

# Views Forecasts Dataset Codebook

Edition 2 - August 2018

Valid from run.07.2018

ViEWS team<sup>1</sup>

<sup>1</sup>Department of Peace and Conflict Research, Uppsala University

September 21, 2018

Please do not cite this document. Instead, cite the recommended citation as listed on the ViEWS webpage.

This document does not contain either definitions or operationalizations for the concepts used here. Please refer to the ViEWS webpage and article set for those.

## 1 Files

Predictions are supplied as a series of six separate folders. Each folder contains predictions for one of the three ViEWS outcome (**sb** for state-based conflict, **ns** for non-state conflict and **os** for one-sided violence) and one level of analysis (**cm** for country-month and **pgm** for PRIO-Grid-month). As such, the six folders are named as such:

- `predictions_sb_pgm` : Predictions for state-based conflict, PRIO-Grid month level
- `predictions_sb_cm` : Predictions for state-based conflict, country month level
- `predictions_ns_pgm` : Predictions for non-state conflict, PRIO-Grid month level
- `predictions_ns_cm` : Predictions for non-state conflict, country month level
- `predictions_os_pgm` : Predictions for one-sided violence, PRIO-Grid month level
- `predictions_os_cm` : Predictions for one-sided violence, country month level

Each folder contains a set of RFC-4180-compatible `csv` files - one such file for each ensemble estimated for the given outcome variable and level of analysis. Further, each folder contains a metadata file describing the ensembles and constituent models in `JSON` format for ease of machine parsing. For example, the `predictions_sb_pgm` folder will contain the following files:

- `average_calib_all_sb.csv` : simple average, calibrated data, full specification, state-based
- `average_calib_cl_sb.csv` : simple average, calibrated data, cross-level layer only, state-based
- `average_calib_wcm_sb.csv` : simple average, calibrated data, with country-month layer, state-based

- `average_calib_nocm_sb.csv` : simple average, calibrated data, no country-month layer, state-based
- `average_calib_select_sb.csv` : simple average, calibrated data, select specification, state-based
- `metadata.json` : metadata for each constituent models

Each `csv` file contains :

- `month_id`: Views month id. Months is a serial integer where 1 is **January 1980**.
  - To obtain the calendar year, use  $\text{floor}(\text{month\_id}/12) + 1980$  where `floor` is the down-rounding formula (verbatim code will work in both R and Python)
  - To obtain the calendar month, use  $\text{month\_id}\%12$  where `%` is the modulus operator (verbatim code will work in both R and Python).
- `[pg_id || country_id]` : Depending on whether the level of analysis is country month or PRIO-grid month, either the PRIO-grid identifier or the country identifier using the Gleditsch and Ward country code system;
- `ensemble_name` : One column, giving the predictions for the ensemble;
- `component_model_name_1 . . . . component_model_name_n` :  $n$  columns, with one column for each of the  $n$  component models of the ensemble. Each of the  $n$  columns contains the predictions for the given component model. Consult the JSON file or the table below for the name of each component model.
- `month`: Calendar month (1–12)
- `year`: Year in YYYY format (e.g. 2018)

## 2 Predictions

Predictions are given as calibrated predicted probabilities, which thus range between 0 and 1. Each variable in the dataset is named using the following convention:

### 2.1 Ensemble naming scheme

Currently, the only ensembles used are calibrated averages. Thus, all ensemble names will start with the `average_calib_` prefix followed by the naming scheme below.

### 2.2 Constituent model naming scheme

- `prefix` : The model name, as described in table 1. This prefix contains the following content:
  - Simulation mode:
    1. `ds`: dynamic simulation
    2. `osa`: one-step ahead simulation

## 2.3 General naming scheme

- **name** : The model or ensemble name, as described above, followed by:
  - Resolution:
    1. **pgm**: PRIO-Grid month
    2. **cm**: Country Month
  - Feature Set : See feature-set column in table 1 for details on the short name of each model based on feature set.
  - Contains cm layer (only applicable for pgm models):
    1. **all**: *only for ensembles* - Ensemble includes all models on all layers'
    2. **select**: *only for ensembles* - Ensemble includes the "select" specification (see JSON file for the construction)
    3. **wcm**: yes
    4. **nocm**: no
    5. **cl**: cross-level (cm\*pgm)
  - Training period: Always entire forecast period, denoted **fcast\_test**
  - Statistical method used:
    1. **logit**: Logistic Regression
    2. **rf**: Random Forest
  - Downsampling:
    1. **downsampled** if downsampling was employed.
    2. **fullsample** if full sample was used.
- **unit of analysis** that is forecasted:
  1. **sb**: state-based conflict
  2. **ns**: non-state one-sided violence
  3. **os**: one-sided violence
- **calibration** :
  1. **calibrated**: Predictions have been calibrated before inclusions in the ensemble using a calibration time period.
  2. **noncalibrated**: Predictions have not been calibrated.

Example:

`osa_pgm_canon_nocm_fcast_test_logit_fullsample_ns_calibrated`

This model:

- estimated using a one-step ahead framework
- forecasts at the PRIO-Grid month resolution
- contains the features of the "canonical" set
- contains no country-month predictors

- is estimated using a logistic regression
- is not downsampled
- predicts non-state violence
- is calibrated on a calibration window

Model name	Dyna- mic mode	Stat. model	Features	Temporal coverage	Weight in en- sem- ble	AUROC (sb)	Brier (sb)	AUPR (sb)
Country-level (cm)								
ds_cm_canon_base_fcast_test*	DS	Logit	cm core	1990– 2016	0.1667	0.94419	0.08766	0.81801
ds_cm_acled_base_fcast_test*	DS	Logit	cm core + ACLED	1997– 2016	0.1667	0.93900	0.09025	0.80143
osa_cm_acled_base_fcast_test_logit*	OSA	Logit	cm core + ACLED	1997– 2016	0.1667	0.91981	0.09851	0.77482
osa_cm_acled_base_fcast_test_rf*	OSA	RF	cm core + ACLED	1997– 2016	0.1667	0.90150	0.09934	0.76869
osa_cm_canon_base_fcast_test_logit*	OSA	Logit	cm core	1990– 2016	0.1667	0.91855	0.09959	0.78842
osa_cm_canon_base_fcast_test_rf*	OSA	RF	cm core	1990– 2016	0.1667	0.89669	0.09771	0.76452
PRIO-Grid level (pgm)								
ds_pgm_canon_nocm_fcast_test*	DS	Logit	pgm core	1990– 2016	0.0833	0.91103	0.00611	0.20776
ds_pgm_canon_wcm_fcast_test*	DS	Logit	pgm core + cm_fcast_test vars.	1990– 2016	0.0833	0.91756	0.00600	0.21749
ds_pgm_acled_wcm_fcast_test*	DS	Logit	pgm core + ACLED + cm_fcast_test vars.	1997– 2016	0.0833	0.91856	0.00607	0.19714
osa_pgm_acled_nocm_fcast_test_rf*	OSA	RF	pgm core + ACLED	1997– 2016	0.0833	0.94875	0.00605	0.20373
osa_pgm_acled_wcm_fcast_test_rf*	OSA	RF	pgm core + ACLED + cm_fcast_test vars.	1997– 2016	0.0833	0.94845	0.00601	0.20507
osa_pgm_canon_nocm_fcast_test_logit*	OSA	Logit	pgm core	1990– 2016	0.0833	0.92290	0.00596	0.21989
osa_pgm_canon_nocm_fcast_test_rf*	OSA	RF	pgm core	1990– 2016	0.0833	0.94506	0.00606	0.19487
osa_pgm_canon_wcm_fcast_test_rf*	OSA	RF	pgm core + cm_fcast_test vars.	1990– 2016	0.0833	0.94858	0.00602	0.19662
cl_ds_pgm_canon_nocm_fcast_test*	DS	Logit	pgm core + cm_fcast_test core	1990– 2016	0.0833	0.91469	0.00598	0.20110
cl_ds_pgm_acled_nocm_fcast_test*	DS	Logit	pgm core + ACLEDv	1997– 2016	0.0833	0.91765	0.00603	0.19961
cl_osa_pgm_canon_nocm_fcast_test_logit*	OSA	Logit	pgm core + cm_fcast_test core	1990– 2016	0.0833	0.91082	0.00593	0.21472
cl_osa_pgm_acled_nocm_fcast_test_rf*	DS	Logit	pgm core + ACLED + cm_fcast_test core	1997– 2016	0.0833	0.94201	0.00608	0.20097

Table 1. Models in the ViEWS ensemble