Promises and Pitfalls in the Spatial Prediction of Ethnic Violence

A Comment*

NILS B. WEIDMANN
Princeton University, USA

MONICA DUFFY TOFT
Harvard University, USA

Despite increasing technical sophistication, the quantitative literature has made little progress in forecasting ethnic violence. Nevertheless, recent efforts in predicting the location of ethnic violence from the spatial ethnic distribution seem to be a major step forward. In 2007, Lim, Metzler, and Bar-Yam proposed an agent-based model that takes as input the ethnic map of a country and derives from it the predicted locations of ethnic violence. The model rests on the assumption that spatial group clusters of a certain critical size are most likely to display ethnic violence. Their model achieves a remarkable level of agreement between predicted and observed locations of violence. Our article scrutinizes this exercise. We show that their analysis suffers from a biased selection of groups and regions, and that the null hypothesis and unit of analysis are inadequate. The proclaimed usefulness of the model for predicting violence in new cases is made difficult by the fact that the model does not generalize from one case to another. We conclude that the model provides little advance on prior research.

KEY WORDS: computational modeling; ethnic settlement patterns; ethnic violence; geographic information systems; spatial prediction

Introduction

In September 2007, Science magazine published an article entitled “Global Pattern Formation and Ethnic/Cultural Violence” (Lim, Metzler, and Bar-Yam, 2007, LMB henceforth). Essentially, the paper advances the argument that inter-group violence occurs at “boundaries between regions that are not sufficiently well defined” (p. 1540) and tests this using a geo-referenced computational model. According to

*We are grateful to Kristian Skrede Gleditsch and Ted Carmichael for comments on an earlier version of this manuscript.
the logic of the model, agents migrate to areas where other agents of the same group live, thus giving rise to the formation of homogenous clusters—a mechanism that has been proposed by Schelling in his famous model of neighborhood segregation (Schelling, 1971). The authors reason that during the process of migration, violence should occur in areas where groups have a certain characteristic cluster size.

Although they constitute valuable contributions to the literature, many computational modeling attempts in the social sciences use models mainly as heuristic tools and do not conduct a thorough empirical validation (Bhavnani, 2006; Lustick et al., 2004; Cederman, 1997). The LMB study is an exception. The authors apply their model to two cases—Yugoslavia and India—using pixelized representations of ethnic maps as input for the model. With each pixel representing an agent, the migration model is run on ethnic maps, and a wavelet filter is used to identify clusters of a critical size. LMB then compare whether those regions predicted by the model as violence prone match regions of observed violence. The degree of correspondence between the predicted and reported locations of violence is high; the authors report correlations of 0.9 and above.

This level of agreement between the model’s predictions and the observed pattern of violence should strike any social scientist as high. A model that predicts as well as the LMB model clearly seems to be a major step forward, not least because the great majority of similar efforts in conflict research have had only limited success so far (Beck et al., 2000; Ward et al., 2010). Indeed, as the authors claim, it is even possible to derive precise recommendations from the model as to how population separation can be employed to prevent conflict: “regions of width less than 10 km or greater than 100 km may provide sufficient mixing or isolation to reduce the chance of violence” (p. 1544). However, what is the simple, yet so powerful, mechanism behind the application of violence that conflict researchers have failed to identify? Led by our curiosity, we investigated further and examined the LMB model and its empirical validation more closely.

This paper reports our efforts and the results we obtained. We identify a number of issues that challenge the results of the LMB paper and its conclusions. Our critical assessment of the LMB paper has a number of implications. Scientifically, our replication and interpretation of the LMB results do not lend support to their simple model of migration and cluster formation as a trigger of violence. We do not argue that the underlying logic of the model is wrong—all we show is that for a convincing empirical test, one would have to (1) revisit the selection of groups and regions, which is problematic in different ways; (2) validate the model on instances of violence that correspond to the postulated mechanism of localized inter-group violence; and (3) resort to an adequate null hypothesis and unit of analysis. More importantly, however, our critique has practical implications. Given our results, we advise caution against implementing the territorial solutions to ethnic violence recommended by the authors.

In order to detail our criticisms against the LMB model, this paper proceeds as follows. The next section summarizes the LMB approach: the theoretical foundations, the model for detecting sites of potential violence, and the empirical evaluation of the model predictions. We then scrutinize this approach more closely, focusing on the data used for the empirical evaluation, the method for measuring the empirical fit of the LMB approach, and the unit of analysis. The conclusion from our re-analysis
of LMB is twofold. First, the high degree of correspondence of the LMB model rests on a problematic empirical approach, and second, despite the major flaws in this approach, there is no evidence of a universal critical patch size that determines where ethnic violence occurs.

The Promises

Ethnic violence has received extensive treatment in the recent literature, and scholarly opinions diverge widely not only regarding the mechanisms that link ethnicity to conflict, but more fundamentally, if ethnicity matters at all (Fearon and Laitin, 2003; Cederman and Girardin, 2007). Most of this literature attempts to explain whether and how the ethnic makeup of a country determines its propensity of being affected by civil war (Toft, 2003). The great majority of works in political science operate at the level of groups and posit that the groups’ interaction in the context of a state under certain conditions leads to violence. Examples of this type of reasoning are Horowitz’s “ranked” vs. “unranked” system of groups (Horowitz, 1985), the intrastate security dilemma between groups (Posen, 1993), or Fearon’s inter-ethnic commitment problem (Fearon, 1998). Even though the mechanisms that lead to conflict differ, what many theories about ethnic conflict have in common is their focus on groups as the primary actors.

LMB offer a fundamentally different perspective. “Rather than as a result of inherent conflicts between the groups themselves”, they argue, “violence arises due to the structure of boundaries between groups” (p. 1541). In other words, what causes the eruption of violence is the spatial placement of group members (or clusters of group members) relative to each other, with violence being more likely if the groups are poorly separated. This approach shifts our attention from the group as the primary actors to a more local level. Here, smaller settlement clusters of groups compete against each other, seemingly detached from the entire group as such. Rather than resulting from ethnic cleavages at the state level, in the LMB model violence originates from cleavages between groups in certain places.

LMB employ a computational agent-based model to test their proposed local mechanism of violence. Like many other agent-based simulations, the LMB model operates on a rasterized, two-dimensional space of “agents”. Each agent has a particular ethnic identity, which in the empirical test is assigned based on real-world data on ethnic spatial distributions (more on this below). It is assumed that similar to Schelling’s (1971) model, agents are satisfied if close to their kin. The model therefore implements simple migration by letting agents move to more convenient locations in their proximity. During this process, the size of group patches increases, creating homogenous ethnic population clusters—in essence, the segregation effect that Schelling’s model aims to explain.

LMB develop a predictor for the location of ethnic violence based on the size of these patches. They assume an inverted U-shape relationship between patch size and violence. In other words, violence occurs in regions where groups have an intermediate patch size, and is unlikely in regions with small or large group patches. The reason for this assumption, according to LMB, is that if groups settle in highly mixed settlement patterns (corresponding to small homogenous patches), collective identities
are weak and groups are not large enough to claim public spaces. However, as patch size increases, this gives rise to an “overlap of cultural domains” (p. 1542) and ultimately, violence. This overlap should not occur in regions with large, homogenous clusters, which are again assumed to remain peaceful.

The question is then how to measure the size of patches. The authors use a “Mexican hat” wavelet filter, a procedure that is used to detect a specified pattern of variation in an underlying signal or in this case, two-dimensional field of agents. A Mexican hat wavelet filter is essentially a circular template of a specified diameter that detects patches of the given diameter in the ethnic map. The template diameter is a model parameter that needs to be fixed before the filter is run. When applied to a particular location, the filter outputs a correspondence score where higher values indicate a better match. This way, for a given diameter of the filter, the authors obtain degrees of correspondence for all locations on a map that indicate the degree to which a group patch of the specified size is present at the respective location.

The empirical fit of the model is then determined by comparing the predicted locations of violence as obtained from the wavelet filter to the locations of actual ethnic violence, information about which was collected by the authors from newspaper reports and Internet sources.

To assess the correspondence between predicted and actual violence, two steps are necessary. First, the continuous scores obtained from the wavelet filter need to be transformed into discrete conflict/peace predictions. This is done by applying a second model parameter, the classification threshold, to the wavelet scores and classifying each location as conflict prediction if the wavelet score exceeds the specified threshold. Second, the correspondence between these predicted locations and the observed ones must be measured. Obviously, we do not expect violence to occur exactly at the predicted locations, but rather somewhere in their proximity. A measure for this imperfect correspondence can be obtained using proximity maps. A proximity map is a map that for each location indicates the distance to the closest conflict event. An example of such a map is shown in Figure 1, with two conflict events placed at arbitrary locations. The shading indicates the distance to the closest event, with darker colors corresponding to larger distances.

The fit between the model predictions and the observed events is now computed as the correlation between two proximity maps, one based on the predicted conflict events, and the other computed on the observed conflict events. This procedure essentially captures whether locations that are close to a predicted event are also close to an observed one. Note that for the computation of the predictions, the model parameters (number of time steps the model is run, wavelet diameter for patch detection, and cutoff for conflict classification) were selected such that the model fit was maximized. Still, the absolute value of this correlation needs to be compared against a null hypothesis in order find out whether it is significant. LMB do this by correlating the proximity map of the model classification with different realizations of a randomized violence map, where the observed number

\footnote{Apart from the correlation coefficient, there are other measures for the match between the two proximity maps. However, we restrict our attention to the former, since it is the one primarily used by LMB.}
of violence events is distributed uniformly across the entire study region. In other words, this procedure gives us the expected agreement between the model and the (random) locations of violence, under the null hypothesis that violence is equally likely at any location in the country.

In summary, then, the LMB analysis works as follows. For each case studied (Yugoslavia and India), a rasterized group map is created based on census data about group populations in a country, such that the initial agent population roughly corresponds to the ethnic map of the respective country. From this initial stage, the model is run for different parameter configurations, and the best fit obtained during this process is reported. This way, the authors achieve correlations of about 0.9 for the Yugoslavia example, and 0.998 for India (p. 1543), all of which differ significantly from the correlation coefficients they achieve under the null hypothesis.

What can the model tell us? There are two main areas where the LMB model could potentially contribute to the current literature on ethnic geography and violence. First, theoretically, it would be interesting to examine local violence-generating mechanisms more closely. Maybe we can learn much more about the emergence of violence from the study of small population clusters, as opposed to the existing group-level theories in political science. For example, LMB emphasize the role of migration in this context, as this is how the (potentially critical) patches form. Migration has frequently been associated with conflict (see e.g. Weiner, 1978), and further insights would constitute a great step forward.

Figure 1. Sample Proximity Map
The map is based on two arbitrary conflict locations. The shading indicates the distance to the closest event (darker colors correspond to longer distances).
A second potential contribution of the LMB study is at the practical level. On one hand, the model could be a tool to identify regions at risk of being affected by group violence. If their propositions are correct, the ethnic map of a country would contain sufficient information to predict where violence might arise. On the other hand, the LMB study entails the question of whether geographic adjustments to the ethnic map could serve as a precautionary measure to reduce the risk of violence. This question has attracted a lot of scholarly attention, following a recommendation to separate ethnic populations into clearly demarcated enclaves (Kaufmann, 1996).

Without a doubt, these questions are interesting and highly relevant, and the LMB article seems to tell us the answers. At the theoretical level, LMB show that the local conflict mechanism triggered by agent migration achieves surprisingly high correspondence with observed conflict events. At the practical level, the policy implication of their model is that population separation into clusters of a certain size might be sufficient to prevent conflict (LMB: 1544). However, these conclusions might be premature. In taking a closer look at the LMB approach, the purpose of the next section is to alert the reader to a number of issues that challenge their findings.

The Pitfalls

In this section, we scrutinize the steps taken by LMB to validate their model. We first start with the research design and case selection and the transformation of census data into the format required for the model. We then proceed by examining the coding of observed incidents of violence and how they are compared to the model predictions. Finally, we consider the output we get from the model, and what we can learn from it.

Selection of Regions and Groups

In order to find out whether fuzzy ethnic boundaries trigger violence, we need to select a study region to test the model on. Without discussion of their selection strategy, LMB identify two countries, Yugoslavia and India—countries with a history of inter-group violence. Even more difficult to understand, however, is the selection of groups for the two cases. Consider the Yugoslavia case. In LMB’s Figure 2A, the authors show the ethnic distribution of ten ethnic groups in Yugoslavia in 1991, which, according to the figure caption, was used as input to the model. Without discussion of the selection criteria, the analysis then drops six of these groups and proceeds with only four: Serbs, Croats, Muslims, and Albanians (LMB, Figure 2B). Not only do peaceful countries, but also peaceful groups in a country with ethnic violence disappear from the analysis. Why? One possible explanation is that this selection is due to limited data availability in the Yugoslav census (Petrovic, 1992), which was used by LMB to assign pixels to groups corresponding to their proportion in a given region. However, this is clearly not the case. The republics of Macedonia and Slovenia are not included, despite (1) still being members of the Socialist Federal Republic of Yugoslavia in 1991 and (2) the availability of population data for these republics.

Other deviations from the group categories in the official census are striking. For example, the model does not include Montenegrins (see Figure 2A in the original
paper and Petrovic, 1992)—instead, they appear to be replaced by Serbs. One could certainly make the argument that Montenegrins are ethnically similar to Serbs, as they are both predominantly members of the Orthodox Church and speak the same language. Clearly, if LMB’s group definition differs from the one employed in the census, we need to know what their definition is, and how the census data have been adapted to conform to it. A similar example is the Vojvodina region in Northern Serbia. Even though the census reports only a 57% Serb population in Vojvodina (Petrovic, 1992: 17), the model integrates a homogenous Serb population in this region. And even more puzzling, the inclusion of Albanians also deviates from the census: Albanians are the majority in the southernmost municipality of Montenegro (even visible in LMB’s Figure 2A), but this settlement was dropped in the model. Given that the model does treat the Albanians as a separate group, why are some Albanian settlements dropped?

Consideration of the group selection criteria is vital to the analysis. The decision to include or exclude groups could have a considerable effect on the predictive accuracy of the model. By construction of the model, predicted locations of conflict appear at the boundaries between large group clusters, for example at the border between Serbia and Kosovo, or between Serbia and Bosnia (see Figure 2D in LMB). The decision to include or exclude certain groups or group settlements therefore partly determines where these predicted conflict regions arise. For example, treating Montenegrins as a separate group from Serbs would have produced predicted locations of conflict along the Montenegrin–Serbian border between Serbs and Montenegrins. Similarly, including Hungarians in Vojvodina would have resulted in conflict predictions in Northern Serbia. According to the authors’ data, there was no violence reported close to the Montenegrin–Serbian border or in Northern Serbia, so these coding decisions would have decreased the model’s agreement with observed incidents of violence.

Absent a thorough explanation of how groups and their locations are selected, one is left with the impression that theoretical criteria did not guide the case selection. Rather, groups and categories seem to have been selected to conform with the model’s predictions, such that peaceful groups (Macedonians, Slovenians, Montenegrins, Hungarians) were dropped. This is of course highly problematic for a number of reasons. First, the incorporation of posterior information about group violence renders any prediction exercise invalid. Second, these exclusions raise the question of how to adapt the ethnic map to a smaller set of groups than there is in reality. As can be seen from Figure 2B-D in LMB, Macedonians and Slovenians were dropped from the map, resulting in a smaller study region. The procedure was different in case of Montenegrins and Hungarians, which were simply replaced by Serbs. This approach creates homogeneity in regions where there is none, thus artificially generating evidence in support of the model, as homogeneity coincides with no violence (see also the section on the null hypothesis below).

**Creation of the Model Input from Census Data**

Having identified a set of countries and ethnic groups on which to validate the model, the next step is to transform the ethnic maps into the raster format that serves as input to the model. Again, a closer inspection reveals some difficulties. We first describe briefly the data source used by LMB as input for their model: Petrovic (1992) summarizes the ethnic composition of the Yugoslav Federation according to population figures from
the 1991 census. More precisely, besides the aggregate ethnic composition of the federal units, the article also reports the percentages of the major ethnic groups (Serbs, Croats, Muslims, and Yugoslavs) at the municipality level. According to LMB (p. 1543), the authors created rasterized maps by assigning “areas of pixilated geographic maps pixel by pixel to ethnic groups at random, but in proportion to their relative population census in their region”. In other words, in a region with 70% Serbs and 30% Croats there would be a homogenous distribution of 70% Serb and 30% Croat agents throughout the region, which obviously fails to account for spatial variation within the region itself. This is not a problem as long as we examine small regions (e.g. municipalities), but this approach clearly becomes problematic as regions increase in size, since we fail to incorporate intra-regional variation.

The difficulty with the LMB approach is that it remains unclear how these regions themselves are defined and there seem to be some serious inconsistencies. Consider, for example, the Muslim population in Southern Serbia, at the border with Montenegro and Kosovo (green patch in LMB Figure 2A). This patch shows up in the LMB model. If the data for Serbia had been constructed using data at the level of federal units, there would have been no variation within Serbia and thus no Muslim patch as the one identified. Therefore, it seems that municipality-level data must have been used to create the ethnic map. However, this stands in contrast to the observation that the remaining parts of Serbia are essentially homogenous, despite the census reporting considerable variation in the Serb population (e.g. Central and Northern Serbia, 87.3% compared to 57.3%, see Petrovic, 1992: 17). The same applies to Montenegro, where the Muslim proportions vary from low percentages to about 80% in the census, a variation that is absent in the LMB map. For example, Muslims make up 50% of the municipality of Plav in Eastern Montenegro (also visible on LMB’s Figure 2A). The LMB model has no Muslims in Plav, rather it appears as populated entirely by Serbs.

These observations suggest that there is no uniform level of detail across the study region as regards the measurement of the group distribution. Again, it seems that ethnic diversity was properly incorporated where violence occurred, but deliberately left out of the picture in peaceful regions. This of course raises similar concerns as we pointed out in the previous section.

**Coding of Violent Events**

Once the model has been run on the rasterized ethnic maps, we need to compare the predicted locations of violence to observed ones. According to LMB, events of violence between ethnic groups in Yugoslavia were coded by collecting newspaper and internet reports on these events. As can be seen from the supplementary material given by LMB, many of the reports mention only the municipality where violence took place. By using the major city of the municipality as spatial reference point (LMB supplementary material: 21), the reports are converted to points on the map that denote the episode of violence. These points are then used to create the map of proximity to observed locations required for the spatial analysis.

---

2 A more detailed screenshot of the model is available at http://www.necsi.edu/research/ethnicviolence/sci317/yugofigure.html.
Whereas LMB acknowledge that their data collection may be biased and incomplete, this is difficult to tell since there are no clear coding rules provided. What constitutes an event of local violence that the model supposedly explains, and what criteria have to be satisfied for the event to be included in their event list? How was the observation period chosen? In the case of Bosnia, for example, the reader is left to wonder why repeated attacks against the Hungarian minority in Vojvodina seem not to qualify as ethnic violence. To be sure, the levels of violence in this region were lower than in those included in the LMB list. Still, this type of violent inter-ethnic behavior is consistent with the LMB model’s focus, so we need to know why these events fail to be included.

The Indian case represents a different problem in terms of the type of violence that is theorized about and the reported violent episodes that are coded for analysis. The data source used by LMB is a report to the Indian government about incidents of terrorism in the states of India. Leaving aside possible problems of reporting bias in governmental data, we note that the report includes casualties that are actually victims of (mostly cross-border) terrorist attacks. Given that the focus of the model is on local violence between ethnic groups, and LMB explicitly exclude cross-border violence (LMB supplementary material: 6), there is a significant mismatch between what these numbers are claimed to mean and their real content. At the same time, the dataset’s focus on terrorist activity causes it to miss the cases of real local violence present in India. For example, Gujarat, a district in the west, is one of the most violence-prone districts in India, having suffered terrible riots for decades—most recently in 2002 between Hindus and Muslims that resulted in at least 1000 deaths. Yet, it fails to rank as one of the top nine for violence between 1999 and 2002 (Figure 3E), and is essentially considered a peaceful location according to LMB. A more general comparison of the numbers presented by LMB to other datasets on local violence in India (Wilkinson, 2004: 60) reveals that their measure of violence has little in common with more systematic data collection efforts.

**Null Hypothesis**

Having coded the observed locations of violence, LMB proceed to compare their model predictions to the observed ones. As described above, the approach proposed by LMB is to create three kinds of proximity maps for the study region: first, a reference map that gives for each pixel the distance to the closest observed conflict location; second, a similar map, but based on the predicted locations; and third, a set of proximity maps based on randomly distributed violence locations that serve as the baseline for comparison. Comparing the correlation between the predicted and the reference map on one hand, and between the randomized maps and the reference map on the other hand gives us an idea of how well the model performs as compared to the assumption that violence is equally likely in the study region.

---


4 Available online at http://164.100.47.4/rsquestion/ShowQn.aspx?qno=65007.

We argue that this null hypothesis—the uniform distribution of ethnic violence—is mistaken. Ethnic violence as postulated by LMB requires at least two ethnic groups as opponents, so by definition, ethnically homogenous regions cannot experience violence of the kind that LMB try to explain. By construction of the wavelet filter used in the LMB paper, high scores of predicted violence can only occur in regions with at least two groups—the same regions in which we would expect ethnic violence a priori. Without comparing the model against a null hypothesis that excludes homogenous regions, there is a considerable risk that the model only identifies mixed regions, and not the fuzzy boundaries inside these regions that supposedly trigger violence.

This criticism is most obvious in the case of India, which according to maps 3A and B in LMB has significant minority populations only in its Northern and Eastern regions. The fact that model predictions match reported conflict locations is therefore predetermined, because in no other place in India are there significant populations of more than one group. However, this criticism also applies to the case of Yugoslavia, where Serbia and Montenegro—according to LMB—appear as largely homogenous. In order to find out whether the high correlation reported in the model is indeed driven by the inclusion of homogenous regions (or regions mistakenly coded as homogenous, see above), we replicated the LMB approach for the Yugoslavia case for a different null hypothesis. After obtaining the dataset of the predicted violence scores from the authors, we geo-referenced this dataset in order to be able to combine it with administrative boundary data. This enables us to vary the inclusion of separate units in the entire study region. We also create a dataset containing the locations of reported violence, as given in the supplementary material of the original article.

We then replicate the findings of the original evaluation. Since we obtained the dataset with pre-computed predictions, we cannot modify the parameters used for generating them (that is, the size of the wavelet filter, and the simulation time $t$). However, we vary the cutoff threshold $h$ for positive classifications within the optimal region (0.3–0.6, in increments of 0.01) as given by LMB. In line with the original findings, our results indicate correlations of 0.85 and above throughout the range of threshold values (compare to LMB’s Figure S4.2 in the supplementary material). All these outcomes are significantly different from the distribution of 1000 simulated maps with the locations of violence distributed randomly across the study region. The best fit we achieve (correlation of 0.884 at $h = 0.38$) comes very close to the best fit reported by the authors (0.89), but due to the geo-referencing procedure and the associated re-projection of the data, we do not expect perfect correspondence.

Having successfully replicated the original findings, we then address our concern about including homogenous regions in the computation of the baseline maps. We therefore repeat the model evaluation, but using random comparison maps with violent events placed only in the mixed regions of Croatia, Bosnia, and Kosovo, and not in Montenegro and Serbia. The full extent of the study region, however, remains unchanged.

The results are striking and can best be displayed graphically. Figure 2 shows two density curves. The solid curve denotes the distribution of correlations between the

---

*Replication data are available from http://nils.weidmann.ws/publications/weidmann10promises.*
reference map and 1000 proximity maps based on randomized violence locations distributed across the entire study region, which is the approach employed in the LMB paper. The grey area shows the level of agreement between the reference map and the predicted maps, for different values of the threshold \( h \) for positive classifications. Clearly, the grey area is far away from the correlation we would observe under random distribution of events in the entire study region, which is exactly what LMB find. However, once we replace the implausible null hypothesis by a more reasonable one (as we argued above), the difference is reduced dramatically. If we allow random locations of violence to occur only in Croatia, Bosnia, and Kosovo, we obtain correlations as given by the dashed line in Figure 2, and the difference between predicted and randomized locations for any cutoff \( h \) is no longer significant at the 0.1% level. Still, the plot shows that the correlations obtained from the model are higher than in the random maps; among the 30 values for \( h \) that we tested, less

**Figure 2. Correlations**

Correlations between the proximity map based on randomized locations of violence when distributing these events across the entire study region (solid line), or only in ethnically mixed regions (dashed line). The grey area indicates the correlations achieved by a proximity map based on the predicted locations.
than one-third (9) are outside the 99% interval, suggesting that the model achieves higher than random predictive accuracy. Hence, we do not dispute that even with this alternative null hypothesis, we can tune the model so that it is significant at some level. The point we want to make is that the result we obtained conveys a radically different picture as compared to the original, highly significant results. Therefore, our confidence in the performance of the model decreases considerably.

Our interpretation of this result is twofold. First, as we suspected, much of the seemingly high predictive capability of the model is driven by the model’s ability to predict the absence of violence in homogenous regions, rather than the successful identification of fuzzy boundaries. As mentioned above, we expect this effect to be much stronger in the case of India, where ethnically diverse regions are small relative to an almost perfectly homogenous region. Second, since the randomized and the predicted correlations are essentially indistinguishable, the model is unable to predict where, in an ethnically mixed region like Bosnia, violence takes place. In other words, regardless of where in Croatia, Bosnia, or Kosovo we place the random events, the model’s empirical fit remains unchanged. This evidence seriously questions the model’s ability to accurately identify locations of violence on a finer scale, and thus challenges the main contribution of the paper, which is to explain localized violence (see also the section on out-of-sample prediction below).

**Unit of Analysis**

Our next consideration is the unit of analysis selected for the model. In order to create an input to the agent model, LMB rasterize the ethnic map at a high resolution (about 3km). Whereas this is certainly required for the model, there is no reason why we should conduct the empirical validation also at the rasterized level. The rasterization suggests a high precision, which is however misleading, as we do not have ethnic distribution data below the municipality level. Similarly, the accounts of ethnic violence listed in LMB frequently report the event as having occurred in a certain municipality or close to a major city, rather than giving a precise location. Therefore, the municipality level seems to be the right way to test the model predictions, because this is where both ethnic distribution and violence can be measured with reasonable confidence. Since we cannot disaggregate the data below this level, we need to aggregate the model predictions to municipalities. More detailed approaches (including the one taken by LMB) claiming finer empirical precision are impossible.

In order to find out if the conclusions of LMB hold when testing at an empirically adequate level, we reanalyze the Bosnia case. Municipality serves as our unit of analysis. In the previous section, we have argued that the presence of multiple groups—rather than the structure of boundaries between them—can explain where violence occurs. We address this point here by proposing the counter-hypothesis that ethnic diversity should be positively related to violence. If the incorporation of ethnic boundaries in the LMB model is of any explanatory value, it should be able to predict violence at the municipality level better than ethnic diversity. Note that in doing so, the LMB model has the advantage of factoring in larger settlement areas of groups, whereas our variable “ethnic diversity” does not draw on information from beyond municipality borders, and thus cannot identify larger patterns. We compute predicted violence scores
according to the LMB model by averaging the wavelet scores to the municipality level (see Figure 3). When comparing peaceful and violence-ridden municipalities with respect to this aggregated wavelet score as shown in Figure 4 (left), we find that they do not differ significantly. This suggests that the LMB model does not seem to be helpful in distinguishing violent units from peaceful ones. We conduct a similar test of our alternative explanation that ethnic diversity is related to violence by computing an indicator of ethno-linguistic fractionalization (Alesina et al., 2003) at the municipality level. In support of our expectations, we find that violence-ridden municipalities are significantly more diverse (Figure 4, right).

We conducted a more extensive test using spatial regression analysis (Ward and Gleditsch, 2008). Since our dependent variable “observed violence” is binary, we estimate logit regression models. The occurrence of violence might be spatially clustered, so we add a spatially lagged dependent variable, based on a row standardized adjacency matrix (Ward and Gleditsch, 2008). Furthermore, larger units might have a higher exposure to violence, so we add the logged size of the unit (measured in the number of raster cells it covers) as a second control variable.

Model 1 tests the impact of predicted violence as a single independent variable. Model 2 uses instead a spatial lag of the predicted violence in order to test whether high predictions of violence in the neighborhood of a municipality can explain its actual occurrence. Model 3 includes ethnic diversity alone, and in Model 4 we compare how the

---

7 The ELF indicator is normally computed following the Herfindahl concentration formula, applied to the country-level shares of ethnic groups. Here, we follow the same approach and apply the concentration formula to the municipality shares of the groups. The resulting indicator ranges between 0 and 1, with higher values corresponding to more diverse units.
predicted level of violence performs against our purely local measure of ethnic diversity. Finally, Model 5 adds the spatially lagged dependent variable to the previous model.

Our analysis provides no empirical support for the assumption that the predicted level of violence can explain the actual occurrence of violence in our sample of 109 Bosnian municipalities. In none of the models does the predicted violence variable receive a positive significant sign, nor does its spatial lag. However, local ethnic diversity at the municipality level seems to be able to explain the occurrence of ethnic violence: it is positively and significantly related to violence in all models. As expected, the observations of violence are spatially clustered, so the spatially lagged dependent variable receives a positive and significant sign.

We explored different robustness checks for the above presented regression models. First, we modified the way the violence predictions were aggregated to the municipality level. When using the maximum predicted violence level, the results do not change substantially. With the municipality median violence as an independent variable, the results remain unchanged, except for Model 4 where the LMB predicted violence variable receives a positive sign and is significant at the 10% level. Second, we varied the spatial weights matrix used to compute the spatial lags: we tested an unstandardized adjacency matrix, and a matrix based on inverse centroid distances (measured in an equidistant conic projection of the municipalities dataset). In none of these models did we receive a positive and significant effect of predicted violence.

**Figure 4.** Kernel Density Estimates
Comparing municipalities that experienced ethnic violence (N = 24) to those that did not (N = 85). Left: Kernel density estimates for municipalities with violence (solid line) and peaceful municipalities (dashed line) across the range of predicted violence levels. The two types of units do not differ with respect to the predicted violence level (t-value 0.01, p-value 0.99). Right: Kernel density estimates for municipalities with violence (solid line) and peaceful municipalities (dashed line) across the range of ethnic diversity, as measured by ELF. Municipalities with reported ethnic violence are significantly more ethnically diverse (t-value −4.93, p-value = 0.00).
Migration and the Computational Model

We now turn to the interpretation of the results from the LMB exercise, in particular the segregation process that leads to the formation of population clusters. As we have motivated above, the question about the role of migration is interesting from both theoretical and practical points of view. What does the model tell us about migration? Whereas the article starts by introducing the Schelling-like migration mechanism, the further discussion largely omits these dynamic aspects. A look at Figure 4.1 (p. 24) in LMB reveals that the position of the predicted locations is largely invariant for different lengths of the simulation (0 to 100 time steps). The authors interpret this as robustness of the model with respect to the number of time steps. Upon closer inspection, however, we suggest a different answer. 100 time steps are simply not sufficient for any significant migration to occur. The reason is that at each time step, there is only one agent selected at random for migration (see description in LMB supplementary material: 3). In a population of more than 20,000 agents, the time period evaluated in LMB’s robustness check thus allows only every 200th agent to move! Therefore, we cannot tell whether the model is robust to variation in simulation time and thus migration, because there essentially is no such variation. Consequently, the reader is left to wonder why a model that is introduced at great length at the beginning of the LMB article does not receive a more systematic treatment.

Using the LMB Model for Prediction

Our last point relates to the possible application of the model for identifying critical locations that have a high risk of conflict because of their ethnic geography. Obviously, before a given model can be applied to provide risk assessments in new cases, we need to be reasonably sure that the effects captured by the model are of some generality. In other words, we need to test whether the model performs well
only on the set of cases on which it was built, or if its insights transfer from the initial sample to other cases. The obvious way to do this is out-of-sample prediction: we fit a model on a case (or set of cases) and test its predictive performance on (one or more) cases that were not used for fitting. Out-of-sample prediction would have been a straightforward exercise to perform for the LMB model. One could determine the model parameters that maximize the fit in the Yugoslav case, and measure how well these parameters fare in the case of India, or vice versa. Yet, no such attempt is made.

We conduct a simple validation exercise by using again the model predictions obtained for the whole of Yugoslavia, as described above. Rather than evaluating these predictions on the same study area, we take the subset of predictions for Bosnia, and conduct our evaluation (i.e. the creation of the observed and random proximity maps) for this subset only. This is not an out-of-sample validation, but it can tell us how the model predicts conflict in regions other than the one it was fitted on. Still, this is a comparatively simple prediction exercise, as the Bosnia data were part of the sample used for fitting the model, so we expect a good predictive performance here. Our results show the contrary. The best correlation between the model predictions and the observed locations is 0.211 (for cutoff \( h = 0.23 \)); at the same time, the correlations obtained under a random distribution of the Bosnian events vary from –0.210 to 0.413 (95% interval over 1000 iterations), so the LMB model predictions are indistinguishable from a random distribution of events. To be sure, it might well be possible to fit the model only on the Bosnia case such that it achieves a high level of agreement. However, here we are interested in whether the model captures some general pattern across cases, and the results show that this does not seem to be the case. Parameter settings that allow the model to identify violent locations across Yugoslavia \((h = 0.38)\) do not help us find these locations if applied to Bosnia only, where an \( h \) parameter of 0.28 maximizes the fit.

As another example for this, compare the patch size obtained in the Yugoslav case (30–60 km) to the patch size in the India case (100 km). Even though LMB state that “the range of filter diameter values for which good agreement is obtained overlaps that of the former Yugoslavia” (p. 1543), there is no mention of the other parameter values for which the overlap was obtained. For example, is there a similar overlap in the cutoff values \( h \)? In fact, the authors explain the difference by the “larger granularity of data” (p. 1543) in the India case. Indeed, there exists a fundamental mismatch between the geographic scale used in both cases: whereas in Yugoslavia the data describe the ethnic composition of municipalities (average size 400 km\(^2\)), in the case of India we deal with states. Even the smallest of these states covers an area more than 25 times as large as a municipality (Tripura, 10,486 km\(^2\)). Leaving aside the question of how to pinpoint a location of violence within these large areas, the mere fact that the data granularity partly determines the critical patch size should worry us considerably.

Obviously, parameters maximizing the fit in one case perform poorly in other cases. This essentially reduces the entire LMB approach to a simple model fitting exercise. All we learn from the results of the model is that it is possible to tune the model to achieve a high agreement, whereas in fact we would be interested in whether the identified patterns are of some generality. Hence, any prediction attempt to identify critical locations in previously unobserved countries will not work.
Conclusions

As we have introduced above, there are two areas where LMB (2007) could contribute to current research on ethnic violence. First, the article presents a local mechanism of violence generation that originates from small population clusters and thus offers a competing theory of ethnic violence that challenges many existing approaches. Second, it provides a tool that could potentially allow us to make risk assessments based on the ethnic map of a country, and at the same time suggests population separation as a possible strategy to contain violence. Both theoretically and practically, there is much promise in the LMB approach. As our critical examination has shown, however, the study fails to live up to it.

The empirical test of the proposed local violence mechanism is problematic in many ways. We show that there are major issues with the case selection that seems to adjust the group map as to better fit the model predictions; that the coding of violence in the case of India has little to do with the type of violence the model theorizes about; we also have reservations against the null hypothesis that assumes violence to be equally likely in a country; and in our re-analysis of Bosnia at the municipality level we find no evidence of a significant effect to the LMB model predictions. We do not claim that there is no mechanism of violence between overlapping cultural domains as proposed by LMB. However, as there is no evidence in support of the proposed mechanism at this point, we prefer to assume that it does not exist until this evidence is presented.

At the same time, this conclusion renders the policy implications to be drawn from the LMB article invalid. Kaufmann’s territorial solution to ethnic conflict has received little empirical support so far (Toft, 2003; Laitin, 2004; Sambanis, 2000), and our analysis shows that this is also true for the LMB strategy. Why should we accept a recommendation about separating populations into “regions of width less than 10 km or greater than 100 km” (p. 1544), if these values (1) have been obtained from a study of two cases out of many ethnic conflicts worldwide; (2) seem to depend on the respective resolution of the ethnic maps used; and (3) even in this limited setting, do not prove to generalize from one case to the other? Not only is such a recommendation premature, it is also dangerous. We hope that our article will be able to draw some attention to the pitfalls in the LMB approach and its questionable conclusions.

References


NILS B. WEIDMANN is a post-doctoral researcher at the Woodrow Wilson School of Public and International Affairs, Princeton University. His research interests include the study of violent conflict at different scales, with an emphasis on geography.

MONICA DUFFY TOFT is Associate Professor of Public Policy at Harvard’s Kennedy School of Government and author of The Geography of Ethnic Violence (Princeton University Press, 2003) and Securing the Peace (Princeton University Press, 2010).