Toward realistic computational models of civil wars*

Lars-Erik Cederman Luc Girardin

International Conflict Research
ETH Zurich
8092 Zurich, Switzerland
{lcederman@ethz.ch, girardin@icr.gess.ethz.ch}

August 21, 2007

Abstract

This paper reports on our ongoing efforts to build increasingly accurate computational models of civil wars, and sketches a computational research program for this purpose. Because our research efforts are still very much work in progress, we focus on methodological problems and developments rather than presenting new substantive results. We propose solutions to the conceptual and methodological obstacles that stand in the way of progress. In particular, we suggest ways to integrate data generated in geographic information systems (GIS) with computational models based on a new agent-based platform called GROWLab. A number of data projects supporting these efforts are described including Geo-Referencing of Ethnic Groups (GREG) and Expert Survey of Ethnic Groups (ESEG).

*) Paper prepared for presentation at the Annual Meeting of the American Political Science Association, to be held in Chicago, August 30-September 2, 2007. We thank Nils Weidmann for his helpful input.

1. Introduction

Traditionally, agent-based modelers have analyzed interstate warfare, but they have been slower to embrace civil wars as their research focus (though see e.g. Epstein 2002; Bhavnani 2006). This paper reports on our ongoing efforts to build increasingly accurate computational models of civil wars, and sketches a computational research program for this purpose. Because our research efforts are still very much work in progress, this paper discusses methodological problems rather than presenting new substantive results. We propose solutions to the conceptual and methodological obstacles that stand in the way of progress. In particular, we suggest ways to integrate data generated in geographic information systems (GIS) with agent-based models.

In political science, the popularity of agent-based modeling has increased. While doubt persists in some quarters, the method has come to be accepted as a useful complement to other formal modeling techniques (Cederman 2001a; Lustick and Miodownik 2007). However, very few models of this type have had a major impact on substantive research in particular subfields of the discipline. With the exception of Schelling's (1978) famous segregation model and Axelrod's (1984) work on evolutionary games, it is hard to name any computational model that has changed the way we think about political phenomena in a profound way.

This insight is the reason why we decided to launch a quest for more realistic models. Whereas the most influential agent-based models have been very abstract, as suggested by the work of Schelling and Axelrod, it would be desirable to generate computational results that lie closer to reality. Using the civil war literature as the main target, our goal is to create a new generation of empirically anchored agent-based models that has the potential to contribute directly to substantive theorizing.

2. Existing computational models of warfare: GeoSim

Geopolitics belongs to the core research agenda of agent-based modelers. Inspired by Realism, the dominant paradigm of International Relations, such modeling efforts typically feature states as autonomous territorial actors embedded in a decentralized system that features balance-of-power equilibria. In a pioneering paper, Bremer and Mihalka (1977) introduced such a model of geopolitical competition. It features state-like organizations with dynamic borders that grow through conquest.

Building on Bremer and Mihalka's original model, Cusack and Stoll (1990) presented a book-length study that analyzes the internal consistency of the realist paradigm. This study includes a systematic set of computational experiments featuring an extended set of decision-making rules. Civil wars were also modeled, but because of the complexity of the model and the limited computational resources available to the authors, the book does not contain any systematic replications of internal conflict.

The GeoSim framework

This research tradition became an important source of inspiration for the GeoSim project. Introduced by Cederman (1997) and modeled from scratch, GeoSim is a family of agent-based models that is based on a dynamic network of interstate relations superimposed on a square grid. In the basic model, all interactions are local, between adjacent states. Each state capital can absorb and dominate a number of provinces in a perfectly Hobbesian fashion. Moreover, their borders are sharply defined. Finally, they derive their power from the number of provinces they control -- thus, the larger a state is, the more powerful it is. Figure 1 illustrates a typical system with dark lines marking state borders and the dots the capitals. The neighborhood relations of a selected state are shown as bidirectional arrows.

The main thrust of the GeoSim research program is to study interstate warfare as a consequence of geopolitical changes affecting the boundaries of states (Cederman 2002). Within this context, it is possible to apply the framework to study the effect of defensive alliances and technology (Cederman 1997, Chap. 4), democracy (Cederman 2001b), and democratization (Cederman and Gleditsch 2004) on interstate conflict processes, as well as the nature of such processes, including war-size distributions based on casualty levels (Cederman 2003).

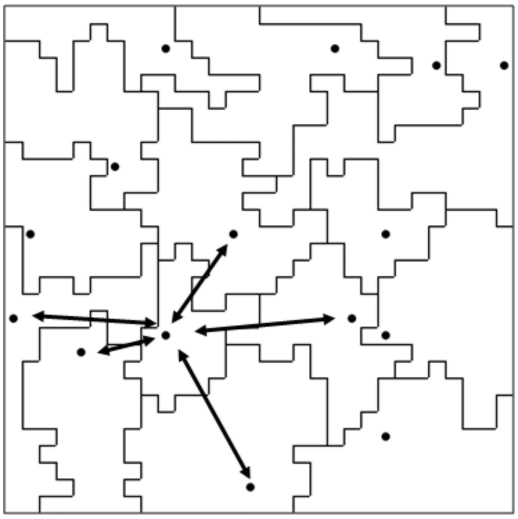

Figure 1. A typical state system in GeoSim with interstate relations shown as arrows

Another stream of research based on the GeoSim framework opens the black box of the state by allowing provinces to stage rebellions against the central power of the state. Focusing on nationalist challenges to the capital within a single state, Cederman (1997: Chap. 8) presents an early attempt to relax the assumption of unitary states. This model introduces a straight-forward center-periphery logic to the model that follows the hierarchical structure of the states. Here the provinces manage to overcome the power balance in favor of the center by subscribing to nationalist platforms. It is shown that the more acute the oppression exerted by the center, the more likely it is that the peripheral actors manage to find a least common denominator based on a “thin” rather than a specific identity.

The nationalist insurgency model

Going a step beyond this simple research design, Cederman (forthcoming) proposes another variation on the GeoSim structure, here labeled the nationalist insurgency model (NIM), in which each simulation run consists of an entire state system, albeit with fixed state borders. This model was developed to reconstruct the mechanisms that drive conflict in center-periphery relationships within states. In an oft-cited study, Fearon and Laitin (2003) suggest that weak states characterized by rough terrain are especially conflict prone because they are incapable of controlling their territories, thus opening a window of opportunity to rebels. However, their findings do not lend any support to hypotheses that connect ethnicity with civil wars.

Challenging this claim, Cederman (forthcoming) manages to reproduce Fearon and Laitin's results on state size and terrain in an artificial world while insisting that ethnicity and nationalism play a key role in generating conflict. This means that the NIM features a center-periphery logic with a cultural dimension rather than being merely materialist. The computational experiments show that civil wars may result where the state is incapable of full cultural penetration resulting in incomplete nationalist mobilization and peripheral collective action based on common identities that challenge the control of the center.

In order to generate findings on nationalist insurgencies, the NIM features a number of additions to the standard interstate framework. Instead of including merely one "layer" pertaining to relations among states, the current model introduces an explicit model of terrain as well as a multi-dimensional cultural map similar to the one proposed in Axelrod (1997: Chap. 7). Inspired by Holland's (1995) schema representation, national identities are modeled as computational coalitions defined in terms of a culture string with "wild cards" that is superimposed on top of the cultural landscape.

Figure 2 shows a 3D snapshot of the model that reveals the rugged physical landscape. Here state borders appear as lines and the state capitals as half spheres. Both capitals and provinces are marked by colored disks if they possess a national identity. Rebellions are shown as vertical "needles". As expected, the projections suggest that most of the fighting takes place in mountainous areas.

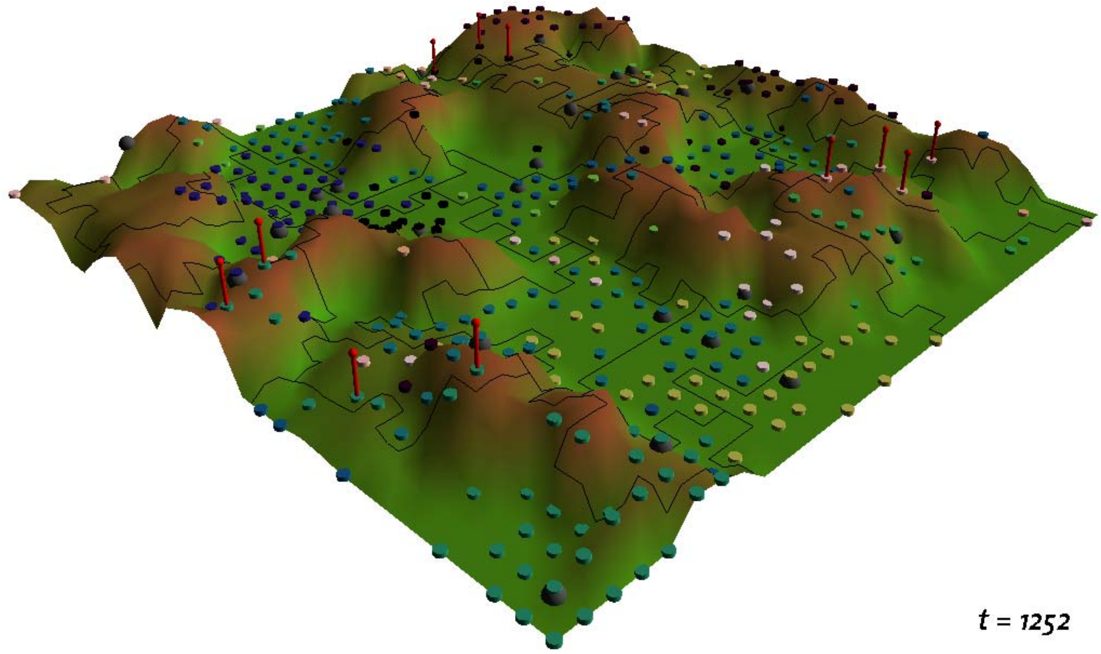

Figure 2. A 3D-projection of the nationalist insurgency model showing rebelling provinces as red needles.

This representation goes beyond models of culture as fixed and immutable properties that influence behavior (Epstein 2002) or as endogenous vectors whose traits all matter (Axelrod 1997, Chap. 7). Identities, however, are more selective, because only politically relevant aspects of culture enter into the power calculus. Allowing for such variation, Lustick's (2000; 2002) agent-based models ABIR and PS-I feature endogenous repertoires composed of sets of identities, but the component identities stand in no specific relationship to each other and thus do not describe a coherent cultural space. Moreover, these, and almost all other models of this type, fail to provide an explicit representation for formal political organizations or terrain.

While the NIM offers considerable flexibility to represent complex phenomena, it is entirely heuristic. Other than reproducing similar macro results as those found by Fearon and Laitin (2003), no attempt was made to calibrate the key objects to real-world conditions. This raises the question of whether it is possible to build agent-based models that conform more closely to real-world phenomena.

3. Data projects: Gathering new empirical information

In the following, we describe the steps that we have taken more recently to put our modeling activities on a more solid empirical footing. While our analytical inspiration comes directly from the GeoSim family of models, the main challenge consists in gathering systematic empirical information in order to render the computational framework more accurate. As was the case in the NIM, this research focuses entirely on ethno-nationalist civil wars following a center-periphery logic, thus setting aside other types of wars that are not ethnic or that involve no hierarchy, such as communal riots.

Beyond fractionalization

In particular, our previous computational research suggests that it is important to determine the relative power of the ethnic groups and their geographic location. Therefore we launched several data projects that serve to investigate these issues. The first one, reported in Cederman and Girardin (2007) replaces the conventional index of ethnic fractionalization with a measure that we call *N**. Directly drawing on the center-periphery of the NIM, this index measures the extent to which peripheral ethnic groups, so-called "marginalized ethnic groups" (MEGs), are excluded from state power in states by comparing their demographic share to that held by the "ethnic group in power" (EGIP). Figure 3 illustrates this basic center-periphery configuration. For the countries that have so far been coded, we have obtained suggestive results when we regress ethnic civil wars on *N** while controlling for the variables used by Fearon and Laitin (2003).

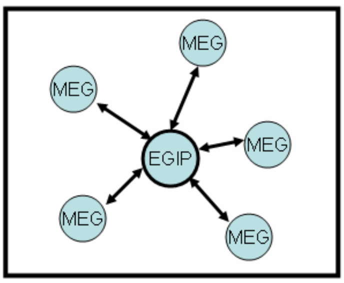

Figure 3. The basic center-periphery configuration

Geo-Referencing of Ethnic Groups (GREG)

These results suggest that the center-periphery logic of the NIM is sound. However, it says little about the geographic dimension. In particular, the demographic approximation of group power leaves much to be desired in cases where small peripheral groups are able to take advantage of weak state reach due to long distances and rough terrain. For this reason, we initiated a second project that puts real empirically observed ethnic groups on the map. This data project that serves to geo-reference ethnic groups around the world (Cederman, Rød and Weidmann 2007).

Relying on maps and data drawn from the classical Soviet *Atlas Narodov Mira*, it is possible to use geographic information systems (GIS) to represent the groups as polygons. The Atlas has several strengths: it is complete and carefully researched, it relies on a uniform group list that is valid across state borders, and it provides high-quality maps. Among the weaknesses, it should be stressed that the Atlas is based on the situation in the 1960s and thus clearly outdated compared to the current situation. However, in most cases, ethnic settlement patterns exhibit considerable inertia, so it seems reasonable to use this dataset as a starting point. It is possible to reconstruct population shares with GIS procedures. Figure 4 provides a snapshot of geo-coded ethnic groups in the former Yugoslavia.

In a recent paper which builds directly on the GREG data, Buhaug, Cederman and Rød (2007) have been able to confirm the results of Cederman and Girardin (2007). As in the computational framework, they focus on relation-specific causes of ethnic conflicts. The goal is to disaggregate both ethnicity and conflict to the level of explicitly geo-coded center-periphery dyads. This is an important goal, because, so far, conflict data has typically been recorded at the country level (Sambanis 2004). Whereas this is a satisfactory simplification in small countries with few conflict groups, the results can be seriously misleading in the case of large countries, such as Russia. The political, cultural and geographic conditions pertaining to Chechnya do not necessarily apply elsewhere in the country.

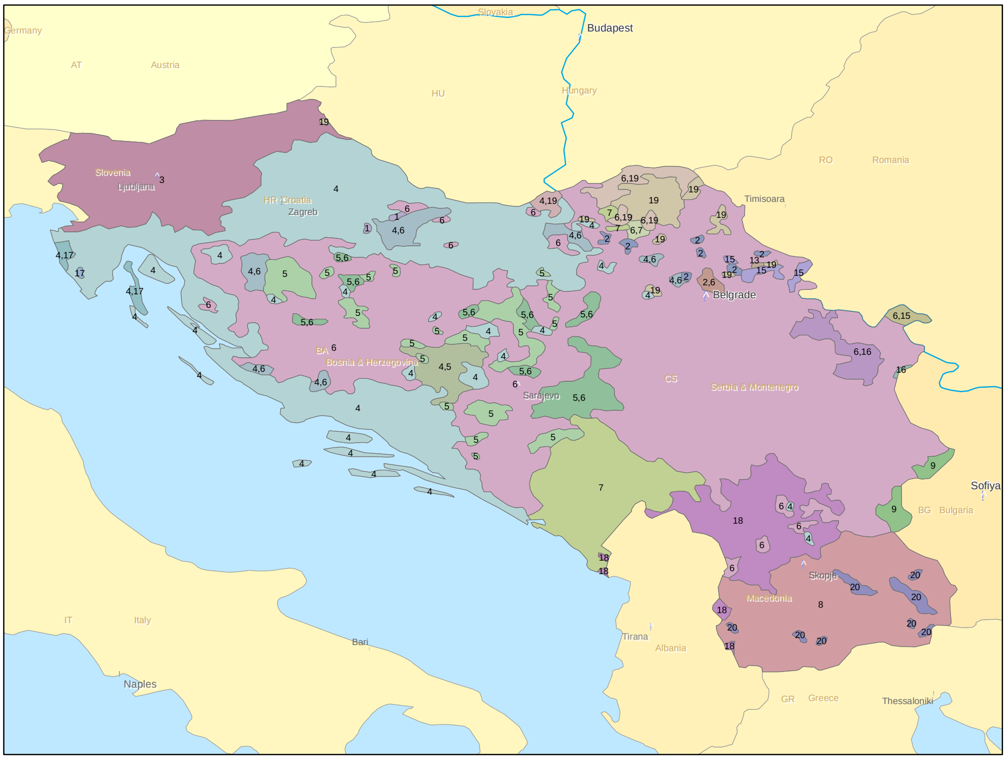

Figure 4. Geo-coded map of the former Yugoslavia with ethnic groups represented by polygons.

Focusing specifically on ethno-nationalist conflicts pitting peripheral ethnic groups against central governments, this disaggregated statistical analysis makes it possible to measure the center-periphery power balance as demographic proxies while controlling for distances and terrain. In addition to establishing a link between the dyads and conflict data, it is essential to pinpoint the location of the ethnic groups themselves.

The results confirm the initial study of Cederman and Girardin (2007). Figure 5 charts the association between the dyadic power balance and the risk of ethno-nationalist conflict. The lower, solid line shows the marginal effect of the excluded group's share of the dyadic population, holding all other covariates at their median values. Obviously, peripheral groups that face vastly superior EGIPs are not likely to rebel, but the risk of conflict increases markedly with the relative power of the marginalized group. In the middle, dashed plot, we changed the distance from the capital to the peripheral group from the 50^{th} to the 95^{th} percentile value. This is associated with a considerably higher overall risk of conflict.

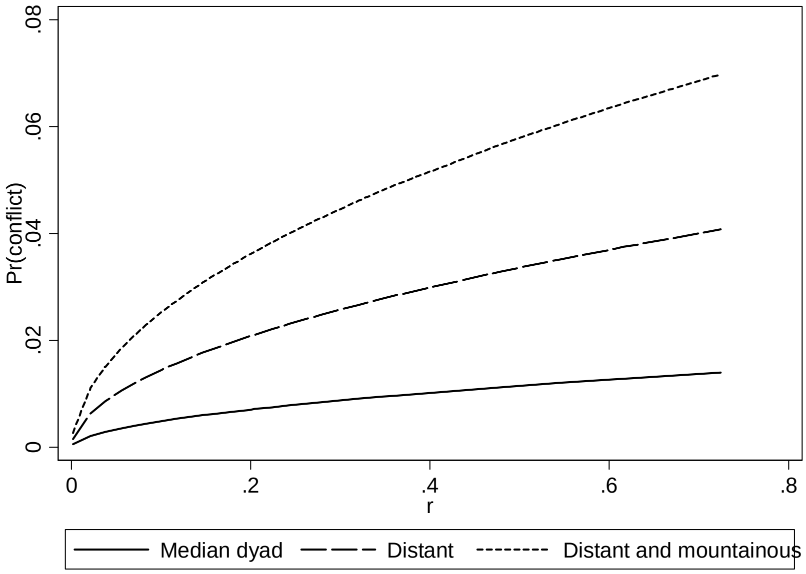

Figure 5: The effect of the dyadic power balance *r* using data from the *ANM*

Expert Survey on Ethnic Groups (ESEG)

Although the results presented in Cederman and Girardin (2007) and Buhaug, Cederman and Rød (2007) are promising, they suffer from many shortcomings. First, the notion of EGIP is quite crude, because it implies that either a group is included in the EGIP or it is not. This makes it hard to capture intermediate situations including those involving different levels of influence of senior and junior partners in power-sharing arrangements. Second, the original coding of EGIP is relatively ad hoc, especially with respect to power sharing arrangements. Furthermore, the EGIP coding was based on Fearon's group list (2003) or on the Soviet *ANM* without further consideration as to whether these are actually politically relevant. Third, because of the difficulties in determining groups' access to power outside Eurasia and North Africa, EGIP coding only exist for this restricted area. Finally, to simplify the coding effort and the construction of the index, Cederman and Girardin (2007) ignored temporal variations despite the existence of historical changes affecting the composition of EGIP in many cases, such as Iraq before and after the fall of Saddam Hussein.

As a response to these problems, we therefore launched a web-based expert survey of ethnic groups (or ESEG for short) in collaboration with Andreas Wimmer at UCLA that aims to address all these four dimensions (see Cederman, Girardin and Wimmer 2006). Our data collection efforts serve to produce a more precise and systematic conceptualization of access to or exclusion from power that does not rely on demographic measurements. Furthermore, by relying on expert coders, the ESEG project promises to be far more accurate and well informed about specific regions than Cederman and Girardin's (2007) initial EGIP coding. Equally importantly, the new instrument extends the coverage to the entire world. Finally, we have designed our survey instrument in such a way that the coders are able to capture major changes in the political relevance of particular groups and their access to power over time. Figure 6 displays the web-based interface of ESEG.

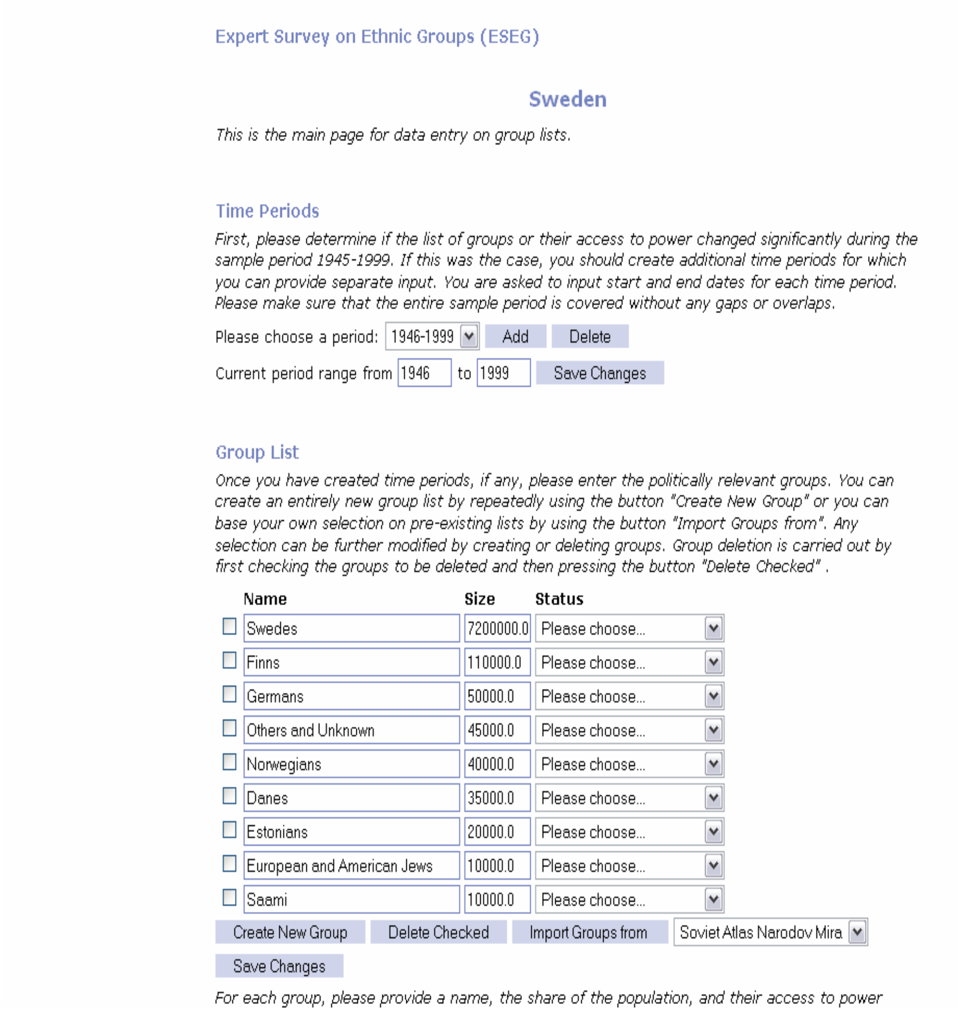

Figure 6. The web-based interface of ESEG

In the summer of 2007, the first version of the complete ESEG dataset was assembled and data on the corresponding dyadic conflicts were collected. We have already been able to conduct preliminary analysis, which indicates that the exclusion perspective holds at both the country and dyadic levels of analysis.

GREG-II -- Putting ESEG on a spatial basis

It would be desirable to conduct spatial analysis based on ESEG. However, so far, GREG is the only geo-coded database on ethnicity that is available because ESEG does not offer information about group locations. Unfortunately, the group lists of GREG and ESEG are not compatible, so there is an urgent need to add spatial information to ESEG. For this reason, we initiated a follow-up data project, that we call GREG-II, as a way to update the geo-coded information of GREG based on the group lists provided by ESEG. In some cases, the differences between the two datasets are minimal, but in most cases, at least some additional cartographic information will be required. In addition, we plan to introduce period-dependent coding for the regions that have seen considerable ethnic resettlement due to voluntary or forced migration, such as the former Yugoslavia. It is our goal to integrate GREG-II as a part of the ESEG data structure.

The empirical work has already produced important insights about ethnicity and civil wars. By disaggregating the conventional country-level analysis to the level of center-periphery dyads, we are now in a position to formulate and validate hypotheses that feature explicit, group-level mechanisms and motivations. However, the empirical analysis has so far been entirely static. In order model conflict processes and their constitutive mechanisms dynamically, it is necessary to return to computational modeling.

4. New Computational Tools: GROWLab

As we have seen, the Nationalist Insurgency Model (NIM) creates an artificial world that can effectively be used for exploratory modeling. On the other hand, the availability of spatially disaggregated data is about to revolutionize the analysis of geopolitical systems. The obvious next step toward a higher degree of realism requires a tighter coupling of the already collected GIS data and our computational tools (Brown et al. 2005).

GROWLab

Rather than continuing to build directly on GeoSim, we therefore decided to create the next generation of frameworks for geopolitical simulation. The result is GROWLab, a software toolbox that facilitates modeling, simulation, analysis, and validation of complex social processes, with a special focus on civil and regional wars, see http://www.icr.ethz.ch/research/growlab. Moreover, it allows the seeding of the model with empirical facts, including GIS-based data, to calibrate the environments and mechanisms to the appropriate level of realism.

The computational framework that we created builds on the same tradition as other agent-based toolkits such as Swarm (Minar et al. 1996) and Repast (Collier et al. 2003). As its predecessors, it aims at supporting three key aspects of the modeling and simulation process, namely:

1. Efficient structuring of the agents relationships and their environment
2. Facilitation of data collection about the state the model and its dynamic
3. Control of scheduling of simulations

The strengths of GROWLab lie in its ability to model complex network and hierarchical relationships between model actors and its versatility to perform statistical and visual analysis of the state of the system and of the unfolding of the processes over time. It is also designed to perform large number of simulation runs on a grid consisting of many independent computers to test the sensitivity of the models.

GROWLab relies on three core concepts to represent agents and their relations: layers, topologies and configurations. A *layer* is a collection of alike and unitary actors. The layers offer general functionality to manage the agents contained in them, but can also be used to collect aggregate data about the entire population. Each layer does not know about the neighborhood relations of its agents – instead, such information can be represented by imposing one or more *topologies* on the layer in question. A topology defines an agent's neighbors and can compute the distance between agents. Whereas topologies can exist only between agents of the same kind, GROWLab offers the possibility to connect agents of different types to create agent hierarchies. This is done using *configurations*, which typically connect agents from two layers. A model structure created with these building blocks is automatically kept synchronized:

an agent removed from a layer is also removed from all the other data structures defined on that layer.

An important type of layer is the *mapping*. It extends the basic layer functionality by adding the possibility to map agents to locations in a *space*. Different types of mappings give the flexibility to map one or many agents to a single location, and vice-versa. A *space* is simply a set of discrete locations. *Spaces* can either be created artificially (e.g. an abstract grid or torus) or can be read from real-world GIS data. Using these data structures, it becomes straightforward to represent neighborhood relations and to carry out geodetic distance computation.

Illustrating the possibilities of applied research, Figure 7 shows a basic structure modeling countries, ethnic groups, and micro-level geographic units. The countries are stored in a Country *layer*. Each country object has attributes such as name, ISO, FIPS, and COW codes. In addition, it is linked to all the locations the country occupies in the space using a *mapping* (one-to-many relationship). The Group *layer* represents the occurrence of ethnic groups in a country. More than one group can be in a given location (many-to-many) and information is provided about the political status of a group in a given country. A *configuration* is used for querying the groups in a country, and the country of a group. Each Cell is linked to one location in the *space* by means of a *mapping* (one-to-one). Geographic information at the cell level is provided as attributes: elevation, population, and GDP. Finally, a *topology* gives allows retrieval of neighbors and the computation of distances between two locations.

For each of the three core concepts introduced above, GROWLab has a set of predefined dynamic graphical visualizations for the inspection of the model. Layers can be portrayed by a list of agents and their attributes. Neighborhood relationships of a topology can be displayed graphically as a network structure, and textually as a paired list of connected partners. The structure of a configuration can be examined as a tree table. A set of two-dimensional graphical displays takes care of visualizing spatial layers and the agents contained in them.

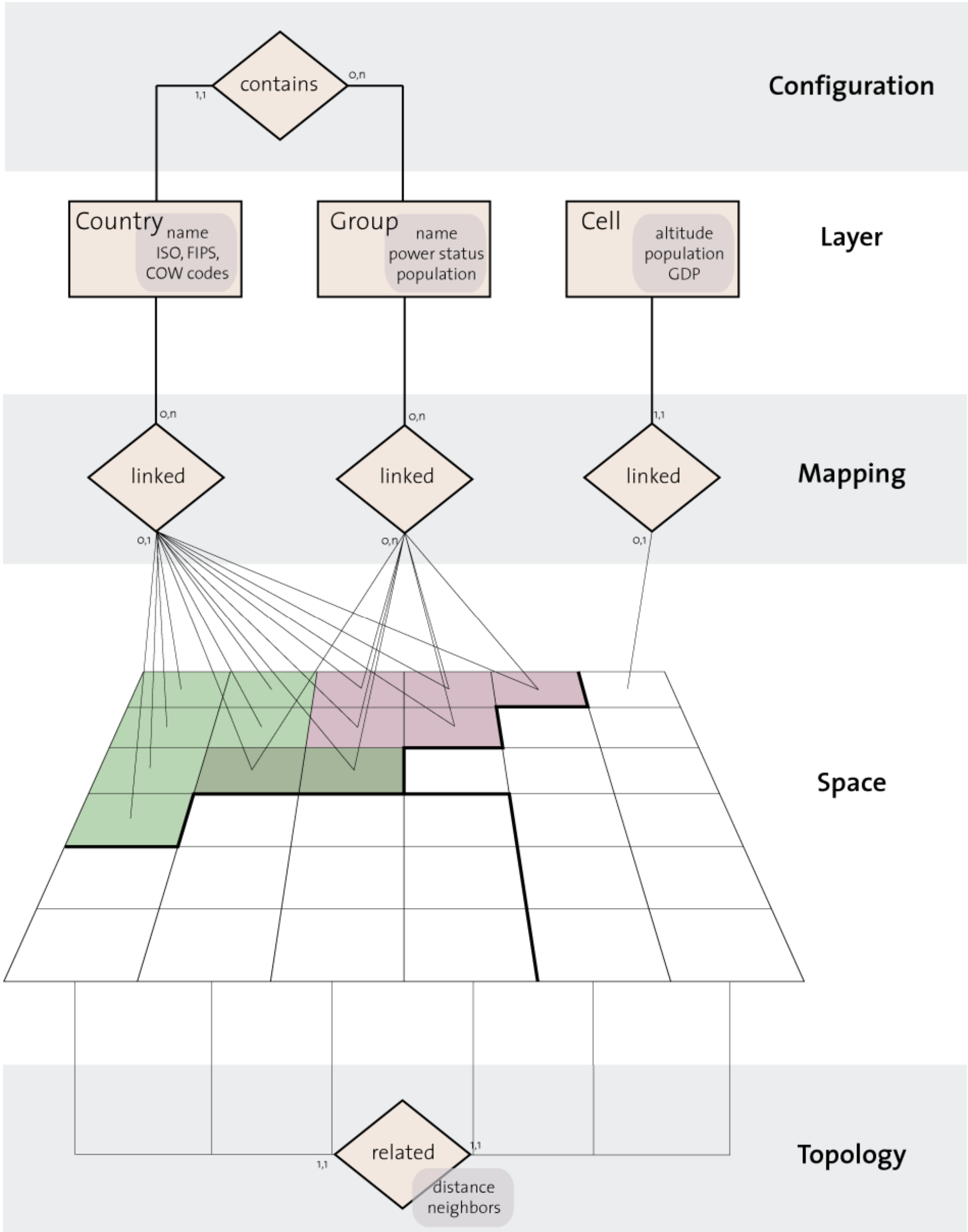

Figure 7. The basic structure of GROWLab applied to a multiethnic state

Figure 8 displays a screenshot of the GROWLab user interface. The upper part of the system is devoted to the control of the simulation. The upper left part shows how agents are linked with one another through various *configurations*; the right part of the user interface illustrates their representation in space. Finally, the lower part of the display is used for a process-oriented view, typically used for showing time series using table or graphs. This particular example informs on the current state of our attempt to capture the link between ethnicity and civil violence. We use this basic setup to test specific causal mechanisms that connect different types of ethno-nationalist configurations with the outbreak and extension of internal conflict. Our goal is to explain such phenomena as integrated parts of spatiotemporal macro processes.

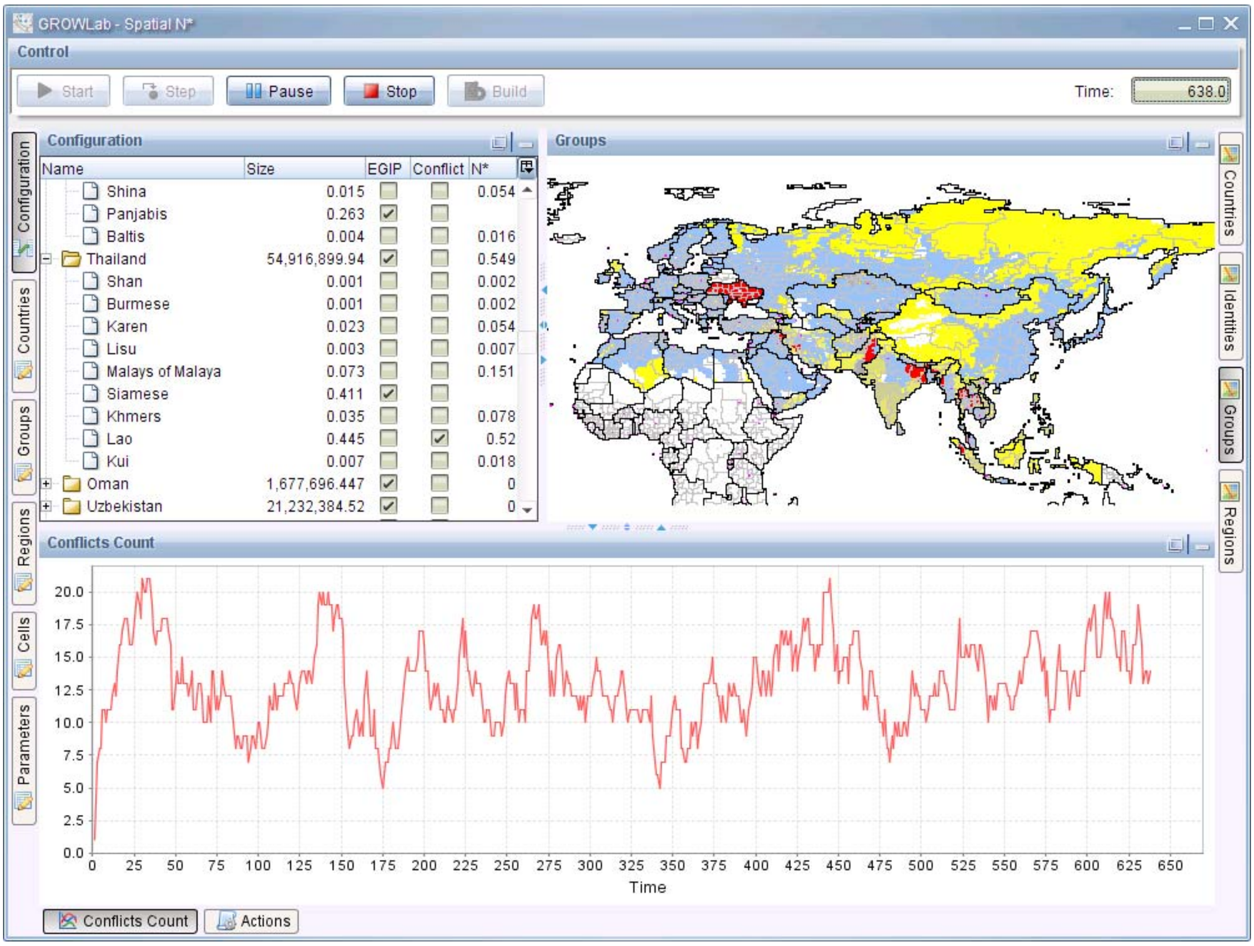

Figure 8. The user interface of GROWLab

GeoModel: A geopolitical template based on GROWLab

While the primary focus of GROWLab is to create simulation models to test various hypotheses, it can also be used for mere data extraction purpose and plain statistical analysis. Based on the template model GeoModel, geo-coded real-world data can be integrated in the modeling process. This template can be extended by inheriting the built-in functionality and by adding some custom behaviors and mechanisms or complement it with additional layers of data.

GeoModel's default space is a rasterized representation of the entire globe, using a WGS84 projection. The raster can be used in two different resolutions: 15'-by-15' (15 arc minute square, ~30km) and 30'-by-30' (~60km). All the geographic data is based on this space. The space has functionality to retrieve locations by latitude/longitude coordinates, and to compute the distance between locations ("as the crow flies" – based on the approximation of a spherical earth).

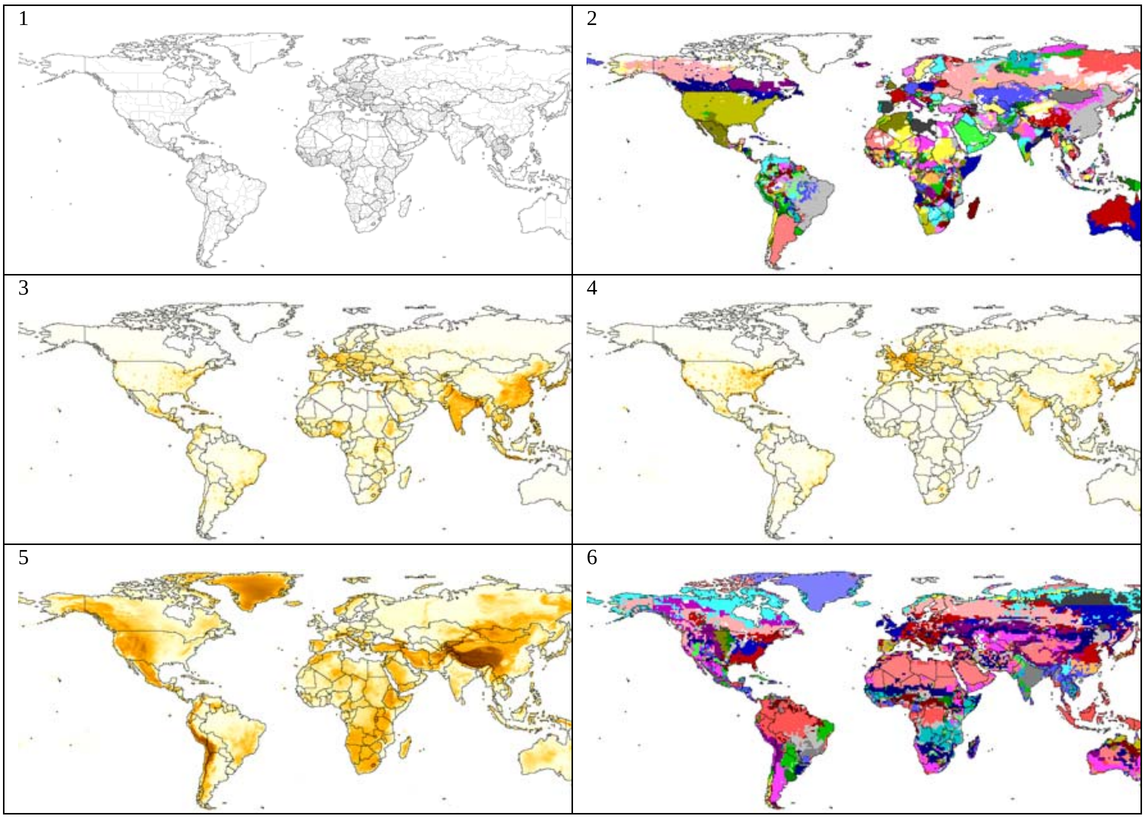

Figure 9. Examples of data contained in GeoModel

Figure 9 shows some of the information contained in the GeoModel template including (1) country border and administrative divisions, (2) ethnic groups across countries, (3) population density, (4) spatial GDP figures, (5) elevation data, and (6) vegetation type. For each country, we provide their borders as of 1964 and 1994, and also try to reconcile their ISO, FIPS and COW codes through customized mapping. To check adjacency of countries, the Minimum Distance Data from Gleditsch and Ward (2001) is also included to query for neighboring countries that are separated by water. At this point, all ethnic groups are directly based on the GREG definitions. For each ethnic group in a country, there is also information about the “ethnic group in power” (EGIP) coding of Cederman and Girardin (2007). In later versions, information from ESEG and GREG-II will be directly integrated.

In addition, we provide disaggregated data for every cell in the system for population (downsampled from the Gridded Population of the World v. 3) and elevation (downsampled from GTOPO30), as well as all the attributes, including GDP estimates, compiled by the G-Econ project (at a 1-degree resolution) from Nordhaus (2006). All of these relieve the modeler from the tedious task of having to collect and merge complicated datasets with one another and provide a prototyping environment where theories can be rapidly tested and validated.

5. Outlook: Future research

Let us conclude with a brief discussion of our future empirical research plans based on GROWLab. Because the main effort has so far centered on developing the infrastructure of the simulation environment, less work has been done as regards actual application of GROWLab to substantive research questions. However, the promise of the new simulation framework should be obvious, especially compared to existing modeling platforms such as GeoSim. One of the most serious shortcomings of previous modeling efforts has been the relatively arbitrary specification of the main causal mechanisms. With the help of our new data projects, and the possibility of integrating information from them directly into the computational model specification, it becomes possible to render conflict processes in a much more realistic way.

An example of such mechanism-based research is Nils Weidmann's (2007) modeling of ethnic groups' settlement patterns. Scholars have frequently mentioned the link between a group's settlement pattern and its involvement in violent conflict. Developed in the context of GROWLab, Weidmann's (2007) mobilization model examines this relationship in detail. It is shows that the settlement pattern of a group has an impact on group mobilization, because it determines how quickly mobilization can spread from a few extremists to the entire group population. All else being equal, groups with a settlement pattern that is favorable to the spread of mobilization should show a higher probability of conflict.

Weidmann proposes a simple geographical mobilization mechanism. Starting with few mobilized individuals located in the major cities of a group, mobilization spreads by means of individuals traveling between cities. The model uses real-world geographic data of group territories and cities. This way, it is possible to artificially mobilize an ethnic group according to the proposed mechanism, and measure the "difficulty" of mobilization of that group. This measure is then used as a predictor of conflict in a statistical analysis.

More generally, future modeling based on GROWLab depends critically on further progress as regards the rendering of dynamic conflict processes. GeoSim relies on an abstract combat model which can be applied to both interstate or internal warfare. However, a new dynamic model is needed in order to represent empirically realistic processes featuring explicit choices by specific actors. Such modeling could draw directly on rational-choice theory, though in a bounded sense.

The introduction of a more realistic core model will enable us to go beyond analysis of conflict onset, which has so far been the primary analytical goal in our statistical work (see Cederman and Girardin 2007; Buhaug, Cederman and Rød 2007). Indeed, the quantitative civil-war literature has so far explored separate dependent variables, such as onset and duration. Such research could analyze the shape of internal conflict as a dynamic process with explicit spatial and demographic extension. Our goal would be to construct an integrated explanation that covers several dimensions of civil violence, including onset, duration, incidence, spatial extension, and casualty levels (on the latter dimension, cf. Cederman 2003; Johnson et al. 2006).

Future computational research could open up another important area of modeling that has so far been relatively understudied in the quantitative and rational-choice literatures, namely that pertaining to complex actor structures. Instead of assuming ethnic groups to be unitary, as we have done so far in our statistical work, GROWLab enables us to represent deeper actor hierarchies including leader-follower dynamics inside such collective entities. It seems reasonable to assume that many civil wars have lasted for a very long time, because group elites have different interests from the masses, especially in those cases where belligerent leaders are able to cling to power by fueling diversionary conflict. As illustrated by Weidmann's recent research, it would be helpful to break up the unitary group structure in an attempt to trace mobilization processes without assuming any automatic solution to collective-choice dilemmas (see also Cederman 1997, Chaps. 7,8).

Ethnic federalism constitutes an especially interesting case of social systems in which conflict has been predicted to follow from multi-level interactions involving ethnically distinct regional sub-units endowed with considerable resources. The spatially explicit modeling framework of GROWLab, supported by geo-coded data from GREG and GREG-II, could serve as ideal test-benches for systematic exploration of detailed causal mechanisms of this type. In this sense, agent-based modeling promises to overcome some of the endogeneity complications marring more conventional quantitative work on ethnic federalism (Hug 2003).

While the goal of increasing the realism of geopolitical simulation provided the original motivation of GROWLab, we are very much aware that fundamental limitations are inherent to such an approach. Some of these difficulties are well known in the literature. As Robert Axelrod (1997) and Steve Bankes (1993) have pointed out, there is an overhanging risk that large, detailed simulation systems grow into unwieldy projects that are so complex that they make causal inference impossible. Successive and careful increase of complexity, as well as micro-level validation promise to reduce these risks. However, after two years of development of GROWLab, we have also discovered an equally troubling dilemma, namely that related to the tradeoff between realism and endogeneity: The more realistic the model, the more closely it has to be linked to real-world data, thus reducing the room for endogeneity. For example, if we decide to freeze the borders of states and ethnic groups, we have to dispense with any attempt to model boundary-formation processes. However, there can be no doubt that such processes are at the heart of geopolitical conflict (Wimmer and Min 2006). In fact, the original GeoSim system was developed precisely to capture such endogenous dynamics, especially those related to nationalism and state formation (Cederman 1997; 2002).

We therefore envisage a research agenda featuring different levels of realism and endogeneity (see also Lustick and Miodownik 2007 for a similar argument). Whereas some research based directly on GROWLab will include large amounts of spatially explicit data, other models will be based on abstract, parametrically created polities that serve as more flexible laboratories for the exploration of macro-historical transformations. The latter type of models can still profit indirectly from research of the former type by relying on empirical calibration of key parameters and by importing versions of the realistic causal mechanisms developed in such settings. Fortunately, GROWLab is flexible enough to allow for modeling in both directions. Rather than arguing for or against model simplicity in the abstract, the specific research question will ultimately have to determine the appropriate level of realism and endogeneity in future applied computational models.

References:

Axelrod, Robert. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Axelrod, Robert. 1997. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton: Princeton University Press.

Bankes, Steve. 1993. Exploratory Modeling for Policy Analysis. Operations Research 41 (3):435-449.

Bhavnani, Ravi. 2006. Ethnic Norms and Interethnic Violence: Accounting for Mass Participation in the Rwandan Genocide *Journal of Peace Research* 43 (6): 651-669.

Bremer, Stuart A., and Michael Mihalka. 1977. Machiavelli in Machina: Or Politics Among Hexagons. In *Problems of World Modeling*, edited by K. W. Deutsch. Boston: Ballinger.

Brown, Daniel G. et al. 2005. Spatial Process and Data Models: Toward Integration of Agent-Based Models and GIS. *Journal of Geographic Systems* 7: 1-23.

Buhaug, Halvard, Lars-Erik Cederman and Jan Ketil Rød. 2007. Disaggregating Ethno-Nationalist Civil Wars: A Dyadic Test of Exclusion Theory. Unpublished paper, PRIO, ETH Zurich, NTNU Trondheim.

Cederman, Lars-Erik. 1997. *Emergent Actors in World Politics: How States and Nations Develop and Dissolve*. Princeton: Princeton University Press.

Cederman, Lars-Erik. 2001a. Agent Based Modeling in Political Science. *The Political Methodologis*t 10 (Fall, 2001): 16-22.

Cederman, Lars-Erik. 2001b. Modeling the Democratic Peace as a Kantian Selection Process. *Journal of Conflict Resolution* 45 (August):470-502.

Cederman, Lars-Erik. 2002. Endogenizing Geopolitical Boundaries with Agent-Based Modeling. *Proceedings of the National Academy* 99 (suppl. 3):7796-7303.

Cederman, Lars-Erik. 2003. Modeling the Size of Wars: From Billiard Balls to Sandpiles. *American Political Science Review* 97 (February):135-150.

Cederman, Lars-Erik. Forthcoming. Articulating the Geo-Cultural Logic of Nationalist Insurgency. In *Order, Conflict, and Violence,* eds. Stathis Kalyvas and Ian Shapiro. Cambridge: Cambridge University Press.

Cederman, Lars-Erik, and Luc Girardin. 2007. Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies. *American Political Science Review* 101: 173-185.

Cederman, Lars-Erik , Luc Girardin, Andreas Wimmer. 2006. Getting Ethnicity Right: An Expert Survey on Power Distributions. Paper presented at the Annual Convention of the American Political Science Association, August 30-September 3.

Cederman, Lars-Erik, and Kristian Skrede Gleditsch. 2004. Conquest and Regime Change: An Evolutionary Model of the Spread of Democracy and Peace. *International Studies Quarterly* 48 (3):603-629.

Cederman, Lars-Erik, Jan Ketil Rød and Nils Weidmann. 2007. Geo-Referencing of Ethnic Groups: Creating a New Dataset. Paper presented at the 48th ISA Convention in Chicago, February 28 - March 3.

Collier, N, T Howe, and M North. 2003. Onward and upward: the transition to Repast 2.0. Paper read at *Proceedings of the First Annual NAACSOS Conference*, at Pittsburgh, PA

Cusack, Thomas R., and Richard Stoll. 1990. *Exploring Realpolitik: Probing International Relations Theory with Computer Simulation*. Boulder: Lynnie Rienner.

Epstein, Joshua M. 2002. Modeling Civil Violence: An Agent-Based Computational Approach. *Proceedings of the National Academy of Sciences* 99 suppl. 3:7243-7250.

Fearon, James D. 2003. Ethnic and Cultural Diversity by Country. *Journal of Economic Growth* 8:195-222.

Fearon, James D., and David D. Laitin. 2003. Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97 (February): 75–90.

Gleditsch, Kristian S., and Michael D. Ward. 2001. Measuring Space: A Minimum Distance Database. *Journal of Peace Research* 38:749-68.

Holland, John H. 1995. *Hidden Order: How Adaptation Builds Complexity*. Reading, Mass.: Addison-Wesley.

Hug, Simon. 2003. Selection Bias in Comparative Research. The Case of Incomplete Datasets. *Political Analysis* 11: 225-274

Johnson, Neil et al. 2006. Universal Patterns Underlying Ongoing Wars and Terrorism. Online archive physics/0605035.

Lustick, Ian S. 2000. Agent-Based Modelling of Collective Identity: Testing Constructivist Theory. *Journal of Artificial Societies and Social Simulation* 3 (1).

Lustick, Ian S. 2002. PS-I: A User-Friendly Agent-Based Modeling Platform for Testing Theories of Political Identity and Political Stability. *Journal of Artificial Societies and Social Simulation* 5 (3).

Lustick, Ian S. and Dan Miodownik 2007. Mechanisms, Ensembles, and Virtualizations: Simplicity and Complexity in Agent-Based Modeling. Paper prepared for presentation at the Annual Meeting of the American Political Science Association, Chicago.

Minar, Nelson, Rogert Burkhart, Chris Langton, and Manor Askenazi. 1996. The Swarm Simulation System: A Toolkit for Building Multi-Agent Simulations. Working Papers 96-06-042, Santa Fe Institute.

Nordhaus, William D. 2006. Geography and Macroeconomics: New Data and New Findings, *Proceedings of the National Academy of Sciences* (US), March 7, 2006, vol. 103, no. 10, 3510-3517.

Sambanis, Nicholas. 2004. Using Case Studies to Expand Economic Models of Civil War. *Perspectives on Politics* 2: 259-279.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior.* New York: W. W. Norton.

Weidmann, Nils B. 2007. From Sparks to Prairie Fires: Spatial Mechanisms of Group Mobilization. Paper prepared for presentation at the Annual Meeting of the American Political Science Association.

Wimmer, Andreas, and Brian Min. 2006. From Empire to Nation-State: Explaining Wars in the Modern World, 1816-2001. *American Sociological Review* 71 (6):867-897.