Conflict Management and Peace Science
http://cmp.sagepub.com/

# Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset

Julian Wucherpfennig, Nils B. Weidmann, Luc Girardin, Lars-Erik Cederman and Andreas Wimmer

*Conflict Management and Peace Science* published online 25 February 2011
DOI: 10.1177/0738894210393217

The online version of this article can be found at:
http://cmp.sagepub.com/content/early/2011/02/21/0738894210393217

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Peace Science Society (International)

**Additional services and information for *Conflict Management and Peace Science* can be found at:**

**Email Alerts:** http://cmp.sagepub.com/cgi/alerts

**Subscriptions:** http://cmp.sagepub.com/subscriptions

**Reprints:** http://www.sagepub.com/journalsReprints.nav

**Permissions:** http://www.sagepub.com/journalsPermissions.nav

Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

Conflict Management and Peace Science
© The Author(s). 2011. Reprints and permissions:
http://www.sagepub.co.uk/journalsPermissions.nav
[DOI:10.1177/0738894210393217]
Vol XX(X): 1–15

cmps

# Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset[1]

JULIAN WUCHERPFENNIG
*Center for Comparative and International Studies, ETH Zürich, Switzerland*

NILS B. WEIDMANN
*Jackson Institute for Global Affairs, Yale University, USA & Center for the Study of Civil War, International Peace Research Institute, Oslo, Norway*

LUC GIRARDIN
*Center for Comparative and International Studies, ETH Zürich, Switzerland*

LARS-ERIK CEDERMAN
*Center for Comparative and International Studies, ETH Zürich, Switzerland*

ANDREAS WIMMER
*Department of Sociology, UCLA, USA*

This article introduces GeoEPR, a geocoded version of the Ethnic Power Relations (EPR) dataset that charts politically relevant ethnic groups across space and time. We describe the dataset in detail, discuss its advantages and limitations, and use it in a replication of Cederman, Wimmer and Min's (2010) study on the causes of ethno-nationalist conflict. We show that territorial conflicts are more likely to involve groups that settle far away from the capital city and close to the border, while these spatial variables have no effect for governmental conflicts.

KEYWORDS: ethnic conflict; ethnic groups; geographic information systems (GIS); group settlement patterns

---

[1] The dataset is available at http://dvn.iq.harvard.edu/dvn/dv/epr. We would like to thank the editor of *CMPS* and three anonymous reviewers for helpful comments and suggestions. Many individuals helped assemble the dataset. We cannot list all country and GIS experts who generously shared their knowledge, but we should like to mention Nils-Christian Bormann, Gustav Brown, Christa Deiwiks, Philippe Duhart, James Flora, Thomas Koblet, Lutz Krebs, Joanne Richards, James R. Scarritt, Veronika Schurmann, Cristina Viehmann, Manuel Vogt and Judith Vorrath. This research was supported by the UCLA International Institute and the European Science Foundation through a grant by the Swiss National Science Foundation (#105511-116795).



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

## 1. Introduction

The recent literature on civil wars has seen a surge of theories, applications and data collection efforts pertaining to the role of geography. While initially focusing on country level characteristics (Fearon and Laitin, 2003), the use of geo-referenced data (GIS) has allowed researchers to test a host of new hypotheses on sub-national conflict processes (for a recent review, see Cederman and Gleditsch, 2009). Conflict location and scope (Buhaug and Gates, 2002), access to lootable resources, such as oil or gemstones (Lujala et al., 2007; Gilmore et al., 2005), population concentration (Buhaug and Rød, 2006; Hegre and Raleigh, 2008), or individual conflict events (Raleigh et al., 2010) are some of the spatial data recently collected and analyzed by conflict researchers. Beyond doubt, research employing these data has significantly furthered our understanding about the causes of, and dynamics within, civil wars.

If analyzed at a disaggregated level, units of analysis for these data are typically grid-cells or spatial points (see e.g. Buhaug and Rød, 2006). From a theoretical viewpoint, however, such units of analysis raise questions about sampling and level of causation, and thereby ultimately also about agency. Put simply, grid-cells do not fight. For example, it remains unclear whether conflict zones are in themselves the object of struggle, or whether they merely represent the battlefield. Likewise, the finding that the presence of petroleum, drugs or gemstones associates with conflict cannot speak to the direction of the causal arrow: do resources merely make conflict feasible by providing the financial means, or are they an end in themselves? More generally, while analyses of spatial data have provided insights to the question of *where* conflict occurs, such analyses have tended to place less emphasis on the question of *who* fights. Seen in this light, there is another spatial dimension that has largely been overlooked. It concerns the location of actors involved in conflict, not of the conflict events themselves. In this article, we show that the spatial location of collective actors is an important element for a proper understanding of conflict dynamics.

Compared to cell-based, spatial approaches to civil wars, group level approaches have the advantage of bringing the analysis closer to the actors in civil wars. So far, most empirical advances in this direction have focused on ethnic categories and groups. To be sure, such communities do not represent political actors in themselves (cf. Brubaker, 2004), nor do entire ethnic communities go to war, but specific organizations, sometimes fighting on opposite sides of an ethnic divide (Kalyvas, 2008). Unfortunately, a global dataset comprising all politically relevant organizations remains unavailable (see the dataset of *rebel* organizations by Cunningham et al., 2009). Data based on ethnic groups as units of analysis thus represent the closest approximation to collective agency in large-N studies of conflict onset.

Existing research has shown that the location and degree of geographic dispersion matters for ethnic conflict (Toft, 2003). While the widely used Minorities at Risk (MAR) dataset (Gurr, 1993) includes estimations of the degree of territorial concentration for each group, only one dataset has so far offered fully geo-coded information to test these argument systematically and on a global scale. Weidmann



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

et al.'s (2010) *Geographic Representation of Ethnic Groups* dataset (GREG) digitally represents settlement patterns of ethnic groups worldwide. In a nutshell, GREG is the GIS version of the *Atlas Narodov Mira* (ANM; Bruk and Apenchenko, 1964), a series of maps collected by Soviet ethnographers charting ethnic groups across space.

ANM (and therefore also GREG) has its limitations, however. Above all, GREG focuses exclusively on the list of ethnic groups given by the ANM authors. Recently, however, it has been recognized that the linguistic differences on which the ANM focuses exclusively do not necessarily correspond to those ethnic cleavages that are politically relevant (Posner, 2004; Cederman and Girardin, 2007; Chandra and Wilkinson, 2008). Moreover, being a one-time snapshot from 1964, the ANM is outdated for many parts of the world because it ignores changes in settlement patterns, for example through migration.

To overcome these limitations, we build on the Ethnic Power Relations dataset (EPR), which identifies all politically relevant ethnic groups and their access to central state power. EPR is a dynamic dataset that codes when ethnic power configurations change over time. Recent studies based on these data have found a strong association between the political exclusion of ethnic groups and their propensity for conflict (Wimmer et al., 2009; Cederman et al., 2010).

So far, the EPR dataset has ignored spatial aspects altogether, although we know from past research that they affect the ethno-political conflict dynamics (Toft, 2003; Weidmann, 2009). Filling this important gap, we introduce GeoEPR, a new dataset that directly builds on the EPR and GREG datasets. GeoEPR combines the strength of geo-referenced data with that of a dynamic group list directly related to ethno-political power structures. Specifically, taking EPR as our basis, we provide ethnic group polygons in GIS shapefile format. Thus, thanks to its immediate linkage to the EPR dataset, GeoEPR includes information on the political status of ethnic groups across the globe.

The article proceeds as follows: we begin with a discussion of the GREG dataset. We then briefly introduce EPR before describing GeoEPR in detail, including a number of examples. Using GeoEPR data, we show its potential in advancing spatially explicit civil-war research, before concluding with a discussion about limitations and avenues for further research.

## 2. State of the Art: GREG

As stated above, GREG is the GIS version of the *Atlas Narodov Mira*, a series of maps collected by Soviet ethnographers charting ethnic groups across space (Weidmann et al., 2010). The GREG project is a natural point of departure when creating a dataset on ethnic group settlements because of its global, comprehensive and detailed coverage (see Weidmann et al., 2010 for a comprehensive review). Moreover, the ANM also provides estimates of demographic group sizes, making it suitable for many types of analyses. It is the standard information source underlying the widely used ethno-linguistic fractionalization index (ELF) (Taylor and Hudson, 1972) and can also be used to study explicitly geographic aspects of conflict processes



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

(Buhaug et al., 2008; Cederman et al., 2009). However, as noted by Weidmann et al. (2010), the ANM (and therefore by implication also GREG) suffers from a series of limitations.

First, given that the ANM is a one-time snapshot, GREG is *time invariant*. As a standalone dataset, GREG therefore does not capture changes in the political and socio-demographic landscape. As states emerge and dissolve, political borders change, oftentimes cutting through ethnic group regions. Moreover, the salience of ethnic identities may also change over time (Laitin and Posner, 2001).

Second, due to its publication date in 1964, the settlement patterns displayed in the ANM are *outdated* for many parts of the world. Indeed, both large-scale internal and external migration, for example in Western Europe, and ethnic cleansing, such as Rwanda or former Yugoslavia, have changed the socio-geographic map in many places.

Third, having been charted by Soviet ethnographers during the Cold War, the atlas exhibits different levels of accuracy and/or resolution depending on the world region in question. Not surprisingly, the resolution displayed in the ANM varies across regions; it is much more detailed across the Eastern Europe and the member states of the Soviet Union than it is, for example, in Africa.

Finally, and perhaps most importantly, the underlying definition of ethnicity is problematic in its own right. Specifically, the ANM uses a linguistic definition of ethnic group, the main criterion for distinguishing between two groups being historic origin of language (Fearon, 2003). This has problematic implications in many countries. For example, Hutus and Tutsis in Rwanda appear as one and the same group, owing to a lack of linguistic differentiation; the same holds true for Sunni and Shi'a Arabs in Iraq.

In sum, while GREG makes a pioneering contribution to spatial analysis of ethnic groups, there is plenty of room for improvement. Our aim is therefore to address these problems by relying on a newer and more dynamic data source to identify settlement patterns of politically relevant ethnic groups.

## 3. Building on a New Dataset on Ethnic Groups: EPR

The Ethnic Power Relations (EPR) dataset identifies all politically relevant ethnic groups around the world and measures how far access to state power differs among them in all years from 1946 to 2005 (Min et al., 2010).[2] Based on an online expert survey, the sample includes *all* politically relevant ethnic groups in all 156 sovereign states with a population of at least one million and a surface area of at least 50,000 square kilometers as of 2005. By offering a more complete list of 733 politically relevant ethnic groups, this dataset improves significantly on previous efforts to code ethnic groups' access to power, such as the Minorities at Risk dataset (Gurr, 1993), which restricts the sample to 282 mobilized minorities and thus largely overlooks the ethnopolitical constellation of power at the center, and Cederman and Girardin (2007), who rely on preliminary, static measures of the political status of ethnic groups and limit their sample to Eurasia and North Africa.

[2] The data can be downloaded from Dataverse, http://dvn.iq.harvard.edu/dvn/dv/epr.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

The EPR dataset defines ethnicity as any subjectively experienced sense of commonality based on the belief in common ancestry and shared culture (Weber, 1978: 385–398). An ethnic group is politically relevant if at least one significant political actor claims to represent its interest in the national political arena, or if its members are systematically discriminated against in the domain of public politics. A "significant" political actor refers to a political organization (not necessarily a party) that is active in the national political arena. Discrimination is defined as political exclusion directly targeted at an ethnic community—thus disregarding indirect discrimination based, for example, on educational disadvantage or discrimination in the labour or credit markets.

Because the politically relevant groups and access to political power may change over time, the EPR dataset provides separate codings for sub-periods from 1946 to 2005 that we refer to as EPR periods. For example, because of the many power shifts in the history of Iraq, EPR features as many as six EPR periods in this case. For each such time period, the access to power enjoyed by representatives of an ethnic group is reported. Focusing on executive power only, that is, representation in the presidency, the cabinet and senior posts in the administration, including the army, the coding rules categorize all politically relevant ethnic groups according to whether their representatives enjoyed access to the executive branch ("included") or not ("excluded"), as well as more fine-grained sub-categories. EPR also codes a different period if the list of relevant groups has changed. Furthermore, EPR allows for nested systems of ethnic categorization: an ethnic category that is relevant during one period may later on fission into several sub-categories that become politically more relevant than the superseding category. Conveniently, such nested structures are captured through a numeric identifier system that links lower-level categories to the common higher-level group. If political relevance shifts from one level of differentiation to another over time, this will again be captured by introducing different periods. Finally, EPR reports demographic population shares for each group (Min et al., 2010).

Whereas GREG includes small language groups, many have little, if any relevance in national politics. As a result, EPR's focus on politically relevant groups leads to a lower number of groups. This is depicted graphically for a snapshot of the 1989 international system in Figure 1. However, owing to its focus on linguistic differences, GREG not does not capture ethnic cleavages along other relevant dimensions, such as religion, and thus leads to a lower number of groups in some cases. This occurs mainly in North Africa and the Middle East, where Arab speakers of different Islamic denominations are not distinguished.

## 4. Putting EPR on a Spatial Footing: GeoEPR

Having outlined the details of our underlying list of ethnic groups, we now turn to the coding of GeoEPR. Most generally, data-collection involved two steps. In a first step, experts were given EPR's list of ethnic groups for their country of expertise and asked to provide for each group (i) a coding of their respective *settlement type*, (ii) a distinction of *geo-periods* (see below), and (iii) *suitable maps* displaying the regional bases and settlement patterns. In a second step, these maps and relevant instructions on their use were passed on to GIS experts who digitized the map



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

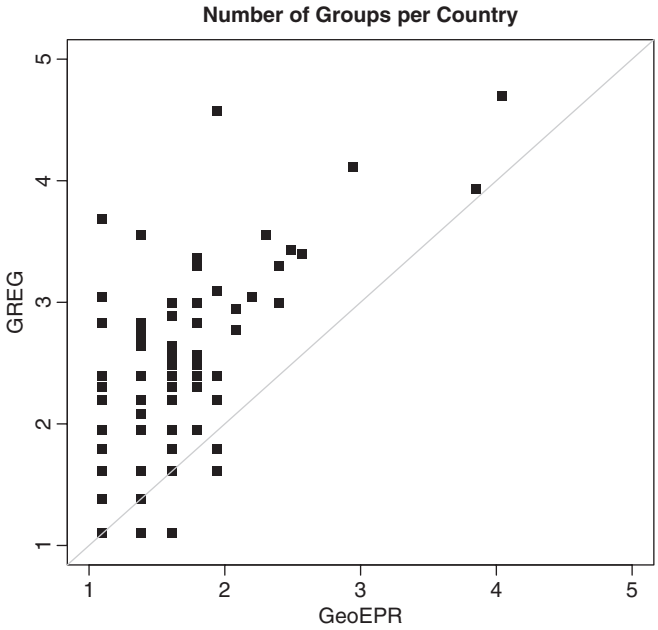


*Figure 1.* Plot of (logged) Number of EPR Groups per Country vs. the Number of GREG Groups (1989 International System)

material provided by the country expert to create a GIS dataset containing the group-referenced polygons of ethnic group settlements. Like EPR itself, GeoEPR is therefore based on an expert survey. Specifically, more than 20 country experts were recruited from both sides of the Atlantic, mainly through personal contacts within the fields of political science, sociology, and anthropology. The codings were subjected to cross-validation by other experts. We discussed each case with the experts directly, reconciled any discrepancies between country specialists, or consulted additional experts when necessary.

### *Settlement Types*

Country experts first coded whether each group in EPR's list has a regional base of settlement. We distinguish between the following (mutually exclusive) types: (1) regionally based, (2) urban, (3) regional and urban, (4) migrant, (5) dispersed or (6) aggregate. Note that we do not code "homelands" defining the territorial origin of a group (Toft, 2003), but actual spatial distribution of group members. The definitions for these patterns are as follows.

> *regionally based:* group members reside in a particular region/in particular regions that are easily distinguishable on a map. In line with MAR, we define regional base as a spatially contiguous region larger than an urban area that is part of the country, in which 25% or more of group members live. Note that overlays between different groups, i.e. co-habitation in the same region, is explicitly possible.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

*urban:* group members are located primarily and on a permanent basis in urban areas and not in a particular region/in particular regions. A group is coded as urban when 60%+ of the group is located in cities.
*regional and urban:* group members live both in a city/in cities and in a particular region/in particular regions.
*migrant:* groups members change location often, as with nomadic or Roma groups.
*dispersed:* group members do not inhabit any particular city/cities or region/regions and are not migrant.
*aggregate:* a particular group which during a geo-period (see below) is aggregated from several smaller ones (provided that their ethnic geography remains constant). In other words, thanks to the nested group structure in EPR (see above), this category significantly reduces the GIS coding efforts, since the ethnic geography of aggregate groups can be computed on the basis of their component groups.

In GeoEPR we provide information about regional bases for groups of categories (1), (3), and (6) only, that is those which are "regionally based" or "regional and urban" or "aggregate". For dispersed groups (5), the regional base corresponds to the country borders, whereas migrant groups (4) by definition do not have a permanent regional base. For simplicity, we leave the geo-coding of primarily urban groups to future research.

### *Geo-Periods*

In contrast to GREG, GeoEPR codes major changes in geographic representation over time. While we do not claim to capture small scale migration trends and other minor transformations of the ethnic map, we attempted to cover at least the most significant changes. Simultaneously to specifying a group's settlement pattern, country experts therefore determined for each group whether there were significant changes in the ethnic geography over time. Where there were important changes in settlement patterns (due to ethnic cleansing, large-scale migration or varying state borders), *geographic periods* have to be identified. We were agnostic with regard to the causes of such changes. Geo-periods thus define periods during which the group's spatial location can be treated as constant. A new geographic period must be introduced in each of the following cases: (i) the type of settlement pattern changes, (ii) group members no longer reside in the same region, (iii) modified state borders cut through an ethnic group region, or (iv) the list of relevant groups changes, especially in cases where different levels of categorical differentiations become politically relevant. Note that geo-periods in GeoEPR are distinct from political periods in EPR. While the latter can refer to changes in access to state power by politically relevant ethnic groups, the former merely assess changes in the ethno-political geography. GeoEPR periods are thus coded independently from the ethno-political dynamic captured in EPR.

Taken together, settlement type and geo-period then provide the basis for coding the regional bases of ethnic groups. Accordingly, country experts were asked to provide suitable map material that represents the spatial location of each EPR group. This also included detailed instructions for the GIS expert. Obviously, multiple maps for multiple geographic periods were required where applicable.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

*Table 1.* Number of Groups by Settlement Type

| *Regionally based* | *Urban* | *Regional and Urban* | *Migrant* | *Dispersed* | *Aggregate* |
|---|---|---|---|---|---|
| 382 | 50 | 247 | 15 | 49 | 12 |

*Table 2.* Dynamic Coding of GeoEPR

| *Coded as* | *Single Period* | *Two Periods* | *Three Periods* |
|---|---|---|---|
| Number of groups | 710 | 18 | 3 |

### *Suitable Map*

Experts were given access to a host of libraries and (online) collections of maps, and asked to provide the most adequate map material, including maps from country reports or regional studies. In a large number of cases, GREG was the appropriate source, either because EPR groups corresponded to GREG groups, or because various GREG groups could be treated as components of an overarching EPR group. In many other cases, it was necessary to consult external sources, such as Gordon (2005) and Levinson (1998).

Once expert coding was completed for a given country, the information consisting of maps and instructions was passed on to GIS experts who subsequently digitized the maps using a standard GIS format. Technically this meant tracing polygons on the map, and assigning polygons representing geographic settlements to the identity codes used for EPR groups. Using a unified online platform with a central database in the background greatly facilitated joining all individual country and group data into the final dataset.

The EPR groups identified for geo-referencing (N = 644) were coded in 755 geographic group-periods in GeoEPR. Table 1 lists the numbers of group-periods by type. Clearly, the majority of group-periods (629 out of 755) are coded as "regionally based" or "regional and urban", and thus has the settlement area represented as one or more polygons in GeoEPR, whereas the "urban" category contains 50 groups that were not geo-coded.

Table 2 shows the dynamic coding of GeoEPR, since one group can be divided into different geographic coding periods. Counting the number of periods per group, we see that the majority of groups are represented by one period, suggesting that major changes in the ethnic geography captured by GeoEPR are relatively rare.[3]

[3] Since the unit is country-group-period, the table does not capture changes in settlement area in cases where they coincide with the emergence of new countries (see the below example of Bosniaks in Yugoslavia and later Bosnia and Herzegovina), and is thus an imperfect measure for the dynamics represented in GeoEPR. For some newly emerging states, the settlement areas of groups were redrawn in addition to adapting them to the new country borders. For these cases, GeoEPR codes dynamics in the settlement pattern of a group, which is not reflected in the number of group periods. However, border changes can also lead to new geo-periods without settlement pattern changes, for example if the new border cuts through an existing group territory.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

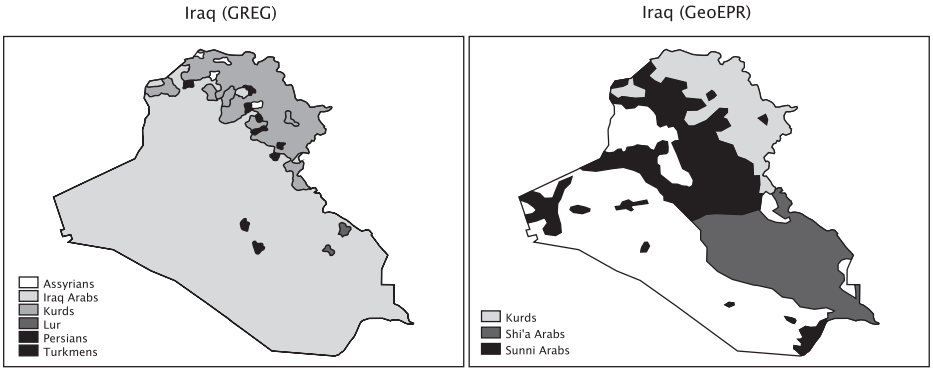


*Figure 2.* Comparison: GREG vs. GeoEPR Coding for Iraq

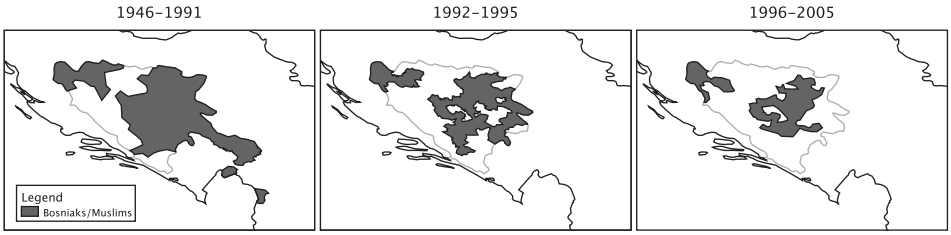


*Figure 3.* GeoEPR Coding for the Bosniak/Muslim Group in Yugoslavia
The borders of the (later independent) constituent Republic of Bosnia and Herzegovina are shown in light grey.

Two examples (Iraq and Yugoslavia) help illustrate the GeoEPR coding and how it differs from GREG (see Figure 2). Based on its linguistic definition of ethnicity, GREG lists a variety of groups for Iraq. Some of these groups have little, if any, political relevance and are therefore excluded from EPR and, therefore, also from GeoEPR. The map in Figure 2 (right-hand panel) shows that according to EPR, there are only three politically relevant ethnic groups in Iraq: Sunni and Shi'a Arabs, and the Kurds. The distinction between Shi'a and Sunni is not included in GREG, because it is not based on language.

Our second example illustrates the temporal coding in GeoEPR. The maps in Figure 3 display the settlement regions of the Bosniak/Muslim group in Yugoslavia and later in Bosnia-Herzegovina, which suffered from large-scale ethnic cleansing during the civil war. There are three geographic periods for this group: 1946–1991, 1992–1995 and 1996–2005, reflecting the "shrinkage" in settlement area as a result of the violent conflict.[4]

[4] Note that this group is listed with two periods in Table 2 during its existence within Yugoslavia, and with a single period as part of Bosnia and Herzegovina.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

## 5. Application

In this section, we demonstrate some potential uses of GeoEPR. While different levels of analysis are possible, including country-level studies (cf. Wimmer et al., 2009), we present a spatial extension of Cederman et al. (2010), who study the outbreak of armed conflict as the result of competing ethno-nationalist claims to state power. The key finding is that ethnic groups excluded from state power are significantly more likely to engage in violent struggle with the government, especially if they experienced a loss of power in the recent past. Thus, certain ethnopolitical power constellations are found to be closely associated with conflict.

Since Boulding's (1962) seminal contribution, however, it has been understood that spatial factors matter for conflict processes. Aspects of geography – distance in particular – affect how power is projected across space. This argument has resonated in the quantitative literature on civil wars, and a host of explanations for this "loss of strength gradient" has emerged. For example, Fearon and Laitin (2003) and others have argued that state capacity, that is, the ability of the state to effectively control territory, is particularly low in the periphery, making rebellion feasible. Building on Wimmer's (2002) theory of ethnic exclusion, Buhaug et al. (2008) suggest that not only material capabilities, but also soft power varies with distance since the state's cultural penetration generally declines in remote areas (see also Cederman et al., 2009). Peripheral groups are therefore less likely to be involved in the process of nation building and therefore most likely to disidentify with the state. Thus, we state our first hypothesis:

> *H1a: The probability of conflict increases with distance between an ethnic region and the capital.*

Buhaug (2006) argues that fighting objectives in civil wars are partly a function of the insurgent's relative capabilities vis-à-vis government forces: since capturing an entire state is more demanding than fighting for local autonomy, peripheral groups are more likely to engage in limited aims and fight on their core territory rather than on the government's "home turf", the capital. Again, however, the argument also extends to motivational factors: peripheral groups excluded from the imagined community of the nation are likely to desire autonomous political control over their region, for example through secession. Thus, we refine our initial hypothesis:

> *H1b: The probability of territorial conflict increases the further away an ethnic group's settlement region is from the capital.*

Whereas these considerations focus on *distance from the capital*, other research has brought forward arguments about the role of *border proximity*. For example, Salehyan (2007) argues that insurgents are likely to find sanctuary in bordering states. Operating from abroad (i.e. outside their home jurisdiction) may allow insurgents to engage in activities that would not be tolerated by the government of their home country. Fighting operations, however, are conditioned by the ability to project force across distance: "the ability to launch and sustain combat operations will critically depend on nearness" (Salehyan, 2009: 37). Extending this argument to ethnic conflicts, Salehyan (2009: 64–65) argues that if the insurgents are operating in the name of, and recruit fighters from, a specific ethnic group, the location of an ethnic region with regard to country borders determines access to sanctuary, as well as other external



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

resources. However, based on aggregate country-level data, Salehyan is unable to test this particular argument statistically (see also Buhaug and Rød, 2006).

Sanctuaries are not the only reason why distance to border matters. For example, Saideman (1997) and others have argued that transnational ethnic ties influence domestic preferences, in particular with respect to secessionist conflict. In this context, we add that a group's relative location is likely to influence the (perceived) feasibility of secession in the first place. After all, if surrounded by the former rump state without shared borders to other states (or sea access), maintaining a new state will be difficult. In sum, these arguments lead to our second hypothesis:

> *H2: The probability of territorial conflict is higher the closer an ethnic region is located to the nearest border.*

## *Data, Measurement and Method*

We test these hypotheses by relying on both EPR and GeoEPR. Thus, the unit of analysis is the group-year, and the dependent variable is group-level conflict onset as coded in Cederman et al. (2010). We restrict our sample to groups that are identifiable in space, including dispersed groups (but excluding nomadic and urban groups).[5] While the hypotheses as such are not new, to date they have not been tested against a spatially explicit dataset that covers all politically relevant groups and avoids the type of sample selection problems discussed above. Thus GeoEPR and EPR together represent ideal data sources to test hypotheses about the relationship between ethnic geography and conflict on a global scale.

We extend the model of Cederman et al. (2010) by adding variables that measure the ethnic groups' distance to the capital and to the nearest international border. We calculate these variables by overlaying GeoEPR with geo-coded data layers drawn from PRIO-GRID (Tollefsen, 2009). This GIS dataset partitions the globe into a global grid of cells. Among others, PRIO-GRID contains information for each grid cell on distance to capital and border, making it ideal for our purposes. We then use GIS software to identify which cells are contained inside the GeoEPR polygons that represent the settlement territories of the various ethnic groups. This in turn allows us to calculate (logged) minimum distance to capital and border at the level of ethnic groups, rather than at the level of grid-cells that are not meaningfully related to relevant actors.[6] We refer the reader to Cederman et al. (2010) for a detailed description of the control variables. Finally, because our dependent variables are categorical, we employ logit and multinomial logit estimators.[7]

---

[5] In other words, we excluded 62 out of 733 groups. Imputing country means for these groups did not alter our results.

[6] We add one unit (km) to each observation to accommodate the log transformation for units with zero distance to capital or border.

[7] Since the multinomial logit relies on the assumption of IIA (independence of irrelevant alternatives), we also validated our results using the computationally more intense multinomial probit estimator which does not assume IIA (Alvarez and Nagler, 1998).



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

## *Results*

Our results are contained in Table 3. Model 1 is a probit model with the dependent variable conflict onset. While also replicating the main findings of Cederman et al. (2010), the model yields strong support for the additional hypotheses introduced in this article. As expected by H1a, the probability of conflict increases the further away a group's settlement region is from the capital, and the closer it is located to borders (H2). Based on the multinomial logit estimator, Model 2 distinguishes between two main conflict types, that is, whether the fighting is over control of limited territory (territorial conflicts) or over control of the state (governmental conflicts). As hypothesized (H1b), distance to the capital and proximity to borders have statistically significant effects only for the territorial conflicts, but not for governmental conflicts. Indeed, a comparison of coefficients across models suggests that the results in Model 1 are likely to be driven by territorial conflicts exclusively. Thus, we conclude that aspects of space matter primarily for territorial conflicts, and less so (if at all) for governmental conflicts.

*Table 3.* (Multinomial) Logit Estimates of Group-Level Conflict Onset

| | *Model* | | |
|---|---|---|---|
| | *1* | *2* | |
| | *All onsets* | *Territorial* | *Governmental* |
| Distance to capital (logged) | 0.175** | 0.297*** | 0.098 |
| | (0.066) | (0.089) | (0.092) |
| Distance to border (logged) | –0.141* | –0.262* | –0.009 |
| | (0.067) | (0.103) | (0.065) |
| Excluded | 1.265*** | 1.162** | 1.330*** |
| | (0.279) | (0.426) | (0.349) |
| Downgraded | 1.764*** | 1.222* | 2.046*** |
| | (0.375) | (0.540) | (0.474) |
| Size (logged) | 0.259** | 0.115 | 0.557*** |
| | (0.083) | (0.102) | (0.121) |
| Previous conflict | 0.773*** | 0.626* | 1.044*** |
| | (0.177) | (0.253) | (0.242) |
| GDP p.c. (lagged, logged) | –0.356** | –0.290 | –0.387** |
| | (0.133) | (0.224) | (0.135) |
| Population (lagged, logged) | –0.101 | –0.003 | –0.297* |
| | (0.115) | (0.147) | (0.145) |
| Peaceyears | –0.100 | –0.165 | –0.055 |
| | (0.075) | (0.086) | (0.128) |
| Constant | –0.000 | –0.001 | 0.000 |
| | (0.001) | (0.001) | (0.001) |
| Observations | 20018 | 20018 | |
| Log likelihood | –682.5 | –732.9 | |

Robust standard errors (clustered by country) in parentheses.
Splines for peaceyears omitted.
***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

Our results provide systematic support for the proposition that relative location of ethnic groups plays an important role in the study of civil war and especially for territorial conflicts. This, in turn, lends support to the theoretical distinction between conflicts fought over control over territory and those fought over control of the state (see Buhaug, 2006; Wimmer et al., 2009).

## 6. Discussion and Conclusion

This article introduced GeoEPR, a GIS-based dataset charting politically relevant ethnic groups across space and time. GeoEPR improves significantly upon GREG, the only comparable dataset. However, it also leaves room for future improvements to address two primary limitations. First, GeoEPR does not provide geo-coding for groups classified as 'urban'. This pertains only to a small number of groups often of immigrant origin, such as Afro-Caribbeans in the United Kingdom, or Koreans in Japan. Second, although GeoEPR is dynamic, it merely captures *major* changes in ethnic geography, but leaves out short-term changes, such as those induced by refugee flows. Indeed, geo-periods by definition can only approximate the underlying dynamics. Future data collection efforts will have to improve upon the level of detail provided in GeoEPR (for studies relying on high resolution data, see e.g. Lyall, 2009; Kalyvas and Kocher, 2009).

Leaving aside these limitations, GeoEPR clearly provides a valuable resource for researchers studying questions related to ethnic geography. By focusing on politically relevant groups as actors, the data allow researchers to test new hypotheses at a level of measurement closer to the location of causation than before. We believe that GeoEPR's full potential can best be tapped when used in conjunction with other spatial data. Thus, by studying particular aspects of the space within which collective actors operate, researchers will be able to move closer to the actual mechanisms at work in conflict processes. Indeed, GIS data from other sources can easily be overlaid with GeoEPR, as illustrated for the case of physical distance by Buhaug et al. (2008), as well as Cederman et al. (2009). Such a "cookie-cutter" approach allows for a host of future applications. For example, by measuring the topography, climate or natural resources located within the settlement area of politically relevant groups, it will be possible to make new measurements of economic context and resources available to ethnic organizations. This in turn will allow researchers to test hypotheses regarding the role of terrain in providing rebel hideouts, local variation in the effects of climate change that may affect conflict propensity, or horizontal economic inequalities between ethnic groups that produce political grievances and encourage armed rebellion.

## References

Alvarez, Michael R., and Jonathan Nagler. 1998. When politics and models collide: Estimating models of multiparty elections. *American Journal of Political Science* 42(1): 55–96.

Boulding, Kenneth E. 1962. *Conflict and Defense: A General Theory*. New York: Harper and Row.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

Brubaker, Rogers. 2004. *Ethnicity without Groups*. Cambridge, MA: Harvard University Press.

Bruk, Solomon I., and V. S. Apenchenko, eds. 1964. *Atlas Narodov Mira*. Moscow: Glavnoe upravlenie geodezii i kartografii gosudarstvennogo geologicheskogo komiteta SSSR and Institut etnografii im. H. H. Miklukho-Maklaia, Akademiia nauk SSSR.

Buhaug, Halvard. 2006. Relative capability and rebel objective in civil war. *Journal of Peace Research* 43(6): 691–708.

Buhaug, Halvard, and Scott Gates. 2002. The geography of civil war. *Journal of Peace Research* 39(4): 417–433.

Buhaug, Halvard, and Jan Ketil Rød. 2006. Local determinants of African civil wars, 1970–2001. *Political Geography* 25(3): 315–335.

Buhaug, Halvard, Lars-Erik Cederman and Jan Ketil Rød. 2008. Disaggregating ethno-nationalist civil wars: A dyadic test of exclusion theory. *International Organization* 62(3): 531–551.

Cederman, Lars-Erik, and Luc Girardin. 2007. Beyond fractionalization: Mapping ethnicity onto nationalist insurgencies. *American Political Science Review* 101(1): 173–185.

Cederman, Lars-Erik, and Kristian Skrede Gleditsch. 2009. Introduction to special issue on disaggregating civil war. *Journal of Conflict Resolution* 53(4): 487–495.

Cederman, Lars-Erik, Halvard Buhaug and Jan Ketil Rød. 2009. Ethno-nationalist dyads. *Journal of Conflict Resolution* 53(4): 496–525.

Cederman, Lars-Erik, Andreas Wimmer and Brian Min. 2010. Why do ethnic groups rebel? New data and analysis. *World Politics* 62(1): 87–119.

Chandra, Kanchan, and Steven Wilkinson. 2008. Measuring the effect of "ethnicity". *Comparative Political Studies* 41(4–5): 515–563.

Cunningham, David E., Kristian Skrede Gleditsch and Idean Salehyan. 2009. It takes two: A dyadic analysis of civil war duration and outcome. *Journal of Conflict Resolution* 53(4): 570–597.

Fearon, James D. 2003. Ethnic structure and cultural diversity around the world: A cross-national data set on ethnic groups. Typescript Stanford University (http://www.stanford.edu/group/ethnic/workingpapers/egroups.pdf).

Fearon, James D., and David D. Laitin. 2003. Ethnicity, insurgency, and civil war. *American Political Science Review* 97(1): 75–89.

Gilmore, Elisabeth, Nils Petter Gleditsch, Paivi Lujala and Jan Ketil Rød. 2005. Conflict diamonds: A new dataset. *Conflict Management and Peace Science* 22(3): 257–272.

Gordon, Raymond G., ed. 2005. *Ethnologue: Languages of the World*, 15th edn. Dallas, TX: SIL International.

Gurr, Ted Robert. 1993. *Minorities at Risk: A Global View of Ethnopolitical Conflict*. Washington, DC: United States Institute of Peace Press.

Hegre, Håvard, and Clionadh Raleigh. 2008. Population size, concentration and civil war: A geographically disaggregated analysis. *Political Geography* 28(4): 224–238.

Kalyvas, Stathis N. 2008. Ethnic defection in civil war. *Comparative Political Studies* 41(8): 1043–1068.

Kalyvas, Stathis N., and Matthew Adam Kocher. 2009. The dynamics of violence in the Vietnam War: An analysis of the Hamlet Evaluation System (HES). *Journal of Peace Research* 46(3): 335–355.

Laitin, David D., and Daniel N. Posner. 2001. The implications of constructivism for constructing ethnic fractionalization indices. *APSA-CP* 12(1): 13–17.

Levinson, David. 1998. *Ethnic Groups Worldwide: A Ready Reference Handbook*. Phoenix, AZ: Oryx.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011

Lujala, Paivi, Jan Ketil Rød and Nadja Thieme. 2007. Fighting over oil: Introducing a new dataset. *Conflict Management and Peace Science* 24(3): 239–256.

Lyall, Jason. 2009. Does indiscriminate violence incite insurgent attacks? *Journal of Conflict Resolution* 53(3): 331–362.

Min, Brian, Lars-Erik Cederman and Andreas Wimmer. 2010. Ethnic exclusion, economic growth, and civil war. Typescript. University of Michigan, ETH Zürich, and UCLA.

Posner, Daniel N. 2004. Measuring ethnic fractionalization in Africa. *American Journal of Political Science* 48(4): 849–863.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research* 47(5): 651–660.

Saideman, Stephen M. 1997. Explaining the international relations of secessionist conflicts: Vulnerability versus ethnic ties. *International Organization* 51(4): 721–753.

Salehyan, Idean. 2007. Transnational rebels: Neighboring states as sanctuary for rebel groups. *World Politics* 59(2): 217–242.

Salehyan, Idean. 2009. *Rebels Without Borders: Transnational Rebels in World Politics*. Ithaca, NY, and London: Cornell University Press.

Taylor, Charles Lewis, and Michael C. Hudson. 1972. *World Handbook of Political and Social Indicators*, 2nd edn. New Haven, CT: Yale University Press.

Toft, Monica Duffy. 2003. *The Geography of Ethnic Violence*. Princeton, NJ: Princeton University Press.

Tollefsen, Andreas F. 2009. PRIO Grid. Presented at the workshop on Environmental Conflicts and Conflict Environments, PRIO, 11–12 June.

Weber, Max. 1978. *Economy and Society: An Outline of Interpretative Sociology*. Berkeley, CA: University of California Press.

Weidmann, Nils B. 2009. Geography as motivation and opportunity: Group concentration and ethnic conflict. *Journal of Conflict Resolution* 53(4): 526–543.

Weidmann, Nils B., Jan Ketil Rød and Lars-Erik Cederman. 2010. Representing ethnic groups in space: A new dataset. *Journal of Peace Research* 47(4): 491–499.

Wimmer, Andreas. 2002. *Nationalist Exclusion and Ethnic Conflicts: Shadows of Modernity*. Cambridge: Cambridge University Press.

Wimmer, Andreas, Lars-Erik Cederman and Brian Min. 2009. Ethnic politics and armed conflict: A configurational analysis of a new global data set. *American Sociological Review* 74(2): 316–337.



Downloaded from cmp.sagepub.com at UCLA on March 30, 2011