Methodological Notes

Jesse Kommandeur, Claudio Prosperini

November 2024

DAMON
Methodological Notes

Authors: Jesse Kommandeur, Claudio Prosperini

Created: November 2024
Updated: February 2025

Cover photo: Canva

This HCSS tool and methodological note have been constructed by the HCSS Datalab. The conclusions and recommendations presented in this tooling and report are the result of in-house independent research and development. Responsibility for the content rests with the authors and the authors alone.

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies
Email: info@hcss.nl
Website: www.hcss.nl

Table of Contents

Key Takeaways 5

1. Introduction 6

2. Conceptualising similarity 7

2. Global Trends 9

3. National Trends 11

4. Events 13

Key takeaways

The **Disturbances and Aggression Monitor (DAMON)** is a HCSS tool that provides near real-time insights into global, national, and event-specific events related to conflict. It leverages data from ACLED to offer a structured analysis of battles, protests, riots, explosions, violence against civilians and strategic developments through three interactive views: Global, National, and Event. These views enable users to track trends across multiple time intervals and navigate between macro-level patterns and granular details.

The **similarity scoring** in DAMON identifies and ranks related events by analysing both textual descriptions and categorical attributes, providing insights at national and international levels. Using methods like TF-IDF and cosine similarity, the algorithms behind the system captures event details while factoring in structured data such as event types and actors. This **hybrid approach** ensures that both descriptive richness and categorical precision contribute to uncovering connections and patterns. By applying these measures, DAMON highlights event relationships across borders and within countries.

The tool's **regular updates**, ranging from weekly to yearly summaries, ensure its relevance for monitoring evolving geopolitical trends. Complementary tools like ranked tables and interactive maps further enhance user engagement, making DAMON a resource for understanding and responding the geopolitical landscape.

1. Introduction

In a dynamic and interconnected global landscape, understanding the complexities of geopolitical relations requires innovative tools and cutting-edge methodologies. At the Hague Centre for Strategic Studies (HCSS), we aim to lead in this domain with pioneering research and solutions that illuminate the intricacies of global, national, and local interactions. A key addition to our suite of tools is the **Disturbances and Aggression Monitor (DAMON)** – a near real-time event dashboard designed to provide comprehensive insights into worldwide political developments.

DAMON leverages data from the Armed Conflict Location & Event Data Project (ACLED) to offer a structured view of political violence, demonstration activities, and other critical political events. The tool supports nuanced analysis through an interactive dashboard with three distinct levels of engagement:

- **Global Trends:** Provides a macro-level overview of all recorded events worldwide, highlighting the most prevalent event types and ranking countries by event frequency over a selected time interval.
- **National Trends:** Offers country-specific insights by drilling down into the nature and distribution of events within a selected country for a chosen time period. Based on the event distribution for a given time interval, similar countries are mapped[1].
- **Event overview:** Allows users to explore detailed information about individual events, including the context, location, and actors involved. Similar national- and international events are also displayed to identify trends within- and across borders.

Designed with flexibility in mind, the dashboard allows users to track trends across various time intervals, including weekly, monthly, quarterly, yearly periods. This adaptability ensures that DAMON serves as a powerful resource for policymakers, analysts, and researchers seeking timely insights into the evolving dynamics of political events worldwide.

This methodological document serves as the foundation for DAMON's analytical framework, offering a clear overview of its data sources, variables, and similarity estimation techniques. It provides detailed insights into how ACLED's event data is integrated into DAMON, enabling systematic analysis of political dynamics. Central to this framework is the monitoring of activities by government forces, rebel groups, militias, and other non-state actors. The document also emphasises the categorisation of violent and non-violent political events, including protests, riots, and strategic developments.

The structure of this document is designed to guide readers through the core functionalities and insights provided by DAMON. The first section provides an overview of **global trends**, highlighting macro-level patterns in political violence and demonstrations. This is followed by an analysis of **national trends**, offering a closer look at detailed event data for specific countries. Next, the document delves into **event-level details**, presenting granular insights into individual events, their contexts, and locations.

1 Both the national and event similarity components of DAMON are still in the beta phase and undergoing continuous development. As a result, the identified similarities between countries may not always accurately reflect their true relationships.

2. Conceptualising Similarity

The concept of **similarity** is a key component of DAMON. Its application on the Event page is what confers novelty to the dashboard. In literature, similarity has been explored across numerous different fields, which led to distinct approaches in understanding and modelling similarity. In the following, we present a conceptual framework for similarity in DAMON, integrating insights from relevant literature and methodologies.

Similarity is a fundamental and widely used concept in computer sciences. The applied notion of similarity should be formulated by someone who is knowledgeable of the specific domain in which the comparison between objects is happening.[2] To flesh out this conceptualisation, a distinction between **categorical[3]** and **textual** similarity should be made.

When looking at **categorical similarity**, geometric and feature-based models[4] are often used to compare objects represented as a collection of categorical variables. The "distances" between objects correspond to their respective dissimilarities. These models align with machine learning techniques, where similarity is represented mathematically as a distance or proximity measure. A common yet simple method is Euclidean Distance, where direct geometric distance is used as a proxy for similarity. **Cosine Similarity** is a more refined version of this approach, which evaluates the angle between the vectors representing the objects of a similarity analysis. Its cosine is then used as a proxy for the similarity between the two objects.[5]

Textual similarity is another concept in literature and computational linguistics, involving the measurement of how closely texts are similar in terms of content, structure, and meaning. These techniques allow for text classification, and the retrieval of information from textual data. Cognitive and information-theoretic models emphasise the mental and probabilistic dimensions of similarity. Studies by Markman and Gentner analyse the nature of the commonalities and differences between structured representations, focusing on the cognitive processes involving comparisons.[6] *An Information-Theoretic Definition of Similarity* represents an effort to define similarity universally in information-theoretic terms, applicable as long as the domain of application is characterised by a probabilistic model, not unlike language itself.[7]

There are various techniques to analyse the similarity of texts, spanning various use cases and complexities. Traditional methods like Levenshtein distance are particularly useful when evaluating spelling differences between short texts of similar length, like movie titles[8]. Term Frequency-Inverse Document Frequency (TF-IDF) can also be used to analyse the importance of words in a document relative to a corpus. Words like "the", "and", or any other that are frequently used in English sentences are assigned a lower weight in similarity calculations;

---

[2] Das, G., & Mannila, H. (2000). Context-Based similarity measures for categorical databases. In *Lecture notes in computer science* (pp. 201–210).
[3] Categorical similarity is based on categorical variables, which refer to qualitative properties, such as country, actor, event type, among others.
[4] Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.
[5] Han, J., Kamber, M., & Pei, J. (2012). Getting to know your data. In *Elsevier eBooks* (pp. 39–82).
[6] Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition*, *24*(2), 235–249.
[7] Lin, D. (1998). An Information-Theoretic definition of similarity. *International Conference on Machine Learning*, 296–304.
[8] Kunde, N. O., Gaikwad, N. O., Kelgandre, N. P., Damodhar, N. R., & Swami, N. P. M. M. M. (2022). The Movie Recommendation System using Content Based Filtering with TF-IDF, Vectorization and Levenshtein Distance. *International Journal of Advanced Research in Science Communication and Technology*, 257–263

domain-specific, less frequent terms (“drive-by”, “drones”, etc.) are instead assigned more value. Generative AI and Large Language Models (LLMs) are also beginning to gain traction in the field of textual similarity, as these techniques have been demonstrated to perform well in domain-specific applications.[9]

We determined that TF-IDF is the most suitable technique for computing similarity in DAMON. This method can reliably analyse semantic texts of different lengths, at a reasonable computational cost. Further work should investigate the possibility of incorporating LLMs into textual similarity; especially if the computational costs of those models keep decreasing.

Both categorical- and textual methods of comparison are used complementarily in DAMON’s analysis of similarity of conflictual events. This is possible thanks to the different types of features describing each event.[10] Categorical similarity is defined by our use of Cosine Similarity between the vectors representing events. This is done with ease when evaluating “truly” categorical features of the events, which can be one-hot encoded[11] and evaluated directly. Textual similarity, on the other hand, requires the generation of “word” columns, which is done with TF-IDF. These complete the vector representing each event: the information-theoretic perspective is expressed by how words that appear more often in the corpus are given a smaller weight in their column, and thus in the similarity calculation.

[9] Gatto, J., Sharif, O., Seegmiller, P., Bohlman, P., & Preum, S. M. (2023). Text encoders lack knowledge: Leveraging generative LLMs for Domain-Specific semantic textual similarity. *Proceedings of the Third Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*, pages 277–288

[10] The following features are used to calculate our similarity score: event types, subtypes, fatality counts, locations, and textual descriptions.

[11] One-hot encoding is a technique used to represent categorical data as numerical vectors. For a given categorical variable with *n* unique values (categories), one-hot encoding creates *n* binary columns, each corresponding to one category. For each data point, the column representing its category is set to 1, while all others are set to 0.

3. Global Trends

The **Global page** provides a macro-level overview of worldwide conflict and political events. Powered by real-time data from ACLED, this page is designed to visually represent the distribution and intensity of events across countries, offering users an at-a-glance understanding of global dynamics. The interactive map is the primary feature of this page, with countries[12] shaded to reflect the intensity of conflict events during the selected timeframe.[13] By hovering over a country, users can quickly view its event count and name, enabling rapid assessments of event intensity.

Figure 1 – Global map showing the number of events per country

Users have the flexibility to select the timeframe for analysis, including *weekly, monthly, quarterly,* or *yearly* periods as described in the table below. This adaptability allows for customisable insights tailored to specific needs or interests. Additionally, the dashboard offers event type filters, enabling users to focus on specific categories such as battles, protests, riots, or strategic developments.[14]

Interval	**Update**	**Days**
Weekly	Every Wednesday	Last 7 days
Monthly	First Wednesday of the month	Last 31 days
Quarterly	First Wednesday of January, April, July, October	Last 91 days
Yearly	First Wednesday of the year	Last 365 days

[12] Displayed countries are based on the 195 UN Member and Observer States, with select abbreviations for commonly referenced states. Additional entities, such as Taiwan and Kosovo, may be included in specific analyses based on economic, diplomatic, or military relevance.
13 The intensity scaling is determined using the square root of the event counts, as this approach reduces the disproportionate impact of numerous low-intensity events while still reflecting their presence in the overall analysis.
14 Armed Conflict Location & Event Data Project (ACLED). "Codebook: Notes." Accessed November 25, 2024. https://acleddata.com/knowledge-base/codebook/#notes.

Complementing the map is a table that ranks countries based on event frequency and displays trends over the chosen timeframe. This provides a clear picture of which countries have experienced the most significant activity, along with insights into whether the volume of events is increasing or decreasing. For example, users can track whether a specific country's event count has surged or declined over the past week or month. The Global Page is designed as an intuitive entry point for exploring DAMON. Users can transition from a macro-level understanding to more granular insights by clicking on a country within the map or table, which redirects them to the National Page.

Figure 2 – Trends per event type of event and country

4. National Trends

The **National** page provides detailed insights into the conflict and political event dynamics within a specific country. This page is designed to allow users to explore and analyse country-level trends, offering a granular understanding of recent developments, event types, and their impacts. Furthermore, this page gives a summary of the total events and fatalities recorded in the selected country over the chosen timeframe. For example, users are informed on the number of events, fatalities, and key patterns such as the percentage of events involving specific types of incidents (e.g., battles, explosions, protests) or civilian targeting.

Another feature of this page is the event count comparison (ECC), followed by a bar chart visualising daily event counts categorised by event types. This breakdown enables users to identify patterns or surges in activity. In addition, the fatality count comparison (FCC) describes fatalities[15] per time interval, providing insights into the human cost of recent events and highlighting changes relative to the average.[16]

An interactive map as shown in figure 2 below highlights the geographic distribution of events across the country, with color-coded markers that represent different event types or their intensity. This allows users to quickly assess where the most significant clusters of activity occurred during the selected timeframe. In addition, users can navigate to the event details from this page.

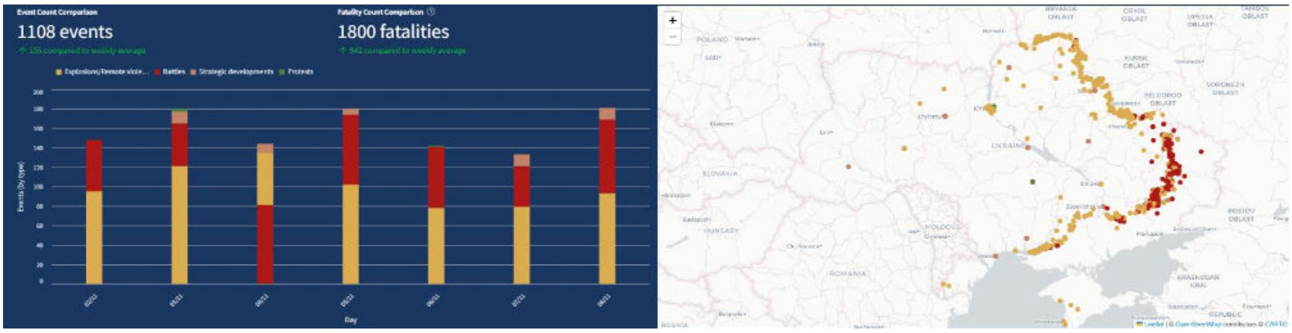


Figure 3 – National map showing the distribution of event in a given country (Ukraine)

To determine similarities between countries based on conflict event distributions, our method first computes the event type distribution for each country as a percentage of total events, creating a vector (e.g., [0.8, 0.1, 0.05, 0.03, 0.02, 0]). These vectors are then compared using cosine similarity to measure distributional resemblance, which quantifies how closely the event types align between countries.

Additionally, an event count similarity metric is calculated based on the normalised absolute difference in total event counts between two countries. The overall similarity between countries is a weighted combination of these metrics[17]. This is represented in the combined country

[15] The number of reported fatalities arising from an event. When there are conflicting reports, the most conservative estimate is recorded. In the case of conflict with other entities, casualties could be on either side.
[16] The average is calculated by subdividing the whole dataset in weeks, months, quarters, or years, based on the event type selection. The average for this timeframe subdivisions Is then calculated, and compared to the current selection.
[17] The distribution of events is assigned 0.9 weight; the total event counts are assigned the remaining 0.1.

similarity (CCS) below, where, A and B are the event distribution vectors for the base and comparison countries, and nA and nB are their respective total event counts.

$$CCS = 0.9 \times \left(\frac{A \cdot B}{\| A \| \| B \|}\right) + 0.1 \times (1 - \frac{| nA - nB |}{\max(nA, nB)})$$

A minimum of 80% similarity is required for a country to be considered similar to the one currently in focus. A maximum of 6 countries are displayed, 3 from the same region of the selected country and 3 from the rest of the world.

To facilitate further analysis, the detailed event view table presents a categorised list of all recorded event types and their counts. Each category includes information on the most active actors, such as military forces, protesters, or government entities. In addition, users can toggle event categories to focus on specific types of incidents and observe their trend lines for a given time interval. By clicking on individual events or selecting events from the map, users can navigate to the Event Page, which offers even more granular insights into specific incidents.

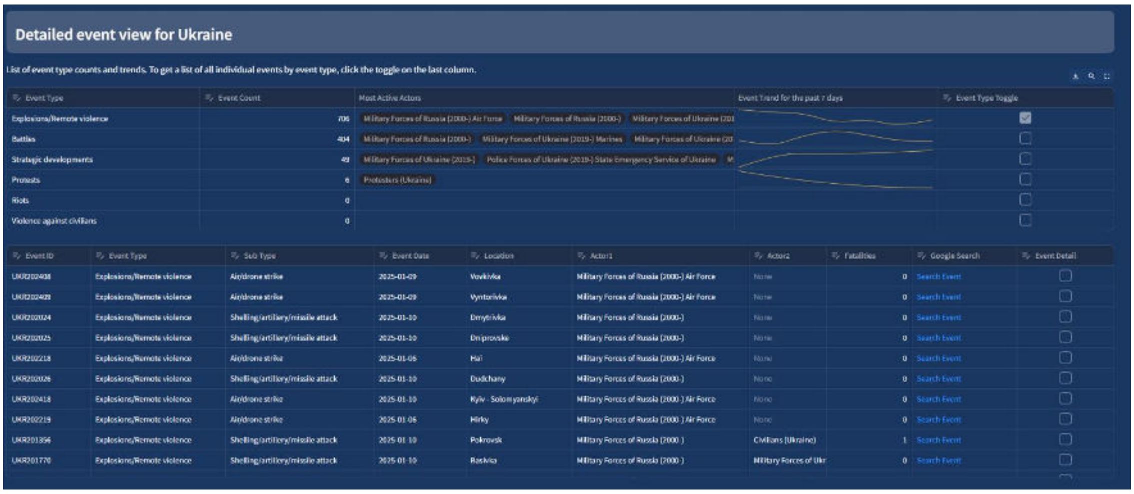


Figure 4 – Detailed event view for a given country (Ukraine)

5. Events

The **Event** page offers an in-depth view of a specific event. It includes attributes such as the context (description of event), the event type, sub-type, involved actors, location, and fatalities. This information is presented alongside an interactive map that situates the event geographically. Another feature of this page is the **Similar Events** section, which highlights events related to the event currently in view. These are grouped into similar national- and similar international categories, allowing users to explore patterns and connections at both regional and global levels.

The system identifies similar events using a hybrid approach that integrates textual similarity and categorical similarity. To analyse textual data, such as event descriptions from the notes column, the system employs TF-IDF to convert text into numerical vectors. Cosine similarity is then used to measure the closeness between these vectors.

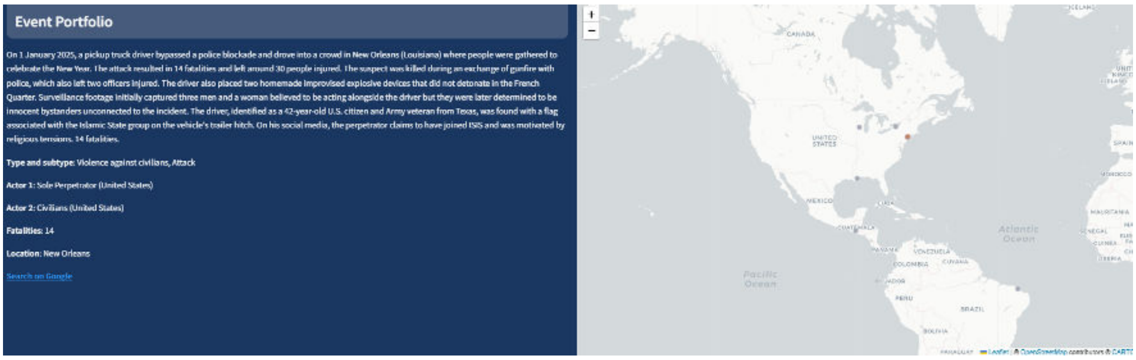


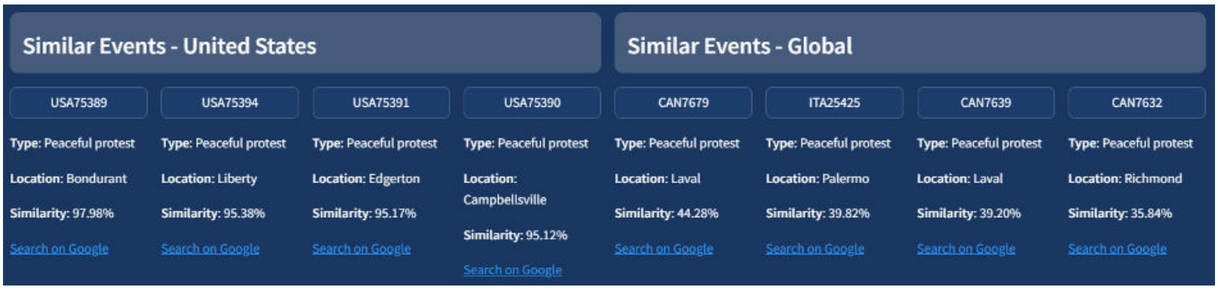
Figure 5 – Event portfolio with map showing similar events

For categorical data, including attributes like event type, sub-event type, actor(s), and location, the system encodes these fields into numerical formats using one-hot encoding. Additionally, fatalities are incorporated as a third metric if the fatality count for the event to compare is > 0, using a lower impact weight. Cosine similarity is again employed to measure the similarity between these encoded attributes. The final similarity score is calculated as a weighted combination of these two (or three) metrics, with an additional consideration for the fatality count when applicable. The formula for the combined event similarity (CES) is:

$$CES = \begin{cases} (w_t \cdot Textual\ Similarity) + (w_c \cdot Categorical\ Similarity) + (w_f \cdot Fatality\ Similarity), & if\ fatalities > 0 \\ (w_t \cdot Textual\ Similarity) + (w_c \cdot Categorical\ Similarity), & Otherwise \end{cases}$$

To determine the optimal weights for combining textual similarity (W_t) and categorical similarity (W_c), a heuristic approach was employed. The process began with equal weights ($W_t = 0.5$) and ($W_c = 0.5$), ensuring an unbiased initial distribution between the two components. Results were then manually inspected to evaluate the performance of different weight configurations, focusing on how well the calculated similarities aligned with expert judgment. This iterative process included the use of feedback loops to refine the weights gradually, adjusting the balance based on observed outcomes and insights.

After several iterations, it was found that assigning a weight of approximately ($W_t \approx 0.63$) for textual similarity and ($W_c \approx 0.37$) for categorical similarity provided the best results if fatalities was $= 0$. If fatalities > 0, the distribution resulted in the following division of weights ($W_t \approx 0.58$), ($W_c \approx 0.37$), ($W_{tf} \approx 0.05$) This weighting reflects the importance of textual descriptions, such as event notes, in capturing nuanced details about events while maintaining the relevance of categorical attributes[18].



The effectiveness of the system is further enhanced by continuous monitoring and updating processes. As new events are added, the system recalibrates its similarity algorithms to incorporate fresh data, ensuring the accuracy and relevance of the similar events suggested.

A "Search on Google" function is also available for each individual event. The search query is formulated by combining an event's subtype, location, and date, which allows users to search for relevant news articles, blogs and other information outside of DAMON.

[18] For events with higher granularity, textual descriptions often carry more nuanced details (e.g., *"Protesters demanded education reforms in the capital"*), making higher weights for textual similarity ($W_t = 0.8$) more effective. In contrast, for events with little textual detail, categorical attributes like event type and actors (e.g., *"IED attack by militants"*) are more relevant, favouring higher weights for categorical similarity ($W_c = 0.7$)

The Hague Centre for Strategic Studies

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies