

The Geometry of Instability

Patterns in Conflict and Migration Flow Forecasting

Thomas Schincariol

A thesis submitted to Trinity College Dublin, the University of Dublin
in fulfilment of the requirements for the degree of

Doctor of Philosophy 2026

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Abstract

The Geometry of Instability: Patterns in Conflict and Migration Flow Forecasting

Thomas Schincariol, Trinity College Dublin

Forecasting in social sciences has progressed rapidly, driven by increased demand from decision-makers. Yet, phenomena like conflict and migration remain difficult to predict due to their complexity of mechanisms, context dependency, and reliance on multi-individual human decision-making. These dynamics often vary in speed, contain time lags, and pose challenges to rigid classical modeling approaches. This thesis introduces ShapeFinder, a shape-based forecasting model that focuses on temporal dynamics rather than static values. It uses flexible distance measures, Dynamic Time Warping (DTW) and Earth Mover’s Distance (EMD), to identify similar historical patterns, even when sequences differ in timing, scale, or context. The model generates scenario-based forecasts with probabilities and returns matched historical cases, providing transparency and interpretability crucial for practical use, unlike black-box models such as deep learning. Across country and regional scales, ShapeFinder outperforms classical statistical methods, as well as machine and deep learning models, in forecasting conflict fatalities and migration flows. In conflict settings, the model versions without covariates outperform complex models with hundreds of variables. In migration flows settings, the model outperforms classical statistical and deep learning models, with covariates showing little added value unless used dynamically. Its lightweight, flexible, and transparent method makes ShapeFinder suited for real-world early warning systems in conflict and migration flow forecasting, but also offers strong potential for academic use.

Keywords: Forecast, Conflict, Migration Flow, Shape-based Model, Pattern Recognition

Acknowledgements

First and foremost, I want to thank my supervisor, Thomas Chadeaux, for the incredible support throughout my PhD. He welcomed me into the world of social science research, which could have been intimidating, but his guidance made the transition smooth and surprisingly enjoyable. Some conspiracists wouldn't believe it, but somehow, he managed not just to stop me from hating the PhD experience, but to actually enjoy it. Beyond the academic side, his everyday support made this whole experience much easier. He was infinitely patient with my late-night work habits, odd questions, and frequent requests. Even if I now realize I have similar needs to a newborn, hopefully, this PhD is just the beginning of our collaboration.

Next, I want to thank the entire PACE team, starting with Hannah. Thank you for supporting me, sharing an office, and still managing to be a great friend and an even better coauthor. I'm also grateful to all the PhDs, postdocs I had the chance to work with, in chronological order: Jian, Adal, Yohan, Junjie, Emmanuel, Chien-Lu, and Jungming. Thanks for your support and feedback during the team meetings or daily talks. Special thanks to Jian for agreeing to play badminton with me, like a parent who agrees to play with his kid even if he doesn't want to. A big thanks also to Jemimah for her admin support, organization, and overall kindness. And thank you to Sarah Smullen for making the entire department's admin work look effortless.

I would also like to thank my PhD cohort for their help and support throughout this experience. A special thanks to Lucas, Andrej, Marina, and Jack for covering for my complete lack of participation during the first classes. Additionally, a big thanks to Liam Kneafsey for the support, especially during the Friday seminars, which can be a tough crowd if your research isn't as ethically solid as car sales forecasts.

I would also like to thank my coauthors, Yani Xue, Derek Groen, and Paola Vesco, for their time, support, and trust in working with me.

I also want to thank the mentors who first set me on the path toward this PhD. First, Ramla Sadeem, Kevin Guelton, and Maxime Colas, thank you for encouraging me to pursue an academic career, even if it's not the most conventional path for someone coming from an engineering school. Next, I want to thank Mamadou Seck and everyone in the Modeling and Simulation Engineering Department at Old Dominion University. You introduced me to a warm, supportive academic environment that made a strong first impression. Finally, I'm deeply grateful to Emmanuel Roux, Thibault Catry, María José Escorihuela, and the entire IRD Espace-Dev and IsardSAT teams. You welcomed me into an interdisciplinary setting where kindness and collaboration make great research. That experience was the final push I needed to be fully convinced to pursue an academic career.

I would also like to thank my family, starting with my parents, for their limitless support, even if they'll never read these lines since they're not in French. I also want to thank my brother Mathieu and partner Justine for welcoming me to Paris more times than I can count to escape from the wonderful Irish weather. I owe you the equivalent of several PhD grants, and I promise I'll pay it back. I'd also like to thank

my partner, Marketa, for her continuous support throughout this entire journey. Pairing with someone who can repeat the same childish joke for four years straight takes real patience, and that, in itself, is a remarkable achievement. Finally, a big thank you to Nico, who isn't actually my second partner, but might look like it given how much time we've spent together, for keeping me entertained, even from abroad, with games that probably only appeal to senior people, like Yahtzee.

Finally, I'd like to thank the Arsenal Football Club for the entertainment and the pure joy, like when Reiss Nelson scored that last-minute winner, or when Declan Rice dropped two free kicks in the Champions League quarter-final. Even without trophies, the club kept me going through tough times. I also want to thank all the writers and producers of *The Office*, which got me through these four years and still made me laugh, even after I'd watched it more times than I've seen my own parents.

Funding

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

Declaration of Generative AI in the writing process

I used ChatGPT to assist with grammar and spelling corrections, translations, and synonym suggestions during the preparation of this work. I reviewed and edited all content afterward and take full responsibility for the final manuscript.

List of Papers

The thesis relies on four papers, which are referred to later by the following numbers:

1. Schincariol, Thomas, Hannah Frank, and Thomas Chadeaux. "Accounting for variability in conflict dynamics: A pattern-based predictive model." *Journal of Peace Research*, 00223433251330790, 2025.
Contribution: Conceptualization, Methodology, Formal Analysis, Coding, Writing
2. Schincariol, Thomas. The Geometry of Conflict: 3D Spatio-Temporal Patterns in Fatalities Prediction, 2025. Manuscript in preparation for submission to the *International Journal of Forecasting*.
3. Schincariol, Thomas. Forecasting Migration Flows to Europe: A Shape-Based Modeling Approach with Covariate Analysis, 2025. Manuscript in preparation for submission to *International Migration*.
4. Schincariol, Thomas, and Thomas Chadeaux. "Temporal Patterns in Migration Flows: Evidence from South Sudan." *Journal of Forecasting* 44.2 (2025): 575-588.
Contribution: Conceptualization, Methodology, Formal Analysis, Coding, Visualization, Writing

Papers 1 and 4 are reproduced here under the terms of the Creative Commons Attribution–NonCommercial 4.0 International License (CC BY-NC 4.0), with appropriate attribution to the original publications.

Contents

1	Introduction	9
1.1	Pattern-based reasoning	10
1.2	Shape-based forecasting models	15
1.2.1	Patterns in Conflict	17
1.2.2	Patterns in Migration Flow	18
1.3	Structure of the Thesis	19
1.4	Findings	21
2	Accounting for variability in conflict dynamics: A pattern-based predictive model	34
3	The geometry of conflict : 3D Spatio-temporal patterns in fatalities prediction	74
4	Forecasting Migration Flows to Europe: A Shape-Based Modeling Approach with Covariate Analysis	115
5	Temporal Patterns in Migration Flows : Evidence from South Sudan	150
A	Appendix	182
A.1	Appendix : Paper 1	182
A.2	Appendix : Paper 2	197
A.3	Appendix : Paper 3	214
A.4	Appendix : Paper 4	219

1 Introduction

Over the past five years, the number of recorded conflict events has doubled (ACLED, 2024), and the number of international migrants has also doubled since the 1990s (UN, 2024). Long-term projections from climate experts suggest that, even under the most optimistic scenarios, these trends will continue, and may even accelerate, with rising numbers of forced migration flows and potential conflict events in the coming decades (IPCC, 2023). The lack of resources and declining land livability are expected to drive major shifts in population distribution, leaving some areas increasingly hostile to human life.

Forecasting models are becoming a key tool in decision-making processes. For example, during the COVID-19 crisis, many governments relied heavily on expert predictions to guide public health responses. Similarly, in the social sciences, policymakers and NGOs are increasingly using conflict and migration forecasting models to inform their strategies. Given the scale of the emerging challenges, continued research into forecasting models for conflict and migration flows is not only relevant but essential for effective preparation and response.

Researchers have made significant progress on forecasting models, particularly over the past few decades. However, many of the methods used appear poorly suited to the specific challenges of predicting conflict and migration flows. A key limitation is the method involved. Most existing models use static values and lack the flexibility needed to capture the complex and dynamic mechanisms underlying these phenomena. While newer approaches such as deep learning offer greater flexibility, they come with two major drawbacks: they require large datasets, which are unavailable for migration flows, and they produce outputs that are difficult to interpret, limiting their usefulness in decision-making and early warning systems. There remains a clear gap for predictive

models that can deliver strong performance while maintaining high interpretability.

In this thesis, the ShapeFinder model is introduced through four variants, adapted for both conflict and migration flow purposes. The core concept of the model is to find similar shapes to the actual ones to predict the future dynamics with the use of scenarios. The future probable scenarios are not the only useful outcome of the model. The similar patterns found in the historical dataset can be extracted for a theory study oriented, but also for an anomaly detection warning system. This element can be used as a validation for a qualitative assessment of the prediction or as an interpretability asset of the model.

The thesis is organized as follows. It begins with a discussion of the theoretical foundations of pattern-based reasoning, followed by a review of the origins and current state of shape-based forecasting models. Next, the applications of these models in conflict and migration studies are discussed. The specific features and methodologies of the four studies conducted are then presented, followed by a summary of the overall findings of the project. Finally, the full versions of the four papers are included.

1.1 Pattern-based reasoning

In 1998, the Democratic Republic of Congo (DRC, formerly Zaire) experienced a sudden rise in violence (Fig. 1, left panel) that marked the start of the Second Congo War. In 2005, Serbia (formerly part of Yugoslavia) experienced a similar event (Fig. 1, right panel), corresponding to the start of the Bosnian War. Both cases followed similar patterns: an initial surge in fatalities, followed by a pause in conflict, and finally a sudden escalation in intensity. This dissertation aims to uncover such patterns in conflict and migration flows. Although the macro-level patterns are similar, the micro-level mechanisms differ. In the DRC, the war began after President Kabila turned against the nations that had helped him rise to power, Rwanda and Uganda, within an already un-

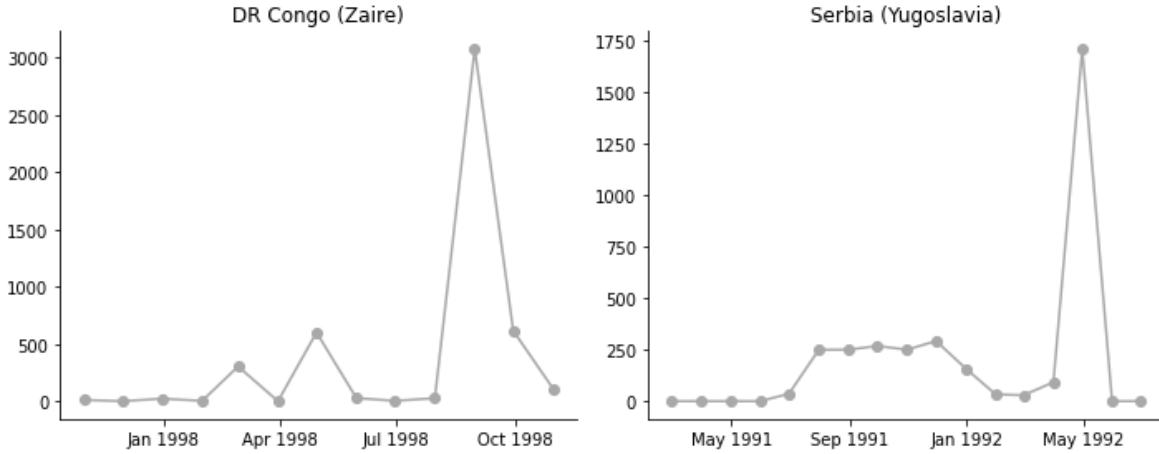


Figure 1: Conflict fatalities in DR Congo, ex-Zaire (left panel) and Serbia, ex-Yugoslavia (right panel).

stable context following the First Congo War. In contrast, the Bosnian War began after Bosnia declared independence, amid the breakup of Yugoslavia. Two different unstable contexts and distinct triggering events led to similar dynamics in conflict fatalities.

While this pattern-based approach may raise questions about its theoretical foundations, it can be initiated in the micro-level mechanisms of human behavior. Both conflict and migration fundamentally start with human decision-making and its interactions. Neuroscience research suggests that the human brain functions as a forecasting machine, constantly working to anticipate future outcomes (Bar, 2007). Drawing on patterns from past experiences and real-time sensory input, the brain aims to reduce environmental uncertainty through predictive processing (Friston, 2010). Decisions emerge from a feedback loop that weighs predicted outcomes against perceived risks and rewards, resulting in habitual behaviors. This predictive nature of human decision-making is well-established in fields like marketing, where anticipating behavior based on past patterns has proven highly effective. Extending these findings from individuals to groups, behavioral sociology has shown that human communities also display structured regularities over time (Eagle and Pentland, 2009). These conclusions sup-

port the relevance of a pattern-based approach in studying collective phenomena such as conflict and migration, where group-level behavior reflects underlying, recurring decision-making dynamics.

However, modeling these micro-level processes remains only theoretically feasible. Accurately simulating the contexts of DRC or Serbia would require an astronomical set of data, and interactions extremely context-dependent. Forecasting the decision of President Kabila to take distance from his former allies would require the entire set of interactions with his environment, combined with his internal neural network information. And it would be just a single piece of all the interactions that led to the onset of the Second Congo War. Therefore, building a mechanistic model like in physics or biology would be impossible due to the immense complexity of neuropsychological drivers behind conflict and migration flow. Simplifying such models would overlook critical elements, including emotional states and social environments (Pfister and Böhm, 2008). Moreover, conflict and migration are not the result of isolated individual choices but emerge from the dynamic interactions of multiple actors. These collective processes create highly complex systems in which individuals influence and are influenced by each other's decisions. This interdependence further complicates attempts to reduce social phenomena to fixed causal models.

Micro-level behaviors can give rise to observable patterns at the macro level, a concept known as emergence theory. Originally introduced by philosopher George Lewes in the 19th century (Lewes, 1877), this theory forms the foundation of agent-based modeling, which is widely used to study complex macro-level phenomena by simulating interactions at the individual level. The relationship between micro- and macro-dynamics has been explored across various disciplines. Emergence theory has demonstrated strong results in fields such as ecology, through predator-prey models, or in epidemiology, with the study of vector-borne diseases. In the malaria example, the micro-level climatic pat-

terns shape the mosquito’s density. And then, the mosquito interaction with human mobility dictates the number of malaria cases, at the macro-level, in time and space.

However, one might argue that emergence theory does not fit all phenomena, especially in social science, that involve strategic actors and not only physical or environmental phenomena. Economics, and notably the stock market prices, seem to be unpredictable, even though markets are driven by human decisions to buy or sell. Although the question of emergence is still actively puzzling the economists (Farmer and Foley, 2009), some researchers suggest that the high speed of market fluctuations (Bouchaud, Farmer, and Lillo, 2009), combined with the influence of algorithmic trading (Johnson et al., 2013), disturb the formation of stable macro-level patterns.

In contrast, social scientists have studied macro-level spatio-temporal patterns, for example, with tools such as cellular automata (CA) to study social interactions (Hegselmann and Flache, 1998). CA is a discrete model of computation focused on cells and their states through time. The most famous CA is Conway’s ‘game of life’ (Conway et al., 1970), where all cells of a two-dimensional grid are dead or alive. Their state at time t is defined based on the state of its eight neighbors at $t - 1$. If a live cell has two or three live neighbors, it stays alive; with more or fewer, it dies. If a dead cell is surrounded by exactly three live neighbors, it becomes a live cell. Even with the simplicity of these micro-level rules, CA can generate complex spatio-temporal patterns. An example is presented in Figure 2, representing the pattern “Achim’s p16” found by Achim Flammenkamp in July 1994.

By applying, step by step, the rules of the ‘game of life’ on the initial state, at $t = 0$ (upper-left panel of the figure), the pattern originally composed of four parts transforms into a flower shape pattern (from $t = 4$ to $t = 6$), to finally comes to a mirror state of the original state at $t = 8$. By repeating the operation until $t = 16$, the pattern is back to its original state, and then can be classified as a 16-period oscillator.

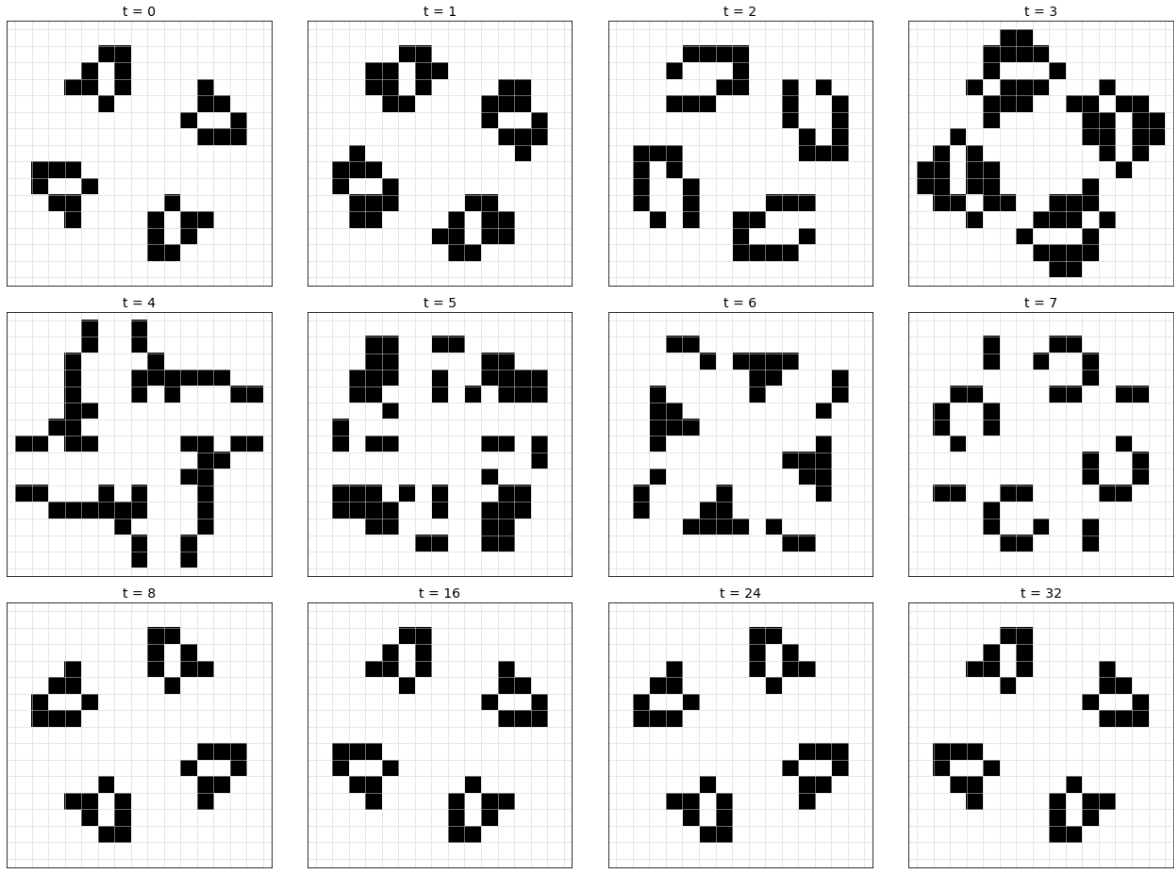


Figure 2: Achim's p16 patterns representation from $t = 0$ (upper left panel) to $t = 32$ (bottom right panel) in the 'game of life'.

This is a relatively simple pattern, but thousands of more complex macro-level patterns have been found, and are found every day, even if the game has basic micro-level rules. Some patterns oscillate (called oscillators), move in space (called spaceships), or even generate spaceships (called guns). In 2023, almost 55 years after the creation of the game, the 'game of life' has been proven to be omniperiodic Brown et al., 2023, meaning it has oscillators for every period possible. In other words, it exists contexts where macro-level patterns repeat themselves with all possible frequency, only ruled by the micro-level mechanisms of the game.

CA models were first used in biology, chemistry, or physics, but also in social science to study social interactions (Sakoda, 1971) or even segregation phenomenon (Schelling,

1971). The presence of emergent dynamics in social science, particularly in conflict and migration studies, is more widely accepted. In conflict research, regularities in strategic behavior (Axelrod and Hamilton, 1981), agent-based modeling for conflict studies (Cederman and Girardin, 2023), or the predictive power of early warning signals such as media coverage (Chadefaux, 2014), offer strong support for macro-level pattern formation. Similarly, in migration studies, researchers have begun employing agent-based models to simulate and forecast migration flows in specific contexts (Ghorbani et al., 2024), further reinforcing the validity of large-scale micro-level patterns formation, resulting from individual modeling.

This study focuses on identifying and distinguishing those macro-level patterns, using shape-based modeling techniques. Unlike traditional point-estimate models, such as linear regression, shape-based approaches capture the dynamics of entire sequences rather than isolated values.

1.2 Shape-based forecasting models

Shape-based forecasting models are largely derived from pattern recognition methods. Originally developed to identify and analyze patterns in data, these methods have been adapted for forecasting. Pattern recognition was initially developed in the 1960s for applications such as speech recognition (Pruzansky, 1963) and optical character recognition (Grimsdale et al., 1959). Over time, it has been extended to broader fields, including fraud detection in finance (Albashrawi, 2016), medical diagnosis using images (Lo et al., 1995), and the identification of extreme climate events (Racah et al., 2017). In the last decade, several advances facilitated their use in forecasting tasks. The matrix profile (Yeh, Zhu, et al., 2016) computes the similarity between all subsequences of a time series, enabling the detection of repeating sequences, known as motifs (Chiu, Keogh, and Lonardi, 2003), as well as unusual patterns, referred to as discords or

anomalies (Keogh, J. Lin, et al., 2007). Based on the core assumption that history tends to repeat itself, shape-based forecasting models follow a straightforward principle: identify similar past patterns as current ones to produce predictions about the future.

One of the main challenges is to accurately measure pattern similarity. For time series subsequences, the most common method to measure the distance between two subsequences is the Euclidean distance. While simple and easy to compute, this measure has limitations. Because it sums point-to-point distances, it only works well when comparing sequences of the same length, and even slight misalignments can result in large distance values. Alternatives like the Manhattan distance (Malkauthekar, 2013), which focuses on differences rather than exact values, also require perfect alignment. These are part of the “lock-step” class of methods, which compare sequences point by point. To address these limitations, more flexible approaches known as “elastic distance” methods were developed. One widely used example is Dynamic Time Warping (DTW) (Berndt and Clifford, 1994), which allows time shifts between points, making it possible to compare misaligned but similar sequences. Other specialized methods compare the parameters of fitted models (Abanda, Mori, and Lozano, 2019), or, more recently, use deep learning approaches such as transformers (Wen et al., 2022). These more complex models are already well established for high-dimensional data like images. Image comparison has evolved from direct pixel-based methods to feature-based comparisons (Scholz et al., 2025), or using distances in embedding space (Barz and Denzler, 2019). Some traditional metrics, like Earth Mover’s Distance (EMD) (Rubner, Tomasi, and Guibas, 2000), assess similarity by computing the minimum cost required to transform one distribution into another.

Once distances are computed, pattern classification is typically performed using unsupervised or semi-supervised methods (Schwenker and Trentin, 2014). In the case of time series subsequences, the effectiveness of clustering techniques is still debated

(Keogh and J. Lin, 2005; Zolhavarieh, Aghabozorgi, and Teh, 2014). Some studies suggest that traditional approaches, such as k-means or hierarchical clustering, often produce clusters that are no better than random. However, newer methods, such as those based on matrix profiles (Yeh, Van Herle, and Keogh, 2016) or embeddings generated by autoencoders (Yu, Kim, and Mechefske, 2021), have renewed interest in classifying and identifying patterns in time series subsequences. Besides this, symbolic representations have been developed to simplify and bring symbolic representations of time series subsequences. One example is Symbolic Aggregate approXimation (SAX) (J. Lin et al., 2007), which converts a numerical time series into a symbolic string, enabling more efficient compact motif discovery or, more generally, classification of subsequences.

Several studies use pattern classification methods as the foundation for prediction. These approaches have been applied across various fields: for example, electric load forecasting using DTW-based clustering (Teeraratkul, O’Neill, and Lall, 2017), the use of discriminative subsequences, called shapelets, for epidemic forecasting (Srivastava, Singh, and Lee, 2022), and shape-based features for predicting solar flares (Deshmukh et al., 2021). The use of shape-based methods is not limited to time series. In a medicine application study, body shape features are extracted from 3D body scans to predict body fat percentage (Y. Lu, McQuade, and Hahn, 2018).

1.2.1 Patterns in Conflict

Pattern recognition has also been explored in conflict studies. Particularly in forecasting applications, Chadeaux was one of the first to apply pattern recognition methods to conflict forecasting (Chadeaux, 2022), introducing a model that later evolved into the ShapeFinder framework. Its model and adaptations are discussed in the first paper. While a range of models has since been tested, most forecasting approaches in this field continue to rely on statistical methods (Brandt, Freeman, and Schrodtt, 2014;

Schrodt, 2006) or machine learning techniques (D’Orazio and Y. Lin, 2022; Colaresi and Mahmood, 2017). In recent years, models have become increasingly complex in an effort to better capture the mechanisms of conflict. However, this added complexity often comes at the cost of interpretability. Newer models, such as transformers or even simpler neural networks, can be difficult for policymakers to understand and trust. A key takeaway from a prediction challenge organized by the Violence & Impacts Early-Warning System (VIEWS) team was the need to go towards more complex and flexible models, but also interpretable (Hegre, Vesco, and Colaresi, 2022). For instance, in the next iteration of the prediction competition (Hegre, Vesco, Colaresi, et al., 2024), the forecasted outputs shifted from point estimates to probabilistic values, more easily actionable for decision-makers.

Although the use of pattern-based approaches in forecasting is still relatively new, patterns have long been studied in conflict research at various levels. The idea of applying pattern recognition to better understand conflict emerged in the early 2000s (Brecke, 1998). Long-term cycles of violence have been analyzed across different regions (Brecke, 2001), alongside more recent trends involving non-state actors and the increasing role of regional and international players (Avis, 2019). More practical patterns have also been studied with more fine-grained studies looking at the pattern of conflict’s diffusion (Schutte and Weidmann, 2011), distinguishing two types of patterns: the relation, or shift of violence, and the escalation of conflict, where conflict expands spatially.

1.2.2 Patterns in Migration Flow

Similar to conflict, patterns in human migration flows have been extensively studied. Research ranges from global analyses of major international migration patterns (Abel and Sander, 2014) to more specific case studies, such as those focused on the United States (Holland and Plane, 2001) or Bangladesh (X. Lu et al., 2016). In addition, mi-

gration patterns linked to specific causes or events have been explored—for instance, the effects of environmental factors (Xu and Famiglietti, 2023) or the impact of epidemics, such as COVID-19, on migration flows (D. Lu et al., 2021).

The forecasting literature on migration flows remains limited, primarily due to the lack of high-quality, continuous data. Most available data are at the country level and reported annually. Finer-grained data, such as monthly or regional, are difficult to obtain given the cost of collection. Typically, data are collected through surveys in refugee camps or border registrations along major migration routes. This data issue is acknowledged as a major challenge in the field (Nicoletti et al., 2023), and several studies investigate the potential of new sources of data (Tjaden, 2021). Given these constraints, it is predictable that shape-based models, which require high-quality time series data to detect informative patterns (though not necessarily large volumes of data like deep learning models), are underused. Instead, more traditional approaches such as classical statistical models (Welch and Raftery, 2022), machine learning (Carammia, Iacus, and Wilkin, 2022), and deep learning techniques (Golenvaux et al., 2020) have been applied. These models have been used across various types of migration, from international movements (Bijak et al., 2019) to internal displacement (Huynh and Basu, 2020). Still, forecasting migration remains highly challenging. A recent review of climate-related migration forecasting models concluded that current approaches are “not yet able to provide reliable numerical estimates” (Schewel et al., 2024), and concludes in the need for new data collection strategies and a more scenario-based approach, aligning with the conflict forecasting field objectives.

1.3 Structure of the Thesis

As mentioned earlier, the model developed in this thesis builds on a previous version introduced in (Chadefaux, 2022), during the first iteration of the VIEWS forecasting

competition (Hegre, Vesco, and Colaresi, 2022). Significant adaptations were made, resulting in the model presented in **Paper 1**. The original model compared an input sequence to the entire historical dataset, resulting in a set of distances. Those distances were used to weigh the importance of historical sequences in the creation of forecasts. This implies a major issue: all patterns were considered in the prediction creation. Even if their weight is small, the noise introduced led the model to forecast mostly flat values, losing the ability to predict complex dynamics. In the adapted versions presented in **Paper 1-4**, the model only focuses on the most similar patterns. Another key change is the prediction creation method. In the original model, the forecast was the average value of the following values of all historical patterns weighted by their similarity. This approach is maintained in **Paper 4**, as the prediction is the average of the following values of only the sequences classified as similar. In contrast, **Paper 1-3** introduce the scenario-based forecast. After identifying similar patterns, their following values are clustered to produce distinct potential scenarios. This allows more interpretability for decision-makers and opens the door to theoretical analysis by comparing different scenario paths. The key attributes of each Paper’s model are displayed in Table 1.

	Paper 1	Paper 2	Paper 3	Paper 4
Field	Conflict	Conflict	Migration	Migration
Spatial unit	Country	PrioGrid	Dyad-Country	Region
Shapes	2D	3D	2D	Multi 2D
Distance	DTW	EMD	DTW	DTW
Covariates	No	No	Static	Dynamic
Benchmark	Views	Views	RF	ARIMA,LSTM
Horizon	12 months	6 months	6 months	3 months

Table 1: Key Attributes Across Papers

Papers 1–2 present purely autoregressive models focused on forecasting conflict

fatalities, while **Papers 3–4** outcome variable is the migration flow. **Paper 3** operates at the dyad-country level, accounting for both the origin and the arrival country in Europe. **Paper 4** takes a more fine-grained approach, forecasting departures from specific regions within South Sudan, covering both internal displacement and international migration. A key methodological innovation appears in **Paper 2**, where the model operates on three-dimensional shapes at the PRIO-Grid spatial level, a worldwide grid of 0.5 x 0.5 decimal degrees (Tollefsen, Strand, and Buhaug, 2012). Here, the Earth Mover’s Distance (EMD) is used to measure similarity, in contrast to the Dynamic Time Warping (DTW) metric used in the other papers. **Paper 4** also applies DTW, but introduces a multilayer variant to handle covariates in the format of shapes. In contrast, **Paper 3** uses a more standard method for incorporating covariates. In terms of benchmark models, since there is currently no widely recognized benchmark model for migration forecasting in the literature, unlike the Views model (Hegre, Akbari, et al., 2022) in conflict forecasting, **Paper 3** is compared against a Random Forest (RF) machine learning model, while **Paper 4** compares its performance to both a statistical model, ARIMA, and a deep learning model, Long Short-Term Memory (LSTM). Finally, the forecast horizon is set at one year for **Papers 1**, six months for **Papers 2–3**, while **Paper 4** uses a shorter three-month horizon due to a more limited input window and restricted availability of historical data.

1.4 Findings

The first key finding is the strong performance of the model in forecasting conflict fatalities and migration flows. It outperforms established benchmarks, but also traditional statistical methods, machine learning models, and even deep learning models. Its accurate predictions make it a strong candidate for forecasting tasks.

This model holds a significant interest and has a distinct place in the forecasting

literature, particularly when transparency is needed and when the data is limited. First, when transparency in forecast production is required, this model stands out. It allows users to fully trace how a forecast is generated and to understand the broader context behind it. In this sense, the model offers a valuable alternative for those who need complete traceability, something that most complex models do not provide. This model also demonstrates strong performance when the unit of interest has limited available data, such as migration flows. Unlike machine learning models, or even large language models, it does not require a formal training phase, meaning it can perform well without a large quantity of data.

In addition, the model performs well when dealing with highly zero-inflated and skewed data, such as conflict data. Because it focuses only on the most similar cases, the forecast values are not influenced by unrelated parts of the dataset. This feature enables the model to produce non-conservative forecasts, even when most of the values are zero or very small values. One of the main challenges in conflict forecasting is the tendency of models to generate overly conservative predictions, as shown in Paper 1. Since most models are optimized using metrics like Mean Squared Error, zero-inflated data often push them toward conservative predictions. The structure of this model, based on selecting and learning from the most similar cases, effectively bypasses this limitation.

Another major challenge in conflict forecasting concerns the prediction of onsets. For decision-makers, identifying the start of new conflicts is one of the most valuable forecasting outcomes. Unfortunately, none of the existing models in the field has proven effective at detecting onsets, and this model is no exception. The versions of the model developed for conflict forecasting (Papers 1 and 2) are autoregressive models. Because the model is purely autoregressive, it is structurally impossible to predict the onset of new conflicts. As discussed in the papers, no model, including the VIEWS model

with its extensive set of covariates, has yet reliably achieved this task. Nonetheless, adaptations, such as the inclusion of covariates, as explored in Papers 3 and 4, could be developed alongside the current model using a hurdle framework. These covariates could include variables that have variation even in peaceful times.

That said, most models using such data, like news reports, tend to face a high rate of false positives or false alarms. In many cases where indicators suggest rising conflict, only a small fraction actually escalates to war. The key future challenge in onset forecasting may therefore lie in distinguishing false positives from true ones. While the ShapeFinder model cannot currently predict onsets, it performs strongly once a conflict is ongoing, particularly in situations with high recent activity, as demonstrated in the results of Paper 1. Future versions of the model should therefore include a hurdle component to address this limitation when the input data contain only zeros.

Additionally, the simplicity of the model also introduces several limitations. The core assumption of the model might be limited theoretically in some cases. The model assumes that the system studied does not vary over time: patterns observed twenty years ago produce similar outcomes nowadays. For example, migration patterns in South Sudan in 2020, shaped by the COVID-19 pandemic, may differ substantially from those in 2023, influenced by the neighboring conflict in Sudan. Similarly, the model does not explicitly account for seasonality. These issues, however, can be addressed by modifying the pattern search process. The model could, for theoretical or seasonal reasons, give more weight to a more recent or seasonally relevant subset of the data to better fit the changes in the system over time.

Linked to these potential theoretical and seasonal issues, comparing patterns of completely different magnitudes could be seen as a clear weakness. Since all sequences are Min–Max normalized, a sequence such as $[0, 1000, 0, 1000, 0]$ would perfectly match $[0, 1, 0, 1, 0]$, even though they are theoretically quite different. As with the seasonality

problem, a possible improvement would be to introduce a magnitude weight to prevent sequences with very different magnitudes from matching. However, we observe that, in practice, the model naturally matches patterns within their respective magnitude categories (see Table in Appendix 1). For instance, most “peak” shapes, where one or two values differ from zero, belong to the small conflict magnitude category. It is highly unlikely to find a sequence alternating between months with 0 and 1000 fatalities in the real world, as high magnitude conflicts gradually increase to such high values.

In addition, switching from Min–Max normalization to z-normalization does not affect the results, likely because the natural patterns already match by magnitude. Similarly, although many conflict forecasting models use log-transformed data to reduce skewness, applying such a transformation here would have little impact. Log transformation would flatten the patterns and alter their shape, but this is unlikely to matter in practice, since real-world conflict data rarely contain sequences that jump abruptly between months with 0 and 1000 fatalities. High-magnitude conflicts typically rise gradually, which limits the influence of extreme values.

The interpretability of the model may also be limited in certain contexts, particularly when dealing with complex patterns such as the large three-dimensional patterns discussed in Paper 2 or the multiple-pattern structures in Paper 4. In the case of three-dimensional patterns, human understanding of spatial relationships is inherently limited, especially when these patterns are presented in two dimensions, such as on paper. This reduces the interpretability of the model, as the full structure of the pattern cannot be easily grasped. In such cases, only the contextual information from the matching patterns, such as their location and timing, remains directly interpretable. Similarly, patterns involving multiple variables, as in Paper 4, are difficult for humans to comprehend, particularly when some variables are lagged or evolve at different speeds. While this flexibility improves the model’s performance by aligning dynamic behaviors,

it simultaneously reduces interpretability.

Finally, one potential limitation of the model lies not in the model itself but in the data. The quality of conflict and migration flow event data has been, and continues to be, a central topic of discussion within the conflict studies community (Eck, 2012; Gleditsch, Metternich, and Ruggeri, 2014). In the case of conflict, datasets are typically derived from news reports, which introduces potential issues that can affect the accuracy of model predictions.

First, event datasets depend heavily on the quality and extent of news coverage. It is well established that conflict event data suffer from reporting bias (Weidmann, 2016). This bias often results in imprecise information regarding both the location and timing of events, with fatality counts frequently reported in aggregated form over several months. Such imprecision poses a problem for models like ours, which are sensitive to the month-to-month dynamics of conflict fatalities. At present, most forecasting models (ShapeFinder included) address this issue by excluding events that span multiple months, as the available information remains too coarse for reliable use.¹

Errors can also arise during the extraction of information from news reports into geo- and time-coded events. The UCDP dataset, used in all three papers employing conflict fatalities data, is among the most reliable and accessible sources available. Nevertheless, given the vast amount of information that must be coded and updated each month, occasional errors are almost inevitable. For instance, in a previous version of the dataset, a single event in Ethiopia was recorded as causing 30,000 fatalities in one day, when in fact these fatalities occurred over a period of five months. Although such errors are typically corrected quickly in subsequent releases ², they highlight the challenges inherent in large-scale data collection and coding.

¹An alternative approach is to distribute the fatalities over the months covered by the event. The VIEWS forecasting model has been using this approach since their last update in October 2025. (Link)

²In this case, in less than a couple of days.

This type of error is particularly problematic for our model, which is highly sensitive to sudden changes in the data. In the example above, the error produces a “peak” effect after normalization: the 30,000 fatalities become an artificial spike, while the preceding and following values are effectively flattened to near zero. As a result, a false pattern is created, one that never occurred in reality, while the genuine temporal variation around the event is lost. Our model is particularly sensitive to these problems, but the data quality affects all forecasting models to some degree, even the most robust ones. Unfortunately, there is little that can be done at the modeling stage to correct for such issues. The best approach is to continue supporting data providers like UCDP in their ongoing efforts to ensure clean and accurate datasets.

Migration flow data face a different set of challenges, often even more complex than those affecting conflict event data. Most migration flow information is collected directly in the field, at refugee camps, along major roads, or at border crossings. This process is extremely costly, as it requires continuous human monitoring of population movements over the period of interest. As a result, the data often lack fine-grained and consistent global coverage, with varying sampling frequencies depending on whether monitoring programs are active in specific areas and time periods.

Consequently, it is difficult to obtain long time series with consistent temporal frequency, especially at sub-annual scales. This poses a particular problem for the ShapeFinder model, as missing values (NAs) are widespread in the dataset, preventing the extraction of a continuous dynamic. Some potential solutions exist, such as resampling at a higher frequency, which, however, leads to a loss of information, or interpolating missing values, which introduces strong assumptions that are acceptable for NAs robust models but problematic for shape-based models.

More broadly, this thesis could also be the starting point of a new discussion on how covariates are used in forecasting conflict and migration flows. While covariates

are essential in causal studies, the pieces of evidence presented in the thesis (conflict autoregressive models performing well, and the Paper 3 conclusion) raise the question of their effectiveness under traditional modeling. Most commonly used covariates are static or change slowly over time, usually yearly, which makes them of limited use for predictive purposes. The extremely complex and context-dependent nature of conflict may explain why such covariates generally add more noise than useful information. The issue probably originates not only in the type of covariates selected but also in how they are incorporated into models. A better approach might be to focus on short-term, fast-changing variables, like protests, news reports, or event-based indexes, that happen shortly before conflict or migration flow. Long-term factors like climate or the economy are the source of these events, but they are certainly too complex and indirect to be useful in a forecasting model. By using signals that show up just before a crisis, we might avoid the complicated, context-specific causes that are almost impossible to model properly.

In addition to its academic contributions, this study has led to several practical applications. The model presented in Paper 1 was adapted into a live monthly prediction framework. The results are published through a monthly report summarizing key findings, as well as a web application³ that provides detailed forecasts. The project has been running for over a year, and the first live performance report⁴ confirms the model's strong performance in real-time settings. The model will also be adapted for use in a warning system as part of an upcoming collaboration with the European Union through the INFORM team, and also potentially in a separate partnership with the Danish Refugee Council (DRC). Additionally, the model and its adaptations are available as a Python package⁵. The full code and documentation are provided in the

³Available at: <https://pace-risk-map-x35exdywcq-ue.a.run.app/>

⁴Available at: <https://www.forecastlab.org/predictionmap>

⁵Available at: <https://github.com/ThomasSchinca/shapefinder>

repository, making it accessible for anyone interested in applying or further developing the model.

References

- Abanda, Amaia, Usue Mori, and Jose A Lozano (2019). “A review on distance based time series classification”. In: *Data Mining and Knowledge Discovery* 33.2, pp. 378–412.
- Abel, Guy J and Nikola Sander (2014). “Quantifying global international migration flows”. In: *Science* 343.6178, pp. 1520–1522.
- ACLED (Dec. 2024). *Conflict Index: December 2024*. Accessed: 2025-04-28. URL: <https://acleddata.com/conflict-index/>.
- Albashrawi, Mousa (2016). “Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015”. In: *Journal of Data Science* 14.3, pp. 553–569.
- Avis, William (2019). “Current trends in violent conflict”. In.
- Axelrod, Robert and William D Hamilton (1981). “The evolution of cooperation”. In: *science* 211.4489, pp. 1390–1396.
- Bar, Moshe (2007). “The proactive brain: using analogies and associations to generate predictions”. In: *Trends in cognitive sciences* 11.7, pp. 280–289.
- Barz, Björn and Joachim Denzler (2019). “Hierarchy-based image embeddings for semantic image retrieval”. In: *2019 IEEE winter conference on applications of computer vision (WACV)*. IEEE, pp. 638–647.
- Berndt, Donald J and James Clifford (1994). “Using dynamic time warping to find patterns in time series”. In: *Proceedings of the 3rd international conference on knowledge discovery and data mining*, pp. 359–370.
- Bijak, Jakub et al. (2019). “Assessing time series models for forecasting international migration: Lessons from the United Kingdom”. In: *Journal of Forecasting* 38.5, pp. 470–487.
- Bouchaud, Jean-Philippe, J Doyne Farmer, and Fabrizio Lillo (2009). “How markets slowly digest changes in supply and demand”. In: *Handbook of financial markets: dynamics and evolution*. Elsevier, pp. 57–160.
- Brandt, Patrick T, John R Freeman, and Philip A Schrodtt (2014). “Evaluating forecasts of political conflict dynamics”. In: *International Journal of Forecasting* 30.4, pp. 944–962.
- Brecke, Peter (1998). “Finding harbingers of violent conflict: Using pattern recognition to anticipate conflicts”. In: *Conflict Management and Peace Science* 16.1, pp. 31–56.
- (2001). “The long-term patterns of violent conflict in different regions of the world”. In: *Uppsala conflict data conference, Uppsala, Sweden*.
- Brown, Nico et al. (2023). “Conway’s Game of Life is Omnipariodic”. In: *arXiv preprint arXiv:2312.02799*.
- Carammia, Marcello, Stefano Maria Iacus, and Teddy Wilkin (2022). “Forecasting asylum-related migration flows with machine learning and data at scale”. In: *Scientific Reports* 12.1, p. 1457.
- Cederman, Lars-Erik and Luc Girardin (2023). “Computational approaches to conflict research from modeling and data to computational diplomacy”. In: *Journal of Computational Science* 72, p. 102112.

- Chadefaux, Thomas (2014). “Early warning signals for war in the news”. In: *Journal of Peace Research* 51.1, pp. 5–18.
- (2022). “A shape-based approach to conflict forecasting”. In: *International Interactions* 48.4, pp. 633–648.
- Chiu, Bill, Eamonn Keogh, and Stefano Lonardi (2003). “Probabilistic discovery of time series motifs”. In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 493–498.
- Colaresi, Michael and Zuhair Mahmood (2017). “Do the robot: Lessons from machine learning to improve conflict forecasting”. In: *Journal of Peace Research* 54.2, pp. 193–214.
- Conway, John et al. (1970). “The game of life”. In: *Scientific American* 223.4, p. 4.
- D’Orazio, Vito and Yu Lin (2022). “Forecasting conflict in Africa with automated machine learning systems”. In: *International Interactions* 48.4, pp. 714–738.
- Deshmukh, Varad et al. (2021). “Shape-based feature engineering for solar flare prediction”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 17, pp. 15293–15300.
- Eagle, Nathan and Alex Sandy Pentland (2009). “Eigenbehaviors: Identifying structure in routine”. In: *Behavioral ecology and sociobiology* 63, pp. 1057–1066.
- Eck, Kristine (2012). “In data we trust? A comparison of UCDP GED and ACLED conflict events datasets”. In: *Cooperation and Conflict* 47.1, pp. 124–141.
- Farmer, J Doyne and Duncan Foley (2009). “The economy needs agent-based modelling”. In: *Nature* 460.7256, pp. 685–686.
- Friston, Karl (2010). “The free-energy principle: a unified brain theory?” In: *Nature reviews neuroscience* 11.2, pp. 127–138.
- Ghorbani, Maziar et al. (2024). “Flee 3: Flexible agent-based simulation for forced migration”. In: *Journal of Computational Science* 81, p. 102371.
- Gleditsch, Kristian Skrede, Nils W Metternich, and Andrea Ruggeri (2014). “Data and progress in peace and conflict research”. In: *Journal of Peace Research* 51.2, pp. 301–314.
- Golenvaux, Nicolas et al. (2020). “An LSTM approach to forecast migration using Google Trends”. In: *arXiv preprint arXiv:2005.09902*.
- Grimsdale, RL et al. (1959). “A system for the automatic recognition of patterns”. In: *Proceedings of the IEE-Part B: Radio and Electronic Engineering* 106.26, pp. 210–221.
- Hegre, Håvard, Forogh Akbari, et al. (2022). *Forecasting fatalities*.
- Hegre, Håvard, Paola Vesco, and Michael Colaresi (2022). “Lessons from an escalation prediction competition”. In: *International Interactions* 48.4, pp. 521–554.
- Hegre, Håvard, Paola Vesco, Michael Colaresi, et al. (2024). “The 2023/24 VIEWS Prediction Challenge: Predicting the Number of Fatalities in Armed Conflict, with Uncertainty”. In: *arXiv preprint arXiv:2407.11045*.
- Hegselmann, Rainer and Andreas Flache (1998). “Understanding complex social dynamics: A plea for cellular automata based modelling”. In: *Journal of Artificial Societies and Social Simulation* 1.3, p. 1.

- Holland, Sandra C and David A Plane (2001). “Methods of mapping migration flow patterns”. In: *Southeastern Geographer* 41.1, pp. 89–104.
- Huynh, Benjamin Q and Sanjay Basu (2020). “Forecasting internally displaced population migration patterns in Syria and Yemen”. In: *Disaster medicine and public health preparedness* 14.3, pp. 302–307.
- IPCC (2023). *Climate Change 2023: Synthesis Report. Longer Report*. Accessed: 2025-04-28. Intergovernmental Panel on Climate Change (IPCC). URL: <https://www.ipcc.ch/report/ar6/syr/>.
- Johnson, Neil et al. (2013). “Abrupt rise of new machine ecology beyond human response time”. In: *Scientific reports* 3.1, p. 2627.
- Keogh, Eamonn and Jessica Lin (2005). “Clustering of time-series subsequences is meaningless: implications for previous and future research”. In: *Knowledge and information systems* 8, pp. 154–177.
- Keogh, Eamonn, Jessica Lin, et al. (2007). “Finding the most unusual time series subsequence: algorithms and applications”. In: *Knowledge and Information Systems* 11, pp. 1–27.
- Lewes, George Henry (1877). *Problems of life and mind*. Vol. 2. Trübner & Company.
- Lin, Jessica et al. (2007). “Experiencing SAX: a novel symbolic representation of time series”. In: *Data Mining and knowledge discovery* 15, pp. 107–144.
- Lo, Shih-Chung B et al. (1995). “Artificial convolution neural network for medical image pattern recognition”. In: *Neural networks* 8.7-8, pp. 1201–1214.
- Lu, Debin et al. (2021). “Spatiotemporal patterns and influencing factors of human migration networks in China during COVID-19”. In: *Geography and Sustainability* 2.4, pp. 264–274.
- Lu, Xin et al. (2016). “Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh”. In: *Global Environmental Change* 38, pp. 1–7.
- Lu, Yao, Scott McQuade, and James K Hahn (2018). “3d shape-based body composition prediction model using machine learning”. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 3999–4002.
- Malkauthekar, MD (2013). “Analysis of euclidean distance and manhattan distance measure in face recognition”. In: *Third International Conference on Computational Intelligence and Information Technology (CIIT 2013)*. IET, pp. 503–507.
- Nicoletti, Alessandro et al. (2023). “Human mobility in the context of environmental and climate change: Recent data collection tools from the International Organization for Migration to address key methodological and conceptual issues”. In: *International Migration* 61.5, pp. 47–59.
- Pfister, Hans-Rüdiger and Gisela Böhm (2008). “The multiplicity of emotions: A framework of emotional functions in decision making”. In: *Judgment and decision making* 3.1, pp. 5–17.
- Pruzansky, Sandra (1963). “Pattern-Matching Procedure for Automatic Talker Recognition”. In: *The Journal of the Acoustical Society of America* 35.3, pp. 354–358.

- Racah, Evan et al. (2017). “Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events”. In: *Advances in neural information processing systems* 30.
- Rubner, Yossi, Carlo Tomasi, and Leonidas J Guibas (2000). “The earth mover’s distance as a metric for image retrieval”. In: *International journal of computer vision* 40, pp. 99–121.
- Sakoda, James M (1971). “The checkerboard model of social interaction”. In: *The Journal of Mathematical Sociology* 1.1, pp. 119–132.
- Schelling, Thomas C (1971). “Dynamic models of segregation”. In: *Journal of mathematical sociology* 1.2, pp. 143–186.
- Schewel, Kerilyn et al. (2024). “How well can we predict climate migration? A review of forecasting models”. In: *Frontiers in Climate* 5, p. 1189125.
- Scholz, Stefan et al. (2025). “Improving Computer Vision Interpretability: Transparent Two-Level Classification for Complex Scenes”. In: *Political Analysis* 33.2, pp. 107–121.
- Schrodt, Philip A (2006). “Forecasting conflict in the Balkans using hidden Markov models”. In: *Programming for peace: Computer-aided methods for international conflict resolution and prevention*. Springer, pp. 161–184.
- Schutte, Sebastian and Nils B Weidmann (2011). “Diffusion patterns of violence in civil wars”. In: *Political Geography* 30.3, pp. 143–152.
- Schwenker, Friedhelm and Edmondo Trentin (2014). “Pattern classification and clustering: A review of partially supervised learning approaches”. In: *Pattern Recognition Letters* 37, pp. 4–14.
- Srivastava, Ajitesh, Satwant Singh, and Fiona Lee (2022). “Shape-based evaluation of epidemic forecasts”. In: *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 1701–1710.
- Teeraratkul, Thanchanok, Daniel O’Neill, and Sanjay Lall (2017). “Shape-based approach to household electric load curve clustering and prediction”. In: *IEEE Transactions on Smart Grid* 9.5, pp. 5196–5206.
- Tjaden, Jasper (2021). “Measuring migration 2.0: a review of digital data sources”. In: *Comparative Migration Studies* 9.1, p. 59.
- Tollefsen, Andreas Forø, Håvard Strand, and Halvard Buhaug (2012). “PRIO-GRID: A unified spatial data structure”. In: *Journal of Peace Research* 49.2, pp. 363–374.
- UN (2024). *International Migrant Stock 2024: Key Facts and Figures*. Advance unedited version. Accessed: 2025-04-28. URL: <https://www.un.org/development/desa/pd/content/international-migrant-stock>.
- Weidmann, Nils B (2016). “A closer look at reporting bias in conflict event data”. In: *American Journal of Political Science* 60.1, pp. 206–218.
- Welch, Nathan G and Adrian E Raftery (2022). “Probabilistic forecasts of international bilateral migration flows”. In: *Proceedings of the National Academy of Sciences* 119.35, e2203822119.
- Wen, Qingsong et al. (2022). “Transformers in time series: A survey”. In: *arXiv preprint arXiv:2202.07125*.

- Xu, Li and James S Famiglietti (2023). “Global patterns of water-driven human migration”. In: *Wiley Interdisciplinary Reviews: Water* 10.4, e1647.
- Yeh, Chin-Chia Michael, Helga Van Herle, and Eamonn Keogh (2016). “Matrix profile III: the matrix profile allows visualization of salient subsequences in massive time series”. In: *2016 IEEE 16th international conference on data mining (ICDM)*. IEEE, pp. 579–588.
- Yeh, Chin-Chia Michael, Yan Zhu, et al. (2016). “Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets”. In: *2016 IEEE 16th international conference on data mining (ICDM)*. Ieee, pp. 1317–1322.
- Yu, Wennian, Il Yong Kim, and Chris Mechefske (2021). “Analysis of different RNN autoencoder variants for time series classification and machine prognostics”. In: *Mechanical Systems and Signal Processing* 149, p. 107322.
- Zolhavarieh, Seyedjamal, Saeed Aghabozorgi, and Ying Wah Teh (2014). “A review of subsequence time series clustering”. In: *The Scientific World Journal* 2014.1, p. 312521.

2 Accounting for variability in conflict dynamics: A pattern-based predictive model

Accounting for variability in conflict dynamics: A pattern-based predictive model

Thomas Schincariol, Hannah Frank and Thomas Chadeaux

Abstract

Existing models for predicting conflict fatalities frequently produce conservative forecasts that gravitate towards the mean. While these approaches have a low average prediction error, they offer limited insights into temporal variations in conflict-related fatalities. Yet, accounting for variability is particularly relevant for policymakers, providing an indication on when to intervene. In this article, we introduce a novel risk-taking methodology, the ‘Shape finder’, designed to capture variability in fatality data, or rather the sudden surges and declines in the number of deaths over time. The method involves isolating historically analogous sequences of fatalities to create a reference repository. Comparing the shape of the input sequence to the historical references, the most similar historical cases are selected. Predictions are then generated using the average future outcomes of the selected matches. The Shape finder is derived from the theoretical understanding that strategic and adaptive interactions between the government and a non-state armed group produce recurring temporal patterns in fatality data, which are indicative of broader developments. In this article, we demonstrate that our approach maintains high accuracy while significantly enhancing the ability to predict shifts, surges, and declines in conflict fatalities over time. We show that combining the Shape finder with existing approaches, the Violence Early-Warning System (ViEWS) ensemble, achieves a lower mean squared error and better accounts for variability in fatality data. The Shape finder methodology performs particularly well for high intensity cases, or rather country-months with substantial armed violence.

Conflict forecasting has emerged as an important subfield of armed conflict studies (Hegre, Bell, et al., 2021; Mueller and Rauh, 2022). Recent advances have harnessed numerous data sources, from geopolitical indicators to digital media, as well as methods from statistics, machine learning, and game theory. Models such as the Violence Early-Warning System (ViEWS), for example, have set benchmarks by integrating complex datasets, including event data and socio-economic indicators, to forecast conflict occurrences and intensities on both national and subnational levels (Hegre, Allansson, et al., 2019; Hegre, Bell, et al., 2021). By analyzing past conflict patterns and integrating various geopolitical, socio-economic, and event-based data, forecasts help identify underlying trends and potential triggers for armed conflict.

Despite these advancements, most models rely on static, cross-sectional data that inadequately address the non-linear and evolving nature of conflicts. As a result, these forecasts, while generally accurate over broad time scales, fail to capture critical short-term fluctuations and sudden escalations. Existing methodological approaches often struggle to predict temporal patterns and fluctuations in conflict dynamics accurately. Such shortcomings are significant, as understanding these temporal variations is crucial for effective conflict mitigation and response strategies. The inability to anticipate these rapid changes can lead to missed opportunities for timely interventions and appropriate allocation of resources (Mueller and Rauh, 2022).

To address these limitations, this article introduces the ‘Shape finder’, designed to enhance the predictive accuracy of conflict-related fatalities by focusing on the temporal sequences and patterns in data. The main goal is to capture variability, rather than producing overly conservative—flat—predictions around the mean. Variability refers to the short-term escalation and de-escalation in the number of fatalities over time. The Shape finder methodology distinguishes itself by using Dynamic Time Warping (DTW) to compare sequences of casualties across different contexts and time periods. This

method allows us to align sequences of varying lengths and scales, making it possible to identify analogous patterns across disparate datasets. By focusing on sequences rather than isolated data points, the Shape finder predicts potential escalations and de-escalations in fatality data with greater precision, instead of returning flat forecasts.

Empirical tests of the Shape finder demonstrate its robust performance when compared to three baselines—the null model, the $t - 1$ model, and the ViEWS ensemble. The model not only predicts the occurrence and intensity of conflict-related fatalities with a comparable degree of accuracy to existing models, but also better captures temporal variability—an area where existing models fall short. Existing models tend to overlook short-term fluctuations and sudden escalations, focusing instead on broader trends—and quickly revert to the mean. In contrast, the Shape finder can detect and forecast critical changes ahead of time.

In addition, we construct a compound model that combines the Shape finder and ViEWS, selecting risk-taking and risk-averse approaches depending on our level of confidence in the predictions. The compound performs better than the constituent models in terms of both prediction error and the ability to capture variability in fatality data. The Shape finder is particularly suitable for high complexity input shapes—casualties time series with high variability—although it does tend to produce overpredictions for low complexity inputs—time series without significant variability, and therefore without meaningful ups-and-downs in the number of fatalities.

The remainder of the article is organized as follows. We first review the state of the art in conflict prediction, and we put forward our theoretical arguments. Then, we describe the data and methodologies used in this study, including a comprehensive description of the Shape finder algorithm and its implementation. Finally, we present the results of our empirical analysis and discuss the implications of our findings.

Conflict prediction

Conflict prediction has emerged as a thriving subfield of peace and conflict studies (Chadefaux, 2017; Colaresi and Mahmood, 2017; Hegre, Metternich, et al., 2017; Ward, Greenhill, and Bakke, 2010). A range of academic early-warning systems exists to forecast armed conflict, complemented by operational systems used by practitioners. Rød, Gåsste, and Hegre, 2024 present a comparative review.

Among these systems, the Violence Early-Warning System (ViEWS) stands out, offering monthly forecasts for the number of fatalities in state-based armed conflict on a global scale, with projections extending up to three years into the future (Hegre, Allansson, et al., 2019; Hegre, Bell, et al., 2021). ViEWS employs a sophisticated ensemble model that incorporates structural variables related to socio-economic, geographic, and political factors. Another approach by Mueller and Rauh, 2022 relies on a large text corpus from over 4 million news articles, and extracts topics using Latent Dirichlet Allocation. Those topics, in addition to a set of variables on conflict history, are used to predict the onset of organized violence per country-month. Finally, the ACLED Conflict Alert System (CAST) predicts organized political violence up to six months ahead, mostly relying on ACLED data as predictors (ACLED, 2023). CAST forecasts are generated on a global scale, encompassing both national and sub-national levels. As the review by Rød, Gåsste, and Hegre, 2024 demonstrates, the models assign a high risk of violent action to a similar set of countries, regardless of the differences across these systems.

Despite advances in methodology and data collection, conflict prediction remains a challenging field. Several key issues continue to hinder progress in this area (e.g. Cederman and Weidmann, 2017). The inherent imbalance in conflict data, with peaceful periods far outnumbering instances of violence, poses significant analytical challenges.

There is also an ongoing debate among scholars about the fundamental causes of armed conflict, leading to uncertainties in model specification. The quality and depth of available data often falls short of what is required for precise predictions, especially in regions prone to conflict, where data collection can be difficult or dangerous. The ‘hard problem of conflict prediction’ (Mueller and Rauh, 2022, p. 2442) are conflict onsets, or rather cases with no prior conflict history. These instances are particularly difficult to anticipate using structural predictors that measure economic performance, demographics, or political institutions. Finally, idiosyncratic events—‘historical accidents’ (Cederman and Weidmann, 2017) such as the end of the Cold War or Brexit—are perhaps impossible to predict.

The rarity of armed conflicts poses one of the most significant methodological challenges. Rare events occur only infrequently, which means that the dataset has a large number of zeros (King and Zeng, 2001). Indeed, it has been noted that the distribution of fatalities is extremely right-skewed. Logit regression and more complex machine learning algorithms are known to underpredict in the context of rare events (Cederman and Weidmann, 2017; King and Zeng, 2001). The right-skewed structure of conflict data encourages risk-averse predictions. For example, a conservative model that exclusively predicts zero fatalities will achieve a low mean squared error (MSE) when dealing with rare events. Consequently, most models tend to produce risk-averse predictions that effectively identify a baseline risk of conflict but struggle to detect temporal variations and sudden shifts in the data. These models produce ‘flat’ predictions close to the mean value of fatalities. We show examples in Figure 3.

However, capturing variability in fatality data and predicting conflict escalation are crucial. Accurately identifying when armed conflict will escalate allows for the design of appropriate responses, rather than assuming a constant level of violence. From a policy perspective, false negatives are more severe than false positives, as the cost of preparing

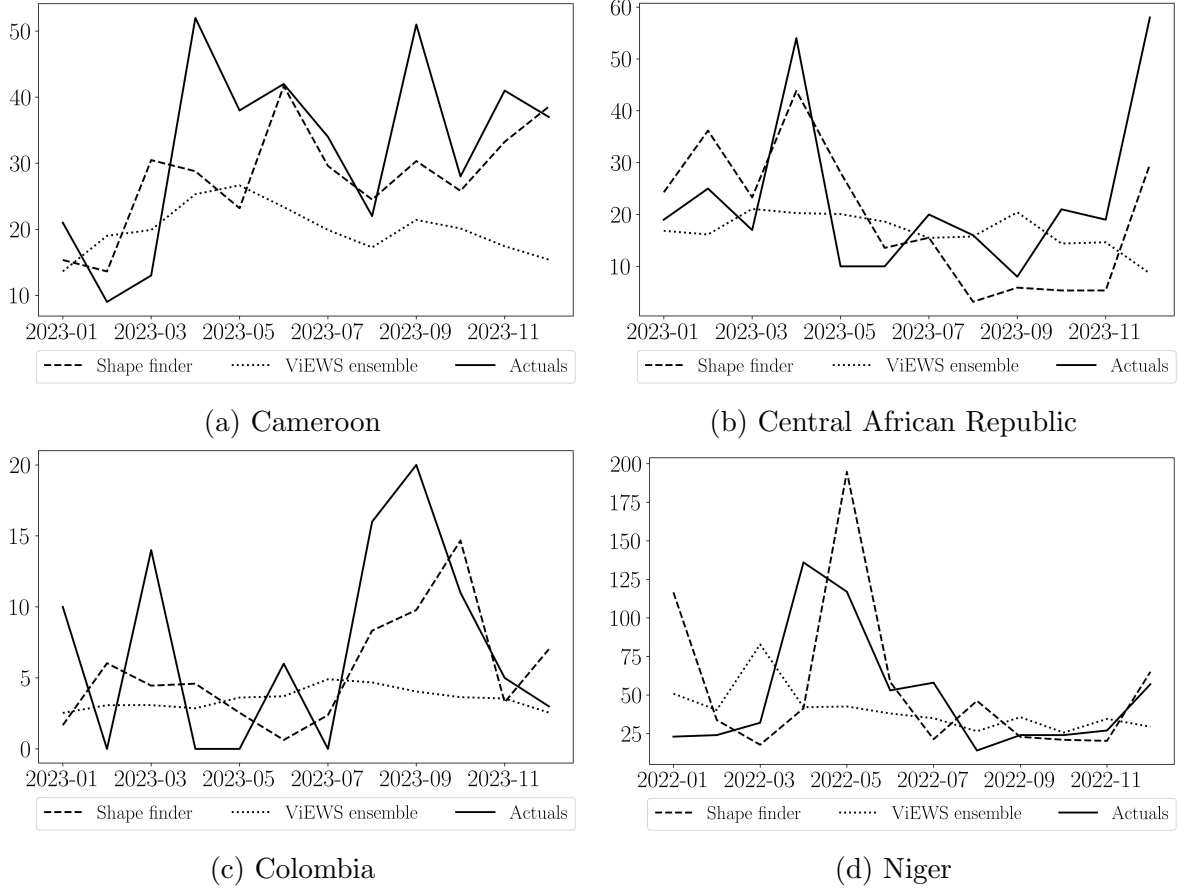


Figure 3: Sample predictions. The actual fatalities are shown in black, with the risk-averse model (ViEWS ensemble) in dotted lines and the risk-taking model (Shape finder) in dashed lines. The risk-averse model generates flat predictions around the mean, missing sudden surges and declines. The risk-taking model captures variability, reflecting the highs and lows more accurately.

for an escalation that does not occur is lower than the human cost of an unmitigated escalation (Mueller, Rauh, et al., 2024). Thus, a risk-taking model, despite occasional overpredictions, may be more relevant for policymakers than a risk-averse model. We define a risk-taking model as a conflict prediction model which can accurately anticipate the variability in fatalities over time—the ups-and-downs in the number of deaths as conflict escalates and de-escalates. While occasional overpredictions are unavoidable, the risk-taking model needs to maintain a low prediction error, comparable to risk-averse approaches.

In this article, we introduce the ‘Shape finder’, a novel approach to conflict prediction that complements existing risk-averse models. The main goal is to capture variability in fatality data, rather than producing ‘flat’ predictions around the mean, thereby making the Shape finder a risk-taking model (Figure 3). This method identifies recurring patterns in historical fatality data to inform future predictions. The Shape finder operates by comparing an input sequence to a comprehensive repository of historical cases, identifying similar shapes or patterns. By focusing on these analogous historical instances, the model generates predictions based on the principle that similar patterns tend to lead to similar outcomes. This approach offers a fresh perspective on conflict forecasting, potentially uncovering insights that traditional risk-averse models might overlook.

Temporal patterns in fatality data are not mere coincidences but reflect underlying data-generating processes common to various social phenomena, including armed conflicts. We posit that these patterns, arising from the dynamic interactions between governments and non-state actors, offer valuable insights into conflict trajectories. Armed conflicts typically evolve from low-intensity engagements, with rebel groups often mobilizing for years before conflicts escalate to full-scale wars (Malone, 2021). Interactions between state and rebel group create discernible patterns in fatality data, which are indicative of future conflict dynamics. By analyzing these patterns, we can infer crucial information about actor-specific characteristics and strategies, such as relative military capabilities or tactical approaches, which affect the shapes of the produced patterns. These insights are challenging to operationalize with existing data.

Our study builds upon and significantly enhances performance of the shape-based model introduced by Chadeaux (2022). While we maintain a similar pattern-finding framework, our approach diverges in crucial ways that yield more actionable predictions, with a particular focus on capturing variability in fatality data. The previous

shape-based model takes one fatality time series as input and obtains the corresponding historical sequences from all other countries. Predicted fatalities are returned as the weighted average, assigning weights based on the similarity between input and the specific historical cases. This method produces overly conservative predictions, thus making it a risk-averse model. In contrast, the updated model considers a limited subset of sequences only when making predictions, selected based on a distance criterion. This targeted approach allows us to generate more risk-taking forecasts by concentrating on the most similar historical patterns, rather than using all historical sequences. Indeed, the Shape finder outperforms the shape-based approach in Chadeaux (2022), especially in terms of the prediction error (Figure A.13 in the Online Appendix).

Another innovation is the extension of the search window for similar sequences. While the previous model uses a fixed 12-month window, the updated approach compares sub-sequences ranging between 8 and 12 months. This enables us to identify similar patterns across different speeds. This flexibility is crucial because conflict dynamics can unfold at different paces across various contexts, and a rigid time frame might miss important recurring patterns that develop over slightly shorter or longer periods.

A final adaptation relates to how the final predictions are produced. Rather than calculating the weighted mean, we apply a hierarchical clustering algorithm to the sequences selected in the comparison process. The centroid of the cluster with the highest number of observations—the majority cluster—is the final prediction. This approach enhances the robustness of the prediction by filtering out outlier cases, and by extension reducing the risk of overpredictions. The clustering process is essential given that the Shape finder seeks to capture variability, while maintaining a low prediction error comparable to risk-averse models.

Patterns in fatality data

The role of patterns in time series data has long been recognized by various disciplines. For example, ecologists use predator-and-prey models, where predator and prey population sizes increase and decrease in reciprocal cycles (Diz-Pita and Otero-Espinar, 2021). Similarly, epidemiologists have proposed compartmental models to anticipate the spread of infectious diseases, which are characterized by reciprocal cycles between susceptible, infectious and recovered rates (Hethcote, 1989). Other disciplines more closely related to the social sciences have likewise developed theoretical arguments on temporal patterns. In macroeconomics, researchers investigate business cycles, which are defined by consecutive economic expansion and recession (Kalecki, 1937). Managing the business cycle, governments can apply countercyclical fiscal policies to navigate the expanding and contracting economy. Finally, demographic transition theory describes the procyclical changes in death and birth rates, where countries transition from high fertility-high mortality to low fertility-low mortality (Kirk, 1996).

What the corresponding theoretical arguments have in common is the emphasis on interactions between ‘actors’, widely understood, which produce countercyclical or procyclical patterns, where the size of actor a increases as the size of actor b increases or decreases, and vice versa. Putting actor size into the context of armed conflict, size indicates the strength of an armed faction in terms of military capabilities, considering troop size, scope of the military equipment, and access to intelligence. The military capabilities of the government and the non-state armed group might develop countercyclical, depending on which side has the upper hand. Alternatively, military capabilities might be procyclical, where an escalation of one side prompts an escalation on the other. The higher the military capabilities in total, the more intense the armed combat and the higher the battle-related fatalities.

Most relevant in this context, theoretical models on social movements investigate the chess-like process of tactical innovation and adaptation taking place between protesters and the state (e.g. Della Porta, 2008; Della Porta et al., 2018; McAdam, 1983; Rød and Weidmann, 2023; Koopmans, 2007; Tarrow, 2011). As protest erupts, the government is likely to respond with state repression (Davenport, 2007). In light of state repression, the protest group needs to change tactics, making them more resilient to the government (McAdam, 1983). In reaction to the tactical innovations on the protesters' side, the government is likely to adapt its repression apparatus, offering a more effective way of keeping protesters in check. Tactical innovations and adaptations are countercyclical processes, and are likely to produce recurring patterns in sequences of protest events, where protest intensity consecutively increases and decreases.

We expect that the countercyclical relationship of tactical innovation and adaptation applies to other forms of political contention as well, such as armed conflict. Continued interactions between the government and a non-state armed group would therefore produce recurring patterns in sequences of fatalities. Leveraging these patterns, we can infer information about actor characteristics, such as relative military capabilities or tactical approaches, which are likely to affect the shapes of the produced patterns. Hence, those patterns contain information about the larger dynamics in armed conflict, and are therefore indicative of the future.

The expectation of temporal patterns in conflict fatality data can also be supported by broader theories of human behavior and organizational dynamics. One foundational concept is the idea of feedback loops in complex systems, which often drive the emergence of predictable patterns over time. In armed conflicts, these feedback loops can manifest as reinforcement or deterrence cycles between opposing forces, where the behavior of one side depends on the behavior of the other side (e.g. Axelrod, 1990; Pierskalla, 2010; Young, 2013). For example, an increase in government military action

may provoke an escalation in retaliatory attacks by insurgents, which in turn could lead to further government responses. This cyclical dynamic, shaped by strategic decision-making and resource availability, creates the conditions for recurring patterns in conflict fatalities.

Furthermore, armed groups and governments operate under constraints of limited information and resources, leading them to make incremental adjustments based on observed outcomes. These adjustments are often reactive and take time to implement, creating lags in the system. Such lags can give rise to periodicity in conflict intensity, as actors respond to perceived threats or opportunities in ways that inadvertently synchronize their actions with the opposing side. This phenomenon parallels oscillatory dynamics and is also observed in other systems, such as labor strikes and employer negotiations (Dosi, Usula, and Virgillito, 2024), where actions by one side spur proportional or mirrored reactions from the other.

Finally, the persistence of patterns in conflict fatalities can also be tied to resource mobilization and exhaustion. Armed conflicts are sustained by material and human resources, which are not infinite. Periods of intense fighting often deplete these resources, necessitating a temporary lull for reorganization and replenishment (Powell, 2004; Slantchev, 2003). This ebb and flow of resources naturally produces periodic peaks and troughs in conflict fatalities. Furthermore, the role of external actors, such as international mediators or arms suppliers, can amplify these patterns by influencing the timing and scale of engagement between warring parties (Beardsley, Cunningham, and White, 2019; Lacina, 2006; Regan, 2002).

Taken together, these theoretical perspectives illustrate how the interplay of strategy, resource dynamics, and adaptive behavior can produce recurring sequences. Identifying and understanding these patterns not only sheds light on past interactions but also offers a basis for anticipating future developments.

Data

The Shape finder is an autoregressive model. Data for the input and output—the number of fatalities in state-based armed conflict—are obtained from the Georeferenced Event Dataset published by the Uppsala Conflict Data Program (UCDP GED) (Sundberg and Melander, 2013). The event data are aggregated at the country-month level, which serves as the unit of analysis. We include fatalities from state-based armed conflict, using the *best* estimate.⁶ State-based armed conflict refers to armed conflict where at least one side is the government of a country (N. P. Gleditsch et al., 2002). The spatial coverage is defined according to the countries listed in K. S. Gleditsch and Ward (1999), and the temporal coverage is based on UCDP methodology—1989 until 2023.⁷

We use two test partitions to validate the performance of the Shape finder, one for the year 2022 and another for the year 2023. The training data spans from 1989 to 2020 for the 2022 test set, and from 1989 to 2021 for the 2023 test partition. This setup allows us to use historical data to predict outcomes for each respective year, while years 2021 and 2022 are used to generate the input sequences. When tuning hyperparameters of the Shape finder, we use the year 2021 as validation data, obtain the input sequences from 2020, and use 1989 to 2019 as training data.

We compare the performance of the Shape finder with the ViEWS ensemble. The ViEWS ensemble has been refined and improved since its first introduction in 2019, and can be seen as the state of the art in conflict prediction. From the updated ViEWS

⁶The UCDP GED dataset offers three estimates of fatalities for each event: a low estimate; a best estimate, reflecting the most reliable count from the source; and a high estimate—the maximum reliable count.

⁷For easy access, the data are downloaded via the ViEWS data API. More specifically, we download predictions for all months in 2022 from the *fatalities001_2022_00.t01* model and predictions for all months in 2023 from the *fatalities001_2023_00.t01* model, thereby using a 12-month prediction horizon. A list of all available data compiled by ViEWS can be found at, https://github.com/prio-data/views_api/wiki/Available-datasets. We use the version of the data downloaded in April 2024.

ensemble (Hegre, Akbari, et al., 2022), which predicts fatality counts, we obtain predictions for all months in 2022 based on data up to 2021, and predictions for all months in 2023 based on data up to 2022, thereby using forecasts for the 12-month prediction horizon. The training-test data split is informed by the ViEWS set-up, in order to enable a comparison between ViEWS and the Shape finder.

Methods

The Shape finder identifies recurring shapes in fatality data to predict future trends. This involves two steps. First, a reference repository is created by comparing each input sequence with all historical cases and storing the closest matches. The similarity between the input sequence and the references is measured using Dynamic Time Warping (DTW), which allows for time-shifts when comparing the shapes of time sequences. Second, this repository is used to predict the next twelve months by clustering the futures of the matched sequences. The prediction is the centroid with the highest number of observations. Figures 4 and 5 illustrate the Shape finder algorithm.

Creating the reference repository. The input sequences are the last 10 months before the testing period for each country.⁸ The reference repository p is constructed by partitioning each country’s historical data into overlapping 10-month sequences using a rolling window approach. For example, Mozambique’s sequences include January to October 1995, February to November 1995, and so on. We also consider changing the window (win) to ± 2 months (i.e. sub-sequences ranging from 8 to 12 months) to account for temporal variations in the speed of events.⁹ This generates a set of N

⁸The input sequence spans March 2021 to December 2021 for the 2022 testing set, and March 2022 to December 2022 for the 2023 set.

⁹Our findings are not substantially affected by slight changes in win (Figure A.13 in the Online Appendix), so we opt for a shorter window length for computational costs reasons.

sequences derived exclusively from the training data. We compare each input sequence s to all sequences in the reference set p using Dynamic Time Warping (discussed below). Only the most similar sequences are retained, creating a filtered reference repository p_{filtered} . This ensures that only the most analogous patterns are considered for future predictions.

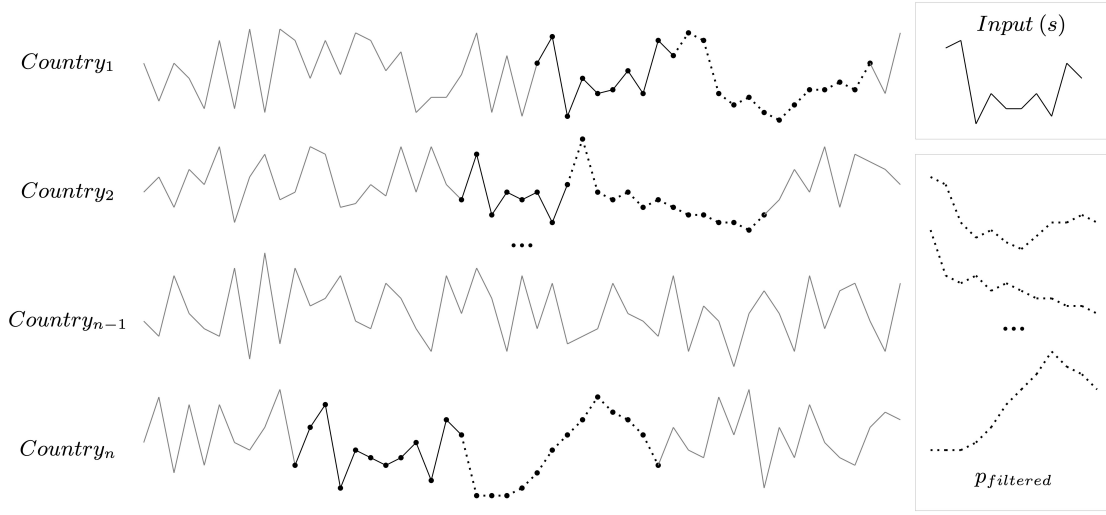


Figure 4: The input sequences consist of the last 10 months prior to the testing period. The reference repository p contains overlapping sub-sequences of 8–12 months. For each input sequence s , only the most similar sub-sequences are retained in the filtered repository p_{filtered} . Similarity is measured using Dynamic Time Warping, and a sequence is included if its distance to s is below a set threshold.

We use Dynamic Time Warping (DTW) to measure the similarity between our input sequence s and the sequences in the reference repository p . DTW is a widely used time series alignment technique that quantifies the similarity, or distance, between time series data (Keogh and Ratanamahatana, 2005). Unlike traditional methods, DTW finds the optimal alignment between two time series without being restricted to fixed temporal indices. It compares each data point in the first time series with all data points in the second, matching each to its closest counterpart. This makes it possible to stretch or compress the axis in time, which produces a new alignment between the

two sequences.¹⁰

Using Dynamic Time Warping (DTW) as a distance metric for sequence comparison offers several advantages—most notably its invariance to time shifting and stretching. Traditional distance metrics, like the Euclidean distance, are sensitive to temporal distortions, making them less suitable for comparing time series data where temporal shifts are common (Keogh and Ratanamahatana, 2005). Consider the following three sequences: $a = [1, 2, 3, 2, 1]$, $b = [2, 3, 2, 0.5, 1]$ and $c = [1, 3, 1]$. All three sequences have the same basic shape—first up, then down. Yet, standard approaches such as regression or the Euclidean distance are unable to find similarities between these sequences due to a strict case-by-case comparison. In contrast, DTW can align sequences that may have similar patterns but occur at different speeds or scales. Additionally, this method accommodates slight variations in sequence lengths when making comparisons, which further increases the flexibility of the approach (T. Wong and M. Wong, 2003). The DTW distance between a and b is 1.118 versus 2.291 for the Euclidean distance. The DTW distance between a and c is 1.414, while the Euclidean distance is undefined in this case due to their differing lengths. DTW is therefore ideal for matching sequences in our dataset of country-month fatalities, where temporal variations are natural occurrences.

After calculating the distance for each pair of input and historical subsequence, we filter out the sequences that do not closely match subsequence s . More precisely, we keep a past subsequence $s_{u \leq t}$ in the set only if $\text{dist}(s, s_{u \leq t})$ is below a certain threshold (dist). The threshold is initiated as $\text{dist}(s, s_{u \leq t}) = 0.1$, and increases by 0.1 until there are at least five reference subsequence in p_{filtered} . The filtered set of past subsequences is then used to make predictions. Considering only a filtered set of past sequences p_{filtered}

¹⁰In technical terms, Dynamic Time Warping calculates the shortest path through a cost matrix, which is generated from the distances between every pair of points in the two time series. This optimal alignment is identified by minimizing the cumulative distance, subject to constraints that ensure continuity and locality in the alignment path. The Dynamic Time Warping distance between the two sequences is determined by the sum of these aligned distances.

yields riskier predictions, given that only the most similar reference series are taken into account.

The threshold *dist* is a hyperparameter of the Shape finder to control the size of the filtered reference repository. A higher threshold results in a more diverse set of references, considering historical cases with a higher dissimilarity in shape. The hyperparameter is optimized by calculating out-of-sample predictions for various values of *dist* in a validation period (the 12 months of 2021) and selecting the value that yields the best performance, *dist* = 0.1 (Figure A.1 in the Online Appendix). Our findings either do not change substantively or performance worsens when increasing the threshold *dist*, which enlarges the size of the filtered repository by construction (Figure A.13 in the Online Appendix).

Predictions. For each input sequence *s* (the last 10 months prior to the testing window), we access the sequences stored in the filtered repository of historical cases p_{filtered} . From each reference sequence, we extract the future, which is the number of fatalities for the next 12 months following that reference sequence.¹¹ We then use these futures to make predictions for our input sequence *s*. The model returns predictions for all input sequences that are not flat, meaning they have at least one month with non-zero fatalities. For flat input sequences, the model predicts zeros for the next 12 months. In those cases, predicting zero is the best guess.¹² We use this step to speed up computation, given that flat inputs are passed on and the Shape finder algorithm is not applied in those instances.

To obtain predictions, we apply the agglomerative hierarchical clustering algorithm,

¹¹To prevent data leakage, we discard all references whose future lies within the test data period.

¹²Indeed, for the majority of flat input sequences in our sample, the future 12 months are flat as well (86%). To ensure a fair comparison with ViEWS, we demonstrate that including a similar approach in the ViEWS forecasts does not significantly change the performance (Figure A.2 in the Online Appendix).

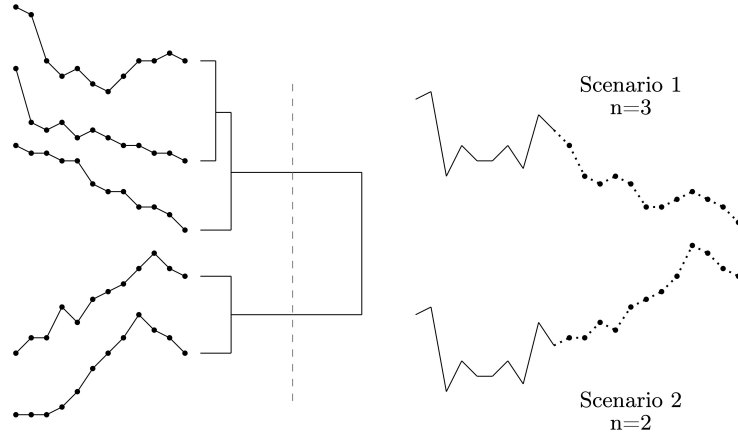


Figure 5: Agglomerative hierarchical clustering. The five sequences on the left, part of $p_{filtered}$, are used as input to the agglomerative hierarchical clustering. On the right, the two scenarios created by the clustering are displayed. The centroid from the majority cluster is selected as the prediction (scenario 1 here).

using the futures of the sequences in the reference repository as input. The predicted outcome is the centroid of the cluster with the highest number of observations (the ‘majority cluster’). This clustering approach has the advantage of grouping similar futures together, which enables the identification of common patterns that might indicate recurring trends. Further, considering clusters rather than individual sequences reduces noise and discards outlier cases.

Agglomerative hierarchical clustering works by repeatedly merging observations into larger clusters based on similarity until all cases are assigned to one single group at the top (Müllner, 2011). A cut-off value (cut) is used to control the depth of the clustering tree, thereby specifying the final number of clusters. Selecting a lower value for cut leads to a lower cut-off value in the clustering tree and by extension a higher number of clusters. We refine the hierarchical structure by optimizing the cut-off value brute-force, which means that cut is another hyperparameter of the Shape finder algorithm.¹³

¹³The threshold with the lowest out-of-sample prediction error is selected, in this context $cut = \frac{horizon}{3} = \frac{12}{3} = 4$, where $horizon$ refers to the prediction window of 12 months. More specifically, we use the 12 months in 2021 as validation data to optimize the hyperparameters of the Shape Finder $dist$ and cut (Figure A.1 in the Online Appendix)

While we optimize *cut* using the validation data (the 12 months in 2021), increasing or decreasing the value of *cut* slightly does not significantly change the performance (Figure A.13 in the Online Appendix).

Based on the clustering solutions, we derive predictions for the next 12 months. We first select the cluster with the largest number of sequences—the majority cluster—and calculate the centroid, which is the average of all sequences in that cluster. That centroid is our prediction for the 12 months following the input sequence.¹⁴ This approach ensures that we select the most robust future—the one which is substantiated by the highest number of observations—thereby circumventing any sparsity concerns due to low numbers of cases assigned to each cluster.¹⁵

Compound model

Risk-averse and risk-taking models offer complementary strengths in conflict prediction. While risk-averse models excel at forecasting high-intensity conflicts based on recent history, they may overlook the potential escalation of medium-intensity cases (Hegre, Akbari, et al., 2022). In other words, risk-averse models are good at predicting a baseline risk of armed conflict accurately, while they cannot capture sudden surges and declines in fatality data. These models are risk-averse because they perform well in terms of average error, and they produce fewer overpredictions. Simultaneously, they cannot address variability and miss important short term developments. Understanding when medium-intensity cases escalate to full-scale war is particularly important for policymakers, potentially allowing them to intervene before a more substantial escalation of violence (Mueller and Rauh, 2022).

¹⁴If the probabilities are uniform, i.e. no clear cluster stands out even after applying the cut-off, we use the average of all future sequences as the prediction without further clustering.

¹⁵The average number of cases contained in the majority cluster is 66, with a standard deviation of 97.

This is the strength of a risk-taking approach—the Shape finder—which seeks to capture variability in fatality time series. However, taking risks does not always pay off, and such models are more prone to overpredictions. The effectiveness of this risk-taking approach depends on the model’s confidence in its predictions. To optimize our forecasting, we must determine when the potential benefits of risk-taking outweigh its costs. We aim to leverage the strengths of both risk-averse and risk-taking models by constructing a compound model.

The Shape finder requires sufficient observations in the historical dataset to produce accurate estimations. An insufficient data basis may lead to predictions being highly affected by outliers, making the model prone to overpredictions. Hence, we propose a compound model that includes predictions from the risk-taking Shape finder when confidence in the prediction is high. The ViEWS ensemble—a risk-averse model—is selected when confidence is low or when the severity of the input is too high, resulting in an elevated risk. An overprediction for high severity cases is likely to drive up the prediction error.

Based on these considerations, the compound employs two selection criteria—the severity of violence and the confidence level of the Shape finder. First, a risk-averse model might perform better for high intensity cases, given that any type of overprediction would lead to a high error. The severity of violence is indicated by the number of fatalities—the higher the fatalities, the higher the severity. Second, we use the clustering results as an indication of the confidence in our prediction. The confidence in the Shape finder is quantified by the relative frequency of cases assigned to the majority cluster, normalized by the logarithm of the number of observations. We select the Shape finder for cases where the observed pattern has occurred more than three times in the past and at least 50% of the historical futures of that pattern fall within the majority cluster. These two components are important for a sufficient understanding

of a specific pattern and its projected outcome, and produce a compound model where 66% of the predictions are from the Shape finder, and 33% from ViEWS.¹⁶

Evaluation

We compare the performance of the Shape finder to three alternatives—the ViEWS ensemble and two baselines. The first baseline—the null model—predicts zero fatalities for all country-months. The second baseline—the $t - 1$ model—uses the input sequence as prediction. These baselines may appear simplistic. However, forecasting that fatalities are zero or will remain consistent with the immediate past is a robust prediction due to the highly skewed nature of fatality data. Indeed, the null model is challenging to beat in terms of prediction error, since a large fraction of observations actually do take a value of zero. The $t - 1$ model may mirror how practitioners make predictions in the field. In the absence of any other information, someone in the field is likely to predict a continuation of the fatalities from the previous month. Ideally, we would also like to compare the Shape finder to an operational conflict prediction system. However, these are run by governments or private entities, and predictions are therefore not publicly available (see Rød, Gåsste, and Hegre, 2024).

The ViEWS ensemble is a more sophisticated approach and combines predictions from a variety of thematic models, using variables on among others conflict history, economic development and political institutions (Hegre, Akbari, et al., 2022). The ViEWS model was first introduced in 2019 and is arguably the most advanced conflict prediction model accessible to the public. The ViEWS ensemble is therefore our primary comparison for assessing the Shape finder’s performance. Our aim is not to position these models as competitors, but rather to explore how risk-averse and risk-taking

¹⁶See Figure A.3 in the Online Appendix for more details on the compound selection.

approaches can complement each other, along the unique strengths of both approaches. The risk-averse model can capture a baseline risk of armed violence, and the risk-taking model is able to mirror variability in fatality time series, despite occasional over-predictions. The goal is to generate predictions that can capture variability, while maintaining a prediction error similar to existing approaches.

We report two evaluation metrics. The first is the mean squared error (MSE), a standard metric which reports the average prediction error. More formally, the MSE is defined as the average of the squared differences between predicted and observed values, $MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$. Despite its ubiquity, the MSE has severe limitations because of its propensity to reward overly conservative forecasts when the data distribution is skewed (Hossin and Sulaiman, 2015). Because conflict fatalities are heavily skewed, optimizing on the MSE can encourage risk-averse predictions. Predicting zero fatalities for all country-months yields a low MSE because most countries are at peace most of the time. This is exacerbated by the square term in the MSE, which severely punishes deviations from observed values, and therefore encourages conservative predictions. Indeed, judging by the MSE alone, there is no significant difference between the predictions of ViEWS and those of a model that only predicts zero (Figure 6).

The MSE does not reward predictions that capture the variability of rare events. Yet, we believe that this variation—the changes over time, rather than the repetition of the last observation—is of main interest. Indeed, relying solely on MSE might mask the model’s shortcomings in identifying variability in fatality data. Therefore, our second evaluation metric is designed to measure the algorithm’s ability to capture *changes* in the number of casualties, rather than capturing a stable value. To this end, our metric quantifies the degree to which our forecasts successfully detect and represent shifts, surges, or declines in casualties. We use the difference of the predicted variation, exponentially and difference weighted, to reward good predictions of peak values. This

metric, which we call the Difference Explained (DE) is defined as

$$DE = \frac{1}{n} \sum e^{-k|\frac{\Delta\hat{y}-\Delta y}{\Delta y}|}, \quad (1)$$

where $\hat{y}$ is the predicted change, y is the actual change, n is the number of predictions, and k is a balancing factor discussed below. Δ refers to the change from $t - 1$ to t (i.e. $y_t - y_{t-1}$).

This metric offers several advantages for evaluating the accuracy of time series predictions. First, DE captures the deviation between the predicted and actual changes in the time series data ($|\Delta\hat{y} - \Delta y|$). Second, the inclusion of an exponential term with balancing factor e^{-k} introduces a nonlinear weighting that penalizes larger deviations between predicted and observed changes. Here we set $k = 5$.¹⁷ Finally, the term Δy assigns greater weight to predictions around larger actual changes and hence rewards good predictions at critical junctures. Larger DE values indicate a model that better captures variability.

In addition, we want to reward accurate predictions with a time shift of a single month—these are cases where a peak in fatalities is predicted either one month too early or one month too late. To allow for this time flexibility, the DE at Δy_{t-1} and Δy_{t+1} is obtained and a penalty of 50% is applied, meaning that the DE at $t - 1$ and $t + 1$ is multiplied by 0.5. The final DE value is the maximum of these three values— $t - 1$, t , or $t + 1$. This approach rewards accurate peak predictions, even with a one-month

¹⁷The metric’s sensitivity to errors at critical peaks requires careful selection of the balancing factor k , here set to 5. Using the results from a data experiment (Figure A.4 in the Online Appendix), factor k is chosen to ensure that the metric rewards risky predictions while maintaining reasonable accuracy. To that end, we created a fictional time series and tested three prediction models—a risk-averse model, a risk-taking model, and an overly risky model. The risk-taking model performs best at capturing variability in the time series, accurately predicting the peaks. Using a balancing factor of 5, rather than 1 or 10, assigns the highest DE to this model. As a robustness check, however, we confirmed that our main findings do not change significantly when increasing or decreasing the balancing factor k (Figure A.5 in the Online Appendix).

lag. Indeed, a one-month lag can be considered a minor deviation, given the model’s 12-month prediction horizon. Overall, the DE metric is particularly useful in scenarios where the prediction of peak values is important.¹⁸

Results

A conflict prediction model would ideally perform well on both tasks—minimizing the prediction error and capturing meaningful variation in fatalities. The null model has a low prediction error, since most country-months experience zero fatalities—the data is highly skewed. However, such a model entirely misses variability—sudden surges or declines. The $t - 1$ model might be able to capture some variability, but makes larger errors because of its simple algorithm. The ViEWS ensemble and the Shape finder address both tasks—having high accuracy and the ability to capture variability. However, ViEWS as a risk-averse model has higher accuracy, while the Shape finder as a risk-taking model has better ability to capture variability (Figure 6).

Compound model. Our main findings are summarized in Figure 6, which shows the difference explained (DE) and mean squared error (MSE) for each model compared to the compound. More specifically, we calculate the log-ratio for each metric, taking the compound as base.¹⁹ For DE, a negative log-ratio signifies that the compound performs better than the reference model. For MSE, a positive log-ratio means that the compound is the better performer. The best performing models are located in the upper right quadrant, since we reversed the MSE axis now showing the accuracy ratio. All model variants—the Shape finder, the ViEWS ensemble and the two baselines—

¹⁸We illustrate the difference between MSE and DE in Figure A.6 in the Online Appendix. In Figure A.7 in the Online Appendix, we demonstrate further that using DTW as a metric does not change our findings.

¹⁹In other words, the log-ratio is defined as $\log\left(\frac{m_{\text{model}}}{m_{\text{compound}}}\right)$, where $m \in \{\text{MSE}, \text{DE}\}$.

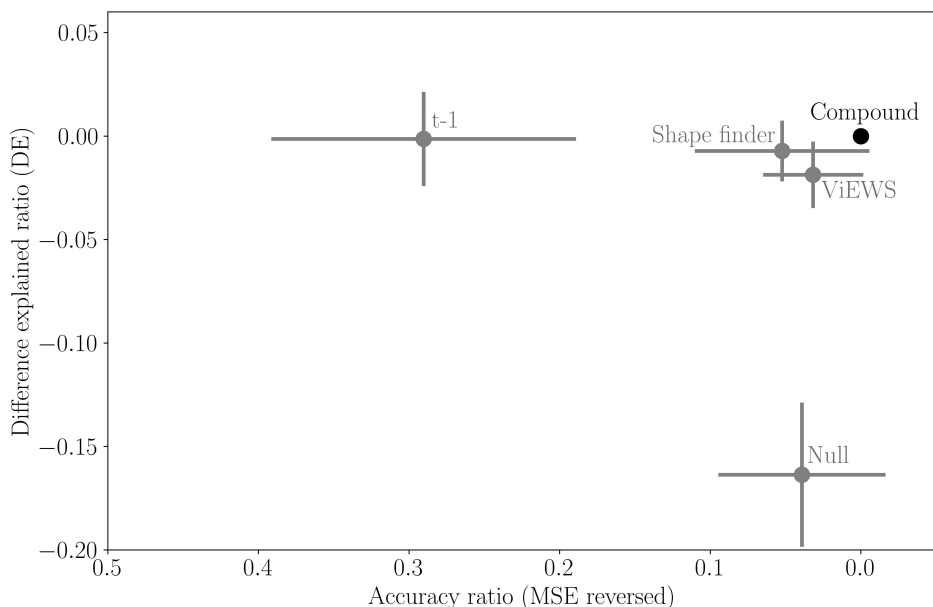


Figure 6: Mean log-ratio for difference explained (DE) and mean squared error (MSE). The log-ratios are computed using the compound as base. For DE, a negative log-ratio signifies that the compound outperforms the reference model. Conversely, for MSE, a positive log-ratio indicates that the compound performs better. For visualization purposes, we reverse the MSE axis, now showing the accuracy ratio. Superior models are located in the top-right quadrant. Vertical and horizontal bars represent the 95% confidence intervals around the mean for each metric. We calculate the confidence intervals as $\pm 1.993 * \text{Standard deviation} / \sqrt{n}$, where 1.993 corresponds to the critical value of t with 73 degrees of freedom.

have a negative log-ratio on DE and a positive log-ratio on MSE. This means that the compound outperforms all other model variants, both in terms of having a lower prediction error and in terms of accurately capturing the variability in fatality data. The compound is able to combine the strength of ViEWS and the Shape finder.²⁰

Next, we examine under which conditions the Shape finder works particularly well or poorly, in comparison to the ViEWS ensemble. We consider three characteristics of the input sequences—the standard deviation of the input sequence, the standard deviation of the differentiated input sequence, and the mean of all data points in the input sequence. These three measures indicate the level of variability—a high mean

²⁰Results for different forecasting horizons are presented in Figure A.8 in the Online Appendix.

value and a low standard deviation of the differentiated series suggests that the fatality sequence has substantial surges and declines. Plotting these characteristics against each other reveals clusters of cases with a positive MSE log-ratio (ViEWS as base and Shape finder as reference), suggesting that the Shape finder performs better, and clusters of cases with a negative MSE log-ratio, implying that ViEWS has a lower error.²¹

The Shape Finder model works particularly well for complex input shapes, which have a high average value and a small standard deviation of the differentiated series. Those two characteristics refer to sequences with substantial variability (Figure A.12 in the Online Appendix), other than flat time series, with single peaks, which would result in a high standard deviation of the differenced series and a low mean value. In these cases, ViEWS often produces either overpredictions or overly conservative forecasts. For input shapes with low complexity, having low average values and high standard deviations—typically time series with single peaks and flat trajectories (Figure A.12 in the Online Appendix)—the Shape Finder tends to overpredict, while ViEWS provides more accurate forecasts.

Well-performing cases. In Figure 7a, we present the cases in the well-performing cluster with the highest MSE log-ratio. For the first two inputs, we observe that the ViEWS ensemble tends to overpredict, likely due to the high mean value in the input shapes. This overprediction might explain why the Shape finder achieves higher accuracy for those cases, in addition to capturing variability.²² This contradicts the idea that risk-averse models systematically underpredict fatalities. For the last two input series, we observe that the ViEWS ensemble produces overly conservative predictions ranging around the mean, while the Shape finder is able to capture the variability. This

²¹For further details on the selection process, see Figure A.9 and A.10 in the Online Appendix.

²²To determine if this observation holds across the entire dataset, we analyze all cases with this type of input. We find a significantly lower MSE for complex inputs when compared to more simplistic patterns, reinforcing our previous interpretation (Figure A.11 in the Online Appendix).

is in line with our discussions on risk-averse and risk-taking models.

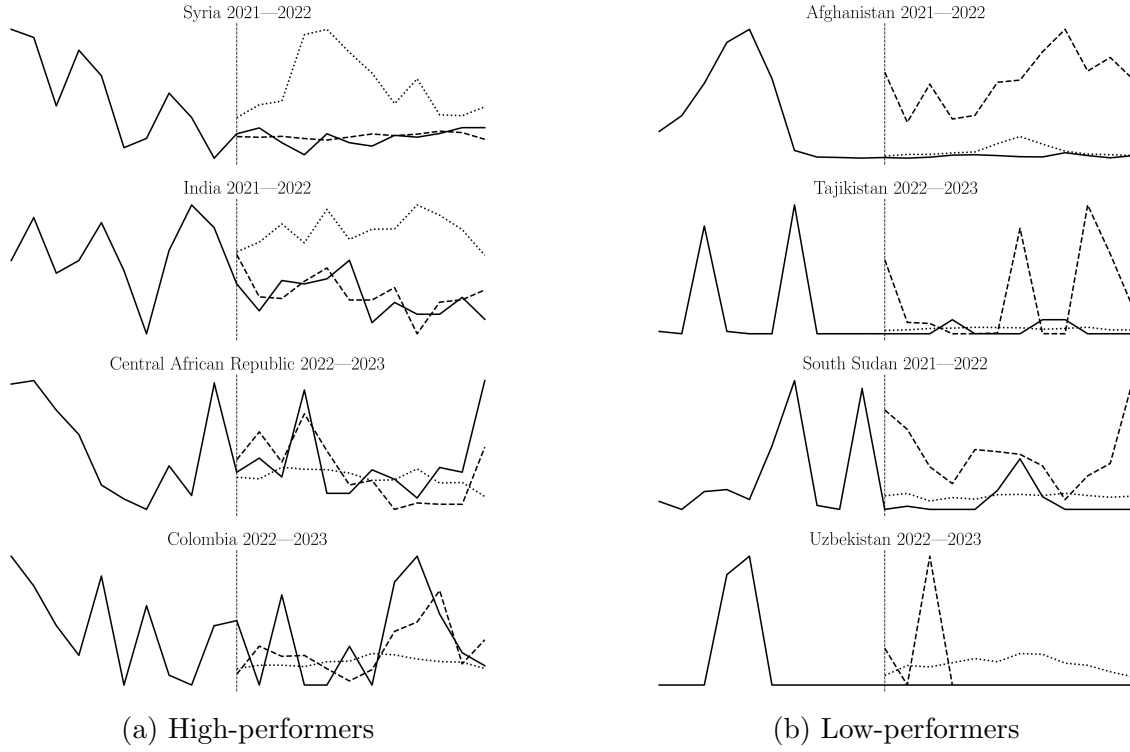


Figure 7: Comparison of input series and predictions for best (a) and worst (b) cases. Input series are to the left of the vertical line. The black line represents actuals, the dotted line represents ViEWS predictions, and the dashed line represents Shape finder predictions. For high complexity input series with substantial variability (a), Shape finder provides more accurate predictions and captures variability, while ViEWS either over- or underpredicts. For low complexity (b), Shape finder produces overpredictions, while ViEWS offers more accurate forecasts.

Poorly-performing cases. In Figure 7b, we present the cases in the poorly-performing cluster with the lowest MSE log-ratio. Low complexity input series lead the Shape finder to overpredict, sometimes significantly. This issue may arise because the model allows a pattern with a low cumulative sum of fatalities to be matched with a high sum case. Patterns with one or two peaks otherwise surrounded by zero fatalities typically occur in low-intensity contexts. When high-intensity cases share these patterns, they are

matched with low-intensity cases, causing overpredictions.²³

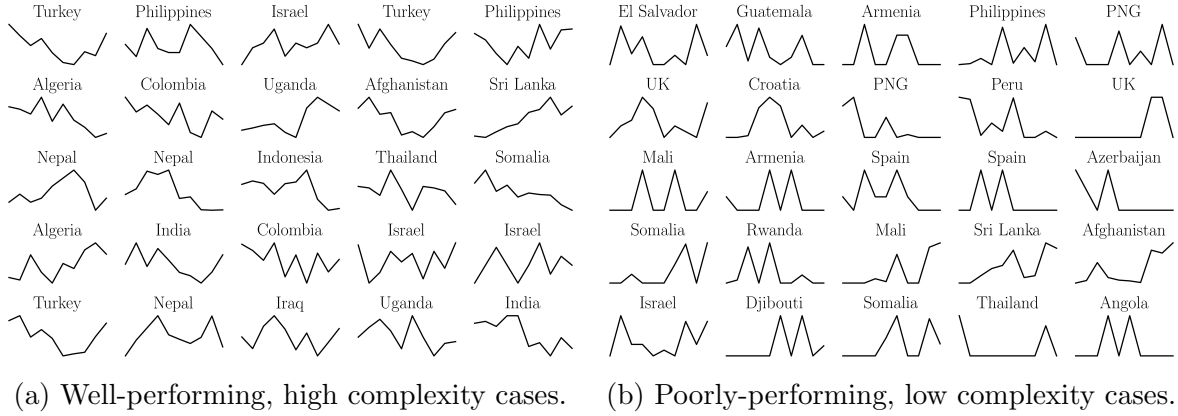


Figure 8: (a) Well-performing input sequences, selected due to having a high mean value and a low standard deviation of the differentiated input sequence. These are high complexity cases, presenting multiple ups and downs in the number of fatalities. The Shape finder seems particularly suitable for those types of cases. (b) Poorly-performing input sequences, selected due to having a low mean value and a high standard deviation (UK=United Kingdom, PNG=Papua New Guinea). These are low complexity cases, presenting single peaks in the middle of zero fatalities. The Shape finder produces overpredictions in those cases, and ViEWS seems more suitable.

Characteristics of well- and poorly-performing cases. Finally, we investigate the characteristics of well- and poorly-performing input sequences. For well-performing cases, we randomly sample sequences with a high mean value and low standard deviation of the differentiated sequence (Figure 8a). For poorly-performing cases, we randomly select observations with a low mean value and a high standard deviation (Figure 8b). These are the characteristics which serve to identify clusters of well- and poorly-performing cases, when comparing the Shape finder to the ViEWS ensemble (for further details, see Figure A.9 and A.10 in the Online Appendix).

The high performing cases correspond to high complexity series, showing high vari-

²³For instance, an increase from 2 to 10 fatalities is not equivalent to an increase from 2,000 to 10,000. This hypothesis is supported by the data in Table A1 in the Online Appendix, where the scale of the input sequences is compared to the scale of the matches. In line with our suggestions, over 50% of the matched patterns have a lower intensity than the input sequence, except for Uzbekistan 2022.

ability with several ups and downs. These shapes have a high mean and a low standard deviation of the differentiated input. Combining those two characteristics reveals sequences with substantial variability, in contrast to flat time series, with single peaks, having a high standard deviation of the differenced series and a low mean. The characteristics of high complexity inputs are most pronounced in Afghanistan, India, Colombia, and the Philippines, all countries with high battle activity (Figure 8a). In other words, the Shape finder performs well when there is substantial armed violence.

For the poorly-performing clusters, we find mostly cases with low complexity, where the input sequences have a relatively low mean value and high standard deviation. These low complexity inputs are characterized by one or two peaks or a small plateau of elevated values followed by a sharp decline, or in other words, low variability. The poorly-performing cluster contains cases such as Spain, Djibouti, the United Kingdom (UK), and Papua New Guinea (PNG) presenting single peaks in fatality counts, and otherwise a value of zero.²⁴

Robustness checks. To ensure the robustness of our findings, we conduct a series of tests in addition to tuning the Shape Finder’s hyperparameters *dist* and *cut*. The tests fall into two main categories—robustness to model specification changes and robustness to variations in model evaluation.

First, we find that changing the window length (*win*) from ± 2 (i.e. 8–12 months) to ± 0 , ± 1 or ± 3 months does not significantly affect performance (Figure A.13 in the Online Appendix). However, larger windows significantly raise computational costs,

²⁴We verify the time series characteristics of well- and poorly-performing cases against the categorization into low- and high-complexity cases, by randomly selecting sequences with a high mean and a low standard deviation of the differentiated series and cases with a low mean and a high standard deviation. Indeed, cases with a high mean and a low standard deviation of the differentiated series are highly complex, showing multiple ups and downs in the number of fatalities (Figure A.12 in the Online Appendix). Conversely, time series with a low mean and a high standard deviation show low complexity, with flat time series and single peaks in the number of fatalities.

such that enlarging the window beyond what we use becomes prohibitively costly. We also explore alternative prediction approaches, including methods that do not involve forecasting based on the majority cluster. One alternative approach involves selecting the closest matches without employing clustering—for instance, directly predicting the future of the input sequence based on the future of the most similar historical pattern. We evaluate the use of the best one, three, and five matches as predictive references; however, this approach significantly underperforms compared to the Shape Finder methodology (Figure A.13 in the Online Appendix). The primary issue with relying on the top matches is the increased likelihood of selecting futures based on outlier cases, which undermines the robustness and reliability of the forecasts.

As a placebo test, we also validate the performance of randomly selecting references from the repository instead of using the matched sequences. As expected, this approach leads to a notable decline in performance, particularly in terms of mean squared error (MSE), as shown in Figure A.13 in the Online Appendix. This outcome underscores the inherent challenge of simultaneously improving both the MSE and the explained variance. Another placebo test involves generating predictions using clusters other than the largest ones. As expected from our premise that larger clusters represent more consistent and reliable patterns, predictions based on smaller or non-dominant (e.g. median) clusters result in less accurate forecasts, which reinforces the importance of focusing on the majority cluster for robust predictions (Figure A.14 in the Online Appendix). This finding aligns with the expectation that when past sequences consistently converge on a particular outcome, that outcome serves as the most reliable predictor for similar patterns.

We also verify that the predictions are well calibrated, meaning that when $\Pr(\text{Cluster A}) = 90\%$, the future belongs to Cluster A about 90% of the time. This would further support the conclusion that the clusters contain meaningful and predictive information.

We therefore assign each future to a cluster and check whether it matches the majority cluster. While calibration is not perfect (Figure A.15 in the Online Appendix), we do observe that higher probabilities for a cluster (e.g. $\Pr(\text{Cluster A}) = 90\%$) correspond to more consistent and predictable futures. Finally, we also confirm that our main findings do not change significantly when increasing or decreasing the balancing factor k in the DE formula (Figure A.5 in the Online Appendix). And we demonstrate that using the DTW distance between predicted and actual futures as a metric instead of DE does not change our findings either (Figure A.7 in the Online Appendix).

Discussion. The Shape finder offers several advantages in comparison to existing conflict prediction models. First, identifying similar historical cases enhances interpretability, enabling researchers to understand how specific predictions are generated—a key characteristic of a trustworthy model (Masís, 2023). Related to this, the inner-workings of the Shape finder are transparent and can easily be comprehended by humans, which is an advantage over similar approaches using deep learning (e.g. Liao et al., 2022; Vafa et al., 2024). This is particularly important in conflict prediction, where practitioners are likely concerned about the trustworthiness of black-box predictions. Second, the Shape finder is entirely autoregressive, requiring no covariates unlike ViEWS and other prediction models. This is particularly advantageous given that most social science data are rarely available at a disaggregated level, making them difficult to use in machine learning applications for complex phenomena such as armed conflict.

Our approach of adjusting the reference repository leads to substantial improvements compared to previous methods that consider all historical cases without filtering (Chadefaux, 2022). By including only the most analogous cases in the reference repository, the Shape finder produces more dynamic and accurate forecasts. While the Shape finder has limitations, such as the inability to predict fatalities following a flat period,

so-called onset cases, its biggest advantage remains its simple autoregressive set-up. Requiring only fatality data and no other correlates, the simple Shape finder achieves performance comparable to much more complex models, such as ViEWS (Figure 6). The ViEWS ensemble combines several constituent models, with at times hundreds of variables (Hegre, Akbari, et al., 2022). This architecture likely requires substantial computational resources. In contrast, predicting 12 months for all countries takes the Shape finder less than 2 hours on a regular desktop computer.

Anticipating conflict onsets—the ‘hard problem of conflict prediction’ (Mueller and Rauh, 2022, p. 2442)—is another area where current conflict prediction models fall short. Based on the autoregressive set-up of the Shape finder, our model cannot forecast conflict onsets either, i.e. the model returns flat predictions based on flat input sequences. Countering this shortcoming, one possible alternative could be a compound model, combining an onset model with the Shape finder. Incorporating variables with more variation, such as news topics (Mueller and Rauh, 2018) or protest indicators, might help address the challenge of predicting conflict onsets.

Conclusion

Conflict prediction has evolved as a thriving subfield of peace research due to its practical relevance (Hegre, Metternich, et al., 2017) and potential for theory validation (Colaresi and Mahmood, 2017). However, forecasting fatalities in armed conflict remains a challenging prediction task. The right skewness of fatalities often leads to overly conservative predictions close to the mean, resulting in risk-averse models which struggle to account for sudden surges and declines in violence. While these models produce accurate predictions when judged by the mean squared error, they fall short in capturing the variability crucial for practitioners seeking to intervene before substantial

escalations of violence (Mueller and Rauh, 2022).

This article introduces the ‘Shape finder’, a risk-taking approach designed to complement existing risk-averse methods in conflict prediction. The Shape finder’s primary advantage lies in its ability to explain variability in fatalities, including sudden escalations or de-escalations. By finding recurring patterns in time series of fatalities and filtering the most similar historical cases into a reference repository, the Shape finder produces more dynamic forecasts. This approach is based on the assumption that similar patterns in the past lead to similar patterns in the future. Patterns in fatality data represent observable implications of interactions between governments and non-state armed groups.

To leverage the unique advantages of both risk-averse and risk-taking approaches, we constructed a compound model combining the Shape finder and the ViEWS ensemble (Hegre, Akbari, et al., 2022)—the most sophisticated prediction model in the field. The compound model is constructed based on the confidence in the Shape finder predictions and the severity of violence, and performs significantly better than both constituent models and two technical baselines—the null model and the $t - 1$ model. Notably, the Shape finder excels in high complexity cases with multiple fluctuations, where ViEWS tends to produce overpredictions or overly conservative forecasts. Conversely, ViEWS performs better in low complexity cases, where the Shape finder tends to overpredict. In simple words, the Shape finder is particularly suitable when there is substantial armed conflict.

In conclusion, the Shape finder offers a valuable complement to existing risk-averse approaches in conflict prediction. Its ability to capture variability in fatalities, coupled with its interpretability and autoregressive nature, makes it a promising tool for researchers and practitioners in the field of peace research and conflict studies. The Shape finder is likely helpful to study social phenomena with a similar data struc-

ture than armed conflict—in particular, rare events, such as protest, riots, and violence against civilian. The theoretical arguments put forward here likely apply to other forms of armed violence and political contention more generally—the interactions between the state and an opposition group produce recurring patterns in sequences of events, which are indicative of the future.

Replication data

The dataset and do-files for the empirical analysis in this article, along with the Online Appendix, are available at <https://www.prio.org/jpr/datasets/> and https://github.com/ThomasSchinca/ShapeFinder_conflict. All analyses were conducted using Python 3.8.5.

Funding

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 101002240), and the Irish Research Council (IRC) in partnership with the Department of Foreign Affairs (DFA) under the Andrew Grene Postgraduate Scholarship in Conflict Resolution (Grant number GOIPG/2023/3883).

Acknowledgements

We would like to thank all participants at the ViEWS Prediction Competition Workshop 2023, in particular Håvard Hegre, Paola Vesco, the German Federal Foreign Office (FFO), and PREVIEW (Prediction, Visualisation, Early Warning), as well as the reviewers and editors at *Journal of Peace Research* for constructive feedback and helpful comments.

References

- ACLED (2023). “ACLED CAST methodology”. Armed Conflict Location and Event Data (ACLED). Available at: https://acleddata.com/acleddatanew/wp-content/uploads/dlm_uploads/2023/07/ACLED-CAST-Methodology—July-2023.pdf (accessed 11 February 2025).
- Axelrod, Robert M. (1990). *The Evolution of Cooperation*. London, New York: Penguin Books.
- Beardsley, Kyle, David E. Cunningham, and Peter B. White (2019). “Mediation, peace-keeping, and the severity of civil war”. In: *Journal of Conflict Resolution* 63(7): 1682–1709.
- Cederman, Lars-Erik and Nils B. Weidmann (2017). “Predicting armed conflict: Time to adjust our expectations?” In: *Science* 355(6324): 474–476.
- Chadefaux, Thomas (2017). “Conflict forecasting and its limits”. In: *Data Science* 1(1–2): 7–17.
- (2022). “A shape-based approach to conflict forecasting”. In: *International Interactions* 48(4): 633–648.
- Colaresi, Michael and Zuhaib Mahmood (2017). “Do the robot: Lessons from machine learning to improve conflict forecasting”. In: *Journal of Peace Research* 54(2): 193–214.
- Davenport, Christian (2007). “State repression and political order”. In: *Annual Review of Political Science* 10(June): 1–23.
- Della Porta, Donatella (2008). “Research on social movements and political violence”. In: *Qualitative Sociology* 31(3): 221–230.
- Della Porta, Donatella et al. (2018). *Social Movements and Civil War: When Protests for Democratization Fail*. Abingdon, Oxon: Routledge.

- Diz-Pita, Érika and M Victoria Otero-Espinar (2021). “Predator–prey models: A review of some recent advances”. In: *Mathematics* 9(15): 1783.
- Dosi, Giovanni, Davide Usula, and Maria Enrica Virgillito (2024). “Increasing returns and labor markets in a predator–prey model”. In: *Journal of Evolutionary Economics* 34(June): 375–402.
- Gleditsch, Kristian S and Michael D Ward (1999). “A revised list of independent states since the Congress of Vienna”. In: *International Interactions* 25(4): 393–413.
- Gleditsch, Nils Petter et al. (2002). “Armed conflict 1946-2001: A new dataset”. In: *Journal of Peace Research* 39(5): 615–637.
- Hegre, Håvard, Forogh Akbari, et al. (2022). “Forecasting fatalities”. Violence Early-Warning System (ViEWS). Available at: <http://uu.diva-portal.org/smash/get/diva2:1667048/FULLTEXT01.pdf> (accessed 11 February 2025).
- Hegre, Håvard, Marie Allansson, et al. (2019). “ViEWS: A political violence early-warning system”. In: *Journal of Peace Research* 56(2): 155–174.
- Hegre, Håvard, Curtis Bell, et al. (2021). “ViEWS₂₀₂₀ : Revising and evaluating the ViEWS political violence early-warning system”. In: *Journal of Peace Research* 58(3): 599–611.
- Hegre, Håvard, Nils W Metternich, et al. (2017). “Introduction: Forecasting in peace research”. In: *Journal of Peace Research* 54(2): 113–124.
- Hethcote, Herbert W (1989). “Three basic epidemiological models”. In: Levin, S, Halam, T, Gross, L. (eds) *Applied Mathematical Ecology*. Berlin, Heidelberg: Springer, 119–144.
- Hossin, M. and M.N. Sulaiman (2015). “A review on evaluation metrics for data classification evaluations”. In: *International Journal of Data Mining & Knowledge Management Process* 5(2): 1–11.

- Kalecki, Michal (1937). “A theory of the business cycle”. In: *The Review of Economic Studies* 4(2): 77–97.
- Keogh, Eamonn and Chotirat Ann Ratanamahatana (2005). “Exact indexing of dynamic time warping”. In: *Knowledge and Information Systems* 7(3): 358–386.
- King, Gary and Langche Zeng (2001). “Explaining rare events in international relations”. In: *International Organization* 55(3): 693–715.
- Kirk, Dudley (1996). “Demographic transition theory”. In: *Population Studies* 50(3): 361–387.
- Koopmans, Ruud (2007). “Protest in time and space: The evolution of waves of contention”. In: In: Snow, DA, Soule, SA, Hanspeter, K (eds) *The Blackwell Companion to Social Movements*. Oxford: Blackwell Publishing Ltd, 19–46.
- Lacina, Bethany (2006). “Explaining the severity of civil wars”. In: *Journal of Conflict Resolution* 50(2): 276–289.
- Liao, Tim F. et al. (2022). “Sequence analysis: Its past, present, and future”. In: *Social Science Research* 107(102772): 1–21.
- Malone, Iris (2021). “The timing of militant violence onset”. Working Paper. Available at: <https://www.dropbox.com/scl/fi/18r4711je60ymrp6y52a9/TimingViolenceOnset.pdf?rlkey=tgsxbgokl5i1z9cz4mlm85&e=1&dl=0> (accessed 11 February 2025).
- Masís, Serg (2023). *Interpretable Machine Learning with Python: Build Explainable, Fair, and Robust High-Performance Models with Hands-On, Real-World Examples*. Birmingham: Packt.
- McAdam, Doug (1983). “Tactical innovation and the pace of insurgency”. In: *American Sociological Review* 48(6): 735–754.
- Mueller, Hannes and Christopher Rauh (2018). “Reading between the lines: Prediction of political violence using newspaper text”. In: *American Political Science Review* 112(2): 358–375.

- Mueller, Hannes and Christopher Rauh (2022). “The hard problem of prediction for conflict prevention”. In: *Journal of the European Economic Association* 20(6): 2440–2467.
- Mueller, Hannes, Christopher Rauh, et al. (2024). “The urgency of conflict prevention – A macroeconomic perspective”. International Monetary Fund (IMF) Working Papers. Working Paper. Available at: <https://www.imf.org/en/Publications/WP/Issues/2024/12/17/TI-Urgency-of-Conflict-Prevention-A-Macroeconomic-Perspective-559143> (accessed 18 February 2025).
- Müllner, Daniel (2011). “Modern hierarchical, agglomerative clustering algorithms”. In: Working Paper. Available at: <https://arxiv.org/abs/1109.2378> (accessed 11 February 2025).
- Pierskalla, Jan Henryk (2010). “Protest, deterrence, and escalation: The strategic calculus of government repression”. In: *Journal of Conflict Resolution* 54(1): 117–145.
- Powell, Robert (2004). “The inefficient use of power: Costly conflict with complete information”. In: *American Political Science Review* 98(2): 231–241.
- Regan, Patrick M. (2002). “Third-party interventions and the duration of intrastate conflicts”. In: *The Journal of Conflict Resolution* 46(1): 55–73.
- Rød, Espen Geelmuyden, Tim Gåsste, and Håvard Hegre (2024). “A review and comparison of conflict early warning systems”. In: *International Journal of Forecasting* 40(1): 96–112.
- Rød, Espen Geelmuyden and Nils B. Weidmann (2023). “From bad to worse? How protest can foster armed conflict in autocracies”. In: *Political Geography* 103(102891): 1–10.
- Slantchev, Branislav L. (2003). “The power to hurt: Costly conflict with completely informed states”. In: *American Political Science Review* 97(1): 123–133.

- Sundberg, Ralph and Erik Melander (2013). “Introducing the UCDP Georeferenced Event Dataset”. In: *Journal of Peace Research* 50(4): 523–532.
- Tarrow, Sidney G (2011). *Power in Movement: Social Movements and Contentious Politics*. Cambridge, New York: Cambridge University Press.
- Vafa, Keyon et al. (2024). “CAREER: A foundation model for labor sequence data”. Working Paper. Available at: <https://arxiv.org/abs/2202.08370> (accessed 11 February 2025).
- Ward, Michael D, Brian D Greenhill, and Kristin M Bakke (2010). “The perils of policy by p-value: Predicting civil conflicts”. In: *Journal of Peace Research* 47(4): 363–375.
- Wong, T.S.F. and MH Wong (2003). “Efficient subsequence matching for sequences databases under time warping”. In: *Seventh International Database Engineering and Applications Symposium, 2003. Proceedings*.
- Young, Joseph K. (2013). “Repression, dissent, and the onset of civil war”. In: *Political Research Quarterly* 66(3): 516–532.

3 The geometry of conflict : 3D Spatio-temporal patterns in fatalities prediction

The geometry of conflict : 3D Spatio-temporal patterns in fatalities prediction

Thomas Schincariol

Abstract

Understanding how conflict events spread over time and space is crucial for predicting and mitigating future violence. However, progress in this area has been limited by the lack of methods capable of capturing the intricate, dynamic patterns of conflict diffusion. The complex nature of those trends needs flexibility in the models to untangle them. This study addresses this gap by analyzing spatio-temporal conflict fatality data using an innovative approach that transforms the data into three-dimensional patterns at the PRIO-grid level. First, our research investigates pattern redundancy and identifies potentially dangerous patterns preceding conflict outbreaks. By applying the Earth Mover's Distance (EMD) algorithm, we detect and classify these patterns, allowing us to compare and match patterns with high adaptive capacity in all dimensions. Using historical similar patterns, we generate predictions of conflict fatalities and compare these with forecasts from the VIEWS ensemble model, a leading benchmark. Our findings demonstrate that recognizing and analyzing conflict diffusion patterns significantly improves predictive accuracy, outperforming the benchmark model. This research contributes to the study of conflict dynamics by introducing a novel pattern recognition framework that enhances the analysis of spatio-temporal data and offers practical applications for early warning systems.

Introduction

The spatio-temporal patterns of conflict have been a central focus for scholars over the past few decades. Understanding the conflict dynamics mechanisms could enhance the performance of fatalities forecasting to improve early warning systems further. Research has shown that one of the most reliable predictors of future fatalities is the lagged values of fatalities in the unit of analysis and its neighboring areas. However, clearly identifying the conditions that favor the diffusion of conflict remains a challenge. Furthermore, most studies have primarily focused on civil conflict (Forsberg, 2014) or specific cases (Saideman, 2012; Kibris, 2021).

Conflict mechanisms are often studied through the lens of diffusion, where fatalities are seen to spread across regions, like an epidemic. Diffusion is mostly studied as a general phenomenon occurring similarly across space and time. However, studying the specific patterns observed in conflict data might bring different information that expands our understanding of conflict dynamics and consequently improves the conflict forecast accuracy. Forecasting models rely on past data to predict future events. Focusing only on historically similar cases can help distinguish valuable information that strengthens forecasting accuracy from unrelated dynamics that introduce noise. For example, (Schutte and Weidmann, 2011) identifies distinct patterns in the spread of conflict across four civil war cases. Similarly, (Kelling and Lin, 2020), using the case of South Sudan, shows how analyzing internal diffusion patterns can deepen our understanding of war dynamics.

The generalizability of this approach and its integration into forecasting models remain unresolved due to the methodologies employed. Most traditional approaches rely on linear models, which lack the flexibility needed to untangle and compare the dynamic diffusion patterns of conflicts. The mechanisms underlying conflict are intrin-

sically complex, requiring models capable of capturing events that, depending on their environment, unfold at varying speeds or scales. Current methods often perform a sophisticated averaging of neighbors, failing to identify specific spatio-temporal patterns. While some machine learning methods offer the potential for the flexibility needed to capture these complex dynamics, their application is restricted by significant drawbacks. These include their dependency on vast amounts of data and limited interpretability, which restricts their capacity to generate clear explanations of findings.

This paper introduces a novel methodology designed to capture complex conflict patterns at the regional level, called ‘ShapeFinder’. Originally developed for country-level conflict forecasting (Schincariol, Frank, and Chadeaux, 2025), the framework is adapted here to accommodate three-dimensional spatio-temporal patterns. Its flexibility enables the comparison and matching of patterns across different geographic scales and intensities. This approach advances our understanding of conflict diffusion mechanisms by identifying significant patterns and contributing to making prediction models more accurate. A notable benefit of this autoregressive model is its simplicity and speed. It does not require any additional variables, which makes it easier to set up and use. The primary strength of the model is its interpretability. By leveraging similar patterns from historical data to construct potential scenarios, it becomes possible to trace the factors driving predictions. The outcome information is twofold. First, the potential future scenarios with their occurrence in the historical data provide transparency on the origins of prediction values. Second, identifying these historical occurrences allows for spatial and temporal pattern analysis. This approach enables the association of specific dynamics with particular regions or time periods.

Additionally, the model can be adapted from a point estimate to a probabilistic outcome, as its initial output consists of multiple potential scenarios, each with an associated probability. This combination of transparency and adaptability ensures that

the model provides clear and comprehensive outcomes, making it accessible even to non-specialists, such as policymakers.

This model contributes by offering a more accessible tool for decision-makers and NGOs, who can use it as an early warning system. Its simplicity and interpretability are key advantages, as complex models like deep learning are often avoided due to their “black box” nature. Thus, this paper adds value both to social science research and to practical, real-world applications. Finally, such a model could be used to focus further on the analysis of conflict patterns and serve as a tool for testing theories. Researchers could investigate whether specific patterns are linked to escalating intensity or the spread of instability. Additionally, the model could assess the impact of interventions by comparing outcomes in similar patterns—one involving an intervention and another without.

Conflict Forecasting at the subnational level

Research over the past decades shows that conflict often spreads across space and time (Black, 2013; Kathman, 2010), especially at the subnational level. Similar to how diseases spread, conflicts can be modeled using diffusion or network models (Dorussen, Gartzke, and Westerwinter, 2016). Studies show that if neighboring regions experience conflict, nearby areas are also at risk in the future (Buhaug and Gleditsch, 2008). As a result, using lagged time and spatial data has become essential for forecasting models (Håvard Hegre et al., 2019). Besides autoregressive factors, variables such as GDP, population, ethnic exclusion (Metternich, Minhas, and Ward, 2017), political instability (Iqbal and Starr, 2008), and variables extracted from new sources like remote sensors (Racek, P. W. Thurner, et al., 2024) also contribute to explain conflict.

To forecast conflict at a regional level, researchers have used advanced statistical

techniques (Fritz et al., 2022), but also machine learning approaches (Bazzi et al., 2022) are increasingly common. Many of these models include spatial and temporal data by incorporating lagged values from neighboring areas using space and time-lagged variables or derived variables like the sum of fatalities of the neighboring areas (Lindholm et al., 2022). However, forecasting conflict remains challenging, especially at finer geographical scales, because conflict data is highly skewed, often with zero values. Despite improvements, most models still face performance issues due to limited flexibility. A recent conflict forecasting competition by the Views team (Vesco et al., 2022) concluded that more flexible models tend to perform better. While newer approaches are reducing model rigidity, such as the non-parametric smoothing method used by (Racek, P. Thurner, and Kauermann, 2024), they remain insufficient for fully capturing conflict dynamics. In the last couple of years, deep learning models have gained popularity, like long short-term memory (LSTM) (Radford, 2022; Maase, 2025), or even temporal fusion transformers (Walterskirchen et al., 2024). Additionally, large language models are adapted to spatio-temporal patterns mining in other fields (Li et al., 2024; Zhang et al., 2024), and offer promising performances. However, those remain poor in interpretability and lack traceability in the predictions.

In recent years, conflict forecasting studies have largely moved beyond country-level analyses. Some studies focus on the first or second-order administrative regions (Aas Rustad et al., 2011), but remain sparse. Most of them that employ finer geographic scales rely on the PRIO-Grid system (Tollefsen, Strand, and Buhaug, 2012), which aggregates data, such as fatalities, into 0.5x0.5 degree cells (about 55x55 km at the equator). This system provides a consistent structure that offers a balance between geographic detail and the limited precision of event location data. Since conflict data is typically extracted from news reports, the exact locations are frequently vague or missing. The appropriate level of aggregation remains a point of debate in the conflict

studies literature (Cook and Weidmann, 2022).

Despite promising results from new methods, predicting conflict at a fine geographic scale remains very difficult. The data is highly zero-inflated and skewed at the country level (Cederman and Weidmann, 2017), and this problem becomes worse at smaller scales, such as the PRIO-grid level. To illustrate how rare these events are, only 0.2% of observations in our study are non-zero. This extreme imbalance makes it hard to gather enough useful information in the training data to build an accurate model. As a result, most prediction models produce vague, smoothed outputs, similar to the results of convolutional smoothing in image processing.

Figure 9 shows conflict predictions for the Kenya–Somalia border during the first half of 2023, from the current best open-source benchmark, the Violence & Impacts Early-Warning System (Views) (Fig. 9b). In the figure, we compare the forecast of the ShapeFinder (Fig. 9c) alongside the previous year’s cumulative fatalities used as model input (Fig. 9a) and the actual observed fatalities (Fig. 9d). Each colored square represents the cumulative fatalities in a PRIO grid cell—darker red indicates more fatalities; transparent cells had none. The VIEWS forecast closely mirrors the input data, predicting high fatalities in the four cells in rows 4-5 and columns 5-6, and medium levels in surrounding cells, resembling the effect of a convolutional blur. This gives a general sense of where conflict might occur, but it results in clear overprediction. In this example, Views predicted 19 cells with more than seven fatalities, ShapeFinder predicted 8, and only 7 were actually observed.

This article introduces a new adaptation of a model called ShapeFinder, which predicts fatalities by identifying and following historically similar patterns. Unlike traditional models that rely on spatial lag variables, ShapeFinder builds its predictions solely from past patterns that closely match the input, allowing for more accurate forecasting. The model is based on the method proposed by Schincariol, Frank, and

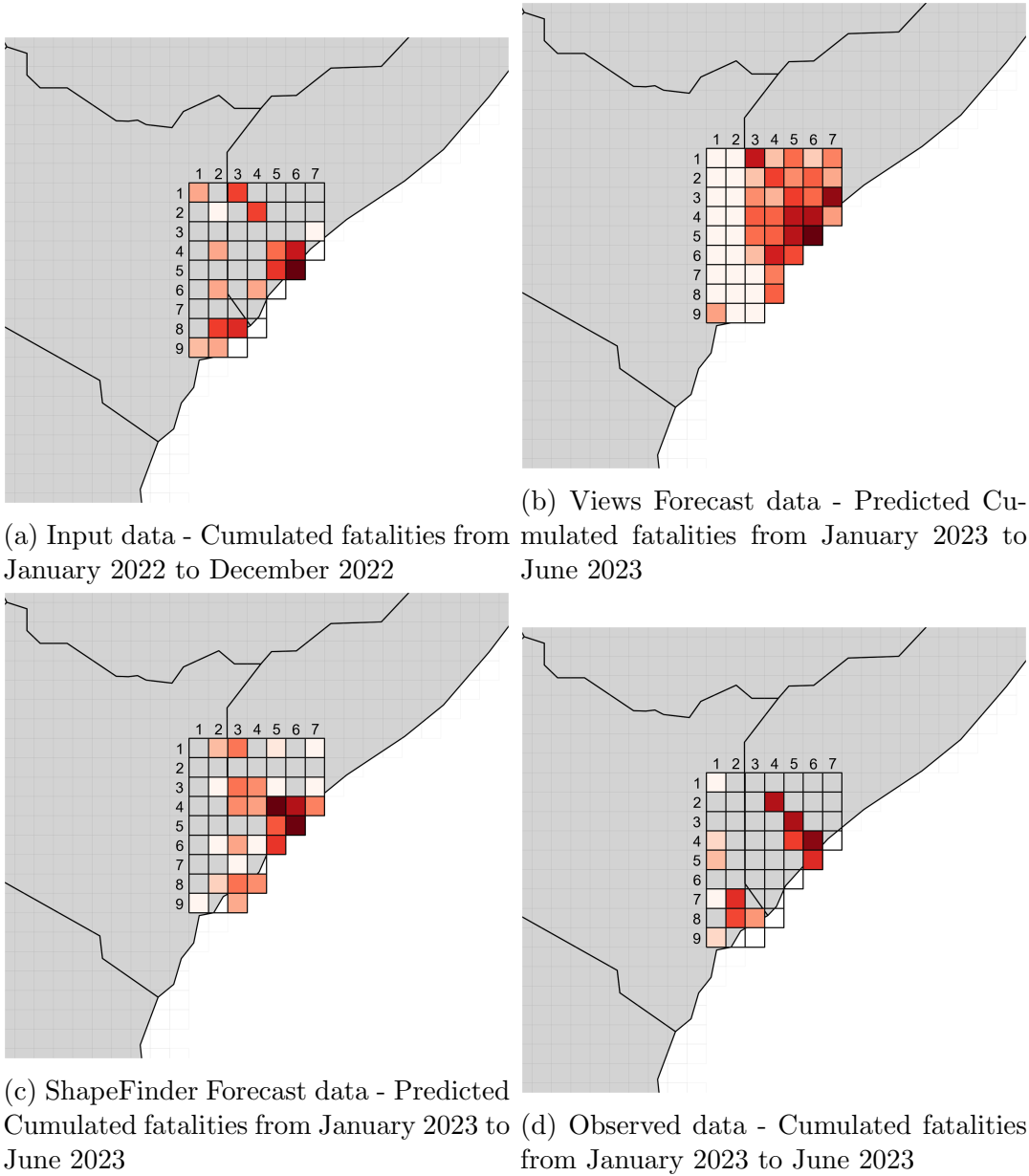


Figure 9: Conflict predictions for the Kenya–Somalia border during the first half of 2023, from Views (b), the ShapeFinder model (c) alongside the previous year’s cumulative fatalities used as model input (a) and the actual observed fatalities (d). Each colored square represents the cumulative fatalities in a PRIO grid cell—darker red indicates more fatalities; transparent cells had none. The numbers in the rows and columns are only included for reference and do not carry any meaning.

Chadefaux (2024), originally developed for country-level time series. Here, it is adapted to work with three-dimensional shapes at the PRIO-grid level. The original approach

uses Dynamic Time Warping (DTW) to measure the similarity between sequences. This distance metric allows comparing two sequences with different lengths. Once similar sequences are found in the historical data, the model predicts future values with the following values of the identified patterns. The model shows strong prediction performance in accuracy but also in the dynamics predicted, highlighting the potential of a shape-based model and the importance of flexibility in such complex applications.

The Prio-Grid problem

The fine-scale geographic detail provided by the Prio-Grid scale demands greater flexibility compared to country-level models. A significant challenge lies in the key assumption of models with spatially lagged variables: the relationships between each Prio-Grid cell and its neighbors are uniform across the grid. This assumption often breaks down in practice. For example, a city like Cairo in Egypt spans several grid cells. If conflict arises within the city, fatalities in two or three neighboring cells might increase rapidly. However, assuming the same relationship between cells within Cairo as between cities like Yaoundé and Douala in Cameroon is difficult to justify. The dynamics of conflict propagation within a single urban area and between two distinct urban centers are not comparable. An effective model should account for these distinctions, treating Cairo as a single cohesive pole while recognizing Yaoundé and Douala as separate but interconnected nodes.

Figure 10 shows an example of an input 3D sequence that happened at the border of Kenya and Somalia in 2022 (left section), corresponding to the three-dimensional representation of the example presented in Figure 9. On the right section, the four most similar cases found by the model within the historical dataset are displayed. The 3D sequence uses longitude on the x-axis, latitude on the y-axis, and time on the z-

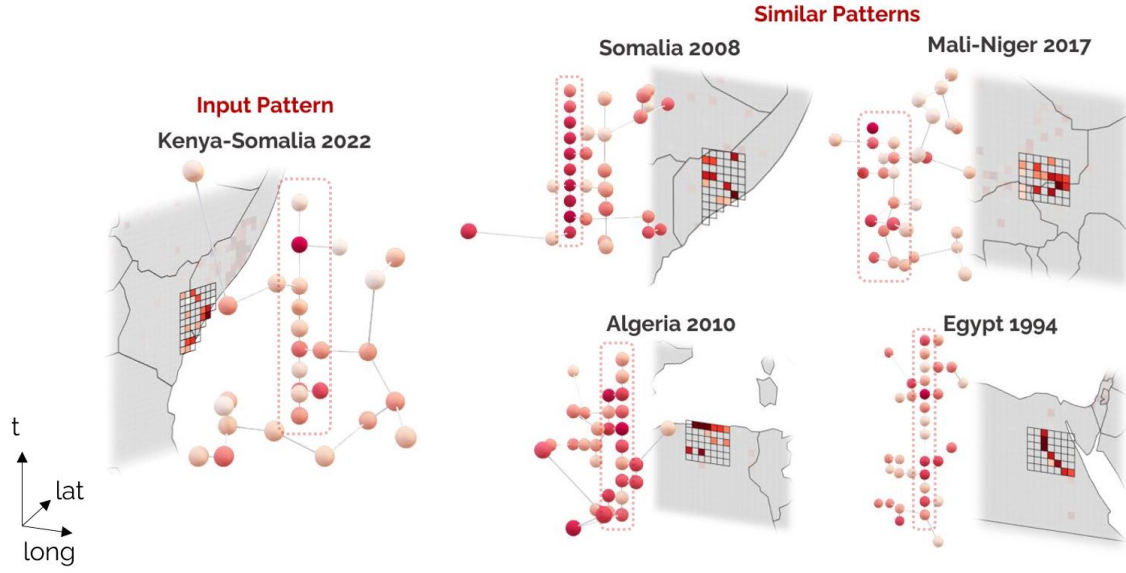


Figure 10: 3D sequences that happened at the border of Kenya and Somalia in 2022 (left section) and four similar cases found in the historical dataset by the model (right section). The 3D sequence uses longitude on the x-axis, latitude on the y-axis, and time on the z-axis. Each point represents the fatality value for a Prio-Grid cell in a specific month. Darker red points indicate higher fatality values, while gray links between points assist in visualizing the 3D structure but carry no interpretative meaning. Accompanying each 3D sequence is a map with accumulated fatalities over the period per Prio-Grid cell. The black grid represents the event’s location, and the colored squares indicate the total fatalities accumulated during the entire period within the Prio-Grid cells.

axis. Each point represents the fatality value for a Prio-Grid cell in a specific month. Darker red points indicate higher fatality values, while gray links between points assist in visualizing the 3D structure but carry no interpretative meaning. Accompanying each 3D sequence is a map with accumulated fatalities over the period per Prio-Grid cell. Similarly to Figure 9 format, the black grid represents the event’s location, and the colored squares indicate the total fatalities accumulated during the period within the Prio-Grid cells.

Across the five patterns, a consistent structure emerges: one cell or a small cluster of cells accounts for most fatalities, maintaining steady intensity over the entire period

(highlighted in dotted red on the figure). Occasionally, spreads of conflict appear in surrounding cells, but they rarely persist for more than a few months. Visually, these patterns resemble a tree, where the core cells with constant fatalities form the trunk, and the short-term spreads constitute branches.

Although the structures appear visually similar, detecting these patterns using a model is significantly challenging. One reason is the varying dimensions in latitude and longitude across the patterns: the Kenya-Somalia case is compact, while others, like Algeria or Mali-Niger, span multiple cells. Another complication is the difference in temporal durations: for example, the Egypt pattern extends over 16 months, whereas the Kenya-Somalia case lasts only 12 months. Additionally, the “trunk” structure varies considerably; in Somalia, it is concentrated in a single cell, while in Algeria, it spreads across two cells, and in Mali-Niger, it involves a group of four cells.

Traditional models, which rely on fixed spatial and temporal lags or neighbor sums, would fail to adequately identify the underlying similarity between these patterns. As a simplified toy model, Figure 11 presents three patterns on a 5×5 grid over three months. As in the previous figure, the x-axis represents longitude, and the y-axis represents latitude. The cell color indicates the number of fatalities per month, ranging from 0 (white) to 9 (dark red), with darker shades representing higher values. Visually, patterns 1 and 2 appear to have more similar dynamics, while pattern 3 is distinct, with constant values throughout the three months. However, when applying classical spatial regression methods to measure distances between patterns, the results indicate a smaller distance between patterns 1 and 3 than between patterns 1 and 2. In contrast, the flexible method introduced in this paper correctly identifies a smaller distance between patterns 1 and 2. Detailed calculations are provided in Appendix A.2.

Incorporating flexibility into models is essential for matching cases with similar dynamics despite differences in speed, scale, or frequency. This flexibility has shown clear

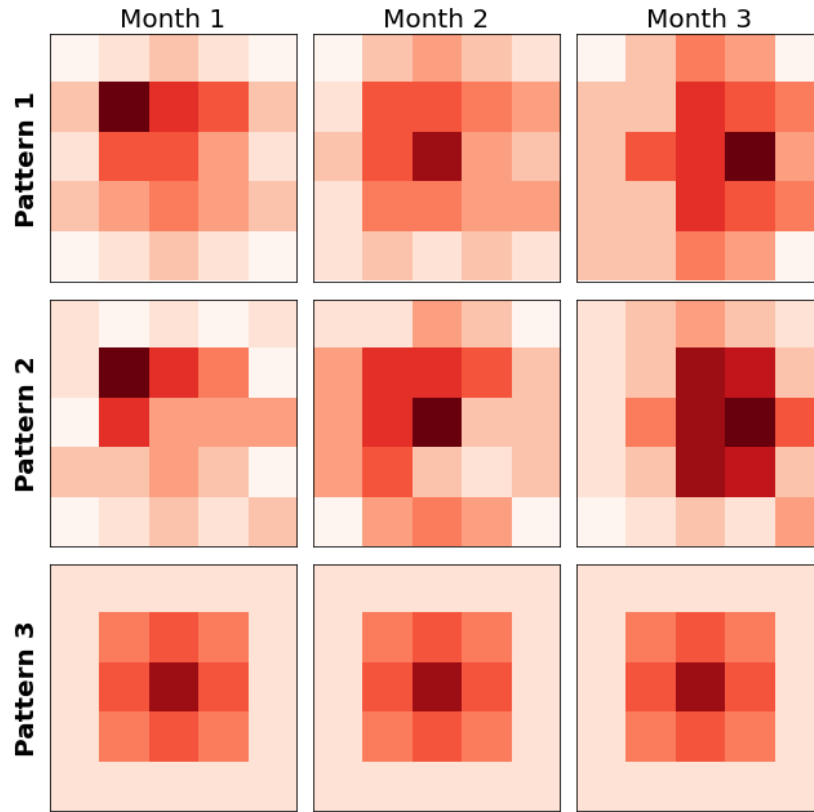


Figure 11: Three patterns on a 5×5 grid over three months. The x-axis represents longitude, and the y-axis represents latitude. The cell color indicates the number of fatalities per month, ranging from 0 (white) to 9 (dark red), with darker shades representing higher values.

improvements even in single-dimensional models (Schincariol, Frank, and Chadeaux, 2024), but when extended to three dimensions, its importance becomes even more pronounced. Events can vary in velocity across three axes—space and time—introducing additional layers of complexity that rigid models cannot capture.

Data Creation

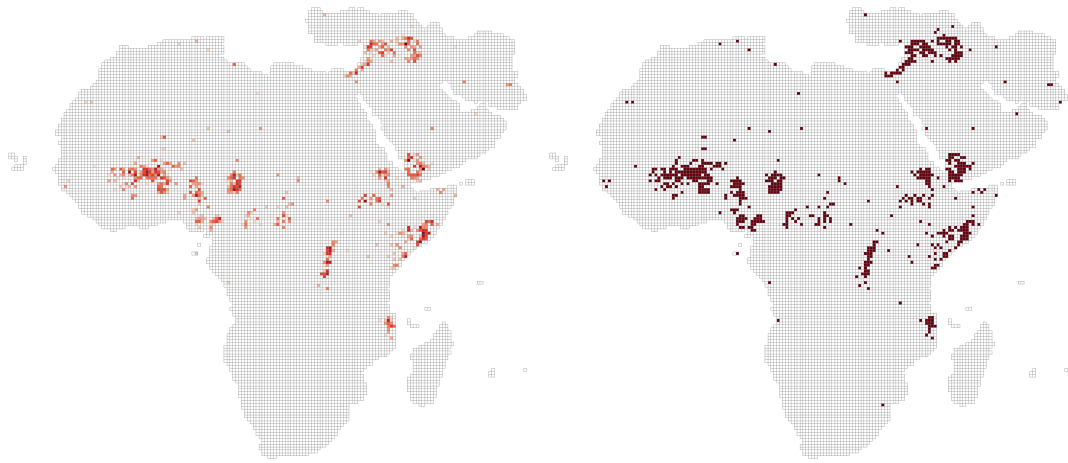
Data source

The data on conflict fatalities used in this study come from the Uppsala Conflict Data Program (UCDP) Georeferenced Event Dataset (GED) (Sundberg and Melander, 2013; Davies, Pettersson, and Öberg, 2022). This dataset provides monthly data at the PRIO-Grid level, covering the period from 1989 to the present. To validate our forecasting results, we compared them with the Views forecast ensemble model (Håvard Hegre et al., 2022), which is a leading benchmark in conflict prediction. The Views model is an ensemble model that includes various machine learning models, including a large number of covariates. Since the model only forecasts the state-based fatalities in the Middle East and Africa, our study focuses on these regions and variable to enable a direct comparison. The data was obtained through the Views API.²⁵

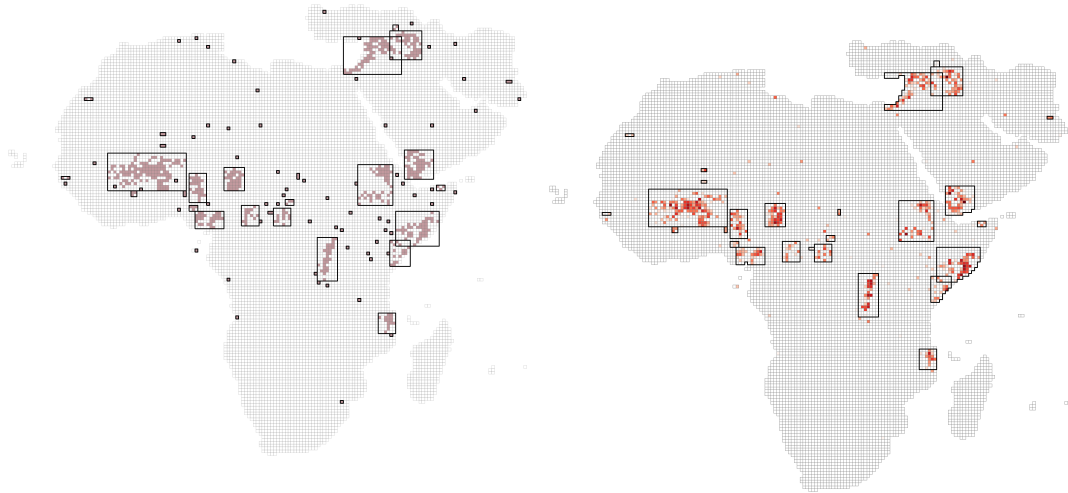
Pattern identification

The first step of our model is to identify potential future zones of danger by applying object-detection techniques from image analysis, like connected-component labeling and morphological erosion. The four steps of the method are represented in Figure 12. This process begins by examining the previous 12 months prior to the forecast period. Fatalities from the previous year are aggregated within each PRIO-Grid cell, resulting in a 2D grid where each cell contains the sum of fatalities for the year (Fig. 12a). The model then scans the grid using a rolling window across latitude and longitude to identify cells with non-zero values, referred to as “active” cells (Fig. 12b).

²⁵The datasets of the four windows of prediction are: *fatalities001_2022_00_t01*, *fatalities001_2022_06_t01*, *fatalities001_2023_00_t01*, *fatalities001_2023_06_t01*. The predictions produced by Views are available here. https://github.com/prio-data/views_api/wiki/Available-datasets



(a) Cumulated fatalities from January 2022 to December 2022 (b) Binary classification of active cell, with at least one cumulated conflict fatalities



(c) Outcome of the Connected Component Labeling. The black borders outline the identified patterns. (d) Union-find patterns without filtered isolated cells by morphological erosion

Figure 12: Four steps of the identification of active conflict zone patterns in 2022 in Africa and the Middle East. The red shading represents the intensity of conflict, with darker red indicating higher numbers of fatalities in (a) and (d).

Once the active cells are identified, Connected Component Labeling (CCL), a well-established technique in image analysis, is applied (Fig. 12c). CCL is commonly used in image analysis to detect and classify objects based on pixel connectivity. Once an

active cell is identified, its neighboring cells within a 2-unit radius are examined. If a neighboring cell is also active, it is grouped into the same cluster. This iterative process continues until no additional active neighboring cells can be found. The resulting networks of interconnected active cells are classified as patterns. The radius is set to 2 units, corresponding to approximately 100km, allowing flexibility of the detection and remaining relatively close enough to have a link. The final step applies a modified version of morphological erosion and dilation (Fig. 12d). Erosion removes small structures by eroding object boundaries, eliminating minor, irrelevant details. Dilation expands objects, fills small gaps, and reconnects structures. These techniques are commonly used in remote sensing, such as for cloud detection, where erosion removes noisy pixels that may be mistaken for small clouds, and dilation merges fragmented cloud pixels into a coherent shape. Here, the erosion is applied by retaining only networks with more than one active cell. Indeed, as shown in Figure A.21, isolated active cells are associated with future conflict only 13% of the time and never diffuse to the neighboring cells. A spatial model would not be appropriate. However, these excluded cells may present opportunities for future work, such as applying a non-spatial autoregressive model as described in Schincariol, Frank, and Chadefaux, 2024, though their exclusion does not significantly impact the overall evaluation (see the Discussion section). The dilation is done by the union of entirely overlapping identified patterns. If two patterns share some cells, the mean of the predictions is considered.

Using this pattern identification method, 99% of the total fatalities and 94% of active cells are considered. The remaining 6% of active cells are isolated and not linked to any larger networks. This aggregation method provides a compromise between maximizing the inclusion of relevant fatalities for effective model training and maintaining computational efficiency. By doing so, the model remains capable of running in just a few hours.

Data transformation

To implement our shape-matching method, the raw data undergoes a transformation process, as illustrated in Figure 13.

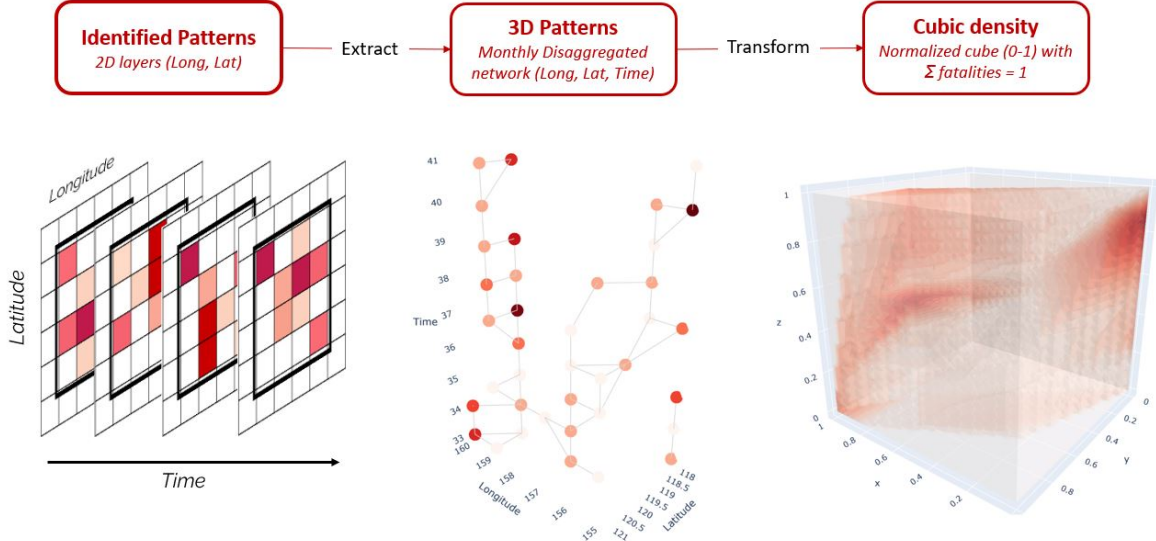


Figure 13: Data transformation process illustrated through three steps. (1-left) Patterns are identified in an aggregated 2D grid network of active cells (as discussed in the Pattern identification section) (2-middle) Patterns of linked neighboring cells with conflict fatalities are extracted and represented in a “molecular” format, where Longitude is on the x-axis, Latitude on the y-axis, and Time on the z-axis. Darker red shades indicate higher fatality values. Gray links between cells aid in 3D visualization but do not signify any meaningful connections. (3-right) The extracted patterns are transformed into density cubes, with each side representing a single normalized value and the intensity of conflict as the density value. As in (2), darker red indicates higher density values.

After the pattern of interest identification, we extract patterns of conflict cells from the network of 2D layers. The 2D grid is then disaggregated into a 3D pattern with Longitude and Latitude as the x-axis and y-axis, and month as the z-axis. An example of such a network is shown in the middle part of the figure, where each colored point represents a Prio-Grid/month cell, with darker red indicating higher fatalities. The gray lines connecting the points are added for visual clarity and do not carry any

specific meaning. In this example, the points are connected, forming a network with two branches—one spanning longitudes 160 to 159 and the other from 157 to 155—both connected by a common base around time 33 to 35.

Finally, the network is transformed into a density cube, as shown on the right side of the figure. The three axes—latitude, longitude, and time—are normalized from 0 to 1. The total number of fatalities within the cells is also normalized so that their sum equals 1, similar to a standard density function. This final step results in a cube with 1 unit per side, containing a 3D density distribution.

Method

The structure of the model is illustrated in Figure 14. The process begins with the extraction of patterns of interest (see Pattern identification), referred to as input patterns in the figure. Next, the model searches for similar patterns in the historical data using a rolling window. This involves calculating two distance metrics: the Earth Mover’s Distance (EMD) and the ratio of active cells (R). If both metrics fall below a defined threshold, the historical pattern is classified as similar to the input pattern. Finally, the subsequent values associated with these similar patterns are clustered to generate the forecast values.

Distance between two 3D Pattern

The distance metrics for two 3D patterns combine two comparison factors: (1) the Earth Moving Distance (EMD) between the cubic density representations of the patterns, and (2) the ratio of cells with at least one conflict between the two patterns.

The Earth Mover’s Distance (EMD) (Andoni, Indyk, and Krauthgamer, 2008) quantifies the similarity between two distributions. Initially developed for transportation

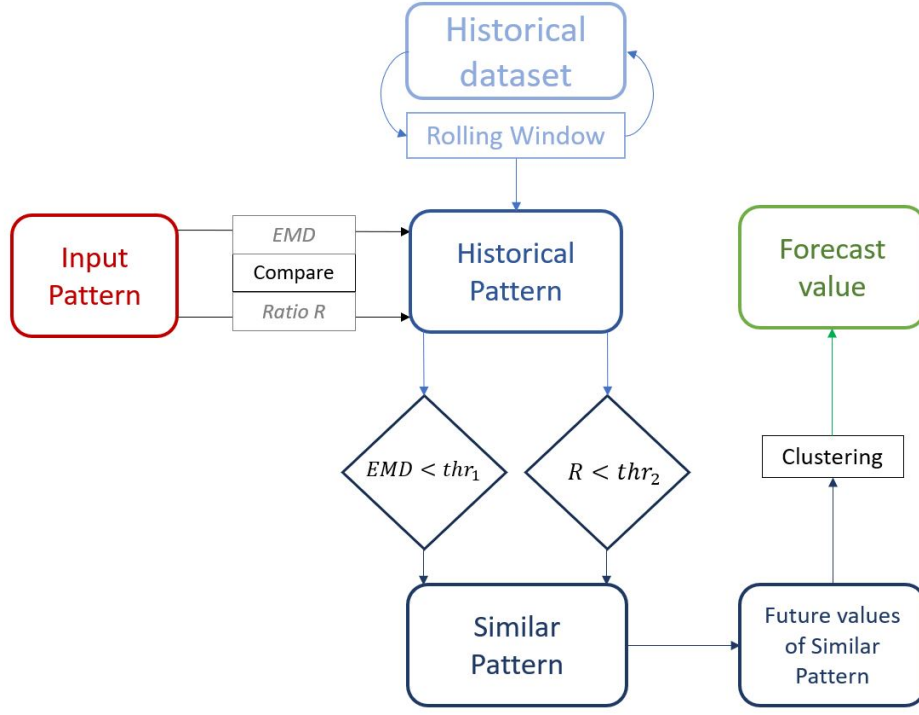


Figure 14: Diagram of the overall method of the model. First, we extracted the patterns of interest as mentioned before, called input patterns in the figure. Then, we look for similar patterns in the historical data using a rolling window, by calculating two distance metrics, Earth Moving Distance (EMD) and ratio of active cells (R). If those two metrics are lower than a threshold, the historical pattern is classified as a similar pattern to the input one. Finally, the values that follow similar patterns are clustered to produce the forecast values.

logistics, it has been adapted for applications in image processing and computer vision. The metric evaluates the amount of work needed to transform one distribution into another. This is achieved by solving an optimal transport problem, minimizing the total cost required to transport the mass. The EMD is defined by the following equation:

$$\text{EMD}(P, Q) = \min_{\gamma \in \Gamma(P, Q)} \sum_{i,j} \gamma_{ij} c(x_i, y_j) \quad (2)$$

where:

- P and Q are the two distributions of points with non-null values for each pattern.

- $\Gamma(P, Q)$ is the set of all possible joint distributions (couplings) with marginals P and Q .
- γ_{ij} represents the difference in normalized fatalities from x_i to y_j .
- $c(x_i, y_j)$ is the Euclidean distance between x_i to y_j .

It is important to note that we compare each pattern not once but four times. The cubic density is rotated 90° around the time axis three times more because a pattern that starts in the southeast and ends in the northwest could be similar to one that starts in the southwest and ends in the northeast.²⁶ The rotation is done around the time axis that remains fixed, as time progresses in only one direction. We take the minimum EMD value of the four pattern orientations. This distance metric allows us to identify similar patterns that may vary in duration and geographic spread—patterns that traditional models often fail to capture. By accommodating variations in both time and space, our method provides a more adaptive understanding of conflict dynamics, enabling the detection of patterns that may be overlooked by classical models due to their rigidity. This flexibility is crucial for accurately analyzing complex phenomena where the timing and geographic extent of events can differ significantly.

The second part of the ratio of active cells (R) of the two patterns. An active cell is a Prio-Grid month observation that has a non-zero value. The R ratio is defined by the following equation:

$$R = \left| \tanh \left(\log \left(\frac{N_{p1}}{N_{p2}} \right) \right) \right| \quad (3)$$

where:

- N_{p1}, N_{p2} is the number of conflict active cells in the first and second pattern

²⁶A future possibility could be to implement a cylinder density instead of cubic and then find the angle that minimizes the EMD.

- R is the Ratio of Active Cells
- $\tanh$ is the hyperbolic tangent function

EMD alone does not account for the counts of conflicts or whether the patterns have a similar number of active cells, those with non-zero fatalities. Thus, two patterns could have similar shapes (low EMD) but very different counts of active cells. Although the model assumes that patterns with different magnitudes are comparable, it does not equate cells with zero fatalities to those with a small number of fatalities. The Ratio of Active Cells function 3 is equal to zero when the ratio $\frac{N_{p1}}{N_{p2}}$ is equal to 1. It increases the same way whether the ratio goes negative or positive, like this we do not have to control the numerator and denominator.

The first factor ensures that the overall shapes of the patterns are similar, while the second ensures that both patterns have a comparable number of active conflict cells. By combining both factors, this method provides a more comprehensive and reliable approach to pattern matching. Two patterns are considered similar if they fall below two critical threshold values. These thresholds were established based on a parameter grid search minimizing the evaluation metrics of the validation set corresponding to a forecasting window from January to June 2021, using data up to December 2020.²⁷

Search for similar patterns in the historical data

To identify similar historical patterns for a given input, we use a rolling window method, as detailed in 15. Input patterns (highlighted in red in the figure) are compared to historical patterns spanning from southwestern Africa in 1989 to the northeastern Middle East in 2023. The rolling window moves sequentially through longitude, latitude, and time.

²⁷The Earth Moving Distance (EMD) threshold is 0.15 and the R ratio one is 0.05.

Forecast creation

The forecast values are generated using the “past future” of similar historical patterns. This process has two main steps: (1) reshaping the past future values to fit the input dimensions, and (2) clustering and identifying the most likely scenario. Because we allow comparing patterns with different dimensions, some similar patterns may differ in size. Extending a pattern is done by multiplying dimensions, while larger patterns are shrunk by dividing dimensions to fit the input size. Forecast values are then adjusted to align with the closest matching grid cells. If multiple values fall in the same cell, they are added together.

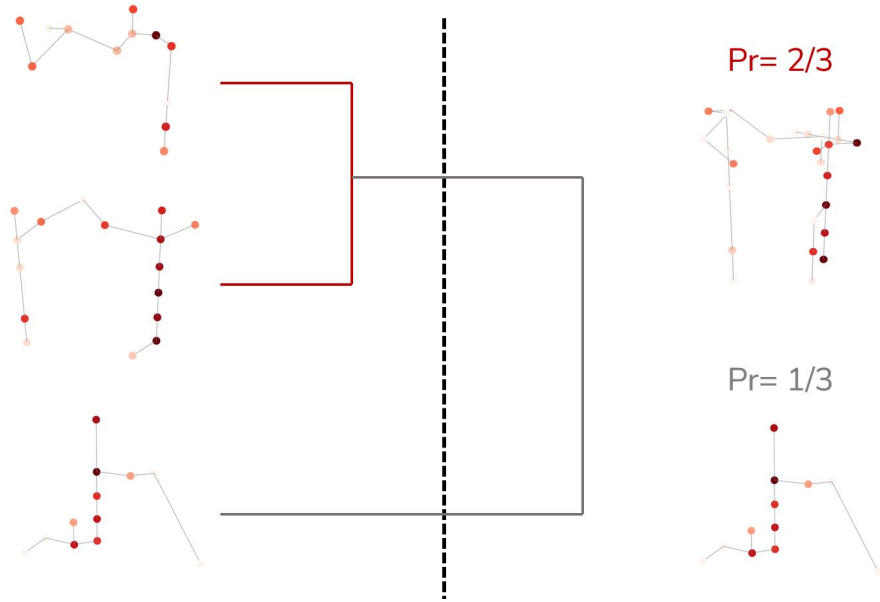


Figure 16: Example of scenario creation using clustering. Three similar patterns were identified in the historical data (left) and created two scenarios (right) with their associated probability, based on the number of observations in each cluster.

Next, we apply clustering on past patterns to find the most probable forecast scenario. After mapping each pattern’s projection to longitude, latitude, and time, we calculate the Euclidean distance between these projections and cluster them based on a set threshold distance. Similarly to the previous distance parameters, the clustering

threshold was established based on a parameter grid search to minimize the evaluation metrics of the validation set. Scenarios are then created by grouping the values by cluster using the mean value for each cell point (longitude, latitude, month). The number of observations per cluster estimates the probability of observing this outcome emerge when a similar input pattern appears. The forecasted value is then the mean value of the cluster with the most observations, representing the most probable future scenario. Figure 16 presents an example of scenario creation using clustering. Three similar patterns were identified in the historical data (left). The two upper patterns have a distance below the defined threshold (indicated by the dotted line), and are then grouped together to form a scenario with a 66% probability. As this is the most probable scenario, it is selected as the forecast value.

This step helps mitigate the influence of outliers, which could heavily weight the prediction with extreme values that rarely have occurred in history. An alternative approach would be to extend this point-estimate forecasting framework into a probabilistic one. Indeed, a distribution could be fitted at each cell (defined by longitude, latitude, and month) using the observations from similar historical patterns.

Evaluation

Benchmark model

We evaluate our model, labeled as ‘ShapeFinder’ (SF), by comparing it to the leading benchmark in the field, the Views forecast ensemble model (Håvard Hegre et al., 2022). This model is one of the few that provides a publicly available monthly point forecast at the grid level. It forecasts up to 36 months ahead for the Middle East and Africa,

starting from January 2022.²⁸ Our evaluation is conducted across four 6-month periods:

- January to June 2022, using data up to December 2021.
- July to December 2022, using data up to June 2022.
- January to June 2023, using data up to December 2022.
- July to December 2023, using data up to June 2023.

These intervals provide a balance between a long enough forecast horizon and sufficient observations. A six-month horizon, half the length of the 12-month input window, offers the best trade-off between dynamic prediction reliability and the potential operational utility. A shorter, three-month horizon may not be sufficient for a warning system, while a 12-month horizon likely exceeds the theoretically temporal limits of reliable dynamic prediction.

Metrics

The models' performance is measured using three metrics. The first is the Mean Squared Error (MSE) at the Prio-Grid level, the most common metric in forecasting. MSE measures the average prediction error but does not account for flexibility in geography or time. It simply checks if the forecast at time t for Prio-Grid p is accurate. The formula for MSE is:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4)$$

With Y_i the observed value and $\hat{Y}_i$ the predicted value.

While the metrics at the Prio-Grid level provide valuable insights, they do not fully capture the primary value of the model. Given that the Prio-Grid operates at a very fine

²⁸It is important to note that their model is optimized with 36-month horizons, and these predicted values are actually live predictions.

spatial scale, it is essential to consider the broader dynamics across entire zones. The true utility of this model lies in its ability to predict the overall amplitude of fatalities and capture the shape of the dynamic patterns within a given area. To better assess the model’s effectiveness in forecasting, we employ two metrics based on the zones defined in Figure 12. First, we calculate the error in predicting the total number of fatalities within each zone over the entire period. Second, we assess the model’s ability to capture the spatio-temporal dynamics using the EMD metric. Together, these metrics evaluate the total fatalities and the spatial pattern of predictions.

To extend the analysis from forecasting evaluation, we need to characterize a pattern and its relationship with its future values. For this, we introduce a metric called ratio of increase (R_{inc}), which quantifies the difference in the number of fatalities between a pattern and its past future. This metric helps assess whether a pattern leads to an escalation, stabilization, or decline in conflict intensity. The formula for R_{inc} is:

$$R_{inc} = \frac{\sum fat_{pf}}{\sum fat_{pattern}} \quad (5)$$

With $\sum fat_{pf}$ the sum of fatalities in the past future section, and $\sum fat_{pattern}$ the sum of fatalities in the pattern.

Results

Patterns and Past Futures

Figure 17 illustrates the Kenya-Somalia example, discussed in Figure 10. The left side of the figure displays four “following values” examples (also referred to as past futures) of similar historical patterns. These examples are extracted from the most probable scenario, which includes ten patterns in total. The right side of the figure compares the

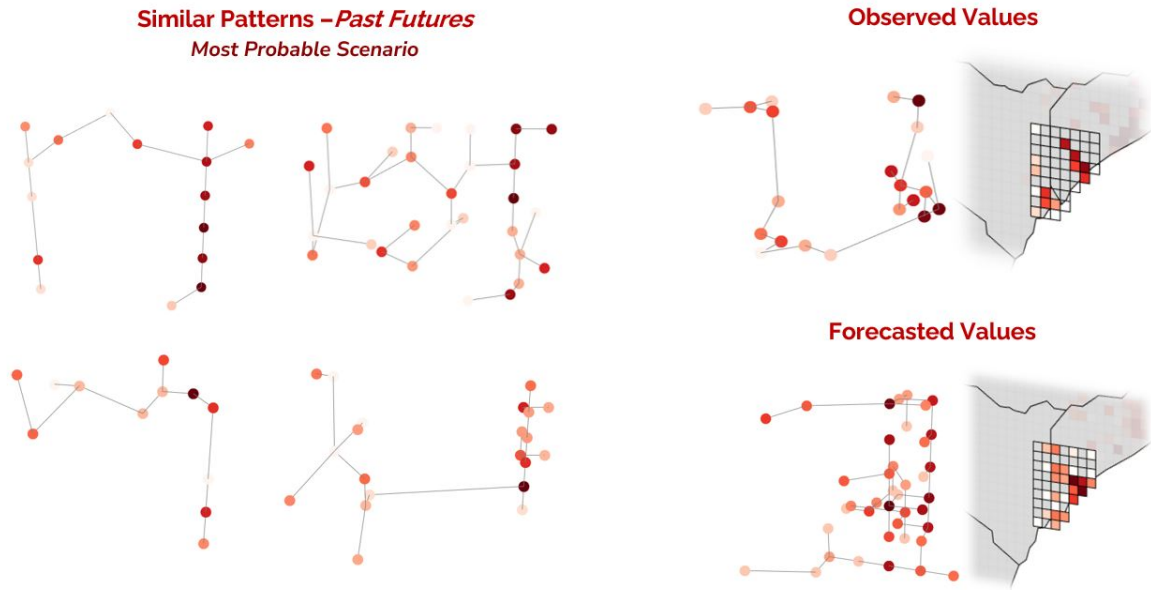


Figure 17: Outcome of the ShapeFinder model for the Kenya-Somalia example, discussed in Figure 10. Four examples of past futures, and the corresponding Forecasted and Observed values. The left side of the figure displays four “following values” examples (also referred to as past futures) of similar historical patterns. The right side of the figure represents the forecasted and actual observed values.

forecasted and actual observed values. The forecasted values are constructed with the following values of similar historical patterns shown on the left. The similar patterns examples share a common dynamic: most fatalities are concentrated in a couple of Prio-Grid cells, located on the right side of the pattern. The patterns also have a secondary zone of fatalities, more minor. In some cases, this secondary zone is concentrated within one or two cells on the left part, while in others, it appears more diffusely distributed around the primary zone of concentration.

The resulting forecast mirrors the observed values to a significant extent, correctly predicting the concentration of fatalities in two cells along the coast near the border. However, while the observed values reveal a cluster of fatalities in the south, this area is slightly underpredicted in the prediction. Instead, the model forecasts a more diffuse spread of fatalities around the two primary cells, failing to capture the precise secondary

cluster. More examples of forecasted projections of fatalities from the SF model and the Views model against the actual observed values are included in Figure A.17.

Meaningful patterns ?

A shape-based forecasting model operates on the assumption that similar patterns lead to similar outcomes. To evaluate this assumption, we examine whether the similarity of past futures between two patterns correlates with their distance. For this, we analyze all patterns in the study and compare them against historically recorded patterns that share comparable dimensions and active cell counts. We measure both the patterns' similarity metric EMD ($EMD_{pattern}$) but also the differences in their ratio of increase (ΔR_{inc}) and EMD for their past futures (EMD_{pf}).

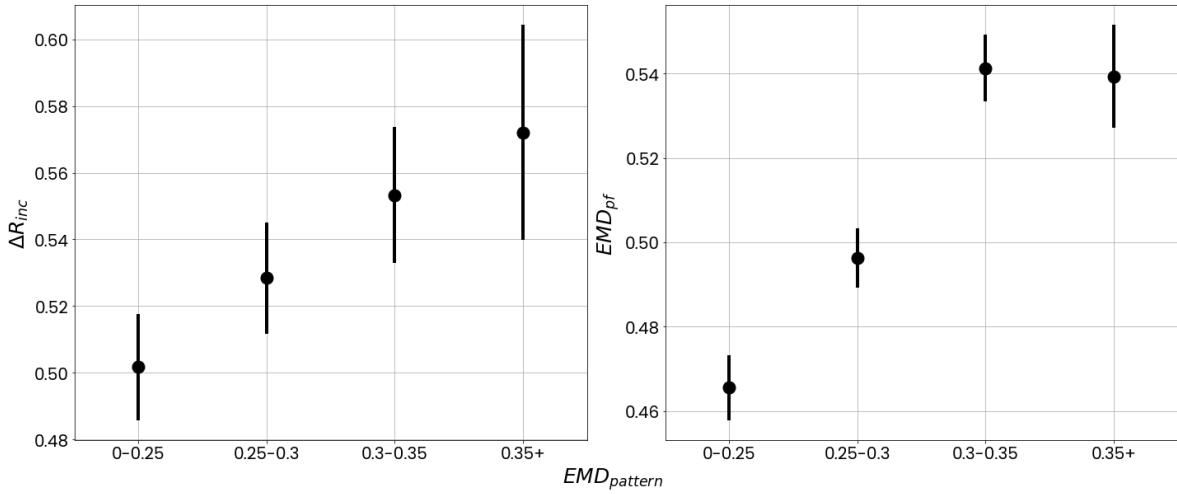


Figure 18: The mean and 95% confidence interval for patterns' similarity metric $EMD_{pattern}$ (x-axis) and the differences in their ratio of increase ΔR_{inc} on the left part and EMD for their past futures EMD_{pf} on the right (y-axis).

Figure 18 presents the results, where values are binned for clarity. The mean and 95% confidence interval for each bin are displayed for R_{inc} (left) and EMD_{pf} (right).²⁹

²⁹This interval is constructed as $\pm t_{0.975, n-1} * SD / \sqrt{n}$ where $t_{0.975, n-1}$ is the critical value of t , SD the standard deviation and n the number of observations. This formula is used for all confidence intervals presented in the remainder of the paper.

Both R_{inc} and EMD_{pf} of the smaller bin are significantly smaller (except the 0.25-0.3 bin for ΔR_{inc}). Generally, when $EMD_{pattern}$ increases, R_{inc} and EMD_{pf} increases too. It confirms that more similar patterns tend to produce more similar outcomes (past futures), validating the model’s core assumption.

Prediction at the Prio-Grid level

First, the Prio-Grid cell is the unit of analysis. In this section, we compare the model results at the micro-level, providing insight into the prediction performance at the smallest geographical grain of analysis.

Figure 19 shows the log ratio of Mean Squared Error (MSE) between the Views model and the ShapeFinder (SF) model at the Prio-Grid level. The units in the two subfigures (bottom left of the figure) are the metrics calculated for the 6-month prediction period and Prio-Grid. Across the four periods analyzed, approximately 1,500 Prio-Grid cells per period (totaling just over 6,000 cells) are included. A positive ratio (in red) indicates that the ShapeFinder model outperforms the Views model. Overall, the SF model demonstrates significantly better performance, with an average log ratio of 0.05. The median log ratio is positive, and in total, the SF model shows improvements in over 62% of cases. However, many of these improvements are minor, as indicated by the substantial number of points included in the histogram bin from 0 to 1. Along with this, a map displays cumulative MSE per cell over four analysis periods. Red indicates better performance by the SF model, blue by the Views model, and gray denotes similar predictions. The Views model performs better in northern Iraq, the Central African Republic, and the Chad-Nigeria border. Mixed results appear in Burkina Faso, the DRC-Rwanda border, and Somalia. Finally, Shapefinder outperforms in Yemen, Nigeria, central Ethiopia, and Syria. Appendix A.2 includes examples of predictions for the Prio-Grid cells with the best and worst performance differences.

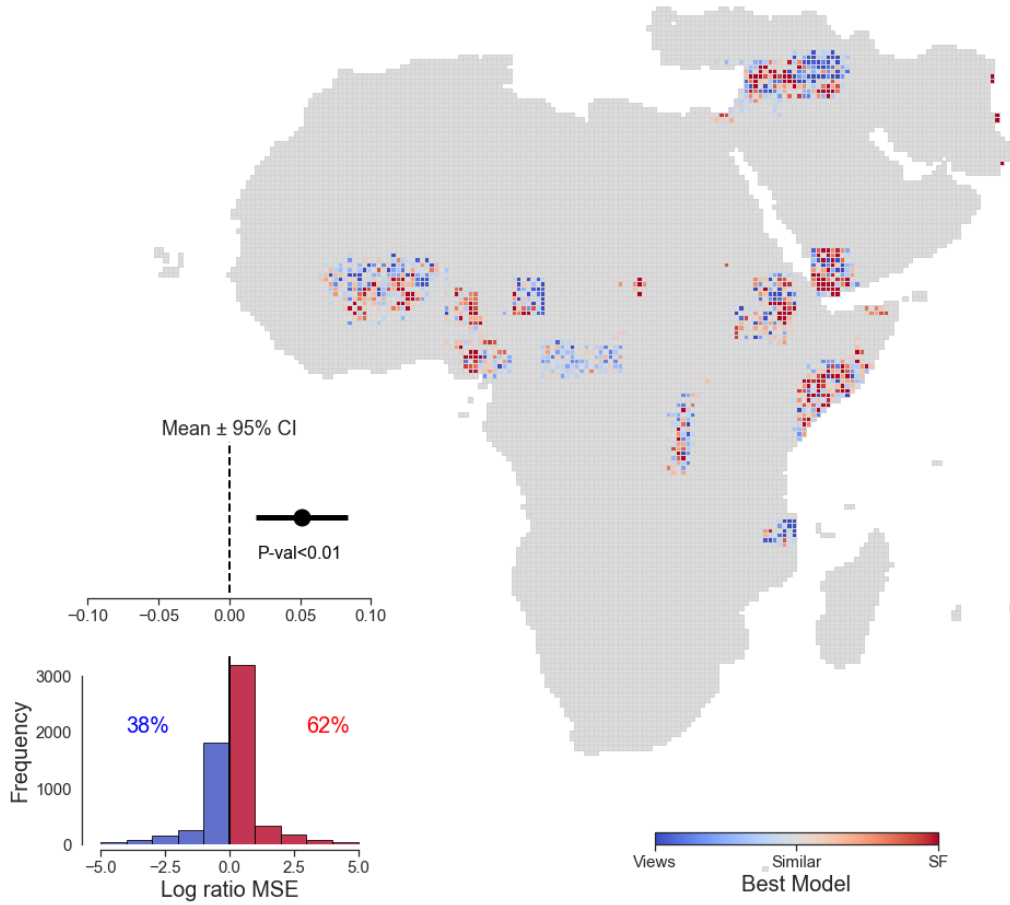


Figure 19: Log ratio of Mean Squared Error (MSE) between the Views model and the ShapeFinder (SF) model at the Prio-Grid level and Mean and 95% of the Confidence level. The units in the two subfigures (bottom left of the figure) are the metrics calculated for the 6-month prediction period and Prio-Grid. Along with this, the map (center of the figure) displays cumulative MSE per cell over four analysis periods. Red indicates better performance by the SF model, blue by the Views model, and gray denotes similar predictions.

Prediction at the Pattern level

In complement, this section focuses on the macro-level results by analyzing the model's performance at the pattern level, as discussed in Figure 12. This approach is more relevant for early warning applications, as it considers entire zones rather than isolated grid cells. By capturing broader conflict dynamics, this level of analysis provides a more

meaningful foundation for decision-making and risk assessment.

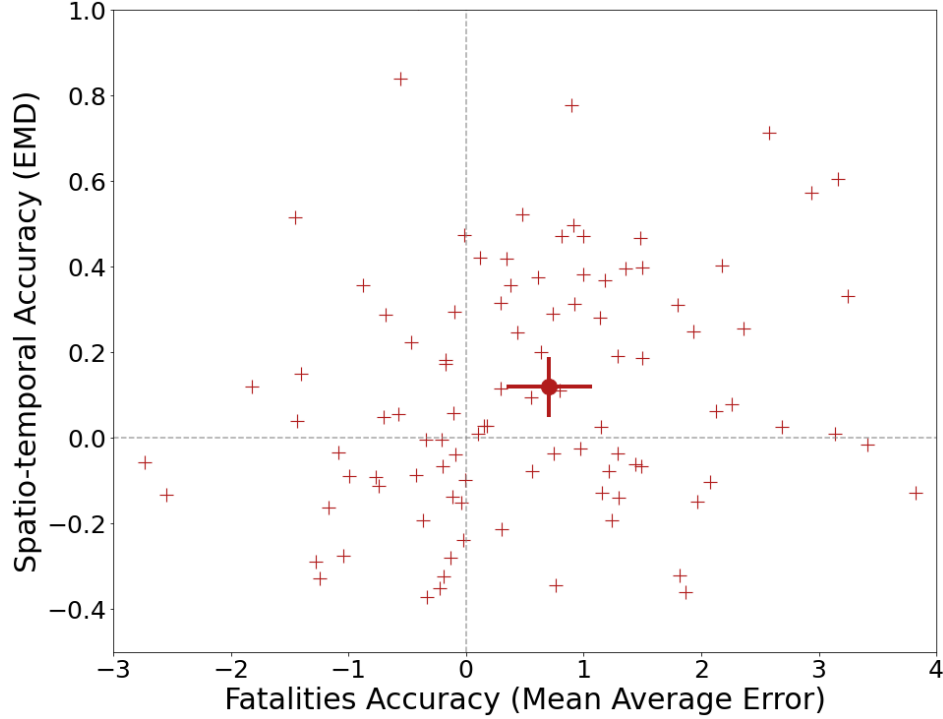


Figure 20: Scatter plot of the log ratio of Raw Error (in x) and EMD (in y) at the pattern level of Views model over Shapefinder model. If the ratio is positive, the Shapefinder model shows better performance. The red point is the mean value with the 95% Confidence Interval.

Figure 20 presents the log ratio of Mean Average Error (MAE) (x-axis) and EMD (y-axis) at the pattern level, comparing the Views model to the ShapeFinder (SF) model. A positive ratio indicates that the SF model outperforms the Views model. The results demonstrate that the SF model outperforms in both MAE (average log ratio = 0.708) and EMD (average log ratio = 0.120), achieving the most accurate predictions for total fatalities and dynamic patterns. In total, 111 patterns are included across the four periods. Appendix A.2 shows examples of projections and forecast maps to further illustrate these findings.

Performance analysis

To assess where and why the SF model outperforms, we examine three characteristics of the 111 identified patterns across the four studied periods. First, we analyze how a pattern evolves in terms of the intensity of fatalities using the ratio of increase R_{inc} . Second, we evaluate two performance measures: the prediction performance of Views (using the Mean Absolute Percentage Error, MAPE) and the relative difference in performance between Views and the SF model (log ratio of error). These metrics allow us to identify regions where Views performs poorly and to determine whether the SF model minimizes these errors, showing clearly areas where the SF model contributes improvements.

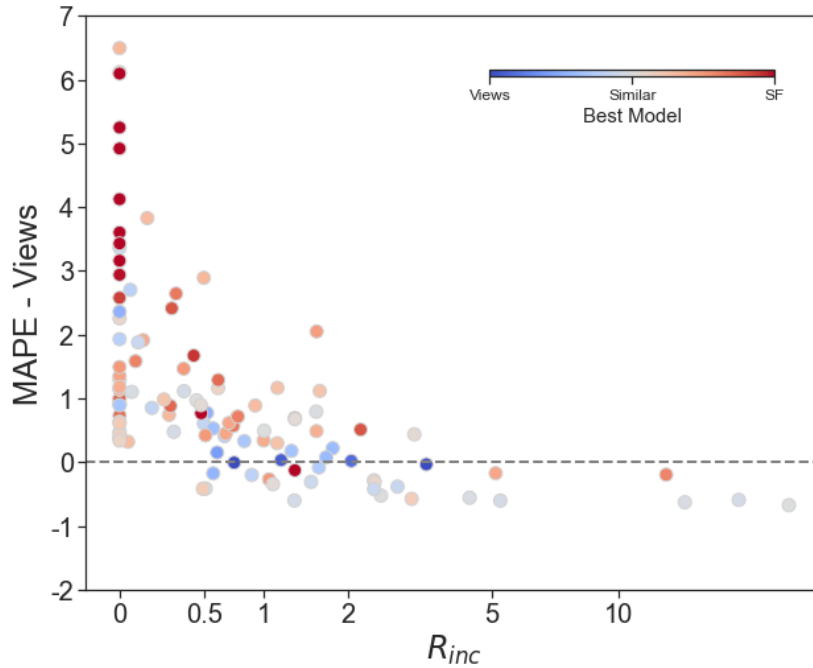


Figure 21: Scatter plot of the ratio of increase R_{inc} (in x) and Mean Average Percentage Error (MAPE) of Views (in y) at the pattern level, with a log-modulus transformation. The color of the point represents the difference in performance between the two models. The darker red, the SF performs better, and the darker blue, the Views model.

Figure 21 illustrates the relationship between the pattern's increase ratio R_{inc} (x-

axis), the Mean Average Percentage Error (MAPE) of the Views model (y-axis) with a log-modulus transformation, and the error ratio of both models (indicated by color). In the figure, darker red areas show where the SF model performs better, while darker blue areas indicate better performance by the Views model. As expected, patterns where the Views model has strong MAPE performances (around the 0-line on the y-axis) also show a negative error ratio (in blue), indicating better performance for Views. Beyond this, we can see three main groups of patterns. First, there is a group of dark red points with R_{inc} value from 0 to 0.5, where the Views model overpredicts significantly (with MAPE values above 2). The second group includes light red points with R_{inc} value up to 2, where Views slightly overpredicts. Finally, there is a group of very pale red and blue points with R_{inc} value above 3, where both models underpredict (negative MAPE values). These are the hardest patterns to forecast, as they involve sudden, large increases in fatalities, and both models underpredict similarly here. Overall, it appears that the SF model's improvement comes mainly in cases where the Views model overpredicts.

An important question is then whether the SF model achieves its strong MSE and Absolute Error scores simply by making conservative predictions. Metrics like these can favor models that produce conservative prediction values, especially in conflict forecasting at a fine scale (monthly, grid-level), where most values are zero. However, Figures A.18 and A.19 show the MAPE and Maximum Error, respectively. These reveal that the SF model strikes a balance between under- and over-predictions. Interestingly, its maximum forecast values are even higher in mean and median than those of the Views model. This suggests that the SF model provides more accurate and higher-scoring predictions without relying on overly conservative estimates.

More detailed plots of the SF performances are displayed in Figure 22. It shows the relationship between pattern characteristics and prediction error at the pattern

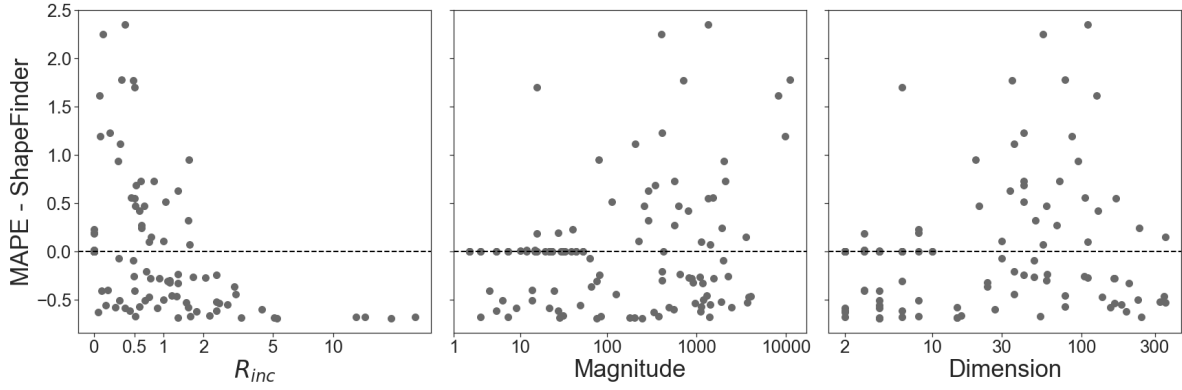


Figure 22: Scatter plots showing the relationship between pattern characteristics and prediction error at the pattern level using the ShapeFinder model. The y-axis in all three plots represents the Mean Absolute Percentage Error (MAPE), with a log-modulus transformation, where positive values indicate overpredictions and negative values indicate underpredictions. The x-axes represent: (left) the ratio of increase R_{inc} , (center) the total number of fatalities in the input pattern (referred to as Magnitude), and (right) the number of active cells in the pattern (referred to as Dimension). The dotted horizontal line at 0 on the y-axis indicates a perfect prediction.

level using the ShapeFinder model. The y-axis in all three plots represents the Mean Absolute Percentage Error (MAPE), with a log-modulus transformation, where positive values indicate overpredictions and negative values indicate underpredictions. The x-axes represent: (left) the ratio of increase R_{inc} , (center) the total number of fatalities in the input pattern (referred to as Magnitude), and (right) the number of active cells in the pattern (referred to as Dimension). The dotted horizontal line at 0 on the y-axis indicates a perfect prediction. Similarly to the Views, all of the overpredictions are situated when R_{inc} is inferior than 2. However, they also correspond to high magnitude input patterns and medium dimension size patterns (from 30 to 100 active cells). The underpredictions are more uniformly distributed and less important in values than overpredictions.

Discussion

In this paper, we introduce a novel method for identifying conflict dynamics at a fine-grained geographic level. It shows strong forecasting performances, overperforming the Views model, one of the strongest benchmarks, at both micro and macro levels. However, this method is based on several assumptions. The primary assumption is that patterns of different magnitudes in geography and intensity are comparable. While studies have shown that this approach works at larger scales, such as country-month analyses (Schincariol, Frank, and Chadeaux, 2024), its effectiveness at smaller scales, as explored in this study, requires further investigation. Future work could adopt a more theory-driven approach, using the model to examine patterns associated with specific variables such as conflict type or commonly used covariates in conflict modeling. Moreover, while this study is limited to Africa and the Middle East, future research could explore cross-regional pattern similarity and extend the model to a broader geographic scope.

Additionally, even if the model is highly flexible, it assumes that the diffusion of conflict to neighboring areas is equally probable. In reality, this is not the case, as factors such as logistics (e.g., roads connecting cells) and human communication networks, which might link distant neighbors more closely than geographically closer ones due to differences in religion, culture, etc., play a significant role. Addressing these factors would require a major overhaul of the model, including the incorporation of road networks and population data for each cell. This would add considerable complexity to the model and necessitate a large amount of data.

A limitation of our model is that it cannot predict an increase in fatalities if the input pattern contains only zeros. However, in 2022 and 2023, 99% of grid cells with only zeros over the previous 12 months continued to have zeros in the next six months.

For the rare cases with a sudden increase, predicting these spikes is challenging for all models, even those with additional data. This is clear in Figure 21, where the Views model underpredicts values when the increase ratio is above 3. To capture such sudden conflict outbreaks, a specialized model would be needed. In fact, for cells outside the patterns studied here, a model that forecasts only zeros would perform better than Views on key metrics.

A deeper analysis of our model’s performance shows that it might offer a more precise forecasted dynamic. It seems that the Views model captures the general area where fatalities might occur but lacks precision within that area, often leading to over-predictions. The EMD metric supports this, indicating that our model better pinpoints specific locations. This difference is also clear in Figure A.17, where aggregated forecasts show that our model provides a more refined pattern, while Views predicts broad, vague clouds of activity. But, as noted in the Benchmark section, the Views model is optimized for a 36-month forecast horizon. If it were optimized for a shorter 6-month horizon, its results might differ significantly.

Conclusion

Our model demonstrates strong forecasting performance with the added benefit of interpretability. Its main asset is the traceability of the source of each forecast. Contrary to highly complex machine learning models, our approach allows us to trace the origins of each forecast, showing when similar patterns occurred in the past and the likelihood of different future scenarios. This makes the model highly interpretable, an important advantage for convincing decision-makers of its reliability. Moreover, it can run quickly (about one hour per period) compared to the much longer processing times of deep learning methods, and only needs fatality input data to run.

Acknowledgements

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

CRedit authorship contribution statement

Thomas Schincariol: Conceptualization; Data curation; Formal analysis; Methodology; Software; Validation; Visualization; Writing - original draft; and Writing - review & editing

Declaration of competing interest

The authors declare that they have no competing interests.

Declaration of Generative AI in the writing process

I used ChatGPT to assist with grammar and spelling corrections, translations, and synonym suggestions during the preparation of this work. I reviewed and edited all content afterward and take full responsibility for the final manuscript.

References

- Aas Rustad, Siri Camilla et al. (2011). “All conflict is local: Modeling sub-national variation in civil conflict risk”. In: *Conflict Management and Peace Science* 28.1, pp. 15–40.
- Andoni, Alexandr, Piotr Indyk, and Robert Krauthgamer (2008). “Earth mover distance over high-dimensional spaces.” In: *SODA*. Vol. 8, pp. 343–352.
- Bazzi, Samuel et al. (2022). “The promise and pitfalls of conflict prediction: evidence from Colombia and Indonesia”. In: *Review of Economics and Statistics* 104.4, pp. 764–779.
- Black, Nathan (2013). “When have violent civil conflicts spread? Introducing a dataset of substate conflict contagion”. In: *Journal of Peace Research* 50.6, pp. 751–759.
- Buhaug, Halvard and Kristian Skrede Gleditsch (2008). “Contagion or confusion? Why conflicts cluster in space”. In: *International studies quarterly* 52.2, pp. 215–233.
- Cederman, Lars-Erik and Nils B Weidmann (2017). “Predicting armed conflict: Time to adjust our expectations?” In: *Science* 355.6324, pp. 474–476.
- Cook, Scott J and Nils B Weidmann (2022). “Race to the bottom: Spatial aggregation and event data”. In: *International Interactions* 48.3, pp. 471–491.
- Davies, Shawn, Therese Pettersson, and Magnus Öberg (2022). “Organized violence 1989–2021 and drone warfare”. In: *Journal of Peace Research* 59.4, pp. 593–610.
- Dorussen, Han, Erik A Gartzke, and Oliver Westerwinter (2016). “Networked international politics: Complex interdependence and the diffusion of conflict and peace”. In: *Journal of peace research* 53.3, pp. 283–291.
- Forsberg, Erika (2014). “Diffusion in the study of civil wars: A cautionary tale”. In: *International Studies Review* 16.2, pp. 188–198.

- Fritz, Cornelius et al. (2022). “The role of governmental weapons procurements in forecasting monthly fatalities in intrastate conflicts: A semiparametric hierarchical hurdle model”. In: *International Interactions* 48.4, pp. 778–799.
- Hegre, Håvard et al. (2019). “ViEWS: A political violence early-warning system”. In: *Journal of peace research* 56.2, pp. 155–174.
- Hegre, Håvard et al. (2022). “Forecasting fatalities”. Violence Early-Warning System (ViEWS). Available at: <http://uu.diva-portal.org/smash/get/diva2:1667048/FULLTEXT01.pdf> (accessed 11 February 2025).
- Iqbal, Zaryab and Harvey Starr (2008). “Bad neighbors: Failed states and their consequences”. In: *Conflict Management and Peace Science* 25.4, pp. 315–331.
- Kathman, Jacob D (2010). “Civil war contagion and neighboring interventions”. In: *International Studies Quarterly* 54.4, pp. 989–1012.
- Kelling, Claire and YiJyun Lin (2020). “Analysis of Conflict Diffusion Over Continuous Space”. In: *Computational Conflict Research*, pp. 201–223.
- Kibris, Arzu (2021). “The geo-temporal evolution of violence in civil conflicts: A micro analysis of conflict diffusion on a new event data set”. In: *Journal of Peace Research* 58.5, pp. 885–899.
- Li, Zhonghang et al. (2024). “Urbangpt: Spatio-temporal large language models”. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5351–5362.
- Lindholm, Andreas et al. (2022). “Predicting political violence using a state-space model”. In: *International Interactions* 48.4, pp. 759–777.
- Maase, Simon P von der (2025). “Next-Generation Conflict Forecasting: Unleashing Predictive Patterns through Spatiotemporal Learning”. In: *arXiv preprint arXiv:2506.14817*.

- Metternich, Nils W, Shahryar Minhas, and Michael D Ward (2017). “Firewall? Or wall on fire? A unified framework of conflict contagion and the role of ethnic exclusion”. In: *Journal of Conflict Resolution* 61.6, pp. 1151–1173.
- Racek, Daniel, Paul Thurner, and Goeran Kauermann (2024). *Integrating Spatio-temporal Diffusion into Statistical Forecasting Models of Armed Conflict via Non-parametric Smoothing*. Tech. rep. Center for Open Science.
- Racek, Daniel, Paul W Thurner, et al. (2024). “Conflict forecasting using remote sensing data: An application to the Syrian civil war”. In: *International Journal of Forecasting* 40.1, pp. 373–391.
- Radford, Benjamin J (2022). “High resolution conflict forecasting with spatial convolutions and long short-term memory”. In: *International Interactions* 48.4, pp. 739–758.
- Saideman, Stephen M (2012). “When conflict spreads: Arab spring and the limits of diffusion”. In: *International Interactions* 38.5, pp. 713–722.
- Schincariol, Thomas, Hannah Frank, and Thomas Chadeaux (2024). *Temporal Patterns in Conflict Prediction: An improved Shape-Based Approach*.
- (2025). “Accounting for variability in conflict dynamics: A pattern-based predictive model”. In: *Journal of Peace Research*, p. 00223433251330790.
- Schutte, Sebastian and Nils B Weidmann (2011). “Diffusion patterns of violence in civil wars”. In: *Political Geography* 30.3, pp. 143–152.
- Sundberg, Ralph and Erik Melander (2013). “Introducing the UCDP georeferenced event dataset”. In: *Journal of Peace Research* 50.4, pp. 523–532.
- Tollefsen, Andreas Forø, Håvard Strand, and Halvard Buhaug (2012). “PRIO-GRID: A unified spatial data structure”. In: *Journal of Peace Research* 49.2, pp. 363–374.
- Vesco, Paola et al. (2022). “United they stand: Findings from an escalation prediction competition”. In: *International Interactions* 48.4, pp. 860–896.

- Walterskirchen, Julian et al. (2024). *Taking time seriously: Predicting conflict fatalities using temporal fusion transformers*.
- Zhang, Zheng et al. (2024). “Large language models for spatial trajectory patterns mining”. In: *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geospatial Anomaly Detection*, pp. 52–55.

4 Forecasting Migration Flows to Europe: A Shape-Based Modeling Approach with Covariate Analysis

Forecasting Migration Flows to Europe: A Shape-Based Modeling Approach with Covariate Analysis

Thomas Schincariol

Abstract

Migration is a significant challenge in the current increasingly violent world. Given the projected effects of climate change in the next decades, the number of individuals that would need to migrate in the next decades is likely to increase. Accurate forecasting is essential to prepare for those migration flows. However, current models face serious limitations due to poor data quality and the complex, multidimensional mechanisms underlying migration decisions. Classical statistical models fail to capture the complexity of migration mechanisms, while newer machine learning models require large volumes of data and lack interpretability, making them less appropriate for decision-making applications. In this article, we use a flexible model, ShapeFinder, based on shape similarity. The model's predictions are derived from analogous cases in historical datasets, offering both interpretability and strong forecasting performance compared to benchmark models. The secondary key finding is the limited impact of push-pull factor covariates in forecasting models. Although these factors were included in this study, they did not significantly improve performance. Three inclusion methods are tested at different stages of the model process, but none lowers the forecasting error. The complexity of human decision-making and the inherent resilience dynamics limit the effectiveness of incorporating covariates as used in traditional forecasting models. Our results demonstrate that a flexible model like ShapeFinder is essential to address the subtleties of migration flows and could serve as a valuable tool in future early warning systems.

Migration studies are emerging as one of humanity's greatest challenges. Over the past 12 years, the number of displaced people has consistently increased, reaching nearly 120 million, equivalent to 1 in every 69 people on Earth (Refugees, 2023). Alarmingly, this phenomenon is accelerating. According to the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6), the ongoing global warming will drive a significant rise in extreme events such as droughts and floods while simultaneously reducing access to food and the habitability of large areas of the planet.

In collaboration with the International Organization for Migration, the IPCC has developed several scenarios to estimate the impact of climate change on migration. In the worst-case scenario, characterized by continued economic practices and reliance on fossil fuels, the global temperature could rise by 4.0°C by the end of the century. This would lead to over 200 million people being displaced due to climate change. This projection surpasses the alarming findings of the World Bank's 2018 Groundswell report, which estimated that by 2050, nearly 150 million people could be forced to migrate due to the slow-onset impacts of climate change.

Whether optimistic or pessimistic, all scenarios highlight the immense challenge humanity faces. This emphasizes the urgent need to forecast migration flows to better prepare for the millions of people who will migrate or be displaced. Achieving this requires identifying the covariates that push or pull individuals toward migration. While scholars have extensively studied these factors, forecasting remains a complex problem due to the intricate nature of human migration. Many push factors are interconnected. For instance, food shortages caused by climate change can trigger political instability, which in turn may lead to economic instability. At the center of this web are individuals who must decide, though often with little real choice, when to leave their homes. Given this complexity, constructing a comprehensive causal framework is nearly impossible. While researchers have identified a range of push and pull factors that influence

migration, the precise formula or mechanism driving these flows remains elusive.

Most studies on migration flows focus on large-scale analyses, typically at the country-year level. The few studies conducted at finer spatiotemporal resolutions are often case-specific and lack generalizability. This limitation can be attributed to the high cost of collecting data continuously across extensive geographical areas. Migration flows remain exceedingly difficult to track, and this lack of consistent, high-resolution data has led to a scarcity of studies on migration flow forecasting. Existing research in this area is limited and typically case-specific.

This paper has a two-fold objective. First, it establishes a benchmark for migration flow forecasting using a novel method based on a shape similarity model, originally developed in (Schincariol and Chadeaux, 2025) for application in South Sudan. The model has been adapted to a broader context, focusing on migration flows into Europe from 2017 to the present. This dataset is one of the few continuous sources of monthly migration flow data that provides both the origin and destination of migrants. Our findings demonstrate that this flexible method outperforms benchmark models in forecasting performance, while maintaining low computational requirements and offering a high level of interpretability. By matching similar historical cases, the model provides insights into the factors driving its predictions, making it easier for users to understand the underlying dynamics.

Second, the paper assesses the impact of migration covariates, or push-pull factors, on forecasting performance. This evaluation is conducted using three distinct inclusion methods within the shape-based model, as well as the machine learning benchmark model. This comprehensive approach allows for a thorough investigation of the potential benefits of incorporating covariates into forecasting. It is important to note that the analysis is limited to forecasting performance, rather than causal explanatory modeling. Since autoregressive terms already capture some underlying information, the study

focuses on determining whether covariates offer additional predictive value, rather than identifying the causal mechanisms of migration.

These findings are significant in multiple ways. First, they validate the effectiveness of shape-based models for forecasting migration flows, when data quality is limited. Second, they have practical implications for early warning systems. The model’s reliance exclusively on autoregressive data, combined with its computational efficiency, makes it a practical and straightforward tool for deployment. Policymakers and NGOs can use this model to better anticipate and prepare for migration challenges in the coming years.

Push-Pull Factors

In migration studies, the push-pull theory is well established. This theory is based on push factors, which make populations want to leave a zone, and pull factors that attract populations to a zone (Everett and Lee, 1966).

The first category of push factors most commonly studied is socio-economic. In (Khan et al., 2023), the authors examine migration triggers in India using survey data, concluding that push factors play a critical role in migration decisions. The primary factors identified are economic, such as “lack of jobs” and “low wages”, and social, including “family debt” and “social insecurity.” Similarly, (Khalid and Urbański, 2021) points out the significant influence of unemployment and poverty on migration decisions based on studies conducted in Thailand and Poland.

Another major push factor is insecurity and conflict. In (Crippa et al., 2024), the authors demonstrate a strong link between conflict and migration, particularly in low-income countries. However, conflict is often intertwined with climate change and food accessibility, where one can trigger the other, both contributing to migration flows. For

example, (Abel et al., 2019) confirms a causal relationship between climate change, conflict, and migration, noting that climate change increases asylum-seeking behavior through its impact on conflict, particularly in recent years. Similarly, (Cottier and Salehyan, 2021) investigates the connection between climate variability and irregular migration to Europe, concluding that drought and rainfall significantly influence irregular migration levels, especially in agrarian countries.

Finally, (Castelli, 2018) explores the factors driving migration, categorizing them into ‘macro’, ‘meso’, and ‘micro’ levels. The macro factors align with those discussed in the previous paragraphs. However, at the meso and micro levels, additional influences come into play, such as communication technology, land grabbing, education, religion, and health systems. These factors can significantly affect the decision to migrate, although they are often challenging to quantify and operationalize within migration models.

Pull factors are closely linked to push factors, as migrants often seek improved conditions to address what is lacking in their place of origin. The primary pull factor is economic opportunity. For instance, (Urbański, 2022) compares push and pull factors in the contexts of Poland and Romania, concluding that pull factors exert a greater influence on migration decisions. Economic considerations, such as higher wages, improved living standards, personal development, and job opportunities, are identified as the dominant pull factors. Similarly, (Ferwerda and Gest, 2021) finds that employment availability is a key pull factor, while liberal democratic rights also play an important role in attracting migrants. A similar conclusion was reached using the civil war refugees example, who mostly seek a democratic and wealthy host country (Turkoglu and Chadeaux, 2019).

Another important pull factor is the cultural aspect, particularly the presence of a diaspora in the destination country. (Collier and Hoeffler, 2018) analyzes the impact of

diaspora and cultural ties on migration using a gravity model, concluding that diasporas in culturally distant countries play a crucial role in reducing the costs of migration. Likewise, (Prieto-Curiel et al., 2024) develops a migration model based on existing diaspora sizes in destination countries. Their model effectively reproduces observed stable migration flows and distributions at fine-grained levels, such as Austrian postal codes and U.S. metropolitan areas. Also, (Johnson and Bellofatto, 2012) investigates the role of religious diasporas, demonstrating that cultural or nationality-based diasporas are not the only significant pull factors; religious ties also play a pivotal role.

Lastly, social factors, such as education opportunities, remain significant pull factors. For instance, (Dago and Barussaud, 2021) explores the motivations of Ivorian students pursuing higher education in France and Switzerland. The study highlights that uncertainties within the Ivorian education system drive many students to view migration as their only viable path to success, emphasizing the critical role of education opportunities in migration decisions.

The challenge of migration flow prediction

Scholars investigate migration mechanisms by examining the factors that push or pull individuals to different destinations. However, a few studies alternatively focus on forecasting migration flows. Given the complexity of migration decision-making, accurate predictions remain a significant challenge. As noted in (Bijak et al., 2019), “Migration is a very complex and multi-dimensional process [...]. Thus, its forecasting is extremely difficult.” Using migration flow data from the UK, this study evaluates classical time series models in the context of migration forecasting, concluding that no single model is universally best suited for different flows. Similarly, (Welch and Raftery, 2022) develops a probabilistic forecast model using a Bayesian hierarchical approach, though

this model operates at the yearly level, limiting its granularity. From a methodological perspective, machine learning has also been explored. For instance, (Carammia, Iacus, and Wilkin, 2022) introduces a sophisticated early warning system employing a machine learning framework for asylum seekers. This study stands out for its use of a finer temporal resolution, offering weekly forecasts with a horizon of just four weeks.

Yet, migration flow forecasting remains particularly challenging in terms of prediction accuracy compared to other fields. This difficulty can be attributed to the poor quality of available data and the rigidity of classical forecasting models, which are not well-suited to capturing the complexities of migration dynamics. Migration decisions are naturally multidimensional and involve highly complex mechanisms. Each dimension may operate at different speeds or hold varying degrees of influence on individual decision-making, depending on the context. Furthermore, different conditions can produce similar migration outcomes. For instance, both natural disasters and economic crises may lead to comparable migration magnitudes. Historical examples include the Great Famine in Ireland and the Syrian Civil War, which, despite vastly different causes, resulted in similar proportions of population displacement, approximately 20–25% of the population migrating abroad. This intricate relationship between inputs and outputs makes it nearly impossible to establish a universally effective forecasting model for migration flows.

This complexity brings into question the utility of covariates in forecasting models. While some new machine learning designs may introduce a degree of flexibility, most models operate on the assumption that relationships established in historical or training data between push-pull factors and subsequent migration flows can reliably predict future patterns. However, when applied to migration decision-making, this assumption becomes theoretically questionable. Current models and data lack the flexibility required to fully capture the subtleties of human decision-making. Relying solely on

push-pull factors identified in migration literature may not be an effective strategy for forecasting models.

Push factors at t				Migration flow		
Cases	Insecurity	Economy	Natural Disaster	t+1	t+2	t+3
A	↗	↘	–	↗	↗	↗
B	↗	↘	–	–	–	–
C	–	–	↗	↗	↗	↗
D	–	–	↗	–	–	↗

Table 2: Push factors and migration flow over time

To illustrate, consider the example displayed in Table 2, which presents four forecasting cases (A to D). On the left, we see the model’s input, or covariates, representing push factors. On the right, the corresponding migration flow values over the next three months. In cases A and B, the input conditions are similar, with rising insecurity and an economic slowdown. However, the resulting migration flows differ significantly. This discrepancy aligns with the theory of livelihood resilience in migration decision-making (Mallick, 2019), which suggests that decisions are influenced not only by push-pull factors but also by the Socio-Ecological System (SES) of individuals. Modeling the SES remains highly complex, as it is deeply context-dependent. For example, (Mallick, 2019) examined rural coastal communities in Bangladesh and found that migration decisions varied based on the type of cultivation practiced—salt-shrimp farming versus rain-fed agriculture. This stresses the immense variability in the SES and the impractical amount of data needed to fully model these mechanisms, exceeding the explanatory power of the classical push-pull set of variables. Similarly, cases C and D have identical input conditions, yet the migration outcomes are delayed by two months in case D. This delay could be attributed to the climatic structure influencing agriculture and, by extension, the SES of the populations in cases C and D.

These examples highlight the critical need for adaptive models, like the one introduced in this paper, that can accommodate such variations in migration decision-making dynamics. Migration decisions are shaped by a tremendous number of factors, including environmental conditions coupled with the cognitive processes of individuals, which together form a dynamic and highly interconnected system. Individuals influence one another, creating feedback loops within social networks. For example, knowing that a neighbor has left may increase an individual’s likelihood of migrating. The complexity and context dependence of these micro-level dynamics make them extremely difficult, if not impossible, to model directly. The collection of push and pull factors is insufficient to capture the overall complexity. General forecasting models that learn from the past and other contexts to generalize the links between variables would not accommodate this. While such models could work for some case-specific situations where the key variables are identified, they could not be extended to a huge variety of contexts.

The analytical focus should instead be placed on the macro-level patterns, corresponding to aggregate outcomes of countless micro-level interactions. To effectively compare and model these macro-level dynamics, forecasting models must be flexible enough to capture similar patterns across diverse contexts. In this study, the primary model is autoregressive, relying on migration flow patterns to learn from the macro-level dynamics to predict. These dynamics are compared with a high flexibility degree, with the Dynamical Time Warping, a distance metric that allows comparison between sequences with different lengths, temporal lags, or speeds. This assures the model’s capacity to identify and generalize similar macro-level patterns across different contexts.

Data

The objective of this study is to forecast migration flows, with a focus on the first monthly arrivals for each origin–entry point dyad in Europe. Migration flow time series often raise challenges due to their short length, discontinuities, or varying frequencies over time. Collecting such data is costly, requiring significant human resources and long-term commitment to maintain consistent data collection efforts. Despite this, the International Organization for Migration (IOM) has recorded data on first arrivals in Europe from January 2017 to June 2024 (Migration (IOM), 2024). This dataset includes monthly records detailing the origins of migrants and their entry points into Europe. Five countries are represented as entry gates to Europe: Spain, Italy, Greece, Cyprus, and Bulgaria. 111 countries are listed as departure points, resulting in 244 routes of migration included in the study. The routes with more than 50% of unavailable values are excluded. Figure 23 displays the migration axes, with departure countries marked in blue and arrival points in red. Thicker lines indicate higher numbers of migrants.

For each migration route, both push factors from the departure country and pull factors from the entry point in Europe country are considered. In practice, push factors appear to play a more significant role, since migrants’ final destinations are not known, European entry points frequently serve as transit points rather than final destinations. The push-pull covariates are summarized in the Table 3.

Traditional pull factors, such as education or standard of living, which typically motivate long-term migration, were excluded from the analysis. Instead, the study focuses on two pull factors that better address the short-term needs of migrants: employment opportunities and potential family links. These factors reflect the immediate support migrants might receive, such as help from family members or acquaintances in the entry-point country, and the ability to quickly find a job to finance their journey to a

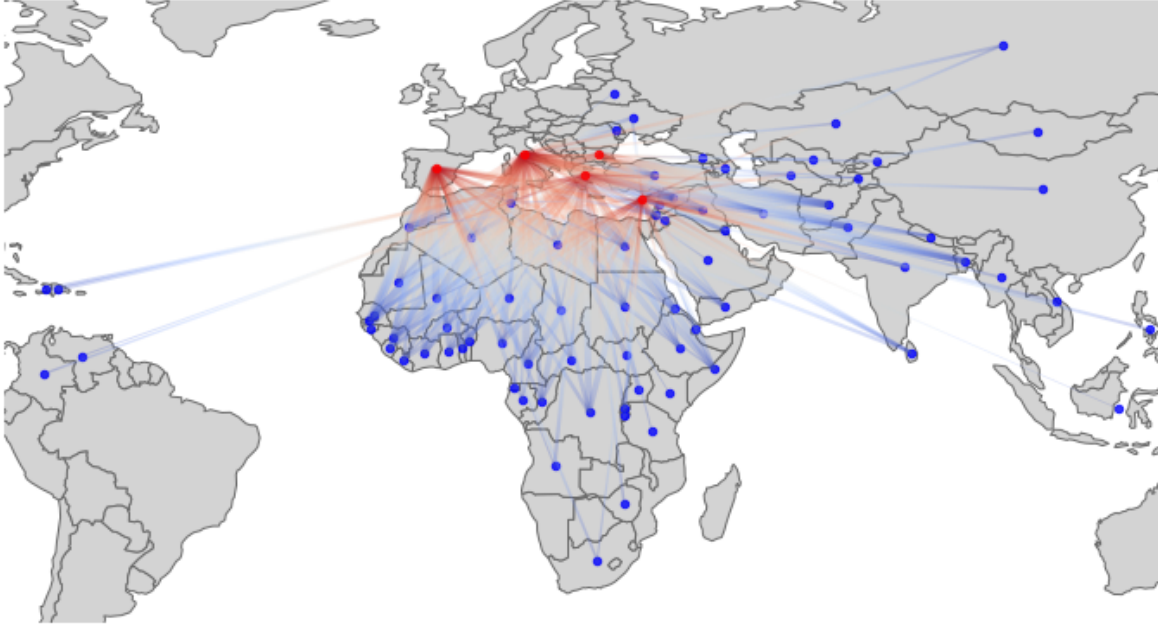


Figure 23: Map of the axes of migration included in the study. The departure is represented in blue and the arrival in red. The width of the line represents the log intensity of flows.

final destination in Europe. To estimate the probability of family or acquaintance links, data from the United Nations' International Migrant Stock dataset is used (Economic and Social Affairs, 2020). This dataset provides information on the number of migrants by origin country for 2020, the latest available year. Employment opportunities are measured using employment rates for foreign-born individuals, based on data from the International Migration Outlook 2023 report by the Organisation for Economic Co-

Covariate	Push/Pul	Category	Frequency	Source
Migrant stock	Pull	Social	Constant	UN
Foreign-born employment rate	Pull	Economic	Constant	OECD
Adult unemployment rate	Push	Economic	Yearly	ILO
IDP (conflict)	Push	Political	Yearly	IDMC
IDP (natural disasters)	Push	Environme	Yearly	IDMC

Table 3: Push-Pull Covariates Summary. The push covariates are extracted from the country of origin and pull from the entry point in Europe.

operation and Development (OECD) (OECD, 2023). The most recent value for both variables is assumed to be the same throughout the entire study period.

In opposition, the study includes classical push factors such as economic hardship, political instability, and environmental degradation. Economic conditions are quantified by the adult unemployment rate (ages 15+), sourced from the Sustainable Development Goals (SDG) Labour Market Indicators database provided by the International Labour Organization (ILO) (International Labour Organization (ILO), 2024). Data on political instability and environmental degradation are obtained from the Internal Displacement Monitoring Centre (IDMC), which tracks internal displacement caused by conflicts and natural disasters. These data offer insights into the direct impact of these push factors on migration. Indeed, the relationship between conflict or environmental factors and migration is complex and non-linear. For instance, conflict or drought events occurring at time t do not necessarily produce immediate effects on migration flows at time $t+1$, as these processes involve more intricate mechanisms. This complexity explains why raw data on conflict or environmental conditions are not directly used in the analysis. Instead, the model relies on the origin country's internal displacement caused by both push factors. The three variables have a yearly frequency from 2017 to 2023.³⁰

Methods

The model used in this study, originally developed for conflict forecasting (Schincariol, Frank, and Chadeaux, 2025), was created for the Violence & Impacts Early-Warning System (VIEWS) prediction competition held in 2023/2024 (Hegre et al., 2024). Known as 'ShapeFinder' (SF), the model employs an intuitive approach by identifying similar patterns in historical data to produce prediction values. Initially, the model was

³⁰The five covariates were also tested in transformed forms. The results of these tests are presented in Appendix A.3.

designed as a purely autoregressive framework. In this study, we adapt the model to include covariates and assess their impact on its performance. To evaluate these adaptations, we compare the performance of the original model and the modified version, also using a Random Forest model as a benchmark. The benchmark is tested in both a purely autoregressive form and with the inclusion of covariates.

The original ShapeFinder

The original model belongs to the family of shape-based models. Its core principle is to identify similar patterns within the historical dataset, which serves as the training set, and use these patterns to generate predictions. The model employs a rolling window over the entire training dataset to detect similar shapes in the input sequence. To measure similarity, the Dynamic Time Warping (DTW) metric (Müller, 2007) is applied between the input sequence and historical patterns. If the DTW metric falls below a defined threshold, the pattern is classified as similar. Once the rolling window is complete, the result is a set of identified similar patterns. The values following these patterns are then clustered to form scenarios with associated frequencies, which are interpreted as probabilities. The final forecast is computed as the mean of the values in the most probable scenarios. The clustering step is crucial, as it eliminates outlier scenarios that may otherwise lead to overpredictions.³¹

The final clustering step is performed using agglomerative hierarchical clustering, as shown in Figure 24. In the figure, the five series on the left represent the following values of similar patterns. The linkage is computed using the Ward method with Euclidean metric, and scenarios are formed by applying a threshold distance. The scenario with the highest number of observations is selected as the most probable. The average

³¹For a more detailed explanation of the model, refer to the Methods section in (Schincariol, Frank, and Chadefaux, 2025).

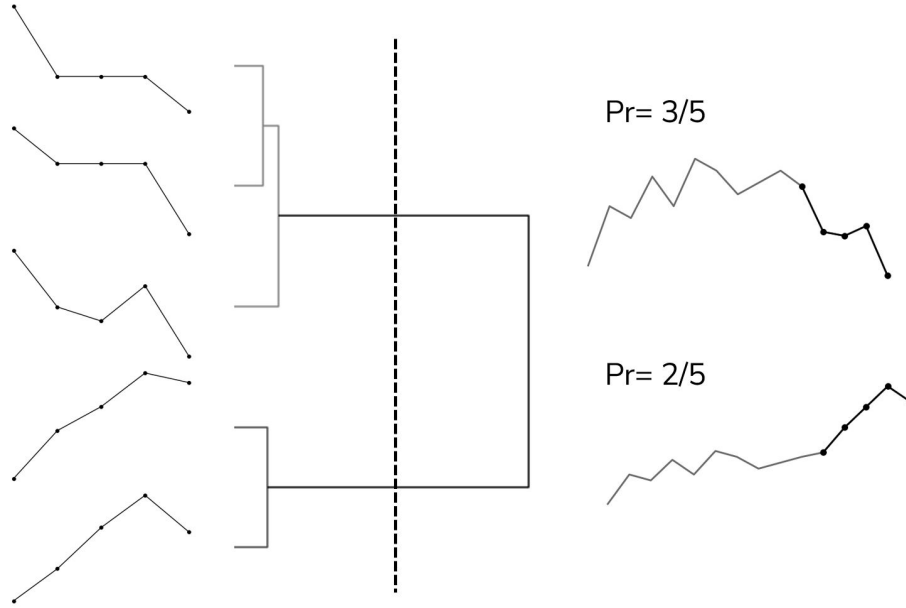


Figure 24: Agglomerative hierarchical clustering for the creation of the scenarios. The five series on the left represent the following values of similar patterns. The linkage is computed using the Ward method, and scenarios are formed by applying a threshold distance. The scenario with the highest number of observations is selected as the most probable. The average value of the cluster for each horizon is then used as the final forecast. Scenario 1 (with three observations) is selected over Scenario 2 (with two observations).

value of the chosen cluster for each horizon is then used as the final forecast. In the example shown, Scenario 1 (with three observations) is selected over Scenario 2 (with two observations). Finally, since all sequences are normalized from 0 to 1, a reverse normalization is applied to return to the input sequence's range.

The model relies on several key parameters. The first is the window length for input patterns. To determine this, we computed the Autocorrelation Function (ACF) plot, which measures the correlation of the time series with itself at different lags. The ACF plot is shown in Figure A.23, where it is evident that at lag $t-9$, the ACF value stabilizes at a low level. Based on this observation, the window length was set to 9. Other parameters consist of the Dynamic Time Warping (DTW) threshold for classifying patterns as similar or non-similar (noted min_{dist}), the minimum number of similar

patterns required when the threshold does not identify enough matches in historical data (noted min_{mat}), and the clustering distance used during scenario creation (noted clu_{dist}). These three parameters were optimized by minimizing the Mean Squared Error (MSE) on the test set, similarly to the benchmark.

Inclusion of Covariates

In this study, three methods of incorporating covariates are tested at different stages of the original model. Figure 25 shows the outline of the original model along with the three adaptations. The original model is represented in black, while the three modifications are displayed by dotted lines in three different colors.

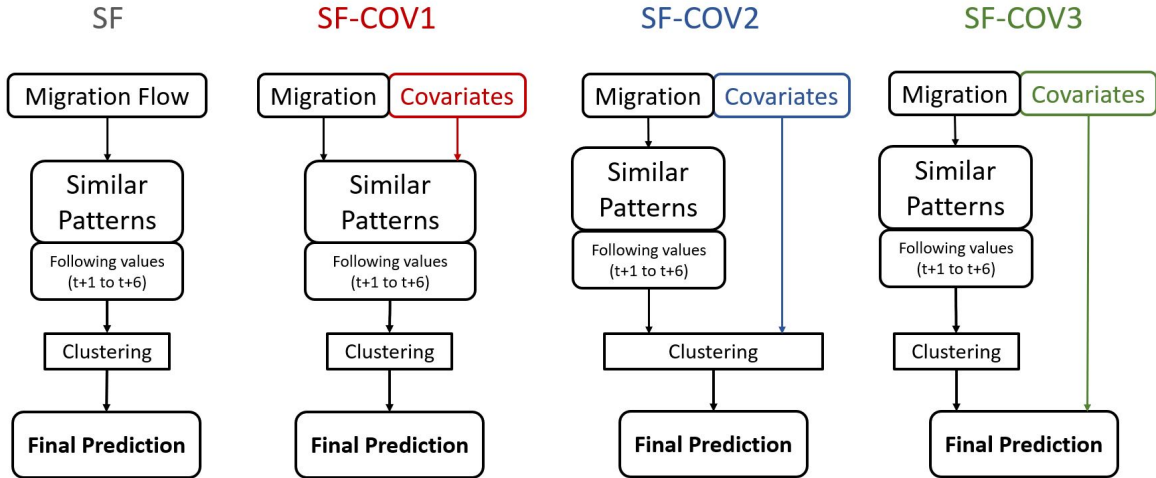


Figure 25: Layout of the original ShapeFinder model (SF) in black. The three modifications, including covariates introduced in this paper, are displayed in dotted lines.

An earlier adaptation of this model incorporated covariates by focusing on their shapes rather than static values (Schincariol and Chadeaux, 2025). In that approach, covariates were integrated into the process of identifying similar patterns in historical data. The model measured the sum of individual Dynamic Time Warping (DTW) values for migration data and covariates, such as food prices and conflict.

Inclusion prior to similar patterns search : SF-COV1

In the first modification model represented in red in the figure, called SF-COV1, we employ a similar method by including the differences in their static values in the overall similarity metric. The distance between two cases is defined by the following equation:

$$d_{i,j} = \frac{\alpha_0 * DTW(seq_i, seq_j) + \sum_{k=1}^n \alpha_k |cov_{k,i} - cov_{k,j}|}{\sum_{l=0}^n \alpha_l} \quad (6)$$

Here, *cov* represents the static value of the covariate³², and α is the weight assigned to each element in the final distance calculation. The weights enable testing the impact of including covariates in the model, and if some have more weight in the improvement of prediction.

Inclusion after similar patterns search

An alternative way of incorporating covariates was also tested. In this approach, covariates are introduced only during the construction of future scenarios, specifically when generating the prediction value from the most probable scenario. In the standard method, clustering is performed using the future values of similar cases to create scenarios with associated probabilities, based on the distribution of similar cases in each cluster. The prediction is then the average of the following values belonging to the most probable scenario. This approach, however, may overlook contextual differences across regions. For instance, a pattern observed in Country A may not produce the same outcome in Country B if job opportunities differ between the two countries at the time the historical pattern occurred.

To address this limitation, two options for integrating covariates at the final step were tested: (a) including covariates in Ward's linkage during clustering to create the

³²The covariates are z-normalized.

scenarios (SF-COV2, blue in Fig25) (b) applying a weighted average to the observations within the selected scenario after the scenarios are formed. (SF-COV3, green in Fig25)

Inclusion in the scenarios creation : SF-COV2

The first option, shown in blue in the figure and denoted SF-COV2, involves a direct modification of the clustering step. Covariates are added to Ward's linkage calculation, producing vectors of 11 elements (6 migration flow sequence points plus 5 covariates). All covariates are z-normalized to prevent them from dominating the sequence points. Once the scenarios are created, the final forecast is obtained in the same way as in the original model: the predicted value at each horizon is the average of the most probable scenario.³³

Inclusion only in the final prediction : SF-COV3

The second option, represented in green in the figure, called SF-COV3, operates after clustering. Instead of directly using the average value of the selected scenario, prediction values are calculated using the following equation:

$$p_h = \frac{\sum \frac{1}{d(covs_{input}, covs_s)} \cdot f_{v_{s,h}}}{\sum \frac{1}{d(covs_{input}, covs_s)}} \quad (7)$$

Here, p_h represents the predicted value for horizon h , $d(covs_{input}, covs_s)$ is the Euclidean distance between the vector of static covariates for the input sequence and those of a similar pattern s , and $f_{v_{s,h}}$ is the future value associated with the similar pattern s at horizon h . This approach adjusts predictions by weighting similar patterns based

³³A second alternative was tested in which clustering was based directly on the distance between covariates, $d(covs_{input}, covs_s)$, instead of on the sequences that follow similar patterns. This approach performed poorly, stressing the importance of clustering on the following sequences, which effectively removes outlier scenarios. The crucial role of this step is discussed in (Schincariol, Frank, and Chade-faux, 2025).

on their proximity in terms of static covariates, accounting for contextual differences between cases.

Benchmark

The ShapeFinder performances are evaluated by comparing our model with a machine learning model, the Random Forest (RF), an ensemble of decision trees. Random Forests are favored for their strong predictive performance, robustness to missing data, and ability to handle shorter time series effectively. They also provide interpretability by identifying the contribution of each feature to the model’s fit. For these reasons, we did not include a deep learning model, such as Long Short-Term Memory (LSTM) networks, which are less interpretable and generally less suitable for shorter time series. Given the relatively short length of each series (90 points), LSTMs may struggle to capture meaningful temporal dependencies without overfitting. In addition, a significant portion of the migration dataset contains missing values. Specifically, 34% of the data is missing across the 244 migration axes series. This excludes the use of classical statistical models such as AutoRegressive Integrated Moving Average (ARIMA), which assumes continuous, stationary, and complete time series data. The large share of missing data would require extensive imputation or interpolation, making ARIMA unsuitable for this study. Even when restricting the analysis to migration routes with the lowest rates of missing values, most routes still suffer from discontinuities, which is a common issue in migration flow data.

For a fair comparison, the RF model is designed with input and output shapes similar to those of ShapeFinder. Specifically, the input consists of lagged values from $t-1$ to $t-9$, and the forecast horizon is set to 6 months. The migration values are log-transformed to handle the skewness in the data. As with SF, two versions of the RF model are implemented: one including covariates and the other without. Both types of

models are trained and tested on the same train/test dataset to ensure comparability. The hyperparameters of the Random Forest model are optimized using a random grid search within the parameter ranges specified in Table A2. Similar to the ShapeFinder (SF) model, the configuration that achieves the lowest MSE on the test set is selected.

Evaluation

The training data spans from 2017 to December 2023, while the evaluation is conducted on predictions for the period from January 2024 to June 2024. Given the relatively short length of each time series (90 points), we maximized the percentage of data used for training to improve model performance. Only axes of migration without missing values in the last 15 timestamps were retained, as the input requires 9 values and the output 6, resulting in a dataset that contains 157 migration axes. This sufficient number of migration axes in the analysis ensures statistical significance. The benchmark includes six models :

1. The original ShapeFinder (SF)
2. The ShapeFinder with covariates inclusion in the similar cases search (SF-COV1)
3. (a) The ShapeFinder with covariates inclusion in the scenarios creation (SF-COV2)
- (b) The ShapeFinder with covariates inclusion only in the final prediction creation (SF-COV3)
4. Random Forest purely autoregressive (RF-AR)
5. Random Forest with covariates (RF-COV)

Similarly to the Random Forrest, the hyperparameters (the minimum number of matches, the minimum distance, and the clustering distance) are optimized on the test

set using a brute-force grid search. The test set is composed of the last six months of the dataset, and the remaining observations form the training set. The hyperparameters selected for the original model are then kept for all model adaptations that incorporate covariates. In the same way, the weights used in the SF-COV models are also optimized on the test set, to ensure a fair comparison.

The performance of the models is evaluated using the Mean Squared Error (MSE), a widely used metric in forecasting. To compare the two models, the log ratio of the MSEs is calculated for each migration axis. A positive log ratio indicates that the model in the denominator performs better, while a negative log ratio signifies superior performance by the model in the numerator.

Results

Similar pattern, similar outcome ?

The core assumption of a shape-based model is that similar shapes in the data tend to lead to similar outcomes. Although our model accounts for outliers through scenario creation and selection of the most probable scenario, it remains essential to verify that a given shape consistently produces similar outcomes. To test this assumption using the migration data, we calculate the average standard deviation of the following values of the similar cases for various similarity intervals, based on the input shapes in the test set.

The “following values” refer to the observations from $t + 1$ to $t + 6$, which are the same values used to form scenarios in the forecasting model (see Figure 24). Similarity is measured using DTW distance, in line with the specification of the forecasting model. For each shape, we identify similar shapes in the historical data that fall within the defined similarity range and compute the standard deviation of the following values for

forecast horizons 1 to 6. The mean standard deviation across all input shapes is then calculated. The results are presented in Figure 26.

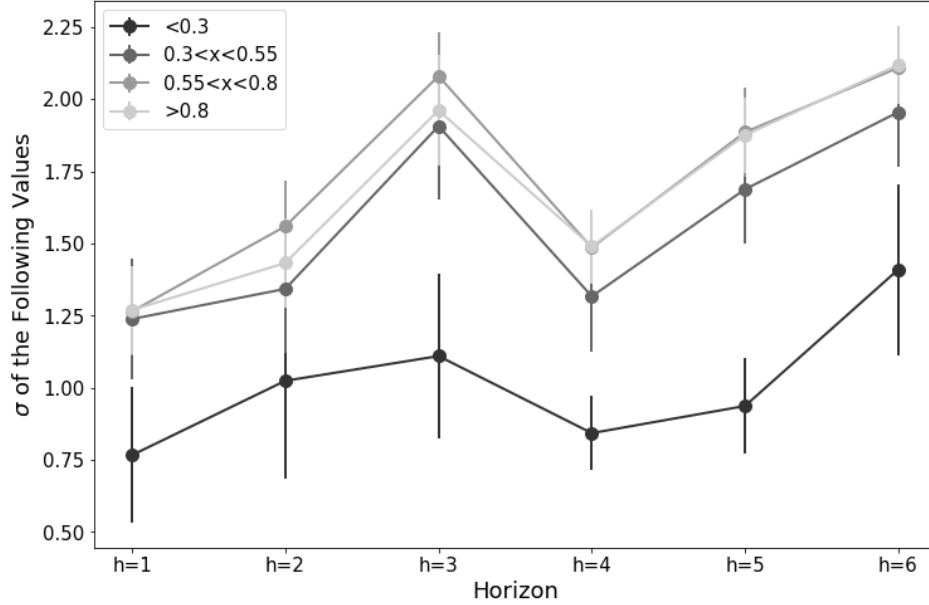


Figure 26: Standard deviation of the Following Values (in y-axis) by horizon of forecast (in x-axis) for the four intervals of similarity: $d < 0.3$, $0.3 < d < 0.55$, $0.55 < d < 0.8$ and $0.8 < d$, with d representing the DTW distance. The darker grey corresponds then to a higher degree of similarity.

The range with the most similar patterns results in a significantly smaller standard deviation in the following points, at the 95% confidence level (except for horizon 2), compared to all other intervals. Furthermore, the second-highest similarity range shows a smaller standard deviation, although not significantly different from the two intervals with the lowest similarity scores. These results provide statistical confirmation of the hypothesis that more similar shapes tend to produce more similar outcomes.

A parallel analysis is conducted using migration data combined with covariates to evaluate the additional information contributed by the covariates. This analysis examines whether similar migration shapes, when coupled with similar covariate conditions, lead to more consistent outcomes and provide more predictive power than migration data alone. The same methodology is applied to assess this, but the similarity intervals

now include the Dynamic Time Warping (DTW) distance of the migration shape combined with the difference in covariate values. Since the analysis involves two distances added together, the interval values were doubled. The results for the five covariates are presented in Figure 27.

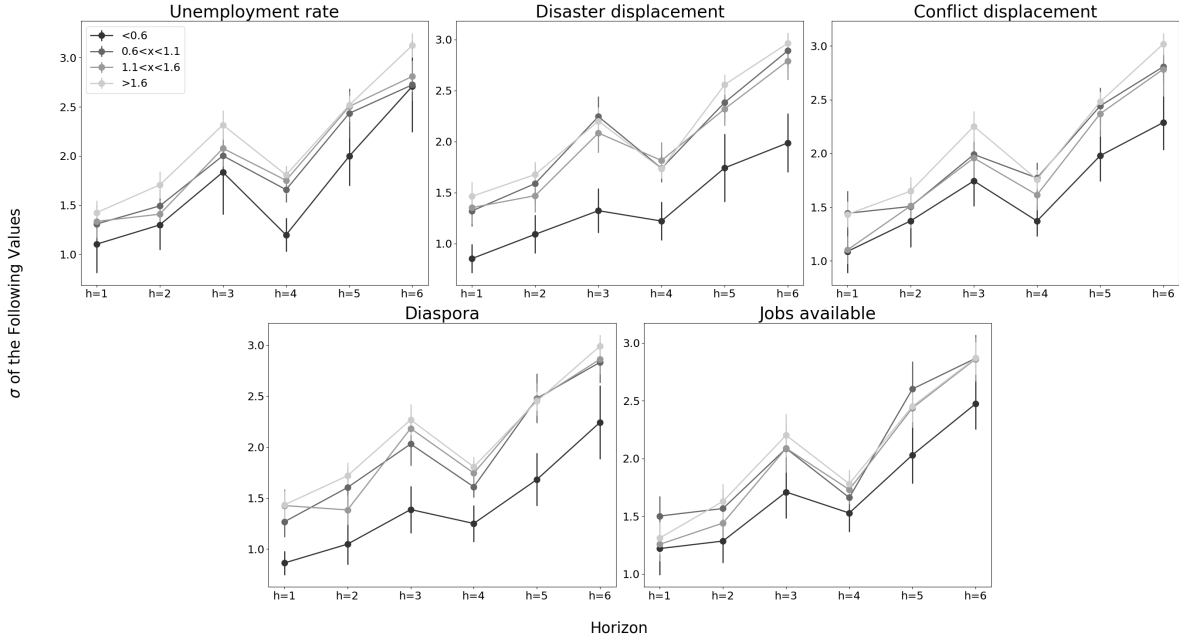


Figure 27: Standard deviation of the Following Values (in y-axis) by horizon of forecast (in x-axis) for the four intervals of similarity: $d < 0.6$, $0.6 < d < 1.1$, $1.1 < d < 1.6$ and $1.6 < d$, with d representing the DTW distance of migration sequences plus the difference of covariate value. The darker grey corresponds then to a higher degree of similarity.

Three out of the five covariates (Unemployment rate, Conflict displacement, and Job availability) produced inconclusive results, with none of the similarity intervals being significantly different from the others. The Diaspora covariate showed minimal impact, yielding results comparable to the purely autoregressive test. This suggests that while the Diaspora covariate does not bring information, it does not introduce additional noise. In contrast, the Disaster displacement covariate demonstrated a significant effect: the most similar interval consistently shows a smaller standard deviation than the other three intervals, with a 95% confidence level across all horizons. How-

ever, the second most similar interval showed the opposite effect, having the highest standard deviation. This could indicate that while the Disaster displacement covariate effectively distinguishes highly similar cases, it may introduce noise when the cases are only moderately similar.

Similar context, similar scenarios?

The second model, designed to evaluate the role of covariates, tests whether these covariates influence the distribution of scenarios once similar migration patterns have been identified. The underlying idea is that covariates capture the context in which migration occurs, and this context may shape the likelihood of one scenario over another. In other words, if two sequences share a similar shape but occur under different contextual conditions, their future evolution might differ. The original model assumes the opposite: similar shapes lead to similar outcomes regardless of context. Here, we examine whether there is statistical evidence of a dependency between context, represented by the covariates, and scenario outcomes.

To test this hypothesis, we apply Pearson’s chi-squared test to measure the statistical association between scenarios and covariates. The procedure follows these steps:

1. For each dyad, record the scenario assignments obtained from the clustering process for all following value sequences.
2. For each sequence, compute the difference in covariate values between the input sequence and the matched sequence, noted as $d(covs_{input}, covs_s)$.
3. Convert these continuous covariate differences into categorical variables by breaking them into bins defined by the 25th, 50th, and 75th percentiles. This ensures a balanced distribution of observations across categories.

4. Construct a contingency table that cross-tabulates (i) scenario assignments and (ii) the categorical covariate bins.
5. Apply Pearson’s chi-squared test to assess whether scenario assignments and covariate categories are statistically independent.

For example, if the model identifies 100 similar patterns and the clustering process results in four distinct clusters, a contingency table for a given covariate s could be like the one presented in Table 4. The operation is repeated for the five covariates for each dyad. However, the test requires assumptions. First, observations must be independent, and the test depends on a large-sample approximation. To satisfy these conditions, only dyads with at least 20 similar patterns are included in the analysis, and the clusters with fewer than 10 observations are removed. Additionally, the assumption that no more than 20% of the contingency table cells have expected counts less than 5 must be met. To address this, clusters exceeding this threshold were filtered out until the condition was satisfied.

$d(covs_{input}, covs_s)$ Bin	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Q1 (Low)	5	11	7	6
Q2	4	4	3	9
Q3	7	7	2	12
Q4 (High)	8	4	4	7

Table 4: Contingency table showing the relationship between the difference covariate $d(covs_{input}, covs_s)$ bins and cluster assignment.

Figure 28 shows the distribution of p-values from the chi-squared test for each variable in the covariate set. The darker bin represents significant results with p-values < 0.05 , indicating dependence between the covariate values and scenarios assigned. This suggests that the covariates bring predictive information relevant to the scenario creation. The grey bin corresponds to p-values < 0.1 , while the light grey bins represent

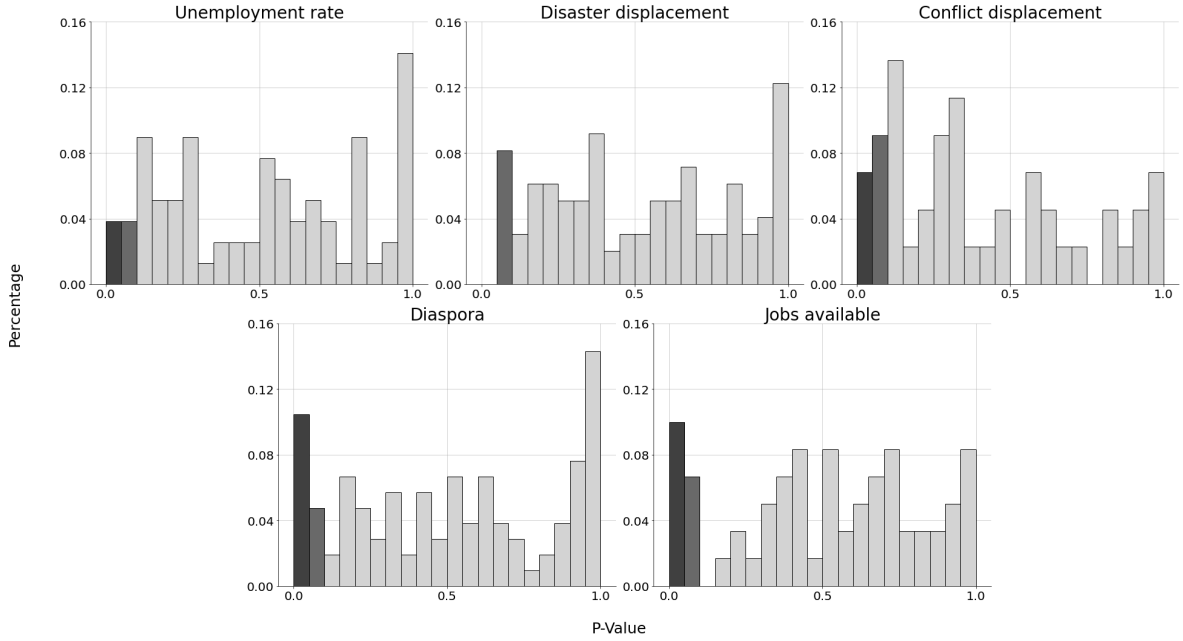


Figure 28: Histogram of the p-value of the Pearson's chi-squared tests between the $d(covs_{input}, covs_s)$ bin values and the scenario clusters. The darker bin represents significant results with p-values < 0.05 , indicating dependence between the covariate values and scenarios assigned. This suggests that the covariates bring predictive information relevant to the scenario creation. The grey bin corresponds to p-values < 0.1 , while the light grey bins represent cases where the null hypothesis of independence cannot be rejected, implying that, in those cases, the covariates mostly bring noise.

cases where the null hypothesis of independence cannot be rejected, implying that, in those cases, the covariates mostly bring noise. The plots show a relatively uniform result across the covariates, with a low percentage of cases indicating dependence between covariates and scenarios. Dependence is observed in 4% of cases for the unemployment rate and 6% for conflict displacement. The Diaspora covariate stands out slightly, as in the previous test, with dependence in 10% of the cases. Similarly, the Jobs availability variable has 10% of highly significant and 7% of weakly significant. These bins represent just under 20% of the cases, suggesting a relatively weak association between this covariate and scenario dependence.

Random Forest Feature Importance

With the Random Forest model, the importance of covariates, or features, can be estimated using feature importance metrics. These metrics quantify how much each feature contributes to the model’s predictive power by measuring the reduction in impurity caused by the feature at the decision tree nodes. In our case, impurity is represented by variance. Feature importance is calculated as the average decrease in impurity attributed to a feature across all trees in the forest. A higher importance value indicates that the feature contributes to a greater average reduction in impurity.

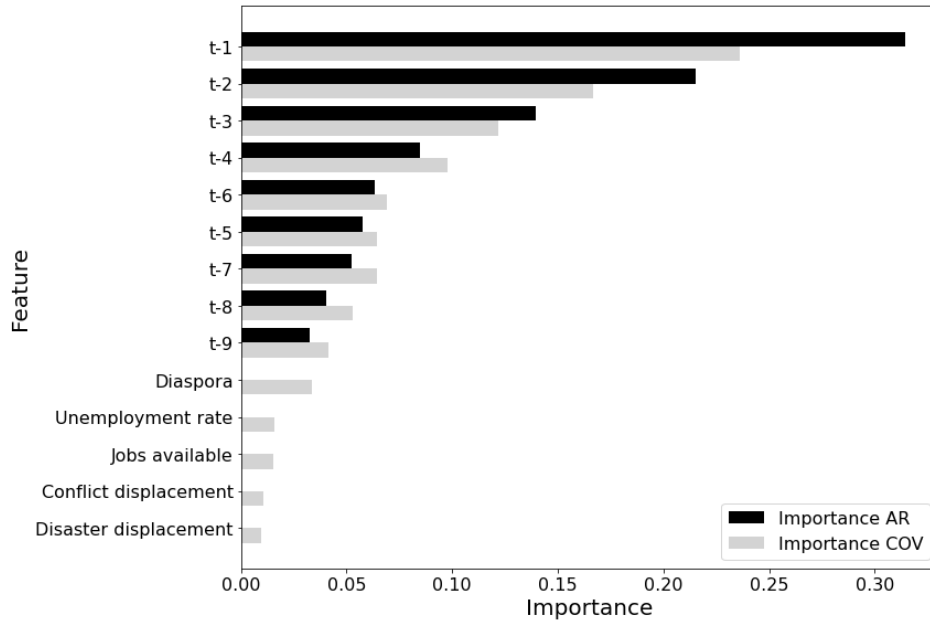


Figure 29: Feature importance score (in x-axis) for the Random Forest purely autoregressive, called “AR” (in black), and the model with covariate called “COV” (in light grey).

Figure 29 presents the feature importance scores for the purely autoregressive Random Forest model (RF) in black and the Random Forest model with covariates (RF+Cov) in light grey. Overlapping scores are shown in grey. The feature importance from the RF model aligns with the autocorrelation function (ACF) dynamics, as expected. For

the RF+Cov model, the results are consistent with previous findings: all covariates exhibit lower importance compared to any autoregressive terms. Also, the importance of the first three autoregressive terms in the RF model compensates for the inclusion of covariates in the RF+Cov model. According to earlier observations, Diaspora stands out as the most important covariate, showing twice the importance of any other covariate.

Forecast evaluation

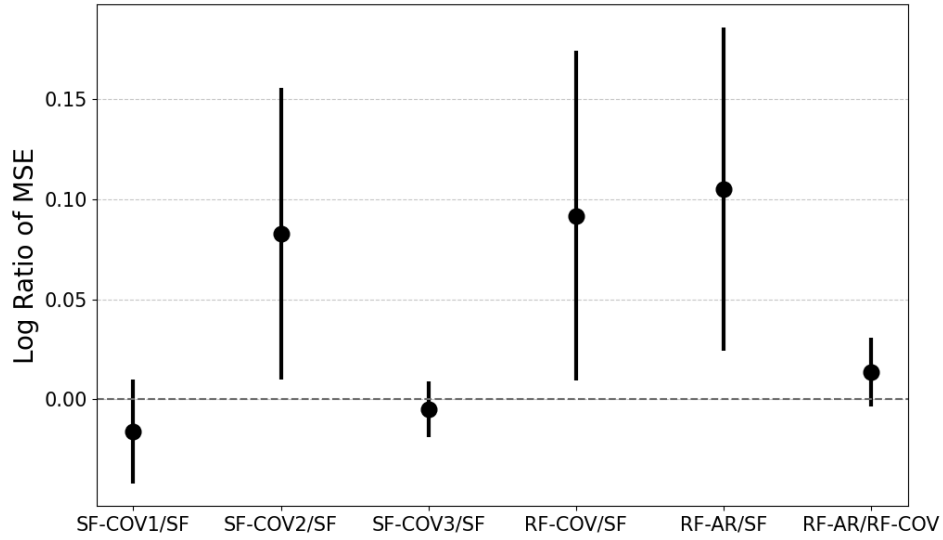


Figure 30: The log ratio of the Mean Squared Error (MSE), using migration axes as individual observations. For each forecast horizon from 1 to 6, the MSE of each model is computed. Performance is then compared across models using the log ratio of their MSEs. A positive log ratio indicates better performance of the model in the denominator of the ratio. The point represents the mean value, and the line is the confidence interval. This interval is constructed as $t_{0.975, n-1} * SD / \sqrt{n}$ where $t_{0.975, n-1}$ is the critical value of t , SD the standard deviation and n the number of observations. The denominator model is ShapeFinder (SF) for the first five error bars and RF-COV for the last one.

The evaluation of benchmark performance provides insight into two aspects: the comparative performance of a shape-based model versus a machine learning model, and the effect of incorporating covariates on forecasting performance. Figure 30 shows the log ratio of the Mean Squared Error (MSE), using migration axes as individual

observations. For each forecast horizon from 1 to 6, the MSE of each model is computed. Performance is then compared across models using the log ratio of their MSEs. A positive log ratio indicates better performance of the model in the denominator of the ratio. In the figure, the denominator model is ShapeFinder (SF) for the first five error bars and RF-COV for the last one. Each pairwise model comparison uses 157 observations, corresponding to the number of migration routes. The point displayed in each case is the mean log ratio, and the error bar represents the confidence interval.³⁴

The results indicate that the ShapeFinder with covariates included during scenario creation (SF-COV2) and both random forest models (RF-AR and RF-COV) produce a comparable, significantly poorer performance compared to the original ShapeFinder. In contrast, the ShapeFinder versions with covariates incorporated during similar-case searches (SF-COV1) and during final prediction creation (SF-COV3) demonstrate slightly better, though not significantly superior, performance, with SF-COV1 performing the best. Finally, the last log ratio point of the figure suggests that adding covariates to the Random Forest models does not lead to significant performance improvements. Additional figures on the models' parameters grid search are presented in Appendix A.3.

Discussion

Beyond the strong performance of the shape-based model, the primary finding is the lack of clear benefits from incorporating covariates into migration autoregressive models. This conclusion is supported by evidence from the shape-based model, but also the Random Forest, where covariates exhibit very low importance scores in the feature rankings. This does not imply that push and pull factors have no impact on migration,

³⁴This interval is constructed as $\pm t_{0.975, n-1} * SD / \sqrt{n}$ where $t_{0.975, n-1}$ is the critical value of t , SD the standard deviation and n the number of observations

far from it. Instead, it suggests that autoregressive components already capture much of the relevant information.

Given that migration reflects complex human decision-making processes, covariates alone cannot fully capture these mechanisms. First, the selected covariates may provide limited information at a monthly scale. Many of them change slowly, which makes them more suitable for explaining yearly migration trends rather than short-term monthly dynamics. Second, capturing the broader socio-ecological system would require an amount of data that cannot be gathered for a general model forecasting monthly flows at the country level. The earlier example of cultivation practices influencing migration in Bangladesh illustrates this issue: such effects are highly context-specific and would require detailed, localized data that are impossible to collect systematically. A similar limitation is observed in conflict forecasting, where complex underlying mechanisms limit the predictive value of slowly changing variables (Schincariol, Frank, and Chadeaux, 2025). For now, autoregressive terms remain the most reliable proxy for short-term migration prediction.

New data sources, such as social media (Yildiz et al., 2024) and online search keywords (Böhme, Gröger, and Stöhr, 2020), offer promising opportunities for migration research. These sources provide two key advantages. First, they operate at a much finer spatiotemporal resolution compared to traditional covariates, which are often limited to country-year scales or even static measures. In contrast, these new data types can offer insights at daily or regional levels, or even finer. Second, they represent the final stage of the human decision-making process. For instance, an individual searching online for “asylum seek Italy” has likely already decided to migrate to Italy, making the subsequent migration relatively straightforward. In contrast, traditional covariates, such as regional instability, may only signal a potential migration decision, which could take anywhere from days to years to materialize, depending on the context.

The limited availability of high-quality migration data constrains our findings. Since migration data is costly to collect, only a small amount is available, particularly at fine spatiotemporal resolutions. Emerging methods, such as those using remote sensing, mobile phone GPS, or even email IP addresses (Tjaden, 2021), offer promising new real-time information sources on millions of individuals. However, these approaches present significant challenges. Ethical and privacy concerns arise, as most of this data is collected by private companies and may not be publicly available. Also, unequal access to technology excludes portions of the population, potentially introducing biases into the data.

Conclusion

This article looked into the challenges and opportunities in migration flow forecasting, especially considering the growing impact of climate change on global migration. While the prediction task remains complex due to poor data quality and the multifaceted nature of migration, the study introduces the ShapeFinder model as a promising approach.

ShapeFinder employs shape similarity for predictions, offering solid performance and much-needed flexibility for modeling migration flows. A standout feature of this model is its interpretability, as its forecasts rely on analogous cases found in historical datasets. This transparency sets it apart from other machine-learning models, which may be perceived as black-box models.

The study also examines the inclusion of covariates, particularly push-pull factors, into migration forecasting models. Despite being common in political science, adding such factors to autoregressive models did not enhance prediction accuracy. This finding aligns with (Schincariol, Frank, and Chadeaux, 2025), which concluded that purely autoregressive models are most effective for modeling fundamentally complex phenomena

like migration or conflict. Incorporating additional covariates might introduce noise and increase data demands, surpassing any potential benefits.

In conclusion, the ShapeFinder model holds promise as a foundation for developing early warning systems for migration flows. However, traditional push-pull factors fail to improve predictive accuracy. Future research could explore new sources of covariates like internet search data, which reflect the final migration decision after the decision-making process, still impossible to model with the current modeling tools.

References

- Abel, Guy J et al. (2019). “Climate, conflict and forced migration”. In: *Global environmental change* 54, pp. 239–249.
- Bijak, Jakub et al. (2019). “Assessing time series models for forecasting international migration: Lessons from the United Kingdom”. In: *Journal of Forecasting* 38.5, pp. 470–487.
- Böhme, Marcus H, André Gröger, and Tobias Stöhr (2020). “Searching for a better life: Predicting international migration with online search keywords”. In: *Journal of Development Economics* 142, p. 102347.
- Carammia, Marcello, Stefano Maria Iacus, and Teddy Wilkin (2022). “Forecasting asylum-related migration flows with machine learning and data at scale”. In: *Scientific Reports* 12.1, p. 1457.
- Castelli, Francesco (2018). “Drivers of migration: why do people move?” In: *Journal of travel medicine* 25.1, tay040.
- Collier, Paul and Anke Hoeffler (2018). “Migration, diasporas and culture: An empirical investigation”. In: *Kyklos* 71.1, pp. 86–109.
- Cottier, Fabien and Idean Salehyan (2021). “Climate variability and irregular migration to the European Union”. In: *Global Environmental Change* 69, p. 102275.
- Crippa, Andrea et al. (2024). “Conflict as a Cause of Migration”. In: *Oxford Economic Papers*, gpae037.
- Dago, Franck and Simon Barussaud (2021). “Push/Pull Factors, Networks and Student Migration from Côte d’Ivoire to France and Switzerland”. In: *Social Inclusion* 9.1, pp. 308–316.
- Economic, United Nations Department of and Population Division Social Affairs (2020). *International Migrant Stock 2020*. Tech. rep. New York: United Nations.

- Everett, S Lee and ES Lee (1966). “A theory of migration”. In: *Demography* 3.1, pp. 47–57.
- Ferwerda, Jeremy and Justin Gest (2021). “Pull factors and migration preferences: Evidence from the Middle East and North Africa”. In: *International Migration Review* 55.2, pp. 431–459.
- Hegre, Håvard et al. (2024). “The 2023/24 VIEWS Prediction Challenge: Predicting the Number of Fatalities in Armed Conflict, with Uncertainty”. In: *arXiv preprint arXiv:2407.11045*.
- International Labour Organization (ILO) (2024). *SDG Indicator 8.5.2 - Unemployment Rate (%)*. <https://ilostat.ilo.org/topics/unemployment-and-labour-underutilization/>. Accessed: 2024-11-14.
- Johnson, Todd M and Gina A Bellofatto (2012). “Migration, religious diasporas, and religious diversity: A global survey”. In: *Mission Studies* 29.1, pp. 3–22.
- Khalid, Bilal and Mariusz Urbański (2021). “Approaches to understanding migration: a mult-country analysis of the push and pull migration trend”. In: *Economics & Sociology* 14.4, pp. 242–267.
- Khan, Md Imran et al. (2023). “Statistical analysis of push and pull factors of migration: A case study of India”. In: *Journal of King Saud University-Science* 35.8, p. 102859.
- Mallick, Bishawjit (2019). “The nexus between socio-ecological system, livelihood resilience, and migration decisions: Empirical evidence from Bangladesh”. In: *Sustainability* 11.12, p. 3332.
- Migration (IOM), International Organization for (2024). *DTM Europe — Mixed Migration Flows to Europe*. Tech. rep. IOM, Europe and Central Asia.
- Müller, Meinard (2007). “Dynamic time warping”. In: *Information retrieval for music and motion*, pp. 69–84.

- OECD (2023). *International Migration Outlook 2023*. Paris: OECD Publishing. DOI: 10.1787/b0f40584-en. URL: <https://doi.org/10.1787/b0f40584-en>.
- Prieto-Curiel, Rafael et al. (2024). “The diaspora model for human migration”. In: *PNAS nexus* 3.5, p. 178.
- Refugees, United Nations High Commissioner for (2023). *Global Trends: Forced Displacement in 2023*. Geneva: UNHCR. URL: <https://www.unhcr.org/global-trends-2023> (visited on 01/10/2025).
- Schincariol, Thomas and Thomas Chadeaux (2025). “Temporal Patterns in Migration Flows Evidence from South Sudan”. In: *Journal of Forecasting* 44.2, pp. 575–588.
- Schincariol, Thomas, Hannah Frank, and Thomas Chadeaux (2025). “Accounting for variability in conflict dynamics: A pattern-based predictive model”. In: *Journal of Peace Research*, p. 00223433251330790.
- Tjaden, Jasper (2021). “Measuring migration 2.0: a review of digital data sources”. In: *Comparative Migration Studies* 9.1, p. 59.
- Turkoglu, Oguzhan and Thomas Chadeaux (2019). “Nowhere to go? Why do some civil wars generate more refugees than others?” In: *International Interactions* 45.2, pp. 401–420.
- Urbański, Mariusz (2022). “Comparing push and pull factors affecting migration”. In: *Economies* 10.1, p. 21.
- Welch, Nathan G and Adrian E Raftery (2022). “Probabilistic forecasts of international bilateral migration flows”. In: *Proceedings of the National Academy of Sciences* 119.35, e2203822119.
- Yildiz, Dilek et al. (2024). “Integrating Traditional and Social Media Data to Predict Bilateral Migrant Stocks in the European Union”. In: *International Migration Review*, p. 01979183241249969.

5 Temporal Patterns in Migration Flows : Evidence from South Sudan

Temporal Patterns in Migration Flows: Evidence from South Sudan

Thomas Schincariol and Thomas Chadeaux

Abstract

What explains the variation in migration flows over time and space? Existing work has contributed to a rich understanding of the factors that affect why and when people leave. What is less understood are the dynamics of migration flows over time. Existing work typically focuses on static variables at the country-year level, and ignores the temporal dynamics. Are there recurring temporal patterns in migration flows? And can we use these patterns to improve our forecasts of the number of migrants? Here, we introduce new methods to uncover temporal sequences—motifs—in the number of migrants over time, and use these motifs for forecasting. By developing a multivariable shape similarity-based model, we show that temporal patterns do exist. Moreover, using these patterns results in better out-of-sample forecasts than a benchmark of statistical and neural networks models. We apply the new method to the case of South Sudan.

What explains the variation in migration flows over time and space? Existing work has contributed to a rich understanding of the factors that determine what pushes people away from their home: war, poverty, or oppression—push factors; and what attracts them elsewhere: economic opportunities or freedom—pull factors. What is less understood is the dynamics of migration flows over time. Existing work typically focuses on static variables at the country-year level. The analysis may involve more fine-grained spatial units, but rarely is the temporal dynamic investigated.

In contrast to this predominantly static approach, our research aims to improve the accuracy of migration flow predictions by identifying dynamic temporal patterns in migration flows. This study investigates whether the sequence of events, rather than just individual events, can provide additional insights into migration patterns. By identifying these temporal patterns, we aim to enhance the accuracy of forecasts regarding migrant numbers and address the shortcomings of current predictive models.

Our novel methodology combines dynamic time warping with pattern recognition algorithms. This approach enables us to discern not just the presence of migration motifs but also their specific characteristics and temporal variations. By applying these methods to fine-grained migration data, we aim to more accurately forecast future migration trends and understand the underlying dynamics driving these flows.

Additionally, our approach benefits from relying on a limited number of covariates rather than employing complex machine learning models with extensive features. This strategy offers significant advantages: it reduces the risk of overfitting, simplifies interpretation, and requires less computational power and data. Complex machine learning models, while powerful, often need large datasets to perform effectively and face issues related to reproducibility and interpretability due to their non-deterministic nature. In contrast, our streamlined approach provides a practical and robust solution, making it more accessible and applicable to various contexts where data may be limited.

To validate our motifs, we take advantage of a unique dataset that is temporally fine-grained. We use data from the IOM’s Displacement Tracking Matrix (DTM), focusing specifically on forced displacement. This includes internal displacement within South Sudan and, where applicable, cross-border refugee movements. The DTM provides fine-grained temporal and spatial data on out-flows from locations within South Sudan, allowing us to analyze short-term fluctuations and localized patterns in forced displacement.

We show that our method, focusing on temporal shapes—distinctive patterns in migration data over time—yields better results than traditional benchmarks set by models like regression models or neural networks. We apply this approach to the case of South Sudan, a context marked by a complex interplay of food prices and conflict fatalities. Our findings offer more accurate predictive models and provide insights that could inform migration policy and aid in understanding similar migration dynamics in other regions.

Our findings are important for a number of reasons. First, migration studies suffer from a lack of fine-grained data. Data are usually coarse and unavailable at the sub-national level. However, we show here that even with a small temporal coverage, we can extract important temporal patterns. These results also have important implications for policy-makers and actors in the field such as NGOs. Predicting migrant flows matters, and although those on the ground may develop over time an intuition for the patterns we uncover here, the approach here has the advantage of extracting patterns that may have been missed in practice.

Patterns in migration flows

The covariates of migration

The question of why and when people leave their homes has received a substantial amount of attention in the literature. The individual decisions leading each person or family to leave are complex. Whether a person flees may depend on their means, their tolerance for risk (Engel and Ibáñez, 2007), or their ability to secure a better alternative elsewhere. These and other factors may in turn vary by age (Kassar and Dourgnon, 2014), education (Van Dalen, Groenewold, and Schoorl, 2005), gender (Lauby and Stark, 1988), or personality (Arcand and Mbaye, 2013).

Beyond individual motives, a number of factors correlate with migration. Most can be classified either as push or pull factors. Conflict (Davenport, W. Moore, and Poe, 2003), human rights violations (W. H. Moore and Shellman, 2004), corruption (Dimant, Krieger, and Meierrieks, 2013) or climate change (Dasgupta et al., 2016), for example, tend to ‘push’ people away from their home. On the other hand, pull factors may attract migrants to other countries: easily accessible neighbors (Turkoglu and Chadeaux, 2019), well-functioning institutions (Ariu, Docquier, and Squicciarini, 2016) or income opportunities (Ortega and Peri, 2009).

While much of the literature focuses on push and pull factors, recent research has begun to explore patterns within migration flows themselves. Some studies at the country-year level have provided initial insights into temporal variations in migration (Melander and Öberg, 2006; Fearon, Shaver, et al., 2020). However, these analyses often miss finer-grained temporal dynamics.

In short, we know that migration—whether forced or not—is affected by a multitude of factors and their interactions (De Haas, 2011). What is less well understood are the

dynamic processes that lead to migration. Indeed, it is likely that it is not solely the raw value of these variables that matters, but also their evolution over time. In other words, it may not be enough to say that the level of a particular variable matters; we also need to understand how its variation over time may affect the dynamic of migration itself.

Why Patterns?

Why would migration flows follow regular temporal patterns? While it is well understood that migration is not random but is driven by various covariates, the idea that there may be recurring temporal patterns in refugee flows has, to our knowledge, not been explored. Of course, existing models do incorporate some elements of temporal dependence, such as lags or first differences. However, these models cannot account (i) for the variety of possible patterns; (ii) for their possible variations over time—the coefficients that are true today may no longer be tomorrow; and (iii) for the fact that these patterns may be warped and stretched versions of each other, such that a one-to-one matching approach will not be able to uncover them.

First, patterns may emerge out of the complex interplay of variables. Patterns emerge spontaneously in natural (Malchow, 2007), animal (Baurmann, Gross, and Feudel, 2007), as well as in social phenomena (Chadefaux, 2021). Indeed, a highly simplistic model alone is enough to generate temporal patterns. Suppose, for example, that the number of migrants and the number of violent armed attacks are linked by a system of two differential equations.

$$\frac{dp}{dt} = \alpha p - \beta p v, \quad (8)$$

$$\frac{dv}{dt} = \delta p v - \gamma v, \quad (9)$$

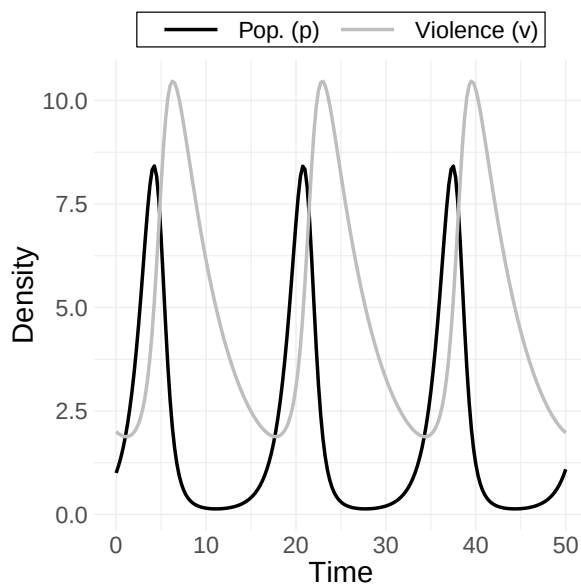


Figure 31: Temporal patterns can emerge from simple differential equations.

where p denotes the population left in the area (and hence the number of migrants), and v the number of violent events in the same area. $\frac{dv}{dt}$ and $\frac{dp}{dt}$ therefore represent the growth rates of migrants and violent events at a particular time t . The underlying idea is that violent events are more likely to take place when the population remains in the area—when there are people to victimize; but people are more likely to flee when there is more violence. This is clearly not a realistic model of migration, and has no intention to represent the underlying dynamic or causes of migration. However, it illustrates that even simple rules of interaction can lead to recurring non-linear temporal patterns (Fig. 31).

Second, migration may be driven by variables that themselves follow repeating temporal patterns. Migration and refugee flows, for example, are frequently associated with conflict (Schon, 2019), which itself can follow repeating temporal patterns (Chadefaux, 2021). Climate-related factors have also been linked to migration (Kaczan and Orgill-Meyer, 2020), and some of these patterns are obviously at least seasonal. Such seasonality is easily captured by an ARIMA model. However, others may be more a

complex and harder to define a priori. It may be the case, for example, that a recurrent pattern of droughts over months or years drive people away. Such complex and nonlinear patterns will be difficult or impossible to uncover for regular approaches relying on lags or differencing.

Another cause for the emergence of migratory patterns is that people are likely to react to real or observed patterns in their environment. Migrants may for example respond to ongoing violence in the country. While the level of violence itself matters, it is likely that they will also base their decision to move on the expected evolution of future trends. These expectations will be based on past trends or mental models. If so, migration itself will mimic the underlying patterns of these variables.

Indeed, humans excel at detecting patterns in their environment—conflict, climate, technology, or their networks. A long literature in psychology and neuroscience shows the ability of various organisms to form complex temporal representations of events (Gallistel, 1990).³⁵ People construct ‘mental models of temporal sequences of events’ (Schaeken, Johnson-Laird, and d’Ydewalle, 1996). “humans build discrete mental simulations of possibilities – mental models – when they reason.” “Reasoners do not represent all of the time points across which an event might endure. Instead, they construct discrete tokens that stand in place of the beginnings and endings of those events (Kelly and Khemlani, 2019). In other words, humans are able to adjust their mental models to events of different duration, as long as these events form a similar pattern delineated by key inflexion points. For example, while a single episode of drought may not prompt widespread migration, a series of such events significantly increases the likelihood of people relocating. This pattern suggests that not only the occurrence of adverse events

³⁵Multiple theories account for our ability to recognize temporal patterns in our environments. Feature detection theory, for example, suggests that we process information by sorting and filtering incoming stimuli. Our feature detectors are individual or groups of neurons that encode specific features. More complex feature detectors can detect more elaborate and specific features. We can identify meaningful sequences of events thanks to our feature detection system (Reed, 2013).

but their frequency and perceived future trend influence migration decisions. Similarly, people will observe the unfolding of conflict events in their regions over time. As a result, their behavior itself is likely to reflect these patterns—or a combination of them. Looking at aggregate number of events or deaths in a given month may therefore not be sufficient to capture the potentially complex patterns of violence that lead people to flee.

Finally, patterns may arise due to factors unique to the migration context. A crucial element of these dynamics is the influence of migration networks and remittances in shaping migration trends over time. Migration networks, established by earlier migrants, can facilitate subsequent migration by reducing costs and risks for new migrants (Massey, Arango, et al., 1993). These networks can lead to cumulative causation, where initial migration creates social structures that increase the likelihood of further migration Massey, Goldring, and Durand, 1994. Similarly, remittances sent by migrants to their home communities can transform the economic landscape, potentially encouraging or discouraging further migration depending on their usage Taylor, 1999. These factors can create self-reinforcing patterns in migration flows that evolve over time, underscoring the importance of examining not only the drivers of migration but also the temporal patterns within the flows themselves.

A methodological problem

There are good reasons, however, for why temporal dynamics have received limited attention. This is due to two limitations of the tools we have worked with so far. First, data is typically limited to national and yearly levels, which does not offer the level of granularity needed to understand the dynamics of migration over time, but rather only lets us compare levels over long periods of time. Some work has started to look at more granular levels (Mayen, Wood, Frazier, et al., 2022), but the data remains scarce.

A second limitation is the difficulty of existing methods in the social sciences to account for temporal dynamics. The methods we use—typically regression—are ill-suited to incorporate the possibly complex ebbs and flows of a complex process such as migration. Existing methods rely on correlation, which compares the raw values of the covariates, and does not allow for possible warping in time and space. Correlation (and therefore regression) relies on the comparison of more or less static covariates, and only allows limited comparisons of time sequences. As a simple example, consider the two sequences presented in Tab.5. Both are clearly similar in that they follow a

Time	1	2	3	4	5	6	7	8	9	10	11
a	5	3	1	3	5	3					
b	5	4	3	2	1	2	3	4	5	4	3

Table 5: Two temporal sequences with varying lengths and speed.

‘down, up, down’ pattern. Yet, a unfolds over 6 time periods, but b over 11. Calculating the correlation between the two series is impossible due to their different lengths. Even if we were to limit our attention to only six observations of b —as a regression would—the correlation between the two is nearly zero. Existing methods that rely on correlation patterns, such as ARIMA or distributed lag models can therefore not adequately account for these patterns. While ARIMA models, for example, can model complex and even nonlinear sequences, they are poor at recognizing similar sequences that may unfold at different speeds. The same is true of standard regression models.

To be sure, there are alternatives to statistical approaches that can model time sequences effectively. Systems dynamics models and agent-based models (ABMs) offer detailed simulations of complex systems, capturing interactions and feedback loops. However, they require significant development time and extensive domain knowledge, making them resource-intensive. Deep-learning models for time series can also uncover complex patterns in data but often require large datasets to perform effectively and

face issues related to reproducibility and interpretability due to their non-deterministic nature. Additionally, machine learning models can be finicky, often needing careful tuning and large amounts of data to yield reliable results. Furthermore, relying on a limited number of covariates, as we have done, offers significant advantages over complex machine learning models with extensive features. This approach reduces the risk of overfitting, simplifies interpretation, and requires less computational power and data, making it more practical for many applications.

This inability to account for variations in length and speed is unfortunate, because events may unfold at different paces. Episodes of violence and their associated trajectories and shape may take place over days in some place, months in others. Ethnic cleansing may unfold over weeks, as was the case of Rwanda, or years, as in Darfur. The resulting patterns of migration may as a consequence be either fast or slow.

However, new data and methods now allow us to take into account not only static factors, such as the number of conflict events or food prices, but also the dynamics—the patterns—of these factors. We take advantage of methods derived mainly from speech recognition to account for the possibility of patterns at different temporal levels.

As an illustration of our approach (detailed below), consider the sequences presented in Fig. 32. They show the number of migrants (left, blue), the food price (center, green) from the Melut region in South Sudan in 2022 (top) and Guit region in 2021 (bottom). The dynamics of monthly migration (blue) in Melut are visually similar to those in Guit. The same applies to food prices (green). Both situations lead to similar migration outcomes (red). Using information from Guit could help forecast events in Melut. Below, we show how to identify similarities across multiple dimensions and find close matches that improve future predictions.

Our approach offers several significant advantages over using complex machine learning models with extensive features. First, by focusing on a limited number of covariates,



Figure 32: Human migration (blue) and Food price (in green) for Melut region in South Sudan (top) and Guit (bottom). The red section represents the migration figures for the three months following the depicted period (blue) for both regions.

we significantly reduce the risk of overfitting. Overfitting occurs when a model is too closely aligned with the training data, capturing noise rather than the underlying pattern. Models with extensive features are particularly prone to this issue because they can fit the training data very well but fail to generalize to new, unseen data. In contrast, our approach with fewer covariates helps ensure that the model captures the essential patterns that are likely to be present in future data, improving its robustness and predictive power.

Second, a simplified model with fewer covariates enhances interpretability. In migration studies, it is crucial for policymakers and practitioners to understand the factors driving migration flows. Complex machine learning models, while powerful, often operate as “black boxes,” making it difficult to extract meaningful insights about the relationship between covariates and migration patterns. Our approach, by limiting the number of covariates, makes it easier to interpret the model’s outputs and understand the impact of each factor on migration flows. This transparency is valuable for developing targeted interventions and informing policy decisions.

Third, models with a limited number of covariates require less computational power and data, making them more practical for many applications. Complex machine learning models with extensive features often demand substantial computational resources and large datasets to perform effectively. In many real-world scenarios, especially in regions with limited data availability, such requirements can be prohibitive. Our approach is more efficient and accessible, allowing for effective migration forecasting even with smaller datasets and limited computational infrastructure.

Data

Migration data are often limited to the country-year level. Most datasets list the flow of migrants from one country to the next in any given year. While this level of analysis is adequate for a static analysis of the factors leading to migration, they are not sufficient to understand the temporal dynamics at play. Yearly, national-level data fail to capture the subtleties of migration trends that occur on a finer temporal scale.

To address this gap, we take advantage of a unique dataset from the International Organization for Migration’s Displacement Tracking Matrix (IOM-DTM) flow monitoring surveys. The dataset offers a detailed view of migration at the sub-national level, providing monthly data on the origins of migrants within South Sudan from January 2020 to October 2022. The IOM-DTM captures displacement, including both internal displacement within South Sudan and cross-border refugee movements. This data encompasses individuals fleeing due to conflict, violence, human rights violations, economic, healthcare, and natural or human-made disasters.³⁶

The IOM-DTM conducts surveys at 31 critical transit points across South Sudan and its borders, providing a level of temporal detail that annual data cannot achieve. The

³⁶Note that the definition of ‘migration’ in this context also include voluntary or economic migration, such as seasonal labor movements or long-term economic relocation.

focus of our analysis is the monthly count of migrants leaving each of the 76 provinces (administrative region 2) in South Sudan.³⁷ For each of these regions, we obtain a time series of 34 data points, each representing the number of migrants leaving each region monthly from January 2020 to October 2022.

Additionally, we incorporate two key exogenous variables into our analysis: food prices and conflict data. The inclusion of these variables is motivated by their significance as major push factors in migration, with both datasets offering the necessary granularity for our analysis at a monthly and regional level. Food price data are derived from the World Bank’s dataset, focusing on the monthly prices of beans in 29 markets across South Sudan (Andrée, 2021). These prices are then associated with the nearest regions to each market, based on the assumption that populations primarily access their nearest markets. Food prices can serve as a valuable proxy for assessing both climatic and economic conditions. Extreme events such as droughts or floods have a direct and significant impact on food prices. Therefore, fluctuations in food prices give insights on the environmental and economic dynamics within a region. The conflict data are sourced from the Uppsala Conflict Data Program (UCDP) Georeferenced Event Dataset (GED) (Sundberg and Melander, 2013; Davies, Pettersson, and Öberg, 2022), which provides us with monthly data on conflict events for each region in the country over the entire period.

Methods

Our approach to uncover and use temporal patterns in migration data is twofold: first, it involves the identification of recurring temporal motifs or patterns in migration data—‘motifs’. These motifs are not predefined but are instead extracted directly from the

³⁷Three additional regions (Luakpiny, Manyo, and Ulang) were excluded due to the absence of migrant data during the study period.

data, allowing for the discovery of patterns that traditional static models may overlook. Second, we use these extracted motifs to forecast migration at the regional level.

This methodological innovation hinges on the application of Dynamic Time Warping (DTW), a technique that allows for the flexible comparison of time-series data, even when these series unfold at different speeds. This flexibility is crucial for accurately matching temporal patterns across different contexts and time frames. The following sections detail the specific steps of this approach, including the extraction of motifs from the data, the application of DTW for pattern matching, and the integration of these elements into a predictive model for migration flows.

Searching for patterns

Identifying patterns in migration data is challenging due to their inherent complexity and variability. There are generally two approaches to uncover these patterns. The first would involve pre-defining patterns based on existing theories. For example, an ‘up-down-up’ pattern in migration could be hypothesized to reflect specific sequences of events leading to increased migrant flow. However, this approach has limitations. First, existing theories do not discuss temporal patterns, but rather focus on static covariates like push and pull factors. As a result, we do not have a good theoretical starting point to hypothesize the existence of particular shapes. Furthermore, relying solely on pre-defined patterns risks missing out on discovering novel, potentially more significant patterns that existing theories do not anticipate.

Our approach is more flexible. Instead of testing predefined patterns, we extract relevant patterns directly from the data. This method involves analyzing historical data to find matching sequences and using these patterns to forecast future events. This approach has two advantages. First, it significantly reduces the risk of overfitting, ensuring the patterns we identify reflect genuine trends rather than random noise.

Secondly, by validating our model’s predictions against actual future events, we can differentiate between true underlying phenomena and mere coincidences in the dataset. Our approach to uncovering migration patterns thus involves two main steps: first, measuring the similarity between time series, and then using these similarity scores to create forecasts. We detail these steps below.

Matching subsequences

Traditional correlation analysis can fail to recognize similarities between time series. For example, time series a and b in figure A.27 look alike, yet their correlation cannot be calculated over the entire range, and is near zero regardless of how we truncate b . We therefore rely on *Dynamic Time Warping* (DTW), a method that allows us to match time series with more flexibility on the time dimension. DTW is a method for measuring the similarity between two time series by examining how well the overall patterns of the time series match, even if they do not align perfectly in time. This is different from traditional methods that only focus on the differences in values at each specific time point.

The essence of DTW lies in its ability to adjust time indices between two series. This involves strategically modifying the time dimension of the series through stretching or compressing, and possibly duplicating or shifting data points, to achieve alignment of similar patterns. Such adjustments allow for a more natural comparison beyond mere point-to-point differences, focusing on the overarching shape or pattern similarity between the series.³⁸ Dynamic time warping allows us to find similarities between time series, even when these series unfold at different speeds. The advantage is that we can uncover time series with similar shapes, in contrast to existing approaches, which focus

³⁸For a more detailed explanation of the DTW algorithm and its implementation, please refer to appendix A.4 and Müller (2007).

on the raw values of the covariates.

We apply DTW to find resemblances between time series representing migration and covariates (food prices and conflict) even when they unfold at different rates. By focusing on shape similarities, we can better anticipate future trends in migration. The distance between two region-months (our unit of analysis) i and j is defined as the sum of the DTW distances across these variables:

$$d_{i,j} = \sum_v \text{DTW}(seq_{i,v}, seq_{j,v}), \quad (10)$$

where v represents each variable of interest—Food prices, Conflict fatalities, and Migration.³⁹ Based on the distance $d_{i,j}$, the observation is classified as a match if the distance is smaller than a threshold θ .⁴⁰ The overall process is described in figure 33.

Forecasts

To forecast the number of migrants originating from a particular region in a given month, we follow three simple steps. First, the most recent time series of region i is extracted. Then time series with a suitably small distance $d_{i,j}$ are extracted in the historical data. To identify these matches, we consider all regions of South Sudan in the past. Finally, the subsequent $h \in \{1, 2, 3\}$ values of these matches—what we call past futures—are collected (squared sequences in figure 33). The predicted values are then

³⁹Note that using a simple sum of these distances may not be the most effective approach. A more refined method could involve a weighted average, where the weights are derived from the data itself, potentially leading to improved accuracy. Nevertheless, our current focus is on maintaining simplicity in our methodology, and as such, we opt for the straightforward summative approach at this stage. Results for different weights do not substantially differ from the ones presented here and are reported in appendix figure A.31.

⁴⁰The model uses an observation window of 6 months for all variables. The threshold to classify two cases as similar is set to 3, using normalized (min-max from 0 to 1) sequences. Several threshold values were evaluated, however, the highest-scoring threshold value of 3 was found to be the most effective, achieving a valid balance between specificity and generalization. The individual flexibility for each variable is approximately 1 under this threshold configuration.

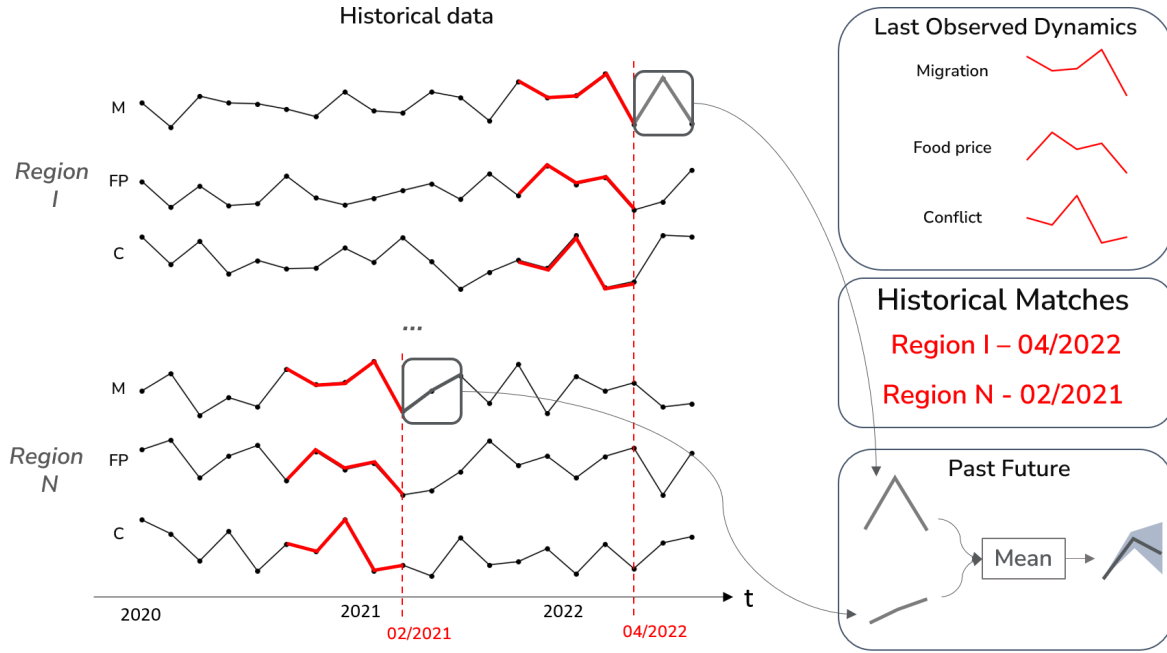


Figure 33: Illustrative representation of the shape similarity model's match-finding and forecasting methodology. Displayed in the top-right quadrant are the most recent shape observations for each variable within the specified region. A rolling window approach is applied to each region to identify instances of similarity. Upon recognizing a similar case (e.g., April 2022 in Region I), the migration values of the ensuing month are recorded and categorized as 'Past Future'. The predicted value is then computed as the average of these 'Past Future' sequences.

the mean values of these past futures. Thus, the prediction for region i and time $t + h$ is the mean of matching units j at time $s + h$, where s is the time at which matching unit j was observed.⁴¹ Note that forecasts are made for all $h \in \{1, 2, 3\}$ simultaneously at time t . As a result, the quality of predictions is expected to decrease as h increases.

⁴¹Note that all subsequences are normalized with Min-Max scaling from 0 to 1 to facilitate the comparison of shapes rather than magnitudes. The 'past futures' are scaled based on their corresponding subsequences, and the forecasted values are determined from normalized mean values. Afterward, the forecasted values undergo a reverse normalization transformation to obtain the final prediction, ensuring its adjustment to the original scale of the data.

Evaluation

We evaluate our forecasts by comparing them to two benchmarks. The first, Auto Regressive Integrated Moving Average (ARIMA), is a widely-used model for forecasting time series. The model is defined by the following equation:

$$\Delta^d Y_t = \mu + \phi_1 \Delta^d Y_{t-1} + \dots + \phi_p \Delta^d Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (11)$$

where $\Delta^d Y_t$ represents the d -th order differenced series, μ is the intercept $\phi_1, \dots, \phi_p$ are the autoregressive coefficients, $\theta_1, \dots, \theta_q$ are the moving average coefficients.

The ARIMA model can handle a variety of time series patterns, including trends and seasonality, by incorporating differencing, which makes it robust in dealing with non-stationary data. The model’s capability to integrate autoregressive and moving average components allows it to capture both short-term and long-term dependencies in the data.

Our second benchmark, Long Short Term Memory (LSTM), is a derivative of Recurrent Neural Networks (RNN), with the added advantage of being able to handle long-term dependencies in sequential data. This makes them ideal for complex time series forecasting tasks where context and history significantly influence future predictions (see appendix A.4 for details).

For each of the three models, the division of data into training and test sets follows a 90/10 ratio. This allocation translates to a test set duration of 3 months. The decision to opt for this distribution is based on the limited amount of data we have at our disposal, yet still leaves us with 228 observations (three observations for each of the 76 regions)—enough to derive statistically significant conclusions.

Results

Patterns

Figure 34 displays sample matches from the data, illustrating specific recurring instances of patterns. These observed patterns are likely the result of the complex interplay between various factors influencing migration, such as economic conditions, conflict, climate patterns, and human behavior as individuals respond to their changing environment.

However, sample repetitions are not enough to reach any conclusions, as these repetitions may occur by chance. After all, even a suitably long white noise series would include repeating patterns. Consequently, a pivotal question is whether these patterns are meaningful. Investigating the outcomes that follow such patterns can provide insights. If patterns do carry information, then we would expect similar patterns to lead to similar outcomes—something we would not expect from white noise. Suppose, for example, that pattern *A* is followed by a wide range of outcomes—many migrants sometimes, few others—whereas pattern *B* is consistently followed by a large number of migrants. Clearly, pattern *B* in this case is more informative.

To put this concept into practice, we gathered the standard deviation of the events following each identified pattern. A substantial standard deviation would imply that the observed patterns are predominantly stochastic and, hence, offer negligible predictive utility. In contrast, a minimal standard deviation signals that the outcomes subsequent to the patterns exhibit a high degree of uniformity. In Figure 35, the y-axis denotes the standard deviation of the ‘post-patterns’—the sequence of three observations that immediately follows the pattern—while the x-axis sorts patterns by their proximity.⁴² A

⁴²This interval is constructed as $\pm t_{0.975, n-1} * SD / \sqrt{n}$ where $t_{0.975, n-1}$ is the critical value of t , SD the standard deviation and n the number of observations.



Figure 34: Illustrative matches of temporal migration patterns (red), food prices (green), and conflict (yellow) from various regions and time periods in South Sudan, identified through dynamic time warping analysis. Each row showcases the data from a distinct region, with time plotted on the x-axis and the number of migrants on the y-axis. The columns represent these three variables for different regions or time periods. The intensity of the grey in the background series indicates the degree of similarity to the colored reference series (darker = more similar).

clear inverse relationship is observed: similar patterns are followed by similar outcomes, whereas dissimilar patterns exhibit larger standard deviations. This trend suggests that closer patterns lead to more consistent and predictable migration flows, suggesting that temporal motifs in migration data do carry important information.

Figure A.29 illustrates instances where patterns of human migration, food prices, and conflict fatalities are combined, along with the visual representation of their corresponding following migration values. The combination of patterns associated with the

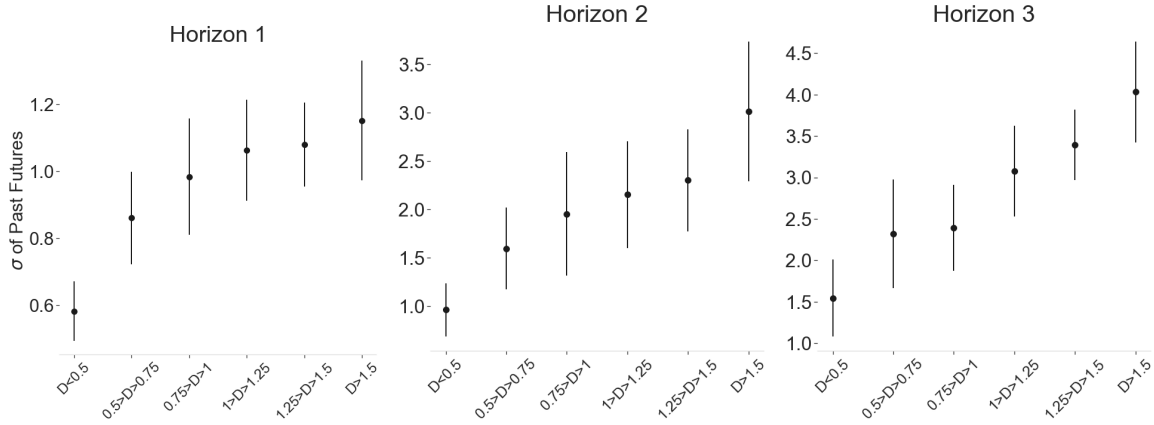


Figure 35: Pattern dissimilarity (D) and post-pattern migration consistency. The graph illustrates the inverse relationship between the degree of similarity in migration patterns (x-axis) and the standard deviation (σ) of subsequent migration flows (y-axis). Lower standard deviations following post-patterns (the three observations that immediately a sequence) indicate more predictable migration outcomes, underscoring the potential of using pattern similarity as a forecasting tool. The data suggest that as the proximity of patterns increases, the predictability of migration trends also increases, indicating that patterns carry important information.

highest standard deviation (top) and the lowest SD (bottom) of the following values are displayed.

Forecasts

We now validate the idea that these patterns carry valuable information by using them to forecast future migration values. To compare the performance of our proposed model against benchmarks, we rely on the log ratio of Mean Squared Error (MSE) as our primary metric. The choice of MSE is motivated by its widespread use and ease of interpretation, making it an ideal standard for assessing and contrasting the efficacy of the different models in our study.

In Figure 36 we present the log ratio of the mean squared errors (MSE) from our predictions.⁴³ Each point on the plot represents the average squared errors across all

⁴³Using ratios directly can be misleading due to their asymmetry: decreases fall between zero and one (e.g., 0.5 means half), while increases can be any number above. So simple ratios are not symmetric

time periods for a particular region.⁴⁴ A positive ratio indicates that the benchmark has larger errors than our model—i.e. that its performance is worse. We find that the ratios for both benchmarks are significantly above zero (1.92 and 0.89 for LSTM and ARIMA respectively, with t-tests significant at the $p < 0.001$ level). These results hold for a variety of metrics (table 6). These results hold for all $h \in \{1, 2, 3\}$ but not beyond. This is to be expected, as outperforming a model that more or less predicts the immediate past (ARIMA) becomes exceedingly difficult without incorporating additional information.

Table 6: Mean and 95% Confidence Intervals for various prediction metrics across different models.

Metric	LSTM	ARIMA	ShapeFinder
RMSE	260 ± 159	149 ± 107	120 ± 86
Med. Abs. Err.	238 ± 148	121 ± 88	81 ± 58
MAPE	$4.5 \times 10^{16} \pm 5.9 \times 10^{16}$	$9.8 \times 10^{15} \pm 1.1 \times 10^{16}$	$1.8 \times 10^{15} \pm 1.3 \times 10^{15}$
MAE	248 ± 155	134 ± 100	103 ± 75

In Figure 37, the geographical distribution of results is illustrated. Each region is color-coded based on the best model for that region. This gives us a sense not only of how much the MSE improves, but how often. Out of the three models, our model, labeled as 'ShapeFinder' (dark blue), has the lowest MSE in 70% of the regions. ARIMA, indicated by the light blue color, and LSTM in purple, are best in significantly fewer regions—15 and 8 respectively. Additionally, in the three regions with the highest count of migration values (Morobo, Juba, and Rubkona), our model has the best performance.

around the point of no change (1.0). By applying the logarithm to ratios, we achieve symmetry: no change corresponds to zero, increases are positive, and decreases are negative, making the log of 2.0 and 0.5 symmetric around zero.

⁴⁴In other words, each point corresponds to $\log\left(\frac{MSE_{benchmark}+1}{MSE_{ShapeFinder}+1}\right)$. The addition of one is a common technique to ensure that the log function is defined even when MSE values are zero.

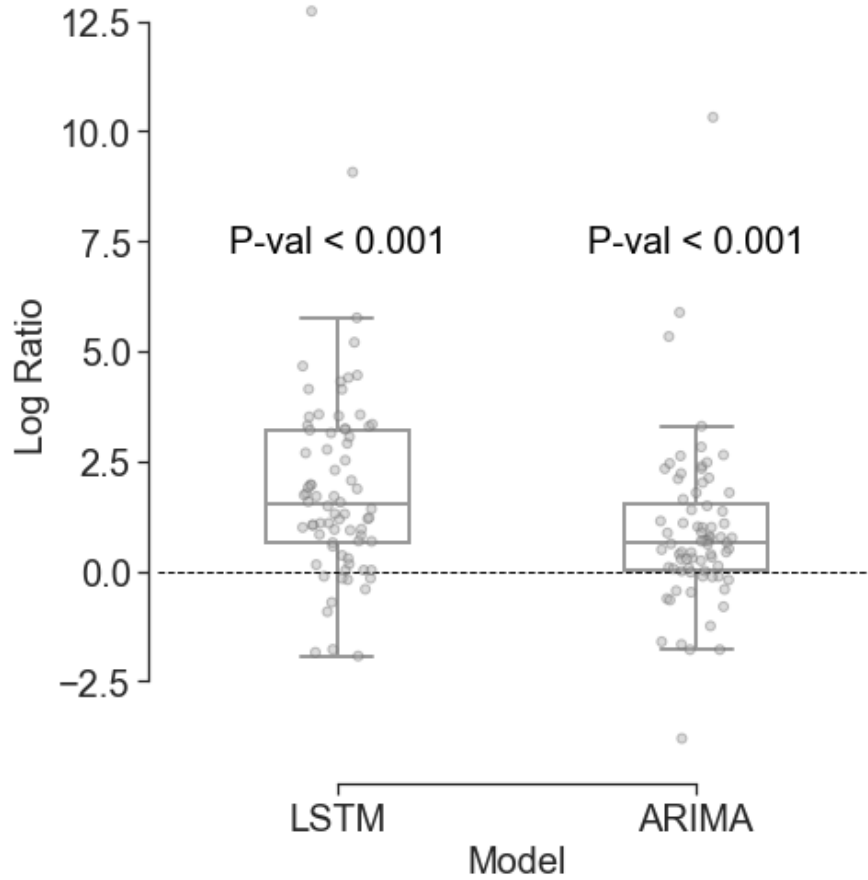


Figure 36: Log Ratio of the Mean Squared Error (MSE) for benchmark models (LSTM and ARIMA) compared to the shape similarity model. Each point represents a region of South Sudan. Points above the zero line indicate regions where the benchmark model has a higher (i.e., worse) MSE than the shape similarity model.

Covariates

Furthermore, regional variations in food prices and conflict fatalities provide further insights. Figure 38b highlights the North-East regions with higher food prices. In these regions, which exhibit the highest mean value in food prices, ARIMA or LSTM models perform best in 80% of the cases. This is further supported by the regression results shown in 38a. The negative regression coefficient implies that as food prices increase, model performance deteriorates. This observation is consistent with the findings of

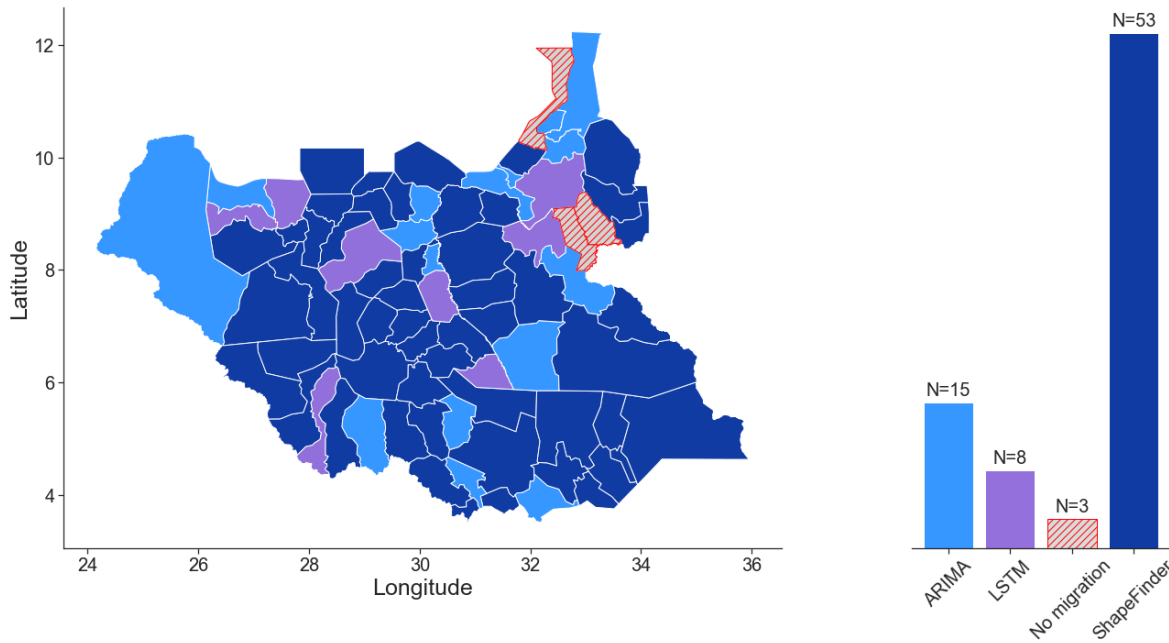


Figure 37: Map of the best-performing model for each region in South Sudan. Different colors indicate the model with the lowest Mean Squared Error (MSE) in each respective region. The accompanying bar plot shows the number of regions where each model achieved the best performance.

Smith and Wesselbaum, 2022, where a linear relationship between food insecurity and migration flows was identified. Given this linearity, simpler models like ARIMA perform well, suggesting that our more complex model may not be necessary.

Conversely, figure 39b shows regions with the highest sum of fatalities due to conflict. In these regions, 80% of the highest fatalities are best modeled using the ShapeFinder model. In 39a, the quadratic regression between conflict fatalities and the MSE log ratio is displayed for the two models, indicating that our model performs better in conflict-affected regions, but its effectiveness decreases as the number of fatalities rises. These findings underscore the importance of considering regional characteristics when selecting and evaluating models. The consistent performance of different models in varying contexts highlights the robustness of the proposed modeling framework, despite changes in the weight configurations.

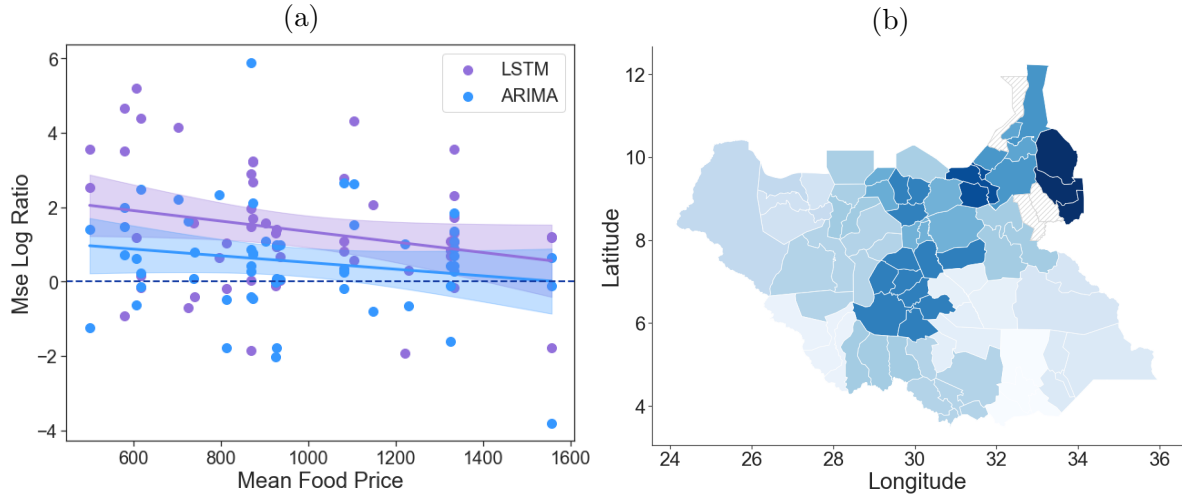


Figure 38: (a) MSE Log Ratio (in y) per Mean Food Price (in x). One point represents one model/region compared to our model. A positive point shows better performance for our model, contrary to negative points. The line represents the linear regression curves for LSTM (purple) and ARIMA (blue). The shaded area is the 95% Confidence Interval. (b) Average Food Price Per Region before the forecasting evaluation period (February 2022 - July 2022). The darker blue represents higher values.

Discussion

By expanding beyond static value-based models, this study enhances our understanding of migration trends in several ways. Static models typically rely on fixed data points, offering a snapshot view that may overlook underlying trends and patterns. In contrast, a dynamic approach analyzes changes over time, revealing how migration patterns—not just raw values—influence overall migration. Additionally, the inclusion of sequences associated with food prices and conflict dynamics allows us to model more complex temporal interactions and understand how they affect migration.

However, the study also faces significant data challenges. Continuous, fine-resolution data on migration are scarce, with most available datasets being limited in temporal and geographical scope. While projects like those led by the International Organization

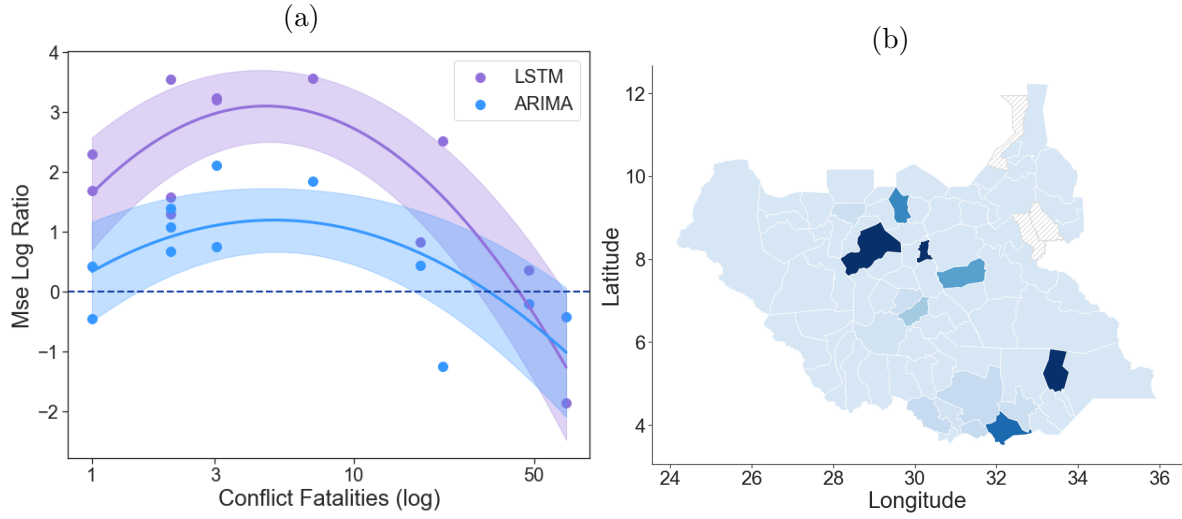


Figure 39: (a) MSE Log Ratio (in y) per Conflict Fatalities (log) (in x). One point represent one model/region comparison to our model. A positive point shows better performance for our model, contrary to negative points. The line represents the quadratic regression curves for LSTM (purple) and ARIMA (blue). The shaded area is the 95% Confidence Interval. (b) Sum of Conflict Fatalities Per Region before the forecasting evaluation period (February 2022 - July 2022). The darker blue represents higher values.

for Migration (IOM) in Iraq offer better granularity, they are hindered by data discontinuities, which pose a challenge for time series analysis and pattern extraction. The cost and logistical complexity of gathering migration data, typically through interviews at key migration points, contribute to the rarity of long-term data projects. Emerging techniques using nightlight satellite imagery or social media data show promise but are still developing and currently lack reliability.

Another limitation is that while our model excels at identifying changes, it cannot predict shifts from long periods of inactivity to activity. This limitation is not specific to our approach but is common in autoregressive models. Indeed, most statistical models often miss the onset of new trends following prolonged stability.

On the other hand, a significant advantage of the methodology presented here is its efficacy in extracting meaningful patterns and making predictions, even with small

3A4&f%5B1%5D=dataset_country%3A81

Conflict of interest statement

All authors declare that they have no conflicts of interest.

References

- Andrée, BPJ (2021). “Monthly food price estimates by product and market”. In: *WLD_2021_RTFP_v02-M; Version*, pp. 12–02.
- Arcand, Jean-Louis and Linguere Mbaye (2013). “Braving the waves: the role of time and risk preferences in illegal migration from Senegal”. In.
- Ariu, Andrea, Frédéric Docquier, and Mara P Squicciarini (2016). “Governance quality and net migration flows”. In: *Regional Science and Urban Economics* 60, pp. 238–248.
- Baurmann, Martin, Thilo Gross, and Ulrike Feudel (2007). “Instabilities in spatially extended predator–prey systems: Spatio-temporal patterns in the neighborhood of Turing–Hopf bifurcations”. In: *Journal of Theoretical Biology* 245.2, pp. 220–229.
- Chadefaux, Thomas (2021). “A shape-based approach to conflict forecasting”. In: *International Interactions* 48 (4), pp. 633–648.
- Dasgupta, Susmita et al. (2016). “Facing the hungry tide: Climate change, livelihood threats, and household responses in coastal Bangladesh”. In: *Climate Change Economics* 7.03, p. 1650007.
- Davenport, Christina, Will Moore, and Steven Poe (2003). “Sometimes you just have to leave: Domestic threats and forced migration, 1964-1989”. In: *International Interactions* 29.1, pp. 27–55.
- Davies, Shawn, Therese Pettersson, and Magnus Öberg (2022). “Organized violence 1989–2021 and drone warfare”. In: *Journal of Peace Research* 59.4, pp. 593–610.
- De Haas, Hein (2011). “The determinants of international migration”. In.
- Dimant, Eugen, Tim Krieger, and Daniel Meierrieks (2013). “The effect of corruption on migration, 1985–2000”. In: *Applied Economics Letters* 20.13, pp. 1270–1274.

- Engel, Stefanie and Ana María Ibáñez (2007). “Displacement due to violence in Colombia: A household-level analysis”. In: *Economic development and cultural change* 55.2, pp. 335–365.
- Fearon, James D, Andrew Shaver, et al. (2020). “Civil war violence and refugee outflows”. In: *Unpublished paper. Stanford University*.
- Gallistel, Charles R (1990). *The organization of learning*. The MIT Press.
- Kaczan, David J and Jennifer Orgill-Meyer (2020). “The impact of climate change on migration: a synthesis of recent empirical insights”. In: *Climatic Change* 158.3-4, pp. 281–300.
- Kassar, Hassène and Paul Dourgnon (2014). “The big crossing: illegal boat migrants in the Mediterranean”. In: *The European Journal of Public Health* 24.suppl.1, pp. 11–15.
- Kelly, Laura Jane and Sangeet Khemlani (2019). “The consistency of durative relations.” In: *CogSci*, pp. 1998–2004.
- Lauby, Jennifer and Oded Stark (1988). “Individual migration as a family strategy: Young women in the Philippines”. In: *Population studies* 42.3, pp. 473–486.
- Malchow, Horst (2007). *Spatiotemporal patterns in ecology and epidemiology: theory, models, and simulation*. Chapman and Hall/CRC.
- Massey, Douglas S, Joaquin Arango, et al. (1993). “Theories of international migration: A review and appraisal”. In: *Population and development review*, pp. 431–466.
- Massey, Douglas S, Luin Goldring, and Jorge Durand (1994). “Continuities in transnational migration: An analysis of nineteen Mexican communities”. In: *American journal of Sociology* 99.6, pp. 1492–1533.
- Mayen, John V, Erik Wood, Tim Frazier, et al. (2022). “Practical flood risk reduction strategies in South Sudan”. In: *Journal of emergency management* 20.8, pp. 123–136.

- Melander, Erik and Magnus Öberg (2006). “Time to go? Duration dependence in forced migration”. In: *International Interactions* 32.2, pp. 129–152.
- Moore, Will H and Stephen M Shellman (2004). “Fear of persecution: Forced migration, 1952-1995”. In: *Journal of Conflict Resolution* 48.5, pp. 723–745.
- Müller, Meinard (2007). “Dynamic time warping”. In: *Information retrieval for music and motion*, pp. 69–84.
- Ortega, Francesc and Giovanni Peri (2009). *The causes and effects of international migrations: Evidence from OECD countries 1980-2005*. Tech. rep. National Bureau of Economic Research.
- Reed, Stephen K (2013). *Psychological processes in pattern recognition*. Academic Press.
- Schaeken, Walter, PN Johnson-Laird, and Gery d’Ydewalle (1996). “Mental models and temporal reasoning”. In: *Cognition* 60.3, pp. 205–234.
- Schon, Justin (2019). “Motivation and opportunity for conflict-induced migration: An analysis of Syrian migration timing”. In: *Journal of Peace Research* 56.1, pp. 12–27.
- Smith, Michael D and Dennis Wesselbaum (2022). “Food insecurity and international migration flows”. In: *International Migration Review* 56.2, pp. 615–635.
- Sundberg, Ralph and Erik Melander (2013). “Introducing the UCDP georeferenced event dataset”. In: *Journal of Peace Research* 50.4, pp. 523–532.
- Taylor, J Edward (1999). “The new economics of labour migration and the role of remittances in the migration process.” In: *International migration* 37.1.
- Turkoglu, Oguzhan and Thomas Chadeaux (2019). “Nowhere to go? Why do some civil wars generate more refugees than others?” In: *International Interactions* 45.2, pp. 401–420.
- Van Dalen, Hendrik P, George Groenewold, and Jeannette J Schoorl (2005). “Out of Africa: What drives the pressure to emigrate?” In: *Journal of Population Economics* 18.4, pp. 741–778.

A Appendix

A.1 Appendix : Paper 1

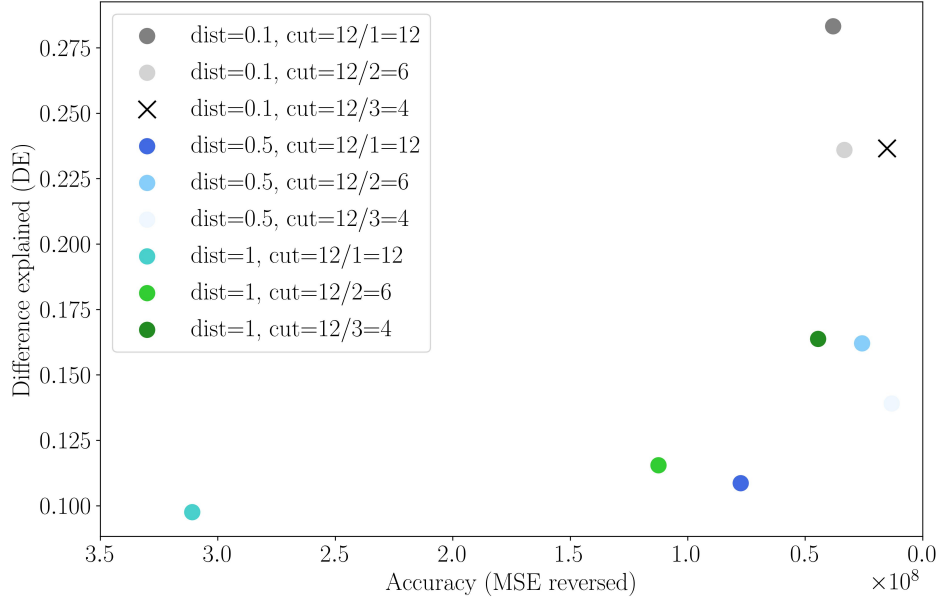


Figure A.1: We use brute-force optimization to tune the hyperparameters of the Shape finder algorithm. We use the 12 months in 2021 as validation data. More specifically, we use all data between 1989 and 2019 to create historical matches for the input sequences in 2020. Hyperparameters are optimized for predictions in 2021. The best set of hyperparameters is used to fit the final model, returning predictions for 2022 and 2023. The hyperparameter combination $dist = 0.1$ and $cut = 12/3 = 4$, highlighted with marker $\times$, has the best performance in the validation data, showing the highest DE and the lowest MSE.

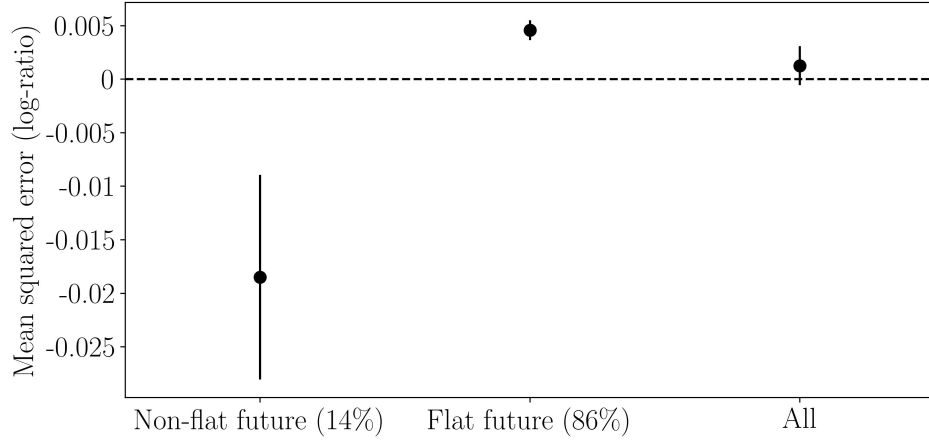


Figure A.2: We perform a robustness check for how we handle flat input sequences. More specifically, we demonstrate that returning flat predictions for flat input sequences does not change the performance of the ViEWS ensemble. We calculate the MSE log-ratio between ViEWS and the null model, taking the null model as baseline, $\log\left(\frac{mse_{ViEWS}}{mse_{Null}}\right)$. A positive value signifies that the null model performs better than ViEWS. For the majority of flat input sequence (86%), the next 12 months are flat as well. For those cases, using the null model instead of ViEWS significantly improves the prediction error. For flat input sequences with a non-flat future (14%), using the null model instead of ViEWS significantly worsens the MSE. Nevertheless, the prediction gain from ViEWS in comparison to the null model is arguably not substantial. This highlights the difficulty of anticipating conflict onsets. For the whole sample (all), using the null model instead of the ViEWS predictions does not significantly change performance.

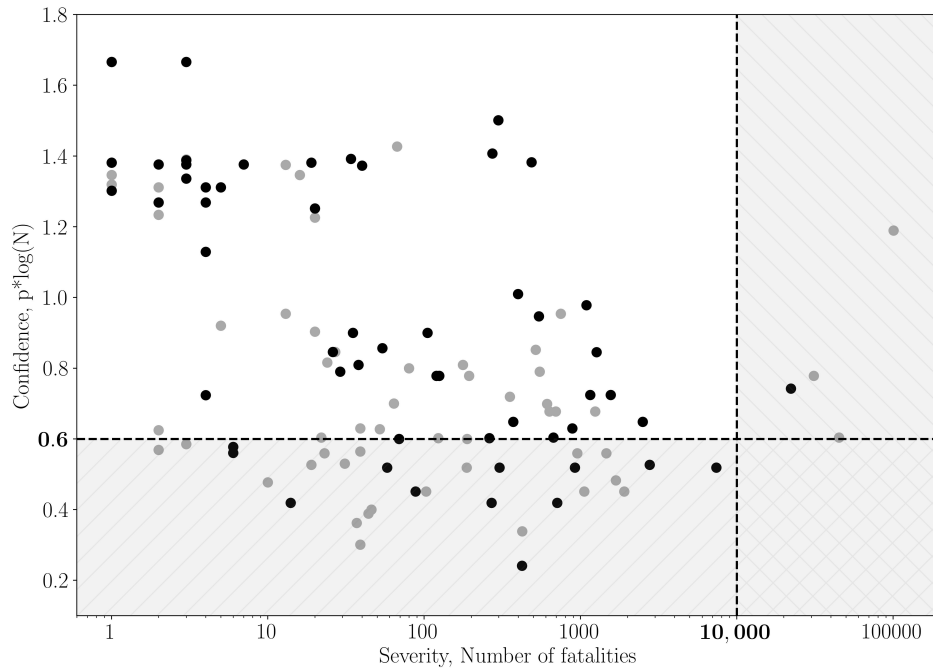


Figure A.3: Selection criteria for the compound model. The compound model uses the Shape Finder prediction for cases above the horizontal line and to the left of the vertical line, representing low- and medium-intensity cases with high confidence. For cases below the horizontal line and to the right of the vertical line, representing high-intensity or low-confidence cases, the ViEWS prediction is selected. The dashed lines reflect theoretical considerations: Shape Finder is preferred for high-confidence cases, while ViEWS is selected for high-intensity cases. Overall, 66% of the predictions are based on Shape Finder, and 33% on ViEWS. Black points indicate that the MSE log ratio is positive (Shape Finder performs better), while gray points indicate a negative ratio (Shape Finder performs worse).

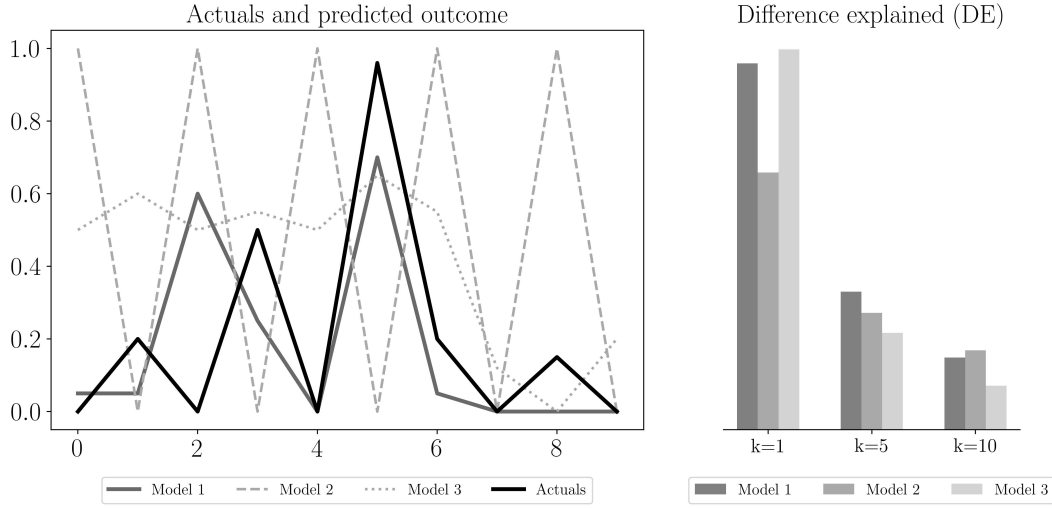


Figure A.4: Fictional data example illustrating how we select the balancing factor $k = 5$ in the difference explained (DE) equation. On the left, the actual outcome is shown in black. A risk-averse prediction is produced by model 3 (dotted line), a high variation prediction is produced by model 2 (dashed line) and model 1 yields the most accurate prediction (solid gray line). On the right the corresponding DE values for k equal 1, 5 and 10 are presented. If a low k factor is chosen ($k = 1$), the risk-averse model (model 3) has the best performance. Conversely, a high k factor ($k = 10$) suggests selecting the high variation model (model 2). The k value rewarding the most accurate model (model 1) is $k = 5$. Based on this experiment, we select 5 as balancing factor.

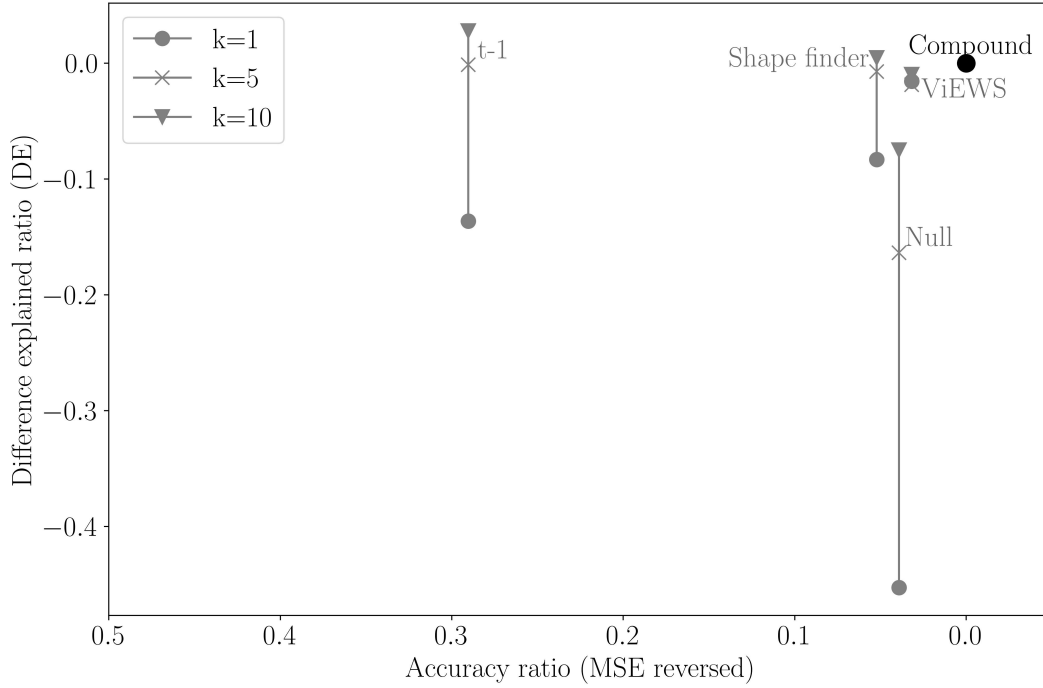


Figure A.5: We examine how varying the balancing factor (k) influences model performance. For the ViEWS ensemble, changing k does not affect the results. For the Shape Finder, $t - 1$ model, and the null model, decreasing k results in worse performance, while increasing k leads to better performance. However, the primary findings remain robust across different values of k .

Table A1: We validate whether the Shape finder overpredictions are due to matching low- with high-intensity cases. For this, we calculate the percentage of matches by scale, considering bins of fatalities <10 , $10-100$, $100-1000$, >1000 . For reference, we also present the scale of the input and the MSE log-ratio. Except Uzbekistan, more than 50% of matches are from cases with a lower intensity in comparison to the input scale. This supports the idea that the overpredictions occur due to matching high-intensity and low-intensity cases.

	Percentage of match by scale				Scale of input	MSE log-ratio
	<10	$10-100$	$100-1000$	>1000		
Afghanistan 2021	0.14	0.14	0.29	0.43	>1000	-4.62
Tajikistan 2022	0.10	0.60	0.10	0.20	$100-1000$	-4.02
South Sudan 2021	0.08	0.46	0.31	0.15	$100-1000$	-2.75
Uzbekistan 2022	0.40	0.60	0.00	0.00	$10-100$	-0.76
Iran 2023	0.62	0.38	0.00	0.00	$10-100$	-0.61

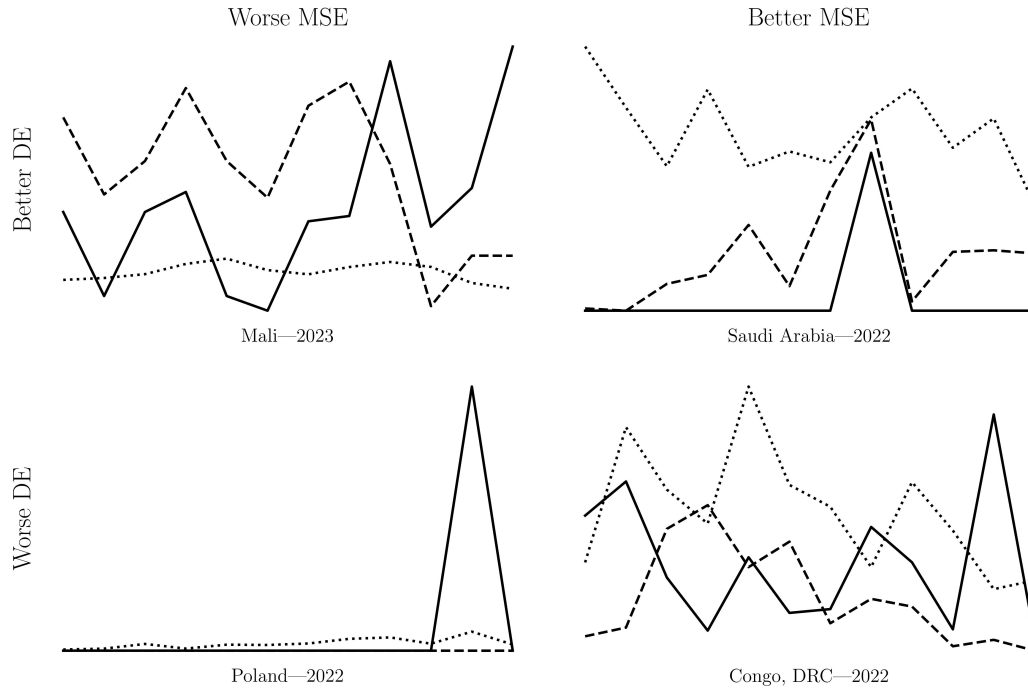


Figure A.6: These examples contrast the difference explained (DE) with the mean squared error (MSE). DE encourages risk-taking and MSE encourages risk-aversion. The black lines represent the true value, the dotted lines represent the ViEWS prediction, and the dashed lines the Shape finder results. The bottom row shows cases where Shape finder has a lower DE compared to ViEWS, and the top row shows cases where Shape finder is better. The same applies to MSE, with worse outcomes in the left column and better ones in the right column. Shape finder has a better DE when ViEWS produces flat predictions, while the MSE is worse in such instances. Shape finder has a lower MSE if ViEWS yields overpredictions, while the DE is worse in those cases.

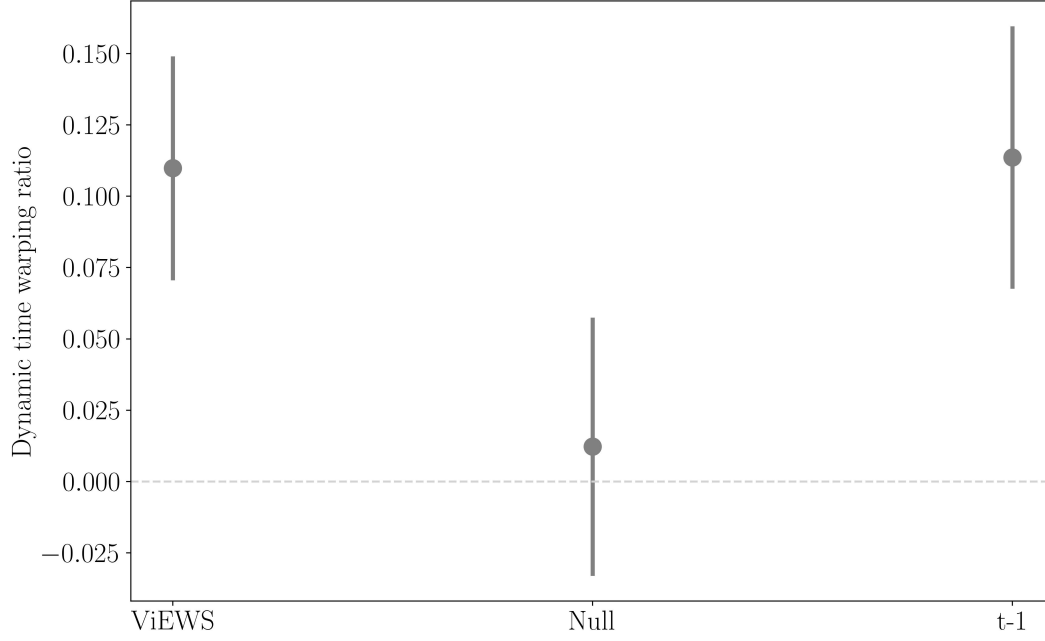


Figure A.7: We validate whether our findings hold when using the Dynamic Time Warping (DTW) distance as evaluation metric. For this purpose, we calculate the DTW ratio, comparing the Shape finder to the reference models, $\log \left(\frac{DTW_{\text{model}}}{DTW_{\text{Shape finder}}} \right)$. A positive value indicates that the Shape finder produces more accurate predictions, when comparing the distance between the actual future and the predicted time series using DTW. We find that the Shaper finder performs better than the ViEWS ensemble, and the $t - 1$ model, when judged by the DTW distance. The difference between Shape finder and the null model is positive, although not statistically significant.

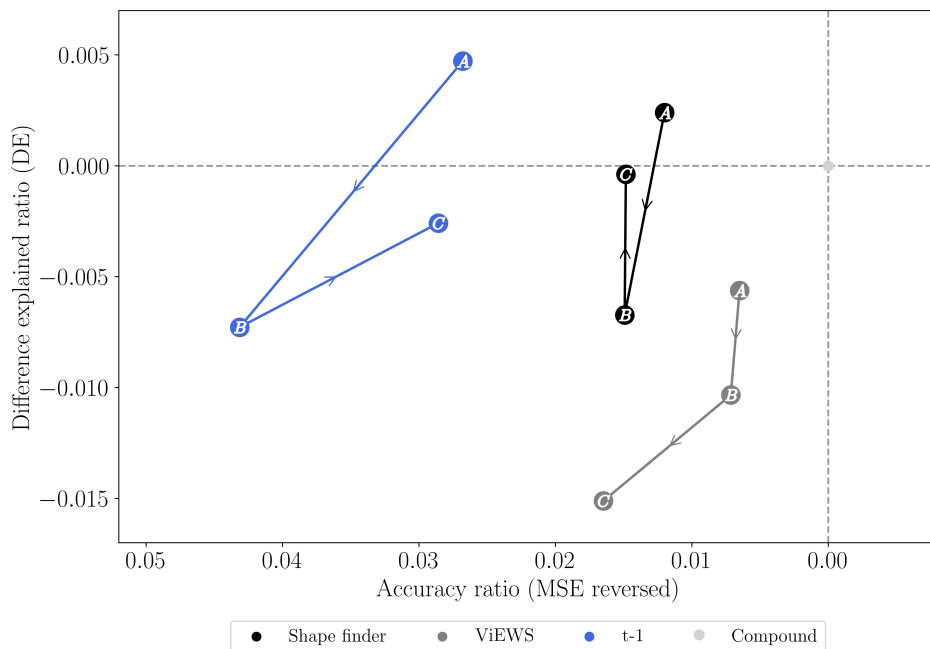


Figure A.8: Mean log-ratio for difference explained (DE) and mean squared error (MSE). The log-ratios are computed using the compound as base. For DE, a negative log-ratio signifies that the compound outperforms the reference model. Conversely, for MSE, a positive log-ratio indicates that the compound performs better. Validating the effect of increasing the forecasting horizon, we present results for 1–4 months ahead (A), 5–8 months ahead (B) and 9–12-months ahead (C). Our findings hold across prediction horizons. For better visibility, we exclude the null model, which has substantially larger log-ratios than the remaining models.

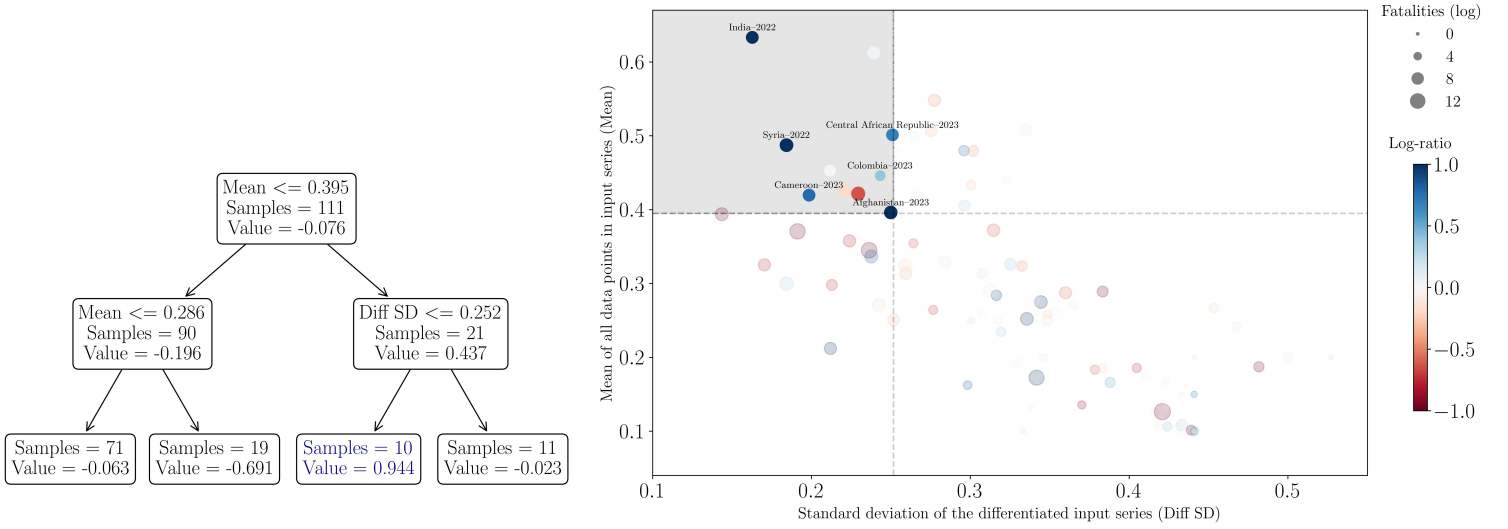


Figure A.9: Using a decision tree (left), we identify a cluster of well-performing cases. The decision tree takes the MSE log-ratio as output and as input a number of time series characteristics. Well-performing cases are located in the upper left quadrant of the scatter plot, indicated by the gray shading. Those are input sequences with a high mean and a low standard deviation of the differentiated input sequence. The scatter plot shows the standard deviation of the differentiated input sequences (x-axis) over the mean value of the input sequences (y-axis). The MSE log-ratio comparing ViEWS and Shape finder is represented in color. A blue shading signifies a positive log-ratio and thereby a more accurate Shape finder forecast. A red shading signals a negative log-ratio and thereby a more accurate ViEWS forecast. The size of each point refers to the log value of the number of fatalities in the forecasting window.

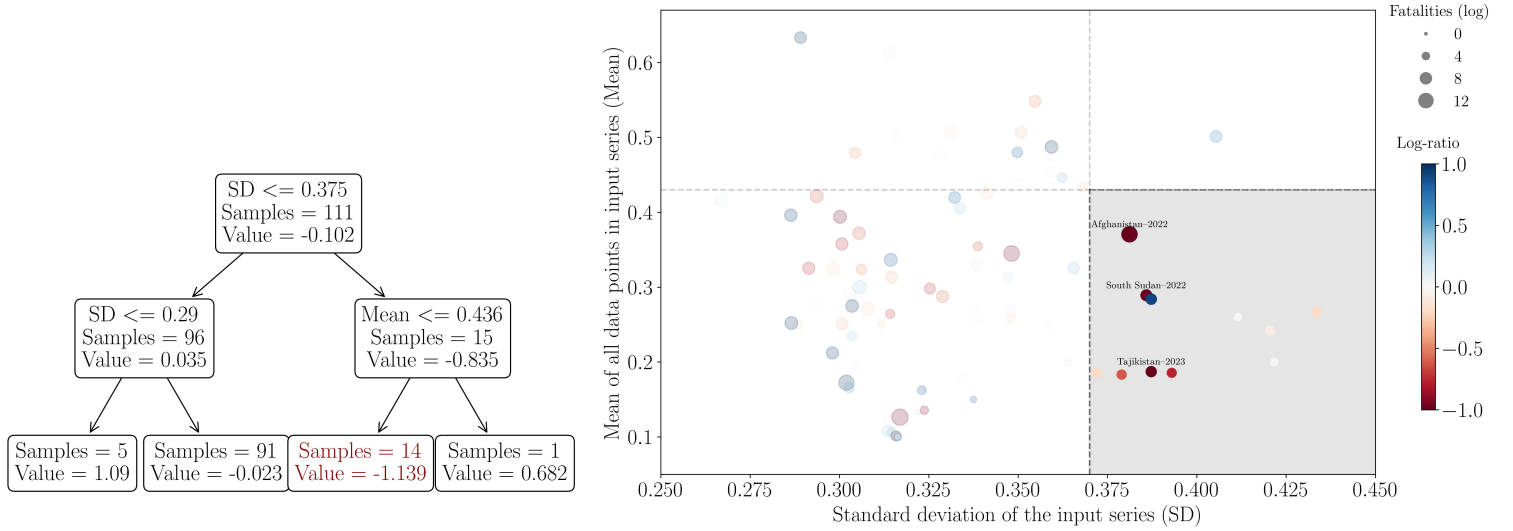


Figure A.10: Using a decision tree (left), we identify a cluster of poorly-performing cases. The decision tree takes the MSE log-ratio as output and as input a number of time series characteristics. Poorly-performing cases are located in the lower right quadrant of the scatter plot, indicated by the gray shading. Those are input sequences with a low mean and a high standard deviation. The scatter plot shows the standard deviation of the input sequences (x-axis) over the mean value of the input sequences (y-axis). The MSE log-ratio comparing ViEWS and Shape finder is represented in color. A blue shading signifies a positive log-ratio and thereby a more accurate Shape finder forecast. A red shading signals a negative log-ratio and thereby a more accurate ViEWS forecast. The size of each point refers to the log value of the number of fatalities in the forecasting window.

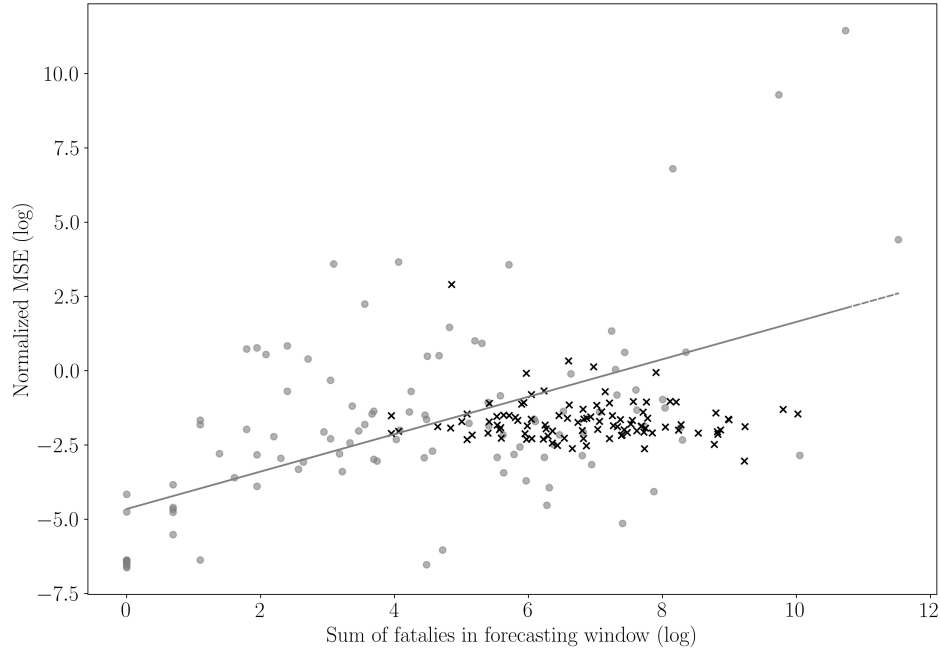


Figure A.11: We validate whether our observation regarding the characteristics of the best performing cases holds across the entire sample. The scatter plot shows the sum of the log-transformed fatalities in the forecasting window over the normalized mean squared error (MSE), likewise considering the log-transformed values. The gray points refer to all input sequences, and the gray line is the corresponding linear regression function. The black crosses are those cases that fit the criteria where the Shape finder performs well—high mean value and low standard deviation of the differentiated input. These cases have a systematically lower MSE, supporting our interpretation.

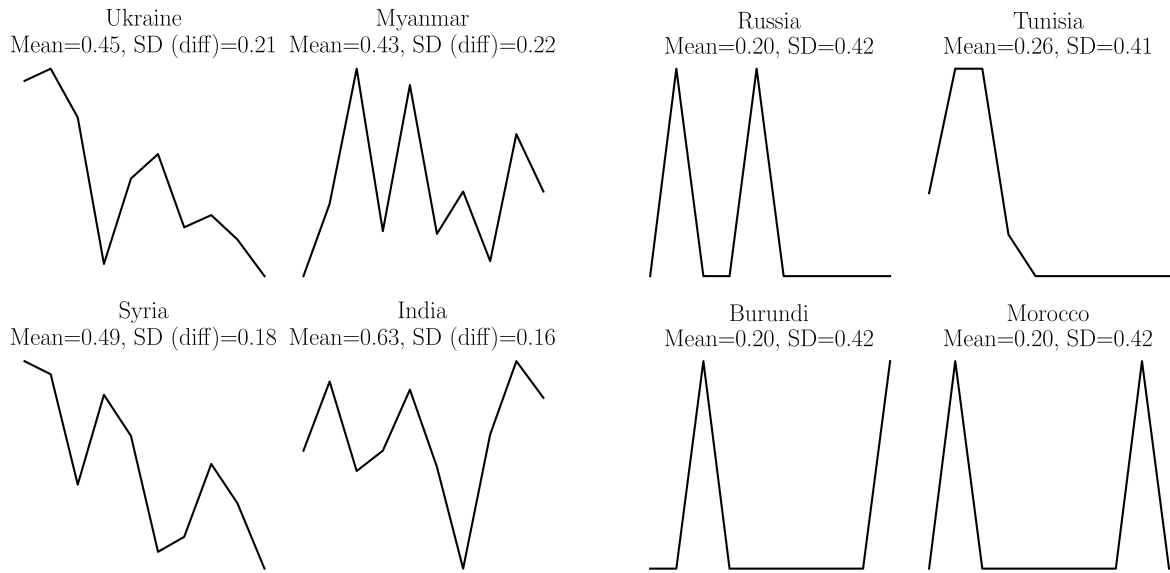


Figure A.12: We verify the time series characteristics of well- and poorly-performing cases against the categorization into low- and high-complexity cases. Time series with a high mean (*Mean*) and low standard deviation of the differentiated series (*SD (diff)*) are highly complex, showing multiple ups and downs in the number of fatalities (left panel). Time series with a low mean (*Mean*) and a high standard deviation (*SD*) are low complexity cases, showing single peaks amidst zero fatalities (right panel).

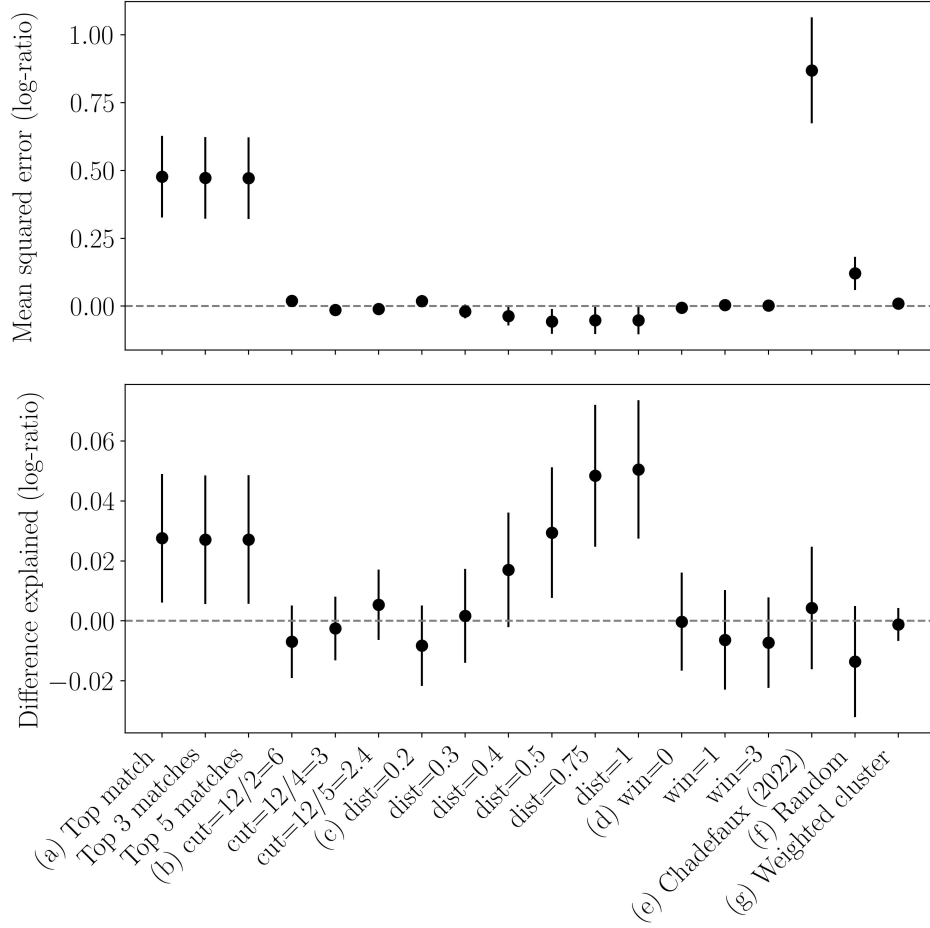


Figure A.13: We perform a range of robustness checks for a set of suggested modifications to the Shape finder. For each modification to the Shape finder algorithm, we report the log-ratio, $\log\left(\frac{m_{\text{Modified Shape finder}}}{m_{\text{Shape finder}}}\right)$, where $m \in \{\text{MSE}, \text{DE}\}$. The original Shape finder model is the baseline, which means that positive values signify that the original algorithm performs better than the modified one. (a) Instead of clustering futures, we select the future of the top closest match, or take the average future of the three or five closest matches. This modification deteriorates performance on both MSE and DE (b) Increasing or decreasing the cut-off value (*cut*) in the hierarchical clustering does not significantly change performance (c) Increasing the distance (*dist*) for the filtered reference repository either does not change performance significantly, or deteriorates performance on DE, while leading to only a small gain in MSE (d) Increasing or decreasing the window (*win*) of the reference subsequences does not significantly change the performance (e) The Shape finder performs significantly better than the shape-based approach presented in Chadeaux (2022), in particular in terms of MSE (f) Using a random selection of non-flat historical references instead of the filtered repository does significantly deteriorate performance when considering the prediction error (g) Using the weighted average of time series within the majority cluster as prediction, instead of the centroid, does not significantly change the performance. The weights are based on the distance to the input sequence.

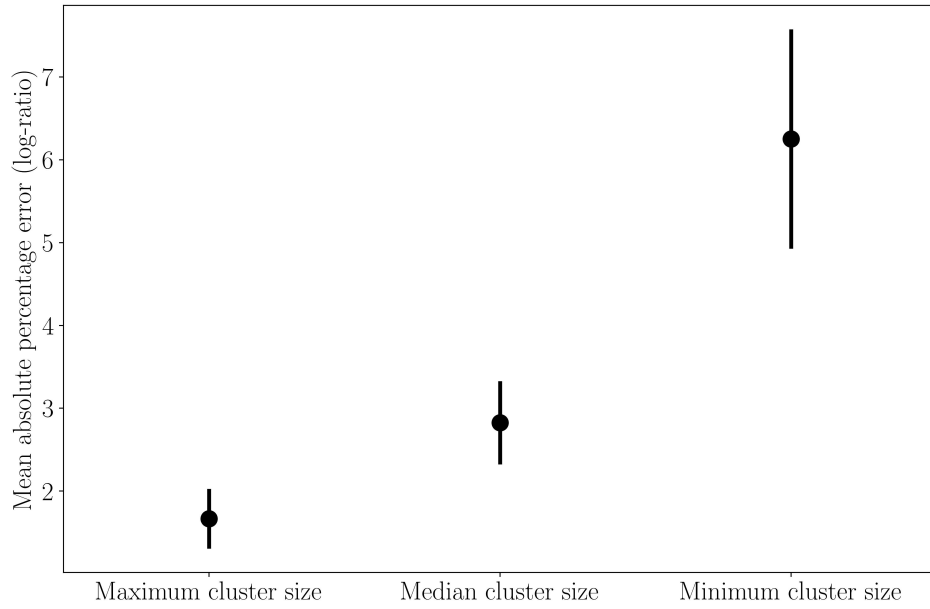


Figure A.14: We validate whether the size of the cluster is associated with the prediction performance. For each case, we select the prediction from the smallest cluster, the median cluster, and the majority cluster and compare the predictions to the actuals, by calculating the log-ratio mean absolute percentage error (MAPE). A larger value signals a higher prediction error, and a smaller MAPE suggests better performance. The Shape finder produces the most accurate predictions for the majority cluster, and the prediction error is largest when using the centroid of the smallest cluster as prediction. The smallest cluster contains outlier cases.

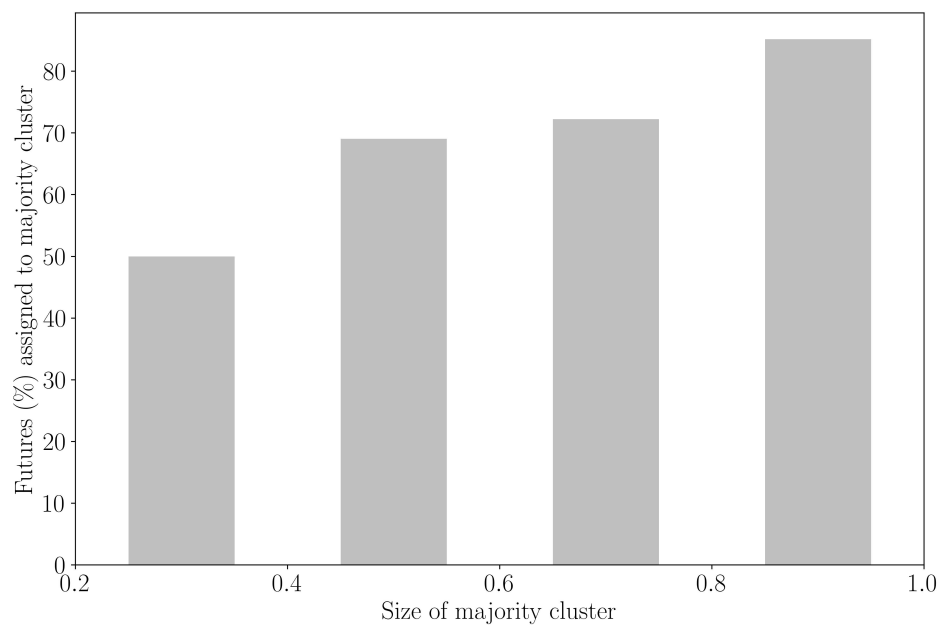
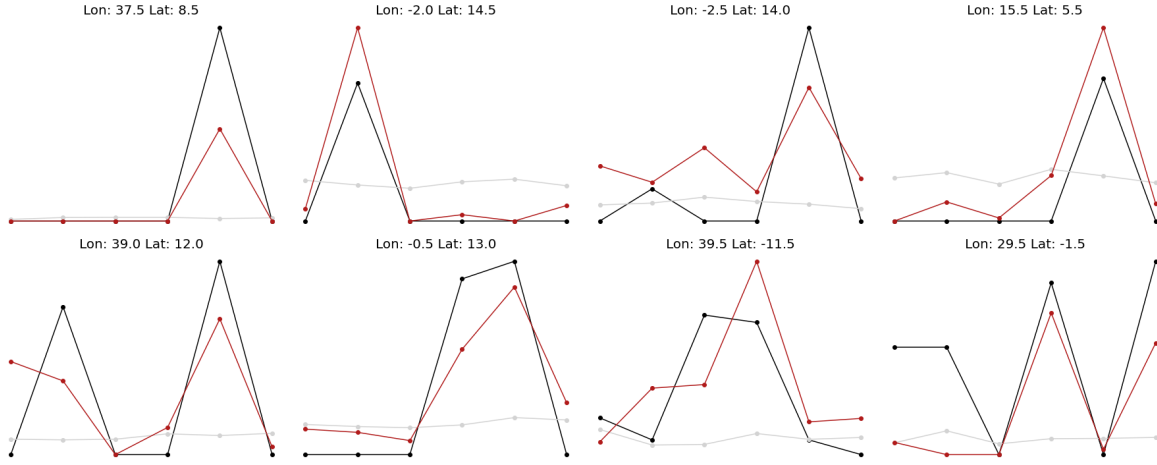


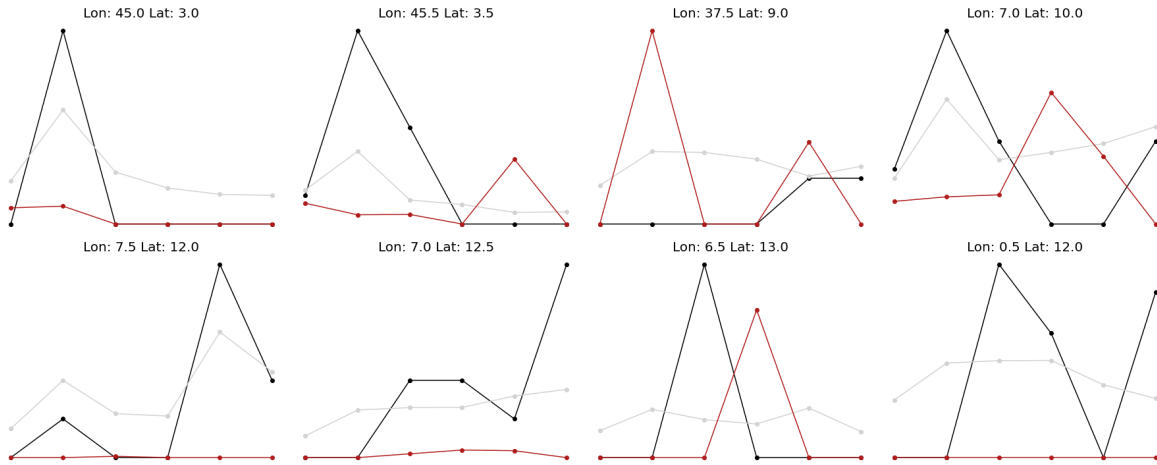
Figure A.15: We validate whether the Shape finder predictions are well-calibrated. More precisely, we calculate the percentage of actual futures that would be assigned to the majority cluster using the Shape finder matching approach, differentiated by the size of the majority cluster. As expected, the larger the size of the majority cluster, the higher the percentage of actual futures assigned to the centroid.

A.2 Appendix : Paper 2

Example of forecast : Prio-Grid level



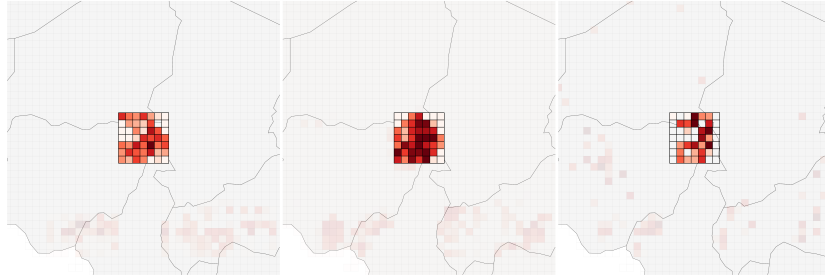
(a) Best performing cases



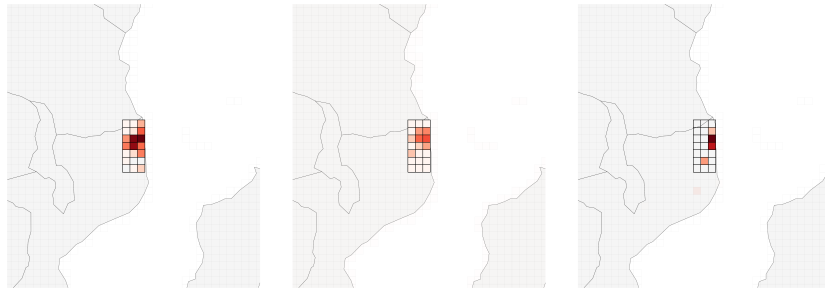
(b) Worse performing cases

Figure A.16: Examples of best (top) and worse (bottom) Prio-Grid MSE scores. The red line is the SF values, the grey is Views, and the black is the observed values.

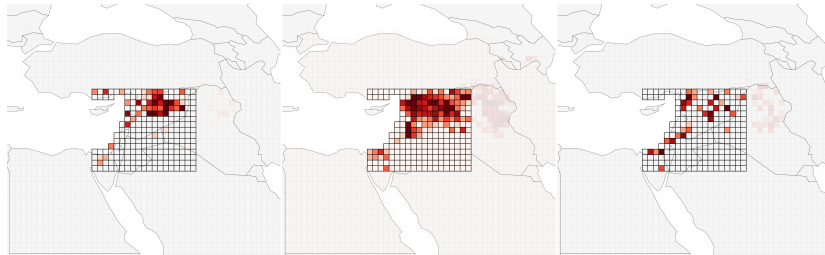
Example of forecast : Pattern level



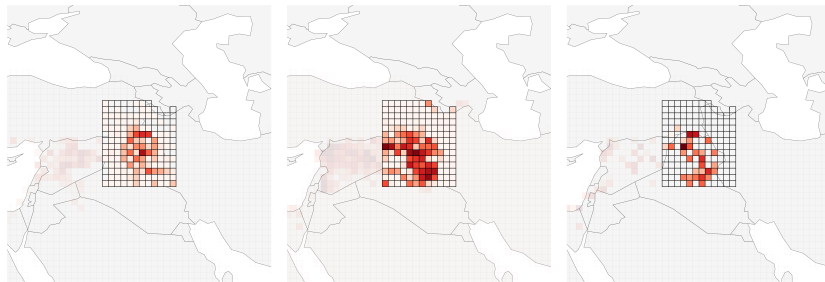
(a) Lake Chad area - Jan-June 2022



(b) North Mozambique - July-Dec 2023



(c) Levant Region - Jan-June 2022



(d) North Iraq - Jan-June 2022

Figure A.17: Examples of four Pattern ShapeFinder Forecast (left), Views Forecast (middle), and Observed values (right). The darker the red, the higher the fatality value. The values are aggregated per Prio-Grid over the 6-month period.

Under/over prediction robustness test

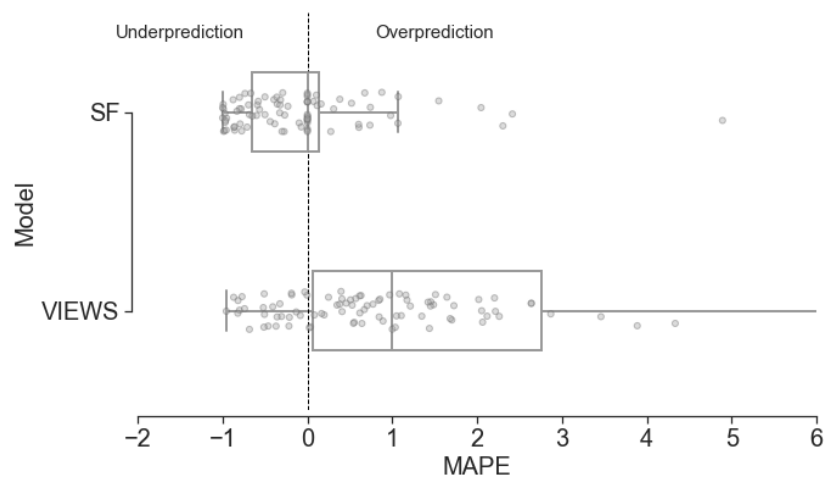


Figure A.18: Boxplot of the Mean Average Percentage Error (MAPE) for both models, SF and Views. A positive value corresponds to an overprediction, in contrast to a negative value.

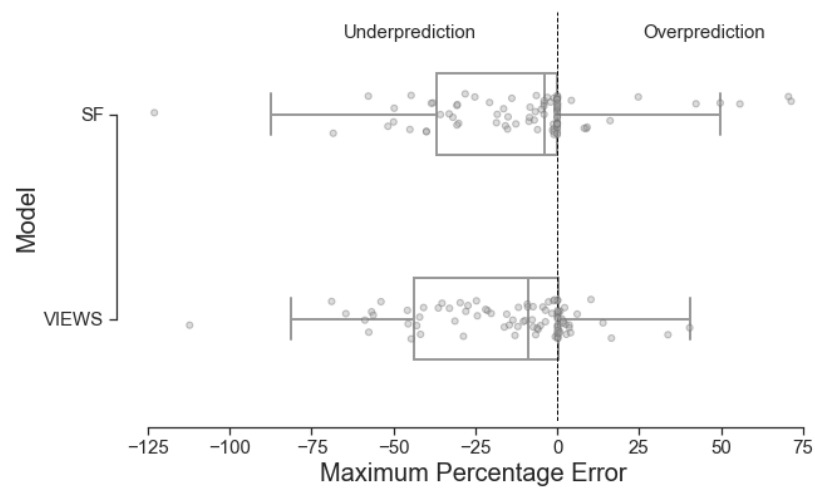


Figure A.19: Boxplot of the Maximum Percentage Error for both models, SF and Views. A positive value corresponds to an overprediction, in contrast to a negative value.

Kenya-Somalia example

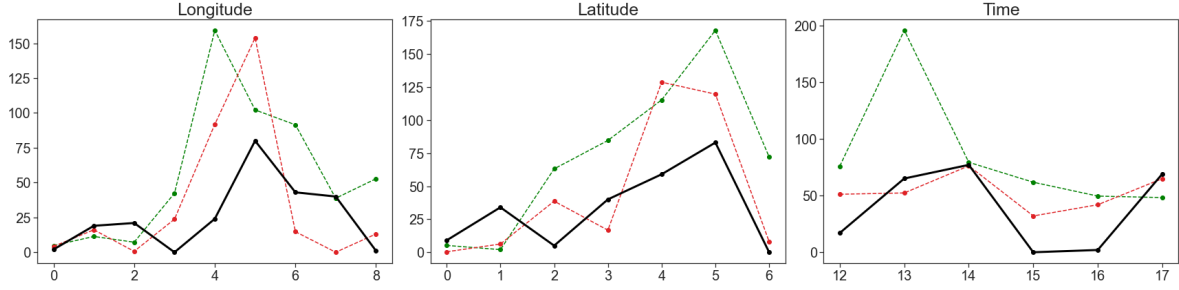


Figure A.20: Projection of fatalities values over the Longitude (left), Latitude (middle), and Time (right). The black line corresponds to the observed values, the green to the Views prediction, and the red to the ShapeFinder prediction.

Figure A.20 compares the forecasted projections of fatalities from the ShapeFinder (SF) model and the Views model against the actual observed values. The black line represents the observed data, while the green and red lines correspond to the Views and ShapeFinder predictions. Our model demonstrates a clear ability to capture the peak of fatalities accurately in both longitude and time dimensions. However, it misses the peak value in latitude by one unit, indicating a minor spatial misalignment. In contrast, the Views model exhibits a consistent overprediction, especially in the first two-month horizon.

Diffusion of fatalities to neighboring cells

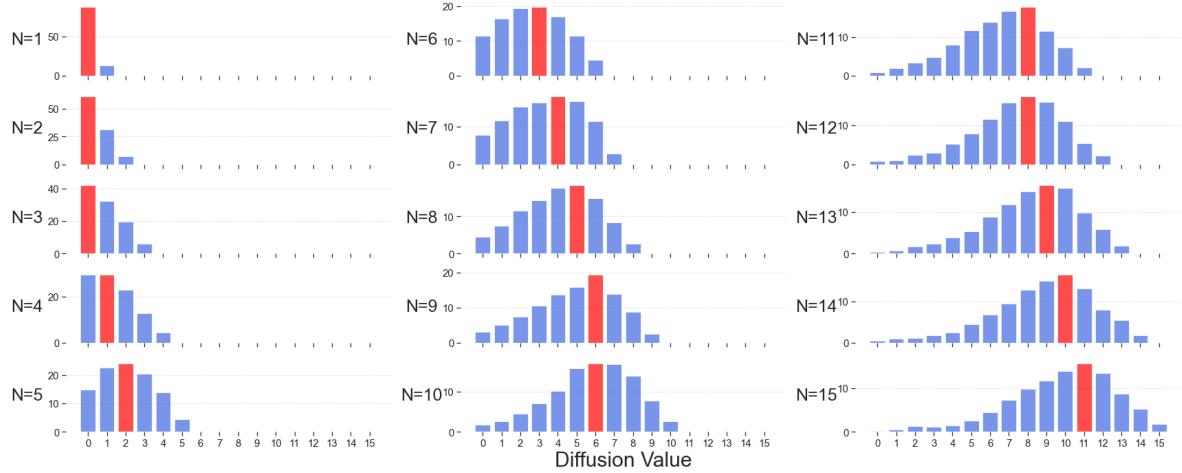


Figure A.21: Histogram of the number of active cells in the spatial neighborhood (5×5 grid) during the future window ($t+1$ to $t+6$), conditioned on the number of active neighboring cells during the past window ($t-12$ to t). Each panel corresponds to a different level of past activity ($N = 1$ to 15), where N is the number of active cells in the 5×5 neighborhood from $t-12$ to t . The red bar in each histogram indicates the most frequent outcome.

Grid example calculation

In this section, the calculations of the Spatial Distance, Earth Mover's Distance (EMD), and Euclidean distance of the example presented in Figure 11 are detailed. As a reminder, the patterns are represented in Figure A.22 with the cell values.

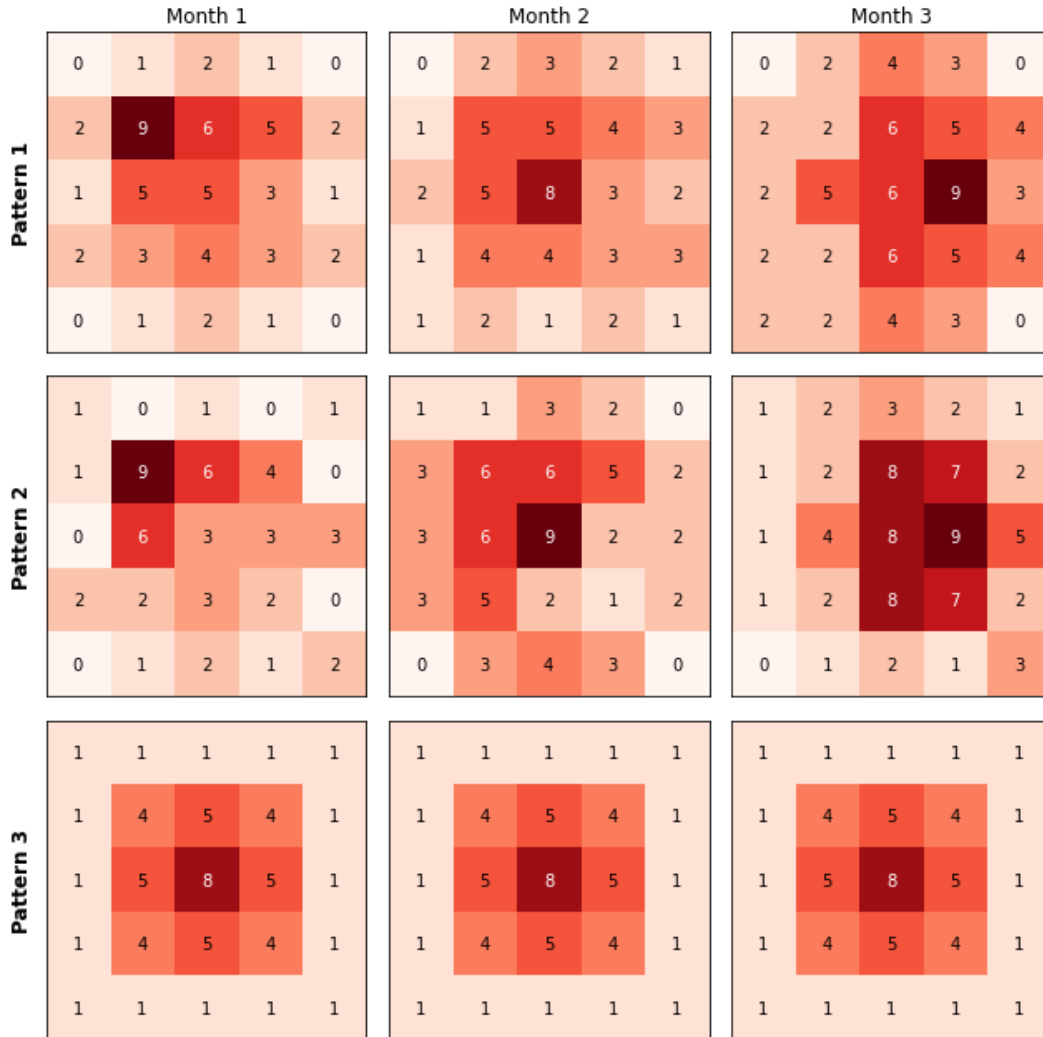


Figure A.22: Representation of the example shown in Figure 11. Three patterns on a 5×5 grid over three months. The x-axis represents longitude, and the y-axis represents latitude. The cell color indicates the number of fatalities per month, ranging from 0 (white) to 9 (dark red), with darker shades representing higher values. The number in the center of the cell indicates the value of the cell.

Spatial Distance

In classical spatial regressions, the model includes a spatial lag, with higher weights assigned to the closest neighbors. In our example, where the central cell is the point of interest, the weights would be:

$$weights = \begin{bmatrix} 0.01 & 0.05 & 0.1 & 0.05 & 0.01 \\ 0.05 & 0.3 & 0.5 & 0.3 & 0.05 \\ 0.1 & 0.5 & 1.0 & 0.5 & 0.1 \\ 0.05 & 0.3 & 0.5 & 0.3 & 0.05 \\ 0.01 & 0.05 & 0.1 & 0.05 & 0.01 \end{bmatrix}$$

In the matrix, the highest weight is placed at the center (cell [2,2] in the matrix), and the weights decrease as the distance from the central cell increases. These weights are proportional to the distance from the center of the grid.

Then, if we define the spatial distance SP_{P_1, P_2} between pattern 1 and pattern 2 with:

$$SP_{P_1, P_2} = \sum_{month} \left| \sum_{lat, long} P_{1, lat, long, m} * weights_{lat, long} - \sum_{lat, long} P_{2, lat, long, m} * weights_{lat, long} \right|$$

Then :

$$\alpha = P_{1,lat,long,m} * weights_{lat,long}$$

$$= \begin{bmatrix} \begin{bmatrix} 0.00 & 0.05 & 0.20 & 0.05 & 0.00 \\ 0.10 & 2.70 & 3.00 & 1.50 & 0.10 \\ 0.10 & 2.50 & 5.00 & 1.50 & 0.10 \\ 0.10 & 0.90 & 2.00 & 0.90 & 0.10 \\ 0.00 & 0.05 & 0.20 & 0.05 & 0.00 \end{bmatrix}, \begin{bmatrix} 0.00 & 0.10 & 0.30 & 0.10 & 0.01 \\ 0.05 & 1.50 & 2.50 & 1.20 & 0.15 \\ 0.20 & 2.50 & 8.00 & 1.50 & 0.20 \\ 0.05 & 1.20 & 2.00 & 0.90 & 0.15 \\ 0.01 & 0.10 & 0.10 & 0.10 & 0.01 \end{bmatrix}, \begin{bmatrix} 0.00 & 0.10 & 0.40 & 0.15 & 0.00 \\ 0.10 & 0.60 & 3.00 & 1.50 & 0.20 \\ 0.20 & 2.50 & 6.00 & 4.50 & 0.30 \\ 0.10 & 0.60 & 3.00 & 1.50 & 0.20 \\ 0.02 & 0.10 & 0.40 & 0.15 & 0.00 \end{bmatrix} \end{bmatrix}$$

$$\beta = P_{2,lat,long,m} * weights_{lat,long}$$

$$= \begin{bmatrix} \begin{bmatrix} 0.01 & 0.00 & 0.10 & 0.00 & 0.01 \\ 0.05 & 2.70 & 3.00 & 1.20 & 0.00 \\ 0.00 & 3.00 & 3.00 & 1.50 & 0.30 \\ 0.10 & 0.60 & 1.50 & 0.60 & 0.00 \\ 0.00 & 0.05 & 0.20 & 0.05 & 0.02 \end{bmatrix}, \begin{bmatrix} 0.01 & 0.05 & 0.30 & 0.10 & 0.00 \\ 0.15 & 1.80 & 3.00 & 1.50 & 0.10 \\ 0.30 & 3.00 & 9.00 & 1.00 & 0.20 \\ 0.15 & 1.50 & 1.00 & 0.30 & 0.10 \\ 0.00 & 0.15 & 0.40 & 0.15 & 0.00 \end{bmatrix}, \begin{bmatrix} 0.01 & 0.10 & 0.30 & 0.10 & 0.01 \\ 0.05 & 0.60 & 4.00 & 2.10 & 0.10 \\ 0.10 & 2.00 & 8.00 & 4.50 & 0.50 \\ 0.05 & 0.60 & 4.00 & 2.10 & 0.10 \\ 0.00 & 0.05 & 0.20 & 0.05 & 0.03 \end{bmatrix} \end{bmatrix}$$

So:

$$\begin{aligned} SP_{P_1,P_2} &= \sum_{month} | \sum \alpha_{month} - \sum \beta_{month} | \\ &= 3.21 + 1.33 + 4.03 = 8.57 \end{aligned}$$

Similarly for SP_{P_1, P_3} ,

$$\delta = P_{3,lat,long,m} * weights_{lat,long}$$

$$= \begin{bmatrix} \begin{bmatrix} 0.01 & 0.05 & 0.10 & 0.05 & 0.01 \\ 0.05 & 1.20 & 2.50 & 1.20 & 0.05 \\ 0.10 & 2.50 & 8.00 & 2.50 & 0.10 \\ 0.05 & 1.20 & 2.50 & 1.20 & 0.05 \\ 0.01 & 0.05 & 0.10 & 0.05 & 0.01 \end{bmatrix}, \begin{bmatrix} 0.01 & 0.05 & 0.10 & 0.05 & 0.01 \\ 0.05 & 1.20 & 2.50 & 1.20 & 0.05 \\ 0.10 & 2.50 & 8.00 & 2.50 & 0.10 \\ 0.05 & 1.20 & 2.50 & 1.20 & 0.05 \\ 0.01 & 0.05 & 0.10 & 0.05 & 0.01 \end{bmatrix}, \begin{bmatrix} 0.01 & 0.05 & 0.10 & 0.05 & 0.01 \\ 0.05 & 1.20 & 2.50 & 1.20 & 0.05 \\ 0.10 & 2.50 & 8.00 & 2.50 & 0.10 \\ 0.05 & 1.20 & 2.50 & 1.20 & 0.05 \\ 0.01 & 0.05 & 0.10 & 0.05 & 0.01 \end{bmatrix} \end{bmatrix}$$

So:

$$\begin{aligned} SP_{P_1, P_3} &= \sum_{month} |\sum \alpha_{month} - \sum \delta_{month}| \\ &= 2.44 + 0.71 + 1.98 = 5.13 \end{aligned}$$

Earth Mover's Distance

The Earth Mover's Distance (EMD) is computed as follows:

$$\text{EMD} = \sum_{i,j} \gamma_{i,j}^* * c_{i,j}$$

where $\gamma_{i,j}^*$ is the Optimal Transport Matrix or the matrix that lowers the transport cost. $c_{i,j}$ is the cost matrix of the Euclidean distance between the points i and j .

The Optimal Transport Matrix is computed from the “Source” distribution called w_1 (i.e., the normalized fatalities of pattern 1) and the “Target” distribution w_2 (i.e., the normalized fatalities of pattern 2).

The normalized patterns are computed as:

$$P_1^{\text{norm}} = \frac{P_1}{\sum P_1}, \quad P_2^{\text{norm}} = \frac{P_2}{\sum P_2}$$

This gives us the following points :

Pattern 1				Pattern 2			
Longitude	Latitude	Month	$Fatatlities_{norm}$	Longitude	Latitude	Month	$Fatatlities_{norm}$
0.0000	0.0000	1.0000	0.0047	0.0000	0.0000	0.0000	0.0048
0.0000	0.0000	2.0000	0.0094	0.0000	0.0000	2.0000	0.0048
0.0000	0.0000	3.0000	0.0047	0.0000	0.0000	4.0000	0.0048
0.0000	1.0000	0.0000	0.0094	0.0000	1.0000	0.0000	0.0048
0.0000	1.0000	1.0000	0.0425	0.0000	1.0000	1.0000	0.0429
0.0000	1.0000	2.0000	0.0283	0.0000	1.0000	2.0000	0.0286
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2.0	4.0	2.0	0.0189	2.0	4.0	3.0	0.0048
2.0	4.0	3.0	0.0142	2.0	4.0	4.0	0.0143

After computation of the optimal transport matrix, we obtain :

$$\gamma_{i,j}^* = \begin{bmatrix} 0.0046 & 0.0 & 0.0 & \dots & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0048 & 0.0 & \dots & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0046 & \dots & 0.0 & 0.0 & 0.0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0.0 & 0.0 & 0.0 & \dots & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & \dots & 0.0095 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & \dots & 0.0 & 0.0048 & 0.009 \end{bmatrix}$$

Then, the Euclidean distance between the coordinates of both patterns is calculated using:

$$d_{i,j} = \sqrt{(long_i - long_j)^2 + (lat_i - lat_j)^2 + (month_i - month_j)^2}$$

This results in a cost matrix $c_{i,j}$ of shape (n_1, n_2) , where n_1 is the number of non-zero cell in pattern1, and n_2 the number of non-zero cell in pattern2.

$$c_{i,j} = \begin{bmatrix} 1.0 & 1.0 & 3.0 & \dots & 4.5826 & 4.899 & 5.3852 \\ 2.0 & 0.0 & 2.0 & \dots & 4.4721 & 4.5826 & 4.899 \\ 3.0 & 1.0 & 1.0 & \dots & 4.5826 & 4.4721 & 4.5826 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 4.5826 & 4.5826 & 5.3852 & \dots & 1.0 & 2.0 & 3.0 \\ 4.899 & 4.4721 & 4.899 & \dots & 0.0 & 1.0 & 2.0 \\ 5.3852 & 4.5826 & 4.5826 & \dots & 1.0 & 0.0 & 1.0 \end{bmatrix}$$

So:

$$\begin{aligned}
EMD_{P_1, P_2} &= \sum_{i,j} \gamma_{i,j}^* * c_{i,j} \\
&= (0.0046 \times 1.0) + (0.0046 \times 1.0) + \dots + (0.009 \times 1.0) \\
&= 0.2290
\end{aligned}$$

Similarly, we repeat the operation with P_1 and P_3 .

$$\gamma_{i,k}^* = \begin{bmatrix} 0.0 & 0.0047 & 0.0 & \dots & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0008 & 0.0056 & \dots & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & \dots & 0.0 & 0.0 & 0.0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0.0 & 0.0 & 0.0 & \dots & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & \dots & 0.0056 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & \dots & 0.0 & 0.0056 & 0.0 \end{bmatrix}$$

$$c_{i,k} = \begin{bmatrix} 1.0 & 0.0 & 1.0 & \dots & 4.5826 & 4.899 & 5.3852 \\ 2.0 & 1.0 & 0.0 & \dots & 4.4721 & 4.5826 & 4.899 \\ 3.0 & 2.0 & 1.0 & \dots & 4.5826 & 4.4721 & 4.5826 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 4.5826 & 4.4721 & 4.5826 & \dots & 1.0 & 2.0 & 3.0 \\ 4.899 & 4.5826 & 4.4721 & \dots & 0.0 & 1.0 & 2.0 \\ 5.3852 & 4.899 & 4.5826 & \dots & 1.0 & 0.0 & 1.0 \end{bmatrix}$$

$$\begin{aligned}
EMD_{P_1, P_3} &= \sum_{i, k} \gamma_{i, k}^* * c_{i, k} \\
&= 0.2984
\end{aligned}$$

Finally,

$$\begin{cases} EMD_{P_1, P_2} < EMD_{P_1, P_3}, \\ SP_{P_1, P_2} > SP_{P_1, P_3} \end{cases}$$

Euclidean Distance

The Euclidean Distance between P_1 and P_2 , here called ed is calculated as:

$$A = P_1 - P_2$$

$$= \left[\begin{bmatrix} -1 & 1 & 1 & 1 & -1 \\ 1 & 0 & 0 & 1 & 2 \\ 1 & -1 & 2 & 0 & -2 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & -2 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 0 & 0 & 1 \\ -2 & -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & 1 & 0 \\ -2 & -1 & 2 & 2 & 1 \\ 1 & -1 & -3 & -1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 1 & 1 & -1 \\ 1 & 0 & -2 & -2 & 2 \\ 1 & 1 & -2 & 0 & -2 \\ 1 & 0 & -2 & -2 & 2 \\ 2 & 1 & 2 & 2 & -3 \end{bmatrix} \right]$$

Taking the absolute values:

$$|A| = \left[\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 2 \\ 1 & 1 & 2 & 0 & 2 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 2 & 1 & 2 & 2 & 1 \\ 1 & 1 & 3 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 2 \\ 1 & 1 & 2 & 0 & 2 \\ 1 & 0 & 2 & 2 & 2 \\ 2 & 1 & 2 & 2 & 3 \end{bmatrix} \right]$$

$$ed_{P_1, P_2} = \sum |A| = 84$$

Similarly, for P_1 and P_3 :

$$B = P_1 - P_3$$

$$= \left[\begin{bmatrix} -1 & 0 & 1 & 0 & -1 \\ 1 & 5 & 1 & 1 & 1 \\ 0 & 0 & -3 & -2 & 0 \\ 1 & -1 & -1 & -1 & 1 \\ -1 & 0 & 1 & 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 \\ 1 & 0 & 0 & -2 & 1 \\ 0 & 0 & -1 & -1 & 2 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 3 & 2 & -1 \\ 1 & -2 & 1 & 1 & 3 \\ 1 & 0 & -2 & 4 & 2 \\ 1 & -2 & 1 & 1 & 3 \\ 1 & 1 & 3 & 2 & -1 \end{bmatrix} \right]$$

Taking the absolute values:

$$|B| = \left[\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 5 & 1 & 1 & 1 \\ 0 & 0 & 3 & 2 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 \\ 1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 3 & 2 & 1 \\ 1 & 2 & 1 & 1 & 3 \\ 1 & 0 & 2 & 4 & 2 \\ 1 & 2 & 1 & 1 & 3 \\ 1 & 1 & 3 & 2 & 1 \end{bmatrix} \right]$$

$$ed_{P_1, P_3} = \sum |B| = 84$$

So:

$$ed_{P_1, P_2} = ed_{P_1, P_3}$$

A.3 Appendix : Paper 3

ACF Curve

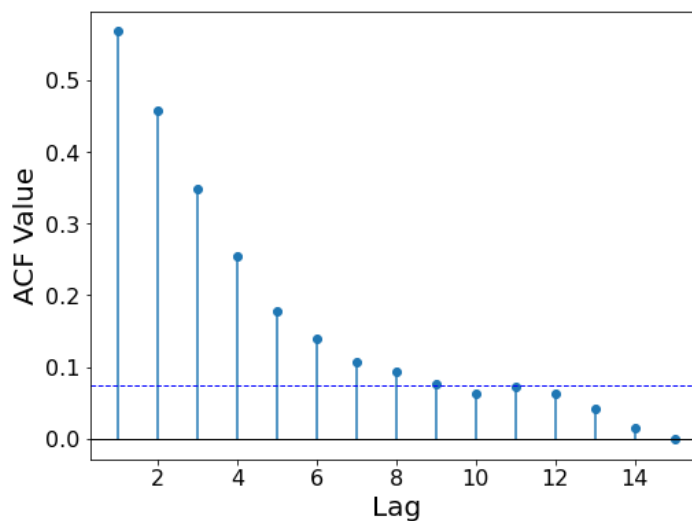


Figure A.23: Autocorrelation Function (ACF) plot, which measures the correlation of the time series with itself at different lags of the migration flow data.

Optimization of the hyperparameters of the Random Forest models

Parameter	Possible Values
n_estimators	{10, 230, 450, 670, 890, 1110, 1330, 1550, 1770, 2000}
max_features	{‘auto’, ‘sqrt’}
max_depth	{10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None}
min_samples_split	{2, 5, 10}
min_samples_leaf	{1, 2, 4}
bootstrap	{True, False}

Table A2: Hyperparameter Search Space for Random Forest models.

Push-Pull Covariates Transformation Test

The five covariates were tested in transformed forms, including differentiation, lagging, and logarithmic transformation. The Table A3 shows the transformation applied to the different covariates.

Covariate	Raw	Logged	Lagged	Squared
Adult unemployment rate	✓	✓	✓	
IDP (natural disasters)	✓	✓	✓	✓
IDP (conflict)	✓	✓	✓	✓
Foreign-born employment/jobs	✓	✓		✓
Diaspora / Migrant stock	✓	✓		

Table A3: Covariates and the transformations tested for each. A checkmark indicates that the given transformation is applied to the covariate.

Ordinary Least Squares (OLS) regression was fitted on the full dataset using all combinations of transformations of the five covariates as independent variables, with migration flows as the dependent variable. The best fit ($R^2=0.0653618$) was obtained with the following transformations, in the order of Table A3: logged, logged, logged, squared, and raw. Using all covariates in their raw form resulted in an R^2 of 0.0638194. Although the improvements are small, the version of each covariate that provides the highest adjusted R^2 is retained for the analysis. All five covariates have statistically significant coefficients at the 99% confidence level.

Models parameters search

SF-COV1 Model

Figure A.24 shows the log ratio of MSE of SF-COV1 over the SF for several weighting vectors α from equation (1). The vector is composed like this :

$$vec_{\alpha} = [\alpha_m, \alpha_{um}, \alpha_{dd}, \alpha_{cd}, \alpha_{dia}, \alpha_{ja}]$$

With m the migration variable, um the unemployment rate, dd the disaster displacement, cd the conflict displacement, dia the diaspora, and ja the jobs availability. The first notable observation is that any combination including all covariates (where $\forall \alpha_{cov} > 0$) consistently performs worse than the original ShapeFinder (SF). Among individual covariates, the two that yield the best performance (where $\alpha_m = 1$ and $\alpha_{cov} = 1$ with all other weights set to 0) are Diaspora and Natural disaster displacement. These findings align with the results presented in Figure 27. The optimal combination of weights involves the inclusion of the unemployment rate and natural disaster displacement, but this improvement remains statistically insignificant.

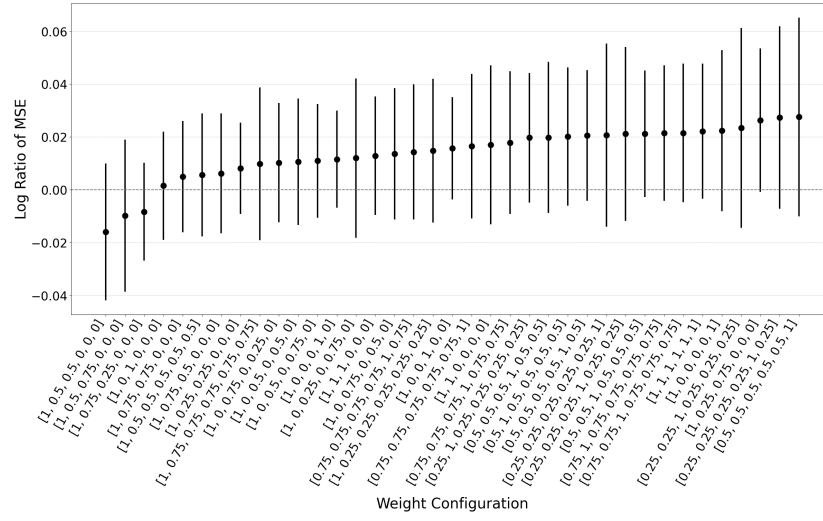


Figure A.24: The log ratio of MSE of SF-COV1 over the SF for several weighting vector α from equation (1). The vector is composed like this : $[\alpha_m, \alpha_{um}, \alpha_{dd}, \alpha_{cd}, \alpha_{dia}, \alpha_{ja}]$ with m the migration variable, um the unemployment rate, dd the disaster displacement, cd the conflict displacement, dia the diaspora and ja the jobs availability. The point represents the mean value, and the line is the confidence interval at 95% confidence.

SF-COV2 Model

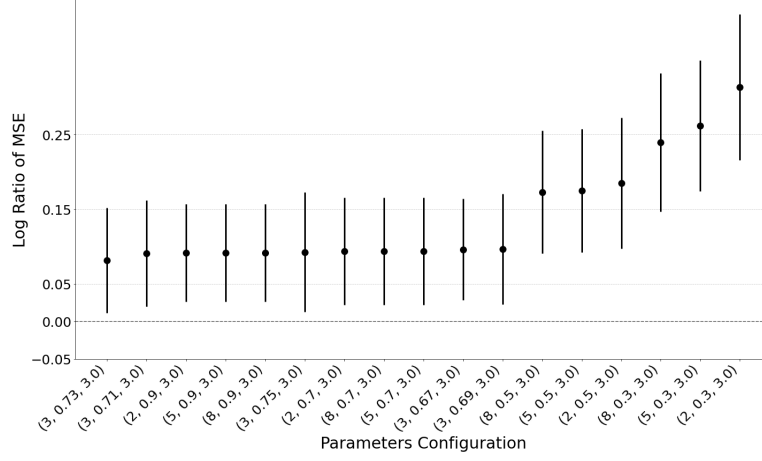


Figure A.25: Log Ratio for the RF-COV2 model compared with SF model for a set of parameters of the minimum matches, the threshold distance to classify sequences as similar, and the clustering distance $[min_{mat}, min_{dist}, clu_{dist}]$

SF-COV3 Model

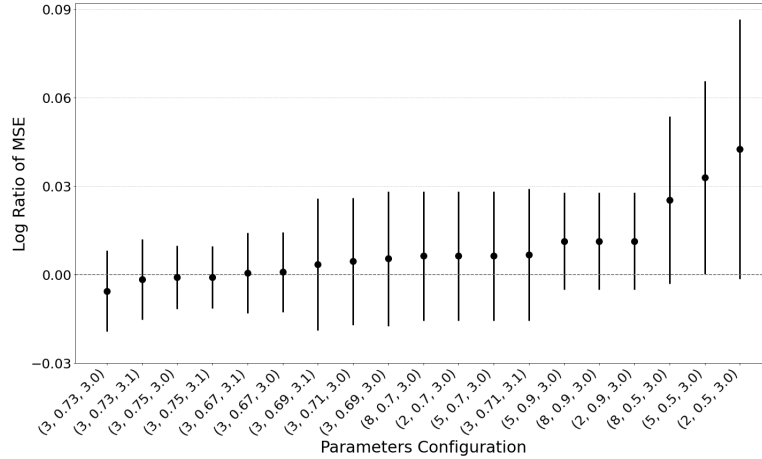


Figure A.26: Log Ratio for the RF-COV3 model compared with SF model for a set of parameters of the minimum matches, the threshold distance to classify sequences as similar, and the clustering distance $[min_{mat}, min_{dist}, clu_{dist}]$

A.4 Appendix : Paper 4

Dynamic Time Warping (DTW): Technical Details

Dynamic Time Warping (DTW) is an algorithm used for measuring the similarity between two temporal sequences which may vary in speed. Originally developed for speech recognition, it has since found applications in a wide range of fields including data mining and time series analysis.

The fundamental principle of DTW is to compare two time series by aligning their indices in a way that minimizes the cumulative distance between them. This comparison is not restricted to matching equivalent time points, allowing DTW to effectively handle sequences that unfold at different rates.

Consider two time series $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_m\}$. DTW aims to find an alignment between these two series that minimizes the total distance. The DTW distance between X and Y is defined as:

$$DTW(X, Y) = \min \sqrt{\sum_{k=1}^K d(w_k)^2} \quad (12)$$

where w_k are the elements of the warping path W , a set of pairs that define a mapping between X and Y , and $d(w_k)$ is the distance between the paired points in X and Y .

The warping path W is typically computed using dynamic programming. The key idea is to build a matrix D where each element $D[i, j]$ represents the distance between x_i and y_j plus the minimum cumulative distance to reach $D[i, j]$ from the starting point $(1, 1)$. The optimal path is then traced back from $D[n, m]$ to $D[1, 1]$.

The distance matrix D is defined as:

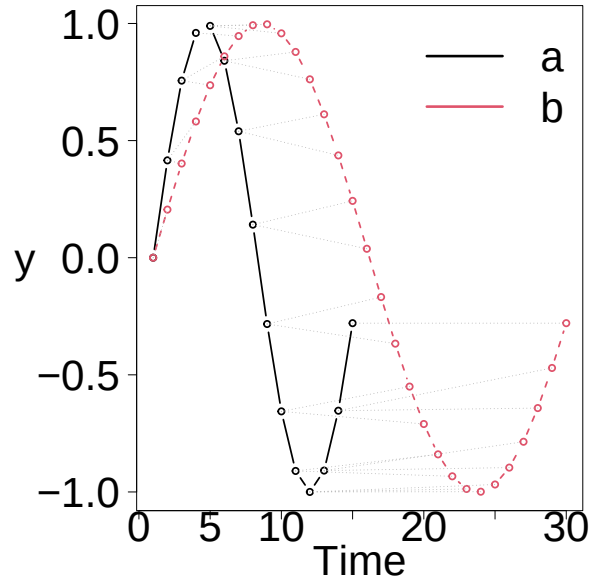


Figure A.27: Dynamic Time Warping

$$D[i, j] = d(x_i, y_j) + \min(D[i - 1, j], D[i, j - 1], D[i - 1, j - 1]) \quad (13)$$

where $d(x_i, y_j)$ is the distance between x_i and y_j , often calculated using the Euclidean distance.

Long Short-Term Memory (LSTM) Network: Detailed Overview

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) particularly suited for learning from sequences of data. They are widely used for their ability to capture long-term dependencies in sequence data, which traditional RNNs often fail to do.

LSTMs have been successfully applied in various domains such as natural language processing, time series forecasting, and sequence generation, showcasing their versatility and effectiveness in handling sequential data with complex temporal dependencies. We present here a very short introduction to this type of neural network. For a more comprehensive review of LSTMs, see Hochreiter and Schmidhuber (1997).

The LSTM network consists of a unique structure with memory cells, as presented in Fig. A.28, and multiple gates controlling the flow of information. The mathematical equations defining the LSTM architecture are as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (\text{Forget gate activation})$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (\text{Input gate activation})$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (\text{Cell input activation})$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (\text{Cell state})$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (\text{Output gate activation})$$

$$h_t = o_t \cdot \tanh(C_t) \quad (\text{Hidden state})$$

Here, f_t , i_t , and o_t represent the activations of the forget gate, input gate, and output gate, respectively. $\tilde{C}_t$ is the cell input activation, C_t is the cell state, and h_t is the hidden state. W_f , W_i , W_C , W_o , b_f , b_i , b_C , and b_o are the respective weights and

biases of the LSTM. The functions σ and $\tanh$ denote the sigmoid activation function and the hyperbolic tangent activation function, respectively.

The LSTM makes decisions about what to store, delete, and output at each step in the sequence through its gates:

- **Forget Gate:** Decides what information is discarded from the cell state.
- **Input Gate:** Updates the cell state with new information.
- **Output Gate:** Determines the next hidden state, influencing the output at the current step based on the cell state.

In our model, the LSTM architecture consists of two layers, each with 10 cells. We utilize an Adam optimizer for training, which is known for its efficiency in handling sparse gradients and adaptive learning rates. The loss function employed is based on the Mean Squared Error (MSE) of the fit. Although validation data is commonly used to prevent overfitting, in our case, it did not yield improved results due to the limited size of the validation dataset.

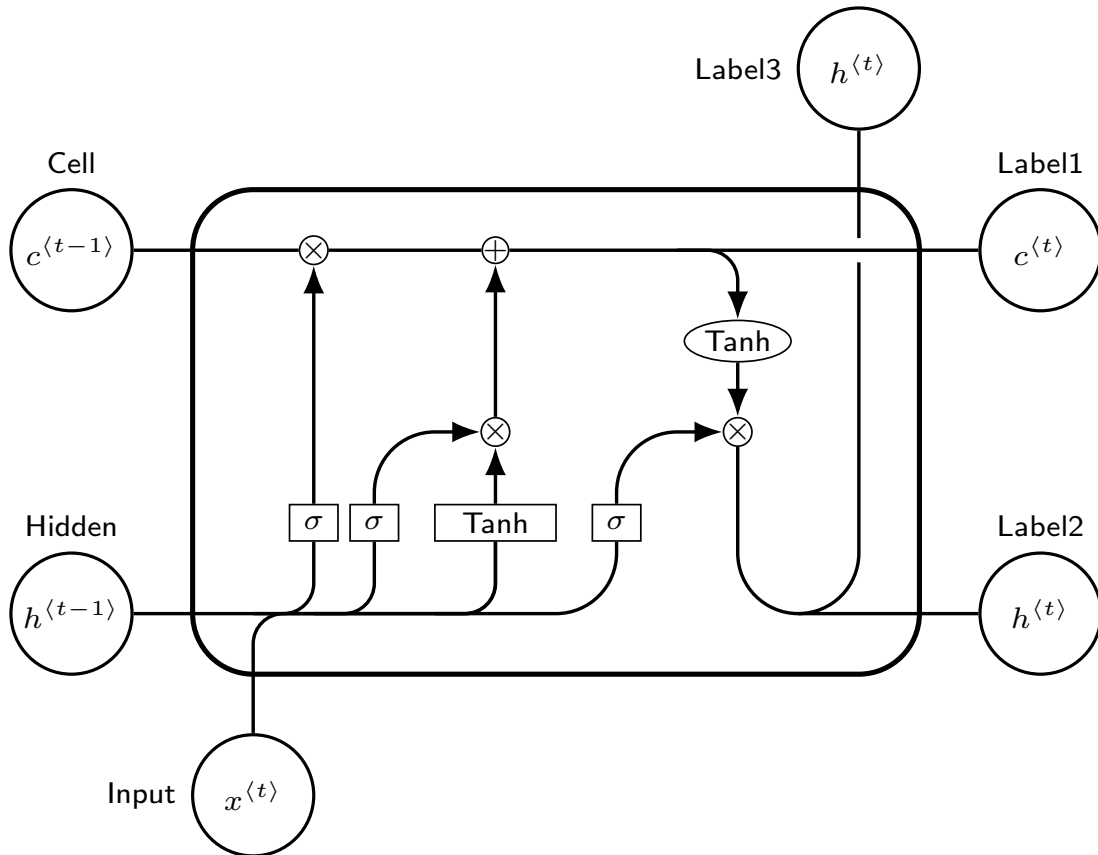


Figure A.28: Schematic representation of an LSTM cell. The cell state C_t is updated based on the inputs from the forget gate, input gate, and output gate, influenced by the previous cell state C_{t-1} and hidden state h_{t-1} .

Sample matches of identified patterns

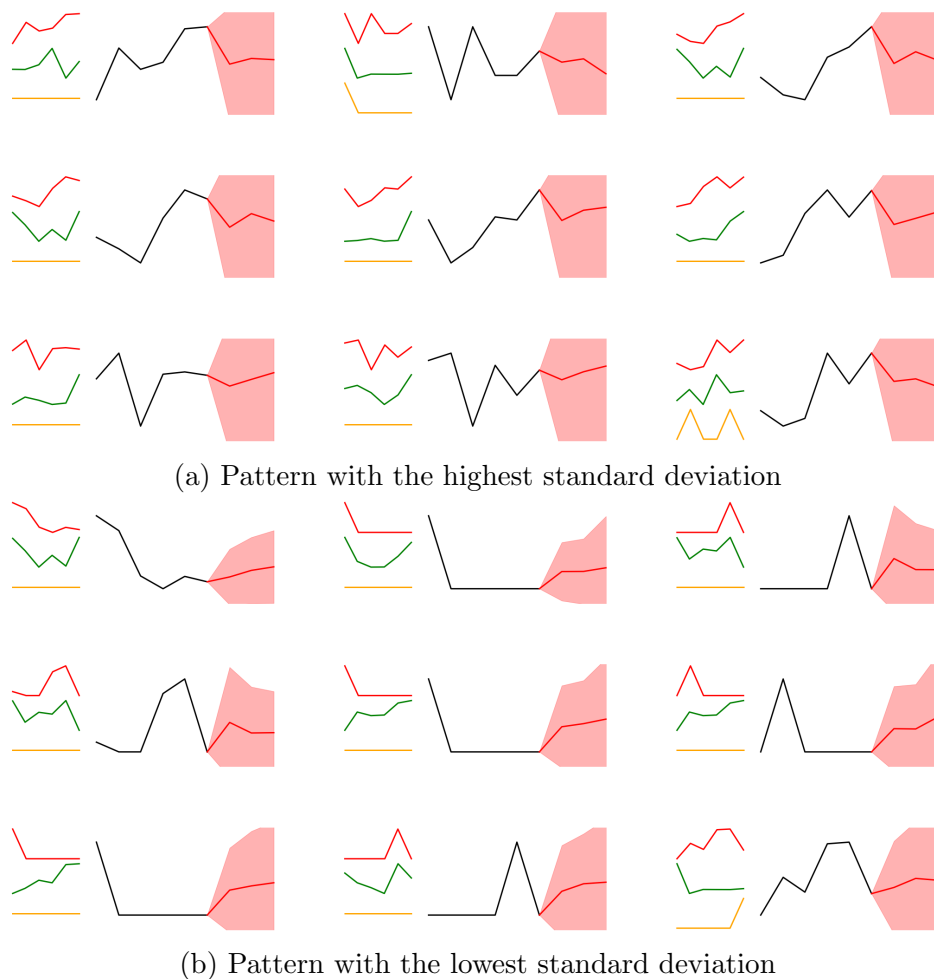


Figure A.29: Illustrative patterns of temporal migration patterns (red), food prices (green), and conflict (yellow) from the nine dynamic combinations with the highest and lowest mean standard deviation of the Past Futures, identified through dynamic time warping analysis. Each cell shows the combination of patterns on the left, with migration dynamic (in black) and the Past Future (in red) with the standard deviation as a confidence interval on the right.

Additional Results

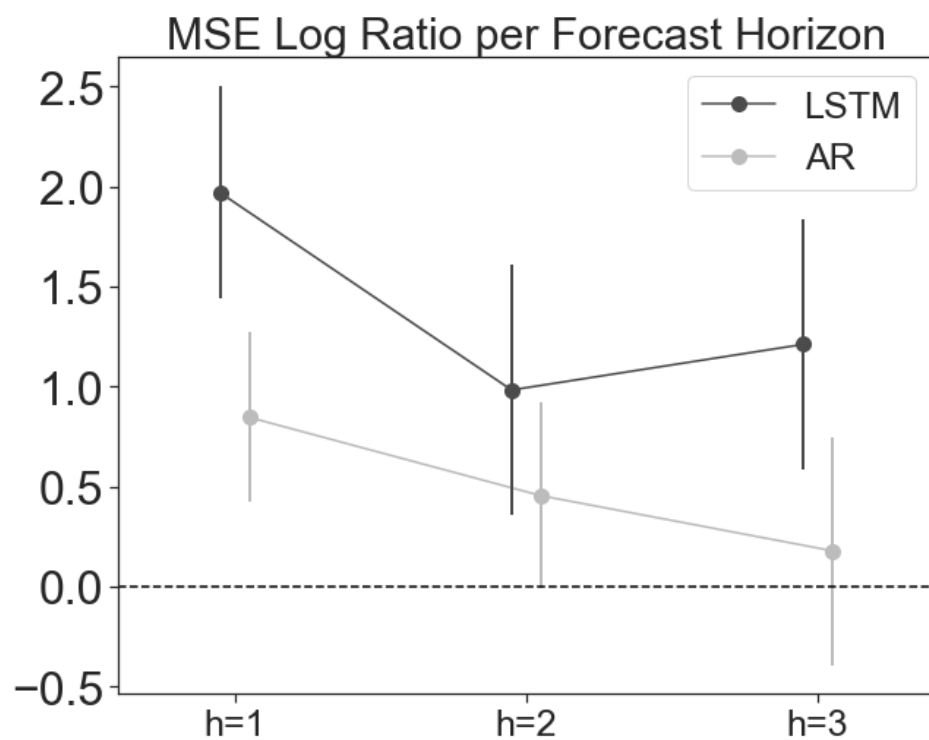


Figure A.30: MSE Log Ratio for forecast horizons $h \in [1, 2, 3]$. Higher values indicate that the model performs worse than the shape-finder.

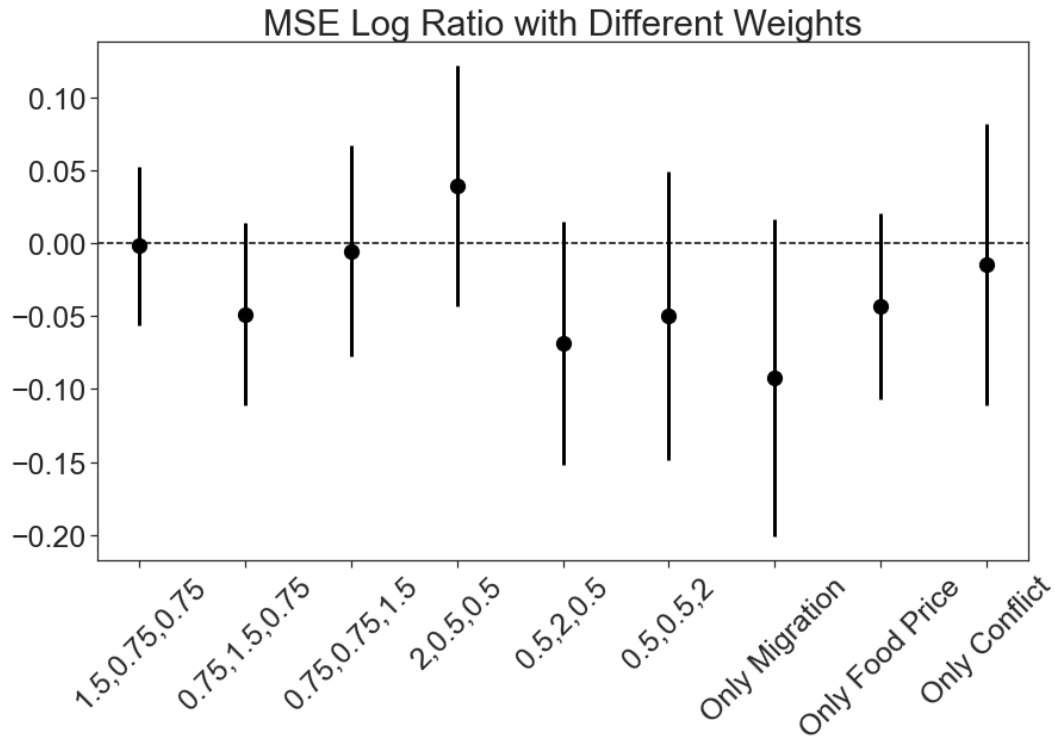


Figure A.31: MSE Log Ratio per model weight configuration compared to the equal weight model. It illustrates the relationship between model weights and the performance outcomes. Notably, variations in model weights do not statistically alter the model's results. The error bars across different weight configurations overlap, indicating that changes in weights do not lead to statistically significant differences in the model's performance.