

Spatial and Temporal Patterns of Conflict Events in Africa: An Exploratory Analysis

Menelisi Maswanganye



**A research report submitted to the Faculty of Science, University
of the Witwatersrand, Johannesburg, in partial fulfilment of the
requirements for the degree of Master of Science**

DECLARATION

I declare that this research report is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

(Signature of candidate)

_____ day of _____ 20_____

ABSTRACT

Conflict events across the world have usually been studied from a qualitative perspective. The current abundance of data pertaining to conflicts now allows for a more statistical approach. This research report examined the spatial distribution of three conflict types (battles, riots and protests, and violence against civilians) throughout mainland Africa. It also considered the evolution of the observed spatial patterns through time by comparing the patterns before 2007 with those from 2007 and onwards. The results show that although Somalia and Sudan are unusually afflicted with conflict, this is only statistically discernible at the more granular areal levels. It was also found that when conflict variables are dichotomised, loss of information occurs, but is somewhat mitigated by considering the analysis at more granular areal levels. Finally, although the number of conflicts has increased, there were mostly no significant differences between spatial patterns before and after 2007.

This research project is dedicated to both of my parents, Eddie and Bonakele
Maswanganye, who encouraged my curiosity about all things African

ACKNOWLEDGMENT

I would like to thank both of my parents, for their unwavering love and support; my supervisor, Sulaiman Salau, for his technical guidance, wisdom and compassion; and the School of Statistics and Actuarial Science – for providing me with the technical foundation to pursue intellectual interests with conviction and analytical rigour.

TABLE OF CONTENTS

LIST OF FIGURES	VII
LIST OF TABLES	VIII
ACRONYMS AND ABBREVIATIONS.....	IX
1 INTRODUCTION.....	1
1.1 Background	1
1.2 Study Objectives	3
1.3 Structure of the report	3
2 LITERATURE REVIEW.....	4
2.1 General overview	4
3 METHODOLOGY.....	9
3.1 Spatial Data Types	9
3.2 Measures of Spatial Autocorrelation.....	9
3.2.1 Global Measures of Spatial Autocorrelation	9
3.2.2 The global Moran's <i>I</i>	10
3.2.3 Join-count statistics (for binary data)	13
3.3 Local measures of Spatial Autocorrelation	14
3.3.1 Local Moran's <i>I</i>	14
3.3.2 Local Join-Count	15
3.4 Spatial Diffusion Analysis	16
4 DATA ANALYSIS AND DISCUSSION.....	19
4.1 Study Area and Data Description.....	19
4.2 Data Analysis	20
4.3 Evaluation	22
4.4 Results.....	23
4.5 Spatial Analysis: Spatial Autocorrelation – Counts.....	24
4.5.1 Summary of Battles from 1997 to 2017	24
4.5.2 Chloropleth Maps for Battles from 1997 to 2017.....	25
4.5.3 Battles from 1997 to 2017- Global Moran's <i>I</i>	27
4.5.4 Battles from 1997 to 2017- LISA.....	29
4.5.5 Summary of Battles from 1997 to 2006 and from 2007 to 2017.....	32
4.5.6 Chloropleth Maps for Battles from 1997 to 2006 and from 2007 to 2017	32
4.5.7 Battles from 1997 to 2006 and from 2007 to 2017- Global Moran's <i>I</i>	34

4.5.8	Battles 1997 to 2006 & 2007 to 2017- LISA.....	36
4.5.9	Battles – Diffusion.....	39
4.5.10	Summary of Riots and Protests from 1997 to 2017.....	40
4.5.11	Chloropleth maps for Riots and Protests from 1997 to 2017	40
4.5.12	Riots and Protests from 1997 to 2017 - Global Moran's I	42
4.5.13	Riots and Protests from 1997 to 2017 - LISA	44
4.5.14	Summary of Riot and Protests from 1997 to 2006 and from 2007 to 2017...	47
4.5.15	Chloropleth Maps for Riots and Protests from 1997 to 2006 and from 2007 to 2017	48
4.5.16	Riots and Protests from 1997 to 2006 and from 2007 to 2017- Global Moran's I.....	50
4.5.17	Riots and Protests from 1997 to 2006 and from 2007 to 2017- LISA.....	51
4.5.18	Riots and Protests – Diffusion	54
4.5.19	Summary of Violence against Civilians from 1997 to 2017	55
4.5.20	Chloropleth Maps for Violence against Civilians from 1997 to 2017.....	55
4.5.21	Violence against Civilians from 1997 to 2017- Global Moran's I	56
4.5.22	Violence against Civilians from 1997 to 2017- LISA	57
4.5.23	Summary of Violence against Civilians from 1997 to 2006 and from 2007 to 2017	58
4.5.24	Chloropleth Maps for Violence against Civilians from 1997 to 2006 and from 2007 to 2017	59
4.5.25	Violence against Civilians from 1997 to 2006 and from 2007 to 2017 Global Moran's I	60
4.5.26	Violence against Civilians from 1997 to 2006 and from 2007 to 2017- LISA	61
4.5.27	Violence Against Civilians – Diffusion.....	63
4.6	Spatial Autocorrelation – Binary Data.....	64
4.6.1	Battles (Join-Counts)	65
4.6.2	Riots and Protests (Join-Counts)	66
	Violence against Civilians (Join-Counts).....	67
5	Summary, Conclusions and Recommendation.....	68
5.1	Discussion	68
5.2	Limitations	70
5.3	Conclusion and Further Work	70
	REFERENCES	72

LIST OF FIGURES

Figure 1: Different definitions of contiguity. The Queen configuration (Q) is on the left, Rook (R) in the middle and Bishop (B) on the right	12
Figure 2 Diagram of Moran Scatterplot.....	15
Figure 3 Depiction of Contagious and Hierarchical Diffusion Patterns	16
Figure 4 Map of mainland Africa	19
Figure 5 Diagram of analysis plan	21
Figure 6 – Percentile maps of battles at admin level 0, 1 and 2	26
Figure 7 Moran scatterplots for battles at admin level 0, 1 and 2.....	28
Figure 8 LISA maps of battles at admin levels 0, 1 and 2	30
Figure 9 Percentile maps of battles, before and after 2007, at admin level 0, 1 and 2	33
Figure 10 Moran scatterplots, before and after 2007, for admin level 0, 1 and 2.....	35
Figure 11 LISA maps, before and after 2007, for admin level 0, 1 and 2	37
Figure 12 Percentile maps of riots and protests for admin levels 0, 1 and 2	41
Figure 13 Moran scatterplots for riots and protests for admin level 0, 1 and 2.....	43
Figure 14 LISA maps of riots and protests for admin level 0, 1 and 2.....	45
Figure 15 Percentile maps of riots and protests, before and after 2007, for admin level 0, 1 and 2	49
Figure 16 Moran scatterplots of riots and protests, before and after 2007, for admin level 1 and 2.....	51
Figure 17 LISA maps of riots and protests, before and after 2007, for admin levels 1 and 2.....	52
Figure 18 Percentile maps of violence against civilians at admin level 0, 1 and 2 ...	55
Figure 19 Moran scatterplots of violence against civilians at admin levels 1 and 2 .	56
Figure 20 LISA maps for violence against civilians for admin levels 1 and 2.....	57
Figure 21 Percentile maps for violence against civilians, before and after 2007, for admin level 0, 1 and 2.....	59
Figure 22 Moran scatterplots of violence against civilians, before and after 2007, at admin level 1 and 2.....	60
Figure 23 LISA maps of violence against civilians, before and after 2007, at admin level 1 and 2.....	61
Figure 24 Significance maps of battles (binary) for admin level 1 (top row) and 2 (bottom row)	65
Figure 25 Significance maps of riots and protests (binary) for admin level 1 (top row) and 2 (bottom row)	66
Figure 26 Significance maps of violence against civilians (binary) for admin level 1 (top row) and 2 (bottom row)	67

LIST OF TABLES

Table 1: Diffusion framework	17
Table 2: Naming convention of variables	23
Table 3: Order of Analysis for count of events	24
Table 4: Summary statistics of battles	24
Table 5: Global Moran's I for battles	27
Table 6: Countries with significant LISA results	31
Table 7: Summary statistics of battles, for period 1 and 2	32
Table 8: Global Moran's I for battles, for period 1 and 2.....	34
Table 9: Countries where LISA for battles were significant, for period 1 and 2.....	38
Table 10: Significant diffusion patterns for battles	39
Table 11: Summary statistics of riots and protests	40
Table 12: Global Moran's I for riots and protests	42
Table 13: Countries where LISA for riots and protest were significant, for period 1 and 2.....	46
Table 14: Summary statistics of riots and protests, for period 1 and 2.....	47
Table 15: Global Moran's I fro riots and protests, for period 1 and 2.....	50
Table 16: Countries where LISA for riots and protests is significant, for period 1 and 2	53
Table 17: Significant diffusion patterns for riots and protests.....	54
Table 18: Summary statistics for violence against civilians.....	55
Table 19: Global Moran's I for violence against civilians	56
Table 20: Countries where LISA for violence against civilians was significant.....	58
Table 21: Summary statistics for violence against civilians, for period 1 and 2	58
Table 22: Global Moran's I for violence against civilians, for period 1 and 2.....	60
Table 23: Countries where LISA for violence against civilians was significant, for period 1 and 2	62
Table 24: Significant diffusion patterns for violence against civilians	63

ACRONYMS AND ABBREVIATIONS

Acronym	Definition
ACLED	Armed Conflict Location and Event Dataset
CSDS	Categorical Spatial Datasets
EDA	Exploratory Data Analysis
ESDA	Exploratory Spatial Data Analysis
GSA	Global Spatial Autocorrelation
LICD	Local Indicators of Categorical Data
LISA	Local Indicators of Spatial Autocorrelation
LMSA	Local Measures of Spatial Autocorrelation
SA	Spatial Autocorrelation

1 INTRODUCTION

1.1 Background

The Global Peace Index (GPI) of the Institute for Economics and Peace (IEP) shows that despite the progress made in fostering peace across the globe, the world has become less peaceful than it was in 2008 (IEP, 2016a). According to the Global Terrorism Index (GTI), the world has also become more violent over the last 10 years (IEP, 2016b). Violence manifests in different forms and thus has multiple definitions. One definition of violence is “*the intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation*” (Krug et al., 2002). Conflict, on the other hand, is defined as “*diverse but recurrent forms of violence between individuals and groups*” (OECD, 2016). A United Nations Refugee Agency (UNHCR) report of 2015, estimated that 65.3 million people across the world were forcibly displaced from their homes, with 21.3 million among them being refugees¹. The global economic impact of violence was estimated at US\$ 13.6 trillion or 13.6% of the world's gross domestic product (GDP) in 2015 (IEP, 2016b), and the increasing levels of violence as well as conflict-induced displacement of people are likely to continue into the near future (IEP, 2016b).

A less peaceful world and persistent conflicts have the potential of hindering the achievement of the United Nations (UN) Sustainable Development Goals (SDGs) which among other things aim to end poverty, protect the planet, and ensure that no one is left behind by 2030 (UN, 2015). The SDGs include a goal on peace, justice and accountability (Goals 16) which stresses the risks of violence to global peace and security as well as the need for a better understanding of violence in the realisation of the SDGs (UN, 2015). It is therefore important to develop tools to better understand when, where and how conflicts occur in order to help governments make effective policies, non-governmental organisations to better channel scarce resources, and to enable peacekeeping and security organisations to prevent and minimise escalation and spread of violence (UN, 2015).

¹ UNHCR Global Trends: Forced Displacement in 2015. Available at www.unhcr.org/576408cd7

The diversion of resources and enforced trade-off between maintaining peace and economic development that results from violent conflict are particularly problematic in Africa where the interminable conflicts such as those brought about by the activities of groups like the Mozambican Resistance Movement (RENAMO) in Mozambique, terrorism (by groups such as Al Shabaab in Somalia and Boko Haram in Nigeria), separatist movements (by groups such as Mai-Mai Kata Katanga in the Democratic Republic of Congo) and the increasingly violent protests such as #FeesMustFall and service delivery protests in South Africa, mean that this is a problem in need of a remedy (IEP, 2016b).

Traditionally, armed conflict research has been a subject matter under the disciplines of international relations and political science. However, as data collection and analysis methods have improved, the analysis of civil and armed conflict data has become more sophisticated (Raleigh, Witmer and O'Loughlin, 2010b). Although conflict data usually exhibit spatial autocorrelation (SA) which is the tendency for spatial data to exhibit similarity as a function of closeness of locations (Ord and Getis, 2001), analytic studies in conflict research often fail to examine such SA (Schutte and Weidmann, 2011). Instead, analytic studies that have made use of conflict datasets employ descriptive measures and are suggestive of spatial clusters but seldom test for such clusters. The implication is that the phenomenon of “violence itself causing violence” is often ignored in inordinate favour of more exogenous explanations (Schutte and Weidmann, 2011).

Conflict and violence studies are often done at the country level whilst neglecting the sub-national level. This is unfortunate as sub-national studies could help with the identification of inherently geographical independent variables, proper assessment of local patterns and therefore lead to correct specification of models (Schutte and Weidmann, 2011; De Juan, 2012). A further drawback of analyses which ignore the sub-national level is that inferences made often exhibit the ecological inference fallacy. This fallacy occurs when a phenomenon that is observed at an aggregated level (for example national level) is incorrectly assumed to also occur or explained by variables at the disaggregated level (for example sub-national level). Researchers, such as Raleigh and Hegre (2009) and Hegre, Ostby and Raleigh (2009), try to circumvent the problem of SA by employing exogenous factors (explanatory variables) in their analysis.

The upsurge in violent events in the last 30 years as a result of rebel groups, terrorist organisations and civil wars (Starr and Iqbal, 2014) requires a rigorous analysis, while the increasing availability of geographically referenced conflict data and spatial methods of analysis makes it possible to holistically study where, how and why conflict events occur. As noted by Schutte and Weidmann (2011), a good understanding of the spatial patterns of data on violence is useful in characterising how clusters of violence evolve spatio-temporally, in setting hypothesis and in specifying and selecting spatial models for the proper analysis of such data. The aforementioned are the reason for this research report.

1.2 Study Objectives

Broadly, the objective of the research report is to analyse the spatial and temporal patterns of violent events in Africa during the time period 1997 to 2017. More specifically, the objectives of the research report are to:

- Test for evidence of global spatial autocorrelation in violent events and fatalities in Africa
- Test for and establish evidence of local spatial clusters in violent events and fatalities in Africa
- Examine the diffusion of violent conflicts and fatalities over space and time in Africa
- Examine the changes that result from analysing conflict data at various spatial and temporal aggregation levels in Africa.

1.3 Structure of the report

Chapter 2 is the literature review which briefly surveys the statistical framework relevant to examining spatial patterns and their evolution. Chapter 3 details the methodology and explores the statistical techniques used in the report. Chapter 4 consists of the analysis of the conflict events through time and space. Chapter 5 is comprised of the summary, conclusions and recommendations for further work.

2 LITERATURE REVIEW

2.1 General overview

Due to the volume of data that is usually dealt with in statistics, a way of gleaning the essence of any data set is required. Tukey (1977) developed the concept of exploratory data analysis (EDA) which offers a way of gaining an initial look into data prior to any formal statistical modelling and inference. EDA, among other things, places emphasis on “graphic representations of data, tentative model-building and hypothesis generation through an iterative process” (Behrens, 1997).

In the particular case of spatial statistics where location information is available, the term for EDA is exploratory spatial data analysis (EDSA) (Anselin, 1995). According to Cressie (1993), spatial data is defined as follows: Let $\mathbf{s} \in \mathbb{R}^d$ be a generic data location in d -dimensional Euclidean space. Suppose the observation $\mathbf{Z}(\mathbf{s})$ at location $\mathbf{s}$ is a random quantity. Let $\mathbf{s}$ vary over index set $D \subset \mathbb{R}^d$ so as to generate the random process $\mathbf{Z} = \{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D\}$. A realisation of $\mathbf{Z}$ is denoted $\{\mathbf{z}(\mathbf{s}) : \mathbf{s} \in D\}$, and is known as an observation. A collection of these observations are a single realisation of the process $\mathbf{Z}(\mathbf{s})$ and are known as spatial data (Cressie, 1993).

A major concern in spatial analysis is the tendency for spatial data to exhibit spatial effects, which manifests in the form of SA and/or spatial heterogeneity (SH) and the need to take such effects into account in data analysis (Anselin, 1995; Ord and Getis, 2001). Tobler’s first law (TFL) of geography states that “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). This law, ultimately, is what allows for the possibility of spatial statistics because it enables the exploitation of observations with full information for the sake of inferring information about nearly-located observations with partial or no information (Tobler, 1970). Thus, EDSA mostly entails describing spatial patterns, identification of spatial outliers, the identification of SA (both at the local and global levels) and detection of spatial heterogeneity (Anselin, 1995; Ord and Getis, 2001).

SA focuses on the association or similarity between information from two locations as a function of the distance between them (Ord and Getis, 2001). Put differently, SA is the relationship between realisations $z(s_1), \dots, z(s_n)$ as a function of the distance and/or direction between locations $s_1, \dots, s_n$. SA, when it exists, is usually positive. Intuitively this means that near things will be more similar than distant things. Negative SA, on the other hand, usually points to the dissimilarity of neighbouring observations (Cressie, 1993). SA can be analysed at the global or local levels.

Global spatial autocorrelation (GSA) looks for an overall relationship between all measured observations in all locations in the spatial data. GSA is quantified using global indicators such as Moran's I (Moran; 1950) and Geary's c (Ord and Getis, 1992). Boots (2003) points out the problem of the average in the context of spatial statistics, meaning: GSAs are constructed using the whole data only to whittle all the information down to a single value. Not only does this result in a loss of information, but placing importance on a single value serve to reinforce the easy but oft-erroneous assumption that the global trend holds throughout the study area.

Local SA (LSA), on the other hand, tries to find evidence for relationship between a subset of the overall spatial data. LSA is usually quantified using local indicators of spatial autocorrelation (LISA) (Anselin, 1995) or local Geary's c (Ord and Getis, 1992). Anselin (1995) notes that "the assumption of stationarity or structural stability" of a variable under investigation "over space may be highly unrealistic", and highlights the occasional divergence of the local patterns of SA and the global one. This divergence is important to keep in mind because the existence of SA globally, makes the assessment of local instabilities all the trickier and is a potential source of error. The complexity stems from the need to take global association into account prior to measuring local SA (Anselin, 1995).

Ord and Getis (2001) emphasises that care has to be taken when adopting certain assumptions (such as the absence of global SA) about the local spatial structure of data since failure to do so will result in unreliable inferences being made. Anselin (1995) sought to map out measures of SA which took local instabilities into account, and demonstrated how this class of measures can be constructed. Anselin (1995) showed that such measures allowed for the decomposition of global indicators such as Moran's I (Moran; 1950) and Geary's c (Ord and Getis, 1992).

Anselin (1995) also highlights two broad insights provided by LISAs:

1. LISAs could serve as indicators of hot spots by revealing areas of significantly high SA
2. LISAs could serve as a means of gauging the weight contributed by certain locations, with highly-weighted locations perhaps representing outliers

Li and Reynolds (1995) give a working definition of SH as “*the complexity in both composition (which is nonspatial) and configuration (which is spatial)*”. An important contribution from Li and Reynolds (1995) is that by formally delineating the two main components of spatial data, namely composition and configuration, the possibility of developing local indicators of categorical data (LICD) was established. Boots (2003) showed that in order to construct LICDs, configuration has to be conditioned on composition. Boots (2003) also provides means of quantifying composition and configuration in constructing LICDs. However, this is only done in the context of gridded, binary data by counting the number of black and white tiles (in the case composition), and using join-counts for configuration. Boots (2006) further explored configurational measures and showed that there are many ways of assessing configurational characteristics (Boots, 2006).

The report by Anselin and Piras (2009) has at its core, an analysis based on adapting continuous data methods to count spatial data, where the variable of interest has integer values. They dichotomize their data before using LICD. They note the loss of information and accuracy inherent in dichotomising the data and made this explicit when comparing the results of their analysis. Although taking advantage of the irregular lattices was not done, a permutation approach was suggested. Due to the ubiquity of global dependence in many data sets, as well as potentially detrimental effect when ignored, the general procedure of developing a measure that takes it into account is invaluable. In a similar vein, Kabos and Csillag (2002) constructed a statistic that is based on join-counts but which can be applied in a context where first-order homogeneity (the absence of GSA) is not assumed. In the related field of ordinal spatial data, Kulldorf (2007) extends his work on *Spatial Scan Statistic* (Kulldorf, 1997) where he developed a scan statistic that detected clusters for count data, to one that detects clusters in ordinal data.

The collation of spatial data that pertains to conflict has allowed for a more quantitative approach to the study of armed conflict. However, the bulk of the work dealing with analysing conflict data done thus far has not wholly embraced the spatial statistical tools that would allow for a more rigorous appraisal of violent phenomena. The importance of quantitatively rigorous studies of violence arises from the fact that the subtle biases that may plague research that has thus far relied overly on qualitative techniques may be rectified. This in turn will ultimately allow for a more nuanced understanding of the factors that drive conflict which may, in conjunction with other actions, hopefully, lead to its eradication (Schutte and Weidmann, 2011).

Raleigh et al. (2010b) examined the spatial and temporal evolution of conflict in the Democratic Republic of Congo by using cluster analysis. However, this method of analysis did not enable the authors describe the actual mechanism through which the clusters evolve over time.

Starr and Iqbal (2014) focused on assessing whether SA existed among states after the fall of the Soviet Union. They considered the periods from 1991 to 2004, and determined that SA was extant and identified spatial diffusion as the mechanism through which domestic conflict spreads. It was also noted that “while the use of ESDA in political science has been growing, the use of the specific technique of “join-counts” is still quite rare”. Their motivation for supporting the use of join-counts was that reliance “on visual inspection alone for evidence of spatial clustering in civil wars” was problematic due to the human eye’s unfortunate propensity to “identifying structure where none may exist”.

LaFree et al. (2011) considered whether the Spanish organisation, ETA, had employed different strategies in two different time periods (before and after 1978). This was determined by analysing the diffusion patterns within those time periods, which then enabled the authors to conclude that, judging from the statistically significant changes in diffusion patterns before and after 1978, ETA had indeed made use of different strategies in their terrorist attacks.

Flores and Villarreal (2015) consider the manner in which drug-cartel-related (DCR) violence has diffused to neighbouring municipalities. The extent of the violence is measured through homicides which have the hallmarks of DCR violence. The major finding is that violence in the DCR spreads to contiguous areas whilst intensifying in the original municipality. The importance of this cannot be overstated. It shows that diffusion processes, which formerly were considered for more conventional armed conflicts can also be used to measure the spread of less conventional ones.

More recently, Mitchell (2018) has used spatial diffusion to examine the spread of public policies (as opposed to events), one example being how smoking regulations spread. Typically, policy diffusion makes use of a number of theories to explain the transference of a policy from one spatial unit to the other. For example, emulation would be said to have occurred if a spatial unit adopted a policy subsequent to its largest neighbour having done so. The distinction between contemporary policy and spatial diffusion research is that the former has been largely concerned with linking spatial units which adopted policies with each other whereas the latter has focused on investigating diffusion properties or propagation patterns. By focusing on one approach while excluding the other, policy diffusion theories are limited unnecessarily. One finding was that policy diffusion, when explored using spatial diffusion seems to exhibit both hierarchical and contagious effects. However, due to the exploratory nature of the study, it was unable to demonstrate this definitively. The paper notes that one way of verifying the results would be to make use of regression modelling and spatial econometric techniques.

3 METHODOLOGY

3.1 Spatial Data Types

There are three main types of spatial data namely point pattern, geostatistical and areal data. The spatial data type that is the focus of this report is the areal or lattice data. It is defined as follows (as described by Cressie 1993):

1. For a given areal unit A , which may be regularly or irregularly shaped, let the attribute of interest, Z , at A be given by $Z(A)$
2. Let D , the domain in which the areal units $A_1, A_2, A_3, \dots$ occur be non-random and countable
3. Let the areal units $A_1, A_2, A_3, \dots$ be mutually exclusive and exhaustive of D . In the case of areal data, spatial dependence is usually measured by assessing the similarity between adjacent areal units.

3.2 Measures of Spatial Autocorrelation

3.2.1 Global Measures of Spatial Autocorrelation

The traditional methods of testing for GSA of continuous variables are the Moran's I (Moran, 1948), the *Geary's c* (Cliff and Ord, 1981) and the Getis and Ord Global G (Getis and Ord, 1992; Ord and Getis, 1995). More appropriate, but informationally-wasteful methods are binary join-count statistics since they involve the dichotomisation of data (Moran, 1948; Anselin and Piras, 2009; Cliff and Ord, 1981). When making inferences using SA measures, the one thing that binds them together is the manner in which the null hypothesis is specified (Boots 2003), namely:

H_0 : *The data are spatially random*

H_1 : *The data has spatial association.*

The Moran's I is the most widely used measure of SA (Getis, 2008) and is detailed in the following section.

3.2.2 The global Moran's I

The global Moran's I is given by:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \mu)(x_j - \mu)}{\sum_i (x_i - \mu)^2} \quad (1)$$

where X is the variable of interest, N is the number of spatial units, $(x_i \ i = 1, \dots, N)$ is the i^{th} observation of X , w_{ij} is the spatial weight matrix and μ is the mean of X over all spatial units.

The Moran's I is essentially a correlation measure with values close to 1 indicating positive SA (similar values are closer to each other); values close to -1 (dissimilar values are closer to each other), negative SA and values close to zero, spatial randomness (values are randomly scattered with no over-arching pattern).

Under H_0 (spatial randomness), the mean and variance of the Moran's I statistic are given as:

$$E[I] = \frac{-1}{N-1}, \quad (2)$$

$$Var(I) = \frac{N^2 \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2 - N \sum_i [\sum_j w_{ij}]^2 + 3 \sum_i \sum_j w_{ij}}{\sum_i \sum_j w_{ij} (N^2 - 1)} - E[I]^2 \quad (3)$$

A z-score can then be calculated as $Z = \frac{I - E[I]}{\sqrt{Var(I)}}$ and can be used to evaluate whether the spatial pattern observed deviates substantially in magnitude from what would be expected under the null hypothesis of spatial. Alternatively, p -values can be generated using two approaches and can be used to evaluate the Moran's I test statistic. These approaches are permutation tests and a parametric approach (Anselin and Piras, 2009).

Permutation tests, which are non-parametric, work as follows:

- They assign each areal unit an observed value with equal probability. For n observations this yields $n!$ possible arrangements.
- From this, an empirical distribution is then constructed.
- The test statistic is calculated by using the actual observed data, whose value is then compared to the reference distribution.
- Values which are to the tail of the distribution are indicative of deviation from spatial randomness (Anselin and Piras, 2009)

On the other hand parametric tests, work as follows:

- Make assumptions about the distribution of the variable under consideration, usually asymptotic normality
- Determine the mean and variance of the test statistic and perform z-tests, using the normal approximation
- Visually assess the Moran scatterplot

3.2.2.1 Spatial Weight Matrix

Since proximity is essential in quantifying SA, a means of measuring it is necessary. One way of doing this is by constructing a spatial weight matrix (SWM). Anselin and Piras (2009) define a SWM as follows: *A SWM (W) is a $n \times n$ “positive and symmetric matrix that defines the structure of the interaction between each pair of spatial units. In each row, the non-zero elements of the matrix express for a given observation which other observations are considered to be its neighbours”*. Proper specification of a weight matrix is important and dependent on the type of problem with different weight matrices yielding different results for the same data (Anselin and Piras, 2009).

The two major means of establishing spatial weight matrices is by way of distance and contiguity. With regards to the spatial distance weights, let d_{ij} be the distance between spatial unit i and j . Then:

$$w_{ij} = \begin{cases} 1 & \text{if } d_{ij} \text{ satisfies some criterion} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

For example, radial distance weights have the following form:

$$w_{ij} = \begin{cases} 1, & 0 \leq d_{ij} \leq d \\ 0, & d_{ij} > d \end{cases} \quad (5)$$

where d is some threshold distance.

There are a number of methods used to measure spatial contiguity of locations. These include Queen, Rook and Bishop criteria (Cliff and Ord, 1969).

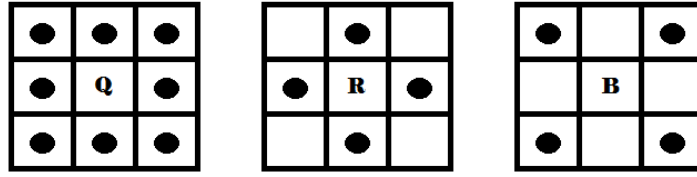


Figure 1: Different definitions of contiguity. The Queen configuration (Q) is on the left, Rook (R) in the middle and Bishop (B) on the right

Regarding the Queen spatial contiguity (with spatial unit of interest labelled Q in the Figure 1), spatial units are deemed to be contiguous with unit Q if they share at least one vertex or edge with Q. With respect to Rook spatial contiguity (R), spatial units are deemed to be contiguous with unit R if they share an edge with R. With respect to Bishop spatial contiguity (B), spatial units are deemed to be contiguous with unit B if they share a vertex with B.

A spatial weights matrix can now be defined as follows: Let two areal units be considered neighbours if they share a common edge. In the weight matrix W , areal unit A_i 's spatial relationship to A_j is denoted by entry(w_{ij}), with a 1 representing adjacency, and 0 non-adjacency (Anselin and Piras 2009). An example of such a matrix is as follows:

$$\begin{bmatrix} 0 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 0 \end{bmatrix}$$

Given the spatial weight matrix W , and ensuring that $\sum_i w_{ij} = 1$, then each row-normalised weight, w_{ij} , can now be interpreted as the fraction of all spatial influence on unit i attributable to unit j (Cliff and Ord, 1969).

3.2.3 Join-count statistics (for binary data)

Let

$$x_i = \begin{cases} 1 & \text{if an event occurred in location } i, \\ 0 & \text{if an event did not occur in location } i. \end{cases} \quad (6)$$

Let $(W)_{ij}$ be the spatial weight matrix and $BB = \frac{1}{2} \sum_i \sum_j w_{ij} x_i x_j$ be the Black-Black statistic, then the local black-black statistic, for the i^{th} location, which can be used to assess positive SA, can be defined as:

$$BB_i = x_i \sum_j w_{ij} x_j \quad (7)$$

A basic interpretation of this statistic is that the higher it is, the greater the degree of SA. For inferential purposes, there are three options: the conditional permutation, asymptotic normality and analytical approaches (Boots, 2003). The latter consists of assuming spatial randomness under the null hypothesis. This supposition implies that in a local region of m_i areal units, the probability of observing r black observations is given by:

$$P(R = r) = \binom{m_i}{r} \widehat{p}_b^r (1 - \widehat{p}_b)^{m_i - r} \quad (8)$$

Thus, determining whether a sub-region has an unusually high number of “events” consists of calculating $P(R \geq r_\alpha)$ where r_α is a critical value corresponding to a significance value α .

Boots (2003) cautioned that considering composition on its own is insufficient, that in fact, to get a clearer picture of local SA, one would have to condition configuration on composition (Boots, 2003). Li (2006) developed a measure for SA for categorical spatial data known as the transiogram. Zhang and Zhang (2008) derived the mean and variance for categorical variables with three levels, that is to say, trinary data. This is vital in that it allows for the testing of SA of trinary data, thus extending the work of Boots (2003) on binary spatial data.

3.3 Local measures of Spatial Autocorrelation

The test statistics that are usually employed to assess local SA are the local Moran's I (LISA), local Geary's c , *Local Getis and Ord's G* and the local join-count statistic. The LISA statistic which is considered below is the most common.

3.3.1 Local Moran's I

As characterised by Anselin (1995), the local Moran's I for spatial unit i is denoted as I_i and is given by:

$$I_i = (x_i - \mu) \sum_{j \in J_i} w_{ij} (x_j - \mu) \quad (9)$$

Where $(x_i \ i = 1, \dots, N), (x_j \ j = 1, \dots, N) . i \neq j$ the i^{th} and j^{th} observations of X , $(W)_{ij}$ is the spatial weight matrix, μ the mean of X over all observations, and J_i is the set of neighbours of spatial unit i .

Essentially the local Moran's I is a covariance measure with high positive values being indicative of positive SA; low negative values are indicative of negative SA and values close to zero, of spatial randomness (Anselin and Piras, 2009). The results of the local Moran's I are mapped and interpreted similar to the Moran's Scatterplot described in 2.4.1.1 below. Significance testing for LISA is done in the following way: assumptions about the distribution of the variable under consideration, usually asymptotic normality. The mean and variance of the test statistic are then calculated and a z-test is performed, using normal approximation (Anselin and Piras, 2009).

3.3.1.1 The Moran scatterplot

The Moran scatterplot (Figure 2) is used to visually assess the results of the LISA test.

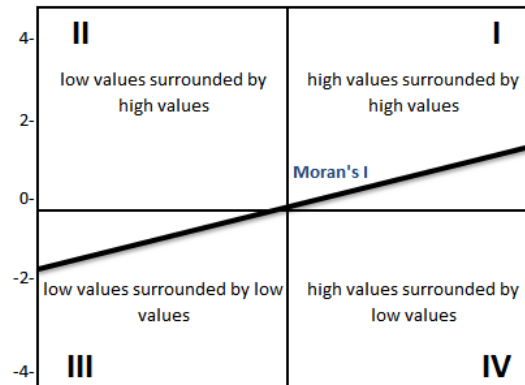


Figure 2 Diagram of Moran Scatterplot

The quadrants of the Moran scatterplot are interpreted as follows: when a spatial unit j which is surrounded by other units is labelled HH it means spatial unit j has a high positive value and is surrounded by units with high positive values. This unit and its surrounding area are deemed a hotspot. When a spatial unit j which is surrounded by other units is labelled LL it means spatial unit j has a low positive value and is surrounded by units with low positive values. This unit and its surrounding area are deemed a coldspot. When a spatial unit j which is surrounded by other units is labelled LH it means spatial unit j has a low positive value and is surrounded by other units with high positive values. This unit is deemed an outlier. These patterns are depicted in the quadrants in Figure 2. The slope of the Moran Scatterplot indicates the global extent of SA. The Moran scatterplot does not provide information about the significance of such patterns (Anselin, 1993).

3.3.2 Local Join-Count

Local join-counts work exactly like global join-counts with the major difference being that instead of using the whole map, a *smaller* spatial unit, is considered sequentially until the whole map has been scanned. For each spatial unit, a black-black statistic is calculated (explained in section 2.3.3). The basic interpretation of this statistic is that the higher it is, the greater the degree of SA in that window (Anselin and Piras, 2009).

3.4 Spatial Diffusion Analysis

Cohen and Tita (1999) extended the LISA to consider the temporal aspect thus providing a detailed framework for exploring spatial diffusion.

1. Calculate a local version of Moran's I or some other LISA for all the locations
2. Decompose the LISA values for each location into High-High (HH), Low-Low (LL), High-Low (HL) and Low-High (LH).
3. Compare the period-to-period transition among different significant LISA pairs in successive period.
4. So a transition from LH at time t to HH at time $t + 1$ would indicate a contagious diffusion pattern, more specifically an expansionary one. Similarly a transition from LH at time t to HL at time $t + 1$ would still indicate a contagious diffusion pattern, although in this case it would be relocation diffusion. They also mention a hierarchical diffusion pattern where the spread does not occur with direct contact but rather by cultural influence, so that subgroups which are not close geographically nevertheless influence each other. One example would be LL to HL, which would be a hierarchical isolated increase.

The diffusion framework is summarised in Table 1 and Figure 2.

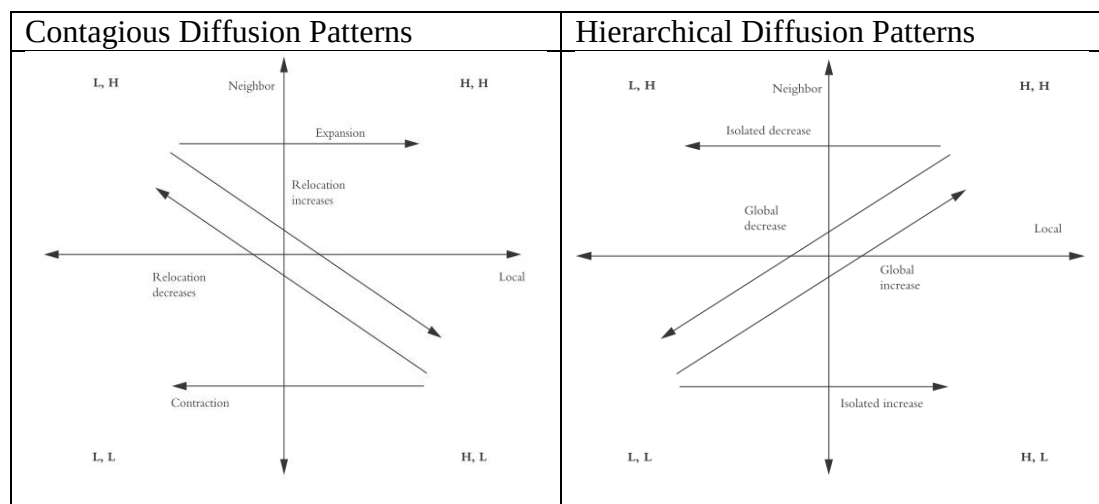


Figure 3 Depiction of Contagious and Hierarchical Diffusion Patterns²

² adapted from Sánchez, Solimano, and Formisano (2005)

Table 1: Diffusion framework

			Time-to-time Change in Local Neighbour Pairs	
Direction of Change	Type of Diffusion	Mechanism of Change	Diffusion Outcome	Diffusion Source
From low to high levels	Contagious	Expansion from neighbours	LH TO HH	HL TO HH
		Relocation from neighbours	LH TO HL	HL TO LH
	Hierarchical	Isolated increase	LL TO HL	LL TO LH
		Global increase	LL TO HH	LL TO HH
From high to low levels	Contagious	Expansion from neighbours	HL TO LL	LH TO LL
		Relocation from neighbours	HL TO LH	LH TO HL
	Hierarchical	Isolated increase	HH TO LH	HH TO HL
		Global increase	HH TO LL	HH TO LL
No change	None	Stationary		

A spatial diffusion analysis would enable us gauge how the HH, LH, HL and LL values have been changing and to draw inferences as to whether violent conflict has been expanding, relocating or not significantly changing at all. In terms of the main diffusion patterns postulated, the *relocation* diffusion is “based on the assumption of clear-cut front lines between the two zones of control” (Schutte and Weidmann, 2011). It characterised by the shift from a zone which is controlled, to yet another zone which becomes contested. The *escalation* diffusion is defined as “the expansion of the geographic scope of violence to new locations during the course of the conflict”. This pattern is characterised by the expansion from one zone which is still contested, to yet another zone which becomes contested (Cohen and Tita, 1999; Schutte and Weidmann, 2011).

LaFree, et al. (2012) suggest a test that can be used to establish whether a statistically significant change has occurred. First, count the number of times that the transition LH to HH (increase from adjacent units) occurs between times t and $t + \delta t$, say n . Then count the total number of spatial units, say N (increases from adjacent and non-adjacent units). In order to perform the test, assume that the probability of there being an increase (L to H) from adjacent units and is the same as that from non-adjacent units (that is: 0.5).

The probability of observing a given sample would then be:

$$\hat{p} = \binom{N}{n} (0.5)^n (0.5)^{N-n} \quad (10)$$

The hypothesis would then be defined as follows:

$$H_0: p = 0.5 \quad vs \quad H_1: p \neq 0.5$$

Thus our test for significance would then be done using the Z-test:

$$Z = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5 \times 0.5}{N}}} \quad (11)$$

If H_0 is rejected, then this would imply that there has been a statistically significant increase from adjacent units, thus indicating an expansion from neighbours.

Regarding the binary approach, Baudains, Johnson and Braithwaite (2013) described its implementation and show how to determine whether a statistically significant evolution (relocations and escalations) through space and time has occurred using permutation testing in a Monte-Carlo framework. Something worth noting is that if count data are used then a unit is deemed as having gone from L to H if the result of the local Moran's I test is statistically significant for the spatial unit j in $t + \delta t$. However, if binary data are being used, it is sufficient if there has been at least one occurrence of a violent event in the spatial unit j in $t + \delta t$ (Baudains, Johnson and Braithwaite, 2013).

4 DATA ANALYSIS AND DISCUSSION

4.1 Study Area and Data Description

The research report focuses on conflict events in countries located on the African continent (Figure 4). Africa faces several of challenges as a result of conflicts and an understanding of these conflicts might help in mitigating them and, as a consequence, the challenges that arise from them.



Figure 4 Map of mainland Africa³

Polygons representing the African continent and the various administrative sub divisions were obtained from the Global Administrative Areas (GADM). The GADM is a spatial database of the location of the world's administrative boundaries which can be used in Geographic Information Systems (GIS) and statistical software. Polygons representing countries (administrative level 0), provinces/states (administrative level 1) and other political administrative units can be downloaded from GADM's website⁴.

Conflict data was collected from the Armed Conflict Location and Event Data Project (ACLED, 2016) available from <http://www.acleddata.com/>. The ACLED is the most comprehensive human-coded public collection of political violence and protest data for developing countries compiled from various sources (Raleigh et al.,

³ GeoDa

⁴ www.gadm.org

2010a). It is a public repository of highly disaggregated conflict data from 1997 to near real-time and allows researchers the flexibility to analyse the data in as fine a detail as possible, or to aggregate the data across dimensions which include events (battles, riots and violence against civilians), actors (civilians, government, rebels or their interactions), time (such as hourly, daily, monthly or yearly) and location (at the location of event, arbitrarily defined aggregated levels such as grid, or administrative units such as countries and states among others). The ACLED has been previously referenced and used by several conflict researchers such as Raleigh (2010), Beardsley (2011), Schutte (2011) and Minoiu and Shemyakina (2014).

While other spatial conflict data repositories such as the Uppsala Conflict Data Program Geo-referenced Events Dataset (UCDP GED) exist, one major advantage that ACLED has over other repositories is that in addition to collecting data on fatality counts, ACLED also has data on non-fatal and non-violent events such as protests (Eck, 2012). One of the advantages of the ACLED dataset is that aggregating the data on the administrative level allows for the exploration of data as function of spatial scale. Another advantage is that it has consistent data that spans two decades, thus allowing for spatial analysis that includes a temporal component. The limitation of the ACLED data comes from the selective inclusion of events as well as the possibility of having inaccurate location and time stamps. To mitigate this limitation, the ACLED includes flags to indicate the levels of geographical and time precision which can be used to minimise the possible misallocation of events in space and time (Raleigh et al., 2010a).

4.2 Data Analysis

Figure 5 gives a summary of the analyses that will be performed. GeoDa software (Anselin, 2017), the R computing environment (R Core Team, 2017) and the spdep package (Bivand, 2013) are employed for the analyses. The types of event that will be considered are riots/protests, violence against civilians, battles and terrorist attacks. It is crucial to note that while different types of events can occur in the same lattice unit, each event will be analysed separately. In order to assess the nature and spread of conflict, two types of variables will be utilised: count variables to measure the number of conflicts or fatalities, and binary variables to measure presence/absence of violence or fatalities.

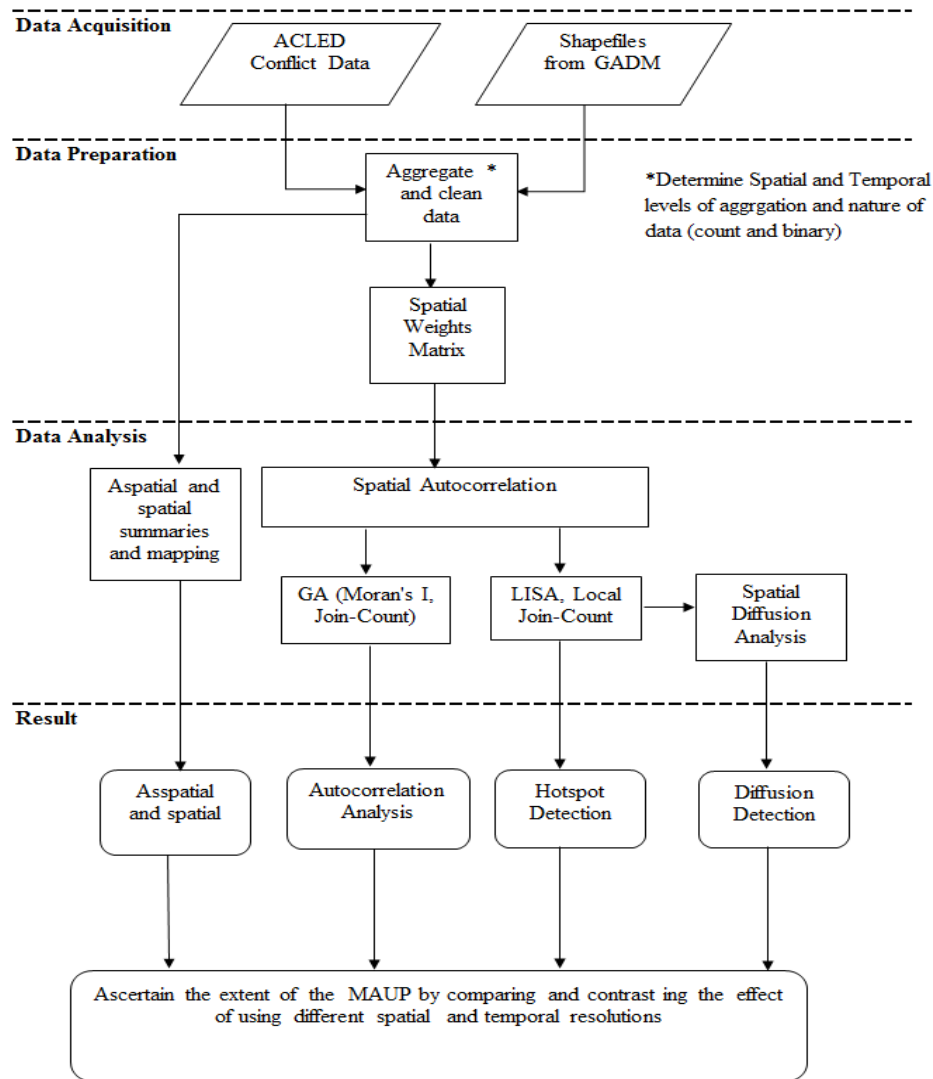


Figure 5 Diagram of analysis plan

The data will be analysed at three spatial resolutions: country (administrative level 0), the first administrative level (consisting of province/states depending on the country) and the second administrative level (consisting of municipalities / districts / constituencies and so on). This will allow for the investigation of whether patterns and associations, which hold at one spatial unit, continue holding at another. More technically it will allow for investigating the Modifiable Areal Unit Problem (MAUP). MAUP is defined as “geographic manifestation of the ecological fallacy in which conclusions based on data aggregated to a particular set of districts may change if one aggregates the same underlying data to a different set of districts” (Waller and Gotway, 2004).

Three temporal resolutions will be used. The time periods selected are from 1997 to 2017, from 1997 to 2006 and from 2007 to 2017. The first time period corresponds to the entire period for which the ACLED data was collected for Africa and the last two periods are arbitrarily defined to split the data into two equal parts to allow for the investigation of possible diffusion.

To measure the presence of GSA, the global version of Moran's I will be used for the count variable, whereas the Join-Count statistic will be used for the binary variable. Queen spatial contiguity weights will be used (see section 2.4) due to its popularity in literature (Anselin and Piras, 2009). To measure the extent of local SA, within spatial units for all time periods, the local version of Moran's I (LISA) will be used for the count variable whereas the local Join-Count statistic will be used for the binary variable. This will yield two separate results for each spatial unit which will enable the determination of similarities in the significance of the local SA.

The evolution of the local SA through time will be determined by way of spatial diffusion. This will enable us to understand whether, firstly, conflict event in question has been increasing or decreasing and, secondly, to determine the mechanics of this change that is, whether it has been expanding, relocating or, perhaps, been spreading but not through direct contact (hierarchically).

4.3 Evaluation

As already described in Section 3.2 three major analyses will be done: determining the SA (that is, clustering), determining the hotspots of conflict event and determining the evolution of the hotspots through time. To initially evaluate the results, a visual assessment, by eye, of patterns will be done as a preliminary step to establish patterns and changes, if they exist. To evaluate the significance of clustering for a given temporal and spatial resolution, the parametric approach for Moran's I will be employed, whereas for join-counts the black-black statistics will be used (as detailed in Section 3.2).

The local Moran's I for the various spatial and temporal units will be further decomposed and their behaviour through time will be considered. In order to determine if there is a statistically significant evolution through time, the test which

makes use of the binomial distribution that was described in Section 3.2 will be used for count data. Likewise, the test which makes use of permutation testing in a Monte-Carlo framework (Baudains, Johnson and Braithwaite (2013)) will be used for joint counts.

4.4 Results

With regards to preparing the data for analysis, the data was assessed for completeness and for outliers. The data used to construct the spatial boundaries (at country, provincial/state and municipal level) was collected from GADM. Although there are plenty of conflict-related variables to choose from, this report focused on number of events related to battles (an aggregation of several battle-related variables), riots and protests, violence against civilians, which account for approximately 84% of the occurrences in the ACLED data. The three variables can also be viewed in terms of time periods. For a given variable and region this will simply be the total number of events that occurred in this period. Lastly, there are three administration levels that are considered. They are: Admin 0 (country level), Admin 1 (state or province/etc. level) and Admin 2 (level equivalent to municipal district). For the sake of convenience, a naming system given in Table 2 was adopted for the variables constructed.

Table 2: Naming convention of variables

Number of Events		Conflict Event		
		Battles	Riots & Protests	Violence Against Civilians
Time Period	1997 to 2017	B_C	RP_C	VAC_C
	1997 to 2006	B_C_0	RP_C_0	VAC_C_0
	2007 to 2017	B_C_1	RP_C_1	VAC_C_1

4.5 Spatial Analysis: Spatial Autocorrelation – Counts

All analyses start by looking at the summary statistics of the variable under consideration. This is followed by a discussion of the spatial distribution of the variable, an exploration and testing for SA at both the global and local levels and finally a test for spatial diffusion is conducted for the variable. A comparative analysis is done for the variable across all three admin levels. In summary, the following will be considered in this exact order:

Table 3: Order of Analysis for count of events

	Variable	Time Period	Admin Level
1	Battles	1997 to 2017	0 1 2
2		1997 to 2006 & 2007 to 2017	
3	Riots & Protests	1997 to 2017	0 1 2
4		1997 to 2006 & 2007 to 2017	
5	Violence Against Civilians	1997 to 2017	0 1 2
6		1997 to 2006 & 2007 to 2017	

4.5.1 Summary of Battles from 1997 to 2017

Considering the number of battles that occurred in in the period 1997 to 2017, Table 4 shows that there were 47 385 battle incidents for the study area. The variations in the totals for the admin levels 1 and 2 could be attributed to data aggregation issues around countries such as Libya, Lesotho Sudan and Western Sahara.

Table 4: Summary statistics of battles

B_C	Total	Average	Minimum	Median	Maximum
Admin 0	47 352	986	2	205	12 053
Admin 1	47 352	63	0	10	4 108
Admin 2	47 352	11	0	1	4 108

The mean number of battles for the period 1997 to 2017 is greater than the median, across all admin levels which implies right skewness in the data. One way of circumventing the effects of skewness will be considered later when join-counts of the dichotomised conflict variables are examined.

4.5.2 Chloropleth Maps for Battles from 1997 to 2017

For the three admin levels, the number of battles in the period 1997 to 2017 were spatially distributed as follows (dark blue to dark red represents low to high). The maps in Figure 6, reveal some evidence of local clustering around the central region (Angola and DRC) and the horn of Africa. The patches of white were due to loss of information that occurred when merging the ACLED and GADM data sets (this is because the names in the two datasets are not always consistent). This was the case for Libya, Lesotho Sudan and Western Sahara.

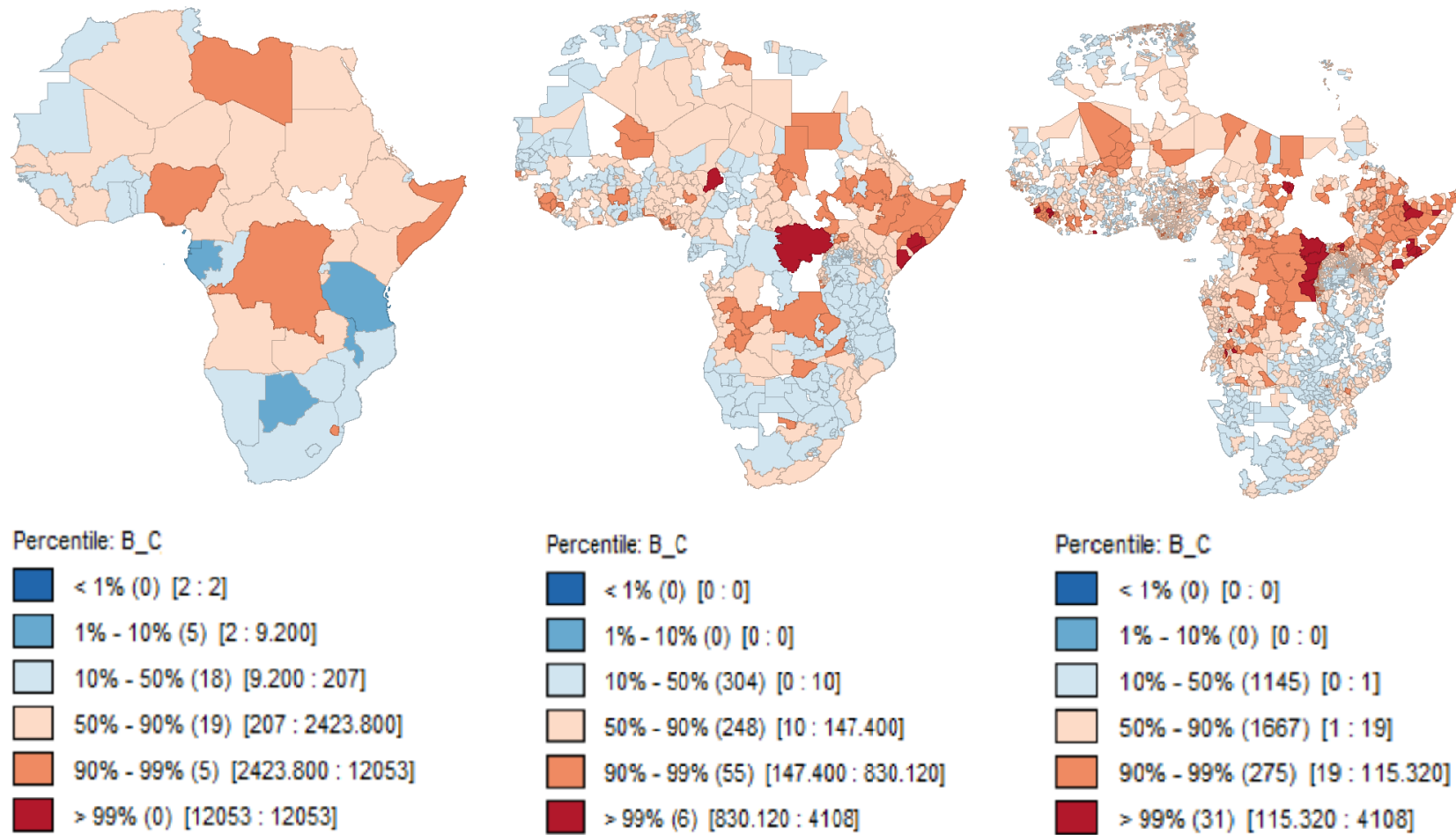


Figure 6 – Percentile maps of battles at admin level 0, 1 and 2

4.5.3 Battles from 1997 to 2017- Global Moran's I

The GSA by way of Moran's I can now be determined for all admin levels. The results are summarised in Table 5 and show that GSA is absent in admin level 0 but positive GSA is present at both admin levels 1 and 2. This implies some degree of clustering of values with similar levels of battles at both admin levels 1 and 2.

Table 5: Global Moran's I for battles

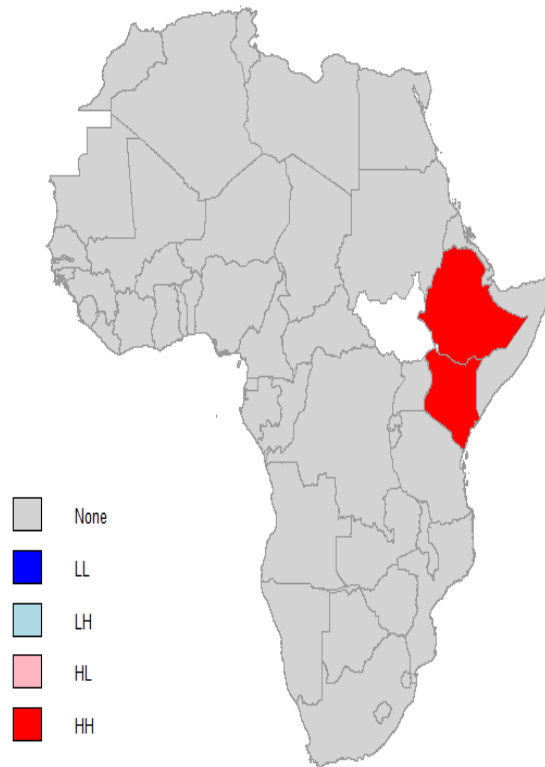
B_C	Moran's I	Expectation	Variance	p-value
Admin 0	0.014	-0.0217	0.00582	0.332
Admin 1	0.374	-0.00163	0.000565	<0.05
Admin 2	0.195	-0.000324	0.000091	<0.05

The GSA only reveals the extent of clustering and a local test has to be conducted to identify clusters and outlying areas of battles. The Moran's scatterplot in Figure 7 however shows that out of the 47 countries under consideration, only a few countries can be considered as outliers (corresponding to the HH and HL quadrants as defined in Section 3.3.1.1).

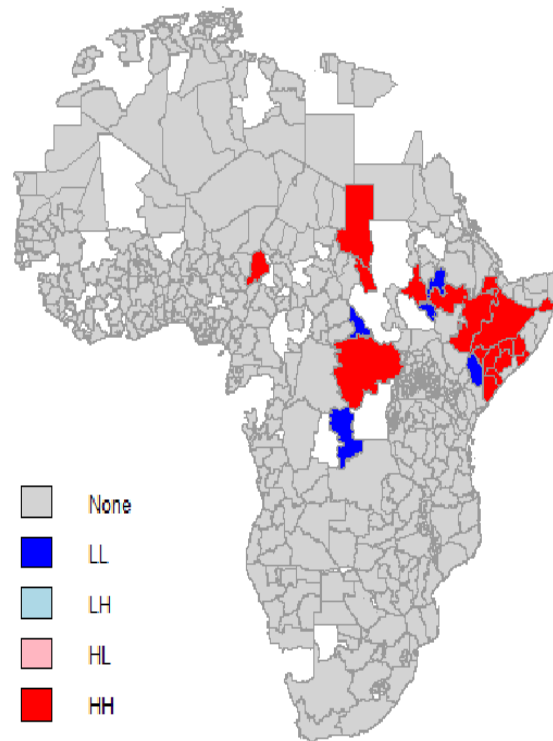
4.5.4 Battles from 1997 to 2017- LISA

When significance tests of LISA are done, the results are presented in the form a LISA map showing areas of significant LL, LH, HL and HH battles. The LISA maps in Figure 7, for all three admin levels, show that the most outlying countries fall into the HH cluster (corresponding to high activity country in a high activity battle neighbourhood). This implies these countries are significant hotspots of battles.

Countries with Significant Activity - B_C



Provinces with Significant Activity - B_C



Municipalities with Significant Activity - B_C

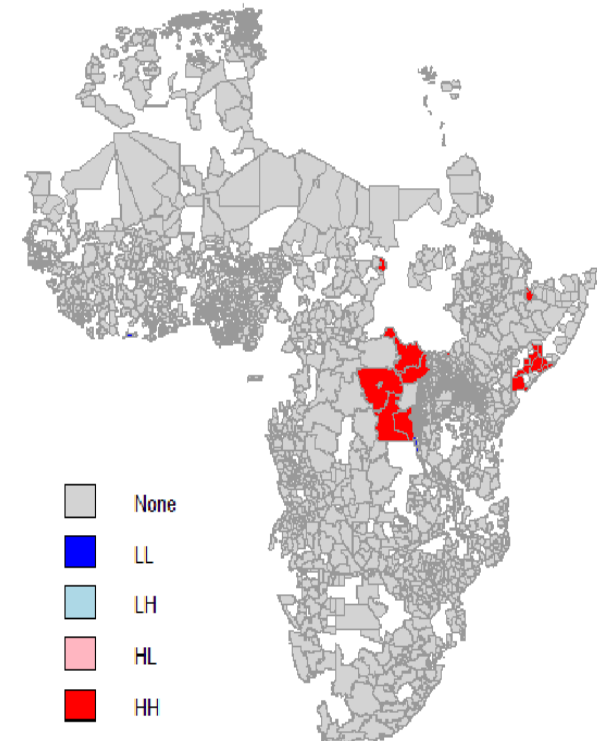


Figure 8 LISA maps of battles at admin levels 0, 1 and 2

A further interrogation of the results (Table 6) reveals that at both admin 1 and 2, several regions in Somalia and Sudan seem to be hot spots for battles. Also, central Africa has several hot spots mostly due to activity in Burundi, the CAR and the DRC. It is worth noting that the CAR, the DRC, Somalia and Sudan have all have been affected by civil wars in the period 1997 to 2017, some of which are still on-going. Notable countries with significant cold-spots are Ethiopia (at admin 1) and Burundi (at admin 2).

Table 6: Countries with significant LISA results

Country	ADMIN 1			ADMIN 2		
	HH	LL	Total	HH	LL	Total
Angola				1		1
Burundi				3	5	8
Central African Republic		1	1	2		2
Côte d'Ivoire					1	1
Democratic Republic of the Congo	1	1	2	6		6
Ethiopia	2	5	7	2		2
Kenya	1	1	2			
Nigeria	1		1			
Rwanda				2	5	7
Somalia	8		8	15		15
South Sudan	1		1			
Sudan	2		2	1		1
Tanzania					1	1
Uganda		1	1	6	11	17
Total	16	9	25	38	23	61

4.5.5 Summary of Battles from 1997 to 2006 and from 2007 to 2017

Focusing on the events that occurred in the periods from 1997 to 2006 and from 2007 to 2017, the total number of events and average number of conflict events along with other summary statistics are provided in Table 7. The results show that for each admin level considered, there was almost double the number of events in the second period spanning 2007 to 2017 compared to the period from 1997 to 2006. Battles also exhibited right skewness for each admin level as indicated by the mean being consistently larger than the median. This is indicative of a few regions with disproportionately higher number of battles.

Table 7: Summary statistics of battles, for period 1 and 2

	Admin Level	Total	Average	Minimum	Median	Maximum
B_C_0	Admin 0	16 786	350	0	51	2 145
	Admin 1	16 786	22	0	2	717
	Admin 2	16 786	4	0	-	717
B_C_1	Admin 0	30 566	636	1	64	10 461
	Admin 1	30 566	41	0	4	3391
	Admin 2	30 566	7	0	0	3391

4.5.6 Chloropleth Maps for Battles from 1997 to 2006 and from 2007 to 2017

From the maps presented in Figure 9, there is some evidence of local clustering specifically around the East Africa and Central Africa for all admin levels in the periods from 1997 to 2006 and from 2007 to 2017. The clustering seems to be due to a relatively higher number of battles in these regions, in particular Somalia and the DRC. Notably, these regions seem to include most of the regions that exhibited clustering when considering the aggregated data from 1997 to 2017.

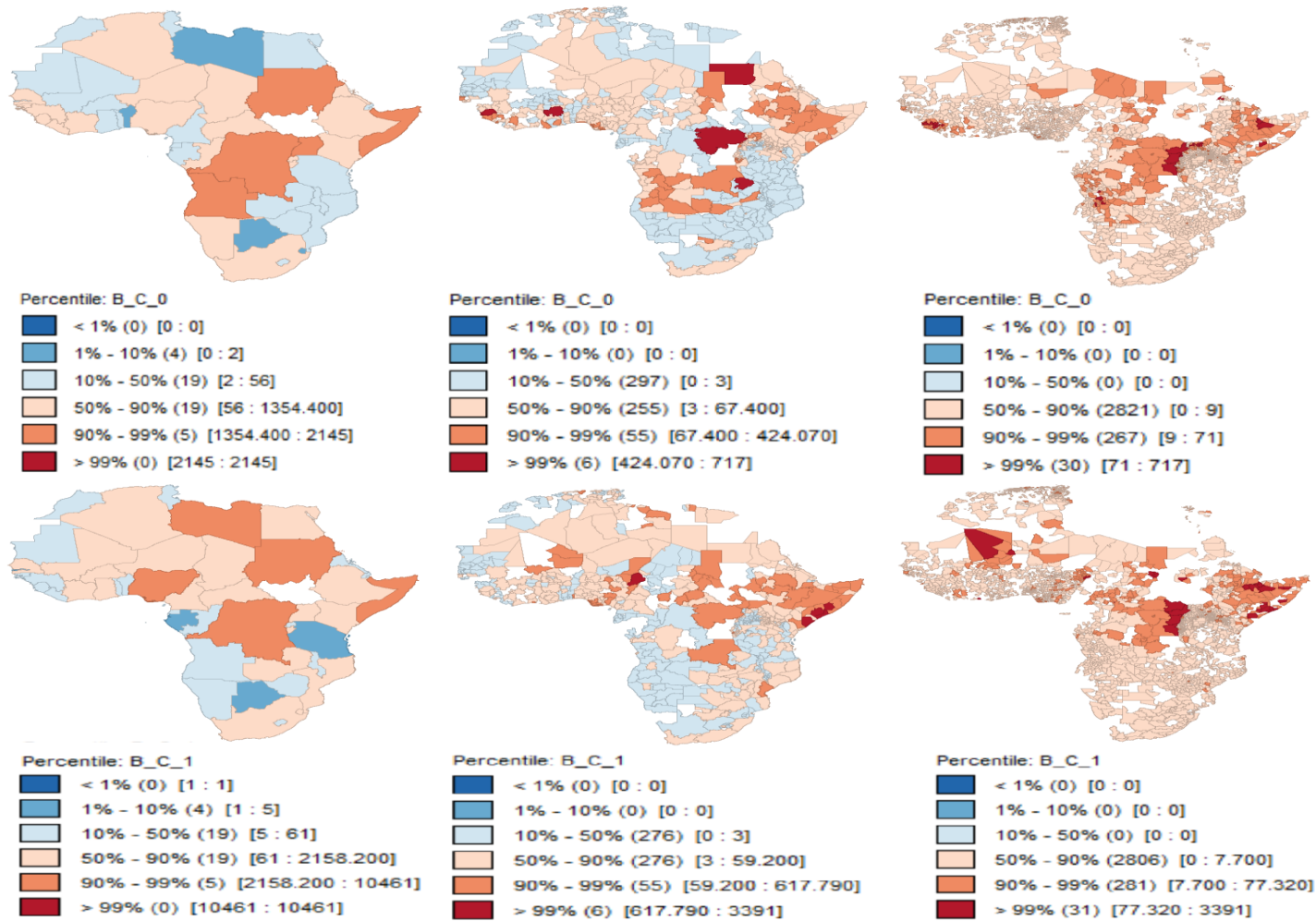


Figure 9 Percentile maps of battles, before and after 2007, at admin level 0, 1 and 2

4.5.7 Battles from 1997 to 2006 and from 2007 to 2017- Global Moran's I

The results of GSA by way of Moran's I for all three admin levels show that with the exception of admin level 0 in period 1, positive GSA is significant in both time periods and across all admin levels (Table 8). While the Moran's I value are very close to zero in all cases, the results of the hypothesis tests imply that battles exhibited a degree of clustering of at admin levels 1 and 2.

Table 8: Global Moran's I for battles, for period 1 and 2

	B_C_0	Moran's I	Expectation	Variance	p-value
B_C_0	Admin 0	1.58E-01	-2.17E-02	9.91E-03	<0.05
	Admin 1	2.30E-01	-1.63E-03	7.78E-04	<0.05
	Admin 2	2.21E-01	-3.21E-04	1.71E-04	<0.05
B_C_1	Admin 0	8.66E-03	-2.17E-02	4.33E-03	3.22E-01
	Admin 1	4.27E-01	-1.63E-03	5.24E-04	<0.05
	Admin 2	2.06E-01	-3.21E-04	8.04E+05	<0.05

The Moran scatterplots presented in Figure 10 reveal that for all 3 admin levels, most outlying observations seem to fall into the HH. This implies these regions are potentially hotspots of battles.

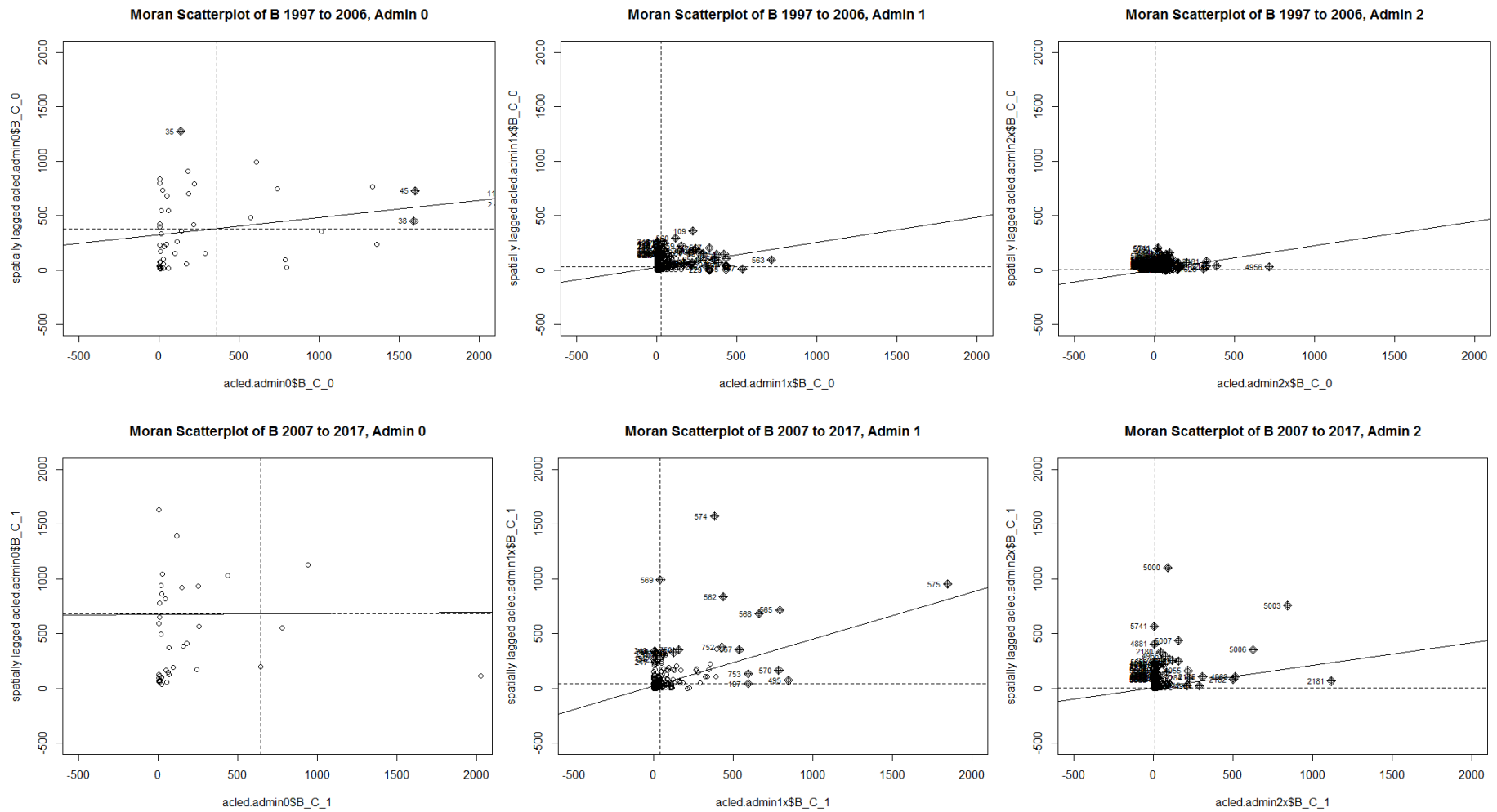


Figure 10 Moran scatterplots, before and after 2007, for admin level 0, 1 and 2

4.5.8 Battles 1997 to 2006 & 2007 to 2017- LISA

The results of significance tests of LISA are presented in Figure 11. Regarding admin levels 1 and 2, there seems to have been a decrease in the number of battle hot spots (judging by the decrease in significant HH regions in Table 9). This decrease was largely driven by a drop in the number of significant HH regions in Angola, Burundi and Sierra Leone. However, the DRC and Somalia have seen a sharp increase in the number of HH regions whereas Uganda seems to have a large number of cold spots and hot spots in period 1 but is dominated by cold spots in period 2.

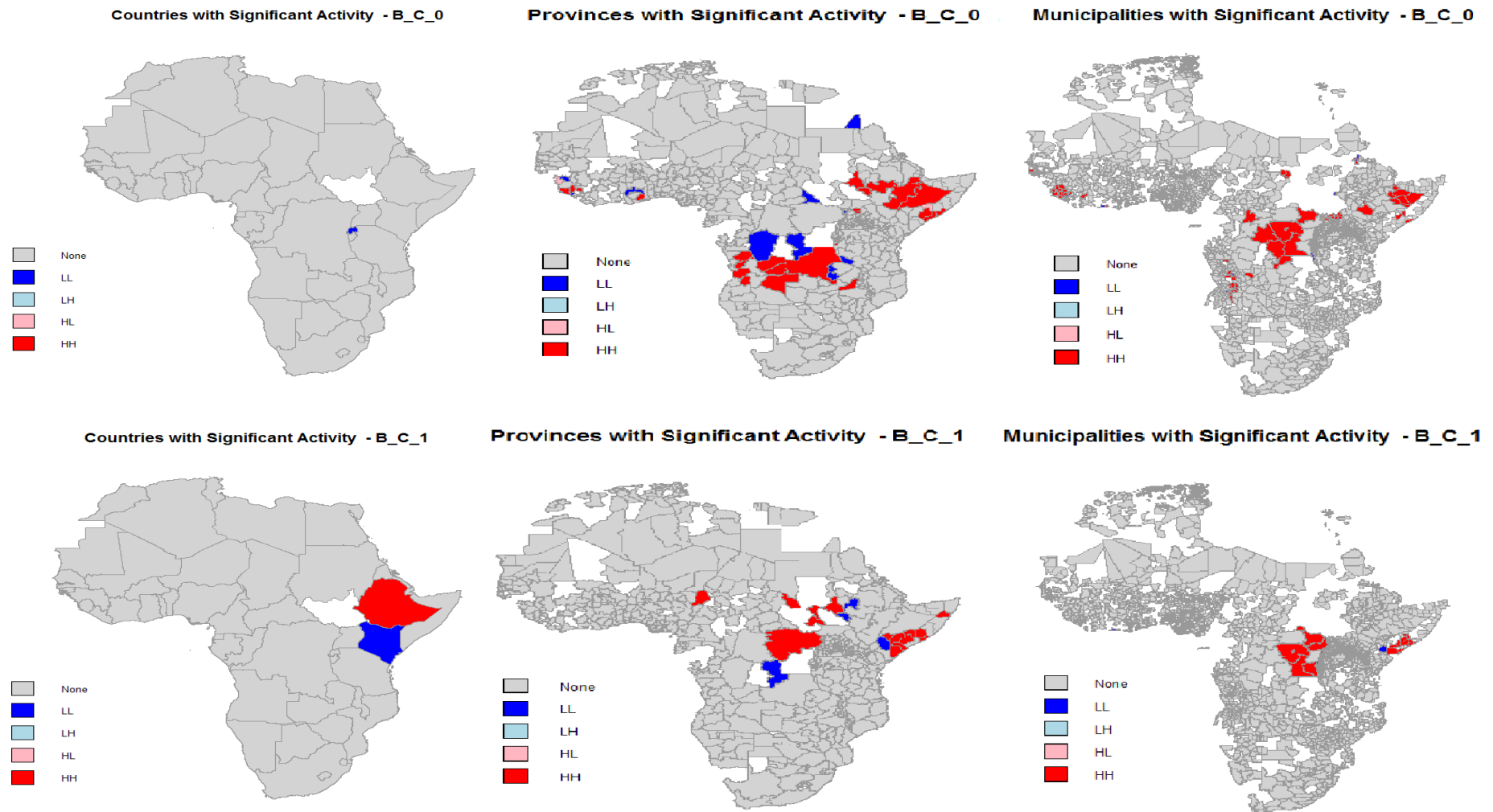


Figure 11 LISA maps, before and after 2007, for admin level 0, 1 and 2

Table 9: Countries where LISA for battles were significant, for period 1 and 2

	ADMIN 1 P1				ADMIN 1 P2			ADMIN 2 P1				ADMIN 2 P2			
	HH	HL	LL	Total	HH	LL	Total	HH	HL	LL	Total	HH	HL	LL	Total
Angola	7			7				16		2	18				
Burundi	3		1	4				11		5	16	2		4	6
Central African Republic			1	1				1			1	2			2
Côte d'Ivoire								2		1	3			1	1
Democratic Republic of the Congo	1		2	3	1	1	2	9			9	6			6
Egypt			1	1											
Eritrea								2		2	4				
Ethiopia	3		3	6		5	5	7		1	8	1			1
Ghana	1		1	2											
Guinea		1	1	2											
Guinea-Bissau								1		1	2				
Kenya	1		1	2	1	1	2							1	1
Liberia								3		1	4				
Nigeria					1		1	1			1				
Republic of Congo								1			1				
Rwanda								2		5	7			5	5
Senegal								3			3				
Sierra Leone	3			3				13			13				
Somalia	3	1		4	8		8	4	1		5	16			16
South Sudan	1			1	3		3								
Sudan					1		1	3			3				
Tanzania			1	1						1	1				
Uganda	2		3	5				19		12	31	1	1	11	13
Zambia	1		1	2											
Total	26	2	16	44	15	7	22	98	1	31	130	28	1	22	51

4.5.9 Battles – Diffusion

To formally determine the nature of the evolution of the spatial patterns for battles through time, diffusion patterns are now considered. To calculate the significance of diffusion, the following is considered:

- Whether there were any changes?
- Whether the change is an increase or decrease?
- If n_1 denotes the number of adjacent increases and n_2 denotes non-adjacent increases, then the increase is significant if $p < 0.05$ where $p = \binom{n_1+n_2}{n_1} (0.5^{n_1+n_2})$ (La Free et al, 2011)

Table 10 summarises the significance of the all the possible diffusion for battles for all 3 admin levels. It is worth nothing that only the non-adjacent decreases in admin level 2 were significant since $p = \binom{5}{5}(0.5^5) = 0.03125$. This implies a hierarchical decrease which is a decrease in the number of battles that is not due to the activity in the surrounding regions (global decrease).

Table 10: Significant diffusion patterns for battles

B_C	ADMIN 0		ADMIN 1		ADMIN 2	
	Count	Total	Count	Total	Count	Total
Stationary	0	0	8	9	30	35
Change	0	0	1	9	5	35
Increases	0	0	0	1	0	5
Decreases	0	0	1	1	5	5
Non-Adjacent Increases	0	0	0	0	0	0
Adjacent Increases	0	0	0	0	0	0
Non-Adjacent Decreases	0	0	1	1	5	5
Adjacent Decreases	0	0	0	1	0	5

4.5.10 Summary of Riots and Protests from 1997 to 2017

Summary statistics for the number of riots and protests (RP) that occurred in the period from 1997 to 2017 are given in Table 11. There were over 41 000 incidents of RP in general and a right skew at all admin levels when the position of the mean is considered relative to the median.

Table 11: Summary statistics of riots and protests

RP_C	Total	Average	Minimum	Median	Maximum
Admin 0	41 204	838	9	336	7 853
Admin 1	41 204	55	0	14	2 167
Admin 2	41 204	10	0	1	1 727

4.5.11 Chloropleth maps for Riots and Protests from 1997 to 2017

For the three admin levels in the period from 1997 to 2017, the number of RP is spatially distributed as shown in Figure 12. The maps show Egypt, Nigeria, South Africa and Ethiopia having high incidents of RP and possible evidence of clustering at admin levels 1 and 2 for these countries.

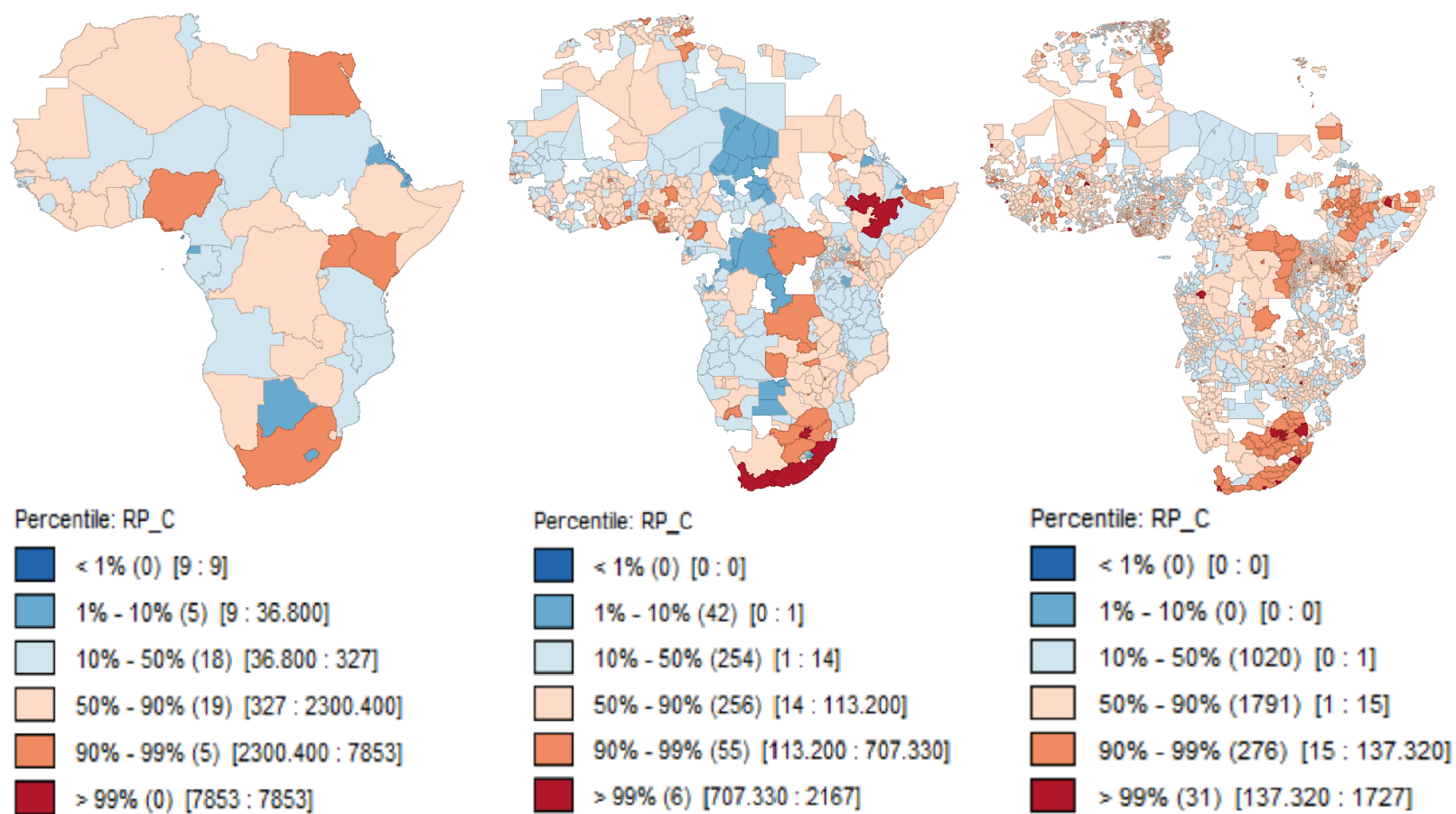


Figure 12 Percentile maps of riots and protests for admin levels 0, 1 and 2

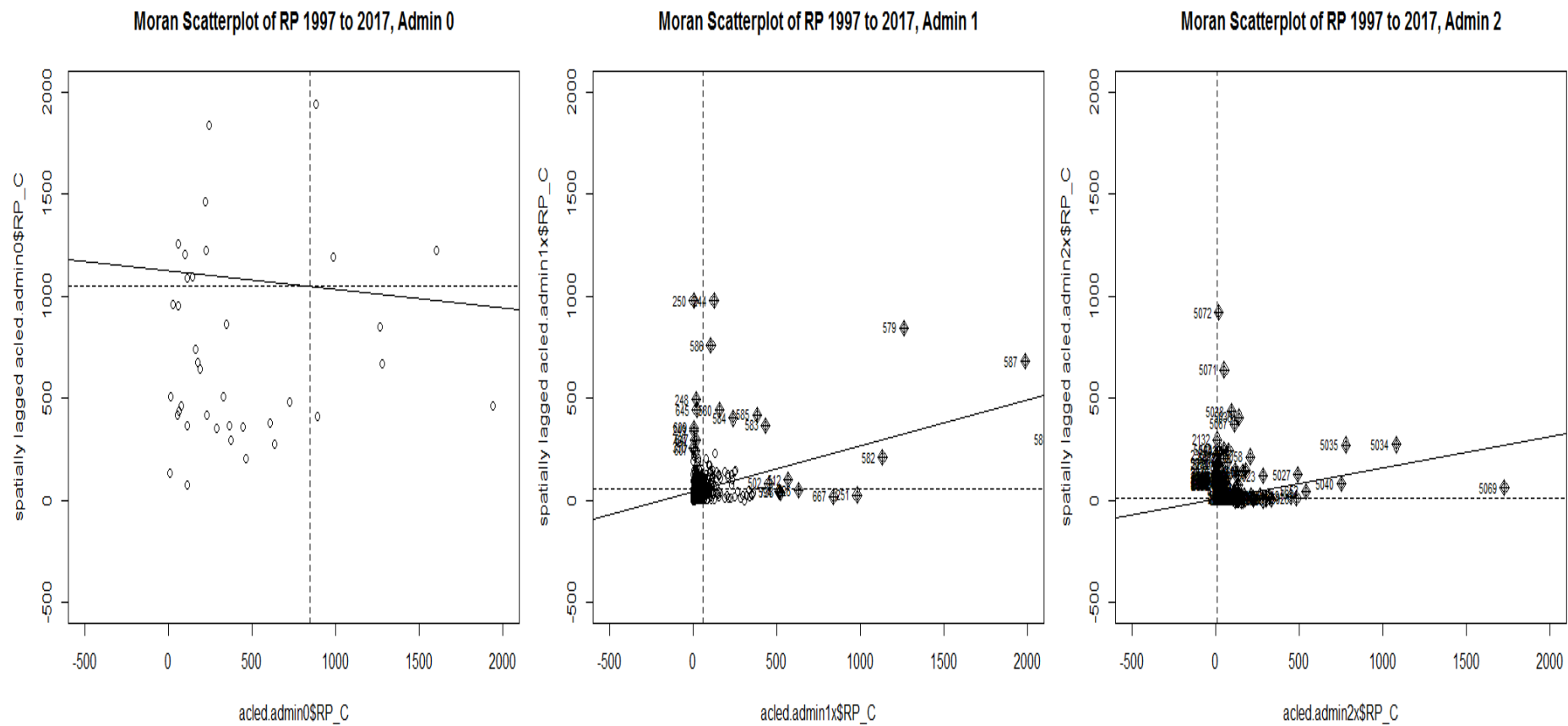
4.5.12 Riots and Protests from 1997 to 2017 - Global Moran's I

Results of the GSA by way of Moran's I determined for all three admin levels are presented in Table 12. While the Moran's I value are close to zero, there is evidence of significant positive GSA at admin levels 1 and 2 which implies some degree of clustering of regions with similar levels of riots and protests. The negative GSA for admin level 0, had it been significant, would have implied the significant presence of clusters of regions with dissimilar levels of riots and protests.

Table 12: Global Moran's I for riots and protests

RP_C	Moran's I	Expectation	Variance	p-value
Admin 0	-9.15E-02	-2.17E-02	7.67E-03	7.87E-01
Admin 1	2.25E-01	-1.63E-03	6.94E-04	<0.05
Admin 2	1.53E-01	-3.21E-04	1.66E-04	<0.05

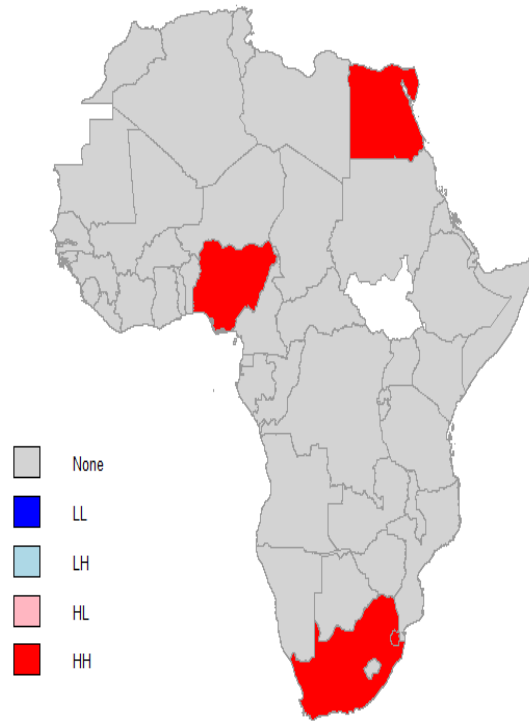
In Figure 13, the Moran scatterplot for admin levels 1 and 2 only show potentially outlying behaviour in the HH quadrant with visibly more areas in the HH quadrants than was seen for battles.



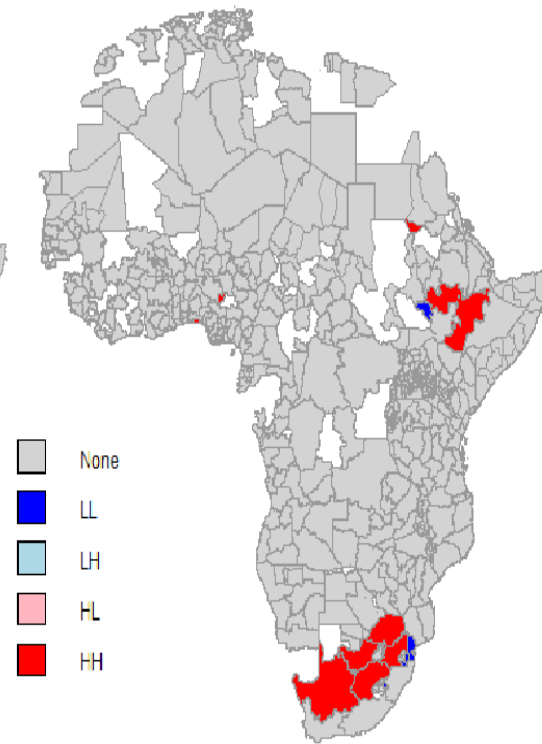
4.5.13 Riots and Protests from 1997 to 2017 - LISA

When the significance of the outlying behaviour described in 4.5.12 is tested, the results (Figure 14 and Table 13) show two countries (Ethiopia and South Africa) as significant hotspots of RP at admin levels 1 and 2. South Africa particularly appears to be a hotbed of riot and protest activity. This can be seen from the fact that it qualifies as HH at admin levels 1 (26 out of 67 riot and protest hotspots are South African regions at admin level 1). From the ACLED data, the protests seem to centre on service delivery.

Countries with Significant Activity - RP_C



Provinces with Significant Activity - RP_C



Municipalities with Significant Activity - RP_C

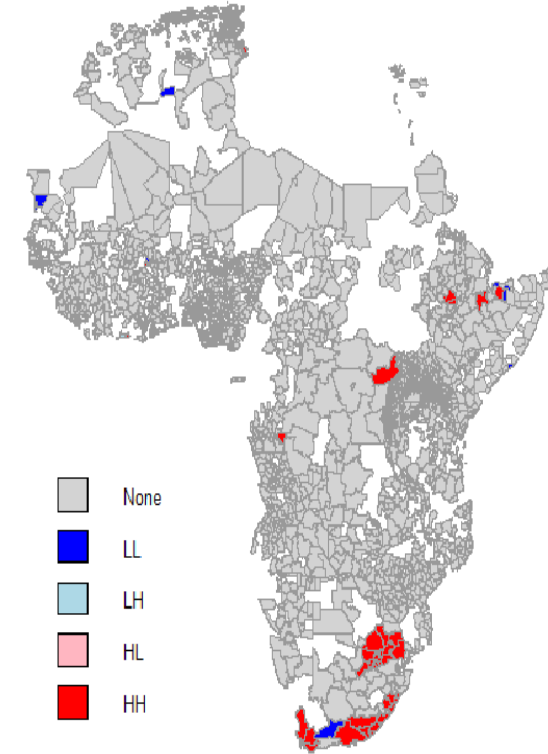


Figure 14 LISA maps of riots and protests for admin level 0, 1 and 2

Table 13: Countries where LISA for riots and protest were significant, for period 1 and 2

	ADMIN 1			ADMIN 2			
	HH	LL	Total	HH	LH	LL	Total
Algeria				1		5	6
Burkina Faso				1		1	2
Burundi				1			1
Central African Republic				1			1
Côte d'Ivoire				1	1		2
Democratic Republic of the Congo				2			2
Egypt				3		2	5
Ethiopia	2	3	5	3		1	4
Guinea				1		1	2
Kenya	1		1	4		3	7
Lesotho		1	1				
Mali				1			1
Mauritania				1		1	2
Morocco				1		1	2
Mozambique		1	1				
Namibia				1			1
Nigeria	2		2	2			2
Republic of Congo				1			1
Senegal				2			2
Somalia				2		4	6
South Africa	5		5	26		1	27
Sudan	1		1				
Swaziland		2	2				
Tunisia	1	2	3	6		2	8
Uganda				1			1
Zambia				2			2
Zimbabwe	1		1	3			3
Grand Total	13	9	22	67	1	22	90

4.5.14 Summary of Riot and Protests from 1997 to 2006 and from 2007 to 2017

Focusing on the periods from 1997 to 2006 and from 2007 to 2017, Table 14 provides summary statistics for the number of RPs in the two time periods for all admin levels. There were almost five times the number of events in the second period spanning 2007 to 2017 compared to the period from 1997 to 2006. There is also right skewness in the data for the three admin levels.

Table 14: Summary statistics of riots and protests, for period 1 and 2

		Total	Average	Minimum	Median	Maximum
RP_C_0	Admin 0	5 432	114	4	53	778
	Admin 1	5 432	7	0	1	321
	Admin 2	5 432	1	0	0	210
RP_C_1	Admin 0	35 772	730	5	294	7 075
	Admin 1	35 772	48	0	11	1 846
	Admin 2	35 772	8	0	1	1 549

4.5.15 Chloropleth Maps for Riots and Protests from 1997 to 2006 and from 2007 to 2017

For all admin levels during the periods from 1997 to 2006 and from 2007 to 2017, RP seem to occur across Africa for both time periods, however, the number of RPs is spatially distributed such that the highest number of incidents occurred in the same three countries identified using the aggregated data for the period 1997 to 2017. These countries are Egypt, Ethiopia, Nigeria and South Africa. Figure 15 also reveals some evidence of local clustering at all admin levels particularly for Ethiopia and South Africa.

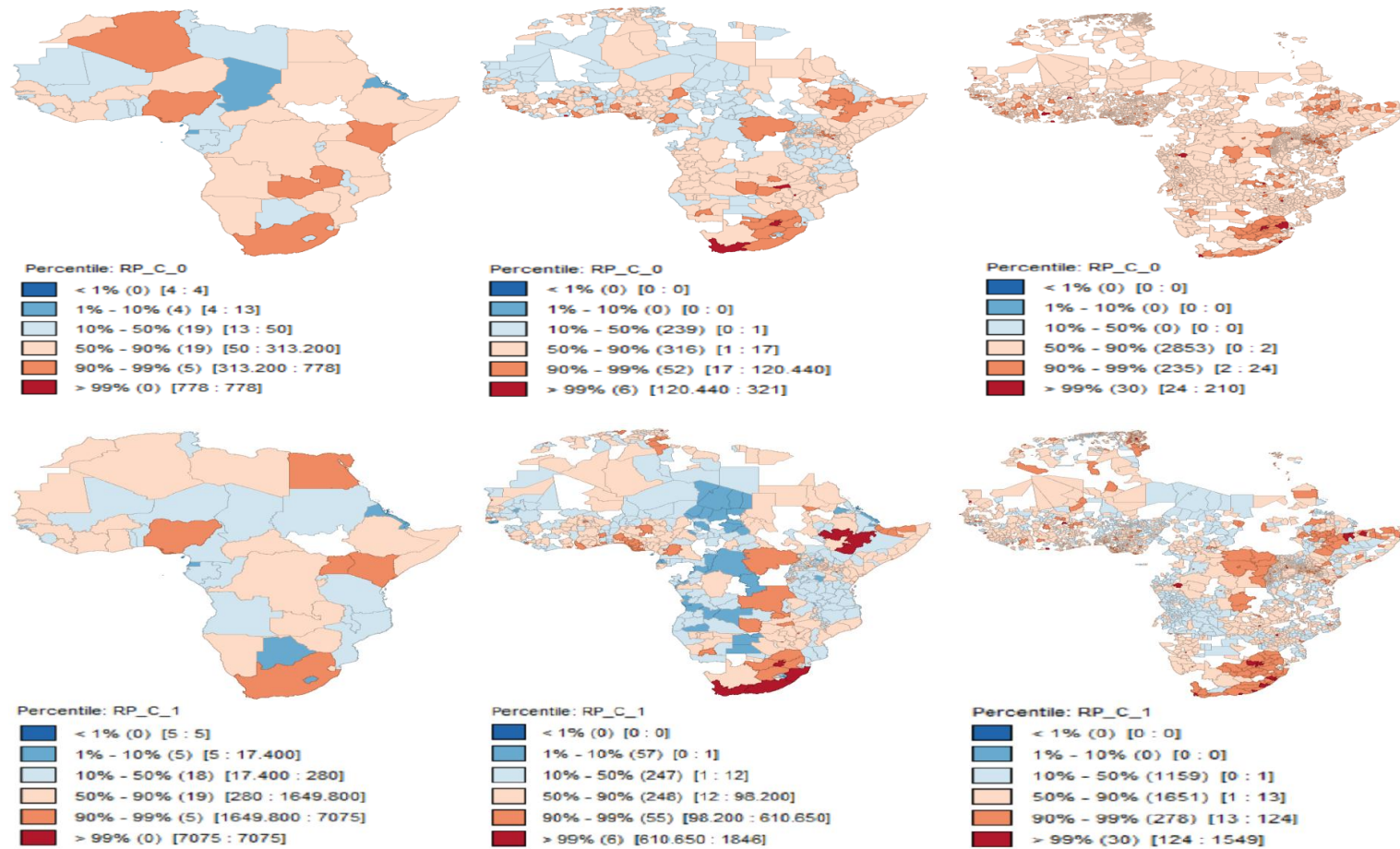


Figure 15 Percentile maps of riots and protests, before and after 2007, for admin level 0, 1 and 2

4.5.16 Riots and Protests from 1997 to 2006 and from 2007 to 2017- Global Moran's I

Using Moran's I, GSA can be determined for all admin levels in both time periods (presented in Table 15). The results are similar to those presented using the aggregate data for time period from 1997 to 2017. In summary, there is significant positive GSA at both admin levels 1 and 2 in both time periods which implies clustering of regions with similar levels of riots and protests.

Table 15: Global Moran's I fro riots and protests, for period 1 and 2

		Moran's I	Expectation	Variance	p-value
RP_C_0	Admin 0	-8.74E-02	-2.17E-02	8.69E-03	7.59E-01
	Admin 1	8.43E-02	-1.63E-03	7.14E-04	<0.05
	Admin 2	8.42E-02	-3.21E-04	1.81E-04	<0.05
RP_C_1	Admin 0	-9.88E-02	-2.17E-02	7.71E-03	8.10E-01
	Admin 1	2.42E-01	-1.63E-03	6.94E-04	<0.05
	Admin 2	1.62E-01	-3.21E-04	1.62E-04	<0.05

From the Moran scatterplots in Figure 16, it appears potentially outlying regions are of the HH variety at admin levels 1 and 2. The associated significance testing results are provided in section 4.5.17.

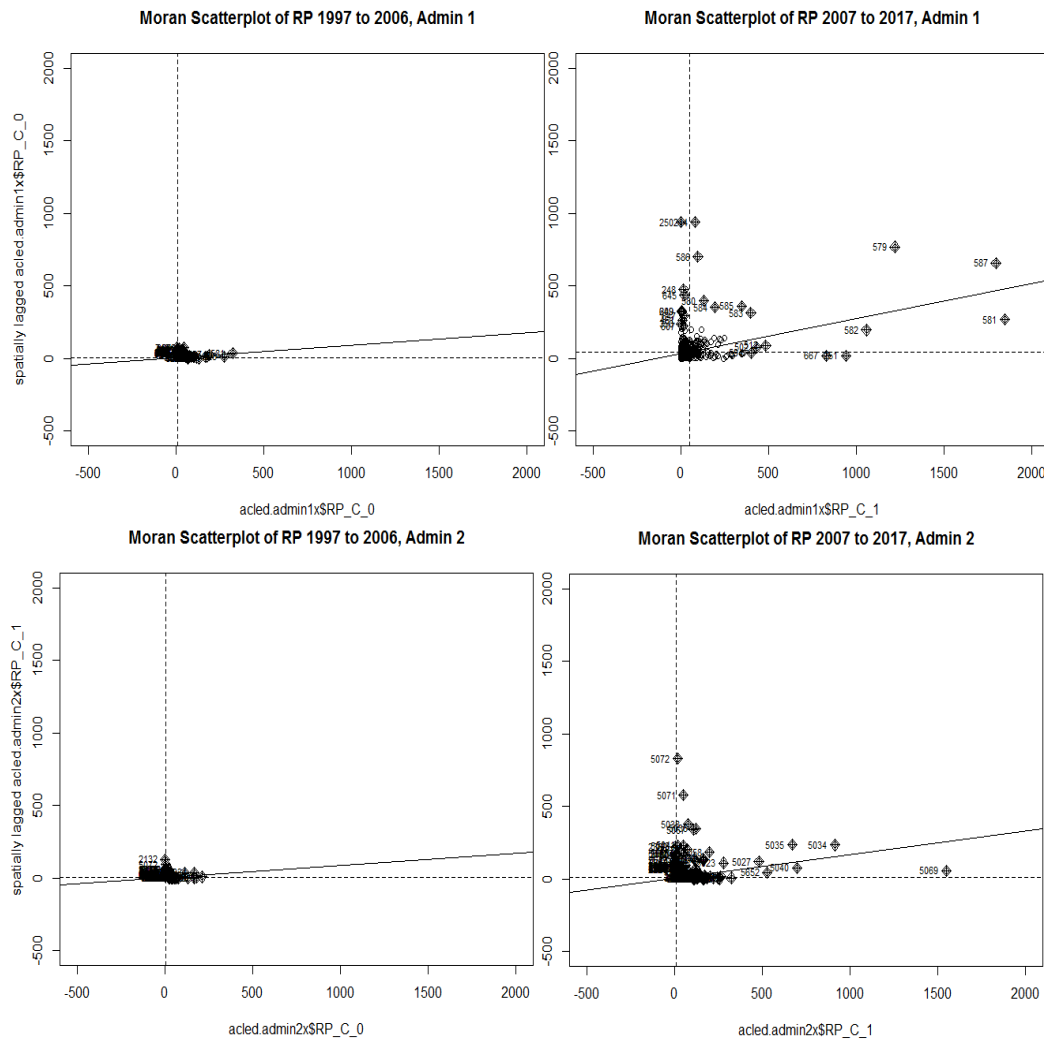
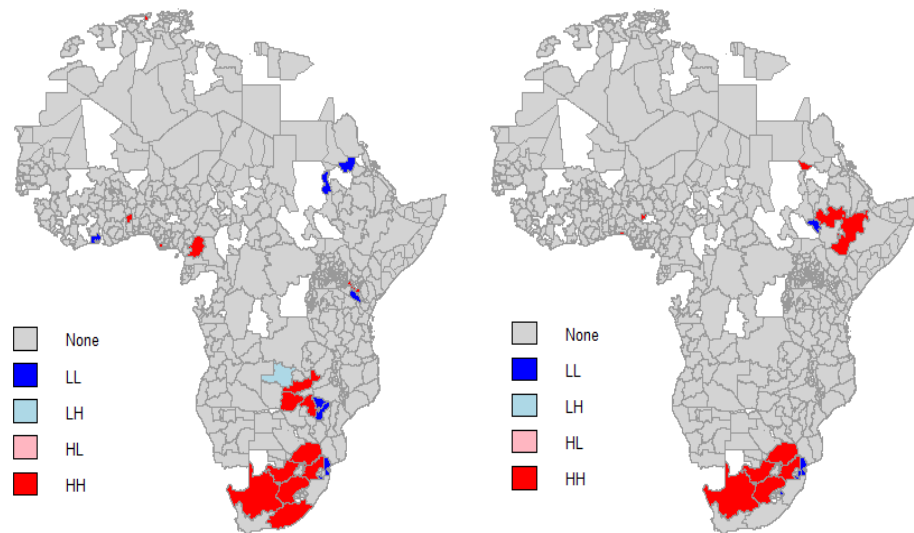


Figure 16 Moran scatterplots of riots and protests, before and after 2007, for admin level 1 and 2

4.5.17 Riots and Protests from 1997 to 2006 and from 2007 to 2017- LISA

There were more areas with significant LISA for RP at admin levels 1 and 2 than there was for Battles in both time periods. As before, South Africa seems to have an unusually high number of RP hot spots, regardless of time period. In fact, the number of hot spots in South Africa doubled between periods 1 and 2 (Figure 17 and Table 16). Next to South Africa, Kenya and Ethiopia had more significant hotspots than other countries on the continent for the two time periods (Table 16).

Provinces with Significant Activity - RP_C_0 Provinces with Significant Activity - RP_C_1



Municipalities with Significant Activity - RP_C_0 Municipalities with Significant Activity - RP_C_1

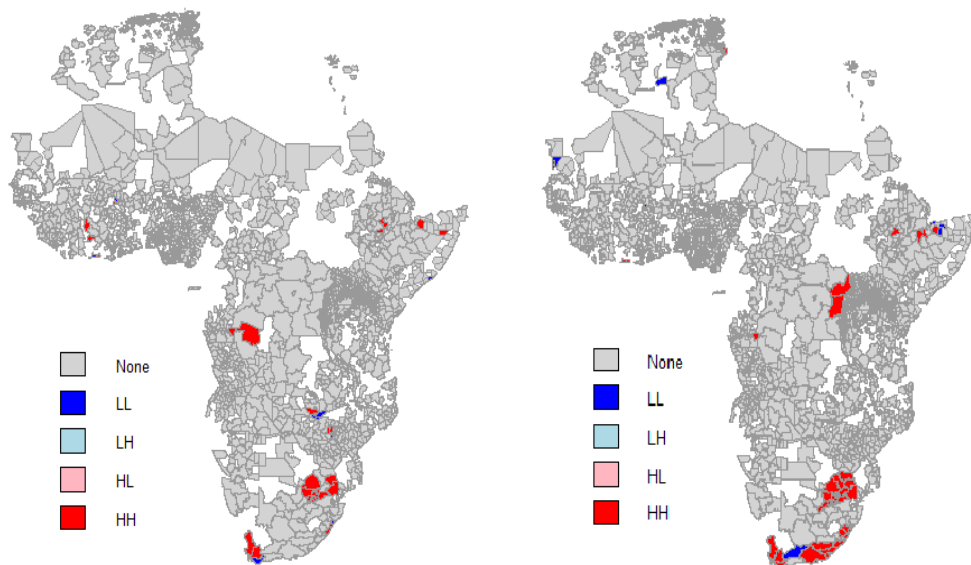


Figure 17 LISA maps of riots and protests, before and after 2007, for admin levels 1 and 2

Table 16: Countries where LISA for riots and protests is significant, for period 1 and 2

	ADMIN 1 P1				ADMIN 1 P2			ADMIN 2 P1			ADMIN 2 P2		
	HH	LH	LL	Total	HH	LL	Total	HH	LL	Total	HH	LL	Grand Total
Algeria	1			1				2	6	8		4	4
Burkina Faso	1			1				1	1	2	1	1	2
Burundi											1		1
Cameroon	1			1									
Central African Republic	1			1				1	1	2	1		1
Côte d'Ivoire			1	1				3	1	4	2		2
Democratic Republic of the Congo								3		3	3		3
Egypt									1	1	3	3	6
Ethiopia	1		1	2	2	3	5	4		4	2	1	3
Guinea								1	1	2	1	1	2
Guinea-Bissau								1	1	2			
Kenya	3		1	4				9	3	12	1	1	2
Lesotho						1	1						
Mauritania								1		1	1	1	2
Morocco									1	1	1	1	2
Mozambique			1	1		1	1	1	1	2			
Namibia								1		1	1		1
Niger								1		1			
Nigeria	1			1	2		2	1		1	2	1	3
Republic of Congo								1		1	1		1
Senegal								1	1	2	2		2
Somalia	1			1				3	1	4	2	3	5
South Africa	6			6	5		5	13	2	15	27	1	28
Sudan			2	2	1		1						
Swaziland			1	1		2	2						
Tanzania								1	1	2			
Togo	1			1									
Tunisia					1	2	3				6	2	8
Uganda								1		1	1		1
Zambia	2	1		3				5	3	8	1		1
Zimbabwe	1		2	3				3	2	5	2		2
Total	20	1	9	30	11	9	20	58	27	85	62	20	82

4.5.18 Riots and Protests – Diffusion

To formally determine the nature of the evolution of spatial patterns for riots and protests through time, diffusion patterns were now considered. Table 17 summarises the significance of the all the possible diffusion patterns for riots and protests for all three admin levels. With regards to protests there no significant diffusion patterns. The closest result was a non-adjacent increase, where $p = \binom{4}{4}(0.5^4) = 0.0625$. The results suggest that despite the almost 5 times difference in the number of cases in the second period (2007 to 2017) relative to the first period (1997 to 2016), the incidents occurred in virtually the same areas. The real difference might be revealed if fatalities were considered.

Table 17: Significant diffusion patterns for riots and protests

RP_C	ADMIN 0		ADMIN 1		ADMIN 2	
	Count	Total	Count	Total	Count	Total
Stationary	1	1	9	9	31	36
Change	0	1	0	9	5	36
Increases	0	0	0	0	4	5
Decreases	0	0	0	0	1	5
Non-Adjacent Increases	0	0	0	0	4	4
Adjacent Increases	0	0	0	0	0	4
Non-Adjacent Decreases	0	0	0	0	1	1
Adjacent Decreases	0	0	0	0	0	1

4.5.19 Summary of Violence against Civilians from 1997 to 2017

The summary statistics with respect to violence against civilians (VAC) that occurred in the period from 1997 to 2017 are presented in Table 18. The total number of VAC is 43 941. The data are right skewed as with the two other variables previously considered.

Table 18: Summary statistics for violence against civilians

VAC_C	Total	Average	Minimum	Median	Maximum
Admin 0	43 941	829	5	269	5 851
Admin 1	43 941	59	0	12	1 879
Admin 2	43 941	10	0	1	1 879

4.5.20 Chloropleth Maps for Violence against Civilians from 1997 to 2017

For all three admin levels in the period from 1997 to 2017, the DRC, Nigeria, Sudan and Zimbabwe have the most number of VAC. From the maps in Figure 18, there is evidence of local clustering across all three admin levels for regions in DRC.

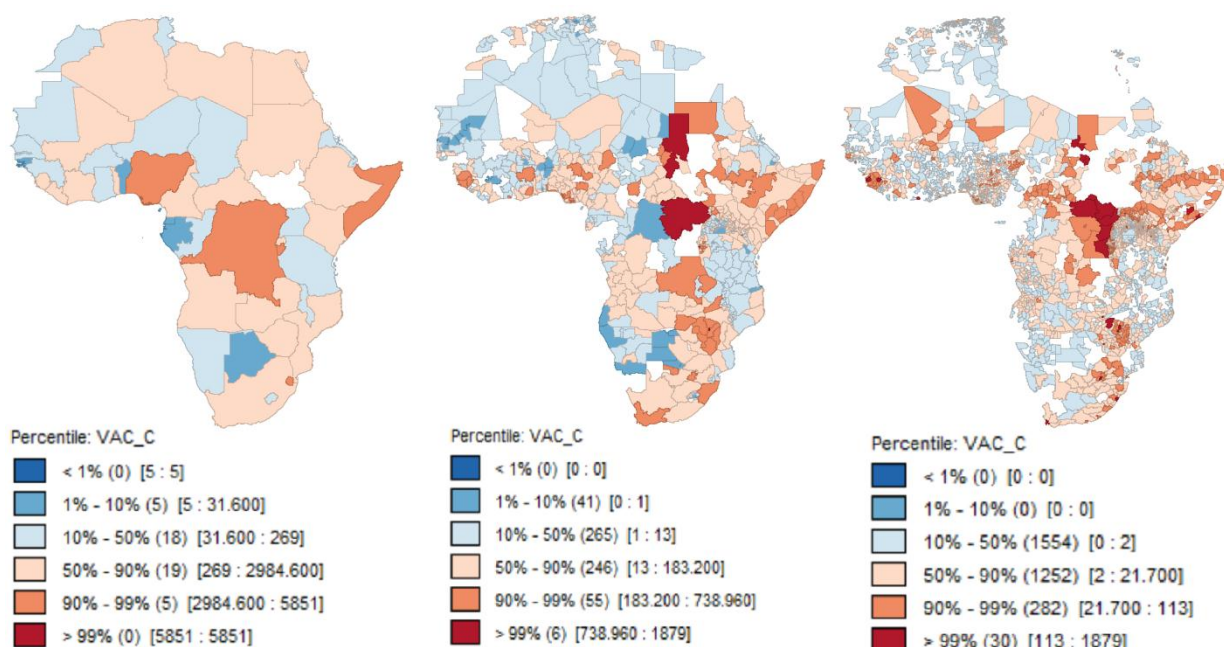


Figure 18 Percentile maps of violence against civilians at admin level 0, 1 and 2

4.5.21 Violence against Civilians from 1997 to 2017- Global Moran's I

The GSA results of all admin levels are presented in Table 19 and indicate that despite the Moran's I values being close to zero, there is evidence of positive GSA at admin levels 1 and 2, which implies a low degree of clustering of regions with similar levels of violence against civilians.

Table 19: Global Moran's I for violence against civilians

VAC_C	Moran's I	Expectation	Variance	p-value
Admin 0	2.81E-02	-2.17E-02	9.40E-03	3.04E-01
Admin 1	2.94E-01	-1.63E-03	7.47E-04	<0.05
Admin 2	1.71E-01	-3.21E-04	1.53E-04	<0.05

Looking at the Moran scatterplots in Figure 19, there are regions with potentially outlying behaviour of the HH variety in admin levels 1 and 2 respectively

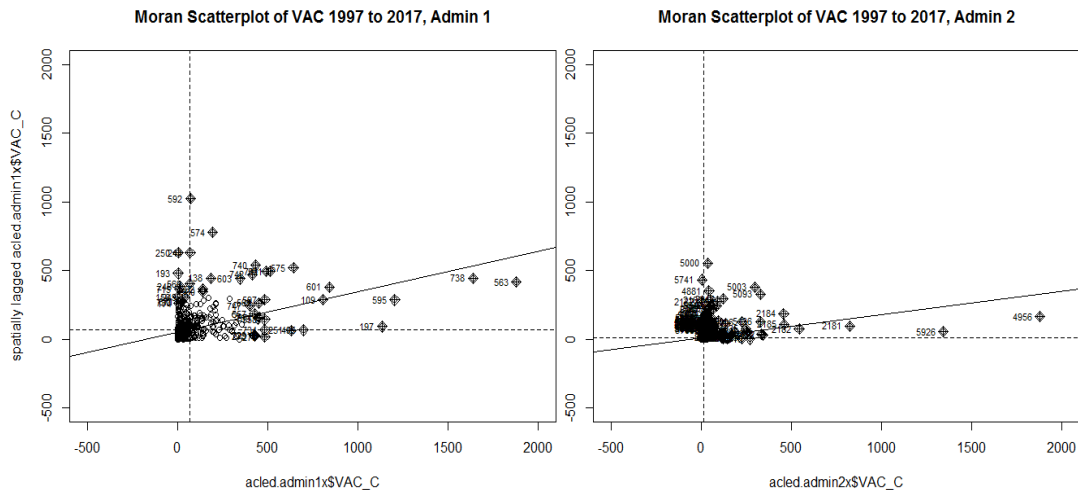


Figure 19 Moran scatterplots of violence against civilians at admin levels 1 and 2

4.5.22 Violence against Civilians from 1997 to 2017- LISA

Significance testing for LISA presented in Figure 20 and Table 20 yield the following results: there were no admin level 0 regions with significant VAC activity. Uganda is an odd mix of hot spots and cold spots (which dominate) at the second admin level. Somalia, Sudan and Zimbabwe consistently have several hot spots of VAC events. The violence against civilians in Zimbabwe was possibly due to the political violence that followed the highly contested elections in 2008 that the Human Rights Watch described as “deeply flawed” (CNN, 23 March 2008).

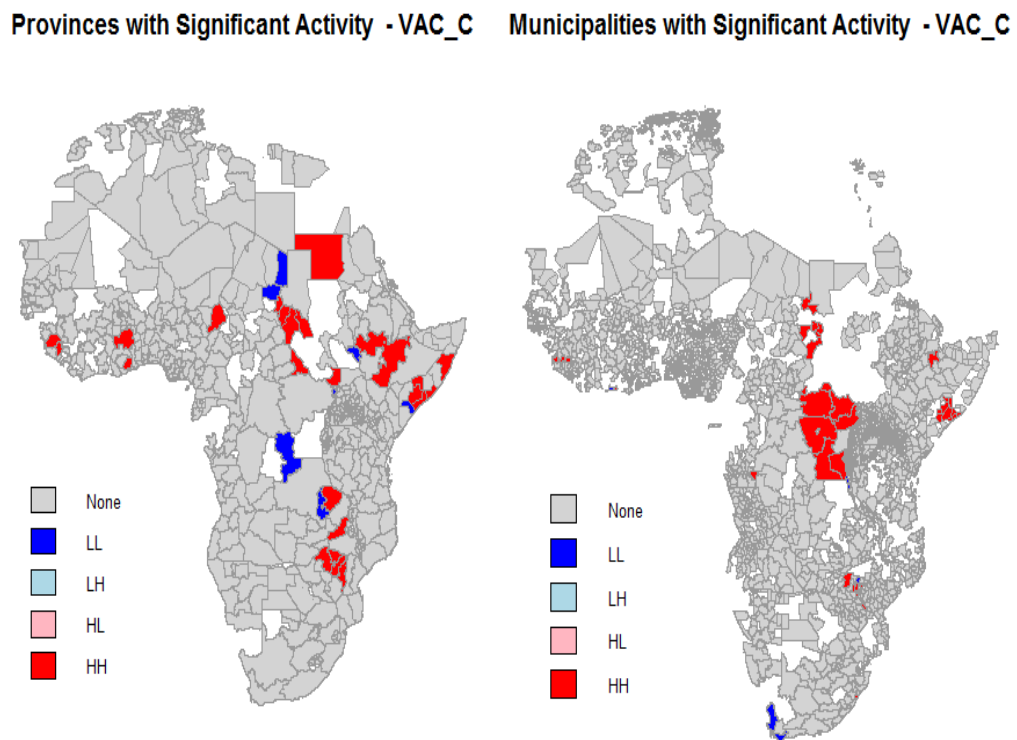


Figure 20 LISA maps for violence against civilians for admin levels 1 and 2

Table 20: Countries where LISA for violence against civilians was significant

	ADMIN 1				ADMIN 2		
	HH	HL	LL	Total	HH	LL	Total
Burundi	1			1	7	4	11
Central African Republic	1			1	4		4
Chad			2	2			
Côte d'Ivoire					1	1	2
Democratic Republic of the Congo			1	1	8		8
Ethiopia	1	1	3	5	1	1	2
Ghana	2			2			
Kenya					1		1
Nigeria	1			1	1		1
Rwanda					3	4	7
Sierra Leone	3			3	5		5
Somalia	5		1	6	12		12
South Africa					3	2	5
South Sudan	1			1			
Sudan	5			5	5		5
Tanzania						1	1
Uganda			2	2	5	13	18
Zambia	2		1	3			
Zimbabwe	4			4	8	1	9
Grand Total	26	1	10	37	64	27	91

4.5.23 Summary of Violence against Civilians from 1997 to 2006 and from 2007 to 2017

Considering the number of conflict events that occurred during the periods from 1997 to 2006 and from 2007 to 2017. There were more than two times the number of events in the second period spanning 2007 to 2017 compared to the period from 1997 to 2006 (Table 21). Comparing the mean to the median also reveals right skewness in the VAC data for all time periods.

Table 21: Summary statistics for violence against civilians, for period 1 and 2

		Total	Average	Minimum	Median	Maximum
VAC_C_0	Admin 0	13 602	284	2	70	2 431
	Admin 1	13 602	18	0	3	1 056
	Admin 2	13 602	3	0	-	895
VAC_C_1	Admin 0	30 339	596	3	131	4 730
	Admin 1	30 339	40	0	6	1 338
	Admin 2	30 339	7	0	1	1 338

4.5.24 Chloropleth Maps for Violence against Civilians from 1997 to 2006 and from 2007 to 2017

For admin level 0 during the periods from 1997 to 2006 and from 2007 to 2017, VAC events are spatially distributed as follows: the maps in Figure 21, show evidence of local clustering across the continent for both periods 1 and 2 which are more pronounced at admin levels 1 and 2.

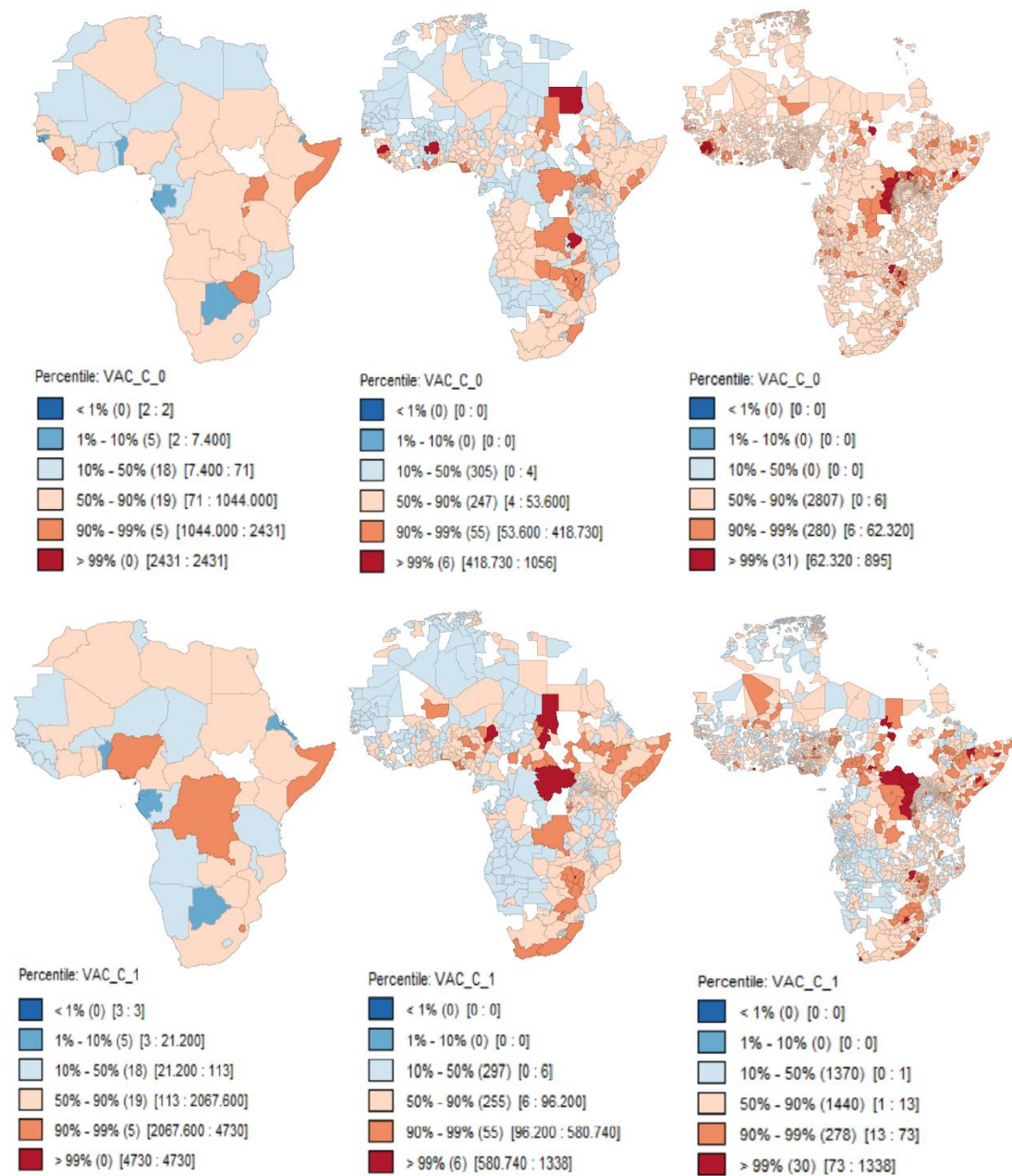


Figure 21 Percentile maps for violence against civilians, before and after 2007, for admin level 0, 1 and 2

4.5.25 Violence against Civilians from 1997 to 2006 and from 2007 to 2017

Global Moran's I

The results of the GSA test by way of Moran's I reveal that as is the case with RP, and the aggregate analysis for VAC, the Moran's I values were close to zero but nonetheless indicate significant clustering at both admin levels 1 and 2 (Table 22).

Table 22: Global Moran's I for violence against civilians, for period 1 and 2

		Moran's I	Expectation	Variance	p-value
VAC_C_0	Admin 0	-2.32E-03	-2.17E-02	8.75E-03	4.18E-01
	Admin 1	2.43E-01	-1.63E-03	7.14E-04	<0.05
	Admin 2	1.48E-01	-3.21E-04	1.37E-04	<0.05
VAC_C_1	Admin 0	4.83E-02	-2.17E-02	9.27E-03	2.33E-01
	Admin 1	3.01E-01	-1.63E-03	7.47E-04	<0.05
	Admin 2	2.02E-01	-3.21E-04	1.49E-04	<0.05

From the Moran scatterplots in Figure 22, there is evidence of potentially outlying regions of the HH regions in all 2 admin levels with significant Moran's I for both time periods.

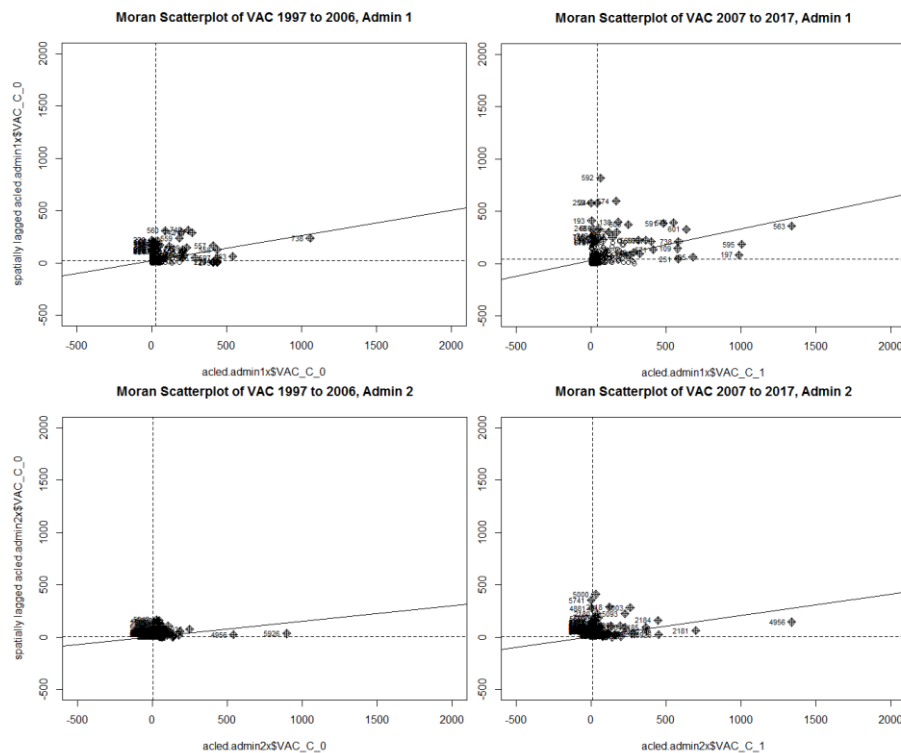


Figure 22 Moran scatterplots of violence against civilians, before and after 2007, at admin level 1 and 2

4.5.26 Violence against Civilians from 1997 to 2006 and from 2007 to 2017- LISA

At both admin levels 1 and 2 and for both time periods, Burundi, Sierra Leone, Somalia, Sudan and Zimbabwe consistently have several hot spots of VAC events (Figure 23 and Table 23). Uganda has an interesting pattern that is worth noting. When it had several significant hot spots violence against civilians in admin level 2, they also occurred alongside several cold spots. This hints at the fact that there are two parts of Uganda, one which is afflicted with conflict events and another which is peaceful.

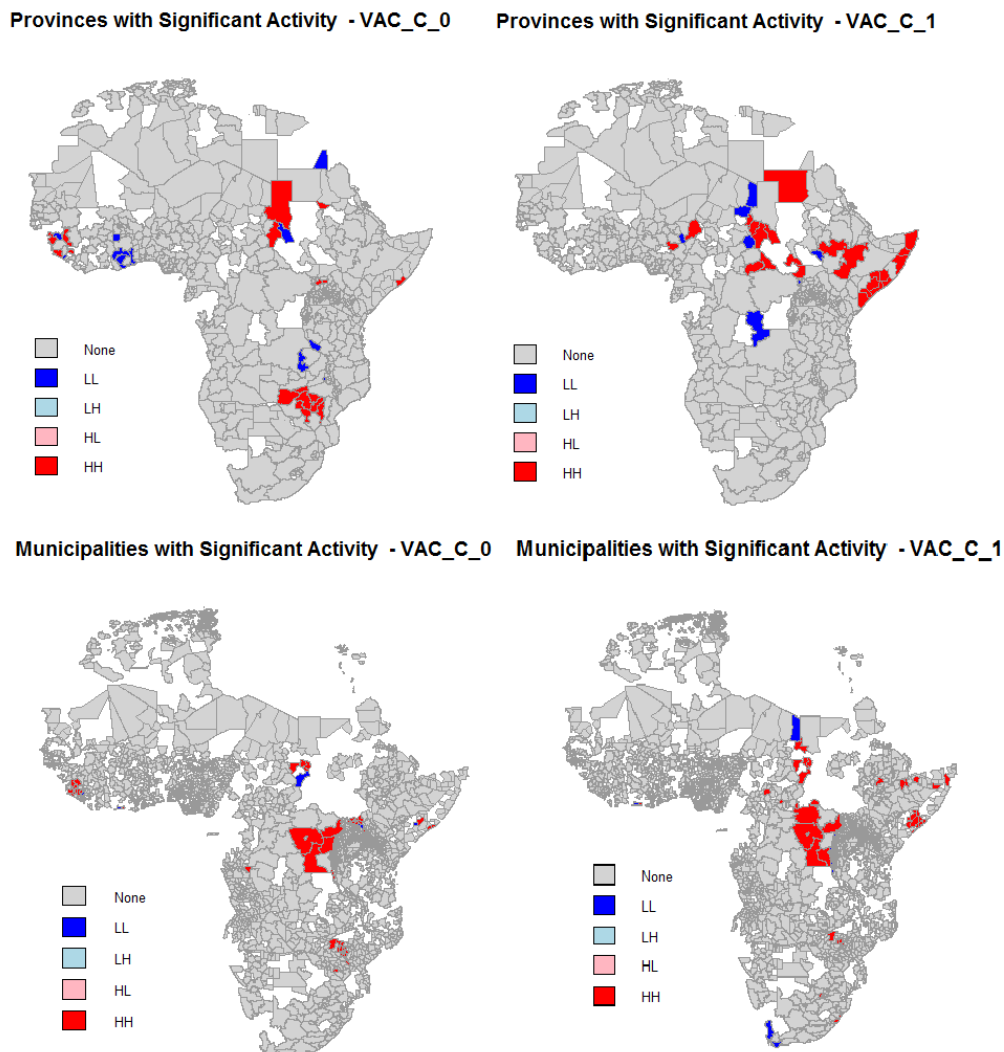


Figure 23 LISA maps of violence against civilians, before and after 2007, at admin level 1 and 2

Table 23: Countries where LISA for violence against civilians was significant, for period 1 and 2

	ADMIN 1 P1			ADMIN 1 P2			ADMIN 2 P1			ADMIN 2 P2		
	HH	LL	Total	HH	LL	Total	HH	LL	Total	HH	LL	Total
Burundi	1		1	1		1	11	1	12	6	4	10
Central African Republic				2	1	3	1		1	6		6
Chad					2	2					1	1
Côte d'Ivoire							1	1	2	1	1	2
Democratic Republic of the Congo					1	1	6		6	6		6
Egypt		1	1									
Ethiopia				1	4	5				3	1	4
Ghana		6	6									
Guinea	2	1	3									
Kenya							1		1			
Liberia	1	1	2					2	2			
Malawi		1	1									
Nigeria				2	1	3	2		2	1		1
Republic of Congo							1		1			
Rwanda							5	2	7	2	5	7
Sierra Leone	2		2				11		11			
Somalia	1		1	10		10	3	1	4	15		15
South Africa										5	3	8
South Sudan				2		2						
Sudan	3	1	4	5		5	3	1	4	5		5
Tanzania		1	1				1		1		1	1
Uganda	2	1	3		3	3	22	11	33	2	15	17
Zambia	2	1	3									
Zimbabwe	5		5	1		1	14		14	4	1	5
Total	19	14	33	24	12	36	82	19	101	56	32	88

4.5.27 Violence Against Civilians – Diffusion

To formally determine the nature of the evolution of the spatial patterns for riots and protests through time, diffusion patterns are now considered. Table 24 summarises the significance of the all the possible diffusion patterns for riots and protests for admin levels 1 and 2. The non-adjacent decreases in admin level 2 were significant since $p = \binom{12}{12}(0.5^{12}) = 0.000244141$. This implies a hierarchical decrease which is a decrease in the number of VAC events that is not due to the activity in the surrounding regions (a global decrease).

Table 24: Significant diffusion patterns for violence against civilians

VAC_C	ADMIN 0		ADMIN 1		ADMIN 2	
	Count	Total	Count	Total	Count	Total
Stationary	0	0	4	5	35	49
Change	0	0	1	5	14	49
Increases	0	0	1	1	2	14
Decreases	0	0	0	1	12	14
Non-Adjacent Increases	0	0	1	1	2	2
Adjacent Increases	0	0	0	1	0	2
Non-Adjacent Decreases	0	0	0	0	12	12
Adjacent Decreases	0	0	0	0	0	12

4.6 Spatial Autocorrelation – Binary Data

Although counts of events were the focus of this research report, it is worth quickly investigating how the conflict events are distributed when they are considered as binary events. To do this, each spatial unit's count of event was dichotomised based on whether the frequency of occurrences exceeded the global mean or not. By constructing the variables in this way, the result should help in the elimination of the right-skewness that was evident for all conflict types across all admin levels.

The maps in Figures 24, 25 and 26, show results for admin 1 and 2 because these levels showed consistently significant results in the analyses of counts of conflict events. On the left of Figures 24 – 26 are maps for the period 1997 to 2017, whereas the middle maps correspond to the period from 1997 to 2006, and the maps on the right are for the 2007 to 2017 period. In all cases, the East Africa and Central Africa region had significantly high numbers of conflict events, regardless of type. This makes sense in two ways. The first is due to the prominence of the DRC and Somalia in the analyses of count data and, tied to this, is the fact that both countries have been afflicted by civil wars during most, if not the whole, of the period from 1997 to 2017. However, this can only be consistently discerned by looking at the admin levels 1 and 2. This is because both countries are surrounded by countries with lots of conflict events, implying that both these countries and their neighbours would evince conflict activity (which would imply a binary variable value of 1). This masking of high conflict areas at admin level 1 is also apparent with South Africa and Kenya with regards to riots and protests. Kenya is the only country at admin level 0 with a significantly high number of riots and protest. However this is due to its neighbours being less susceptible to riots and protests.

The loss of information that occurs as a result of transforming a variable into binary form is mitigated by the granularity of the admin level that is being dealt with. Hence there is the expectation that the more granular the admin level, the more the results of the binary and count methods will converge.

4.6.1 Battles (Join-Counts)

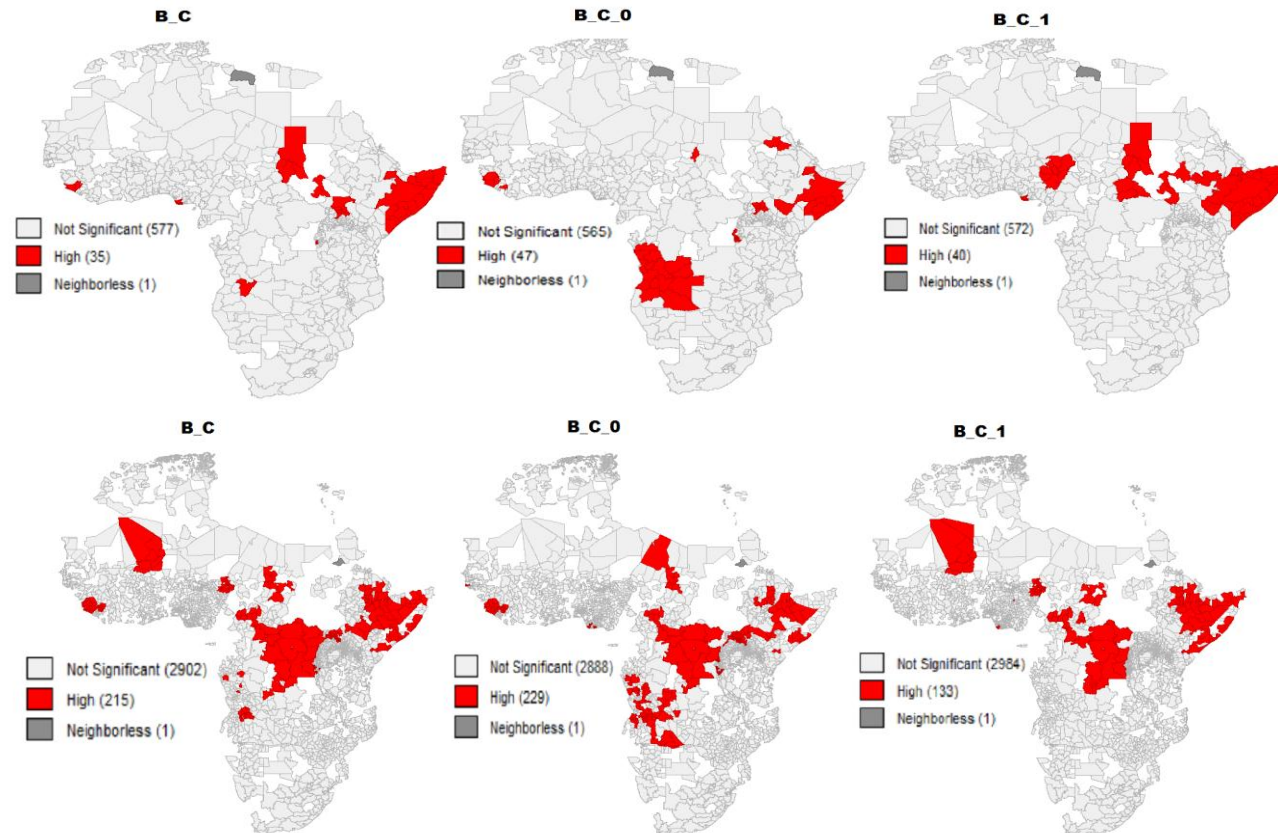


Figure 24 Significance maps of battles (binary) for admin level 1 (top row) and 2 (bottom row)

4.6.2 Riots and Protests (Join-Counts)

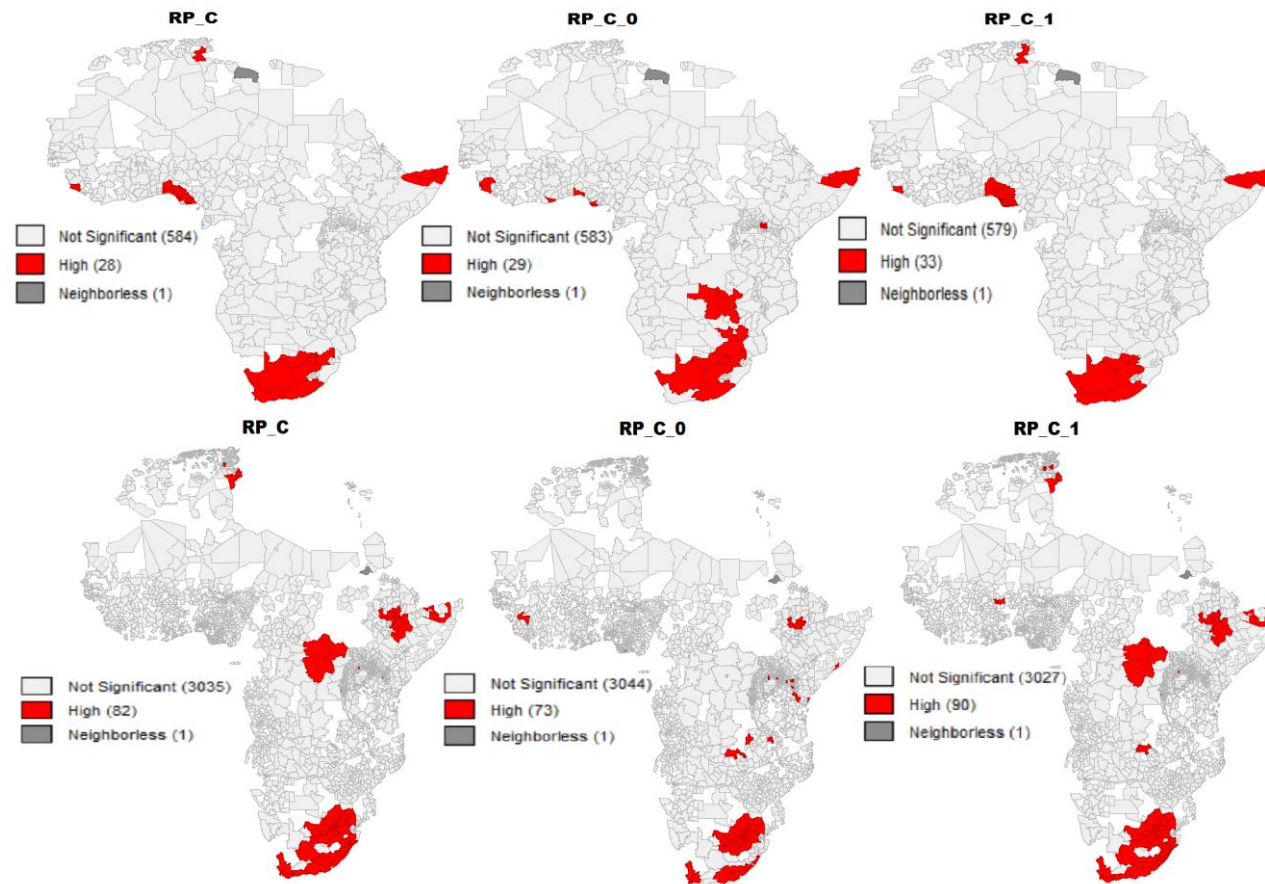


Figure 25 Significance maps of riots and protests (binary) for admin level 1 (top row) and 2 (bottom row)

Violence against Civilians (Join-Counts)

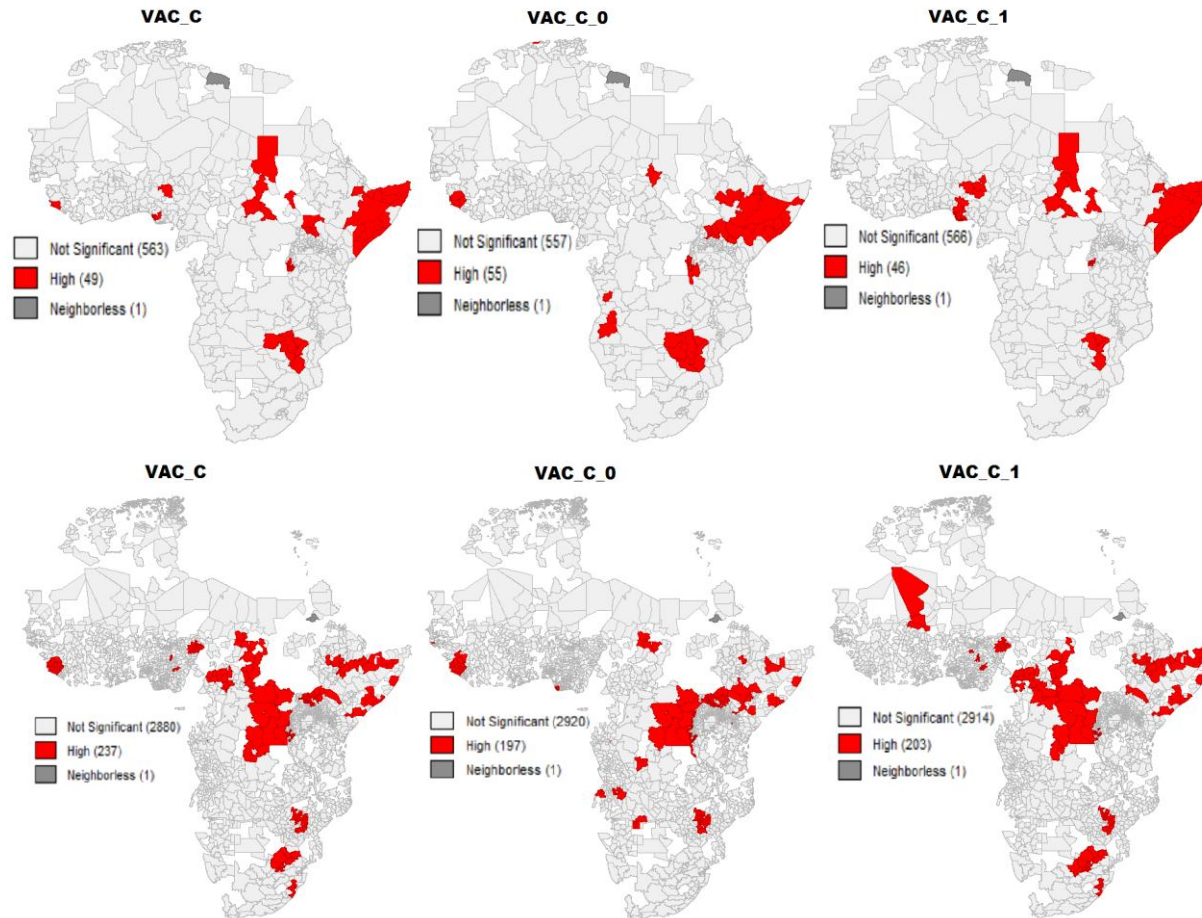


Figure 26 Significance maps of violence against civilians (binary) for admin level 1 (top row) and 2 (bottom row)

5 Summary, Conclusions and Recommendation

5.1 Discussion

This research project explored:

- The global spatial autocorrelation of conflict events in Africa from 1997 to 2017
- Local spatial clusters of conflict events in Africa from 1997 to 2017
- The diffusion/spread of conflicts events over space and time in Africa from 1997 to 2017
- The changes that result from altering the spatial and temporal aggregation level of the data

The examination of all four objectives was centred on count data, with a brief exploration of binary data in order to determine whether the count data results held even after dichotomising the variables of interest.

The most noteworthy result of the analysis is the confirmation of the DRC, Somalia and Sudan as particularly troubled regions in Africa. All three countries have experienced civil wars in the period under consideration (1997 to 2017). This translated to a number of hot spots, regardless of the type of conflict under investigation. In relation to this, it is worth commenting on why both countries' significance with regard to conflict events, was only consistently evident at admin levels 1 and 2. Local SA is, essentially, a measure that quantifies statistical significance of a variable in a region in relation to its neighbouring regions. When tests of local SA encounter a neighbourhood where a variable occurs at a high (or low) level, for a region to qualify as being a hot spot (or cold spot) it would also need to be rather unusual. A good example is South Africa with regards to riots and protests. It was a hot spot at admin level 0, and comprised almost half of the hot spots in admin levels 1 and 2. Whereas Somalia and Sudan, although undoubtedly afflicted with various conflict types, did not individually comprise even a quarter of hot spots in admin levels 1 and 2 at any point.

The relationship between results in all admin levels considered seems to follow the ensuing pattern:

- Admin levels 1 and 2 tend to yield similar results. That is, if a region was significant at admin level 1, then some of its sub-regions will tend to be significant in admin level 2.
- There is no clear-cut relation between admin levels 0, 1 and 2. The only instance where some consistency in results was found was with regard to South Africa when examining riots and protests. As already stated, this consistency likely hinges on the large number of hot spots that South Africa had at admin levels 1 and 2.
- The inconsistency in results between admin level 0, on the one hand, and admin levels 1 and 2, on the other, is a manifestation of the MAUP. As one goes from one level of granularity to the next, discrepancies in results arise.

Closely tied to the MAUP, is the discrepancy in results that arises from dichotomising the variable of interest. When the results of the count and binary data approaches are compared, the following was noted:

- For admin levels 1 and 2, there is a semblance of similarity between the results of the count and binary variables.
- Where inconsistency exists in the results, they seem to be tied to two issues, both which centre around loss of information. When a count variable is dichotomised, anything that falls below the mean is now a 0 and everything else, a 1 – thus the variability that exists below and above the mean is lost. Likewise, when a variable is only considered at the level 1, the variability that exists at the more granular level 2 is lost when one aggregates the data.

Lastly, the examination of diffusion revealed that virtually no significant changes occurred in spatial patterns through time. However, this result should be regarded with caution because diffusion tests are highly dependent on there being numerous significant regions which experienced changes in the nature of their spatial patterns.

Although there were plenty of significant regions in both time periods (1997 to 2006 and 2007 to 2017), the spatial patterns were mostly of the HH variety.

The consequence is that most of the regions exhibit the HH to HH pattern during both periods, which would be classified as stationary and thus would not qualify as either an expansion or relocation. One explanation is that although the conflict events were observed to have increased from period 1 to 2, the evidence suggests that they mostly increased in the same regions, as can be seen by the low number of expansions and relocations.

5.2 Limitations

The following were not considered in the research project:

- Exogenous explanations of the observed patterns of SA as they evolve through space and time. While the modelling of the observed patterns so that the effect of altering the causal factors can be determined, this was not considered in the research report.
- Extrapolation of the spread of the observed patterns into the future was also not considered.
- Furthermore, by using points within a polygon (be it a country, or lower admin level region) rather than area names, the ACLED and GADM data sets can be more effectively merged so that a loss of information does not occur. Countries with non-Roman characters in the names were particularly impacted by merging via names which resulted in a disproportionate loss of information for North African countries.

5.3 Conclusion and Further Work

The analysis of conflict events in Africa yielded 3 noteworthy results, the first of which can be summarised as follows: areal resolution matters. Troubled regions like Somalia and Sudan would remain unnoticed if the analysis had only been restricted to administration level 0. The second result is that although the total number of conflict events has increased, they have not been necessarily spreading to

neighbouring regions. Rather, the areas which previously experienced a conflict type at a significantly high level continued to experience more of that conflict type.

The current work could be expanded in the following ways:

- By considering the number of fatalities in conjunction with the number of events that occurred. This could be done by creating composite variables which takes both variables into account.
- By having more than two time periods in the analysis. The information loss that results from aggregating areal data or dichotomising variables was evident in the report. By only aggregating and examining the data from using only two broad time periods, it is likely that variability that occurs with more granular time periods was lost.
- Exogenous explanations of the observed patterns of SA as they evolve through space and time and possible extrapolation of the spread of the observed patterns into the future can also be considered.

REFERENCES

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, 27(2), pp93-115.

Anselin, L. (1993). *The Moran scatterplot as an ESDA tool to assess local instability in spatial association*. Morgantown, WV: Regional Research Institute, West Virginia University.

Anselin, L., Bera, A., Florax, B., and Yoon, M. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*. 26(1), pp77 – 104.

Anselin, L and Piras, G. (2009). Approaches towards the identification of patterns in violent events, Baghdad, Iraq. Defence Technical Information Centre Document, pp76. Available at: <http://www.dtic.mil/dtic/tr/fulltext/u2/a500195.pdf>. First accessed: November 2016.

Armed Conflict and Location Event Data Project ACLED. (2016). Version 7 data, Available at: <http://www.acleddata.com/data/acled-version-7-1997-2016/>. First accessed: February 2017.

Baudains, P., Johnson, S.D. and Braithwaite, A.M. (2013). Geographic patterns of diffusion in the 2011 London riots. *Applied Geography*, 45, pp211 – 219.

Beardsley, K. (2011). Peacekeeping and the contagion of armed conflict. *The Journal of Politics*, 73(4), pp1051 – 1064.

Behrens, J.T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), pp131.

Bivand R. (2013). *spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*.

R package version 0.5-56. Available at: <http://CRAN.R-project.org/package=spdep>

Brundson, C., Fotheringham, S., and Charlton, M. (2002). Geographical weighted local statistics applied to binary data. *Geographic Information Science*. 2478, pp38 – 50.

Boots, B. (2003). Developing local measures of spatial association for categorical data. *Journal of Geographical System*, 5, pp139 – 160.

Boots, B. (2006). Local configuration measures for categorical spatial data: binary regular lattices. *Journal of Geographical System*, 8, pp1 – 24.

Boots, B and Csillag, F. (2006). Categorical maps, comparisons, and confidence. *Journal of Geographical System*, 8, pp109 – 118.

Cliff, A. D. and J. K. Ord. (1969). “The problem of Spatial autocorrelation.” In *London Papers in Regional Science 1, Studies in Regional Science*, pp25 – 55.

Cohen, J. and Tita, G. (1999). Diffusion in homicide: Exploring a general method for detecting spatial diffusion processes. *Journal of Quantitative Criminology*, 15(4), pp451 – 493.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley and Sons, New York.

De Juan, A. (2012). Mapping political violence: The approaches and conceptual challenges of subnational geospatial analyses of intrastate conflict. *GIGA Working Papers*, 211. Available at: <https://www.giga-hamburg.de/en/publication/mapping-political-violence-%E2%80%93-the-approaches-and-conceptual-challenges-of-subnational>. First accessed: November 2017.

Eck, K. (2012). In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Cooperation and Conflict*, 47(1), pp124 – 141.

Flores, M., Villarreal, A. (2015). Exploring the spatial diffusion of homicides in Mexican municipalities through exploratory spatial data analysis. *Cityscape*, 17(1), pp35 – 49.

Getis, A and Ord, J. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*. 24, pp189 – 206.

Getis, A and Ord, J. (1995). Local spatial autocorrelation statistics: Distribution issues and application. *Geographical Analysis*. 27(4), pp286 – 306.

Getis, A and Ord, J. (2001). Testing for local spatial autocorrelation in the presence of global autocorrelation. *Journal of Regional Science* 41(3), pp411 – 432.

Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37(4), pp491 – 496.

Hegre, H., Østby, G. and Raleigh, C. (2009). Poverty and civil war events: A disaggregated study of Liberia. *Journal of Conflict Resolution*, 53(4), pp598 – 623.

Index, Global Peace. (2016a). Ten Years of Measuring Peace, Institute for Economics and Peace. Available at: http://economicsandpeace.org/wp-content/uploads/2016/06/GPI-2016-Report_2.pdf. First accessed: July 2017.

Index, Global Terrorism (2016b). Measuring and Understanding the Impact of Terrorism, Institute for Economics and Peace. Available at: <http://economicsandpeace.org/wp-content/uploads/2016/11/Global-Terrorism-Index-2016.2.pdf>. First accessed: July 2017.

Kabos, S. and Csillag, F. (2002). The analysis of spatial association on a regular lattice by join-count statistics without the assumption of first-order homogeneity. *Computers & Geosciences*, 28(8), pp901 – 910.

Krug, E.G., Mercy, J.A., Dahlberg, L.L., Zwi, A.B.. (2002). The world report on violence and health. *The lancet*, 360(9339), pp1083-1088.

Kulldorff, M. (1997). A spatial scan statistic. *Communication in Statistics – Theory and Methods*. 26, pp1481 – 1496.

Kulldorff, M. (2007). A spatial scan statistic for ordinal data. *Statistics in Medicine*. 26, pp1594 – 1607.

LaFree, G., Dugan, L., Xie, M. and Singh, P. (2012). Spatial and temporal patterns of terrorist attacks by ETA 1970 to 2007. *Journal of Quantitative Criminology*, 28(1), pp7 – 29.

Li, H. and Reynolds, J. (1994). A simulation experiment to quantify spatial heterogeneity in categorical maps. *Ecology*. 75, pp2446 – 2455.

Li, H. and Reynolds, J. (1995). On definition and quantification of heterogeneity. *Oikos*. 73(2), pp280 – 284.

Li, W. (2006). Transiogram: A spatial relationship measure for categorical data. *International Journal of Geographical Information Science*. 20(6), pp693 – 699

Minoiu, C. and Shemyakina, O.N. (2014). Armed conflict, household victimization, and child health in Côte d'Ivoire. *Journal of Development Economics*, 108, pp.237 – 255.

Mitchell, J.L. (2018). Does Policy Diffusion Need Space? Spatializing the Dynamics of Policy Diffusion. *Policy Studies Journal*, 46(2), pp.424 – 451.

Moran, P., (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37, pp17 – 23.

Organisation for Economic Co-Operation and Development. (2016). *States of Fragility 2016: Understanding Violence*. OECD Publishing. pp 53. Available at: <http://www.oecd.org/dac/states-of-fragility-2016-9789264267213-en.htm>. First accessed: January 2018.

Raleigh, C. and Dowd, C. (2015). *Armed Conflict Location and Event Data Project (ACLED) Codebook*. Armed Conflict Location and Event Data. Available at: https://www.acleddata.com/wpcontent/uploads/2015/01/ACLED_Codebook_2015.pdf. First accessed: January 2017.

Raleigh, C. and Hegre, H. (2009). Population size, concentration, and civil war. A geographically disaggregated analysis. *Political Geography*, 28(4), pp.224 – 238.

Raleigh, C., Linke, A., Hegre, H. and Joackim K. (2010a). Introducing ACLED: An Armed Conflict Location and Event Dataset, *Journal of Peace Research* 47, pp1 – 10.

Raleigh, C., Witmer, F., O'Loughlin, J. and Denemark, R.A. (2010b). A review and assessment of spatial analysis and conflict: The geography of war. *The International Studies Encyclopedia*, 10, pp6534 – 6553.

Sánchez, F., Solimano, A., Formisano, M. (2005), Conflict, violence and crime in Colombia, *Understanding Civil War*, 2, pp119 – 159.

Sandor, K and Csillag, F. (2002). The analysis of spatial association on a regular lattice by join-count statistics without the assumption of first-order homogeneity. *Computers and Geosciences*, 28, pp901 – 910.

Schutte, S. Weidmann, N. (2011). Diffusion pattern of violence in civil wars. *Political Geography*. 30, pp143 – 152.

Starr, H. Darmofal, D. Iqbal, Z. (2008). Civil War: Spatiality, Contagion, and Diffusion. Conference on “Spatial and Network Analysis of Conflict”, September 25 – 27, 2008. pp25 - 27.

Tobler, W.R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1), pp234 – 240.

Tukey, J.W. (1977). *Exploratory data analysis*. Addison-Wesley, Reading, Massachusetts.

United Nations. (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York: United Nations. Available at: <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>. First accessed: October 2017.

United Nations Refugee Agency. (2015). Available at: <http://www.unhcr.org/576408cd7.pdf>. First accessed: October 2017.

Wallensteen, P. and Sollenberg, M. (2001). Armed Conflict, 1989 – 2000. *Journal of Peace Research*, 38(5), pp629 – 644.

Waller, L.A. and Gotway, C.A. (2004). *Applied spatial statistics for public health data*. John Wiley & Sons.

Zhang, S and Zhang, K. (2008). The study on Trinary Join-Counts for Spatial Autocorrelation. *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, pp111 – 118.