



PROCEDURA PER LA DIVISIONE E RIUNIONE DI FILE
NELL'AMBITO DEL PROGETTO MYOCEAN
REVISIONE 1

A. BUSSANI e G. NOTARSTEFANO

Approved by:

Dr. Paola Del Negro

Indice

Abstract.....	3
Pre requisiti.....	3
Contenuto.....	3
Estrarre dati dai file netcdf, matlab e uso delle relative funzioni di libreria C.....	4
Descrizione funzionalità funzione split.....	7
Descrizione funzionalità funzione unione.....	9
Bibliografia:.....	10

Abstract

Nell'ambito del progetto MyOcean, uno dei WP in cui è coinvolto l'OGS prevede il reprocessing dei dati di temperatura e salinità da piattaforme in-situ nel Mar Mediterraneo e Mar Nero. Il periodo considerato comprende gli ultimi 20 anni di dati (1990-2010).

I file in questione sono archiviati sui portali ufficiali dei due bacini che sono rispettivamente localizzati in Grecia (HCMR) ed in Bulgaria (IOBAS). I formati dei file sono NetCDF (convenzioni OceanSites 1.1) e sono strutturati in modo da avere racchiuse in un unico file tutti i dati misurati da ciascuna piattaforma nel proprio periodo di attività. Al fine di poter usare i suddetti file nella procedura di validazione sviluppata e utilizzata dall'OGS trimestralmente per il controllo di qualità ordinario dei dati in-situ (Notarstefano et al., 2011), si è reso necessario creare una procedura automatica di divisione dei file sia temporalmente che spazialmente

A tale proposito sono state create due funzioni in matlab che effettuano la divisione e l'unione.

Le funzioni sono state inserite all'interno del server Oceano nella cartella:

```
/storage/sire/work/drifter/matlab/
```

La funzione di **divisione** è la seguente:

```
split_netcdf_by_date([filename|dir],destpath)
```

La funzione di **unione** è la seguente:

```
union_netcdf_by_date([filename|dir],destpath)
```

L'utenza drifter è quella preposta per eventuali successivi cambiamenti all'interno delle funzioni.

Pre requisiti

Matlab 7.9.0.529 (R2009b) e successivi.

Contenuto

La divisione dei dati in modo temporale è stata fatta dividendo ciascuna singola piattaforma in tanti file quanti sono i mesi nei quali la piattaforma ha fornito i dati.

All'interno di ciascun file netcdf (sia mensili che senza limiti di tempo), esistono all'interno degli attributi globali che contengono i limiti spaziali/temporali corrispondenti:

```
geospatial_lat_min
```

geospatial_lat_max

geospatial_lon_min

geospatial_lon_max

ed inoltre vengono inclusi limiti temporali

time_coverage_start

time_coverage_end

La funzione di unione va a recuperare i file mensili di ciascuno strumento e riunisce i dati, andando a modificare gli attributi globali in maniera opportuna.

Gli attributi che vengono modificati (all'interno dei limiti temporali mensili) sono:

geospatial_lat_min

geospatial_lat_max

geospatial_lon_min

geospatial_lon_max

ed inoltre vengono inclusi limiti temporali aggiornati per:

time_coverage_start

time_coverage_end

Estrarre dati dai file netcdf, matlab e uso delle relative funzioni di libreria C

Per la lettura dei file è stata creata una funzione matlab, ncread.

```
test=ncread('/storage/sire/dati/drifter/MyOcean_History/mooring/MO_TS_MO_IF000232.nc')
test =
  Attributes: [1x1 struct]
    DEPH: [2x382 single]
    DEPH_DM: [2x382 char]
    DEPH_QC: [2x382 int8]
    DEPTH: [2x1 single]
    DEPTH_QC: [2x1 int8]
    LATITUDE: [382x1 single]
    LONGITUDE: [382x1 single]
    POSITION_QC: [382x1 int8]
    PSAL: [2x382 single]
    PSAL_DM: [2x382 char]
    PSAL_QC: [2x382 int8]
    TEMP: [2x382 single]
    TEMP_DM: [2x382 char]
    TEMP_QC: [2x382 int8]
    TIME: [382x1 double]
    TIME_QC: [382x1 int8]
```

```
>> test.Attributes
```

```
ans =
```

```
Global: [1x1 struct]
  DEPH: [1x1 struct]
  DEPH_DM: [1x1 struct]
  DEPH_QC: [1x1 struct]
  DEPTH: [1x1 struct]
  DEPTH_QC: [1x1 struct]
  LATITUDE: [1x1 struct]
  LONGITUDE: [1x1 struct]
  POSITION_QC: [1x1 struct]
  PSAL: [1x1 struct]
  PSAL_DM: [1x1 struct]
  PSAL_QC: [1x1 struct]
  TEMP: [1x1 struct]
  TEMP_DM: [1x1 struct]
  TEMP_QC: [1x1 struct]
  TIME: [1x1 struct]
  TIME_QC: [1x1 struct]
```

```
>> test.Attributes.Global
```

```
ans =
```

```
comment: ' '
distribution_statement: [1x299 char]
  author: 'codac'
geospatial_vertical_max: '3'
  citation: [1x125 char]
  conventions: 'OceanSITES Manual 1.1, CF-1.1'
quality_control_indicator: '6'
  references: [1x70 char]
data_assembly_center: 'EUROSITES'
  wmo_platform_code: 'IF000232'
  id: 'MO_TS_MO_IF000232'
naming_authority: 'OceanSITES'
  qc_manual: 'OceanSITES User's Manual v1.1'
institution_references: 'http://www.nocs.uk'
geospatial_lat_max: '36.2573'
  format_version: '1.1'
  area: 'North Atlantic Ocean'
netcdf_version: '3.5'
  date_update: '2013-01-29T10:27:31Z'
  source: 'BUOY/MOORING: SURFACE, MOORED : observation'
platform_code: ' '
update_interval: 'daily'
  data_type: 'OceanSITES time-series data'
geospatial_lat_min: '36.2524'
time_coverage_start: '2010-04-08T00:00:00Z'
  data_mode: 'R'
  institution: 'HELLENIC CENTER FOR MARINE RESEARCH (HCMR)'
geospatial_lon_max: '25.5054'
geospatial_lon_min: '25.5005'
  time_coverage_end: '2010-05-26T18:00:00Z'
summary: ' '
```

```
cdm_data_type: 'Time-series'
contact: 'codac@noacs.uk'
quality_index: 'A'
geospatial_vertical_min: '0'
title: ' '
pi_name: ' '
site_code: ' '
history: [1x117 char]
NCO: '4.0.5'
```

In matlab esistono delle funzioni di base relative all'estrapolazione dei dati dei file netcdf.

```
help netcdf
```

NETCDF Summary of MATLAB NETCDF capabilities.

MATLAB provides low-level access to netCDF files via direct access to more than 30 functions in the netCDF library. To use these MATLAB functions, you must be familiar with the netCDF C interface. The "NetCDF C Interface Guide" for version 3.6.2 may be consulted at http://www.unidata.ucar.edu/software/netcdf/old_docs/docs_3_6_2/.

In most cases, the syntax of the MATLAB function is similar to the syntax of the netCDF library function. The functions are implemented as a package called "netcdf". To use these functions, one needs to prefix the function name with package name "netcdf", i.e.

```
ncid = netcdf.open ( ncfile, mode );
```

The following table lists all the netCDF library functions supported by the netCDF package.

abort	- Revert recent netCDF file definitions.
close	- Close netCDF file.
create	- Create new netCDF file.
endDef	- End netCDF file define mode.
inq	- Return information about netCDF file.
inqLibVers	- Return netCDF library version information.
open	- Open netCDF file.
reDef	- Set netCDF file into define mode.
setDefaultFormat	- Change default netCDF file format.
setFill	- Set netCDF fill mode.
sync	- Synchronize netCDF dataset to disk.
defDim	- Create netCDF dimension.
inqDim	- Return netCDF dimension name and length.
inqDimID	- Return dimension ID.
renameDim	- Change name of netCDF dimension.
defVar	- Create netCDF variable.
getVar	- Return data from netCDF variable.
inqVar	- Return information about variable.
inqVarID	- Return ID associated with variable name.
putVar	- Write data to netCDF variable.
renameVar	- Change name of netCDF variable.
copyAtt	- Copy attribute to new location.
delAtt	- Delete netCDF attribute.
getAtt	- Return netCDF attribute.
inqAtt	- Return information about netCDF attribute.

```
inqAttID      - Return ID of netCDF attribute.
inqAttName    - Return name of netCDF attribute.
putAtt        - Write netCDF attribute.
renameAtt     - Change name of attribute.
```

The following functions have no equivalents in the netCDF library.

```
getConstantNames - Return list of constants known to netCDF library.
getConstant      - Return numeric value of named constant
```

Please read the files netcdfcopyright.txt and mexncccopyright.txt for more information.

Contents of netcdf:

```
createnetcdf_new_DEPH          - function createnetcdf(M)
createYesterdayNetCDF         - M is a structure as follows
createnetcdf                   - function createnetcdf(M)
createnetcdf_new_DEPH_POSITIONqc - function createnetcdf(M)
ncgetExtremes                  - function M=ncread(filename)
aggregate_mission              - new_C = C;
ncread                          - function M=ncread(filename)
createnetcdf_new               - function createnetcdf(M)
reinterp3hMilena                - chiamata da "createMonthlyNetCDF" reinterpolata
ogni 3 ore i dati del
SplitTimeSeriesAndGenerateNetCDF - function SplitTimeSeriesAndGenerateNetCDF(M)
ncread_coriolis                 - function M=ncread_coriolis(filename)
createnetcdf_delayedmode        - function createnetcdf(M)
createMonthlyNetCDF             - function createMonthlyNetCDF(M)
createEachMonthNetCDF           - function createMonthlyNetCDF(M)
```

Descrizione funzionalità funzione split

La funzione di split (split_netcdf_by_date.m) prevede 2 parametri:

Il **primo** parametro puo' essere un singolo file, un file con caratteri "jolly" (*), oppure una cartella. Il programma elaborerà e dividerà ciascun file che riuscirà a trovare dal parametro inserito.

Esempi riguardanti il primo parametro:

- '/storage/sire/dati/drifter/MyOcean_History/mooring/MO_TS_MO_EXHC0020.nc'
Elaborazione del singolo file.
- '/storage/sire/dati/drifter/MyOcean_History/mooring/'
Elaborazione di tutti i file .nc che sono contenuti nella cartella
- '/storage/sire/dati/drifter/MyOcean_History/mooring/MO_TS_*.nc'
Elaborazione di tutti i file nella cartella fornita che iniziano per MO_TS_ e finiscono per .nc

Il **secondo** parametro indica la directory di destinazione.

Esempi:

- './"
Salvataggio nella directory corrente.
- '/storage/sire/dati/drifter/MyOcean_History_month/mooring/'
Salvataggio nella directory relativa.

Nella directory di destinazione verranno creati tanti file quanti i mesi nei quali la piattaforma ha dati.

Quindi per MO_TS_MO_EXHC0020.nc avremo i file MO_YYYYMM_TS_MO_EXHC0020.nc, l'anno e il mese sono nel formato YYYYMM.

es:

```
MO_201003_TS_MO_EXHC0020.nc
MO_201004_TS_MO_EXHC0020.nc
MO_201005_TS_MO_EXHC0020.nc
MO_201006_TS_MO_EXHC0020.nc
```

I file prodotti hanno solo i dati relativi al mese di riferimento, le variabili globali sono relative ai dati nel mese, vengono considerati in questo caso solo i dati che hanno un quality flag 1. Il limite controllato dalla funzione è solo quello di tempo. Se il dato non è corretto viene considerato il dato corretto di tempo precedente, in modo da non perderne la consequenzialità temporale e quindi includerlo comunque nel file mensile (anche se il dato di tempo non rientra nei parametri del mese).

Ecco la sezione di codice relativa a questo dettaglio:

```
%le date con QC~=1 devono essere sostituite con la data precedente (per
%il conteggio dei mesi)
indici_dati_errati=find(data_TQC~=1);
if isempty(indici_dati_errati)
    dataTIME_valid=data;
else
    %identifica i vari "blocchi" di date errate
    blocchi_errati=diff(indici_dati_errati);
    inizio_blocchi=find(blocchi_errati~=1);
    inizio_blocchi=[1; inizio_blocchi+1; numel(indici_dati_errati)+1];
    numero_inizio_blocchi=numel(inizio_blocchi);
    for i_num_blocchi=1:numero_inizio_blocchi-1
        indice_inizio_blocco=indici_dati_errati(inizio_blocchi(i_num_blocchi
));
        indice_fine_blocco=indici_dati_errati(inizio_blocchi(i_num_blocchi+1
))-1);
        %primo_dato_errato_nel_blocco=data(indice_inizio_blocco); %per debug
        %ultimo_dato_errato_nel_blocco=data(indice_fine_blocco); %per debug
        if (indice_inizio_blocco==1)
            %si sceglie il valore seguente!
            dato_corretto=data(indice_fine_blocco+1);
        else
            %si sceglie il valore precedente
            dato_corretto=data(indice_inizio_blocco-1);
        end
        %riempio quel singolo blocco (nuova array) con i dati corretti
        dataTIME_valid(indice_inizio_blocco:indice_fine_blocco)=dato_corrett
o;
    end
end
```

Descrizione funzionalità funzione unione

La funzione di `union_netcdf_by_date.m` prevede 2 parametri:

Il **primo** parametro puo' essere un singolo file, un file con caratteri "jolly" (*), oppure una cartella.

Il programma elaborerà e unirà ciascun file mensile che riuscirà a trovare dal parametro inserito.

Esempi riguardanti il primo parametro:

- `'/storage/sire/dati/drifter/MyOcean_History_month/mooring/'`
Elaborazione di tutti i file `.nc` che sono contenuti nella cartella
- `'/storage/sire/dati/drifter/MyOcean_History_month/mooring/MO_*_TS_MO_61417.nc'`,
Elaborazione di tutti i file nella cartella fornita che iniziano per `MO_` e finiscono per `_TS_MO_61417.nc`

Il **secondo** parametro indica la directory di destinazione.

Esempi:

- `'/'`
Salvataggio nella directory corrente.
- `'/storage/sire/dati/drifter/MyOcean_History_month/mooring/'`
Salvataggio nella directory relativa.

Nella directory di destinazione verrà creato un file per ciascuna piattaforma senza le date.

Nel caso del parametro `MO_*_TS_MO_EXHC0020.nc`, se i file presenti sono questi:

`MO_201003_TS_MO_EXHC0020.nc`
`MO_201004_TS_MO_EXHC0020.nc`
`MO_201005_TS_MO_EXHC0020.nc`
`MO_201006_TS_MO_EXHC0020.nc`

verrà creato il file

`MO_TS_MO_EXHC0020.nc`

I file prodotti hanno solo l'unione dei dati dei vari mesi, e le variabili globali sono relative ai dati completi. I limiti controllati dalla funzione (ed immesse nelle variabili globali) sono: relativi alle variabili di longitudine, latitudine, attraverso la variabile globale `position_qc`, la variabile di tempo attraverso la variabile `time_qc`.

Lista variabili globali:

`geospatial_lat_max`

`geospatial_lat_min`

`geospatial_lon_max`

`geospatial_lon_min`

`time_coverage_start`

time_coverage_end

```
%POSITION_QC
position_qc=dati_netcdf_var(ncid_out,'POSITION_QC');
data_all_qc=identify_good_array(position_qc);
union_all=dati_netcdf_var(ncid_out,'LATITUDE');
geospatial_lat_max=max(union_all(data_all_qc));
geospatial_lat_min=min(union_all(data_all_qc));
union_all=dati_netcdf_var(ncid_out,'LONGITUDE');
geospatial_lon_max=max(union_all(data_all_qc));
geospatial_lon_min=min(union_all(data_all_qc));
%end POSITION_QC
clear position_qc;

%TIME
time_qc=dati_netcdf_var(ncid_out,'TIME_QC');
data_all_qc=identify_good_array(time_qc);
union_all=dati_netcdf_var(ncid_out,'TIME');
%dataumanomatlab=dataTIME_valid+datenum('01/01/1950');
time_coverage_start=data_oceansite(datestr(min(union_all(data_all_qc))
+datenum('01/01/1950'),31));
time_coverage_end=data_oceansite(datestr(max(union_all(data_all_qc))
+datenum('01/01/1950'),31));
```

Vengono considerati corretti i dati aventi quality control: 1, 2, 5, 8.

A tale scopo è stata creata la sub function all'interno del file union_netcdf_by_date.m

```
function aaa=identify_good_array(array_QF)
    aaa=(array_QF==1 | array_QF==2 | array_QF==5 | array_QF==8);
end
```

Bibliografia:

http://www.oceansites.org/docs/oceansites_user_manual_version1.2.pdf

Notarstefano G., Bussani A. and Bolzon G. (2011). Assessment of temperature and salinity data obtained from in-situ platforms in the Mediterranean Sea (2011) 2011/67 OGA 27 SIRE dd. 3/8/2011