



ESSAY

Predicting armed conflict: Time to adjust our expectations?

Lars-Erik Cederman^{1*} and Nils B. Weidmann^{2*}

This Essay provides an introduction to the general challenges of predicting political violence, particularly compared with predicting other types of events (such as earthquakes). What is possible? What is less realistic? We aim to debunk myths about predicting violence, as well as to illustrate the substantial progress in this field.

If “big data” can help us find the right partner, optimize the choice of hotel rooms, and solve many other problems in everyday life, it should also be able to save lives by predicting future outbreaks of deadly conflict (1). This is the hope of many researchers who apply machine learning techniques to new, vast data sets extracted from the Internet and other sources. Given the suffering and instability that political violence still inflicts on

the world, this vision is conflict researchers’ ultimate frontier in terms of policy impact and social control.

Despite this promise, however, prediction remains highly controversial in academic conflict research. Relatively few conflict experts have attempted explicit forecasting of conflicts. Furthermore, no system of early warning has established itself as a reliable tool for policy-making, although major efforts are currently under way (2).

Recent years have seen the emergence of a series of articles that attempt to address this void by leveraging the latest advances in large-scale data collection and computational analysis. The task in these studies is to predict whether interna-

tional or internal conflict is likely to occur in a given country and year, thus creating yearly “risk maps” for violent conflict around the world. The first prediction models were based on the emerging quantitative methodology in political science at the time and relied on simple linear-regression models.

However, it was soon recognized that these models cannot capture the varying effects and complex interactions of conflict predictors. This realization led to the introduction of machine learning techniques such as neural networks (3), an analytical trend that continues to the present day. In these models, the interactions of risk factors generating violent outcomes are inductively inferred from the data, and this process typically requires highly complex models. Today, country-level analyses with resolution at the level of a year still constitute the majority of the work on conflict prediction, with some studies having pushed the time horizon of their predictions several decades into the future (4).

More recently, newly available data and improved models have allowed conflict researchers to disentangle the temporal and spatial dynamics of political violence. Some of this research produces monthly or daily forecasts. Such temporal disaggregation requires adaptations of existing prediction models. For example, the approach presented in (5) is based on conflict event data for the Israel-Palestine conflict. Using a model that

¹ETH Zurich, Zurich, Switzerland. ²University of Konstanz, Konstanz, Germany.

*Corresponding author. Email: lcederman@ethz.ch (L.-E.C.); nils.weidmann@uni-konstanz.de (N.B.W.)



The costs of an inability to predict political violence

distinguishes between high- and low-intensity conflict, the analysis generates predictions for the year 2010 based on data from 1996 to 2009. Alternative approaches have aimed to leverage new kinds of predictors, such as war-related news reports (6). Due to their ability to capture political tension at a much higher temporal resolution, these reports prove to be stronger predictors of war onset than conventional structural variables (such as the level of democracy).

Other studies have attempted to explore sub-national variation in violence, trying to predict not only when but also where conflict is likely to break out. Spatial disaggregation allows conflict predictions to be produced for administrative units, such as districts or municipalities, or arbitrary grid-based locations. Existing work in this area has focused on specific countries and conflicts. Weidmann and Ward (7), for example, generate predictions at the municipality level for the civil war in Bosnia, as illustrated in Fig. 1. Similar forecasts have been made for violence across spatial grid cells in Africa (8). Again, the complexity of spatial prediction models can vary, ranging from spatial regression models (7) to more flexible yet complex machine learning models (8).

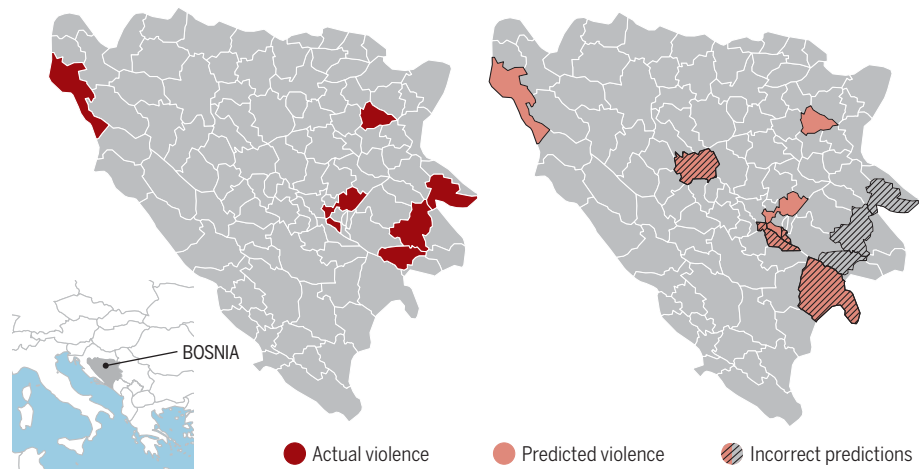


Fig. 1. Prediction of civil war violence at the municipality level in Bosnia. (Left) Actual occurrence of violence (dark red) in seven municipalities in June 1995. (Right) Predicted violence (light red) according to the spatial-temporal model described in (7). The striped pattern highlights incorrect predictions. Although conflict in four municipalities was forecasted correctly, the model missed three actual outbreaks and falsely predicted violence in four municipalities. As is often the case in conflict prediction, many areas remain peaceful and are predicted as such (shown in gray).

Promise and pitfalls of prediction

It is clear that considerable advances have been made in the area of conflict prediction. Using clear and objective statistical criteria, newer approaches attain higher levels of out-of-sample accuracy than conventional, explanatory models. In contrast to causal explanation of past instances of violence, out-of-sample forecasting enables the prediction of events that were not used for fitting the model. Researchers relying on advanced quantitative techniques have also scored specific forecasting successes. For example, in a report commissioned by the Political Instability Task Force, Ward and his team were able to forecast the military coup in Thailand 1 month before its actual occurrence on 7 May 2014 (9).

Moreover, some progress has been made in dealing with the challenge posed by the prediction of rare events. Standard, off-the-shelf machine learning models are typically applied to problems in which the different outcomes are relatively balanced. This is not the case for predictions of violence and peace, in which the units examined are peaceful most of the time. This problem can be addressed by different resampling techniques, which result in a much higher overall predictive accuracy of the model. Muchlinski *et al.* (10) applied such techniques to the problem of predicting civil war out of sample from 2001 to 2014. Their model predicted 9 out of 20 civil wars correctly, whereas conventional regression models predicted none.

The literature has also established that a focus on out-of-sample prediction helps guard against the inclusion of long lists of explanatory factors that may worsen predictive performance (11). More generally, such analysis also serves as a useful reminder that causal explanation of past events and prediction of future ones are two distinct, though related, standards of empirical performance (12).

Despite this progress, however, it would be overly optimistic to say that the dream scenario of life-saving conflict prevention has become im-

minently realizable. Moreover, the field is still far from the policy impact that pollsters and economic forecasters enjoy. Why is this so?

Perhaps the most pernicious problem pertains to the common failure to fully appreciate the fundamental complexity surrounding processes of peace and conflict. As opposed to relatively structured institutional decision-making settings, such as voting and consumer behavior at the micro level, conflict processes typically encompass an unwieldy set of actors interacting in surprising and, by definition, rule-breaking ways (13). Such situations are characterized by fundamental and inherent complexity that allows for “pattern prediction” (14) rather than precise empirical forecasting of specific events. In the absence of full knowledge of how all theoretical components interact and sufficient data to measure the relevant variables, all that can be hoped for is risk assessment on the basis of structural features that increase the probability of conflict. Thus, at least at the macro level, it is futile to pin one’s hopes of future predictive performance on extrapolations from previous successes in much less complex areas, such as billiards, planetary movements, or traffic systems (1), or for that matter in simpler political settings such as electoral competition (13), where both the theoretical principles are well known and events of interest occur with high frequency.

Though machine learning techniques, such as neural networks, are able to capture nonlinearities in underlying data, geopolitical changes altering the very units of analysis, such as states and their borders, pose a much more fundamental challenge, especially to long-term macro predictions (15). Most macro models tend to trace properties of a given set of existing states into the future while ignoring the possibility of territorial change, such as secession and unification. Yet, as illustrated by the changes brought about by the end of the Cold War in the former Soviet Union

and Yugoslavia, country-level data on these states offer little guidance for prediction after the end of the Cold War. Beyond territorial changes, these implicit assumptions of constancy also apply more generally to interaction between units and the effect of causal mechanisms. This problem haunts the use of “cross-validation,” which divides the data set into parts, some of which are used to “train” the forecasting algorithm before testing it on the remaining “holdout” parts (12). In cases where such a practice cuts historical sequences into pieces, valuable information on long-term trends will be lost because this approach mixes historical periods as if they were equivalent.

Data quality further impedes progress in the prediction of political violence. Unlike the movement of billiard balls or planetary trajectories, measuring the onset, location, and timing of conflict is much more difficult and is associated with considerable uncertainty. Similar issues exist for many of the determinants of violence, such as economic conditions (16). Even if measurement error is not a problem for statistical explanation of past events, it constitutes a challenge for the prediction of future violence and typically reduces the confidence that violence occurs at the predicted place and time. More pernicious types of errors occur if the measurement of violence is systematically related to one or more predictor variables. Because political violence is often coded from secondary sources such as news articles, a high level of observed violence can be due to a high level of actual violence or a higher probability of reporting (or both) (17). This makes prediction difficult. If anything, scaling up the size of the data set—as in several projects that made use of automated event coding—is likely to exacerbate this problem, due to reliance on the same secondary sources (18).

Although recent advancements in prediction are promising, we caution against a tendency to overvalue its importance for both theory and policy. As argued above, out-of-sample forecasts can contribute to theory building, but this does not imply that valid explanations must always be predictive. As illustrated by Darwinian theory, some highly path-dependent processes allow for only post-hoc explanations of specific cases. Given the complexity characterizing conflict processes, especially at the macro level, such explanations can still provide crucial information about the effectiveness of specific mechanisms and policies. Furthermore, it would be unwise to view predictive performance as the only valid standard of empirical assessment, particularly in cases where the predictive model is so complicated and opaque that it remains unclear what drives predictive success. For example, Bayesian averaging over model ensembles is an elegant inductive technique that pools large amounts of data from competing models, but unless disentangled in theoretical terms, the overall outcome may amount to little more than a theoretical black box.

There are also reasons to be cautious with regard to the policy relevance of predictive studies. Scholars producing forecasts typically assume that policy-makers want predictive risk assessments

more than anything else because this would allow them to reduce potential conflict through preventive resource allocation and intervention. However, these hopes presuppose that the effects of policy intervention are well known. In fact, theory-free prediction does little to guide intervention without knowledge about the drivers of conflict. Therefore, carefully executed policy analyses assessing the causal effectiveness of conflict-reducing measures are a prerequisite for politically effective macro forecasting. Given the difficulties of obtaining reliable information on key social indicators, especially in developing countries, basic description and explanatory modeling may, in many instances, be more urgently needed than forecasting.

Recommendations

There are a number of ways in which existing work on conflict prediction can be improved—for example, when it comes to the communication of methodology and results. In some cases, this calls for more user-friendly methods of presenting results, such as reporting existing and projected trends rather than merely ROC (receiver operating characteristic) curves based on fancy estimation techniques. Transparency also requires that crucial assumptions about sampling periods and uncertainty measures be stated explicitly and tested for robustness in scenarios based on alternative assumptions. Otherwise, researchers' error estimates may convey a false sense of certainty.

To assess the added value of new approaches, analysts need to do a better job comparing their forecasts from complex prediction machinery to simple baseline models. In its purest form, such a baseline model simply predicts no change from the past. For instance, Lim *et al.* (19) purported to predict the location of ethnic violence in the former Yugoslavia with a complex agent-based model. Although the model's predictive accuracy looks impressive at first blush, further scrutiny shows that this performance is very close to a model that places incidents of violence randomly on the map, except in Serbia and Montenegro (20).

Ultimately, the hope that big data will somehow yield valid forecasts through theory-free “brute force” is misplaced in the area of political violence. Automated data extraction algorithms, such as Web scraping and signal detection based on social media, may be able to pick up heightened political tension, but this does not mean that these algorithms are able to forecast low-probability conflict events with high temporal and spatial accuracy. Large, automatically coded data sets are helpful as long as researchers account for their limitations regarding the possible lack of data quality and representativeness. It is thus hardly surprising that human “superforecasters” working in teams are still able to beat not only more specialized experts when it comes to the prediction of political events in general, but also prediction markets and other automated methods (21).

Overall, we strongly believe that conflict prediction is useful and worth investing in. Yet, future forecasting research needs to recognize the inherent limitations imposed by massive historical complex-

ity and contingency in human systems. As illustrated by the end of the Cold War and more recent events, such as “Brexit” and Donald Trump's electoral triumph, historical “accidents” often make a mockery of decontextualized out-of-sample extrapolation. Discussing the difficulty of long-run forecasting of economic development, Milanovic (22) reminds us that “the number of variables that can and do change, the role of people in history (‘free will’), and the influence of wars and natural catastrophes are so great that even forecasts of broad tendencies made by the best minds of a generation are seldom correct.”

At the same time, however, forecasts with much more limited spatial and temporal scope—such as projected short-term trajectories of violence in a given city in an ongoing civil war—are perfectly possible, as they are less likely to be affected by these developments. Therefore, the challenge for the field is to find the right balance between the inherent complexity of the social and political world and the associated limitations on our ability to accurately forecast political violence. Within a limited spatiotemporal radius, policy-relevant prediction is feasible and potentially extremely useful, as illustrated by recent efforts to accelerate collection of disaggregated and spatially explicit data on conflict events (2). Beyond these limits, however, massive theoretical and empirical uncertainty tends to overwhelm the attempts to forecast. In such cases, predictive modeling may be more useful as a heuristic tool for generating possible scenarios rather than as a producer of specific policy advice.

REFERENCES AND NOTES

1. D. Helbing *et al.*, *J. Stat. Phys.* **158**, 735–781 (2015).
2. S. P. O'Brien, *Int. Stud. Rev.* **12**, 87–104 (2010).
3. N. Beck, G. King, L. Zeng, *Am. Polit. Sci. Rev.* **94**, 21–35 (2000).
4. H. Hegre *et al.*, *Environ. Res. Lett.* **11**, 054002 (2016).
5. P. T. Brandt, J. R. Freeman, P. A. Schrodt, *Confl. Manage. Peace Sci.* **28**, 41–64 (2011).
6. T. Chadefaux, *J. Peace Res.* **51**, 5–18 (2014).
7. N. B. Weidmann, M. D. Ward, *J. Conflict Resolut.* **54**, 883–901 (2010).
8. C. Perry, *Stab. Int. J. Secur. Dev.* **2**, 1–18 (2013).
9. Ward Lab, “The coup in Thailand and progress in forecasting.” Blog: <https://predictiveheuristics.com/2014/05/22/the-coup-in-thailand-and-progress-in-forecasting/>.
10. D. Muchlinski, D. Siroky, J. He, M. Kocher, *Polit. Anal.* **24**, 87–103 (2016).
11. M. D. Ward, B. D. Greenhill, K. Bakke, *J. Peace Res.* **47**, 363–375 (2010).
12. G. Shmueli, *Stat. Sci.* **25**, 289–310 (2010).
13. J. Ulfelder, *Foreign Policy*, 8 November 2012; <http://foreignpolicy.com/2012/11/08/why-the-world-cant-have-a-nate-silver/>.
14. F. A. Hayek, in *The Critical Approach to Science and Philosophy* (Univ. of Chicago Press, 1964), pp. 332–349.
15. L.-E. Cederman, *Emergent Actors in World Politics: How States and Nations Develop and Dissolve* (Princeton Univ. Press, 1997).
16. M. Jerven, *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It* (Cornell Univ. Press, 2013).
17. N. B. Weidmann, M. D. Toft, *Confl. Manage. Peace Sci.* **27**, 159–176 (2010).
18. W. Wang, R. Kennedy, D. Lazer, N. Ramakrishnan, *Science* **353**, 1502–1503 (2016).
19. M. Lim, R. Metzler, Y. Bar-Yam, *Science* **317**, 1540–1544 (2007).
20. N. B. Weidmann, M. D. Toft, *Confl. Manage. Peace Sci.* **27**, 159–176 (2010).
21. P. E. Tetlock, B. A. Mellers, J. P. Scoblic, *Science* **355**, 481–483 (2017).
22. B. Milanovic, in *Global Inequality: A New Approach for the Age of Globalization* (Harvard Univ. Press, 2016), p. 21.

ACKNOWLEDGMENTS

We thank H. Hegre, P. Hunziker, D. Lazer, and J. Wucherpfennig for comments on a draft of this paper and L. Kaweraw for research assistance. N.B.W. acknowledges financial support from the Alexander von Humboldt Foundation (Sofja Kovalevskaja Award).

10.1126/science.aal4483

Predicting armed conflict: Time to adjust our expectations?

Lars-Erik Cederman and Nils B. Weidmann

Science **355** (6324), 474-476.
DOI: 10.1126/science.aal4483

ARTICLE TOOLS

<http://science.sciencemag.org/content/355/6324/474>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/355/6324/468.full>
<http://science.sciencemag.org/content/sci/355/6324/470.full>
<http://science.sciencemag.org/content/sci/355/6324/477.full>
<http://science.sciencemag.org/content/sci/355/6324/481.full>
<http://science.sciencemag.org/content/sci/355/6324/483.full>
<http://science.sciencemag.org/content/sci/355/6324/486.full>
<http://science.sciencemag.org/content/sci/355/6324/489.full>
<http://science.sciencemag.org/content/sci/355/6324/515.full>

REFERENCES

This article cites 16 articles, 3 of which you can access for free
<http://science.sciencemag.org/content/355/6324/474#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)