Documentation for the HZG Evaluation Suite Version 0.92.2

Ronny Petrik, Christoph Braun and Beate Geyer

March 7, 2025

Contents

1 General introduction to HZG Evaluation Suite **3**

2 Quick start of EvaSuite on Levante **4**

3 Which datasets can be analyzed? **5**
3.1 Requirements on the datasets . . . 5
3.2 The regular rotated grid . . . 5
3.3 The regular lon-lat grid . . . 5
3.4 The Lambert Conic conformal grid . . . 6
3.5 Single locations - regular rotated or lon-lat grid . . . 6
3.6 Time dimension . . . 6
3.7 NetCDF data variables . . . 7

4 How does it work? - Internal workflow of EvaSuite **8**

5 How to obtain and install EvaSuite **9**
5.1 Requirements on the software . . . 9
5.2 Installing the stand-alone version of the HZG Evaluation Suite Version 0.92.2 . . . 9
5.2.1 Obtain program code . . . 9
5.2.2 Installation on Levante . . . 9
5.2.3 Update evasuite . . . 11
5.2.4 Uninstall evasuite . . . 11

6 How to configure EvaSuite **12**
6.1 Controlling the evaluation - main config file . . . 12
6.1.1 General configuration . . . 12
6.1.2 Data mining . . . 16
6.1.3 Professional settings for the Suite and netcdf files . . . 19
6.2 Settings for the datasets - model and proof data files . . . 20
6.3 Controlling the visualisation . . . 21

7 How to run EvaSuite **24**
7.1 Running the stand-alone version of EvaSuite . . . 24
7.1.1 Configure and run EvaSuite via config-files and sbatch . . . 24
7.1.2 Automisation: Run evasuite via a wrapper script . . . 24
7.1.3 Examples . . . 25

8 Supplementary tools **29**
8.1 Adding pre-defined and custom subregion shapes . . . 29
8.1.1 Command-line based geopackage exploration and modification . . . 29
8.1.2 Jupyter notebook-based geopackage exploration and modification . . . 30

A Code examples for the general configuration file **31**
A.1 General configuration, dictionary Visual_program . . . 31
A.2 General configuration, dictionary Masking_program . . . 31
A.3 General configuration, dictionary DataDumpProcess_programs . . . 31

A.4 General configuration, dictionary datamining, subdictionary target_EvalFrequency . . 31
A.5 General configuration, dictionary datamining, subdictionary subregions_BoxCorner_files 32

B Tables - overview about possible options **33**
B.1 General configuration, dictionary datamining, supported resample frequencies 33
B.2 General configuration, dictionary datamining, supported resample methods 33
B.3 General configuration, dictionary datamining, supported temporal aggregation methods 33
B.4 General configuration, dataset settings, Correction of Parameter 34

C Overview of available subregions in the geopackage file **35**

D Jupyter notebook to visualize and add custom subregions **45**

E Available metrics in EvaSuite **49**

Chapter 1

General introduction to HZG Evaluation Suite

The HZG Evaluation Suite is a software tool developed for the evaluation of climate model simulations by comparing model data against gridded reference datasets, which can be either observation or model-based. The recent version of the HZG Evaluation Suite is v0.92.2. Within this document we refer to the HZG Evaluation Suite as EvaSuite.

EvaSuite enables the application of various metrics over user-defined temporal selections, providing flexibility for diverse research applications. EvaSuite can operate as a fully automated component, for instance, as part of the Starter Package for ICON-CLM Experiments (SPICE), or as a stand-alone application.

In this documentation you will find guidance, how to quickly get started with EvaSuite in chapter 2. We describe, which datasets can be analyzed using EvaSuite in chapter 3 and the internal workflow in chapter 4. In the subsequent chapters you will find how to obtain and install (chapter 5), configure (chapter 6), and run EvaSuite (chapter 7). Chapter 7 also includes hands-on examples to familiarize with EvaSuite. Finally, we provide supplementary tools related to using and defining subregions in chapter 8.

In the appendices you will find further helpful information regarding configurations (appendices A and B), subregions (appendices C and D), as well as the metrics included in EvaSuite (appendix E).

Chapter 2

Quick start of EvaSuite on Levante

In this chapter we describe how to quickly get started using EvaSuite on the computing cluster Levante hosted by the German Scientific Computing Center (DKRZ) in Hamburg.

First, obtain the source code of EvaSuite as described in section 5.2.1. Second, follow the steps in section 5.2.2 to install EvaSuite on Levante. This section will guide you through installing required R components (section 5.2.2), setting up a pre-defined conda environment (section 5.2.2), and running the installation (section 5.2.2). Third, obtain prepared example scripts including small sample datasets (section 7.1.3) and finally adjust and run an example evaluation (section 7.1.3). We also provided two further examples in sections 7.1.3 and 7.1.3 in case you want to familiarize yourself with running EvaSuite via different pathways.

Chapter 3

Which datasets can be analyzed?

3.1 Requirements on the datasets

The data format which is supported by HZG Evaluation Suite Version 0.92.2 is 'netCDF'.

Due to different operations (like spatial interpolation), it is needed to use HZG Evaluation Suite Version 0.92.2 only with datasets, which fulfill minimum requirements:

- The horizontal grid should be a regular lon-lat grid, a regular rotated grid, a single geographical location (as in the case of measurement data), a Gaussian grid, or an irregular grid containing lon-lat information for each grid point.
- The vertical grid should be coded in each file and for each variable in the dataset.
- Each variable in a dataset should possess a variable name without special characters except '_' and '-' and '.'.
- Each variable in a dataset should hold the attributes 'long_name', 'units' and, for the sake of perfectness 'cell_methods'.
- The time dimension should be always contained. The time bounds are very important to make the program understand the 'cell_methods'. It is not allowed to leave out the time coordinate and to code the temporal information in the filename.
- As in HZG Evaluation Suite Version 0.92.2 all data files related to a specific dataset have to be located in such a way, that a regular expression can be used to catch all of them, i.e. for each dataset only a specifc regular expression is searched for data for the evaluation.

3.2 The regular rotated grid

This grid is often used by regional models (COSMO, REMO, and ICON as output option). All files relevant for evaluation have to fulfill the following requirements on metadata

- The variables 'rlon' and 'rlat' should be contained in each file. Other names should be replaced by the standard names using h nco's. 'rlon' and 'rlat' have to be only one-dimensional.
- Each variable should contain the attribute 'grid_mapping' having the value 'rotated_pole'. Otherwise interpolations cannot be treated in a correct way.
- It is good practice for netcdf-conventions to also put lon's and lat's in a file and to deliver a coordinates attribute for each data variable. If this is not the case, the program could fail w.r.t. horizontal interpolations.

3.3 The regular lon-lat grid

This grid is ofen used by regional models or global models. All files relevant for evaluation have to fulfill the following requirements on metadata

- The coordinate variables 'lon' and 'lat' should be contained in each file. Other names should be replaced by the standard names using nco's. 'lon' and 'lat' have to be only one-dimensional.
- 'grid_mapping' or 'coordinate' attributes are superfluous because the grid is already meant to be located on the sphere with geographical locations.

3.4 The Lambert Conic conformal grid

This grid is sometimes used by regional models like WRF or by providers of measurement data. Starting with version HZG Evaluation Suite Version 0.23.0 , this grid type is allowed but only centre grid points can be examined. The support for the grid is realized by setting it to a curvilinear grid, i.e. the cdo's are only evaluating the longitudes and latitudes contained in the file. All files relevant for evaluation have to fulfill the following requirements on metadata

- The coordinate variables 'lon' and 'lat' should be contained in each file. Other names should be replaced by the standard names using nco's.
- With respect to each data variable, 'grid_mapping' attribute is set to 'lambert_conformal_conic', and the 'coordinate' attribute is set to 'lon lat'.

In addition, a variable 'lambert_conformal_conic' is contained with the following attributes having exemplary values:

```
int lambert_conformal_conic ;
  lambert_conformal_conic:grid_mapping_name = "lambert_conformal_conic" ;
  lambert_conformal_conic:standard_parallel = 46L, 49L ;
  lambert_conformal_conic:latitude_of_projection_origin = 47.5 ;
  lambert_conformal_conic:longitude_of_central_meridian = 13.3333333333333
     ;
  lambert_conformal_conic:false_easting = 400000L ;
  lambert_conformal_conic:false_northing = 400000L ;
  lambert_conformal_conic:earth_radius = 6367470.f ;
  lambert_conformal_conic:radius = 6367470.f ;
```

To use the LCC2 grid, one has to tell the EvaSuite that during data mining with cdo's the grid-describing variable above-mentioned should be repaired, i.e. the string 'lambert_conformal_conic' should be added to the list 'RepairNetcdf_variables' given in the file 'config_SUITEgeneral.py'. Otherwise, the variable is thrown away by cdo's.

3.5 Single locations - regular rotated or lon-lat grid

This grid is often used for point measurements. All files relevant for evaluation have to fulfill one of the following requirements on metadata

1. The coordinate variables 'lon' and 'lat' (singleton dimension) are contained in each file.
2. The coordinate variables 'rlon' and 'rlat' are contained in each file. A 'grid_mapping' and 'coordinates' attribute have to be contained in each file. Moreover, the rotated pole has to be given as netCDF variable.

3.6 Time dimension

The time variable has to have at least a dimension of one. It is not allowed to have no time variable / dimension in a netCDF file. The time units attribute has to be contained as well as the calendar attribute. If the latter is not given then the calendar 'proleptic_gregorian' is taken. Time bounds are not required but it is good practice to code them for each data variable in addition to the time coordinate. Otherwise the 'cell_methods' are not interpreted correctly.

Up to version HZG Evaluation Suite Version 0.92.2, only a relative time coordinate is allowed. The format 'days as ...' leads to conflicts when interpreting the date stamps. One needs to use cdo's to set a relative time axis prior to the evaluation run.

3.7 NetCDF data variables

The best option is that data follow the NetCDF Climate and Forecast (CF) Metadata Conventions. It is a must that the data variables (excluding the coordinate variables) have the attribute standard_name. In addition, the attributes 'long_name' and 'units' should be contained in the netCDF. Otherwise, the program provides the results of evaluations (nc-files and figures) with some strange units and variable descriptions. Moreover, the evaluation can crash if 'units' are missing.

Chapter 4

How does it work? - Internal workflow of EvaSuite

The general workflow of the Evaluation Suite is as follows (this is described here for the mode Eval-RAW):

1. Assemble all necessary input datasets for the evaluation of the indicated variables and periods (corresponds to CDO selvar and seltime)
2. Remap data to target grid and region (corresponds to CDO remap...).
3. Sample/Select time steps that will provide the basis for the evaluation (corresponds to e.g. CDO monmean for resampling time series to monthly mean values)
4. Calculation of the specified evaluation statistics (this can involve spatial or further temporal averaging). This step can be repeated without going through the 3 prior steps by using the option `eval_StartingPoint = 'StartEvaCompute'`.

The data produced during the above steps can be found in the output directory DataMining, which contains the following 4 subdirectories:

1. WP01_Preperation
2. WP02_UniqueGrid
3. WP03_EvalProcess
4. WP04_EvalFinal

Chapter 5

How to obtain and install EvaSuite

5.1 Requirements on the software

As of HZG Evaluation Suite Version 0.92.2, the following packages or software need to be installed on the computer:

- The statistical software 'R',
- The R-package 'ncdf4Utils' by Burkhardt Rockel (needed to create 2D-x-y plots). The package must be compatible with the R version running - check by starting 'R' and typing in 'library(ncdf4Utils)'. If the library cannot be found, search for the path regarding 'ncdf4Utils' and add the autmoatic load in your '∼/.Rprofile'.
- The python-package 'pypeR' (needed to communicate with 'R')
- The python-package 'Xarray' (needed for the whole datamining. Version 0.9.6 is required at minimum, because earlier version were not tested. Version 0.10.0 is required for using the 'resample' functionality in combination with strange formulae - as it is the case for spatial or temporal correlations.
- The climate data operators 'cdo',
- The python-package 'Pycdo' for having nice communication with the climate data operators,
- The ncar netCDF Operator 'nco'.

We describe in the following sections how to install the required software packages and EvaSuite itself.

5.2 Installing the stand-alone version of the HZG Evaluation Suite Version 0.92.2

5.2.1 Obtain program code

Download the tar-file from `https://swift.dkrz.de/v1/dkrz_c0725fe8741c474b97f291aac57f268f/clmcom/spice/HZG_Evaluation_Suite-v0.92.2.tar` and untar using

`tar -xf HZG_Evaluation_Suite-v0.92.2.tar` .

On Levante, we recommend to untar in your home directory.

5.2.2 Installation on Levante

In the following sections it is explained how to install HZG Evaluation Suite Version 0.92.2 on the high performance computing cluster Levante at the Deutsche Klimarechenzentrum, Hamburg. Although the description is cluster specific, it may provide helpful information for the installation on other clusters, too.

Preparation: Installation of R components of Evasuite

The following steps are required independently of how you obtained evasuite.

First, load the required modules via

`module purge`

`module load gcc r` .

Second, define and create the required directories. PATH_TO_EVASUITE_CODE_DIR needs to be adjusted to the directory, where the installation code of HZG Evaluation Suite Version 0.92.2 is located. E.g. $HOME/HZG_Evaluation_Suite-v0.92.1beta if you untared the above file in your home directory.

`INSTALLDIR=$HOME/R/x86_64-pc-linux-gnu-library/4.1`

`LIBDIR=PATH_TO_EVASUITE_CODE_DIR/R-packages/Rlibs`

`mkdir -p $INSTALLDIR`

Third, install the required R-packages, which are not yet installed on Levante. All other R-packages should already be installed.

`R --no-init-file CMD INSTALL -l $INSTALLDIR $LIBDIR/mapdata_2.2-6.tar.gz`

`R --no-init-file CMD INSTALL -l $INSTALLDIR $LIBDIR/PCICt_0.5-4.tar.gz`

`R --no-init-file CMD INSTALL -l $INSTALLDIR $LIBDIR/ncdf4Utils-2.1.4.tar.gz`

Finally, create an R-profile. Create the file $HOME/.R/libraries.Rprof with the content below (if it does not exist yet). This may require to first create the directory $HOME/.R/. Adjust the path given for ncdf.lib.dir to your account number.

```
ncdf.lib.dir="/home/b/bYOURACCOUNTNUMBER/R/x86_64-pc-linux-gnu-library/4.1/"
#------------------------------------------------------------------------
#--------- adopt the library settings to your machine
#            if a library is already installed you can skip it
#------------------------------------------------------------------------

#example for the dkrz-levante system (only the ncdf4utils is enough for
    plotmap)
  library(PCICt, lib.loc=ncdf.lib.dir)
  library(mapdata, lib.loc=ncdf.lib.dir)
  library(ncdf4Utils, lib.loc=ncdf.lib.dir)
```

Adjust the path to R-profile in the configuration file
PATH_TO_EVASUITE_CODE_DIR/hereon_evasuite/system/evasuite.sysconf.yaml‘

`profile: "/home/b/bYOURACCOUNTNUMBER/.R/libraries.Rprof"`

Set up a conda environment for installing evasuite on Levante

Using this way of setting up a conda environment makes sure that the environment suites the current version of evasuite and that no additional packages need to be downloaded during the installation.

Download the files to set up the conda environment from `https://swift.dkrz.de/v1/dkrz_c0725fe8741c474b97f291aac57f268f/clmcom/spice/evasuite_conda_environment.tar` to your cluster and untar the archive. Afterwards change into the untared directory and run

`./deploy-env.sh EvaSuite_conda_environment.tar.gz` .

This will automatically install the conda environment ”EvaSuite_conda_environment”.

Run installation

Load the required modules via

`module purge`

`module load gcc r`

and activate the environment created in the previous section

`conda activate EvaSuite_conda_environment` .

Change to the directory, which contains the file setup.py. E.g. $HOME/HZG_Evaluation_Suite-v0.92.1beta if you untared the compressed code in your home directory. Then run the python package installation `pip install ./` . The installation output should end with:

```
Successfully built hereon-evasuite

WARNING: The script evasuite is installed in '/home/b/bYOURACCOUNTNUMBER/.
    local/bin' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress this
    warning, use --no-warn-script-location.

Successfully installed hereon-evasuite-0.93
```

You can test whether the installation was succesful by running the command `evasuite --help`.

Optionally: Add the path, where evasuite was installed to, to your $PATH variable by adding the line `PATH=$PATH:~/.local/bin` to the file ~/.profile .

Note: In order to later execute evasuite, you need to activate the environment created above.

5.2.3 Update evasuite

In order to update evasuite, simply run the installation as described in 5.2.2 again. This will uninstall the existing version first and install evasuite again. If updating evasuite this way you can skip the installation of R-packages. However, you might need to adjust the path in 'hereon_evasuite/system/evasuite.sysconf.yaml' as described there.

5.2.4 Uninstall evasuite

You can uninstall evasuite via pip: `pip uninstall hereon_evasuite`. Confirmation with Y should e.g. yield 'Successfully uninstalled hereon-evasuite-0.92.0b0'.

Uninstalling evasuite removes the following directories in the conda environment, into which you installed evasuite

```
../bin/evasuite

../lib/python3.10/site-packages/hereon\_evasuite-0.92.0b0.dist-info/*

../lib/python3.10/site-packages/hereon\_evasuite/*
```

Chapter 6

How to configure EvaSuite

Three configuration files control the behaviour of the HZG Evaluation Suite Version 0.92.2: a configuration for the evaluation itself, a configuration for the visualization settings, and a configuration file for advanced users specifying all evaluation quantities and evaluation modes. The evaluation suite reads these files as so-called dictionaries and extracts all the information needed to make an evaluation run.

6.1 Controlling the evaluation - main config file

The first line of this file is commented and contains the program version: `# !v0.16`. This line should not be removed because since Version 0.17 the program is checking for the consistency between configuration files and program version. As of HZG Evaluation Suite Version 0.92.2, a warning is printed out if the program version cannot be detected due to missing git-repository.

All the options are explained in the following table.

6.1.1 General configuration

Option (reject hyphen)	Type	Description + Example
eval_StartingPoint	String	Define where to start with the evaluation. Possible values are 1. 'cleanall': Everything located in the directory 'eval_workdir' is cleaned and the Evaluation run starts from scratch 2. 'StartEvaCompute': The computation starts from all datasets being interpolated to the same grid. This is useful for changing something like data aggregation frequency without doing the full evaluation run. `eval_StartingPoint = 'StartEvaCompute'`
eval_workdir	String	The directory where the evaluation process takes place including all operations on file and intermediate results are stored. The directory contains a subdir 'Datamining' and a subdir 'Visualizations'. Also, the configuration files are written in the workdir. `eval_workdir = "/scratch/m/m221054/HZG_WORK/Suite-Out/"`
eval_backup	Boolean	Configurations, final results and visualizations are backuped after evaluation run into a hashtagged directory named as the 'eval_workdir' but adding a description (datasets, evaluation parameters, evaluation period, temporal aggregation) and a hashtag. The hashtag is computed from the content of the three configuration files. `eval_backup = True`
eval_backup_dir	String	The 'eval_backup' switch activates the backup of results into a hashtagged directory but it can be overwritten by setting explicitly here a directory name.

eval_Mode	String	Define the type of evaluation and the related visualization. As of HZG Evaluation Suite Version 0.92.2 two types are implemented, i.e. the user can choose between 1. 'CompareRAW': All the datasets are compared to each other w.r.t. the raw data. 2. 'EvalRAW' : The datasets specified in 'data2proof_acronyms' are evaluated against the datasets specified in 'evaldata_acronyms'. The user is free in specifying the list of metric quantities (see 'eval_metrics') `eval_Mode= "EvalRAW"`
eval_Recomp-Var_type	String	Allow for the complex manipulation of data prior evaluation or comparison. This setting is predefined with 'False', i.e. the user does not has to define it in the configuration. As of HZG Evaluation Suite Version 0.92.2 only one setting is possible 1. 'grad_z': The gradient in the vertical direction is calculated but only in case of levels with a height coordinate. 2. 'setvalidrange;min;max' : All values in the datasets outside the range 'min' and 'max' are set to nan (e. g. radiation at night not considered) 3. 'setmaskedrange;min;max': All values in the datasets within the range 'min' and 'max' are set 1 otherwise to nan. 4. False: No manipulation of the raw data is done. `eval_RecompVar_type = 'grad_z'`
eval_parameters	List of one string	Define which parameters to evaluate. As of HZG Evaluation Suite Version 0.92.2 only one parameter is allowed but the format has to be a list. The user is free in defining a name for a parameter, but for each dataset the user has to specify which variable name is associated with the parameter. `eval_parameters = ["temperature"]`

eval_metrics	List of strings	The metric that should be considered for evaluation. If the 'eval_Mode' is set to 'CompareRAW', this settings in nullified. If the 'eval_Mode' is set to 'EvalRAW', the user is free to choose a list of evaluation measures. At the moment, the following metrics are supported 1. 'BIAS_temporal' : the mean error (bias) for each location over time, 2. 'BIASnorm_temporal' : the mean error (bias) for each location over time, but normalized by the temporal mean, 3. 'BIAS_spatial' : the mean error (bias) for each time over the 2D field, but normalized by the spatial mean, 4. 'MAE_temporal' : the mean absolute error for each location over time, 5. 'MAE_spatial' : the mean absolute error for each time over the 2D field, 6. 'RMSE_temporal' : the root mean square error for each location over time, 7. 'RMSE_spatial' : the root mean square error for each time over the 2D field, 8. 'MAD_temporal' : the mean absolute deviation for each location over time, $MAE(x,y) = MEDIAN(MOD(x,y,t) - MEDIAN(OBS(x,y,t)))$, 9. 'MAD_spatial' : the mean absolute deviation for each time over the 2D field, 10. 'LinCorr_temporal' : the linear correlation for each location over time, 11. 'LinCorr_spatial' : the Pearson correlation coefficient over the 2D field and for each time, 12. 'KlingGupta_temporal': the Kling-Gupta efficiency for each location over time 13. 'quantile_bias_temporal' : the bias of quantiles for each location, computed from the PDF of all values in time. 14. 'quantile_mae_temporal' : as for 'quantile_bias_temporal', but the mean absolute error of quantiles for each location, computed from the PDF of all values in time. 15. 'quantile_iqd_temporal' : as for 'quantile_bias_temporal', but the inverse quadratic distance of quantiles for each location, computed from the PDF of all values in time. 16. 'pdf_pss_temporal' : the Person-Skill-score for each location, computed from temporal distribution. 17. 'AMSESS_temporal' : the advanced mean squared error skill score for each location over time, which is a normalization of the original mean squared error skill score. 18. 'AMSESSsigm_temporal' : the significance for RMSE-based skill score (AMSESS) of model A to be better than those of model B (significance level of 0.10). 19. 'AMSESSsigs_temporal' : the significance for RMSE-based skill score (AMSESS) of model A to be better than those of model B (significance level of 0.05). 20. 'BIASsigBSm_temporal' : the significance for the BIAS of model A to be smaller than those of model B (significance level of 0.10). 21. 'BIASsigBSs_temporal' : the significance for the BIAS of model A to be smaller than those of model B (significance level of 0.05). [In future releases with additional evaluation modes implemented (like Taylor-Diagram), the value here is overwritten with the setting given in the configuration file 'config_EVALmode_metrics'] `eval_metrics = ["RMSE_temporal","AME_temporal"]`

data2proof_acronyms	List of strings	A list of datasets which are proved by the evaluation tool (often model data). One has to use acronyms without special characters except '_' and '-'. The datasets given here have to be specified in more detail by the settings in 'data2proof_settings', i.e. all acronyms given here have to appear also in 'data2proof_settings'. In the case of skill score related evaluation measures the first dataset in the list is taken as the 'skill-score-reference' where all other datasets are then the 'skill-score' test data. `data2proof_acronyms = ["CEL202","CEL204"]`
evaldata_acronyms	List of strings	A list of datasets which are the reference data for the evaluation tool (often observation, but also a reference model is possible). One has to use acronyms without special characters except '_' and '-'. The datasets given here have to be specified in more detail by the settings in 'data2proof_settings', i.e. all acronyms given here have to appear also in 'data2proof_settings'. `evaldata_acronyms = ["TRY_v001"]`
evaldata_combine_method	String	Specify the treatment of multiple reference data sets ('evaldata_acronyms'). The user can choose to combine the reference data sets during the evaluation process by specifying one of the following options 1. 'mean' : The reference data sets are averaged to a new datasetroot mean square error in temporal space. 2. 'minmax' : The range of the reference datasets is determined and as a result two reference data sets are calculated (a minimum and a maximum data. However, as of HZG Evaluation Suite Version 0.92.2 this feature does not work with the evaluation mode 'EvalRAW'. 3. '-' : The reference datasets are taken as they are. Thus, for multiple reference data sets the evaluation is done for each of them. `evaldata_combine_method = "-"`
Visual_program	Dictionary	Programs to use for visualization and related settings. The dictionary contains for each program a key with the program name and a value which is a new dictionary with the following items: 1. 'default-version' : (boolean) This is an indicator that decides to use the standard version of the program read from the environmental variable 'PATH'. If set to .false., the user has to give a 'binary-location'. 2. 'version' : (string) The program version (not used). 3. 'binary-location' : (string) The location of the binary which should be used for the evaluation run. An example for this dictionary is given here. Usually, only one program is used but a mixture is possible in theory.
Masking_program	Dictionary	Programs to use for masking sub-domains and related settings. The dictionary has the same structure as for the 'Visual_program' but with two exceptions regarding the items: 1. 'parent-prog' : (string) Indicates if this program has a dependency on another program (e.g. the standard settings uses a mask maker has written at Hereon which requires the statistical software 'R'. 2. 'RScriptname' : (string) Indicates the name of the script which has to be called. An example for this dictionary is given here. Usually, only one program is used.

DataDump-Process_programs	Dictio-nary	Programs to use for data processing and related settings. It is required to provide settings for cdo's and nco's. Ncdump is not needed at all as of HZG Evaluation Suite Version 0.92.2. The detailed settings for Python-xarray are not evaluated, but the key must exist. The dictionary has the same structure as for the 'Visual_program' but with one exception regarding the item: 1. 'ompthreads' : (integer) Number of threads used by the program. However, only a small amount of operations are really parallelized. An example for this dictionary is given here. Usually, a mixture of programs is used. For earlier versions, it was important to use nco's at least on version 4.6.7. Otherwise important functions could not be executed.
DataMain-ProgramToUse	String	The name of the program should be used in first order for data operations. The name must be consistent with one of the keys in 'DataDump-Process_programs' related to nco's or cdo's. This setting only affects a small number of functions (extracting time stamps, ...) for which the user can choose its preferred tool. `DataMainProgramToUse = "cdo"`

6.1.2 Data mining

The option 'datamining' is a huge dictionary controlling the whole data processing. The dictionary contains many settings explained below in the table. It is most important to mention that the settings can be subdivided into two groups.

One group contains all the settings up to 'interpol_method' and they are fixed for all evaluation runs using the option `eval_StartingPoint = 'StartEvaCompute'`. In other words, if the user wants to evaluate another period of interest or wants to redefine the reference spatial grid, he has to rerun the full evaluation suite using `eval_StartingPoint = 'cleanall'`.

The second group contains all settings starting at 'target_EvalFrequency' and they can be changed for rerunning the evaluation using the option `eval_StartingPoint = 'StartEvaCompute'`. Thus, many evaluations can be performed without always treating a huge amount of data processing steps.

Dictionary key (reject hyphen)	Type	Description + Example
targettime_full-span	List of two strings	Define the time period to examine. Specifying a list of two comma-separated strings containing the beginning and the end in format "YYYY-MM-DD HH:MM:SS". If the user wants only a specific date, he has to give '[begin,end]'. `["2007-01-01 00:00:00", "2007-12-31 23:00:00"]`
targettime_season	List with one String	Define which period of the year to examine, i.e. cut out from 'target-time_fullspan' only a specific time of the year. The following settings are allowed 1. 'complete' : All data are taken between the two dates given by 'targettime_fullspan'. 2. 'Month-XX': A specific month is cut out, where 'XX' has to be replaced by two two-digit month number. 3. 'winter', 'autumn', 'summer', 'spring' : A specific season is cutted out (this option is buggy and should not be used `"targettime_season": ["complete"]`

targetgrid_horizontal	String	Indicates the reference horizontal spatial grid on which the evaluation is based. This value has to be consistent with one of the 'data2proof_acronyms' and 'evaldata_acronyms'. All datasets given in 'data2proof_acronyms' and 'evaldata_acronyms' are interpolated to that grid, which is contained by the dataset given here. The interpolation method is specified with the option 'interpol_method' of the dictionary 'datamining'. The program recognizes automatically equal grids, i.e. if the data are already on the reference grid, no interpolation is done. If the user gives a '-' then nothing is done. `"targetgrid_horizontal": "CEL202"`
addDS_on_targrid_hori	List of strings	Indicates which datasets given in 'data2proof_acronyms' and 'evaldata_acronyms' are on the same grid as the reference horizontal spatial grid. This value has to be consistent with one of the 'data2proof_acronyms' and 'evaldata_acronyms'. This option allows to control which datasets are finally interpolated to the reference horizontal grid. It is required for those cases, where a lot of datasets are on the same grid as the reference grid. There is no automatic detection to find out about equal horizontal grids in the various datasets. In case that no dataset has the same grid as the dataset with the reference grid, set to 'None'. `"addDS_on_targrid_hori": ["CEL203","CEL204","CEL205"]`
target_RollingMean	Integer	Defines the number of consecutive time steps over which the reference data and the model data are averaged in a temporal window shifted through the time dimension. The number of time steps is related to the setting in 'target EvalFrequency'. If the frequency for evaluation is daily values, then a setting of 'target_RollingMean=5' leads to 5 day rolling means for evaluations or comparisons. If no rolling mean should be computed, set to 'None'. `"target_Rollingmean": 5`
targetgrid_vertical	String	The same as for the option 'targetgrid_horizontal', but now for a reference vertical grid. Only height levels are supported in HZG Evaluation Suite Version 0.92.2. The type of vertical levels has to be given for each dataset using the switch 'levtype' (see dataset dictionaries). The interpolation to height levels is done linearly. `"targetgrid_vertical": "-"`
targetdom_horizontal	Dictionary	Cut out only a part of the whole domain covered by the data set with the horizontal reference grid (given in 'targetgrid_horizontal'). Thus, the settings are related to the horizontal reference grid and only the related data set is cut according to the user-setting and prior interpolation. The setting requires the key 'False' or a dictionary with items corresponding to the dimension to cut along and a list of two values indicating the start and end grid index. For instance, in the case of two horizontal dimensions the first item can be 'x-dimension:[10,20]'. The switch is predefined with 'False', i.e. the user does not have to define it in the configuration. `"targetdom_horizontal": 'rlon': [100,300], 'rlat': [50,250]`
targetdom_vertical	Dictionary	The same as for the option 'targetdom_horizontal' but with respect to the vertical coordinates. Thus, some vertical slices are cut out of the dataset holding the vertical reference grid (given in 'targetgrid_vertical'). The setting requires the key 'False' or a dictionary with the item corresponding to the level to cut along and a list of two values indicating the start and end grid index. The switch is predefined with 'False', i.e. the user does not has to define it in the configuration. `"targetdom_vertical": 'height': [2,4]`

interpol_method	String	The method used for horizontal interpolation, but as of HZG Evaluation Suite Version 0.92.2 stuck to 'linear'. [In future versions this would be a list of strings indicating for each evaluation parameter the interpolation method.] `"interpol_method": "bilinear"`
target_EvalFrequency	Dictionary	Define the temporal resolution on which the evaluation measures or the comparison is based on, i.e. the raw data are recomputed to the frequency given here and afterward the evaluation starts. There are 3 possible options for the dictionary: 1. 'method' : '-': Nothing is done with the raw data. 2. 'method' : 'reselect','targetgrid': gridtime, 'tolerance': tol: The date stamps in all raw data are reselected consistent to the date stamps in the dataset specified by 'gridtime'. The value 'gridtime' has to be consistent with one of the acronyms given in 'data2proof_acronyms' and 'evaldata_acronyms'. The reselection process searches with the nearest neighbour method (no interpolation is done and the final time stamps in the netcdf's are the same as in gridtime). The tolerance for the reselection can be specified using option tol (e.g. '1H' means one hour, '30M' means thirty minutes). This option is usually utilized in case of measurement data with data gaps in time. 3. 'method' : 'resample','freq' : freqv, 'how' : howv: Define for each data set the new temporal frequency and the method how to get there. Details about the supported keywords for 'freqv' and 'howv' can be found in the tables here and here An example for the three possibilites is given here.
aggregate_temporal	String	Indicates how to aggregate the evaluation statistics calculated. Please do not mix it up with the setting for 'target_EvalFrequency'. As an example: You have one dataset with hourly data and two datasets B and C with 10-minute data. Using the option 'target_EvalFrequency' to select for all datasets only the time stamps of dataset A, i.e. all data are now hourly data. Now you can use `"aggregate_temporal": "weekly"` to determine the BIAS of hourly data on a weekly aggregation period, i.e. if the raw data contained one year of data, you get a timeseries with 53 values for the BIAS. Possible values for the setting 'aggregate_temporal' can be found here. A special keyword is 'fullperiod' which means to aggregate the statistics over the whole time (in the afore-mentioned case only one BIAS for the whole year). `"aggregate_temporal": "3-daily"`
subregion_acronyms	List of strings / None	Defines the sub-domains of interest or one gives 'False' as a keyword. If this setting is not equal to 'False' then evaluation statistics are prepared for each sub-domain given here in the list. The names given here must be consistent to the header line of the files given in the setting 'subregions_BoxCorner_files'. If the program find a keyword in the sub-domain list also in the header line of a specific file, then the polygon in this file is associated with the subdomain. If not, the subdomain is automatically removed from further evaluation. `"subregions_acronyms": ["NI"]`

subregions_BoxCorner_files	Dictionary	The key 'polyfile' is a string and defines the location of the polygone file (format geopackage) which contains the polygons given in the list 'subregion_acronyms'. The key 'attribregion' is a string and defines the column of the attribute table in the polygon geopackage file which contains the names of the polygons. Thus, the names of polygons has to match with the 'subregions_acronyms' `"subregions_BoxCorner_files": {"polyfile": "polygons.gpkg", "attribregion": "NAME_0"}`
mask_Orography	Dictionary	Allows to analyze regions dependent on their altitude (height above sea level). Key 'method' can be set to 'threshold' or 'ranges'. If 'threshold' is chosen, each of the subregions given in 'subregion_acronyms' is divided further into two subsubregions dependent on the altitude given by the key 'value'. I. e. all points located below the altitude of 'value' are considered as subsubregion1 and all values above are considered as subsubregion2. If the key 'method' is set to 'ranges', each of the subregions is devided further into many subsubregions dependent on the altitudes given by the key 'limits', which is a list of real or integer numbers. I. e. all values in the list define an altitude range for which data points are filtered and analyzed: '[200,1000,2000]' separates between data points located between 0 and 200 m altitude, 200 and 1000 m altitude, 1000 and 2000 m altitude and points located higher than 2000 m altitude. Set this parameter to 'None' or 'False' to deactivate the orographic based masking. `"mask_Orography": {"method": "ranges", "value": False, "limits": [0,200,1000,2000]}`
mask_dataf_Orography	String	Defines the location of the data with orography in order to analyze subsubregions based on orography. `"mask_dataf_Orography": "orography.nc"`
mask_landsea		
mask_dataf_landsea	...	...
operator_time	String	Considered only in case of `eval_Mode= "CompareRAW"`. For simple comparison, this setting defines with which method the temporal aggregation given in 'aggregate_temporal' is done. For an evaluation with measure 'RMSE' it is predefined but in the case of a simple comparison, the temporal aggregation is free of choice. Possible values are 'mean', 'sum', 'max', and 'std' and '-'. `"operator_time": "mean"`
operator_spatial_x	String	Considered only in case of `eval_Mode= "CompareRAW"`. As for 'operator_spatial_t' but for the spatial aggregation along the x-axis. `"operator_spatial_x": "mean"`
operator_spatial_y	String	Considered only in case of `eval_Mode= "CompareRAW"`. As for 'operator_spatial_x' but for the spatial aggregation along the y-axis. Only in case of no aggregation along the x-axis, the user is free of choice. If there is already an operation along the x-axis, both operators are coupled. `"operator_spatial_y": "mean"`
operator_spatial_z	String	Not implemented as of HZG Evaluation Suite Version 0.92.2.

6.1.3 Professional settings for the Suite and netcdf files

At the end of the general configuration file there is a section where the names of directories created by the program can be manipulated: 'directory_visualize', 'directory_datamining', 'directory_maskregions', 'directoryDM_WPprepare' and 'directoryDM_WPcompeval' (the last are lists with two entries).

Moreover, the user has to define which variables and attributes have to be preserved during data processing: 'RepairNetcdf_variables' and 'RepairNetcdf_attributes' are lists of strings. The preserved variables and attributes are searched in the raw data and if they get lost due to data processing, an automatic repair process is called prior interpolation, grid masking and visualization. 'RepairNetcdf_targetgrid' is of particular interest because here the variable names are given which contain coordinate information of the target grid.

In addition, 'PossibleDimNames' specifies which coordinate variables have to be associated with the spatial, temporal, and vertical dimensions. Otherwise the program, for instance, does not know how to compute a mean on the spatial field.

6.2 Settings for the datasets - model and proof data files

The two dictionaries 'data2proof_settings' and 'evaldata_settings' contain all information about the raw data of these datasets defined by the acronyms in the settings 'data2proof_acronyms' and 'evaldata_acronyms'. That is all acronyms given have to be specified in the dictionaries 'data2proof_settings' and 'evaldata_settings'. Both dictionaries contain the following settings:

Dictionary key (reject hyphen)	Type	Description + Example
gridtype	String	Defines the type of grid for the raw data. As of HZG Evaluation Suite Version 0.92.2 possible values are 'gridded' and 'station'. The last option is for the point observations. `"gridtype": "gridded"`
format_data	String	Defines in which format the raw data are given. Only 'netcdf' is supported. `"format_data": "netcdf"`
levtype	String	The vertical coordinate type of the raw data. As of HZG Evaluation Suite Version 0.92.2 only 'height is supported. This setting only matters if it comes to a vertical interpolation. In that case complex 3D-height coordinates or pressure levels are not allowed in the recent version. `"levtype": "height"`
storagedir	String	The location where the data are stored. If the data are distributed over many subdirectories, one should only give the root directory here. Everything else is handled by the option 'storagefil' `"storagedir": "/pool/data/CCLM-CRCS/output_CEL/"`
storagefil	String	The name of the files. One can set single files or multiple files using regular expressions not only for files but also subdirectories. As of HZG Evaluation Suite Version 0.92.2 only one evaluation parameter is allowed. This parameter can be distributed over many files (e.g. monthly timeseries). However, in future more than one evaluation parameter is allowed: the parameters should then be packed in one file but the time can be distributed over many files. `"storagefil": "CEL*/CEL202.*.2007.nc"`
paramdict	Dictionary	Defines which eval parameter belongs to which variable name in the netcdf file. Using this setting the evaluation program knows how to associate the 'eval_parameters' with netcdf data. `"paramdict": {"temperature": "T_2M", "precip": "TOT_PREC"}`

paramcorr	Dictionary	Defines for each variable name given in 'paramdict' the possibility to correct the data w.r.t. to units. If the key 'unitcor' is true, one can specify with key 'corr_vals : [oper, number, newunit]' how to manipulate the data using '*', '-', '+' or '/' as a mathematical operator and 'number' as value as well as 'newunit' as new unit. The 'paramcorr' dictionary also allows for time shifts. If the key 'timeshift' is true, one can specify with the key 'timcor_val : [oper,number,unit]' how to shift the time axis using '+' or '-' as operator. 'number' as well as 'unit' define how much hours (unit='h') or minutes (unit='m') the shift should be. An example for the setting is given here.

6.3 Controlling the visualisation

The visualization configuration is read via a YAML-file which allows for comments using '#' at the beginning of the line. The YAML-file is internally handled as a dictionary.

Option (reject hyphen)	Type	Description + Example
output_ld0D1D	string	Define the output type for 0d and 1d results. Possible values are 1. 'ascii': If the final result of calculations is only point data or 1D-data, then the output is written in ASCII-Format. 2. 'netcdf': If the final result of calculations is only point data or 1D-data, then the output is written in NETCDF-Format. `output_ld0D1D = 'netcdf'`
add_maps2D	Logical	Decide if the final results with two-dimensional data are visualized (True) or not (False) `add_maps2D = True`
settings_density	dictionary	controls the behaviour of the probability density plots or Q-Q-plots 1. key 'bins': define the bins of the probability density function, i.e. a list of numerical values 2. key 'cumulative': a logical switch for getting a cumulative probability density function (True) or not (False) 3. key 'Widthperctil': in case of Q-Q plots one can control the spacing between the percentiles to analyze. Choosing '1' means that each percentile form 1 to 100 is determined and plotted. `"settings_density": {"bins": [2,3,4,5,6], "cumulative": False, "Widthperctil": 2}`

settings_1D	dictionary	controls the behaviour of the 1D-plots which encloses also the probability density plots 1. key 'plot_type': define the type of plot in case of one-dimensional data are plotted or if a probability density plot is desired. The option 'line' is for all plots related to usual evaluation with 1D-results (yearly cycles, ...). The option 'scatter' produces a scatter for the data points of two data sets. All other switches are related to a probability density analysis. 'hist' for histograms using bars for visualization, 'hist_line' for histograms using lines, 'qqplot' for Q-Q plots showing percentile comparison of two datasets, 'qqplot_scatter' showing in addition to 'qqplot' the scatter of data points in two datasets. The setting 'percentile_diff' shows the difference between the percentiles of two datasets. 2. key 'colors': a list of predefined color names (see matplotlib documentation) or hexadecimal codes used for the different lines in a 1D-plot. In case of comparisons, the list has to be as long as the number of datasets to be plotted (model datasets plus the reference dataset). In case of evaluations, the list has to be as long as the number of model datasets (data sets to be proofed). 3. key 'marker': Defines the type of how to mark single data points in the 1D-plots. One single value (a string) can define a unique marking for all lines. Possible values are '' for no marker, 'o' for circles as markers and 'x' for crosses as markers. If you want to control each line of the plot individually give a list of values that is at least the same length as the list of colours (or the number of data sets to be analyzed). 4. key 'linestyle': Define the line style in the 1D-plots in a manner equal to the key 'marker'. Give a single string value for a unique linestyle or a list of values. Values can be '' for no line as well as 'solid', 'dotted', 'dashed' and 'dashdot'. 5. key 'markersize': Defines the size of the marker in a manner equal to the key 'marker'. Give a single real number for a unique marker size or a list of real values. The behaviour of the marker size setting can be read in the matplotlib documentation.

settings_1D continued	dictionary	1. key 'changingYAxis': In case of complex evaluations with many datasets and many regions the number of 1D plots can be huge. The final results are plotted over many PDF-pages. The logical switch here allows you to change the y-axis on each pdf-page individually (setting 'True') or to keep the y-axis constant over all PDF-pages ('False'). 2. key 'max_datasets': Defines how many datasets are plotted into a single 1D-plot (integer value). In case of a comparison between model datasets and reference datasets, setting 'max_datasets' to '4' means that, in case the number of datasets is larger than 4, the plots are spread over many single plots. I. e. the first figure contains the reference dataset along with 3 model datasets and on the next plot the figure contains the reference dataset again and the remaining datasets. In case of an evaluation of model datasets on a reference dataset, the plots are also spread over many single figures if the number of model datasets exceeds the value of 'max_datasets'. I. e. setting 'max_datasets' to '4', the first plot contains the evaluation metric related to the reference data set and the first 4 model datasets. The next plot contains the evaluation metric related to the reference data set and the remaining model datasets. 3. key 'MultipleFigures_maxrows' and 'MultipleFigures_maxcols': define the number of plots next to each other (columns) and the number of plots below each other (rows) on a single PDF-page. Values are given as integer numbers. `"settings_density": {"bins": [2,3,4,5,6], "cumulative": False, "Widthperctil": 2}`

Chapter 7

How to run EvaSuite

7.1 Running the stand-alone version of EvaSuite

We here describe two options how to run evasuite. The first option directly operates on the configuration files of evasuite. The second option uses a wrapper script to set the configuration parameters of the evasuite and allows to execute multiple evaluations in a loop (e.g. for different variables or metrics). Furthermore, we provide three examples, how to use these different options.

7.1.1 Configure and run EvaSuite via config-files and sbatch

You can directly configure evasuite via the two configuration files for general settings and visualization. To do so, you can create and fill copies of the files "config_SUITEgeneral_TEMPLATE.yaml" and "config_visualization_TEMPLATE.yaml", which you find in the subdirectory "configuration" in the evasuite code directory that you created (section 5.2.1).

Furthermore, you need to make sure that the model and/or reference data that you want to analyze is contained in the model and reference configuration files you refer to in the general configuration file. To do so, you can copy and extend the model and reference configuration files "modeldata_settings_COPAT2_ex.yaml" and "reference_datasets_v2.yaml" in the subdirectory "inputs" in the evasuite code directory. We provide an example containing the four filled configuration files in section 7.1.3.

The command to run EvaSuite is (you need to replace the path to and the specific names of the configuration files):

```
evasuite --cfggen=path_to_EvaSuite_code_dir/configuration/your_GENERAL_namelist.yaml
--cfgvis=path_to_EvaSuite_code_dir/configuration/your_VISUALIZATION_namelist.yaml
--inprefdat=path_to_EvaSuite_code_dir/inputs/your_Reference_datasets.yaml
--inpproofdat=path_to_EvaSuite_code_dir/inputs/your_Modeldata_settings.yaml .
```

Note: Make sure to activate the conda environment, into which you installed evasuite 5.2.2.

Important note: The evaluation carried out by evasuite can be quite comprehensive. Therefore, we recommend running evasuite in an interactive session on a compute node or via submitting a batch script via sbatch as presented in the example in section 7.1.3.

7.1.2 Automisation: Run evasuite via a wrapper script

You can also use a wrapper script to run evasuite. This provides the option to submit multiple evaluations at the same time. In this case the template configuration files GENERAL_namelist.yaml and VISUALIZATION_namelist.yaml in the subdirectory "configuration" in the evasuite code directory are filled during the execution of the wrapper script. Afterwards, evasuite is executed using the resulting configuration files. We recommend to run the wrapper script as a batch script and submit it via sbatch as presented in the examples in the sections 7.1.3 and 7.1.3. If you use and adapt one of the wrapper scripts, the output of multiple variables or metrics will be stored in a common output directory, which you need to specify in the script.

7.1.3 Examples

We here describe how to run three short example evaluations with HZG Evaluation Suite Version 0.92.2. Two evaluations are based on one variable only. These evaluations are identical and only differ in the way, in which evasuite is configured and executed. For one evaluation (section 7.1.3) the configuration is done directly via the configuration namelists as described in section 7.1.1. For the other evaluation (section 7.1.3) the configuration is done via a wrapper script as described in section 7.1.2. The third evaluation is based on two variables and multiple metrics (section 7.1.3). This evaluation also is based on a wrapper script. In the next section, we provide a description, how to obtain the scripts and data to run the example evaluations.

Obtain scripts and data for example evaluations

Download the example scripts and data from
`https://swift.dkrz.de/v1/dkrz_c0725fe8741c474b97f291aac57f268f/clmcom/spice/evasuite_examples.tar`. The content has a total size of 1.2 Gb, so we recommend to download it to your work directory (assuming a similar architecture as on the DKRZ cluster Levante).

Within the copied directory you will find the two subdirectories

- example_scripts
- example_data

.

The subdirectory example_scripts contains a subdirectory for each example evaluation, that are described in further detail in the following sections

1. single evaluation via direct modification of configuration files:
   EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal_direct (section 7.1.3)
2. single evaluation via wrapper script:
   EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal (similar evaluation as 1.) (section 7.1.3)
3. multiple evaluations via wrapper script:
   EXAMPLE_tas_pr_CCLM_vs_Eobs_BIAStemporal_norm (section 7.1.3)

The subdirectory example_data contains daily model and reference data for the year 2014 as well as a geopackage file with subregion shapes (EvaSuite_subregion_shapes.gpkg). The model data is based on simulations with the regional climate model COSMO on the CORDEX-domain EUR-11 driven by the global climate models EC-Earth3-Veg and MIROC6. The reference data was extracted from EObs v25 with a grid spacing of 0.1 degree. For all three datasets daily-mean surface temperature is provided. For the two simulations precipitation is provided as a daily-mean flux. For EObs, precipitation is provided as the daily-accumulated precipitation.

The COSMO simulations were conducted by Hendrik Feldmann, KIT.

General preparations to run the examples

Install the stand-alone version of EvaSuite (section 5.2). Copy the files
EXAMPLE_modeldata_settings.yaml, EXAMPLE_reference_datasets.yaml, and
VISUALIZATION_namelist.yaml from "example_scripts" to the subdirectory "inputs" in your EvaSuite code directory (your directory HZG_Evaluation_Suite-v0.92.1beta you untared while installing EvaSuite). In the copied configuration files
EXAMPLE_modeldata_settings.yaml and EXAMPLE_reference_datasets.yaml replace all instances of YOURPATH_TO_DATA_FILES with the path to the downloaded example data. Copy the three subdirectories from "example_scripts" to a subdirectory in your home, e.g.
$HOME/run_HZG_Evaluation_Suite.

Run example evaluation with one variable: EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal_direct

In this example we calculate the bias of the daily-mean near-surface air temperature (tas) at each grid point over the entire period and additionally provide the spatial average for specified prudence subregions and Germany.

Before submitting the evaluation via sbatch you need to set the correct paths in the following files.

- GENERAL_EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal.yaml:
  replace YOUR_WORK_DIR and YOUR_BACKUP_DIR and create the corresponding directories (look up the meaning of eval_workdir and eval_backup_dir in section
  e.g. YOUR_WORK_DIR: /scratch/b/YOUR_DKRZ_ACCOUNT/evasuite_out/EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal
  e.g. YOUR_BACKUP_DIR: /work/YOUR_PROJECT_ACCOUNT/evasuite_backup/EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal
  replace YOURPATH_TO_DATA_FILES for accessing the subregion masks via polyfile

- sbatch_EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal.sh
  replace PROJECT_ACCOUNT
  replace YOURPATH_TO_EVASUITE_CODE by the path to the EvaSuite code directory (your directory HZG_Evaluation_Suite-v0.92.1beta you untared while installing EvaSuite)., e.g. /home-/b/YOUR_DKRZ_ACCOUNT/HZG_Evaluation_Suite-v0.92.1beta

Afterwards, you can submit a job via `sbatch sbatch_EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal.sh` . The job is expected to finish within 1 to 2 minutes.

Once the job finished you will find several logfiles in the directory, in which you submitted the batch script.

- log_EvaSuite.txt: internal log file written by evasuite with information about the evaluation process. Finishes with "BYE" if the evaluation ran successfully.

- sbatch_EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal.log: external log file of the batch job with console output provided by evasuite. This will e.g. contain python errors that occurred while executing evasuite. Finishes with "BYE" if the evaluation ran successfully.

- sbatch_EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal.oXXXXXXXX: output file of the batch job. States the amount of resources used for the evaluation.

- sbatch_EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal.eXXXXXXXX: error file of the batch job. This will contain errors that occurred during the call of EvaSuite. This file is empty in case of successful execuation of the evaluation.

You will find the final results in the working and backup directory. In the working directory the results are located in the subdirectory EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal/Visualization. Within this subdirectory you will find

- figures in pdf format in EvalRAW_plot2D-x-y

- data in netcdf format in EvalRAW_data0d (In this example, each netcdf file contains only one value, which is the time-mean domain-mean bias for the specific subregion.).

Furthermore, you will find intermediate results of the evaluation process in the subdirectory DataMining. In the backup directory you will also find the two subdirectories EvalRAW_plot2D-x-y and EvalRAW_data0d.

EXAMPLE_tas_CCLM_vs_Eobs_BIAStemporal

In this example we perform the same evaluation as in the first example, i.e. we calculate the bias of the daily-mean near-surface air temperature (tas) at each grid point over the entire period and additionally provide the spatial average for specified prudence subregions and Germany. The idea is to showcase a different approach of configuring and running evasuite than the direct configuration via the namelists as presented in the previous example. Here, we run evasuite via a wrapper script as described in section 7.1.2).

Before submitting the wrapper script via sbatch you need to do the following adjustments in EvaSuite_wrapper_script.sh:

- replace PROJECT_ACCOUNT
- replace YOURPATH_TO_EVASUITE_INSTALLATION by the path into which you installed evasuite, e.g. /home/b/YOUR_DKRZ_ACCOUNT/.local/bin; you can figure out this path via "which evasuite"
- replace YOURPATH_TO_EVASUITE_CODE; see the first example
- replace YOUR_WORK_DIR; see the first example
- replace YOUR_BACKUP_DIR; see the first example
- replace YOURPATH_TO_DATA_FILES; see the first example

Create the corresponding directories YOUR_WORK_DIR and YOUR_BACKUP_DIR. Afterwards, you can submit a job via `sbatch EvaSuite_wrapper_script.sh`. The job is expected to finish within 1 to 2 minutes.

While the batch job is running, information will be written to logfile called slurm-XXXXXXXX.out. Within this file, you will find information about the evaluation performed and where to find the corresponding evasuite logfiles. This logfile also contains information about the resource consumption of the evaluation performed.

Finally, you will find the results of all evaluations in the working directory and the backup directory specified above. The results are (should be!) identical to the ones in the first example.

EXAMPLE_tas_pr_CCLM_vs_Eobs_BIAStemporal_norm

In this example we calculate the bias and normalized bias of the daily-mean near-surface air temperature (tas) and the daily precipitation sum (pr) at each grid point over the entire period and additionally provide the spatial average for specified the prudence subregion P_Alps and Germany. Furthermore, we calculate absolute values for the reference and the model datasets as monthly mean values and also spatially average the over the same subregions. To initiate all these evaluations at once, we use a wrapper script.

Before submitting the wrapper script via sbatch you need to apply the same adjustments in EvaSuite_wrapper_script.sh as in the previous example:

- replace PROJECT_ACCOUNT
- replace YOURPATH_TO_EVASUITE_INSTALLATION by the path into which you installed evasuite, e.g. /home/b/YOUR_DKRZ_ACCOUNT/.local/bin; you can figure out this path via "which evasuite"
- replace YOURPATH_TO_EVASUITE_CODE; see the first example
- replace YOUR_WORK_DIR; see the first example
- replace YOUR_BACKUP_DIR; see the first example
- replace YOURPATH_TO_DATA_FILES; see the first example

Create the corresponding directories YOUR_WORK_DIR and YOUR_BACKUP_DIR. Afterwards, you can submit a job via `sbatch EvaSuite_wrapper_script.sh`. The job is expected to finish within 5 minutes.

While the batch job is running, information will be written to logfile called slurm-XXXXXXXX.out. Within this file, you will find information about each evaluation performed and where to find the corresponding evasuite logfiles. This logfile also contains information about the resource consumption of all the evaluations performed.

Finally, you will find the results of all evaluations in the backup directory specified above. In this example, you will find

- figures in pdf format in CompareRAW_plot1D (line plots based on absolute values), CompareRAW_plot2D-x-y (maps based on absolute values), and EvalRAW_plot2D-x-y (maps based on the absolute and relative bias)
- data in netcdf format in CompareRAW_data1D (time series of absolute values averaged over each subregion) and in EvalRAW_data0d (In this example, each netcdf file contains again only one value, which is the time-mean domain-mean bias for the specific subregion.)

In this example, intermediate results are only available in the specified working directory for the last evaluation performed, since the same working directory is used for each evaluation.

Chapter 8

Supplementary tools

8.1 Adding pre-defined and custom subregion shapes

EvaSuite allows for calculation of different statistics in different subregions. The shape of this subregions is treated by the Evasuite using so-called geopackage files which is an extension of the former known shape files. If you want to execute the Evasuite, you need to specify a geopackage file and the ID of the attribute used to identify the names of the shapes. Both settings need to be configured in the general configuration file.

We provide a geopackage file with an long list of predefined subregions, which can easily be used with EvaSuite. You can download the geopackage file here:
`https://swift.dkrz.de/v1/dkrz_c0725fe8741c474b97f291aac57f268f/clmcom/spice/EvaSuite_allpolygones.v1.gpkg`. The list of the available subregions is given in C.

We here describe two options of checking the content of existing geopackage files and adding further subregions. The first option is command-line based and suitable for adding already existing polygons to the geopackage file. This option is ready to be used after installing EvaSuite. The second option is based on a jupyter notebook, which allows the visualization of subregions. Based on the second option, we also demonstrate how to add custom subregions with simple geometries. The jupyter notebook can be obtained from the EvaSuite git repository as described in the corresponding subsection.

8.1.1 Command-line based geopackage exploration and modification

So, starting with a geopackage-file one has e.g. received with the installation of SPICE, type

`evasuite --showpolygons ${geopackage-file}`

to receive a list of the content. Please investigate the header of the table shown and identify the attribute name for the region names (e.g. 'NAME_0').

To get the whole list of regions stored in the geopackage file type

`evasuite --showpolygons ${geopackage-file} NAME_0`.

This are regions which can be handled by the EvaSuite. If you want to handle more regions, you need to extend the geopackage file. There are two possibilities.

Add further subregions using a geopackage file

This can be done using the following syntax

`evasuite --addpolygondb ${newgeopackage} ${newattrib} ${newregionnames} ${oldgeopackage}${oldattrib} ${result}`,

where the old geopackage file with regions stored under the attributes ${oldattrib} is extended by the content of the new geopackage file. Regarding the new geopackage file one has to specify the attribute ${newattrib} and the values of the attribute ${newregionnames} (the names of the regions stored in the new geopackage file). The final file is stored in ${result} and the final attribute is the same as the old attribute (${oldattrib}).

Add further subregions from an ascii file

The ascii file needs to be in the so-called wkt-format. This can be done using the following syntax

`evasuite --addpolygonascii ${newwktfile} ${newattrib} ${oldgeopackage} ${oldattrib} ${result}`.

It is the nearly the same procedure as for case A), but the final file is extended by ALL regions stored in the wkt file (with names given by the attribute $newattrib).

8.1.2 Jupyter notebook-based geopackage exploration and modification

A pdf copy of the jupyter notebook is provided in chapter D. Descriptions are included within the notebook itself.

You can find and download the jupyter notebook from `https://gitlab.dkrz.de/clm-community/evasuite/HZG_Evaluation_Suite/-/tree/add_custom_subregion_creation/scripts/create_custom_subregion?ref_type=heads`.

Please see the DKRZ resources on the usage of jupyterhub and jupyter notebooks here: `https://docs.dkrz.de/doc/software%26services/jupyterhub/index.html`.

Appendix A

Code examples for the general configuration file

A.1 General configuration, dictionary Visual_program

```
Visual_program          = {
     "Rstat"       : { "default-version" : True, "version": "3.2.0", "binary-
        location" : "/sw/rhel6-x64/r/r-3.2.0-gcc48/bin/R" }
}
```

A.2 General configuration, dictionary Masking_program

```
Masking_program     = {
    "HZG-masker" : { "default-version" : False, "version": "2.0", "parent-
       prog" : "Rstat",
    "binary-location" : "/mnt/lustre01/pf/zmaw/m221054/HZG-WORK/
       project_OpenFRED/EvaSuite/HZG_Evaluation_Suite/Addons/MaskCreate/
       polymask_v2.Rs",
    "RScriptname" : "polymask"}
}
```

A.3 General configuration, dictionary DataDumpProcess_programs

```
DataDumpProcess_programs  = {
  "cdo"         : { "default-version" : True, "version": "1.8.0", "binary-
     location" : "/sw/rhel6-x64/cdo/cdo-1.8.0-magicsxx-gcc48/bin/cdo" , "
     ompthreads" : 2},
  "nco"         : { "default-version" : True, "version": "1.0.0", "binary-
     location" : "" , "ompthreads" : 2},
  "ncdump"      : { "default-version" : True, "version": "1.0.0", "binary-
     location" : "" },
  "pyxarray"    : { "default-version" : True, "version": "1.0.0", "binary-
     location" : "" },
}
```

A.4 General configuration, dictionary datamining, subdictionary target_EvalFrequency

```
"target_EvalFrequency" : {"method" : "resample", "freq" : "daily", "how" : "
   mean"}

"target_EvalFrequency" : {"method" : "reselect", "targetgrid" : "PAULAT"}
```

```
"target_EvalFrequency" : {"-"}
```

A.5 General configuration, dictionary datamining, subdictionary subregions_BoxCorner_files

```
"subregions_BoxCorner_files" : {
        "directory" : "/pf/m/m221054/HZG-WORK/project_OpenFRED/EvaSuite/
           Subregion_BoxCorners/BeateGeyer_Lowlands/",
        "filepattern" : "poly*.dat"},
```

Appendix B

Tables - overview about possible options

B.1 General configuration, dictionary datamining, supported resample frequencies

keyword	Description
'yearly'	yearly aggregation
'seasonal'	seasonal aggregation (starting in December)
'monthly'	monthly aggregation
'weekly'	weekly aggregation (starting on Mondays)
'daily'	daily aggregation
'hourly'	hourly aggregation

B.2 General configuration, dictionary datamining, supported resample methods

keyword	Description
'mean'	calculating an aggregated mean (e.g. daily mean)
'sum'	calculating an aggregated sum (Bug 1: NaN values are not considered, i.e. a daily sum is '1' if 23 hourly values are NaN and one value is '1'; Bug 2: sum over NaN-values is 0)
'min'	calculating a minimum during aggregation (e.g. daily minimum)
'max'	calculating a maximum during aggregation (e.g. daily maximum)

B.3 General configuration, dictionary datamining, supported temporal aggregation methods

keyword	Description
'fullperiod'	aggregation in total
'yearly'	yearly aggregation
'seasonal'	seasonal aggregation (starting in December)
'monthly'	monthly aggregation
'weekly'	weekly aggregation (starting on Mondays)

'daily'	daily aggregation
'hourly'	hourly aggregation
'monthlyoverperiod'	grouping all data to single months
'weeklyoverperiod'	grouping all data to single business weeks
'dailyoverperiod'	grouping all data to single days of the year
'hourlyoverperiod'	grouping all data to single hours of the day

B.4 General configuration, dataset settings, Correction of Parameter

```
"paramcorr"   : {"T_2M" :
{ "unitcor"    : True, "corr_vals"   : ["+","273.15","K"],
  "timeshift" : True, "timcor_val" : ["-","6","h"]
}}
```

Appendix C

Overview of available subregions in the geopackage file

We provide a geopackage file with an long list of predefined subregions, which can easily be used with EvaSuite. You can download the geopackage file here: `https://swift.dkrz.de/v1/dkrz_c0725fe8741c474b97f291aac57f268f/clmcom/spice/EvaSuite_allpolygones.v1.gpkg`. The list of the available subregions is given in C.

No	NAME_0	category
1	S_Mediterranean_Sea	Sea
2	S_Black_Sea	Sea
3	S_Baltic_Sea	Sea
4	S_North_Sea	Sea
5	S_Marmara_Sea	Sea
6	S_North_Sea_extended	Sea
7	S_North_Atlantic_Ocean	Sea
8	S_South_Atlantic_Ocean	Sea
9	P_Alps	Prudence box
10	P_BritishIslands	Prudence box
11	P_East-Europe	Prudence box
12	P_France	Prudence box
13	P_IberianPeninsula	Prudence box
14	P_Mediterranean	Prudence box
15	P_Mid-Europe	Prudence box
16	P_Scandinavia	Prudence box
17	P_North-West-Africa	Prudence box
18	Albania	country
19	Austria	country
20	Belgium	country
21	Bulgaria	country
22	Bosnia and Herzegovina	country
23	Belarus	country
24	Switzerland	country
25	Cyprus	country
26	Czech Republic	country
27	Germany	country
28	Denmark	country
29	Spain	country
30	Estonia	country
31	Finland	country
32	France	country
33	United Kingdom	country
34	Georgia	country

35	Greece	country
36	Croatia	country
37	Hungary	country
38	Ireland	country
39	Iceland	country
40	Italy	country
41	Kosovo	country
42	Lithuania	country
43	Luxembourg	country
44	Latvia	country
45	Moldova	country
46	Macedonia	country
47	Montenegro	country
48	Netherlands	country
49	Norway	country
50	Poland	country
51	Portugal	country
52	Romania	country
53	Russia	country
54	Serbia	country
55	Slovakia	country
56	Slovenia	country
57	Sweden	country
58	Turkey	country
59	Ukraine	country
60	AustriaCzech	summarized countries
61	Baltics	summarized countries
62	Benelux	summarized countries
63	HungarySlovakia	summarized countries
64	MiddleBalkan	summarized countries
65	NorthWesternBalkan	summarized countries
66	SouthernBalkan	summarized countries
67	EZG_EU_Rhein	catchment
68	EZG_EU_Rhein_N	catchment
69	EZG_EU_Rhein_S	catchment
70	EZG_G_Donau	catchment German part
71	EZG_G_Elbe	catchment German part
72	EZG_G_Ems	catchment German part
73	EZG_G_Havel-Spree	catchment German part
74	EZG_G_Issel	catchment German part
75	EZG_G_Maas	catchment German part
76	EZG_G_Oder	catchment German part
77	EZG_G_Rhein	catchment German part
78	EZG_G_Schwarze-Elster	catchment German part
79	EZG_G_Spree	catchment German part
80	EZG_G_Weser	catchment German part

No	NAME_0	category
81	Stuttgart, Stadtkreis	Landkreis Deutschland
82	Böblingen	Landkreis Deutschland
83	Esslingen	Landkreis Deutschland
84	Göppingen	Landkreis Deutschland
85	Ludwigsburg	Landkreis Deutschland
86	Rems-Murr-Kreis	Landkreis Deutschland
87	Heilbronn, Stadtkreis	Landkreis Deutschland

88	Heilbronn, Landkreis	Landkreis Deutschland
89	Hohenlohekreis	Landkreis Deutschland
90	Schwäbisch Hall	Landkreis Deutschland
91	Main-Tauber-Kreis	Landkreis Deutschland
92	Heidenheim	Landkreis Deutschland
93	Ostalbkreis	Landkreis Deutschland
94	Baden-Baden, Stadtkreis	Landkreis Deutschland
95	Karlsruhe, Stadtkreis	Landkreis Deutschland
96	Karlsruhe, Landkreis	Landkreis Deutschland
97	Rastatt	Landkreis Deutschland
98	Heidelberg, Stadtkreis	Landkreis Deutschland
99	Mannheim, Stadtkreis	Landkreis Deutschland
100	Neckar-Odenwald-Kreis	Landkreis Deutschland
101	Rhein-Neckar-Kreis	Landkreis Deutschland
102	Pforzheim, Stadtkreis	Landkreis Deutschland
103	Calw	Landkreis Deutschland
104	Enzkreis	Landkreis Deutschland
105	Freudenstadt	Landkreis Deutschland
106	Freiburg im Breisgau, Stadtkreis	Landkreis Deutschland
107	Breisgau-Hochschwarzwald	Landkreis Deutschland
108	Emmendingen	Landkreis Deutschland
109	Ortenaukreis	Landkreis Deutschland
110	Rottweil	Landkreis Deutschland
111	Schwarzwald-Baar-Kreis	Landkreis Deutschland
112	Tuttlingen	Landkreis Deutschland
113	Konstanz	Landkreis Deutschland
114	Lörrach	Landkreis Deutschland
115	Waldshut	Landkreis Deutschland
116	Reutlingen	Landkreis Deutschland
117	Tübingen, Landkreis	Landkreis Deutschland
118	Zollernalbkreis	Landkreis Deutschland
119	Ulm, Stadtkreis	Landkreis Deutschland
120	Alb-Donau-Kreis	Landkreis Deutschland
121	Biberach	Landkreis Deutschland
122	Bodenseekreis	Landkreis Deutschland
123	Ravensburg	Landkreis Deutschland
124	Sigmaringen	Landkreis Deutschland
125	Ingolstadt, Kreisfreie Stadt	Landkreis Deutschland
126	München, Kreisfreie Stadt	Landkreis Deutschland
127	Rosenheim, Kreisfreie Stadt	Landkreis Deutschland
128	Altötting	Landkreis Deutschland
129	Berchtesgadener Land	Landkreis Deutschland
130	Bad Tölz-Wolfratshausen	Landkreis Deutschland
131	Dachau	Landkreis Deutschland
132	Ebersberg	Landkreis Deutschland
133	Eichstätt	Landkreis Deutschland
134	Erding	Landkreis Deutschland
135	Freising	Landkreis Deutschland
136	Fürstenfeldbruck	Landkreis Deutschland
137	Garmisch-Partenkirchen	Landkreis Deutschland
138	Landsberg am Lech	Landkreis Deutschland
139	Miesbach	Landkreis Deutschland
140	Mühldorf a. Inn	Landkreis Deutschland
141	München, Landkreis	Landkreis Deutschland
142	Neuburg-Schrobenhausen	Landkreis Deutschland
143	Pfaffenhofen a.d. Ilm	Landkreis Deutschland

144	Rosenheim, Landkreis	Landkreis Deutschland
145	Starnberg	Landkreis Deutschland
146	Traunstein	Landkreis Deutschland
147	Weilheim-Schongau	Landkreis Deutschland
148	Landshut, Kreisfreie Stadt	Landkreis Deutschland
149	Passau, Kreisfreie Stadt	Landkreis Deutschland
150	Straubing, Kreisfreie Stadt	Landkreis Deutschland
151	Deggendorf	Landkreis Deutschland
152	Freyung-Grafenau	Landkreis Deutschland
153	Kelheim	Landkreis Deutschland
154	Landshut, Landkreis	Landkreis Deutschland
155	Passau, Landkreis	Landkreis Deutschland
156	Regen	Landkreis Deutschland
157	Rottal-Inn	Landkreis Deutschland
158	Straubing-Bogen	Landkreis Deutschland
159	Dingolfing-Landau	Landkreis Deutschland
160	Amberg, Kreisfreie Stadt	Landkreis Deutschland
161	Regensburg, Kreisfreie Stadt	Landkreis Deutschland
162	Weiden i.d. Opf, Kreisfreie Stadt	Landkreis Deutschland
163	Amberg-Sulzbach	Landkreis Deutschland
164	Cham	Landkreis Deutschland
165	Neumarkt i.d. OPf.	Landkreis Deutschland
166	Neustadt a.d. Waldnaab	Landkreis Deutschland
167	Regensburg, Landkreis	Landkreis Deutschland
168	Schwandorf	Landkreis Deutschland
169	Tirschenreuth	Landkreis Deutschland
170	Bamberg, Kreisfreie Stadt	Landkreis Deutschland
171	Bayreuth, Kreisfreie Stadt	Landkreis Deutschland
172	Coburg, Kreisfreie Stadt	Landkreis Deutschland
173	Hof, Kreisfreie Stadt	Landkreis Deutschland
174	Bamberg, Landkreis	Landkreis Deutschland
175	Bayreuth, Landkreis	Landkreis Deutschland
176	Coburg, Landkreis	Landkreis Deutschland
177	Forchheim	Landkreis Deutschland
178	Hof, Landkreis	Landkreis Deutschland
179	Kronach	Landkreis Deutschland
180	Kulmbach	Landkreis Deutschland
181	Lichtenfels	Landkreis Deutschland
182	Wunsiedel i. Fichtelgebirge	Landkreis Deutschland
183	Ansbach, Kreisfreie Stadt	Landkreis Deutschland
184	Erlangen, Kreisfreie Stadt	Landkreis Deutschland
185	Fürth, Kreisfreie Stadt	Landkreis Deutschland
186	Nürnberg, Kreisfreie Stadt	Landkreis Deutschland
187	Schwabach, Kreisfreie Stadt	Landkreis Deutschland
188	Ansbach, Landkreis	Landkreis Deutschland
189	Erlangen-Höchstadt	Landkreis Deutschland
190	Fürth, Landkreis	Landkreis Deutschland
191	Nürnberger Land	Landkreis Deutschland
192	Neustadt a.d. Aisch-Bad Windsheim	Landkreis Deutschland
193	Roth	Landkreis Deutschland
194	Weißenburg-Gunzenhausen	Landkreis Deutschland
195	Aschaffenburg, Kreisfreie Stadt	Landkreis Deutschland
196	Schweinfurt, Kreisfreie Stadt	Landkreis Deutschland
197	Würzburg, Kreisfreie Stadt	Landkreis Deutschland
198	Aschaffenburg, Landkreis	Landkreis Deutschland
199	Bad Kissingen	Landkreis Deutschland

200	Rhön-Grabfeld	Landkreis Deutschland
201	Haßberge	Landkreis Deutschland
202	Kitzingen	Landkreis Deutschland
203	Miltenberg	Landkreis Deutschland
204	Main-Spessart	Landkreis Deutschland
205	Schweinfurt, Landkreis	Landkreis Deutschland
206	Würzburg, Landkreis	Landkreis Deutschland
207	Augsburg, Kreisfreie Stadt	Landkreis Deutschland
208	Kaufbeuren, Kreisfreie Stadt	Landkreis Deutschland
209	Kempten (Allgäu), Kreisfreie Stadt	Landkreis Deutschland
210	Memmingen, Kreisfreie Stadt	Landkreis Deutschland
211	Aichach-Friedberg	Landkreis Deutschland
212	Augsburg, Landkreis	Landkreis Deutschland
213	Dillingen a.d. Donau	Landkreis Deutschland
214	Günzburg	Landkreis Deutschland
215	Neu-Ulm	Landkreis Deutschland
216	Lindau (Bodensee)	Landkreis Deutschland
217	Ostallgäu	Landkreis Deutschland
218	Unterallgäu	Landkreis Deutschland
219	Donau-Ries	Landkreis Deutschland
220	Oberallgäu	Landkreis Deutschland
221	Berlin	Landkreis Deutschland
222	Brandenburg an der Havel, Kreisfreie Stadt	Landkreis Deutschland
223	Cottbus, Kreisfreie Stadt	Landkreis Deutschland
224	Frankfurt (Oder), Kreisfreie Stadt	Landkreis Deutschland
225	Potsdam, Kreisfreie Stadt	Landkreis Deutschland
226	Barnim	Landkreis Deutschland
227	Dahme-Spreewald	Landkreis Deutschland
228	Elbe-Elster	Landkreis Deutschland
229	Havelland	Landkreis Deutschland
230	Märkisch-Oderland	Landkreis Deutschland
231	Oberhavel	Landkreis Deutschland
232	Oberspreewald-Lausitz	Landkreis Deutschland
233	Oder-Spree	Landkreis Deutschland
234	Ostprignitz-Ruppin	Landkreis Deutschland
235	Potsdam-Mittelmark	Landkreis Deutschland
236	Prignitz	Landkreis Deutschland
237	Spree-Neiße	Landkreis Deutschland
238	Teltow-Fläming	Landkreis Deutschland
239	Uckermark	Landkreis Deutschland
240	Bremen, Kreisfreie Stadt	Landkreis Deutschland
241	Bremerhaven, Kreisfreie Stadt	Landkreis Deutschland
242	Hamburg	Landkreis Deutschland
243	Darmstadt, Kreisfreie Stadt	Landkreis Deutschland
244	Frankfurt am Main, Kreisfreie Stadt	Landkreis Deutschland
245	Offenbach am Main, Kreisfreie Stadt	Landkreis Deutschland
246	Wiesbaden, Kreisfreie Stadt	Landkreis Deutschland
247	Bergstraße	Landkreis Deutschland
248	Darmstadt-Dieburg	Landkreis Deutschland
249	Groß-Gerau	Landkreis Deutschland
250	Hochtaunuskreis	Landkreis Deutschland
251	Main-Kinzig-Kreis	Landkreis Deutschland
252	Main-Taunus-Kreis	Landkreis Deutschland
253	Odenwaldkreis	Landkreis Deutschland
254	Offenbach, Landkreis	Landkreis Deutschland
255	Rheingau-Taunus-Kreis	Landkreis Deutschland

256	Wetteraukreis	Landkreis Deutschland
257	Gießen, Landkreis	Landkreis Deutschland
258	Lahn-Dill-Kreis	Landkreis Deutschland
259	Limburg-Weilburg	Landkreis Deutschland
260	Marburg-Biedenkopf	Landkreis Deutschland
261	Vogelsbergkreis	Landkreis Deutschland
262	Kassel, Kreisfreie Stadt	Landkreis Deutschland
263	Fulda	Landkreis Deutschland
264	Hersfeld-Rotenburg	Landkreis Deutschland
265	Kassel, Landkreis	Landkreis Deutschland
266	Schwalm-Eder-Kreis	Landkreis Deutschland
267	Waldeck-Frankenberg	Landkreis Deutschland
268	Werra-Meißner-Kreis	Landkreis Deutschland
269	Rostock, Kreisfreie Stadt	Landkreis Deutschland
270	Schwerin, Kreisfreie Stadt	Landkreis Deutschland
271	Mecklenburgische Seenplatte	Landkreis Deutschland
272	Landkreis Rostock	Landkreis Deutschland
273	Vorpommern-Rügen	Landkreis Deutschland
274	Nordwestmecklenburg	Landkreis Deutschland
275	Vorpommern-Greifswald	Landkreis Deutschland
276	Ludwigslust-Parchim	Landkreis Deutschland
277	Braunschweig, Kreisfreie Stadt	Landkreis Deutschland
278	Salzgitter, Kreisfreie Stadt	Landkreis Deutschland
279	Wolfsburg, Kreisfreie Stadt	Landkreis Deutschland
280	Gifhorn	Landkreis Deutschland
281	Goslar	Landkreis Deutschland
282	Helmstedt	Landkreis Deutschland
283	Northeim	Landkreis Deutschland
284	Peine	Landkreis Deutschland
285	Wolfenbüttel	Landkreis Deutschland
286	Göttingen	Landkreis Deutschland
287	Diepholz	Landkreis Deutschland
288	Hameln-Pyrmont	Landkreis Deutschland
289	Hildesheim	Landkreis Deutschland
290	Holzminden	Landkreis Deutschland
291	Nienburg (Weser)	Landkreis Deutschland
292	Schaumburg	Landkreis Deutschland
293	Region Hannover	Landkreis Deutschland
294	Celle	Landkreis Deutschland
295	Cuxhaven	Landkreis Deutschland
296	Harburg	Landkreis Deutschland
297	Lüchow-Dannenberg	Landkreis Deutschland
298	Lüneburg, Landkreis	Landkreis Deutschland
299	Osterholz	Landkreis Deutschland
300	Rotenburg (Wümme)	Landkreis Deutschland
301	Heidekreis	Landkreis Deutschland
302	Stade	Landkreis Deutschland
303	Uelzen	Landkreis Deutschland
304	Verden	Landkreis Deutschland
305	Delmenhorst, Kreisfreie Stadt	Landkreis Deutschland
306	Emden, Kreisfreie Stadt	Landkreis Deutschland
307	Oldenburg (Oldenburg), Kreisfreie Stadt	Landkreis Deutschland
308	Osnabrück, Kreisfreie Stadt	Landkreis Deutschland
309	Wilhelmshaven, Kreisfreie Stadt	Landkreis Deutschland
310	Ammerland	Landkreis Deutschland
311	Aurich	Landkreis Deutschland

312	Cloppenburg	Landkreis Deutschland
313	Emsland	Landkreis Deutschland
314	Friesland (DE)	Landkreis Deutschland
315	Grafschaft Bentheim	Landkreis Deutschland
316	Leer	Landkreis Deutschland
317	Oldenburg, Landkreis	Landkreis Deutschland
318	Osnabrück, Landkreis	Landkreis Deutschland
319	Vechta	Landkreis Deutschland
320	Wesermarsch	Landkreis Deutschland
321	Wittmund	Landkreis Deutschland
322	Düsseldorf, Kreisfreie Stadt	Landkreis Deutschland
323	Duisburg, Kreisfreie Stadt	Landkreis Deutschland
324	Essen, Kreisfreie Stadt	Landkreis Deutschland
325	Krefeld, Kreisfreie Stadt	Landkreis Deutschland
326	Mönchengladbach, Kreisfreie Stadt	Landkreis Deutschland
327	Mülheim an der Ruhr, Kreisfreie Stadt	Landkreis Deutschland
328	Oberhausen, Kreisfreie Stadt	Landkreis Deutschland
329	Remscheid, Kreisfreie Stadt	Landkreis Deutschland
330	Solingen, Kreisfreie Stadt	Landkreis Deutschland
331	Wuppertal, Kreisfreie Stadt	Landkreis Deutschland
332	Kleve	Landkreis Deutschland
333	Mettmann	Landkreis Deutschland
334	Rhein-Kreis Neuss	Landkreis Deutschland
335	Viersen	Landkreis Deutschland
336	Wesel	Landkreis Deutschland
337	Bonn, Kreisfreie Stadt	Landkreis Deutschland
338	Köln, Kreisfreie Stadt	Landkreis Deutschland
339	Leverkusen, Kreisfreie Stadt	Landkreis Deutschland
340	Düren	Landkreis Deutschland
341	Rhein-Erft-Kreis	Landkreis Deutschland
342	Euskirchen	Landkreis Deutschland
343	Heinsberg	Landkreis Deutschland
344	Oberbergischer Kreis	Landkreis Deutschland
345	Rheinisch-Bergischer-Kreis	Landkreis Deutschland
346	Rhein-Sieg-Kreis	Landkreis Deutschland
347	Städteregion Aachen	Landkreis Deutschland
348	Bottrop, Kreisfreie Stadt	Landkreis Deutschland
349	Gelsenkirchen, Kreisfreie Stadt	Landkreis Deutschland
350	Münster, Kreisfreie Stadt	Landkreis Deutschland
351	Borken	Landkreis Deutschland
352	Coesfeld	Landkreis Deutschland
353	Recklinghausen	Landkreis Deutschland
354	Steinfurt	Landkreis Deutschland
355	Warendorf	Landkreis Deutschland
356	Bielefeld, Kreisfreie Stadt	Landkreis Deutschland
357	Gütersloh	Landkreis Deutschland
358	Herford	Landkreis Deutschland
359	Höxter	Landkreis Deutschland
360	Lippe	Landkreis Deutschland
361	Minden-Lübbecke	Landkreis Deutschland
362	Paderborn	Landkreis Deutschland
363	Bochum, Kreisfreie Stadt	Landkreis Deutschland
364	Dortmund, Kreisfreie Stadt	Landkreis Deutschland
365	Hagen, Kreisfreie Stadt	Landkreis Deutschland
366	Hamm, Kreisfreie Stadt	Landkreis Deutschland
367	Herne, Kreisfreie Stadt	Landkreis Deutschland

368	Ennepe-Ruhr-Kreis	Landkreis Deutschland
369	Hochsauerlandkreis	Landkreis Deutschland
370	Märkischer Kreis	Landkreis Deutschland
371	Olpe	Landkreis Deutschland
372	Siegen-Wittgenstein	Landkreis Deutschland
373	Soest	Landkreis Deutschland
374	Unna	Landkreis Deutschland
375	Koblenz, Kreisfreie Stadt	Landkreis Deutschland
376	Ahrweiler	Landkreis Deutschland
377	Altenkirchen (Westerwald)	Landkreis Deutschland
378	Bad Kreuznach	Landkreis Deutschland
379	Birkenfeld	Landkreis Deutschland
380	Mayen-Koblenz	Landkreis Deutschland
381	Neuwied	Landkreis Deutschland
382	Rhein-Lahn-Kreis	Landkreis Deutschland
383	Westerwaldkreis	Landkreis Deutschland
384	Cochem-Zell	Landkreis Deutschland
385	Rhein-Hunsrück-Kreis	Landkreis Deutschland
386	Trier, Kreisfreie Stadt	Landkreis Deutschland
387	Bernkastel-Wittlich	Landkreis Deutschland
388	Eifelkreis Bitburg-Prüm	Landkreis Deutschland
389	Vulkaneifel	Landkreis Deutschland
390	Trier-Saarburg	Landkreis Deutschland
391	Frankenthal (Pfalz), Kreisfreie Stadt	Landkreis Deutschland
392	Kaiserslautern, Kreisfreie Stadt	Landkreis Deutschland
393	Landau in der Pfalz, Kreisfreie Stadt	Landkreis Deutschland
394	Ludwigshafen am Rhein, Kreisfreie Stadt	Landkreis Deutschland
395	Mainz, Kreisfreie Stadt	Landkreis Deutschland
396	Neustadt an der Weinstraße, Kreisfreie Stadt	Landkreis Deutschland
397	Pirmasens, Kreisfreie Stadt	Landkreis Deutschland
398	Speyer, Kreisfreie Stadt	Landkreis Deutschland
399	Worms, Kreisfreie Stadt	Landkreis Deutschland
400	Zweibrücken, Kreisfreie Stadt	Landkreis Deutschland
401	Alzey-Worms	Landkreis Deutschland
402	Bad Dürkheim	Landkreis Deutschland
403	Donnersbergkreis	Landkreis Deutschland
404	Germersheim	Landkreis Deutschland
405	Kaiserslautern, Landkreis	Landkreis Deutschland
406	Kusel	Landkreis Deutschland
407	Südliche Weinstraße	Landkreis Deutschland
408	Rhein-Pfalz-Kreis	Landkreis Deutschland
409	Mainz-Bingen	Landkreis Deutschland
410	Südwestpfalz	Landkreis Deutschland
411	Regionalverband Saarbrücken	Landkreis Deutschland
412	Merzig-Wadern	Landkreis Deutschland
413	Neunkirchen	Landkreis Deutschland
414	Saarlouis	Landkreis Deutschland
415	Saarpfalz-Kreis	Landkreis Deutschland
416	St. Wendel	Landkreis Deutschland
417	Dresden, Kreisfreie Stadt	Landkreis Deutschland
418	Bautzen	Landkreis Deutschland
419	Görlitz	Landkreis Deutschland
420	Meißen	Landkreis Deutschland
421	Sächsische Schweiz-Osterzgebirge	Landkreis Deutschland
422	Chemnitz, Kreisfreie Stadt	Landkreis Deutschland
423	Erzgebirgskreis	Landkreis Deutschland

424	Mittelsachsen	Landkreis Deutschland
425	Vogtlandkreis	Landkreis Deutschland
426	Zwickau	Landkreis Deutschland
427	Leipzig, Kreisfreie Stadt	Landkreis Deutschland
428	Leipzig	Landkreis Deutschland
429	Nordsachsen	Landkreis Deutschland
430	Dessau-Roßlau, Kreisfreie Stadt	Landkreis Deutschland
431	Halle (Saale), Kreisfreie Stadt	Landkreis Deutschland
432	Magdeburg, Kreisfreie Stadt	Landkreis Deutschland
433	Altmarkkreis Salzwedel	Landkreis Deutschland
434	Anhalt-Bitterfeld	Landkreis Deutschland
435	Jerichower Land	Landkreis Deutschland
436	Börde	Landkreis Deutschland
437	Burgenlandkreis	Landkreis Deutschland
438	Harz	Landkreis Deutschland
439	Mansfeld-Südharz	Landkreis Deutschland
440	Saalekreis	Landkreis Deutschland
441	Salzlandkreis	Landkreis Deutschland
442	Stendal	Landkreis Deutschland
443	Wittenberg	Landkreis Deutschland
444	Flensburg, Kreisfreie Stadt	Landkreis Deutschland
445	Kiel, Kreisfreie Stadt	Landkreis Deutschland
446	Lübeck, Kreisfreie Stadt	Landkreis Deutschland
447	Neumünster, Kreisfreie Stadt	Landkreis Deutschland
448	Dithmarschen	Landkreis Deutschland
449	Herzogtum Lauenburg	Landkreis Deutschland
450	Nordfriesland	Landkreis Deutschland
451	Ostholstein	Landkreis Deutschland
452	Pinneberg	Landkreis Deutschland
453	Plön	Landkreis Deutschland
454	Rendsburg-Eckernförde	Landkreis Deutschland
455	Schleswig-Flensburg	Landkreis Deutschland
456	Segeberg	Landkreis Deutschland
457	Steinburg	Landkreis Deutschland
458	Stormarn	Landkreis Deutschland
459	Erfurt, Kreisfreie Stadt	Landkreis Deutschland
460	Gera, Kreisfreie Stadt	Landkreis Deutschland
461	Jena, Kreisfreie Stadt	Landkreis Deutschland
462	Suhl, Kreisfreie Stadt	Landkreis Deutschland
463	Weimar, Kreisfreie Stadt	Landkreis Deutschland
464	Eichsfeld	Landkreis Deutschland
465	Nordhausen	Landkreis Deutschland
466	Unstrut-Hainich-Kreis	Landkreis Deutschland
467	Kyffhäuserkreis	Landkreis Deutschland
468	Schmalkalden-Meiningen	Landkreis Deutschland
469	Gotha	Landkreis Deutschland
470	Sömmerda	Landkreis Deutschland
471	Hildburghausen	Landkreis Deutschland
472	Ilm-Kreis	Landkreis Deutschland
473	Weimarer Land	Landkreis Deutschland
474	Sonneberg	Landkreis Deutschland
475	Saalfeld-Rudolstadt	Landkreis Deutschland
476	Saale-Holzland-Kreis	Landkreis Deutschland
477	Saale-Orla-Kreis	Landkreis Deutschland
478	Greiz	Landkreis Deutschland
479	Altenburger Land	Landkreis Deutschland

480	Eisenach, Kreisfreie Stadt	Landkreis Deutschland
481	Wartburgkreis	Landkreis Deutschland
482	Germany_Baden-Wuerttemberg	Bundesland Deutschland
483	Germany_Bayern	Bundesland Deutschland
484	Germany_Berlin	Bundesland Deutschland
485	Germany_Brandenburg	Bundesland Deutschland
486	Germany_Bremen	Bundesland Deutschland
487	Germany_Hamburg	Bundesland Deutschland
488	Germany_Hessen	Bundesland Deutschland
489	Germany_Mecklenburg-Vorpommern	Bundesland Deutschland
490	Germany_Niedersachsen	Bundesland Deutschland
491	Germany_Nordrhein-Westfalen	Bundesland Deutschland
492	Germany_Rheinland-Pfalz	Bundesland Deutschland
493	Germany_Saarland	Bundesland Deutschland
494	Germany_Sachsen-Anhalt	Bundesland Deutschland
495	Germany_Sachsen	Bundesland Deutschland
496	Germany_Schleswig-Holstein	Bundesland Deutschland
497	Germany_Thueringen	Bundesland Deutschland
498	Germany_East	Region Deutschland
499	Germany_North	Region Deutschland
500	Germany_West	Region Deutschland
501	Germany_South	Region Deutschland
502	IAWAK-EE	Region NUKLEUS-project
503	KlimaKonform	Region NUKLEUS-project
504	KlimaKonform+EZG_WE	Region NUKLEUS-project
505	EZG_WE	Region NUKLEUS-project
506	KlimaKonform_BBOX	Region NUKLEUS-project
507	KARE	Region NUKLEUS-project
508	ISAP	Region NUKLEUS-project
509	WAKOS_all	Region NUKLEUS-project
510	WAKOS_Festland	Region NUKLEUS-project
511	WAKOS_Inseln	Region NUKLEUS-project
512	Duisburg	Region NUKLEUS-project
513	Duisburg_3x3	Region NUKLEUS-project

Appendix D

Jupyter notebook to visualize and add custom subregions

We here provide a pdf copy of a jupyter notebook to visualize subregions and add custom subregions. In this example we show how to add a rectangular subregion, which covers entire Europe. You can find and download the jupyter notebook from `https://gitlab.dkrz.de/clm-community/evasuite/HZG_Evaluation_Suite/-/tree/add_custom_subregion_creation/scripts/create_custom_subregion?ref_type=heads`.

Example: Plotting regions from gpkg file and adding polygons

Christoph Braun, KIT, July 2024

Load required modules

In [1]:
```
import geopandas as gpd
from shapely.geometry import Polygon
```

Load data

In [2]:
```
proj = "PROJECT"
user = "USER"
```

In [3]:
```
path_to_example_data = "/work/"+proj+"/"+user+ \
                       "/evasuite_demofiles/example_data/"
```

In [4]:
```
input_file = path_to_example_data+"EvaSuite_subregion_shapes.gpkg"
```

In [5]:
```
data = gpd.read_file(input_file)
```

Explore data

In [6]:
```
data
```

Out[6]:

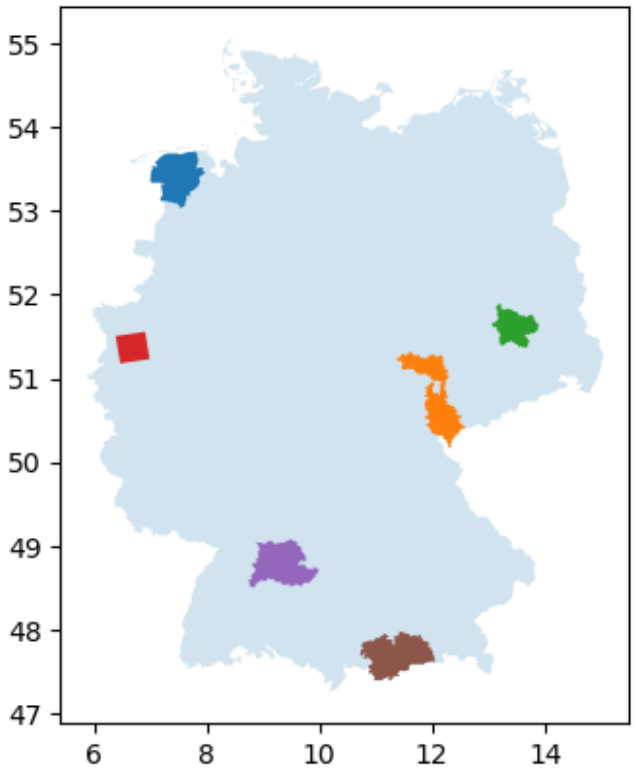
	ID	NAME_0	ID_0	ID_1	geometry
0	1.0	P_Alps	NaN	NaN	POLYGON ((5.00000 44.00000, 5.00000 48.00000, ...
1	2.0	P_BritishIslands	NaN	NaN	POLYGON ((-10.00000 50.00000, -10.00000 59.000...
2	3.0	P_East-Europe	NaN	NaN	POLYGON ((16.00000 44.00000, 16.00000 55.00000...
3	4.0	P_France	NaN	NaN	POLYGON ((-5.00000 44.00000, -5.00000 50.00000...
4	5.0	P_IberianPeninsula	NaN	NaN	POLYGON ((-10.00000 36.00000, -10.00000 44.000...
...	...	...	...	...	...
89	NaN	Duisburg_3x3	NaN	NaN	MULTIPOLYGON (((6.46907 51.17857, 6.98961 51.2...
90	NaN	EZG_Rhein	NaN	NaN	MULTIPOLYGON (((7.36470 51.96069, 7.37073 51.9...
91	NaN	EZG_Rhein_N	NaN	NaN	MULTIPOLYGON (((7.36470 51.96069, 7.37073 51.9...
92	NaN	EZG_Rhein_S	NaN	NaN	MULTIPOLYGON (((6.39623 49.77450, 6.39775 49.7...
93	NaN	DTL	NaN	NaN	MULTIPOLYGON (((9.12593 47.66864, 9.12577 47.6...

94 rows × 5 columns

... e.g. plot NUKLEUS regions

In [7]:
```
ax = data[data['NAME_0'] == 'Germany'].plot(alpha=0.2)
data[data['NAME_0'] == 'WAKOS_Festland'].plot(ax=ax)
data[data['NAME_0'] == 'KlimaKonform'].plot(ax=ax,color='C1')
data[data['NAME_0'] == 'IAWAK-EE'].plot(ax=ax,color='C2')
data[data['NAME_0'] == 'Duisburg_3x3'].plot(ax=ax,color='C3')
data[data['NAME_0'] == 'ISAP'].plot(ax=ax,color='C4')
data[data['NAME_0'] == 'KARE'].plot(ax=ax,color='C5')
```

Out[7]:
```
<AxesSubplot:>
```



Adding a new subdomain and writing to file

Create new polygon covering entire European region

In [8]:
```
lat_point_list = [35, 35, 70, 70, 35]
lon_point_list = [-10, 25, 25, -10, -10]

polygon_geom = Polygon(zip(lon_point_list, lat_point_list))
```

Add new polygon to data

In [9]:
```
# get last position of dataframe
entry_position = len(data)
```

In [10]:
```
data.loc[entry_position] = [entry_position,
                            "EUROPE_RECT",
                            "NaN",
                            "NaN",
                            polygon_geom]
```

Check resulting dataframe

In [11]:
```
data
```

Out[11]:

	ID	NAME_0	ID_0	ID_1	geometry
0	1.0	P_Alps	NaN	NaN	POLYGON ((5.00000 44.00000, 5.00000 48.00000, ...
1	2.0	P_BritishIslands	NaN	NaN	POLYGON ((-10.00000 50.00000, -10.00000 59.000...
2	3.0	P_East-Europe	NaN	NaN	POLYGON ((16.00000 44.00000, 16.00000 55.00000...
3	4.0	P_France	NaN	NaN	POLYGON ((-5.00000 44.00000, -5.00000 50.00000...
4	5.0	P_IberianPeninsula	NaN	NaN	POLYGON ((-10.00000 36.00000, -10.00000 44.000...
...	...	...	...	...	...
90	NaN	EZG_Rhein	NaN	NaN	MULTIPOLYGON (((7.36470 51.96069, 7.37073 51.9...
91	NaN	EZG_Rhein_N	NaN	NaN	MULTIPOLYGON (((7.36470 51.96069, 7.37073 51.9...
92	NaN	EZG_Rhein_S	NaN	NaN	MULTIPOLYGON (((6.39623 49.77450, 6.39775 49.7...
93	NaN	DTL	NaN	NaN	MULTIPOLYGON (((9.12593 47.66864, 9.12577 47.6...
94	94.0	EUROPE_RECT	NaN	NaN	POLYGON ((-10.00000 35.00000, 25.00000 35.0000...

95 rows × 5 columns

Write dataframe to gkpg file

In [12]:
```
data.to_file(path_to_example_data+"EvaSuite_subregion_shapes_extended.gpkg",
             driver="GPKG")
```

Appendix E

Available metrics in EvaSuite

This chapter was written by Sergei Petrov, Ronny Petrik, and Beate Geyer.

1 General Information

1.1 One model evaluation

Let **OBS** and **MOD** be some observational data and the model output for a specific variable, which is stored as a time series at the moments $t_k, k = \overline{1, N_T}$ on a structured 2D-grid $(x_i, y_j),\ i = \overline{1, N_x},\ j = \overline{1, N_y}$

$$\mathrm{OBS}(x_i, y_j, t_k),\ \mathrm{MOD}(x_i, y_j, t_k). \tag{1}$$

The list of available metrics is shown in the following table.

Eva-Suite name	Brief description	Output
'BIAS_temporal'	Temporal bias & Mean over time temporal bias for each location	$\mathbf{e_t}(x_i, y_j, t_k)$, $\mathbf{me_t}(x_i, y_j)$
'BIASnorm_temporal'	Normalized bias over time & Mean over time bias normalized by the temporally averaged observation values for each location	$\mathbf{e_{t,n}}(x_i, y_j, t_k)$, $\mathbf{me_{t,n}}(x_i, y_j)$
'BIAS_spatial'	Spatial bias & Mean over time spatial bias	$\mathbf{e_s}(t_k)$, $\mathbf{me_s}$
'BIASnorm_spatial'	Normalized spatial bias & Mean spatial bias normalized by spatially and temporally averaged observation values	$\mathbf{e_{s,n}}(t_k)$, $\mathbf{me_{s,n}}$
'MAE_temporal'	Temporal absolute bias & Mean over time temporal absolute bias for each location	$\mathbf{ae_t}(x_i, y_j, t_k)$, $\mathbf{mae_t}(x_i, y_j)$
'MAE_spatial'	Spatial absolute bias & Mean over time spatial absolute bias	$\mathbf{ae_s}(t_k)$, $\mathbf{mae_s}$
'MAD_temporal'	Temporal absolute bias & Median absolute deviation for each location over time	$\mathbf{ae_t}(x_i, y_j, t_k)$, $\mathbf{mad_t}(x_i, y_j)$
'MAD_spatial'	Spatial median absolute deviation & Mean over time spatial median absolute deviation	$\mathbf{mad_s}(t_k)$, $\mathbf{mmad_s}$
'RMSE_temporal'	The root of mean square error for each location over time	$\mathbf{rmse_t}(x_i, y_j)$
'RMSEbcn_temporal'	Normalized by mean over time observations unbiased root mean square error	$\mathbf{rmse_{t,n}}(x_i, y_j)$
'RMSE_spatial'	The spatial root mean square error & Mean over time spatial root mean square error	$\mathbf{rmse_s}(t_k)$, $\mathbf{mrmse_s}$
'LinCorr_temporal'	the linear correlation for each location over time	-
'LinCorr_spatial'	the Pearson correlation coefficient over in 2D space for each time step and temporal aggregation according to user-settings	-
'KlingGupta_temporal'	the Kling-Gupta efficiency for each location over time (composition of Linear correlation, BIAS and variability)	-
'quantile_bias_temporal'	the bias of quantiles for each location, computed from the PDF of all values in time	-
'quantile_mae_temporal'	the absolute mean error of quantiles for each location, computed from the PDF of all values in time	-
'quantile_iqd_temporal'	the inverse quadratic distance of quantiles for each location, computed from the PDF of all values in time	-
'pdf_pss_temporal'	The Persson-Skill-score for each location, computed from the PDF of all values in time	-

1.2 Evaluation of two models

Let **OBS** and **MOD_A** as well as **MOD_B** be some observational data and the output of two different models for a specific variable, which is stored as a time series at the moments $t_k, k = \overline{1, N_T}$ on a structured 2D-grid $(x_i, y_j),\ i = \overline{1, N_x},\ j = \overline{1, N_y}$

$$\mathrm{OBS}(x_i, y_j, t_k),\ \mathrm{MOD_A}(x_i, y_j, t_k),\ \mathrm{MOD_B}(x_i, y_j, t_k). \tag{2}$$

Then, a so-called skill score compares the performance of model A and model B against the reference data set (OBS). The following table lists the available skill scores.

Eva-Suite name	Brief description	Output
'AMSESS_temporal'	the advanced mean squared error skill score for each location over time, which is a normalization of the original mean squared error skill score.	-
'AMSESSsigm_temporal'	the significance for RMSE-based skill score (AMSESS) of model A to be better than those of model B (significance level of 0.10)	-
'AMSESSsigs_temporal'	the significance for RMSE-based skill score (AMSESS) of model A to be better than those of model B (significance level of 0.05)	-
'BIASsigm_temporal'	the significance for the BIAS of model A to be smaller than those of model B (significance level of 0.10). -	
'BIASsigs_temporal'	the significance for the BIAS of model A to be smaller than those of model B (significance level of 0.05)	-

2 Full descriptions of metrics

2.1 BIASes

2.1.1 Temporal bias

The **temporal bias** is 2D-field of difference between model output and observations for each time step. The **mean temporal bias** is the averaged over time 2D-field of difference between model output and observations

$$\mathbf{e_t}(x_i, y_j, t_k) = \mathrm{MOD}(x_i, y_j, t_k) - \mathrm{OBS}(x_i, y_j, t_k) \tag{3}$$

$$\mathbf{me_t}(x_i, y_j) = \frac{\sum_{k=1}^{N_T} \mathbf{e_t}(x_i, y_j, t_k)}{N_T} \tag{4}$$

2.1.2 Normalized temporal bias

The **normalized temporal bias** is a 2D-field of difference between model output and observations divided by the observations at each time step. The **mean normalized temporal bias** is the averaged over time 2D-field of difference between model output and observations divided by the temporally mean value of observations

$$\mathbf{e_{t,n}}(x_i, y_j, t_k) = \frac{\mathbf{e_t}(x_i, y_j, t_k)}{\mathrm{OBS}(x_i, y_j, t_k)} \tag{5}$$

$$\mathbf{me_{t,n}}(x_i, y_j) = \mathbf{me_t}(x_i, y_j) \cdot \left(\frac{\sum_{k=1}^{N_T} \mathrm{OBS}(x_i, y_j, t_k)}{N_T} \right)^{-1} \tag{6}$$

2.1.3 Spatial bias

The **spatial bias** is the averaged over spatial dimensions 1D-function of difference between model output and observations. The **mean spatial bias** is a single number that represents a mean over time spatial bias

$$\mathbf{e_s}(t_k) = \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} \mathbf{e_t}(x_i, y_j, t_k)}{N_x \cdot N_y} \tag{7}$$

$$\mathbf{me_s} = \frac{\sum_{k=1}^{N_T} \mathbf{e_s}(t_k)}{N_T} \tag{8}$$

2.1.4 Normalized spatial bias

The **normalized spatial bias** is the spatial bias normalized by spatially averaged observational data. The **mean normalized spatial bias** is the mean spatial bias normalized by spatially and temporally averaged observational data

$$\mathbf{e_{s,n}}(t_k) = \mathbf{e_s}(t_k) \cdot \left(\frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} \mathrm{OBS}(x_i, y_j, t_k)}{N_x \cdot N_y} \right)^{-1} \tag{9}$$

$$\mathbf{me_{s,n}} = \mathbf{me_s} \cdot \left(\frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_T} \mathrm{OBS}(x_i, y_j, t_k)}{N_x \cdot N_y \cdot N_T} \right)^{-1} \tag{10}$$

2.2 Mean Absolute Errors (MAEs)

2.2.1 Temporal mean absolute error

The **temporal absolute error** is 2D-field of absolute difference between model output and observations. The **temporal mean absolute error** is the averaged over time temporal mean absolute error

$$\mathbf{ae_t}(x_i, y_j, t_k) = \left| \mathbf{e_t}(x_i, y_j, t_k) \right| \tag{11}$$

$$\mathbf{mae_t}(x_i, y_j) = \frac{\sum_{k=1}^{N_T} \mathbf{ae_t}(x_i, y_j, t_k)}{N_T} \tag{12}$$

2.2.2 Spatial mean absolute error

The **spatial absolute error** is the averaged over spatial dimensions 1D-function of the absolute difference between model output and observations. The **spatial mean absolute error** is the single number that represents a mean over time spatial absolute error

$$\mathbf{ae_s}(t_k) = \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} \mathbf{ae_t}(x_i, y_j, t_k)}{N_x \cdot N_y} \tag{13}$$

$$\mathbf{mae_s} = \frac{\sum_{k=1}^{N_T} \mathbf{ae_s}(t_k)}{N_T} \tag{14}$$

2.3 Median Absolute Deviations (MADs)

We introduce the sample median function **MED** as the central value of the same sample in ascending order. Taking **MED** value over time for 2D time-series means taking the median value over time for each point of the grid. Taking **MED** value over space for 2D time-series means taking the central value (for 2D-sample with fixed time) for each time step.

2.3.1 Temporal median absolute deviation

The **temporal mean absolute deviation** is the median over time for the absolute difference between model output and the median over time observations

$$\mathbf{mad_t}(x_i, y_j) = \underset{t_k}{\mathrm{MED}}\Big(\Big|\mathrm{MOD}(x_i, y_j, t_k) - \underset{t_k}{\mathrm{MED}}\big[\mathrm{OBS}(x_i, y_j, t_k)\big]\Big|\Big) \tag{15}$$

2.3.2 Spatial median absolute deviation

The **spatial median absolute deviation** is the median over space dimensions for the absolute difference between model output and the median over space dimensions of observations. The **mean spatial mean absolute deviation** is the single number that represents the time-averaged spatial median absolute deviation

$$\mathbf{mad_s}(t_k) = \underset{x_i, y_j}{\mathrm{MED}}\Big(\Big|\mathrm{MOD}(x_i, y_j, t_k) - \underset{x_i, y_j}{\mathrm{MED}}\big[\mathrm{OBS}(x_i, y_j, t_k)\big]\Big|\Big) \tag{16}$$

$$\mathbf{mmad_s} = \frac{\sum\limits_{k=1}^{N_T} \mathbf{mad_s}(t_k)}{N_T} \tag{17}$$

2.4 Root mean square errors (RMSEs)

2.4.1 Temporal root mean square error

The **temporal root mean square error** is the root of the averaged over time squared bias for each location

$$\mathbf{rmse_t}(x_i, y_j) = \sqrt{\frac{\sum\limits_{k=1}^{N_T} \mathbf{e}_t(x_i, y_j, t_k)^2}{N_T}} \tag{18}$$

2.4.2 Normalized temporal root mean square error

The **normalized temporal root mean square error** is unbiased temporal root mean square error divided by mean over time observation data

$$\mathbf{rmse_{t,n}}(x_i, y_j) = \sqrt{\frac{\sum\limits_{k=1}^{N_T} \Big(\mathbf{e_t}(x_i, y_j, t_k) - \mathbf{me_t}(x_i, y_j)\Big)^2}{N_T}} \cdot \left(\frac{\sum\limits_{k=1}^{N_T} \mathrm{OBS}(x_i, y_j, t_k)}{N_T}\right)^{-1} \tag{19}$$

2.4.3 Spatial root mean square error

The **spatial root mean square error** is the root of the averaged over space dimensions squared bias for each time step. The **mean spatial root mean square error** is the temporal mean of spatial root mean square errors

$$\mathbf{rmse_s}(t_k) = \sqrt{\frac{\sum\limits_{i=1}^{N_x}\sum\limits_{j=1}^{N_y} \mathbf{e_t}(x_i, y_j, t_k)^2}{N_x \cdot N_y}} \tag{20}$$

$$\mathbf{mrmse_s} = \frac{\sum\limits_{k=1}^{N_T} \mathbf{rmse_s}(t_k)}{N_T} \tag{21}$$

2.5 Linear correlation

2.5.1 The temporal linear correlation

2.5.2 The spatial linear correlation / Pearson correlation

2.6 The Kling-Gupta efficiency

The efficiency is widely used for river discharge and reads

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{22}$$

involving the Pearson-Correlation coefficient r and the ratio α between the variability of modelled values and observed values. β is the so-called bias term, i. e. the ratio between the mean of the modelled values and the observed values. A perfect score would be equal to one and a considerably good performance is reached for values above 0.7.

2.7 Metrics based on probability density functions

2.7.1 BIAS of quantiles

The distribution of model and observations are analyzed for the quantiles 0.2, 0.4, 0.6, and 0.8 (standard settings). The bias for each of these quantiles is summed up. There is the option to remove the systematic bias from the calculation.

2.7.2 Mean absolute error of quantiles

The distribution of modelled values and observations are analyzed for the quantiles 0.2, 0.4, 0.6, and 0.8 (standard settings). The mean absolute error of these quantiles is summed up. There is the option to remove the systematic bias from the calculation.

2.7.3 Inverse quadratic distance

The distribution of modelled values and observations are analyzed for the quantiles 0.2, 0.4, 0.6, and 0.8 (standard settings). The inverse quadratic distance of these quantiles is summed up. There is the option to remove the systematic bias from the calculation.

2.7.4 Persson Skill-score

The PDF from modelled values and observed values is computed based on user-defined bins and for each location. Then, both PDFs are compared by computing

2.8 Skill score based on the RMSE, AMSESS

The advanced mean squared error skill score compares the mean squared errors between model A and model B based on a reference data set. It reads A score of 1 means that model A is much better than model B, and -1 means that model B is much better than model A. A score equal to zero means that the performance of models A and B are close to each other.

2.9 Significance of model skill based on the RMSE

2.9.1 Moderate Significance of the AMSESS

The PDF of the advanced mean squared error skill score (see section 2.8) is analyzed using bootstrapping, i. e. 500 hundred samples of observational data and model data (models A and B) are created to compute the AMSESS. If 90 percent of the distribution is below zero, model B is defined to be significantly better than model A and vice versa (i. e. a one-sided significance test with a significance level of 0.10).

2.9.2 Strong Significance of the AMSESS

The PDF of the advanced mean squared error skill score (see section 2.8) is analyzed by means of bootstrapping, i. e. 500 hundred samples of observational data and model data (model A and B) are created to compute the AMSESS. If 95 percent of the distribution is below zero, model B is defined to be significantly better than model A and vice versa (i. e. a one-sided significance test with a significance level of 0.05).

2.10 Significance of model skill based on the BIAS

2.10.1 Moderate Significance of the BIAS

The PDF of the BIAS is analyzed by means of bootstrapping, i. e. 500 hundred samples of observational data and model data are created to compute the BIAS again and again for model A and model B. Then, the distribution of the BIAS in model A and model B are compared by computing the difference between both distributions. If 90 percent of the distribution is below zero, the BIAS of model B is defined to be significantly smaller than those of model A and vice versa (significance test with a significance level of 0.10).

2.10.2 Strong Significance of the BIAS

The PDF of the BIAS is analyzed by means of bootstrapping, i. e. 500 hundred samples of observational data and model data are created to compute the BIAS again and again for model A and model B. Then, the distribution of the BIAS in model A and model B are compared by computing the difference between both distributions. If 95 percent of the distribution is below zero, the BIAS of model B is defined to be significantly smaller than those of model A and vice versa (significance test with a significance level of 0.05).