



Celebrating 30 years of
integrated climate impact research
at the Potsdam Institute.

Guide through the ISIMIP QC-Tool

PROCLIAS TG 1.2 Automatic QC/QA tool

Matthias Büchner, Iliusi Vega, Jochen Klar

Prague, June 5, 2023

The ISIMIP QC-tool

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



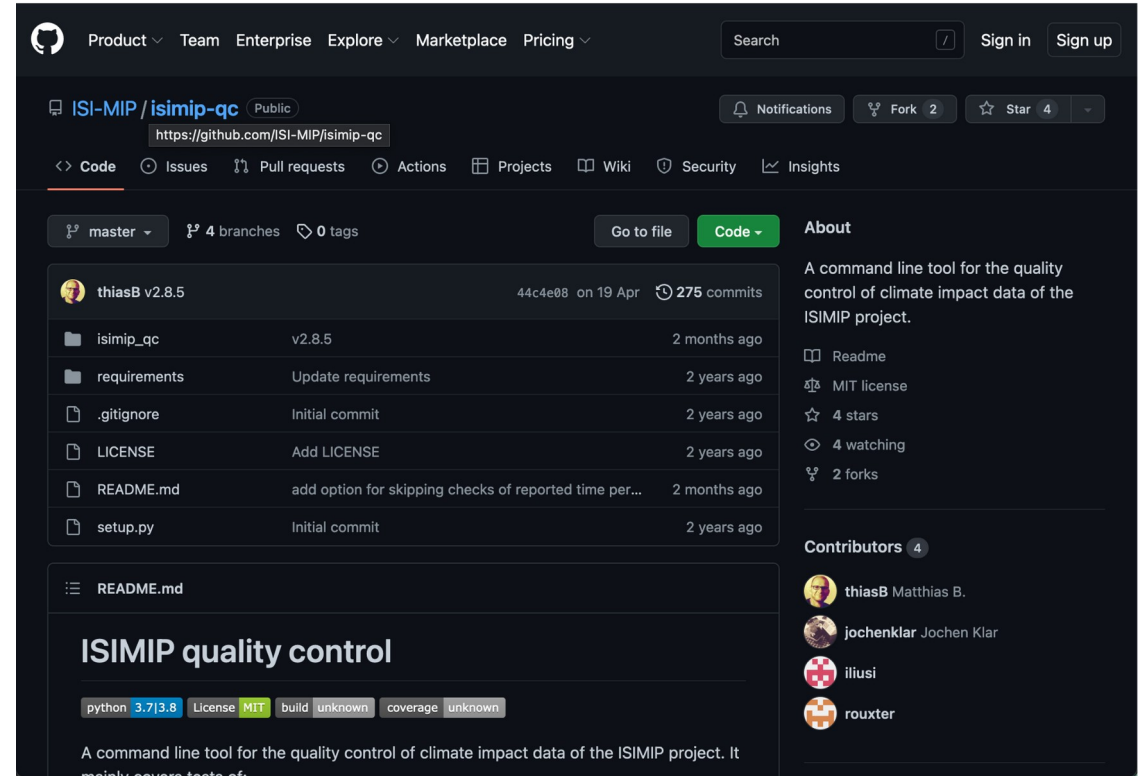
Celebrating 30 years of
integrated climate impact research
at the Potsdam Institute.

Leibniz
Leibniz
Association

What is it?

Command line tool for QC of ISIMIP climate impact data

- › Works on Linux, macOS, Windows
- › The ISIMIP QC-tool tests against protocol schemas:
 - ▶ File name
 - ▶ Variables, dimensions and global attributes
 - ▶ Data model and types
 - ▶ Consistency of NetCDF time axis
 - ▶ Data values range (when defined)



Celebrating 30 years of integrated climate impact research at the Potsdam Institute.



Installation

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




Celebrating 30 years of
integrated climate impact research
at the Potsdam Institute.



Requirements

- › Check modules already installed:
 - module list (Ubuntu/Debian)
 - pip list (macOS)
- › Python3:
 - sudo apt-get install python3 python3-dev python3-venv (Ubuntu/Debian)
 - brew install python (macOS)
- › Git:
 - sudo apt-get install git (Ubuntu/Debian)
 - brew install git (macOS)

```
(env) @mistralpp4% emacs isimip2-qc_test.sh
(env) @mistralpp4% ./isimip2-qc_test.sh
/work/bb0820/scratch/ISIMIP2b/unchecked/lakes_global/VIC-LAKE/
/work/bb0820/ISIMIP/ISIMIP2b/OutputData/lakes_global/VIC-LAKE/
/mnt/lustre01/work/bb0820/ISIMIP/ISIMIP2b/UploadArea/lakes_global/VIC-LAKE/qc-logs
CHECKING : /work/bb0820/scratch/ISIMIP2b/unchecked/lakes_global/VIC-LAKE/gfdl-esm2m/pre-industrial/vic-lake_gfdl-esm2m_ewembi_picontrol_1860soc_co2_watertemp_global_daily_1661_1670.nc4
INFO : File matched naming scheme: {'model': 'vic-lake', 'climate_forcing': 'gfdl-esm2m', 'bias_adjustment': 'ewembi', 'climate_scenario': 'picontrol', 'soc_scenario': '1860soc', 'sens_scenario': 'co2', 'variable': 'watertemp', 'region': 'global', 'time_step': 'daily', 'start_year': 1661, 'end_year': 1670}.
INFO : Global attribute "contact" looks good.
INFO : Global attribute "isimip_id" found (e252eaf2-23f2-4b8b-a092-269b87c1e660).
INFO : Global attribute "isimip_protocol_version" (c4c3403a2563a4a0201076537f00d1beaa438162) does not match tool current protocol version (3023fcd908f450c5dae5b399da8a78b52be77bc7).
INFO : Global attribute "isimip_qc_pass_date" is set to "2021-08-04 10:21 UTC".
INFO : Global attribute "isimip_qc_version" (2.6.1) does not match tool current tool version (2.7.3).
WARNING : Data model is NETCDF4 (not NETCDF CLASSIC).
INFO : Variable "watertemp" compression level looks good (9)
INFO : Dimensions for variable "watertemp" look good.
INFO : 360 latitudes defined.
INFO : 720 longitudes defined.
INFO : Latitude index order looks good (N to S).
INFO : Valid time unit found (days since 1661-01-01 00:00:00)
INFO : Valid calendar found (proleptic_gregorian)
INFO : Time period covered by this file matches the internal time axis (1661-1670)
INFO : Correct number of time steps (3652) given the defined calendar (proleptic_gregorian)
WARNING : watertemp.chunking=[1, 1001, 360, 720]. Should be [1, 1, 360, 720] (with proleptic_gregorian)
INFO : Variable unit matches protocol definition (K).
INFO : Missing value attribute "FillValue" is properly set.
INFO : Missing value attribute "missing_value" is properly set.
INFO : Depths order looks good (positive down).
INFO : File passed all checks without unfixable issues.
```



In the ISIMIP website you will find a video with detailed instructions on “How to use the ISIMIP Quality checking tool”, at <https://www.isimip.org/gettingstarted/>



Celebrating 30 years of integrated climate impact research at the Potsdam Institute.



Setup

```
# setup venv on Linux/macOS/Windows WSL
python3 -m venv env
source env/bin/activate

# setup venv on Windows cmd
python -m venv env
call env\Scripts\activate.bat

# install from GitHub
pip install git+https://github.com/ISI-MIP/isimip-qc

# update from Github
pip install --upgrade git+https://github.com/ISI-MIP/isimip-qc
```



Celebrating 30 years of
integrated climate impact research
at the Potsdam Institute.

Leibniz
Association

Usage

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



Celebrating 30 years of
integrated climate impact research
at the Potsdam Institute.



Usage

› Usage command:

```
isimip-qc [-h] [--config-file CONFIG_FILE]
  [-c] [-m] [--unchecked-path UNCHECKED_PATH]
  [--checked-path CHECKED_PATH]
  [--protocol-location PROTOCOL_LOCATIONS]
  [--log-level LOG_LEVEL]
  [--log-path LOG_PATH] [-f] [-w] [-e]
  [-r [MINMAX]] [--fix]
  [--fix-datamodel [FIX_DATAMODEL]]
  [--check CHECK]
  schema_path
```

› The only mandatory argument is the `schema_path`

› Options are described at <https://github.com/ISI-MIP/isi-mip-qc#usage>

```
Check ISIMIP files for matching protocol definitions

positional arguments:
  schema_path          ISIMIP schema_path, e.g. ISIMIP3a/OutputData/water_global

optional arguments:
  -h, --help          show this help message and exit
  --config-file CONFIG_FILE
                    File path of the config file
  -c, --copy          Copy checked files to CHECKED_PATH if no warnings or errors were found
  -m, --move          Move checked files to CHECKED_PATH if no warnings or errors were found
  -O, --overwrite    Overwrite files in CHECKED_PATH if present. Default is False
  --unchecked-path UNCHECKED_PATH
                    base path of the unchecked files
  --checked-path CHECKED_PATH
                    base path for the checked files
  --protocol-location PROTOCOL_LOCATIONS
                    URL or file path to the protocol when different from official repository
  --log-level LOG_LEVEL
                    Log level (ERROR, WARN, INFO, or DEBUG)
  --log-path LOG_PATH
                    base path for the individual log files
  --include VARIABLES_INCLUDE
                    include only this comma-separated list of variables
  --exclude VARIABLES_EXCLUDE
                    exclude this comma-separated list of variables
  -f, --first-file    only process first file found in UNCHECKED_PATH
  -w, --stop-on-warnings
                    stop execution on warnings
  -e, --stop-on-errors
                    stop execution on errors
  -r [MINMAX], --minmax [MINMAX]
                    test values for valid range (slow, argument MINMAX defaults to show the top
  -nt, --skip-time-span-check
                    skip check for simulated time period
  --fix              try to fix warnings detected on the original files
  --fix-datamodel [FIX_DATAMODEL]
                    also fix warnings on data model found using NCCOPY or CDO (slow). Choose pro
  --check CHECK      perform only one particular check
```



Celebrating 30 years of
integrated climate impact research
at the Potsdam Institute.



Try the demo script

- › Download the demo script at https://www.isimip.org/documents/625/isimip_qc-tool.txt
- › Check our recommended fixes in the ISIMIP page “*Preparing simulation files*”
- › Contact the ISIMIP data management team, they’ve gathered a long list of scripts with quick fixes

```
#!/bin/bash

# This^^ command sets bash, in the '/bin/' directory, as the default shell for executing
# commands in this file. Check bash location with: $ which bash

. /YOUR-PATH-TO-TOOL/qc-env/bin/activate

# Schema definitions:
ROUND='ISIMIP2b'
DATATYPE='OutputData'
SECTOR='water_global'

# Define working directories:
UNCHECKED_DIR=/YOUR-PATH/TO/DOWNLOADED-FILES
CHECKED_DIR=/YOUR-PATH/TO/PROCESSED-FILES
LOG_DIR=/YOUR-PATH/TO/LOG-FILES
echo $UNCHECKED_DIR
echo $CHECKED_DIR
echo $LOG_DIR

# Uncomment to only show the working directories:
#exit

# Create working directories if needed:
[ ! -d $CHECKED_DIR ] && mkdir -p $CHECKED_DIR
[ ! -d $LOG_DIR ] && mkdir -p $LOG_DIR

# Additional definitions for options
CHECKONE=check_time_dimension # check_variable
PROTOCOL_LOC=/YOUR-PATH/qc-env/isimip-protocol-2
VARNAME=csoil

# Command line:
isimip_qc $ROUND/$DATATYPE/$SECTOR/ \
  -f \
  --include $VARNAME \
  --log-level=INFO \
  --log-path=$LOG_DIR \
  --unchecked-path=$UNCHECKED_DIR \
  --checked-path=$CHECKED_DIR \

# Other options:
# --check $CHECKONE
# --protocol-location=$PROTOCOL_LOC \
# --skip-time-span-check \
# --fix --move # --copy
# --fix --fix-datamodel=cdo -copy
# --minmax

# NOTE: the '-fix-datamodel=cdo' option also fixes warnings on data model found using
# NCCOPY or CDO (this can be slow).

# NOTE: I store this script as an executable '*.sh' file (-rwxr-xr-x) at '/YOUR-PATH-TO-
# TOOL/qc-env/'
```



Questions? Write us!

- › General inquiries: info@isimip.org
- › ISIMIP data management team: isimip-data@pik-potsdam.de
- › The ISIMIP data management team are:
 - ▶ Matthias Büchner: buechner@pik-potsdam.de
 - ▶ Iliusi Vega: delvalle@pik-potsdam.de
 - ▶ Jochen Klar: klar@pik-potsdam.de



Celebrating 30 years of
integrated climate impact research
at the Potsdam Institute.



POTSDAM INSTITUTE FOR
CLIMATE IMPACT RESEARCH

