

# ITERATIVE OUTLIER ANALYSIS HEURISTIC TO STUDY HIGH IMPACT TERROR ATTACKS OF THE MENA REGION AND EUROPE

PRABAL PRATAP SINGH\* & DEEPU PHILIP†

Department of Industrial and Management Engineering, IIT Kanpur, India

## ABSTRACT

Terrorism perpetrated by both internal and external actors destabilizes the economy, hampers social harmony, and jeopardizes any nation's internal security. While various forms of terror attacks occur globally, most terrorist organizations focus on high impact attacks (HIA) that maximize human fatalities and associated damages to precipitate fear within the community. Realistic modeling of terrorism is essential to identify critical trends and patterns to devise appropriate countermeasures. global terrorism database (GTD) is the most comprehensive terrorism-related database available for research purposes. Comparing terrorist attacks require specific benchmarks. A specific attribute from the GTD (nkill) and a derived composite parameter from GTD (global terrorism index (GTI)) are compared to establish their efficacy in successfully segregating HIAs from other terrorist attacks. This research focuses on the Middle East and North Africa (MENA) region and Europe. A sequential filtering heuristic based on local point outlier analysis utilizes both measures to extract HIAs of the MENA region. The proposed heuristic using nkill reports 1,055 unique HIAs in comparison with GTI and 522 vice versa. A simple geospatial analysis of these HIAs indicates the most fertile regions for terror organizations. Terrorism proliferation plots are developed to visually identify the migration timelines of major terrorist organizations that are capable of HIAs from the MENA region to Europe. The study concludes that virulent terrorist groups of the MENA region like Al-Qaida in the Arabian Peninsula (AQAP), Islamic State of Iraq and the Levant (ISIL), Muslim Extremists (ME), etc. migrated to Europe successfully.

*Keywords: high impact attack, global terrorism database (GTD), global terrorism index (GTI), outlier analysis, sequential filtering heuristic, MENA region, terrorist migration, geospatial analysis.*

## 1 INTRODUCTION

Terrorism is a global threat that jeopardizes the peace, development, and economic progress of a nation. From 2000 to 2018, terrorism cost the world economy an estimated \$US 855 billion [1]. In these 18 years, countries in the Middle East and North Africa (MENA) region have witnessed the highest level of terrorism due to various armed conflicts. MENA region is the cradle of many notorious terrorist groups like Islamic State (IS), Al-Qaida in the Arabian Peninsula (AQAP), etc., which are responsible for many devastating domestic and transnational terror attacks [2]. For example, ISIS (Islamic State of Iraq and Syria), also known as ISIL (Islamic State of Iraq and the Levant), is a Sunni jihadist group with extreme violent ideology. Al-Qaida inspired ISIS origins. Agencies are actively analyzing ISIS's financing, management, and organization to overcome various threats posed by it. The European Union's expenditure to strengthen law enforcement agencies has increased over the years to tackle this menace from its immediate south [3].

A database like global terrorism database (GTD) [4] can foster studies on terrorism trends to develop various insights that can help to devise appropriate countermeasures. GTD follows an annual database release schedule and reports an ensemble of terror incidents from 1970

\* ORCID: <https://orcid.org/0000-0002-0738-7629>

† ORCID: <http://orcid.org/0000-0002-4607-9020>



using various sources, as shown in Fig. 1. GTD contains details of 181,691 terrorism incidents with 135 attributes for each terror incident collected by different organizations such as Pinkerton Global Intelligence Services (PGIS), Centre for Terrorism and Intelligence Studies (CETIS), Institute for the Study of Violent Groups (ISVG), and the National Consortium for the Study of Terrorism and Responses to Terrorism (START) [5].

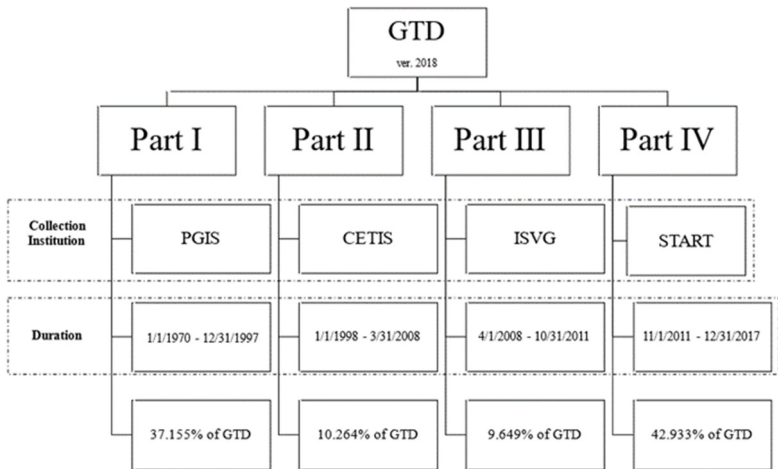


Figure 1: Constituents of global terrorism database (GTD).

Terrorism is a contested concept, and hence, its definition changes from one region to another [6]. GTD utilizes its own definition to classify and include an event as a terrorist attack. It defines terrorism as the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation [5].

Global terrorism index (GTI) [7] utilizes the latest 5-year incident information from the GTD to rank nations based on the level of terrorism faced by them. GTI is a composite score synthesized using GTD's three main attributes: (i) number of fatalities (nkill); (ii) number of people wounded (nwound); and (iii) the property damage reported in the attack (propextent) as shown in eqn (1):

$$GTI - IS = 3 \times nkill + 1 \times nwound + 2 \times propextent + 1 \times incidence. \quad (1)$$

Next, a weighted average of the GTI impact score (GTI-IS) for the last 5 years is calculated to obtain a weighted impact raw score (WIRS). Then, the WIRS is scaled to a 1–10 interval to rank a nation, thereby creating GTI index.

The main aim of any terrorist attack is to instill fear in the target population by inflicting pain and misery on its victims. An event that results in a large magnitude of pain and misery, known as a high impact attack (HIA), requires coordination, planning, sophisticated weapons, trained and motivated individuals. Such attacks are infrequent and hence usually considered as anomalies or outliers in usual statistical analysis.

Statistically, an outlier is an observation that deviates significantly from other observations in the dataset, thereby arousing suspicion that it was generated by a different mechanism [8]. A survey on outlier detection methods by Blázquez-García et al. [9] reports

that these techniques vary based on input data type, the outlier extraction process, and the technique's nature.

Box-and-Whisker plot is a statistical tool that utilizes five number diagrammatic summary: (i) minimum; (ii) first quartile (25th percentile); (iii) median (50th percentile); (iv) third quartile (75th percentile); and (v) maximum on any univariate data series, which can also retrieve outliers [10]. Box-and-Whisker plot provides two sets (lower and upper) of fences: (i) inner fences; and (ii) outer fences constructed using the interquartile range (IQR). The data points beyond the lower and upper outer fences are considered as outliers, while those in between the inner and outer fences are considered potential outliers. However, this study considers data points beyond the upper inner fence (UIF) as outliers, representing the HIAs. Parrinello et al. [11] proposed a similar iterative approach to detect outliers in the laboratory recalibration for removing the drift in Uric acid measurements from the atherosclerosis risk in communities (ARIC) study. However, they considered data points greater than the three standard deviations from the mean as outliers. Their procedure iteratively determines anomalies until no further outliers are reported during a particular iteration. A similar yet heavily modified iterative approach using the UIF value of the Box-and-Whisker plot was developed for this work.

## 2 METHODOLOGY

Each record in the GTD is a terrorism incident that contains information about casualties, perpetrator information, and details of the attack type, weapon type, target type, etc. Terrorist organization that aims to perpetrate maximum fear tends to maximize fatalities associated with HIAs, which is captured by the *nkill* attribute. The primary aim of this work is to extract HIAs from GTD for a particular region of interest. For this, an iterative outlier analysis (IOA) heuristic that utilizes the UIF value of the Box-and-Whisker plot is developed. With a univariate data series as input, the heuristics extract local point outliers for each year, which are considered as HIAs. The iterative extraction stops when the recalculated UIF values converge for two consecutive iterations, resulting in no event labeled as HIA.

It should be noted that the *nkill* attribute has the highest weight in the composite index GTI-IS as per eqn (1), which inspires an efficacy analysis of both *nkill* and GTI-IS in extracting HIAs. Thus, the proposed IOA heuristic is implemented separately using Python ecosystem, as shown in Fig. 2 on GTI-IS and *nkill* univariate data series to extract HIAs.

The incident records for the MENA region from GTD part IV were used in this study. Using either *nkill* or GTI-IS univariate data series, IOA iteratively computes the UIF values for each year until convergence while extracting local outliers simultaneously. Finally, the algorithm combines these annual outliers to obtain the final HIA dataset. A simple geospatial analytics of HIAs was conducted using the location information of the attack to visualize the fertility of a particular region towards terrorism. Additionally, terrorist attacks by the most active organizations reported in the Europe region are compared with the HIAs of the MENA region to elicit probable migration patterns of such organizations.

## 3 RESULTS

GTD 2018 part IV contains data from 2011 to 2017 that amounts to 42.93% of the total database. As mentioned earlier, this study focuses on attacks reported in the MENA and Europe region. GTD part IV reports 78,005 total incidents, out of which the MENA region contributes 30,246 incidents. Table 1 summarizes the countries of the MENA region along with their respective number of reported terror incidents. Iraq tops the list with 56.68% of all reported attacks of the region, followed by Yemen, Libya, and Syria reporting 9.69%, 7.38%, and 6.71% of the attacks, respectively.



---

**Algorithm 1** Iterative Outlier Analysis

---

**Input:** GTD part IV Dataframe - MENA region  
**Output:** Two datasets containing outliers based on nkill and GTI Impact score  
**Require:** nkill, GTI score

```

procedure IOA(Dataframe with nkill and gti score)
  for each attribute  $attr \in \{nkill, GTI\}$  do
    Divide the dataframe annually
    repeat
      for each year  $i$  with attribute  $attr$  do
         $uif\_val_i = Q3 + 1.5 * (Q3 - Q1)$ 
         $uof\_val_i = Q3 + 3 * (Q3 - Q1)$ 
         $uif\_val\_steady_i \leftarrow uif\_val_i$ 
      end for
    until  $uif\_val$  becomes steady at  $n$  iteration
    repeat
      for each year  $i$  with attribute  $attr$  do
        Initialize  $Set1_i$  &  $Set2_i$ 
        for each incidence  $j$  of current dataset do
          if  $attr_j < uif\_val\_steady$  then
             $Set1_i \leftarrow Set1_i + incidence_j$ 
          else
             $Set2_i \leftarrow Set2_i + incidence_j$ 
          end if
        end for
      end for
    until done
     $OutlierDataframe = \sum_i Set2_i$ 
  end for
end procedure

```

---

Figure 2: Iterative outlier analysis algorithm.

Table 1: Number of terrorist incidents in MENA countries.

Country	Incident count	Percentage
Iraq	17,144	56.682
Yemen	2,930	9.687
Libya	2,232	7.379
Syria	2,028	6.705
Egypt	1,975	6.530
Turkey	1,475	4.877
West Bank and Gaza Strip	710	2.347
Israel	535	1.769
Lebanon	453	1.498
Saudi Arabia	307	1.015
Bahrain	163	0.539
Algeria	116	0.384
Tunisia	88	0.291
Iran	55	0.182
Jordan	27	0.089
United Arab Emirates	3	0.010
Kuwait	3	0.010
Qatar	1	0.003
Morocco	1	0.003



Also, GTD reports attacks in Eastern and Western Europe separately, accounting for 2,249 and 1,576 events, respectively. This study combined both and then compared them with the HIAs extracted from the MENA region to obtain possible insights into the influence of terrorism from one neighboring region to another.

Additionally, the nkill, nwound, and propextent attributes of the GTD were analyzed due to their usage in the GTI-IS equation. Both nkill and nwound attributes exhibit a positive skew of 54.07 and 50.40, respectively, for the MENA region. An empirical cumulative distribution (ECDF) [12] plot of nkill and nwound is given in Fig. 3, indicating that around 80% of the reported fatalities/wounds are in single digits.

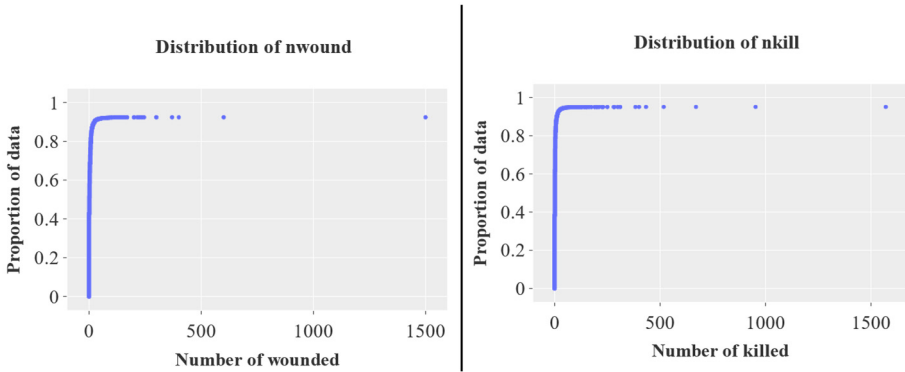


Figure 3: ECDF of nkill and nwound.

Table 2: Category levels for extant property damage in the MENA region.

Damage code		Description	Incident counts
Codebook	GTI		
1	3	Catastrophic	0
2	2	Major	8
3	1	Minor	7,550
4	0	Unknown	22,688

Table 2 summarizes the levels of the propextent attribute with category codes from the GTD Codebook [5] and GTI [7]. No catastrophic property damage from the MENA region is reported for the duration of this HIA analysis. Also, 75.01% of all incident records from the MENA region have an unknown amount of property damage.

The proposed IOA heuristic extracted local point outliers continuously until the UIF value converged between two successive iterations. Typical convergence occurred around the fourth iteration for every annual dataset. The confirmatory fifth iteration ensures convergence of fence values, thereby terminating the heuristic. The five-statistic data summary of the Box-and-Whisker plot for the first and fourth iteration of the heuristic using both nkill and GTI-IS datasets is provided in Table 3.

A comparative analysis of HIAs extracted using nkill and GTI-IS data series is depicted in Fig. 4, which indicates that the proposed IOA can find more HIAs using the nkill attribute than the composite score GTI-IS. About 1,055 HIAs were exclusively captured by the IOA

Table 3: Iteration details of IOA.

Iteration details	nkill outlier iter 1	nkill outlier iter 4	GTI-IS outlier iter 1	GTI-IS outlier iter 4
Count	3,849	4,162	3,029	3,629
Mean	15.67	14.74	73.96	65.66
Standard deviation	39.61	38.23	139.57	128.88
Min	3	3	13	14
25th percentile	6	5	34	30
50th percentile	9	8	47	41
75th percentile	15	14	73	65
Max	1,570	1,570	4,711	4,711

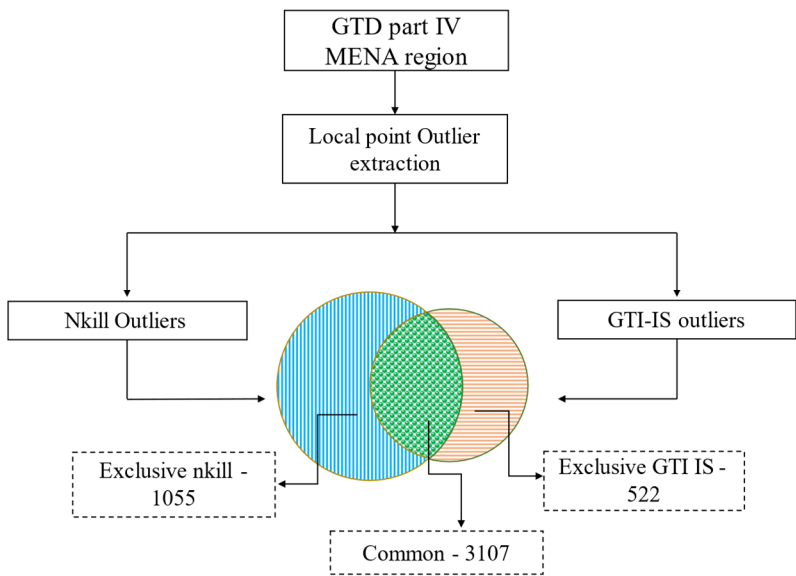


Figure 4: Overview of comparative analysis of outliers.

when using the nkill attribute as input, whereas only 522 HIAs were exclusively captured by the same heuristics while using GTI-IS as the input data. Also, 3,107 HIAs were commonly identified from both data series by the heuristic.

Table 4 compares the Pearson and Spearman’s rank correlation analysis between the nkill and nwound attributes within the three HIAs datasets generated by the IOA: (i) exclusive nkill dataset; (ii) exclusive GTI-IS dataset; and (iii) common HIAs dataset along with the input data. It is evident that the nkill and nwound attributes in the input data are positively correlated, but the exclusive GTI-IS HIAs indicate a negative correlation w.r.t other HIAs.

Geospatial visualization of the extracted HIAs is provided in Fig. 5 for exclusive nkill, Fig. 6 for exclusive GTI-IS, and Fig. 7 for common HIAs to facilitate the visual identification of fertility patterns. It is evident that the exclusive HIAs from nkill (Fig. 5) have similar patterns as common HIAs (Fig. 7).

Table 4: Correlation analysis of three HIAs datasets against input data.

Correlation analysis	Pearson	Spearman
Input data	0.284	0.425
Exclusive GTI	−0.041	<b>−0.365</b>
Exclusive nkill	−0.132	−0.163
Common	0.063	−0.162

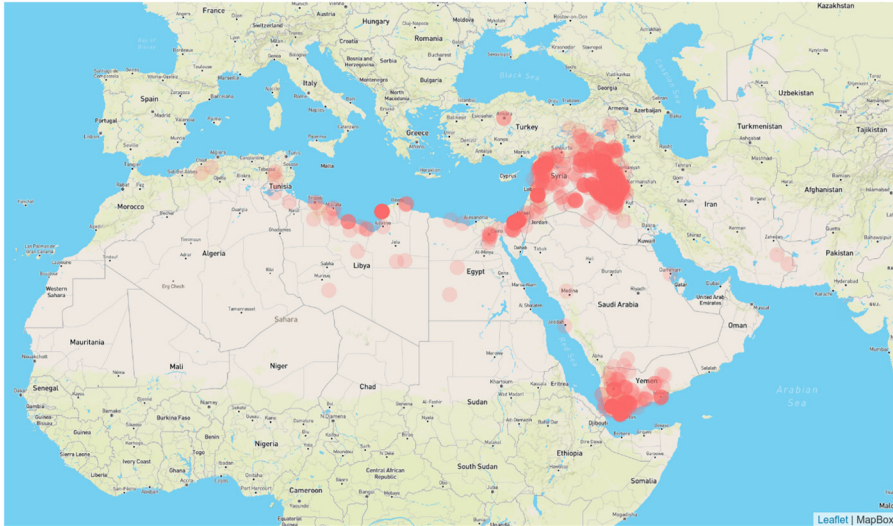


Figure 5: Geographical visualization of exclusive nkill HIAs.

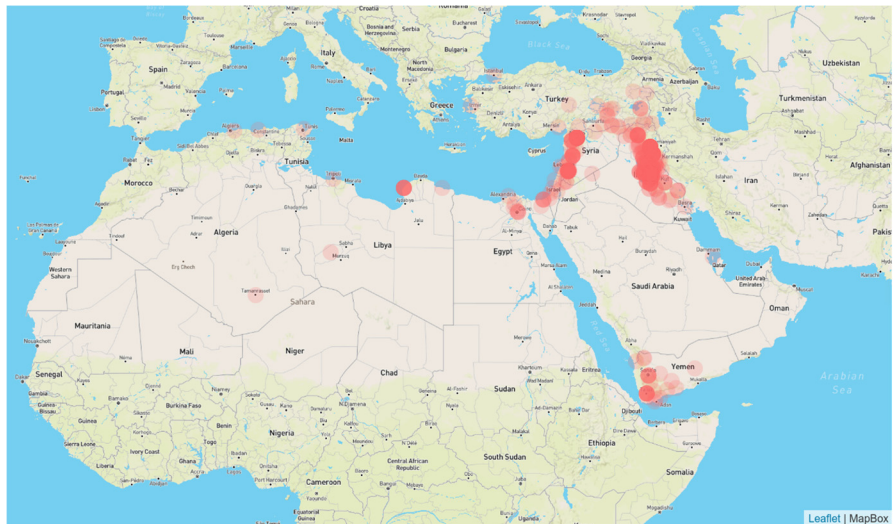


Figure 6: Geographical visualization of exclusive GTI-IS HIAs.



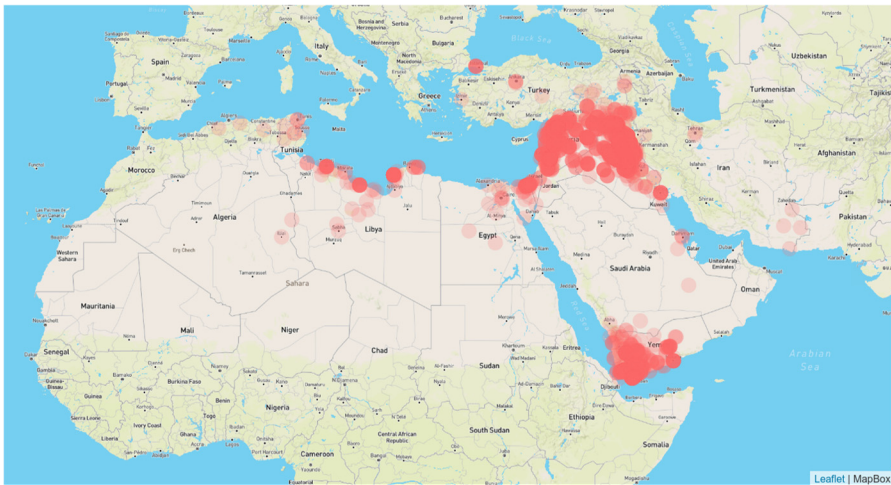


Figure 7: Geographical visualization of common HIAs.

The organizations orchestrating HIAs in the MENA region are of interest. To identify such organizations, aggregating and sorting of the nkill/GTI-IS values for each year was conducted from 2011 to 2017. Each year’s top ten organizations were extracted and combined, resulting in 37 unique active organizations in both datasets. Tables 5 and 6 are the truncated list of the top ten most active organizations obtained from the aforementioned 37 groups against nkill and GTI-IS, respectively. The most active organizations in both Tables 5 and 6 are quite similar, with the top four groups: (i) ISIL; (ii) ANF; (iii) HE; and (iv) AQAP, remaining the same.

Aliases of all the organizations tabulated in Tables 5–8 are as follows: ANF, Al-Nusrah Front; AQAP, Al-Qaida in the Arabian Peninsula; AQI, Al-Qaida in Iraq; Adan IS, Adan-Abyan Province of the Islamic State; CE, Caucasus Emirate; CPIS, Caucasus Province of the Islamic State; DPR, Donetsk People’s Republic; DR, Dissident Republicans; HE, Houthi extremists (Ansar Allah); HTAS, Hay’at Tahrir al-Sham; ISB, Imam Shamil Battalion; ISIL, Islamic State of Iraq and the Levant; JIE, Jihadi-inspired extremists; LPR, Luhansk People’s

Table 5: Most active organizations in HIAs of MENA inflicting highest fatalities (nkill).

Organizations	Year							Total nkill
	2011	2012	2013	2014	2015	2016	2017	
ISIL	0	0	1,045	9,532	7,379	10,770	6,521	35,247
ANF	0	298	676	751	919	193	0	2,837
HE	0	0	54	186	1091	711	417	2,459
AQAP	15	814	258	588	279	115	157	2,226
Sinai IS	0	0	0	0	487	205	588	1,280
AQI	103	729	319	0	0	0	0	1,151
PKK	3	173	0	0	0	288	72	536
Adan IS	0	0	0	0	0	323	71	394
Tripoli IS	0	0	0	0	117	205	0	322
HTAS	0	0	0	0	0	0	313	313





Table 6: Most active organizations in HIAs of MENA reporting highest aggregated GTI-IS.

Organizations	Year							Total GTI-IS
	2011	2012	2013	2014	2015	2016	2017	
ISIL	0	0	5,682	34,214	28,169	39,837	23,194	131,096
ANF	0	1,889	2,284	2,744	3,078	718	0	10,713
HE	0	0	254	757	4,928	2,555	1,682	10,176
AQAP	51	3,067	1,138	2,380	1,095	483	591	8,805
AQI	605	4,175	1,730	0	0	0	0	6,510
Sinai IS	0	0	0	0	1,816	738	2,097	4,651
PKK	11	836	0	0	0	1,857	287	2,991
Adan IS	0	0	0	0	0	1,272	295	1,567
Sanaa IS	0	0	0	0	1,534	0	0	1,534
ME	0	492	0	349	617	0	0	1,458

Table 7: Most active organizations in Europe region inflicting highest fatalities (nkill).

Organizations	Year							Total nkill
	2011	2012	2013	2014	2015	2016	2017	
DPR	0	0	0	1,006	597	30	24	1,657
LPR	0	0	0	173	77	0	0	250
ISIL	0	0	0	4	137	40	27	208
JIE	0	0	0	0	10	112	25	147
CE	0	23	42	28	0	3	0	96
ME	0	28	3	10	9	4	29	83
RS	0	0	0	44	7	0	0	51
CPIS	0	0	0	0	0	12	30	42
ISB	0	0	0	0	0	0	16	16
PRM	0	0	0	11	0	0	2	13

Table 8: Most active organizations in Europe region reporting highest aggregated GTI-IS.

Organizations	Year							Total GTI-IS
	2011	2012	2013	2014	2015	2016	2017	
DPR	0	0	0	3,998	3,079	122	105	7,304
ISIL	0	0	0	0	838	397	243	1,478
LPR	0	0	0	980	431	0	0	1,411
JIE	0	0	0	0	59	878	185	1,122
ME	0	109	39	56	51	39	268	562
CE	0	188	248	122	0	0	0	558
RS	0	0	0	155	38	0	0	193
CPIS	0	0	0	0	0	70	113	183
UN	0	0	0	0	156	0	0	156
DR	0	0	38	22	48	31	16	155

Republic; ME, Muslim extremists; PKK, Kurdistan Workers' Party; PRM, Pro-Russia Militia; RS, Right Sector; Sanaa IS, Sanaa Province of the Islamic State; Sinai IS, Sinai Province of the Islamic State; Tripoli IS, Tripoli Province of the Islamic State; UN, Ukrainian nationalists.

Tables 7 and 8 summarize similar data as that of Tables 5 and 6, respectively, but for the European region. It should be noted that the data for Europe did not undergo the HIA extraction process using the IOA heuristic due to the sparsity in data for Europe in comparison with the MENA region. Usually, a terrorist attack in the European region attracts global attention, resulting in a larger impact compared to the MENA region.

Organizations common in the complete list of the most active organizations for both regions are virulent. ISIL, AQAP, ME, SME, and Gunmen are active in both regions, while the GTI-IS dataset additionally identifies PKK. Fig. 8 depicts the proliferation plot of the activities of these virulent organizations from their host to the migrated region against the nkill attribute. It can be seen from Fig. 8 that ISIL took an additional year before becoming active in Europe.

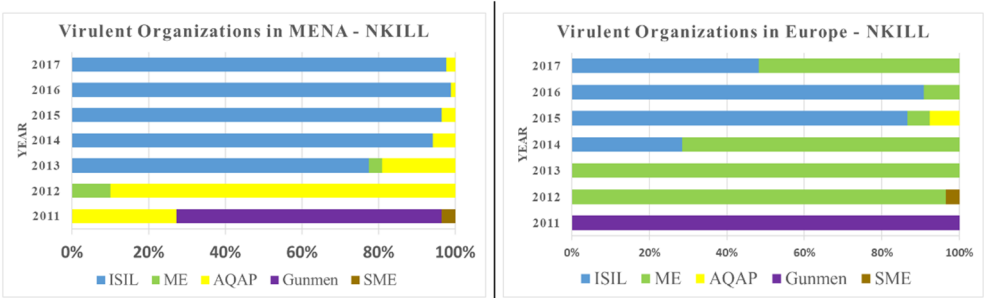


Figure 8: Virulent organization in MENA and Europe based on nkill.

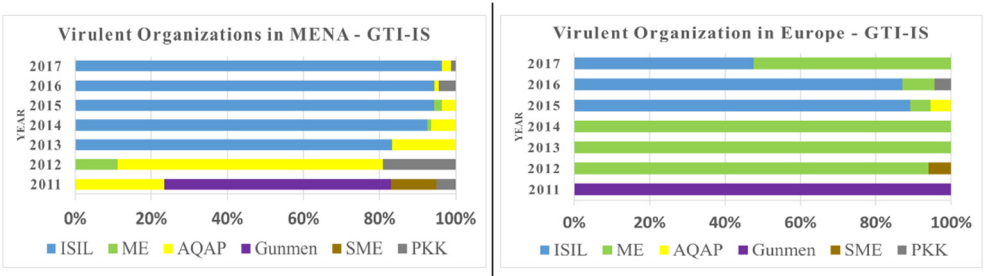


Figure 9: Virulent organization in MENA and Europe based on GTI-IS.

Fig. 9 is a similar proliferation plot as of Fig. 8 but by utilizing GTI-IS as the input. Contrary to the analysis using nkill, Fig. 9 indicates that ISIL took two additional years to emerge in the Europe region.

#### 4 CONCLUSION

This study suggests that using a pure database attribute (nkill) provided a richer set of HIAs than the composite derived GTI-IS value. In totality, the IOA heuristic extracted 14.69% more HIAs using nkill than GTI-IS. We attribute this to the averaging effect of GTI-IS value along with the choice of the coefficients in the GTI-IS calculation in eqn (1). The exclusive nkill HIAs and the common HIAs maintained the same fertility patterns whereas, exclusive GTI-IS HIAs displayed somewhat deviant patterns. Proliferation plots of most active organizations is an effective tool to identify migrational patterns. Further, time-series analysis of HIAs for the MENA region indicates the global threat of ISIS and its potential to migrate to neighboring regions.

#### REFERENCES

- [1] Bardwell, H. & Iqbal, M., The economic impact of terrorism from 2000 to 2018. *Peace Econ. Peace Sci. Public Policy*, **27**, pp. 227–261, 2021. <https://doi.org/10.1515/peps-2020-0031>.
- [2] Kim, W. & Sandler, T., Middle East and North Africa: Terrorism and conflicts. *Glob. Policy*, **11**, pp. 424–438, 2020. <https://doi.org/10.1111/1758-5899.12829>.
- [3] Sofija, V., *The Fight Against Terrorism*, p. 12.
- [4] LaFree, G. & Dugan, L., Introducing the global terrorism database. *Terror Polit. Violence*, **19**, pp. 181–204, 2007. <https://doi.org/10.1080/09546550701246817>.
- [5] GTD Codebook, 2018. <https://www.start.umd.edu/gtd/downloads/Codebook.pdf>. Accessed on: 2 Aug. 2021.
- [6] Schmid, A.P. ed., *The Routledge Handbook of Terrorism Research*, Routledge: London and New York, 2013.
- [7] Hyslop, D. & Morgan, T., Measuring terrorism with the global terrorism index. *Contrib. Confl. Manag. Peace Econ. Dev.*, **22**, pp. 97–114, 2014. [https://doi.org/10.1108/S1572-8323\(2014\)0000022010](https://doi.org/10.1108/S1572-8323(2014)0000022010).
- [8] Hawkins, D.M., *Identification of Outliers*, Springer Netherlands: Dordrecht, 1980. <https://doi.org/10.1007/978-94-015-3994-4>.
- [9] Blázquez-García, A., Conde, A., Mori, U. & Lozano, J.A., A review on outlier/anomaly detection in time series data. ArXiv200204236, Cs Stat, 2020.
- [10] Ramachandran, K.M. & Tsokos, C.P., *Mathematical Statistics with Applications in R*, 3rd ed., Elsevier: Philadelphia, 2020.
- [11] Parrinello, C.M. et al., Iterative Outlier Removal: A method for identifying outliers in laboratory recalibration studies. *Clin. Chem.*, **62**, pp. 966–972, 2016. <https://doi.org/10.1373/clinchem.2016.255216>.
- [12] University of Virginia Library Research Data Services and Sciences, Understanding Empirical Cumulative Distribution Functions. <https://data.library.virginia.edu/understanding-empirical-cumulative-distribution-functions/>. Accessed on: 2 Aug. 2021.

