Universität der Bundeswehr München

CISS MUNICH WORKING PAPER SERIES
WORKING PAPER 010/2026

Center for Intelligence & Security Studies

THE ROLE OF THEORY IN CONFLICT PREDICTION: COMPARING COMPETING EXPLANATIONS FOR COMMUNAL VIOLENCE WITH MACHINE LEARNING

Further publications: https://www.unibw.de/ciss/working-paper-series

ISSN 3053-7800

Summary

This paper discusses the role of theory in conflict predicting. After describing the motivation behind the shift from explanation to prediction in conflict research, we give a brief overview of how conflict prediction works. We use forecasts of communal violence in West Africa as an example to illustrate the shortcomings of a common approach to compare theories using models with different sets of explanatory variables. We show that differences between such "theory-based" models in conflict prediction only appear in artificially limited model specifications that are poor reflections of reality. In contrast, we highlight the potential role of interpretability methods in connecting prediction models with theories. In sum, theory plays a limited role in optimizing predictive performance but remains essential for interpreting predictions. Conflict forecasts are thus best understood as risk assessment tools whose theoretical insights must be contextualized and complemented by qualitative analysis.

Daniel Mittermaier & Vanessa Gottwick

Why do we study the causes of conflict? More specifically, shouldn't a scientific theory on the risk factors of conflict also be able to make predictions? Should we not base policy recommendations on studies that extrapolate well into the future, rather than only explaining the past? If their predictive power has not been evaluated, how else can researchers expect policymakers to act on their findings?[1]

This return to the essentials motivates a shift from explanation towards predictive modelling in quantitative conflict research. Over the last decade and a half, the prediction of violent conflict has thus grown from a rarity to an established field of research.

A brief overview of conflict prediction

The goal of any statistical modelling is to create a simplified but accurate representation of the world. This broadly encompasses connecting outcomes with one or multiple explanatory variables. One reason for the hesitation of early conflict research to make concrete predictions certainly lies in the inherent complexity of conflicts and their genesis. In one of the earliest attempts at prediction, conflict is fittingly described as "highly nonlinear, massively interactive, and heavily context dependent or contingent"[2], notoriously difficult to capture in statistical models. Yet, producing forecasts requires identifying generalizable patterns between explanatory variables and outcome, which hold true across different contexts and across time.

Enabled by a leap in available data, both on conflict itself and its presumed causes, and the removal of many computational barriers through technological progress, machine learning (ML) methods have become the de facto standard in the field. These are explicitly designed to capture non-linear effects and interactions[3]. Rather than making a-priori assumptions about the impact of potential explanatory variables (often called *features* in a ML context), ML models infer their relationships with the outcome (*target*), violent conflict, from the data. They are also more robust to

[1] Philip A Schrodt, "Seven Deadly Sins of Contemporary Quantitative Political Analysis," *Journal of Peace Research* 51, no. 2 (2014): 287–300; Michael D Ward et al., "The Perils of Policy by P-Value: Predicting Civil Conflicts," *Journal of Peace Research* 47, no. 4 (2010): 363–75; Nils B. Weidmann and Michael D. Ward, "Predicting Conflict in Space and Time," *Journal of Conflict Resolution* 54, no. 6 (2010): 883–901.

[2] Nathaniel Beck et al., "Improving Quantitative Studies of International Conflict: A Conjecture," *American Political Science Review* 94, no. 1 (2000): 22.

[3] Lars-Erik Cederman and Nils B. Weidmann, "Predicting Armed Conflict: Time to Adjust Our Expectations?," *Science* 355, no. 6324 (2017): 474–76.

intercorrelations among features than traditional regression models, a common challenge of working with real-world data.

For the *features* in conflict prediction, various data from a multitude of providers are used, mostly international organizations, government agencies, NGOs, or research institutions. They broadly encompass socio-economic information, demographics, the political system, geographical characteristics, climate data, and of course the history of violence. Data on conflict itself is collected at the event level based mostly on news reports and aggregated to the desired unit of analysis. In practice, selecting and harmonizing the data often represents the lion's share of work. The *target* – the definition of conflict – can e.g. be based on all armed violence or a specific type of violence, and predictions typically refer either to the risk of occurrence (*classification problem*) or the expected intensity (*regression problem*), for example measured as the number of fatalities.

Prediction models have been developed at many spatial and temporal resolutions. Most commonly, predictions are made at a yearly or monthly level with a relatively short time horizon of one to three years. Individual forecasts most often either cover whole countries – both for their obvious importance as political entities shaping their socio-economic and regulatory contexts and for their relevance as policy units to users of prediction systems – or are broken down to small grid-cells in a geographic raster independent of country borders.

To test the predictive power of a model without having to wait for the future to arrive, the data is split during development, with a small portion (e.g. the last year available) held back to evaluate a model's predictive performance on a known "future" (*out of sample testing*). The optimal model configuration is then developed based on the remaining data. This is often split up multiple times again with the model only seeing subsets each time, to avoid simply memorizing the data rather than learning generalizable patterns.

To evaluate the quality predictions, several different metrics are available, depending on the type of prediction problem. Using probabilities of conflict occurrence as an example, a simple approach would be calculating the percentage of correct predictions (*accuracy*) after converting the output to "yes" or "no" predictions based on a probability threshold. In practice, more complex metrics are usually chosen, which take the performance across the whole probability range into account. As any predictive system makes mistakes, it is important to define which mistakes can be tolerated and which should be avoided if possible and choose a metric that reflects this. To some extent this generally involves a trade-off: is it more important to recognize all potential instances of conflict at the cost of false alarms, or should the latter be avoided at the risk of missing some instances of conflict.

Despite the variety in data, conflict history has by far the highest predictive power in current models, especially in short term forecasts. Given the prevalence of ongoing and recurring conflict, this is most likely due to (localized) conflict dynamics being more determined by cycles of violence and temporal dependencies than by the structural factors captured by most other data. The strongest remaining indicator thus becomes the resulting violence itself[4], regardless of the specific cause. The most promising avenues of research to improve insights into the short-term processes determining violence lie in the analysis of unstructured text data continuously generated in high frequency in the form of (news) media[5] or internet content[6]. This is especially important, since

[4] Thomas Chadefaux and Thomas Schincariol, "Endogenous Conflict and the Limits of Predictive Optimization," *EPJ Data Science* 14, no. 1 (2025): 82.

[5] Hannes Mueller and Christopher Rauh, "The Hard Problem of Prediction for Conflict Prevention," *Journal of the European Economic Association* 20, no. 6 (2022): 2440–67.

[6] Christian Oswald, "I Still Haven't Found What I'm Looking for: Predicting Security-Related Incidents and Conflict Fatalities with Google Trends and Wikipedia Data," *Journal of Conflict Resolution* 70, no. 2-3 (2025): 499–524.

previous violence naturally cannot predict the rare but critical cases of newly emerging conflict.

As of today, conflict predictions exhibit significant uncertainty, caused by inherently random components of conflict, an ever-changing world, and imperfect or non-existent data on many potential risk factors[7]. This results in a trade-off for users: The higher the resolution of the predictions and thus more detailed the information, the higher the accompanying uncertainty around this information. While efforts to at least quantify this uncertainty are underway[8], in practice this means that conflict predictions are currently best understood and used as risk assessments rather than concrete predictions on what is going to happen.

Despite being a motivator for the resurgence in prediction, theory often plays only a limited role in the design of prediction models. While the occasional argument around theory-based prediction systems is raised[9], the integration of theory remains mostly centered on the pre-selection of potential features and algorithms, with the final selection and modelling architecture more determined by what works best. ML-based "tests" of theory consequently often focus on performance improvements through the inclusion of theory-specific variables over a base model. However, the comparison with base models artificially constrains a model's performance, making it easier to achieve such improvements.

In the following section, we demonstrate this effect through a model predicting communal violence in West Africa. Additionally, we highlight a different approach to think about the connection between theory and prediction using ex-post model interpretation techniques.

Connecting predictions with explanations: communal violence in West Africa

Communal violence is a recurring phenomenon, with the ongoing farmer-herder clashes in Nigeria perhaps the most prominent example. It is characterized as localized, mostly low-intensity "violent conflict between non-state groups that are organised along a shared communal identity"[10]. Existing explanations for communal violence can be broadly categorized into environmental factors and governance factors. The former centers around resource competition, potentially exacerbated by climate factors such as rainfall variability, droughts, and long-term climate-related changes in resource distributions[11]. The latter highlights the role of (local) institutions in shaping and moderating communal tensions, from economic and political favoritism over land use/ownership regulation and regulatory uncertainty to traditional local leadership[12]. It is important to note, however, that these are not necessarily competing, but potentially

[7] Daniel Mittermaier et al., "Forests of Uncertaint(r)Ees: Using Tree-Based Ensembles to Estimate Probability Distributions of Future Conflict," arXiv:2512.06210, preprint, arXiv, December 5, 2025.
[8] Håvard Hegre et al., "The 2023/24 VIEWS Prediction Challenge: Predicting the Number of Fatalities in Armed Conflict, with Uncertainty," *Journal of Peace Research* 62, no. 6 (2025): 2070–87.
[9] Robert A. Blair and Nicholas Sambanis, "Forecasting Civil Wars: Theory and Structure in an Age of 'Big Data' and Machine Learning," *Journal of Conflict Resolution* 64, no. 10 (2020): 1885–915; Andreas Beger et al., "Reassessing the Role of Theory and Machine Learning in Forecasting Civil Conflict," *Journal of Conflict Resolution* 65, nos. 7–8 (2021): 1405–26; Robert A. Blair and Nicholas Sambanis, "Is Theory Useful for Conflict Prediction? A Response to Beger, Morgan, and Ward," *Journal of Conflict Resolution* 65, nos. 7–8 (2021): 1427–53.
[10] J. Brosché and E. Elfversson, "Communal Conflict, Civil War, and the State: Complexities, Connections, and the Case of Sudan," *African Journal on Conflict Resolution* 12, no. 1 (2012): 33.
[11] e.g. Hanne Fjelde and Nina von Uexkull, "Climate Triggers: Rainfall Anomalies, Vulnerability and Communal Conflict in Sub-Saharan Africa," *Political Geography* 31, no. 7 (2012): 444–53; Jonas Nordkvelle et al., "Identifying the Effect of Climate Variability on Communal Conflict through Randomization," *Climatic Change* 141, no. 4 (2017): 627–39; Stefan Döring and Katariina Mustasilta, "Spatial Patterns of Communal Violence in Sub-Saharan Africa," *Journal of Peace Research* 61, no. 5 (2024): 858–73.
[12] e.g. Solveig Hillesund, "Choosing Whom to Target: Horizontal Inequality and the Risk of Civil and Communal Violence," *Journal of Conflict Resolution* 63, no. 2 (2019): 528–54; Alexander De Juan et al., "The Pacifying Effects of Local Religious Institutions: An Analysis of Communal Violence in Indonesia," *Political Research Quarterly* 68, no. 2 (2015): 211–24; Kristine Eck, "The Law of the Land: Communal Conflict and Legal Authority," *Journal of Peace Research* 51, no. 4 (2014): 441–54; Tor A Benjaminsen and Boubacar Ba, "Farmer–Herder Conflicts, Pastoral Marginalisation and Corruption: A Case Study from the Inland Niger Delta of Mali," *The Geographical Journal* 175, no. 1 (2009): 71–81.

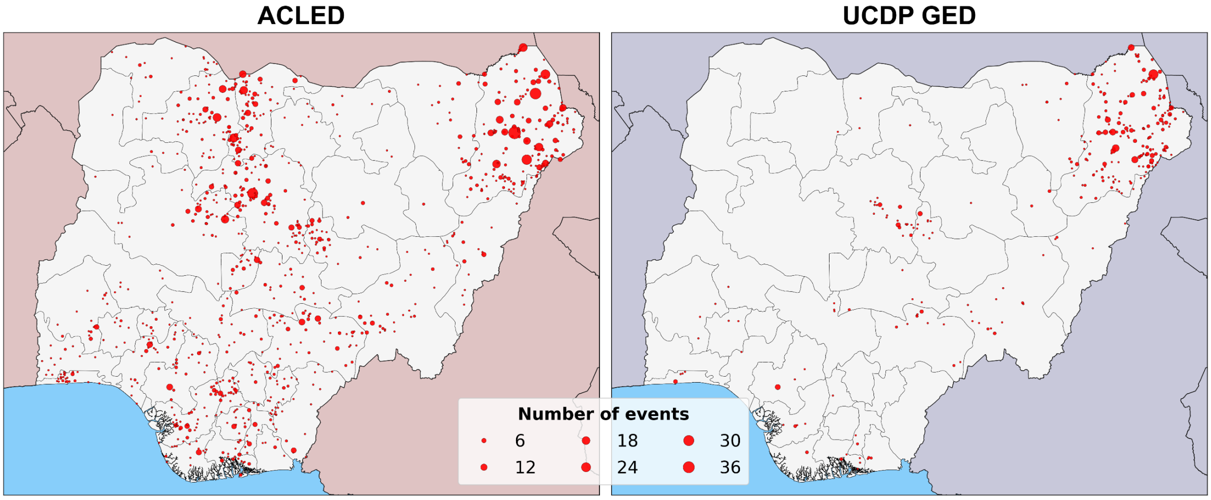

Figure 1: Comparison of armed violence event counts recorded in ACLED and UCDP GED for Nigeria in 2020. We exclude protests, riots and strategic developments from ACLED, which are event types not captured by UCDP GED.

interacting explanations for communal violence[13]. A ML-based approach is well suited to capture such interactions.

Modelling approach

Given the prominence of the farmer-herder conflicts, we focus on West Africa in this exercise. Prompted by the localized nature of communal violence and the cross-border migration patterns and identities of herders, we build our model not at the country-level, but on a subnational 0.5° x 0.5° grid, which corresponds to roughly 55 x 55 km at the equator[14]. Accounting for the delayed nature of many climate-related effects, such as water scarcity building up over time with a lack of precipitation, we chose a seasonal approach with a three-month temporal resolution. Our model for this exercise is developed on data from 2012-2019, with resulting predictions for 2020 used for the evaluation. However, data was selected so predictions could also be generated for the true future.

To measure violence at this spatial resolution, we have the choice between two conflict event datasets, UCDP GED[15] and ACLED[16]. While following the same general goal, they differ somewhat in their approach to measuring violence: UCDP GED focusses more on organized actors and has a minimum level of violence threshold for their inclusion, while ACLED casts a wider net. For some contexts this can lead to differences in the patterns of violence observed. Given the often low-intensity nature of communal violence and the potential lack of organized actors, ACLED seems the better choice for this application, which is confirmed when examining the data for Nigeria, the epicenter of the farmer-herder conflict: ACLED captures a lot more violence in central and North-East Nigeria than UCDP GED, which corresponds to the hotspot areas of communal violence in qualitative descriptions of the conflict (Figure 1). Accordingly, our target is based on ACLED and measures the occurrence of communal violence in a given grid cell and season. Our predictions therefore estimate the probability of this occurrence.

As alluded to above, we create several submodels, which all include specific groups of features:

[13] e.g. Kristina Petrova, “Floods, Communal Conflict and the Role of Local State Institutions in Sub-Saharan Africa,” *Political Geography* 92 (January 2022): 102511; Stefan Döring, “Come Rain, or Come Wells: How Access to Groundwater Affects Communal Violence,” *Political Geography* 76 (January 2020): 102073.
[14] Andreas Forø Tollefsen et al., “PRIO-GRID: A Unified Spatial Data Structure,” *Journal of Peace Research* 49, no. 2 (2012): 363–74.
[15] Ralph Sundberg and Erik Melander, “Introducing the UCDP Georeferenced Event Dataset,” *Journal of Peace Research* 50, no. 4 (2013): 523–32.
[16] Clionadh Raleigh et al., “Introducing ACLED: An Armed Conflict Location and Event Dataset,” *Journal of Peace Research* 47, no. 5 (2010): 651–60.

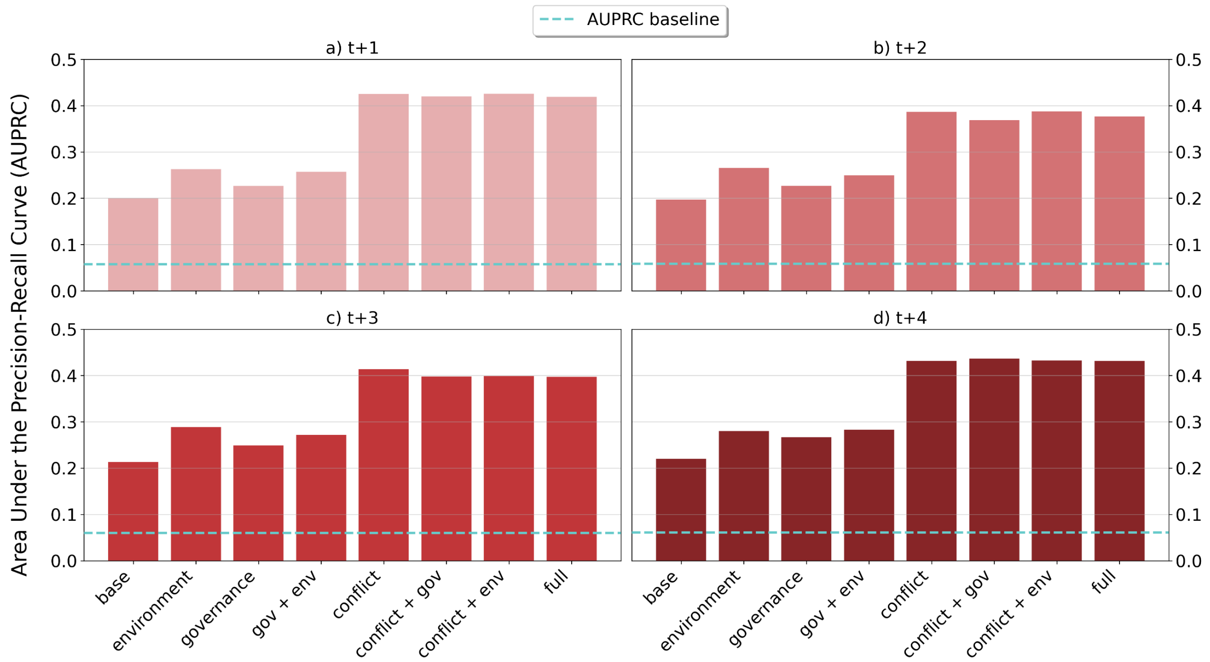

Figure 2: Comparison of performance evaluation according to the area under the precision-recall curve (AUPRC) for 2020 forecasts and different feature groups. The AUPRC baseline for random predictions corresponds to the share of grid cells where communal violence occurred in the given prediction window (around 6% in all four windows).

1. *Base* features consisting of population density and nighttime lights, the latter of which corresponds reasonably well with local wealth. These are included with all model specifications.
2. Features representing *environmental* explanations. These include vegetation indices, temperature and precipitation measures, soil moisture, agricultural production monitoring data, survey data on climate impacts and food insecurity, and several features derived from land cover classifications.
3. Features representing *governance* explanations. These are comprised of national governance and regime characteristics measures, localized survey data on trust in traditional leaders, (local) governments and group identities, data on ethnic groups, and a grid cell's remoteness (representing government reach).
4. Features constructed from conflict data capturing current, surrounding and historic *conflict*.

Predictive performance and its limits

Figure 2 shows the performance of different combinations of feature groups, measured as area under the precision-recall curve (AUPRC). The AUPRC is a score for probabilistic classification predictions, which evaluates the predictions across the whole range of probabilities by varying the threshold for conversion to binary predictions. Importantly, in doing so it only takes cases where violence was either predicted or violence occurred into account. We chose this metric as it is not influenced by the large number of zeroes in the data, with corresponding observations arguably not adding useful information on the quality of the models. Communal violence on the model's resolutions is only recorded in roughly 3% of observations, making no violence a default assumption without much insight in most cases.

Figure 2 shows the evaluation metrics for the models across the various feature sets. A simple comparison of performance without *conflict* shows both improvements for the *environment* and *governance* models over the baseline individually as well as better comparative performance for the *environment* model. In contrast, once *conflict* features are added to the mix, there is no real difference in performance regardless of the inclusion of any of the *environment* or *governance* features. In fact, the

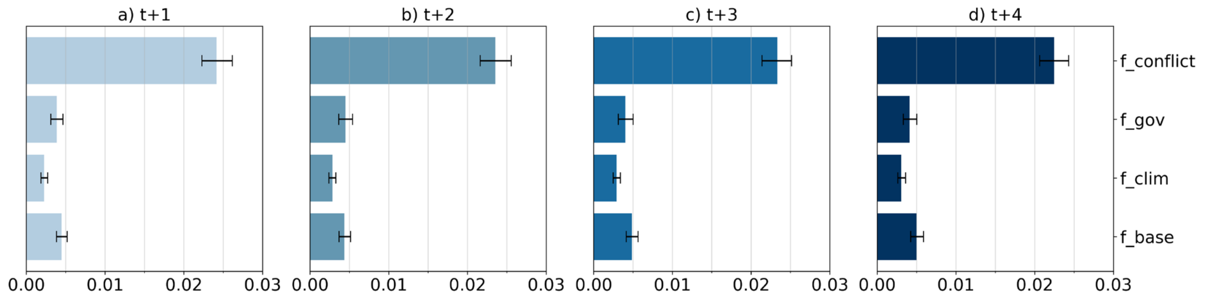

Figure 3: Mean absolute contribution to predicted probability per feature group and prediction window based on grouped SHAP values for 2020 predictions. Error bars represent bootstrapped 95% confidence intervals.

variance in performance from repeated model runs due to the random component of the Random Forests algorithm was often greater than the variance between the different models including *conflict* features.

Interpretation as an alternate approach

While the models without conflict demonstrate some predictive power, these results confirm that an approach based on performance comparison with varying features only shows differences in artificially limited models. While any model by necessity is only an approximation, its usefulness depends on how well it captures its subject. Simply ignoring the presence of conflict or other components results in a poor representation of the real world. Any insights gained from such limited models thus may not be trustworthy.

We therefore highlight an alternate approach: using model interpretation methods to break down the predictions into the impact of individual feature groups. We use SHAP (Shapley Additive exPlanations) values in this example, an approach rooted in game theory, as the additive character allows for simply summing the impact of individual features to calculate the joint impact of a feature group[17].

Figure 3 shows the grouped impact of the features in the full model based on SHAP values. In our model, both *environmental* and *governance* have some predictive power, with *governance* being significantly more impactful than the *environmental* group, at least in the short-term, making this the stronger of the two explanations. For *governance*, trust in the local government and trust in traditional leaders had the highest individual impacts on Q1 2020 predictions, while for climate, this was the share of agricultural areas lost over the past 5 years.

As SHAP values are calculated for each prediction individually, we can also use this approach to highlight the spatial impact of the different explanations. Figure 4 provides an example, showing where the observed *base*, *environmental*, *governance* and *conflict* features increase or decrease the predicted likelihood of communal violence in our model. This way, we can extract the spatial patterns of risk resulting from the different explanations in our model.

The role of theory in conflict prediction

Why then does the inclusion of these features not improve the forecasts compared to conflict-only models? As discussed above, this is only surprising at first glance, since the outcome of environmental and governance factors in terms of conflict is already measured through past conflict itself. On the contrary, one should be careful whenever strong increases in predictive power through the inclusion of new features are reported, as the marginal differences between the model specifications with conflict features highlight. Rather than actual gains this might reflect a model

[17] Scott M. Lundberg and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Red Hook, NY, USA), NIPS'17, December 4, 2017, 4768–77.

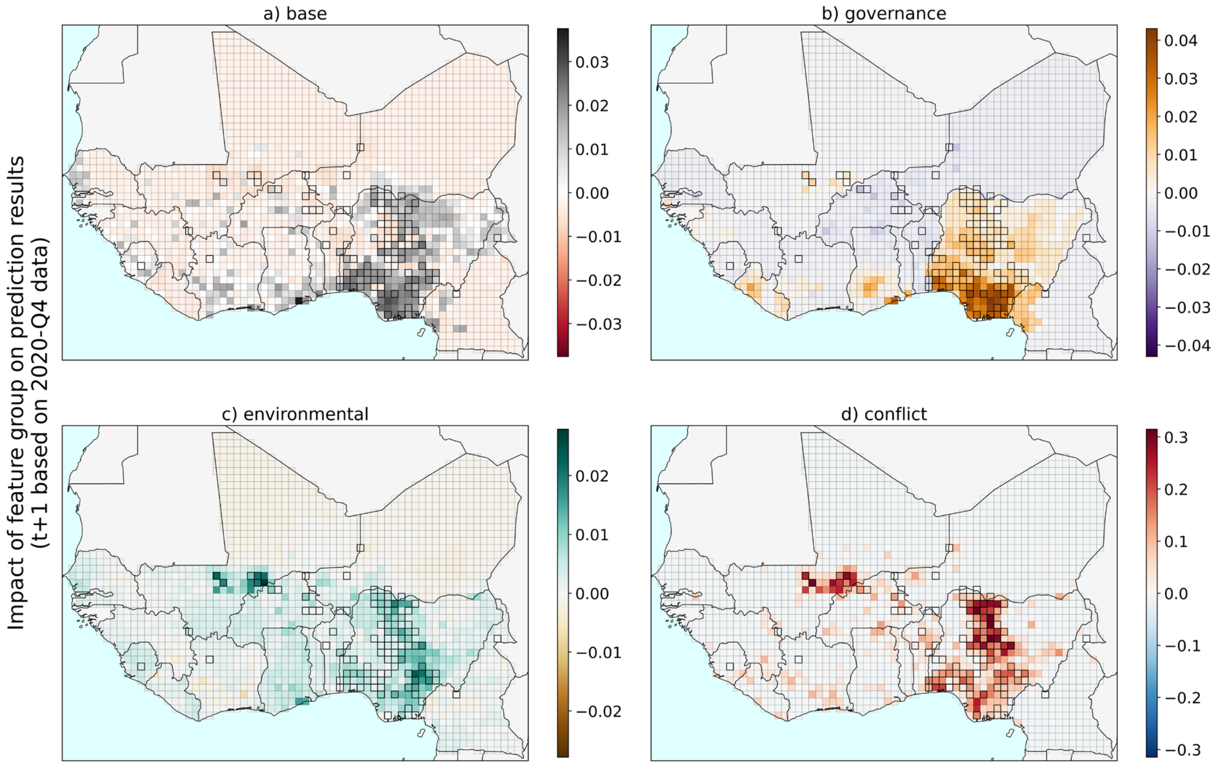

Figure 4: Spatial distribution of aggregated impact on predictions per feature group, based on SHAP values for Q1 2021 predictions. Bold squares reflect observations where communal violence occurred.

specification not fully optimized on predictive power and thus a flawed representation of reality. The usefulness of explanatory theories for conflict predictions thus depends on the use case: If the goal is simply to create the best risk score, the reasons for conflict and violence are likely less important than the best possible data on conflict itself.

If the goal is to evaluate theories, however, our approach shows how we can still learn from the inclusion of additional features and their impact on the model. Their inclusion transfers some of the predictive power from observed conflict back to its root causes, allowing us to judge them in the model's context. Additionally, where conflict newly erupts without prior history, models only trained on conflict itself are bound to be blind as mentioned above, while broader models are potentially able to pick up on warning signs.

However, any insight into theory that can be gained from forecasting efforts comes with some caveats: as is the case with all models, resulting insights can only be as accurate as the model's reflection of the real world. Accordingly, such explanations are only true in the context of a model's world view. As this inherently means a simplified view of the world, models need to be carefully evaluated based on their outcomes, to understand their capabilities and shortcomings. While using quantitative models as risk assessment tools can in fact provided valuable insights and highlight risk in unexpected cases, their warnings should thus always be complemented by qualitative analyses or other "sanity checks" in practice.

Daniel Mittermaier is a research associate at the Center for Early Warning (CCEW) at the University of the Bundeswehr Munich. His recent research has focused on the integration of uncertainty in prediction models and its pitfalls in evaluation, as well as the quantification of risks in the climate-conflict nexus.

Vanessa Gottwick, PhD, is a program manager at IABG and former head of the department on conflict research at the CCEW.