TEST
https://doi.org/10.1007/s11749-020-00722-2

ORIGINAL PAPER

Modelling informative time points: an evolutionary process approach

Andreia Monteiro[1] · **Raquel Menezes**[2] · **Maria Eduarda Silva**[3]

Received: 5 June 2019 / Accepted: 9 June 2020

Abstract
Real time series sometimes exhibit various types of "irregularities": missing observations, observations collected not regularly over time for practical reasons, observation times driven by the series itself, or outlying observations. However, the vast majority of methods of time series analysis are designed for regular time series only. A particular case of irregularly spaced time series is that in which the sampling procedure over time depends also on the observed values. In such situations, there is stochastic dependence between the process being modelled and the times of the observations. In this work, we propose a model in which the sampling design depends on all past history of the observed processes. Taking into account the natural temporal order underlying available data represented by a time series, then a modelling approach based on evolutionary processes seems a natural choice. We consider maximum likelihood estimation of the model parameters. Numerical studies with simulated and real data sets are performed to illustrate the benefits of this model-based approach.

Keywords Evolutionary processes · Informative time points · Continuous-time autoregressive process

Mathematics Subject Classification 62M10

1 Introduction

Analysis of experimental data that have been observed at different points in time leads to specific problems in statistical modelling and inference. In traditional time series, the main emphasis is on the case when a continuous variable is measured at discrete

✉ Andreia Monteiro
andreiaforte50@gmail.com

1 CIDMA, University of Minho, Braga, Portugal

2 CBMA, University of Minho, Braga, Portugal

3 CIDMA, Faculty of Economics, University of Porto, Porto, Portugal

Published online: 24 June 2020

equispaced time points, and there is an extensive body of literature on analysing equally spaced time series data, see for example Box et al. (2015) and Brockwell and Davis (2002). Nevertheless, unevenly spaced (also called unequally or irregularly spaced) time series data naturally occur in many scientific domains. For example, data related to natural disasters such as earthquakes, floods, or volcanic eruptions which typically occur at irregular time intervals, give rise to irregularly or unevenly spaced time series. A particular situation of irregularly spaced data is that in which the sampling design depends also, for practical constraints, on the observed values. Examples occur in fisheries where the data are observed when the resource is available, in sensoring when sensors keep only some records in order to save memory and in clinical studies, when a worse clinical condition leads to more frequent observations of the patient. In all such situations, there is stochastic dependence between the process under study and the times at which the observations are made, and the observation times are informative on the underlying process. Ignoring this dependence can lead to biased estimates and misleading inferences.

In this context, Monteiro et al. (2018) introduce the concept of Preferential Sampling in the temporal dimension and propose a model-based approach to make inference and prediction. The suggested framework considers the observed time points as the realization of a time point process stochastically dependent on an underlying latent process (e.g. an individual health indicator, when subjected to regular monitoring).

Monteiro et al. (2018) assumed that the variable of interest is sampled in time according to a sampling design that depends on the values of the underlying process, ignoring the past of the observation processes. However, this kind of assumption of a memoryless process for the observations process having an evolution without aftereffects is sometimes unrealistic and useless in real contexts, where the dependence on the past is crucial.

In this work, we consider that the sampling design may depend on entire past history of the process, meaning all the times of the observations as well as the values of these observations. In these situations, the observed time points can be considered informative to the process being studied. Within the scope of longitudinal studies, the importance of joint modelling informative times and data was already recognised by Ryu et al. (2007) and Liang et al. (2009), who proposed joint modelling and analysis of longitudinal data with possibly informative observation times via latent variables. In these studies, the follow-up time process is considered dependent on the longitudinal outcome process and it should not be regarded deterministic in the design of the study. The analogous problem in the context of longitudinal clinical trial data has been studied too in the context of issues concerning missing values and dropouts, in the sense that a missing observation conveys partial information about the value that would have been observed. See, for example, Diggle and Kenward (1994), Hogan and Laird (1997) and Daniels and Hogan (2008).

Our framework considers joint models for data indexed by informative observation times, assuming a continuous time underlying process observed at irregular and stochastic points. To represent the underlying process, we opt for a continuous time series model such as the continuous time autoregressive (CAR) model, which is mathematically and computationally tractable and yet sufficiently flexible to represent a wide range of phenomena. The assumption that the observation times are informative

and stochastic is equivalent to assuming that they are a realisation of a random process, which is stochastically dependent on the underlying process. This dependence is specified via a model-based approach that relies on point processes, namely marked evolutionary processes.

Point processes provide a very useful theoretical tool to represent the evolution of some random value, or system, over time. In such processes, it is assumed that what happens now may depend on the past, but not on the future. This identifies a natural ordering for temporal point processes. Our interest is to consider a point process that specifies a stochastic model for the time of the next event given we know all the times of previous events. Such processes are termed evolutionary point processes.

The paper is organized as follows. In Sect. 2, we provide some theoretical background on evolutionary point processes. In Sect. 3, we consider that the sampling design may depend on all past history of the process and we propose a model, based on evolutionary processes that take into account that the times and values of the observations contain important information for the underlying process (informative and stochastic time points). We proceed with likelihood inference to estimate the parameters of this model and we consider a numerical method based on a Laplace approach to optimize the likelihood. In Sect. 4, using numerical studies, we document the performance of this approach comparing the results of main parameter estimates with those obtained from the traditional approach for irregularly spaced data. In Sect. 5, we show the application of the previously described methodology to a real data set related to monitoring the level of a biomedical marker, after a cancer patient undergoes a bone marrow transplant. Section 6 is devoted to make some concluding remarks.

2 Background on evolutionary processes

We start by reviewing some concepts on evolutionary point processes (Daley and Vere-Jones 2003). To set the notation, let $(T_n)_{n\in\mathbb{N}}$ denote an increasing sequence of positive random times. An important concept in evolutionary processes is the history of the process, denoted by H_t which represents the entire history of the point process (T_n) prior to time t, meaning that H_t specifies the times of all point events in the interval $(-\infty, t)$. We refer to $\widetilde{H}_t$ as the observed history of the process over the interval $[0, t)$, that is the history consistent with an observation on the process.

In this work, the point process is assumed to be simple point processes, meaning that no points coincide and therefore the points can be ordered strictly in time. Furthermore, the specification of the point process conditional on its history is via the conditional intensity function.

The conditional intensity can be written directly in terms of the hazard functions since hazard function has a natural interpretation as the conditional instantaneous event rate. Following Daley and Vere-Jones (2003), given a sequence t_i with $0 < t_1 < \cdots < t_n < \cdots$ the hazard functions are defined by

$$\lambda^*(t) = \begin{cases} h_1(t), & 0 < t \leq t_1 \\ h_n(t|t_1, \ldots, t_{n-1}), & t_{n-1} < t \leq t_n, n \geq 2 \end{cases}$$

The intuitive content of the notion of a conditional intensity function is well-expressed through the suggestive relation

$$\lambda^*(t)\mathrm{dt} = E\left[N(\mathrm{dt})|\widetilde{H}_t\right]$$

where dt is an infinitesimal interval around t, $N(\cdot)$ denotes the number of points falling in an interval and $\widetilde{H}_t$ is the σ-algebra of events occurring at times up to but not including t. Thus, the conditional intensity can be interpreted as the conditional risk of the occurrence of an event at t, given the realization of the process over the interval $[0, t)$.

2.1 Conditional intensity function

The conditional intensity function of the point process (T_n), $\lambda^*(t) = \lambda(t|\widetilde{H}_t)$, is defined by

$$\lambda^*(t) = \frac{f_T(t|\widetilde{H}_t)}{1 - F_T(t|\widetilde{H}_t)},\ t_1 < \cdots < t_{n-1} < t < t_n < \cdots \tag{1}$$

where $f_T(t|\widetilde{H}_t)$ is the conditional density and $F_T(t|\widetilde{H}_t)$ is the corresponding cumulative distribution function.

Intuitively, the conditional intensity at t gives the conditional "risk" of a point event occurring at that instant in time, given the observed history of the process prior to time t.

Examples of point processes in which the conditional intensity has a particular functional form are the following:

- The (inhomogeneous) Poisson process. In this process the number of points in disjoint sets is independent and the conditional intensity function inherits this property. The Poisson process is quite simply the point process in which the conditional intensity function is independent of the past, i.e. the conditional intensity function is equal to the intensity function of the Poisson process, $\lambda^*(t) = \lambda(t)$.
- The conditional intensity function of a Hawkes process, Hawkes (1971), with an exponential decay function has the form

$$\lambda^*(t) = \eta + \psi \sum_{i:t_i \in (0,t)} \exp(-\gamma(t - t_i))$$

where $\eta > 0$, $\psi \geq 0$, $\gamma > 0$ and $\psi < \gamma$ for the process to be stationary. Note that each time a new point arrives in this process, the conditional intensity grows by ψ and then decreases exponentially back towards η. In other words, a point increases the chance of getting other points immediately after (self-exciting). Setting $\psi = 0$, return us to the homogeneous Poisson process.

2.2 Marked point processes

In addition to the times of the point events, there may be additional variables that are of interest associated with each point event. This information is known as marks and the mark space (M) can be of many different types, but it is often (a subset of) $\mathbb{R}$ or $\mathbb{N}$. The marks may have an independent interest or may be included to make a more realistic model of the event times. For example, in the analysis of a particular medical indicator, it is relevant to know its value and not only when it was observed. In addition, the value of the indicator influences how often measurements are taken.

More formally, a marked point process, with point event times in $\mathbb{R}$ and marks in M, is a point process $\{(T_n, Y_n)_{n\in\mathbb{N}}\}$ on $\mathbb{R} \times M$ with the additional property that the process associated with times $t_1, t_2, \ldots$, the ground process, is itself a point process on $\mathbb{R}$. We specify a marked point process by defining the conditional intensity $\lambda(\cdot|\widetilde{H}_t)$ of the ground process, and then, for a given point event and observed history at time t, we define the conditional distribution function for the marks. The later may be represented as $f_Y^*(y|t) = f_Y(y|t, \widetilde{H}_{t,y})$, specifying the density of the mark Y given t and the history of the process that now includes information of times and marks of past events. This means that the definitions of the complete and observed histories, $H_{t,y}$, and $\widetilde{H}_{t,y}$, and the conditional intensity function was extended for marked point processes. We can now define the conditional intensity function for the marked case as

$$\lambda^*(t, y) = \lambda^*(t) f_Y^*(y|t) \tag{2}$$

$\lambda^*(t)$ is called the ground intensity and is defined exactly as the conditional intensity function for the unmarked case, except that it is allowed to depend on the marks of the past events. In addition, the marks are assumed to be conditionally independent given the history of the marked point process and unanticipated. A process is said to have unanticipated marks if the distribution of the mark at t_i is independent of all previous point event times and marks.

Thus, we can rewrite (2) as

$$\lambda^*(t, y) = \frac{f_{T,Y}(t, y|\widetilde{H}_{t,y})}{1 - F_T(t|\widetilde{H}_t)}$$

where $f_{T,Y}(t, y|\widetilde{H}_{t,y})$ is the joint density of the time and the mark, conditional on past times and marks, and $F_T(t|\widetilde{H}_{t,y})$ is the conditional cumulative distribution function of T also conditional on the past times and marks.

A marked point process $(\mathbf{T},\mathbf{Y}) = \{(T_1, Y_1), (T_2, Y_2) \cdots\}$ is strongly stationary if all the shifted marked point processes $\{(T_1 - s, Y_1), (T_2 - s, Y_2), \ldots\}$ have the same distribution with $s \in \mathbb{R}$.

An example of a marked point process is the marked Hawkes process. This process is a generalization of the unmarked Hawkes process, such that each point event time

now has a mark associated with it. The conditional intensity of the ground process is given by

$$\lambda(t|\widetilde{H}_t) = \lambda^*(t) = \eta + \psi \sum_{t_i : t_i \in (0,t)} \exp(\beta_1 y_i) \exp(-\gamma(t - t_i)) \tag{3}$$

where $\eta, \gamma > 0$, $\psi, \beta_1 \geq 0$ and y_i denotes the observed value at time t_i.

Equivalently, we could define it by its conditional intensity function including both marks and times

$$\lambda^*(t, y) = \left(\eta + \psi \sum_{t_i : t_i \in (0,t)} \exp(\beta_1 y_i) \exp(-\gamma(t - t_i)) \right) f^*(y|t) \tag{4}$$

The idea behind using this model is that every new event increases the intensity by $\psi \exp(\beta_1 y_i)$ and large events increase the intensity more than small.

2.3 Inference

Daley and Vere-Jones (2003) note that for point processes described as having an evolutionary character, their conditional intensities and likelihoods are relatively simple. The evolutionary character of such point processes allows the likelihood to be found by successively conditioning on the past. Explicitly, the likelihood of a realization $((t_1, y_1), \ldots, (t_n, y_n))$ on $[0, T) \times \mathbb{R}$, of a marked point process is given by

$$L_E = \left(\prod_{i=1}^{n} \lambda^*(t_i) \right) \exp\left(- \int_0^T \lambda^*(u) \mathrm{d}u \right) \left(\prod_{i=1}^{n} f_Y^*(y_i|t_i) \right) \tag{5}$$

See Daley and Vere-Jones (2003, p. 246–256) for a development of the likelihood. The third factor on the right-hand side of (5) is the contribution to the likelihood from the observed marks.

The use of the corresponding log-likelihood implies bearing in mind some practical considerations. A point process is only observed for a finite interval $[0, T]$ and time 0 is some time after the origin of the process. For evolutionary point processes, there may be effects from point events occurring before time 0. Daley and Vere-Jones (2003) referred such effects as edge or boundary effects. An approach often taken in the literature is ignoring the effects from point events occurring before the start of the observation period. In this case, the conditional intensity can be regarded as approximate for some initial part of the observation period, and as such, there is likely to be some effect on the estimated model. Rasmussen (2013) highlights that the estimate of η is likely to be too high, however, he noted that the effects on the estimated model will be negligible if the data set being used is large.

3 An evolutionary model for informative time points

Consider an unobserved stochastic process in time $S(t)$, represented by a CAR(1) that satisfies the differential equation

$$\mathrm{d}S(t) + \alpha_0 S(t)\mathrm{d}t = \mathrm{d}W(t)$$

where α_0 is the autoregressive coefficient and $W(t)$ is a Wiener process with variance parameter σ_w^2. $S(\cdot)$ is a stationary Gaussian process if $\alpha_0 > 0$, with $E[S(t)] = 0$. Now admit that $S(t)$ is observed at times t_i, $i = 1, \ldots, n$, yielding a data set (t_i, y_i), where the corresponding $Y_i = Y(t_i)$ is the noisy version of $S(t_i)$, $Y(t_i) = \mu + S(t_i) + N(0, \tau^2)$. Since our goal is to infer on $S(t)$, admitting that the sampling times are stochastic and the sampling design may depend on all past history of the process (both the actual times and values of the observations) then a model able to deal with this evolutionary character must specify the joint distribution of S, $T = (t_1, \ldots, t_n)$ and $Y = (Y_1, \ldots, Y_n)$, $[S, T, Y]$. Considering that $[S, T, Y] = [S][T, Y\,|S]$ let $\{(T, Y)|S\}$ be an evolutionary marked point process[1] with ground intensity similar to (3)

$$\lambda_S^*(t) = \lambda(t|\widetilde{H}_t, S) = \eta + \psi \sum_{t_i : t_i \in (0,t)} \exp(\beta_1 y_i) \exp(-\gamma(t - t_i)) \tag{6}$$

with $\eta, \gamma > 0$, $\psi < \gamma$ and $\psi, \beta_1 \geq 0$.

Admitting the conditional mark density, $f_S^*(y|t) = f^*(y|t, S)$, then according to (2), the conditional intensity function including both marks and times is

$$\lambda_S^*(t, y) = \lambda(t, y|\widetilde{H}_t, S) = \lambda_S^*(t) f_S^*(y|t) \tag{7}$$

The main purposes behind this model are

- every new event increases the intensity by $\psi \exp(\beta_1 y_i)$ and large events increase the intensity more than small events;
- observations that are more distant in time have less influence, considering on γ parameter;
- the initial value of the conditional intensity equals η and we ignore effects from events occurring before the first observation.

This model, henceforth EVOL, allows to take into account the history of the process, capture the evolutionary character of the process and deal with irregularly spaced time series.

3.1 Maximum likelihood estimation

To obtain estimates for the parameters of the model we use maximum likelihood estimation. For the shared latent process model, the likelihood function for data T and

[1] This point process is stationary under the conditions given by Proposition 6.4. VII. from Daley and Vere-Jones (2003)

Y can be expressed as

$$L(\boldsymbol{\theta}) = [T, Y] = \int_S [T, Y, S]\mathrm{d}S = \int_S [S][T, Y|S]\mathrm{d}S \tag{8}$$

where $\boldsymbol{\theta} = \left(\mu, \sigma = \sqrt{\frac{\sigma_w^2}{2\alpha_0}}, \phi = \frac{1}{\alpha_0}, \tau, \beta_1, \gamma, \psi, \eta\right)$ represents all the model parameters.

Considering that the likelihood of a marked point process is given by (5), $[T, Y|S]$ in (8) takes the form

$$[T, Y|S] = \left(\prod_{i=1}^{n} \lambda_S^*(t_i)\right) \exp\left(-\int_0^T \lambda_S^*(u)\mathrm{d}u\right) \left(\prod_{i=1}^{n} f_S^*(y_i|t_i)\right)$$

The associated log-likelihood function is given by

$$\log([T, Y|S]) = \sum_{i=1}^{n} \log \lambda_S^*(t_i) - \int_0^T \lambda_S^*(u)\mathrm{d}u + \sum_{i=1}^{n} \log f_S^*(y_i|t_i) \tag{9}$$

Substituting in (9), the conditional (ground) intensity, $\lambda_S^*(\cdot)$, and the conditional mark density $f_S^*(y_i|t_i)$, specified as $N(S_i, \tau^2)$, then the log-likelihood can be rewritten as

$$\begin{aligned}\log([T, Y|S]) = &\sum_{i=1}^{n} \log\left(\eta + \psi \sum_{j:t_j<t_i\in(0,t)} \exp(\beta_1 y_j - \gamma(t_i - t_j))\right) \\ &-\eta T - \frac{\psi}{\gamma}\sum_{i=1}^{n} \exp(\beta_1 y_i)(1 - \exp(-\gamma(T - t_i))) \\ &-\frac{n}{2}\log(2\pi\tau^2) - \frac{1}{2\tau^2}\sum_{i=1}^{n}(y_i - S_i)^2\end{aligned} \tag{10}$$

3.2 Computational procedures

The calculation of the integral in (8) is perfomed in three steps. In a first step, we need to calculate the log of the joint bivariate distribution of the observations conditional on the underlying process S as given in expression (10). To overcome the computational burden resulting from the nested sum in the first term

$$\sum_{i=1}^{n} \log\left(\eta + \psi \sum_{j:t_j<t_i\in(0,t)} \exp(\beta_1 y_j - \gamma(t_i - t_j))\right),$$

we use a compiled C++ subroutine. Additionally, that we set $\lambda(0|\widetilde{H}_0) = \eta$ and ignore the effects from point events occurring before time 0.

In a second step we need to approximate $[S]$, the distribution of the unobserved underlying process. For this purpose we use a technique based on stochastic partial differential equations (SPDE). Following Lindgren et al. (2011), we represent a Gaussian process with Matérn covariance structure as the solution of the following SPDE,

$$\left(\phi^{-2} - \Delta\right)^{\alpha/2} (\omega S(t)) = \epsilon(t), \ t \in \mathbb{R}^+, \tag{11}$$

where $\epsilon(t)$ is Gaussian white noise, Δ is the Laplacian and ϕ is the range parameter of the Matérn covariance function $\gamma(u)$ in its standard parametrization,

$$\gamma(u) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} (u/\phi)^{\nu} K_{\nu} (u/\phi) : u \geq 0$$

where K_ν is the modified Bessel function of second kind and order $\nu > 0$ and σ^2 is the marginal variance. The integer value of ν determines the mean square differentiability of the underlying process, which matters for predictions made using such a model. However, ν is usually fixed since it is poorly identified in typically applications. The remaining parameters in (11) are $\alpha = \nu + 1/2$, from this we can identify the exponential covariance with $\nu = 1/2$, and ω that controls the variance,

$$\omega^2 = \frac{\Gamma(1/2)}{\Gamma(1)(4\pi)^{1/2}\phi^{-1}\sigma^2} \tag{12}$$

Finally, S is approximated by $\widetilde{S}$, where

$$\widetilde{S}(t) = \sum_{k=1}^{m} \psi_k(t) W_k, \ t \in \mathbb{R}^+$$

with $\psi_k(\cdot)$ being piecewise linear basis functions at a set of time knots and $W = W_1, ..., W_m$ is a zero-mean multivariate Gaussian variate with covariance matrix Q^{-1}. The construction is done by projecting the SPDE onto the basis representation in what is essentially a Finite Element method. For $\alpha = 1$ the required form of Q is

$$Q = \omega^2(\phi^{-2}C + G_2)$$

where C and G_2 are sparse matrices whose explicit expressions can be found in Lindgren et al. (2011).

In the last step, to compute the integral in the likelihood (8), we utilize automatic differentiation of a Laplace approximation to the marginal likelihood, following Kristensen et al. (2016).

Note that the likelihood function for $L(\boldsymbol{\theta})$ can be written as

$$L(\boldsymbol{\theta}) = \int_S \exp(-f(S, \boldsymbol{\theta}))\mathrm{d}S \tag{13}$$

where $f(S, \boldsymbol{\theta})$ denote the negative joint log-likelihood of the data, $\boldsymbol{\theta}$ is the vector of parameters (fixed effects) and S the random effects. Thus the Laplace approximation for $L(\boldsymbol{\theta})$ is

$$L^*(\boldsymbol{\theta}) = (2\pi)^{N/2} \det(H(\boldsymbol{\theta}))^{-1/2} \exp(-f(\widehat{S}(\boldsymbol{\theta}), \boldsymbol{\theta}))$$

where

$$\widehat{S}(\boldsymbol{\theta}) = \arg_S \min \mathrm{f}(S, \boldsymbol{\theta}) \tag{14}$$

and $H(\boldsymbol{\theta})$ is the Hessian of f with respect to S evaluated at $\widehat{S}(\boldsymbol{\theta})$,

$$H(\boldsymbol{\theta}) = \frac{\partial^2}{\partial S^2} f(S, \boldsymbol{\theta})|_{S=\widehat{S}(\boldsymbol{\theta})}$$

The estimate of $\boldsymbol{\theta}$ minimizes the negative of the logarithm of the Laplace approximation,

$$-\log L^*(\boldsymbol{\theta}) = -\frac{N}{2}\log(2\pi) + \frac{1}{2}\log\det(H(\boldsymbol{\theta})) + f(\widehat{S}(\boldsymbol{\theta}), \boldsymbol{\theta}) \tag{15}$$

This objective function and its derivatives acquired by using automatic differentiation, are required to apply standard nonlinear optimization algorithms (e.g. nlmimb) to optimize the objective function and obtain the estimate for $\boldsymbol{\theta}$.

Uncertainty of the estimate $\widehat{\boldsymbol{\theta}}$ or of any differentiable function of the estimate $\zeta(\widehat{\boldsymbol{\theta}})$ that the user specifies, is obtained by the δ-method:

$$\mathrm{Var}(\zeta(\widehat{\boldsymbol{\theta}})) = -\left\{ \frac{\partial \zeta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \left[\frac{\partial^2(\log L^*(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} \frac{\partial \zeta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \tag{16}$$

These uncertainty calculations also require derivatives of (15). However, derivatives are straight-forward to obtain using automatic differentiation in this context.

In particular, using the R package TMB, short for Template Model Builder, (Kristensen et al. 2016), the user has to define the joint log-likelihood of the data and (i.e. conditional on) the random effects as a C++ template function. The other operations such as integration and calculation of the marginal score function, are done directly in R language. The package evaluates and maximizes the Laplace approximation of the marginal likelihood, where the random effects are automatically integrated out. This approximation, and its derivatives, are obtained using automatic differentiation (up to order three) of the joint likelihood. In the case of sampling design that may depend on entire past history of the process, we simply have to define the joint negative log-likelihood as

$$f(S, \boldsymbol{\theta}) = -log([S][T, Y|S])$$

and allow TMB package, (Kristensen et al. 2016), to integrate out the latent field S to evaluate approximately (8).

4 Numerical studies

We now intend to proceed with the assessment of the EVOL model, comparing the results of its parameter estimates and those of the traditional Kalman filter approach to irregularly spaced data (cts package (Wang 2013)). We use simulated time series, so we start by describing the procedure needed to simulate a marked point process.

4.1 Simulation design

The classic method to simulate an inhomogeneous Poisson process is the thinning method of Lewis and Shedler (1979). This method requires that the conditional intensity to be bounded above, i.e. there is a finite M such that for all t, $\lambda(t|\widetilde{H}_t, S) \leq M$. This method was generalised by Ogata (1981) and this generalisation only requires that the intensity to be locally bounded. The algorithm is described as follows. Suppose we can find a piecewise constant process $M(\cdot|\widetilde{H}_t, S)$, conditional on the history of the point process, such that for $t \in [0, T)$,

$$\lambda(t|\widetilde{H}_t, S) \leq M(\cdot|\widetilde{H}_t, S)$$

Given that we can find a suitable $M(\cdot|\widetilde{H}_t, S)$, we can simulate a realisation of the point process of interest in this way: define an inhomogeneous Poisson process N^* which has a piecewise constant intensity $M(\cdot|\widetilde{H}_t, S)$ that changes value according to the history $\widetilde{H}_t$ and decide on the termination condition, for example, the simulation interval is $[0, T)$, then simulate the points $0 \leq t_1^* < t_2^* < \cdots < t_{N^*[0,T)}^* < T$ from the process N^*. Each t_i^* is then selected with probability $\lambda(t_i^*|\widetilde{H}_{t_i^*}, S_{t_i^*})/M(t_i^*|\widetilde{H}_{t_i^*}, S_{t_i^*})$ to form part of the simulated realisation of the point process of interest, where the history $H_{t_i^*}$ and $S_{t_i^*}$ give the simulated history of the point process of interest up to time t_i^*. For each point t_i that is selected to the simulated realisation of the point process of interest we simulate a mark y_i from $Y(t) = \mu + S(t) + N(0, \tau^2)$.

In practice, the function $M(\cdot|\widetilde{H}_t, S)$ changes value each time a point event is added to the simulated realisation of the process of interest, and so it will not be known before carrying out the simulation.

To generate a time series under a preferential sampling design that depends on all past history of the process, we adapt the R code used by Lapham (2014, p.124–125).

As follows, we start to generate a realization of S, a CAR(1) process[2] with $\alpha_0 = 0.2$ and $\sigma_w^2 = 1$. These values correspond to $\text{Var}[S(\cdot)] = \sigma^2 = \frac{\sigma_w^2}{2\alpha_0} = (1.581)^2$ and $\phi = \frac{1}{\alpha_0} = 5$, being the latter related to the lag beyond which there is no correlation for practical purposes. The parameter values used to generate the marked point process are

$$\eta = 0.05,\ \psi = 0.025,\ \beta_1 = 0.6,\ \gamma = 0.1$$

and to generate the marks y_i, we consider $\mu = 0$ and $\tau = 0.1$.

[2] We use package yuima, with $S_0 = 0$ and a discretization of time domain in 1600 points equally spaced.

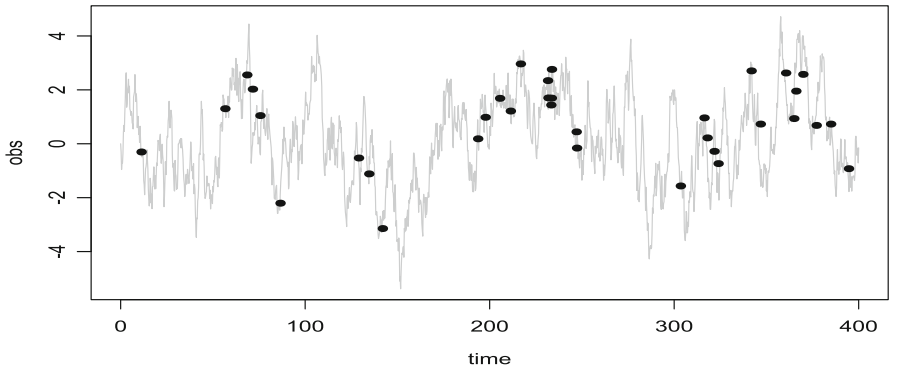

Fig. 1 Sample times with dependency on all past history of the process and underlying process S (gray line)

Table 1 Maximum likelihood estimates, under EVOL approach and by Kalman filter approach (CTS): with mean (standard errors) obtained from a total of 500 independent samples

	True	EVOL	CTS
μ	**0**	**0.196** (0.267)	**0.225** (0.304)
σ	**1.581**	**1.567** (0.204)	**1.606** (0.209)
ϕ	**5**	**5.995** (1.647)	**6.188** (1.617)
τ	**0.1**	**0.456** (0.197)	**0.483** (0.194)
β_1	**0.6**	**0.618** (0.128)	
γ	**0.1**	**0.095** (0.026)	

Bold values are highlight parameter estimates from standard errors

To illustrate the results of these sampling procedure, we represent in Fig. 1 a realization of the process S (gray line) and the resulting data set.

4.2 Estimation results

For EVOL model, η and ψ parameters have a tuning role. Relatively to η, we ignore effects from point events occurring before the start of the observation period and we assume that the initial value of the conditional intensity equals η. Regarding ψ, it controls the sum value in the ground intensity. Thus, in a first simulation study the parameters μ, σ, ϕ, τ, β_1 and γ are the target of estimation and we set η and ψ values at the true ones. For the simulation study we consider a total of 500 independent samples with at least 50 points over the interval $[0, 400]$. The results of the mean and the standard errors for each parameter, obtained from EVOL model, under (9), and from Kalman filter approach implemented via cts package (Wang 2013), are summarized in Table 1. In Fig. 2 we have the corresponding boxplots, with true parameter values marked as red line.

By analysing Table 1 and Fig. 2, we conclude that EVOL model presents more accurate estimates than Kalman filter approach. The parameter τ seems to be overestimated in both approaches. For β_1 and γ the estimates are quite reasonable and we believe that the inclusion of these two parameters in the model is more realistic in real contexts.

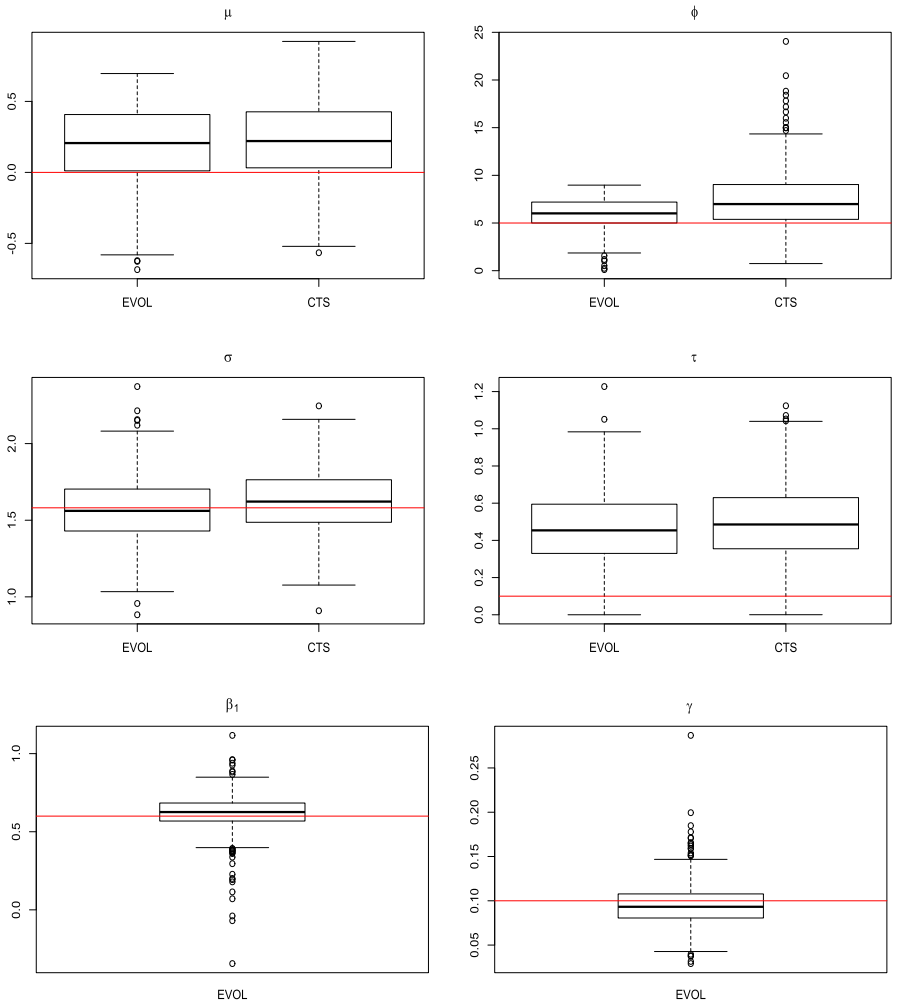

Fig. 2 Boxplots for models parameters estimated over 500 independent samples with true parameter values marked as red line, under EVOL and Kalman filter (CTS) approaches

Further studies with different combinations of the parameters, namely for β_1 and γ were analysed. When $\beta_1 > \gamma$ the conclusions are similar, but when $\beta_1 < \gamma$ or $\beta_1 > 1$ it is necessary to do some calibration work with parameter ψ in order to obtain samples with a reasonable dimension.

4.3 Sensitivity analysis

We conduct a second simulation study aiming: to analyse the impact of also estimating parameters η and ψ; and to investigate the sensitivity in parameter estimation to initial values, needed by the iterative procedure supporting the likelihood method.

Table 2 MLE's, mean (standard errors) obtained from a total of 200 independent samples, considering as initial values for EVOL approach the parameters estimated by traditional Kalman filter

	True	EVOL (True $\boldsymbol{\theta_0}$)	EVOL ($\boldsymbol{\theta_0}$ from Kalman filter)
μ	**0**	**0.239** (0.276)	**0.239** (0.276)
σ	**1.581**	**1.526** (0.224)	**1.526** (0.224)
ϕ	**5**	**5.720** (1.429)	**5.720** (1.429)
τ	**0.1**	**0.477** (0.202)	**0.477** (0.202)
β_1	**0.6**	**0.782** (0.325)	**0.782** (0.325)
γ	**0.1**	**0.114** (0.065)	**0.114** (0.065)
ψ	**0.025**	**0.025** (0.023)	**0.025** (0.023)
η	**0.05**	**0.065** (0.021)	**0.065** (0.021)

Bold values are highlight parameter estimates from standard errors

Thus, we first consider as initial values ($\boldsymbol{\theta_0}$) the "true" values. We then consider: for μ, ϕ, σ and τ, the estimates obtained by the traditional Kalman filter approach; and, for the other parameters, $\beta_1 = 0.4$, $\gamma = 0.2$, $\eta = 0.07$ and $\psi = 0.035$.

The results of the mean and standard errors for each parameter, obtained from a total of 200 independent samples are summarized in Table 2.

The proposed method seems to be quite robust to initial values and the inclusion of parameters η and ψ do not cause identifiability issues, only parameter β_1 is a little overestimated. In real data applications η and ψ are the calibration parameters that need to be tunned, for example, iteratively during the estimation of the model. The final ML estimates of the model parameters are then obtained for pre-defined values of η and ψ.

5 Application to real data

We now consider the problem of monitoring the level of two biomedical markers, platelet (PLT) and hematocrit (HTC), after a cancer patient undergoes a bone marrow transplant. The data, composed by 54 measurements over 91 days of log(PLT) and log(HTC) shown in Fig. 3, is studied by Shumway and Stoffer (2017) as missing data problem. These data are made available in package astsa Stoffer (2017) with the name of "blood".

The biomedical marker PLT was also studied by Monteiro et al. (2018), who present a model to deal with irregularly spaced time series in which the sampling design only depends on the contemporaneous value of the underlying process where conditional on S, T is an inhomogeneous Poisson process with intensity $\lambda(t) = \exp\{a + \beta S(t)\}$ and conditional on S and T, Y is a set of mutually independent Gaussian variates with τ^2 being the measurement error variance. We now intend to relate the results of the two models, both targeting preferential sampling issues. We need to have in mind that the convergence of the algorithm proposed by Monteiro et al. (2018) is very slow and the running time becomes burdensome for longer time series and a large number of Monte Carlo samples. Besides these, the large variability between likelihood values in each

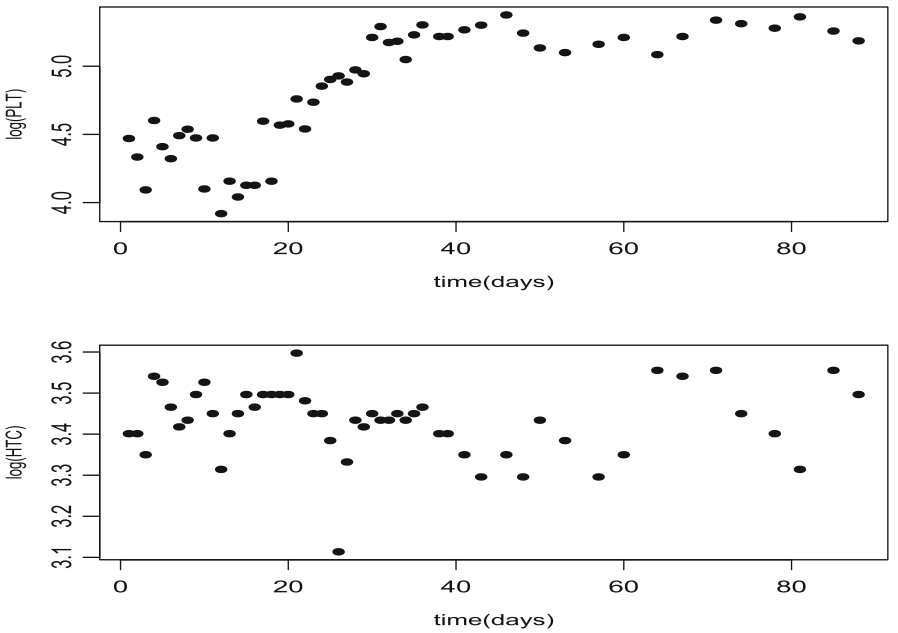

Fig. 3 Measurements of biomedical markers platelet and hematocrit, in the logarithm scale, log(PLT) and log(HCT)

Monte Carlo iteration makes the likelihood difficult to optimize. So, we here suggest an alternative method (henceforth LAP), detailed in Monteiro et al. (2019), that uses automatic differentiation of a Laplace approximation to the marginal likelihood, as described in Sect. 3.2, to approximate the integral in the likelihood

$$L(\boldsymbol{\theta}) = [T, Y] = \int_S [T, Y, S]\mathrm{d}S = \int_S [S][T, Y|S]\mathrm{d}S = \int_S [S][T|S][Y|T, S]\mathrm{d}S$$

The approximation of the Gaussian process S through the SPDE technique is also done as explained in Sect. 3.2.

The proposed evolutionary model assumes that $\beta_1 \geq 0$, since we are considering the situation where sampled times are concentrated, predominantly, near the maxima. As PLT observations are near the minima, we perform an axial reflection around the mean.

The estimated parameters, for log(PLT) biomedical marker, together with estimated standard errors are summarized in Table 3.[3] Comparing the parameter estimates, we confirm that for μ, ϕ, σ and τ the EVOL and LAP approaches are in accordance.

Parameter β_1 in (6) is not statistically significant, however γ is significant reinforcing the fact that the distance to the previous observed times is relevant.

The estimated parameters, for log(HTC) marker, together with estimated standard errors are summarized in Table 4. Parameters β_1 and β are not statistical significant.

[3] Recall ω is a reparametrization of σ, defined in Eq. (12).

Table 3 Maximum likelihood estimates under LAP and EVOL for log(PLT)

Parameter	LAP model		EVOL model	
	Estimate	Standard error	Estimate	Standard error
μ	4.993	0.290	4.986	0.287
$\log(\omega)$	2.545	0.198	2.530	0.198
σ	0.329		0.329	
$\log(\phi)$	3.559	0.710	3.529	0.708
ϕ	35.115		34.079	
$\log(\tau)$	-2.086	0.132	-2.091	0.133
τ	0.124		0.124	
β	-0.936	0.316	–	
β_1	–		1.170	0.712
γ	–		0.091	0.039
η	–		0.002	
ψ	–		0.077	

Table 4 Maximum likelihood estimates under LAP and EVOL for log(HTC)

Parameter	LAP model		EVOL model	
	Estimate	Standard error	Estimate	Standard error
μ	3.428	0.019	3.429	0.019
$\log(\omega)$	2.372	0.150	2.378	0.147
σ	0.088		0.087	
$\log(\phi)$	0.567	0.369	0.570	0.366
ϕ	1.763		1.768	
$\log(\tau)$	-11.306	3991.8	-8.523	$\gg 0$
τ	0.00001		0.0002	
β	0.491	1.923	–	
β_1	–		0.475	1.084
γ	–		0.149	0.039
η	–		0.002	
ψ	–		0.028	

The lack of significance for parameter ϕ is in accordance with the low temporal correlation revealed by the initial analysis of the autocorrelation function.

Prediction

A practical question relevant in the monitoring of biomedical markers, or any other such quantity, is to predict the time of the next observation of a pre-defined value of the marker or variable under study. Such prediction may allow a more precise schedule of visits to the hospital, for example. The theoretical complexity underlying point

Table 5 Mean absolute percentage error (MAPE) and mean absolute error (MAE) for prediction time t_{54}, given $Y(t_{54}) = 0.356$ and all history of the process up to t_{53}. The true t_{54} is 88

	Prediction of time t_n
MAPE	1.66%
MAE	1.46

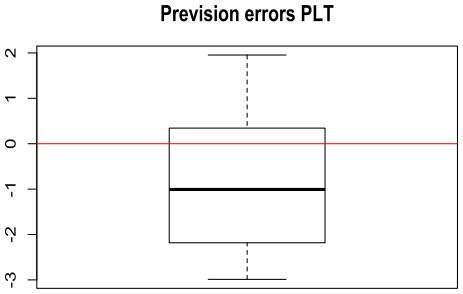

Fig. 4 Boxplot with prediction errors of predicted t_n versus true $t_n = 88$

processes prediction exercises and the unavailability of explicit numerical solution (Daley and Vere-Jones 2008) suggest resorting to Monte Carlo approaches. Here we focus on the ability of our model to predict the hitting time for a given threshold for the mark based on simulation. In fact, since the expression of the conditional intensity function is known, simulation of the marked point process is straightforward.

To illustrate the approach we assume that the process is observed until time $t_{53} = 85$ days and predict the time t_{54} when the mark $Y(t_{54})$ reaches the level assumed of interest 0.356. First we estimate the model EVOL using observations up t_{53} and proceed using the simulation design described in Sect. 4.1. The Monte Carlo study comprised 500 replications (R) and the corresponding results are assessed by MAPE[4] and MAE[5] errors, presented in Table 5. Figure 4 shows the boxplot for prediction errors.

The results indicate that the proposed model may be useful for predictive purposes.

6 Conclusions

In this work, we present a model approach that allows to deal with sampling designs that depend on all past history of the process. This model allows to take into account the evolutionary character of the process and is, in our opinion more realistic, since it also considers the previous observations and the temporal distance to which they occurred. To specify a process conditional on the past we considered the intensity function and a marked point process for the times T and marks Y. The results for the parameter estimation are quite satisfactory, the algorithm is computationally efficient and provides user high levels of flexibility, due to the direct specification of the joint likelihood. Further, the proposed model-based approach can be used to obtain pre-

[4] $\text{MAPE} = \frac{1}{R}\sum_{r=1}^{R}\frac{|t_{54}-\widehat{t}_{54,r}|}{t_{54}} \times 100\%$

[5] $\text{MAE} = \frac{1}{R}\sum_{r=1}^{R}|t_{54}-\widehat{t}_{54,r}|$

dictions on the next event, given a possible value for a future mark, allowing a better knowledge of future events.

Nonetheless, the discussed model presents some difficulties when applied to real data, namely in the definition of the initial values of the calibration parameters (η and ψ) and we intend to define a simple method to choose suitable starting values for these parameters.

It is also of our interest to apply this model in other scientific areas, for example in the context of financial markets, where the volume of transactions may depend on the history of the process.

Acknowledgements The authors acknowledge Foundation FCT (Fundação para a Ciência e Tecnologia) as members of the research project PTDC/MAT-STA/28243/2017 and Center for Research & Development in Mathematics and Applications of Aveiro University within project UID/MAT/04106/2019.

References

Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. Wiley, Hoboken

Brockwell PJ, Davis RA (2002) Introduction to time series and forecasting. Springer, Berlin

Daley DJ, Vere-Jones D (2003) An introduction to the theory of point processes. Elementary theory and methods, vol 1, 2nd edn. Springer, New York

Daley DJ, Vere-Jones D (2008) An introduction to the theory of point processes. General theory and structure, vol 1, 2nd edn. Springer, New York

Daniels MJ, Hogan JW (2008) Missing data in longitudinal studies: strategies for Bayesian modeling and sensitivity analysis. CRC Press, Boca Raton

Diggle P, Kenward MG (1994) Informative drop-out in longitudinal data analysis. J R Stat Soc Ser C (Appl Stat) 43:49–93

Hawkes AG (1971) Spectra of some self-exciting and mutually exciting point processes. Biometrika 58(1):83–90

Hogan JW, Laird NM (1997) Model-based approaches to analysing incomplete longitudinal and failure time data. Stat Med 16(3):259–272

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

Lapham BM (2014) Hawkes processes and some financial applications. PhD thesis, University of Cape Town

Lewis PAW, Shedler GS (1979) Simulation of nonhomogeneous poisson processes by thinning. Naval Res Logist Q 26(3):403–413

Liang Y, Lu W, Ying Z (2009) Joint modeling and analysis of longitudinal data with informative observation times. Biometrics 65(2):377–384

Lindgren F, Rue H, Lindström J (2011) An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. J R Stat Soc Ser B (Stat Methodol) 73(4):423–498

Monteiro A, Menezes R, Silva ME (2018) Modelling irregularly spaced time series under preferentail sampling. RevStat Stat J **(accepted)**

Monteiro A, Menezes R, Silva ME (2019) Modelling preferential sampling in time. BEIO 35(3):180–196

Ogata Y (1981) On Lewis' simulation method for point processes. IEEE Trans Inf Theory 27(1):23–31

Rasmussen JG (2013) Bayesian inference for Hawkes processes. Methodol Comput Appl Probab 15(3):623–642

Ryu D, Sinha D, Mallick B, Lipsitz SR, Lipshultz SE (2007) Longitudinal studies with outcome-dependent follow-up: models and Bayesian regression. J Am Stat Assoc 102(479):952–961

Shumway RH, Stoffer DS (2017) Time series analysis and its applications: with R examples, 4th edn. Springer texts in statistics. Springer, New York

Stoffer D (2017) astsa: Applied statistical time series analysis. https://CRAN.R-project.org/package=astsa. Accessed Jan 2018
Wang Z (2013) cts: An R package for continuous time autoregressive models via Kalman filter. J Stat Softw 53(5):1–19