



Contribution ID: 66

Type: Poster

## Manual data review and quality control –An add-on to SaQC

Wednesday 26 February 2025 19:40 (20 minutes)

The growing volume of high-resolution time series data in Earth system science requires the implementation of standardised and reproducible quality control workflows to ensure compliance with the FAIR data standards. Automated tools such as SaQC[1] address this need, but lack the capacity for manual data review and flagging. It is therefore the intention of this project to develop a Python-based tool with an intuitive graphical user interface (GUI) for local machines, thereby enhancing the functionality of SaQC. It is anticipated that the tool will be user-friendly, even for those with limited experience of Python. The GUI will therefore be capable of interactively visualising the time series data, highlighting the data that has already been automatically flagged. The selection of data points may be accomplished by clicking on them, and a flag may be assigned via a dropdown menu. An optional comment field may be utilised to record supplementary information, such as details of pollution events. Moreover, the option to unflag data that has failed the automated quality control process, but which is considered valid by the scientist, will be available.

The manual flagging tool will be based on SaQC, thereby facilitating future integration. Consequently, integration into an existing SaQC workflow will be straightforward. It should be noted, however, that this is not exclusive to SaQC users; it can be easily applied to data created by another tool for automatic quality control. A simple conversion of the data via the pandas library will be sufficient for utilisation of the manual flagging tool. The flagging schemes can either be adopted from SaQC or own schemes can be integrated. Following the flagging process, the user is then able to decide how to export the data set.

The manual flagging tool represents a valuable addition to existing toolkits for all scientists handling time-series datasets, effectively completing the data quality control process. From a scientific perspective, the benefits of this tool include increased efficiency and traceability in the data flow, as well as improved data quality through the fine-tuning of automatic controls based on experience and contextual knowledge.

[1] Schäfer, David, Palm, Bert, Lünenschloß, Peter, Schmidt, Lennart, & Bumberger, Jan. (2023). System for automated Quality Control - SaQC (2.3.0). Zenodo. <https://doi.org/10.5281/zenodo.5888547>

### I want to participate in the youngRSE prize

**Primary author:** BÜTTNER, Nicole (Institute of Meteorology and Climate Research –Atmospheric Aerosol Research, Karlsruhe Institute of Technology, Karlsruhe, Germany)

**Co-authors:** LOUISOT, Benjamin (Institute of Meteorology and Climate Research –Atmospheric Environmental Research, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany); Dr LORENZ, Christof (Institute of Meteorology and Climate Research –Atmospheric Environmental Research, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany); SCHÄFER, David (Research Data Management Team, Helmholtz Centre for Environmental Research GmbH –UFZ, Leipzig, Germany); Dr FÖSIG, Romy (Institute of Meteorology and Climate Research –Atmospheric Aerosol Research, Karlsruhe Institute of Technology, Karlsruhe, Germany)

**Presenter:** BÜTTNER, Nicole (Institute of Meteorology and Climate Research –Atmospheric Aerosol Research, Karlsruhe Institute of Technology, Karlsruhe, Germany)

**Session Classification:** Poster and Demo Session together with Reception

**Track Classification:** Research Software: visualisations and analysis