WORLD RESOURCES INSTITUTE

wps
Water, Peace and Security

TECHNICAL NOTE

LEVERAGING WATER DATA IN A MACHINE LEARNING–BASED MODEL FOR FORECASTING VIOLENT CONFLICT

SAMANTHA KUZMA, PETER KERINS, ELIZABETH SACCOCCIA, CAYLA WHITESIDE, HANNES ROOS, AND CHARLES ICELAND

ABSTRACT

We present a methodology to forecast conflict (defined as organized violence resulting in at least 10 fatalities over a 12-month period) up to a year in advance using a random forest model. When applied to test data, the model captures 86 percent of future conflicts. The model's conflict signal is noisy, with half of conflict predictions representing false positives. We also explore whether water-related indicators are useful predictors of conflict. Water-related variables are assessed to be correlated with conflict outcomes, but not empirically significant for model decision-making. However, adjusting the definition of conflict, such as by lowering the fatality threshold or examining only emerging conflict, increases the significance of water variables. A web-based tool that houses the model allows users to explore forecasts and indicators spatially and through time, providing additional information on underlying vulnerabilities as a first step toward enabling timely, effective water-related interventions to mitigate conflict and/or build peace.

BACKGROUND

Is Water a Contributor to Conflict?

Political instability and conflicts are rarely caused by a single factor. More commonly, violent conflict emerges from a confluence of circumstances and actions. Any serious attempt to address instability must consider as wide a range of potential contributing factors as possible. Yet development projects in conflict-prone areas have often neglected the relevance of water resources (Swain 2015; Conca and Wallace 2009; Machlis and Hanson 2008).

CONTENTS

Technical notes document the research or analytical methodology underpinning a publication, interactive application, or tool.

Suggested Citation: Kuzma, S., P. Kerins, E. Saccoccia, C. Whiteside, H. Roos, and C. Iceland. 2020. "Leveraging Water Data in a Machine Learning–Based Model for Forecasting Violent Conflict." Technical Note. Washington, DC: World Resources Institute. Available online at: www.wri.org/publication/leveraging-water-data.

While water-related risks are rarely the cause of conflict, researchers agree that they should be seen as a contributing factor (Gleick and Iceland 2018; Swain 2015). Water serves as a "threat multiplier" to conflict through a number of pathways, including diminished water supply, increased demand, extreme floods, and poor management (Gleick and Iceland 2018; Swain 2015). Where conflict has only recently ended, poor water management can negate investments made toward recovery (Swain 2015).

Though water-related problems are not new, the compounding threats of climate change and a booming population have brought a new sense of urgency to addressing them (Gleick and Iceland 2018). Organizations like the United Nations, the World Economic Forum, and the High Level Panel on Water have called on the international community to investigate the interlinkages between water, conflict, and human security. For example, on October 26, 2018, the Kingdom of the Netherlands organized an Arria-formula meeting of the UN Security Council to explore ways for the UN system to address water scarcity as a contributor of conflict (UN Web TV 2018).

A complicating factor is that there is no consensus on the degree to which water plays a role in conflict. Even with the most frequently touted examples of water-driven conflict, such as the Syrian civil war, experts debate whether water should be considered a causal driver (Selby et al. 2017). The truth is that the causal relationship between water and conflict is nebulous. Yet it is not necessary for experts to untangle this web in order for thoughtful, effective interventions to be possible (ECC Factbook 2015). Research shows that interventions targeting water issues are effective at addressing environmental degradation, forging ties between stakeholder groups, and catalyzing shifts toward peace (Ide and Detges 2018). Water cooperation engagements last, too. A study found that 145 water-related treaties created through cooperative negotiations have proved to be resilient over time, even as hostilities over other issues persist (Wolf 1998).

For water-focused interventions to be most effective, practitioners need to help identify and prioritize opportunities to intervene *before* the outbreak of conflict, as well as provide assistance to understand local conditions and engage with local actors. This technical note introduces a web-based tool, powered by a new conflict forecasting model, built to help meet that first need: identifying and prioritizing opportunities for water-related interventions. The tool is just one offering of the Water, Peace and Security (WPS) partnership, described below; other initiatives of the partnership offerings address the remaining needs.

Existing Conflict Tools and Models

Conflict modeling is a rapidly growing field. Some modelers within the field prioritize *understanding* existing conflicts, while others aim to *forecast* them. Our model falls into the latter category, although it incorporates ideas from both. Some work within the *understanding* line of research supports the notion that water can be an important contributor to conflict. This research is often qualitative, hyperlocal, or both. For example, Adelphi's ECC Factbook (Adelphi 2019) uses a curated collection of past and present conflicts to map causal pathways linking water and climate (among other drivers) to conflict. The Pacific Institute's Water Conflict Chronology also provides a curated set of conflicts, although these only comprise water-related conflicts. For both projects, qualitative analysis on known conflicts can be easily explored on a web-based tool. However, these tools are not intended to forecast emerging conflicts.

Uppsala University publishes one of the best open-source conflict forecasting models, the Violence Early-Warning System (ViEWS)[1]. ViEWS releases monthly forecasts for the African continent. Using an ensemble method, the forecasts are produced for three types of conflicts: state-based, one-sided, and non-state-based. ViEWS's forecasts are released along with an updated analysis in a PDF format, or as a data download. While its results are descriptive and transparent, they are currently not available on a web-based interactive platform. Differences between ViEWS and the WPS model exist, including the latter's inclusion of environmental data, but these differences create a valuable opportunity for benchmarking. As Hegre et al. (2018) note, there is no universally valid metric to assess model performance. We hope that by comparing two disparate models (see Benchmarking the Model section p. 20), we can not only improve conflict forecasting but also build legitimacy and trust among potential users.

WPS Tool and Model

Our research is focused on enabling timely, effective, water-related interventions to mitigate conflict and/or build peace. To that end, we have released a web-based tool to help our audience prioritize locations that might benefit the most from water-focused interventions. The overall objective is to offer a platform where actors from the defense, development, diplomacy, and disaster relief sectors (among others) can identify conflict hot spots before violence erupts, begin to understand the local context, and prioritize opportunities for water interventions. This requires information that is timely, accurate,

and actionable. To meet user demand for early warning information, we will release updated 12-month forecasts every 3 months. The tool allows users to explore these forecasts of ongoing and emerging conflicts, as well as underlying model inputs and contextual indicators across a variety of domains.

The forecasting model was designed to be as useful as possible to our core audience, while still prioritizing performance. Fundamental aspects of the modeling paradigm, including how conflict is defined, the temporal and spatial resolutions, and the forecast time horizon, were chosen to match user needs. For example, given that our audience is concerned with the impacts of conflict, we did not limit our investigation to conflicts driven by water challenges. Instead, we forecast all instances of organized violence (fully defined in the Dependent Variable section p. 7). User engagement also led us to prioritize identifying as many potential hot spots as possible, analogous to an initial medical screening.

The foundation of our forecasting is an expansive library of quantitative indicators potentially related to conflict. These indicators, our "predictor variables," are available for exploration as both interactive maps and time series. Some of these indicators are not fit for use in a quantitative model but are nonetheless useful for decision-making. The tool includes these contextual indicators alongside the model inputs. All tool functionality—access to datasets, metadata, and geospatial visualization specifications—is powered by the Resource Watch API, an open-source service designed to easily integrate into user workstreams.

As described below, the machine learning model cannot attribute specific causes to its forecasts. Our tool, therefore, is not intended to elucidate causal relationships between the predictor variables and conflict. The tool does, however, highlight instances of water shocks (i.e., heavy rains or drought) reflecting our audience's interest in water-related interventions. While we do not know or claim that water shocks drive conflict, we do believe they are important for screening when on-the-ground adaptation measures are needed. It is in places that require adaptation, but lack the resources to effectively respond, that water interventions may be most impactful. On the tool, the user can highlight areas exhibiting both a high risk of conflict and near-real-time water shocks like severe precipitation anomalies (as described by the Standardized Precipitation Index [McKee et al. 1993]) to prioritize areas of interest. The user can then examine contextual data to learn about local underlying conditions. Filtering by water shocks is optional, but encouraged.

Partnership

The model and web-mapping interface are products of the WPS partnership, which is pioneering the development of public information tools and approaches that can support evidence-based actions to reduce security risks and promote water cooperation. The partnership is composed of organizations with expertise in water and conflict, including the IHE-Delft Institute for Water Education (IHE-Delft, lead), the World Resources Institute (WRI), Deltares, The Hague Centre for Strategic Studies (HCSS), Wetlands International, and International Alert. The partnership is funded by the Netherlands Ministry of Foreign Affairs. WPS's work is strengthened by affiliated partners like New America, Pacific Institute, and Oregon State University. Guided by the global model forecasts, WPS will closely investigate areas of concern to better understand the possible role of water in those places. Specifically, WPS's local assessments will examine potential human responses to changing water availability and the consequences of those responses for local cooperation. On the ground, WPS will offer trainings and capacity development to support stakeholders in building coalitions and facilitating dialogue processes. WPS will work with key stakeholders to consider and implement early interventions, and, hopefully, prevent conflict.

METHODOLOGY

Approach

Machine learning techniques are adept at detecting subtle or complex patterns within large volumes of data without requiring *a priori* instructions or a theoretical framework. Since conflict literature does not offer a clear consensus on what factors contribute to violence (Cederman and Weidmann 2017), nor on the causal pathways to conflict, machine learning was identified as a viable means of establishing quantitative linkages between local conditions and the subsequent emergence (or continuation, or cessation) of conflict. We elected to use a machine learning approach after a literature review indicated its predictive power outperformed both subject matter experts and standard statistical approaches (Tetlock 2005; Mullainathan and Spiess 2017).

Model Type

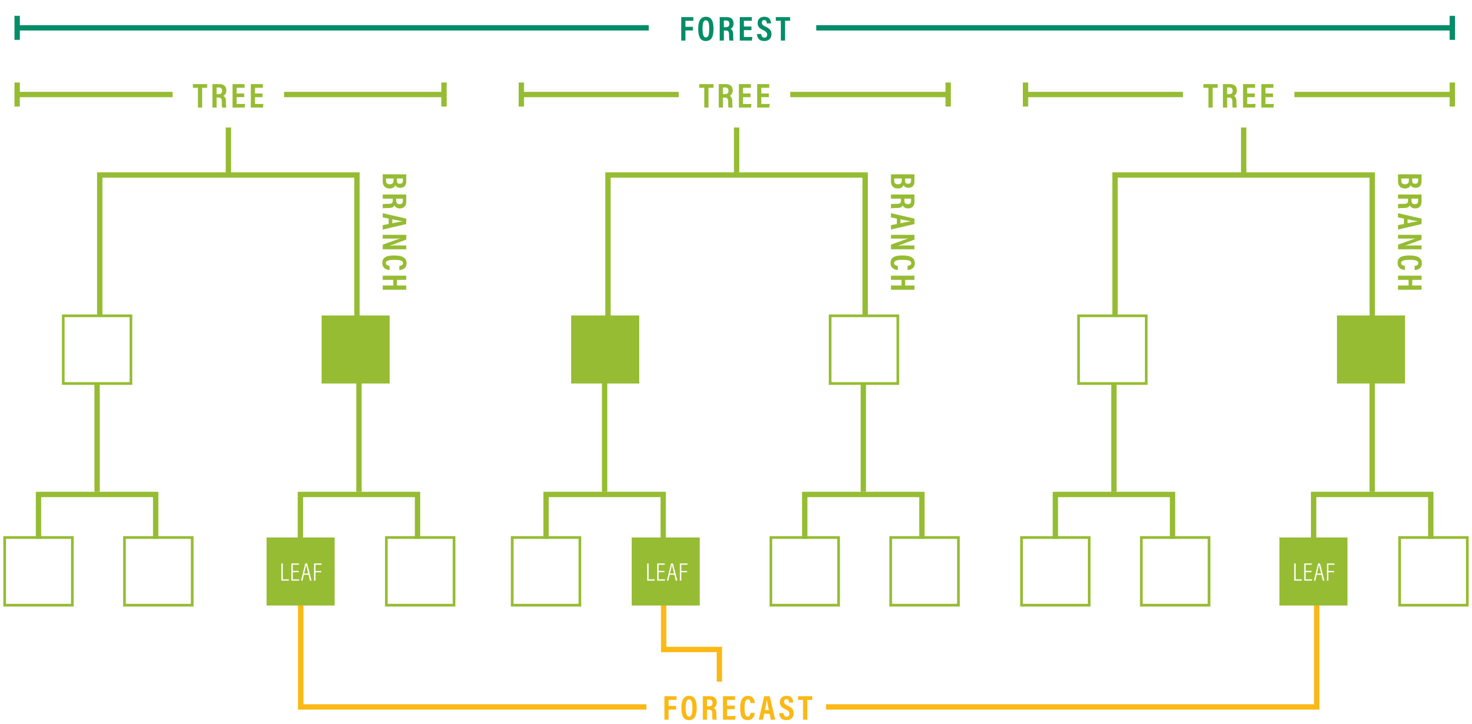
Machine learning encompasses a huge range of mathematical structures and techniques. We employed the random forest (RF) model type, a vehicle for ensemble supervised learning. In supervised learning, models are exposed to concrete examples of inputs and corresponding outputs in order to detect patterns linking the two.[2] In the context of this project, the inputs are the aforementioned "predictor variables," while the model output is referred to as the "dependent variable" or "forecast."

An RF algorithm generates an ensemble of decision trees (aptly named after their branching structure, illustrated in Figure 1). A "tree" uses a random subset of the predictor variables to make decisions about the value of the dependent variable. Each decision, or "branch," leads the tree closer to its final prediction, the "leaf" (Keller and Evans 2019). The "forest" then tabulates the individual tree's predictions in a "vote," with the most popular becoming the overall model's prediction (in our case, either Yes/Conflict or No/Peace) (Breiman 2001). Once the RF model is trained using known inputs and outputs, we can "run" a new sample of predictor variable values through the model to generate a forecast. We can even use the leaves to calculate the confidence of the forecast (i.e., the fraction of trees that voted for the winning category).

We chose RF for its potential to generate high-quality conflict forecasts. Perry explored the feasibility of forecasting conflict using RF and found that an RF model created using a diverse set of predictor variables performed better than a baseline model using prior violence as the sole predictor (2013). ViEWS also outperformed a baseline model (Hegre et al. 2018). Although this latter piece of research was published only after we made our selection, it supports the viability of ensemble-based models[3] for forecasting conflict.

Additional features made RF particularly suitable for our work. First, many models created through machine learning constitute "black boxes" whose inner workings are difficult or impossible to understand, such as neural networks. While RF, too, can be labeled a "black box," the structure of the trees that constitute an RF model is relatively intuitive, and it is easy to follow how each tree makes its predictions, even though the aggregate model, the forest, may remain somewhat opaque. RF is thus more apprehensible to a general audience than many alternatives, which may help bolster its credibility in the eyes

Figure 1 | **Random Forest Schematic**



Source: Authors. Based on Figure 2 from Keller and Evans 2019.

of potential users—a critical prerequisite for its deployment and application in the real world. Furthermore, there are established, if imperfect, ways of characterizing the relative importance of different variables to an RF model's decision-making. Lastly, RF offers flexibility. The structure can accommodate highly dissimilar predictor variables without extensive preprocessing—important for a project incorporating indicators from a wide variety of domains and in a range of formats. It is important to note that these qualities are shared by other analytical techniques. While we did not test other model structures at this stage in our research, our Next Steps section explores how we may expand our methods in future iterations.

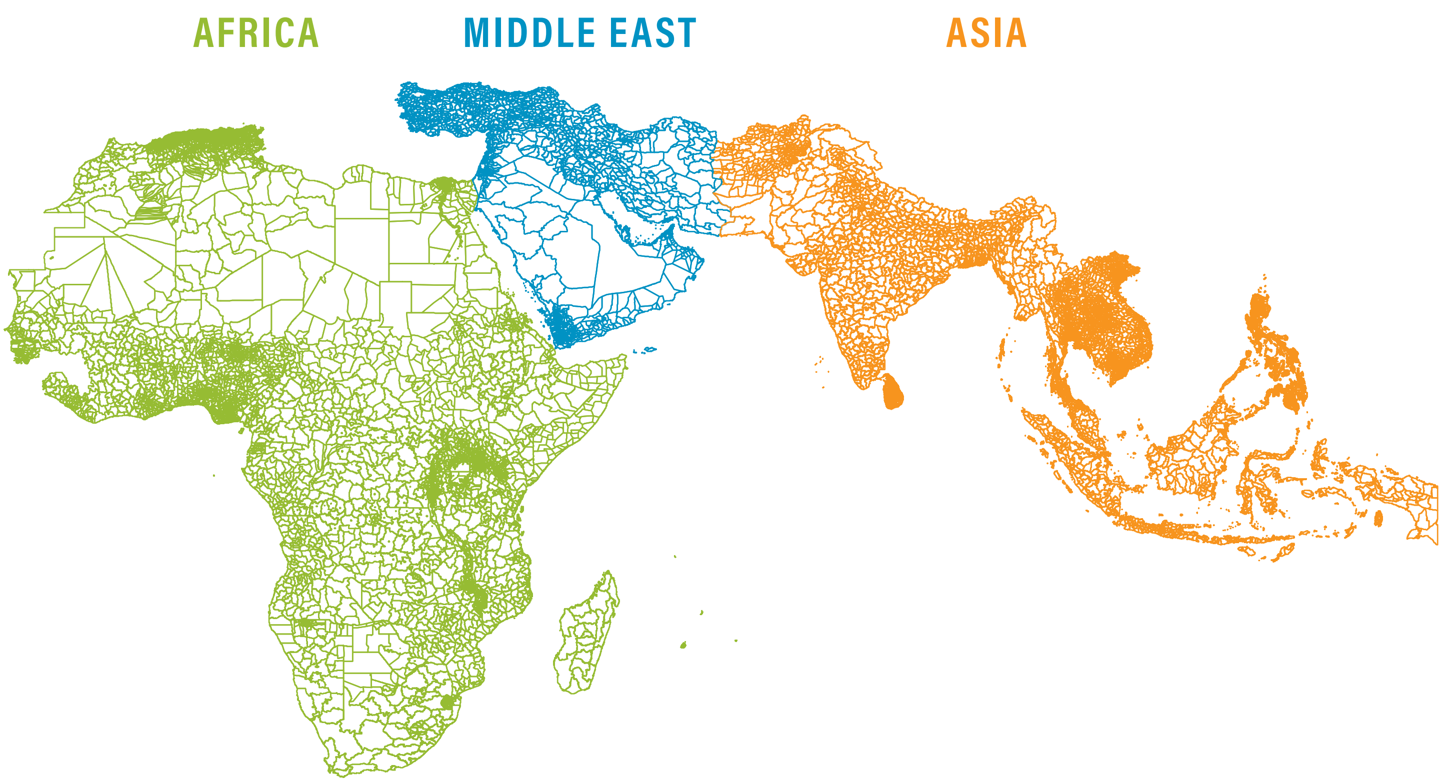
Unit of Analysis

The data used as indicators in our model capture a wide range of factors—from socioeconomics to governance to biophysical phenomena—and are correspondingly collected at a variety of geographic and temporal resolutions. In order to utilize and analyze them in concert, we had to bring these disparate types of data within a shared framework. We achieved this by reconciling all data to a common unit of analysis.

RESOLUTION

Understanding that a useful forecast must specify both where and when something is expected to happen, our unit of analysis has two dimensions: the "district-month" represents a given second-level administrative unit[4] (hereafter, a "district") in a given month.

Figure 2 | **Districts Used to Train, Validate, and Test the Model**



Note: The inclusion of certain regions does not imply a higher conflict risk; it simply reflects the research interest of the data producers.
Source: Authors. Data from GADM 2018. Regions defined by ACLED 2019.

For the spatial dimension, no single set of borders can perfectly accommodate the range of datasets, which can correspond to watersheds, aquifers, political borders, or fine or coarse grids, among others. Administrative boundaries are widely recognized, well understood by all audiences, and offer multiple levels of specificity, thus representing an acceptable compromise between familiarity and precision. Since governments, companies, international organizations, and NGOs alike often plan and execute on the basis of administrative units, datasets that correspond to those same boundaries may be more easily incorporated into decision-making processes. Modeling at the district level may have little precedent in the literature, and introduces its own difficulties—for example, the huge variation in size of districts within and across countries—but reflects the paramountcy of user needs in our model design. Our district boundaries, based on the Database of Global Administrative Areas (GADM 2018), were created using the methodology described in Appendix A.2.

The situation is similar with the temporal dimension: some indicators can be measured daily; others only yearly, if that. For many indicators, monthly updates represent the fastest possible tempo. At the same time, any frequency greater than monthly risks masking critical signals in the data, such as anomalous weather during a short but vital window of time for planting crops. Preserving this temporal precision is important enough that we chose to use a monthly rather than annual timestep as the temporal unit of analysis.

COVERAGE

We eventually plan to produce forecasts for every district around the world. However, only districts with at least three years of recorded organized violence (in our data source) were eligible to train, validate, and test the model. At the time of this publication, the geographic scope was limited to the following regions: Africa, Asia, and the Middle East, shown in Figure 2 (regions are defined by ACLED [2019]).

Performance Metrics

For each month in each district, our model aims to predict the occurrence of organized violence resulting in at least 10 fatalities over the next 12 months (full definition in Dependent Variable section on the next page). When our model correctly forecasts the presence of conflict, that successful prediction is called a true positive. A failure to forecast the absence of conflict is referred to as a false negative. Table 1 arranges the four possible outcomes into a table known as a confusion matrix. When evaluating model performance, we prioritized the successful identification of actual occurrences of conflict (i.e., true positive scenarios).[5]

Table 1 | **Confusion Matrix for Assessing Model Results**

	CONFLICT PREDICTED	CONFLICT NOT PREDICTED
CONFLICT OCCURRED	True Positive	False Negative
CONFLICT DID NOT OCCUR	False Positive	True Negative

Source: Authors

We characterize the performance of our model via three metrics, all available in the scikit-learn module (Pedregosa et al. 2011). While these are just three out of many options, they help us describe the strengths and weaknesses of our model in a transparent, straightforward way. They can also be applied to other model types for comparison. We used these metrics not only to assess model performance but also to monitor its sensitivity. As we tested a new model configuration, we would train models using different sets of samples to ensure our results were not due to the chance selection of a peculiar subset of training data. The scikit-learn documentation describes the three performance metrics as follows (Pedregosa et al. 2011):

Recall: the ability of the model to identify all positive samples

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}$$

Precision: the ability of the model not to mislabel as positive a sample that is negative

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

F2 score: the weighted harmonic mean of recall and precision, with recall having twice as much influence

$$F2 = (1 + 2^2)\frac{Precision * Recall}{(2^2 * Precision) + Recall}$$

Like an initial medical screening, our model is optimized to flag *all* concerning cases for further analysis. In other words, we would rather wrongly forecast the presence of conflict than incorrectly forecast its absence (i.e., "peace," in the strictly negative sense). For this reason, we prioritize recall over the other metrics. The downside to this decision is that our model is likely to overestimate conflict. While we believe this shortcoming is acceptable for our model's purpose, our continuing research will explore improving model precision.

Model Data

Dependent Variable

DATA SOURCE

In order to employ a machine learning approach, we needed a transparent dataset of organized violence with high temporal and spatial resolution to use as the dependent variable. We chose the Armed Conflict Location and Event Database (ACLED) (Raleigh et al. 2010). Each event in this database is manually classified and verified by multiple researchers. Every event is assigned a precise date and location (i.e., coordinates); event classification follows a clear codebook (ACLED 2017) ensuring comparability across time and space. Updated data are released weekly.

PROCESSING STEPS

A qualifying conflict event was any instance of an event type listed in Table 2. Nonviolent events such as strategic developments and peaceful protests were discarded. The location of each remaining event was matched to a second-level administrative unit as described in Appendix A.3, and dates were aggregated by month. In this way, a set of events (possibly empty) was associated with each district-month.

For each district-month, we then constructed a corresponding binary value by examining all qualifying events occurring in that district over the following 12 months. "Yes" was assigned to district-months with 10 or more fatalities over the next year, "No" to the others to represent peace. For example, the dependent variable for all district-months in October 2017 would sum the fatalities from qualifying violent conflicts from October 2017 to September 2018, as shown in Figure 3. The binary value associated with any sum greater than 10 would be "Yes." The minimum 10-fatality threshold for conflict helps filter out immaterial events (such as a heart attack at a protest)

Table 2 | **List of ACLED Events Included in Encapsulation of Conflict as Our Dependent Variable**

EVENT TYPE	SUB-EVENT TYPE
Battles	Armed clash
	Government regains territory
	Non-state actor overtakes territory
Explosions/remote violence	Air/drone strike
	Chemical weapon
	Grenade
	Remote explosive/landmine/IED
	Shelling/artillery/missile attack
	Suicide bomb
Violence against civilians	Abduction/forced disappearance
	Attack
	Sexual violence
Riots	Mob violence
	Violent demonstration
Protests	Excessive force against protestors

Source: Authors. Based on ACLED 2017.

Figure 3 | **Temporal Schematic of the Dependent Variable**

Run date for training sample | Twelve month moving window

	2017					2018									
	AUG	SEPT	OCT	NOV	DEC	JAN	FEB	MARCH	APRIL	MAY	JUNE	JULY	AUG	SEPT	OCT
Fatalities from qualifying organized violent conflict	10	9	0	0	0	5	4	0	0	0	0	0	0	0	15
Dependent variable value for training sample			NO												

Note: The dependent variable for each training sample is created by summing fatalities from organized violence 12 months into the future. If 10 or more deaths occurred, the dependent variable is set as "Yes" (binary 1). If less than 10 deaths occurred, the variable is set to "No" (binary 0). In the example above, 9 fatalities occurred over the 12 month window, so the dependent variable was set to No.
Source: Authors.

from our training set while still capturing emerging, small-scale conflicts that are of interest to our users. The threshold chosen reflected feedback we received during our user engagement phase.

Predictor Variables

DATA SOURCE

Research shows that a diverse set of predictor variables yields a better forecast (Perry 2013). We, therefore, adopted a variation of the United Nations Development Programme's (UNDP) human security framework to capture the full spectrum of security-relevant indicators in our tool and model. Defined by the UNDP as "freedom from want and freedom from fear," human security emphasizes the welfare of individuals through seven main categories: community, economy, food, environment, health, personal, and political (UNDP 1994). The relationship between human security and conflict is strong and complex, as conflict can be both a cause and effect of human insecurity (Barnett and Adger 2007). Human security is a useful paradigm for clarifying how diverse issues—ranging from hunger to repression to environmental destruction—interact and require comprehensive, context-specific solutions (UN n.d.).

For our own framework, we made a few structural shifts based on the research priorities of WPS. Our themes are intended to improve our model's performance by identifying gaps and biases in our database and provide the user with an intuitive data scheme to facilitate his/her own context-specific research. We dissolved the "health" and "personal" themes into the other categories, relabeled "political" as "governance," and added a category for past conflict. We also chose to focus the environmental category solely on water during this initial phase. Our final data themes are community, conflict, economy, food, governance, and water. In the future, we intend to add an "energy" theme and expand our "environment" theme from just water to a wider natural resource lens. The categories (as conceptualized by WPS) are described thus:

Community contains information about local society and its demographics. Insecure communities grapple with distrust and competition for resources, exposing them to outbreaks of violence.

Conflict describes how peaceful or violent an area was over the past 12 months. Areas with a recent history of conflict are more likely to experience conflict in the future (Perry 2013).

Economy provides details on household and country-level economic security. Insecure economies may lack the capacity to manage the volatility of resources like water and food, making them more vulnerable to destabilizing events.

Food characterizes an area's relationship to food through its production, dependence on agricultural livelihoods, and access to markets. Disruption of food availability and livelihoods may increase unrest.

Governance examines the capacity of a government to adequately provide basic services for its citizens (such as drinking water or electric power) and the ability of citizens to interact with their government. Like economic security, good governance can provide a buffer to its people during times of instability. Poor governance and a lack of political security may engender unrest.

Water monitors local water resources. Water resources are an integral part of food security, economic security, personal health, and ecosystem health. Disturbances in water supply—whether through droughts, floods, or poor management—may trigger insecurity through any of these pathways and contribute to the outbreak of violence (Gleick and Iceland 2018).

Table 3 | **Common Data Themes Observed in Literature Review**

RESEARCH ON CONFLICT AND MIGRATION FORECAST	COMMUNITY	CONFLICT	ECONOMY	FOOD	GOVERNANCE	WATER
Auclair 1999	■			■	■	■
Farinosi et al. 2018	■		■	■	■	■
Halkia et al. 2017	■	■	■	■	■	■
Hegre et al. 2018	■	■	■	■	■	
Muchlinski et al. 2016	■		■		■	
Perry 2013	■	■	■		■	■
Robinson and Dilkina 2017	■			■		
Stamatia et al. 2017	■	■	■	■	■	■

Note: All papers listed in this table explore forecasting conflict and/or migration. We tracked critical data themes to help prioritize which datasets to include in our own analysis. The categorizations are subjective and informal, and the list is by no means comprehensive.
Source: Authors.

We identified a number of global datasets to test in the model as predictor variables via our literature review (a sample of which is outlined in Table 3) and internal expertise. Datasets must be global in coverage, open-source, and methodologically transparent. While we favored datasets with a history dating back to at least the early 2000s, we also accepted datasets without a historical record in the absence of a viable alternative. In Table 4, we list every predictor variable tested in our model, organized by our six thematic categories. As Table 4 illustrates, many datasets cut across multiple themes. For example, child malnutrition falls under the community category as a proxy for gender equality, but also reflects the status of local economic and food security, as well as the government's ability to provide for its citizens. All datasets were assigned a primary category. In the tool, we also tag datasets by their cross-cutting themes to encourage data discovery. Users simply click on a tag, such as "water," to see a list of all datasets related to the water theme (regardless of whether that is their main category).

PROCESSING STEPS

All predictor variables used in this analysis were processed to align with the district-month unit of analysis. We matched most spatial data to the district scale in two ways:

- For data available at a finer resolution (such as gridded data), values within a district were statistically aggregated using an appropriate transformation (for example, calculating the mean of all grid elements within a district).
- For data available at a coarser resolution (such as country-level data), all districts within a country were assigned the country value.

Temporal data preparation followed a similar pattern. For indicators available at a daily or weekly tempo, we aggregated values to the monthly scale, typically using the sum (for details, see Appendix A.4). Coarser data were upsampled; for example, values reported annually were simply assigned to all months within the corresponding year (i.e., each month was given the annual value). We were cautious about overextending data and decided against interpolating annual data to create monthly values. Around half of our datasets are annual, meaning a large portion of our analysis would be based on unvetted interpolations. In comparison, datasets with a five-year time step are all rooted in one source, United Nations Department of Economic and Social Affairs (UN DESA) statistics. We, therefore, felt it reasonable to interpolate five-year data, after consulting UN DESA methodology, to create annual estimates.

The final step was to prepare these data for machine learning ingestion by constructing samples, where each sample is a single set of input values drawn from each of the various predictor variables (for training samples, these are then paired with the corresponding dependent variable value: Yes/Conflict/1 or No/Peace/0). This requires more than simply reconciling the predictor variables to the unit of analysis. Our system is intended to be used on

Table 4 | **List of Datasets Tested in Model**

	DATA TESTED IN MODEL	SOURCE	SPATIAL RESOLUTION	TEMPORAL RESOLUTION	CROSSCUTTING THEMES					
COMMUNITY	Ethnic fractionalization	Cline Center for Democracy 2012	National	Annual	✓					
	Religious fractionalization	Cline Center for Democracy 2012	National	Annual	✓					
	Child malnutrition	World Health Organization 2019	National	Annual	✓		✓	✓	✓	
	Percentage of males that are aged 0–14	United Nations 2017	National	Annual	✓					
	Percentage of males that are aged 15–24	United Nations 2017	National	Annual	✓					
	Percentage of males that are aged 25–64	United Nations 2017	National	Annual	✓					
	Percentage of males that are aged 65+	United Nations 2017	National	Annual	✓					
	Percentage of population aged 0–14 that are male	United Nations 2017	National	Annual	✓					
	Percentage of population aged 15–24 that are male	United Nations 2017	National	Annual	✓					
	Percentage of population aged 25–64 that are male	United Nations 2017	National	Annual	✓					
	Percentage of population aged 65+ that are male	United Nations 2017	National	Annual	✓					
	National population	United Nations 2017	National	Annual	✓					
	Local population count	CIESIN 2016a	1 km	5-year	✓					
	Local population density	CIESIN 2016b	1 km	5-year	✓					
	Rural to urban ratio	PBL 2018; van Vuuren et al. 2007	1 km	Baseline	✓					
	Rural population	PBL 2018; van Vuuren et al. 2007	1 km	Baseline	✓					
	Urban population	PBL 2018; van Vuuren et al. 2007	1 km	Baseline	✓					
	Urbanization rate	United Nations 2019	National	5-year	✓		✓			
CONFLICT	Battles (count, fatalities, binary)	Raleigh et al. 2010	Point	Daily		✓			✓	
	Remote violence (count, fatalities, binary)	Raleigh et al. 2010	Point	Daily		✓				
	Violence against civilians (count, fatalities, binary)	Raleigh et al. 2010	Point	Daily		✓			✓	
	Violent protests and riots (count, fatalities, binary)	Raleigh et al. 2010	Point	Daily		✓			✓	
	Agreements (count, fatalities, binary)	Raleigh et al. 2010	Point	Daily		✓			✓	
	Peaceful protests (count, fatalities, binary)	Raleigh et al. 2010	Point	Daily		✓			✓	
ECONOMY	Gross domestic product per capita	World Bank Group 2017	National	Annual			✓		✓	
	Inflation, consumer prices (annual %)	International Monetary Fund 2019	National	Annual			✓		✓	
	Infant mortality	UN IGME 2019	National	Annual			✓		✓	
	Poverty headcount at $1.90/day	World Bank Group 2019	National	Annual	✓		✓		✓	
	Poverty headcount at $3.20/day	World Bank Group 2019	National	Annual	✓		✓		✓	

Table 4 | **List of Datasets Tested in Model (Cont'd)**

	DATA TESTED IN MODEL	SOURCE	SPATIAL RESOLUTION	TEMPORAL RESOLUTION	CROSSCUTTING THEMES					
ECONOMY	Poverty headcount at $5.50/ day	World Bank Group 2019	National	Annual	✓		✓		✓	
ECONOMY	Unemployment, total (% of total labor force)	International Labour Office 2017	National	Annual	✓		✓		✓	
ECONOMY	Employment in agriculture (% of total employment)	International Labour Office 2017	National	Annual			✓	✓	✓	
FOOD	Agriculture value added to GDP (% of total GDP)	World Bank Group 2016	National	Annual			✓	✓		
FOOD	Value of irrigated crops	International Food Policy Research Institute 2019	10 km	2010			✓	✓		✓
FOOD	Value of rainfed crops	International Food Policy Research Institute 2019	10 km	2010			✓	✓		✓
FOOD	Value of all crops	International Food Policy Research Institute 2019	10 km	2010			✓	✓		✓
FOOD	Percentage of crops that are rainfed	International Food Policy Research Institute 2019	10 km	2010				✓		✓
GOVERNANCE	Access to drinking water	WHO and UNICEF 2017	National	Annual					✓	✓
GOVERNANCE	Access to sanitation	WHO and UNICEF 2017	National	Annual					✓	✓
GOVERNANCE	Corruption index	Standaert 2015	National	Annual					✓	
GOVERNANCE	Political corruption index	Coppedge et al. 2019; Pemstein et al. 2018	National	Annual					✓	
GOVERNANCE	Executive corruption index	Coppedge et al. 2019; Pemstein et al. 2018	National	Annual					✓	
GOVERNANCE	Type of government (16 regime types)	Bell 2016	National	Annual					✓	
GOVERNANCE	Recent election	Bell 2016	National	Annual					✓	
WATER	Standard precipitation index 3-month anomalies	Johnson et al. 2019	1 km	Monthly				✓		✓
WATER	Standard precipitation index 6-month anomalies	Johnson et al. 2019	1 km	Monthly				✓		✓
WATER	Standard precipitation index 12-month anomalies	Johnson et al. 2019	1 km	Monthly				✓		✓
WATER	Standard precipitation index 24-month anomalies	Johnson et al. 2019	1 km	Monthly				✓		✓
WATER	Actual evapotranspiration	Senay et al. 2011	1 km	Monthly				✓		✓
WATER	Evapotranspiration 1-month anomalies	Senay et al. 2011	1 km	Monthly				✓		✓
WATER	Evapotranspiration 12-month anomalies	Senay et al. 2011	1 km	Annual				✓		✓
WATER	Baseline water stress	Hofste et al. 2019	Admin 1	Baseline				✓		✓
WATER	Riverine flood risk	Hofste et al. 2019	Admin 1	Baseline					✓	✓
WATER	Seasonal variability	Hofste et al. 2019	Admin 1	Baseline				✓		✓
WATER	Interannual variability	Hofste et al. 2019	Admin 1	Baseline				✓		✓

Note: Each dataset is listed along with its source, original spatial and temporal resolution, and primary category.
Source: Authors.

a real-time basis, generating fresh, up-to-date forecasts every quarter. Ideally, this would simply be a matter of feeding equally up-to-date predictor variable values into the model. However, many of our datasets are not and will not be available in near real time. For example, GDP data for 2018 were not available until late 2019. Therefore, we cannot expect to be able to run our model in January 2019 using 2018 GDP data. We compensated for this delay by "lagging" predictor variables: using datasets on a delayed basis in order to ensure that a value is reliably available for use as model input. Thus, the predictor variable attached to a given district-month may in fact represent a measurement made considerably earlier. In the above example, that January 2019 forecast would utilize GDP data from 2017. Importantly, the model is trained and applied using a consistent lag for each dataset: not only does the January 2019 forecast utilize 2017 GDP data, but the training samples representing January 2018 use GDP data from 2016, and so on. The length of this lag for a dataset depended on the delay in its availability. Lags range anywhere from zero months (i.e., new data is available at the beginning of the month) to 48 months (i.e., new data is published every two years). Finally, some indicators do not change over time, but rather represent a baseline condition, such as Aqueduct's Baseline Water Stress (Hofste et al. 2019). No lag is applied to these baseline datasets. We provide a more detailed explanation of all processing steps for each dataset in Appendix A.1.3.

There are, of course, limitations to these processing decisions. For example, we may be blunting the decision-making impact of our annual data by not interpolating monthly values. For gridded data that were downsampled to the district level, it is possible that we are smoothing over local differences, particularly in the larger districts. In the future, we will experiment with alternative processing, such as applying more interpolations and using a quantile approach when downsampling instead of taking the average.

Model

The objective of our modeling is to forecast the risk of conflict as effectively as possible. To that end, we created models utilizing different combinations of input variables and hyperparameters. Each model used historical data (described in the Model Data section p. 7) to learn to "predict" past conflict outcomes based on past conditions. We then applied each model to a separate tranche of more recent historical data and used its performance on this validation data to estimate the model's predictive power on "true" forecasts made based on current conditions.

Structure

We implemented our modeling using the well-known Python data analysis library scikit-learn[6] and executed the code via Jupyter notebooks housed within Google's Colaboratory environment.[7] Since nonparametric and nonlinear model structures like RF can fit training data very closely, overfitting[8] was a concern. Therefore, model hyperparameter selection centered on minimizing overfitting while still optimizing for performance. Hyperparameter selection was conducted via manual parameter search, again measuring model performance by predictive success on reserved validation data. Tree depth was limited to 32 decision nodes (where a branch forks, see Figure 1), and the minimum number of samples needed to trigger a node split was increased to six.[9] The forest consisted of 200 trees.

Training, Validation, and Test Data

To better evaluate model performance, we simulated application conditions by leaving a gap[10] in time between our training data and validation data (i.e., no data from June 2016 through May 2017 were used). Data from the latest year available (June 2017–June 2018) were split into validation and test tranches and used to measure the performance of our model. Data tranches are summarized in Table 5.

Table 5 | **Overview of Data Tranches Used to Train, Calibrate, and Assess the Model**

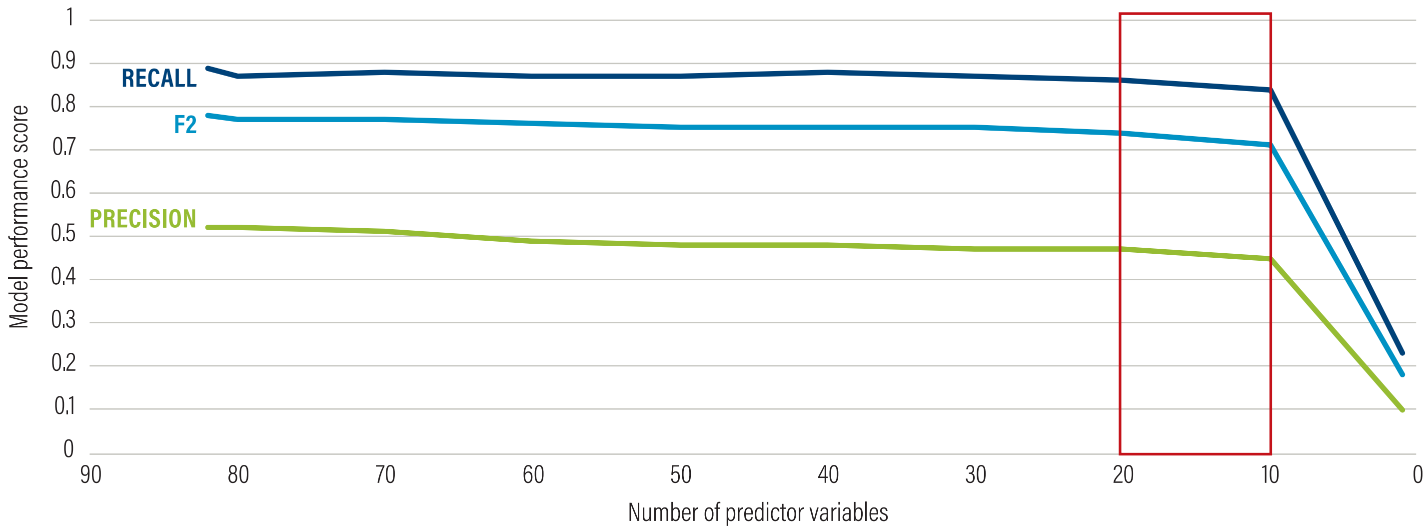
TRANCHE	PURPOSE	DATE RANGE
Training	Fit model to data	January 2004–May 2016
Validation	Estimate prediction error	June 2017–December 2017
Test	Assess final model	January 2018–June 2018

Source: Authors. The descriptions used in "Purpose" column are based on Hastie et al. 2009.

Balancing Training Data

In our full dataset, conflict is relatively rare: approximately 3 percent of district-months[11] experienced a qualifying number of fatalities over the following year (for any given year, this number ranges from 2 to 6 percent). Given such rarity, a machine learning–based model might try to maximize its accuracy simply by never predicting conflict. However, such a model would obviously be useless. (This also highlights the limitations of accuracy score[12] in assessing model quality).

Figure 4 | **Model Performance as a Function of the Number of Predictor Variables**



Note: Model performance stays relatively stable even as the number of indicators decreases until it reaches a cliff at 10 variables. From these results, we found that a model with around 10–20 indicators would perform almost as well as a model with 80.
Source: Authors.

Table 6 | **Defining the 'Direction' of Conflict**

CONFLICT DIRECTION	PAST 12 MONTHS	FORECASTED 12 MONTHS
Ongoing peace	Less than 10 fatalities	Less than 10 fatalities
Emerging conflict	Less than 10 fatalities	At least 10 fatalities
Ongoing conflict	At least 10 fatalities	At least 10 fatalities
End to conflict	At least 10 fatalities	Less than 10 fatalities

Note: "Past 12 months" refers to conflicts that actually took place. "Forecasted 12 months" are the forecasted results.
Source: Authors.

In order to ensure meaningful predictions, we constructed a balanced training dataset. We took all the district-months where conflict was present and added an equal number of randomly selected district-months where conflict was absent—a process known as random undersampling, since it omits samples from the more prevalent category. Undersampling is a straightforward method of dealing with imbalanced datasets that is popular across many fields (Liu et al. 2009; Effendy et al. 2014; Khalilia et al. 2011). We did not test alternative techniques for class-balancing, some of which are elaborations on undersampling. However, we elected not to employ techniques incorporating artificial samples, such as Synthetic Minority Over-sampling Technique (SMOTE, [Chawla et al. 2002]), in part to maintain methodological credibility in the eyes of consumers unfamiliar with modeling or machine learning, which includes almost all of our key audiences. In a distinct, but related, modification, we tried training with weighted classes—in essence, assigning greater value to samples of the less prevalent category—but observed no improvement in performance. Ultimately, we used a balanced subset of samples to train our model without class weighting.

A sensitivity analysis was conducted to test the impact of using different proportions of yes/conflict and no/peace samples in the training dataset. We found that a 1:1 ratio yielded the best model performance. This means we utilized only 7 percent of the total pool of available training data: all 32,000 of the yes/conflict samples, and the same number of the no/peace samples, randomly selected. We used this balanced subset of samples to train our model.

Reducing Model Inputs

As described in the Predictor Variables section, we constructed a large pool of 82 predictor variables from which to draw the actual model inputs. Utilizing fewer variables can help reduce input covariance, which in turn can improve the quality of measurements of indicator significance. Importantly, using fewer variables also increases the interpretability of the model. Theoretically, we could build a model using thousands of predictor variables, and it is possible we would see some improvement in performance (the use of all 82 indicators increased the precision by 7 percent relative to our final model, which uses 19 indicators). However, our model is intended to inform intervention decisions, and we believe that including too many indicators would reduce understanding and thus usability of our forecast. To compromise between a high-performing model and an interpretable one, we chose to reduce the number of predictor variables to a subset that would perform almost as well as the model using all 82. Working in the absence of clear theoretical guidance—in part due to poorly understood causal relationships—we turned to empirical methods for selecting which variables from the pool of available datasets should be included. We utilized Recursive Feature Elimination (RFE) to probe the relationship between the number of predictor variables and model performance, and to identify the variables most effective for predicting conflict. Starting with the full set of features, RFE[13] repeatedly retrains the model, in each iteration eliminating the least important feature (defined by a coefficient attribute of feature importance) until only one feature remains—in theory, the most important one (Pedregosa et al. 2011).

To find the overall top-performing predictor variables, we ran the RFE analysis using data from all available regions (Figure 2). We also ran the RFE separately for each region to explore the impact of geographical differences (four RFE experiments in total: general, Africa-specific, Asia-specific, and Middle East-specific). To ensure stability in the results, we repeated each RFE experiment three times, reshuffling the balanced samples used to train the model each time. Finally, we created a list of top-performing indicators per experiment by assigning predictor variables their highest achieved ranking (i.e., the minimum numeric rank of the three runs). We found that a model with as few as 10–20 predictor variables can achieve similar performance to a model with all 82 indicators (see Figure 4). Therefore, any variable that achieved a ranking of 15 or lower (i.e., more important) in any of the trials described immediately above was considered a top-performing variable for a given experiment.

Creating the Models

The first model we created, our general model, was trained using samples from all regions, and the predictor variables were limited to the top-performing indicators from the general RFE run (19 variables[14] in all). We then created region-specific models, which were trained using region-specific samples only, with the predictor variables utilized similarly dictated by the region-specific RFE results. In the end, four different models were trained. Each model was then applied sequentially to the samples from each region, with performance evaluated using the metrics listed in the Performance Metrics section.

To better understand model performance, we also chose to evaluate how each model performed in different scenarios, defined by past and future conflict status. For example, in a place where there has been peace, how well can a model forecast the emergence of conflict? Or, in a place experiencing conflict, how well can it forecast whether that conflict will end? We, therefore, split our dataset into four subsets based on the definitions in Table 6—every district-month (i.e., every sample) represents one (and only one) of these four scenarios. We again used the same three metrics to evaluate model performance.

RESULTS

Variable Selection

The top-performing predictor variables for each regional subset from our RFE analysis are displayed in Table 7. The results establish a correlation between these variables and our ACLED-derived conflict outcomes. While these results do not establish *causal* relationships, these indicators proved to be the most useful for accurately sorting events within the model. Results from trials using the full, multi-region dataset aligned most closely with the results from trials using only data from Africa, in that similar sets of indicators were identified by RFE as important. Africa has the longest ACLED record and is the largest region by far, meaning more samples are drawn from that region than any other (81 percent of training). This alignment suggests, unsurprisingly, that the relative proportion of samples has a major impact on the relationships

Table 7 | Predictor Variables Ranked by RFE Importance for General and Regional Models

Color scale denotes more important conflicts

1	2	3	4	5	6	7	8	9	10
MOST IMPORTANT									LEAST IMPORTANT

	DATA TESTED IN MODEL	GENERAL	AFRICA	ASIA	MIDDLE EAST
COMMUNITY	Local population count	1	1	2	9
CONFLICT	Battles (count)	2	2	1	4
COMMUNITY	Local population density	3	3	8	1
CONFLICT	Violence against civilians (count)	3	4	9	26
GOVERNANCE	Access to sanitation	3	6	39	49
ECONOMY	Gross domestic product per capita	5	4	41	47
CONFLICT	Battles (fatalities)	6	7	1	4
WATER	Seasonal variability	6	6	23	20
COMMUNITY	Percentage of males that are aged 65+	8	16	43	30
COMMUNITY	Rural population	9	9	14	3
COMMUNITY	Percentage of population aged 25–64 that are male	10	11	33	33
CONFLICT	Violence against civilians (fatalities)	10	10	5	28
FOOD	Agriculture value added to GDP (% of total GDP)	11	3	42	56
FOOD	Value of rainfed crops	12	15	13	18
WATER	Standard precipitation index 24-month anomalies	12	14	7	5
WATER	Interannual variability	12	12	16	19
COMMUNITY	Percentage of population aged 65+ that are male	13	16	7	29
WATER	Riverine flood risk	15	18	21	8
ECONOMY	Employment in agriculture (% of total employment)	16	33	11	50
WATER	Evapotranspiration 12-month anomalies	16	16	23	9
CONFLICT	Battles (binary)	17	19	18	13
CONFLICT	Remote violence (count)	18	38	5	2
COMMUNITY	Child malnutrition	20	20	39	48
FOOD	Value of all crops	21	20	20	11
WATER	Baseline water stress	22	19	27	12
ECONOMY	Poverty headcount at $5.50/day	23	25	57	44
FOOD	Standard precipitation index 12-month anomalies	25	24	15	10
COMMUNITY	Urbanization rate	27	31	12	37
COMMUNITY	Percentage of population aged 15–24 that are male	28	26	35	30
COMMUNITY	Urban population	28	26	36	25
ECONOMY	Poverty headcount at $3.20/day	29	41	57	62
FOOD	Percentage of crops that are rainfed	29	35	2	11
COMMUNITY	Rural to urban ratio	30	30	36	25
FOOD	Value of irrigated crops	31	30	10	11
GOVERNANCE	Corruption index	31	43	45	72
WATER	Actual evapotranspiration	31	34	27	20
WATER	Standard precipitation index 6-month anomalies	32	32	24	12
COMMUNITY	Percentage of population male ages 24-64	34	21	36	32
ECONOMY	Unemployment, total (% of total labor force)	36	35	41	80
GOVERNANCE	Access to drinking water	36	31	11	78
COMMUNITY	National population	38	38	37	29
GOVERNANCE	Poverty headcount at $1.90/day	40	53	54	44
COMMUNITY	Percentage of males that are aged 0–14	41	23	28	29
WATER	Standard precipitation index 3-month anomalies	41	40	31	23
CONFLICT	Peaceful protests (count)	42	39	27	30
WATER	Evapotranspiration 1-month anomalies	43	40	34	24
COMMUNITY	Religious fractionalization	45	48	22	58
COMMUNITY	Ethnic fractionalization	46	51	25	44
CONFLICT	Remote violence (fatalities)	46	54	12	6
COMMUNITY	Percentage of males that are aged 15–24	48	48	34	32
GOVERNANCE	Political corruption index	48	45	44	46
CONFLICT	Violence against civilians (binary)	49	44	55	40
COMMUNITY	Percentage of population aged 0–14 that are male	51	46	52	31
CONFLICT	Violent protests and riots (count)	54	49	40	41

Table 7 | **Predictor Variables Ranked by RFE Importance for General and Regional Models (Cont'd)**

	DATA TESTED IN MODEL	GENERAL	AFRICA	ASIA	MIDDLE EAST
ECONOMY	Inflation, consumer prices (annual %)	54	53	31	48
GOVERNANCE	Executive corruption index	55	50	50	56
CONFLICT	Violent protests and riots (fatalities)	57	57	52	43
GOVERNANCE	REIGN: Military	57	58	63	48
CONFLICT	Remote violence (binary)	59	62	52	22
GOVERNANCE	Recent election	59	58	56	45
GOVERNANCE	REIGN: Dominant party	60	60	68	51
GOVERNANCE	REIGN: Party presidential democracy	60	59	69	45
CONFLICT	Agreements (count)	63	63	54	41
GOVERNANCE	REIGN: Party personal dictatorship	63	61	67	52
GOVERNANCE	REIGN: Party provisional civilian	65	64	67	58
GOVERNANCE	REIGN: Party provisional military	65	64	66	57
CONFLICT	Violent protests and riots (binary)	67	66	62	60
GOVERNANCE	REIGN: Military personal	67	67	66	59
GOVERNANCE	REIGN: Warlordism	68	74	66	66
GOVERNANCE	REIGN: Party military	70	72	61	68
GOVERNANCE	REIGN: Party personal	70	69	67	67
GOVERNANCE	REIGN: Party personal military hybrid	71	69	71	69
GOVERNANCE	REIGN: Parliamentary democracy	72	73	64	70
GOVERNANCE	REIGN: Monarchy	74	71	74	71
CONFLICT	Peaceful protests (fatalities)	75	75	62	73
CONFLICT	Peaceful protests (binary)	76	76	75	74
GOVERNANCE	REIGN: Foreign occupied	77	77	77	75
GOVERNANCE	REIGN: Indirect military	78	78	78	76
CONFLICT	Agreements (fatalities)	79	79	79	77
GOVERNANCE	REIGN: Oligarchy	80	80	80	79
GOVERNANCE	Agreements (binary)	81	81	81	81

Note: The rankings are sorted by the RFE results for the general (i.e., all data) experiment; region-specific RFE results are also shown. These were created by a model trained using only samples for the given region. The first fifteen predictor variables listed were selected for the final model.
Source: Authors.

Table 8 | **General vs. Regional Model Performance Comparison**

		ALL REGIONS	AFRICA		ASIA		MIDDLE EAST	
		GENERAL	GENERAL	REGION-SPECIFIC	GENERAL	REGION-SPECIFIC	GENERAL	REGION-SPECIFIC
Overall	Recall	0.86	0.85	0.86	0.90	0.83	0.83	0.78
	Precision	0.47	0.40	0.39	0.49	0.55	0.62	0.69
	F2	0.74	0.70	0.69	0.77	0.75	0.77	0.76
Ongoing peace	Recall	0.93	0.93	0.92	0.94	0.95	0.95	0.97
	Precision	0.99	0.99	0.99	0.99	0.99	0.98	0.97
	F2	0.94	0.94	0.93	0.95	0.96	0.96	0.97
Emerging conflict	Recall	0.60	0.65	0.65	0.64	0.48	0.38	0.23
	Precision	0.20	0.22	0.21	0.17	0.17	0.21	0.22
	F2	0.43	0.47	0.46	0.41	0.35	0.33	0.23
Ongoing conflict	Recall	0.98	1.00	1.00	0.99	0.95	0.94	0.92
	Precision	0.75	0.67	0.67	0.81	0.87	0.78	0.80
	F2	0.92	0.91	0.91	0.95	0.93	0.90	0.89
End of conflict	Recall	0.11	0.08	0.08	0.07	0.42	0.18	0.29
	Precision	0.63	0.97	1.00	0.58	0.69	0.49	0.55
	F2	0.13	0.10	0.10	0.09	0.46	0.21	0.32

Note: The general model (trained with data from all regions) was tested in each region and then compared against the region-specific models. The results represent model performance on the test data (i.e., forecasts for January–June 2018). The number of samples per test: All Regions: 40,052; Africa: 18,438; Asia: 16,277; and Middle East: 5,337.
Source: Authors.

elucidated by the machine learning process. Additional measures of feature importance and coefficient correlation are provided in Appendix A.1.

Model Selection

General versus Regional Models

Four distinct models were trained: one using all data (general), and three using region-specific data (i.e., Africa, Asia, and the Middle East). The performance of the general model in each region is provided, along with the performance of the region-specific models for comparison, in Table 8. Performance is characterized by the three key metrics (see Performance Metrics section) calculated from the application of each model to the test tranches of samples (i.e., forecasts for January–June 2018). Models were assessed based on their overall performance as well as their success in forecasting emerging conflicts, which user engagement identified as the most important class of conflict for our audience.

Overall, the general model performed better than the region-specific models. This is especially true for emerging conflicts: Asia and the Middle East saw around a 15 percentage point increase in recall with the general model, relative to region-specific models, with little-to-no change in precision. The trade-off was a decrease in performance for the "end to conflict" category. While this effect is particularly pronounced in Asia, the improvement in forecasting emerging conflict makes this an acceptable trade-off, given our model's intended use.

Final Model: Error Rate and Suggested Application

Our final, "general" model was trained on a balanced dataset comprising of 74,118 samples from all regions, ranging from January 2004 to June 2016. The model generated forecasts based on 19 predictor variables, shown in Table 9.

Overall, the model captures 86 percent of future conflicts, successfully forecasting 98 percent of ongoing conflicts and 60 percent of emerging conflicts (Table 8). The trade-off for this high recall is low precision. The model's conflict signal is noisy, with half of conflict predictions representing false positives; that is, instances where conflict was forecast but did not actually occur. To examine precision further, it is possible to look at the error rate as a function of forecast confidence. Recall that a forecast is determined by tallying the "votes" of the decision trees in the forest; a forecast of conflict requires at least 50 percent of the trees voting that way. This ratio—conflict votes to total votes—can be described as the forecast confidence. As Figure 5 clearly shows, most incorrect predictions represent forecasts with a 50–75 percent confidence.

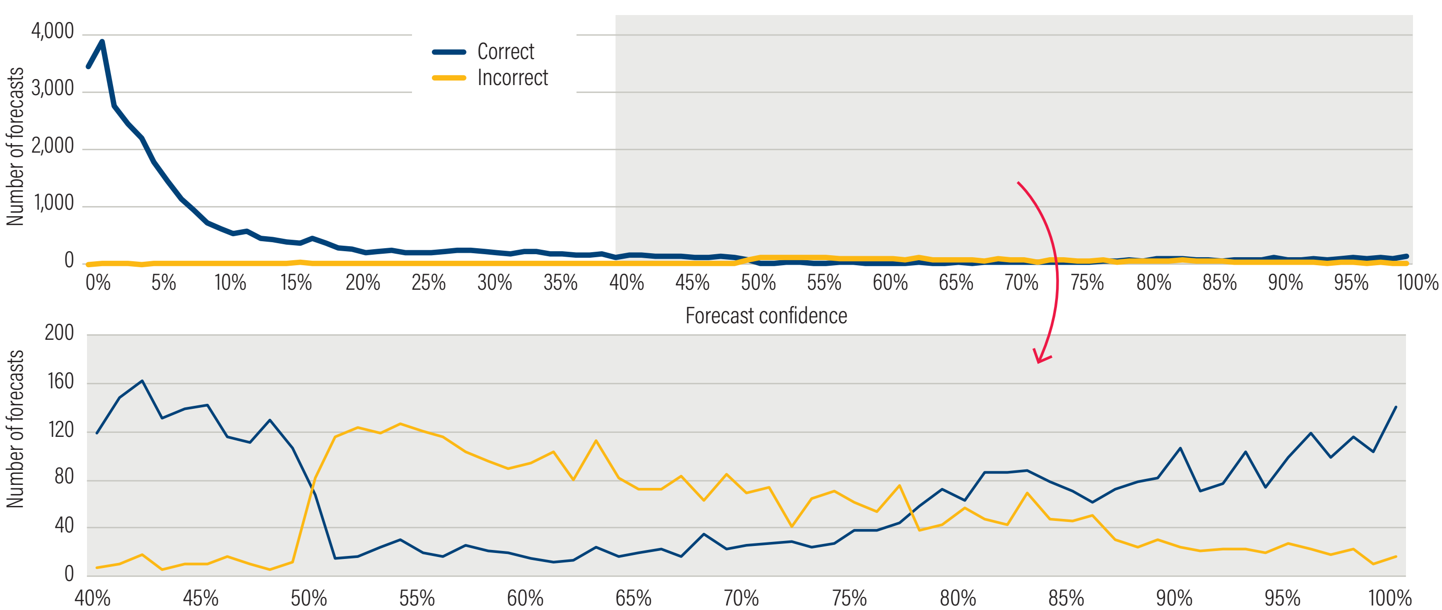
Our model's error rate varies inversely with forecast confidence. As Figure 6 shows, emerging conflicts are the most common class in the 50–70 percent confidence range, while ongoing conflicts greatly outnumber emerging ones when the confidence is more than 80 percent. These results reveal that the model is better at forecasting ongoing conflicts than emerging ones, and, intuitively, this makes sense. It is harder for the model to predict new activity compared to activity already under way.

This precision analysis has significant implications for the application of model forecasts and illustrates why the tool distinguishes between emerging conflicts and ongoing

Table 9 | **List of Predictor Variables Utilized by Final Model**

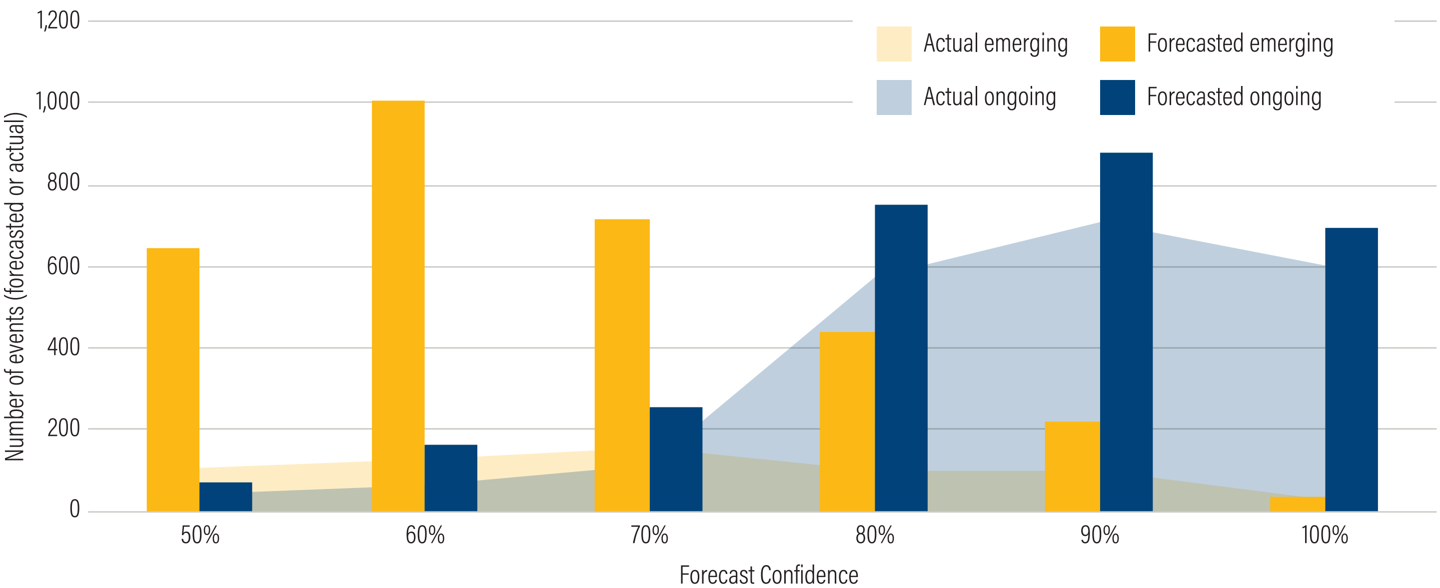
	DATA TESTED IN MODEL	SOURCE	SPATIAL RESOLUTION	TEMPORAL RESOLUTION	CROSSCUTTING THEMES					
COMMUNITY	Percentage of males that are aged 65+	United Nations 2017	National	Annual	✓					
	Percentage of population aged 25–64 that are male	United Nations 2017	National	Annual	✓					
	Percentage of population aged 65+ that are male	United Nations 2017	National	Annual	✓					
	Local population count	CIESIN 2016a	1 km	5-year	✓					
	Local population density	CIESIN 2016b	1 km	5-year	✓					
	Rural to urban ratio	PBL 2018; van Vuuren et al. 2007	1 km	Baseline	✓					
CONFLICT	Battles (count)	Raleigh et al. 2010	Point	Daily		✓			✓	
	Battles (fatalities)	Raleigh et al. 2010	Point	Daily		✓			✓	
	Violence against civilians (count)	Raleigh et al. 2010	Point	Daily		✓			✓	
	Violence against civilians (fatalities)	Raleigh et al. 2010	Point	Daily		✓			✓	
EC.	Gross domestic product per capita	World Bank Group 2017	National	Annual			✓		✓	
FOOD	Agriculture value added to GDP (% of total GDP)	World Bank Group 2016	National	Annual			✓	✓		
	Value of rainfed crops	IFPRI 2019	10 km	2010			✓	✓		✓
GOV.	Access to sanitation	WHO and UNICEF 2017	National	Annual					✓	✓
WATER	Standard precipitation index 24-month anomalies	Johnson et al. 2019	1 km	Monthly				✓		✓
	Riverine flood risk	Hofste et al. 2019	Admin 1	Baseline					✓	✓
	Seasonal variability	Hofste et al. 2019	Admin 1	Baseline				✓		✓
	Interannual variability	Hofste et al. 2019	Admin 1	Baseline				✓		✓

Figure 5 | **Correct and Incorrect Forecasts per Forecast Confidence**



Note: The confidence refers to the percentage of trees in the forest that forecasted conflict. The results shown represent the application of the model to the test data (i.e., forecasts for January–June 2018). Most of the incorrect forecasts had a confidence between 50 and 75 percent.
Source: Authors.

Figure 6 | **Number of Ongoing and Emerging Conflicts Displayed per Forecast Confidence**



Note: Chart shows all instances of forecasted conflict (i.e., forecast confidence above 50 percent). Both forecasted and actual conflicts are shown. The data shown represent application of the model to the test dataset (i.e., January–June 2018).
Source: Authors.

conflicts. Users interested in the ongoing conflict forecasts can have high confidence in the forecast, and may feel comfortable acting on this information immediately. For emerging conflicts, users can view these results as a "first screening," feeling confident that our "net" has caught most emerging conflicts, but acknowledging they are interspersed with many instances of peace. These users can then use the tool to investigate the local context in the potential hot spots before engaging with the WPS partnership to prioritize intervention opportunities.

Influence of Water Indicators

Given WPS's focus on water, we were interested in understanding the impact of water variables on model predictions. We hypothesized that water indicators would be influential for two reasons. First, water is integral to human security: food, health, economic, and ecosystem security all depend on the availability of high-quality water. Second, the structure of the water data may exaggerate the impact of water-related factors. Machine learning performs better when it has more information to make decisions, and variation in the input data can be linked to variation in the target variable.

The presence of water predictor variables in the top performing indicators suggests that there is at least a correlation between water variables and conflict. To test this more directly, we trained an otherwise identical model without any water-related predictor variables. The absence of water indicators resulted in no change in recall, and a negligible decrease in precision (1 percent, within the model's margin of error). While these results do not support our hypothesis, they do align with the notion that water is not a driver of conflict, but rather a contributor. Interestingly, water-related predictor variables had a much more significant impact when the threshold for conflict was set below 10 fatalities (i.e., +5 percentage points in recall for overall conflict, +12 percentage points for emerging conflict). Further analysis is needed to parse out the relationship between our predictor variables and the *severity* of conflict.

One final analysis was performed to assess the influence of water. In the previous section, we establish that the model forecasts ongoing conflicts better than emerging ones. Despite our interest in forecasting emerging conflict for early warning, our model did not weigh emerging conflicts as more important during training. Rather, the model strives to be "right" as often as possible. Given that there are more instances of ongoing conflict than emerging conflict, it is possible that the model prioritized predictor variables that help correctly identify ongoing conflict. Therefore, we applied the RFE process (see the Reducing Model Inputs section) once more, this time only

using samples of emerging conflict in order to see whether the model prioritized other variables. Table 10 shows that models trained exclusively on emerging conflicts ranked many of the water-related variables (like access to drinking water, value of rainfed agriculture, and precipitation anomalies) higher than for overall conflicts (identical to the "general" column in Table 7). In fact, access to drinking water was ranked as the most important indicator for emerging conflicts. Again, the model does not explain *why* it ranks so highly, and we cannot and do not infer that a lack of drinking water *causes* conflict, only that there is a strong correlation between the two.

Benchmarking the Model

MOMENTUM MODEL

The value of the model can be measured in part according to its marginal utility compared to what already exists. It is generally accepted, from both the academic and the common-sense perspective, that the presence or absence of conflict in the recent past is a key determinant and predictor of the presence or absence of conflict in the near future (Perry 2013). This is in part due to the fact that many of the factors that caused a conflict are often still present after the conflict ends (Collier et al. 2003)—or that the conflict has not ended at all. Simply put, conflict yesterday makes conflict tomorrow more likely.

We can leverage this insight to construct a simple "momentum" model, which always expects the future to match the past: in this case, for any district-month where there was a qualifying conflict event in the preceding 12 months, the model will forecast conflict; where there was not, it will forecast peace.

VIEWS MODEL

We can also benchmark our performance against other machine learning-based forecasts. The ViEWS forecast most resembles our own. However, key differences do exist (see Hegre et al. 2018 for more details on ViEWS model):

- **Coverage:** ViEWS forecasts conflict in Africa, while our current model generates forecasts for Africa, Asia, and the Middle East.
- **Resolution:** ViEWS provides forecasts at the country and gridded level (55 kilometers x 55 kilometers at the equator). Our model forecasts at the district level.
- **Underlying conflict data:** The ViEWS dependent variable is based on Uppsala Conflict Data Program (UCDP) data, while ours is derived from ACLED.
- **Definition of conflict:** ViEWS creates separate forecasts for three different conflict types: state-based, one-sided, and non-state conflict. None perfectly align with our definition of organized violence (see the Dependent Variable section), but one-sided conflict (deliberate and direct targeting of unarmed civilians) and non-state conflict (e.g., inter-communal violence, fighting between rebel organizations, fighting between political parties) together are most similar.
- **Model:** We both use an ensemble method: ViEWS uses Bayesian model averaging (BMA), and we use random forest.

BENCHMARKING AGAINST MOMENTUM

When constructed as described—most closely analogous to our machine learning modeling—and applied to the same test and validation data as the machine learning model, the momentum model exhibits a recall of about 71 percent (compared to an 86 percent recall in the WPS model, as shown in Table 11). Shortening the forecast horizon increases accuracy, as does shortening the lookback period.

To some extent, there is less to this than meets the eye. For starters, performance is essentially a function of the average duration of periods of sustained conflict or peace. Longer lookback periods can smooth this effect somewhat, but at the cost of decreased predictive power. Performance may vary widely based on the frequency of transitions between conflict and peace.

More importantly, the momentum model does not and cannot ever forecast the end of a conflict, nor the outbreak of conflict in a currently peaceful area—and these are the exact scenarios most critical to an early warning system. To predict more conflict in a war-torn province or more peace in a stable, wealthy democracy is largely useless. To forecast the *outbreak* of conflict, however, opens up vital opportunities for intervention and risk reduction.

BENCHMARKING AGAINST VIEWS

Despite the multitude of differences between the WPS model and ViEWS, Table 11 suggests performance to be roughly comparable. There are differing strengths and weakness between the two: ViEWS has a better precision, so it is likely that their conflict labels are more likely to be right; WPS has better recall, so it is less likely to miss conflicts. What is unknown is how well ViEWS forecasts the emergence of new conflicts. This gap may be filled sometime soon, as Hegre et al. (2018). discuss their interest in

Table 10 | **Predictor Variables Ranked by RFE Importance for Overall vs. Emerging Conflicts**

Color scale denotes more important conflicts

1	2	3	4	5	6	7	8	9	10
MOST IMPORTANT									LEAST IMPORTANT

	DATA TESTED IN MODEL	OVERALL CONFLICTS	EMERGING CONFLICTS
GOVERNANCE	Access to drinking water	36	1
CONFLICT	Violent protests and riots (fatalities)	57	2
CONFLICT	Violent protests and riots (count)	54	3
CONFLICT	Violent protests and riots (binary)	67	4
CONFLICT	Violence against civilians (fatalities)	10	5
CONFLICT	Violence against civilians (count)	3	6
CONFLICT	Violence against civilians (binary)	49	7
FOOD	Value of rainfed crops	12	8
FOOD	Value of irrigated crops	31	9
FOOD	Value of all crops	21	10
GOVERNANCE	Executive corruption index	55	11
GOVERNANCE	Political corruption index	48	12
COMMUNITY	Urbanization rate	27	13
COMMUNITY	Urban population	28	14
WATER	Standard precipitation index 6-month anomalies	32	15
WATER	Standard precipitation index 3-month anomalies	41	16
WATER	Standard precipitation index 24-month anomalies	12	17
WATER	Standard precipitation index 12-month anomalies	25	18
ECONOMY	Unemployment, total (% of total labor force)	36	19
ECONOMY	Employment in agriculture (% of total employment)	16	20
ECONOMY	Poverty headcount at $5.50/day	23	21
ECONOMY	Poverty headcount at $3.20/day	29	22
ECONOMY	Poverty headcount at $1.90/day	40	23
COMMUNITY	Child malnutrition	20	24
COMMUNITY	Percentage of population aged 65+ that are male	13	25
COMMUNITY	Percentage of population aged 25–64 that are male	10	26
COMMUNITY	Percentage of population aged 15–24 that are male	28	27
COMMUNITY	Percentage of population aged 0–14 that are male	51	28
WATER	Seasonal variability	6	29
GOVERNANCE	Access to sanitation	3	30
COMMUNITY	Rural to urban ratio	30	31
COMMUNITY	Rural population	9	32
WATER	Riverine flood risk	15	33
CONFLICT	Remote violence (fatalities)	46	34
CONFLICT	Remote violence (count)	18	35
CONFLICT	Remote violence (binary)	59	36
FOOD	Percentage of crops that are rainfed	29	37
CONFLICT	Peaceful protests (fatalities)	75	38
CONFLICT	Peaceful protests (count)	42	39
CONFLICT	Peaceful protests (binary)	76	40
ECONOMY	Gross domestic product per capita	5	41
FOOD	Agriculture value added to GDP (% of total GDP)	11	42
COMMUNITY	National population	38	43
COMMUNITY	Local population density	3	44
COMMUNITY	Local population count	1	45
WATER	Interannual variability	12	46

Table 10 | **Predictor Variables Ranked by RFE Importance for Overall vs. Emerging Conflicts (Cont'd)**

	DATA TESTED IN MODEL	OVERALL CONFLICTS	EMERGING CONFLICTS
GOVERNANCE	REIGN: Warlordism	68	47
GOVERNANCE	REIGN: Party provisional military	65	48
GOVERNANCE	REIGN: Party provisional civilian	65	49
GOVERNANCE	REIGN: Party presidential democracy	60	50
GOVERNANCE	REIGN: Party personal dictatorship	63	51
GOVERNANCE	REIGN: Party personal military hybrid	71	52
GOVERNANCE	REIGN: Party personal	70	53
GOVERNANCE	REIGN: Party military	70	54
GOVERNANCE	REIGN: Parliamentary democracy	72	55
GOVERNANCE	REIGN: Oligarchy	80	56
GOVERNANCE	REIGN: Monarchy	74	57
GOVERNANCE	REIGN: Military personal	67	58
GOVERNANCE	REIGN: Military	57	59
GOVERNANCE	REIGN: Indirect military	78	60
GOVERNANCE	REIGN: Foreign occupied	77	61
GOVERNANCE	REIGN: Dominant party	60	62
ECONOMY	Inflation, consumer prices (annual %)	54	63
WATER	Evapotranspiration 12-month anomalies	16	64
WATER	Evapotranspiration 1-month anomalies	43	65
WATER	Actual evapotranspiration	31	66
GOVERNANCE	Recent election	59	67
WATER	Baseline water stress	22	68
GOVERNANCE	Corruption index	31	69
CONFLICT	Battles (fatalities)	6	70
CONFLICT	Battles (count)	2	71
CONFLICT	Battles (binary)	17	72
COMMUNITY	Religious fractionalization	45	73
COMMUNITY	Ethnic fractionalization	46	74
CONFLICT	Agreements (fatalities)	79	75
CONFLICT	Agreements (count)	63	76
CONFLICT	Agreements (binary)	81	77
COMMUNITY	Percentage of males that are aged 65+	8	78
COMMUNITY	Percentage of population male ages 24-64	34	79
COMMUNITY	Percentage of males that are aged 15–24	48	80
COMMUNITY	Percentage of males that are aged 0–14	41	81

Note: The rankings are sorted by the RFE results for emerging conflicts.
Source: Authors.

Table 11 | **Benchmarking the WPS Model**

		WPS MODEL	BASELINE	ViEWS (NON-STATE)		ViEWS (ONE-SIDED)	
		ALL REGIONS	ALL REGIONS	AFRICA		AFRICA	
		DISTRICT	DISTRICT	COUNTRY	GRID	COUNTRY	GRID
		RANDOM FOREST	MOMENTUM	BMA	BMA	BMA	BMA
Overall	Recall	0.86	0.71	0.59	0.00	0.67	0.04
	Precision	0.47	0.73	0.63	--	0.71	0.84
	F2	0.74	0.71	0.60	--	0.68	0.05
	ROC AUC	0.89	0.84	0.92	0.89	0.92	0.95
	AUPR	0.42	0.55	0.68	0.05	0.79	0.20
	Brier	0.084	0.057	0.062	0.004	0.076	0.005
Emerging conflict	Recall	0.60	0	--	--	--	--
	Precision	0.20	0	--	--	--	--
	F2	0.43	--	--	--	--	--

Note: We compare our results to a baseline "momentum" model, and to another machine learning–based model (ViEWS). Three additional performance metrics are provided as these are the metrics of choice for ViEWS. These are: Area under the Receiver Operating Characteristic (ROC AUC),[15] Area Under the Precision-Recall Curve (AUPR),[16] and the Brier score.[17] They are not analyzed in this report. ViEWS data from Hegre et al. 2018. ViEWS does not provide recall, precision, or F2 scores. We calculated those using the confusion matrices provided in Hegre et al.'s appendix (2018), assuming that predictions over 0.5 were of conflicts. We chose results for the non-state and one-sided conflicts as these were closest in content to our own forecast. The confusion matrix for grid-level non-state forecasts did not provide information for prediction probability over 25 percent, thus the "not available" placeholder.
Source: Authors.

a similar type of presentation in the future.We encourage other conflict forecast researchers to distinguish between emerging and ongoing conflicts so potential users of the information can be fully aware of the strengths and weaknesses of a forecast before they use it to inform a decision.

LIMITATIONS

Unit of Analysis

As described above, we utilize the district-month as our unit of analysis. This sets a hard limit on the precision of our model's predictions: it cannot specify a location beyond a second-level administrative district, or in any way look closer within or between those boundaries. The model's outputs are strictly tied to the unit of analysis.

This is relevant to the predictor variables as well. Each input dataset is necessarily distilled to a single figure per district-month. If the data are in fact collected at a much finer resolution, or with a frequency much greater than monthly, significant information may be destroyed in the course of this numeric aggregation—for example, if we take the average value of every grid cell in a district, then a district containing both very high and very low values may appear quite similar to one containing only medium values. It is likely that this reduces the sensitivity of the resulting model, relative to a system somehow able to truly use all the data.

In addition, the precise boundaries of these higher-level administrative units are not always accurate or consistent globally. If a conflict event's geographic coordinates fall right outside the border of a unit as defined by a geospatial file, that event will be matched to a different administrative unit. In addition, these units are treated atomically, but, in reality, conflicts and instability may flow across borders, especially within a country.

Subnational Event Data

ACLED's organized violence data was selected for many reasons, including its inclusion of non-fatal organized violence, procedural transparency, and growing geographic coverage. Still, we acknowledge that there are limitations associated with subnational event data such as the quality of the location assigned to each event. For example, a study found that ACLED had trouble distinguishing villages with the same name (Eck 2012). As a result, it

is possible that events are placed in the wrong district. This is particularly pronounced in rural areas, as media coverage in rural villages is often sparse. Conflict events in rural areas far from a major city are known to be poorly geocoded (Weidmann 2015).

Correlation, Not Causation

As is often the case with machine learning, our model represents and captures only correlations. Notions of causation lie strictly beyond the scope of our quantitative analysis, as encapsulated by the model. While inspection of model inputs and outputs in conjunction with human expertise may yield subjective insights into the nature or causes of a given instance (or absence) of conflict, the model itself does not and cannot make any claims on these issues. This abstract point becomes clear when considering the model inputs themselves: an indicator like access to drinking water clearly does not *cause* conflict; at best it is a proxy for some underlying dynamic that impacts water and conflict alike.

For these reasons, predictive performance should not be the standard of empirical assessment used to understand causation (Cederman and Weidmann 2017). The WPS partnership, acknowledging that the model cannot give rigorous answers to questions of causation, will use the model as a hot spot detection tool to identify high-risk areas. Then, using local tools, knowledge, and relationships, WPS will examine the underlying causes of conflicts in these high-risk areas and propose interventions.

Precision

Due to the high uncertainty of the emerging conflict forecasts, users should approach the results as a screening tool rather than an alarm bell. That is, users should use these forecasts to prioritize locations for closer inspection before allocating resources.

NEXT STEPS

While this technical note describes the first iteration of our conflict forecasting model, we plan to release additional, improved versions of the model and analysis. Primarily, we hope to test alternative model structures like neural networks. In addition, we will use statistical methods to assess casual relationships between the predictor variables and conflict. We also plan to test additional datasets for inclusion in the model, as well as alternative data processing methods. For example, while using RFE to select features, we noticed that variables encapsulating longer periods of time were consistently identified as more predictive than shorter-term alternatives: 24-month and 12-month Standard Precipitation Index (SPI) figures were more useful than the six-, three-, or one-month variations, for example. We, therefore, plan to introduce water datasets incorporating even longer histories, such as a 48-month SPI. We also hope to utilize indicators that are currently available, but difficult to make amenable to machine learning ingestion, such as a novel dataset of reservoir surface areas over time, or rainfall forecasts.

While not included in the UNDP's list of human security themes, researchers have found energy security to be a driver of conflict. Energy price spikes have led to protests and violent events within and between states (Varigonda 2013), while significant dependence on importing and exporting energy increases vulnerability (Månsson et al. 2014). We, therefore, plan to add additional energy-related datasets in order to test their predictive value in forecasting conflict and better understand the water-food-energy nexus.

Finally, we also plan to investigate the relationship between our predictor variables and the *severity* of conflict (i.e., number of fatalities). Our analysis suggests that water had a much more significant impact when the threshold for "qualifying" violent conflict was set below 10 fatalities. These results cannot be used to establish water's role as a driver of conflict. However, we could speculate that macro socioeconomic factors like GDP, urbanization, and past conflict may create a stronger signal for larger conflicts, while more biophysical indicators like water and irrigation type present a stronger signal for smaller scale violence (this may also reflect a distinction between transboundary and intercommunal conflicts). We hope these modeling efforts can help us improve our model's predictive power and establish causal relationships.

APPENDIX A

Figure A1 | **Feature Importance for the Predictor Variables**

Note: The overall normalized feature importance score is shown in yellow, and the standard deviation of the feature importance is shown as a blue line. Values below 0.01044 were removed from this graph.
Source: Authors. Source code from sci-kit learn documentation, Pedregosa et al. 2011.

Table A1 | **Additional Measures of Feature Importance for the Predictor Variables**

	DATA TESTED IN MODEL	RFE BEST SCORE	NORMALIZED FEATURE IMPORTANCE	PEARSON'S CORRELATION COEFFICIENT (ABSOLUTE VALUE)
COMMUNITY	Ethnic fractionalization	46.0	0.01	0.32
	Religious fractionalization	45.0	0.01	0.16
	Child malnutrition	20.0	0.01	0.29
	Percentage of population males age 0-14	41.0	0.01	0.31
	Percentage of population male ages 15-24	48.0	0.01	0.05
	Percentage of population male ages 25-64	34.0	0.01	0.30
	Percentage of population male ages 65+	8.0	0.01	0.27
	Ages 0-14 that are male	51.0	0.01	0.07
	Ages 15-24 that are male	28.0	0.01	0.05
	Ages 25-64 that are male	10.0	0.02	0.02
	Ages 65+ that are male	13.0	0.01	0.08
	National population	38.0	0.01	0.08
	Local population count	1.0	0.03	0.19
	Local population density	3.0	0.02	0.01
	Rural to urban ratio	30.0	0.01	0.09
	Rural population	9.0	0.02	0.18
	Urban population	28.0	0.01	0.12
	Urbanization rate	27.0	0.01	0.04
CONFLICT	Battles (binary)	17.0	0.04	0.47
	Battles (count)	2.0	0.10	0.23
	Battles (fatalities)	6.0	0.08	0.23
	Remote violence (binary)	59.0	0.00	0.26
	Remote violence (count)	18.0	0.02	0.15
	Remote violence (fatalities)	46.0	0.01	0.10
	Violence against civilians (binary)	49.0	0.01	0.37
	Violence against civilians (count)	3.0	0.08	0.24
	Violence against civilians (fatalities)	10.0	0.06	0.16
	Violent protests and riots (binary)	67.0	0.00	0.11
	Violent protests and riots (count)	54.0	0.00	0.15
	Violent protests and riots (fatalities)	57.0	0.00	0.05
	Agreements (binary)	81.0	0.00	0
	Agreements (count)	63.0	0.00	0.15
	Agreements (fatalities)	79.0	0.00	0
	Peaceful protests (binary)	76.0	0.00	0.01
	Peaceful protests (count)	42.0	0.01	0.10
	Peaceful protests (fatalities)	75.0	0.00	0.03
ECONOMY	Gross domestic product per capita	5.0	0.02	0.36
	Inflation, consumer prices (Annual %)	54.0	0.01	0.13
	Poverty headcount at $1.90/day	40.0	0.01	0.22
	Poverty headcount at $3.20/day	29.0	0.01	0.30
	Poverty headcount at $5.50/day	23.0	0.01	0.35
	Unemployment, total (% of total labor force)	36.0	0.01	0.12
	Employment in agriculture (% of total employment)	16.0	0.01	0.31

Table A1 | **Additional Measures of Feature Importance for the Predictor Variables (Cont'd)**

	DATA TESTED IN MODEL	RFE BEST SCORE	NORMALIZED FEATURE IMPORTANCE	PEARSON'S CORRELATION COEFFICIENT (ABSOLUTE VALUE)
FOOD	Agriculture value added to GDP (% of total GDP)	11.0	0.02	0.32
	Value of irrigated crops	31.0	0.01	0.08
	Value of rainfed crops	12.0	0.02	0.11
	Value of all crops	21.0	0.02	0.11
	Percentage of crops that are rainfed	29.0	0.01	0.01
GOVERNANCE	Access to drinking water	36.0	0.01	0.30
	Access to sanitation	3.0	0.02	0.34
	Corruption index	31.0	0.01	0.27
	Political corruption index	48.0	0.01	0.14
	Executive corruption index	55.0	0.01	0.05
	REIGN: Dominant Party	60.0	0.00	0.10
	REIGN: Foreign Occupied	77.0	0.00	0
	REIGN: Indirect Military	78.0	0.00	0
	REIGN: Military	57.0	0.00	0.33
	REIGN: Military Personal	67.0	0.00	0.01
	REIGN: Monarchy	74.0	0.00	0.07
	REIGN: Oligarchy	80.0	0.00	0
	REIGN: Parlimentary Democracy	72.0	0.00	0.15
	REIGN: Party Military	70.0	0.00	0.06
	REIGN: Party Personal	70.0	0.00	0.03
	REIGN: Party Personal dictatorship	63.0	0.00	0.16
	REIGN: Party Personal Military Hybring	71.0	0.00	0.10
	REIGN: Party Presidential Democracy	60.0	0.00	0.08
	REIGN: Party Provisional Civilian	65.0	0.00	0.14
	REIGN: Party Provisional Military	65.0	0.00	0.03
	REIGN: Warlordism	68.0	0.00	0.10
	Recent election	59.0	0.00	0.00
WATER	Standard Precipitation Index 3-month anomalies	41.0	0.01	0.03
	Standard Precipitation Index 6-month anomalies	32.0	0.01	0.05
	Standard Precipitation Index 12-month anomalies	25.0	0.01	0.02
	Standard Precipitation Index 24-month anomalies	12.0	0.01	0.02
	Actual Evapotranspiration	31.0	0.01	0.02
	Evapotranspiration 1-month anomalies	43.0	0.01	0.19
	Evapotranspiration 12-month anomalies	16.0	0.02	0.19
	Baseline Water Stress	22.0	0.02	0.09
	Riverine Flood Risk	15.0	0.02	0.25
	Seasonal Variability	6.0	0.02	0.11
	InterAnnual Variability	12.0	0.02	0.02

A.1 Additional Measures of Feature Importance

The Recursive Feature Elimination (RFE) approach used to select variables in the analysis is just one of many ways to assess the relative importance of the predictor variables on the forecast. Another approach, feature importance, examines how much each variable decreases the accuracy of the forecast when removed. Variables that have little impact on accuracy have low feature importance and are considered unimportant to model performance. Figure A1 shows the feature importance for the predictor variable (scores are normalized so that the sum of the scores is one). The results suggest that only the first few variables, all related to historic conflict, are informative. Figure A1 also shows the standard deviation of feature importance across the trees (i.e., trees in a forest may find different variables important). There is much more variability in importance across trees, although the historic conflict data remains most informative.

While not used to make decisions in this analysis, the correlation coefficient can be useful for understanding the statistical relationship among the predictor variables and the dependent variable. Pearson's Correlation Coefficient is shown in Table A1. Results are compared to the RFE and feature importance.

A.2 Processing District Boundaries

	GADM ADMINISTRATIVE LEVEL 2 VERSION 3.6
Source organization	GADM
Time covered in dataset	2018
Original resolution	Point
URL	https://gadm.org/data.html
Citation	GADM 2018

Processing Steps

Administrative Level 2 (hereafter *district*) data was downloaded from GADM, along with Level 1 (hereafter *province*) and Level 0 (hereafter *country*). Some countries (commonly island nations) lack provincial and/or district boundaries and, therefore, were missing from the Level 2 database. We filled in gaps by merging missing polygons from the Level 1 and Level 0 databases. We then created unique identifiers for district and provincial fields using the same format that GADM uses.

A.3 Processing the Dependent Variable

	VIOLENT EVENTS
Source organization	ACLED
Time covered in dataset	1997–current (country-specific)
Original resolution	Point
URL	https://www.acleddata.com/data/
Citation	Raleigh et al. 2010

Processing Steps

First, every event in the catalog was matched to a district using its coordinates. Next, the number of events per event type (battles, violent riots and protests, violence against civilians, remote violence) were summed per district/month. This was repeated to find the number of fatalities per event type per district/month. Finally, a rolling window was created to capture the summation of the next 12 months of data per district/month—i.e., number of events and number of fatalities over the next 12 months, per event type. For example, a forecast run in October 2018 would use the event data associated with October 2018–September 2019. Any district/month with 10 fatalities or more in its rolling window was deemed a "conflict" and given a binary score of one. Places that did not meet this threshold were given a binary score of zero.

A.4 Processing the Predictor Variables

A.4.1 Community

CHILD MALNUTRITION

	PREVALENCE OF STUNTING, HEIGHT FOR AGE (% OF CHILDREN UNDER 5)
Source organization	JME. Accessed via the World Bank
Time covered in dataset	1990–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/SH.STA.STNT.ZS
Citation	WHO 2019

Processing steps

First, data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Next, we filled in missing data using a regional/income-level estimation. This was created by grouping countries by their region and income level as defined by the World Bank. The data were then averaged per regional income level per year and used to replace missing data.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the child malnutrition data associated with October 2017.

URBANIZATION RATE

	AVERAGE ANNUAL RATE OF CHANGE OF THE PERCENTAGE URBAN BY COUNTRY
Source organization	UN DESA/Population Division
Time covered in dataset	1950–2100
Original resolution	National, 5-year increments
Time lag	0 months
URL	https://population.un.org/wup/
Citation	UN DESA, Population Division 2019

Processing steps

All districts were matched to the national-level data using country codes. Next, we interpolated between five-year increments to generate annual estimates. All months within a year were assigned the year's score.

No lag was necessary for these data because they have modeled results through 2100.

PERCENTAGE OF MALE POPULATION BY AGE GROUP

	INTERPOLATED POPULATION BY BROAD AGE GROUPS (MALE)
Source organization	UN DESA/Population Division
Time covered in dataset	1950–2100
Original resolution	National, annual
Time lag	0 months
URL	https://population.un.org/wpp/Download/Standard/Interpolated/
Citation	UN DESA, Population Division 2017

Processing steps

Data for 2015 and earlier were pulled from the "Estimates" tab; projected data for after 2015 were pulled from the "Medium variant" tab. All districts were matched to the national-level data using country codes. All months within a year were assigned the year's score. Finally, age ratio was calculated using the following formula:

$$r_{age} = \frac{male_{age}}{male_{tot}}$$

Where:

r_{age} = *ratio of males in broad age group to total male population*

$male_{age}$ = *number of men in broad age group (i.e., ages 25–64)*

$male_{tot}$ = *total number of men in population (i.e., all ages)*

No lag was necessary for these data because they have modeled results through 2100.

MALE SEX RATIO BY AGE GROUP

	INTERPOLATED POPULATION BY BROAD AGE GROUPS (TOTAL)
Source organization	UN DESA/ Population Division
Time covered in dataset	1950–2100
Original resolution	National, 5-year increments
Time lag	0 months
URL	https://population.un.org/wpp/Download/Standard/Interpolated/
Citation	UN DESA, Population Division 2017

Processing steps

Data for 2015 and earlier were pulled from the "Estimates" tab; projected data for after 2015 were pulled from the "Medium variant" tab. All age groups provided by the UN were used. All districts were matched to the national-level data using country codes. Next, we interpolated between five-year increments to generate annual estimates. All months within a year were assigned the year's score. Finally, the male-to-female ratio provided by the UN was converted into the percentage of males per age group using the following formula:

$$r_{sex} = \frac{male_{age}}{pop_{age}}$$

Where:

r_{sex} = *percentage of males in broad age group*

$male_{age}$ = *number of men in broad age group (i.e., ages 25–64)*

pop_{age} = *total number of people in broad age group (i.e., ages 25–64)*

No lag was necessary for these data because they have modeled results through 2100.

NATIONAL POPULATION

	ANNUAL NATIONAL POPULATION
Source organization	UN DESA/Population Division
Time covered in dataset	1950–2100
Original resolution	National, annual
Time lag	0 months
URL	https://population.un.org/wpp/Download/Standard/Population/
Citation	UN DESA Population Division 2017

Processing steps

Data for 2015 and earlier were pulled from the "Estimates" tab; projected data for after 2015 were pulled from the "Medium variant" tab. All districts within a country were assigned the country's score; all months within a year were assigned the year's score.

No lag was necessary for these data because they have modeled results through 2100.

LOCAL POPULATION COUNT

	LOCAL POPULATION COUNT
Source organization	Center for International Earth Science Information Network (CIESIN)
Time covered in dataset	2000–2020
Original resolution	30 arcsecond, 5-year increments
Time lag	0 months
URL	https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-adjusted-to-2015-un-wpp-country-totals-rev11
Citation	CIESIN 2016a

Processing steps

First, the total population per district was found by summing the gridded data within each boundary. This was repeated for all years (i.e., 2000, 2005, 2010, 2015, 2020). Next, we interpolated between the five-year increments to generate annual estimates. All months within a year were assigned the year's score.

No lag was necessary for these data because they have modeled results through 2020.

LOCAL POPULATION DENSITY

	LOCAL POPULATION DENSITY
Source organization	CIESIN
Time covered in dataset	2000–2020
Original resolution	30 arcsecond, 5-year increments
Time lag	0 months
URL	https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-adjusted-to-2015-un-wpp-country-totals-rev11
Citation	CIESIN 2016b

Processing steps

First, the population density per district was found by averaging the gridded data within each boundary. This was repeated for all years (i.e., 2000, 2005, 2010, 2015, 2020). Next, we interpolated between the 5-year increments to generate annual estimates. All months within a year were assigned the year's score.

No lag was necessary for these data because they have modeled results through 2020.

URBAN/RURAL POPULATION

	URBAN EXTENT
Source organization	PBL Netherlands Environmental Assessment Agency
Time covered in dataset	2010
Original resolution	30 arcseconds, 2010
Time lag	0 months
URL	https://www.pbl.nl/en/publications/towards-an-urban-preview
Citation	PBL 2018

	GRIDDED POPULATION
Source organization	PBL
Time covered in dataset	2010
Original resolution	30 arcseconds, 2010
Time lag	0 months
URL	http://www.sciencedirect.com/science/article/pii/S0959378006000501
Citation	van Vuuren et al. 2007

Processing steps

First, the gridded population data were classified into rural and urban categories using the gridded urban extent data. Next, the number of rural and urban citizens were found by summing the classified data within each district boundary. Finally, the percentage of rural population within each district was calculated using the following formula:

$$p = \frac{r}{(r+u)}$$

Where:

r = rural population (-)

u = urban population (-)

p = percentage rural from total population (%)

The results were applied ubiquitously through time due to the lack of temporal resolution.

Note: we do capture urbanization through time using the "Urbanization Rate," but this is only available at the country level.

FRACTIONALIZATION

	ETHNIC AND RELIGIOUS COMPOSITION
Source organization	Cline Center
Time covered in dataset	1945–2013
Original resolution	National, annual
Time lag	24 months
URL	https://clinecenter.illinois.edu/project/Religious-Ethnic-Identity/composition-religious-and-ethnic-groups-creg-project
Citation	Cline Center for Democracy 2012

Processing steps

Fractionalization rates are "HCSS in-house" calculations based on the Cline Center's Composition of Religious and Ethnic Groups Project, which provided the demographic proportions of ethnic groups in countries. Ethnic fractionalization was calculated based on these data using the following formula:

$$FRACT_j = 1 - \sum_{i=1}^{N} s_{ij}^2$$

Where:

s_{ij} = *share of group i (i = 1...N) in country j*

Conceptually, it refers to "the probability that two randomly selected individuals belong to two different ethnic groups" (Alesina et al. 2003).

Data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Missing data were replaced using a forward fill method—results from the more recent year with data were propagated forward. When the data were missing for a particular country for all the years, the original NSD data were used (Alesina et al. 2003).

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the fractionalization data associated with October 2017.

A.4.2 Conflict

SUBNATIONAL EVENT DATA

	HISTORY OF VIOLENCE AND PEACE
Source organization	ACLED
Time covered in dataset	1997–current (country-specific)
Original resolution	Point
Time lag	0 months
URL	https://www.acleddata.com/data/
Citation	Raleigh et al. 2010

Processing steps

First, every event in the catalog was matched to a district using its coordinates. Next, the number of events per event type (agreements, battles, peaceful protests, violent riots and protests, violence against civilians, remote violence) were summed per district/month. This was repeated to find the number of fatalities per event type per district/month. Finally, a rolling window was created to capture the summation of the past 12 months of data per district/month—i.e., number of events and number of fatalities in the past 12 months, per event type. For example, a forecast run in October 2019 would use the event data associated with October 2018–September 2019. Any district/month with 10 fatalities or more in its rolling window was deemed a "conflict" and given a binary score of one. Places that did not meet this threshold were given a binary score of zero.

A.4.3 Economy

GROSS DOMESTIC PRODUCT PER CAPITA

	GDP PER CAPITA, PPP (CONSTANT 2011 INTERNATIONAL $)
Source organization	World Bank, International Comparison Program database. Accessed via the World Bank.
Time covered in dataset	1990–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD
Citation	World Bank Group 2017

Processing steps

First, data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Next, we filled in missing data using a regional/income-level estimation. This was created by grouping countries by their region and income level as defined by the World Bank. The data were then averaged per regional income level per year and used to replace missing data.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the GDP per capita data associated with October 2017.

NATIONAL INFANT MORTALITY

	MORTALITY RATE, INFANT (PER 1,000 LIVE BIRTHS)
Source organization	UN Inter-agency Group for Child Mortality Estimation (UNICEF, WHO, World Bank, UN DESA Population Division).
Time covered in dataset	1990–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/SP.DYN.IMRT.IN
Citation	(UN IGME 2019)

Processing steps

Data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Missing data were replaced using a backward fill method—results from the more recent year with data were propagated backward.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the mortality rate data associated with October 2017.

INFLATION

	INFLATION, CONSUMER PRICES (ANNUAL %)
Source organization	International Monetary Fund, International Financial Statistics. Accessed via the World Bank.
Time covered in dataset	1981–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG
Citation	IMF 2019

Processing steps

First, data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Next, we filled in missing data using a regional/income-level estimation. This was created by grouping countries by their region and income level as defined by the World Bank. The data were then averaged per regional income level per year and used to replace missing data.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the inflation data associated with October 2017.

UNEMPLOYMENT

	UNEMPLOYMENT, TOTAL (% OF TOTAL LABOR FORCE)
Source organization	International Labour Organization, ILOSTAT database. Accessed via the World Bank.
Time covered in dataset	1991–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS
Citation	International Labour Office 2017

Processing steps

First, data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Next, we filled in missing data using a regional/income-level estimation. This was created by grouping countries by their region and income level as defined by the World Bank. The data were then averaged per regional income level per year and used to replace missing data.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the unemployment data associated with October 2017.

POVERTY HEADCOUNT

	POVERTY HEADCOUNT RATIO AT $1.90/DAY (2011 PPP; % OF POPULATION)
Source organization	World Bank, Development Research Group. Accessed via the World Bank.
Time covered in dataset	1977–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/SI.POV.DDAY
Citation	World Bank Group 2019

	POVERTY HEADCOUNT RATIO AT $3.20/DAY (2011 PPP; % OF POPULATION)
Source organization	World Bank, Development Research Group. Accessed via the World Bank.
Time covered in dataset	1977–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/SI.POV.LMIC
Citation	World Bank Group 2019

	POVERTY HEADCOUNT RATIO AT $5.50/DAY (2011 PPP; % OF POPULATION)
Source organization	World Bank, Development Research Group. Accessed via the World Bank.
Time covered in dataset	1977–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/SI.POV.UMIC
Citation	World Bank Group 2019

Processing steps

First, data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Next, we filled in missing data using a regional/income-level estimation. This was created by grouping countries by their region and income level as defined by the World Bank. The data were then averaged per regional income level per year and used to replace missing data.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the poverty data associated with October 2017.

A.4.4 Food

EMPLOYMENT IN AGRICULTURE

	EMPLOYMENT IN AGRICULTURE (% OF TOTAL EMPLOYMENT)
Source organization	International Labour Organization, ILOSTAT database. Accessed via the World Bank.
Time covered in dataset	1991–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS
Citation	International Labour Office 2017

Processing steps

First, data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Next, we filled in missing data using a regional/income-level estimation. This was created by grouping countries by their region and income level as defined by the World Bank. The data were then averaged per regional income level per year and used to replace missing data.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the employment data associated with October 2017.

AGRICULTURE VALUE ADDED TO GDP

	AGRICULTURE, FORESTRY, AND FISHING, VALUE ADDED (% OF GDP)
Source organization	World Bank and OECD. Accessed via the World Bank.
Time covered in dataset	1960–2018
Original resolution	National, annual
Time lag	24 months
URL	https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS
Citation	World Bank Group 2016

Processing steps

First, data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Next, we filled in missing data using a regional/income-level estimation. This was created by grouping countries by their region and income level as defined by the World Bank. The data were then averaged per regional income level per year and used to replace missing data.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the data associated with October 2017.

VALUE OF CROP PRODUCTION PER IRRIGATION METHOD

	VALUE OF CROP PRODUCTION (IRRIGATED, RAINFED, ALL TECHNOLOGIES)
Source organization	International Food Policy Research Institute (IFPRI)
Time covered in dataset	2010
Original resolution	5 arcminutes, 2010
Time lag	0 months
URL	http://mapspam.info/data/
Citation	IFPRI 2019

Processing steps

The 2010 SPAM database contains 42 crops. SPAM aggregates the values of all 42 crops into one gridded dataset representing the value of crop production per irrigation method (i.e., irrigated, rainfed, and both irrigated and rainfed methods). For each irrigation method, the total value of crops per district was found by summing the gridded data within each boundary.

The results were applied ubiquitously through time due to the lack of time series data.

Note: IFPRI does provide SPAM data for 2000 and 2005, but we were unable to certify whether these data are appropriate to use in time series (for example, the year 2000 only featured 20 crops while 2010 featured 42).

PERCENTAGE OF RAINFED CROPS

	CROP PRODUCTION (IRRIGATED, RAINFED, ALL TECHNOLOGIES)
Source organization	IFPRI
Time covered in dataset	2010
Original resolution	5 arcminutes, 2010
Time lag	0 months
URL	http://mapspam.info/data/
Citation	IFPRI 2019

Processing steps

The 2010 SPAM database contains 42 crops. The production (in metric tons) of each crop is available for both irrigated and rainfed methods. In all, we used 84 gridded crop production datasets (42 irrigated, 42 rainfed) to calculate this variable. Each dataset was summed per district boundary. The crops were then aggregated by irrigation method (irrigated vs. rainfed) to find the total crop production per irrigation method per district. Finally, the percentage of rainfed crops was calculated using the following formula:

$$p = \frac{r}{(r+i)}$$

Where:

r = total rainfed crop production (metric tons)

i = total irrigated crop production (metric tons)

p = percentage rainfed from total production (%)

The results were applied ubiquitously through time due to the lack of time series data.

Note: IFPRI does provide SPAM data for 2000 and 2005, but we were unable to certify whether these data are appropriate to use in time series (for example, the year 2000 only featured 20 crops while 2010 feature 42).

A.4.5 Governance

ACCESS TO WASH

	ACCESS TO DRINKING WATER AND ACCESS TO SANITATION
Source organization	JMP
Time covered in dataset	2000–2015
Original resolution	National, annual
Time lag	48 months
URL	https://washdata.org/data/household
Citation	WHO and UNICEF 2017

Processing steps

We used the national "At least basic" rate of access to measure a country's WASH service level. All districts within a country were assigned the country's score; all months within a year were assigned the year's score.

We do not anticipate these data to be ready in near real time. Updates to these data occur once every few years. To account for a delay in data availability, we "lagged" this indicator by 48 months. For example, a forecast run in October 2019 would use the drinking water and sanitation data associated with October 2015.

CORRUPTION

	BAYESIAN CORRUPTION INDEX
Source organization	Ghent University
Time covered in dataset	1984–2017
Original resolution	National, annual
Time lag	24 months
URL	https://users.ugent.be/~sastanda/BCI/BCI.html
Citation	Standaert 2015

	POLITICAL AND EXECUTIVE CORRUPTION
Source organization	Varieties of Democracy Institute, the Department of Political Science at the University of Gothenburg, Sweden
Time covered in dataset	1789–2018
Original resolution	National, annual
Time lag	24 months
URL	https://doi.org/10.23696/vdemcy19
Citation	Coppedge et al. 2019; Pemstein et al. 2018

Processing steps

Data were matched to the analysis's district/month scale. All districts within a country were assigned the country's score; all months within a year were assigned the year's score. Missing data were replaced using a backward fill method—results from the more recent year with data were propagated backward.

We do not anticipate these data to be ready in near real time. To account for a delay in data availability, we "lagged" this indicator by 24 months. For example, a forecast run in October 2019 would use the corruption data associated with October 2017.

TYPE OF GOVERNMENT

	RULERS, ELECTIONS, AND IRREGULAR GOVERNANCE DATA (REIGN)
Source organization	One Earth Future
Time covered in dataset	1950–2018
Original resolution	National, monthly
Time lag	0 months
URL	https://oefresearch.org/datasets/reign
Citation	Bell 2016

Processing steps

The categorical regime type data were converted into binary variables—that is, every type was made into its own field and corresponding countries were assigned a value of "1." All districts within a country were assigned the country's score. Missing data were replaced using a backward fill method—results from the more recent year with data were propagated backward.

REIGN data is available in near real time, so no time lag was necessary.

A.4.6 Water

STANDARD PRECIPITATION INDEX (SPI)

	ECMWF SEAS5 SEASONAL FORECAST DATA
Source organization	European Centre for Medium-Range Weather Forecasts (ECMWF)
Time covered in dataset	2001–present
Original resolution	1 degree, daily
URL	https://www.ecmwf.int/en/newsletter/154/meteorology/ecmwfs-new-long-range-forecasting-system-seas5
Citation	Johnson et al. 2019

Processing steps

Precipitation anomalies for 1-, 3-, 6-, 12- and 24-month rolling windows were created by Deltares using precipitation data from the ECMWF SEAS5 seasonal weather forecasts. Specifically, we used the Standard Precipitation Index approach to create the anomalies. The baseline used ranges from January 2001 to January 2019. The resulting data is provided at the same grid as the source data—a regular grid with a resolution of one degree.

SPI data is available in near real time, so no time lag was necessary.

EVAPOTRANSPIRATION (ET)

	SSEBOP EVAPOTRANSPIRATION PRODUCTS (VERSION 4)
Source organization	USGS; FEWS NET
Time covered in dataset	2003–2019
Original resolution	30 arcseconds, monthly and annually
Time lag	0 months
URL	https://earlywarning.usgs.gov/fews/product/460
Citation	Senay et al. 2011

Processing steps

For evapotranspiration (ET) data, we selected monthly and annual anomalies, as well as monthly actual ET. Gridded data were downloaded for every timestamp in our analysis. To match data to the district level, we found the minimum gridded ET value within each district boundary.

ET data is available in near real time, so no time lag was necessary.

AQUEDUCT WATER RISK INDICATORS

	PROVINCIAL-LEVEL WATER RISK INDICATORS
Source organization	WRI
Time covered in dataset	1960–2014 (baseline)
Original resolution	Provincial, baseline
Time lag	0 months
URL	https://www.wri.org/resources/data-sets/aqueduct-30-country-rankings
Citation	Hofste et al. 2019

Processing steps

Available indicators include Baseline Water Stress, Seasonal Variability, Interannual Variability, and Riverine Flood Risk. All districts within a province were assigned the province's score. The results were applied ubiquitously through time because they represent a long-term, chronic condition.

ENDNOTES

1. Model details: ViEWS uses a Bayesian Model Averaging approach to forecast conflict (state-based, non-state-based, and one-sided conflict) 36 months in advance (Hegre et al. 2018).

2. For a more detailed discussion of supervised learning, please see https://www.sciencedirect.com/topics/computer-science/supervised-learning.

3 Random Forest is a type of ensemble model. ViEWS uses another type called Bayesian Model Averaging.

4. The delineation of administrative units varies from country to country. The zero administrative unit is the country, while the first administrative level corresponds to the subunits of the country, the second administrative level corresponds to the subunits of the first-level administrative unit, and so on. For example, in the United States, the first administrative level is the state, while the second administrative level is (generally) the county.

5. The recall and precision metrics can also be used to measure the performance of "no conflict" predictions.

6. For more information on Python's scikit-learn, please visit: https://scikit-learn.org/stable/.

7. For more information on Google's Colaboratory environment, please visit: https://colab.research.google.com/notebooks/welcome.ipynb.

8. "Overfitting" is when a model comes to fit a particular set of training data very closely, at the expense of failing to fit unseen data well. In other words, an overfit model is one that performs well on training data but fails to generalize.

9. For an explanation of these and other available parameters, please see the documentation: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

10. The dependent variable used to train the model requires 12 months of forward-looking data. For example, the most recent dependent variable that could be used to train a model run in October 2019 would be from October 2018 (i.e., events from October 2018–September 2019).

11. Statistic refers to all available district-months within the geographic study area, shown in Figure 2, from 2004 through 2018. Note that outside of Africa, data were not available for that full time period.

12. Accuracy is calculated as: True / (True + False).

13. For the scikit-learn web page on Recursive Feature Elimination, please see: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html.

14. Nineteen variables achieved a ranking of 15 or lower in the three-run experiment.

15. Hegre et al. (2018) on ROC AUC: "rewards models for increasing detection of actual conflict (true positives) relative to 'false alarms.'"

16. Hegre et al. (2018) on AUPR: "is a relative measure of how precisely a model predicts true positives and the true positive rate."

17. Hegre et al. (2018) on Brier: "measures the accuracy of probabilistic predictions. It favors sharp, accurate probabilistic predictions (near 0 or 1)."

REFERENCES

ACLED (Armed Conflict Location & Event Data Project). 2017. *Armed Conflict Location & Event Data Project (ACLED) Codebook*. https://www.acleddata.com/wp-content/uploads/dlm_uploads/2017/10/ACLED_Codebook_2019FINAL_pbl.pdf.

ACLED. 2019. "ACLED Coverage to Date." ACLED. https://www.acleddata.com/wp-content/uploads/dlm_uploads/2018/02/Country-and-Time-Period-coverage_updOct2019.pdf.

Adelphi. 2019. "The Environment, Conflict, and Cooperation (ECC) Factbook." *Factbook/Map*. https://factbook.ecc-platform.org/conflict-map.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8. http://dx.doi.org/10.1023/A:1024471506938.

Auclair, Allan N.D. 1999. "Population Displacement and Migration in Developing Countries: An Exploratory Study in the Development of an Early Warning Decision Support System." RAND Science and Technology National Defense Research Institute, November. https://www.rand.org/pubs/monograph_reports/MR1264z0.html.

Barnett, Jon, and W. Neil Adger. 2007. "Climate Change, Human Security and Violent Conflict." *Political Geography*, *Climate Change and Conflict* 26 (6): 639–55. doi:10.1016/j.polgeo.2007.03.003.

Bell, Curtis. 2016. *The Rulers, Elections, and Irregular Governance Dataset (REIGN)*. Broomfield, CO: OEF Research. Available at oefresearch.org.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Cederman, Lars-Erik, and Nils Weidmann. 2017. "Predicting Armed Conflict: Time to Adjust Our Expectations?" *Science* 355 (6324): 474–76. doi:10.1126/science.aal4483.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-sampling Technique." *Journal of Artificial Intelligence Research* 16 (June): 321–57. doi:10.1613/jair.953.

CIESIN (Center for International Earth Science Information Network). 2016a. *Gridded Population of the World (GPW), v4: UN-Adjusted Population Count, v4 (2000, 2005, 2010, 2015, 2020)*. Palisades, NY: SEDAC. https://doi.org/10.7927/H4PN93PB.

CIESIN. 2016b. *Gridded Population of the World (GPW), v4: UN-Adjusted Population Density, v4 (2000, 2005, 2010, 2015, 2020)*. Palisades, NY: SEDAC. https://dx.doi.org/10.7927/H4HX19NJ.

Cline Center for Democracy. 2012. *The Composition of Religious and Ethnic Groups (CREG) Project*. Urbana-Champaign: University of Illinois. https://uofi.app.box.com/s/ayrhyzii5rp27lubg19ild2q0bpsxrmg.

Collier, Paul, V.L. Elliott, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis. 2003. *Breaking the Conflict Trap: Civil War and Development Policy*. 26121. The World Bank. http://documents.worldbank.org/curated/en/908361468779415791/Breaking-the-conflict-trap-civil-war-and-development-policy.

Conca, Ken, and Jennifer Wallace. 2009. "Environment and Peacebuilding in War-Torn Societies: Lessons from the UN Environment Programme's Experience with Postconflict Assessment." *Global Governance* 15 (4): 485–504. https://www.jstor.org/stable/27800777.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard et al. 2019. *V-Dem Global Dataset V9*. Varieties of Democracy (V-DEM). doi.org/10.23696/vdemcy19.

ECC Factbook. 2015. "What This Factbook Is All About." *Dashboard*. https://www.ecc-platform.org/factbook/documentation/background.

Eck, Kristine. 2012. "In Data We Trust? A Comparison of UCDP GED and ACLED Conflict Events Datasets." *Cooperation and Conflict* 47 (1): 124–41. doi:10.1177/0010836711434463.

Effendy, Veronikha, Adiwijaya, and Z.K.A. Baizal. 2014. "Handling Imbalanced Data in Customer Churn Prediction Using Combined Sampling and Weighted Random Forest." In *2014 2nd International Conference on Information and Communication Technology (ICoICT)* 325–30. doi:10.1109/ICoICT.2014.6914086.

Farinosi, Fabio, Carlo Giupponi, Arnaud Reynaud, Guido Ceccherini, César Carmona-Moreno, Ad De Roo, D. Gonzalez-Sanchez et al. 2018. "An Innovative Approach to the Assessment of Hydro-political Risk: A Spatially Explicit, Data Driven Indicator of Hydro-political Issues." *Global Environmental Change* 52 (September): 286–313. doi:10.1016/j.gloenvcha.2018.07.001.

GADM. 2018. *GADM Database of Global Administrative Areas 3.6* (version 3.6). Global Administrative Areas. https://gadm.org/download_world.html.

Gleick, Peter, and Charles Iceland. 2018. "Water, Security, and Conflict." World Resources Institute, August. https://www.wri.org/publication/water-security-and-conflict.

Halkia, Stamatia, Stefano Ferri, Ines Joubert-Boitat, Francesca Saporiti, and Mayeul Kauffmann. 2017. *The Global Conflict Risk Index (GCRI): Regression Model, Data Ingestion, Processing and Output Methods*. EUR—Scientific and Technical Research Reports. Publications Office of the European Union. JRC108767. doi:10.2760/303651.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Edition. Springer Series in Statistics. New York: Springer. https://web.stanford.edu/~hastie/ElemStatLearn/.

Hegre, Håvard, Marie Allansson, Matthias Basedau, Michael Colaresi, Mihai Croicu, Hanne Fjelde, Frederick Hoyles et al. 2018. "ViEWS: A Political Violence Early-Warning System." *Journal of Peace Research* 56 (2): 155–74. doi:10.1177/0022343319823860.

Hofste, Rutger, Samantha Kuzma, Paul Reig, Sara Walker, Edwin H. Sutanudjaja, Marc F.P. Bierkens, M. Kuijpers-Linde et al. 2019. *Aqueduct 3.0: Updated Decision-Relevant Global Water Risk Indicators*. Technical Note. Washington, DC: World Resources Institute. https://www.wri.org/publication/aqueduct-30.

Ide, Tobias, and Adrien Detges. 2018. "International Water Cooperation and Environmental Peacemaking." *Global Environmental Politics* 18: 63–84. doi:10.1162/glep_a_00478.

IFPRI (International Food Policy Research Institute). 2019. "Global Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 1.0." Harvard Dataverse. doi:10.7910/DVN/PRFF8V.

International Labour Office. 2017. *ILO Labour Force Estimates and Projections: 1990–2030* (2017 Edition). ILO Department of Statistics. https://www.ilo.org/ilostat-files/Documents/LFEP.pdf.

IMF (International Monetary Fund). 2019. "World Economic Outlook, October 2018: Challenges to Steady Growth." IMF. https://www.imf.org/en/Publications/WEO/Issues/2019/10/01/world-economic-outlook-october-2019.

Johnson, Stephanie J., Timothy N. Stockdale, Laura Ferranti, Magdalena A. Balmaseda, Franco Molteni, Linus Magnusson, Steffen Tietsche et al. 2019. "SEAS5: The New ECMWF Seasonal Forecast System." *Geoscientific Model Development* 12 (3): 1087–1117. doi.org/10.5194/gmd-12-1087-2019.

Keller, Christoph A., and Mat J. Evans. 2019. "Application of Random Forest Regression to the Calculation of Gas-Phase Chemistry within the GEOS-Chem Chemistry Model V10." *Geoscientific Model Development* 12 (3): 1209–25. doi:https://doi.org/10.5194/gmd-12-1209-2019.

Khalilia, Mohammed, Sounak Chakraborty, and Mihail Popescu. 2011. "Predicting Disease Risks from Highly Imbalanced Data Using Random Forest." *BMC Medical Informatics and Decision Making* 11 (1): 51. doi:10.1186/1472-6947-11-51.

Liu, Xu-Ying, Jianxin Wu, and Zhi-Hua Zhou. 2009. "Exploratory Undersampling for Class-Imbalance Learning." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39 (2): 539–50. doi:10.1109/TSMCB.2008.2007853.

Machlis, Gary E., and Thor Hanson. 2008. "Warfare Ecology." *BioScience* 58 (8): 729–36. doi:10.1641/B580809.

Månsson, André, Bengt Johansson, and Lars J. Nilsson. 2014. "Assessing Energy Security: An Overview of Commonly Used Methodologies." *Energy* 73 (August): 1–14. doi:10.1016/j.energy.2014.06.073.

McKee, Thomas, Nolan Doesken, and John Kleist. 1993. "The Relationship of Drought Frequency and Duration to Time Scale." In *Proceedings of the Eighth Conference on Applied Climatology* 179–84. Anaheim, California. https://scinapse.io/papers/2153179024.

Muchlinski, David, David Siroky, Jingrui He, and Matthew Kocher. 2016. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data." *Political Analysis* 24 (1): 87–103. doi:10.1093/pan/mpv024.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106. doi:10.1257/jep.31.2.87.

PBL. 2018. *Towards an Urban Preview: Modelling Future Urban Growth with 2UP*. 3255. The Hague: PBL Netherlands Environmental Assessment Agency. https://www.pbl.nl/en/publications/towards-an-urban-preview.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (October): 2825–30. http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, Joshua Krusell, and Farhad Miri. 2018. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data." SSRN Scholarly Paper ID 3167764. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3167764.

Perry, Chris. 2013. "Machine Learning and Conflict Prediction: A Use Case." *Stability: International Journal of Security and Development* 2 (3): Art. 56. doi:10.5334/sta.cr.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47 (5): 651–60. doi:10.1177/0022343310378914.

Robinson, Caleb, and Bistra Dilkina. 2017. "A Machine Learning Approach to Modeling Human Migration." In *COMPASS*. doi:10.1145/3209811.3209868.

Selby, Jan, Omar S. Dahi, Christiane Fröhlich, and Mike Hulme. 2017. "Climate Change and the Syrian Civil War Revisited." *Political Geography* (September) 60: 232–44. doi:10.1016/j.polgeo.2017.05.007.

Senay, Gabriel B., Michael E. Budde, and James P. Verdin. 2011. "Enhancing the Simplified Surface Energy Balance (SSEB) Approach for Estimating Landscape ET: Validation with the METRIC Model." *Agricultural Water Management* 98 (4): 606–18. doi:10.1016/j.agwat.2010.10.014.

Standaert, Samuel. 2015. "Divining the Level of Corruption: A Bayesian State-Space Approach." *Journal of Comparative Economics* 43 (3): 782–803. doi:10.1016/j.jce.2014.05.007.

Swain, Ashok. 2015. “Water and Post-conflict Peacebuilding.” *Hydrological Sciences Journal*, August, 150814115356005. doi:10.1080/02626667.2015.1081390.

Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* STU-Student edition. Princeton University Press. https://www.jstor.org/stable/j.ctt7spbt.

UN (United Nations). n.d. “Human Security and Agenda 2030.” United Nations Trust Fund for Human Security. https://www.un.org/humansecurity/wp-content/uploads/2017/10/Human-Security-and-the-SDGs.pdf. Accessed on February 3, 2020.

UN IGME (United Nations Inter-agency Group for Child Mortality Estimation). 2019. *Explanatory Notes: Child Mortality Trend Series to 2018*. https://childmortality.org/wp-content/uploads/2019/09/UNIGME-Explanatory-Notes_ENGLISH.pdf.

UN Web TV. 2018. *Arria-Formula Meeting on Water Peace and Security.* New York, NY. http://webtv.un.org/meetings-events/treaty-bodies/watch/arria-formula-meeting-on-water-peace-and-security/5854096417001/?term=http://webtv.un.org/meetings-events/treaty-bodies/watch/children-on-the-move-preventing-child-trafficking-by-implementing-the-un-global-plan-of-action/5588262537001/.

UNDP (United Nations Development Programme), ed. 1994. *Human Development Report 1994*. New York: Oxford University Press. http://hdr.undp.org/en/content/human-development-report-1994.

UN DESA, Population Division (United Nations Department of Economic and Social Affairs, Population Division). 2017. *World Population Prospects: Methodology of the United Nations Population Estimates and Projections, the 2017 Revision*. ESA/P/WP.250. New York: United Nations. https://population.un.org/wpp/Publications/Files/WPP2017_Methodology.pdf.

UN DESA, Population Division. 2019. *World Urbanization Prospects: The 2018 Revision.* New York: United Nations Department of Economic and Social Affairs, Population Division. https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

van Vuuren, Detlef P., Paul L. Lucas, and Henk Hilderink. 2007. “Downscaling Drivers of Global Environmental Change: Enabling Use of Global SRES Scenarios at the National and Grid Levels.” *Global Environmental Change* 17 (1): 114–30. doi:10.1016/j.gloenvcha.2006.04.004.

Varigonda, Kesava Chandra. 2013. “An Assessment of the Impact of Energy Insecurity on State Stability in India.” *Energy Policy* 62 (November): 1110–19. doi:10.1016/j.enpol.2013.06.091.

Weidmann, Nils B. 2015. “On the Accuracy of Media-Based Conflict Event Data.” *Journal of Conflict Resolution* 59 (6): 1129–49. doi:10.1177/0022002714530431.

WHO (World Health Organization) and UNICEF (United Nations Children’s Fund). 2017. *Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines.* Geneva: World Health Organization (WHO) and the United Nations Children’s Fund (UNICEF). https://washdata.org/.

WHO. 2019. “Global Health Observatory Data Repository: Joint Child Malnutrition Estimates (UNICEF-WHO-WB).” http://apps.who.int/gho/data/node.main.ngest?lang=en.

Wolf, Aaron T. 1998. “Conflict and Cooperation along International Waterways.” *Water Policy* 1 (2): 251–65. doi:10.1016/S1366-7017(98)00019-1.

World Bank Group. 2016. “Agriculture, Forestry, and Fishing, Value Added (% of GDP).” https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS.

World Bank Group. 2017. “GDP per capita, PPP (constant 2011 international $).” https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD.

World Bank Group. 2019. “PovcalNet: Methodology.” http://iresearch.worldbank.org/PovcalNet/methodology.aspx.

ACKNOWLEDGMENTS

This publication was made possible thanks to the support of the Netherlands Ministry of Foreign Affairs. The authors would also like to thank New America's Resource Security program for providing additional funds and research in support of this publication.

We are pleased to acknowledge WRI's institutional strategic partners, who provide core funding: the Netherlands Ministry of Foreign Affairs, the Royal Danish Ministry of Foreign Affairs, and the Swedish International Development Cooperation Agency.

The authors would like to thank the following people for providing invaluable insight and assistance in the development of this paper: Susanne Schmeier (IHE-Delft), Betsy Otto (WRI), Meredith Hess (WRI), Laura Malaguzzi Valeri (WRI), and Emilia Suarez (WRI); as well as our internal reviewers John Brandt, Mikaela Weisse, Brookie Guzder-Williams, and Lihuan Zhou, and external reviewers Kefyalew Abera (Mercy Corps), Sharon Burke (New America), and Hans Korving (Deltares).

The authors would also like to thank the WPS research partners at IHE-Delft, Deltares, the Hague Center for Strategic Studies, Wetlands International, and International Alert, as well as our affiliated research partners at the Pacific Institute, New America, and Oregon State University. Finally, we would like to thank Green River, Romain Warnault, Carni Klirs, and Billie Kanfer for their extensive guidance and feedback during the design and development of this publication and website.

Opinions or points of view expressed in this report are those of the authors and do not necessarily reflect the position of the reviewers, research partners, or the organizations they represent.

ABOUT THE AUTHORS

Samantha Kuzma is an associate at the World Resources Institute.
Contact: samantha.kuzma@wri.org.

Peter Kerins is a research analyst at the World Resources Institute.

Elizabeth Saccoccia is a research analyst at the World Resources Institute.

Cayla Whiteside is an environmental engineering master's candidate and Schneider Fellow from Stanford University.

Hannes Roos is a data scientist at The Hague Centre for Strategic Studies.

Charles Iceland is the director of Global and National Water Initiatives at the World Resources Institute.

ABOUT WRI

World Resources Institute is a global research organization that turns big ideas into action at the nexus of environment, economic opportunity, and human well-being.

Our Challenge

Natural resources are at the foundation of economic opportunity and human well-being. But today, we are depleting Earth's resources at rates that are not sustainable, endangering economies and people's lives. People depend on clean water, fertile land, healthy forests, and a stable climate. Livable cities and clean energy are essential for a sustainable planet. We must address these urgent, global challenges this decade.

Our Vision

We envision an equitable and prosperous planet driven by the wise management of natural resources. We aspire to create a world where the actions of government, business, and communities combine to eliminate poverty and sustain the natural environment for all people.

Water, Peace and Security

info@waterpeacesecurity.org
@WaterPeaceSec

BUILT ON

RESOURCEWATCH

IN COLLABORATION WITH

Ministry of Foreign Affairs

WITH ADDITIONAL SUPPORT FROM

NEW AMERICA