A classical approach for cetacean classification

- Summarizing our efforts to develop a classical approach for cetacean classification with PAMGuard, using a dataset of striped dolphins (SD) and sperm whales (SW) audio recordings.
- Brainstorming on how to proceed, in order to compare our results with a deep learning cetacean classification model that uses the same dataset.

Introducing the basics of the deep learning model

.The model input is the spectrogram image of a single audio file.

.The dataset consists of 531 wav files (347 striped dolphins SD and 184 sperm whales SW).

.The time duration of each file is approximately 4s-6s.

.The model output is a label prediction for the audio file that was given as input.

.The prediction labels are two, the same as in the dataset, either SD or SW.

.The output is determined by the model parameters.

.We can’t know what each parameter expresses (the model is a “black box”).

.We can assume that the cetacean vocalizations (clicks, whistles) are taken into consideration, along with other aspects of the audio (e.g. noise).

Introducing the basics of the deep learning model

.The values for the model parameters are determined via the training process, performed on a subset of the dataset.

.The remaining dataset samples are used to test the model.

.The model doesn’t know what a click or a whistle is.

.It classifies events (i.e. files) as either striped dolphins or sperm whales.

.It does not recognize other labels if they have not been present during training.

.It may be sensitive to the audio time duration, which means that we don’t know how it will perform for audio inputs of varying time duration (e.g. 2-week audio).

.The model’s performance is evaluated with confusion matrices, accuracy, etc.

What is PAMGuard

.PAMGuard is “an integrated passive acoustic monitoring (PAM) software infrastructure”.

.It is an open-source software, written in Java and available for Windows, Linux and macOS.

.It is a “world standard software infrastructure for acoustic detection, localisation and classification (of marine mammals)”.

.It offers a set of plug-in modules that are accessed through a graphical user interface (GUI).

.These modules are combined to create a data model that manipulates acoustic data (input) and produces useful information (output).

Stating our goals

1. Use PAMGuard for marine mammal detection on unprocessed audio files.
   - → Big duration of the audio stream (e.g. 2–week period).
   - → Possibility to load the file or to connect
   - → Performing species classification alongside the detection is a strong plus.
   - → Parts of the audio where detection is present, could be given to a marine biologist expert for classification.

2. Develop a non-learning approach for cetacean species classification and compare its results with a deep learning model that performs the same task.

PAMGuard – Click Detector

.The Click Detector is a PAMGuard module that is used for the detection of clicks in marine mammal audio.

.First, the signal is passed through a filter, in order to eliminate the irrelevant frequencies (e.g. low noisy frequencies).

.Then, the detection is performed according to signal to noise ratio (SNR). The sample bins whose SNR surpasses a given threshold, i.e. the loud sample bins, are detected as clicks.

.At sample bin i, the signal is computed as: $S_i = a_S \cdot |x_i| + (1 - a_S) \cdot S_{i-1}$

.At sample bin i, the noise is computed as: $N_i = a_N \cdot |x_i| + (1 - a_N) \cdot N_{i-1}$

PAMGuard – Click Detector

.The Click Detector parameters that can be adjusted by the user are:

–The pre-detection filter characteristics (e.g. type, order, cutoff frequency). Here, we use a high pass filter with a cutoff frequency of 1kHz

–The threshold over which clicks are detected (default is 10dB).

–The maximum duration of a detected click in sample bins (default is 1024).

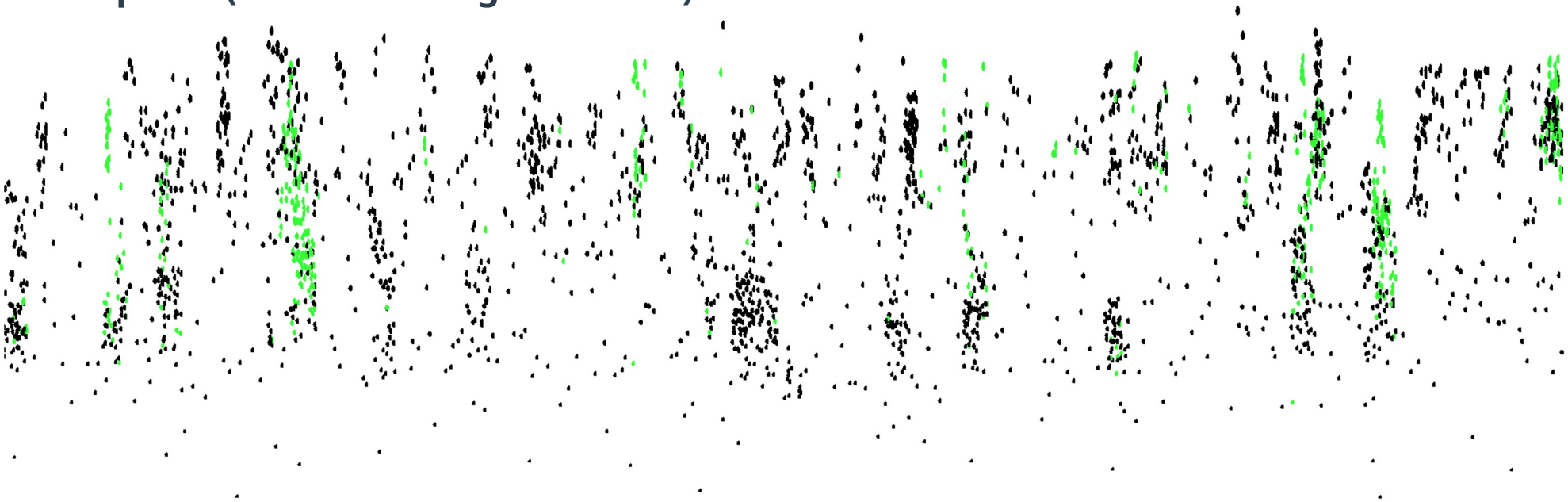
–The minimum distance between detected clicks in sample bins (default is 100).

.The detected clicks can be labeled/classified according to their time domain (e.g. time duration) and spectral (e.g. peak frequency) characteristics.

PAMGuard – Click Detector

.Click detection on overall dataset.

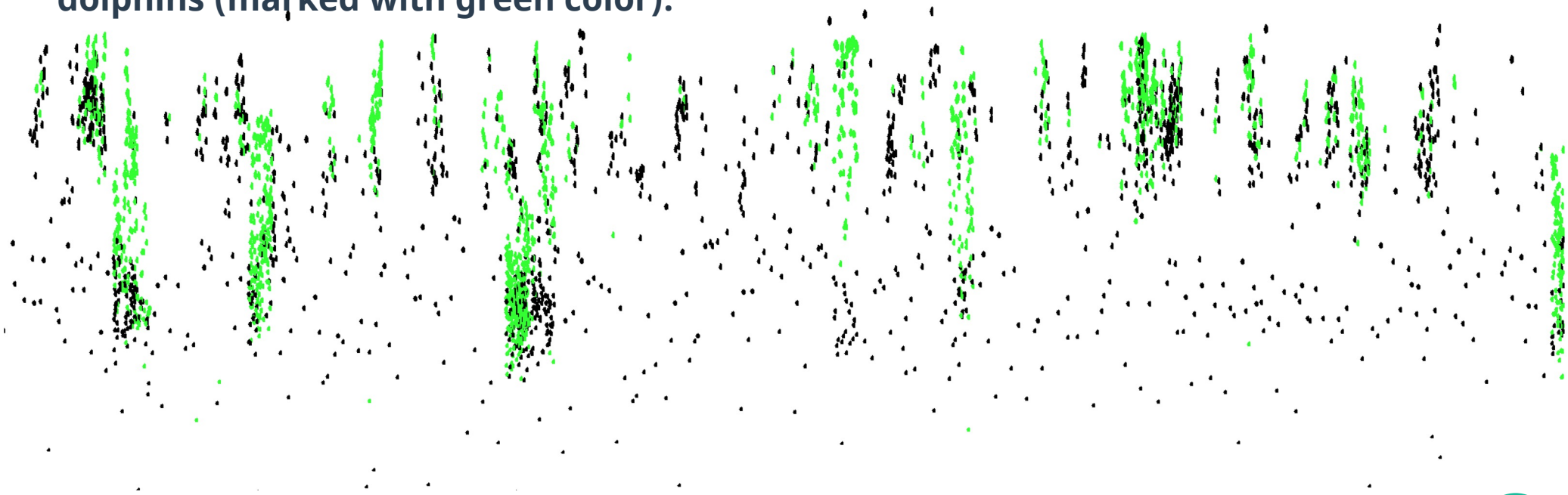
.Clicks with peak frequency between 15kHz and 50kHz are classified as striped dolphins (marked with green color).



PAMGuard – Click Detector

.Click detection on striped dolphin dataset.

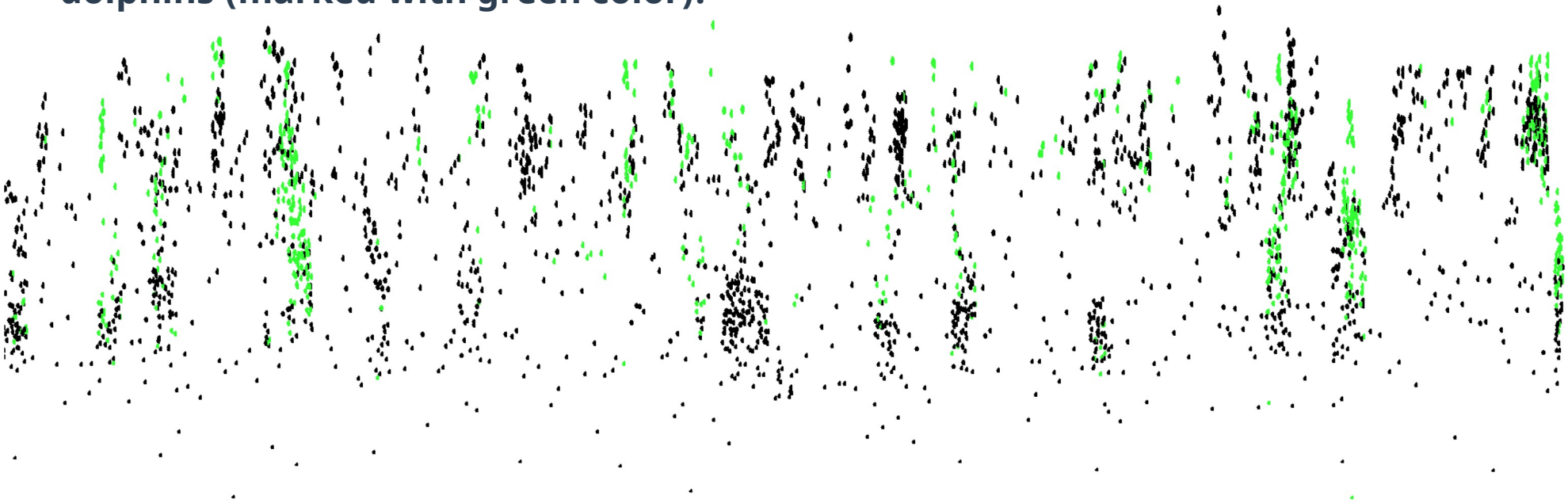
.Clicks with peak frequency between 15kHz and 50kHz are classified as striped dolphins (marked with green color).



PAMGuard – Click Detector

.Click detection on overall dataset.

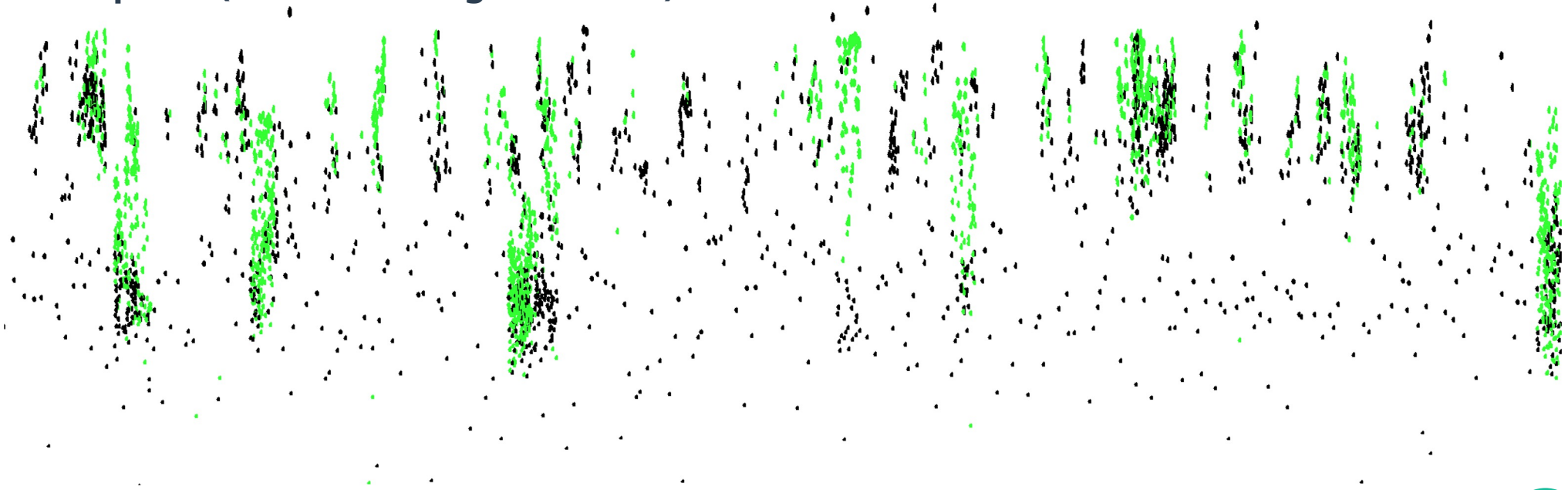
.Clicks with peak frequency between 10kHz and 50kHz are classified as striped dolphins (marked with green color).



PAMGuard – Click Detector

.Click detection on striped dolphins dataset (α_N 0.01 for no active click region).

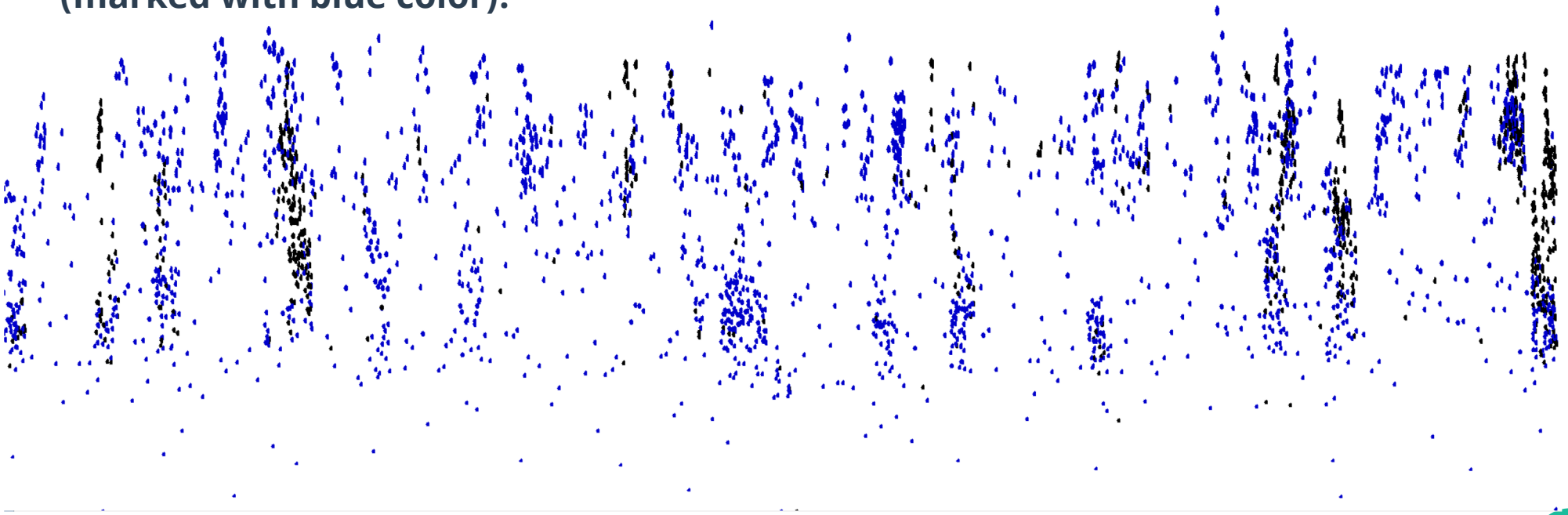
.Clicks with peak frequency between 10kHz and 50kHz are classified as striped dolphins (marked with green color).



PAMGuard – Click Detector

.Click detection on overall dataset (α_N 0.01 for no active click region).

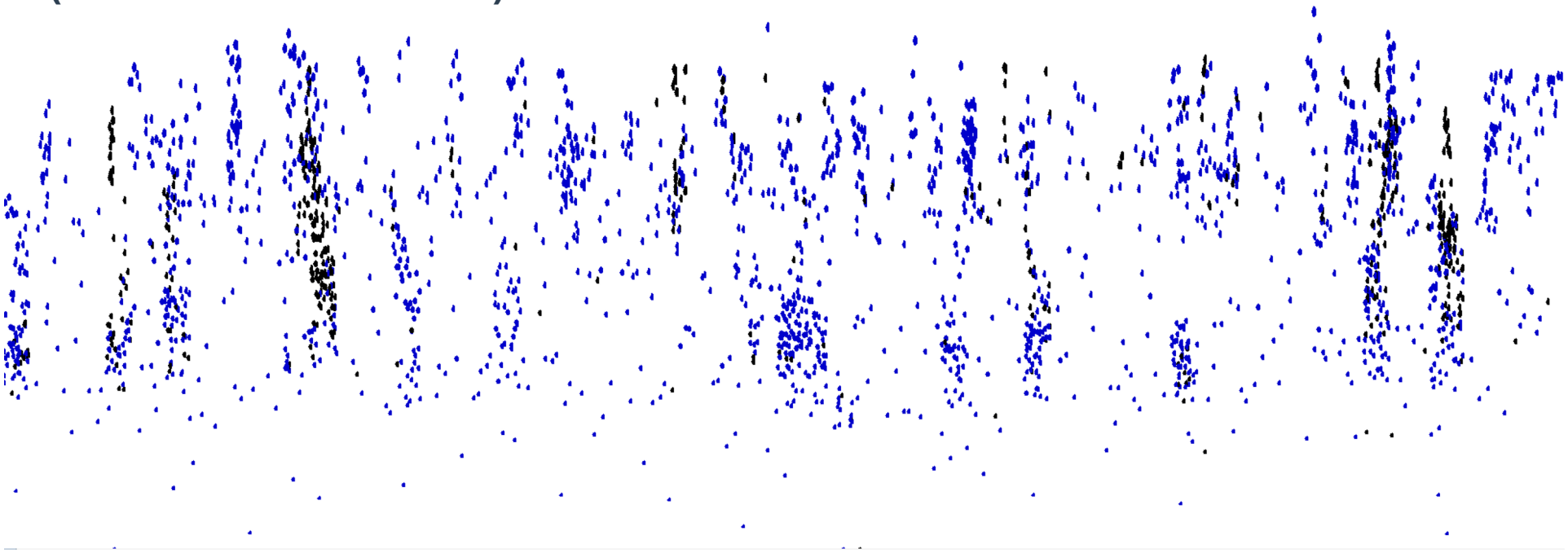
.Clicks with peak frequency between 1kHz and 15kHz are classified as sperm whales (marked with blue color).



PAMGuard – Click Detector

.Click detection on sperm whales dataset (α_N 0.01 for no active click region).

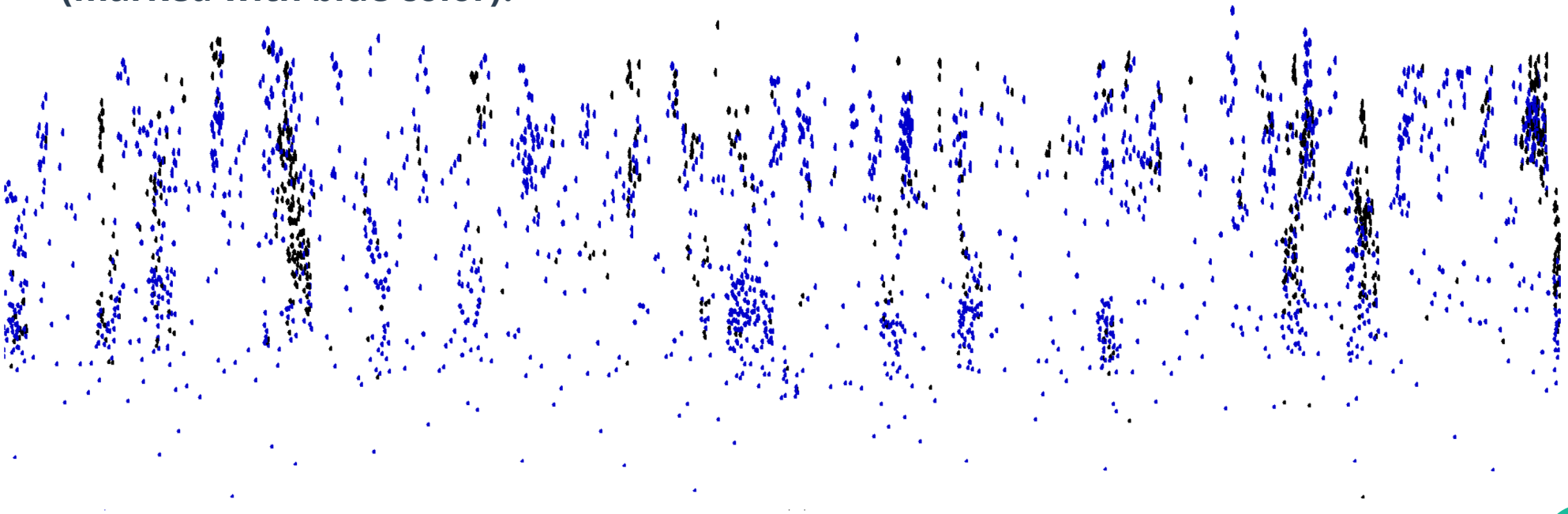
.Clicks with peak frequency between 1kHz and 15kHz are classified as sperm whales (marked with blue color).



PAMGuard – Click Detector

.Click detection on overall dataset (α_N 0.01 for no active click region).

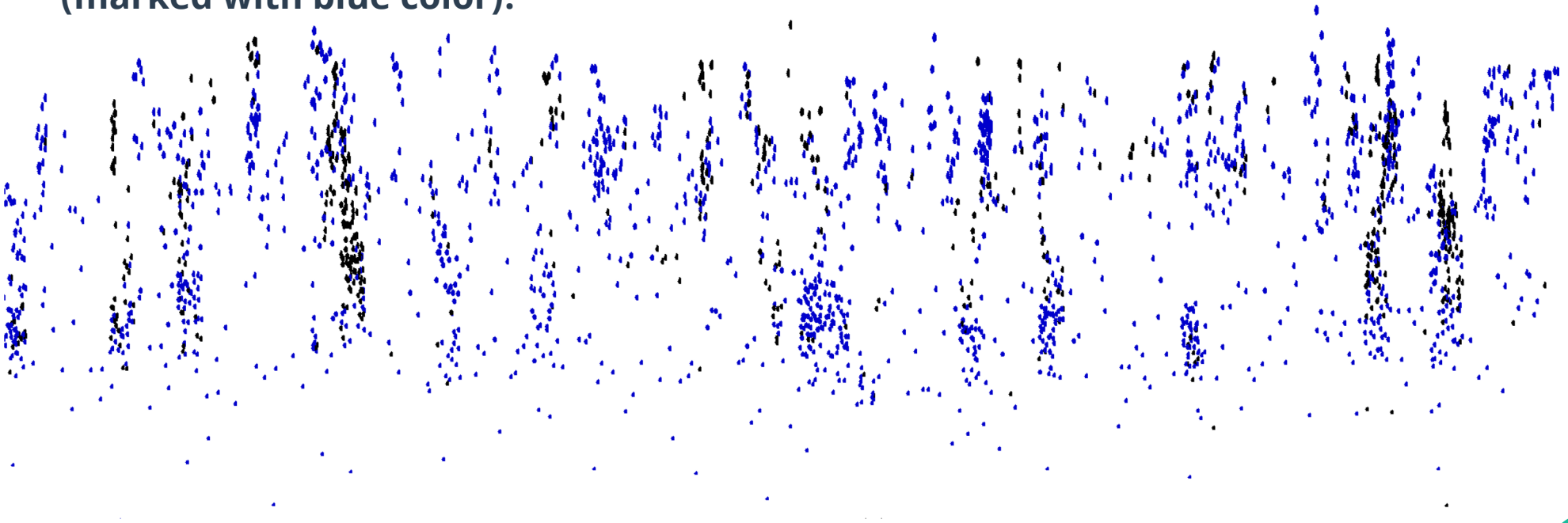
.Clicks with peak frequency between 1kHz and 10kHz are classified as sperm whales (marked with blue color).



PAMGuard – Click Detector

.Click detection on sperm whales dataset (α_N 0.01 for no active click region).

.Clicks with peak frequency between 1kHz and 10kHz are classified as sperm whales (marked with blue color).



PAMGuard – Click Detector

.What it can do:

–Detect clicks on a streaming audio (from either a single file or a batch of files).

–Label the clicks according to key click characteristics.

–Using multiple labels (e.g. both SD and SW) is possible, but when overlap and ambiguity is present, the order of the labels is important (the first one is favored).

–Provide a visual representation of the results.

. What it cannot do:

–Quantify the results as classification metrics (e.g. accuracy, confusion matrix).

–Infer a label for the entire stream or a section of it (i.e. a file/event).

–In order to utilize the PAMGuard click detections/classifications, we need to load them to Matlab (or R) and execute our own (simple or complex) algorithm.

Multiple Labels Classification

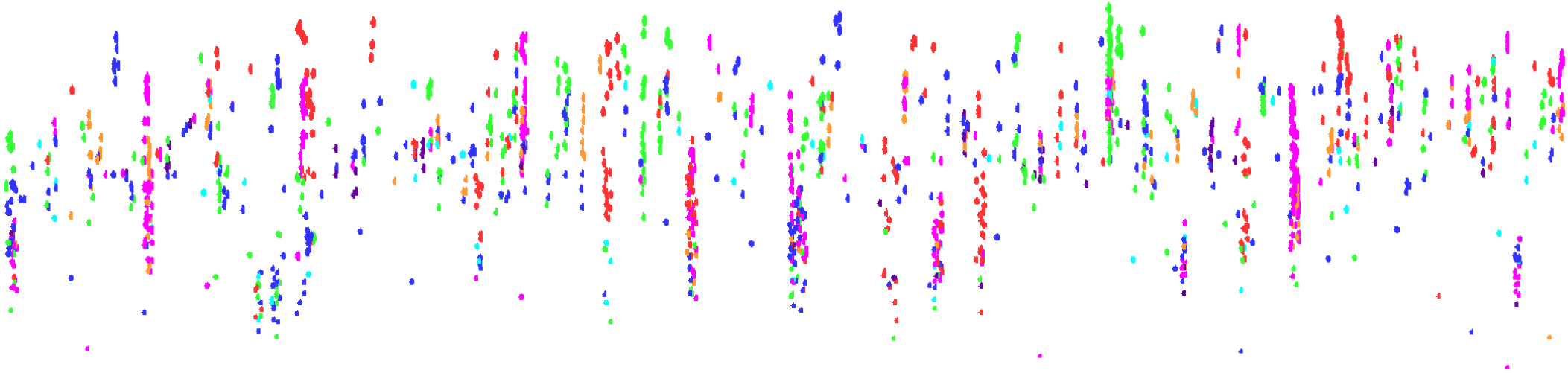
. To see if it was possible to classify two species with close sounds patterns, we tried to find some frequencies where one specie is more represented than the other :

4 labels have been set for SW :
Label 1(navy blue) = Peak frequency (PF) (3500-4000Hz)
Label 2(green) = PF (5000-7000Hz)
Label 3(red) = PF (10000-15000Hz)
Label 4(sky blue) = PF (1000-3000Hz) and Energy Bands (5000-7000Hz and 10000-15000Hz)

3 labels have been set for SD :
Label 5(orange) = PF (19000-25000Hz)
Label 6(pink) = PF (27000-36000Hz)
Label 7(purple) = PF (1000-3000Hz) and Energy Band (19000-25000Hz)

Multiple Labels Classification

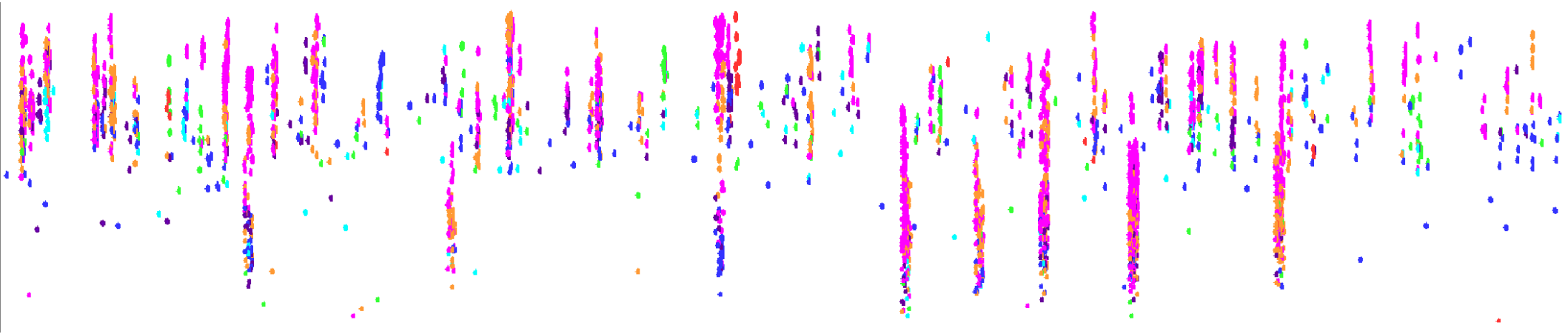
Clicks classification on sperm whales dataset :



SW : Navy blue / Green / Red / Sky blue
SD : Orange / Pink / Purple

Multiple Labels Classification

Clicks classification on striped dolphin dataset :



SW : Navy blue / Green / Red / Sky blue
SD : Orange / Pink / Purple

Multiple Labels Classification

```
Score Stripped Dolphin files
SD_countSW = 39
SD_countSD = 308
SD_TIE = 0
SD_NO_DETECTION = 0
SD_NO_IDENTIFICATION = 1



Score Sperm Whale files
SW_countSW = 70
SW_countSD = 114
SW_TIE = 0
SW_NO_DETECTION = 0
SW_NO_IDENTIFICATION = 0
```

```
Score Overall
Total_countSW = 109
Total_countSD = 422
Total_TIE = 0
Total_NO_DETECTION = 0
Total_NO_IDENTIFICATION = 1


Stats
Classification
True_Positive = 71.1864
False_Positive = 28.8136
```

.First result with Matlab post processing.

What is PAMpal

.PAMpal is a CRAN package that is used for loading, organizing and processing data produced by PAMGuard.

.It was developed with Matlab during 2020-2021 and first cited on nature, in a work that performs acoustic differentiation and classification of wild belugas and narwhals in Baffin Bay, West Greenland.

.Zahn et al. Acoustic differentiation and classification of wild belugas and narwhals using echolocation clicks. (2021).

.In this work, PAMpal was used to:

- **load click detections organized in events, as produced by PAMGuard**
- **calculate key click parameters (e.g. peak frequency)**
- **perform multivariate analysis to examine potential species differentiation**
- **perform species classification on an event basis**

Why use PAMpal

.It makes loading and organizing PAMGuard outputs on R easy and offers a set of useful functions for manipulating them.

.E.g., we can load click detections, calculate their key characteristics, calculate the average of each click characteristic per event/file and get an idea of each click characteristic distribution per species for our dataset (Figure S1).

.We can use multivariate analysis on chosen characteristics to examine potential differentiation between species and plan classification accordingly (Figure 2).

.We can use the classifiers that are offered in the PAMpal package to perform classification, instead of coding our own classification algorithm. The output will be comparable to the neural network model output (Table 3).

.We can easily integrate output from other PAMGuard modules, e.g. whistles.

Open questions

.Do we need to translate the click classification as produced by PAMGuard to event classification, in order to compare with the deep learning model?

.If not, how are we going to do the comparison? (We need to somehow infer accuracy, confustion matrices, etc.)

.Do we want to use Matlab and code our own translation of click classification to event classification?

.Or R with PAMpal and use the offered functions/methods for click classification to event classification translation?

.Do we want to consider more information, e.g. whistles?

.How are we going to address the limitations introduced by the low sampling rate?