LETTER • OPEN ACCESS

Projecting armed conflict risk in Africa towards 2050 along the SSP-RCP scenarios: a machine learning approach

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Projecting armed conflict risk in Africa towards 2050 along the SSP-RCP scenarios: a machine learning approach

Jannis M Hoch[1,*] ⓘ, Sophie P de Bruin[1,2] ⓘ, Halvard Buhaug[3,4] ⓘ, Nina Von Uexkull[3,5] ⓘ, Rens van Beek[1] ⓘ and Niko Wanders[1] ⓘ

1 Department of Physical Geography, Utrecht University, Utrecht, The Netherlands
2 PBL Netherlands Environmental Assessment Agency, The Hague, The Netherlands
3 Peace Research Institute Oslo (PRIO), Oslo, Norway
4 Department of Sociology and Political Science, Norwegian University of Science and Technology, Trondheim, Norway
5 Department of Peace and Conflict Research, Uppsala University, Uppsala, Sweden
* Author to whom any correspondence should be addressed.

E-mail: j.m.hoch@uu.nl

Keywords: conflict risk, climate change, machine learning, water security, scenarios

Abstract

In the past decade, several efforts have been made to project armed conflict risk into the future. This study broadens current approaches by presenting a first-of-its-kind application of machine learning (ML) methods to project sub-national armed conflict risk over the African continent along three Shared Socioeconomic Pathway (SSP) scenarios and three Representative Concentration Pathways towards 2050. Results of the open-source ML framework CoPro are consistent with the underlying socioeconomic storylines of the SSPs, and the resulting out-of-sample armed conflict projections obtained with Random Forest classifiers agree with the patterns observed in comparable studies. In SSP1-RCP2.6, conflict risk is low in most regions although the Horn of Africa and parts of East Africa continue to be conflict-prone. Conflict risk increases in the more adverse SSP3-RCP6.0 scenario, especially in Central Africa and large parts of Western Africa. We specifically assessed the role of hydro-climatic indicators as drivers of armed conflict. Overall, their importance is limited compared to main conflict predictors but results suggest that changing climatic conditions may both increase and decrease conflict risk, depending on the location: in Northern Africa and large parts of Eastern Africa climate change increases projected conflict risk whereas for areas in the West and northern part of the Sahel shifting climatic conditions may reduce conflict risk. With our study being at the forefront of ML applications for conflict risk projections, we identify various challenges for this arising scientific field. A major concern is the limited selection of relevant quantified indicators for the SSPs at present. Nevertheless, ML models such as the one presented here are a viable and scalable way forward in the field of armed conflict risk projections, and can help to inform the policy-making process with respect to climate security.

1. Introduction

Without effective climate change mitigation measures and with continuing human-induced ecological degradation, environmental pressures on livelihoods are expected to worsen in many regions around the world (Adger *et al* 2014, IPCC 2019). A more contested impact of climate change is an increased risk of violent conflict (Hsiang *et al* 2013, Buhaug *et al* 2014, Koubi 2019, Mach *et al* 2019). Political concern as well as scientific and security interests have hence been rising during the last decades. This has resulted

in a maturing body of academic literature on climate-conflict connections (Von Uexkull and Buhaug 2021), also feeding decision-making of intergovernmental institutions, such as the UN Security Council (Scott 2015, Conca 2019).

However, the scientific consensus is still limited regarding the relevance and strength of specific mechanisms linking climate, the environment, and armed conflict risk (Koubi 2019). Recent conclusions differ due to, *inter alia*, the use of different data proxies, timescales, geographical scales as well as definitions of conflict, and the field is further challenged

by concerns about sampling bias in climate-conflict research (Adams *et al* 2018).

Nevertheless, several conditions—including low socioeconomic development and economic shocks, weak governmental capacity, and a recent history of armed conflict—are generally accepted as important contextual risk factors (Mach *et al* 2019). Under these conditions, climatic and environmental drivers are most likely to increase conflict risk (see Buhaug and Von Uexkull (2021) and Mach *et al* (2019) for potential linkages). Already conflict-prone countries, which lack good governance systems and depend on climate-sensitive economic activities such as rain-fed agriculture are found to be the most vulnerable to the adverse effects of climate change (Von Uexkull 2014, Almer *et al* 2017, Otto *et al* 2017).

Gaining more insights into the role of water-related environmental stress for future armed conflict risk is therefore needed. One way to do so is quantitative forecasting. Relevant recent attempts have focused mostly on developing early warning models for armed conflict for a limited time horizon (Hegre *et al* 2017, 2019, 2021, WPS Partnership 2021). For those instruments, accuracy and forecasting skills are paramount. With their prediction horizon, they are suited to inform, for example, short-term policy making and interventions. They are not, however, intended to explore security implications of plausible long-term scenarios aiding capacity building and long-term policy processes.

Incomplete knowledge about the relations between conflict drivers and the lack of sufficient observational data make it challenging to project long-term conflict risk (Cederman and Weidmann 2017). Nevertheless, making projection ensembles without claiming to make absolute and accurate predictions is a viable way towards better estimate uncertainties (Maier *et al* 2016). The main aim of the projections is to assess plausible developments along with alternative scenarios rather than predict the onset of an event. This approach is already successfully adopted in other scientific disciplines such as flood and drought risk projections (Hirabayashi *et al* 2013, Wanders *et al* 2015). The insights obtained from these long-term projections can then facilitate hotspot identification, development of adaptive policy options, and the preparation for rare events (Mahmoud *et al* 2009, van Beek *et al* 2020).

Thus far, few studies address the long-term future risk of conflict (de Bruin *et al* 2021, Von Uexkull and Buhaug 2021). Examples are Hegre *et al* (2013) predicting conflict towards 2050; Witmer *et al* (2017) projecting future regions at conflict risk under until 2065 using various Representative Concentration Pathways (RCPs) and Shared Socioeconomic Pathways (SSPs); and Hegre *et al* (2016) projecting conflict towards 2100 under alternative SSPs. Up until today, Witmer *et al* (2017) is the only conflict projection study engaging with the SSP-RCP framework (van Vuuren *et al* 2014).

Machine learning (ML) models have already been identified as a viable way forward in conflict risk projections (Colaresi and Mahmood 2017). Here, we use CoPro, a novel open-source ML model (Hoch *et al* 2021a), to disentangle historical relations between socioeconomic as well as hydro-climatic indicators and armed conflict. Compared to the above-mentioned examples, using ML has the distinct advantage that it is data-driven and can deal with non-linearity between indicator and conflict data without pre-defining theoretically assumed interactions. With this first, flexible, data-driven analysis of future conflict risk we aim to (a) advance the currently under-studied field of long-term conflict risk projections (de Bruin *et al* 2021), (b) evaluate model ability to quantify future changes of regions-at-risk using ML techniques, (c) evaluate the changes in conflict risk across scenarios, and (d) (re)assess the importance of socioeconomic and hydro-climatic drivers for future changes in armed conflict risk.

To which extent an ML approach can help projecting climate change impacts, including possible knock-on effects on livelihood insecurity and resource competition, is thus the central question of this paper. Understanding how different future pathways will develop can facilitate shaping sustainable, fair, and peaceful policies, and the use of data-driven approaches may be an important cornerstone in this.

2. Data and methods

2.1. Spatio-temporal properties

We applied CoPro over the entire continent of Africa (Hoch *et al* 2021a). The analysis was conducted at an annual temporal resolution which suffices for long-term outlooks of conflict risk. As spatial aggregation level we employed sub-national water provinces, which are defined by hydrological boundaries of river basins intersected with the administrative boundaries of countries (Straatsma *et al* (2020); see figure 1). By estimating and projecting conflict risk by water province, we are able to account for important within-country variation in hydrological characteristics that shape climate change impacts and, possibly, conflict risk. Also, their use mitigates challenges associated with alternative high-resolution gridded designs, such as high spatial dependence.

To train, test, and evaluate CoPro, we focused on the period 1995 until 2015, the longest intersect of available historic hydro-climatic, socioeconomic, and conflict data. We then projected conflict risk forward in time until 2050. Projections follow three alternative pathways of societal development included in the SSP scenario framework (O'Neill *et al* 2017). As not all SSPs are compatible with all RCPs, the following SSP-RCP combinations were employed to reflect a range of
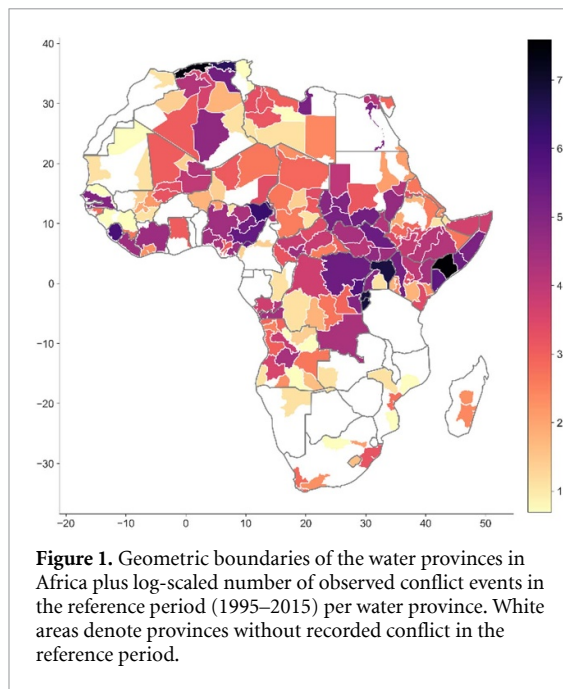
Figure 1. Geometric boundaries of the water provinces in Africa plus log-scaled number of observed conflict events in the reference period (1995–2015) per water province. White areas denote provinces without recorded conflict in the reference period.

Table 1. Overview of indicators used to project water province-level conflict risk. All indicators are sampled with a one-year time lag. See appendix B for more details.

SSP indicators	RCP indicators	Conflict-related indicators
Education	Precipitation	Conflict in year (outcome)
Population count	Evaporation	Conflict in previous year
Gross domestic product (GDP) (Purchasing power parity, PPP) per capita	Flood volume	Conflict in neighbouring province in previous year
Governance	Upper soil water storage	

socioeconomic and climate developments: SSP1 with RCP 2.6, SSP2 with RCP 4.5, and SSP3 with RCP 6.0. Details are given in appendix A.

To assess the relative importance of hydro-climatic drivers, we performed an attribution experiment: one simulation including both hydro-climatic and socioeconomic data ('SSP-RCP run'), and another one with socioeconomic data only ('SSP run').

2.2. Data description

For our analysis, we used indicators already quantified in the SSP projections that can theoretically and empirically be established as drivers of conflict risk (see table 1). An important guiding factor for data selection was the availability of consistent historical *and* projected data. Some commonly employed indicators in empirical conflict studies, such as ethnopolitical exclusion and political instability, could not be included due to the absence of SSP-consistent projections for these variables. In other cases, the parametrization of projected variables fails to capture dimensions that are salient for conflict risk. For example, the extended portfolio of SSPs includes within-country income inequality projections (Rao *et al* 2019) but these reflect inequalities between individuals whereas what mainly matters for armed conflict risk are systematic inequalities across identity groups (Cederman *et al* 2013). As more data becomes available in the future, follow-up attempts can aim at expanding the number of explanatory input variables used. We bias-corrected all variables to ensure that the statistical properties do not change between the historical period and projections. A more elaborated overview of data properties and processing can be found in appendix B. From each pixel-scale or

country-scale variable, we calculated average values by water province and year, except for population for which the annual sum was applied. We performed an additional sensitivity analysis (SA) run to test alternative sampling methods of the RCP indicators. In this run (named 'SSP-RCP (SA) run'), the minimum (i.e. worst-case) pixel-scale value of all RCP indicators per water province was used, except for flood volume for which the maximum was taken. For all runs, a one-year time lag was applied for all indicators to mitigate reverse causality between drivers and effect.

For conflict event observations, we employed the UCDP Georeferenced Event Dataset (GED) v20.1 (Sundberg and Melander 2013, Pettersson and Öberg 2020). We selected data on 'state-based armed conflict' and 'non-state conflict' events, indicating deadly conflict between the government and one or more non-state actors or between non-state actors, respectively. Conflict events between countries were not included as they remain exceptionally rare, and accounting for this conflict type would require a different research design. Conflict was coded as a binary variable, obtaining the value '1' if at least one conflict event was reported in the given water province during the year and '0' if not.

To account for history of conflict, a well-established driver of conflict occurrence (Hegre and Sambanis 2006, Mach *et al* 2019), we sampled whether armed conflict took place in the same province during the previous year. Additionally, we sampled whether a conflict event occurred in any of the neighbouring provinces in the previous year to account for 'spill-over effects' (Buhaug and Gleditsch 2008, Schutte and Weidmann 2011). A binary value was assigned depending on the outcome.

2.3. Set-up of the ML model

By means of using ML methods, we determine the historic relation between the indicators ('sample data') and conflict risk ('target data'). It is in the nature of the ML algorithm applied that this relation is stationary in time. The established link does hence

not change between the historical training period and future projections, a general limitation common with previous work on projections (Bowlsby *et al* 2020). Although we are unable to explore alternative assumptions of dynamic predictive power here, using ML has the distinct advantage that it can flexibly deal with non-linearity between the sample and target data and without pre-defining theoretically assumed interactions between the indicators. We defer this important challenge to future research.

2.3.1. The reference period 1995-2015

To derive a stable relationship between the indicators and conflict, we employed the open-source software package CoPro v0.1.1 (Hoch *et al* 2021a, 2021b) to train a Random Forest classifier (RFC) model with 21 years of data (1995–2015). See appendix C for a detailed model description.

For each year of the reference period 1995–2015, values were extracted for all indicators. The resulting sample data and target data were appended annually, yielding a 'master matrix'. To minimize the risk of overfitting our model, 100 RF trees were initialised. For each tree, 70% of the master matrix data were randomly drawn to train the model and the remaining 30% were preserved to evaluate the predictions. The RFC model, therefore, follows a different approach than for example Witmer *et al* (2017), who uses a conventional regression model framework.

These predictions were subsequently evaluated against observed conflict events and a range of evaluation metrics was computed (see section 3.1). Additionally, the relative importance of each indicator was assessed to improve our understanding of their relation with conflict occurrence. While the evaluation metrics focus on the accuracy of all data points, it is also important to assess accuracy per water province. Hence, the fraction of correct predictions (FOPs) per water province polygon *i* was determined as follows:

$$\text{FOP}_i = \frac{\sum \text{cp}_i}{N_i}$$

where *cp* denotes a correct prediction and *N* the number of predictions made for a given polygon *i*. FOP can thus range between 0 (no correct prediction) and 1 (only correct predictions). Computing the FOP allows for identifying provinces where model output is more likely to be correct.

2.3.2. Projections of conflict risk until 2050

From the end of the reference period until 2050, we make annual out-of-sample forward projections. To maintain the internal consistency of each projection pathway, this was done for each selected SSP-RCP combination and each of the 100 RF trees separately.

The last reference year (here: 2015) is used to initialize the conflict risk projections, since all projections are based on indicator and conflict values of the

previous year due to the 1 year time lag. All projections after 2016 draw upon the projected binary maps of conflict occurrence in the previous time step of each individual RF tree, while the remaining indicator values are provided by SSP- and RCP-specific input data.

At the end of each projection year, the outcomes of all trees are combined per water province. We therefore not only obtain a final projection for the year 2050 but also all years in between, yielding the possibility to track conflict risk development over time (see appendix D for conflict risk development over the entire African continent).

As a quantitative validation of the long-term projections against true outcomes is not possible, model output is evaluated by comparing projections across all scenarios for the SSP-RCP and SSP run separately. Therefore, the probability of conflict (POC) per polygon *i* was annually determined over all RF trees (*T*) as:

$$\text{POC}_i = \frac{\sum_t^T P(c)_{i,t}}{T}$$

where *P(c)* denotes the projected probability of conflict per polygon *i* and RT tree *t*.

3. Results and discussion

3.1. Model validation

The reference period 1995–2015 is used to evaluate the performance of the SSP-RCP run and the SSP run. In general, only marginal differences in predictive performance are reported between both runs (table 2, figure 2). Including hydro-climatic information has thus only limited effect on the model's ability to correctly predict conflict risk across the African continent for the historical sample in the current study design.

For both runs, the overall model performance is good as indicated by ROC-AUC scores above 0.9, with the SSP run showing a slightly better performance. The computed ROC-AUC score is in line with previous studies (Hegre *et al* 2013, Colaresi and Mahmood 2017). The mean Brier-score, measuring the mean squared difference between the predicted probability and the actual outcome, is slightly higher than that computed by Witmer *et al* (2017) but comparable with Hegre *et al* (2019).

Overall accuracy—that is, the fraction of correct classifications—is good in both runs. Mean precision (the ability of the classifier not to label an observation as 'conflict' that is 'non-conflict') is slightly higher in the SSP-RCP run whereas recall, which expresses the ability of the classifier to find all positive observations, is lower than in the SSP run. The relatively low recall in both runs is most likely rooted in the imbalanced training dataset due to the small fraction of conflict observations (∼22%).

Table 2. Overview of computed model evaluation scores for both SSP-RCP run and SSP run for the period 1995–2015.

	SSP-RCP run	SSP run
Average ROC-AUC score	0.90	0.91
Average precision-score	0.74	0.77
Average accuracy	0.87	0.87
Average precision[a]	0.76	0.75
Average recall	0.62	0.64
Average Brier-score	0.10	0.09
Average FOP	0.87	0.87

[a] Note that precision and precision-score are different. While the former computes actual precision of model output, the latter is based on the precision-recall curve.

Figure 2. (A) ROC-curve averaged over 100 model repetitions for both SSP-RCP run and SSP run. (B) probability density functions for accuracy, precision, recall for SSP-RCP (darker shade; solid line) and SSP runs (lighter shade; dashed line).
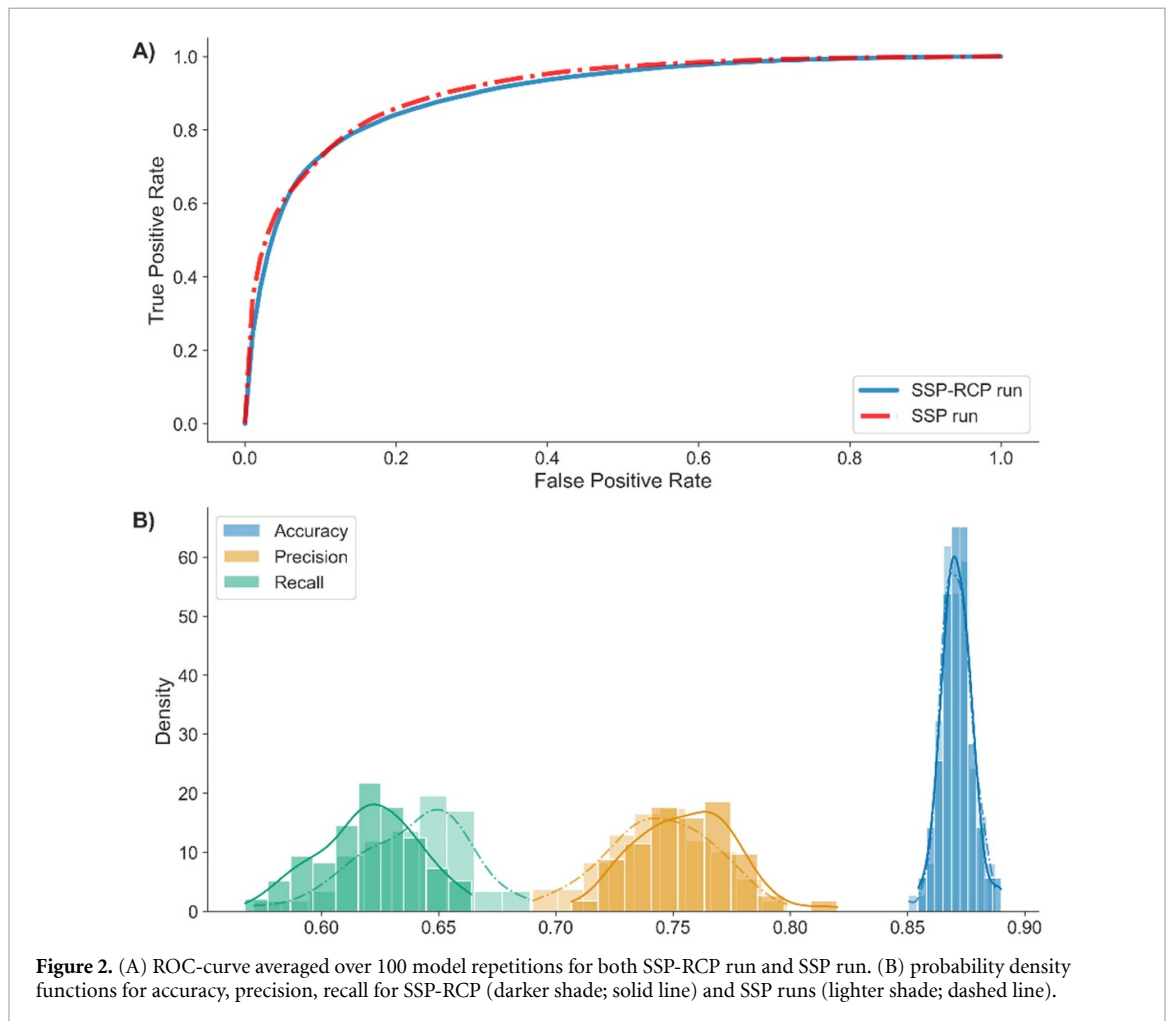
The spatial model performance strongly depends on the number of conflict events between 1995 and 2015 per water province (see figure 3(A)). Predictions of conflict are more accurate in very conflict-rich provinces and in provinces with no or little conflict observations. In contrast, polygons with an intermediate number of reported conflict events tend to be less accurately predicted. Overall FOP is nevertheless high with a sample average of 0.87 in both runs (see table 2). Identical values are obtained in the SSP-RCP (SA) run (see appendix E), indicating robust model performance across settings.

Areas with low model accuracy in the reference situation as expressed by low FOP values include southern Algeria as well as parts of the Sahel and Sahara, the Democratic Republic of the Congo (DRC), Somalia, and Ethiopia (figure 4(A)). In these areas, only an intermediate number of conflict events is observed (figures 1 and 3). There is, however, not a single country for which all water provinces are poorly modelled—an advantage of using a sub-national aggregation level. Conflict-prone regions identified with high POC in the out-of-sample validations are, *inter alia*, the Horn of Africa, South
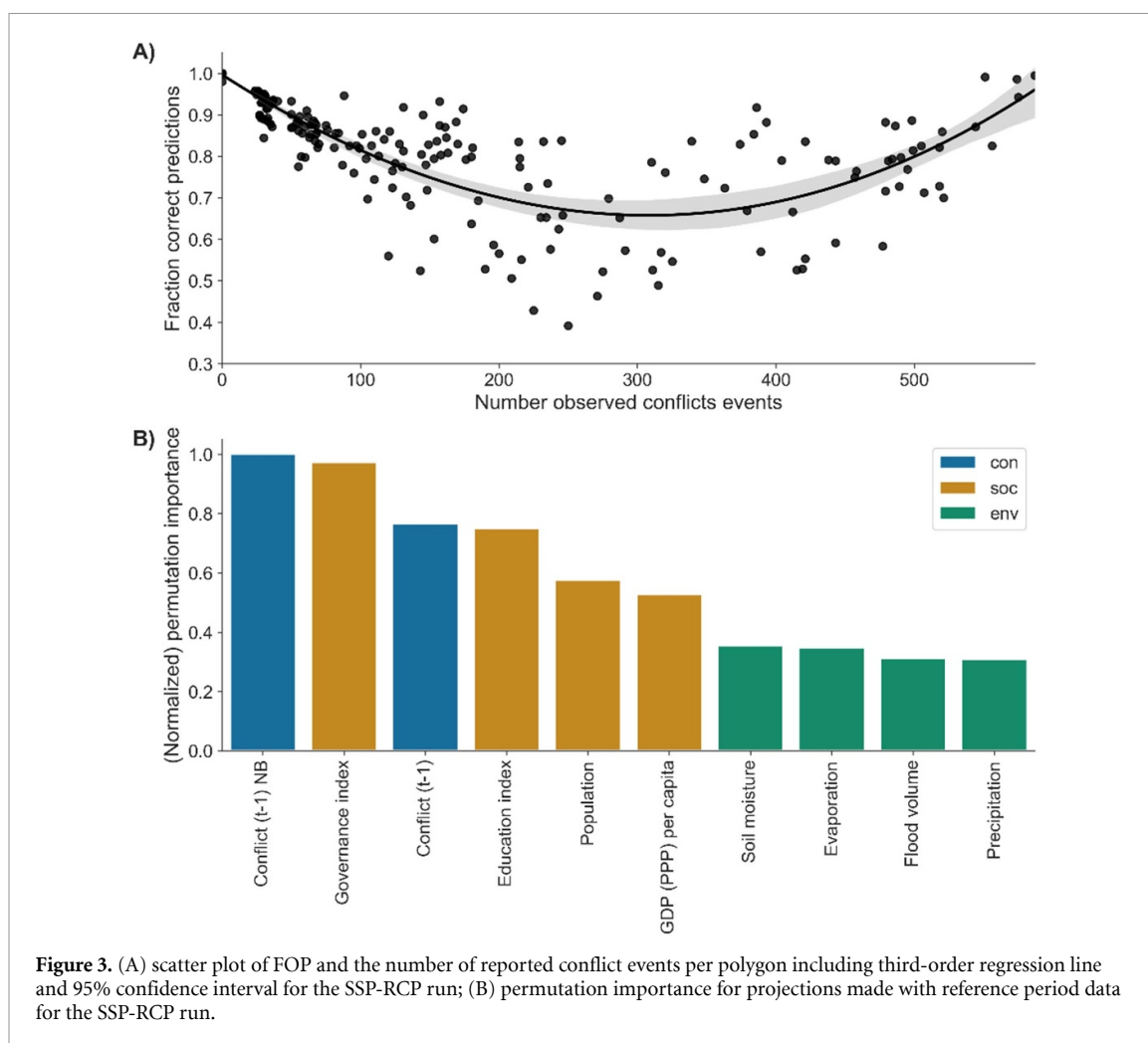
Figure 3. (A) scatter plot of FOP and the number of reported conflict events per polygon including third-order regression line and 95% confidence interval for the SSP-RCP run; (B) permutation importance for projections made with reference period data for the SSP-RCP run.

Sudan, Nigeria, and the north-eastern part of DRC (figure 4(B)). Projections for these areas largely agree with observations of current conflict as reported in the conflict database (figure 1). By comparing FOP and POC values obtained by the SSP only and SSP-RCP run, we find that for the reference period the inclusion of hydro-climatic variables both regionally improves and reduces accuracy as indicated by high FOC values (figure 4(C)) and that particularly eastern Africa and Nigeria are predicted by the SSP-RCP run to be more conflict-prone than in the SSP run (figure 4(D)). The distribution of provinces where inclusion of RCP indicators improved FOP values is 54%, again quantifying their overall limited impact. Detailed maps of FOP, FOP difference, and a number of observed conflict events for selected regions can be found under appendix F.
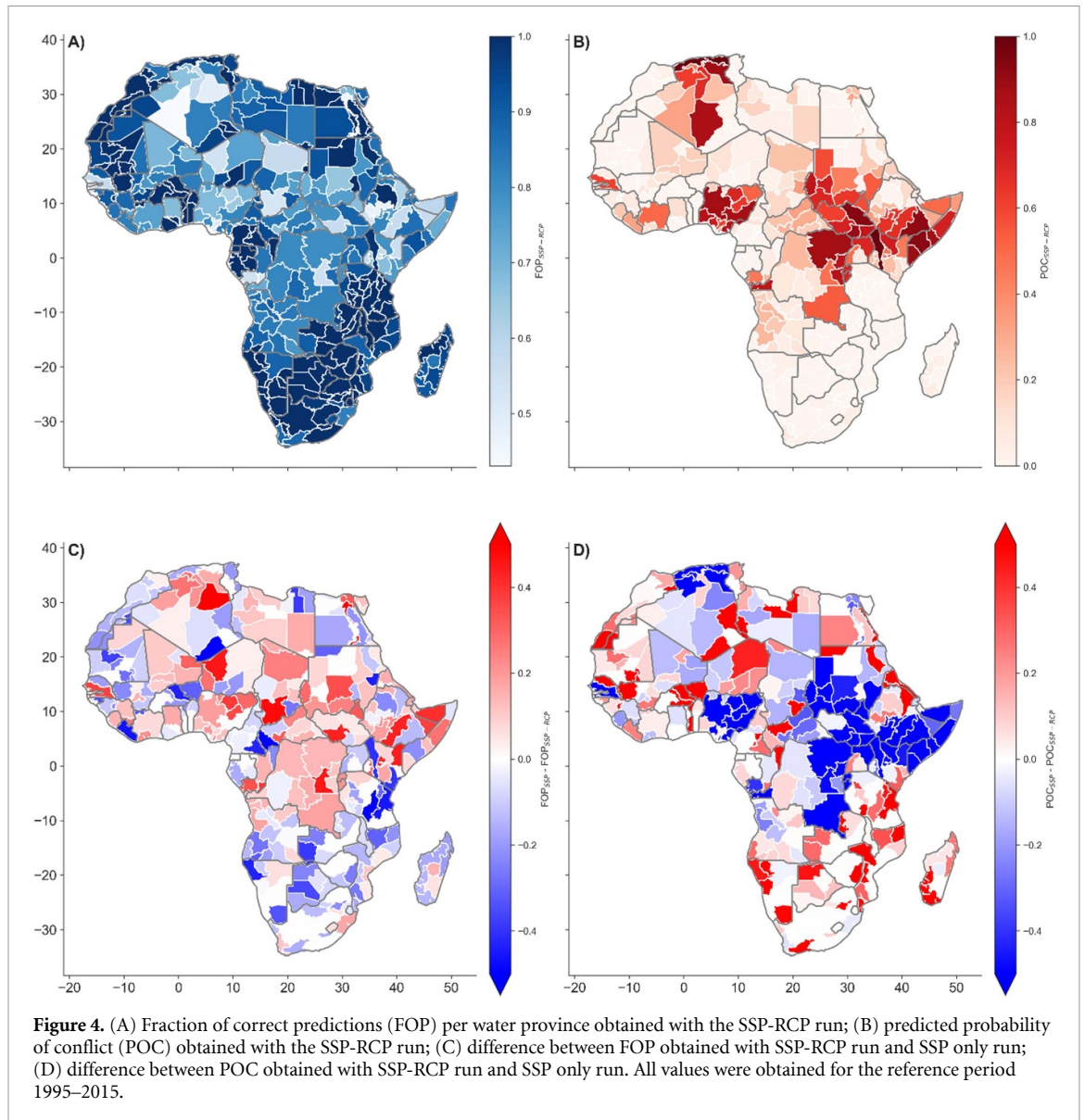
3.2. Major predictors of conflict

To assess the indicator importance in RF models, there are multiple approaches (Tyralis *et al* 2019). Here, we computed the permutation importance per indicator, that is, the decrease in model score when the original relation between indicator and dependent values is broken (Breiman 2001). The

permutation importance was subsequently normalized relative to the indicator with the highest value to improve comparability. It is important to note that the permutation importance does not provide information on whether a variable increases or decreases conflict risk. Aggregating importance is therefore not sensible as different variables can have countervailing effects.

The indicator with the highest importance is conflict in the previous year (figure 3(B)). A recent history of conflict is an important, well-documented driver of conflict (Hegre and Sambanis 2006, Goldstone *et al* 2010, Bara 2014, Mach *et al* 2019). Previous conflict in neighbouring water provinces also plays an important role and is ranked third (Buhaug and Gleditsch 2008, Schutte and Weidmann 2011).

The second-ranked indicator is quality of governance, whose relevance again is supported by earlier empirical studies (Goldstone *et al* 2010, Besley and Persson 2011, Walter 2015).

Education and population count are ranked fourth and fifth. Education may have indirect impacts via socioeconomic divisions as well as varying degrees of political inclusion (Barakat and Urdal 2009, Brown 2011). A high population count is found to amplify

Figure 4. (A) Fraction of correct predictions (FOP) per water province obtained with the SSP-RCP run; (B) predicted probability of conflict (POC) obtained with the SSP-RCP run; (C) difference between FOP obtained with SSP-RCP run and SSP only run; (D) difference between POC obtained with SSP-RCP run and SSP only run. All values were obtained for the reference period 1995–2015.

the risk of conflict through multiple processes, including by increasing the likelihood of finding a critical mass of prospective combatants (Raleigh and Hegre 2009).

GDP per capita (PPP) is found to be of less importance than other socioeconomic indicators. This may be surprising since low economic development is often mentioned as a major risk factor for conflict (Fearon and Laitin 2003, Mach *et al* 2019). The modest explanatory power in our model is partly a product of also accounting for human development (education), which often is ignored in conflict studies, and because our spatial sample includes mostly low-income countries (Vestby *et al* 2021).

Overall, the hydro-climatic indicators are found to be the least influential among the indicators but still add to the explanatory power of the ML model. This is in line with prevalent findings, underlining that climate anomalies themselves are unlikely to lead to conflict in the absence of adverse socioeconomic

conditions (Mach *et al* 2019). Here, soil moisture and evaporation are of slightly higher importance than flood volume and precipitation although the overall differences are marginal.

The overall picture therefore shows that CoPro can capture the main historical spatial and temporal variability of conflict occurrence over Africa well. With respect to indicator importance, model results follow the current understanding of contemporaneous literature by assigning higher importance to the history of conflict and socioeconomic drivers than to hydro-climatic variables.

3.3. Scenario projections
3.3.1. Output analysis
After validating CoPro model output for the historical period, we first explore the projections made with multiple SSP-RCP combinations and subsequently compare them with output from SSP only runs. The volatile and somewhat stochastic pattern of conflict

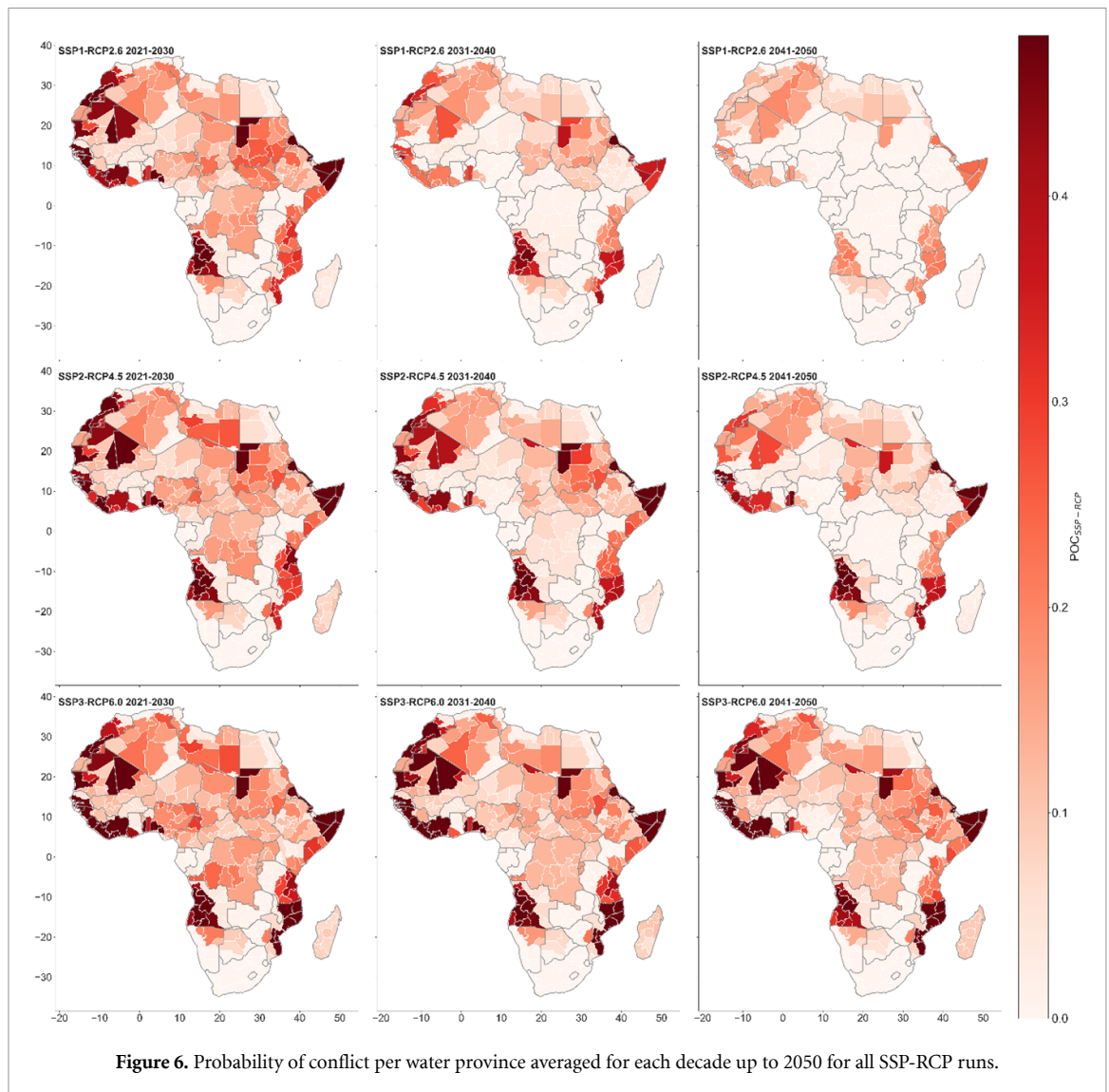**Figure 5.** (left) projected probability of conflict (POC) per water province averaged over period 2041–2050 for the SSP-RCP run; (right) absolute difference between simulated POC for SSP run and SSP-RCP run per water province for corresponding SSP-RCP combinations. Blue colours correspond to higher risk without the hydro-climatic projections used in the SSP-RCP run (i.e. climate change contributes to reducing simulated conflict risk in these areas). Note that for the right panel the legend values are manually set for improved visualization of the spatial patterns.

onset and ending suggests that evaluating projections for a single year may yield rather arbitrary results (see appendix D). We therefore decided to average output over the final decade 2041–2050 to obtain a more robust picture.

Projections made reflect the scenario storylines and show greater conflict probability in SSP3-RCP6.0 compared to SSP1-RCP2.6 (figure 5). This difference between scenarios is consistent over time (figure 6). For all projections, the spatial spread is less than in the reference situation. Given the more sustainable

development of SSP1-RCP2.6 compared to today, a reduction of conflict-prone areas can be expected, in line with earlier research (Hegre *et al* 2016). Even so, the simulated drop in overall conflict propensity is also driven by the overly optimistic quantitative projections for future socioeconomic development in Africa even under SSP3 (see Buhaug and Vestby 2019) that depresses future modelled conflict risk particularly for low-income countries. Governance projections may also be overly optimistic in the SSPs as its future development is modelled as a function of

Figure 6. Probability of conflict per water province averaged for each decade up to 2050 for all SSP-RCP runs.
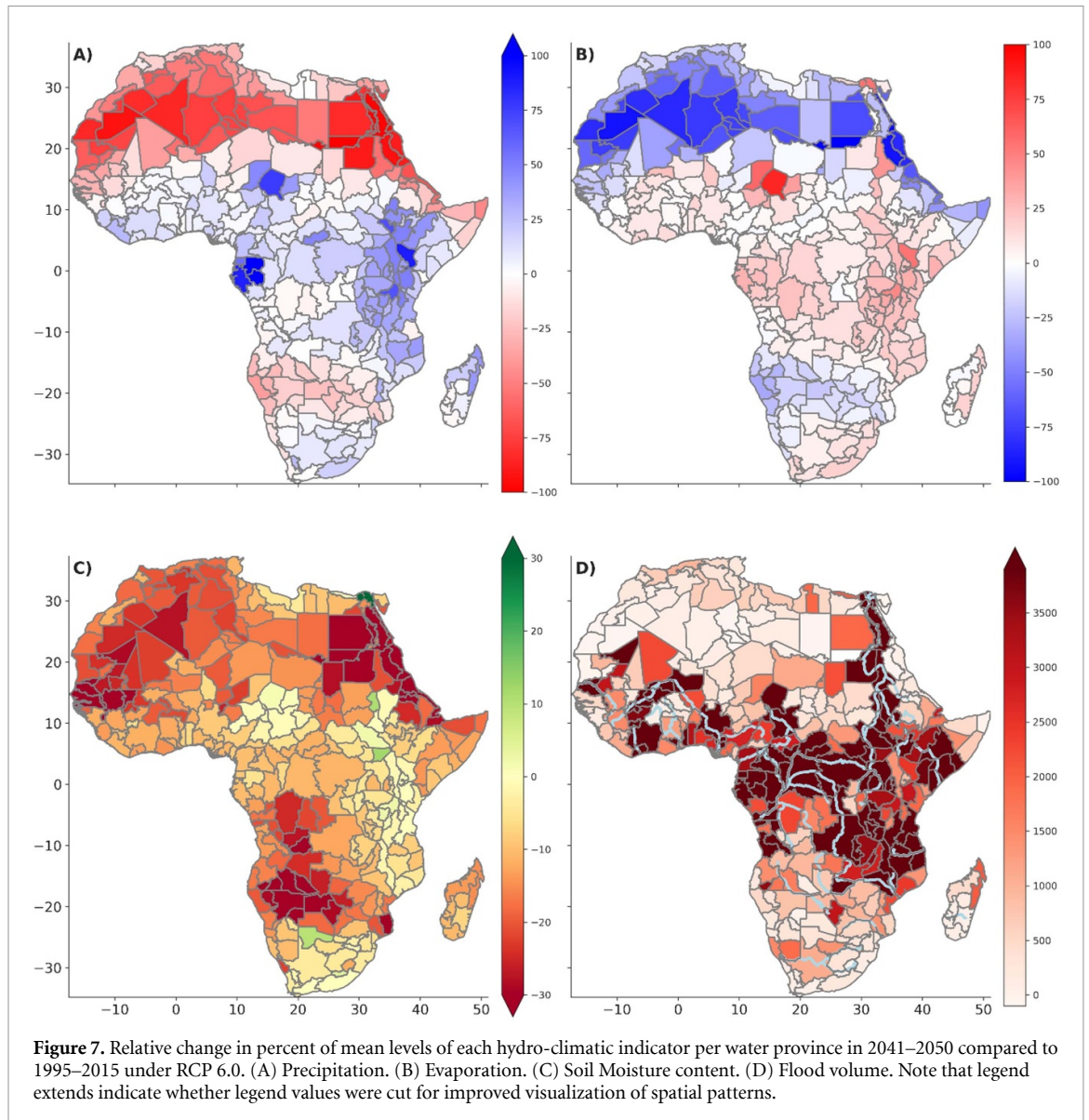
economic growth, implying that overall conflict prevalence may be higher than what these risk projections indicate, especially in less optimistic development futures.

Figure 5 shows the distribution of and divergence in projected POC over Africa for the SSP-RCP runs compared to the reference scenarios. For SSP1-RCP2.6, the highest POC is obtained for North and West Africa as well as for (parts of) Mozambique, Tanzania, Kenya, and Angola. For SSP3-RCP6.0, and to a lesser extent for SSP2-RCP4.5, almost the entire Sahara and Sahel zone and the Horn of Africa face substantial armed conflict risk. Other areas projected to experience increased POC in SSP3-RCP6.0 compared to the other scenarios are large parts of Angola, DRC, Northern Mali and coastal West Africa. Conflict risk also increases in southern Morocco and Mauritania. These areas overlap only partly with those having a high POC for the reference situation (see figure 4(B)).

Comparing output from the SSP-RCP and SSP only runs, several patterns can be observed. For the

SSP1 scenario, overall differences are small, owing to the relatively modest changes in projected hydroclimatic conditions until 2050 in the associated RCP 2.6 pathway. In SSP2, especially Northern and parts of Central Africa are projected to be more conflict-prone when climate effects are not accounted for, whereas parts of the Sahel and southern Africa are projected to have a decreased POC. In SSP3, differences are found to increase especially in the Sahel, showing both a higher POC (northern Sahel) and a lower POC (southern Sahel) when not considering RCP indicators. In general, and as expected, results depict that the influence, both negative and positive, of climate change becomes more pronounced with higher RCPs.

To explore this in more detail, a closer look at RCP 6.0 reveals that in Northern Africa, projected decreases in precipitation and evaporation (figure 7) correspond with higher POCs in the SSP-RCP run compared to the SSP run. In DRC, increases in flood volume may add to an increased POC. Meanwhile, increased levels of precipitation in Western Africa and southern parts of the Sahel could explain a lower

Figure 7. Relative change in percent of mean levels of each hydro-climatic indicator per water province in 2041–2050 compared to 1995–2015 under RCP 6.0. (A) Precipitation. (B) Evaporation. (C) Soil Moisture content. (D) Flood volume. Note that legend extends indicate whether legend values were cut for improved visualization of spatial patterns.

POC in the SSP-RCP scenario. For other regions, the hydro-climatic patterns are too ambiguous to make a substantive influence on the projections.

3.3.2. Projection uncertainties

A clear caveat in making these projections is the implicit (but common) modelling assumption that the shape and strength of relationships between the predictors and the outcome remain stationary across the training and projection periods. Relations will most likely not remain stable over time; especially when climate change impacts worsen, its role is likely to increase with respect to the reference situation due to non-linear sensitivities and potential social tipping points (Mach *et al* 2019). Also, Bowlsby *et al* (2020) point out that the drivers of instability are not constant over time and that care must be taken when interpreting projection studies based on historical relations. This limitation could be partly overcome by using more advanced deep-learning and

self-learning ML models or by altering the historical relation between indictors and conflict to explore an ensemble of possible futures. However, such more complex models also would make it more difficult to understand the input–output relations between drivers and conflict risk.

When testing the output sensitivity to different sampling methods of the RCP indicators, results of the SSP-RCP (SA) run indicate an overall agreement of projected trends at the regional scale (see appendix E). Locally, projections of the climate sensitivity run show, however, both negative and positive deviations, indicating that the way climate variables are sampled may affect projection outcomes at the water province scale.

Furthermore, the impact of hydro-climatic data must be assessed carefully as the RCPs are quantified differently in different GCMs. The applied IPSL model provides only one of multiple possible realizations of future climate. IPSL was selected as it

projects changes that are in the mean of the full ensemble of CMIP5 GCM models (Warszawski *et al* 2014, Wanders *et al* 2015). However, the direction and magnitude of change for specific climate indicators vary across GCMs in some parts of the African continent.

Still, the results exhibit consistency in space and time across the outputs for various SSP-RCP combinations as projected POC values agree with the underlying scenario storylines (appendix A). In the end, we cannot claim with certainty how interactions and relations will develop in the future, and how armed conflict risk will be distributed in space and time. As projections in general can at best work as realizations of imaginable futures (de Bruin *et al* 2021), it would not be credible to pretend that we hold this knowledge, nor that it can be accurately included into models. As such, the conflict maps shown represent a limited number of plausible realizations among an infinitely imaginable set of possible futures.

4. Conclusions and recommendations

To project future areas at risk of armed conflict, we employed the open-source ML model CoPro to produce maps of regions-at-risk for various scenarios in Africa until 2050. Also, we compared the relative impact of hydro-climatic variables on conflict occurrence. To our knowledge, this study represents the first attempt to use ML for long-term conflict risk projections. By using data-driven approaches, existing model designs can be complemented and theoretical insights can be contributed to the ongoing debate on the potential impacts of climate change on armed conflict.

Results indicate a more peaceful future compared to current conditions for SSP1-RCP2.6, and in many areas also under SSP2-RCP4.5. In the SSP3-RCP6.0 scenario, conflict risk will increase in many regions that already suffer from high prevalence of conflict, particularly in the Horn of Africa and parts of West Africa and East Africa (figure 6). These results are consistent with the underlying scenario storylines and other studies. Besides, our results indicate that hydroclimatic indicators may both increase and decrease conflict risk, depending on the location: in Northern Africa and large parts of Eastern Africa climate change increases projected conflict risk whereas for areas in the West and northern part of the Sahel shifting climatic conditions may reduce conflict risk. Since the runs performed are more experiments than depictions of the real world with all its complexity, these findings must, however, be interpreted carefully.

A wider range of quantified SSP indicators would allow for ensemble projections and thus for mapping their uncertainties (O'Neill *et al* 2020). Until then, we are limited to available sources, including too bright projections of economic growth for low-income countries that also affect the governance

projections (Buhaug and Vestby 2019). Currently, ensemble projections are only possible for RCP indicators derived from GCMs. In follow-up studies, their ensemble output should be used to confirm (or dismiss) our findings of the projected impact of hydroclimatic indicators.

We also recommend investigating the role of on-the-ground impact of the meteorological drivers precipitation and temperature. Changes thereof cannot be translated directly to changes in conflict, but it is rather the local impact that is decisive. Example candidates are the impact of climate change on groundwater levels (Döring 2020), actual flood and drought risk (Von Uexkull 2014, Ide *et al* 2021), crop production (Von Uexkull *et al* 2016), and food prices (Raleigh *et al* 2015).

This study merely focused on climate change impact of hydrology-related indicators. Other climate-related factors that might inform conflict risk, such as heatwaves, are not considered. Besides, the use of annual averages does not capture changes in, for example, timing and intensity of the rainy season, and cumulative effects building up over time. Future work should hence try to include these intra- and inter-annual effects. With the flexible structure of CoPro and the implemented ML approach, new insights and novel data sources can be included as they become available.

We found that data availability is a major constraint for advancing data-driven projections of armed conflict risk. Since the distribution of areas with observed conflict events versus areas without conflict is imbalanced towards the latter, transitional areas that have seen violence only sporadically or in parts of the training period are more difficult to predict. Furthermore, only drivers that have been projected within the SSP framework (plus the extended governance data) could be employed, whereas empirical conflict literature offers additional contextual variables of importance, such as political discrimination and grievances (Cederman *et al* 2013) and agricultural dependence (Von Uexkull *et al* 2016). When improved quantitative data under the various SSPs becomes available, data-driven projections can be advanced. Another avenue for future work is considering potential differences in responses for different conflict types, as well as the unique scope conditions under which these might materialize (Von Uexkull and Buhaug 2021).

Adverse climate change impacts intensifying in many regions raise concerns for peace and security. As precise knowledge about 'where' and 'when' of conflict onset is impossible to obtain for long-term projections, following various scenarios and producing consistent maps of possible conflict risk realizations may facilitate informing policy-making processes. Based on these conflict maps, the potential consequences of today's decision-making on long-term conflict development can become tangible. This study

points to the benefits for peace of investing in economic, human, and political development and maintaining sustainable demographic change (resulting in a SSP1 world with decreasing radiative forcing) over nationalism and protectionism (resulting in a SSP3 world with stabilizing radiative forcing). Our study also shows that projecting conflict risk with ML approaches may be a viable way forward towards more insights into the delicate interplay of climate change and conflict.

Data availability statement

Acknowledgments

Appendix A

In deciding which SSP-RCP combinations to use, we followed the matrix of possible combinations as provided by van Vuuren *et al* (2014). Within these possibilities we included the more divergent combinations. Table 3 provides brief descriptions of the SSP and RCP scenarios used.

Appendix B

The following socioeconomic indicators were used: pixel-scale log-transformed population count (Jones and O'Neill 2016), pixel-scale log-transformed gross domestic product per capita based on purchasing power parity (GDP per cap (PPP); Murakami and Yamagata 2019), country-scale education expressed as the mean number of schooling years at age 25 or older (Wittgenstein Centre for Demography and Global Human Capital 2018), and country-scale

estimates of quality of governance (Andrijevic *et al* 2020)1F[5], with the latter representing an extension to the basic SSP projections based on the World Bank's Worldwide Governance Indicators.

As hydro-climatic indicators we selected yearly anomalies of precipitation, evaporation, flood volume, and upper soil water storage per water province. These pixel-scale indicators were selected to represent overall climate variability (precipitation and evaporation) and on-the-ground hydrological effects (floods and soil water storage as proxy for droughts (Basche *et al* 2016, Silva 2017)). All environmental variables were simulated with the global hydrological model PCR-GLOBWB (Sutanudjaja *et al* 2018). For climate projections under the various RCPs, the model was forced with CMIP5 output from the global climate model (GCM) IPSL, derived from the ISIMIP ensemble (Warszawski *et al* 2014) to ensure consistency between the historical and future records.

Additional details are provided in table 4.

In line with common approaches in climate science (Teutschbein and Seibert 2012), we bias-corrected all variables to ensure that the statistical properties are not altered significantly moving from the historical record to the projections as such alterations could potentially weaken the relation between the projected indicators and conflict events. Thus, we used the last available observation and compared it to the first year from the projection. We assumed that the computed additive bias remains constant throughout the projections and corrected all future years accordingly.

All indicators are gridded and were conservatively resampled to a 5 arc-min spatial resolution (that is, around 10 km by 10 km). For those indicators with discontinuous temporal coverage of the simulation period (both reference and projection period), linear interpolation was applied between available data points. The same data sources were used for both the reference period and the projection period. This list gives only the variables that are exogenously entered into the projections. Conflict in neighbouring provinces and history of conflict are based on the dependent conflict variable.

Appendix C

This appendix outlines the main characteristics of CoPro, the machine learning (ML) framework developed to project long-term conflict risk. More specific and detailed details can be found on the online documentation, which also contains interactive examples of the various steps taken throughout the simulation (https://copro.readthedocs.io/en/latest/).

---

[5] Although governance is typically not subsumed under socioeconomic indicators, we do so here for the sake of easier readability and traceability throughout the manuscript.

Table 3. Overview and descriptions of scenarios used in the study.

	RCP	Scenario description (from O'Neill *et al* (2017))
SSP1 Sustainability	2.6	SSP1 is characterised by a gradual shift towards a more sustainable and inclusive path than today's. International cooperation, higher levels of health care and education accelerate a downward demographic trend. Challenges for mitigation and adaptation are low. Under RCP 2.6, total radiative forcing increases to 3.0 W m^{-2} until mid-century before a decline begins. It is the low end of the scenario literature in terms of emissions and radiative forcing (van Vuuren *et al* 2011). For this scenario, greenhouse gases emissions need to be collectively reduced.
SSP 2 Middle of the road	4.5	SSP2 follows the current trends in environmental and socioeconomic developments without fundamental breakthroughs. Challenges for mitigation and adaptation are moderate. Under RCP 4.5, total radiative forcing will have increased relatively steeply to around 3.8 W m^{-2} before stabilization begins. To reach RCP 4.5, changes in the energy system are needed and cost-efficient technologies to lower net emissions must be implemented (Thomson *et al* 2011).
SSP 3 Regional Rivalry	6.0	SSP3 is characterised by an increase in nationalism, degrading environmental developments and declining investments in healthcare and education, leading to high population growth in lower income countries. Challenges for mitigation and adaptation are high. Total radiative forcing under RCP 6.0 increases steadily to 3.5 W m^{-2} in 2050. Stabilization only begins in the end of the century. RCP 6.0 implies explicit climate policy intervention and greenhouse gas emissions peak around 2060 and then decline until 2100 (Masui *et al* 2011).

Details on the specific application in this paper, such as input variables and division of training and test data, are found in sections 2.2 and 2.3 in the main manuscript.

CoPro software requirements and installation

CoPro is a computational framework specifically designed to project conflict risk using ML methods. It is entirely written in Python and makes use of the latest geospatial and ML packages. During development, emphasis was put in usability which was reviewed in a separate software publication (Hoch *et al* 2021a). CoPro can be installed on both Windows, MacOS, and Linux. Installation is possible either from source code, giving users the possibility to further develop the software, or as compiled software for immediate use (see https://copro.readthedocs.io/en/latest/Installation.html). In both cases, but particularly the first, minimum Python experience is necessary.

Once installed, CoPro can be executed from command line alongside a text file (hereafter named 'config-file') containing information about data sources and settings for a run. The user thus only needs to fill the config-file with run-specific input data and settings, but does not need to adapt anything on the software side of things.

Input data requirements and settings

To run CoPro, data sources and settings need to be provided in a text file, hereafter named 'config-file'. A template can be found at https://copro.readthedocs.io/en/latest/Settings.html. CoPro can be run with any input indicator dataset as long as

it meets the following requirements: (a) it has a clear indicator variable name; (b) it has continuous annual data along the time axis, (c) it is gridded with longitude and latitude information, and (d) it is provided in netCDF-format. In a nutshell, the input netCDF-file needs to have three dimensions: longitude, latitude, and time. Also, the spatial aggregation level (e.g. water provinces, counties, states, countries and so forth) can be user-defined by providing a file with corresponding polygons which altogether define the overall study area. The only input dataset that is not flexible is the conflict event data. Here, CoPro is currently still limited to UCDP GED (Sundberg and Melander 2013, Pettersson and Öberg 2020). The only flexibility with respect to the conflict event data is the type of violence which can be user-defined. Future work will aim at including other conflict event databases such as ACLED. Additional settings that need to be provided to CoPro are:

- the historical time period;
- the year until which projections are ought to be made;
- optionally, climate zones can be specified which will work as masks for the study domain. That way, only the overlay area between the study area and the selected climate zones will be considered in the simulations;
- a location where to store model output.

Machine learning settings

In addition to the input data requirements and settings, a couple of settings need to be specified in the config-file with respect to the ML methods to be employed.

Table 4. Overview of indicators used plus their data source and additional notes.

Indicator	Unit	Source	Notes
Population count	—	Jones and O'Neill (2016)	A downscaling model was used to produce projections of spatial population change that are quantitatively consistent with national population and urbanization projections for the SSPs and qualitatively consistent with assumptions in the SSP narratives regarding spatial development patterns.
Gross domestic product (GDP) per capita (purchasing power parity (PPP))	Billion USD (2005)/capita	Murakami and Yamagata (2019)	The GDP (PPP) is determined by downscaling urban and non-urban population by using multiple auxiliary variables, yielding gridded values until 2100 by 10 years. GDP per capita (PPP) is obtained per water province by dividing the mean GDP (PPP) with population count averaged over each water province.
Education	years	Wittgenstein Centre Human Capital Data Explorer (2021)	Mean number of schooling years at age 25 or older for all sexes at the country level.
Governance index	—	Andrijevic *et al* (2020)	This composite governance index is computed at the country level taking the Worldwide Governance Indicators (WGI) as the starting point. Determinants for the projections are, inter alia, GDP per capita, education, and gender gap in education. A more elaborated outline can be found in the methods section of Andrijevic *et al* (2020) and its supplement.
Upper soil water storage	m	PCR-GLOBWB (Sutanudjaja *et al* 2018)	All hydro-climatic variables were continuously simulated with the global hydrological model PCR-GLOBWB at 10 arc-minutes spatial resolution. Annual averages of the model output formed the input for this study. Anomalies were determined for both the reference and projected period based on mean reference value.
Total evaporation	m		
Flood volume	m^3		
Precipitation	M		
Simulated conflict	—	Simulated by CoPro	Per water province, a Binary indicator whether the machine learning model projects conflict (value 1) or non-conflict (value 0). Is accompanied by the probability of conflict which is computed as outlined in section 2.3.2.
Conflict in previous year	—	UCDP GED v20.1/CoPro	For the reference period, this Binary indicator was determined per water province by checking whether there was at least 1 conflict event reported in the UCDP GED dataset (Sundberg and Melander 2013, Pettersson and Öberg 2020) in the previous year. For the projection period, conflict occurrence simulated by each classifier was used instead.
Conflict in neighbouring province in previous year	—	UCDP GED v20.1/CoPro	Idem, but specified to capture spatial instead of temporal proximity to other events, i.e. whether there was at least one neighbouring water province where conflict was reported (reference run) or simulated (projection) in the previous year.

- The scaling method. The scaling method defines how the values sampled from the different indicator datasets are converted to standard normally distributed data. This is needed to avoid indicators with higher variance to dominate the ML estimation. CoPro currently supports four different scaling methods;
- the ML method. CoPro currently supports three methods: NuSVC (Nu-Support Vector Classification), KNeighborsClassifier, and Random Forest Classifier. They all belong to the class of supervised learning classifiers (see below for additional information). Each model differs how the relation between indicators and conflict event data is established as well as in additional tuning parameters;
- the fraction of datapoints used for training the ML model. This inherently defines the fraction that can be used to validate the out-of-sample predictions;
- The number of ML model instances to be used. This is a measure to account for the uncertainty and

arbitrariness how the ML method chooses the data used for training and evaluating the predictions.

Supervised learning classification

As we distinguish our ML target (that is, the variable whose prediction we try to optimize) as either 'conflict' or 'no-conflict', we can speak of classification. And as we know these labels *a priori* and feed the ML model with this information, we employ supervised learning classification methods: methods that learn under user-supervision using upfront-known classifiers for the target data. The ability to learn is then also the main difference compared to more conventional statistical methods such as (linear) regression as for instance used by Witmer *et al* (2017).

Within supervised learning classification, there is a plethora of ML routines. We included three of these routines into CoPro and briefly explain them here in more detail.

Nu-SVC is a classification method from the group of support vector machines (SVMs). These SVMs separate labelled target data using a hyperplane, which in case of two indicators is a line, as decision boundary. Depending on which side of the hyperplane the indicator values fall, a SVM returns the corresponding label. SVMs have the advantage of low computational demand if drawing a hyperplane is feasible.

The KNeighbors Classifier is a classification method from the group of nearest neighbors. To predict the label P (e.g. conflict or not) of a point in a two-dimensional case with two indicators A and B, the KNeighborsClassifier would first calculate the distance between the indicator value pair (A_P, B_P) to all other known indicator value pairs. Depending on the value provided for *k*, the classifier selects all labels within the radius *k* for a decision and assigns the majority of the labels found. By changing the value for *k*, the search radius and thus number of labels in the search can be increased or decreased.

The Random Forest Classifier (applied in this manuscript) belongs to the group of ensemble algorithms. It randomly selects from the known indicator values and corresponding labels to create the so-called decision trees. Each tree is further branched up to a certain depth or until there is no additional information gain. To predict the label from indicator values, the average vote from all decision trees is employed. In a binary example, the predicted label would be 1 if the average from all decision trees is above 0.5. This method is suitable if the labelled data cannot be easily divided by a hyperplane or if the nearest neighbors do not provide a clear estimate. Further (mathematical) information can be found in Breiman (2001).

CoPro workflow in a nutshell

Once all data and settings are provided, the simulations can be commenced following the workflow depicted in figure 8. Additional information and an interactive Python notebook can be found at https://copro.readthedocs.io/en/latest/examples/index.html. In a first step, the relation between indicators and conflict event data needs to be established. To that end, CoPro initially defines the study area and conflict events to be considered by applying the different model settings in a filtering step. Subsequently, CoPro will go through each year t_h of the historical period. Per year and polygon, the indicator and conflict datasets will be read, applying a 1 year time lag for the indicator data plus the variables 'Conflict in previous year' and 'Conflict in neighboring province in previous year' (see appendix B). This implies that the first year has to be skipped and merely serves as input to the second. Hence, the indicator data associated to t_h consists of the data observed for t_{h-1}. For the target conflict data itself no time lag is applied. Per polygon, CoPro produces a Binary value per indicator dataset representative for a given water province. This value is determined using common statistical methods such as the mean, max or min. It can also be opted for log-transforming the data. Both settings, that is the statistical method and whether indicator values should be log-transformed, must be provided in the config-file.

Once indicator and target data were sampled for the entire historical period, the scaling method is applied. Then, a user-specified number of model instances is trained with a user-defined fraction of this scaled data. The trained instances are stored to be used again for the projections. The other part of the scaled data is then used to make out-of-sample predictions of conflict occurrence and evaluate them using multiple metrics. By using and averaging across multiple model instances, a robust picture can be obtained of the accuracy of conflict risk predictions.

In a second step, CoPro projects conflict risk per year t_p between the end of the historical period until the year until which projections are ought to be made. Due to the 1 year time lag, the first projection year can still draw upon historical data. Afterwards, CoPro employs the scaled projected annual indicator data at t_{p-1} as model input together with simulated conflict risk at t_{p-1}. This is again executed for each model instance separately to output projected conflict risk at t_p. By again averaging across all model instances per year, CoPro yields one overall projection for each t_p. These out-of-sample forward projections are continued until the last year of the projection period is reached.

Appendix D

The development over time of projected mean probability of conflict over the entire continent of Africa is depicted in figure 9. Even for these continent-average aggregations, which mask large sub-continental differences (see figure 6), the 'erratic' nature of conflict onset is visible, mostly driven by variations in
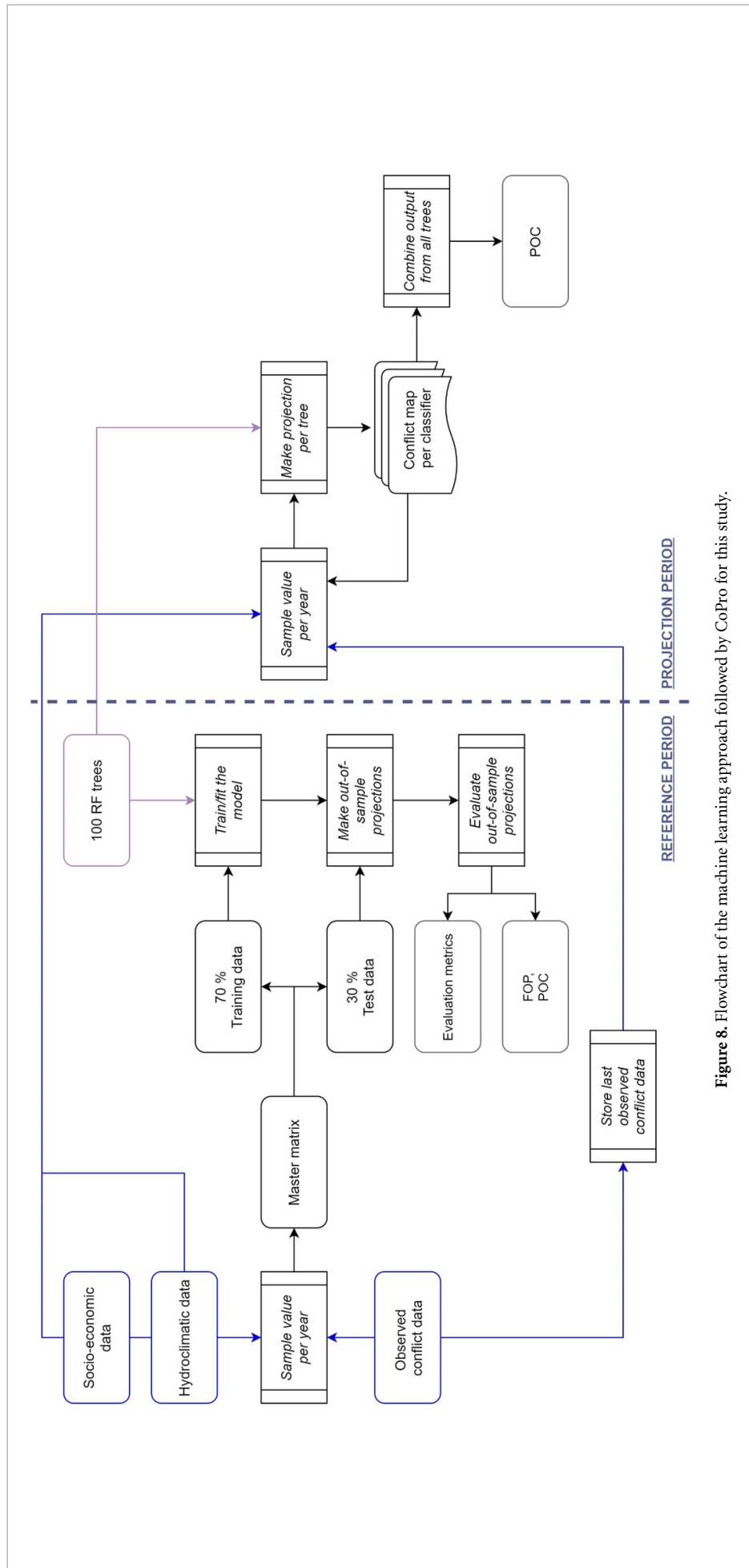
Figure 8. Flowchart of the machine learning approach followed by CoPro for this study.
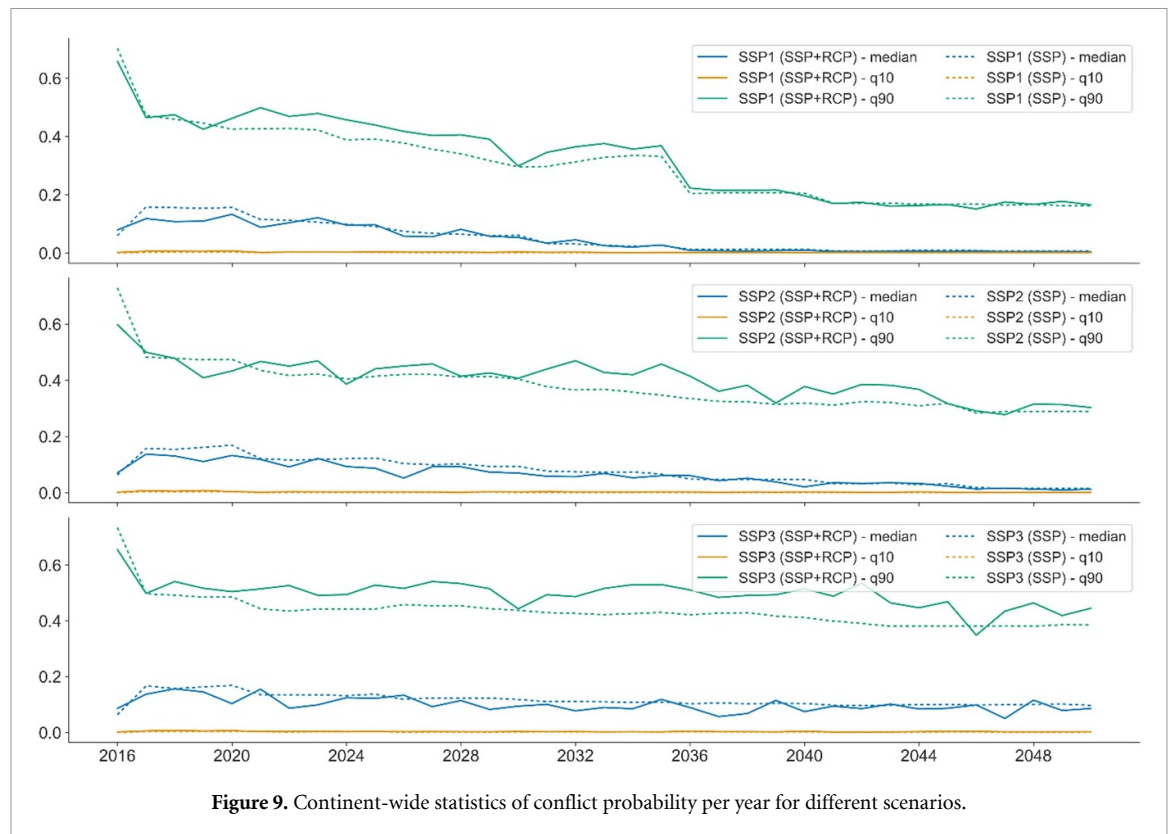
Figure 9. Continent-wide statistics of conflict probability per year for different scenarios.

the hydro-climatic variables which show a higher degree of year to year variability. Clear differences in trends can be observed between a sustainable development scenario (SSP1) and a scenario characterized by regional rivalry (SSP3). For the latter, results indicate that no pacification will occur until 2050, whereas the other scenarios show a clear downward trend. Additionally, continent averages show that the evolution of socioeconomic development is more influential for future conflict risk than hydro-climatological change, which is in line with other research (Witmer *et al* 2017, Koubi 2019, Mach *et al* 2019).

Appendix E

To assess how sensitive the model is to changes in the sampling method of the climate variables, we performed an additional 'sensitivity analysis' run (hereafter: SSP-RCP (SA) run), aggregating the climate variables in a different way. In these models, we use the minimum pixel-level values of the climate variables within water province boundaries for all climate variables, except for flood for which we take the maximum. These aggregations follow a 'weakest link' logic where groups may mobilize for violence anywhere within the water province when exposed to adverse impacts of climate variability and unfavorable conditions in some places cannot be made up for by favorable conditions in other locations. For the historical period, results differ only

very marginally from the default SSP-RCP run. Similarly, projected conflict risk for the period 2041–2050 differs primarily at the water province level (figure 10) while regional trends and model performance remain stable across sampling techniques (table 5).

Appendix F

Figure 11 provides detailed maps of the fraction of correct predictions (FOPs) for the SSP-RCP run in the reference period 1995–2015, the difference of FOP values between SSP only run and SSP-RCP run, and the number of observed conflict events per water province in this period. The FOP is determined as the number of correct predictions over number of total predictions. For more detail, see section 2.3.1. By comparing FOP values with the number of observed conflict events, it is possible to obtain an idea why CoPro yields high respectively low FOP values. Overall, the model yields highest FOP values where either a lot of conflict events are observed (in the south West of Somalia, for instance) or where only few or no events are observed, such as in the north West of Ethiopia. This aligns with the findings presented in figure 3 and section 3.1. Furthermore, we find that there is no FOP difference mostly for provinces with a low or high number of observed conflicts. Except from this, a clear relation between number of conflicts observed and FOP difference between runs cannot be derived from these data.
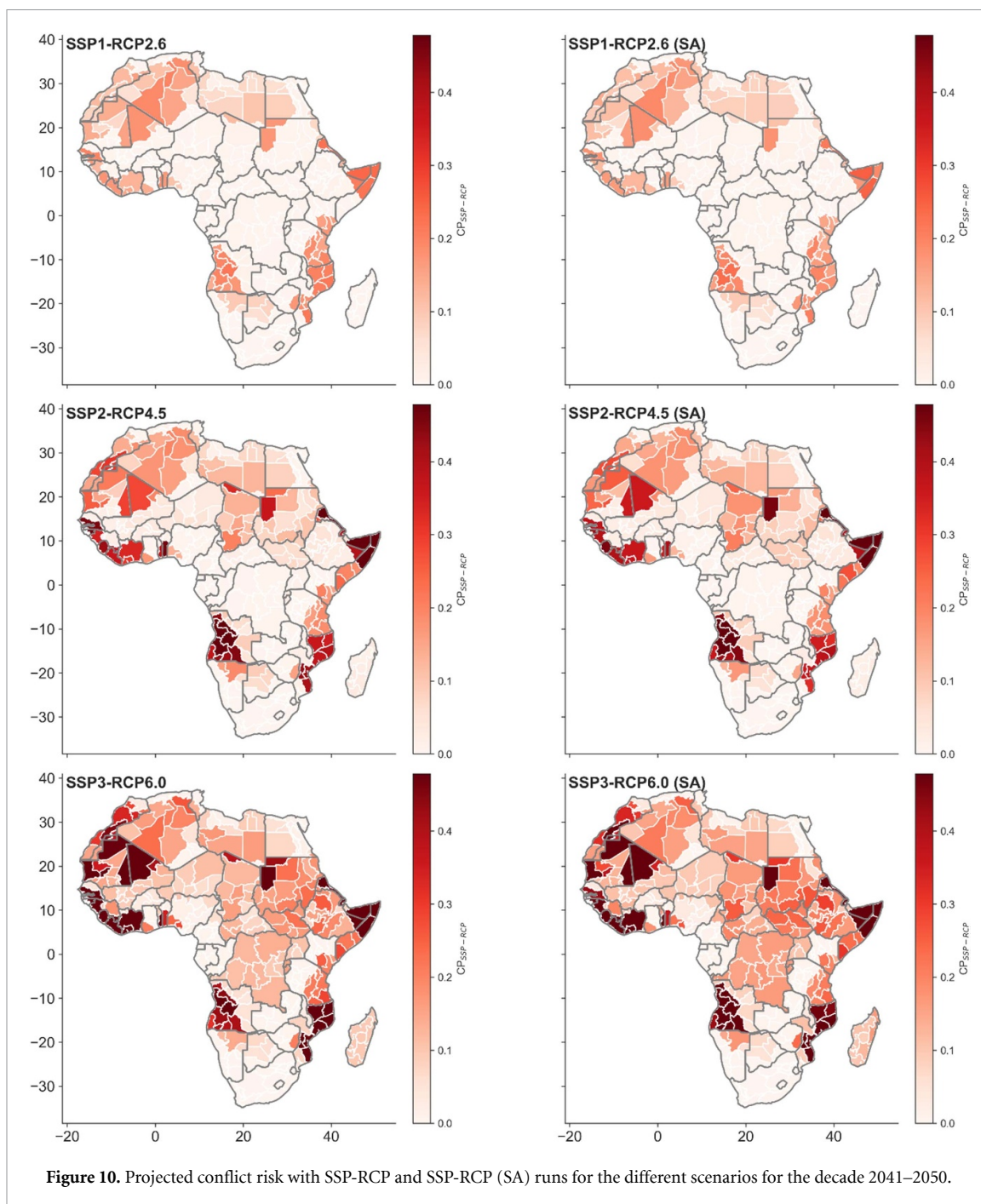
Figure 10. Projected conflict risk with SSP-RCP and SSP-RCP (SA) runs for the different scenarios for the decade 2041–2050.

Table 5. Overview of computed model evaluation scores for SSP-RCP and SSP-RCP (SA) run for the period 1995–2015.

	SSP-RCP run	SSP-RCP (SA) run
Average ROC-AUC score	0.90	0.90
Average precision-score	0.74	0.75
Average accuracy	0.87	0.87
Average precision	0.76	0.76
Average recall	0.62	0.61
Average Brier-score	0.10	0.09
Average FOP	0.87	0.87

Figure 11. (Top to bottom) maps of fraction of correct predictions (FOP) for SSP-RCP run, FOP difference between SSP run and SSP-RCP run, and log-scaled number of observed conflict events for the SSP-RCP run in the historical reference period, 1995–2015, for selected regions (from left to right): Nigeria, Somalia, Ethiopia, and the Sahel zone as defined by UN-OCHA. White areas indicate that no conflict events are observed.

ORCID iDs

Jannis M Hoch ⬡ https://orcid.org/0000-0003-3570-6436

Sophie P de Bruin ⬡ https://orcid.org/0000-0003-3429-349X

Halvard Buhaug ⬡ https://orcid.org/0000-0002-6432-5985

Nina Von Uexkull ⬡ https://orcid.org/0000-0001-9492-1596

Rens van Beek ⬡ https://orcid.org/0000-0002-4758-108X

Niko Wanders ⬡ https://orcid.org/0000-0002-7102-5454

References

Adams C, Ide T, Barnett J and Detges A 2018 Sampling bias in climate–conflict research *Nat. Clim. Change* **8** 200–3

Adger W N *et al* 2014 Human Security *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* Field C B *et al* (Cambridge: Cambridge University Press) pp 755–91 (https://ar5-syr.ipcc.ch/ipcc/resources/pdf/WGII/WGIIAR5-Chap12_FINAL.pdf)

Almer C, Laurent-Lucchetti J and Oechslin M 2017 Water scarcity and rioting: disaggregated evidence from Sub-Saharan Africa *J. Environ. Econ. Manage.* **86** 193–209

Andrijevic M, Crespo Cuaresma J, Muttarak R and Schleussner C-F 2020 Governance in socioeconomic pathways and its role for future adaptive capacity *Nat. Sustain.* **3** 35–41

Bara C 2014 Incentives and opportunities: a complexity-oriented explanation of violent ethnic conflict *J. Peace Res.* **51** 696–710

Barakat B and Urdal H 2009 Breaking the waves? Does education mediate the relationship between youth bulges and political violence ? Policy research working papers The World Bank (https://doi.org/10.1596/1813-9450-5114)

Basche A D, Kaspar T C, Archontoulis S V, Jaynes D B, Sauer T J, Parkin T B and Miguez F E 2016 Soil water improvements

with the long-term use of a winter rye cover crop *Agric. Water Manage.* **172** 40–50

Besley T and Persson T 2011 The logic of political violence *Q. J. Econ.* **126** 1411–45

Bowlsby D, Chenoweth E, Hendrix C and Moyer J D 2020 The future is a moving target: predicting political instability *Br. J. Polit. Sci.* **50** 1405–17

Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32

Brown G K 2011 The influence of education on violent conflict and peace: inequality, opportunity and the management of diversity *Prospects* **41** 191–204

Buhaug H *et al* 2014 One effect to rule them all? A comment on climate and conflict *Clim. Change* **127** 391–7

Buhaug H and Gleditsch K S 2008 Contagion or confusion? Why conflicts cluster in space1 *Int. Stud. Q.* **52** 215–33

Buhaug H and Vestby J 2019 On growth projections in the shared socioeconomic pathways *Glob. Environ. Polit.* **19** 118–32

Buhaug H and Von Uexkull N 2021 Vicious circles: violence, vulnerability, and climate change *Annu. Rev. Environ. Resour.* **46** 545–68

Cederman L-E, Gleditsch K S and Buhaug H 2013 *Inequality, Grievances, and Civil War* (Cambridge: Cambridge University Press)

Cederman L-E and Weidmann N B 2017 Predicting armed conflict time to adjust our expectations? *Science* **355** 474–6

Colaresi M and Mahmood Z 2017 Do the robot: lessons from machine learning to improve conflict forecasting *J. Peace Res.* **54** 193–214

Conca K 2019 Is there a role for the UN security council on climate change? *Environ. Sci. Pol. Sustain. Dev.* **61** 4–15

de Bruin S P, Hoch J M, Von Uexkull N, Buhaug H, Demmers J, Visser H and Wanders N 2021 Projecting long-term armed conflict risk: an underappreciated field of inquiry? *Glob. Environ. Change* **72** 102423

Döring S 2020 Come rain, or come wells: how access to groundwater affects communal violence *Polit. Geogr.* **76** 102073

Fearon J D and Laitin D D 2003 Ethnicity, insurgency, and civil war *Am. Polit. Sci. Rev.* **97** 75–90

Goldstone J A, Bates R H, Epstein D L, Gurr T R, Lustik M B, Marshall M G, Ulfelder J and Woodward M 2010 A global model for forecasting political instability *Am. J. Polit. Sci.* **54** 190–208

Hegre H *et al* 2019 ViEWS: a political violence early-warning system *J. Peace Res.* **56** 155–74

Hegre H *et al* 2021 ViEWS2020: revising and evaluating the ViEWS political violence early-warning system *J. Peace Res.* **58** 599–611

Hegre H, Buhaug H, Calvin K V, Nordkvelle J, Waldhoff S T and Gilmore E 2016 Forecasting civil conflict along the shared socioeconomic pathways *Environ. Res. Lett.* **11** 054002

Hegre H, Karlsen J, Nygård H M, Strand H and Urdal H 2013 Predicting armed conflict, 2010–2050 *Int. Stud. Q.* **57** 250–70

Hegre H, Metternich N W, Nygård H M and Wucherpfennig J 2017 Introduction: forecasting in peace research *J. Peace Res.* **54** 113–24

Hegre H and Sambanis N 2006 Sensitivity analysis of empirical results on civil war onset *J. Conflict Resolut.* **50** 508–35

Hirabayashi Y, Mahendran R, Koirala S, Konoshima L, Yamazaki D, Watanabe S, Kim H and Kanae S 2013 Global flood risk under climate change *Nat. Clim. Change* **3** 816–21

Hoch J M, de Bruin S and Wanders N 2021a CoPro: a data-driven modelling framework for conflict risk projections *J. Open Source Softw.* **6** 2855

Hoch J M, de Bruin S and Wanders N 2021b CoPro v0.1.1. Zenodo (https://doi.org/10.5281/zenodo.4972327)

Hsiang S M, Burke M and Miguel E 2013 Quantifying the influence of climate on human conflict *Science* **341** 1235367

Ide T, Kristensen A and Bartusevičius H 2021 First comes the river, then comes the conflict? A qualitative comparative analysis of flood-related political unrest *J. Peace Res.* **58** 83–97

IPCC 2019 *Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems* (Geneva: IPCC) (www.ipcc.ch/srccl)

Jones B and O'Neill B C 2016 Spatially explicit global population scenarios consistent with the shared socioeconomic pathways *Environ. Res. Lett.* **11** 084003

Koubi V 2019 Climate change and conflict *Ann. Rev. Polit. Sci.* **22** 343–60

Mach K J, Kraan C M, Adger W N, Buhaug H, Burke M, Fearon J D, Field C B, Hendrix C S, Maystadt J-F and O'Loughlin J 2019 Climate as a risk factor for armed conflict *Nature* **571** 193–7

Mahmoud M, Liu Y, Hartmann H, Stewart S, Wagener T, Semmens D, Stewart R, Gupta H, Dominguez D and Dominguez F 2009 A formal framework for scenario development in support of environmental decision-making *Environ. Model. Softw.* **24** 798–808

Maier H R, Guillaume J H, van Delden H, Riddell G A, Haasnoot M and Kwakkel J H 2016 An uncertain future, deep uncertainty, scenarios, robustness and adaptation: how do they fit together? *Environ. Model. Softw.* **81** 154–64

Masui T, Matsumoto K, Hijioka Y, Kinoshita T, Nozawa T, Ishiwatari S, Kato E, Shukla P R, Yamagata Y and Kainuma M 2011 An emission pathway for stabilization at 6 Wm^{-2} radiative forcing *Clim. Change* **109** 59

Murakami D and Yamagata Y 2019 Estimation of gridded population and gdp scenarios with spatially explicit statistical downscaling *Sustainability* **11** 2106

O'Neill B C *et al* 2017 The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century *Glob. Environ. Change* **42** 169–80

O'Neill B C *et al* 2020 Achievements and needs for the climate change scenario framework *Nat. Clim. Change* **10** 1074–84

Otto I M, Reckien D, Reyer C P O, Marcus R, Le Masson V, Jones L, Norton A and Serdeczny O 2017 Social vulnerability to climate change: a review of concepts and evidence *Reg. Environ. Change* **17** 1651–62

Pettersson T and Öberg M 2020 Organized violence, 1989–2019 *J. Peace Res.* **57** 597–613

Raleigh C, Choi H J and Kniveton D 2015 The devil is in the details: an investigation of the relationships between conflict, food price and climate across Africa *Glob. Environ. Change* **32** 187–99

Raleigh C and Hegre H 2009 Population size, concentration, and civil war. A geographically disaggregated analysis *Polit. Geogr.* **28** 224–38

Rao N D, Sauer P, Gidden M and Riahi K 2019 Income inequality projections for the shared socioeconomic pathways (SSPs) *Futures* **105** 27–39

Schutte S and Weidmann N B 2011 Diffusion patterns of violence in civil wars *Polit. Geogr.* **30** 143–52

Scott S V 2015 Implications of climate change for the UN security council: mapping the range of potential policy responses *Int. Aff.* **91** 1317–33

Silva L L 2017 Are basin and reservoir tillage effective techniques to reduce runoff under sprinkler irrigation in mediterranean conditions? *Agric. Water Manage.* **191** 50–56

Straatsma M, Droogers P, Hunink J, Berendrecht W, Buitink J, Buytaert W, Karssenberg D, Schmitz O, Sutanudjaja E H and van Beek L P H 2020 Global to regional scale evaluation of adaptation measures to reduce the future water gap *Environ. Model. Softw.* **124** 104578

Sundberg R and Melander E 2013 Introducing the UCDP georeferenced event dataset *J. Peace Res.* **50** 523–32

Sutanudjaja E H *et al* 2018 PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model *Geosci. Model Dev.* **11** 2429–53

Teutschbein C and Seibert J 2012 Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods *J. Hydrol.* **456–457** 12–29

Thomson A M *et al* 2011 RCP4.5: a pathway for stabilization of radiative forcing by 2100 *Clim. Change* **109** 77

Tyralis H, Papacharalampous G and Langousis A 2019 A brief review of random forests for water scientists and practitioners and their recent history in water resources *Water* **11** 910

van Beek L, Hajer M, Pelzer P, van Vuuren D and Cassen C 2020 Anticipating futures through models: the rise of integrated assessment modelling in the climate science-policy interface since 1970 *Glob. Environ. Change* **65** 102191

van Vuuren D P *et al* 2011 RCP2.6: exploring the possibility to keep global mean temperature increase below 2 °C *Clim. Change* **109** 95

van Vuuren D P *et al* 2014 A new scenario framework for climate change research: scenario matrix architecture *Clim. Change* **122** 373–86

Vestby J, Buhaug H and Von Uexkull N 2021 Why do some poor countries see armed conflict while others do not? A dual sector approach *World Dev.* **138** 105273

Von Uexkull N 2014 Sustained drought, vulnerability and civil conflict in Sub-Saharan Africa *Polit. Geogr.* **43** 16–26

Von Uexkull N and Buhaug H 2021 Security implications of climate change: a decade of scientific progress *J. Peace Res.* **58** 3–17

Von Uexkull N, Croicu M, Fjelde H and Buhaug H 2016 Civil conflict sensitivity to growing-season drought *Proc. Natll Acad. Sci.* **113** 12391–6

Walter B F 2015 Why bad governance leads to repeat civil war *J. Conflict Resolut.* **59** 1242–72

Wanders N, Wada Y and Van Lanen H A J 2015 Global hydrological droughts in the 21st century under a changing hydrological regime *Earth Syst. Dyn.* **6** 1–15

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O and Schewe J 2014 The inter-sectoral impact model intercomparison project (ISI–MIP): project framework *Proc. Natll Acad. Sci.* **111** 3228–32

Witmer F D, Linke A M, O'Loughlin J, Gettelman A and Laing A 2017 Subnational violent conflict forecasts for sub-Saharan Africa, 2015–65, using climate-sensitive models *J. Peace Res.* **54** 175–92

Wittgenstein Centre for Demography and Global Human Capital, 2018 Wittgenstein centre human capital data explorer [WWW Document] (available at: http://dataexplorer. wittgensteincentre.org/wcde-v2/) (Accessed 2 December 2021)

WPS Partnership 2021 Water, peace and security [WWW Document]. About the water, peace and security partnership (available at: https://waterpeacesecurity.org/ info/about-wps) (Accessed 2 December 2021)