A Curated Dataset for the Methane Emissions Technology Evaluation Center (METEC)

Jerry Duggan, Anindita Chakraborty, Bryan Rainwater
CSU Energy Institute, Colorado State University

METEC
COLORADO STATE UNIVERSITY

Background

The objective of providing a curated data set is to create a publicly available data ecosystem which anyone can access to support their development efforts.

Scientific importance
- High-resolution release data provide a critical foundation for developing and refining atmospheric methane transport and detection models.
- Enable independent evaluation of sensor detection limits, response times, and long-term stability.

Data streams captured every day (~70 MB/day)
- Controlled releases: above ground and below ground emission rates and setpoints.
- Meteorology: wind speed and direction, temperature, humidity, and air pressure.
- Reference sensors: background methane and ancillary gases to provide site-wide context.

Current and emerging applications
- Open access for atmospheric modelers, technology developers, and policy makers.
- Operational support, including routine site monitoring with real-time alarms and early warnings.

Methodology

Integrated data pipeline
- SCADA and field sensors transmit raw observations to a centralized database for secure long-term archiving.
- Automated scripts ingest daily operator logs and release schedules, ensuring synchronization of physical and digital records.

Automated processing and packaging
- Conversion of raw engineering data into standardized, analysis-ready files with complete metadata.
- Daily synthesis transforms detailed operator records into a concise, researcher-friendly data product.

Quality control (QC)
- Continuous automated checks compare captured values against expected physical ranges and operational setpoints.
- Flagged anomalies are carefully reviewed through targeted manual inspections supported by diagnostic QC graphs and dashboards.

Data product structure
- Organized into clearly defined components: Controlled Releases, Meteorology, and Reference Sensors, with consistent file formats.

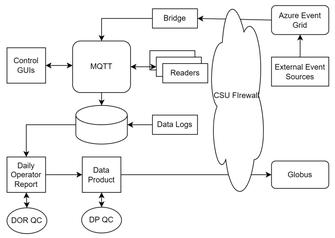

Figure 1: Data enters the system through real-time messaging, external data ingress, or batch upload of offline sensor systems. Reports for operations data and publicly accessible data are created daily. QC steps ensure high-quality data.

Results

Reliable, high-quality daily data product
- Seamless integration of release, weather, and reference sensor data from multiple instruments.
- Consistent timestamp alignment ensures comparability across datasets.
- Human review at critical production stages.

Enhanced reproducibility and transparency
- Each release is documented with a complete operational context and automated QC summaries.
- Reproducible processing pipeline reduces human error and supports long-term sustainability.

Scalability and flexibility
- System architecture readily incorporates additional sensors, new release rigs, and future experimental designs.
- Modular design allows rapid adjustments to meet evolving research needs

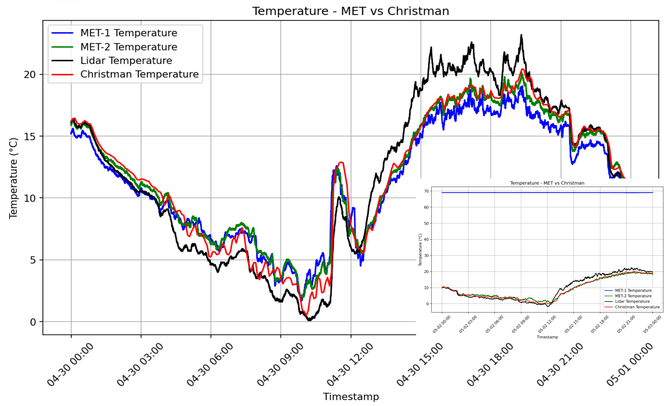

Figure 2: Comparison of meteorological temperature data (°C) from MET stations and LIDAR measurements with corresponding observations from the Christman Field weather station. Christman Field is a long-term National Weather Service station located near the METEC site. Its high-quality, continuously monitored records provide an independent reference for validating on-site meteorological measurements. The inset in the lower-right corner highlights a case where one MET station was reporting erroneous data, illustrating how these plots quickly reveal such issues

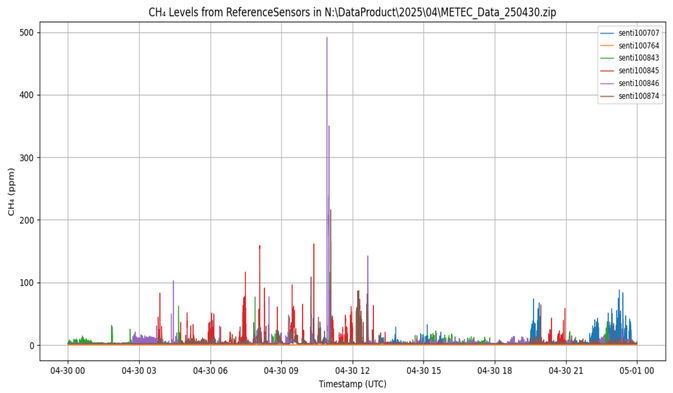

Figure 3: Methane concentration measurements (ppm) recorded by six independent Project Canary sensors

Conclusions and Next Steps

Expand the scope of METEC
- Incorporate data from ongoing and planned sub-projects, including satellite-based release detection, autonomous mobile methane measurement units (AMMMU) experiments, remote release rigs, and investigator-led campaigns.

Integrate advanced analytics
- Embed automated atmospheric transport and emission-rate models developed by the METEC team to provide near-real-time derived products.

Improve data access and usability
- Develop web-based portals and APIs for finer-grained queries, on-demand data selection, and easy integration with external modeling frameworks.

Sustain data quality and adaptability
- Maintain rigorous automated QC and responsive manual oversight as infrastructure, instrumentation, and scientific data evolve.

Operational challenges
- Continuous automated capture of multiple data types in near real time.
- Processing, validating, and packaging large volumes of heterogeneous data without manual intervention.
- Maintaining consistent quality while accommodating evolving site infrastructure and experimental designs.

Acknowledgements and Contact Information

Acknowledgements:
- This material is based upon work supported by the Department of Energy under Award Number(s) DE-FE0032276.
- Colorado State University METEC team for site operations, sensor maintenance, and data support.

Jerry Duggan
Research Associate
CSU Energy Institute, Colorado State University
Jerry.Duggan@colostate.edu

Anindita Chakraborty
Simulation Software Engineer
CSU Energy Institute, Colorado State University
anindita.chakraborty@colostate.edu