

Syntaxe pour les données *SAFRAN* hydro-climatiques à la station à intégrer dans le portail DRIAS

Ce document décrit le format attendu des fichiers NetCDF qui seront distribuées via le portail DRIAS. Ceci permet de normaliser les fichiers afin de réduire les efforts requis à leur utilisation. Le format souhaité, c'est-à-dire la structure du fichier et un nombre suffisant de métadonnées renseignées, suit la norme CMOR (*Climate Model Output Rewrit*) imposé de plus en plus dans les projets de modélisation internationaux et souhaite être conforme à la norme CF. Pour toutes questions n'hésitez pas à contacter la DCSC de Météo-France.

Voici les grandes caractéristiques :

- **Les noms de fichier** sont composés d'arguments décrivant la simulation et le projet.
- Chaque fichier contient une seule **variable** de sortie de simulation, en plus des **axes : temps et station**, qui se rapporte à la simulation d'un seul modèle et une seule expérience (scénario ou partie historique).
- Chaque fichier contient toute la **période** modélisée pour l'expérience (scénario ou partie historique) et couvre uniquement le **domaine** décrit par le projet.
- Les **unités et dimensions** s'accordent à ce qui est requis par les tables MIP (*model intercomparaison project*).
- Un certain nombre de **métadonnées** sont attendues. Leurs formats et pour certains leur contenu correspondent aux standards fournis par les tables de références. D'autres métadonnées peuvent être ajoutées pour compléter l'information. Tout comme les tables MIP peuvent être enrichies.
Les attributs globaux renseignent sur la réalisation de la simulation. Les attributs des coordonnées spécifient la grille, le calendrier et les unités à partir desquels la variable est défini. Enfin les attributs de la variable permettent de préciser son format, son unité, son type, mais aussi d'informer sur la méthode d'extraction de la valeur et le format des valeurs manquantes.

1. Nom du fichier :

Les éléments composant le nom du fichier fournissent rapidement des informations sur la simulation et s'écrit comme ce qui suit :

Variable_Domain_Reanalysis_HYDRO-Inst-Model_TimeFrequency_StartTime-EndTime_Suffix.nc

Le séparateur est un underscore '_'

- position 1 – Variable : *Nom de la variable (avec Adjust si les données sont corrigées), sera à la lettre près identique au nom de la variable du fichier.*

- 'debit'

- position 2 – Domain : *Couverture spatiale des données*

- 'France'
- 'Garonne'
- 'Loire'
- 'Seine'

- position 3 – Reanalysis : *Nom de la Réanalyse utilisée en entrée*

- 'SAFRAN-France-2022'

- position 4 – HYDRO-Inst-Model : *Identifiant du HYDRO = Institut-Modèle*

- 'MF-SIM2'
- 'IPSL-ORCHIDEE'
- 'INRAE-J2000'
- 'BRGM-Marthe'
- 'BRGM-EROS'
- 'BRGM-AquiFR'

- position 5 – TimeFrequency : *Le pas de temps du jeu de données*

- 'day'
- '1hr'

- position 6 – Startyear-Endyear: *Couverture temporelle des données sous forme YYYYMMDD-YYYYMMDD. Les fichiers doivent être en année hydro et couvrir la période 01/08/1976-31/07/2022*

- '19760801-20220731'

- position 7 – Suffix: **Exceptionnellement** – toute information permettant de distinguer des fichiers lorsque les éléments précédents ne le permettent pas. Exemple sur le calcul de l'ETP par des méthodes de calcul différentes.

- 'FAO'
- 'Hg0175'

Attention toutefois le nombre de caractères autorisés dans le nom des fichiers (comportant tout le chemin d'accès) peut être limité à 255 caractères sous Windows. Pour vérifier : `echo path/filename | wc -m`

quelques exemples :

debit_France_SAFRAN-France-2022_MF-ISBA-CTIP_day_19760801-20220731.nc

debit_France_SAFRAN-France-2022_EDF-MORDOR-SD_day_19760801-20220731.nc

2. Attributs globaux :

Les attributs globaux sont souvent récupérés des fichiers sources. Ils renseignent sur la réalisation de la simulation, de la descente d'échelle dynamique à la correction de biais, tous essentiels à la traçabilité. Il est demandé de recopier sans modifier les entrées. Les informations concernant la modélisation hydrologique : le modèle, l'institut, la date de mise en œuvre, référence, ... seront spécifiées par de nouveaux attributs avec le préfixe "hy_" (attributs repérés en vert).

Les attributs globaux attendus, en bleu concernant le couplage GCM/RCM, en orange la correction de biais atmosphérique et en vert pour la modélisation hydrologique :

- **projet_id**, identification du projet.

- :project_id = "EXPLORE2" ;
- project_id = "DRIAS-2020"

On mettra project_id = "DRIAS-2020" pour les couples de modèles de l'ensemble DRIAS-2020 ; pour les couples de modèles ajoutés dans le cadre du projet EXPLORE2 (dont tous ceux corrigés par la méthode CDF-t) on mettra :project_id = "EXPLORE2" ;

- **observation_id**, une chaîne de caractères de la base d'observation utilisée pour corriger les données.

- :observation = "SAFRAN-France and SAFRAN-Montagne Quintana-Segui P., Le Moigne P., Durand Y., Martin E., Habets F., Baillon M., Canellas C., Franchisteguy L., Morel S., 2008, Analysis of Near-Surface Atmospheric Variables : Validation of the SAFRAN Analysis over France, Journal of Applied Meteorology and Climatology, 47, 92-107.
<https://doi.org/10.1175/2007JAMC1636.1> "
- :observation_id = "SAFRAN-France-2022" ;

- **institute_id**, un nom court du centre de production de la réanalyse SAFRAN

- :institution = "Meteo-France" ;
- :institute_id = "MF" ;

- **frequency**, l'intervalle de temps d'échantillonnage de la série de données.

- :frequency = "day" ;

- **contact**, fournit l'adresse électronique de la personne responsable des données.

- :contact = "jean-michel.soubeyroux@meteo.fr" ;

- **creation_date**, la date à laquelle la réanalyse SAFRAN-France a été réalisée.

- :creation_date = "2022-10-13T14:25:48Z" ;

- **product**, une chaîne de caractères indiquant la méthodologie pour créer cet ensemble de données.

- :product = "hydro-climatique" ;

- **hy_projet_id**, identification du projet

- :hy_projet_id = "EXPLORE2"

- **hy_institute_id**, un nom d'identification court du centre de modélisation contribuant aux données.

- :hy_institute_id = "Meteo-France" ;

- **hy_model_id**, un acronyme qui identifie le modèle hydrologique

- :hy_model_id = "SIM2" ;

- **hy_version_id**,

- :hy_version_id = "V8F" ;

- **hy_creation_date**, la date à laquelle la simulation a été réalisée

- :hy_creation_date = "2023-01-20T17:53:28Z" ;

- **hy_contact**, fournit le nom ou l'adresse électronique de la personne responsable des données.

- :hy_contact = "driascontact@meteo.fr" ;

3. Les dimensions :

Les séries temporelles de chaque station ayant le même nombre d'instance et des valeurs temporelles identiques pour toutes les instances, la représentation en tableau bidimensionnel est idéale. Celle-ci comporte deux dimensions : celle du temps 'time' et celle des éléments ici appelé 'station'. De préférence ces dimensions apparaissent dans l'ordre suivant 'time', puis 'station'.

L'axe temporel est défini à partir d'une variable unidimensionnelle : time(time), la période couverte coïncide à celle annoncée par les métadonnées. Le nombre de valeur vérifie l'information sur la fréquence temporelle des données. Attention toute fois au type de calendrier et aux variations de début et de fin de période, des problèmes récurrents qui sont issus des modèles eux-mêmes. D'où l'importance d'une bonne documentation des attributs.

L'axe temporel doit être défini comme « UNLIMITED », c'est-à-dire de dimension 1 et sans restriction. Cela permet de pouvoir concaténer des fichiers NetCDF si besoin. Pour cela, il est possible de les générer sans dimension fixée (record) et de les « degenerate ». Une commande ncks existe pour cela :

```
ncks -O --mk_rec_dmn time in.nc out.nc # Change "time" to record dimension
```

Le temps doit toujours inclure explicitement l'attribut "units" ; il n'y a pas de valeur par défaut. **L'unité de temps** attendue est : "days since YYYY-MM-DD hh:mm:ss" ; où YYYY définit l'année, MM le mois, DD le jour, hh l'heure, mm les minutes et ss les secondes.

L'attribut "units" prend une valeur selon de codage suivant : "days since 1950-01-01 00:00:00" qui indique les jours depuis le 1er janvier 1950.

La chaîne date/heure de référence (qui apparaît après l'identifiant since) est obligatoire. Elle peut inclure la date seule, ou la date et l'heure, ou la date, l'heure et le fuseau horaire. Si le fuseau horaire est omis, la valeur par défaut est UTC, et si l'heure et le fuseau horaire sont omis, la valeur par défaut est 00:00:00 UTC.

Le choix du calendrier définit l'ensemble des dates (combinaisons année-mois-jour) qui sont autorisées. Il spécifie donc le nombre de jours entre deux dates quelconques. **Le calendrier de temps** attendu est : "**standard**"; c'est le calendrier par défaut le calendrier grégorien. Dans ce calendrier, les dates/heures sont dans le calendrier grégorien, dans lequel une année est bissextile si (i) elle est divisible par 4 mais pas par 100 ou (ii) elle est divisible par 400.

Naturellement les valeurs des dimensions sont croissantes et n'ont pas de valeur manquante.

Chaque élément 'station' est identifié à partir d'au moins 2 variables qui sont des chaînes de caractères de longueur fixe et définie :

- 'code(station, code_strlen)' un code unique issu d'une banque de données qui le rattache à un site hydrométrique bien identifiés.
- 'name(station, name_strlen)' un libellé de la station décrivant le cours d'eau, la ville la plus proche et un complément sur le site en question (ex : La Vienne à Limoges - Pont-Neuf).
- 'code_type(station, code_type_strlen)' un libellé identifiant le type de la station.
- 'network_origin(station, network_origin_strlen)' un libellé identifiant le réseau auquel se rattache de code d'identification.

Un exemple :

dimensions:

```
time = UNLIMITED ; // (34698 currently)
station = 143 ;
code_strlen = 8 ;
name_strlen = 64 ;
code_type_strlen = 5 ;
network_origin_strlen = 11 ;
```

variables:

```
double time(time) ;
    time:standard_name = "time" ;
    time:long_name = "time" ;
    time:units = "days since 1950-01-01 00:00:00" ;
    time:calendar = "standard" ;
    time:axis = "T" ;
```

```
char code(station, code_strlen)
    code:long_name = "code of station" ;
    c("AAAAAAA", "BBBBBBBB", "CCCCCCCC")
```

```
char name(station, name_strlen) ;
    name:long_name = "name of station" ;
    ex : c("La a sur a", "La b sur b", "La c sur c")
```

```
char code_type(station, code_type_strlen)
    code_type:long_name = "code of station" ;
    ex : c("SANDRE", "BSS", "MESO")
```

SANDRE : Service d'administration nation des données et référentiels sur l'eau (site hydrométrique)

BSS : Banque du sous sol

MESO : Masse d'eau souterraine

```
char network_origin(station, network_origin_strlen) ;
    station:long_name = "station" ;
```

ex : c("ONDE", "HYDRO", "RCS")

4. Les coordonnées station :

Les variables de coordonnées sont indispensables à la localisation des stations, elles sont donc unidimensionnelles fonction de l'élément 'station' :

- L93_X(station) et L93_Y(station)
- LII_X(station) et LII_Y(station)

On identifie la localisation des stations, suivant la norme des portails hydrologiques, par des coordonnées en Lambert 93. Pour assurer la succession d'utilisation des données des utilisateurs DRIAS, les coordonnées seront également à renseigner en Lambert II étendu, l'avantage c'est que cette dernière projection donne une grille régulière.

Les stations doivent être incluses dans la couverture spatiale annoncée dans les métadonnées et nom du fichier.

Les variables représentant la latitude ou la longitude doivent toujours inclure l'attribut units ; il n'y a pas de valeur par défaut. **L'attribut units** est une chaîne de caractères et l'unité attendue : **L93_X:units = "m" ; L93_Y:units = "m" ;**

Naturellement les valeurs des dimensions sont croissantes et n'ont pas de valeur manquante.

Les attributs des coordonnées attendus :

- **standard_name**, un nom d'identification court de la coordonnée.
- **long_name**, un nom d'identification long de la coordonnée.
- **units**, spécifie l'unité de la variable coordonnée.

Un exemple :

```
double L93_X(station) ;
  L93_X:standard_name = "X Lambert-93" ;
  L93_X:long_name = "horizontal coordinate in Lambert-93" ;
  L93_X:units = "m" ;
double L93_Y(station) ;
  L93_Y:standard_name = "Y Lambert-93" ;
  L93_Y:long_name = "vertical coordinate in Lambert-93" ;
  L93_Y:units = "m" ;
int L93 ;
  L93:standard_name = "Lambert-93" ;
  L93:long_name = "RGF93 / Lambert-93" ;
  L93:grid_mapping_name = "Lambert_Conformal_Conic_2SP" ;
  L93:standard_parallel_1 = "49" ;
  L93:standard_parallel_2 = "44" ;
  L93:latitude_of_origin = "46.5" ;
  L93:central_meridian = "3" ;
  L93:false_easting = "700000" ;
  L93:false_northing = "6600000" ;
  L93:EPSG = "2154" ;
  L93:references = "https://spatialreference.org/ref/epsg/2154/html/" ;

double LII_X(station) ;
  LII_X:standard_name = "X Lambert-II" ;
  LII_X:long_name = "horizontal coordinate in Lambert-II" ;
  LII_X:units = "m" ;
double LII_Y(station) ;
  LII_Y:standard_name = "Y Lambert-II" ;
  LII_Y:long_name = "vertical coordinate in Lambert-II" ;
  LII_Y:units = "m" ;
int LII ;
  LII:standard_name = "Lambert-II" ;
```

```
LII:long_name = "NTF (Paris) / Lambert zone II" ;
LII:grid_mapping_name = "Lambert_Conformal_Conic_1SP" ;
LII:latitude_of_origin = "52" ;
LII:central_meridian = "0" ;
LII:scale_factor = "0.99987742" ;
LII:false_easting = "600000" ;
LII:false_northing = "2200000" ;
LII:epsg = "27572" ;
LII:references = "https://spatialreference.org/ref/epsg/ntf-paris-lambert-zone-ii/html/" ;
```

Pour certain modèle, la localisation du point de station peut être assez différent par rapport à la réelle position de la station identifié par le code hydro. On peut compléter les coordonnées par le ou les éléments suivant :

- Lat(station) et Lon(station)
- L93_X_model(station) et L93_Y_model(station)
- LII_X_model(station) et LII_Y_model(station)

5. La surface :

Aux variables de coordonnées s'ajoutent aussi les variables caractérisant la surface topologique des bassins associés à chaque station.

Les stations doivent être incluses dans la couverture spatiale annoncée dans les métadonnées et nom du fichier.

Paramètres :

- standard_name : un nom d'identification court de la coordonnée
- long_name : un nom d'identification long de la coordonnée
- units : spécifie l'unité de la variable coordonnée

Un exemple :

```
double topologicalSurface(station) ;
topologicalSurface:long_name = "topological surface of the watershed" ;
topologicalSurface:units = "km2" ;
```

Les surfaces des bassins versant modélisés peuvent être parfois différentes par rapport aux observations. On invite donc à compléter par les éléments suivant :

```
double topologicalSurface_model(station) ;
topologicalSurface_model:long_name = "topological surface of the watershed in the model world" ;
topologicalSurface_model:units = "km2" ;
```

6. La variable :

Les dimensions de la variable sont au même nombre que les axes spécifiés précédemment et sont définies de la même manière (nom et grandeur). L'ordre des indices souhaité dépend du type des dimensions (voir ci-dessus) :

- var (time, station)

Le nom de la variable est typiquement un acronyme qui suit les références des tables MIP et auquel est associé un nom court ou long et une unité standard. Bien sûr il est identique au nom de la variable dans le nom du fichier.

Les attributs de la variable attendus :

- **standard_name**, un nom d'identification court de la variable.

- `debit:standard_name = "debit" ;`
- `debit:long_name = "debit_modcou" ;`

- **units**, spécifie l'unité de la variable coordonnée.

- `debit:units = "m3.s-1" ;`

- **missing_value** et **_Fillvalue**, spécifie comment sont identifiés respectivement les valeurs manquantes et les valeurs non valides.

- `debit:_FillValue = NaNf ;`
- `debit:missing_value = NaNf ;`

- **cell_methods**, fournit l'information concernant le calcul ou l'extraction de la variable.

- `debit:cell_methods = "time: mean" ;`

- **comment**, complément d'information sur l'extraction, le calcul de la variable

- `evspsblpotAdjust:comment = "Potential evapotranspiration calculated using the Hargreaves method with unique coefficient 0.175 from DRIAS-2020 corrected data set (the variable rsdsAdjust is not used). Source :"`

Un exemple :

variables:

```
float debit(time, station) ;
    debit:standard_name = "debit" ;
    debit:long_name = "debit_modcou" ;
    debit:units = "m3.s-1" ;
    debit:_FillValue = NaNf ;
    debit:missing_value = NaNf ;
    debit:cell_methods = "time:sum" ;
```

7. Cohérence croisée :

La mise en place d'un double niveau d'information (éléments du nom du fichier et métadonnées) nécessite de contrôler la cohérence entre les deux, mais est primordiale car contribue à la qualité du jeu de données. Tout comme la standardisation des unités et des noms est primordiale pour éviter les confusions et simplifier le traitement des données par les utilisateurs.

Exemple de variables hydrologiques :

Accronyme	standard name	long name	units	cell_methods
debit	debit	debit_modcou	m ³ .s ⁻¹	time:sum

CF Standard Name Table = <http://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html>

IPCC Standard Output from Coupled Ocean-Atmosphere GCMs = <https://pcmdi.llnl.gov/mips/cmip3/variableList.html>

CMIP5-CMOR-Tables = https://wcrp-cmip.github.io/WGCM_Infrastructure_Panel/cmip_and_mip_tables.html

Data Reference Syntax (DRS) for bias-adjusted CORDEX = http://is-enes-data.github.io/CORDEX_adjust_drs.pdf

Code R pour formater les fichiers NetCDF 1D pour DRIAS

https://github.com/super-lou/Ex2D_toolbox/tree/main/Rtools/help/DRIAS_export/DRIAS_export_1D

8. Compression :

Afin de réduire les volumes de données, une compression est effectuée sur les fichiers NetCDF ainsi produits. Nous utilisons la commande suivante :

```
nccopy -d5 -c time/1,x/,y/ fic\_IN.nc fic\_OUT.nc
```

Celle-ci permet de réduire d'environ 50 % la taille du fichier.

LES FICHIERS SAFRAN sont disponibles sous :

https://climatedata.umr-cnrm.fr/public/dcsc/projects/EXPLORE2_SAFRAN_H/