

A többdimenziós adatstruktúrák értelmezése és a ncdf4 R package alkalmazása

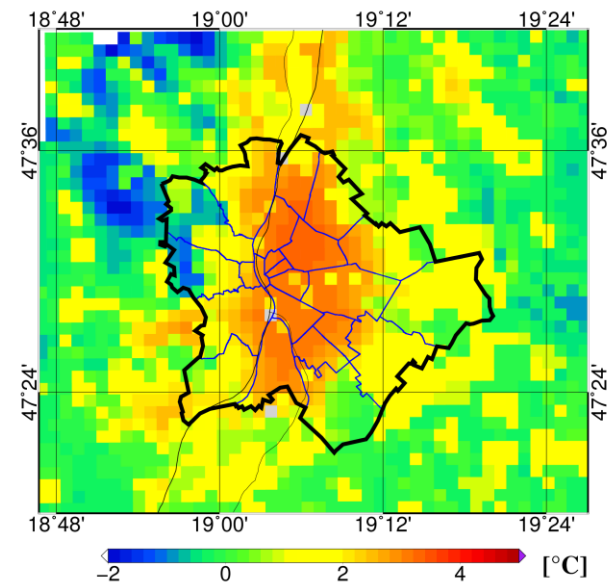


Göndöcs Julcsi
Meteorológus, PhD hallgató

R-Ladies meetup, 2017. 12.12.

A többdimenziós adatstruktúrák értelmezése

- ▶ Legtöbb esetben 2D adattáblák
 - ▶ fejléc és soronként rekordok
- ▶ Meteorológiában 3D és (4D tömbök)
 - ▶ meteorológiai modell outputok -> eltérő dimenziójú változók sokaságát tartalmazzák (2,3,4D)
 - ▶ 2D tömbök: felszíni adatok kiterítve (t = 1 adott időpillanat)
 - ▶ Nem csak felszíni hanem magassági szintenként is



Származtatott mező:
városi hősziget intenzitás

0.689	0.706	0.118	0.884	...
0.535	0.532	0.653	0.925	...
0.314	0.265	0.159	0.101	...
0.553	0.633	0.528	0.493	...
0.441	0.465	0.512	0.512	...
0.398	0.401	0.421	0.398	...
0.342	0.647	0.515	0.816	...
0.111	0.300	0.205	0.526	...
0.523	0.428	0.712	0.929	...
0.214	0.604	0.918	0.344	...
0.100	0.121	0.113	0.126	...
0.112	0.986	0.234	0.432	...
0.765	0.128	0.863	0.521	...
1.000	0.985	0.761	0.698	...
0.455	0.783	0.224	0.395	...
0.021	0.500	0.311	0.123	...
1.000	1.000	0.867	0.051	...
1.000	0.945	0.998	0.893	...
0.990	0.941	1.000	0.876	...
0.902	0.867	0.834	0.798	...
...
...
...
0.204	0.175
0.760	0.531
0.997	0.910
0.995	0.726

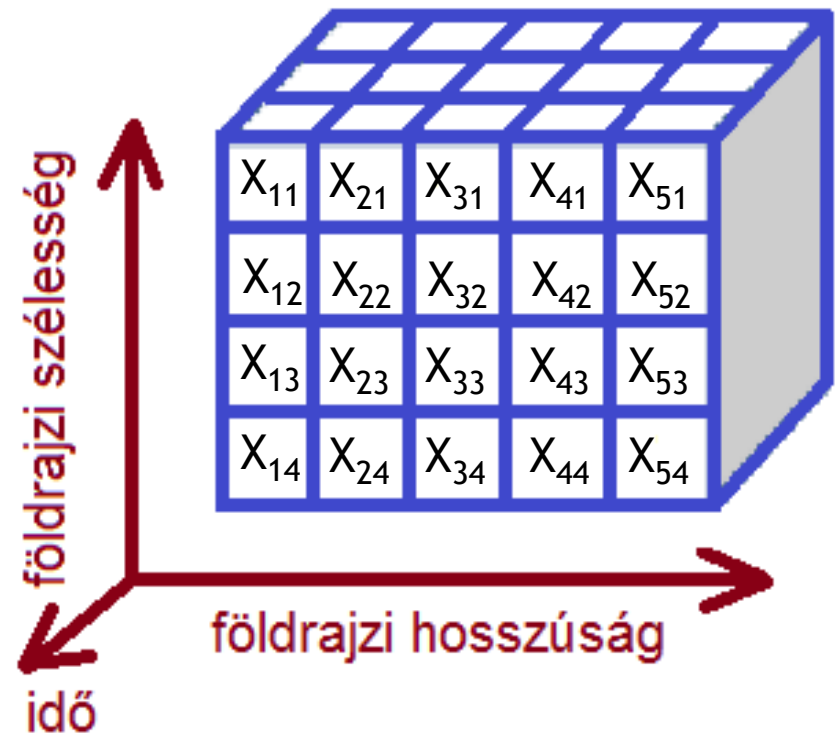
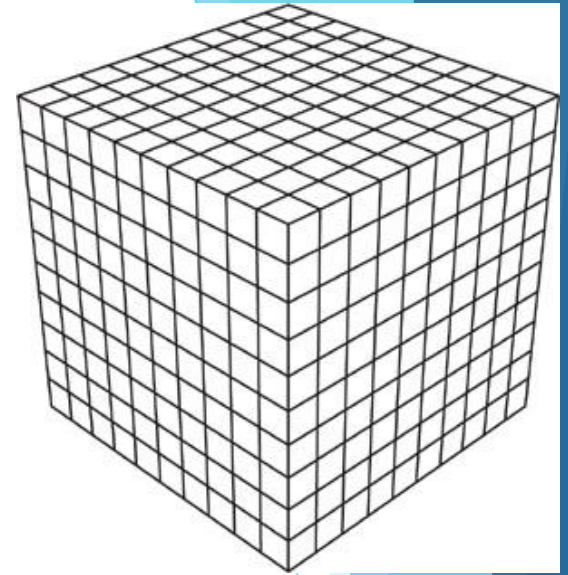
A többdimenziós adatstruktúrák értelmezése

▶ 3D tömbök

- ▶ Minden egyes időpontra tartalmazza a 2D rács értékeit

▶ 4D tömbök

- ▶ A tér minden (x,y,z) pontjára, minden időpillanatban tartalmazzák az adatokat



NetCDF fájlok a földtudományokban (Network Common Data Form)

- ▶ Platformfüggetlen
- ▶ Bináris formátumban tárolja a fájlokat -> gyors elérés
- ▶ Önleíró, mivel tartalmazza az adatokat és a hozzájuk tartozó metaadatokat, amelyek meghatározzák az adatok struktúráját, is egy helyen
- ▶ Változóknak előredefiniált tulajdonságai vannak
- ▶ Adatok tárolása úgy, hogy könnyen darabolható legyen



ncdf4 R package alkalmazása

- ▶ `install.packages("ncdf4")`
- ▶ `library(ncdf4)`

ncdf4 R package alkalmazása

- ▶ `proba <- nc_open("T2.nc")` | fájl megnyitása
- ▶ `print(proba)` | metaadatok kinyerése

```
Console ~/
> print(proba)
File T2.nc (NC_FORMAT_64BIT):

  1 variables (excluding dimension variables):
    float T2[west_east,south_north,Time]
      FieldType: 104
      MemoryOrder: XY
      description: TEMP at 2 M
      units: K
      stagger:
      coordinates: XLONG XLAT XTIME

  3 dimensions:
    Time Size:121 *** is unlimited ***
    south_north Size:93
    west_east Size:108

122 global attributes:
  TITLE: OUTPUT FROM WRF V3.8 MODEL
  START_DATE: 2000-07-14_00:00:00
  SIMULATION_START_DATE: 2000-07-14_00:00:00
  WEST-EAST_GRID_DIMENSION: 109
  SOUTH-NORTH_GRID_DIMENSION: 94
  BOTTOM-TOP_GRID_DIMENSION: 44
  DX: 1111.11096191406
  DY: 1111.11096191406
  SKEBS_ON: 0
  SPEC_BDY_FINAL_MU: 1
  USE_Q_DIABATIC: 0
  GRIDTYPE: C
  DIFF_OPT: 0
  KM_OPT: 4
```

T2 [hosszúság, szélesség, idő]

ncdf4 R package alkalmazása

- ▶ `attributes(proba)$names` | milyen tulajdonságokkal rendelkezik
- ▶ `attributes(proba$var)$names` | attribútumok nevei
- ▶ `valtozo <- ncvar_get(proba, varid="T2")` | változó adatainak kinyerése

```
Console ~/  
> attributes(proba)$names
[1] "filename" "writable" "id" "safemode" "format" "is_GMT" "groups" "fqgn2"
Rindex"
[9] "ndims" "natts" "dim" "unlimdimid" "nvars" "var"
> attributes(proba$var)$names
[1] "T2" "XLAT" "XLONG"
> |
```

values	
proba	List of 14
valtozo	Large array (1215324 elements, 9.3 ...
um	[1:108, 1:93, 1:121] 298 298 298 298 298 ...

T2 [hosszúság, szélesség, idő]

ncdf4 R package alkalmazása

- ▶ `nc_close(proba)` | bezárjuk a fájlt
- ▶ R segítségével NetCDF fájl definiálása is lehetséges
- ▶ Következőkben:
 - ▶ 3D légnyomási adatok statisztikai vizsgálata-> [hosszúság, szélesség, idő]
 - ▶ Adat vizualizáció

Köszönöm a figyelmet!