

# Chapter 17

## Dynamical Evolution of Anti-social Phenomena: A Data Science Approach



Syed Shariq Husain and Kiran Sharma

**Abstract** Human interactions can be either positive or negative, giving rise to different complex social or anti-social phenomena. The dynamics of these interactions often lead to certain spatio-temporal patterns and complex networks, which can be interesting to a wide range of researchers—from social scientists to data scientists. Here, we use the publicly available data for a range of anti-social and political events like ethnic conflicts, human right violations and terrorist attacks across the globe. We aggregate these anti-social events over time, and study the temporal evolution of these events. We present here the results of several time-series analyses like recurrence intervals, Hurst R/S analysis, etc., that reveal the long memory of these time-series. Further, we filter the data country-wise, and study the time-series of these anti-social events within the individual countries. We find that the time-series of these events have interesting statistical regularities and correlations. Using multi-dimensional scaling technique, the countries are then grouped together in terms of the co-movements with respect to temporal growths of these anti-social events. The data science approaches to studying these anti-social phenomena may provide a deeper understanding about their formations and spreading. The results can help in framing public policies and creating strategies that can check their spread and inhibit these anti-social phenomena.

### Introduction

Humans prefer to form groups and act collectively. These groups have evolved from simple settlements in ancient times to huge nations in modern times; defined by multiple causes like languages, common heritage, geographical boundaries, and even

---

S. S. Husain · K. Sharma (✉)  
School of Computational and Integrative Sciences,  
Jawaharlal Nehru University, New Delhi 110067, India  
e-mail: [kiransharma1187@gmail.com](mailto:kiransharma1187@gmail.com)

S. S. Husain  
e-mail: [shariq.iitk@gmail.com](mailto:shariq.iitk@gmail.com)

© Springer Nature Switzerland AG 2019  
F. Abergel et al. (eds.), *New Perspectives and Challenges in Econophysics and Sociophysics*, New Economic Windows,  
[https://doi.org/10.1007/978-3-030-11364-3\\_17](https://doi.org/10.1007/978-3-030-11364-3_17)

ideologies. Often human cooperation has been the motivating force behind the rapid progress of man. This cooperation [1] has extended from blood relatives to totally unrelated individuals. Contrarily, evolution has been responsible for drawing distinctions among themselves in their bids for the “survival of the fittest”. The segregation [2, 3] can be seen in various forms of race, caste, class, religion, political ideology, etc. The assortment of positive and negative aspects of human social behavior makes it extremely complex and convoluted with multiple parameters playing crucial roles. Thus, it is extremely difficult to assess and model the complexity of human social behavior, ranging from bonding, co-operation, support to greed, jealousy, conflict, aggression, coup, war, etc.

Entire world has seen time and again different forms of conflicts, aggression, war, and terrorism, which have plagued mankind from antiquity. Anti-social phenomenon notably possesses very different characteristics than normal social behavior. The interactions among anti-social agents are very low and the occurrences of events tend to be independent of each other. A conflict is an activity which takes place between conscious (not necessarily rational) beings when their interests are mutually inconsistent with each other. A conflict is usually associated with violent activities. The human society has been riddled with conflicts. The first known conflict, a case of inter-group violence, was in eastern Africa around 10,000 years ago as an attempt to seize resources - territory, women, food stored in pots, which resulted in the killing of over two dozen prehistoric men, women, and children [4]. However, as there has been more progress in civilization, humans have become more materialistic and self-centered, and gone beyond competition for tangible resources; they have adopted causes like religion, racial superiority, etc., as pretexts for killing others.

Many people, including Karl Marx and Friedrich Engels, have proposed theories of social conflicts. Apart from social scientists, physicists and data scientists have recently tried to perform in-depth studies and provide mathematical models, statistical and time series analysis of the empirical data and tried to propose potential solutions to the menaces of terrorism, conflicts and other social phenomena, leading to the development of the field of sociophysics [5–8]. Sociophysics is marked by the belief that large-scale statistical measurement of social variables reveals underlying relational patterns that can be explained by theories and laws found in natural sciences, and physics in particular.

In this chapter, we focus on the data dependent statistical analyses of three major anti-social phenomena, viz., ethnic conflicts (EC), human right violations (HR), and terrorism (GTD) [9–11]. An ethnic conflict is a conflict between two or more contending ethnic groups where each group fights for its position within the society on the basis of ethnicity, derived from common descent, culture, language and sometimes, even a common identity. Similarly, a human right violation is said to occur when the basic fundamental rights of a person or a group of persons are infringed upon. Both these anti-social phenomena are based on the conflicts between one or more contesting parties (two sets of actors). However, terrorism differs from these conflicts in the sense that the casualties occurring in a terrorist event are direct or indirect targets of the terrorist groups (sources). The aim of terrorism is not limited to eliminating the target group or destruction of resources, rather it is specifically

carried out to send out a psychological message to the adversary [12]. In other words, unlike in ethnic conflicts and human rights violations, the terrorist attacks are carried out to send across a message to the opponent [13].

Here, we use the publicly available data from: (a) GDELT database [14, 15], which has news reports in media consisting of records of a wide range of socio-economic and political events, viz. ethnic conflicts and human rights violations, over a long period of time, and (b) GTD project [16, 17], which has recorded the terrorist attack incidents that occurred in the last half-century across the globe. We aggregate these anti-social events over time, and study the temporal evolution of these events. We present here the results of several analyses like recurrence intervals, Hurst R/S analysis, etc., that reveal the long memory of these time series [18]. Further, we filter the data country-wise, and study the correlations of these anti-social events within the individual countries. Using the multi-dimensional scaling, we cluster the countries together in terms of the co-movements with respect to temporal growths of these anti-social events. The time series of these events reveal interesting statistical regularities and correlations.

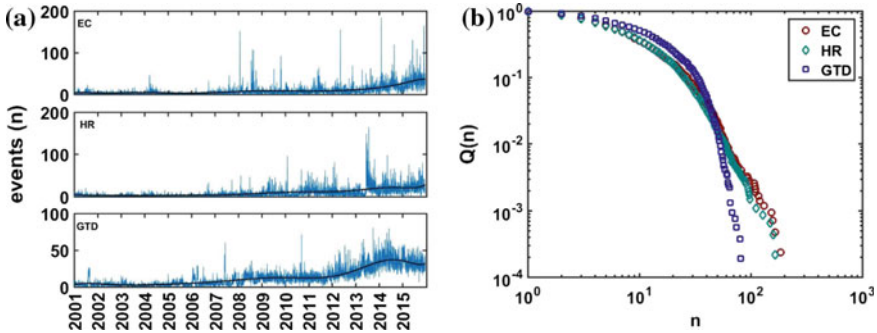
The article is organized as follows. Section “Data description, Methodology and Results” describes the data description, methodology and results in detail. Section “Concluding Remarks” contains the concluding remarks.

## **Data Description, Methodology and Results**

### ***Data Description***

We have used the Global Database of Events, Language, and Tone (GDELT) [14, 15] which is an open source database hosted and managed by GDELT project through *Google Cloud*. GDELT monitors the world’s news media from nearly every corner of every country in print, broadcast, and web formats, in over 100 languages, every moment of every day. The GDELT project is a real-time open database, where the human society is seen through the eyes of the world’s news media, reaching deeply into local events, reaction, discourse, and emotions of the most remote corners of the world. The entire GDELT event database is available and can be extracted using *Google BigQuery*. We filtered all events related to ethnic conflicts (EC) and human rights violations (HR) happening around the world spanning over a large time scale. We procured 45, 942 events for EC and 48, 295 for HR for a 15 year period, 2001–2015.

We have also analysed the data on terrorism events. For the analysis we have used the Global Terrorism Database (GTD) which is an open-source database provides a detailed account of terrorist events around the world from 1970–2017 [16, 17]. The event database is hosted by the National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland. We procured 72, 521 events for the same 15 year period, 2001–2015.



**Fig. 17.1** **a** Plots for the time evolution of the number of events  $n$  reported daily for EC, HR and GTD during the period 2001–2015 with trends (black solid curves). **b** The complementary cumulative density function (CCDF)  $Q(n)$  that  $n$  or more events are reported on a particular day. The data seems to fit well to a stretched exponential of the form ( $\exp[-an^b]$ ) for EC (red circles), HR (green diamonds) and GTD (blue squares), with exponents given in Table 17.2

The list of all the countries analysed, containing names along with their three letter ISO codes, is given in Table 17.1.

## Methodology and Results

The GDELT and GTD data sets contain detailed information about the anti-social events, viz. ethnic conflicts (EC), human rights violations (HR) and terrorist attacks (GTD), on the scale of a day. Our overarching aim is to observe nature of the memory of each of the time-series (EC, HR and GTD) and the cross-correlations among them, within a country. Further, we would like to group the countries together on the basis of their long-term evolution trends and correlations. First, we study simple statistics of auto-correlations, Hurst R/S analysis and recurrence intervals distribution, of the detrended time series. Later, we study the co-movements of the countries on the events spaces using the multidimensional scaling technique.

We have considered the data for the period 2001–2015, and generated daily time series of EC, HR and GTD, as shown in Fig. 17.1a. The black curves show the long time trends, which imply that the time-series are not stationary. To see the spread of the events, we computed the complementary cumulative density function (CCDF) of the events: For the probability density function (PDF)  $P(n)$  as  $n$  reported events per day, the cumulative density function (CDF) is  $F(n) = P(N \leq n)$ ; then, the CCDF is  $Q(n) = 1 - F(n)$ , such that it estimates the probability of the events are above a particular level  $n$ ,  $P(N > n)$ . As the empirical PDFs are often too noisy (specially toward the tails) to be relied upon for statistics, it is known that integrating a signal improves its “Signal-to-Noise ratio”. So, we plot the CCDF as it reduces the noise content and makes the information contained by the signal clearer.

**Table 17.1** List of countries and their 3-letter ISO codes

S. No.	Code	Country	S. No.	Code	Country	S. No.	Code	Country
1	<b>AFG</b>	Afghanistan	36	<b>GEO</b>	Georgia	71	<b>NLD</b>	Netherlands
2	<b>AGO</b>	Angola	37	<b>GHA</b>	Ghana	72	<b>NOR</b>	Norway
3	<b>ALB</b>	Albania	38	<b>GMB</b>	Gambia	73	<b>NPL</b>	Nepal
4	<b>ARG</b>	Argentina	39	<b>GRC</b>	Greece	74	<b>PAK</b>	Pakistan
5	<b>ARM</b>	Armenia	40	<b>HKG</b>	Hong Kong	75	<b>PER</b>	Peru
6	<b>AUS</b>	Australia	41	<b>HRV</b>	Croatia	76	<b>PHL</b>	Philippines
7	<b>AZE</b>	Azerbaijan	42	<b>HTI</b>	Haiti	77	<b>POL</b>	Poland
8	<b>BDI</b>	Burundi	43	<b>HUN</b>	Hungary	78	<b>PRK</b>	North Korea
9	<b>BEL</b>	Belgium	44	<b>IDN</b>	Indonesia	79	<b>RUS</b>	Russia
10	<b>BGD</b>	Bangladesh	45	<b>IND</b>	India	80	<b>RWA</b>	Rwanda
11	<b>BGR</b>	Bulgaria	46	<b>IRL</b>	Ireland	81	<b>SAU</b>	Saudi Arabia
12	<b>BIH</b>	Bosnia-Herzegovina	47	<b>IRN</b>	Iran	82	<b>SDN</b>	Sudan
13	<b>BLR</b>	Belarus	48	<b>IRQ</b>	Iraq	83	<b>SEN</b>	Senegal
14	<b>BRA</b>	Brazil	49	<b>ISR</b>	Israel	84	<b>SLE</b>	Sierra Leone
15	<b>BTN</b>	Bhutan	50	<b>ITA</b>	Italy	85	<b>SLV</b>	El Salvador
16	<b>CAN</b>	Canada	51	<b>JOR</b>	Jordan	86	<b>SOM</b>	Somalia
17	<b>CHE</b>	Switzerland	52	<b>JPN</b>	Japan	87	<b>SRB</b>	Serbia
18	<b>CHL</b>	Chile	53	<b>KEN</b>	Kenya	88	<b>SWE</b>	Sweden
19	<b>CHN</b>	China	54	<b>KGZ</b>	Kyrgyzstan	89	<b>SYR</b>	Syria
20	<b>CIV</b>	Cote D'ivoire	55	<b>KHM</b>	Cambodia	90	<b>TCO</b>	Chad
21	<b>COG</b>	Democratic Republic of the Congo	56	<b>KOR</b>	South Korea	91	<b>THA</b>	Thailand
22	<b>COL</b>	Colombia	57	<b>KWT</b>	Kuwait	92	<b>TUN</b>	Tunisia
23	<b>CUB</b>	Cuba	58	<b>LBN</b>	Lebanon	93	<b>TUR</b>	Turkey
24	<b>CYP</b>	Cyprus	59	<b>LBR</b>	Liberia	94	<b>TWN</b>	Taiwan
25	<b>CZE</b>	Czech Republic	60	<b>LBY</b>	Libya	95	<b>UGA</b>	Uganda
26	<b>DEU</b>	Germany	61	<b>LKA</b>	Sri Lanka	96	<b>UKR</b>	Ukraine
27	<b>DNK</b>	Denmark	62	<b>LVA</b>	Latvia	97	<b>USA</b>	United States
28	<b>DZA</b>	Algeria	63	<b>MDA</b>	Moldova	98	<b>UZB</b>	Uzbekistan
29	<b>ECU</b>	Ecuador	64	<b>MEX</b>	Mexico	99	<b>VEN</b>	Venezuela
30	<b>EGY</b>	Egypt	65	<b>MKD</b>	Macedonia	100	<b>VNM</b>	Vietnam
31	<b>ESP</b>	Spain	66	<b>MMR</b>	Myanmar	101	<b>YEM</b>	Yemen

(continued)

**Table 17.1** (continued)

S.No.	Code	Country	S.No.	Code	Country	S.No.	Code	Country
32	<b>ETH</b>	Ethiopia	67	<b>MNP</b>	Northern Mariana Islands	102	<b>ZAF</b>	South Africa
33	<b>FJI</b>	Fiji	68	<b>MYS</b>	Malaysia	103	<b>ZMB</b>	Zambia
34	<b>FRA</b>	France	69	<b>NAM</b>	Namibia	104	<b>ZWE</b>	Zimbabwe
35	<b>GBR</b>	United Kingdom	70	<b>NGA</b>	Nigeria			

**Table 17.2** Exponent values for events of EC, HR, and GTD

Series	Exponents	
	a	b
EC	0.50 ± 0.01	2.49 ± 0.01
HR	0.76 ± 0.01	2.91 ± 0.01
GTD	0.48 ± 0.01	3.17 ± 0.01

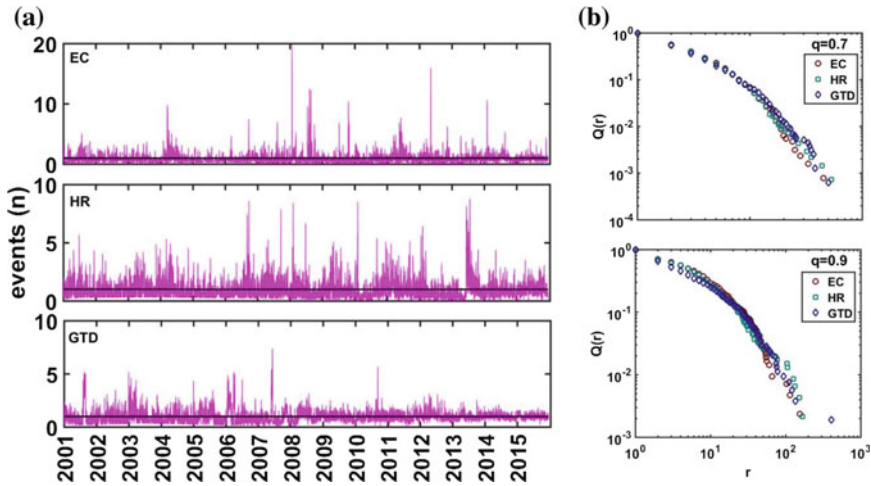
As the number of news entries  $n$  per day is a stochastic variable, we often see bursts of activities for all the three anti-social phenomena: EC, HR and GTD. Due to large inter-day fluctuations in the number of reports and the bursty nature, the CCDF shows a broad distribution. Figure 17.1b shows the plots for the CCDF  $Q(n)$  that  $n$  or more events are reported on a particular day, for the three time-series. Each of the curves is well-fitted by a stretched exponential of the form,  $(\exp[-an^b])$  with exponents given in Table 17.2.

The auto-correlation is the correlation of a signal with a time-delayed copy of itself, as a function of delay or lag. In simple words, it is the similarity between observations as a function of the time lag between them, which can be used for finding repeating patterns or periodicity obscured by noise. The Hurst exponent is a popular measure of long-term memory in a time series, which relates to the auto-correlations of the same and the rate at which these auto-correlations decrease as the time-lag between the pair of values increases.

Extreme events are rare in natural as well as social phenomena, but it is essential to study their properties as the consequences of extreme events are often enormous [19–21]. As researchers, we are often interested in the question that how long would we have to wait for extreme events of a certain magnitude to recur. We thus fix a threshold  $X^{(q)}$  and consider only the events of magnitude higher than  $X^{(q)}$ , where  $q$  denotes the quantile. We define the *recurrence interval* as the time interval between two consecutive extreme events:

$$R_t = \begin{cases} \text{NA,} & X(t) < X^{(q)} \\ \inf \{ \tau > 0 \mid X(t + \tau) \geq X^{(q)} \}, & X(t) \geq X^{(q)} \end{cases}, \quad (17.1)$$

where  $X(t)$  is an event occurring at time  $t$ ,  $X^{(q)}$  is the threshold, and  $\tau$  is a time lag.



**Fig. 17.2** **a** Detrended time series for EC, HR and GTD with a flat trend (black solid line). **b** shows the plots of complementary cumulative density function  $Q(r)$  that  $r$  events recurred at quantiles  $q = 0.7$  and  $q = 0.9$ . The data for quantiles  $q = 0.7$  and  $q = 0.9$  have been found to fit well to stretched exponentials,  $(a \exp[-bn^c])$  with parameters given in Tables 17.3 and 17.4

**Table 17.3** Parameter values for recurrence CCDF at  $q = 0.7$

Series	Parameters		
	a	b	c
EC	$3.53 \pm 0.37$	$1.27 \pm 0.10$	$0.49 \pm 0.03$
HR	$3.69 \pm 0.46$	$1.31 \pm 0.12$	$0.49 \pm 0.03$
GTD	$6.95 \pm 0.71$	$1.94 \pm 0.10$	$0.38 \pm 0.01$

A real data series usually exhibit non-stationarity of various forms such as “seasonal effects”, “trends”, etc. Though it is very difficult to completely eliminate non-stationarity, its effect can be reduced by introducing some corrective measures. Each type of non-stationarity requires a different type of correction. Here, none of the original time series of events is stationary, as can be seen in Fig. 17.1, where the black dashed line shows the inherent trend of the time series. We calculate the trend with a polynomial of degree 10, and then divide the original signal by the computed trend, resulting in a detrended series. Figure 17.2a shows the events time series after detrending it. The black dashed line, which shows trend of this detrended time series, is thus flat. The detrended events time series can be assumed to be weakly stationary.

We hence analyse the CCDF of the recurrence time intervals on the detrended time series (see Fig. 17.2a) to quantify the extreme events or duration of recurrences of an event. Figure 17.2b shows the plots of CCDF  $Q(r)$  that  $r$  events recurred at quantiles  $q = 0.7$  and  $q = 0.9$ . The data for quantiles  $q = 0.7$  and  $q = 0.9$  seem to fit well to stretched exponentials,  $(a \exp[-bn^c])$ , with parameters given in Tables 17.3 and 17.4. It should be mentioned that as the quantile  $q$  increases, the distribution becomes

**Table 17.4** Parameter values for recurrence CCDF at  $q = 0.9$

Series	Parameters		
	a	b	c
EC	$1.67 \pm 0.09$	$0.54 \pm 0.04$	$0.50 \pm 0.02$
HR	$1.69 \pm 0.10$	$0.56 \pm 0.05$	$0.51 \pm 0.02$
GTD	$4.72 \pm 0.72$	$1.58 \pm 0.15$	$0.28 \pm 0.02$

fatter, i.e., lower recurrence time intervals occur less frequently. This observation is quite obviously explained by the fact that at higher values of  $q$  there are fewer extreme events and they are spread apart.

It is often not possible to comprehend certain effects using empirical data. Thus, the results obtained by analyses of empirical data generally need to be compared against standard benchmarks. In such situations, artificial data can be simulated according to required specifications and the simulated data can then serve as reliable benchmarks. Therefore, we first use Gaussian noises (white and fractional) to understand certain effects and use them as benchmarks for comparing the empirical statistics.

Gaussian noise is a statistical noise having a probability density function equal to that of the Normal (or Gaussian) distribution; a special case is the white Gaussian noise (wGn) or Brownian motion, in which the increments (values at any pair of times) are identically distributed and statistically independent (and hence uncorrelated). Thus, it has no auto-correlation for positive lags, and an exponentially decreasing recurrence interval distribution. We illustrate a white Gaussian noise in Fig. 17.3a.

A fractional Brownian motion is a generalization of Brownian motion. The main difference between fractional Brownian motion and regular Brownian motion is that the increments in Brownian motion are independent, whereas increments for fractional Brownian motion are not. A fractional Gaussian noise (fGn) with Hurst exponent  $0 \leq H \leq 1$ , is defined as a continuous-time Gaussian process  $B_H(t)$  on  $[0, T]$ , which starts at zero, has expectation zero for all  $t$  in  $[0, T]$ , and has a covariance function. Mathematically,

$$\forall(t, s) \in \mathbb{R}_+^2, \quad \mathbb{E}[B_H(t)] = 0 \tag{17.2}$$

$$\mathbb{E}[B_H(t)B_H(s)] = \frac{|t|^{2H} + |s|^{2H} - |t - s|^{2H}}{2}. \tag{17.3}$$

The auto-correlation function (ACF) of a fractional Gaussian noise with Hurst exponent  $H$  is given by:

$$ACF(\tau) \rightarrow \frac{|\tau + 1|^{2H} + |\tau - 1|^{2H} - 2|\tau|^{2H}}{2}. \tag{17.4}$$



For a stationary process with auto-correlations decaying  $ACF(\tau) \sim \tau^{-\gamma}$  (long-memory processes), it can be shown mathematically  $\gamma = 2 - 2H$  [22].

Figure 17.3b shows the ACF and PDF of recurrence time intervals for fractional Gaussian noise with Hurst exponent  $H = 0.8$ . If the underlying time series has auto-correlation, then the extreme events are auto-correlated as well. Evidently the presence of the auto-correlation renders the probability density function of the recurrence time intervals to be a stretched exponential, instead of a pure exponential as observed in the case of white Gaussian noise.

The Hurst exponent is a useful statistical method for inferring the properties of a time series. There are various methods to calculate Hurst exponent, which measures the existence of trend or ‘persistence’ or long-range memory present in the time series. We used the rescaled range (R/S) method to compute the Hurst exponent [23]. The rescaled range (R/S) method is calculated for a time series,  $X_1, X_2, \dots, X_T$ , as follows [24]: We first break the long time series with  $T$  data points, into shorter windows of  $n$  data points, such that there are  $m = T/n$  windows. For each of the  $m$  windows of size  $n$ , we have the partial time series  $X_1, X_2, \dots, X_n$ , for which we calculate the rescaled range:

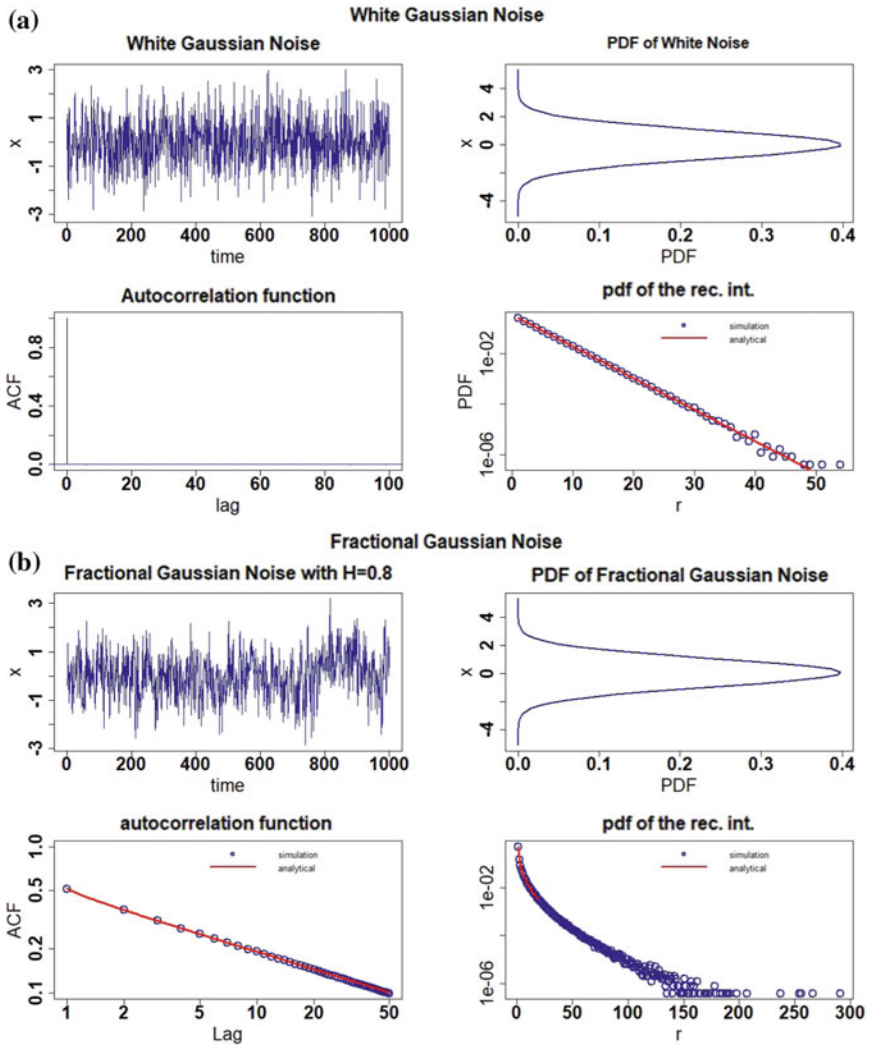
1. Calculate the mean  $\mu = \frac{1}{n} \sum_{i=1}^n X_i$ .
2. Create a mean adjusted series  $Y_i = X_i - \mu$  for  $i = 1, 2, \dots, n$ .
3. Calculate the cumulative deviate series  $Z_t = \sum_{i=1}^t Y_i$  for  $t = 1, 2, \dots, n$ .
4. Compute the range  $R(n) = \max(Z_1, Z_2, \dots, Z_n) - \min(Z_1, Z_2, \dots, Z_n)$ .
5. Compute the standard deviation  $S(n) = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2}$ , where  $\mu$  is the mean for the partial time series  $X_1, X_2, \dots, X_n$ .
6. Calculate the rescaled range  $R(n)/S(n)$  and average over all the partial time series of length  $n$ .

The Hurst exponent is estimated by fitting the power law:  $\mathbb{E}[R(n)/S(n)] = Cn^H$  to the empirical data, where  $C$  is a constant. This can be done by plotting  $\log[R(n)/S(n)]$  as a function of  $\log n$ , and fitting a straight line. The value of the slope gives the Hurst exponent  $H$ , such that

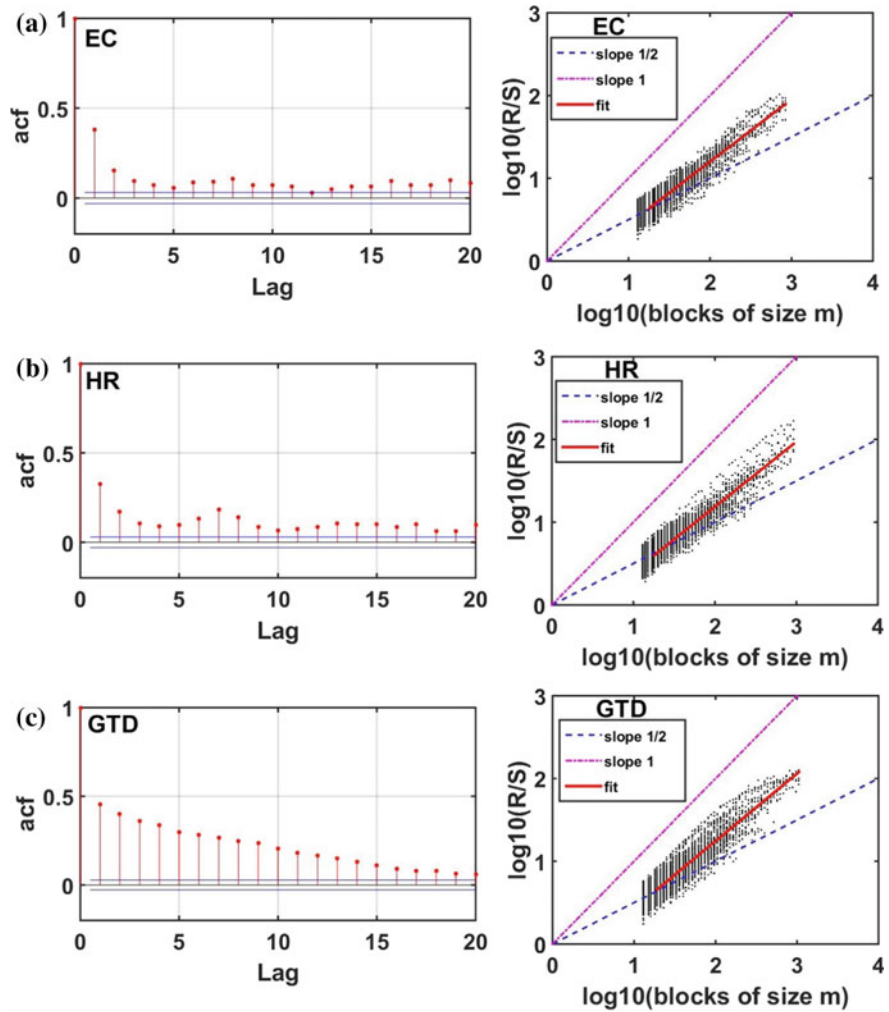
- A value in the range  $0 \leq H < 0.5$  indicates a time series with ‘anti-persistent’ behavior,
- a value in the range  $0.5 < H \leq 1$  indicates a time series with long-term positive auto-correlation (‘persistent’ behavior),
- a value of  $H = 0$  indicates a pink noise,
- a value of  $H = 0.5$  indicates a completely uncorrelated series (Brownian motion).

Figure 17.4 shows the auto-correlation of the detrended time series and Hurst exponent based on R/S analysis having exponent for (a) EC  $0.75 \pm 0.01$ , (b) HR  $0.78 \pm 0.01$  and (c) GTD  $0.82 \pm 0.01$ . The auto-correlation for GTD is decaying exponentially. As the value of the exponent is greater than 0.5 so the time series shows the persistence behavior for all EC, HR and GTD.

Next, we study the co-movements among the different countries across the globe. Figure 17.5a shows the time series of  $n$  events (EC, HR and GTD) during the period of



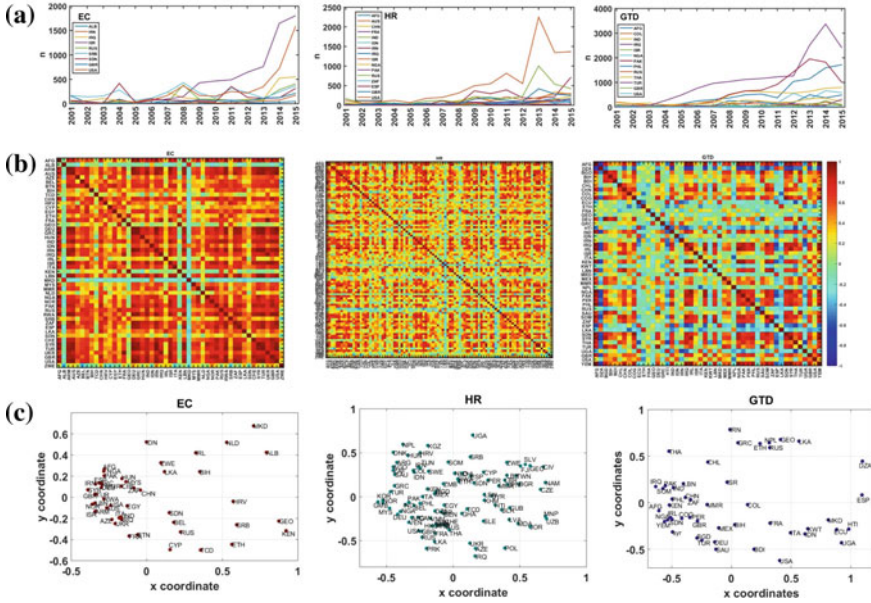
**Fig. 17.3** **a** Time series for white Gaussian noise, probability density function for white noise (independent and identically distributed variables), auto-correlation function of time series, probability density function of recurrence time intervals at  $q = 0.75$ . The time series was generated using the *rnorm()* function in R-software for statistical computing. **b** Time series for fractional Gaussian noise with Hurst index  $H = 0.8$ , probability density function of the fractional Gaussian noise (dependent and identically distributed variables), auto-correlation function of the time series, probability density function for recurrence intervals at  $q = 0.75$ . The time series was generated using the *simFGN()* function in R-software for statistical computing



**Fig. 17.4** Plot for the auto-correlation of detrended time series and Hurst exponent based on R/S analysis having exponent for **a** EC  $0.75 \pm 0.01$ , **b** HR  $0.78 \pm 0.01$  and **c** GTD  $0.82 \pm 0.01$ . As the value of the exponent is greater than 0.5 so the time series shows the persistence behavior for all EC, HR and GTD

2001–2015 for a few countries chosen arbitrarily. We take  $N$  countries and aggregate the events over a year, producing  $T = 15$  data points for the period 2001–2015. To build the correlation matrices, we define the equal-time Pearson cross-correlation coefficient for the time series of the number of events per year  $c_i$  as

$$\rho_{ij}(\tau) = \frac{\langle c_i c_j \rangle - \langle c_i \rangle \langle c_j \rangle}{\sigma_i \sigma_j}. \tag{17.5}$$



**Fig. 17.5** **a** Time series plots for number of events  $n$  of different countries for EC, HR and GTD during the period of 2001–15. The list of country names can be seen in Table 17.1. **b** Correlation matrices for EC, HR and GTD. As number of countries  $N$  are more than length of time series  $T$ , i.e.  $N \gg T$ , the power mapping technique with a distortion  $\varepsilon = 0.6$  is applied to the correlation matrix to suppress the noise. **c** 2D MDS plots for EC, HR and GTD during the period 2001–2015. The MDS plots show the co-movement of the countries: similar countries are grouped together and dissimilar ones placed far apart

where  $\sigma_i = \sqrt{\langle c_i^2 \rangle - \langle c_i \rangle^2}$  is the standard deviation of  $c_i$ ,  $i, j = 1, \dots, N$ , and  $\langle \dots \rangle$  denotes average over the time period  $\tau$ . The elements  $\rho_{ij}$  are restricted to the domain  $-1 \leq \rho_{ij} \leq 1$ , where  $\rho_{ij} = 1$  signifies perfect correlations,  $\rho_{ij} = -1$  perfect anti-correlations, and  $\rho_{ij} = 0$  corresponds to uncorrelated pairs.

It is difficult to estimate the exact correlation among  $N$  time series, each of length  $T$ , as spurious correlations or ‘noise’ are present in the finite time series (see Ref. [25]). The quality of the estimation of true correlation in a matrix strongly depends upon the ratio of the length of the time series  $T$  and the number of time series  $N$ ,  $Q = T/N$ . Correlation matrices are less noisy for higher value of  $Q$ . As  $N > T$ , the corresponding cross-correlation matrices are also singular with  $N - T + 1$  zero eigenvalues, which leads to poor eigenvalue statistics. Thus, we use the power map technique [25–27] to break the degeneracy of eigenvalues at zero and suppress the noise. In this method, a non-linear distortion is given to each cross-correlation coefficient ( $\rho_{ij}$ ) of the correlation matrix  $\rho$  by:  $\rho_{ij} \rightarrow (\text{sign } \rho_{ij})|\rho_{ij}|^{1+\varepsilon}$ , where  $\varepsilon$  is the distortion parameter; here we used  $\varepsilon = 0.6$  (see Refs. [25, 27] for choice of the parameter).

Figure 17.5b shows the correlation matrices (after using the power mapping method), computed over different time series across the countries by using Eq. 17.5. The correlation matrix for EC shows more correlations (colored red) among the countries as compared to anti-correlations (colored blue). The correlation matrices for HR and GTD look very different. In order to visualize the correlations, we apply the multidimensional scaling (MDS) technique. First, we transform the correlation matrix  $\rho$  into distance matrix  $\mathbf{D}$ , as

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}, \quad (17.6)$$

such that  $2 \geq d_{ij} \geq 0$ . After transforming the correlation matrix into distance matrix, we generate the MDS map. The MDS algorithm is used to display the structure of similarity in terms of distances, as a geometrical map where each country corresponds to a set of coordinates in the multidimensional space. MDS arranges different countries in this space according to the strength of the pairwise distances between them. Two similarly behaving countries are represented by two points that are close to each other, and two dissimilarly behaving countries are placed far apart in the map. In general, we choose the embedding dimension to be 2, so that we are able to plot the coordinates in the form of a map. It may be noted that coordinates are not necessarily unique, as we can arbitrarily translate and rotate them, as long as such transformations leave the distances unaffected. Figure 17.5c shows the 2D MDS plots for EC, HR and GTD based on the similarities/ distances among them.

At the end, we also calculated the correlation among the different time series for individual countries. The correlations are computed among EC-HR, EC-GTD and HR-GTD. Few countries like ESP, IDN, ITA and RUS have low correlations among EC-HR, whereas ESP, FRA, ITA, LKA and RUS show anti-correlations for EC-GTD; countries like ESP, FRA, GRC, IDN, ITA, LKA show anti-correlations for HR-GTD. For further details, see Table 17.5. It must be noted that these are just linear correlations, and causal relations cannot be inferred.

## Concluding Remarks

In this paper, our goal was to do the time series analysis and apply data science approaches to the study of the daily anti-social events like ethnic conflicts (EC), human right violations (HR) and terrorist attacks (GTD). As the time series were non-stationary, so we made them stationary by detrending them. We computed the recurrence interval distribution of events and made attempts to relate it with its auto-correlation function. Then we computed the Hurst exponent using the rescaled range (R/S) analyses, which gives the information about whether long memory is present or not. Further, our interest was to study the co-movements of the countries in the respective events spaces. To visualize the co-movements, we computed the cross-correlations among different countries, transformed the correlations into distances and then projected the distances into 2D multidimensional scaling maps.

**Table 17.5** Cross-correlation among events across countries

S.No.	Country	EC-HR	EC-GTD	HR-GTD	S.No.	Country	EC-HR	EC-GTD	HR-GTD
1	<b>AFG</b>	0.72	0.80	0.84	11	<b>IRQ</b>	0.61	0.73	0.85
2	<b>CHN</b>	0.81	0.64	0.89	12	<b>ISR</b>	0.90	0.43	0.33
3	<b>DEU</b>	0.65	0.95	0.49	13	<b>ITA</b>	0.21	-0.05	-0.09
4	<b>ESP</b>	0.47	-0.52	-0.28	14	<b>LKA</b>	0.57	-0.22	-0.39
5	<b>FRA</b>	0.91	-0.08	-0.10	15	<b>NGA</b>	0.73	0.77	0.77
6	<b>GBR</b>	0.81	0.72	0.84	16	<b>PAK</b>	0.83	0.65	0.82
7	<b>GRC</b>	0.91	0.18	-0.01	17	<b>RUS</b>	0.29	-0.07	0.04
8	<b>IDN</b>	0.42	0.69	-0.10	18	<b>TUR</b>	0.97	0.95	0.90
9	<b>IND</b>	0.94	0.85	0.89	19	<b>USA</b>	0.61	0.38	0.11
10	<b>IRN</b>	0.80	0.11	0.45	20	<b>ZAF</b>	0.96	0.21	0.17

**Acknowledgements** The authors would like to thank Anirban Chakraborti, Vishwas Kukreti, Arun S. Patel and Hirdesh K. Pharasi for critical discussions and inputs. KS acknowledges the University Grants Commission (Ministry of Human Resource Development, Govt. of India) for her senior research fellowship. SSH and KS acknowledge the support by University of Potential Excellence-II grant (Project ID-47) of JNU, New Delhi, and the DST-PURSE grant given to JNU by the Department of Science and Technology, Government of India.

## References

1. Perc, M., Jordan, J.J., Rand, D.G., Wang, Z., Boccaletti, S., Szolnoki, A.: Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51 (2017)
2. Schelling, T.: Models of segregation. *Am. Econ. Rev.* **59**(2), 488–93 (1969)
3. Schelling, T.: Dynamic models of segregation. *J. Math. Sociol.* **1**, 143–186 (1971)
4. Lahr, M.M., Rivera, F., Power, R.K., Mounier, A., Copsey, B., Crivellaro, F., Edung, J.E., Fernandez, J.M.M., Kiarie, C., Lawrence, J., Leakey, A., Mbua, E., Miller, H., Muigai, A., Mukhongo, D.M., Van Baelen, A., Wood, R., Schwenninger, J.L., Grn, R., Achyuthan, H., Wilshaw, A., Foley, R.A.: Inter-group violence among early holocene hunter-gatherers of west Turkana, Kenya. *Nature* **529**, 394–398 (2016)
5. Abergel, F., Aoyama, H., Chakraborti, B.K., Chakraborti, A., Deo, N., Raina, D., Vodenska, I.: *Econophysics and Sociophysics: Recent Progress and Future Directions*. Springer, Berlin (2017)
6. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**(2), 591 (2009)
7. Chakraborti, B.K., Chakraborti, A., Chatterjee, A.: *Econophysics and Sociophysics: Trends and Perspectives*. Wiley-VCH, Germany (2006)
8. Sen, P., Chakraborti, B.K.: *Sociophysics: An Introduction*. Oxford University Press, Oxford (2014)
9. Clauset, A., Young, M., Gleditsch, K.S.: On the frequency of severe terrorist events. *J. Confl. Resolut.* **51**(1), 58–87 (2007)

10. Sharma, K., Sehgal, G., Gupta, B., Sharma, G., Chatterjee, A., Chakraborti, A., Shroff, G.: A complex network analysis of ethnic conflicts and human rights violations. *Sci. Rep.* **7**(1), 8283 (2017)
11. Husain, S.S., Sharma, K., Kukreti, V., Chakraborti, A.: Identifying the global terror hubs and vulnerable motifs using complex network dynamics (2018). [arXiv:1802.01147](https://arxiv.org/abs/1802.01147)
12. Richardson, L.: *The Roots of Terrorism*. Routledge, London (2013)
13. Cutter, S.L., Richardson, D.B., Wilbanks, T.J.: *The Geographical Dimensions of Terrorism*. Routledge, London (2014)
14. GDELT - data format codebook v 1.03, as on 25 Aug 2013 (2016). [http://data.gdeltproject.org/documentation/GDELT-Data\\_Format\\_Codebook.pdf](http://data.gdeltproject.org/documentation/GDELT-Data_Format_Codebook.pdf)
15. The global database of events, language and tone (GDELT) (2016). [www.gdeltproject.org/](http://www.gdeltproject.org/)
16. Global terrorism database (GTD)-codebook: inclusion criteria and variables, as on 3 Jan 2013 (2018). <https://www.start.umd.edu/gtd/downloads/codebook.pdf>
17. Global terrorism database (GTD) (2018). <https://www.start.umd.edu/gtd/contact/>
18. Tilak, G.: Studies of the recurrence-time interval distribution in financial time-series data at low and high frequencies (2012)
19. Chicheportiche, R., Chakraborti, A.: Copulas and time series with long-ranged dependencies. *Phys. Rev. E* **89**, 042,117 (2014)
20. Chicheportiche, R., Chakraborti, A.: A model-free characterization of recurrences in stationary time series. *Phys. A Stat. Mech. Appl.* **474**, 312–318 (2017)
21. Santhanam, M., Kantz, H.: Return interval distribution of extreme events and long-term memory. *Phys. Rev. E* **78**(5), 051,113 (2008)
22. Tarnopolski, M.: On the relationship between the hurst exponent, the ratio of the mean square successive difference to the variance, and the number of turning points. *Phys. A Stat. Mech. Appl.* **461**, 662–673 (2016)
23. Chakraborti, A., Santhanam, M.: Financial and other spatio-temporal time series: long-range correlations and spectral properties. *Int. J. Mod. Phys. C* **16**(11), 1733–1743 (2005)
24. Torres, L.R.M., Rojas, A.R., Luévano, J.R., Hernández, R.T.P.: Exponente de hurst en series de tiempo electrosísmicas
25. Pharasi, H.K., Sharma, K., Chakraborti, A., Seligman, T.H.: Complex market dynamics in the light of random matrix theory (2018). [arXiv:1809.07100](https://arxiv.org/abs/1809.07100)
26. Chakraborti, A., Sharma, K., Pharasi, H.K., Das, S., Chatterjee, R., Seligman, T.H.: Characterization of catastrophic instabilities: market crashes as paradigm (2018). [arXiv:1801.07213](https://arxiv.org/abs/1801.07213)
27. Pharasi, H.K., Sharma, K., Chatterjee, R., Chakraborti, A., Leyvraz, F., Seligman, T.H.: Identifying long-term precursors of financial market crashes using correlation patterns (2018). *New J. Phys.* **20**, 103041 (2018)