Temporal evolution of the Mediterranean fin whale songs

Paul Best[1,*], **Ricard Marxer**[1], **Sébastien Paris**[1], **and Hervé Glotin**[1,2]

[1]Université de Toulon, Aix Marseille Univ, CNRS, LIS, DYNI, Marseille, France
[2]Pôle INPS
[*]paul.best@univ-tln.fr

ABSTRACT

In this paper, we present the analysis of fin whale (*Balaenoptera Physalus*) songs on passive acoustic recordings from the Pelagos Sanctuary (Western Mediterranean Basin). The recordings were gathered from 2008 to 2018 with 2 different antennas. On the task of fin whale 20Hz pulse detection, we first show how low complexity convolutional neural networks (CNN) help to cope with data diversity, and compete against much larger architectures. With the post analysis of the automated detection, 20Hz pulses were classified into two types and inter pulse interval (IPI) were measured. The results confirm previous observations on the relationship between pulse type and IPI in Mediterranean fin whale songs, and extends it to larger quantitative and temporal scales. From the latter emerge insights on inter-annual shifts of stereotypical IPI as well as intra-annual trends of centroid frequencies.

1 Introduction

The fin whale (*Balaenoptera Physalus*) is commonly found in the western basin of the Mediterranean sea, with an estimated population of approximately 3500 individuals[1]. As cetaceans, they are highly vocal animals, making the most out of the favorable underwater sound propagation (especially compared to light propagation). Their vocalizations supposedly serve group cohesion[2,3], food signaling[4], and mate attraction[5,6]. They all are very low frequency sounds, some around 20Hz, barely noticeable to the human hear.

This study focuses solely on the sequenced 20Hz pulses of the fin whales. The high predictability of those sequences, as well as their potential reproductive function, makes us refer to it as 'songs'. The term song has been commonly used in the scientific community to describe bird, primate, or mysticete vocalizations, when they fill out those two latter conditions. The function of the song in the animal kingdom is commonly described as a means of territorial defence and/or mate attraction / selection, it being a marker of health (sound amplitude), body size (pitch), and cognitive skills (ability to learn and reproduce sequences). Nonetheless, its evolutionary function(s) is still a matter of debates and yet to be proven rigorously[7].

The study of the function of songs in Atlantic and Pacific fin whales are most likely applicable to the Mediterranean population. However, it is not the case for the study of the songs' structure. Indeed, alike other cetacean species, fin whales show geographical acoustic differentiation[8–11], hypothesised to be cultural in some cases. The cultural hypothesis holds if a phenomenon is shown to be learned and taught by peers, not genetically determined, and not triggered solely by environmental factors[12]. Moreover, cultural behaviours are community specific, and cannot be found in a whole specie. The divergence of mysticetes songs in different populations are presumably a result of drifts emerging from the conformity and creativity constraints of song production[13]. Moreover, the character displacement theory with songs serving as a discrimination marker for allopatric populations has been hypothised for fin whales of Northern Atlantic by Delarue et al.[14]. As for the Mediterranean population, it has been shown to be resident and genetically dissociated with the North Atlantic population[15], and their song, especially the IPI, were shown to be a relevant discrimination metric[11,16]. The Mediterranean fin whales do not follow strict migration patterns or reproduction periods alike their oceanic conspecifics[1], so their song can be heard all year round.

The base unit of the songs, the 20hz pulse, is shared by all fin whales. The main differentiation of songs lies in the IPI (sometimes called INI for inter note interval) and pulse spectrums[17,18]. Alike for some other fin whale populations (e.g. in Northern Pacific[8,9]), Mediterranean 20Hz pulses fall into 2 two distinct types, one with a slightly higher pitch than the other[19,20] (see Fig. 1). These two categories are sometimes labelled as 20Hz pulse and back-beat, we will refer to them as type A and B for short, with A being the higher pitched pulse. Songs that consist of a regular alternation between type A and type B, called doublet patterns (as opposed to singlets where only one type occurs), are commonly found in the Pacific and Atlantic oceans. In these songs, there is a strong relationship between IPI and pulse type : two characteristic IPIs are found, one from A to B, and another from B to A[9,10,21–23] (singlet songs also have their own stereotypical IPI). As for the Mediterranean population, the

songs don't seem to follow the singlet or doublet patterns as strictly (see Fig.1). Nonetheless, two studies present stereotypical IPIs. Based on recordings from 1999, Clark et al.[19] observe a link between pulse type and IPI in the Mediterranean sea for two pulse sequences (about 100 pulses). About ten year later, Castellote et al.[11] observe a common IPI around 14.9, but do not mention its relationship with pulse types.

This paper intends to extend the present knowledge on the Mediterranean fin whale song structure, following the Passive Acoustic Monitoring (PAM) approach. Those song patterns being population specific[24], understanding them allows for population dynamics monitoring and stock structure assessment via the non invasive and cost efficient PAM method. The following analysis makes use of recordings coming from 2 different PAM stations : Boussole, and Bombyx, that recorded between 2008 and 2018. Employing a combination of manual, machine learning, and signal processing methods, a database of fin whale pulses was built, referencing time positions, centroid frequencies and pulse types. For the analysis of such large databases, automatic methods help reduce human effort. Moreover, machine learning algorithms such as neural networks infer discriminating rules that can cope with noise diversity and low signal to noise ratio (SNR), with a stochastic approach. Here, we make use of convolutionnal neural networks, which learn relevant filters to screen images (here spectrograms). The filters are optimized for a given task, in our case the binary classification of spectrograms (whether it contains a fin whale pulse or not). The convolutionnal approach allows time independent feature detection. Such machine learning algorithms highly depend on the data they are trained on. Regularization methods are put in place to enforce that models do not just find discrimination filters that best fit the training data, but rather find a solution robust to new data (e.g. data from new recording systems, or data that includes new kinds of noise). We call this the generalization capabilities of a model.

Following the latter detection mechanisms, signal processing and statistical analysis served the inference of song structure, and its evolution through time. Both inter-annual and intra-annual trends were analysed regarding IPI and centroid frequencies, for comparison with other populations of fin whales as well as other mysticete species. Indeed, the end goal of this study is to expand the current knowledge on song structures and their evolutionary trends, in hope for insights on their cause which remains unknown.

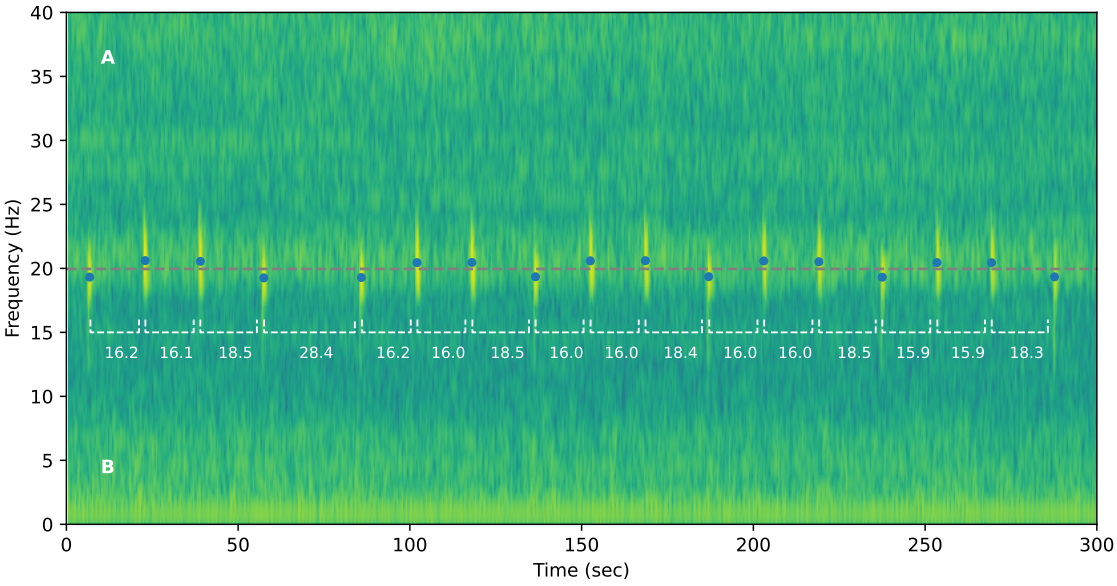

Figure 1. Spectrogram of a fin whale pulse sequence recorded by the Bombyx buoy in October 2018. Spectrogram parameters are described in section 2.7. Dots show the centroid frequencies of the detected pulses, with white dashed lines showing the IPIs. The grey dashed line denotes the discrimination threshold between type A and B pulses, at 19.9Hz.

Data source	Magnaghi[25]	Boussole[26]	Bombyx[27]	Total
Location	Tyrrhenian Sea	South of Sanremo	Port-Cros Island	Tyrrhenian Sea
Depth (m)	1	10-25	25	1-25
Recording year	1999	2008-2009	2015-2018	1999-2018
Sampling rate (kHz)	6	32	50	
ON/OFF protocole (min)	continuous	5/10	1/5 until Oct. 17, then 5/15	
Recorded time (hours)	0.75	1,752	3,533	5,286
Positive annotations	78	430	282	790
Negative annotations	396	4,098	292	4,786
Detection threshold		0.15	0.68	
Detected pulses		1,418	2,272	3,690
Detected A pulses		1,182	1,980	3,162
Detected B pulses		292	236	523
Detected sequences		214	530	744
Detected bouts		43	203	246

Table 1. Summary of the recording characteristics for each source. The data from Magnaghi was only used in the CNN training, not in the post analysis.

2 Material and Methods

2.1 Recorders

Recording characteristics for the 3 sources used in this study are summarized in Tab.1. Detailed recording dates can be visualised in Fig.2. The diversity of the data gathered poses great challenges for the automated analysis but also offers relatively robust performance measures, especially on the generalization capabilities of the machine learning models. The different recording stations being in the same region (smaller than the range fin whales travel[28]), we consider the observed individuals to belong to the same population.

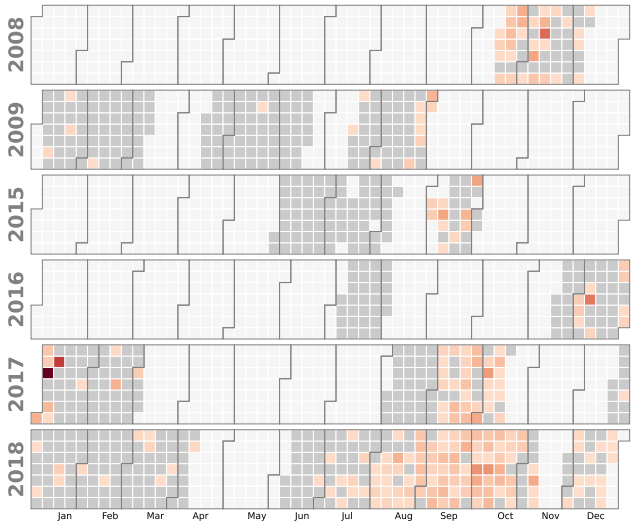

Figure 2. Calendar of the recorded days (grey cells). Shades of red denote the number of detected sequences normalized by the number of recorded hours (ranging from 0 to 8).

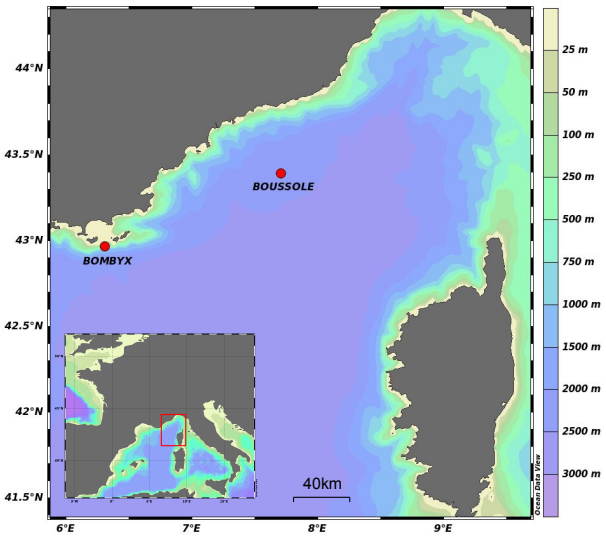

Figure 3. Map showing the two recording stations used in the analysis. This map was made using Ocean Data View[29].

2.2 Database constitution

An iterative annotation process was conducted to gather a database of fin whale pulses to train machine learning models on. The objective task is the detection of the 20Hz pulses in the signal, or in other words, the binary classification of the signal between 20Hz pulses and any other sounds. Starting from a single annotated song (see Magnaghi data in Tab.1), models (see

section 2.4) were trained and forwarded on unlabeled data. Positive and negative predictions were randomly sampled, manually annotated, and added to the training set for the next iteration. For positive examples picked for annotation, when possible, all other pulses of the surrounding bout were annotated. Like so, we avoid 'iterative over-fitting' (the model specializing to detect one specific type of pulse despite iterations to increase the database size). The resulting database is described in Tab.1, with the number of annotated positive and negative samples for each data source. To cope with the imbalance of the two classes, the positive examples were over-sampled by a factor 4 during training.

2.3 Data Preprocessing

We describe in the following section each step from loading the sound waveform to the actual forward pass of the CNN (see section 2.4). We first select a 5 seconds window surrounding the annotation and downsample it to 200Hz (using the Fourier method). The fin whale pulse lying around 20Hz, a nyquist frequency of 100Hz is enough to describe the pulse as a whole. Besides, the spectrum until 100Hz can contain information that helps to discriminate negative annotations. Including higher frequencies would be more costly in computation and might induce the network to overfit on features that are not related to the fin whale pulse. The waveform is then standardized (subtracting the mean and dividing by the standard deviation), making the network's input relatively stable in sound exposure level (SEL).

To enforce generalization and better low SNR performances, we add brown noise to the input signals at train time. Brown noise was chosen for its similarity to the sea's ambient noise. We standardize the generated brown noise and multiply it by 1.41 before adding it to the input signal. This means, considering the SEL as the variance of the signal, that the resulting SNR is -3dB ($10\log_{10}(\frac{1}{1.41^2}) \approx -3$).

We then compute the short time fourier transform (STFT) of the waveform using a window size and hop size of 256 and 32 points respectively. We apply a mel transform to the spectrogram resulting in 128 logarithmically spaced frequency bins ranging from 0 to 100Hz. This transformation is useful for the detection of animal vocalization in higher frequencies as it yields relevant representations of the spectrum. For this range of frequencies, it has a negligible impact.

We eventually apply $\log_{10}(1+x*10^a)$ with a being a trainable parameter of the model (inspired from Grill and Schlüter[30]). The resulting 128x24 image is then batched and given as an input to the model.

2.4 CNN based pulse detection

We designed a relatively low complexity CNN architecture (from 6 to 40 thousand parameters depending on the number of features and the kernel size) using 3 depth-wise convolution layers[31]. Depth-wise convolution layers were chosen instead of regular convolutions since they reduce the architecture's complexity (in terms of number of trainable parameters and multiplications needed for a forward pass), thus enforcing a better generalisation and allowing the model to be embedded in a low power micro-controller[32]. The assumption for the use of this type of layer is that filters of the input features can be independent from their combination to the output features. Taking for example 64 input and 128 output features with a kernel of size 5, a regular convolution layer would need $64*5*128 = 40,960$ parameters. A depth-wise convolution layer on the other hand, would consist of $64*5+64*128 = 8,512$ parameters. The number of multiplications for a forward pass are proportional to the latter (the number of parameters multiplied by the number of convolution strides).

The convolution occurs on the time dimension only, as the frequency bins are considered as input features for the model. We do not convolve on the frequency dimension since frequency invariant features are not to be expected in fin whale vocalisations (this does not impeach the model from learning several pulse types laying at different pitches).

The two first layers are followed by batch normalization, leaky linear rectifier unit, and dropout ($p = 0.25$). The model ends with a maximum pooling layer, and is trained as a binary classifier using a binary cross entropy loss. The model is trained for 50 epochs with a batch size of 16, a learning rate of 0.001, which decays by 3% at each epoch, an Adam optimizer[33], and a weight decay L2 loss of 0.04. We studied the effect of varying number of features per layer and kernel sizes (see Supplementary Fig.11). To measure at best the generalization performance of the model for hyper-parameter selection, we used two sources for training and the third for testing, in a cross-fold manner. Maximum performance and stability across folds was achieved using 128 features per layer and kernels of size 5. Fig.4 shows the receiving operating characteristics (ROC) curves for each fold of the latter architecture. The area under the ROC curves (AUC) are 0.992, 0.943, and 0.997 for Bombyx Magnaghi and Boussole test sets respectively. The generalisation capabilities are further demonstrated on a totally different dataset in Section 2.5.2. The performance of the model (as described in 2.6) facing added brown noise are shown in Fig.13.

2.5 Baseline comparison for pulse detection

2.5.1 Comparison on the same dataset with a different method

We chose template matching (also termed as matched filter) as a baseline method to compare the performance of our detection model against, as it is a common approach to the mysticete sound event detection task[8,34,35]. Our template is the average of all the annotated pulses in the training set in the Fourier domain. Pixels below 10 of power spectral density (PSD) were set to 0. We then threshold on the cross-correlation product of samples with the template. The resulting detection performances are

presented in Fig.4. The AUC of the template matching method is 0.898 (5 to 10 points less than the CNN model, depending on the fold).

2.5.2 Comparison on a different dataset with a similar method

We report the results of our model on the dataset published by a study of CNN based fin whale 20Hz pulse detection[36]. We ran our model (described in section 2.6), trained on our data only, on the data published by the latter study. The resulting AUC and peak F1-score are 0.93 and 0.88, when their reported best overall performances are 0.95 and 0.91. In comparison, our model has 33% less parameters than the base CNN model, and the training framework is considerably simpler.

2.6 Detection preceding the song analysis

The model used for the following analysis was trained on all annotations with 128 features per layer and kernels of size 5. For the detection threshold, we took the thresholds at which the true negative rate (TNR) equals the true positive rate (TPR), for each data source separately (see Tab.1). Those thresholds give TPR / TNR values of 0.96 and 0.97 for the Bombyx and Boussole data respectively. The Magnaghi data was not included in the analysis since multiple fin whale songs are overlapping in the available segments, and the following method does not cope with this.

Discarding the max pooling layer at the end of the CNN enables us to have a sequence of predictions as a function of time. We retained as pulse time the highest predictions (peak) in sliding 4 second windows, when above the given detection threshold. The center of the receptive field of the network (0.8 seconds wide) at the given prediction peak gives us an approximate time position for the pulse. Pulses at a distance of less than 45 seconds were considered as being part of the same sequence. Sequences are considered as being part of the same bout if separated by less than 2 hours (following Watkins et al.[5]).

2.7 Pulse analysis

Following the detection process, the pulse analysis allowed us to extract a more detailed description of each pulse, such as the exact time position, the centroid frequency, the bandwidth, and the SNR. We start by selecting a 8 second window surrounding the prediction peak ($T = [0,8]$). We then apply a bandpass butterworth filter of order 3 between 10 and 30Hz, and resample at 250Hz. The STFT is then computed with a window size of 1024 points (including 75% of 0 padding), and 97% of overlap (yielding spectral and temporal resolutions of 0.24Hz and 0.03sec respectively). The resulting spectrogram is a matrix $\mathbf{S}_{i,j}$ with the following relation between indices and frequency / time values : $f = i\frac{125}{512}$, $t = \frac{8j+128}{250}$.

The position of the pulse $\hat{t}$ is estimated as the column of the maximum value in the 18-22Hz frequency band (Eq.1). The envelope $E(f)$ of the pulse (Eq.2) is then computed with a maximum pooling of 1.2sec surrounding $\hat{t}$, and subtracting an estimate of the background spectrum via the median of each frequency bin (the median being less sensitive the fin whale pulses as using a mean would). This envelope is used to compute the left and right boundaries of the pulse spectrum, with $(\max_f E(f))/4$ as a threshold (equivalent to -6 dB). Left and right intersection frequencies are linearly interpolated. The bandwidth and centroid frequencies of the pulse are given by the difference and the mean of the left / right boundaries respectively. Eventually, an estimate of the SNR is given by Eq.3, with the median used again to avoid including the pulse's energy into the background energy. Peak frequency or spectrum weighted mean are often used to estimate a pulse's pitch[8,37]. We rather chose the centroid frequency as it appeared to be a better discriminating metric for the separation of the two pulse types. Indeed, the Kullbak-Leibler divergence (KL) between A and B centroid frequency distributions is significantly higher than for peak frequency distributions (113 and 30 respectively).

$$\hat{t} = \underset{t\in T}{\operatorname{argmax}} \max_{f\in[18,22]} \mathbf{S}_{i,j} \tag{1}$$

$$E(f) = \max_{t\in[\hat{t}-0.6,\hat{t}+0.6]} \mathbf{S}_{i,j} \; - \; \underset{t\in T}{\operatorname{median}}\, \mathbf{S}_{i,j} \tag{2}$$

$$E_{Background} = \underset{\substack{f\in[15,25] \\ t<\hat{t}-1 \,\cup\, t>\hat{t}+3}}{\operatorname{median}} \mathbf{S}_{i,j}, \qquad E_{Pulse} = \max_f E(f), \qquad SNR = 10\log_{10}\left(\frac{E_{Pulse}}{E_{Background}}\right) \tag{3}$$

2.8 Pre-analysis filtering

To filter out false positives, pulses with a bandwidth higher than 6Hz, or with a centroid frequency outside the $[18.5, 22.5]$ interval were withdrawed. Besides, only sequences with a mean SNR of at least 8dB, and with at least 3 pulses were kept for the following analysis. Sequences containing IPIs below 11sec or above 45sec were discarded as well. Number of registered pulses and sequences are shown in a calendar Fig.2 and in Tab.1.

To classify between A and B types, a two component gaussian mixture model (GMM) was fitted on the centroid data (see Fig.5) using the expectation-maximization (EM) algorithm. This lead to a centroid threshold of 19.96Hz to discriminate between the two types (see Fig. 1). We note that the pulse centroid frequency evolving through time (discussed in section 3.2), a better discrimination threshold would have to be fitted for each period. However, we chose data quantity over data quality for a better estimate.

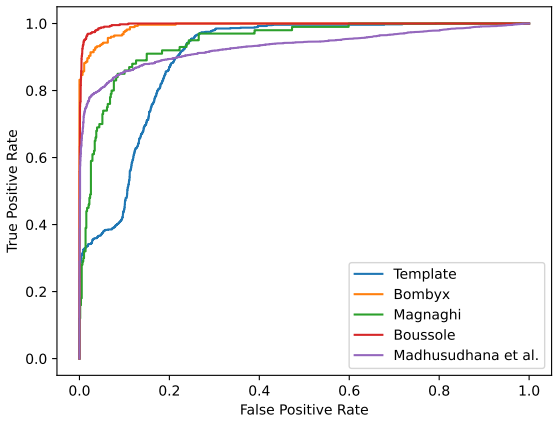

Figure 4. ROC curves for each test set (the two remaining sources serving as training set) and for the template matching method. The ROC curve of the model over the dataset published in[36] is also displayed.

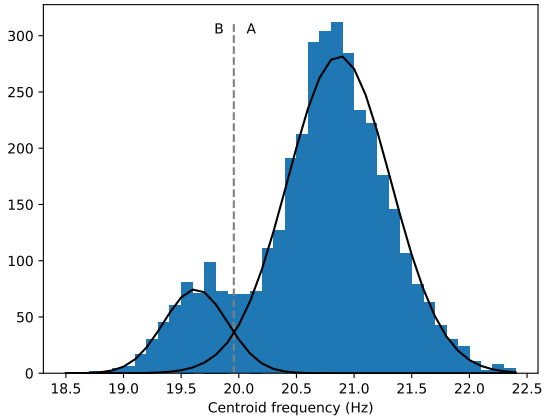

Figure 5. Histogram of the centroid frequencies of the filtered detected pulses. Black lines denote the fitted GMM.

3 Results

The following results are taken from a database of 744 sequences with 3690 pulses in total (see Tab.1). This database is available online at `http://sabiod.lis-lab.fr/pub/fin_whale_songs/`.

3.1 Stereotypical IPI

The IPI appears to be strongly determined by the type sequence (see Fig. 6). The typical interval for a 'AB' sequence is 2sec longer than for the 'AA' or 'BA' sequences. On the other hand, the 'BB' sequences (less frequent but still present) are 11sec longer on average, but present less stability than the others.

We show how those distributions evolve with respect to time in Fig.7, with a similar approach than conducted by Weirathmueller et al.[8]. Points denote the most frequent IPI for a given type sequence and a quarter of the year (IPIs were quantized with a resolution of 0.1sec, dates were quantized with a resolution of 3 months). Only bins with a frequency of at least 5% among a group of at least 100 pulse sequences were kept. We also added the points measured in 1999 by Clark et al.[19] (the only study to our knowledge that references IPI depending on type sequence in the Mediterranean sea), and a point measured in 2008 by Castellote et al.[11] (assuming it describes the 'AA' sequence by default, as it is not specified in the article). The 'BB' sequence did not provide enough occurrences for the statistical tests to be relevant. For sequences 'AA', 'AB', and 'BA', we plot fitted linear models, whose coefficients of determination are 0.86, 0.97, and 0.92 respectively. The p-value for the null-hypothesis that the slope of the latters is not significantly different from 0 are all inferior to 0.001. The estimated slopes for the 'AA', 'AB', and 'BA' type sequences are 0.09, 0.10, and 0.09 respectively (in seconds/year).

3.2 Centroid frequency

For the study of the evolution of pulse pitch, we quantized the centroid frequencies with a resolution of 0.1Hz. Contrarily to the stereotypical IPIs, the centroid frequencies did not show any inter-annual evolution, but rather an intra-annual evolution (see supplementary Fig. 12). We show how this distribution evolves with respect to months of the year in Fig. 8, and try to model it linearly. Only bins with a frequency of at least 10% among a group (column) of at least 50 pulses were kept as peak points. The coefficients of determination for the model is 0.74, with an estimated slope of -0.09 (in Hz per month). The p-values for the null-hypothesis that the slope is significantly different from 0 is 0.03.

We show the distribution of type B pulses with respect to months of the year in Fig. 8, below the dashed line. There is not enough data to draw an analysis alike the one conducted for the type A pulses.

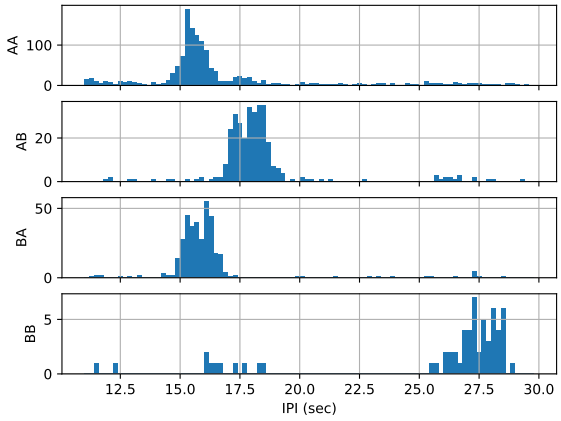

Figure 6. Histogram of the IPI for each type sequence.

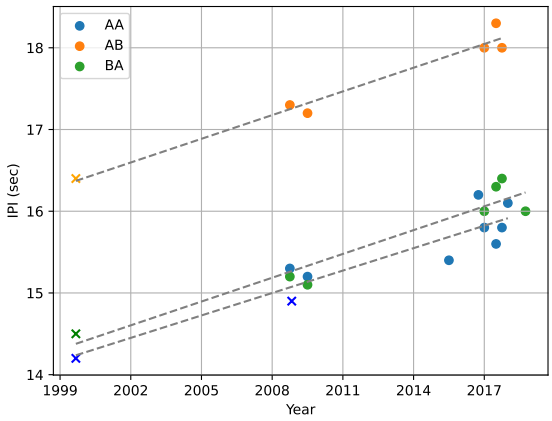

Figure 7. Scatter plot of the most frequent IPI per year for each type sequence. Fitted linear models are shown as grey dashed lines. Points extracted from Clark et al.[19] and Castellote et al.[11] are denoted by crosses.

3.3 Correlation between centroid frequency and IPI

With the observation of synchronous intera-annual shifts of both IPI and centroid frequencies (in Pacific fin whales), the hypothesis of a link between the two arises. Weirathmueller et al.[8] states that the augmentation of the IPI through the years could be explained by the simultaneous decrease in pulse centroid frequencies (lower frequency pulses presumably requiring a bigger effort to produce, a bigger gap between them could be needed). The observed stereotypical IPIs of Mediterranean fin whales also support this idea (‘AA’ showing the shortest IPI in average). We thus test further this hypothesis in this section, analysing the correlation between IPI and the centroid frequency (for pulses with IPIs between 14 and 20 seconds).

To dissociate this analysis from the link between pulse types and IPI (see section 3.1), we fitted a 3 component gaussian mixture model on the bi-dimensional representation of pulses (centroid frequency versus time until the next pulse). Like so, we were able to group pulse sequences (’AA’, ’AB’, and ’BA’) together, and conduct a correlation analysis on each group independently. Fig.9 shows the scatter plot of the pulses, and their assignation to each mixture component. For each of the latter, we computed the pearson correlation coefficient, which output -0.37, -0.22, and -0.35 for ‘BA’, ‘AB’, and ‘AA’ respectively (all p-values are below 0.001).

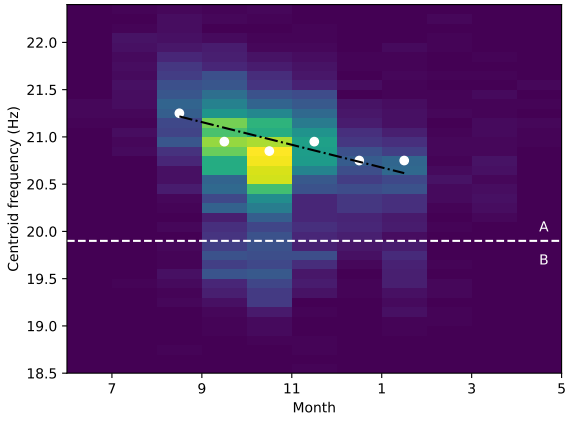

Figure 8. Bi-histogram of the centroid frequencies against months of the year. The horizontal line shows the separation between type A and type B pulses. The fitted linear model is shown as a black dashed line.

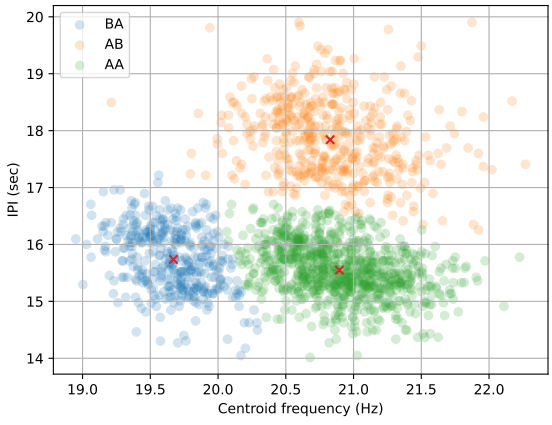

Figure 9. Scatter plot of pulses centroid frequency against the time until the next pulse (IPI). Colors denote the GMM assignation, whose means are marked with crosses.

4 Discussion

Our study draws the structure of the Mediterranean fin whale songs with a statistical approach, extending the previous analysis conducted for this population[11,19]. The CNN used for 20Hz pulse detection showed robustness facing data diversity, performances comparable to the state of the art, with a relatively simple framework. Besides, such architectures are light enough to be embedded in low-power micro-processors for applications such as real time alert systems for collision risk mitigation[32]. The automatic approach led to the confirmation of the local stereotypical IPIs determined by the pulse type sequence. Those results were previously shown with 100 pulses[19], we confirm them with several thousands, spread over 10 years.

Making use of the data sprawl over a decade, we show how those stereotypical IPIs evolved, following a linear growth of approximately 0.1sec/year for 20 years. Such trends have been shown with fin whales of the North-East Pacific[8], with slopes between 0.5 and 0.9 sec/year, and with fin whales of the Central-North Pacific[9], with slopes between 0.6 and 1.3 sec/year. The Mediterranean fin whales thus take part in inter-annual IPI shifts, also observed in fin whales of the Pacific.

Inter-annual shifts in IPI are rather recent and poorly documented. Weirathmueller et al. states that the downward shift in frequency might be linked with the increase in IPI, lower frequency pulses potentially being more demanding in energy. We measured a relatively low correlation coefficient between the two variables, and our data did not show any evidence for an inter-annual frequency decrease. Those observations thus go against this hypothesis, but more data is still needed to firmly state that the phenomenon does not exist.

As for the IPI shift slopes, it seems plausible that the differences between Pacific and Mediterranean populations arise culturally. Whether they are originally caused by the same factors or not, the singing patterns drift independently, with song conformity only taking place within a given population. If environmental factors are the sole responsible for those patterns, they have to be present in the Pacific and in the Mediterranean sea, but operating at different rates.

Inter-annual shifts in vocalization frequencies are well documented with blue whales[37,38], and bowhead whales[39]. Fin whales also showed similar trends in the Pacific[8] (for 20Hz pulses, between -0.1 and -0.2Hz/year) and in the Indian Ocean[40] (for 99Hz pulses, -0.21Hz/year). Numerous hypothesis were formulated and tested on the cause of this phenomenon, such as the increase in population density or body sizes (following the cease of commercial whaling), the increase in calling depth[41], the augmentation of noise from melting icebergs[40], the acidification of the oceans affecting sound propagation[42] among others. So far, no explanation has been confirmed, a consensus is yet to be achieved.

No inter-annual frequency shift was found in our data, Mediterranean fin whales could thus be an exception to this widespread trend. Nonetheless, our data showed an intra-annual decrease in centroid frequencies (0.09Hz/month). Such phenomenon was previously observed in the Indian Ocean among large mysticetes including fin whales[40]. Besides, in other studies of the Atlantic[5,10] and Pacific Oceans[8,21], IPI increases were observed during the reproductive season (winter), resetting to the previous starting value in autumn. In a similar fashion than the IPI seems directly linked to the reproductive season in those populations[21] (hormonal activity, progressive dilution of the competition), pulse pitch could be the feature that Mediterranean fin whale modify every winter.

Mediterranean fin whale songs show both similarities and dissimilarities compared to other populations of the world, in terms of structure and temporal evolutionary trends. The populations share genomes, but differ in the environments they evolve in (potentially affecting their migratory patterns, or reproductive seasonality[1]). Gathering this kind of data for multiple species and populations is key to advance on yet unsolved questions (origin of inter/intra-annual trends of acoustic features), and perhaps one day point out whether the answer is culture, genetics, environmental factors or a combination of the three.

References

1. Notarbartolo-Di-Sciara, G., Zanardelli, M., Jahoda, M., Panigada, S. & Airoldi, S. The fin whale balaenoptera physalus (l. 1758) in the mediterranean sea. *Mammal Rev.* **33**, 105–150, DOI: https://doi.org/10.1046/j.1365-2907.2003.00005.x (2003).
2. Payne, R. & Webb, D. Orientation by means of long range acoustic signaling in baleen whales. *Annals New York Acad. Sci.* **188**, 110–141 (1971).
3. Watkins, W. A. Activities and underwater sounds of fin whales. *Sci. Rep. Whales Res. Inst* **33**, 83–117 (1981).
4. Romagosa, M. *et al.* Food talk: 40-hz fin whale calls are associated with prey biomass. *Proc. Royal Soc. B* **288**, 20211156 (2021).
5. Watkins, W. A., Tyack, P., Moore, K. E. & Bird, J. E. The 20-hz signals of finback whales (balaenoptera physalus). *The J. Acoust. Soc. Am.* **82**, 1901–1912, DOI: 10.1121/1.395685 (1987).
6. Croll, D. A. *et al.* Only male fin whales sing loud songs. *Nature* **417**, 809–809 (2002).

7. Fitch, W. T. The biology and evolution of music: A comparative perspective. *Cognition* **100**, 173–215 (2006).
8. Weirathmueller, M. J. *et al.* Spatial and temporal trends in fin whale vocalizations recorded in the ne pacific ocean between 2003-2013. *Plos one* **12**, e0186127 (2017).
9. Helble, T. A. *et al.* Fin whale song patterns shift over time in the central north pacific. *Front. Mar. Sci.* **7**, 907 (2020).
10. Morano, J. L. *et al.* Seasonal and geographical patterns of fin whale song in the western north atlantic ocean. *The J. Acoust. Soc. Am.* **132**, 1207–1212 (2012).
11. Castellote, M., Clark, C. W. & Lammers, M. O. Fin whale (balaenoptera physalus) population identity in the western mediterranean sea. *Mar. Mammal Sci.* **28**, 325–344 (2012).
12. Whitehead, H. & Rendell, L. *The cultural lives of whales and dolphins* (University of Chicago Press, 2014).
13. Payne, K. The progressively changing songs of humpback whales: a window on the creative process in a wild animal. *The origins music* 135–150 (2000).
14. Delarue, J., Todd, S. K., Van Parijs, S. M. & Di Iorio, L. Geographic variation in northwest atlantic fin whale (balaenoptera physalus) song: Implications for stock structure assessment. *The J. Acoust. Soc. Am.* **125**, 1774–1782 (2009).
15. Bérubé, M. *et al.* Population genetic structure of north atlantic, mediterranean sea and sea of cortez fin whales, balaenoptera physalus (linnaeus 1758): analysis of mitochondrial and nuclear loci. *Mol. ecology* **7**, 585–599 (1998).
16. Pereira, A., Harris, D., Tyack, P. & Matias, L. Fin whale acoustic presence and song characteristics in seas to the southwest of portugal. *The J. Acoust. Soc. Am.* **147**, 2235–2249 (2020).
17. Thompson, P. O., Findley, L. T. & Vidal, O. 20-hz pulses and other vocalizations of fin whales, b alaenopteraphysalus, in the gulf of california, mexico. *The J. Acoust. Soc. Am.* **92**, 3051–3057 (1992).
18. Hatch, L. & Clark, C. Acoustic differentiation between fin whales in both the north atlantic and north pacific oceans, and integration with genetic estimates of divergence. *Unpubl. paper to IWC Sci. Comm.* (2004).
19. Clark, C. W., Borsani, J. F. & Notarbartolo-Di-sciara, G. Vocal activity of fin whales, balaenoptera physalus, in the ligurian sea. *Mar. Mammal Sci.* **18**, 286–295, DOI: https://doi.org/10.1111/j.1748-7692.2002.tb01035.x (2002).
20. Sciacca, V. *et al.* Annual acoustic presence of fin whale (balaenoptera physalus) offshore eastern sicily, central mediterranean sea. *PloS one* **10**, e0141838 (2015).
21. Oleson, E. M., Širović, A., Bayless, A. R. & Hildebrand, J. A. Synchronous seasonal change in fin whale song in the north pacific. *PLoS ONE* **9**, e115678, DOI: 10.1371/journal.pone.0115678 (2014).
22. Constaratas, A. N., McDonald, M. A., Goetz, K. T. & Giorli, G. Fin whale acoustic populations present in new zealand waters: Description of song types, occurrence and seasonality using passive acoustic monitoring. *Plos one* **16**, e0253737 (2021).
23. Furumaki, S., Tsujii, K. & Mitani, Y. Fin whale (balaenoptera physalus) song pattern in the southern chukchi sea. *Polar Biol.* DOI: 10.1007/s00300-021-02855-y (2021).
24. Castellote, M., Clark, C. W. & Lammers, M. O. Acoustic and behavioural changes by fin whales (balaenoptera physalus) in response to shipping and airgun noise. *Biol. Conserv.* **147**, 115–122, DOI: 10.1016/j.biocon.2011.12.021 (2012).
25. Pavan, G., Fossati, C., Manghi, M. & Priano, M. Passive acoustics tools for the implementation of acoustic risk mitigation policies. *Eur. Cetacean Soc. Newsl. No* 52–58 (2004).
26. Laran, O., Castellote, M., Caudal, F., Monnin, A. & H., G. Suivi acoustique des cetaces au nord du sanctuaire pelagos. Tech. Rep., Pelagos (2009). https://www.sanctuaire-pelagos.org/fr/sensibilisation/bulletins-de-liaison-annuel/com_docman/doc_download/70/170/fr-FR/?task=doc_download&gid=70.
27. Glotin, H. *et al.* Projet VAMOS : Visées aeriennes de mammifères marins jointes aux obervations acoustiques sous-marines de la bouée BOMBYX et antares: nouveaux modèles en suivis et lois allométriques du *Physeter macrocephalus*, *Ziphius Cavirostris* et autres cétacés. Tech. Rep., Pelagos (2017). https://www.sanctuaire-pelagos.org/fr/tous-les-telechargements/etudes-scientifiques-studi-scientifici-studies/etudes-francaises/789-14-037-vamos.
28. Panigada, S. *et al.* Satellite tagging of mediterranean fin whales: working towards the identification of critical habitats and the focussing of mitigation measures. *Sci. reports* **7**, 1–12 (2017).
29. Schlitzer, R. & Mieruch, S. Ocean data view goes online. *Bollettino di Geofisica* **12**, 167 (2021).
30. Grill, T. & Schlüter, J. Two convolutional neural networks for bird detection in audio signals. In *2017 25th European Signal Processing Conference (EUSIPCO)*, 1764–1768 (IEEE, 2017).

31. Bai, L., Zhao, Y. & Huang, X. A cnn accelerator on fpga using depthwise separable convolution. *IEEE Transactions on Circuits Syst. II: Express Briefs* **65**, 1415–1419 (2018).

32. Best, P. *et al.* Stereo to five-channels bombyx sonobuoys: from four years cetacean monitoring to real-time whale-ship anti-collision system. In *e-Forum Acusticum 2020* (2020).

33. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

34. Mellinger, D. K. & Clark, C. W. Recognizing transient low-frequency whale sounds by spectrogram correlation. *The J. Acoust. Soc. Am.* **107**, 3518–3529 (2000).

35. Bouffaut, L., Dréo, R., Labat, V., Boudraa, A.-O. & Barruol, G. Passive stochastic matched filter for antarctic blue whale call detection. *The J. Acoust. Soc. Am.* **144**, 955–965 (2018).

36. Madhusudhana, S. *et al.* Improve automatic detection of animal call sequences with temporal context. *J. Royal Soc. Interface* **18**, 20210297 (2021).

37. Malige, F. *et al.* Inter-annual decrease in pulse rate and peak frequency of southeast pacific blue whale song types. *Sci. reports* **10**, 1–11 (2020).

38. McDonald, M. A., Hildebrand, J. A. & Mesnick, S. Worldwide decline in tonal frequencies of blue whale songs. *Endangered species research* **9**, 13–21 (2009).

39. Thode, A. M., Blackwell, S. B., Conrad, A. S., Kim, K. H. & Michael Macrander, A. Decadal-scale frequency shift of migrating bowhead whale calls in the shallow beaufort sea. *The J. Acoust. Soc. Am.* **142**, 1482–1502 (2017).

40. Leroy, E. C., Royer, J.-Y., Bonnel, J. & Samaran, F. Long-term and seasonal changes of large whale call frequency in the southern indian ocean. *J. Geophys. Res. Ocean.* **123**, 8568–8580 (2018).

41. Gavrilov, A. N., McCauley, R. D. & Gedamke, J. Steady inter and intra-annual decrease in the vocalization frequency of antarctic blue whales. *The J. Acoust. Soc. Am.* **131**, 4476–4480 (2012).

42. Hester, K. C., Peltzer, E. T., Kirkwood, W. J. & Brewer, P. G. Unanticipated consequences of ocean acidification: A noisier ocean at lower ph. *Geophys. research letters* **35** (2008).

43. Zipf, G. K. Human behaviour and the principle of least effort. (1950).

44. Kershenbaum, A. *et al.* Shannon entropy as a robust estimator of zipf's law in animal vocal communication repertoires. *Methods Ecol. Evol.* **12**, 553–564 (2021).

Acknowledgements

We first thank Gianni Pavan for providing the Magnaghi data, which gave us the first reference of Mediterranean fin whale pulses, necessary to start to learn the analysis of the rest of the data. We are grateful for the precious feedback provided by Franck Malige and Julie Patris. We would like to thank Osean SAS le Pradet, its director O. Philippe and F. Fayet who participated in the instrumentation of BOMBYX and provided their recorders. We thank the Parc national of Port-Cros, PMS SAS for their help in the logistics, and Prefecture Maritime de la Méditerranée. We thank G. Rougier for his help in the installation of BOMBYX.
This research is partly funded by:

- Institut Universitaire de France (H.G. Chair), TPM, CG83, University of Toulon, Pole INPS and LIS Dyni which financed construction and maintenance of BOMBYX for 4 years,
- MARITTIMO Intereg European FEDER GIAS project, and Engie Fondation, for co-funding,
- MI CNRS MASTODONS sabiod.org, for storage of the data and GPU computation,
- Region PACA and GIAS Marittimo cofund P.B. PhD,
- ANR-20-CHIA-0014-01 national Chair in Artificial Intelligence for Bioacoustics (ADSIL)(H.G), and ANR-18-CE40-0014 SMILES, for support in massive data algorithmic,
- FUI ABYSSOUND for support in acoustic processing,
- The Pelagos Sanctuary which granted the studies on BOUSSOLE and BOMBYX.

Author contributions statement

H.G. conceived and conducted the Bombyx and Boussole the experiments that yielded the acoustic data. R.M., S.P. and H.G. advised on the algorithms and analysis. P.B. conceived the algorithms, analysed the results, and wrote the manuscript. All authors reviewed the manuscript.

5 Additional information

Data accessibility

All the data used in 3 can be found at http://sabiod.lis-lab.fr/pub/fin_whale_songs/

Competing interests

The authors declare no competing interests.

6 Supplementary Material

6.1 Sequence analysis and Zipfs law

To our knowledge, most of the fin whale vocalization sequences are analysed as singlets or doublets[9,10,21]. Meaning a succession of the same pulse type, or an alternation between 2 pulse types. The Mediterranean fin whale songs apparently do not strictly follow those two patterns, but rather present a mixture of them. To analyse the sequences occurrence, with a system that is generic to their length, we thus use the number of consecutive A pulses (the type A pulse is chosen since it is largely more frequent than the type B pulse). A singlet sequence 'AAAAA' thus becomes '5', a doublet sequence 'ABABABAB' becomes '1111', and a mixture sequence 'BAABBAAABABA' becomes '20311'. Such a system gives us insights on potential patterns and tendencies on the fin whale songs. We plot the histogram of the occurrences of these number of consecutive pulses in Fig. 10, sorted from the most frequent to the least frequent.

Such a distribution seems appropriate for an estimation of the Zipf law power coefficient. Zipf law[43] is often used in language analysis, as it describes one common feature to all human languages. It is expressed by the following equation : $f \propto r^{-c}$, with f the frequency of a word, c the power law coefficient (PLC), and r the rank of the word (1 being the most frequent). The PLC describes how stereotyped the studied phenomenon is. When close to 0, the distribution is uniform, and each word has the same probability of occurrence. On the other hand, a high value of c means that a few words occur a lot when the others are rare. All human languages show a PLC of approximately 1, and any optimal communication canal would follow this characteristic (following the "principle of least effort").

Zipf's law has been used to characterize animal communication systems[44] on their potential language features. We thus fitted a linear model on the log frequencies against the log rank for our number of consecutive A pulses distribution (see Fig. 10). The found PLC is 0.6, which can be interpreted as "the number of consecutive A pulses is more uniformly distributed than an optimal communication canal"[43].

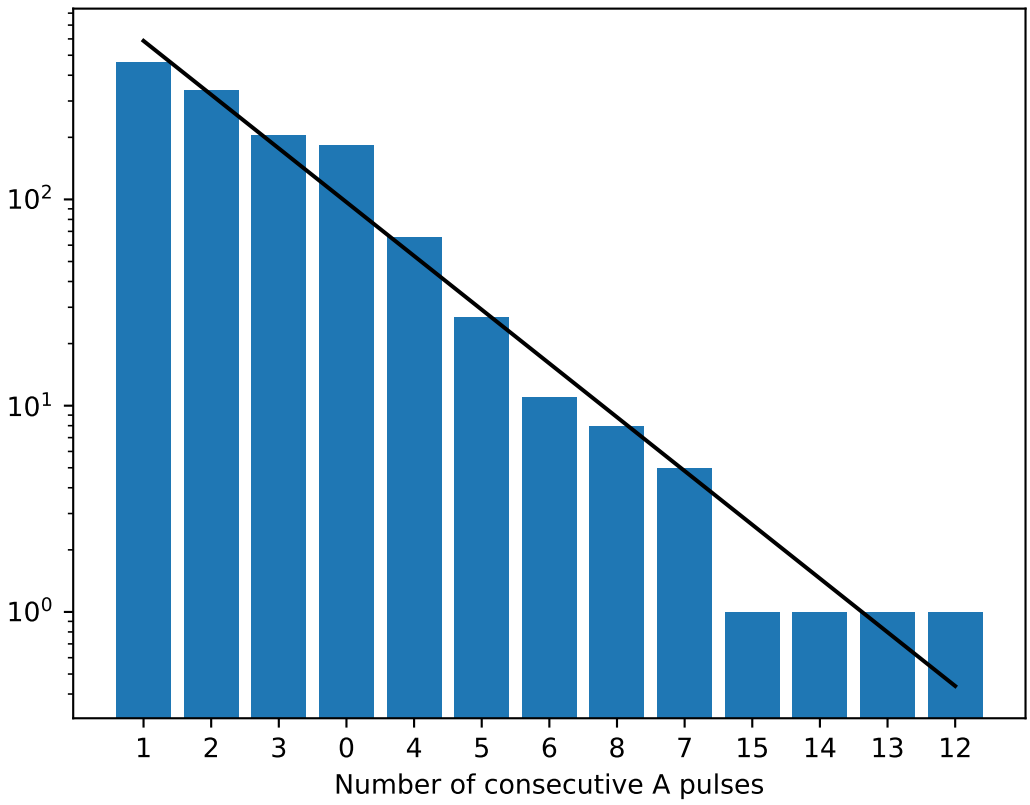

Figure 10. Distribution of the number of type A pulses in between type B pulse. The black line denotes the fitted slope for PLC estimation.

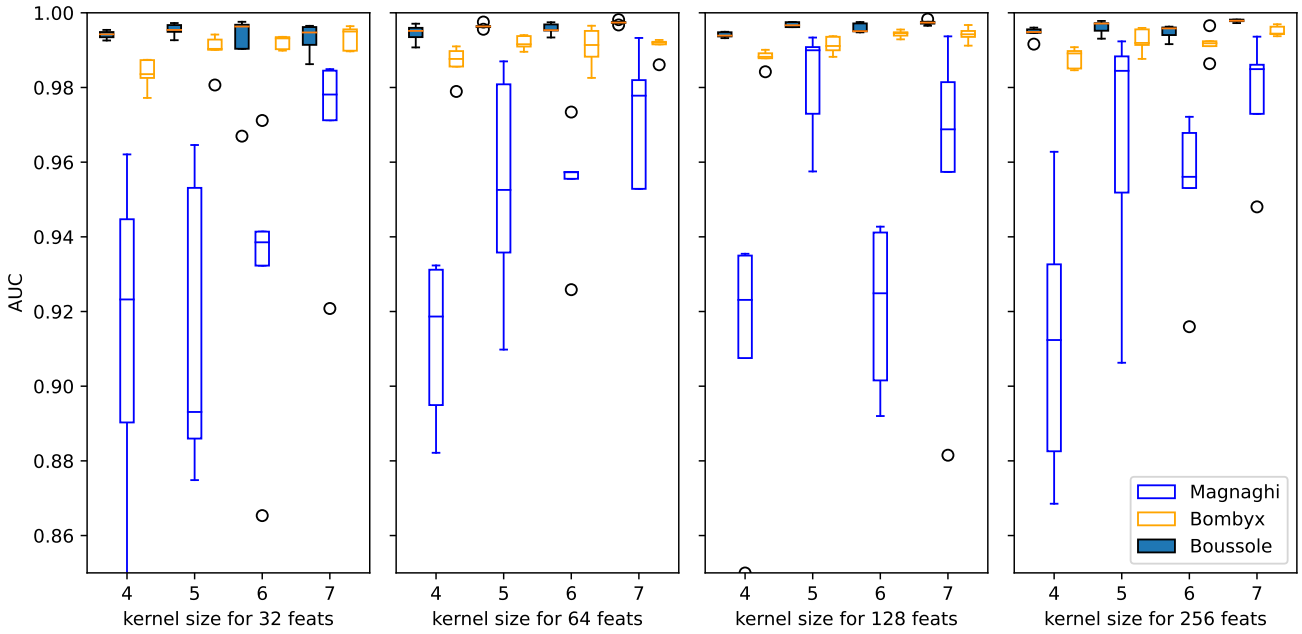

Figure 11. Boxplots of the AUC of several combination of hyper-parameters. For each number of feature per layer, kernel size, and train/test fold, 5 runs were conducted. Folds are labelled with their test set (meaning that Bombyx scores report the performance of models trained on Magnaghi and Boussole).

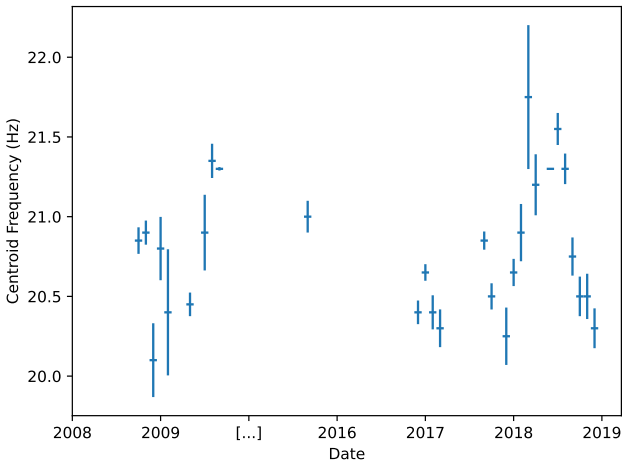

Figure 12. Most frequent centroid frequencies for each month of the dataset (horizontal bars). Vertical bars denote the mean of the square difference with the given most frequent centroid

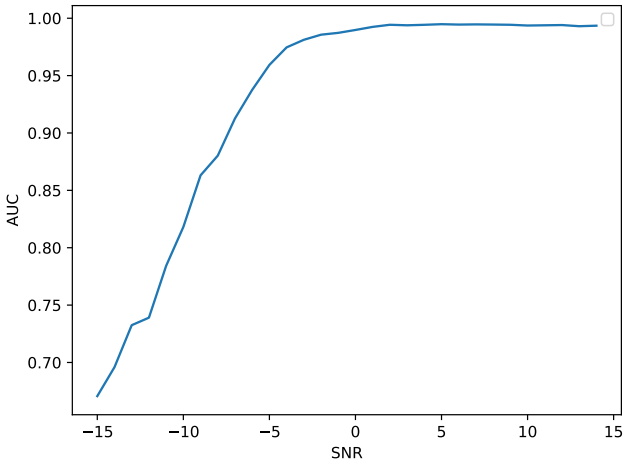

Figure 13. AUC of the model as a function of the added brown noise level (measured in SNR as defined in 2.3)

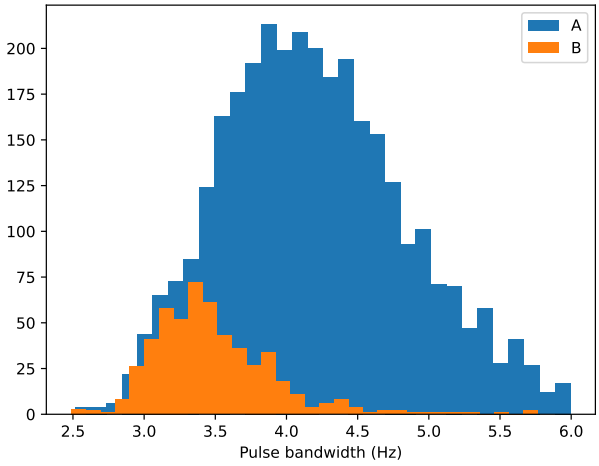

Figure 14. Bandwidth of the detected pulses (at the peak energy -6dB), following the method described in 2.7