

Forecasting counting and time statistics of compound Cox processes: a focus on intensity phase type process, deletions and simultaneous events

- Paula R. Bouzas, Nuria Ruiz-Fuentes, Carmen Montes-Gijón, Juan Eloy Ruiz-Castro
- Forecasting counting and time statistics of compound Cox processes: a focus on intensity phase type process, deletions and simultaneous events
- *Statistical Papers* (2021) 62:235–265
- DOI: <https://doi.org/10.1007/s00362-019-01092-0>





Forecasting counting and time statistics of compound Cox processes: a focus on intensity phase type process, deletions and simultaneous events

Paula R. Bouzas¹ · Nuria Ruiz-Fuentes² · Carmen Montes-Gijón¹ · Juan Eloy Ruiz-Castro¹

Received: 26 February 2018 / Revised: 24 October 2018 / Published online: 7 February 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Compound Cox processes (CCP) are flexible marked point processes due to the stochastic nature of their intensity. This paper states closed-form expressions of their counting and time statistics in terms of the intensity and of the mean processes. They are forecast by means of principal components prediction models applied to the mean process in order to reach attainable results. A proposition proves that only weak restrictions are needed to estimate the probability of a new occurrence. Additionally, the phase type process is introduced, which important feature is that its marginal distributions are phase type with random parameters. Since any non-negative variable can be approximated by a phase-type distribution, the new stochastic process is proposed to model the intensity process of any point process. The CCP with this type of intensity provides an especially general model. Several simulations and the corresponding study of the estimation errors illustrate the results and their accuracy. Finally, an application to real data is performed; extreme temperatures in the South of Spain are modeled by a CPP and forecast.

Keywords Compound Cox process · Estimation · Principal components prediction · Phase type process

Mathematics Subject Classification 60G25 · 60G51 · 60G55 · 62M20 · 62M99 · 90C15

✉ Paula R. Bouzas
paula@ugr.es

¹ Department of Statistics and Operations Research, University of Granada, Granada, Spain

² Department of Statistics and Operations Research, University of Jaén, Jaén, Spain

1 Introduction

The compound Cox process (CCP) is the natural extension of the compound Poisson process and the Cox process (CP) since it is a Poisson process with stochastic intensity and also a marked point process. More precisely, a CCP is a marked point process in which the point process is a CP and the marks of the events are independent and identically distributed as well as independent of the CP Snyder and Miller (1991).

The CCP has been used in several fields including geology, as in the classical application to earthquakes by Ogata (1998) or Gospodinov and Rotondi (2001); demography (Economou 2003); risk theory (Lin and Pavlova 2006); econometrics (Chertok et al. 2016); astrophysics (Si 2001), etc. However, although this is quite a general and flexible model, it is rare to find systematic studies of this type of process. On the contrary, studies on CP or CCP often use a limited variety of examples where the intensity or the mean process has an assumed stochastic structure that can fit the phenomena under consideration (Bouzas et al. 2002; Genaro and Simonis 2015). The main reason may lay in the difficulty of estimating the parameters of the CCP (mean or intensity processes) that characterize it, specially without any assumption on the process. Bouzas et al. (2006, 2007, 2010b, 2012) made the effort of studying a Cox process or a CCP only from the sample paths using Functional Data Analysis, which is a very powerful technique.

It is remarkable that a CCP generalizes some particular type of processes that are found in the literature as completely different or unrelated. A CCP is a CP if the mark space is denumerable with one mark $U = 1$ or a CP with random deletions if the mark space is $\{0, 1\}$ marking the deleted points with 0. Relaxing the property of orderliness, a CCP can represent a CP with simultaneous occurrences when the mark space is $\{1, 2, \dots\}$ (positive integers) indicating the number of occurrences. A multichannel CP can be considered a CCP in which the mark indicates the region where the point occurs. A time-space Cox process can also be modeled by a CCP in time where the mark on a point is continuously distributed and indicates the spatial position of the point.

It can be very important to study the occurrences of a CCP which marks are in a given subset of the mark space, for example, the points not deleted of a CP with random deletions, the points with more than one occurrence from a CP with simultaneous occurrences, the points in a given region within the mark space of a time-space Cox process, etc. Thus, this paper studies in depth the statistics, estimation and forecasting of this type of process.

Bouzas et al. (2007) proposed a methodology for estimating the mean process and some counting statistics of a CCP based on the representation theorems extended to CCP and assuming raw data as the only available information. The present paper tries to go further in the study of CCP in the same context of observation. It provides expressions for some other counting statistics and for some important time statistics. This is interesting not only to actually have the expressions but in order to address the inference of the CCP. Subsequently, these statistics are forecast in a future instant of time making use of functional data analysis techniques. The estimations presented are calculated in terms of the mean process, but they could be expressed in terms of the intensity process using its own estimation from Bouzas et al. (2012).

During the course of this paper another issue will also be addressed. A new kind of process will be defined in order to be proposed as a new model for the intensity process. This model achieves a high level of generalization and flexibility and is based on phase type distributions. This type of distributions, which were first introduced by Neuts (1975), are positive valued. Therefore, they are appropriate to model an intensity of any counting process. As this paper deals with a counting process which intensity is stochastic, a doubly stochastic phase type distribution is proposed to be the intensity for each instant of time. Doubly stochastic phase type distribution means that the distribution changes, i.e. it is not the same for each instant of time because its parameters are also random. Therefore, it is obtained a stochastic process which will be called *phase type process*. This proposal provides a compound Cox process with a stochastic intensity which marginals are phase type distributions. Then, it is a wider alternative for those models found in the literature which intensity is a linear combination of exponential distributions or similar.

The paper is structured in the following way. In Sect. 2, closed-form expressions for counting and time statistics of a CCP having specific marks are derived and the forecast of those statistics are calculated in Sect. 3. Section 4 proposes the CCP which intensity process is a phase type process. Section 5 presents simulations of the CCP for which the forecasting method has been applied in order to illustrate the results stated in the previous sections. The simulation examples are also necessary to be able to assess the theoretical results so, the estimations can be compared with the true or sample values within different types of CCPs and in several scenarios. These simulations are notable examples of CCPs, Cox processes with deletions, simultaneous occurrences or time-space Cox processes. Section 6 addresses the model and forecast of extreme maximum and extreme minimum values of temperatures in the South of Spain modeled by a CCP, obtaining the principal components prediction model for the mean process of the two corresponding counting processes and the forecast of the statistics. The main conclusions of the paper are gathered in Sect. 7.

2 Statistics of a CCP having specific marks

Let $\{N(t); t \geq t_0\}$, the number of points until instant t , be a CCP with intensity $\{\lambda(t, x(t)); t \geq t_0\}$ where $\{x(t); t \geq t_0\}$ is the information process and with i.i.d. marks associated to the arrival times and independent of the point process. The n -th arrival time will be denoted by w_n and its mark by u_n , which is a realization of the random variable U in \mathcal{U} with probability distribution P_u . Let $\{N(t, B); t \geq t_0\}$ be the counting process of the points from the former CCP whose marks are in $B \subseteq \mathcal{U}$.

In this section, using the representation theorems of a CCP, the expression for the basic counting statistics of $N(t, B)$ taking into account that the mark space is nondenumerable (analogous in the denumerable case) are provided in terms of the intensity and in terms of the mean process. The key idea of these theorems is that the process $\{N(t, B); t \geq t_0\}$ is also a Cox process with intensity $\int_B \lambda(\sigma, x(\sigma)) P_u(dU) d\sigma$ (Bouzas et al. 2007). Then, it is possible to derive its statistics extending the knowledge of the Cox process in terms of the intensity and in terms of the mean process when it is possible. That is interesting given that it is not usual to find in the literature the statistics

of a CCP explicitly written, specially the time statistics. Let us point out the the case of the original Cox process regardless the marks is included in the CCP just taking $B = U$.

From now on, in order not to complicate the expressions of the equations and being aware that the intensity and mean processes depend on the information process and that $t \geq t_0$, the intensity process of the CCP, $\lambda(t, x(t))$, will be denoted by $\lambda(t)$ and the mean process, $\Lambda(t, x(t))$, by $\Lambda(t)$.

2.1 Counting statistics

The counting statistics of a counting process are based on its probability mass function; therefore, its expression is basic. Nevertheless, other counting statistics are more or less important according to the framework, survival analysis, reliability, counting point of view, etc. The probability mass function, the mean and the mode (Bouzas et al. 2010b) are only included because they are basic statistics and also to present the set of interesting counting statistics. The novelty of this subsection is to take into account the expression of the cumulative distribution function as it plays an important role in reliability theory and which will be used later on.

• The *probability mass function* (pmf) of the CCP, the probability of having n points with mark in B until t , is

$$\begin{aligned} P [N(t, B) = n] &= E_x \left[\frac{1}{n!} \left(\int_{t_0}^t \int_B \lambda(\sigma) P_u(dU) d\sigma \right)^n \exp \left[- \int_{t_0}^t \int_B \lambda(\sigma) P_u(dU) d\sigma \right] \right] \\ &= E_x \left[\frac{1}{n!} \left(\Lambda(t) \int_B P_u(dU) \right)^n \exp \left[- \Lambda(t) \int_B P_u(dU) \right] \right]. \end{aligned} \quad (1)$$

where E_x is the expectation conditional to the information process (t).

• The *mean* of the CCP, that is the expected number of points in B until t , is expressed as

$$\begin{aligned} E [N(t, B)] &= E_x \left[\int_{t_0}^t \int_B \lambda(\sigma) P_u(dU) d\sigma \right] \\ &= E_x \left[\Lambda(t) \int_B P_u(dU) \right] = E_x [\Lambda(t)] \int_B P_u(dU). \end{aligned}$$

• The *mode* of the CCP with marks in B denoted by $n_{\max}^B(t) \in \mathbb{N}$, is the most probable number of occurrences in B until a certain time point t so $n_{\max}^B(t) = \{n \mid \max P [N(t, B) = n]\}$. As proved in Bouzas et al. (2010b),

$$n_{\max}^B(t) = \begin{cases} \Lambda(t) \int_B P_u(dU) - 1, & \Lambda(t) \int_B P_u(dU) \in \mathbb{N} \\ \left\{ \begin{array}{l} \text{int} \left(\Lambda(t) \int_B P_u(dU) - 1 \right) \\ \text{or} \\ \text{int} \left(\Lambda(t) \int_B P_u(dU) - 1 \right) + 1 \end{array} \right\}, & \Lambda(t) \int_B P_u(dU) \notin \mathbb{N} \\ 0, & \Lambda(t) \int_B P_u(dU) < 1 \end{cases} \quad (2)$$

note: $int =$ integer part.

- The *cumulative distribution function* of $N(t, B)$ is the probability of having observed less than a given number of occurrences in B before a fixed instant of time. Its expression, using the pmf, is clearly the following

$$\begin{aligned}
 P [N(t, B) < n] &= P [N(t, B) \leq n - 1] \\
 &= E_x \left[\sum_{i=0}^{n-1} \frac{1}{i!} \left(\Lambda(t) \int_B P_u(dU) \right)^i \exp \left[-\Lambda(t) \int_B P_u(dU) \right] \right] \\
 &= E_x \left[\frac{\Gamma \left(n, \Lambda(t) \int_B P_u(dU) \right)}{\Gamma(n)} \right] \tag{3}
 \end{aligned}$$

where $\Gamma(n) = \int_0^\infty \sigma^{n-1} e^{-\sigma} d\sigma$ is the usual gamma function and $\Gamma(n, y) = \int_y^\infty \sigma^{n-1} e^{-\sigma} d\sigma$ is the incomplete gamma function.

2.2 Time statistics

A key theoretical novelty of this paper is the derivation of the expressions for some important time statistics of a CCP with marks in a given subset of the mark space \mathcal{U} . Once more, one of the keys to derive them, is the fact that $N(t, B)$ is a Cox process itself.

- The *survival or reliability function* of $N(t, B)$ is the probability that the n -th point happens after a given time T which is connected with the cumulative distribution function the following way

$$P^B [W_n > T] = P [N(T, B) < n] \tag{4}$$

where P^B denotes the probability in a subset B . Therefore, its expression is the one of Eq. (3).

- Due to the conditioning method (Snyder and Miller 1991), the *conditional probability density* of the n -th occurrence time of $N(t, B)$, the probability of having a new point in B after having occurred the $(n - 1)$ -th point, is

$$\begin{aligned}
 &P_{w_n/w_{n-1}}^B (W_n/W_{n-1}) \\
 &= E_x \left[\lambda(W_n) \int_B P_u(dU) \exp \left[-\int_{W_{n-1}}^{W_n} \int_B \lambda(\sigma) P_u(dU) d\sigma \right] \right] \\
 &= E_x \left[\lambda(W_n) \int_B P_u(dU) \exp \left[-(\Lambda(W_n) - \Lambda(W_{n-1})) \int_B P_u(dU) \right] \right]. \tag{5}
 \end{aligned}$$

- The *probability of having a new specific occurrence within an interval of time* (the probability of having the following occurrence with mark in B in the time interval $[W_{n-1}, T)$) is also interesting

$$\begin{aligned}
 P_{w_n/w_{n-1}}^B(W_n < T/W_{n-1}) &= E_x \left[1 - \exp \left[- \int_{W_{n-1}}^T \int_B \lambda(\sigma) P_u(dU) d\sigma \right] \right] \\
 &= E_x \left[1 - \exp \left[- (\Lambda(T) - \Lambda(W_{n-1})) \int_B P_u(dU) \right] \right]. \quad (6)
 \end{aligned}$$

• As a consequence, the *survival or reliability conditional function* can be written as

$$P_{w_n/w_{n-1}}^B(W_n > T/W_{n-1}) = E_x \left[\exp \left[- (\Lambda(T) - \Lambda(W_{n-1})) \int_B P_u(dU) \right] \right]. \quad (7)$$

Taking into account the relation between the occurrence times and the interarrival times, $w_n = w_{n-1} + t_n$, t_n , the n -th interarrival time, it is straightforward to derive similar densities and statistics in terms of the latter. Let us give an example using Eq. (5).

• The *forward-occurrence density* is identical to the conditional density of t_n and it is expressed as

$$\begin{aligned}
 p_{t_n/w_{n-1}}^B(T/W_{n-1}) &= p_{w_n/w_{n-1}}^B(W_n + T/W_{n-1}) \\
 &= E_x \left[\lambda(W_n + T) \int_B P_u(dU) \exp \left[- (\Lambda(W_n + T) - \Lambda(W_{n-1})) \int_B P_u(dU) \right] \right].
 \end{aligned}$$

A common feature of the statistics studied in both Sects. 2.1 and 2.2 is that they all are expressed in terms of the intensity or the mean of the Cox process regarding the marks. This will allow us to use any further study on that intensity or mean for any particular $N(t, B)$. Furthermore, predicting the intensity or the mean, the statistics will be also predicted as it will be done in the following section.

3 Forecasting the statistics of the CCP

As pointed out above, the forecast of the intensity or the mean process would allow to forecast the statistics of any $N(t, B)$. Different techniques of forecasting a stochastic process could be contemplated to apply to the intensity or the mean. An aim of this paper is to forecast the CCP assuming that the observed sample paths of the process is the only available information. Therefore, the framework chosen for the estimation and further forecast is functional principal components analysis (FPCA). Bouzas et al. (2006, 2012) proposed corresponding *ad hoc* FPCA's to estimate the intensity and the mean processes of a Cox process. It is well known that principal components prediction (PCP) developed by Aguilera et al. (1997) within FPCA is a useful and precise methodology. The PCP of a stochastic process provides a continuous prediction of the process in a future time interval from discrete observations of the process in the past. The method to perform a PCP model is based on a double FPCA of the process (FPCA of the past and of the future) and the correlation between the principal

components of both. Bouzas et al. (2010b) used this powerful technique to derive a forecasting method for the CCP which will provide estimators for its statistics.

In this paper, the forecast of a CCP will be done in terms of the mean process using the method mentioned (Bouzas et al. 2010b). Let us assume that sample paths of the Cox process have been observed in a discrete set of time points in $[T_0, T_1]$ and a forecast in a time point $s \in (T_1, T_2]$ is needed with $T_0 < T_1 < s \leq T_2$. The method consists on two main steps. Firstly, estimate sample paths of the mean process in a discrete set of time points (each one from several sample paths of the Cox process).

Secondly, obtaining a continuous estimation of the mean process in the past and in the future by means of FPCA. The estimations are denoted as follows

$$\begin{aligned} \widehat{\Lambda^1(t)} &= \Lambda^{q_1}(t) = \mu_{\Lambda}^1(t) + \sum_{j=1}^{q_1} \xi_j f_j(t); & t \in [t_0 = T_0, T_1] \\ \widehat{\Lambda^2(s)} &= \Lambda^{q_2}(s) = \mu_{\Lambda}^2(s) + \sum_{j=1}^{q_2} \eta_j g_j(s); & s \in (T_1, T_2] \end{aligned} \tag{8}$$

where μ_{Λ}^1 and μ_{Λ}^2 are the mean functions of Λ in the past and future intervals, ξ_i are the principal components of the past, $g_j(s)$, the principal functions of the future and where q_1 and q_2 are chosen so that an important percentage of variance is explained. The choice is sometimes made to compromise the amount of explained variance attained with the complication of the final model. In this paper, the percentage of explained variance imposed to the model is 99%.

Afterwards, the correlation between the past and the future has to be established. Denoting by $\tilde{\eta}_j^{p_j} = \sum_{i=1}^{p_j} b_i^j \xi_i$ the estimator of η_j , $j = 1, \dots, q_2$ in terms of the p_j principal components ξ_j , Eq. (8) can be rewritten so that the prediction of the mean process in a future interval $(T_1, T_2]$ given the past interval $[T_0, T_1]$ is calculated by

$$\tilde{\Lambda}^{q_2}(s) = \mu_{\Lambda}^2(s) + \sum_{j=1}^{q_2} \left(\sum_{i=1}^{p_j} b_i^j \xi_i \right) g_j(s); \quad s \in (T_1, T_2] \tag{9}$$

In order to indicate the number of principal components of the past and future considered in the final estimation, the PCP model will be denoted as $PCP(q_2; p_1, \dots, p_j)$.

Having finished the prediction of the mean process, the forecasting expressions of the mean and the mode of the CCP using Eq. (9) can be found in Bouzas et al. (2010b).

- *Mean* of $N(T_1, s, B)$, s in the interval $(T_1, T_2]$, where $N(T_1, s, B)$ is the number of points between T_1 and s ,

$$E [N(T_1, s, B)] \simeq E_x \left[\Lambda(s) \int_B P_u(dU) \right] = \mu_{\Lambda}^2(s) \int_B P_u(dU). \tag{10}$$

- *Mode* of $N(T_1, s, B)$, s in $(T_1, T_2]$. The number of occurrences with maximum probability until time s is

$$n_{\max}^B(s) \simeq \begin{cases} \tilde{\Lambda}^{q_2}(s) \int_B P_u(dU) - 1, & \tilde{\Lambda}^{q_2}(s) \int_B P_u(dU) \in \mathbb{N} \\ \left\{ \begin{array}{l} \text{int} \left(\tilde{\Lambda}^{q_2}(s) \int_B P_u(dU) - 1 \right) \\ \text{or} \\ \text{int} \left(\tilde{\Lambda}^{q_2}(s) \int_B P_u(dU) - 1 \right) + 1 \end{array} \right. & \tilde{\Lambda}^{q_2}(s) \int_B P_u(dU) \notin \mathbb{N} \\ 0, & \tilde{\Lambda}^{q_2}(s) \int_B P_u(dU) < 1 \end{cases} \quad (11)$$

From here on, this paper presents the forecasting expressions of other statistics, the cumulative distribution function within the counting statistics and all the time statistics from Sect. 2.

- The *cumulative distribution function* of the process $N(s, B)$ with $s \in (T_1, T_2]$ is the probability of having observed less than a given number of occurrences in B before the instant of time s . By means of Eqs. 3 and 9 and denoting $N(T_1, B) = n^{T_1}$, it is obtained

$$P [N(s, B) < n] \simeq \begin{cases} 0, & n \leq n^{T_1} \\ E_x \left[\frac{\Gamma(n - n^{T_1} - 1, \Lambda(s) \int_B P_u(dU))}{\Gamma(n - n^{T_1} - 1)} \right], & n > n^{T_1} \end{cases}$$

where $\Gamma(n)$ is the gamma function and $\Gamma(n, x)$ is the incomplete gamma function defined previously.

- The *probability of having a new specific occurrence within the interval* $(W_{n-1} = T_1, T_2]$, i.e. the probability of Eq. (6).

The plain application of the forecasting of the mean given in Eq. (9) provides an intractable expression of this statistic as it is usual dealing with a CP. Despite that, if very weak restrictions are verified, it is possible to derive an explicit expression able to be implemented. Let us present it in the following proposition.

Proposition 1 *Given the estimation of the mean process of $N(t)$ in Eq. (9), let us assume that the p.c.'s, ξ_i , used to estimate the p.c.'s of the future, $\eta_j, j = 1, \dots, q_2$ are independent, normally distributed (or $k \geq 30$) and their distributions are (or can be estimated as) $\mathcal{N}(0, \sigma_i)$ where the σ_i^2 are the variances previously estimated in the corresponding FPCA.*

Fixing $s \in (W_{n-1} = T_1, T_2]$, the probability of having the following occurrence with mark in B within the interval $(W_{n-1} = T_1, s)$ is

$$P_{w_n/w_{n-1}}^B(W_n < s/W_{n-1}) \simeq 1 - \exp \left[\left(\Lambda^{q_1}(W_{n-1}) - \mu_{\Lambda}^2(s) \right) \int_B P_u(dU) \right] \\ \times \exp \left[\frac{\left(\int_B P_u(dU) \right)^2}{2} \sum_{j=1}^{q_2} \left(\sum_{i=1}^p \left(\sigma_i b_i^j \right)^2 \right) (g_j(s))^2 \right] \quad (12)$$

where p is the largest $p_j, \forall j$.

Proof In order to simplify the notation in the proof, let us denote $a = \int_B P_u(dU)$ and $b = \Lambda^{q_1}(W_{n-1})$ both constants. Then, Eq. (6) is rewritten like this

$$P_{w_n/w_{n-1}}^B(W_n < s/W_{n-1}) = E_x \left[1 - \exp \left[-(\Lambda(s) - \Lambda(W_{n-1})) \int_B P_u(dU) \right] \right] \\ = E_x \left[1 - \exp [-a(\Lambda(s) - b)] \right] = 1 - E_x [e^{-a\Lambda(s)} e^{ab}] = 1 - e^{ab} M_{\Lambda(s)}(-a)$$

where $M_{\Lambda(s)}(\sigma)$ is the generating function of $\Lambda(s)$.

Using the prediction of Eq. (9), it is obtained

$$P_{w_n/w_{n-1}}^B(W_n < s/W_{n-1}) \simeq 1 - e^{ab} E \left[\exp \left[-a \left(\mu_{\Lambda}^{(2)}(s) + \sum_{j=1}^{q_2} \left(\sum_{i=1}^{p_j} b_i^j \xi_i \right) g_j(s) \right) \right] \right]. \tag{13}$$

Substituting a and b by their values, the expression of the probability would be achieved but, as mentioned above, it is intractable.

In order to obtain a practical expression for this probability, let us first introduce the following notation. Let us call p to the largest $p_j, \forall j$. In other words, p is the largest subindex of the p.c.'s from the past used to estimate the p.c.'s of the future. Therefore, Eq. (13) can be written and transformed as the following

$$1 - e^{ab} E \left[\exp \left[-a \left(\mu_{\Lambda}^{(2)}(s) + \sum_{j=1}^{q_2} \left(\sum_{i=1}^p b_i^j \xi_i \right) g_j(s) \right) \right] \right] \\ = 1 - e^{ab} e^{-a\mu_{\Lambda}^{(2)}(s)} E \left[\exp \left[-a \sum_{j=1}^{q_2} \left(\sum_{i=1}^p b_i^j \xi_i \right) g_j(s) \right] \right] \\ = 1 - e^{ab} e^{-a\mu_{\Lambda}^{(2)}(s)} E \left[\exp \left[\sum_{i=1}^p \xi_i \left(\sum_{j=1}^{q_2} -a b_i^j g_j(s) \right) \right] \right] \\ = 1 - e^{ab} e^{-a\mu_{\Lambda}^{(2)}(s)} E \left[\prod_{i=1}^p \prod_{j=1}^{q_2} \exp \left[-a b_i^j \xi_i g_j(s) \right] \right]. \tag{14}$$

It should be noticed that some ξ_i may be missing because $b_i^j = 0, \forall j$, which means that they do not appear in the estimation of the p.c.'s of the future. If the remaining p.c.'s, ξ_i , are independent, the expression of Eq. (14) follows this way

$$P_{w_n/w_{n-1}}^B(W_n < s/W_{n-1}) \simeq 1 - e^{ab} e^{-a\mu_{\Lambda}^{(2)}(s)} \left[\prod_{i=1}^p \prod_{j=1}^{q_2} M_{\xi_i} \left(-a b_i^j g_j(s) \right) \right].$$

Finally, let us assume that the remaining p.c.'s ξ_i are distributed as $\mathcal{N}(0, \sigma_i^2)$ where σ_i^2 have been estimated in the corresponding FPCA. If this does not hold but the sample size was greater than 30, we can approximate their distributions by those Normals. Then, as the generating function of ξ_i is $M_{\xi_i}(u) = \exp\left[\frac{\sigma_i^2 u^2}{2}\right]$, the equation can be written

$$\begin{aligned}
 &P_{w_n/w_{n-1}}^B(W_n < s/W_{n-1}) \\
 &\simeq 1 - \exp\left[\left(\Lambda(W_{n-1}) - \mu_\Lambda^2(s)\right) \int_B P_u(dU)\right] \left[\prod_{i=1}^p \prod_{j=1}^{q_2} \exp\left[\frac{(\sigma_i a b_i^j g_j(s))^2}{2}\right]\right] \\
 &= 1 - \exp\left[\left(\Lambda(W_{n-1}) - \mu_\Lambda^2(s)\right) \int_B P_u(dU)\right] \left[\prod_{i=1}^p \prod_{j=1}^{q_2} \exp\left[\frac{(\sigma_i b_i^j g_j(s) \int_B P_u(dU))^2}{2}\right]\right]. \tag{15}
 \end{aligned}$$

It is then straightforward to obtain the expression stated in the proposition. □

It is remarkable that it has been found out from the simulations and several examples of real data (Bouzas et al. 2010a) that the p.c.'s usually verify to have a joined Normal distribution so, as they are uncorrelated, they become independent as well as they are normally distributed. Thus, the restrictions imposed above are not difficult to fulfill.

• An obvious consequence of Proposition 1 is that, in the same scenario, the **survival conditional function in s** is expressed as

$$\begin{aligned}
 &P_{w_n/w_{n-1}}^B(W_n > s/W_{n-1}) \\
 &\simeq \exp\left[\left(\Lambda^{q_1}(W_{n-1}) - \mu_\Lambda^2(s)\right) \int_B P_u(dU)\right] \\
 &\quad \times \exp\left[\frac{(\int_B P_u(dU))^2}{2} \sum_{j=1}^{q_2} \left(\sum_{i=1}^p (\sigma_i b_i^j)^2\right) (g_j(s))^2\right]. \tag{16}
 \end{aligned}$$

4 Phase type process

The intensity of any counting process needs to be non-negative valued so, exponential, gamma, Weibull or some generalizations of them are the most common distributions to model it (see, e.g., Asha and Nair 2010; Russell and Engle 2010; Chen and Hall 2013; Lefebvre and Belsalma 2015; Bieniek and Goroncy 2017; Sephefirar and Yarahmadian 2017; Singh and Tripathi 2018). These same references also prove that the selection of the distribution is not simple and it is important to find new models which are more adaptable to real data. The purpose of this section is to propose another and even more general model suitable for the intensity of a doubly stochastic counting process.

Phase type (PH) distributions were first introduced by Neuts (1975) and described in detail by Neuts (1981). Another recommended and recent reference is He (2014). This type of distributions has been mainly applied in areas such as survival, reliability and queuing theory. Recently, Ruiz-Castro (2016a, b) considered PH distributions to model the behavior of complex reliability systems by counting several events.

This distribution class is a generalization of the exponential one with a matrix algebraic structure, by means of which the main measures associated with a reliability system can be modeled and calculated in an algorithmic and well-structured form, thus simplifying computational implementation. Several well known distributions such as degenerate distribution, Erlang, Coxian, hyperexponential, hypoexponential, etcetera, are particular cases of PH distributions (Neuts 1981; He 2014; Tank and Eryilmaz 2015).

One of the main features of PH distributions is their good properties; a notable property is that, not only known distributions are phase type but, in the continuous case, this class is dense in the non-negative class of probability distributions, while in the discrete case, any probability distribution with finite support is a PH distribution (Neuts 1975). Therefore, given a non-negative distribution, this one can be approximated as much as wanted. Consequently, these distributions are perfect candidates to model the intensity of a counting process. It is important to remark that this type of distributions is generally associated with time between failures in reliability theory; however, their use in this proposal is completely different.

Another aspect to deal with is the stochastic nature of the intensity of a doubly stochastic counting process (it is not only an unchanging random variable). For this reason, a PH distribution with random parameters, let us name it as *phase type process*, is proposed to model the intensity of the counting process. Thus, the intensity process could follow a different phase type distribution at each instant of time. Let us define this new process using the notation of the PH distributions.

Definition 1 A *phase type process* is a stochastic process with the representation $(\boldsymbol{\pi}, \mathbf{T})$ where

1. $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)$ and $\mathbf{T} = (T_{ij})_{p \times p}$
2. with any π_i and/or T_{ij} random variables
3. such that for each realization of the random variables, the distribution is a PH distribution.

The last item of the definition imposes on $\boldsymbol{\pi}$ and \mathbf{T} all the restrictions of the PH distributions, i.e. $\sum_{i=1}^p \pi_i = 1$, negative diagonal elements of \mathbf{T} and non-negative T_{ij} , $i \neq j$, non-positive sum of the elements of each row and being invertible.

Using a phase type process as a model for an intensity, a very general and flexible intensity process is obtained and consequently, a general doubly stochastic counting process is derived as well. It is notable that any particular type of counting process has not been assumed throughout this section. Thus, this new model could be used as intensity for any counting or point process and in any knowledge field. In order to continue within the framework of this paper, the counting process taken into account from now on is the CCP.

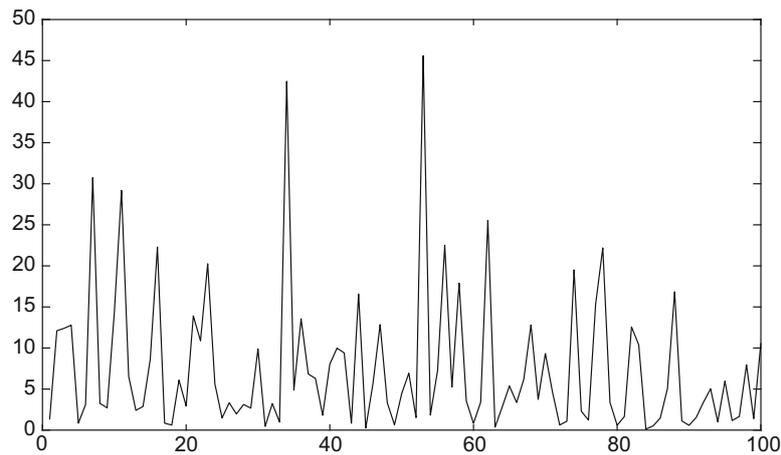


Fig. 1 Sample path of the intensity as a phase type process

Let us give an example of a CCP with a phase type process as intensity. The intensity process of this example has the following representation

$(\boldsymbol{\pi}, \mathbf{T}(\beta_1, \beta_2, \beta_3))$ where

$$\boldsymbol{\pi} = (0.2, 0.4, 0.4) \quad \text{and} \quad \mathbf{T}(\beta_1, \beta_2, \beta_3) = \begin{pmatrix} -1/\beta_1 & 0 & 0 \\ 0 & -1/\beta_2 & 0 \\ 0 & 0 & -1/\beta_3 \end{pmatrix}$$

with $\beta_1 \rightsquigarrow \mathcal{B}(10, 0.8) + 1$, $\beta_2 \rightsquigarrow \mathcal{B}(20, 0.5) + 1$ and $\beta_3 \rightsquigarrow \mathcal{B}(1, 0.5) + 1$ (17)

and where \mathcal{B} denotes the Binomial distribution.

Figure 1 shows a sample path of this intensity process given in the representation above (17).

Let us consider that the marks of this example of CCP are two-dimensional, following the normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean $\boldsymbol{\mu} = (4, 8)$ and covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} 1.5 & 0.7 \\ 0.7 & 1.5 \end{pmatrix}$.

If the set of marks of interest within the mark space is chosen to be $B = \{(x_1, x_2); 2 \leq x_1 \leq 4, 7 \leq x_2 \leq 9\}$ then, the probability of having a mark in this set is $\int_B P_u(dU) = 0.2717$. Figure 2 represents the mark space with its distribution and the set B .

Having defined the intensity process and the mark space, the CCP is characterized. Taking into account the set of marks B , the counting process extracted from the CCP is defined as the CP formed by the points with marks in B . Intuitively, the initial sample paths lose points when considering only the points with marks in B . Figure 3 shows three sample paths of the CP regardless the marks (initial sample paths) and the corresponding sample paths with marks in the set B .

This example can be interpreted as a time-space CP where the mark of a point represents its position in the plane. It will be used as a simulation case in the next section.

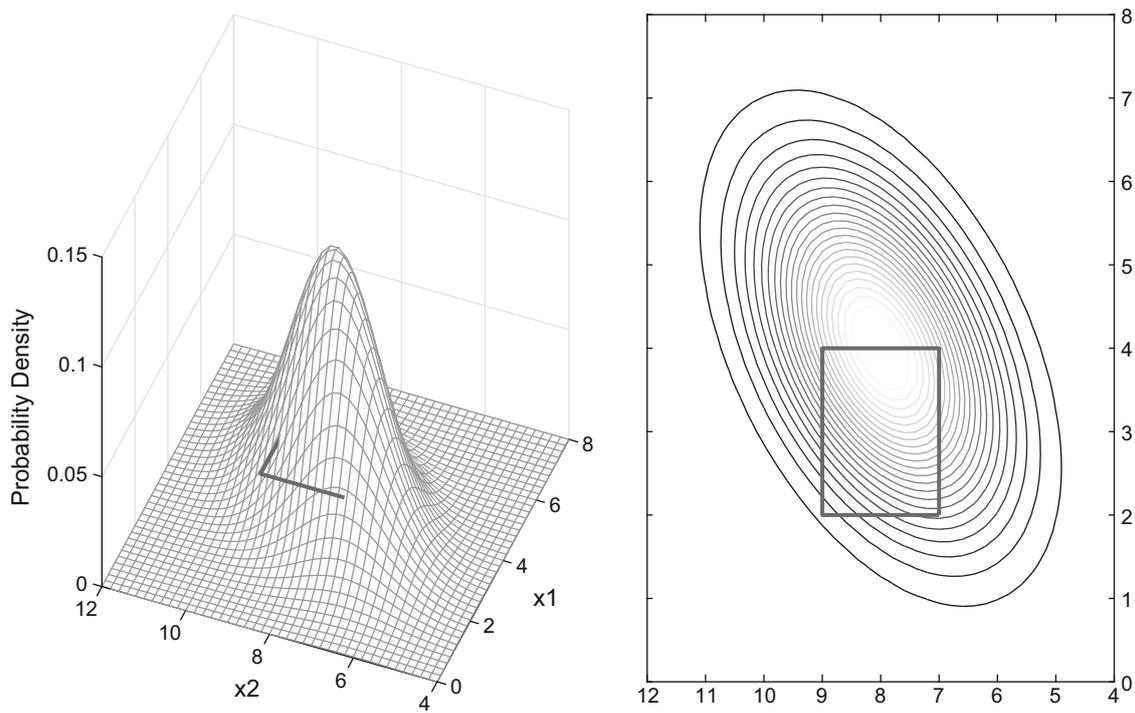


Fig. 2 Graphic of the mark space with its distribution and the set of marks B

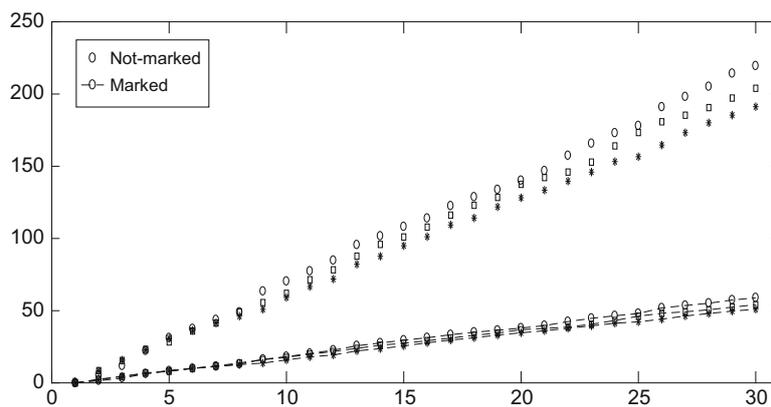


Fig. 3 Three sample paths of the CP regardless the marks (without dashes) and the corresponding ones with marks in the set B (symbols joined by dashes)

5 Simulations

This section illustrates the theoretical calculations presented in the former sections by means of several simulations. As usual, they will also allow to carry out the assessment of the results.

The simulations will consist on a number of CCP with different types of intensity processes and mark distributions. In the literature, it is usual to find papers studying counting processes which could be modeled as a special class of marked process. For example, multi-channel processes can be thought as marked processes whose marks are the channels. The simulations chosen for the present section illustrate some interesting cases such as a CCP with deletions, simultaneous occurrences or time-space CP's.

The selection of the intensities and mark spaces for the first and second simulations has been inspired by examples found in articles from different knowledge areas. The most usual models have gamma or geometric Brownian motion intensities and lognormal, binomial and gammas as mark distributions. These type of models can be found in Gospodinov and Rotondi (2001), Barta et al. (2005), Park and Padgett (2005) and, in general, in the articles provided in the References section. The Binomial as mark distribution allows us to illustrate the case of possible deletions of points, an interesting feature to take into account. Additionally, the problem of simultaneous sources of point occurrences is also interesting in many fields (see Snyder and Miller 1991; Dousse et al. 2005; Greenberg et al. 2008). On the other hand, the values of the parameters of the distributions have not been chosen due to specific reasons, they are mere examples.

Each simulation case consists on $k = 100$ sample paths in $[0, 20]$ of a CCP. As explained in Sect. 3, the mean process has been estimated in the future interval $(T_1, T_2]$ and it has been used to estimate the mean and the mode in an instant of time $s \in (T_1, T_2]$, as well as the probability of having a new point before s . Several values of T_1 and s will be chosen to illustrate the calculation of the mentioned statistics.

The simulations and all the calculations were implemented in MATLAB R2016b and all the figures were created with the same program.

Simulation case 1 The first simulation is a CCP whose intensity process is a $\text{Gamma}(4, 0.5)$ with marks distributed as a $\text{Binomial}(1, 0.6)$ and $B = \{u; u = 1\}$ so $\int_B P_u(dU) = p(1) = 0.6$. This CCP can be interpreted as a CP with deletions (points with mark $u = 0$).

Applying the corresponding PCP technique (see Eq. 9) and fixing the future interval as $(T_1 = 5, T_2 = 20]$, the mean process has been modeled as a $\text{PCP}(q_2; p_1, \dots, p_j) \equiv \text{PCP}(6; 4, 4, 0, 4, 3, 4)$. Initially, this model implies that 6 components of the future are used for the estimation and each one is explained with 4, 4, 0, 4, 3 and 4 of the past, respectively. However, examining the parameters of the model in detail, it is remarkable that the third future component is not actually explained by any from the past. Consequently, only 5 components of the future are kept in the model. Therefore, having computed the coefficients and taking into account the null and non-null coefficients, the mean process for $s \in (T_1 = 5, T_2 = 20]$ can be written the following way

$$\tilde{A}^6(s) = \mu_{\Lambda}^2(s) + \begin{pmatrix} -1.024 & 0.962 & -1.654 \\ -2.903 & 1.488 & -3.965 \\ -0.425 & 0.162 & -0.263 \\ -0.061 & 0.115 & 0 \\ -0.457 & 0.212 & -0.404 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_3 \\ \xi_4 \end{pmatrix} (g_1(s), g_2(s), g_4(s), g_5(s), g_6(s)).$$

The same type of calculations should be done if another future interval is chosen. Therefore, changing the time T_1 , another PCP model would be obtained. As $T_1 \in (T_0, T_2)$, it is not possible to show all the cases, however, three different values will

be used to calculate the estimations. It is interesting to point out that if the instant T_1 is very close to T_0 , the past interval is very small and so, the information available in order to estimate and forecast the future, might not be enough to obtain good results. On the other hand, having a fixed T_1 , the further the estimation time s is, the biggest estimation error can be expected. Then, the estimation error tends to increase as s gets closer to the end of the future interval, T_2 .

Table 1 shows the results of the estimations of the mean (see Eq. 10), mode (see Eq. 11) and the probability of having a new point before a fixed instant of the future (see Proposition 1) for this example. In order to be more illustrative, the estimations are calculated for three different scenarios, when the future interval is considered $(T_1 = 5, T_2 = 20]$, $(T_1 = 10, T_2 = 20]$ or $(T_1 = 15, T_2 = 20]$. Furthermore, the estimations are calculated for several instants s within the future interval.

Simulation case 2 The second simulation is a CCP formed by four randomly simultaneous Poisson processes with intensities 5, 4, 6 and 3, respectively. Each of these processes is operating with probability 0.5 at any instant of time. The marks are distributed as a Binomial(10,0.6) and $B = \{u; 6 \leq u \leq 8\}$ so $\int_B P_u(dU) = 0.5867$. This CCP can be interpreted as the mixture of Poisson processes with simultaneous occurrences where the mark at an instant of time is the number of occurrences at that moment.

In this case, fixing again the future interval as $(T_1 = 5, T_2 = 20]$, the mean has been modeled as a $PCP(q_2; p_1, \dots, p_j) \equiv PCP(9; 4, 4, 0, 4, 4, 4, 4, 0, 4)$ (see Eq. (9)). Reading carefully the notation and discarding the components with none explicative components from the past, finally, only seven components of the future are used. Having computed the coefficients, the mean process for $s \in (T_1 = 5, T_2 = 20]$ is written as

$$\tilde{\Lambda}^9(s) = \mu_{\Lambda}^2(s) + \begin{pmatrix} -0.696 & -4.007 & -3.017 & -2.367 \\ 1.876 & 1.099 & 1.427 & 0.650 \\ -0.988 & 0 & -0.365 & -0.327 \\ -0.194 & -0.387 & 0 & 0.004 \\ 0.021 & 0.134 & 0 & -0.134 \\ -0.594 & 0.093 & 0 & 0.107 \\ 0.019 & 0 & 0 & 0.364 \end{pmatrix} \times \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} (g_1(s), g_2(s), g_4(s), g_5(s), g_6(s), g_7(s), g_9(s)).$$

Again, the same type of calculations should be done if any other future interval is chosen.

Table 2 shows the results of the estimations for this simulation case using Eqs. (10), (11) and Proposition 1. As in the first case, three different intervals of the future are considered.

Table 1 Estimation of the mean, mode and the probability of having a new point before s (Prob) for *simulation 1*

T_1	s	Mean	Mode	Prob	
5	6	7.217	6	0.511	
	7	8.438	7	0.865	
	8	9.651	8	0.959	
	9	10.865	9	0.988	
	10	12.072	11	0.996	
	11	13.300	12	0.998	
	12	14.496	13	0.999	
	13	15.685	14	0.999	
	14	16.900	15	0.999	
	15	18.109	17	0.999	
	16	19.286	18	0.999	
	17	20.492	19	0.999	
	18	21.684	20	0.999	
	19	22.884	21	0.999	
	20	24.061	23	0.999	
	10	11	13.300	12	0.661
		12	14.496	13	0.904
		13	15.685	14	0.971
		14	16.900	15	0.991
15		18.109	17	0.997	
16		19.286	18	0.999	
17		20.492	18	0.999	
18		21.684	20	0.999	
19		22.884	21	0.999	
20		24.061	23	0.999	
15	16	19.286	18	0.579	
	17	20.492	19	0.886	
	18	21.684	20	0.965	
	19	22.884	21	0.989	
	20	24.061	23	0.996	

Simulation case 3 The third simulation is the case presented in Sect. 4 used as an example of CCP whose intensity is the phase type process given in Eq. (17), with the marks following the same two-dimensional Normal distribution and choosing the same set B .

Following the same steps as in the previous simulations, in this case, the mean process has been modeled as a $PCP(9; 4, 4, 0, 0, 4, 0, 0, 0, 4)$ [see Eq. (9)] for $s \in (T_1 = 5, T_2 = 20]$. Similarly as in the previous simulation cases, taking into account the components of the future that have correlations with some of the past, this model will finally use four components of the future and the expression is

Table 2 Estimation of the mean, mode and the probability of having a new point before s (Prob) for simulation 2

T_1	s	Mean	Mode	Prob	
5	6	31.882	31	0.983	
	7	37.194	36	0.999	
	8	42.420	41	0.999	
	9	47.738	46	0.999	
	10	52.989	52	0.999	
	11	58.271	57	0.999	
	12	63.578	62	1	
	13	68.879	67	1	
	14	74.226	73	1	
	15	79.495	78	1	
	16	84.770	83	1	
	17	90.085	89	1	
	18	95.436	94	1	
	19	100.715	99	1	
	20	105.875	104	1	
	10	11	58.271	57	0.999
		12	63.578	62	0.999
		13	68.879	67	0.999
		14	74.226	73	0.999
		15	79.495	78	0.999
16		84.770	83	0.999	
17		90.085	89	1	
18		95.436	94	1	
15	19	100.715	99	1	
	20	105.875	104	1	
	16	84.770	83	0.999	
	17	90.085	89	0.999	
	18	95.436	94	0.999	
	19	100.715	99	0.999	
	20	105.875	104	0.999	

$$\tilde{\Lambda}^9(s) = \mu_{\Lambda}^2(s) + \begin{pmatrix} 0 & 0 - 0.449 - 0.525 \\ 0.615 - 0.501 & 0.073 & 0.350 \\ -0.013 - 0.123 & 0.161 & 0.312 \\ 0 & 0 & 0.262 & 0.251 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} (g_1(s), g_2(s), g_5(s), g_9(s)).$$

Table 3 shows the results of the estimations using Eqs. (10), (11) and Proposition 1 for this case using the three different future intervals chosen as examples.

Table 3 Estimation of the mean, mode and the probability of having a new point before s (Prob) for *simulation 3*

T_1	s	Mean	Mode	Prob	
5	6	11.041	10	0.831	
	7	12.941	11	0.977	
	8	14.769	13	0.996	
	9	16.629	15	0.999	
	10	18.452	17	0.999	
	11	20.265	19	0.999	
	12	22.183	21	0.999	
	13	24.085	23	0.999	
	14	25.911	24	0.999	
	15	27.751	26	0.999	
	16	29.559	28	0.999	
	17	31.259	30	0.999	
	18	33.285	32	0.999	
	19	35.166	34	0.999	
	20	36.936	35	0.999	
	10	11	20.265	19	0.583
		12	22.183	21	0.972
		13	24.085	23	0.995
		14	25.911	24	0.999
		15	27.751	26	0.999
16		29.559	28	0.999	
17		31.359	30	0.999	
18		33.285	32	0.999	
19		35.166	34	0.999	
20		36.936	35	0.999	
15	16	29.559	28	0.95	
	17	31.359	30	0.970	
	18	33.285	32	0.996	
	19	35.166	34	0.999	
	20	36.936	35	0.999	

5.1 Assessment of results

The estimations obtained for the simulation examples used the theoretical results of previous sections. Let us assess the accuracy of the calculations.

The estimation of the mean of $N(t, B)$ in a fixed future instant s , Eq. (10), is the deterministic part of the estimation of the mean process given in Eq. (9). It has been obtained by a PCP model within the functional data analysis techniques which are well known and already proved to be a good methodology for estimation. Therefore, it is not needed to prove that this estimation is accurate. However, it is illustrative to compare the estimations of the mean given in Tables 1, 2 and 3 with the sample mean

Table 4 Comparison of the estimated mean and the sample mean of $N(s, B)$ from all the simulated sample paths for *simulation 1*

s	Estimation of the mean			Sample mean
	$T_1 = 5$	$T_1 = 10$	$T_1 = 15$	
6	7.217			7.224
7	8.448			8.450
8	9.651			9.675
9	10.865			10.890
10	12.072			12.102
11		13.300		13.292
12		14.496		14.512
13		15.685		15.727
14		16.900		16.918
15		18.109		18.123
16			19.286	19.320
17			20.492	20.520
18			21.684	21.733
19			22.884	22.919
20			24.061	24.117

of $N(t, B)$ for the same instants s computed with all the simulated sample paths. Tables 4, 5 and 6, respectively for the simulation cases 1–3, present the estimated means with the different three future intervals and the sample means.

The comparison between the estimations of the mean and the sample values of it (Tables 4, 5, 6) shows that the similarity or small error is remarkable for all the cases. It is also notable, as well as expected and explained in Sect. 5, that the estimations are more similar to the sample values if the instant s is closer to the instant T_1 . Each estimation is made in a different setting (different past intervals $T_1 = 5, 10$ or 15) and for different time points ($s = 6, \dots, 20$) so they are not comparable. Nevertheless, it can be illustrative that the average errors are 0.026, 0.102 and 0.278 (this last one is larger because of the last estimation value) for simulations 1, 2 and 3, respectively.

In order to study the estimation error of the mode and the probability of having a new point before s , 100 new sample paths were simulated. The accuracy of the estimations can be evaluated for any T_1 , the instant of time which splits the observation interval in past and future, and for any instant s of the future interval. This section will present the study of the errors for $T_1 = 10$ and $s = 11, 15$ and 20 as it would be impossible to present it for every T_1 and s .

The sample percentages of the number of points before an instant s are calculated from the 100 new sample paths. Tables 7, 8 and 9 show, for three examples of s , the most frequent values as well as their frequencies, respectively for the three simulation cases. Comparing the values of these tables with the estimated mode in Tables 1, 2 and 3, respectively, it can be observed that the estimated mode is always one of the most probable values. This observation was found to be a general behaviour for any s and T_1 . Thus, the estimation of the mode can be considered reliable.

Table 5 Comparison of the estimated mean and the sample mean of $N(s, B)$ from all the simulated sample paths for *simulation 2*

s	Estimation of the mean			Sample mean
	$T_1 = 5$	$T_1 = 10$	$T_1 = 15$	
6	31.882			31.900
7	37.194			37.155
8	42.420			42.466
9	47.738			47.712
10	52.271			53.038
11		58.271		58.264
12		63.578		63.532
13		68.879		68.842
14		74.226		74.126
15		79.495		79.470
16			84.770	84.718
17			90.085	89.997
18			95.436	95.312
19			100.715	100.646
20			105.875	105.955

Table 6 Comparison of the estimated mean and the sample mean of $N(s, B)$ from all the simulated sample paths for *simulation 3*

s	Estimation of the mean			Sample mean
	$T_1 = 5$	$T_1 = 10$	$T_1 = 15$	
6	11.041			11.147
7	12.941			12.993
8	14.769			14.891
9	16.629			16.716
10	18.452			18.588
11		20.265		20.414
12		22.183		22.225
13		24.085		24.118
14		25.919		26.016
15		27.751		27.843
16			29.559	29.711
17			31.359	31.528
18			33.285	33.308
19			35.166	35.246
20			39.936	37.112

Table 7 Most frequent values of number of points until s (Np) and percentage of sample paths with such number of points until s (%) for *simulation 1*

	$s = 10$		$s = 15$		$s = 20$	
	Np	%	Np	%	Np	%
8	9		14	10	19	8
9	12		15	18	20	10
10	17		16	16	21	13
11	18		17	12	22	13
12	20		18	7	23	13
13	5		19	14	24	8
14	4		20	8	25	11

Table 8 Most frequent values of number of points until s (Np) and percentage of sample paths with such number of points until s (%) for *simulation 2*

	$s = 10$		$s = 15$		$s = 20$	
	Np	%	Np	%	Np	%
47	7		69	7	99	2
48	6		70	3	100	3
49	1		71	3	101	6
50	4		72	5	102	2
51	6		73	5	103	4
52	3		74	4	104	6
53	1		75	10	105	7
54	9		76	2	106	3
55	2		77	7	107	1
56	5		78	5	108	9
57	1		79	2	109	2
58	8		80	5	110	1

Finally, let us examine the behaviour of the estimations of the probability of having a new point before s . Figures 4, 5 and 6 present the box-and-whisker plots of the estimations of the mentioned probabilities given the same instants used above and show the accuracy of the estimations, respectively for the three simulation cases.

6 Extreme values of temperatures in the South of Spain

Given the enormous socioeconomic and environmental impact of extreme meteorological events, providing predictive methods in this area has become increasingly prevalent in recent decades. As in many other phenomena in a wide variety of frameworks, the points which exceed some chosen limits are relevant and lead to a new series of events. A series of this type of events may be modeled as a point process or as the resulting counting process.

The data available for the example studied in this section are the maximum and minimum temperatures collected at some weather stations in the South of Spain (Seville,

Table 9 Most frequent values of number of points until s (N_p) and percentage of sample paths with such number of points until s (%) for *simulation 3*

$s = 10$		$s = 15$		$s = 20$	
N_p	%	N_p	%	N_p	%
7	1	14	6	26	8
8	5	15	6	27	5
9	4	16	4	28	3
10	5	17	7	29	5
11	10	18	6	30	11
12	7	19	8	31	4
13	8	20	6	32	8
14	11	21	5	33	4
15	8	22	3	34	3
16	7	23	7	35	5
17	10	24	4	36	6
18	5	25	9	37	3
19	3	26	4	38	4

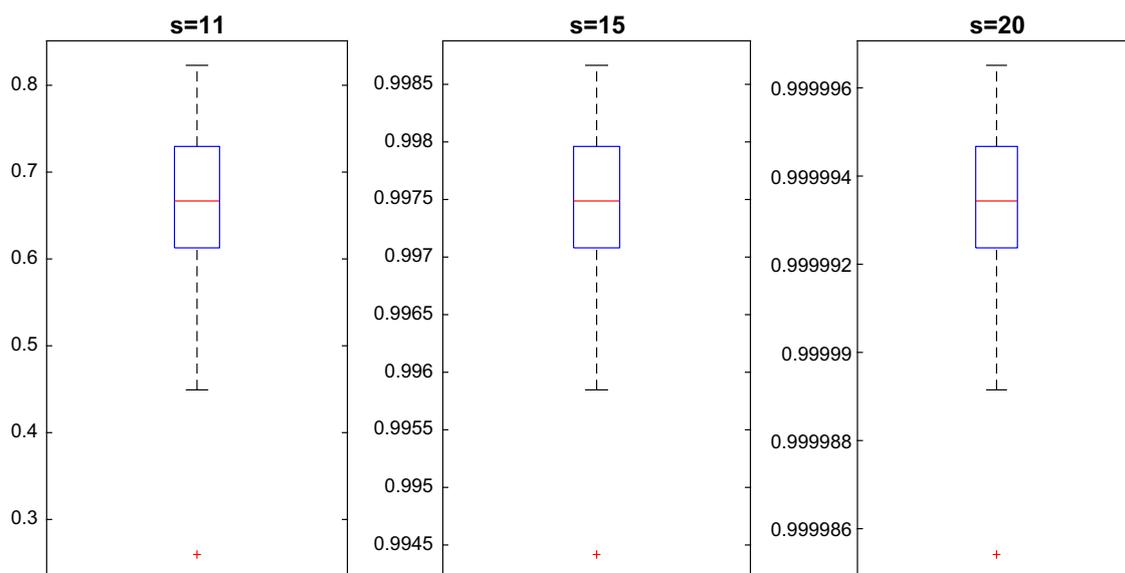


Fig. 4 Box-and-whisker plot of the estimations of the probability of having a new point before s for $T_1 = 10$ for *simulation 1*

Granada, Córdoba, etc.) belonging to the same climate zone. Let us consider the extreme maximum temperatures as the events of primary interest, i.e. the maximum temperatures over and under given thresholds. Then, these extreme values form a counting process. A second counting process arises applying the same reasoning to the extreme minima. In this paper, the CCP is proposed to model the both counting processes.

Let us begin developing carefully the study for the extreme maximum temperatures; afterwards, due to the parallelism, the case of the minima can be straightforward presented.

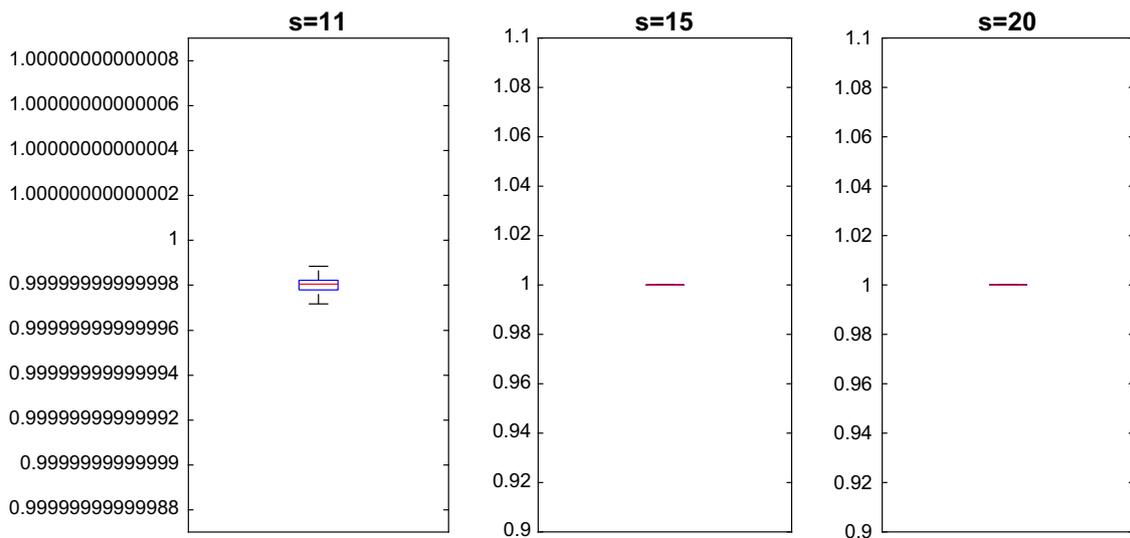


Fig. 5 Box-and-whisker plot of the estimations of the probability of having a new point before s for $T_1 = 10$ for *simulation 2*.

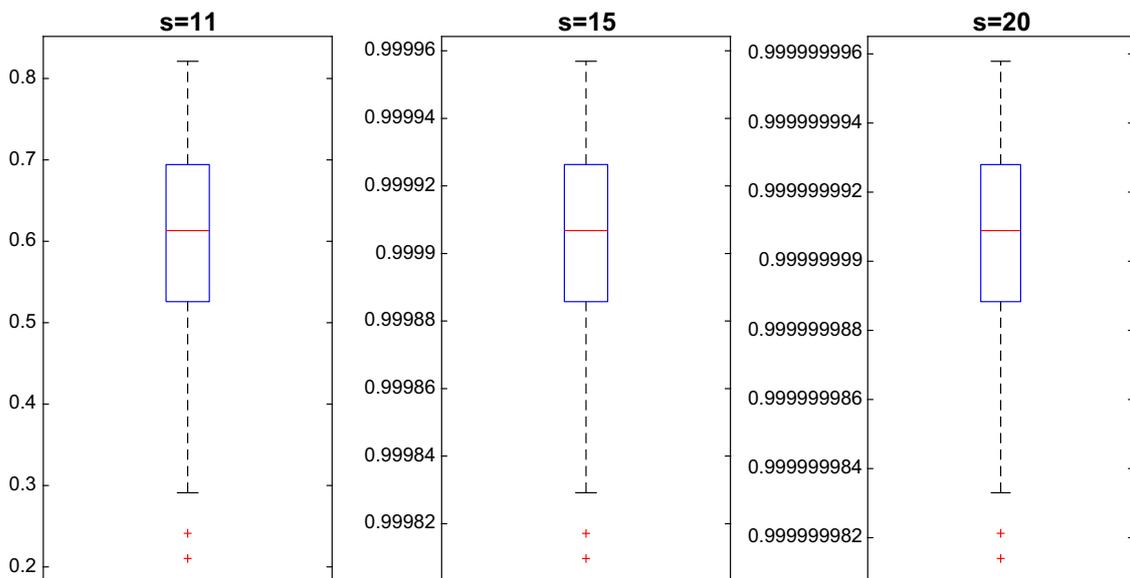


Fig. 6 Box-and-whisker plot of the estimations of the probability of having a new point before s for $T_1 = 10$ for *simulation 3*.

6.1 Extreme maximum temperatures

First of all, the extremes need to be calculated from the raw data and will be considered as a point process with the associated counting process $N(t)$ which mean process will be modeled. Later, the subset of marks B will be defined so that the counting process $N(t, B)$ is obtained. Finally, the forecast of the statistics will be calculated.

The detailed description of the data is the following. The maximum temperatures available belong to 15 weather stations within the same climate zone in the South of Spain from 1991 to 2011. The thresholds chosen to consider the temperature as an extreme value for each weather station were the limits around the sample mean

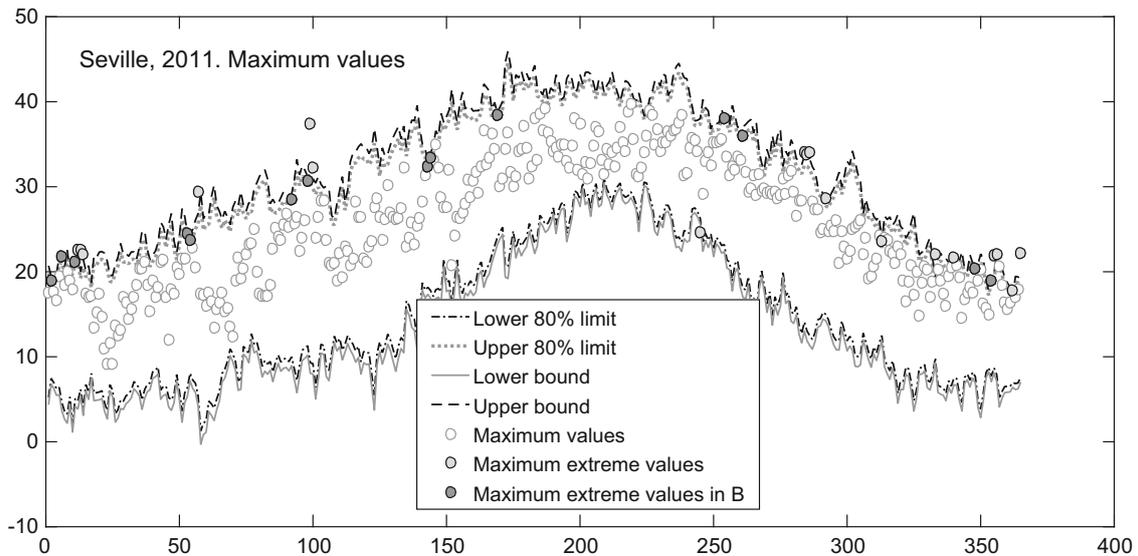


Fig. 7 Daily maximum temperatures, extreme maxima and extreme maxima in B in the weather station in Seville in 2011

accumulating the 80% of the observations for each day of the year using Chebyshev's inequality. The interval with that condition results to be $(\bar{x} - 2.361\sigma, \bar{x} + 2.361\sigma)$.

Studying the extremes, it can be observed that there are some extremely high or low values. Discarding these excessive extreme points, which can be caused by strange circumstances, the remaining points represent the usual behaviour of extreme temperatures in the area. Therefore, considering the actual value of the temperatures as the marks of the points, the process of final interest is formed by the extreme values in the subset of marks B , where B is the union of the intervals $(\bar{x} + 2.361\sigma, \bar{x} + 2.5\sigma)$ and $(\bar{x} - 2.361\sigma, \bar{x} - 2.5\sigma)$. As mentioned above, 2.361 is the value given by the Chebyshev's condition and 2.5 has been chosen to determine an upper and lower additional bounds for the marks. The probability associated to the subset B was estimated as $\int_B P_u(dU) = 0.5752$.

The weather station in Seville has been chosen as an example to illustrate the selection of the extreme values and the subset of marks B . Figure 7 shows the daily maximum temperatures in Seville along 2011, the values considered extremes and the values remaining in B . Figure 8 enhances the previous one for the first part of the year in order to facilitate a more detailed visualization.

Repeating the procedure of selecting the data in B for every weather station and considering the counting process generated, $N(t, B)$, 15 sample paths are obtained. The sample paths of $N(t)$ and $N(t, B)$ for all the weather stations are plotted in Fig. 9.

Concerning the time intervals involved in the performing of the FPCA and further results given in this paper, the interval $[T_0, T_2]$ is a year denoted by $[T_0 = 0, T_2 = 12]$ where 1, 2, ... are the months. Following the steps mentioned in Sect. 3 and as the temperatures have been observed during 21 years, a sample path of the mean process will be estimated from 21 of the counting process for each of the 15 weather stations. Then, 15 sample paths of the mean process have been obtained and will be used to derive the corresponding PCP model. The mean process of the extreme maximum temperatures in B has been modeled as a $PCP(1; 2)$ and for $s \in (T_1 = 5, T_2 = 12]$, it can be written the following way

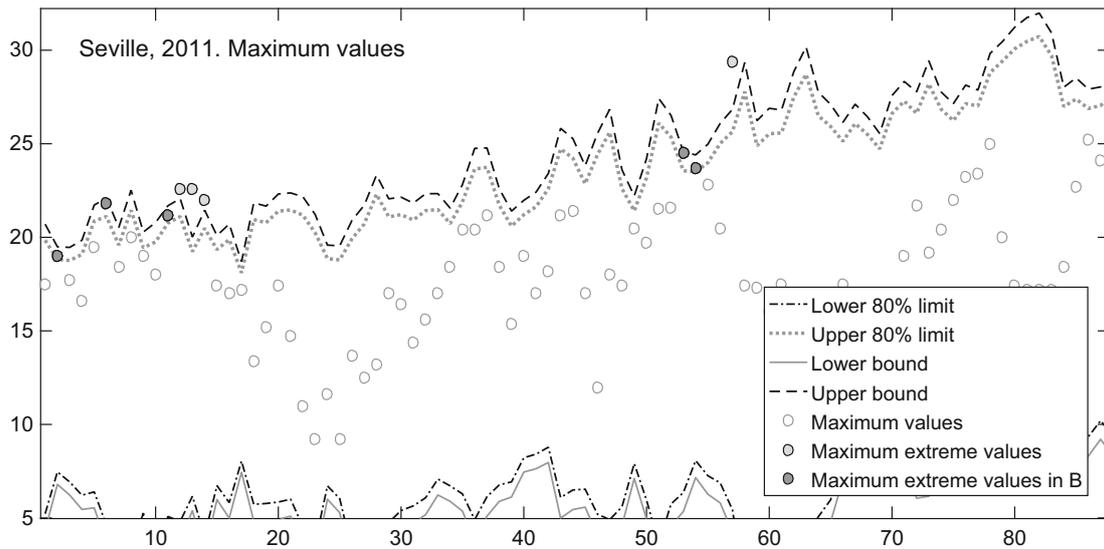


Fig. 8 Daily maximum temperatures, extreme maxima and extreme maxima in B in the weather station in Seville in the first 80 days of 2011

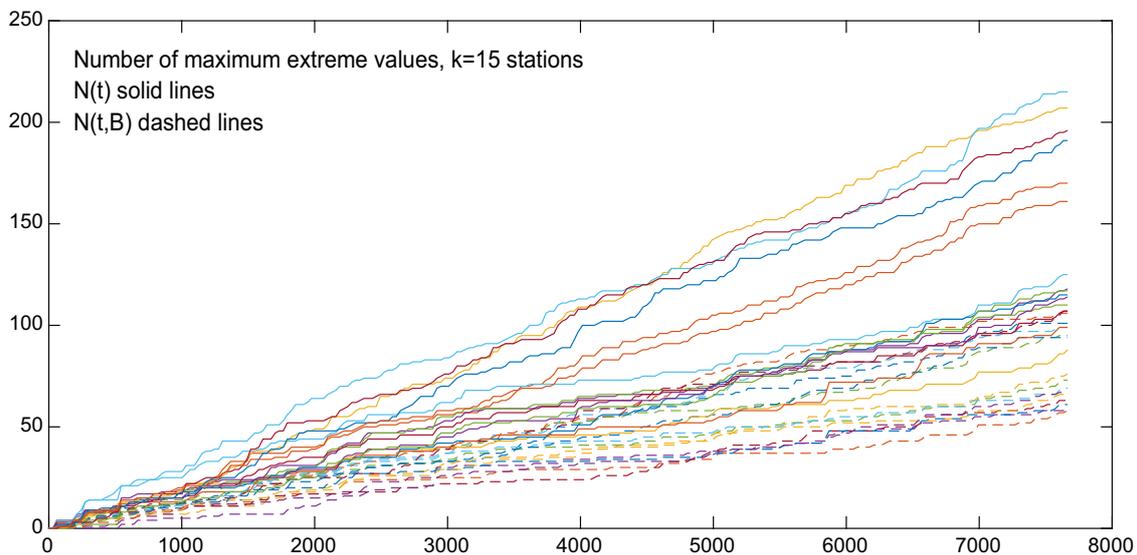


Fig. 9 Sample paths of the counting process of extreme maximum temperatures and for extreme maxima in the subset B for every weather station

$$\tilde{\Lambda}^1(s) = \mu_{\Lambda}^2(s) + (-0.705 \ 1.9792) \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} g_1(s).$$

Having derived the prediction of the mean process, the statistics can be forecast using Eqs. (10), (11) and Proposition 1. Table 10 shows the results of the forecast of the mean, mode and the probability of having a new point before s (Prob) for the extreme maximum temperatures in the subset B using three different future intervals chosen as examples. The future intervals used begin in $T_1 = 4$ which is the month of April, 5 which is May and 6, June and the instants of forecast are the following months.

As this is a real data example, an assessment of the results is not applicable, however the comparison between the estimations of the mean and the mean sample values can

Table 10 Estimation of the mean, mode and the probability of having a new point before s (Prob) for the extreme maximum temperatures in the subset B

T_1	s	Mean	Mode		Prob
4	5	1.561	0	1	0.226
	6	1.876	0	1	0.435
	7	2.221	1	2	0.599
	8	2.473	1	2	0.689
	9	2.8	1	2	0.776
	10	3.077	2	3	0.830
5	11	3.455	2	3	0.883
	6	1.876	1	1	0.270
	7	2.221	2	1	0.483
	8	2.473	2	3	0.598
	9	2.8	2	3	0.710
6	10	3.077	3	3	0.780
	11	3.455	3	3	0.843
	7	2.221	2	1	0.088
	8	2.473	2	3	0.438
	9	2.8	2	3	0.602
6	10	3.077	3	3	0.699
	11	3.455	3	3	0.794

Table 11 Comparison of the estimated mean and the sample mean of $N(s, B)$ from all the observed sample paths for the extreme maximum temperatures in the subset B

s	Estimation of the mean			Sample mean
	$T_1 = 4$	$T_1 = 5$	$T_1 = 6$	
5	1.561			1.524
6		1.876		1.857
7			2.220	2.191
8			2.473	2.667
9			2.8	3.191
10			3.077	3.667
11			3.455	4

be helpful. In order to illustrate that the estimations are also good in this real case, not only in the simulations, that comparison has been made in three scenarios of past and future intervals. The first scenario consists of choosing $T_1 = 4$, so the past is the interval from the beginning of the year until April, the second considers the past interval until May and the third, until June. The outcomes are shown in Table 11. It can be observed that the differences are small and, as explained in the previous sections, increasing as the instant s is further from the past interval.

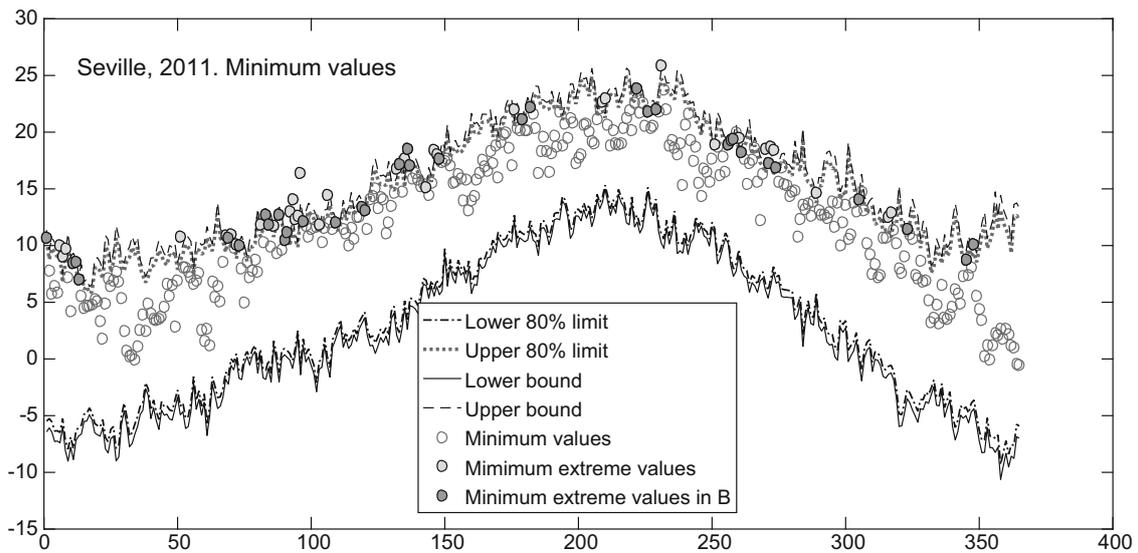


Fig. 10 Daily minimum temperatures, extreme minima and extreme minima in B in the weather station in Seville in 2011

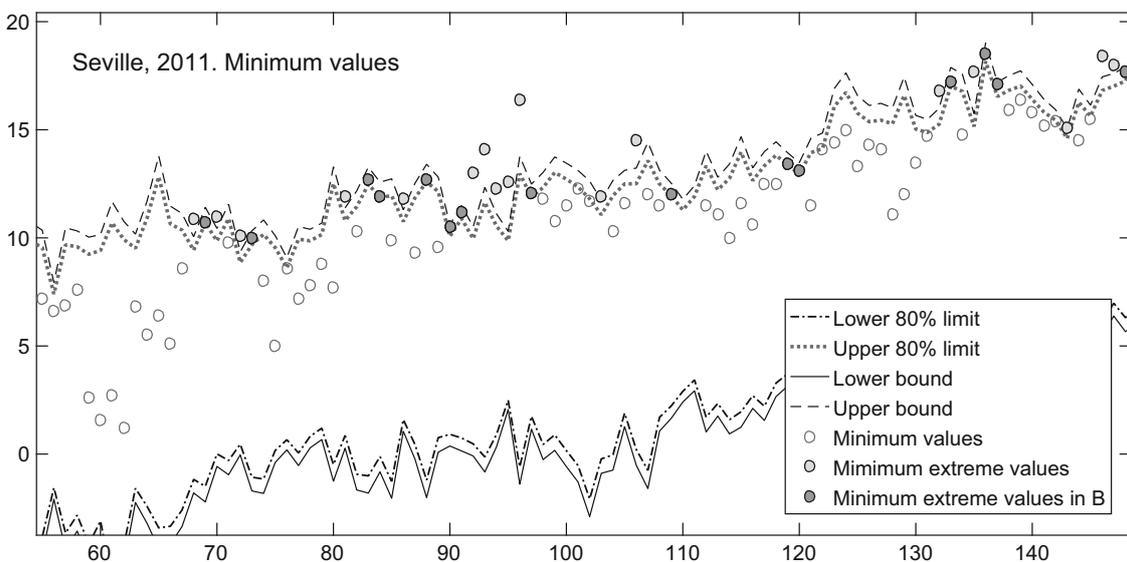


Fig. 11 Daily minimum temperatures, extreme minima and extreme minima in B in the weather station in Seville in the first 80 days of 2011

6.2 Extreme minimum temperatures

Having described in detail the process of applying the results of this paper from the raw data in the case of the extreme maximum temperatures, there is no more to explain for the case of the minima. The steps taken are completely parallel, including the same boundaries to define the extremes and the subset B . It only remains to provide the outcomes.

Once again, the weather station in Seville illustrates the selection of the extreme minimum values and those in the subset of marks B . Figure 10 shows the whole year 2011 and Fig. 11 gives a closer look at the first part of the year.

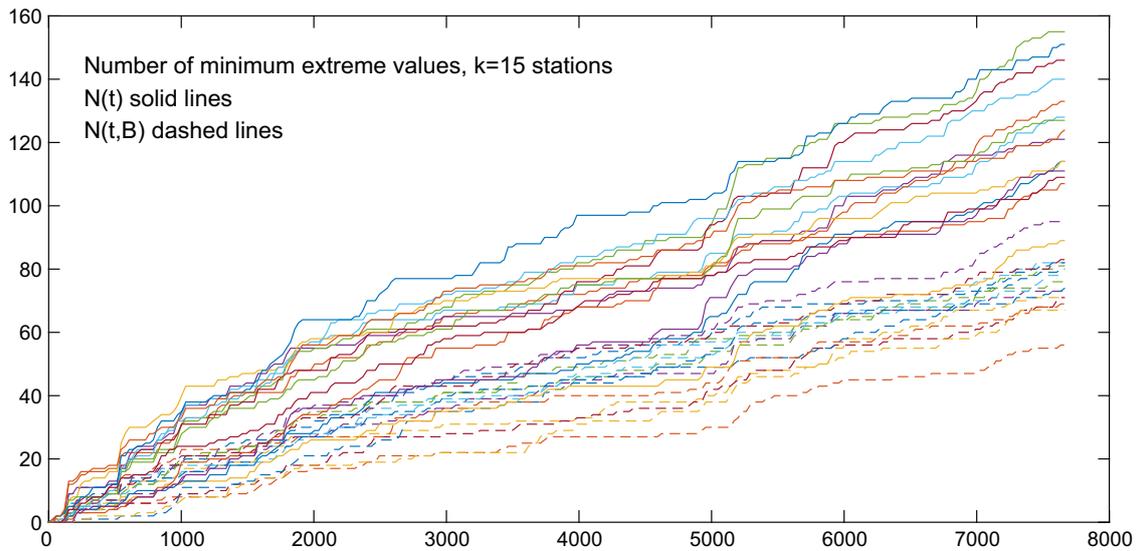


Fig. 12 Sample paths of the counting process of extreme minimum temperatures and for extreme minima in the subset B for every weather station

Selecting the data for all the weather stations and considering the counting processes associated with the points, the sample paths of $N(t)$ and $N(t, B)$ are obtained and plotted in Fig. 12.

Using the sample paths of $N(t)$, the mean process of the extreme minimum temperatures has been modeled as a $PCP(2; 1, 1)$ and for $s \in (T_1 = 5, T_2 = 11]$, it can be represented by

$$\tilde{\Lambda}^2(s) = \mu_{\Lambda}^2(s) + \begin{pmatrix} 0.2277\xi_1 \\ -1.3651\xi_1 \end{pmatrix} (g_1(s), g_2(s)).$$

Having the PCP model of the mean process, the statistics forecast for the extreme minimum temperatures in the subset B using three different future intervals are presented in Table 12.

As in the maxima, the predictions on Table 12 cannot be assessed but the comparison between the estimations of the mean and the mean sample values suggests that the results can be considered reliable (Table 13).

7 Concluding comments

The CCP and the CCP formed only by those points which have specific fixed marks are clearly interesting. This article derives closed-form expressions for the counting and time statistics of these processes in terms of the intensity and/or mean which are also stochastic. Then, it is notable that the mean or the intensity process that needs to be taken into account in order to forecast the CCP with specific marks, is the one of the counting process regardless the marks. This is important, firstly, because it allows to use the information of all the occurrences; secondly, because only that intensity

Table 12 Estimation of the mean, mode and the probability of having a new point before s (Prob) for the extreme minimum temperatures in the subset B

T_1	s	Mean	Mode		Prob
4	5	1.417	0	1	0.265
	6	1.8	0	1	0.499
	7	2.223	1	2	0.674
	8	2.564	1	2	0.769
	9	2.855	1	2	0.828
	10	3.122	2	3	0.866
	11	3.361	2	3	0.879
5	6	1.8	0	1	0.335
	7	2.223	1	2	0.577
	8	2.564	1	2	0.699
	9	2.855	1	2	0.775
	10	3.122	2	3	0.828
	11	3.361	2	3	0.864
6	7	2.223	1	2	0.372
	8	2.564	1	2	0.556
	9	2.855	1	2	0.669
	10	3.122	2	3	0.746
	11	3.361	2	3	0.788

Table 13 Comparison of the estimated mean and the sample mean of $N(s, B)$ from all the observed sample paths for the extreme minimum temperatures in the subset B

s	Estimation of the mean			Sample mean
	$T_1 = 4$	$T_1 = 5$	$T_1 = 6$	
5	1.417			1.571
6		1.800		1.714
7			2.223	2.143
8			2.564	2.381
9			2.856	2.714
10			3.122	2.905
11			3.361	3.191

or mean process must be forecast in order to forecast the statistics of every counting process of points in any given subset of marks.

The forecasts of the statistics are performed using PCP models which implies that the methodology used is only based on the sample observations of the counting process. Then, no assumptions on the intensity process, or the mean, need to be done to forecast the CCP.

The intensity process of a CCP is a non-negative stochastic process. Section 4 already recalled the importance of PH distributions within the framework of non-negative ones. Combining the topics of the last two sentences, this paper defines the phase type process, a process whose marginals are PH distributions. Then, a new and

versatile model for an intensity is provided and, as a result, a new model for a counting process. The phase type process is used to implement one of the simulation examples.

Three simulation examples have been performed to illustrate the results obtained throughout this paper. These examples were not chosen randomly; instead, they were deliberately selected in order to present an example of CCP which is a generalization of a CP with deletions (case 1), a CP with simultaneous occurrences (case 2) and a time-space CP (case 3). The latter case includes the singularity of having a phase type process as intensity. These diverse simulations are a small representation of all the counting processes that can be modeled by a CCP.

The accuracy of the theoretical expressions derived to forecast the counting and time statistics is assessed. The estimation of the mean is achieved by means of PCP models. The abundant literature on PCP models have already proved that this technique provides accurate estimations. The mode and the probability of having a new point before a fixed instant of the future are estimated for the three simulation examples. It would be impossible to present the estimation for every instant of time and any possible future interval so, this paper presents them in three distinct scenarios of the past and future intervals and for several instants of the future. As explained in the corresponding section, the assessment shows that the estimations can be considered accurate and robust for all the simulation cases and scenarios.

Finally, examining available data from different weather stations in the South of Spain, the extreme values of maximum and minimum temperatures have been modeled as CPP's. Considering the point process of the extremes within a given subset, the results obtained along this paper have been applied to the data. The PCP model for the mean processes were obtained, the mean and the mode in a given future instant of time and the probability of obtaining another point before an instant of time have been calculated. Even it is not possible or applicable the assessment of the results in a real data example, the comparison between the forecast of the mean and the mean sample values suggests a good performance of the model and the forecast in both extreme maximum and minimum temperatures.

Acknowledgements This work was supported by Ministerio de Economía y Competitividad (project MTM2013-47929-P) and Consejería de Innovación de la Junta de Andalucía (Grants FQM-307 and FQM-246), all in Spain.

References

- Aguilera AM, Ocaña FA, Valderrama MJ (1997) An approximated principal component prediction model for continuous-time stochastic processes. *Appl Stoch Models Data Anal* 13:61–72
- Asha G, Nair UN (2010) Reliability properties of mean time to failure in age replacement models. *Int J Reliab Qual Saf Eng* 17:15–26
- Barta P, Miller M, Qiu A (2005) A stochastic model for studying the laminar structure of cortex from mri. *IEEE Trans Med Imaging* 24:728–742
- Bieniek M, Goroncy A (2017) Sharp lower bounds on expectations of GOS based on DGFR distributions. *Stat Pap*. <https://doi.org/10.1007/s00362-017-0972-y>
- Bouzas PR, Aguilera AM, Valderrama MJ (2002) Forecasting a class of doubly stochastic Poisson processes. *Stat Pap* 43:507–523
- Bouzas PR, Valderrama MJ, Aguilera AM, Ruiz-Fuentes N (2006) Modelling the mean of a doubly stochastic poisson process by functional data analysis. *Comput Stat Data Anal* 50:2655–2667

- Bouzas PR, Ruiz-Fuentes N, Ocaña FM (2007) Functional approach to the random mean of a compound Cox process. *Comput Stat* 22:467–479
- Bouzas PR, Ruiz-Fuentes N, Matilla A, Aguilera AM, Valderrama MJ (2010a) A cox model for radioactive counting measure: inference on the intensity process. *Chemom Intell Lab Syst* 103:116–121
- Bouzas PR, Ruiz-Fuentes N, Ruiz-Castro JE (2010) Forecasting a compound cox process by means of PCP. In: Lechevallier Y, Saporta G (eds) *Proceedings of COMPSTAT'2010*. Physica-Verlag, Heidelberg, pp 839–846
- Bouzas PR, Aguilera AM, Ruiz-Fuentes N (2012) Functional estimation of the random rate of a Cox process. *Methodol Comput Appl Probab* 14:57–69
- Chen F, Hall P (2013) Inference for a non-stationary self-exciting point process with an application in ultra-high frequency financial data modeling. *J Appl Probab* 50:1006–1024
- Chertok AV, Korolev VY, Korchagin AY (2016) Modeling high-frequency non-homogeneous order flows by compound Cox processes. *J Math Sci* 214:44–68
- Dousse O, Baccelli F, Thiran P (2005) Impact of interferences on connectivity in ad hoc networks. *IEEE/ACM Trans Netw* 13:425–436
- Economou A (2003) On the control of a compound immigration process through total catastrophes. *Eur J Oper Res* 147:522–529
- Genaro AD, Simonis A (2015) Estimating doubly stochastic Poisson process with affine intensities by Kalman filter. *Stat Pap* 56:723–748
- Gospodinov D, Rotondi R (2001) Exploratory analysis of marked Poisson processes applied to Balkan earthquake sequences. *J Balkan Geophys Soc* 4:61–68
- Greenberg D, Houweling A, Kerr J (2008) Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat Neurosci* 11:749–751
- He Q (2014) *Fundamentals of matrix-analytic methods*. Springer, New York
- Lefebvre M, Belsalma F (2015) Modeling and forecasting river flows by means of filtered Poisson processes. *Appl Math Model* 39:230–243
- Lin XS, Pavlova KP (2006) The compound Poisson risk model with a threshold dividend strategy. *Insur Math Econ* 38:57–80
- Neuts MF (1975) *Probability distributions of phase type*. Liber. Amicorum Prof. Emeritus. H. Florin. Department of Mathematics, University of Louvain, Belgium
- Neuts MF (1981) *Matrix geometric solutions in stochastic models: an algorithmic approach*. Dover, New York
- Ogata Y (1998) Space–time point-process models for earthquake occurrences. *Ann Inst Stat Math* 50:379–402
- Park C, Padgett W (2005) Accelerated degradation models for failure based on geometric Brownian motion and gamma processes. *Lifetime Data Anal* 11:511–527
- Ruiz-Castro JE (2016a) Complex multi-state systems modelled through marked markovian arrival processes. *Eur J Oper Res* 252:852–865
- Ruiz-Castro JE (2016b) Markov counting and reward processes for analyzing the performance of a complex system subject to random inspections. *Reliab Eng Syst Saf* 145:155–168
- Russell JR, Engle RF (2010) Analysis of high-frequency data. In: *Handbook of financial econometrics: tools and techniques*. Elsevier, Amsterdam, pp 383–426
- Sepehrifar M, Yarahmadian S (2017) Decreasing renewal dichotomous Markov noise shock model with hypothesis testing applications. *Stat Pap* 58:1115–1124
- Si S (2001) Random irreversible phenomena: entropy in subordination. *Chaos Solitons Fractals* 12:2873–2876
- Singh S, Tripathi YM (2018) Estimating the parameters of an inverse Weibull distribution under progressive type-I interval censoring. *Stat Pap* 12:2873–2876
- Snyder DL, Miller MI (1991) *Random point processes in time and space*, 2nd edn. Springer, New York
- Tank F, Eryilmaz S (2015) The distributions of sum, minima and maxima of generalized geometric random variables. *Stat Pap* 56:1191–1203