Predictive time-rescaling for goodness-of-fit assessment of self-exciting temporal point processes

M.-A. El-Aroui

RBS, International University of Rabat, Technopolis, Sala-Al-Jadida, Morocco.

ABSTRACT

This work introduces a new predictive time-rescaling methodology designed to provide an asymptotically unbiased goodness-of-fit (Gof) framework for general self-exciting temporal point processes. Specifically addressing the analysis of single observed trajectories, the proposed approach mitigates the inherent bias introduced by *plugged-in* parameter estimates by focusing on forecasting accuracy. Using a predictive-sequential procedure, model validation is centred on the precision of forecasted arrival times of events. We demonstrate that these times, when transformed via sequentially estimated parameters, converge in probability to vectors of iid. exponential random variables with unit mean (Exp(1)) under standard regularity conditions. The framework's efficacy is validated through numerical simulations comparing standard and predictive time-rescaling for non-homogeneous Poisson and Hawkes processes, further supported by an application to Japanese seismic data.

KEYWORDS

Model checking; Gof-tests; KS-plots; Hawkes processes; Non-homogeneous Poisson processes.

1. Introduction

Self-exciting temporal point processes (SETPPs) are frequently used to model, analyse and forecast events' arrival times in several fields: seismology, neuroscience, cyber-security, finance, infectious diseases, social networks modelling, criminology, etc.

As it was noted by many authors point process models are getting more complex but corresponding tools for goodness-of-fit (Gof) assessment "*have lagged behind, restricted mostly to the spatial statistics literature*" according to [15]. In neuroscience, where SETPPs are used to analyse spike-train data and infer the coding properties of neural systems across different brain areas, [13] noted an increasing "*need for statistical modelling and goodness-of-fit tools that can address neural coding problems at the population level*".

This work studies the use of time-rescaling to assess the Gof of SETPP models. It first highlights a significant, yet frequently overlooked, limitation in the conventional application of the time-rescaling theorem: the systematic bias introduced when model parameters are replaced by their estimates. A predictive time-rescaling methodology is introduced yielding an asymptotically unbiased goodness-of-fit test for general self-exciting temporal point processes, contingent upon standard regularity conditions.

Let us assume that a recurrent phenomenon is observed at times $T_1, T_2, \ldots$ modelled

Contact: M.-A. El-Aroui. Email: Mhamed-Ali.Elaroui@uir.ac.ma and Mhamed.Elaroui@gmail.com

as arrival times of a point process N which counting process is $\{N_t\}_{t\geq 0}$ is defined on a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. The point process N is assumed *self-exciting*.

It will be assumed here that a single point process (or a single trajectory) is observed. The conditional intensity defines entirely the probability distribution of the self-exciting process:

$$\lambda(t|\mathcal{F}_{t-}) := \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}\{N_{t+\Delta t} - N_{t-} = 1 | \mathcal{F}_{t-}\}$$

where $\mathcal{F}_{t-}$ denotes the sigma-algebra defined by the past of the process N before t. The cumulative intensity is $\Lambda(t|\mathcal{F}_{t-}) := \int_0^t \lambda(u|\mathcal{F}_{u-})\, du$. The notations will be simplified below by replacing $\lambda(t|\mathcal{F}_{t-})$ and $\Lambda(t|\mathcal{F}_{t-})$ by $\lambda(t)$ and $\Lambda(t)$.

The process of events arrivals is modelled using the following parametric formulation for the conditional intensity:

$$\mathcal{M} = \{\lambda(.,\theta);\ \theta \in \Theta \subset \mathbb{R}^k\}. \tag{1}$$

Standard regularity conditions on $\lambda(.,\theta)$ and $\Lambda(\bullet,\theta) := \int_0^\bullet \lambda(u,\theta)\, du$ (existence and continuity of first and second derivatives wrt θ as stated for example in [12] or [10]) will be assumed in the following.

This work advances the field by introducing a predictive time-rescaling approach that compensates for the bias inherent in *plug-in* parameter estimation. By addressing these distributional shifts, we provide a more robust framework for SETPP model validation, specifically improving the reliability of standard tools like Kolmogorov-Smirnov plots (KS-plots) and Gof tests for $\mathcal{M}$, i.e. statistical tests for the null hypothesis:

$$\mathrm{H}_0 :\ \text{`}\lambda(.) \in \mathcal{M}\text{'} \qquad \text{vs} \qquad \mathrm{H}_1 :\ \text{`}\lambda(.) \notin \mathcal{M}\ \text{'}. \tag{2}$$

Notations –

(1) The studied phenomenon will be observed until n events are recorded.
(2) Let $T_1, \ldots, T_n$ denote the first n occurrence times with $T_0 = 0$.
(3) Let $X_1, \ldots, X_n$ denote the first n inter-events times: $X_i = T_i - T_{i-1}$ for $i = 1, \ldots, n$.
(4) Let θ_0 denote the *true* value of the unknown parameter θ under the null hypothesis H_0.
(5) The Kolmogorov-Smirnov distribution is denoted $\mathcal{L}_{ks}$.

Since the initial work of [9] time-rescaling has been frequently used to assess the Gof of SETPP models by checking the iid-Expon.(1) assumption of the rescaled arrival times. This check is usually based on KS-plots or Gof tests (mainly KS or Chi-square). Generally users of this approach use exponentiality checking tools after replacing the unknown parameters by their estimates. Nevertheless this plug-in step invalidates the iid-Expon.(1) property for the rescaled times. The practical effects of this plug-in step have been most often overlooked by practitioners. In the neuroscience field, [11] mentioned this problem when predicting neuron spike occurrences and developed an alternative solution (based on random sub-sampling) in the case of several observed

trajectories.

Three different approaches of time-rescaling will be compared in this paper:

(1) `TR0`: time-rescaling with the true parameters θ_0 (unavailable in real life applications) used in the data-generating process.
(2) `TR1`: standard time-rescaling with estimated parameters $\widehat{\theta}_n$ using the whole dataset.
(3) `TR2`: predictive time-rescaling with sequentially estimated parameters. This approach will be introduced in section 3.

The previous procedures of time-rescaling will be analysed through intensive simulations for two families of SETPPs: non-homogeneous Poisson processes (NHPP) and Hawkes processes. The most used models for NHPP are the power-law process (PLP) and the log-linear process (LLP). The conditional intensities of these two NHPP write $\lambda(t) = \alpha\, h(t)$ where $\alpha > 0$ is a scale parameter and h is a baseline hazard rate function (given by $h(t) = \beta t^{\beta-1}$ for PLP and by $h(t) = \exp(\beta t)$ for LLP). Hawkes processes (see for example [7]) are self-exciting point processes which conditional intensities write: $\lambda(t) = \mu + \sum_{i:t_i<t} \phi(t - t_i)$, where μ is called the *background* intensity and ϕ is the *triggering* function (given by $\phi(t) = k/(c+t)^p$ for Power-law and by $\phi(t) = \alpha \exp(-\beta t)$ for exponential decay Hawkes processes).

Section 2 introduces the use of standard time-rescaling to assess the Gof of NHPPs and Hawkes processes. An alternative unbiased predictive version of time-rescaling is presented in section 3. This predictive time-rescaling is used to develop KS-plots and KS Gof tests for SETPPs taking into consideration both the stochastic impact of parameters' estimation and models' predictive accuracy. A numerical analysis of the predictive time-rescaling approach is presented in section 4. A final discussion is presented in the last section.

2. Time-rescaling in Gof assessment of SETPPs

The literature provides several Gof tests in the case of multi-trajectory observed point processes (see for example [1] and [4]). Gof tests based on single observed trajectories are rather scarce (see for example [2] or [8] in the case of NHPPs or for more general SETPPs: [15] and [14] who developed interesting but rather algorithmically tedious Gof approaches).

Users of SETPPs generally check their model-Gof using the time-rescaling theorem with the usual diagnosis tools of the iid-Expon.(1) assumption after replacing the unknown parameters by their estimates when transforming arrival times. But this plug-in step invalidates the main result of the time-rescaling theorem and leads to biased model-checking diagnoses.

Model checking of SETPPs is most frequently based on the following theorem ([9]).

Theorem 2.1. *Let $0 \leq T_1 < T_2 < \cdots < T_{n-1} < T_n \leq T$ be the set of event times of a SETPP observed on a period $[0, T]$. Define, for $i \in \{1, \ldots, n\}$, the transformation*

$$\Lambda(T_i) = \int_0^{T_i} \lambda(u)\, du$$

Then the $\Lambda(T_i)$'s are distributed as the times of the first n events of a Poisson process

with unit rate.

$$\textit{Let} \qquad E_i := \Lambda(T_i) - \Lambda(T_{i-1}) \qquad \textit{for } i \in \{1, \ldots, n\} \tag{3}$$

denote the rescaled inter-arrival times *are iid-Expon.(1) rv's.* □

As mentioned by Brown et al. 2002 [3], "*The time-rescaling theorem generates a history-dependent rescaling of the time axis that converts a point process into a Poisson process with a unit rate*".

Practitioners usually assess the Gof of SETPP models by replacing in Theorem 2.1 the true (but unavailable) rescaled inter-arrival times $E_i = \Lambda(T_i) - \Lambda(T_{i-1})$ by their estimated counterparts: $\widehat{E}_i := \Lambda(T_i, \hat{\theta}_n) - \Lambda(T_{i-1}, \hat{\theta}_n)$ for $i \in \{1, \ldots, n\}$.

Following [3], two specific diagnosis tools are frequently used to check whether the estimated rescaled inter-arrival times $\widehat{E}_i$ behave like iid-Expon.(1) rv's:

(1) Kolmogorov-Smirnov (KS) Gof-tests: calculate the following KS-statistic:

$$\widehat{K}_n := \sqrt{n} \max \left[\max_{1 \le i \le n} \left(\frac{i}{n} - \widehat{U}_{(i)} \right) \ , \ \max_{1 \le i \le n} \left(\widehat{U}_{(i)} - \frac{i-1}{n} \right) \right] \tag{4}$$

where $\{\widehat{U}_{(i)}\}_{1 \le i \le n}$ are the (increasing) order statistics of the transformed rv's:

$$\widehat{U}_i := 1 - \exp(-\widehat{E}_i) = 1 - \exp\big(- \Lambda(T_i, \hat{\theta}_n) + \Lambda(T_{i-1}, \hat{\theta}_n)\big) \qquad \text{for } i \in \{1, \ldots, n\}. \tag{5}$$

The tested model will be rejected when $\widehat{K}_n$ exceeds the appropriate quantile of Kolmogorov-Smirnov distribution $\mathcal{L}_{ks}$.

(2) KS-plots or quantile-quantile plots: the quantiles of the Uniform[0,1] distribution approximated for $i \in \{1, \ldots, n\}$ by $b_i = (i - 0.5)/n$ are plotted against $\hat{u}_{(i)}$'s: the sorted observed values of the transformed rv's $\widehat{U}_i$. If the model has a good fit, then the points $(\hat{u}_{(i)}, b_i)$ should lie close to the diagonal line. Confidence bounds for the degree of agreement between the models and the data may be constructed using $\mathcal{L}_{ks}$. For moderate to large sample sizes the 95% (99%) confidence bounds are well approximated by $b_i \pm 1.36/n^{1/2}$ ($b_i \pm 1.63/n^{1/2}$).

Practitioners in several fields assess their models' Gof by checking the iid-Expon.(1) property on the estimated rescaled inter-arrival times $\widehat{E}_i$'s. This Gof procedure supposes that the iid-Expon.(1) property of Theorem 2.1 remains approximately valid when the estimated parameters $\hat{\theta}_n$ are plugged-in. More precisely it assumes that the iid-Expon.(1) property of the theoretical rescaled inter-arrival times E_i's can be extended to their estimated counterparts $\widehat{E}_i$'s. Several simulation results (see [6]) show that this assumption does not hold and leads systematically to biased diagnosis results.

It is clear that the standard time-rescaling `TR1` used to test the assumption iid-Expon(1) on $\widehat{E}_i$'s is biased since the rejection-percentages are systematically less than the nominal sizes δ. This is not the case for time-rescaling `TR0` based on the true parameters nor the case for the predictive time-rescaling `TR2` that will be described in section 3.

3. A predictive time-rescaling with sequentially estimated parameters

This section examines a predictive formulation of the time-rescaling theorem designed to mitigate the systematic bias identified previously. This asymptotically unbiased framework leverages Dawid's predictive-sequential (prequential) approach [5] to ensure more robust model assessment. As the events arrive sequentially, after each arrival i the available observed arrival times $t^{(i)} \equiv (t_1, \ldots, t_i)$ are used to update the estimate of the unknown parameter θ and then predict the time of the next event T_{i+1}. The Gof approach is based on the statistical assessment of the accuracy of these successive predictions.

Notations –

- For $i \leq n$ let $\hat{\theta}_i$ denote the ML estimator of θ using the sample of observed arrival times $(t_1, \ldots, t_i)$ available after the i-th event.
- For all $t \in]t_{i-1}, t_i]$, let $\hat{\theta}_{[t^-]} := \hat{\theta}_{i-1} = \hat{\theta}(t_1, \ldots, t_{i-1})$.
- m will denote a minimal sample-size below which the ML estimator of θ is inefficient.

Definition 3.1. At each time t, let the *prequential intensity* be $\widetilde{\lambda}(t) := \lambda_t(\hat{\theta}_{[t^-]})$.

The *prequential cumulative intensity* is: $\widetilde{\Lambda}(t) := \displaystyle\int_0^t \widetilde{\lambda}(u)\, du = \int_0^t \lambda(u, \hat{\theta}_{[u^-]})\, du.$

Definition 3.2. For $i \in \{m+1, \ldots, n\}$ the *predictive rescaled inter-arrival times* are defined by

$$\begin{aligned} \widetilde{E}_i &:= \widetilde{\Lambda}(T_i) - \widetilde{\Lambda}(T_{i-1}) \\ &= \int_{T_{i-1}}^{T_i} \lambda(u, \hat{\theta}_{[u^-]})\, du = \Lambda(T_i, \hat{\theta}(T_1, \ldots, T_{i-1})) - \Lambda(T_{i-1}, \hat{\theta}(T_1, \ldots, T_{i-1})). \quad (6) \end{aligned}$$

3.1. A predictive asymptotically unbiased time-rescaling

The following assumptions and regularity conditions will be assumed:

A1. The studied SETPP will be supposed orderly and non-explosive with $T_n \xrightarrow[\text{prob.}]{n\to\infty} \infty$ and $n^{-1/2} X_n \xrightarrow[\text{prob.}]{n\to\infty} 0$.

A2. $\lambda(t, \theta)$ is $\mathcal{F}_t-$predictable, continuous in θ, non-negative almost-surely for all $\theta \in \Theta$ and all $t > 0$. $\displaystyle\int_0^t \lambda(u, \theta)\, du < +\infty$ for all $t \geq 0$ with probability 1.

A3. For all i and j in $\{1, \ldots, k\}$, $\dot{\lambda}_i(t, \theta) := \frac{\partial \lambda(t,\theta)}{\partial \theta_i}$ and $\ddot{\lambda}_{ij}(t, \theta) := \frac{\partial^2 \lambda(t,\theta)}{\partial \theta_i \partial \theta_j}$ exist and are continuous in θ for all $\theta \in \Theta$ and all $t > 0$.

A4. For all j in $\{1, \ldots, k\}$ let $\dot{\Lambda}_j(t, \theta) := \frac{\partial \Lambda(t,\theta)}{\partial \theta_j}$. There exists an integer i_0 such that for all $i \geq i_0$, with probability 1, $\left\| \dot{\Lambda}_j(T_{i-1}, \theta_0) - \dot{\Lambda}_j(T_{i-1} + X_i, \theta_0) \right\| \leq M_1$ for some $M_1 < \infty$.

A5. The ML estimator $\hat{\theta}$ satisfies the following two conditions:
   a. $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$.

b. $\lambda(t,\hat{\theta}_n) - \lambda(t,\theta_0) \overset{n\to+\infty}{\underset{\text{Prob.}}{\longrightarrow}} 0$ uniformly in t.

Similarly to Theorem 2.1, it is proven that for any fixed integer d, under H_0 and the previous assumptions, $d-$vectors of the predictive rescaled inter-arrival times $\{\widetilde{E}_i\}_{i=m+1,\ldots,m+d}$ converge (in probability when $m \to \infty$) to iid-Expon.(1) rv's.

Proposition 3.3. If Assumptions $(A1,\ldots,A5)$ hold, then under H_0 and for any fixed $d \in \mathbb{N}^*$:

$$(\widetilde{E}_{m+1},\ldots,\widetilde{E}_{m+d}) - (E_{m+1},\ldots,E_{m+d}) \overset{m\to+\infty}{\underset{\text{Prob.}}{\longrightarrow}} 0 \qquad \text{where } E_i's \text{ are iid-Expon.(1).} \quad (7)$$

Proof – The proof is detailed in [6].

4. Gof checking of SETPPs based on predictive time-rescaling

4.1. Predictive KS-plots and KS Gof-tests

The predictive Gof checking is based on testing the iid-Expon.(1) hypothesis for the sequence $\{\widetilde{E}_i\}_{m<i\leq n}$ or equivalently the iid-Unif.[0,1] hypothesis for the sequence $\{\widetilde{U}_i := 1-\exp(-\widetilde{E}_i)\}_{m<i\leq n}$. This could be done using a KS-plot i.e. an Expon.(1) QQ-plot for $\{\widetilde{E}_i\}_{m<i\leq n}$ or a Unif.[0,1] QQ-plot for $\{\widetilde{U}_i\}_{m<i\leq n}$. A KS Gof-test can also be used by comparing (using KS distances denoted in the following $\widetilde{K}_{m,n}$) the empirical distribution function (Edf) of $\widetilde{U}_i$'s to the Cdf of the Unif.[0,1] distribution. The predictive Gof-test will therefore calculate $\widetilde{K}_{m,n}$ and compare it to upper quantiles of $\mathcal{L}_{ks}$. $\widetilde{K}_{m,n}$ are given by the following equation:

$$\widetilde{K}_{m,n} := (n-m)^{1/2} \max\left[\max_{1\leq i\leq n-m}\left(\frac{i}{n-m} - \widetilde{U}_{(i)}\right) \;,\; \max_{1\leq i\leq n-m}\left(\widetilde{U}_{(i)} - \frac{i-1}{n-m}\right)\right] \quad (8)$$

where m is a suitably chosen integer ($m = n/2$ and $m = n/5$ are used for the numerical experiments of subsection 4.3) and $\{\widetilde{U}_{(i)}\}_{1\leq i\leq n-m}$ are the (increasing) order statistics obtained from the transformed rv's $\{\widetilde{U}_i\}_{m<i\leq n}$.
The predictive KS Gof-test calculates $\widetilde{K}_{m,n}$ and compares it to upper quantiles of $\mathcal{L}_{ks}$. H_0 will be rejected when $\widetilde{K}_{m,n}$ exceeds the appropriate quantile of $\mathcal{L}_{ks}$.

4.2. Implementation for NHPP and Hawkes processes

For NHPP models with intensity function $\lambda(.)$, the predictive model checking based on `TR2` is easily derived using equation (6) where $\Lambda(t,\theta) = \alpha t^\beta$ for NHPP-PLP and $\Lambda(t,\theta) = \alpha\beta^{-1}(\exp(\beta t)-1)$ for NHPP-LLP.

For Hawkes processes with a power-law function the predictive rescaled inter-arrival

times are given (for $i = m+1, \ldots, n$) by

$$\begin{aligned}\widetilde{E}_i &= \int_{T_{i-1}}^{T_i} \lambda(s, \hat{\theta}_{[s^-]})\, ds \qquad (9)\\ &= \hat{\mu}_{i-1}(T_i - T_{i-1}) + \frac{\hat{k}_{i-1}}{1-\hat{p}_{i-1}} \sum_{j=1}^{i-1} \Big[\big(\hat{c}_{i-1} + T_i - T_j\big)^{1-\hat{p}_{i-1}} - \big(\hat{c}_{i-1} + T_{i-1} - T_j\big)^{1-\hat{p}_{i-1}} \Big].\end{aligned}$$

Hawkes processes with an exponential decay have the following predictive rescaled inter-arrival times:

$$\begin{aligned}\widetilde{E}_i &= \int_{T_{i-1}}^{T_i} \lambda(s, \hat{\theta}_{[s^-]})\, ds \qquad (10)\\ &= \hat{\mu}_{i-1}(T_i - T_{i-1}) - \frac{\hat{\alpha}_{i-1}}{\hat{\beta}_{i-1}} \sum_{j=1}^{i-1} \Big[\exp\big(-\hat{\beta}_{i-1}(T_i - T_j)\big) - \exp\big(-\hat{\beta}_{i-1}(T_{i-1} - T_j)\big) \Big].\end{aligned}$$

4.3. Numerical Experiments

The finite-sample validity of Proposition 3.3 is checked by the following four step simulation approach:

(1) A fixed value θ_0 of θ is chosen.
(2) M data sets (of size n each, with $n \in \{50, 100, 200\}$) are simulated using the model with intensity $\lambda(., \theta_0)$.
(3) For the j-th simulated data-set, the realization $\widetilde{k}_{m,n}^{(j)}$ of $\widetilde{K}_{m,n}$ is calculated.
(4) The Edf of $\{\widetilde{k}_{m,n}^{(j)}\}_{1 \le j \le M}$ is compared to the Cdf of $\mathcal{L}_{ks}$ by comparing their upper quantiles denoted respectively $\tilde{q}_\delta$ and q_δ^{ks} for a set of seven significance levels $\delta \in \{0.50, 0.25, 0.15, 0.10, 0.05, 0.025, 0.01\}$. The mean absolute relative error MARE $= \frac{1}{7}\sum_{\delta \in \{0.50,\ldots,0.01\}} \big(|\tilde{q}_\delta - q_\delta^{ks}|\big)/q_\delta^{ks}$ measures the distance between the Edf of $\{\tilde{k}_{m,n}^{(j)}\}_{j \le M}$ and $\mathcal{L}_{ks}$, and therefore the finite-sample relevance of Proposition 3.3.

Results presented in figures 1 and 2 suggest that Proposition 3.3 holds in the finite-sample context since $\widetilde{K}_{m,n}$'s Edf (in dotted blue) are very close to $\mathcal{L}_{ks}$'s Cdf in black solid lines. This was not the case with the standard time-rescaling `TR1` based on a single estimation step using the whole data set since the curves of $\widehat{K}_n$'s Edf (in dotdash red) are shifted from to the Cdf of $\mathcal{L}_{ks}$.
Table 1 (for NHPP) and table 2 (for Hawkes) compare numerically the upper quantiles of $\widetilde{K}_{m,n}$'s Edf and those of $\mathcal{L}_{ks}$. The last columns of these tables give small MARE's (between 2% and 3%) suggesting that assumptions A1-A5 used to prove Proposition 3.3 hold for the studied NHPP and Hawkes processes. For the standard time-rescaling `TR1` these MARE values were around 18% for NHPP and 30% for Hawkes processes. Table 3 compares the MAREs between the standard $\mathcal{L}_{ks}$ distribution and the $\widehat{K}_n$'s Edf (for the usual `TR1` approach) to the MAREs between the $\mathcal{L}_{ks}$ and $\widetilde{K}_{m,n}$'s Edf. Table 3 shows that the predictive time-rescaling `TR2` is much less biased than the usually used time-rescaling `TR1` approach.
In these numerical experiments $M = 1000$ replications were simulated for each one of the considered processes. When different parameters' values were experimented the results were globally very similar to those presented in tables 1 and 2 confirming that

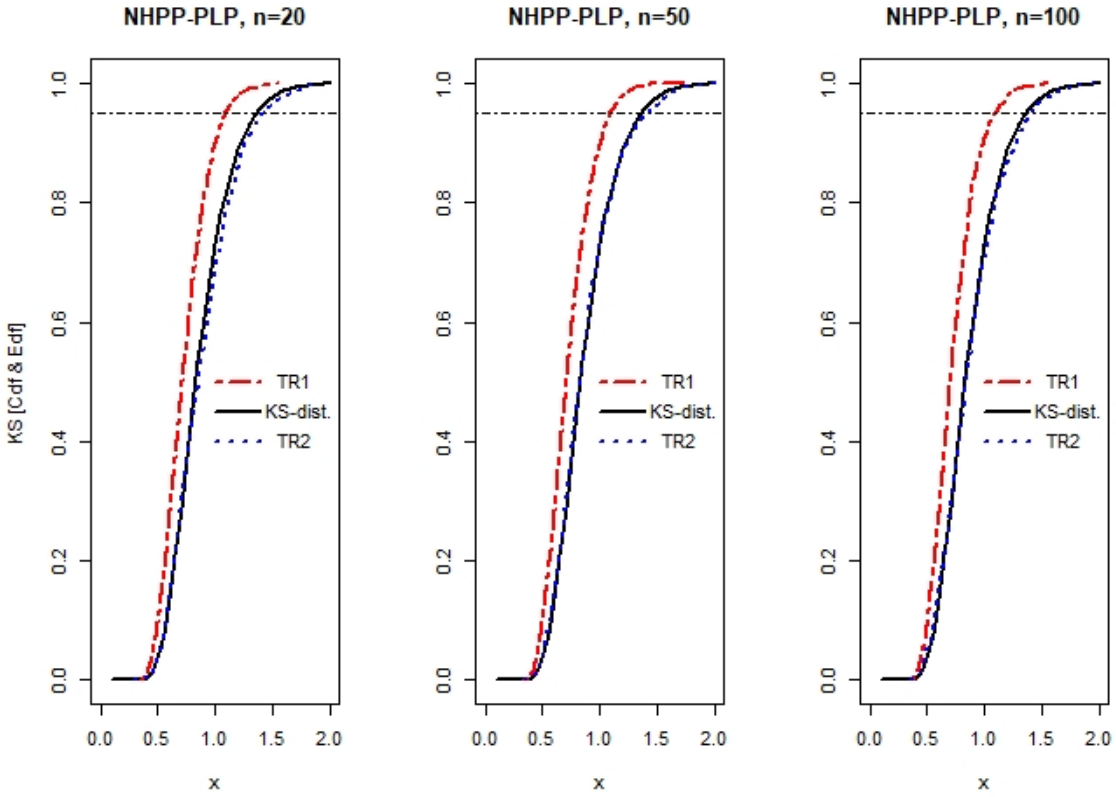

Figure 1. Cdf of $\mathcal{L}_{ks}$ (in black) with Edf of the KS-statistics $\widehat{K}_n$ (in dotdash red) and Efd of the predictive $\widetilde{K}_{m,n}$ (in dotted blue). NHPP-PLP. $(\alpha, \beta) = (1, 3)$.

Proposition 3.3 is valid for a wide range of models and parameter values.

Table 4 shows the size of KS-Gof tests (percentage of rejecting the right model) with the three time-rescaling approaches `TR0`, `TR1` and `TR2` (with $m = n/2$). It is clear that the predictive time-rescaling `TR2` is much less biased than the standard time-rescaling `TR1` since the rejection-percentages for `TR2` are much closer to the nominal levels δ.

While these experimental results are promising, further investigation is required to enhance the statistical power of these predictive Gof procedures. Specifically, future research should aim to establish a theoretical basis for model identifiability determining which self-exciting point process models can be effectively distinguished within a single-trajectory observational framework and which remain indistinguishable given the available data constraints.

4.4. Modelling arrivals of earthquakes in Japan (end of 2019)

The proposed predictive Gof framework is validated using an empirical dataset of seismic activity recorded near the Japanese islands in late 2019, as documented by [7]. This sample consists of 200 seismic events occurring between October and December 2019 (visualized in Figure 3). Hawkes processes are conventionally employed for seismic modelling due to their capacity to capture temporal clustering, specifically the

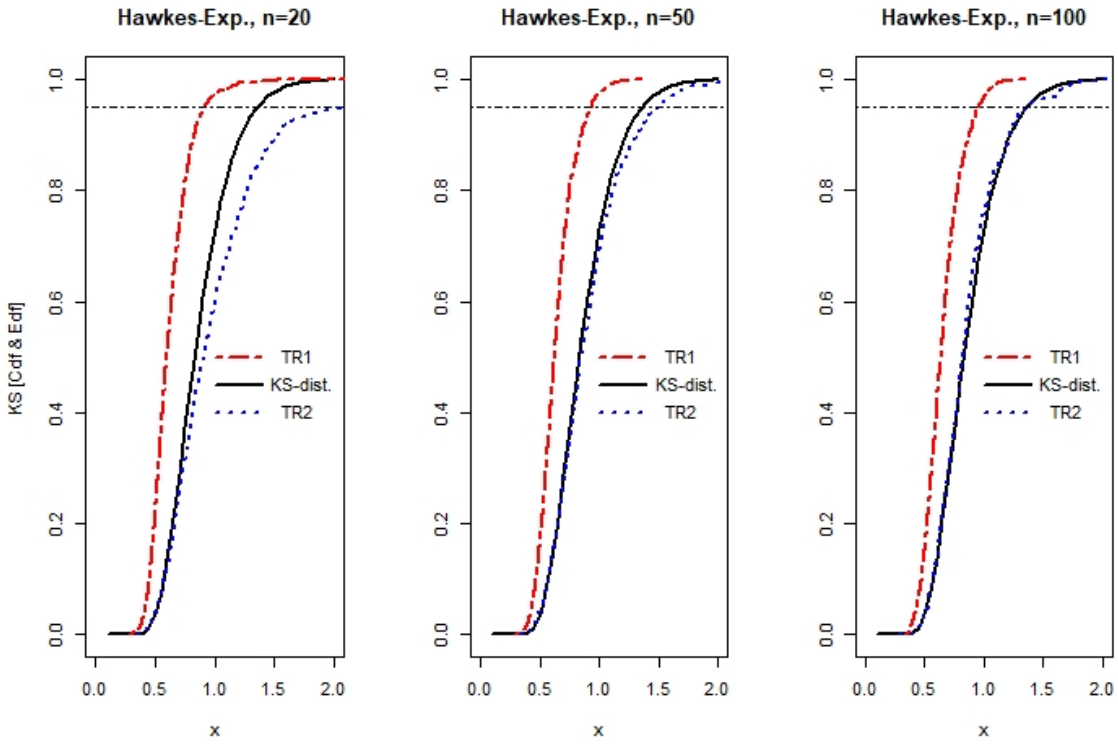

Figure 2. Cdf of $\mathcal{L}_{ks}$ (in black) with Edf of the KS-statistics $\widehat{K}_n$ (in dotdash red) and Efd of the predictive $\widetilde{K}_{m,n}$ (in dotted blue). Hawkes-Exp. $(\mu, \alpha, \beta) = (0.3, 1.2, 1, 5)$.

Signif. level δ	0.50	0.25	0.15	0.10	0.05	0.025	0.01	MARE
$\mathcal{L}_{ks}$ quantiles	0.822	1.019	1.138	1.224	1.358	1.480	1.628	
				NHPP-PLP. $(\alpha, \beta) = (1, 3)$				
$n = 50$	0.829	1.032	1.149	1.257	1.428	1.568	1.720	
	(0.9%)	(1.3%)	(1.0%)	(2.7%)	(5.2%)	(5.9%)	(5.7%)	3.3%
$n = 100$	0.817	1.030	1.167	1.278	1.393	1.492	1.688	
	(0.5%)	(1.1%)	(2.5%)	(4.4%)	(2.6%)	(0.8%)	(3.7%)	2.3 %
$n = 200$	0.817	1.017	1.158	1.250	1.346	1.468	1.595	
	(0.6%)	(0.2%)	(1.8%)	(2.1%)	(0.8%)	(0.8%)	(2.0%)	1.2 %
				NHPP-LLP. $(\alpha, \beta) = (0.01, 0.05)$				
$n = 50$	0.833	1.051	1.191	1.285	1.415	1.540	1.689	
	(1.4%)	(3.2%)	(4.6%)	(5.0%)	(4.2%)	(4.1%)	(3.8%)	3.7%
$n = 100$	0.848	1.025	1.164	1.245	1.414	1.543	1.643	
	(3.2%)	(0.6%)	(2.3%)	(1.7%)	(4.1%)	(4.3%)	(0.9%)	2.4%
$n = 200$	0.824	1.032	1.158	1.237	1.357	1.461	1.612	
	(0.2%)	(1.3%)	(1.7%)	(1.1%)	(0.1%)	(1.3%)	(1.0%)	0.9%

Table 1. NHPP models: upper quantiles of $\widetilde{K}_{m,n}$ distribution. Absolute relative errors between upper quantiles of $\mathcal{L}_{ks}$ and those of the sample distributions of $\widetilde{K}_{m,n}$ are given between brackets. MARE for each line is given in the last column.

Signif. level δ	0.50	0.25	0.15	0.10	0.05	0.025	0.01	MARE
$\mathcal{L}_{ks}$ quantiles	0.822	1.019	1.138	1.224	1.358	1.480	1.628	
		Hawkes-Pow. $(\mu, k, c, p) = (0.1, 0.2, 0.3, 1.5)$						
$n = 50$	0.807	1.009	1.146	1.293	1.359	1.421	1.557	
	(1.9%)	(0.9%)	(0.7%)	(5.6%)	(0.1%)	(4.0%)	(4.3%)	2.0%
$n = 100$	0.816	1.097	1.194	1.268	1.355	1.403	1.499	
	(0.7%)	(7.7%)	(4.9%)	(3.6%)	(0.2%)	(5.2%)	(7.9%)	4.3 %
		Hawkes-Exp. $(\mu, \alpha, \beta) = (0.3, 1.2, 1.5)$						
$n = 50$	0.837	1.045	1.145	1.218	1.375	1.539	1.566	
	(1.8%)	(2.6%)	(0.6%)	(0.5%)	(1.3%)	(4.0%)	(3.8%)	2.0%
$n = 100$	0.810	1.031	1.162	1.243	1.421	1.515	1.583	
	(1.5%)	(1.2%)	(2.1%)	(1.5%)	(4.7%)	(2.4%)	(2.8%)	2.3 %

Table 2. Hawkes processes: upper quantiles of $\widetilde{K}_{m,n}$ distribution. Absolute relative errors between upper quantiles of $\mathcal{L}_{ks}$ and those of the sample distributions of $\widetilde{K}_{m,n}$ are given between brackets. MARE for each line is given in the last column.

	NHPP-PLP		NHPP-LLP		Hawkes-Pow.		Hawkes-Exp.	
	Time-rescaling approach							
	TR1	TR2	TR1	TR2	TR1	TR2	TR1	TR2
Sample size								
$n = 50$	18.3%	3.3%	19.2%	3.7%	36.8%	2.0%	29.2%	2.0%
$n = 100$	17.5%	2.3%	18.8%	2.4%	39.2%	4.3%	27.7%	2.3%
$n = 200$	18.1%	1.2%	15.9%	0.9%	40.3%		27.5%	

Table 3. MAREs assessing differences between upper quantiles of $\mathcal{L}_{ks}$ and those of the sample distributions of $\widehat{K}_n$ (for TR1) and $\widetilde{K}_{m,n}$ (for TR2). 1000 replications were simulated for each case. TR1: standard time-rescaling with estimated parameters. TR2: predictive time-rescaling [$m = n/2$].

	NHPP-PLP			NHPP-LLP			Hawkes-Pow.			Hawkes-Exp.		
	Time-rescaling approach											
	TR0	TR1	TR2	TR0	TR1	TR2	TR0	TR1	TR2	TR0	TR1	TR2
Sig.lev												
δ=1%	1.1%	0.1%	1.1%	0.9%	0.0%	1.2%	1.1%	0%	1.3%	0.5%	0.0%	2.6%
δ=5%	5.3%	0.7%	5.5%	4.8%	0.6%	6.2%	4.7%	0%	6.1%	4.3%	0.2%	8.6%
δ=10%	10.2%	1.9%	10.4%	10.5%	2.0%	11.3%	9.6%	0%	11.2%	9.0%	0.2%	13.8%

Table 4. Sizes of KS Gof-tests: percentages of H_0 rejection on data simulated from the tested model. Sample size $n = 50$ and 10000 replications. TR0: time-rescaling with the true parameters. TR1: time-rescaling with estimated parameters. TR2: predictive time-rescaling [$m = n/2$].

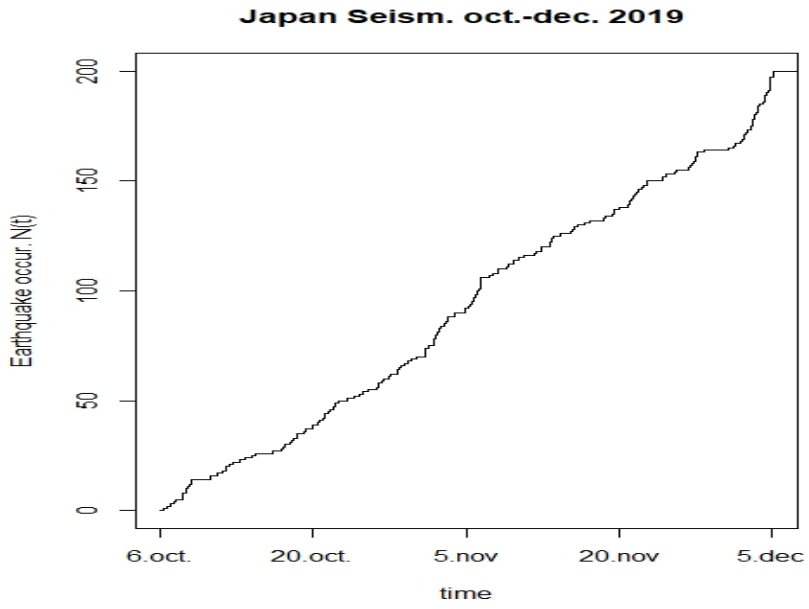

Figure 3. Event times of 200 Japanese seismic episodes between October 6 and December 5, 2019.

triggering of aftershocks by primary seismic events. This dataset is a subset of the broader catalog analysed in [7] (Chap. 11), where Q-Q plot diagnostics indicated that the Hawkes-Exponential model achieves a superior fit compared to both NHPP and Hawkes-Power law formulations.

Figure 4 and Table 5 present the p-values of the predictive KS Gof assessment for the selected seismic sequence [`TR2` with $m = n/2$]. The predictive test results indicate a definitive rejection of NHPP models from the 115th event onward. Analysis of the last events suggests that while both Hawkes specifications outperform the NHPP, the Hawkes-Exponential model shows signs of diminishing fit. In contrast, the Hawkes-Power law model exhibits increasing p-values during the final stages of the sequence, suggesting it is the most robust candidate for modelling the dynamic evolution of this seismic activity.

Earthquake occurrence i	190	191	192	193	194	195	196	197	198	199
	P-values									
NHPP-PLP.	2.10e-03	2.6e-03	3.2e-03	3.8e-03	4.5e-03	5.2e-03	6.0e-03	4.4e-03	5.4e-03	6.5e-03
NHPP-LLP.	5.7e-04	6.8e-04	8.2e-04	9.7e-04	1.1e-03	1.3e-03	1.6e-03	1.8e-03	2.1e-03	2.5e-03
Hawkes-Pow.	0.925	0.892	0.962	0.991	0.969	0.929	0.870	0.928	0.869	0.796
Hawkes-Exp.	0.323	0.389	0.318	0.257	0.206	0.163	0.128	0.161	0.126	0.098

Table 5. P-values of the predictive Gof-test for the 10 last Japanese earthquake occurrences during the studied period [October 6 - December 5, 2019].

5. Conclusion

This research evaluates the efficacy of time-rescaling for goodness-of-fit (Gof) testing in self-exciting temporal point processes under single-trajectory observation. We

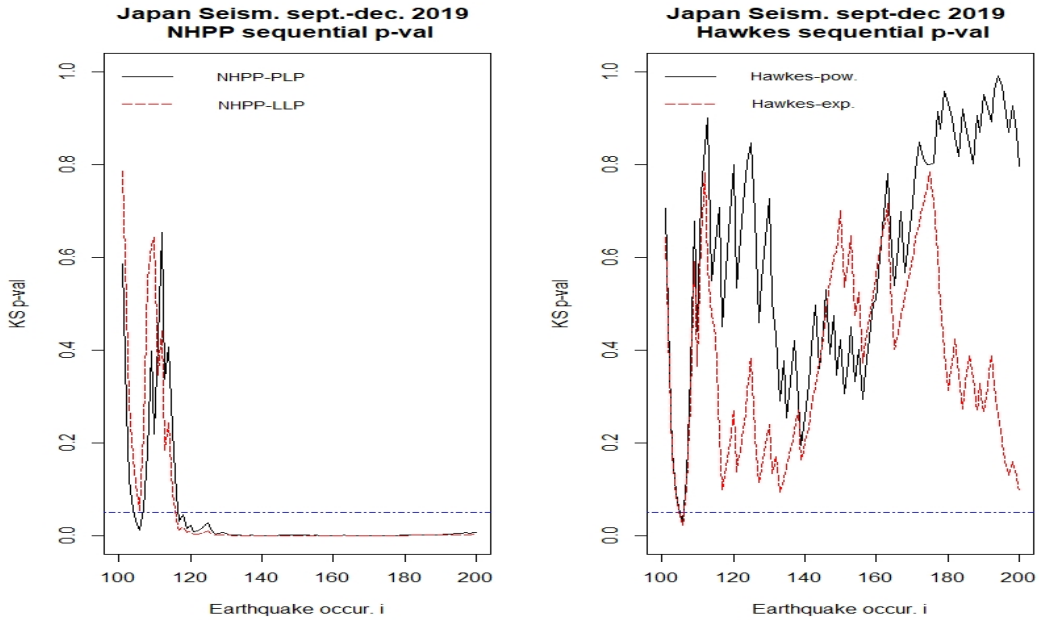

Figure 4. P-values of the predictive KS Gof-test on the last 200 Japanese seismic episodes of 2019 [`TR2` with $m = n/2$].

first quantified the systematic bias inherent in the standard plug-in approach, finding discrepancies of approximately 18% for NHPP and 30% for Hawkes models between empirical and theoretical Kolmogorov-Smirnov (KS) quantiles. These results confirm that standard time-rescaling is frequently unreliable for model validation.

To resolve this, we introduced a predictive time-rescaling framework based on a sequential inferential procedure. We proved that the resulting transformed inter-arrival times converge to i.i.d. Exp(1) variables, effectively eliminating the systematic bias. Numerical simulations demonstrate that this predictive-sequential approach significantly outperforms the standard method; for instance, it achieves rejection rates of up to 91% for misspecified models where the standard test was powerless.

Despite these promising results, several avenues for future research remain:

- Model Separability: we aim to theoretically determine which SETPP classes can be distinguished within a single-trajectory framework, potentially leveraging the *amount of information* concept.
- Optimization of m: further work is required to determine the optimal training window m. This involves a bias-variance trade-off between parameter estimation accuracy and the convergence quality of the rescaled times.
- Theoretical Expansion: We intend to extend Proposition 3.3 to prove that the predictive-sequential empirical process converges to a Brownian bridge, providing a more rigorous foundation for a wider array of Gof tests for SETPP models.

Acknowledgements

The author wants to thank Dr. Patrick Laub for data and information sharing, Dr. Kyungsub Lee for fruitful exchanges about his R package `emhawkes`[1] and Pr. Olivier Gaudoin for his continuous support.

[1] https://cran.r-project.org/web/packages/emhawkes/index.html

References

[1] Andersen, P., Borgan, O., Gill, R., and Keiding, N. *Statistical Models Based on Counting Processes*. Springer, 1993.

[2] Bhattacharjee, M., Deshpande, J., and Naik-Nimbalkar, U. Unconditional tests of goodness of fit for the intensity of time-truncated nonhomogeneous poisson processes. *Technometrics 46*, 3 (2004), 330–338.

[3] Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., and Frank, L. M. The Time-Rescaling Theorem and Its Application to Neural Spike Train Data Analysis. *Neural Computation 14*, 2 (2002), 325–346.

[4] Cook, R., and Lawless, J. *The Statistical Analysis of Recurrent Events*. Springer, 2007.

[5] Dawid, A. Statistical theory: The prequential approach. *J R Stat Soc Ser A Stat Soc 147* (1984), 278–292.

[6] El-Aroui, M.-A. On the use and misuse of time-rescaling to assess the goodness-of-fit of self-exciting temporal point processes. *Journal of Applied Statistics 52*, 12 (2025), 2247–2270.

[7] Laub, P., Lee, Y., and Taimre, T. *The Elements of Hawkes Processes*. Springer, 2021.

[8] Lindqvist, B., and Rannestad, B. Monte carlo exact goodness-of-fit tests for nonhomogeneous poisson processes. *Appl. Stochastic Models Bus. Ind. 27* (2011), 329–341.

[9] Papangelou, F. Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc. 165* (1972), 483–506.

[10] Rathbun, S. Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *Journal of Statistical Planning and Inference 51* (1996), 55–74.

[11] Reynaud-Bouret, P., Rivoirard, V., Grammont, F., and Tuleau-Malot, C. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *J. Math. Neurosc. 4*, 3 (2014), 1–41.

[12] Svensson, A. Asymptotic estimation in counting processes with parametric intensities based on one realization. *Scandinavian Journal of Statistics 17*, 1 (1990), 23–33.

[13] Tao, L., Weber, K., Arai, K., and Eden, U. A common goodness-of-fit framework for neural population models using marked point process time-rescaling. *Journal of Computational Neuroscience 45* (2018), 147–162.

[14] Wei, S., Zhu, S., Zhang, M., and Xie, Y. Goodness-of-fit test for mismatched self-exciting processes. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)* (San Diego, California, USA, 2021), vol. 130, PMLR, pp. 1243–1251.

[15] Yang, J., Rao, V., and Neville, J. A stein–papangelou goodness-of-fit test for point processes. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* (2019), K. Chaudhuri and M. Sugiyama, Eds., vol. 89 of *Proceedings of Machine Learning Research*, PMLR, pp. 226–235.