

v3.3.1 12/1/14

CVDP created by Adam Phillips (NCAR/CGD/CAS)

Note: Before attempting to run the Climate Variability Diagnostics Package (CVDP) please read the following document. If you need assistance please visit the CVDP support page:
<http://www2.cesm.ucar.edu/working-groups/cvcwg/cvdp/support>

General notes

The CVDP is completely written in NCL, but knowledge of NCL is not required. NCL v6.1.2 or newer should be used, while v6.2.0 or newer should be used if operating on netCDF4 files. The files must be on local disk or on OPeNDAP servers. (Note: If accessing OPeNDAP data one must use an OPeNDAP enabled version of NCL.)

The CVDP creates plots that are based on the following variables:
TREFHT (tas), TS (ts), PRECT (pr), PSL (psl), and SNOWDP (snd).
Not all variables need to be present for a model simulation to be analyzed.

The CVDP operates on 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". Note: "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 softlinks. The CVDP does not read the time variable out of input netCDF files; it relies on the times specified in the file names. The number of timesteps specified in the file name must match 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.190001-190912.nc

modelA.TS.191001-191912.nc

modelA.TS.192001-192912.nc

but a directory that contained these TS files would not be acceptable:

modelA.TS.190001-190912.nc

modelA.TS.190001-192912.nc

modelA.TS.191001-191912.nc

modelA.TS.192001-192912.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.)

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 | generic path to data files | analysis start year | analysis end year

Note that " | " is used as a delimiter. Important: 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.

Note #1: The paths specified in the namelist should not end in with the syntax "YYYYMM-YYYYMM.nc".

Note #2: If a directory path is specified it should end with a "/"

Note #3: The CVDP can only analyze complete years. You can read in a dataset that starts or ends in an incomplete year, but those years cannot be set as being analyzed.

Individual namelist row examples:

- An example of reading in the CCSM4 control run where all files are in the stated directory:

CCSM4 Control | /project/mdata/b40.1850.track1.1deg.006/ | 800 | 1199

- An example of reading in a CCSM4 simulation from NCAR's CMIP5 archive disk:

CCSM4 | /project/cmip5/ETH/cmip5/historical/Amon/*/CCSM4/r1i1p1/ | 1970 | 2005

(Note the "*/" syntax where the variable names are usually specified.)

- An example of reading in the 3rd run from the CCSM4 Large Ensemble when all the data is in one directory:

CCSM4 LE #1 | /project/yampa03/dlocal/b40.20th.005_ens/b40.20th.005_ens03.* | 1979 | 2005

(Note that a partial file name is given to distinguish the 3rd simulation from other simulations in the directory.)

Numerous example namelists are provided in the example_namelists directory.

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 | 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.

An example namelist_obs file is shown here:

```
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_10.190101-201012.nc | 1920 | 2011
SNOWDP
```

The order of the rows should not change. (1st row = TS, 2nd row = PSL, etc) If you do not have an observational dataset for a particular variable, simply leave the row blank after the variable name (shown in example above with SNOWDP).. Note that TS refers to a SST dataset.

The CVDP can only analyze complete years. You can read in a dataset that starts or ends in an incomplete year, but those years cannot be set as being analyzed. The following syntax (for example) is allowed as 1870 and 2013 are not set to be analyzed:

```
TS | HadISST | /project/cas/DATA/hadisst.187002-201307.nc | 1871 | 2012
```

Metric Table note: In order for the CVDP to produce a metric table containing pattern correlations and rms differences, namelist_obs must be set up and used. In order to get the complete metric table with overall model scores, the specified analysis start and end years for the TS, SST, and PSL observational datasets must match.

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 ready to go, to start the CVDP enter "ncl driver.ncl" in the terminal window. The command can be put into background mode and the output sent to a file:

```
"ncl driver.ncl >&! a.out &"
```

When you run the package with a new namelist for the first time it is highly recommended that you set namelists_only = "True" in driver.ncl. 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 namelist_byvar/namelist_\$variable file 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 in namelist_\$variable is listed

as "missing" the CVDP is not finding the file. Check your namelist settings and verify if the specified path syntax is correct.

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.html to see CVDP output.

Analyzing multiple observational datasets at once

While it is not possible to specify multiple observational datasets for each variable within namelist_obs, it is possible to trick the CVDP into using more than 1 set of observational datasets.

An example of this is stored in the example_namelist/multiple_obs directory. A 2nd set of observational data are placed in a separate directory, and soft links are used (that include the variable name within the file name). This tricks the CVDP into treating the 2nd set of observations as a model run. One limitation of this method is that you cannot write out the CVDP data to netCDF files.

Writing out CVDP data to netCDF files

Within driver.ncl there is an option, output_data, that allows the user to output various CVDP calculations to netCDF files. If output_data is set to "True", all atmospheric variables for a particular simulation must be on the same grid. This statement applies to all simulations listed in namelist. The land variable snd can be on a different grid than the atmospheric variables. All output netCDF files are CF-conforming.

Known limitations of the CVDP

- Does not work with data that is on a spectral element or any other curvilinear grid. (=contains 2D lats/lons)

- File names cannot start with the variable name followed by a period.

(ex. tas.mymodel.198001-201212.nc) Use soft links or rename the files for use in the CVDP.

- Only certain variable names (within the .nc files) are accepted:

TS = (/ "TS", "ts", "sst" /)

PSL = (/ "PSL", "psl", "slp", "SLP", "prmsl" /)

TREFHT = (/ "TREFHT", "tas", "temp", "air", "temperature_anomaly" /)

PRECT = (/ "PRECC", "PRECL", "PRECT", "pr", "PPT", "ppt", "p", "P", "precip", "PRECIP" /)

SNOWDP = (/ "SNOWDP", "snd" /)

If you wish to read in a different variable name you will have to alter lines 42-56 of ncl_scripts/functions.ncl.

Converting CESM history files to the timeseries format required by the CVDP

The CVDP expects netCDF timeseries files which contain one variable per file with all the timesteps included. The netCDF Operators can be used to accomplish this.

The following csh script was contributed by Jerry Potter (NASA Goddard).

```
#####  
An example script for PSL assuming you have a set of h0 files in once directory:  
#####  
#!/bin/csh  
#  
set START=`date`  
foreach am (*.nc)  
#  
  ncks -3 -F -v PSL $am `basename $am nc`PSL.nc  
#  echo $am  
#  echo `basename $am nc4`subset.nc  
#  echo $am  
#  echo cloud.$am  
end  
  
#rm -rf *cloud*  
set FINISH=`date`  
echo $START[4] to $FINISH[4] > timing.txt  
echo "all done"  
#####  
then concatenate these files something like:  
ncrcat -4 -h -L 2 *PSL.nc PSL_197901_201312.nc  
#####
```