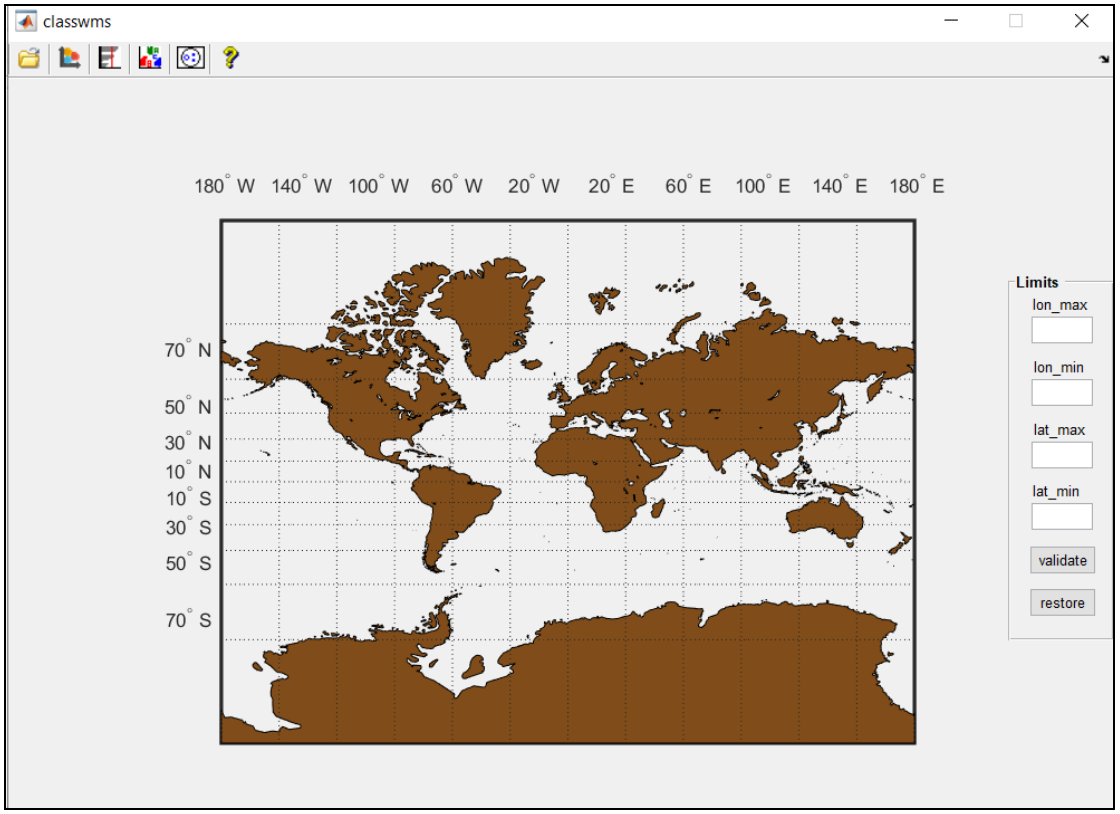



# CLASSWMS GUI

## HELP DOCUMENTATION

*CLASSWMS is a Graphical User Interface developed under MATLAB, to classify water masses using clustering analysis and KNN classifier, according to the process presented by (Ayoub BELATTMANIA et al 2023).*

Elaborated by: Ayoub BELATTMANIA
February 2024

# 1. OVERVIEW

The classwms tool enables geoscientists, particularly oceanographers, to classify water masses automatically on the basis of hydrological profiles gathered in situ. The application consists of two classification methods:

- The clustering analysis:

This approach is employed on a T-S diagram and relies on the iterative k-means algorithm, a conventional method used to define the characteristics of water masses.
To assess the effectiveness of the clustering analysis, users have the option to calculate the silhouette coefficient.

- K Nearest Neighbors Classification:

The method, extensively explained in the investigation by Ayoub BELATTMANIA et al 2023, relies on a novel supervised classification approach applied to the potential density and potential spicity (σ-π) diagram.

# 2. USAGE AND SUPPORT MATERIAL

CLASSWMS has been meticulously crafted and rigorously tested to operate seamlessly across Windows versions 7, 8, 10, and 11, all in their 64-bit versions. The utilization of the classwms source code mandates the presence of MATLAB R2021a version or higher and the Gibbs-SeaWater (GSW) Oceanographic Toolbox. Before the interface displays on the screen, users are prompted to specify the path to the toolbox upon launching the main code.
The target system running the standalone application (executable file) does not require a licensed copy of MATLAB. Nevertheless, it requires the MATLAB Runtime installation R2021a (https://www.mathworks.com/products/compiler/matlab runtime.html) to run the application. Additionally, the GSW toolbox is integrated within the executable file.

# 3. MAIN FUNCTIONALITIES

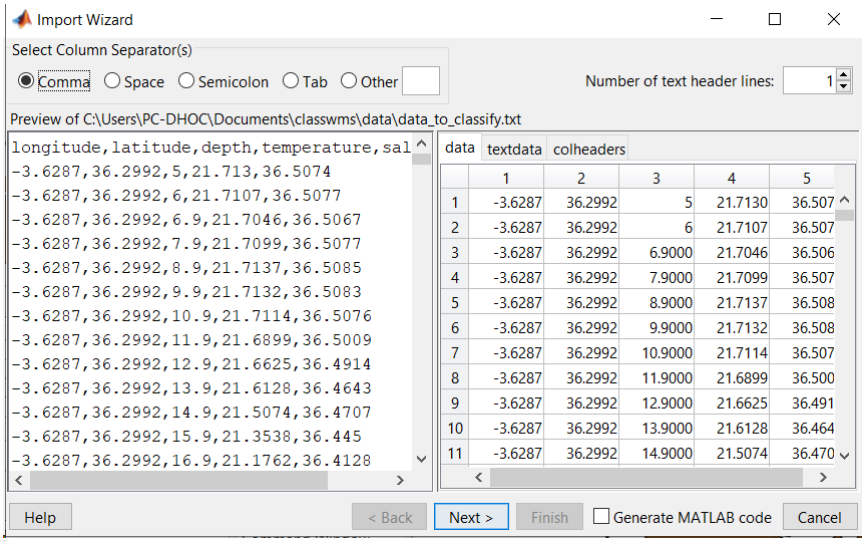
**Import data**

The Import data option is used to import the data file that the user wishes to classify. This file, called 'data_to_classify.txt', should conform to the following specifications:

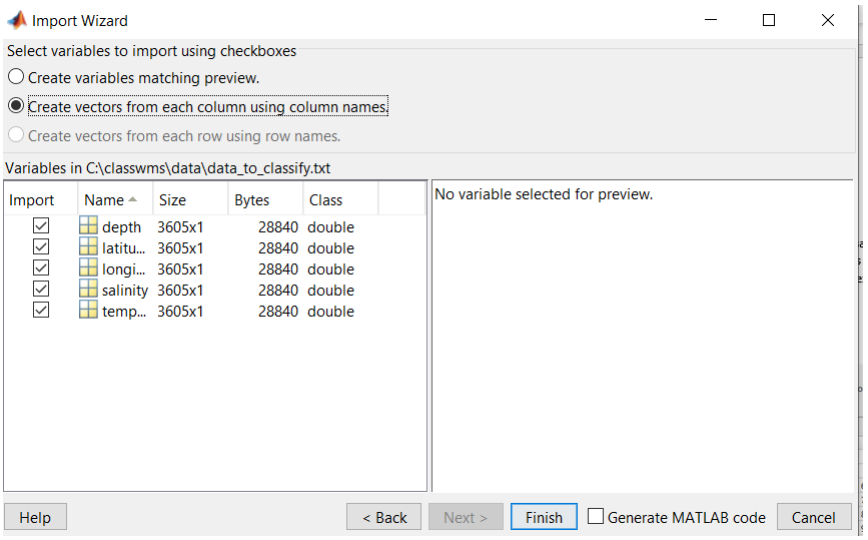
- ✓ Line 1 (file header): longitude, latitude, depth, temperature and salinity.
- ✓ Line >1 (data): Coordinates should be in Decimal Degrees and range from -90 to 90 for latitude and -180 to 180 for longitude. Outliers must be replaced with NANs.

Here are the steps involved in this operation:

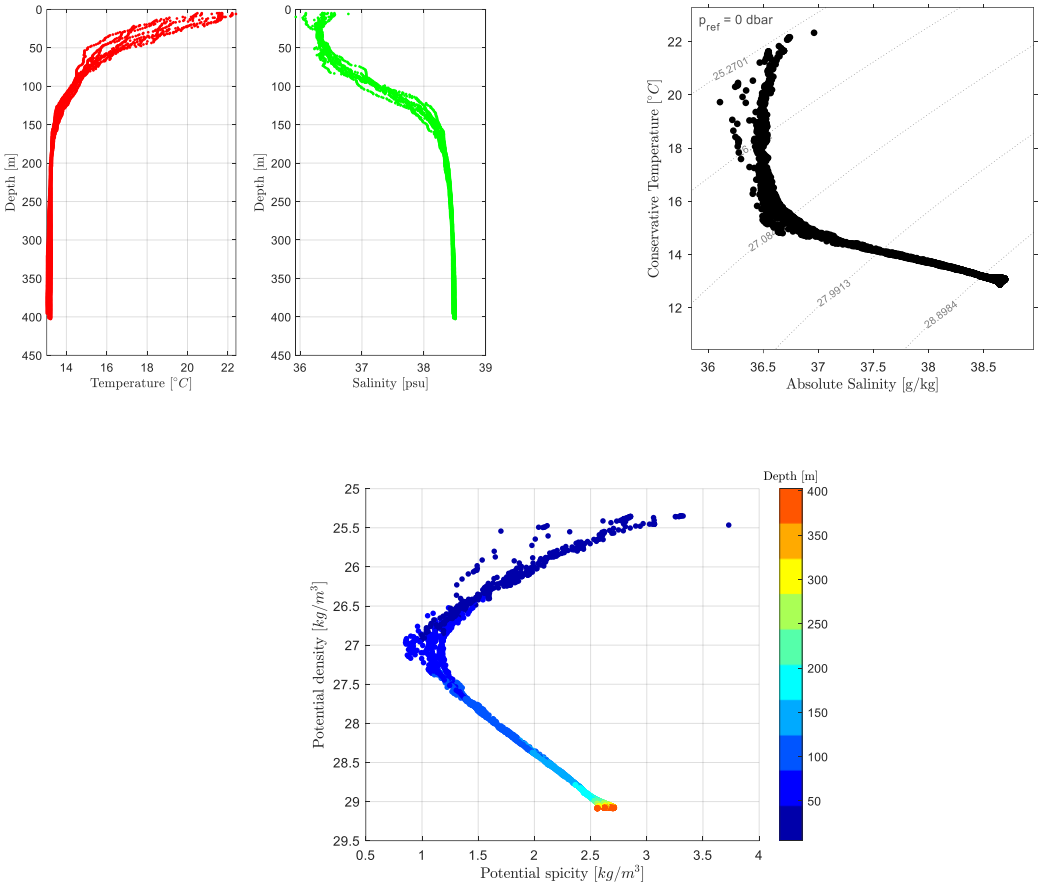
- **Separator selection**



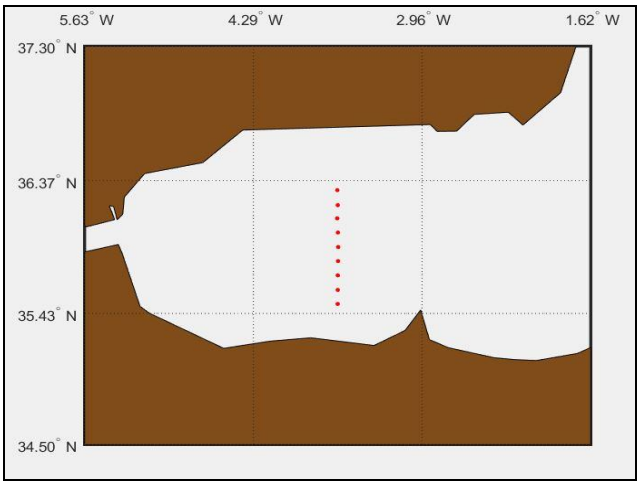

- **Assigning columns to fields**




After the process is finished, the ensuing visuals will be showcased. These encompass TS profiles, Conservative Temperature-Absolute Salinity diagram, and their corresponding σ-π diagram.

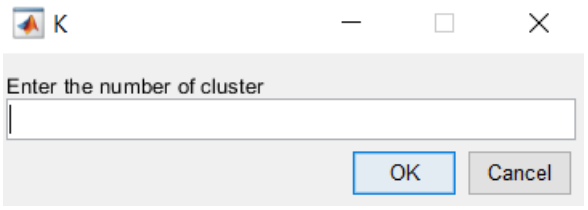


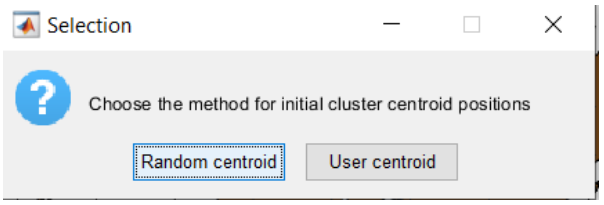

Also, the geographic coordinates of the profiles are depicted on the map. To modify the geographical scope, the user enters the boundary parameters and confirms the selection. The global map is re-displayed upon clicking the restore button.

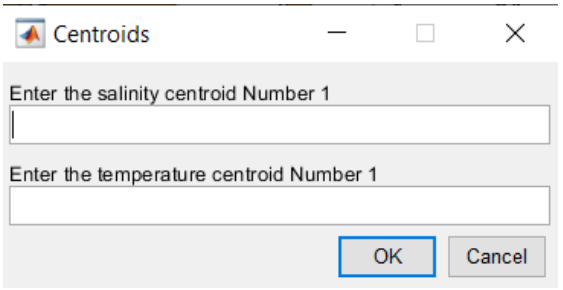


## k-means

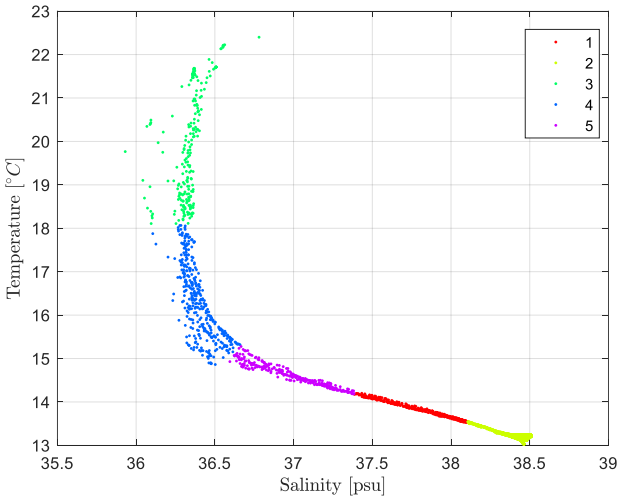
This functionality is employed to implement the k-means algorithm on the imported dataset. The user selects the desired number of clusters and subsequently specifies the method for centroid selection. Opting for the 'Random centroid' entails initializing k- means with randomly chosen centroids. Conversely, opting for 'User centroid' offers the user the opportunity to define temperature and salinity values for centroids.



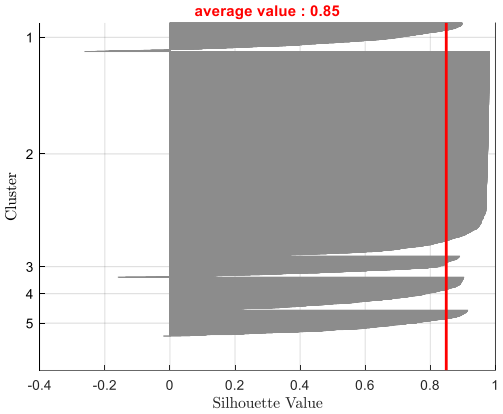




Following confirmation, the interface produces the outcome of water mass classification, presented in the format of a TS diagram.



## Silhouette plot

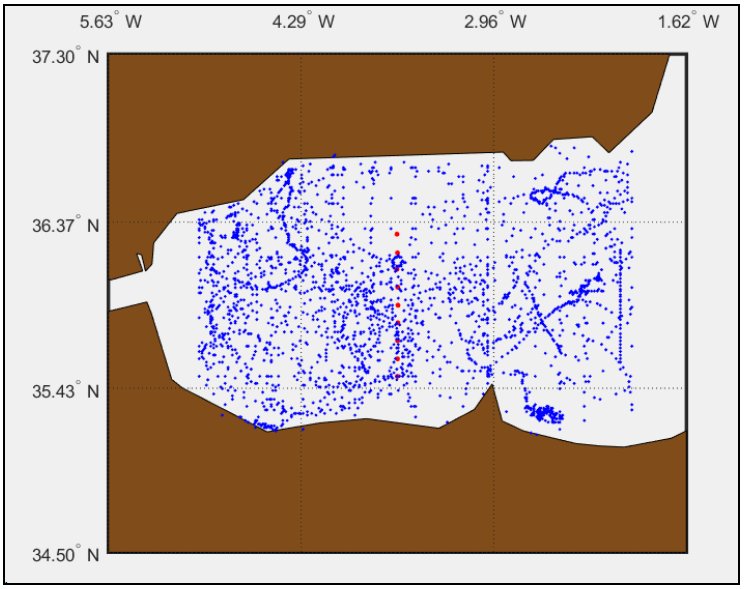
Silhouette values can be graphed trough ‘Silhouette plot’ button, to visually evaluate the selected number of clusters.



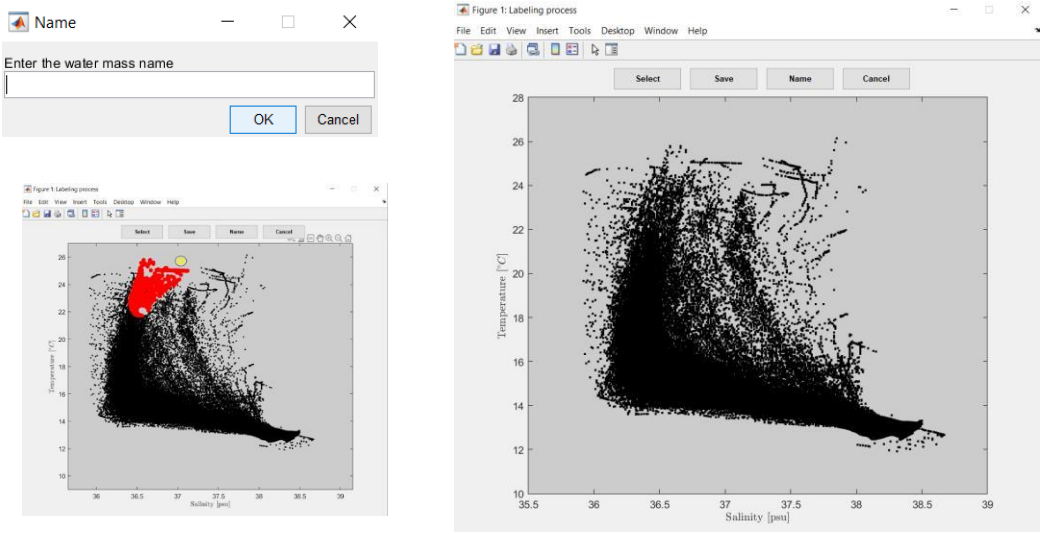

## Labeling process

This functionality is employed for importing the user-designated database file. The file, referred to as 'database_not_labeled.txt', comprises the subsequent fields: longitude, latitude, depth, temperature, and salinity. The file should adhere to the same format as 'data_to_classify.txt'.

Also, the geographic coordinates of the database profiles are ploted on the map. To modify the geographical scope, the user enters the boundary parameters and confirms the selection. The global map is re-displayed upon clicking the restore button.



After that, an interactive window appears to allow labeling.

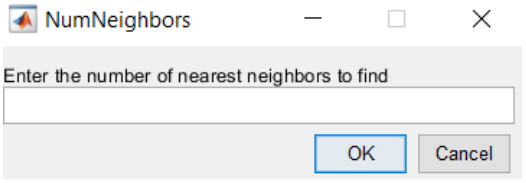


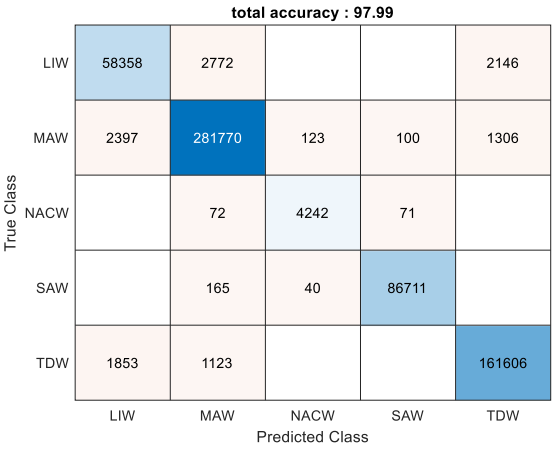

Initially, the user selects the water mass's name for labeling and subsequently initiates the process of sample selection associated with that particular water mass. Upon clicking the 'save' button, the data undergoes conversion into a potential density and potential spicity (σ-π) diagram, after which it is stored in a file encompassing the subsequent attributes: spi, sigma, WM. This procedure needs to be reiterated based on the count of distinct water masses.

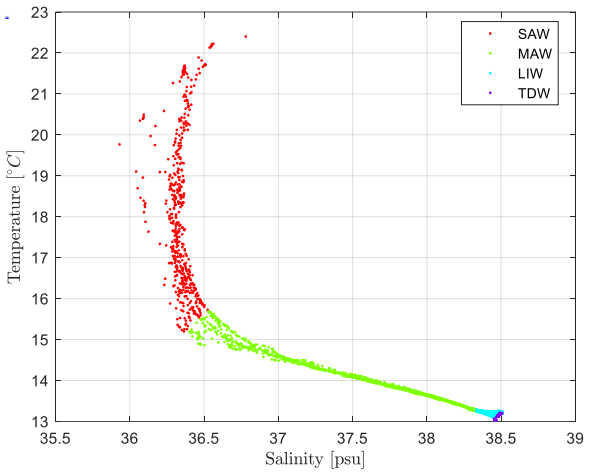
Upon completing this process, the user should combine the individual files for each water mass into a single file named 'database_labeled.txt', which should include a header comprising the fields: PI, SIGMA, NAME.

## KNN classifier

This feature is utilized to implement the k-nearest neighbors (knn) algorithm for the classification of water masses. The prepared labeled database file is imported automatically, initiating the training process. Subsequently, confusion matrix charts are generated, and the interface produces the results of water mass classification in the form of a TS diagram.




## 4. DATA REFERENCES

- The 'data_to_classify.txt' file, contains a hydrographic (CTD) cast of an intensive oceanographic survey (BIOMEGA) collected on board of the Spanish R/V Garcia del Cid during October 2003. Data were provided through SeaDataNet Pan European infrastructure for ocean and marine data management (https: //www.seadatanet.org).

- The 'database_not_labeled.txt' file, contains in-situ observations from oceanographic databases such as World Ocean Database 2018 (WOD18) and the Global Data Assembly Centers (GDACs).