Bats Monitoring: A Classification Procedure of Bats Behaviors based on Hawkes Processes

Christophe Denis, Charlotte Dion-Blanc, Romain Edmond Lacoste, Laure Sansonnet, Yves Bas

▶ **To cite this version:**

Christophe Denis, Charlotte Dion-Blanc, Romain Edmond Lacoste, Laure Sansonnet, Yves Bas. Bats Monitoring: A Classification Procedure of Bats Behaviors based on Hawkes Processes. 2023. hal-04345822

HAL Id: hal-04345822
https://hal.science/hal-04345822

Preprint submitted on 14 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bats Monitoring: A Classification Procedure of Bats Behaviors based on Hawkes Processes

Christophe Denis[(2,3)], Charlotte Dion-Blanc[(3)],

Romain E. Lacoste[(2)], Laure Sansonnet[(4,3)], Yves Bas[(1,5)]

(1) CESCO, UMR 7204, MNHN, CNRS, Sorbonne Université

(2) LAMA, UMR 8050, Université Gustave Eiffel

(3) LPSM, UMR 8001, Sorbonne Université

(4) Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay

(5) PatriNat, OFB, MNHN*

December 14, 2023

Abstract

We are interested in the problem of classifying commuting and foraging behavior of bats at a site. To this extent, we use echolocation calls data detected by acoustic sensors during a night at this site. As the temporal distribution of calls is a relevant indicator of behavior, it is natural to model the calls sequences by point processes. Given the self-exciting dynamics observed in foraging behavior, we propose to model the calls sequences of bats by Hawkes processes. Specifically, we consider that the start time of each call emitted on a site is a jump of a Hawkes process. For the classification task, we use a suitable procedure that relies on the empirical risk minimization principle. Then, we assess the performance of the procedure on synthetic data and also provide a comparison with the random forests algorithm. The overall methodology is evaluated with a goodness-of-fit test. Finally, we present the obtained results on a real data set, collected as part of Vigie-Chiro project. The classification results are convincing and show the relevance of our method.

Keywords Supervised learning, Point process model, Bat monitoring.

1 Introduction

In this paper, we tackle the problem of classifying foraging and commuting behavior of bats at a site using echolocation calls data recorded by acoustic sensors during a night at this site. The increasing affordability and storage capacity of acoustic recorders has encouraged the development of Passive Acoustic Monitoring (P.A.M.) for the study of species behavior. This growing accessibility has given rise to participatory project such as Vigie-Chiro[1] project. Vigie-Chiro is a France-wide bats acoustic monitoring project. As the calls bats emit are ultrasonic and therefore inaudible

*The authors are in alphabetic order, except the last one who provides the application context and the data set.

[1]Website of the project here

to humans, P.A.M. is a relevant analysis tool to study their call sequences. In addition, the nocturnal nature of bats makes them ideal subjects for P.A.M. studies, as they are less subject to anthropogenic disturbance that can add noise to recordings. In the present work, we propose a supervised method to classify the behavior of a specimen on a site during one night, between a commuting behavior, meaning that the animal is just passing by the monitored site, and foraging behavior, meaning that animal is hunting on the site. Being able to compare sites according to their majority behavior, foraging sites or commuting sites, would open large opportunities in spatial ecology such as evaluating landscape connectivity and foraging habitat selection.

State of the art. Research shows the temporal distribution of the calls emitted by a bat is a relevant indicator of its behavior. In particular, the feeding buzz proportion has been used to detect foraging on a site, for example in [15, 20, 29], where a feeding buzz characterizes the capture of a prey. Thus, the presence of feeding buzzes is a strong indicator of foraging behavior. Although a classifier based on the feeding buzzes proportion achieves good classification score [27], the detection of buzzes require heavy computational resources. Furthermore, some bats, such as gleaners (e.g. *Plecotus auritus* or *Myotis myotis*) rarely emit feeding buzzes, as they often rely on the acoustic cues emitted by their prey to feed [12] . For these species, it is therefore not possible to rely on this proxy to classify bat behavior. Finally, it is important to notice that the raw data recorded on each site (as temperature, wind, etc.) are not sufficient to characterize the behavior of the bats. It therefore appears that building a classifier based solely on the information contained in the calls sequences, such as the start time of each calls sequence, represents a major challenge in this context.

In the literature, conventional metrics for quantifying bat activity involve averaging the number of pulses emitted in each segment of time over the course of a night [see 9]. The novelty of the present work is to model the temporal dynamic of bats calls and to take advantage of this modeling to classify their behavior. Precisely, in the case of foraging behavior, we expect to observe a self-exciting dynamic. Indeed, the discovery of prey at a site may encourage the surrounding bats to stay there and continue to forage and thus to emit calls. For this reason, we propose to model the temporal calls sequence of bats by a special type of point process, namely a Hawkes processes.

Hawkes processes, introduced in [16], are proposed to model events sequence where the occurrence of events increase the chances of getting another in the future. Thanks to their intrinsic ability to model self-exciting dynamics, Hawkes processes are extremely versatile in terms of application domains. Although historically applied to model earthquakes in seismology in [17, 22], their sphere of application has rapidly expanded to many other fields. For instance, in neuroscience in [26], in mathematical finance in [3], in social network studies in [24], in football [4]. From a theoretical point of view, numerous statistical methods of inference have been proposed for Hawkes processes. On can cite recently for example [2, 13, 30, 7] and references there in.

Main contributions. We propose a methodology dedicated to discriminate the majority behavior of bats at a site, between foraging and commuting, based on sequence of echolocation calls data. One of the main improvement of the present paper regarding to the existing literature is the proposed modeling of the sequence of echolocation calls by a point process. Indeed, rather to consider some global feature such as mean, we provide a model which is designed to handle the temporal dependency of the data. The resulting modeling has two appealing properties, it is interpretable and can be easily implemented for practical purpose. This thinner description of the sequence echolocation calls data allows us to build an efficient classification procedure tailored to the proposed model.

More precisely, we model sequences of bats calls as a realization of a linear exponential Hawkes

processes. According to this model the classes are discriminated by the baseline and the parameter of the exponential kernel of the Hawkes process. We then adapt, in the binary classification setup, the classification procedure presented in the seminal work of [11]. This algorithm is a surrogate risk minimization procedure that is dedicated to the classification of Hawkes processes. A salient point of our procedure is that we do not only consider the usual misclassification risk to assess the performance of our methodology, but also a goodness-of-fit test. It allows us to evaluate both modeling and classification procedure. The goodness-of-fit test relies on the Time-Scaling Theorem [10] and a Kolmogorov-Smirnov test.

We show that the proposed classification exhibits good performance on synthetic data and promising results on real data. In particular, we implement our classification procedure on a labeled data set whose classes, commuting and foraging, are well separated. We then evaluate the classification performance using a cross-validation scheme and provide a comparison with the popular random forests algorithm. The percentage of correct classification is around 68.13% which is better than the one obtained with the random forests procedure. Besides, we emphasize that our method provides a model for the temporal sequence recorded. This is a key point of our procedure by comparison with black-box algorithms. Furthermore, the results obtained with the goodness-of-fit test highlights the relevance of our modeling. Lastly, we apply our classification algorithm on an unlabeled data set whose classes are much overlapped that leads to a difficult classification problem. We again evaluate the resulting predictions with the goodness-of-fit and obtain similar results as for labeled data.

Finally, beyond the classification task, our proposed procedure outputs an index of the foraging behavior which do not depend on environmental covariates. In particular, we can consider this index to characterize a monitored site and investigate scale effects. We compare this index of prediction with covariates that measure the presence of artificial light and water. As expected, according to our prediction rule, light or water impact the foraging behavior.

Content. This paper is organized as follows. Data are presented in Section 2. The statistical modeling as well as the classification method that we propose are described in Section 3. The performance of our procedure is investigated through numerical experiments on synthetic data in Section 4. In Section 5, we apply our procedure on the real data that motivated this study. Finally, we draw some perspectives in Section 6.

2 Data

The data set, collected as part of Vigie-Chiro project, is described in Section 2.1 while the modeling of the data is investigated in Section 2.2. Then, in Section 2.3 we present the construction of the data sets used to build and perform our classification procedure for dealing with the problem of classifying commuting and foraging behavior of bats.

2.1 Echolocation call data

The original data set consists of sequences of bats echolocation calls recorded over one or more nights at 755 different sites in France. More precisely, for each observation, namely each observed site, we have access to the start and end time of each sequence emitted which may come from different species. In addition, through additional data processing (developed in [27]), we have access to the number of feeding buzzes detected on each site. Feeding buzz (see [15] for instance) is a very specific pattern in which the bat emits a burst of echolocation calls at a very high rate in an attempt of prey capture. Therefore, The presence of feeding buzzes is a strong indicator

of foraging behavior at a site. A visual representation of such a pattern is given in Figure 1. Finally, other environmental covariates are also available, and summarized in Table 1. Note that the variables Light and Water are used in Section 6 to discuss the predictions of our method.

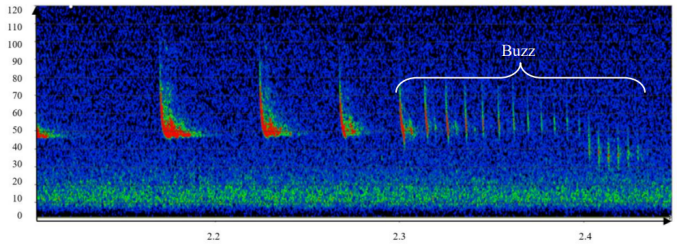

Figure 1: Sonogram containing a feeding buzz, frequency as a function of time.

Name	Description
N_{buzz}	Number of feeding buzzes detected
D_{tot}	Total duration of all recorded calls
Long	Longitude
Lat	Latitude
Light	Artificial light in a 500 m buffer zone
Water	Body of soft water in a 500 m buffer zone

Table 1: Overview of the available covariates

In the following, for a given site, we call an *event*, denoted T_ℓ, the starting time of a recorded sequence of calls. We refer to a *sequence of events* several events $(T_\ell)_{\ell\geq 1}$ on a given observation interval. Finally, we choose to use the denomination *sequence of calls* for the sequence of events on one site rather than *sequences of starting time of sequences of calls* for the sake of simplicity.

Figure 2 displays the histogram of the numbers of events per site as well as their spatial distribution on the map of France. In particular, it can be seen that the data set contain a wide variety of sites profiles, making it representative of ecological realities. Note also that an important part of the sites present a moderate activity.

2.2 Modeling the sequence of calls

In this section, we discuss the modeling of the sequences of calls. The echolocation system leads us to consider the temporal distribution of the calls with a view to obtain a good characterization of bat behavior. We propose to model a sequence of calls $(T_\ell)_{\ell\geq 1}$ as a realization of a point process N. Some justification of our approach are given hereafter.

Figure 3 displays an illustration of the temporal dynamic of a sequence of calls. Four temporal sequences are represented in this figure on the left part, each line corresponding to one site. Besides, the autocorrelation as the function of the lag for the same four nights is displayed in Figure 3 on the right part. Note that the darker the square, the stronger the autocorrelation. In particular, for the four nights, the autocorrelation is not zero. Furthermore, for the two first nights, it does not vanish quickly. In view of this observation, for a sequence of calls $(T_\ell)_{\ell\geq 1}$, it suggests that a temporal

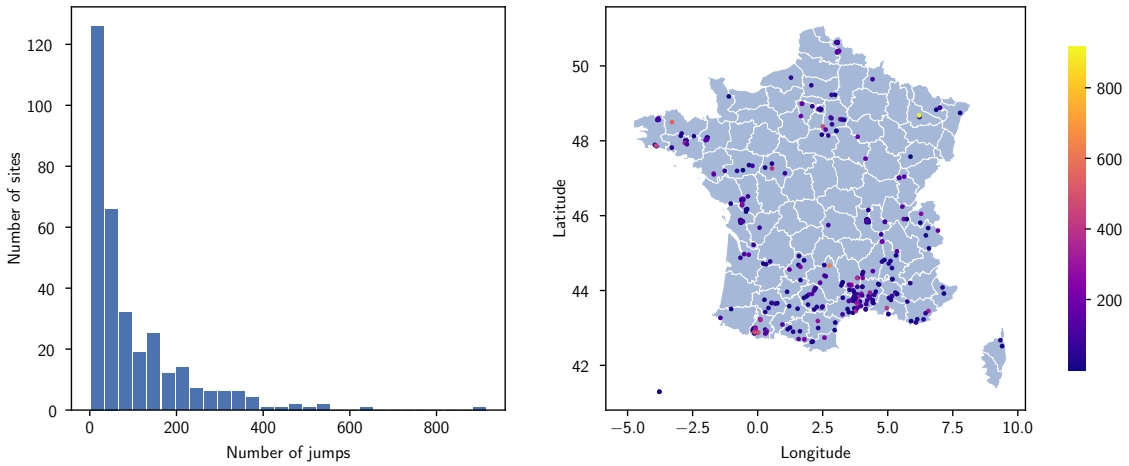

Figure 2: *(Bats data set).* Histogram of the number of events for each sites (left). Visual representation on a map of France: each point on the map represents a site and its color refers to the number of events in the temporal sequences (right).

dependency between the events T_ℓ exists. Therefore, it leads us to consider the observed sequences of events as realizations of Hawkes processes [16]. Indeed, Hawkes process is a continuous point process for which the occurrence of random events depends on past events. In particular, the use of such a model to describe bats behavior seems quite relevant due to the underlying self-exciting dynamic. Indeed, the discovery of prey at a site, indicated by the emission of echolocation calls, may encourage surrounding bats to stay there and continue to forage and thus continue to emit calls. Namely, if there is a prey to catch on a site, it promotes the arrival of other animals. The proposed modeling is detailed in Section 3.1.

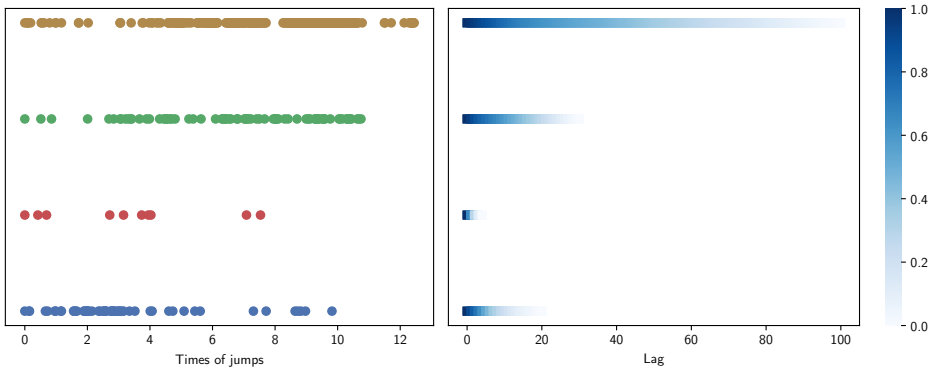

Figure 3: *(Bats data set).* Visualization of the temporal dependence for four nights. On the left are represented the start times of echolocation calls sequences. On the right it is the autocorrelation as a function of the lag for the same four nights.

2.3 Construction of the data set

In this section, we describe the construction of the data set that we used to perform our classification procedure. We focus on the Common Pipistrelle (*Pipistrellus pipistrellus*) since this is a species that emits buzzes and for which a large amount of data is available. Besides, for each site,

we choose to only keep the first night of recording to maintain parity between sites. It also ensures controlled recording quality, since the sensors are installed in good weather conditions whereas poor conditions (such as rain or wind) may add unwanted noise, in addition to the fact that it may also disrupt the behavior of the bats, for the other nights. Moreover, only calls in the time slot from sunset to 1:00 am are kept, since this is the period when bats are more active. Hence, for each observation, the considered feature is the time series of the start times of each call emitted on each site.

Then, we build two data sets: a labeled data set which serves as a training set, and an unlabeled data set that we consider as the test set. More precisely, a part of the observations are labeled according to the proportion of buzzes detected on the site (see [28]). Thus, to each site is associated the following index:

$$\mathcal{I}(s) = 3600 \times \frac{N_{\text{buzz}}(s)}{D_{\text{tot}}(s)} \tag{1}$$

where $s \in \{1, \dots, 755\}$ is a given site, $N_{\text{buzz}}(s)$ is the number of feeding buzzes detected on this site, and $D_{\text{tot}}(s)$ the total duration of all recorded calls on the site. When the buzz index $\mathcal{I}$, defined by Equation (1), is very low (very few buzzes by hour) or very large (clear foraging behavior), this characterizes easily the two classes (commuting and foraging).

Therefore, for $s \in \{1, \dots, 755\}$, we consider the following labeling scheme:

- if $\mathcal{I}(s) < \tau_1$ then commuting is the majority behavior on s;
- if $\mathcal{I}(s) > \tau_2$ then foraging is the majority behavior on s.

The threshold values are fixed to $\tau_1 = 3$ and $\tau_2 = 90$. Since our data set is well representative of ecological realities of behavior (see Figure 4), these threshold values are sufficiently marked to characterize the two behavior and therefore the training data set is correctly labeled. In the following, the label 1 refers to the class of foraging sites when $I(s) > \tau_2$, and label 0 to the class of commuting sites for which $I(s) < \tau_1$. Performing such a partition induces a quite balanced distribution of 168 observations for class 0 and 164 sites for class 1. These observations for the labeled data denoted $\mathcal{D}_n^L$ in the following, with $n = 332$. The remaining observations, that is to say the sites for which the index belongs to $[\tau_1, \tau_2]$, form the unlabeled data set denoted $\mathcal{D}_m^U$ in the following with $m = 423$. Figure 4 displays the distribution of the index distinguishing the labeled data (and the two classes 0 and 1) and the unlabeled data. The unlabeled data correspond to the central values of the index (in $[\tau_1, \tau_2]$) which is the zone where the index does not give a clear characterization of the commuting (class 0) or foraging (class 1) behavior.

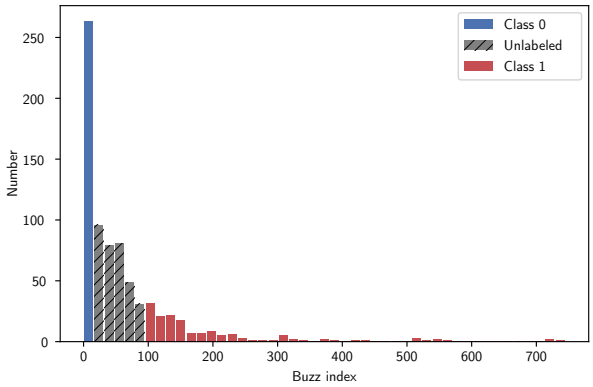

Figure 4: *(Bats data set).* Visualization of the labeling scheme using feeding buzzes proportion.

3 Statistical methodology

This section is devoted to the statistical methodology developed for our classification procedure. The temporal model for the observations is introduced in Section 3.1. The classification method that relies on our modeling is formally described in Section 3.2. Finally, in Section 3.3, we give details on goodness-of-fit tests dedicated to assess the performance of our procedure.

3.1 Data modeling with Hawkes processes

In Section 2.2, we have motivated the modeling of the sequence of calls by a Hawkes process [16, 18, 22]. Indeed, in the context of echolocation call data, mainly in the presence of the foraging buzz pattern, we expect the events times (or jumps times) to be linked and that some events triggered other event.

Formally, we consider that a sequence of calls $(T_\ell)_{\ell\geq 1}$ is a sequence of events coming from a Hawkes process. Such point process, denoted N, can be characterized by its conditional intensity, which gives the probability to have a jump in a small interval after t conditional on the past:

$$\lambda(t) := \lim_{h\to 0^+} \frac{\mathbb{P}\left(N_{t+h} - N_t = 1 \mid \mathcal{F}_{t^-}\right)}{h}$$

where $\mathcal{F}_{t^-}$ denoted the information of the past strictly before time t. We focus here on the linear exponential Hawkes process. In this particular case, the intensity is given for $t \geq 0$ by

$$\lambda(t) := \mu + \int_0^t \alpha\beta e^{-\beta(t-s)} dN_s = \mu + \sum_{T_\ell < t} \alpha\beta e^{-\beta(t-T_\ell)},$$

where

- sequence $(T_\ell)_{\ell\geq 1}$ are the time jumps of the process;
- parameter $\mu > 0$ is the baseline intensity, driving the dynamic of the process before the arrival of jumps;
- parameter $\alpha \in [0, 1)$ is the arrival intensity, describing the influence of a jump on future jumps;
- parameter $\beta \geq 0$ is the rate of the decay which controls the arrival intensity and dictates how fast this influence vanishes *w.r.t.* the time.

Hence, we assume that the conditional intensity of the underlying process N depends on an unknown parameter which belongs to the following family of parameters

$$\Theta = \{\mu > 0, 0 \leq \alpha < 1,\ \beta \geq 0\}.$$

The condition on α (sub-critical regime) ensures the non explosion of the intensity process (*i.e.* $\mathbb{E}[\lambda_t]$ bounded). Note that the case where $\alpha = 0$ corresponds to the definition of a Poisson process (see [10] for instance). The exponential kernel is widely studied, and its choice is mainly motivated by the recursive writing of the likelihood which serves the purpose of fast computation. Furthermore, for practical purpose, the resulting model is easily interpretable.

Hereafter, we formally present the statistical framework which consists in a mixture of two Hawkes models.

Mixture model. The observed features are assumed to come from a *mixture* of Hawkes processes whose jump times lie on the interval $[0, T]$, with $T > 0$. The counting process is denoted $(N_t)_{(0\leq t\leq T)}$ and the associated jumps times are $T_1, \ldots, T_{N_T}$. Let us introduce the label Y which is distributed according to a Bernoulli distribution with parameter $\mathfrak{p}^* = \mathbb{P}(Y = 1)$ assumed to be unknown. Then, conditional on Y, we assume that N has a conditional intensity writing as

$$\lambda_{\theta_Y^*}(t) := \mu_Y^* + \int_0^t \alpha_Y^* \beta_Y^* e^{-\beta_Y^*(t-s)}\, dN_s = \mu_Y^* + \sum_{T_\ell < t} \alpha_Y^* \beta_Y^* e^{-\beta_Y^*(t-T_\ell)}, \tag{2}$$

where $\theta_Y^* = (\mu_Y^*, \alpha_Y^*, \beta_Y^*) \in \Theta$ is unknown. In particular, the intensity depends now on the label of the observations. Therefore, a generic observation is the random pair $\{\mathcal{T}_T, Y\}$, where the feature $\mathcal{T}_T = (T_1, \ldots, T_{N_T})$ are jumps times of a Hawkes process with intensity $\lambda_{\theta_Y^*}$. The training data set $\mathcal{D}_n^L := \left\{\left(\mathcal{T}_T^1, Y^1\right), \ldots, \left(\mathcal{T}_T^n, Y^n\right)\right\}$ consists of independent copies of $\{\mathcal{T}_T, Y\}$. In this supervised setting, based on the learning sample $\mathcal{D}_n^L$, our goal is to build an empirical classifier denoted by $\hat{g}$. Hence, from a new observation $\mathcal{T}_T$, $\hat{g}(\mathcal{T}_T)$ is a prediction of the label Y.

Bayes classifier. The performance of a predictor $\hat{g}$ is usually measured through its misclassification risk $\mathbb{P}\left(\hat{g}(\mathcal{T}) \neq Y\right)$. In particular, we expect that this risk is closed to the one of the Bayes classifier g^* defined as

$$g^* \in \arg\min_g \mathbb{P}\left(g(\mathcal{T}_T) \neq Y\right).$$

In [11], the authors show that g^* can be characterized as $g^*(\mathcal{T}_T) = \mathbb{1}_{\{\pi_{\mathfrak{p}^*,\theta^*}(\mathcal{T}_T)\geq 1/2\}}$, with

$$\pi_{\mathfrak{p}^*,\theta^*}(\mathcal{T}_T) = \mathbb{P}\left(Y = 1|\mathcal{T}_T\right) = \frac{\mathfrak{p}^* \exp\left(F_1^*(\mathcal{T}_T)\right)}{\mathfrak{p}^* \exp\left(F_1^*(\mathcal{T}_T)\right) + (1-\mathfrak{p}^*) \exp\left(F_0^*(\mathcal{T}_T)\right)},$$

and

$$F_k^*(\mathcal{T}_T) = -\int_0^T \lambda_{\theta_k^*}(s) ds + \sum_{T_\ell \in \mathcal{T}_T} \log\left(\lambda_{\theta_k^*}(T_\ell)\right).$$

Hence, the characterization of the Bayes classifier relies on the likelihood of the observations conditionally to the labels.

3.2 Classification algorithm

In this section, we briefly describe the classification procedure previously proposed in [11] for the supervised multiclass classification of Hawkes processes. The procedure is here detailed in the binary classification setup. In the first step of the algorithm, the distribution $\mathfrak{p}^*$ of labels is estimated by its empirical counterpart $\hat{\mathfrak{p}}$. In the second step, we build a classifier $\hat{g}$ that relies on a surrogate risk minimization [5]. In particular, it involves a score function $\hat{f}$ that maps a feature $\mathcal{T}_T$ onto $\mathbb{R}$. The score function $\hat{f}$ is defined as the minimizer of the empirical quadratic loss over a suitable class of functions $\hat{\mathcal{F}}$ defined below, that is

$$\widehat{f} \in \operatorname*{argmin}_{f\in\widehat{\mathcal{F}}} \frac{1}{n}\sum_{i=1}^n \left(U_i - f(\mathcal{T}_T^i)\right)^2, \tag{3}$$

with $U_i = 2\mathbb{1}_{\{Y_i=1\}} - 1$. Conditional on the label, the choice of the class of functions $\widehat{\mathcal{F}}$ is based on the likelihood of the observations and is defined as $\hat{\mathcal{F}} = \{2\pi - 1 :\ \pi \in \widehat{\Pi}\}$ where

$$\widehat{\Pi} = \left\{\pi_{\widehat{\mathfrak{p}},\theta},\ \boldsymbol{\theta} = (\theta_0, \theta_1) \in \Theta^2, \pi_{\widehat{\mathfrak{p}},\boldsymbol{\theta}}\left(\mathcal{T}_T\right) = \frac{\widehat{\mathfrak{p}} \exp\left(F_1(\mathcal{T}_T)\right)}{\widehat{\mathfrak{p}} \exp\left(F_1(\mathcal{T}_T)\right) + (1-\widehat{\mathfrak{p}}) \exp\left(F_0(\mathcal{T}_T)\right)}\right\}$$

with, for $k \in \{0,1\}$

$$F_k(\mathcal{T}_T) := -\int_0^T \lambda_{\theta_k}(s)ds + \sum_{T_\ell \in \mathcal{T}_T} \log\left(\lambda_{\theta_k}(T_\ell)\right), \text{ with } \lambda_{\theta_k}(s) = \mu_k + \sum_{T_\ell < t} \alpha_k \beta_k e^{-\beta_k(t-T_\ell)}.$$

Hence the score function $\hat{f}$ satisfies $\hat{f} = 2\pi_{\hat{\mathfrak{p}},\hat{\boldsymbol{\theta}}} - 1$. Note that $\pi_{\widehat{\mathfrak{p}},\hat{\boldsymbol{\theta}}}(\mathcal{T}_T)$ is then an approximation of $\pi^*(\mathcal{T}_T) = \mathbb{P}(Y = 1|\mathcal{T}_T)$. It implies that the score function $\hat{f}$ can be viewed as an estimator of the function $2\pi^* - 1$. Besides, the estimated parameter $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1)$ provides estimate of parameter $\boldsymbol{\theta}^* = (\theta_0^*, \theta_1^*)$. Finally, the resulting classifier $\hat{g}$ is then naturally defined as

$$\widehat{g}(\mathcal{T}_T) = \mathbb{1}_{\{\widehat{f}(\mathcal{T}_T))\geq 0\}}.$$

Interestingly, under mild assumptions, the classifier $\hat{g}$ is consistent (see [11] for more details). This algorithm is referred in the following as `ERM`.

3.3 Goodness-of-fit test

One of the main strength of our method is the interpretability of the results. Indeed, beyond the classification task, our algorithm also provides a model for the behavior of the bats within each class. More precisely, our procedure provides estimates of the parameters of interests (θ_0^*, θ_1^*). To assess the quality of our procedure, given these estimates, we can perform a goodness-of-fit test as follows. The procedure is based on the Time-Rescaling Theorem given in [10]. It mainly implies that for the compensator of a process N with intensity λ, given for $t \geq 0$ by $\Lambda(t) = \int_0^t \lambda(s)\ ds$, that a.s., the transformed sequence $\{\tau_l = \Lambda(T_l)\}$ is a realization of a unit-rate Poisson process if and only if the original sequence $\{T_l\}$ is a realization from the point process N. This result can be applied to provide a goodness-of-fit test, and its proof can be found in [23]. The resulting test is performed for example in [14], and [6]. Let us explain the procedure. We fix a class $k \in \{0,1\}$ and consider the obtained estimates $\widehat{\theta}_k$. Then, the estimated compensator is

$$\Lambda_{\widehat{\theta}_k}(t) = \int_0^t \left(\lambda_{\widehat{\theta}_k}(s)\right)_+ ds. \tag{4}$$

Hence, we want to build the goodness-of-fit test defined by the null hypothesis H_0: "*the sequence of observations is a realization of the point process with intensity* $\lambda_{\widehat{\theta}_k}$". According to Time-Rescaling Theorem, under H_0, the sequence $\{\Lambda_{\widehat{\theta}_k}(T_{l+1}) - \Lambda_{\widehat{\theta}_k}(T_l)\}$ are increments of a unit-rate Poisson process, then are distributed according to an exponential distribution with parameter 1. Therefore to test H_0, we can perform the classical Kolmogorov-Smirnov test. In order to ensure that the Time-Rescaling Theorem applies, it is important to note that the estimation must be done on a part of the sample and the goodness-of-fit performed on the remaining data (see [25] for more details). For each class, we use this goodness-of-fit test to validate the obtained model. This procedure is described in Section 4.2.

Nevertheless, we are aware of the fact that it is difficult to conclude from one realization (*e.g.* one path) whether a point process is a Poisson process, mainly because, as it is explained in [25], the variations of the points can either be due to non-stationarity or to more complex dependency structures that cannot be studied on just one repetition. Besides, as the estimation of the model parameters is done on a sample and the test is realized on another sample, with the same distribution, it is expected that the rejection rate of H_0 is higher than it should.

4 Numerical experiments

The goal of this section is to investigate the performance of our method. Specifically, we evaluate both the modeling with Hawkes processes and the classification procedure, using synthetic data.

Our classification procedure, presented in Section 3.2, is referred as `ERM` in the following. Our method is implemented with `Python`. In particular, the optimization of the objective function given in Equation (3) is carried out by using the Stochastic Gradient Descent (`SGD`) optimizer of the `torch.optim` library.

As a benchmark, we compare the `ERM` algorithm with the random forests algorithm (named `RF`). Introduced in [8], the random forests procedure is a popular method that can be used in a wide range of applications, including time series prediction. The main issue in our framework to apply `RF` is that each observation does not have the same length. For overcoming this problem, a zero-padding step is performed, which means that each sequence whose length is less than the maximal length of the data set is zero-filled. Note that we used the function `RandomForestClassifier` implemented in the `scikit-learn` library with default parameters.

The simulation scheme is described in Section 4.1 while the evaluation of the procedure is presented in Section 4.2. Finally, the obtained results are provided and discussed in Section 4.3.

4.1 Simulation scheme

For the most comprehensive study possible, we consider three scenarios referred to as *Scenario 1* to *3*. In each scenario, the values of the parameters associated with the two classes are more or less close, which varies the difficulty of the corresponding classification task. For example, in *Scenario 1*, the dynamics materialized by the two classes are very distinct and therefore easy to dissociate. Note that *Scenario 3* corresponds to the case where class 0 describes a Poisson model (*e.g.* $\alpha_0 = 0$) whereas class 1 is a Hawkes model. Indeed, it is crucial to notice that the Poisson model is included in the model and thus the user does not have to choose between Poisson or Hawkes model in advance. Through this choice of scenarios, we propose to simulate a dynamics close to those expected in real data. The values of the parameters associated to each scenario are given in Table 2.

	Scenario 1	*Scenario 2*	*Scenario 3*
μ_0	1.0	1.0	1.0
μ_1	1.0	1.0	0.5
α_0	0.2	0.4	0.0
α_1	0.7	0.5	0.5
β_0	3.0	2.5	0.0
β_1	1.5	2.0	1.5

Table 2: *(Simulated data).* The panel of scenarios used to study the performance of the procedure. In each scenario, the associated set of parameters is displayed for both classes.

To assess the difficulty of the classification task, for each scenario, we derive the evaluation of the misclassification risk of the Bayes classifier with a data set of size 10000 over 20 repetitions. The results are further provided in Table 3. We also provide on Figure 5 a graph of the process with the conditional intensity process through time, for each scenario. In particular, from Figure 5, we

can see that the conditional intensities functions are quite different *w.r.t.* the classes in *Scenario 1* and *Scenario 3*. In these scenarios, the classes should be easy to discriminate. This is confirmed by the Bayes risk, which is respectively evaluated to 0.08 and 0.17. On the contrary, in *Scenario 2*, the conditional intensities functions are close, and the classes are overlapped which yields to a more difficult classification problem. It is confirmed by the Bayes risk, which is equal to 0.38.

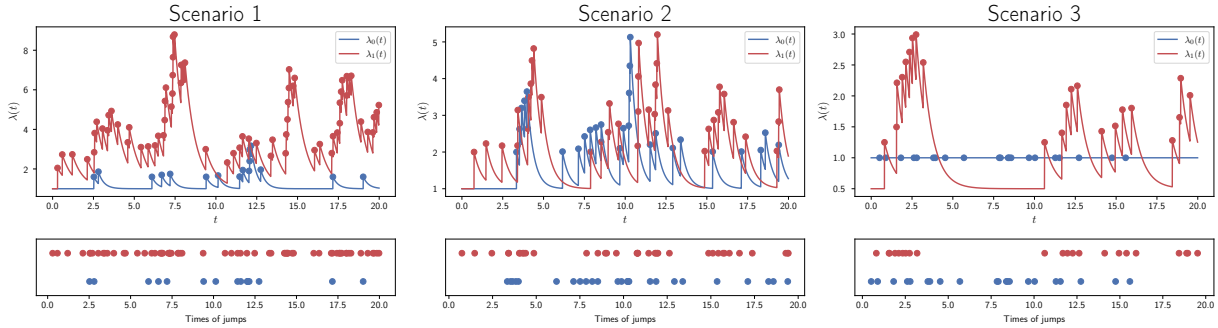

Figure 5: *(Simulated data).* Visual representation of the three scenarios by plotting for each class the conditional intensity and the path of the associated Hawkes process.

Concerning synthetic data generation, each path is simulated by cluster process representation algorithm. Introduced in [18], this method uses the branching properties of the self-exciting Hawkes process. Given a Hawkes process of baseline term μ and kernel function h (an exponential function here), the algorithm is the following. First, we simulate the arrivals of the immigrants, as a Poisson process of intensity function μ. Then for each immigrant arrived in T_ℓ, we generate its offsprings at future times $t > T_\ell$ which follow a Poisson process of intensity $h(t - T_\ell)$. Then, in the next step, the offsprings become the ancestors, and we continue iterating until there are no more successors. This algorithm is detailed in [21] for instance.

4.2 Performance evaluation

In this section, we investigate the performance of our procedure. Our methodology is evaluated according to the goodness-of-fit test presented in Section 3.3. In addition, the classification performance of `ERM` and `RF` are assessed by their error rate and compared to the one of the Bayes classifier.

Hereafter, we present the evaluation scheme that relies on Monte-Carlo repetitions. We fix $n_{\text{train}} = 300$, $n_{\text{test}} = 1000$, $T = 20$, $\mathfrak{p}^* = 0.5$. For each scenario described in Table 2, we repeat independently 20 times the following steps:

1. We simulate the data set $\mathcal{D}_{n_{\text{train}}}$ and $\mathcal{D}_{n_{\text{test}}}$;
2. From $\mathcal{D}_{n_{\text{train}}}$, we compute the estimates $\hat{\theta}_0$, $\hat{\theta}_1$ and build the classifiers $\hat{g}_{\texttt{ERM}}$, $\hat{g}_{\texttt{RF}}$;
3. Based on $\mathcal{D}_{n_{\text{test}}} = \{(\mathcal{T}_T^i, Y^i), i = 1, \dots, n_{\text{test}}\}$,
   (a) We evaluate the error rate of the classifiers `ERM` and `RF` using

$$\text{Err}_{\texttt{ERM}} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \mathbb{1}_{\{\hat{g}_{\texttt{ERM}}(\mathcal{T}_T^i) \neq Y^i\}}, \quad \text{and} \quad \text{Err}_{\texttt{RF}} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \mathbb{1}_{\{\hat{g}_{\texttt{RF}}(\mathcal{T}_T^i) \neq Y^i\}};$$

   (b) for each $i = 1, \dots, n_{\text{test}}$, we apply the goodness-of-fit test on the sequence $\mathcal{T}_T^i = \left(T_l^i\right)_{1 \leq l \leq N_T}$ by

i. Computing the estimated compensator, defined by Equation (4), based respectively on the true label Y_i and the predicted label $\widehat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)$:

$$\Lambda_{\widehat{\theta}_{Y_i}}(T_T^i) \text{ and } \Lambda_{\widehat{\theta}_{\widehat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)}}(T_T^i)$$

ii. Considering the test described in Section 3.3, from the increments of $\Lambda_{\widehat{\theta}_{Y_i}}(T_T^i)$, we get its associated p-value p_{Y_i}; from $\Lambda_{\widehat{\theta}_{\widehat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)}}(T_T^i)$, we obtain $p_{\widehat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)}$

(c) Finally, we consider the mean of the obtained p-values conditionally to the true classes or predicted classes; more precisely, for $k \in \{0,1\}$, let $n_k = \sum_{i=1}^{n_{\text{test}}} \mathbb{1}_{\{Y^i=k\}}$ and $\hat{n}_k = \sum_{i=1}^{n_{\text{test}}} \mathbb{1}_{\{\hat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)=k\}}$; we compute

$$\bar{p}_k = \frac{1}{n_k}\sum_{i=1}^{n_{\text{test}}} p_{Y_i}\mathbb{1}_{\{Y_i=k\}} \text{ and } \widehat{\bar{p}}_k = \frac{1}{\hat{n}_k}\sum_{i=1}^{n_{\text{test}}} p_{\widehat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)}\mathbb{1}_{\{\widehat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)=k\}}$$

(d) From the p-values, for each k, we also evaluate the *acceptance* rates for the 5% significance level test

$$\mathcal{A}_k = \frac{1}{n_k}\sum_{i=1}^{n_{\text{test}}} \mathbb{1}_{\{p_{Y_i}\geq 0.05\}}\mathbb{1}_{\{Y_i=k\}} \text{ and } \hat{\mathcal{A}}_k = \frac{1}{\hat{n}_k}\sum_{i=1}^{n_{\text{test}}} \mathbb{1}_{\{p_{\widehat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)}\geq 0.05\}}\mathbb{1}_{\{\widehat{g}_{\texttt{ERM}}(\mathcal{T}_T^i)=k\}}.$$

From Step 3 of the above procedure, we provide the mean of standard deviations of the obtained error rates, p-values, and acceptance rates.

	Bayes	ERM	RF
Scenario 1	0.08 (0.00)	0.08 (0.01)	0.09 (0.01)
Scenario 2	0.38 (0.00)	0.40 (0.02)	0.44 (0.02)
Scenario 3	0.17 (0.00)	0.17 (0.01)	0.27 (0.01)

Table 3: *(Simulated data).* Empirical error Err over 20 Monte-Carlo repetitions for each classifier in the three scenarios. The standard deviation is provided between parentheses.

		p-value $\bar{p}$	Acceptance Rate $\mathcal{A}$	p-value $\widehat{\bar{p}}$	Acceptance Rate $\hat{\mathcal{A}}$
Scenario 1	Class 0	0.49 (0.01)	0.94 (0.01)	0.50 (0.02)	0.96 (0.01)
	Class 1	0.49 (0.02)	0.95 (0.02)	0.52 (0.02)	0.96 (0.02)
Scenario 2	Class 0	0.46 (0.01)	0.92 (0.01)	0.54 (0.02)	0.97 (0.01)
	Class 1	0.48 (0.02)	0.93 (0.02)	0.46 (0.04)	0.91 (0.03)
Scenario 3	Class 0	0.45 (0.03)	0.92 (0.02)	0.46 (0.02)	0.93 (0.02)
	Class 1	0.40 (0.02)	0.89 (0.01)	0.43 (0.02)	0.90 (0.02)

Table 4: *(Simulated data).* Mean p-values and acceptance rate for a 5% significance level test over 20 Monte-Carlo repetitions in the three scenarios. The standard deviation over repetitions is displayed.

4.3 Results

This section is devoted to the discussion of the obtained results provided in Table 3, and 4.

First, from Table 3, we observe that for all scenarios `ERM` performs as well as the Bayes classifier and better than `RF` in terms of error rate. Besides, the standard deviations are small. In particular, for *Scenario 3*, the error rate of `RF` evaluated at 0.3 is much larger than the one of `ERM` evaluated at 0.17. In this scenario, the difficulty comes from the fact that the class 0 is distributed according to a Poisson process ($\alpha_0 = 0$) while the class 1 comes from a Hawkes process ($\alpha_1 > 0$). It highlights that our procedure is also able to detect Poisson dynamics (which corresponds to $\alpha_0 = 0$) which is not the case for `RF`.

Secondly, from the results provided in Table 4, we can see that the mean of the p-values conditionally to the true labels $(\bar{p}_k)_{k\in\{0,1\}}$ are closed to 0.5 and the acceptance rates are closed to 0.95, especially for Scenario 1. Indeed, under H_0 with the true parameters θ_0^*, θ_1^*, the p-values of goodness-of-fit test should be distributed according to a uniform distribution on $[0, 1]$. Since we apply the goodness-of-fit test with estimated parameters $\hat{\theta}_0$, $\hat{\theta}_1$, we only expect that those values are closed to the theoretical ones. We can notice that the p-valuers are also close to 0.5 on *Scenario 2* with the estimated label, which should be the most difficult task of classification. However, we can note that for *Scenario 3* those values are slightly biased (approximately 0.4 for the p-value and 0.89 for the acceptance rate). This could be due to the fact that, in this scenario, the estimation of the parameter is more difficult. Finally, for *Scenario 1* and 3, in Table 4, the obtained values with the predicted labels are almost the same as with true labels. It assesses the good performance of our procedure on these scenarios.

5 Application to the data

In this section, we apply the methodology described in Section 3 to the bats data set. First, in Section 5.1 we assess the performance of our procedure on labeled data. Then, in Section 5.2 we focus on the unlabeled sample and present the obtained results.

5.1 Classification on labeled data

In this section, we investigate the performance of our methodology on the labeled data set $\mathcal{D}_n^L$. In particular, as in Section 4.2 we compare our classification procedure with `RF` and apply the goodness-of-fit test presented in Section 3.3. The evaluation scheme, carried out through Monte-Carlo repetitions, is as follows. We repeat independently 20 times the following steps:

1. we randomly choose 75% of the data set $\mathcal{D}_n^L$ for training and the remaining 25% for test. We denote respectively those subsample by $\mathcal{D}_{n_{\text{train}}}^L$ and $\mathcal{D}_{n_{\text{test}}}^L$;
2. based on $\mathcal{D}_{n_{\text{train}}}^L$, we compute the estimates $\hat{\theta}_0$, $\hat{\theta}_1$ and build the classifiers $\hat{g}_{\texttt{ERM}}$, $\hat{g}_{\texttt{RF}}$;
3. on $\mathcal{D}_{n_{\text{test}}}^L$, we apply the step 3) of the procedure describe in Section 4.2.

The results are provided in Table 5 and in Figure 6 where the confusion matrix is displayed. Several comments can be made from these results.

Firstly, we obtain a satisfactory rate of correct classification equal to 0.68 (0.04). Let us point out that the `RF` algorithm achieves a rate of 0.67 (0.02) of correct classification which is similar to the `ERM` classifier. However, the lack of interpretation of the `RF` procedure is an important limitation to consider this method for practical purpose.

Secondly, Figure 6 shows that the predictions are significantly better on class 1 than on class 0. In particular the true positive rate is of order 0.79 which is quite good. Besides, the `ERM` classifier

tends to predict more often class 1, that is foraging behavior. It can be interpreted as the fact that a transit site with a bit of foraging is assigned as moderate hunting site by our procedure.

Lastly, from Table 5, we can see that the means of the acceptances rates over the Monte-Carlo simulation are around 0.4 except for $\hat{\mathcal{A}}$ where the rate is of order 0.66. It is obviously less than on synthetic data but it is sufficiently high to confirm that the proposed modeling is relevant to discriminate the sites between commuting and foraging. Similar comment can be made regarding to the means of the p-values. In particular, looking at the values of the estimates of the parameters θ_0^*, and θ_1^*, we observe that the two classes are mainly discriminated by the arrival intensity. We obtain $\hat{\alpha}_0 = 0.26$ for the class 0, and $\hat{\alpha}_1 = 0.98$ for the class 1. Hence, as expected, the emission of echolocation calls tends to promote the foraging. Finally, for the class 0, we also can notice that the p-value $\hat{\bar{p}}$ obtained with the predicted labels is much larger than the p-value $\bar{p}$ obtained with the true labels. The same comment holds for the acceptance rate. It is mainly due to the fact that our classification procedure predicts the class 1 more often, especially for the sites which have ambiguity between the classes, for example high traffic commuting sites or commuting sites with some foraging behavior. Thus, these sites may not be included in the computation of $\hat{\bar{p}}$ and $\hat{\mathcal{A}}$, which could explain the higher values.

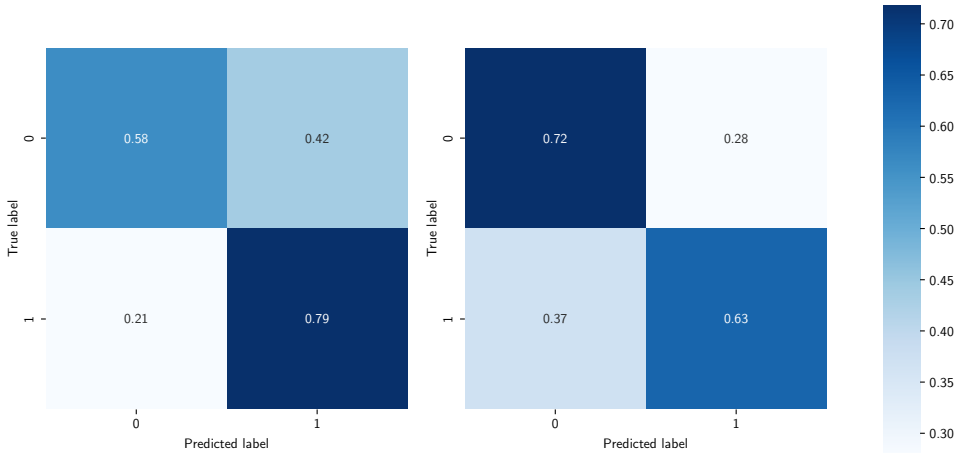

Figure 6: *(Bats data set).* On $\mathcal{D}_n^L$. Confusion matrix of prediction error over 20 Monte-Carlo repetitions on labeled data. The true label are those given by the metric $\mathcal{I}$. On the left are represented the predictions of the `ERM` procedure, percentage of correct classification: 68.13% (4.15). On the right are represented the predictions of the `RF` procedure, percentage of correct classification: 67.35% (2.21).

	p-value $\bar{p}$	Acceptance Rate $\mathcal{A}$	p-value $\widehat{\bar{p}}$	Acceptance Rate $\hat{\mathcal{A}}$
Class 0	0.17 (0.03)	0.42 (0.06)	0.26 (0.06)	0.66 (0.11)
Class 1	0.14 (0.04)	0.44 (0.08)	0.15 (0.03)	0.45 (0.07)

Table 5: *(Bats data set).* Mean p-values and acceptance rate over 20 Monte-Carlo repetitions on labeled data $\mathcal{D}_n^L$. The standard deviation over repetitions is displayed. For each classes, the average values of the estimated parameters are $\widehat{\theta}_0 = (\widehat{\mu}_0 = 1.02, \widehat{\alpha}_0 = 0.26, \widehat{\beta}_0 = 0.92)$ and $\widehat{\theta}_1 = (\widehat{\mu}_1 = 1.09, \widehat{\alpha}_1 = 0.98, \widehat{\beta}_1 = 1.08)$.

5.2 Prediction on unlabeled data

In this section, we provide predictions for the observations of the unlabeled data $\mathcal{D}_m^U$ ($m = 423$). We recall that, as illustrated in Figure 4, the unlabeled sites are the sites for which the value of the metric $\mathcal{I}$ belongs to $[\tau_1, \tau_2]$. Let us also emphasize that on these sites, bats have mixed behavior between commuting and foraging. Therefore prediction of the majority behavior on these sites is a tricky task.

We apply our methodology on $\mathcal{D}_m^U$. For each site i we return the predicted label $\hat{Y}_i$. The classifier `ERM` predicts foraging behavior (class 1) for 320 sites and commuting behavior (class 0) for 103 sites. Then, with the predicted label $\hat{Y}_i$, we perform the goodness-of-fit test described in Section 4.2. Note that, since the true label is unknown, we only compute $\widehat{\widehat{p}}$ and $\widehat{\hat{\mathcal{A}}}$. The results are provided in Table 6. We observe that the $\widehat{\widehat{p}}$ and $\widehat{\hat{\mathcal{A}}}$ are slightly better for the class 1 which corresponds to the foraging behavior. It is also interesting to note that we obtain similar values as on the labeled sample.

	p-value $\widehat{\widehat{p}}$	Acceptance Rate $\hat{\mathcal{A}}$
Class 0	0.15	0.43
Class 1	0.21	0.49

Table 6: *(Bats data set).* Mean p-values and acceptance rate for a 5% significance level test on unlabeled real data $\mathcal{D}_n^U$.

Discussion. To discuss the obtained results, we also propose to investigate the predictions given by our method *w.r.t.* two environmental covariates of interest, Light and Water (see Table 1). This choice seems to be appropriate insofar as they represent factors likely to influence bat behavior. Indeed, artificial light spot may create prey aggregation and thus favor foraging behavior for fast-flying species such as the Common Pipistrelle (see [1]). Same comment holds for soft water. Indeed the presence of soft water is a highly favorable condition for insect reproduction.

To illustrate the interplay between the obtained predictions and the two covariates of interest, for each site $i \in \mathcal{D}_m^U$ with observed sequence of calls $\mathcal{T}_T^i$, we represent in Figure 7 the *predictive* probability $\pi_{\widehat{\mathfrak{p}},\hat{\boldsymbol{\theta}}}(\mathcal{T}_T^i)$ (estimation of $\mathbb{P}(Y_i = 1|\mathcal{T}^i)$) as a function of covariates values on this site. For both covariates, we observe that, above a threshold, only foraging behavior is predicted by our procedure. The predictions on those sites seems relevant. Indeed, as already said, when the covariates Light and/or Water take high value, it strongly implies foraging behavior. For the sites for which the covariates values are moderate, we can see that the predictive probabilities are close to 0.5, thus indicating uncertainty of the predictions.

6 Conclusion and perspectives

In the present work, we propose a novel procedure for ecologists that is designed to classify the majority behavior of a site into two classes: foraging sites and commuting sites. We apply our methodology on both synthetic and real data sets. The proposed method exhibits good performance on synthetic data while the obtained results on real data set are promising. In addition to the predicted classes, our procedure outputs predictive probabilities that provide an index of confidence for the foraging behavior. Such an index is a major point of interest for ecologists, as it enables to interpret the activity of sites with mixed behavior between pure foraging and pure commuting. It is important to note that predictions as well as predictive probabilities do not

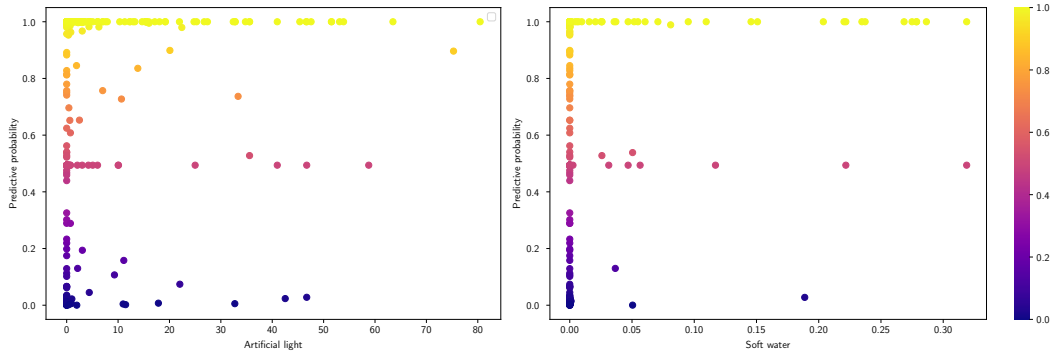

Figure 7: *(Bats data set)*. For each site $i \in \mathcal{D}_m^U$, Predictive probability $\pi_{\widehat{\mathfrak{p}},\hat{\boldsymbol{\theta}}}(\mathcal{T}_T^i)$ represented as a function of Artificial light for the Figure on the left and as a function of Soft water for the Figure on the right.

depend on environmental covariates. Therefore, our procedure may be used to highlight scale effects. Furthermore, our proposed methodology could be extended to any supervised classification problem for temporal dependent data that can be modeled using a continuous time point process.

Home range size is known to affect population dynamics [see 19]. In particular, as the common pipistrelle has a small home range, leading to short periods of commuting, it is difficult to discriminate between foraging and commuting. Although our procedure shows promising result with this species, it could be very interesting to apply it to other species of bats which have a larger home range. As a consequence, on a one-night scale, we expect to observe a more marked majority behavior for these species. Keeping this in mind, we could consider the western barbastelle (*Barbastella barbastellus*) living in forest habitat or Daubenton's myotis (*Myotis daubentonii*) whose roosts are close to water sources.

Besides, given that several species are recorded at each site, it could be a guideline for further research to model the call sequence of each species using multivariate Hawkes processes. More precisely, each component of the multivariate Hawkes process would model the temporal distribution of calls associated with a species. Such a model would not only take into account the past activity of a single species but also the effects of inter-species cooperation. To model the effect of competition between species, inhibition influences may be considered in the model. To this end, we can allow the intensity parameter α to be negative by adding a link function in the intensity. The extension of our procedure to this framework could be the object of further works.

Finally, from ecological perspective, it would be interesting to consider the ranking problem. Indeed, we could rank the sites according to the foraging behavior intensity. It would allow direct comparisons across sites while separating habitats selected for foraging from those selected for commuting. It would be particularly helpful when a stressor could have opposite effects on those two behaviors, e.g. artificial light facilitating foraging for some species but having negative effect on commuting activity [see 1].

Data availability statement

Both data sets and code can be found at here.

Acknowledgements

This work has been supported by the Chair "Modélisation Mathématique et Biodiversité" of Veolia-École polytechnique-Museum national d'Histoire naturelle-Fondation X, through a Ph.D. scholarship. The project is also part of the 2022 DAE 103 EMERGENCE(S) - PROCECO project supported by Ville de Paris.

References

[1] C. Azam, I. Le Viol, Y. Bas, G. Zissis, A. Vernet, J.F. Julien, and C. Kerbiriou. Evidence for distance and illuminance thresholds in the effects of artificial lighting on bat activity. *Landscape and Urban Planning*, 175:123–135, 2018.

[2] E. Bacry, M. Bompaire, S. Gaïffas, and J-F Muzy. Sparse and low-rank multivariate hawkes processes. *Journal of Machine Learning Research*, 21(50):1–32, 2020.

[3] E. Bacry, I. Mastromatteo, and J.F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 01(01):1550005, 2015.

[4] A. Baouan, S. Coustou, M. Lacome, S. Pulido, and M. Rosenbaum. Crediting football players for creating dangerous actions in an unbiased way: the generation of threat (got) indices. *arXiv preprint arXiv:2304.05242*, 2023.

[5] P.L. Bartlett, M.I. Jordan, and J.D. Mcauliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156, 2006.

[6] A. Bonnet, C. Dion-Blanc, F. Gindraud, and S. Lemler. Neuronal network inference and membrane potential model using multivariate hawkes processes. *Journal of Neuroscience Methods*, 372:109550, 2022.

[7] A. Bonnet, M. Martinez Herrera, and M. Sangnier. Inference of multivariate exponential hawkes processes with inhibition and application to neuronal activity. *Statistics and Computing*, 33(4):91, 2023.

[8] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[9] E.R. Britzke, E.H. Gillam, and K.L. Murray. Current state of understanding of ultrasonic detectors for the study of bat ecology. *Acta Theriologica*, 58:109–117, 2013.

[10] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer New York, NY, 2003.

[11] C. Denis, C. Dion-Blanc, and L. Sansonnet. Multiclass classification for hawkes processes. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pages 539–547. PMLR, 2022.

[12] C. Dietz, O. von Helversen, D. Nill, P. Lina, and A. Hutson. *Bats of Britain, Europe and Northwest Africa*. A & C Black London, 2009.

[13] S. Donnet, V. Rivoirard, and J. Rousseau. Nonparametric bayesian estimation for multivariate hawkes processes. *Annals of Statistics*, 48(5):2698–2727, 2020.

[14] P. Embrechts, T. Liniger, and L. Lin. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48(A):367–378, 2011.

[15] D.R. Griffin, F.A. Webster, and C.R. Michael. The echolocation of flying insects by bats. *Animal Behaviour*, 8(3):141–154, 1960.

[16] A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[17] A.G. Hawkes. Cluster models for earthquakes-regional comparisons. *Bull. Int. Stat. Inst.*, 45(3):454–461, 1973.

[18] A.G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[19] A. Laforge, F. Archaux, A. Coulon, C. Sirami, J. Froidevaux, N. Gouix, S. Ladet, H. Martin, K. Barré, C. Roemer, F. Claireau, C. Kerbiriou, and L. Barbaro. Landscape composition and life-history traits influence bat movement and space use: Analysis of 30 years of published telemetry data. *Global Ecology and Biogeography*, 30(12):2442–2454, 2021.

[20] C.F. Moss and A. Surlykke. Auditory scene analysis by echolocation in bats. *The Journal of the Acoustical Society of America*, 110:2207–2226, 2001.

[21] J. Møller and J.G. Rasmussen. Perfect simulation of Hawkes processes. *Advances in Applied Probability*, 37(3):629–646, 2005.

[22] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[23] F. Papangelou. Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society*, 165:483–506, 1972.

[24] Z. Qu, C. Lyu, and C. Chi. Mush: Multi-stimuli hawkes process based sybil attacker detector for user-review social networks. *IEEE Transactions on Network and Service Management*, 2022.

[25] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience*, 4(1):3, 2014.

[26] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in neurosciences via hawkes processes. In *2013 IEEE global conference on signal and information processing*, pages 317–320. IEEE, 2013.

[27] C. Roemer, J.F. Julien, P.P. Ahoudji, J.M. Chassot, M. Genta, R. Colombo, G. Botto, C. A Negreira, B.A. Djossa, R.K. Ing, et al. An automatic classifier of bat sonotypes around the world. *Methods in Ecology and Evolution*, 12(12):2432–2444, 2021.

[28] E.G. Rowse, S. Harris, and G. Jones. The switch from low-pressure sodium to light emitting diodes does not affect bat activity at street lights. *PLOS ONE*, 11:1–14, 2016.

[29] H.U. Schnitzler and E.K.V. Kalko. Echolocation by insect-eating bats. *BioScience*, 51(7):557–569, 2001.

[30] J. Worrall, R. Browning, P. Wu, and K. Mengersen. Fifty years later: new directions in hawkes processes. *SORT (Statistics and Operations Research Transactions)*, 46(1):3–38, 2022.