Basic workflow to work with ERA5land data with stars

Carles Milà

2021-10-22

Contents

Getting ready

This notebook contains a basic example of a R-based workflow for data management and analysis of ERA5land data with the stars package. The example data consist on 2m and skin temperature ERA5land data for the study area comprised between 40ºN, 44ºN, 0ºE, 4ºE at an hourly resolution for January 2020. These can be easily downloaded from the climate data store and are also included in the repo.

```
library("stars")
library("sf")
library("readr")
library("dplyr")
library("viridis")
library("ggplot2")
```

Read data and stars 101

We easily read ERA5land data using `read_stars`.

```
(airtemp <- read_stars("data/ERA5Land_2mtemp.grib"))
```

```
## stars object with 3 dimensions and 1 attribute
## attribute(s), summary of first 1e+05 cells:
##  ERA5Land_2mtemp.grib
##  Min.   :268.7
##  1st Qu.:275.3
##  Median :277.9
##  Mean   :278.0
##  3rd Qu.:280.5
##  Max.   :287.4
##  NA's   :39895
## dimension(s):
##      from  to offset delta                       refsys point
## x       1  41  -0.05   0.1 Coordinate System importe...    NA
```

```
## y       1  41  44.05  -0.1 Coordinate System importe...    NA
## band    1 744     NA    NA                             NA    NA
##                                                                    values
## x                                                                    NULL
## y                                                                    NULL
## band 0[-] SFC (Ground or water surface),...,0[-] SFC (Ground or water surface)
##      x/y
## x    [x]
## y    [y]
## band
```

```
(skintemp <- read_stars("data/ERA5Land_skintemp.grib"))
```

```
## stars object with 3 dimensions and 1 attribute
## attribute(s), summary of first 1e+05 cells:
##  ERA5Land_skintemp.grib
##  Min.   :258.1
##  1st Qu.:272.1
##  Median :276.2
##  Mean   :276.4
##  3rd Qu.:280.4
##  Max.   :290.8
##  NA's   :39895
## dimension(s):
##      from  to offset delta                       refsys point
## x       1  41  -0.05   0.1 Coordinate System importe...    NA
## y       1  41  44.05  -0.1 Coordinate System importe...    NA
## band    1 744     NA    NA                             NA    NA
##                                                                    values
## x                                                                    NULL
## y                                                                    NULL
## band 0[-] SFC (Ground or water surface),...,0[-] SFC (Ground or water surface)
##      x/y
## x    [x]
## y    [y]
## band
```

We use read our AOI that was stored as a rds object.

```
(catalonia <- read_rds("data/cat.rds"))
```

```
## Simple feature collection with 1 feature and 0 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 260160.2 ymin: 4488767 xmax: 526553.9 ymax: 4747976
## Projected CRS: ETRS89 / UTM zone 31N
## # A tibble: 1 x 1
##                                                                  geometry
##                                                               <POLYGON [m]>
## 1 ((386439.4 4561272, 386437.2 4561272, 386428.8 4561272, 386422.2 4561272, 386~
```

```
plot(catalonia)
```

We can merge the two products and fix their names.

```
alltemp <- c(airtemp, skintemp)
names(alltemp) <- gsub(".grib", "", names(alltemp))
alltemp
```

```
## stars object with 3 dimensions and 2 attributes
## attribute(s), summary of first 1e+05 cells:
##  ERA5Land_2mtemp  ERA5Land_skintemp
##  Min.   :268.7    Min.   :258.1
##  1st Qu.:275.3    1st Qu.:272.1
##  Median :277.9    Median :276.2
##  Mean   :278.0    Mean   :276.4
##  3rd Qu.:280.5    3rd Qu.:280.4
##  Max.   :287.4    Max.   :290.8
##  NA's   :39895    NA's   :39895
## dimension(s):
##      from  to offset delta                       refsys point
## x       1  41  -0.05   0.1 Coordinate System importe...    NA
## y       1  41  44.05  -0.1 Coordinate System importe...    NA
## band    1 744     NA    NA                            NA    NA
##                                                                values
## x                                                                NULL
## y                                                                NULL
## band 0[-] SFC (Ground or water surface),...,0[-] SFC (Ground or water surface)
##      x/y
## x    [x]
## y    [y]
## band
```

We subset stars objects by using []. 1st is the attribute, next are the dimensions in order.

```
alltemp[1] # subset attribute
```

```
## stars object with 3 dimensions and 1 attribute
## attribute(s), summary of first 1e+05 cells:
##  ERA5Land_2mtemp
```

```
##  Min.   :268.7
##  1st Qu.:275.3
##  Median :277.9
##  Mean   :278.0
##  3rd Qu.:280.5
##  Max.   :287.4
##  NA's   :39895
## dimension(s):
##      from  to offset delta                       refsys point
## x       1  41  -0.05   0.1 Coordinate System importe...    NA
## y       1  41  44.05  -0.1 Coordinate System importe...    NA
## band    1 744     NA    NA                            NA    NA
##                                                                values
## x                                                                NULL
## y                                                                NULL
## band 0[-] SFC (Ground or water surface),...,0[-] SFC (Ground or water surface)
##      x/y
## x    [x]
## y    [y]
## band
```

```
alltemp[, 1] # subset 1st dimension
```

```
## stars object with 3 dimensions and 2 attributes
## attribute(s):
##  ERA5Land_2mtemp  ERA5Land_skintemp
##  Min.   :261.8    Min.   :259.1
##  1st Qu.:275.1    1st Qu.:273.1
##  Median :278.7    Median :277.3
##  Mean   :278.6    Mean   :277.5
##  3rd Qu.:282.0    3rd Qu.:281.6
##  Max.   :292.2    Max.   :294.7
## dimension(s):
##      from  to offset delta                       refsys point
## x       1   1  -0.05   0.1 Coordinate System importe...    NA
## y       1  41  44.05  -0.1 Coordinate System importe...    NA
## band    1 744     NA    NA                            NA    NA
##                                                                values
## x                                                                NULL
## y                                                                NULL
## band 0[-] SFC (Ground or water surface),...,0[-] SFC (Ground or water surface)
##      x/y
## x    [x]
## y    [y]
## band
```

```
alltemp[,, 1] # subset 2nd dimension
```

```
## stars object with 3 dimensions and 2 attributes
## attribute(s):
##  ERA5Land_2mtemp  ERA5Land_skintemp
##  Min.   :269.6    Min.   :267.1
##  1st Qu.:277.5    1st Qu.:276.7
##  Median :280.0    Median :279.7
##  Mean   :279.7    Mean   :279.2
```

```
##  3rd Qu.:281.9    3rd Qu.:281.8
##  Max.   :289.2    Max.   :290.0
## dimension(s):
##      from  to offset delta                       refsys point
## x       1  41  -0.05   0.1 Coordinate System importe...    NA
## y       1   1  44.05  -0.1 Coordinate System importe...    NA
## band    1 744     NA    NA                            NA    NA
##                                                                values
## x                                                                NULL
## y                                                                NULL
## band 0[-] SFC (Ground or water surface),...,0[-] SFC (Ground or water surface)
##      x/y
## x    [x]
## y    [y]
## band
```

```
alltemp[,,, 1] # subset 3rd dimension
```

```
## stars object with 3 dimensions and 2 attributes
## attribute(s):
##  ERA5Land_2mtemp  ERA5Land_skintemp
##  Min.   :270.7    Min.   :261.9
##  1st Qu.:274.3    1st Qu.:270.7
##  Median :276.3    Median :273.7
##  Mean   :276.3    Mean   :273.7
##  3rd Qu.:278.4    3rd Qu.:277.4
##  Max.   :284.5    Max.   :281.1
##  NA's   :674      NA's   :674
## dimension(s):
##      from to offset delta                      refsys point
## x       1 41  -0.05   0.1 Coordinate System importe...    NA
## y       1 41  44.05  -0.1 Coordinate System importe...    NA
## band    1  1     NA    NA                           NA    NA
##                                  values x/y
## x                                  NULL [x]
## y                                  NULL [y]
## band 0[-] SFC (Ground or water surface)
```

We now fix the time dimension with the correct times we downloaded. For these data, it's hourly for Jan 2020.

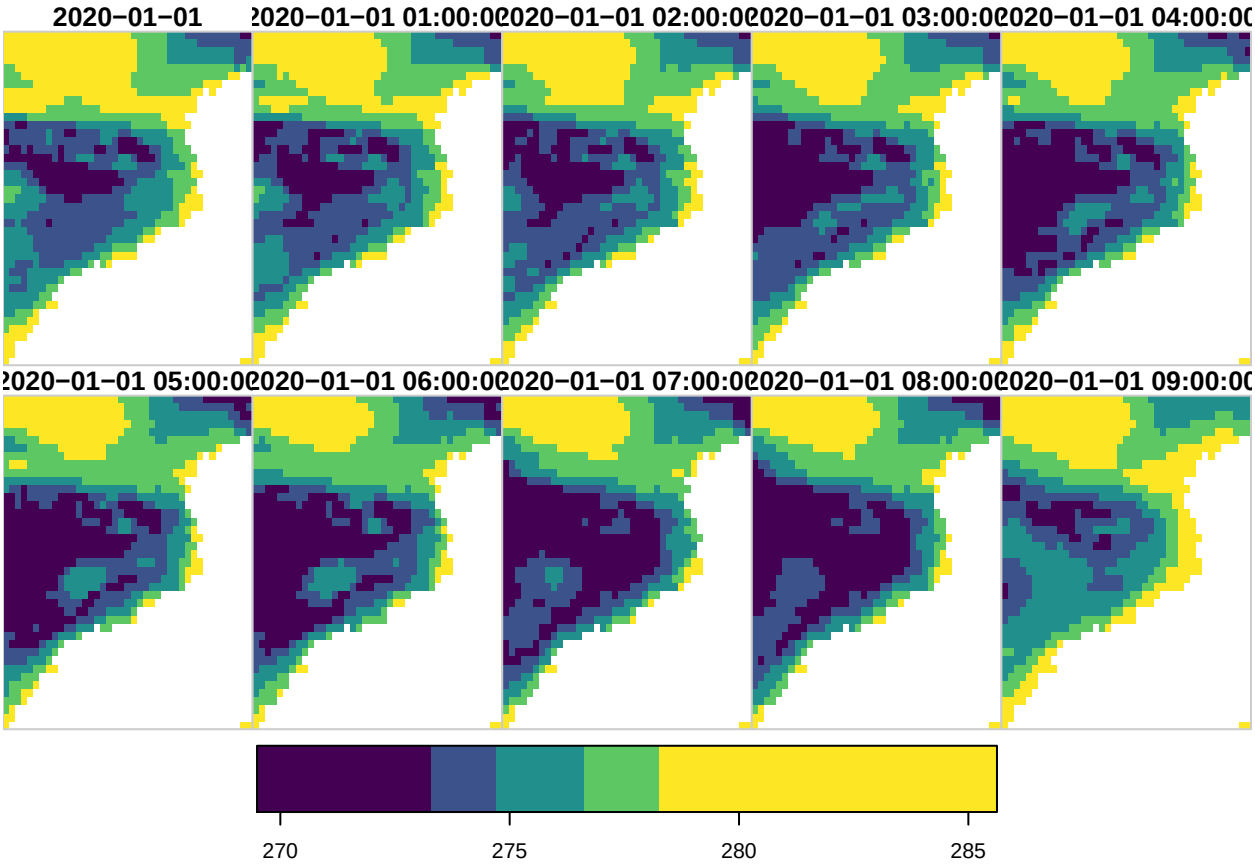
```
datetime_era5 <- seq.POSIXt(ISOdatetime(2020, 1, 1, 0, 0, 0),
                            ISOdatetime(2020, 1, 31, 23, 0, 0), by="1 hour")
dates_era5 <- seq.Date(as.Date("2020-01-01"), as.Date("2020-01-31"), "1 day")
(alltemp <- st_set_dimensions(alltemp, 3, names="datetime", values=datetime_era5))
```

```
## stars object with 3 dimensions and 2 attributes
## attribute(s), summary of first 1e+05 cells:
##  ERA5Land_2mtemp  ERA5Land_skintemp
##  Min.   :268.7    Min.   :258.1
##  1st Qu.:275.3    1st Qu.:272.1
##  Median :277.9    Median :276.2
##  Mean   :278.0    Mean   :276.4
##  3rd Qu.:280.5    3rd Qu.:280.4
##  Max.   :287.4    Max.   :290.8
##  NA's   :39895    NA's   :39895
```

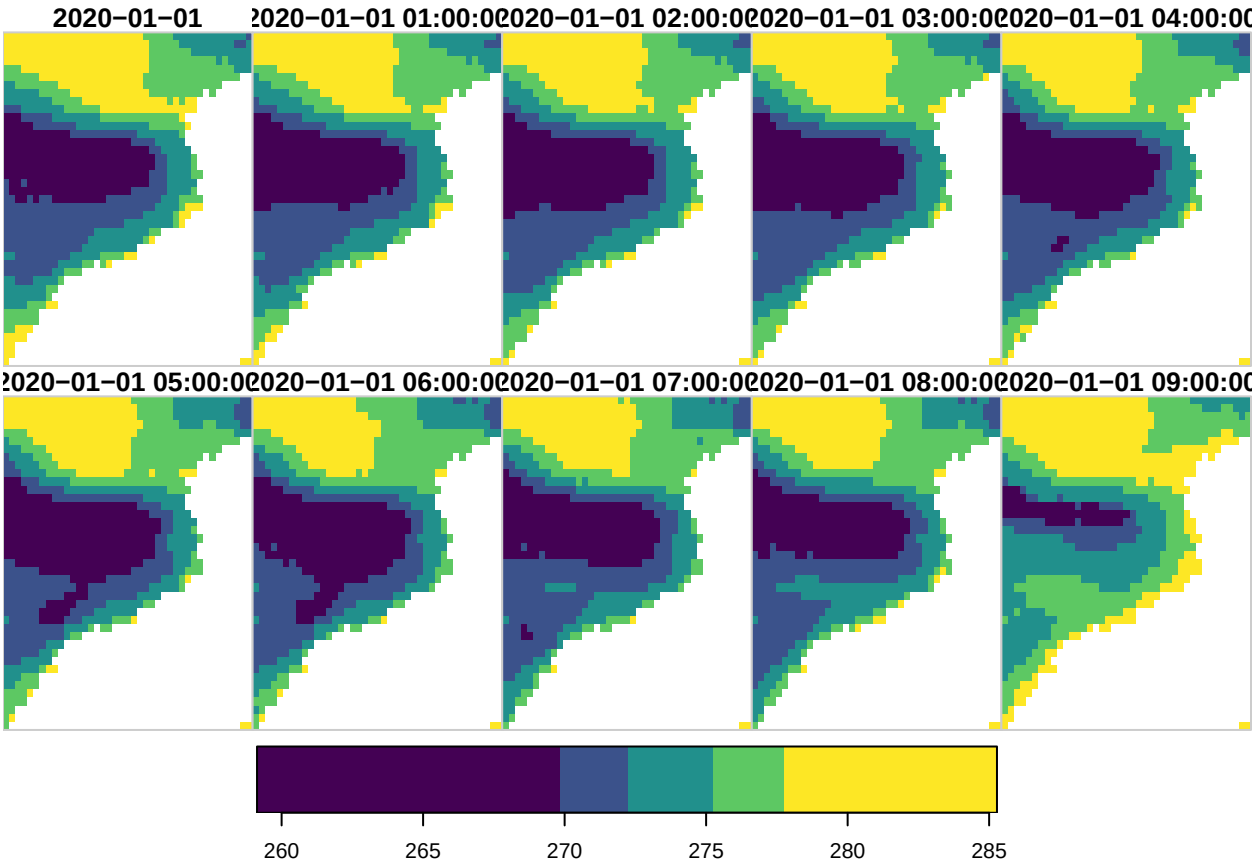
```
## dimension(s):
##          from  to         offset   delta                        refsys point
## x           1  41          -0.05     0.1 Coordinate System importe...    NA
## y           1  41          44.05    -0.1 Coordinate System importe...    NA
## datetime    1 744 2020-01-01 CET 1 hours                       POSIXct    NA
##          values x/y
## x          NULL [x]
## y          NULL [y]
## datetime   NULL
```

We can use the basic plotting capabilities of `stars` to check the data.

```
plot(alltemp["ERA5Land_2mtemp",,,1:10], col=viridis(5))
```




```
plot(alltemp["ERA5Land_skintemp",,,1:10], col=viridis(5))
```



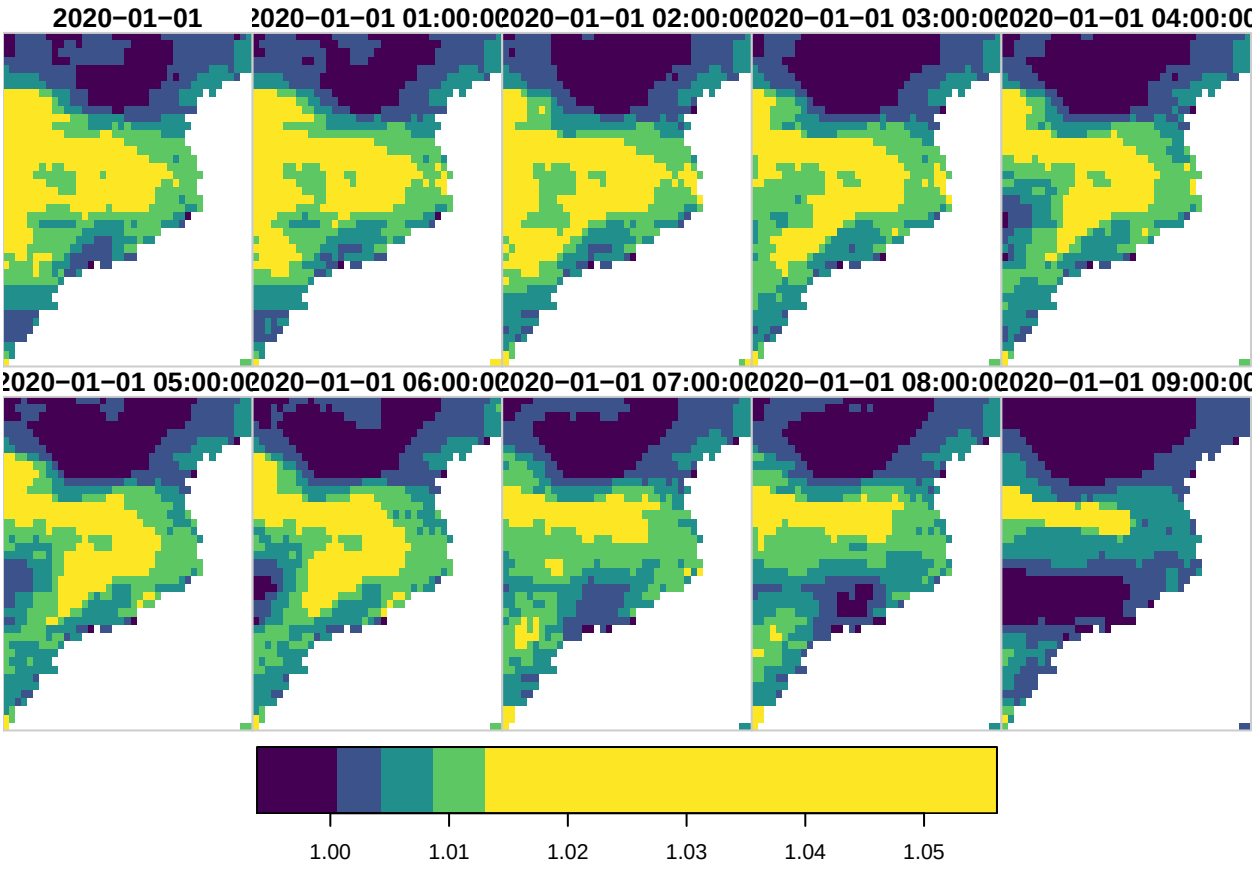
Band calculations

We can now make easy calculations within/between bands now. Here we do a ratio of air/skin temp.

```
(alltemp <- mutate(alltemp, ratio = ERA5Land_2mtemp/ERA5Land_skintemp))
```

```
## stars object with 3 dimensions and 3 attributes
## attribute(s), summary of first 1e+05 cells:
##  ERA5Land_2mtemp  ERA5Land_skintemp      ratio
##  Min.   :268.7    Min.   :258.1      Min.   :0.98
##  1st Qu.:275.3    1st Qu.:272.1      1st Qu.:1.00
##  Median :277.9    Median :276.2      Median :1.00
##  Mean   :278.0    Mean   :276.4      Mean   :1.01
##  3rd Qu.:280.5    3rd Qu.:280.4      3rd Qu.:1.01
##  Max.   :287.4    Max.   :290.8      Max.   :1.06
##  NA's   :39895    NA's   :39895      NA's   :39895
## dimension(s):
##          from  to         offset   delta                       refsys point
## x           1  41          -0.05     0.1 Coordinate System importe...    NA
## y           1  41          44.05    -0.1 Coordinate System importe...    NA
## datetime    1 744 2020-01-01 CET 1 hours                        POSIXct    NA
##          values x/y
## x          NULL [x]
## y          NULL [y]
## datetime   NULL
```

```
plot(alltemp[3,,,1:10], col=viridis(5))
```



Time aggregation and filtering

Now we can easily calculate daily averages, we convert the time dimension to date and aggregate.

```
(alltemp <- aggregate(alltemp, "day", FUN=mean))
```

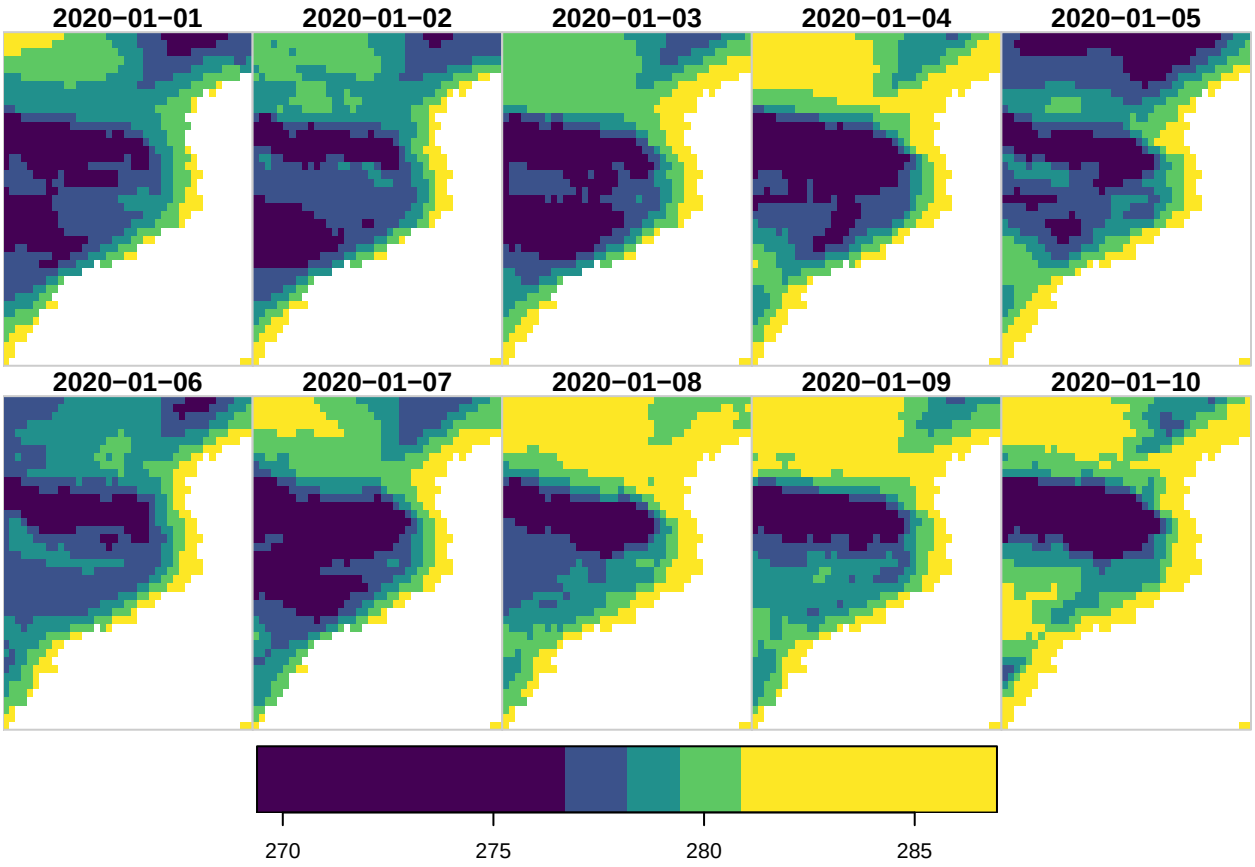
```
## stars object with 3 dimensions and 3 attributes
## attribute(s):
##  ERA5Land_2mtemp  ERA5Land_skintemp      ratio
##  Min.   :264.9    Min.   :260.2      Min.   :0.996
##  1st Qu.:276.6    1st Qu.:275.2      1st Qu.:1.002
##  Median :279.0    Median :278.2      Median :1.003
##  Mean   :278.7    Mean   :277.6      Mean   :1.004
##  3rd Qu.:281.0    3rd Qu.:280.4      3rd Qu.:1.005
##  Max.   :288.3    Max.   :288.2      Max.   :1.049
##  NA's   :20894    NA's   :20894      NA's   :20894
## dimension(s):
##      from to         offset  delta                       refsys point values
## time    1 31 2020-01-01 CET 1 days                      POSIXct    NA   NULL
## x       1 41          -0.05    0.1 Coordinate System importe...    NA   NULL
## y       1 41          44.05   -0.1 Coordinate System importe...    NA   NULL
##      x/y
## time
## x    [x]
```

```
## y    [y]
```

```
(alltemp <- st_set_dimensions(alltemp, "time", names = "date", values = dates_era5))
```

```
## stars object with 3 dimensions and 3 attributes
## attribute(s):
##  ERA5Land_2mtemp  ERA5Land_skintemp      ratio
##  Min.   :264.9    Min.   :260.2      Min.   :0.996
##  1st Qu.:276.6    1st Qu.:275.2      1st Qu.:1.002
##  Median :279.0    Median :278.2      Median :1.003
##  Mean   :278.7    Mean   :277.6      Mean   :1.004
##  3rd Qu.:281.0    3rd Qu.:280.4      3rd Qu.:1.005
##  Max.   :288.3    Max.   :288.2      Max.   :1.049
##  NA's   :20894    NA's   :20894      NA's   :20894
## dimension(s):
##      from to     offset  delta                       refsys point values x/y
## date    1 31 2020-01-01 1 days                         Date    NA   NULL
## x       1 41      -0.05    0.1 Coordinate System importe...    NA   NULL [x]
## y       1 41      44.05   -0.1 Coordinate System importe...    NA   NULL [y]
```

```
plot(alltemp[,1:10,,], col=viridis(5))
```



We can also filter by dimensions, for example date, if we want to:

```
(alltemp_8days<- dplyr::filter(alltemp, date <= as.Date("2020-01-08")))
```

```
## stars object with 3 dimensions and 3 attributes
## attribute(s):
```

```
##  ERA5Land_2mtemp  ERA5Land_skintemp      ratio
##  Min.   :270.7    Min.   :264.5      Min.   :0.997
##  1st Qu.:277.0    1st Qu.:275.2      1st Qu.:1.002
##  Median :278.5    Median :277.6      Median :1.004
##  Mean   :278.6    Mean   :277.1      Mean   :1.005
##  3rd Qu.:280.1    3rd Qu.:279.5      3rd Qu.:1.007
##  Max.   :287.0    Max.   :286.3      Max.   :1.041
##  NA's   :5392     NA's   :5392       NA's   :5392
## dimension(s):
##      from to     offset  delta                       refsys point values x/y
## date    1  8 2020-01-01 1 days                         Date    NA   NULL
## x       1 41      -0.05    0.1 Coordinate System importe...    NA   NULL [x]
## y       1 41      44.05   -0.1 Coordinate System importe...    NA   NULL [y]
```

Transforming CRS and cropping to AOI

Now we transform from the geographic to our target CRS. To do that, we want to use bilinear interpolation using GDAL. To do that, we need to create a template first and then transform using GDAL one attribute at a time.

```
alltemp_template <- st_warp(alltemp, crs = st_crs(catalonia)) # This uses nearest neighbour
# When using GDAL, we can only transform one attribute at a time
alltemp <- c(st_warp(alltemp["ERA5Land_2mtemp"], alltemp_template, method = "bilinear", use_gdal = T),
             st_warp(alltemp["ERA5Land_skintemp"], alltemp_template, method = "bilinear", use_gdal = T)
             st_warp(alltemp["ratio"], alltemp_template, method = "bilinear", use_gdal = T))
# When using GDAL, names and time dimension info are lost, so we redefine them
names(alltemp) <- names(alltemp_template)
alltemp <- st_set_dimensions(alltemp, 3, names = "date", values = dates_era5)
```

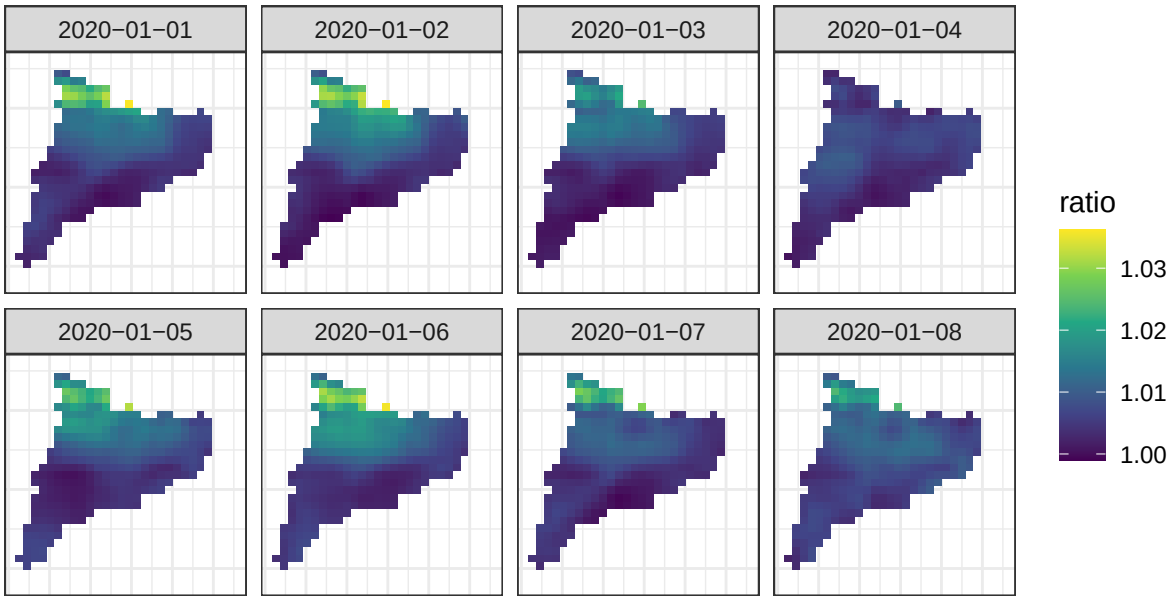
Now we can crop to our study area.

```
alltemp <- st_crop(alltemp, catalonia)
```

Using ggplot2

We can use stars objects in ggplot2 using `geom_stars` and use time dimension for faceting. Using `tmap` is also possible.

```
ggplot() +
  geom_stars(data=alltemp["ratio",,,1:8]) +
  facet_wrap(~ date, nrow = 2) +
  scale_fill_continuous(type = "viridis", na.value = "#FFFFFF00") +
  xlab("") + ylab("") +
  theme_bw() +
  theme(aspect.ratio=1, axis.ticks = element_blank(), axis.text = element_blank())
```



Extract pixel values

To extract pixel values at certain points, the raster stars objects is converted to a stars vector object, which we can easily transform into a data frame.

```
catalonia_centroid <- st_centroid(catalonia)
(alltemp_pxls <- st_extract(alltemp, catalonia_centroid))
```

```
## stars object with 2 dimensions and 3 attributes
## attribute(s):
##  ERA5Land_2mtemp  ERA5Land_skintemp      ratio
##  Min.   :275.0    Min.   :273.4      Min.   :1.000
##  1st Qu.:276.7    1st Qu.:275.0      1st Qu.:1.002
##  Median :277.6    Median :276.3      Median :1.004
##  Mean   :277.8    Mean   :276.6      Mean   :1.005
##  3rd Qu.:278.5    3rd Qu.:277.7      3rd Qu.:1.007
##  Max.   :281.6    Max.   :281.0      Max.   :1.013
## dimension(s):
##          from to     offset  delta                refsys point
## geometry    1  1         NA     NA ETRS89 / UTM zone 31N  TRUE
## date        1 31 2020-01-01 1 days                   Date    NA
##                          values
## geometry POINT (377366 4628291)
## date                       NULL
```

```
alltemp_pxls <- as.data.frame(alltemp_pxls)
head(alltemp_pxls)
```

```
##                 geometry       date ERA5Land_2mtemp ERA5Land_skintemp    ratio
## 1 POINT (377366 4628291) 2020-01-01        277.7917          275.8592 1.007229
## 2 POINT (377366 4628291) 2020-01-02        277.3808          274.9778 1.009005
## 3 POINT (377366 4628291) 2020-01-03        275.6949          274.6414 1.003984
## 4 POINT (377366 4628291) 2020-01-04        275.9717          275.0677 1.003448
## 5 POINT (377366 4628291) 2020-01-05        277.3432          275.6577 1.006277
## 6 POINT (377366 4628291) 2020-01-06        277.7176          275.6904 1.007603
```

Zonal statistics

If we want to calculate statistics by area, we'll use the `aggregate` function. Here we compute the mean value in our study area.

```
(zonal_means <- aggregate(alltemp, catalonia, FUN=mean, na.rm=T))
```

```
## stars object with 2 dimensions and 3 attributes
## attribute(s):
##  ERA5Land_2mtemp  ERA5Land_skintemp      ratio
##  Min.   :275.3    Min.   :273.4      Min.   :1.001
##  1st Qu.:277.0    1st Qu.:275.1      1st Qu.:1.004
##  Median :277.5    Median :276.0      Median :1.006
##  Mean   :277.8    Mean   :276.2      Mean   :1.006
##  3rd Qu.:278.3    3rd Qu.:277.1      3rd Qu.:1.008
##  Max.   :281.6    Max.   :280.4      Max.   :1.012
## dimension(s):
##          from to     offset  delta                refsys point
## geometry    1  1         NA     NA ETRS89 / UTM zone 31N FALSE
## date        1 31 2020-01-01 1 days                   Date    NA
##                                values
## geometry POLYGON ((386439 4561272, 3...
## date                             NULL
```

```
zonal_means <- as.data.frame(zonal_means)
head(zonal_means)
```

```
##                        geometry       date ERA5Land_2mtemp ERA5Land_skintemp
## 1 POLYGON ((386439.4 4561272,... 2020-01-01        277.3961          275.2150
## 2 POLYGON ((386439.4 4561272,... 2020-01-02        277.3204          275.0393
## 3 POLYGON ((386439.4 4561272,... 2020-01-03        276.9265          275.1874
## 4 POLYGON ((386439.4 4561272,... 2020-01-04        277.1000          275.7076
## 5 POLYGON ((386439.4 4561272,... 2020-01-05        277.9732          275.6854
## 6 POLYGON ((386439.4 4561272,... 2020-01-06        277.9402          275.3512
##      ratio
## 1 1.008173
## 2 1.008557
## 3 1.006492
## 4 1.005193
## 5 1.008518
## 6 1.009714
```