Outil de comparaison du contenu des fichiers NetCDF mono-profils

Cette note concerne l’outil *Matlab nc\_compare\_mono\_profile\_files* utilisé pour comparer le contenu des fichiers NetCDF mono-profils.

[1. Outil de comparaison du contenu des fichiers NetCDF mono-profils 1](#_Toc383681427)

[1.1. Table des matières 1](#_Toc383681428)

[1.2. Description de la livraison 1](#_Toc383681429)

[1.3. Pré-requis 1](#_Toc383681430)

[1.4. Installation/configuration de l’outil 1](#_Toc383681431)

[1.5. Mise en oeuvre 1](#_Toc383681432)

[1.6. Sorties 2](#_Toc383681433)

# Description de la livraison

L’outil est composé du seul fichier *nc\_compare\_mono\_profile\_files.m*.

# Pré-requis

L’outil a été écrit et testé avec Matlab R2013b.

# Installation/configuration de l’outil

Aucune installation particulière n’est nécessaire.

L’outil se configure en renseignant les premières lignes du fichier *nc\_compare\_mono\_profile\_files.m* :

* DIR\_INPUT\_BASE\_NC\_FILES : répertoire du premier ensemble de fichiers à comparer,
* DIR\_INPUT\_NEW\_NC\_FILES : répertoire du second ensemble de fichiers à comparer,
* DIR\_LOG\_CSV\_FILE : répertoire de stockage des fichiers *.log* et *.csv* produits,
* FLOAT\_LIST\_FILE\_NAME : liste des numéros WMO des flotteurs à traiter par défaut,
* PRINT\_DIFF\_DATA\_FLAG : flag d’affichage, dans le log, des comparaisons des données des profils,
* FIRST\_PROF\_DATE\_TO\_STOP\_COMPARISON : restriction temporelle des profils comparés (mettre à vide (FIRST\_PROF\_DATE\_TO\_STOP\_COMPARISON= '') pour ne pas utiliser cette possibilité).

La structure des répertoires DIR\_INPUT\_BASE\_NC\_FILES et DIR\_INPUT\_NEW\_NC\_FILES doit être identique à celle du FTP Argo, à savoir :

* Un répertoire par flotteur (nommé du numéro WMO du flotteur),
* Pour chaque répertoire flotteur, un sous répertoire « *profiles* » contenant les fichiers NetCDF mono-profils.

# Mise en oeuvre

La mise en œuvre se fait en exécutant le programme *nc\_compare\_mono\_profile\_files* :

* Avec arguments (Ex : nc\_compare\_mono\_profile\_files(2901496, 6901027)) pour traiter les flotteurs dont les numéros WMO sont fournis en argument,
* Sans argument pour traiter tous les flotteurs de la liste FLOAT\_LIST\_FILE\_NAME.

La comparaison se fait entre un premier ensemble de fichiers appelé **Base** et un second appelé **New**.

L’association des profils Base aux profils New se fait sur les dates des profils (à défaut sur les dates des positions des profils).

Un profil non daté est ignoré.

Pour chaque profil Base, de date *ProfDateBase*, on recherche le profil New associé dont la date *ProfdateNew* est telle que :

*| ProfDateBase – ProfdateNew | ≤ 10 heures* (INTERVAL\_HOUR = 10).

En conséquence, il est souhaitable que l’ensemble Base soit celui dont les dates des profils sont les plus fiables (i.e. conformes aux spécifications du cookbook Traj). **Nous conseillons donc de mettre les mono-profils issus du décodeur Matlab en tant que Base**.

# Sorties

Le programme produit 2 fichiers : un fichier *log* et un fichier *csv*.

Le fichier *log* contient les messages d’information (type : INFO, WARNING ou ERROR) générés lors du traitement.

Lorsque le flag PRINT\_DIFF\_DATA\_FLAG est à 1, le fichier log contient également la comparaison des profils des 2 ensembles Base et New **lorsque ceux-ci sont différents**.

Le fichier *csv* contient les colonnes suivantes :

* Colonne A : le numéro de ligne,
* Colonne B : le numéro WMO du flotteur,
* Colonne C : la direction du profil (‘D’ ou ‘A’)
* Colonnes D à K : le résultat des comparaisons sur chaque élément comparé. Ce résultat peut valoir : 0 (éléments identiques), 1 (éléments différents) ou -1 (la comparaison n’a pas lieu d’être).

Les éléments comparés sont dans l’ordre :

* + Les numéros de cycles,
  + Les dates des profils,
  + Les localisations des profils (dates et positions),
  + Les modes des profils (‘R’, ‘A’ ou ‘D’),
  + Les découpages des profils (pour séparer les données pompées et non pompées),
  + Les nombres de niveaux des profils,
  + Les mesures des profils,
  + Les versions des formats des fichiers NetCDF mono-profils.
* Colonnes L à T et U à AC : les éléments du Profil Base (Là T) d’une part, du profil New (U à AC) d’autre part.

Les éléments fournis pour chaque profil sont dans l’ordre :

* + Le numéro de cycle du profil,
  + La date du profil,
  + La date de la localisation du profil,
  + La longitude du profil,
  + La latitude du profil,
  + Le mode du profil (‘R’, ‘A’ ou ‘D’),
  + Un flag de détection d’un profil découpé (pour séparer les données pompées et non pompées),
  + Le nombre de niveaux du profil,
  + La version du format du fichier NetCDF mono-profil.

Remarque :

La détection d’un profil découpé se fait par les critères suivants :

* N\_PROFF = 2,
* Les deux profils sont à la même date,
* Les deux profils ont la même liste de paramètres.

Attention :

* Ce critère ne garantit pas la détection d’un profil possiblement coupé (s’il n’y a pas de mesures entre la surface et PCutOff),
* Ce critère sera à reprendre pour les fichiers C (core) lorsque les règles de gestion de ces derniers auront été définies.