Echoes of Conflict
Quantifying Redundancy in Fatality Time Series

Thomas Schincariol[1*] and Thomas Chadefaux[1]

[*]Corresponding author: Thomas Schincariol, schincat@tcd.ie
[1]Department of Political Science, Trinity College Dublin

Abstract

Do wars follow predictable patterns, or are they chaotic and inherently unpredictable? Conflicts might be guided by underlying regularities shaped by political, economic, and social forces, or they could unfold as chaotic systems driven by randomness and local contingencies. We explore this question by quantifying temporal redundancy in conflict-related fatalities. Using entropy-based analysis and Dynamic Time Warping, we compare conflict dynamics with other complex systems such as seismology and finance. Our findings reveal that the dynamics of fatalities within conflicts exhibit high levels of entropy—they are largely unpredictable. Conflicts generally display higher entropy than climate or earthquake data, highlighting the challenges of predictive modeling. Moreover, conflict patterns are not region or decade-specific, which suggests a degree of universality. Our framework provides a quantitative basis for distinguishing between stable and chaotic phases in conflict dynamics, and assess the feasibility of predictive modeling in different conflict contexts.

Keywords: Temporal Redundancy; Conflict Data; Fatalities; Entropy; Temporal Patterns

The adage "History repeats itself" has long intrigued scholars and laypeople alike. It suggests that historical events and patterns tend to recur over time and that human behavior, societal trends, and global events often follow recognizable sequences. As Mark Twain famously quipped, "History doesn't repeat itself, but it often rhymes." This idea remains contentious. Critics argue that viewing history through repetitive patterns oversimplifies complex realities, as each historical moment is unique and shaped by specific contexts. Proponents, on the other hand, contend that identifying patterns is crucial for developing explanatory frameworks and understanding underlying mechanisms in historical events. This debate is especially relevant to conflict studies, where recurring patterns of violence can help to identify underlying causes and improve predictive models.

This paper seeks to explore whether temporal patterns in history do indeed repeat. Specifically, we aim to determine the extent of repetition in conflict-related events and identify patterns across various historical periods and geographical boundaries. Conflicts throughout history have exhibited considerable diversity, presenting a unique opportunity to investigate the potential for consistent recurrence of violence and fatalities over time and space. By examining the temporal progression of conflict-related fatalities, we aim to uncover inherent cycles or motifs within these events. Identifying such patterns would suggest the existence of intrinsic temporal rhythms in conflicts, potentially shedding light on the fundamental mechanisms driving the dynamics of conflict.

Measuring the repetition of temporal patterns in conflict is valuable for several reasons. Firstly, it helps us to understand the stability and predictability of conflict dynamics. If certain conflict patterns are found to repeat over time, it suggests that there are underlying factors or conditions consistently driving these events. This understanding of the temporal aspects that contribute to the recurrence of violence can inform theoretical models of conflict. Secondly, identifying redundant patterns enables the development of more effective intervention strategies. By recognizing when specific patterns are likely to reoccur, policymakers and practitioners can time their interventions to pre-empt and mitigate the impacts of conflict. Lastly, measuring redundancy in temporal patterns allows for a comparative analysis across different domains. By comparing the results obtained from conflict data with those in other fields such as seismology, finance, and epidemiology, we can better understand the uniqueness or commonality of conflict dynamics relative to other complex systems.

We recognize that the question of identifying recurring patterns in history is ambitious and complex, so the scope of this paper is intentionally narrow—we focus specifically on conflict fatalities. By focusing solely on temporal sequences of conflict fatalities, we are limited in what we can measure and conclude. Our analysis thus does not account for the broader social, political, or economic factors that often influence conflicts, nor does it address the potential recurrence of non-fatality-related events. We recognize that this begins to address how historical patterns might repeat. Despite this narrow scope, we aim to offer a methodological blueprint for investigating similar questions across different domains. Additionally, our inquiry contributes to the broader discussion on the predictability of historical events. We aim to assess the frequency and extent of temporal sequences' recurrence, questioning whether these sequences recur more frequently than would be expected by chance.

Our methodological approach centers on the concept of entropy—a measure of randomness or uncertainty in a dataset. By calculating the entropy of temporal sequences in conflict data, we can estimate the level of redundancy within these sequences. Low entropy would indicate higher predictability and suggest an underlying data-generating process favoring certain patterns. Conversely, high entropy would point to a more random distribution of patterns and low redundancy. We apply this entropy-based analysis to various aspects of conflicts, including conflict type, region, and temporal resolution. Additionally, we compare our results with findings from other fields such as seismology, finance, and epidemiology to contextualize the predictability of conflict events within the broader predictive landscape. Finally, we also focus on whether certain patterns tend to be consistently followed by certain outcomes. By calculating the entropy of these outcomes, we can quantify the information content available in these patterns.

While we draw methodological tools from complexity science and information theory, our central theoretical contribution is to political science. Specifically, we engage three longstanding debates in the study of conflict.

First, our work speaks to research on the temporal dynamics of political violence, including the escalation and de-escalation of conflict over time. Scholars have long debated whether conflict trajectories are strategic and path-dependent or largely stochastic (Fearon 2004, Cederman, Wimmer & Min 2010). By measuring the recurrence of specific fatality patterns, we provide empirical evidence that escalation often follows discernible structures—supporting the idea that conflict unfolds in temporally organized ways.

Second, we contribute to the literature on conflict forecasting and early warning. Existing models typically focus on structural risk factors such as political institutions, economic inequality, or demographic pressures (Goldstone, Bates, Epstein, Gurr, Lustik, Marshall, Ulfelder & Woodward 2010, Michael D. Ward & Bakke 2010). In contrast, our approach captures endogenous temporal dynamics, identifying patterns that precede different types of conflict trajectories. This allows us to assess not only whether conflict occurs, but how its intensity and duration might evolve—an increasingly relevant concern in both academic and policy forecasting (Chadefaux 2014).

Third, we engage the debate over whether conflict dynamics are universal or context-specific. Some scholars emphasize the local, idiosyncratic nature of political violence, while others identify recurring "conflict cycles" or macro-patterns that transcend regional boundaries (Kriesberg 1998*a*, Harish & Little 2017*a*). Our findings show that certain temporal structures repeat across countries and decades, suggesting that while triggers of violence may differ, the unfolding of conflict often follows shared rhythms.

In all three areas, our contribution is to provide a quantitative method for evaluating temporal regularities in conflict data. Rather than assuming cyclical or patterned behavior, we use entropy-based analysis and Dynamic Time Warping to test whether—and under what conditions—such structure exists. In doing so, we offer a methodological bridge between statistical forecasting and qualitative insights about historical repetition in conflict.

Our analysis reveals that conflict patterns are dominated by periods of peace, but when focusing on active conflict periods, these patterns exhibit a spectrum of predictability, ranging from highly unpredictable sequences (similar to white noise) to more stable patterns resembling earthquake data. Predictability varies by conflict intensity, with low-intensity conflicts showing more regular, predictable patterns and high-intensity conflicts exhibiting greater unpredictability. Despite differences in magnitude, conflict patterns remain consistent across regions and decades, with no significant region-specific or time-specific trends. Temporal resolution also affects predictability, with weekly data showing slightly higher entropy than monthly data. Overall, conflict patterns show higher entropy than other fields like inflation or temperature, though less than random data, highlighting both the potential and limits of forecasting conflict dynamics using historical data. Furthermore, our findings suggest a degree of universality in conflict dynamics, as no region or decade-specific patterns were detected. This suggests that conflict patterns may be driven by global processes rather than localized phenomena.

The paper is structured as follows: We begin with a review of relevant literature, followed by a detailed description of our methodology. We then present our findings, categorized by various conflict attributes. The discussion section interprets these results in the context of historical repetition and conflict predictability. We conclude by exploring the implications of our findings for conflict prevention and management strategies, as well as suggesting avenues for future research.

Patterns of Conflict

The assertion that "those who cannot remember the past are condemned to repeat it" (Santayana 1905) suggests that events and patterns in human behavior, societal trends, and global phenomena tend to repeat over time. Thucydides, in his analysis of the Peloponnesian War, observed recurring patterns in state and individual behavior (Thucydides 2000). Similarly, Machiavelli argued that political states cycle through periods of order and disorder, driven by inherent human nature and societal dynamics (Machiavelli 1532), while Marx famously remarked that "history repeats itself, first as tragedy, second as farce" (Marx 1852).

The debate over historical repetition is central to historical and political analysis, with significant implications for understanding historical processes. Proponents of cyclical history argue that recognizing patterns helps uncover causal relationships (Goodin & Tilly 2006, Mahoney & Rueschemeyer 2003), while critics argue that this oversimplifies unique events shaped by specific contexts (Collier & Mazzuca 2006). This tension reflects a fundamental challenge in historical analysis: the need to balance the particularities of each historical moment with the search for generalizable insights.

For conflict specifically, there are clear reasons to expect temporal redundancy, including historical patterns, structural issues, and predictable trajectories. Much of the literature on conflict focuses on identifying recurring conditions that may lead to violence. From economic instability and inequality (Collier 2009) to political transitions and weak governance (Fearon & Laitin 2003), research often centers on these repeating factors as drivers of conflict. The concept of "conflict cycles" has been widely discussed in political science and related disciplines. It emphasizes the repetitive patterns of conflict escalation and de-escalation. For example, Kriesberg (1998*b*) describes how conflicts typically transition through phases of intensification and resolution, shaped by internal dynamics and external interventions. Similarly, (Acemoglu & Wolitzky 2014) offer an economic perspective, modeling how cyclical patterns of distrust and conflict arise due to incomplete information and coordination failures. In political contexts specifically, (Harish & Little 2017*b*) identify recurring cycles of electoral violence, with predictable phases of heightened conflict followed by periods of relative calm. This literature underscores the importance of understanding temporal dynamics in conflicts. In particular, it highlights predictable phases of escalation, peak violence, and de-escalation. Our study complements this literature by empirically quantifying the degree of redundancy or repetition within conflict-related fatalities using entropy-based measures and Dynamic Time Warping. By doing so, we directly assess the extent to which theoretically suggested conflict cycles manifest consistently across different conflicts and temporal contexts, providing a systematic method for evaluating the predictability implied, among others, by cyclical theories.

However, there are also reasons why conflict patterns may not repeat, including strategic behavior, data limitations, and stochastic events. Conflicts involve complex strategic, dynamic, and stochastic human actions that are not entirely dissimilar to stock markets. Any discernible pattern in behavior risks being exploited by actors, in such a way that the repeating pattern disappears (Tversky & Kahneman 1974). Even with a comprehensive understanding of conflict causes, understanding the timing, intensity, and evolution of conflicts can be an uphill task. Conflict dynamics are often nonlinear, subject to minute changes in conditions. Furthermore, data is often incomplete or unreliable.

Additionally, the fickle nature of alliances and international relations can drastically alter the conflict landscape. The secretive and delicate nature of diplomatic negotiations, peace treaties, or shifting geopolitical interests can transform a potential conflict into a situation of relative stability or vice versa (Holsti 1996). These complexities are challenging to encapsulate within simple temporal patterns. Furthermore, individual decision-making plays an important role in conflicts as well. Influential figures, political leaders, or even single actors can trigger, escalate, or defuse conflicts. Influenced by personal biases, beliefs, or misinformation, these decisions are highly stochastic (Tetlock 2005*a*). 'Black swan' events—unforeseeable, highly impactful incidents such as sudden technological advancements, natural disasters, or pandemics, can significantly alter conflict dynamics (Taleb 2007*a*).

The extent to which data "repeats" varies widely across scientific fields, driven by differences in the nature of phenomena, data availability, and the complexity of models. In seismology, predicting earthquakes is notoriously difficult (Geller, Jackson, Kagan & Mulargia 1997). Similarly, in finance, the efficient market hypothesis suggests that market prices reflect all available information, so that consistently achieving returns above the market average is almost impossible. In contrast, climate models, which simulate interactions between the atmosphere, oceans, land, and ice, are effective at predicting broad, long-term patterns such as global warming. Epidemiological models, while useful for predicting disease spread, also come with substantive uncertainty, as seen during the COVID-19 pandemic.

To illustrate the implications of redundancy for prediction, Figure 1 presents a concrete example of a recurring pattern in conflict fatalities. The red line shows a 12-month sequence from the test set characterized by a declining trend with intermittent rebounds. A classical time series model, such as ARIMA, trained on the full dataset would likely extrapolate a continued decrease. However, our pattern similarity-based model identifies a small number of highly similar historical patterns (also in red), all of which are followed by an increase in conflict fatalities (black). The resulting prediction, based on this localized matching, successfully captures this trend reversal. This example demonstrates how redundancy—i.e., the reappearance of similar temporal structures—can support accurate forecasts even with limited data, in contrast to global autoregressive approaches that average over many dissimilar sequences.

Each field faces unique predictive challenges. Earthquakes and conflicts have identified partial mechanisms for prediction, but only under limited conditions (e.g., yearly forecasts for conflicts or low-magnitude earthquakes) (Banna, Taher, Kaiser, Mahmud, Rahman, Hosen & Cho 2020). Financial markets' unpredictability stems from

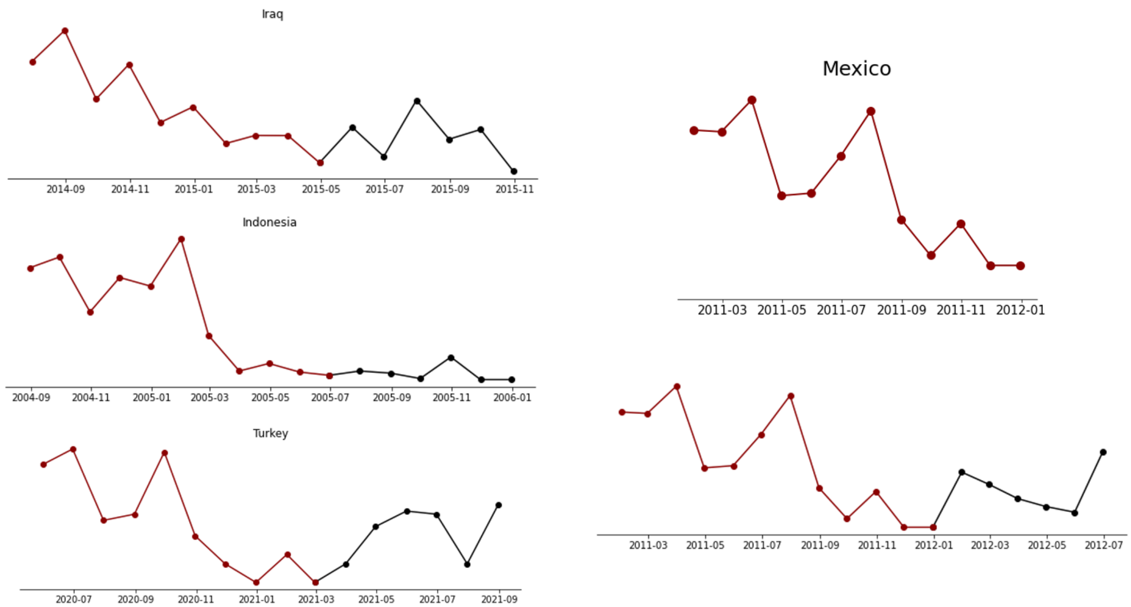


Figure 1: Illustration of prediction using a pattern similarity-based model. The top-right subplot shows the input pattern (red), drawn from the test set. The three subplots on the left display the closest historical matches (in red) along with their true future values (in black). The bottom-right subplot shows the actual future trajectory of the input pattern (black). This example highlights how high redundancy—i.e., repeated historical patterns—enables accurate prediction based on a small number of matched analogs, even when a classical autoregressive model would fail.

their strategic nature (Hill & Motegi 2019), while yearly climatic patterns are highly predictable due to strong seasonality (Papacharalampous, Tyralis & Koutsoyiannis 2018), though short-term climate events can still be chaotic.

Data

We draw upon conflict casualty data sourced from the Uppsala Conflict Data Program's (UCDP) Georeferenced Event Dataset (GED). This data collection provides an extensive overview of global conflict incidents, each accounting for at least one reported fatality. The dataset encompasses a worldwide scope, capturing daily occurrences from 1989 onwards. Our analysis focuses on the fluctuation of fatality numbers over time.[1]

To compare redundancy across various fields, we rely on a range of temporal data sources. Climate data, specifically country-averaged temperature readings, is obtained from ERA-5, a satellite data product of the ECMWF Integrated Forecast System (IFS) under the Copernicus European program (Hersbach, Bell, Berrisford, Hirahara, Horányi, Muñoz-Sabater, Nicolas, Peubey, Radu, Schepers et al. 2020). For seismological data, we turned to the comprehensive earthquake catalog maintained by the United States Geological Survey (USGS). This catalog provides crucial information on seismic events, including location, magnitude, and depth. To ensure meaningful analysis, we extracted country-level earthquake data with a magnitude exceeding five, considering lower magnitude events to be generally predictable (Banna et al. 2020). In the context of epidemiology, we used data detailing daily Covid cases per country.[2] Historical financial stock return data was obtained from Yahoo Finance and inflation data from the World Bank data, encompassing 192 countries and spanning from January 1970 to December 2022. (Ha, Kose & Ohnsorge 2023).

Methods

In this section, we outline the steps taken to extract and analyze patterns from conflict data to understand temporal redundancy. Our approach follows three main steps. First, we extract the relevant subsequences from larger time series data. Second, using Dynamic Time Warping, we identify and compare similar patterns, accommodating variations in the speed of events. Finally, we use these measures of similarity to quantify temporal redundancy at multiple levels of aggregation and across different fields. We detail each of these steps below.

[1] We underscore that our study focuses on temporal patterns in conflict fatality figures, which imposes certain limitations on our conclusions. Conflict, as a sociopolitical phenomenon, is shaped by a myriad of observable and non-observable variables beyond just fatalities. We have chosen to focus on fatalities as a key measurable output of these multifaceted conflict dynamics. Though this focus might only give a partial view, fatality numbers, as a tangible expression of conflict severity, offer a starting point for our exploration of temporal redundancy and predictability. Our study is not an attempt to simplify conflict to its death toll but rather a starting point for a broader conversation about redundancy in conflict data.

[2] The data was sourced from Our World In Data for the period between March 2020 and April 2023 (Mathieu, Ritchie, Rodés-Guirao, Appel, Giattino, Hasell, Macdonald, Dattani, Beltekian, Ortiz-Ospina & Roser 2020).

Pattern Extraction

Our first step is the extraction of subsequences from each of the larger time series. We used a sliding window approach to extract 12-month subsequences from each country's conflict data, beginning in January 1989 and continuing until 2023. This yielded 4,480 subsequences per country. For the other fields, we randomly extracted 300 subsequences from each dataset. The sample size is limited to 300 because our goal is to compare it with conflict data, rather than conducting a detailed analysis of other variables. Analyzing additional fields would require significantly more computation time, without significantly affecting the entropy scores. Each subsequence was normalized to values between 0 and 1 to allow for comparison across different scales of conflict.[3] These normalized subsequences form the basis for our subsequent analysis.

Due to pronounced zero-inflation in the data—57% of the extracted patterns consist of 12 zeros ('flat' patterns)—we conducted two analyses: one using the full dataset and another focused on sequences with at least one non-zero value ('conflict-filtered'). This approach helps to examine the dynamics within conflicts, without being skewed by long periods of peace

Additionally, since patterns can vary significantly across different temporal scales, we examined temporal redundancy at multiple levels of aggregation, including weekly and monthly intervals. For example, in weather forecasting, predictability is higher on an annual scale—summers are generally warmer than winters, and certain regions consistently experience more rainfall during specific months. However, this predictability diminishes at the weekly scale due to numerous variables affecting day-to-day weather. Similarly, in conflict data, the dynamics may differ depending on the temporal scale. By adopting this multi-scale approach, we are able to explore and compare the predictability of patterns across various levels of temporal granularity.

Once extracted, the subsequences are compared using Dynamic Time Warping (DTW), a metric that captures similar shapes by accommodating variations in the timing of events (Keogh & Ratanamahatana 2005). In essence, DTW measures the similarity between two temporal sequences by allowing for non-linear alignments in time. This means that two patterns do not need to unfold at the same speed to be considered similar. For example, a sudden spike in conflict fatalities that occurs over a shorter time period in one conflict could still be matched with a more gradual increase in another, provided the overall shape of the sequences is similar. DTW accomplishes this by "warping" the time axis, aligning points in the two sequences that are similar in value, even if they occur at different times. Details of the DTW

[3] The normalization is performed using the following formula: $\text{normalized value} = \frac{\text{original value} - \text{min value}}{\text{max value} - \text{min value}}$ where min value and max value are the minimum and maximum values in the specific 12-month subsequence.

methods can be found in appendix A.

The use of Dynamic Time Warping (DTW) in our analysis is motivated by the need to capture recurring patterns in conflict trajectories that may not align precisely in time. Traditional methods that compare sequences point by point (e.g., Euclidean distance) assume that escalation occurs at the same pace across cases, which is unlikely in real-world conflict dynamics. DTW, by contrast, allows for temporal distortion—matching sequences that are structurally similar but unfold at different speeds or include lags. This flexibility is essential when studying political conflict, where escalation and de-escalation may be triggered by endogenous cycles (e.g., election timing), exogenous shocks (e.g., international intervention), or strategic delays. DTW thus allows us to compare 12-month patterns with subsequences ranging from 10 to 14 months, as similar patterns may unfold over different time spans across regions or periods (Chadefaux 2022).[4] Subsequences that fall below a predetermined distance threshold are considered part of the same pattern.[5]

Assessing the Repetition of Conflict Patterns

Our initial aim is to determine the extent of redundancy in conflict patterns. We use entropy, a concept from information theory, to measure the randomness or uncertainty in the dataset (Rabiner 1989). Lower entropy indicates higher redundancy, meaning that specific patterns repeat frequently with similar outcomes, suggesting a more structured and predictable dataset. In contrast, higher entropy reflects greater variability and less predictability.

Maximum entropy occurs in a uniform distribution where all outcomes are equally likely, while minimum entropy occurs when an outcome is nearly certain. Thus, high-entropy sequences indicate greater unpredictability, whereas low-entropy sequences suggest more predictable patterns. Calculating entropy for temporal sequences in the conflict data allows us to assess and compare redundancy across fields, providing insights into the unique or shared dynamics of conflict relative to other complex systems.

[4]The DTW algorithm measures similarity between temporal sequences of varying lengths using specific parameters. Euclidean distance was used as the base metric between individual points, with a Sakoe-Chiba band of 10% of sequence length to constrain the warping path and reduce computational complexity. To account for different sequence lengths, the DTW distance was normalized by dividing the raw DTW distance by the warping path length. A symmetric step pattern allowed equal movement in diagonal, horizontal, and vertical directions, with strict boundary conditions ensuring that the warping path started and ended at the beginning and conclusion of both sequences.

[5]We set this threshold at 0.3 to balance the similarity of sequences within patterns and the number of patterns identified. To address the issue of parameter sensitivity and clustering validity, we conducted a systematic robustness analysis, the results of which are now included in Appendix C (Figures 8–10)

Entropy complements DTW by measuring the level of uncertainty in the outcomes that follow these matched patterns. Low entropy indicates that similar trajectories often lead to similar results, which is crucial for understanding how much predictive content is embedded in conflict sequences. Together, DTW and entropy allow us to assess both the recurrence of temporal motifs and their relevance for forecasting—two core concerns in the study of conflict dynamics.

Traditional methods like Approximate Entropy (ApEn) and Sample Entropy (SampEn) (Pincus 1991) are often used to assess short-term predictability but struggle to capture broader, long-term patterns. Conflict dynamics, influenced by political, economic, and social changes, often unfold over weeks, months, or years, making localized entropy measures insufficient. Our approach shifts the focus to clustering sequences that capture entire temporal structures, allowing us to identify patterns that span longer time frames. In conflict data, significant escalations or de-escalations are typically driven by long-term trends rather than short-term fluctuations, requiring a more comprehensive analysis of temporal structures. For a more detailed discussion of entropy, see appendix B.

Evaluating the Consistency of Patterns' Outcomes

Our second key question is whether a pattern consistently leads to the same outcome in the future. To assess the consistency of outcomes that follow specific patterns, we examine "pattern futures"—the six values following each subsequence—using hierarchical clustering. By employing hierarchical clustering, we are able to group similar "futures" based on the subsequent trajectory of conflict fatalities. This approach helps identify potential future scenarios that could follow a pattern and assigns probabilities to each.[6] We calculate the Shannon entropy of each cluster using the following equation:

$$H(X) = \frac{-\sum_{i=1}^{p} P(x_i) \log P(x_i)}{\log N_{\text{sub}}} \tag{1}$$

$$E(X) = \sum_{j=1}^{np} \frac{f_j H(X)}{f_j} \tag{2}$$

[6] Hierarchical clustering offers several advantages over other techniques such as k-means clustering or principal component analysis (PCA) in the context of our study. First, unlike methods such as k-means clustering, where the number of clusters must be predefined, hierarchical clustering does not require us to specify how many distinct clusters exist beforehand. This is particularly advantageous for our analysis of conflict patterns, as we do not have prior knowledge of the number of recurring "futures" that follow each pattern. Second, conflict dynamics, especially the future trajectories following a particular pattern, are often nonlinear and complex. Hierarchical clustering, with its ability to consider nested relationships between data points, captures these complexities more effectively than methods that assume linear relationships, such as PCA.

Here, $H(X)$ is the normalized entropy value, $P(x_i)$ is the probability of cluster i, p is the number of clusters of pattern futures, and N_{sub} represents the number of subsequences classified as a pattern. The first equation calculates the normalized entropy for each individual pattern. The second equation, $E(X)$ (called the Future entropy) aggregates the entropy values and weighs them by the frequency f_j of each np pattern. Thus, $E(X)$ captures in a single number the average predictability of the future outcomes of the patterns.

Results

How Often Do Patterns Repeat? The recurrence of specific patterns across different domains offers insight into the underlying complexity and predictability of each system. Figure 2 presents the frequency of pattern matches within a variety of datasets, ranging from conflict and earthquakes to economic and environmental data. A high concentration of patterns within a few dominant sequences points to repetitive and relatively simple systems. Conversely, datasets like white noise, characterized by randomness, show little to no pattern repetition.

Unsurprisingly, we find that the distribution of conflict data is highly skewed and dominated by 'flat' patterns. This skewness is expected, as most regions of the world are at peace most of the time, leading to numerous sequences with zero conflict events. To gain a clearer understanding of within-conflict dynamics, it is more insightful to analyze a dataset that filters out the non-conflict periods. We refer to this dataset as 'conflict-filtered'. This filtered dataset focuses on sequences with at least one conflict event, thereby providing a more accurate representation of the patterns and trends within active conflict periods. The conflict-filtered distribution shares similarities with other datasets characterized by recurring events, such as rainfall and filtered earthquakes. Like conflict-filtered, these datasets exhibit patterns where certain sequences repeat with a degree of regularity. In the case of rainfall, the patterns are influenced by seasonal cycles and climatic conditions, while filtered earthquakes reflect the periodicity of seismic activities without the noise of minor tremors. The similarity between conflict-filtered and these natural phenomena suggests that conflicts, like rainfall and significant earthquakes, are governed by underlying factors that drive their recurrence.

Do Patterns Lead to Predictable Outcomes? The distribution of patterns informs us about the likelihood of observing a specific pattern at a given time. However, a system with few patterns can still be unpredictable if their outcomes are random, while a system with many patterns can be highly predictable if each leads to consistent outcomes. For example, traffic flow on a few routes (low diversity) can be

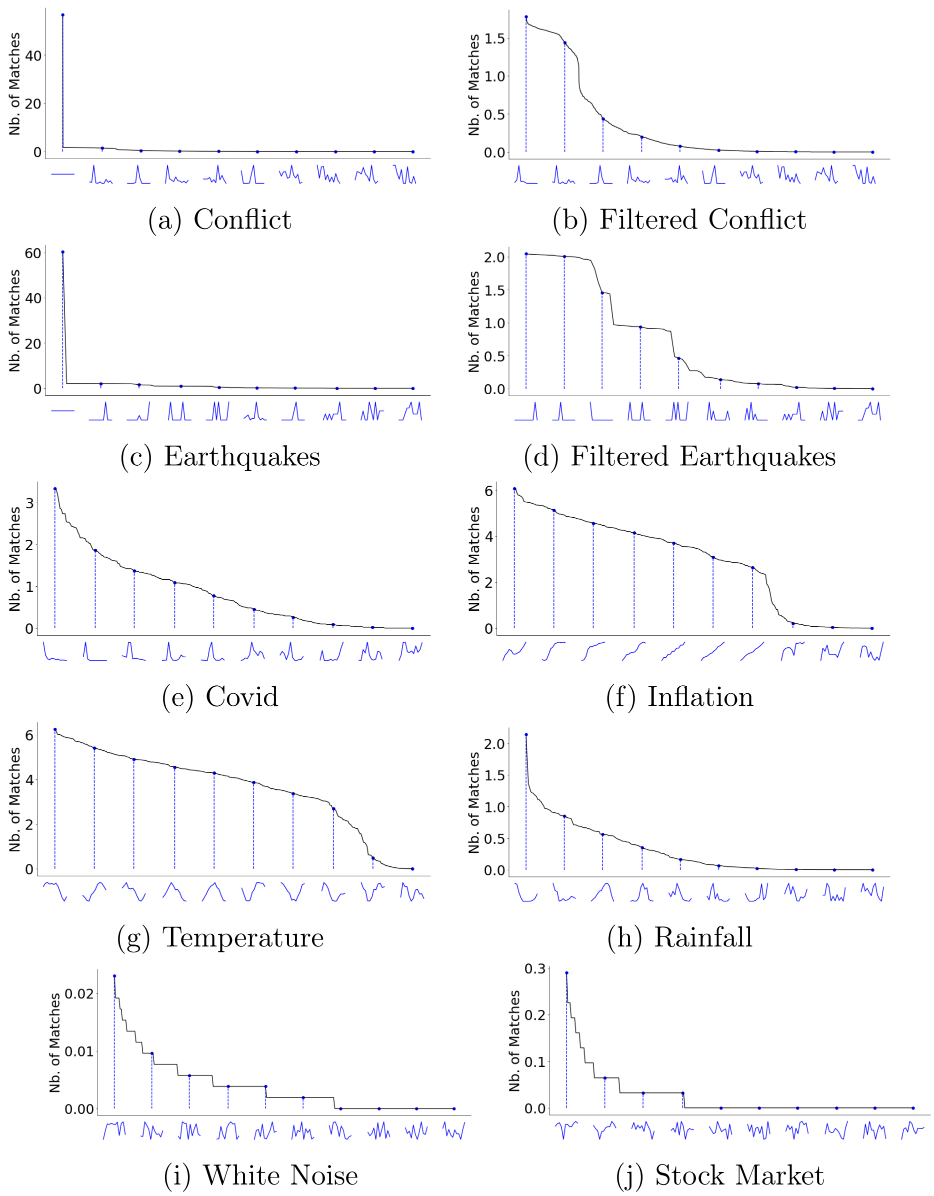


Figure 2: Frequencies of patterns for various datasets with 12-month window length. Sample patterns are displayed along the x-axis for illustration. In the "filtered" datasets, only sequences with at least one non-zero value were retained.

unpredictable due to random events, while a busy train schedule with many routes (high diversity) can remain predictable if trains run consistently on time.

Figure 3 displays both the pattern distribution and the distribution of their future outcomes. The x-axis represents pattern frequency, indicating how often each pattern appears in the data. A point on the far right suggests that all patterns are identical, which is nearly the case for conflict data, where most patterns are 'flat'. Conversely, points on the left indicate unique patterns that occur only once—as seen with many white noise patterns, which appear infrequently, as expected in random draws. While short temporal patterns can occasionally repeat by chance, the expected frequency of repetition is limited, serving as a useful reference point.

On the y-axis, we plot the entropy of each pattern's future, which is calculated by collecting the future six values for each pattern and determining the entropy of their distribution (see above). High entropy on the y-axis indicates that these patterns lead to widely varying outcomes, while low entropy suggests more consistent and predictable outcomes.

We observe a clear correlation between the frequency of patterns and the predictability of their future states, which is expected given the finite size of the dataset.[7] However, differences across domains remain clear. For example, temperature data shows a distribution similar to earthquakes and inflation, with most observations clustered in the top right, along with some patterns exhibiting lower entropy along the diagonal. This top-right position corresponds to a high frequency of subsequences per pattern and low entropy in their futures. Conflict data is highly concentrated in this corner, mainly due to the prevalence of flat patterns, resembling the pattern seen in earthquakes, though high-magnitude earthquakes occur more frequently than conflict fatalities. Inflation, meanwhile, shows clear and consistent patterns due to its "sticky" nature, while temperature sequences are influenced by seasonal trends, enhancing their predictability.

Additionally, the density distribution for conflict-filtered reveals mixed similar patterns characterized by two separate zones: one in the top-right corner similarly to the top right corner of Earthquakes, and another in the bottom-left corner similarly to White Noise and Stock Market. The top-right density mass suggests good predictability, as it corresponds to patterns with one or two peak values—commonly found in the earthquake dataset. These patterns generally lead to flat subsequent points, typically representing only a few fatalities that do not escalate into long-term conflict, contributing to low entropy. The second density mass corresponds to patterns with high entropy and few subsequences, as in White Noise and Stock Market dataset (concentrated in the bottom-left density). High entropy suggests these patterns are associated with low predictability. In contrast, Covid data and Rainfall reveal a more dispersed density with spread entropy scores through the diagonal,

[7] To see why, consider the extreme case where the dataset consists of a single repeating pattern—in this case, each pattern consistently leads to the same outcome, appearing in the top right of the figure. Conversely, if each pattern is unique (one sequence per pattern), their futures become unpredictable since no comparisons can be made, positioning them in the bottom left. This creates a natural constraint, clustering data along the diagonal.

reflecting the varied dynamics in those domains.

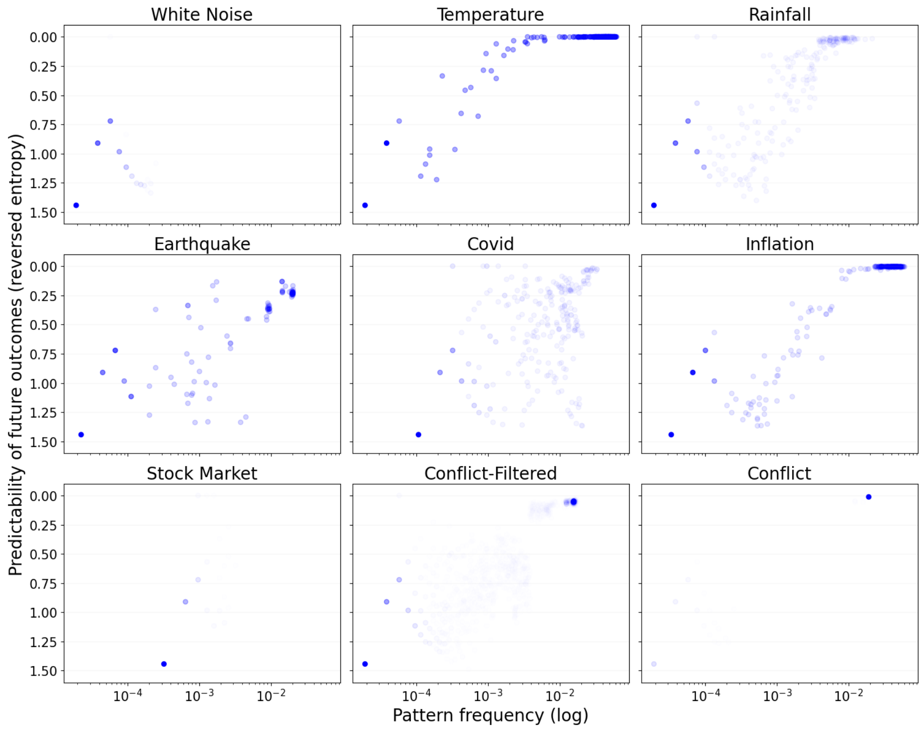


Figure 3: Density of patterns with normalized entropy (reversed scale) on the y-axis (predictability) and the log number of subsequences per pattern on the x-axis (pattern frequency). Patterns in the top right show frequent and predictable outcomes, while those in the bottom left are rare and unpredictable.

We also examined the impact of two parameters: window size and data aggregation frequency. Window size, which determines the number of observations in each sequence (e.g., 6, 12, or 18), directly affects our ability to capture patterns. Smaller windows reveal short-term, detailed fluctuations but may introduce noise, reducing overall predictability. Larger windows, while better suited for capturing long-term trends, can obscure finer details. Data aggregation frequency, which controls how often data is grouped (e.g., weekly or monthly), similarly shapes our analysis. High-frequency data captures rapid changes but increases volatility and unpredictability, while lower-frequency data smooths out short-term noise, highlighting more stable, recurring trends and improving predictability.

Figure 4 presents the Future Entropy values (y-axis in Figure 3) for various fields across different data aggregation configurations, including weekly and monthly frequencies. White noise and stock market data consistently have the highest Future

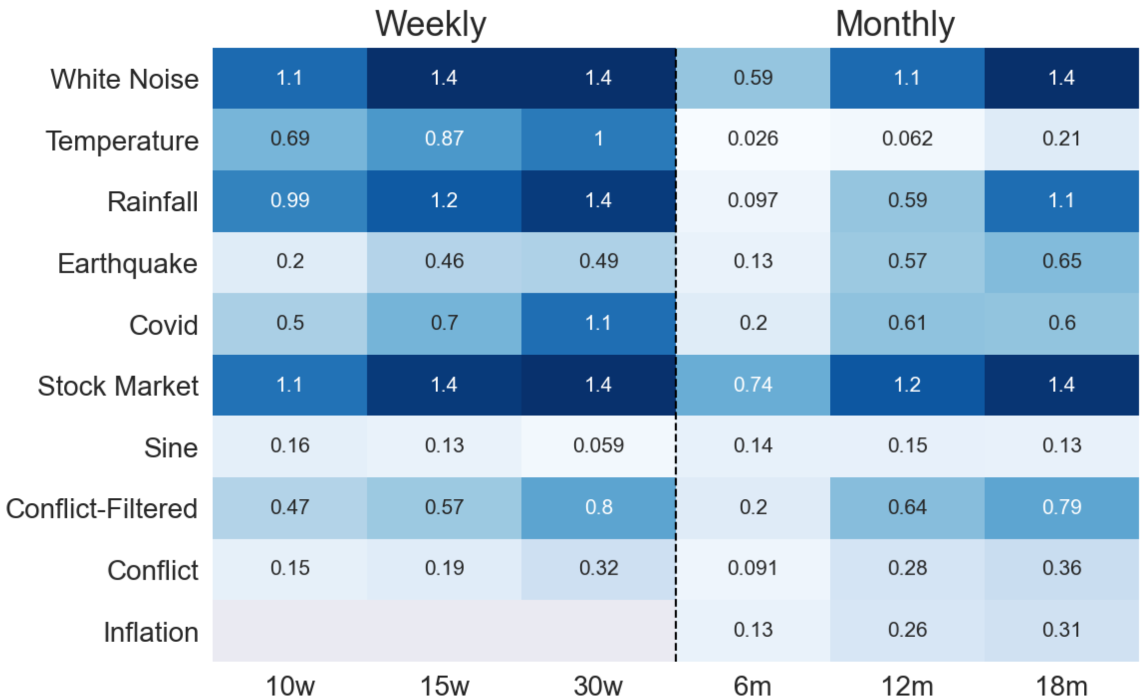


Figure 4: Future Entropy of different fields for various temporal resolutions and window lengths. Conflict corresponds to the fatalities dataset extracted from UCDP and conflict-filtered is the conflict dataset with filtered sequences.

Entropy scores, reflecting their stochastic nature and inherent unpredictability. These results align with previous findings that highlight the random behavior of these time series. A Sine curve was added to establish a lower boundary on the Future Entropy scale, as it is highly predictable.[8] Even with minor random noise introduced, the entropy values remain low across all conditions, demonstrating its high level of predictability.

Conflict has low entropy scores, similar to those of the Sine patterns, mostly due to the prevalence of flat patterns. Rainfall and temperature data have significantly higher entropy scores at weekly frequencies compared to monthly, highlighting the increased variability in short-term weather patterns. Finally, filtered conflict patterns display entropy scores higher than earthquake ones. In particular, the weekly entropy values for Conflict-filtered are close to its monthly values and similar to Covid scores, indicating a relatively low level of predictability.

[8]The Sine dataset was generated by simulating sinusoidal functions with consistent frequency and adding random noise. The number of series matches the number of countries in the conflict dataset, and the length of each series corresponds to the number of months in the conflict dataset.

Assessing the Universality of Conflict Patterns: Regional, Temporal, and Scale Variations

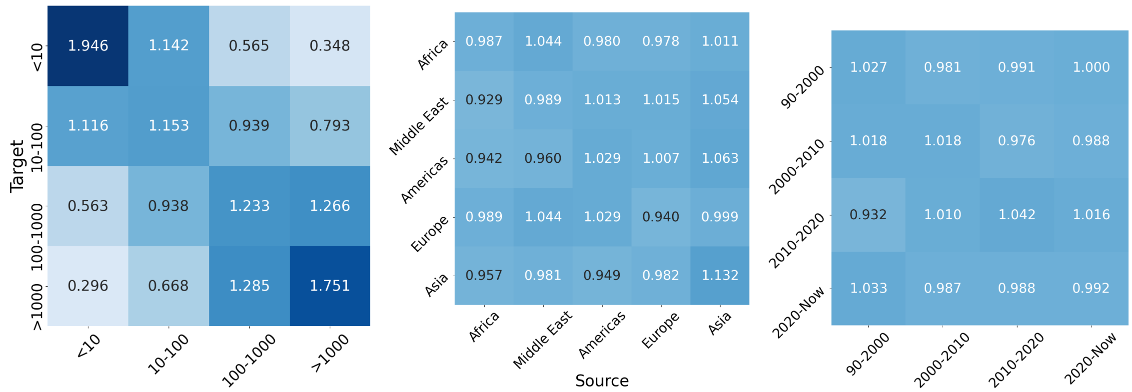
We have so far analyzed patterns globally, but it is possible that significant variation exists across regions, time periods, and scales of conflict. Are these patterns universal, or do they change depending on the context? The answer matters because it affects how we study and model conflicts. In particular, can a single, overarching model capture conflict dynamics, or do we need region- or magnitude-specific approaches to better understand and predict conflict behavior. We therefore now break down our results for conflict-filtered by magnitude, period, and region.

Of particular interest is whether certain patterns are specific to particular contexts, such as regions, time periods, or conflict sizes. For example, are there patterns unique to Africa or the 1990s, or specific to large-scale conflicts? If so, we would expect sequences of fatalities from Africa to cluster with other African sequences, rather than those from other regions. Similarly, patterns from large wars should differ from those in smaller skirmishes. To test this, we calculated the probability that a sequence (e.g., from an African country) would match the same pattern as a sequence from another context (e.g., a country in Asia). If patterns are not specific to regions or periods, we would expect this probability to be consistent across regions. In other words, the likelihood of two sequences belonging to the same pattern would not depend on where or when they occurred. Fig. 5 represents the distribution of these probabilities by conflict magnitude, decade, and region. For example, the top-right cell of Fig.5b corresponds to:

$$\frac{P(\text{pat}(i) = \text{pat}(j) \mid i \in \text{Asia}, j \in \text{Africa})}{P(j \in \text{Africa})},$$

where $pat(i)$ denotes the pattern to which sequence i belongs. This ratio expresses the probability that a sequence from Asia shares the same pattern as one from Africa, normalized by the probability that sequence j belongs to Africa. Figure 5a shows that the highest proportion of matching subsequences is found within the same magnitude pattern category. This indicates that patterns derived from low-magnitude conflicts predominantly match with other low-magnitude conflict subsequences. This finding suggests that distinct conflict patterns are specific to each magnitude category and are not generally shared across different magnitudes. On the contrary, in Fig. 5b and 5c, the distribution is consistent regardless of the region or decade where the pattern is extracted. African and the 90s decade subsequences are prevalent, aligning with the distribution of non-zero subsequences. It suggests that conflict patterns are comparable across space and time.

These results are supported by the distribution of entropies (of futures) presented in Fig. 6. Decadal and regional densities have similar multimodal distribution, aligning with the general conflict-filtered density. In contrast, the scale category shows



(a) Scale (b) Region (c) Decade

Figure 5: Distribution of probabilities observed in a given category in similar occurrences (rows) knowing the original pattern feature (columns), normalized by the probability of observing this given category in similar occurrences. We classify patterns based on four fatalities magnitude scales (left) (low: < 10, mid-low: 10-100, mid-high: 100-1000, and high: > 1000 per year), the five regions/continents coded in the UCDP dataset (center), and the available four decades in the dataset. (right)

disparate results. Low-intensity conflict patterns are characterized by low entropy rates, mainly due to the high number of patterns that only have non-zero values in one or two months, followed by zero values. These are similar in density and pattern shape to Earthquakes. Mid-high and high-magnitude patterns have comparable densities, concentrated around high entropy values. This could be partly explained by the small number of observations in history with a large number of fatalities, but also by the diversity of the few pattern futures that lead to uncertain following values. Average values for each magnitude subgroup are shown in the Appendix in Fig. 14.

Discussion

These findings have significant implications for both resource allocation and research focus in conflict studies. In terms of resource allocation, understanding the predictability of different conflict types can guide more efficient distribution of efforts and funds. Regions prone to low-intensity conflicts with more predictable patterns, such as South Thailand or the Mindanao region in the Philippines, might benefit from sustained, long-term peacebuilding efforts. These areas could be targeted for consistent investment in infrastructure, education, and community-building initiatives that address the root causes of conflict. In contrast, areas at risk of high-intensity conflicts,

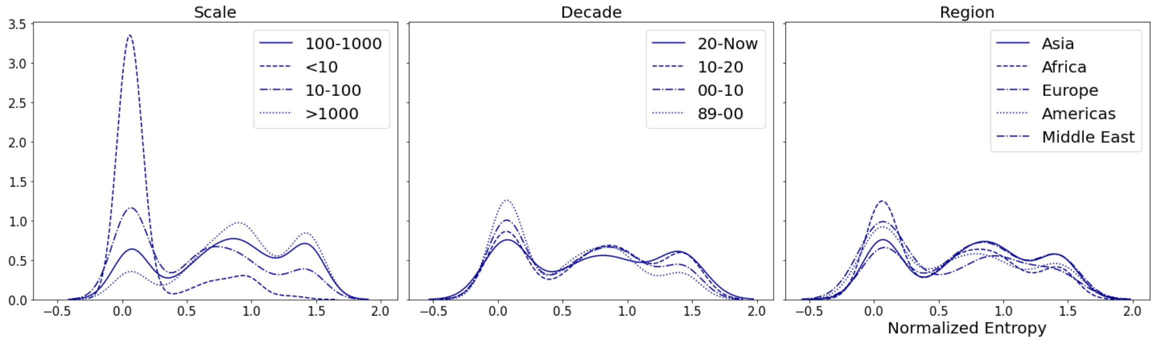


Figure 6: Density of normalized entropy of futures (Conflict-filtered data), per magnitude (left), decade (center), and region (right).

such as the Levant or Sudan, less predictable conflicts might require more flexible, rapid-response resources. Supplementing autoregressive models with real-time inputs like news signals or event-based data can help overcome the few similar historical occurrences necessary to build a robust early warning system. Moreover, maintaining readily deployable peacekeeping forces or establishing quick-response humanitarian aid systems is essential.

For conflict researchers, our findings highlight the need for different analytical approaches depending on the pattern features. Conflicts associated with rare or unique historical patterns, linked to high-intensity events, have high entropy values. In such cases, autoregressive models alone are insufficient, and additional real-time or contextual variables are necessary to improve forecasting accuracy. Researchers studying these conflicts should prioritize real-time data collection and analysis of rapidly changing situations. They might also explore the use of complex systems theory or chaos theory to better understand these highly volatile scenarios. In contrast, more predictable conflicts could be studied with an emphasis on long-term, structural factors. Here, researchers might employ longitudinal studies, focusing on gradual changes in socio-economic conditions, political structures, or demographic shifts that contribute to conflict patterns over time.

While our findings confirm that conflict dynamics are often difficult to predict, we stress that this unpredictability is not uniform. Some conflict sequences—particularly those involving low-intensity violence or recurring temporal motifs—exhibit high redundancy and are substantially more forecastable. Others, especially high-intensity or escalation-prone conflicts, display far greater entropy and lower pattern recurrence.

This heterogeneity is theoretically meaningful. It echoes long-standing arguments in conflict research that distinguish between slow-building, structurally driven violence and sudden-onset or strategically reactive escalation (Fearon 2004, **?**). Our analysis provides empirical grounding for this distinction, offering a framework to assess when conflict trajectories follow recognizable paths and when they resemble "black swan" events—outcomes that are rare, impactful, and difficult to anticipate (Taleb 2007*b*).

Moreover, by quantifying the temporal predictability of conflict events, our approach contributes to the growing literature on bounded rationality and limits of foresight in complex political environments (Tetlock 2005*b*). The ability to distinguish between forecastable and non-forecastable sequences may help refine the use of early warning systems, improve resource targeting, and generate testable hypotheses about the political and institutional factors that shape these temporal dynamics.

In this sense, our contribution is not just that conflict is difficult to predict in general—a point long recognized, at least theoretically—but that patterns of predictability vary systematically across conflict types. Recognizing and measuring this variation offers a path forward for both theoretical modeling and policy-relevant forecasting in political science.

Yet, important limitations that should be addressed. One is our focus on autoregressive patterns, without fully exploring the broader range of covariates—such as political, economic, and social factors—that may influence conflict dynamics. Future research could extend this study by incorporating such variables to develop more comprehensive models, particularly for the first class of unpredictable patterns. This could lead to a better understanding of the factors that drive sudden escalations in conflict fatalities and improve forecasting accuracy in such scenarios.

The quality and source of the data also present challenges. The reliance on news reports for conflict fatalities data introduces uncertainties, particularly regarding the accuracy of reported dates and death tolls. These potential discrepancies could obscure the detection of true patterns and may affect the robustness of our findings. Additionally, the timeframe of the study, beginning in 1989, may not fully capture long-term shifts in conflict dynamics or account for pre-existing trends.

A key limitation of our study is its reliance solely on conflict-related fatalities. While this choice excludes non-lethal forms of violence or protest, it enables us to focus on a consistently reported and substantively meaningful outcome: the severity of conflict. Fatalities reflect escalation in a way that is both observable and temporally structured, allowing for robust comparisons across space and time. Nonetheless, future extensions could incorporate broader event data, such as those from ACLED, to capture a wider range of conflict dynamics, including protests, riots, or threats. Doing so would enrich the analysis but also introduce new challenges, such as greater heterogeneity in event types and reduced comparability in magnitude.

A key takeaway from our findings is the consistency of conflict patterns across time and space, suggesting that while conflicts may differ in scale, their underlying dynamics are often comparable. This offers promise for enhancing predictive accuracy by leveraging historical data to anticipate future trends in conflict regions. However, the unpredictability of high-variation patterns remains a significant challenge for forecasting efforts. This finding underscores the need for further refinement of models that can account for sudden escalations in conflict intensity.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 101002240.

References

Acemoglu, Daron & Alexander Wolitzky. 2014. "Cycles of Conflict: An Economic Model." *American Economic Review* 104(4):1350–1367.

Banna, Md. Hasan Al, Kazi Abu Taher, M. Shamim Kaiser, Mufti Mahmud, Md. Sazzadur Rahman, A. S. M. Sanwar Hosen & Gi Hwan Cho. 2020. "Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges." *IEEE Access* 8:192880–192923.

Cederman, Lars-Erik, Andreas Wimmer & Brian Min. 2010. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62(1):87–119.

Chadefaux, Thomas. 2014. "Early Warning Signals for War in the News." *Journal of Peace Research* 51(1):5–18.

Chadefaux, Thomas. 2022. "A shape-based approach to conflict forecasting." *International Interactions* 48(4):633–648.

Collier, P. 2009. *Wars, Guns, and Votes: Democracy in Dangerous Places*. HarperCollins.

Collier, Ruth Berins & Sebastian Mazzuca. 2006. Does History Repeat? In *The Oxford Handbook of Contextual Political Analysis*, ed. Robert E. Goodin & Charles Tilly. Oxford University Press pp. 472–489.

Fearon, J. D. & D. D. Laitin. 2003. "Ethnicity, insurgency, and civil war." *American Political Science Review* 97(1):75–90.

Fearon, James D. 2004. "Why Do Some Civil Wars Last So Much Longer Than Others?" *Journal of Peace Research* 41(3):275–301.

Geller, RJ, DD Jackson, YY Kagan & F Mulargia. 1997. "Earthquakes cannot be predicted." *Science* 275(5306):1616–1617.

Goldstone, Jack A, Robert H Bates, David L Epstein, Ted Robert Gurr, Michael B Lustik, Monty G Marshall, Jay Ulfelder & Mark Woodward. 2010. "A global model for forecasting political instability." *American journal of political science* 54(1):190–208.

Goodin, Robert E. & Charles Tilly, eds. 2006. *The Oxford Handbook of Contextual Political Analysis*. Oxford University Press.

Ha, Jongrim, M Ayhan Kose & Franziska Ohnsorge. 2023. "One-stop source: A global database of inflation." *Journal of International Money and Finance* p. 102896.

Harish, S. P. & Andrew T. Little. 2017*a*. "The Political Violence Cycle." *American Political Science Review* 111(2):237–255.

Harish, S. P. & Andrew T. Little. 2017*b*. "The Political Violence Cycle." *American Political Science Review* 111(2):237–255.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers et al. 2020. "The ERA5 global reanalysis." *Quarterly Journal of the Royal Meteorological Society* 146(730):1999–2049.

Hill, Jonathan B & Kaiji Motegi. 2019. "Testing the white noise hypothesis of stock returns." *Economic Modelling* 76:231–242.

Holsti, K. J. 1996. *The State, War, and the State of War*. Cambridge University Press.

Keogh, Eamonn & Chotirat Ann Ratanamahatana. 2005. "Exact indexing of dynamic time warping." *Knowledge and information systems* 7:358–386.

Kriesberg, Louis. 1998*a*. *Constructive Conflicts: From Escalation to Resolution*. Rowman & Littlefield.

Kriesberg, Louis. 1998*b*. *Constructive Conflicts: From Escalation to Resolution*. Lanham, MD: Rowman Littlefield.

Machiavelli, Niccolò. 1532. *The Prince*. Penguin Classics.

Mahoney, James & Dietrich Rueschemeyer, eds. 2003. *Comparative Historical Analysis in the Social Sciences*. Cambridge University Press.

Marx, Karl. 1852. *The Eighteenth Brumaire of Louis Bonaparte*. International Publishers Co.

Mathieu, Edouard, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina & Max Roser. 2020. "Coronavirus Pandemic (COVID-19)." *Our World in Data* . https://ourworldindata.org/coronavirus.

Michael D. Ward, Brian D. Greenhill & Kristin M. Bakke. 2010. "The Perils of Policy by Prediction: Forecasting Civil Conflicts." *Journal of Peace Research* 47(4):363–375.

Papacharalampous, Georgia, Hristos Tyralis & Demetris Koutsoyiannis. 2018. "Predictability of monthly temperature and precipitation using automatic time series forecasting methods." *Acta Geophysica* 66(4):807–831.

Pincus, S. M. 1991. "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences* 88(6):2297–2301.

Rabiner, L. R. 1989. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77(2):257–286.

Santayana, George. 1905. *The Life of Reason: The Phases of Human Progress*. Charles Scribner's Sons.

Taleb, N. N. 2007*a*. *The Black Swan: The Impact of the Highly Improbable*. Random House.

Taleb, Nassim Nicholas. 2007*b*. *The Black Swan: The Impact of the Highly Improbable*. Random House.

Tetlock, P. E. 2005*a*. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press.

Tetlock, Philip E. 2005*b*. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press.

Thucydides. 2000. *History of the Peloponnesian War*. Penguin Classics.

Tversky, A. & D. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185(4157):1124–1131.

Appendix

A Dynamic Time Warping

Dynamic Time Warping (DTW) is an algorithm designed to align temporal sequences that may differ in speed or length. It aims to identify an optimal alignment between two sequences by warping the time axis. This enables the matching of similar patterns despite variations in timing. Whereas traditional distance metrics (e.g., Euclidean distance) require sequences of identical length and assume a direct one-to-one correspondence, DTW allows time shifts and distortions by allowing for "warping" along the time axis.

Consider for example two sequences, $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_m)$, where X and Y may differ in length. The DTW algorithm first constructs a cost matrix D of size $n \times m$, where each element $D(i, j)$ represents the squared Euclidean distance between the i-th point in sequence X and the j-th point in sequence Y. A cumulative cost matrix C is then calculated, where each element $C(i, j)$ is the minimum cumulative cost required to reach the point (i, j) from the starting point $(1, 1)$. The cumulative cost at a given point is then the sum of the current distance $d(i, j)$ and of the minimum cumulative cost from the previous step, using the following recurrence relation:

$$C(i, j) = d(i, j) + \min(C(i-1, j), C(i, j-1), C(i-1, j-1)),$$

where $d(i, j)$ denotes the distance between x_i and y_j.

Dynamic Time Warping's ability to accommodate temporal variations makes it particularly suitable for analyzing data where events unfold at different rates. For example, in conflict studies, fatalities in different regions may follow similar patterns over time, but these patterns might occur at different rates or have different durations. By applying DTW, we can align fatality numbers to detect common patterns despite variations in timing or intensity.

To illustrate DTW in practice, consider the two sequences depicted in Figure 7. These sequences are visibly misaligned due to differences in length and timing. By applying DTW, we can warp the time axis such that the sequences are aligned in a way that minimizes the total cumulative distance. This process involves constructing the cumulative cost matrix, calculating the minimal cost path, and identifying the optimal alignment despite the discrepancies between the sequences. The resulting alignment provides insights into the similarities and differences between the two time series, making it possible to compare patterns that may otherwise seem unrelated.

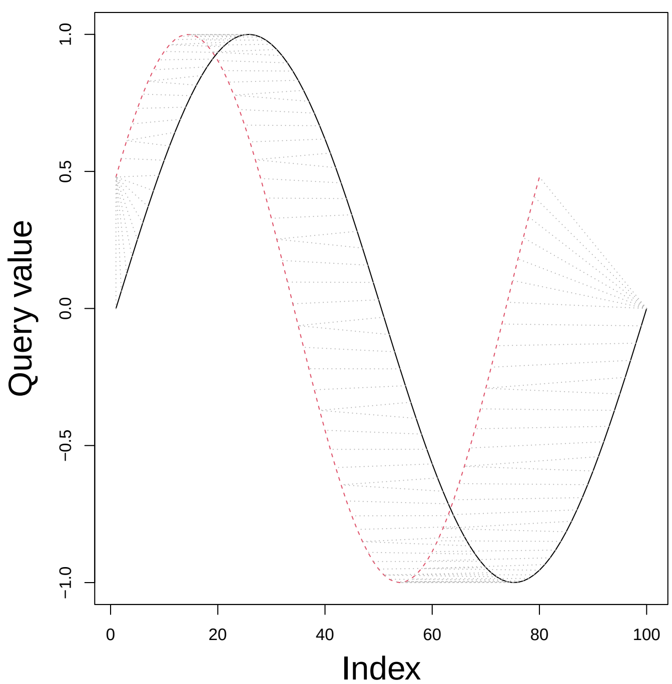


Figure 7: Example of two misaligned time series.

B Entropy

Entropy serves as a measure of uncertainty or randomness in time series data. The fundamental form of entropy used in information theory is Shannon entropy, which measures the average level of uncertainty or information content in a random variable. Given a discrete random variable X with possible outcomes $\{x_1, x_2, \ldots, x_n\}$ and probabilities $P(x_i)$, the Shannon entropy $H(X)$ is defined as:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i),$$

where $P(x_i)$ is the probability of occurrence of outcome x_i.[9] Higher entropy values suggest greater unpredictability (closer to random or white noise).

When applied to time series data, entropy-based methods such as Approximate Entropy (ApEn) and Sample Entropy (SampEn) are commonly used to assess the regularity and complexity of sequences. These methods quantify how often patterns of data points repeat within a time series, offering a measure of the system's predictability or randomness. ApEn calculates the probability that patterns of length m that are close to each other will remain close when extended by an additional point. SampEn refines this approach by excluding self-matches, reducing bias in the calculation.

However, both ApEn and SampEn are designed to assess the internal regularity of a single time series and are not suited for comparing two different sequences. Their focus on identifying local patterns and short-term repetitions within one sequence limits their ability to detect similarities between different time series, especially when these sequences evolve at different rates or durations. Since neither method accounts for temporal distortions or varying speeds, they cannot capture repeating patterns or structural similarities between sequences that may be misaligned in time. This makes them inadequate for contexts like conflict data, where events often unfold unevenly across time.

In contrast, clustering time series data offers a more effective approach for uncovering patterns that recur across different sequences. By grouping similar sequences into clusters, we can identify underlying structural similarities, even when those sequences differ in speed or duration. Dynamic Time Warping (DTW) further enhances this process by allowing for the non-linear alignment of sequences. DTW stretches or compresses time to account for variations in the timing of events, enabling a more accurate comparison between sequences. This is particularly important in the analysis of conflict data, where the timing and intensity of events can vary significantly across different regions and time periods.

[9] The logarithm is typically taken in base 2, so that the entropy is in bits.

C Parameters Robustness Tests

Our framework relies on two parameters: the maximum clustering distance (clu), which determines how closely future sequences must align to be grouped together, and the similarity threshold (d), which defines whether two sequences share the same temporal pattern. In choosing these parameters, we aimed to satisfy two criteria: first, that patterns should not be so specific that matches become rare or so general that they include dissimilar cases; and third, that the results should be interpretable both statistically and visually.

Figure 8 presents two clustering evaluation metrics: the Silhouette score (left y-axis, in black) and the Calinski-Harabasz Index (right y-axis, in grey), plotted across various clustering distance thresholds. The Silhouette score measures the degree of cohesion within clusters relative to their separation from other clusters, and tends to penalize configurations where clusters are close to each other, even if they are internally compact. However, the Calinski-Harabasz Index quantifies the ratio of between-cluster dispersion to within-cluster dispersion, and can yield high values for tightly grouped clusters, even when they are in close proximity. In both cases, higher values indicate better-defined clusters. As shown in the figure, a clustering distance of $clu = 0.3$ offers the best trade-off between the two metrics, balancing intra-cluster compactness with inter-cluster separation.

Similarly, Figure 9 examines the number of similar patterns found at various d thresholds. When $d = 0.1$, matches are sparse, which limits the ability to generalize from historical analogs. In contrast, $d = 0.5$ and $d = 1.0$ identify too many matches, some of which clearly diverge in behavior. A threshold of $d = 0.3$ results in a stable match distribution that avoids both extremes.

To visually validate the selected parameters, we present example matches in Figure 10, using the Mexico case introduced earlier. While no patterns matched at $d = 0.1$, the patterns that matched at $d = 0.3$ show clear structural and behavioral similarity to the reference sequence, while higher d values introduce clear mismatches. These results suggest that $d = 0.3$ effectively captures the intended shape similarity without collapsing meaningful distinctions.

Taken together, these analyses suggest that our chosen values for clu and d achieve a careful balance between overfitting and underfitting. They produce clusters that are internally coherent, predictively useful, and robust to small perturbations. This balance ensures that the entropy metric meaningfully reflects outcome uncertainty without being distorted by arbitrary or overly rigid pattern definitions.

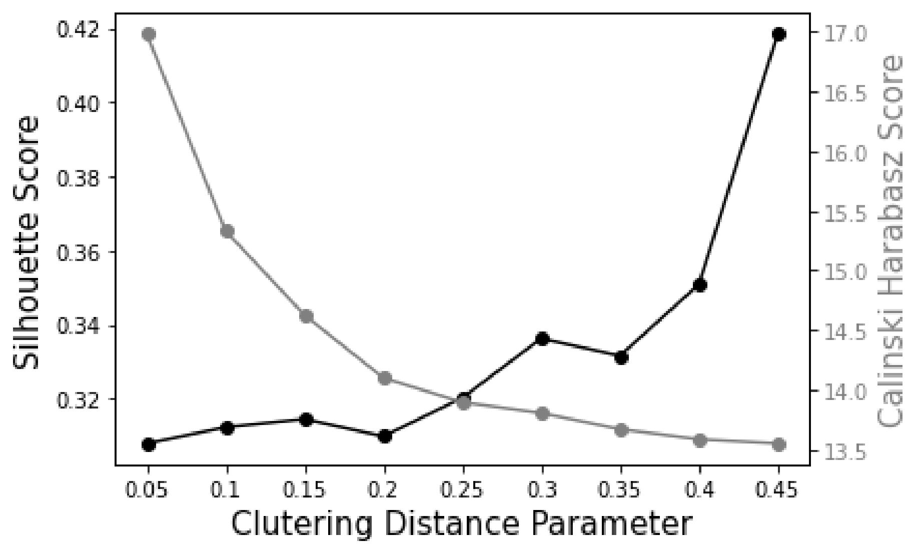


Figure 8: Silhouette score (left y-axis, in black) and the Calinski-Harabasz Index (right y-axis, in grey), plotted across various clustering distance thresholds.

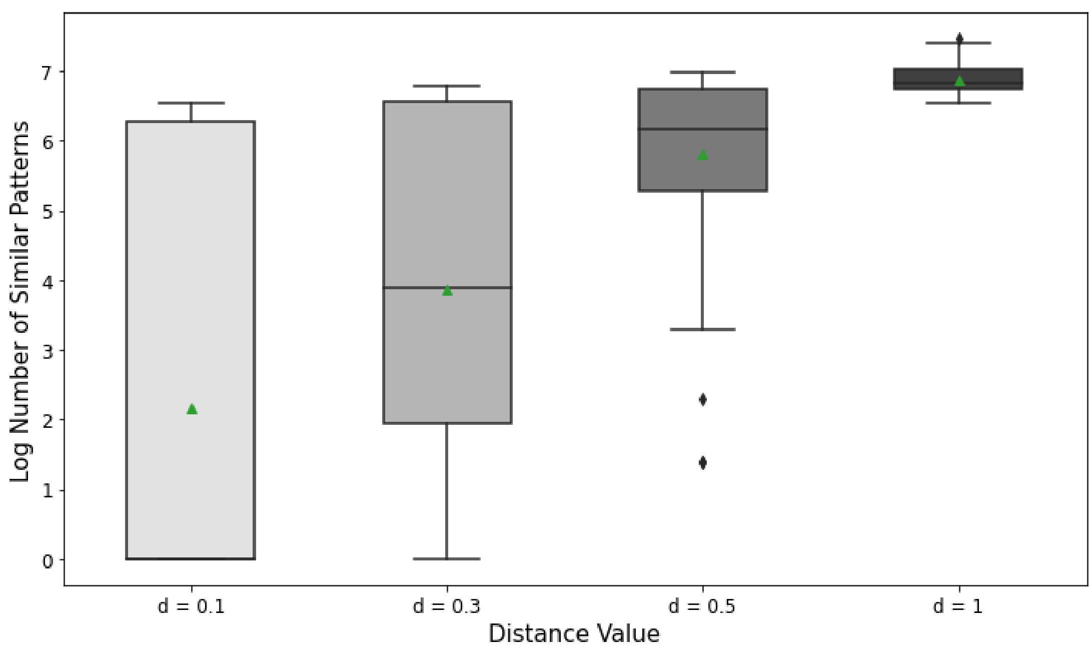


Figure 9: Box plot of the number of similar patterns found for four different d values (log).

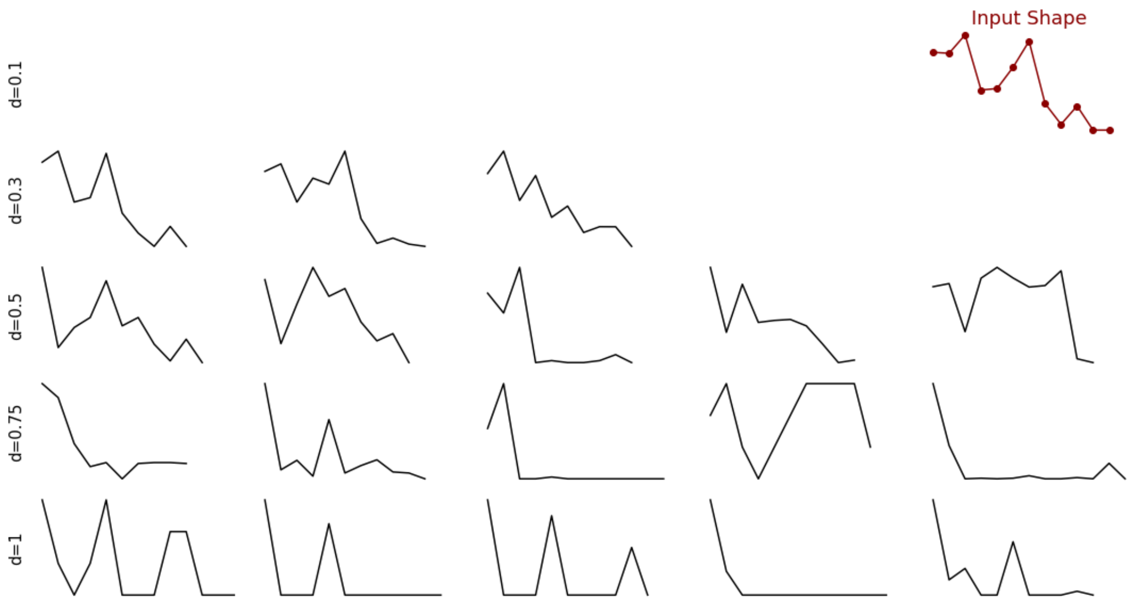


Figure 10: Examples of matched patterns for different d values, using the Mexico input shape (in red on the top right corner).

D Simulation-based power tests

First Test

In the first test, we generate a synthetic 10-point time series along with nine perturbed versions, each with linearly increasing levels of added noise, with the following process:

1. Generate an original random pattern as a vector:

$$x = [x_1, x_2, \ldots, x_{10}] \sim \mathcal{U}(0, 1)$$

2. Generate the corresponding nine perturbed patterns:

$$x^{(j)} = x + \epsilon^{(j)}, \quad \text{where } \epsilon^{(j)} \sim \mathcal{N}(0, \sigma_j^2)$$

Where σ_j increases gradually from 0 (original pattern) to 0.5 (maximum noise introduced)

3. Normalize and then calculate the DTW between the original pattern and the perturbed patterns

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad \tilde{x}^{(j)} = \frac{x^{(j)} - \min(x^{(j)})}{\max(x^{(j)}) - \min(x^{(j)})}$$

$$d_j = \text{DTW}\left(\tilde{x}, \tilde{x}^{(\sigma_j)}\right)$$

We then assess whether the model captures the expected relationship between noise level and DTW distance. As shown in Figure 11, higher noise levels correspond to higher DTW distances, indicating lower similarity. DTW scores for noise levels below 0.25 are significantly different, while confidence intervals begin to overlap beyond that point, likely due to increased uncertainty at higher noise levels. Additionally, Figure 12 provides visual examples to evaluate the model's classification. Within each row, red patterns are those identified as similar by the model, while blue patterns are classified as different. Transparency indicates the degree of similarity. The more visible, the lower the DTW distance value.

Second Test

In the second test, we generate an ARMA (AutoRegressive Moving Average) time series of 100 points using specified parameters p and q. In parallel, we generate two additional time series of 1000 points each: one ARMA series with the same parameters and one purely Random series. The initial 100-point ARMA series is divided into 10 subsequences, each with 10 time steps. For each subsequence, we

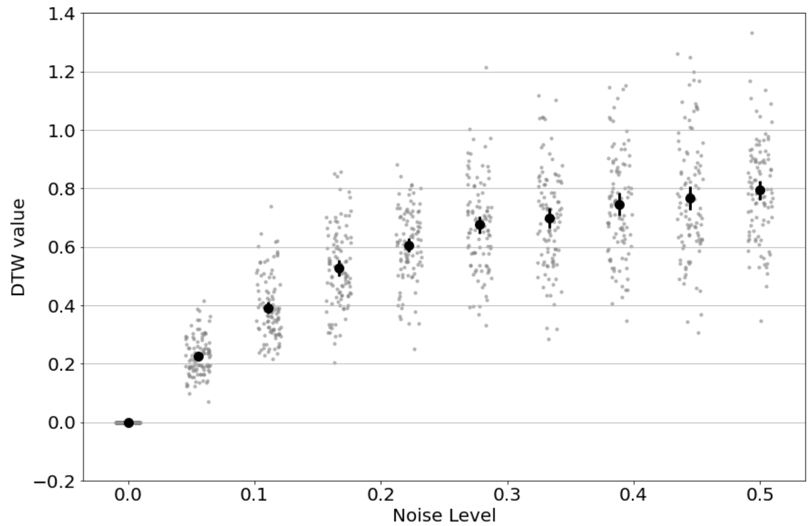


Figure 11: Mean DTW distance value (in y) for the noise introduced level (in x). The dot is the mean value of 100 sequences, and the lines represent the 95% Confidence Interval.

search for the most similar pattern within both the 1000-point ARMA series and the Random series, using DTW as the distance metric. The smallest DTW distance found in each series is kept. This procedure is iterated 1000 times to get the distributions of minimum DTW distances for both the ARMA and Random series. Figure 13 displays these distributions: ARMA in blue and Random in red. The results indicate that the ARMA series has lower minimum DTW distances. A Wilcoxon rank-sum test, appropriate due to the skewed distribution, confirms that the median minimum DTW distance is significantly lower for the ARMA series compared to the Random series.

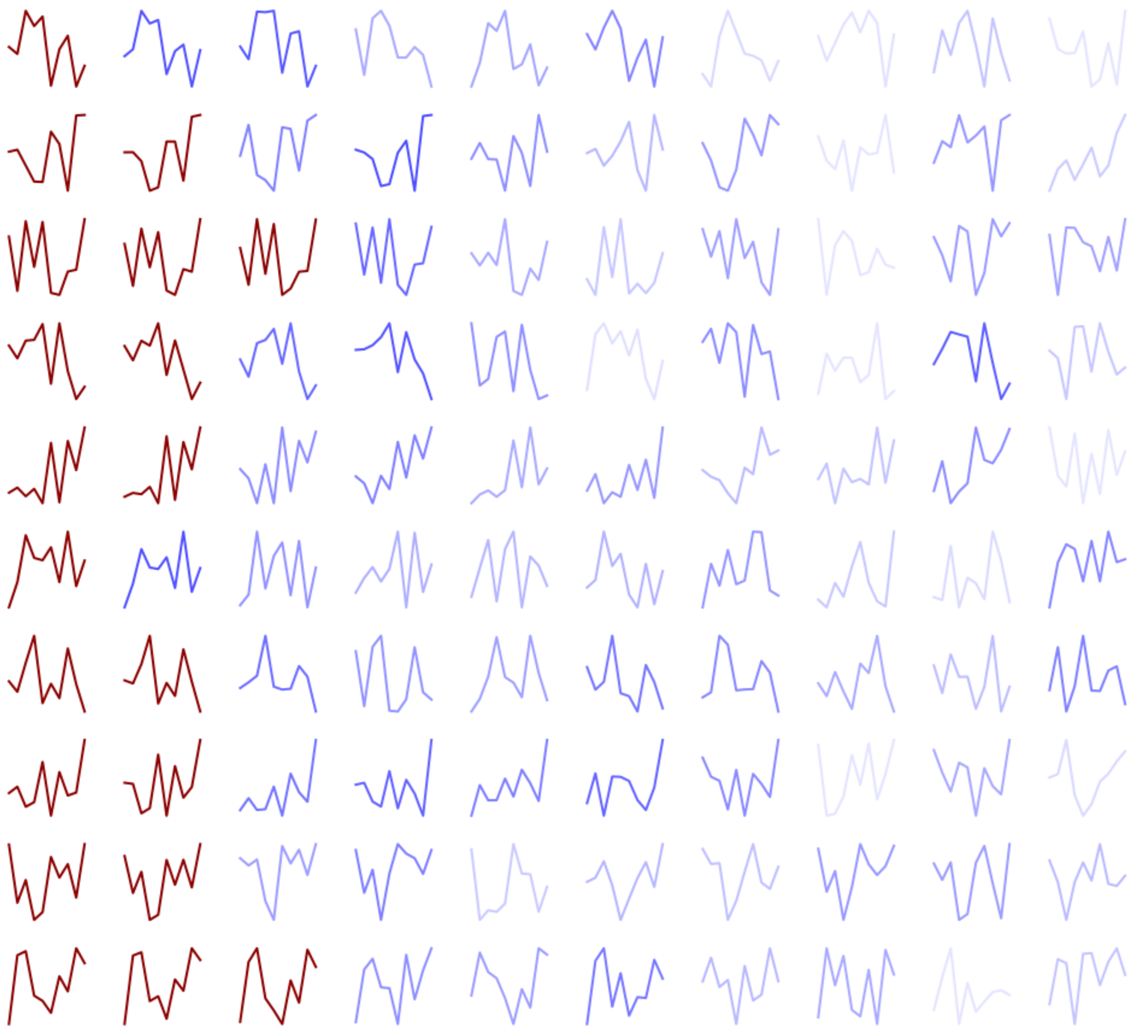


Figure 12: Ten examples of synthetic time series and their perturbed versions. The first column shows the original dynamics, while each following column to the right introduces increasing levels of noise. Red patterns are classified by the model as similar, while blue patterns are not. Transparency indicates the degree of similarity. The more visible, the lower the DTW distance value.

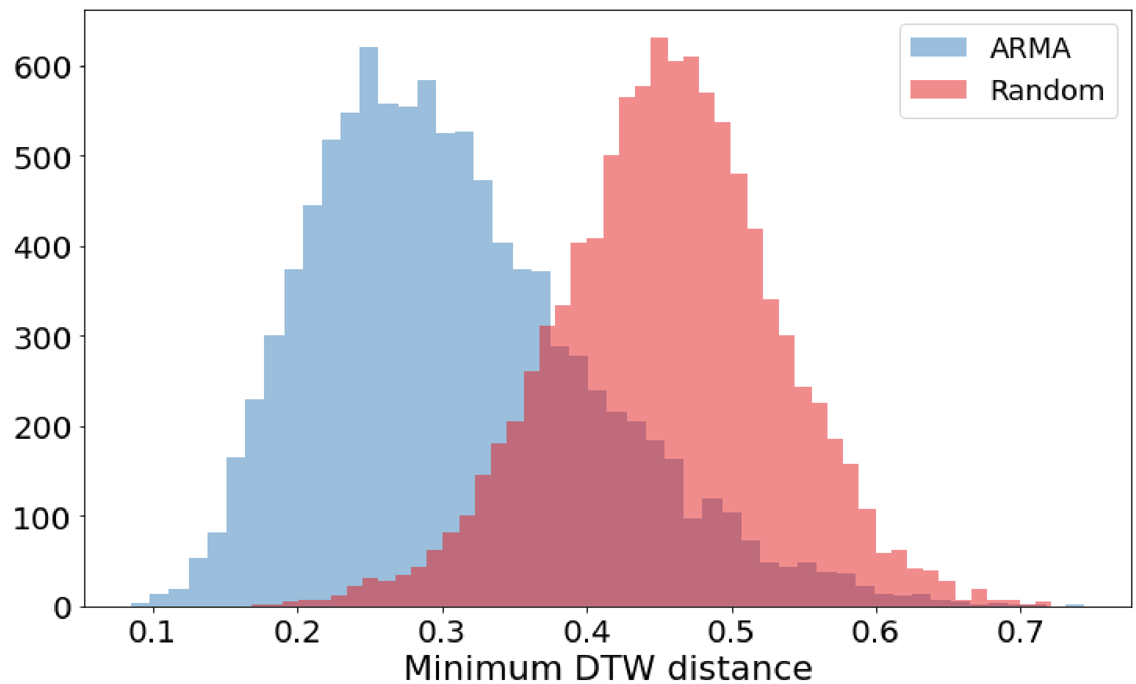


Figure 13: Distribution of the minimum DTW distance found from an ARMA 10-points subsequence to another ARMA series (in blue) and a Random series(in red).

E Additional Figures

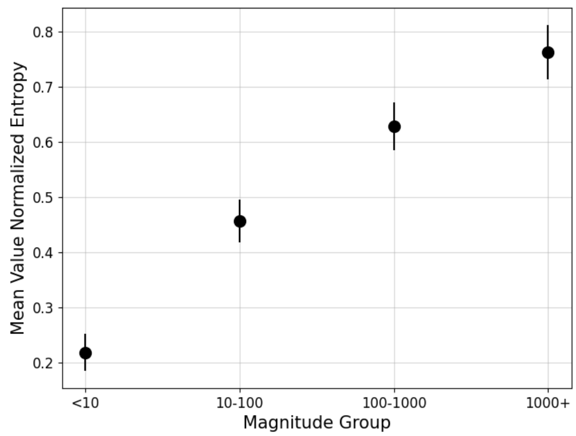


Figure 14: Mean value with 95% confidence interval of normalized entropy grouped by magnitude interval.