PROCEEDINGS OF SCIENCE

QualPipe: a quality control pipeline for the CTAO Data Processing and Preservation System

Matteo Balbo,* Mykhailo Dalchenko and Teresa Montaruli on behalf of the CTAO Consortium

Département de Physique Nucléaire et Corpusculaire, Université de Genève, 1211 Genève 4, Switzerland

E-mail: Matteo.Balbo@unige.ch, Mykhailo.Dalchenko@unige.ch, Teresa.Montaruli@unige.ch

The Cherenkov Telescope Array Observatory (CTAO) is a large-scale astrophysics project that will generate several petabytes of compressed data annually. Ensuring the quality of this vast data stream is critical for reliable scientific analyses.

QualPipe is the automated quality control pipeline integrated into the Data Processing and Preservation System (DPPS) of CTAO. It assesses data from the raw archival level (Data Level 0; DL0) to the science-ready stage (DL3), evaluating whether observation conditions were nominal, whether the hardware functioned correctly, and whether the data meet the necessary standards for scientific analysis. The pipeline evaluates a wide range of monitoring variables against a set of customizable criteria defined in configuration files, and produces quality reports for each set of telescopes participating in an observation (subarray), as well as for individual array elements.

A key feature of QualPipe is its design: the core code is entirely independent of the specific data quality criteria, which are defined externally through configuration files. This makes it highly adaptable to evolving quality requirements, including those of the early-stage commissioning of CTAO. The system includes a graphical user interface for QualPipe users, who can inspect current and past quality assessment reports and navigate through numerous diagnostic plots for each array element.

This document presents the architecture of QualPipe, its role in CTAO's quality control workflow, and how it supports the efficient validation of observational data at scale.

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland

ICRC 2025
The Astroparticle Physics Conference
Geneva July 15-24, 2025

*Speaker

PoS(ICRC2025)559

1. Introduction

The Cherenkov Telescope Array Observatory (CTAO) is set to become the world's leading ground-based gamma-ray observatory, with sites in both hemispheres and a diverse array of telescopes. The scientific output of CTAO critically depends on the quality of its data, which must be validated at every stage of processing. Given the scale and complexity of the observatory, manual quality control is impractical. To address this, the CTAO Data Processing and Preservation System (DPPS) integrates an automated data quality pipeline, `QualPipe`, designed to ensure that only data meeting rigorous standards are delivered for scientific analysis.

2. Project goals and motivation

The primary goal of `QualPipe` is to provide a robust, automated, and configurable framework for the quality assessment of CTAO data. The pipeline is responsible for:

- Collecting and synthesizing data quality measurements from other pipeline subsystems.
- Assessing the quality of data from raw (DL0) to science-ready (DL3) stages.
- Verifying that observation conditions and hardware performance are within nominal parameters.
- Applying a wide range of customizable criteria to monitoring quantities.
- Producing comprehensive, automated data quality reports and summaries.
- Providing a user interface to explore current and archival reports and quality plots.
- Supporting the needs of both routine operations and early commissioning phases.

`QualPipe` assigns an overall data quality status for each observation block, which is currently the minimum time unit used to express data quality. Although finer time units may be implemented in future releases, data quality reports can already be generated for arbitrarily longer intervals by combining the data quality reports of the individual observation blocks. `QualPipe` is designed to be modular, flexible, and extensible, allowing data quality criteria to evolve as the observatory matures and new requirements emerge.

3. QualPipe architecture

`QualPipe` is designed as a modular system with a clear separation between the automatic pipeline that produces quality metrics and the web application used to explore them. The main components are:

- Automatic data quality pipeline
- Storage system
- Web application

```yaml
metrics:
  - Metric:
      name: "atmosphere_height_measurement"
      input_source:
        "/simulation/service/atmosphere_density_profile/height"
      array_element: WEATHER_STATION
      data_category: DL1_EVENT
      telescope_specific: False
      criteria:
        - RangeCriterion:
            min_value: 0.0
            max_value: 120.0

  - Metric:
      name: "atmosphere_height_median"
      input_source:
        "/simulation/service/atmosphere_density_profile/height"
      array_element: WEATHER_STATION
      data_category: DL1_EVENT
      telescope_specific: False
      descriptor: MedianDescriptor
      criteria:
        - RangeCriterion:
            min_value: 30
            max_value: 40

- Metric:
    name: "interleaved_pedestals_charge_std"
    input_source: "/dl1/monitoring/telescope/pedestal/charge"
    array_element: LST
    data_category: DL1_EVENT
    telescope_specific: True
    descriptor: StandardDeviationDescriptor
    criteria:
      - TelescopeRangeCriterion:
          min_value:
            - [type, LST*, 5]
            - [id, 3, 4.3]
          max_value:
            - [type, LST*, 8]
            - [id, 3, 7.5]

- Metric:
    name: "interleaved_flatfield_charge_mean"
    input_source: "/dl1/monitoring/telescope/flatfield/charge"
    array_element: LST
    data_category: DL1_EVENT
    telescope_specific: True
    descriptor: MeanDescriptor
    criteria:
      - TelescopeThresholdCriterion:
          threshold: 102
          above: False
```

Figure 1: Generic example of a YAML configuration file describing how to derive quality metrics from an ingested file, and the criteria types and values to evaluate them.

3.1 Automatic data quality pipeline

Automation in `QualPipe` is essential for handling the large data volumes and operational complexity of CTAO. The pipeline is triggered automatically as new data become available, ensuring timely quality assessment without manual intervention. The `QualPipe` automatic pipeline receives ingested data, derives quality metrics via the application of quality descriptors, and evaluates the products against quality criteria. All instructions are customizable via configuration files (YAML or JSON). The core implements a set of functional descriptors and evaluation criteria independent of the specific data metrics to which they are applied.

A distinctive feature of `QualPipe` is its use of external configuration files for defining quality metric derivation and the application of quality criteria. This approach offers several advantages:

- **Flexibility:** new quality metrics can be added or existing ones updated as operational experience grows, without changing the pipeline code.
- **Adaptability:** criteria values can be defined globally, customized by telescope type, or specified for unique array element IDs (adopting the `ctapipe` tuple convention [1, 2]).
- **Transparency:** all metrics and criteria applications are documented and version-controlled, facilitating reproducibility.

This approach ensures that the system remains adaptable to evolving quality requirements, supporting both early-stage commissioning and later, more mature phases of the CTAO experiment, while maintaining robustness and reliability.

A generic example of a YAML configuration file is shown in Fig. 1. As illustrated, functional descriptors may include mean, median, standard deviation, first derivative, logical combinations, or more complex algorithms, while quality criteria may include simple thresholds and range checks.

PoS(ICRC2025)559

Useful glossary definitions can be found in the appendix (see page 7). The evaluation modules interpret the instructions in the configuration file and apply the appropriate logic to the data. The results of the quality assessment are then compiled into structured reports for each observation, including pass/fail status and detailed logs.

These reports are accessible via the web application, which supports filtering, searching, and navigation across different data levels and array elements.

3.2 Storage system

The newly produced quality metrics, applied criteria, evaluated results, and detailed reports are all stored in a dedicated storage environment, typically a relational database. The database is continuously updated as new products become available. It is made accessible to the DPPS data quality exploration user interface in read-only mode, allowing users to retrieve and display both recent and archival data quality information.

3.3 The `QualPipe` web application

The `QualPipe` web application is a fundamental component of the data quality ecosystem, providing an intuitive and powerful interface for users to interact with quality assessment results. The web application consists of two main parts: a backend service responsible for data retrieval and processing, and a modern frontend user interface for data exploration and visualization.

The backend is implemented using FastAPI, exposing a set of RESTful API endpoints for querying the database. These endpoints enable efficient retrieval of quality metrics, criteria, and reports, as well as support for filtering and aggregation based on user-selected parameters such as site, observation date, observation block, data level, array element type, and array element ID. The backend also handles basic data manipulation and formatting, ensuring that the frontend receives data ready for visualization.

The frontend is built with Jinja2 for server-side rendering of template pages and D3.js for dynamic, interactive data visualization. The user interface is designed to be responsive and user-friendly, enabling users to quickly access both summary and detailed information. Key features of the web application include:

- **Dashboard:** a comprehensive overview of recent and historical quality reports, with summary tables and pass/fail indicators for each observation and data level.
- **Interactive plots:** visualization of quality metrics, applied criteria, and diagnostic results using interactive charts and plots, allowing users to explore trends and anomalies.
- **Drill-down navigation:** the ability to navigate from high-level summaries to detailed reports for individual telescopes, subarrays, or auxiliary elements, with filtering by site, date, observation block, and data level.
- **Direct comparison:** side-by-side comparison of the same quality metrics across different array elements, facilitating the identification of outliers or systematic issues.

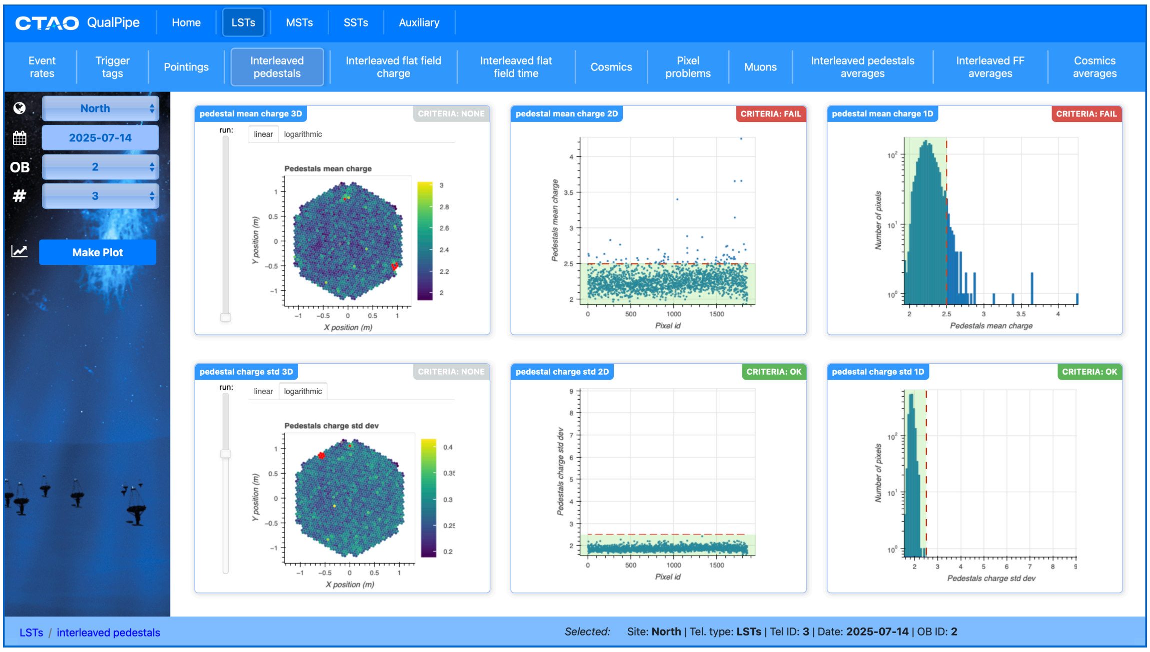

Figure 2: Example page of the data quality exploration user interface.

- **User feedback and validation:** (planned) mechanisms for users to flag issues, provide feedback, or validate quality reports, supporting continuous improvement of the quality control process.

The home page presents a summary table of the latest available quality reports across all data levels. Users can navigate to dedicated pages for each array element type (e.g., LSTs, MSTs, SSTs, Auxiliary) via a primary navigation bar, and access predefined sets of diagnostic plots through a secondary navigation bar. The sidebar allows users to select specific sites, dates, observation blocks, array element IDs, and data levels, dynamically updating the displayed data and plots.

The seamless integration between the backend and frontend ensures that users can efficiently retrieve and visualize various levels of quality data, supporting both routine monitoring and in-depth investigations. The use of modern web technologies guarantees scalability, maintainability, and a smooth user experience.

An example page of the data quality exploration user interface is shown in Fig. 2. In this example, the user has selected the *interleaved pedestals* page for the *LST* array elements, filtered by the *North* site, date *July 14th 2025*, *observation block 2*, and *array element ID 3*. The quality criteria are depicted in the figure as dashed red lines, while the permissible value ranges are highlighted with green shaded areas. The outcome of each metric's evaluation against its criteria is clearly indicated by a badge label positioned in the top right corner of every diagnostic plot.

4. Summary workflow overview

The typical workflow of `QualPipe` is as follows:

1. **Data ingestion:** new monitoring data and metadata are collected from various DPPS subsystems at stages from DL0 to DL3, as they become available.

2. **Metrics production:** new quality metrics are produced by the automatic pipeline, applying quality descriptors to the newly ingested data, following the instructions in the configuration file.

3. **Criteria application:** the pipeline checks the newly produced quality metrics against the appropriate quality criteria, as described in the configuration file. These criteria include thresholds and ranges, and their values can be defined globally or customized for specific array element types or unique individual IDs. Evaluation is performed independently for each array element and for the subarray as a whole.

4. **Report generation:** results are aggregated into structured quality reports, which can include summary tables, pass/fail indicators, and detailed logs.

5. **Output storage:** all produced quality metrics, applied quality criteria, and compiled quality reports are stored in a database, which is served to the `QualPipe` web application.

6. **User interaction:** through the data quality exploration user interface, users can browse current and historical quality reports, inspect diagnostic plots, analyze trends, and access detailed information for troubleshooting or validation.

5. Conclusion and outlook

`QualPipe` represents a robust and flexible solution for automated quality control in the CTAO Data Processing and Preservation System. Its modular architecture, externalized criteria configuration, and comprehensive reporting capabilities make it well-suited to the evolving needs of the observatory. As CTAO transitions from commissioning to full operations, `QualPipe` will continue to adapt, incorporating new quality metrics and criteria, and supporting efficient, reliable data validation at scale.

Future developments may include integration with additional monitoring systems, enhanced diagnostic capabilities, and further automation of feedback loops to support rapid response to quality issues.

Acknowledgments

We gratefully acknowledge financial support from the agencies and organizations listed here: CTAO Consortium's Acknowledgments.

APPENDIX: `QualPipe` glossary

- **Quality descriptor:** a function that defines how a quality metric is derived from input data. Examples of quality descriptors include mean, median, standard deviation, first derivative, or more complex calculations.

- **Quality metric:** a quantifiable measure that reflects an aspect of data quality or performance in the telescope system. Quality metrics are derived from input data (DL0-DL3) using quality descriptors and are evaluated against one or more quality criteria. They help assess data quality, verify the proper operation of array common elements, and identify anomalies. A quality metric must have an associated time interval.

- **Quality criterion:** a rule that should be applied to a quality metric to evaluate whether the telescope's data meets operational requirements. Examples of quality criteria types include 'threshold-based', 'range-based', 'multi-region', and 'categorical'.

- **Data quality report:** a structured summary of data quality metrics covering observations taken over a specific time interval, which could be an observation block, a lunar cycle, a year period, or user-defined. It includes the metric names, the associated instruments or external factors, the criteria types, and the results of their evaluations. They are stored in a designated system and can be inspected in the DPPS data quality exploration UI.

- **Data quality production:** a data reduction containing the quality metrics data, the applied quality criteria types and values, and the results of their evaluation. It includes also the necessary metadata required for their plot representation in the DPPS data quality exploration UI. Data quality productions are generated for individual array elements and for subarrays.

References

[1] K. Kosack, M. Linhoff, J. Watson, J. Jacquemier, L. Beiske, T. Bylund et al., *cta-observatory/ctapipe: v0.26.0*, June, 2025. 10.5281/zenodo.3372210.

[2] M. Linhoff, L. Beiske, N. Biederbeck, S. Fröse, K. Kosack and L. Nickel, *ctapipe – prototype open event reconstruction pipeline for the cherenkov telescope array*, in *Proceedings, 38th International Cosmic Ray Conference*, vol. 444, 2023, DOI.