

**Εργαστήριο Ανώτερης Γεωδαισίας
Μεταπτυχιακό Πρόγραμμα ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
«Αναλυτικές Μέθοδοι στη Γεωπληροφορική»
(Ακαδ. Έτος 2022 -23)**

ΟΝΟΜΑΤΕΠΩΝΥΜΟ

ΕΞΑΜΗΝΟ

Ημερομηνία Παράδοσης : **10/1/2023**

ΠΡΟΕΡΓΑΣΙΑ #3 (Προαιρετική αλλά συνιστάται)

Σκοπός: Το παρόν υπόδειγμα εξάσκησης αποσκοπεί στην πρώτη γνωριμία σας με το πώς να διαβάζετε, να εισάγετε και να επεξεργάζεστε αρχεία δεδομένων σε μορφότυπο Netcdf χρησιμοποιώντας βασικές εντολές και συναρτήσεις του R και πακέτων του.

(1) Πριν αρχίσετε την προεργασία σας, συμβουλευτείτε τους ακόλουθους ιστότοπους για ένα πλήρες παράδειγμα χρήσης αρχείων NetCDF από το έγκριτο διαδικτυακό κόμβο (portal) γεωδεδομένων στο *Εθνικό Εργαστήριο Oak Ridge Distributed Active Archive Center (ORNL DAAC) των ΗΠΑ*:

- https://github.com/ornl daac/netCDF_data_in_R/blob/master/README.md
- (ολόκληρο το R script της φροντιστηριακής άσκησης βρίσκεται στον ακόλουθο σύνδεσμο:
https://github.com/ornl daac/netCDF_data_in_R/blob/master/netCDF_in_r_ornl daac_tutorial.Rmd

Συγκεκριμένα καλείστε να αναπαράγατε τα βήματα της ανάλυσης ενός συνόλου δεδομένων που περιέχει δείκτες κανονικοποιημένης διαφοράς βλάστησης (*NDVI, Normalized Difference Vegetation Index*) στις αρκτικές περιοχές. Τα δεδομένα προέρχονται κυρίως από αισθητήρες προηγμένου ραδιομέτρου πολύ υψηλής ανάλυσης (AVHRR) σε αρκετούς δορυφόρους NOAA κατά τα έτη 1982 έως 2012. Εν προκειμένω, αρχικά θα χρειαστεί να ανακτήσετε το αρχείο **gimms3g_ndvi_1982-2012.nc4** ακολουθώντας τον σύνδεσμο <https://doi.org/10.3334/ORNLDAAC/1275> αφού προηγουμένως κάνετε μια απλή εγγραφή (με τα στοιχεία και το email σας, στο <https://urs.earthdata.nasa.gov/users/new>) για την πρόσβασή σας στην τεράστια ποικιλία των δεδομένων που παρέχει η NASA. Σημειώστε, ότι το User ID και το Password που θα δηλώσετε, θα μπορείτε να τα χρησιμοποιήσετε μελλοντικά ένα χρειαστείτε να έχετε πρόσβαση σε διάφορους κόμβους δεδομένων της NASA (συνεπώς, συνιστάται να κρατήσετε αυτές τις πληροφορίες σας για μελλοντική χρήση).

(2) Ακολουθήστε βήμα-βήμα ένα έτοιμο παράδειγμα χρήσης του R για την επεξεργασία και ανάλυση μιας εικόνας σε μορφή .TIFF, με γεωαναφορά, αναφερόμενη στη βαθυμετρία της λίμνης Victoria, τη μεγαλύτερη σε έκταση λίμνη της Αφρικής που είναι ταυτόχρονα η

μεγαλύτερη τροπική και η δεύτερη σε έκταση λίμνη γλυκού νερού της Γης. Τα αναγκαία δεδομένα βρίσκονται στον ιστοχώρο του Πανεπιστημίου Harvard

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/SOEKNR> →

Access Data

και συγκεκριμένα στο αρχείο TIFF με την ονομασία [LV Bathy V7.tif](#). Ο απευθείας σύνδεσμος του αρχείου είναι

<https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/SOEKNR/J3ESQR&version=10.2> .

Το πλήρες παράδειγμα (βήμα προς βήμα εντολές του R) θα το βρείτε στην ιστοσελίδα

<https://semba-blog.netlify.app/04/24/2020/lake-victoria-bathymetry/>

Για τους σκοπούς της άσκησης θα χρειαστεί να φορτώσετε τα πακέτα του R **tidyverse**, **sf**, και **leaflet**. Σημειώστε ότι καθώς τα πακέτα **raster** και **tidyverse** χρησιμοποιούν κάποιες συναρτήσεις με κοινά ονόματα, οι εκάστοτε συναρτήσεις του **raster** πρέπει να καλούνται χρησιμοποιώντας εντολές της μορφής **raster::function()** με τα απαραίτητα ορίσματα.

(3) Ο ιστότοπος της NASA <https://data.giss.nasa.gov/gistemp/> περιέχει ατμοσφαιρικά δεδομένα, σε πλέγμα κανάβου $2^{\circ} \times 2^{\circ}$, για την εκτίμηση της μεταβολής της παγκόσμιας θερμοκρασίας πάνω από τη γήινη επιφάνεια. Στη συγκεκριμένη ιστοσελίδα, ανατρέξτε στην ενότητα **'Gridded Monthly Temperature Anomaly Data'** (Πλέγμα Μηνιαίων Δεδομένων Ανωμαλιών Θερμοκρασίας) → **'Compressed NetCDF Files (regular $2^{\circ} \times 2^{\circ}$ grid)'** (Συμπίεσμένα αρχεία NetCDF (κανονικό πλέγμα $2^{\circ} \times 2^{\circ}$)) και από εκεί κατεβάστε στον χώρο εργασίας του R το netCDF αρχείο [Land-Ocean Temperature Index, ERSSTv5, 1200km smoothing](#) του οποίου ο πλήρης σύνδεσμος είναι

https://data.giss.nasa.gov/pub/gistemp/gistemp1200_GHCNV4_ERSSTv5.nc.gz.

Για τους σκοπούς της Προεργασίας θα χρειαστεί να έχετε εγκαταστήσει και να φορτώσετε τα πακέτα του R **fields**, **ncdf4**, **R.utils**, **ggplot2**, **maps**, ...

Με ένα απλό κλικ στο σύνδεσμο με την ονομασία του αρχείου θα μπορούσατε να το κατεβάσετε στον Η/Υ σας. Ωστόσο, αντί αυτού, κατεβάστε το ζητούμενο αρχείο απευθείας στο χώρο εργασίας σας του R, χρησιμοποιώντας τη συνάρτηση **download.file()** με μια εντολή της μορφής **download.file(URL, destfile, mode = 'wb')**, όπου URL είναι η διαδικτυακή διεύθυνση του ζητούμενου αρχείου εν προκειμένω, η διεύθυνση του ζητούμενου αρχείου θα είναι

https://data.giss.nasa.gov/pub/gistemp/gistemp1200_GHCNV4_ERSSTv5.nc.gz),

destfile είναι η ονομασία του αρχείου προορισμού στο οποίο θα αποθηκευθεί ο αρχείο που θα κατεβάσετε στο χώρο εργασίας του R (εν προκειμένω, χρησιμοποιήστε μια συντομότερη ονομασία, όπως **gistemp1200_ERSSTv5.nc.gz**). Η παράμετρος **mode = 'wb'**, είναι απαραίτητη προκειμένου να το αρχείο προορισμού να αποθηκευθεί σε binary μορφή.

Ελέγξτε ότι το αρχείο **gistemp1200_ERSSTv5.nc.gz** έχει όντως κατέβει στο χώρο εργασίας σας του R.

Μπορείτε να πάρετε πληροφορίες για το περιεχόμενο του ενός συμπίεσμένου NetCDF αρχείου, π.χ. [compressed netCDF file.gz](#) , με μια εντολή της μορφής:

```
print(file.info("compressed netCDF file.gz"))
```

Αντίστοιχα, η αποσυμπίεση του ίδιου αρχείου μπορεί να γίνει κάνοντας χρήση της συνάρτησης **gunzip()**, από το πακέτο **R.utils**, με μια εντολή της μορφής:

```
gunzip("compressed netCDF file.gz", destname="sorterfilename.nc", remove=FALSE)
```

όπου το όρισμα **destname** μπορεί να δώσει μια συντομότερη ονομασία στο αποσυμπιεσμένο NetCDF αρχείο (εάν παραληφθεί το αποτέλεσμα θα είναι το αποσυμπιεσμένο NetCDF αρχείο να διατηρήσει την αρχική ονομασία που είχε εντός του αρχικού συμπίεσμένου NetCDF αρχείου εισόδου **compressed netCDF file.gz**), ενώ το όρισμα **remove=FALSE** είναι απαραίτητο προκειμένου να μην διαγραφεί, μετά την αποσυμπίεση, το αρχείο εισόδου **compressed netCDF file.gz**. Η αποσυμπίεση, εάν έγινε επιτυχώς, θα έχει δώσει στο χώρο εργασίας του R ένα αποσυμπιεσμένο αρχείο netCDF με την ονομασία **sorterfilename.nc**.

Η ανάγνωση δεδομένων που είναι αποθηκευμένα σε αρχεία netCDF είναι αρκετά απλή στο R, με την προϋπόθεση ότι έχετε φορτώσει προηγουμένως τα κατάλληλα πακέτα. Ίσως χρειαστεί να εγκαταστήσετε αυτά τα πακέτα αν δεν τα έχετε χρησιμοποιήσει στο παρελθόν.

Διαβάστε τα περιεχόμενα του αρχείου netCDF **sorterfilename.nc** (τα οποία είναι τιμές ανωμαλιών θερμοκρασίας) χρησιμοποιώντας τη συνάρτηση **nc_open()** και εκχωρήστε τα δεδομένα σε μια δομή/μεταβλητή δεδομένων με την ονομασία **nctemp**. Εκτυπώστε τα μεταδεδομένα σχετικά με αυτό το αρχείο σε ένα αρχείο κειμένου. Χρησιμοποιήστε εντολές της μορφής

```
nctemp <- nc_open("sorterfilename.nc")
print(nctemp)
```

Μπορείτε επίσης να αποθηκεύσετε τα μεταδεδομένα σχετικά με αυτό το αρχείο σε ένα αρχείο κειμένου στο χώρο εργασίας του R, με κάποια ονομασία της επιλογής σας, π.χ. ***_metadata.txt**, χρησιμοποιώντας τις ακόλουθες εντολές μιας δικής σας συνάρτησης

```
{
  # Save the print(nc) dump to a text file '*_metadata.txt'
  sink('*_metadata.txt')
  print(nctemp)
  sink()
}
```

Από το αποτέλεσμα της εκτύπωσης των μεταδεδομένων παρατηρήστε ότι υπάρχουν δύο μεταβλητές: **time_bnds**, που περιέχει την ημερομηνία έναρξης και λήξης κάθε παρατήρησης και **tempanomaly**, που είναι η μεταβλητή που μας ενδιαφέρει. Το **tempanomaly** έχει 3 διαστάσεις: [lon, lat, time] - γεωγραφικό μήκος, γεωγραφικό πλάτος και χρονική εποχή παρατήρησης. Παρατηρήστε ότι οι διαστάσεις lat και lon είναι αντίστοιχα 90 και 180 γιατί ο κানাβος των δεδομένων σε όλη τη γήινη σφαίρα είναι σε ισοδιάσταση 2° x 2°. Συνολικά, υπάρχουν τέσσερις διαστάσεις στο αρχείο: lat, lon, time και "nv", που χρησιμοποιείται για την καταγραφή της αρχής και του τέλους του χρονικού εύρους των παρατηρήσεων. Μπορείτε να αγνοήσετε το "nv" και να εστιάσετε στις τρεις διαστάσεις που χρησιμοποιούνται για την οργάνωση των δεδομένων **tempanomaly**. Υπάρχουν επίσης 5 καθολικά (global) χαρακτηριστικά που παρέχουν πληροφορίες μεταδεδομένων για το αρχείο.

Ενδιαφέρει να εξάγετε αυτά τα δεδομένα στις διαστάσεις lat, lon και χρόνου time κάθε παρατήρησης **tempanomaly** και να αποθηκευθούν. Για αυτό, χρησιμοποιήστε εντολές της μορφής:

```
lon <- ncvarget(nc_data, "lon")
head(lon) # look at the first few entries in the longitude vector

lat <- ncvarget(nc_data, "lat")
head(lat) # look at the first few entries in the latitude vector

t <- ncvarget(nc_data, "time")
head(t) # look at the first few entries in the time vector
```

όπου nc_data είναι το όνομα του netCDF αρχείου ενδιαφέροντος, εν προκειμένω το **nctemp**.

Παρατηρήστε ότι οι χρονικές εποχές δεν δίνονται σε κάποια γνώριμη μορφή, αλλά είναι σε **‘ημέρες από 01/01/1800 00:00:00** (days since 1800-01-01 00:00:00)’, όπως φαίνεται και από τα μεταδεδομένα που εκτυπώσατε προηγουμένως (29233 29264 29293 29324 ...). Επομένως, είναι κατανοητό ότι οι χρονικές ενδείξεις πρέπει να μετατραπούν σε τρέχουσες ημερομηνίες και χρονικές εποχές, χρησιμοποιώντας εντολές της μορφής

```
tm = as.Date( t, origin="1800-01-01", tz="UTC")
tm
```

όπου tz υποδηλώνει την επιθυμητή χρονική ζώνη (εν προκειμένω, προσδιορίζουμε να μετατραπούν σε Παγκόσμια Συντονισμένη Ώρα / Universal Coordinated Time).

Παρατηρήστε ότι μετά τη μετατροπή, οι χρονικές ενδείξεις είναι τις μορφής π.χ., 1880-06-15 (έτος-μήνας-μέρα).

Εξάγετε τις τιμές των ανωμαλιών θερμοκρασίας σε όλα τα σημεία του πλέγματος κανάβου των δεδομένων, για την πρώτη χρονική εποχή "1880-01-15". Επίσης μπορείτε να εξακριβώσετε εάν στα δεδομένα περιλαμβάνεται κάποια τιμή πλήρωσης κενών τιμών που μπορεί να χρησιμοποιήθηκε για δεδομένα που λείπουν. Για αυτό το σκοπό χρησιμοποιήστε εντολές της μορφής

```
anom = ncvarget(nctemp, "tempanomaly", start=c(1,1,1), count=c(-1,-1,1) )
dim(anom)

fillvalue <- ncatt_get(nctemp, "tempanomaly", "_FillValue")
fillvalue # should be as shown also in the metadata for the specific variable

head(anom)
tail(anom)
```

Στο όρισμα **count=c(-1,-1,1)** στην κλήση της συνάρτησης **ncvarget()**, οι τιμές -1, -1 υποδηλώνουν ότι θα ληφθούν όλες οι διαθέσιμες τιμές κατά lon και lat στα δεδομένα **tempanomaly**, ενώ η τρίτη τιμή 1 υποδηλώνει ότι θα ληφθεί υπόψη μόνο το πρώτο χρονική εποχή.

Για να δημιουργήσετε ένα τυπικό παγκόσμιο χάρτη απεικόνισης των ανωμαλιών θερμοκρασίας κατά τη χρονική εποχή "1880-01-15" (για την ποία έχουν επιλεγεί τα δεδομένα στο διάγραμμα τιμών **anom**), χρησιμοποιήστε μια-μια ξεχωριστά τις ακόλουθες εντολές:

```
image.plot(lon,lat,anom) # function image.plot() is in package 'fields'
```

```
map(add=T) # function map() is in package 'maps'
```

Το όρισμα **add=T** ή **add=TRUE** στην κλήση της συνάρτησης **map()** υποδηλώνει ότι οι οριογραμμές ακτών και συνόρων μεταξύ κρατών θα προστεθούν στο γράφημα που παράγει η κλήση της συνάρτησης **image.plot()**.

Ακολουθώς δημιουργήστε ένα χάρτη ανωμαλιών θερμοκρασίας για τη 10^η ("1880-10-15") χρονική εποχή στα επίπεδα των χρονικών δεδομένων, τις ακόλουθες τυπικές εντολές:

```
tm[10]
anom10 = ncvr_get(nctemp, "tempanomaly", start=c(1,1,10), count=c(-1,-1,1) )
dim(anom10)
image.plot(lon,lat,anom10) # in package 'fields'
map(add=T) # in package 'maps'
```

Κατά παρόμοιο τρόπο δημιουργήστε ένα χάρτη ανωμαλιών θερμοκρασίας για την 25^η ("1882-01-15") χρονική εποχή στα επίπεδα των χρονικών δεδομένων.

Δημιουργήστε έναν αντίστοιχο χάρτη ανωμαλιών θερμοκρασίας για την περιοχή με γεωγραφικά πλάτη lat[31°, 45°] και γεωγραφικά μήκη lon[-7°, 41°]. Χρησιμοποιήστε τις ενδεικτικές εντολές

```
lat_rng=c(31,45) # lat limits for area
lon_rng=c(-7,41) # lon limits for area
lat_rng
lon_rng

lat_ind = which( lat >= lat_rng[1] & lat <= lat_rng[2] )
lat_ind # vector of indices for lat values
lat[lat_ind] # confirm the lat values in the positions indicated by lat_ind

lon_ind = which( lon >= lon_rng[1] & lon <= lon_rng[2] )
lon_ind # vector of indices for lon values
lon[lon_ind] # confirm the lon values in the positions indicated by lon_ind

lat_area = lat[lat_ind]
lat_area # show lat values for the area
lon_area = lon[lon_ind]
lon_area # show lon values for the area
```

Ακολουθώς εκχωρήστε τις αντίστοιχες τιμές ανωμαλιών θερμοκρασίας σε ένα πίνακα τιμών με την ονομασία **anom_tseries1**, χρησιμοποιώντας τις ακόλουθες εντολές

```
anom_tseries1 = ncvr_get(nctemp, "tempanomaly",
  start = c(lon_ind[1],lat_ind[1],1),
  count = c(length(lon_ind),length(lat_ind),1) )
is.matrix(anom_tseries1)
dim(anom_tseries1)

image.plot(lon_area, lat_area,anom_tseries1)
map(add=T)
```

Ενδιαφέρει να εξάγετε την χρονοσειρά των μέσων τιμών ανωμαλιών θερμοκρασίας στην περιοχή που περικλείεται από τις γεωγραφικές συντεταγμένες lat[35°, 37°] και lon[21°, 23°] (~ Νότια Πελοπόννησος) για όλη τη χρονική περίοδο των παρατηρήσεων.

```

lat_area=c(35,37)
lon_area =c(21,23)

lat_area
lon_area

lat_ind = which( lat >= lat_area[1] & lat <= lat_area[2] )
lat_ind
lat[lat_ind]

lon_ind = which( lon >= lon_area[1] & lon <= lon_area[2] )
lon_ind
lon[lon_ind]
lat_rgn = lat[lat_ind]
lat_rgn
lon_rgn = lon[lon_ind]
lon_rgn

#      Extract the temperature anomalies for the area of interest
anom_tseries = ncvr_get(nctemp, "tempanomaly",
                        start = c(lon_ind[1], lat_ind[1], 1),
                        count = c(length(lon_ind),length(lat_ind),-1) )

is.array(anom_tseries)      # check if the data is in the form of an array
dim(anom_tseries)          # print the dimensions of the array

#      Compute the means of the temperature anomalies over the area
#      Ignore NA values
anom_ts = apply(anom_tseries, 3, mean, no.na=T)

length(anom_ts)             # There should be 1702 values
plot(tm,anom_ts,type="l")   # Plot a line graph

```

ΣΗΜΕΙΩΣΗ – Με τα παραδοτέα σας θα πρέπει να συμπεριληφθούν τα εκάστοτε *.Rhistory* αρχεία που θα προκύψουν από τις ξεχωριστές συνεδρίες του R που διεκπεραιώσατε για την υλοποίηση κάθε μέρους της Θεματικής Εργασίας.