

# Predicting Escalating and De-Escalating Violence in Africa Using Markov Models

David Randahl  and Johan Vegelius

Uppsala University

## ABSTRACT

This contribution to the ViEWS prediction competition 2020 proposes using Markov modeling to model the change in the logarithm of battle-related deaths between two points in time in a country. The predictions are made using two ensembles of observed and hidden Markov models, where the covariate sets for the ensembles are drawn from the ViEWS country month constituent models. The weights for the individual models in the ensembles were obtained using a genetic algorithm optimizing the fit on the TADDA-score in a calibration set. The weighted ensembles of visible and hidden Markov models outperform the ViEWS prediction competition benchmark models on the TADDA score in the test period of January 2017 to December 2019 for all time steps. Forecasts until March 2021 predict increased violence primarily in Algeria, Libya, Tchad, Niger, and Angola, and decreased or unchanged levels of violence in most of the remaining countries in Africa. An analysis of the model weights in the ensembles shows that the conflict history constituent model provided by ViEWS was dominant in the ensembles.

Esta contribución a la competencia de predicciones 2020 del Sistema de Alerta Temprana de Violencia (Violence Early Warning System, ViEWS) propone utilizar la modelización de Márkov para elaborar un modelo del cambio en el logaritmo de las muertes relacionadas con batallas entre dos puntos temporales en un país. Las predicciones se elaboran con dos conjuntos de modelos observados y ocultos de Márkov, en los que los grupos de covariables de los conjuntos se obtienen de los modelos constituyentes mensuales de los países del ViEWS. La relevancia de los modelos individuales en los conjuntos se obtuvo mediante un algoritmo genético que optimiza el ajuste de la puntuación TADDA en un grupo de calibración. Los conjuntos ponderados de los modelos visibles y ocultos de Márkov superan los modelos de referencia de la competencia de predicciones del ViEWS en relación con la puntuación TADDA (Distancia absoluta orientada con aumento de dirección) en el período de prueba de enero de 2017 a diciembre de 2019 para todos los intervalos de tiempo. Las

## KEYWORDS

Forecast; Markov model; prediction; state-based violence

**CONTACT** David Randahl  [david.randahl@pcr.uu.se](mailto:david.randahl@pcr.uu.se)  Department of Peace and Conflict Research, Uppsala University, Uppsala, Sweden.

 Supplemental data for this article can be accessed [here](#).

This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

predicciones hasta marzo de 2021 pronostican un aumento en la violencia principalmente en Argelia, Libia, Chad, Níger y Angola, y niveles de violencia disminuidos o sin variaciones en la mayoría de los países restantes en África. Un análisis de la relevancia de los modelos en los conjuntos demuestra que los modelos constituyentes de la historia de conflictos que proporciona el ViEWS fueron dominantes en dichos conjuntos.

Cette contribution au concours de prévision ViEWS (Violence early-warning system, système d'alerte précoce sur la violence) 2020 propose d'utiliser la modélisation de Markov pour modéliser l'évolution du logarithme des décès liés aux conflits entre deux moments de l'histoire d'un pays. Les prédictions sont effectuées à l'aide de deux ensembles de modèles de Markov cachés et de modèles de Markov observés, et les jeux de covariables de ces ensembles sont tirés des modèles constituants par mois et pays du système ViEWS. Les pondérations des modèles individuels des ensembles ont été obtenues en utilisant un algorithme génétique optimisant l'ajustement sur le score TADDA (Distance absolue ciblée avec augmentation de direction) dans un jeu de calibration. Les ensembles pondérés de modèles de Markov visibles et cachés sont plus performants que les modèles de référence du concours de prédiction ViEWS pour ce qui est du score TADDA de la période de test de janvier 2017 à décembre 2019, et ce pour tous les pas de temps. Les prévisions jusqu'à mars 2021 ont permis de prédire une augmentation de la violence principalement en Algérie, en Libye, au Tchad, au Niger et en Angola, et une diminution ou un maintien des niveaux de violence dans la plupart des autres pays d'Afrique. Une analyse des pondérations des modèles dans les ensembles montre que le modèle constituant basé sur l'histoire des conflits fourni par ViEWS serait dominant dans les ensembles.

## Introduction

In this contribution to the 2020 ViEWS prediction competition, we explore the usefulness of Markov models in forecasting change in conflict-related fatalities across Africa. In the Markov model an individual moves through different states over time, and the probability of transitioning from one state to another depends on the current state and is independent of previous states. The Markov model can be extended to include covariates on the transition probabilities so that the probability of any future state is conditional on the current state of the individual and the covariate values of the individual at the current time. Additionally, the states in the Markov model can either be directly observable, here referred to as the observed Markov model (OMM), or hidden, where the underlying, latent, state can only be observed through some other observable representation of the latent state. The Markov model with hidden states is referred to as the hidden Markov model (HMM).<sup>1</sup>

---

<sup>1</sup>For a more comprehensive description of both the OMM and HMM, see Jackson (2011).

Markov models have been proven to be useful in a wide range of scientific fields where sequences of information need to be categorized. For instance in speech recognition, where the aim is to interpret sound waves as sequences of bits of information, and in protein structure determination, where sequences of amino acids of unknown length can be classified as different proteins (see for instance Rabiner and Juang 1986; Rabiner 1989; Stultz, White, and Smith 1993; White, Stultz, and Smith 1994; Krogh et al. 2001). In these examples, the object to be classified, i.e. individual phonemes or proteins, are not directly observable. Instead, frequencies of sound and chains of amino acids are the observable representation of those objects. Modeled as a hidden Markov model the phonemes and proteins are seen as the hidden states that are classified based on the observed sound frequencies and amino acids. Given a specific state, i.e. phoneme or protein, the likelihood of transitioning to another state can also be determined, as certain phonemes or proteins are more likely to follow a given phoneme or protein than others.

In terms of conflict forecasting, Markov models provide a useful tool as conflicts can be seen as moving through different *phases* of conflict dynamics over time (see for instance Diehl 2006; Brandt et al. 2012), where the different phases of conflict (or peace) can be thought of as Markov states. Given a specific, latent, state the conflict will then generate conflict events or fatalities based on the probability distribution of that state and will have a specific likelihood of transitioning to another state in the next time period. As the transition probabilities for the states can be estimated, it is possible to make long-term forecasts based on simulated Markov chains where each observation is simulated as moving through the forecast period a large number of times with the specific transition probabilities for each time step. Markov models have been used to forecast a variety of different conflict outcomes, including net conflictual actions between China and Taiwan conditional on the conflict dynamic (state) in the dyad (Brandt et al. 2012), and forecasts of conflict events in Lebanon and the Balkans conditional on the current dynamic in those specific conflicts (Schrodt 2006, 1997).

Our belief is that the real-world process which generates a change in fatalities from conflict closely resembles a Markov process with several states corresponding to different dynamics or phases of conflict. In this process, there would likely be a large number of states which would be unobserved or potentially even mixed, due to the unknown preferences and strategies of the actors involved. Such a process would therefore best be modeled through the hidden Markov model, where the observable representation of the latent state is the change in fatalities generated by each individual (country) between two time periods. However, as we do observe

the actual change in fatalities, an alternative to the hidden Markov model would be to approximate the true latent states using our domain knowledge of conflict processes to artificially create observable Markov states which can then be modeled as an observed Markov model. Both of these approaches have advantages and disadvantages. From a theoretical perspective, we believe that the hidden Markov model provides a more honest representation of the true data generating process as the model is free to identify the latent states based on the data rather than based on any preconceived notions of how the process behaves. At the same time, we believe that this freedom and the unknown nature of the states, and the number of states, make forecasts into the future less stable. It is also possible that the hidden Markov model fails to identify the correct number of states if data is not plentiful enough, or if there are complex transition patterns where transitions may only happen between certain states. The observed Markov model, on the other hand, suffers from artificially imposed states which may be crude representations of the true latent state but which are likely to deviate more from the true process. Such artificial and crude states may, however, provide a structure that helps produce more stable forecasts. Consequently, in this paper, we explore both of these approaches for forecasting the change in the number of fatalities from armed conflict.

There are three main reasons that we expect the Markov modeling approach to outperform other prediction methods for this problem. First, the conflict-state structure of the Markov model will allow the covariates to have different effects on the dependent variable in the different conflict-states. This should allow the models to better estimate the true effect of the covariates in each state, and thereby reduce the amount of random noise in the predictions. Second, separating the estimation of the effect of covariates on the change in fatalities given a specific state, and the effect of covariates on the transition between states, allows the model to capture complex non-linear relationships between the covariates and the outcome. This aspect is especially important for covariates which may have no impact on one of the two, i.e. either transitions or the actual outcome, or covariates which have effects in divergent directions, i.e. for instance increasing the likelihood of transitioning to a conflict state but lowering the expected change in fatalities given a conflict state. Third, for the observed Markov model(s) the existence of observable states allows us to fix the predictions to specific state-based values, such that no noisy predictions need to be estimated for those states.

The rest of the paper proceeds as follows. We begin our contribution by describing the forecasting problem and the mathematical formulations of the observed and hidden Markov models in more detail. We then proceed to the methodology and model specifications for our forecasting efforts. Finally, the

results of the models are presented and possible extensions of our framework are discussed.

### Markov Models for Forecasting Change in Fatalities

The forecast target for the ViEWS 2020 prediction competition is the change in the natural logarithm of battle-related deaths (BRDs) from state-based conflicts between the target month,  $t$ , and  $s$  months previously, using the data that is available at time  $t - s$  (Hegre, Vesco, and Colaresi 2022). Our belief is that this target closely resembles a process that is conditional on the underlying Markov state. This implies that when using Markov models to forecast this target the Markov states themselves, i.e. the conflict dynamics, are only interesting in the sense that we base our forecast of the target on the observed or estimated hidden Markov state of each underlying observation.

#### The Observed Markov Model

In the observed Markov model, the Markov states are directly observed and the propagation of states is defined by letting  $p(m_{i,t}|m_{i,t-1})$  be the probability of Markov state  $m$  for country  $i$  at time  $t$  conditioned on state  $m_{i,t-1}$  at time  $t - 1$ . This transition probability is modeled by multinomial logistic regression with covariates  $\mathbf{x}_{i,t}$  (including 1 for the intercept) and a set of parameter vectors corresponding to respective allowed transitions.

The prediction target, i.e.  $y_{i,t} = \Delta \log(\text{brd})_{i,t}$  is then modeled as conditional on the current state  $m_{i,t}$  such that  $y_{i,t}|m_{i,t} \sim F(\cdot)$ . For this prediction problem, we assume that  $F(\cdot)$  is either a constant or Normally distributed such that  $y_{i,t}|m_{i,t} \sim N(\mathbf{x}_{i,t}^T \beta_m, \sigma_m^2)$  where  $\beta_m$  is a vector of parameters corresponding to state  $m$  and  $\sigma_m^2$  is the variance corresponding to state  $m$ . In the case where  $y_{i,t}|m_{i,t} \sim N(\mathbf{x}_{i,t}^T \beta_m, \sigma_m^2)$  this is the equivalent to modeling  $y_{i,t}$  as an OLS regression with state-specific OLS estimators

$$\hat{\beta}_m = \left[ \sum_{i=1}^n \sum_{t=1}^{T_i} \mathcal{I}(m_{i,t} = m) (\mathbf{x}_{i,t} \mathbf{x}_{i,t}^T) \right]^{-1} \sum_{i=1}^n \sum_{t=1}^{T_i} \mathcal{I}(m_{i,t} = m) \mathbf{x}_{i,t} y_{i,t}$$

$$\hat{\sigma}_m^2 = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} \mathcal{I}(m_{i,t} = m) (y_{i,t} - \mathbf{x}_{i,t}^T \beta_m)^2}{n_m}.$$

where  $T_i$  is the number of time points of country  $i$ ,  $n$  is the number of countries in the sample,  $\mathcal{I}(m_{i,t} = m)$  is the indicator function which is 1 when  $m_{i,t} = m$  and 0 otherwise, and  $n_m = \sum_{i=1}^n \sum_{t=1}^{T_i} \mathcal{I}(m_{i,t} = m)$  is the number of observations in state  $m$  (across time and countries). Given the data until time  $t$  and these assumptions, the distribution of the dependent

variable at time  $t + 1$  is known and its expected value is available as an estimate for predictions. The same is true for later observations as well.

### The Hidden Markov Model

As in the observed Markov model, the propagation of states is defined by modeling  $p(m_t|m_{t-1})$  by multinomial regression. In contrast to the observed Markov model, in the hidden Markov model, the states are unobserved, i.e. latent. The density function of the forecasting target of observation  $i$  at time  $t$  can be viewed as a mixture of those states with proportions determined by the likelihood that the observation was produced by the respective states. Let  $p(y_{i,t}|m_{i,t})$  be the distribution of the target variable  $y_{i,t}$  conditioned on the state  $m_{i,t}$ . Define the vector of model parameters to be  $\theta$  which includes the state-specific parameters  $\beta_m$  and  $\sigma_m^2$  for all  $m$  and the transition-specific parameters in the logistic regression. Given  $\theta$  and  $p(y_{i,t}|m_{i,t})$ , the state proportions are

$$p(m_{i,t}|y_{i,t}) = \frac{p(y_{i,t}|m_{i,t})p(m_{i,t})}{\sum_{m_{i,t}} p(y_{i,t}|m_{i,t})p(m_{i,t})}. \quad (1)$$

However, neither the parameters  $\theta$ , nor the state proportions  $p(m_{i,t}|y_{i,t})$ , are known in advance. The EM algorithm of Dempster, Laird, and Rubin (1977) is an appropriate method to estimate  $\theta$  here and is implemented as follows: Given some initial vector  $\theta^{(0)}$  the proportions  $p(m_{i,t}|y_{i,t})$  are calculated from Equation (1) for each observation. Considering  $p(m_{i,t}|y_{i,t})$  as the distribution of the latent states, the expected value of the log-likelihood is maximized with respect to  $\theta$ . The vector that maximizes the expected value of the log-likelihood is  $\theta^{(1)}$  and the procedure is repeated until convergence. As previously, it is assumed that  $y_{i,t}|m_{i,t} \sim N(x_{i,t}^T \beta_m, \sigma_m^2)$ . After choosing the initial parameter vector  $\theta^{(0)}$ , the closed-form solution to the maximization problem with respect to  $\beta_m$  and  $\sigma_m^2$  is

$$\beta_m^{(1)} = \left[ \sum_{i=1}^n \sum_{t=1}^{T_i} p(m_{i,t}|y_{i,t}) x_{i,t} x_{i,t}^T \right]^{-1} \sum_{i=1}^n \sum_{t=1}^{T_i} p(m_{i,t}|y_{i,t}) x_{i,t} y_{i,t}$$

$$\sigma_m^{2(1)} = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} p(m_{i,t}|y_{i,t}) (y_{i,t} - x_{i,t}^T \beta_m)^2}{\sum_{i=1}^n \sum_{t=1}^{T_i} p(m_{i,t}|y_{i,t})},$$

where  $\sum_{i=1}^n \sum_{t=1}^{T_i} p(s_{it}|y_{it})$  can be interpreted as the effective sample size in state  $s$ . When the largest absolute value of the elements in  $\beta_m^{(k)} - \beta_m^{(k-1)}$  and  $\sigma_m^{2(k)} - \sigma_m^{2(k-1)}$ , for some positive integer  $k$ , is less than some critical value  $\epsilon$ , typically  $\epsilon = 1 \cdot 10^{-4}$ , the EM algorithm is assumed to have converged. Then the estimates are  $\hat{\beta}_m = \beta_m^{(k)}$  and  $\hat{\sigma}_m^2 = \sigma_m^{2(k)}$ .

Notice that the observed Markov model is, in some sense, a special case of the hidden Markov model in that  $p(m_i|y_i)$  is either 0 or 1, meaning that the states of each observation are known and the estimates above reduce to the group-specific OLS estimates of corresponding state parameters. There is no closed-form solution for the state-transition parameters. Therefore, the Newton-Raphson method was used to update those.

## Forecasting Setup

Taking a Markov approach to forecasting the change in  $\log(brd)$  between months, in essence, reduces the forecasting problem to a set of classification (on Markov states) and regression (on change in  $\log(brd)$ ) problems. As the classification and regression problems are conditional on the Markov states, both the number of Markov states and the covariates entering the model(s) need to be defined before forecasting.

## Markov States

The hidden Markov model allows us to pre-specify the number of Markov states that exist within the model. The model then finds these states in the data based on the observed realization (change in  $\log brd$ ) of the process and the covariates. For the forecasting approach, we tried running the hidden Markov model with 2, 3, and 4 hidden states. However, regardless of whether we ran the model with 2, 3, or 4 states, the model only seemed to find two stable states irrespective of the covariates which we entered into the model. We, therefore, settled for two hidden Markov states which we labeled as “peaceful” and “conflict.”

In the observed Markov model specification, on the other hand, we had to explicitly define discreet and directly observable Markov states. As discussed briefly in the Introduction, these states are in essence artificial since it is we as forecasters who define these states based on our domain knowledge of the process we are modeling. Ocular inspection of individual countries time-series of the forecast target initially suggested that 2 or 3 Markov states would yield a good estimate of the process, with an initial suggestion of states labeled “peaceful,” “at-risk,” and “conflict,” where transitions between states would be restricted such that a country would need to transition through the “at-risk” state to move from “peaceful” to “conflict” or vice-versa.

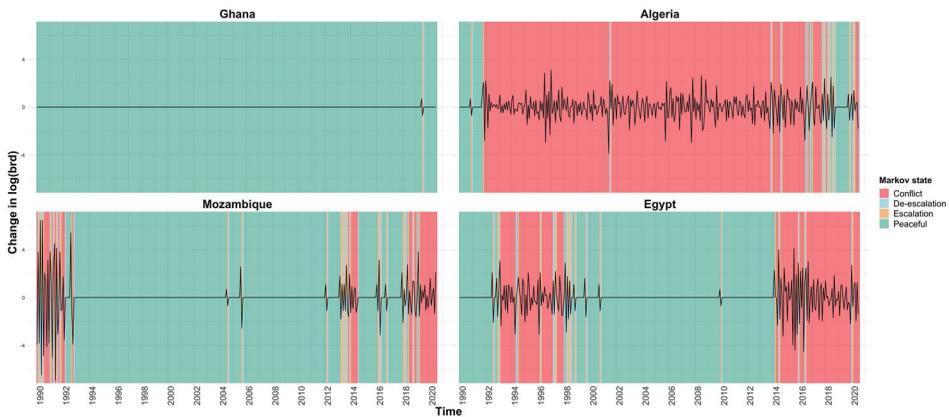
The three Markov-state specifications seem to capture the processes of the forecasting target well, but further theoretical consideration led to the division of the “at-risk” state into two separate states labeled “escalation” and “de-escalation.” Formally, these four states were defined as:

1. “Peaceful” if the observed number of battle related deaths at time  $t$  and at time  $t - 1$  were 0
2. “Escalation” if the observed number of battle related deaths at time  $t$  were  $>0$  and at time  $t - 1$  were 0
3. “De-escalation” if the observed number of battle related deaths at time  $t$  were 0 and at time  $t - 1$  were  $>0$
4. “Conflict” if the observed number of battle related deaths at time  $t$  and at time  $t - 1$  were  $>0$

Individual time-series of four countries, Ghana, Algeria, Mozambique, and Egypt, with the observed Markov state (four states), superimposed can be seen in [Figure 1](#) below.

Explicit in this definition of states is that when moving one time-step ahead each Markov state only allows for the transition to two states, i.e. transitions are allowed from “peaceful” to either “peaceful” or “escalation,” from “escalation” to “de-escalation” or “conflict,” from “de-escalation” to “escalation” or “peaceful, and from “conflict” to “conflict” or to “de-escalation.” This dynamic is visualized in Table 4 in the [Supplementary Appendix](#). When moving more than one step ahead, an individual country can move to any other state, a dynamic visualized in Table 5.

Forcing the “escalation” and “de-escalation” to be inherently transient states which can only last for one time period may seem an odd choice, but this specification has substantial upsides. Primarily, this allows for the two more common states, “peaceful” and “conflict” to be more stable as the estimation can discount transition between these states. Additionally, including “escalation” and “de-escalation” as transient states allow us to set the value of the forecasting target to a constant conditional on certain transitions. For instance, a transition from “de-escalation” to “peaceful” would imply that the forecasting target, change in  $\log brd$  would be zero.



**Figure 1.** Markov states over time for Ghana, Algeria, Mozambique, and Egypt.

Similarly, a transition from “conflict” to “de-escalation” would imply that the forecasting target would be  $-\log brd_{i,t-s}$  where  $s$  refers to the step-length of the target. Thus, including “escalation” and “de-escalation” as transient states reduce the number of “from-to” transition pairs where the forecast target needs to be estimated using OLS from 16 to 8, where the remaining 8 are constants. These expected changes in  $\log brd$  given the transition between states can be seen in Table 6 in the Appendix.

### **Covariate Sets**

To obtain predictions from the Markov models, we need to select appropriate covariates to include in the models. We have chosen a relatively agnostic approach to the covariates included by selecting 15 covariate sets from the ViEWS *country-month* constituent models (Hegre et al. 2021, Appendix B). We then train the models individually for each covariate set, yielding a total of 15 predicted values each for the hidden and observed Markov models. Finally, we present the forecasts as weighted ensembles of these 15 models for the observed and hidden Markov models, respectively. The 15 models specifications included are shown in Table 7 in the [Supplementary Appendix](#).

### **Model Averaging by Genetic Optimization**

As we expected the performance of the 15 constituent models included in the ensembles to have varying degrees of performance, we decided to set aside calibration data for each set of forecasts and use it to optimize the model weights based on the predictive performance in the calibration data. For set 1, the true forecasts for October 2020–March 2021, June 2016–December 2019 were used as calibration data. For set 2, the forecasts for January 2017–December 2019, June 2013–December 2016 were used as calibration data, and for set 3, May 2010–May 2013 were used as calibration data.<sup>2</sup> We then trained the constituent models on all data prior to the calibration data, i.e. on months January 1990–May 2016 for the first calibration set, January 1990–May 2013 for the second set, and January 1990–May 2010 for the third set, and assessed the predictive performance in the calibration sets.

The weights for the constituent models in each of the ensembles for each of the steps 1–7 were then optimized using a simple genetic algorithm (for an introduction to the concept of genetic algorithms, see for instance

---

<sup>2</sup>The seeming overlap between the calibration and test sets in the months June 2016–December 2016 is due to the time shifting, i.e. since the forecasts are shifted 2–7 months, no data that is part of the test data is actually used in the calibration set after time-shifting.

Mitchell 1998; Kumar, Jain, and Sharma 2018), which assessed the fitness of the weights based on the weighted ensembles TADDA-score in the calibration data. The genetic algorithm was constructed with 100 initial sets of weights that were allowed to evolve over 500 generations to optimize the weights.<sup>3</sup> To produce the actual ensemble predictions, the constituent models were then re-trained using the full training data, i.e. both the training data and the calibration data, and forecasts into the test periods were made by weighting the 15 models by the genetically optimized weights for each of the ensembles and steps 1–7.

### **Performance Metrics and Benchmark Comparison**

The weighted ensembles of hidden and observed Markov models are evaluated against two different performance metrics. First, the ensembles are evaluated against the *mean squared error* (MSE). MSE may, however, be a problematic metric when the outcome has a high proportion of identical values, such as zeroes, as models which only predict these values may receive a low MSE without necessarily being able to make good predictions outside this singular value. Because of these limitations with the MSE, the ensembles are also evaluated on the *Targeted Absolute Distance with Direction Augmentation* (TADDA) score, which has been developed by the ViEWS team as an alternative evaluation metric when forecasting political violence. The TADDA score improves on the MSE by taking into account the magnitude and signs of the predicted and observed outcomes. For technical details on the TADDA score, see Vesco, Hegre, and Colaresi (2022).

The two weighted ensembles of hidden and observed Markov models are also compared to the benchmark model of the prediction competition. The benchmark model on the *country-month* level is a random forest model with more than 400 features from the ViEWS database, trained on the global level and predicting on Africa. More details about the benchmark model can be found in Hegre, Vesco, and Colaresi (2022) and Vesco, Hegre, and Colaresi (2022).

## **Results**

The results with regards to the predictive performance of the observed and hidden Markov ensembles can be seen in Table 1 below. These results show that both Markov ensembles outperform the benchmark model for

---

<sup>3</sup>The number of initial weights was selected to produce a reasonable initial coverage of the parameter space to evolve from, and the number of generations were set to be sufficiently large for the algorithm to converge. In practice, the genetic algorithm converged much faster than the allowed 500 generations for every ensemble. A larger set of initial weights was also tried, which did not meaningfully change the estimated weights but which increased the computational time substantially.

**Table 1.** TADDA and MSE scores for the weighted ensembles in the test period January 2017–December 2019.

Step	TADDA			MSE		
	OMM	HMM	Benchmark	OMM	HMM	Benchmark
1	0.344	0.391		0.545	0.570	
2	0.362	0.427	0.535	0.605	0.628	0.736
3	0.374	0.465	0.520	0.640	0.700	0.727
4	0.396	0.486	0.501	0.688	0.754	0.689
5	0.394	0.496	0.554	0.694	0.801	0.793
6	0.406	0.503	0.520	0.720	0.814	0.748
7	0.406	0.511	0.533	0.730	0.827	0.788

the competition in the test period January 2017 to December 2019 on the TADDA scores for all of the time steps. With regards to MSE, the observed Markov ensemble outperforms the benchmark on all time steps, and HMM performs approximately on par with the benchmark model. As the genetic algorithm was optimized on the TADDA score, it is not unexpected that the Markov models show better performance relative to the benchmark model on TADDA than on the MSE.

Tables 8–13 in the [Supplementary Appendix](#) show the individual model weights in the ensemble for the observed and hidden Markov ensembles in each of the three sets of forecasts. These weights show that it is more or less exclusively the conflict history models, *cfshort*, *confhist\_2019*, and *neibhist* models that seem to contribute to the ensembles. This is not unexpected as these conflict history models should reasonably be better at picking up a signal for escalation and de-escalation of conflicts than models focusing on more static or slow-moving economic and social factors, such as the *demog*, *econ*, *inst*, and *vdem* models.<sup>4</sup>

As the observed Markov ensemble is the best performing of our three approaches we submit this specification to the ViEWS prediction competition. Section *Forecasts Until March 2021*, discussing specific forecasts until March 2021, is therefore only based on the results from the observed Markov ensemble.

### ***Comparison of the Observed and Hidden Markov Ensembles***

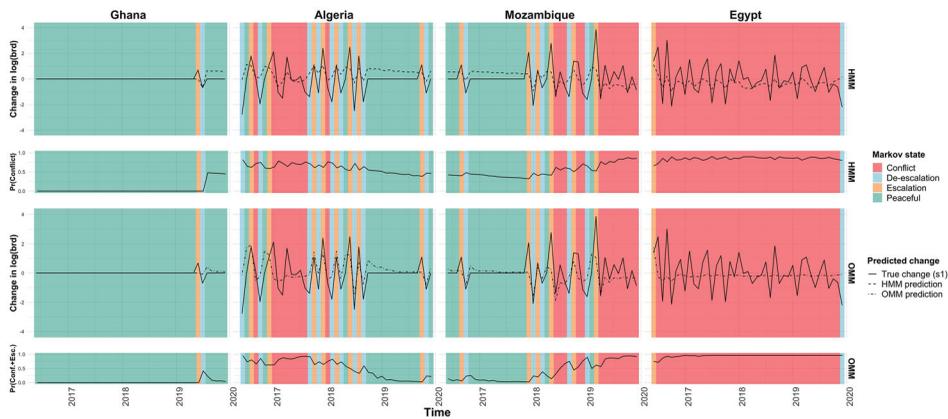
The results in [Table 1](#) above clearly show that the observed Markov ensemble outperforms the hidden Markov ensemble. This result may seem surprising, as we previously argued that we believe that the true process which generates a change in fatalities from conflicts closely resemble a hidden Markov process. To bring some clarity into why the OMM outperforms the HMM, we decided to compare the performance of the two models under the different observed Markov states. [Figure 2](#) below show the

<sup>4</sup>For more information on the constituent models, see Hegre et al. (2021), and Appendix B in particular.

predicted and actual change of fatalities ( $s = 1$ ) for four selected countries in the test period 2017–2019, as well as the observed Markov states and the estimated state probabilities for the OMM and HMM.

While this figure only shows the results from four of the countries on which the models are evaluated, it highlights some interesting differences between the models. In general, there seems to be an agreement between the OMM and HMM which state is dominating. However, the HMM seems to react much slower to the change between the true states compared to the OMM where transitions in the truly observed state yield a faster response in the probabilities of the conflict and escalation states. This discrepancy between the HMM and the OMM is likely to cause predictions with higher variability in the HMM as there is more uncertainty involved with regard to the state the process is in. This can, for instance, be seen in the case of Algeria which after a period of on-and-off violence returned to a more stable peaceful state in mid-2018. In this case, the probability of a return to conflict decreased much more rapidly in the OMM than the HMM. This pattern is also evident in Mozambique after a single violent month in 2016, and in Ghana after a single violent month in 2019. We believe that the more rapid reaction in predictions from the OMM compared to the HMM following changes in the true observed state is due to the sharper transition boundaries provided by the structure imposed by the OMM.

Table 2 below shows the out-of-sample performance of the observed and hidden Markov ensembles aggregated to the differences observed Markov



**Figure 2.** Predicted and actual change in  $\log(brd)$   $s = 1$ , observed Markov state, and estimated state probabilities for Ghana, Algeria, Mozambique, and Egypt in the out-of-sample test period January 2017–December 2019. The two upper panes refer to the Hidden Markov Model, and the two lower panes to the Observed Markov Model. For clarity, the probabilities of states “Conflict” and “Escalation” are summed for the OMM.

**Table 2.** TADDA and MSE scores for the weighted ensembles given the true observed Markov state in the test period January 2017–December 2019 predicting  $s = 1$ .

Observed state	TADDA		MSE	
	HMM	OMM	HMM	OMM
Conflict	1.198	1.367	0.878	0.895
Escalation	2.927	2.865	1.410	1.389
De-escalation	3.362	2.445	1.838	1.233
Peace	0.058	0.044	0.118	0.069

states during the test period 2017–2019 ( $s = 1$ ). This table shows that the OMM outperforms the HMM on both TADDA and MSE in all observed states except the conflict state where the HMM is slightly better. This finding is interesting as we had expected the OMM to perform better than HMM in the conflict state but worse in the escalation and de-escalation states, as one of the reasons for introducing the escalation and de-escalation states was to capture the most unstable time periods to better capture the “peace” and “conflict” states. The fact that the OMM is substantively better at predicting the de-escalation and peace states is likely due to the faster reaction time of the OMM with regards to the transition between states, which allows the model to more aptly capture such changes. This pattern is also evident in [Table 3](#) which shows the predicted state probabilities from the OMM and HMM given the truly observed states. This table shows that for both the conflict and peace states, the OMM offers higher estimated probabilities of that state compared to the HMM. This may matter less for the conflict state, where an additional prediction of the actual change in  $\log(\text{Brd})$  still needs to be done, it may matter more for the performance in the peace state where the change in  $\log(\text{Brd})$  is known to be zero.

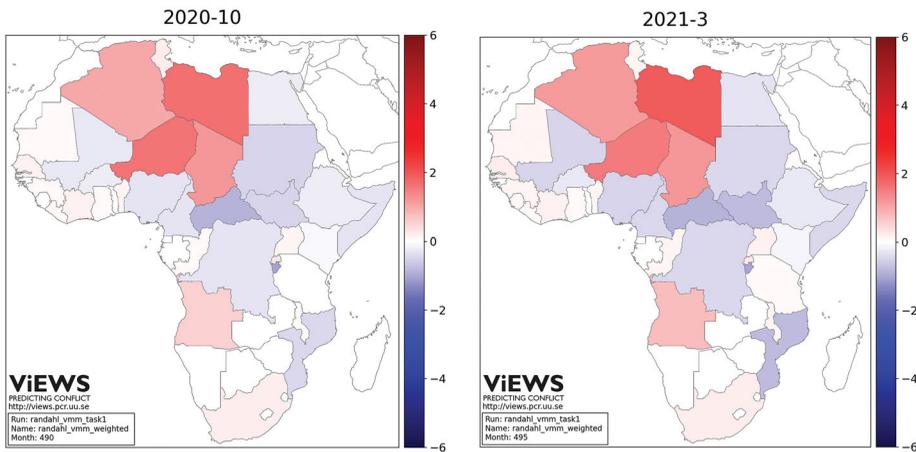
Taken together, this comparison of the HMM and OMM shows that as we had expected, the structure provided by the artificially imposed states in the OMM likely contributes to a more stable estimation of the state probabilities. Our expectation that the OMM would perform better than the HMM in the conflict state due to the exclusion of the more difficult escalation and de-escalation states, was, however, not met.

### **Forecasts Until March 2021**

As part of the ViEWS prediction competition, we provided forecasts for Africa based on data up until August 2020 and 7 months forward until March 2021. [Figure 3](#) below shows the forecasted change in  $\log(\text{brd})$  for all countries from August 2020 to October 2020 and March 2021, respectively using the observed Markov ensemble. The map shows that we forecast increased violence in a cluster of countries in the Sahara and Sahel region of Africa, primarily in Libya and Niger. All four countries in this cluster

**Table 3.** Predicted state probabilities for the weighted ensembles given the true observed Markov state in the test period January 2017–December 2019 predicting  $s = 1$ .

Observed state	Pr(Conflict)		Pr(Esc.) OMM	Pr(De-esc.) OMM	Pr(Peace) OMM
	HMM	OMM			
Conflict	0.810	0.886	0.000	0.114	0.000
Escalation	0.403	0.000	0.315	0.000	0.685
De-escalation	0.532	0.561	0.000	0.439	0.000
Peace	0.085	0.000	0.033	0.000	0.967

**Figure 3.** Forecasted change in  $\log brd$ . On the left, forecasted change in  $\log brd$  between August 2020 and October 2020. On the right, forecasted change in  $\log brd$  between August 2020 and March 2021.

are countries which in August 2020 were in the “de-escalation” Markov state in the observed Markov specification, which means that they are countries that observed 0 battle-related deaths in that specific month, but that they had observed more than zero battle-related deaths in July 2020. The observed Markov ensemble then predicts that it is likely that these countries will return to a state of conflict, or transition back and forth between the “conflict,” “escalation,” and “de-escalation” states in the coming 7 months. As the constituent model with the largest weight in the observed Markov ensemble is the conflict history 2019 constituent model, this effect of conflict history is amplified.

Among positive cases, the model predicts that several countries which were in the observed Markov state “conflict” in August 2020 state are likely to return to the “peaceful” state in the coming seven months. Countries expected to follow this pattern include Egypt, Cameroon, and Mozambique, whose individual predicted change in  $\log brd$  over the entire forecast period can be seen in Figure 4 below. The forecasted decrease in  $\log(brd)$  over the period can therefore primarily be attributed to a forecasted likelihood of a

transition in Markov states, from “conflict” to “peaceful” between the two time periods.

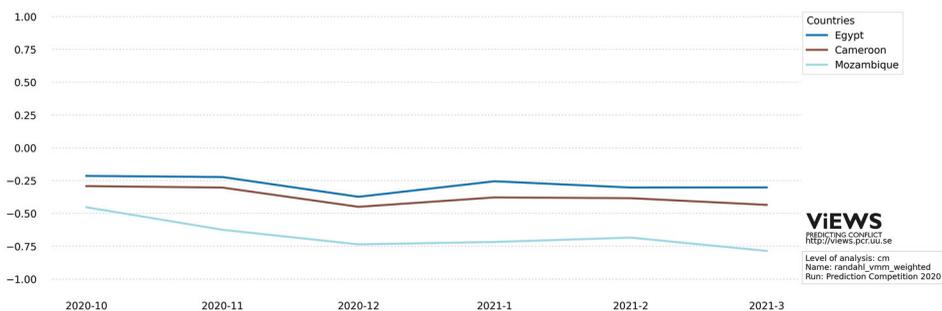
In addition to the countries which are predicted to move from the “conflict” state to a peaceful state, there are also several countries that escalated from the “peaceful” to the “escalation” state between July and August 2020 are expected to return to their peaceful state. Among these countries are Burundi and the Central African Republic.

## Extensions

The framework we have proposed in this submission to the ViEWS predictions competition seems to yield good results compared to the prediction competition’s benchmark models. Yet, there are several extensions that would be easy to implement that could improve the performance further.

In Markov models, the distribution of the state at the next time point is given by the current state. This is a strong assumption which can be relaxed with the more flexible conditional random field models (Lafferty, McCallum, and Pereira 2001; Wallach 2004; and Quattoni et al. 2007), where all pairs of states are assumed to be independent conditioned on the states in between (and connected to the two states) which could potentially yield similarly good results. Investigating the effectiveness of introducing conditional random fields is left as a future project.

Other extensions are possible to integrate using the Markov modeling framework used in this submission. For instance, the estimation of the observed Markov model reduces to a set of classification and regression problems. In this submission, we have chosen to use binary and multinomial logistic regression to estimate the transition between the states in the Markov chain, and simple OLS to calculate the expected change in the natural logarithm of fatalities. However, the choice of GLM-models for this modeling was simply a convenience choice, and it would be entirely possible to use for instance random



**Figure 4** Predicted change in  $\log(\text{brd})$  for Egypt, Cameroon, and Mozambique from August 2020 until March 2021.

forests or other classifiers to make the predictions of both the transition between states and the change in fatalities given the transition between states.

Additionally, we have simply chosen to include a set of 15 constituent models from the ViEWS country-month specification in our ensembles, but these constituent models may not be the optimal models for the prediction problem. Changing the composition of these ensembles could therefore also yield better performance. Similarly, the weighting of the ensemble has in this contribution been based on the predictive performance with regards to the prediction target. An alternative specification could, for the observed Markov ensemble, base the weights on the constituent models' ability to predict the artificial Markov states. Another possibility would be to vary the weights based on the constituent models' ability to predict the forecasting target conditional on the Markov states and thus allow different constituent models to have different weights conditional on the state. Last but not least, we have chosen to make this submission on the country-month level, but this same methodology could easily be adapted to the prio-grid month level as well.

## Funding

This research note is a submission for the 2020 ViEWS prediction competition. The research was funded by the European Research Council, project H2020-ERC-2015-AdG 694640 (ViEWS) and the Swedish Research Council, grant number 2017-01175. For more information on the project see <http://www.pcr.uu.se/research/views/>.

## ORCID

David Randahl  <http://orcid.org/0000-0003-1069-6067>

## References

- Brandt, Patrick T., John R. Freeman, Tse-Min Li, and Phillip A. Schrodt. 2012. "A Bayesian Time Series Approach to the Comparison of Conflict Dynamics." Paper Presented at the 2012 Annual Meeting of the American Political Science Association, New Orleans, August 31-September 2.
- Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1): 1–22.
- Diehl, Paul F. 2006. "Just a Phase? Integrating Conflict Dynamics over Time." *Conflict Management and Peace Science* 23 (3): 199–210. doi:10.1080/07388940600837490.
- Hegre, Håvard, Curtis Bell, Michael Colaresi, Mihai Croicu, Frederick Hoyles, Remco Jansen, Maxine Ria Leis, et al. 2021. "ViEWS2020: Revising and Evaluating the ViEWS Political Violence Early-Warning System." *Journal of Peace Research* 58 (3): 599–611. doi:10.1177/0022343320962157.

- Hegre, Håvard, Paola Vesco, and Michael Colaresi. 2022. "Lessons from an Escalation Prediction Competition." *International Interactions* 48 (4).
- Jackson, Christopher. 2011. "Multi-State Models for Panel Data: The Msm Package for R." *Journal of Statistical Software* 38 (8): 1–28. doi:10.18637/jss.v038.i08.
- Krogh, Anders, Björn Larsson, Gunnar Von Heijne, and Erik L. L. Sonnhammer. 2001. "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes." *Journal of Molecular Biology* 305 (3): 567–580. doi:10.1006/jmbi.2000.4315.
- Kumar, Sandeep, Sanjay Jain, and Harish Sharma. 2018. "Genetic Algorithms." In *Advances in Swarm Intelligence for Optimizing Problems in Computer Science*, 27–52. CRC Press.
- Lafferty, John, Andrew McCallum, and Fernando C. N. Pereira. 2001. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data."
- Mitchell, Melanie. 1998. *An Introduction to Genetic Algorithms*. MIT Press.
- Quattoni, Ariadna, Sybor Wang, Louis-Phillipe Morency, Michael Collins, Trevor Darrell, and Mit Csail. 2007. "Hidden Conditional Random Fields." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (10): 1848–1852. doi:10.1109/TPAMI.2007.1124.
- Rabiner, Lawrence R. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE* 77 (2): 257–286. doi:10.1109/5.18626.
- Rabiner, Lawrence, R., and Biinghwang Juang. 1986. "An Introduction to Hidden Markov Models." *IEEE ASSP Magazine* 3 (1): 4–16. doi:10.1109/MASSP.1986.1165342.
- Schrodt, Philip A. 1997. *Early Warning of Conflict in Southern Lebanon Using Hidden Markov Models*. American Political Science Association.
- Schrodt, Philip A. 2006. "Forecasting Conflict in the Balkans Using Hidden Markov Models." In *Programming for Peace*, 161–184. Springer.
- Stultz, Collin M., James V. White, and Temple F. Smith. 1993. "Structural Analysis Based on State-Space Modeling." *Protein Science* 2 (3): 305–314. doi:10.1002/pro.5560020302.
- Vesco, Paola, Håvard Hegre, and Michael Colaresi. 2022. "Comparing and Evaluating the Contributions." *International Interactions* 48 (4).
- Wallach, Hanna M. 2004. "Conditional Random Fields: An Introduction." *Technical Reports (CIS)*, 22.
- White, James V., Collin M. Stultz, and Temple F. Smith. 1994. "Protein Classification by Stochastic Modeling and Optimal Filtering of Amino-Acid Sequences." *Mathematical Biosciences* 119 (1): 35–75. doi:10.1016/0025-5564(94)90004-3.