

Change detection in non-stationary Hawkes processes through sequential testing

Moinak Bhaduri^{1*}, Dhruva Rangan¹, and Anurag Balaji¹

¹Department of Mathematical Sciences, Bentley University, 175 Forest Street, Waltham, Massachusetts, United States of America

Abstract. Detecting changes in an incoming data flow is immensely crucial for understanding inherent dependencies, formulating new or adapting existing policies, and anticipating further changes. Distinct modeling constructs have triggered varied ways of detecting such changes, almost every one of which gives in to certain shortcomings. Parametric models based on time series objects, for instance, work well under distributional assumptions or when change detection in specific properties - such as mean, variance, trend, etc. are of interest. Others rely heavily on the "at most one change-point" assumption, and implementing binary segmentation to discover subsequent changes comes at a hefty computational cost. This work offers an alternative that remains both versatile and untethered to such stifling constraints. Detection is done through a sequence of tests with variations to certain trend permuted statistics. We study non-stationary Hawkes patterns which, with an underlying stochastic intensity, imply a natural branching process structure. Our proposals are shown to estimate changes efficiently in both the immigrant and the offspring intensity without sounding too many false positives. Comparisons with established competitors reveal smaller Hausdorff-based estimation errors, desirable inferential properties such as asymptotic consistency and narrower bootstrapped margins. Four real data sets on NASDAQ price movements, crude oil prices, tsunami occurrences, and COVID-19 infections have been analyzed. Forecasting methods are also touched upon.

1 Introduction

A rich class of natural and artificial phenomenon may be described adequately through the language of point processes - a probabilistic description summarizing their salient features. These occurrences may evolve over time, space, their combinations or over more abstract domains and often exhibit complex dependencies. Examples include the arrival of spam emails in one's mailbox, the distribution of stars in our galaxy, and several others in between. Of interest in this work is a special subcategory termed Hawkes processes through which, the interconnectedness among events can be conveniently understood and modeled. This is in contrast to simpler alternatives such as the homogeneous Poisson processes, where the

*Corresponding author: mbhaduri@bentley.edu

arrival shocks are assumed to be independent *a priori*. A sampling of relevant mathematical descriptions is offered in the following section.

Change-point detection, on the other hand, stays a fascinating problem confronting modern statistical inference, where one wants to check whether a stable arrival flow has deviated in some fashion and estimate any point of departure, objectively and accurately. Mathematical tractability often demands simplified constructs - detecting changes only in means or variances, for instance. The purpose of this brief communication is to examine the applicability of some of the techniques proposed recently by the first author to detect more general changes under a deterministic Poissonian framework in a stochastic intensity-driven Hawkes process scenario.

Section 2 surveys the relevant literature and formalizes the mathematical structure, the next reports simulation results that demonstrate how the proposed technique triumphs over its more established competitors, and section 4 implements the analyses on various data sets with different intricacies.

2 Theory and methods

2.1 Hawkes process

A continuous time stochastic process $\{N(t)\}_{t \geq 0}$ is frequently employed to model a bunch of ordered arrival points $t_1 < t_2 < \dots < t_n < \dots$ 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 $\lambda(\cdot)$, termed the intensity, given through

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

is taken to exist, quantifying the instantaneous probability of observing at least one shock. A $\lambda(\cdot)$ free of time leads to a stationary point process. An increasing $\lambda(\cdot)$ leads to a deteriorating process (where shocks occur more and more frequently as time goes on), while a decreasing $\lambda(\cdot)$ implies an improving sequence (with shocks happening less and less frequently). Regularity conditions on $\lambda(\cdot)$ 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

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

A detailed description can be had from Rigdon and Basu [1], for instance. Our change point detection proposals, elaborated in Bhaduri (2018) [2] surveys processes of the above kind, with purely deterministic choices of $\lambda(\cdot)$. Such a choice, however, assumes the (conditional) intensity is independent of the history - a condition relaxed through a stochastic choice of $\lambda(\cdot)$. A random, data-dependent intensity enables one event 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

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

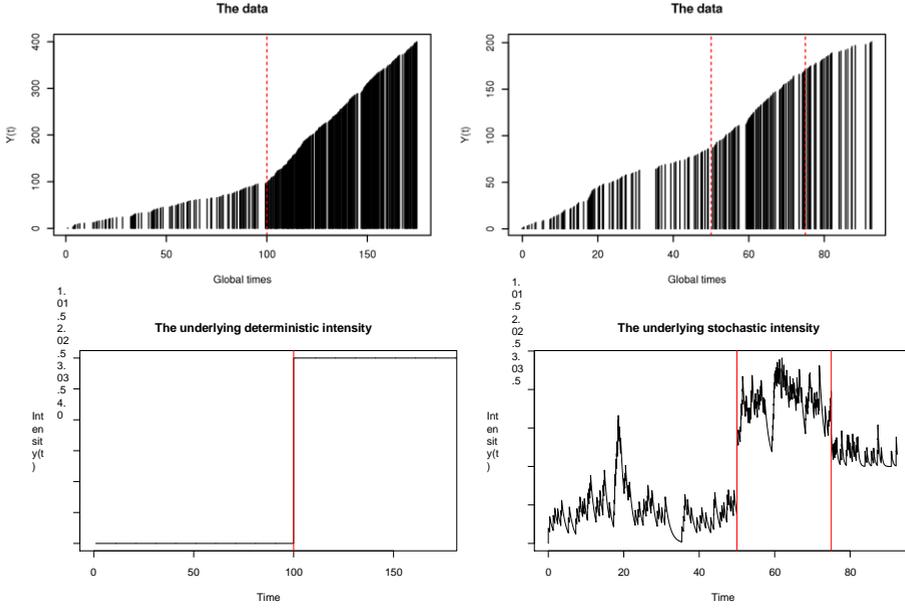


Fig. 1. Data reaction to changes under deterministic (left panel) and random (right panel) intensities.

This choice leads to a Hawkes process (Hawkes (1971) [3]) where the intensity is composed of one (typically) data independent baseline $\lambda_0(\cdot)$, modeling the rate of occurrence of the “major” events, and a data dependent memory kernel $\omega(\cdot)$, modeling how much of an influence one shock has on the “minor” aftershocks to follow. Hawkes (1971) [3] 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 analyze the arrival times of spam emails (Prince and Heard (2020) [4]), modeling intensity bursts in financial data sets [5], and various others.

Time independent choices of $\lambda_0(\cdot)$ and $\omega(\cdot)$ 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. Fig. 1 demonstrates the differences between the environments implied by deterministic and stochastic intensities. A prominent jump in the deterministic step intensity from 1 to 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 a 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) [2], Zhan et al. (2019), [6], Bhaduri and Ho (2018) [7], Bhaduri and Zhan (2018) [8], Ho and Bhaduri (2017) [9], Bhaduri, Zhan and Chiu (2017) [10], Bhaduri et al. (2017) [11], Ho et al. (2016) [12], Ho and Bhaduri (2015) [13], Tan, Bhaduri, and Ho (2014) [14]. These articles elaborate the batch and sequential detection scenarios and each one of the CPM-based options to follow. For the purpose of the current short communication, we briefly touch upon Hawkins et al. (2003)’s [15] general approach. Given a bunch of discrete-time variables X_i s, change-locations τ_i s may update the underlying probability distributions as:

A detection problem, therefore, comprises of choosing one of

$$H_0: X_i \sim F_0(x; \theta_0), i = 1, 2, \dots, n$$

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

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

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

and a change is signaled through

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

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

$$\hat{\tau} = \operatorname{argmax}_{k=2,3,\dots,n-1} D_{k,n}$$

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.). The tables below 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.

Table 1. Choices for the two-sample $D_{k,n}$ statistics under the CPM framework (Hawkins et al. (2003)[15]).

Competitor	Construction	Choice
CPM-Exp (Ross (2014))	$M_{k,n} = -2\log\left(\frac{L_0}{L_1}\right)$	$D_{k,n} = M_{k,n}$
CPM-Adjusted Exp (Ross (2014))	$M_{k,n}^c = \frac{M_{k,n}}{E(M_{k,n})}$	$D_{k,n} = M_{k,n}^c$
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))	$M = \sum_{X_i} \left(\sum_{i \neq j} I(X_i \geq X_j) - \frac{n+1}{2} \right)^2$	$D_n = M$ (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 \hat{F}_{S_1}(x) - \hat{F}_{S_2}(x) $	$D_{k,n} = M_{k,n}$
CPM-Cramer-von-Mises (Ross, Adams (2012))	$M_{k,n} = \int_{-\infty}^{\infty} \hat{F}_{S_1} - \hat{F}_{S_2} dF_t(x)$	$D_{k,n} = M_{k,n}$

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

Competitor	Working
E-divergence (Matteson, James (2013, 2014))	$D(X, Y; \alpha) = \int_{R^d} \phi_X(t) - \phi_Y(t) ^2 \left(\frac{2\pi^{d/2} \Gamma(1 - \alpha/2)}{\alpha 2^\alpha \Gamma((d + \alpha)/2)} t ^{d+\alpha} \right)^{-1} dt > C$
Parametric (Chen, Gupta (2011))	$L_k = -2\log \frac{L_0(\hat{\lambda})}{L_1(\hat{\lambda}, \hat{\lambda})} < C$
Pettitt (Pettitt (1979))	$K_T = \max_{1 \leq t \leq T} \left \sum_{i=1}^t \sum_{j=t+1}^T sgn(X_i - X_j) \right > C$
Buishand (Buishand (1982))	$U = \frac{1}{n(n+1)} \sum_{k=1}^{n-1} \left(\frac{S_k}{D_x} \right)^2$, where $S_k = \sum_{i=1}^k (X_i - \bar{X})$, $D_x = sd(X)$

2.2.2 Our proposal

Our approach (Bhaduri (2018) [2], Bhaduri (2020) [16]) towards detecting changes offers an algorithm that can operate on continuous time (i.e., the conversion to discrete-time X_s 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: $\{H_1, H_2, \dots, H_m\}$, p-values: p_1, p_2, \dots, p_m .
- H_i tests stationarity on the first $i + 1$ events.

- Order the p-values: $p_{(1)} < p_{(2)} < \dots < p_{(m)}$.
- 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 restated 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) [17]. The actual testing is done through an array of novel statistics offered by Bhaduri (2018) [2]. Their definitions and crucial properties are summarized below.

Table 3. Test statistic proposals for multiple testing (Bhaduri (2018) [2]).

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

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_α^R and c_α^L are summarized in Bhaduri (2018) [2].

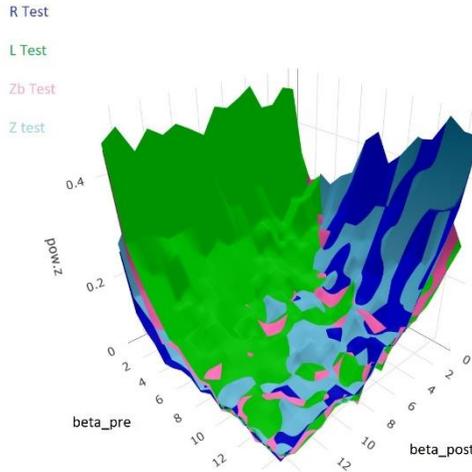


Fig. 2. Power study among Z , Z_B , R , and L showing asymmetry.

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 Fig. 2 where the immigrant intensity $\lambda_0(\cdot)$ was taken to be time-inhomogeneous and the only change was brought through the offspring kernel $\omega(\cdot)$. 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 fairly identical, detection gets tougher, leading to a lower power, while 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 Fig. 3 through Fig. 5, the closeness of the change points estimated by our sequential proposals and their competitors to the true ones through a larger scale 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 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. Fig. 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 Fig. 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.

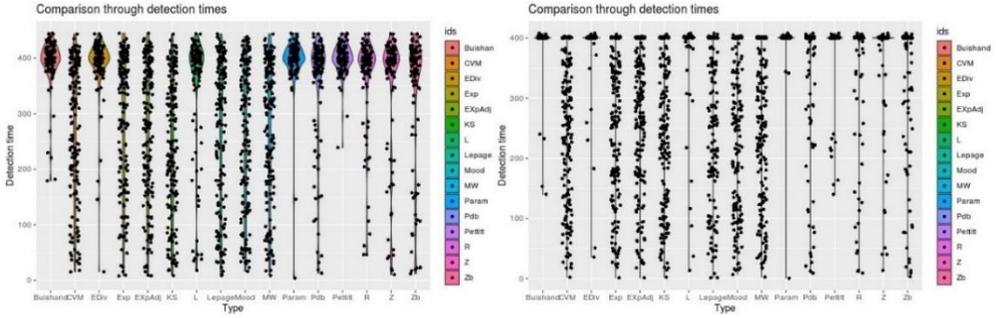


Fig. 3. Change identification under stationarity: left panel - failure truncation, right panel - time truncation.

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 Fig. 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) [18] 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.

With our simulation cases, calculation of the long-run probabilities of the type (i) is shown through Fig. 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 points will be picked, i.e., one won’t over or underestimate with a high chance.

4 Real data analyses

4.1 NASDAQ-100 movements, crude oil prices, and tsunami occurrences

The NASDAQ-100 is an index made up of the top 100 companies listed on the NASDAQ exchange, excluding firms in the financial sector. The data for the NASDAQ-100 analysis has been collected from the National Association of Securities Dealers Automated Quotations (NASDAQ - <https://www.nasdaq.com/>). With 5 years of historical data, ranging

from 6/3/2015 to 6/2/2020, we identified 315 days where opening prices jumped \$40.39 or more, a value equivalent to the third quartile of price changes. The subsequent dates were then used to create a vector of global times.

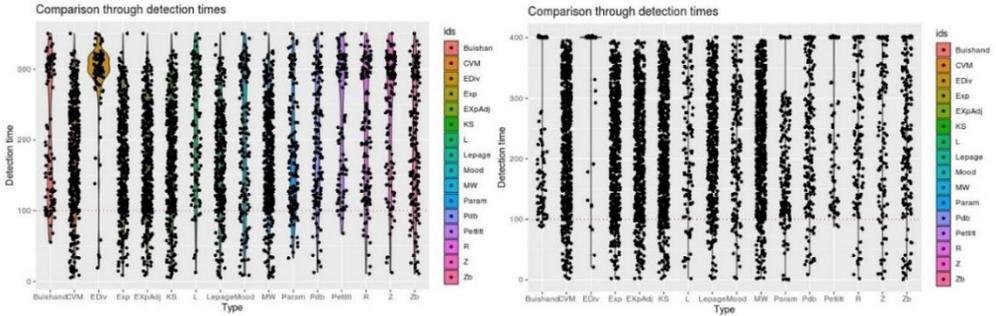


Fig. 4. Change identification under $\beta : 3.8 \rightarrow 0.8$: left panel - failure truncation, right panel - time truncation.

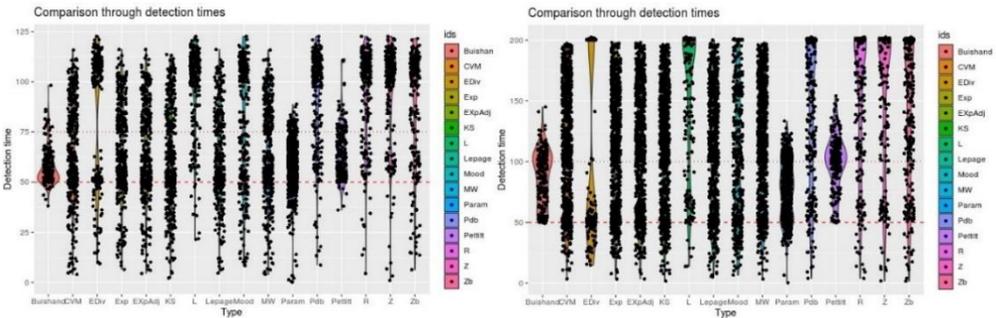


Fig. 5. Change identification under $\lambda_0 : 1 \rightarrow 2, \beta : 3.8 \rightarrow 0.8$: left panel - failure truncation, right panel - time truncation.

The raw data for the crude oil analysis has been collected from the National Association of Securities Dealers Automated Quotations (NASDAQ - <https://www.nasdaq.com/>). This institution collects daily data on stocks including, but not limited to, their opening and closing prices. Using 5 years of historical data, ranging from 6/4/2015 to 6/3/2020, we identified 314 days where opening prices jumped \$0.67 or more, a value equivalent to the third quartile of price changes. The subsequent dates were then used to create a vector of global times.

The data for the tsunami analysis has been collected from The National Centers for Environmental Information (NCEI - <https://www.ngdc.noaa.gov/ngdc.html>), a reliable source that maps real time data of natural phenomena such as earthquakes, tsunamis, and volcanoes. Using this source, we collected data on all tsunami instances between the years of 1900 and 2019. While all the data was accurately collected by the NCEI, some of the tsunami instance timings were missing exact dates. To remedy this issue we evenly sprayed the data while ensuring that each year maintained the exact number of tsunami instances it originally saw.

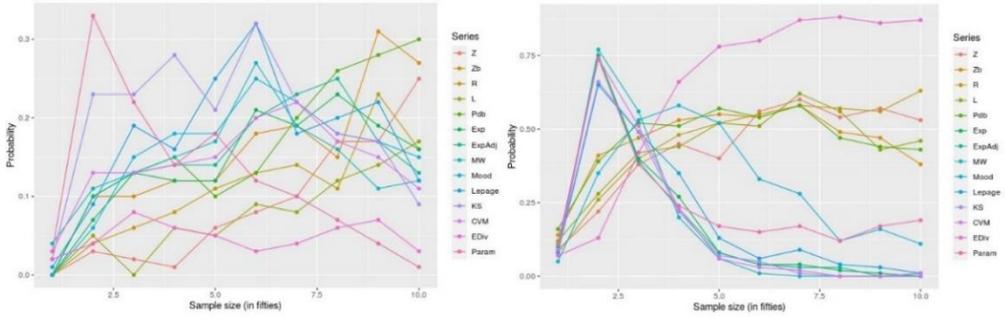


Fig. 6. Checking asymptotic consistency with $K = 2$, left panel - failure truncation, right panel time truncation.

Using the sprayed data we created a vector filled with the corresponding global times for the 1,316 tsunami occurrences.

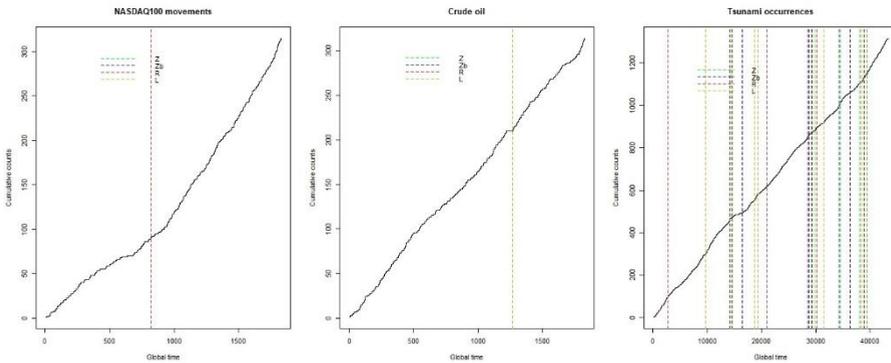


Fig. 7. Change point detection on the point processes storing (from left to right) NASDAQ oscillations, crude oil price fluctuations, and tsunami occurrences.

4.1.1 Bootstrapping

Through the NASDAQ 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) [19], 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) [9] 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) [20]’s block resampling algorithm to generate point processes similar to the original one.

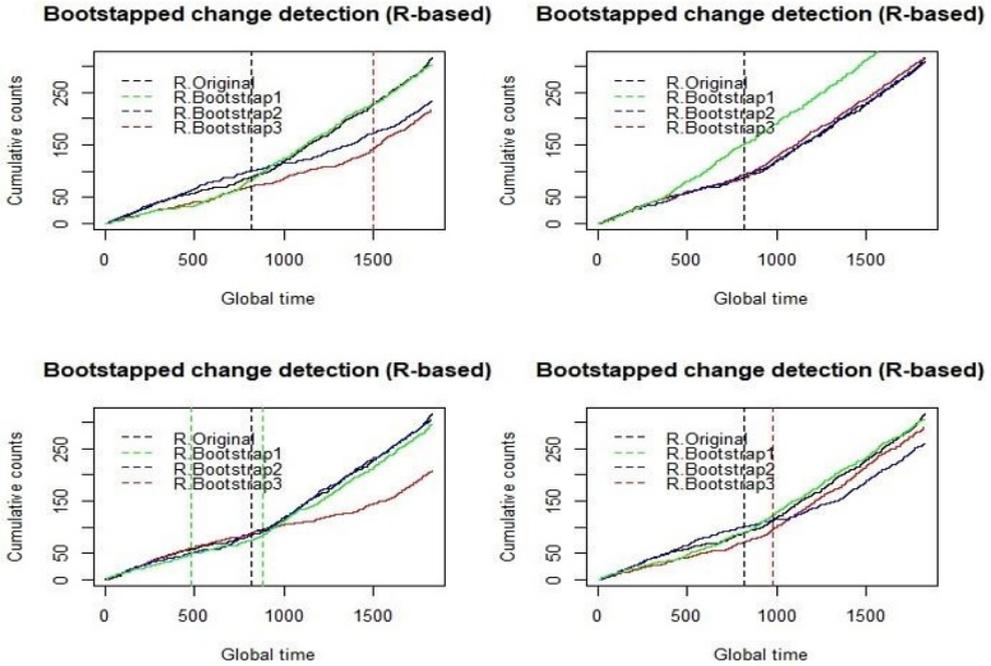


Fig. 8. Bootstrapping NASDAQ-100 and subsequently, detecting change-points, through R .

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, \dots, U_b uniformly and independently from $(0, T - \frac{T}{b}]$.
- for $j = 1, 2, \dots, b$:
 - 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 $X_j^*(\cdot) := |A_j^* \cap \cdot|$
- the simulated point process X^* is given by $X^* := \sum_{i=1}^b X_i^*$

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 $(0, T]$.

The method works best with roughly stationary point processes such as our NASDAQ or crude oil price examples. So, we choose the former 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 Fig. 8. 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 lower panel diagrams of Fig. 8, 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.

Table 4. 95% bootstrapped intervals from different change-point options.

Tool	Interval
Z	[207.7534,1408.9309]
ZB	[794.8049, 854.8034]
R	[843.5838,1263.2371]
L	[300.0461,1437.6212]
CPM-Exp	[204.3672,1445.7812]
CPM-ExpAdj	[378.2436,1446.7218]
CPM-MW	[319.5833,1552.2955]
CPM-Mood	[658.6435,1755.2883]
CPM-Lepage	[434.029, 1762.224]
CPM-KS	[333.9317,1589.5221]
CPM-CVM	[262.1857,1749.9598]
EDiv	[147.0139,1170.8838]
Pettitt	[515.5954,1430.2563]
Buishand	[405.9311,1377.8027]
Parametric	[465.0303,1483.3931]

4.2 Coronavirus infections in India

The initial COVID-19 data for India was obtained from Our World in Data1, a reputable source for coronavirus data. The data contained values on the number of infections per day in the country. India had observed its first case of covid-19 on January 30th 2020 and the virus has spread rapidly across the country since then. The covid-19 data was obtained on July 10th 2020 and by then the country had observed 793,802 infections. As the data contained values on the number of infections per day in the country, it had to be “sprayed” first to obtain the global times of each infection to be able to conduct the change point analysis. This was done by using the uniform distribution to randomly distribute the number of infections on each day across the 24-hour period. The final dataset that was used for the analysis contains 793,802 entries, one entry or global time for each individual infection in the country (As of July 10th 2020).

A few observations are in order:

- The estimates of the true change points are visually acceptable. Very often, the optimal guesses are placed at around points in time where the underlying process undergoes significant improvements or deteriorations, seen through prominent edges. This is especially vivid in the first panel figures where relatively sparse crowding of change estimates enable one to visualize the twists and turns. In particular, the sharp rise in cases around a global time of 800 is detected by all proposals.
- Our algorithm picks up the complexity of the data. As the number of cases rise (please notice the increase in the vertical scale over time), the likelihood of the underlying point process suffering changes either through exogenous (e.g., visitors arriving from other countries) or endogenous (e.g., community transmission) impacts rises as well. It is natural, therefore, to expect the number of change detections in such a case to correlate with the data volume. This is evidenced through the graphs in the lower panel analyzing data over the more recent months, where a heavy crowding of detected change points suggests the underlying process must have gone through several regime shifts.

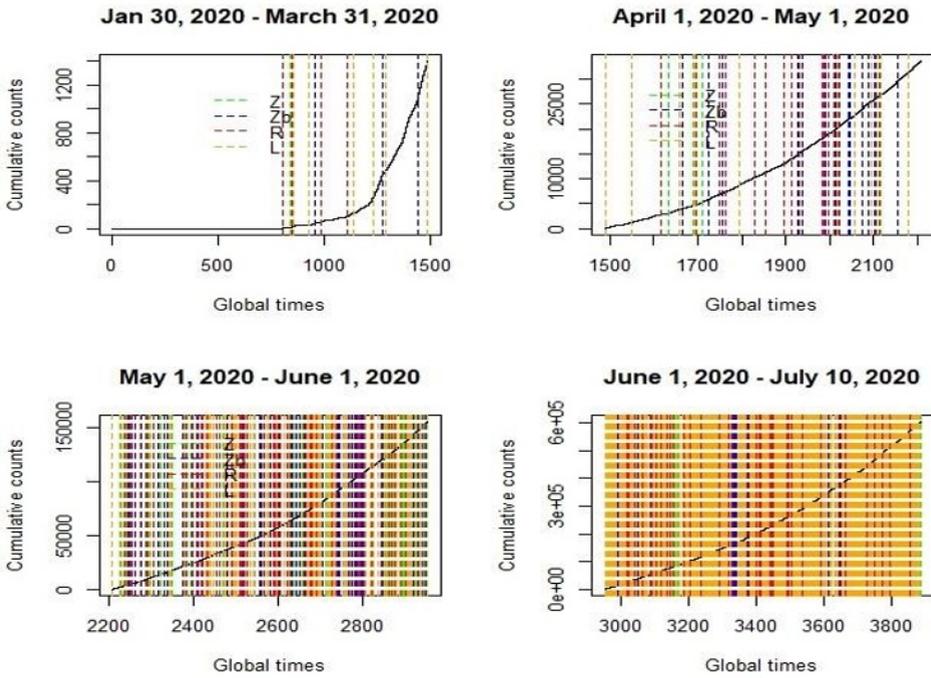


Fig. 9. Change point detection on the point processes storing COVID-19 infections in India over different months.

- Too many detections often raise a valid concern over false alarms. The asymptotic consistency result described in the previous section, however, establishes our method is considerably free of such pitfalls. Consequently, the change points we detect may, with some confidence, be claimed as valid estimates, not through random noise.
- Despite the crowding in the bottom panel, gaps are noticeable. Those periods are ones over which the sequence is expected to stay relatively stable. Probing deeper, we found these gaps make intuitive sense. Those are times around which different states or the country as a whole, imposed preventive measures such as "stay-at-home" requests or preventing travelers from other countries from visiting.

4.2.1 A note on forecasting

We observe in passing that under a simplifying assumption of $\alpha = 0$ and $\lambda_0(t) := \sum_{i=1}^2 \lambda_i I_{(0,t]}(t)$, change point-based forecasts can be had too. We exploit the connection between a homogeneous Poisson process and the exponential distribution: a simple point process is a homogeneous Poisson with rate λ if and only if the inter-event times are exponentially distributed with rate $1/\lambda$. Using an estimated change point $\hat{\tau}$, we can break up the data into a pre-change and a post-change piece and calculate the mean inter-event time on each. If the post-change step height is estimated to be $\hat{\lambda}_2$, then to wait for say k future events, one needs to wait for $k/\hat{\lambda}_2$ time units. Using $\hat{\tau} = 1135.1752$ from our L -based test, we found $\hat{\tau}_2 = 3.5968$. So to wait for around 4 more infections, one needs to wait roughly one time unit. In practice, we found this count over the next time unit to be close at 3. Such

an analysis may aid health providers estimate the number of hospital beds or respirators to arrange over the immediate future. The validity of the two-step model can be confirmed through a goodness-of-fit test.

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. Ordinary point processes of the stationary Poisson type frequently fall short to model such 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 first 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-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 NASDAQ-100 movements, tsunami occurrences, crude-oil prices, and COVID-19 infections in India demonstrate their easy applicability and graphic appeal. Model-based accurate forecasting tools are briefly described, with data-dependent ones in the works.

References

1. S.E. Rigdon, A.P. Basu, Statistical methods for the reliability of repairable systems (John Wiley and Sons, 2000)
2. M. Bhaduri, Bi-directional testing for change point detection in Poisson processes UNLV Theses, Dissertations, Professional Papers, and Capstones. 3217 (2018)
3. A. Hawkes, *Biometrika* **58**(1), 83 – 90, doi:10.2307/2334319 (1971)
4. M. Prince-Williams, N.A. Heard, *Stat. Comput.* **30**, 209 – 220 (2020)
5. M. Rambaldi, V. Filimonov, F. Lillo, Fabrizio, *Phys. Rev. E* **97**(3), 032 – 318 (2018)
6. F. Zhan, et al., *IEEE Access* **7**, 140860 – 140874 (2019)
7. M. Bhaduri, C.-H. Ho, *Environ. Model. Assess.* **24**, 495 – 507 (2018)
8. M. Bhaduri, J. Zhan, *IEEE Access* **6**, 30855 – 30864 (2018)
9. C.-H. Ho, M. Bhaduri, *Math. Geosci.* **49**, 893 – 911 (2017)

10. M. Bhaduri, J. Zhan, C. Chiu, *IEEE Access* **5**, 27354 – 27365 (2017)
11. M. Bhaduri, J. Zhan, C. Chiu, F. Zhan, *IEEE Access* **5**, 15883 – 15892 (2017)
12. C.-H. Ho, G. Zhong, F. Cui, M. Bhaduri, *J. Finance and Bank Management* **4**, 15 – 33 (2016)
13. C.-H. Ho, M. Bhaduri, *Natural Hazards* **78**, 669 – 679 (2015)
14. S. Tan, M. Bhaduri, C.-H. Ho, *J. Geoscience and Environment Protection* **02**, 16 – 26 (2014)
15. D.M. Hawkins, P.H. Qiu, C.W. Kang, *J. Qual. Technol.* **35**(4), 355 – 366 (2003)
16. M. Bhaduri, *Sci Rep.* **10**, 15889, <https://doi.org/10.1038/s41598-020-72803-z>. (2020)
17. Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc. Series B (Methodological)* **57**(1), 289 – 300 (1995)
18. C. Truong, L. Oudre, N. Vayatis, *Signal Processing*, **167**, 107 – 299 (2020)
19. B. Efron, *Ann. Statist.* **7**(1), 1 – 26 (1979)
20. W. Braun, R.J. Kulperger, *J. Stat. Comput. Simul.* **60**, 10.1080/00949659808811878 (1998)