Bias-Corrected CMIP5 CESM Data in WRF/MPAS Intermediate File Format

Cindy L. Bruyère
Andrew J. Monaghan
Daniel F. Steinhoff
David Yates

NCAR Technical Notes
NCAR/TN 515+STR

National Center for Atmospheric Research
P. O. Box 3000
Boulder, Colorado
80307-3000
www.ucar.edu

NCAR | UCAR
National Center for Atmospheric Research

National Science Foundation
NCAR IS SPONSORED BY THE NSF

NCAR TECHNICAL NOTES

http://library.ucar.edu/research/publish-technote

The Technical Notes series provides an outlet for a variety of NCAR Manuscripts that contribute in specialized ways to the body of scientific knowledge but that are not yet at a point of a formal journal, monograph or book publication. Reports in this series are issued by the NCAR scientific divisions, serviced by OpenSky and operated through the NCAR Library. Designation symbols for the series include:

EDD – Engineering, Design, or Development Reports
Equipment descriptions, test results, instrumentation, and operating and maintenance manuals.

IA – Instructional Aids
Instruction manuals, bibliographies, film supplements, and other research or instructional aids.

PPR – Program Progress Reports
Field program reports, interim and working reports, survey reports, and plans for experiments.

PROC – Proceedings
Documentation or symposia, colloquia, conferences, workshops, and lectures. (Distribution maybe limited to attendees).

STR – Scientific and Technical Reports
Data compilations, theoretical and numerical investigations, and experimental results.

The National Center for Atmospheric Research (NCAR) is operated by the nonprofit University Corporation for Atmospheric Research (UCAR) under the sponsorship of the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

National Center for Atmospheric Research
P. O. Box 3000
Boulder, Colorado 80307-3000

NCAR/TN-515+STR
NCAR Technical Note

March 2015

Bias-Corrected CMIP5 CESM Data in WRF/MPAS Intermediate File Format

Cindy L. Bruyère
Mesoscale and Microscale Meteorology Laboratory, NCAR, Boulder, CO, USA
Andrew J. Monaghan
Research Applications Laboratory, NCAR, Boulder, CO, USA
Daniel F. Steinhoff
Research Applications Laboratory, NCAR, Boulder, CO, USA
David Yates
Research Applications Laboratory, NCAR, Boulder, CO, USA

Mesoscale and Microscale Meteorology Laboratory

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
P. O. Box 3000
BOULDER, COLORADO 80307-3000
ISSN Print Edition 2153-2397
ISSN Electronic Edition 2153-2400

Table of Contents

1. List of Figures

2. List of Tables

3. Overview

This technical note describes the generation of global bias corrected climate model output files from version 1 of NCAR's Community Earth System Model (CESM; Hurrell et al. 2013) that participated in phase 5 of the Coupled Model Intercomparison Experiment (CMIP5), which supported the Intergovernmental Panel on Climate Change Fifth Assessment Report (IPCC AR5). The dataset contains all the variables needed for the initial and boundary conditions for simulations with the Weather Research and Forecasting model (WRF; Skamarock et al. 2008) or the Model for Prediction Across Scales (MPAS; Ringler et al. 2008, Thuburn et al. 2009, and Skamarock et al. 2012), provided in the Intermediate File Format specific to WRF and MPAS. The data are interpolated to 26 pressure levels, have a horizontal resolution of approximately 1**°,** and are provided in files at six hourly intervals. The variables have been bias corrected using the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim; Simmons et al. 2006 and Dee et al. 2011) fields for 1981-2005, following the method in Bruyère et al. (2014). Files are available for a 20th Century simulation (1951-2005) and three concomitant Representative Concentration Pathway (RCP) future scenarios (RCP4.5, RCP6.0 and RCP8.5) spanning 2006-2100, and are freely available from Monaghan et al. (2014).

Section 5 describes the dataset and its creation; section 6 describes how to use this dataset as input to WRF; while section 7 describes the software that was used to create the dataset, and a software user guide to create new datasets.

4. Acknowledgements

NCAR is funded by the National Science Foundation. The CESM project is supported by the National Science Foundation and the Office of Science of the U.S. Department of Energy, and is maintained by the Climate and Global Dynamics Division at NCAR. ECWMF, the producer of ERA-Interim, is supported by contributions from its 34 Member and Co-operating states. The version of ERA-Interim used within was reformatted and provided by the Research Data Archive within the Computational and Information Systems Laboratory at NCAR. This work was partially supported by the Environment Agency of Abu Dhabi under Contract EAD/SL/01/2014 with the Climate Change Research Group; the Research Partnership to Secure Energy for America (RPSEA); NSF EASM Grants AGS-1048841 and AGS-1048829; and DNV GL.

5. Data Description

a) CMIP5 CESM Data

Intermediate files that can be used as the initial and boundary conditions for the Weather Research and Forecasting model (WRF; Skamarock et al. 2008) and the Model for Prediction Across Scales (MPAS; Ringler et al. 2008, Thuburn et al. 2009, and Skamarock et al. 2012) are constructed using simulations from a subset (*viz., Version 4 of the Community Climate System Model - CCSM4; Gent et al. 2011*) of Version 1 of the Community Earth System Model (CESM; Hurrell et al. 2013). CESM is a coupled global climate model (GCM) comprised of four component models that simulate the atmosphere, ocean, land surface and sea-ice. The CESM simulations used to generate the present dataset were performed in support of the Coupled Model Intercomparison Experiment Phase 5 (CMIP5; Taylor et al. 2012) and the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC 2013). CESM ranks at the top of all CMIP5 GCMs in its ability to simulate global patterns of observed temperature and rainfall (Knutti et al. 2013).

Model fields are obtained from the National Center for Atmospheric Research's Globally Accessible Storage Environment (GLADE). The dataset is also available from NCAR's CISL Research Data Archive (http://rda.ucar.edu/datasets/ds316.0), and from the Earth System Grid - Program for Climate Model Diagnosis and Intercomparison (ESG-PCMDI) gateway at Lawrence Livermore National Laboratory (http://pcmdi3.llnl.gov/esgcet/home.htm). The CESM simulations used to construct the present bias-corrected dataset include a historical simulation and three future projections. A 20th century simulation ("20THC") was forced by observed natural and anthropogenic atmospheric composition changes spanning 1861-2005. The future projections are generated under Representative Concentration Pathways (RCPs; Moss et al. 2010) 4.5, 6.0 and 8.5,spanning 2006-2100. RCP4.5 is a low-to-moderate emissions scenario with GHG radiative forcing reaching 4.5 W m^{-2} near 2100. It represents a trajectory that may be plausible if, for instance, GHG emissions pricing were introduced in order to limit radiative forcing (Thomson et al. 2011). RCP6.0 is a moderate GHG emissions scenario that is similar to RCP4.5 in that a variety of strategies for reducing GHGs would be applied to eventually stabilize radiative forcing near the end of the 21st century (Masui et al. 2011). RCP8.5 is a high-emissions scenario with greenhouse-gas (GHG) radiative forcing reaching 8.5 W m^{-2} near 2100. It represents a plausible trajectory if little is done to curb greenhouse gas emissions (Riahi et al. 2011).

While an ensemble of CESM simulations was performed for each scenario to characterize model uncertainty, only output from ensemble member #6 -- also known as the "Mother of All Runs (MOAR)" -- was used to construct the present files, because that is the only member that has the full three-dimensional fields required to force WRF or MPAS available at 6-hourly intervals. The specific CESM ensemble member #6 simulations used were: b40.20th.track1.1deg.012 (20th Century), b40.rcp4_5.1deg.006 (RCP4.5), b40.rcp6_0.1deg.006 (RCP6.0) and b40.rcp8_5.1deg.007 (RCP8.5). Additional details at: http://www.cesm.ucar.edu/experiments/cesm1.0/.

b) Bias-Correcting CMIP5 CESM Data

Like all GCMs, CESM contains regional-scale biases due to having coarse spatial resolution and a limited representation of some physical processes. Such biases can adversely affect the dynamical downscaling process and contribute to uncertainty. To remedy these biases, it is common to *bias correct* the climate model output before using it to drive regional-scale models like WRF (e.g., Rasmussen et al. 2011, Xu and Yang 2012, Done et al. 2013). The present Intermediate Formatted Files (referred to as CESM-Int; Monaghan et al. 2014) were constructed with a recently-developed bias correction method (Bruyère et al. 2014) that uses global atmospheric and surface reanalysis to correct for the mean bias in the CESM 3-dimensional temperature, geopotential height, wind, and humidity fields, as well as the surface pressure, sea level pressure, sea surface temperature (SST), skin temperature, and soil temperature and moisture fields (Bruyère et al. 2014). The bias correction method corrects the mean state while retaining synoptic-scale and climate-scale variability as simulated by CESM. The reanalysis used for the bias correction is the European Centre for Medium-Range Weather Forecasting (ECMWF) Interim Reanalysis (ERAI; Simmons et al. 2006 and Dee et al. 2011), obtained from NCAR's Research Data Archive for surface and pressure levels (ECMWF 2009, 2012). ERAI is considered one of the most accurate atmospheric reanalyses available at the present time (e.g., Lorenz and Kunstmann 2010). The bias-corrected CESM output is produced by combining a 25-year (1981-2005) mean annual cycle from ERAI and a 6-hourly perturbation term from CESM, as follows:

$$\boldsymbol{CESM} = \overline{\boldsymbol{CESM}} + \boldsymbol{CESM}'$$
$$\boldsymbol{ERAI} = \overline{\boldsymbol{ERAI}} + \boldsymbol{ERAI}'$$
$$\boldsymbol{CESM_R} = \overline{\boldsymbol{ERAI}} + \boldsymbol{CESM}'$$

where overbar terms are the 25-year mean annual cycle, primed terms are perturbations from the mean annual cycle, and $CESM_R$ is the revised (bias-corrected) CESM output at 6-hourly intervals. $CESM_R$, which can be used as the initial and boundary conditions for WRF and MPAS climate simulations, combines a base, seasonally varying climate provided by ERAI with day-to-day weather, climate variability and change provided by CESM. Fig. 1 shows a schematic of the bias-correction method. Here, as an example, a single grid point from the SST field is shown. The solid black line depicts the ERAI 25-year mean, showing the slow varying mean value of SST over an annual cycle. Similarly the dashed grey line depicts the CESM historical long-term mean. The CESM model has a mean cold bias, and although CESM has a similar annual cycle as the ERAI data, the mean bias varies from month to month. The colored lines show random historical CESM years before (dashed lines) and after (solid lines) bias correction. It is clear that the bias correction method just removes the mean annual bias, while retaining the day-to-day climate variability from CESM.

Using the bias-correction technique, 6-hourly Intermediate files were generated for the historical period and the three RCP scenarios. These bias corrected Intermediate files (CESM-Int; Monaghan et al. 2014) are available from NCAR's CISL Research Data Archive (http://rda.ucar.edu/datasets/ds316.1), and also can be accessed directly from NCAR's GLADE system (/glade/p/rda/data/ds316.1).

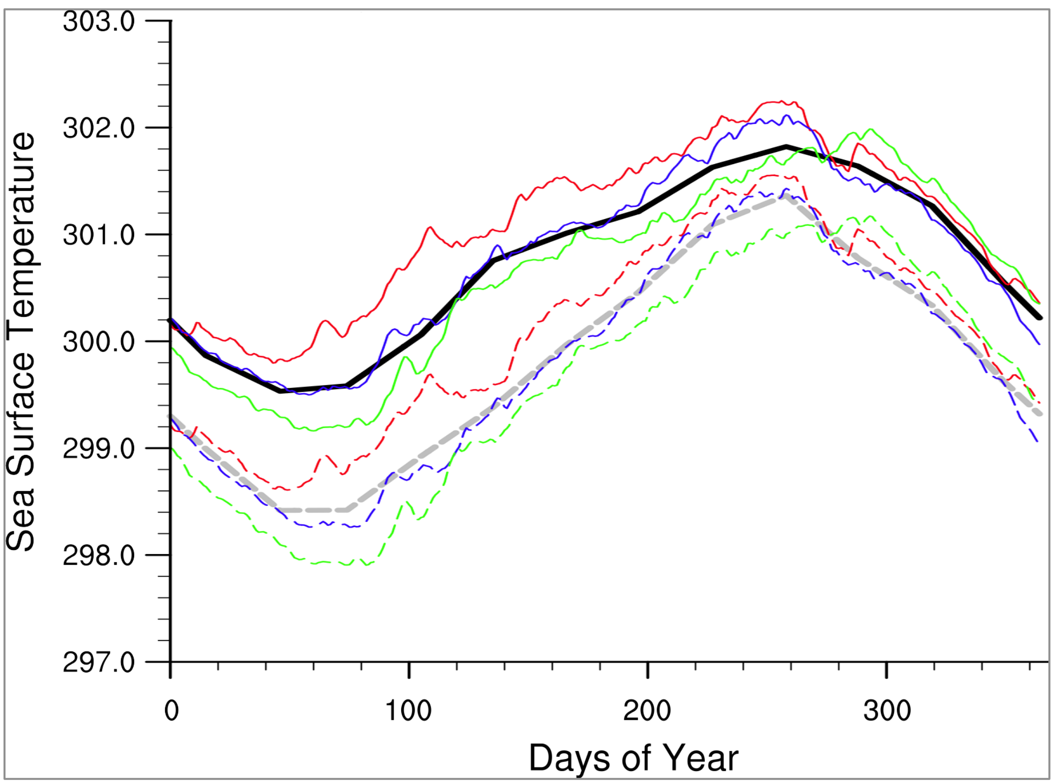

Figure 1: Schematic of the bias-correction method. The solid black line is the ERAI 25-year mean annual cycle; the dashed grey line is the CESM historical long-term mean annual cycle; while the colored lines are random historical CESM years before (dashed lines) and after (solid lines) bias correction.

6. Using CESM Intermediate files with WPS/WRF and MPAS

Directly using CESM-Int provided on the RDA site (http://rda.ucar.edu/datasets/ds316.1), or processing your own via the software described in section 7, replaces the **ungrib.exe** program in WPS. All other steps of WPS and WRF are still required. A typical workflow for WPS/WRF would be:

--Run **geogrid.exe** after specifying all of your namelist.wps parameters. namelist.wps (and possibly GEOGRID.TBL) should be modified to allow for the inland lake surface type, which in turn will use the "TAVGSFC" variable to initialize lake surface temperatures. Otherwise, lake surface temperatures should be checked to ensure they look reasonable. Section "***Alternative Initialization of Lake SSTs***" in Chapter 3 of the WRF Users' Guide: http://www.mmm.ucar.edu/wrf/users/docs/user_guide_V3/users_guide_chap3.htm provides instructions on how to do this.

If you choose the inland lakes option, you ***do not*** have to run tavg_sfc.exe per the instructions in the WRF Users' Guide, as this is already done in the processing scripts, and TAVGSFC is already available to you within the CESM-Int files (therefore, you also ***do not*** need to add "TAVGSFC" to the 'constants' section of namelist.wps as instructed in the WRF Users' Guide). You may also want to set "geog_data_res = 'modis_lakes+30s'," or "geog_data_res = 'usgs_lakes+30s'," in the **&geogrid** section of namelist.wps so that you get the inland lakes surface type. This is essential if you plan to use the TAVGSFC field to diagnose inland lake temperatures.

-- Instead of running ***ungrib.exe*****,** either 1) directly use the Intermediate files (CEMS-Int) provided on NCAR's Research Data Archive (http://rda.ucar.edu/datasets/ds316.1), or from NCAR's glade system (/glade/p/rda/data/ds316.1); or 2) run the CESM-to-Intermediate-Files software described in section 7 to create your own Intermediate files.

-- Run **metgrid.exe**, properly specifying the names of the Intermediate files in namelist.wps. You may also want to modify METGRID.TBL to optimize interpolation of SSTs.

Modifying METGRID.TBL might require some experimentation with some of the interpolation options to refine your subsequent ingest of the Intermediate files into **metgrid.exe**, and to make sure you get fields in your met_em files that look right. In particular, masked data -- SST and SEAICE -- interpolation can be tricky near the coastlines when using the comparatively low-resolution data from CESM. Here are some options (modifications to METGRID.TBL) that seem to work well:

```
========================================================
name=SST
        interp_option=sixteen_pt+wt_average_16pt+search
        masked=land
        interp_mask = LANDSEA(1)
        fill_missing=-1.E+30
        flag_in_output=FLAG_SST
        missing_value=-1.E+30
========================================================
```

```
========================================================
name=SEAICE
        interp_option=four_pt+average_4pt
        interp_mask=LANDSEA(1)
        masked=land
        missing_value=-1.E30
        fill_missing=0.
========================================================
```

The "search" interpolation option for SST will mean that any *inland* lake will take on the SST value from the nearest ocean grid point - this is usually inaccurate. Therefore, if you use this option, make sure to use the inland lakes option ("TAVGSFC") discussed above in the running **geogrid.exe** section.

The CESM-Int data (Monaghan et al. 2014) have the following naming conventions:

Historical Period (1951-2005): CCSM4_CMIP5_MOAR_BC_20THC:YYYY-MM-DD_HH
RCP4.5 Scenario (2006-2100): CCSM4_CMIP5_MOAR_BC_RCP45:YYYY-MM-DD_HH
RCP6.0 Scenario (2006-2100): CCSM4_CMIP5_MOAR_BC_RCP60:YYYY-MM-DD_HH
RCP8.5 Scenario (2006-2100): CCSM4_CMIP5_MOAR_BC_RCP85:YYYY-MM-DD_HH

where YYYY is the year, MM the month, DD the day and HH the hour. The fields available in each file are described in Table 1.

In your namelist.wps file, the following is the correct way to specify the **&metgrid** section (note that everything you should need to run the model is in the Intermediate files listed above, unless you want to use some other fields of your own. Note also that "CASE" means "20THC", "RCP45", "RCP60", or "RCP85"):

```
&metgrid
 fg_name= ' CCSM4_CMIP5_MOAR_BC_CASE' ,
 constants_name = ,
 io_form_metgrid = 2,
/
```

-- Run **real.exe** from the WRF directory once you have successfully created your met_em files. Note that you may have to modify "NUM_LAND_CAT" in namelist.input to reflect the new "lake" land surface type in your dataset. You can find the value of NUM_LAND_CAT from issuing ncdump -h on any met_em file.

-- Run **wrf.exe**. CESM data have ***no leap years***. For simulations that span leap years, add -DNO_LEAP_CALENDAR to "ARCH_LOCAL" in the configure.wrf file before compiling WRF. Similarly, for WPS add the -DNO_LEAP_CALENDAR flag to the list of CPPFLAGS in the configure.wps file before compiling. Alternatively, run **metgrid.exe** up to 18Z February 28, and then restart **metgrid.exe** again at 00Z March 1 for leap years.

Downscaling GCM data has a different set of simulation choices than using WRF for forecasting or hindcasting, since those simulations involve driving WRF with reanalysis or model forecast data rather than GCM data. Firstly, some users prefer not to do cold-starts (re-initializations) of WRF every few days as is common for downscaling reanalysis or model forecast data. Instead, they might run an entire 20-year period continuously, without any cold starts (in which case they use intermittently-written "RESTART" files to get around wall clock constraints). One advantage of running continuous simulations is to allow the model soil state to spin up via the land surface model (e.g., the Noah LSM that is coupled to WRF). It typically takes about a year for soil fields to fully spin up. Another advantage of continuous simulations is that they don't require as much effort for preprocessing, because there are fewer cold-starts. Since these are climate model runs, the typical reasons for doing frequent cold starts (e.g., using high fidelity initial conditions to constrain model accuracy with respect to the large-scale driving fields) are not generally first-order considerations. However, you should keep a few things in mind if you choose to do long-term (i.e., longer than a month) simulations:

1. Make sure you regularly update SST using the "sst_update" flag in your WRF namelist.input file. Otherwise, you might be using January SSTs in July if you initialized a run in January of that year. The CESM-Int data provide 6-hourly TAVGSFC fields so that the inland lakes can be updated along with SSTs (note that although TAVGSFC is written out at 6-hourly intervals for convenience and consistency with other fields, TAVGSFC is actually the *monthly average* skin temperature and therefore only changes on the first day of each month; this approach

prevents spuriously large diurnal or day-to-day fluctuations of the lake surface temperatures that can otherwise occur).

2. You may want to run a "spin-up" year prior to your period-of-interest and then throw it out. This will allow full spin up of the soil state.

3. Fractional sea ice from CESM has been included in the Intermediate files. As mentioned below, interpolation of sea ice from the comparatively coarse resolution of the CESM domain to your WRF domain can be tricky. Therefore, be sure to check your sea ice fields in the wrflowbdy_d0X file once you have interpolated everything onto the WRF domain, to make sure they look realistic, and especially to make sure there aren't issues near coastlines, where masking differences between CESM and WRF can be problematic. If you choose not to use fractional sea ice, make sure to set " seaice_threshold = 271.35 " (-1.8 C) in namelist.input, which diagnoses the existence of sea ice at SSTs less than 271.35. Note that the SEAICE and SNOW fields have not been bias corrected.

4. For very long simulations, consider increasing the boundary width to 10 (from the default of 5), and use a combined linear-exponential decay in the relaxation zone (i.e., set spec_exp = 0.33).

5. Consider using adaptive time-stepping to keep the model stable, while also allowing for longer time-steps during convectively stable periods. This will optimize run times, while preventing model blow-ups during convectively active periods.

6. The model uses a climatological value for deep soil, which might be unrealistic for long simulations. For multi-year simulations, consider updating the soil moisture *(tmn_update = 1),* using a lag of approximately 150 days *(lagday).*

7. SST input typically has coarse temporal resolution. As SST changes very slowly, this is not a problem in of itself. But this does mean that the input SST lacks any diurnal signal. Using the *sst_skin=1* option in the namelist.input file will calculate and add an appropriate diurnal signal to the input SST data, allowing for better interaction with clouds and radiative terms in the model.

8. Multi-year simulations could result in heat build-up in the upper atmosphere if the input data does not extend high enough into the troposphere. Ensuring that the input data extends to at least 10 hPa will prevent this issue. Also make sure this is explicitly set in the namelist.input file, and set *ptop_requested to at least 1000* (10 hPa). Note that *ptop_requested* cannot be set to a pressure level that is not present in the input data.

9. Finally, consider using buckets for precipitation (bucket_mm) and energy (bucket_J) accumulation. This will prevent loss of precision for long simulations due to adding very small amounts to large numbers.

Some of the programs in the WPS/util directory may be useful for checking your Intermediate files: rd_intermediate.exe (for file contents) and plotfmt.ncl (for plotting Intermediate data for a quick look).

A typical workflow for MPAS would be:

-- Compile the MPAS source code with the option *CORE=init_atmosphere*. This will create a single executable (*init_atmosphere_model*) that is used to create initial conditions for any MPAS mesh.

-- Interpolate the MPAS static fields available in WPS_GEOG to the MPAS mesh.

-- Then, directly use the Intermediate files (CEMS-Int) provided on NCAR's Research Data Archive (http://rda.ucar.edu/datasets/ds316.1), or from NCAR's glade system (/glade/p/rda/data/ds316.1), and horizontally and vertically interpolate the 2-dimensional and 3-dimensional meteorological fields to the MPAS grid *(for more information see the subroutine init_atm_case_gfs in module mpas_init_atm_cases.F).*

All basic namelist parameters are contained in the file namelist.init_atmosphere. For more information obtain the MPAS User's Guide at: http://mpas-dev.github.io

Table 1. Description of variables in the 6-hourly bias-corrected Intermediate files.

FIELD	LEVEL			UNITS	DESCRIPTION	BIAS-CORRECTED?
	200100	201300	PLEVS*			
LANDSEA	X			--	The time-invariant land sea mask (1=land or 0=water) from CESM.	N
SOILHGT	X			m	The time-invariant terrain elevation data from CESM.	N
SKINTEMP	X			K	Monthly average skin temperature expressed every six hours (e.g., all time steps in Jan 2002 are the same, then SKINTEMP changes for Feb 2002, etc). This variable is relatively unimportant as it is used only to initialize the land surface model.	Y
TAVGSFC	X	1.		K	This field is identical to SKINTEMP but is used to initialize surface temperatures for inland lakes if that option is used. The use of a monthly average temperature field (expressed 6-hourly) can keep the lake surface temperatures from undergoing large diurnal fluctuations.	Y
SST	X			K	Daily sea surface temperature expressed every 6 hours. (e.g., all time steps on Jan 1, 2002 are the same, then SST changes for Jan 2, 2002). The SST field is interpolated from the ocean model (POP) output from CESM.	Y
SEAICE	X			fraction	Daily sea ice fraction expressed every 6 hours. (e.g., all time steps on Jan 1, 2002 are the same, then SEAICE changes for Jan 2, 2002). The SEAICE field is interpolated from the ocean model (POP) output from CESM.	N

SNOW	X			kg m^{-2}	Monthly average snow water equivalent expressed every six hours (e.g., all time steps in Jan 2002 are the same, then SNOW changes for Feb 2002, etc). This variable is relatively unimportant as it is used only to initialize the land surface model.	N
ST000010	X			K	Monthly average soil temperature from 0->10 cm depth expressed every six hours (e.g., all time steps in Jan 2002 are the same, then ST000010 changes for Feb 2002, etc). This variable is relatively unimportant as it is used only to initialize the land surface model.	Y
ST010040	X			K	Same as above, but for 10->40 cm depth.	Y
ST040100	X			K	Same as above, but for 40->100 cm depth.	Y
ST100200	X			K	Same as above, but for 100->200 cm depth.	Y
SM000010	X			fraction	Monthly average soil water fraction from 0->10 cm depth expressed every six hours (e.g., all time steps in Jan 2002 are the same, then SM000010 changes for Feb 2002, etc). This variable is relatively unimportant as it is used only to initialize the land surface model.	Y
SM010040	X			fraction	Same as above, but for 10->40 cm depth.	Y
SM040100	X			fraction	Same as above, but for 40->100 cm depth.	Y
SM100200	X			fraction	Same as above, but for 100->200 cm depth.	Y
PMSL		X		Pa	Six-hourly mean sea level pressure from CESM. PMSL is diagnosed from surface geopotential and pressure and the lowest model level temperature and pressure from CESM using the ECMWF methodology *(see function 'pslec' in the NCL software).*	Y

PSFC	X			Pa	Six-hourly surface pressure from CESM.	Y
UU	X		X	m s^{-1}	Six-hourly zonal wind speed from CESM. Pressure level (PLEV) winds are directly from CESM output. The 10-m winds (level 200100) are diagnosed from the lowest CESM model level winds using a power law that assumes the winds diminish to zero at the surface.	Y
VV	X		X	m s^{-1}	Six-hourly meridional wind speed from CESM. Pressure level (PLEV) winds are directly from CESM output. The 10-m winds (level 200100) are diagnosed from the lowest CESM model level winds using a power law that assumes the winds diminish to zero at the surface.	Y
TT	X		X	K	Six-hourly temperature fields from CESM. Pressure level (PLEV) temperatures are directly from CESM output. The 2-m temperature (level 200100) is diagnosed from the lowest 2 CESM model levels by linear extrapolation.	Y
RH	X		X	%	Six-hourly relative humidity fields from CESM. The pressure level (PLEV) RH fields are calculated from CESM temperature/ pressure/ specific humidity via the 'relhum' function in NCL. The 2-m relative humidity (level 200100) is assumed to be the same as the RH on the lowest CESM model level.	Y
GHT			X	m	Six-hourly geopotential height from CESM, diagnosed from the 3-dimensional virtual temperature and pressure fields via the function 'cz2ccm' in NCL.	Y

*Pressure levels in hPa: 1000, 975, 950, 925, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10

7. Description of Software

This section describes in detail how the data were processed, the software, and how to use the software to create similar bias-corrected datasets. The software can be obtained from NCAR's CISL Research Data Archive: http://rda.ucar.edu/datasets/ds316.1/#!software. Unzip and untar the code (CESM_to_Intermediate.tar.gz) in your working directory.

The software package consists of three sections *(described in detail below):*

a) a section for obtaining the CESM data and converting it to the Intermediate File Format (contained in the directory **CESM_TO_INT)**;

b) a section for obtaining the ERAI data and converting it to the Intermediate File Format (directory **ERAI_TO_INT**); and

c) a section for bias correcting the CESM data using ERAI data (directory **BIAS_CORRECTION**).

The software was written in NCL and FORTRAN, so if users want to run this code on their own computers, they will need to have a FORTRAN compiler and NCL (http://www.ncl.ucar.edu - version 6.3.0 of higher) installed. The software also makes use of Climate Data Operators (CDO; https://code.zmaw.de/projects/cdo) for file manipulation, so users also need to ensure these operators are installed in their computers before running this software. NCL and CDO are freely available, and some FORTRAN compilers (e.g., gnu) can also be obtained freely.

a) Conversion of CESM to Intermediate Format

Required Data

The filenames for the different input datasets have been hardcoded in the NCL conversion script. If you would like to use this script you need to provide data in the same format and in files with the same names. Table 2 lists the data requirements and filename conventions used in the NCL conversion script. The script can process data at multiple times, but we recommend no more than 3 months at a time, else the scripts' memory footprint will be too large.

Table 2: Data requirements and filename conventions used in the NCL conversion script

Field	**Variable Name**	**Units**	**Frequency**	**Vertical coordinate**	**NCL File name**	**Projection**
Temperature	T	K	6-hourly	hybrid	atmos_ta.nc	Lat-Lon
Zonal Wind	U	m/s	6-hourly	hybrid	atmos_ua.nc	Lat-Lon
Meridional Wind	V	m/s	6-hourly	hybrid	atmos_va.nc	Lat-Lon
Specific Humidity	Q	kg/kg	6-hourly	hybrid	atmos_hus.nc	Lat-Lon
Surface Pressure	PS	Pa	6-hourly		atmos_ps.nc	Lat-Lon
Surface Geopotential	PHIS	m^2/s^2	Invariant		atmos_zsfc.nc*	Lat-Lon
Landmask	LANDMASK	0/1	Invariant		atmos_lmask.nc*	Lat-Lon
SWE	snw	kg/m^2	Monthly		atmos_snw_1.nc**	Lat-Lon
Soil Temperature	soil_temperature	K	Monthly		atmos_tsl_1.nc**	Lat-Lon
Soil Water Content	mrlsl	kg/m^2	Monthly		atmos_mrlsl_1.nc**	Lat-Lon
Surface Temperature	ts	K	Monthly		atmos_ts_1.nc**	Lat-Lon
SST	tos	K	Daily		atmos_tos_1.nc**	Gaussian
Ice Cover	aice_d	%	Daily		atmos_sic_1.nc***	Gaussian

**Provided with software*

***Currently only available from the NCAR supercomputer*

****Currently only available on NCAR tape storage*

Invariant data are supplied with the conversion scripts and are stored in the directory *Invariant_Data.*

On the NCAR supercomputer the 6-hourly, monthly and SST data are available on the glade file system. The 6-hourly data are in the directory */glade/p/rda/data/ds316.0/,* while the monthly and SST data are in the directory */glade/p/vetssg/data/CMIP5/output1/NCAR/CCSM4*. The daily sea ice data are currently only available on NCARs' tape storage under the directory */CCSM/csm.*

All the 6-hourly data are available from the NCAR RDA web site (http://rda.ucar.edu/datasets/ds316.0/). The monthly and daily data are only available from the NCAR supercomputing system. If you want to run this code on a different computer platform, please contact the NCAR RDA help desk for assistance in obtaining these data files.

How to Run the CESM to Intermediate File Software

1. Make sure to load all of the modules you may need:
On the NCAR supercomputer, type:

```
module load ncl
module load cdo
```

If you are running on your own computer, make sure the NCL and CDO libraries are installed and loaded.

The NCL scripts make use of the function ***wrf_wps_write_int*** to create the Intermediate files. This function is available in NCL since version 6.3.0. Please ensure that you have at least this version of the NCL libraries installed before running the code.

2. Obtain and process the data
All the software necessary to read the CESM data and convert it to the Intermediate file format is contained in the directory "**CESM_TO_INT**". Within this directory, the following NCL script reads in CESM data and writes out the required fields in Intermediate File Format:

convert_cesm_hybrid_nc_to_pressure_int.ncl

This NCL script has been specifically written to process one member (Member #6, or 6i1p1) of the CESM *(http://rda.ucar.edu/datasets/ds316.0)* dataset (*viz., Version 4 of the Community Climate System Model - CCSM4; Gent et al. 2011*). Note, this is the only member that you can process, as it is the only member that has the full three-dimensional fields required to force WRF or MPAS available at 6-hourly intervals. You can run this script on your computer if you have the CESM data available. The script expects specific data in predefined files *(which is described in the Data Requirements section above).* If you are running this code on the NCAR supercomputer, you can use the driver script (*process_cesm_data.csh*) to link in the correct data and run the NCL script. You can get file-naming conventions by looking at the driver script. You can also get an idea of where the exact directories of interest are by looking in the driver script.

The NCL script writes out all required fields at 6-hourly intervals to Intermediate files called *"IM_root_name"_"CASE"*:*YYYY-MM-DD_HH*, where; *"IM_root_name"* by default is **CESM_CMIP5_MOAR**, but can be set by the user; "*CASE*" is either 20THC, RCP45, RCP60 or RCP85; while *YYYY, MM, DD* and *HH* have their usual time conventions. Users also have the option to set an output directory. The Intermediate files will be written to "*outDIR*"/*YYYY*. If a user does not specify the output directory, the default (**./OUTPUT**) output directory will be used.

If you are running the code on the NCAR supercomputing platform, you may use the following scripts to obtain the correct data:

get_seaice.csh
process_cesm_data.csh

Since the high-frequency SEAICE is currently only available on NCAR tape storage, download the SEAICE first by using the script *get_seaice.csh* inside the SEAICE/ directory. You can use the *submit_job_to_queue.csh* script to submit the job to the NCAR queue (*recommended*). Unfortunately, this step is necessary, because only the monthly average sea ice fraction data are available on GLADE, and at least daily varying sea ice fields are needed in order to have consistent data for our lower boundaries. The only way to get the sea ice is to download it from the HPSS tape storage, which is what the *get_seaice.csh* script does. To run this script type:

./get_seaice.csh -c *CASE* –y *YYYY* –m *XX*

where, *CASE* is the RCP simulation you are running, i.e., 20THC, RCP85, RCP60, or RCP45; *YYYY* is the start year; and *XX* is the number of years of data you want to download. *CASE* and *YYYY* are required; if *XX* is not set, a single year (*YYYY*) will be downloaded.

Secondly, link in all the required data and run the NCL script to process the data.

./process_cesm_data.csh -c *CASE* –y *YYYY* –m *XX* –f *FILENAME* –o *DIRECTORY*

where, *CASE* (required) is the RCP simulation you are running, i.e., 20THC, RCP45, RCP60, or RCP85; *YYYY* (required) is the start year; *XX* (optional; default is 1 year) is the number of years of data you want to process; *FILENAME* (optional; CESM_CMIP5_MOAR is default) is the root name for the Intermediate files; and *DIRECTORY* (optional; default is ./OUTPUT) is the Intermediate file output directory.

This script links all the required data and runs the NCL script (*convert_cesm_hybrid_nc_to_pressure_int.ncl*) in three-month sections. This script takes approximately 30 min to process 1 year of data and writes out all Intermediate files. Due to the lengthy run time, it is recommended to submit the job to the queue. You can do so with the *submit_job_to_queue.csh* script. A year of data is roughly 40 Gb. Output files are 6-hourly and named per the convention described above.

Make sure the following file is in your *CESM_TO_INT* directory, which allows the SST and SEAICE fields to be interpolated from the POP grid to the CESM grid:
    map_gx1v6_to_fv0.9x1.25_aave_da_090309.nc.

b) Conversion of ERA-Interim to Intermediate Format

This module is called "**ERAI_TO_INT** ". The purpose of this software package is to convert the surface and pressure-level ERAI fields to 1) the same horizontal and vertical domain as the CESM data that was processed in the step above; and 2) to the Intermediate File Format. The purpose of this step is to facilitate the bias-correction step, which is described in the following section. This step can be done by either; first processing the 6-hourly ERA-Interim data and then calculating monthly means, or by directly processing the ERA-Interim monthly mean data. Here we use the available monthly mean ERA-Interim data.

The ERA-Interim monthly mean data used are stored as dataset ds627.1 on GLADE courtesy of NCAR's Research Data Archive:
    /glade/p/rda/data/ds627.1/ei.moda.an.pl, and
    /glade/p/rda/data/ds627.1/ei.moda.an.sfc

The data are also available from the NCAR RDA web site for users that would like to process the data on their own systems (*http://rda.ucar.edu/datasets/ds627.1*). The data files required are:
    ei.moda.an.pl.regn128sc.DATE.grb,
    ei.moda.an.pl.regn128uv.DATE.grb, and
    ei.moda.an.sfc.regn128sc.DATE.grb
where *DATE* is for a given year and month.

The NCL script that reformats ERA-Interim (*convert_era_grib_to_cesm_pressure_int.ncl*) uses the gaussian-to-fixed global grid functions that are available in NCL in order to do the horizontal grid transformation. No vertical interpolation is necessary as all 26 of the vertical pressure levels that are needed to match the CESM vertical levels are already available. *Note that since the ERA-Interim data are transformed to the CESM grid, the CESM grid specifications are required by the NCL script and have therefore been hardcoded into this script.*

Run the NCL script by typing:

ncl convert_era_grib_to_cesm_pressure_int.ncl <options>

The available command line options are:

1. YS=*YYYY*
   Start processing data from this year (*required*)
2. YE=*YYYY* or numYY=*NUM*
   Controls the number of years that will be used to create the monthly mean values.
   Either explicitly specify the range by setting YE, or specify the number of years to use for the average with the option numYY.
   This is an optional setting. If not set, a 20-year average will be used.
   It is recommended to always use at least 20 years or more to calculate the average.
3. MMs and MMe
   Control the months to generate. This is mainly used for workflow management, as all 12 monthly means are needed.
   This is an optional setting. If not set all 12 months will be processed.
4. 'IM_root_name="*New-Output-Root-Name*"'
   Root name to use for the Intermediate files.
   Optional. It not set ERAI_*YS-YE*_AVG will be used.
   Note the syntax required for this option.
5. ‘outDIR="*New-Output-Directory*"'
   Directory to write the Intermediate files to.
   Optional. If not set ./OUTPUT will be used.
   Note the syntax required for this option.

Note that since the output data are only month specific at this point, and not associated with any given year, the files will have a generic “yyyy” in the date information, rather than any actual year. As the code is relatively memory intensive, it is recommended to submit it to the queue if running on the NCAR supercomputer.

c) Bias-Correction

This module (contained in the **BIAS_CORRECTION** directory) is written primarily in FORTRAN 90 and performs the Bruyère et al. (2014) bias correction by reading in the CESM and ERA-Interim intermediate files that were created using the two software packages described above. The bias correction is done in three steps: 1) calculation of monthly means, 2) interpolation of the monthly means into 6-hourly means, and 3) bias correction. Correspondingly there are three FORTRAN 90 routines for these tasks:

monthly_means.f90
interp_6hr.f90
bias_correct.f90

In addition, there is a FORTRAN 90 module common across all the above main FORTRAN programs:

module_basic.f90

Input to all the programs is done through a common *namelist.input* file.

<u>How to Run the Bias-Correction Software</u>

1. Compile the three FORTRAN 90 routines

Compile scripts have been provided for PGI and Intel. The compile scripts are somewhat NCAR centric, but can be easily adapted to any other computing platform.

2. Create monthly means

First edit the **&mean** section of the namelist.input file:

i. change the paths to the input and output directories (*the output directory will be created if it does not exists*);
ii. specify the input and output Intermediate file root names;
iii. specify the years that should be used to generate the monthly mean averages (**note**: this should be the same years as you used for the ERA-Interim monthly means).

Second, generate the CESM monthly means by typing:

monthly_means.exe

This step is extremely time consuming, with one 20-year average monthly mean taking approximately 4 hours on the NCAR supercomputer, so be sure to submit this job to the queue (i.e,. do not run this job on the login node because it will fail due to consuming excess memory).

As we generated ERA-Interim monthly means directly in the previous step, there is no need to run the monthly mean program for the ERA-Interim data. If you have 6-hourly ERA-Interim data in Intermediate file format, then you need to repeat this step for these data as well. Since the output data at this point are only month specific, and not associated with a specific year, the files will have a generic “yyyy” in the date information, rather than any actual year.

3. Interpolate the monthly means to 6-hourly means

This step must be done for both the CESM and ERA-Interim datasets. The program requires that all 12 monthly mean files for each dataset are available. The executable is controlled via the **&interp** section of the namelist.input file. In the namelist, change the input and output directories (*the output directory will be created if it does not exist*), and specify the input and output Intermediate file root names.

Run the program (once for the ERA-Interim data and once for the CESM data) by typing:

interp_6hr.exe

This code runs very fast and can be run on the command line.

Since the data are interpolated from mean monthly to mean 6-hourly, the output files will, as before, use a generic "yyyy" specification, rather than any actual year.

4. Perform bias correction of the CESM 6-hourly files

Edit the **&bc** section of the namelist.input file. This section of the namelist requires 4 path/root name combinations; one set for the 6-hourly CESM input files; one set for the desired 6-hourly CESM *bias corrected output*; one set for the 6-hourly monthly mean CESM files; and one set for the 6-hourly monthly mean ERA-Interim files. The times you want to process are also controlled by the namelist.input file.

Run the bias correction step by typing:

bias_correct.exe

The bias correction program is not very memory intensive, but if a long time slice is processed, it is desirable to submit the job to a queue rather than running the job interactively.

This is the last step in the CESM-to-Intermediate File Format software. You are now ready to use the created Intermediate files directly as input to the WPS/metgrid.exe program.

Notes on the Bias-Correction Software

1. On the NCAR supercomputer, processing time can be reduced if multiple versions of each processing script are run simultaneously (through multiple job submissions), breaking up the jobs by month (for monthly_means) and by smaller ranges of years (for bias_correct).

2. Details of the bias correction software can be found in the *bias_correct.f90* program - specifying whether variables are bias corrected or just passed along, how to handle negative values, etc. A summary is also available in Table 1.

8. References

Bruyère C.L., J.M. Done, G.J. Holland, and S. Fredrick, 2014: Bias Corrections of Global Models for Regional Climate Simulations of High-Impact Weather. *Climate Dynamics,* **43***,* 1847-1856*, doi:10.1007/s00382-013-2011-6.*

Dee D.P., with 35 co-authors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. R. Meteorol. Soc.,* **137**, 553-597, doi:10.1002/qj.828.

Done J.M., G.J. Holland, C.L. Bruyère, L.R. Leung, and A. Suzuki-Parker, 2013: Modeling high-impact weather and climate: lessons from a tropical cyclone perspective. *Climatic Change*, doi:10.1007/s10584-013-0954-6.

ECMWF, 2009:. ERA-Interim Project. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. http://rda.ucar.edu/datasets/ds627.0/. Accessed 2014.

ECMWF, 2012: ERA-Interim Project, Monthly Means. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. http://rda.ucar.edu/datasets/ds627.1/. Accessed 2014.

Gent P.R., G. Danabasoglu, L.J. Donner, M.M. Holland, E.C. Hunke, S.R. Jayne, D.M. Lawrence, R.B. Neale, P.J. Rasch, M. Vertenstein, P.H. Worley, Z.-L. Yang, and M. Zhang: 2011. The Community Climate System Model version 4. *J. Climate,* **24**: 4973-4991, doi: 10.1175/2011JCLI4083.1.

Hurrell J.W., and co-authors, 2013: The Community Earth System Model: A Framework for Collaborative Research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360.

IPCC (Intergovernmental Panel on Climate Change), 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschufng, A. Nauels, Y. Xia, V. Bex, and P.M Midgley, eds). Cambridge, UK and New York, NY:Cambridge University Press.

Knutti R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.*, **40**, 1194–1199, doi:10.1002/grl.50256.

Lorenz C., and H. Kunstmann, 2012: The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis. *J. Hydrometeor*, **13**, 1397–1420.

Masui T., K. Matsumoto, Y. Hijioka, T. Kinoshita, T. Nozawa, S. Ishiwatari S, and co-authors, 2011: An emission pathway for stabilization at 6 Wm^{-2} radiative forcing. *Clim Change,* **109***, 59-76.*

Monaghan, A.J., D.F. Steinhoff, and C.L. Bruyère, 2014: *NCAR CESM Global Bias-corrected CMIP5 Output to Support WRF/MPAS Research. Research Data Archive at the National Center for Atmospheric Research*, Computational and Information Systems Laboratory. http://dx.doi.org/10.5065/D6DJ5CN4.

Moss, R.H., and co-authors, 2010: The next generation of scenarios for climate change research and assessment. *Nature,* **463,** 747-756, doi:10.1038/nature08823.

Rasmussen R., and co-authors, 2011: High-Resolution Coupled Climate Runoff Simulations of Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate. *J. Climate*, **24**, 3015–3048.

Riahi, K., S. Rao, V. Krey, C. Cho, V. Chirkov, G. Fischer, G. Kindermann, and N. Nakicenovic, 2011: RCP8.5 – A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, **109**, 33–57.

Ringler, T., L. Ju and M. Gunzburger, 2008: A multiresolution method for climate system modeling: application of spherical centroidal Voronoi tessellations, *Ocean Dynamics*, **58** (5-6), 475-498, doi:10.1007/s10236-008-0157-2.

Simmons, A., S. Uppala, D. Dee, and S. Kobayashi, 2006: ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter*, **110**, ECMWF, Reading, United Kingdom, 26–35.

Skamarock, W., J.B. Klemp, J. Dudhia, D.O. Gill, D. Barker, M.G. Duda, X.-Y. Huang, and W. Wang, 2008: A Description of the Advanced Research WRF Version 3. NCAR Technical Note NCAR/TN-475+STR, doi:10.5065/D68S4MVH.

Skamarock, W.C., J.B. Klemp, M.G. Duda, L. Fowler, S-H. Park, and T.D. Ringler, 2012: A Multi-scale Nonhydrostatic Atmospheric Model Using Centroidal Voronoi Tesselations and C-Grid Staggering. *Monthly Weather Review*, **240**, 3090-3105, doi:10.1175/MWR-D-11-00215.1.

Taylor K.E., R.J. Stouffer, and G.A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteorol. Soc.,* **93**: 485–498 doi:10.1175/BAMS-D-11-00094.1.

Thomson, A.M., K.V. Calvin, S.J. Smith, G.P. Kyle, A. Volke, P. Patel, S. Delgado-Arias, B. Bond-Lamberty, M.A. Wise, L.E. Clarke, and J.A. Edmonds, 2011: RCP4.5: a pathway for stabilization of radiative forcing by 2100. *Climatic Change*, **109**, 77-94.

Thuburn, J., T. Ringler, J. Klemp and W. Skamarock, 2009: Numerical representation of geostrophic modes on arbitrarily structured C-grids, *Journal of Computational Physics*, 2009: **228** (22), 8321-8335, doi:10.1016/j.jcp.2009.08.006.

Xu Z., Z.-L. Yang, 2012: An improved dynamical downscaling method with GCM bias corrections and its validation with 30 years of climate simulations. *J. Climate* **25**: 6271-6286, doi:10.1175/JCLI-D-12-00005.1.