

PAM: Large Dataset Analysis Course Overview

Passive acoustic monitoring is a well-established method for long-term monitoring of terrestrial and marine environments. Cost effective equipment and data storage costs further allow for large volumes of recordings to be continuously collected over extended periods of time. However, the processing of large acoustic datasets can be cumbersome and time consuming. “PAM: Large Dataset Analysis” is an asynchronous, online technical training course designed to provide bioacousticians with an approach to processing large passive acoustic datasets using a combination of accessible software including PAMGuard and several R programming packages.

In this approach to passive acoustic dataset analysis, we provide an end-to-end workflow that you can use for annotating and summarizing sounds within your data. We suggest this approach can be used for marine or terrestrial acoustic datasets and we include activity datasets from both environments. The course starts with the semi-automated approach to dataset annotation using a combination of machine learning algorithms, and post-automation grouping of calls within PAMGuard. PAMGuard has long been touted as a real-time monitoring tool but is also capable of effective processing of archival data using a suite of automated detectors and review modules. Once data are annotated in the binary and database format used by PAMGuard, we extract measurement and from the R package ‘PAMPal’ to format, query and visualize the annotated data. The course then guides you through the process of generating representative figures of example calls in your data using the ‘seewave’ package in R. Finally, we provide guidance on descriptive statistics for your data and examples of inferential statistics that can be used to compare measurements of your data using the statistical platform jamovi (R-based graphical user interface, or GUI). Marine and terrestrial datasets are provided for data processing practice (selecting one dataset to use throughout the course).

The course is designed for users whose experience levels include:

- **Bioacoustics:** familiarity with the topic and understand fundamental elements of passive acoustic monitoring, mitigation or research.
- **PAMGuard & SQLite Studio:** some experience with PAMGuard is expected. We do not step through the core elements of PAMGuard, but the “PAM Software Basics” course is available if users require this training, as are several resources on the PAMGuard and OSA websites. PAMGuard is installed on local machine so a Windows PC with a minimum of 8 GB of RAM is required. NOTES: processing large acoustic datasets is

computationally intensive; we mitigate this to accommodate a range of computing power by having you process shorter segments of data in the examples. However, we recommend applying these techniques to months of data, that you have access to a powerful computer or virtual machine. *If access to computer is limited, inquire with us regarding availability of a virtual machine at an additional cost.* SQLite Studio is downloaded to accompany PAMGuard detection summary extraction from database. Microsoft Excel is used to format extracted csv file.

- **R:** We assume no-to-novice experience level in R, and the design of the modules based in R as such that you will not need to have programming experience to complete them. We use R notebooks and Posit Cloud during these modules. Please note, this course is not intended to teach you how to program in R, nor extensively train you in all aspects of the ‘ggplot2’ and ‘seewave’ packages – it is intended to provide you with some example data visualization techniques and code you can use in your own work. If you are desiring to learn more regarding the basics of R and analytical uses, see our “R for Ocean Science Data Analysis” course.
- **jamovi:** this tool is a statistical analysis GUI that incorporates R under the hood. It offers an easy user interface for conducting statistical evaluation and analyses and is easy to install on local computers.

Given the likely variability in experience for the content in the final two modules, participants may select from a “Full Course Complement” (including all modules below) or a “PAMGuard and PAMpal Only” option that include Modules 1-4 from the list below.

Access to the course is provided for a period of one month to enable ample time to review all guided videos and complete activities at your own pace. The semi-interactive format includes email feedback on your module activity submission by an experienced course instructor. Upon completion of all modules, participants will receive a certificate of completion (see details below)

Course Topics

MODULE 1: Overview

- 1.1: Course Overview
- 1.2: Characteristics and challenges of large acoustic datasets
- 1.3: Additional bioacoustics software & datasets

MODULE 2: PAMGuard: Automated Processing

- 2.1: Data selection & basic configuration prep
- 2.2: Automated detection: configuration of detectors
- 2.3: Creation of User Defined Form for annotation
- 2.4: Accessory modules: LTSA, Noise Band Monitor, etc.
- 2.5: Processing with example dataset

MODULE 3: PAMGuard: Post-automation Annotation

- 3.1: Navigation of panels in PAMGuard Viewer Mode
- 3.2: Annotating data
- 3.3: Best practices for expediting review
- 3.4: Processing with example dataset

MODULE 4: Data Extraction (SQLite Tables & PAMPal)

- 4.1: Extraction of Detection Group Summary Table from SQLite
- 4.2: PAMPal processing: data ingestion, management & measurements
- 4.3: PAMPal: associating recordings & adding environmental data
- 4.4: Additional PAMPal capabilities
- 4.5: Processing with example dataset

MODULE 5: Data Visualization

- 5.1: Figure creation in PAMPal
- 5.2: Visualizing data with base plot & ‘ggplot2’
- 5.3: Creating a spectrogram using ‘seewave’
- 5.4: Processing with example dataset

MODULE 6: Data Summarization & Statistical Analysis

- 6.1: Overview of descriptive and inferential statistics for univariate data
- 6.2: Tabular and graphical descriptive statistics in jamovi
- 6.3: Basic statistical tests for detections/events in jamovi
- 6.4: Processing with example dataset

Certificate of Completion: This course is recognized as a continuing professional development (CPD) course as assessed by the [Institute of Marine Engineering, Science and Technology](#) (IMarEST). This independent internationally recognized organization attests to the high standards and quality of this marine scientific focused training. Upon completion of this course, you will receive a certificate indicating the course topics completed.