INDRA-ASTRA

Hindu mythology

INDRA Deity of lightning, thunder, rains and river flows

INDRA-ASTRA Indra's weapon

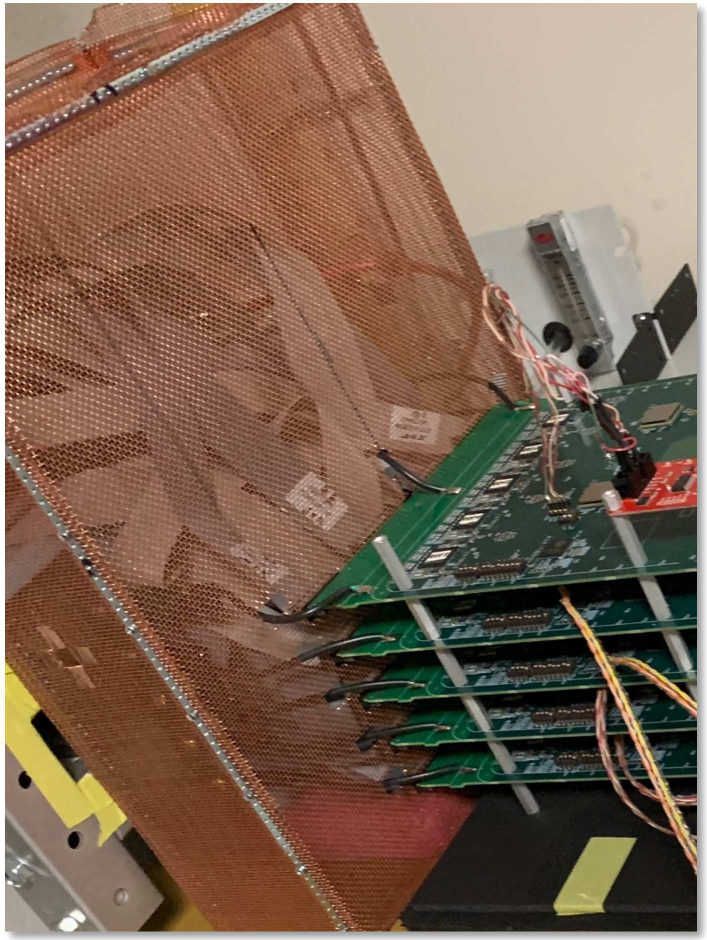
Jefferson Lab

INDRA Facility for **I**nnovations in **N**uclear **D**ata **R**eadout and **A**nalysis

INDRA-ASTRA Evaluation & Development of Algorithms & Techniques for Streaming Detector Readout

Markus Diefenthaler

Address Challenges of Autonomous Control and Experimentation



INDRA-ASTRA

Develop a prototype for an autonomous, responsive detector system as a first step towards an autonomous, self-conscious experiment.

R&D integrated with streaming readout and AI/ML efforts at Jefferson Lab

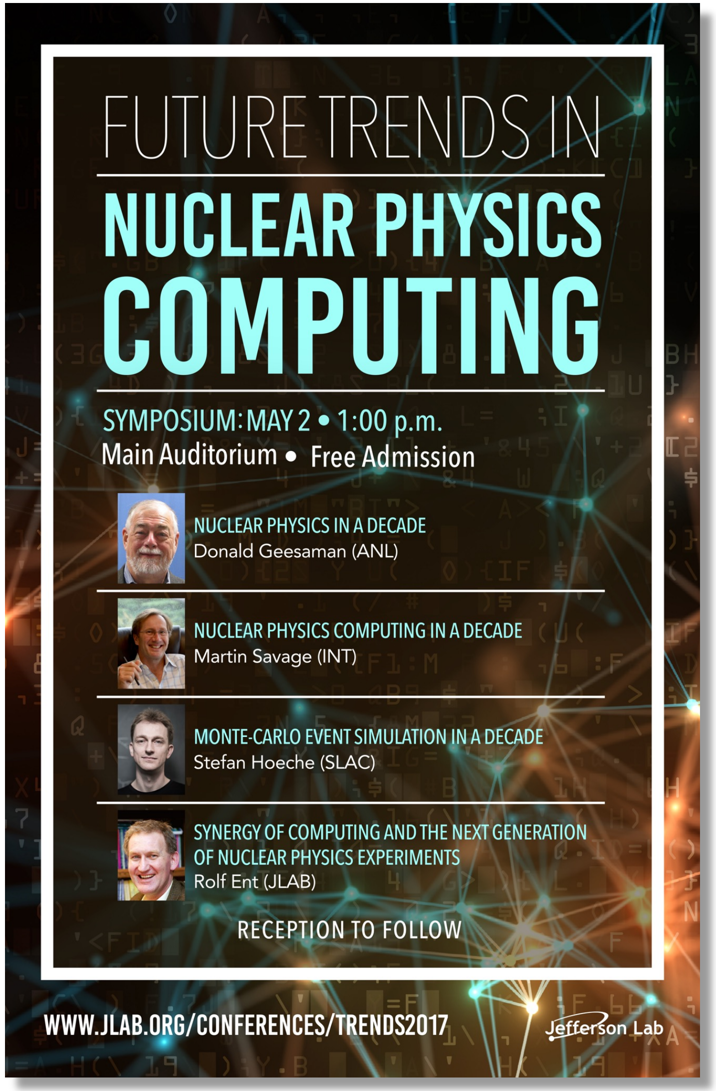
Team

Jefferson Lab
- **ENP** M. Diefenthaler, E. Jastrzembski, H. Szumila-Vance
- **CST** D. Lawrence, V. Gyurjyan

Old Dominion University
- **Applied Numerical Mathematics** R. Fang, A. Farhat, Y. Xu

Databricks
- S. Rajamohan

Future Trends



Details on https://www.jlab.org/FTNPC

Donald Geesaman (ANL, former NSAC Chair) “*It will be **joint progress of theory and experiment** that moves us forward, not in one side alone*”

One path
- Sharing data early with theory
- Sharing event-level information, not only histograms
- Comparing experiment and theory at the event level

Martin Savage (INT) “*The next decade will be looked back upon as a **truly astonishing period in Nuclear Physics** and in our understanding of fundamental aspects of nature. This will be **made possible by advances in scientific computing** and in how the Nuclear Physics community organizes and collaborates, and how DOE and NSF supports this, to take full advantage of these advances.*”

We can make a difference
- AI/ML for autonomous control and experimentation is a tremendous opportunity.
- AI4EIC is vital part how the NP community organizes and collaborates.

Our Vision for Software & Computing at the EIC

Rapid turnaround of data for the physics analysis and to start the work on publications:

- **Problem** Aligned, calibrated, reconstructed, and validated data for physics analyses and the resulting publications available after O(1year) due to complexity of NP experiments (and their organization).
- **Goal** Analysis-ready data from the DAQ system.
- **Solution** Compute-detector integration with AI at the DAQ and analysis level.

More details https://eic.github.io/activities/principles.html

Principle 2:
We will have an unprecedented compute-detector integration:

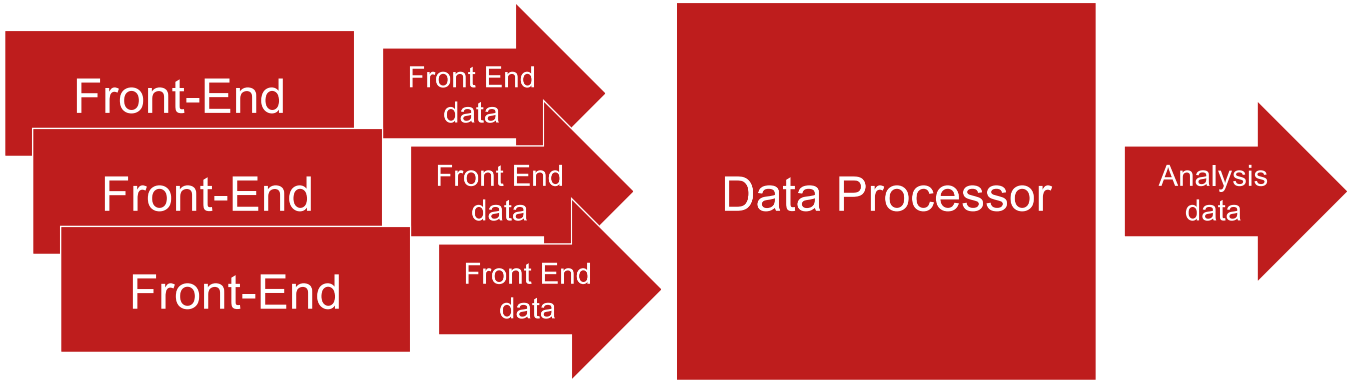
- We aim for autonomous alignment and calibration.
- We aim for a rapid, near-real-time turnaround of the raw data to online and offline productions.

Integration of DAQ, analysis and theory to optimize physics reach

Research model with seamless data processing from DAQ to data analysis

- Building the best detector that fully supports streaming readout and AI/ML:
  - FastML for alignment, calibration, and reconstruction in near real time.
    - Applications and Techniques for Fast Machine Learning in Science (*Front.Big Data* 5 (2022) 787421)
  - AI for intelligent decisions
- For rapid turnaround of data for the physics analysis and to start the work on publications.

Streaming Readout and (near) real-time processing



Data Processor

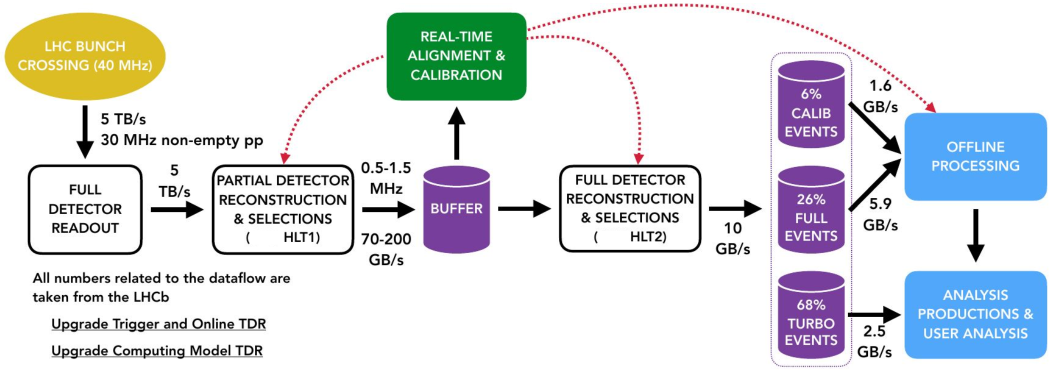
- Assembles data into physics events
- Outputs data suitable for physics analyses and the resulting publications

Features

- **FastML**
  - Autonomous alignment
  - **Autonomous calibration (INDRA-ASTRA)**
  - Reconstruction
  - Event filtering based on full event information
  - Autonomous anomaly detection
- **AI**
  - **Responsive detectors (INDRA-ASTRA)**
  - Conscious experiment

Allen Example from LHCb

LHCb Upgrade Dataflow



HLT1 challenge: reduce 5 TB/s to 70-200 GB/s in real-time with high physics efficiency

Autonomous calibrations

To deal with time-changing data, one needs strategies, at least, for the following

- detecting when a change occurs
- determining which examples to keep and which to drop (if any)
- updating calibrations models when significant change is detected

"In most challenging data analysis applications, data evolve over time and must be analyzed in near real time. Patterns and relations in such data often evolve over time, thus, models built for analyzing such data quickly become obsolete over time. In machine learning and data mining this phenomenon is referred to as **concept drift**." (I. Žliobaitė, M. Pechenizkiy, J. Gama , An Overview of Concept Drift Applications)

	Machine Learning	A method by which we abstract information from raw data
Alternative	**Multiscale Method**	A way that we represent a function, a signal or a system.

Autonomous calibrations in INDRA-ASTRA

OUR APPROACH

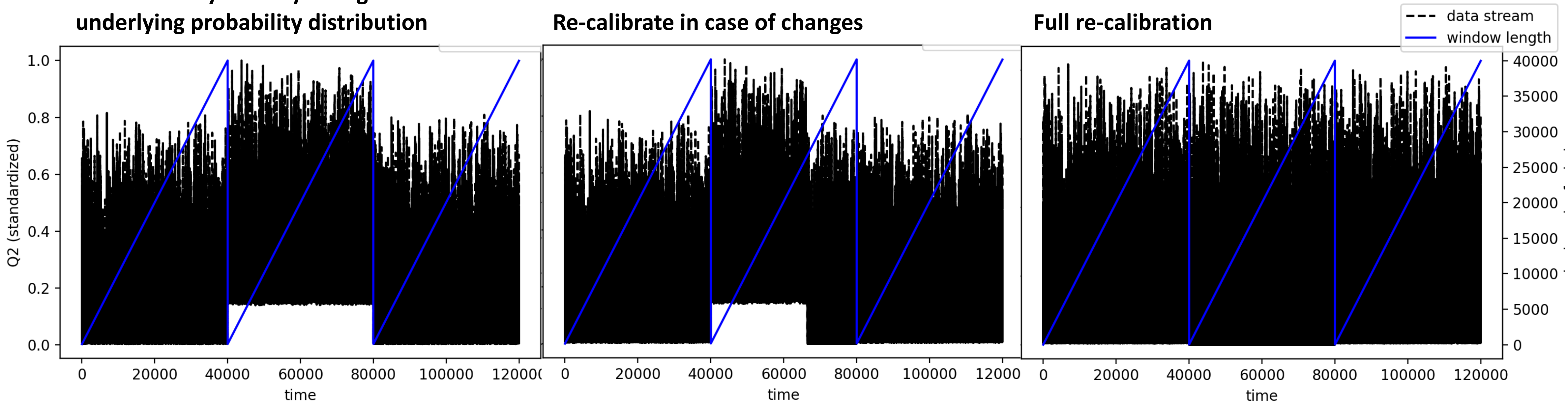
1. **Identify different data-taking periods** Use the start of distinct data-taking periods based on changes in the mean of the data stream. Explored ADWIN2 and multiscale algorithms for change detection.

2. **Calibrate different data-taking periods to a baseline** Calibrate w.r.t. baseline calibration due to lack of features for fast calibrations in DIS. measurement.

Automatically identify changes in the underlying probability distribution

Re-calibrate in case of changes

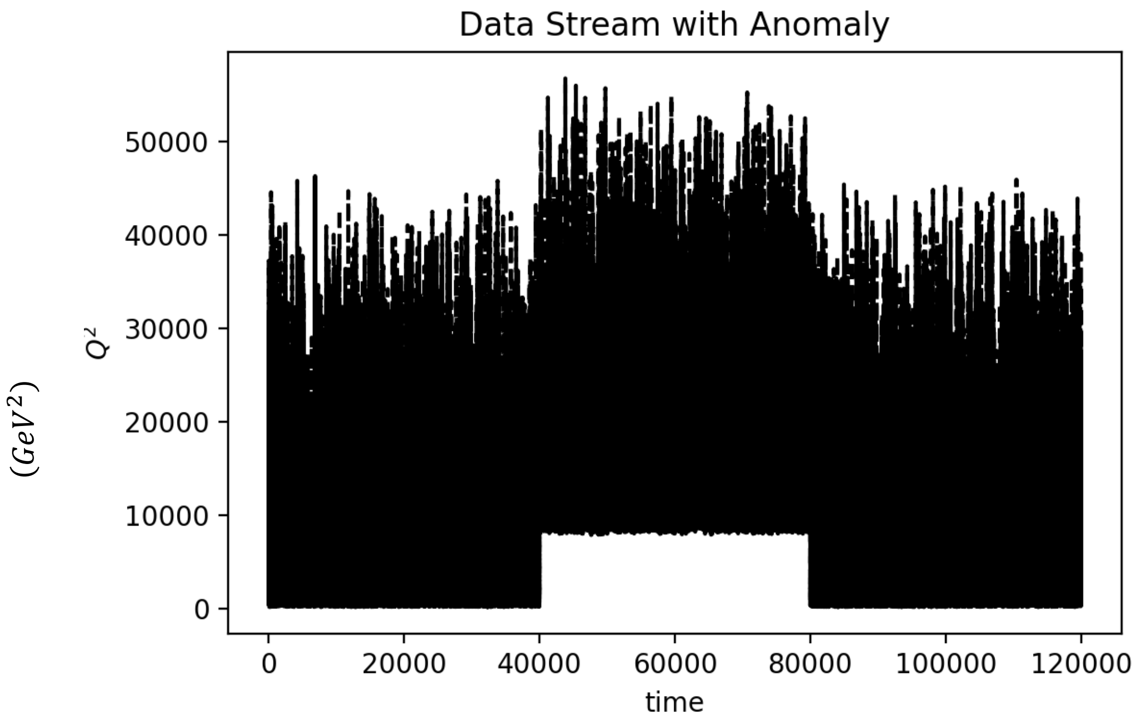
Full re-calibration



Toy Example

To represent the data stream we use a sample of 120,000 Inclusive Deep Inelastic Scattering Monte Carlo events

- generated in the context of the ZEUS experiments
- Includes full detector simulation
- Reconstructed kinematics with all detector effects.



We observe a stream of x and Q^2, reconstructed by the electron method [3] based on the measurement of the (x, y, z) position and energy E of the outgoing lepton in the calorimeter.

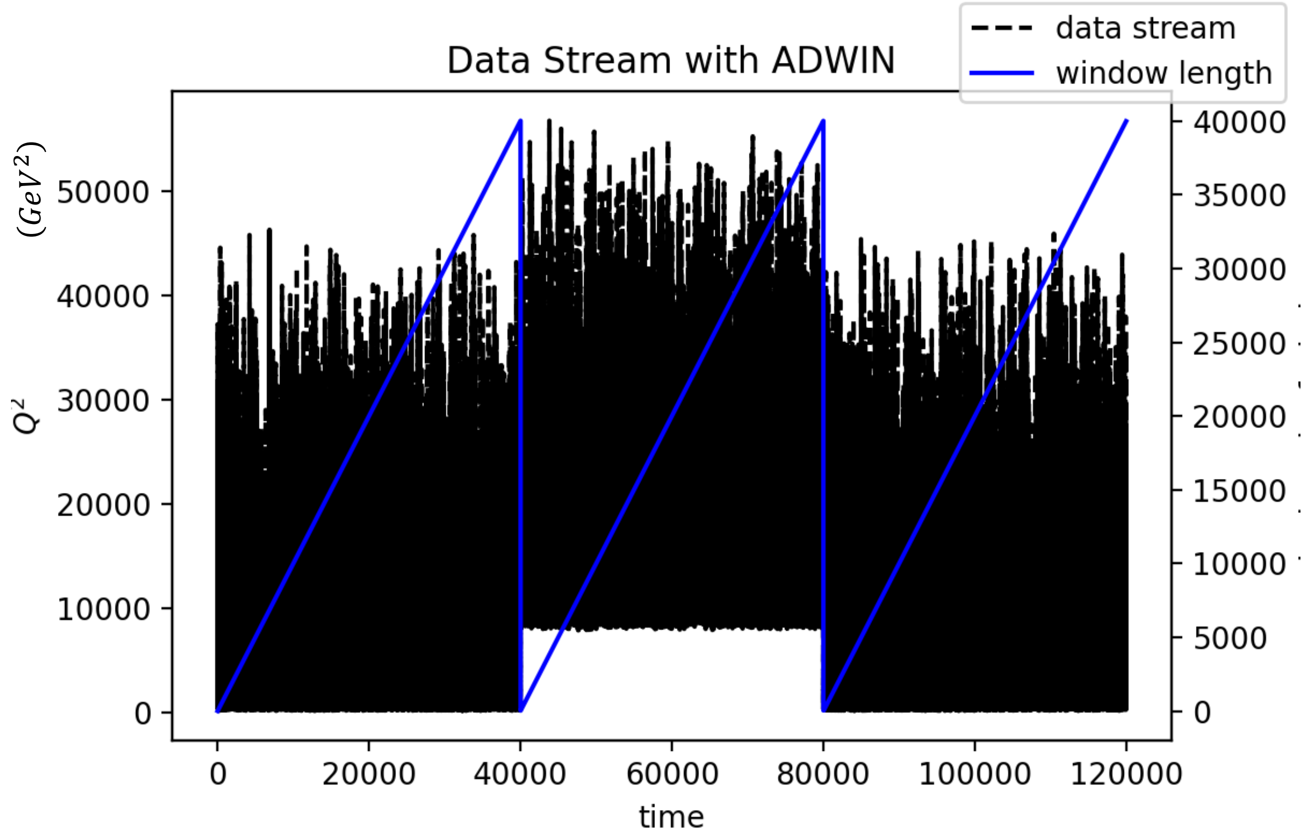
We subdivide the stream into 3 data-taking periods of equal parts and apply a constant shift of two standard deviations to each (x, y, z) position and energy E measurements in the second data taking period.

Change detection

ADWIN is an **ADaptive WINdowing technique** used for detecting distribution changes, concept drift, or anomalies in data streams with established guarantees on the rates of false positives and false negatives

(A. Bifet and R. Gavalda, *Learning from time-changing data with adaptive windowing*, in Proceedings of the 2007 SIAM international conference on data mining, SIAM, 2007, pp. 443–448)

Data Period	Start Time	Time ADWIN Detects Change
2	40000	40020
3	80000	80012



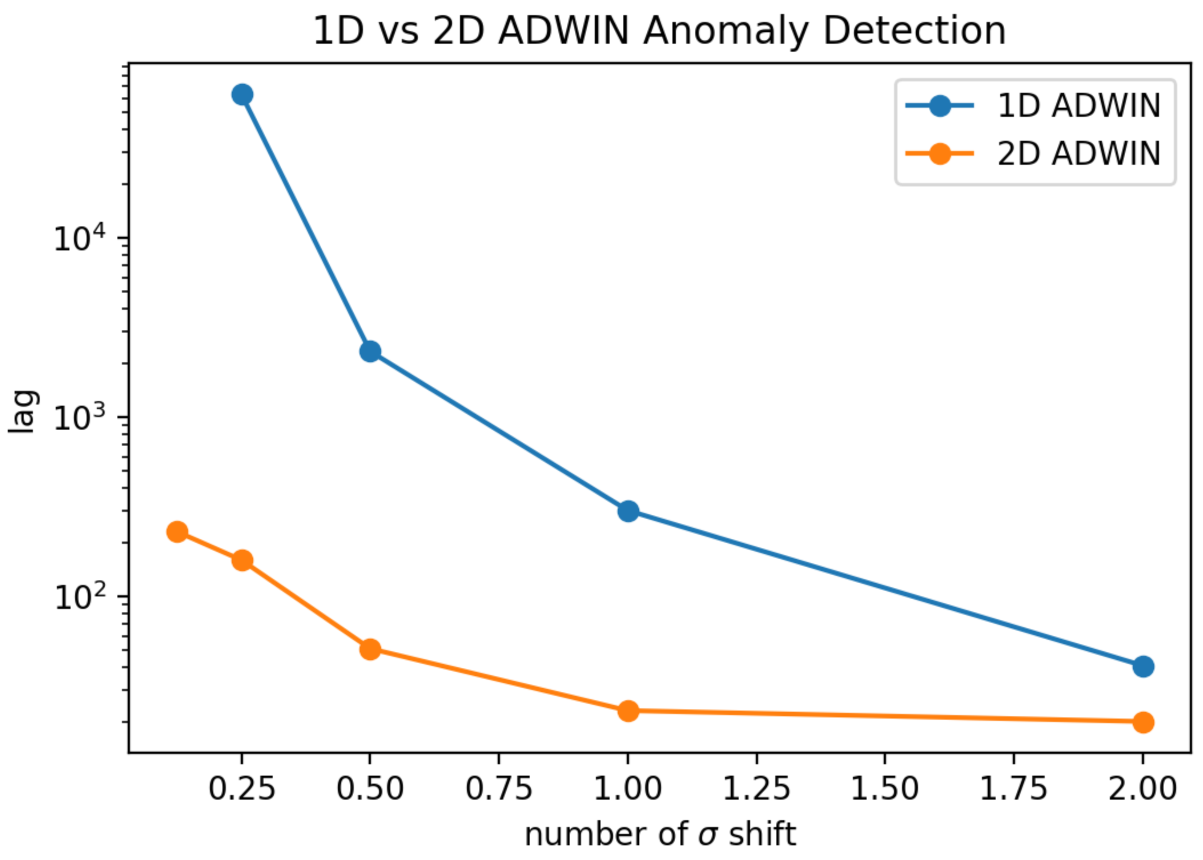
Jefferson Lab

Improving ADWIN

A higher-dimensional extension of ADWIN improves its ability to find changes in the data distribution.

Two cases:

- 1D: only use information from Q^2
- 2D: use information from (x, Q^2)

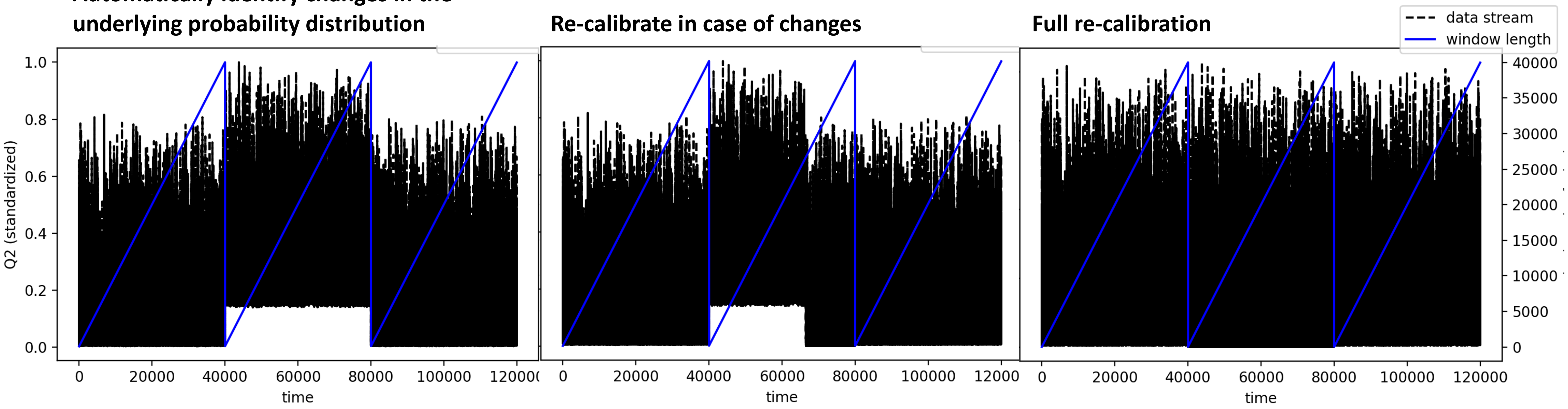


Calibrating each data-taking period to baseline

Automatically identify changes in the underlying probability distribution

Re-calibrate in case of changes

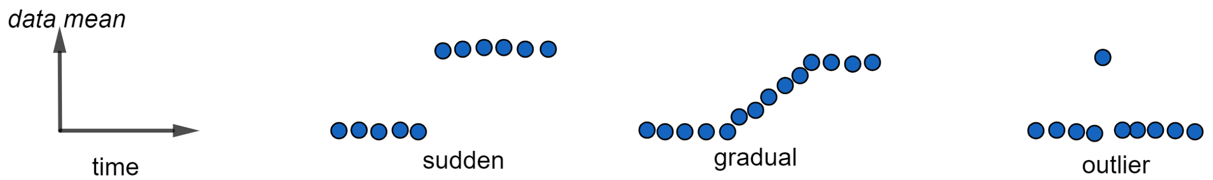
Full re-calibration



Hoeffding's Inequality For a confidence level of 0.01 and a margin of error of 0.01, a minimum sample of 26492 observations is needed to estimate of the mean in each data-taking period.

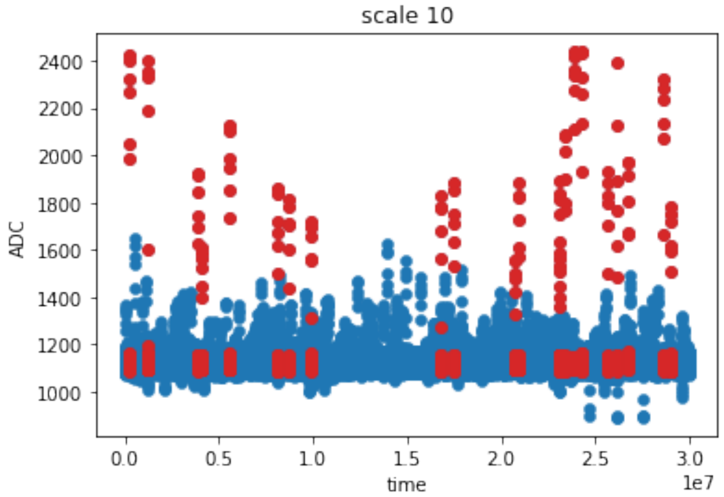
Multiscale method

- ADWIN2 suitable for detection of sudden changes but not gradual changes



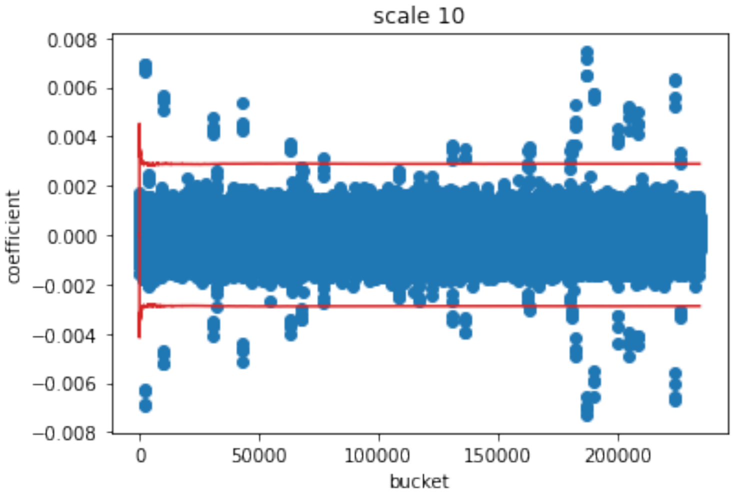
- Apply **multiscale method** for detection of sudden changes and gradual changes
  - Various test functions for various changes
  - **Fast algorithm for signal analysis**
    - Represent data in multiscale basis:
    - Base coefficients close to zero → No change.
    - Base coefficients far away from zero → Change

Reference Z. Chen, C.A. Micchelli, C., Y. Xu, *Multiscale Methods for Fredholm Integral Equations*, Cambridge Monographs on Applied and Computational Mathematics (2015)



Transform to **coefficient space**:

- Outliers in the distribution → Change.
- **Change detection: Detect outliers**
  - e.g., using IQR



Jefferson Lab

Online Multiscale Method

$$\underbrace{d_0, d_1}_{a_0^1}$$

$$\overbrace{d_0, d_1, \underbrace{d_2, d_3}_{a_2^1}}^{a_0^2}$$

$$d_0, d_1, d_2, d_3, \underbrace{d_4, d_5}_{a_4^1}$$

$$\underbrace{d_0, d_1, d_2, d_3, \overbrace{d_4, d_5, \underbrace{d_6, d_7}_{a_6^1}}^{a_4^2}}_{a_0^3}$$

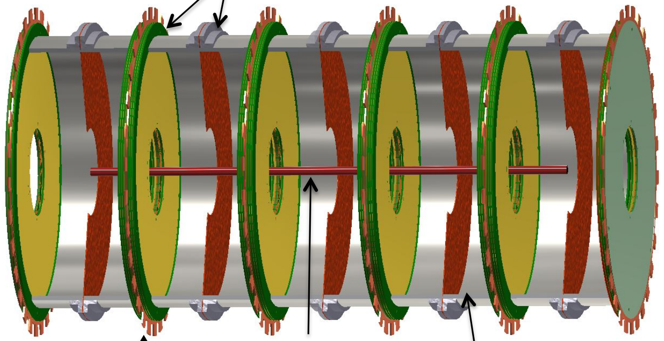
For the sequential data, we begin with the small scale, and when we obtain more data, we increase the scale. The design of the algorithm is suitable for monitoring the sequential data.

If the change detected in a small scale, then we can conclude the change happens in a small set, which means the detection results is more accurate. So we start with accurate scale, and then detect for next scale.

TDIS Streaming Readout Prototype

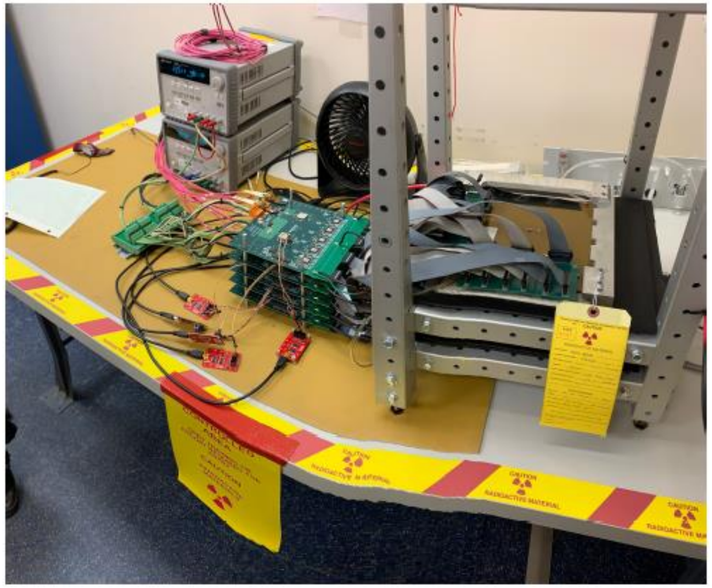
Tagged Deep Inelastic Scattering (TDIS)

- **Hall A** Super Big-Bite (SBS)
- **measurement** tagging for meson structure via the Sullivan process
- **science goals** meson structure functions and PDF
- **detection of low momentum spectators** GEM based multiple TPC (mTPC), reduced drift time in mTPC allows for triggered or streaming readout

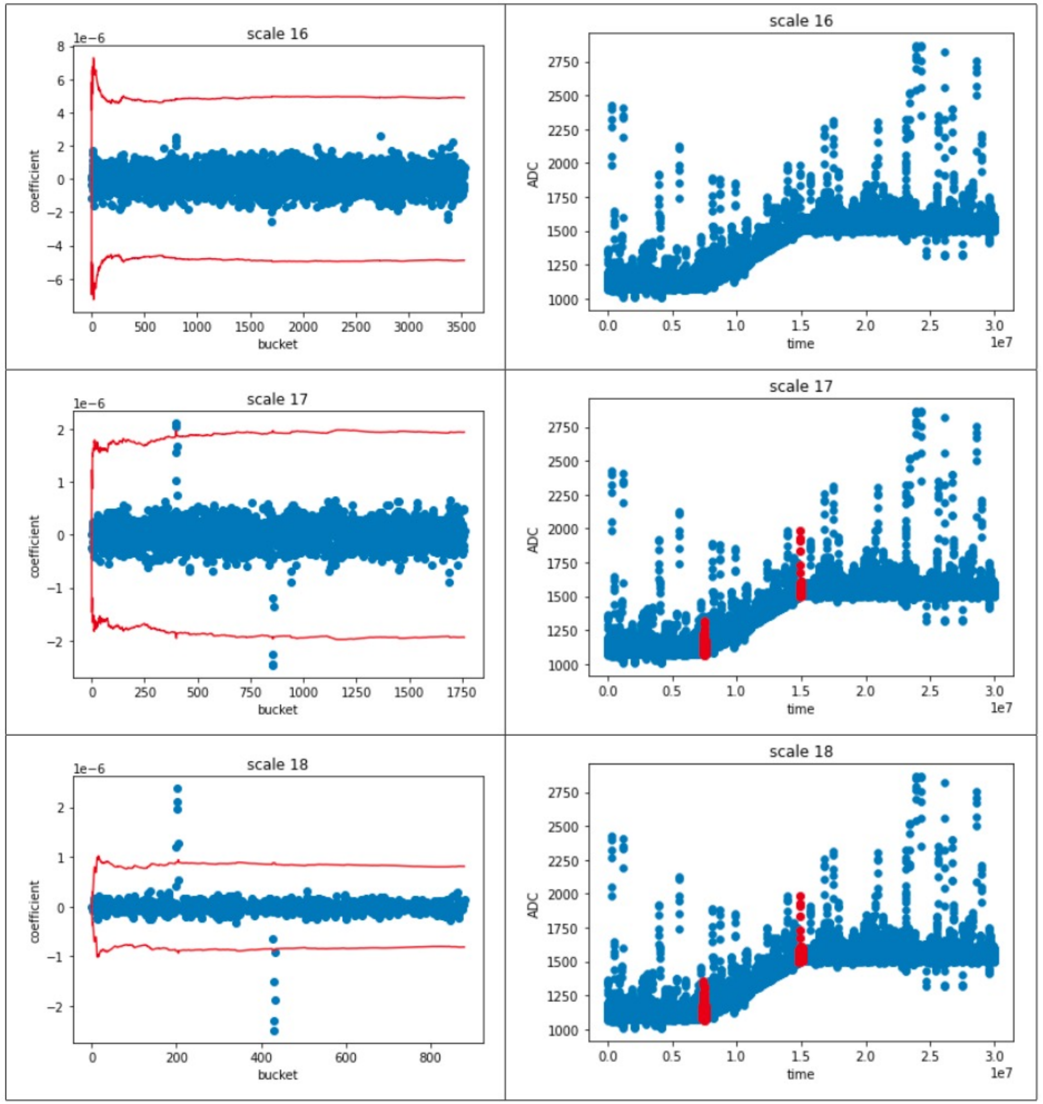


TDIS Streaming Readout Prototype

- **SAMPA** novel front-end ASIC developed for streaming readout of GEM based ALICE TPC
- **ongoing tests** study GEM pulse data and stream continuously
- **preliminary results** stream trigger-less GEM data (768 channels) in DAS and DSP modes at 45 Gb/s via 5 ALICE front-end cards (FECs)
- **next steps**
  - integrate FELIX hardware and software into streaming system

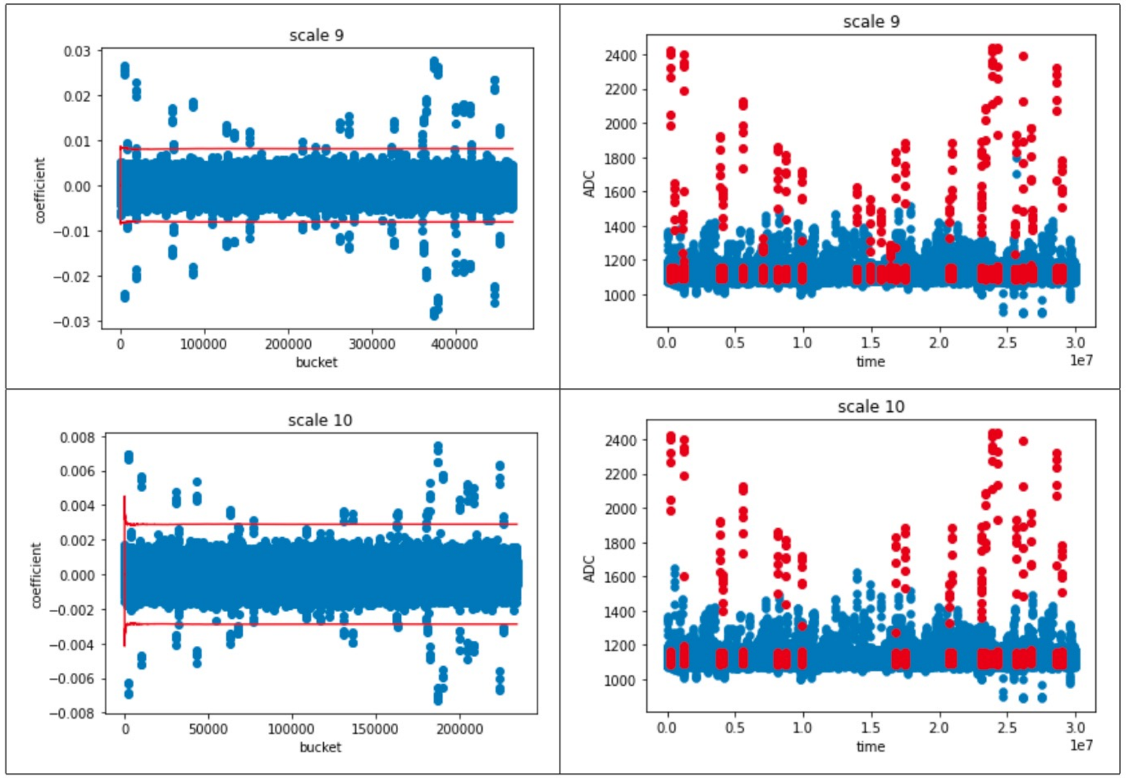


Monitoring the TDIS Streaming Readout Prototype



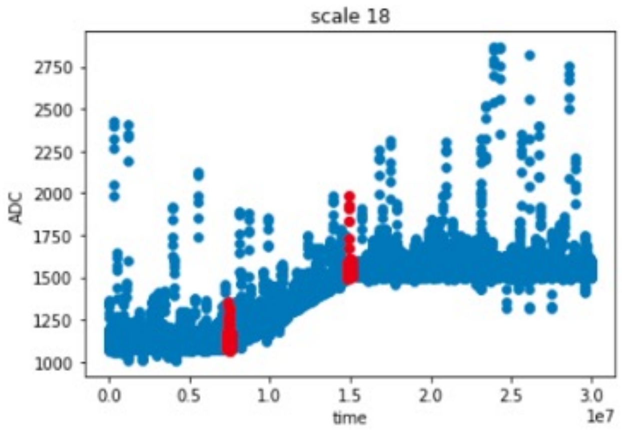
Multi Scale Method: **Various test functions for various changes**

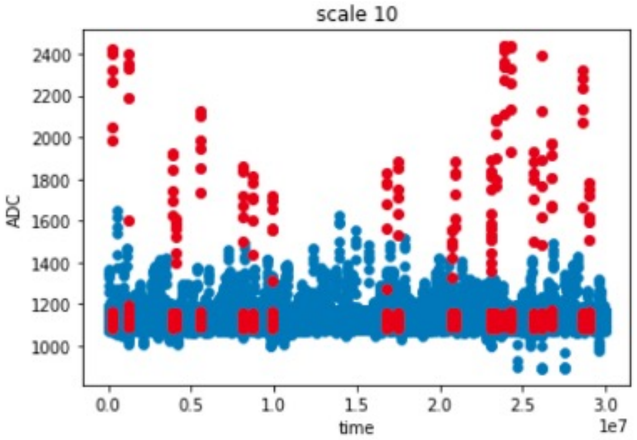
- Represent data in multiscale basis:
  - Increase of base coefficients → Change.
- Transform to **coefficient space**:
  - Outliers in the distribution → Change.
- Detect Changes → **Detect outliers using IQR**, symbolized in **red**.



Apply multiscale method

- Use **multiscale method** at various scales for change detection





- Analyze results from **multiscale method** (in **red**) and decide on response
  - Issue alarm
  - Restart calibration
  - Start user-defined process

Markus Diefenthaler

mdiefent@jlab.org

INDRA-ASTRA

- Part of R&D on streaming readout and AI/ML
- Work towards prototype for a fully automated, responsive detector system

Status

- Developed method for autonomous calibration of DIS experiments using baseline calibrations and autonomous change detection.
- Developed ADWIN2 and multiscale methods for autonomous change detection.
- Versatile multiscale method can be used to increase reliability of data and find and fix issues on time.

Next steps

- Test multiscale method in running SBS experiment