INSTRUCTIONS TO USE THE FPO\_CLASSIFICATION\_FUNCTION

Author: Sophie de Grissac - DioMed&a Science

COMPANY N° (SIRET): 882 532 245 00017

last update: 28/03/2020

email: s.degrissac@gmail.com

Contact me for any update, question, problem related to this function.

**Preliminary note**

It is assumed that the user has an updated version of the software R and know how to open a R session, a text editor, and run some lines.

The function is made so that the user only enters the system paths to the needed files and folders.

**Overview of the program**

This program takes as input a file with raw VMS locations of fishing vessels (only potters so far) and uses a Hidden Markov Model (HMM) and a speed filter to classify locations by potential activity, i.e. fishing or not fishing (transit) based on the step length and angles between successive locations. The model is asked to discriminate between 3 modes that are later interpreted as 1-Fishing, 2-Uncertain activity (slow transit or mixture of transit and fishing), 3-Transit. The choice of a 3 states model has been made based on tests.

The model is set as a conservative one so that it is more likely to underestimate fishing locations rather than find false positives.

Because the model needs perfectly regular GPS pings to converge properly, the VMS locations are interpolated. Therefore, the resulting file with VMS locations and modelled activity does not contain the **exact REAL VMS locations but a linear interpolation of these locations.**

**Limitations**

The program checks for various problems in the file that may impair the model results (number of locations per trip, ping irregularities, duplicated locations) and will remove from the analyse tracks that will compromise the function. A notification with the number of trips removed will appear on the screen.

Despite this automatic checking some exceptions may arise and the model may give unrealistic results for some vessels. For this reason, it is possible to visually check all results by quickly going through the plots provided by the program.

Without validation data the model cannot identify with 100% certainty the fishing location.

At this point, the program discriminates between different ping rate (a different model set in run according to each ping rate) but cannot give satisfying results with ping rate over 5min (typically 10 min). If vessels with ping rate min are included in the input file, they will be removed from analyse.

**FILES PROVIDED**

3 files are provided:

* **FPO\_classification\_function.R** is the classification program itself, you shouldn’t have to open it unless able to modify the code of the function if needed.
* **User\_Run\_FPO\_Classification.R** is the small script you will use to run the program.
* **harbourLoc.csv** is a file with harbour locations used by the program to filter VMS locations. You can add harbours in this file but do not change the name of the file.

**UDER DATA INPUT**

The user provides 1 file containing the VMS locations of potting vessels extracted from the database.

There are some restrictions and required formatting:

* You can put any vessels in the file but in the end, **only vessel trips with a VMS ping rate of 5 min or less (1 and 3 min) will be kept.** The model does not accommodate properly ping rate 10 min for now.
* File must be a CSV (coma separation) with the first 4 columns as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **vessel identity** | **Date + time** | **longitude** | **latitude** |

* Total number of columns and names of the columns is not important.
* **The format of the Date+Time MUST BE dd/dd/yyy hh:mm:ss.** If it isn’t you can change it in Excel (select the date column -> right clic -> “cell format -> Custom -> type *dd/dd/yyy hh:mm:ss -> save the file in .csv if it isn’t already*)

**STEPS TO FOLLOW**

1. Copy the files **User\_Run\_FPO\_Classification.R**, **FPO\_classification\_function.R** and **harbourLoc.csv** in a folder. Ideally you can copy the file with the VMS location to classify in the same folder (optional).
2. Open an R session and an R editor (or any text editor such as Notepad++)
3. Open the file **User\_Run\_FPO\_Classification.R**
4. In the file **User\_Run\_FPO\_Classification.R** follow the embedded instructions preceded by the sign **###**.

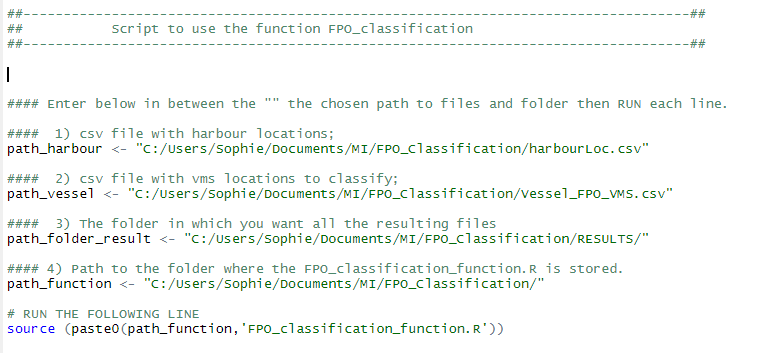
You have to enter the correct paths in between the “ “.

Orientation of the slash ( **/** ) is important. It has to be either:

* **1 forward slash** (e.g.: “C:/Users/Sophie/Documents/”)
* **2 backward slashes** (e.g.: “C:\\Users\\Sophie\\Documents\\”)

Note that paths 1 and 2 include the name of the file with extension .csv whereas paths 3 and 4 are only folders (end by a **/**).

Below is an example where I entered the paths to the different files and folders on my computer.



1. Once the paths have been properly entered, you can run all the lines (Ctrl-A to select everything then copy in the R session or Run from your R editor).

At this point if you’ve entered the paths correctly the program should run and analyse the VMS file provided. You will be notified of the progresses and when the task is done.

**DATA OUTPUT**

Within the folder you chose to store the results in you will find:

* 1 CSV file named **FPO\_vms\_resultClassification.csv** containing the VMS locations of all trips analysed along with new columns corresponding to the program outputs including the activity classification. (note: trips that were excluded from the analyse are also excluded from the resulting file).
* 1 folder for each ping rate containing:
  + 1 csv file with the results for this ping rate class only.
  + 1 R object containing the raw model output for checking if required (for advanced R users)
  + 1 folder per vessel containing:
    - 1 result csv file for this vessel
    - **1 folder with individual tracks plotted for fast visual checking**

**MODEL OUTPUTS INTERPRETATION in the resulting csv file**

|  |  |  |
| --- | --- | --- |
| **HMM3** | **HMM3\_SpeedFilter** | **Activity\_classification** |
| Raw output interpretation of the model by a Viterbi algorithm that calculated the most probable state/activity for each location: state 1, 2 or 3. | A small speed filter enables to recategorize (quite conservatively) some uncertain locations (HMM state 2) into transit locations (HMM state 3). | **State 1: FISHING**  High probability that this locations correspond to fishing events.  **State 2: MIX-TRANSIT**  Uncertain activity often corresponding to slow transit, relocations between fishing event or mixture of transit and fishing.  **State 3: TRANSIT**  High probability that this locations correspond to transit. |