V4.0.0 1/27/16
CVDP created by Adam Phillips (NCAR/CGD/CAS)

Before attempting to run the Climate Variability Diagnostics Package (CVDP) please read the following document. For more information about the CVDP project please see the CVDP website: http://www2.cesm.ucar.edu/working-groups/cvcwg/cvdp

Required software

- NCL v6.1.2 or newer (v6.2.0 or newer if operating on netCDF4 files.)
- Image Magick
- Python 2.7.X or newer (if running in parallel; see the run_parallel option in Step 3 below.)

General notes

The CVDP is completely written in NCL, but knowledge of NCL is not required. The input files must be on local disk or on OPeNDAP servers. The CVDP creates plots that are based on the following variables: TREFHT (tas), TS (ts), PRECT (pr), PSL (psl), SNOWDP (snd), MOC (msftmyz/stfmmc) and aice (sic). Not all variables need to be present for a model simulation to be analyzed.

The CVDP operates on monthly time series files (observational or model-based), such as those found in the CMIP5 archive or those distributed by NCAR that contain one variable and a number of timesteps. *The CVDP does not work with CESM history files.* The CVDP expects the file names of input model and observational data to end in a specific format of "YYYYMM-YYYYMM.nc" denoting the start year and month and end year and month of the dataset. "YYYY" MUST have 4 digits. This naming format is used for files in the CMIP5 archive along with NCAR's CESM post-processed files. If your model data file names do not end in this manner it is suggested that you either rename the files or use soft links. The CVDP does not read the time variable out of input netCDF files; it relies on the times specified in the file names and checks that the number of timesteps indicated in the file name matches the number of timesteps in the file.

Multiple files containing the period of study are completely acceptable. The CVDP does not expect overlapping time slices of data to be present in an identified data directory. For instance a directory that contained these TS files would be acceptable:
modelA.TS.190001-191912.nc
modelA.TS.192001-193912.nc
A directory that contained these TS files would not be acceptable:
modelA.TS.190001-191912.nc
modelA.TS.190001-193912.nc
modelA.TS.192001-193912.nc

Non-monthly (ex. daily, 6-hourly) time series files should not be kept in the same directory as the monthly data. If they are, make sure the syntax used in namelist explicitly excludes non-monthly files. It is generally preferable for each model run to have its own directory, but it is not a necessity. (See the examples under *Step 1: set the namelist* below.)

Observational or model file names cannot start with the variable name followed by a period. (ex. tas.m1.198001-201212.nc) Use soft links or rename the files for use in the CVDP.

Instructions for running the Climate Variability Diagnostics Package (CVDP)

There are 2 files that must be set up to run the package (namelist, driver.ncl). A 3rd file can be set up if one wishes to include observations in the analysis (namelist_obs). These three files must be in the same directory. The CVDP codebase (ncl_scripts/*) can be located anywhere and can be pointed to using the driver.ncl option zp.

Step 1: set the namelist

The namelist file contains information about which set of model data you would like to pass in to the CVDP. You may enter as many simulations as you would like, but only specify one simulation per row.

Within the namelist file each row should follow the following format:
model name (arbitrary) | generic path to data files | analysis start year | analysis end year

" | " is used as a delimiter. The path should be as generic and simple as possible to capture the necessary files. The use of "*" or "{}" syntax is recommended. One can test that the specified path works to (only) identify the requested files by doing a "ls $path" via the command prompt.

namelist Example 1:
CCSM4 Control | /project/mdata/b40.1850.track1.1deg.006/ | 800 | 1199
CCSM4 Hist #1 | /project/cmip5/historical/*/*/CCSM4/r1i1p1/ | 1970 | 2005

Two simulations are specified in Example 1. The first is a CCSM4 control simulation where all the necessary files are in the stated directory. The second is a CMIP4 CCSM4 historical run. Note that /*/ syntax is used to span multiple directories.

namelist Example 2:
CESM1 LE #21 | /project/yampa03/dlocal/CESM1-LENS/b.e11.B*.f09_g16.021.* | 1979 | 2013
GFDL-ESM2M #1 | /project/cmip5/historical/Amon/r1i1p1/ | 1985 | 2005

Two simulations are specified in Example 2. The first is CESM1 Large Ensemble member #21. Note that a partial file name is provided to distinguish the 21st member data from other data in the specified directory. The second simulation specified is a GFDL CMIP5 historical simulation.

Numerous other example namelists are provided in the example_namelists directory.

General namelist notes:
- The paths specified in namelist should not end in with the syntax "YYYYMM-YYYYMM.nc".
- If a directory path is specified it should end with a "/"
- The CVDP can only analyze complete years. You can read in a simulation that starts or ends in an incomplete year, but those years cannot be set as being analyzed.
- Model data on curvilinear grids (such as the CESM spectral element grid) cannot be read into the CVDP. Curvilinear data should be regridded before CVDP input. One can use NCL's ESMF regridding tools to accomplish this. (See https://www.ncl.ucar.edu/Applications/ESMF.shtml)

Step 2: set the namelist_obs (optional)

The namelist_obs file contains information about which observational datasets (if any) are to be used. This file is only used if the driver.ncl option obs is set to "True".

The namelist_obs file is formatted as follows:
variable | observation name (arbitrary) | path to file(s) | analysis start year | analysis end year

Note that " | " is used as a delimiter. The paths specified in namelist_obs (contrary to those in namelist) should be as specific as possible. One dataset should be specified per row, but multiple datasets can be specified for each variable. If an observational dataset is not specified for a particular variable that variable should be left off of the namelist_obs file.

namelist_obs Example 1:
TS | HadISST | /project/cas/DATA/hadisst.187001-201312.nc | 1920 | 2011
PSL | 20thC_ReanV2 | /project/cas/DATA/prmsl.mon.mean.187101-201112.nc | 1920 | 2011
TREFHT | MLOST | /project/cas/DATA/mlost.v3.5.2.188001-201212.nc | 1920 | 2011
PRECT | GPCC | /project/cas/DATA/full_data_v6_precip_10.190101-201010.nc | 1920 | 2009
SNOWDP | UDel | ~/Data/snow_depth.195001-201012.nc | 1960 | 2005
MOC | Obs_MOC1 | ~/Data/moc.observed.198101-201212.nc | 1981 | 2012
aice_nh | NASA B v2 | /project/cas/DATA/seaice.nsidc.nasa.bv2.197011-201412.nc | 1971 | 2014
aice_sh | NASA Team | /project/cas/DATA/seaice.nsidc.nasa.team.197011-201412.nc | 1979 | 2012

Example #1 shows how to set an observational dataset for each of the 8 CVDP variables. Note that the analysis period can be different for each observed dataset.

namelist_obs Example 2:
PSL | 20thC_ReanV2 | /project/cas/DATA/prmsl.mon.mean.187101-201112.nc | 1920 | 2011
TREFHT | MLOST | /project/cas/DATA/mlost.v3.5.2.188001-201212.nc | 1920 | 2011
TS | HadISST | /project/cas/DATA/hadisst.187001-201312.nc | 1920 | 2011
TS | ERSSTv3b | /project/cas/DATA/ersstv3b.185401-201312.nc | 1920 | 2012
TREFHT | HadCRUT3v | /project/cas/DATA/mlost.v3.5.2.188001-201212.nc | 1920 | 2011
TREFHT | MLOST | /project/cas/DATA/mlost.v3.5.2.188001-201212.nc | 1950 | 2012

Example #2 shows how to specify multiple datasets per variable, while not setting observational datasets for PRECT, SNOWDP, MOC, aice_nh or aice_sh. The ordering of the rows is completely irrelevant, and one can specify a dataset twice but specify different periods of analysis (as was done for rows 2 and 6).

General namelist_obs notes:
- The CVDP can only analyze complete years. You can read in an observational dataset that starts or ends in an incomplete year, but those years cannot be set as being analyzed. (See the PRECT row above for an example.)
- Each dataset can start and end at different years, and the period of analysis can be different for each dataset.
- In order to get ENSO SST/TS/PSL composites (and the metrics table) the start and end years must match between the specified SST, TS and PSL datasets. In order to get ENSO PR composites the start and end years must match between the specified SST and PRECT datasets.
(continued on next page)

- TS refers to a specified SST dataset.
- aice_nh / aice_sh refers to sea ice concentration in the northern / southern hemispheres. If one has an observational file that spans both hemispheres that file can be specified twice in both aice_nh and aice_sh rows.
- Due to compositing requirements the CVDP will set each variable namelist to be the same length. The length will be equal to the number of model simulations plus the maximum number of observational datasets per variable. Thus, if three PSL observational datasets are set and one observational dataset is set for every other variable, then each variable namelist will have three observational rows in addition to the specified model simulation. In this scenario the CVDP will attempt to fill in the (non-PSL) namelists with the first specified observational dataset. If an observational dataset is not provided for a variable, the CVDP would set each observational row to missing.

Step 3: modify and run driver.ncl

---

driver.ncl is the driving script of the CVDP. There are user-adjustable options located at the top of driver.ncl. Each option has comments on the right explaining the various settings. Once driver.ncl is set, one can start the CVDP by entering "ncl driver.ncl" in the terminal window. The command can also be put into background mode and the output sent to a file:
"ncl driver.ncl >&! a.out &"

Notable driver.ncl options:
namelists_only – Set to “True” when running with a new model or observational dataset in namelist or namelist_obs. This will allow you to examine the variable namelists that the CVDP set up (based on your namelist and optional namelist_obs files). Within each file in namelist_byvar/ you will find a path for each dataset. You can execute a "ls $path" to see if the set path(s) are correct. If the path is listed as "missing" the CVDP is not finding the file. Check your namelist/namelist_obs settings and verify that the specified path syntax is correct.

output_data – Set to “True” to output all calculations to netCDF files. If “True” all atmospheric variables for a particular model run must be on the same grid. Land and ocean variables can be on different grids than atmospheric variables. All output netCDF files are CF-conforming.

run_parallel – When set to “True”, the CVDP will run in parallel mode and submit multiple CVDP calculation scripts at once. This option can significantly reduce the CVDP run time. To use this option python needs to be installed on your local machine along with the subprocess, sys, time, and os modules (all common). The number of scripts that can run concurrently is set via the *max_num_tasks* option. Note that terminal output may get intermixed when running in parallel mode. When set to “False” the CVDP will submit calculation scripts serially.

machine_casesen – Set this option = “True” if your filesystem is case sensitive. If your filesystem is case insensitive (as most Macs are) you can set this option to “False”. In developing the namelists the CVDP searches for files by using specific syntax that may cause case insensitive systems to see a file twice, thus causing the CVDP to error out.

Step 4: Examining CVDP output

Final CVDP output is written to the directory specified via the outdir directory in driver.ncl. The output is displayed via HTML files. Open a browser and point to $outdir/index.hmtl to see CVDP output. If you set the driver.ncl option tar_output = “True” you will have to untar the file prior to viewing the contents.

Known limitations of the CVDP

- Does not run on the spectral element or any other curvilinear grid. (=contains 2D lats/lons)
- Only certain variable names (within the .nc files) are accepted:

```
 TS     = (/"TS","ts","sst",”t_surf”/)
 PSL    = (/"PSL","psl","slp","SLP","prmsl",”msl”,”slp_dyn”/)
 TREFHT = (/"TREFHT","tas","temp","air","temperature_anomaly",\
                  ”temperature”,”t2m”,”t_ref”,”T2”,”tempanomaly”/)
 PRECT  = (/"PRECC","PRECL","PRECT","pr","PPT","ppt","p","P","precip",\
                "PRECIP",”tp”,”prcp”,”prate”/)
 SNOWDP = (/"SNOWDP","snd"/)
 MOC = (/”MOC”,”msftmyz”,”stfmmc”/)
 aice_nh = (/”aice_nh”,”aice”,”sic”,”SIC”,”CN”,”ice”,”icec”/)
 aice_sh = (/”aice_sh”,”aice”,”sic”,”SIC”,”CN”,”ice”,”icec”/)
```

  If you wish to read in a different variable name you can alter lines 57-71, 372-374 and 613-618 of ncl_scripts/functions.ncl as necessary.

- Minimum length of simulation/observational data required: 6 years