UPPSALA UNIVERSITET

ViEWS PREDICTING CONFLICT

Constituent models, *pgm*.

Online appendix C to ViEWS$_{2020}$: Revising and evaluating the ViEWS political Violence Early-Warning System.

Håvard Hegre[1,2], Curtis Bell[1,3], Michael Colaresi[1,4], Mihai Croicu[1], Frederick Hoyles[1], Remco Jansen[1], Maxine Ria Leis[1], Angelica Lindqvist-McGowan[1], David Randahl[1], Espen Geelmuyden Rød[1], and Paola Vesco[1]

[1]Department of Peace and Conflict Research, Uppsala University
[2]Peace Research Institute Oslo
[3]U.S. Naval War College
[4]University of Pittsburgh

January 21, 2021

Abstract

This appendix documents all the constituent models in the ensembles at the PRIO-GRID-month (*pgm*) level. We give an overview of the main estimation/machine-learning techniques. We describe and motivate each model, report the importance of the main features, and show prediction maps for selected steps for the 2020–2022 period.

CONTENTS

C-1 ESTIMATION OF CONSTITUENT MODELS

ViEWS relies on logistic regression, random forest (RF) and gradient boosting machines (GBM) models. The logit model is a generalized linear model (GLM) that performs well compared to many machine-learning techniques (Géron, 2017). Computational costs are low, and with large datasets like ours overfitting is not a serious concern. Both RF models (Breiman, 2001; Muchlinski et al., 2016) and GBM models (Friedman, 2001; Chen and Guestrin, 2016) are machine-learning techniques employing classification and regression trees (CART). In CART, the response variable Y is predicted using a decision tree and some predictor variables X. The tree consists of a number of 'splits' into different branches. Each split is found by searching all values in X to find the constant which maximally separates between the categories of Y. The tree continues to be split until some threshold is achieved (to avoid overfitting).

In RF models, CART trees are combined with bootstrap-aggregating (bagging), and random feature selection. Bagging creates an ensemble of trees, each slightly different. These trees are, however, correlated as some variables are especially good at discriminating Y. To avoid this, a random subset of variables (predictors or features) are selected for each 'split', solving the correlation problem and creating a forest of uncorrelated 'random' trees.

In GBM models, an ensemble of CART trees are trained in an additive, iterative fashion. At each iteration, a new CART tree (a new learner) is added to the ensemble, grown in such a way as to minimize a chosen, derivable, loss function capturing remaining classification or regression error. In practice, this is done by training each new tree on a regularized form of the residuals computed between the ensemble predictions at the previous step and the observed values. Each such new tree is added to the ensemble, and the iterative process is restarted (Friedman, 2001; Chen and Guestrin, 2016). This iterative improvement is referred to as 'boosting'; this iterative minimization of residual-based loss functions is equivalent with gradient descent. ViEWS uses the XGBoost implementation of GBM, an implementation that adds a regularized learning objective adding a complexity penalty to tree growing and smoothing final learned weights, as well as implementing the same random feature selection and sub-sampling of data (Chen and Guestrin, 2016).

RF and XGB models are memory-intensive to estimate given the computational capacity available for ViEWS. Hence, at the *pgm* level, we estimate them using a 'downsampled' dataset which includes all conflict events and a random sample of non-events (see details in the 'Overview of models' section in this appendix).

C-2 HANDLING FORECASTING DYNAMICS

ViEWS employs two alternative strategies to compute forecasts for each of $s \in (1, .., 36)$ months into the future. We call these Dynamic simulation (*ds*) and One-step-ahead (*osa*).

Dynamic simulation

The dynamic simulation (Dynasim) procedure builds on Hegre et al. (2013) and Hegre et al. (2016), and is discussed at length there.[1] The procedure involves simulating model parameters based on the estimated coefficients and the variance-covariance matrix from a model. In addition, we compute predicted probabilities for the outcomes for the first month t, draw outcomes, recalculate the history variables so that the input predictor matrix X_{t+1} at $t+1$ reflects that draw. This is repeated for each month s in the forecasting window, and for each simulated set of parameters.

If we are interested in forecasting two months into the forecasting period, we first train the constituent models, estimate the weights, and produce our ensemble one-month-ahead forecast. To produce forecasts

[1]The first author's original script 'PRIOsim' was rewritten in C# and C++ by Joakim Karlsen. The Python routines underlying the current projects are based on the 'Dynasim' reimplementation of this, written by Jonas Vestby and Frederick Hoyles.

for the next month, we need the input predictor matrices $X_{t+1}^{(k)}$. For many constituent models, these input predictors will themselves be functions of actual conflict (e.g., lagged conflict indicators, time since last conflict, spatial distance to nearest conflict). Since these do not exist for the next month (after the training window), we use the prediction as the probability of an unobserved predictor, for example for conflict at time $t+1$, when forecasting conflict at $t+2$. A simulated value is drawn from this probability, and recorded within a new simulated set of predictors $\widetilde{X}_{t+1}^{(k)}$.

The predictions for the three outcomes are obtained simultaneously within each time step. For each of these, we compute the predicted probability at $t+1$ as a function of information available at t, including the status for the other two outcomes. This procedure repeats for every month to the end of the forecasting window.

'One-step-ahead' modeling

In the 'one-step-ahead' modeling, we predict each step into the future ($t+1$, (...), $t+12$, (...) $t+36$) independently, as opposed to dynamic simulation which moves forward sequentially. We do this by estimating a set of models of the form $f_s(X_{t-s})$ where s denotes the number of months into the future to forecast. In regression notation these take the form $y_{t+s} = X_t\beta_t,\ for\ s \in (1, 36)$. The 'one-step-ahead' model does this by time-shifting the right-hand side variables with respect to the outcome before models are trained, making the model a link function between the future (y_{t+s}) and the present (X_t).

C-3 OVERVIEW OF *PGM* MODELS

Pgm models are built according to themes which can be combined in increasing levels of complexity. The basic themes include History, Sptime, Institutional, Oil, Natural Geography, Social Geography and SPEI (drought).

Model name
hist_legacy
sptime
onset24_100_all
onset24_1_all
pgd_natural
pgd_social
spei_full
ds_25
ds_dummy
allthemes
crosslevel
xgb

Table C-1. Models in the *cm* ensemble as of r.2020.02.01

We have defined a set of 'models' characterized by a set of features as well as an estimation procedure. All models are trained for each of the three different types of conflict, i.e. state-based **(sb)**, non-state **(ns)** and one-sided **(os)** violence, as well as for all steps s.

If nothing else is stated, the estimator is the random forest classifier (Breiman, 2001) from the scikit-learn package (Pedregosa et al., 2011). All parameters used are the package defaults, except for $n\_estimators$ which controls the number of trees in the forest and is set to 1000.

With these settings the procedure works as follows: for each of 1000 trees, draw a bootstrapped sample of the same size as the training data. On this sample, fit a decision tree using the gini split quality criterion, considering the square root of the number of features for each split. Predicted probabilities from the forest are computed as the average of the predicted probability from each tree, which is the fraction of samples for the class in the current leaf of the tree.

Feature importances are a measure of the importance of each feature in the forest. Below, we present impurity-based features importances.[2] The importances of all features sum to one and are assigned according to the position of the feature in the trained trees, with features appearing higher up in the trees receiving a higher score as the expected fraction of samples they contribute to predicting is higher.

Below, we present a description of the model features used as predictors for the different dependent variables. The following sections will provide further details on the types of outcome which is predicted.

C-3.1 History (hist_legacy)

Description

The history model traces the conflict history of a grid-cell, based on the evidence showing a higher conflict incidence in locations which experienced conflict in the past (Collier, Hegre, and Raknerud, 2003). The model includes 1-12 temporal lags of the specific dependent variable, such as incidence of state-based violence. Moreover, the specific outcome is temporally lagged by 1-3 months for the cells adjacent to the conflictual grid cell. Also included are the other dependent variables lagged by 1 month, both for each grid cell and all the neighboring cells. Finally, the model includes a 12-months half-life decay function of the time since a conflict event occurred for all dependent variables.

Estimation details

The model uses the 'one-step-ahead' approach (see Section C-2) and is estimated on data for Africa only.

Feature importances

The tables below report the feature importances for each type of violence and different time steps ($s = 1, 3, 6, 12, 24, 36$). Please note that for broad models like *hist_legacy*, the tables report only the first 25 most important features. The 12-months decay function of the last conflict episode is the most important feature to predict violence of any type. The most important feature is not strictly outcome-specific; time since last **sb** conflict is the most relevant variable to predict both **sb** and **os**, while time since last **os** and **ns** violence are the most relevant variable to predict **ns** conflict.

[2] For a critical review of impurity-based feature importances, see Strobl et al. (2007) and Strobl et al. (2008)

sb_hist_legacy	1	3	6	12	24	36
decay_12_time_since_ged_dummy_sb	0.121231	0.119443	0.117021	0.124543	0.127912	0.134342
decay_12_time_since_ged_dummy_os	0.112707	0.117378	0.121663	0.128593	0.137806	0.151205
decay_12_time_since_ged_dummy_ns	0.068922	0.072056	0.075678	0.081365	0.091652	0.102848
tlag_1_ged_dummy_sb	0.067318	0.065875	0.062689	0.054182	0.041584	0.039771
decay_12_time_since_acled_dummy_pr	0.061117	0.063602	0.065700	0.069556	0.074375	0.081557
tlag_3_ged_dummy_sb	0.051518	0.039368	0.048652	0.045230	0.040391	0.026612
tlag_2_ged_dummy_sb	0.050512	0.055556	0.046442	0.044761	0.047562	0.038962
tlag_4_ged_dummy_sb	0.043669	0.043064	0.040543	0.033727	0.034569	0.030552
tlag_3_splag_1_1_ged_dummy_sb	0.037127	0.036443	0.036217	0.036648	0.037098	0.039536
tlag_1_splag_1_1_ged_dummy_sb	0.036771	0.036650	0.035561	0.036323	0.037681	0.038045
tlag_2_splag_1_1_ged_dummy_sb	0.036729	0.036236	0.036115	0.035924	0.037551	0.038349
tlag_9_ged_dummy_sb	0.036539	0.029030	0.034238	0.035560	0.030200	0.026418
tlag_8_ged_dummy_sb	0.036501	0.034741	0.029745	0.027167	0.029051	0.029455
tlag_7_ged_dummy_sb	0.034952	0.036619	0.034124	0.036397	0.026527	0.029480
tlag_5_ged_dummy_sb	0.034915	0.037552	0.039093	0.034391	0.033162	0.031315
tlag_6_ged_dummy_sb	0.034385	0.035929	0.035640	0.030899	0.029772	0.028529
tlag_10_ged_dummy_sb	0.029613	0.027659	0.027620	0.034612	0.026498	0.022936
tlag_1_splag_1_1_ged_dummy_os	0.026441	0.026612	0.026404	0.026331	0.028392	0.029392
tlag_11_ged_dummy_sb	0.026185	0.026413	0.029094	0.024950	0.027053	0.021653
tlag_12_ged_dummy_sb	0.022991	0.028948	0.027576	0.028740	0.029322	0.026476
tlag_1_splag_1_1_acled_dummy_pr	0.015094	0.015875	0.015183	0.015667	0.016526	0.017133
tlag_1_splag_1_1_ged_dummy_ns	0.006802	0.007045	0.007094	0.006503	0.007159	0.007094
tlag_1_ged_dummy_os	0.006197	0.006164	0.006226	0.006192	0.006434	0.006518
tlag_1_ged_dummy_ns	0.001765	0.001741	0.001684	0.001740	0.001723	0.001823

Table C-2. Feature importances for the history model, **sb**

os_hist_legacy	1	3	6	12	24	36
decay_12_time_since_ged_dummy_sb	0.147846	0.155234	0.162215	0.172716	0.185132	0.196111
decay_12_time_since_ged_dummy_os	0.123162	0.126667	0.129812	0.136369	0.146676	0.156496
decay_12_time_since_ged_dummy_ns	0.120364	0.126126	0.133293	0.142658	0.151221	0.162635
decay_12_time_since_acled_dummy_pr	0.092971	0.096245	0.098423	0.100444	0.104508	0.109688
tlag_1_splag_1_1_ged_dummy_sb	0.048805	0.047756	0.047556	0.045033	0.044509	0.040315
tlag_1_splag_1_1_ged_dummy_os	0.048682	0.046253	0.043965	0.042208	0.041186	0.038556
tlag_3_splag_1_1_ged_dummy_os	0.046852	0.046058	0.044464	0.042540	0.040347	0.037783
tlag_2_splag_1_1_ged_dummy_os	0.046820	0.045120	0.044127	0.043404	0.040565	0.036952
tlag_1_splag_1_1_acled_dummy_pr	0.029159	0.028074	0.027072	0.028041	0.027987	0.026682
tlag_1_ged_dummy_os	0.028104	0.025499	0.023043	0.017774	0.013709	0.012631
tlag_2_ged_dummy_os	0.026436	0.024043	0.020746	0.018228	0.015578	0.012837
tlag_5_ged_dummy_os	0.023677	0.019453	0.019823	0.017167	0.014255	0.013309
tlag_3_ged_dummy_os	0.023588	0.024263	0.021586	0.018947	0.014648	0.013357
tlag_4_ged_dummy_os	0.023281	0.021775	0.018347	0.017288	0.015333	0.014145
tlag_1_splag_1_1_ged_dummy_ns	0.021444	0.020896	0.020785	0.019284	0.018218	0.015913
tlag_6_ged_dummy_os	0.020947	0.019773	0.021593	0.016769	0.014789	0.014504
tlag_11_ged_dummy_os	0.019951	0.017989	0.017307	0.017951	0.016400	0.013850
tlag_7_ged_dummy_os	0.019010	0.017970	0.017357	0.016928	0.015343	0.013167
tlag_8_ged_dummy_os	0.018790	0.018987	0.018010	0.017048	0.016768	0.014295
tlag_10_ged_dummy_os	0.018149	0.017651	0.017923	0.018861	0.014983	0.013977
tlag_9_ged_dummy_os	0.017813	0.019777	0.018331	0.016505	0.015767	0.013995
tlag_12_ged_dummy_os	0.017578	0.018570	0.018072	0.018173	0.016870	0.015414
tlag_1_ged_dummy_sb	0.010938	0.010681	0.010882	0.010409	0.010528	0.009396
tlag_1_ged_dummy_ns	0.005633	0.005140	0.005270	0.005257	0.004682	0.003990

Table C-3. Feature importances for the history model, **os**

Change history

The model has been in use since r.2020.02.01

C-3.2 Spacetime (sptime)

Description

The spacetime model includes a set of features that reflect both time and space proximity to conflict and thus comprehensively captures the tendency of conflicts to cluster in space and time (Buhaug and Gleditsch, 2008). A 'spacetime' feature maps the distance to a conflict episode across three dimensions, i.e. longitude,

ns_hist_legacy	1	3	6	12	24	36
decay_12_time_since_ged_dummy_os	0.146669	0.154492	0.159657	0.167275	0.177262	0.181365
decay_12_time_since_ged_dummy_ns	0.126797	0.128753	0.135104	0.142163	0.155602	0.166558
decay_12_time_since_ged_dummy_sb	0.115306	0.118566	0.120641	0.120185	0.123179	0.127904
decay_12_time_since_acled_dummy_pr	0.104075	0.106462	0.106806	0.107641	0.106343	0.113010
tlag_1_ged_dummy_ns	0.040552	0.028038	0.025057	0.023947	0.017046	0.015778
tlag_1_splag_1_1_ged_dummy_ns	0.037741	0.038204	0.034405	0.039750	0.045939	0.048320
tlag_3_splag_1_1_ged_dummy_ns	0.036659	0.033044	0.037391	0.039577	0.044778	0.047125
tlag_2_splag_1_1_ged_dummy_ns	0.034698	0.033552	0.035295	0.044930	0.046209	0.044969
tlag_1_splag_1_1_acled_dummy_pr	0.032681	0.030610	0.029228	0.029617	0.031671	0.028809
tlag_2_ged_dummy_ns	0.030904	0.032474	0.025427	0.019968	0.018908	0.018265
tlag_4_ged_dummy_ns	0.030325	0.023098	0.028586	0.023489	0.015699	0.016597
tlag_1_splag_1_1_ged_dummy_os	0.026589	0.025999	0.025442	0.025598	0.024778	0.026362
tlag_9_ged_dummy_ns	0.025401	0.025758	0.017888	0.023880	0.021565	0.014023
tlag_8_ged_dummy_ns	0.024421	0.028771	0.021433	0.018288	0.018579	0.015360
tlag_3_ged_dummy_ns	0.024212	0.027350	0.026309	0.021018	0.019326	0.015155
tlag_5_ged_dummy_ns	0.023981	0.022973	0.026937	0.017003	0.016282	0.013546
tlag_10_ged_dummy_ns	0.023721	0.022652	0.027341	0.022045	0.017884	0.015825
tlag_11_ged_dummy_ns	0.022546	0.019509	0.018841	0.022052	0.017724	0.015942
tlag_12_ged_dummy_ns	0.021999	0.017345	0.020502	0.024026	0.018679	0.018482
tlag_6_ged_dummy_ns	0.021803	0.026471	0.025325	0.019579	0.017503	0.015826
tlag_7_ged_dummy_ns	0.019736	0.027091	0.024413	0.020364	0.019195	0.014202
tlag_1_splag_1_1_ged_dummy_sb	0.017531	0.017532	0.017384	0.017482	0.016071	0.016460
tlag_1_ged_dummy_os	0.006798	0.006557	0.006032	0.006082	0.006009	0.006296
tlag_1_ged_dummy_sb	0.004854	0.004697	0.004556	0.004043	0.003769	0.003822

Table C-4. Feature importances for the history model, **ns**

latitude (the spatial coordinates, measured in spatial degrees) and the time-span. The spacetime feature is therefore a synthetic, multidimensional representation of distance, where each of the three dimensions contribute to the final measure of proximity/distance to conflict. The model includes a number of features with different weightings of time proximity for model space proximity. In one such spacetime feature, a violent event that occurred in cell *x* at time $t-1$ is considered to be equally distant in spacetime as a violent event that occurred in the same time-period t in a location which is 1 spatial degree away from the reference point. In each case, the spacetime distance is equal to 1. In other features included in the model, we vary the scale of the time dimension by multiplying the time dimension of this event space by a scalar (we use 0.01, 1 and 10). By shrinking this time scaling factor we can put temporally distant events relatively closer to each grid month, making the space-time tend toward the time since event in the same cell by making spatial distances comparatively large. Conversely, scaling time up makes the feature tend closer to a pure spatial distance where events further back in time are rarely closer than spatially proximate events. The spacetime feature algorithm is essentially a nearest neighbor algorithm and so in principle supports computing the average distance to the k nearest events. We have limited ourselves to the $k = 1$ case of the single nearest neighbor. For maps demonstrating the shifted time scale, see Figure C-1. Included is two variants: space time only (sptime) and one including all themes except history (sptime_and_all_themes).

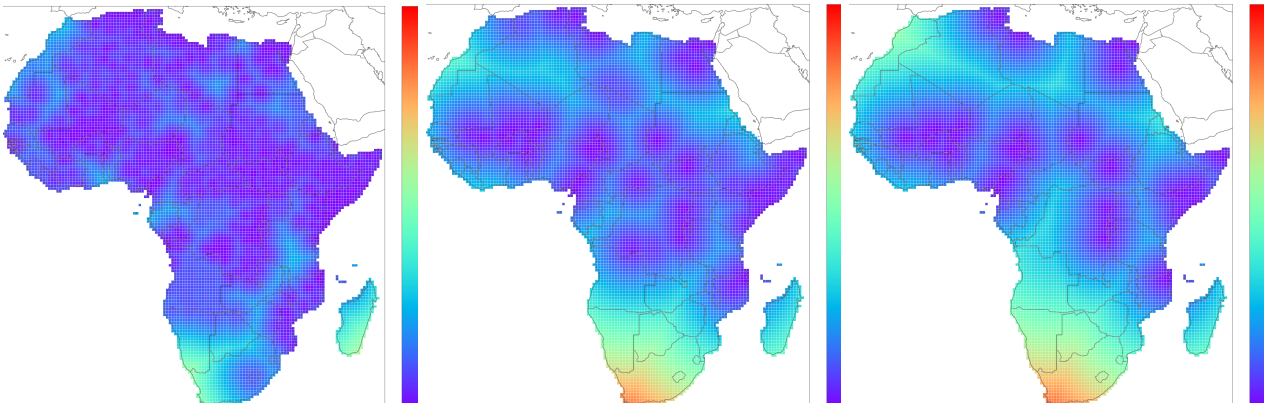
Estimation details

The model uses the 'one-step-ahead' approach (see Section C-2) and is estimated on data for Africa only.

Feature importances

The table below reports the feature importances for each type of violence and different time steps ($s = 1, 3, 6, 12, 24, 36$). Time distance is the most important feature to predict **sb** violence, slightly overcoming the importance of spatial lags. Spatial proximity to the outcome-specific violent is more relevant to predict **os** and **ns** violence.

Figure C-1. Maps illustrating scaling of the time dimension of the spacetime 'distance to event' feature.



(a) Time scale factor = 0.01 *(b)* Time scale factor = 1 *(c)* Time scale factor = 10

sb_sptime	1	3	6	12	24	36
time_since_ged_dummy_sb	0.107512	0.107484	0.123020	0.100964	0.094118	0.094741
stdist_k1_t001_ged_dummy_sb	0.105428	0.098914	0.088711	0.089833	0.083591	0.076906
stdist_k1_t10_ged_dummy_sb	0.096320	0.089038	0.080690	0.078302	0.072322	0.068418
stdist_k1_t001_ged_dummy_ns	0.094259	0.096035	0.097025	0.097669	0.101416	0.101881
stdist_k1_t001_ged_dummy_os	0.087747	0.092190	0.096659	0.095350	0.092889	0.087289
time_since_ged_dummy_ns	0.086342	0.090762	0.094679	0.101856	0.117246	0.121385
stdist_k1_t1_ged_dummy_sb	0.084704	0.079387	0.067897	0.074801	0.066150	0.062713
stdist_k1_t10_ged_dummy_ns	0.077513	0.078521	0.078629	0.078363	0.078601	0.079716
stdist_k1_t1_ged_dummy_ns	0.075472	0.077185	0.078402	0.079455	0.078083	0.079257
time_since_ged_dummy_os	0.066706	0.068698	0.072131	0.079068	0.088923	0.097605
stdist_k1_t1_ged_dummy_os	0.060491	0.063593	0.062384	0.063927	0.064310	0.065569
stdist_k1_t10_ged_dummy_os	0.057506	0.058193	0.059772	0.060412	0.062352	0.064519

os_sptime	1	3	6	12	24	36
stdist_k1_t001_ged_dummy_ns	0.112812	0.114766	0.115158	0.115528	0.115045	0.112240
time_since_ged_dummy_ns	0.109756	0.111741	0.110372	0.108476	0.107868	0.107639
stdist_k1_t10_ged_dummy_ns	0.100421	0.100341	0.100017	0.098784	0.096962	0.094419
stdist_k1_t1_ged_dummy_ns	0.093822	0.094733	0.095137	0.093507	0.093538	0.089860
time_since_ged_dummy_os	0.093048	0.090691	0.085692	0.084298	0.084892	0.087321
time_since_ged_dummy_sb	0.081286	0.083167	0.084491	0.086063	0.089744	0.092364
stdist_k1_t001_ged_dummy_sb	0.079351	0.083042	0.085573	0.088174	0.086615	0.087632
stdist_k1_t001_ged_dummy_os	0.078056	0.075618	0.075214	0.074504	0.075406	0.076562
stdist_k1_t10_ged_dummy_os	0.067830	0.064476	0.065007	0.064119	0.064749	0.065973
stdist_k1_t10_ged_dummy_sb	0.064353	0.066450	0.066428	0.067369	0.066254	0.065090
stdist_k1_t1_ged_dummy_os	0.061382	0.054244	0.055631	0.055922	0.057189	0.059770
stdist_k1_t1_ged_dummy_sb	0.057883	0.060731	0.061282	0.063257	0.061738	0.061129

ns_sptime	1	3	6	12	24	36
stdist_k1_t001_ged_dummy_sb	0.109862	0.117961	0.119419	0.118146	0.116368	0.113810
time_since_ged_dummy_ns	0.106852	0.100830	0.098596	0.095580	0.092569	0.091319
time_since_ged_dummy_sb	0.088366	0.093323	0.094842	0.095905	0.098549	0.101799
stdist_k1_t001_ged_dummy_ns	0.086110	0.082499	0.081475	0.084789	0.084629	0.085001
stdist_k1_t10_ged_dummy_ns	0.086016	0.077823	0.078586	0.076756	0.077869	0.075872
time_since_ged_dummy_os	0.082231	0.080108	0.080875	0.081867	0.081638	0.085535
stdist_k1_t10_ged_dummy_sb	0.079340	0.078824	0.077734	0.076918	0.076755	0.077052
stdist_k1_t001_ged_dummy_os	0.079320	0.081461	0.082591	0.083248	0.084462	0.085120
stdist_k1_t1_ged_dummy_sb	0.077264	0.078349	0.077160	0.076874	0.076067	0.076416
stdist_k1_t10_ged_dummy_os	0.070087	0.071555	0.071920	0.071985	0.072631	0.071620
stdist_k1_t1_ged_dummy_ns	0.068134	0.068698	0.067232	0.069085	0.068662	0.068289
stdist_k1_t1_ged_dummy_os	0.066419	0.068568	0.069570	0.068848	0.069800	0.068167

Table C-5. Feature importances for the Spacetime model, **sb**, **os**, **ns**

Change history

The model has been in use since r.2020.02.01

C-3.3 Onset (onset24_1_all, onset24_100_all)

Description

The Onset models are run for two dependent outcomes: at least one BRD and at least 100 BRDs. An onset is defined as the first time a specific grid, or its neighbors, reached the threshold in a 24-month sliding window. Uses the features from the *all* model.

Estimation details

The model uses the 'one-step-ahead' approach (see Section C-2) and is estimated on data for Africa only.

Change history

The model has been in use since r.2020.02.01

Feature importances

The tables report the feature importance for each type of violence and different time steps ($s = 1, 3, 6, 12, 24, 36$). Please note that for broad models such as *Onset*, the tables include only the first 25 most important features. Spatial proximity to conflict episodes is the most important feature to predict any type of violence. The occurrence of **ns** violence in neighboring cells is especially important to forecast not only **ns** but also **sb** and **os** conflicts, suggesting that violent episodes which do not involve the state may be important to capture early signs of any type of violence. This may also indicate that **ns** violence can transition into other forms of more structured conflict such as **sb** violence. This suggests that the spatial setting is particularly relevant to predict violence at the *pgm* level, while the time dimension is more relevant to predict conflict at the *cm* level. Likewise, geographical context and climate-related factors are among the most important features to predict the outbreak of any violence types, including the type of land, the presence of natural resources and the value of the SPEI indicator of drought. This indicates that environmental-related factors may be relevant in predicting early signs of violence.

sb_onset24_1_all	1	3	6	12	24	36
stdist_k1_t001_ged_dummy_ns	0.018598	0.018721	0.019086	0.019308	0.019544	0.019709
stdist_k1_t10_ged_dummy_ns	0.017685	0.017761	0.018193	0.018006	0.018319	0.018306
stdist_k1_t1_ged_dummy_ns	0.017220	0.017960	0.018163	0.018048	0.017681	0.018208
stdist_k1_t001_ged_dummy_os	0.016927	0.017356	0.017359	0.017737	0.017745	0.018085
stdist_k1_t001_ged_dummy_sb	0.016277	0.016268	0.016250	0.016787	0.016938	0.017166
ln_pgd_bdist3	0.015931	0.015588	0.014851	0.014816	0.014574	0.014939
pgd_pasture_ih	0.015563	0.015229	0.014878	0.014901	0.014745	0.015204
ln_pgd_capdist	0.015244	0.014526	0.014458	0.014412	0.014256	0.014485
stdist_k1_t10_ged_dummy_os	0.015195	0.016050	0.016572	0.016935	0.017351	0.017441
ln_pgd_pop_gpw_sum	0.015126	0.014356	0.014219	0.014106	0.014370	0.014472
pgd_agri_ih	0.014963	0.014717	0.014428	0.014192	0.014418	0.014798
ln_pgd_ttime_mean	0.014832	0.014141	0.013805	0.013750	0.013566	0.013806
time_since_ged_dummy_sb	0.014728	0.014370	0.014460	0.014247	0.014424	0.014048
stdist_k1_t1_ged_dummy_os	0.014727	0.015213	0.015657	0.016019	0.016382	0.016779
spdist_pgd_diamsec	0.013466	0.013040	0.012774	0.012652	0.012542	0.012943
time_since_ged_dummy_os	0.013093	0.013086	0.013308	0.013211	0.013539	0.013389
stdist_k1_t10_ged_dummy_sb	0.012884	0.013566	0.014547	0.014563	0.015680	0.015639
spei_1	0.012626	0.013553	0.013445	0.013604	0.013327	0.013359
time_since_ged_dummy_ns	0.012469	0.012495	0.012934	0.012734	0.013034	0.013307
spdist_pgd_petroleum	0.012393	0.011829	0.011702	0.011893	0.011783	0.012000
spei_3	0.012196	0.012435	0.012801	0.012250	0.012301	0.012170
spei_2	0.012189	0.012833	0.013094	0.012605	0.012520	0.012373
stdist_k1_t1_ged_dummy_sb	0.012178	0.012860	0.013540	0.014164	0.014570	0.014839
spei_4	0.011924	0.011866	0.012170	0.012016	0.012287	0.011699
pgd_urban_ih	0.011408	0.010638	0.010445	0.010208	0.010243	0.010560

Table C-6. Feature importances for model onset_1, **sb**

sb_onset24_100_all	1	3	6	12	24	36
stdist_k1_t10_ged_dummy_ns	0.017877	0.017514	0.018129	0.020543	0.020008	0.017569
stdist_k1_t001_ged_dummy_ns	0.017102	0.018185	0.018018	0.019094	0.018173	0.018612
pgd_pasture_ih	0.016470	0.016402	0.016078	0.015805	0.016246	0.017017
stdist_k1_t1_ged_dummy_ns	0.015509	0.016682	0.018007	0.017848	0.017159	0.018003
ln_pgd_bdist3	0.015051	0.014315	0.013900	0.015072	0.014706	0.015503
ln_pgd_capdist	0.014640	0.014315	0.014673	0.014963	0.014792	0.015704
pgd_agri_ih	0.014068	0.014479	0.014350	0.014440	0.014700	0.015560
ln_pgd_pop_gpw_sum	0.014020	0.014417	0.014397	0.014337	0.014517	0.014947
stdist_k1_t001_ged_dummy_os	0.013946	0.014300	0.015245	0.015956	0.015676	0.017252
ln_pgd_ttime_mean	0.013921	0.013207	0.013155	0.013412	0.013235	0.013533
spei_1	0.013277	0.014222	0.013672	0.012360	0.013622	0.013530
spdist_pgd_diamsec	0.013044	0.012635	0.013285	0.013635	0.013135	0.013921
stdist_k1_t10_ged_dummy_os	0.012700	0.013377	0.013427	0.014420	0.015592	0.015064
spei_4	0.012019	0.011496	0.011588	0.011827	0.012595	0.011639
spei_3	0.011621	0.012744	0.011888	0.011134	0.012547	0.011636
stdist_k1_t1_ged_dummy_os	0.011550	0.012980	0.012543	0.014200	0.015581	0.014892
spei_2	0.011353	0.013514	0.012542	0.011936	0.011997	0.011857
spei_6	0.011330	0.010791	0.011460	0.010913	0.012008	0.011833
spei_5	0.011206	0.011146	0.011710	0.011362	0.012406	0.010976
time_since_ged_dummy_ns	0.011196	0.012099	0.011560	0.011931	0.011365	0.010927
pgd_urban_ih	0.011127	0.010970	0.010996	0.011480	0.011761	0.012236
pgd_imr_mean	0.011014	0.010794	0.010933	0.010974	0.010799	0.010717
spei_10	0.010978	0.010597	0.010817	0.011544	0.011025	0.011504
spdist_pgd_petroleum	0.010891	0.011792	0.011100	0.011738	0.011543	0.012637
spei_13	0.010791	0.010477	0.010717	0.010492	0.011404	0.011314

Table C-7. Feature importances for model onset_100, **sb**

os_onset24_1_all	1	3	6	12	24	36
stdist_k1_t10_ged_dummy_ns	0.017171	0.016318	0.017487	0.017205	0.017446	0.017160
stdist_k1_t001_ged_dummy_ns	0.016921	0.017348	0.017808	0.017943	0.018247	0.018509
stdist_k1_t001_ged_dummy_sb	0.015953	0.016276	0.016427	0.016960	0.016857	0.017079
stdist_k1_t1_ged_dummy_ns	0.015642	0.015975	0.016532	0.017138	0.017108	0.016980
stdist_k1_t001_ged_dummy_os	0.015546	0.015717	0.016257	0.016751	0.016678	0.017493
ln_pgd_pop_gpw_sum	0.015359	0.014745	0.014216	0.014264	0.014335	0.014740
ln_pgd_capdist	0.015118	0.014728	0.014234	0.014150	0.013942	0.014311
ln_pgd_bdist3	0.015080	0.014354	0.013812	0.013669	0.013650	0.013875
pgd_agri_ih	0.015041	0.014367	0.014001	0.014170	0.014514	0.014716
pgd_pasture_ih	0.014898	0.014120	0.013953	0.014173	0.014499	0.014584
time_since_ged_dummy_os	0.014462	0.013873	0.013720	0.013384	0.013084	0.012796
stdist_k1_t10_ged_dummy_sb	0.014218	0.014572	0.014590	0.014937	0.015457	0.015300
ln_pgd_ttime_mean	0.014214	0.013603	0.013253	0.013277	0.013054	0.013323
stdist_k1_t1_ged_dummy_sb	0.013287	0.013524	0.014255	0.014566	0.014513	0.014874
spei_1	0.012986	0.013484	0.013943	0.013629	0.012821	0.013965
spdist_pgd_petroleum	0.012852	0.012306	0.012107	0.012077	0.011959	0.012242
stdist_k1_t10_ged_dummy_os	0.012631	0.013789	0.014424	0.014874	0.014921	0.015968
time_since_ged_dummy_sb	0.012562	0.012826	0.013079	0.012975	0.013047	0.013218
time_since_ged_dummy_ns	0.012295	0.012664	0.012622	0.012789	0.012656	0.012598
spdist_pgd_diamsec	0.012178	0.011893	0.011644	0.011554	0.011691	0.011959
stdist_k1_t1_ged_dummy_os	0.012072	0.012978	0.013369	0.014310	0.014310	0.015266
spei_2	0.011998	0.013084	0.012996	0.012907	0.012758	0.013075
spei_3	0.011972	0.012879	0.012238	0.012327	0.012225	0.012289
pgd_urban_ih	0.011914	0.011288	0.010594	0.010727	0.010679	0.010862
pgd_savanna_ih	0.011608	0.010884	0.010816	0.010794	0.010675	0.010918

Table C-8. Feature importances for model onset_1, **os**

os_onset24_100_all	1	3	6	12	24	36
stdist_k1_t10_ged_dummy_ns	0.017445	0.016699	0.019595	0.016829	0.020981	0.019646
ln_pgd_capdist	0.016681	0.016653	0.017028	0.016816	0.017821	0.017378
stdist_k1_t001_ged_dummy_ns	0.016211	0.017199	0.017178	0.018765	0.018691	0.019507
ln_pgd_bdist3	0.016001	0.015803	0.015389	0.015019	0.015042	0.014536
stdist_k1_t1_ged_dummy_ns	0.015875	0.016125	0.017777	0.016499	0.017971	0.018110
pgd_pasture_ih	0.015186	0.015188	0.015461	0.014780	0.015536	0.015304
stdist_k1_t001_ged_dummy_sb	0.014311	0.015115	0.016156	0.016099	0.016511	0.016749
spdist_pgd_petroleum	0.014217	0.013745	0.013762	0.013586	0.014029	0.014032
pgd_agri_ih	0.014016	0.014361	0.014899	0.014500	0.015111	0.015141
ln_pgd_pop_gpw_sum	0.013901	0.014202	0.014266	0.013925	0.014927	0.015349
pgd_forest_ih	0.013655	0.012587	0.012459	0.012613	0.013375	0.012820
stdist_k1_t10_ged_dummy_sb	0.013006	0.013660	0.014266	0.014295	0.015172	0.016371
ln_pgd_ttime_mean	0.012962	0.013216	0.013235	0.013528	0.013421	0.013696
spei_1	0.012480	0.012381	0.014132	0.013508	0.013455	0.013442
spei_3	0.012243	0.011852	0.013150	0.011688	0.011228	0.012210
stdist_k1_t1_ged_dummy_sb	0.011943	0.012117	0.013121	0.013668	0.013829	0.014960
spdist_pgd_diamsec	0.011938	0.011380	0.010855	0.011311	0.011561	0.012129
spei_5	0.011892	0.010904	0.011936	0.011066	0.011267	0.011635
spei_7	0.011666	0.010222	0.011201	0.011722	0.011117	0.010862
stdist_k1_t001_ged_dummy_os	0.011517	0.012830	0.014735	0.016146	0.016341	0.017237
time_since_ged_dummy_sb	0.011388	0.011912	0.012362	0.012453	0.013403	0.013674
spei_4	0.011371	0.010809	0.011440	0.011132	0.012242	0.011442
spei_10	0.011309	0.010728	0.010777	0.010492	0.010347	0.011264
spei_6	0.011200	0.010705	0.011270	0.010794	0.011379	0.011102
spei_8	0.011190	0.011080	0.010974	0.011227	0.010830	0.010705

Table C-9. Feature importances for model onset_100, **os**

ns_onset24_1_all	1	3	6	12	24	36
stdist_k1_t001_ged_dummy_os	0.017490	0.017397	0.017192	0.018048	0.018350	0.018232
stdist_k1_t001_ged_dummy_ns	0.017099	0.017095	0.017120	0.017401	0.018091	0.018627
stdist_k1_t001_ged_dummy_sb	0.016631	0.016601	0.017072	0.017141	0.017900	0.017852
ln_pgd_pop_gpw_sum	0.015234	0.014336	0.014073	0.013969	0.014579	0.014486
time_since_ged_dummy_ns	0.015202	0.014847	0.014601	0.014739	0.014185	0.014129
stdist_k1_t10_ged_dummy_sb	0.014849	0.015039	0.015171	0.015582	0.015653	0.016315
pgd_pasture_ih	0.014780	0.013997	0.013844	0.014053	0.014548	0.014745
ln_pgd_bdist3	0.014778	0.013969	0.013706	0.013809	0.013899	0.014019
pgd_agri_ih	0.014634	0.014054	0.013576	0.013731	0.014412	0.014768
stdist_k1_t1_ged_dummy_os	0.014279	0.014857	0.014687	0.015218	0.015783	0.016090
stdist_k1_t1_ged_dummy_sb	0.014256	0.014400	0.014673	0.014963	0.015672	0.015760
spei_1	0.014057	0.013138	0.013190	0.012961	0.013192	0.013146
stdist_k1_t10_ged_dummy_os	0.014048	0.014750	0.014969	0.015433	0.016152	0.016007
stdist_k1_t10_ged_dummy_ns	0.014019	0.014733	0.015603	0.015788	0.016845	0.017245
ln_pgd_capdist	0.013941	0.013157	0.013329	0.013262	0.013719	0.013665
ln_pgd_ttime_mean	0.013799	0.013291	0.012685	0.013087	0.013250	0.013380
stdist_k1_t1_ged_dummy_ns	0.013703	0.014095	0.014700	0.015260	0.016019	0.016566
time_since_ged_dummy_os	0.012725	0.012911	0.013239	0.013322	0.013095	0.013058
spei_2	0.012713	0.012330	0.012039	0.012795	0.013007	0.012053
spdist_pgd_diamsec	0.012638	0.012340	0.012047	0.012215	0.012562	0.012550
time_since_ged_dummy_sb	0.012506	0.013015	0.012957	0.012949	0.013102	0.013068
spdist_pgd_petroleum	0.012287	0.011679	0.011411	0.011676	0.011873	0.011806
spei_3	0.011783	0.011797	0.011842	0.012655	0.012221	0.011270
spei_4	0.011476	0.011445	0.011618	0.012523	0.011732	0.011535
spei_5	0.011156	0.011430	0.011329	0.012066	0.011529	0.010996

Table C-10. Feature importances for the onset_1 model, **ns**

ns_onset24_100_all	1	3	6	12	24	36
stdist_k1_t001_ged_dummy_os	0.017630	0.016404	0.017484	0.016799	0.019009	0.019260
pgd_pasture_ih	0.017079	0.015784	0.015300	0.015925	0.017052	0.018078
ln_pgd_capdist	0.015788	0.013898	0.014944	0.013748	0.016458	0.016175
stdist_k1_t10_ged_dummy_sb	0.015688	0.016772	0.013790	0.015368	0.015932	0.016008
stdist_k1_t001_ged_dummy_sb	0.015664	0.015934	0.016015	0.016979	0.017503	0.017469
ln_pgd_bdist3	0.015645	0.014754	0.014990	0.014658	0.016136	0.016316
stdist_k1_t001_ged_dummy_ns	0.015445	0.016674	0.015411	0.016362	0.017385	0.019255
stdist_k1_t1_ged_dummy_os	0.015222	0.014107	0.014275	0.014156	0.015592	0.015142
stdist_k1_t10_ged_dummy_os	0.014695	0.013016	0.016443	0.013675	0.015059	0.014587
stdist_k1_t10_ged_dummy_ns	0.014687	0.014640	0.014056	0.014184	0.017644	0.020033
ln_pgd_pop_gpw_sum	0.014001	0.013293	0.013143	0.013729	0.015337	0.014278
stdist_k1_t1_ged_dummy_ns	0.013974	0.013230	0.014105	0.013697	0.015852	0.016030
spdist_pgd_diamsec	0.013754	0.013375	0.012657	0.013276	0.014618	0.014444
pgd_agri_ih	0.013275	0.012720	0.013183	0.013798	0.015252	0.016437
ln_pgd_ttime_mean	0.013200	0.012339	0.011968	0.012370	0.012922	0.013395
stdist_k1_t1_ged_dummy_sb	0.013161	0.013813	0.014298	0.012923	0.015712	0.015142
spei_4	0.012803	0.010187	0.011763	0.011516	0.011284	0.009066
time_since_ged_dummy_ns	0.012457	0.012048	0.011572	0.011520	0.010420	0.011226
spei_5	0.012270	0.010680	0.010868	0.011328	0.011011	0.009865
time_since_ged_dummy_sb	0.011655	0.012158	0.011257	0.010658	0.011124	0.011639
spei_8	0.011592	0.011490	0.010685	0.010605	0.010843	0.010661
spei_1	0.011473	0.013267	0.011328	0.012519	0.015526	0.012030
spei_3	0.011449	0.011675	0.012484	0.010901	0.011644	0.010076
spei_9	0.011443	0.011023	0.011110	0.010896	0.010459	0.010363
spei_7	0.011418	0.012154	0.010821	0.011796	0.010552	0.010657

Table C-11. Feature importances for the onset_100 model, **ns**

C-3.4 Natural geography (pgd_natural)

Description

The grid level natural geography model includes the spatial distance to diamonds and petroleum deposits. In fact, research has shown that the location of natural resources such as minerals and hydrocarbons play an important role in affecting the probability and duration of conflict, due to the opportunities and incentives it creates for rebel groups (Lujala, 2005). The model also contains indicators referring to the main land type: cultivated areas, barren, forest, mountains, svanna, shrub, pasture and urban areas. The type of land may affect the risk of conflict in two schematic ways; first, mountainous terrain offers an ideal setting for rebels to hide and fight (Fearon, 2003); second, arable, fertile lands may encourage deprived people to fight for seizing the resource (Hidalgo et al., 2010).

Estimation details

The model uses the 'one-step-ahead' approach (see Section C-2) and is estimated on data for Africa only.

Feature importances

The tables below report the feature importances for each type of violence and different time steps ($s = 1, 3, 6, 12, 24, 36$). The type of land and land use are the most important features to predict any type of violence at the *pgm* level. The availability of profitable natural resources such as diamond and petroleum are also relatively important in predicting conflict, although less than agricultural-related variables.

sb_pgd_natural	1	3	6	12	24	36
pgd_pasture_ih	0.186816	0.185835	0.184290	0.185486	0.182741	0.180676
pgd_agri_ih	0.167282	0.168836	0.168543	0.170958	0.174737	0.177455
pgd_savanna_ih	0.131608	0.130035	0.130735	0.128957	0.128530	0.126468
pgd_forest_ih	0.119029	0.117281	0.118641	0.115980	0.113359	0.111252
pgd_urban_ih	0.107489	0.108503	0.108658	0.110872	0.113624	0.116976
spdist_pgd_diamsec	0.077449	0.078355	0.078250	0.076679	0.076750	0.074341
spdist_pgd_petroleum	0.070585	0.071312	0.070267	0.069365	0.067015	0.066040
pgd_mountains_mean	0.060417	0.059164	0.058636	0.057457	0.055587	0.053299
pgd_shrub_ih	0.054167	0.055438	0.056488	0.058680	0.060591	0.065352
pgd_barren_ih	0.025160	0.025241	0.025494	0.025565	0.027065	0.028142

os_pgd_natural	1	3	6	12	24	36
pgd_pasture_ih	0.183822	0.179077	0.182034	0.181357	0.183577	0.187210
pgd_agri_ih	0.181070	0.181687	0.181193	0.183699	0.185448	0.186256
pgd_urban_ih	0.146354	0.146887	0.146047	0.145173	0.143767	0.142298
pgd_savanna_ih	0.127994	0.130804	0.128486	0.130554	0.131086	0.129931
pgd_forest_ih	0.105601	0.106859	0.106696	0.108127	0.108669	0.107498
spdist_pgd_petroleum	0.062052	0.061115	0.061823	0.061091	0.061040	0.063095
spdist_pgd_diamsec	0.053217	0.054172	0.055439	0.053238	0.054484	0.054258
pgd_barren_ih	0.051318	0.049931	0.049354	0.047853	0.046319	0.043922
pgd_mountains_mean	0.048376	0.048647	0.048635	0.047571	0.046205	0.045055
pgd_shrub_ih	0.040197	0.040820	0.040293	0.041338	0.039405	0.040479

ns_pgd_natural	1	3	6	12	24	36
pgd_pasture_ih	0.171874	0.170917	0.171462	0.175321	0.186234	0.189402
pgd_agri_ih	0.156985	0.158277	0.159376	0.159589	0.166675	0.167907
pgd_urban_ih	0.139961	0.142161	0.142807	0.142327	0.138348	0.136953
pgd_savanna_ih	0.117082	0.121270	0.120925	0.120535	0.121039	0.118961
pgd_forest_ih	0.114341	0.112620	0.113523	0.113052	0.106766	0.104710
pgd_mountains_mean	0.083464	0.081336	0.080100	0.077500	0.071491	0.067845
spdist_pgd_petroleum	0.060687	0.060967	0.059910	0.059129	0.057419	0.056478
spdist_pgd_diamsec	0.055932	0.055709	0.053864	0.054932	0.053980	0.053694
pgd_shrub_ih	0.050606	0.050776	0.052162	0.052317	0.054016	0.056466
pgd_barren_ih	0.049070	0.045967	0.045871	0.045298	0.044032	0.047584

Table C-12. Feature importances for model natural geography, **sb**, **os**, **ns**

Change history

The model has been in use since r.2020.02.01

C-3.5 Social geography (pgd_social)

Description

The grid level social geography model contains a set of human geography features that may affect conflict, such as the (logged) distance to the national border, the nearest urban center and the capital city (Collier and Hoeffler, 2002). As conflict is deeply influenced by the degree of development—to the point that it has been defined as *development in reverse* (Collier et al., 2003)—the model also includes proxies of development akin the logged grid-level population size, GDP, infant mortality rate and an indicator of the share of excluded ethnic groups in each cell.

Estimation details

The model uses the 'one-step-ahead' approach (see Section C-2) and is estimated on data for Africa only.

Feature importances

The tables below report the feature importances for each type of violence and different time steps ($s = 1, 3, 6, 12, 24, 36$). Population size and gross cell product are the most important features to predict any type of violence, and their influence remains constant over the time steps. Infant mortality rate is relatively more important to predict **ns** violence, while distance to the capital city is more relevant to the other types of conflict (**sb**, **os**).

sb_pgd_social	1	3	6	12	24	36
ln_pgd_pop_gpw_sum	0.359695	0.359046	0.360514	0.363542	0.370824	0.382364
pgd_gcp_mer	0.239964	0.239668	0.239571	0.240666	0.237039	0.230920
ln_pgd_capdist	0.101451	0.101760	0.101729	0.099645	0.096991	0.095909
ln_pgd_bdist3	0.097956	0.098165	0.097535	0.095225	0.093435	0.090468
ln_pgd_ttime_mean	0.095284	0.096109	0.094523	0.093705	0.090578	0.088147
pgd_imr_mean	0.091934	0.091676	0.091956	0.092296	0.094897	0.095399
greq_1_pgd_excluded	0.013716	0.013575	0.014172	0.014922	0.016235	0.016793
os_pgd_social	**1**	**3**	**6**	**12**	**24**	**36**
ln_pgd_pop_gpw_sum	0.437880	0.435807	0.431703	0.433201	0.429984	0.429416
pgd_gcp_mer	0.251570	0.253885	0.251193	0.253239	0.255931	0.254437
ln_pgd_capdist	0.082636	0.083257	0.085232	0.084385	0.084478	0.085307
pgd_imr_mean	0.073763	0.074185	0.077619	0.075317	0.076407	0.077824
ln_pgd_bdist3	0.064367	0.064574	0.064457	0.063951	0.063948	0.065249
ln_pgd_ttime_mean	0.062949	0.061638	0.062680	0.062702	0.062593	0.062397
greq_1_pgd_excluded	0.026834	0.026654	0.027115	0.027205	0.026659	0.025370
ns_pgd_social	**1**	**3**	**6**	**12**	**24**	**36**
ln_pgd_pop_gpw_sum	0.358971	0.361053	0.360526	0.374131	0.379946	0.397179
pgd_gcp_mer	0.286361	0.282628	0.283776	0.276408	0.283145	0.271006
pgd_imr_mean	0.076293	0.078422	0.076658	0.074843	0.072730	0.072527
ln_pgd_ttime_mean	0.074648	0.073601	0.076416	0.075244	0.072876	0.073298
ln_pgd_capdist	0.071577	0.070141	0.072063	0.068652	0.070395	0.068590
greq_1_pgd_excluded	0.066944	0.067168	0.066648	0.064186	0.055601	0.052179
ln_pgd_bdist3	0.065206	0.066988	0.063913	0.066536	0.065307	0.065221

Table C-13. Feature importances for the social geography model, **sb**, **os**, **ns**

Change history

The model has been in use since r.2020.02.01

C-3.6 SPEI (spei_full)

Description

The SPEI model refers to the Standardized Precipitation Evapotranspiration Index (SPEI) global drought monitor data. The fully specified model includes SPEI data to model all 48 time scales (s = 1 up to s = 48) (*spei_full*). Drought occurrence is likely to increase the incidence of conflict, especially in vulnerable communities which lack the resources to cope with the environmental shock (von Uexkull et al., 2016). The SPEI index is a commonly used measure of drought (Vincente-Serrano et al., 2010) and is computed from both precipitation and temperature data, which makes it more suitable to climatic drought monitoring than other indices such as the SPI. Since the index is based on a water balance, higher SPEI values correspond to greater availability of water, while lower values correspond to water scarcity. A subseted model containing SPEI time scales 1, 3, 12, 24, 36, and 48 together with the history model's features is also included (speisubset_leghist)

Estimation details

The model uses the 'one-step-ahead' approach (see Section C-2) and is estimated on data for Africa only.

Change history

The model has been in use since r.2020.02.01

Feature importances

The tables below report the feature importances for each type of violence and different time steps ($s = 1, 3, 6, 12, 24, 36$). The SPEI index is a dynamic index which can be computed at different time-scales, from 1 up to 48 months. SPEI index computed at short time scale (one and two months) are the most important features to predict violence of any type, followed by the SPEI value averaged across four years. Although this is not an indication of causality, the feature importances suggests that drought can have both short-term and long term.

sb_spei_full	1	3	6	12	24	36
spei_1	0.024547	0.025041	0.024803	0.024773	0.024616	0.024869
spei_2	0.023013	0.023240	0.023187	0.022973	0.022880	0.023335
spei_48	0.022656	0.022896	0.022957	0.022965	0.023417	0.023061
spei_3	0.022391	0.022301	0.022126	0.022136	0.022160	0.022372
spei_4	0.021911	0.021656	0.021472	0.021629	0.021946	0.021942
spei_5	0.021583	0.021445	0.021219	0.021278	0.021350	0.021455
spei_13	0.021565	0.021438	0.021339	0.021030	0.021283	0.020823
spei_47	0.021541	0.021801	0.021823	0.021730	0.022079	0.022056
spei_12	0.021517	0.021372	0.021478	0.021317	0.021235	0.020897
spei_11	0.021437	0.021333	0.021241	0.021268	0.021233	0.020993
spei_6	0.021377	0.021280	0.021176	0.021065	0.020949	0.021299
spei_14	0.021202	0.021103	0.020930	0.020892	0.021043	0.020791
spei_7	0.021147	0.020837	0.020755	0.020872	0.020873	0.021153
spei_10	0.021101	0.020969	0.021018	0.020963	0.021075	0.020826
spei_8	0.021068	0.020903	0.020610	0.020770	0.020954	0.021014
spei_15	0.021057	0.020921	0.020700	0.020638	0.020700	0.020670
spei_46	0.020838	0.020892	0.021142	0.020914	0.021341	0.021173
spei_9	0.020818	0.020930	0.020726	0.020787	0.020918	0.021058
spei_36	0.020750	0.020649	0.020810	0.020611	0.020496	0.020765
spei_16	0.020639	0.020628	0.020464	0.020544	0.020470	0.020382
spei_17	0.020588	0.020348	0.020361	0.020406	0.020268	0.020240
spei_24	0.020578	0.020742	0.020791	0.020985	0.020691	0.020480
spei_22	0.020552	0.020183	0.020595	0.020676	0.020584	0.020263
spei_23	0.020547	0.020531	0.020822	0.020750	0.020541	0.020502
spei_37	0.020537	0.020612	0.020851	0.020801	0.020545	0.020604

Table C-14. Feature importances for the Spei model, **sb**

os_spei_full	1	3	6	12	24	36
spei_1	0.025448	0.025100	0.024559	0.025051	0.024332	0.026255
spei_2	0.023646	0.023565	0.022834	0.023783	0.023077	0.023893
spei_48	0.022752	0.023161	0.023183	0.023154	0.022578	0.022515
spei_3	0.022516	0.022295	0.022009	0.022902	0.022754	0.023034
spei_4	0.022256	0.021675	0.021346	0.022105	0.022230	0.022371
spei_47	0.022005	0.021935	0.022010	0.022052	0.021678	0.021529
spei_5	0.021603	0.021335	0.021107	0.021485	0.022031	0.022092
spei_13	0.021557	0.021259	0.021400	0.020701	0.021352	0.020536
spei_12	0.021413	0.021241	0.021476	0.021057	0.021246	0.021016
spei_6	0.021316	0.020839	0.020996	0.020992	0.021393	0.021474
spei_11	0.021225	0.021156	0.021105	0.020907	0.021313	0.021115
spei_14	0.021204	0.021078	0.020968	0.020669	0.021344	0.020642
spei_46	0.021118	0.021083	0.021240	0.021167	0.020869	0.020720
spei_9	0.020963	0.020968	0.020833	0.020827	0.021181	0.021408
spei_10	0.020864	0.020881	0.021014	0.020985	0.021441	0.020935
spei_7	0.020852	0.020805	0.020707	0.020642	0.021360	0.021559
spei_15	0.020851	0.020766	0.020743	0.020762	0.020987	0.020536
spei_8	0.020844	0.020846	0.020918	0.020783	0.021086	0.021273
spei_45	0.020616	0.020604	0.020714	0.020767	0.020799	0.020579
spei_16	0.020590	0.020769	0.020542	0.020592	0.020848	0.020311
spei_24	0.020578	0.020546	0.020934	0.020739	0.020094	0.020799
spei_26	0.020496	0.020691	0.020765	0.020730	0.020078	0.020292
spei_17	0.020452	0.020353	0.020318	0.020706	0.020659	0.020560
spei_27	0.020449	0.020741	0.021065	0.020744	0.019807	0.020053
spei_44	0.020431	0.020343	0.020253	0.020277	0.020319	0.020182

Table C-15. Feature importances for the Spei model, **os**

ns_spei_full	1	3	6	12	24	36
spei_1	0.025039	0.024984	0.024143	0.023927	0.023880	0.024623
spei_2	0.023432	0.023477	0.023069	0.023142	0.022814	0.023150
spei_48	0.022931	0.022348	0.022762	0.023218	0.023332	0.021694
spei_3	0.022300	0.022481	0.022225	0.022424	0.022727	0.022376
spei_5	0.021875	0.021751	0.021130	0.021781	0.022008	0.021540
spei_7	0.021658	0.021254	0.020635	0.021540	0.021372	0.021397
spei_4	0.021574	0.021470	0.021568	0.021642	0.022146	0.021896
spei_8	0.021529	0.021247	0.020709	0.021321	0.021434	0.021560
spei_11	0.021444	0.020679	0.021244	0.021413	0.021163	0.021094
spei_6	0.021355	0.022010	0.020983	0.021539	0.021483	0.021206
spei_47	0.021340	0.021228	0.021338	0.021962	0.021985	0.020797
spei_10	0.021308	0.020654	0.020965	0.021645	0.021248	0.021251
spei_9	0.021226	0.020897	0.020945	0.021445	0.021274	0.021378
spei_12	0.021032	0.021344	0.021606	0.021511	0.021397	0.021150
spei_13	0.020985	0.021296	0.021702	0.021008	0.021350	0.020685
spei_14	0.020980	0.020991	0.021647	0.020851	0.020799	0.020806
spei_46	0.020913	0.020538	0.020329	0.021150	0.021175	0.020287
spei_23	0.020864	0.020549	0.021065	0.021237	0.020556	0.020945
spei_21	0.020808	0.020581	0.020724	0.021233	0.020735	0.021043
spei_16	0.020715	0.020873	0.021501	0.020882	0.020627	0.020617
spei_17	0.020698	0.021281	0.021224	0.020787	0.020881	0.020551
spei_18	0.020671	0.020888	0.020936	0.021019	0.020423	0.020403
spei_24	0.020641	0.020672	0.021170	0.020937	0.020405	0.020727
spei_19	0.020582	0.020958	0.020209	0.020786	0.020206	0.020568
spei_20	0.020575	0.020604	0.020318	0.020919	0.020938	0.020814

Table C-16. Feature importances for the Spei model, **ns**

C-3.7 Dynasim (ds_dummy, ds_25)

Description

The Dynasim models are specified as the canonical *pgm* models in Hegre et al. (2019) and they include the endogenously simulated violence types, with the addition of a new outcome of at least 25 BRDs in a grid cell. Both variants of simulated outcomes are included.

Estimation details

The model uses the dynamic simulation model illustrated in (see Hegre et al., 2019) and is estimated on data for Africa only. The single-death and 25 BRDs models are simulated together to inform each other.

Change history

The model has been in use since r.2020.02.01

C-3.8 All Themes (all_themes)

Description

The All Themes models include all the outcome-specific thematic models, i.e. the Natural Geography, Social Geography, Demographic, as well as the outcome-specific sets of conflict history features. We also include the Institutional and Oil Rent model, which have been removed from the ensemble due to their poor performance but can still contribute to the general RF setup. In fact, the advantage of broad models like this is due to their ability to capture interactions between different features and thus take into adequate account the complex relationships, feedback and conditional factors that may all together make conflict more likely.

Estimation details

The model uses the 'one-step-ahead' approach (see Section C-2) and is estimated on data for Africa only.

Change history

The model has been in use since r.2020.02.01

Feature importances

The tables below report the feature importances for each type of violence and different time steps ($s = 1, 3, 6, 12, 24, 36$). Please note that for broad models like *all_themes*, the tables report only the first 25 most important features. Similarly to the trend observed at the *cm* level, the most important features to predict violence at the *pgm* level are those related to each outcome-specific conflict history. Land type and use are especially relevant to predict **os** and **ns**. Population size is also among the most important features to predict violence of any type.

sb_allthemes	1	3	6	12	24	36
decay_12_time_since_ged_dummy_sb	0.099209	0.094488	0.099113	0.101902	0.101993	0.103407
tlag_1_ged_dummy_sb	0.045804	0.042179	0.041325	0.033777	0.021375	0.019716
tlag_2_ged_dummy_sb	0.040205	0.040595	0.034238	0.028400	0.023053	0.018418
decay_12_time_since_ged_dummy_os	0.036827	0.040734	0.044385	0.048703	0.053020	0.056226
tlag_3_ged_dummy_sb	0.034080	0.026570	0.032917	0.027193	0.024604	0.015232
tlag_4_ged_dummy_sb	0.030862	0.030819	0.026634	0.023300	0.020891	0.015122
tlag_6_ged_dummy_sb	0.027330	0.023795	0.025342	0.017765	0.014526	0.014466
tlag_7_ged_dummy_sb	0.025321	0.026431	0.022920	0.020138	0.013497	0.013724
ln_pgd_pop_gpw_sum	0.024481	0.024154	0.024079	0.024373	0.025550	0.027417
tlag_5_ged_dummy_sb	0.024219	0.027388	0.024543	0.020296	0.016707	0.015041
pgd_agri_ih	0.023875	0.023793	0.023632	0.024580	0.025200	0.026094
tlag_9_ged_dummy_sb	0.021505	0.021068	0.021115	0.019590	0.015197	0.012352
pgd_pasture_ih	0.021497	0.021110	0.020719	0.020664	0.021410	0.022828
tlag_8_ged_dummy_sb	0.021017	0.023561	0.019207	0.018002	0.014997	0.015787
fvp_ssp2_edu_sec_15_24_prop	0.020887	0.021132	0.023333	0.026306	0.033412	0.033882
ln_fvp_timeindep	0.020825	0.021266	0.022792	0.034416	0.037612	0.037417
pgd_gcp_mer	0.020526	0.020353	0.020092	0.020120	0.021065	0.022113
ln_fvp_timesincepreindepwar	0.019858	0.020096	0.021214	0.026958	0.032233	0.035233
ln_pgd_ttime_mean	0.019766	0.019367	0.018493	0.018180	0.018419	0.018807
ln_pgd_bdist3	0.019479	0.018860	0.018240	0.017985	0.018091	0.018218
ln_pgd_capdist	0.018940	0.018531	0.018200	0.018558	0.019730	0.021258
decay_12_time_since_ged_dummy_ns	0.018788	0.020886	0.023508	0.026639	0.030859	0.034208
tlag_10_ged_dummy_sb	0.018759	0.018260	0.018098	0.018384	0.015761	0.010946
fvp_ssp2_urban_share_iiasa	0.018592	0.018404	0.019612	0.022095	0.022554	0.024419
tlag_11_ged_dummy_sb	0.018205	0.017407	0.018189	0.014488	0.014873	0.010524

Table C-17. Feature importances for the All_themes model, **sb**

os_allthemes	1	3	6	12	24	36
decay_12_time_since_ged_dummy_ns	0.152481	0.154399	0.151617	0.153147	0.148890	0.140953
decay_12_time_since_ged_dummy_os	0.062788	0.068855	0.072782	0.076277	0.077504	0.079672
decay_12_time_since_ged_dummy_sb	0.052166	0.058146	0.062706	0.068009	0.074263	0.076117
decay_12_time_since_acled_dummy_pr	0.051206	0.055557	0.059492	0.060714	0.063938	0.068115
tlag_1_splag_1_1_ged_dummy_ns	0.043436	0.040343	0.038555	0.038965	0.036988	0.036046
tlag_1_ged_dummy_ns	0.040614	0.033263	0.030270	0.025602	0.020571	0.016812
ln_pgd_pop_gpw_sum	0.028577	0.027418	0.026561	0.026806	0.026588	0.027284
pgd_pasture_ih	0.025929	0.024793	0.024375	0.024872	0.024916	0.026469
pgd_agri_ih	0.025543	0.024306	0.023993	0.024325	0.024988	0.026333
pgd_gcp_mer	0.025513	0.024362	0.023496	0.023192	0.023593	0.022399
pgd_urban_ih	0.024430	0.023837	0.023034	0.023209	0.023089	0.023579
ln_pgd_ttime_mean	0.021234	0.019680	0.019121	0.018783	0.018656	0.018616
ln_pgd_capdist	0.021189	0.020126	0.019786	0.019697	0.019486	0.019535
ln_pgd_bdist3	0.020630	0.019954	0.019052	0.019133	0.018731	0.018866
ln_fvp_population200	0.020579	0.020351	0.020711	0.020821	0.022059	0.022540
ln_fvp_timesinceregimechange	0.020168	0.022718	0.023282	0.022298	0.021005	0.019978
ln_fvp_timesincepreindepwar	0.020143	0.020597	0.021190	0.019551	0.020924	0.021198
ln_fvp_timeindep	0.020046	0.020187	0.019514	0.019083	0.019722	0.019761
fvp_ssp2_urban_share_iiasa	0.019985	0.019025	0.019269	0.018713	0.019212	0.019446
fvp_ssp2_edu_sec_15_24_prop	0.019962	0.019214	0.019294	0.018458	0.018358	0.018306
fvp_lngdpcap_nonoilrent	0.019481	0.019018	0.018403	0.019134	0.022884	0.024924
tlag_1_splag_1_1_acled_dummy_pr	0.019043	0.018892	0.019819	0.020099	0.021198	0.020521
spdist_pgd_diamsec	0.018742	0.017795	0.017330	0.016642	0.016623	0.016910
pgd_savanna_ih	0.017465	0.016637	0.016467	0.016710	0.017060	0.017983
spdist_pgd_petroleum	0.017351	0.016241	0.016057	0.016033	0.015456	0.015735

Table C-18. Feature importances for the All_themes model, **os**

ns_allthemes	1	3	6	12	24	36
decay_12_time_since_ged_dummy_ns	0.091485	0.093564	0.096816	0.104425	0.107150	0.107883
decay_12_time_since_ged_dummy_os	0.050372	0.059655	0.064912	0.073138	0.075634	0.077771
decay_12_time_since_ged_dummy_sb	0.043207	0.049537	0.055706	0.063783	0.073248	0.074407
decay_12_time_since_acled_dummy_pr	0.038182	0.044118	0.049943	0.052971	0.056319	0.062063
ln_pgd_pop_gpw_sum	0.027875	0.026274	0.025245	0.025066	0.025025	0.025920
pgd_pasture_ih	0.026739	0.024736	0.023965	0.023680	0.023891	0.025019
pgd_agri_ih	0.026462	0.025107	0.024338	0.024196	0.024976	0.026226
tlag_1_ged_dummy_ns	0.024739	0.018435	0.016955	0.014378	0.009763	0.007853
pgd_gcp_mer	0.022803	0.021465	0.020526	0.019586	0.019332	0.019038
ln_pgd_bdist3	0.022536	0.021309	0.020254	0.019279	0.018871	0.019347
ln_pgd_ttime_mean	0.022480	0.020761	0.019703	0.018495	0.018383	0.018236
ln_pgd_capdist	0.022265	0.021221	0.020380	0.020093	0.019648	0.019578
pgd_urban_ih	0.021541	0.020649	0.020289	0.020408	0.020569	0.021555
ln_fvp_population200	0.020238	0.019905	0.020275	0.020611	0.021985	0.023309
tlag_2_ged_dummy_ns	0.020152	0.022000	0.015321	0.012933	0.010491	0.009440
ln_fvp_timesincepreindepwar	0.019930	0.020255	0.019819	0.020023	0.020561	0.021169
tlag_1_splag_1_1_ged_dummy_ns	0.019852	0.019125	0.017172	0.020365	0.024215	0.024326
spdist_pgd_diamsec	0.019753	0.018257	0.017259	0.016703	0.016337	0.016469
fvp_ssp2_edu_sec_15_24_prop	0.019563	0.018824	0.018627	0.017663	0.018076	0.018598
tlag_3_ged_dummy_ns	0.019465	0.016909	0.016043	0.012241	0.009486	0.007542
ln_fvp_timesinceregimechange	0.019221	0.020828	0.023264	0.021604	0.019785	0.019485
tlag_4_ged_dummy_ns	0.019151	0.014512	0.015464	0.014020	0.009075	0.008537
fvp_ssp2_urban_share_iiasa	0.018698	0.018319	0.018021	0.018028	0.019041	0.019687
ln_fvp_timeindep	0.018418	0.017673	0.017478	0.017875	0.018727	0.019663
spdist_pgd_petroleum	0.018204	0.017182	0.016345	0.015742	0.015262	0.015412

Table C-19. Feature importances for the All_themes model, **ns**

C-3.9 Cross Level (crosslevel)

Description

The risk of conflict in a given area is affected by both sub-national and national factors. The cross-level model builds on this notion and allows the two spatial levels of analysis to inform each other.

Estimation details

The model is estimated by using the product of the predicted probabilities at the *cm* and *pgm* levels. Specifically, the product of the *xgb_model* at the *pgm* level and the *all_glob* model at the *cm* level is used.

Change history

The model has been in use since r.2020.02.01

C-3.10 XGBoost All (xgb_all)

Description

The XGBoost model is a Gradient Boosting Machine (GBM) model. The model uses the same feature set as the *all_themes* model described above, with the notable addition of fatality estimates and event counts for all *spacetime* elements described above.

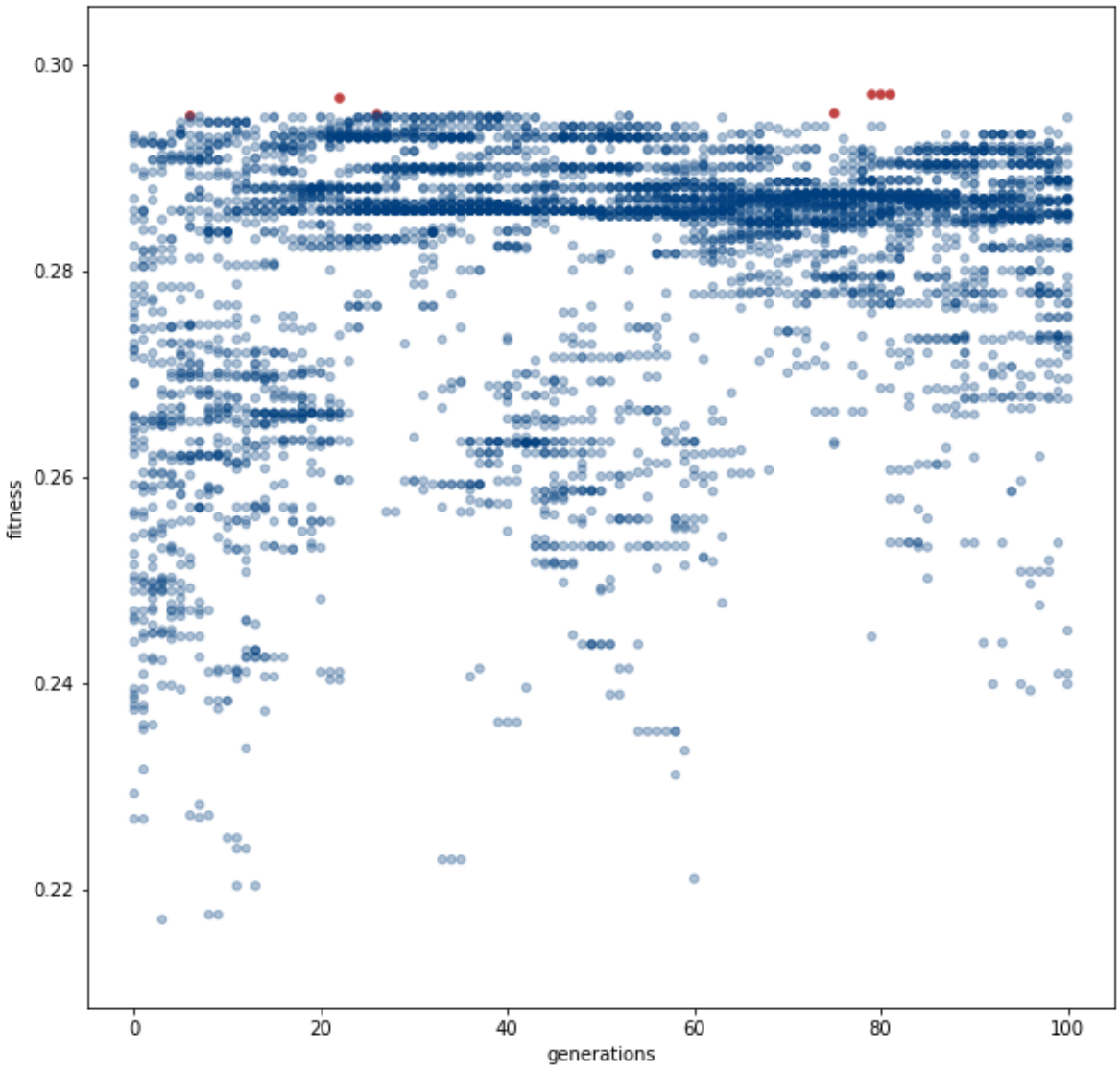
Estimation details

The model is estimated using the XGBoost algorithm (Chen and Guestrin, 2016), using both row and column sub-sampling as well as tree regularization. We performed hyper-parameter tuning on the following parameters: minimum loss reduction (or γ in XGB parlance), learning rate (or η in XGB parlance), maximum tree depth, minimum child weight and proportion of feature-space to sample (or column sample percentage per tree in XGB parlance). To prevent overfitting, common with boosted tree models, early stopping was implemented, with the model stopping the boosting procedure if the AUPR metric would not improve out-of-sample for 10 boosting iterations (see Appendix D for the reasoning behind AUPR as a metric). All out-of-sample evaluations were done on the subsequent partition of the data, in a similar manner with the rest of the ViEWS infrastructure. Hyper-parameters were tuned in the following manner:

1. A reasonable set of possible values were pre-selected for each hyper-parameter—resulting in a pool of 115,200 potential models.
2. The number of possible options, coupled with the size of the dataset, made it impossible to tune hyper-parameters exhaustively through grid-searching. Instead, for each constituent model, we optimized hyper-parameters through the use of a genetic algorithm (Russell and Norvig, 2016). This was applied in the following way: an initial sample of 50 possible hyper-parameter combinations were randomly sampled from the pool. 50 XGBoost models were trained with these hyper-parameters (using the 'train' partition), and scored based on the AUPR metric (out of sample, using the 'calibration' partition). 50 models were iterated across 100 generations. Figure C-2 shows the behavior of the 5,000 such estimated models across the 100 generations. Based on this score, models were allowed to probabilistically "mate", i.e., two high performing models will be merged taking some hyper-parameters (randomly) from the first the model and other hyper-parameters from the second model. 50 such mated "offspring" were produced. After this stage, some hyper-parameters were mutated from the mated "offspring", randomly being replaced with other hyper-parameters from the pool (with a probability of .1). Resulting models were then scored, and the mating and mutation procedure repeated for a total of 100 generations.
3. In order to avoid over-fitting of the hyper-parameter tuning routine on the calibration partition, we selected the best five performing hyper-parameter sets in any generation.

For each time-step in the infrastructure, an ensemble of five XGBoost models was trained, one with each set of hyper-parameters identified in the above procedure. A simple ensemble was constructed, taking the

Figure C-2. Performance (AUPR) of the genetic-optimized xgb models across the 100 generations of optimization. Each dot represents a model: the horizontal axis represents the generation where the model was produced, and the vertical axis the performance of that model. Note that the genetic algorithm was specified in such a way as to produce maximum variability as to increase the probability of identifying the best performing models.



simple average of predictions as the final prediction of the ensemble. As early stopping is required, the following solution was implemented:

- For models trained only on the 'train' partition, early stopping was done on 'calibration';
- For models trained on 'train' and 'calibrate' partitions, early stopping was done on the 'evaluate' partition;
- For models used for final forecasting, trained on all partitions, training ended at the last month with full UCDP data (currently December 2018); early stopping was done on data containing UCDP candidate events as targets (January 2019 - November 2019).

Note that hyper-parameter tuning is always done only once—with models trained on the 'train' partition and evaluated on the 'calibration' partition.

Figure C-3. Feature importances scores (in the gain domain) in a 6-month ahead state-based (**sb**) xgb model.

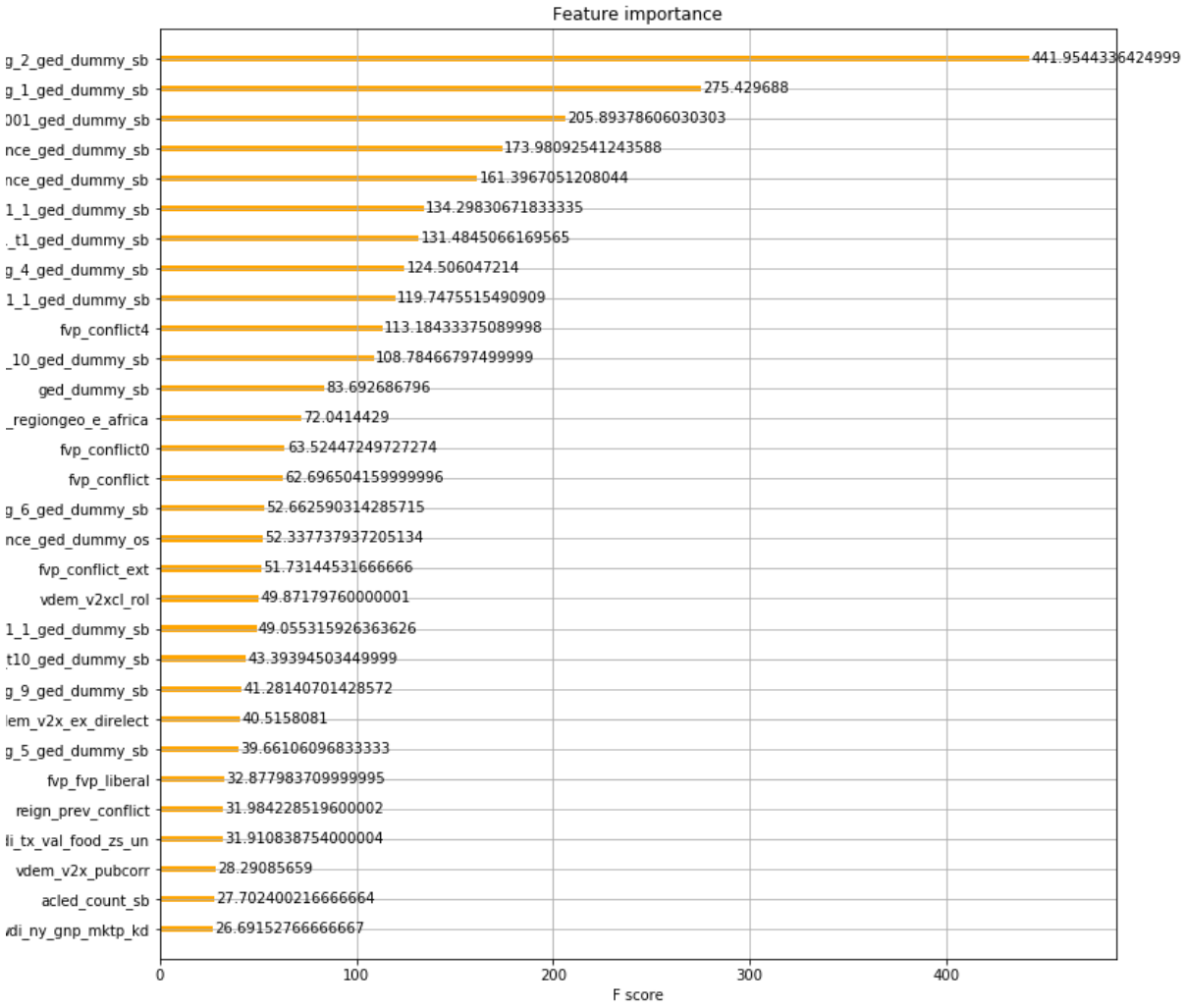


Figure C-3 shows feature importances scores for the most important 30 features from the xgb state-based (**sb**) 6-month ahead model in the gain domains (features that generate the highest total improvement in the F score).

Change history

The model has been in use since r.2020.02.01

C-4 CHANGES IN *PGM* MODELS

The *pgm* ensembles contain several broad models including all features mentioned below (a description of each *cm* model is reported in Appendix B). *Pgm* models are based on spatially disaggregated data, in line with recent empirical research which increasingly uses Geographical Information System (GIS) to investigate conflict at a sub-national level (Tollefsen, Strand, and Buhaug, 2012). Although *cm* models are especially useful for predictions of new conflicts, the higher spatial resolution of conflict predictors enable the *pgm* models to better capture differences across space that are leveled out by national average. Results of *pgm* models can not only contribute with new insight on the current empirical research on the causes of conflict,

but also provide a more thorough indication to policy-makers.

The new *pgm* models mainly include features which reflect recent empirical research findings, such as a drought model (*spei_full*). SPEI refers to the Standardized Precipitation Evapotranspiration Index, a commonly used measure of drought (Vicente-Serrano, Beguería, and López-Moreno, 2010). The SPEI index is computed from both precipitation and temperature data, which makes it more suitable to climatic drought monitoring than other indices. The model captures the non-linear relationship, featuring a higher risk of conflict at the two extremes of the scale, i.e. when SPEI is either higher or lower than average.

The *sptime* model focuses on the distance to conflict episodes both in space and time, by means of a single metric that measures the distance from a given grid-cell month to the most proximate conflict observation across both space and time. The model contains a number of these features representing various relative weightings of spatial and temporal distance. This model enables us to comprehensively account for the role of spatial and temporal context in affecting the probability of conflict (Buhaug and Rød, 2006; Raleigh and Hegre, 2009). Finally, three *pgm* models are trained using only onsets of conflict as the outcome to forecast.

C-5 DETAILED DESCRIPTIONS, CONSTITUENT MODELS

The following sections delve deeper into the three different types of violence (**sb**, **os**, **ns**) and present the results of predictions thereof.

C-5.1 State-based conflict (sb)

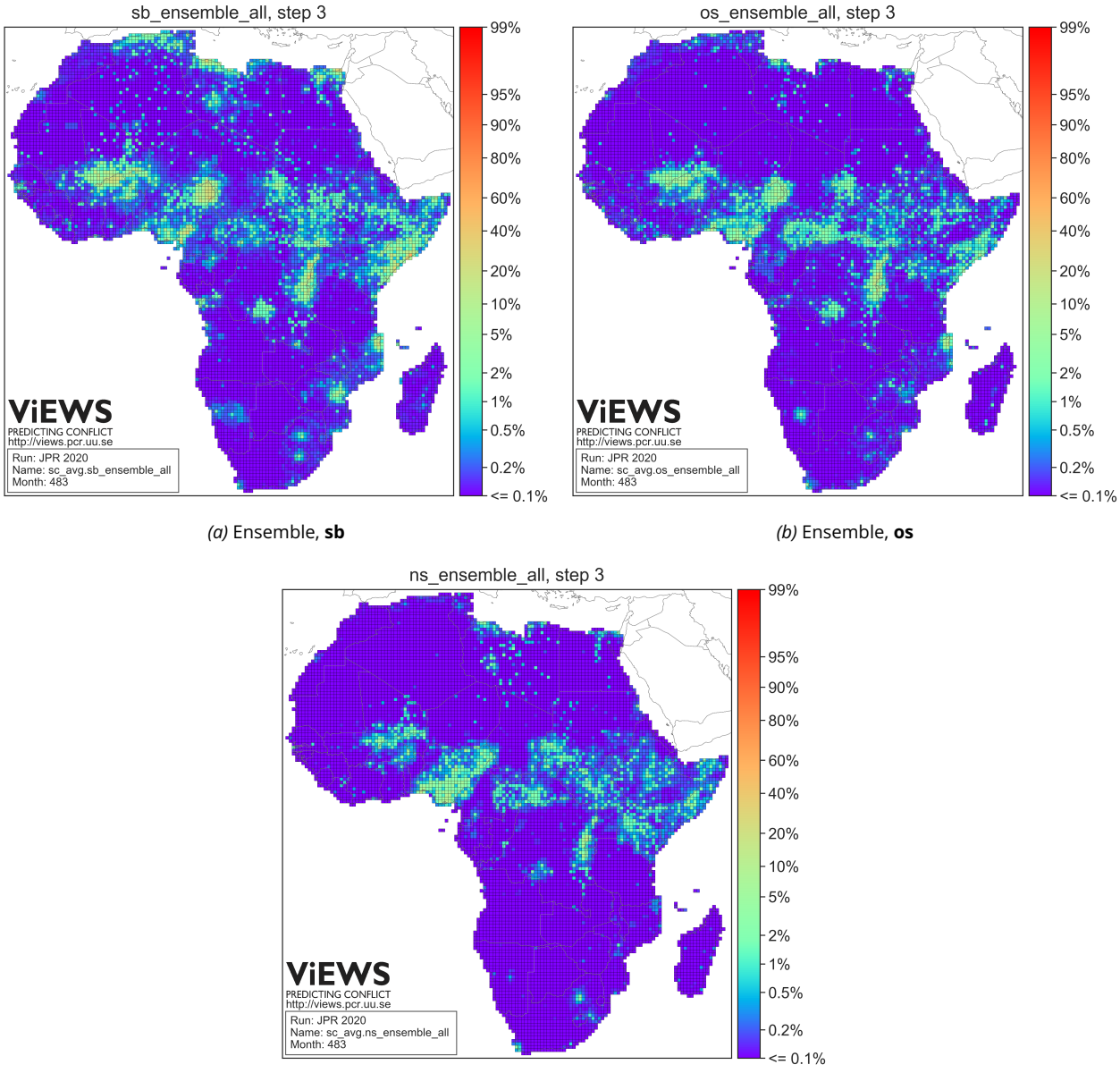
State-based violence is defined according to the standard UCPD definition, as a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year (Pettersson and Eck, 2018). State-based conflicts are thus characterized by the use of the force by the government as an active side of conflict, as opposed to non-state conflicts where none of the active sides are a government. Predictions of these types of violence may be of interest for all scholars of civil wars and those concerned with the main drivers of conflict against the state. By contrast, the following sections may be more relevant for researchers of other forms of violence which do not involve the government as an active part.

As follows, we present the prediction maps at the grid-cell level (*pgm*) for **sb**.

Prediction maps, forecasts (sb)

Figure C-5 shows the forecasts for the geographical *pgm* level for March 2020 based on the $s = 3$ model. All models are calibrated, i.e. rescaled using a logit procedure so that mean predicted probability in the calibration period is similar to the observed proportion of conflict. Models that separate poorly at this level has a bluish color, as the mean probability is much lower than for *cm*. Figure C-6 show the same for December 2022, i.e. $s = 36$ months into the future based on data up to and including December 2019.

Figure C-4. Predicted probabilities, ensemble models, **sb**, **os** and **ns**, $s = 3$ (month 483, March 2020), based on data up to December 2019

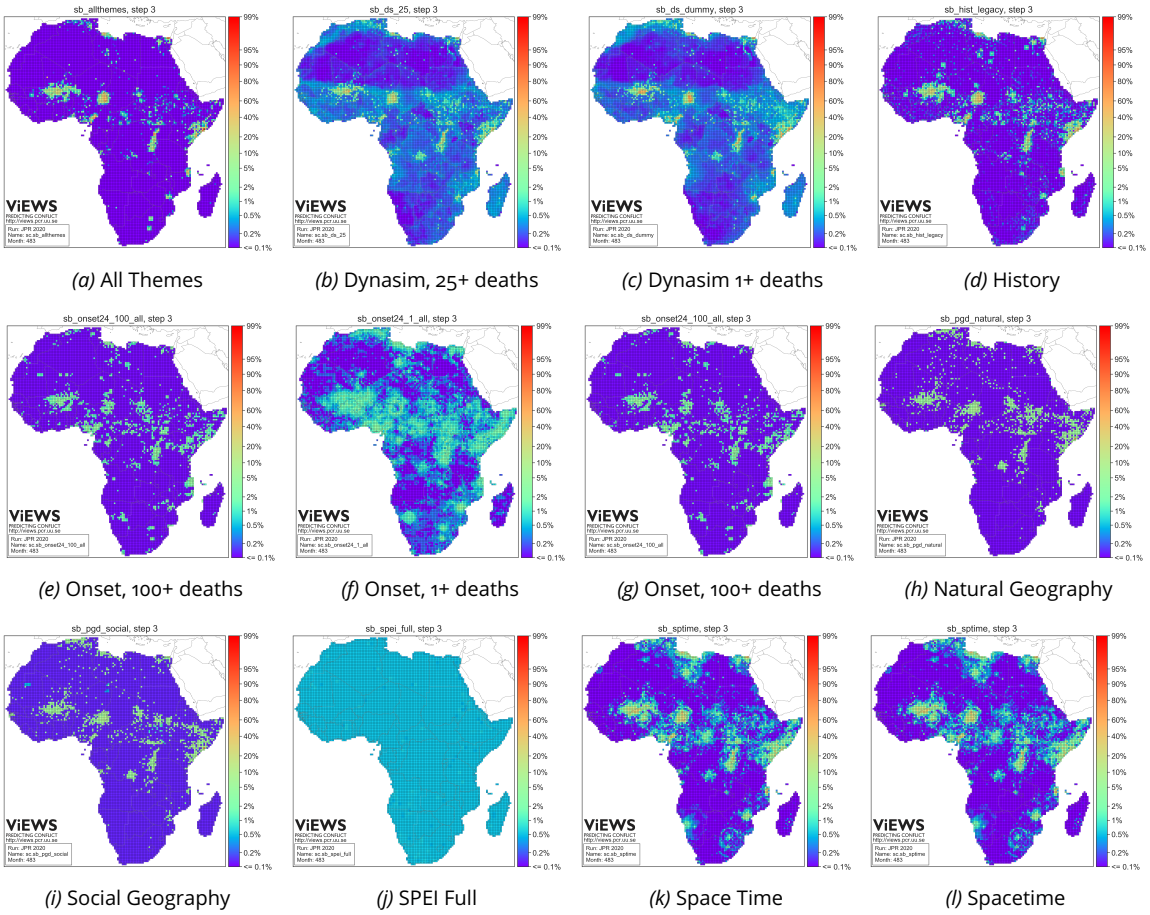


(a) Ensemble, **sb**

(b) Ensemble, **os**

(c) Ensemble, **ns**

Figure C-5. Predicted probabilities, constituent models, **sb**, $s = 3$ (month 483, March 2020), based on data up to December 2019

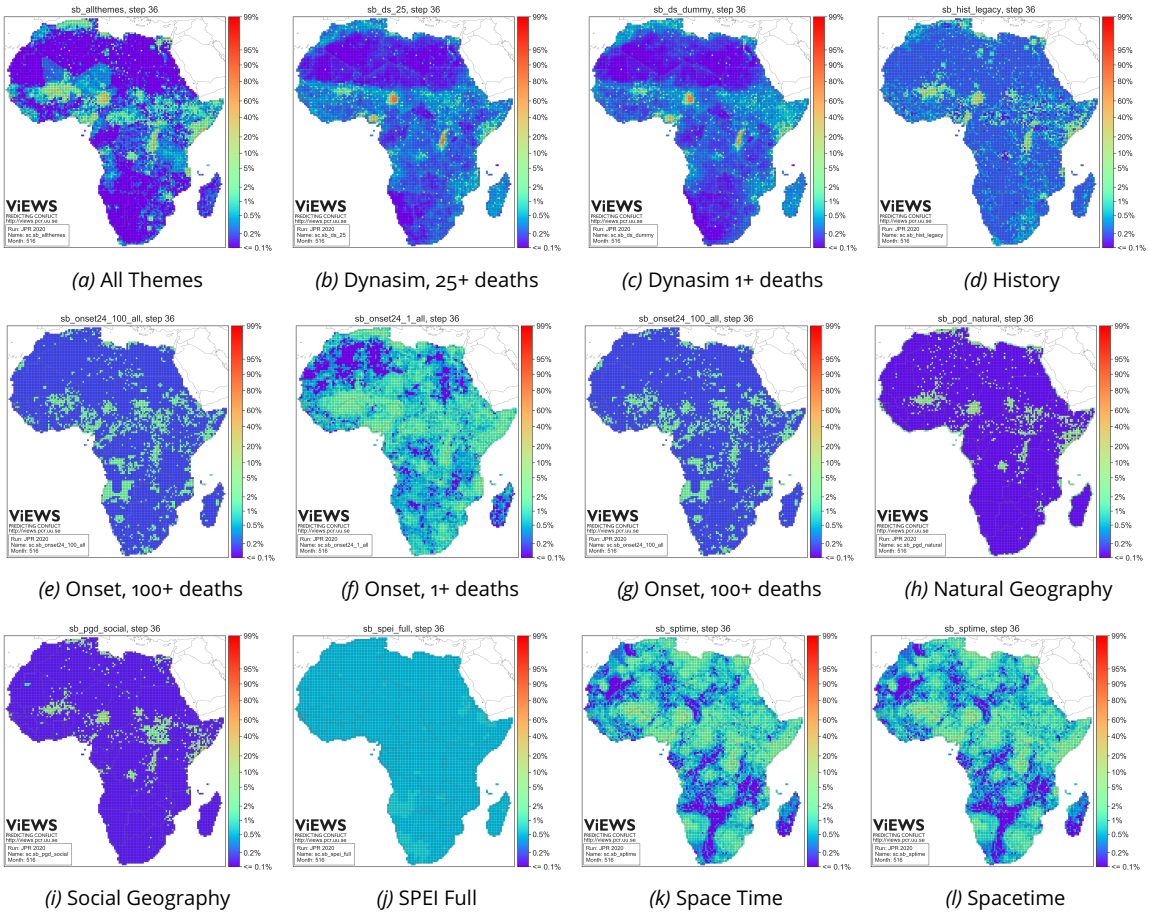


(a) All Themes *(b)* Dynasim, 25+ deaths *(c)* Dynasim 1+ deaths *(d)* History

(e) Onset, 100+ deaths *(f)* Onset, 1+ deaths *(g)* Onset, 100+ deaths *(h)* Natural Geography

(i) Social Geography *(j)* SPEI Full *(k)* Space Time *(l)* Spacetime

Figure C-6. Predicted probabilities, constituent models, **sb**, $s = 36$ (month 516, December 2022), based on data up to December 2019



(a) All Themes *(b)* Dynasim, 25+ deaths *(c)* Dynasim 1+ deaths *(d)* History

(e) Onset, 100+ deaths *(f)* Onset, 1+ deaths *(g)* Onset, 100+ deaths *(h)* Natural Geography

(i) Social Geography *(j)* SPEI Full *(k)* Space Time *(l)* Spacetime

C-5.2 One-sided violence (os)

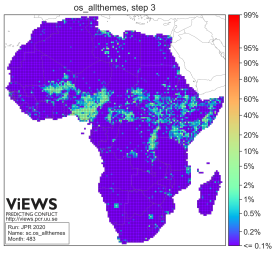
One-sided violence is defined as the deliberate use of armed force by the government of a state or by a formally organized group against civilians which result in at least 25 BRDs in a year (Pettersson and Eck, 2018). Unlike state-based conflict, *os* violence involves the use of the armed force by one actor against the civilian population. Predictions of **os** can be of particular interest to scholars of atrocities and violence against civilians.

As follows, we present the prediction maps at the grid-cell level (*pgm*) for **os**.

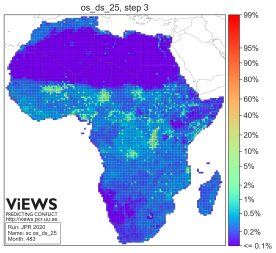
Prediction maps, forecasts (os)

Figure C-5 shows the forecasts for the geographical *pgm* level for March 2020 based on the $s = 3$ model. All models are calibrated, i.e. rescaled using a logit procedure so that mean predicted probability in the calibration period is similar to the observed proportion of conflict. Models that separate poorly at this level has a blueish color, as the mean probability is much lower than for *cm*. Figure C-8 show the same for December 2022, i.e. $s = 36$ months into the future based on data up to and including December 2019.

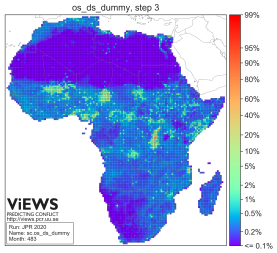
Figure C-7. Predicted probabilities, constituent models, **os**, $s = 3$ (month 483, March 2020), based on data up to December 2019



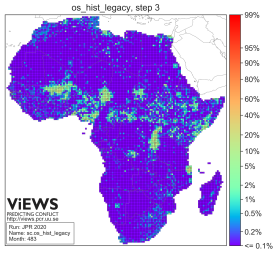
(a) All Themes



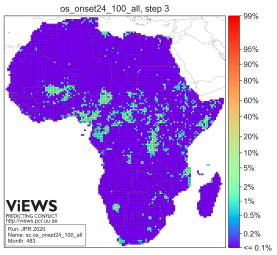
(b) Dynasim, 25+ deaths



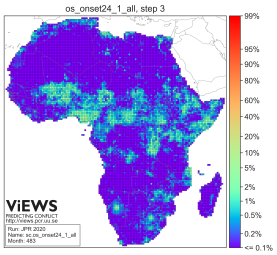
(c) Dynasim 1+ deaths



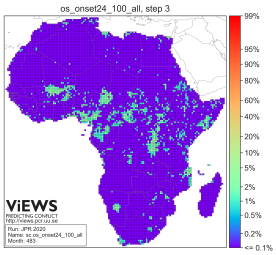
(d) History



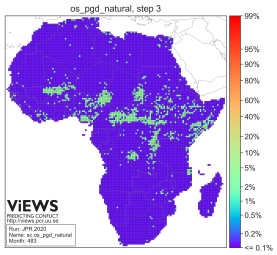
(e) Onset, 100+ deaths



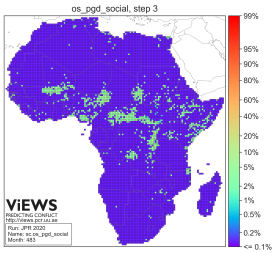
(f) Onset, 1+ deaths



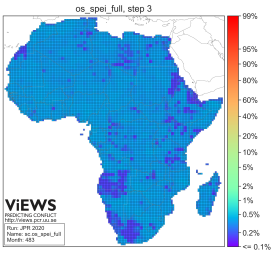
(g) Onset, 100+ deaths



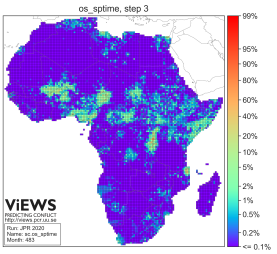
(h) Natural Geography



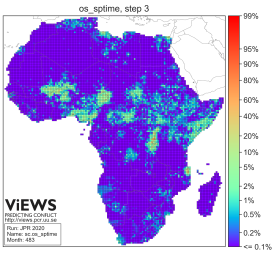
(i) Social Geography



(j) SPEI Full

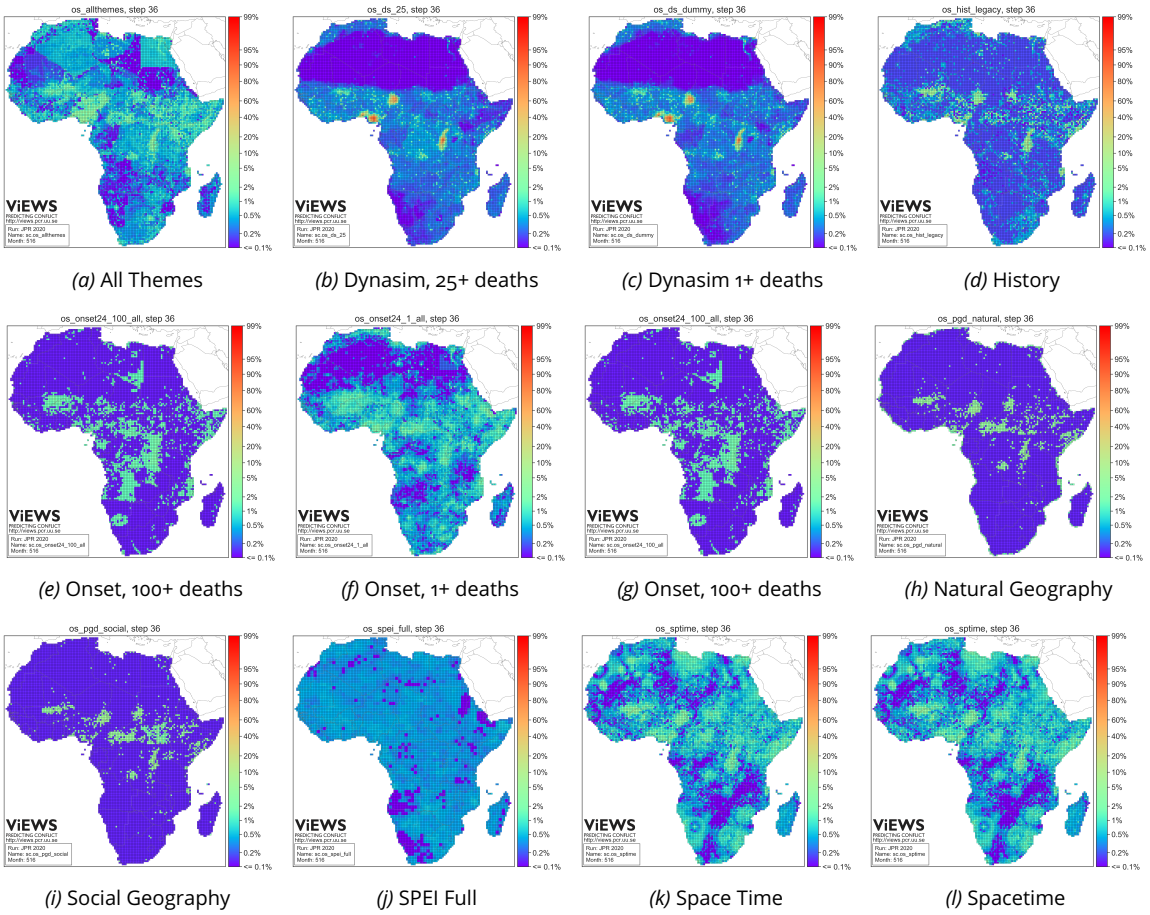


(k) Space Time



(l) Spacetime

Figure C-8. Predicted probabilities, constituent models, **os**, $s = 36$ (month 516, December 2022), based on data up to December 2019



(a) All Themes *(b)* Dynasim, 25+ deaths *(c)* Dynasim 1+ deaths *(d)* History

(e) Onset, 100+ deaths *(f)* Onset, 1+ deaths *(g)* Onset, 100+ deaths *(h)* Natural Geography

(i) Social Geography *(j)* SPEI Full *(k)* Space Time *(l)* Spacetime

C-5.3 Non-state conflict (ns)

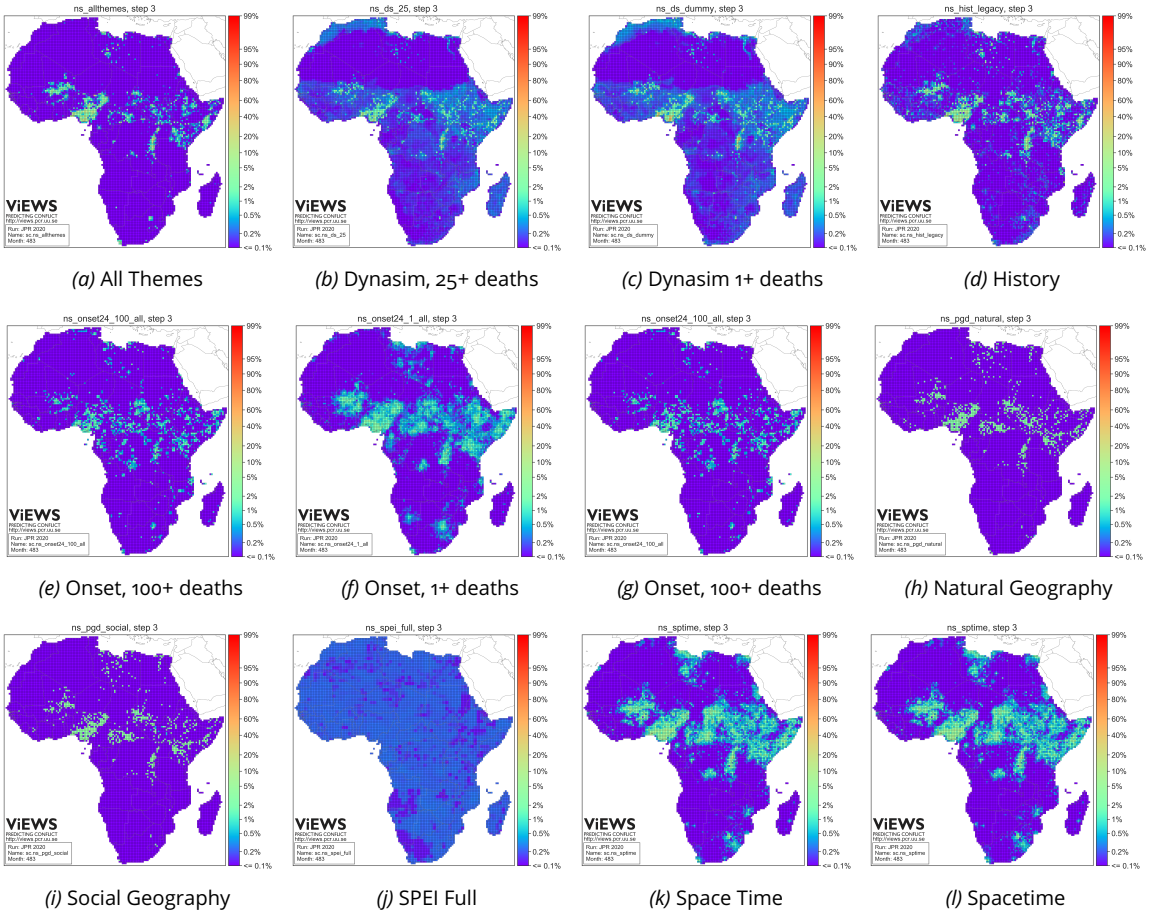
Non-state conflict is defined as the use of armed force by two formally organized groups, none of which is the government of a state, which result in at least 25 BRDs in a year (Pettersson and Eck, 2018). Conflicts between rebel or ethnic groups fall in this category.

As follows, we present the prediction maps for **ns**.

Prediction maps, forecasts (ns)

Figure C-9 shows the forecasts for the geographical *pgm* level for March 2020 based on the $s = 3$ model. All models are calibrated, i.e. re-scaled using a logit procedure so that mean predicted probability in the calibration period is similar to the observed proportion of conflict. Models that separate poorly at this level has a bluish color, as the mean probability is much lower than for *cm*. Figure C-10 show the same for December 2022, i.e. $s = 36$ months into the future based on data up to and including December 2019.

Figure C-9. Predicted probabilities, constituent models, **ns**, $s = 3$ (month 483, March 2020), based on data up to December 2019

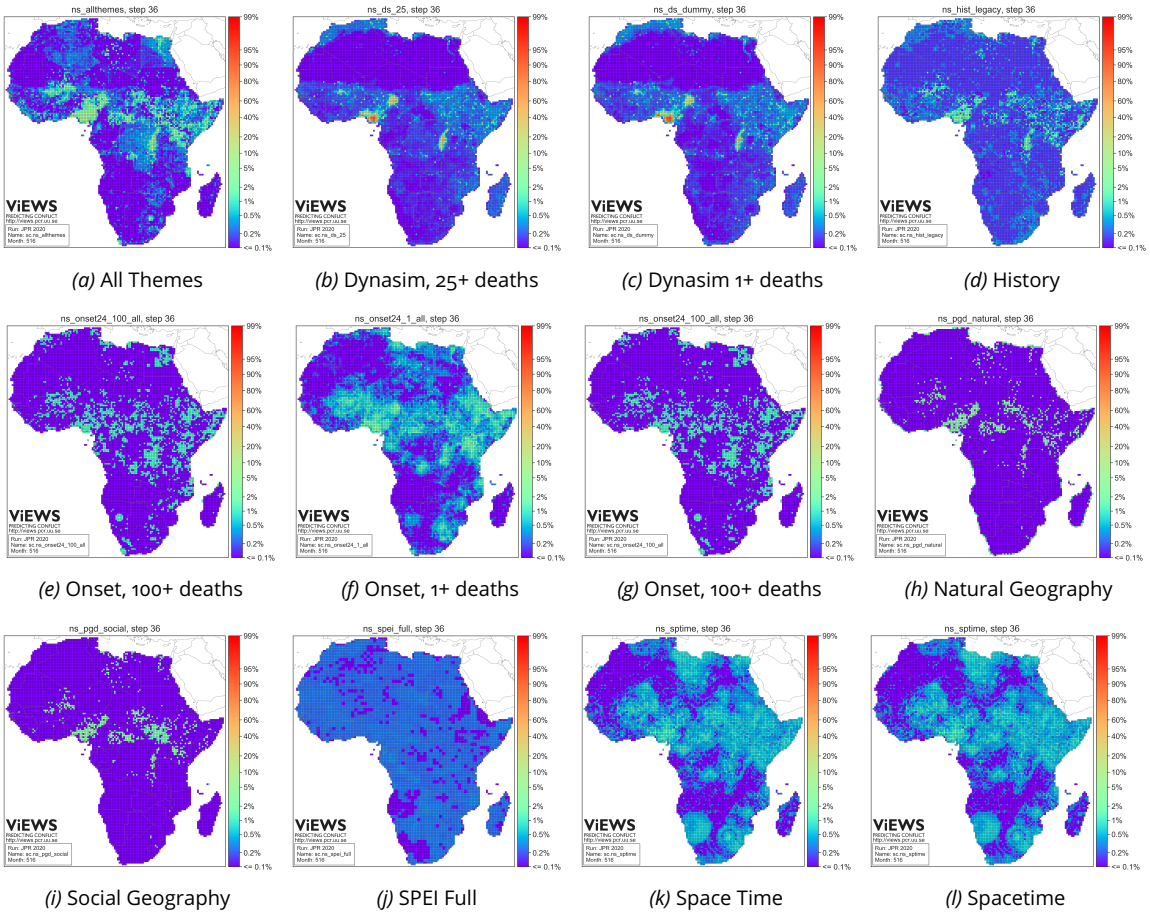


(a) All Themes *(b)* Dynasim, 25+ deaths *(c)* Dynasim 1+ deaths *(d)* History

(e) Onset, 100+ deaths *(f)* Onset, 1+ deaths *(g)* Onset, 100+ deaths *(h)* Natural Geography

(i) Social Geography *(j)* SPEI Full *(k)* Space Time *(l)* Spacetime

Figure C-10. Predicted probabilities, constituent models, **ns**, $s = 36$ (month 516, December 2022), based on data up to December 2019



(a) All Themes *(b)* Dynasim, 25+ deaths *(c)* Dynasim 1+ deaths *(d)* History

(e) Onset, 100+ deaths *(f)* Onset, 1+ deaths *(g)* Onset, 100+ deaths *(h)* Natural Geography

(i) Social Geography *(j)* SPEI Full *(k)* Space Time *(l)* Spacetime

REFERENCES

Breiman, Leo (2001). "Random Forests". In: *Machine learning* 45.1, pp. 5–32.

Buhaug, Halvard and Kristian Skrede Gleditsch (2008). "Contagion or Confusion? Why Conflicts Cluster in Space". In: *International Studies Quarterly* 52, pp. 215–233.

Buhaug, Halvard and Jan K. Rød (2006). "Local Determinants of African Civil Wars, 1970–2001". In: *Political Geography* 25.3, pp. 315–335.

Chen, Tianqi and Carlos Guestrin (2016). "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Collier, Paul, Lani Elliot, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis (2003). *Breaking the Conflict Trap. Civil War and Development Policy*. Oxford: Oxford University Press. URL: https://openknowledge.worldbank.org/handle/10986/13938.

Collier, Paul, Håvard Hegre, and Arvid Raknerud (2003). *ƒ War*.

Collier, Paul and Anke Hoeffler (2002). "On the Incidence of Civil War in Africa". In: *Journal of Conflict Resolution* 46.1, pp. 13–28.

Fearon, James D. (2003). "Presentation". In: *Workshop on Conflict Data*. Oslo.

Friedman, Jerome H. (2001). "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics*, pp. 1189–1232.

Géron, Aurélien (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.

Hegre, Håvard, Marie Allansson, Matthias Basedau, Mike Colaresi, Mihai Croicu, Hanne Fjelde, Frederick Hoyles, Lisa Hultman, Stina Högbladh, Remco Jansen, Naima Mouhleb, Sayeed Auwn Muhammad, Desirée Nilsson, Håvard Mokleiv Nygård, Gudlaug Olafsdottir, Kristina Petrova, David Randahl, Espen Geelmuyden Rød, Gerald Schneider, Nina von Uexkull, and Jonas Vestby (2019). "ViEWS: A political Violence Early Warning System". In: *Journal of Peace Research* 56.2, pp. 155–174. URL: https://doi.org/10.1177/0022343319823860.

Hegre, Håvard, Halvard Buhaug, Katherine V. Calvin, Jonas Nordkvelle, Stephanie T. Waldhoff, and Elisabeth Gilmore (2016). "Forecasting civil conflict along the shared socioeconomic pathways". In: *Environmental Research Letters* 11.5, p. 054002. DOI: 10.188/1748-9326/11/5/054002.

Hegre, Håvard, Joakim Karlsen, Håvard Mokleiv Nygård, Håvard Strand, and Henrik Urdal (2013). "Predicting Armed Conflict 2010–2050". In: *International Studies Quarterly* 57.2, pp. 250–270. DOI: 10.1111/isqu.12007.

Hidalgo, F. Daniel, Suresh Naidu, Simeon Nichter, and Neal Richardson (2010). "Economic Determinants of Land Invasions". In: *The Review of Economics and Statistics* 92.3, pp. 505–523. URL: https://EconPapers.repec.org/RePEc:tpr:restat:v:92:y:2010:i:3:p:505-523.

Lujala, PØivi (2005). "The Spoils of Nature – Oil, Gemstones, and Civil Conflict". In: *13th National Conference in Political Science*. HurdalsjØen.

Muchlinski, David, David Siroky, Jingrui He, and Matthew Kocher (2016). "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data". In: *Political Analysis* 24.1, pp. 87–103. DOI: 10.1093/pan/mpv024.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pettersson, Therése and Kristine Eck (2018). "Organized violence, 1989–2017". In: *Journal of Peace Research* 55.4, pp. 535–547. DOI: 10.1177/0022343318784101. URL: https://doi.org/10.1177/0022343318784101.

Raleigh, Clionadh and Håvard Hegre (2009). "Population, Size, and Civil War. A Geographically Disaggregated Analysis". In: *Political Geography* 28.4, pp. 224–238.

Russell, Stuart J and Peter Norvig (2016). *Artificial intelligence: a modern approach*. London: Pearson Education Limited,

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis (2008). "Conditional variable importance for random forests". In: *BMC Bioinformatics* 9.1, p. 307. DOI: 10.1186/1471-2105-9-307. URL: https://doi.org/10.1186/1471-2105-9-307.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn (2007). "Bias in random forest variable importance measures: Illustrations, sources and a solution". In: *BMC Bioinformatics* 8.1, p. 25. DOI: 10.1186/1471-2105-8-25. URL: https://doi.org/10.1186/1471-2105-8-25.

Tollefsen, Andreas Forø, Håvard Strand, and Halvard Buhaug (2012). "PRIO-GRID: A unified spatial data structure". In: *Journal of Peace Research* 49.2, pp. 363–374. DOI: 10.1177/0022343311431287.

Vicente-Serrano, Sergio M., Santiago Beguería, and Juan I. López-Moreno (2010). "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index". In: *Journal of Climate* 23.7, pp. 1696–1718. DOI: 10.1175/2009JCLI2909.1. URL: https://doi.org/10.1175/2009JCLI2909.1.

von Uexkull, Nina, Mihai Croicu, Hanne Fjelde, and Halvard Buhaug (2016). "Civil conflict sensitivity to growing-season drought". In: *Proceedings of the National Academy of Sciences* 113.44, pp. 12391–12396.

FUNDING

ViEWS is funded by the European Research Council, grant number 694640 and Uppsala University.

European Research Council
Established by the European Commission

COLLABORATIONS

ViEWS has an active interaction with other projects, including CLIMSEC, CAVE and CROP at PRIO (https://prio.org/), the MISTRA Geopolitics project, and most importantly the Uppsala Conflict Data Program (https://ucdp.uu.se/) at Uppsala University.

CODEBASE & PUBLICATIONS

ViEWS' codebase is available at:

https://github.com/UppsalaConflictDataProgram/OpenViEWS2

The full pool of publications are accessible at:

https://pcr.uu.se/research/views/publications/