A shape-based approach to conflict forecasting

Thomas Chadefaux
Trinity College Dublin
Department of Political Science
chadefat@tcd.ie

October 27, 2021

Abstract

Do conflict processes exhibit repeating patterns over time? And if so, can we exploit the recurring shapes and structures of the time series to forecast the evolution of conflict? Theory has long focused on the sequence of events that precedes conflicts (e.g., escalation or brinkmanship). Yet, current empirical research is unable to represent these complex interactions unfolding over time because it attempts to match cases on the raw value of covariates, and not on their structure or shape. As a result, it cannot easily represent real-world relations which may, for example, follow a long alternation of escalation and détente, in various orders and at various speeds. Here, I aim to address these issues using recent machine-learning methods derived from pattern recognition in time series to study the dynamics of casualties in civil war processes. I find that the methods perform well on out-of-sample forecasts of the count of the number of fatalities per month from state-based conflict. In particular, our results yield Mean Squared Errors that are lower than the competition benchmark. We discuss the implication for conflict research and the importance of comparing entire sequences rather than isolated observations in time.

Do conflict processes exhibit repeating patterns over time? And if so, can we exploit the recurring shapes and structures of the time series to forecast the evolution of conflict? Theory has long focused on the sequence of events that precedes conflicts (e.g., escalation or brinkmanship). Yet, current empirical research is unable to represent these complex interactions unfolding over time because it attempts to match cases on the value of covariates, and not on their structure or shape. In other words, it cannot easily represent real-world relations which may, for example, follow a long alternation of escalation and détente, in various orders and at various speeds. Consider for example the sometimes elaborate patterns of guerilla warfare, in which rebels conduct repeated small attacks before retreating after a more or less prolonged engagement.[1] Or consider the issue of power transitions which may take place rapidly or over extended durations, making it particularly challenging to measure empirically. Deescalation is also rarely linear and can follow complex patterns of decreased violence followed by pauses and even re-escalation, as evidenced by the US's protracted retreat from Vietnam. Lags of a few orders might be included in a regression to account for issues of autocorrelation, but they cannot represent the potentially complex trajectories followed by armed conflicts and their associated casualties.

Here, I aim to address these issues using recent machine-learning methods derived from pattern recognition in time series to study the dynamics of casualties in civil war processes. The novelty is to move away from the current reliance in the social sciences on covariance structures of the raw signals and to supplement these typical approaches with methods to extract shapes and better understand the patterns of escalation into violence.

Methodologically, existing approaches in the social sciences typically rely on the relationship between individual observations (e.g., country-year). They do so by matching a given situation with another that shares similar covariate values. Researchers may, for example, find that countries with a combination of high ethnic fractionalization, low GDP, and weak

[1] **?**, for example, describe typical ambush attack by the the Viet Cong.

institutions are particularly prone to civil war onset. Most of the recent work on conflict forecasting relies on these methods (**?**, **?**, **?**, **?**)

However, this 'matching' on the value of covariates does not make use of time dependence in the sequence. These approaches are unable to incorporate the more complex dynamics of escalation that typically arise (**?**, **?**). This is unfortunate, as most real-world interactions cannot be understood simply as a function of the current state of a variable, or even of many variables and their interactions. Modeling the type of complex back and forth, up and down, aggression–appeasement–escalation patterns that are common in civil conflicts is difficult.

Some methods (e.g., ARIMA) may include additional components aimed at modeling temporal dynamics. Lags, first differences, or regression splines, for example, may account for time dependence. However, even the most complex lag structure can miss important temporal patterns, for two main reasons. First, lag structures become increasingly complex as the number of lags increases. Such models can quickly become unwieldy in highly nonlinear patterns—ones we expect to be the norm rather than the exception in complex strategic interactions such as the ones preceding conflict. Second, even assuming that we knew the correct autoregressive structure, these models are unable to deal with accelerations or slowdowns. The same type of events can take place at different speeds. For example, a pattern of escalation may take place over a few months or an entire year, even though they are generated by the same underlying process. Lag models will still attempt to match observations one-to-one—e.g., the first lag of the dependent variable for Burundi with the first lag for Rwanda, even though the relevant comparison may be the first lag for Rwanda and the third for Burundi.

This inability to incorporate complex nonlinear temporal patterns may explain why, in existing approaches, the lagged dependent variable is almost always the best predictor—the immediate past best explains the current state of the system (**?**, **?**). These challenges are likely to be even more pronounced in the context of conflict dynamics, where the static value

of covariates is unlikely to be enough to predict the evolution of conflict.

Instead, I use here entire sequences of events—and not single observations—as units of analysis. This makes it possible to extract potentially complex motifs and highly nonlinear patterns, as opposed to only basic trends or static values. It also makes it possible to address potentially distorted and time-warped patterns.

[Figure 1 about here.]

To better understand the importance of studying sequences as a whole, rather than as a set of more or less independent observations, consider a simple illustrative example. Suppose that we observe a conflict-related process, such as the number of casualties in a particular unit-month (top plot in fig. 1; vertical dotted lines denote the timing of a particular event to be predicted—e.g. the change in the number of deaths). We are then interested in finding similar patterns in the future, with the expectation that the similarities will help us anticipate the likely outcome of that sequence of events.

Unfortunately, a correlation-based approach—whether it be a regression, random forest, neural network, etc.—applied to the data points is unable to detect any pattern here. This is regardless of the number of lags, first differences, splines, or "time since the last event" variables we include (**?**). The problem is not the model itself, but rather that we do not treat sequences as the unit of observation.

If instead we decompose the series into sequences, we then realize that the same pattern repeats three times, albeit over different durations (second row). Isolating each one, it then becomes clear that they are all noisy versions of a general *motif*, of a form that could be summarized as 'up 1, down 1/2, up 1'.

Moreover, a correlation-based approach would fail here for another reason, because it requires the same number of observations in each sequence. This is unfortunate, as the same underlying conflict dynamic may take place over different time spans—e.g., 3, 6, 12

months—and yet share a general underlying pattern. That pattern would be missed using standard approaches.

Methods

In line with the ViEWS competition's guidelines (see introductory article to this special issue), our goal is to "predict substantial change, upwards or downwards, in the count of the number of fatalities per month from state-based conflict—as defined by the Uppsala Conflict Data Program [...]—and within the parameters set by the ViEWS project, over a subset of the forecasting horizon." More specifically, the task is to correctly predict the change in the log of the number of casualties from a period t to another period $t + s$. An 'observation' therefore refers to the value of $\Delta_s \ln(Y_{i,t} + 1))$, where Y is the number of fatalities recorded in the UCDP dataset and aggregated to the ViEWS units of analysis. Here we make forecasts for both country- and PRIO-grid months.[2] Extending the analysis to the PRIO-grid level has two important advantages: it allows us to analyze subnational patterns, and it also allows us to avoid the constraints of administrative boundaries when the operations of many violent actors are in fact cross-national in nature.

Comparing Sequences

Each unit (i.e., either country or PRIO-grid) has a total of 378 months of observations from 1989 to 2020. The first step is to split these 378 observations into sequences of one year (i.e., 12 observations per sequence).[3] A particular sequence is denoted by $s_{i,t}$, where i refers to the unit (country or PRIO-grid) from which this sequence was extracted, and t to the last month of the 12-month sequence. For example, $s_{\text{Nigeria},380}$ refers to the set of observations of

[2]Details of the competition are not reproduced in this short note but are available in the introductory article.

[3]Admittedly, the choice of a sequence of length 12 is arbitrary and driven by computational cost. Future work could vary this meta-parameter to analyze longer sequences.

$\Delta_s \ln(Y_{i,t}+1)$) for Nigeria that starts in Sept. 2010 (month 369) and ends in August 2011 (month 380). That particular sequence is displayed in fig. **??**.

The next step is to compare these 12-month sequences to each other. Intuitively, sequences that are similar (discussed below) are expected to lead to the same outcome (escalation/de-escalation), and dissimilar ones to different outcomes (see eqn. 1). Practically, each sequence is first z-normalized and then compared to other sequences from the past and other countries.[4] More specifically, a distance metric is calculated between each pair of sequences $s_{i,t}$ and $s_{j\neq i,u<t}$ occurring at times t and u respectively.[5] For example, the 12-month sequence for Nigeria's casualties discussed above is compared to the 12-month sequence for Chad from Sept. 2009 to August 2010, as well as to the 12-month sequence for Tunisia from Jan. 1995 to December 1995, and so on.

Dynamic Time Warping

To compare sequences, I rely on Dynamic Time Warping (**?**), which is an algorithm that measures the similarity between two temporal sequences that may vary in speed. These sequences may have approximately the same shape, but do not line up on the X-axis. To find their similarity, we will therefore need to first preprocess them by 'warping' the time axis of either (or both) sequences in order to better align them.

This is particularly appropriate for the problem of escalation and the ups and downs of conflict, as these may take place over shorter or longer periods but still present the same pattern. The basic idea is to match two series by allowing one observation in series i to be matched with several in series j. Series are shifted, scaled, or stretched in a way that

[4] I discuss below what is meant by "compared". Normalizing is important to avoid matches solely due to similar raw values, and also because it is known empirically to lead to fewer classification errors (**?**).

[5] We only compare to past observations for obvious reasons. We also exclude observations from the source country to avoid any possible contamination (e.g., sequences $s_{\text{Nigeria},380}$ and $s_{\text{Nigeria},381}$ overlap over 11 months.)

minimizes their distance (see fig. **??** for illustration[6]). This type of problem cannot easily be addressed by standard approaches to time series, as they fail to detect patterns that do not match one-to-one.

More precisely, consider two sequences: a query $X = (x_1, \ldots, x_N)$ and a reference $Y = (y_1, \ldots, y_N)$. We calculate a *warping curve* $\phi(k), k = 1, \ldots, T$ as:

$$\phi(k) = (\phi_x(k), \phi_y(k)) \tag{1}$$

Intuitively, the warping functions $\phi_x(k)$ and $\phi_y(k)$ form a new mapping of the indices of X and Y—one that allows a distortion in the time axes. Based on that mapping, an average accumulated distortion is calculated between the two (warped) series:

$$d_\phi^w(X, Y) = \sum_{k=1}^{T} d\left(\phi_x(k), \phi_y(k)\right) m_\phi(k)/M_\phi,$$

where d is the Euclidean distance, $m_\phi(k)$ is a (user-defined) per-step weighting coefficient and M_ϕ is a normalization constant to allow for comparisons. In short, the Euclidean distance is calculated between the two realigned series, as opposed to the two original series.

Consider for example the initial values of the following two sequences (the full sequences are displayed in fig. **??**):

$$s_{\text{Nigeria},380} = (0.51, 0.34, -1.95, 1.95, \ldots)$$
$$s_{\text{Pakistan},197} = (0.28, -1.63, -2.08, 1.10, \ldots)$$

Note that the first observation in both sequences is similar, with a (Euclidean) distance of only 0.23 between them. On the other hand, the distance between the second observations is large ($d(0.34, 1.63) \approx 2$). However, the second observation of Nigeria and the first of

[6]Illustration by the Wikimedia Commons, distributed under a CC-BY-SA-4.0 license.

Pakistan are still close to one another, and so will be paired. In turn, the third observation for Nigeria is close to the second and third observation for Pakistan, and so are paired. We continue in this manner until we reach the end of both series.

Figures **??**–**??** illustrate the process visually. In fig. **??**, we represent the warping path between the sequence $s_{\text{Nigeria},380}$ and its closest match. The red dotted lines indicate which observations are matched together. We see, as discussed above, that the first and second observations for Nigeria are paired with the first observation for Pakistan, and so on. The cost matrix, displayed in fig. **??**, shows the pairwise distance between each observation of the two sequences (e.g., 0.2 at the bottom corner corresponds to the distance between 0.51 and 0.28). The warping path is then obtained by aligning each sequence's indices in a way that minimizes the cumulative distance between the two sequences.[7]

[Figure 2 about here.]

This process of comparing the distance between two warped series is repeated for every pair of sequences in our data. This yields a distance d_{ij}^w between i and j, which is a measure of how dissimilar two sequences are after aligning them using the DTW algorithm.[8] I now discuss how these distance measures d_{ij}^w are combined to generate a forecast.

From similarity to predictions

It will be convenient to define $f_{i,t}$ as the 'future' of sequence $s_{i,t}$. For example, $f_{i,t}$ for $s_{\text{Nigeria},380}$ is the observation of the change in the number of fatalities in Nigeria from month 380 to month 381 (i.e., what we observe in month 381). Then the uncalibrated predicted change in fatalities for sequence $s_{i,t}$, which we denote by $\hat{f}_{i,t}$, is simply calculated as the average of the futures of past sequences, weighted by their similarity (i.e., the inverse of

[7] A good review of DTW is **?**.

[8] For more on the DTW algorithm, see (Berndt & Clifford, 1994).

their distance):

$$\hat{f}_{i,t} = \frac{1}{N} \sum_{j \neq i, t_j < t_i} \left(\frac{1}{d_{ij}^w} f_{j,t_j} \right) \tag{2}$$

Intuitively, $\hat{f}_{i,t}$ is obtained by comparing sequence $s_{i,t}$ to past sequences $s_{j,t_j<t_i}$, and forecasting the future of $s_{i,t}$ as a weighted average of the future of those comparison sequences. Past sequences $s_{j,\cdot}$ that are similar (i.e., those with a small d_{ij}^w) are expected to lead to similar outcomes and hence are assigned a large weight, whereas dissimilar sequences (large d_{ij}^w) receive a low weight (i.e., $1/d_{ij}^w$ is small).

Figure **??** illustrates the process of estimating a forecast for Egypt, Mozambique, and Cameroon (country-level). As described above, the source data (in black) is compared with past sequences. Sequences that are similar—i.e., those with a low time warped distance—are assigned a large weight (in black), whereas dissimilar ones receive a low weight. Predictions for October 2020, for example, are a weighted average of the first "future" point in the dotted line on the red graphs (as well as those of hundreds of other sequences that cannot be included in the figure).

[Figure 3 about here.]

Finally, we obtain a calibrated version of $\hat{f}_{i,t}$ by regressing by OLS $f_{i,t}$ on $\hat{f}_{i,t}$ for all $t < 2014$. Using the estimated coefficients, we then calculate fitted values for $t \in [2014, 2016]$, and so on for all relevant forecast sets.

Results

By applying these methods to all sequences, we can make true forecasts for the period that immediately follows the competition (i.e., October 2020–March 2021), as well as 'backward'

looking forecasts for the 2017–19 sample (i.e., using information until, for example, Dec. 2016 to predict Jan. 2017).

We predict a small overall increase in the number of casualties over the period Oct. 2020–March 2021 (fig. **??**). More deaths are expected in November and December than in February or March, and we expect a rise in the number of deaths in Mozambique, but a decline in Egypt and Cameroon (fig. **??**). Maps of our predictions broken down by country and PRIO-grid, are also displayed in figure **??**. For October 2020, for example, the largest expected increase is in Benin and the largest decrease in Egypt.

[Figure 4 about here.]

[Figure 5 about here.]

To estimate the quality of our forecasts and compare them to other approaches, we also calculated the mean-squared error (MSE) for both country-month and PRIO-grid month forecasts over the period 2017–19 (Africa only). We report results for three models for each step.[9] The three models are: (i) our DTW model described above (also referred to as 'shape'); (ii) a simple constant model, where $f_{i,t}$ is simply estimated to be a constant (equal to the past average change in casualties); and (iii) the benchmark model reported in the ViEWS paper.

Overall, the results are encouraging. Qualitatively, we find that 83% of the actual outcomes fall within our 25%–75% prediction range (obtained by bootstrapping). In other words, there is an 83% chance that the outcome will indeed be within the given range. More generally, the MSE for the Dynamic Time Warping model significantly improves upon both the benchmark and the constant models (Fig. **??**). This is true for all values of 'step', and at both country-month and PRIO-grid month level. Similarly, we also find that the Pseudo-Earth Mover Divergence score for our approach significantly improves upon the

[9] As a reminder, 'step 2' means for example using data up to January to forecast March

benchmark (Appendix table **??**).[10] The Targeted Absolute Distance With Direction Augmentation (TADDA) score also significantly improves upon the benchmark, but is worse than the 'constant' model (Appendix Fig. **??**). This is likely due to the fact that the constant model is very conservative in its estimate (i.e., it always forecasts a change close to zero), and therefore never crosses the threshold needed to trigger any penalty (see introductory chapter for details on the TADDA score). Finally, for illustration purposes, we also report in the appendix the prototypes of the most common shapes observed in the overall data and in the observations with the highest expected (de-)escalation (Appendix Fig. **??**).

[Figure 6 about here.]

Discussion

The results are overall encouraging and underline the importance of incorporating more complex temporal dynamics into existing forecasting efforts. Methods that allow temporal matching on the basis of distorted or warped sequences offer a flexible approach to time series that may allow to uncover dynamics that would otherwise remain undetected.

In addition, the current results arise only from the use of the shape-based approach. The predictive gains are achieved without adding new information or data to the model—simply by using fluctuations in the dependent variable itself. Combining the predictions obtained here with models with additional covariates would further improves the result. One way to do this is to incorporate the predictions made here into an ensemble model of other predictions. The results of so doing are reported in the introductory chapter of this issue. They show that the forecasts that arise from the present shape-based approach do contribute to improve upon other methods.

[10] pEMDiv scores were obtained from the work of the ViEWS team and are as reported in the introductory article to this journal's issue.

There are, however, limitations to the current approach. First, we have limited ourselves to the comparison of sequences of the same size (12 months). This still lets us compare dynamics that may take place at different speeds—e.g., a few months vs. an entire year. However, it does not allow us to compare potentially much slower processes (e.g., taking place over 2 years) that may exhibit the same pattern. Ideally, we would therefore first vary the 12-month window. Furthermore, we could compare windows of different sizes to each other—e.g., sequences of 12 months to others of 24 or 37 months. However, one issue with varying the duration of the sequence is the computational cost. At the moment, we are already applying dynamic time warping to millions of pairs of series. One obvious solution for the future would be parallelization, to which this problem lends itself well since our unit of analysis—the sequence-pair—is small and each pair of sequences can be analyzed on a different core.

Another technical limitation of the approach taken here is our choice of the weight to apply to past sequences in determining the likely future of a sequence of interest. Sequences are weighted by their proximity to the 'query' sequences using a simple inverse function. This may result in stronger predictions being washed out by other predictions, and may explain our narrow set of predictions, which is closely centered around zero. Risky predictions are rarely made. While this leads to good results in terms of out-of-sample metrics such as the Mean Squared Error, the approach may be poor at identifying radical changes. In future work, we will therefore aim to change the weights assigned to past sequences. For example, it may be best to simply assign a weight of one to the most similar sequence, and zero to all the other sequences. This would likely result in more extreme predictions, though possibly at the cost of lower accuracy.

Finally, we plan in future work to complement this approach with others that incorporate geography more explicitly. In particular, dynamic *space* warping would allow us to compare geographic patterns by warping them to match others. Other avenue of future research

include the temporal clustering of patterns. For example, do patterns in the early 1990s differ from those of the late 2020s? Such temporal disaggregation would potentially allow us to weigh past sequences by their temporal distance, as well as to possibly understand how patterns evolve over time—possibly as a response to the systemic or structural environment, or to changes in military tactics and technology or policy preferences. Patterns may also vary not only temporally, but also by type of violence, by actor, or by tactic.[11]

Overall, the method presented here offers a number of advantages in the present case, but also for further applications. First, it is radically different in its approach to existing correlation-based approaches. As such, it is likely to add information to existing forecasts. Furthermore, it requires little in terms of data. In fact, all we use here is the variable to be predicted itself. No additional covariates are needed. This is not to say that covariates are not useful, as shown in the introductory article of this journal's issue, but rather that covariates are not *necessary*. This could prove particularly useful for policy-makers who may not have access to the right socio-economic data at the relevant level of prediction.[12] Intuitively, we all already perform similar mental forecast based on existing patterns. People on the field may 'expect' things to get worse when they recognize a familiar pattern of escalation and deescalation. The advantage of the present method is to formalize such intuitions in a way that takes advantage of more data and aggregates it more effectively.

[11] I thank an anonymous reviewer for this suggestion.
[12] I am grateful to an anonymous reviewer for this suggestion.

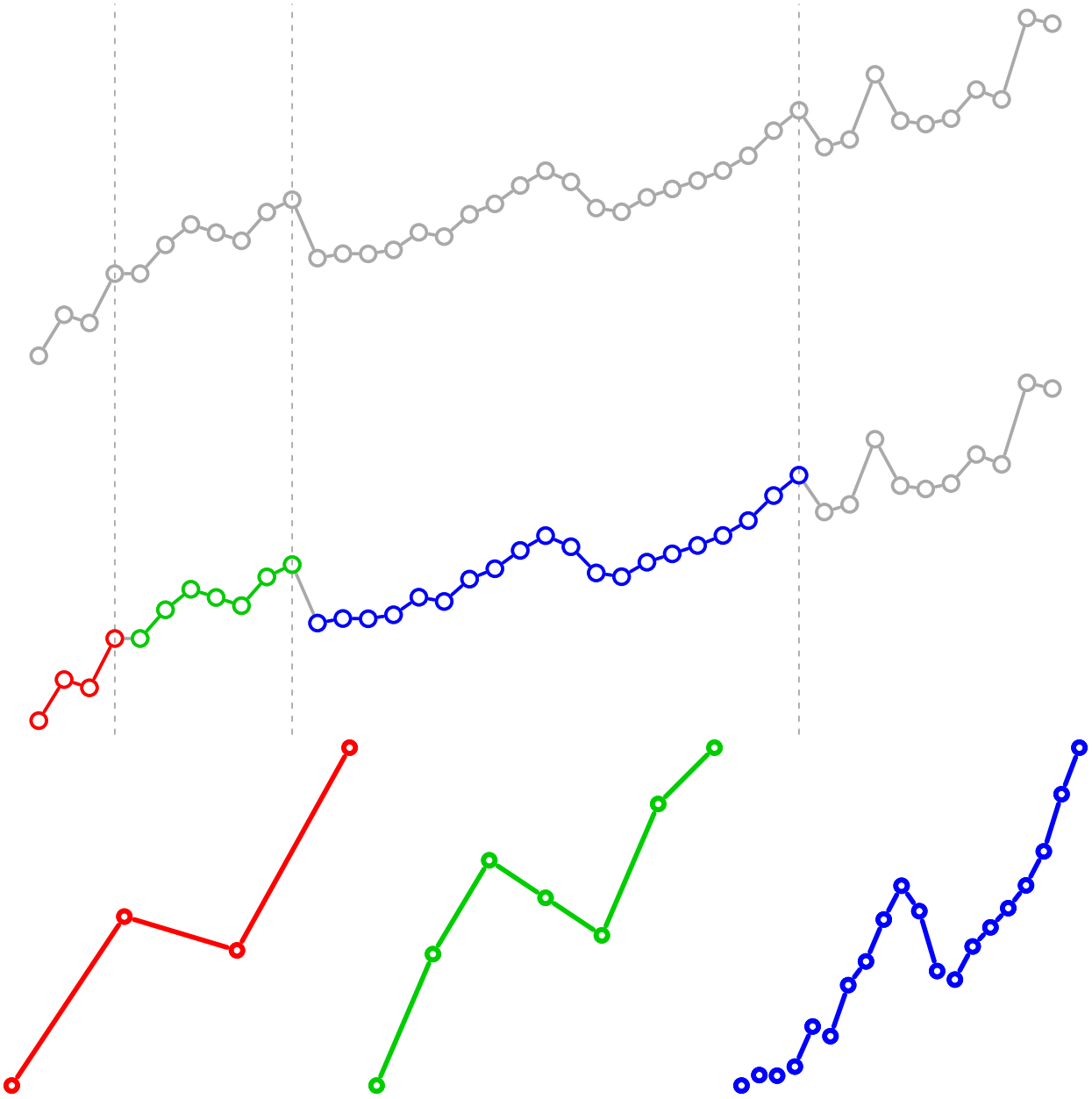
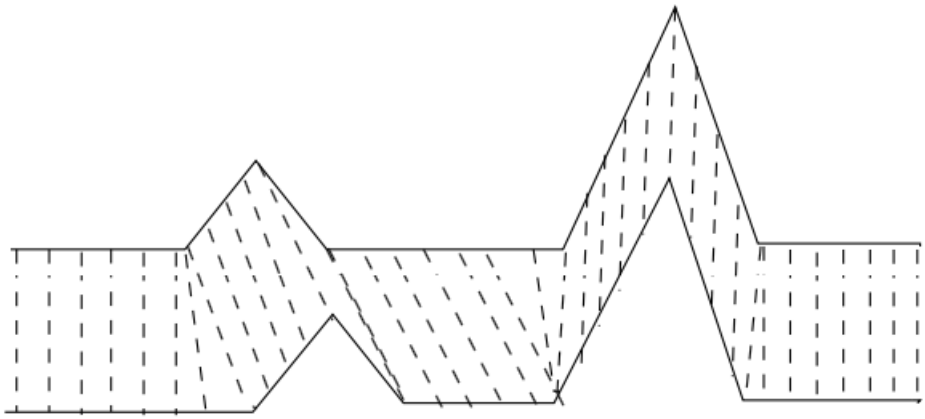
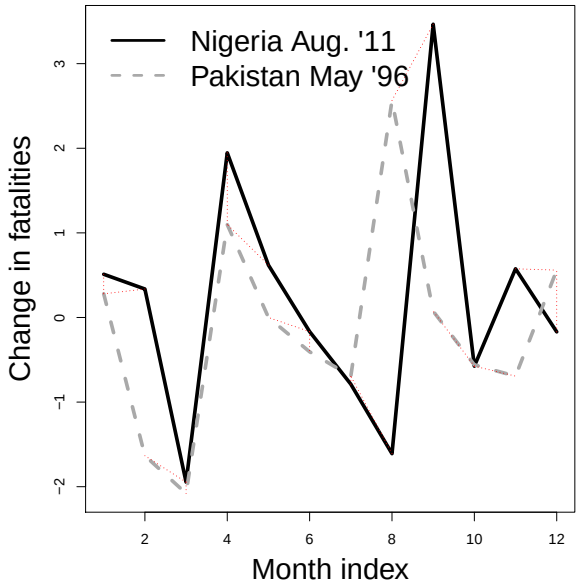


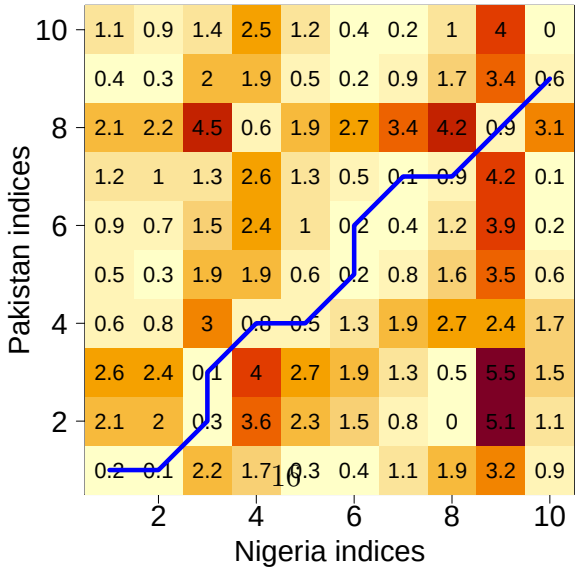
Figure 1. *One-to-one measures of distance fail to discern patterns.*



(a)


Nigeria Aug. '11
Pakistan May '96
Change in fatalities
Month index
2
4
6
8
10
12
−2
−1
0
1
2
3

(b)


Pakistan indices
Nigeria indices
2
4
6
8
10

(c)

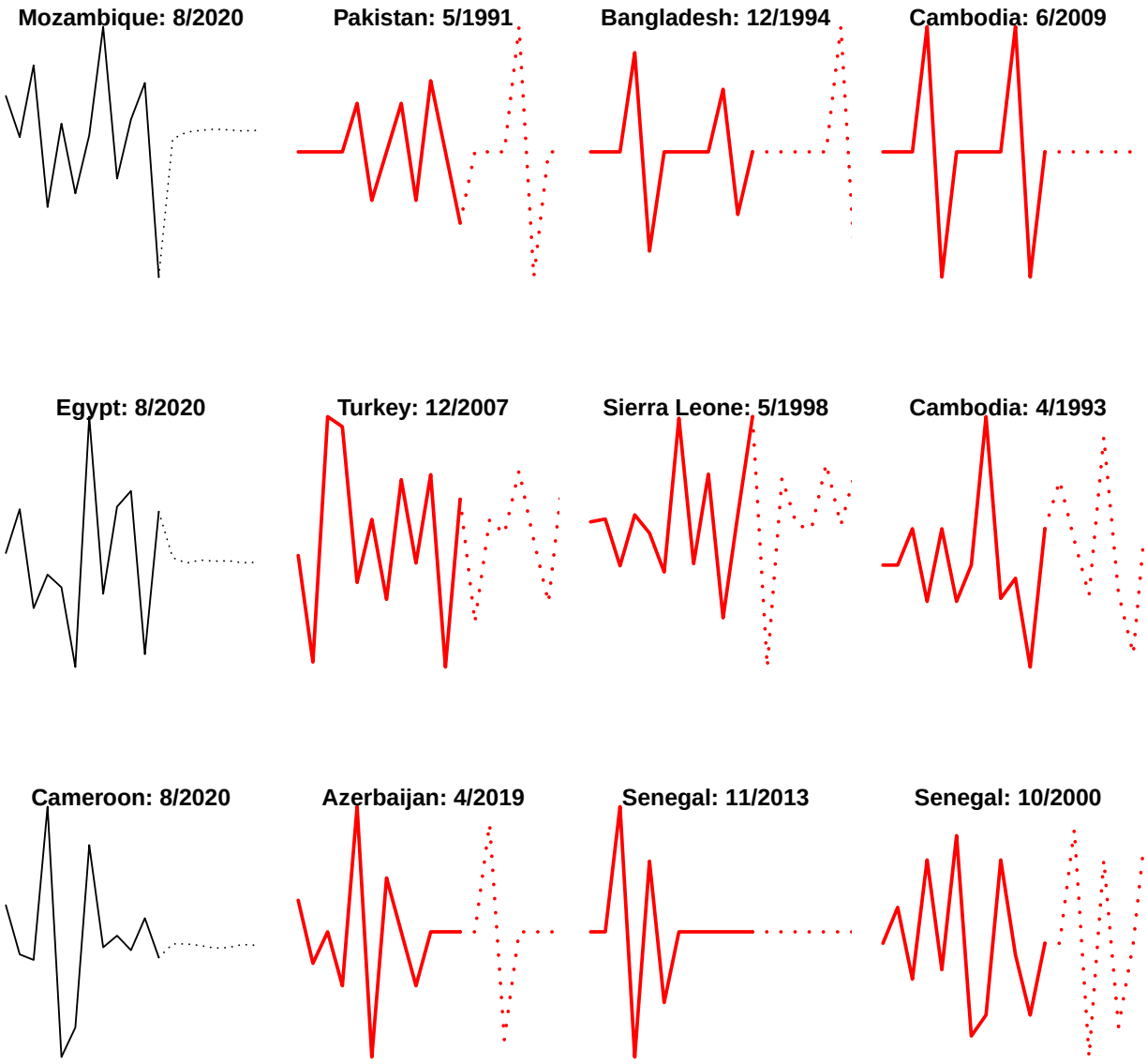
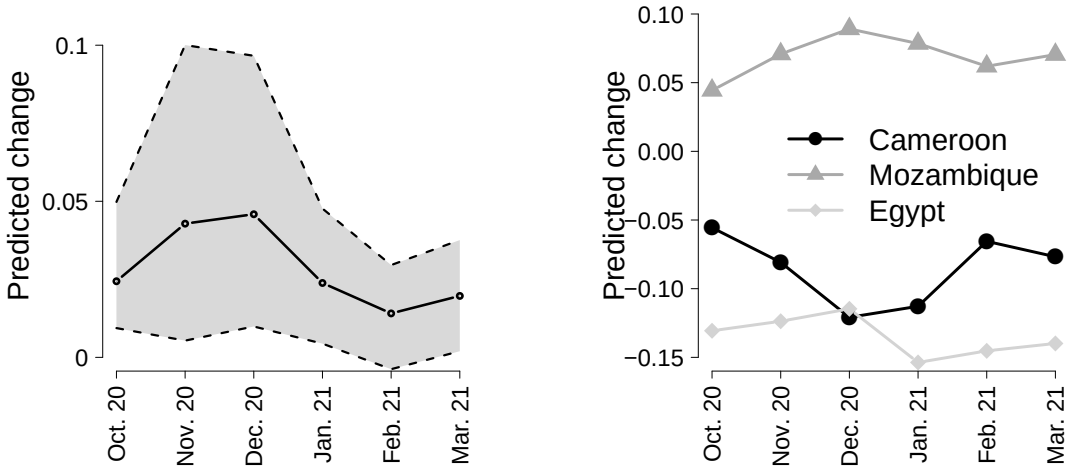


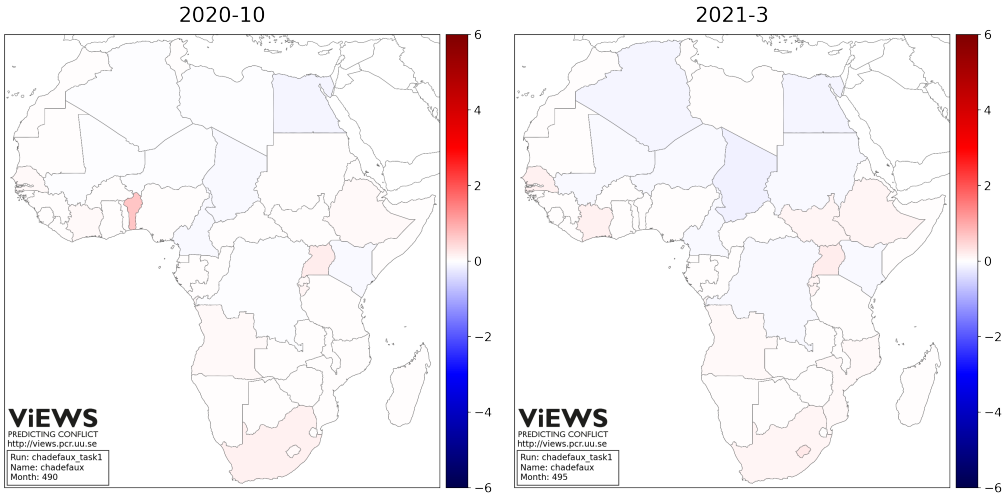
Figure 3. *Sample predictions and best matches. Source sequences (in black) are matched with past sequences. Best matches (i.e., those with small time warped distances) are displayed in red. Predicted values for Oct. 2020–March 2021 (black dashed lines) are obtained as the weighted average of the future of other past sequences (red dashed lines), weighted by their similarity.*



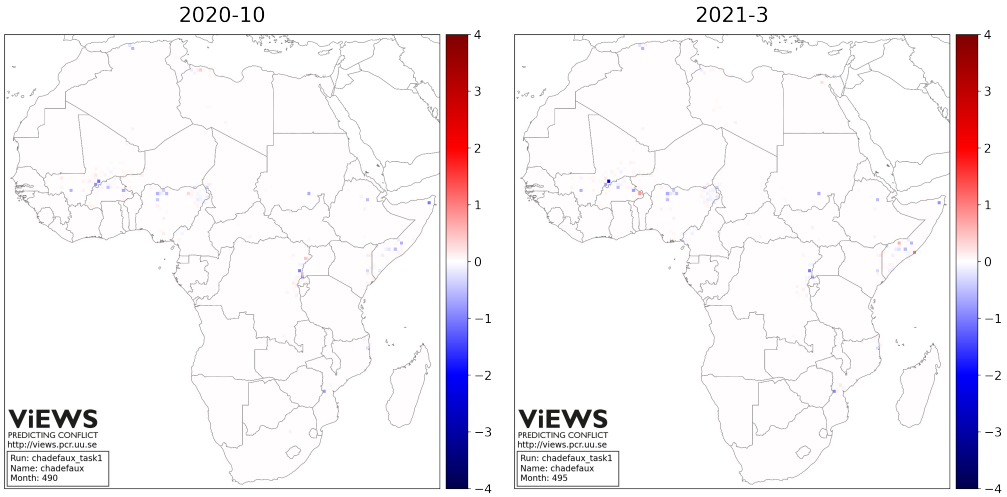
(a) *Africa (average and 95% confidence interval [bootstrapped])*

(b) *Selected country-months, Oct. 2020–March 2021*

Figure 4. *Predicted change in violence, Oct. 2020–March 2021*

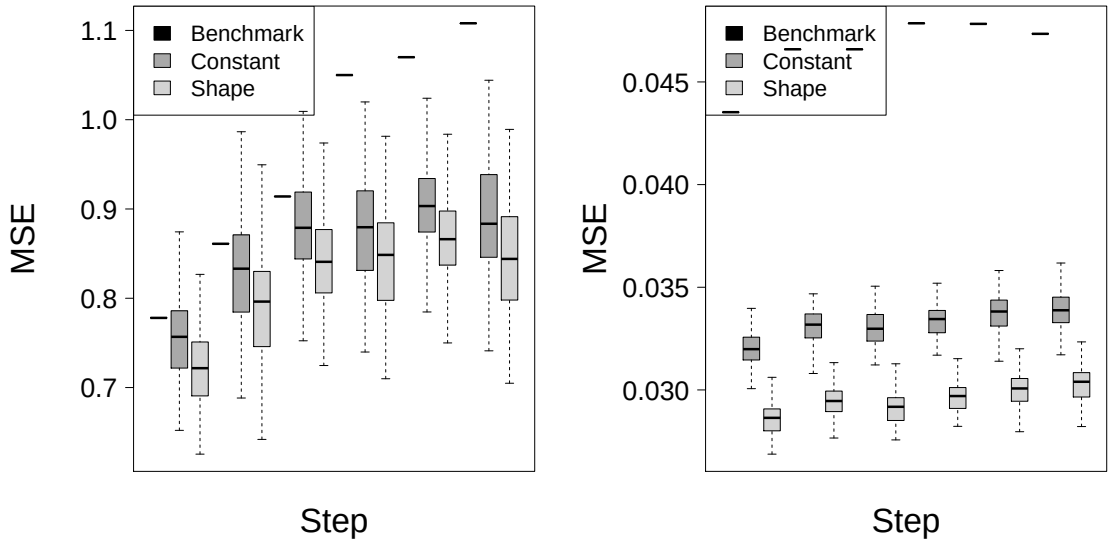


(a) *Country-month, October 2020* ***(b)*** *Country-month, March 2021*



(c) *PRIO-grid-month, October 2020* ***(d)*** *PRIO-grid-month, March 2021*

Figure 5. *Prediction maps for October 2020 and March 2021, country- and PRIO-grid months (courtesy of the ViEWS team).*



(a) *Country-month (cm) level* ***(b)*** *Prio-grid-month (pgm) level*

Figure 6. *Mean Squared Error for* country-month *and* PRIO-grid-month *predictions of the change in log fatalities (test partition: 2017-19, Africa). 'Shape' is the model based on Dynamic Time Warping; 'Constant' forecasts the past average value of casualties (i.e., regression on a constant). Benchmark estimates are represented by a single line. Lower values indicate better predictions. All pairwise differences are significant at* $p < 0.01$ *using either a t-test or Wilcoxon test.*