An Inhomogeneous Weibull–Hawkes Process to Model Underdispersed Acoustic Cues

Alec B. M. Van Helsdingen⬤, Tiago A. Marques⬤, and Charlotte M. Jones- Todd⬤

A Hawkes point process describes self-exciting behaviour where event arrivals are triggered by historic events. These models are increasingly becoming a popular choice in analysing event-type data. Like all other inhomogeneous Poisson point processes, the waiting time between events in a Hawkes process is derived from an exponential distribution with mean one. However, as with many ecological and environmental data, this is an unrealistic assumption. We, therefore, extend and generalise the Hawkes process to account for potential under- or overdispersion in the waiting times between events by assuming the Weibull distribution as the foundation of the waiting times. We apply this model to the acoustic cue production times of sperm whales and show that our Weibull–Hawkes model better captures the inherent underdispersion in the interarrival times of echolocation clicks emitted by these whales.

Key Words: Overdispersion; Point patterns; Self-excitement; Underdispersion; Sperm whale.

1. INTRODUCTION

A temporal point pattern comprises a series of observed event times over some time period (Daley and Vere-Jones 2003). Such events often occur in clusters, and through using point process methodology, we are able to infer the inherent pattern in their sequences (Diggle 2013). A temporal Hawkes process (Hawkes 1971a,b) is a type of self-exciting model where the occurrence of one event triggers events in the near future. This means that each event immediately increases the rate at which future events occur; the influence of the event diminishes over time.

The self-exciting behaviour of the Hawkes process makes it a particularly useful model for phenomena where events tend to cluster in time. Applications are wide-ranging and to date

A. B. M. Helsdingen (✉) · C. M. Jones-Todd, Department of Statistics, University of Auckland, 38 Princes St, Auckland 1010, New Zealand, (E-mail: *alec.van.helsdingen@auckland.ac.nz*). T. A. Marques, Centre for Research into Ecological and Environmental Modelling, University of St Andrews, Buchanan Gardens, KY16 9LZ St Andrews, Scotland, UK. T. A. Marques, Departamento de Biologia Animal Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.

include seismology (Ogata 1988), finance (Bacry et al. 2015; Hawkes 2018), criminology (Zhuang and Mateu 2019; Park et al. 2021), neuroscience (Reynaud-Bouret et al. 2013) and social media (Zhao et al. 2015). The common theme amongst these applications is that the occurrence of an event (e.g. an earthquake, a large financial trade, or a Tweet) is known to trigger future events (i.e. an aftershock, further transactions, retweets). By using a Hawkes model, the potential causal relationship between events may be inferred. This is because the self-exciting component captures the heightened risk or rate of occurrences immediately following an event.

The phenomenon of event occurrences inducing others is also prevalent throughout environmental, biological and agricultural sciences. Modelling these data as a Hawkes process aids in detecting the clustering and the potentially triggering nature of these events. Accordingly, the Hawkes process has recently begun to see use in these fields, including modelling the spread of invasive species (Balderama et al. 2012; Gupta et al. 2018), forest fires (Tonini et al. 2017; Holbrook et al. 2022), and fisheries stocks (Nakagawa et al. 2019).

The events of a classic Hawkes process are distributed according to an inhomogeneous Poisson process according to a rate that consists of some baseline intensity and a term that accounts for the self-excitement (Hawkes 1971a). However, many natural and ecological events do not occur according to the typically assumed inhomogeneous Poisson process, but are either underdispersed (less variance in the times between events) or overdispersed. For example, the spatial locations of termite mounds (Pringle et al. 2010), pine trees (Kenkel 1988) and pollinated plants (Herrera 2021) are all more regularly spread out than a Poisson point pattern, and the counts of offspring (Brooks et al. 2019) often have less variance than a Poisson distribution. Conversely, overdispersion is common in much ecological count data such as counts of migrating birds (Lehikoinen et al. 2010) and plant species richness (Kleijn et al. 2008).

For an inhomogeneous temporal Poisson process, the compensator at a given point in time point is the expected number of events that have occurred since time zero, see Sect. 2.1. The differences in compensator values between consecutive events are exponentially distributed. Previous work by Berman (1981) allowed the consecutive compensator differences of a temporal point process to instead follow a gamma distribution. There is also a long tradition of using the Weibull distribution to model inter-event times in renewal processes (Lomnicki 1966; Yannaros 1994; Ong et al. 2015). More generally, Lindqvist et al. (2003) developed their trend renewal process, which allows an inhomogeneous temporal point process to use any distribution for the compensator differences whose support is non-negative in place of the exponential. The Weibull distribution has been a common choice as the compensator difference distribution in trend renewal processes, being applied in diverse applications such as medicine (Pietzner and Wienke 2012), volcanic eruptions (Bebbington 2010), and battery reliability (Wang et al. 2019). These previous models, however, do not describe the potentially self-exciting nature of events, as the inhomogeneous "trend function" is not a Hawkes process.

The RHawkes model developed by Wheatley et al. (2016) allows the compensator differences for the baseline, i.e. non-self-excited events only, to come from any distribution with non-negative support. The Weibull–Hawkes process developed by Zhang et al. (2020) again only relaxes the assumption for the baseline events; they define the baseline rate as a

power-law function. In both cases, the remaining (self-excited) events are assumed to adhere to the Poisson assumption regarding compensator differences mentioned above.

The different assumptions about baseline and non-baseline events are computationally expensive and suppose some fundamental difference between these two classes of events. This, alongside the exponential waiting time assumption itself, is unrealistic for many situations. This is because (1) in many examples there is no way to reliably distinguish a baseline from a non-baseline event, even when applying methods such as stochastic declustering, and (2) real-world events are often either overdispersed or underdispersed compared to what a inhomogeneous Poisson assumes.

In this paper, we extend the traditional Hawkes process to account for under- or overdispersion in the waiting times between events by modelling all (both baseline and non-baseline) compensator differences using the Weibull distribution, and as an example we fit this model to the acoustic cue production times of sperm whales (*Physeter macrocephalus*). In Sect. 2 we present our Weibull–Hawkes process and demonstrate, via simulation, the model's ability to capture both under- and overdispersion. In Sect. 3 we introduce the acoustic cue data and show that the extended Hawkes model captures the inherent underdispersion in the interarrival times of echolocation clicks. In Sect. 4 we discuss how this extension leads to a more realistic and flexible self-exciting model better suited to the typical data structures seen throughout the environmental, biological, and agricultural sciences.

2. MATERIALS AND METHODS

2.1. A SELF-EXCITING TEMPORAL POINT PROCESS

A Hawkes process assumes that each event immediately increases the rate at which future events may occur, i.e. the occurrence of an event *excites* others. This self-excitement effect diminishes over time and the rate of events decays to some long-term baseline rate if no further events are observed. For current time t the conditional rate of occurrence is given by an intensity function, $\lambda(t; \cdot)$, which comprises a baseline intensity and a self-excitement term. The self-exciting term typically consists of a summation over all historic events ($\tau_i < t$ where $i = 1, ..., N$) that is weighted by some specified decay kernel. The conditional intensity of the Hawkes process proposed by Hawkes (1971a) for current time t is given by

$$\lambda(t; \gamma, \cdot) = \mu(t; \cdot) + \gamma \sum_{i:\tau_i < t} \nu(t - \tau_i). \tag{1}$$

Here, $\mu(t; \cdot) > 0$ is some temporally varying background (baseline) rate and $\nu(t - \tau_i)$ is the historic dependence kernel (which integrates to one). The parameter γ, termed the branching ratio, gives the expected number of events triggered by any single event. The baseline events are customarily called *immigrants* and the triggered events *children*, though in practice it may not be possible to distinguish between the two. Where applicable, this is the terminology we will use during the remainder of the manuscript.

The compensator of a temporal point process, $\Lambda(\tilde{t})$, evaluated at some time $\tilde{t}$ gives the expected number of events in the interval $[0, \tilde{t}]$. For any inhomogeneous Poisson process such as a Hawkes process, the compensator is defined as

$$\Lambda(\tilde{t}) = \int_0^{\tilde{t}} \lambda(t) \, dt. \tag{2}$$

The random time change theorem (Daley and Vere-Jones 2003) states that if the set of events $[\tau_1, \tau_2, ..., \tau_N]$ is a realisation of an inhomogeneous Poisson process, then $[\Lambda(\tau_1), \Lambda(\tau_2), ..., \Lambda(\tau_N)]$ is a realisation of a homogeneous Poisson process with unit rate. We denote the compensator differences as $\delta\Lambda_i = \Lambda(\tau_i) - \Lambda(\tau_{i-1})$ for $i = 2, ..., N$ and $\delta\Lambda_1 = \Lambda(\tau_1)$. Under the assumptions above $\delta\Lambda_i \sim \text{Exp}(1)$. This is an unrealistic and restrictive assumption for many real-world data, which rarely occur in an independent Poisson manner.

As was discussed in Sect. 1, the exponential compensator assumption for inhomogeneous temporal Poisson processes has previously been relaxed by using Gamma processes (Berman 1981), and more generally, any distribution with non-negative support and unit mean (Lindqvist et al. 2003). For the self-exciting case, Wheatley et al. (2016) relaxes the exponential waiting time assumption for the immigrant points only. That paper developed a renewal Hawkes model where for the immigrant points $\delta\Lambda_i \sim \text{Weibull}\,(\rho, k)$ with scale parameter ρ and shape parameter k, see Sect. 2.2. However, the children remain restricted by the assumption of exponential compensator differences. In Sect. 2.2 we extend the Hawkes model to account for under- and overdispersion in the waiting times of all events, both immigrants and children, using the Weibull distribution.

2.2. AN INHOMOGENEOUS WEIBULL–HAWKES MODEL

In Sect. 2.2.1 we discuss the Weibull distribution and how it can be used to create a family of distributions with mean one. In Sect. 2.2.2 we review the classic Hawkes process with an exponentially decaying kernel, and in Sect. 2.2.3 we present our Weibull–Hawkes model.

2.2.1. The Weibull Distribution

If $T \sim \text{Weibull}\,(\rho, k)$ (for $t \geq 0$), then

$$f(t; \rho, k) = \left(\frac{k}{\rho}\right) \left(\frac{t}{\rho}\right)^{k-1} \exp\left(-\frac{t}{\rho}\right)^k \tag{3}$$

where $\rho > 0$ is the scale parameter and $k > 0$ is a shape parameter that control the spread (stretching/shrinking) and peakedness (peak roundness) of the distribution, respectively. The mean is given by $\rho\Gamma(1 + 1/k)$ where $\Gamma(\cdot)$ is the Gamma function. The hazard (failure) rate function follows a power-law relationship with time and is given by $h(t; \rho, k) = \left(\frac{k}{\rho}\right) \left(\frac{t}{\rho}\right)^{k-1}$. Note that when $k = 1$, the hazard rate is constant (ρ^{-1}).

Setting $g(k) = \Gamma(1 + 1/k)$ and $\rho = \frac{1}{g(k)}$, the Weibull probability density function given by Eq. (3) becomes

$$f(t; k) = (k \, g(k)) \, (t \, g(k))^{k-1} \, \exp(-t \, g(k))^k. \tag{4}$$

Here, k is a dispersion parameter; when $k = 1$ the distribution reduces to an exponential, when $k > 1$ the variance is smaller than that of an exponential, and when $k < 1$ it is larger. Thus, the parameter k allows us to model both overdispersion ($k < 1$) and underdispersion ($k > 1$). By setting ρ appropriately, we have ensured that the mean is always one and thus eliminate the need for the second parameter. For $\delta \Lambda_i \sim \text{Weibull}(1/g(k), k)$, the log-likelihood, $l(g(k), k \mid \delta \boldsymbol{\Lambda})$, is given by

$$l(g(k), k \mid \delta \boldsymbol{\Lambda}) = N \left[\log(k) + k \log(g(k))\right] \\ + \sum_{i=1}^{N} (k-1) \log(\delta \Lambda_i) - (g(k) \delta \Lambda_i)^k \tag{5}$$

where N is the number of events.

2.2.2. The Hawkes Process

The intensity of a classic Hawkes process detailed in Hawkes (1971a) with an exponential decay kernel is

$$\lambda(t; \alpha, \beta, \cdot) = \mu(t; \cdot) + \alpha \sum_{i: \tau_i < t} \exp(-\beta(t - \tau_i)). \tag{6}$$

Here, α is the immediate increase in intensity following an event, and β is the exponential decay of the intensity over time. The branching ratio [γ in Eq. (1)] is α/β. The log-likelihood for the version of the Hawkes process presented in Eq. (6) is given as follows:

$$l_H(\alpha, \beta, \cdot \mid \boldsymbol{\tau}) = \sum_{i=1}^{N} \log(\lambda(\tau_i; \alpha, \beta, \cdot)) - \int_0^T \lambda(t; \alpha, \beta, \cdot) \, dt. \tag{7}$$

As noted by Ozaki (1979), this log-likelihood is not always convex and therefore a maximisation algorithm should be run from multiple starting points to ensure that the true maximum has been found.

2.2.3. Proposed Weibull–Hawkes Model

We consider two extension to the Hawkes process, both relaxing the restrictive assumption of an exponential compensator difference distribution. The extensions we propose both make use of the Weibull distribution and allow us to capture/model both under- and overdispersion in waiting times to varying degrees of flexibility. The first model we propose lets $\delta \Lambda_i \sim \text{Weibull}(1/g(k), k)$ as in Sect. 2.2.1; that is, the compensator differences follow the Weibull distribution given by Eq. (4), while having an event rate $\lambda(t; \cdot)$ which is the same as the Hawkes process [see Eq. (6)].

The second model we propose is an extension of the one described above where now we use a mixture of two Weibull distributions to model the compensator differences. We denote the first Weibull distribution as having dispersion parameter k_1 and mixture weight p. The second Weibull has dispersion parameter k_2 and by necessity mixture weight $1 - p$. We constrain p to be greater than a half to ensure the model is identifiable. We define the contribution of the first Weibull to the mean as $m_1 = p \, g \, (k_1) \, \rho_1$, where $g(\cdot)$ is defined as in Sect. 2.2.1. Given that the mean of the mixture must be one, we can compute the scale parameter of the first Weibull as $\rho_1 = m_1 / (p \, g(k_1))$ and that of the second Weibull as $\rho_2 = (1 - m_1) / ((1 - p) \, g(k_2))$. Thus the parameter space is defined by $0 < m_1 < 1$, $0.5 < p < 1, k_1, k_2 > 0$.

The likelihood of the first Weibull–Hawkes model is the probability of observing no events between time 0 and τ_1, τ_1 and τ_2... τ_{N-1} and τ_N multiplied by the conditional intensities at $\boldsymbol{\tau}$ (Daley and Vere-Jones 2003; Lindqvist et al. 2003). The probability that there are zero events in the interval $[\tau_{i-1}, \tau_i]$ is

$$1 - F(\Lambda(\tau_i; \alpha, \beta, \cdot) - \Lambda(\tau_{i-1}; \alpha, \beta, \cdot); 1/g(k), k)$$

where F is the c.d.f. of the Weibull distribution given in Eq. (3). Following Eq. 4 of Lindqvist et al. (2003), the conditional intensity at the time τ_i can be written as

$$\lambda_{CI}(\tau_i; \alpha, \beta, \cdot \mid H_t) = \lambda(\tau_i; \alpha, \beta, \cdot) \, h(\Lambda(\tau_i; \alpha, \beta, \cdot) - \Lambda(\tau_{i-1}; \alpha, \beta, \cdot); 1/g(k), k) \quad (8)$$

where $h(t; 1/g(k), k)$ is the hazard function of the previously mentioned Weibull distribution. Thus, we can write

$$L_{WH}(g(k), k, \alpha, \beta, \cdot \mid \boldsymbol{\tau}) = \prod_{i=1}^{N} [1 - F(\Lambda(\tau_i; \alpha, \beta, \cdot) - \Lambda(\tau_{i-1}; \alpha, \beta, \cdot); 1/g(k), k)]$$
$$\prod_{i=1}^{N} \lambda(\tau_i; \alpha, \beta, \cdot) \, h[\Lambda(\tau_i; \alpha, \beta, \cdot) - \Lambda(\tau_{i-1}; \alpha, \beta, \cdot); 1/g(k), k] \quad (9)$$

and by noting the definition of hazard $h(t) = f(t)/(1 - F(t))$, this simplifies to

$$L_{WH}(g(k), k, \alpha, \beta, \cdot \mid \boldsymbol{\tau}) = \prod_{i=1}^{N} \lambda(\tau_i; \alpha, \beta, \cdot) f(\Lambda(\tau_i; \alpha, \beta, \cdot) \quad (10)$$
$$-\Lambda(\tau_{i-1}; \alpha, \beta, \cdot); 1/g(k), k)$$

and so we can write the log-likelihood as

$$l_{WH}(g(k), k, \alpha, \beta, \cdot \mid \boldsymbol{\tau}, \boldsymbol{\delta\Lambda}) = l(g(k), k \mid \boldsymbol{\delta\Lambda}) + \sum_{i=1}^{N} \log(\lambda(\tau_i; \alpha, \beta, \cdot)) \quad (11)$$

where $l(g(k), k \mid \boldsymbol{\delta\Lambda})$ is defined by Eq. (5) and $\lambda(t; \alpha, \beta, \cdot)$ by Eq. (6).

Alternatively, if one were to use the conditional intensity given by Eq. (8) and substitute this into the likelihood given by Eq. (7), one obtains Eq. (5) in Lindqvist et al. (2003). This equation can be rearranged to give Eq. (6) of the same article, whose first term is equivalent to Eq. (10). We do not need the second term because we have assumed throughout that the process stops at the time of the last event τ_N, which means the second term is always one.

The derivation of the likelihood for our second, mixture model follows the same line of reasoning, except that the distribution referred to by its c.d.f $F(\cdot)$, p.d.f. $f(\cdot)$ and hazard $h(\cdot)$ is now the mixture model parameterised above. Thus, in the final likelihood given by Eq. 11, we substitute $l_M(g(k_1), g(k_2), k_1, k_2, p, m_1)$ for $l(g(k), k \mid \delta\Lambda)$. Using the definition of $f(t; p, k)$ in Eq. (3) this can be written as

$$
\begin{aligned}
l_M&(g(k_1), g(k_2), k_1, k_2, p, m_1, \alpha, \beta, \cdot \mid \delta\Lambda) \\
&= \sum_{i=1}^{N} \log[p \ f(\delta\Lambda_i, m_1/(p \ g(k_1)), k_1) \\
&\quad + (1-p) \ f(\delta\Lambda_i, (1-m_1)/((1-p)g(k_2)), k_2)]
\end{aligned} \tag{12}
$$

Our model is self-exciting and accounts for either under- or overdispersion in waiting times. The self-exciting component captures the heightened instantaneous average rate of occurrences immediately following an event, and the Weibull interarrival times give the model a huge degree of flexibility over a standard Hawkes process.

Figure 1 shows three realisations of this Weibull–Hawkes process, all with $\alpha = 0.5$, $\beta = 1$, and $\mu = 1$. The pattern shown in panel A is simulated with $k = 1$ (i.e. with the typically exponentially distributed waiting times of a Hawkes process); panel B shows the start of simulated point processes with $k = 5$ (underdispersed) and $k = 0.5$ (overdispersed) while panels C and D show the histograms of compensator differences of the point patterns depicted on panel B. An exponential is a poor fit for both, highlighting the flexibility of our model.

2.3. SIMULATION STUDY

To assess the model performance, we carried out two simulation studies. The two simulation studies use a common baseline intensity $\mu(t) = \mu + B \ s(t) - C \ \sin(\frac{t}{P})$. We defined $s(t)$ as 1 when $P\pi < t < 2P\pi, 3P\pi < t < 4P\pi, 5P\pi < t < 6P\pi...$ and 0 at all other times. We included a constraint $0 < C < \mu$. Thus, the baseline rate consists of a constant μ, a sinusoidal component $C \ \sin(\frac{t}{P})$ and a square wave $B \ s(t)$. The peaks and troughs of the sinusoidal curve and square wave align. For our simulation study we set $\mu = 0.5$, $B = 1, C = 0.25$, and $P = 10$. As will be seen in Sect. 3.2, this choice of baseline function has similarities with the function we use in the applied example. The self-excitement had parameter values $\gamma = 0.25$ and 0.75 and $\beta = 0.01$ and 0.1.

For the first study, we simulated from the first proposed Weibull–Hawkes model with values of $k = 0.5, 2/3, 1.5$ and 2 to test under- ($k > 1$) and overdispersion ($k < 1$) relative to the exponential assumption ($k = 1$). The second simulation study tested the proposed
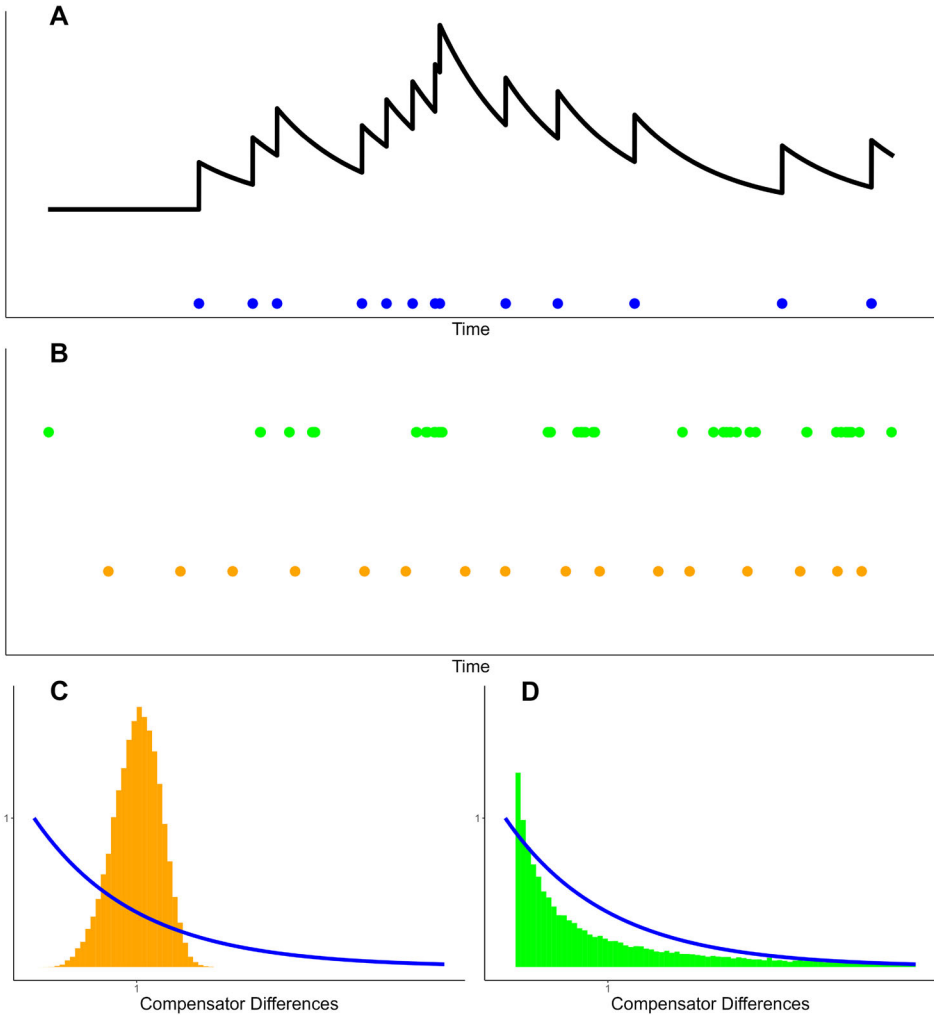
Figure 1. Top: portion of a sequence of times (blue) from a Hawkes process with $\lambda(t) = 1 + 0.5 \sum_{i:\tau_i < t} \exp(-(t - \tau_i))$ and the corresponding intensity (black line). Middle: sequence of times from our Weibull–Hawkes process (see Sect. 2.2.3) with the same formula for $\lambda(t)$, but with dispersion parameters $k = 5$ (orange) and $k = 0.5$ (green). Bottom: simulated distributions of the compensator differences of both processes. The exponential distribution (blue) is clearly inconsistent with both. See Appendix C for the simulation method (Color figure online).

mixture Weibull–Hawkes model. The same values for the parameters of $\mu(t)$, α, β and T were used. We set $k_1 = 1.5, 2$ and 3, $k_2 = 0.5$ and 0.75, $p = 0.65$ and 0.85. In each simulation, $p = m_1$.

For every set of parameters, we simulated four different time periods: $T = 400, 1000,$ 2500 and 6250. For the first study, there are 64 combinations of parameters and time periods and for each set we generated 250 realisations of our Weibull–Hawkes process for a total of 16,000 simulations. For the second study, we simulated 100 realisations for each of the 192 parameter and time period sets for a total of 19,200 simulations.

Details of the simulation algorithm can be found in Appendix C. Model fitting was carried out using the R package TMB (Kristensen et al. 2016), which allows a user to write

Table 1. Median biases (%) for each parameter across the four time periods used in the two simulation studies

Model	Weibull				Weibull mixture			
Parameter	400	1000	2500	6250	400	1000	2500	6250
μ	5.91	1.82	-3.63	-6.22	14.3	20.0	25.8	28.6
B	-3.03	-3.39	-3.55	-3.37	-12.0	-10.0	-9.9	-10.0
C	7.07	11.4	13.2	14.2	14.6	15.9	13.0	12.3
$\frac{\alpha}{\beta}$	1.25	1.71	1.95	3.18	-3.64	-1.07	-0.735	-1.15
β	1.04	5.63	4.91	3.31	73.1	53.5	52.1	53.3
k_1	0.512	0.182	0.0392	0.0241	-2.69	-1.77	-1.55	-1.16
k_2	–	–	–	–	2.38	0.605	-0.141	-0.0683
p	–	–	–	–	3.42	2.89	2.57	2.48
m_1	–	–	–	–	3.06	2.83	2.98	2.80

The second row gives the study duration T

a log-likelihood function in C++. Due to the non-convex nature of the likelihood (Ozaki 1979), we maximised the log-likelihood given by Eq. (11) from 20 random starting points. We used the New Zealand eScience Infrastructure (NeSI) server (https://www.nesi.org.nz/) to perform the simulations and fitting, which required 482 h of computing time. The results of this study are given in Sect. 2.4 (Table 1).

2.4. SIMULATION STUDY RESULTS

Our simulation studies showed that the bias in the estimates of the parameters of the compensator difference distribution: k_1 (or k), k_2, p and m_1 are low and decrease with sample size. The parameters for the baseline rate and β show greater bias, especially in the second mixture model simulation study. The model has some difficulty determining how much of the baseline intensity is comprised of the constant μ and how much is time-varying (B and C). Some biases do not decrease with sample size; perhaps the period of the pulse wave (B s(t)) and sinusoidal wave $\left(C \sin\left(\frac{t}{P}\right)\right)$ were too short for these waves to be distinguishable. However, on the whole, biases are reasonable and the results give us confidence in the simulation methodology, the likelihood that we derived in Sect. 2.2.3 and that the model can be fitted in a reasonable amount of time. The complete results of our two simulation studies can be found in the GitHub repository https://github.com/ABMvanHelsdingen/WHP. Selected results of the simulation study are plotted in Appendix D, and complete results of our two simulation studies can be found in the GitHub repository https://github.com/ABMvanHelsdingen/WHP.

3. APPLIED EXAMPLE

3.1. ACOUSTIC CUE RATE DATA

Acoustic cues are emitted by cetaceans for a variety of reasons, including hunting by echolocation (Johnson et al. 2004), communication (Deecke et al. 2005), and mating (Smith et al. 2008). Acoustic cues of cetaceans can be passively recorded by either hydrophones

Table 2. Summary statistics for the whale cues, depths, and dive cycles

Quantity	Value
Number of clicks	50,031
Tagging time (s)	39,394
Mean interclick interval (s)	0.787
Maximum interclick interval (s)	1010
Mean depth (m)	457
Maximum depth (m)	880
Deep dives	13

(Zimmer et al. 2008) or by tagging individual whales (Madsen et al. 2002). One example of the latter is a digital acoustic recording tag (DTAG), which are motion and acoustic recording tags that are attached to cetaceans via suction cups (Johnson and Tylack 2003). The tag records sound and also has sensors that measure the dive depth and the animal's orientation. The sound data are then processed (Shamir et al. 2014), so that sounds emitted by the animal can be distinguished from background noise. The acoustic properties of the cues themselves such as frequency harmonics and power spectra can be used to infer specific behaviours (Mohl et al. 2003; Au et al. 2006) of cetaceans. In addition, the rate of acoustic cues can be used to infer the impact of anthropogenic sounds on behaviour (Tyack et al. 2011; Hawkins and Popper 2016).

A resourceful way of using these data is to estimate animal abundance. To do this, we need to estimate the average cue production rate (Marques et al. 2013) and cues are often considered as having a constant long-term average rate that averages out any other factors such as depth. In previous studies analysing the relationships between cue rates and covariates, the cue data have been binned or aggregated across dive cycles (Stimpert et al. 2015; Warren et al. 2017). This aggregation leads to a huge loss of information. Therefore, we propose treating the acoustic cue times as a temporal point process directly, considering the timestamps of cues as a realisation of some point process. In addition, we propose the use of a self-exciting model to better capture the inherent clustering and potentially contagious nature of echolocation cues.

Figure 2 shows the temporal point pattern of acoustic cues from a single sperm whale alongside its recorded depth (m). A summary of the data over the entire time period considered (approximately 11 h) is given in Table 2. These data were recorded using DTAGs and are part of the ACCURATE project (https://accurate.st-andrews.ac.uk/), which has collated data from over 100 sperm whales. To illustrate our proposed methods, we consider an individual, tag code sw03_253b, tagged in the Mediterranean in 2003. Figure 4 in Appendix A shows the distribution of the interclick intervals.

The number of acoustic cues emitted by sperm whales is known to increase with dive depth (Stimpert et al. 2015; Warren et al. 2017). Furthermore, there is evidence to suggest that cue rate changes alongside the whale's rate and direction of descent (Watwood et al. 2006). We found that acoustic cues were more frequent during descents and when the whale was at the bottom of each dive, see Fig. 6 in Appendix B. Acoustic cues occurred at a lower
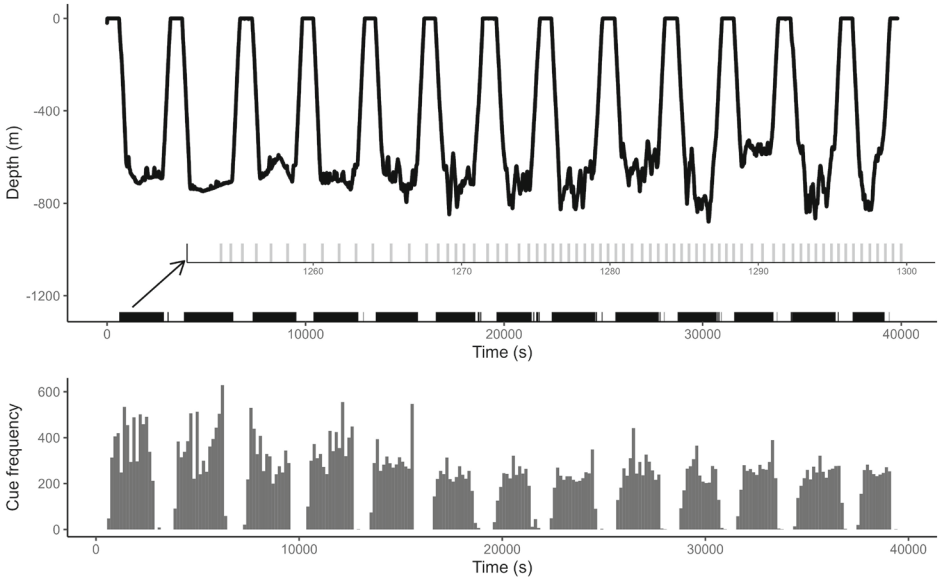
Figure 2. Top: depth profile of the whale. Middle: rug plot showing cue times with inset showing individual cue times in a 50 s interval. Bottom: histogram of cue times (each bin is about 160 s long). Note that barely any cues are emitted when the whale is on the surface .

frequency during ascents and virtually never when the whale was at the surface, as the lower panel of Fig. 2 clearly shows. Cues are clearly clustered, see Fig. 6; this is most obvious during the ascent where there were several bursts followed by longer periods of silence.

3.2. MODELLING ACOUSTIC CUE RATES

Letting $d(t)$ be the dive depth in kilometres at time t (i.e. $d(t) > 0$ when the whale is underwater, available at a frequency of 1 Hz), $r(t)$ denotes the rate of descent calculated using numerical differentiation (i.e. $r(t) > 0$ when descending), and $s(t) = 0$ if $d(t) > 0.02$ (i.e. > 20 metres below the surface), $s(t) = 1$; otherwise, we set $\mu(t; \cdot)$ in Eq. (6) to be

$$\mu(t; \boldsymbol{\eta}) = \exp(\eta_0 + \eta_1 s(t) + \eta_2 d(t) + \eta_3 r(t)). \tag{13}$$

Here each $\eta_j (j = 0, ..., 3)$ is a coefficient of the inhomogeneous baseline rate of cues. As in Eq. (6) the parameter α is the instantaneous increase in intensity when an event occurs and β is the decay over time of this self-exciting effect. The branching ratio of events is now α/β, and we have $\beta \geq \alpha$ so that $\alpha/\beta \leq 1$. We also set a constraint $\eta_1 > 0$ so that the baseline rate when the whale is at the surface ($s(t) = 0$) is less than when it is underwater ($s(t) = 1$). Similar to other studies, we excluded the first complete dive from the point pattern; this is to avoid possible short-term behaviour changes induced by the tagging process (Barlow et al. 2013; Hildebrand et al. 2015). Estimated parameter values for this model are given in Sect. 3.3.

Table 3. Parameter priors, posterior means and 95% credible intervals for the baseline rate parameters in Eq. (13), the self-excitement parameters from Eq. (6) and the Weibull parameters of Eq. (11) and Eq. (12)

Model		Weibull		Weibull mixture	
Parameter	Prior	Posterior mean	95% CI	Posterior mean	95% CI
η_0	U$(-30,0)$	-13.2	$(-16.2, -9.8)$	-13.01	$(-14.8, -9.63)$
η_1	U$(0,100)$	9.16	$(5.79, 12.2)$	8.74	$(5.33, 10.4)$
η_2	U$(-20,20)$	2.72	$(2.48, 2.98)$	0.0254	$(-0.125, 0.172)$
η_3	U$(-1000,1000)$	620	$(576, 663)$	148	$(114, 180)$
$\frac{\alpha}{\beta}$	U$(0,1)$	0.923	$(0.916, 0.930)$	0.990	$(0.987, 0.992)$
β	U$(10^{-20}, 10)$	1.08	$(1.06, 1.10)$	4.17	$(4.13, 4.21)$
k_1	U$(0.01,100)$	1.53	$(1.52, 1.54)$	41.9	$(41.1, 42.7)$
k_2	U$(0.01,100)$	–	–	1.43	$(1.40, 1.45)$
p	U$(0.5,1)$	–	–	0.892	$(0.888, 0.895)$
m_1	U$(0,1)$	–	–	0.880	$(0.876, 0.884)$

3.3. Modelling Results

We fitted our model to the whale cue data in both frequentist and Bayesian frameworks. For the frequentist results, see Appendix E. To perform model fitting in a Bayesian framework, we used the R package NIMBLE (de Valpine et al. 2017) to run Markov chain Monte Carlo (MCMC) chains. We fitted both the single Weibull and Weibull mixture model in NIMBLE for 50,000 MCMC iterations. We set a burn-in of 10,000 iterations and thinned the chain so that every 4th iteration was retained; thus, we finished with 10,000 samples. This required 58 min of compute time on the NeSI server.

The estimates for both models and the priors we used for Bayesian inference are shown in Table 3. The estimates for $\boldsymbol{\eta}$ are broadly concordant. At the surface, the cue rate ($\exp(\eta_0)$) is close to zero and jumps up when the whale dives below the surface. Estimates of the effect of being underwater (η_1), depth (η_2) and rate of descent (η_3) differ, but are all of the same sign, though the mixture model estimate for η_2 is indistinguishable from zero.

As is evident in Fig. 3, modelling the compensator differences as a single Weibull distribution results in a poor fit, albeit still much better than an exponential. This can be explained by there being a few influential outliers (the maximum value of $\delta\hat{\Lambda}_i$ is 46.71), which necessitate that the value of k be smaller (i.e. a higher variance) than what would be a good fit around the centre of the distribution. With $\hat{k} = 1.53$, we expect $\sim 1.8 \times 10^{-8}$ compensator differences over 10, but we observe 156 such values. When two Weibull distributions are used, the fit is much improved, with about 90% of the mixture being very sharply peaked ($k_1 = 41.9$) and the remaining weight being more dispersed so as to capture the tail. The tail is now much less extreme, with the maximum value of $\delta\hat{\Lambda}_i$ a more reasonable 16.12. However this model does make two markedly different conclusions. Firstly, the estimate of η_2 is now indistinguishable from zero, i.e. holding the whale's vertical speed constant, depth may have no direct impact on cue production rate so as long as the whale is at least 20 ms deep. The second is that $\frac{\alpha}{\beta}$ is now much closer to one. In the first model, the expected number of *descendants* of each cue, $\hat{\beta}/(\hat{\beta} - \hat{\alpha})$, is 12.0, whereas in the second model it is around 100. This insight from the second model would imply that only about 1 % of the cues
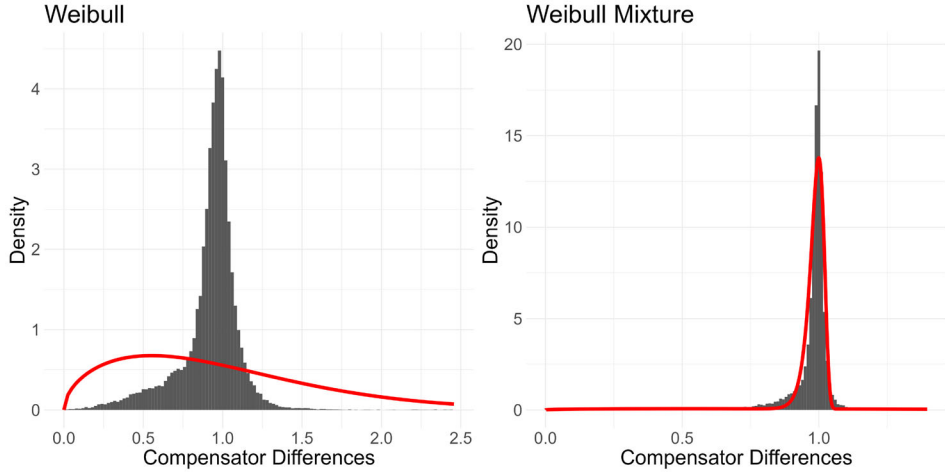
Figure 3. Compensator differences ($\delta\Lambda_i$) of the whales cues for both the Weibull and Weibull-Mixture models. The red line is the density of the fitted compensator difference distributions. For the model using a single Weibull distribution, a very long tail, with the maximum value being 46.71, necessitates a lower value of $\tilde{k}$ than what would be the best fit around the centre of the distribution (Color figure online).

are baseline, with the rest as the result of self-excitement. It is worth noting, however, that these models do not perform stochastic declustering and therefore we cannot make direct inferences about which cues are baseline and which cues triggered/caused the others.

4. DISCUSSION

Hawkes processes are typically the model of choice for self-exciting event-type data. The trend renewal process introduced by Lindqvist et al. (2003) is a temporal point process that can assume any positive distribution for the waiting times of events; however, there is generally no self-exciting component to the event arrivals. Our proposed Weibull–Hawkes models incorporate both self-excitement and under- or overdispersion in the waiting times, combining these two insights into one integrated framework that adds flexibility to the Hawkes process by applying a specific case of the trend renewal process.

Our simulation studies confirm our model is correctly formulated and that parameter recovery by MLE is feasible and works well. We simulated both under- and overdispersion, different rates of decay of the self-excitement and different branching ratios and got broadly the same results with respect to bias for all these scenarios. While some bias was seen even in the samples with $T = 6250$, it should be pointed out that the whales dataset is about an order of magnitude larger, so any issues related to small sample sizes observed in the simulations should not affect our results in Sect. 3.3 to the same degree.

Our model improves upon previous studies of whale cues by using the exact cue times, as opposed to aggregated counts. This additional information makes discerning the level of self-excitement far easier; it is possible to fit Hawkes processes to binned data but this poses various challenges (Shlomovich et al. 2022). Use of deaggregated data also means that it is possible to explore the level of dispersion in the waiting times. Understanding the

fine scale details of the sound producing mechanisms may shed light into the ways whales explore and interact with their surroundings. We are able to explore the effects of depth and rate of descent directly, as parameters in the model, rather than more informally through comparing cue rates at different values of d(t) and r(t).

Our extensions to the Hawkes process are only a first step and could be refined in many ways. Various different formulations for all three of the baseline rate, self-exciting kernel and compensator differences are imaginable. We have only considered a log-linear baseline intensity function for simplicity, but any other strictly positive function could be considered, for example, nonlinear relationships or the incorporation of other covariates. The self-exciting kernel need not be exponential, though any other function (e.g. a gamma distribution) would prove computationally costly as it would be of $O(N^2)$ time complexity rather than $O(N)$. Finally, functions other than the exponential, Weibull and mixtures of two Weibulls could be used. We have demonstrated that the Weibull distribution can be used as a part of trend renewal process for self-exciting data, but this does not preclude the use of other distributions for both this dataset and more generally.

In many cases, passive acoustic monitoring studies have been interested in a single whale population density estimate for a given time period and area. In that case, a single average cue rate that applies to that time period may be sufficient, even if it might be difficult to estimate the applicable cue rate (Marques et al. 2023). However, if one is interested in making comparisons across time or space, understanding spatio-temporal differences in cue rates is crucial, as assuming a constant cue rate when that is not the case could result in biased inferences, with existing differences in population density being masked, or with spurious differences in density being found. Our model makes progress in this direction by making the cue rate variable across time.

We have presented a model for purely temporal point processes. Thus, spatial information from spatial-temporal point patterns is discarded when using our framework. We anticipate however that our model could be extended to a spatiotemporal point process so that the waiting times between events are under- or overdispersed. This would prove a useful extension to spatiotemporal Hawkes processes (Reinhart 2018) and be especially useful in fields such as agriculture and biology where most point pattern data is spatiotemporal rather than purely temporal.

In conclusion, we have presented a new class of Hawkes processes that relax the Poisson assumptions made in all previous versions of the Hawkes process. Our model has potential to be useful in fields where Poisson assumptions are unrealistic including many natural and environmental sciences. By using our model, researchers can verify if clustering is the result of self-excitement or overdispersion. Conversely, our model can expose self-excitement that might be obscured by the more regular spacing of an underdispersed point pattern.

ACKNOWLEDGEMENTS

Data Availability All code is openly available in the GitHub repository https://github.com/ABMvanHelsdingen/WHP, which has DOI https://doi.org/10.5281/zenodo.11136815

Declarations

Conflict of interest The authors declare no conflict of interest.

A. INTERARRIVAL TIMES OF ECHOLOCATION CLICKS

Figure 4 shows the raw waiting times between cues. The distribution is clearly not consistent with an exponential distribution. A main peak at about 0.5 s can be seen, a second peak at about 0.05 s, and a small third peak at about 0.3 s. Evidently, a homogeneous Poisson process will be inadequate for this data.

Figure 5 shows the compensator differences after fitting a traditional Hawkes process. The distribution has a sharp peak around 1 and is completely inconsistent with an exponential distribution with mean 1 as assumed by the random time change theorem (Daley and Vere-Jones 2003).
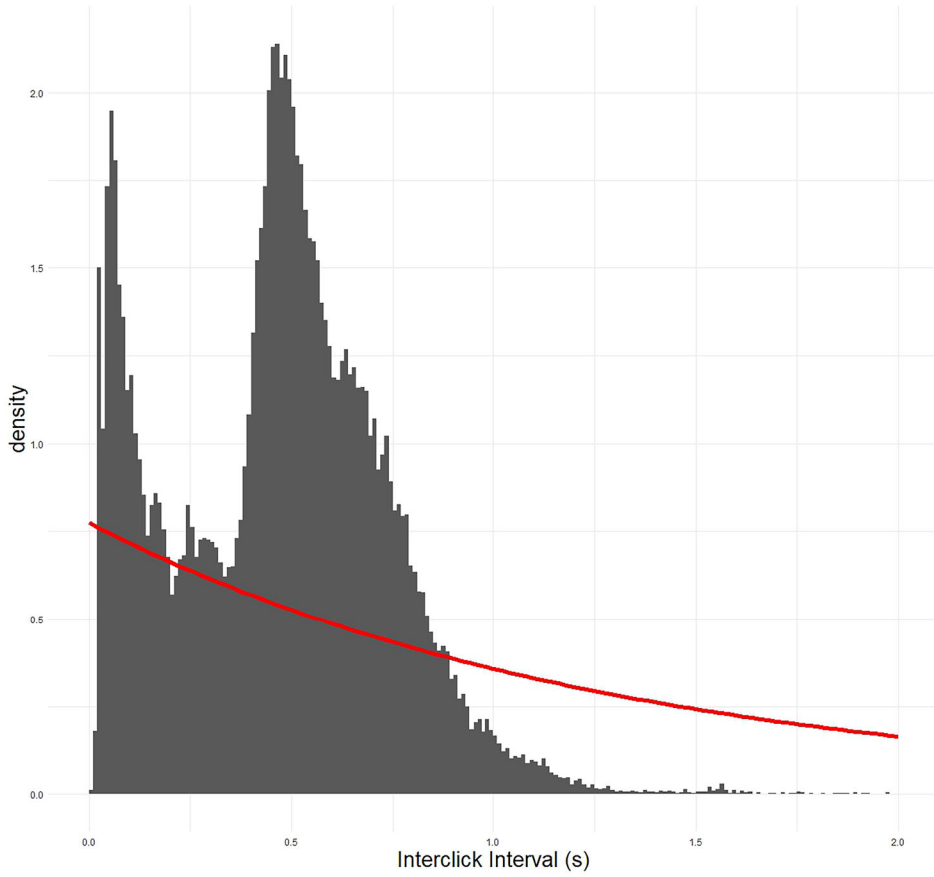
Figure 4. Interclick intervals, with an exponential distribution of the same mean overlaid (red line). The exponential has a much fatter tail than the observed intervals (Color figure online).
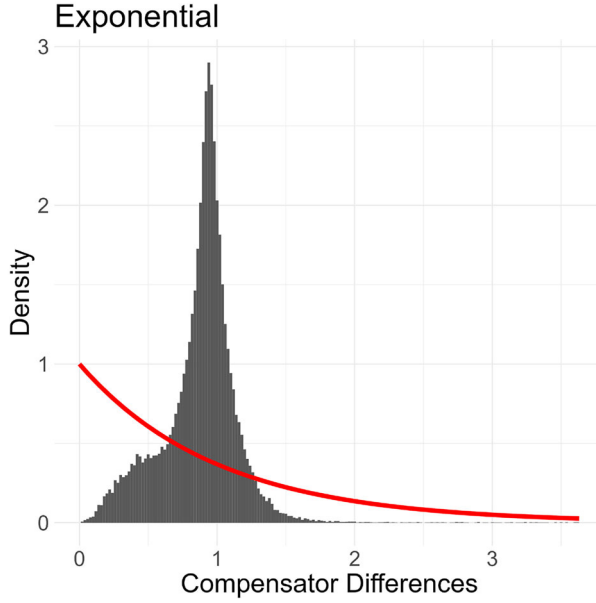
Figure 5. Compensator differences ($\delta\Lambda_i$) when a conventional Hawkes process is fitted to the data. The red line is the density of the assumed exponential distribution (Color figure online).

B. DEPTHS AND CUE TIMES

Figure 6 shows a more granular view of the whale's behaviour. It is based on Fig. 2, and shows a single, typical dive of the whale, allowing for a closer inspection of how cue rate varies across the dive cycle. Note the clustering evident in the rug plot.

C. SIMULATION METHODOLOGY

Our simulation method has two steps at each iteration i in order to generate an event time τ_i.

1. Generate a random value w_i for the compensator difference of the current interval.

2. Choose a value for τ_i so that $w_i = \delta\Lambda_i = \int_{\tau_{i-1}}^{\tau_i} \lambda(t)\, dt$

Step 1

To simulate from the Weibull distribution, we generate $u_i \sim$ Uniform $(0, 1)$. We then compute $w_i = g(k)\,(-\log(u_i))^{1/k}$ where $g(k) = \Gamma(1 + 1/k)$ as defined in Sect. 2.2.1. This can be easily adapted for a mixture of two Weibull distributions, by generating a further variable $v_i \sim$ Uniform $(0, 1)$ to determine which of the two Weibull distributions the current interval comes from.

Step 2—constant $\mu(t)$

The equation in Step 2 must be solved numerically. We denote the constant baseline rate $\mu(t)$ as μ. We start by noting that in the period between τ_{i-1} and τ_i the minimum possible value of $\lambda(t)$ is μ and the maximum value occurs immediately after τ_{i-1}, which we shall designate as $\lambda(\tau_{i-1} + \epsilon)$. For convenience, we use the notation $T = \tau_i - \tau_{i-1}$.
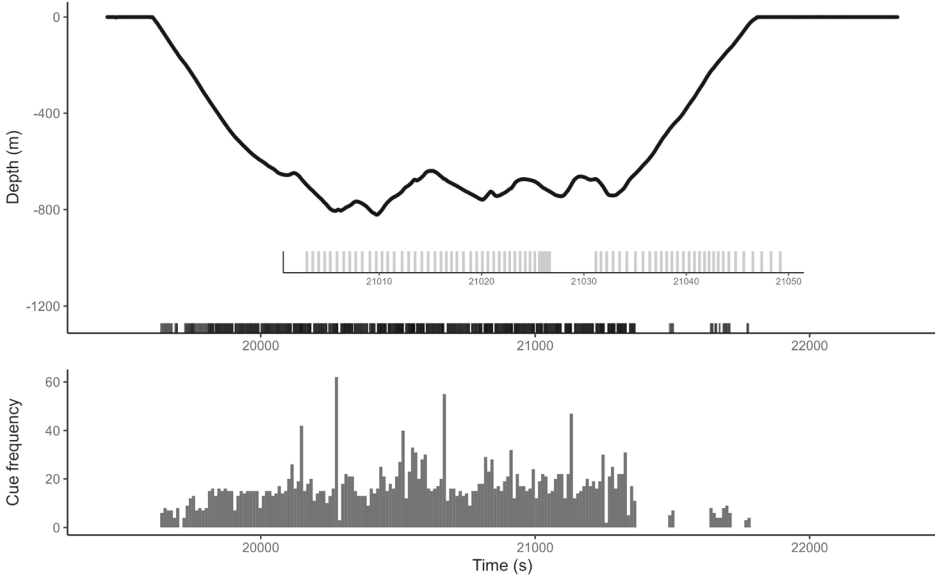
Figure 6. Top: depths of the whale against time for one of the dives middle: rug plot showing cue times with inset bottom: histogram of cue times. Note that virtually no cues are emitted when the whale is on the surface .

Using these minimum and maximum values of $\lambda(t)$, we can say that $T < \frac{w_i}{\mu}$ and $T > \frac{w_i}{\lambda(\tau_{i-1}+\epsilon)}$. Denote these lower and upper bounds for T as $T_L^{(1)}$ and $T_U^{(1)}$ respectively.

By applying the Newton–Raphson method, we can write:

$$T_L^{(k+1)} = T_L^{(k)} - \left(\int_{\tau_{i-1}}^{\tau_{i-1}+T_L^{(k)}} \lambda(t) \, dt - w_i \right) / \lambda\left(\tau_{i-1} + T_L^{(k)}\right) \tag{14}$$

This equation updates our lower bound, bringing it closer to the true value of T. A similar equation can be written for the upper bound estimates.

If we continually iterate through Eq. (14) for both the lower and upper bounds, eventually T_L^k and T_U^k get extremely close. In our experience it often takes ten or fewer iterations for the difference between the bounds to be indistinguishable from zero. The code used to run the simulation study in Sect. 2.3 runs Newton's method for 10 iterations, and if the difference exceeds a tolerance (10^{-8}) runs further batches of 10 iterations until the difference is less than the tolerance.

This works because $\lambda(t)$ is strictly decreasing between events, ensuring that the lower bound is always an underestimate and the upper bound always an overestimate. The two bounds can never cross.

At iteration 1, this method converges immediately (since $\lambda(t) = \mu$ and thus is constant), giving $\tau_1 = w_1/\mu$.

Step 2—variable $\mu(t)$

When $\mu(t)$ is not constant, we cannot use the Newton–Raphson method, as $\lambda(t)$ may increase without an event occurring. But since $\Lambda(t)$ is strictly increasing, we can use the Bisection method to numerically solve for τ_i.

As before, we need maximum and minimum values of T, and we can write $T > \frac{w_i}{\max \mu(t) + \lambda(\tau_{i-1}+\epsilon) - \mu(\tau_{i-1})}$ and $T < \frac{w_i}{\min \mu(t)}$. As before we denote these as $T_L^{(1)}$ and $T_U^{(1)}$, respectively. Note that our method requires $\mu(t)$ to be strictly positive.

At each iteration we compute the mean, which we denote as x, of $T_U^{(k)}$ and $T_L^{(k)}$, and then $\int_{\tau_{i-1}}^{x} \lambda(t)\, dt$. If this integral exceeds w_i, we know $x > T$ and we set $T_U^{(k+1)} = x$. Otherwise, $T_L^{(k+1)} = x$. Iterations continue until the difference is less than a tolerance (10^{-9}). Because the difference halves at each iteration, the number of iterations needed can be calculated in advance.

D. ADDITIONAL SIMULATION STUDY RESULTS

Figures 7, 8, 9 and 10 show, respectively, the bias in estimates of k and α/β for the first simulation study, and the bias in k_1 and in α/β for the second simulation study.
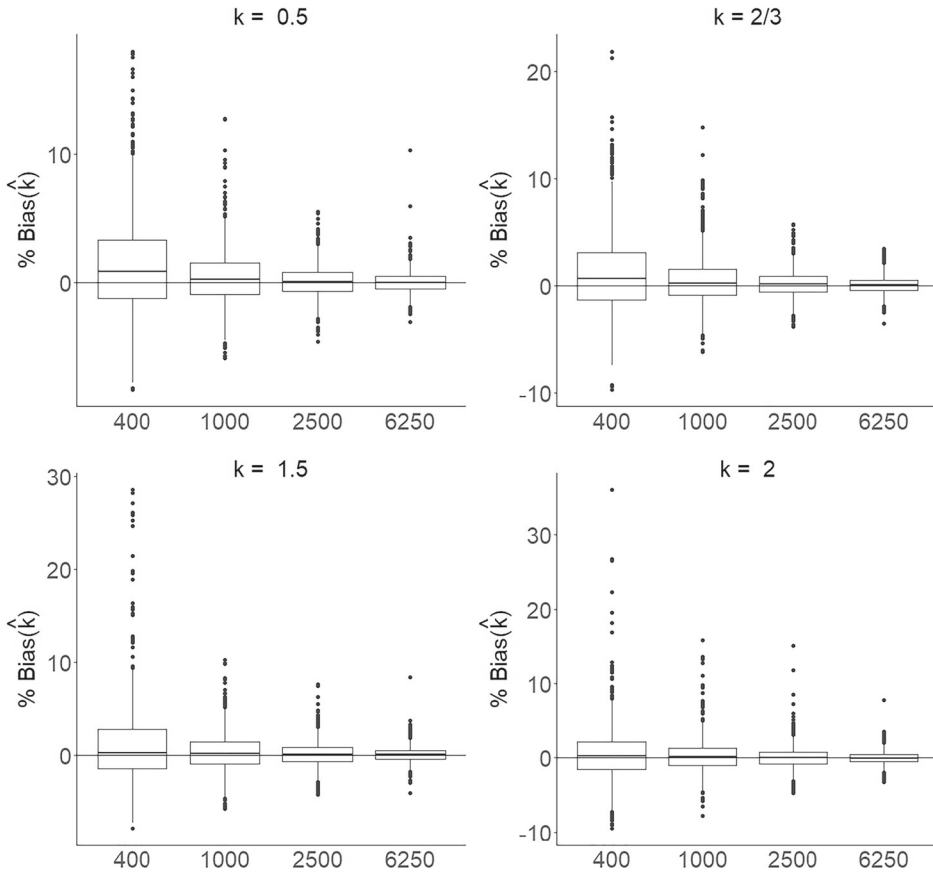

Figure 7.   Bias in maximum likelihood estimates of k (y-axes) for the four different time periods across the four values of k used in the first simulation study. The estimates appear to be asymptotically unbiased and efficient .
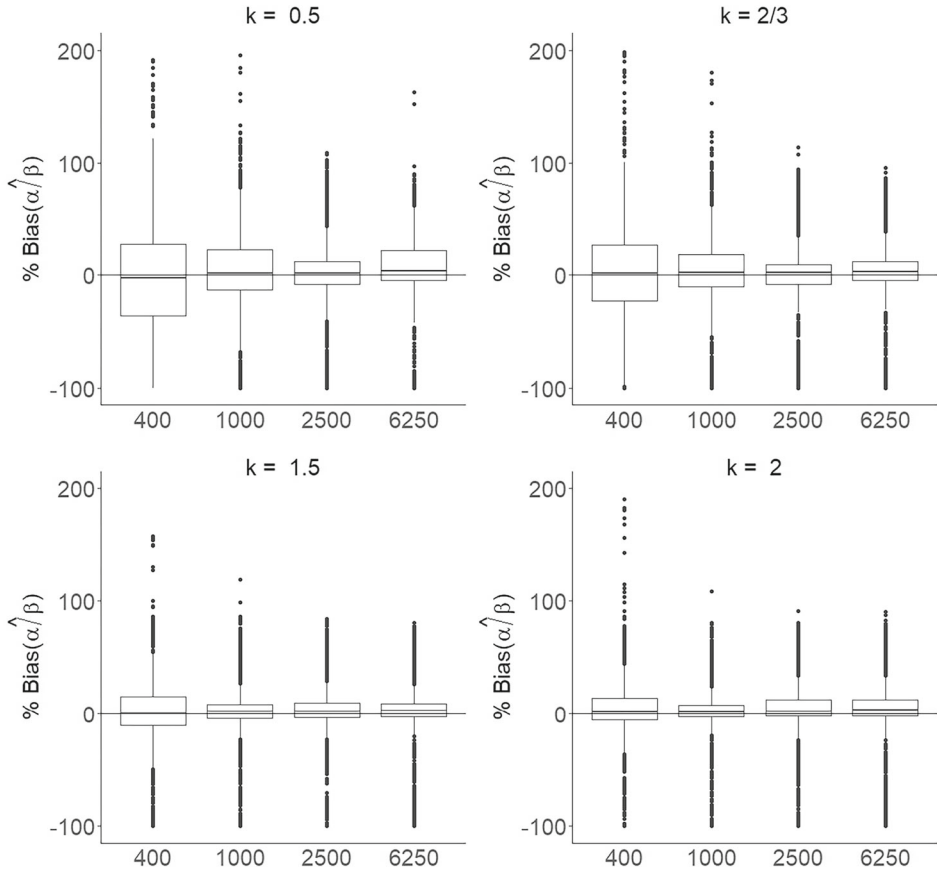
Figure 8. Bias in maximum likelihood estimates of $\frac{\alpha}{\beta}$ (y-axes) for the four different time periods across the four values of k used in the first simulation study. The y-axis is truncated at 200% .
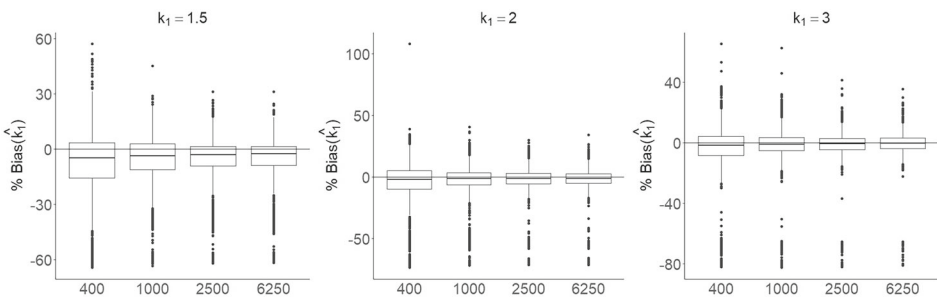

Figure 9. Bias in maximum likelihood estimates of k_1 (y-axes) for the four different time periods across the three values of k_1 used in the second simulation study. The estimates are negatively biased in smaller sample sizes, but have little bias in the largest samples $(T = 6250)$.
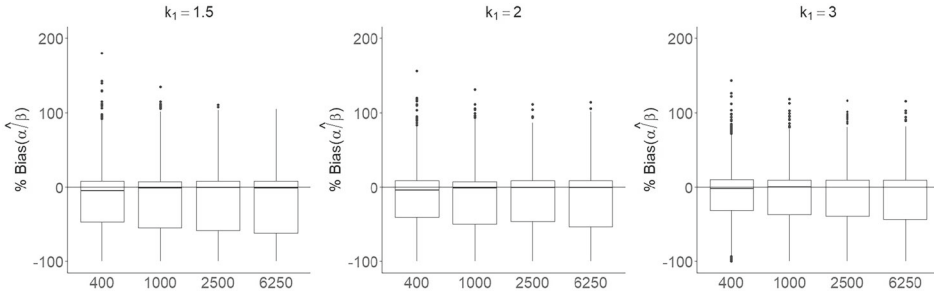
Figure 10. Bias in maximum likelihood estimates of $\frac{\alpha}{\beta}$ (y-axes) for the four different time periods across the three values of k_1 used in the second simulation study. The estimates are unbiased but have a negative skew .

E. FREQUENTIST ANALYSIS

We fitted both a conventional Hawkes process and the model with a single Weibull distribution for the compensator differences in a frequentist framework as well as a Bayesian framework. For our frequentist analysis, we maximised the log-likelihood from 500 different starting points. As with our simulation study, we used the NeSI server and required 59 min of compute time. The maximised log-likelihood was $l = -7119$ for the Weibull–Hawkes process and $l = -13042$ for a conventional Hawkes process.

Parameters for the Weibull–Hawkes model are given in Table 4. The estimates are broadly concordant with the Bayesian analysis presented in Sect. 3, especially for k, α and β. Estimates for the parameters of the baseline intensity function $\boldsymbol{\eta}$ were less consistent but similar.

Table 4. Parameter MLEs and standard errors for the baseline rate parameters in Eq. (13), the self-excitement parameters from Eq. (6) and the dispersion parameter of Eq. (11)

Parameter	MLE	SE
η_0	-9.80	0.221
η_1	4.67	NA
η_2	4.25	0.477
η_3	175	78.2
$\frac{\alpha}{\beta}$	0.930	NA
β	1.13	0.0117
k	1.56	0.00523

NA indicates that the value could not be computed

REFERENCES

Au WWL, Pack AA, Lammers MO, Herman LM, Deakos MH, Andrews K (2006) Acoustic properties of humpback whale songs. J Acoust Soc Am 120(2):1103–1110. https://doi.org/10.1121/1.2211547

Bacry E, Mastromatteo I, Muzy J-F (2015) Hawkes processes in finance. Mark Microstruct Liq 1(1):1550005. https://doi.org/10.1142/S2382626615500057

Balderama E, Schoenberg FP, Murray E, Rundel PW (2012) Application of branching models in the study of invasive species. J Am Stat Assoc 107(498):467–476. https://doi.org/10.1080/01621459.2011.641402

Barlow J, Tyack PL, Johnson MP, Baird RW, Schorr GS, Andrews RD, Aguilar de Soto N (2013) Trackline and point detection probabilities for acoustic surveys of Cuvier's and Blainville's beaked whales. J Acoust Soc Am 134(3):2486–2496. https://doi.org/10.1121/1.4816573

Bebbington MS (2010) Trends and clustering in the onsets of volcanic eruptions. J Geophys Res 115:B01203. https://doi.org/10.1029/2009JB006581

Berman M (1981) Inhomogeneous and modulated gamma processes. Biometrika 68(1):143–152. https://doi.org/10.2307/2335815

Brooks M, Kristensen K, Rosa Darrigo M, Rubim P, Uriarte M, Bruna E, Bolker BM (2019) Statistical modeling of patterns in annual reproductive rates. Ecology 100(7):e02706. https://doi.org/10.1002/ecy.2706

Daley DJ, Vere-Jones D (2003) An introduction to the theory of point processes. Springer, New York

de Valpine P, Turek D, Paciorek C, Anderson-Bergman C, Temple Lang D, Bodik R (2017) Programming with models: writing statistical algorithms for general model structures with NIMBLE. J Comput Graph Stat 26(2):403–413. https://doi.org/10.1080/10618600.2016.1172487

Deecke VB, Ford JKB, Slater PJB (2005) The vocal behaviour of mammal-eating killer whales: communicating with costly calls. Anim Behav 69(2):395–405. https://doi.org/10.1016/j.anbehav.2004.04.014

Diggle PJ (2013) Statistical analysis of spatial and spatio–temporal point patterns. Chapman and Hall/CRC, London

Gupta A, Farajtabar M, Dilkina B, Zha H (2018) Discrete Interventions in Hawkes Processes with Applications in Invasive Species Management. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence, IJCAI-18, pp 3385–3392. International Joint Conferences on Artificial Intelligence Organization

Hawkes AG (1971) Spectra of some self-exciting and mutually exciting point processes. Biometrika 58(1):83–90. https://doi.org/10.1093/biomet/58.1.83

Hawkes AG (1971) Point spectra of some mutually exciting point processes. J R Stat Soc Series B Stat Methodol 33(3):438–443

Hawkes AG (2018) Hawkes processes and their applications to finance: a review. Quant Financ 18(2):193–198. https://doi.org/10.1080/14697688.2017.1403131

Hawkins AD, Popper AN (2016) A sound approach to assessing the impact of underwater noise on marine fishes and invertebrates. ICES J Mar Sci 74(3):635–651. https://doi.org/10.1093/icesjms/fsw205

Herrera C (2021) Unclusterable, underdispersed arrangement of insect-pollinated plants in pollinator niche space. Ecology 102(6):e03327. https://doi.org/10.1002/ecy.3327

Hildebrand JA, Baumann-Pickering S, Frasier KE, Trickey JS, Merkens KP, Wiggins SM, McDonald MA, Garrison LP, Harris D, Marques TA, Thomas L (2015) Passive acoustic monitoring of beaked whale densities in the Gulf of Mexico. Sci Rep 5:16343. https://doi.org/10.1038/srep16343

Holbrook AJ, Ji X, Suchard MA (2022) Bayesian mitigation of spatial coarsening for a Hawkes model applied to gunfire, wildfire and viral contagion. Ann Appl Stat 16(1):573–595. https://doi.org/10.1214/21-AOAS1517

Johnson M, Madsen PT, Zimmer WMX, Aguilar de Soto N, Tyack PL (2004) Beaked whales echolocate on prey. Proc R Soc B 271:S383–S386. https://doi.org/10.1098/rsbl.2004.0208

Johnson MP, Tylack PL (2003) A digital acoustic recording tag for measuring the response of wild marine mammals to sound. IEEE J Ocean Eng 28(1):3–12. https://doi.org/10.1109/JOE.2002.808212

Kenkel N (1988) Pattern of self-thinning in jack pine: testing the random mortality hypothesis. Ecology 69(4):1017–1024. https://doi.org/10.2307/1941257

Kleijn D, Kohler F, Báldi A, Batáry P, Concepción ED, Clough Y, Díaz M, Gabriel D, Holzschuh A, Knop E, Kovács A, Marshall EJP, Tscharntke T, Verhulst J (2008) On the relationship between farmland biodiversity and land-use intensity in Europe. Proc R Soc B 276:903–909. https://doi.org/10.1098/rspb.2008.1509

Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM (2016) TMB: automatic differentiation and Laplace approximation. J Stat Softw 70 (5): 1–21. 10.18637/jss.v070.i05

Lehikoinen A, Saurola P, Byholm P, Lindén A, Valkama J (2010) Life history events of the Eurasian Sparrowhawk *Accipiter nisus* in a changing climate. J Avian Biol 41:627–636. https://doi.org/10.1111/j.1600-048X.2010.05080.x

Lindqvist BH, Elvebakk G, Heggland K (2003) The trend-renewal process for statistical analysis of repairable systems. Technometrics 45(1):31–44. https://doi.org/10.1198/004017002188618671

Lomnicki ZA (1966) A note on the Weibull renewal process. Biometrika 53(3):375–381. https://doi.org/10.2307/2333644

Madsen PT, Payne R, Kristiansen NU, Wahlberg M, Kerr I, Møhl B (2002) Sperm whale sound production studied with ultrasound time/depth-recording tags. J Exp Biol 205(13):1899–1906. https://doi.org/10.1242/jeb.205.13.1899

Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D, Tyack PL (2013) Estimating animal population density using passive acoustics. Biol Rev 88:287–309. https://doi.org/10.1111/brv.12001

Marques TA, Marques CS, Gikkopoulou KC (2023) A sperm whale cautionary tale about estimating acoustic cue rates for deep divers. J Acoust Soc Am 154(3):1577–1584. https://doi.org/10.1121/10.0020910

Møhl B, Wahlberg M, Madsen PT (2003) The monopulsed nature of sperm whale clicks. J Acoust Soc Am 114(2):1143–1154. https://doi.org/10.1121/1.1586258

Nakagawa T, Subbey S, Solvang HK (2019) Integrating Hawkes process and biomass models to capture impulsive population dynamics. Dyn Cont Discrete Impuls Syst Series B Appl Algorithms 26(3):153–170

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Stat Assoc 3(401):9–27. https://doi.org/10.1080/01621459.1988.10478560

Ong SH, Biswas A, Peiris S, Low YC (2015) Count distribution for generalized Weibull duration with applications. Commun Stat Theory Methods 44(19):4203–4216. https://doi.org/10.1080/03610926.2015.1062105

Ozaki T (1979) Maximum likelihood estimation of Hawkes' self-exciting point processes. Ann Inst Stat Math 31:145–155. https://doi.org/10.1007/BF02480272

Park J, Schoenberg FP, Bertozzi AL, Brantingham PJ (2021) Investigating clustering and violence interruption in gang-related violent crime data using spatial-temporal point processes with covariates. J Am Stat Assoc 116(536):1674–1687. https://doi.org/10.1080/01621459.2021.1898408

Pietzner D, Wienke A (2012) The trend-renewal process: a useful model for medical recurrence data. Stat Med 32(1):142–152. https://doi.org/10.1002/sim.5503

Pringle RM, Doak DF, Brody AK, Jocqué R, Palmer TM (2010) Spatial pattern enhances ecosystem functioning in an African Savanna. PLOS Biol 8(5):e1000377. https://doi.org/10.1371/journal.pbio.1000377

Reinhart A (2018) A review of self-exciting spatio–temporal point processes and their applications. Stat Sci 33(3):299–318. https://doi.org/10.1214/17-STS629

Reynaud-Bouret P, Rivoirard V, Tuleau-Malot C (2013) Inference of functional connectivity in Neurosciences via Hawkes processes. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 317–320, https://doi.org/10.1109/GlobalSIP.2013.6736879

Shamir L, Yerby C, Simpson R, von Benda-Beckmann AM, Tyack P, Samarra F, Miller P, Wallin J (2014) Classification of large acoustic datasets using machine learning and crowdsourcing: application to whale calls. J Acoust Soc Am 135(2):953–962. https://doi.org/10.1121/1.4861348

Shlomovich L, Cohen EAK, Adams N, Patel L (2022) Parameter estimation of binned Hawkes processes. J Comput Graph Stat 31(4):990–1000. https://doi.org/10.1080/10618600.2022.2050247

Smith JN, Goldizen AW, Dunlop RA, Noad MJ (2008) Songs of male humpback whales, *Megaptera novaeangliae*, are involved in intersexual interactions. Anim Behav 76(2):467–477. https://doi.org/10.1016/j.anbehav.2008.02.013

Stimpert AK, DeRuiter SL, Falcone EA, Joseph J, Douglas AB, Moretti DJ, Friedlaender AS, Calambokidis J, Gailey G, Tyack PL, Goldbogen JA (2015) Sound production and associated behavior of tagged fin whales (*Balaenoptera physalus*) in the Southern California Bight. Anim Biotelem 3:23. https://doi.org/10.1186/s40317-015-0058-3

Tonini M, Gonzalez Pereira M, Parente J, Vega Orozco C (2017) Evolution of forest fires in Portugal from spatio–temporal point events to smoothed density maps. Nat Hazards 85:1489–1510. https://doi.org/10.1007/s11069-016-2637-x

Tyack PL, Zimmer WMX, Moretti D, Southall BL, Claridge DE, Durban JW, Clark CW, D'Amico A, DiMarzio N, Jarvis S, McCarthy E, Morrissey R, Ward J, Boyd IL (2011) Beaked whales respond to simulated and actual navy sonar. PLOS One 6(3):e17009. https://doi.org/10.1371/journal.pone.0017009

Wang YF, Tseng ST, Lindqvist BH, Tsui KL (2019) End of performance prediction of lithium-ion batteries. J Qual Technol 51(2):198–213. https://doi.org/10.1080/00224065.2018.1541388

Warren VE, Marques TA, Harris D, Thomas L, Tyack PL, Aguilar de Soto N, Hickmott LS, Johnson MP (2017) Spatio-temporal variation in click production rates of beaked whales: Implications for passive acoustic density estimation. J Acoust Soc Am 141(3):1962–1974. https://doi.org/10.1121/1.4978439

Watwood SL, Miller PJO, Johnson M, Madsen PT, Tyack PL (2006) Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). J Anim Ecol 75(3):814–825. https://doi.org/10.1111/j.1365-2656.2006.01101.x

Wheatley S, Filimonov V, Sornette D (2016) The Hawkes process with renewal immigration and its estimation with an EM algorithm. Comput Stat Data Anal 94:120–135. https://doi.org/10.1016/j.csda.2015.08.007

Yannaros N (1994) Weibull renewal processes. Ann Inst Stat Math 46:641–648. https://doi.org/10.1007/BF00773473

Zhang L, Liu J, Zuo X (2020) Survival analysis of failures based on Hawkes process with Weibull base intensity. Eng Appl Artif Intell 93:103709. https://doi.org/10.1016/j.engappai.2020.103709

Zhao Q, Erdogdu MA, He HY, Rajaraman A, Leskovec J (2015) SEISMIC: a self-exciting point process model for predicting tweet popularity. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1513–1522. https://doi.org/10.1145/2783258.2783401

Zhuang J, Mateu J (2019) A Semiparametric Spatiotemporal Hawkes-Type Point Process Model with Periodic Background for Crime Data. J R Stat Soc Series A (Stat Soc) 182(3):919–942. https://doi.org/10.1111/rssa.12429

Zimmer WMX, Harwood J, Tyack PL, Johnson MP, Madsen PT (2008) Passive acoustic detection of deep-diving beaked whales. J Acoust Soc Am 124(5):2823–2832. https://doi.org/10.1121/1.2988277