# Building and improving VIEWS, a political violence early-warning system

Håvard Hegre[1,2]

[1]Peace Research Institute Oslo (PRIO)
[2]Department of Peace and Conflict Research, Uppsala University

Presentation at the German Data Science Days 2023
LMU Munich, 10 March 2023

# Aim: Forecasting the incidence of organized armed violence

Number of fatalities in state-based armed conflict

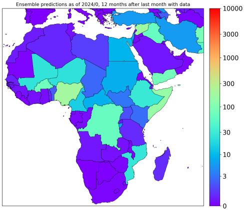
- a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year.
- Data from Uppsala Conflict Data Program
  - Monthly candidate ⇒ annual final events data

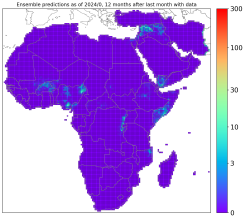
Update schedule, production system

- Monthly updates
- Forecasts for all of 1–36 months into the future
- Country and geographical/PRIO-GRID level

Developed since 2017 at PRIO (Peace Research Institute Oslo) and Uppsala University

Forecasts based on data up to December 2022, Country (top) and PRIO-GRID (bottom)



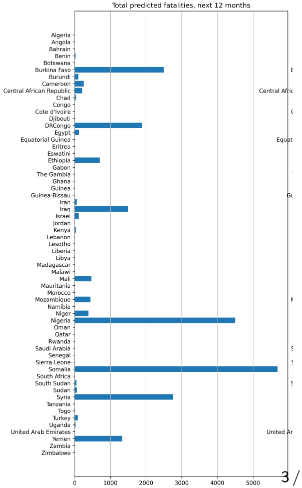



# Transparency

**VIEWS strives for maximal transparency**

- Open data sources
- Open source code
  - https://github.com/prio-data/viewser/wiki
  - https://github.com/viewsforecasting
- Publicly available forecasts
  - https://viewsforecasting.org
- Publicly available evaluation
  - https://doi.org/10.1177/0022343320962157
  - http://uu.diva-portal.org/smash/get/diva2:1667048/FULLTEXT01.pdf

Predicted number of fatalities in 2023, Africa and the Middle East



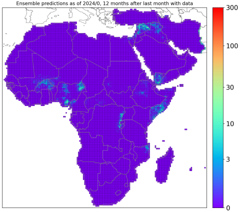
# The prediction problem: Major challenges

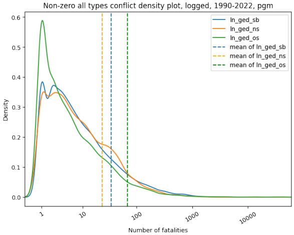
Data sparseness:

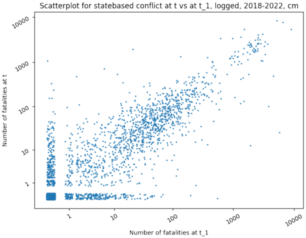
- Most observations are zero
  - 87% of country months
  - 99% of PRIO-GRID months
- Non-zeros strongly right-skewed
  - $\Rightarrow$ power-law distribution
- Strong autocorrelation

Theoretical challenges:

- Armed conflicts have multiple causes
- Latent risk unobservable until outbreak
- War initiation decisions fraught with fundamental uncertainty



# Methodology, current system

Building-block constituent models

- Separately by:
    - Country and geographical level
    - Each step forward
- Combinations of feature sets and algorithms

Ensembles

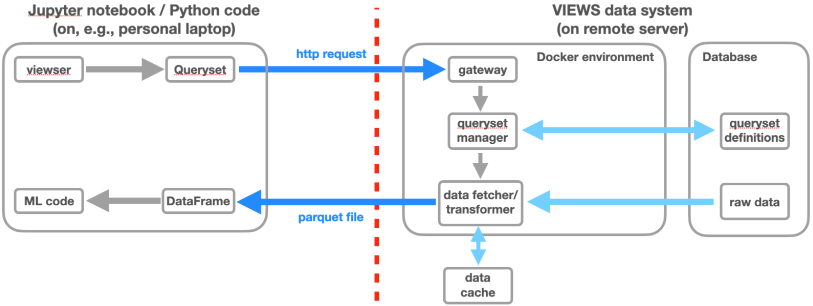
- Calibration
- Weighting

Partitioning: 'Leave the future out'

- 1990–2013: Train models
- 2014–2017: Estimate weights, hyper-parameters, calibration
- 2018–2021: Test
    - Repartition for true future forecasts

# Infrastructure: main components

- Data in SQL database
- Bespoke code to ingest data into database
- 'viewser': queryset system to:
  - Retrieve data
  - Transform features
- viewser API can be imported into any Python code
- Model organization in Python scripts/notebooks

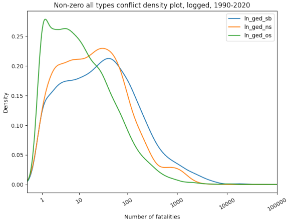


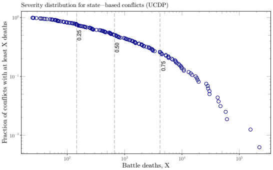
# Machine-learning models

**Core models:** Decision-tree models

- Random forests (XGB implementation)
- Gradient boosting models (XGB/LGB/sklearn implementations)

---




**Distribution of outcome challenge – Solutions:**

- Predicting $log(Y+1)$
- Hurdle models (Fritz et al. 2022)
  - Learn probability of non-zero observations $\hat{p}_{nz} = p(Y > 0)$
  - Learn number of fatalities if non-zero $\hat{Y}_{nz} = Y | Y > 0$
  - Combined prediction $\hat{Y} = \hat{p}_{nz} \times \hat{Y}_{nz}$
- Markov models (Randahl and Vegelius 2022)

# Predictors: country level

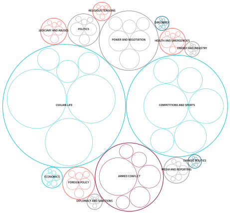
**'Feature sets':**

- Conflict history
  - Lots of lags, decay functions
  - Spatially and temporally
- Political institutions (V-Dem)
- Development (WDI)
- News monitoring, topic model (Mueller and Rauh 2018)

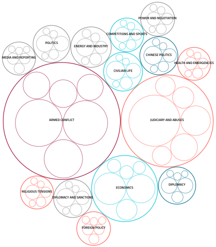
**Selected indicators:**

- Conflict in country previous months
- Conflict in neighboring countries
- Liberal democracy
- Infant mortality rates

Burkina Faso, January 2017




Burkina Faso, April 2017



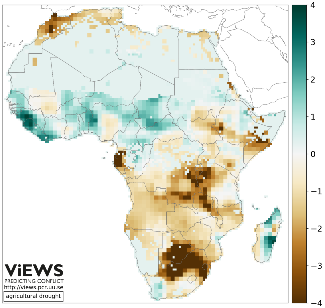
# Predictors: grid-cell level

**'Feature sets':**

- Conflict history
  - Lots of lags, decay functions
  - Spatially and temporally
- Natural geography features (terrain, resources)
- Social geography features (cities, borders, demography)
- Climate: drought, growing season
- Protests (from ACLED)

**Selected indicators:**

- Conflict in cell and neighboring cells
- Distance to oil extraction
- Protests

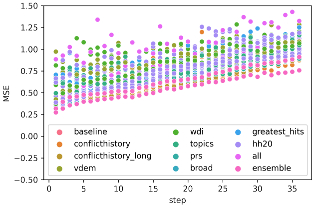


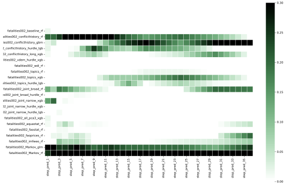

Agricultural drought, June 2018

# Ensembling

Ensemble of models that perform well but are reasonably diverse

- ‘Wisdom of the crowd’ – most wise when diverse and competent
- Ensembling safeguards against overfitting
    - But can be over-fit to calibration partition
- Country-level model weights trained using a genetic algorithm
- Optimizing on mean squared error of prediction across all cases



# Evaluation metrics

**Out-of-sample evaluation and development**

- What constitutes a good prediction?
- Current main metric: Mean Squared Error (of $ln(Y+1)$ )
- The square of the difference between what we predict and what actually happened
- Favors models well calibrated at large

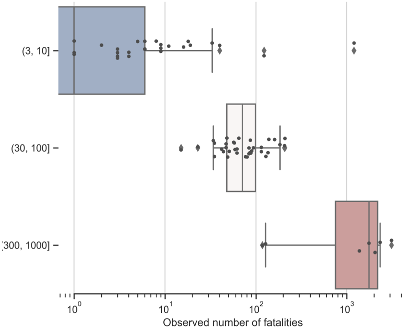
**Useful at cm level, not so useful at pgm**

- Tends to favor 'nihilistic' models – e.g. models that predict no change
- Bin-by-bin evaluation inappropriate
  - Alternative metric based on 'earth-mover distance'

**We may be more interested in the probability of extreme events than in the point prediction**

# How well do we predict?

MSEs at country level between .25 and .75

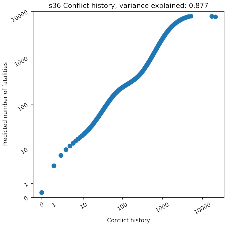



How many were killed per country if we predict the following 12 months into the future:

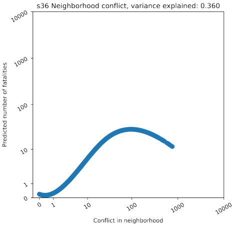
- 3–10 fatalities:
  - 50% are 1 or higher, median observation is 1, and 95% are below 30
- 30–100 fatalities:
  - 90% are between 30 and 200
- 300–1000 fatalities:
  - all are above 100, and 90% are above 800

# Which features are important?

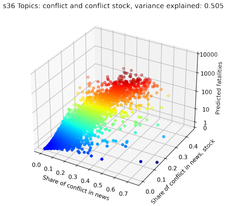
*Conflict history (.877)*



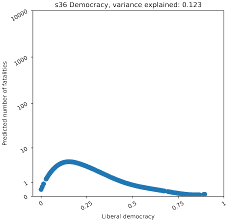

*Neighboring conflict (.360)*



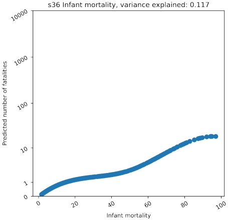

Topic 10: *Conflict (.505)*



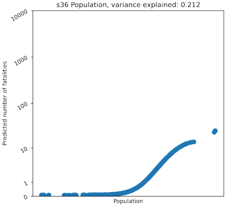

*Democracy (.123)*




*Infant mortality rate (.117)*




*Population size (.212)*



# Strengths, shortcomings, next steps

Strengths:

- Model works well for current conflict situations
- Most violence occurs in these

Shortcomings and suggested solutions:

- Forecasting onset of conflict is very hard
    - Improve models and input features
    - Rephrase optimization criteria to give more weights to onsets
- Fatality model yields only point predictions
    - Model prediction uncertainty

Other next steps:

- Neural nets
- Actor layer

# Sources of uncertainty in prediction modeling

Aim:

- To produce VIEWS forecasts as probability distributions over possible fatality counts

Why?

- Users want to know the uncertainty of predictions
- Point predictions yield the most likely outcome, but we are interested in low-probability, catastrophic events

How to reach this?

1. Create 'draws' by combining forecasts from:
    - Constituent models
    - Bootstraps of input data
    - Realizations from measurement models
2. Bootstraps from predictions/conformal predictions

# Uncertainty regarding input data: measurement models

1. How many did really die in each conflict?
    - What is documentable (and when) versus what really happened
        - Solution: Complementing UCDP's 'best' estimates with probability distributions over the true values
        - Distribution obtained through an expert elicitation
2. When and where did violence occur?
    - 'Known geographic imprecision' – UCDP notes location is imprecise and assigns placeholder location
        - Estimate the spatial probability distribution for each conflict
        - Randomly draw location based on distribution
3. Candidate data are imperfect approximations to final data
    - Solution: 'now-cast' final GED data using a ML model

# Uncertainty in model evaluation

Our test dataset is just a sample

- Statistical uncertainty regarding the evaluation metric
- Solution: bootstrapping

More fundamentally:

- What are the best evaluation metrics?
- Test window seen as a sample

Evaluation metrics designed for predictions as probability distributions

- CRPS
- Interval scores
- Ignorance score

# VIEWS 2023/24 prediction competition

Prediction competition:

- Predicting the number of fatalities from organized political violence as probability distributions
- Broad set of contributors working on a common, well-defined challenge
- Predecessor:
  - *International Interactions*
  - `https://www.tandfonline.com/doi/full/10.1080/03050629.2022.2029856`
  - Ideas are becoming incorporated in ViEWS

INTERNATIONAL INTERACTIONS
https://doi.org/10.1080/03050629.2022.2029856

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

## United They Stand: Findings from an Escalation Prediction Competition

Paola Vesco[a], Håvard Hegre[a,b], Michael Colaresi[a,c], Remco Bastiaan Jansen[a], Adeline Lo[d], Gregor Reisch[e], and Nils B. Weidmann[f]

[a]Uppsala University; [b]Peace Research Institute Oslo; [c]University of Pittsburgh; [d]University of Wisconsin-Madison; [e]Formerly German Federal Foreign Office; [f]University of Konstanz

**ABSTRACT**

This article presents results and lessons learned from a prediction competition organized by ViEWS to improve collective scientific knowledge on forecasting (de-)escalation in Africa. The competition call asked participants to forecast *changes* in state-based violence for the true future (October 2020–March 2021) as well as for a held-out test partition. An external scoring committee, independent from both the organizers and participants, was formed to evaluate the models based on both qualitative and quantitative criteria, including performance, novelty, uniqueness, and replicability. All models contributed to advance the research frontier by providing novel methodological or theoretical insight, including new data, or adopting innovative model specifications. While we discuss several facets of the competition that could be improved moving forward, the collection passes an important test. When we build a simple ensemble prediction model—which draws on the unique insights of each contribution to differing degrees—we can measure an improvement in the prediction from the group, over and above what the average individual model can achieve. This wisdom of the crowd effect suggests that future competitions that build on both the successes and failures of ours, can contribute to scientific knowledge by incentivizing diverse contributions as well as focusing a group's attention on a common problem.

Este artículo presenta los resultados y las enseñanzas extraídas en el marco de un certamen de predicción organizado por los responsables del proyecto Sistema de Alerta Temprana de Violencia (Violence Early-Warning System, ViEWS) con el propósito de mejorar los conocimientos científicos colectivos sobre la previsión de la (des)escalada en el continente africano. En el certamen se pidió a los participantes que desarrollaran una previsión con respecto a los cambios en la violencia estatal para el futuro real (de octubre de 2020 a marzo de 2021), así como para una muestra de prueba que se

**KEYWORDS**
Conflict; escalation; forecasting; political violence; prediction

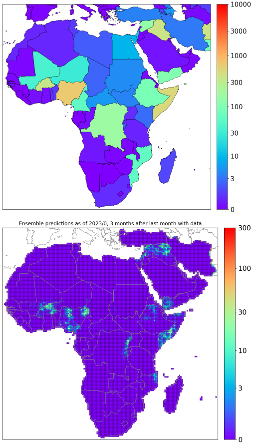
CONTACT Paola Vesco paola.vesco@pcr.uu.se Department of Peace and Conflict Research, Uppsala University, Uppsala, 751 20, SE, Sweden.

Replication data can be found at https://dataverse.harvard.edu/dataverse/internationalinteractions. Please address all questions regarding replication to paola.vesco@pcr.uu.se

Supplemental data for this article can be accessed here.

# Overview

- For state-based conflict
  - Number of fatalities as recorded by the UCDP
- For two levels of analysis
  - Per country
  - Per ‘grid cell’
  - Global coverage at country level
  - Africa and the Middle East at grid level (13,000 cells)
- Two prediction windows
  - True future: The year of 2024
  - Test set: Each of the years 2019–2022
  - All forecasts based on data up to and including October the year before
- Main evaluation metric: CRPS




Predicted fatalities in Dec-22

# Structure of contributions

- Short summary of model/contribution, basis for introduction article
- 3–5 page write-up of model/contribution, for workshop and as working paper
- Forecasts:
  - The true future:
    - Predictions for all months of 2024, based on data up to October 2023
    - Month-by-month evaluation updates on VIEWS website

  Four sets of test predictions:
    - Predictions for all months of 2019, 2020, 2021, 2022, for data up to October the month before

  Unit of analysis:
    - Either country-month or PRIO-GRID-month

  Format:
    - Up to 1,000 draws from the prediction distribution
    - or, point estimates (we will generate samples)

# Time frame, tentative

https://viewsforecasting.org/prediction-competition-2/

- Early March 2023: formal invitation to participate
- 15 May 2023: Deadline for abstracts for participants; data and code to participants
- About 1 October 2023: Workshop for (selected) contributors. Contributors submit preliminary forecasts and summary papers just before
- 1 December: Providing all participants with updated data
- 10 December: Contributors submit the final predictions
- 1 January 2024: Start of prediction window
- 31 December 2024: End of forecasting window

# Questions?

**Contact:**

hhegre@prio.org or views@pcr.uu.se

**Websites:**

`http://viewsforecasting.org`

`https://www.prio.org/projects/1976`

**Newsletter:**

Email `views@pcr.uu.se` to register

Thanks to the VIEWS team for all the work
on data, modeling and presentation!