

Aumento risoluzione

Michy Alice

November 19, 2016

Introduzione: caricamento pacchetto.

Il seguente script funziona con qchlorophyll versione 1.1 o superiore.

Prima di installare qchlorophyll è necessario installare i seguenti pacchetti:

1. sp
2. gstat
3. stringi

```
#install.packages("/home/qchlorophyll_1.1.tar.gz", repos = NULL, type = "source")  
require(qchlorophyll)
```

Nota preliminare! Attenzione!

Siccome il procedimento è molto oneroso dal punto di vista della memoria RAM, consiglio di lavorare su un tipo di file per volta e di non lavorare su applicazioni terze mentre R sta elaborando i dati.

Esempio: inizio a lavorare sulla variabile qnet con i dati giornalieri, carico i dati, effettuo l'interpolazione e poi salvo il dataframe ottenuto in formato .csv (o altro). Pulisco la memoria con **rm(list=ls())** e procedo con la prossima variabile.

Per una macchina con almeno 8GB di memoria RAM, non è strettamente necessario effettuare quanto richiesto sopra, però il seguire questa procedura “garantisce” che le operazioni vengano portate a termine e che la macchina “non si incanti”.

Ulteriore consiglio: quando si aumenta la risoluzione, non fissare una risoluzione troppo elevata (ossia non fissare un parametro **step** troppo basso).

Caricamento dati.

Di seguito alcune indicazioni sul caricamento dati. L'output delle funzioni di caricamento dati è una lista con un dataframe per ogni anno.

- Le funzioni per il caricamento dati ricavano l'anno di riferimento delle osservazioni dal nome del file (il nome del file deve quindi contenere un numero di 4 cifre) e, se non è specificato diversamente (vedere sotto), assumono come data iniziale il primo gennaio dell'anno di riferimento.

Dati giornalieri

Selezionare la directory contenente i file .nc che si desidera caricare

```
setwd("/home/mich/Dropbox/cartella_condivisa_linux/quantide/packages_R/qchlorophyll_/dati/nuovo_lavoro/1
```

Utilizzare la funzione `load_nc_to_resize` senza specificare nessuna data per caricare tutti i file.

- E' necessario specificare la variabile da estrarre.
- E' possibile estrarre anche solo alcuni anni specificando gli argomenti *from* e *to*.
- Il ritaglio dell'area desiderata avviene immediatamente prima del caricamento allo scopo di risparmiare memoria. Specificare l'area geografica da ritagliare attraverso i parametri *lower_left_lat_lon* e *upper_right_lat_lon*
- Di default, la variabile tempo contenuta nel file .nc è chiamata "time" (argomento *time_variable*)
- Verificare gli altri valori di default della funzione con `?load_nc_to_resize`.

```
path <- getwd()

# Estrazione di tutti i dati giornalieri dal 1998 al 2009
dati_giornalieri <- load_nc_to_resize(path = path, variables = c("qnet"))

# Estrazione dei dati giornalieri dal 2000 al 2009 (inclusi)
dati_giornalieri <- load_nc_to_resize(path = path,
                                     from = 2000,
                                     to = 2009,
                                     variables = c("qnet"),
                                     lower_left_lat_lon = c(52, -65),
                                     upper_right_lat_lon = c(67, -42))
```

Dati mensili

Per i dati mensili è necessario specificare la frequenza mensile fissando l'argomento *monthly* al valore *TRUE*.

```
setwd("/home/mich/Dropbox/cartella_condivisa_linux/quantide/packages_R/qchlorophyll_/dati/nuovo_lavoro/1
path <- getwd()

# Estrazione di tutti i dati mensili
dati_mensili <- load_nc_to_resize(path = path,
                                  variables = c("qnet"),
                                  lower_left_lat_lon = c(52, -65),
                                  upper_right_lat_lon = c(67, -42),
                                  monthly = TRUE)

# Estrazione dei dati mensili dal 2000 al 2009 (inclusi)
dati_mensili <- load_nc_to_resize(path = path,
                                  from = 2000,
                                  to = 2009,
                                  variables = c("qnet"),
                                  lower_left_lat_lon = c(52, -65),
                                  upper_right_lat_lon = c(67, -42),
                                  monthly = TRUE)
```

Dati con la data in formato "ore da data di riferimento"

Per i file `shftl.nc` (dati con frequenza giornaliera) la data è fornita nel formato "ore trascorse da una data di riferimento". La data di riferimento in questione è 1800-01-01 (formato: yyyy-mm-dd). La data di riferimento

va specificata nella funzione `load_nc_to_resize` come segue:

```
setwd("/home/mich/Dropbox/cartella_condivisa_linux/quantide/packages_R/qchlorophyll_/dati/nuovo_lavoro/1")
path <- getwd()

# Caricamento di tutti i dati giornalieri shftl
dati_giornalieri_shftl <- load_nc_to_resize(path = path,
                                           variables = c("shftl"),
                                           lower_left_lat_lon = c(52, -65),
                                           upper_right_lat_lon = c(67, -42),
                                           date_origin = "1800-01-01")

# ATTENZIONE: specificare la date_origin SOLO se è necessario e si è sicuri che il formato della data n
```

Nota 1: Prestare attenzione alla directory di lavoro. La funzione di caricamento è abbastanza generale da non mostrare errori (oppure mostrare solo avvisi) nel caso in cui si carichino dati con frequenza mensile ma senza specificarlo, oppure non si specifichi la data di riferimento qualora sia necessario. Chiaramente in tali casi l'output non sarà affidabile.

Interpolazione (aumento risoluzione dei file)

Di seguito l'aumento di risoluzione sui dati caricati.

L'aumento di risoluzione viene effettuato con l'algoritmo [inverse distance weighting](#) in modo da garantire che i valori stimati siano entro il massimo e il minimo dei valori effettivi noti.

Il procedimento non dipende dalla frequenza delle osservazioni per cui è identico sia che si tratti di dati mensili che annuali. Si rende però necessario specificare il nome della variabile da interpolare (qnet oppure shftl negli esempi).

Il dataframe di riferimento deve contenere almeno le variabili longitudine e latitudine (lon e lat). Il parametro **step** determina il livello di risoluzione. Si assume che longitudine e latitudine siano equispaziate di un valore pari al parametro **step**. Vedere la variabile **grd** per un esempio. Consiglio di non fissare **step** a valori troppo bassi (ossia non fissare una risoluzione troppo elevata). L'interpolazione è effettuata giorno per giorno e nonostante sia relativamente veloce, il procedimento è abbastanza oneroso in termini di tempo proprio per la necessità di lavorare su dati giornalieri ad ogni interpolazione (le interpolazioni effettuate sono 365 per il numero di file caricati). Per i dati mensili i tempi sono notevolmente più corti siccome le interpolazioni da effettuare sono solo 12 per il numero di file caricati.

La funzione `interpolate_grid` restituisce un dataframe dplyr pronto per essere aggregato o elaborato ulteriormente.

```
# Interpolation test

# Dataframe di riferimento (deve contenere almeno le variabili lon e lat)
# lon lat range (x, y)
x_range <- as.numeric(c(-65, -42))
y_range <- as.numeric(c(52, 67))
step_ <- 0.25
# Reference dataframe
ref_grd <- expand.grid(lon = seq(from = x_range[1], to = x_range[2], by = step_), lat = seq(from = y_range[1], to = y_range[2], by = step_))
```

```
# Interpolazione  
df_interpolato <- interpolate_grid(dati_giornalieri, variable = "qnet", step = 0.25, reference_df = ref.
```
