INTEGRATING CONFLICT DATA FOR ENHANCED ANALYSIS: A MULTI-SOURCE APPROACH TO SUPPORT VULNERABILITY FRAMEWORKS

Authors: Wesley Douglas, Ashleigh Basel, Onivola Minoarivelo, Alessandro Craparo, Grazia Pacillo, Peter Läderach

Affiliations: CGIAR FOCUS Climate Security, The CGIAR Initiative on Fragility, Conflict and Migration, The Alliance of Bioversity International and CIAT

PHOTO: WFP/ALFREDO ZUNIGA

ABSTRACT

The future of conflict and food security analysis depends on moving beyond isolated datasets to embrace integrated, multi-dimensional insights. This study presents an innovative approach to integrating conflict event data from multiple sources, utilising a customised version of the Matching Event Data by Location, Time, and Type (MELTT) framework and the Validation and Exploration of the Hierarchical Integration of Conflict Event Data (VEHICLE) methodology. The analysis focuses on regions monitored by the Famine Early Warning System and Global Fragility Act priority countries, aiming to improve vulnerability frameworks within fragile regions. The resulting enhanced regional conflict datasets offer a more comprehensive understanding of conflict dynamics and are a valuable tool for organisations, supporting more effective crisis response and conflict impact assessments.

Key messages and recommendations

- The integration of conflict data from multiple sources addresses the limitations of individual datasets.
- By adapting and customising established approaches to conflict data integration, the accuracy of conflict event categorisation can be enhanced. Resulting in greater precision when matching conflict events across datasets.
- The resulting integrated datasets support more accurate forecasting of the downstream effects of conflict as well as the analysis of violence patterns and dynamics over time and across regions.
- Through the use of the improved dataset the link between conflict and food security is clear, conflict acts as a crisis multiplier, compounding vulnerabilities in regions such as Mozambique and Haiti.
- The ongoing integration of multiple conflict datasets is recommended to improve the depth and accuracy of conflict analysis.
- In the effort to improve conflict data precision, the proposed custom approach to event matching and classification is advocated for, to ultimately ensure richer integrated datasets.
- The application of the integrated datasets should be especially considered within broader risk and vulnerability frameworks to support targeted interventions.
- The insights gained from integrated conflict datasets should be prioritised for incorporation in fragile regions, such as those identified by the GFA.

1. INTRODUCTION

The increasing availability of event-level conflict data has greatly advanced the analysis of violence patterns and dynamics over time and across regions. Traditionally, researchers have relied on individual datasets to explore specific types of conflict-related events, such as state-based violence or insurgencies. However, relying on a single dataset often limits the scope and depth of the analysis due to incomplete coverage, missing events, and potential biases in data collection processes. To address these limitations, a comprehensive conflict data integration process was undertaken.

This approach integrates conflict event data from multiple sources, focusing on regions monitored by the Famine Early Warning System (FEWS NET) and the priority countries of the Global Fragility Act (GFA), therefore focusing on areas highly vulnerable to conflict-induced food insecurity and fragility. The integration of these datasets was built upon a customised version of Matching Event Data by Location, Time, and Type framework (MELTT; Donnay et al., 2019) as well as Validation and Exploration of the Hierarchical Integration of Conflict Event Data framework (VEHICLE; Mayer et al., 2021).

Recognising the critical importance of robust conflict data, this integration effort seeks to enhance the accuracy and comprehensiveness of multi-source conflict information. This can facilitate impact-based forecasting, providing critical insights into the downstream effects of conflict, particularly in regions vulnerable to food insecurity. By incorporating conflict dynamics into broader risk frameworks, this method enhances the vulnerability component, improving the predictive capacity of systems like FEWS NET. It allows for a more nuanced understanding of how conflict interacts with other stressors and the potential impacts on focal population, thereby offering more effective tools for crisis anticipation, mitigation and targeted intervention.

Several prominent conflict event datasets are available, each offering unique insights into different aspects of violence and unrest. As demonstrated by Donnay et al. (2019) and further supported by the complementary work of Mayer et al. (2021), event data can be effectively matched across location, time and type of violence using sources such as the Armed Conflict Location & Event Data (ACLED; Raleigh & Hegre, 2009), the Uppsala Conflict Data Project (UCDP; Davies et al., 2024; The UCDP-GED dataset provides detailed records of organised violence and state-based conflicts. By focusing on both state-based and non-state violence, UCDP- Sundberg & Melander, 2013), the Global Terrorism Database (GTD; START 2013), and the Social Conflict Analysis Database (SCAD; Salehyan et al., 2012).

For this analysis, a systematic approach was employed to integrate conflict event data from two key sources, ACLED and the UCDP, specifically the UCDP-GED (Georeferenced Event Dataset). ACLED captures a wide range of conflict-related events, including political violence, protests, riots, and violence against civilians. It is renowned for its comprehensive coverage of non-state violence and political unrest. GED is particularly valuable for analysing larger, organised armed conflicts.

While the Global Terrorism Database (GTD) and Social Conflict Analysis Database (SCAD) were part of the original methodological frameworks (Donnay et al., 2019; Mayer et al., 2021), they are excluded for this research approach due to their data collection limitations, as both heavily depend on media sources. As noted by Weidmann (2015), the dominant reliance on media sources can introduce biases and gaps in spatial coverage, especially for datasets like SCAD and GTD that concentrate on social conflict and terrorism. Depending on the region, these biases might result in the insufficient representation of either state-based or non- state armed conflicts.

Through this approach, more comprehensive, complete and granular datasets are created that can capture the spatiotemporal overlap of violent conflict events. The method employed here not only minimises duplicate entries during dataset merging but also improves event coverage and the accuracy of conflict data, especially in underreported or volatile regions. The resulting integrated datasets serve as valuable tools for guiding operations and decision-making processes for organisations such as USAID as well as in broader areas of conflict prevention and security. By offering more detailed insights into patterns of violence, the integrated datasets can better support preparedness and operations around food insecurity, improve the understanding of conflict impact and provide insights into the relationship to other hazards such as climate-related extremes. After providing deeper insights into the building blocks of the integrated dataset, we provide a summary of the resulting datasets for FEWS NET and GFA countries. We then highlight the improvements our approach offers over previous methods, concluding with a demonstration of how the integrated datasets can be applied to assess conflict-related food insecurity in Mozambique and Haiti.

2. METHODOLOGY

2.1 Countries of Interest

This project focuses on six GFA countries and three countries monitored by FEWS NET (Figure 1). The six GFA countries -Benin, Guinea, Papua New Guinea, Ghana, Côte d'Ivoire, and Libya- are considered critical for U.S. foreign policy under the Global Fragility Act, which seeks to reduce fragility and prevent conflict in key regions. These countries face a range of conflict-related challenges, including political instability, ethnic tensions, and insurgent activity, all of which are compounded by other non-conflict related hazards and socio-economic vulnerabilities. Through integrated datasets from ACLED and UCDP-GED, this project aims to provide a clearer understanding of violence patterns and conflict triggers across these diverse regions.

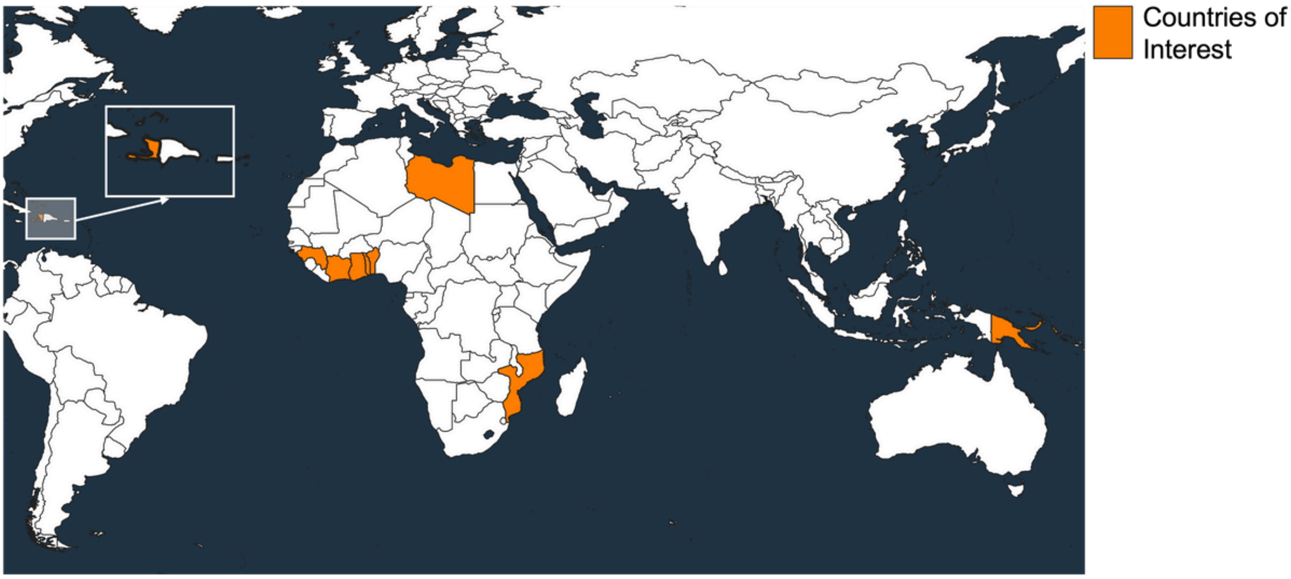


Figure 1: Nine countries of interest including Benin, Guinea, Papua New Guinea, Ghana, Côte d'Ivoire, Libya, Togo, Mozambique, and Haiti.

Table 1 provides a comparison of conflict event data from ACLED and UCDP-GED across the nine countries of interest. There are substantial differences in the number of events recorded between the two datasets. These differences highlight the variation in the scope and event classification between the datasets.

*Table 1. Data Summary by Countries of Interest**

Affiliation	Country	ACLED Data Start Date	UCDP Data Start Date	ACLED Event Count	UCDP Event Count
GFA	Benin	1997	2019	1462	36
	Guinea	1997	1994	2836	49
	Papua New Guinea	2021	1989	1031	138
	Ghana	1997	1991	2291	52
	Côte d'Ivoire	1997	1990	2940	360
	Libya	1997	2008	11614	1262
FEWS NET	Togo	1997	1991	569	49
	Mozambique	1997	1989	4031	1326
	Haiti	2018	1989	7919	231

**From UCDP-GED Global version 24.1 (Davies et al., 2024; Sundberg & Melander, 2013) & ACLED Database, via API access, 10 Oct '24.*

The process of merging ACLED and UCDP data is essential in creating a more comprehensive dataset that overcomes the limitations of individual sources. For instance, in Papua New Guinea, merging increases the temporal scope significantly, as UCDP data begins in 1989, whereas ACLED data only starts in 2021 (Table 1). By combining these datasets, we obtain a more complete picture of conflict trends across a longer historical period.

While the discrepancies in event counts may seem large (Table 1), they highlight the different focuses of each dataset. ACLED tends to capture a broader range of conflict events, including protests and local incidents, whereas UCDP often focuses on larger, more organised violent conflicts.

These differences can be seen as complementary, as merging the datasets allows for a richer and more nuanced understanding of conflict. ACLED's detailed recording of smaller-scale events can add granularity to UCDP's coverage of more severe conflicts, ensuring that both major and minor events are accounted for in analyses.

2.2 Matching Event Data by Location, Time and Type

As part of the preprocessing steps, we excluded ACLED entries classified as "strategic developments" since they fall outside the scope of this project, which focuses specifically on direct conflict events. While strategic developments are relevant in broader geopolitical analysis, they do not represent the immediate, event-level conflict dynamics that this analysis seeks to capture. The integration of ACLED and UCDP-GED datasets was performed using a customised version of the Matching Event Data by Location, Time, and Type methodology (MELTT; Donnay et al., 2019). This approach builds on the original MELTT process by incorporating the hierarchical structure introduced in the Validation and Exploration of the Hierarchical Integration of Conflict Event Data framework (VEHICLE; Mayer et al., 2021). This ensures that entries occurring at similar times and locations across datasets are accurately matched and consolidated, while enabling more refined categorisation of events and actors. The matching and consolidation processes are manged within the MELTT structure and is designed to address common challenges in conflict data integration, such as overlapping event coverage, inconsistencies in event reporting, and the need to remove duplicate entries. This approach is implemented by using the meltt package v0.4.3 (Donnay et al., 2022) in R (R Core Team 2022) and follows three primary stages:

2.3 Spatio-temporal Matching

Using the MELTT protocol, recorded events in the ACLED and UCDP-GED datasets are systematically compared based on their location (latitude and longitude) and date of occurrence. A spatio-temporal "blocking" approach is then applied, where only entries occurring within a defined space (Δs) and time window (Δt), are flagged for potential matching. This method minimises unnecessary comparisons, streamlining the matching process. For this research, the spatio-temporal matching criteria are set to 5 km and 1 day. This choice is guided by the VEHICLE (Mayer et al., 2021) approach and reflects a balanced compromise to account for the common uncertainties in event reporting. The uncertainty resulting from discrepancies in the data is addressed through the MELTT substructure, which is further refined during the attribute comparison and event type matching.

2.4 Attribute Comparison and Event Type Matching

Once potential matches are identified based on spatio-temporal proximity, they are compared using additional entry attributes, such as event type and actors involved. To resolve instances where multiple events from different datasets might be matched to the same entry, the stable marriage problem algorithm is utilised (Gale & Shapley. 1962). This approach is particularly well suited for conflict data integration, as it systematically pairs events by maximising the likelihood accurate matches based on event attributes. The stable marriage problem ensures that each event in one dataset is matched with only one event from another dataset, preventing duplicate matches and minimising errors.

The MELTT framework accommodates uncertainty or 'fuzziness' in its matching process, allowing entries to be matched even when not all attributes perfectly align. This is particularly useful in conflict data integration, where discrepancies in reporting can arise due to different coding protocols across datasets. A hierarchical taxonomy is utilised to evaluate whether entries are likely to be the same by categorising them into broader groups.

For both VEHICLE and MELTT, this taxonomy is divided into three components: actor taxonomy, event taxonomy, and precision taxonomy. The actor taxonomy categorises the participants involved in conflict events, such as government forces, rebel groups, or civilians (Figure 2a). The event taxonomy classifies the type of conflict activity, such as battles, violence against civilians, territorial disputes or violent (and nonviolent) protests/demonstrations (Figure 2b). Lastly, the precision taxonomy accounts for the level of geographic accuracy in the reported event location, ranging from precise coordinates to broader areas (Figure 2c).

In our approach, we retain the original MELTT process for the precision taxonomy, as it aligns with the VEHICLE precision taxonomy. However, we adopt the hierarchical structures from VEHICLE for the actor and event taxonomies, enabling more granular classification and improved matching across datasets. This ensures that entries are consistently categorised and matched, even when discrepancies exist in how actors or event types are coded across different datasets.

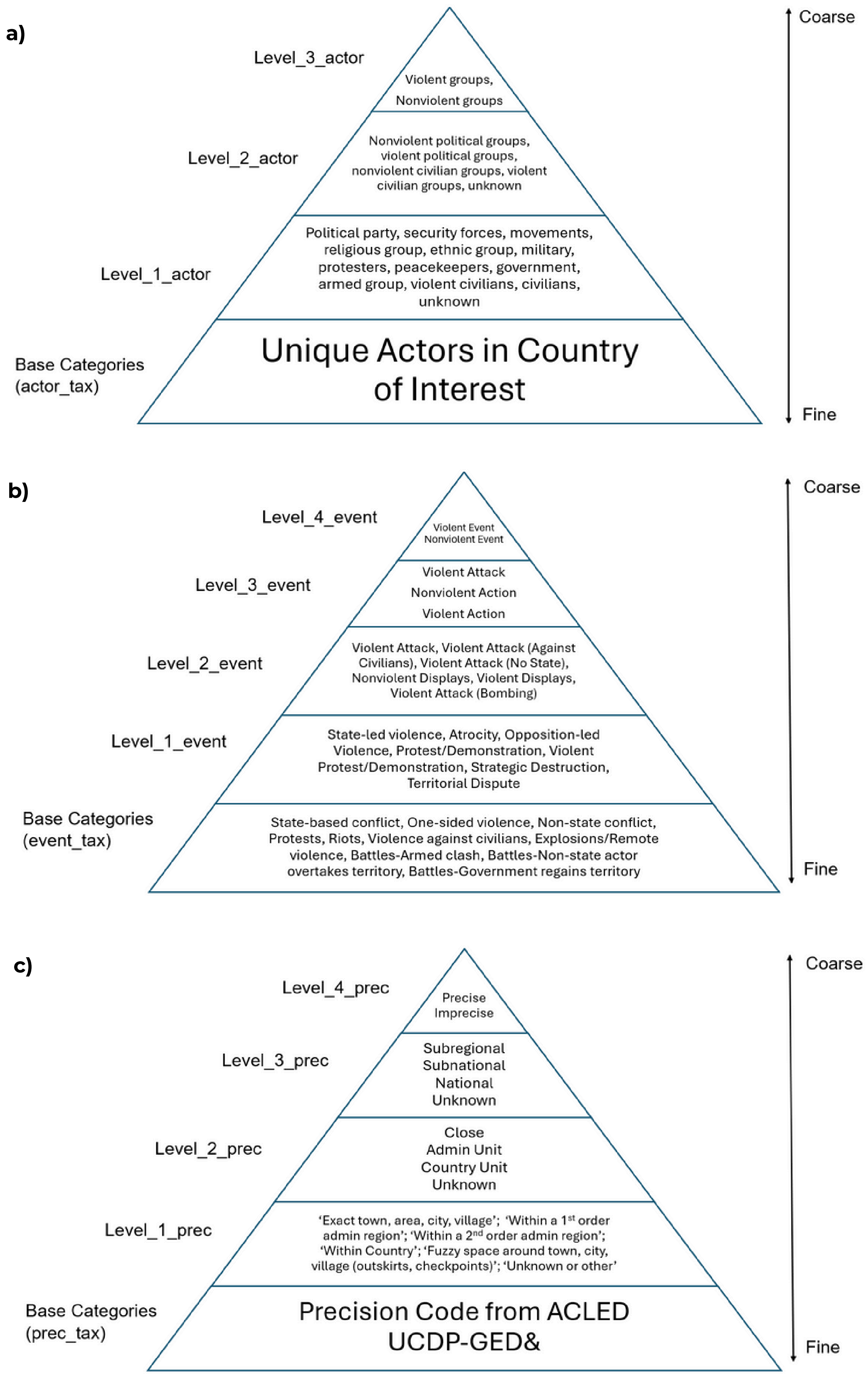


Figure 2: This figure illustrates the hierarchical structure of the actor (a), event (b) and precision (c) classification used in this project. The pyramid represents different levels of classification, ranging from the most granular categories at the base to more generalised categories at the top.

2.5 Actor taxonomy

The original MELTT framework utilises a keyword-based classification system to categorise actors within a predefined actor hierarchy. For instance, actors associated with keywords such as "militia" are automatically assigned to the ‘armed group’ category. While this automated approach is efficient for large-scale datasets, it can introduce inconsistencies, particularly when datasets use different terminologies to describe the same actors.

In the present study, a manual actor classification approach is adopted. Specifically, actors from each dataset were individually reviewed and assigned to the 'Level_1_actor' categories, as illustrated in Figure 2a. This method, although labour-intensive, ensures a higher degree of consistency when categorising actors across both ACLED and UCDP-GED, refer to Table 2 for actor count per country of interest.

This approach was chosen due to its ability to address ambiguities and context-specific actor roles, which may not be captured effectively through an automated, keyword-based system. This manual process also allows for more nuanced decision-making, particularly when dealing with regional variations or dataset differences in actor names and behaviours, where actor roles may not be immediately discernible through keyword match alone.

*Table 2. Summary of Actor Count per Country of Interest**

Affiliation	Country	ACLED Actor Count	UCDP Actor Count
GFA	Benin	81	3
	Guinea	93	8
	Papua New Guinea	144	11
	Ghana	156	4
	Côte d'Ivoire	125	16
	Libya	284	21
FEWS NET	Togo	30	3
	Mozambique	63	6
	Haiti	114	9

**From UCDP-GED Global version 24.1 (Davies et al., 2024; Sundberg & Melander, 2013) & ACLED Database, via API access, 10 Oct '24.*

2.6 Event and Precision Taxonomies

Within the MELTT process, the event taxonomy ensures that similar conflict activities, even when coded with different labels across datasets, can be grouped into broader event categories, allowing for flexible matching. The base categories, as illustrated in Figure 2 a-c, are taken directly from ACLED and UCDP event labels. They are then grouped into broader categories - Level_1_event, Level_2_event, and so on- until the most general classification, Level_4_event, where each entry is designated as either a violent or non-violent event. This hierarchical structure is key during the Attribute Comparison stage, as it enables the identification of corresponding events across datasets.

Similarly, the precision taxonomy categorises the level of geographic accuracy for each event, ranging from exact coordinates to broader regional approximations. The base categories, as illustrated in Figure 2c, are taken directly from ACLED and UCDP precision codes and are then grouped into broader categories - Level_1_prec, Level_2_prec, and so on- until the most general classification, Level_4_prec, where each entry is designated as either precise or imprecise in terms of geographic location. Within the MELTT process, this taxonomy allows the integration to account for varying degrees of spatial precision in event data, ensuring that events with different levels of geospatial accuracy can still be compared and matched appropriately.

2.7 Final Event Integration

We consolidate the matched entries into a single event record, after matching events across datasets. Events that could not be matched across datasets were retained as unique entries, ensuring comprehensive coverage of all conflict events within a chosen region. This process allows us to leverage complementary information from both datasets, thereby enhancing the overall quality of the data.

3. SUMMARY OF THE INTEGRATED CONFLICT DATASET

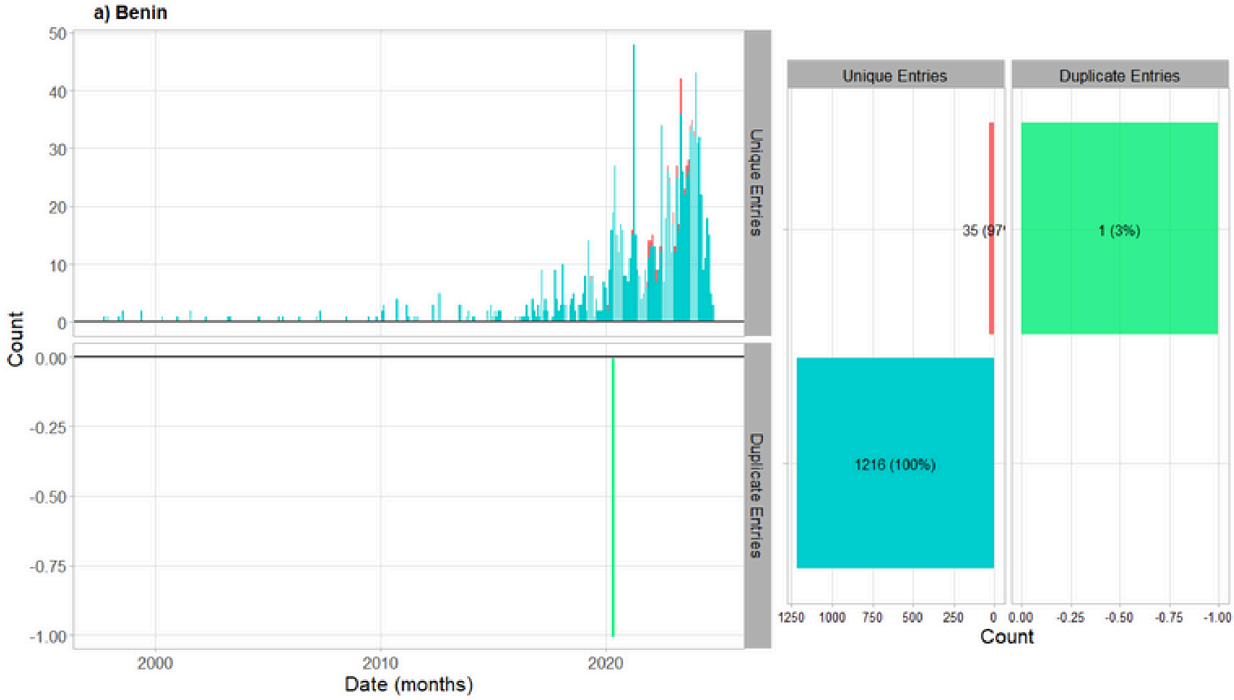
A critical component of the integration process is the hierarchical classification of actors and events. By leveraging the actor and event taxonomies derived from the VEHICLE framework, it enables the categorisation and integration of events more consistently, particularly in cases where actors or event types were coded differently across the original datasets. This manual classification approach was especially relevant given the specific regions chosen for the project, where actor nuances played a crucial role in identifying correct matches.

The Monte Carlo analysis from Donnay et al. (2019) demonstrated that increasing the richness of taxonomies improves performance, particularly when using additional levels within the hierarchical taxonomy. As a result, by opting for the more detailed VEHICLE taxonomies, in turn provided more granular distinctions between actors and event types. This strategy was essential for capturing the complexity of the data, particularly in regions where actors and event types might be misclassified by more generic categories.

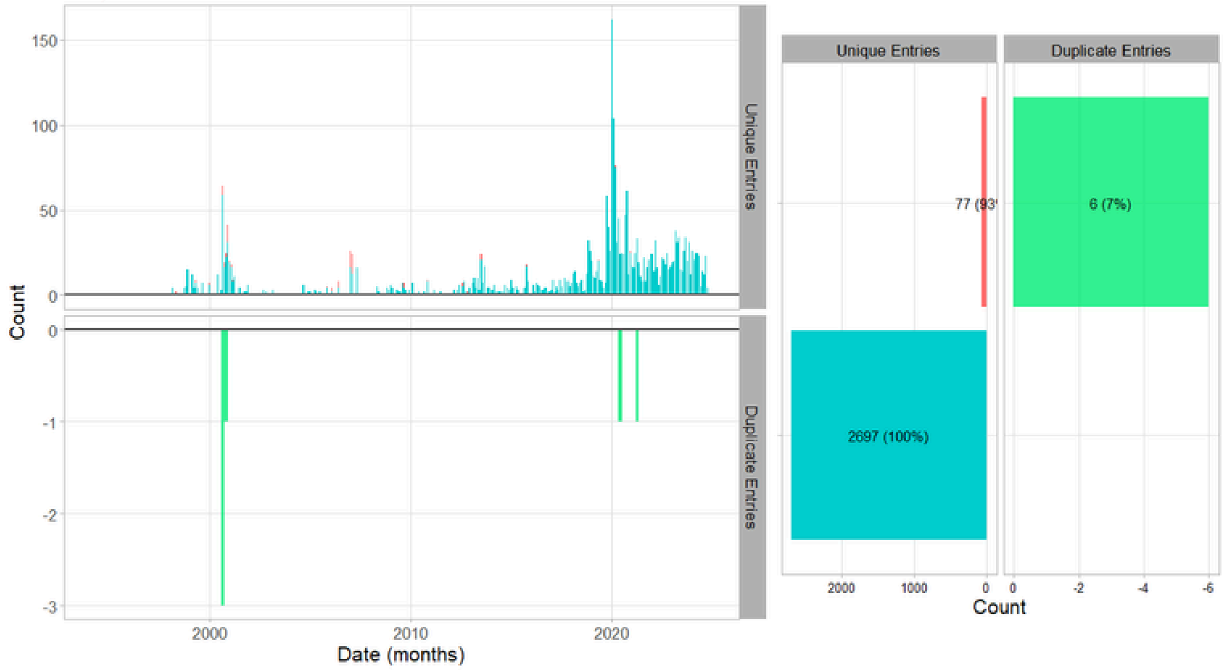
Although manual classification of actors was appropriate for this study, it becomes resource intensive when applied to datasets with a large number of actors. For instance, in Nigeria (not included in this study), there are around 880 distinct actors. Nonetheless, a well-constructed taxonomy is crucial for effective data matching, serving as the primary mechanism by which a user’s assumptions -regarding how data fits together- are made transparent. The manual classification used in this analysis provided a clearer alignment between the conceptual framework and the empirical data, allowing for more accurate and meaningful event integration. This process ultimately produces finalised datasets that included the taxonomic classifications, enabling better analysis and more reliable insights in the subsequent stages of the research and use cases. This effort ensured that events involving key actors were consistently represented, enabling a more reliable analysis of conflict dynamics.

As a result, the final datasets provide a clearer picture of conflict events, particularly where data overlap between ACLED and UCDP-GED. All detected duplicates were successfully matched and removed, ensuring that no event is double counted in the final datasets.

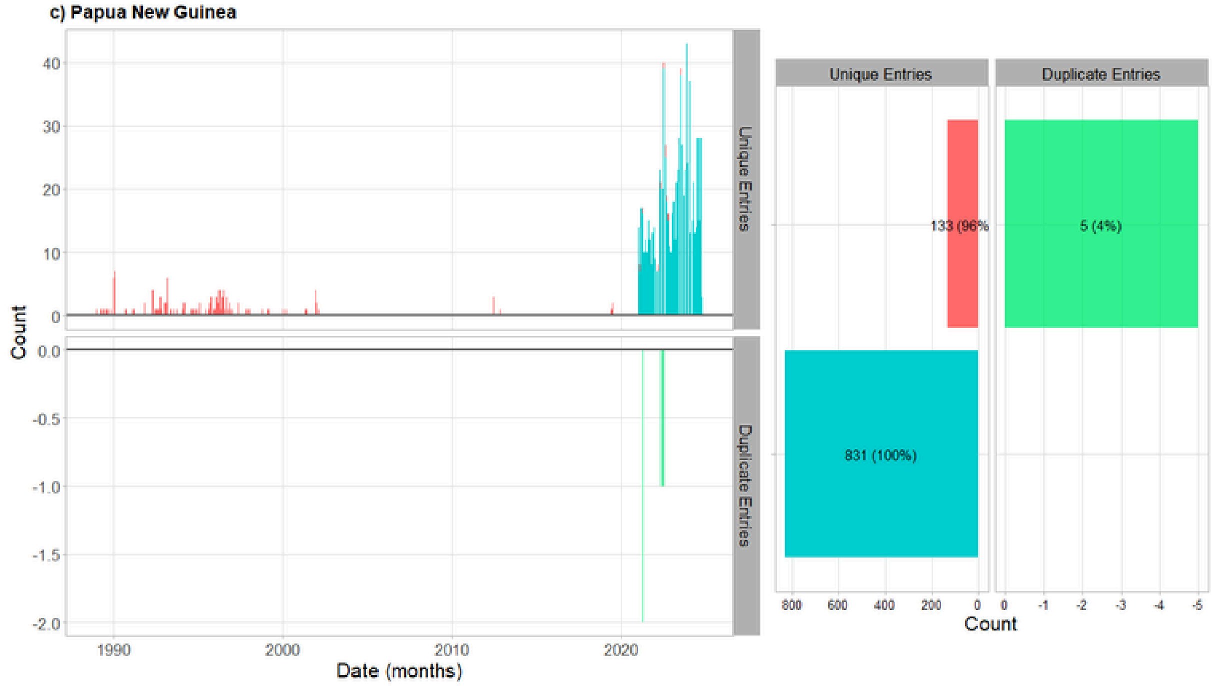
a) Benin



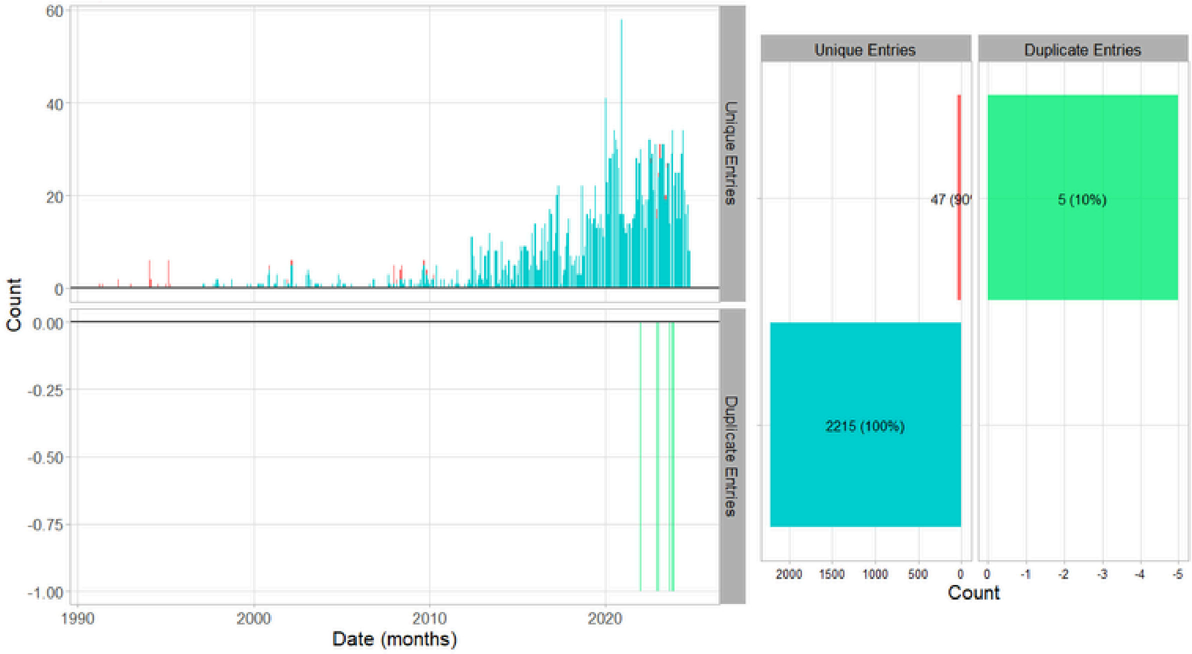
b) Guinea



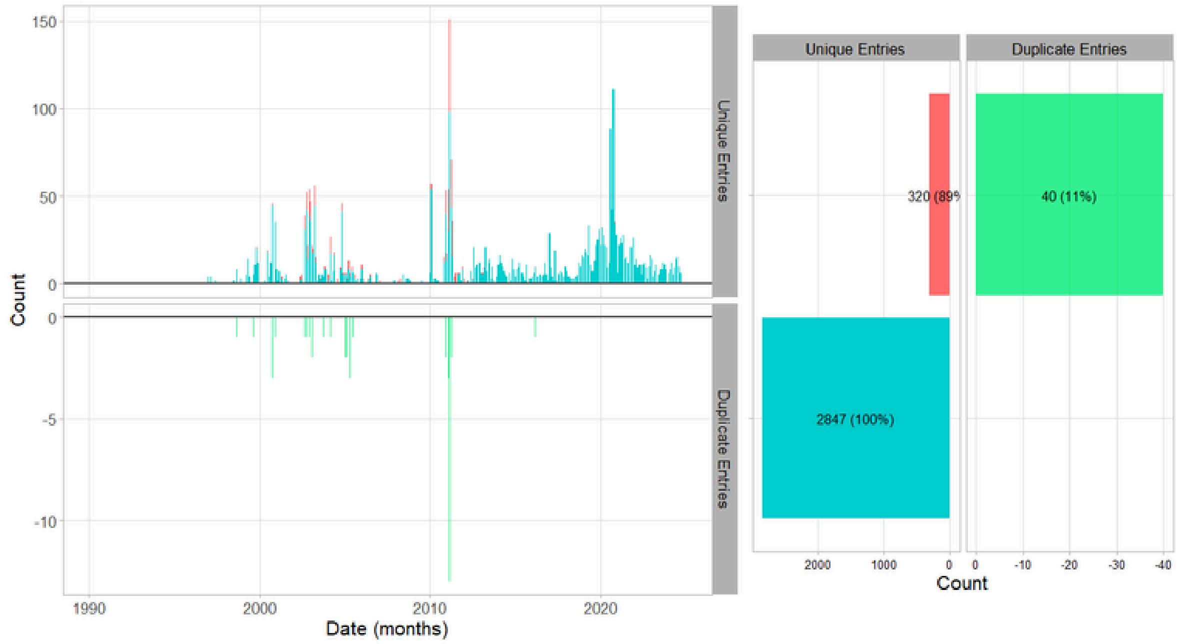
c) Papua New Guinea



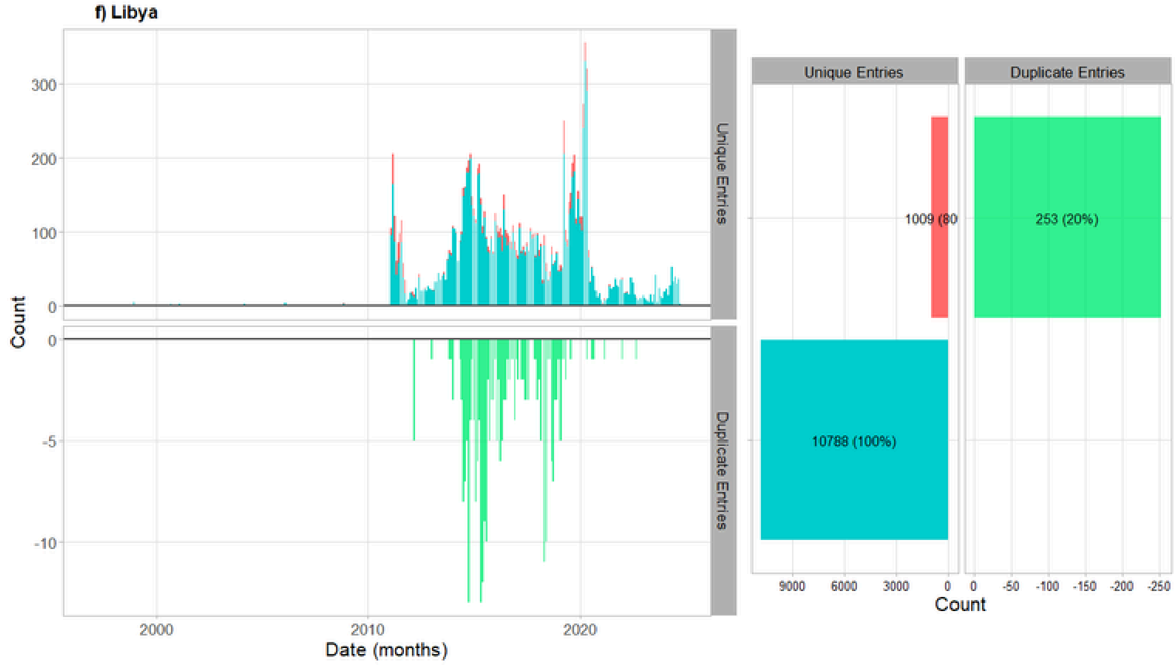
d) Ghana



e) Côte d'Ivoire



f) Libya



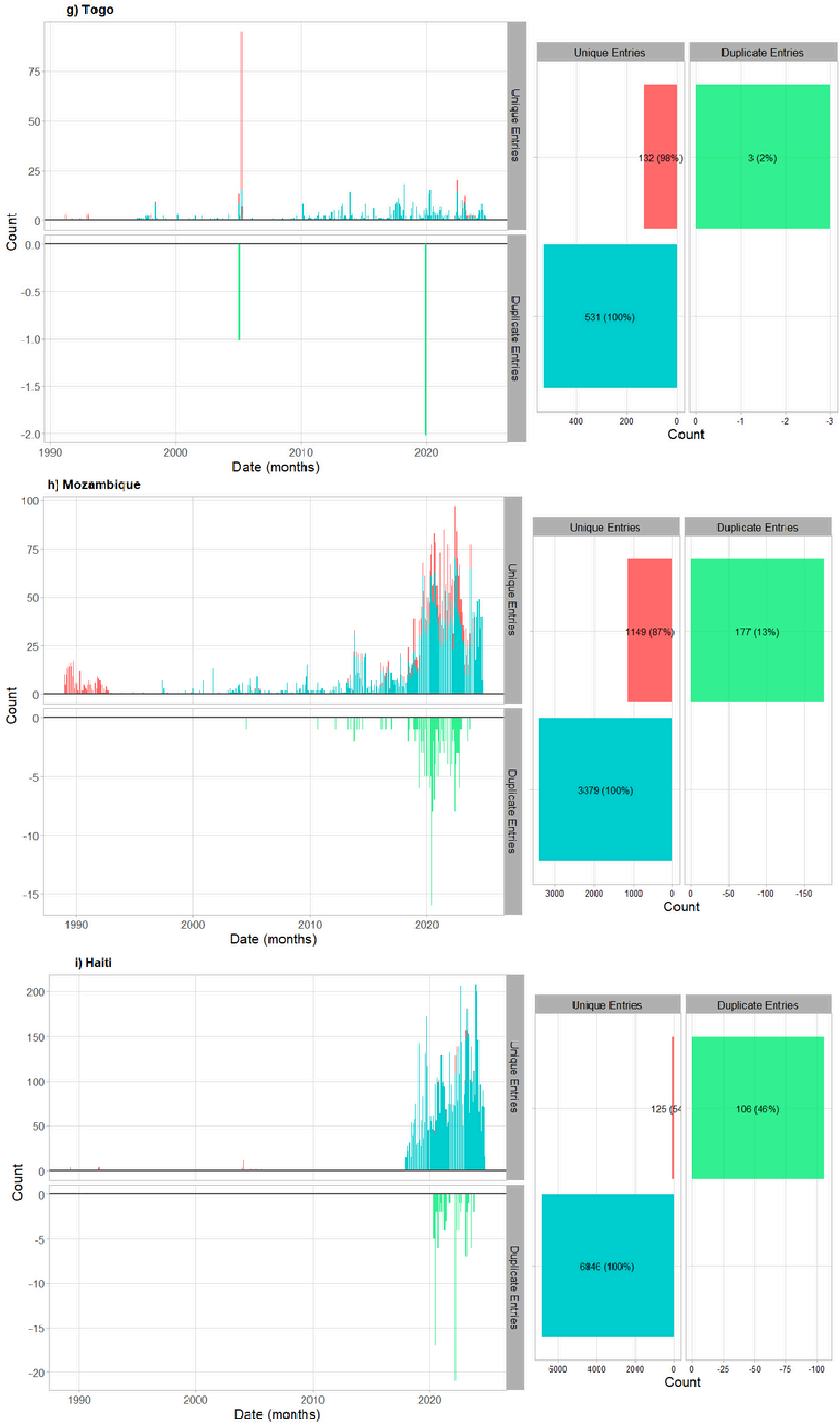


Figure 3: Time series and bar plots for the selected countries. In subfigure panels a) to i), the bar plot or the right visually represents the unique and overlapping entries. The entries from the primary dataset, ACLED, are depicted in blue. All matching (or duplicate) entries are shown with respect to the preceding dataset. Therefore, any matches identified in UCDP-GED (which is indicated in red) are relative to ACLED. On the left side of sub-figures a) to i), the time series plots display the customised MELTT output. These plots provide a mirrored view: unique entries are plotted above the axis, while removed duplicates are shown below, giving a clear snapshot of when duplicates were identified.

Table 3. Summary of MELTT Output per Country of Interest

Affiliation	Country	Total Number of Input Observations *	Number of Unique Observations (after deduplication)	Number of Unique Matches	Number of Duplicates Removed
GFA	Benin	1251	1250	1	1
	Guinea	2778	2772	6	6
	Papua New Guinea	965	960	5	5
	Ghana	2263	2258	5	5
	Côte d'Ivoire	3196	3156	40	40
	Libya	12049	11796	253	253
FEWS NET	Togo	666	663	3	3
	Mozambique	4700	4523	177	177
	Haiti	7059	6953	106	106

**The total number of input observations may differ from the sum of event counts in ACLED and UCDP-GED as shown in Table 1, due to the exclusion of entries categorised as 'Strategic Developments'*

Figure 3 and Table 3 provide a summary and visual of the data processing outputs for the countries of interest. Among the GFA countries, Benin processed the third lowest number of input observations, removing only one duplicate. Guinea yielded 2,772 unique events after eliminating six duplicates. Papua New Guinea removed five duplicates from the second lowest number of input observations, while Ghana also removed five duplicates from 2,263 input observations. Côte d'Ivoire removed 40 duplicates from 3,196 observations, ranking fourth in the number of duplicates removed, with a notable spike in duplicate entries around 2011 (see Figure 3e). Libya had the highest number of input observations and duplicates at 12,049 and 253 respectively, with most entries and duplicates concentrated between 2010 and 2020 (see Figure 3f).

For the FEWS NET countries, Togo processed 666 input observations, removing three duplicates. Mozambique had 4,700 input observations -the third highest among these countries- and removed 177 duplicates, primarily from 2018 onward (see Figure 3h). Haiti had the second highest number of input observations at 7,059 and removed 106 duplicates, with the majority identified from 2020 onward.

4. ADVANCEMENTS OF THE INTEGRATED CONFLICT DATASET OVER PREVIOUS APPROACHES

As outlined in this report, during the integration of conflict event data for Mozambique, a manual actor classification approach was applied, utilising additional levels within the hierarchical taxonomy inspired by VEHICLE. This approach was compared to the original MELTT methodology described in Donnay et al. (2019), which relies on more general, keyword-based classifications. The comparison highlighted the importance of taxonomical richness, particularly in regions like Mozambique, where actor naming conventions differ significantly across datasets.

As shown in Table 2, there are 63 unique actors from ACLED and 6 unique actors from UCDP for Mozambique. Using the approach outlined in the supplementary material from Donnay et al. (2019), these actors are grouped into three broad categories: violent groups, nonviolent groups, and unknown, which on totality, represent a taxonomy only two levels deep.

By contrast, the manual refinement of taxonomies in this project, as shown in Figure 2 a-c, offers a much more detailed categorisation with additional hierarchical levels. The use of more levels within taxonomies allowed for a more nuanced differentiation of actors beyond the simplistic groupings employed in the original MELTT protocol. For example, ACLED’s actor naming conventions for the political party and militant group RENAMO are significantly more detailed than those in UCDP-GED. ACLED lists actors such as "RMJ: Renamo Military Junta", "RENAMO: Mozambican National Resistance", and "Former RMJ: Renamo Military Junta", whereas UCDP labels them simply as "Renamo". This discrepancy serves as an example and illustrates the limitations of the keyword-based classification approach.

When applying the original MELTT procedure with its keyword algorithm, the UCDP-GED actor "Renamo"would likely be categorised as unknown. In addition, the presence of the word "Military" in ACLED's actor labels (e.g., "RMJ: Renamo Military Junta") would incorrectly lead to classification as government, conflating military actors with state-led violence. These misclassifications underscore the inadequacy of the keyword-based approach in cases where detailed actor names are necessary for accurate matching.

Donnay et al. (2019) attempted to address this issue by adjusting for UCDP's nuances, proposing the use of the side-b actor (non-state actor) in state-based violence to handle such inconsistencies. However, even with this adjustment, the results were not optimal for the Mozambique dataset. The broad classification scheme still misrepresented actors like RENAMO, which reduced the accuracy of event matching.

In contrast, by manually refining and adjusting the taxonomies for this analysis, a more accurate actor classification was achieved. This is especially evident in the detection of duplicate entries. The manual approach led to a substantial increase of 10% in identifying duplicate entries, as illustrated in Figure 4. This improvement demonstrates the value of a more granular taxonomy, which enables more precise identification and matching of events, particularly when actor names vary considerably between datasets.

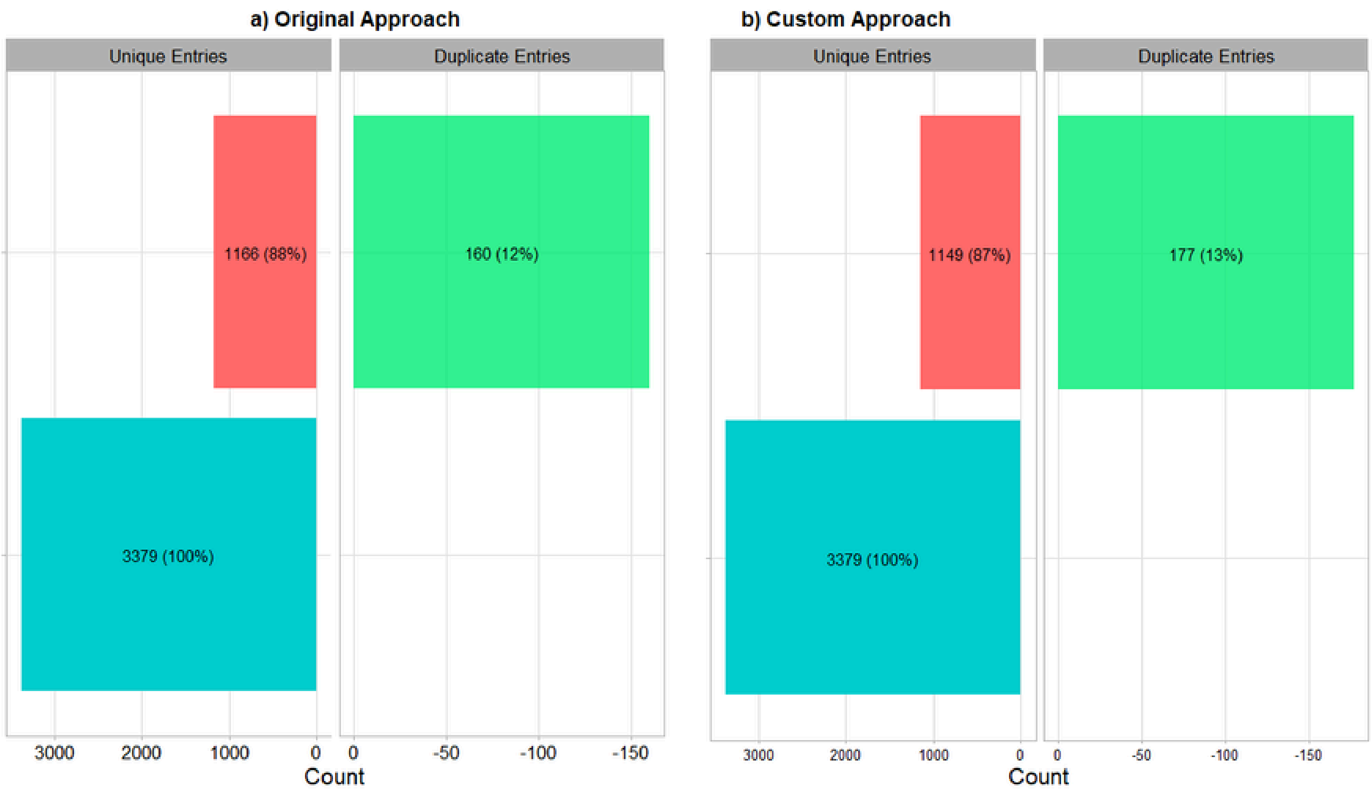


Figure 4: Bar plots for manually refining and adjusting the taxonomies vs. original MELTT methodology. In subfigure panels a) & b), the bar plot visually represents the unique and overlapping entries. The entries from the primary dataset, ACLED, are depicted in blue. All matching (or duplicate) entries are shown with respect to the preceding dataset. Therefore, any matches identified in UCDP-GED (which is indicated in red) are relative to ACLED.

In summary, even though the MELTT procedure's keyword-based classification provides a scalable method for conflict event integration, the manually refined and more detailed VEHICLE taxonomies applied in this project offered a more accurate representation of actor data for Mozambique. This highlights the critical role of taxonomy adjustment and refinement, which, as Donnay et al. (2019) note, is an essential part of the MELTT procedure. However, the results of this analysis demonstrate that a more detailed, manually curated taxonomy can significantly enhance accuracy and deliver denser and more accurate datasets, particularly in regions with complex actor dynamics like Mozambique.

5. STRENGTHENING FOOD SECURITY PREPAREDNESS WITH INSIGHTS FROM THE INTEGRATED CONFLICT DATA

The integrated dataset is especially valuable for both operational and analytical applications in conducting conflict impact assessments within fragile contexts. Tracking hotspots of conflict intensity across space and time enables the monitoring of potential pathways leading to the emergence of food insecurity in specific locations. Monitoring of conflict dynamics, particularly their resurgence, can serve as an early warning system to prevent conflict-induced food insecurity.

5.1 The case of Mozambique

In Mozambique, a spatiotemporal hotspot analysis of the co-occurrence of major conflict events and food insecurity can help identify potential characteristics of conflict-related food insecurity. While in April 2013 a clear return of armed conflict was observed in Sofala province, central Mozambique where RENAMO attacked a police post, leading to the killing of several police officers and to further multiple clashes between RENAMO fighters and government forces throughout that year at multiple other locations (Körtl 2015), that conflict escalation was not directly associated with a food insecurity scenario (Figure 5 a-c).

In February 2016, the Southern part of Mozambique experienced a crisis-level food insecurity. A below-average rainfall in the preceding months exacerbated by the 2015-2016 El Niño event led to drought conditions that severely affected agricultural production and impacting food security in these locations (FEWS NET 2016). That year also saw an escalation of conflict events between RENAMO and the government military counter offensives. The violence and conflict escalated to the point where civilians were displaced from their homes, particularly in the central provinces. The movement of people and goods is constrained, particularly affecting the flow of surplus food from the North to the deficit areas, and exacerbating food insecurity. Key transportation routes were disturbed by conflict attacks, hampering the movement of food supplies and agricultural products to markets, leading to shortages and price increases (FEWS NET 2016; Vhumbunu 2017). In contrast to the 2013 scenario, the 2016 situation illustrates how a resurgence of conflict acts as a crisis multiplier, exacerbating the vulnerabilities of already food-insecure regions and extending food insecurity crisis to neighbouring locations. (Figure 5 d-f).

In 2020, Mozambique experienced significant conflict events primarily driven by the intensification of ongoing violence in the northern province of Cabo Delgado. Violent conflict events were characterised by armed groups attacks on villages, facilities and military outposts, particularly in the Town of Palma in March 2020. An escalation of violence followed when the Mozambican government responded to the attacks by intensifying military operations in Cabo Delgado (IPSS 2020). Food insecurity in that region was caused by multiple interrelated factors such as the Covid-19 pandemic, economic challenges but most importantly the ongoing conflict and violence. The conflict led to the destruction of livelihoods and to forced displacement further diminishing food production capacity. Agricultural activities were severely disrupted, leading to immediate and long-term impacts on food security (FEWS NET 2020a). This situation illustrates a scenario where conflict intensification was the main driver of food insecurity (Figure 5 g-i).

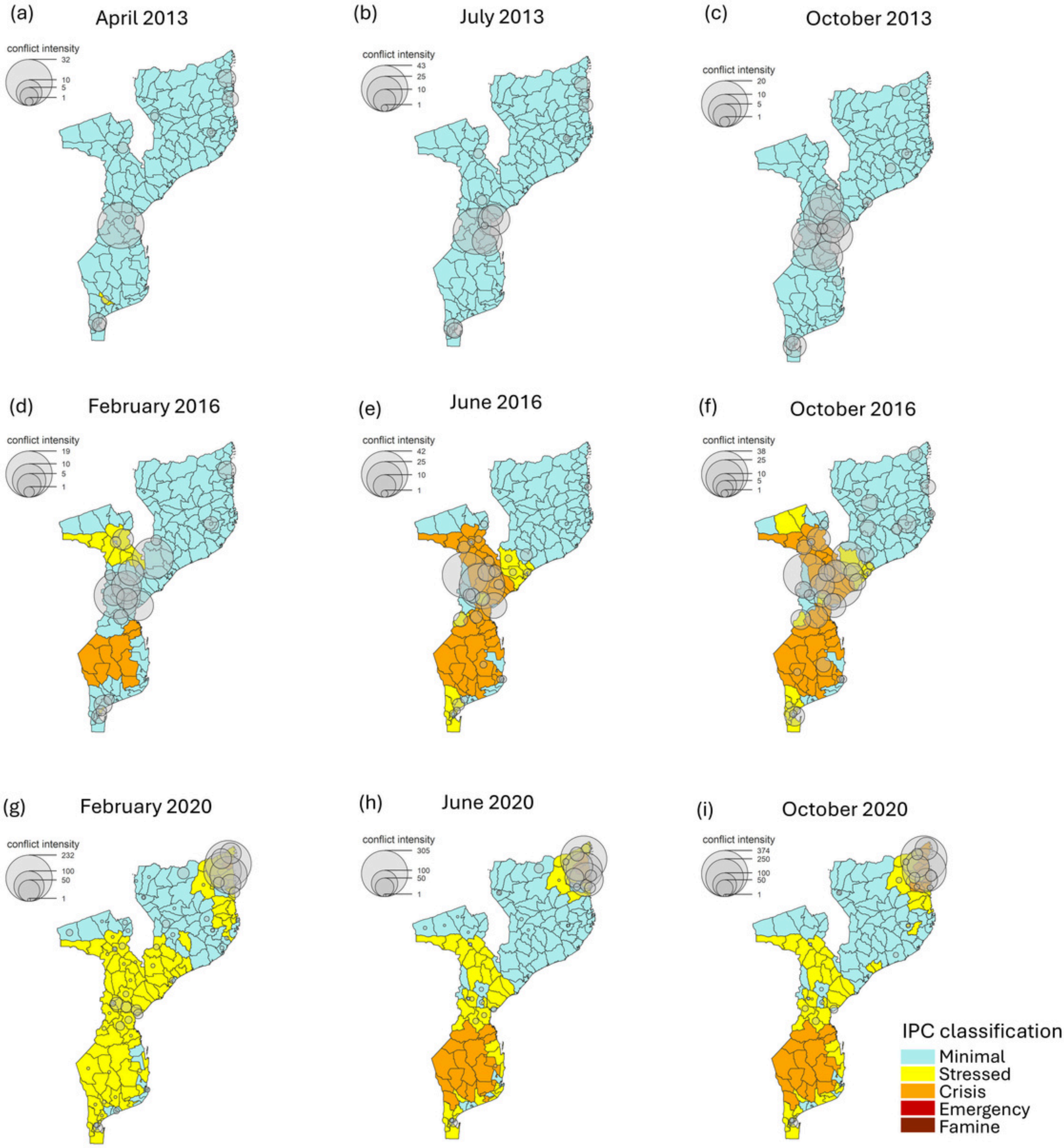


Figure 5: Food insecurity continues to worsen across Mozambique while conflict hotspots are localised mainly in the north. Conflict intensity measures the number of conflict events weighted by their number of fatalities during the past 6 months. Conflict intensity was aggregated at the second-level administrative unit (district). IPC (Integrated food security phase classification) categories by FEWS NET indicate the severity and magnitude of food insecurity and malnutrition in a specific area.

5.2 The case of Haiti

The conflict landscape in Haiti differs from that of Mozambique as it is defined by widespread gang violence and territorial control, political instability, violent demonstrations and escalating economic and humanitarian crises (Patel et al 2023). While the most intense conflict events are concentrated in and around the capital, Port-au-Prince, significant violence has also been reported across nearly all second-level administrative units (communes) since 2018 (Figure 6). FEWS NET has reported a worsening widespread food insecurity across Haiti, with many regions experiencing crisis-level food insecurity since 2018 (Figure 6). A combination of environmental, social, political, and economic factors is driving food insecurity in Haiti (FEWS NET 2020b; FEWS NET 2021). First, due to frequent below-average rainfall during the agricultural seasons, and pluvial flooding caused by frequent hurricane and tropical storms, particularly in the northern and southern regions, key staple production is reduced, further exacerbating food insecurity particularly in rural areas. Secondly, widespread gang violence and the political instability contributed significantly to Haiti's food insecurity. While armed gangs gained territorial control beyond urban areas, disturbing agricultural activities and causing cropland and farming abandonment in rural areas, market shortages due to frequent disruption of transport networks further strained food access and increased food prices. Third, food insecurity was further exacerbated by Haiti's long term economic insecurity mainly due to political instability. An ongoing depreciation of the Haitian Gourde causes inflation and makes basic food items increasingly unaffordable for many households. Urban areas, such as Port-au-Prince are particularly affected by the rising cost of living. Hence, for Haiti, worsening food insecurity (Figure 6) is driven by multiple intertwined factors mainly related to the country's political and social insecurity.

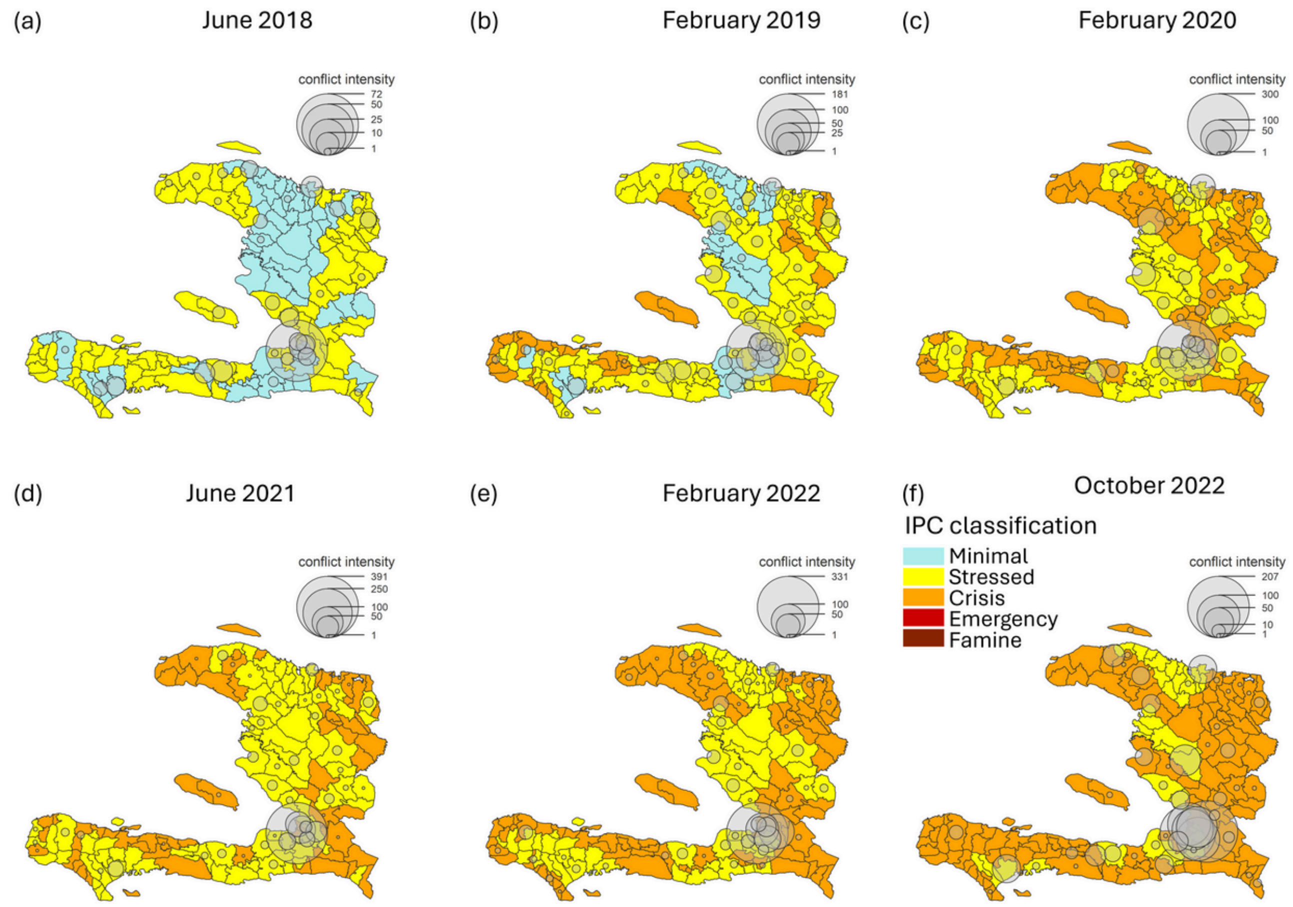


Figure 6: Food insecurity and conflict hotspots are widespread across Haiti and continue to worsen. Conflict intensity measures the number of conflict events weighted by their number of fatalities during the past 6 months. Conflict intensity was aggregated at the second-level administrative unit (commune). IPC (Integrated food security phase classification) categories by FEWS NET indicate the severity and magnitude of food insecurity and malnutrition in a specific area.

6. DISCUSSION

When examining the improved merged dataset, an interesting finding emerges when comparing the number of input observations and the number of duplicates removed. It might be expected that datasets with a higher number of input observations would result in a higher number of duplicates being removed. However, this is not always the case. For example, Haiti had more input observations (7059) than Mozambique (4700), yet only 106 duplicates were removed for Haiti compared to 177 for Mozambique. This discrepancy suggests that the number of input observations alone does not directly determine the number of duplicates removed.

For instance, Mozambique's higher number of duplicates removed suggests that there might have been more overlapping or redundant entries across its datasets. In contrast, Haiti may have had more unique event records, resulting in fewer duplicates being identified. The number of duplicates is likely influenced by other factors, such as the degree of overlap between the different datasets being integrated, the nature of the events captured, and how those events were recorded. In some cases, larger datasets may include more unique events, reducing the likelihood of duplicates, while smaller datasets may capture more similar events across multiple sources, increasing the likelihood of duplication. This highlights the importance of not only looking at raw input counts but also considering the context-specific characteristics, nature, and then the overlap of the datasets being merged.

The integrated dataset is particularly useful for operational and analytical purposes related to conflict impact assessments in fragile contexts. While individual datasets like ACLED and UCDP-GED provide valuable insights into conflict dynamics, their limitations can hinder comprehensive analysis of how conflict impacts food security. Merging datasets from multiple sources enables a broader scope of conflict events, for instance, while ACLED excels in capturing political violence including protest events and non-state actors, UCDP-GED focuses more on organized armed conflicts between state and non-state actors. Together they provide a more nuanced understanding of the conflict landscape that affects food systems. This enables a layered analysis of conflict types, actor interactions and event intensities, enabling better pattern recognition. For instance, distinguishing between frequent low-level incidents and infrequent high-intensity conflicts reveals their distinct impacts on food availability, access and stability.

Furthermore the integration of longer time-series data allows researchers to analyse trends in conflict and food security over extended periods. This temporal dimension is critical for understanding the cyclical or prolonged nature of conflict and its lasting impacts on food insecurity that may not be apparent when analysing shorter or individual datasets. Furthermore, by mitigating report biases present in individual datasets, the merged dataset offers a more reliable representation of conflict dynamics. This comprehensive approach not only enhances predictive insights and forecasting capabilities but also supports more informed operational and analytical purposes in fragile contexts.

By providing a more complete understanding of the actors involved, the types of conflict, and their spatial and temporal characteristics, this dataset supports the needs of organisations such as USAID in quantifying conflict-induced crises and prioritising interventions for applications such as FEWS NET. For example, in a region such as Haiti, where conflict events have direct implications for food insecurity, the integrated dataset helps provide a more nuanced understanding of how conflict impacts the access to resources and the overall stability of affected communities.

In summary, the output of this analysis goes beyond simply merging two conflict event datasets. The purpose is to deliver a valuable resource for improving the quality of conflict analysis and supporting decision-making for stakeholders working in conflict prevention, crisis response, and stability operations. This effort highlights the importance of multi-source integration, actor consistency, and systematic classification to improve the reliability of conflict event data, which is essential for effective monitoring and response strategies in vulnerable regions.

The final output of this integration is a comprehensive dataset that merges conflict event data from ACLED and UCDP-GED, with actor and event data systematically integrated using a customised and improved MELTT framework. This integrated dataset enhances the coverage, accuracy, and granularity of conflict data, specifically for the GFA and FEWS NET countries of interest. By combining these datasets, we mitigate individual dataset limitations, such as incomplete coverage and reporting biases, offering a more nuanced representation of conflict dynamics across these regions.

REFERENCES

- *Davies, Shawn, Garoun Engström, Therese Pettersson & Magnus Öberg (2024). Organized violence 1989-2023, and the prevalence of organized crime groups. Journal of Peace Research 61(4).*

- *Donnay, K., Dunford, E. T., McGrath, E. C., Backer, D., & Cunningham, D. E. (2019). Integrating Conflict Event Data. Journal of Conflict Resolution, 63(5), 1337-1364.* https://doi.org/10.1177/0022002718777050

- *Donnay K, Dunford E (2022). _meltt: Matching Event Data by Location, Time and Type_. R package version 0.4.3, <https://CRAN.R-project.org/package=meltt>.*

- *FEWS NET (2016). Mozambique food security outlook: June 2016 to January 2017. Famine Early Warning Systems Network.* https://fews.net/sites/default/files/documents/reports/Mozambique_FSO_2016_06_Final.pdf

- *FEWS NET (2020a). Mozambique food security outlook: October 2019 to May 2020. Famine Early Warning Systems Network.* https://fews.net/sites/default/files/documents/reports/MOZAMBIQUE_Food_Security_Outlook_October%202019_Final.pdf

- *FEWS NET (2020b). Haiti food security outlook: June 2019 to January 2020. Famine Early Warning Systems Network. https://fews.net/sites/default/files/documents/reports/HT_OL_06_2019_June_EN.pdf*

- *FEWS NET (2021). Haiti food security outlook: June 2020 to January 2021. Famine Early Warning Systems Network. https://fews.net/sites/default/files/documents/reports/HT_FSO_June_January_June30_Final_EN_0.pdf*

- *Gale, D., & Shapley, L. S. (1962). College Admissions and the Stability of Marriage. The American Mathematical Monthly, 69(1), 9–15. https://doi.org/10.2307/2312726*

- *Institute for Peace and Security Studies (IPSS) (2020). Mozambique Conflict Insights: Volume 1, Conflict Insight and Analysis. Better Care Network. https://bettercarenetwork.org/sites/default/files/2021-04/MOZAMBIQUE-Conflict-Insights-vol-1-Conflict-Insight-and-Analysis-1.pdf.*

REFERENCES

- *Körtl, Cornelia (2015). Armed Conflict in Mozambique in 2013: Another Resource Curse? Politikon: The IAPSS Journal of Political Science 27 (July): 36-67. https://doi.org/10.22151/politikon.27.2.*

- *Mayer, B., Lawonn, K., Donnay, K., Preim, B. and Meuschke, M. (2021), VEHICLE: Validation and Exploration of the Hierarchical Integration of Conflict Event Data. Computer Graphics Forum, 40: 1-12.* https://doi.org/10.1111/cgf.14284

- *Patel, R., Gleason, K., Polzin, S., Ripamonti, D., Garvey, A., Dowd, C. (2023) Conflict and Food Systems - Synthesis Report. Concern Worldwide: Dublin https://www.concern.net/knowledge-hub*

- *Raleigh, C., Kishi, R. & Linke, A. Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. Humanit Soc Sci Commun 10, 74 (2023). https://doi.org/10.1057/s41599-023-01559-4*

- *R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL* https://www.R-project.org/.

- *Salehyan, Idean, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. "Social conflict in Africa: A new database." International Interactions 38, no. 4 (2012): 503-511.*

- *START (National Consortium for the Study of Terrorism and Responses to Terrorism). (2022). Global Terrorism Database 1970 - 2020 [data file]. https://www.start.umd.edu/gtd*

- *Sundberg, Ralph and Erik Melander (2013) Introducing the UCDP Georeferenced Event Dataset. Journal of Peace Research 50(4).*

- *Vhumbunu, Clayton Hazvinei. (2017). Suppressing the Revival of Confict in Mozambique through Inclusive National Dialogue.* https://www.accord.org.za/confict-trends/ suppressing-revival-confict-mozambique-inclusive-national-dialogue/

- *Weidmann, Nils B. (2015). On the Accuracy of Media-based Conflict Event Data. Journal of Conflict Resolution 59(6): 1129*

SUGGESTED CITATION

Douglas, W., Basel, A., Minoarivelo., O, Craparo, C., Pacillo, G., Läderach, P. (2024). Integrating Conflict Data for Enhanced Analysis: A Multi-Source Approach to Support Vulnerability Frameworks. Rome, Italy: CGIAR FOCUS Climate Security.

This work was carried out with support from the CGIAR Initiative on Climate Resilience, and the CGIAR Initiative on Fragility, Conflict, and Migration. We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund: https://www.cgiar.org/funders/