Modeling and Inferencing for Activity Profile of Terrorist Groups

Vasanthan Raghavan*

Abstract

There has been a continued interest in modeling the activity profile of terrorist groups over the last few decades. Pioneered by Enders and Sandler, initial work in terrorism modeling focused on time-series analysis techniques such as the threshold vector auto-regression (TAR) model. More recent developments in this area have been along two directions. The first framework leverages a self-exciting hurdle model (SEHM), popularized in diverse applications such as seismology and gang warfare modeling, for terrorist activity. The second framework builds a hidden Markov model (HMM) framework to capture terrorist group dynamics. The focus of this work is on a comparative analysis of the SEHM and HMM frameworks in terms of their explanatory and predictive powers. Specific attention is then paid to the inferencing capability of the HMM framework for the early detection of spurts and downfalls in activity.

Key Words: Explanation and prediction, inferencing, spurt detection, changepoint detection, model comparison, terrorism analysis, threshold vector auto-regression model, self-exciting hurdle model, hidden Markov model

1. Introduction

Given a surge in terrorist activity over the last few decades and a corresponding growing interest in modeling and inferencing of the activity of terrorist groups, this paper studies different aspects of these two problems.

Early works on modeling the activity of terrorist groups adopt different classical time-series analyses techniques resulting in diverse modeling frameworks such as the threshold vector auto-regression (TAR), Cox proportional hazards and zero-inflated Poisson models. While terrorist activity can be reasonably accurately captured with these models, these models have been primarily studied in the context of worldwide terrorism trends, rather than for specific terrorist groups. The focus of this paper is on two recent innovations: self-exciting hurdle modeling (SEHM) framework motivated by modeling efforts in seismology and gang warfare, and a hidden Markov modeling (HMM) framework motivated by the need to capture abrupt switches in terrorism dynamics.

From a modeling perspective, this paper addresses the distinctions between the SEHM and the HMM frameworks. It studies the fine nuances in terms of modeling as well as the impact of these nuances on the explanatory and predictive powers of these frameworks on terrorist activity. While the two frameworks appear to have their own unique advantages in terms of explanation, the HMM framework appears to be superior in terms of prediction.

From an inferencing perspective, this paper studies three different approaches for the quick detection of spurts and downfalls in the activity. The first (and simplest) approach exploits the HMM structure and is of parametric nature. While this approach appears to be excellent in terms of inferencing performance, it suffers from disadvantages that render it difficult to be adopted from a practical standpoint. These disadvantages include model learning delays and retrospective (non-causal) state classification. Motivated by these issues, the second approach adopts a changepoint detection view of the problem and uses

*Qualcomm Flarion Technologies, Inc., 3 Petunia Drive, Apt 1F, North Brunswick, NJ 08902, E-mail: vasanthan_raghavan@ieee.org

an Exponentially Weighted Moving-Average (EWMA) algorithm to repeatedly declare a change when a spurt exceeds an appropriate threshold. Though non-parametric, this approach suffers from a significant performance deterioration (relative to the first approach) that allows it to classify only major spurts and/or downfalls. To overcome this difficulty, the third approach further nuances the notion of a spurt by associating it with changes in the resilience or coordination in the group and detecting such changes. This task is eased by developing a majorization theory-based ordering of attack frequency vectors. This approach is not only non-parametric and hence easily adoptable, but also comparable in performance with the parametric scheme.

2. Temporal Modeling of Activity Profiles

The observations capturing terrorist group dynamics come from a complex network that bestows correlations in both time and spatial (network) structure. In general, these observations are multivariate and are of mixed type. Specifically, observations in terrorism modeling are made of categorical, ordinal and interval variables, e.g., time, location, type of ammunition used, (apparent) sub-group of the group involved, intensity and impact of the attacks, etc. In addition, the observations can suffer from non-idealities such as missing data, mislabeled data, temporal and attributional ambiguity, etc.

The first step in terrorism modeling is the development of a temporal model for the activity profile of a terrorist group by discarding the categorical and ordinal variables. In this direction, let the first and last day of the time-period of interest be denoted as Day 1 and Day N, respectively. Let M_i denote the number of terrorism incidents on the ith day of observation, $i = 1, \cdots, N$. Note that M_i can take values from the set $\{0, 1, 2, \cdots\}$ with $M_i = 0$ corresponding to no terrorist activity on the ith day of observation. On the other hand, there could be multiple terrorism incidents corresponding to independent attacks on a given day reflecting a high level of coordination between various sub-groups of the group. Let H_i denote the history of the group's activity till (and including) day i. That is, $H_i = \{M_1, \cdots, M_i\}$, $i = 1, 2, \cdots, N$ with $H_0 \triangleq \varnothing$. The temporal point process model is completely specified if $\mathsf{P}\left(M_i = r | H_{i-1}\right)$ is known as a function of H_{i-1} for all $i = 1, \cdots, N$ and $r = 0, 1, 2, \cdots$.

2.1 Classical time-series methods

Different versions of interrupted time-series analyses have been used to study whether certain strategic policy interventions lead to statistically significant reduction in certain types of attacks and/or if different types of attacks act as substitutes for (or complements of) one another. In particular, the main focus of works such as [1–4] is the study of the efficacy of interventions such as strengthening airport security barriers, fortification of US embassies/missions abroad, US' anti-terrorism laws, international conventions on hijackings, retaliatory bombings, etc.

To be specific, a simple first-order threshold vector auto-regression (TAR) model studying the impact of a certain policy intervention (captured by the indicator function where the policy is in effect and denoted as p_1) on two types of attacks (denoted by the time-series $\{M_{1,i}\}$ and $\{M_{2,i}\}$, respectively) is given as:

$$M_{1,i} = a_1 M_{1,i-1} + b_1 M_{2,i-1} + c_1 \mathsf{p}_1 + \mathsf{Other\ components}, \quad (1)$$

$$M_{2,i} = a_2 M_{2,i-1} + b_2 M_{1,i-1} + c_2 \mathsf{p}_1 + \mathsf{Other\ components}. \quad (2)$$

In general, the two types of attacks are cross-correlated with appropriately chosen model coefficients (a_j, b_j and c_j, $j = 1, 2$) capturing the interdependence between them.

The main conclusion from the TAR modeling approach is that certain policy interventions result in an unanticipated increase in certain types of substitution attacks. For example, installation of metal detectors and airport security barriers that render certain types of attacks more costly for the terrorist group (such as skyjackings) tend to result in the substitution of these attacks with other types of attacks that are less costly for the group (such as other types of hostage events not protected by metal detectors). Another example of this substitution effect is the rise in assassinations of protected persons as a consequence of increased security barriers at US missions abroad, even as kidnappings and hostage events decrease. The net consequence of this study is the identification of a rough 4 and $1/2$-year cycle in terrorism events corresponding to increased terrorist activity (perhaps of a different kind) in response to certain interventions that then results in depletion of terrorist group resources leading to a subsequent phase of low activity.

Other examples of the use of the TAR model include [5–7] and Cox proportional hazards or zero-inflated Poisson models [8,9] for the short- and long-run behavior of worldwide terrorist activity.

2.2 Self-exciting hurdle model (SEHM)

A theoretical foundation for the above-described phenomenon of attack clustering and contagion is provided by the SEHM framework developed in [10, 11]. In its simplest form, the hurdle component of the SEHM creates data sparsity by ensuring a pre-specified density of zero counts, while the self-exciting component induces clustering of data. Self-exciting models have become increasingly popular in diverse fields such as seismology [12], gang behavior modeling [13], and insurgency dynamics [14]. The SEHM used in [10] is described as

$$\mathsf{P}\left(M_i = r | H_{i-1}\right) = \begin{cases} e^{-(B_i + SE_i(H_{i-1}))}, & r = 0 \\ \frac{r^{-s}}{\zeta(s)} \cdot \left(1 - e^{-(B_i + SE_i(H_{i-1}))}\right), & r \geq 1 \end{cases} \tag{3}$$

where B_i is a baseline process, and $SE_i(\cdot)$ is the self-exciting component given as

$$SE_i(H_{i-1}) = \sum_{j\,:\,j<i,\, M_j>0} \alpha_j g(i-j) \tag{4}$$

for an appropriate choice of decay function $g(\cdot)$ and influence parameters $\{\alpha_j\}$. On the other hand, $s \in (1, \infty)$ is an appropriately chosen parameter of the zeta distribution, and $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ is the Riemann-zeta function. While a constant s parameter leads to the simplest modeling framework, s can in general be driven by another self-exciting process. A class described by eight parameters is studied in [10] and it is shown that a four parameter model optimizes an Akaike Information Criterion (AIC) metric for terrorism data from Indonesia/Timor-Leste over the period from 1994 to 2007. This model is shown to accurately capture terrorism data (especially the extreme outliers such as days with $36, 11$, and 10 attacks).

2.3 Hidden Markov model (HMM)

An alternate modeling framework based on HMMs is proposed for the activity profile in [15], where it is hypothesized that M_i depends only on certain hidden states $\boldsymbol{S}_i$ (such as *Intentions*, *Tactics*, or *Capabilities*) in the sense that M_i is conditionally independent of H_{i-1} and $\boldsymbol{S}_1, \cdots, \boldsymbol{S}_{i-1}$ given $\boldsymbol{S}_i$. Further, [15] also hypothesizes a time-homogenous one-step Markovian evolution for $\boldsymbol{S}_i$ with a d-state model to capture the dynamics of the group over time. That is, $\boldsymbol{S}_i \in \{0, 1, \cdots, d-1\}$ with each distinct value corresponding to

a different level in the underlying attribute of the group. Using these two hypotheses, the temporal point process model can be simplified as

$$\begin{aligned}
&\mathsf{P}\left(M_i = r | H_{i-1}\right) \\
&\quad = \sum_{j=0}^{d-1} \sum_{k=0}^{d-1} \mathsf{P}\left(M_i = r,\, \boldsymbol{S}_i = j,\, \boldsymbol{S}_{i-1} = k | H_{i-1}\right) && (5) \\
&\quad = \sum_{j=0}^{d-1} \sum_{k=0}^{d-1} \mathsf{P}\left(M_i = r | \boldsymbol{S}_i = j,\, \boldsymbol{S}_{i-1} = k,\, H_{i-1}\right) \cdot \mathsf{P}\left(\boldsymbol{S}_i = j,\, \boldsymbol{S}_{i-1} = k | H_{i-1}\right) && (6) \\
&\quad = \sum_{j=0}^{d-1} \sum_{k=0}^{d-1} \mathsf{P}\left(M_i = r | \boldsymbol{S}_i = j\right) \cdot \mathsf{P}\left(\boldsymbol{S}_i = j,\, \boldsymbol{S}_{i-1} = k\right). && (7)
\end{aligned}$$

The trade-off between accurate modeling of the group's attributes (larger d is better for this goal) versus estimating more model parameters[1] (smaller d is better for this goal) is resolved in [15] by focussing on mature terrorist groups (where the *Intentions* and *Tactics* attributes remain stable) and by considering a $d = 2$ setting. This trade-off corresponds to a binary quantization of the group's *Capabilities* into *Active* and *Inactive* states. For the observations, a simple model such as the two-parameter *hurdle-based geometric* model, defined as,

$$\mathsf{P}(M_i = r | \boldsymbol{S}_i = j) \triangleq \mathsf{HBG}\left(\mu_j, \gamma_j\right) = \begin{cases} 1 - \gamma_j, & r = 0 \\ \gamma_j (1 - \mu_j) \cdot (\mu_j)^{r-1}, & r \geq 1 \end{cases} \tag{8}$$

can be hypothesized. The intuition behind the hurdle-based geometric model is that the terrorist group remains *oblivious* of its past activity and continues to attack with the same *Tactics* as before, as long as its objective is met, provided a certain group resistance/hurdle has been overcome. The special case where there is no group resistance to this aforementioned strategy is obtained by setting $\mu_j = \gamma_j$, resulting in a geometric observation density. From a class of many one- and two-parameter observation models, [15] shows that the hurdle-based geometric model fits the FARC dataset from RAND Database on Worldwide Terrorism Incidents (RDWTI) [16] best. The FARC dataset captures the activity of the group in Colombia over the nine-year period from 1998 to 2006.

3. Model Comparison: Explanation and Prediction

3.1 Qualitative comparison

While the TAR, SEHM and HMM frameworks assume that the current observation/activity in a terrorist group is dependent on the past history of the group, the models differ in how this dependence is realized. In the TAR model, the current observation is explicitly dependent on the past observations along with (possibly) the impact from other independent variables corresponding to certain geopolitical events/interventions. In the SEHM framework, the probability of a future attack is enhanced by the history of the group according to the formula:

$$\frac{\mathsf{P}(M_i > 0 | H_{i-1})\Big|_{\mathsf{SEHM}}}{\mathsf{P}(M_i > 0 | H_{i-1})\Big|_{\mathsf{Non-SEHM}}} = 1 + \frac{e^{-B_i}}{1 - e^{-B_i}} \cdot \left(1 - e^{-SE_i(H_{i-1})}\right) \geq 1. \tag{9}$$

[1]The model parameters in the HMM framework include the transition probability matrix parameters and the observation density parameters. Thus, the number of parameters is $d(d - 1 + \ell)$ where ℓ is the number of observation density parameters in each state.

The HMM framework combines both these facets by introducing a hidden state sequence. The state sequence depends explicitly on its most immediate past (one-step Markovian structure), whereas the probability of an attack is enhanced based on the state realization.

The TAR model and the HMM are similar from the viewpoint of regime switching as these features are modeled explicitly. However, the mechanism of regime switching is different in the two cases: the former assumes a change in the auto-regressive process, whereas the latter assumes a state transition in the HMM. The SEHM also incorporates a switch between states (induced by the self-exciting component), but this switch is more of an implicit feature of the model rather than an explicit component. More importantly, the TAR model considers global terrorism trends rather than trends constrained to a specific region or a specific group. Similarly, the Indonesia/Timor-Leste dataset considered by [10] is a collation of *all* attacks in Indonesia and Timor-Leste from diverse groups with significantly different activity profiles such as *Dar-ul-Islam*, *Gerakan Aceh Merdeka*, *Jemaah Islamiyah*, etc. On the other hand, the FARC dataset considered in [15] is exclusively the action of the many sub-groups of FARC.

These subtle (yet important) differences lead to distinctive abilities for each framework in terms of the explanatory power (of past attacks) and the predictive power (of future attacks). These aspects are studied next and the power of each framework is illustrated with the FARC dataset studied in [15] and the Indonesia/Timor-Leste dataset studied in [10]. To ease model learning[2], we let T_k, $k = 1, 2, \cdots$ denote the time to the kth day of terrorist activity (with T_0 set to $T_0 = 0$) and define $\Delta T_k \triangleq T_k - T_{k-1}$ to denote the time to the subsequent day of activity (inter-arrival duration of attack days).

3.2 Model learning

For the HMM framework, three one-parameter models[3] (viz. Poisson, shifted zeta and geometric), as well as three two-parameter models (viz. Pòlya, non-self-exciting hurdle-based zeta and hurdle-based geometric) are considered for $\{M_i\}$. With a two-state HMM as an overlay over $\{M_i\}$, model parameters (denoted by the simplistic notation $\boldsymbol{\lambda}_{\mathsf{HMM}}$) are learned with the classical Baum–Welch algorithm [17] to locally maximize the likelihood function of the inter-arrival durations, $\mathsf{P}\left(\Delta T_1^n | \boldsymbol{\lambda}_{\mathsf{HMM}}\right)$. Of these six models, the geometric and the hurdle-based geometric models allow simple recursions for estimates of model parameter(s) via the Baum–Welch algorithm, while the shifted zeta and the hurdle-based zeta distributions capture heavy tails; see [15] for details. Further, the geometric model turns out to be generally the best from a parsimonious sense, whereas the hurdle-based geometric model turns out to be a good fit from among the six models from an AIC perspective.

For the SEHM approach, the different baseline and self-exciting models considered in [10] are used to model $\{\Delta T_1^n\}$. The `fmincon` function in `MATLAB` is used to learn model parameters that maximize the likelihood function, $\mathsf{P}\left(\Delta T_1^n | \boldsymbol{\lambda}_{\mathsf{SEHM}}\right)$ (see [10, Eq. 8]). It turns out that a four parameter model (one parameter for the trend component and three parameters for the negative binomial self-exciting component) is a good model for both datasets.

3.3 Explanatory power

For the explanatory power, we focus on SEHM's and HMM's ability to explain the times to the subsequent day of activity $\{\Delta T_1^n\}$. This is captured by the AIC for the two models,

[2]Model learning with $\{M_i\}$ is problematic since the solution to the subsequent inferencing problem mirrors the randomness in $\{M_i\}$, instead of exposing the macroscopic features of the terrorist group.

[3]These models have support on the non-negative integers.

defined as,

$$\mathsf{AIC}(n)\Big|_{\mathsf{HMM}} \triangleq 2\,k_{\mathsf{HMM}} - 2\,\mathsf{P}\left(\Delta T_1^n|\boldsymbol{\lambda}_{\mathsf{HMM}}\right) \tag{10}$$

$$\mathsf{AIC}(n)\Big|_{\mathsf{SEHM}} \triangleq 2\,k_{\mathsf{SEHM}} - 2\,\mathsf{P}\left(\Delta T_1^n|\boldsymbol{\lambda}_{\mathsf{SEHM}}\right). \tag{11}$$

Note that the AIC score captures the negative of the log-likelihood and thus a model with a smaller AIC score is better than a model with a larger AIC score. Table 1 shows the AIC score comparison between the optimal four parameter SEHM and the optimal HMM for the two datasets. From this study, we see that in terms of explanatory power, both HMM and SEHM frameworks perform reasonably well, with neither framework clearly outperforming the other. The HMM framework is better for the FARC dataset, whereas the SEHM is seen to be better for the Indonesia/Timor-Leste dataset. An explanation for this observation is that the Indonesia/Timor-Leste dataset has a heavier tail than the FARC dataset, which is better captured with the SEHM framework.

Table 1: Comparison between AIC scores with the SEHM and HMM frameworks for the FARC and Indonesia/Timor-Leste datasets.

FARC			Indonesia/Timor − Leste		
n	SEHM	HMM	n	SEHM	HMM
100	671.68	671.06	100	723.78	729.47
200	1117.40	1112.07	165	1091.78	1116.92
300	1521.93	1521.36	200	1283.08	1305.27
400	2127.55	2121.81	250	1589.43	1615.87
450	2333.88	2327.02	300	2018.92	2041.35

3.4 Predictive power

For the predictive power, we focus on each approach's ability to predict ΔT_{n+1} given $\{\Delta T_1^n\}$. For this, we use the conditional mean estimator of the form $\widetilde{\Delta} T_{n+1} = \mathsf{E}\left[\Delta T_{n+1}|\Delta T_1^n\right]$. For the HMM framework, it can be checked that

$$\widetilde{\Delta} T_{n+1}\Big|_{\mathsf{HMM}} = \sum_{i=0}^{1} \beta_i \mathsf{E}\left[\Delta T_{n+1}|\boldsymbol{S}_{n+1} = i)\right], \tag{12}$$

where $\alpha_n(j) \triangleq \mathsf{P}(\Delta T_1^n, \boldsymbol{S}_n = j)$ is updated via the forward procedure [17] and

$$\beta_i = \frac{\sum_j \alpha_n(j)\mathsf{P}(\boldsymbol{S}_{n+1} = i|\boldsymbol{S}_n = j)}{\sum_j \alpha_n(j)}. \tag{13}$$

For the SEHM framework, from (3), we have

$$\widetilde{\Delta} T_{n+1}\Big|_{\mathsf{SEHM}} = \frac{1}{1 - e^{-(B_n + SE_n(H_{n-1}))}}. \tag{14}$$

For the sake of comparison, we also use a sample mean estimator as a baseline:

$$\widetilde{\Delta} T_{n+1}\Big|_{\mathsf{Baseline}} = \frac{1}{n}\sum_{i=1}^{n} \Delta T_i. \tag{15}$$

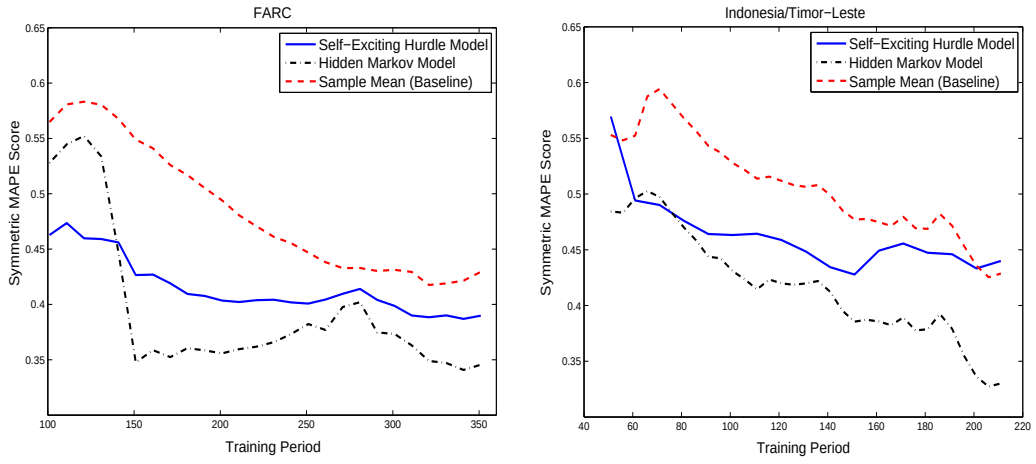
To compare prediction with the three approaches, we use the Symmetric Mean Absolute Percentage Error (SMAPE) score, defined as,

$$\mathsf{SMAPE}(n) \triangleq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\Delta T_i - \widetilde{\Delta} T_i}{\Delta T_i + \widetilde{\Delta} T_i}\right|. \tag{16}$$

Note that the SMAPE score captures the relative error in prediction and is a number between 0% and 100% with a smaller value indicating a better prediction algorithm. The SMAPE scores of the time to the next day of activity for the three estimators (HMM, SEHM, and baseline) are plotted as a function of the training period for model learning in Fig. 1(a) for the FARC dataset and in Fig. 1(b) for the Indonesia/Timor-Leste dataset. Table 2 also shows the SMAPE comparison between the two frameworks for the two datasets. It can be seen from these results that for both the datasets, the HMM framework results in a better prediction than the SEHM and the baseline frameworks provided the training period is long to ensure accurate model learning for the HMM. Further, for the Indonesia/Timor-Leste dataset, even the baseline sample mean estimator outperforms the SEHM estimator for large n.

Table 2: Comparison between SMAPE scores with the SEHM and HMM frameworks for the FARC and Indonesia/Timor-Leste datasets.

FARC			Indonesia/Timor − Leste		
n	SEHM	HMM	n	SEHM	HMM
100	46.27%	52.78%	100	46.33%	43.32%
150	42.95%	35.75%	125	45.47%	41.89%
200	40.40%	35.61%	150	42.84%	38.75%
250	40.09%	38.14%	175	45.23%	38.00%
300	39.92%	37.35%	200	43.46%	33.99%



(a) (b)

Figure 1: SMAPE scores for the three models with a) the FARC dataset and b) the Indonesia/Timor-Leste dataset.

4. Inferencing: Detecting Spurts and Downfalls in Activity

Motivated by the above studies, in the sequel, we hypothesize that the observations (corresponding to terrorist activity) can be accurately described by a $d = 2$-state HMM framework with observations following the hurdle-based geometric model as in (8). Our goal is the early detection of abrupt spurts and downfalls in the activity profile. This is a problem of significant bearing in counterinsurgency operations as well as policy framing.

4.1 Parametric approach: Viterbi algorithm

The simplest approach to leverage the underlying HMM structure is to develop a parametric scheme to classify the hidden states (*Capabilities*) via the use of the Viterbi algorithm [17] with the converged model parameter estimates from the Baum–Welch algorithm on $\{M_i\}$. A notable disadvantage of this approach is that inferencing on the group's *Capabilities* on a daily basis could lead to a performance mirroring the potential rapid fluctuations in the observations. This is particularly disadvantageous in making global policy decisions based on local inferencing of group dynamics.

To overcome this difficulty, we propose inferencing over a $\delta > 1$ day time-window. For this, we decompose the time-period of interest into disjoint time-windows, $\Delta_n,\ n = 1, 2, \cdots, K$, where $\Delta_n = \{(n-1)\delta + 1, \cdots, n\delta\}$ and $K = \lceil \frac{N}{\delta} \rceil$. The appropriate choice of δ is determined by the group dynamics and the timelines for inferencing decisions with typical choices being 7 or 15 days corresponding to a weekly or a fortnightly decision process. We then assume that the hidden state remains fixed over Δ_n:

$$\boldsymbol{S}_i\Big|_{i\in\Delta_n} = s_n,\ s_n \in \{0, 1\}, \tag{17}$$

and our goal is to infer s_n with the aid of an appropriate set of observations corresponding to Δ_n.

To aid in inferencing, we associate a spurt in activity to either a change in the resilience of the group or a change in the level of coordination in the group [18–23] (or perhaps both features). We focus on a set of *attack metrics* that capture the resilience and coordination in the group: i) X_n, the number of days of terrorist activity, and ii) Y_n, the total number of attacks, both within the Δ_n time-window,

$$X_n = \sum_{i\,\in\,\Delta_n} \mathbb{1}\left(\{M_i > 0\}\right); \quad Y_n = \sum_{i\,\in\,\Delta_n} M_i, \quad n = 1, 2, \cdots, \tag{18}$$

where $\mathbb{1}(\cdot)$ denotes the indicator function of the set under consideration. Note that Y_n/δ is the average number of attacks per day and thus Y_n is a reflection of the intensity of attacks launched by the group. In general, X_n is more indicative of resilience in the group, whereas Y_n captures the level of coordination better.

With the hurdle-based geometric model in (8), it can be seen that [15]

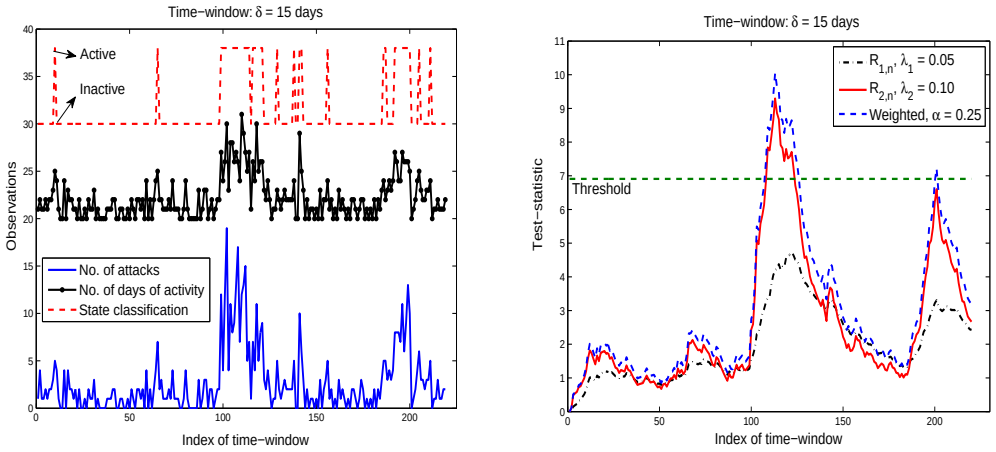
$$\mathsf{P}\left(X_n = k, Y_n = r \Big| \boldsymbol{S}_i|_{i\in\Delta_n} = j\right) = \binom{\delta}{k}\binom{r-1}{r-k}\cdot (1-\gamma_j)^{\delta-k}(\gamma_j)^k \cdot (1-\mu_j)^k(\mu_j)^{r-k},\ r \geq k. \tag{19}$$

Model parameters learned with the Baum–Welch algorithm with $\{(X_n, Y_n)\}$ as observations are then used *retrospectively* (or non-causally) with the Viterbi algorithm for state classification. The output of the Viterbi algorithm is a state estimate for the period of interest

$$\Big\{\boldsymbol{S}_i = \widehat{s}_n \in \{0, 1\} \text{ for all } i \in \Delta_n \text{ and } n = 1, \cdots, K\Big\}. \tag{20}$$

A state estimate of 1 indicates that the group is *Active* over the period of interest, whereas an estimate of 0 indicates that the group is *Inactive*. Transition between states indicates spurt/downfall in the activity.

This approach is applied to the FARC dataset with a $\delta = 15$ day time-window and the results are illustrated in Fig. 2(a). As can be seen from this study, the state classification approach detects even *small* and *non-persistent* changes. However, this performance comes at the cost of model learning (which implicitly assumes model stationarity) and retrospective state classification (that renders it almost impractical from an applications standpoint).



(a) (b)

Figure 2: (a) State classification with the hurdle-based geometric model for the observation sequence $\{(X_n, Y_n)\}$ for the FARC dataset. (b) Performance of the three EWMA tests in spurt detection for the FARC dataset.

4.2 Non-parametric approach: EWMA algorithm

The changepoint detection problem of detecting sudden and abrupt changes in the statistical nature of observations has been studied for over sixty years; see, for example, [24–26] and the references therein for a summary of the state-of-the-art of the area.

Motivated by the rich literature of changepoint detection, we now propose a spurt detection approach based on the Exponential Weighted Moving-Average (EWMA) algorithm. The EWMA algorithm was first introduced by [27] for (continuously) tracking and detecting a change in the mean of a sequence of observations. Here, the test-statistic (R_n) is a first-order auto-regressive version of the observation process (Z_n) to be tracked with smoothing effected by an appropriately-chosen parameter (λ):

$$R_n = (1 - \lambda)R_{n-1} + \lambda Z_n, \ n \geq 1 \tag{21}$$

and $R_0 = 0$. The test-statistic is tested continuously against a threshold A_γ and change is declared at the first instant the test-statistic exceeds the threshold:

$$\tau_{\mathsf{EWMA}} = \inf\{n \geq 1 : R_n \geq A_\gamma\}. \tag{22}$$

A_γ is chosen to ensure that the average run length (ARL) to false alarm is at least γ. Small values of the smoothing parameter λ usually work best in changepoint detection [28, 29] as they smoothen small changes and enhance big changes.

The EWMA framework can be particularized to spurt detection in the activity profile of a terrorist group by repeatedly applying (22) with X_n and Y_n as observations. Two

parameters $\{\lambda_1, \lambda_2\} \in [0, 1]$ are chosen appropriately and used to update the variables $R_{1,n}$ and $R_{2,n}$ as follows:

$$R_{1,n} = (1-\lambda_1)R_{1,n-1} + \lambda_1 X_n \quad (23)$$

$$R_{2,n} = (1-\lambda_2)R_{2,n-1} + \lambda_2 Y_n \quad (24)$$

for $n \geq 1$ with $R_{1,0} = 0 = R_{2,0}$. The best choices of smoothing parameters λ_1 and λ_2 for changepoint detection are obtained experimentally/numerically since the state-of-the-art in EWMA design is such that smoothing parameter design is still open, even for simple models such as Gaussian and exponential densities [29]. We propose three stopping-times for declaring change: one based on $R_{1,n}$, another based on $R_{2,n}$, and the third on a weighted combination (with weights α and $\sqrt{1-\alpha^2}$, $\alpha \in [0, 1]$) of the two test-statistics:

$$\tau_1 = \inf\{n \geq 1 \ : \ R_{1,n} \geq A_1\} \quad (25)$$

$$\tau_2 = \inf\{n \geq 1 \ : \ R_{2,n} \geq A_2\} \quad (26)$$

$$\tau_{\mathsf{weighted}} = \inf\left\{n \geq 1 \ : \ \alpha R_{1,n} + \sqrt{1-\alpha^2} R_{2,n} \geq A\right\}, \quad (27)$$

where the thresholds A_1, A_2, and A are chosen to meet the corresponding ARL constraints. While design of the threshold requires further work, experimental studies suggest that a threshold of the form

$$\{A_1, \ A_2, \ A\} = \mathcal{O}\left(\log(\gamma)\right) \quad (28)$$

ensures that $\{\mathsf{ARL}(\tau_1), \ \mathsf{ARL}(\tau_2), \ \mathsf{ARL}(\tau_{\mathsf{weighted}})\} = \mathcal{O}(\gamma)$. Since τ_{weighted} combines the information contained in both $\{X_n\}$ and $\{Y_n\}$, it should empirically be a better test than both τ_1 and τ_2. Nevertheless, all the three tests could be of potential utility depending on the nature of the terrorist group.

In Fig. 2(b), we plot the test-statistics: $R_{1,n}$ with $\lambda_1 = 0.05$, $R_{2,n}$ with $\lambda_2 = 0.10$, and $\alpha R_{1,n} + \sqrt{1-\alpha^2} R_{2,n}$ with $\alpha = 0.25$. The threshold is designed as $\{A_1, A_2, A\} = 3\log(\gamma)$ for $\gamma = 10$. From Fig. 2(b), we see that an appropriate weighted combination of the metric that captures resilience and the level of coordination in the group performs better than either test-statistic taken separately (with the same threshold for all the three tests). While with the FARC dataset, the weighted sum performs only marginally better than the resilience-based metric, in general, we expect τ_{weighted} to significantly improve the performance over either τ_1 or τ_2. But more importantly, the EWMA algorithm-based approach detects only *persistent* changes or changes that last for a sufficiently long duration such that the changepoint detection methodology can work accurately. In other words, the major spurts in FARC activity are detected, whereas the minor spurts *cannot* be detected with this approach. This is because the method does not incorporate or exploit the underlying statistical information of $\{X_n\}$ or $\{Y_n\}$.

4.3 Non-parametric approach: Majorization theory

We now consider an alternate non-parametric approach for spurt detection. To illustrate this approach, consider two extreme scenarios: i) a group conducting δ attacks on a specific day over a δ-day time-window and no other attacks in this period, and ii) a group conducting one attack on each day of the δ-day period. The former setting correlates well with a group having a high-degree of coordination, whereas the latter setting would be more amenable with the belief that the group has a high-degree of resiliency. Rephrasing the above, a metric that measures the degree of "well-spreadness" of attacks (or its lack thereof) over an

appropriately chosen time-window can be used as an indicator of high resilience (or coordination). On this note, majorization theory provides a theoretical framework to compare two vectors on the basis of their "well-spreadness" [30].

We apply the theoretical framework of catalytic majorization and the existence of certain functionals that *bijectively* capture this relationship, developed in [31], to detect changes in resilience and coordination. Let $\underline{\boldsymbol{M}} = [M_1, \cdots, M_\delta]$ capture the distribution of frequency of attacks over a certain time-window. We call $\underline{\boldsymbol{M}}$ the *attack frequency vector* and note that by definition $\underline{\boldsymbol{M}} \in \mathcal{P}(\delta)$, provided that there is at least one attack over this time-window.

Motivated by the discussion in [31], we consider the following functionals in comparing two different attack frequency vectors: i) number of attacks over the time-window (denoted as Z_n), ii) normalized power mean for some $\alpha > 1$, defined as,

$$\mathsf{NPM}\left(\underline{\boldsymbol{M}}\big|_{\Delta_n}, \alpha\right) \triangleq \frac{\left(\sum_i \mathsf{M}_i^\alpha\right)^{1/\alpha}}{\sum_i \mathbb{1}\left(\{\mathsf{M}_i > 0\}\right)}, \tag{29}$$

and iii) Shannon entropy, defined as,

$$\mathsf{SE}\left(\underline{\boldsymbol{M}}\big|_{\Delta_n}\right) \triangleq -\sum_i \mathsf{M}_i \log(\mathsf{M}_i). \tag{30}$$

Rephrasing the main conclusion of [31], a vector that corresponds to a large Z_n and is more spread-out (indicating a high resilience in the group) results in a larger value for $\mathsf{SE}\left(\underline{\boldsymbol{M}}\big|_{\Delta_n}\right)$. On the other hand, a vector that corresponds to a large Z_n and is less spread-out (indicating a high coordination in the group) results in a larger value for $\mathsf{NPM}\left(\underline{\boldsymbol{M}}\big|_{\Delta_n}, \alpha\right)$. Finally, a small value for Z_n suggests that the group is an *Inactive* state.

We now propose a simplistic birth-death process model to track changes in resilience and coordination. For this, we define two functions that compare the Shannon entropy and the normalized power mean over Δ_n with the corresponding running sample means as follows:

$$\widetilde{X}_n = \frac{\mathsf{SE}\left(\underline{\boldsymbol{M}}\big|_{\Delta_n}\right)}{\frac{1}{\Delta}\sum_{i=1}^{\Delta} \mathsf{SE}\left(\underline{\boldsymbol{M}}\big|_{\Delta_{n-i}}\right)}; \qquad \widetilde{Y}_n = \frac{\mathsf{NPM}\left(\underline{\boldsymbol{M}}\big|_{\Delta_n}, \alpha\right)}{\frac{1}{\Delta}\sum_{i=1}^{\Delta} \mathsf{NPM}\left(\underline{\boldsymbol{M}}\big|_{\Delta_{n-i}}, \alpha\right)}. \tag{31}$$

We then update two functions that capture the two facets of interest, $R(n)$ and $C(n)$, as follows:

$$R(n) = R(n-1) + \tau_{\mathcal{R}},\ n \geq 1,\ R(0) = 0, \tag{32}$$
$$C(n) = C(n-1) + \tau_{\mathcal{C}},\ n \geq 1,\ C(0) = 0, \tag{33}$$

where $p_{\mathcal{R}}$ and $p_{\mathcal{C}}$ are appropriately chosen *Inactive* state penalties, and

$$\tau_{\mathcal{R}} = \mathbb{1}\left(\widetilde{X}_n > \overline{\gamma}_{\mathcal{R}},\ Z_n > \tau\right) - \mathbb{1}\left(\widetilde{X}_n < \underline{\gamma}_{\mathcal{R}},\ Z_n > \tau\right) - p_{\mathcal{R}} \cdot \mathbb{1}\left(Z_n \leq \tau\right) \tag{34}$$
$$\tau_{\mathcal{C}} = \mathbb{1}\left(\widetilde{Y}_n > \overline{\gamma}_{\mathcal{C}},\ Z_n > \tau\right) - \mathbb{1}\left(\widetilde{Y}_n < \underline{\gamma}_{\mathcal{C}},\ Z_n > \tau\right) - p_{\mathcal{C}} \cdot \mathbb{1}\left(Z_n \leq \tau\right). \tag{35}$$

To restate, $\tau_{\mathcal{R}}$ and $\tau_{\mathcal{C}}$ take four possible values: 1, -1, 0, and $p_{\mathcal{R}}$ (or $p_{\mathcal{C}}$), depending on whether the group is resilient/coordinating, non-resilient/non-coordinating, neither resilient nor coordinating, and *Inactive*, respectively. More importantly, the proposed approach quickly detects changes in resilience and coordinaton (and allows these changes to

be categorized) without suffering from explicit model learning delays. Thus, the proposed approach is of tremendous advantage in practice.

We now consider state classification with the FARC dataset. We use the following parameters in our study: $\delta = 15$ days, $\Delta = 5$, $\alpha = 2.5$, $\tau = 4$, $p_{\mathcal{R}} = 0.2$, $p_{\mathcal{C}} = 0$, $\overline{\gamma}_{\mathcal{R}} = \overline{\gamma}_{\mathcal{C}} = 0.6770$, and $\underline{\gamma}_{\mathcal{R}} = \underline{\gamma}_{\mathcal{C}} = 0.4513$. Fig. 3(a) plots the two statistics, $R(n)$ and $C(n)$, against the backdrop of $Z(n)$. It can be seen that $R(n)$ decreases initially before starting to rise in early 2002 (coinciding with *Plan Columbia*) and again in 2006 coinciding with the re-election period. On the other hand, $C(n)$ shows only minor spurts over the same period indicating that FARC was a more resilient group than a group coordinating multiple attacks.

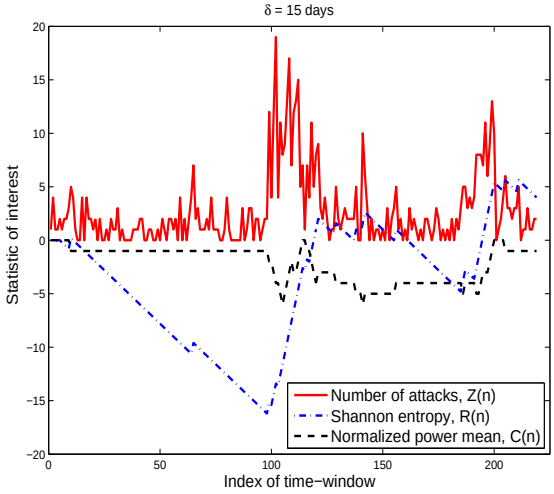


Figure 3: Resilience and level of coordination functions for the FARC dataset.

5. Concluding Remarks

The main focus of this paper is in model comparison between the SEHM framework and the HMM framework that are used to capture the activity of terrorist groups. In particular, the explanatory and predictive powers of the two frameworks are compared and contrasted, with specific attention to the FARC and the Indonesia/Timor-Leste datasets. The comparison study illustrated the distinctive advantage of the HMM framework in prediction. Building on this modeling study, we then considered the problem of quick detection of spurts in the activity profile. We developed a non-parametric majorization theory-based approach for this task and showed that this approach compares favorably relative to an (impractical) parametric approach as well as a non-parametric approach based on the EWMA algorithm. Future work will consider the application of this approach to a broad swathe of terrorist groups' activity profiles.

References

[1] W. M. Landes, "An economic study of U.S. aircraft hijackings, 1961-1976," *Journal of Law and Economics*, vol. 21, no. 1, pp. 1–31, Apr. 1978.

[2] J. Cauley and E. I. Im, "Intervention policy analysis of skyjackings and other terrorist incidents," *The American Economic Review*, vol. 78, no. 2, pp. 27–31, May 1988.

[3] W. Enders, T. Sandler, and J. Cauley, "U.N. conventions, technology and retaliation in the fight against terrorism: An econometric evaluation," *Terrorism and Political Violence*, vol. 2, no. 1, pp. 83–105, 1990.

[4] B. Brophy-Baermann and J. A. C. Conybeare, "Retaliating against terrorism: Rational expectations and the optimality of rules versus discretion," *American Journal of Political Science*, vol. 38, no. 1, pp. 196–210, Feb. 1994.

[5] W. Enders and T. Sandler, "The effectiveness of antiterrorism policies: A vector autoregression-intervention analysis," *The American Political Science Review*, vol. 87, no. 4, pp. 829–844, Dec. 1993.

[6] W. Enders and T. Sandler, "Is transnational terrorism becoming more threatening? A time-series investigation," *Journal of Conflict Resolution*, vol. 44, no. 3, pp. 307–332, June 2000.

[7] W. Enders and T. Sandler, "Patterns of transnational terrorism, 1970-1999: Alternative time-series estimates," *International Studies Quarterly*, vol. 46, no. 2, pp. 145–165, June 2002.

[8] G. LaFree, N. A. Morris, and L. Dugan, "Cross-national patterns of terrorism, comparing trajectories for total, attributed and fatal attacks, 1970-2006," *British Journal of Criminology*, vol. 50, no. 4, pp. 622–649, 2010.

[9] L. Dugan, G. LaFree, and A. Piquero, "Testing a rational choice model of airline hijackings," *Criminology*, vol. 43, no. 4, pp. 1031–1065, Nov. 2005.

[10] M. D. Porter and G. White, "Self-exciting hurdle models for terrorist activity," *Annals of Applied Statistics*, vol. 6, no. 1, pp. 106–124, 2012.

[11] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, no. 1, pp. 83–90, Apr. 1971.

[12] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical Association*, vol. 83, no. 401, pp. 9–27, Mar. 1988.

[13] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, Mar. 2011.

[14] E. Lewis, G. O. Mohler, P. J. Brantingham, and A. Bertozzi, "Self-exciting point process models of civilian deaths in Iraq," *Security Journal*, vol. 25, pp. 244–264, 2012.

[15] V. Raghavan, A. Galsytan, and A. G. Tartakovsky, "Hidden Markov models for the activity profile of terrorist groups," *Annals of Applied Statistics*, vol. 7, no. 4, pp. 2402–2430, Dec. 2013.

[16] "RAND Database of Worldwide Terrorism Incidents (RDWTI)," Available: [Online]. `http://www.rand.org/nsrd/projects/terrorism-incidents.html`.

[17] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[18] M. Sageman, *Understanding Terror Networks*, University of Pennsylvania Press, 2004.

[19] K. Cragin and S. A. Daly, *The Dynamic Terrorist Threat: An Assessment of Group Motivations and Capabilities in a Changing World*, RAND Corporation, Santa Monica, CA, 2004, Available: [Online]. `http://www.rand.org/pubs/monograph_reports/MR1782`.

[20] D. N. Santos, "What constitutes terrorist network resiliency?," *Small Wars Journal*, vol. 7, no. 5, May 31 2011.

[21] M. Lindberg, "Factors contributing to the strength and resilience of terrorist groups," *Grupo de Estudios Estrategicos (GEES) Publication*, May 9 2010.

[22] B. S. Blomberg, K. Gaibulloev, and T. Sandler, "Terrorist group survival: Ideology, tactics, and base of operations," *Public Choice*, vol. 149, no. 3-4, pp. 441–463, Dec. 2011.

[23] R. M. Bakker, J. Raab, and H. B. Milward, "A preliminary theory of dark network resilience," *Journal of Policy Analysis and Management*, vol. 31, no. 1, pp. 33–62, Winter 2012.

[24] A. N. Shiryaev, *Optimal Stopping Rules*, Springer-Verlag, NY, 1978.

[25] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*, Prentice Hall, Eniglewood Cliffs, NJ, 1993.

[26] H. V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, 2008.

[27] S. W. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 1, no. 3, pp. 239–250, Aug. 1959.

[28] M. S. Srivastava and Y. Wu, "Evaluation of optimum weights and average run lengths in EWMA control schemes," *Communications in Statistics - Theory and Methods*, vol. 26, no. 5, pp. 1253–1267, 1997.

[29] A. S. Polunchenko, G. Sokolov, and A. G. Tartakovsky, "Optimization and efficiency analysis of the EWMA procedure for detecting changes in the exponential distribution," *Proceedings of Quality and Productivity Research Conference, Long Beach, CA*, 2012.

[30] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic Press, NY, 1979.

[31] V. Raghavan, "Detecting changes in resilience and level of coordination in terrorist groups," *JSM Proceedings, Quality and Productivity Section, Alexandria, VA: American Statistical Association*, pp. 1193–1203, Aug. 2014.