

# Population size, concentration, and civil war. A geographically disaggregated analysis

Clionadh Raleigh<sup>a,\*,c</sup>, Håvard Hegre<sup>b</sup>

<sup>a</sup> Trinity College Dublin, International Peace Research Institute/Center for the Study of Civil War, Ireland

<sup>b</sup> University of Oslo, International Peace Research Institute/Center for the Study of Civil War, Norway

## ABSTRACT

**Keywords:**  
Civil war  
Central Africa  
Political geography

Why do larger countries have more armed conflict? This paper surveys three sets of hypotheses put forward in the conflict literature regarding the relationship between the size and location of population groups. These hypotheses are based on population mass and concentrations, distances, and some residual state-level characteristics. The hypotheses are tested with a new dataset – ACLED (Armed Conflict Location and Events Dataset) – that disaggregates internal conflicts into individual events. The analysis covers 14 countries in Central Africa. The conflict event data are juxtaposed with geographically disaggregated data on populations, distance to capitals, borders, and road networks. The paper develops a statistical method to analyze these types of data. The analysis shows that the risk of conflict events increases with local population size, and is particularly large in highly populous locations within a country. We find only a moderate difference in conflict-proneness between areas located close to or far from a country's capital, except when combined with population concentration: Conflict events cluster particularly strongly in larger population concentrations that are distant from the capital, such as the Eastern provinces of the Democratic Republic of Congo.

© 2009 Elsevier Ltd. All rights reserved.

## Civil war and country size

The most robust empirical finding in country-level studies of civil war is that large countries have more civil war than small countries (Collier & Hoeffler, 2004; Fearon & Laitin, 2003; Hegre & Sambanis, 2006). A country with a population of 10 million inhabitants has an estimated risk that is twice as high as one of 1 million inhabitants.<sup>1</sup>

As Sambanis (2003) points out, however, there is little agreement as to why large countries have more civil war than small ones (2003: p. 26). The conflict literature has suggested several competing explanations for this relationship. Are large countries more conflict-prone because of the difficulty of projecting power over long distances? Do large countries experience governance problems due to the multiple layers of authority necessary to control populations? Do conflicts stem from the ethno-regional heterogeneity typical of large countries? Or is it simply because there are so many people with the potential to start a bloody quarrel?

In most previous studies, country size has been measured in terms of the size of the population – a notable exception is the study of Buhaug and Rød (2006). But the sample of explanations above demonstrates that there are several relevant manifestations of size that are closely related at the national level: Most fundamentally, countries may be large in terms of population or in terms of territory. But countries may also be large in terms of the number of distinct cultural or ethnic groups living in the territory, or the distances over which a government must be able to deploy forces, or the length of an international border. It is difficult to distinguish between these variables in empirical studies at the national level, since they are highly correlated when aggregated. The reliance on national-level studies presents an ecological inference problem as the nature of populations and population density in particular areas is assumed to be homogeneous across a state. In this study, we measure both the dependent variable of conflict occurrence and the measure of population density. Our unit of analysis is a 'grid square' with size  $8.6 \times 8.6$  km. The precise description is given below. In Section 2, we refer to such grid squares as 'locations'. The ecological inference issue is alleviated as we directly test the propensity to experience conflict of any population group resident in a location.

If we know at which locations conflicts occur, we may succeed in supporting explanations based on variables such as the distance from the capital and the overall size of the country's population. If conflicts are located mainly at some distance from countries'

\* Corresponding author. Tel.: +353 1 896 3482.

E-mail address: [raleigh@tcd.ie](mailto:raleigh@tcd.ie) (C. Raleigh).

<sup>c</sup> Dr. Raleigh funded by Irish Research Council for the Humanities and Social Sciences.

capitals, we might infer that large countries have more conflicts because of the difficulties of projecting governmental power. If they are located in population concentrations irrespective of location relative to the capita, other explanations should be sought.

Related to the size of populations is its distribution. For instance, the Democratic Republic of Congo is not only characterized by being enormously large, but also shows tremendous variation in population density. The capital is located in the far west of the country, and there is a large population concentration in the far east. Our disaggregated research design also allows exploring how such patterns of population dispersion – and the resultant political geography of states – affect the risk of conflict events.

Our analysis also addresses the larger issue of how conflict formations can be predicted and managed. We do not suppose that population clusters are an underlying cause of conflict, but that particular areas of a state may be more or less susceptible to civil war violence due to the local characteristics and contexts. Critical population masses are attractive to both rebels and governments for public support, economic and strategic importance. Grievances applied to larger populations have greater impact than those experienced by distant and small groups. This analysis means to clarify the weight of these demographic geographies on the propensity for conflict.

The paper uses a new dataset called ACLED (Armed Conflict Location and Events Dataset) that is designed for this type of disaggregated analysis. The dataset codes the location of all reported conflict events in 14 countries in Central Africa in the 1960–2004 period. Central Africa is chosen for three reasons: Many of its component states have experienced multiple conflicts since Independence; the included states are of various sizes, population masses and densities, providing a microcosm of the continent at large; and finally, ACLED has collected conflict information on all of the included states. We believe that our framework, model and conclusions can be extended beyond Central Africa, and that it presents a rigorous test of our hypotheses.

The conflict event data are juxtaposed with geographically disaggregated data on populations, distance to capitals, borders, and road networks. The paper suggests a framework to analyze statistically data at this level of analysis that avoids that temporal and statistical dependence between observations. This model focuses on the more static aspects of controlling for the (dynamic) history of the conflict. The propositions developed below, then, state how population size and distribution affects when and where conflicts begin, as well as where they expand.

#### *Population, geography and conflict: location-specific factors*

We will look into three groups of size-related explanations of civil war. The first simply posits that every citizen has a constant and uniform propensity for becoming involved in a rebellion. The second relates to heterogeneity and military constraints posed by distance. The third set of explanations highlights the importance of population concentrations.

We will review the arguments and formulate empirical implications for them at the appropriate level of analysis: either at the local level (for a village or a small piece of territory), or at the national level. The empirical analysis will attempt to discriminate between the explanations by testing them at the local level and see how much variance between locations is left to be explained by the national-level population size.

#### *Population-mass explanations*

The simplest explanation of the national-level relationship between population size and the risk and extent of conflict is based

on the assumption of a constant and homogenous ‘per-capita conflict propensity’. If there is a given probability that a randomly picked individual starts or joins a rebellion, then the risk of rebellion increases with population. Collier and Hoeffler state the ‘per-capita propensity’ mechanism explicitly:

“Population is likely to be correlated with conflict risk. If two identical areas, each with a conflict risk of  $p$ , are treated as a single area, the conflict risk for the single area rises to  $2p - p^2$ . Since  $p$  is small [...], this effect alone would yield an elasticity of conflict risk with respect to population of slightly less than unity” (2002: 11).

There are several mechanisms through which a constant per-capita risk of rebellion may emerge. First, a potential rebel group leader can only recruit up to a certain fraction of a population. The larger the recruitment pool, the greater the chances of recruiting a sufficiently large group to initiate a rebellion. To the extent that rebellions occur in the recruits’ home locations, the risk of conflict events should be proportional to the population mass at that location. Second, even when insurgents operate away from their home locations, they are likely to target locations that can provide supplies such as food, tools, etc. Such supplies are, on average, richer where the population is larger, as these areas are ‘economically self-sufficient’ (McCull, 1969). Third, a major part of a rebel group’s strategy is to inflict damage upon the government. The group will therefore target the most valuable locations controlled by the government. The rebels will attempt to hinder the government from benefiting economically from the large support base in a populous area, or make the government suffer the psychological and reputational loss from demonstrating an inability to control urban areas.

At the location (grid square) level, the implication of the population-mass explanations is:

**Proposition 1 Constant Per-Capita Risk of Conflict:** The risk of civil war events at a location increases with the size of the population at that location but it is independent of the size of the national population of the country to which the location belongs. The ‘constant per-capita propensity’ explanation implies that a continent with few countries would have the same number of conflict events as a similarly sized continent with many countries. At the national level, it suggests that the risk of conflict in a country is exactly proportional to the size of its total population.

A weaker formulation of the population-mass explanation states that the country-level conflict risk increases, but not proportionally with population. In typical logistic cross-national regression models, estimates for a log population variable are around 0.3 (e.g. Hegre & Sambanis, 2006) where the ‘elasticity’ of civil war risk with respect to population is 0.3: a 1% increase in population leads to a 0.3% increase in the risk of conflict onset.<sup>2</sup> This means that at the country level, the per-capita risk decreases with country size: a country the size of Nigeria has an estimated risk that is three times higher than a country the size of Liberia, although the population of Nigeria is twenty times that of Liberia. At the location level, the elasticity of civil war event risk may also be less than one, hence not strictly proportional to absolute population density. Below we test whether the per-capita risk of conflict is constant or decreasing.

**Distance, transportation costs, and borders:** The probability of conflict events at a location is likely to be dependent on where the location is situated relative to the capital, a rebel group’s headquarters, and to international borders. Two aspects are particularly important: The relationship between relative locations and preferences, and military factors. We will discuss these in turn.

**Heterogeneity of preferences:** Alesina and Spolare (2003: pp. 40–45) develop a model based on the assumption that there are

economies of scale in the production of public goods and that the utility individuals derive from the public good decrease with distance. Distance is conceived of in terms of preferences and of physical distance, since cultural and economic factors such as religion, ethnic affiliation, or dominant occupation tend to be geographically clustered. These background factors determine at least a part of an individuals' preferences. Hence, geographical distance between two populations is correlated with distance in terms of preferences regarding public policies. Since populous countries are often geographically extensive, they are also likely to be heterogeneous. Alesina and Spolaore argue that economies of scale allow large countries to provide better public goods, but at a given distance from the center of the country, the distance in terms of preferences outweighs the efficiency of the large government. Beyond this distance, population groups will have an incentive to secede. Although not mentioned explicitly by Alesina and Spolaore, such secession attempts may turn into violent conflicts.

Geographic peripherality is often linked with ethnic and political peripherality (see Gurr, 1970). If civil wars are caused by differences over public policies, we may infer that insurgencies should originate in locations far from the capital. Even when the conflict is purely over political power per se, rebel groups are likely to attempt to exploit a local population's resentment toward the government to win their 'hearts and minds'.

**Military factors:** The dispersion of population also has military implications. Although there clearly are economies of scale in defense, (Collier & Hoeffler, 2002: p. 15; 2004: p. 572) these are counteracted by the challenge of controlling large territories. States with limited reach may not be able to control activities in territories beyond the established infrastructure of the state (Clapham, 1986; Gurr, 1970; Herbst, 2000). Warfare becomes more difficult as the front extends and transportation takes longer and requires better organization, supply lines become more vulnerable to guerrilla attacks, and the local population may become more hostile the further from the capital they are located. Even though a government may establish military bases throughout the territory to minimize the impact of distance, vulnerable supply lines support these bases, and the bases themselves become targets for rebel group activities. Locations far from the capital, then, may be conflict-prone both because the population is likely to have preferences that diverge markedly from the government and because it is difficult for the governments to control distant territories.<sup>3</sup>

Lichbach (1995: p. 156) details the role of geographic isolation for dissident communities as one that results from distance, poor transportation, inaccessible terrain, and fluid boundaries. In addition, it should follow that as distance from national authorities increases, the collective dissent should increase – and states with poor transportation networks should experience higher levels of dissent. Relatedly, Herbst (2000) describes the political geography of a small country as 'favorable' to a government.

Yet, the nature of and logic behind territorial control during civil wars is actively debated. Kalyvas (2006) claims that the distribution of territorial control during war is a function of strength and success where "military effectiveness as determined by geography generally trumps pre-war political and social support in spawning control....incumbents tend to control cities, even when these cities happen to be social, religious, or ethnic strongholds of their opponents, whereas the insurgents' strongholds tend to be in inaccessible rural areas, even when the rural populations are inimical to them" (Kalyvas, 2006: p. 133). Further, he asserts that "rurality" is a proxy for various causal mechanisms, including the ability of combatants to hide without being denounced because of rural norms of solidarity and honor; higher levels of tolerance among rural people to threats of violence; a tradition of rebellion

reinforced by the norms of reciprocity, which leads to mass participation in anti-state activities ranging from contraband to banditry to full-fledged rebellion; and the fact that an economy based on subsistence farming tends to favor armed resistance more than one based on wage labor" (Kalyvas, 2006: p. 136). McColl, however, notes that many practitioners of civil war regarded the physical geography of a country of secondary importance to specific pre-war political and social contexts in determining the location and establishment of rebel bases (1967, 1969).

The heterogeneity and military explanations have the same empirical implication:

**Proposition 2 Distance from Capital:** The risk of civil war events at a location increases with the distance from the location to the capital of the country.

Related to this aspect of geographical distribution is the importance of national borders. rebel groups may operate more easily in border areas since neighboring countries may provide (actively or tacitly allow) safe zones for rebels. Again, it is impossible to distinguish fully between military and cultural factors as conflicts that begin in border areas may also be linked to irredentist movements in defiance of, for instance, the state's ethnic character.

**Proposition 3 Distance from Border:** The risk of civil war events at a location is higher in border zones.

A third location aspect that impacts where conflicts take place is the location of rebel group headquarters. Even though rebel groups often rely on hit-and-run strategies, they are more constrained by logistics and find it difficult to operate far from their headquarters. Locations far from rebel group headquarters should therefore experience fewer conflict events. In fact, rebel groups are likely to be more constrained by distance than governments since governments typically have had the opportunity to establish military bases throughout the territory before the conflict started.

**Proposition 4 Distance from Rebel Group Headquarters:** The risk of civil war events at a location decreases with the distance from the location to the rebel group's headquarters.

#### *Population concentration*

A third set of explanations relate to population concentrations: The conflict-proneness of a local population may be larger the more concentrated the population is in the neighborhood. In addition to the size or mass of the population in the location, the size of the population in the immediate neighborhood is important. There are at least three reasons why this might be so: First, population concentrations help solve coordination problems; second, population concentrations may be more homogenous; and third, concentrated populations are more autonomous than dispersed ones.

Prospective rebel groups face trade-offs between economies of scale and geographical extent just as governments do. Collier (2000) and Lichbach (1995) stress the importance of the collective action and coordination problems prospective rebels face. Accordingly, Lichbach (1995) contends that as the geographic concentration of dissidents increases, collective dissent should also increase. This is directly due to the ability of dissidents to communicate, coordinate mutual expectations, and reduce organizational costs: 'As the concentration of dissidents increases, the extent and intensity of interactions among dissidents increases, which in turn increases their communications (e.g. of grievances). [...] [W]ith reduced distance between dissidents, it is easier to administer rewards for compliance and punishment for noncompliance' (Lichbach, 1995: pp. 158–159).

In the absence of community, or rebels socialized into common norms (1995: p. 126), Lichbach argues that rebels will employ a 'contract' as one way to overcome the 'rebel's dilemma'.

Communities that are autonomous, stable and concentrated can forge a number of different contracts to assure collective dissent. Of those types of contracts, homogeneity in social background allows for lower transaction costs and hence increased cooperation for joint collective dissent – ‘homogeneity, moreover, facilitates the development of information, trust and norms, and hence reduces the bargaining, monitoring, and enforcement costs of social contracts’ (p. 139). As noted above, homogeneity in social background is also a function of geographical concentration. Heterogeneity is likely to vary more with geographical distance than with population size. This serves to increase the importance of local population concentrations. Herbst (2000: pp. 152–154) argue that ‘hinterland’ countries which are large but where most of the population is concentrated in a small number of areas have a relatively low conflict propensity. In the hinterland where population density is low, governments can relatively easily control it. Herbst contends that most political battles will occur in capital areas since this is where the important stakes are located.

Densely populated areas are often so because they are resource-rich at the outset or are strategically located, e.g. close to a major harbor or a navigable river. Moreover, economies of scale often help such locations to become even richer. Population concentrations mark locations that are particularly valuable both for the rebels and for the government, and should attract conflict events if hostilities break out. Distance from the capital and population concentrations are factors that are likely to reinforce each other with respect to risk of conflict events.

Yet concentration can also work against a movement as nationally based dissident movements, or small movements in large countries, are more prone to failure because of their inability to permeate the remainder of the state. Lichbach (1995: p. 160) also notes that a ‘wide geographic scope can work to the advantage of a dissident movement as it works against the government’s ability to repress’. The perceived increase of collective dissent in urban areas is related to the concentration argument as the ability of dissidents to organize is lessened due to proximity. The fewer urban areas in which dissents may gather, the higher the collective dissent in those areas (Tilly, 1964 as cited in Lichbach, 1995: p. 162).

These arguments have the same empirical implication as posited in Proposition 1. In addition, however, given our fine-grained dataset, locations in highly concentrated areas such as large cities and fertile valleys are characterized by also having large populations in the neighboring locations. Hence:

**Proposition 5 Population Concentration:** The risk of civil war events at a location increases in the size of population in the immediate geographical neighborhood.

#### *Population concentration and distance-related factors*

‘Political geography’ in part refers to how a government controls its territory. The political geographies of developing countries are often at odds with traditional understandings of sovereignty. Herbst (2000) describes the political geographies of states with large territories and non-contiguous areas of high population as the most difficult to control for a government, and Lichbach (1995) argues that the combination of large distances to the capital and population concentrations particularly facilitate dissent. The population geography in such countries can, in some cases, be associated with ethnic complexity, as 70% of African ethnic communities live in spatially distinct ethnic ‘homelands’ (Scarritt & McMillan, 1995). Further, such population patterns are at odds with typical understandings of territorial sovereignty – these states are the most susceptible to fracturing as an uncontrolled population can choose to deny the legitimacy of governments (Herbst, 2000: p. 146).

A few empirical studies find evidence that concentration and distance reinforce each other. Toft (2003) investigates whether different settlement types lead to increased motives for separatist conflict by employing the MAR ethnic concentrations (urban, concentrated majority, concentrated minority and dispersed). Each group is found to have a different capability of armed rebellion. Urban groups are most able to create networks to mobilize the populous and dominate necessary resources; concentrated majorities have similar capabilities. For concentrated minorities, rebel capabilities are indeterminate and dependent on the context and region. Dispersed minorities are the weakest in terms of ability to create conditions suitable for separatist conflict.

Collier and Hoeffler (2004) test Herbst’s hypothesis in their national-level study by calculating a gini coefficient for population distribution. Countries in which the population is evenly distributed throughout the territory will have a score of 0, whereas a country where all the population is concentrated in one of their  $20 \times 20 \text{ km}^2$  will have a score of 1. They consistently obtain a negative estimate for this variable, controlling for other factors, and conclude that population concentration reduces the risk of conflict. Their measure is not likely to capture all aspects of population diffusion, however, since they cannot distinguish between a country where the population is concentrated in one cluster covering 10% of the territory and one where the population is concentrated in two clusters of 5% each, but with a considerable geographical distance between them.

Population and conflict geography in the Democratic Republic of Congo (DRC) corresponds to these arguments regarding population concentration and dispersion. Concentrations of language-based groups are evident throughout eastern DRC. Due to the limited access of the government, multiple international borders, and the dense population concentrations, certain concentrated ethnic homelands have a higher potential of conflict than other, more accessible, sparsely populated areas of DRC. Fig. 1 shows the population concentrations in 1990 (CIESIN data) for Central Africa. Heavily populated areas are shaded in darker tones. Civil conflict in DRC has overwhelmingly occurred in the eastern portion of the state, which is the most densely populated area and also geographically peripheral to the capital, Kinshasa. Of the eleven Congolese rebel groups accounted for in the dataset used in this paper, all have operated either exclusively or partially in the eastern and southern areas of DRC.

**Proposition 6 Concentration and Dispersion:** The risk of civil war events at a location increases more strongly in local population concentrations in locations distant from the capital of countries.

#### *Residual state-level mechanisms*

We have pointed out a set of location-specific factors, each of which imply a positive relationship between country size and national-level risk of armed conflict. But it is not certain that such location-specific factors are the only relevant ones. The size of a country itself may affect risk over and beyond what is implied by sheer population size, distance, and population distributions. If the economies of scale with respect to defense are sufficiently large, the risk of conflict events at a location at a given distance from the capital may be lower the larger the country (Collier & Hoeffler, 2002). Moreover, large countries may be rather more conflictual than small ones for several reasons. Fearon and Laitin (2003: p. 81) note that insurgency will be favored when potential rebels face a ‘large country population, which makes it necessary for the center to multiply layers of agents to keep tabs on who is doing what at the local level’. Furthermore, the economies of scale in defense do not unambiguously decrease with the risk of conflict, since they also raise the stakes of the political contest. Since controlling the



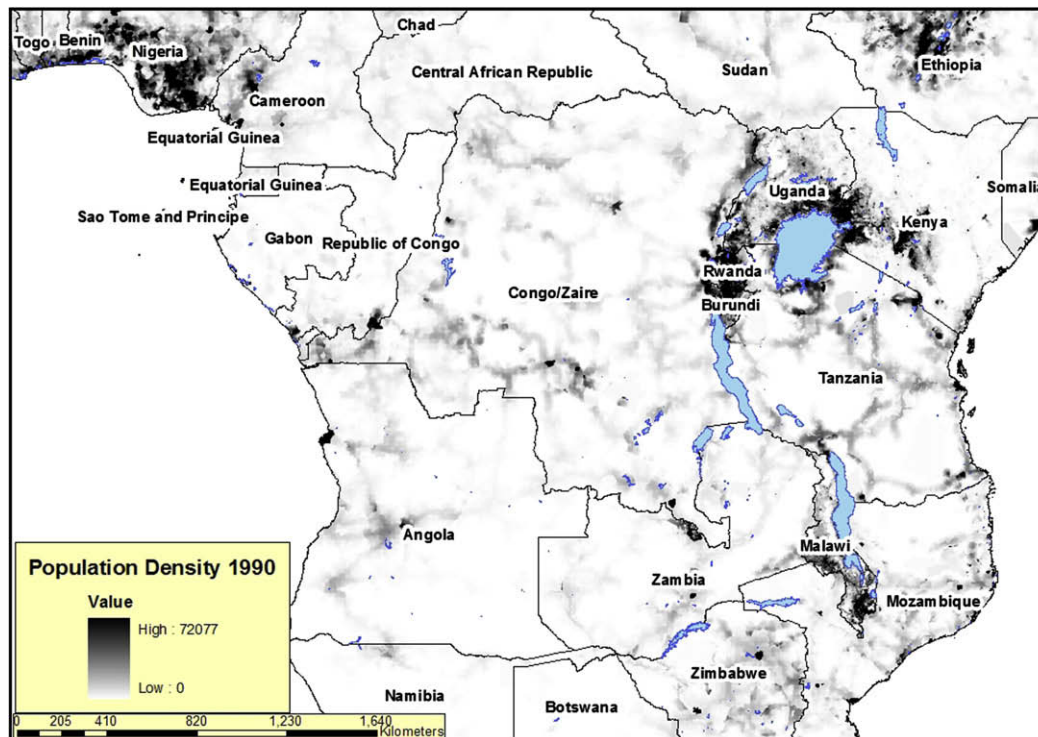


Fig. 1. Population concentration in Central Africa, 1990. Source: CIESIN.

government in a large, wealthy, and powerful country is more attractive to would-be heads of states, they will be more willing to initiate an insurgency to take control over the government.

We consequently acknowledge the possibility that the risk of conflict events at a location depends on the size of the country it is located within:

**Proposition 7 State-Level Effects of Population Size:** The risk of civil war events at a location varies with the size of the population of the country in which the location belongs, controlling for the local effects.

## Research design

### Unit of analysis

To distinguish between the different theoretical statements regarding how population sizes, population concentrations and locations relate to risk of conflict, we need to investigate exactly where conflicts occur. We have created a dataset using a Geographic Information Systems (GIS) program which converted large territories into smaller portions of  $8.6 \text{ km} \times 8.6 \text{ km}$ , totaling 74 square kilometers. These small grid squares are meant to proxy a 'locality'. Individual grid squares are our units of observation (hereafter 'squares'). We exclude all water body squares from the dataset. This approach is similar to that of Buhaug and Rød (2006), with two important differences. First, their squares are much larger ( $100 \times 100 \text{ km}$ ). Second, they code the dependent variable considerably more crudely than is done in the ACLED dataset described below. Buhaug and Rød (2006) use the 'scope' and 'location' variables in the Uppsala/PRIO dataset. Although much more suited to geographically disaggregated analysis than other datasets, this location dataset has some limitations. It does not record changes over time in the center location and extent of conflicts, and it reports the total extent of the conflict zone without distinguishing between areas that saw repeated and extensive fighting and those that only experienced scattered activities or individual events far from the center of the conflict.

### Disaggregated dependent variable: ACLED

The ACLED dataset (Raleigh, Linke, Hegre, & Karlsen, in press) deals specifically with these problems as it disaggregates conflicts into a series of singular events within larger wars. ACLED collects information on reported conflict events that occur within the context of a civil war. Reported events include battles resulting in no change of territory, battles resulting in a transfer of territory to the rebel actor, battles resulting in government forces recapturing rebel-held territory, establishment of a rebel base or headquarters, rebel activity that is not battle related (e.g. presence or recruitment), and civilian killings. Events always involve two actors, typically a rebel group and a government, although rebel-rebel battles occasionally occur. Events are discrete in that they occur at a specific georeferenced point location and on a specific day.

ACLED data are collected from secondary information sources – primarily news reports, periodical information, books, humanitarian reports and information collected from the Uppsala Armed Conflict Project archives. The data combine information from the available reported literature to determine the actors, type, locations, and dates of individual conflict events.<sup>4</sup>

The unit of observation in ACLED is the event. In Fig. 2, the ACLED data for each location of a recorded conflict event are displayed for six Central African states. In several of these locations, multiple events occurred over the 1960–2004 period covered in ACLED.

For this paper, the occurrence of one of the 2530 events in a square at a given date constitutes the dependent variable. Spatial and temporal modeling is possible, as each conflict event is georeferenced by coordinates and assigned a date of occurrence. Raleigh et al. (in press) give an explanation of ACLED data and provide some basic exploratory statistics.

The dataset used in this article covers 14 countries in Central Africa. Six had a conflict in the 1960–2004 period according to the Uppsala/PRIO Armed Conflict Dataset (Gleditsch et al., 2002):

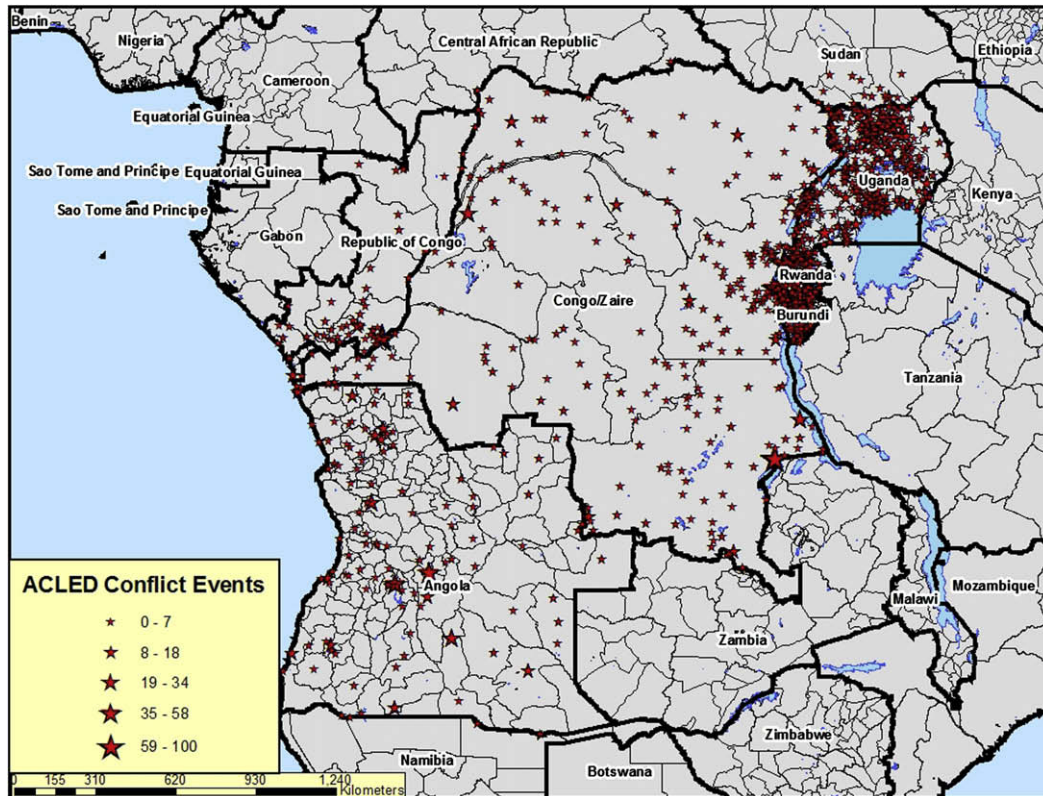


Fig. 2. ACLED conflict events in Central Africa, 1960–2004.

Angola, Burundi, Republic of Congo (Brazzaville), Democratic Republic of Congo (Zaire), Rwanda, and Uganda. The remaining 8 did not have a conflict: Cameroon, Central African Republic, Equatorial Guinea, Kenya, Malawi, Tanzania, and Zambia.

#### Statistical model to handle dependence between units

Both the squares and the conflict events are obviously not fully independent – all events within one conflict are related to each other as an action in one location leads to a later retaliation by the opposing party or to further advances in proximate locations. Events in one conflict may also affect the likelihood of other conflicts, such as the spillover of the conflict in Rwanda into Eastern DRC.

The statistical model employed to analyze these data must handle the dependence between observations. We will do this by explicitly modeling the probability of an event in a location as a function of preceding events in the same and in adjacent squares. We can do this since we know both the precise date and the precise geographic location of each event. Our information about the previous history of the conflict is coded as explanatory variables that are described in detail below.

We use an adaption of the calendar-time Cox regression model presented in [Raknerud and Hegre \(1997\)](#) for this purpose. In Cox regression, the dependent variable is the transition between ‘states of nature’ – the transition from peace to conflict in a square. A central concept is the hazard function,  $\lambda(t)$ , which is closely related to the concept of transition probability:  $\lambda(t)\Delta t$  is approximately the probability of a transition in the ‘small’ time interval  $(t, \Delta t)$  given that the subject under study is at risk of transition at  $t$ . The main idea of Cox regression is the assumption that the hazard of war  $\lambda_d(t)$  for square  $d$  can be factorized into a parametric function of

(time-dependent) variables and a non-parametric function of time itself (the baseline hazard):

$$\lambda_d(t) = \alpha(t) \exp \left( \sum_{j=1}^p \beta_j X_j^d(t) \right) \quad (1)$$

In (1),  $\alpha(t)$  is the baseline hazard: an arbitrary function of calendar time reflecting unobserved variables at the system level.  $X_j^d(t)$  is a (possibly time-dependent) explanatory variable for square  $d$ ;  $\beta_j$  is the corresponding regression coefficient; and  $p$  is the number of explanatory variables. All legitimate explanatory variables are known prior to  $t$  – they must be a part of the history up until immediately before  $t$ .

In contrast to ordinary survival analysis,  $t$  is calendar time here. The model is useful because it allows handling observations that are recorded on the finest possible time-scale to keep track of the succession of events. It also allows for non-stationarity in the underlying baseline probability of conflict events due to changes in latent variables at the system level. Such non-stationarity may be due to several causes: the end of the Cold War, changes over time in the prices and availability of arms, changes over time in the reporting of conflict events in Western media, etc.

Estimating this model involves (i) estimation of the regression coefficients  $\beta_j$  and (ii) estimation of the baseline hazard of war  $\alpha(t)$ . These two tasks are quite different, since the latter is an unknown function – not a parameter. However, for the specific purpose of inference about conflict, we are mainly interested in the ‘structural’ parameters  $\beta$ . Inferences about  $\beta$  can efficiently be made by conditioning on the time-points of outbreaks of war,  $\{t_1, t_2, \dots, t_n\}$ . This means that we can consider  $\{t_1, t_2, \dots, t_n\}$  as fixed rather than stochastic, without losing any information about the parameters.

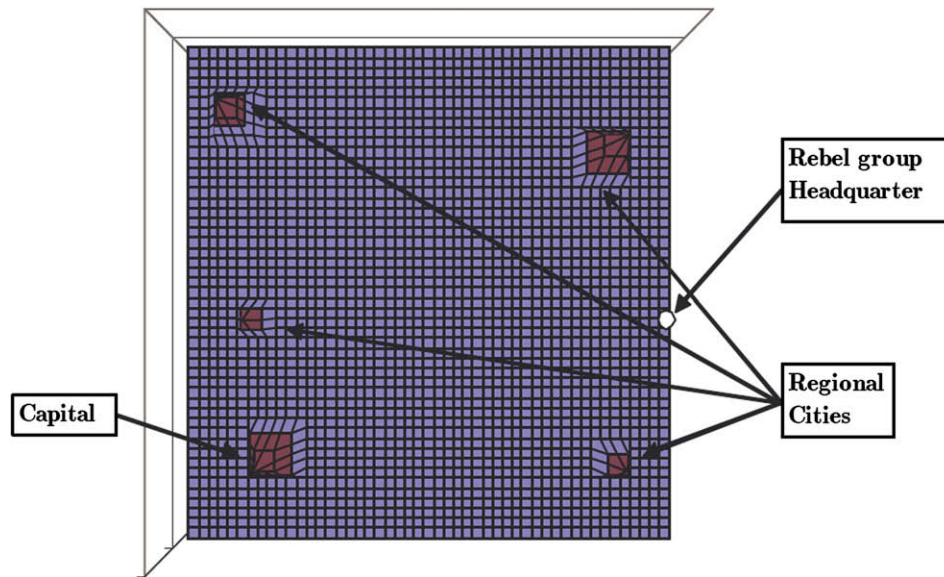


Fig. 3. Map of hypothetical country, with locations of capital, rebel group headquarters, and main population concentrations.

Given that there is an outbreak of war at time  $t_w$ , the probability that this war outbreak will happen in square  $d$  is:

$$\Pr(\text{war in a square } d | \text{a war breaks out at } t_w) = \frac{\exp\left(\sum_{j=1}^p \beta_j X_j^d(t)\right)}{\sum_{i \in R_{t_w}} \exp\left(\sum_{j=1}^p \beta_j X_j^i(t)\right)} \quad (2)$$

where  $R_{t_w}$  is the risk set at  $t_w$ : the set of squares that are at peace immediately before  $t_w$ . The parameters can be interpreted in terms of a relative probability of war.

To perform an analysis with this model, we need a data file constructed in the following way: For each  $t_w$  – i.e. each day a square war breaks out somewhere – we take a ‘snapshot’ of the system; we note for all squares the values of the explanatory variables at that particular day. As is seen from expression (2), the square that did have an event at  $t_w$  is compared to all squares that were at risk of doing so. Thus, all information for the time between different  $t_w$ ’s is ignored. From the combined information about all outbreaks in the period under study, we can estimate the hazard function (1).

A dataset comparing all the 100,000 squares 2530 times would be forbiddingly large – approximately 250,000,000 lines. It is therefore necessary to analyze a sample of the observations. The observations of positive events contain more information than the non-event observations. We therefore sample asymmetrically including all of the transition events and 1.0% of the non-transition events.

#### Disaggregated independent variables

Local-level data on land, population, and elevation are available in the geospatial format of raster files with a resolution of 1 km. Using Geographic Information systems (GIS), attributes from raster and point data are associated with the grid square in which they lie. In this way, spatial data is georeferenced to a location that is defined by the grid cell. This process results in a data structure in which each row has within it combined information on a square defined by the grid, the national-level information in which it is

located, and the local data on physical geography and population from the raster data. These data can then be imported into statistical programs for analysis.

We aggregate all data up to a grid of  $8.6 \times 8.6 \text{ km}^2$ . Each grid square is assigned attributes of the country it is in along with information from data disaggregated to the level of the individual squares. Fig. 3 illustrates this grid as a fictive country somewhat smaller than the average size in our dataset ( $50 \times 50$  squares, or  $430 \times 430 \text{ km}$ ) with a fairly representative but stylized population distribution. The country has three major cities, one of which is the capital, and two smaller ones. A rebel group has its headquarters at the eastern border.

The ACLED data for the Central African conflicts were aggregated up to the  $8.6 \times 8.6 \text{ km}^2$  and merged with information on other explanatory variables aggregated to the same level. Squares in which conflict events (not shown in Fig. 3) are located will be coded as conflict squares. The 14 countries in the dataset cover 8 million square kilometers or just over 127,000 squares.

#### Log population in square

High-resolution population data are created by CIESIN in collaboration with United Nations Environmental Program.<sup>5</sup> The database population count assessments are from a compilation of existing data sources. The sources of error for population counts and distribution are admittedly substantial, particularly in developing areas. In this version, population figures are transformed into a distributive surface to use in spatial analysis. A rasterized (gridded) format is based on interpolating ‘accessibility weights’ by administrative units and assigning population totals based on the assumption that population distribution and densities in Africa are strongly correlated with accessibility. The accessibility index is the sum of population totals of the towns in the vicinity weighted by distance. Each raster unit is assigned a population based on interpolating population densities from the accessibility weight and a distance decay function for surrounding areas. Adjustments to the accessibility grid were made for inland water bodies, elevation and protected areas. Although by no means a perfect account of local population, the data remain the most comprehensive and sophisticated spatially referenced population data available. We use estimated total population figures, observed once for every decade since 1960.<sup>6</sup>



In original form, the population count has a resolution of 1 km × 1 km. For this paper, the population counts were aggregated to the 8.6 × 8.6 km grid square. For observations in the period 1.1.1960–31.12.1964 we used the population figure for 1960. For observations in the period 1.1.1965–31.12.1974 we used the population figure for 1970, etc.

Since the area of each square is identical, the variable also indicates the local population density. The variable was log-transformed in the analysis reported.<sup>7</sup> 0.5 were added to all observations to avoid non-defined transformations.

#### *Log population in neighboring squares*

To test propositions 5 and 6 concerning population concentrations in the immediate neighborhood of the squares, information is coded in three variables. The first 'ln (population in neighborhood 1st order)' is the sum of the population sizes in the eight immediately contiguous squares. The second is the 2nd order neighborhood population – the sum of the populations in the 16 squares contiguous to these again. The third is the sum of the populations in the 24 3rd order neighbors. The variables were thereafter log-transformed. These variables allow capturing whether squares located close to other populous squares have a higher risk of events than more isolated squares in terms of populations.

#### *Population growth in square*

Population growth data were derived from the 'log population in square' variable. We coded local population growth as the difference in square population from one period to the next. For observations in the period 1.1.1960–31.12.1964 we coded population growth as missing. For observations in the period 1.1.1965–31.12.1974 we used log population for 1970 minus log population for 1960, etc.

#### *Log population in country*

To allow distinguishing between the local-level and national-level mechanisms, we added information for the total population of the country, log-transformed. The data were taken from the *Times Concise Atlas of the World* (2000).

#### *Log area of country*

We also added information on the total extent of the country in log square kilometers. The data were taken from the *Times Concise Atlas of the World* (2000).

#### *Distance from capital*

To test hypotheses regarding the distribution of populations, we estimate the distance from each square to the capital of the country. The variable was coded as the distance in terms of squares and log-transformed.

#### *Distance from rebel group headquarters*

We coded the location of the headquarters of the rebel groups participating in the conflicts under study, and calculated the distance from each square to the most proximate rebel group headquarters (we do not know a priori which rebel group or government will act in a particular square). As the 'distance from capital' variable, it was coded as the distance in terms of squares and log-transformed.

#### *Border square*

A square is considered a 'border square' if it is located up to 20 km from the national borderline or a national border runs through it. Squares belonging to more than one country are not straightforward to code, so we coded national-level information for border squares according to the following rule: A border square was considered to belong to the country that was most frequent among the eight neighboring squares. In tie cases, we assigned nationality randomly between the tied countries.

#### *Interaction country-square population*

This variable was created to test the population settlement pattern hypothesis. It is an interaction between population count at a location (square) as a portion of the country's total population.

#### *Road type*

Road type is a variable by ESRI that is available in the Digital Chart of the World Data. It is a high-resolution dataset at 1:1,000,000 scale and consists of arcs which indicate road mass. Road type is defined by the following: The reference category (0) squares are those with dual lane/divided highways, other primary roads, or road connectors within urban areas (types 1 or 8 in the ESRI dataset). The second category is secondary roads (type 2), and the third combines squares with informal or tertiary roads (tracks, trails or footpaths) or no road registered at all (types 3 and 0, respectively, in the ESRI dataset). Fig. 4 overlays the types of roads in the original dataset before our recategorization. The shaded area represents the portion of Africa for which we code information for the explanatory variables.<sup>8</sup>

#### *Variables modeling temporal and spatial dependence*

We coded three variables to account for temporal and spatial dependence:

##### *Temporal proximity of event in square*

A fundamental issue is the dependence between events and previous events in the same location. We calculated the number of days since a conflict event happened in the same location, analogous to the 'peace years' variable in country-year models (e.g. Beck, Katz, & Tucker, 1998). As in Raknerud and Hegre (1997), we assume that the effect of the previous event decreases at a constant rate, and compute a decay function with a half-life of  $\alpha$ :  $pes = 2^{(-days/\alpha)}$ . The variable is called 'Proximity of event in square'.

We estimated a set of models with different values for  $\alpha$ , corresponding half-lives of ¼, ½, 1, 2, and 4 years.  $\alpha = 1430$  days or 4 years yielded the highest log likelihood, so we estimated all the models reported below with that as the half-life parameter. We expect a positive parameter estimate, as events are more likely in locations where conflict has already started than in other locations.

##### *Proximity of event in neighboring square (1st and 2nd order)*

Events in a location are dependent not only on previous events in the same locations, but also on previous events in other locations. Events in the most proximate locations are presumably the most important. We calculated the number of days since a conflict event happened in 1) first-order neighborhoods – locations immediately adjacent to the unit of observation, and 2) the number of days since conflict in second-order neighborhoods – the squares adjacent to these again. We calculated the decay function with 1430 days as the half-life



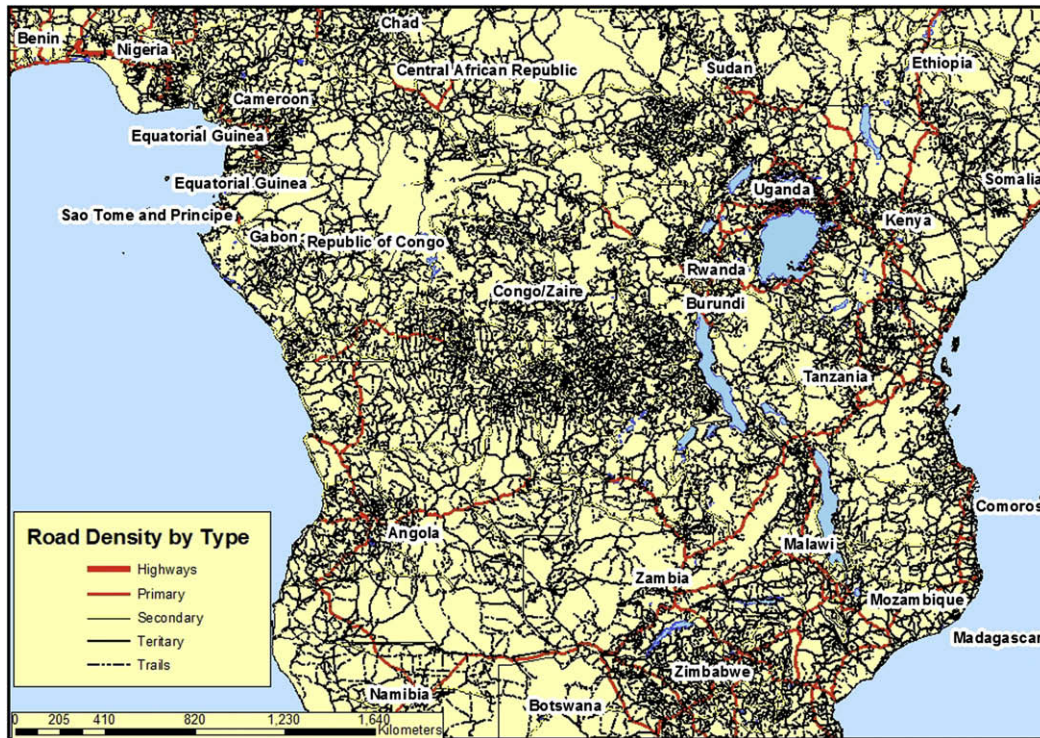


Fig. 4. Road density in Central Africa, 1994.

parameter for both. We refer to these variables as 'Proximity of event in neighboring square (1st)' and 'Proximity of event in neighboring square (2nd)', respectively.

We expect positive parameter estimates for the 'Proximity of event in neighboring square (1st, 2nd)' variables, as events are more likely in locations close to where conflict has started than in other locations.

#### *Distance to most recent previous event*

The 'Proximity of event in neighboring square' variable cannot capture the extent to which events are dependent on geographically more distant events – i.e. events outside the immediate neighborhood. To capture these relationships we calculate the distance (in square units, e.g. 8.6 km) from the unit of observation to the most recent event in the dataset. We log-transform the variable, and expect a negative relationship between the variable and the risk of observing events: Events are most likely to be followed by events in proximate squares, so the risk decreases with distance from the most recent event.

#### *Dynamics*

With simple extensions, our calendar-time Cox regression model allows a detailed study of the dynamics of conflict. Since information about preceding events is included in the set of explanatory variables, our results are interpreted as the relative probability of event in a location given the history of the location. If we obtain positive estimates for population size at the location and for the proximity of event in square, we must conclude that large populations both increase the risk of the initial event and indirectly increase the risk of subsequent events. With simple extensions, our model has a potential for dynamic modeling that we do not exploit for space reasons. For instance, the dynamic development at an

aggregate level implied by the estimates could be explored by means of simulations.

## **Results**

### *Testing the hypotheses*

In Table 1, we present results for three models omitting information on the locations of rebel group headquarters. The three models vary in terms of how we model population concentrations. In Table 2 we include the headquarters variable for the same three models.<sup>9</sup> We will refer to the models in Table 1 as Models A1, A2, and A3, and those in Table 2 as B1, B2, and B3.

In Models A1 and B1, the estimates for the 'population in square' variable are positive and significant. Conflict events tend to occur more often in squares that are relatively populous, as posited in Hypothesis 1. The estimate implies that increasing populations in squares by a factor of 2.7 increases the risk of conflict events by about 13%.<sup>10</sup>

Hypothesis 2 is also clearly supported in Models A1 and B1. Increasing distance from the capital by a factor of 2.7 increases the risk of conflict events by 8–10%. Hypothesis 3 receives strong support from the analyses as well. Controlling for other variables, including the distance from capital, border squares are more than three times as likely to have conflict events as other squares.<sup>11</sup>

Models A2 and B2 allow for investigating the effect of population concentrations. The three 'population in neighborhood' variables are significant both individually and jointly in both models. Hypothesis 5 is clearly supported.

In Table 2, the 'distance from rebel group headquarters' variable is also included. In Model B1, the estimate for this variable is  $-0.107$  – roughly the equivalent to the estimate for the 'distance to capital' variable. Conflict events are most frequent close to rebel group headquarters, as stated in Hypothesis 4.

**Table 1**

Cox regression estimation results. Risk of conflict event, population size and concentration models, 1960/2004.

Variable	Model A1	Model A2	Model A3
	$\beta$ (s.e.)	$\beta$ (s.e.)	$\beta$ (s.e.)
Log (population in sq)	0.132*** (0.014)	–0.157*** (0.038)	–0.215*** (0.082)
Log (population in neighborhood)		0.158** (0.081)	0.784*** (0.19)
Log (population in neighborhood 2nd order)		0.333*** (0.098)	–1.58*** (0.26)
Log (population in neighborhood 3rd order)		–0.123*** (0.064)	1.87*** (0.17)
Log (population in sq $\times$ log (distance to capital))			0.0257 (0.026)
Log (population in neighborhood $\times$ log (distance to capital))			–0.197*** (0.056)
Log (population in neigh. 2nd order $\times$ log (distance to capital))			0.572*** (0.077)
Log (population in neigh. 3rd order $\times$ log (distance to capital))			–0.570*** (0.048)
Log (population in country)	0.0493 (0.0343)	0.0492 (0.034)	0.123 (0.039)
Log (distance to capital)	0.0834*** (0.024)	0.143*** (0.024)	2.18*** (0.18)
Border square	1.13*** (0.072)	1.10*** (0.074)	1.14*** (0.078)
Road type 0: primary roads	Ref	Ref	Ref
Road type 1: secondary roads	category –0.628*** (0.062)	category –0.750*** (0.063)	category –0.708*** (0.067)
Road type 2: tertiary or no roads	–1.39*** (0.081)	–1.64*** (0.084)	–1.74*** (0.086)
Proximity of previous event in square	6.11*** (0.077)	6.18*** (0.076)	6.11*** (0.078)
Proximity of previous in neighborhood (1st order)	–0.904*** (0.071)	–0.814*** (0.073)	–0.619*** (0.077)
Proximity of previous in neighborhood (2nd order)	–0.0097 (0.069)	–0.0054 (0.071)	–0.176 (0.075)
Log distance to previous event	–0.494*** (0.0343)	–0.494*** (–0.019)	–0.484*** (–0.019)
No. of failures	2291	2291	2291
II Null Model	–30,571.32	–30,571.32	–30,571.32
II Full Model	–22,188.49	–22,127.27	–22,019.07

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Many of the variables in the model are closely related, and the individual coefficients cannot be interpreted on their own.<sup>12</sup> Predictions for the fictive country represented in Fig. 3 help interpret these estimates. Fig. 5 shows how distance alone affects the estimated risk for each location in this map, based on the estimates in Model B1 (this figure disregards variation in population densities).<sup>13</sup> Here, only the distances from the capital and the rebel group headquarters make a difference. The figure shows clearly how the risk increases as the distance from the capital increases and the distance to the rebel group headquarters increases. Both Hypothesis 2 and 4 are supported by these estimates. Note, however, that the magnitude of the effect is relatively small. The largest difference in relative risk is less than 3 – the per-capita risk of conflict events is two to three times more frequent in the area with a dark shade close to the rebel base than in the capital.

Fig. 6 shows the predictions from the same model, but taking border squares and population into account. The effect of the population-mass variable is reflected as peaks at the population concentrations. Note how the risk of events is much higher along the borders, and that the risk of events are highest where 'ln (population in square)' is high. Although the model clearly shows that the risk of events increase with the distance from the capital, the population variable is substantially more important.

**Table 2**

Cox regression estimation results. Risk of conflict event, population size and concentration models, 1960–2004. Including distance to rebel group headquarters.

Variable	Model B1	Model B2	Model B3
	$\beta$ (s.e.)	$\beta$ (s.e.)	$\beta$ (s.e.)
Log (population in sq)	0.122*** (0.015)	–0.179*** (0.038)	–0.280*** (0.083)
Log (population in neighborhood)		0.192** (0.082)	0.945*** (0.19)
Log (population in neighborhood 2nd order)		0.326*** (0.099)	–1.80*** (0.27)
Log (population in neighborhood 3rd order)		–0.141** (0.065)	2.00*** (0.18)
Log (population in sq $\times$ log (distance to capital))			0.0378 (0.026)
Log (population in neighborhood $\times$ log (distance to capital))			–0.236*** (0.057)
Log (population in neighborhood. 2nd order $\times$ log (distance to capital))			0.635*** (0.079)
Log (population in neigh. 3rd order $\times$ log (distance to capital))			–0.614*** (0.0494)
Log (population in country)	0.0736** (0.0353)	0.0789** (0.036)	0.177 (0.041)
Log (distance to capital)	0.0952*** (0.024)	0.156*** (0.024)	2.29*** (0.178)
Log (distance to rebel group headquarters)	–0.107*** (0.034)	–0.115*** (0.034)	–0.173*** (0.035)
Border square	1.15*** (0.072)	1.11*** (0.034)	1.14*** (0.079)
Road type 0: primary roads	Ref	Ref	Ref
Road type 1: secondary roads	category –0.650*** (0.062)	category –0.770*** (0.063)	category –0.737*** (0.067)
Road type 2: tertiary or no roads	–1.41*** (0.081)	–1.65*** (0.084)	–1.76*** (0.067)
Proximity of previous event in square	6.05*** (0.077)	6.12*** (0.078)	6.02*** (0.079)
Proximity of previous in neighborhood (1st order)	–0.973*** (0.073)	–0.898*** (0.077)	–0.738*** (0.081)
Proximity of previous in neighborhood (2nd order)	–0.053 (0.071)	–0.0953 (0.072)	–0.243 (0.076)
Log distance to previous event	–0.482*** (–0.019)	–0.484*** (–0.019)	–0.468*** (–0.020)
No. of failures	2291	2216	2216
II Null Model	–30,571.32	–30,571.32	–30,371.32
II Full Model	–22,183.54	–22,121.63	–22,006.83

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

In Model B2, we test Hypothesis 5, adding the three 'ln (population in neighborhood)' variables. The first- and second-order variables are positive and significant, whereas the third-order variable is negative and significant. With these variables in the model, the estimate for the population in square variable is negative. This result, again, may be due to collinearity: Since populations are clustered, populous squares often have populous neighborhoods, and the estimates should be interpreted jointly. Together, they provide ample support for the population concentration hypothesis: Conflict events happen disproportionately in squares close to population centers. The negative estimate may be interpreted to mean the events occur in the outskirts of such population centers. Note the increase in log likelihood relative to Model B1, from –22,183.54 to 22,121.63 – including information about the neighborhood population clearly increases the model's fit to the data.

Fig. 7 presents the estimates in Model B2. Again, the figure plots the estimated log relative risk for each cell in the fictive country based on Model B2. The inclusion of the population size of the cells in the immediate neighborhood further increases the importance of population relative to distance. The squares close to the major cities are 3–5 times more likely to experience conflict events than other squares. Conflict events clearly tend to occur close to population concentrations, not in the hinterlands. In particular, the city

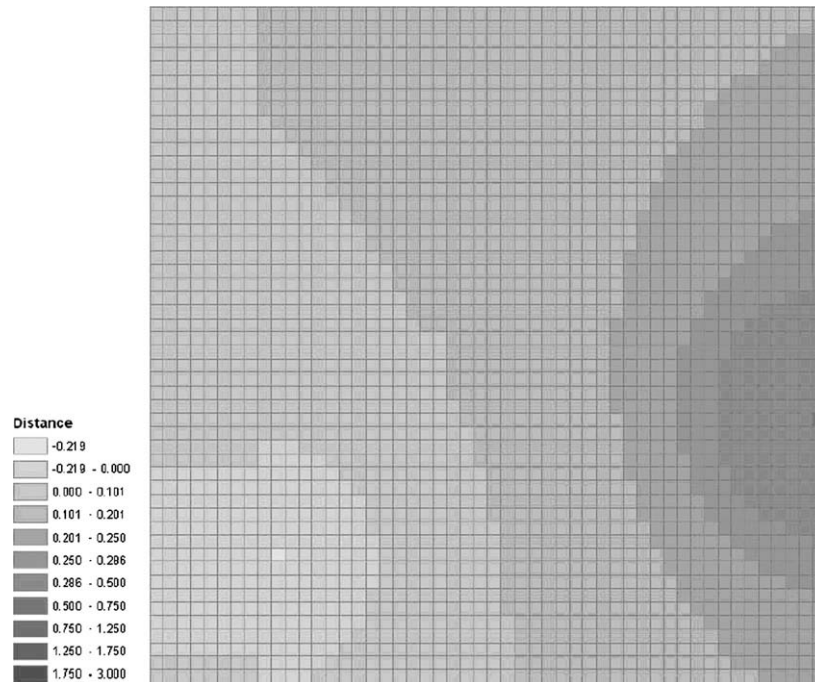


Fig. 5. Estimated per-capita relative risk of conflict events for locations in Fig. 3, distance variables only.

in the North East region is predicted to be particularly conflict-prone, since it combines a population concentration and a large distance from the capital. This city is the most conflict event-prone location on the map, rivaling even the border squares far from the capital. This testifies to the importance of including information on the population in the neighborhood in the model, not just the population at the location. These results provide strong support for Hypothesis 6, and reflect the conflict patterns in Herbst's 'difficult' political geographies.

In Model B3, we test Hypothesis 8 by adding the interaction terms between the four local population variables and the distance from the capital to the model. The interaction terms are significant, and the log likelihood of the model is further reduced to  $-22,006.83$ .

The predictions from Model B3 are plotted in Fig. 8. The population-distance interaction model further strengthens the importance of population concentrations relative to distance.<sup>14</sup>

The estimate for  $\ln(\text{population in country})$  is not statistically significant in most of the models in Tables 1 and 2, controlling for the other variables in the model. There is not much support for Hypothesis 7: The risk of conflict events is clearly larger in populous squares, but is independent of the size of the population in the country. We take this as evidence that we have succeeded in modeling the most important explanations of why large countries have more armed conflicts.

#### Results for control variables

The three road type variables in Table 1 seek to refine the measure of the government's ability to control the territory at the location. The 'primary roads' category was set as the reference category. Conflict events are 47% less likely to happen in squares with secondary roads than in the reference category squares, and 75% less likely to happen in squares with no roads or only informal roads than in squares with primary roads.<sup>15</sup>

The results run counter to our initial expectations – conflicts are assumed to occur in faraway and inaccessible regions. However, the

finding may not be so counter-intuitive after all. First, battle events occur where rebel group and army units encounter each other. Such meeting places are normally reached by road. Second, rebel groups tend to target high-value places (villages, military installations, pipelines, mines, etc.), and roads also often connect such places. Third, there is also a reporting bias at play here – media report incidences primarily in accessible areas.

The variables designed to capture spatial and temporal dependence largely obtain the expected estimates. The estimate for the 'proximity of previous event in square' variable shows that a square that has experienced conflicts one year ago has a risk of another event 167 times higher than squares with no conflict history. The ' $\ln(\text{distance to previous event})$ ' variable shows that squares that are close to the location of the most recent event in the dataset are much more likely to see events than more distant squares – increasing the distance from an event by one percent decreases the risk of the next event occurring by 1.18 percent.

The estimate for 'proximity of previous event in neighborhood, 1st order' is anomalous, on the other hand. The estimate is negative, indicating that the risk of conflict events decrease for a time in the immediate neighborhood of an event. The reason why the estimate turns out this way is not clear, but is most likely due to collinearity.

Both population variables are also associated with the road network variable, since population totals figures were interpolated assuming that population distribution and densities are correlated with accessibility. This introduces another potential for collinearity. The estimate for the road variable pulls in the same direction, however: Civil war events are most frequent in the most accessible squares containing primary roads and high population concentrations.

The results for other variables in the model are mainly consistent with our expectations and previous studies such as Buhaug and Gates (2002) and Buhaug and Rød (2006). One exception is the weak effect of distance from the capital in our analysis relative to theirs. This may be explained by their choice to use the distance from the capital to what they define as the



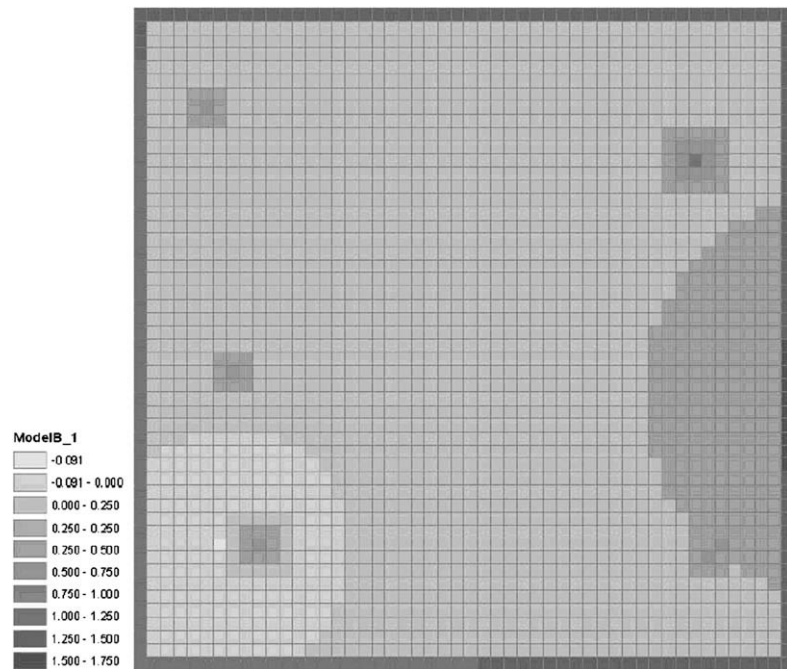


Fig. 6. Estimated relative risk of conflict events for locations in Fig. 3, based on Model B1.

geographical center of the conflict area as the distance measure. We, in contrast, use the distance from the capital to each individual conflict event. A more likely explanation for the difference in finding may be that they include data for all of Africa whereas we only analyze Central Africa. In particular, the conflicts in densely populated Rwanda and Burundi weigh more heavily in our analysis than in theirs.

#### *Implications for country-level studies*

One of the motivations for this study was to use the disaggregated analysis to discern between different explanations for why more populous countries have more frequent civil wars. To do so, we need a method to aggregate our estimates of the risk of conflict events in a grid square up to the country level. We assume

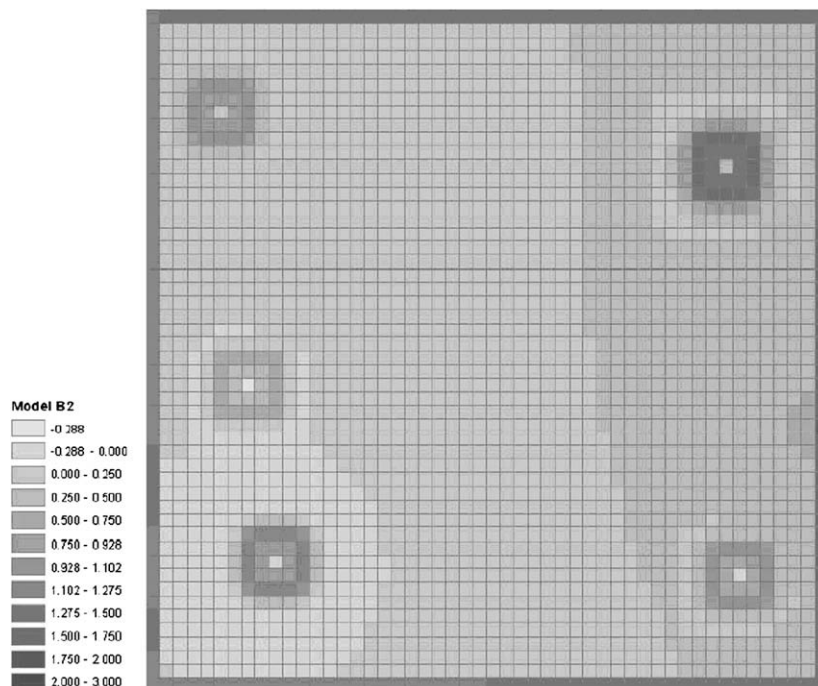


Fig. 7. Estimated relative risk of conflict events for locations in Fig. 3, based on Model B2.

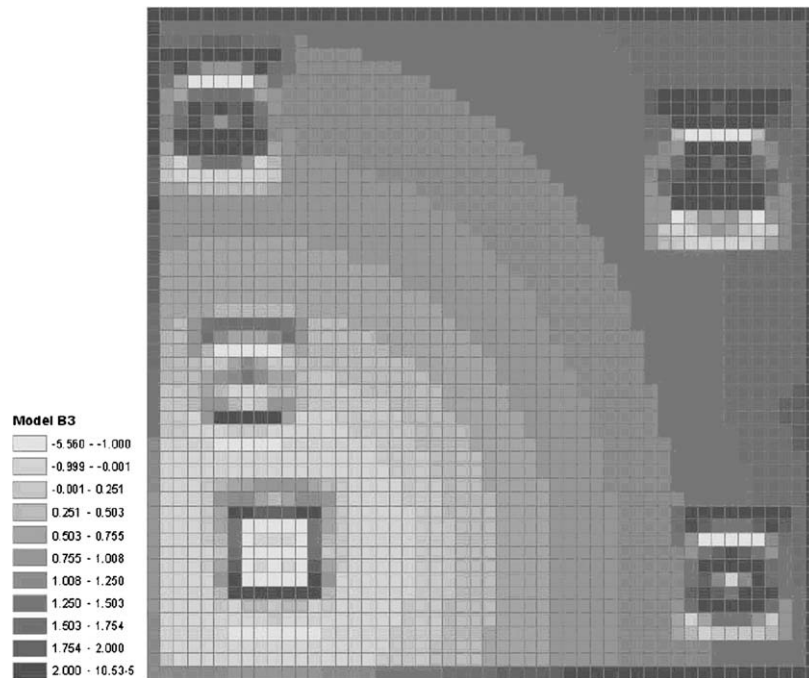


Fig. 8. Estimated relative risk of conflict events for locations in Fig. 3, based on Model B3.

that the probability  $p_c$  of observing a civil war in a country is proportional to the probability  $p_d$  of observing at least one event within the  $S$  squares constituting its territory. This probability is given by  $p_c = 1 - \prod_{d=1}^S (1 - p_d)$ .

There are four parameters that affect how the square-level probabilities in our model aggregate up to the country level. The first is the area of the country or the number of squares  $S$ . If the square-level probabilities are independent of the number of squares in the country, the country-level risk must increase in the area of the country.<sup>16</sup> Comparing two countries with similar population densities, the most extensive country has the highest country-level probability of conflict. The second results from countries with larger areas having higher average distances to the capital. We have shown that the risk of conflict events increase with distance from the capital. This result reinforces the conflict-inducing effect of area. Our third parameter runs counter to this country-level effect: We demonstrated that border squares are more prone to conflict events. Geographically small countries have a larger proportion of border squares. This counteracts the risk-increasing effect of size.

Table 3 shows the predicted risk of conflict aggregated to the country level for a set of nine stylized countries. The predictions are

**Table 3**

Estimated country-level risk of conflict as function of population density and country size.

Grid square density of population				
Number of grid squares	Square kilometers (appr.)	200	2000	20,000
18 × 18	24,000	1.0 (65,000)	1.2 (650,000)	1.8 (6,500,000)
57 × 57	240,000	8.5 (650,000)	10.4 (6,500,000)	15 (65,000,000)
180 × 180	2,400,000	71 (6,500,000)	85 (65,000,000)	120 (650,000,000)

Figures in cells are estimated risks relative to the baseline (the upper left case; a country of size 18 × 18 squares and population density of 200) – 10.4 in the middle cell reflects that the estimated risk of conflict events is 10.4 times higher than the baseline. Figures in parentheses are approximate total populations.

based on the estimates in Model A1. The rows of the table identify three different sizes in terms of number of grid cells or square kilometers. The smallest example country is 18 by 18 squares, or 24,000 square km, about the size of Rwanda and Burundi. The largest is 180 by 180 squares, or 2,400,000 km<sup>2</sup>, the size of D.R. Congo. The columns represent three different population densities – 200, 2000, and 20,000 persons per grid square, respectively. We aggregate the estimated square-level probabilities up to these countries, and report these risks relative to the risk of the smallest example country with the lowest population density. If we keep the area constant but increase population density from 200 to 2000 per grid square, the country-level risk increases by about 20%. If we increase it further to 20,000 per square, the risk increases by further 50%. If we, on the other hand, increase the area of the country keeping population density constant, the estimated risk increases by a factor of 8.5.

In country-level studies, the correlation between area and population size is too strong to allow distinguishing between them. Our disaggregated analysis, however, indicates that the positive effect of long distances on the risk of conflict events contributes much more than population per se to the finding that large, populous countries have more conflicts at the country level.

Combining these two effects, we find an elasticity of conflict risk with respect to population that is lower than one, just as country-level studies do.

## Conclusion

This paper represents an initial analysis of the ACLED data (Raleigh et al., in press), which codes the dates and exact locations of individual events within a set of internal armed conflicts. We have specified a statistical model that handles both spatial and temporal dependence in a setup of very fine-grained data and that allows analyzing the dynamics internal to civil wars. Since we analyze both initial events in a given conflict and the diffusion in

time and space given by the subsequent events, our analysis bridges the gap between studies of the onset and the duration of civil war.

The analysis has illuminated some aspects of the relationship between country size and the risk of internal conflict that cannot be analyzed in national-level studies: Conflict events tend to have frequencies in proportion to the size of the population in a given location, as indicated by a 'per-capita propensity' hypothesis. We also find that conflicts happen predominantly where populations cluster locally. In the sample of Central African conflicts we have studied, conflict events happen more frequently in locations far from the capital of a country and close to the border.

Moreover, we find the importance of distance to be considerably less than that of population mass, except when combined with population concentration. Conflict events do happen in peripheral regions, but the analysis indicates that the picture of African internal conflicts as primarily rural events is inexact. The risk of conflict in a location depends on the value of the location, here proxied location by size of the population that resides within it.

Although the effect of distance is moderate relative to population concentration, the results still indicate that countries with populations that are largely concentrated around the capital have fewer internal conflict events than countries with populations that are spread out, or, even more strongly, are also concentrated in locations far from the capital.

Our results might contribute to the discussion on the usefulness of partition to solve internal armed conflicts (see Kaufmann, 1996 and Sambanis, 2000 for discussions). The relatively slight importance of distance from the capital relative to population mass indicates that the total risk of conflict events in the two separated countries will only be marginally smaller in the original one. However, territories with large and clearly separated population concentrations might conceivably have perceptibly lower risks of conflict events as two or more countries surrounding each concentration than as one 'difficult' political geography (our analysis does not say anything about the risk of interstate war between such entities, however).

The analysis presented here points to the immense potential inherent in disaggregated analysis. The model presented here is well suited to be extended to test hypotheses regarding the availability of 'lootable resources', patterns in the distribution of ethnic groups and/or the geographical distribution of income, or the impact of the location of refugee camps or of loutable resources. With continued data collection on development, political marginalization and local ethnic relationships, our understanding of civil war causes and dynamics can be further refined.

## Acknowledgments

Thanks to Joakim Karlsen for writing a program to create the dataset used in the analysis, to Siri Aas Rustad for research assistance, and to Kristian Gleditsch, Anke Hoeffler, Pat Regan, Mike Ward, Nils Weidmann, Jen Ziemke, and John O'Loughlin for valuable comments. The research has been funded by the Research Council of Norway, grant no. 163115/V10.

## Endnotes

<sup>1</sup> Calculated on the basis of a parameter estimate for  $\ln(\text{population})$  of 0.28, as in Hegre and Sambanis (2006). Other studies obtain larger estimates for this variable.

<sup>2</sup> By 'increased population', we are comparing the size of population both across countries and across time.

<sup>3</sup> If the rebel group is sufficiently strong, it will be able to push the frontline to the capital. In that case, we might expect to see a large number of events close to the capital. Simulations reported in Weidmann, Hegre, and Raleigh (2006) indicate, however, that in these cases the rebel group tends to be so strong relative to the government that the war ends quickly, with few observable events.

<sup>4</sup> See [acleddata.com](http://acleddata.com) for more information and to download data.

<sup>5</sup> <http://na.unep.net/globalpop/africa/>. We use the fourth version of population data.

<sup>6</sup> The UNDP/CIESIN data are available by continent at <http://grid2.cr.usgs.gov/globalpop/Africa/part2.html>.

<sup>7</sup> The population variable is strongly skewed to the right. Log-transformation of such variables reduces the influence of individual extreme observations. Moreover, there are good reasons to expect the marginal effect of population is decreasing with size.

<sup>8</sup> Further information on the road measure can be obtained at <http://atlas.geo.cornell.edu>.

<sup>9</sup> Note that the estimates in Tables 1 and 2 are very similar. The inclusion of the 'distance to rebel group headquarters' does not affect the results much apart from reducing the log likelihood with more than 5 points.

<sup>10</sup> The population and distance variables are measured as natural logarithms of the original variables. A one-unit increase in the variables as they enter the estimation, then, corresponds to multiplying the original variable by  $e = 2.718$ . We therefore interpret the estimates by calculating the effect of increasing the original variable by a factor of 2.7. The estimates may also be interpreted as an elasticity: Model A1 implies that increasing distance from the capital by one percent, for instance, increases risk of conflict events by 0.0834 percent relative to the baseline.

<sup>11</sup> Increasing the Border square variable by one unit increases relative risk by a factor of  $\exp(1.13) = 3.1$ .

<sup>12</sup> Inspection of the variance-covariance matrix of the estimates reveals only low correlations between estimates of the simplest models. In models A3 and B3, however, some estimates for the neighborhood variables and the interaction terms have correlations larger than 0.5. Given this collinearity, individual estimates will be unstable. The effects visualized in Figs. 5–8 take all the estimates jointly into account, however, and the conclusions drawn from these are much less sensitive to collinearity problems.

<sup>13</sup> The shades in Figs. 5–8 reflect the relative risk of conflict event, calculated as the exponential of the linear expression  $\beta X$ . The different values for the  $X$  variables are given by each location's distance from the borders, the rebel group HQ, and the capital of the 'country', the populations of each location as specified in Fig. 3, and the populations of the neighboring locations.

<sup>14</sup> Models A3 and B3 suffer somewhat from collinearity problems, reflected as pairs of large positive and large negative estimates for variables that are highly correlated.

<sup>15</sup>  $\exp(-0.628) = 0.53$  and  $\exp(-1.39) = 0.25$ .

<sup>16</sup> In an extension of Model A1 (not reported here) we tested whether the log of the number of squares in the country improve the goodness-of-fit of the model. It does not.

## References

- Alesina, A., & Spolaore, E. (2003). *The size of nations*. Cambridge, MA: MIT Press. pp. 40–45.
- Beck, N., Katz, J., & Tucker, R. (1998). Taking time seriously in binary time-series cross-section analysis. *American Journal of Political Science*, 42, 1260–1288.
- Buhaug, H., & Gates, S. (2002). The geography of civil war. *Journal of Peace Research*, 39, 417–433.
- Buhaug, H., & Rød, J. K. (2006). Local determinants of African civil wars, 1970–2001. *Political Geography*, 25, 315–335.
- Clapham, C. (1986). *Third world politics*. London: Croom Helm.
- Collier, P. (2000). Doing well out of war: an economic perspective. In M. Berdal, & D. M. Malone (Eds.), *Greed & grievance: Economic agendas in civil wars* (pp. 91–111). Boulder, CO: Lynne Rienner.
- Collier, P., & Hoeffler, A. (2002). Greed and grievance in civil war. Centre for Economic Policy Research Working Paper, London.
- Collier, P., & Hoeffler, A. (2004). Greed and grievance in civil war. *Oxford Economic Papers*, 56, 563–595.
- Fearon, J., & Laitin, D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, 97, 75–90.
- Gleditsch, N. P., Mikael E., Margareta S., & Håvard S. (2002). Armed Conflict 1946–2001: A New Dataset. *Journal of Peace Research*, 39, 615–637.
- Gurr, T. R. (1970). *Why men rebel*. Princeton, NJ: Princeton University Press.
- Hegre, H., & Sambanis, N. (2006). Sensitivity analysis of the empirical literature on civil war onset. *Journal of Conflict Resolution*, 50, 508–535.
- Herbst, J. (2000). *States and power in Africa: Comparative lessons in authority and control*. Princeton, NJ: Princeton University Press. p. 146, 152–154.
- Kalyvas, S. (2006). *The Logic of Violence in Civil War*. Cambridge: Cambridge University Press.
- Kaufmann, C. (1996). Possible and impossible solutions to ethnic civil wars. *International Security*, 20, 136–175.
- Lichbach, M. (1995). *The rebel's dilemma*. Ann Arbor, MI: University of Michigan Press. p. 126, 156, 158–159, 160, 162.
- McColl, R. (1967). A Political Geography of Revolution: China, Vietnam and Thailand. *The Journal of Conflict Resolution*, 11, 153–167.
- McColl, R. (1969). The insurgent state: territorial bases of revolution. *Annals of the Association of American Geographers*, 59, 613–631.



- Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. Introducing ACLED: an armed conflict location and event dataset. *Journal of Peace Research*, in press. (Available from: [www.acleddata.com](http://www.acleddata.com)).
- Raknerud, A., & Hegre, H. (1997). The hazard of war: reassessing the evidence for the democratic peace. *Journal of Peace Research*, 34, 385–404.
- Sambanis, N. (2000). Partition as a solution to ethnic war: an empirical critique of the theoretical literature. *World Politics*, 52, 437–483.
- Scarritt, J., & McMillan, S. (1995). Protest and rebellion in Africa: explaining conflicts between minorities and the state in the 1980s. *Comparative Political Studies*, 28, 323–349.
- Times Concise Atlas of the World*. (2000) (8th ed.). London: Times Books.
- Toft, M. (2003). *The geography of ethnic violence*. Princeton, NJ: Princeton University Press.
- Weidmann, N., Hegre, H., & Raleigh, C. (2006). *Modeling spatial and temporal patterns of civil war*. Typescript. PRIO.