**Scripts for producing and analysing seasonal hydrological hindcasts**

While using these script should save a lot of time (compared to developing them from scratch), effort will be needed to adapt them to your own specific case. Especially it will be necessary to adapt the names of directories and set up the corresponding directory structure.

**There are several directories with relevant scripts:**

1. /home/WUR/greue002/CLIMAX/scripts; this directories mainly contains scripts, etc., for the *production of the hindcasts and the preparation of files for the analysis*
2. /home/WUR/greue002//R/myscripts/analysisCLIMAX/; this directories mainly contains scripts for the *analysis of the hindcasts*
3. Windows: greue002\M\Documents\R\R\_scripts\_Windows\CLIMAX; scripts for the *evaluation of hindcasts at locations with discharge observations*; also *post-processing* (to reduce errors in bias and spread in the hindcasts of discharge).
4. Other sub-directories of /home/WUR/greue002/R/myscripts contain the following possibly useful scripts:

* analysisEUPORIAS: documentation in greue002\M\Documents\overdracht\2016\ Programs R , cdo, grads, idl and bash.docx
* biascorrWouter: the most useful file is biasCorr.R, which performs a bias correction following Themeßl, M. J., Gobiet, A., & Leuprecht, A. (2011), reference is in script.
* disObs: see greue002\M\Documents\overdracht\2022\Scripts discharge observation.docx
* Functions: ccAtPValue.R computes the correlation coefficient as a function of the number of samples and the p-value. Can be used to determine threshold values of significance / non-significance of the correlation coefficient.

**1) /home/WUR/greue002/CLIMAX/scripts (mainly production of hindcasts)**

The central script: copyAndRun.sh

Several of the other scripts can be called from this central script. Those other scripts are automatically run by first copying the so-called empty version (“...empty.sh”) and then replacing so-called placeholder strings (e.g. finiyear\_ph, representing the first year) in the new version by specific values for the actual procedure to be carried out by the new version (e.g. 1981 for the first year).

acq\_singlevar\_CDS.sh download data for a single variable from the CDS

calc\_monthmean\_....... 4 scripts for calculating monthly means of the meteorological forcing / hydrological output of the reference simulation / hindcasts. This script is called from copyAndRun.sh.

comp\_mean\_initstates.sh compute the mean of initial states. This script is called from copyAndRun.sh. Mean initial states are needed for running hindcasts that isolate the effect of the meteorological forcing (by eliminating the effect of yearly variations in the initial conditions).

compute\_climate.sh compute climate data (average over many years) of the forcing or the hydrological output. This script is not connected to copyAndRun.sh and therefore not as generally useable as some other scripts.

conv\_to\_VIC\_forc\_empty.sh converts the hindcasts data downloaded from the CDS to meteorological forcing for VIC. This script is called from copyAndRun.sh. Use make\_WFDEIforc\_refrun\_empty.sh for the forcing of the reference simulation.

copyAndRun.sh the central script that can be used to call other scripts to perform tasks needed to produce and analyse seasonal hydrological hindcasts.

forms directory containing the forms for downloading data from the CDS

from24to3h\_empty.R script for resampling in the temporal domain the meteorological forcing of VIC. I used it to convert daily means to 3-hourly values but the script is probably able to also e.g. convert daily means to 6-hourly values and vice versa. Only temperature and short-wave incoming radiation are affected.

globparfile\_CLIMAX\_template.txt VIC global parameter file with placeholders

grids directory with grids. Perhaps not used anywhere.

make\_WFDEIforc\_refrun\_empty.sh converts the WFDEI forcing to VIC forcing for the reference simulation. This script is called from copyAndRun.sh. Use conv\_to\_VIC\_forc\_empty.sh for the hindcasts.

oldscripts this directory contains old version of scripts and obsolete scripts

rearr\_pred\_leadtime\_empty.sh rearrange the monthly output so each resulting file contains the data for a single lead month. This script is called from copyAndRun.sh.

run\_Rscript.sh script needed for running R scripts in batch mode. Do not remove it. This script is called from copyAndRun.sh.

runVIC\_empty.sh for running VIC. This script is called from copyAndRun.sh.

select\_data\_station\_empty\_v7.R this script produces two types of files needed for the analysis of the model output at the locations of the discharge observation stations: 1) for each observation station: the hindcasts, the output of the reference simulation and the observations; for three variables, namely precipitation, evapotranspiration and discharge, where precipitation and evapotranspiration are averaged over all upstream cells; discharge is provided in m3/s and as specific discharge (mm/day). The output file also contains metadata. 2) metadata for all observation stations.

This script can only be run if at least the following files are available: GLEAM dataset (“observations” of evapotranspiration), a file with the number of upstream cells for each model cell, a file with the indices of all upstream cells for each model cell, a netcdf file with gridded (same grid as VIC output) discharge observations.

**The directory oldscripts** is probably irrelevant

acq\_forcing\_CDS\_empty\_v3.sh This script is obsolete. CDS data can be downloaded with acq\_singlevar\_CDS.sh and then be converted to the correct input files for VIC with conv\_to\_VIC\_forc\_empty.sh.

biascorr this directory is related to bias correction. I did not make these scripts, etc., nor did I use this material. I also forgot who gave me this material; perhaps Wietse Franssen. My own bias correction script (biasCorr.R) is located in /home/WUR/greue002//R/myscripts/biascorrWouter/

rearr\_modout\_empty.sh script for rearranging the output of the hydrological model. This script is probably obsolete.

**2) /home/WUR/greue002//R/myscripts/analysisCLIMAX/ (mainly analysis of the hindcasts)**

The most important script is make\_verif\_plots.R. Many of the other scripts can only be run after their input data were generated with make\_verif\_plots.R. Note that running make\_verif\_plots.R is quite time-consuming, e.g. it takes some 20-30 minutes to produce a plot with 5 (metrics) x 7 (lead times) = 35 maps, mainly because the metrics have to be computed for each grid cell.

ann\_cycle\_regio.R make plot of the annual cycle for a region. Script is probably obsolete. Rather use plot\_annual\_cycle\_v2.R.

compareRestrHCLead.R This script generates a plot that compares the skill of different types of hydrological hindcasts (e.g. FullSH, InitSH and MeteoSH) and of precipitation hindcasts, as in Fig. 3 of Greuell and Hutjes (2022). First make\_verif\_plots.R must be run to generate the input data for this script.

compare\_twosimuls.R This script compares the skill of two or more variables, of hindcasts with and without bias correction in different terciles (ROC), for different types of hydrological hindcasts and precipitation, of detrended versus undetrended data, for different hydrological models and for pseudo- and real observations. Examples are figure 6c, 8 and 10 in Greuell et al. (2018), Figures 2c, 3, 4, 5, 9b and c in Greuell et al. (2019). First make\_verif\_plots.R must be runned to generate the input data for this script.

makefilesregbound.R script for creating a .rds and a .txt file with boundaries of regions within a domain. This script must be adapted for each domain. These .rds and .txt file then form the input for several of the other scripts.

make\_verif\_plots.R This script produces maps of the skill of the hindcasts, either for meteorological or for hydrological hindcasts, e.g. Fig. 3 in Greuell et al. (2018). The following metrics can be plotted: correlation coefficient, bias, CRPS, reliability component of the CRPSReliab, skill score of the CRPS, ROC area (Below Normal and Above Normal and the difference between these two) and RPSS. Other metrics could relatively easily be added.

The script also generates and saves arrays with data for further analysis.

map\_nmonth\_sign.R generate a map with the fraction of cases (e.g. all combinations of target and lead months) with significant skill. An example is Figure 4 in Greuell and Hutjes (2022).

maps\_climate.R Generates a map of the climate, e.g. Fig. 5 in Greuell and Hutjes (2022).

map\_which\_type\_SH\_v2.R generates a map in which different colours indicate where the different types of hindcasts should or could be used and where skill is low for all types of hindcasts, see Figure 10 in Greuell and Hutjes (2022).

plot\_annual\_cycle.R Generates plot of the annual cycles of precipitation, evapotranspiration, net precipitation and runoff for the entire domain or sub-regions, see e.g. Fig. 6a of Greuell and Hutjes (2022).

summary\_skill.R This script produces a graph that summarizes the skill of a hindcasts (one point for each combination of a target and a lead month). An example is figure 5 in Greuell et al. (2018). First make\_verif\_plots.R must be run to generate the input data for summary\_skill.R.

**3) Windows: greue002\M\Documents\R\R\_scripts\_Windows\CLIMAX: evaluation at observation stations.**

In this directory call\_plots\_verif.R is the central script from which most of the other scripts are called. The most time-consuming script is postProcessing.R

alphaRandom.R compute alpha, a metric that quantifies reliability, see Greuell and Hutjes (2022), for random hindcasts.

call\_plots\_verif.R central script for calling other scripts

ccAtPvalue.R correlation coefficient as a function of the number of samples and the p-value

crps\_normal\_mean.R obsolete

crps\_normal\_samples.R obsolete

crpsDisFuncMean.R compute the mean CRPS for three different distribution functions (normal, log-normal and gamma)

effectPPOnStd.R generate a histogram of the ratio of the standard deviation of the ensembles of discharge before and after post-processing

fitDistributions.R fit distribution function to the ensembles and determine the p-values of the observations with respect to the distribution functions, i.e. the location of the observation with respect to the cumulative distribution function of the hindcasts

MakeArraysCatchNrInd.R generate three lat-lon arrays to be derived from the routing network: 1) upstream of the cells. 2) number of upstream cells. 3) indices of all upstream cells.

mapKlingGupta.R make maps for the whole domain of the Kling Gupta Efficiency of the reference simulation

MapMetrics.R generates maps of metrics (averages over all combinations of target and lead months) for the stations in the catchments of the Parana and the Sao Francisco catchments

postProcessing.R post-process the hindcasts (method of Gneiting at al., 2015)

relationsMetrics.R generates scatter plots of different metrics versus each other. An example is Fig A3 in Greuell and Hutjes (2022).

selectGridCells.R select grid cells with discharge observations within the catchment of a river

stationDischarge.R plot of the performance of the reference simulation for a single station

stationSkill.R generates graph showing different metrics for the evaluation of hindcasts before and after post-processing at the location of a discharge observation station. An example is Fig A2 in Greuell and Hutjes (2022).

TalagrandDiagram.R generate a Talagrand diagram, see Mason and Stephenson, 2008.