Online Appendix for
"ViEWS: A political Violence Early Warning System"
Article published in *Journal of Peace Research* 56(2)
(https://journals.sagepub.com/home/jpr)

Håvard Hegre[1,3], Marie Allansson[1], Matthias Basedau[1,4], Michael Colaresi[1,2], Mihai Croicu[1], Hanne Fjelde[1], Frederick Hoyles[1], Lisa Hultman[1], Stina Högbladh[1], Remco Jansen[1], Naima Mouhleb[1], Sayyed Auwn Muhammad[1], Desirée Nilsson[1], Håvard Mokleiv Nygård[1,3], Gudlaug Olafsdottir[1], Kristina Petrova[1], David Randahl[1], Espen Geelmuyden Rød[1], Gerald Schneider[1,5], Nina von Uexkull[1], and Jonas Vestby[3]

[1]Department of Peace and Conflict Research, Uppsala University
[2]University of Pittsburgh
[3]Peace Research Institute Oslo
[4]German Institute of Global and Area Studies
[5]University of Konstanz

February 14, 2019

Abstract

This online appendix supports the article 'ViEWS: A political Violence Early Warning System' (Hegre et al., 2019), and further documents the ViEWS early-warning system. It details the various predictors used in ViEWS, defines the units of analysis in use, describes the statistical modeling, the use of downsampling, calibration, and ensemble modeling, and explain how missing and incomplete data were treated. The document continues to provide more details on the evaluation of the system, presents some additional forecasting results, and describes data management procedures in use.

Funding

ViEWS receives funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 694640) as well as from Uppsala University. ViEWS computations are performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

Contents

A Predictors

Here, we present all the predictors used in ViEWS, organized by the themes presented in Tables 2 and 3 in the main paper. We state the name the variables have in the database, report the sources for the information, and summarize how we have processed the data.

A.1 Country-level predictors

The country-level (cm) predictors are organized in six themes as detailed in Table 2 in the main paper.

Baseline

For each dependent variable the baseline model is the country-level mean of the dependent variable itself in the training period. This model is intuitively similar to a model with only individual fixed effects and no other predictors but easier for us to estimate.

Mean of state based conflict computed for the training period (`mean_ged_dummy_sb`) Country-level mean of the incidence of state based violence computed for the training period.

Mean of non-state conflict computed for the training period (`mean_ged_dummy_ns`) Country-level mean of the incidence of non-state conflict computed for the training period.

Mean of one sided violence computed for the training period (`mean_ged_dummy_os`) Country-level mean of the incidence of one sided violence computed for the training period.

Mean of protest incidence computed for the training period (`mean_acled_dummy_pr`) Country-level mean of the incidence of protests computed for the training period.

Conflict history theme

Conflict history models for each of the outcomes includes all decay functions of time since conflict events, the first order temporal lag of all outcomes and all temporal lags from 1 to 12 months of the dependent variable itself.

Decay function of time since state-based conflict event (`decay_12_cw_ged_dummy_sb_0`) Exponential decay function applied to months since state-based conflict in the country. The decay function is

$$2^{\frac{-tse}{12}}$$

where tse is time since event in months. It has a half-life of 12 months.

Decay function of time since non-state conflict event (`decay_12_cw_ged_dummy_ns_0`) Exponential 12-month half-life decay function applied to months since non-state conflict in the country.

Decay function of time since one-sided violence event (`decay_12_cw_ged_dummy_os_0`) Exponential 12-month half-life decay function applied to months since one-sided violence in the country.

Lagged state-based conflict event (`l`i`_ged_dummy_sb`). Lagged state-based conflict in the country. 1 if conflict occurred i months ago, zero otherwise. Computed for each $i \in (1, 2, .., 12)$

Lagged non-state conflict event (`l`i`_ged_dummy_ns`). Lagged non-state conflict in the country. 1 if conflict occurred i months ago, zero otherwise. Computed for each $i \in (1, 2, .., 12)$

Lagged one-sided violence event (`l`i`_ged_dummy_os`). Lagged one-sided violence in the country. 1 if conflict occurred i months ago, zero otherwise. Computed for each $i \in (1, 2, .., 12)$

Demography theme

Population size (`ln_fvp_population200`) The total population of a given country. This variable uses historical data stitched together in order to gather information from 1900 until today, and continuing with projections into the future. The variable is constructed with data from Maddison (2007), World Bank (2017) and Samir & Lutz (2008). First, we harmonize candidate data to the WDI series. We then assemble a time-series based on the following preference: WDI > Maddison WP-4. Third, the SSP projection (second scenario 'Middle of the Road') is added for the future projections. Lastly, we interpolate the logged data in order to remove missing values where this is possible. In order to have a gradual transition from one series to another we calculate a gradual exchange between the WDI value and the SSP value from 2007 until 2017. This variable is included in the natural logarithm form.

Proportion of population between 15 and 24 with at least lower secondary education (`ssp2_edu_sec_15_24_p`
The proportion of the population between 15 and 24 that has completed at least lower secondary schooling implies those that have completed lower or upper secondary school. Those that have attained tertiary education are included in this number. This variable is constructed using Samir & Lutz (2008) historical data from IIASA.

Proportion of population living in urban areas (`ssp2_urban_share_iiasa`) The proportion of the total population living in an urban area. This variable is taken from Samir & Lutz (2008) IIASA data, using historical data collected by the UN.

Economy theme

GDP per capita (`fvp_lngdppercapita200`) The natural log of GDP per capita in a given country-year. This variable uses historical data stitched together in order to gather information from 1900 until today, and continuing with projections into the future. The variable is constructed with data from Maddison (2007), World Bank (2017) and Samir & Lutz (2008). In order to be able to put the data together into a series we needed to create divisors in order to convert them all into purchasing power parity adjusted 2005 US dollars. First, we harmonize all candidate data to the WDI2005PPP (NY.GDP.MKTP.PP.KD) series (we want our measurement in 2005PPP units). We here assume that the ratio between the candidate data and WDI2005PPP series is constant over time within a country. We then assemble a time-series based on the following preference: WDI2005PPP > WDI2005 Constant US > Maddison WP-4. This preference is based on looking at the data and finding the series that fit best with the WDI data. Third, the SSP projection (second scenario 'Middle of the Road') is added for the future projections. Lastly, we interpolate the log(gdp/cap) data in order to remove missing values where this is possible.

Log GDP per capita, oilrents only (`fvp_lngdpcap_oilrent`) The natural log of GDP per capita from oilrents in a given country-year. It is calculated using the World Bank Development indicator named "Oil rents" (NY.GDP.PETR.RT.ZS) (World Bank, 2017), our GDP per capita measure and the population data noted below. For the projections into the future we calculate the average percentage of GDP per capita constituted by oil rents for each country over the last five years of data and assume that the same rate will continue into the future.

Log GDP per capita, excluding oilrents (`fvp_lngdpcap_nonoilrent`) The natural log of GDP per capita excluding oilrents in a given country-year. Computed using the same methodology as Log GDP per capita including oilrents.

Growth in Log GDP per capita, oilrents only (`fvp_grlngdpcap_oilrent`) Yearly change in the natural log GDP per capita from oilrent sources in a given country-year. Growth computed as $x_t - x_{t-1}$.

Growth in Log GDP per capita, excluding oilrents (`fvp_grlngdpcap_nonoilrent`) Yearly change in natural log of GDP per capita excluding oilrents in a given country-year. Growth computed as $x_t - x_{t-1}$.

Institutions theme

Democracy index (`fvp_democracy`) This index uses three dimensions from V-Dem 7.1 Coppedge et al. (2017) to classify the level of democracy in each country. These dimensions are centered on: free and fair elections (v2x_polyarchy), democratic participation (v2x_partip) and constraints on the executive (v2x_liberal). The higher the performance on any or each of these leads to a higher score on the democracy index, and vice versa. The variable ranges from 0 to one, where one is the highest possible level of democracy.

Democracy (`fvp_demo`) This variable is a dummy variable, indicating whether or not this country is to be regarded a democracy or not. If it does fall into the definition of democracy it is coded one, if not, it is coded 0. This classification is constructed on the basis of the Democracy index variable. To construct the categories we rely on a cube produced by plotting the three dimensions against each other. We calculate the Euclidean distance from the democracy and autocracy corners and standardize the distance to fall between 0 and 1. We then define as a democracy any country-year that is no further from the democracy corner than 0.40, and as autocratic any country that is less than 0.40 from the autocracy corner. The semi-democracies, thus, are those that are in between.

Semi-democracy (`fvp_semi`) This variable is a dummy variable, indicating whether or not this country is to be regarded a semi-democracy or not. If it does fall into the definition of semi-democracy it is coded one, if not, it is coded 0. This classification is constructed on the basis of the Democracy index variable. To construct the categories we rely on a cube produced by plotting the three dimensions against each other. We calculate the Euclidean distance from the democracy and autocracy corners and standardize the distance to fall between 0 and 1. We then define as a democracy any country-year that is no further from the democracy corner than 0.40, and as autocratic any country that is less than 0.40 from the autocracy corner. The semi-democracies, thus, are those that are in between.

Autocracy (`fvp_auto`) This variable is a dummy variable, indicating whether or not this country is to be regarded an autocracy or not. If it does fall into the definition of autocracy it is coded one, if not, it is coded 0. This classification is constructed on the basis of the Democracy index variable. To construct the categories we rely on a cube produced by plotting the three dimensions against each other. We calculate the Euclidean distance from the democracy and autocracy corners and standardize the distance to fall between 0 and 1. We then define as a democracy any country-year that is no further from the democracy corner than 0.40, and as autocratic any country that is less than 0.40 from the autocracy corner. The semi-democracies, thus, are those that are in between.

Time since pre-independence war (`ln_fvp_timesincepreindepwar`) The total number of years since the country experienced a pre-independence war. This is based on the entrance dates set by Gleditsch & Ward (1999). Included in natural logarithm form.

Time since regime change (`ln_fvp_timesinceregimechange`) The total number of years since the country experienced regime change. Defined by changes between the dummy variables fvp_demo, fvp_semi and fvp_auto. Included in natural logarithm form.

Proportion of population excluded from power (`fvp_prop_excluded`) Number of excluded ethnic groups (discriminated or powerless) in the country for a given year. Derived from the EPR 2014 update 2 dataset (Vogt et al., 2015). Missing data is filled in through linear interpolations. This is used for country-level using EPR's non-georeferenced data.

Time since independence (`ln_fvp_timeindep`) The total number of years since the country became an internationally recognized sovereign state. This is based on the entrance dates set by Gleditsch & Ward (1999). Included in natural logarithm form.

Protest theme

Decay function of time since protest event (`decay_12_cw_acled_dummy_pr_0`) Exponential 12-month half-life decay function applied to months since non-state conflict in the same country. The decay function is

$$2^{\frac{-tse}{12}}$$

where *tse* is time since event in months. It has a half-life of 12 months.

Lagged protest event (l*i*_acled_dummy_pr). Lagged ACLED protest events in the country. 1 if conflict occurred i months ago, zero otherwise. Computed for each $i \in (1, 2, .., 12)$

A.2 PRIO-GRID-level predictors

The PRIOGRID-level (*pgm*) predictors are organized in six themes as detailed in Table 3 in the main paper.

Baseline

For each dependent variable the baseline model is the country-level mean of the dependent variable itself in the training period. This model is intuitively similar to a model with only individual fixed effects and no other predictors but easier for us to estimate.

Mean of state based conflict computed for the training period (mean_ged_dummy_sb) Grid-cell mean of the incidence of state based violence computed for the training period.

Mean of non-state conflict computed for the training period (mean_ged_dummy_ns) Grid-cell mean of the incidence of non-state conflict computed for the training period.

Mean of one sided violence computed for the training period (mean_ged_dummy_os) Grid-cell mean of the incidence of one sided violence computed for the training period.

Mean of protest incidence computed for the training period (mean_acled_dummy_pr) Grid-cell mean of the incidence of protests computed for the training period.

Conflict history theme

For each dependent variable's model the conflict history theme includes:

- Decay function of time since all outcomes.
- First order temporal lags of all other outcomes.
- First to twelth order temporal lag of the dependent variable.

- First order spatial lag of first order temporal lag of all other outcomes.
- First order spatial lag of first, second and third order temporal lag of dependent variable.

State based conflict event (ged_dummy_sb) Dummy variable indicating whether there was a conflict event (at least one battle related death) in a given grid cell in a given month. State based conflicts, that is intra-state, inter-state, internationalized and extrasystemic conflicts (Sundberg & Melander, 2013). In the country models this is aggregated up to the country level using cShapes v.0.4-2. (Weidmann, Kuse & Gleditsch, 2010). Included in various transformations as a predictor and as the dependent variable for models of state based conflict.

Non-state conflict event (ged_dummy_ns) Dummy variable indicating whether a non-state conflict event (at least one battle related death) had occurred within the given grid cell. Non-state conflicts, where non-state armed groups are in conflict with one another (Sundberg & Melander, 2013). Included in various transformations as a predictor and as the dependent variable for models of non-state conflict.

One-sided conflict event (ged_dummy_os) Dummy variable indicating whether there was a conflict event (at least one death) in a given grid cell in a given month. One-sided violence regards cases where government forces or non-state armed groups engage in violence against civilians (Sundberg & Melander, 2013). Included in various transformations as a predictor and as the dependent variable for models of one-sided violence.

Lagged state-based conflict event (l*i*_ged_dummy_sb) Lagged state-based conflict in the grid cell. 1 if conflict occurred i months ago, 0 otherwise. Computed for each $i \in (1, 2, .., 12)$

Lagged non-state conflict event (l*i*_ged_dummy_ns) Lagged non-state conflict in the grid cell. 1 if conflict occurred i months ago, 0 otherwise. Computed for each $i \in (1, 2, .., 12)$

Lagged one-sided violence event (l*i*_ged_dummy_os) Lagged one-sided violence in the grid cell. 1 if conflict occurred i months ago, 0 otherwise. Computed for each $i \in (1, 2, .., 12)$

Decay function of time since state-based conflict event (`decay_12_cw_ged_dummy_sb_0`) Exponential decay function applied to months since state-based conflict in the same grid cell. The decay function is

$$2^{\frac{-tse}{12}}$$

where tse is time since event in months. It has a half-life of 12 months.

Decay function of time since non-state conflict event (`decay_12_cw_ged_dummy_ns_0`) Exponential 12-month half-life decay function applied to months since non-state conflict in the same grid cell.

Decay function of time since one-sided violence event (`decay_12_cw_ged_dummy_os_0`) Exponential 12-month half-life decay function applied to months since non-state conflict in the same grid cell.

Spatial lag of state-based conflict event (`q_1_1_l`*t*`_ged_dummy_sb`). Sum of t month time-lagged state-based conflict events in the neighbouring grid cells. Neighbouring is defined by a queens movement in chess meaning cells horizontally, vertically or diagonally adjacent. Computed for first order neighbors, meaning only directly adjacent cells are considered. Computed for $t \in (1, 2, 3)$. If all neighbouring cell had state-based conflict events one month ago q_1_1_l1_ged_dummy_sb takes the value 8. If only one neighbouring cell had state-based conflict q_1_1_l1_ged_dummy_sb takes the value 1.

Spatial lag of non-state conflict event (`q_1_1_l`t`_ged_dummy_ns`). Sum of t month time-lagged non-state conflict events in the neighbouring grid cells. Neighbouring is defined by a queens movement in chess meaning cells horizontally, vertically or diagonally adjacent. Computed for first order neighbors, meaning only directly adjacent cells are considered. Computed for $t \in (1, 2, 3)$.

Spatial lag of one-sided conflict event (`q_1_1_l`t`_ged_dummy_os`). Sum of t month time-lagged one-sidede violence events in the neighbouring grid cells. Neighbouring is defined by a queens movement in chess meaning cells horizontally, vertically or diagonally adjacent. Computed for first order neighbors, meaning only directly adjacent cells are considered. Computed for $t \in (1, 2, 3)$.

Natural geography theme

Distance to nearest secondary diamonds resource (`ln_dist_diamsec`). Captures the distance from the grid cell to the nearest secondary diamonds resource (static data). The distance is measured in logged WGS86 units (decimal degrees), a unit closely resembling the design choices for the overall grid. The original variable is named diamsec_s and is collected from PRIO-GRID, based on Klein Goldewijk et al. (2011).

Distance to nearest petroleum resource (`ln_dist_petroleum`). Captures the distance from the grid cell to the nearest petroleum resource (static data, and only onshore production). The distance is measured in logged WGS86 units (decimal degrees), a unit closely resembling the design choices for the overall grid. The original variable is named petroleum_s and is collected from PRIO-GRID, based on Klein Goldewijk et al. (2011)

Proportion of mountainous terrain (`mountain_ih_li`). Percentage area of the cell covered by mountains. Collected from PRIO-GRID, based on ISAM-HYDE landuse data (Meiyappan & Jain, 2012). Missing data is filled in through linear interpolation.

Agricultural area (`agri_ih_li`). Percentage area of the cell covered by agricultural area. Collected from PRIO-GRID, based on ISAM-HYDE landuse data(Meiyappan & Jain, 2012). Missing data is filled in through linear interpolation.

Barren area (`barren_ih_li`). Percentage area of the cell covered by barren area. Collected from PRIO-GRID, based on ISAM-HYDE landuse data (Meiyappan & Jain, 2012). Missing data is filled in through linear interpolation.

Forest area (`forest_ih_li`). Percentage area of the cell covered by forest area. Collected from PRIO-GRID, based on ISAM-HYDE landuse data (Meiyappan & Jain, 2012). Missing data is filled in through linear interpolation.

Grasslands (`savanna_ih_li`). Percentage area of the cell covered by grasslands. Collected from PRIO-GRID, based on ISAM-HYDE landuse data (Meiyappan & Jain, 2012). Missing data is filled in through linear interpolation.

Shrublands (`shrub_ih_li`). Percentage area of the cell covered by shrublands. Collected from PRIO-GRID, based on ISAM-HYDE landuse data (Meiyappan & Jain, 2012). Missing data is filled in through linear interpolation.

Pasture land (`pasture_ih_li`). Percentage area of the cell covered by pasture area, based on ISAM-HYDE land-use data. In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010 (Meiyappan & Jain, 2012). Missing data is filled in through linear interpolation.

Social geography theme

Distance to neighboring country (`ln_bdist1`). The spherical distance in kilometer from the cell centroid to the border of the nearest land-contiguous neighboring country, based on country border data using cShapes v.0.4-2. (Weidmann, Kuse & Gleditsch, 2010). Included in natural logarithm form of the original variable.

Travel time (`ln_ttime`). Log-transformed estimate of the travel time to the nearest major city, derived from a global high-resolution raster map of accessibility developed for the EU (Uchida, 2009). Collected from PRIO-GRID. Included in natural logarithm form of the original variable.

Distance to capital city (`ln_capdist`). The spherical distance in kilometers from the cell centroid to the national capital city in the corresponding country, based on coordinate pairs of capital cities derived from the cShapes dataset v.0.4-2. It captures changes over time wherever relevant (Weidmann, Kuse & Gleditsch, 2010). Included in natural logarithm form of the original variable.

Population size (`ln_pop`). Population size for each populated cell in the grid, taken from the History Database of the Global Environment (HYDE) version 3.1. Population estimates are available for 1950, 1960, 1970, 1980, 1990, 2000, and 2005. The original pixel value is number of persons. Included in natural logarithm form of the original variable. Collected from PRIO-GRID, based on Klein Goldewijk et al. (2011).

Gross cell product (`gcp_li_mer`). The gross cell product, measured in USD, based on the G-Econ dataset v4.0, last modified May 2011. The original G-Econ data represent the total economic activity at a 1x1 degree resolution, so when assigning this to PRIO-GRID we distribute the total value across the number of contained PRIO-GRID land cells. In border areas, the G-Econ 1x1 degree cells might overlap with PRIO-GRID cells allocated to a neighboring country.To minimize bias, PRIO-GRID only extracts G-Econ data for cells that have the same country code as the G-Econ cell represents. This variable is only available for five-year intervals since 1990 (Nordhaus, 2006). Missing data is filled in through linear interpolations.

Infant mortality rate(`imr_mean`). Mean infant mortality rate. Collected from PRIO-GRID, based on raster data from the SEDAC Global Poverty Mapping project (Storeygard et al., 2008).

Proportion of mountainous terrain (`mountains_mean`). Proportion of mountainous terrain within the cell based on elevation, slope and local elevation range. Collected from PRIO-GRID, taken from a high-resolution mountain raster developed for UNEPś Mountain Watch Report. (Blyth, 2002).

Urban area (`urban_ih_li`). Percentage area of the cell covered by urban area. Collected from PRIO-GRID, based on ISAM-HYDE landuse data(Meiyappan & Jain, 2012). Missing data is filled in through linear interpolations.

Excluded groups (`excluded_li`). Number of excluded ethnic groups (discriminated or powerless) in the grid cell for the given year. Collected from PRIO-GRID, derived from the GeoEPR 2014 update 2 dataset (Vogt et al., 2015). Missing data is filled in through linear interpolations. This is used for country-level using EPR's non-georeferenced data.

CM theme

Democracy (`fvp_demo`) This variable is a dummy variable, indicating whether or not this country is to be regarded a democracy or not. If it does fall into the definition of democracy it is coded one, if not, it is coded 0. This classification is constructed on the basis of A.1. To construct the categories we rely on a cube produced by plotting the three dimensions against each other. We calculate the Euclidean distance from the democracy and autocracy corners and standardize the distance to fall between 0 and 1. We then define as a democracy any country-year that is no further from the democracy corner than 0.40, and as autocratic any country that is less than 0.40 from the autocracy corner. The semi-democracies, thus, are those that are in between.

Semi-democracy (`fvp_semi`) This variable is a dummy variable, indicating whether or not this country is to be regarded a semi-democracy or not. If it does fall into the definition of semi-democracy it is coded one, if not, it is coded 0. This classification is constructed on the basis of A.1. To construct the categories we rely on a cube produced by plotting the three dimensions against each other. We calculate the Euclidean distance from the democracy and autocracy corners and standardize the distance to fall between 0 and 1. We then define as a democracy any country-year that is no further from the democracy corner than 0.40, and as autocratic any country that is less than 0.40 from the autocracy corner. The semi-democracies, thus, are those that are in between.

Time since pre-independence war (`ln_fvp_timesincepreindepwar`) The total number of years since the country experienced a pre-independence war. This is based on the entrance dates set by Gleditsch & Ward (1999). Included in natural logarithm form.

Time since regime change (`ln_fvp_timesinceregimechange`) The total number of years since the country experienced regime change. Defined by changes between the dummy variables fvp_demo, fvp_semi and fvp_auto. Included in natural logarithm form.

Proportion of population excluded from power (`fvp_prop_excluded`) Number of excluded ethnic groups (discriminated or powerless) in the country for a given year. Derived from the EPR 2014 update 2 dataset (Vogt et al., 2015). Missing data is filled in through linear interpolations. This is used for country-level using EPR's non-georeferenced data.

Time since independence (`ln_fvp_timeindep`) The total number of years since the country became an internationally recognized sovereign state. This is based on the entrance dates set by Gleditsch & Ward (1999). Included in natural logarithm form.

GDP per capita (`lnGDPpc200`) The GDP per capita in a given country-year. This variable uses historical data stitched together in order to gather information from 1900 until today, and continuing with projections into the future. The variable is constructed with data from Maddison (2007), World Bank (2017) and Samir & Lutz (2008). In order to be able to put the data together into a series we needed to create divisors in order to convert them all into purchasing power parity adjusted 2005 US dollars. First, we harmonize all candidate data to the WDI2005PPP (NY.GDP.MKTP.PP.KD) series (we want our measurement in 2005PPP units). We here assume that the ratio between the candidate data and WDI2005PPP series is constant over time within a country. We then assemble a time-series based on the following preference: $WDI2005PPP > WDI2005ConstantUS > MaddisonWP - 4$. This preference is based on looking at the data and finding the series that fit best with the WDI data. Third, the SSP projection (second scenario 'Middle of the Road') is added for the future projections. Lastly, we interpolate the log(gdp/cap) data in order to remove missing values where this is possible. This variable is coded in the natural logarithm form of the original variable, and serves as a base for calculating the oil, and non-oil GDP per capita for each country.

Log GDP per capita, oilrents only (`fvp_lngdpcap_oilrent`) The natural log of GDP per capita from oilrents in a given country-year. It is calculated using the World Bank Development indicator named "Oil rents" (NY.GDP.PETR.RT.ZS) (World Bank, 2017), our GDP per capita measure and the population data noted below. For the projections into the future we calculate the average percentage of GDP per capita constituted by oil rents for each country over the last five years of data and assume that the same rate will continue into the future.

Log GDP per capita, excluding oilrents (`fvp_lngdpcap_nonoilrent`) The natural log of GDP per capita excluding oilrents in a given country-year. Computed using the same methodology as Log GDP per capita including oilrents.

Growth in Log GDP per capita, oilrents only (`fvp_grlngdpcap_oilrent`) Yearly change in the natural log GDP per capita from oilrent sources in a given country-year. Growth computed as $x_t - x_{t-1}$.

Growth in Log GDP per capita, excluding oilrents (`fvp_grlngdpcap_nonoilrent`) Yearly change in natural log of GDP per capita excluding oilrents in a given country-year. Growth computed as $x_t - x_{t-1}$.

Population size (`ln_fvp_population200`) The total population of a given country. This variable uses historical data stitched together in order to gather information from 1900 until today, and continuing with projections into the future. The variable is constructed with data from Maddison (2007), World Bank (2017) and Samir & Lutz (2008). First, we harmonize candidate data to the WDI series (we want our measurement in thousands). We then assemble a time-series based on the following preference: $WDI > MaddisonWP-4$. Third, the SSP projection (second scenario 'Middle of the Road') is added for the future projections. Lastly, we interpolate the logged data in order to remove missing values where this is possible. In order to have a gradual transition from one series to another we calculate a gradual exchange between the WDI value and the SSP value from 2007 until 2017. This variable is included in the natural logarithm form of the original variable.

Proportion of population between 15 and 24 with at least lower secondary education (`ssp2_edu_sec_15_24_p` The proportion of the population between 15 and 24 that has completed at least lower secondary schooling implies those that have completed lower or upper secondary school. Those that have attained tertiary education are included in this number. This variable is constructed using Samir & Lutz (2008) historical data from IIASA.

Proportion of population living in urban areas (`ssp2_urban_share_iiasa`) The proportion of the total population living in an urban area. This variable is taken from Samir & Lutz (2008) IIASA data, using historical data collected by the UN.

Protest theme

Protest event (`acled_dummy_pr`). Dummy variable indicating whether there was a protest event as defined by ACLED (Armed Conflict Location and Event Dataset) in a given grid cell in a given month (Raleigh et al., 2010). Included in various transformations as a predictor and as the dependent variable for models of protest event incidence. Models of protest are used as auxilliary models in our dynamic simulation forecasts.

Decay function of time since protest event (`decay_12_cw_acled_dummy_pr_0`) Exponential 12-month half-life decay function applied to months since non-state conflict in the same country. The decay function is

$$2^{\frac{-tse}{12}}$$

where tse is time since event in months. It has a half-life of 12 months.

Lagged protest event (li_acled_dummy_pr). Lagged ACLED protest events in the country. 1 if conflict occurred i months ago, zero otherwise. Computed for each $i \in (1, 2, .., 12)$

Spatial lag of protest event (`q_1_1_l`t`_acled_dummy_pr`). Sum of t month time-lagged state-based conflict events in the neighbouring grid cells. Neighbouring is defined by a queens movement in chess meaning cells horizontally, vertically or diagonally adjacent. Computed for first order neighbors, meaning only directly adjacent cells are considered. Computed for $t \in (1, 2, 3)$.

B Levels of analysis and dependent variables

B.1 Levels of analysis

ViEWS generates forecasts at two levels of analysis: country months (Gleditsch & Ward, 1999, abbreviated *cm* in ViEWS), and sub-national geographical location months (*pgm*). The *cm* level is particularly useful to provide predictions for entirely new conflicts where no known actors exist, and to model tensions and processes at the government level. The set of countries is defined by the Gleditsch-Ward country code (Gleditsch & Ward, 1999, with later updates), and the geographical extent of countries by the latest version of CShapes (Weidmann, Kuse & Gleditsch, 2010). Note that the *cm* and *pgm* definitions are not fully compatible with each other. PRIO-GRID provides a 1:1 cell-to-country correspondence by assigning the grid cell to the country taking up the largest area (Tollefsen, 2012). When PRIO-GRID cells span two or more countries, all events contained in that PRIO-GRID cell are aggregated, ignoring which country they actually took place in. In the country-month dataset, such events are assigned to the country where the event took place. Moreover, PRIO-GRID cells exist for the entire duration of the dataset, but only those months in which a country has existed in the Gleditsch & Ward (1999) country list are included in the *cm* datasets.

For the subnational forecasts, ViEWS relies on the PRIO-GRID (version 2.0 Tollefsen, Strand & Buhaug, 2012), a standardized spatial grid structure consisting of quadratic grid cells that jointly cover all areas of the world at a resolution of 0.5 x 0.5 decimal degrees. Near the equator, a side of such a cell is 55 km. This resolution is close to the precision level of the data we have for the outcomes. Investigating the spatial error of the UCDP-GED in Afghanistan, Weidmann (2014, p.1143) found that most events were "located within 50 km of where they actually occured". Given this, a finer resolution might not yield more precise forecasts.

We have retrieved the grid-level structure directly from the PRIO-GRID API to ensure full compatibility.

B.2 The dependent variable: recent history

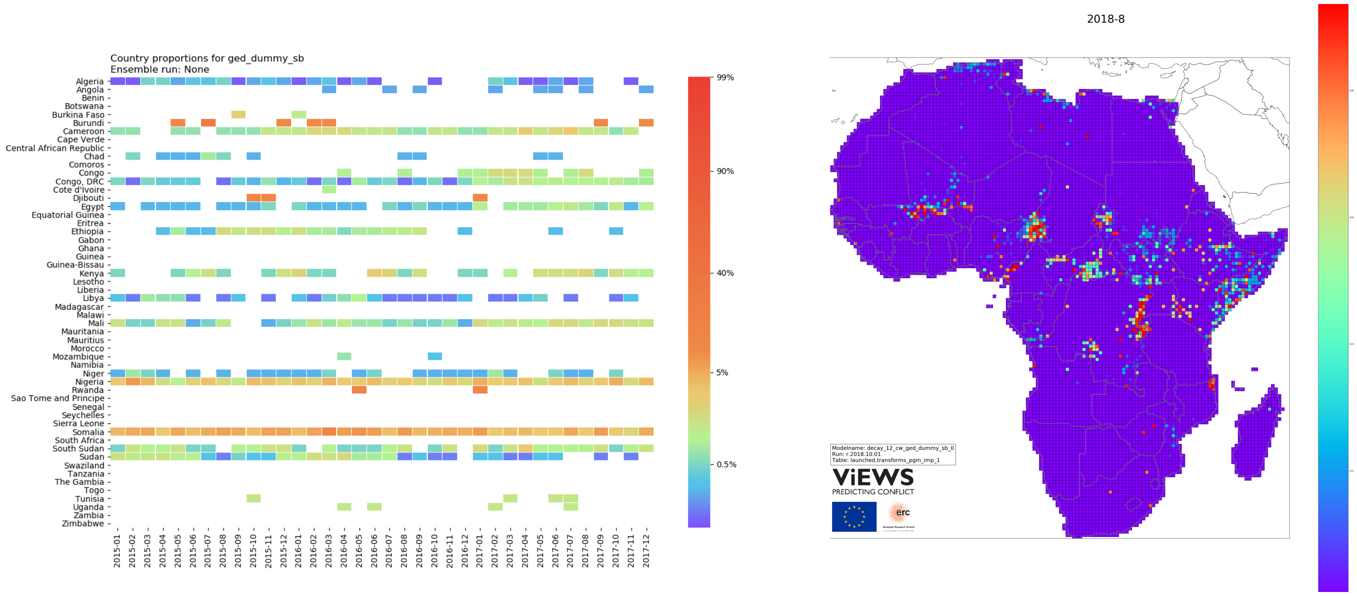
ViEWS generates predictions for the three forms of organized violence coded by the UCDP (Melander, Pettersson & Themnér, 2016): State-based conflict (**sb**), one-sided violence against civilians (**os**), and non-state conflict (**ns**).[1] Figure 1 summarizes the most recent observations for the **sb** outcome variables. Figure 1a shows the proportion of PRIO-GRID cells with at least one event for each month in the 2015–2017 period. Figure 1b depicts the recent history of violence in each PRIO-GRID cell. Red cells had conflict in August 2018, and purple ones have not seen conflict in many years. Figure 2 and 3 show the same for the other two outcome variables.

Conflict data are primarily obtained from UCDP-GED and take the form of events (Sundberg & Melander, 2013). Historical data covering 1989–2017 are extracted from the UCDP GED version 18.1 (Croicu & Sundberg, 2013; Allansson, Melander & Themnér, 2017; Pettersson & Eck, 2018).[2] Newer data are provided by the new UCDP-Candidate dataset which is updated monthly (see Hegre et al., 2018, for an introduction). This allows using conflict events up to one month before the forecasting window. Here, we use data including August 2018.

We aggregate the UCDP-GED events up to our two levels of analysis.

[1] See Melander, Pettersson & Themnér (2016) and `https://www.pcr.uu.se/research/ucdp/definitions/` for detailed definitions.

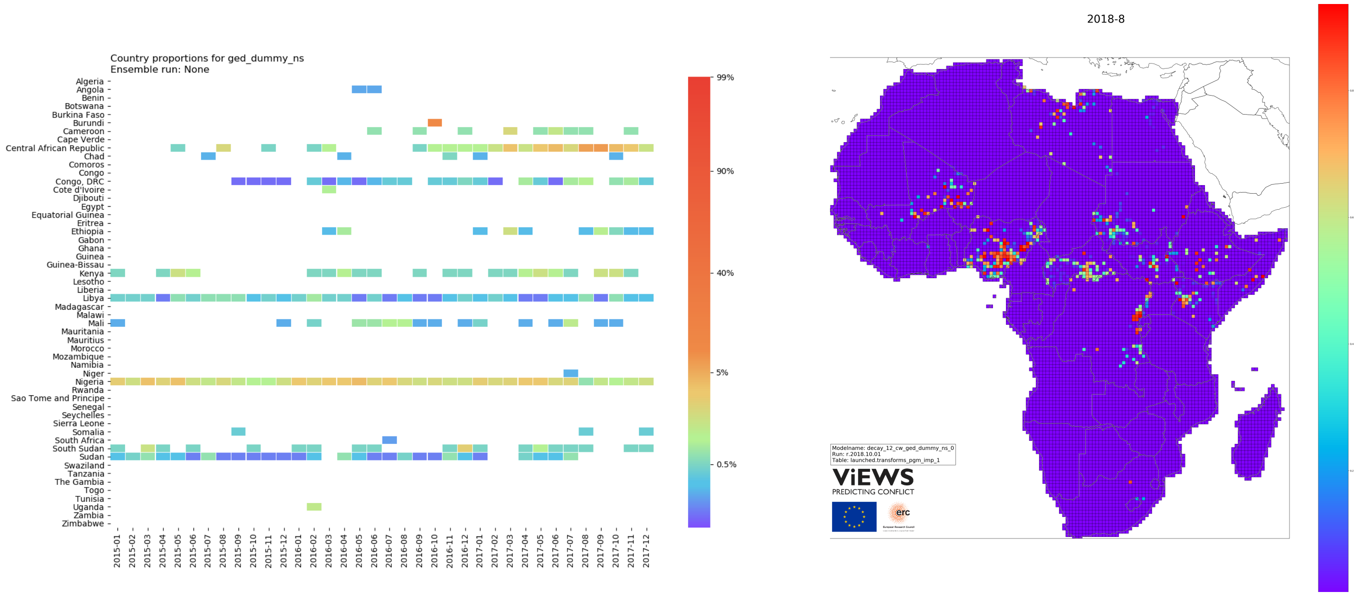
[2] The UCDP-GED raw data are publicly available through the UCDP-GED API.(Croicu & Sundberg, 2013). ViEWS automatically retrieves these data from the API each month and aggregate to our units of analysis as described in Hegre et al. (2018). Usage of the API is described in `http://ucdp.uu.se/apidocs/`; the data are available as version 18.1 (1989–2017).



(a) Observed proportion of PRIO-GRID cells with conflict, by country and month, 2015–2017

(b) Decay function of time since most recent event up to August 2018, halflife 12 months

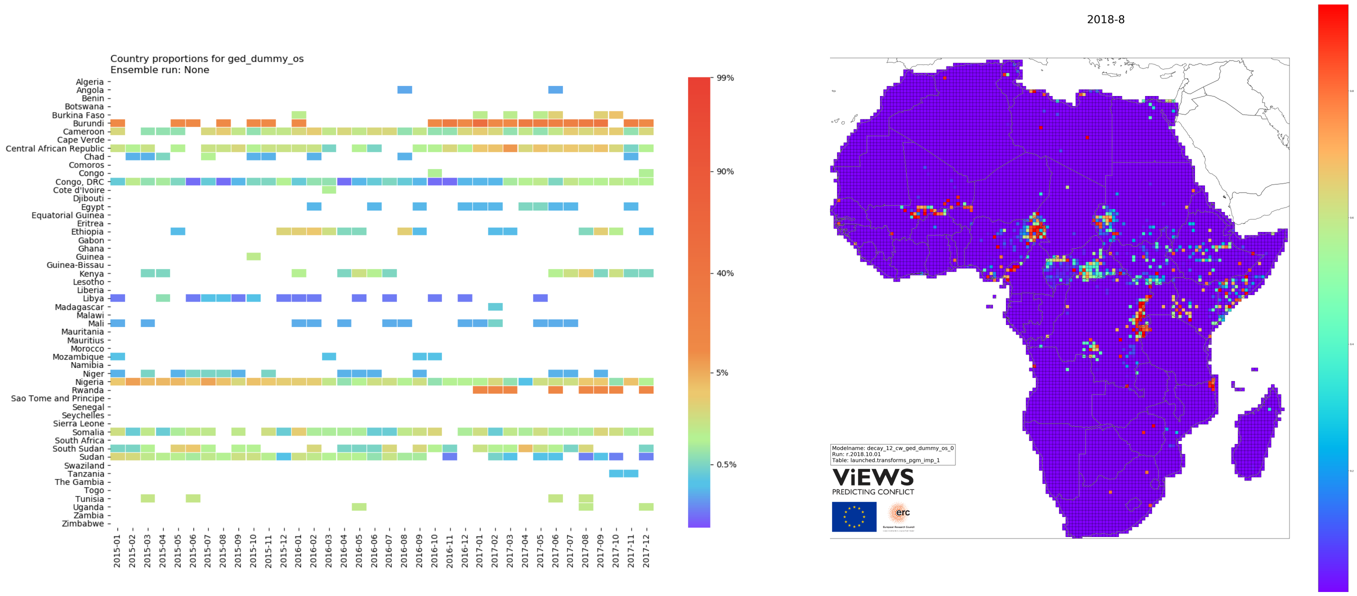
Figure 1. State-based conflict (**sb**), *pgm* level, as recorded in Sundberg & Melander (2013) and Hegre et al. (2018).



(a) Observed proportion of PRIO-GRID cells with conflict, by country and month, 2015–2017

(b) Decay function of time since most recent event up to August 2018, halflife 12 months

Figure 2. Non-state conflict (**ns**), *pgm* level, as recorded in Sundberg & Melander (2013) and Hegre et al. (2018).



(a) Observed proportion of PRIO-GRID cells with conflict, by country and month, 2015–2017

(b) Decay function of time since most recent event up to August 2018, halflife 12 months

Figure 3. One-sided violence (**os**), *pgm* level, as recorded in Sundberg & Melander (2013) and Hegre et al. (2018).

B.3 The persistence of conflict in Africa

Our forecasts for state-based conflict in Africa are very stable over time (Figure 8 in main article and Figure 13 below). This to a large degree reflect that conflict patterns have been very persistent over the past decade.

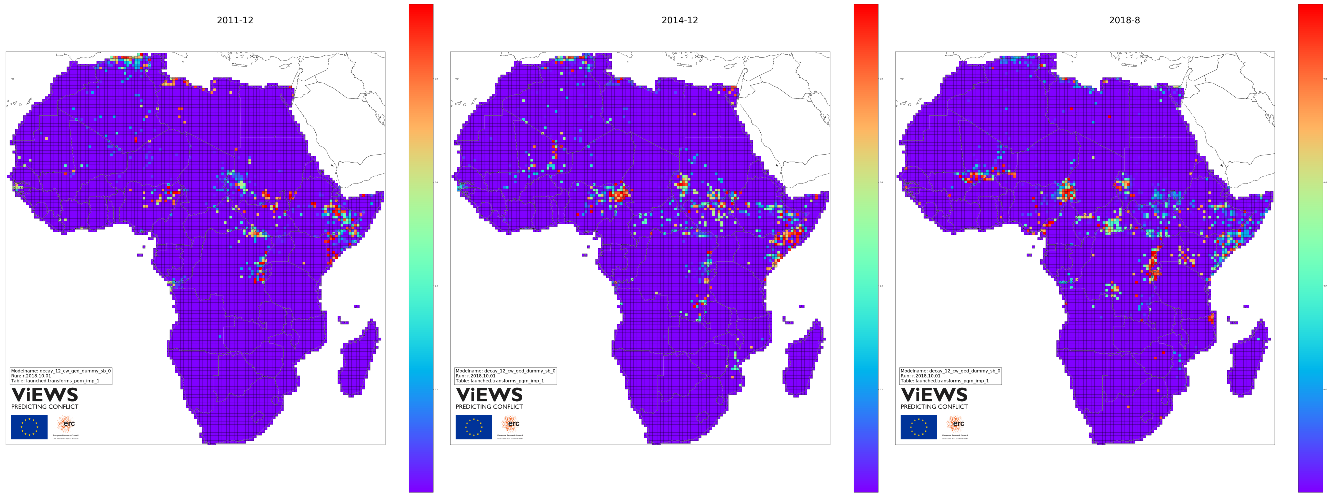


Figure 4. Conflict history at three points in time: December 2011 (left), December 2014 (middle), August 2018 (right)

Figure 4 shows the recent conflict history at the end of the training period (December 2011) as well as the end of the calibration period (December 2014). At both points in time, conflicts in Algeria, Northern Nigeria, Sudan, South Sudan, Somalia, and DR Congo were dominating. By 2014, Mali had entered the fray. These conflict hot spots were the same as those dominating the test period (2015–2017), as shown in the right-most map of the situation in August 2018. These conflicts both influenced the training and calibration of the ensemble and contributed heavily to the conflict history available in December 2014, and our predictions for these locations were depressingly accurate. Our slow-moving predictors (political institutions and geography) reinforced this pattern.

B.4 Descriptive statistics

The evaluation metrics we use are dependent on the data they are evaluated for. In particular, they are all in varying ways dependent on class balance. Table 1 shows that the *pgm* dataset has very strong class imbalance. 0.3% of pgms had state-based conflict events, 0.1% had non-state conflict events, and 0.2% had one-sided violence.

avg_sb	0.003227
avg_ns	0.001123
avg_os	0.002112
avg_decay_sb	0.044849
avg_decay_ns	0.033029
avg_decay_os	0.039404
stdev_decay_sb	0.129578
stdev_decay_ns	0.099688
stdev_decay_os	0.116185

Table 1. Descriptive statistics of dependent variables 1990-2018. Also includes decay functions of those dependent variables

C Statistical modeling

C.1 Estimation of constituent models

ViEWS relies on logistic regression and random forest 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. The random forest model (Breiman, 2001; Muchlinski et al., 2016) is a machine-learning technique based on a combination of classification and regression trees (CART), bootstrap-aggregating (bagging), and random feature selection. 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). CART can be combined with bagging to create 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. Because random forest models are computationally intensive, we estimate them using a 'downsampled' dataset which includes all conflict events and a random sample of non-events (see details in Section E.1 in this appendix).

D Ensemble Bayesian Model Averaging

In ViEWS, we rely on the average prediction from all models to create our ensemble. However, we also implemented an Ensemble Bayesian Model Averaging (EBMA) approach (Montgomery, Hollenbach & Ward, 2012; Beger, Dorff & Ward, 2014; Raftery & Lewis, 1992; Raftery et al., 2005). EBMA weighs the constituent models before combining them in an ensemble. The weights are based on the performance of the constituent models in the calibration period. Overall, we have found that the two approaches yield very similar results (for similar conclusions in a different context, see Graefe et al., 2015).

That the weighted approach of EBMA does not outperform the unweighted (or equally-weighted) average prediction is potentially because the relative performance of the constituent models varies over time. The

EBMA computes weights so that the performance of the ensemble of constituent models is optimized for the actual outcomes in the calibration period. However, these weights may not be optimal for other time periods (the testing/forecasting period). The conflicts that happened to occur in the calibration period may be driven by certain factors represented in one or two of our themes (say for example ethnic cleavages or coup d'etats). If these factors are not equally important in the conflicts that occur in the testing/forecasting period, the ensemble suffers from over-fitting to the calibration data. Such clustering of conflict with similar causes in time (and space) is not unusual, consider for example the numerous conflicts in the immediate aftermath of the breakdown of the Soviet Union or the Arab Spring uprisings. In contrast, the unweighted average ensembles are not affected by such temporal fluctuations since the weights are not given by the data at all.

Another weakness that the EBMA approach shares with the equally-weighted average of predicted probabilities is that the weights are applied uniformly to all cases. It is possible that some themes are more important in some subsets of the countries or geographical locations in Africa. In that case, models should be weighted differently in different sub-regions. ViEWS will look into various stacking techniques to see whether these approaches can yield better aggregated forecasts.

Since the EBMA results are very similar to the unweighted average, and EBMA involves considerable additional complexity, the ViEWS forecasts are currently based on the latter. However, we will continue to combine models using both ensemble methods and expand to other ensemble methods to accumulate evidence on their relative performance for conflict forecasting.

D.1 Combining three types of political violence outcomes

The figures depicting our forecasts (Figures 5 and 7 in the main article, and additional figures in Section J), are relatively similar – locations with a high predicted probability of one type of violence also has high risk of the other two. This partly reflects that the various forms of organized violence occur through related processes and are constrained by the same factors. In addition, there is considerable spillover between the forms of violence. One-sided violence, for instance, is most frequently perpetrated in the context of a state-based conflict. We believe there is ample scope for improving the system by modeling more carefully how the three outcomes affect each other and how they are distinct.

Combining these three outcomes in a single system brings several advantages. First, they together constitute a reasonable definition of political violence that subsumes conflicts such as the ongoing war in Syria, the 1994 genocide in Rwanda, and drug-related organized violence in Mexico. The system allows them to be modeled separately since they follow different dynamics and involve different types of actors. At the same time, ViEWS allows the various types of violence to serve as early-warning indicators for each other.

E Downsampling and calibration

E.1 Downsampling

A majority of the models in ViEWS were trained on all available observations. All our random forests, however, were trained on a downsampled dataset. When downsampling, we keep all conflict events and randomly sample a proportion of the remaining observations. This serves two purposes.

First, it reduces the computational burden. The *pgm* unit of analysis consists of about 11,000 units for Africa only. Sampled monthly over the 1990–2014 period for the training dataset, this amounts to a dataset with about 3.21M rows. Only 13,739 of these – 0.4% – contain observations of UCDP-GED events. To facilitate the estimation of the computationally intensive random forest models using this data, we trained them on a dataset containing all *pgm* units with at least one UCDP-GED event and a random sample of 10% of the remaining observations.

Second, downsampling is a way of inducing an asymmetric cost-function into our computation. Assuming that incorrectly predicting peace when there is violence is more costly than predicting violence when there is actually peace, then we would like our forecasts to hue more closely to predicting the events of violence, even at the cost of over-estimating some violence in peaceful circumstances. By reducing the proportion of non-events, we also reduce their influence on the fitted models relative to the instances where events occurred. If observations that result in events and non-events are weighted equally, then any unique signal in the rare minority class may be lost. For example, a distinct, but rare, data-generation process might lead to a higher probability of an event as compared to most observations, which will have a lower probability of an event. In this case, downsampling will help our algorithms learn the patterns in cases where violent events occurred, instead of those patterns being treated as random, rare, noise deviating from the more frequent non-events. This should produce higher precision, for example, as compared to non-weighted training where events are highly infrequent because the model has been trained to work harder to predict events, as compared to non-downsampled cases (Ricardo Barandela & Rangel, 2003; Chao Chen & Breiman, 2004).

This procedure leads to more predictions of events by artificially shifting the mean upwards. Our calibration procedure, described below, transform these predictions so that they in aggregate yield a predicted conflict intensity that is as close to the actual as possible.

E.2 Calibration

Forecasting requires that each model is well calibrated: that the average predicted outcome probabilities for a set of cases is similar to the actual relative frequency for that set. Models that were trained on a downsampled dataset do not have this property, and require calibration. The same applies to the models that are constructed as the product of *cm* and *pgm* probabilities. We therefore calibrate the constituent model before entering them into the ensemble.

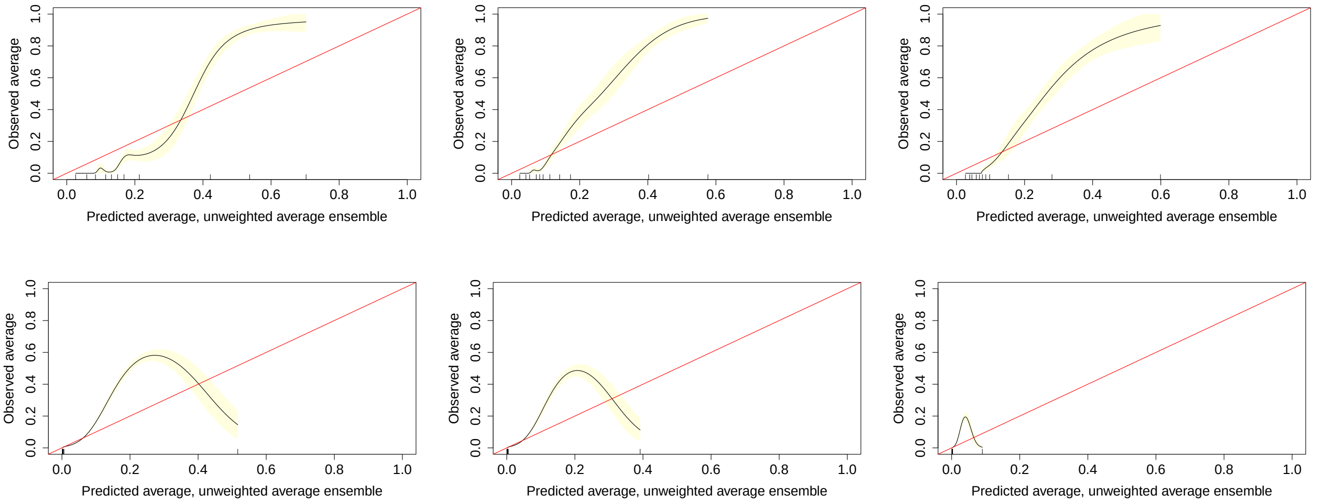


Figure 5. Calibration plots, *cm* (top) and *pgm* (bottom) level. Left: **sb**. Centre: **os**. Right: **ns**. Note that the x-axis and y-axis is different in the bottom right plot.

We use the calibration partition to calibrate the models. We obtain recentering and rescaling parameters γ_{0i}, γ_{1i} by estimating logistic regression models for each constituent model on the calibration period:

$$logit(p(Y^c_v = 1)) = \hat{\gamma}_{0i} + \hat{\gamma}_{1i} z^c_{iv}$$

where z^c_{iv} is the log odds of conflict for model i on conflict type v. The rescaling parameters $\hat{\gamma}_{0i}, \hat{\gamma}_{1i}$ are then used to shift and strengthen the probabilities in the forecasting period by

$$\hat{p}_{cal}(Y^c_v = 1) = \frac{e^{\hat{\gamma}_{0i} + \hat{\gamma}_{1i} z^c_{iv}}}{e^{(\hat{\gamma}_{0i} + \hat{\gamma}_{1i} z^c_{iv})} + 1}$$

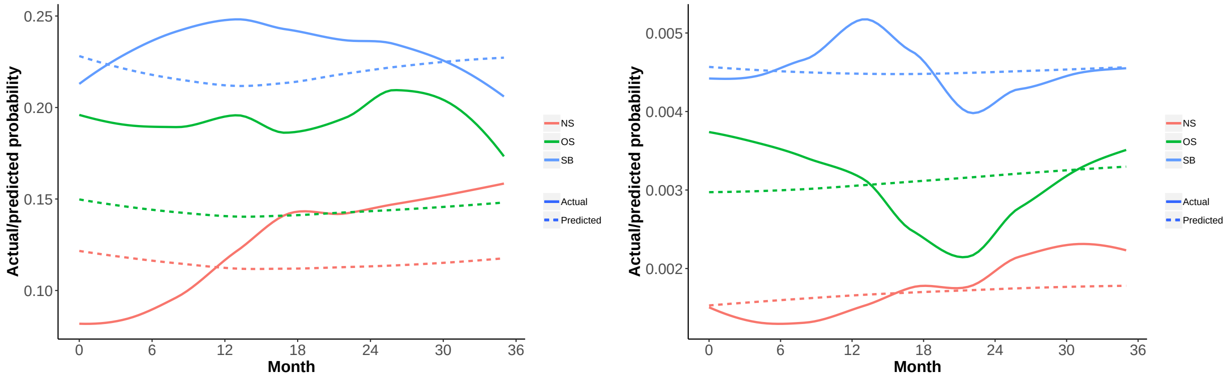


Figure 6. Calibration over time, *cm* (left) and *pgm* (right). The solid lines are smoothed with a loess function. Note that the y-axis differs in each plot.

where $\hat{p}_{cal}(Y_v^c = 1)$ is our calibrated predicted probability of conflict.

If a model is well calibrated, then an event occurs approximately x percent of the time when the model suggests that there is an x percent chance of an event occurring. This can be gauged visually with calibration plots. In calibration plots, the predicted probabilities are binned on the x-axis and the frequency of actual events within the observations in each bin is plotted on the y-axis. A perfectly calibrated model follows a 45 degree angle. Deviations indicate that the model underpredicts or overpredicts. We show calibration plots for our six ensembles in Figure 5. The top panel plots the *cm* ensembles. Overall, the *cm* ensembles assign both too low and too high probabilities. On the left hand side of each plot, we can see that the predicted probabilities are lower than the actual probability. In the middle of the plots, however, the predicted probability is too high. The bottom panel plots the *pgm* ensembles. Here, we can also see that all three ensembles assign both too low and too high probabilities.

We can also evaluate how the calibration of models change over time. In Figure 6, we display the mean actual/predicted probability of conflict on the y-axis, and months in the testing/forecasting period on the x-axis. The colors indicate the conflict type, blue for **sb**, green for **os**, and red for **ns**. Moreover, solid lines are the observed relative frequencies and the dotted lines the predicted probabilities from the unweighted average ensembles. Figure 6 shows that the ensembles are relatively well-calibrated. However, it striking that the predicted probabilities are relatively constant over time, while the observed relative frequencies fluctuate. This is to be expected, given that many of the inputs used for forecasting are constant over time.

F Handling missing or incomplete data in ViEWS

F.1 Dependent variables

UCDP-GED includes high-resolution temporal and geographical references. In about 15% of the cases, however, UCDP has been unable to identify the location more precisely than for instance a given second-order administrative region. In such cases, the UCDP assigns the center point of the region as a place-holder location and marks the event with a precision score. For prediction purposes, the place-holder solution is not optimal. Hence, ViEWS has developed a method for multiple imputation of their locations, as documented in Croicu & Hegre (2018). This method employs the locations of precisely known events within the same conflict and within close temporal proximity to determine an empirical spatial probability distribution of latent conflict propensity for each uncertain event. We then sample this empirical probability distribution multiple times. All the forecasts reported by ViEWS are based on a set of 5 imputed location datasets. Croicu & Hegre (2018) show that this improves the predictive performance of the system considerably.

F.2 Predictors

Some variables in the predictor data used by the ViEWS project contains a certain amount of missing data. This is problematic for several reason. Firstly, making predictions require the data to be complete, i.e. all values for the variables we use to make predictions must be known. Secondly, not using appropriate methods for handling the missing data in the training models risk creating biased parameter estimates and/or standard errors (Allison, 2009) which may have an adverse affect on the predictive capabilities of the models.

Depending on the mechanisms for how the missingness appear, different missing data methods have different advantages. The most common method for handling missing data is to simply remove all observations which are not complete in *listwise deletion* (Lall, 2016). Listwise deletion does, however, assume that the data are missing completely at random, i.e. that the reason for the missingness is independent both of the value of the variable itself *and* independent of the values of the other variables in the model. In social sciences this is most often not a reasonable assumption. If it does not hold, the results from the analysis will be biased (Allison, 2009; **?**; **?**).

An alternative method for handling missing data which does not require this assumption is imputation. Here, the missing data are replaced (imputed) with some plausible values. Several possible imputation methods exist. Among the more naive are to use either the mean or the predicted values from a linear regression of the variable on all other variables. Mean and regression imputation do, however, also bias the results unless some strong assumptions are fulfilled (Buuren, Buuren). The solution to these problems is to use multiple imputation. [3]

The ViEWS project uses Multiple Imputation with the Amelia II package in R to replace the missing data. Amelia II uses bootstrapping and the expectation-maximization (EM) algorithm to impute each missing value *(*m) times to create *m* complete datasets. The number of imputations, *m*, affects a number of statistical quantities, including power and efficiency. To achieve reasonable statistical efficiency, as few as 5 complete imputed datasets are needed. A higher number of imputations do, however, lead to both higher efficiency and higher power and precision (see for instance Graham, Olchowski & Gilreath, 2007; White, Royston & Wood, 2011). As imputation is computationally intensive in large datasets, and the ViEWS project due to its focus on prediction is less concerned with statistical power, five imputations are currently used. Our dynamic simulation procedure uses all five datasets simultaneously and the results are aggregated using the Rubin Rules (Rubin, 1987). In the one-step-ahead forecasts, only one imputed dataset is used at this time. Using one dataset instead of five will not bias the results, but will reduce the statistical efficiency (Buuren, Buuren).

The ViEWS project will conduct further tests on how different missing data techniques affect prediction and simulation, in order to create a best practice for missing data in predictive studies.

G Detailed evaluation, state-based conflict (sb)

In this section, we present evaluation results for each individual model in the ensemble for **sb**, biseparation plots, and confusion matrices.

The following metrics are computed (Géron, 2017, pp.86–95):

Precision (Pr):

$$Pr = \frac{TP}{TP + FP}$$

Recall or Sensitivity (R):

$$R = \frac{TP}{TP + FN}$$

Accuracy (A):

[3] For a more comprehensive test of missing data methods see Randahl (2016).

$$A = \frac{TN + TP}{TN + FP + FN + TP}$$

Brier Score:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - A_i)^2$$

where $\hat{p}_i$ is the prediction for observation i and A_i what actually occurred.

The precision, recall/sensitivity, and specificity measures can be used for single thresholds but also aggregated to apply to all thresholds possible by the data at hand. ViEWS makes use of the Area Under the curve of the Receiver Operating Characteristic (AUROC), Area Under the Precision-Recall curve (AUPR), Brier score, and Accuracy. The AUROC and AUPR metrics range from 0 to 1, with high values signifying good predictive performance. AUROC is based on the ROC curve, which plots the true positive rate[4] ($TPR = \frac{TP}{TP+FN}$) over the true negative rate ($TNR = \frac{TN}{FP+TN}$) for each possible threshold.[5] AUROC scores are high for models that correctly recall a large fraction of the positives for any given level of false alarms. AUPR is based on the PR plot, which plots precision ($Pr = \frac{TP}{TP+FP}$) over recall ($R = \frac{TP}{TP+FN}$).[6] AUPR scores are high for models that are correct in a large fraction of the positive predictions for any given level of recall or true positive rate. The Brier score is defined as $BS = \frac{1}{N}\sum\limits_{i=1}^{N}(\hat{p}_i - A_i)^2$ where $\hat{p}_i$ is the prediction for observation i and A_i what actually occurred. Brier scores range from 0 to 1, and lower scores correspond to better performance.

Accuracy is the proportion of cases that are correctly classified: $A = \frac{TN+TP}{TN+FP+FN+TP}$. The accuracy metric differs from the other three in that it is defined for only a single threshold. We select the threshold that minimizes the misclassification costs in the calibration period of the models we evaluate. We assign a cost of 1 for each false positive and a cost of 10 to each false negative in the test period, as we see failing to predict conflicts that happened as more problematic than failing to predict non-conflicts. The reward for TP and TN is zero. The confusion matrices for each of the models we evaluate are reported below, as well as the thresholds for each of them. We also report the F_1-score which is the geometric mean of precision and recall for that threshold, as well as the threshold probability.

We choose to rely on this suite of performance metrics because model performance is multidimensional. In many model comparisons, one model outperforms others in terms of all measures, so we can safely conclude on the best model. In other situations, the picture is less consistent and multiple metrics reflect this. In particular, while the Brier score favors sharp, accurate probabilistic predictions (near 0 or 1), the relative ordering of the forecasts are used for AUROC and AUPR. Moreover, since the AUROC captures the trade-off between producing a large number of true positives versus the expense of many false alarms, the metric favors models that are good at correctly predicting no-conflict cases. The AUPR, on the other hand, only focuses on the positive cases, since it captures the trade-off between maximizing the proportion of positive predictions that are correct versus identifying as many of the actual conflicts as possible. Consequently, the AUPR is much less likely to reward models that excel at predicting non-conflict cases.

Since we are more interested in predicting instances of political violence than the absence of such, we give priority to the AUPR over the AUROC, as the former rewards models more for accurately predicting a one, as compared to a zero. Accuracy and the associated contingency matrices are intuitive since they refer to specific predictions for a given threshold. Still, they may not be the best metric in our case since our models typically are excellent in identifying cases of no-conflict but we are more interested in the conflict cases.

[4] We use the conventional notation that TN, FN, TP, FP refer to true negatives, false negatives, true positives and false positives respectively.

[5] A 'threshold' defines a probability $p*$ over which the system yields a positive prediction. A threshold of 0.5, for instance, means we predict a positive if $\hat{p} > 0.5$ and a negative if not. Specificity is $1 - TPR$, where TPR is the true positive rate.

[6] Note that 'sensitivity', 'true positive rate' and 'recall' are synonyms.

G.1 AUROC, Brier, and AUPR, *cm* level

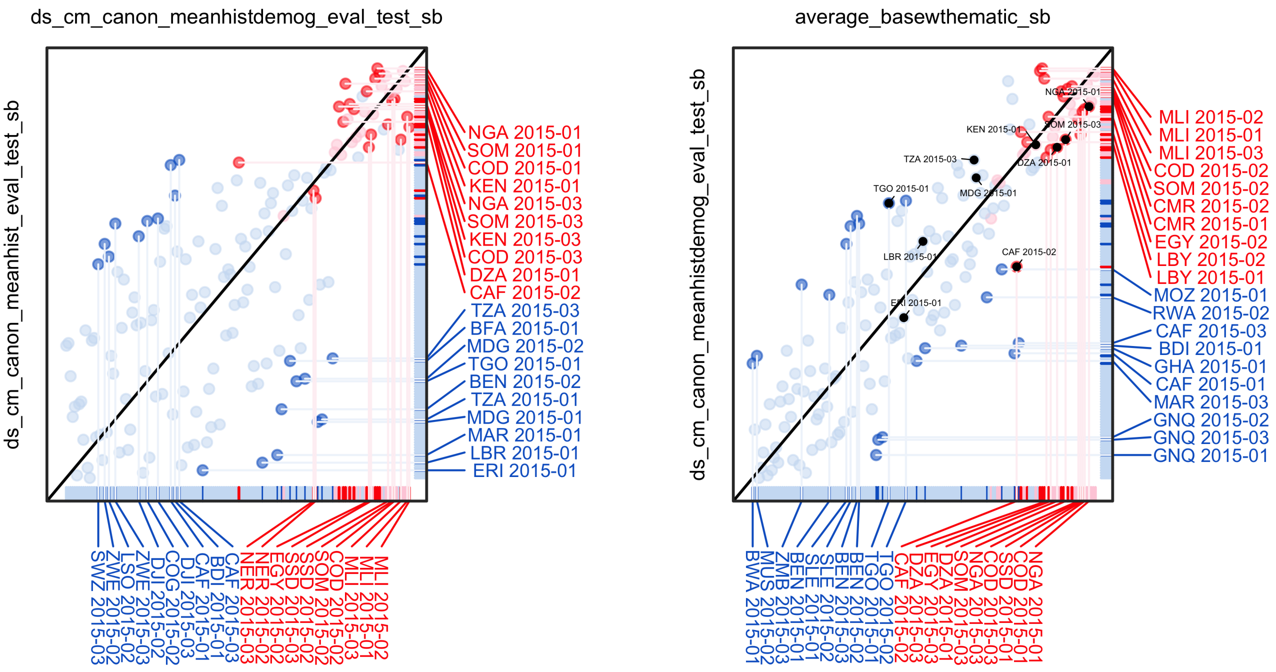
Table 2 shows the main evaluation metrics for all the constituent models in the *cm*-level ensemble, as well as for the ensemble itself. Models are named according the following convention:

- 'ds' and 'osa' refers to the way the model handles dynamics (see methodology section in main paper).
- 'cm' refers to the level of analysis.
- 'acled' or 'canon' denotes whether the models are trained on the period for which we have ACLED data (1997–) or the complete dataset (1990–).
- 'base', 'mean', 'demog', 'eco', 'hist', 'inst', 'protest' refers to a theme as listed in Table 2 in the main paper. 'base' refers to all themes, 'mean' to the baseline, 'demog' to demography theme, 'eco' to economy theme, 'hist' to conflict history theme, 'inst' to institution theme, and 'protest' to protest theme.
- 'logit_fullsample' and 'rf_downsampled' refers to the logit and random forest specifications of 'osa'.
- 'sb' refers to conflict type.
- The final line, 'average_basewthematic' is the ensemble of all the other models.

	ROC_AUC	Brier_Score	PR_AUC
ds_cm_acled_base_eval_test_sb	0.94790	0.06578	0.84879
ds_cm_canon_base_eval_test_sb	0.94624	0.06537	0.84145
ds_cm_canon_mean_eval_test_sb	0.82375	0.12524	0.67472
ds_cm_canon_demog_eval_test_sb	0.75827	0.16109	0.41894
ds_cm_canon_eco_eval_test_sb	0.69813	0.16647	0.42729
ds_cm_canon_hist_eval_test_sb	0.94680	0.06918	0.85272
ds_cm_canon_inst_eval_test_sb	0.66490	0.16757	0.39481
ds_cm_canon_protest_eval_test_sb	0.70354	0.22020	0.35786
osa_cm_acled_base_eval_test_logit_fullsample_sb	0.95197	0.06883	0.84474
osa_cm_canon_base_eval_test_logit_fullsample_sb	0.95483	0.06569	0.85701
osa_cm_canon_mean_eval_test_logit_fullsample_sb	0.85661	0.13868	0.57983
osa_cm_canon_eco_eval_test_logit_fullsample_sb	0.69165	0.16340	0.46114
osa_cm_canon_hist_eval_test_logit_fullsample_sb	0.94946	0.07051	0.86608
osa_cm_canon_inst_eval_test_logit_fullsample_sb	0.67326	0.16576	0.48685
osa_cm_canon_protest_eval_test_logit_fullsample_sb	0.68583	0.20764	0.34425
osa_cm_acled_base_eval_test_rf_downsampled_sb	0.95488	0.07184	0.86883
osa_cm_canon_base_eval_test_rf_downsampled_sb	0.95287	0.07362	0.85618
osa_cm_canon_demog_eval_test_logit_fullsample_sb	0.77214	0.15714	0.42471
osa_cm_canon_mean_eval_test_rf_downsampled_sb	0.86299	0.13158	0.59163
osa_cm_canon_demog_eval_test_rf_downsampled_sb	0.84481	0.13794	0.61618
osa_cm_canon_eco_eval_test_rf_downsampled_sb	0.76901	0.15406	0.52316
osa_cm_canon_hist_eval_test_rf_downsampled_sb	0.93942	0.08453	0.83937
osa_cm_canon_inst_eval_test_rf_downsampled_sb	0.78731	0.15359	0.62807
osa_cm_canon_protest_eval_test_rf_downsampled_sb	0.57568	0.17813	0.24507
average_basewthematic_sb	0.95549	0.09318	0.86930

Table 2. SB constituent models and ensembles, cm level, 36 months

Overall, Table 2 shows that there are big differences between the models. As expected, the theme models perform poorer than the models with all predictors included. Importantly, while there are a few models that perform similar to the ensemble, the performance of these are likely to fluctuate depending on the time period we use for evaluation. Combining all models in the ensemble therefore adds robustness to the forecast. That said, the ensemble suffers somewhat with respect to the sharpness of predictions in comparison to a number of other models.



(a) Baseline + history + demography model (horizontally) vs. baseline + history only (vertically)

(b) Ensemble model (horizontally) vs. Baseline + history + demography model (vertically)

Figure 7. Bi-separation plots, *cm* level

Bi-separation plots, *cm* level

To evaluate the difference in performance in more detail, we show bi-separation plots (Colaresi & Mahmood, 2017). Figure 7a shows a bi-separation plot for the model combining the baseline, history, and demography themes on the horizontal axis and the same for the simpler baseline and history model on the vertical, and a scatter plot for each observation in the *cm* evaluation period. Country-months with observed conflict are represented with red color, and no-conflict ones with blue. A number of interesting observations are labeled in the plot. The cluster of blue dots above and to the left of the diagonal are no-conflict observations that have been ranked much lower in terms of conflict probability by the more extensive model. They refer to a number of small, relatively well-educated, peaceful countries (e.g., Swaziland and Botswana). The cluster of blue dots beneath and to the right of the diagonal are another set of peaceful observations that the model with demographic characteristics ranks higher than the history-only model. All these refer to Tanzania, a very populous but poor country that has been remarkably peaceful.

The cluster of red dots beneath and to the right of the diagonal are conflict observations that the extensive model yielded a higher predicted probability for, contributing to an improved performance. These cases include Tchad in 2015 and Niger in 2015–16, both medium-size countries with low education rates. Finally, the red dots to the left and above the diagonal are conflict observations for which the extensive model yielded a lower predicted probability. They include Tunisia, which has high education rates relative to the African average.

Figure 7b demonstrates the performance of the ensemble model compared to the model including the conflict history and demography themes. The horizontal separation plot shows that the ensemble is much better at sorting out the conflict cases. Adding information from all the other themes in particular improves predictions of the conflicts in Ethiopia, Tchad, and Tunisia, as well as the absence of violence in Benin, Burkina Faso, and Zambia. Some of the conflict cases that the history + demography model did well on

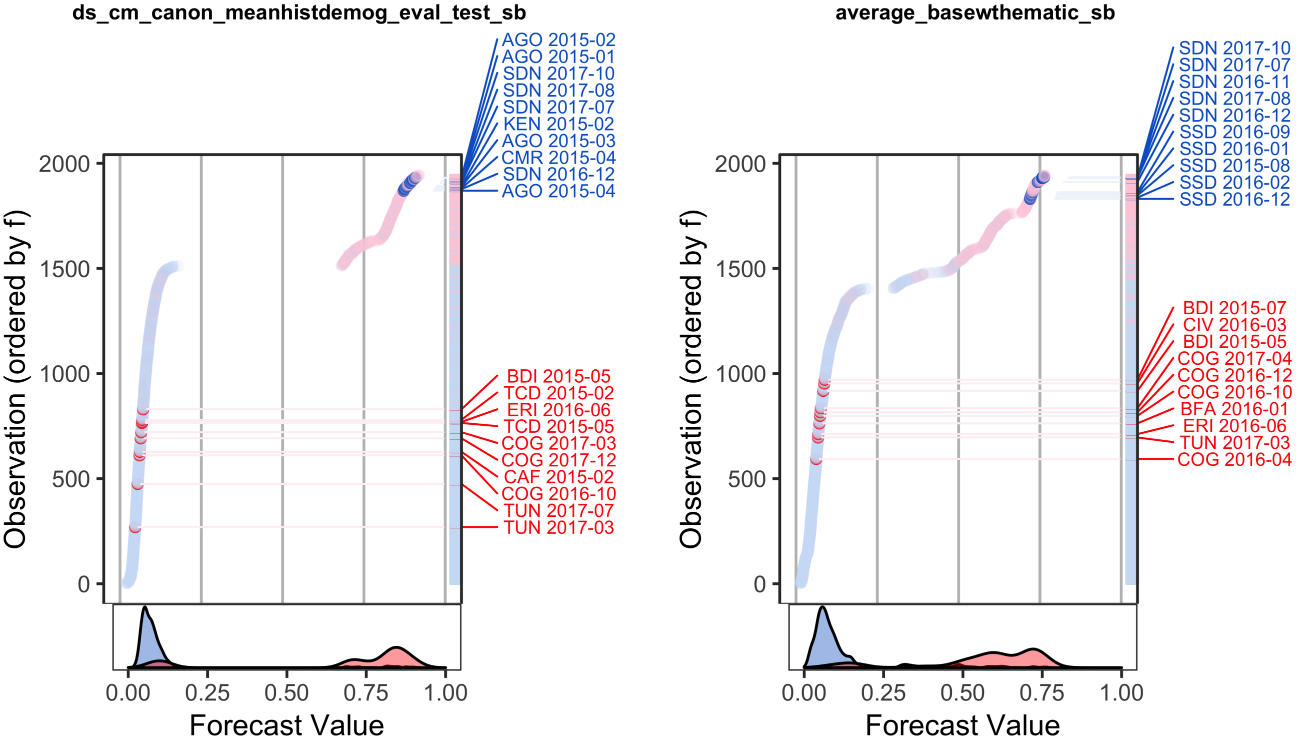


Figure 8. Model criticism plots

(e.g., in Angola) are less well pointed out by the ensemble, though. Similarly, the ensemble model suggests a high likelihood of conflict in Zimbabwe, but this conflict did not occur in the evaluation period.

Figure 8 show model criticism plots (also developed by Colaresi & Mahmood, 2017) for the demography and ensemble models at the cm level. They show that the ensemble is better at identifying positive cases in the test partition – none of the actual conflict months are ranked in the bottom third of predicted probabilities, a clear improvement relative to the demography model. At the same time, the distribution of predicted probabilities are less sharp for the ensemble model.

Figure 9 shows two more bi-separation plots that complement Figure 7. The left-most figure shows changes in ranking of observations when adding the economics theme to the baseline + history only model. The economics predictors increase the predicted probabilities of

Confusion matrices, *cm* level

Confusion matrices show predicted values of conflict as opposed to observed values (actuals) for the models described in Tables 5 and 6 of the article. Since these matrices present binary predictions (positives and negatives), rather than probabilities, a threshold (`cutoff`) needs to be chosen. A 'threshold' is defined as a probability $p*$ over which the model yields a positive prediction. A threshold of 0.5, for instance, means we predict a positive if $\hat{p} > 0.5$ and a negative if not.

	Observed		
Predicted	Pos	Neg	Sum
Pos	True Positive (TP)	False Positive (FP)	TP + FP
Neg	False Negative (FN)	True Negative (TN)	FN + TN
Sum	TP + FN	FP + TN	Total

Table 3. Confusion matrix definitions

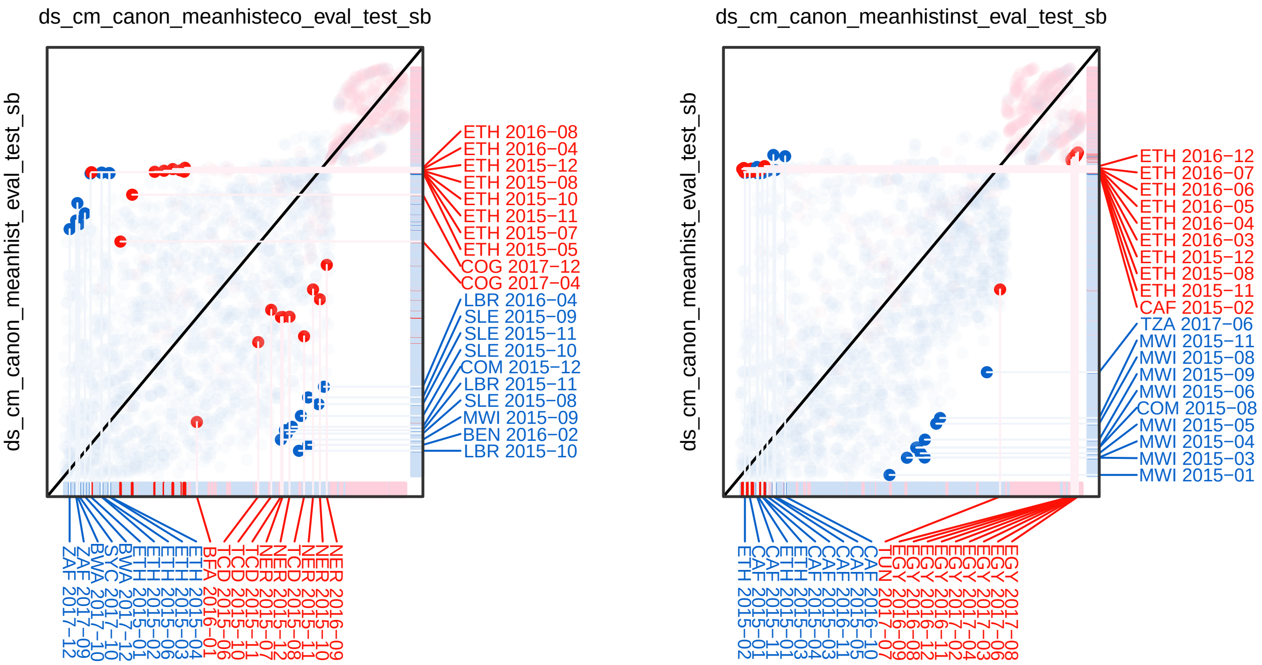


Figure 9. Bi-separation plots: Contributions from economics and institutional themes

This cutoff was chosen by minimizing the misclassification costs in the calibration period of the models we evaluate. We assign a cost of a false-negative to 10 times the cost of a false-positive, as we consider that failing to predict conflict is more problematic than failing to predict non-conflict for both methodological reasons (as the classes are highly imbalanced) and practical reasons.

Confusion matrices contain the information described in Table 3 (Géron, 2017, pp. 86–95).

Tables 4–10 show confusion matrices for all the models in Table 5 in the main article.

	Observed		
Predicted	Pos	Neg	Sum
Pos	397	672	1069
Neg	57	818	875
Sum	454	1490	1944

Note. State-based conflict at cm-level, January 2015 to December 2017. Accuracy = 0.625, F1 = 0.521, precision = 0.371, recall = 0.874, threshold = 0.132.

Table 4. Baseline

G.2 AUROC, Brier, and AUPR, *pgm* level

Table 11 shows the main evaluation metrics for all the constituent models in the *pgm*-level ensemble, as well as for the ensemble itself. Models are named according the following convention:

- 'ds' and 'osa' refers to the way the model handles dynamics (see methodology section in main paper).
- 'pgm' refers to the level of analysis.
- 'acled' or 'canon' denotes whether the models are trained on the period for which we have ACLED data (1997–) or the complete dataset (1990–) for all models except the thematic ones. The thematic

	Observed		
Predicted	Pos	Neg	Sum
Pos	411	245	656
Neg	43	1245	1288
Sum	454	1490	1944

Note. State-based conflict at cm-level, January 2015 to December 2017. Accuracy = 0.852, F1 = 0.741, precision = 0.627, recall = 0.905, threshold = 0.055.

Table 5. Baseline + conflict history theme

	Observed		
Predicted	Pos	Neg	Sum
Pos	399	191	590
Neg	55	1299	1354
Sum	454	1490	1944

Note. State-based conflict at cm-level, January 2015 to December 2017. Accuracy = 0.873, F1 = 0.764, precision = 0.676, recall = 0.879, threshold = 0.083.

Table 6. Baseline + history + economics themes

models have acled in the name even though they are estimated on the complete dataset. This strictly a labelling issues due to a technicality in the database processing.

- 'nocm', 'wcm', 'cl', 'mean', 'soc', 'nat', 'histonly', 'cm', 'protest' refers to a theme as listed in Table 3 in the main paper. 'nocm' refers to all themes without cm level predictors, 'wcm' and 'cl' refer to all themes including cm level predictors either in the pgm model specification or multiplied probabilities as explained in the methodology section of the main article, 'mean' to the baseline, 'soc' to social geography theme, 'nat' to natural geography theme, 'histonly' to conflict history theme, 'cm' to cm theme, and 'protest' to protest theme.

- 'logit_fullsample' and 'rf_downsampled' refers to the logit and random forest specifications of 'osa'.

- 'sb' refers to conflict type.

- The final line, 'average_allwthematic' is the ensemble of all the other models.

Similar to the cm level evaluation, Table 11 shows that there are big differences between the models at the pgm level. As expected, the theme models perform poorer than the models with all predictors included. Interestingly, the sharpness of predictions is very similar across models here, as compared to the cm level evaluation.

Confusion matrices, *pgm* level

Tables 12–16 show the confusion matrices for the models we evaluate at the *pgm* level (Table 6 in main article).

	Observed		
Predicted	Pos	Neg	Sum
Pos	426	354	780
Neg	28	1136	1164
Sum	454	1490	1944

Note. State-based conflict at cm-level, January 2015 to December 2017. Accuracy = 0.803, F1 = 0.69, precision = 0.546, recall = 0.938, threshold = 0.09.

Table 7. Baseline + history + demography themes

	Observed		
Predicted	Pos	Neg	Sum
Pos	374	58	432
Neg	80	1432	1512
Sum	454	1490	1944

Note. State-based conflict at cm-level, January 2015 to December 2017. Accuracy = 0.929, F1 = 0.844, precision = 0.866, recall = 0.824, threshold = 0.724.

Table 8. Baseline + history + institutions themes

	Observed		
Predicted	Pos	Neg	Sum
Pos	436	512	948
Neg	18	978	996
Sum	454	1490	1944

Note. State-based conflict at cm-level, January 2015 to December 2017. Accuracy = 0.727, F1 = 0.622, precision = 0.46, recall = 0.96, threshold = 0.089.

Table 9. All themes

	Observed		
Predicted	Pos	Neg	Sum
Pos	437	282	719
Neg	17	1208	1225
Sum	454	1490	1944

Note. State-based conflict at cm-level, January 2015 to December 2017. Accuracy = 0.846, F1 = 0.745, precision = 0.608, recall = 0.963, threshold = 0.126.

Table 10. Ensemble

	ROC_AUC	Brier_Score	PR_AUC
ds_pgm_canon_nocm_eval_test_sb	0.91639	0.00670	0.23576
ds_pgm_canon_wcm_eval_test_sb	0.90922	0.00646	0.25480
ds_pgm_acled_nocm_eval_test_sb	0.92204	0.00638	0.24544
cl_ds_pgm_canon_nocm_eval_test_sb	0.92663	0.00640	0.21426
ds_pgm_acled_soc_eval_test_sb	0.75885	0.00657	0.01878
ds_pgm_acled_nat_eval_test_sb	0.74278	0.00657	0.01383
ds_pgm_acled_mean_eval_test_sb	0.63240	0.00657	0.04874
ds_pgm_acled_histonly_eval_test_sb	0.89488	0.00685	0.21931
ds_pgm_acled_protest_eval_test_sb	0.66140	0.00657	0.01248
ds_pgm_acled_cm_eval_test_sb	0.59921	0.00657	0.00585
osa_pgm_acled_nocm_eval_test_logit_fullsample_sb	0.93033	0.00612	0.25824
osa_pgm_canon_nocm_eval_test_logit_fullsample_sb	0.92955	0.00610	0.26148
osa_pgm_canon_wcm_eval_test_logit_fullsample_sb	0.92510	0.00617	0.24607
cl_osa_pgm_canon_nocm_eval_test_logit_fullsample_sb	0.93727	0.00612	0.25479
osa_pgm_acled_soc_eval_test_logit_fullsample_sb	0.76909	0.00657	0.01721
osa_pgm_acled_nat_eval_test_logit_fullsample_sb	0.75525	0.00657	0.01318
osa_pgm_acled_mean_eval_test_logit_fullsample_sb	0.86800	0.00664	0.07008
osa_pgm_acled_histonly_eval_test_logit_fullsample_sb	0.92198	0.00616	0.24726
osa_pgm_acled_protest_eval_test_logit_fullsample_sb	0.70775	0.00657	0.01588
osa_pgm_acled_cm_eval_test_logit_fullsample_sb	0.55844	0.00657	0.00501
osa_pgm_acled_nocm_eval_test_rf_downsampled_sb	0.94519	0.00623	0.23495
osa_pgm_canon_nocm_eval_test_rf_downsampled_sb	0.94481	0.00623	0.22896
osa_pgm_canon_wcm_eval_test_rf_downsampled_sb	0.94781	0.00618	0.24163
cl_osa_pgm_canon_nocm_eval_test_rf_downsampled_sb	0.94615	0.00624	0.22618
osa_pgm_acled_soc_eval_test_rf_downsampled_sb	0.91338	0.00646	0.17248
osa_pgm_acled_nat_eval_test_rf_downsampled_sb	0.90216	0.00654	0.15135
osa_pgm_acled_mean_eval_test_rf_downsampled_sb	0.88153	0.00657	0.07407
osa_pgm_acled_histonly_eval_test_rf_downsampled_sb	0.93301	0.00626	0.23359
osa_pgm_acled_protest_eval_test_rf_downsampled_sb	0.67398	0.00657	0.01360
osa_pgm_acled_cm_eval_test_rf_downsampled_sb	0.78972	0.00657	0.02489
average_allwthematic_sb	0.94839	0.00623	0.27689

Table 11. SB constituent models and ensembles, pgm level, 36 months

	Observed		
Predicted	Pos	Neg	Sum
Pos	234	1046	1280
Neg	1426	381558	382984
Sum	1660	382604	384264

Note. State-based conflict at pgm-level, January 2015 to December 2017. Accuracy = 0.994, F1 = 0.159, precision = 0.183, recall = 0.141, threshold = 0.017.

Table 12. Baseline

	Observed		
Predicted	Pos	Neg	Sum
Pos	797	3623	4420
Neg	863	378981	379844
Sum	1660	382604	384264

Note. State-based conflict at pgm-level, January 2015 to December 2017. Accuracy = 0.988, F1 = 0.262, precision = 0.18, recall = 0.48, threshold = 0.066.

Table 13. Baseline + history themes

	Observed		
Predicted	Pos	Neg	Sum
Pos	605	2085	2690
Neg	1055	380519	381574
Sum	1660	382604	384264

Note. State-based conflict at pgm-level, January 2015 to December 2017. Accuracy = 0.992, F1 = 0.278, precision = 0.225, recall = 0.364, threshold = 0.141.

Table 14. Baseline + history + social and natural geography themes

	Observed		
Predicted	Pos	Neg	Sum
Pos	639	2726	3365
Neg	1021	379878	380899
Sum	1660	382604	384264

Note. State-based conflict at pgm-level, January 2015 to December 2017. Accuracy = 0.99, F1 = 0.254, precision = 0.19, recall = 0.385, threshold = 0.11.

Table 15. All themes

	Observed		
Predicted	Pos	Neg	Sum
Pos	735	2706	3441
Neg	925	379898	380823
Sum	1660	382604	384264

Note. State-based conflict at pgm-level, January 2015 to December 2017. Accuracy = 0.991, F1 = 0.289, precision = 0.214, recall = 0.443, threshold = 0.064.

Table 16. Ensemble

H Evaluation, ns

In this section, we present evaluation and forecast results for **ns**. Tables 17, 18 19, 20, and Figure 10 show the equivalent evaluation for **ns** as we have shown for **sb** in the main article and in section G. For details on the naming conventions in the tables, please consult section G.

	ROC_AUC	Brier_Score	PR_AUC
ds_pgm_acled_mean_eval_test_ns	0.54452	0.00409	0.01110
ds_pgm_acled_meanhist_eval_test_ns	0.85411	0.00415	0.04193
ds_pgm_acled_meansocnathist_eval_test_ns	0.80084	0.00409	0.02554
ds_pgm_acled_meansocnathistcm_eval_test_ns	0.79150	0.00409	0.02538
average_allwthematic_ns	0.90395	0.00409	0.05515

Table 17. NS constituent models and ensembles, pgm level, 36 months

	ROC_AUC	Brier_Score	PR_AUC
ds_pgm_canon_nocm_eval_test_ns	0.76993	0.00409	0.03517
ds_pgm_canon_wcm_eval_test_ns	0.76368	0.00409	0.03479
ds_pgm_acled_nocm_eval_test_ns	0.78320	0.00409	0.03032
cl_ds_pgm_canon_nocm_eval_test_ns	0.81765	0.00409	0.04289
ds_pgm_acled_soc_eval_test_ns	0.71915	0.00409	0.00949
ds_pgm_acled_nat_eval_test_ns	0.69538	0.00409	0.00384
ds_pgm_acled_mean_eval_test_ns	0.54452	0.00409	0.01110
ds_pgm_acled_histonly_eval_test_ns	0.84234	0.00409	0.04131
ds_pgm_acled_protest_eval_test_ns	0.70320	0.00409	0.00782
ds_pgm_acled_cm_eval_test_ns	0.64237	0.00409	0.00535
osa_pgm_acled_nocm_eval_test_logit_fullsample_ns	0.82539	0.00409	0.04255
osa_pgm_canon_nocm_eval_test_logit_fullsample_ns	0.81525	0.00409	0.04097
osa_pgm_canon_wcm_eval_test_logit_fullsample_ns	0.78580	0.00409	0.03903
cl_osa_pgm_canon_nocm_eval_test_logit_fullsample_ns	0.88727	0.00409	0.04275
osa_pgm_acled_soc_eval_test_logit_fullsample_ns	0.73401	0.00409	0.01035
osa_pgm_acled_nat_eval_test_logit_fullsample_ns	0.71350	0.00409	0.00411
osa_pgm_acled_mean_eval_test_logit_fullsample_ns	0.82988	0.00410	0.02091
osa_pgm_acled_histonly_eval_test_logit_fullsample_ns	0.89896	0.00417	0.04650
osa_pgm_acled_protest_eval_test_logit_fullsample_ns	0.78732	0.00409	0.01246
osa_pgm_acled_cm_eval_test_logit_fullsample_ns	0.64649	0.00409	0.00449
osa_pgm_acled_nocm_eval_test_rf_downsampled_ns	0.91044	0.00409	0.05640
osa_pgm_canon_nocm_eval_test_rf_downsampled_ns	0.90585	0.00409	0.06643
osa_pgm_canon_wcm_eval_test_rf_downsampled_ns	0.92898	0.00408	0.06443
cl_osa_pgm_canon_nocm_eval_test_rf_downsampled_ns	0.92793	0.00409	0.06149
osa_pgm_acled_soc_eval_test_rf_downsampled_ns	0.84842	0.00409	0.05216
osa_pgm_acled_nat_eval_test_rf_downsampled_ns	0.83972	0.00409	0.05413
osa_pgm_acled_mean_eval_test_rf_downsampled_ns	0.79747	0.00409	0.02867
osa_pgm_acled_histonly_eval_test_rf_downsampled_ns	0.89505	0.00409	0.04890
osa_pgm_acled_protest_eval_test_rf_downsampled_ns	0.69820	0.00409	0.00791
osa_pgm_acled_cm_eval_test_rf_downsampled_ns	0.83200	0.00409	0.00982
average_allwthematic_ns	0.90395	0.00409	0.05515

Table 18. NS constituent models and ensembles, pgm level, 36 months

	ROC_AUC	Brier_Score	PR_AUC
ds_cm_canon_mean_eval_test_ns	0.79469	0.08691	0.54263
ds_cm_canon_meanhist_eval_test_ns	0.96404	0.05837	0.77836
ds_cm_canon_meanhisteco_eval_test_ns	0.82576	0.07331	0.59396
ds_cm_canon_meanhistdemog_eval_test_ns	0.89656	0.07837	0.68546
ds_cm_canon_meanhistinst_eval_test_ns	0.81799	0.06981	0.62667
ds_cm_acled_mndmgecohstnst_eval_test_ns	0.87963	0.07915	0.57454
average_basewthematic_ns	0.94837	0.07318	0.72198

Table 19. NS constituent models and ensembles, cm level, 36 months

	ROC_AUC	Brier_Score	PR_AUC
ds_cm_acled_base_eval_test_ns	0.94707	0.06547	0.72842
ds_cm_canon_base_eval_test_ns	0.95328	0.06335	0.74951
ds_cm_canon_mean_eval_test_ns	0.79469	0.08691	0.54263
ds_cm_canon_demog_eval_test_ns	0.70675	0.09694	0.40876
ds_cm_canon_eco_eval_test_ns	0.70228	0.11048	0.23194
ds_cm_canon_hist_eval_test_ns	0.96519	0.06132	0.78931
ds_cm_canon_inst_eval_test_ns	0.39043	0.11780	0.12344
ds_cm_canon_protest_eval_test_ns	0.73217	0.10393	0.21929
osa_cm_acled_base_eval_test_logit_fullsample_ns	0.93679	0.06975	0.64575
osa_cm_canon_base_eval_test_logit_fullsample_ns	0.94566	0.06629	0.68087
osa_cm_canon_mean_eval_test_logit_fullsample_ns	0.87932	0.08863	0.43815
osa_cm_canon_eco_eval_test_logit_fullsample_ns	0.75232	0.10738	0.32756
osa_cm_canon_hist_eval_test_logit_fullsample_ns	0.95552	0.05608	0.68121
osa_cm_canon_inst_eval_test_logit_fullsample_ns	0.60653	0.10899	0.19035
osa_cm_canon_protest_eval_test_logit_fullsample_ns	0.76011	0.10359	0.25260
osa_cm_acled_base_eval_test_rf_downsampled_ns	0.94828	0.06033	0.68120
osa_cm_canon_base_eval_test_rf_downsampled_ns	0.94254	0.06533	0.63387
osa_cm_canon_demog_eval_test_logit_fullsample_ns	0.71601	0.09895	0.40929
osa_cm_canon_mean_eval_test_rf_downsampled_ns	0.88006	0.09173	0.46776
osa_cm_canon_demog_eval_test_rf_downsampled_ns	0.87773	0.08971	0.50941
osa_cm_canon_eco_eval_test_rf_downsampled_ns	0.63579	0.10680	0.23110
osa_cm_canon_hist_eval_test_rf_downsampled_ns	0.93470	0.06095	0.66823
osa_cm_canon_inst_eval_test_rf_downsampled_ns	0.75085	0.09821	0.40758
osa_cm_canon_protest_eval_test_rf_downsampled_ns	0.74649	0.10442	0.22113
average_basewthematic_sb	0.91552	0.08258	0.52357

Table 20. NS constituent models and ensembles, cm level, 36 months

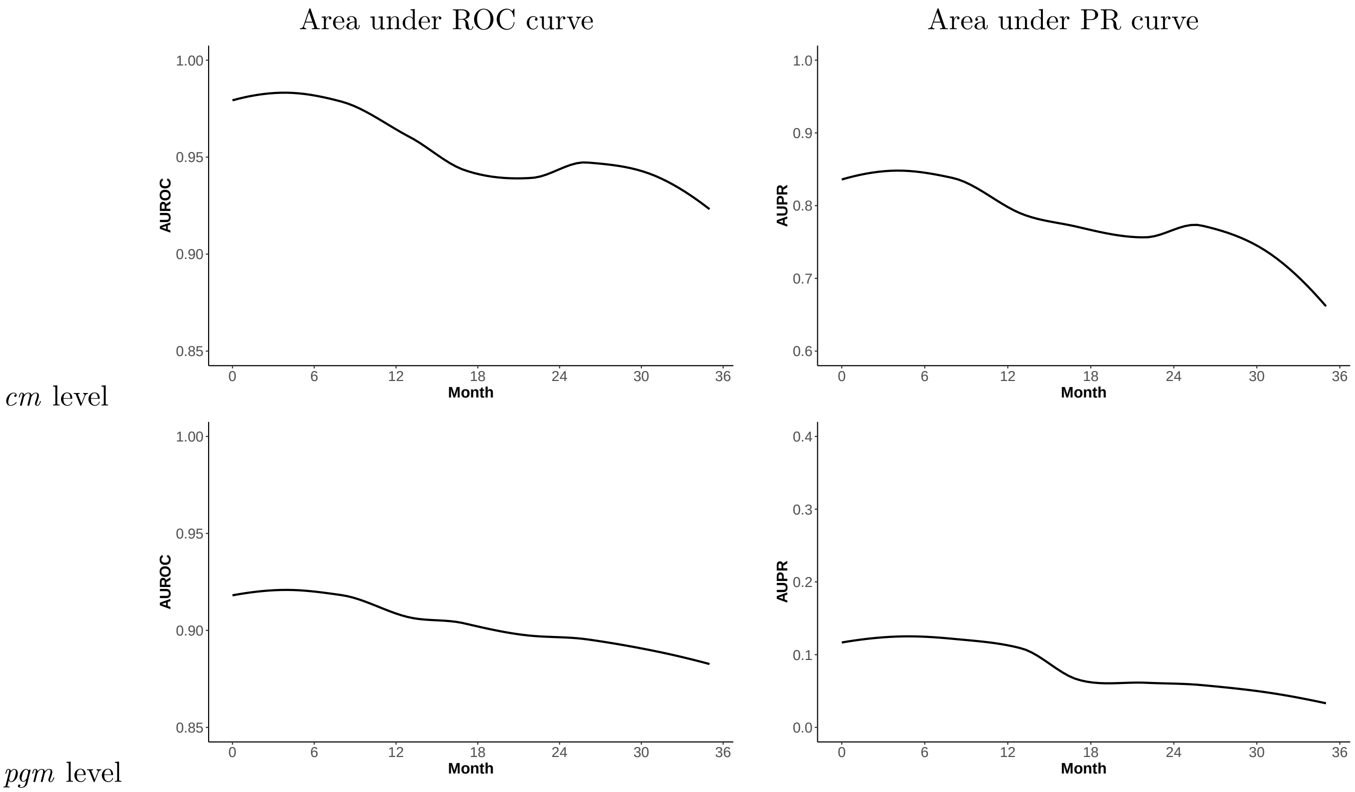


Figure 10. Performance over time, **ns**, *cm* (top) and *pgm* (bottom). AUROC (left) and AUPR (right), by month in forecasting window. The lines are smoothed with a loess function. Note that the y-axis differs in each plot.

I Evaluation, os

In this section, we present evaluation and forecast results for **os**. Tables 21, 22, 23, 24, and Figure 11 show the equivalent evaluation for **os** as we have shown for **sb** in the main article and in section G. For details on the naming conventions in the tables, please consult this section.

	ROC_AUC	Brier_Score	PR_AUC
ds_pgm_acled_mean_eval_test_os	0.61381	0.00527	0.04988
ds_pgm_acled_meanhist_eval_test_os	0.87787	0.00572	0.15748
ds_pgm_acled_meansocnathist_eval_test_os	0.90279	0.00578	0.14525
ds_pgm_acled_meansocnathistcm_eval_test_os	0.90997	0.00542	0.15596
average_allwthematic_os	0.94798	0.00513	0.20727

Table 21. OS constituent models and ensembles, pgm level, 36 months

	ROC_AUC	Brier_Score	PR_AUC
ds_pgm_canon_nocm_eval_test_os	0.88843	0.00554	0.16020
ds_pgm_canon_wcm_eval_test_os	0.90436	0.00552	0.15702
ds_pgm_acled_nocm_eval_test_os	0.90142	0.00523	0.16867
cl_ds_pgm_canon_nocm_eval_test_os	0.90993	0.00527	0.15201
ds_pgm_acled_soc_eval_test_os	0.80585	0.00527	0.01504
ds_pgm_acled_nat_eval_test_os	0.77734	0.00527	0.01245
ds_pgm_acled_mean_eval_test_os	0.61381	0.00527	0.04988
ds_pgm_acled_histonly_eval_test_os	0.87696	0.00547	0.15438
ds_pgm_acled_protest_eval_test_os	0.69981	0.00527	0.01166
ds_pgm_acled_cm_eval_test_os	0.72482	0.00527	0.00606
osa_pgm_acled_nocm_eval_test_logit_fullsample_os	0.92599	0.00509	0.17936
osa_pgm_canon_nocm_eval_test_logit_fullsample_os	0.91357	0.00506	0.18852
osa_pgm_canon_wcm_eval_test_logit_fullsample_os	0.92163	0.00506	0.18571
cl_osa_pgm_canon_nocm_eval_test_logit_fullsample_os	0.92489	0.00507	0.17957
osa_pgm_acled_soc_eval_test_logit_fullsample_os	0.81256	0.00527	0.01463
osa_pgm_acled_nat_eval_test_logit_fullsample_os	0.79434	0.00527	0.01259
osa_pgm_acled_mean_eval_test_logit_fullsample_os	0.83820	0.00525	0.07241
osa_pgm_acled_histonly_eval_test_logit_fullsample_os	0.91773	0.00520	0.17616
osa_pgm_acled_protest_eval_test_logit_fullsample_os	0.68571	0.00527	0.01409
osa_pgm_acled_cm_eval_test_logit_fullsample_os	0.73657	0.00527	0.00582
osa_pgm_acled_nocm_eval_test_rf_downsampled_os	0.94177	0.00515	0.18749
osa_pgm_canon_nocm_eval_test_rf_downsampled_os	0.93976	0.00517	0.18518
osa_pgm_canon_wcm_eval_test_rf_downsampled_os	0.94159	0.00511	0.17709
cl_osa_pgm_canon_nocm_eval_test_rf_downsampled_os	0.94163	0.00515	0.17067
osa_pgm_acled_soc_eval_test_rf_downsampled_os	0.90169	0.00527	0.13447
osa_pgm_acled_nat_eval_test_rf_downsampled_os	0.87855	0.00527	0.10655
osa_pgm_acled_mean_eval_test_rf_downsampled_os	0.84481	0.00527	0.08377
osa_pgm_acled_histonly_eval_test_rf_downsampled_os	0.91981	0.00527	0.16809
osa_pgm_acled_protest_eval_test_rf_downsampled_os	0.66915	0.00527	0.00990
osa_pgm_acled_cm_eval_test_rf_downsampled_os	0.80703	0.00527	0.01583
average_allwthematic_os	0.94798	0.00513	0.20727

Table 22. OS constituent models and ensembles, pgm level, 36 months

	ROC_AUC	Brier_Score	PR_AUC
ds_cm_canon_mean_eval_test_os	0.75262	0.13773	0.55828
ds_cm_canon_meanhist_eval_test_os	0.91740	0.07919	0.77562
ds_cm_canon_meanhisteco_eval_test_os	0.87062	0.08672	0.72667
ds_cm_canon_meanhistdemog_eval_test_os	0.87695	0.08902	0.64605
ds_cm_canon_meanhistinst_eval_test_os	0.83577	0.09006	0.64326
ds_cm_acled_mndmgecohstnst_eval_test_os	0.90402	0.08603	0.70878
average_basewthematic_os	0.93448	0.10264	0.80914

Table 23. OS constituent models and ensembles, cm level, 36 months

	ROC_AUC	Brier_Score	PR_AUC
ds_cm_acled_base_eval_test_os	0.92992	0.07827	0.80586
ds_cm_canon_base_eval_test_os	0.92682	0.08238	0.79209
ds_cm_canon_mean_eval_test_os	0.75262	0.13773	0.55828
ds_cm_canon_demog_eval_test_os	0.74827	0.14640	0.37405
ds_cm_canon_eco_eval_test_os	0.76830	0.14212	0.42117
ds_cm_canon_hist_eval_test_os	0.92318	0.07741	0.75591
ds_cm_canon_inst_eval_test_os	0.63214	0.15888	0.30793
ds_cm_canon_protest_eval_test_os	0.70661	0.14508	0.30143
osa_cm_acled_base_eval_test_logit_fullsample_os	0.92139	0.07868	0.75031
osa_cm_canon_base_eval_test_logit_fullsample_os	0.92389	0.07903	0.79501
osa_cm_canon_mean_eval_test_logit_fullsample_os	0.81506	0.12884	0.53952
osa_cm_canon_eco_eval_test_logit_fullsample_os	0.73579	0.14509	0.44790
osa_cm_canon_hist_eval_test_logit_fullsample_os	0.91101	0.07905	0.79067
osa_cm_canon_inst_eval_test_logit_fullsample_os	0.61018	0.15861	0.34459
osa_cm_canon_protest_eval_test_logit_fullsample_os	0.68593	0.14667	0.29725
osa_cm_acled_base_eval_test_rf_downsampled_os	0.91851	0.08234	0.77968
osa_cm_canon_base_eval_test_rf_downsampled_os	0.91659	0.08405	0.77511
osa_cm_canon_demog_eval_test_logit_fullsample_os	0.76597	0.14606	0.38501
osa_cm_canon_mean_eval_test_rf_downsampled_os	0.83186	0.13876	0.57761
osa_cm_canon_demog_eval_test_rf_downsampled_os	0.79102	0.13662	0.50908
osa_cm_canon_eco_eval_test_rf_downsampled_os	0.76807	0.14320	0.43615
osa_cm_canon_hist_eval_test_rf_downsampled_os	0.90722	0.08258	0.76523
osa_cm_canon_inst_eval_test_rf_downsampled_os	0.74552	0.14753	0.50227
osa_cm_canon_protest_eval_test_rf_downsampled_os	0.61866	0.15853	0.22503
average_basewthematic_sb	0.91384	0.10069	0.69488

Table 24. OS constituent models and ensembles, cm level, 36 months

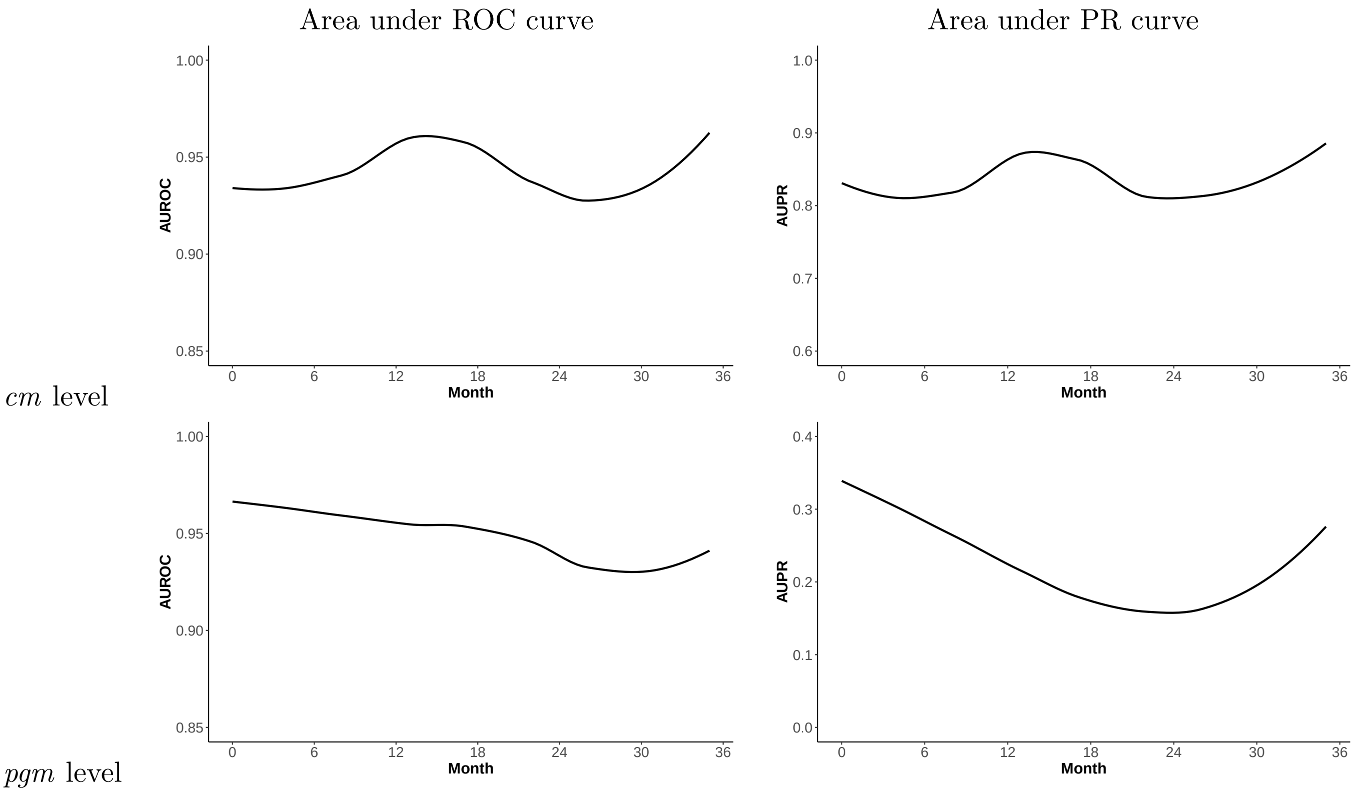
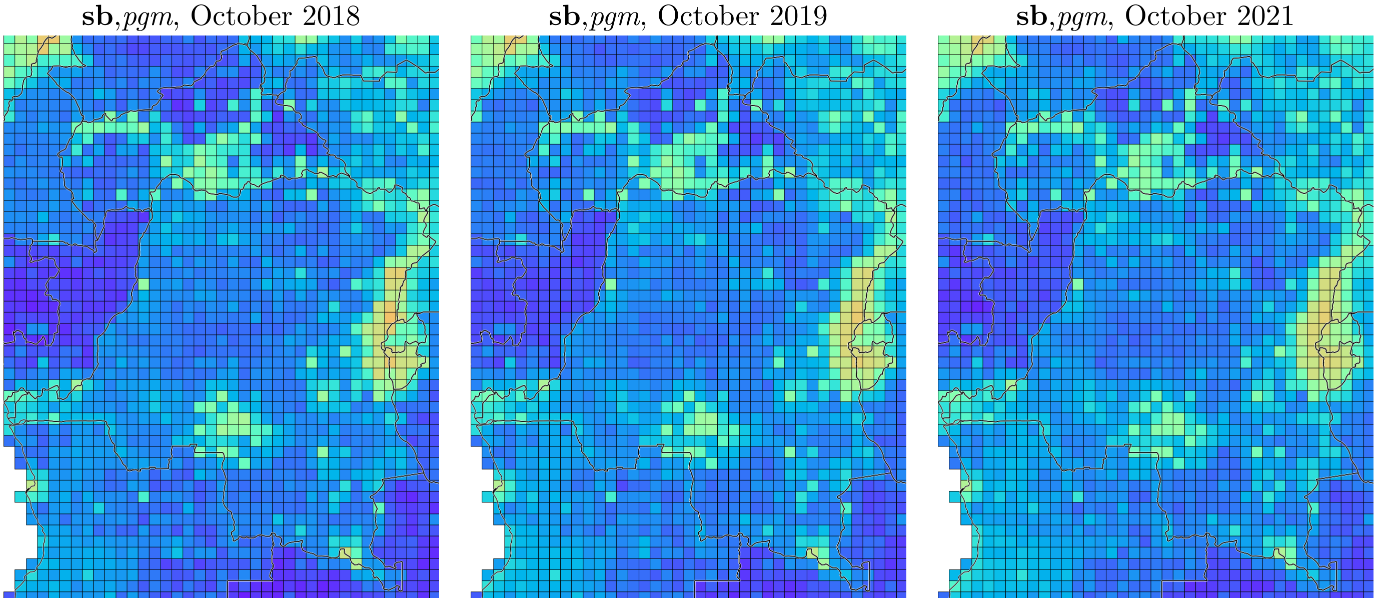


Figure 11. Performance over time, **os**, *cm* (top) and *pgm* (bottom). AUROC (left) and AUPR (right), by month in forecasting window. The lines are smoothed with a loess function. Note that the y-axis differs in each plot.

J Additional forecasting results

Figure 12 takes a closer look at Western Central Africa over time. There is a sizeable orange cluster, with probabilities around 30–50% in each *pgm* in the Eastern DRC and Burundi for June 2018. This cluster is slowly expanding over the forecasting period, and spills into Rwanda. As clear from Figure 8 in the main article, forecasted conflict patterns are very stable.



sb,*pgm*, October 2018

sb,*pgm*, October 2019

sb,*pgm*, October 2021

Figure 12. Forecasts October 2018 – August 2021, state-based conflict, DRC, CAR, Rwanda, Burundi

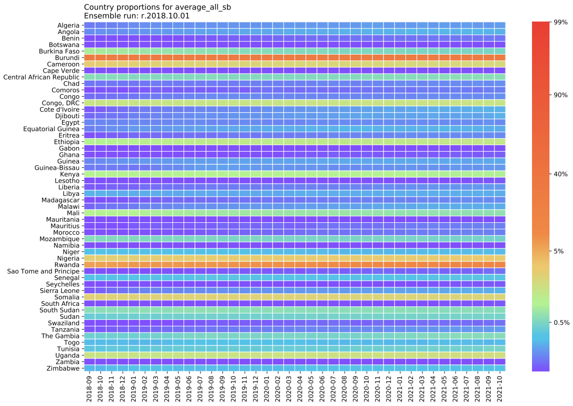


Figure 13. Predicted proportion of PRIO-GRID cells with conflict, by country and month, September 2018–October 2021, **sb**

Country name	2018-10	2019-10	2020-10	2021-10
Algeria	0.0016	0.0019	0.0021	0.0021
Angola	0.0018	0.0022	0.0025	0.0025
Benin	0.0009	0.0012	0.0013	0.0015
Botswana	0.0003	0.0003	0.0004	0.0004
Burkina Faso	0.0080	0.0058	0.0050	0.0048
Burundi	0.2507	0.2680	0.2757	0.2691
Cameroon	0.0257	0.0218	0.0203	0.0193
Cape Verde	0.0005	0.0006	0.0007	0.0008
Central African Republic	0.0058	0.0054	0.0055	0.0059
Chad	0.0015	0.0015	0.0017	0.0018
Comoros	0.0010	0.0013	0.0013	0.0015
Congo	0.0014	0.0017	0.0017	0.0018
Congo, DRC	0.0141	0.0145	0.0147	0.0145
Cote d'Ivoire	0.0012	0.0017	0.0021	0.0026
Djibouti	0.0014	0.0018	0.0022	0.0025
Egypt	0.0017	0.0017	0.0017	0.0018
Equatorial Guinea	0.0018	0.0024	0.0025	0.0025
Eritrea	0.0011	0.0015	0.0016	0.0017
Ethiopia	0.0098	0.0122	0.0127	0.0133
Gabon	0.0004	0.0005	0.0006	0.0006
Ghana	0.0007	0.0010	0.0011	0.0012
Guinea	0.0018	0.0022	0.0023	0.0023
Guinea-Bissau	0.0015	0.0020	0.0018	0.0019
Kenya	0.0095	0.0097	0.0091	0.0089
Lesotho	0.0006	0.0008	0.0009	0.0010
Liberia	0.0011	0.0015	0.0016	0.0020
Libya	0.0024	0.0023	0.0024	0.0024
Madagascar	0.0010	0.0014	0.0016	0.0017
Malawi	0.0013	0.0019	0.0022	0.0024
Mali	0.0100	0.0074	0.0068	0.0062
Mauritania	0.0004	0.0005	0.0006	0.0006
Mauritius	0.0009	0.0012	0.0013	0.0016
Morocco	0.0008	0.0012	0.0014	0.0016
Mozambique	0.0055	0.0050	0.0048	0.0045
Namibia	0.0002	0.0003	0.0003	0.0003
Niger	0.0029	0.0028	0.0027	0.0027
Nigeria	0.0241	0.0264	0.0266	0.0270
Rwanda	0.0623	0.0716	0.0890	0.1115
Sao Tome and Principe	0.0009	0.0011	0.0012	0.0013
Senegal	0.0030	0.0033	0.0033	0.0035
Seychelles	0.0007	0.0008	0.0010	0.0010
Sierra Leone	0.0012	0.0018	0.0021	0.0025
Somalia	0.0226	0.0225	0.0217	0.0216
South Africa	0.0008	0.0009	0.0010	0.0011
South Sudan	0.0057	0.0060	0.0059	0.0056
Sudan	0.0041	0.0045	0.0047	0.0047
Swaziland	0.0009	0.0011	0.0012	0.0014
Tanzania	0.0011	0.0015	0.0017	0.0020
The Gambia	0.0046	0.0052	0.0049	0.0059
Togo	0.0027	0.0030	0.0027	0.0028
Tunisia	0.0037	0.0040	0.0042	0.0049
Uganda	0.0144	0.0147	0.0160	0.0169
Zambia	0.0005	0.0006	0.0006	0.0007
Zimbabwe	0.0027	0.0027	0.0029	0.0030

Table 25. Country proportions (sb)

Figure 14 and Table 26 shows the forecasts in the same form for **ns** conflict, and Figure 15 and Table 27 for one-sided violence.

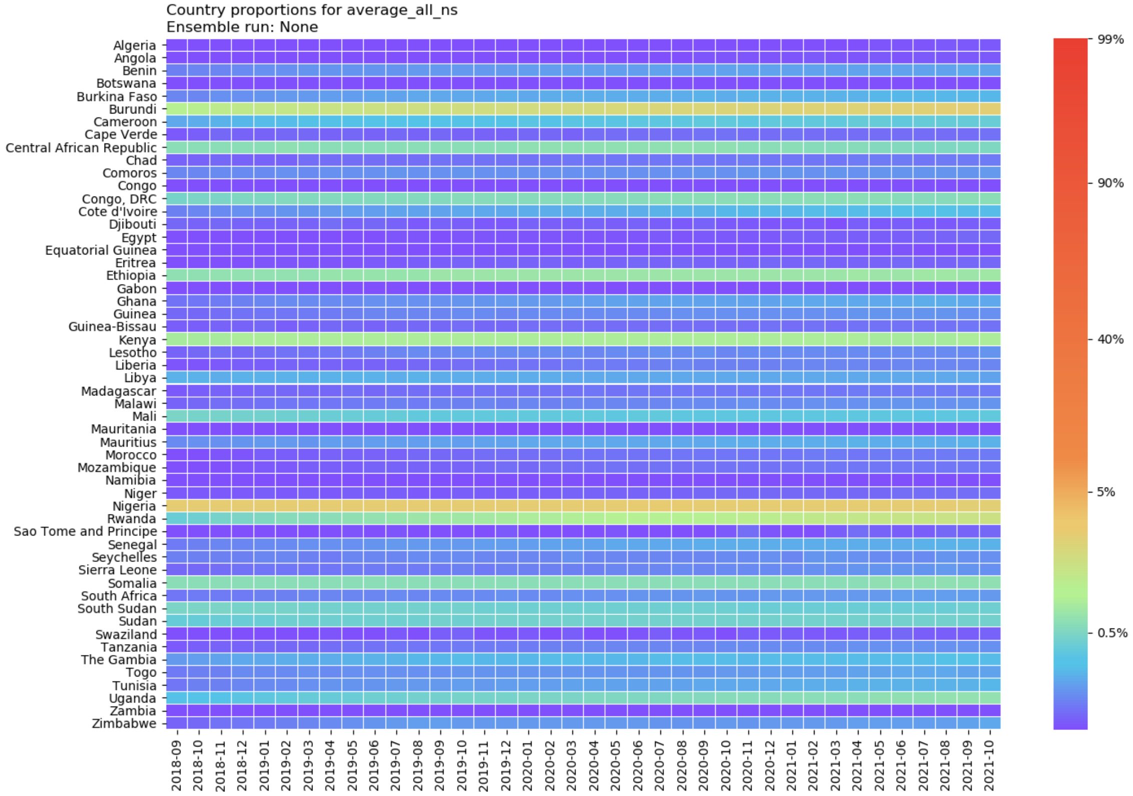


Figure 14. Predicted proportion of PRIO-GRID cells with conflict, by country and month, September 2018–October 2021, **ns**

Figure 13 shows the forecasted share of PRIO-GRID cells in conflict by country and by month for all three outcomes (as in Figure 8 in the main article). Table 25 shows the same information in table form for selected months. The forecasted proportion of grid cells with at least one event of state-based conflict in for instance Burundi is .25 in October 2018 and remains stable until October 2021. The forecasted proportion in Ethiopia starts at 0.010 in October 2018 and increases up to 0.013 at the end of the forecasting window.

Country name	2018-10	2019-10	2020-10	2021-10
Algeria	0.0006	0.0008	0.0010	0.0011
Angola	0.0007	0.0009	0.0011	0.0011
Benin	0.0017	0.0020	0.0022	0.0022
Botswana	0.0007	0.0008	0.0008	0.0008
Burkina Faso	0.0017	0.0023	0.0024	0.0026
Burundi	0.0112	0.0170	0.0209	0.0246
Cameroon	0.0024	0.0032	0.0037	0.0040
Cape Verde	0.0012	0.0012	0.0014	0.0014
Central African Republic	0.0059	0.0060	0.0061	0.0052
Chad	0.0012	0.0014	0.0015	0.0015
Comoros	0.0017	0.0019	0.0018	0.0020
Congo	0.0006	0.0007	0.0007	0.0007
Congo, DRC	0.0049	0.0055	0.0059	0.0060
Cote d'Ivoire	0.0017	0.0022	0.0026	0.0027
Djibouti	0.0012	0.0012	0.0011	0.0012
Egypt	0.0008	0.0011	0.0011	0.0013
Equatorial Guinea	0.0007	0.0009	0.0010	0.0010
Eritrea	0.0010	0.0012	0.0012	0.0013
Ethiopia	0.0063	0.0069	0.0071	0.0072
Gabon	0.0004	0.0004	0.0005	0.0005
Ghana	0.0015	0.0019	0.0022	0.0023
Guinea	0.0013	0.0017	0.0018	0.0019
Guinea-Bissau	0.0012	0.0014	0.0013	0.0014
Kenya	0.0082	0.0085	0.0085	0.0085
Lesotho	0.0013	0.0018	0.0017	0.0019
Liberia	0.0011	0.0013	0.0016	0.0017
Libya	0.0025	0.0024	0.0023	0.0022
Madagascar	0.0012	0.0014	0.0015	0.0015
Malawi	0.0013	0.0016	0.0018	0.0019
Mali	0.0047	0.0038	0.0037	0.0038
Mauritania	0.0005	0.0005	0.0006	0.0007
Mauritius	0.0018	0.0021	0.0023	0.0025
Morocco	0.0010	0.0013	0.0014	0.0015
Mozambique	0.0010	0.0012	0.0014	0.0015
Namibia	0.0006	0.0006	0.0007	0.0007
Niger	0.0011	0.0012	0.0013	0.0013
Nigeria	0.0253	0.0257	0.0258	0.0247
Rwanda	0.0044	0.0075	0.0108	0.0154
Sao Tome and Principe	0.0010	0.0010	0.0011	0.0013
Senegal	0.0016	0.0020	0.0024	0.0025
Seychelles	0.0016	0.0018	0.0018	0.0018
Sierra Leone	0.0013	0.0016	0.0018	0.0019
Somalia	0.0059	0.0059	0.0058	0.0063
South Africa	0.0015	0.0017	0.0019	0.0020
South Sudan	0.0048	0.0044	0.0043	0.0043
Sudan	0.0040	0.0044	0.0045	0.0045
Swaziland	0.0009	0.0012	0.0010	0.0012
Tanzania	0.0012	0.0015	0.0017	0.0018
The Gambia	0.0021	0.0027	0.0027	0.0028
Togo	0.0016	0.0019	0.0020	0.0022
Tunisia	0.0016	0.0021	0.0023	0.0026
Uganda	0.0030	0.0046	0.0055	0.0063
Zambia	0.0007	0.0009	0.0010	0.0010
Zimbabwe	0.0013	0.0020	0.0020	0.0023

Table 26. Country proportions (ns)

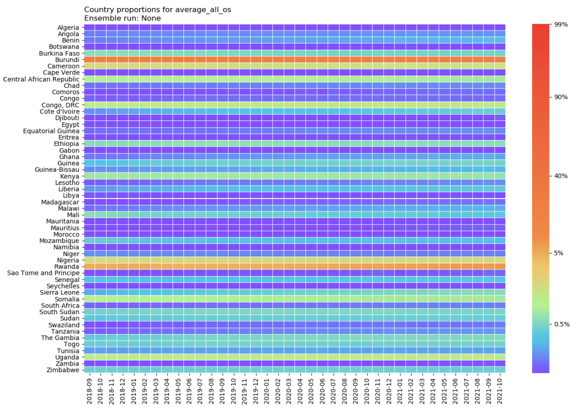


Figure 15. Predicted proportion of PRIO-GRID cells with conflict, by country and month, September 2018–October 2021, **os**

Country name	2018-10	2019-10	2020-10	2021-10
Algeria	0.0006	0.0006	0.0007	0.0007
Angola	0.0016	0.0018	0.0019	0.0019
Benin	0.0016	0.0020	0.0023	0.0027
Botswana	0.0004	0.0004	0.0005	0.0005
Burkina Faso	0.0059	0.0052	0.0047	0.0048
Burundi	0.1906	0.1844	0.1769	0.1659
Cameroon	0.0192	0.0168	0.0152	0.0148
Cape Verde	0.0005	0.0006	0.0007	0.0008
Central African Republic	0.0092	0.0096	0.0087	0.0079
Chad	0.0014	0.0016	0.0017	0.0019
Comoros	0.0008	0.0011	0.0012	0.0014
Congo	0.0012	0.0013	0.0013	0.0013
Congo, DRC	0.0121	0.0125	0.0130	0.0126
Cote d'Ivoire	0.0020	0.0029	0.0034	0.0041
Djibouti	0.0007	0.0009	0.0009	0.0012
Egypt	0.0008	0.0008	0.0009	0.0010
Equatorial Guinea	0.0014	0.0015	0.0017	0.0017
Eritrea	0.0007	0.0009	0.0010	0.0010
Ethiopia	0.0056	0.0062	0.0057	0.0057
Gabon	0.0004	0.0005	0.0005	0.0005
Ghana	0.0015	0.0020	0.0022	0.0023
Guinea	0.0032	0.0043	0.0046	0.0047
Guinea-Bissau	0.0018	0.0023	0.0025	0.0027
Kenya	0.0073	0.0077	0.0071	0.0065
Lesotho	0.0011	0.0016	0.0017	0.0018
Liberia	0.0020	0.0028	0.0031	0.0037
Libya	0.0011	0.0010	0.0010	0.0010
Madagascar	0.0009	0.0011	0.0013	0.0014
Malawi	0.0014	0.0020	0.0023	0.0025
Mali	0.0055	0.0043	0.0038	0.0036
Mauritania	0.0004	0.0004	0.0004	0.0005
Mauritius	0.0007	0.0009	0.0010	0.0012
Morocco	0.0006	0.0007	0.0008	0.0009
Mozambique	0.0040	0.0033	0.0031	0.0030
Namibia	0.0003	0.0003	0.0003	0.0003
Niger	0.0018	0.0017	0.0016	0.0016
Nigeria	0.0179	0.0193	0.0186	0.0189
Rwanda	0.0489	0.0470	0.0568	0.0675
Sao Tome and Principe	0.0008	0.0011	0.0012	0.0012
Senegal	0.0033	0.0033	0.0034	0.0034
Seychelles	0.0006	0.0005	0.0005	0.0007
Sierra Leone	0.0024	0.0039	0.0048	0.0057
Somalia	0.0106	0.0088	0.0081	0.0078
South Africa	0.0012	0.0014	0.0015	0.0016
South Sudan	0.0045	0.0051	0.0052	0.0048
Sudan	0.0037	0.0041	0.0041	0.0041
Swaziland	0.0009	0.0014	0.0014	0.0015
Tanzania	0.0012	0.0016	0.0018	0.0020
The Gambia	0.0040	0.0049	0.0053	0.0053
Togo	0.0041	0.0047	0.0044	0.0045
Tunisia	0.0022	0.0022	0.0022	0.0022
Uganda	0.0115	0.0115	0.0120	0.0123
Zambia	0.0008	0.0009	0.0010	0.0011
Zimbabwe	0.0043	0.0043	0.0043	0.0043

Table 27. Country proportions (os)

K Data management

Data are currently stored in a large, highly normalized (3NF) Postgres database to avoid redundancy and ensure consistency. Each unit of analysis (*pgm, py, pg, cm, cy, c, am, ay, a*) thus has its own individual table and storespace; each piece of data is stored once and only once; each individual relation is unique across the entire 90 GB database. These measures eliminate errors stemming from data duplication across various datasets as well as mitigate potential human errors.

Quantities of interest are computed and stored back in the database automatically through an organized data ingestion process with each monthly update, as are imputations and monthly estimation results.

The database is completely versioned; new versions are automatically produced through a custom-built backup-and-store mechanism every week as well as after each data update cycle. For security, a unique hash of each backup is created with each backup and stored separately from the very beginning, preventing data tampering or accidental corruption.

L Future extensions

As we continue to develop ViEWS, we will follow the guidelines of Colaresi & Mahmood (2017). They suggest a workflow inspired by machine-learning procedures where the performance of forecasts are iteratively critiqued and the successes and failures of previous forecasts inform the next generation. To facilitate this workflow, ViEWS relies on extensive documentation of monthly forecasts. Change logs are provided at `https://github.com/UppsalaConflictDataProgram/OpenViEWS/blob/master/CHANGELOG.md`.

We envision future extensions of ViEWS in multiple areas. A first important extension will be to incorporate the ‘early-warning’ component more explicitly. As we have seen above, the current forecasts are mainly driven by past conflict. To facilitate ‘early warning’ of new conflicts the top priority for ViEWS moving forward is the collection of data and estimation of models that can improve forecasts of conflict onset. Current efforts involve including data on elections and large-scale changes to political institutions. ViEWS will also improve models for how the three different outcomes affect each other, and explore how to model escalation and de-escalation within units beyond the conflict/non-conflict dichotomy currently in use.

Another extension is to include actors as units, specifying *who* are involved in events. ViEWS will focus on all relevant pairings of actors identified by the UCDP as participants in the political violence events they record, as well as all governments and selected other actors, such as protest movements. This level allows tracing actor-related escalation, termination, recurrence, transformation, and external involvement in conflicts, complementing the geographically defined system. An early attempt to set up actor-level predictions appears in Croicu & Hegre (2018).

A third extension is to solicit structured input from a set of area experts that will sketch the main political conflict issues and relevant actors’ position in their countries of expertise as input to game-theoretic models that generate forecasts (as in Schneider, Finke & Bailer, 2010). The experts will also provide own assessments of the likelihood of conflict escalation or de-escalation.

ViEWS also aims to refine existing and develop new performance metrics, such as measures based on a principled analysis of misclassification costs, and visual tools for criticism based on geographic context. Finally, ViEWS continually seeks to improve forecasts by comparing other ensemble methods to the current average forecast, in particular tree-based models.

To document these extensions, we will follow up with a series of annual short articles to present the future development of ViEWS. In these, we will present new methodological, theoretical, and data-related innovations in ViEWS relative to the previous articles, and show a brief extract of the most recent forecasts for the year following the previous publication. We will also present a formalized comparison between the forecasts published the year before and the events that actually happened the year after. As such, the series format would ensure that the forecasts are evaluated not only out-of-sample, but as true forecasts of events that were unknown to the researchers at the time their models were trained.

ViEWS will also work to develop and adapt a number of other performance metrics. We will develop a domain-specific evaluation measure based on differential classification rewards and misclassification costs. We will adapt the concept of 'Earth Mover Distance' (EMD) as a score with which to compare models (Pele & Werman, 2008, 2009). EMD calculates the minimum amount of work necessary to move one distribution of values arrayed in n-dimensional space, such as our predictions across space and time, to a target distribution, such as the actual observations of conflict or its absence across *pgms*.

References

Allansson, Marie; Erik Melander & Lotta Themnér (2017) Organized violence, 1989–2016. *Journal of Peace Research* 54(4): 574–587.

Allison, Paul D. (2009) Missing Data. In: Roger E. Millsap & Alberto Maydeu-Olivares (eds) *The SAGE handbook of quantitative methods in psychology*. Sage Publications.

Beger, Andreas; Cassy L Dorff & Michael D Ward (2014) Ensemble forecasting of irregular leadership change. *Research & Politics* 1(3).

Blyth, Simon (2002) *Mountain watch: environmental change & sustainable developmental in mountains*. Number 12. UNEP/Earthprint.

Breiman, Leo (2001) Random forests. *Machine learning* 45(1): 5–32.

Buuren, Stef van *Flexible imputation of missing data*. CRC press.

Chao Chen, Andy Liaw & Leo Breiman (2004). Using random forests to learn imbalanced data. In: University of California-Berkley Tech Report 666.

Colaresi, Michael & Zuhaib Mahmood (2017) Do the robot: Lessons from machine learning to improve conflict forecasting. *Journal of Peace Research* 54(2): 193–214.

Coppedge, Michael; John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Frida Andersson, Michael Bernhard, Steven M. Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Pamela Paxton, Daniel Pemstein, Laura Saxer, Brigitte Seim, Rachel Sigman & Jeffrey Staton (2017) V-dem codebook v7.1.

Croicu, Mihai & Håvard Hegre (2018) A fast spatial multiple imputation procedure for imprecise armed conflict events.

Croicu, Mihai & Ralph Sundberg (2013) Ucdp georeferenced event dataset codebook version 4.0 (http://www.pcr.uu.se/research/ucdp/datasets/ucdp_ged/).

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

Gleditsch, Kristian S. & Michael D. Ward (1999) A revised list of independent states since the congress of vienna. *International Interactions* 25(4): 393–413.

Graefe, Andreas; Helmut Küchenhoff, Veronika Stierle & Bernhard Riedl (2015) Limitations of ensemble bayesian model averaging for forecasting social science problems. *International Journal of Forecasting* 31(3): 943–951.

Graham, John W.; Allison E. Olchowski & Tamika D. Gilreath (2007) How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science* 8(3): 206–213.

Hegre, Håvard; Marie Allansson, Matthias Basedau, Mike Colaresi, Mihai Croicu, Hanne Fjelde, Frederick Hoyles, Lisa Hultman, Stina H́ogbladh, 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 & Jonas Vestby (2019) Views: A political violence early warning system. *Journal of Peace Research* 56(2).

Hegre, Håvard; Mihai Croicu, Kristine Eck & Stina Högbladh (2018) Ucdp-monthly. monthly updated organized violence data in event and aggregated forms.

Klein Goldewijk, Kees; Arthur Beusen, Gerard Van Drecht & Martine De Vos (2011) The hyde 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography* 20(1): 73–86.

Lall, Ranjit (2016) How Multiple Imputation Makes a Difference. *Political Analysis* 24(4): 414–433 (https://www.cambridge.org/core/journals/political-analysis/article/how-multiple-imputation-makes-a-difference/8C6616B679EF8F3EB0041B1BC88EEBB9).

Maddison, Angus (2007) *Contours of the World Economy 1–2030*. Essays in Macroeconomic History. Oxford: Oxford University Press.

Meiyappan, Prasanth & Atul K Jain (2012) Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science* 6(2): 122–139.

Melander, Erik; Therése Pettersson & Lotta Themnér (2016) Organized violence, 1989–2015. *Journal of Peace Research* 53(5): 727–742.

Montgomery, Jacob M; Florian M Hollenbach & Michael D Ward (2012) Improving predictions using ensemble bayesian model averaging. *Political Analysis* 20(3): 271–291.

Muchlinski, David; David Siroky, Jingrui He & Matthew Kocher (2016) Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. *Political Analysis* 24(1): 87–103.

Nordhaus, William D. (2006) Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences of the United States of America* 103(10): 3510–3517.

Pele, Ofir & Michael Werman (2008). A linear time histogram metric for improved sift matching. In: Computer Vision–ECCV 2008 , 495–508. Springer.

Pele, Ofir & Michael Werman (2009). Fast and robust earth mover's distances. In: 2009 IEEE 12th International Conference on Computer Vision , 460–467. IEEE.

Pettersson, Therése & Kristine Eck (2018) Organized violence, 1989–2017. *Journal of Peace Research* 55(4): 535–547 (https://doi.org/10.1177/0022343318784101).

Raftery, Adrian E.; Tilmann Gneiting, Fadoua Balabdaoui & Michael Polakowski (2005) Using bayesian model averaging to calibrate forecast ensembles. *American Meteorological Society* 133(1): 1155–1173.

Raftery, Adrian E & Steven M Lewis (1992) Comment: One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Statistical Science* 7(4): 493–497.

Raleigh, Clionadh; Håvard Hegre, Joakim Karlsen & Andrew Linke (2010) Introducing acled: An armed conflict location and event dataset. *Journal of Peace Research* 47(5): 651–660.

Randahl, David (2016) Raoul: An r-package for handling missing data (http://www.diva-portal.org/smash/get/diva2:940656/FULLTEXT01.pdf).

Ricardo Barandela, Jose Salvador Sanchez, Vicente Garcıa & Edgar Rangel (2003) Strategies for learning in class imbalance problems. *Pattern Recognition* 36(3): 849–851.

Rubin, Donald B. (1987) *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Samir, KC; Bilal Barakat, Vegard Skirbekk & Wolfgang Lutz (2008) *Projection of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 2005–2050. IIASA IR-08-xx.* Laxenburg, Austria: International Institute for Applied Systems Analysis.

Schneider, Gerald; Daniel Finke & Stefanie Bailer (2010) Bargaining power in the european union: An evaluation of competing game-theoretic models. *Political Studies* 58(1): 85–103 (https://doi.org/10.1111/j.1467-9248.2009.00774.x).

Storeygard, Adam; Deborah Balk, Marc Levy & Glenn Deane (2008) The global distribution of infant mortality: a subnational spatial view. *Population, Space and Place* 14(3): 209–229.

Sundberg, Ralph & Erik Melander (2013) Introducing the ucdp georeferenced event dataset. *Journal of Peace Research* 50(4): 523–532.

Tollefsen, Andreas Forø (2012) Prio-grid codebook (http://file.prio.no/ReplicationData/PRIO-GRID/PRIO-GRID_codebook_v1_01.pdf).

Tollefsen, Andreas Forø; Håvard Strand & Halvard Buhaug (2012) Prio-grid: A unified spatial data structure. *Journal of Peace Research* 49(2): 363–374.

Uchida, Hirotsugu (2009) Agglomeration index: towards a new measure of urban concentration.

Vogt, Manuel; Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker & Luc Girardin (2015) Integrating data on ethnicity, geography, and conflict the ethnic power relations data set family. *Journal of Conflict Resolution*: 0022002715591215.

Weidmann, Nils B. (2014) On the accuracy of media-based conflict event data. *Journal of Conflict Resolution Online first, DOI: 10.1177/0022002714530431*: 1–21.

Weidmann, Nils B; Doreen Kuse & Kristian Skrede Gleditsch (2010) The geography of the international system: The cshapes dataset. *International Interactions* 36(1): 86–106.

White, Ian R.; Patrick Royston & Angela M. Wood (2011) Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* 30(4): 377–399 (https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4067).

World Bank (2017) World development indicators.