

EcoTransLearn R-package

Version 1.0-1

USER MANUAL

G. WACQUET

*IFREMER
LABORATOIRE ENVIRONNEMENT RESSOURCES
CENTRE MANCHE MER DU NORD*

TABLE OF CONTENTS

INTRODUCTION	4
INSTALLATION AND EXECUTION	4
R-package installation	4
Python installation	4
Launching the Graphical User Interface	5
USE OF THE GRAPHICAL USER INTERFACE	6
DATA SELECTION button	6
SETTINGS button	6
CLASSIFY button	12
VIEW button	13
MORE... (+) button.....	14
Recommended folder tree	17

INTRODUCTION

In recent years, Deep Learning (DL) has been increasingly used in many fields, in particular in image recognition, due to its ability to solve problems where traditional machine learning algorithms fail. However, building an appropriate DL model from scratch especially in the context of ecological studies, such as monitoring marine ecosystems, is a difficult task due to the dynamic nature and morphological variability of living organisms, as well as the high cost in terms of time, human resources, and skills required to label a large number of training images. To overcome this problem, Transfer Learning (TL) can be used to improve a classifier by transferring information learnt from many domains thanks to a very large training set composed of various images, to another domain with a smaller amount of training data. To compensate the lack of “easy-to-use” software optimized for ecological studies, we propose the *EcoTransLearn* R-package, which allows greater automation in classification of images acquired with various devices, thanks to different TL methods pre-trained on the generic ImageNet dataset.

INSTALLATION AND EXECUTION

R-package installation

The version **1.0-0** of the *EcoTransLearn* package needs a recent version of R (version **4.0.x** or upper). It can be directly downloaded on GitHub at: <https://github.com/IFREMER-LERBL/EcoTransLearn> (see ‘Releases’ page to download the .tar file.)

By double-clicking on the R icon on the desktop, or by selecting R in the start menu, a window appears on the screen: this is the R console. This allows to control R directly by command lines. It also allows to display the main results and messages of the different actions performed with *EcoTransLearn*.

The R-packages needed by *EcoTransLearn* (*colorRamps* **2.3**, *ggplot2* **3.3.5**, *grid* **4.0.5**, *jpeg* **0.1-9**, *mapplots* **1.5.1**, *maps* **3.4.0**, *randomForest* **4.6-14**, *reticulate* **1.23**, *SDMTools* **1.1-221.2**, *shapefiles* **0.7**, *stringr* **1.4.0**, *svDialogs* **1.0.3**, *svMisc* **1.2.3**, *tcltk2* **1.2-11**, *tiff* **0.1-10**, *zooimage* **5.5.2** or upper versions) can be installed directly from the R console, by typing:

```
install.packages(c("colorRamps",  
                  "ggplot2",  
                  "grid",  
                  "jpeg",  
                  "mapplots",  
                  "maps",  
                  "randomForest",  
                  "reticulate",  
                  "SDMTools",  
                  "shapefiles",  
                  "stringr",  
                  "svDialogs",  
                  "svMisc",  
                  "tcltk2",  
                  "tiff",  
                  "zooimage"))
```

Then choose a mirror (default: 0-cloud) to start download and installation.

Python installation

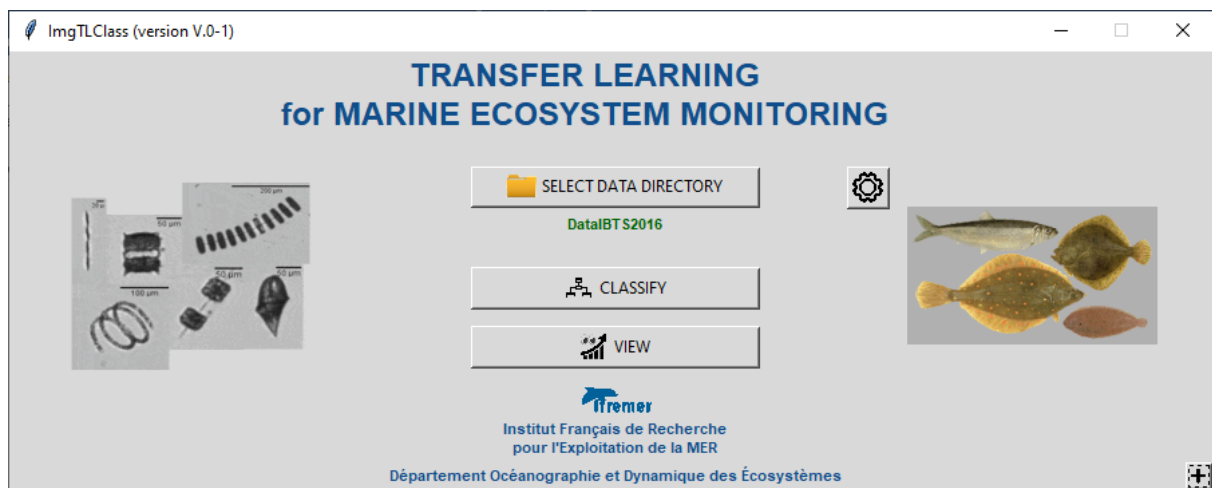
Anaconda is a scientific distribution of Python, which allows to use several applications (such as Spyder, Jupyter Notebook, ...) and to manage different libraries. It can be downloaded on the website: <https://docs.anaconda.com/anaconda/navigator/install/>. The *EcoTransLearn* package needs the version **3.7** (or upper) of Python.

Once the distribution installed, the Python libraries needed by *EcoTransLearn* (*matplotlib 3.3.4*, *numpy 1.19.2*, *pandas 1.1.5*, *sklearn 0.24.2*, *tensorflow 2.2.0* or upper versions) can be installed directly by opening an Anaconda Prompt, and by entering the commands:

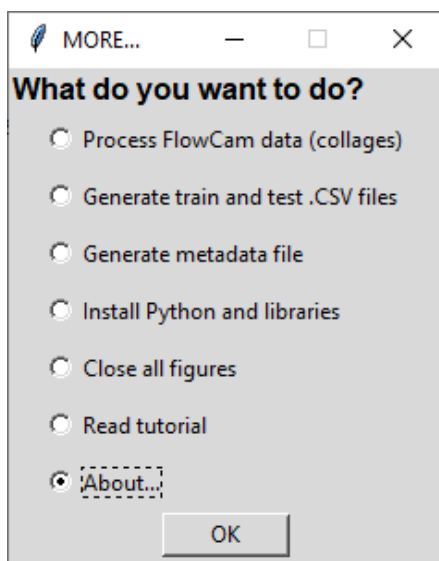
```
conda install matplotlib
conda install numpy
conda install pandas
conda install sklearn
conda install tensorflow
```

Launching the Graphical User Interface

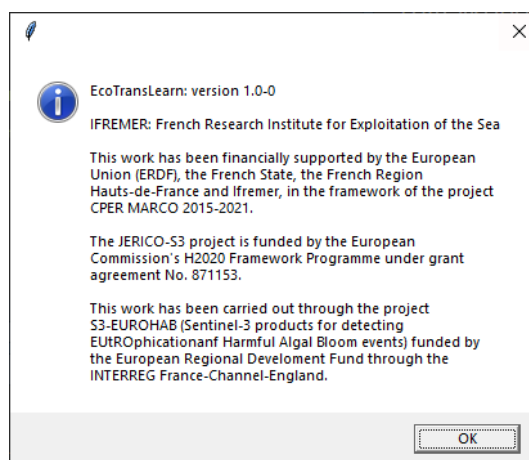
Once the installation of the packages is finished, it is possible to make sure that the previous steps run smoothly by checking that the installed version is **1.0-0**. To do this, first type in the R console `library(EcoTransLearn)`, to load the package, then `EcoTransLearn()`, to launch the Graphical User Interface (GUI):



Click on the + button (bottom right). A new window appears:



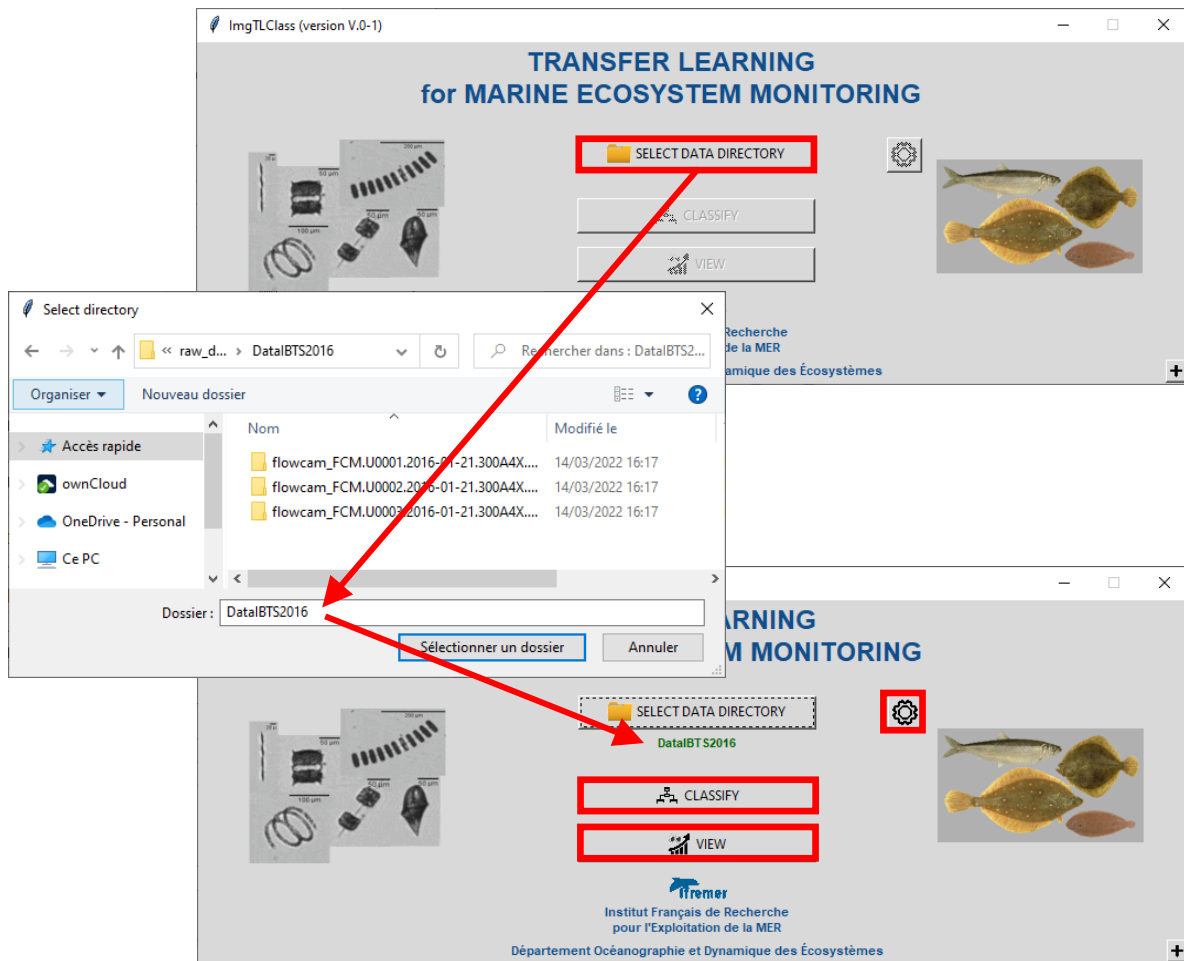
By selecting **About...**, a dialog box appears and informs the user of the *EcoTransLearn* version.



USE OF THE GRAPHICAL USER INTERFACE

For routine use of *EcoTransLearn*, an ergonomic Graphical User Interface is available.

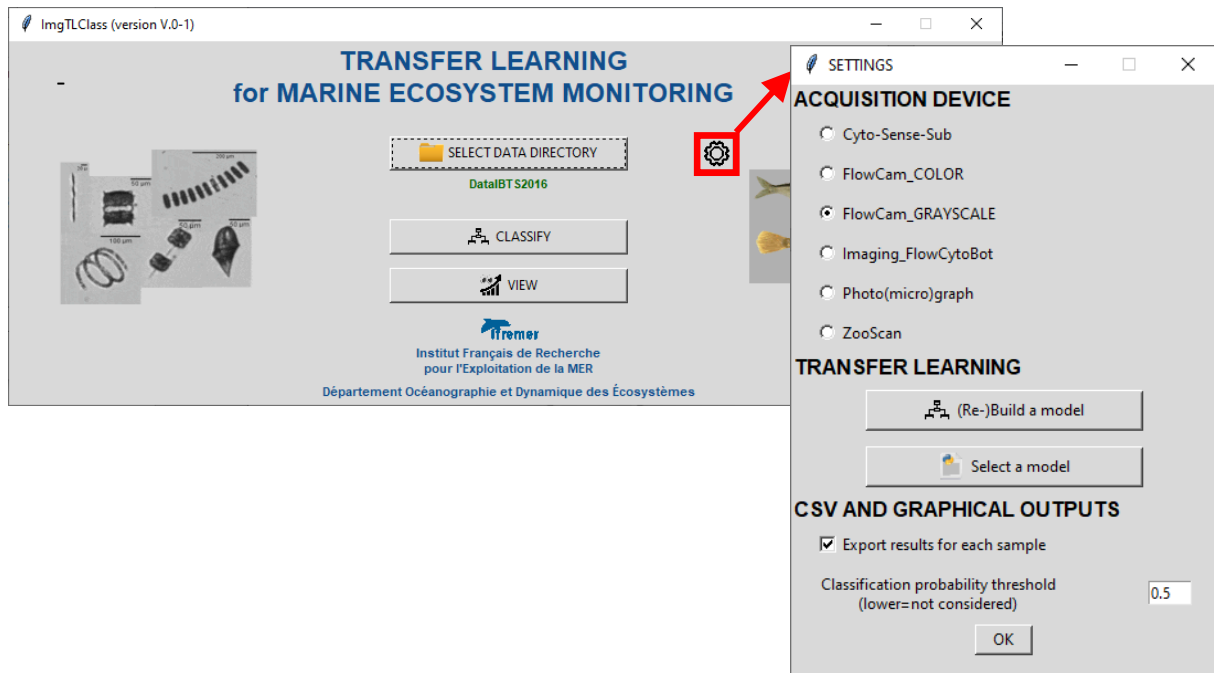
DATA SELECTION button



Click on the **SELECT DATA DIRECTORY** button, select the directory containing the images to classify, then validate by clicking on **OK**: the name of the selected directory is displayed below the selection button, and the **SETTINGS**, **CLASSIFY** and **VIEW** buttons become active.

SETTINGS button

After selection of the input data, the settings window is automatically displayed (without clicking on the **SETTINGS** button). However, it is possible to redisplay this window at any time by clicking on this button.



❖ ACQUISITION parameters

Choose the kind of instrument used for image acquisition.

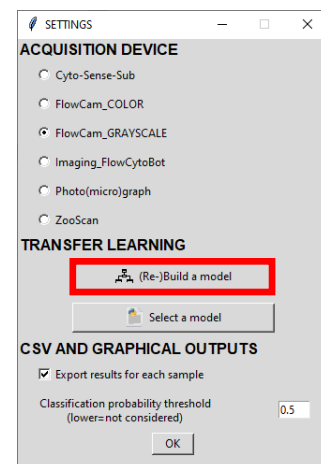
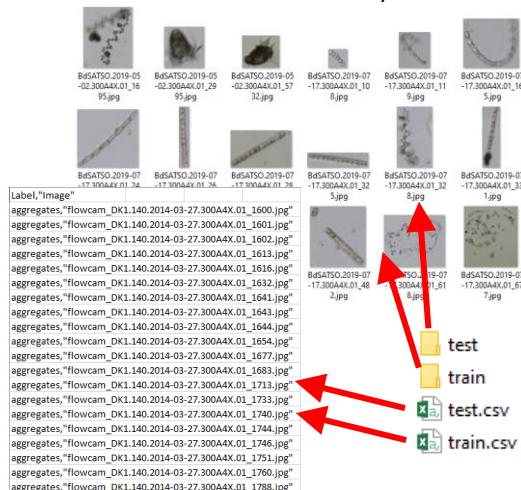
Note: for the FlowCam_COLOR and the FlowCam_GRAYSCALE, it is possible to directly output the raw data from the instruments (collages). The package allows to directly create the vignettes (1 image per particle) from the collages and the 1st file.

❖ CLASSIFICATION parameters

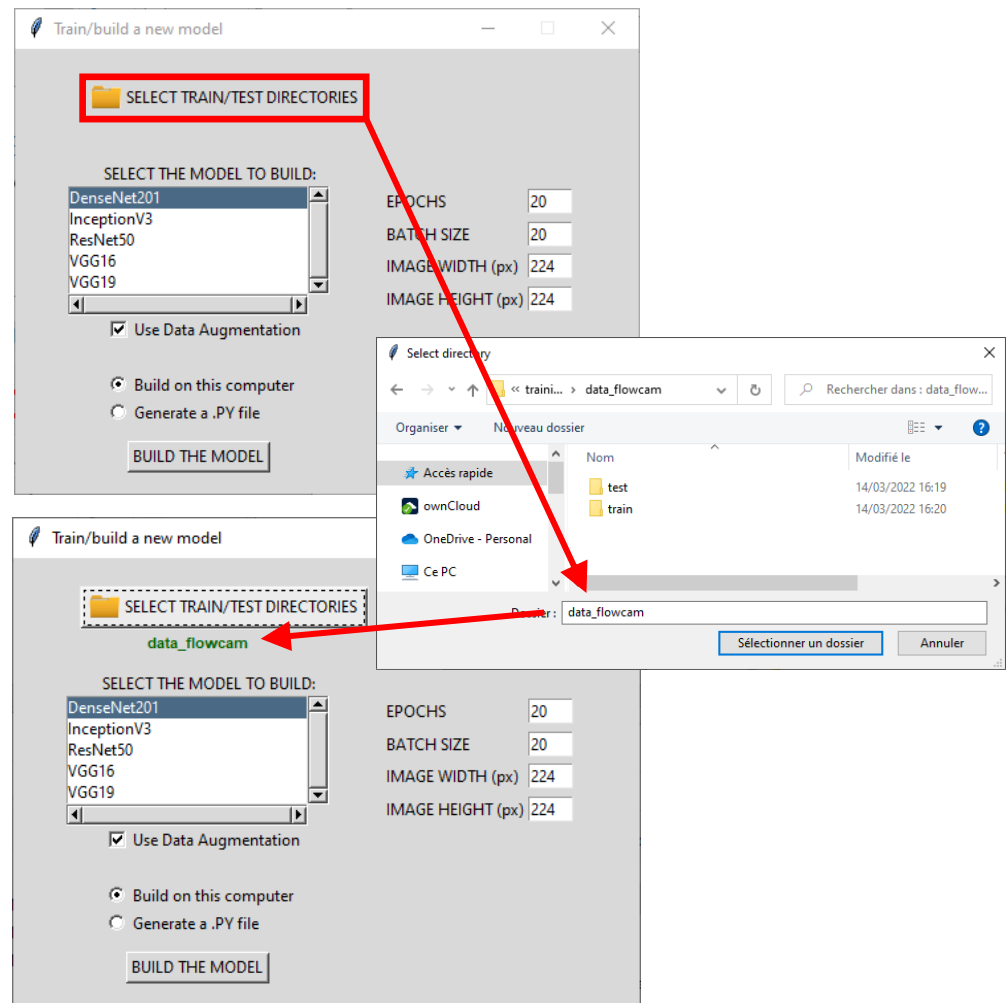
▪ (Re-)Build a model

This button allows to build (or rebuild) a classification model by Transfer Learning using a data set including a training set (directory must be named **train**) for learning, and a test set (directory must be named **test**) for validation and evaluation.

All unsorted images are bulked into these directories. Two CSV files (named 'train.csv' and 'test.csv') allows listing the group of each image (see **Generate train and test .CSV files** section).



By clicking on this button, a new window appears.



Click on the **SELECT TRAIN/TEST DIRECTORIES** button, select the directory containing the two sub-folders **train** and **test**, then confirm by clicking **OK**: the name of the selected directory is then displayed below the selection button

It is then possible to select different Convolutional Neural Network (CNN) architectures in the **SELECT THE MODEL TO BUILD** list, among:

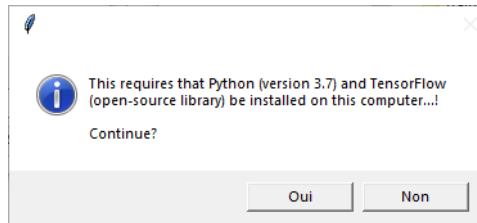
- DenseNet201
- InceptionV3
- ResNet50
- VGG16
- VGG19

and adjust parameters related to images and training step, by setting values for **EPOCH** (default=20), **BATCH SIZE** (default=20), **IMAGE WIDTH** (default=224) and **IMAGE WEIGHT** (default=224), but also choose the possibility of using the technique of data augmentation (**Use Data Augmentation**). In the case of a training set with few images, this option is used to automatically generate additional images from the original images in the training set, by applying geometric transformations such as rotations (by default, rotation_range=45) and horizontal and vertical flips (by default, horizontal_flip=True and vertical_flip=True).

The last step is to choose the material on which to build and adapt the classification model. Depending on the selected option, and after clicking on the **BUILD THE MODEL** button:

- **Build on this computer**

A dialog box appears:



Warning: the training duration can be long (several hours) depending on the number of images in the training set and according to the parameters defined in the previous step.

- **Generate a .PY file**

A script is automatically created, and can be run on other hardware (dedicated computer, calculation server, etc.).

Warning: three variables must be modified according to the data location ('trainPath', 'testPath' and 'save_dir').

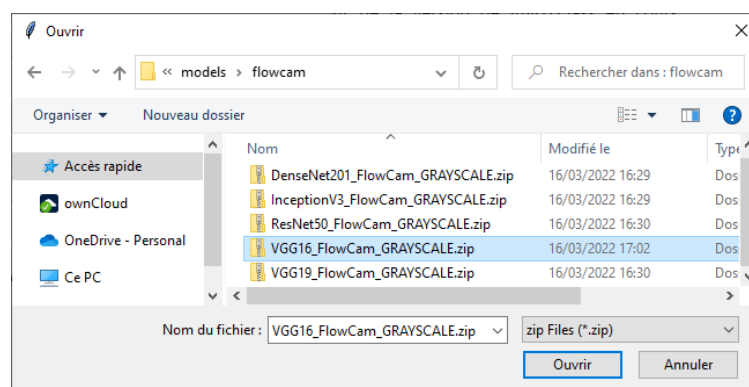
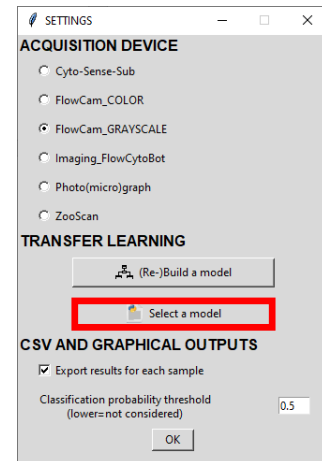
At the end of the execution of the script, a new folder is created at the root of the training/test set. In this directory (named '**saved_model**'), a ZIP file is saved and contains all the information necessary for the classification of a new set of images.

- VGG16_classnames.csv
- VGG16_history.csv
- VGG16_model.h5

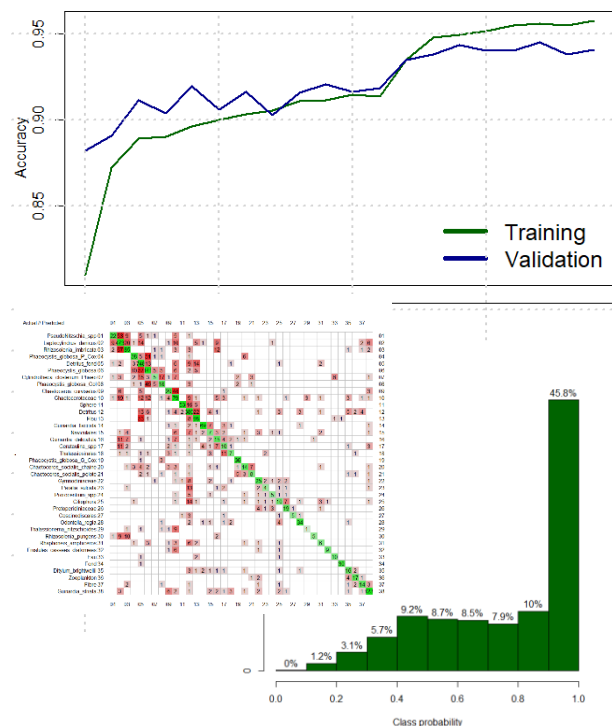
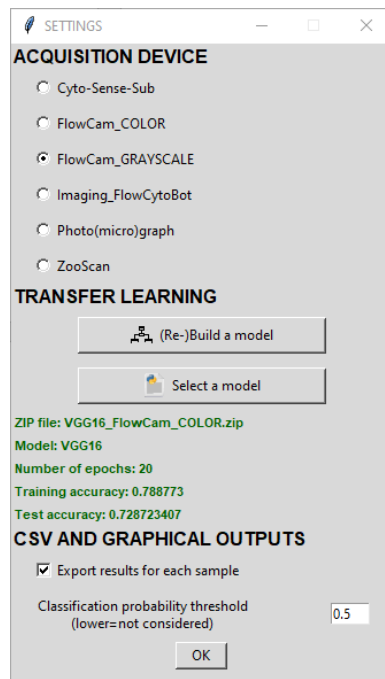
modelName_classNames.csv contains the class names in the training set.
modelName_history.csv contains the evaluation scores of the built model.
modelName_model.h5 is the model built with Python.

- **Select a model**

This button allows to choose a model for the automated classification of the images contained in the directory selected during the first step (see section “Selection of input data”).



Select the ZIP file generated during the previous step, then validate by clicking on **Open**: information on the model performance (calculated during the training and validation steps) is then displayed below the selection button, as well as **Accuracy** (≈ percentage of correctly classified data) and **Loss** (≈ distance between actual data and predicted data) curves, confusion matrix, etc.



The proposed graphical tools provide an overall view of recognition performance of the model through:

- the distribution of class probabilities, in order to have an overview of the degree of confidence for each identification ;
- the accuracy and loss curves computed during the training phase (by cross-validation), as well as on a validation dataset, as the epochs progress (the epoch is a hyperparameter that defines the number times that the training algorithm will work through the entire training dataset) ;
- the confusion matrix which is a square contingency table opposing the actual identification to the predicted ones (on the diagonal : correctly identified, off-diagonal: erroneous predictions), and which allows to define the groups presenting the most important confusions in order to be able to optimize the training set (by merging or separating groups, adding images, etc.).

❖ OUTPUTS parameters

▪ Export results for each sample

This option allows to export the results for each sub-folder of the directory selected during the first step (see section "Selection of input data").

For each processed sample, three CSV files can be generated:

sampleName_CLASSIF.csv

Filename	Class
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_10.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_100.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1000.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1001.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1002.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1003.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1004.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1005.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1006.jpg	Ciliophora
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1007.jpg	Odontella
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1008.jpg	dark
unknown\flowcam_FCM.U0003.2016-01-21.300A4X.01_1009.jpg	fiber

with Filename: the name of the image file
Class: the predicted class of the image

sampleName_PRED.csv

Asterionellopsis	C_curvisetus	C_danicus	C_socialis	Ciliophora	Dactyliolen	Dytilum	G_flaccida
3.52E-09	1.24E-08	4.22E-10	3.65E-11	0.000285448	1.65E-06	9.95E-09	7.16E-10
1.34E-11	2.78E-11	7.02E-14	1.16E-16	4.84E-10	5.95E-14	3.10E-13	3.12E-16
1.86E-12	2.97E-09	1.78E-13	8.65E-14	8.47E-05	1.23E-07	4.80E-09	1.05E-11
9.69E-17	1.54E-11	3.29E-18	7.69E-20	3.97E-08	1.30E-10	1.50E-12	2.34E-17
6.27E-08	5.92E-07	3.55E-09	2.57E-09	7.62E-06	1.90E-06	1.73E-07	1.10E-07
1.42E-08	1.77E-08	5.07E-08	7.74E-11	1.00E-05	1.52E-07	1.58E-08	1.10E-11
1.01E-24	1.02E-15	9.25E-27	4.89E-27	3.53E-14	3.10E-17	2.96E-18	2.73E-22
3.96E-15	1.95E-12	1.79E-16	2.53E-18	6.86E-06	1.27E-08	2.07E-13	3.37E-14
6.43E-15	2.69E-09	8.91E-14	5.48E-17	0.001769048	2.66E-08	1.20E-12	4.52E-14

This file contains the class probabilities for each image file and group.

sampleName.csv

Sample	Group	Relative	Count	Prob_threshold	Percent_used
flowcam_FCM.U0001.2016-01-21.300A4X.01	Asterionellopsis	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	C_curvisetus	5.882352941	2	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	C_danicus	8.823529412	3	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	C_socialis	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Ciliophora	5.882352941	2	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Dactyliolen	8.823529412	3	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Dytilum	8.823529412	3	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	G_flaccida	8.823529412	3	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	G_striata	5.882352941	2	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Gymnodinium	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Lauderia	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Leptocylindrus	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Odontella	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	P_globosa	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Pleuro_Gyrosigma	5.882352941	2	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Prorocentrum	14.70588235	5	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	PseudoNitzschia	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Rhizosolenia	11.76470588	4	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Thalassionema	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Thalassiosira	2.941176471	1	0.75	89.88476312

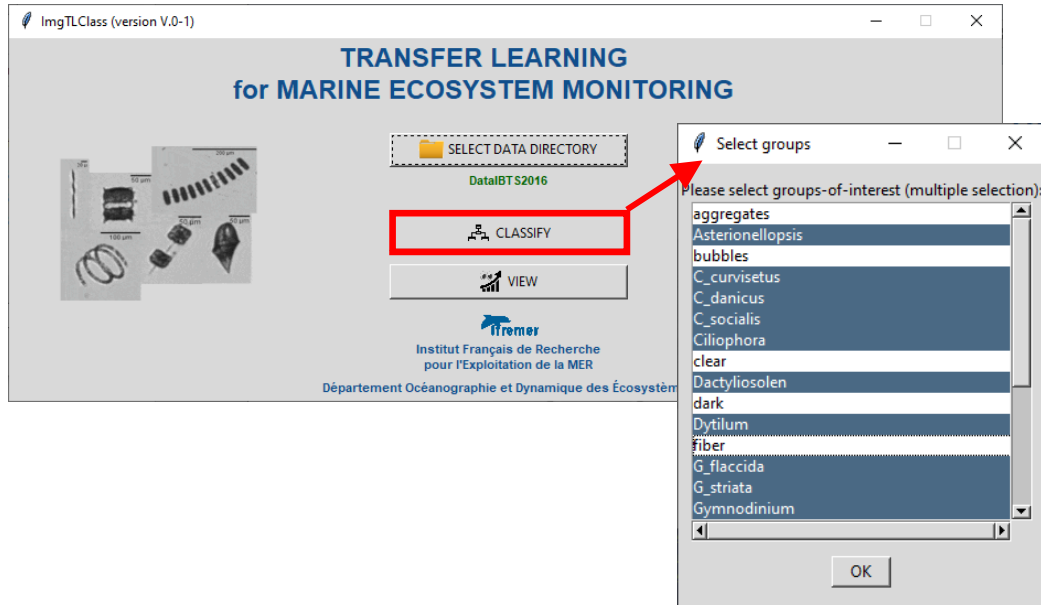
with Sample: the name of the sample
Group: list of all groups in the training set
Relative: relative abundance (in %) of each group
Count: absolute abundance of each group
Prob_threshold: threshold for class probability (see next section)
Percent_used: percentage of images with class probability upper than threshold

Classification probability threshold

The threshold value defined at this step allows to take into account only the images having a probability of “good” classification greater than this threshold. To take into account all the images, this value must be set to 0.

CLASSIFY button

To classify new images, click on the **CLASSIFY** button. A new window appears for group selection.



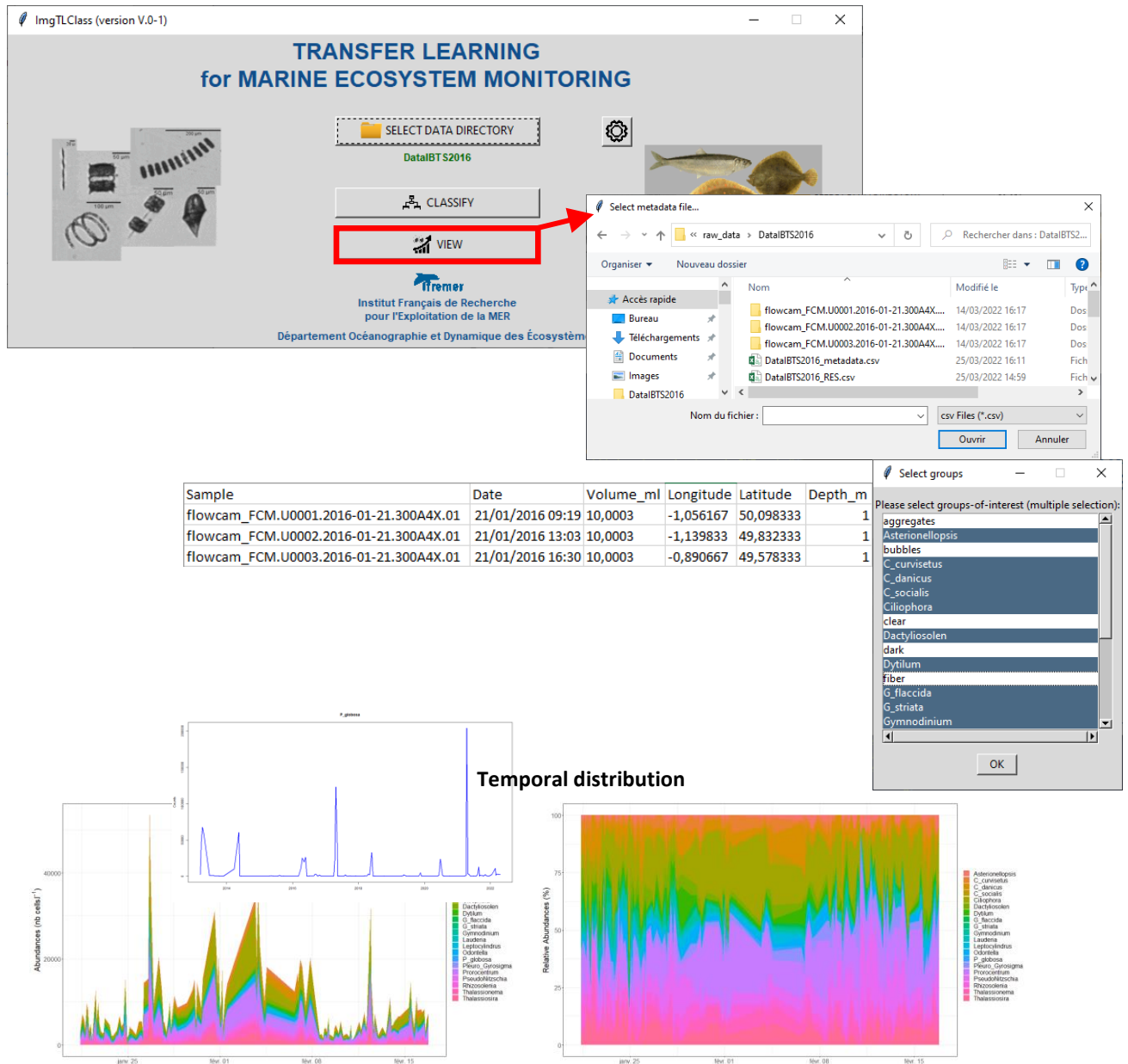
Sample	Group	Relative	Count	Prob_threshold	Percent_used
flowcam_FCM.U0001.2016-01-21.300A4X.01	Asterionellopsis	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	C_curvisetus	5.882352941	2	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	C_danicus	8.823529412	3	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	C_socialis	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Ciliophora	5.882352941	2	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Dactyliosolen	8.823529412	3	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Dytium	8.823529412	3	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	G_flaccida	8.823529412	3	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	G_striata	5.882352941	2	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Gymnodinium	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Lauderia	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Leptocylindrus	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Odontella	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	P_globosa	0	0	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Pleuro_Gyrosigma	5.882352941	2	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Prorocentrum	14.70588235	5	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	PseudoNitzschia	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Rhizosolenia	11.76470588	4	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Thalassionema	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0001.2016-01-21.300A4X.01	Thalassiostra	2.941176471	1	0.75	89.88476312
flowcam_FCM.U0002.2016-01-21.300A4X.01	Asterionellopsis	1.388888889	1	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	C_curvisetus	1.388888889	1	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	C_danicus	12.5	9	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	C_socialis	0	0	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Ciliophora	6.944444444	5	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Dactyliosolen	5.555555556	4	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Dytium	6.944444444	5	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	G_flaccida	4.166666667	3	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	G_striata	5.555555556	4	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Gymnodinium	5.555555556	4	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Lauderia	0	0	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Leptocylindrus	0	0	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Odontella	4.166666667	3	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	P_globosa	0	0	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Pleuro_Gyrosigma	4.166666667	3	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Prorocentrum	15.27777778	11	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	PseudoNitzschia	8.333333333	6	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Rhizosolenia	6.944444444	5	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Thalassionema	5.555555556	4	0.75	90.75812274
flowcam_FCM.U0002.2016-01-21.300A4X.01	Thalassiostra	5.555555556	4	0.75	90.75812274

with

- Sample: the name of the sample
- Group: list of all groups in the training set
- Relative: relative abundance (in %) of each group
- Count: absolute abundance of each group
- Prob_threshold: threshold for class probability (see next section)
- Percent_used: percentage of images with class probability upper than threshold

VIEW button

By clicking on this button, a new window appears for group selection.



**TRANSFER LEARNING
for MARINE ECOSYSTEM MONITORING**

SELECT DATA DIRECTORY
DatalBTS2016

CLASSIFY

VIEW

Institut Français de Recherche
pour l'Exploitation de la MER
Département Océanographie et Dynamique des Écosystèmes

Select metadata file...

raw_data > DatalBTS2016

Rechercher dans : DatalBTS2...

Organiser > Nouveau dossier

Accès rapide

Bureau

Téléchargements

Documents

Images

DatalBTS2016

Nom

Modifié le

Type

flowcam_FCM.U0001.2016-01-21.300A4X...

14/03/2022 16:17

Dos

flowcam_FCM.U0002.2016-01-21.300A4X...

14/03/2022 16:17

Dos

flowcam_FCM.U0003.2016-01-21.300A4X...

14/03/2022 16:17

Dos

DatalBTS2016_metadata.csv

25/03/2022 16:11

Fich

DatalBTS2016_RES.csv

25/03/2022 14:59

Fich

Nom du fichier :

csv Files (*.csv)

Ouvrir

Annuler

Sample	Date	Volume_ml	Longitude	Latitude	Depth_m
flowcam_FCM.U0001.2016-01-21.300A4X.01	21/01/2016 09:19	10,0003	-1,056167	50,098333	1
flowcam_FCM.U0002.2016-01-21.300A4X.01	21/01/2016 13:03	10,0003	-1,139833	49,832333	1
flowcam_FCM.U0003.2016-01-21.300A4X.01	21/01/2016 16:30	10,0003	-0,890667	49,578333	1

Select groups

Please select groups-of-interest (multiple selection):

aggregates

Asterionellopsis

bubbles

C. curvisetus

C. danicus

C. socialis

Ciliophora

clear

Dactyliosolen

dark

Dytilum

fiber

G. flaccida

G. striata

Gymnodinium

OK

Temporal distribution

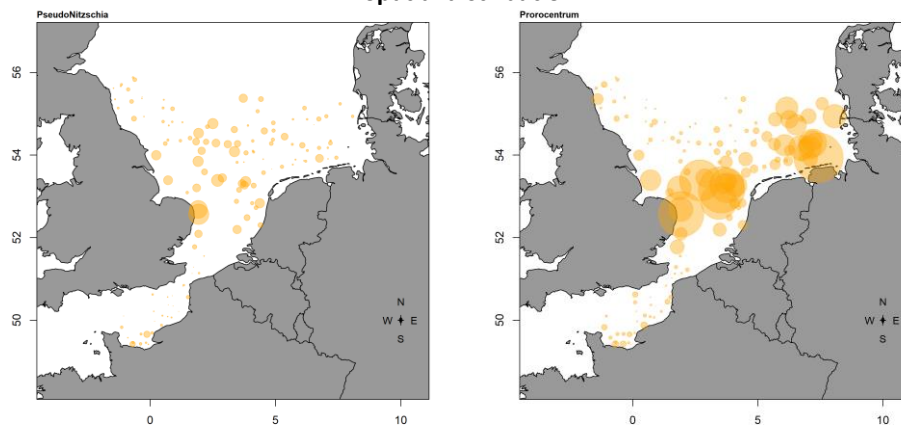
Abundances (nb cells⁻¹)

Relative Abundances (%)

Legend:

- Dactyliosolen
- Dytilum
- G. flaccida
- G. striata
- Gymnodinium
- Leptocylindrus
- Odontella
- P. globosa
- PseudoNitzschia
- Prorocentrum
- Rhodomonas
- Thalassiosira

Spatial distribution

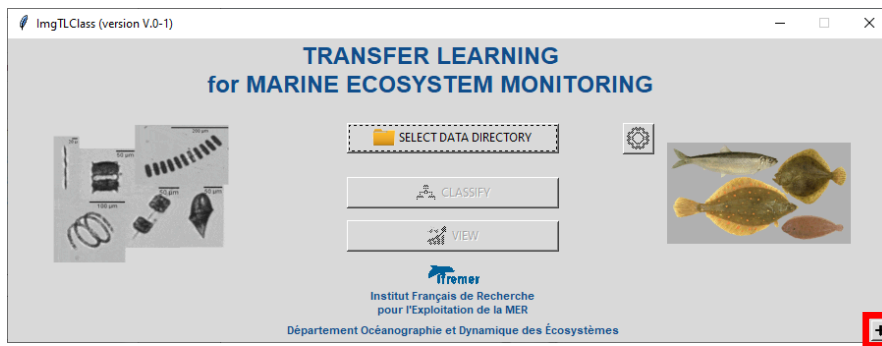


Thanks to the selection of a CSV file containing metadata on the samples (GPS coordinates, sampling date, imaged volume, ...), it can be possible to obtain different graphical outputs for temporal and spatial distribution.

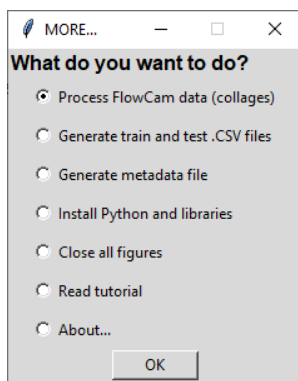
For temporal distribution, the absolute and relative counts are computed from the results file ('sampleName.csv', see 'OUTPUTS parameters' section), then plotted on graphs (a global graph with all groups, and individual graphs for each selected group). Concerning the spatial distribution, a mapping tool is proposed. Abundances of each selected group is projected on the map, thanks to the GPS coordinates provided in the metadata file (the diameter of the circles is depending on the obtained counts).

MORE... (+) button

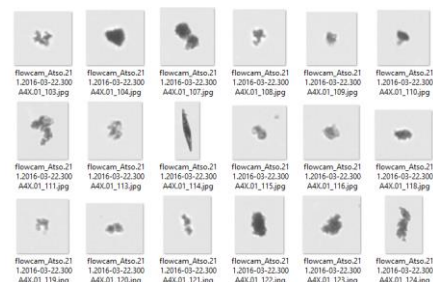
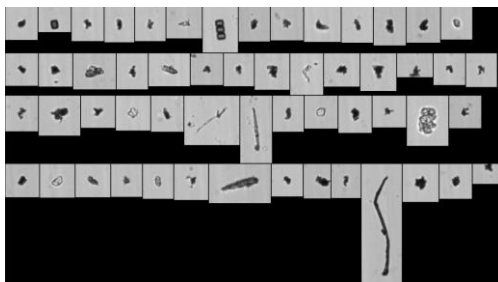
To help the user in the different data formatting steps for image analysis, several options are available. To view the list of these additional tools, click on the + button (**MORE...**, located at the bottom right of the main window).



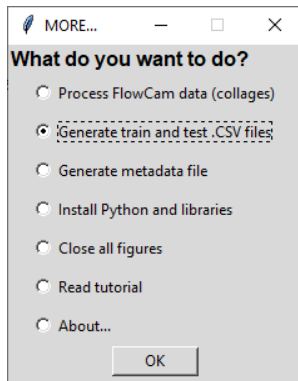
❖ Process FlowCam data (collages)



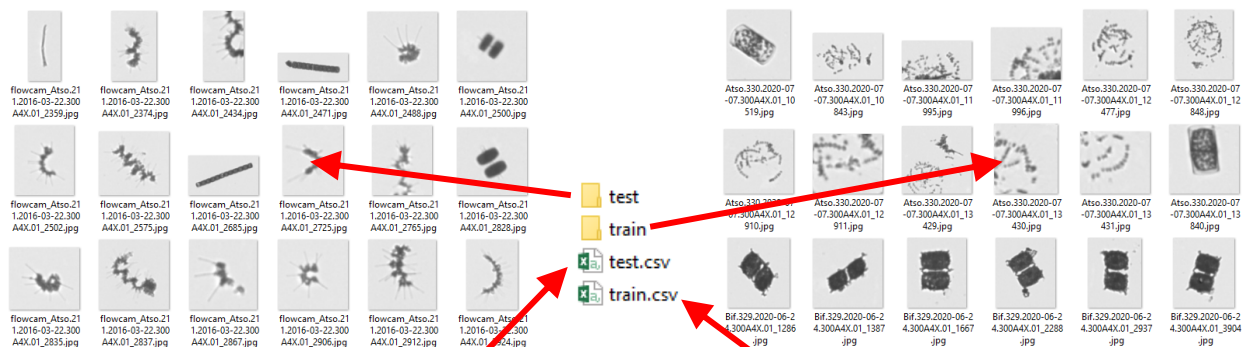
Raw data from the FlowCam is often presented as collages (one file with multiple images). It is then possible to cut and save vignettes (one image file per particle) from these collages. To do this, choose **Process FlowCam data (collages)**.



❖ Generate train and test .CSV files



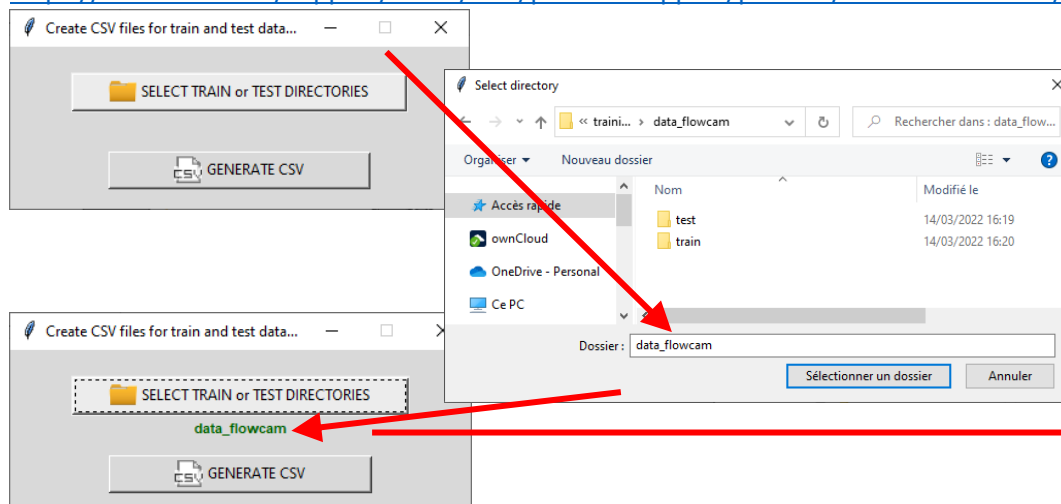
The images for training and test sets can be sorted into different subfolders (corresponding to groups). However, *Ecotranslearn* works with unique directories for **train** and **test**, in which images are put in bulk. To find out which group each image belongs to, a CSV file must be created. This is possible by selecting this option.



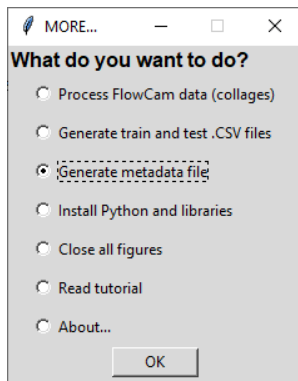
Label	Image
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1600.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1601.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1602.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1613.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1616.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1632.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1641.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1643.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1644.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1654.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1677.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1683.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1713.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1733.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1740.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1744.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1746.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1751.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1760.jpg"
aggregates	"flowcam_DK1.140.2014-03-27.300A4X.01_1788.jpg"

Label	Image
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_10.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_1041.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_1082.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_1113.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_1138.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_130.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_132.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_144.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_1512.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_154.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_1635.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_167.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_20.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_2065.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_2099.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_233.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_2449.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_2480.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_2485.jpg"
aggregates	"flowcam_ATSO.027.2013.03.14.300A4X.01_2532.jpg"

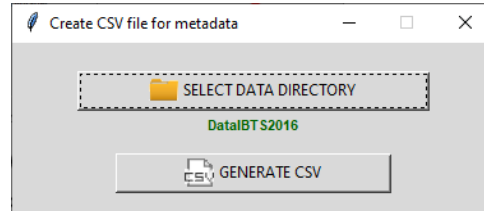
<https://www.dell.com/support/home/fr-fr/product-support/product/dell-wd19tb-dock/drivers>



❖ Generate metadata file



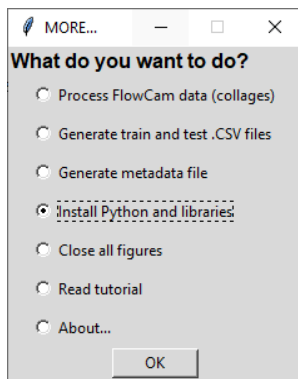
For the visualisation of classification results, a metadata file is required. This CSV file contains information on each sample, such as Date, Volume, GPS coordinates, and other metadata. A template can be generated thanks to this option.



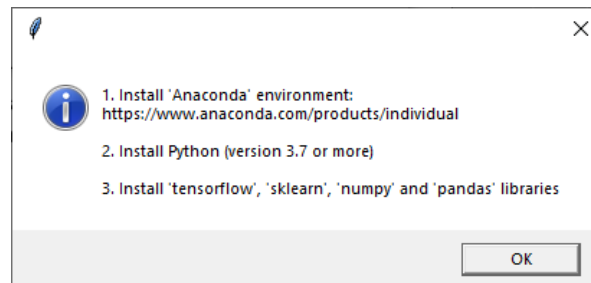
Sample	Date	Volume_ml	Longitude	Latitude	Depth_m
R1.2021-03-04.300A4X.01	2021-03-04 16:58:14	2,9994	NA	NA	1
R1.2021-03-17.300A4X.01	2021-03-17 14:55:44	5,0002	NA	NA	1
R1.2021-03-25.300A4X.02	2021-03-25 14:31:37	2,7657	NA	NA	1

with Sample: the sample name
 Date: the date of sampling (format: YYYY-mm-dd HH:MM:SS)
 Volume_ml: imaged volume (ml)
 Longitude: sample longitude (ex: -1,056167)
 Latitude: sample latitude (ex: 50,098333)
 Depth: sampling depth (m)

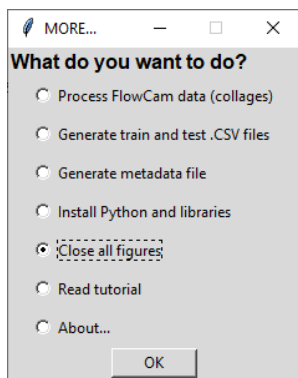
❖ Python and libraries requirements



This option allows having information on Python installation and the required libraries.



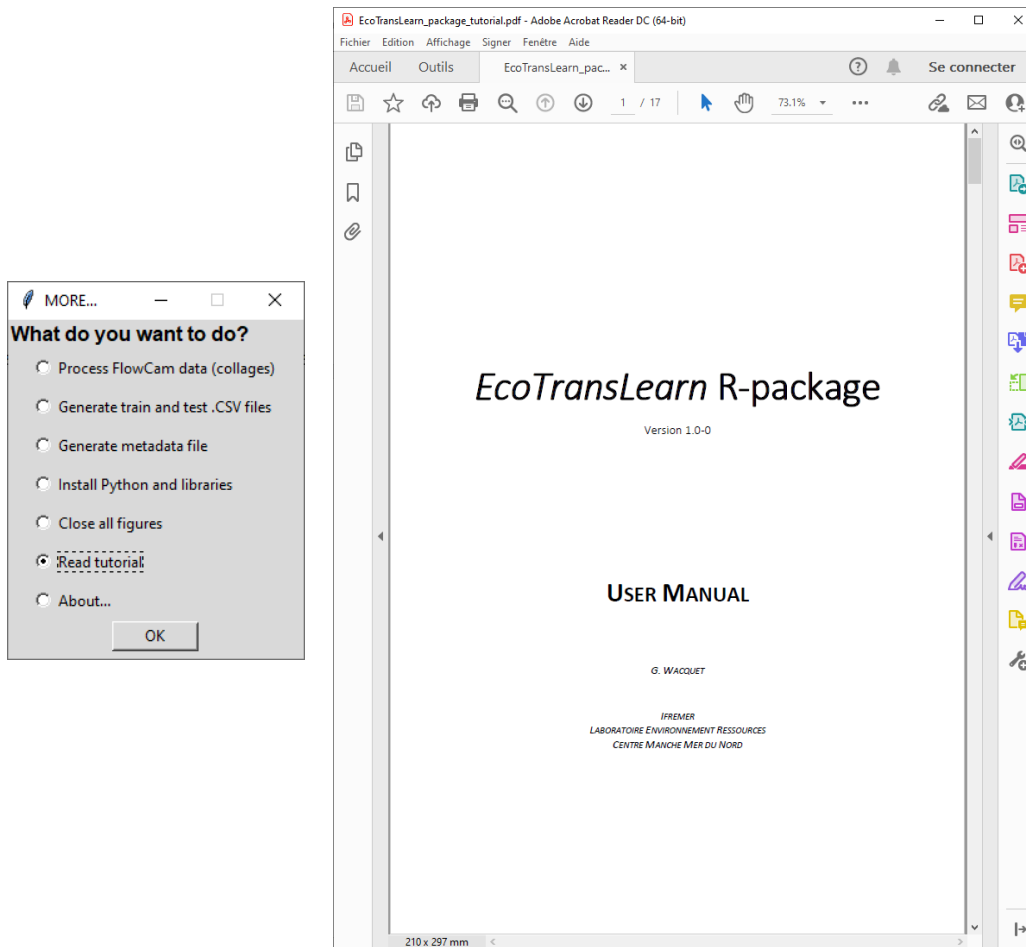
❖ Close all figures



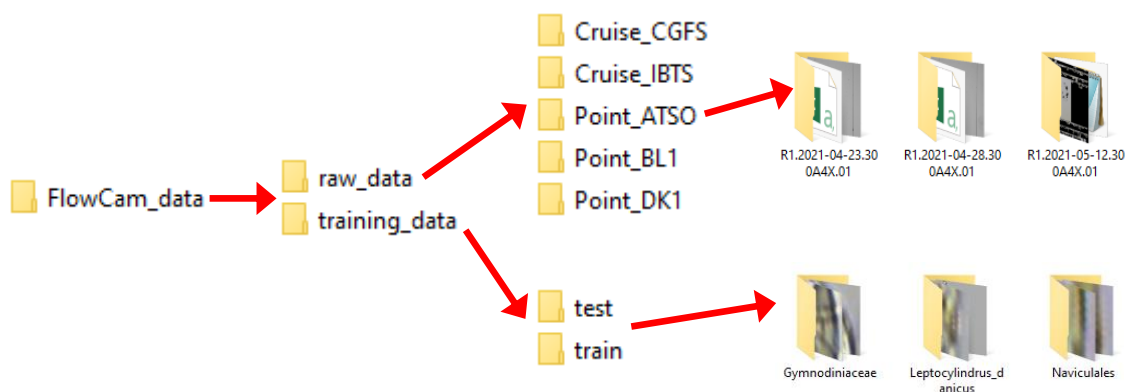
This option allows closing all figures.

❖ Read tutorial

A user manual is available (PDF file).



Recommended folder tree



For each acquisition device, it is recommended to create a dedicated folder, containing two sub-folders: one for the raw data to be processed (**'raw_data'**), and one for the training data (**'training_data'**).

'raw_data' should contain all samples with images to classify, or samples with collages files (for the FlowCam device).

'training_data' should contain a 'train' folder and a 'test' folder, where the images will be sorted into different subdirectories (which will be bulked up using the **Generate train and test .CSV files** option).