




## RESEARCH ARTICLE

# Accounting for automated identification errors in acoustic surveys

Kévin Barré<sup>1,2</sup>  | Isabelle Le Viol<sup>1,2</sup> | Romain Julliard<sup>1</sup> | Julie Pauwels<sup>1</sup> |  
 Stuart E. Newson<sup>3</sup> | Jean-François Julien<sup>1</sup> | Fabien Claireau<sup>1,4,5</sup>  |  
 Christian Kerbiriou<sup>1,2</sup>  | Yves Bas<sup>1,6</sup>

<sup>1</sup>Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum national d'Histoire naturelle, Centre National de la Recherche Scientifique, Sorbonne Université, Paris, France; <sup>2</sup>Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum national d'Histoire naturelle, Concarneau, France; <sup>3</sup>British Trust for Ornithology, The Nunnery, Thetford, Norfolk, UK; <sup>4</sup>University of Greifswald, Zoology Institute and Museum, Greifswald, Germany; <sup>5</sup>Naturalia Environnement, Site Agroparc, Avignon, France and <sup>6</sup>Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), UMR 5175, CNRS - Université de Montpellier - Université Paul-Valéry Montpellier - EPHE, Montpellier, France

## Correspondence

Kévin Barré

Email: kevin.barre@edu.mnhn.fr

## Funding information

AgroSolutions; DIM ASTREA grants from  
 Region Ile-de-France

Handling Editor: Veronica Zamora-Gutierrez

## Abstract

1. Assessing the state and trend of biodiversity in the face of anthropogenic threats requires large-scale and long-time monitoring, for which new recording methods offer interesting possibilities. Reduced costs and a huge increase in storage capacity of acoustic recorders have resulted in an exponential use of passive acoustic monitoring (PAM) on a wide range of animal groups in recent years. PAM has led to a rapid growth in the quantity of acoustic data, making manual identification increasingly time-consuming. Therefore, software detecting sound events, extracting numerous features and automatically identifying species have been developed. However, automated identification generates identification errors, which could influence analyses which look at the ecological response of species. Taking the case of bats for which PAM constitutes an efficient tool, we propose a cautious method to account for errors in acoustic identifications of any taxa without excessive manual checking of recordings.
2. We propose to check a representative sample of the outputs of a software commonly used in acoustic surveys (Tadarida), to model the identification success probability of 10 species and two species groups as a function of the confidence score provided for each automated identification. Using this relationship, we then investigated the effect of setting different false positive tolerances (FPTs), from a 50% to 10% false positive rate, above which data are discarded, by repeating a large-scale analysis of bat response to environmental variables and checking for consistency in the results.
3. Considering estimates, standard errors and significance of species response to environmental variables, the main changes occurred between the naive (i.e. raw data) and robust analyses (i.e. using FPTs). Responses were highly stable between FPTs.

Christian Kerbiriou and Yves Bas are equal contribution as last authors.

4. We conclude it was essential to, at least, remove data above 50% FPT to minimize false positives. We recommend systematically checking the consistency of responses for at least two contrasting FPTs (e.g. 50% and 10%), in order to ensure robustness, and only going on to conclusive interpretation when these are consistent. This study provides a huge saving of time for manual checking, which will facilitate the improvement in large-scale monitoring, and ultimately our understanding of ecological responses.

#### KEYWORDS

bioacoustic, cautious threshold, chiroptera, error rate, false positives, passive acoustic monitoring, semi-automated identification, survey methods

## 1 | INTRODUCTION

With few exceptions, the rate of biodiversity loss does not appear to be slowing down (Butchart et al., 2010). In 2010, the 10th Conference of Parties to the Convention on Biological Diversity adopted a new 2011–2020 global Strategic Plan for Biodiversity, and in turn, the European Union launched a new Biodiversity Strategy (2011/2307). This strategy aims to halt biodiversity loss and the degradation of ecosystem services by 2020. Such objectives require large-scale and long-time studies using adapted monitoring methods for surveying and understanding biodiversity changes (Fisher, Frank, & Leggett, 2010) in response to anthropogenic pressures and environmental policies. The implementation of such studies is highly constrained by the time and cost induced. Interestingly, the development of new recording methods, such as passive acoustic monitoring (PAM), offers interesting possibilities and is taking an increasing place in monitoring (Gibb, Browning, Glover-Kapfer, & Jones, 2018).

The reduced costs of acoustic recorders and the huge increase in storage capacity have resulted in an exponential increase in the use of PAM on a very wide range of species groups within a few years (e.g. Froidevaux, Zellweger, Bollmann, & Obrist, 2014; Frommolt, 2017; Jeliaskov et al., 2016; Kalan et al., 2015; Nowacek, Christiansen, Bejder, Goldbogen, & Friedlaender, 2016; Stahlschmidt & Brühl, 2012). Such approaches are already widely used by researchers as well as by people working for environmental consultancies and government agencies for various biodiversity evaluations (Adams, Jantzen, Hamilton, & Fenton, 2012). PAM can be particularly useful to carry out surveys on cryptic taxa such as nocturnal fauna (Delpont, Kemp, & Ferguson, 2002; Jeliaskov et al., 2016; Newson, Evans, & Gillings, 2015), and to monitor pristine habitats which are otherwise difficult to access and survey by other approaches (Gasc, Sueur, Pavoine, Pellens, & Grandcolas, 2013). PAM is also mobilized in citizen science programs, for which it is an efficient tool for the implementation of large-scale biodiversity monitoring (Newson et al., 2015; Jeliaskov et al., 2016; Kerbiriou, Azam et al., 2018; Penone, Kerbiriou, Julien, Marmet, & Le Viol, 2018).

Despite rapid and exciting developments in acoustic monitoring, there have been substantial challenges in developing this technology

into a cost-effective, scalable monitoring tool. Perhaps the biggest and most complex issue facing acoustic monitoring has been the objective and statistical taxonomic identification of bioacoustic signals. With the arrival on the market of a new generation of affordable acoustic recorders, which allow for continuous recordings over several days, such volumes of acoustic data cannot be processed manually (Bas, Bas, & Julien, 2017; Newson et al., 2015).

In parallel to the development of PAM, several methods for detecting sound events, extracting numerous features and automatically identifying species have been developed (Adams et al., 2012; Bas et al., 2017; Britzke, Duchamp, Murray, Swihart, & Robbins, 2011; Ovaskainen, Moliterno de Camargo, & Somervuo, 2018; Parsons & Jones, 2000). However, automated identification software has been criticized due to significant error rates, suggesting cautious and limited use (Russo & Voigt, 2016; Rydell, Nyman, Eklöf, Jones, & Russo, 2017), which heavily reduces the advantages of automated algorithms. Nonetheless, authors have highlighted the potential for combining automated classifiers with manual validation to help overcome error risks associated with automated identification, and so saving a huge amount of work in reducing the extent of manual checking required (López-Baucells et al., 2019). Moreover, most available software provides confidence scores associated with each automated identification in the form of probabilities or other numerical indexes (Obrist, Boesch, & Fluckiger, 2004; Waters & Barlow, 2013), which unlike the error rate is not dependent of the relative abundance of the species. The confidence scores provided by software aim to be an indicator of the true success probabilities of automated identifications, and are strongly species-dependant. There is thus an implicit relationship between the error rate and confidence scores and most software manuals advocate using confidence thresholds below which data should be discarded to minimize the error rate, for example Tadarida (Bas et al., 2017), SonoChiro (Biotope, 2013) and BatClassify (Scott & Altringham, 2017). Regardless of the software used, the relationship between the error rate and confidence scores is an important part of the automated identification performance, yet it has never been directly assessed in previous methodological studies (Fritsch & Bruckner, 2014; Rydell et al., 2017). Consequently, the level at which confidence thresholds should be

set is unclear to most users, which has limited the use of automated identification in ecological studies. A threshold that is too cautious could lead to high-generated false negative rates (i.e. by discarding a large proportion of data containing true positives below a given confidence score), which could result in a lack of statistical power. In contrast, a threshold that is not cautious enough could lead to high false positive rates (i.e. fails in automated identifications), particularly through the inclusion of records of species which are most similar acoustically, which involve statistical noise. Moreover, errors (generated false negative rates or false positive rates) could also be spatially clustered by environmental conditions that alter the quality of the signal (Denzinger & Schnitzler, 2013), which potentially induce statistical biases in relation with confidence measure provided by the software. False positive rates and generated false negative rates thus induce different caveats for which there is not a unique way to set confidence thresholds. Given the wide range of taxa for which PAM is increasingly being used, there is a need to account for these caveats using a method generalizable to any acoustically surveyed taxa.

In this study, we propose a method for assessing the effect of using confidence thresholds in acoustic automated identification on the detection of species responses to environmental variables. This method can be applied to any acoustic taxa for which automated identification software and acoustic signature knowledge are already developed, and where confidence scores are provided. Taking the case of bats, we first manually checked a representative sample of a large number of bat recordings identified using an automated identification software (Tadarida; Bas et al., 2017) commonly used in bat studies (Barré, Le Viol, Bas, Julliard, & Kerbiriou, 2018; Barré, Le Viol, Julliard, Chiron, & Kerbiriou, 2017; Claireau et al., 2019; Pauwels et al., 2019; Pinaud, Claireau, Leuchtman, & Kerbiriou, 2018). Using this sample, we then modelled the identification success for 10 species and two species groups of bats in relation to the confidence score provided by the software. This allowed us to define the minimum confidence score needed to ensure a given false positive tolerance (FPT). We then examined how setting different FPTs, from 50% to 10% maximum false positive rate, above which data are discarded, may affect a statistical inference by repeating a large-scale analysis of the response of species and species groups activity to five environmental variables, and looking at consistency of the results among FPTs.

## 2 | MATERIALS AND METHODS

### 2.1 | Bat survey

We used an acoustic dataset collected previously to study the effect of wind turbines on bat activity (Barré et al., 2018) because it was based on a random sampling design with high variability and no confounding effects in terms of environmental variables (Figure S1). The following environmental variables are known as good predictors of bat activity: type of site that is, hedgerow versus open area habitat located at an average of 86 m (*SD*: 70 m) away from any

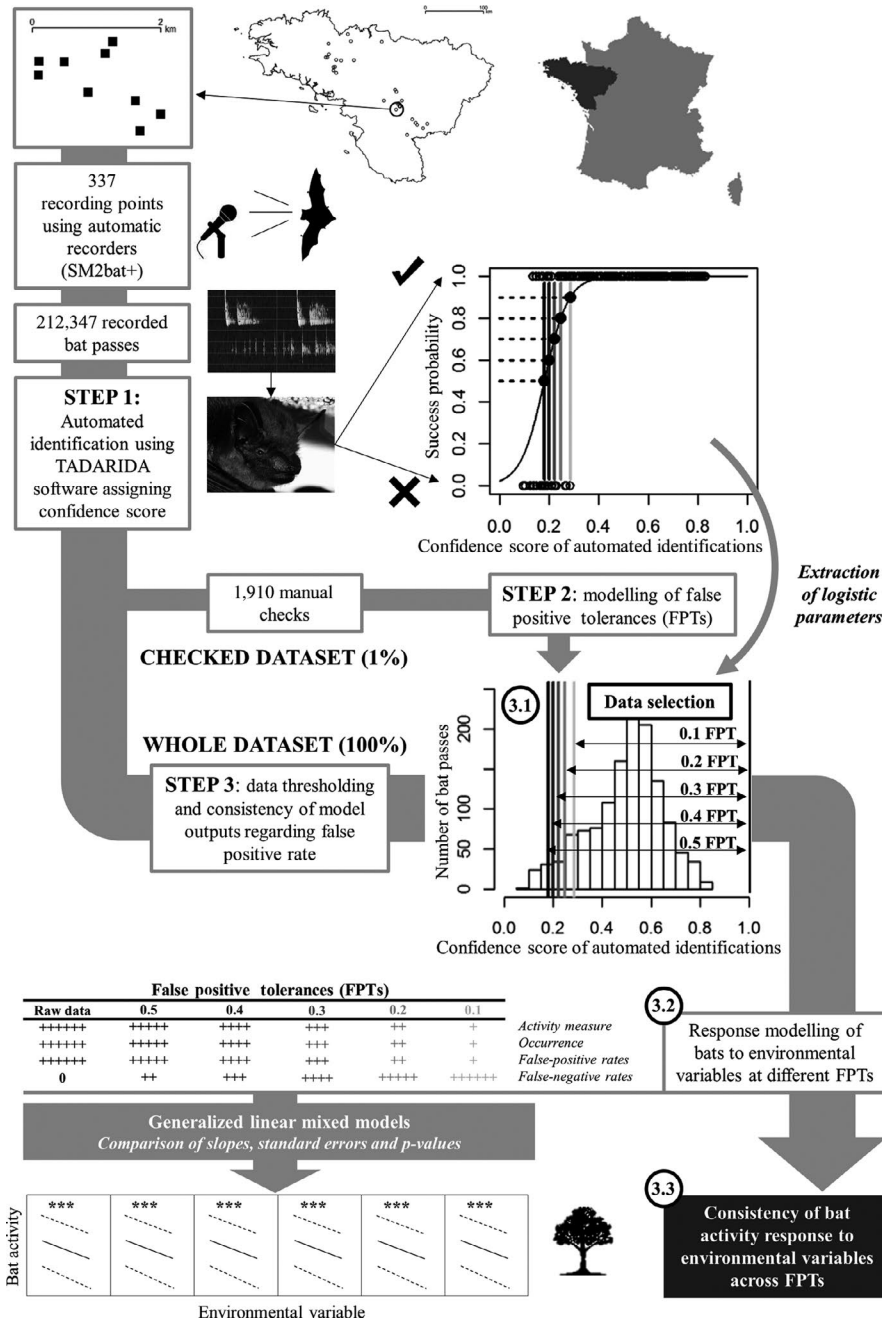
hedgerow (Lacoeuilhe, Machon, Julien, & Kerbiriou, 2016; Verboom & Huitema, 1997), the distance in meters to a forest ( $M = 700$ ,  $SD = 506$ ; Boughey, Lake, Haysom, & Dolman, 2011; Frey-Ehrenbold, Bontadina, Arlettaz, & Obrist, 2013), the distance to an urban area ( $M = 335$ ,  $SD = 170$ ; Azam, Le Viol, Julien, Bas, & Kerbiriou, 2016), the distance to a wetland ( $M = 579$ ,  $SD = 363$ ; Sirami, Jacobs, & Cumming, 2013; Santos, Rodrigues, Jones, & Rebelo, 2013) and the total length of hedgerows in meters within a 1,000 m radius ( $M = 3,439$ ,  $SD = 1,622$ ; Verboom & Huitema, 1997; Lacoeuilhe et al., 2016). The latter four variables presented important environmental variability, and a similar gradient between sites located close to hedgerows and those in open areas (Figure S1).

Bats were recorded at 337 sites (one complete night per site, with 207 sites close to hedgerows and 130 sites in open area) in northwest France (Figure 1) dominated by agriculture (82%) and forest (11%) areas. Recordings were carried out over 23 complete nights, recording from 30 min before sunset until 30 min after sunrise, from 7 September to 8 October 2016.

We simultaneously sampled 11–15 survey sites per night separated by at least 300 m (Figure 1). Echolocation calls were recorded using one automatic acoustic recorder per site survey (Song Meter SM2Bat+, Wildlife Acoustics Inc., Concord, MA, USA). The detectors automatically recorded all ultrasounds using predefined settings as recommended by the French bat monitoring program 'Vigie-Chiro' (trigger level set to 6 dB Signal Noise Ratio and set to continue recording until 2.0 s after last trigger event, 384 kHz sampling rate; for further details see Azam et al., 2018; Barré et al., 2018; Claireau et al., 2019; Pauwels et al., 2019). Whilst continuous recording is typically used for monitoring birds and several other species groups, for bats which echolocate at high frequency, and so produce heavy sound files, it is necessary to use triggered recording, to be able to manage and store the data and process the recordings. In addition, these trigger settings are very sensitive (6 dB of signal-to-noise ratio) and detect the majority of bats which would have been detected if recording were continuous. As recommended by Millon, Julien, Julliard, and Kerbiriou (2015), Kerbiriou, Azam et al. (2018) and Kerbiriou, Bas et al. (2018), we retained one bat pass per 5-s interval, which is the mean duration of all bat species passes.

### 2.2 | Step 1: manual checking of a subset of the data

The identification process performed in the first step was divided in two sub-steps (Figure 1). In the first sub-step, echolocation calls were detected and classified to the closest taxonomic level using the Tadarida software (Bas et al., 2017) (hereafter named as primary identification), which assigns a species and confidence score (continuous values between 0 and 1) to each recorded bat pass (212,347 in total). In the second sub-step, we selected a representative sample by a stratified random sampling of 25 primary identifications for each 0.1 class of confidence score (i.e. 10 classes in total) for each species and groups for manual checking, except for *Rhinolophus* species for which all identifications were selected due to their low number. We performed a double manual checking



**FIGURE 1** Schematic and chronological representation of the steps used to study the relationship between automated identification errors in acoustic data and the detected relationship between bat activity and environmental variables

(KB and YB) on this stratified random selection of 1,910 bat passes (hereafter named as checked dataset or manual checking), using BatSound© software (Pettersson Elektronik AB, Sweden) and Syrinx software (John Burt, Seattle, WA, USA) for 10 species and two groups (*Myotis* spp. and *Plecotus* spp.) (Table 1), by visual inspection and measurement of discriminating characteristics of calls on spectrograms (Barataud, 2015). Species groups were used for genera within which species are difficult to identify from one another, except for one species of *Myotis* spp., *Myotis nattereri*, for which echolocation calls are very characteristic (Barataud, 2015; Obrist et al., 2004). We made the choice to separate two species which are commonly grouped because of their frequency overlap: *Pipistrellus kuhlii* and *Pipistrellus nathusii*. We manually separated

these species by combining measurements of energy peak, final frequency, call duration, bandwidth and time between calls as discussed in Barataud (2015). In relatively open habitats like in our study, *P. nathusii* emit very commonly very short bandwidth, and higher frequencies than *P. kuhlii* when emitting such kind of calls (i.e. quasi-constant frequency). *P. kuhlii* very often use a short frequency modulation at the end of the call and this is very rare in *P. nathusii* calls. Finally, we randomly checked 500 sound files identified as not containing bats to assess missed bat events.

We assumed that manual checking provided the most conservative species assignments, which allowed us to accurately assign to each primary identification a true positive (i.e. a correct automated identification of the species), a false positive (i.e. a fail in automated

**TABLE 1** Total bat passes assigned to each species by the automated identification per confidence score classes, number of bat passes manually double checked and false positives noted (step 1 in Figure 1). See Table S1 for species composition in false positives

Species	Upper limits of confidence score classes of the automated identification										Total
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	
<i>Barbastella barbastellus</i>											
Total passes	3	52	144	242	297	671	940	1,312	1,596	578	5,835
Checked passes	3	25	25	25	25	25	25	25	25	25	228
False positives	3	5	1	0	0	0	0	0	0	0	3.9%
<i>Eptesicus serotinus</i>											
Total passes	1	55	102	149	268	461	218	79	10	0	1,343
Checked passes	1	25	25	25	25	25	25	25	9	0	185
False positives	1	13	7	0	0	0	0	0	0	0	11.4%
<i>Myotis nattereri</i>											
Total passes	9	166	211	223	225	411	269	180	247	47	1,988
Checked passes	9	9	3	6	8	2	2	10	23	25	97
False positives	8	5	1	2	1	0	0	0	0	0	17.5%
<i>Myotis spp</i>											
Total passes	20	534	815	770	701	1,708	1,132	445	258	47	6,430
Checked passes	20	25	25	25	25	25	25	25	25	25	245
False positives	19	14	6	6	4	0	0	0	0	0	20.0%
<i>Nyctalus leisleri</i>											
Total passes	3	47	41	33	11	8	9	1	0	0	153
Checked passes	3	25	25	25	11	8	9	1	0	0	107
False positives	2	16	14	13	4	0	0	0	0	0	45.8%
<i>Nyctalus noctula</i>											
Total passes	0	113	110	82	24	43	16	6	1	0	395
Checked passes	0	25	25	25	24	25	16	6	1	0	147
False positives	0	25	23	24	23	7	0	0	0	0	69.4%
<i>Pipistrellus kuhlii</i>											
Total passes	12	223	401	667	1,142	4,026	6,654	10,222	5,240	2	28,589
Checked passes	12	25	25	25	25	25	25	25	25	2	214
False positives	11	10	8	4	2	2	1	0	0	0	17.8%

(Continues)

TABLE 1 (Continued)

Species	Upper limits of confidence score classes of the automated identification										Total
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	
<i>Pipistrellus nathusii</i>											
Total passes	0	12	33	37	93	183	153	61	5	0	577
Checked passes	0	12	25	25	25	25	25	25	5	0	167
False positives	0	11	20	20	19	17	15	9	1	0	67.1%
<i>Pipistrellus pipistrellus</i>											
Total passes	2	303	760	1,636	3,298	8,311	14,221	27,205	83,744	28,024	167,504
Checked passes	2	25	25	25	25	25	25	25	25	25	227
False positives	1	2	0	1	1	0	0	1	0	0	2.6%
<i>Plecotus</i> spp											
Total passes	8	139	176	194	174	250	206	145	56	4	1,352
Checked passes	8	30	26	25	28	25	25	25	25	4	221
False positives	5	19	8	2	1	1	0	0	0	0	16.3%
<i>Rhinolophus ferrumequinum</i>											
Total passes	0	0	0	0	1	6	5	28	1	0	41
Checked passes	0	0	0	0	1	6	5	28	1	0	41
False positives	0	0	0	0	0	0	0	0	0	0	0.0%
<i>Rhinolophus hipposideros</i>											
Total passes	0	1	1	10	8	16	26	62	4	0	128
Checked passes	0	1	1	10	8	16	26	62	4	0	128
False positives	0	1	1	7	1	0	0	0	0	0	7.8%

identification of the species) or a false negative (i.e. defined in this study as a pass of the species automatically identified as another one) in the checked dataset.

The efficiency of the automated identification may be spatially heterogeneous due to habitat structure (Denzinger & Schnitzler, 2013). We tested for the dependence of false positives (i.e. a binomial response variable: failure or success of the automated identification) and false negative ones (i.e. a binomial response variable: automatically identified as another species or correct identification) on the five tested environmental variables. We performed generalized linear mixed models (binomial response variables; *logit* link) with the environmental variables as explanatory variables, using date as random effect to control for inter-night variations.

### 2.3 | Step 2: false positive rate modelling

The success probability, defined as the success or failure of the automated species identification, was used as the response variable to perform generalized linear models (binomial response variable; *logit* link) using the confidence score provided by the automated identification software as the explanatory variable (see step 2 in Figures 1 and 2). Using these models, we could predict the confidence score corresponding to a given success probability of the automated identification. Thus, predicted confidence score constitutes the minimum one required to ensure a given false positive tolerance (FPT, i.e. one minus the success probability) in the whole dataset (i.e. including all checked and non-checked primary identifications; Figure 1; Table 2). We selected all FPTs starting from the highest acceptable one (0.5, i.e. a maximum false positive rate of 50%, which expected to give an approximately balanced number of false negatives and false positives) to the lower one (0.1, i.e. a maximum false positive rate of 10%) by 0.1 classes (i.e. 0.5, 0.4, 0.3, 0.2 and 0.1 FPTs).

### 2.4 | Step 3: data thresholding and consistency of model outputs regarding false positive rate

After predicting the required confidence score to ensure a given FPT in the automated identification, we filtered the whole dataset on the five predicted confidence scores corresponding to the five FPT (see step 3 in Figures 1 and 3; Table 2). This allowed us to calculate for each FPT in the whole dataset, the remaining number of bat passes, occurrences and an estimation of false positive rate and generated false negative rate by reducing the FPT (Table 2). In order to assess the trade-off between false positive rates and generated false negative rates generated by reducing FPT, for each FPT, we estimated for the whole dataset the false positive rate (i.e. incorrect primary identifications) and generated false negative rate (i.e. as a consequence of discarding true positives because of reducing FPT) from equations used to model the false positive rate in step 2. For each bat pass BP of a given species *S*, we first computed the probability of there being a true positive (TP, equation 1) and a false positive (FP, equation 2) as follows:

$$TP_{BP,S} = \frac{\exp(ax+b)}{1 + \exp(ax+b)}, \quad (1)$$

$$FP_{BP,S} = 1 - TP_{BP,S}, \quad (2)$$

where *a* corresponds to the estimated parameter from the logistic regression between manual checking (i.e. the response variable: success/fail in automated identification; step 2 in Figures 1 and 2) with the confidence score provided by the software (i.e. the explanatory variable), *x* is the confidence score of the bat pass provided by the automated identification software and *b* is the intercept of the logistic regression (Figure S2).

This allowed us to estimate the generated false negative rate (FNR, Equation 3) for a given species *S* and a given threshold of false positive tolerance FPT in the whole dataset, by averaging all probabilities to have a true positive TP from bat passes BP discarded by reducing FPT (i.e. between the targeted FPT and the maximum FPT of 1) as follows:

$$FNR_{S,FPT} = \frac{\sum_{FPT}^{FPT=1} TP_{BP,S}}{n_{BP,S}}, \quad (3)$$

where *n* is the total number of bat passes BP of the species *S*.

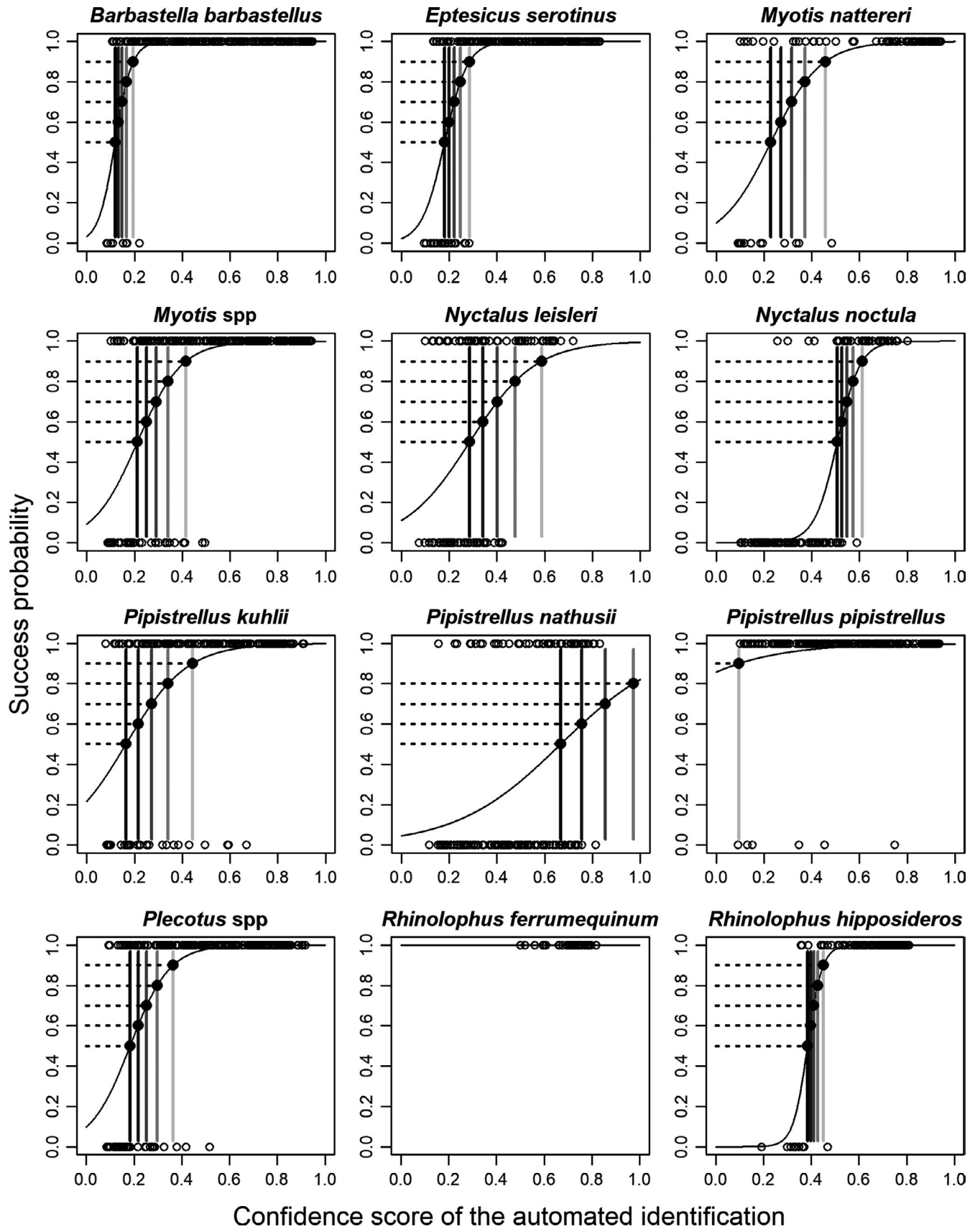
We were also able to estimate the false positive rate (FPR, Equation 4) for a given species *S* and a given threshold of false positive tolerance FPT in the whole dataset, by averaging probabilities to have a false positive FP from bat passes BP between the minimum FPT (i.e. zero) and the targeted FPT as follows:

$$FPR_{S,FPT} = \frac{\sum_{FPT=0}^{FPT} FP_{BP,S}}{n_{BP,S,FPT}}, \quad (4)$$

where *n* is the number of bat passes BP between the minimum FPT (i.e. zero tolerance of false positives) and the targeted FPT of a given species *S*.

Finally, we evaluated the automated classification efficiency by drawing receiver operating characteristic (ROC) curves between confidence scores of presences and absences of each species, and computing area under curve (AUC) with the R package `PRROC` (Figure S3).

For each species and species groups, we then performed generalized linear mixed models (GLMM, R package `LME4`) using as a response variable the number of bat passes filtered on one of the five FPTs or the raw number of primary identifications without thresholding (i.e. whole dataset) (six GLMMs in total performed on 0.5, 0.4, 0.3, 0.2, 0.1 FPTs and on the whole dataset). Environmental variables were included as fixed effects, among which quantitative ones were scaled. According to the sampling design (i.e. 11–15 simultaneous recording sites per night), we included date as a random effect to control for inter-night variation in weather conditions and landscape context. We applied a Poisson error or a Negative binomial error distribution to GLMMs in order to minimize issues in the overdispersion ratio in models (i.e. as close as possible to 1; Zuur, Ieno, Walker, Saveliev, & Smith, 2009). All explanatory variables showed



**FIGURE 2** Logistic regressions between the success probability and the confidence score of the automated identification. The success probability was predicted from a subset manually checked assigning a success or a failure in automated identifications. Horizontal dotted lines show success probabilities in automated identification used for thresholding (i.e. false positive tolerances: 0.5, 0.4, 0.3, 0.2 and 0.1) to remove data in the total dataset below the corresponding confidence scores (vertical lines)

**TABLE 2** Minimum confidence scores needed to ensure false positive tolerances (step 2 in Figure 1), associated changes in the number of bat passes, the occurrence (presence rate among sites), the estimated false positive rate and the generated false negative rate estimated for the whole dataset (212,347 bat passes; step 3 in Figure 1)

Species	False positive tolerance					
	Raw data	0.5	0.4	0.3	0.2	0.1
<i>Barbastella barbastellus</i>						
Confidence score	/	0.119	0.133	0.148	0.167	0.195
No. of bat passes	5,835	5,828	5,824	5,822	5,809	5,787
Occurrences	0.694	0.694	0.694	0.694	0.694	0.694
Estimated false positive rate	0.003	0.002	0.002	0.002	0.001	0.001
Estimated false negative rate	0	<0.001	<0.001	0.001	0.003	0.006
<i>Eptesicus serotinus</i>						
Confidence score	/	0.180	0.200	0.221	0.246	0.285
No. of bat passes	1,343	1,297	1,287	1,273	1,255	1,205
Occurrences	0.373	0.339	0.336	0.333	0.324	0.312
Estimated false positive rate	0.044	0.022	0.019	0.015	0.012	0.006
Estimated false negative rate	0	0.011	0.016	0.023	0.031	0.065
<i>Myotis nattereri</i>						
Confidence score	/	0.229	0.271	0.317	0.373	0.458
No. of bat passes	1,986	1,759	1,659	1,562	1,436	1,239
Occurrences	0.688	0.648	0.624	0.609	0.578	0.529
Estimated false positive rate	0.136	0.081	0.064	0.049	0.034	0.021
Estimated false negative rate	0	0.036	0.059	0.087	0.132	0.199
<i>Myotis spp.</i>						
Confidence score	/	0.212	0.250	0.291	0.341	0.416
No. of bat passes	6,428	5,783	5,483	5,135	4,747	4,173
Occurrences	0.798	0.792	0.786	0.774	0.765	0.716
Estimated false positive rate	0.145	0.092	0.073	0.054	0.038	0.024
Estimated false negative rate	0	0.036	0.062	0.099	0.145	0.219
<i>Nyctalus leisleri</i>						
Confidence score	/	0.286	0.342	0.402	0.476	0.587
No. of bat passes	153	67	43	28	22	12
Occurrences	0.211	0.138	0.104	0.070	0.055	0.031
Estimated false positive rate	0.502	0.305	0.222	0.149	0.115	0.075
Estimated false negative rate	0	0.193	0.279	0.337	0.370	0.425
<i>Nyctalus noctula</i>						
Confidence score	/	0.507	0.527	0.548	0.574	0.613
No. of bat passes	395	61	50	41	29	22
Occurrences	0.220	0.080	0.067	0.058	0.046	0.040
Estimated false positive rate	0.850	0.212	0.158	0.120	0.066	0.042
Estimated false negative rate	0	0.029	0.044	0.054	0.082	0.097
<i>Pipistrellus kuhlii</i>						
Confidence score	/	0.164	0.216	0.272	0.341	0.444
No. of bat passes	28,588	28,456	28,305	28,077	27,737	26,854
Occurrences	0.899	0.899	0.890	0.884	0.881	0.875
Estimated false positive rate	0.033	0.030	0.028	0.026	0.023	0.019
Estimated false negative rate	0	0.002	0.005	0.010	0.019	0.045

(Continues)

TABLE 2 (Continued)

Species	False positive tolerance					
	Raw data	0.5	0.4	0.3	0.2	0.1
<i>Pipistrellus nathusii</i>						
Confidence score	/	0.668	0.756	0.853	0.971	/
No. of bat passes	577	101	18	0	0	0
Occurrences	0.404	0.116	0.031	0.000	0.000	0.000
Estimated false positive rate	0.623	0.437	0.370	/	/	/
Estimated false negative rate	0	0.275	0.355	0.377	/	/
<i>Pipistrellus pipistrellus</i>						
Confidence score	/	0.000	0.000	0.000	0.000	0.096
No. of bat passes	167,503	167,503	167,503	167,503	167,503	167,502
Occurrences	0.954	0.954	0.954	0.954	0.954	0.954
Estimated false positive rate	0.007	0.007	0.007	0.007	0.007	0.007
Estimated false negative rate	0.000	0.000	0.000	0.000	0.000	0.000
<i>Plecotus</i> spp.						
Confidence score	/	0.184	0.217	0.253	0.298	0.364
No. of bat passes	1,352	1,229	1,185	1,129	1,034	909
Occurrences	0.615	0.599	0.596	0.596	0.584	0.544
Estimated false positive rate	0.128	0.079	0.065	0.051	0.034	0.019
Estimated false negative rate	0	0.034	0.053	0.080	0.131	0.211
<i>Rhinolophus ferrumequinum</i>						
Confidence score	/	0.000	0.000	0.000	0.000	0.000
No. of bat passes	41	41	41	41	41	41
Occurrences	0.046	0.046	0.046	0.046	0.046	0.046
Estimated false positive rate	0.000	0.000	0.000	0.000	0.000	0.000
Estimated false negative rate	0.000	0.000	0.000	0.000	0.000	0.000
<i>Rhinolophus hipposideros</i>						
Confidence score	/	0.385	0.398	0.411	0.427	0.452
No. of bat passes	128	117	116	116	116	113
Occurrences	0.113	0.107	0.104	0.104	0.104	0.104
Estimated false positive rate	0.078	0.011	0.007	0.007	0.007	0.003
Estimated false negative rate	0	0.018	0.022	0.022	0.022	0.199

a variance inflation factor value under 1.5, meaning there was no strong evidence of multicollinearity (Chatterjee & Hadi, 2006).

We then compared the estimates of each environmental variable among fitted models to check the consistency in the response of bats to environmental variables in relation to the different FPTs.

### 3 | RESULTS

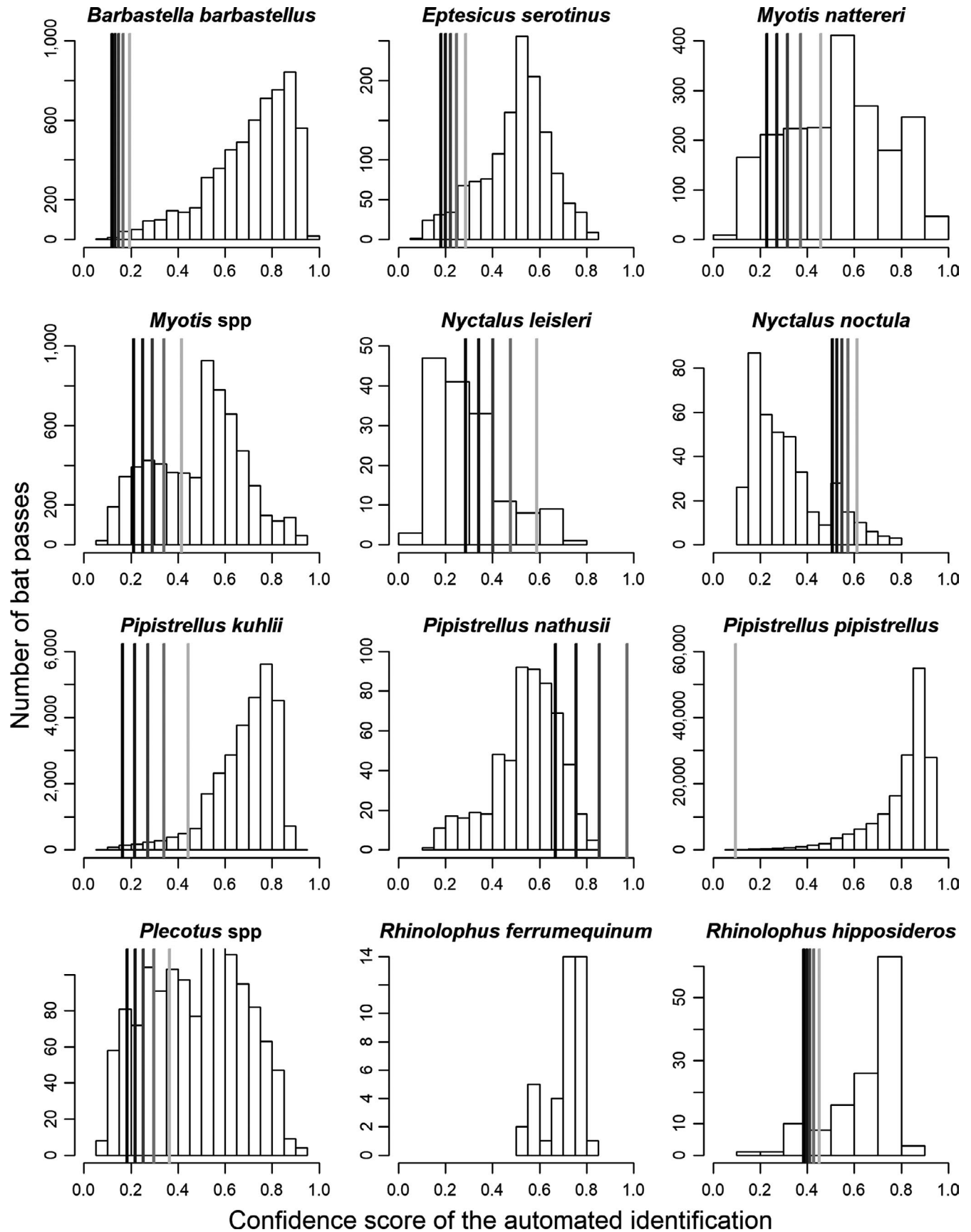
#### 3.1 | Automated identification and manual checking

Over the 23 nights sampled, among the 212,347 bat passes recorded, 167,504 (79%) were assigned to *Pipistrellus pipistrellus*, 28,589 (13%) to *Pipistrellus kuhlii*, 6,430 (3%) to *Myotis* spp. and 5,835 (3%) to *Barbastella barbastellus* (Table 1). A stratified random sample of 1,910 bat passes were manually checked (Table 1). False

positive rates varied a lot among species, from 0.0% for *Rhinolophus ferrumequinum* to 69.4% for *Nyctalus noctula* (Table 1). The largest number of errors detected in manual checks was for *N. noctula* confused with social calls of *P. pipistrellus* (only one location involved) and non-bat noises, and calls of *Pipistrellus nathusii* were confused with *P. kuhlii*, *P. pipistrellus* and non-bat noises (Table S1). Concerning the random checking of 500 sound files identified as non-bat by the software, we found that three (0.6%) contained bat events.

#### 3.2 | Checking for environmental biases in identification errors

Using the dataset on which manual checks were carried out, we investigated a potential variation in automated identification errors due the environmental variables. The probability of these being false



**FIGURE 3** Number of bat passes in the total dataset according to confidence scores provided by the automated identification. Vertical lines show the threshold below which data were removed to ensure a given false positive tolerance (from black to grey: 0.5, 0.4, 0.3, 0.2 and 0.1)

positives was significantly affected by only one environmental variable (habitat type of survey sites: hedgerow vs. open area) and for only one species, *N. noctula* ( $p < 0.001$ ; Table S2). All other environmental variables were not found to affect the probability of there being false negatives for any species (Table S3).

### 3.3 | False positive rate modelling

Success and failure in automated identification assessed through manual checking were modelled in relation to the confidence score provided by the software, allowing us to predict the required confidence score to ensure a given FPT (Figure 2). Confidence scores required to ensure FPTs (i.e. 0.5, 0.4, 0.3, 0.2 and 0.1) did not vary much for species such as *B. barbastellus* (0.12–0.20), *Eptesicus serotinus* (0.18–0.29) and *Rhinolophus hipposideros* (0.39–0.45), but more for others, for example *Nyctalus leisleri* (0.29–0.59), *P. kuhlii* (0.16–0.44) and *Plecotus* spp. (0.18–0.36) (Table 2). In addition, these FPTs confidence scores were lower for *B. barbastellus*, *E. serotinus*, *P. kuhlii*, *Plecotus* spp., *Myotis* spp., and higher for *P. nathusii* and *N. noctula* (Table 2).

For *P. pipistrellus*, errors were rare thus the lowest possible confidence score (0.096) corresponded to a FPT lower than 0.2. In contrast, for *P. nathusii*, the highest possible confidence score (0.971) corresponded to a FPT greater than 0.1, that is more than one in ten chance of failure (Table 2). Moreover, no errors were found in the sample for *R. ferrumequinum*, which prevented the modelling of error rate for this species (Table 2).

Low FPTs (i.e. removing data below a high confidence score) often led to an important decrease in activity measures (Table 2). For example, *Myotis* spp. and *N. leisleri* activity decreased by 27.8% and 82.1%, respectively, between 0.5 FPT and 0.1 FPT (Table 2). However, such high decreases in activity resulted in a little decrease in occurrence for these species: 6.7% for the *Myotis* spp. group and 10.7% for *N. leisleri* (Table 2). For other species, the activity and occurrence were more stable across FPTs, including for *B. barbastellus*, *E. serotinus*, *P. kuhlii*, *Plecotus* spp., and *R. hipposideros* (Table 2).

At the highest FPT (0.5), the estimated false positive rate was high (>21%) for three species (*N. leisleri*, *N. noctula* and *P. nathusii*), and very low (<5%) for six species (*B. barbastellus*, *E. serotinus*, *P. kuhlii*, *P. pipistrellus*, *R. ferrumequinum* and *R. hipposideros*) (Table 2). However, at the lowest FPT (0.1), all species showed an estimated false positive rate under 0.05, except for *N. leisleri* (0.08) and *P. nathusii* for which no data satisfied a FPT lower than 0.1 (Table 2).

Estimating the generated false negative rate (i.e. true positives discarded by reducing the FPT) was very low (<4%) at 0.5 FPT for most species except *N. leisleri* (0.19) and *P. nathusii* (0.28) (Table 2). This rate became more important at 0.1 FPT, with null values for *P. pipistrellus* and *R. ferrumequinum*; with very low values (<10%) for five species (*B. barbastellus*, *E. serotinus*, *N. noctula*, *P. kuhlii* and *R. hipposideros*); and with high values for *N. leisleri* (0.425) and *P. nathusii* (0.377) (Table 2). The average AUC from ROC curves was 0.93 (range: 0.73–1.00; Figure S3).

### 3.4 | Consistency of activity patterns across error rate tolerance gradient

To study the influence of confidence score thresholding according to FPTs below which data were discarded (i.e. changes in amount of data, species occurrence, estimated false positive rate and estimated rate of generated false negative), modelling of the bat response (i.e. the number of bat passes according to selected FPT) to environmental variables was performed at all FPTs.

When comparing model outputs from naive (i.e. raw data) to robust analyses (i.e. FPTs), a loss or a gain of significance was occurred for the open areas versus hedgerows variable for *N. leisleri*, the distance to forest for *Myotis* spp. and *N. leisleri*, the length of hedgerows for *N. leisleri* and the distance to urban areas for *N. noctula* (Table 3). In addition, for significant variables, an inversion of the direction of the estimate for the open areas versus hedgerows variable occurred for *N. noctula* and *P. nathusii* (Table 3). In all other cases, no changes were found (Table 3).

However, we did not detect any major changes in model outputs between the 0.5, 0.4, 0.3, 0.2 and 0.1 FPTs for which response estimates and standard errors remained highly stable (Table 3). In only two cases, we detected a loss of significance: for *N. noctula* with FPTs lower than 0.2 and 0.3 for the distance to forests and the length of hedgerows variables respectively (Table 3). However, for this species, the open areas versus hedgerows variable remained significant and highly stable at all FPTs (Table 3).

All species had at least one significant habitat variable response irrespective of the used FPTs, except *N. leisleri*. Hedgerows had a significantly higher bat activity (i.e. number of bat passes) associated with them than open areas for seven species or groups (*B. barbastellus*, *M. nattereri*, *Myotis* spp., *P. kuhlii*, *P. pipistrellus*, *Plecotus* spp. and *R. hipposideros*) and a significantly lower bat activity for two species (*N. noctula* and *P. nathusii*) (Table 3). We also found a significant negative relationship between bat activity and (a) the distance to urban areas variable for two species or groups (*E. serotinus* and *Plecotus* spp.; Table 3); (b) with the distance to forest variable for two species (*N. noctula* and *R. ferrumequinum*; Table 3); (c) with the distance to wetlands variable for *R. ferrumequinum*; and (d) with the length of hedgerows variable for *N. noctula*, *P. pipistrellus* and *R. ferrumequinum* (Table 3) but a significant positive relationship with the distance to wetlands variable for *P. kuhlii* (Table 3).

## 4 | DISCUSSION

This study demonstrates that automated acoustic identification of bats, as well as by extension all other taxa acoustically identifiable by software, coupled with partial manual checking and false positive rate modelling (i.e. semi-automated identification; Newson et al., 2015), is a key tool for improving reliability of studies based on acoustic data. Indeed, robust ecological responses could be produced even in cases where false positive rates were so far considered too high (Rydell et al., 2017). This new and robust framework takes advantage of confidence scores provided by the automated

**TABLE 3** Species response to environmental variables (estimates, standard errors and *p*-values) according to the false positive tolerances (\*\**p* < 0.001, \*\*\**p* < 0.01, \**p* < 0.05, *p* < 0.1)

Species	Environmental variables	False positive tolerance					
		Raw data	0.5	0.4	0.3	0.2	0.1
<i>Barbastella barbastellus</i>	Open areas versus hedgerows	-2.81 ± 0.24***	-2.81 ± 0.24***	-2.81 ± 0.24***	-2.81 ± 0.24***	-2.81 ± 0.24***	-2.81 ± 0.24***
	Dist. to forest	0.08 ± 0.12	0.08 ± 0.13	0.08 ± 0.13	0.08 ± 0.13	0.08 ± 0.13	0.08 ± 0.13
	Dist. to wetland	-0.03 ± 0.12	-0.03 ± 0.12	-0.03 ± 0.12	-0.03 ± 0.12	-0.03 ± 0.12	-0.04 ± 0.12
	Dist. to urban	0.01 ± 0.1	0.01 ± 0.1	0.01 ± 0.1	0.01 ± 0.1	0.01 ± 0.1	0.02 ± 0.1
	Length of hedgerows	0.17 ± 0.12	0.17 ± 0.12	0.17 ± 0.12	0.17 ± 0.12	0.17 ± 0.12	0.17 ± 0.12
<i>Eptesicus serotinus</i>	Open areas versus hedgerows	-0.57 ± 0.38	-0.43 ± 0.4	-0.44 ± 0.4	-0.45 ± 0.41	-0.43 ± 0.42	-0.35 ± 0.42
	Dist. to forest	-0.07 ± 0.23	-0.15 ± 0.24	-0.15 ± 0.25	-0.16 ± 0.25	-0.15 ± 0.25	-0.13 ± 0.26
	Dist. to wetland	0.08 ± 0.19	0.12 ± 0.2	0.12 ± 0.2	0.12 ± 0.2	0.12 ± 0.21	0.08 ± 0.21
	Dist. to urban	-0.7 ± 0.19***	-0.8 ± 0.21***	-0.79 ± 0.21***	-0.78 ± 0.21***	-0.77 ± 0.21***	-0.77 ± 0.22***
	Length of hedgerows	0.2 ± 0.23	0.2 ± 0.24	0.21 ± 0.24	0.21 ± 0.24	0.19 ± 0.25	0.16 ± 0.25
<i>Myotis nattereri</i>	Open areas versus hedgerows	-1.16 ± 0.21***	-1.14 ± 0.22***	-1.12 ± 0.23***	-1.05 ± 0.23***	-1.01 ± 0.24***	-1.03 ± 0.27***
	Dist. to forest	0.16 ± 0.13	0.13 ± 0.13	0.14 ± 0.13	0.15 ± 0.14	0.1 ± 0.14	0.11 ± 0.15
	Dist. to wetland	0.17 ± 0.11	0.21 ± 0.12	0.23 ± 0.12	0.24 ± 0.12	0.22 ± 0.13	0.21 ± 0.13
	Dist. to urban	0.07 ± 0.1	0.08 ± 0.11	0.09 ± 0.11	0.11 ± 0.12	0.11 ± 0.12	0.13 ± 0.13
	Length of hedgerows	0.18 ± 0.12	0.22 ± 0.13	0.24 ± 0.13	0.27 ± 0.14	0.32 ± 0.14*	0.3 ± 0.16
<i>Myotis</i> spp	Open areas versus hedgerows	-1.66 ± 0.19***	-1.64 ± 0.19***	-1.6 ± 0.19***	-1.55 ± 0.19***	-1.54 ± 0.19***	-1.61 ± 0.26***
	Dist. to forest	0.24 ± 0.12*	0.22 ± 0.12	0.22 ± 0.12	0.22 ± 0.12	0.22 ± 0.13	0.20 ± 0.13
	Dist. to wetland	0.1 ± 0.1	0.11 ± 0.11	0.1 ± 0.11	0.11 ± 0.11	0.1 ± 0.11	0.10 ± 0.11
	Dist. to urban	-0.07 ± 0.09	-0.08 ± 0.09	-0.08 ± 0.1	-0.06 ± 0.1	-0.05 ± 0.1	-0.03 ± 0.1
	Length of hedgerows	0.13 ± 0.12	0.15 ± 0.12	0.15 ± 0.12	0.17 ± 0.12	0.18 ± 0.12	0.21 ± 0.13
<i>Nyctalus leisleri</i>	Open areas versus hedgerows	-0.8 ± 0.22**	-0.26 ± 0.29	-0.23 ± 0.35	0.43 ± 0.4	0.69 ± 0.45	1.1 ± 0.64
	Dist. to forest	0.34 ± 0.13**	0.16 ± 0.17	0.21 ± 0.21	0.08 ± 0.26	0.14 ± 0.28	0.49 ± 0.35
	Dist. to wetland	0.07 ± 0.1	-0.09 ± 0.15	-0.02 ± 0.19	-0.12 ± 0.26	-0.21 ± 0.3	-0.17 ± 0.42
	Dist. to urban	-0.1 ± 0.1	-0.19 ± 0.15	-0.01 ± 0.18	0.08 ± 0.23	0.23 ± 0.26	0.43 ± 0.35
	Length of hedgerows	0.35 ± 0.12**	0.23 ± 0.16	0.23 ± 0.21	0.27 ± 0.25	0.28 ± 0.29	0.22 ± 0.41

(Continues)

TABLE 3 (Continued)

Species	Environmental variables	False positive tolerance						
		Raw data	0.5	0.4	0.3	0.2	0.1	
<i>Nyctalus noctula</i>	Open areas versus hedgerows	-1.19 ± 0.17***	1.46 ± 0.31***	1.7 ± 0.36***	1.83 ± 0.4***	1.37 ± 0.44**	1.28 ± 0.49*	
	Dist. to forest	-0.55 ± 0.11***	-0.68 ± 0.23**	-0.66 ± 0.26*	-0.7 ± 0.29*	-0.26 ± 0.32	-0.12 ± 0.35	
	Dist. to wetland	-0.07 ± 0.06	0.02 ± 0.18	0.16 ± 0.21	0.25 ± 0.24	0.3 ± 0.27	0.34 ± 0.34	
	Dist. to urban	0.25 ± 0.07***	-0.07 ± 0.18	-0.1 ± 0.21	-0.12 ± 0.23	-0.01 ± 0.25	-0.04 ± 0.29	
	Length of hedgerows	0.34 ± 0.08***	0.43 ± 0.21*	0.49 ± 0.25*	0.52 ± 0.28.	0.16 ± 0.31	-0.03 ± 0.36	
<i>Pipistrellus kuhlii</i>	Open areas versus Hedgerows	-1.98 ± 0.26***	-1.98 ± 0.26***	-1.98 ± 0.27***	-1.98 ± 0.27***	-1.98 ± 0.27***	-1.98 ± 0.27***	
	Dist. to forest	0.09 ± 0.13	0.09 ± 0.13	0.09 ± 0.13	0.09 ± 0.14	0.09 ± 0.14	0.1 ± 0.14	
	Dist. to wetland	0.25 ± 0.13*	0.25 ± 0.13*	0.26 ± 0.13*	0.25 ± 0.13*	0.26 ± 0.13*	0.26 ± 0.13*	
	Dist. to urban	0.07 ± 0.13	0.07 ± 0.13	0.07 ± 0.13	0.08 ± 0.13	0.08 ± 0.13	0.08 ± 0.13	
	Length of hedgerows	0.07 ± 0.15	0.06 ± 0.15	0.06 ± 0.15	0.06 ± 0.15	0.06 ± 0.15	0.06 ± 0.15	
<i>Pipistrellus nathusii</i>	Open areas versus Hedgerows	-0.37 ± 0.24	1.02 ± 0.38**	2.57 ± 0.84**	/	/	/	
	Dist. to forest	0.1 ± 0.16	0.28 ± 0.23	0.81 ± 0.46.	/	/	/	
	Dist. to wetland	0.06 ± 0.13	0.02 ± 0.2	0.53 ± 0.42	/	/	/	
	Dist. to urban	-0.05 ± 0.13	0.09 ± 0.21	0 ± 0.44	/	/	/	
	Length of hedgerows	0.11 ± 0.16	0.42 ± 0.24.	0.88 ± 0.54	/	/	/	
<i>Pipistrellus pipistrellus</i>	Open areas versus Hedgerows	-2.87 ± 0.19***	-2.87 ± 0.19***	-2.87 ± 0.19***	-2.87 ± 0.19***	-2.87 ± 0.19***	-2.87 ± 0.19***	
	Dist. to forest	0.13 ± 0.13	0.13 ± 0.13	0.13 ± 0.13	0.13 ± 0.13	0.13 ± 0.13	0.13 ± 0.13	
	Dist. to wetland	0.04 ± 0.11	0.04 ± 0.11	0.04 ± 0.11	0.04 ± 0.11	0.04 ± 0.11	0.04 ± 0.11	
	Dist. to urban	-0.13 ± 0.1	-0.13 ± 0.1	-0.13 ± 0.1	-0.13 ± 0.1	-0.13 ± 0.1	-0.13 ± 0.1	
	Length of hedgerows	0.35 ± 0.12**	0.35 ± 0.12**	0.35 ± 0.12**	0.35 ± 0.12**	0.35 ± 0.12**	0.35 ± 0.12**	
<i>Plecotus</i> spp.	Open areas versus Hedgerows	-0.91 ± 0.19***	-0.85 ± 0.19***	-0.87 ± 0.19***	-0.87 ± 0.19***	-0.85 ± 0.19***	-0.79 ± 0.2***	
	Dist. to forest	0.08 ± 0.12	0.1 ± 0.12	0.11 ± 0.12	0.1 ± 0.12	0.09 ± 0.12	0.08 ± 0.13	
	Dist. to wetland	-0.16 ± 0.11	-0.14 ± 0.11	-0.15 ± 0.11	-0.15 ± 0.11	-0.14 ± 0.11	-0.17 ± 0.12	
	Dist. to urban	-0.25 ± 0.1**	-0.25 ± 0.1*	-0.26 ± 0.1**	-0.25 ± 0.1**	-0.25 ± 0.1*	-0.23 ± 0.1*	
	Length of hedgerows	0.1 ± 0.12	0.09 ± 0.12	0.09 ± 0.12	0.08 ± 0.12	0.11 ± 0.12	0.11 ± 0.13	

(Continues)

TABLE 3 (Continued)

Species	Environmental variables	False positive tolerance					
		Raw data	0.5	0.4	0.3	0.2	0.1
<i>Rhinolophus ferrumequinum</i>	Open areas versus Hedgerows	0.26 ± 0.39	0.26 ± 0.39	0.26 ± 0.39	0.26 ± 0.39	0.26 ± 0.39	0.26 ± 0.39
	Dist. to forest	0.74 ± 0.25**	0.74 ± 0.25**	0.74 ± 0.25**	0.74 ± 0.25**	0.74 ± 0.25**	0.74 ± 0.25**
	Dist. to wetland	-1.20 ± 0.29***	-1.20 ± 0.29***	-1.20 ± 0.29***	-1.20 ± 0.29***	-1.20 ± 0.29***	-1.20 ± 0.29***
	Dist. to urban	-0.21 ± 0.26	-0.21 ± 0.26	-0.21 ± 0.26	-0.21 ± 0.26	-0.21 ± 0.26	-0.21 ± 0.26
	Length of hedgerows	0.83 ± 0.29**	0.83 ± 0.29**	0.83 ± 0.29**	0.83 ± 0.29**	0.83 ± 0.29**	0.83 ± 0.29**
<i>Rhinolophus hipposideros</i>	Open areas versus Hedgerows	-3.08 ± 0.74***	-2.92 ± 0.74***	-2.92 ± 0.74***	-2.92 ± 0.74***	-2.92 ± 0.74***	-2.89 ± 0.73***
	Dist. to forest	0.09 ± 0.3	-0.47 ± 0.36	-0.5 ± 0.37	-0.5 ± 0.37	-0.5 ± 0.37	-0.51 ± 0.36
	Dist. to wetland	-0.33 ± 0.26	-0.45 ± 0.26	-0.49 ± 0.27	-0.49 ± 0.27	-0.49 ± 0.27	-0.46 ± 0.28
	Dist. to urban	-0.18 ± 0.26	-0.17 ± 0.26	-0.14 ± 0.27	-0.14 ± 0.27	-0.14 ± 0.27	-0.15 ± 0.27
	Length of hedgerows	0.03 ± 0.3	0.06 ± 0.3	0.07 ± 0.3	0.07 ± 0.3	0.07 ± 0.3	0.08 ± 0.3

identification software and its ability for distinguishing true positives and false positives (Figure S3), controlling for false positive tolerances (FPTs), and checking for potential biases induced by identification errors.

#### 4.1 | Using confidence thresholding

Minimum confidence scores required to ensure a given FPT according to species exhibited low to moderate variation across the 0.5 to 0.1 FPTs (Table 2). To investigate the effect of the automated identification errors on bat activity patterns in relation with FPTs, we studied the response of bat activity to several environmental variables known to impact bats. Depending on species, the most significant responses to environmental variables were consistent with known patterns of bat activity: a negative effect of open areas versus hedgerows and of decreasing length of hedgerows (Lacoeuilhe et al., 2016; Verboom & Huitema, 1997), of distance to forest (Boughey et al., 2011; Frey-Ehrenbold et al., 2013), to urban areas (Jung & Threlfall, 2016; Mckinney, 2005) and to distance to wetlands (Santos et al., 2013; Sirami et al., 2013).

A comparison of the relationship between environmental variables and bat activity between using the raw data (i.e. using the whole dataset regardless of the confidence score) and FPTs selected data (i.e. removing data above defined FPT to minimize the false positive rate) showed some discrepancies. We sometimes found opposite significant responses, for example the effect of open areas versus hedgerows on *N. noctula* and *P. nathusii*, when comparing results from raw data and FPTs (Table 3). This demonstrates that analyses conducted on raw automated identification data could be severely biased. In this respect, removing data above a 0.5 FPT (i.e. removing data with a low success probability) is essential, in accordance with concerns expressed by Russo and Voigt (2016).

Logically these biases due to false positives mostly seem to impact uncommon species which are acoustically similar to commoner ones. Here the most impacted species is *P. nathusii* which suffers from a high false positive rate due to the local abundance of *P. kuhlii* and *P. pipistrellus* (Tables 1 and 2). Consequently, an analysis conducted on raw automatically identified data for this species seems to be driven by the response of the two other Pipistrelles.

#### 4.2 | Assessing robustness of ecological inferences

We assessed the robustness of ecological inferences by studying the consistency of bat responses to environmental variables among FPTs. However, for *P. nathusii* it was not possible to ensure such a robustness due a lack of data from 0.4 FPT (Table 2). This framework thus showed that this was not possible to produce robust ecological inferences on this species due to a high false positive rate in this dataset. In addition, for *N. noctula*, we lost significance of the response to the distance to the forest and the length of hedgerows from 0.2 and 0.3 FPTs respectively (Table 3). Such loss of significance could be linked to a high loss of bat passes and occurrences by reducing the FPTs, or linked to environmental biases affecting spatial

distribution of false positive or generated false negative rates. Thus, given the uncertainty about the mechanism involved, it was also not possible to produce robust inferences for this species given high losses of bat passes and occurrence, and high estimated false positive rates by reducing the FPTs (Table 2).

At the other end of the spectrum, the estimated false positive rate was always extremely low or even zero whatever the confidence score in the automated identification for *P. pipistrellus* and *R. ferrumequinum* (Table 1), thus not raising any problem of error risk.

For all nine other species or species groups, 15 of the 18 significant responses to environmental variables were robust with a high stability of model outputs while reducing the FPT from 0.5 to 0.1 (Table 3). In addition, despite a decrease in bat activity measures due to thresholding at FPTs, the occurrence of species remained highly stable whilst retaining statistical power among FPTs. Our study thus demonstrates that using our approach many ecological inferences could be robust against identification errors.

### 4.3 | Survey recommendations and limitations

This study proposes a cautious method to account for identification errors in acoustic surveys aimed at studying the response of bats in relation to environmental variables, such as anthropogenic pressures, without the need for exhaustive checking of recordings.

The FPT of 0.5 is a threshold for which false negatives and false positives are expected to be approximately balanced. However, false positives are more likely to produce biases because their rate is strongly driven by the activity pattern of other species. In contrast, the FPT of 0.1 minimizes the false positive rate, but at the cost of losing potentially a lot of data, so a high generated false negative rate by discarding true positives (Table 2). Rather than looking for a possible optimal threshold, we recommend that researchers systematically check the consistency of responses for at least two significantly different thresholds (e.g. 0.5 and 0.1 FPTs), in order to assess the robustness of the results and only going on to conclusive interpretation when these are consistent.

A lack of consistency is most likely to occur for rare species with very low abundance/occurrence, and for uncommon species which are acoustically similar to commoner ones such as *P. nathusii* here which is acoustically similar to *P. kuhlii* (Obrist et al., 2004). The efficiency of the automated identification of *P. nathusii* and *N. lesleiri* was lowest (AUC of 0.73 for both; Figure S3) due to particular context of the study where these species were much rarer than their acoustically closest relative (*P. kuhlii* and *E. serotinus*, respectively; Table 1). For these species, either systematic manual checking or an important improvement in automated identification efficiency is needed to conduct robust analyses. However, our framework of error rate modelling is already sufficient to effectively identify these problematic species and should prevent users of automated identification to draw conclusions that are not robust. In addition, another prerequisite for drawing robust conclusions from this framework is to ensure that error types (i.e. false

negatives and false positives) are not correlated with the variables tested in the study. In our study case, we only detected one significant dependence for the open area versus hedgerows for the false positives of *N. noctula* (Table S2). For this species, automated identification was more efficient (i.e. lower number of false positives) for survey sites located in open areas than close to hedgerows where calls are more difficult to identify due to frequency modulation (Barataud, 2015; Obrist et al., 2004). It is not surprising that the false positive rate of a rare species like *N. noctula* could be influenced by local habitat type because this variable is expected to have different effects on other species, and thus influence false positive rate through the relative density between *N. noctula* and other bat species. Thus, we expect a bias in the measure of activity towards open areas in this case. Hence, the significant positive response of this species to open areas compared to hedgerows should be considered unreliable to make any ecological inference (Table 3).

This method can be applied to any ecological studies with standardized sampling but, of course, cannot help for surveys where no error can be tolerated, for example for producing species inventories for protected species, as required for environmental impact assessments (Russo & Voigt, 2016). However, in this case, automated identification can still indicate what bat passes should be manually checked in order to identify species presence at the site scale, by selecting passes with the highest confidence scores, and thus saving time for the user.

Finally, the proposed method can be applied to any acoustic taxa for which automated identification software is developed and where confidence scores are provided. A crucial advantage of this method is that manual checking of a relatively small subset of the dataset (<1% in this study) is sufficient to assess error rates associated with species identification. This is especially true given that checking all data is very time-consuming and virtually impossible for such a large dataset.

### ACKNOWLEDGEMENTS

This work was supported by DIM ASTREA grants from Region Ile-de-France. We sincerely thank Agrosolutions for funding field study fees. We thank IN2P3 Computing Centre for providing facilities to process and archive in the long-term all the recordings of this study, and Didier Bas for help in this process. We also thank the two anonymous reviewers for their constructive comments improving this work.

### AUTHORS' CONTRIBUTIONS

K.B., C.K. and Y.B. conceived the ideas; K.B. and Y.B. designed the methodology; K.B. collected the data; K.B. and Y.B. manually checked bat passes; K.B. and J.P. analysed the data and wrote the R scripts; all the authors led the writing of the manuscript. All authors critically contributed to the drafts and gave their final approval for publication.

## DATA ACCESSIBILITY

All R codes and data used in the study are available from the following github website link: <https://github.com/KevBarre/Semi-automated-method-to-account-for-identification-errors-in-biological-acoustic-surveys> <https://doi.org/10.5281/zenodo.2646482>.

## ORCID

Kévin Barré  <https://orcid.org/0000-0001-5368-4053>

Fabien Claireau  <https://orcid.org/0000-0003-3848-7427>

Christian Kerbiriou  <https://orcid.org/0000-0001-6080-4762>

## REFERENCES

- Adams, A. M., Jantzen, M. K., Hamilton, R. M., & Fenton, M. B. (2012). Do you hear what I hear? Implications of detector selection for acoustic monitoring of bats. *Methods in Ecology and Evolution*, 3(6), 992–998. <https://doi.org/10.1111/j.2041-210X.2012.00244.x>
- Azam, C., Le Viol, I., Bas, Y., Zissis, G., Vernet, A., Julien, J.-F., & Kerbiriou, C. (2018). Evidence for distance and illuminance thresholds in the effects of artificial lighting on bat activity. *Landscape and Urban Planning*, 175, 123–135. <https://doi.org/10.1016/j.landurbplan.2018.02.011>
- Azam, C., Le Viol, I., Julien, J. F., Bas, Y., & Kerbiriou, C. (2016). Disentangling the relative effect of light pollution, impervious surfaces and intensive agriculture on bat activity with a national-scale monitoring program. *Landscape Ecology*, 31(10), 2471–2483. <https://doi.org/10.1007/s10980-016-0417-3>
- Barataud, M. (2015). *Acoustic ecology of european bats: Species identification, study of their habitats and foraging behaviour*. (p. 352). Paris: Muséum national d'Histoire naturelle & Mèze: Biotope. (Inventaires & biodiversité; 8).
- Barré, K., Le Viol, I., Bas, Y., Julliard, R., & Kerbiriou, C. (2018). Estimating habitat loss due to wind turbine avoidance by bats: Implications for European siting guidance. *Biological Conservation*, 226(April), 205–214. <https://doi.org/10.1016/j.biocon.2018.07.011>
- Barré, K., Le Viol, I., Julliard, R., Chiron, F., & Kerbiriou, C. (2017). Tillage and herbicide reduction mitigate the gap between conventional and organic farming effects on foraging activity of insectivorous bats. *Ecology and Evolution*, 1–11. <https://doi.org/10.1002/ece3.3688>
- Bas, Y., Bas, D., & Julien, J. (2017). Tadarida: A toolbox for animal detection on acoustic recordings. *Journal of Open Research Software*, 5(6), 1–8. <https://doi.org/10.5334/jors.154>
- Biotope. (2013). Manual of SonoChiro 3.0. Retrieved from [http://www.leclub-biotope.com/index.php?controller=attachment&id\\_attachment=13](http://www.leclub-biotope.com/index.php?controller=attachment&id_attachment=13)
- Boughhey, K. L., Lake, I. R., Haysom, K. A., & Dolman, P. M. (2011). Effects of landscape-scale broadleaved woodland configuration and extent on roost location for six bat species across the UK. *Biological Conservation*, 144(9), 2300–2310. <https://doi.org/10.1016/j.biocon.2011.06.008>
- Britzke, E. R., Duchamp, J. E., Murray, K. L., Swihart, R. K., & Robbins, L. W. (2011). Acoustic identification of bats in the eastern United States: A comparison of parametric and nonparametric methods. *Journal of Wildlife Management*, 75(3), 660–667. <https://doi.org/10.1002/jwmg.68>
- Butchart, S. H. M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., ... Watson, R. (2010). Global biodiversity: Indicators of recent declines. *Science*, 328(5982), 1164–1168. <https://doi.org/10.1126/science.1187512>
- Chatterjee, S., & Hadi, A. S. (2006). *Regression analysis by example*, 5th ed.. New Jersey: John Wiley & Sons Inc.. <https://doi.org/10.1002/0470055464>
- Claireau, F., Bas, Y., Puechmaille, S. J., Julien, J. F., Allegrini, B., & Kerbiriou, C. (2019). Bat overpasses: An insufficient solution to restore habitat connectivity across roads. *Journal of Applied Ecology*, 56, 573–584. <https://doi.org/10.1111/1365-2664.13288>
- Delport, W., Kemp, A. C., & Ferguson, J. W. H. (2002). Vocal identification of individual African Wood Owls *Strix woodfordii*: A technique to monitor long-term adult turnover and residency. *Ibis*, 144(1), 30–39. <https://doi.org/10.1046/j.0019-1019.2001.00019.x>
- Denzinger, A., & Schnitzler, H. U. (2013). Bat guilds, a concept to classify the highly diverse foraging and echolocation behaviors of microchiropteran bats. *Frontiers in Physiology*, 4, 1–16. <https://doi.org/10.3389/fphys.2013.00164>
- Fisher, J. A. D., Frank, K. T., & Leggett, W. C. (2010). Dynamic macroecology on ecological time-scales. *Global Ecology and Biogeography*, 19(1), 1–15. <https://doi.org/10.1111/j.1466-8238.2009.00482.x>
- Frey-Ehrenbold, A., Bontadina, F., Arlettaz, R., & Obrist, M. K. (2013). Landscape connectivity, habitat structure and activity of bat guilds in farmland-dominated matrices. *Journal of Applied Ecology*, 50(1), 252–261. <https://doi.org/10.1111/1365-2664.12034>
- Fritsch, G., & Bruckner, A. (2014). Operator bias in software-aided bat call identification. *Ecology and Evolution*, 4(13), 2703–2713. <https://doi.org/10.1002/ece3.1122>
- Froidevaux, J. S. P., Zellweger, F., Bollmann, K., & Obrist, M. K. (2014). Optimizing passive acoustic sampling of bats in forests. *Ecology and Evolution*, 4(24), 4690–4700. <https://doi.org/10.1002/ece3.1296>
- Frommolt, K. H. (2017). Information obtained from long-term acoustic recordings: Applying bioacoustic techniques for monitoring wetland birds during breeding season. *Journal of Ornithology*, 158(3), 659–668. <https://doi.org/10.1007/s10336-016-1426-3>
- Gasc, A., Sueur, J., Pavoine, S., Pellens, R., & Grandcolas, P. (2013). Biodiversity sampling using a global acoustic approach: Contrasting sites with microendemism in New Caledonia. *PLoS ONE*, 8(5), e65311. <https://doi.org/10.1371/journal.pone.0065311>
- Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2018). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10, 169–185. <https://doi.org/10.1111/2041-210X.13101>
- Jeliazkov, A., Bas, Y., Kerbiriou, C., Julien, J. F., Penone, C., & Le Viol, I. (2016). Large-scale semi-automated acoustic monitoring allows to detect temporal decline of bush-crickets. *Global Ecology and Conservation*, 6, 208–218. <https://doi.org/10.1016/j.gecco.2016.02.008>
- Jung, K., & Threlfall, C. (2016). Urbanization and its effect on bats: A global meta-analysis. In T. Kingston, & C. Voigt (Eds.), *Bats in the anthropocene: Conservation of bats in a changing world* (pp. 13–33). Basel, Switzerland: Springer.
- Kalan, A. K., Mundry, R., Wagner, O. J. J., Heinicke, S., Boesch, C., & Kühl, H. S. (2015). Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators*, 54, 217–226. <https://doi.org/10.1016/j.ecolind.2015.02.023>
- Kerbiriou, C., Azam, C., Touroult, J., Marmet, J., Julien, J. F., & Pellissier, V. (2018). Common bats are more abundant within Natura 2000 areas. *Biological Conservation*, 217, 66–74. <https://doi.org/10.1016/j.biocon.2017.10.029>
- Kerbiriou, C., Bas, Y., Le Viol, I., Lorrilliere, R., Mougnot, J., & Julien, J. F. (2018). Potential of bat pass duration measures for studies of bat activity. *Bioacoustics*, 4622, 1–16. <https://doi.org/10.1080/09524622.2017.1423517>
- Lacoeuilhe, A., Machon, N., Julien, J. F., & Kerbiriou, C. (2016). Effects of hedgerows on bats and bush crickets at different spatial scales. *Acta Oecologica*, 71, 61–72. <https://doi.org/10.1016/j.actao.2016.01.009>
- López-Baucells, A., Torrent, L., Rocha, R., Bobrowiec, P. E. D., Palmeirim, J. M., & Meyer, C. F. J. (2019). Stronger together: Combining automated classifiers with manual post-validation optimizes the workload

- vs reliability trade-off of species identification in bat acoustic surveys. *Ecological Informatics*, 49, 45–53. <https://doi.org/10.1016/j.ecoinf.2018.11.004>
- Mckinney, M. L. (2005). Urbanization as a major cause of biotic homogenization. *Biological Conservation*, 127(3), 247–260. <https://doi.org/10.1016/j.biocon.2005.09.005>
- Millon, L., Julien, J.-F., Julliard, R., & Kerbiriou, C. (2015). Bat activity in intensively farmed landscapes with wind turbines and offset measures. *Ecological Engineering*, 75, 250–257. <https://doi.org/10.1016/j.ecoleng.2014.11.050>
- Newson, S. E., Evans, H. E., & Gillings, S. (2015). A novel citizen science approach for large-scale standardised monitoring of bat activity and distribution, evaluated in eastern England. *Biological Conservation*, 191, 38–49. <https://doi.org/10.1016/j.biocon.2015.06.009>
- Nowacek, D. P., Christiansen, F., Bejder, L., Goldbogen, J. A., & Friedlaender, A. S. (2016). Studying cetacean behaviour: New technological approaches and conservation applications. *Animal Behaviour*, 120, 235–244. <https://doi.org/10.1016/j.anbehav.2016.07.019>
- Obrist, M. K., Boesch, R., & Fluckiger, P. F. (2004). Variability in echolocation call design of 26 Swiss bat species: Consequences, limits and options for automated field identification with a synergetic pattern recognition approach. *Mammalia*, 68(4), 307–322. <https://doi.org/10.1515/mamm.2004.030>
- Ovaskainen, O., Moliterno de Camargo, U., & Somervuo, P. (2018). Animal Sound Identifier (ASI): Software for automated identification of vocal animals. *Ecology Letters*, 21(8), 1244–1254. <https://doi.org/10.1111/ele.13092>
- Parsons, S., & Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *The Journal of Experimental Biology*, 203(Pt 17), 2641–2656. <https://doi.org/10.1007/s00114-005-0622-4>
- Pauwels, J., Viol, I. Le, Azam, C., Valet, N., Julien, J.-F., Bas, Y., ... Kerbiriou, C. (2019). Accounting for artificial light impact on bat activity for a biodiversity-friendly urban planning. *Landscape and Urban Planning*, 183, 12–25. <https://doi.org/10.1016/j.landurbplan.2018.08.030>
- Penone, C., Kerbiriou, C., Julien, J.-F., Marmet, J., & Le Viol, I. (2018). Body size information in large-scale acoustic bat databases. *PeerJ*, 6, e5370. <https://doi.org/10.7717/peerj.5370>
- Pinaud, D., Claireau, F., Leuchtman, M., & Kerbiriou, C. (2018). Modelling landscape connectivity for greater horseshoe bat using an empirical quantification of resistance. *Journal of Applied Ecology*, 55(6), 2600–2611. <https://doi.org/10.1111/1365-2664.13228>
- Russo, D., & Voigt, C. C. (2016). The use of automated identification of bat echolocation calls in acoustic monitoring: A cautionary note for a sound analysis. *Ecological Indicators*, 66, 598–602. <https://doi.org/10.1016/j.ecolind.2016.02.036>
- Rydell, J., Nyman, S., Eklöf, J., Jones, G., & Russo, D. (2017). Testing the performances of automated identification of bat echolocation calls: A request for prudence. *Ecological Indicators*, 78, 416–420. <https://doi.org/10.1016/j.ecolind.2017.03.023>
- Santos, H., Rodrigues, L., Jones, G., & Rebelo, H. (2013). Using species distribution modelling to predict bat fatality risk at wind farms. *Biological Conservation*, 157, 178–186. <https://doi.org/10.1016/j.biocon.2012.06.017>
- Scott, C., & Altringham, J. (2017). *BatClassify*. Retrieved from <https://bitbucket.org/chrisccott/batclassify>
- Sirami, C., Jacobs, D. S., & Cumming, G. S. (2013). Artificial wetlands and surrounding habitats provide important foraging habitat for bats in agricultural landscapes in the Western Cape, South Africa. *Biological Conservation*, 164, 30–38. <https://doi.org/10.1016/j.biocon.2013.04.017>
- Stahlschmidt, P., & Brühl, C. A. (2012). Bats as bioindicators—The need of a standardized method for acoustic bat activity surveys. *Methods in Ecology and Evolution*, 3(3), 503–508. <https://doi.org/10.1111/j.2041-210X.2012.00188.x>
- Verboom, B., & Huitema, H. (1997). The importance of linear landscape elements for the pipistrelle *Pipistrellus pipistrellus* and the serotine bat *Eptesicus serotinus*. *Landscape Ecology*, 12(2), 117–125. <https://doi.org/10.1007/BF02698211>
- Waters, D., & Barlow, K. (2013). Bat detectors: Past, present and future. *British Wildlife*, 87, 86–92.
- Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer Science & Business Media: New York. <https://doi.org/10.1007/978-0-387-87458-6>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Barré K, Le Viol I, Julliard R, et al. Accounting for automated identification errors in acoustic surveys. *Methods Ecol Evol*. 2019;10:1171–1188. <https://doi.org/10.1111/2041-210X.13198>