Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Sebastian Schutte

Regions at Risk
Estimating Conflict Zones in African Civil Wars

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts
in Comparative and International Studies

April 2010

Center for Comparative and International Studies, ETH Zurich

Supervisors: Lars-Erik Cederman - ETH Zurich
Kay W. Axhausen - ETH Zurich

Abstract

A method for estimating conflict zones in civil wars based on spatial Point Patterns Analysis (Baddeley, 2008) is presented in this paper. The study region is delimited to eight countries in Sub-Saharan Africa that have either experienced full-blown civil wars or strong communal violence. Based on a stylized model of guerrilla activity, several covariates were identified that hold information on distances to strategic targets. Moreover, the theoretical concepts of 'opportunity' and 'willingness' (Starr, 1978) are incorporated into the theoretical discussion. The results allow for a more general insight into the spatial extent of civil wars – independent of their declared political aims. In addition, the predictive capabilities of the statistical techniques are put to a rigid test. Generally, estimating regions that face an increased risk of becoming conflict zones in civil war seems possible based on the presented approach.

Acknowledgments

I would like to thank Prof. Lars-Erik Cederman for comments on this thesis and for providing the opportunity for me to kick-start my life as a political scientist with an entire Master's program. Moreover, I would like to thank Prof. Kay W. Axhausen for his comments and helpful suggestions as well as for his readiness to supervise a thesis outside his areas of official responsibility. Finally, Elin Arbin deserves gratitude for giving important linguistic advice.

Contents

1 Introduction

Classic research on political violence has mostly focused on the state. Although theoretical contributions have analyzed the origins of wars on lower levels (Waltz, 1959), quantitative studies usually focused on country-level observations. The strategic bipolarity and overkill capacity of the Cold War gave rise to a very clean top-down understanding of global war and peace, leaving little room for incorporating sub-national factors. With the sudden end of the Cold War, a shift in attention towards civil conflict took place in the scholarly community.

Civil wars were the most frequent and most devastating form of large-scale political violence in the second half of the twentieth century (Harbom and Wallensteen, 2005). Not surprisingly, the conceptual landscape of interstate conflict heavily influenced early studies on civil wars. Country–year observations of political regime type and socio-economic variables formed the basis of quantitative civil war research in the late 1990s. Especially two contributions triggered both interest and controversy with regard to why civil wars occur (Collier and Hoeffler, 2004; Fearon and Laitin, 2003). Both studies used highly aggregated data to show the relative importance of geographic and economic variables in comparison to political exclusion and ethno-linguistic divide. Both their findings and their methodology have been criticized in a number of subsequent contributions. First, the operationalization of grievance-related variables has been rejected as insufficient. In-

stead of ethno-linguistic diversity, for example, political exclusion along ethnic lines has been suggested as a more meaningful concept (Cederman and Girardin, 2007). Second, the exclusive focus on highly aggregated data as a basis for sweeping claims has been questioned. Instead, the inclusion of qualitative studies in the analysis of civil wars has been suggested (Sambanis, 2003). Although different in theory and methodology, these critical accounts point to the problem of over-aggregated observations.

With the unequal distribution of prosperity within nations, the infinite variety of geographic constellations, and the complex power-play of sub-national actors pressed into one-dimensional annual observations, it is easy to forget that not all wars are alike. A large variance with regard to intensity, severity, spatial extent, and – most importantly – causes can be observed in civil conflict. In several attempts to disaggregate civil wars along geographic lines, the spatial extent of the fighting (conflict zone) has become the focus of attention (Buhaug and Gates, 2002; Buhaug and Gleditsch, 2008). One important finding from the geographical literature is that conflict zones in civil wars vary considerably in size and location inside war-torn countries. Several attempts have been made to identify systematic variations of conflict zone sizes with regard to explanatory variables drawn from stylized spatial models (Boulding, 1962). Moreover, sets of explanatory factors along the lines of 'opportunity' and 'willingness' (Starr, 1978) have been adapted to work in space. For example, Hegre, Østby and Raleigh (2009) used both the distance to diamond mines and road networks as explanatory concepts in their spatial analysis, explicitly following Starr's conceptual distinction.

The shift in attention towards smaller units of analysis and disaggregated explanatory concepts has certainly contributed to our understanding of civil wars. But for all the progress in disaggregated studies on civil conflict, some obvious questions remain largely unanswered. Most importantly, a somewhat exclusive focus on confirming theoretically

derived hypotheses has neglected outcome-centric questions. For example, a straightforward answer to which regions are at risk of becoming conflict zones in civil wars is missing in all recent contributions. This is regrettable for both practical and academic reasons. While policy makers and practitioners might care less about rationalist versus constructivist schools in conflict research, they nevertheless need easily interpretable predictions. Since fighting in civil wars varies considerably with regard to location, spatial extent, and intensity across cases, reliable estimations of conflict zones can help to inform the work of policy makers and relief organizations. From a scientific point of view, easily interpretable predictions allow us to assess a model's predictive performance regardless of the underlying technology. With prediction being a general measure of scientific success across disciplinary divisions, prediction can inform scholars about the relative performance of their models.

The research question in this study therefore asks for the location of conflict zones in civil wars. Answers will be provided in the form of spatial statistical predictions of the densities of conflict events and an assessment of their reliability. Moreover, several hypotheses regarding the role of geographic factors will be tested in order to contribute to the theoretical discussion. Especially the literature on guerrilla warfare will be used as a basis for the quantitative analysis.

1.1 Research Question: Where is the Fighting?

The primary research question can be quickly summarized in one sentence: *Where does the fighting take place in a country torn by civil war?* Several related questions have been asked before, and these will be laid out in Section 1.3. In the following, a brief discussion about the relevance of the present research question is given. Much of contemporary conflict research sets out to explain the causes of war. These causes are usually

drawn from a wide array of theoretical considerations provided by the classics and recent research. While the achievements of the causal approach cannot be overlooked and the empirical tests of the proposed explanations become ever more sophisticated, it also poses problems. Most importantly, the onsets of wars in Sub-Saharan Africa may simply arise from an increase in intensity of smoldering conflicts. Corresponding examples might be the continuous struggles in East Ethiopia's Ogaden province or the regular recurrence of Tuareg rebellions in Mali and Niger (Tareke, 2000; Keita, 1998).

As Section 1.2 will explain, political violence in Africa often arises in a complex actor typology which is not delimited by state borders. While this insight might complicate questions of causality and onset, it also provides opportunities to approach the resulting spatial conflict patterns independent of war onset. This study contributes to our understanding of spatial variations of violence intensity inside war-torn countries.

Whether and to what extent the military logic of guerrilla warfare leaves its footprint in the empirical record contributes to the ongoing scholarly discussion. In addition to this theoretical aspect, this study also makes concrete predictions. What can we learn from the spatial patterns in observed cases? Can we generalize beyond the horizon of our examples? Can we help to inform policy makers with statistical insights on regions at risk? These questions will be pursued throughout the study. A quick overview of related projects will be given in the following Section.

1.2 Political Violence in Today's Africa

The empirical analysis in this study focuses on eight countries in Sub-Saharan Africa. A short overview of the political and military situation in the last decade will be given at this point. This overview is important since many of the theoretical considerations below are based on the most recent political situation in the region.

Although the Second Congolese War officially ended in 2003 with a transitional government taking power in the Democratic Republic of the Congo, the echo of this enormous confrontation still reverberates in the region. Hostilities between the complex alliances continue to this day. At least seven state actors were involved in the conflict. These include the governments of Angola, Zimbabwe, Namibia, and Congo on the one side, and Rwanda, Uganda, and Burundi on the other. Moreover, numerous armed rebel groups have taken sides in the conflict. While territorial gains along the logic of classic interstate warfare played only a secondary role in much of the conflict, targeted attacks on strategic locations were carried out regularly (Reyntjens, 2001, 311-312). A central problem in the region is the complex overlap of borders, ethnicities, and the free movement of armed groups across porous boundaries. Prunier describes the situation between the Rwandan Genocide and the first Congo War thus:

> Borders were porous, populations were highly heterogeneous, and their distribution did not correspond to the border limits; conflicts overlapped and intermingled in ways that made them influence each other even if they were of completely different nature (Prunier, 2009, 73).

The implosion of Mobutu's 32-year rule and the successful march on Kinshasa by Kabila's forces in 1997 set the stage for a wider conflict in the region. Over the course of the following six years, an estimated total of 3.8 million perished due to combat, violence against civilians, and war-related food shortage (Notholt, 2008, section 2.28).

The current situation in the Congo Basin is still largely influenced by this catastrophe. While the interstate dimension of the conflict seems resolved for the moment, an impressively large array of armed non-state actors remains active in the region. These rebel groups fight adversary governments using a small set of stereotypical tactics, including the build-up of bases outside the state's reach, point attacks on strategic targets, and the application of guerrilla warfare. As the empirical part of this study will show, operationalizing these activities in a spatial statistical model allows for the reliable prediction

of regions at risk across different civil conflicts.

1.3 Related Work

Extended knowledge on the socio-economic determinants of civil wars exists today. But as the research questions and available datasets become more detailed over time, the standard country–year level regression analysis inherited both from interstate war research and econometric methodology can be too coarse to provide suitable answers alone (Sambanis, 2003, 1). Consequently, several disaggregated approaches have been applied in the past to provide answers in two directions that are vital to the identification of the social mechanisms that drive the statistical regularities.

The first direction is agency. Fearon and Laitin's (2003) contribution possibly prematurely rejected the assumption that ethnicity and identity play at best secondary roles in the onset of civil wars. It has consequently been challenged in a number of studies (e.g. (Cederman and Girardin, 2007; Wimmer, Cederman and Min, 2009)). Disaggregating states as the classic actors in international relations research allows for a better understanding of the processes and dynamics at the sub-national level. Moreover, a wide array of long-standing theoretical claims regarding power status and national identity can only be tested on the basis of finer units of analysis. For example, cultural homogeneity in space as one principle of nationalist integration (Gellner, 1983) cannot be tested if nation states are the atomic unit of analysis.

The second direction is space. A considerable amount of classic work has focused on problems related to the role of geography in war.[1] Nevertheless, the last decade has witnessed an increased interest in geographic constellations that facilitate intrastate conflict. Starting from a country–year level, Fearon and Laitin (2003) observed that the percentage

[1] For a comprehensive review see Diehl (1990).

of rough terrain in a country is a significant predictor of the onset of civil wars. Nevertheless, the observation that the presence of mountains *somewhere* in the country of interest increases the risk of conflict leaves much to be desired. Especially Buhaug and Gates (2002) contributed to a better understanding of the exact role of geographical factors within the state. They introduced the notion of "conflict zones" into the academic discourse, thereby offering an alternative to coarse country-level measurements. These conflict zones were introduced as "the circular area centered around the conflict center and covering all significant battle zones" (Buhaug and Gates, 2002, 421). Estimating their size as a function of geographic factors, the authors were able to show the importance of distance to the nearest border and presence of natural resources as explanatory variables. Circular conflict zones have the clear disadvantage of including areas that are unaffected by civil war, such as international waters and parts of neighboring states. Nevertheless, a first convincing step towards finer spatial units of analysis had been made.

Buhaug (2008) took a closer look at the difference between separatist insurgencies and wars over the state with regard to their spatial properties. The intuitive result is that separatist conflicts tend to be fought in remote regions of the state while wars over the state target the capital. While this result does not surprise in isolation, the conceptual contribution of this study cannot be overlooked. The political aims of rebellion can be found in the spatial signatures of the corresponding civil wars. A more detailed view on the locations of violent clashes is offered by Raleigh and Hegre (2009). Using data on single conflict events, their study identifies a series of important variables that affect conflict risk at a fine-grained spatial level. Most importantly, population mass and concentration were found to drive the probability of conflict in specific locations. Moreover, road networks, distance to the capital, and distance to the nearest border were found to also influence the results. Controlling for events in the spatio-temporal proximity of the locations under

investigation enabled the model to take contagious diffusion effects into account.[2]

In a similar study, Hegre, Østby and Raleigh (2009) contributed to the political understanding of the spatial dynamics of conflict events. By focusing on the Liberian civil war, the authors were able to identify factors that enabled them to predict conflict events. Among those were the relative wealth of a specific region and distances to diamond mines. Their confirmed assumption that relatively wealthy regions are at greater risk is in line with the theoretical concept of relative deprivation (Gurr, 1970). This finding marks another milestone in the understanding of political motivations and military opportunities based on the analysis of conflict events. Inspired by these contributions, the following Section will review the theoretical literature on fighting in civil wars that forms the basis of the empirical part of this study.

[2]The statistical tools for dealing with spatial situations in which the assumption of independence between observations is clearly violated has received increasing attention in recent years. Especially Ward and Gleditsch (2008) contributed to the establishment of the required methodological standards.

2 Theoretical Considerations – Fighting in Civil Wars

An accurate estimation of conflict zones in civil wars critically relies on an understanding of the processes that generate the spatial spread of conflict events. The conceptual dichotomy of 'willingness' and 'opportunity' has been used to identify key variables that determine the onset and location of armed conflict in numerous studies. Usually, sets of variables belonging to either family of explanations are tested in multivariate regressions (see for example Fearon and Laitin, 2003; Hegre, Østby and Raleigh, 2009). While this theoretical dichotomy certainly helps to narrow the empirical analysis, it sometimes leads to over-interpreted findings.

A cautious interpretation of statistical results is presented in Hegre, Østby and Raleigh (2009). Although geographic factors reliably predict conflict locations, the authors stress that their partial non-findings are not conclusive concerning the causes of war. This reminder echos Carl Sagan's basic rule: "absence of evidence is not evidence of absence". A statistical non-finding usually translates into a failed rejection of the null hypothesis. A failed rejection is nevertheless not a confirmation: The statistical model, the underlying functional form, or the available data might simply be too limited to capture existing real-world dynamics. This basic rule has not made its way into the top ranks of quantitative

conflict research. Bold statements based on non-findings violate the probabilistic nature of hypothesis testing, and yet they are frequently presented as central theoretical findings.

The concepts of 'willingness' and 'opportunity' are used in this study to identify potential locations of violent events in civil wars. Nevertheless, a rigid differentiation of scenarios or even explicit causal mechanisms based on these concepts appears problematic. Rebels might be motivated ideologically and yet aim for low-hanging strategic fruits in inaccessible terrain. In such a case, the conduct of the military operations and the resulting locations of conflict events are shaped by some geographic opportunity structure. This geographic opportunity structure does not allow us to pinpoint the underlying political aims of the rebels: Motives as different as economic gain or socialist revolution can leave similar spatial footprints of conflict events.

Nevertheless, these spatial patterns are fundamentally related to the type of military campaign that is carried out by the political actors in place. For example, a military coup against the highest echelons of a central government will affect only specific areas in the capital city. On the other hand, modern maneuver warfare is likely to spread destruction across vast areas and along the paths of sweeping frontlines. Yet another scenario are the frozen lines of Flanders and Verdun in World War I, or the year-long fighting for Khorramshahr during the Iran–Iraq War. Clearly, the conflict zones emerging from these scenarios are very different.

Therefore, answering the general question of where conflict zones might be located in civil wars requires a deeper look into the primary type of fighting in civil conflict. A wide array of partially overlapping and sometimes fuzzy definitions surround the theoretical debate. Especially "insurgency", "new war", "asymmetric fighting", and "guerrilla warfare" are terms typically used to describe warfare that deviates from "conventional" fighting. Especially the categories of new versus old civil wars appears less helpful and

has been criticized accordingly (Kalyvas, 2001). Nevertheless, systematic differences between conventional and unconventional fighting can be elaborated. Kalyvas (2005) suggested a typology of three main types of civil war. "Irregular" or "unconventional" warfare are usually associated with intrastate conflict. A less familiar category is "symmetric non-conventional", referring to settings where both political actors are comparatively weak and employ irregular tactics. In most African civil wars, weak actors can be assumed to rely on irregular tactics. Since the spatial properties of symmetric and asymmetric non-conventional war are supposedly similar, Kalyvas' distinction between these categories will be dropped in this study. Nevertheless, the remaining distinction between conventional and non-conventional warfare is critical for modeling the spatial properties of conflict events in civil wars. The next two Sections will discuss these spatial properties in more detail.

2.1 Conventional Warfare

Of these two terms, "conventional warfare" is certainly the easier concept to agree on. The state is the principal actor in this context. Clearly delimited frontlines, combined use of heavy arms, and major battles that result in large numbers of casualties are typical attributes of this type of war. A comprehensive theory of modern conventional warfare has been provided by Biddle (2006). Key terms in his theory are the evasive elements of modern combat, collectively referred to as "cover" and "concealment". Biddle argues that the increased range and lethality of modern arms enforced drastic changes in interstate warfare at the dawn of the twentieth century. The most visible change is the abolition of large and slow formations exposed in the open. Instead, various methods were introduced to either protect (cover), or hide (conceal) troops from direct attack. A fundamental transition took place from the proud and colorful appearance of marching

armies in the Napoleonic Age to the crawling, field-gray trench inhabitants of World War I. This transition was enforced by the radically increased range, accuracy, rate-of-fire, and mass availability of modern arms. After World War I, strategies for survival under the new circumstances were developed, including the introduction of various new means of movement, camouflage, and protection. One effect of the traumatizing trench stalemate was the reestablishment of maneuver warfare. Biddle's theory underlines that World War I marked the last drastic and sudden change in warfare in the twentieth century. As Biddle puts it:

> The technological changes most often cited as revolutionary – the increased lethality of precision guided weapons, the increased range of deep strike air and missile systems, and the increased ability to gather and process information – are all extensions of very longstanding trends (Biddle, 2006, 197).

Cover and concealment between strong military opponents is usually achieved by adopting new technology and tactics in the field. Two important spatial implications arise from conventional warfare. First and most obviously, modern conventional warfare allows for long-distance dependencies: A cruise missile that is fired in the Mediterranean Sea can reach any country in the greater Middle East. Second, the reasons for theaters of war emerging in specific regions are not necessarily locally determined. The Allied campaigns in North Africa during World War II, for example, were based on assessments of the global strategic situation and not on North African rivalries. Based on these considerations, the notion of conflict zones as introduced by Buhaug and Gates (2002) and attempts to estimate them through the analysis of local determinants seems not entirely applicable to conventional, interstate war. Interstate campaigns in multiple theaters of war arise from centralized strategic considerations and not from local determinants. Modeling conflict zones in conventional warfare with all its long distance dependencies therefore seems very difficult. The next Section will compare the discussed theoretical concepts and their spatial properties to non-conventional warfare. After that, a stylized model of

the spatial dynamics of civil wars will be derived.

2.2 Unconventional Warfare

Usually associated with revolutionary movements, non-conventional and guerrilla warfare received major attention during the twentieth century. Dating back to the Napoleonic Wars, "guerrilla" is the diminutive form of "la guerra", the Spanish word for war. This type of "little war" has played a significant role not only in particular episodes of armed conflict, but in an entire class of anti-colonial wars. Moreover, virtually all major interstate wars that witnessed the long-term occupation of foreign territory experienced episodes of guerrilla fighting. Interestingly, Biddle's (2006) core concepts can be found in the theory of guerrilla warfare. Based on his experiences in the Cuban revolution, Guevara wrote a textbook-style guide on the conduct of guerrilla war. Referring to the "fundamental postulates" of guerrilla war, Guevara observes:

> As we have already said, guerrilla fighting will not always take place in the country most favorable to the employment of its tactics; but when it does, that is, when the guerrilla band is located in zones difficult to reach, either because of dense forests, steep mountains, impassable deserts or marshes, the general tactics, based on the fundamental postulates of guerrilla warfare, must always be the same (Guevara, 1961, 10).

Clearly, "cover" and "concealment" are major concerns in this text passage. While conventional forces might employ different technical implementations of these concepts, they are clearly emulated in unconventional warfare in terms of terrain utilization. Another element of guerrilla tactics are attacks on critical infrastructure and supply lines (von Dach, 1957). While the irregular forces are comparatively weak, surprise attacks outside the well-protected centers are still feasible. Stressing the importance of these attacks, theories on guerrilla warfare usually refer to them as initial steps in a wider insurgency. This insurgency eventually launches an assault on the capital, trying to overthrow the gov-

ernment. Due to the absence of large numbers of heavy arms in guerrilla warfare, the above-mentioned long-distance dependencies are of lesser concern. Since guerrilla wars require rough terrain or nearby borders to allow the build-up of bases outside the state's reach, areas at risk are arguably easier to identify. Overall, the conflict zones in guerrilla war seem to be predominantly defined by local determinants. Since unconventional fighting targets specific locations and requires the presence of certain geographic conditions, conflict zones in civil wars might be shaped systematically by infrastructure and terrain. The next Section summarizes the main points of this analysis to derive a stylized model.

2.3 A Stylized Model

The basic assumption of this model can be summarized as default rebel initiative: Based on guerrilla tactics, rebels determine the times and locations of attacks, choose the risks they are going to take, decide on the scale of the entire campaign and leave the state with very little initiative of its own. This scenario assumes the military asymmetry of non-conventional warfare. Although states can launch attacks against rebels in specific regions of the country, these attacks are usually reactions to previous rebel activities.

Since complete territorial control is not practically achievable for the state, the protection of 'vital' regions is the state's primary goal. Areas to protect are the capital city, the population centers, access routes to guerrilla strongholds, and areas of increased resource extraction. Consequently, clashes between rebels and government forces are likely to occur in the vicinity of these static strategic targets.

At first glance, this stylized model completely neglects the active role of the state in fighting rebels, but this impression is inaccurate. One important activity of the state is the reestablishment of control over strategic locations that have previously fallen into the hand of the rebels. By focusing on strategic locations, fighting as a result of repeated

changes in control are accounted for.

2.4 Hypotheses

A number of inherently spatial hypotheses can be derived from the theoretical discussion. Areas of strategic importance to the rebels are considered to be at risk. These areas include population centers, strategic elements in the road network, regions that are prone to cross-border attacks, and those that are also far away from the capital. Alternative explanations might stress the importance of locations related to monetary or political aims. Diamond mines that allow quick access to wealth could attract fighting activities. Similarly, wealthy regions could be targeted in guerrilla raids. According to the above discussion, several hypotheses will be tested.

Population

Since war is a social phenomenon, it requires humans to be present and engage in armed conflict. In interstate wars and the age of arms with global reach, dyads of senders and receivers of lethal payloads can be established rapidly. This insight leads Boulding to the conclusion that "war as a system can be defined roughly as men throwing things at each other with malicious intent" (Boulding, 1962, 266). The range of things that are thrown defines who can engage whom in aggressive actions. It defines the spatial opportunity structure in terms of Starr's conceptual framework: "Opportunity, as the possibility of interaction, creates also the possibility of conflictual interaction" (Starr, 1978, 368).

While interstate war can be fought at large distances, the attacks of lightly armed rebels usually requires the physical presence of the actors. Therefore, the presence of population is an important independent variable. Raleigh and Hegre (2009) found strong statistical support for a positive effect of population density on the probability of conflict events in

a given location. Based on these considerations, hypothesis one focuses on population:

- H1: Population density has a positive effect on the probability of conflict events.

Borders

Rebels often use the borderlands to neighboring countries to escape the state's reach. The Vietcong, for example, moved their vital supply lines partially to Laos and Cambodia, and Afghan mujahideen traditionally fight superpowers from bases in the borderlands to Pakistan. Most importantly for this study, rebels crossing porous borders is a very common phenomenon in Africa. The importance of borders has also been shown empirically: Hegre, Østby and Raleigh (2009) found the distance to the nearest international border to be a highly significant explanatory variable in a disaggregated study on the Liberian civil war. Hypothesis two pays tribute to this finding:

- H2: The distance to the nearest border has a negative effect on the probability of conflict events.

The Capital City

In his seminal work, Boulding (1962) outlined a simple formal model that proved to have impressive explanatory power. The effect of decreasing power over growing distance is captured in this model (Boulding, 1962, 245). Since the basic model assumes a linear decline, it can be expressed as a loss of strength gradient, or LSG for short. This model offers interesting insights into the basic strategic constellations between two military opponents, their LSGs, and the distance between them.

This formalism also yields a predictive element for the analysis of specific war scenarios: War is supposed to occur in locations where the strength of two states is comparable.

A state that is completely within the reach of a foreign and superior military force is unlikely to launch a military offense due to the predictable and unfavorable outcome. The political effect of a looming military superiority is therefore part of this model.

Buhaug (2008, 3) made an explicit reference to Boulding's approach and observed: "In a domestic setting, the ability of a state to exert authority throughout its territory is determined by the government's capability and its LSG". Following this adoption of the LSG concept to internal conflict, the geodesic distance between a state's capital and single locations inside its territory will be used as an explanatory variable. Buhaug and Rød (2006) distinguished among two scenarios in their analysis of the location of conflict events. In wars over the state, fighting events usually take place closer to the capital than in secessionist wars. While this finding is intuitive, it requires a distinction between revolutionary versus secessionist wars for understanding the effect of distance to the capital. Therefore, a twofold effect for distance to the capital city is postulated as the third hypothesis:

- H3a: In secessionist wars, distance to the capital has a positive effect on the probability of conflict events.
- H3b: In wars over the state, distance to the capital has a negative effect on the probability of conflict events.

Cost-Optimal Paths

Both the theory and the history of guerrilla warfare focus on targeted attacks against supply lines and key elements of the state's infrastructure. Identifying these lines automatically is not a straightforward procedure. First of all, realistic starting points and destinations have to be identified. Following the stylized model, the state's capital always represents the starting point of these routes. The other end are areas of strategic importance that the state tries to control. Many possible strategic locations spring to mind as

possible candidates. Protecting large cities or natural resources are certainly concerns of the state. For the purpose of this study, however, the geographic centers of the ethnic groups belonging to the state are taken into account. The rationale behind this second, less obvious geographic location is the following: in an insurgent war, the state needs to maintain control over the different ethnic fractions. The employed access routes guarantee shortest access in time to these centers. Expressed in a hypothesis, the role of center-periphery routes is assumed to be the following:

- H4: Distance to cost-optimal center-periphery paths has a negative effect on the probability of conflict events

Wealth

The connection between poverty and the probability of civil war onset is a robust finding on the macro level and therefore controlled for in many studies. Whether this relationship also holds on the micro-level is yet to be tested. Nordhaus et al. (2006) provided disaggregated GDP data on a global scale for the years 1990 and 2000. The data set provides estimates for the "Gross Cell Product" (GCP) – the local fraction of a state's GDP for 100 km² cells. The comparatively low resolution of the dataset called for the inclusion of additional data sources. Therefore, data on the location of diamond mines was included from Gilmore et al. (2005). Obviously, diamond mines provide opportunities for fast resource extraction and are therefore considered strategically important in civil wars. The inclusion of this data was intended to supplement the coarse grained GCP estimates.

Generally, less developed countries often face large inequalities in the distribution wealth as reflected in high Gini coefficients. But this measurement does not express the spatial distribution of wealth. The GCP measurement fills this gap by providing disaggregated and geocoded data on wealth. Following the theoretical discussion and the

assumption of rebel initiative, wealthy regions should be at a higher risk of seeing armed conflict. Rebels might attack them to loot and steal or to deny the state sources of income. Independent of the political goals of the rebellion, wealthy regions should attract fighting. Therefore, hypothesis five pays tribute to the role of wealth in specific locations.

- H5a: GCP has a positive effect on the probability of conflict events
- H5b: Distance to the nearest diamond mine has a positive effect on the probability of conflict events

Terrain

One essential doctrine of Guerrilla Warfare is the utilization of rough and inaccessible terrain (von Dach, 1957). Operationalizing this idea requires data on 'accessibility' which is hard to define. While terrain elevation is a more easily established measure, the variety of factors that can increase or decrease the accessibility of a given location provides problems to a coherent coding. Terrain and soil conditions, road and railroad networks, bodies of water, and forested regions all affect the maximal possible traveling speeds. As discussed in Section 3.3, a suitable data set could be obtained for this study. The effect of terrain accessibility is assumed to be similar to the one of spatially distributed wealth: although rebels might use inaccessible terrain to escape the state's reach, they preferably attack in the infrastructurally developed regions. The reason is simply the increased strategic importance of these areas. Moreover, the state's interest in these more important areas allows for clashes between the opposing forces and thereby a higher number of conflict events than in the less accessible areas. Consequently, hypothesis six points to the importance of terrain accessibility:

- H6: Terrain accessibility has positive effect on the probability of conflict events

3 Underlying Data

This Section gives an overview of the data that was used for the statistical analysis. Following the theoretical discussion above, variables are drawn from a variety of categories to test the hypotheses. Some of the measurements are straightforward, like population size in certain regions, while others were established based on more subtle characteristics.

Since the research question is inherently spatial, the covariates are operationalized as spatial layers. Instead of a small number of observations for the independent variables, covariates for the entire window of observation are used to explain the observed point pattern. The next Section describes the properties of the underlying data in detail. All covariate layers were converted to the base resolution of the 'Gridded Population of the World' dataset, which is about 30 km² per cell. This resolution was chosen as a compromise between computational tractability and spatial accuracy. To allow for comparability, most data layers were chosen to reflect the situation of the 2000 to 2010 decade. One exception is the Gridded Cell Product (Nordhaus et al., 2006) which reflects the situation of 1990.

3.1 Case Selection

The selection of suitable cases for this study was severely complicated by a number of factors. First and most importantly, the cross-sectional design of the study was based on

independent observations from the last decade. Infrastructure, for example, is likely to change over time and the available snapshot reflected the year 2000. Therefore, cases of civil conflict in this study had to take place in a similar time frame. Effectively, only the last decade qualifies for a suitable time window, assuming a causal link between preexisting spatial conditions leading to specific patterns of conflict events. Moreover, several theoretical assumptions presuppose weak states, large countries, and porous borders – typical attributes of African civil wars. Systematic deviations from these attributes could have induced a mismatch between theory and empirical analysis. Needless to say, data on single conflict events is only available for a minority of civil wars.

With the selection of cases narrowed down to the last decade and the African continent, eight countries from the ACLED data set (Raleigh and Hegre, 2005) were included in the analysis. These are Burundi, Democratic Republic of the Congo, Ethiopia, Guinea, Kenya, Liberia, Niger, and Sudan. Several other countries were excluded, because they either hold to little information (less than ten conflict events) or they only referred to communal violence with unclear political objectives. ACLED provides information on single conflict events, the corresponding dates, and locations for about 50 countries.

3.2 Population

Data from the "Gridded Population of the World" project was taken as one covariate layer. It offers accurate information on population density and absolute population size for reasonably small grid cells (about 30 km²). Figure 3.1 shows absolute population numbers for the Democratic Republic of the Congo and neighboring states. The color scheme ranges from light (low) to dark (high). The region in Figure 3.1 serves to illustrate the various data layers.

Figure 3.1: Gridded Population of the Word for the example region.

3.3 Accessibility

Accessibility is defined in this study as a combined measurement of several factors that affect possible traveling speeds. A suitable coding of accessibility requires the creation of a so-called friction map: a data structure that holds information on traveling speeds for single cells. The factors that influence traveling speeds can be manifold: Road networks, soil conditions, terrain slope, elevation, and forestation certainly play a role. A comprehensive aggregation of these factors has been performed by Nelson (2008). His global friction-map for traveling times between all cities with more than 50.000 inhabitants offers a suitable data basis for this study. Figure 3.2 shows gridded travel times values for the example region. Colors range from light (low) to dark (high).

Figure 3.2: The Global Map of Accessibility.

3.4 Cost-Optimal Paths

Identifying the exact progression of strategic center-periphery routes requires knowledge on obstacles and traveling conditions along the way. This information is extracted from the global map of accessibility (Nelson, 2008). Moreover, an algorithm is needed to calculate a global optimum for a sequence of steps between neighboring locations that connects start and destination. A well-known path-finding algorithm was reimplemented for this purpose Dijkstra (1959). This algorithm identifies shortest paths in weighted graphs. By turning the "Global Map of Accessibility" into a graph structure that connects neighboring units with weighted edges corresponding to the travel times values, this algorithm could be used to identify shortest paths in time based on realistic information on infrastructure. The often used geodesic distances and the corresponding routes along great circle bearings are of little help in the identification of realistic traveling routes. By focusing on routes that connect locations on the surface of the plant directly, great circle bearings are only applicable to sea transport or aviation, but not to movements on land. Land-based transportation routes are intimately connected to irregular geographical features and available infrastructure. Consequently, optimal paths in time also show irregular features. The distance to these path is encoded as an additional, explanatory variable. Figure 3.3 shows the distance to the cost-optimal paths with colors ranging from light (low) to dark (high).

Figure 3.3: Distance to cost-optimal paths between center and periphery.

3.5 Distance to Capital

The geodesic distance to the capital for every location in the country was included in the statistic. Following the 'Loss-of-Strength' model (Boulding, 1962), fighting should occur in a specific distances to the capital city, in areas where the military strength of the opponents evens out. These geodesic distances were calculated for every cell in the observational window. Figure 3.4 shows the distance-to-capital values ranging from dark (low) to light (high).

Figure 3.4: Distance to capital for all locations in the example region.

3.6 Distance to Border

The logical counterpart to the LSG is a measurement of distance to the nearest border. As discussed in Section 2.4, rebel bases outside the state's reach are a frequent phenomenon in African civil wars. Figure 3.5 shows the border distance values for the sample region. Colors range from light (low) to dark (high).

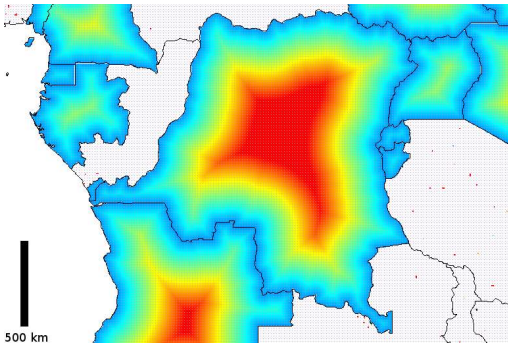


Figure 3.5: Distances to the nearest border.

3.7 Diamonds

Natural resources play an important role in both the onset and the recurrence of civil wars. In several countries like Liberia and Sierra Leone, diamond mines have been of strategic importance to the rebels. Unlike oil or metal, diamonds are easily transported and can provide extreme revenue when sold on the world market. Financing rebellion through diamond exports is a common pattern in African civil wars. To account for this factor, a layer that holds distances to the nearest diamond mine has been added to the analysis. Figure 3.6 shows these distance for the example region. Color range from dark (low) to light (high).

Figure 3.6: Distances to the nearest diamond mine.

3.8 Gross Cell Product

GDP per capita is frequently used in macro-level statistics on civil war onset. A disaggregated measure that reflects wealth in specific locations was provided by Nordhaus et al. (2006). The unit of observation in this data set is the Gross Cell Product (GCP) – an estimate for the market value of all goods and services in a geographic region. Cells with a maximal size of 60 nautical square miles (about 111 kilometers) are coded in this dataset. Figure 3.7 shows the "Gridded Cell Product" for the example region. Colors range from dark (low) to light (high). The resolution of the data was artificially improved by interpolation to make it comparable with the remaining covariates.

Figure 3.7: Gross Cell Product for the example region.

4 Analysis and Modeling

This chapter describes in detail the quantitative part of this study. Guided by the previous theoretical discussion, the aim of the quantitative analysis is to estimate a series of models that capture the spatial distribution of civil war conflict events in Sub-Saharan Africa. A series of country-based model estimates will be presented and evaluated. Moreover, the more ambitious goal of creating a generalized model with out-of-sample predictive capabilities will be pursued in a second step.

4.1 Modeling the Theory

As stated above, the presented explanation for the spatial distribution of conflict events is centered around the assumption of rebel initiative: Small groups of combatants can easily travel the vast and sparsely populated areas of the African hinterland. Their mobility is not seriously limited by any state's ability to control its own territory. Random encounters and skirmishes between rebel combatants and the state's military are therefore relatively rare. Instead, the rebels are free to exploit their tactical initiative when it comes to attacking military targets. Once these targets are taken, the state is likely to respond, leading to further encounters in the region.

Since the analysis is inherently spatial, the applied regression analysis differs from classic econometric approaches. Instead of independent variables belonging to differ-

ent observations, independent covariates in terms of spatial raster layers will be used to estimate the models. These covariates carry information about infrastructure, spatially distributed wealth, or vital routes for the state's projection of power.

Clearly, the functional relationship between covariates and conflict events has to be specified in advance. The formal specification of the stochastic process that supposedly generated the observed point pattern has to strike a balance between mathematical tractability and theoretical adequacy. Very much in line with the first requirement is the spatial Poisson Process that will be discussed in detail in Section 4.2.2. In essence, the Spatial Poisson Process makes two simplifying assumptions about the mechanisms that bring about conflict events in civil wars.

These assumptions are the spatial and temporal independence of events. Spatio-temporal independence of events may seem contrary to common theoretical wisdom, but it is a justified starting point for this analysis. In econometric studies on civil war onset on the macro level, the assumption of independence between individual observations is clearly violated in many instances. One example for such a situation might be the coeval outbreak of civil wars in two neighboring countries. If the neighboring countries are perceived as independent observations, well-known diffusion effects might remain unaccounted for. Both Salehyan and Gleditsch (2006) and Gleditsch (2007) describe the potentially destabilizing effects of sudden refugee influx on neighboring states. Moreover, several guerrilla groups in central African countries operate against governments from bases in neighboring states, thereby escaping the opposed government's reach. These two phenomena clearly undermine any assumption of independence in the observations. The statistical solution in spatial econometrics is based on the inclusion of a "spatial lag": the value of the dependent variable in neighboring units (Ward and Gleditsch, 2008, 18). The effect of civil war on a neighboring country is thereby explicitly taken into account in the regression model.

With regard to this study, however, the assumed diffusion processes are less clear. In comparison to civil wars that last for years and affect the lives of hundreds of thousands of people, conflict events in the ACLED data are mere blinks of an eye, historically speaking. Skirmishes between rebels and government forces might last for only a few hours instead of years. These micro events do not necessarily affect each other. Neighboring conflict sites in a cross sectional view might be brought about by different conflict events between different actors. Years might pass in between their occurrence. Moreover, the ACLED coding of civil wars is clearly a very incomplete record of violent events. Hundreds of thousands were reported dead from the civil war in Burundi and only 109 minor conflict events are encoded in ACLED for this particular conflict. These obvious gaps in the data could effectively destroy existing spatio-temporal dependencies. Moreover, assessing the spatial correlation based on measurements like Moran's I does not allow for the conclusion that spatial diffusion is actually taking place. Since the independent variables (like population density) are also clustered in space, spatial correlation of the dependent variable is not a clear indicator for diffusion as such.[1] Defining an arbitrary scope for the diffusion dynamics (e.g. 30 days and 30 kilometers) would include another ad hoc decision into the modeling process. A theoretical derivation of such a scope seems provisional at best: Why should 30 kilometers be the general maximum distance for conflict spillovers?

Despite this methodological problem, spatial dependency can play a role in the spread of conflict events. After a violent clash in a specific location, fighting forces are usually still in the region and can attack nearby targets. Locations that have a history of armed combat might be reinforced to withstand future attacks. These examples briefly illustrate that the data-generating process features elements of spatial dependency that a Poisson Process lacks. Instead of explicitly accounting for the history of armed clashes in the

[1] For a more elaborate discussion of this problem, see Buhaug and Gleditsch (2008).

region, the spatial Poisson Process assumes independence between the events. The events are brought about by exogenous factors, such as the proximity of diamond mines, and not by endogenous dynamics.

But this apparent incompatibility between dynamic theory and static method can be resolved based on the following assumption: Even though the spread of fighting events displays endogenous dynamics, the strategic aims of violent clashes are unlikely to change. If the rebels succeed in seizing selected targets, the government forces might counterattack in the same region. If the rebels fail to seize their objective, they might try again. These repeated battles for relevant targets are likely to exhibit endogenous dynamics: Forces might relocate repeatedly in evasive maneuvers, chase after each other, or ambush in favorable terrain. Nevertheless, these battles are still likely to occur in the proximity of strategic locations. Therefore, these strategic locations remain good predictors during a conflict that yields potentially complex spatial dynamics.

To further account for conflict in neighboring regions, the unit of observation is chosen somewhat modestly. It is not the individual conflict event, but a density estimate of events in a specific region. A continuous density estimate is by definition driven by events in the neighboring locations. This view on the data is oblivious to whether or not these neighboring events were brought about by endogenous or exogenous conflict dynamics. Although they are less specific than point predictions, event densities still allow for an identification of the general regions at risk.

To sum up the above discussion, it is important to recognize that there is no ready-to-use test for spatio-temporal interactions between conflict events. Assuming independence seems the appropriate procedure both in light of obvious gaps in the available data and with regard to the unclear spatio-temporal scope of diffusion events. Moreover, a focus on event density instead of single events as the unit of analysis effectively exploits the

equifinality of endogenous and exogenous conflict dynamics: Both types of conflict lead to a series of clustered events in the proximity of strategic locations.

4.2 Statistical Analysis

The statistical analysis serves two main purposes. First, the relevance of the covariates is tested. This first step is important for testing the theoretical assumptions about the occurrence of conflict in specific locations. The second purpose of the analysis is a rigid test of the predictive performances of the fitted models, both in-sample and out-of-sample. The knowledge gained in this analysis should allow for a geographic generalization of areas at risk beyond the included cases. Ideally, counterfactual scenarios for civil wars could be explored at a later point on the basis of this research. To estimate the reliability of the out-of-sample predictive capabilities of a generalized model, cross-validation is performed for the eight cases.

The applied spatial statistics approach in this study is usually referred to as "Point Pattern Analysis" in the literature. Ripley (1977) in particular contributed to the development and promotion of the underlying formal techniques of this approach. A general distinction between spatial econometrics and spatial statistics is not easy, since they employ a variety of modeling and fitting techniques that partially overlap. A clear difference however is the type of the dependent variable. While spatial econometrics allows for continuous dependent variables, Point Pattern Analysis assumes point events to be the outcome of a stochastic process. Moreover, spatial econometrics allows for time series analysis, while Point Pattern Analysis is strictly cross sectional. These limitations are accounted for in this analysis by focusing on a narrow time window of less than ten years. Data on single conflict events for the 2002 to 2009 period were obtained from the ACLED data set.

4.2.1 Testing for Complete Spatial Randomness

An appropriate starting point for analyzing spatial point patterns is a test for complete spatial randomness. Two statistics are suitable to test the assumption of randomness. The classic chi-square test can be adapted for Point Pattern Analysis in the following way: The spatial observation window is divided into quadrats of known size. These quadrats each contain a number of points. These numbers are treated as the random variables that are included in the chi-square test. Significant deviations from the chi-square distribution are then calculated as in classic econometrics. More generally, the Kolmogorov-Smirnov (KS) test can be used to compare probability distributions. Comparing the observed quadrat counts to those from a uniform random distribution provides a test for spatial randomness. Table 4.1 holds p-values for both tests. Spatial randomness is rejected in all cases.

4.2.2 Fitting Spatial Poisson Processes

As mentioned in 4.1, the inhomogeneous spatial Poisson Process provides the mathematical basis for modeling the spatial patterns of conflict events. The underlying assumptions deserve a closer look at this point. Since the inhomogeneous spatial Poisson Process is a variant of the simpler homogeneous one, the homogeneous variant will be discussed first. Subsequently, the inhomogeneous form will be analyzed.

Properties of Spatial Poisson Processes

The Spatial Poisson Process is a random process. In its simple homogeneous form, only a few assumptions are made with regard to the nature of the data generating process. First and most importantly, the spatial points are assumed to be distributed in space independently of one another. Second, the number of points falling into any region follows

Country	Kenya	DRCongo	Sudan	Ethiopia	Burundi	Guinea	Liberia	Niger
ChiSq	1.1e-33	6.1e-278	4.1e-57	1.9e-10	1.4e-36	0.00056	1.1e-07	0.039
KS	0	0	5.4e-12	2.8e-06	0	0.01809	0.00098	0.0012

Table 4.1: Tests for complete spatial randomness.

a Poisson distribution. The formal properties of this homogeneous Poisson Process of intensity $\lambda > 0$ are defined in Baddeley (2008, 72):

- (PP1): the number $N(X \cap B)$ of points falling in any region B is a Poisson random variable
- (PP2): the expected number of points falling in B is $E[N(X \cap B)] = \lambda \cdot area(B)$
- (PP3): if B1, B2 are disjoint sets then $N(X \cap B1)$ and $N(X \cap B2)$ are independent random variables
- (PP4): given that $N(X \cap B) = n$, the n points are independent and uniformly distributed in B

where λ is the intensity of the point process, e.g. the expected number of points falling into an area of certain size. Calculating the probability of the occurrence of a given number of points in a specific region only requires two parameters: the number of points n and the intensity of the point pattern λ. The probability is then given by:[2]

$$f(n;\lambda) = \frac{\lambda^n e^{-\lambda}}{n!},$$

where

- e is the base of the natural logarithm
- n is the number of points in a specific area
- λ is the expected number of points for that area

Clearly, simulating point patterns based on the resulting probability distribution does not usually resemble real-world conflict patterns. Instead, this procedure would effectively implement complete spatial randomness – an assumption that we have rejected for the

[2] See NIST/SEMATECH (2003, sec 3, 331).

empirical record above. Therefore, the homogeneous Poisson Process is a good conceptual starting point, but it does not realistically approximate the empirical sample. Nevertheless, the inhomogeneous Poisson Process deserves a closer look as a formal extension. The inhomogeneous Poisson Process allows for the intensity λ to vary within the observational window. This is especially important for this study, since the variation of intensity can be made dependent on one or more covariates. This formal adaption requires the replacement of two properties of the homogeneous Poisson Process (Baddeley, 2008, 79):

- (PP2'): the number $N(X \cap B)$ of points falling in a region B has expectation $E[N(X \cap B)] = \int_B \lambda(u)du$.
- (PP4'): given that $N(X \cap B) = n$, the n points are independent and identically distributed, with common probability density $f(u) = \lambda(u)/I$, where $I = \int_B \lambda(u)du$.

This brings us much closer to a realistic model: Some regions in the countries of interest might be crowded with conflict events, while other areas remain comparatively peaceful. Assuming that the variance in intensity is brought about by the values of the introduced covariates allows for fitting $\hat{\beta}$ coefficients for the covariate layers to reproduce the empirical distribution.

The fitting can be performed based on a maximum likelihood procedure (Berman and Turner, 1992). Translating this formalism back into theory leads to the following simplification: Fighting events occur independently of one another and at somewhat random locations. However, the number of fighting events in different regions of the observational window is based on specific values of the covariates. The covariates are weighted based on the observed number of events for these values. Therefore, the target distribution of event intensity is approximated in the model.

What can we gain from fitting such a model? First of all, the distribution of covariate values can be compared to the intensity distribution in the model. This allows for a basic

test of goodness-of-fit. Moreover, we can use the estimated intensities to simulate event distributions in a Monte Carlo procedure. While the single event distribution with its specific event locations might be largely driven by random variability, the average density of a series of simulations should be able to capture regions at risk. Extrapolating from the sample to other countries allows for out-of-sample predictions.

4.2.3 Simulation and Prediction

Tests of statistical significance are important for the theoretical discussion. Identifying the factors that determine the locations of armed clashes allows for a critical review of the theoretical assumptions. Nevertheless, this study also aims at prediction. In order to test the predictive performance of the model, densities of empirical and predicted events were compared. Point patterns are by no means continuous functions: A point is either absent or present in a given location. The density of points in a certain area can be computed, nevertheless, based on a Gaussian kernel function. Applying the same techniques to the average density in a series of simulated runs allows for the comparison of two continuous estimates.

Before discussing these results in more detail, the general procedure of simulating point patterns and assessing their predictive correctness should be discussed. After fitting the model coefficients to the empirical sample, the estimated intensity ($\hat{\lambda}$) of the point pattern can be calculated for each location of the map. Based on this estimated intensity, Monte Carlo simulations of point patterns can be calculated. The individual simulation run is largely driven by random variability and does not allow for reliably testing the underlying model. Nevertheless, the mean result for a larger number of simulations reveals the model's specific spatial probability distribution more accurately. For each country, 100 simulation runs were performed. The mean density value from these simulations series

was used to assess the predictive accuracy both numerically and visually. Moreover, the empirical density distribution allowed for the establishment of a standard deviation and confidence intervals.

These confidence intervals enable the graphical representation of areas in the country in which the average predicted density is within an empirically established confidence interval. The necessary accuracy for an in-sample prediction is defined as the simulated density value within the 10% confidence interval of the empirical distribution. Section 4.3.3 discusses graphical representations of the model predictions. The prediction accuracy for the out-of-sample setting can be assessed similarly to the in-sample approach.

4.3 Results

4.3.1 Statistical Results

To get an overview of the statistical relevance of the covariates in a quasi-econometric sense, a test for significance was employed. As mentioned above, the Kolmogorov-Smirnov test compares two probability distributions. It can be utilized in Point Pattern Analysis to compare a covariate's cumulative probability distribution to the probability distribution of a spatial Poisson Process (Berman, 1986). The advantage of the Kolmogorov-Smirnov test is that it is oblivious to the type of underlying probability distribution – it is non-parametric and distribution-free. Based on these properties, tests for statistical significance can be performed for a series of models. Table 4.2 summarizes the number of significant effects found for each of the covariates and the eight cases. The tested model includes all seven covariates: population density, distance to the capital city, distance to the nearest diamond mine, the Gross Cell Product, distance to the nearest border, distance to the nearest optimal center-periphery path, and terrain accessibility.

Covariate	Pop.	Dist. to Cap.	Diamonds	GCP	Border Dist.	Optimal Paths	Accessibility
# Significant	4/8	2/8	4/6	1/8	5/8	3/8	3/8

Table 4.2: Number of significant effects ($p <= 0.05$) for the covariates based on the Kolmogorov-Smirnov Test.

The Table shows that for all covariates, significant effects can be observed in some of the cases. Nevertheless, a closer look at the results is necessary for a detailed discussion. Therefore, Table 4.3 shows the coefficient estimates for each country.

4.3.2 Simulation and Prediction

Tests of statistical significance are important for the theoretical discussion. Identifying the factors that determine the locations of armed clashes allows for a critical review of the theoretical assumptions. Nevertheless, this study also pursues a secondary objective: the identification of regions at risk. In order to test the predictive performance of the model, densities of empirical and predicted events were compared. Point patterns are by no means continuous functions: A point is either absent or present in a given location. The density of points in a certain area can be computed, nevertheless, based on a non-parametric kernel function.[3] Applying the same techniques to the average density in a series of simulated runs based on the predicted probabilities of a fitted model enables the comparison of two continuous estimates.

A visual inspection of the densities reveals general attributes of the fitted model; these will be discussed in Section 4.3.3. The model's ability or inability to identify the general "heat zones" of armed rebellion can be seen with the naked eye. With the exception of Ethiopia, the fitted model correctly reflects the approximate conflict zones. Moreover, the approximate magnitude of conflict events in these zones is correctly identified.

However, in order to establish a general metric of in-sample prediction, eyeballing is not sufficient. Instead, the model's ability to reproduce the empirical findings within a certain confidence interval should be regarded as a correct prediction. The predictions for different areas in the country might differ, so the test has to be performed for multiple

[3] Note the the bandwidth parameter is essential for the exact shape of the density estimate. In all following estimates, Silverman's 'rule of thumb' (Silverman, 1986, 48)) was used to estimate a suitable bandwidth parameter for each country.

	N	Intercept	Population	Opt. Paths	Loss of Strength	Border Dist.	Diamonds	Cell Product	Accessibility
Burundi	109	5.775e+00***	2.437e-07*	-4.932e-02	-1.929e-02**	-2.033e+00**	4.791e-02*	2.006e-09	-1.124e-03*
DRC	210	-2.981e+00***	1.436e-06*	-3.971e-03	1.853e-03*	-1.549e-04*	1.494e-02**	9.592e-10	-4.84e-03
Ethiopia	154	9.551e-01***	3.195e-05	-2.098e-03*	2.375e-03	-6.627e-03**	-	-1.624e-09	-1.97e-03
Guinea	21	3.636e+00***	1.44e-06	-1.72e-03	4.028e-03	-2.362e+00	-2.323e-02*	-5.241e-09	-1.497e-02
Kenya	180	1.619e+00***	1.875e-06***	6.681e-03*	2.166e-03	-3.622e-03*	-	7.777e-10	-3.622e-03**
Niger	17	3.345e-01***	-2.688e-05	-3.622e-03	-2.518e-03	-4.275e-04	-2.01e-02*	-7.922e-10	-5.003e-04
Liberia	43	2.86e+00***	3.556e-06	-1.037e-02	1.282e-02	-1.168e-03	-7.254-e03	6.917-e10	-8.475e-03
Sudan	214	1.68e-01***	2.91e-06***	-7.479e-03***	3.571e-03***	2.658e-03***	-6.069e-02	8.428e-10**	-3.674e-03***

Table 4.3: For each country, the results from the Kolmogorov-Smirnov test for statistical significance of the covariates are presented in this Table. Significance codes should be read as follows: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

locations inside the observation window. The Figures in Section 4.3.3 show areas of correct (pink) and incorrect (blue) predictions. Note that in most cases, the high-intensity conflict zones are approximately correctly predicted.

Table 4.4 shows the performance of the different predictions numerically and can be interpreted in the following way. The country's area for which correct predictions were made is given in percent. Correct predictions were defined above as simulated density estimates inside an empirical confidence interval. The in-sample predictions were evaluated in a straightforward procedure: Models were fitted for the individual countries and then used in the Monte Carlo simulation of conflict events. For the out-of-sample prediction, a different procedure was used: Based on the model estimates of the remaining seven countries, the eighth country was predicted.[4] The $\hat{\beta}$ estimates of the training cases were weighted by the number of corresponding observations. A weighted average of their $\hat{\beta}$ coefficients was used to predict the remaining unseen case.[5]

This procedure was repeated for all eight cases with the exception of the Democratic Republic of the Congo which had to be skipped due to memory limitations.

Country	In-sample 10%	Out-of-sample 20%
Sudan	18.97%	59.31%
Ethiopia	54.41%	69.85%
Kenya	48.84%	61.02%
DR Congo	27.06%	-
Liberia	47.96%	55.87%
Burundi	77.93%	74.47%
Guinea	27.24%	58.88%
Niger	0%	1.71%

Table 4.4: Prediction performance for the various models: in-sample and out-of-sample.

To further illustrate the predictive performances, Table 4.1 shows the accuracy of out-of-sample predictions for different confidence intervals. For required accuracies between

[4] This procedure largely resembles k-fold cross-validation as described by Mitchell (1997, 147). One notable difference is that the training set is not randomly selected. Instead, seven out of the eight countries were used to predict the remaining one.

[5] In order to preserve the approximate number of conflict events to predict, the intercept was actually fitted on the case itself.

20% and 30%, most predictions are comparatively accurate. This accuracy only refers to the correct prediction of event densities somewhere in the country. The next Section will show that the high intensity conflict zones are also approximated correctly in six out of eight cases.

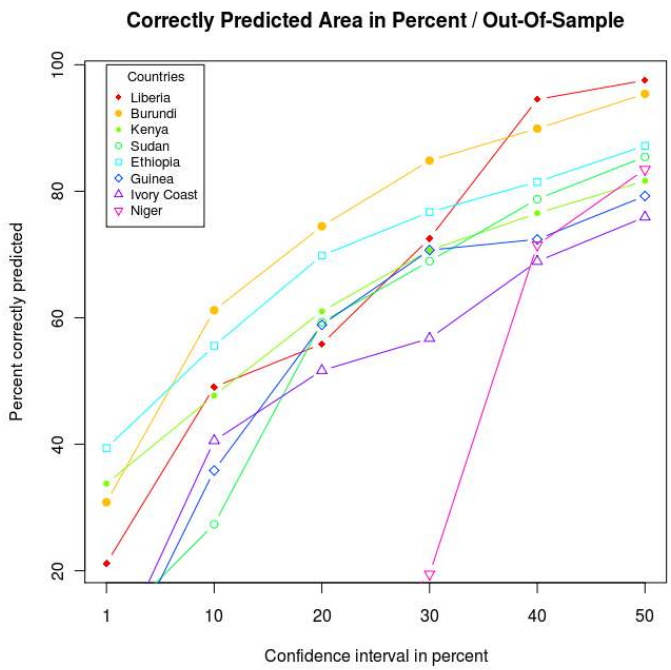


Figure 4.1: Prediction accuracy for different confidence intervals. Note that confidence intervals of 20% to 30% allow for the reliable prediction of 60% to 70% of the country area.

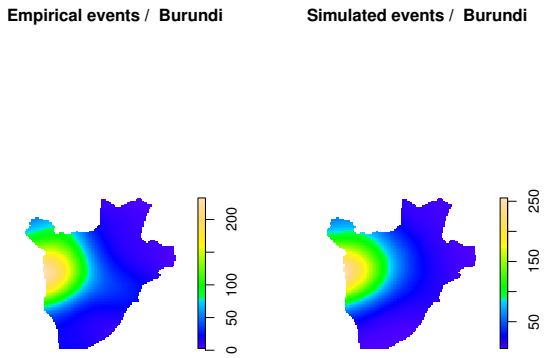
4.3.3 Graphical Predictions

The spatial predictions can also be assessed visually, giving an important and intuitive insight into the model's abilities and limitations. A series of graphics will be presented and discussed for this purpose. Comparisons of in-sample predictions and empirical densities are presented for each country. Moreover, out-of-sample predictions compared to the empirical record are shown in a second series of plots.

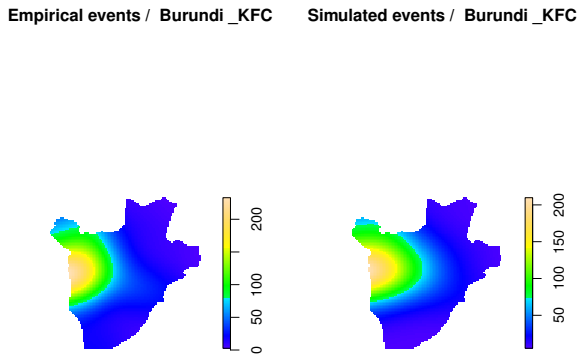
The following graphics directly compare predictions and empirical densities in the two

topmost maps. These predictions were made in-sample, i.e. based on a model that was fitted on the empirical events. Differences in densities are illustrated in the lower graphics. On the left, deviations between empirical and simulated densities of more than 10% are shown in dark grey (in-sample). The remaining light grey regions of the country area are predicted correctly. On the lower right, deviations between empirical and simulated densities of more than 20% are shown in dark grey (out-of-sample/cross validation). Again, the remaining regions are predicted correctly.

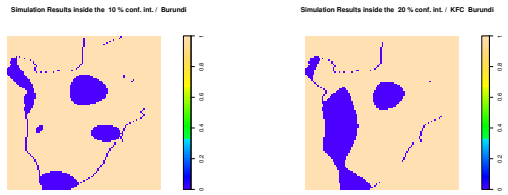
Burundi



(1)



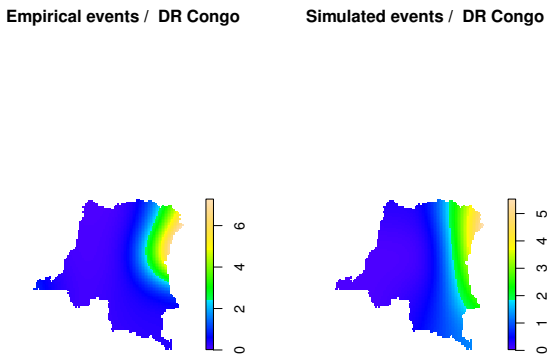
(2)



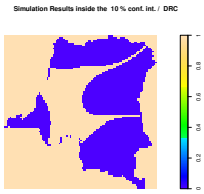
(3)

Figure 4.2: From top to bottom: Visual comparison between empirical and simulated event densities in-sample (1) and out-of-sample (2). Areas that were predicted correctly at a given confidence interval both in-sample and out-of-sample (3).

Democratic Republic of the Congo



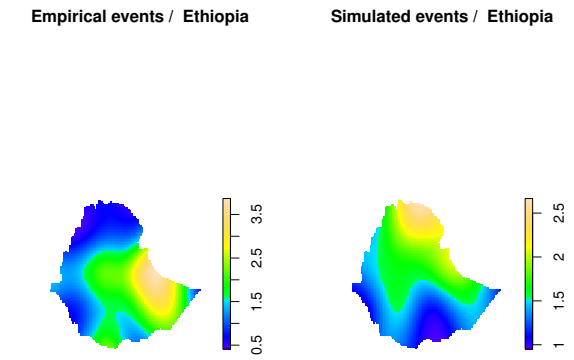
(1)



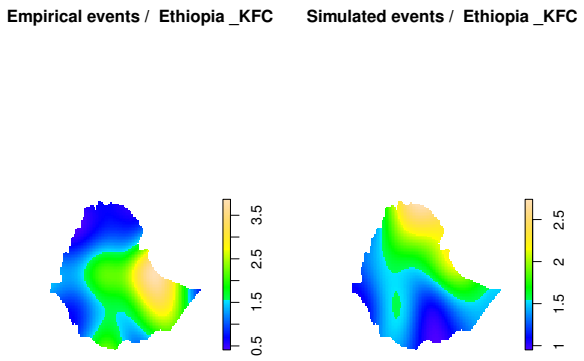
(2)

Figure 4.3: From top to bottom: Visual comparison between empirical and simulated event densities in-sample (1). Areas that were predicted correctly at a given confidence interval in-sample (2). Please note that no out-of-sample prediction could be performed for DRC based on memory limitations.

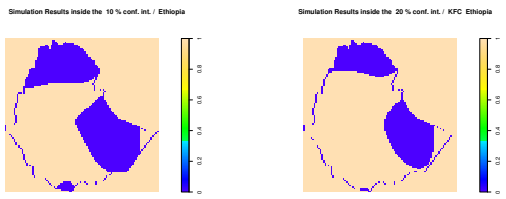
Ethiopia



(1)



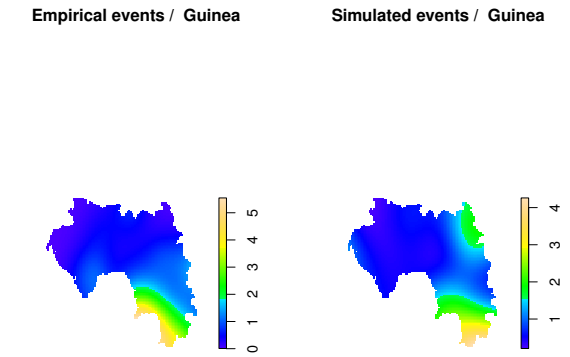
(2)



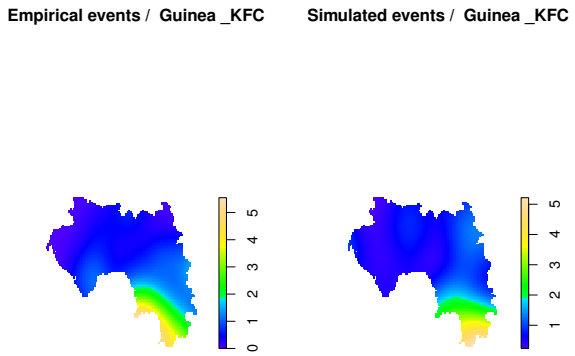
(3)

Figure 4.4: From top to bottom: Visual comparison between empirical and simulated event densities in-sample (1) and out-of-sample (2). Areas that were predicted correctly at a given confidence interval both in-sample and out-of-sample (3).

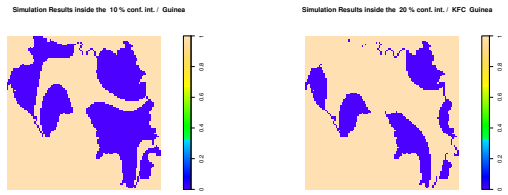
Guinea



(1)



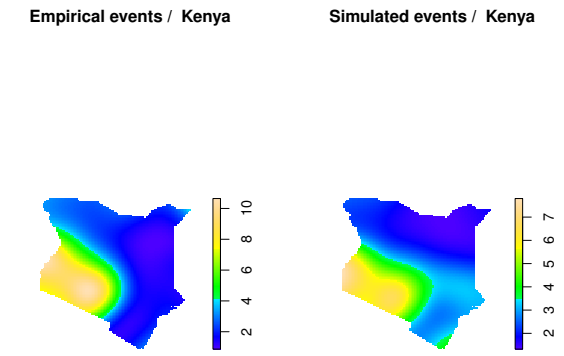
(2)



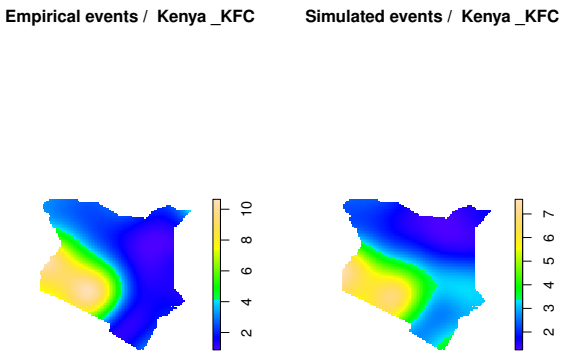
(3)

Figure 4.5: From top to bottom: Visual comparison between empirical and simulated event densities in-sample (1) and out-of-sample (2). Areas that were predicted correctly at a given confidence interval both in-sample and out-of-sample (3).

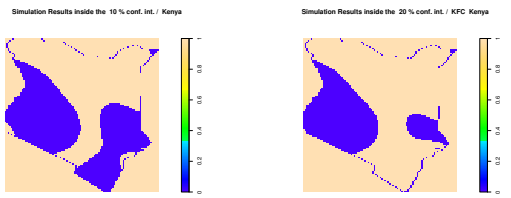
Kenya



(1)



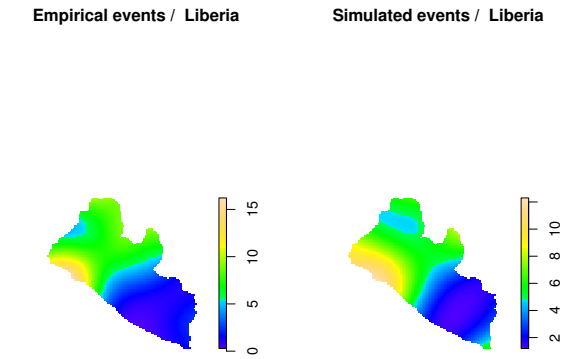
(2)



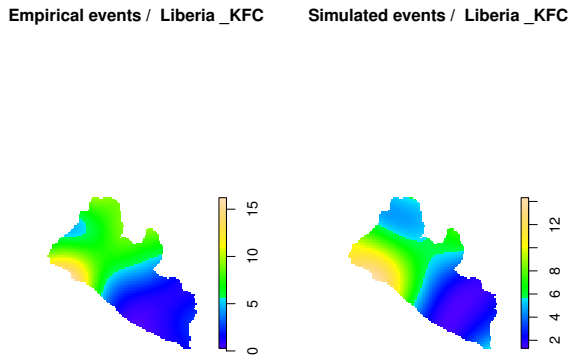
(3)

Figure 4.6: From top to bottom: Visual comparison between empirical and simulated event densities in-sample (1) and out-of-sample (2). Areas that were predicted correctly at a given confidence interval both in-sample and out-of-sample (3).

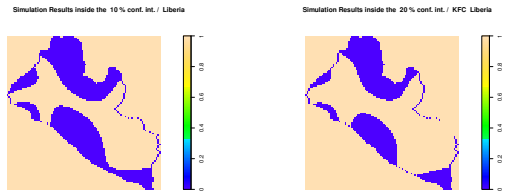
Liberia



(1)



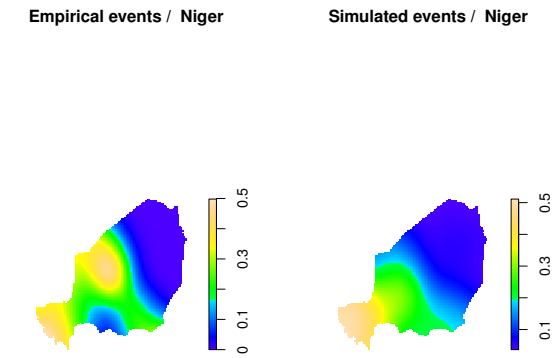
(2)



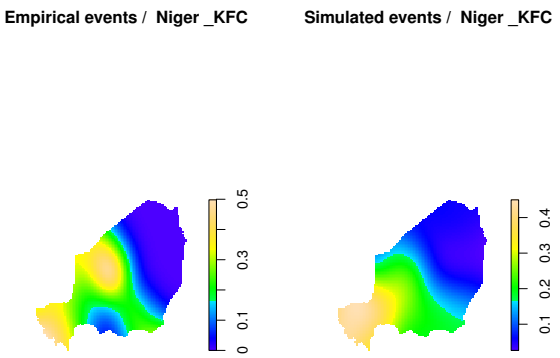
(3)

Figure 4.7: From top to bottom: Visual comparison between empirical and simulated event densities in-sample (1) and out-of-sample (2). Areas that were predicted correctly at a given confidence interval both in-sample and out-of-sample (3).

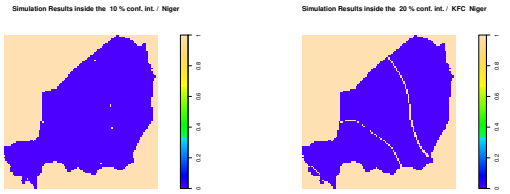
Niger



(1)



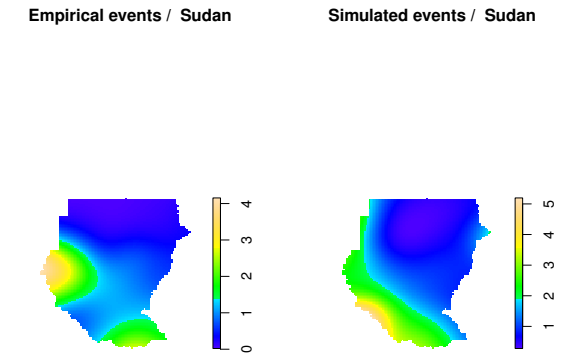
(2)



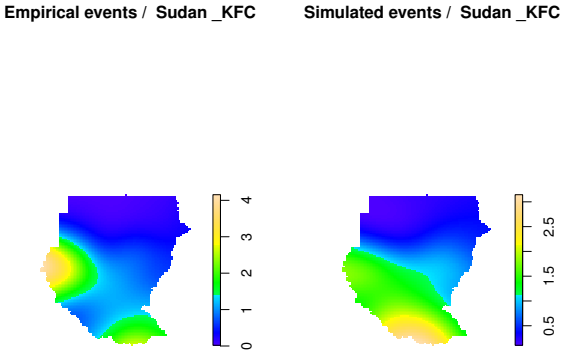
(3)

Figure 4.8: From top to bottom: Visual comparison between empirical and simulated event densities in-sample (1) and out-of-sample (2). Areas that were predicted correctly at a given confidence interval both in-sample and out-of-sample (3).

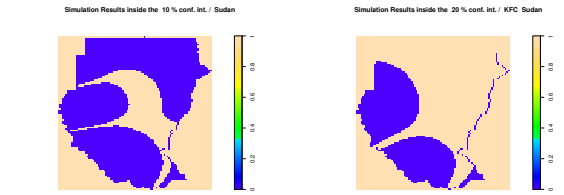
Sudan



(1)



(2)



(3)

Figure 4.9: From top to bottom: Visual comparison between empirical and simulated event densities in-sample (1) and out-of-sample (2). Areas that were predicted correctly at a given confidence interval both in-sample and out-of-sample (3).

5 Conclusion and Discussion

5.1 Discussion

The presented study provides new insights into the spatial dimensions of civil conflict in Africa. Based on theoretical expectations that pay tribute to the tactics of guerrilla warfare, a formal model was designed to explain the density of conflict events in specific regions. Similarly to t-statistics in econometric analysis, the Kolmogorov-Smirnov test was used to evaluate the statistical significance of the explanatory covariates and the results deserve to be discussed in more detail. After that, a discussion of the predictive performances and their limitations will follow.

Theoretical Expectations and Statistical Findings

The theoretical expectations as presented in Section 2.4 could be largely confirmed on the basis of the statistical analysis. Table 4.3 on page 44 shows the generally positive effect of population density on the probability of conflict events. This observation is well in line with Hypothesis 1. Similarly, the negative effect of distance to the nearest border on the probability of conflict events is shown in the Table, confirming Hypothesis 2. This latter finding confirms the observation by Hegre, Østby and Raleigh (2009) for the case of the Liberian civil war. Hypotheses 3a and 3b cannot be confirmed on the

basis of the presented study. Distance to the capital is found to be a significant factor in the prediction of conflict events in three out of eight tested cases. Nevertheless, the main goal of the insurgency in terms of secession or overthrowing the government does not reliably result in a positive or negative effect of distance to the capital city. Even though the average distance to the capital across all events might be driven by the political incompatibility, this logic cannot be reversed to explain the location of a single event.[1] A more fine-grained analysis of local conditions is necessary to fully explain the occurrence of violent clashes. Hypothesis 4 refers to the negative effect of distance to cost-optimal center-periphery paths on the probability of conflict events. This hypothesis is largely confirmed by the analysis with the exception of Kenya. The remaining hypotheses, 5a and 5b, cannot be considered confirmed or rejected by the analysis. Although the proximity of diamond mines is found to be a significant indicator in four out of six cases, two of the corresponding coefficients are positive and the remaining two are negative. Similarly, the Gross Cell Product does not allow for a reliable prediction of conflict events. In only one out of eight tested cases was the measurement found to be significant, and the estimated coefficients are extremely small. Therefore, no definitive conclusion can be drawn on the role of wealth as a spatial predictor for conflict events based on this study.

All in all, the focus on guerrilla tactics, which determine a specific kind of terrain utilization and target locations, helped to identify meaningful covariates. A rejection of willingness-related explanations is impossible within this study for two reasons. First and most importantly, failed rejections of null hypotheses never translate into confirmations, as mentioned before in chapter 2. This is especially true if the underlying data is as coarse as the Gross Cell Product by Nordhaus et al. (2006), featuring only one data point

[1] Two sources were used to identify and cross check the political aims of the civil wars under investigation. The Armed Conflict Dataset (Gleditsch et al., 2002) holds information about the main incompatibility for the selected cases. Moreover, Notholt (2008) provided summary accounts for many recent wars in the region.

for more than 100 square kilometers. Second, the behavioral logic of insurgency does not always allow for a side-by-side comparison of willingness and opportunity. With the goal of controlling wealthy regions in the long run, rebels might fight an irregular war anywhere in the country. The political aims of civil conflict do not translate as such into military operations. Cuban communists, Afghan mujahideen, and Ugandan child soldiers might display similar tactical behavior on the battlefield while following different ideologies. This fact can be utilized for the prediction of conflict zones beyond the horizon of empirical evidence, as demonstrated in terms of predictive evaluation of k-fold-cross validation in this study.

In three out of eight cases, the applied statistical test found terrain accessibility to be a significant predictor for conflict events. Apart from that, all fitted coefficients have a negative sign, which is well in line with Hypotheses 6. The assumed positive effect of terrain accessibility translates into a negative sign due to the underlying data. The global map of accessibility stores traveling times in each cell, so that high values indicate rough terrain or the lack of infrastructure.

Evidence and Beyond: The Predictive Performance

Both in-sample and out-of-sample, the applied method produces more or less accurate predictions. As the visual predictions in Section 4.3.3 indicate, the general "conflict zone" as described by Buhaug and Gates (2002) is predicted correctly in six out of eight cases (with Ethiopia and Sudan being the incorrect predictions). The magnitude of conflict densities is not always captured correctly, but the peak areas correspond fairly well to the empirical findings. The success in out-of-sample predictions is largely explained by fairly homogeneous signs for five important covariates across almost all cases: Population and loss of strength have a positive impact on conflict event probability (in 7/8 and 6/8 cases,

respectively). Distance to optimal paths and nearest borders are associated with a negative effect (7/8 cases each). Moreover, terrain accessibility is found to have a positive effect in all cases. The derived average model for the out-of-sample prediction also displays these properties which seem to be robust characteristics of the underlying processes. This general finding allows for the identification of regions at risk beyond the scope of the empirical record and can inform policy makers and scholars alike.

What Drives the Results?

Although the overall results of this study are encouraging, a closer look at the civil war cases is needed to understand why certain countries are approximated so much better than others. Both the statistical results and the theoretical considerations in Section 1.2 indicate that most violence in today's Democratic Republic of the Congo, Burundi, and probably Rwanda can be understood in the context of the Second Congolese War. With a complex actor topology, a variety of material and ideological agendas, and a vast array of proxy battles, the post-2003 situation in the region is far from stable. In this largely contingent situation, robust spatial patterns of conflict events emerge in areas of either strategic importance or tactical opportunity. Cross-border attacks and operations in remote areas of war-torn countries mark one essential element of this type of political violence. Moreover, control over populated areas and roads of strategic importance seem to be general military goals of the warring parties. For the Democratic Republic of the Congo, Burundi, Kenya, and Sudan, the employed test of statistical significance revealed the importance of the corresponding covariates. Consequently, the graphical predictions for the Democratic Republic of the Congo, Burundi, and Kenya correspond well to the empirical findings with regard to the centers of armed conflict and high intensity conflict zones. The prediction for Sudan is not fully accurate in that it unifies the two separate

conflict zones. Similarly, the situation in Ethiopia is not captured correctly, since the hot spot in the eastern part of the country is largely unaccounted for in the prediction.

Sudan and Ethiopia deserve a closer look as outlier cases in the analysis. Ethiopia's eastern Ogaden province has been the scene of intense cultural and military tension for decades. With the population being predominantly Somali, an unsuccessful attempt was made to incorporate the territory in a greater Somalia during the Ogaden War of 1977/78 (Tareke, 2000, 635). The stunning military defeat of the Somali regular forces did not end the political demand for removing Ethiopian reign over the territory, nor did it prevent guerrilla forces from carrying out further attacks in the region. In other words, "Ethiopians won the war, but not the peace" (Tareke, 2000, 663). The irredentist Somali claim proved to be harder to silence than the arms of their regular forces. Major counterinsurgency efforts were unable to stop the fighting for several years. A recurrence of the hostilities in terms of irregular forces fighting Ethiopian rule took place in 1995. The Ogaden National Liberation Front (ONLF) emerged as a new political actor and recruited many of the former rebels in the province. During the entire last decade, irregular forces of the ONLF fought a guerrilla campaign with the goal of removing Ethiopian rule from the Ogaden province. This political agenda has left its spatial footprint in the empirical record in terms of a highly localized conflict zone within the contested territory. The statistical prediction failed to account for the longstanding ethno-nationalist claim of localized pro-Somali rebels.

Similarly, the single-zone prediction for Sudan fails to account for two separate conflict zones emerging from two separate conflicts, while correctly predicting the approximate area of the fighting. These two conflicts are the civil war in Darfur in the east and the cross-border attacks by the LRA in the south (Notholt, 2008). The former conflict started in 2003 and has since spread into neighboring Chad and the Central African Republic.

Independent of this development, Ugandan LRA rebels have repeatedly extended their operations into southern Sudan (Schomerus, 2007). Separated by military scenario and geographic location, the two conflicts in Darfur and South Sudan have resulted in concentrated conflict zones. The latter example – cross-border attacks from a neighboring country – is more in line with the theoretical considerations.

Both Ethiopia and Sudan therefore yield scenarios that are delimited within specific geographic regions for reasons that are outside the introduced considerations. Darfur and Ogaden are high-intensity conflicts, delimited to contested soil. This makes it hard to capture the exact zones of conflict in a spatial model that is uninformed by history. Nevertheless, the majority of the conflict zones are predicted correctly based on the general theoretical assumptions.

5.2 Outlook

The presented study has shown that a simple spatial model can be fitted to approximate an empirical distribution based on a set of theoretically derived covariates. Adding point-to-point interaction in the future and a temporal component could increase our understanding of how civil wars unfold in space. In addition to the point pattern analysis techniques pioneered by Ripley (1977), spatial econometrics could help us to understand the importance of the different covariates in bringing about conflict events. Ward and Gleditsch (2008) provided a suitable framework for adding spatial dependencies to econometric analysis.

Since porous borders and mobile rebel organizations are taken into account in the underlying theory, future versions of the model should also account for cross-border diffusion of violence. Instead of assuming a “closed polity” for every case, a broader focus on the situation on both sides of the border could be part of both model and theory. A unified risk map for the entire region would certainly be an interesting application for the

approach.

The Achilles' heel of any project that aims at explaining conflict events remains the underlying data: ACLED only covers a very small fraction of the actual events, and its sole reliance on international media sources might lead to an urban reporting bias. In a worst case scenario, ACLED might primarily inform scholars about where conflict events are reported from instead of where they actually happened. Assessing this bias will be necessary for future extensions of the model. Cross checking the results with an additional event dataset could be a suitable next step.

The general formal approach of this study could be used to inform practitioners in the future. Near real-time incidence reporting as pioneered by the ushahidi project (`www.ushahidi.com`) might allow for the fast build-up of spatial risk models in emerging crises. Extrapolating from these models could help guide the activities of relief operations. From a more academic point of view, the introduced covariates could be combined with additional information about natural resources or contested territory to further account for the political motivations of the actors involved. The overall scheme of fitting models and testing them in out-of-sample predictions certainly allows for the inclusion of additional data, although a more sophisticated approach might be needed to account for spatio-temporal dependencies. Graphical predictions and out-of-sample testing enables an intuitive side-by-side comparison of different formal models. Such an intuitive and scalable approach could shorten innovation circles in an otherwise fragmented methodological landscape. The increasingly accurate prediction of regions at risk will hopefully inform scholars and decision makers alike in the future.

Bibliography

Baddeley, A. 2008. "Analysing spatial point patterns in R, Workshop Notes, Version 3.".

Berman, M. 1986. "Testing for Spatial Association Between a Point Process and Another Stochastic Process." *Applied Statistics* (35):54–62.

Berman, M. and T. R. Turner. 1992. "Approximating point process likelihoods with GLIM." *Applied Statistics* 41:31–38.

Biddle, S. 2006. *Military Power*. Princeton University Press.

Boulding, K. 1962. *Conflict and Defense: A General Theory*. Harper.

Buhaug, H. 2008. "Dude, Where's My Conflict? LSG, Relative Strength, and the Location of Civil War." Paper presented at the 48th International Studies Convention, Chicago, IL, USA, 28 FEB - 3 MARCH, 2007.

Buhaug, H. and J-K. Rød. 2006. "Local Determinants of African Civil Wars, 1970-2001." *Political Geography* 25:315–335.

Buhaug, H. and K. S. Gleditsch. 2008. "Contagion or Confusion? Why Conflicts Cluster in Space." *International Studies Quarterly* 52:215233.

Buhaug, H. and S. Gates. 2002. "The Geography of Civil War." *Journal of Peace Research* 39:417–433.

Cederman, L.-E. and L. Girardin. 2007. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies." *American Political Science Review* 101(1):173–185.

Collier, P. and A. Hoeffler. 2004. "Greed and Grievance in civil war." *Oxford Economic Papers* 56:563–595.

Diehl, P. F. 1990. "Geography and war: A Review and Assessment of the Empirical Literature." *International Interactions* 17(1):11–27.

Dijkstra, E. W. 1959. "A Note on two Problems in Connection with Graphs." *Numerische Mathematik* 1:269–271.

Fearon, J. D. and D. D. Laitin. 2003. "Ethnicity, Insurgency and Civil War." *American Political Science Review* 97(1):75–90.

Gellner, E. 1983. *Nations and Nationalism*. Cornell University Press.

Gilmore, E., N. P. Gleditsch, P. Lujala and J. K. Rød. 2005. "Conflict Diamonds: A New Dataset." *Conflict Management and Peace Science* 22(3):257–292.

Gleditsch, K. S. 2007. "Transnational Dimensions of Civil War." *Journal of Peace Research* 44(3):293–309.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg and H. Strand. 2002. "Armed Conflict 1946-2001: A New Dataset." *Journal of Peace Research* 39(5):615–637.

Guevara, E. C. 1961. *Guerrilla Warfare: Authorized Edition*. Ocean Press.

Gurr, T. 1970. *Why Men Rebel*. Princeton University Press.

Harbom, L. and P. Wallensteen. 2005. "Armed Conflict and its International Dimensions, 1946-2004." *Journal of Peace Research* 42(5):623–635.

Hegre, H., G. Østby and C. Raleigh. 2009. "Poverty and Civil War Events: A Disaggregated Study of Liberia." *Journal of Conflict Resolution* 53(4):598–623.

Kalyvas, S. 2001. ""New" and "Old" Civil Wars: A Valid Distinction?" *World Politics* 54.

Kalyvas, S. 2005. Warfare in Civil Wars. In *Rethinking the Nature of War*, ed. Duyvesteyn I. and J. Angstrom. Abingdton pp. 88–108.

Keita, Kalifa. 1998. "Conflict and conflict resolution in the Sahel: The Tuareg insurgency in Mali." *Small Wars and Insurgencies* 9(3):102–128.

Mitchell, T. M. 1997. *Machine Learning*. McGraw-Hill International Editions.

Nelson, A. 2008. "Estimated travel time to the nearest city of 50,000 or more people in year 2000.".

NIST/SEMATECH. 2003. "e-Handbook of Statistical Methods." http://www.itl.nist.gov/div898/handbook/.

Nordhaus, W., Q. Azam, D. Corderi, K. Hood, M. N. Victor, M. Mohammed, A. Miltner and J. Weiss. 2006. "The G-Econ Database on Gridded Output: Methods and Data.".

Notholt, S. 2008. *Fields of Fire – An Atlas of Ethnic Conflict*. BM Atlas.

Prunier, G. 2009. *Africas World War*. Oxford University Press.

Raleigh, C. and H. Hegre. 2005. "Introducing ACLED: An Armed Conflict Location and Event Dataset.".

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

Reyntjens, F. 2001. "Briefing: The Democratic Republic of Congo, from Kabila to Kabila." *African Affairs* 100:311–317.

Ripley, B. D. 1977. "Modelling Spatial Patterns." *Journal of the Royal Statistical Society* 39(2).

Salehyan, I. and K. S. Gleditsch. 2006. "Refugees and the Spread of Civil War." *International Organization* 60(2):335–366.

Sambanis, N. 2003. "Expanding Economic Models of Civil Wars Using Case Studies." *Perspectives on Politics* (10).

Schomerus, M. 2007. "The Lord's Resistance Army in Sudan: A History and Overview." *HSBA Working Paper 8 / Small Arms Survey* .

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Starr, H. 1978. "'Opportunity' and 'Willingness' as Ordering Concepts in the study of War." *International Interactions* 4(4):363–387.

Tareke, G. 2000. "The Ethiopia-Somalia War of 1977 Revisited." *The International Journal of African Historical Studies* 33(3):635–667.

von Dach, H. 1957. *Total Resistance: A Guerrilla War Manual for Everyone*.

Waltz, K. 1959. *Man, the State, and War*. Columbia University Press.

Ward, M. D. and K. S. Gleditsch. 2008. *Spatial Regression Models*. Sage.

Wimmer, A., L.-E. Cederman and B. Min. 2009. "Ethnic Politics and Violent Conflicts, 1946-2005: A Configurational Approach." *American Sociological Review* 74(2):316–337.