

the ISIMIP QC-Tool

basic format checks for ISIMIP simulation data

Matthias Büchner, Iliusi Vega, Jochen Klar

Prague, June 5, 2023

Overview

- Command line tool for QC of ISIMIP climate impact data
- Python package (Linux, macOS, Windows)
- tests against protocol schemas:
 - } File name patterns, ISIMIP controlled vocabulary (CV)
 - } Variables, dimensions and attributes
 - } Data model and types
 - } Consistency of NetCDF time axis (file name vs. Internal time axis)
 - } Data value ranges (when defined)

Open Source

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

The screenshot shows the GitHub repository page for 'thiasB v3.0.2' of the 'isimip-qc' project. The repository is on the 'master' branch, has 5 branches, and 3 tags. It contains 26b2386 commits, with 329 total commits. The repository description is: 'A command line tool for the quality control of climate impact data of the ISIMIP project.' The repository is licensed under MIT and has 4 stars, 4 watchers, and 2 forks. The repository is reported as 'Report repository'.

The repository files list includes:

File Name	Commit Message	Commit Date
isimip_qc	v3.0.2	last month
.gitignore	Prepare PyPI release	4 months ago
LICENSE	Add LICENSE	3 years ago
README.md	introduce forced copy/move	last month
requirements.txt	Prepare PyPI release	4 months ago
setup.cfg	Prepare PyPI release	4 months ago
setup.py	Prepare PyPI release	4 months ago

The README.md file content is:

ISIMIP quality control

python `>=3.6` License MIT

A command line tool for the quality control of climate impact data of the ISIMIP project. It mainly covers tests of:

- the file name against the protocol schemas and patterns
- variables, dimensions and global attributes
- data model and types
- some consistency checks on the NetCDF time axis and
- if the data is within a valid value range (when defined in the ISIMIP protocol)

This is still work in progress.

Releases 3

- 3.0.2 (Latest) last month
- + 2 releases

Contributors 4

- thiasB Matthias B.
- jochenklar Jochen Klar
- iliusi
- rouxter

Installation

- Python \geq 3.6
- Virtual environment recommended
- # install from the Python Package Index (PyPI), recommended
- `pip install isimip-qc`
- # update from PyPI
`pip install --upgrade isimip-qc`

Usage

- `isimip-qc [-h] [--config-file CONFIG_FILE] [-c] [-m] [-O]`
- `[--unchecked-path UNCHECKED_PATH] [--checked-path CHECKED_PATH]`
- `[--protocol-location PROTOCOL_LOCATIONS] [--log-level LOG_LEVEL]`
- `[--log-path LOG_PATH] [--log-path-level LOG_PATH_LEVEL] [--include INCLUDE_LIST]`
- `[--exclude EXCLUDE_LIST] [-f] [-w] [-e] [-r [MINMAX]] [-nt] [--fix]`
- `[--fix-datamodel [FIX_DATAMODEL]] [--check CHECK] [--force-copy-move] [-V]`
- `schema_path`

- Pulls protocol specifications from Github or local protocol repository
- Only mandatory argument is the `schema_path`
- Options described at <https://github.com/ISI-MIP/isimip-qc#usage>
- Log files per NetCDF data file

Valid ranges

- Helps identify wrong unit conversions and model inadequacies
- Variable ranges collected from coordinated effort among sector coordination
- Defined in protocol variable definitions (Github)
- log level VRDETAIL returns top lists of violations with time step and location (time consuming)

Fixing

- Capable of fixing certain aspects of metadata
 - - Data model, internal compression
 - - Variable long_names, units taken from ISIMIP protocol
 - - Removing global and variable attributes (per white list)
- Global attributes added:
 - - QC-tool version
 - - Protocol git hash
 - - Date of check
 - - Unique file id (uuid)

Questions, issues, development

- 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