

ISIMIP/PROCLIAS workshop Potsdam 2024

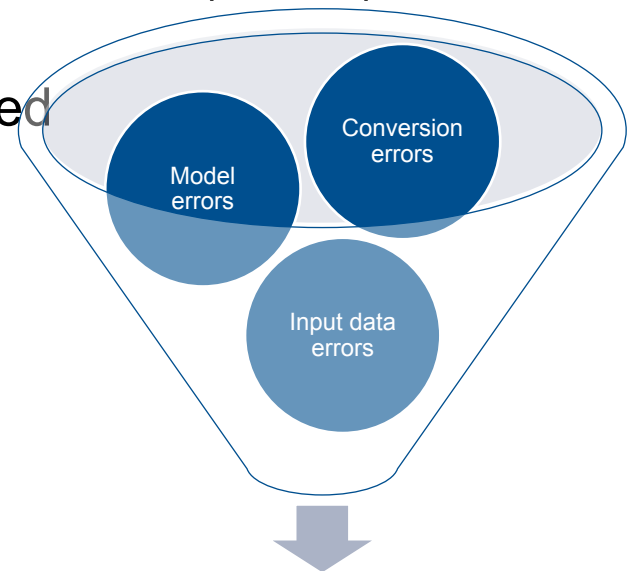
Automatic quality check / quality assessment of impact model output – results from PROCLIAS TG1.2

Hannes Müller Schmied^{1,2}, Jochen Klar³, Matthias Büchner³, Gerbrand Koren⁴ + PROCLIAS TG1.2

¹Goethe-University Frankfurt, ²Senckenberg SBIK-F Frankfurt ³Potsdam Institute for Climate Impact Research, Potsdam, ⁴Copernicus Institute of Sustainable Development, Utrecht



- Process-based impact models as valuable tools for simulating processes in a changing world (e.g., socio-economy, climate)
- Inter-Sectoral Model Intercomparison Project (www.isimip.org) act as umbrella for multiple sectors and ~ 100 modelling groups that follow a simulation protocol
- allows model intercomparison, evaluation and (cross-) sectoral (multi-) model impact assessments
- But: high quality model output is needed

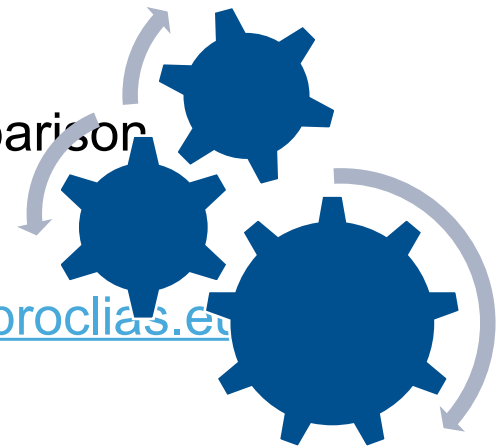


Objective

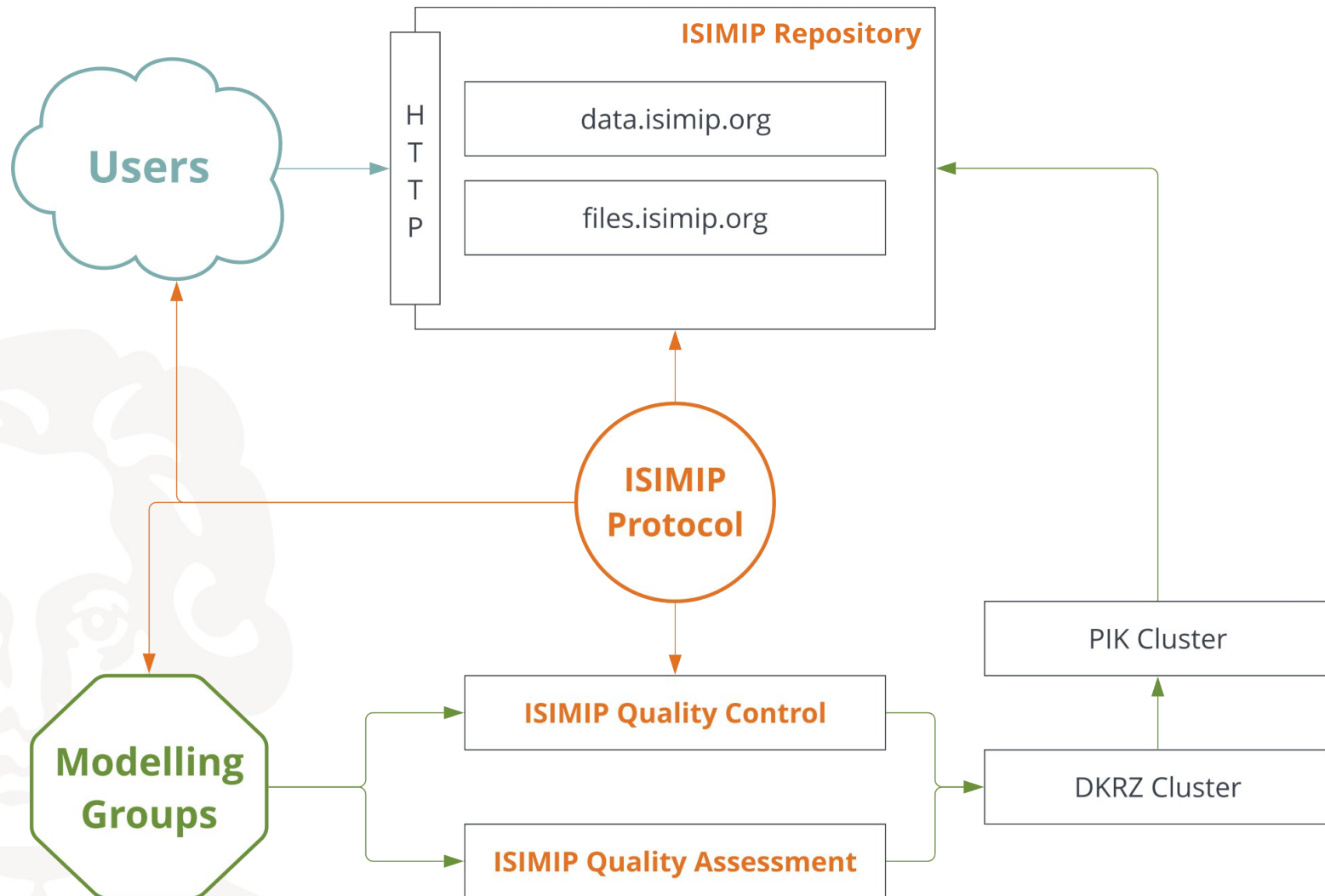
- **Establish a QC/QA workflow with tools to**
 - Formally check the model output for
 - data format and plausibility (Quality control, QC), and conduct
 - Quality assessment (QA) to
 - assess model output in comparison to existing output and (later) to
 - assess key variables with benchmark data
- à Avoids publishing incorrect data
- à Serves model development and model intercomparison
- à Integrated into the ISIMIP data infrastructure



n & TG1.2 in PROCLIAS <https://proclias.eu>



Overview and role of QC/QA-Tool in ISIMIP data infrastructure



- QC-tool is operational and used with each data submission
- Python command line tool (`pip install isimip-qc`)
- It checks (among others):
 - file name against the protocol patterns
 - NetCDF structure (variables, dimensions ...)
 - if the data is within a valid value range (collection within PROCLIAS)
 - **easy-to-use at your institution**



<https://github.com/ISI-MIP/isimip-qc>

```
$ isimip-qc ISIMIP3a/OutputData/biomes
CHECKING : example_gswp3-w5e5_counterclim_2015soc_default_clitterag-total_global_annual_1901_2019.nc
INFO      : Data model looks good (NETCDF4_CLASSIC).
INFO      : Variable "clitterag-total" compression level looks good (5)
INFO      : Dimensions for variable "clitterag-total" look good: ('time', 'lat', 'lon').
INFO      : 360 latitudes defined.
INFO      : 720 longitudes defined.
...
ERROR     : "units" attribute for "time" is "years since 1901-1-1 00:00:00". Should be ...
ERROR     : "calendar" attribute for "time" is "360_day". Must be one of ...
ERROR     : 360_day calendar is not allowed for monthly and annual data anymore. Use one of ...
INFO      : Variable properly chunked [1, 360, 720].
INFO      : Variable unit matches protocol definition (kg m-2).
INFO      : Missing value attribute "_FillValue" is properly set.
INFO      : Missing value attribute "missing_value" is properly set.
CRITICAL  : File did not pass all checks. Unfixable issues detected.
```

- QA-tool is in development and case-by-case on model output data
- Python command line tool (`pip install isimip-qa`, *planned*)
- Two-step approach:
 - **Extraction:** Create CSV files from NetCDF files, e.g. Time series of mean values for a (user provided) region or globally, count values, create maps
 - **Plot:** Create plots from the extractions, with or without additional computations, e.g. `dayofyear` from global or regional means.
- Usable with sensible defaults to be integrated into (semi-) automatic workflows
- Extractions can be fetched via the internet, e.g. to compare a local model against an ensemble of published models
- Modular design and friction-less extension to sector specific assessments.



<https://github.com/ISI-MIP/isimip-qa>

Example:

Time series for ISIMIP3a (water_global, qtot, monthly, global mean)

isimip-qa

```
{model}_{climate_forcing}_{climate_scenario}_histsoc_default_qtot_global_monthly
climate_forcing=gswp-w5e5,... climate_scenario=obsclim,counterclim model=CWatM,...
```

gswp3-w5e5

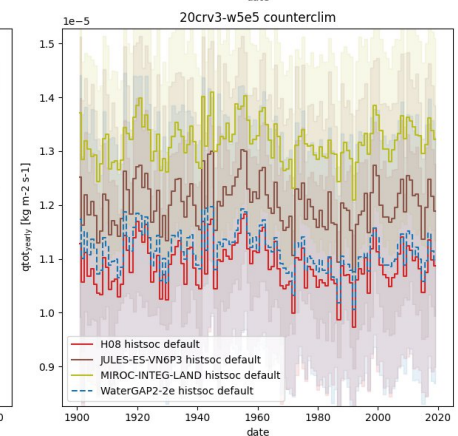
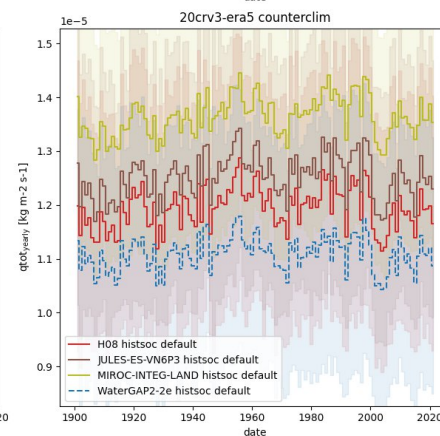
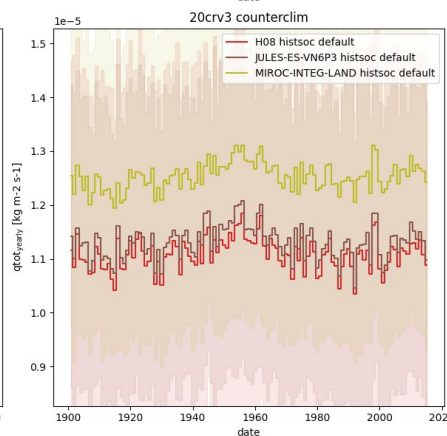
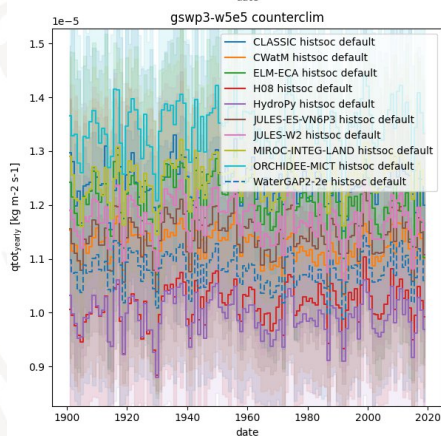
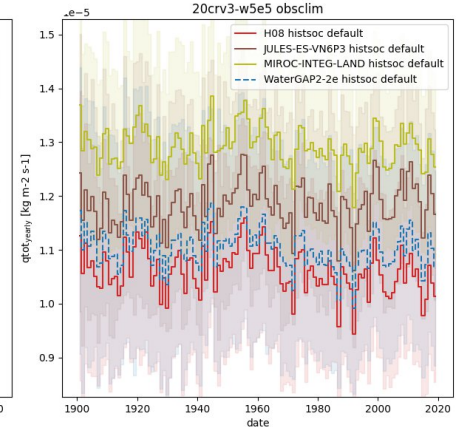
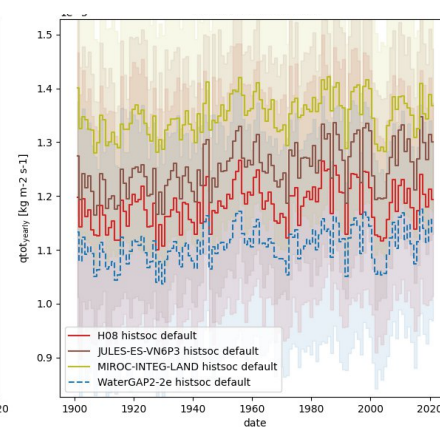
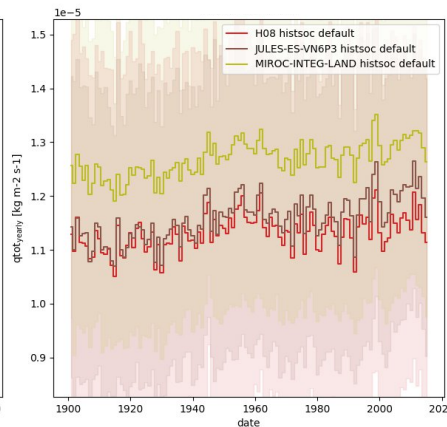
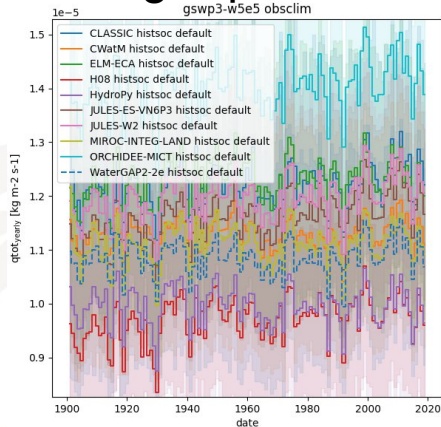
various_gswp3-w5e5- 20crv3

_histsoc 20crv3-era5

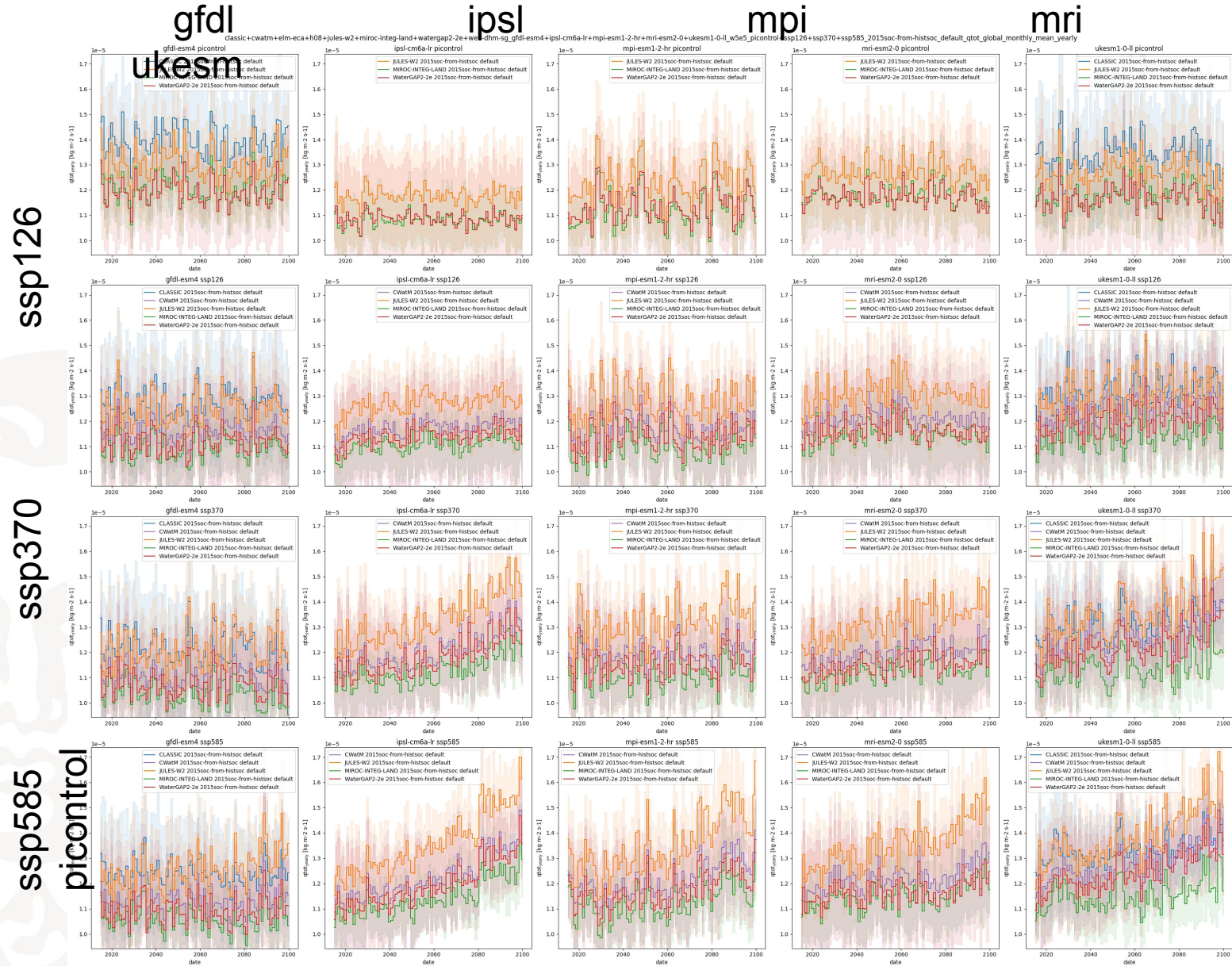
20crv3-w5e5

obsclim

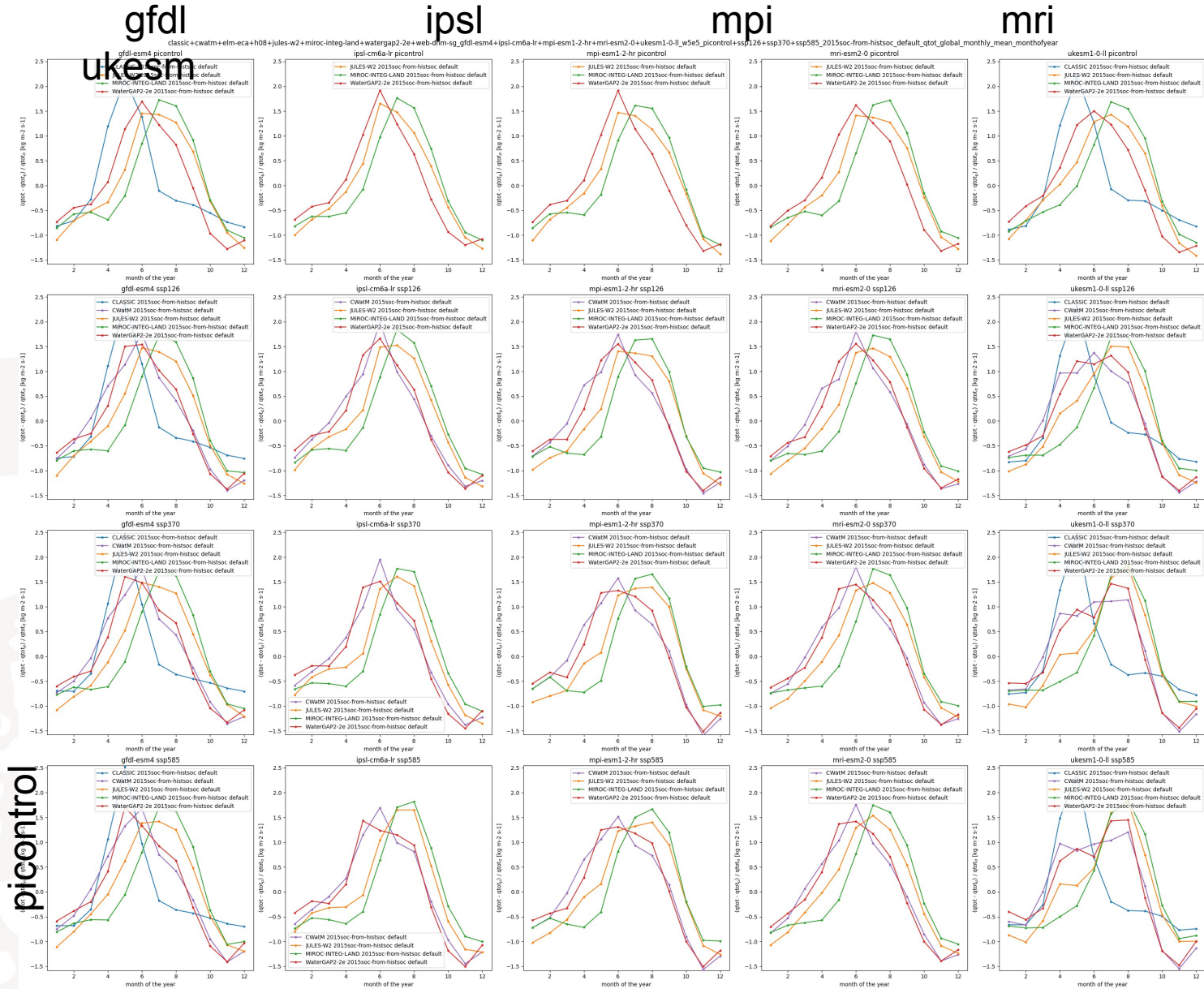
counterclim



Example:
Time series for ISIMIP3b (water_global, qtot, monthly, future, global mean)

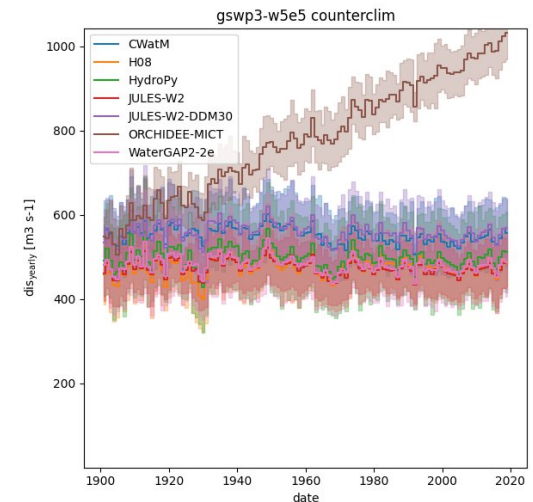
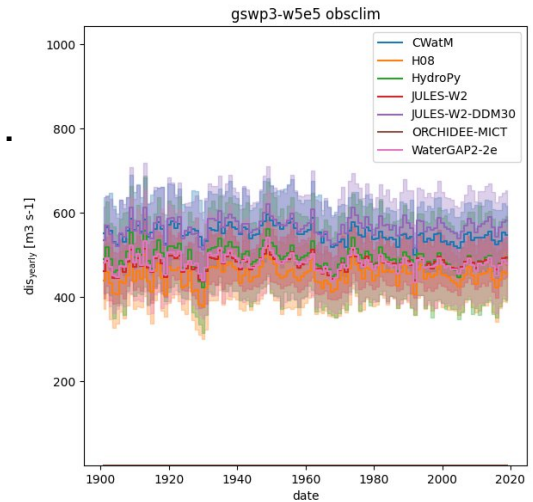
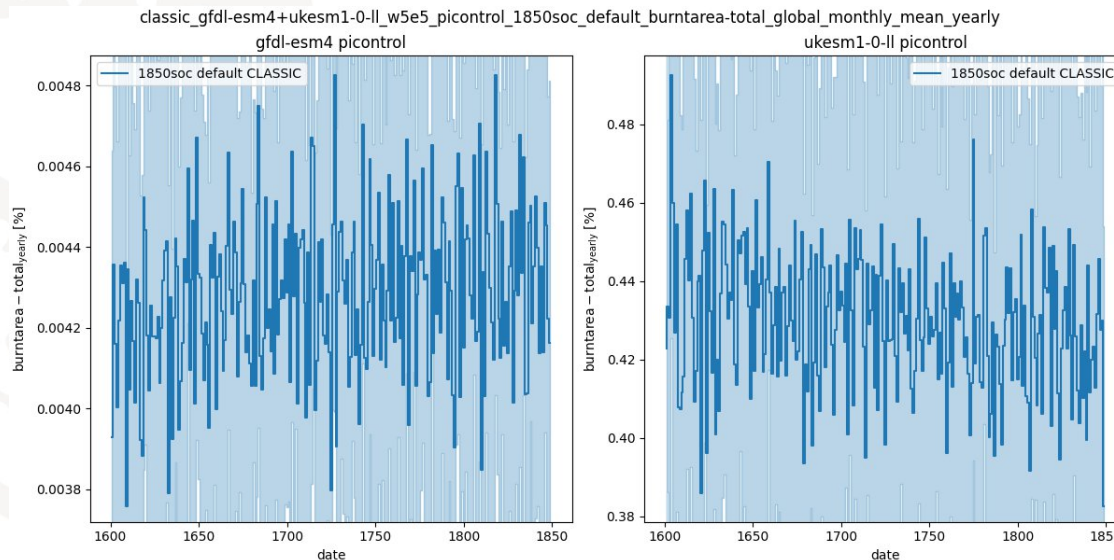


Example:
Montly means for ISIMIP3b (water_global, qtot, monthly, future, global mean)



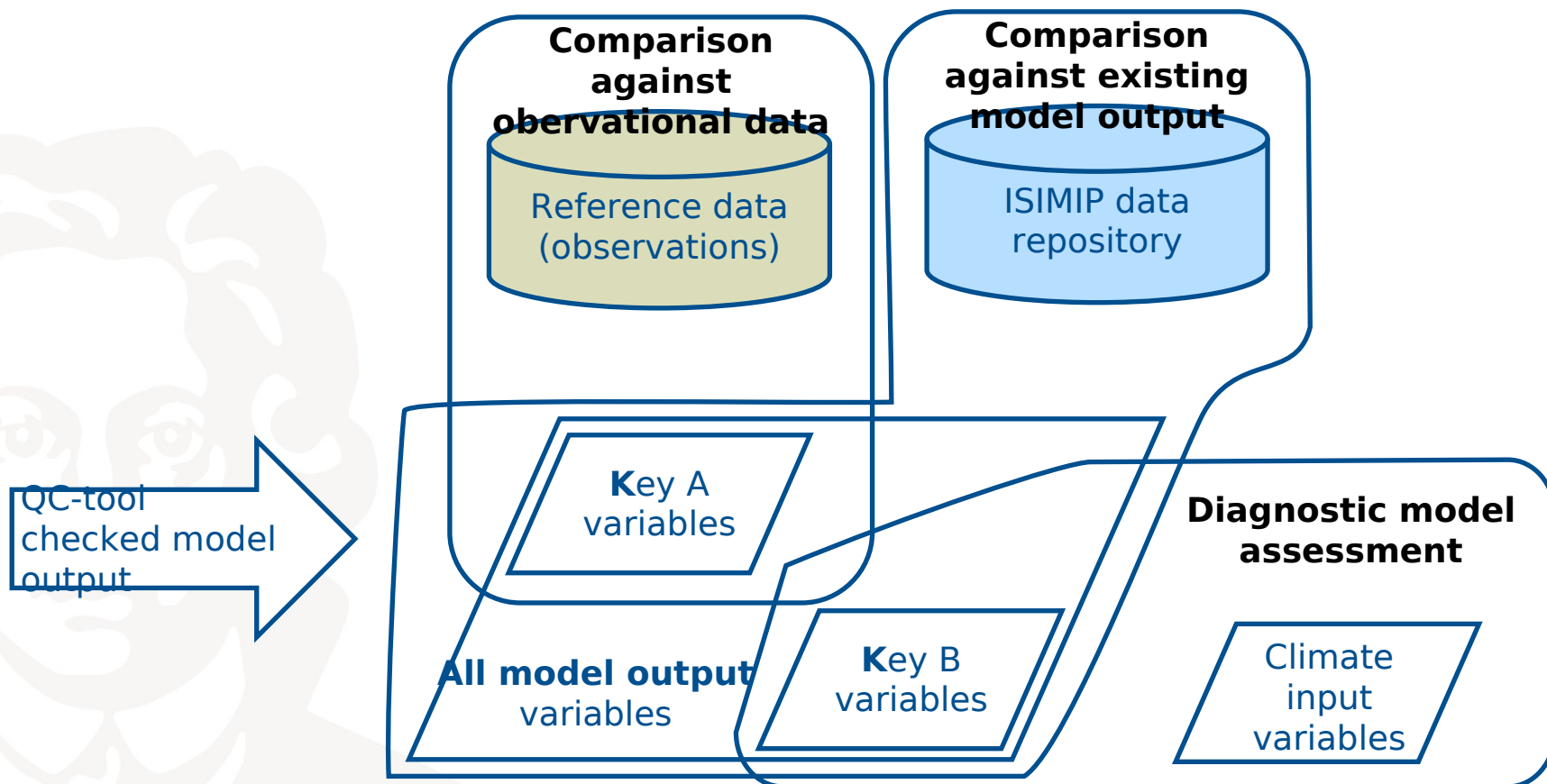
Example: Errors

- Right now, errors are found after visual inspection.
- The modellers are contacted and a solution is found.
- If the data is already public on data.isimip.org, a caveat is distributed.
- If not, the data is retracted/replaced at DKRZ and the modellers are informed.



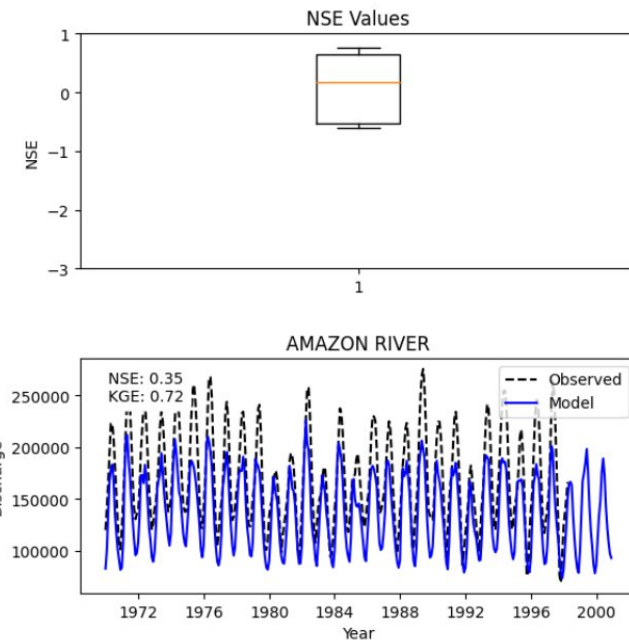
<https://data.isimip.org/caveats/48/>

- Several members of TG1.2 are closely related to ISIMIP global water
- ISIMIP/PROCLIAS hackathon Utrecht Sept 2023



Example for comparison against streamflow and terrestrial water storage anomalies; and model diagnostics

- **ISIMIP QA:** Extraction & preparation of model data
- **Sector specific:** performance assessment, visualization
- Streamflow: 1509 stations (GRDC, GSIM, ADHI)
- TWSA: 148 large river basins (> 200.000 km²)
- Model diagnostics (e.g. R/P), Gnann et al. 2023
- Output: e.g. a pdf / a html website, a traffic light system
- Each sector knows at best their benchmark data and metrics



river	station	id	lat	lon	ddm30_lat	ddm30_lon	country	catchment
Amazon river	Obidos, Porto	3629000	-1.95	-55.51	-1.75	-55.75	BR	4680000
Mississippi river	Vicksburg, MS	4127800	32.32	-90.91	32.25	-91.25	US	2964255
Ob	Salekhard	2912600	66.57	66.53	66.25	66.25	RU	2949998
Nile river	El ekhsase	2900000	29.70	31.28	29.75	31.25	EG	2900000
Lena	Stolb	2903430	72.37	126.80	72.25	126.75	RU	2460000
Orinoco	Puente Angostura	3206720	8.15	-63.60	8.25	-63.75	VE	836000

QC-tool

- Small improvements, tool is finished and in operational use; extensions possible (e.g. integration of new variables / variable ranges via ISIMIP protocol)

QA-tool

- Deployment in the quality control / publication workflow
- Selection of variable and regions
- Automatic quantitative assessment
- Finalization of global water pilot
 - Relation of performance indicators to basin characteristics
- Extension to other sectors
- Further coordination beyond PROCLIAS

QC/QA-Tool:

ISIMIP data management team (Jochen Klar, Matthias Büchner, Lisa Novak): isimip-data@pik-potsdam.de

Participation / Continuation within PROCLIAS TG 1.2:

Hannes Müller Schmied:

hannes.mueller.schmied@em.uni-frankfurt.de

