# Mapping Violent Presence of Armed Actors in Colombia

Javier Osorio[a, *], Mohamed Mohamed [b], Viveca Pavon [a], Susan Brewer-Osorio[c]

[a] *School of Government and Public Policy, University of Arizona, josorio1@email.arizona.edu, vivecapavon@email.arizona.edu*
[b] *Master in Geographic Information Systems Technology, University of Arizona, engmoh888@email.arizona.edu*
[c] *Center for Latin American Studies, University of Arizona, susanosorio@email.arizona.edu*
* Corresponding author

**Abstract**: Heat maps and Early Warning Systems have traditionally contributed to identifying and managing risks associated with crime trends, health hazards, and natural disasters. However, their application to analyzing civil war dynamics still is at an early stage. To address this need, this research integrates Natural Language Processing (NLP) tools and Geographic Information Systems (GIS) to generate an interactive map of the violent presence of armed actors in the Colombian civil war between 1988 and 2017. The NLP component generates fine-grained geo-location data of armed actors' violent presence. The GIS component then uses the geo-referenced data to present dynamic clusters of four main types of actors: Government forces, Insurgent organizations, Paramilitary groups, and Criminal organizations. Each type of actors is further disaggregated into a multitude of specific armed organizations. Based on anomalies in the spatio-temporal trends identified in the data, we develop an EWS methodology to detect "emerging", "intense", and "critical" cases. This application contributes to the efforts of the academic and policy communities to understanding the spatio-temporal dynamics of political violence and promoting sustainable peace in civil war settings.

**Keywords:** armed actors, civil war, conflict, heat map, interactive map, Colombia

## 1. Introduction

The analytic power of Geographic Information Systems (GIS) assists in monitoring, preventing, and responding to destructive forces such as natural disasters, epidemics, and crime. However, their application remains underdeveloped for analyzing another deadly force: civil wars. This paper integrates Natural Language Processing technology (NLP) and GIS tools to generate an interactive online visualization of the violent presence of armed actors in the Colombian civil war between 1988 and 2017. By focusing on this country, the research helps to understand the dynamics of conflict in one of the world's oldest civil wars and contributes to ongoing peace building efforts. The online application renders spatio-temporal heat maps of different types of actors including government forces, insurgent organizations, paramilitary groups, and criminal organizations over time and across space. The web map also allows disaggregating these main actors into a variety of specific armed organizations. Inputs from the NLP and GIS components are then used to inform an Early Warning System (EWS) capable of detecting "emerging," "intense," and "critical" anomalies in armed actor activity. These alerts could be used to inform practitioners and researchers about specific security challenges in Colombia.

Even though there is a vast political science literature on the causes and dynamics of civil war, few studies offer rigorous and systematic conflict mapping and warning systems. The Armed Conflict Location and Event Data Project (ACLED) provides an interactive map and regular analyses of a variety of conflict scenarios in the Middle East and a (Armed Conflict Location & Event Data Project, 2018). A recent project of Uppsala University, ViEWS (Hegre et al., 2018), offers conflict warnings in Africa but does not include interactive mapping. Despite their contributions, these projects do not cover the Latin American region or the Colombian case. This research joins ACLED and ViEWS in their efforts to use GIS and EWS to analyze conflict dynamics. This study is also part of the recent expansion of EWS applications to a variety of academic and industry sector s (Scheffer et al., 2009, Basher, 2006, Grover-Kopec et al., 2005, Griffiths et al., 2009, International Strategy for Disaster Reduction, 2005).

A central contribution of this research is the integration of NLP, GIS, and EWS within the same framework. The NLP component enables the possibility of processing large amounts of text quickly and with minimal effort. The GIS component allows an intuitive visualization in a user-friendly interface. Finally, the EWS methodology provides concrete information of relevant cases to inform analysts in the academic and policy sectors.

The paper is organized in the following manner. The next section presents the conceptual constructs informing the measures of this study. The subsequent segment provides contextual information about the Colombian conflict. This is followed by a description of the NLP procedure implemented for geo-coding armed actors. Then, the paper presents the interactive online map that shows violent presence of armed actors and identifies major trends. Next, we outline the early warning methodology

and report results. The final section presents some conclusions.

## 2. Armed Actors

This application focuses on mapping armed actors in Colombia to identify temporal and spatial dynamics of conflict. The term *Armed Actor* refers to state and non-state armed entities willing and capable to exercise the organized use of violence against other armed actors or the civilian population in a specific territory in order to achieve political or economic goals (International Committee of the Red Cross, 2010, Art. 1, p. 84). For this research, the main types of armed actors include:

- Government forces
- Insurgent organizations
- Paramilitary groups
- Criminal organizations

The main actors are further disaggregated in a variety of specific organizations as indicated in Table 1. State armed groups refer to the formal institutions of the state in charge of providing internal and external security. Non-state armed actors refer to a broad range of armed organizations outside the direct control of government authorities or in opposition to it that are able to use force. Since the distinction between state and non-state armed actors is becoming conceptually and empirically blurred (Berti, 2016, Arjona, 2016, Arias, 2017), this project focuses on both state and non-state actors. Such approach helps to understand the interactions between different types of actors.

The data indicates the *Violent Presence* of armed actors at the municipality-year level. Given the nature of the information source, this data reflects the location of armed actors involved in violent incidents of both lethal (e.g. assassinations) or non-lethal nature (e.g. threats, displacement, etc.). However, it does not reflect cases in which armed actors are present in a territory but exercise no violence (Arjona, 2011). Such cases would correspond to areas of dominant or monopolistic control of armed actors in which the use of violence might not be unnecessary (Kalyvas, 2006). However, the information source might not report such cases, thus inhibiting the software form detecting non-violent presence of armed groups. In consequence, inferences drawn from this data should consider a narrow approach focused on the violent presence of armed actors.

## 3. The Colombian Conflict

Colombia's protracted armed conflict has endured for six decades. During that time, the nature and intensity of violence evolved in response to economic booms of licit and illicit commodities, and the emergence of new armed actors. The onset of the conflict is long rooted in rural inequality and political exclusion pervasive in this country. Historically, Colombian society was strongly divided between Conservative and Liberal party supporters. In 1948, a civil war broke out following the assassination of Jorge Eliecer Gaitan, a progressive presidential candidate of the Liberal party. Gaitan's death

provoked an urban riot that spread to the countryside as Liberal elites mobilized their peasant base into guerrilla bands (Braun, 2003). In 1954, the war known as *La Violencia* ended with a power-sharing pact between Conservative and Liberal elites that formally excluded other political parties from participating in government and failed to address social inequality (Dix, 1980).

| Type | Organization |
|---|---|
| Government | Generic mentions of government |
| | Army |
| | Airforce |
| | Navy |
| | Police |
| | Attorney's General Office |
| | Intelligence Agency |
| Insurgents | Generic mentions of insurgents |
| | Revolutionary Armed Forces of Colombia (FARC) |
| | National Liberation Army (ELN) |
| | Popular Liberation Army (EPL) |
| | People's Revolutionary Army (ERP) |
| | 9th of April Movement (M-19) |
| | Ernesto Rojas Commandos (CER) |
| | Simon Bolivar Guerrilla Coordinating Board (CGSB) |
| Paramilitaries | Generic mentions of paramilitaries |
| | United Self-defense forces of Colombia (AUC) |
| | Popular Revolutionary Anti-Terrorist Army of Colombia (ERPAC) |
| | United Gaitan Self-defense forces (AGU) |
| | Peasant Self-defense forces of Cordoba and Uraba (ACCU) |
| | Social Cleansing |
| | Independent paramilitaries |
| Criminals | Generic mentions of criminals (BACRIM) |
| | Los Urabenos |
| | Los Rastrojos |
| | Aguilas Negras |
| | Los Paisas |
| | Envigado Office |
| | Los 400 |
| | Independent criminal groups |

Table 1. Armed actor by type and organization

In the aftermath of *La Violencia*, multiple insurgent groups emerged to challenge the government. Lingering peasant militias established independent "peasant republics" that were attacked by the Colombian military in 1963. The remaining population formed the Revolutionary Armed Forces of Colombia (FARC) in 1964. That year, the National Liberation Army (ELN) formed, and in 1967 the Popular Liberation Army (EPL) joined the conflict (Offstien and Aristizabal, 2003). Last, the group 19th of April Movement (M-19), emerged in response to a fraudulent presidential election in 1970 (Boudon, 2001). In 1985, these four groups briefly joined forces as the Simón Bolívar Guerrilla Coordinating Board (CGSB), but otherwise operated autonomous from each other, in addition to multiple smaller insurgent groups such as the People's Revolutionary Army (ERP) and the Ernesto Rojas Commandos (CER). The M-19 and the EPL demobilized in 1991 and formed political parties,

while an EPL faction turned into a paramilitary group. After a long peace negotiation process, the FARC demobilized in 2016, thus leaving the ELN as the only active insurgency.

Early on, Colombian insurgent groups engaged in sporadic and mostly defensive operations against government forces. However, as in many other civil wars, natural resources played a role in intensifying the dynamics of conflict (Fearon, 2005, Lujala, 2009, Ross, 2006). In Colombia, illicit coca booms after 1978 and the discovery of major oil reserves in 1983 provided resources for insurgents to build their military capacity. The FARC extracted protection rents from narcotics traffickers and coca farmers, expanding into new territories where they carried out extortion and ransom kidnappings. During the 1990s, coca production increased in Colombia in tandem with the expansion of the FARC. By 2001, the FARC had about 21,000 armed combatants and was present in over 70\% of Colombia a (Norman, 2018). The ELN owes its expansion to the 1983 discovery of the Caño Limón oil field. This gave ELN the opportunity to extort oil contractors and kidnap their employees for ransom. Oil also attracted the national military, FARC, and paramilitaries to oil-rich areas (Dunning and Wirpsa, 2004).

During the 1980s and 1990s, right-wing paramilitary groups organized to confront the armed left with the support of state security forces and narcotics traffickers. In 1981, the Medellin Cartel endorsed the paramilitary organization, Death to Kidnappers (MAS) as a rejoinder to widespread insurgent extortion against wealthy landowners, including prominent drug capos. After that, paramilitary groups proliferated with support from drug traffickers, an alliance that contributed to the intensification of violence. Between 1994 and 1997, local paramilitary groups gathered under the United Self-Defense Forces of Colombia (AUC). The AUC displaced, massacred, and selectively assassinated non-combatants as a strategy for breaking civilian support for the insurgency (Romero, 2003).

In 2005, President Alvaro Uribe (2002-2010) negotiated the disarmament of about 31,000 paramilitaries and launched a successful military campaign against insurgents leading to significantly improvements in security conditions. Nevertheless, about 3,000 paramilitaries remained active following the demobilization, and former AUC mid-level commanders formed a new generation of paramilitary groups called Emerging Criminal Bands (BACRIM). BACRIM groups are criminal organizations that lack political goals and engage in narcotics trafficking and other black- market activities (Nussio, 2018). In 2016, another significant peace agreement was signed between President Juan Manuel Santos (2010-2018) and the FARC, ending six decades of warfare and leading to the successful reintegration of about 7,000 ex-FARC combatants. However, some dissident factions of the FARC remain active.

## 4. Natural Language Processing for Identifying Armed Actors

### 4.1 Information Source

The information used in this study comes from *Noche y Niebla*, a publication created by CINEP comprising a collection of human rights violations narratives (Centro de Investigacion y Educación Popular, 2018). CINEP has been gathering information related to human rights incidents taking place within the context of the Colombian conflict since 1987. For more than three decades, CINEP has been systematically reviewing, gathering, systematizing, and categorizing dispersed information from dozens of national and local newspapers on a daily manner, as well as from verbal and written testimonies issued by victims of human rights violations, their relatives, community leaders, other civil society organizations, and other relevant stakeholders. Being closely affiliated with the Catholic church, CINEP counts with a vast network of information nodes throughout Colombia, which provides a broad territorial network reaching even the most remote areas.

CINEP follows strict ethical and methodological standards for building the collection of narratives comprised in *Noche y Niebla*. In this way, CINEP reviews multiple sources of documentation about each incident, and only publishes verified events based on cross-referenced sources. Three decades of producing high quality data contributed to CINEP's prestige as a key organization advancing the respect for human rights, increasing accountability and reparation, documenting historic memory, and promoting peace. In addition, CINEP has long provided data for academic research analyzing the Colombian conflict (Holmes et al., 2006, Franz, 2016, Bagozzi et al., 2018).

Based on *Noche y Niebla* 's narratives, the corpus used in this research comprises a collection of 45,375 human rights violations reports between 1988 and 2017 written in Spanish. After a simple pre-processing procedure of the text, each line of the corpus contains four key elements: (i) the date of occurrence in YYYYMMDD format; (ii) a unique identifier for each narrative; (iii) a brief description of the incident; (iv) and the location identified in the narrative. The following narrative present an example of information contained in the corpus:

```
Original in Spanish:
20060812 20060812_NYN_P1_P30376 | Presuntos
insurgentes del ELN asesinaron al joven
Alfredo Torres, el hecho se presento en la
vereda Patanguejo, corregimiento de
Cajamarca.
Segun la fuente, hace pocos dias la victima
habia terminado de prestar servicio militar,
Cauca, Mercaderes, Cajamarca.
```

```
English Translation (by the authors):
20060812 20060812_NYN_P1_P30376 | Presumed
insurgents of the ELN killed young
Alfredo Torres.  The event took place in the
Patanguejo community, in Cajamarca district.
According to the source, the victim finished
the military service a few days ago,
Cauca, Mercaderes, Cajamarca.
```

## 4.2 Computerized Textual Annotation

A central component of this study rests on the use of NLP to geo-reference the presence of armed actors by processing large volumes of text in a systematic manner. To generate the database of violent presence of armed actors in Colombia, the project relied on Eventus ID, a software for supervised event coding from text written in Spanish (Osorio and Reyes, 2016). Eventus ID belongs to a family of coders that rely on sparse parsing to process the text such as KEDS, TABARI, and Petrarch (Schrodt et al., 1994, Schrodt, 2009, Schrodt et al., 2014). As such, Eventus ID uses a set of dictionaries of actors and locations to processes the text and identify the actors mentioned in the corpus, as well as their related locations. Figure 1 describes the general protocol of geo-locating actors using Eventus ID. The algorithm implements the following steps:

1. The software starts reading the corpus.
2. Eventus ID uses the actors dictionary as search criteria to look for a match in the corpus.
3. If there is an actor match, Eventus ID stores the matching text and identifies the corresponding numeric code.
4. For each corpus line containing an actor match, the program uses the location dictionaries to look for a toponym.
5. If there is a match with the locations dictionaries, the program then verifies the location filters to identify whether or not this location name is a false positive.
6. If there is no match in the location filters, Eventus ID saves the coded actor and locations in the output.
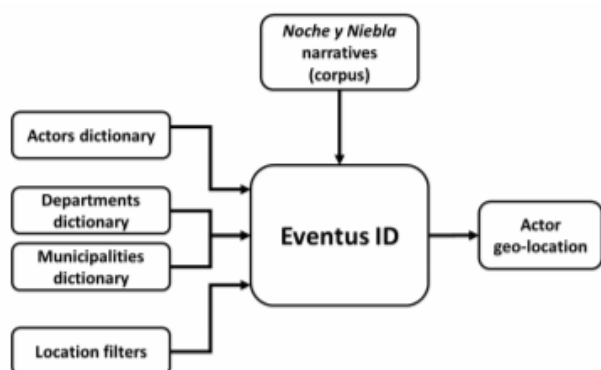


Figure 1. Actor geo-location protocol using Eventus ID

To develop an encompassing as possible dictionary of actors, researchers relied on Named Entity Recognition (NER) tools to identify previously unknown actors (The Stanford Natural Language Processing Group, 2014). The resulting dictionary comprises a list of 7,560 actor names including organization names, their acronyms, branches or sub-groups, and main leaders. Actors are categorized in four main types of armed groups: Government, Insurgents, Paramilitaries, and Criminals as discussed in Section 2. To identify actors, Eventus ID uses the actors dictionary as search criteria to detect actors in the corpus. For example, if the actors dictionary includes the name "FARC", the system looks for it in the text. Once it finds a match for "FARC" in the corpus, the system records the

date of occurrence, extracts the textual match, and assigns the corresponding numeric code for this actor. These annotations are stored in a coding output database.

To identify the locations, Eventus ID relies on two dictionaries, one for departments (equivalent to states) and another one for municipalities (similar to counties). In this way, the software geo-references actors at the sub-national level. A common challenge in event coding relates to correctly identifying the place of occurrence of an event (Lee et al., 2016, Chalabi, 2014). To address the geographic disambiguation challenge, Eventus ID uses a locations filter to eliminate matches that might look as toponyms, but do not actually refer to locations. For example, the locations filter includes "Universidad del Magdalena" to prevent the program from erroneously identify this name as the "Magdalena" department.

For each data point, the database includes the longitude and latitude coordinates of the municipality center. This allows projecting the coded data into a map. Due to limitations in the information source, the projection does not provide the exact location of an event as based on a specific street address or a geo-tag marked in the field. In this case, the projection is just an approximation of the location of each incident based on the municipality's coordinates.

After coding actors and locations, the methodology considers a post-coding process for cleaning the data in Stata. This protocol includes a deduplication routine to eliminate multiple matches of the same actor in the same municipality day. In this way, the analysis avoids artificial inflation and considers a conservative approach for including in the cleaned database one single actor per organization, for each municipality-day. The protocol also includes a process for location deduplication and cleaning to reduce concerns of false positives. The final data set includes the type of actor, the specific organization recorded, the date of occurrence, as well as the department and municipality.

## 5. Mapping Armed Actors

### 5.1 Interactive Web Map Application

To visualize the data points, we developed a geographic representation framework with ArcGIS Pro v2.2 using the WGS84 Web Mercator (Auxiliary Sphere) coordinates system. After developing the map, we used ArcGIS Online to deploy the interactive application in the web, which can be accessed at www.colombiaarmedactors.org.

The visualization presents heat maps of violent presence of armed actors as defined in Section 2. A heat map is a GIS analysis technique that provides a graphic representation of high-density occurrence or clustering of a phenomenon. In contrast to hot-spot analysis tools that rely on statistical estimates of spatial auto-correlation, heat maps provide a more flexible and intuitive way of clustering the observations and assigning a color gradient to indicate high activity concentration. The raw information used for rendering the heat map comes from the data points coded in the NLP procedure discussed in Section 4.2.

To build a heat map, ArcGIS uses a Gaussian distribution to assess the strength of the influence of each point over a specific area as determined by a radius. The program renders a high color intensity to areas concentrating high point density. In contrast to a fixed raster framework, ArcGIS Online dynamically adjusts the radius for rendering the heat map based on the user's projection scale. In this way, the heat map provides a more aggregate or disaggregate clustering as users zoom out or into the map.

The web application includes a variety of functions to allow users interact with the map in different ways. The system is able to display heat maps over time with yearly data between 1988 and 2017. The time slider allows pausing or adjusting the time lapse speed, as well as selecting the range of years being displayed. Users can visualize clusters of all the actors in the map or explore particular spatio-temporal behaviours by selecting specific actors by type or organization as indicated in Table 1.

## 5.2 Data Visualization

Heat maps of violent presence of armed actors in Colombia are useful in understanding the spatio-temporal dynamics of multi-actor conflict. The interactive online map allows users to visualize trends of armed actor's violent presence. As a proxy to the projections rendered in the web application, Figure 2 presents a series of annual maps displaying the concentration of all armed actors at five-year intervals from 1990 to 2015.

As Panels (a) and (b) in Figure 2 show, the violent presence of armed actors was confined to a reduced set of locations in 1990 and 1995. During this period, the FARC and ELN were active in remote regions, but mostly unchallenged by paramilitary groups. Meanwhile, President César Gaviria (1990-1994) entered peace talks with multiple insurgent groups resulting in the demobilization of the EPL and M-19, and other smaller guerrilla organizations. Panel (c) shows a dramatic increase in intensity and scope of armed actor activity in 2000. This escalation corresponds to the collapse of peace talks between President Andres Pastrana (1998-2002) and the FARC, resulting in the re-militarization of counterinsurgency policy. The AUC paramilitary group also formed in 1997 and launched an aggressive expansion into insurgent-controlled areas. As Panel (d) shows, the intensity of armed actor activity slightly decreased in 2005, yet the extension of the conflict kept affecting a substantial part of the country. The decline is attributable to the demobilization of paramilitaries and President Alvaro Uribe's military campaign against insurgents, which reduced their numbers and territorial reach.

Panel (e) shows that in 2010 there was a continued, slight downward trend of de-escalation and re-concentration of armed actor activity. This reflects the continued insurgent retraction in a context of increased state control. Following up to 2015, there were marginal oscillations in the intensity and scope of armed actor activity as Panel (f) shows. These fluctuations relate to the expansion of

BACRIMs, and the cease-fire during the FARC-government peace talks.
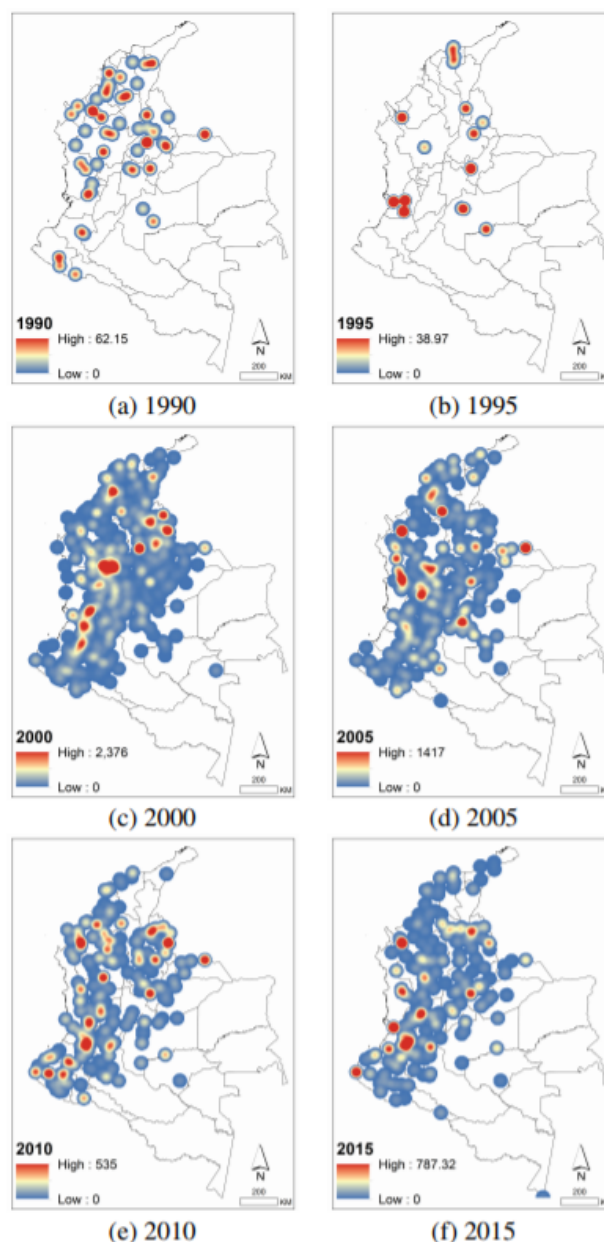


Figure 2. Violent presence of all armed actors

In addition to presenting temporal variations in the dynamics of armed actor violent presence, the online application allows visualizing the concentration of activity by type of actor. As a proxy to the heat maps by actor type available in the online interface, Figure 3 presents heat maps of government forces, insurgent organizations, paramilitary groups, and criminal organizations with data aggregated between 1988 and 2017. Panel (a) shows the government's capacity to deploy a violent presence with high degree of intensity in large parts of the territory. Panel (b) also shows the challenge and resistance that multiple insurgent organizations pose to the state in vast areas of the country. Panel (c) also shows the broad territorial reach that paramilitary organizations have despite their late entry into the conflict. Finally, panel (d) reveals that the

6 of 9

emerging criminal bands have less intense activity compared to the other actors, yet they their presence is also ubiquitous.
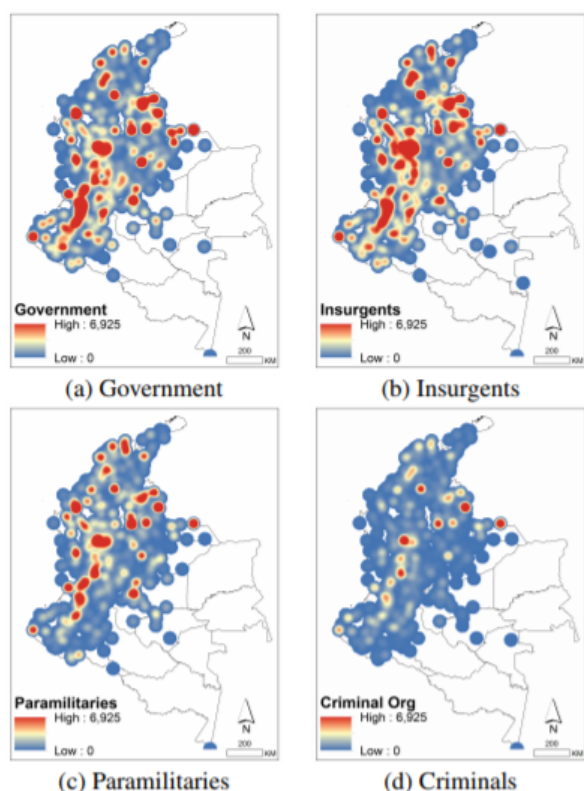


Figure 3. Violent presence per type of actor (1988-2017)

Another feature of the web application is the possibility of selecting specific organizations from a menu of 30 armed groups to visualize their violent presence (see Table 1). As a proxy to this feature, Figure 4 presents the concentration areas of violent presence of two Colombian insurgent organizations: the FARC and the EPL. This example reveals distinct areas of violent guerrilla presence. Panel (a) show that the FARC, a larger and stronger organization, concentrates its violent activity along the West. In contrast, Panel (b) shows that the EPL, a smaller guerrilla group, concentrated its violent activity in the North of the country.
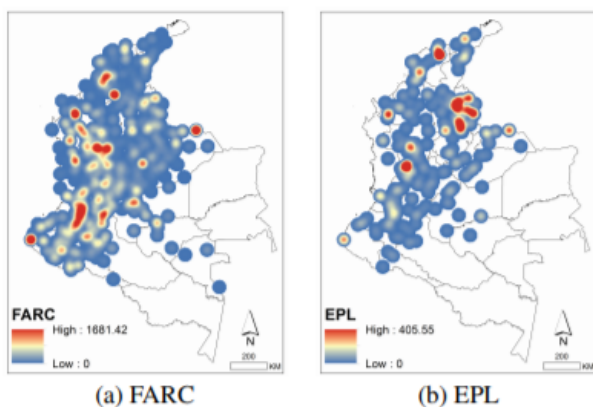


Figure 4. FARC and EPL insurgent groups (1988-2017)

## 6. Early Warning System

Based on the temporal trends of violent presence of armed actors in each municipality, we developed an Early Warning System (EWS) to detect spikes in armed actor activity. The EWS decision rules consider three levels of alert based on the extent of deviation above the moving average of previous years as depicted in Table 2. For each municipality $i$ in year $t$, the decision rule assigns an alert level ($\alpha=\{emerging, intense, critical\}$) by considering a minimum number of incidents ($\beta_{i,\tau}=\{2,4,6\}$) and if the number of incidents of armed actor violent presence ($n_{i,t}$) is greater than $x$ standard deviations ($\sigma_{i,\mu\alpha}$) above the moving average in the four previous years ($\mu_{i,ma}$), where $x_{i,ma}=\{2,4,6\}$. This decision rule allows detecting *emerging* municipalities that reach two standard deviations above the moving average, *intense* cases that spike up to four standard deviations above the moving average, and *critical* cases that have six or more standard deviations above the moving average.

| Alert level | Baseline $\beta_{i,t}$ | Decision rules |
|---|---|---|
| Emerging | 2 | $n_{i,t} > (\mu_{i,ma} + (2 \times \sigma_{i,ma}))$ |
| Intense | 4 | $n_{i,t} > (\mu_{i,ma} + (4 \times \sigma_{i,ma}))$ |
| Critical | 6 | $n_{i,t} > (\mu_{i,ma} + (6 \times \sigma_{i,ma}))$ |

Table 2. EWS decision rule for alert levels

Applying this EWS decision rules allow issuing 1,307 alerts, which correspond to 3.8\% of the total municipality-years considered in this analysis. The EWS detected 1,273 "emerging" cases of increasing armed actor activity. The system also identified 473 cases as "intense" anomalies in their respective trajectories. Finally, the warning system detected 561 "critical" municipalities. Figure 5 presents the frequency of alert cases between 1992 and 2017. The EWS shows a remarkable spike of alerts between 1998 and 2000. After this period, the number of alerts fluctuated with moderate oscillations below 100 cases per year.
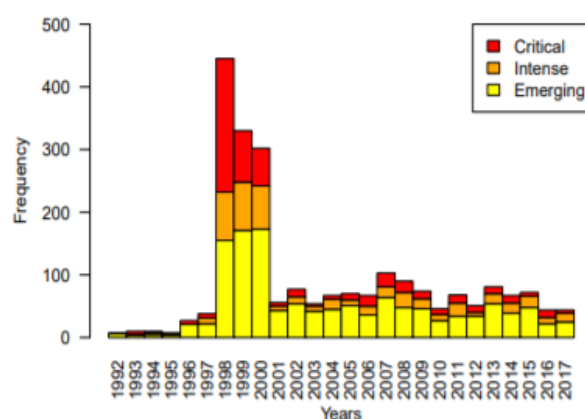


Figure 5. Early Warning System alerts by level

For 2017, the EWS issued "critical" alerts for five municipalities: Riosucio, Chocó, Carmen del Darien, Chocó; Mutatá, Antioquia; Purace, Cauca; and Santa Bárbara, Nariño. Fine grained analysis as depicted in Table 3 reveals the nature of the threat and provides

actionable information for different types of decision makers. The high levels of armed actor activity in Riosucio should call the attention of the international community. The data reveals that this municipality is intensely contested between government forces, insurgents, and paramilitaries. Such level of conflict might pose a threat to the Colombian peace process due to its proximity to the FARC-deconcentration camp of Brisas, also known as La Florida. The data should also call the attention of government officials to curtail the high levels of paramilitary activity in Carmen del Darien and Mutatá. Human Rights organizations should also pay close attention to government forces in Purace due to their disproportionate violent presence. Finally, government forces should reinforce their presence in Santa Bárbara where the system detects insurgent intense activity.

| Municipality | Gov. | Ins. | Para. | Crim. | Total |
|---|---|---|---|---|---|
| Riosucio | 27 | 20 | 35 | 1 | 83 |
| C. del Darién | 12 | 1 | 10 | 1 | 24 |
| Mutatá | 12 | 0 | 7 | 0 | 19 |
| Puracé | 10 | 1 | 0 | 0 | 11 |
| Santa Bárbara | 3 | 6 | 2 | 0 | 11 |

Table 3. EWS alerts of "critical" cases in 2017

## 7. Conclusion

The rapidly changing dynamics of political violence require powerful tools to represent and analyze complex data in an easy-to-use manner that can effectively inform researchers and policy makers. This need is particularly pressing for sub-national level analyses of violence in conflicts involving multiple warring parties. To address such need, this research integrates Natural Language Processing technology with Geographic Information Systems to inform an Early Warning System in the context of the Colombian conflict.

The NLP component relies on Eventus ID, a computerized event coding protocol, to process a large collection of human rights violations narratives developed by CINEP. This software allows analysts to identify the locations of the violent presence of multiple armed actors in Colombia. The geo-coded data is then used as an input for the GIS component of this research. The GIS application consists of an easy-to-use interactive web interface that generates dynamic heat maps of armed actor violent presence. The web application is freely available at www.colombiaarmedactors.org. The interactive map allows users to visualize the dynamic concentration of armed actors over time and across space. The system also disaggregates the behaviour of four main types or armed actors (government, insurgents, paramilitaries, y criminals) and a variety of individual armed organizations. The data is then used to inform an Early Warning System capable of detecting relevant anomalies that substantially depart from the temporal trends of each municipality. The EWS categorizes relevant cases as "emerging", "intense", or "critical" and provides detailed information to inform policy and research analysts.

For future developments, this EWS has the potential of delivering near-real-time data to identify emerging threats in Colombia. The integration of NLP and GIS provide a powerful platform to escalate this EWS for quickly processing large volumes of data and providing valuable and timely recommendations. The key requirement to enable this potential is to rely on a daily flow of information. Such input could come from newspaper publications or field reports of government or international agencies. In this way, the system can effortlessly update the continuous stream of information to better inform researchers and practitioners. In addition, the analytic power of this EWS can be extended to forecasting models to improve the ability to identify possible scenarios and anticipate emerging security challenges (Brandt et al., 2011, Ward et al., n.d.).

## 8. Acknowledgements

## 9. References

Arias, D. E., 2017. Criminal Enterprises and Governance in Latin America and the Caribbean. Cambridge University Press, New York.

Arjona, A., 2016. Rebelocracy: Social Order in the Colombian Civil War. Cambridge University Press, New York.

Arjona, A. M., 2011. Presencia vs. violencia: problemas de medicion de la presencia de actores armados en Colombia. Available online at https://tinyurl.com/y776k3vh, (accessed December, 2018).

Armed Conflict Location & Event Data Project, 2018. Armed conflict location & event data project (acled). Available online at https://www.acleddata.com/, (accessed December, 2018).

Bagozzi, B. E., Brandt, P. T., Freeman, J. R., Holmes, J. S., Kim, A., Palao Mendizabal, A. and Potz-Nielsen, C., 2018. The Prevalence and Severity of Underreporting Bias in Machine and Human-Coded Data. Political Science Research and Methods pp. 1–9. Available online at https://tinyurl.com/y8o3oshj, (accessed December, 2018).

Basher, R., 2006. Global early warning systems for natural hazards: systematic and people-centered. Philosophical Transactions of The Royal Society. Available online at https://www.unisdr.org/2006/ppew/info-eesources/docs/RSTA20061819p.pdf, (accessed December, 2018).

Berti, B., 2016. What's in a name? Re-conceptualizing non-state armed groups in the Middle East. Nature pp. 1–8. Available online at https://www.nature.com/articles/palcomms201689.pdf, (accessed December, 2018).

Boudon, L., 2001. Colombia's M-19 democratic alliance: A case study in New-Party self-destruction. Latin American Perspectives 28(1), pp. 73–92.

Brandt, P., Freeman, J. R. and Schrodt, P. A., 2011. Real Time, Time Series Forecasting of Inter- and Intra-State Political Conflict. Conflict Management and Peace Science 28(1), pp. 41–64.

Braun, H., 2003. The Assassination of Gaitan: Public Life and Urban Violence in Colombia. University of Wisconsin Press.

Centro de Investigacion y Educacion Popular, 2018. Banco de datos de derechos humanos y violencia Política.Available online athttps://base.nocheyniebla. org/consulta_web.php, (accessed December, 2018).

Chalabi, M., 2014. Mapping Kidnappings in Nigeria. Available on line at https://fivethirtyeight.com/features/ mapping-kidnappings-in-nigeria/, (accessed December, 2018).

Dix, R. H., 1980. Consociational democracy: The case of colombia. Comparative Politics 12(3), pp. 303–321.

Dunning, T. and Wirpsa, L., 2004. Oil and the political economy of conflict in Colombia and beyond: A linkages approach. Geopolitics 9(1), pp. 81–108.

Fearon, J. D., 2005. Primary commodity exports and civil war. Journal of Conflict Resolution 49(4), pp. 483–507.

Franz, T., 2016. Plan Colombia: illegal drugs, economic development and counterinsurgency a political economy analysis of Colombia's failed war. development Policy Review 34(4), pp. 563–591.

Griffiths, P., Vingoe, L., Mounteney, J. and Hartnoll, R., 2009. Drug Information Systems, Early Warning, and New Drug Trends: Can Drug Monitoring Systems Become More Sensitive to Emerging Trends in Drug Consumption? Substance Use and Misuse. Available online at https://www.tandfonline.com/doi/abs/10.3109/1082608 0009148423, (accessed December,2018).

Grover-Kopec, E., Kawano, M., Klaver, R., Blumenthal, B., Ceccato, P. and Connor, S., 2005. An online operational rainfall-monitoring resource for epidemic malaria early warning systems in africa. Malaria Journal. Available online at https://malariajournal.biomedcentral.com/articles/10.11 86/1475-2875-4-6, (accessed December, 2018).

Hegre, H., Croicu, M., Eck, K. and Hogbladh, S., 2018. Introducing the UCDP-Candidate Events Dataset and the ViEWS Outcomes dataset. Available online at https://www.pcr.uu.se/digitalAssets/653/c_653796-l_1-k_ucdp_candidate_views_outcomes.pdf, (accessed December, 2018).

Holmes, J. S., Amin Gutierrez de Pineres, S. and Curtin, K. M., 2006. Drugs, Violence, and Development in Colombia: A Department-Level Analysis. Latin American Politics & Society 48(3), pp. 157–184.

International Committee of the Red Cross, 2010. Protocols Additional to the Geneva Conventions of 12 August 1949. Technical report. Available online at https://www.icrc.org/en/doc/assets/files/other/icrc_002_ 0321.pdf, (accessed December, 2018).

International Strategy for Disaster Reduction, 2005. Platform for the promotion of early warning. Available online at https://www.unisdr.org/2006/ppew/whats-ew/basics-ew.htm, (accessed December, 2018).

Kalyvas, S. N., 2006. The Logic of Violence in Civil War. Cambridge University Press, Cambridge, MA.

Lee, S., Liu, H. and Ward, M., 2016. Lost in Space: Geolocation in Event Data. Available online at https://arxiv.org/abs/1611.04837, (accessed December, 2018).

Lujala, P., 2009. Deadly Combat over Natural Resources. Gems, Petroleum, Drugs, and the Severity of Armed Civil Conflict. Journal of Conflict Resolution 53(1), pp. 50–71.

Norman, S. V., 2018. Narcotization as Security Dilemma: The FARC and Drug Trade in Colombia. Studies in Conflict and Terrorism 41(8), pp. 638–659.

Nussio, E., 2018. Ex-Combatants and violence in Colombia: are yesterday's villains today's principal threat? Third World Thematics: A TWQ Journal 3(1), pp. 135–152.

Offstien, N. and Aristizabal, C., 2003. An Historical Review and Analysis of Colombian Guerrilla Movements: FARC, ELN, and EPL. Desarrollo Y Sociedad 52, pp. 99–142.

Osorio, J. and Reyes, A., 2016. Supervised Event Coding From Text Written in Spanish: Introducing Eventus ID. Social Science Computer Review. Available online at https://journals.sagepub.com/doi/abs/10.1177/08944393 15625475, (accessed December, 2018).

Romero, M., 2003. Paramilitares y autodefensas, 1982-2003.

Ross, M., 2006. A Closer Look At Oil, Diamonds, and Civil War. Annual Review of Political Science 9(1), pp. 265–300.

Scheffer, M., Bascompte, J., Brock, W., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E., Rietkerk, M. and Sugihara, G., 2009. Early-warning signals for critical transitions. Nature. Available online athttp://www.uvm.edu/pdodds/files/papers/others/2009/ scheffer2009a.pdf, (accessed December, 2018).

Schrodt, P. A., 2009. TABARI. Textual Analysis by Augmented Replacement Instructions. Available online at http://eventdata.parusanalytics.com/software.dir/ tabari.html,(accessed December, 2018).

Schrodt, P. A., Beieler, J. and Idris, M., 2014. Three's a Charm?: Open Event Data Coding with EL:DIABLO, PETRARCH, and the Open Event Data Alliance. In: International Studies Association, Toronto. Available online at http://parusanalytics.com/eventdata/papers.

dir/Schrodt-Beieler-Idris-ISA14.pdf,(accessed December, 2018).

Schrodt, P. A., Davis, S. G. and Weddle, J. L., 1994. Political Science: KEDS - A program for the machine coding of event data. Social Science Computer Review 12(4), pp. 561–587.

The Stanford Natural Language Processing Group, 2014. Stanford Named Entity Recognizer. Available online at http://nlp.stanford.edu/software/CRF-NER.shtml, (accessed December, 2018).

Ward, M. D., Metternich, N. W., Dorff, C. L., Gallop, M., Hollenbach, F. M., Schultz, A. and Weschle, S., n.d. Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction. International Studies Review (4), pp. 473–490.