

# prepare\_CMIP6\_data

November 4, 2024

## 1 Prepare CMIP6 Data on Capser

```
[1]: import pandas as pd
import xarray as xr
import numpy as np
import os.path
import glob

import matplotlib.pyplot as plt
# Useful for plotting maps.
import cartopy.crs as ccrs

# This can be useful for working with multiple processors - to be explored
  ↪ later on.
# from dask.distributed import Client, LocalCluster.
```

```
[2]: def get_MIP(experiment):
    """
    Utility function to get the activity associated with a particular
    ↪ experiment.
    """
    if experiment == 'ssp245-covid':
        return 'DAMIP'
    elif experiment == 'ssp370-lowNTCF':
        return 'AerChemMIP'
    elif experiment.startswith('ssp'):
        return 'ScenarioMIP'
    elif experiment.startswith('hist-'):
        return 'DAMIP'
    else:
        return 'CMIP'
```

```
[3]: overwrite = False

model = 'NorESM2-LM'
experiments = [
    '1pctCO2', 'abrupt-4xCO2', 'historical', 'piControl', # CMIP.
```

```

        'hist-GHG', 'hist-aer', # DAMIP (hist-aer not found in DAMIP).
        'ssp126', 'ssp245', 'ssp370', 'ssp370-lowNTCF', 'ssp585' #
    ↪ScenarioMIP.
]
variables = [
    'tas', 'pr', 'tasmin', 'tasmax',
]

```

```

[4]: def get_data(variable, experiment, member):
    """
    Read a particular CMIP6 (Amon) variable from NorESM2.
    """
    files = glob.glob(f"/glade/collections/cmip/CMIP6/{get_MIP(experiment)}/NCC/
    ↪NorESM2-LM/{experiment}/{member}/Amon/{variable}/gn/v20190815/{variable}/*.
    ↪nc")
    if files == []:
        return (True, f"{variable} is not found.")
    return (False, xr.open_mfdataset(files)[variable])

def get_daily_data(variable, experiment, member):
    """
    Read a particular CMIP6 (day) variable from NorESM2.
    """
    files = glob.glob(f"/glade/collections/cmip/CMIP6/{get_MIP(experiment)}/NCC/
    ↪NorESM2-LM/{experiment}/{member}/day/{variable}/gn/v20190815/{variable}/*.
    ↪nc")
    if files == []:
        return (True, f"{variable} is not found.")
    return (False, xr.open_mfdataset(files)[variable])

```

```

[5]: for experiment in experiments:
    # Just take three ensemble members (there are more in the COVID simulation,
    ↪but we don't need them all).
    for i in range(3):
        physics = 2 if experiment == 'ssp245-covid' else 1 # The COVID
        ↪simulation uses a different physics setup.
        # TODO - check the differences...
        member = f"r{i+1}i1p1f{physics}"
        print(f"Processing {member} of {experiment}...")
        outfile = f"{model}_{experiment}_{member}.nc"
        if (not overwrite) and os.path.isfile(outfile):
            print("File already exists, skipping.")
            continue

    # Call function to get variable data.

```

```

tasmin = get_daily_data('tasmin', experiment, member)
tasmax = get_daily_data('tasmax', experiment, member)
tas = get_data('tas', experiment, member)
pr = get_data('pr', experiment, member)

# Array containing booleans for which variables have data.
avail_arr = [tasmin[0], tasmax[0], tas[0], pr[0]]

# Give information about which variables/directory could not be found.
if any(avail_arr):
    if all(avail_arr):
        print('No variables/directory found.')
    else:
        for var in [tasmin, tasmax, tas, pr]:
            if var[0]:
                print(var[1])
        print("Skipping this realisation as no data present")
        continue

# Obtain only the data without boolean flag.
tasmin = tasmin[1]
tasmax = tasmax[1]
tas = tas[1]
pr = pr[1].persist() # Since we need to process it twice.

# Derive additional vars.
dtr = tasmax-tasmin
pr = pr.chunk(dict(time=-1)) # Rechunk time into a single chunk.

# Process dataset.
ds = xr.Dataset({'diurnal_temperature_range': dtr.groupby('time.year').
↪mean('time'),
                'tas': tas.groupby('time.year').mean('time'),
                'pr': pr.groupby('time.year').mean('time'),
                'pr90': pr.groupby('time.year').quantile(0.9,
↪skipna=True)})
ds.to_netcdf(f"{model}_{experiment}_{member}.nc")

```

```

Processing r1i1p1f1 of 1pctCO2...
tasmin is not found.
tasmax is not found.
Skipping this realisation as no data present
Processing r2i1p1f1 of 1pctCO2...
No variables/directory found.
Skipping this realisation as no data present
Processing r3i1p1f1 of 1pctCO2...
No variables/directory found.

```

Skipping this realisation as no data present  
 Processing r1i1p1f1 of abrupt-4xC02...  
 tasmin is not found.  
 tasmax is not found.  
 Skipping this realisation as no data present  
 Processing r2i1p1f1 of abrupt-4xC02...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of abrupt-4xC02...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of historical...  
 File already exists, skipping.  
 Processing r2i1p1f1 of historical...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of historical...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of piControl...  
 File already exists, skipping.  
 Processing r2i1p1f1 of piControl...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of piControl...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of hist-GHG...  
 tasmin is not found.  
 tasmax is not found.  
 Skipping this realisation as no data present  
 Processing r2i1p1f1 of hist-GHG...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of hist-GHG...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of hist-aer...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r2i1p1f1 of hist-aer...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of hist-aer...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of ssp126...  
 No variables/directory found.

Skipping this realisation as no data present  
 Processing r2i1p1f1 of ssp126...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of ssp126...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of ssp245...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r2i1p1f1 of ssp245...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of ssp245...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of ssp370...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r2i1p1f1 of ssp370...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of ssp370...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of ssp370-lowNTCF...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r2i1p1f1 of ssp370-lowNTCF...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of ssp370-lowNTCF...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r1i1p1f1 of ssp585...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r2i1p1f1 of ssp585...  
 No variables/directory found.  
 Skipping this realisation as no data present  
 Processing r3i1p1f1 of ssp585...  
 No variables/directory found.  
 Skipping this realisation as no data present