Detecting changes in self-exciting Point Processes through trend reversal

Anuja Das* Moinak Bhaduri†

Abstract

Point processes, often of the Poisson type, offer a framework to model random events evolving in continuous time, and the self-exciting subclass tackles cases where the occurrence of one inflates or deflates the occurrence probability of another in an immediate neighbourhood. Examples include an earthquake and its aftershocks, a sequence of landslides, a tweet and its re-tweets, and several others. There exist combinations of underlying intensities, especially close similarities between the pre- and post-change flow that make the detection and estimation of changes in the first or second generation sub-process quite hard. Through extensive simulations and real applications, this work examines how a newly developed statistic, formed through switching the usual flow of time, aids the detection exercise. Improvements over established competitors have been quantified, optimal intensity classes have been characterized, and generalizations have been described.

Key Words: point processes, self-exciting process, Hawkes process, trend permutation, change-point detection

1. Introduction

Point processes throng the modelling of random tessellations at a pace both unprecedented and justified. Such patterns evolve over time, space, their combinations, or more complicated topologies, and frequently signal complex dependencies. Instances cover the times at which COVID-19 patients arrive at a health clinic, the distribution of stars in our galaxy, and several others in between. In stark contrast to simpler alternatives like the homogeneous Poisson process, where the numbers of shocks over disjoint compact supports are assumed to be independent *a priori*, of interest in this work is a specific subclass termed Hawkes processes, through which those numbers are allowed to be connected, enabling both realistic modelling and a firmer understanding of the observed trajectory. The following section offers the necessary grasp of certain stochastic structures.

Change point detection, on the other hand, remains an exercise of intrigue confronting modern statistical inference, where one questions whether a stable arrival flow has deviated somehow, and estimates any point of departure, objectively and precisely. Mathematical tractability often compels uncomplicated constructs - change estimations only in means, or variances, for instance. This brief communication is poised to scrutinize the applicability of certain techniques proffered by the second author (in the context of deterministic Poissoninan intensities) in conditions when the driving intensity is dictated by the data, and thus, is random.

Section 2 surveys the relevant literature and organises some mathematical conventions, the next reports simulation-based findings that illustrate how the proposed technique triumphs over its established competitors, and section 4 implements the analyses on data sets of varied flavours and intricacies.

*Graduate student, MSBA, Bentley University, 175 Forest St, Waltham, Massachusetts 02452

†Assistant Professor, Department of Mathematical Sciences, Bentley University, 175 Forest St, Waltham, Massachusetts 02452

2. Theory and methods

2.1 Hawkes process

A continuous time stochastic process $\{N(t)\}_{t\geq 0}$ is routinely deployed to model a bunch of ordered arrival points $t_1 < t_2 < ... < t_n < ...$ representing the global occurrence times of some random phenomenon of interest. $\{N(t)\}_{t\geq 0}$ is also referred to as a counting process with the understanding that $N(t)$ at a given time t, will count the number of observations in $(0, t]$. Please notice the (almost sure) strict ordering in the arrival times, ensuring a simple point process, the type we are examining now, as opposed to an explosive one. Additionally, a function $\lambda(.)$, termed the intensity, given through

$$\lambda(t) = \lim_{h\to 0} \frac{P\{N(t, t+h] \geq 1\}}{h}, t \geq 0, \tag{1}$$

is taken to exist, quantifying the instantaneous probability of observing at least one shock. A $\lambda(.)$ free of time leads to a stationary point process. An increasing $\lambda(.)$ leads to a deteriorating process (where shocks occur more and more frequently as time goes on), while a decreasing $\lambda(.)$ implies an improving sequence (with shocks happening less and less frequently). Regularity conditions on $\lambda(.)$ lead to the independent increment property and ultimately, to a Poisson process, i.e., when $N(t) \sim \text{Pois}(\int_0^t \lambda(x)dx)$, with probability calculations done through

$$P[N(t) = n] = \exp(-\int_0^t \lambda(x)dx)\frac{\{\int_0^t \lambda(x)dx\}^n}{n!}, n = 0, 1, 2, 3.... \tag{2}$$

A detailed description can be had from Rigdon and Basu [11], for instance. Our change point detection proposals, elaborated in Bhaduri (2018) [3] survey processes of the above kind, with purely deterministic choices of $\lambda(.)$. Such a choice, however, assumes the (conditional) intensity is independent of the history - a condition relaxed through a stochastic choice of $\lambda(.)$. A random, data-dependent intensity enables one event to influence another following, an apt requirement to model such events like earthquakes where one major shock inflates the occurrence probabilities of several aftershocks over a close neighbourhood. We opt for

$$\lambda_Y(t) = \lambda_0(t) + \sum_{i=1+max(0,N(t)-r)}^{N(t)} \omega(t - t_i). \tag{3}$$

This choice leads to a Hawkes process (Hawkes (1971) [12]) where the intensity is composed of one (typically) data independent baseline $\lambda_0(.)$, modelling the rate of occurrence of the "major" events, and a data-dependent memory kernel $\omega(.)$, modeling how much of an influence one shock has on the "minor" aftershocks to follow. Hawkes (1971) [12] chose the exponential memory of the form $\omega(u) := \alpha \exp(-\beta u)$, where $\beta > 0$ controls the rate of "forgetting the past". We stick to the exponential choice with $\alpha < \beta$ to ensure stability. Hawkes processes of this type have been used to analyse the arrival times of spam emails (Prince and Heard (2020) [14]), modelling intensity bursts in financial data sets [15], and various others.

Time independent choices of $\lambda_0(.)$ and $\omega(.)$ lead to a stationary Hawkes process. For our analyses, to create a non-stationary version, we have corrupted both the baseline intensity and the memory kernel in our simulation studies (section 3 below) at pre-determined locations and measured the performance of our proposals against established competitors. Figure 1 demonstrates the differences between the environments implied by deterministic and stochastic intensities. A prominent jump in the deterministic step intensity from 1 to

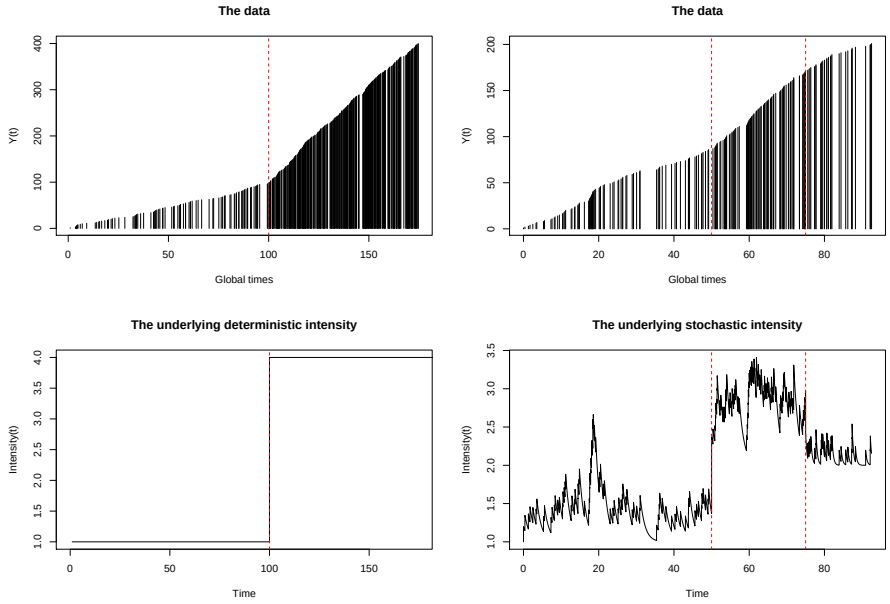

Figure 1: Data reaction to changes under deterministic (left panel) and random (right panel) intensities

4 at $t = 100$ leads to an obvious heavier crowding on the arrival times while changes (in the baseline $\lambda_0(t)$ from 1 to 2 at $t = 50$ and in the memory kernel $\omega(t)$ from 0.8 to 2.8 at $t = 75$) in the stochastic intensity lead to less discernible changes in the data leading to a more complex change detection exercise.

2.2 Change detection

2.2.1 Known tools

For a thorough discussion on change detection in a longitudinal context, we direct interested readers to some of our previous work: Bhaduri (2018) [3], Zhan et al. (2019), [18], Bhaduri and Ho (2018) [8],Bhaduri and Zhan (2018) [7], Ho and Bhaduri (2017) [5], Bhaduri, Zhan and Chiu (2017) [9], Bhaduri et al. (2017) [10], Ho et al. (2016) [4], Ho and Bhaduri (2015) [17], Tan, Bhaduri, and Ho (2014) [16]. These articles elaborate on the batch and sequential detection scenarios and each one of the CPM-based options to follow. For our current short communication, we briefly touch upon Hawkins et al. (2003)'s [13] general approach. Given a bunch of discrete-time variables X_is, change-locations τ_is may update the underlying probability distributions as:

$$X_i \sim \begin{cases} F_0 & \text{if } i \leq \tau_1 \\ F_1 & \text{if } \tau_1 < i \leq \tau_2 \\ F_2 & \text{if } \tau_2 < i \leq \tau_3 \\ ... & \end{cases}$$

A detection problem, therefore, comprises of choosing one of

$$H_0 : X_i \sim F_0(x; \theta_0), i = 1, 2, .., n \tag{4}$$

$$H_1 : X_i \sim \begin{cases} F_0(x; \theta_0), & i = 1, 2, ..., k \\ F_1(x; \theta_1), & i = k+1, k+2, ..n \end{cases} \tag{5}$$

Table 1: Choices for the two-sample $D_{k,n}$ statistics under the CPM framework (Hawkins et al. (2003))

Competitor	Construction	Choice
CPM-Exp (Ross (2014))	$M_{k,n} = -2log(\frac{L_0}{L_1})$	$D_{k,n} = M_{k,n}$
CPM-Adjusted Exp (Ross (2014))	$M^c_{k,n} = \frac{M_{k,n}}{E(M_{k,n})}$	$D_{k,n} = M^c_{k,n}$
CPM-Mann-Whitney (Hawkins, Deng (2010))	$U_{k,n} = \sum_{i=1}^{k}\sum_{j=k+1}^{n} sgn(X_i - X_j)$	$D_{k,n} = U_{k,n}$ (scaled)
CPM-Mood (Ross et al. (2011))	$M = \sum_{X_i}((\sum_{i\neq j}^{n} I(X_i \geq X_j)) - \frac{n+1}{2})^2$	$D_n = M$ (standardized)
CPM-Lepage (Ross et al. (2011))	$L = U^2 + M^2$	$D_n = L$
CPM-Kolmogorov-Smirnov (Ross, Adams (2012))	$M_{k,n} = sup_x\lvert\hat{F}_{S_1}(x) - \hat{F}_{S_2}(x)\rvert$	$D_{k,n} = M_{k,n}$
CPM-Cramer-von-Mises (Ross, Adams (2012))	$M_{k,n} = \int_{-\infty}^{\infty}\lvert\hat{F}_{S_1} - \hat{F}_{S_2}\rvert dF_t(x)$	$D_{k,n} = M_{k,n}$

Table 2: Other parametric options based on likelihood ratio tests, energy divergence, and trend tests

Competitor	Working
E-divergence (Matteson, James (2013, 2014))	$D(X,Y;\alpha) = \int_{R^d}\lvert\phi_X(t) - \phi_Y(t)\rvert^2(\frac{2\pi^{d/2}\Gamma(1-\alpha/2)}{\alpha 2^\alpha\Gamma((d+\alpha)/2)}\lvert t\rvert^{d+\alpha})^{-1}dt > C$
Parametric (Chen, Gupta (2011))	$L_k = -2log\frac{L_0(\hat{\lambda})}{L_1(\hat{\lambda},\hat{\lambda'})} < C$
Pettitt (Pettitt (1979))	$K_T = max_{1\leq t\leq T}\lvert\sum_{i=1}^{t}\sum_{j=t+1}^{T} sgn(X_i - X_j)\rvert > C$
Buishand (Buishand (1982))	$U = \frac{1}{n(n+1)}\sum_{k=1}^{n-1}(\frac{S_k}{D_x})^2$, where $S_k = \sum_{i=1}^{k}(X_i - \bar{X}), D_x = sd(X)$

With a given sample size n, some statistic $D_{k,n}$ is constructed, that measures the "difference" between the pre- and post-chang blocks for an arbitrary choice of an initial change estimate at k. These $D_{k,n}$s, in turn, lead to

$$D_{k,n} \Rightarrow D_n = \max_{k=2,3,..,n-1} D_{k,n} \tag{6}$$

and a change is signaled through

$$\phi(D_n) = \begin{cases} 1 & \text{if } D_n > h_n \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where the threshold h_n is estimated from the null-distribution of D_n, with the estimated change location at

$$\hat{\tau} = argmax_{k=2,3,..,n-1} D_{k,n} \tag{8}$$

Different choices of $D_{k,n}$ lead to different options, working well under different assumptions (changes, for instance, only in the mean or the variance structure, the trend, etc.). Tables 1 and 2 lay them out while the references above provide details. While comparing these options with our proposals below in a continuous time point process setting, we take the X values as the inter-event times, which leads to a discrete time series.

2.2.2 Our proposal

Our approach (Bhaduri (2018) [3]) towards detecting changes offers an algorithm that can operate on continuous time (i.e., the conversion to discrete-time Xs is not needed). Essentially, a test involving a block of n neighbouring event times needs to be conducted. If this test signals stationarity and if a similar test involving a block of $n+1$ neighbouring event times (the n-many from the previous stage and the immediate next) signals non-stationarity, we estimate a change-point between the n-th and the $n+1$-th event times. More formally, it runs thus:

- Set series of hypotheses: $\{H_1, H_2, ...H_m\}$, p-values: $p_1, p_2, ...p_m$.

Table 3: Test statistic proposals for multiple testing (Bhaduri (2018))

Proposal	Critical regions	Comments
$Z = -2\sum_{i=1}^{n} log(t_i/t_n)$	$Z \leq \chi^2_{1-\alpha/2,2n-2}$ or $Z \geq \chi^2_{\alpha/2,2n-2}$	UMPU (Bain and Engelhardt (1991)) in power law setting: $\lambda(t) = \frac{\beta}{\theta}(\frac{t}{\theta})^{\beta-1}$, $t > 0$
$Z_B = -2\sum_{i=1}^{n} log(1 - t_i/t_n)$	$Z_B \leq \chi^2_{1-\alpha/2,2n-2}$ or $Z_B \geq \chi^2_{\alpha/2,2n-2}$	More powerful than Z in detecting increasing step intensities (Ho (1993)). Under further analysis.
$R := max(Z, Z_B)$	$R \geq c^{\alpha}_R$	Powerful under deterministic intensities (Bhaduri (2018)).
$L := min(Z, Z_B)$	$L \leq c^{\alpha}_L$	Powerful under deterministic intensities (Bhaduri (2018)).

- H_i tests stationarity on the first $i + 1$ events.
- Order the p-values: $p_{(1)} < p_{(2)} < ... < p_{(m)}$.
- Set $S_i := \{k : p_{(k)} < \frac{k}{m}\alpha\}$
- $$\hat{\tau}_i := \begin{cases} \min\{k : p_{(k)} < \frac{k}{m}\alpha\}, & S_i \neq \emptyset \\ \infty, & S_i = \emptyset \end{cases}$$

Once the earliest significant test (if any) is detected, the algorithm may be restarted with the detected change point as the fresh time origin to discover subsequent changes, if any. The technique is, therefore, free of the "at-most-one-change-point (AMOC)" assumption under which several parametric proposals work. Declaring significance through ordered p-values is done cautiously since performing multiple correlated tests is known to inflate the type-I error probability. We follow the false discovery rate control suggested by Benjamini and Hochberg (1995) [6]. The actual testing is done through an array of novel statistics offered by Bhaduri (2018) [3]. Their definitions and crucial properties are summarized in Table 3. A deteriorating sequence inflates the value of Z_B and deflates the value of Z. Randomized versions of the two through the maximum and the minimum signal general non-stationarities (i.e., both improvement and deterioration) through significantly large or small values. The critical thresholds c^R_α and c^L_α are summarized in Bhaduri (2018) [3].

3. Simulation studies

As a preliminary test to identify which one of the four statistics shown in Table 3 has the highest classification accuracy, we have conducted a power study summarized in Figure 2 where the immigrant intensity $\lambda_0(.)$ was taken to be time-inhomogeneous and the only change was brought through the offspring kernel $\omega(.)$. The pre- and post-change β values are placed along the x and the y axes, while the power of each test, the estimated probability of correctly identifying a non-stationary sequence as a non-stationary sequence, is plotted along the z axis. The power surface appeals to intuition. Along the diagonal, when the pre- and the post-change recollections are reasonably identical, detection gets tougher, leading to lower power. In contrast, along the edges and the corners, when the difference is stark, the power rises. We found there is no one statistic that shows the uniformly best power (although the minimum based L test occupies a larger region) and about the diagonal, there exists an asymmetry in the power surfaces.

Next, we investigated, through Figures 3 through 5, the closeness of the change points estimated by our sequential proposals and their competitors to the true ones through a more massive simulation study conducted with 10^4 runs. Each dot signifies a global time (plotted along the vertical axis) of change detection, and under each setting, we have conducted tests on both failure truncated (i.e., when we wait for a specified *number* of shocks, regardless

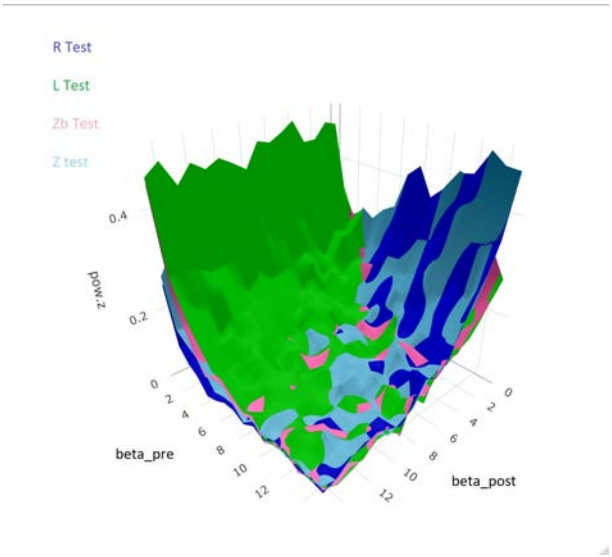

Figure 2: Power study among Z, Z_B, R, and L showing asymmetry

of the time it takes to wait that long) and time truncated (i.e., when we wait for a specified time, regardless of the number of events seen by then) cases. Figure 3 shows the average run length comparisons under the assumption of no change. The heaviest crowding is observed towards the end of the process. This, owing to how non-detections are expressed (please see the previous section), confirms that all our proposals and most of the rest pick up stationarity adequately. The next scenario, graphed in Figure 4 shows the results under a true change in the memory kernel's β value from 3.8 to 0.8 at time 100 (shown through the broken horizontal line), with the baseline intensity held constant. Most of our sequential proposals, especially L and R generate estimations that crowd around this true change point. Another observation is that quite a few of our competitors estimate a change point prior to the true change - clearly a false alarm. The sequential proposals are largely free of such a problem. Finally, in Figure 5, we bring about changes in both the baseline intensity (from 1 to 2) at time point 50 (shown through the heavy broken horizontal line) and the memory kernel (from 3.8 to 0.8) at time points 75 or 100 (shown through the fainter broken horizontal line). Again, we find it is our sequential proposals that demonstrate the strongest and clearest clustering around these two true change points without sounding too many false alarms.

A natural question to ask at this stage is how will these estimators react to a large influx of data? Asymptotic consistency of these types of estimators, as pointed out by Troung et al. (2020) [19] can be examined through:

- i) $P(|\hat{\tau}| = K)$
- ii) $\frac{1}{T}||\hat{\tau} - \tau^*||_\infty$

where $||\hat{\tau} - \tau^*||_\infty := \max\{\max_{\hat{t}\in\hat{\tau}} \min_{t^*\in\tau^*} |\hat{t} - t^*|, \max_{t^*\in\tau^*} \min_{\hat{t}\in\hat{\tau}} |\hat{t} - t^*|\}$, with τ^* representing the set of true change points, $\hat{\tau}$ representing the set of estimated change points, and k representing the size of τ^* (i.e., the true number of changes). A change point algorithm is said to be asymptotically consistent if the first probability converges to 1 and the second norm converges to 0 as the terminal time of the process is pushed to ∞. The norm in (ii) represents the Hausdorff norm needed to quantify the "distance" between two sets not necessarily of the same size. It isn't hard to verify this distance penalizes overestimation quite harshly.

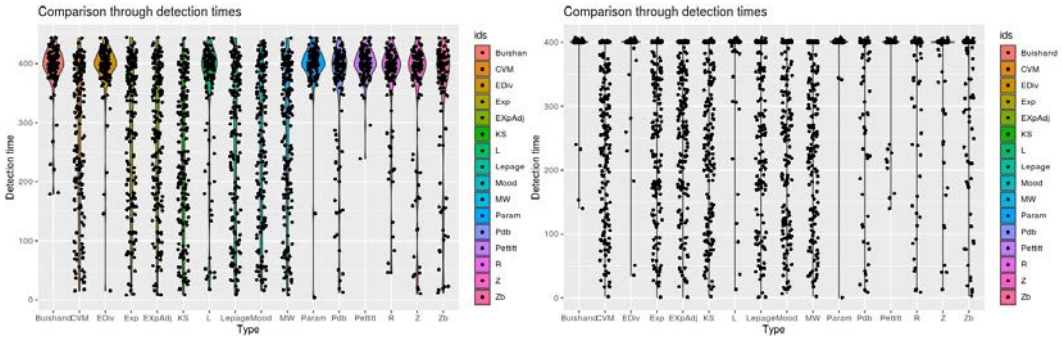

Figure 3: Change identification under stationarity: left panel - failure truncation, right panel - time truncation

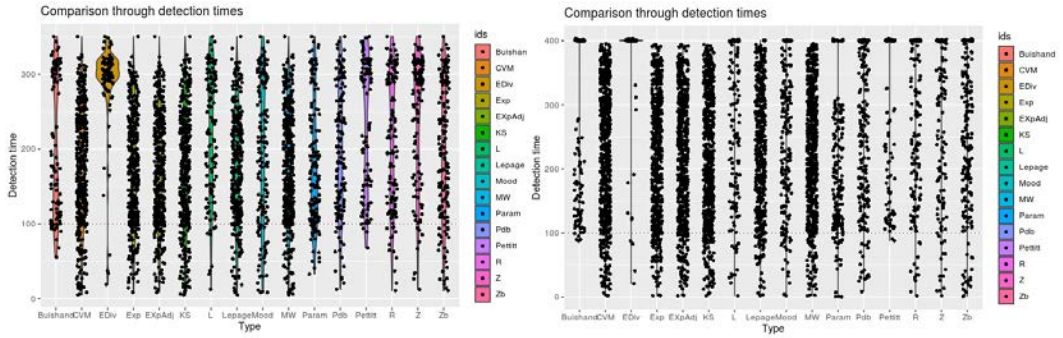

Figure 4: Change identification under $\beta : 3.8 \rightarrow 0.8$: left panel - failure truncation, right panel - time truncation

With our simulation cases, calculation of the long-run probabilities of the type (i) is shown through Figure 6. We observe (especially in the time-truncated scenario) this asymptotic probability for most of our competitors drop fast (primarily due to their sounding too many false alarms), while ours either hold steady or increase. This suggests with our sequential offerings, if one waits sufficiently long (either in terms of time or in terms of data), the right number of change pints will be picked, i.e., one won't over or underestimate with a high chance.

4. Real data analyses

4.1 Financial announcements

The webpage at *https://www.dailyfx.com/* keeps a record of financial announcements coming out of forty one different countries. The announcements are classed into "low", "medium", and "high" categories according to the impact they are likely to cause on the financial market. Information on their actual clock times are also available. We have chosen to analyse announcements coming out of the United States since the beginning of 2020. Figure 7 shows the change points guessed by our proposals.

We observe that detections are scaled with data volume. Understandably, the number of estimated change points increase with a lowering of the impact category. It is crucial to grasp, however, through the asymptotic consistency demonstrated through Figure 6, that estimation is likely to represent a true change and not a false alarm.

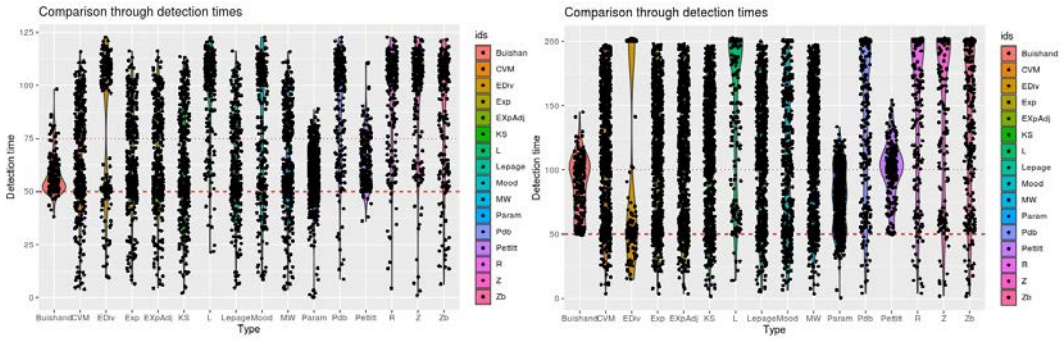

Figure 5: Change identification under $\lambda_0 : 1 \to 2, \beta : 3.8 \to 0.8$: left panel - failure truncation, right panel - time truncation

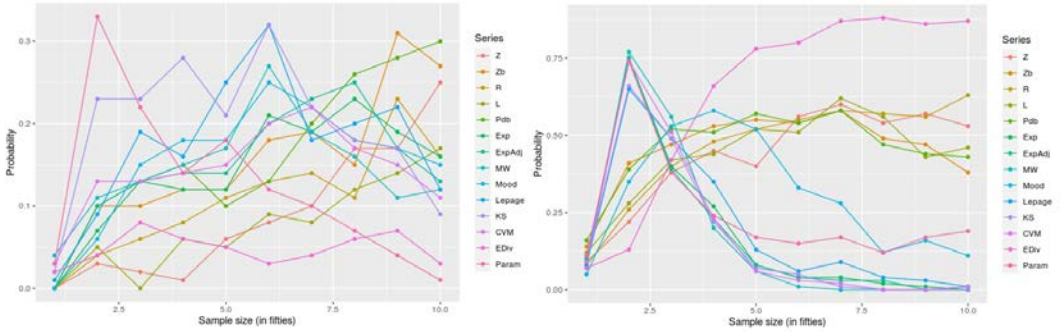

Figure 6: Checking asymptotic consistency with $K = 2$, left panel - failure truncation, right panel - time truncation

4.2 Global terrorism

The Global Terrorism Database (GTD)TM is an unclassified collection of domestic and international terror incidents that have occurred globally since 1970. The database records a multitude of variables including the date, location, nature of the attack, and targeted parties. There exists a lack of consensus regarding what formally constitutes terrorism; the database, nonetheless, classes each attack into one of nine predefined categories. Examples of these categories include armed assaults, bombing/ explosions, and assassinations. It is crucial to note that the database also contains attacks that failed upon implementation but does not include attacks where an attempt was never made. At the time of this writing, the GTD consisted of 191,464 records ranging from January, 1970 to December, 2018. For the purposes of our research, only records pertaining to the USA were used, and three of the nine attack types were analysed. These were Armed Assaults, Bombing/ Explosion, and Facility/ Infrastructure, comprising 272, 940, and 720 records, respectively. The estimated change points from our proposals are graphed in Figure 8.

Moreover, the fifteen change point detection methods (the established eleven shown through tables 1 and 2 and our proposals shown in table 3) are clustered in Figure 9 through the proximities of their estimated change points. The Hausdorff metric described in section 3 has been used to implement the clustering. We find that in both categories of attacks, there appear to be two distinct partitions, with our proposals being clustered together. Additionally, the methods under infrastructure attacks are separated at a low height (around 20) compared to a higher height (around 60) for armed assaults. This confirms our findings in Figure 8 where the methods agree and tend to overlap more often in the middle panel than in the last.

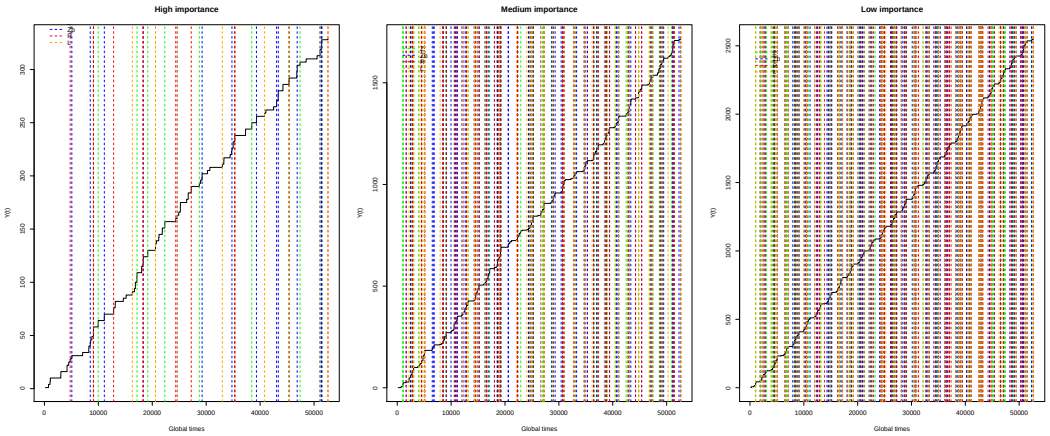

Figure 7: Change point detection, financial announcements data set

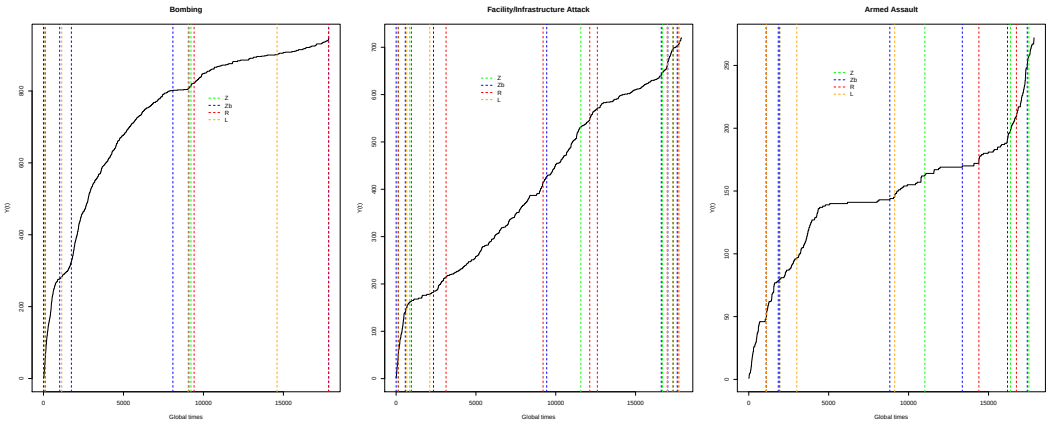

Figure 8: Change point detection, global terrorism data set

A natural curiosity may lead one question the proximity values when *different* techniques are applied to analyse armed assaults and infrastructure attacks. Figure 10 answers that query. The colour gradient at the intersection of a row and a column indicates the amount of discrepancy (measured again, through the Hausdorff metric) between change point locations estimated through the "row" test applied on armed assaults and the "column" test applied on infrastructure attacks. Denser the colour, more different are the locations. Lighter the colour, more similar they are. The fact that these two specific attack types progress roughly similarly appeals to intuition.

4.3 Tsunami

The National Centers for Environmental Information (NCEI) that sits under the National Oceanic and Atmospheric Administration (NOAA) maintains a Global Historic Tsunami Database with records dating back to 2,000 BC. Numerous variables describing each incident are documented, including the date, location, primary magnitude, and focal depth. Records have grown more thorough during the later years. Consequently, the database was subset for the purpose of this analysis to only include tsunamis that occurred on or after January 1, 1900. The original database consisted of 2,961 incidents which shrunk to 1,252 incidents after eliminating tsunamis that were observed prior to the 20th century. Fig 11 documents the estimated change points.

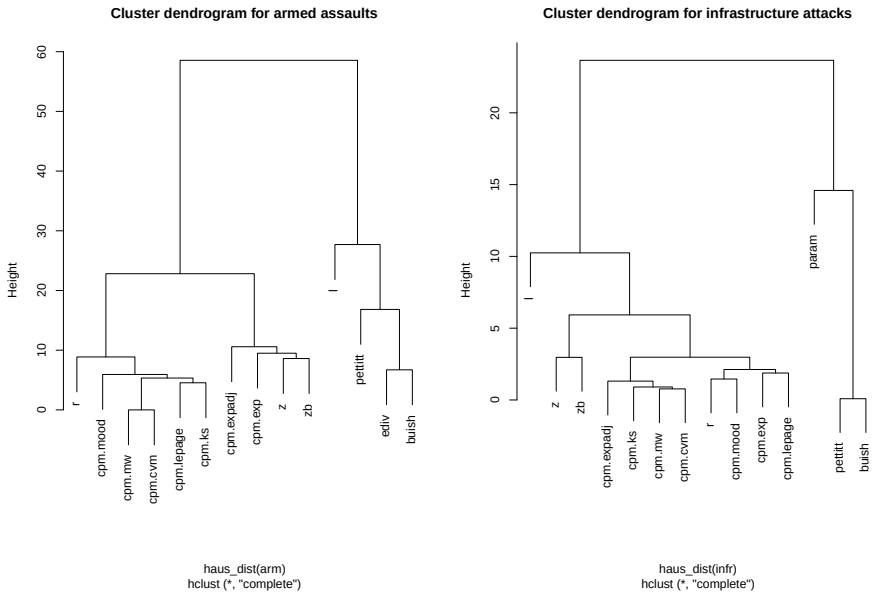

Figure 9: Clustering change point detection methods through Hausdorff distance

4.3.1 Bootstrapping

Through the armed assault example, we now proceed to describe a way of finding interval estimates in addition to the point estimates shown through the vertical lines. Such an exercise will quantify the volatility in our point guesses and aid us gauge the inherent uncertainty. We use bootstrapping to accomplish such an end. Purely random bootstrap estimates of the kind initially introduced by Efron (1971) [1], however, will not be applicable since any underlying auto-correlation will be destroyed. Research into *discrete*-domain time series offers a remedy. Bootstrapping here is done in blocks of some ideal size and such methods have been used by Ho and Bhaduri (2017) [5] and others to estimate standard errors of certain statistics. In the current context, however, the presence of a *continuous*-time point process necessitates a further modification. We follow Braun and Kulperger (1998) [2]'s block resampling algorithm to generate point processes similar to the original one. With an observed point pattern A, with terminal time T, the method runs thus:

- set an ideal number b of blocks.
- simulate $U_1, U_2, ..., U_b$ uniforly and independently from $(0, T - \frac{T}{b}]$.
- for $j = 1, 2, ..., b$:
  - define $A_j := (U_j, U_j + \frac{T}{b}] \cap A$
  - define $A_j^* := A_j - U_j + \frac{(j-1)T}{b}$
  - define $X_j^*(.) := |A_j^* \cap .|$
- the simulated point process X^* is given by $X^* := \sum_{i=1}^{b} X_j^*$

Thus, points from b non-overlapping blocks of size $\frac{T}{b}$ each are joined to create X^*, a point process similar (in terms of properties, the number of shocks could be different) to the original X, ensuring X^* is a restriction on $(0, T]$.

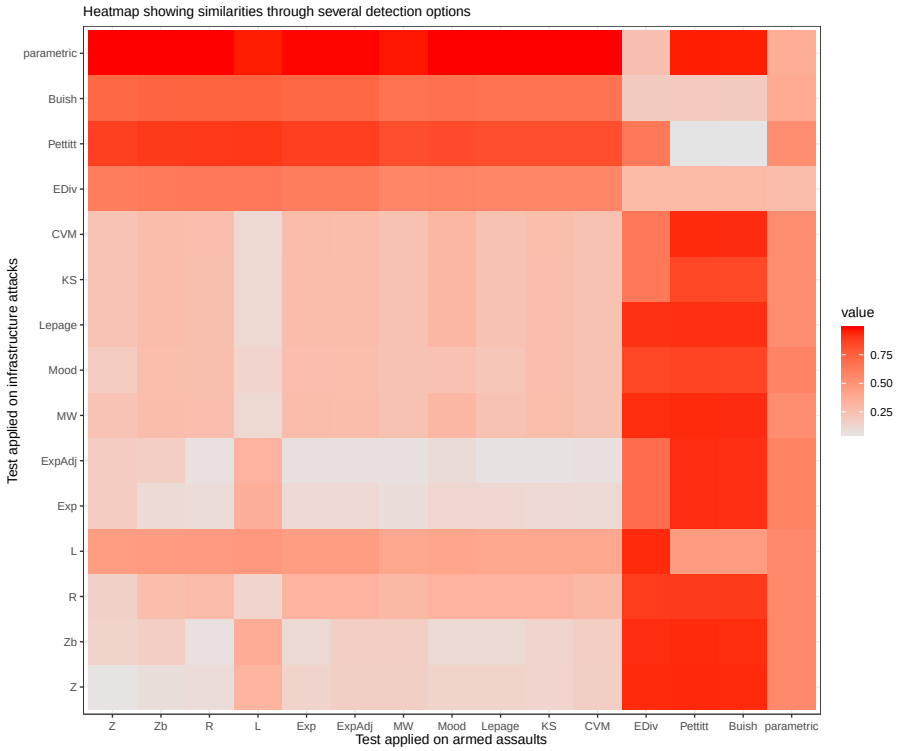

Figure 10: Change point detection similarities with (possibly) different tools applied on different processes

The method works best with roughly stationary point processes. Still, for a visual understanding, we choose the armed assault cases to replicate 100 resamples, each with $b = 15$, apply our sequential testing approach with Z, Z_B, R, L, along with the competitors described in the previous section on each, and generate a bootstrapped change-point distribution. For illustration purposes, we separately show the first three replications over four runs in Figure 12. The black curve represents the observed process and the others are simulated through the above algorithm. It is a reasonable expectation, confirmed for instance, through the upper panel diagrams of Figure 12, that when the replicates are extremely similar to the original, the resulting change point estimates should also be close to the ones from the actual. Table 4 reports the 95% bootstrapped intervals from each method. We notice that the ones from our sequential proposals are consistently tighter than the rest, while covering the most noticeable bend.

The bootstrapping approach outlined above is ideal in case we have dense unimodal crowding of estimated change points. We note that in case of multimodal clustering, one may adopt an approach motivated by the Bayesian highest posterior density credible sets - finding out the narrowest interval that covers 95% of the estimated points as an estimate of the most prominent change point and then applying a similar approach on 95% of the remaining to estimate the next and so on.

5. Conclusions

Random tessellations are plentiful in natural and social sciences. Spatio-temporal mathematical tools that narrate the evolution of earthquakes and hurricanes, often, with certain variations, condense a friendship-network propagation over time quite effortlessly. Conventional point processes of the stationary Poisson type frequently fall short to model such

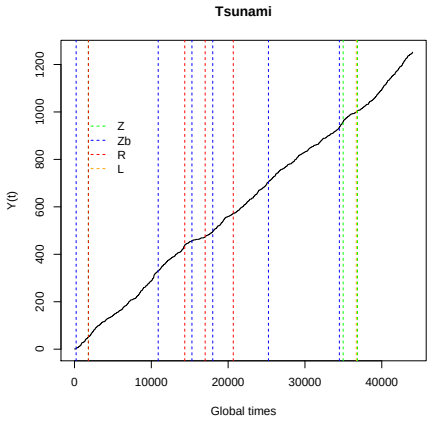

Figure 11: Change point detection, tsunami data set

Table 4: 95% bootstrapped intervals from different change-point options, bombing example

Tool	Interval
Z	[153.694, 10066.385]
Z_B	[332.0148, 11622.5463]
R	[441.3436, 11520.1330]
L	[278.2487, 10375.6471]
CPM-Exp	[300.0683, 15108.9024]
CPM-ExpAdj	[313.1459, 15089.5402]
CPM-MW	[221.695, 13925.299]
CPM-Mood	[65.26729, 14764.16921]
CPM-Lepage	[67.9426, 14205.0223]
CPM-KS	[77.70617, 13770.44373]
CPM-CVM	[78.35929, 14070.67880]
EDiv	[3971.361, 14434.789]
Pettitt	[2631.349, 8775.837]
Buishand	[3650.69, 10076.39]
Parametric	[359.6613, 17397.9730]

count patterns due to several reasons, primary among which is their inherent "without aftereffect" feature, which guarantees help from relevant history would be ignored. This work studies more realistic Hawkes models where the occurrence probability of any shock is influenced by its (potentially infinite) history. The underlying intensity process, in a sense, is thus random in itself. Such a data-dependent intensity is known to generate a branching process structure, where observations in the first generation come from some deterministic intensity describing the ongoing exogenous (i.e., external) environment - the arrivals, for instance, of COVID-19 patients from neighbouring countries, and those in the second (offspring) generation come from some random intensity depicting the current endogenous (i.e., internal) situation - the community spread for example, of the virus within a given country. Detecting changes in these intensities remain crucial for incorporating new measures or anticipating a larger influx of patients. This work popularizes a way of estimating possible changes in both types of intensities through conducting a sequence of hypothesis tests using variations of trend permutation statistics. Originally proposed by the second author in the context of deterministic intensities, the technique, with minor modifications, is now demonstrated to work under more hostile stochastic processes, offering reliable, distribution-free change-estimates without sounding too many false alarms, and without restrictions of the "at most one change point" type. Comparisons with time series-

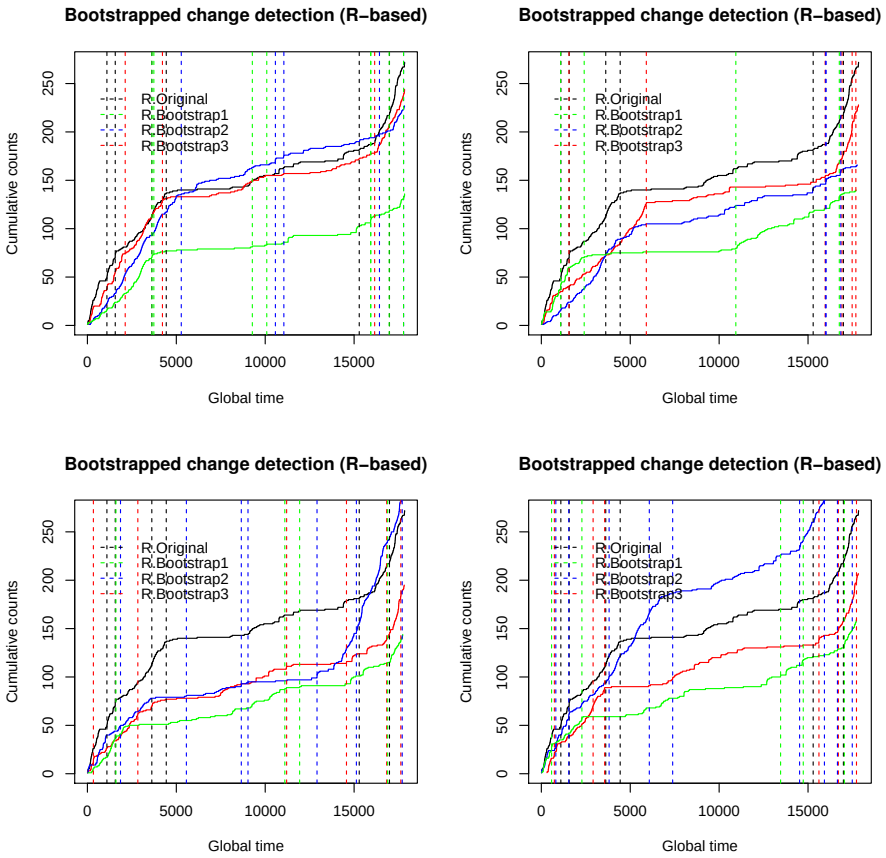

Figure 12: Bootstrapping the armed assault point process, and subsequently, detecting change-points, through R

based change detection methods document smaller estimation errors (viewed through Hausdorff distances), better inferential properties, including asymptotic consistency and tighter bootstrapped estimation intervals. Real examples sampled from financial announcements, global terrorism instances, and tsunami occurrences demonstrate their easy applicability and graphic appeal. Model-based accurate forecasting tools are briefly described, with data-dependent ones in the offing.

References

[1] Efron, B. Bootstrap Methods: Another Look at the Jackknife. Ann. Statist. 7, no. 1, 1–26 (1979).

[2] Braun, Willard & Kulperger, R.J. A bootstrap for point processes. Journal of Statistical Computation and Simulation. 60. 10.1080/00949659808811878. (1998).

[3] Bhaduri, M (2018) *Bi-Directional Testing for Change Point Detection in Poisson Processes* UNLV Theses, Dissertations, Professional Papers, and Capstones. 3217.

[4] Ho, C.-H., Zhong, G., Cui, F., & Bhaduri, M. Modeling interaction between bank failure and size. *Journal of Finance and Bank Management* **4**, 15–33 (2016).

[5] Ho, C.-H. & Bhaduri, M. A Quantitative Insight into the Dependence Dynamics of the Kilauea and Mauna Loa Volcanoes Hawaii. *Mathematical Geosciences* **49**, 893–911 (2017).

[6] Benjamini, Y., & Hochberg, Y. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society. Series B (Methodological), 57(1), 289-300. (1995).

[7] Bhaduri, M. & Zhan, J. Using Empirical Recurrence Rates Ratio for Time Series Data Similarity. *IEEE Access* **6**, 30855–30864 (2018).

[8] Bhaduri, M. & Ho, C.-H. On a Temporal Investigation of Hurricane Strength and Frequency. *Environmental Modeling & Assessment* **24**, 495–507 (2018).

[9] Bhaduri, M., Zhan, J. & Chiu, C. A Novel Weak Estimator For Dynamic Systems. *IEEE Access* **5**, 27354–27365 (2017).

[10] Bhaduri, M., Zhan, J., Chiu, C. & Zhan, F. A Novel Online and Non-Parametric Approach for Drift Detection in Big Data. *IEEE Access* **5**, 15883–15892 (2017).

[11] Rigdon, S.E., and Basu, A.P. (2000). *Statistical Methods for the Reliability of Repairable Systems*, Wiley series in Probability and Statistics, John Wiley and Sons.

[12] Hawkes, A. Spectra of Some Self-Exciting and Mutually Exciting Point Processes. Biometrika, 58(1), 83-90. doi:10.2307/2334319 (1971)

[13] Hawkins, D.M., Qiu, P.H., and Kang, C.W. *The Changepoint Model for Statistical Process Control.* Journal of Quality Technology, 35(4):355–366, (2003)

[14] Prince-Williams, M, Heard, N. A. *Nonparametric self-exciting models for computer network traffic*, Stat Comput 30, 209–220, (2020)

[15] Rambaldi, Marcello and Filimonov, Vladimir and Lillo, Fabrizio (2018) *Detection of intensity bursts using Hawkes processes: An application to high-frequency financial data*, Phys. Rev. E, 97(3), 032-318

[16] Tan, S., Bhaduri, M. & Ho, C.-H. A Statistical Model for Long-Term Forecasts of Strong Sand Dust Storms. *Journal of Geoscience and Environment Protection* **02**, 16–26 (2014).

[17] Ho, C.-H. & Bhaduri, M. On a novel approach to forecast sparse rare events: applications to Parkfield earthquake prediction. *Natural Hazards* **78**, 669–679 (2015).

[18] Zhan, F. *et al.*. Beyond Cumulative Sum Charting in Non-Stationarity Detection and Estimation. *IEEE Access* **7**, 140860–140874 (2019).

[19] Truong, C, Oudre, L, Vayatis, N *Selective review of offline change point detection methods*, Signal Processing, 167, 107-299. (2020)