

StableClim - Process CMIP5 Historical and RCP

Stuart C Brown

Contents

Processing CMIP5 Historical and RCP simulations	1
Set-up	2
Creating the common grid	3
Splicing the historical simulations to the future forecasts	4
Convert the outputs to short and add scale_factor to the files	9

Processing CMIP5 Historical and RCP simulations

The code below splices the CMIP5 historical and RCP simulations together to provide a spatiotemporally unified modelled climate between 1850 and 2100.

The data is spliced following [Santer *et al.*](#), with the exception of the CNRM-CM5 model which we decided to drop from the analysis due to the splicing of the historical, historicalExt, and RCP simulations.

The code processes the data using [CDO tools](#) and [Windows Subsystem for Linux](#).

Some minor post-processing is performed using [NCO](#).

The code performs the following functions for each model:

- 1) Subset the data to the required years (as appropriate following Santer *et al.*)
- 2) Append the historical to the start of the RCP data
- 3) Take the ensemble average across all realisations for a given model
- 4) Regrid to a common 2.5°x2.5° grid using bilinear interpolation
- 5) Convert the data to annual averages/annual totals
- 6) Convert the units
- 7) Append some meta-data to the output NetCDF files

To achieve this requires matching the realisations that are available for each model under each RCP scenario to the corresponding historical simulation.

Set-up

Set up the various directories etc. that will be used in the processing. Need a list of model names, RCP scenarios, and variables to iterate through

```
library(raster)
library(ncdf4)
model_list <- list.dirs("../data/CMIP5/historical/ts/",
  full.names = FALSE,
  recursive = FALSE
)
# Remove the CNRM-CM5 model
## Issue with volcanic forcing (see https://doi.org/10.1073/pnas.1210514109)
model_list <- model_list[!grepl("CNRM-CM5", model_list, perl = TRUE)]
rcp_scenarios <- c("rcp26", "rcp45", "rcp60", "rcp85")
clim_vars <- c("ts", "pr")
```

Here we have access to 19 models. Each of the model outputs is named and identified with a triad of integers (**N**, **M**, **L**) for example “r1i1p1”. This identifier is used to distinguish among closely related simulations from a given model. The models have a different number of ensemble members (hereafter, realisations) available. We opted to use all available realisations (e.g. “r1”, “r2”, “r3”), provided they were from the same initial conditions (“i1”), and had the same physics (“p1”). This decision was made due to differences in prescribed and interactive physics with regards to aerosols and ozone which can impact air temperatures (Santer *et al.*).

Creating the common grid

This sections creates the common 2.5°x2.5° grid for the bilinear interpolation

```
r <- raster(res = 2.5, crs = crs(raster()))
r[] <- 1
xvals <- unique(values(init(r, "x")))
yvals <- unique(values(init(r, "y")))
nx <- length(xvals)
ny <- length(yvals)
lon <- ncdim_def("longitude", "degrees_east", xvals)
lat <- ncdim_def("latitude", "degrees_north", yvals)
mv <- 0
var_temp <- ncvar_def(
  name = "grid",
  units = "",
  dim = list(lon, lat),
  longname = "grid",
  missval = mv,
  prec = "byte"
)
ncout <- nc_create(
  filename = "C:/tmp/cdo_processing/dest_grid.nc",
  list(var_temp), force_v4 = TRUE
)
## put the data in the file
ncvar_put(
  nc = ncout,
  varid = var_temp,
  vals = values(r),
  start = c(1, 1),
  count = c(-1, -1)
)
nc_close(ncout)
```

Splicing the historical simulations to the future forecasts

This next section iterates through the code and matches the future forecast to the corresponding historical simulation. The files are then processed following the steps outlined above

```
baseComm <- "wsl cd /mnt/c/tmp/cdo_processing/;"

for (scen in rcp_scenarios) {
  for (var in clim_vars) {
    for (model in model_list) {
      ## find and copy the rcp files to the processing directory
      realDirs <- list.dirs(paste("../data/CMIP5", scen, var, model,
        sep = "/"
      ))[-1]
      if (length(realDirs) == 0) {
        ## not all models have data for all RCP's. Skip if missing.
        next()
      }
      realisations <- gsub("^.*/", "", realDirs)
      realisations <- realisations[grepl("i1p1$", realisations)]
      realFiles <- list.files(realDirs, "\\\\.nc$", full.names = TRUE)
      realFiles <- realFiles[grepl(
        x = realFiles, pattern = paste(realisations, collapse = "|")
      )]
      file.copy(realFiles, to = "C:/tmp/cdo_processing/")
      ## find and copy the historical files to the processing directory
      ## only copy those files with the same realisations
      histDirs <- list.dirs(paste("../data/CMIP5/Historical", var, model,
        sep = "/"
      ))[-1]
      histDirs <- histDirs[grepl(histDirs,
        pattern = paste(paste0(realisations, "$"),
          collapse = "|"
        )
      )]
      stopifnot(length(histDirs) == length(realisations))
      histFiles <- list.files(histDirs, "\\\\.nc$", full.names = TRUE)
      file.copy(histFiles, to = "C:/tmp/cdo_processing/")
      ## Now use CDO to process the files
      for (real in realisations) {
        files <- list.files("C:/tmp/cdo_processing/", real, full.names = FALSE)
        histFiles <- files[grepl("historical", files)]
        futureFiles <- files[grepl(scen, files)]
        ## historical
        ### Hadley and GFDL models have different sim length.
        ### Hadley historical sims run from 1859-12:2005-11
        ### need to add empty layers to start of file to bring it back
        ### to 1850-01 (9 years 11 months, 119 months)
      }
    }
  }
}
```

```

if (grepl(x = model, pattern = c("Had")))) {
  comm <- paste(
    baseComm, "cdo -setreftime,1850-01-16,00:00,1month
    -settaxis,1850-01-16,00:00,1month -setctomiss,-999 -setvals,0,-999
    -mulc,0 -copy -selimestep,1/119", histFiles[1],
    paste0("ts_Amon_", model, "_historical_", real, "_185001-185911.nc")
  )
  shell(comm, mustWork = TRUE)
  files <- list.files("C:/tmp/cdo_processing/", real,
    full.names = FALSE
  )
  histFiles <- files[grepl("historical", files)]
}

### GFDL sims run from 186001-200512
### Need to add in 10 years (120 months)
if (grepl(x = model, pattern = c("GFDL")))) {
  comm <- paste(
    baseComm, "cdo -cat",
    paste(histFiles[1:5], collapse = " "),
    "historical_temp.nc"
  )
  shell(comm, mustWork = TRUE)
  comm <- paste(
    baseComm, "cdo -setreftime,1850-01-16,00:00,1month
    -settaxis,1850-01-16,00:00,1month -setctomiss,-9999
    -setvals,0,-9999 -mulc,0 -copy -selimestep,1/120",
    "historical_temp.nc",
    paste0(
      "ts_Amon_", model, "_historical_",
      real, "_185001-185912.nc"
    )
  )
  shell(comm, mustWork = TRUE)
  file.remove("C:/tmp/cdo_processing/historical_temp.nc")
  files <- list.files("C:/tmp/cdo_processing/", real,
    full.names = FALSE
  )
  histFiles <- files[grepl("historical", files)]
}

## Deal with models that have historicalExt runs
## limits historical sims to 1850:2005
if (grepl(x = model, pattern = c("BCC-CSM1.1|MIROC5")))) {
  comm <- paste(
    baseComm, "cdo selyear,1850/2005", histFiles,
    paste0(
      "ts_Amon_", model, "_historical_", real,
      "_185001-200512.nc"
    )
  )
  shell(comm, mustWork = TRUE)
  file.remove("C:/tmp/cdo_processing/historical_ext.nc")
  files <- list.files("C:/tmp/cdo_processing/", real,
    full.names = FALSE
  )
  histFiles <- files[grepl("historical", files)]
}

```

```

    )
  )
  shell(comm, mustWork = TRUE)
  file.remove(paste0("C:/tmp/cdo_processing/", histFiles))
  files <- list.files("C:/tmp/cdo_processing/", real,
    full.names = FALSE
  )
  histFiles <- files[grepl("historical", files)]
}
if (length(histFiles) > 1) {
  comm <- paste(baseComm, "cdo -cat",
    paste(histFiles, collapse = " "),
    paste0(
      paste("historical", var, model, real, sep = "_"),
      ".nc"
    ),
    sep = " "
  )
  ### escape any brackets
  comm <- gsub(pattern = "\\(", replacement = "\\\\(", x = comm)
  comm <- gsub(pattern = "\\)", replacement = "\\\\)", x = comm)
  shell(comm, mustWork = TRUE)
  histFiles <- list.files("C:/tmp/cdo_processing/", real,
    full.names = FALSE
  )
  histFiles <- histFiles[grepl(
    x = histFiles,
    pattern = paste0("historical_", var, "_", model, "_", real)
  )]
}
## Splice the historical and RCP simulations
inFiles <- paste(c(histFiles, futureFiles), collapse = " ")
outFile <- paste0(paste("spliced_historical", scen, var, model,
  real,
  sep = "_"
), ".nc")
comm <- paste(baseComm, "cdo cat", inFiles, outFile, sep = " ")
comm <- gsub(pattern = "\\(", replacement = "\\\\(", x = comm)
comm <- gsub(pattern = "\\)", replacement = "\\\\)", x = comm)
shell(comm, mustWork = TRUE)
}
## Now we take the ensemble average across all realisations
## for a given model.
## Can swap ensmean for ensctl, 50 for median
splicedSims <- paste(
  list.files("C:/tmp/cdo_processing/", "spliced", full.names = FALSE),
  collapse = " "

```

```

)
outFile <- paste0(paste("ensAvg_origGrid_historical", scen, var, model,
  sep = "_"
), ".nc")
comm <- paste(baseComm, "cdo ensmean -selyear,1850/2100", splicedSims,
  outFile,
  sep = " ")
)
comm <- gsub(pattern = "\\(", replacement = "\\\\(", x = comm)
comm <- gsub(pattern = "\\)", replacement = "\\\\)", comm)
shell(comm, mustWork = TRUE)
## Regrid to common 2.5°x2.5° grid and convert to annual averages
## convert units at the same time
ensFile <- outFile
outFile <- paste0(paste("ensAvg_regridAnnAvg_historical", scen, var, model,
  sep = "_"
), ".nc")
fConv <- if (var == "ts") {
  ### convert ts from K to °C
  "-subc,273.15 -selname,ts"
} else if (var == "pr") {
  ### Convert pr from kg m2 s1 to mm/day
  "-mulc,86400 -selname,pr"
}
stUnit <- if (var == "ts") {
  "-setunit,'degC' -setname,ts"
} else {
  "-setunit,'mm/day' -setname,pr"
}
## The command for all of the processing steps
comm <- paste(baseComm, "cdo -f nc4 --cmor -k grid -b F32
  setmissval,-9999 -settaxis,1850-06-16,00:00,1year",
  stUnit, "-yearmean",
  fConv, "-remapbil,dest_grid.nc", ensFile, outFile,
  sep = " ")
)
comm <- gsub(pattern = "\\(", replacement = "\\\\(", x = comm)
comm <- gsub(pattern = "\\)", replacement = "\\\\)", comm)
shell(comm, mustWork = TRUE)
## Now clean up the files
## Copy the two output files to a new directory
moveFiles <- paste0("C:/tmp/cdo_processing/", c(ensFile, outFile))
file.copy(moveFiles, "C:/tmp/cdo_outputs/", overwrite = TRUE)
## delete all files from processing directory
rmFiles <- list.files("C:/tmp/cdo_processing/", "\\*.nc$",
  full.names = TRUE
)

```

```
rmFiles <- rmFiles[!grepl("dest_grid.nc", rmFiles)]  
file.remove(rmFiles)  
}  
}  
}
```

All the CMIP5 RCP simulations have now been spliced with the corresponding historical simulation and converted to ensemble averages giving 19 simulations of climate at annual time-steps between 1850 and 2100. Importantly the data are on a common-grid enabling comparisons between simulations, and to simulated climate in the past (from the TraCE-21ka simulation).

Convert the outputs to short and add scale_factor to the files

To save disk-space the output files are converted to short integers using NCO tools, with the `scale_factor` attribute added to the header for each of the files.

This process reduces the output file size from ~36GB to ~18GB, and retains 2 decimal places of precision.

```
cd /mnt/c/tmp/cdo_outputs
ls | egrep '_ts_' | egrep 'regridAnnAvg' | while read file; do
    ncap2 -O -s 'ts=short(ts/0.01);ts@scale_factor=0.01' $file $file
    ncks -L 5 -O $file $file
done
```

```
cd /mnt/c/tmp/cdo_outputs
ls | egrep '_pr_' | egrep 'regridAnnAvg' | while read file; do
    ncap2 -O -s 'pr=short(pr/0.01);pr@scale_factor=0.01' $file $file
    ncks -L 5 -O $file $file
done
```