Ecological Indicators 90 (2018) 346–355

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

# Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis

Juan Sebastian Ulloa[a,b,*], Thierry Aubin[b], Diego Llusia[a,b], Charles Bouveyron[c], Jérôme Sueur[a]

[a] Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, Sorbonne Université CNRS-MNHN-UPMC-EPHE, 57 rue Cuvier, CP 50, F-75005, Paris, France

[b] Equipe Communications Acoustiques, UMR 9197 Neuro-PSI-CNRS, Université Paris Sud, bat. 446, F-91405 Orsay, France

[c] Laboratoire MAP5, UMR CNRS 8145, Université Paris Descartes & Sorbonne Paris Cité, Paris, France

A R T I C L E I N F O

*Keywords:*
Ecoacoustic monitoring
Acoustic community
Unsupervised machine learning
Wavelets
Nocturnal soundscape

A B S T R A C T

Ecoacoustic monitoring has proved to be a viable approach to capture ecological data related to animal communities. While experts can manually annotate audio samples, the analysis of large datasets can be significantly facilitated by automatic pattern recognition methods. Unsupervised learning methods, which do not require labelled data, are particularly well suited to analyse poorly documented habitats, such as tropical environments. Here we propose a new method, named Multiresolution Analysis of Acoustic Diversity (MAAD), to automate the detection of relevant structure in audio data. MAAD was designed to decompose the acoustic community into few elementary components (soundtypes) based on their time–frequency attributes. First, we used the short-time Fourier transform to detect regions of interest (ROIs) in the time–frequency domain. Then, we characterised these ROIs by (1) estimating the median frequency and (2) by running a 2D wavelet analysis at multiple scales and angles. Finally, we grouped the ROIs using a model-based subspace clustering technique so that ROIs were automatically annotated and clustered into soundtypes. To test the performance of the automatic method, we applied MAAD to two distinct tropical environments in French Guiana, a lowland high rainforest and a rock savanna, and we compared manual and automatic annotations using the adjusted Rand index. The similarity between the manual and automated partitions was high and consistent, indicating that the clusters found are intelligible and can be used for further analysis. Moreover, the weight of the features estimated by the clustering process revealed important information about the structure of the acoustic communities. In particular, the median frequency had the strongest effect on modelling the clusters and on classification performance, suggesting a role in community organisation. The number of clusters found in MAAD can be regarded as an estimation of the soundtype richness in a given environment. MAAD is a comprehensive and promising method to automatically analyse passive acoustic recordings. Combining MAAD and manual analysis would maximally exploit the strengths of both human reasoning and computer algorithms. Thereby, the composition of the acoustic community could be estimated accurately, quickly and at large scale.

## 1. Introduction

The diversity of life forms is an invaluable biological resource threatened by anthropogenic environmental change (Pimm et al., 1995; Thomas et al., 2004). Given the pace of this change, there is an imperative need to develop quantitative indicators that provide specific information on the state of biodiversity (Pereira et al., 2013). With the advent of new sensor technology it is possible to remotely collect environmental data, assisting to determine, and eventually buffer, the pressures on biological diversity and ecosystem services (Petrou et al., 2015). In particular, the use of passive acoustic sensors in ecological research, or ecoacoustics (Sueur and Farina, 2015), has proved to be a viable method for biodiversity assessment that can be scaled up at multiple spatial and temporal scales (Towsey et al., 2014). The environmental sounds collected by these automated sensors usually include a large combination of both biotic and abiotic sounds, which are mixed down into a single time series. Such interlaced audio data needs to be unravelled in order to extract and to decipher ecological meaningful information, which represents to date a prominent bottleneck for the application of acoustic sensors in biodiversity monitoring.

A significant proportion of animal species produce sounds for social interaction, navigation or predator-prey encounters (Fletcher, 2014).

* Corresponding author at: Institut de Systématique, Évolution, Biodiversité, ISYEB, UMR 7205 CNRS-MNHN-UPMC-EPHE, Muséum national d'Histoire naturelle, Sorbonne Universités, 57 rue Cuvier, CP 50, F-75005 Paris, France.

*E-mail address:* juan.ulloa@mnhn.fr (J.S. Ulloa).

https://doi.org/10.1016/j.ecolind.2018.03.026
Received 10 November 2017; Received in revised form 7 March 2018; Accepted 12 March 2018
1470-160X/ 

Most of these acoustic signals have a species-specific signature that can be exploited for the remote identification of species. The use of these signatures is a direct way to retrieve ecological data about species presence, abundance, status and distribution. Manual species identification by experts can be carried on audio datasets, but for large collections, the analysis can be facilitated by automatic pattern recognition methods such as supervised learning (Kershenbaum et al., 2016). Supervised learning is a method to build a statistical classifier based on labelled training data (Webb and Copsey, 2011). An increasing number of supervised learning tools have been adapted to identify automatically single species (Dugan et al., 2013; Ganchev et al., 2015; Ulloa et al., 2016) or several species (Briggs et al., 2012; Potamitis, 2014; Heinicke et al., 2015; Dong et al., 2015; Xie et al., 2016; Ruiz-Muñoz et al., 2016). The application of supervised learning is limited by the large reference datasets required to 'train' the classifiers and the high acoustic similarity sometimes observed between closely related taxa. The available sound libraries, even if providing thousands of samples, still cover only a small fraction of the animal sound diversity, at both population and species scales.

An alternative to species identification consists in characterising the acoustic community or the soundscape with the use of acoustic indices (Sueur et al., 2014). Rather than focusing on target species, acoustic indices aim to describe the global structure of the soundscape. A variety of indices have been proposed and applied to terrestrial (Lellouch et al., 2014; Farina et al., 2015; Fuller et al., 2015) and underwater habitats (Parks et al., 2014; Desjonquères et al., 2015; Harris et al., 2016; Buscaino et al., 2016). These indices revealed, for example, changes in bird species richness among woodland habitats (Depraetere et al., 2012) or dynamics of the soundscape across different temporal scales (Rodriguez et al., 2014). However, they also showed to be sensitive to transitory or permanent background noise, variation in the distance of the animals to the sensor, and the relative sound amplitude or the calling rate of the signalling animal (Gasc et al., 2015; Kendrick et al., 2016).

More recently, methods based on unsupervised learning have been adapted to audio recordings achieved in natural environments. Unsupervised learning searches for structures or patterns in a dataset without using labels. This approach has been extensively used to draw inferences in areas where labelled data is inaccessible or too expensive, such as astronomy (Way, 2012), genetics and genomics (Libbrecht and Noble, 2015). In an innovative work, Eldridge et al. (2016) adapted sparse-coding and source separation algorithms to extract shift-invariant spectro-temporal "atoms" from environmental recordings. However, the authors did not establish a clear link between the spectro-temporal "atoms" and ecological or biological processes. Unsupervised learning has also been used as a pre-processing step for the classification task, significantly improving the classification performance on species recognition (Stowell and Plumbley, 2014). In their approach, Stowell and Plumbley (2014) first decomposed the sounds into "atoms" with spherical k-means, and then used the "atoms" as features for the supervised learning framework. Thus, unsupervised learning offers new means to characterise sounds and may provide insights on the acoustic communities of diverse and threatened ecosystems, such as those of tropical regions (Pekin et al., 2012; Rodriguez et al., 2014).

The present work emerges from the question: how to best measure, quantify and characterise environmental sounds (from biotic and abiotic sources) in passive acoustic recordings to get valuable ecological indicators? We propose a new data-driven method, named Multiresolution Analysis of Acoustic Diversity (MAAD), to automate the discovery of plausible and interpretable patterns in passive acoustic recordings. To build a generalized method for multiple conditions and environments, we adapted methods from the unsupervised learning field. We estimated acoustic diversity by detecting regions of interest in sound recordings and grouping them into soundtypes based on the value of their time-frequency attributes. To test the flexibility and robustness of the method, we applied MAAD to two distinct night tropical environments in French Guiana, a lowland high rainforest (HF) and a rock savanna (RS). The RS is inhabited by a distinct and likely less diverse animal community in comparison with the HF (Bongers et al., 2001) so that it was expected to find contrasting acoustic communities between these two tropical environments. We compared manual and automated annotations to (1) evaluate the model selection procedure; (2) assess the relevance of different features in the clustering process; and (3) quantify the overall similarity between manual and MAAD soundtypes. To conclude, we give practical advices and discuss how MAAD can potentially be transferred to other environments in order to track the state and dynamics of animal communities for biodiversity studies.

## 2. Material and methods

The workflow of the proposed method (MAAD) followed four main steps: (1) passive acoustic recordings were transformed into the time-frequency domain using the windowed short-time Fourier transform and the Fourier coefficients were filtered to remove noise and to highlight sounds that can be delimited in time and frequency, here defined as regions of interest (ROIs); (2) each ROI was then characterised by features in the time-frequency domain using 2D wavelets; (3) the ROIs with their attributes were used to automatically estimate clustering hyper-parameters; and (4) the hyper-parameters and the attributes of the ROIs were passed to a clustering algorithm that formed homogenous groups of ROIs, namely soundtypes (Fig. 1). This led to an automatic partitioning and characterization of soundtypes, which can be used to determine their presence, relative abundance and diversity within acoustic communities. To validate the proposed approach, the automatic partitioning provided by MAAD was compared to expert manual annotations using the adjusted Rand index (ARI).

### 2.1. Audio dataset

Audio data were collected in French Guiana at the CNRS Nouragues Research Station (4°05′N; 54°40′W). The station is mainly occupied by a lowland high rainforest (HF) and a rock savanna (RS), among other ecosystems. The HF dominates on lower parts of the station (40–100 metres above sea level), has a fairly open understory and is closed on top by a dense canopy elevating at 25–35 m. The tree density in HF is

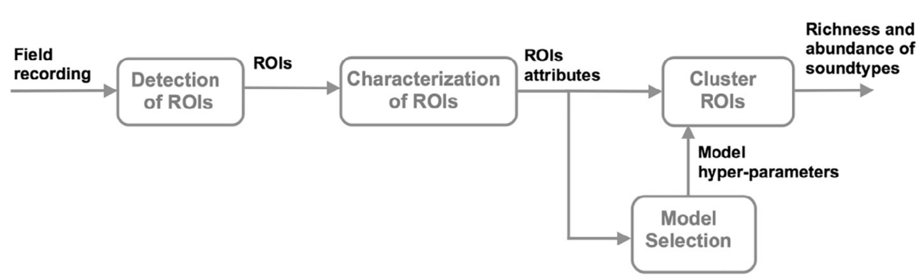


**Fig. 1.** Block diagram of MAAD. Each step of the workflow is depicted as a grey box. Input and output after each step are indicated in black. Model selection is an optional step. Model hyper-parameters can also be set based on prior information about the acoustic community.

high, with a total basal area ranging from 30 to 45 $m^2$/ha, and the floristic composition is heterogeneous, since no species dominates the site. The RS is found on a granite hill about 400 m above sea level that rises abruptly and overtops the forest. Due to high microclimatic fluctuations and poorly developed soils, the RS is only partially colonised by vegetation, being its floristic composition drastically different from the surrounding forest. Small trees, xerophytic herbs and shrubs in scattered patches separated by rock areas covers nearly half of the area of the granite hill (Sarthou, 2001).

Environmental sounds were gathered using automated acoustic sensors (Songmeter SM2, Wildlife Acoustics Inc., Concord, MA, USA) equipped with omnidirectional microphones (PUI Audio POM-3535L-3-R, frequency response 50 Hz–20 kHz ± 4 dB). A single acoustic sensor was placed at each environment, HF (04°05′15″N; 52°40′42″W) and RS (04°05′33″N; 52°40′40″W), and recorded one minute every 30 min from sunset to sunrise for 10 consecutive nights (5–15 December 2014). Each sensor was set to sample the audio at 44.1 kHz with a 16-bit resolution (mono, WAV format). This audio database was subsampled by selecting two one-minute samples per night, one four hours after sunset (22 h 17 min UTC/GMT-3 h) and one four hours before sunrise (02 h 24 min UTC/GMT-3 h). These environmental audio recordings were deposited at the sound library of the Muséum national d'Histoire naturelle (www.sonotheque.mnhn.fr, Table S2). At both sites, two files recorded during heavy rain were removed. The final audio database included 36 one-minute files.

### 2.2. Detection of regions of interest (ROIs)

A region of interest is an isolated region in the time-frequency domain with a high density of energy. The automated detection of ROIs followed a four-step process computed with MATLAB (The MathWorks, Inc., Natick, MA) using the signal processing toolbox. First, we computed a spectrogram of the audio signal using the windowed short-time Fourier transform, (1024 FFT length, 50% overlap, Hamming window). Second, we applied a denoising method, namely *spectral subtraction* (Boll, 1979; Yu et al., 2008), which allows to highlight transitory sounds by removing stationary noise found in the background. Third, we used a 2D rotationally symmetric Gaussian filter to remove small impulsive noise and to join close-by regions of high-density energy (5 by 5 element size, 0.5 standard deviation). As a final step, we applied a linear amplitude threshold to select the regions that were in the foreground. Since the spectrogram gives a sparse representation of the acoustic environment, regions of high density of energy can be identified as observations distant from the low-density background. Hence, the linear threshold (*lth*) was set for each recording by evaluating the dispersion of the spectrogram values and selecting values of the spectrogram distributed one-and-a-half times the inter-quartile range (IQR) above the third quartile ($lth = Q_3 + 1.5 \times IQR$). The use of quartiles gives a robust measure of central tendency and spread effective to non-normal data (Tukey, 1977).

Thereby, each detected ROI was a frame of variable size in the time-frequency domain, delimited by a start and end time, and a minimum and maximum frequency. The number of ROIs found in the RS and the HF audio files were respectively 4028 and 5375, for a total of 9403.

### 2.3. Characterization of ROIs

Automated measurements on the frequency and the time-frequency shape of each ROI were performed. To measure the frequency, a single feature was calculated: the median frequency, which is the value that divides the ROI into two frequency intervals of equal energy. This is a robust measurement that does not vary much based on the exact time-frequency bounds.

To measure the shape of the ROI in the time-frequency domain a wavelet analysis was used. The purpose of this procedure was to decompose the signal into coefficients that can be saved and manipulated to better represent the information in the signal. The wavelet transform is the result of filtering the signal with a bank of specific filters (or 'wavelets'). Each analysing wavelet can be visualised as a kernel of fixed scale that moves along the data. When the wavelet encounters a feature in the data with similar shape and scale, the analysis returns a high value for the wavelet coefficient. Then, the operation is repeated at a different scale with a new dilated or contracted wavelet. In this way the wavelet transform allows a multiresolution analysis and can represent hierarchical structures of the data. This scale-by-scale analysis is particularly suited for the detection of local features in aperiodic data. Wavelets can be extended to the two dimensional case (2D), in particular to process images (Mallat, 2009). In 2D, wavelets are dilated as in the one-dimensional analysis, and in addition rotated. A 2D wavelet transform of a spectrogram allows finding co-occurrence of time–frequency elements at different scales.

First, the high frequencies were recovered by convolution with the wavelet filters. By rotating and dilating the wavelet, we obtained rotation and scale covariant coefficients, which allowed discriminating the differences in shape of the different ROIs. Then, each filtered signal was averaged with a rotation-invariant low-pass filter. The rotation-invariant low-pass filter removed small differences between similar ROIs, forming homogeneous groups. The operation on a 2D signal $x$ is formalised as:

$$Sx = (|x * \psi_{j,\theta}| * \phi) \tag{1}$$

where the symbol $*$ denotes spatial convolution, $\phi$ is a gaussian low-pass filter and $\psi_{j,\theta}$ is a wavelet dilated by $2^j$ and rotated by an angle $\theta$. The filter bank used consisted of wavelets of the Morlet family, at 16 scales and 8 angles: horizontal (0°), vertical (90°) and diagonals (22.5°, 45°, 67.5°, 112.5°, 135°, 157.5°). In this way, a total of 128 shape features were calculated. An illustrative subset of the 2D filters is presented in Fig. 2. The filter bank and the coefficients were computed with MATLAB (The MathWorks, Inc., Natick, MA) using the ScatNet toolbox (http://www.di.ens.fr/data/software/scatnet/; Sifre and Mallat, 2013).

### 2.4. High dimensional clustering

Clustering is an unsupervised learning analysis that aims at grouping objects into homogenous groups or clusters. As opposed to supervised learning, clustering is more flexible since no groups need to be defined a priori, *i.e.* the groups are formed based on the value of the attributes of the data. If available, labelled data can be used to estimate whether the groups found are suitable classes. To group the ROIs in homogeneous groups, a method suited to the multidimensional attributes of the ROIs was used. This method, named *High Dimensional Data Clustering* (HDDC), is a clustering technique based on a family of twelve parsimonious Gaussian mixture models adapted to multivariate high-dimensional data (Bouveyron et al., 2007). The mixture model-based

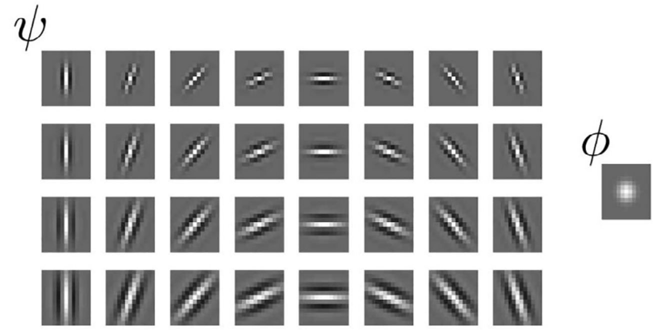


**Fig. 2.** Subset of the 2D wavelet filter bank used to capture spectro-temporal features of the signal. On the left, Morlet wavelets $\psi$ at four scales (along rows) and eight angles (along columns) are illustrated. On the right, the gaussian low-pass filter $\phi$ is represented.

clustering (on which HDDC is based) is defined in a probabilistic framework and has two particular advantages: (1) it is known to be a robust approach to deal with unbalanced datasets and (2) it is interpretable from a statistical point of view (Fraley and Raftery, 2002). The mixture model is naturally robust to unbalanced data sets because of the parameter $\pi_k$, which correspond to the weight of the group component in the mixture (see Eq. (1), Text S6). The additional advantage of the mixture model is that it is a comprehensible statistical model and therefore allows to use model selection techniques, such as the Slope heuristics which we use later in the proposed framework.

The models proposed in HDDC have different regularizations that control the complexity of the clustering. The most complex model is $a_{kj}b_kQ_kd_k$, all the parameters are class-specific and the dimension is specific to each cluster. The simplest model is $abQ_kd$, all the parameters are common between classes and the dimension of the class subspace is common. The properties of the parsimonious models in HDDC are detailed in Text S6.

A model selection procedure was implemented to estimate the hyper-parameters that control the complexity of the model. These hyper-parameters are the model *M,* the number of groups *K,* and the - threshold value *th* to find the intrinsic dimensionality of each class. Classical model selection methods, namely AIC (Akaike, 1974) and BIC (Schwarz, 1978) criteria, are asymptotic (they assume that *n* tends to infinity) and therefore might not be appropriate. More recently, Birgé and Massart (2006) proposed a data-driven technique that alleviates this assumption and was used in different situations (Baudry et al., 2011), including model-based clustering (Bouveyron et al., 2015). The method, called *slope heuristics* (SHC), of the model *M* is defined as follows:

$$SHC(M) = l(\hat{\theta}) - 2\hat{s}\xi(M) \tag{2}$$

where $\theta$ is the set of parameter values that maximize the log-likelihood function $l(\hat{\theta})$, $\xi(M)$ is the number of free parameters of the model, and $\hat{s}$ is the slope of the linear part of $l(\hat{\theta})$ with regard to the number of parameters. SHC follows the same rationale than other model selection criteria such as BIC and AIC, the likelihood of the fitted model is penalised by a function. Yet, SHC criterion has been found to be more consistent than BIC for model selection in HDDC (Bouveyron et al., 2015). A detailed overview and practical implementation advice of the SHC can be found in Baudry et al. (2011).

Slope heuristics were calculated for the twelve models implemented in HDDC (see Text S6), at ten different thresholds (0.0001, 0.0005, 0.001, 0.01, 0.03, 0.05, 0.07, 0.1, 0.15, 0.2), for 39 values of *K* (from 2 to 40, by steps of one). Since HDDC has a random initialization, the returned log-likelihood can vary between executions. Hence, the slope heuristics value was calculated ten times for each combination of hyper-parameters. The mean value was stored and the maximum was selected to find the hyper-parameters of the HDDC models.

With the hyper-parameters fixed, the model was fitted ten times with random initialisation. Random initialisation is a standard method to initiate the Expectation-maximization algorithm. This method correctly explores the parameter space to reach the global maximum of the likelihood (Biernacki et al., 2003). Among the ten trials, only the model with the highest likelihood was selected for feature importance analysis and validation. Feature importance was calculated by multiplying the vector of estimated variances by the corresponding orthogonal matrix of orientations $Q_k$ on each cluster. The total weight of the features is the average of the feature importance on all clusters. HDDC and slope heuristics were both computed using the R package HDclassif (Bergé et al., 2012).

### 2.5. Validation of system performance

To evaluate and determine the performance of MAAD, proper ground truth was established and a quantitative method to compare ground truth and the system output was used. ROIs ($n = 9403$) of sound recordings, which were automatically detected, were examined manually using the software Raven (Bioacoustic Research Program, 2014). Aural and visual inspection of spectrograms, plus manual measurements on the temporal (duration and pulse rate) and spectral (median frequency and bandwidth) domain were made. Based on this combined examination, ROIs were categorised into distinct homogeneous groups, here referred as soundtypes. If the amplitude of the sound was too low and the features could not be inspected correctly, the ROI was marked as undetermined.

The automatic annotations output by MAAD were compared with the manual annotations using the adjusted Rand index (ARI). The ARI is a similarity measure between two partitions (Hubert and Arabie, 1985). Given two partitions, U and V, derived from a set of *n* objects, the ARI is computed according to:

$$ARI = \frac{\binom{n}{2}(a+d) - [(a+b)(a+c)+(c+d)(b+d)]}{\binom{n}{2}^2 - [(a+b)(a+c)+(c+d)(b+d)]} \tag{3}$$

where *a* denotes objects in a pair placed in the same group in U and in the same group in V; *b* denotes objects in a pair placed in the same group in U and in different groups in V; *c* denotes objects in a pair placed in the same group in V and in different groups in U and; and *d* denotes objects in a pair placed in different groups in U and in different groups in V. This index, bounded between ± 1, was derived from the popular Rand index but has the advantage of being adjusted for chance with respect to the null hypothesis and can be interpreted as the difference between probabilities of concordance and discordance. Independently of the number of clusters and samples, the ARI has a value of -1 when the partitions are opposed, close to 0 for random labelling, and exactly 1 when the two partitions are identical.

Clustering analyses, cluster validation and graphs were achieved with R version 3.2.0 (R Core Team, 2017).

## 3. Results

We manually identified 35 soundtypes in the HF and 18 in the RS dataset. The relative soundtype abundance was unbalanced in both datasets (Fig. S1). On average manual annotation required 25–35 min per file. Manual annotations were used only for performance validation purposes, that is, to interpret the return of MAAD at different settings. Two main tests were performed. The first one consisted in changing the hyper-parameter *K* of the model, from 2 to 40 by unitary steps. The second one consisted in using different subsets of features: diagonal wavelets (16 scales × 6 = 96 features), horizontal and vertical wavelets (16 scales × 2 = 32 features), shape (32 + 96 = 128 features), median frequency (1 feature), and the full set (128 + 1 = 129 features).

### 3.1. Model selection

To begin with the cluster analysis, the most adequate model hyper-parameters were identified by observing the trend of the slope heuristics criterion. On both datasets, RS and HF, slope heuristics attained a maximum value with the model $a_{kj}b_kQ_kd_k$, the most complex one (a full covariance matrix for each group), and a threshold value of 0.0005. As expected, the suitable number of clusters *K* was different for each habitat. The curve showing the evolution of the slope heuristics value for different K peaked between 10 and 15 with a maximum at 11 on the RS dataset and peaked between 15 and 20 with a maximum at 17 on the HF dataset (Fig. 3).

Using the manual annotations, different settings of the hyper-parameter *K* (*i.e.* the number of clusters) were tested to analyse the response of MAAD. The response at different values of K was similar in both datasets (Fig. 4), the performance of the clustering increments at the

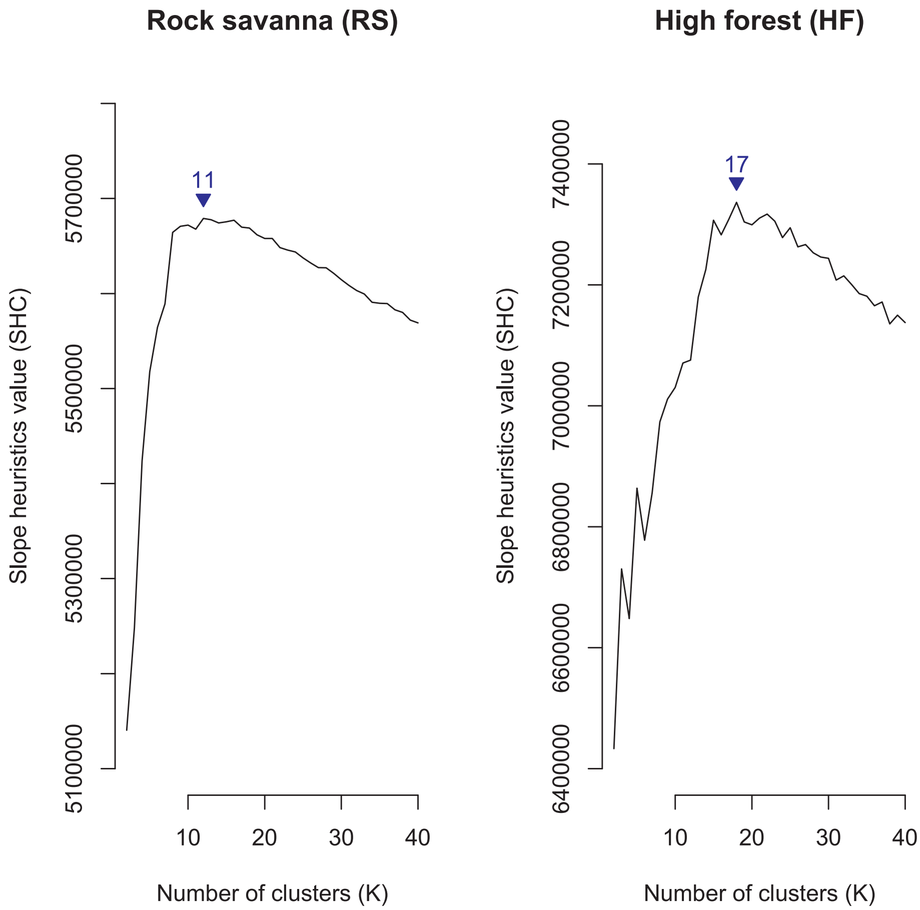


**Fig. 3.** Model selection using slope heuristics. Variation of the slope heuristics criterion with respect to the number of clusters ($K$) for the RS and HF datasets. Slope heuristics find its maximum for RS at 11 clusters, and for HF at 17 clusters. This maximum is found for RS and HF with the same mixture model ($a_{kj}b_kQ_kd_k$) and threshold value (0.0005).

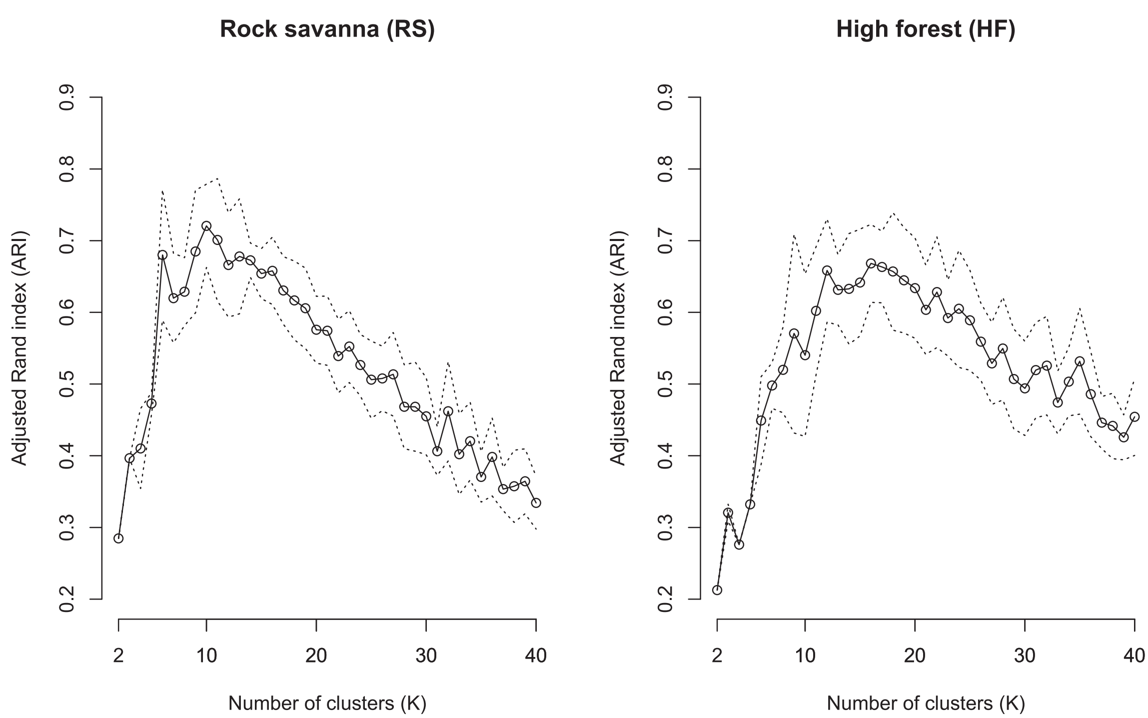


**Fig. 4.** Variation of the performance of MAAD with respect to different values of the hyper-parameter K, number of clusters. The performance of MAAD is measured with the adjusted Rand index (ARI). The returned ARI value was calculated for 10 random initialisations of the clustering. The solid line represents the mean value and the dashed lines indicate the standard deviation of the results.

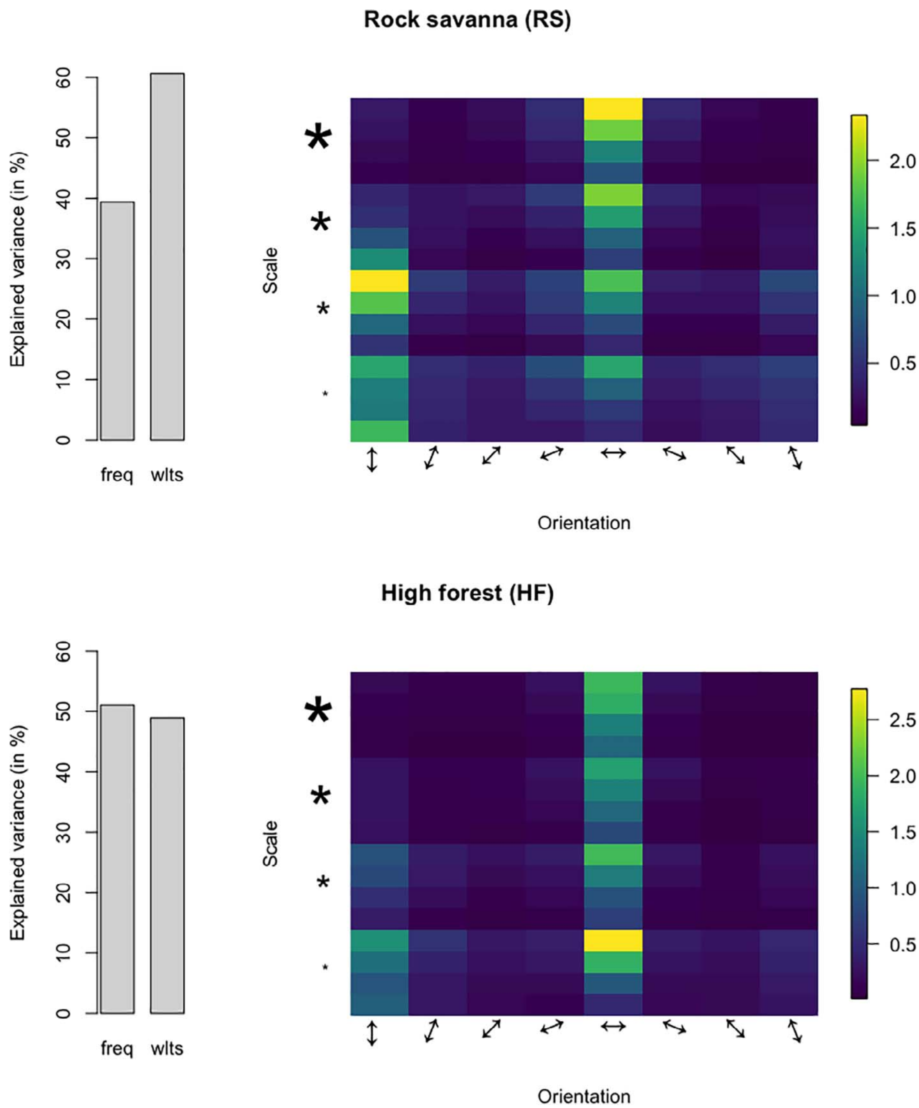


**Fig. 5.** Representation of the amount of variance accounted for each of the 129 features used on the clustering process. The bar diagram (left) compares the median frequency (freq) and the sum of the 128 wavelets features (wlts). The intensity map (right) compares the relative importance of wavelets features at different angles and sizes, with dark blue indicating lowest value, and bright yellow the highest value.

beginning and after reaching a peak, the performance begins to drop progressively. The peak value differs for the two habitats, 9 for the RS and 15 for the HF.

### 3.2. Feature relevance

Using the hyper-parameters found with the slope heuristics, the ROIs were automatically clustered based on their computed time-frequency attributes (129 features). Before evaluating the clustering results, the weight of the features estimated by the clustering process were analysed. Interestingly, a single feature, the median frequency, accounted for most of the variation in the data, 39.4% and 51.0% for the RS and HF respectively (Fig. 5). The rest of the variation was associated to the combined wavelet features (n = 128) related to the time-frequency shape of the sound. The relative importance of each of the 128 wavelet features was plotted on an intensity map, a graphical representation of a matrix where each cell is highlighted according to its value. The intensity map showed that in both habitats the same two orientations explained best the data variance, the vertical (90°) and horizontal (0°) orientations (Fig. 5). However, different scales are emphasized in each habitat, medium and large scales in the RS, and small scales in the HF. Wavelet features at diagonal angles (22.5°, 45°, 67.5°, 112.5°, 135° and 157.5°), which are related in the time-frequency domain to upsweeps and downsweeps, explained the residual data variance.

In light of the weights of the features estimated during the clustering process, different subsets of features were tested (diagonal wavelet orientations, perpendicular wavelet orientations, all wavelets, median frequency and the full set) and contrasted with the manual annotations to further examine the response of MAAD. The global return on both datasets, RS and HF, was the same (Fig. 6). The model including only the diagonal wavelets (22.5°, 45°, 67.5°, 112.5°, 135° and 157.5°) gave the lowest ARI value. A higher ARI was obtained when using the horizontal and vertical wavelets (0° and 90°). By including all the wavelet features the result was improved again. By using only the median frequency of the ROIs, the results were even better than using all the 128 wavelet features. Finally, the best result (ARI value of 0.77 for RS and 0.85 for HF) was obtained by combining all the features, shape and median frequency.

### 3.3. Clustering results

ROIs were grouped into soundtypes through an unsupervised framework using the hyper-parameters returned by the slope heuristics

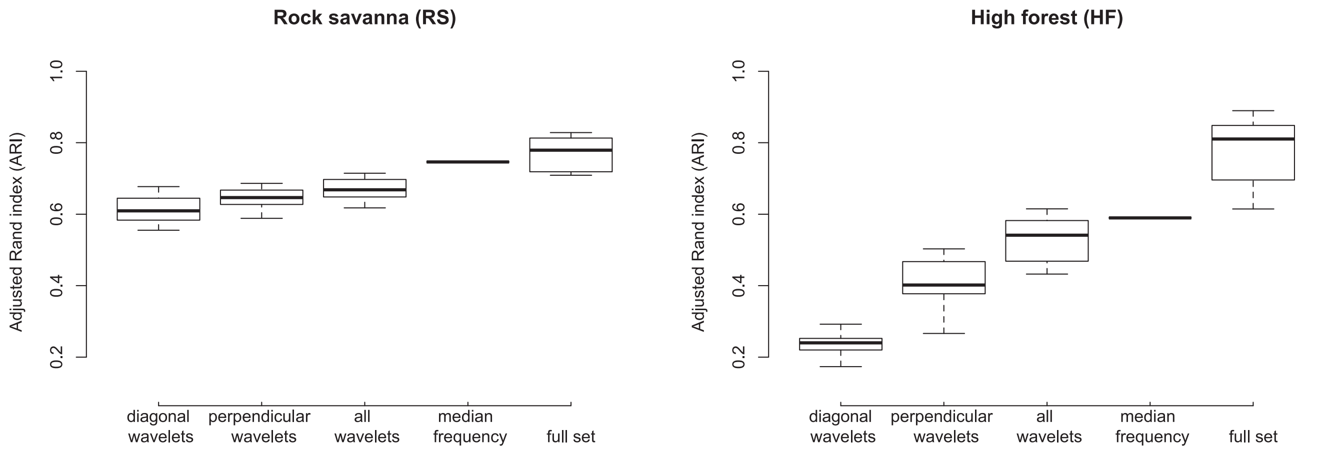


**Fig. 6.** Response of MAAD using different feature sets: time frequency shape described by diagonal wavelets (96 non-perpendicular features), perpendicular wavelets (32 features), all wavelets (96 + 32 = 128 features), the median frequency (1 feature), and the full set of features (129 features). The performance was measured with the ARI metric computed over 10 trials. All but one feature set, the median frequency, had random initialisation. There is no box for the median frequency because univariate clustering had deterministic initialisation.

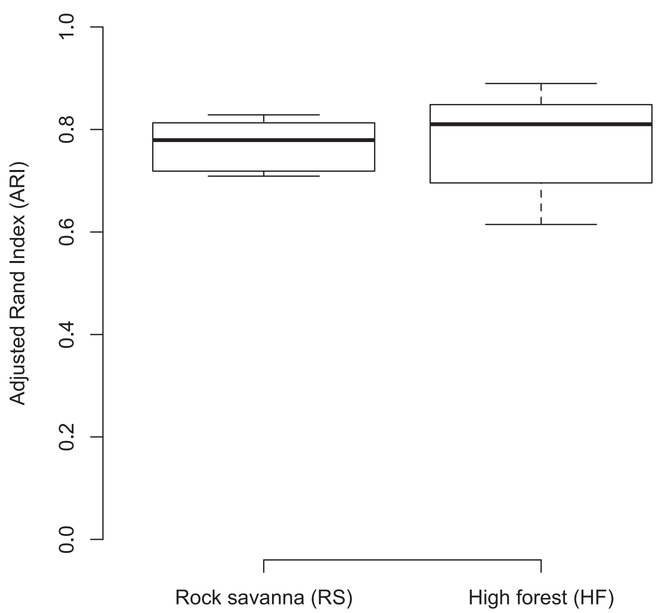


**Fig. 7.** Global classification performance of MAAD for RS and HF datasets measured with the ARI metric computed over 10 trials with random initialisation. The ARI is bounded between ± 1, has a value close to 0 for random labelling and exactly 1 when the two partitions are identical.

criterion and providing the full set of spectro-temporal features. Comparative analysis showed a high concordance between manual and automatic partitions with an ARI of 0.77 and 0.85 for the RS and the HF environments respectively (Fig. 7). In general, the random initialisation of the clustering algorithm induced a relatively small variation on the result (s.d. < 0.13), compared to the possible variation of the ARI index (from −1 to +1). Detailed analysis by soundtype showed that most of the errors were due to clusters splitting (Table S4). A visual example of the final output is depicted on Fig. S3.

The average computing time to process a one-minute file through the complete pipeline was 45.67 s on a desktop computer (3.4 GHz Intel Core i5 processor, 8 GB memory). Automatic annotation was on average forty times faster than human annotation.

## 4. Discussion

The animal acoustic diversity is known to potentially carry relevant ecological information related to the species diversity (Riede, 1993; Krause and Farina, 2016). However, it is still challenging to use automated statistical tools to analyse and extract ecological meaningful information from passive acoustic recordings. MAAD was designed to overcome this barrier enabling to analyse environmental audio recordings by automatically decomposing the acoustic community into few elementary components based on their time-frequency attributes. Our experiments showed that the partitions derived by MAAD in distinct tropical acoustic communities were highly similar to the ones obtained by meticulous manual (aural and visual) inspection. In addition, MAAD showed that some specific features were more informative for the clustering model, revealing potential structures that partition the acoustic community.

### 4.1. Model selection

The number of soundtypes in an assemblage (*i.e.* the acoustic richness) is a common measure of the acoustic diversity. Slope heuristics indicated that the most appropriate model for decomposing the HF dataset had to include more clusters ($K = 17$) than the RS dataset ($K = 11$). A higher hyper-parameter $K$ represents higher acoustic diversity in the HF, which is a result that matches our expectations and manual annotations. However, more soundtypes were found manually than automatically. In a closer look, we observed that common soundtypes were clustered correctly (e.g. $A_1$, $A_5$ on the RS, and $A_3$, $A_5$ on the HF, see Supplementary data), while rare soundtypes with less than 20 samples were not identified (e.g. $A_{14}$, $A_{16}$ on the RS, and $A_{24}$, $A_{33}$ on the HF, see Supplementary data). Slope heuristics makes a balance between the likelihood and the complexity of the model. As rare soundtypes are represented by a small number of samples (less than 20 samples), they do not increment the likelihood considerably to represent new clusters; instead, they are absorbed by larger clusters. Therefore, the number of clusters found in MAAD has to be regarded as the richness of common soundtypes in a given environment. In other words, soundtypes with infrequent presence in the recorded time series are expected to have low likelihood to be detected. As in many other sampling techniques in ecology, rare and elusive species are difficult to detect.

It is also important to note the resemblance between the slope heuristics trend and the response of the system with respect to increment of the hyper-parameter K, the number of clusters. In particular, the value of K selected by slope heuristics (11 and 17 for RS and HF respectively) is close to the value of K with the highest ARI value (9 and 15 for RS and HF respectively). Slope heuristics allows finding automatically a plausible number of clusters in relation to clustering performance, meaning that this criterion seems to be a suitable alternative to the human supervision.

### 4.2. Feature relevance

Generative modelling, such as HDDC, builds a full model of the distribution of features in each group. These models can be analysed to understand what group properties are the most important for clustering the objects. In our framework, the weight of the features estimated by the clustering process revealed important information about the structure of the acoustic community. The median frequency had the strongest effect on modelling the clusters. In other words, frequency predicted soundtype identity better than all the shape features. Partitioning the transmission channel in different frequency bands appears to be a common strategy to avoid masking, although other mechanisms may also generate the same pattern. Our results are congruent with frequency partitioning, which has been previously observed on assemblages of crickets (Schmidt et al., 2013), cicadas (Sueur, 2002) and amphibians (Villanueva-Rivera, 2014). Frequency over dispersion allows a great number of co-occurring signals to be accommodated in a limited acoustic space. Formulated under the acoustic niche hypothesis, organisms would have evolved to occupy specific spectro-temporal 'niches', decreasing the risk of heterospecific mating and information masking (Krause, 1993). Alternatively, many other selective pressures might be responsible for signal divergence and acoustic partitioning, such as those related to body size or female preferences (Gerhardt, 1994).

The acoustic space has multiple dimensions and the frequency is just one of them. Other dimensions, such as the shape features, had a lower but significant impact. The shape features derived with the 2D wavelets were also important features to derive the clusters. In particular, vertical and horizontal wavelets (0° and 90°) had a significant effect on the clustering process. These features are based on the spectrogram representation of the signal; therefore, most of the sounds were clustered based on variations in the duration of the sound and variations in the frequency bandwidth. Diagonal wavelets had less importance in the model learned. This outcome was expected since insects and amphibians, which dominated the studied acoustic communities, are known to produce sounds with few frequency modulations (Gerhardt and Huber, 2002).

### 4.3. Clustering results

Signalling animals produce redundant and species-specific sounds, which result in intuitive clusters. Based on this observation, MAAD was designed to give a representation based on a combination of elementary components (soundtypes) to form a whole (the acoustic community). To our knowledge, only Eldridge et al. (2016) attempted a similar approach to characterise the acoustic communities or soundscape. Both approaches, Eldridge et al. (2016) and ours, are based on unsupervised learning techniques, however, the aim and the evaluation of the result differ significantly. Eldridge et al. (2016) measured the ability to reconstruct a soundscape based on few spectro-temporal 'atoms' as a way to measure the 'decomposability' of a scene, and the evaluation was completely visual. In contrast, we aimed at finding ecologically plausible and interpretable 'atoms' or soundtypes. We evaluated our approach by comparing manual versus automated partitioning by using an objective measure of similarity, the ARI. Unfortunately, the differences in the objectives and on the evaluation procedure make our work hard to be confronted to Eldridge's and colleagues work.

MAAD was tested under two contrasting scenarios and gave robust clustering results, with high and consistent similarity between manual and automated partitions. This suggests that the elementary time-frequency components found by MAAD are interpretable and that the output can be used in further analysis for studies in ecology and evolution. For instance, the number of items in each cluster corresponds to the relative abundance of each soundtype. This information can be used to compute diversity indices such as Shannon, Simpson or Whittaker indices (Magurran, 2004), returning an estimation of local acoustic diversity. Alternatively, after processing with MAAD, a manual inspection of detected soundtypes may enable to establish a direct link between MAAD clusters and species. For example, HF cluster number $B_5$ could be identified as stridulations of the cricket *Lernecella minuta*, and RS cluster number $B_6$ could be identified as vocalisations of the amphibian *Hypsiboas boans* (Supplementary data, Fig. S5). This semi-supervised framework would allow to annotate efficiently large sound databases for deeper analyses.

The clustering errors were mainly due to the division of major ground-truth clusters into homogeneous subgroups. The marked unbalanced nature of the dataset played an important role on this clustering subdivision. Clusters with many observations have a stronger weight maximising the overall likelihood of the model than clusters with rare observations. Since the parameters of the model were estimated so as to maximise the global likelihood of the model, the likelihood was incremented by splitting large clusters instead of creating new small clusters. Cluster splitting was also observed in the study of the response of the model to the variation of *K*. After reaching a peak, at a lower *K* than the true number of groups, the performance measured by the ARI dropped in both datasets. The ARI measures the number of ROIs correctly partitioned and hence the performance measure was mainly impacted by the splitting of large clusters and less by the wrong clustering of small ones. This also explains why the clustering results were highly accurate even if the soundtype richness found by the unsupervised procedure was lower than that by the manual one. Interestingly, the division of clusters with large observations still resulted in homogeneous groups, which could be assessed and combined by manual inspection. Further research is necessary to evaluate the performance on other scenarios in order to validate this method across the diverse acoustic communities found in practice. These tests would also be valuable to assess the error propagation of the system, identifying the potential sources of error and exploring how they influence the results.

MAAD is an adaptable framework that can be coupled with expert knowledge. An advantage of model-based clustering, which is used by MAAD, is that the uncertainty for an observation to belong to a cluster is measured by a posterior probability. Observations with probabilities drifting from 1 could be subsequently flagged and assessed by an expert. Combining MAAD and manual analysis would maximally exploit the strengths of both human reasoning and computer algorithms. Thereby, the composition of the acoustic community could be estimated accurately, quickly and at large scale.

## Data accessibility

The environmental audio recordings were deposited at the sound library of the Muséum national d'Histoire naturelle (www.sonotheque.mnhn.fr). The collection number of each file is presented on Table S2. Source codes (Matlab and R) are available at: https://github.com/juansulloa/maad_matlab. A step by step instruction to run the analysis is provided in Text S7.

## Acknowledgements

This research was supported by the Labex CEBA (Centre d'Étude de la Biodiversité Amazonienne) and the SABIOD MASTODONS Big Data (CNRS MI project 2012–2017). We would like to warmly thank Philippe Gaucher, Elodie Courtois and Patrick Châtelet for their help during our stay at the Nouragues research station. We specially thank Julio Pedraza who provided valuable advices on the use of high performance computing clusters to run batch process. We also would like to thank Laure Desutter-Grandcolas and Elodie Courtois for their help in identifying clusters to a species-specific acoustic signal. DLL was supported by the Fyssen Foundation (Post-doctoral Grant). JSU was supported by COLCIENCIAS (Doctoral Scholarship of the Colombian government). The authors declare no conflict of interest. We finally wish to acknowledge the valuable help of two anonymous referees for their remarks and comments on the manuscript.

### Author contribution statement

JSU, TA, CB, DLL and JS conceived the ideas and designed methodology; JSU and DLL collected the acoustic data; JSU, TA, CB and JS analysed the data; JSU and JS led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolind.2018.03.026.

### References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723.
Baudry, J.-P., Maugis, C., Michel, B., 2011. Slope heuristics: overview and implementation. Stat. Comput. 22, 455–470.
Bergé, L., Bouveyron, C., Girard, S., 2012. HDclassif: an R package for model-based clustering and discriminant analysis of high-dimensional data. J. Stat. Software 46.
Biernacki, C., Celeux, G., Govaert, G., 2003. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. Comput. Stat. Data Anal. 41, 561–575.
Bioacoustic Research Program, 2014. Raven Pro: Interactive Sound Analysis Software. The Cornell Lab of Ornithology, Ithaca, NY.
Birgé, L., Massart, P., 2006. Minimal penalties for gaussian model selection. Probab. Theory Relat. Fields 138, 33–73.
Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. 27, 113–120.
Bongers, F., Charles-Dominique, P., Forget, P.M., Théry, M., 2001. Nouragues: Dynamics and Plant-animal Interactions in a Neotropical Rainforest. Kluwer Academic Publishers, Dordrecht Boston.
Bouveyron, C., Girard, S., Schmid, C., 2007. High-dimensional data clustering. Comput. Stat. Data Anal. 52, 502–519.
Bouveyron, C., Côme, E., Jacques, J., 2015. The discriminative functional mixture model for a comparative analysis of bike sharing systems. Annal. Appl. Stat. 9, 1726–1760.
Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J., Hadley, A.S., Betts, M.G., 2012. Acoustic classification of multiple simultaneous bird species: a multi-instance multi-label approach. J. Acoust. Soc. Am. 131, 4640–4650.
Buscaino, G., Ceraulo, M., Pieretti, N., Corrias, V., Farina, A., Filiciotto, F., Maccarrone, V., Grammauta, R., Caruso, F., Giuseppe, A., Mazzola, S., 2016. Temporal patterns in the soundscape of the shallow waters of a Mediterranean marine protected area. Sci. Rep. 6, 34230.
Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., Sueur, J., 2012. Monitoring animal diversity using acoustic indices: implementation in a temperate woodland. Ecol. Ind. 13, 46–54.
Desjonquères, C., Rybak, F., Depraetere, M., Gasc, A., Le Viol, I., Pavoine, S., Sueur, J., 2015. First description of underwater acoustic diversity in three temperate ponds. PeerJ 3, e1393.
Dong, X., Towsey, M., Truskinger, A., Cottman-Fields, M., Zhang, J., Roe, P., 2015. Similarity-based birdcall retrieval from environmental audio. Ecol. Inf. 29 (Part 1), 66–76.
Dugan, P., Pourhomayoun, M., Shiu, Y., Paradis, R., Rice, A., Clark, C., 2013. Using high performance computing to explore large complex bioacoustic soundscapes: case study for right whale acoustics. Procedia Comput. Sci. 20, 156–162.
Eldridge, A., Casey, M., Moscoso, P., Peck, M., 2016. A new method for ecoacoustics? toward the extraction and evaluation of ecologically-meaningful soundscape components using sparse coding methods. PeerJ 4, e2108.
Farina, A., Ceraulo, M., Bobryk, C., Pieretti, N., Quinci, E., Lattanzi, E., 2015. Spatial and temporal variation of bird dawn chorus and successive acoustic morning activity in a Mediterranean landscape. Bioacoustics 24, 269–288.
Fletcher, N.H., 2014. Animal bioacoustics. In: Rossing, T.D. (Ed.), Springer Handbook of Acoustics. Springer, New York, pp. 821–841.
Fraley, C., Raftery, A.E., 2002. Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. 97, 611–631.
Fuller, S., Axel, A.C., Tucker, D., Gage, S.H., 2015. Connecting soundscape to landscape: Which acoustic index best describes landscape configuration? Ecol. Ind. 58, 207–215.
Ganchev, T.D., Jahn, O., Marques, M.I., de Figueiredo, J.M., Schuchmann, K.-L., 2015. Automated acoustic detection of Vanellus chilensis lampronotus. Expert Syst. Appl. 42, 6098–6111.
Gasc, A., Pavoine, S., Lellouch, L., Grandcolas, P., Sueur, J., 2015. Acoustic indices for biodiversity assessments: analyses of bias based on simulated bird assemblages and recommendations for field surveys. Biol. Conserv. 191, 306–312.
Gerhardt, H.C., 1994. The evolution of vocalization in frogs and toads. Annu. Rev. Ecol. Syst. 25, 293–324.
Gerhardt, H.C., Huber, F., 2002. Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions. University of Chicago Press, Chicago.
Harris, S.A., Shears, N.T., Radford, C.A., 2016. Ecoacoustic indices as proxies for biodiversity on temperate reefs. Methods Ecol. Evol. 7, 713–724.
Heinicke, S., Kalan, A.K., Wagner, O.J.J., Mundry, R., Lukashevich, H., Kühl, H.S., 2015. Assessing the performance of a semi-automated acoustic monitoring system for primates. Methods Ecol. Evol. 6, 753–763.
Hubert, L., Arabie, P., 1985. Comparing partitions. J. Classification 2, 193–218.
Kendrick, P., Lopez, L., Waddington, D., Young, R. (2016). Assessing the robustness of soundscape complexity indices. International Congress on Sound & Vibration (ICSV).
Kershenbaum, A., Blumstein, D.T., Roch, M.A., Akçay, Ç., Backus, G., Bee, M.A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S.L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T.M., Garland, E.C., Gustison, M., Harley, H.E., Huetz, C., Hughes, M., Hyland Bruno, J., Ilany, A., Jin, D.Z., Johnson, M., Ju, C., Karnowski, J., Lohr, B., Manser, M.B., McCowan, B., Mercado, E., Narins, P.M., Piel, A., Rice, M., Salmi, R., Sasahara, K., Sayigh, L., Shiu, Y., Taylor, C., Vallejo, E.E., Waller, S., Zamora-Gutierrez, V., 2016. Acoustic sequences in non-human animals: a tutorial review and prospectus: acoustic sequences in animals. Biol. Rev. 91, 13–52.
Krause, B.L., 1993. The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats. Soundscape Newslett. 6, 4–6.
Krause, B., Farina, A., 2016. Using ecoacoustic methods to survey the impacts of climate change on biodiversity. Biol. Conserv. 195, 245–254.
Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., Sueur, J., 2014. Monitoring temporal change of bird communities with dissimilarity acoustic indices. Methods Ecol. Evol. 5, 495–505.
Libbrecht, M.W., Noble, W.S., 2015. Machine learning applications in genetics and genomics. Nat. Rev. Genet. 16, 321–332.
Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell Pub, Malden, Ma.
Mallat, S.G., 2009. A Wavelet Tour of Signal Processing: The Sparse Way, third ed. Elsevier/Academic Press, Amsterdam Boston.
Parks, S.E., Miksis-Olds, J.L., Denes, S.L., 2014. Assessing marine ecosystem acoustic diversity across ocean basins. Ecol. Inf. 21, 81–88.
Pekin, B.K., Jung, J., Villanueva-Rivera, L.J., Pijanowski, B.C., Ahumada, J.A., 2012. Modeling acoustic diversity using soundscape recordings and LIDAR-derived metrics of vertical forest structure in a neotropical rainforest. Landscape Ecol. 27, 1513–1522.
Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. Science 339, 277–278.
Petrou, Z.I., Manakos, I., Stathaki, T., 2015. Remote sensing for biodiversity monitoring: a review of methods for biodiversity indicator extraction and assessment of progress towards international targets. Biodivers. Conserv. 24, 2333–2363.
Pimm, S.L., Russell, G.J., Gittleman, J.L., Brooks, T.M., 1995. The future of biodiversity. Science 269, 347–350.
Potamitis, I., 2014. Automatic classification of a taxon-rich community recorded in the wild. PLoS One 9, e96936.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Riede, K., 1993. Monitoring biodiversity: analysis of amazonian rainforest sounds. Ambio 22, 546–548.
Rodriguez, A., Gasc, A., Pavoine, S., Grandcolas, P., Gaucher, P., Sueur, J., 2014. Temporal and spatial variability of animal sound within a neotropical forest. Ecol. Inf. 21, 133–143.
Ruiz-Muñoz, J.F., Castellanos-Dominguez, G., Orozco-Alzate, M., 2016. Enhancing the dissimilarity-based classification of birdsong recordings. Ecol. Inf. 33, 75–84.
Sarthou, C., 2001. Plant Communities on a Granitic Outcrop. Nouragues: Dynamics and Plant-animal Interactions in a Neotropical Rainforest. Kluwer Academic Publishers, Dordrecht Boston, pp. 64–77.
Schmidt, A.K.D., Römer, H., Riede, K., 2013. Spectral niche segregation and community organization in a tropical cricket assemblage. Behav. Ecol. 24, 470–480.
Schwarz, G., 1978. Estimating the dimension of a model. Annal. Stat. 6, 461–464.
Sifre, L., Mallat, S., 2013. Rotation, scaling and deformation invariant scattering for texture discrimination. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 1233–1240.
Stowell, D., Plumbley, M.D., 2014. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. PeerJ 2, e488.
Sueur, J., 2002. Cicada acoustic communication: potential sound partitioning in a multispecies community from Mexico (Hemiptera: Cicadomorpha: Cicadidae). Biol. J. Linn. Soc. 75, 379–394.
Sueur, J., Farina, A., 2015. Ecoacoustics: the ecological investigation and interpretation of environmental sound. Biosemiotics 8, 493–502.
Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acust. United Acust. 100, 772–781.
Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Townsend Peterson, A., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145–148.
Towsey, M., Parsons, S., Sueur, J., 2014. Ecology and acoustics at a large scale. Ecol. Inf. 21, 1–3.
Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley Pub, Co, Reading, Mass.
Ulloa, J.S., Gasc, A., Gaucher, P., Aubin, T., Réjou-Méchain, M., Sueur, J., 2016. Screening large audio datasets to determine the time and space distribution of

Screaming Piha birds in a tropical forest. Ecol. Inf. 31, 91–99.
Villanueva-Rivera, L.J., 2014. Eleutherodactylus frogs show frequency but no temporal partitioning: implications for the acoustic niche hypothesis. PeerJ 2, e496.
Way, M.J. (Ed.), 2012. Advances in Machine Learning and Data Mining for Astronomy. Chapman and Hall/CRC, Boca Raton, Fla.
Webb, A.R., Copsey, K.D., 2011. Statistical Pattern Recognition, third ed. Wiley, Hoboken.
Xie, J., Towsey, M., Zhang, J., Roe, P., 2016. Adaptive frequency scaled wavelet packet decomposition for frog call classification. Ecol. Inf. 32, 134–144.
Yu, G., Mallat, S., Bacry, E., 2008. Audio denoising by time-frequency block thresholding. IEEE Trans. Signal Process. 56, 1830–1839.