary indicators of whether ICEWS recorded the presence (= 1) or absence (= 0) of at least one FARC-directed HRV within a particular municipality.

We next formatted the GED (Sundberg and Melander 2013) in a comparable manner to the ICEWS data described above. The GED is a (near-global) human-coded event dataset that draws on both news(wire) sources and non-governmental organization reports for its coding of individual events. We started by subsetting the GED to encompass only Colombia-based events for the years 2002-2009. For these Colombian events, we next retained all non-state perpetrated cases of violence against civilians (i.e., "one-sided violence"), while taking care to exclude any instances of violence against civilians that were perpetrated explicitly by Colombian drug cartels, the Colombian military or police, and government-affiliated militia groups. We then retained the subset of those events that had GED's standardized "FARC" source actor designation. These retained GED events were then aggregated and merged to

Colombian municipality templates for our period of interest, while taking care to omit any GED events whose levels of geocoding accuracy were determined to be too ambiguous to fit within the municipality administrative level.[13] After these formatting and aggregation tasks were complete, we combined our GED measures to our aforementioned municipality-level template for the 2002-2009 period using latitude-longitude coordinates. The mean (median) for these 2002-2009 municipality-level GED-based event counts is 0.278 (0.00). For the three municipality-subperiods of 2002-2004, 2005-2007, and 2008-2009 that we also consider, the corresponding mean (median) FARC event counts for our GED data are 0.168, 0.073, and 0.037 (0.00, 0.00, and 0.00), respectively. As above, we then dichotomized our municipality event counts to binary indicators of whether GED recorded the presence (= 1) or absence (= 0) of at least one FARC-directed HRV within a given municipality.

Finally, we aggregated and merged our CINEP validation data (CINEP 2008) to our municipality-level ICEWS and GED events for the 2002-2009 period. The CINEP data are originally stored at the event level, with information attached to each event for that particular event's perpetrator (source actor), year, and municipality—among other variables. The recorded events in our initial sample include only directed rebel (source) to citizen (target) violence events. Directed dyad interactions of this sort (1) facilitate the comparison of the events with the directed dyadic event information contained in ICEWS and GED, and (2) ensure that our analysis closely parallels the most common approach to event data coding and analysis within the field (i.e., dyadic relational interactions). Within CINEP's data, *source* actors are designated by the specific rebel group perpetrating a given human rights violation and the *target* of each event is inferred to be a civilian or group of civilians. To combine these data with our formatted ICEWS and GED events, we first collapse CINEP's recorded rebel-perpetrated HRV events to the unique event-ID level. We then subset CINEP's events to only include actual instances of "material," *FARC*-directed human rights violations, rather

[13]Specifically, we only retained events that GED indicated were either (i) geolocated within 25km of a known location or (ii) whose exact location was recorded with latitude and longitude coordinates.

than both material and verbal human rights violations.[14] After subsetting the CINEP data in these manners, we aggregated all remaining FARC perpetrated CINEP events to the municipality level for our years of interest, and merged these event counts to our formatted GED and ICEWS data. For our 2002-2009 period, these CINEP event counts saw a mean (median) of 0.733 (0.00); whereas for our respective subperiods we obtained municipality averages (medians) of 0.515, 0.144, and 0.073 (0.00, 0.00, and 0.00). We then dichotomized these CINEP event counts in the same manner as described for ICEWS and GED above.

6 Auxiliary Analysis

6.1 Selected municipalities for External Validity Assessments

A.1 presents reported FARC events from each of our three measures (CINEP, ICEWS & GED) for municipalities classified as proximate or remote from Bogota based on the distance an international journalist would need to travel from the country's primary international airport. These 10 municipalities were selected by identifying the areas with the highest CINEP-reported FARC activity and subsetting for the top- and bottom-5 based on municipality centroid distance from Bogota. Here we have classified those 5 municipalities most near Bogota as exhibiting "Journalistic Proximity" while those most distant as exhibiting "Journalistic Remoteness".

6.2 External Validity and Design Overview

This section presents a number of supplemental model results that are intended to serve as points of comparison to our main article's cross-sectional geostatistical (SPDE) models. Table A.2 summarizes these comparisons and extensions, with reference to not only the cross-sectional SPDE models reported in the main article but also the cross-sectional base-

[14] That is, we remove all non-material violence events (e.g., threats), including categories such as 'Threatens', 'Recruitment', and 'Collective Threats,' which altogether constituted roughly 63% of all FARC perpetrated violence events in CINEP for our period of analysis.

Department	Municipality	CINEP	ICEWS	GED	Pop. Density	Capital Dist.	Classification
Tolima	Alvarado	6	0	1	24.61	106.48	Journalistic Proximity
Tolima	Dolores	7	1	2	12.05	135.42	Journalistic Proximity
Caldas	Samana	13	0	1	29.28	144.71	Journalistic Proximity
Antioquia	San Francisco	6	0	0	15.95	168.78	Journalistic Proximity
Antioquia	San Luis	8	0	1	18.65	177.69	Journalistic Proximity
Norte de Santander	Tibu	13	7	4	11.14	447.93	Journalistic Remoteness
Arauca	Arauca	6	17	2	11.75	482.43	Journalistic Remoteness
Sucre	Ovejas	9	1	2	43.03	522.24	Journalistic Remoteness
Bolivar	El Carmen de Bolivar	13	0	1	62.66	535.94	Journalistic Remoteness
Cesar	Valledupar	8	0	1	73.51	593.07	Journalistic Remoteness

Table A.1: FARC Reporting for 10 Journalistically Remote and Proximite Municipalities

	Neighborhood Model (SPEM)	Neighborhood Model (SPEM)	Geostatistical Model (SPDE)	Geostatistical Model (SPDE)
Internal Validity Assessment	Human Coded Data [GED]	Machine Coded Data [ICEWS]	Human Coded Data [GED]	Machine Coded Data [ICEWS]
External Validity Assessment	Human Coded Data [GED,CINEP]	Machine Coded Data [ICEWS,CINEP]	Human Coded Data [GED,CINEP]	Machine Coded Data [ICEWS,CINEP]

Table A.2: Research Design

line (SPEM) models reported below.[15] The first row of Table A.2 illustrates our *internal validation* assessment. For this assessment, we separately estimate a SPEM (neighborhood model) for human coded (GED) and machine coded (ICEWS) event data in the subsections below. The main article then repeats this GED vs. ICEWS comparison for the (geostatistical) models described therein. Our assessment of internal validity evaluates how closely one's spatial inferences align when modeling human coded and machine coded data in these two manners. However, this analysis does not tell us the degree to which our spatial inferences, and any differences therein, are at all valid in relation to an *external* record of actual events. The second row in Table A.2 describes our *external validation* assessment. Here we again separately compare inferences based on SPEM and SPDE models of our human coded (GED) and machine coded (ICEWS) data. These comparisons are based on spatial models of a gold standard event database, namely, a collection of events collected independently by the aforementioned CINEP data.

[15]We then separately compare our cross-sectional SPDE results to our time period SPDE results further below.

For the incidence of conflict, human rights violations, and other dichotomous aggregations of event data, spatial binary outcome models are typically employed in analyses of the conflict data mentioned above and in the main article. The interpretation of these models are similar to those for interval measured data. But spatial binary outcome models are much more challenging to estimate. For example, the probit spatial error model for binary data addresses the "mismatch between the spatial delineation of the measurement and empirical presence of [a] variable of interest ... [or the presence of] an omitted variable that is itself spatially correlated" (Calabrese and Elkink 2014: 867).[16] The binary probit spatial error model can be expressed as:[17]

$$y = X\beta + u, \quad u = \lambda M u + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n). \tag{1}$$

where y is an n×1 vector of binary dependent variables, X is an n×k matrix of independent variables, I_n is the identity matrix of size n, β is a k×1 vector of coefficients, λ is a parameter in the interval [-1,1] to be estimated, and M is the pre-specified connectivity matrix for the errors. Substituting for u we have:

$$y = X\beta + (I_n - \lambda M)^{-1}\epsilon \tag{2}$$

so the variance of the error term, $v = (I_n - \lambda M)^{-1}\epsilon$ can be written as

$$\Sigma = E(vv') = \sigma^2((I_n - \lambda M)^{-1}((I_n - \lambda M)^{-1})'). \tag{3}$$

[16] Calabrese and Elkink give this example: "the presence of a particular natural resource in particular countries, the geographical zones in which the resource is present do not exactly match with the country borders. A measurement of the presence of these resources in countries is necessarily spatially correlated but as a nuisance rather than in a theoretically interesting sense" (20-14: 667). In our comparisons of two types of event data and ground truth data, this "nuisance" is an important indicator of the validity of machine vs. human coded geocoded data.

[17] This rendition of the binary probit spatial error model is taken from Martinelli and Geniaux (2017: 31). They argue that the multivariate normal covariance structure of binary spatial probit models makes them easier to estimate than spatial logit models.

Consistent and efficient estimates of β and λ are obtained by maximizing the following likelihood function:

$$L(\beta, \lambda) = \Phi_n(x \in A|\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \int_{A_1} \int_{A_2} \dots \int_{A_n} e^{-(\frac{1}{2}x'\Sigma^{-1}x)} \tag{4}$$

where the integral intervals are defined by $A = [A_i]_{i\in(1,\dots n)} = (a_i, b_i)_{i\in 1,\dots,n)}$,

$$a_i = \begin{cases} X\beta & \text{if } y_i = 0 \\ -\infty & \text{if } y_i = 1 \end{cases}$$

and

$$b_i = \begin{cases} \infty & \text{if } y_i = 0 \\ X\beta & \text{if } y_i = 1 \end{cases}$$

The EM algorithm, Bayesian Gibbs sampler, recursive importance sampling algorithm, and general methods of moments all have been used to estimate binary probit spatial models. On the basis of the results of recent Monte Carlo experiments, we employ a technique based on the conditional log-likelihood with univariate conditional approximation of the MVN probabilities and variance-covariance matrix of the model (Martinetti and Geniaux 2017). We are interested in learning if the estimates for λ are comparable for human and machine coded datasets, if the existence and magnitudes of model spatial error correlation are the same in both kinds of data. The estimates of the coefficients for (a)spatial covariates in the binary probit spatial error models[18] for the two kinds of data should tell us if, once the correlation of errors across units is accounted for, human and machine geocoding yield the same inferences about determinants of our dependent variable.[19]

[18]Note that the binary spatial probit model is formulated in terms of a latent variable as advocated by Franzese et al. (2016).

[19]To our knowledge, only one study has employed a neighborhood-type spatial model to evaluate human and machine geo-tagged events. Hammond and Weidmann (2014) use a logit SDL model with a temporal lag to explain a binary violence indicator at the grid cell-month level. The occurrence of such violence in

6.3 Neighborhood Spatial Data Analysis

Because the Colombia testbed closely resembles the design features of Martinelli and Geniaux's (2017) Monte Carlo experiments for the SPEM—for example, they include the case of 1,000 units and set the number of neighbors equal to four—we used the conditional log-likelihood estimator with a univariate conditional approximation of the MVN probabilities and variance-covariance matrix. We employed a queen connectivity matrix for the spatial error and exactly the same tuning parameters used by Martinelli and Gerniaux.[20] We report two sets of comparison models below, in each case using the same covariates that were used in our main article's geostatistical models. First, as a point of reference, Table A.3 briefly demonstrates the robustness of our results in situations where one employs a nonspatial probit, as opposed to an SPEM. Second, Table A.4 then reports the estimates for the SPEMs for each dataset and for underreporting in ICEWS- and GED-HRVs relative to CINEP.

	ICEWS	GED	CINEP	ICEWS Underreporing	GED Underreporting
Intercept	-5.275	-4.834	-5.598	-3.920	-3.875
	[-6.502, -4.049]	[-5.976, -3.692]	[-6.606, -4.590]	[-4.885, -2.956]	[-4.895, -2.855]
Dist. Bogota, km (log)	0.187	0.088	0.233	0.190	0.189
	[0.040, 0.334]	[-0.038, 0.213]	[0.125, 0.341]	[0.078, 0.302]	[0.067, 0.311]
Population (log)	0.332	0.329	0.357	0.192	0.174
	[0.248, 0.417]	[0.248, 0.409]	[0.280, 0.434]	[0.120, 0.263]	[0.102, 0.245]
TRI	0.001	0.006	0.008	0.004	0.005
	[-0.003, 0.006]	[0.002, 0.010]	[0.004, 0.011]	[0.000, 0.008]	[0.001, 0.009]
LogLik	-445.257	-456.720	-585.001	-515.653	-470.937
N	1116	1116	1116	1116	1116

Note: 95% confidence interval in brackets estimated with robust standard errors.

Table A.3: Probits: 2002-2009

Africa was recorded by humans in ACLED and GED and by machine in GDELT. Briefly, they find that only the ACLED and GED models indicate spatial dependence in patterns of violence. Moreover, the fitted SDL models for ACLED and GED confirm the conventional wisdom that violence is more likely in remote parts of countries. GDELT suggests the opposite. Hence they conclude that "there is clear evidence for a capital-centric geocoding pattern [bias] in GDELT" (2014: 5). We further discuss Hammond and Weidmann's study in the main article's Discussion section.

[20]Our sample size is 1,116 municipalities. With the queen connectivity matrix the average number neighbors for the municipalities is 5.7. Like Matinelli and Geniaux (2017: 32) we employ a sixth order Taylor approximation of the spatial term $(I - \lambda M)^{-1}$. This tuning parameter is important for managing the computational time for the Choleski decomposition of the variance covariance matrix.

	ICEWS	GED	CINEP	ICEWS Underreporing	GED Underreporting
Intercept	-5.475	-5.272	-6.850	-4.471	-4.222
	[-6.057, -4.893]	[-5.319, -5.224]	[-6.988, -6.712]	[-4.633, -4.309]	[-4.372, -4.073]
Dist. Bogota, km (log)	0.192	0.091	0.283	0.213	0.203
	[0.139, 0.245]	[0.071, 0.112]	[0.273, 0.294]	[0.179, 0.246]	[0.165, 0.240]
Population (log)	0.346	0.360	0.439	0.221	0.191
	[0.289, 0.403]	[0.359, 0.361]	[0.429, 0.448]	[0.211, 0.231]	[0.169, 0.212]
TRI	0.001	0.007	0.009	0.005	0.005
	[0.001, 0.002]	[0.006, 0.008]	[0.009, 0.010]	[0.004, 0.006]	[0.005, 0.006]
λ	0.347	0.492	0.684	0.579	0.493
	[0.335, 0.360]	[0.479, 0.506]	[0.665, 0.702]	[0.559, 0.599]	[0.471, 0.515]
LogLik	-436.898	-431.704	-494.029	-467.298	-446.376
N	1116	1116	1116	1116	1116

Note: 95% confidence interval in brackets.

Table A.4: SPEM Models: 2002-2009

Turning to the SPEMs in Table A.4, the model results for each dataset confirm the existence of spatial error dependence. In each case the spatial parameter, λ, is a reliable predictor. This implies, following the interpretation in Ward and Gleditsch (2019) and Anselin (2006), that the SPEM model errors are spatially correlated and there is mismatch between the actual spatial scale of FARC activity and the subnational (discrete) units of observation. That being said, the current SPEM results do *not* show that the (machine coded) ICEWS events are plagued by a remoteness problem. The coefficient on logged distance from Bogota for ICEWS is positive as is comparable to the coefficient for GED (and CINEP), indicating that the machine coded data are comparably detecting remote FARC HRVs on the periphery of Colombia. The coefficients for population are also positive and reliable across all models.

A number of key differences emerge when comparing the SPEM results from Table A.4 against the standard probit results in Table A.3. First, likelihood ratio tests always support the addition of the spatial dependence parameter, λ in the SPEM models over the non-spatial probit specifications. Additionally, by modeling the spatial dependence in the errors as we have done here, the overall efficiency of the estimates also increases as indicated by the smaller confidence intervals presented in the SPEM models. This also has direct implications for the interpretation of substantive covariates. For example, in the non-spatial probits, the GED model specification indicates no reliable relationship between our distance from Bogota indicator and GED-reported FARC activity. However, after accounting for spatial dependence the SPEM model results with its more accurately estimated standard errors reveals that the relationship between GED-reported FARC activity and Bogota distance much more closely matches ICEWS and CINEP (although, the GED measure is smaller with its upper bound estimate smaller than either of ICEWS' or CINEP's lower bound). As well, the improved precision of estimates for TRI in our ICEWS model exhibit a similar change with that variable now more closely matching the results in the GED and CINEP models.

Figure A.4 depicts the predicted probabilities of FARC events for the ICEWS and GED models. They are remarkably similar. Using these predictions, Figure A.3 presents the respective two receiver operating characteristic (ROC) curves using the CINEP FARC events as ground truth. These ROC curves are also indistinguishable with very similar areas under the curve of 69.2% [95% CI: 65.8, 72.5] for the ICEWS model and 69.5% [95% CI: 66.0, 73.0] for GED. As illustrated further below (Figure A.5), we reach similar conclusions if we only consider underreporting within our SPEM framework.

An alternative means of evaluating our ICEWS and GED data is to analyze a new dependent variable that represents when each event dataset failed to register (i.e., underreported) an actual FARC event, as recorded by CINEP. Columns 4 and 5 of Table A.4 report these results for cases when $ICEWS = 0$ and $CINEP = 1$ and when $GED = 0$ and $CINEP = 1$. Again, there is clear evidence of spatial error correlation; both of the relevant λ's in this case are reliable predictors. However, there is *no* evidence of a *unique* remoteness problem with ICEWS. The GED and ICEWS coefficient estimates for distance to Bogota are also very similar. These estimates imply that, as distance from Bogota increases, ICEWS and GED each demonstrate an increasing tendency to underreport FARC events relative to our gold standard CINEP data. Yet, controlling for this aspect of remoteness, we also find that

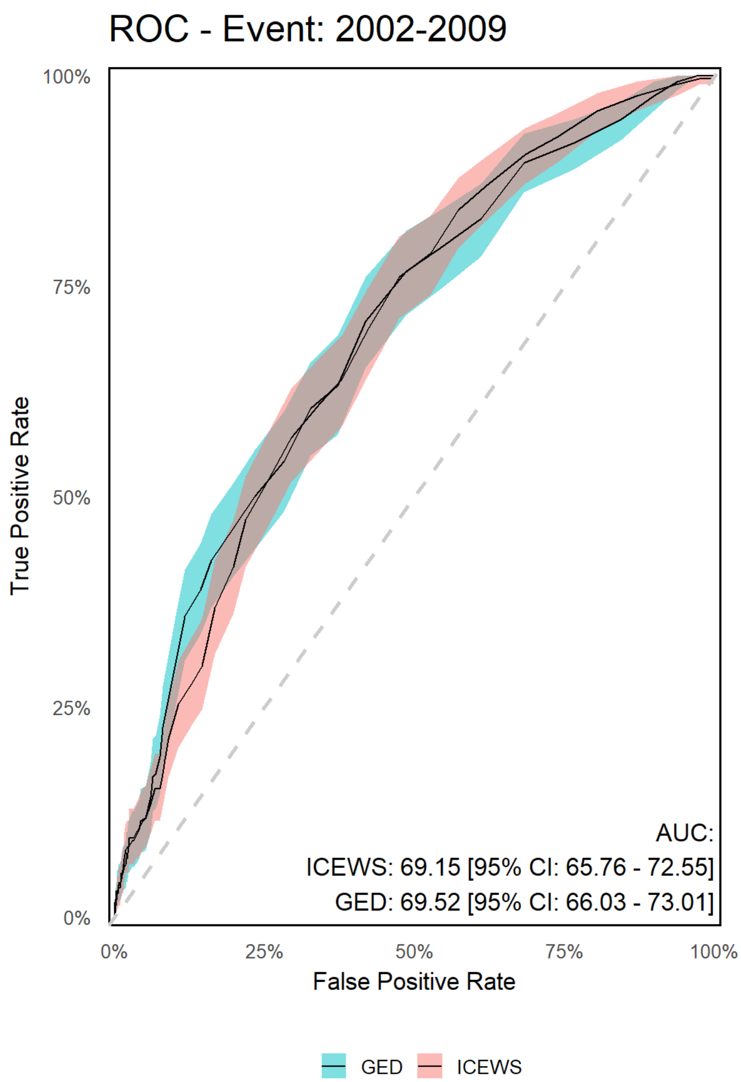


Figure A.3: SPEM ICEWS, GED Accuracy (ROCs)

underreporting in both ICEWS and GED is reliably higher for more populous Colombian municipalities. Finally, although TRI does have a positive effect on underreporting, in each instance it is small and likely negligible from a substantive perspective. In sum, columns 4-5 suggest the unsampled events in ICEWS and GED have a similar data generating process to the sampled events, and that there is no (severe) sample selection discrepancies that arise from analyzing the sampled events from the two datasets.[21]

[21] If we think about selection in terms of the law of total probability $E(Y|X) = E(Y|X, Z = 1)Pr(Z|X) + E(Y|X, Z = 0)(1 - Pr(Z|X))$ we have $E(Y|X)$ in column 3 of Table A.4, $E(Y|X = 1)$ in columns 1 or 2 of the same Table, and $E(Y|X, Z = 0)$ in columns 4 and 5. Therefore the results show that geocoding

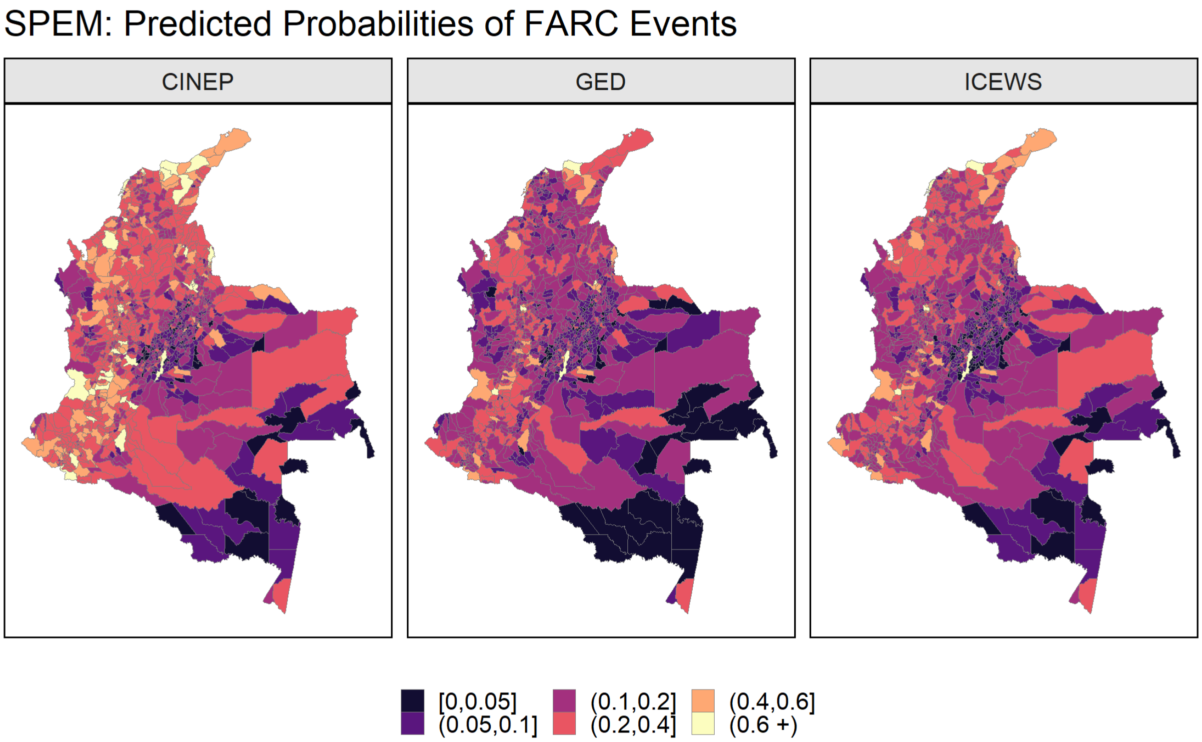


Figure A.4: Predicted Probability of FARC Events from SPEMs

error isn't substantially affecting inference in this case since columns 1, 2, and 3 indicate that the sampled relationships are equivalent to the population relationship and columns 1 vs. 2 and 2 vs. 5 show that $E(Y|X, Z = 1) \approx E(Y|X, Z = 0)$. We thank Scott Cook for this interpretation.

6.4 Auxiliary SPDE Estimates

	ICEWS	GED	CINEP	ICEWS Underreporing	GED Underreporting
Intercept	-5.662	-7.239	-8.324	-4.223	-3.336
	[-7.809, -3.512]	[-11.281, -3.917]	[-13.707, -4.154]	[-8.652, -0.771]	[-6.367, -0.588]
Dist. Bogota, km (log)	0.211	0.408	0.595	0.264	0.156
	[-0.111, 0.518]	[-0.081, 0.980]	[-0.040, 1.388]	[-0.266, 0.935]	[-0.261, 0.610]
Population (log)	0.355	0.368	0.376	0.140	0.110
	[0.260, 0.454]	[0.267, 0.473]	[0.275, 0.479]	[0.039, 0.241]	[0.008, 0.208]
TRI	0.006	0.014	0.018	0.010	0.007
	[0.000, 0.013]	[0.007, 0.020]	[0.011, 0.024]	[0.003, 0.016]	[0.001, 0.013]
Kappa	5.675	0.994	1.258	1.542	1.943
	[0.832, 24.057]	[0.387, 1.898]	[0.576, 2.301]	[0.617, 3.035]	[0.572, 4.727]
Sigma	1.265	0.792	2.021	1.039	0.621
	[0.064, 8.324]	[0.302, 1.565]	[0.928, 3.591]	[0.481, 1.861]	[0.196, 1.402]
Range	55.046	315.849	249.654	203.601	161.467
	[3.375, 156.224]	[123.612, 605.266]	[108.404, 436.257]	[76.615, 381.758]	[39.859, 348.192]
LogLik	-465.646	-452.414	-527.630	-492.083	-475.855
N	1116	1116	1116	1116	1116

Note: Point estimates reflect posterior median, 95% HPD in brackets.

Table A.5: SPDE Models of ICEWS, GED, and CINEP, 2002-2009

	ICEWS	GED	CINEP	ICEWS Underreporing	GED Underreporting
Intercept	-3.820	-5.566	-7.125	-5.881	-3.897
	[-6.177, -0.888]	[-9.154, -2.564]	[-11.724, -3.420]	[-10.661, -2.154]	[-7.183, -1.152]
Dist. Bogota, km (log)	-0.018	0.161	0.417	0.392	0.197
	[-0.453, 0.317]	[-0.262, 0.640]	[-0.148, 1.097]	[-0.177, 1.102]	[-0.218, 0.688]
Population (log)	0.278	0.339	0.355	0.227	0.137
	[0.177, 0.382]	[0.237, 0.445]	[0.254, 0.458]	[0.126, 0.328]	[0.036, 0.237]
TRI	0.004	0.007	0.016	0.011	0.009
	[-0.002, 0.011]	[0.001, 0.014]	[0.010, 0.023]	[0.005, 0.018]	[0.003, 0.015]
Kappa	1.189	0.886	1.293	1.186	1.707
	[0.217, 3.822]	[0.269, 1.924]	[0.562, 2.394]	[0.488, 2.323]	[0.513, 4.015]
Sigma	0.284	0.489	1.395	1.094	0.579
	[0.059, 0.692]	[0.142, 1.101]	[0.622, 2.544]	[0.454, 2.084]	[0.200, 1.214]
Range	263.306	354.148	242.868	264.714	183.858
	[33.073, 690.542]	[106.760, 765.170]	[102.152, 437.736]	[101.292, 488.888]	[48.321, 393.481]
LogLik	-369.389	-392.993	-504.872	-477.819	-456.209
N	1116	1116	1116	1116	1116

Note: Point estimates reflect posterior median, 95% HPD in brackets.

Table A.6: SPDE Models of ICEWS, GED, and CINEP, 2002-2004

	ICEWS	GED	CINEP	ICEWS Underreporing	GED Underreporting
Intercept	-8.326	-9.877	-5.214	-3.228	-3.877
	[-11.461, -5.813]	[-23.989, 5.736]	[-11.642, 0.316]	[-8.950, 2.097]	[-9.494, 1.561]
Dist. Bogota, km (log)	0.495	0.497	0.045	-0.082	0.017
	[0.131, 0.934]	[-0.232, 1.454]	[-0.883, 1.032]	[-0.978, 0.803]	[-0.900, 0.882]
Population (log)	0.409	0.420	0.300	0.163	0.184
	[0.295, 0.530]	[0.279, 0.570]	[0.162, 0.438]	[0.011, 0.310]	[0.033, 0.330]
TRI	0.005	0.019	0.007	0.004	0.002
	[-0.002, 0.014]	[0.009, 0.030]	[-0.002, 0.016]	[-0.005, 0.014]	[-0.007, 0.011]
Kappa	5.112	0.446	2.829	3.465	4.124
	[0.800, 15.876]	[0.107, 0.955]	[1.236, 5.809]	[1.122, 9.154]	[1.371, 11.183]
Sigma	1.334	1.679	4.141	4.547	5.764
	[0.089, 4.628]	[0.182, 6.215]	[1.829, 8.377]	[0.766, 16.012]	[1.211, 19.988]
Range	61.258	703.923	110.986	90.525	76.053
	[7.967, 172.096]	[209.010, 1744.483]	[40.597, 197.060]	[20.202, 185.206]	[16.474, 153.653]
LogLik	-290.302	-192.089	-299.203	-265.236	-273.844
N	1116	1116	1116	1116	1116

Note: Point estimates reflect posterior median, 95% HPD in brackets.

Table A.7: SPDE Models of ICEWS, GED, and CINEP, 2005-2007

	ICEWS	GED	CINEP	ICEWS Underreporing	GED Underreporting
Intercept	-7.026	-9.540	-7.999	-7.866	-7.260
	[-10.008, -4.852]	[-19.052, -2.996]	[-14.084, -3.107]	[-13.958, -3.126]	[-12.097, -3.044]
Dist. Bogota, km (log)	0.334	0.496	0.316	0.429	0.272
	[0.019, 0.696]	[-0.471, 1.776]	[-0.463, 1.209]	[-0.322, 1.331]	[-0.408, 0.969]
Population (log)	0.359	0.389	0.384	0.306	0.322
	[0.236, 0.502]	[0.220, 0.567]	[0.237, 0.539]	[0.154, 0.461]	[0.172, 0.480]
TRI	-0.005	0.013	0.016	0.015	0.015
	[-0.013, 0.004]	[0.002, 0.027]	[0.005, 0.027]	[0.004, 0.026]	[0.005, 0.027]
Kappa	5.260	0.753	2.282	2.532	4.066
	[0.340, 27.123]	[0.222, 1.642]	[0.883, 5.176]	[0.908, 5.840]	[1.135, 11.926]
Sigma	1.406	1.964	2.704	2.628	3.736
	[0.011, 23.437]	[0.621, 4.281]	[0.885, 6.330]	[0.876, 5.928]	[0.535, 15.674]
Range	59.129	416.541	137.521	123.964	77.105
	[2.063, 241.721]	[123.925, 912.523]	[41.722, 259.041]	[35.620, 242.569]	[13.525, 168.488]
LogLik	-192.443	-150.976	-212.514	-208.315	-184.795
N	1116	1116	1116	1116	1116

Note: Point estimates reflect posterior median, 95% HPD in brackets.

Table A.8: SPDE Models of ICEWS, GED, and CINEP, 2008-2009

6.5 Auxiliary Plots

6.5.1 SPEM - Underreporting model Receiver Operator Curves

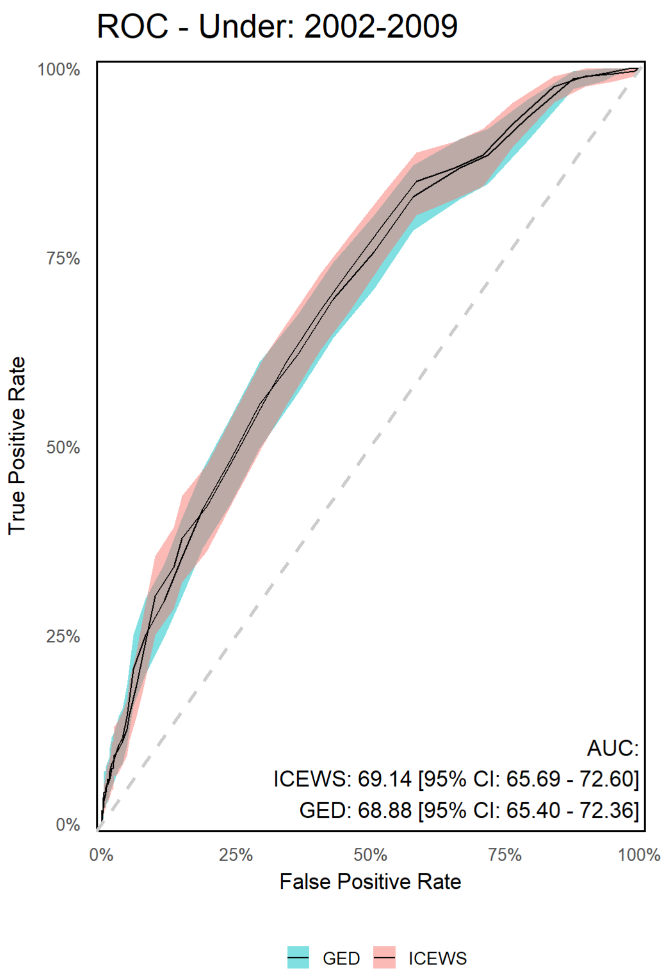


Figure A.5: SPEM ROCs for Underreporting, Cross-Sectional

6.5.2 SPDE - Receiver Operator Curves

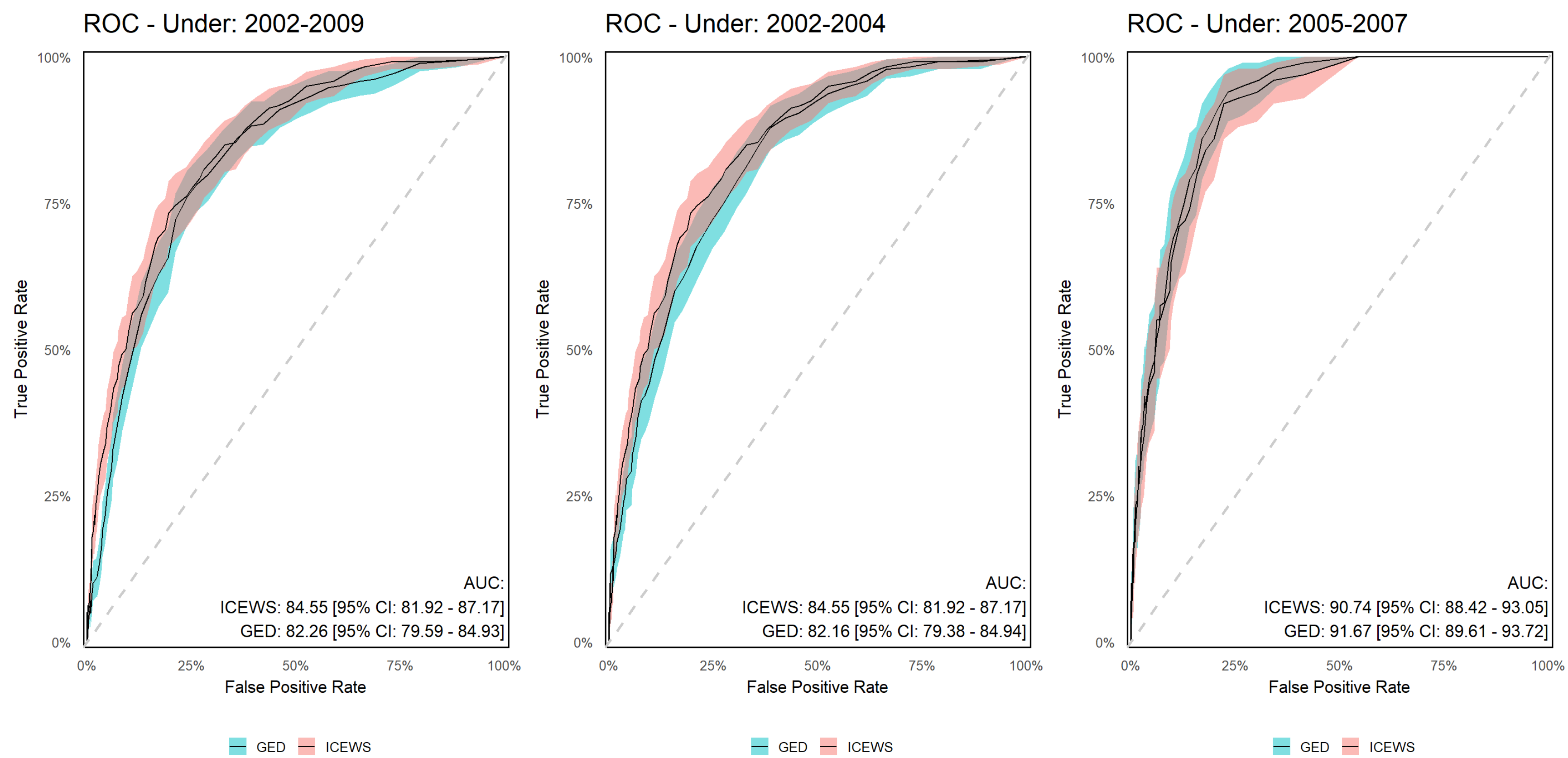
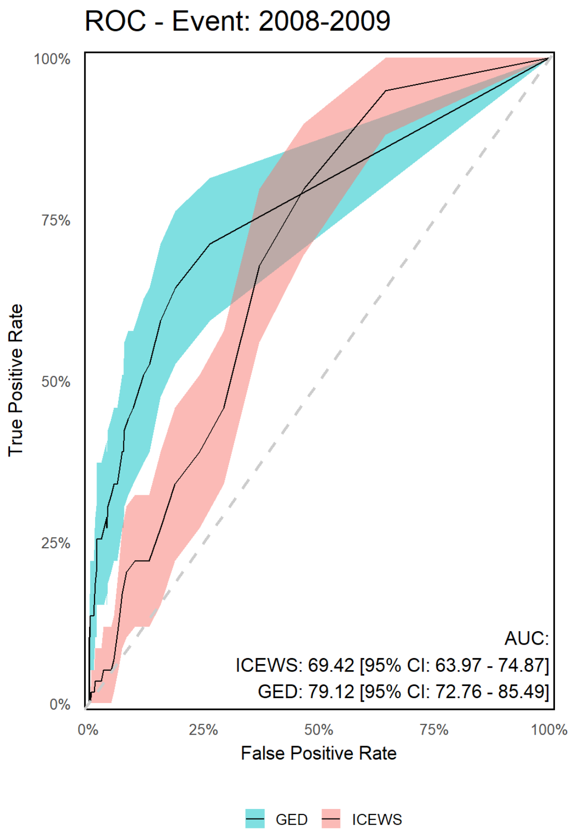


Figure A.6: SPDE ROCs for Underreporting



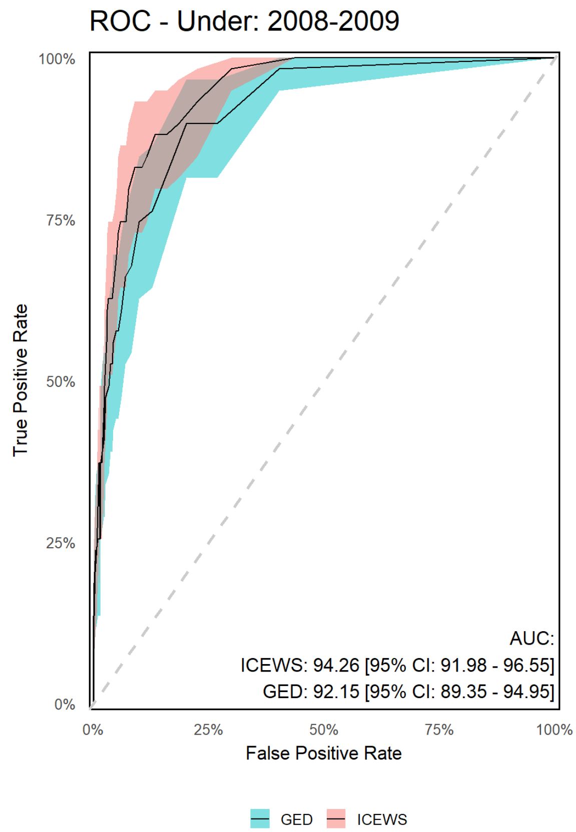


Figure A.7: SPDE ROCs, 2008-2009 Cross-Section

6.5.3 GMRF Estimates

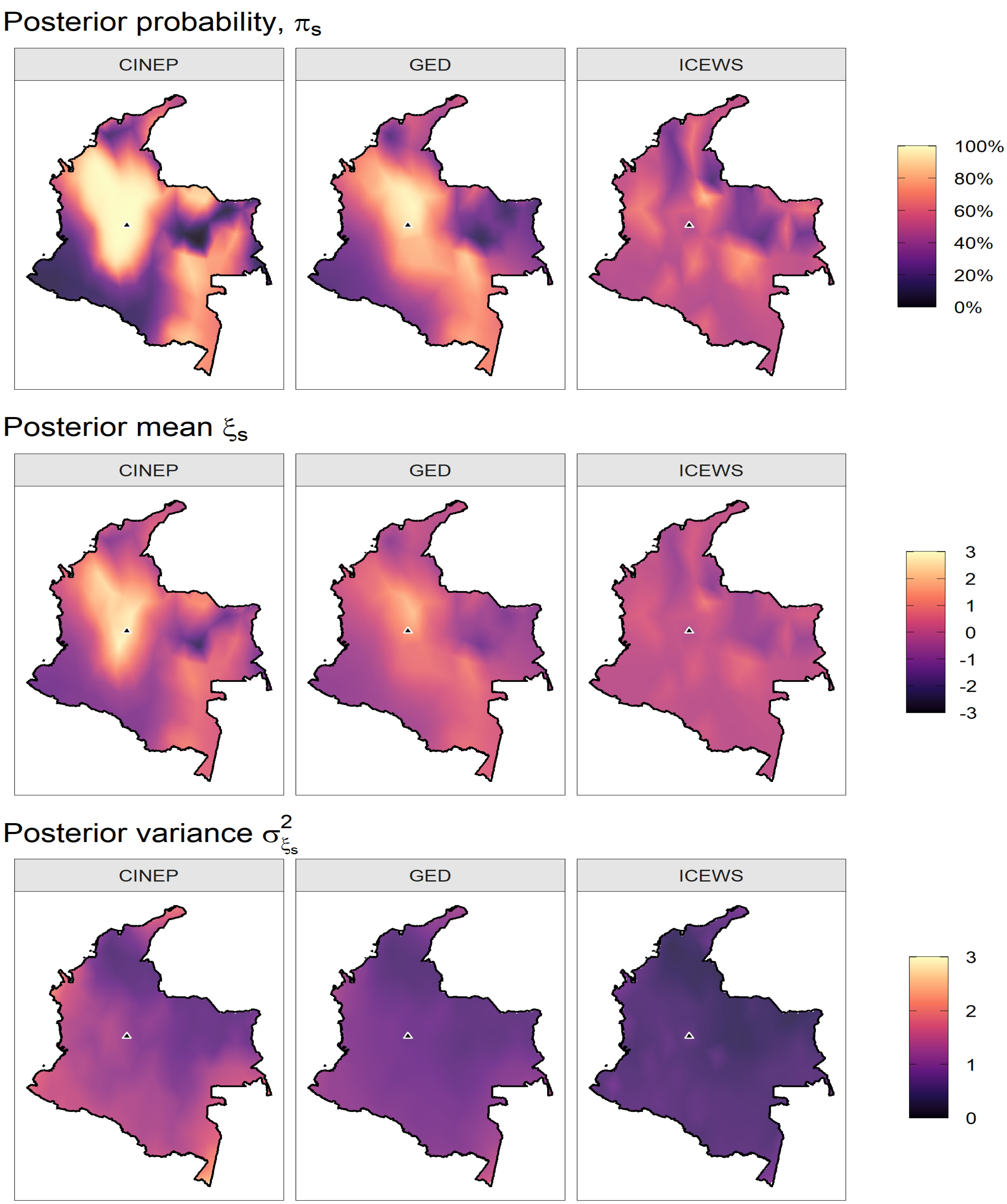


Figure A.8: SPDE: GMRF Estimates, Cross-Sectional

6.5.4 SPDE Range Estimates: 2002-2009 Results

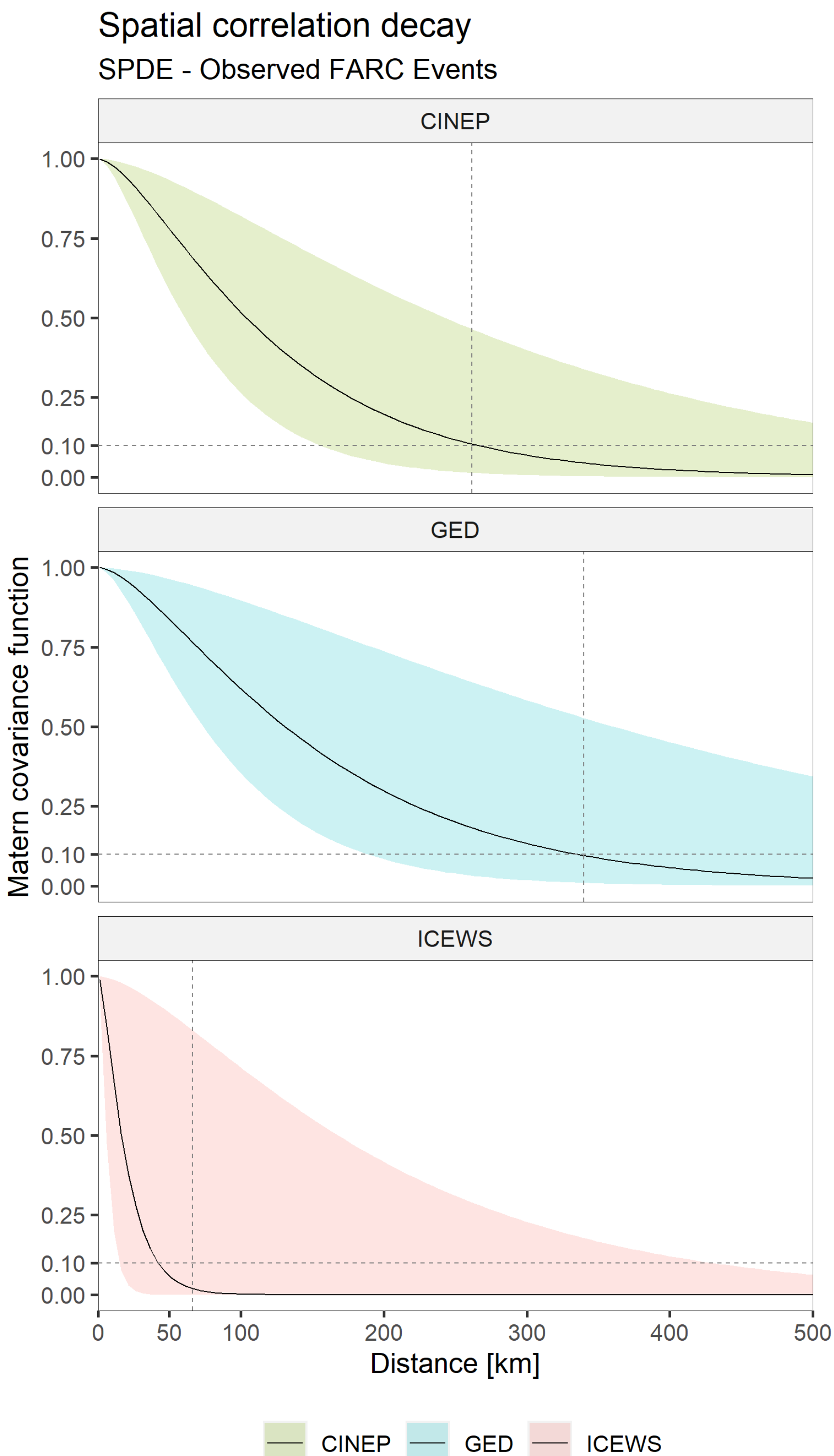


Figure A.9: SPDE: GMRF - Spatial Error Correlation

7 References

Althaus, S., B. Peyton, and D. Shalmon. 2018. "Spatial and temporal dynamics of Boko Haram activity across six event data generation pipelines: testing a new approach to event data validation." Paper presented at the APSA annual meeting, Boston, MA.

Althaus, S., J. Bajjalieh, J. F. Carter, B. Peyton, and D. A. Shalmon. 2017. "Cline Center Historical Phoenix Event Data. v.1.0.0." Distributed by the Cline Center for Advanced Social Research. http://www.clinecenter.illinois.edu/data/event/phoenix/.

Anselin, L. 2006. "Spatial econometrics." Chapter 26 in *Palgrave Handbook of Econometrics Volume 1 Econometric Theory* Basingstone: Palgrave MacMillan, pps. 901-969.

Raytheon BBN Technologies. 2015. "BBN ACCENT Event Coding Evaluation." Version: August 28, 2015. `https://dataverse.harvard.edu/dataverse/icews`.

Boshee, E., J. Lautenschalger, S. O'Brien, S. Shellman, J.Starz, and M. Ward. 2016. "ICEWS Coded Event Data." `http://dx.doi.org/10.7910/DVN/28075.HarvardDataverse`.

Calabrese, R. and J.A. Elkink. 2014. "Estimators of binary spatial autoregressive models: a Monte Carlo study." *Journal of Regional Science* 54(4): 664-687.

CINEP. 2008. "Marco Conceptual: Banco de Datos de Derechos Humanos y Violencia Política." Centro de Investigación y Educaión Popular.

Cook, S. J. and N. B. Weidmann 2019. "Lost in aggregation: improving event analysis with report level data." *American Journal of Political Science* 63(1): 250-264.

Croicu, M. and R. Sundberg. 2015. "UCDP GED Codebook version 4.0," Department of Peace and Conflict Research, Uppsala University.

Davenport, C. and P. Ball. 2002. "Views to a kill: Exploring the implications of source selection in the case of Guatemalan state terror, 1977-1995." *Journal of Conflict Resolution* 46(3): 427-450.

D'Ignazio, D. R. Bhargava, E. Zuckerman, and L. Beck (2014). "Cliff-clavin: Determining geographic focus for news." *NewsKDD: Data Science for News Publishing at KDD.*

Eck, K. 2012. "In data we trust? A comparison of UCDP GED and ACLED conflict events datasets." *Cooperation and Conflict* 47(1): 124-141.

Franzese, R.J., J.C. Hays, and S.J. Cook. 2016. "Spatial- and spatial temporal-autoregressive probit models of interdependent binary outcomes." *Political Science Research and Methods* 4(1): 151-173.

Halterman, A. 2017. "Mordecai: full text geoparsing and event geocoding." *Journal of Open Source Software* 2(9), 91.

Halterman, A. 2019. "Geolocating political events in text." recarXiv: 1905.12713v1 [cs.CL] 29 May 2019.

Hammond, J. and N. B. Weidmann. 2014. "Using machine coded event data for the micro level study of political violence." *Research and Politics* July-September: 1-8.

Hellmeier, S., N. B. Weidmann, and E. G. Rod (2018) "In the spotlight: analyzing sequential attention effects in protest reporting." *Political Communication* 35: 587-611.

ICEWS. 2015. "ICEWS Coded Event Data Readme" URL: `https://dataverse.harvard.edu/dataverse/icews`, Accessed on: 3/8/2021.

King, G. and W. Lowe. 2004. "An automated information extraction tool for international conflict data with performance as good as human coders." *International Organization* 57(3): 617-642.

Lautenschlager, J., J. Starz, and I. Warfield. 2017. "A statistical approach to the subnational geolocation of event data." In *Advances in Cross-Cultural Decision Making,* S. Schatz and M. Hoffman, Editor. Switzerland: Springer International Publishing.

Leetaru, K. and Philip Schrodt. 2013. "GDELT: Global data on events, language and tone, 1979-2002." Presented at the Annual Meeting of the International Studies Association, San Francisco, California.

Lockheed Martin. 2021. "ICEWS: iDATA." URL: `https://www.lockheedmartin.com/en-us/capabilities/research-labs/advanced-technology-labs/icews.html`, Accessed on: 3/8/2021.

Martinetti, D. and G. Geniaux. 2017. "Approximate likelihood estimation of spatial probit models." *Regional Science and Urban Economics* 64: 30-45.

Nardulli, P.F., A. Singh, M. Martin, D. Shalmon, B. Peyton, and S. Althaus (2019) "Social Political Event Data Set (SPEED): Liberia, Philippines, Sierra Leone (1979-2008) Codebook. v1.0.0. Champaign, Illinois. The Cline Center.

Norris, C., P. A. Schrodt, and J. Beieler. 2012. "PETRARCH-2: another event coding program." *The Journal of Open Source Software*, 9(2).

OEDA. 2016. "Real Time Event Data/Phoenix." Available at http://eventdata.utdallas.edu/.

Otto, S. 2013. "Coding one-sided violence from media reports." *Cooperation and Conflict*, 48(4).

PITF. 2009. "Political instability task force worldwide atrocities event data collection codebook version 1.0B2." Version pitf.world.19950101-20130930.csv.

Schrodt, P.A. 2015. "Comparison metrics for large scale political event data sets." Paper presented at the Text as Data Meeting. NY. New York University, October 16-17.

Schrodt, P. and D. Gerner. 1994. "Validity assessment of machine-coded event data set for the middle east, 1982-1992." *American Journal of Political Science* 38: 825-854.

Schrodt, P., D. Gerner, and O. Yilmaz. 2009. "Conflict and Mediation Event Observations (CAMEO): An Event Data Framework for a Post Cold War World." In International Conflict Mediation: New Approaches and Findings. J. Bercovitch and S. Gartner (ed.). Rouledge: New York.

Sundberg, R. and E. Melander. 2013. "Introducing the UCDP georeferenced event dataset." *Journal of Peace Research* 50(4): 523-532.

Wang, W. R. Kennedy, D. Lazer, and N. Ramakrishnan. 2016. "Growing Pains For Global Monitoring of Societal Events." *Science* 353(6307): 1502-1503

Ward, M.D. and K.S. Gleditsch. 2019. *Spatial Regression Models* Second edition. Sage.