AI for Experimental Control

Torri Jeske | roark@jlab.org | SRO-XII | 12.3.2024

Jefferson Lab EPSCI Carnegie Mellon University

Certain aspects of experiments are costly and error prone.

Calibration

Usually done after data taking, with a timescale on the order of months to years. As a result, there is a *significant* delay between data collection and publication.

Subsystem operation

Can require human attention and manual intervention.

Detector/Data Monitoring

Requires *constant* human attention.

Can we develop and *deploy* an AI system to autonomously adjust detector controls during data acquisition in order to reduce or eliminate the need for offline calibrations?

AI for Experimental Controls Proposal

Main objectives

Uniform gain

Can we train a model to determine calibration constants as quickly as possible? From those calibration constants, can we recommend a HV setting to stabilize the chamber gain?

Reduced expert time

Reduce the number of offline iterations required for satisfactory dE/dx and timing resolution

Extend to other detector systems

Build an overall control system that can be used for other detector components

Challenges with deploying ML based software

Higher technical risk

How to define “good enough” performance? What model architecture to use? Variable data needs and quality? Limitations of AI/ML based solutions

Expanded skill set/workforce

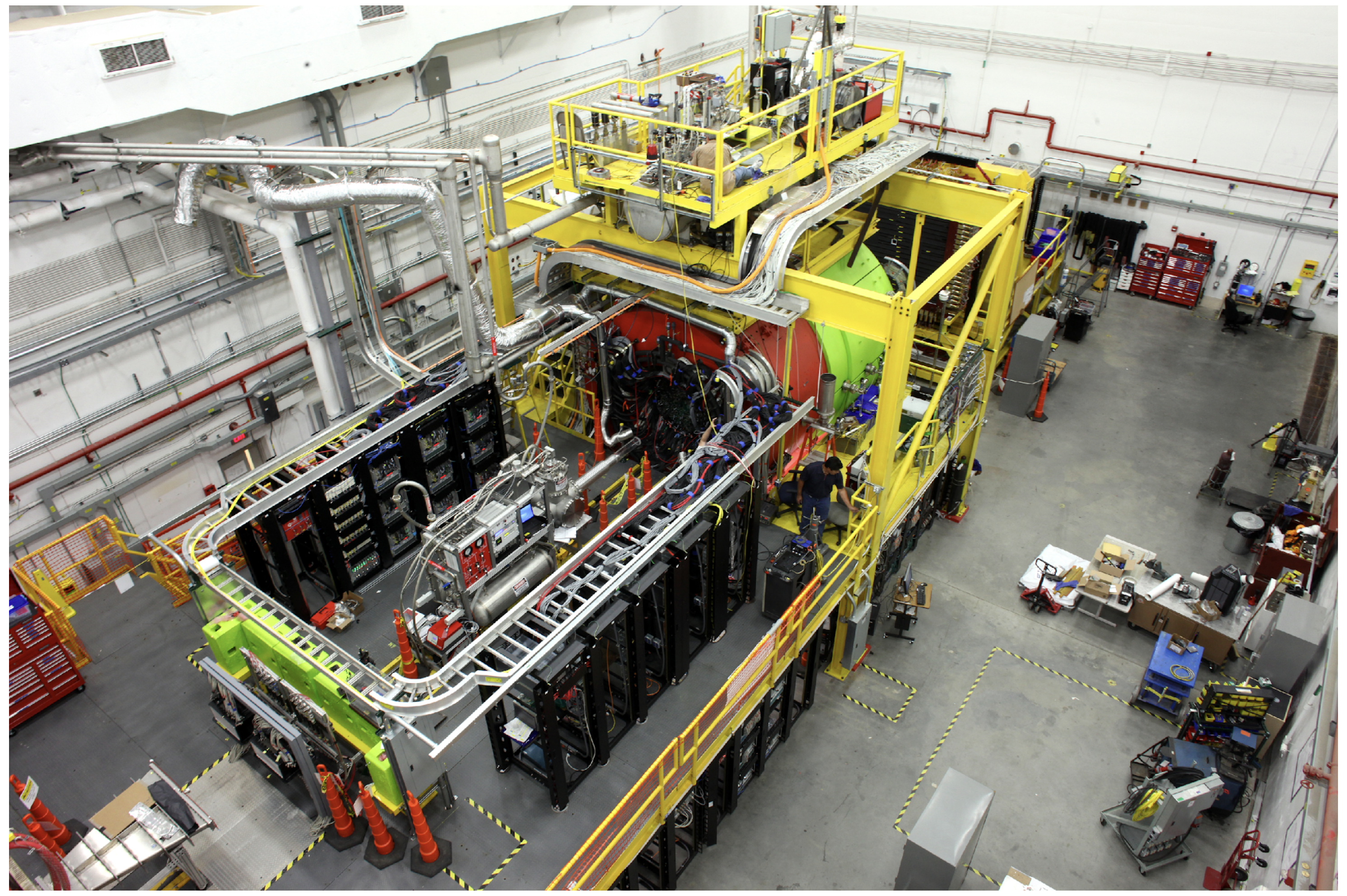
Introducing ML based solutions for NP tasks requires a broader set of skills and team members

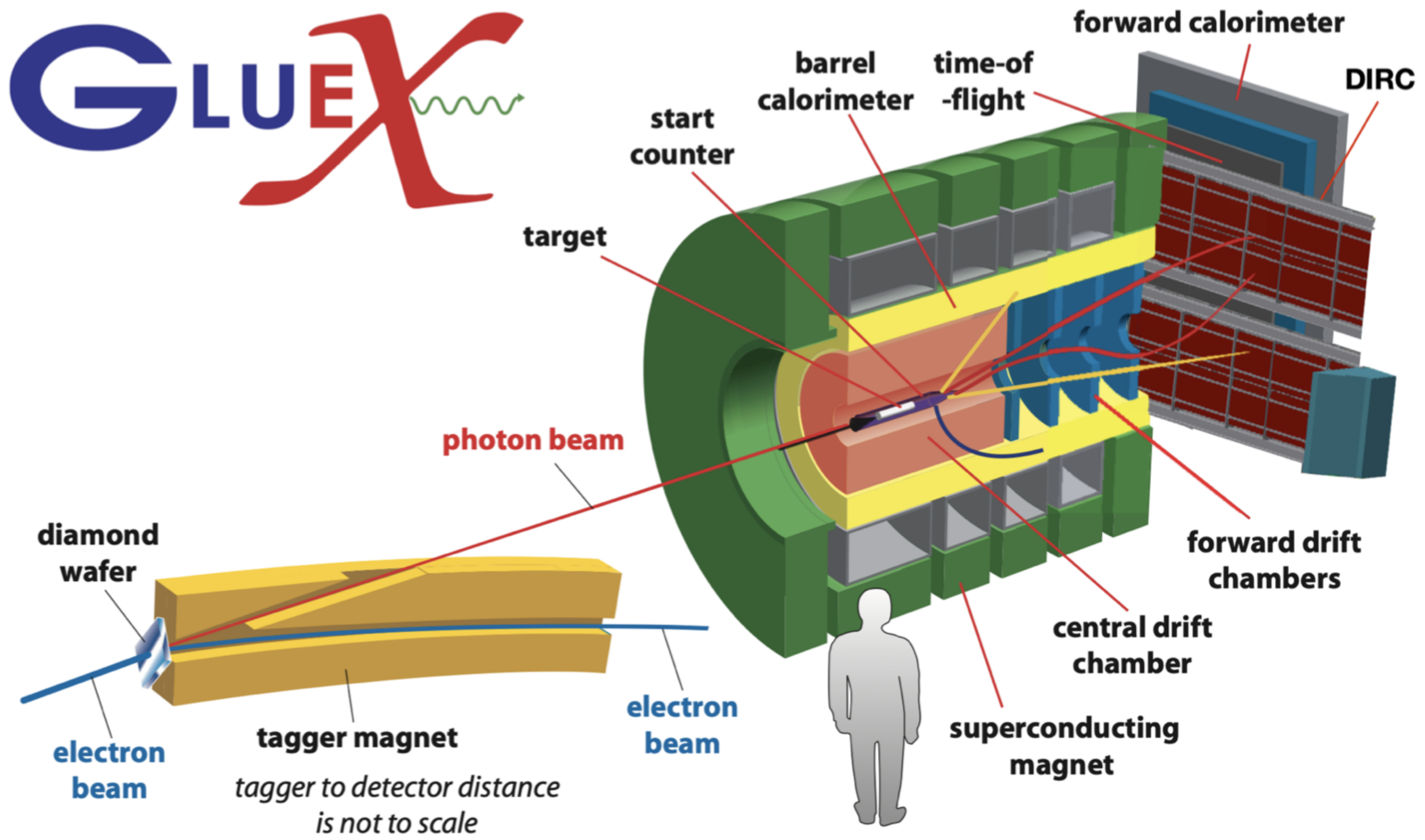
Require change

The control system we develop would change the standard operation of the detector

Gluonic Excitations Experiment

Located in Hall-D at Jefferson Lab, GlueX was designed to search for and measure exotic hybrid mesons produced in photo-production reactions





GlueX Central Drift Chamber

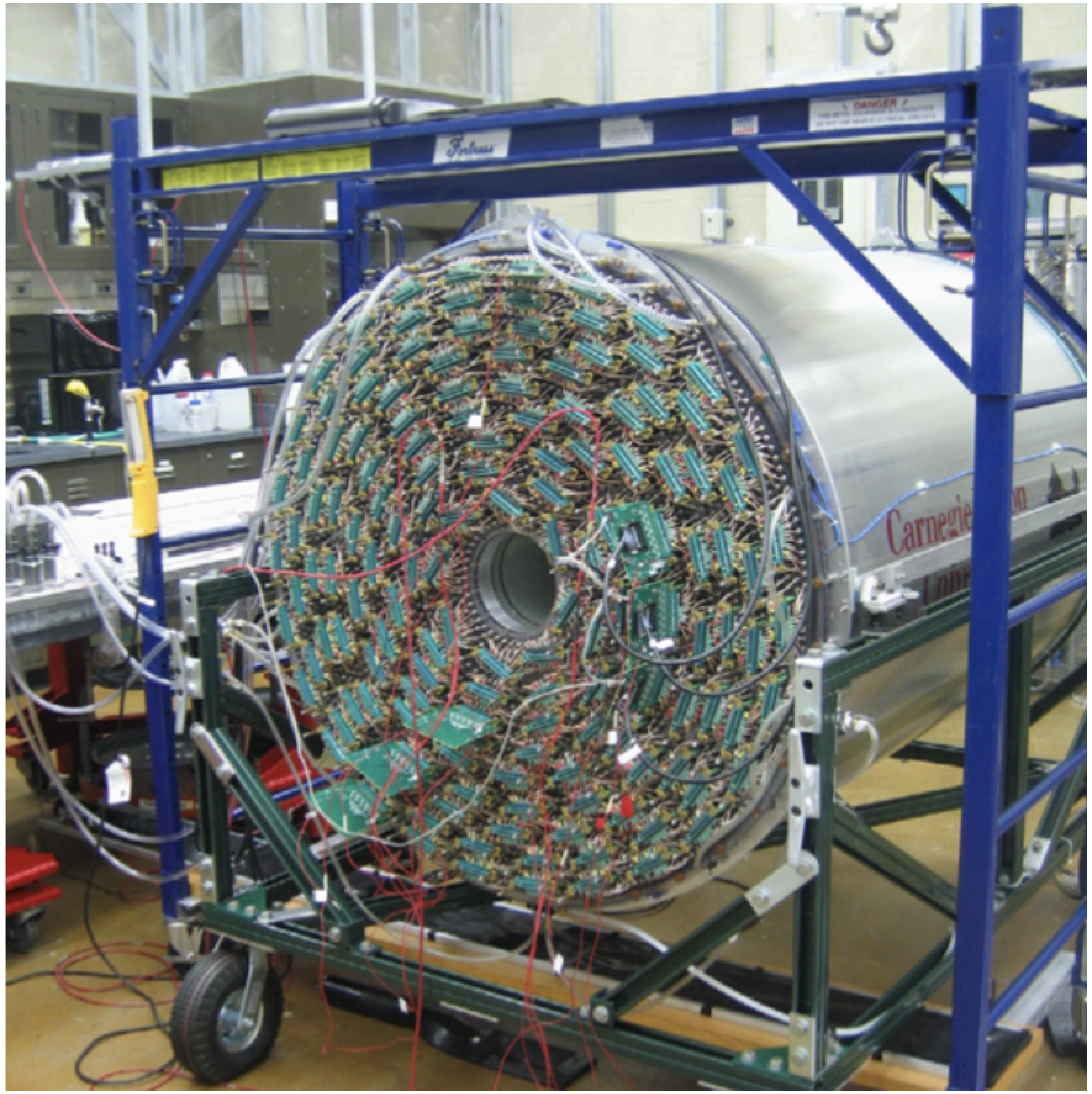
Used to detect and track charged particles as they traverse the detector

Specs:

- 1.5 m long x 1.2 m diameter, cylindrical straw tube chamber
- 3522 anode wires traditionally held at 2125 V
- 50:50 Ar:CO2 gas mixture at 30 Pa above atmospheric pressure

Calibrations (run-by-run):

- Chamber gain
- Drift time to drift distance

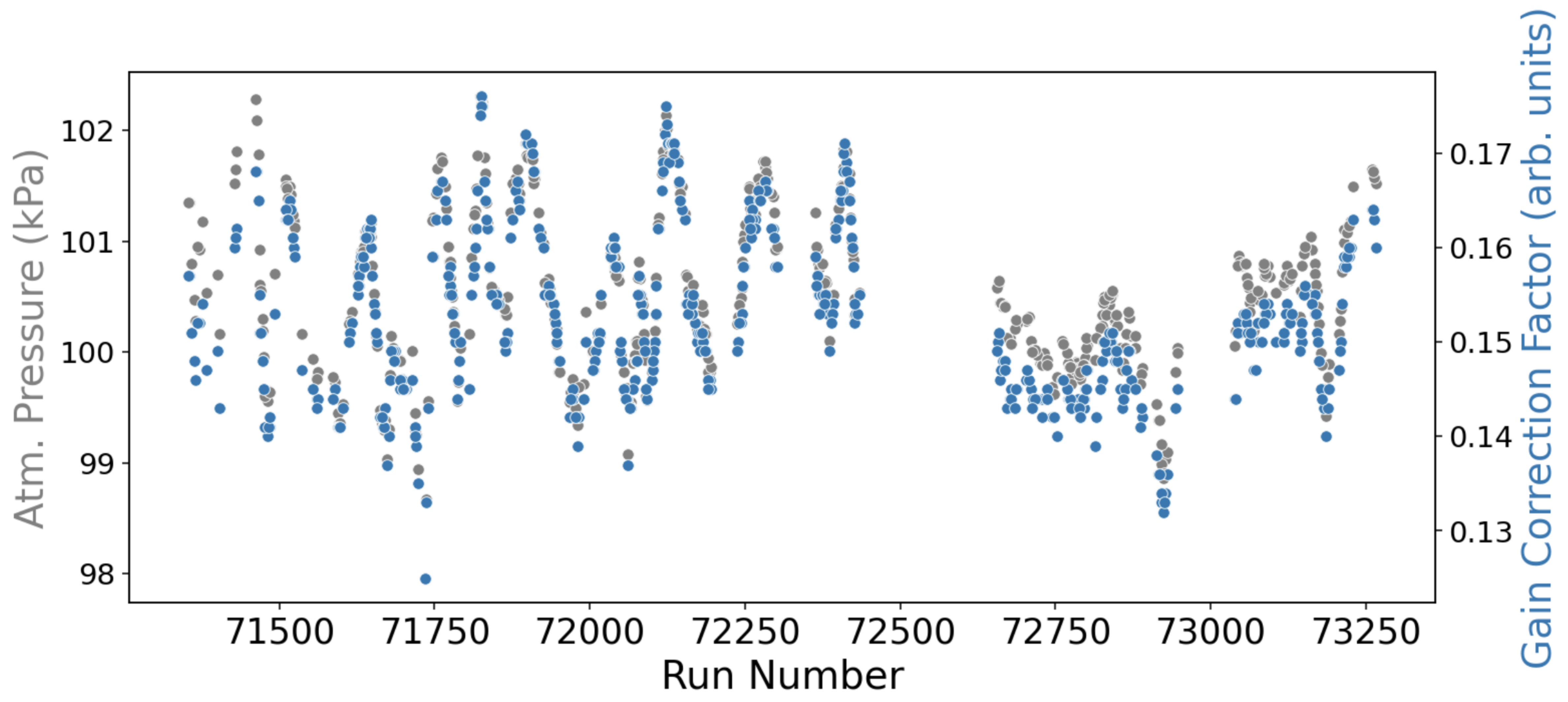


Jefferson Lab EPSCI Carnegie Mellon University

Chamber gain

The gain determines the size of the avalanche and, therefore, the height of the pulse recorded.

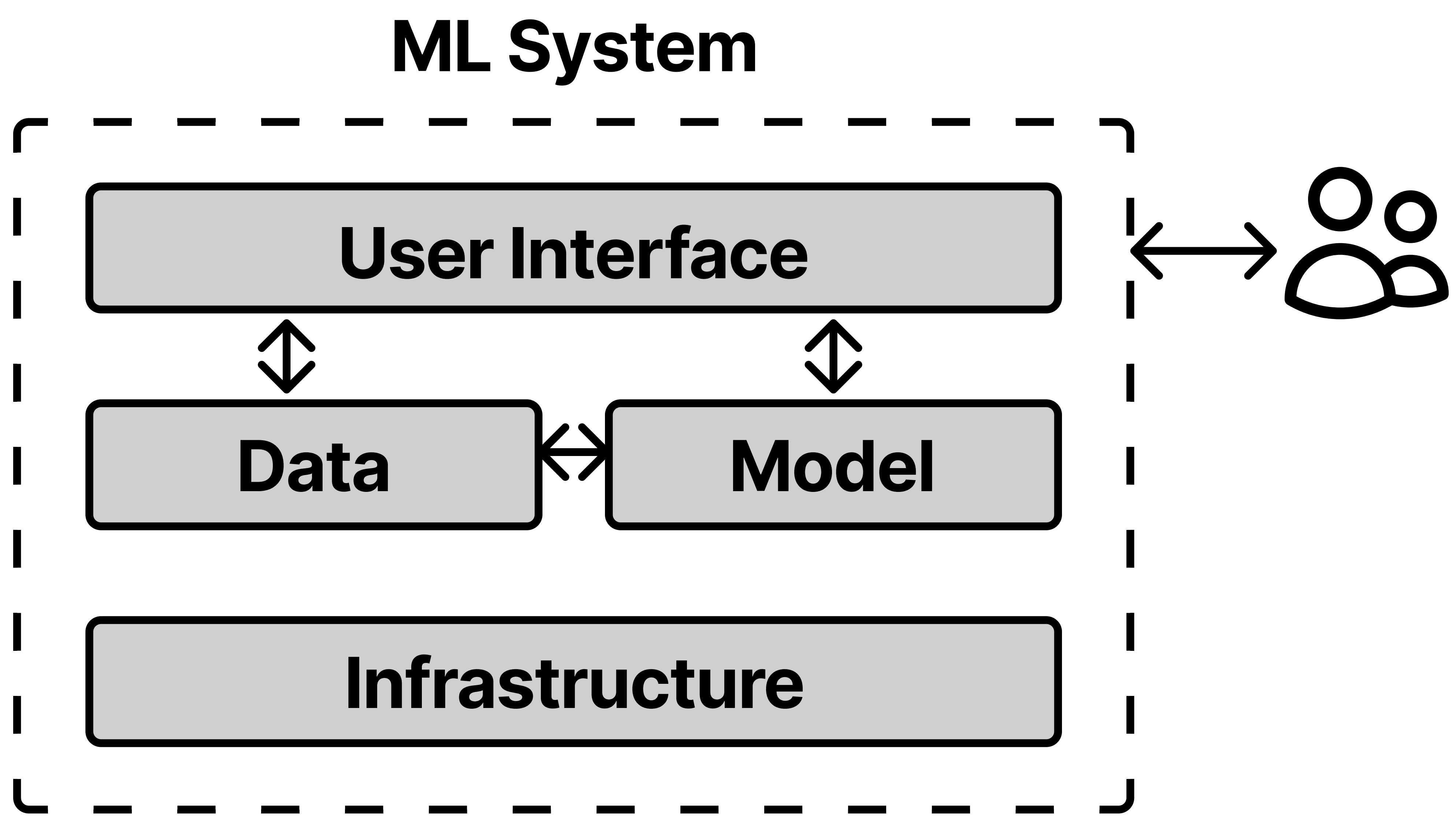
This affects both the measured amplitude used in particle identification and the measured drift time used to determine the particle's momentum.



ML Control System Design

There are several design choices and requirements that determine what the ML system might look like

For our use case, we want quick inference times and readily accessible input features. For monitoring purposes, we want to clearly convey input features, inferences, and control decisions to the shift crew and detector experts.



ML system status

601 **Observations**

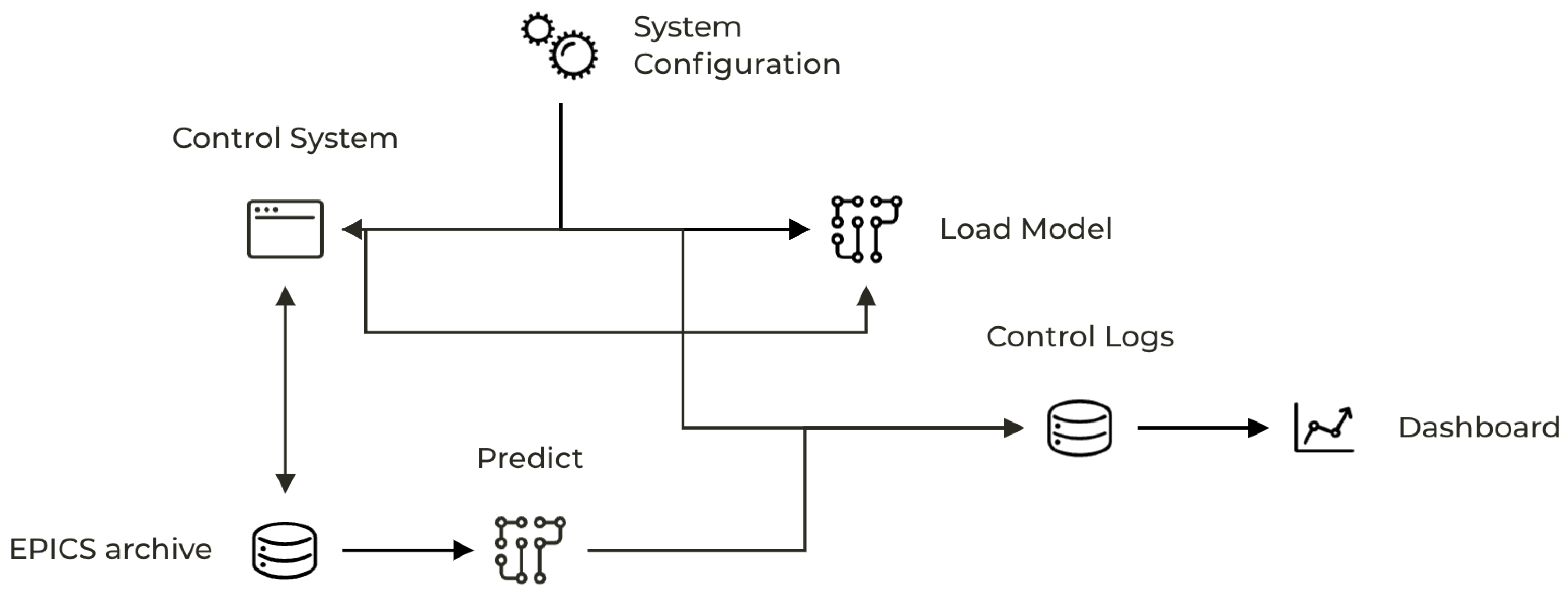
Balanced for High, Medium, and Low pressure runs
80/20 train test split

5 **Model architectures**

LR, ANNs, XGBoost, GPs
Offline training
Online inference

5 **Iterations of control system**

Actively maintained by EPSCI group



Architectures + Performance

GlueX has an extensive calibration and conditions database which can be used for training models to predict calibration values

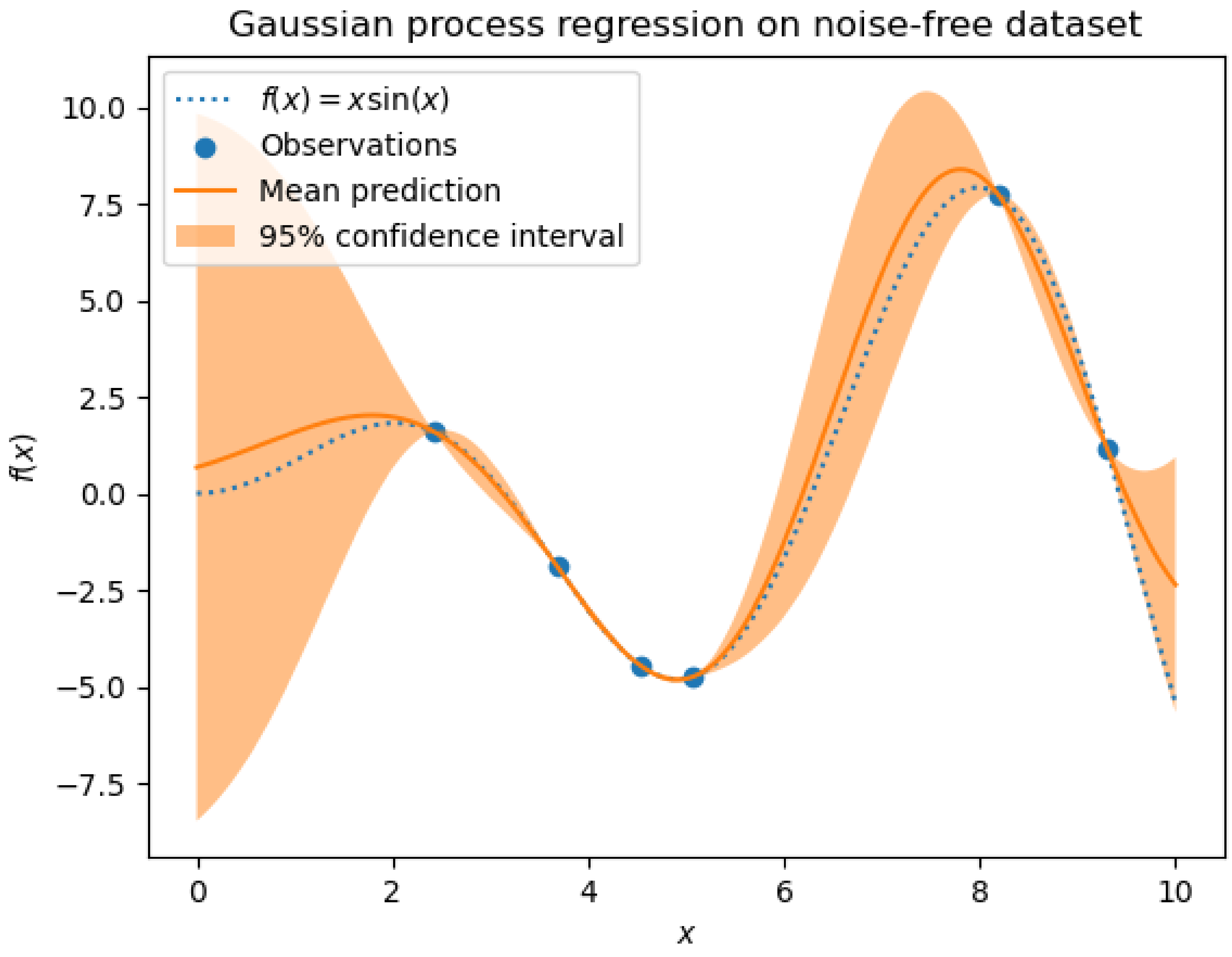
Initial feature set included slow controls data, reconstructed quantities, and engineered features

Model type	Run period(s)	# features	MAPE	Max % err	Ratio > 1% err
LR	2018 & 2020	11	1.3%	19.1%	97/164
LR	2020	11	0.7%	2.0%	30/106
LR	2018 & 2020	5	2.3%	20.3%	96/164
LR	2020	5	0.7%	2.6%	26/106
RF	2018 & 2020	82	1.7%	18.5%	83/164
ANN-7 layers	2018 & 2020	122	1.8%	11.4%	74/164
ANN-4 layers	2018 & 2020	122	1.9%	10.8%	84/164
ANN-3 layers	2018 & 2020	122	1.9%	11.9%	90/164
XGBoost	2018 & 2020	122	1.6%	10.2%	76/164
XGBoost	2018 & 2020	82	1.4%	11.8%	68/164
XGBoost	2018 & 2020	71	1.6%	11.2%	71/164
XGBoost	2018 & 2020	12	1.8%	11.1%	72/164
GPR	2018 & 2020	25	1.7%	10.9%	80/164
GPR	2018 & 2020	14	1.5%	9.7%	66/164
GPR	2018 & 2020	11	1.5%	10.1%	72/164
GPR	**2020**	**11**	**0.5%**	4.1%	**17/106**
GPR	2018 & 2020	5	1.5%	9.1%	70/164
GPR	2020	5	0.7%	3.6%	28/106
GPR	2018 & 2020	3	1.8%	12.0%	74/164
GPR	**2020**	**3**	0.7%	**1.9%**	25/106
GPR	2018 & 2020	1	3.4%	18.6%	129/164
GPR	2020	1	1.1%	4%	49/106

Gaussian processes

Gaussian process was chosen based on

- Performance
- Quick training and inference times
- **“Out-of-the-box” uncertainty quantification**



Jefferson Lab EPSCI Carnegie Mellon University

Training

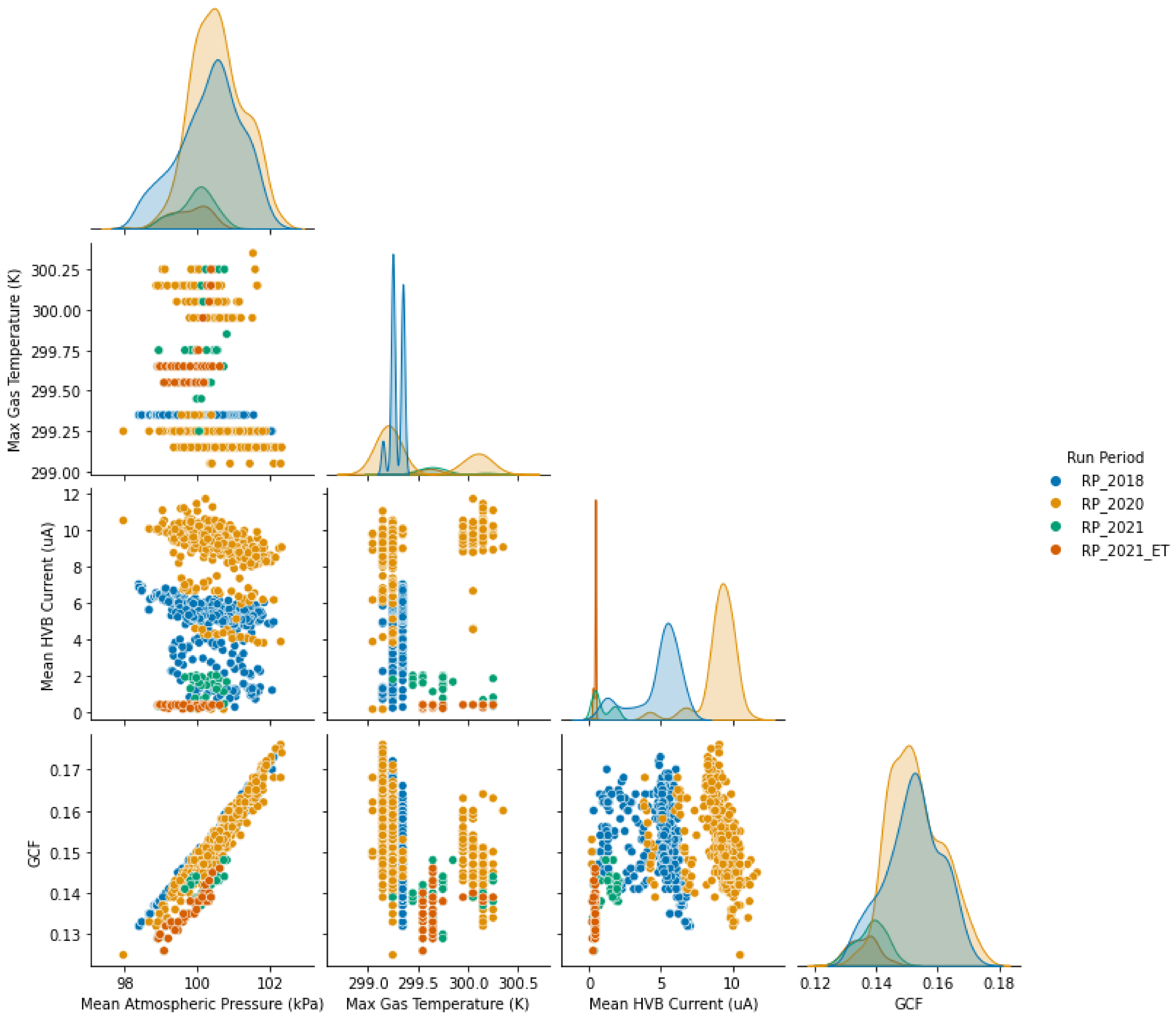
Gaussian process

Input features: environmental and experimental data from EPICS archive

- Gas temperature
- Current drawn from the HV boards
- Atmospheric pressure

Target: Gain calibration constants from previous experiments

Kernel: Radial basis function + White Noise



Jefferson Lab EPSCI Carnegie Mellon University

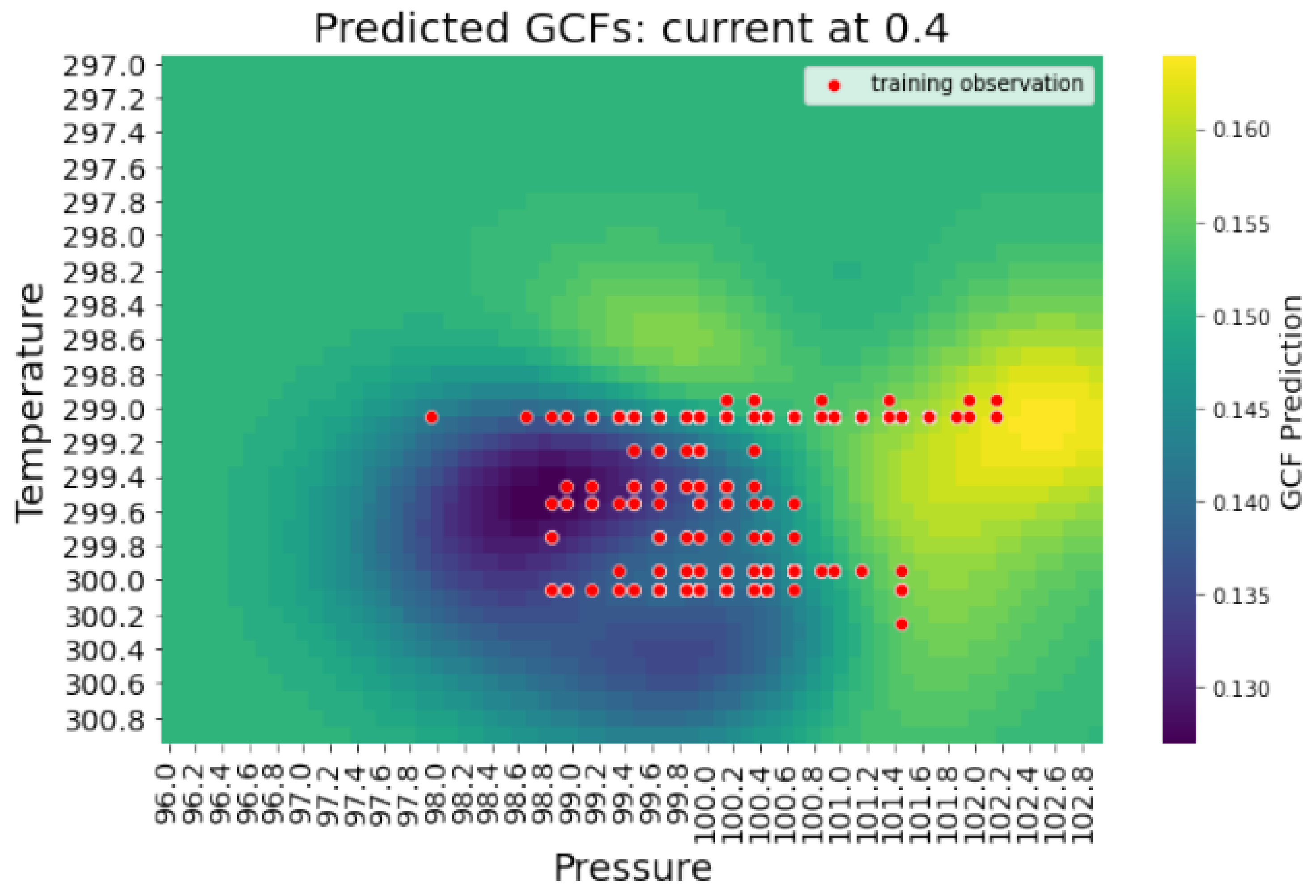
Control

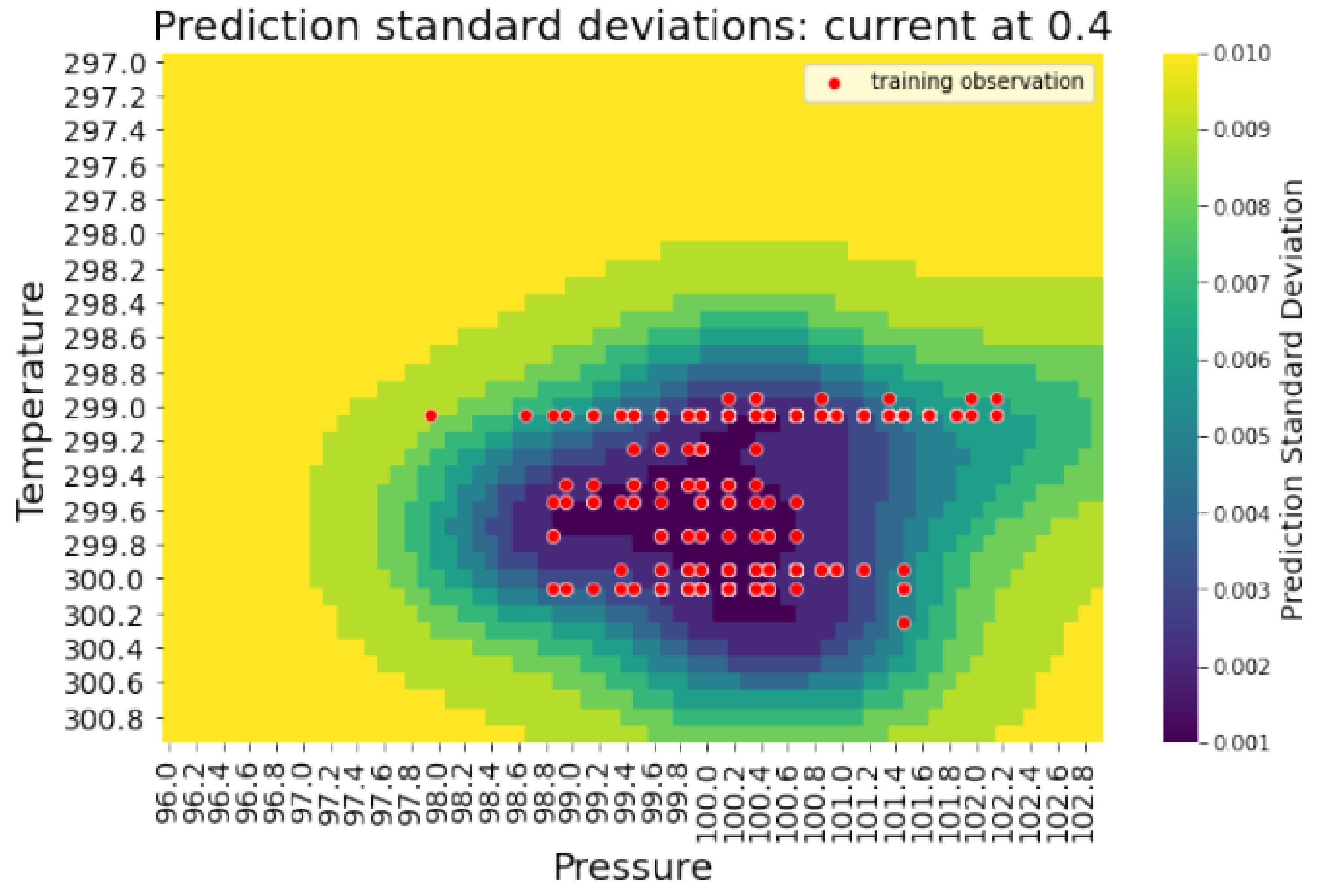
Policies

Having an uncertainty estimate on our predictions is a requirement

- Control policies
  - when model is uncertain
  - when HV setting is detected outside of allowed operational zone
- Informs when the model may need to be retrained

From an operations standpoint, we do not distinguish between *epistemic* or *aleatoric* uncertainty



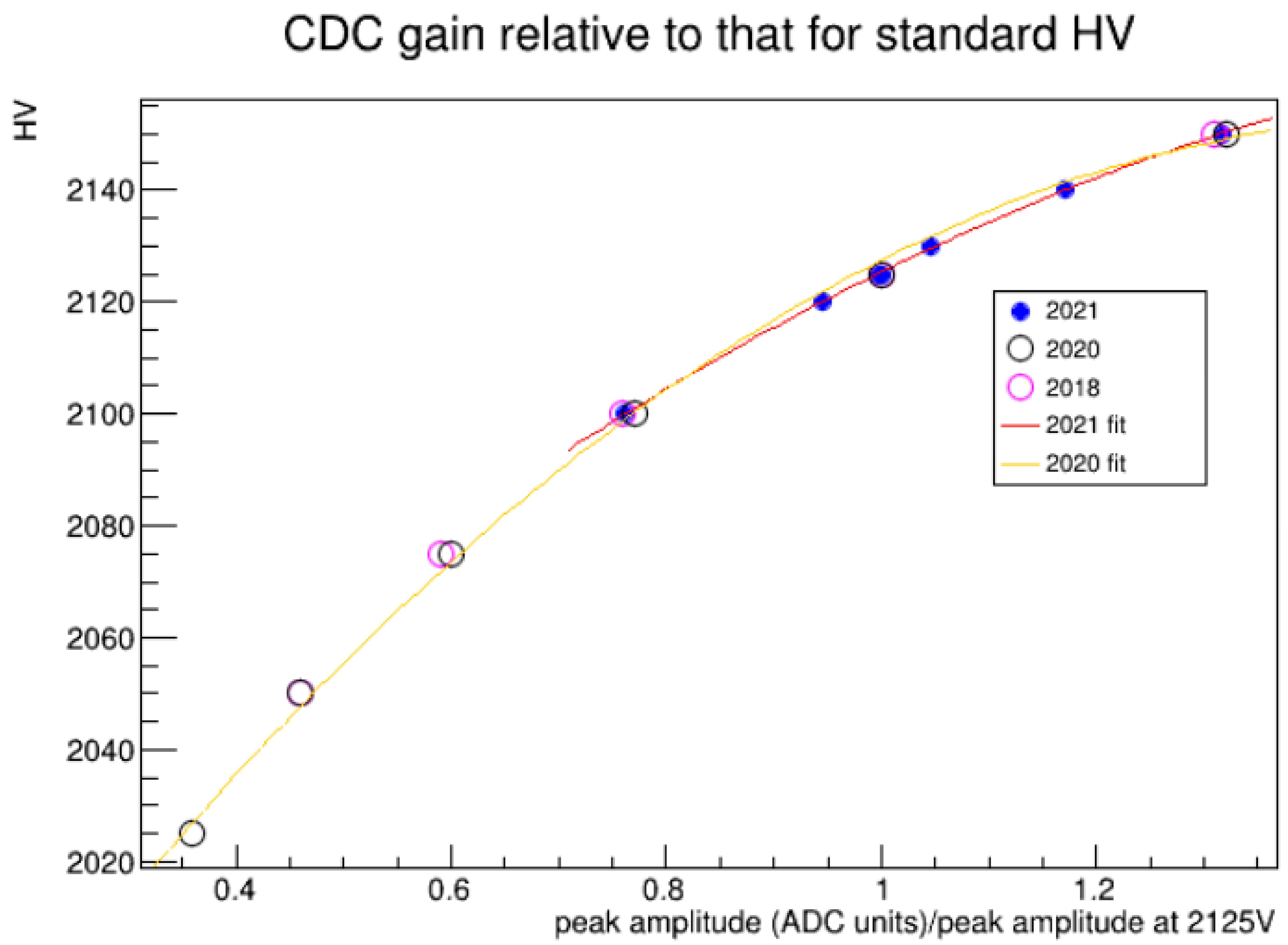


Control

High Voltage Recommendation

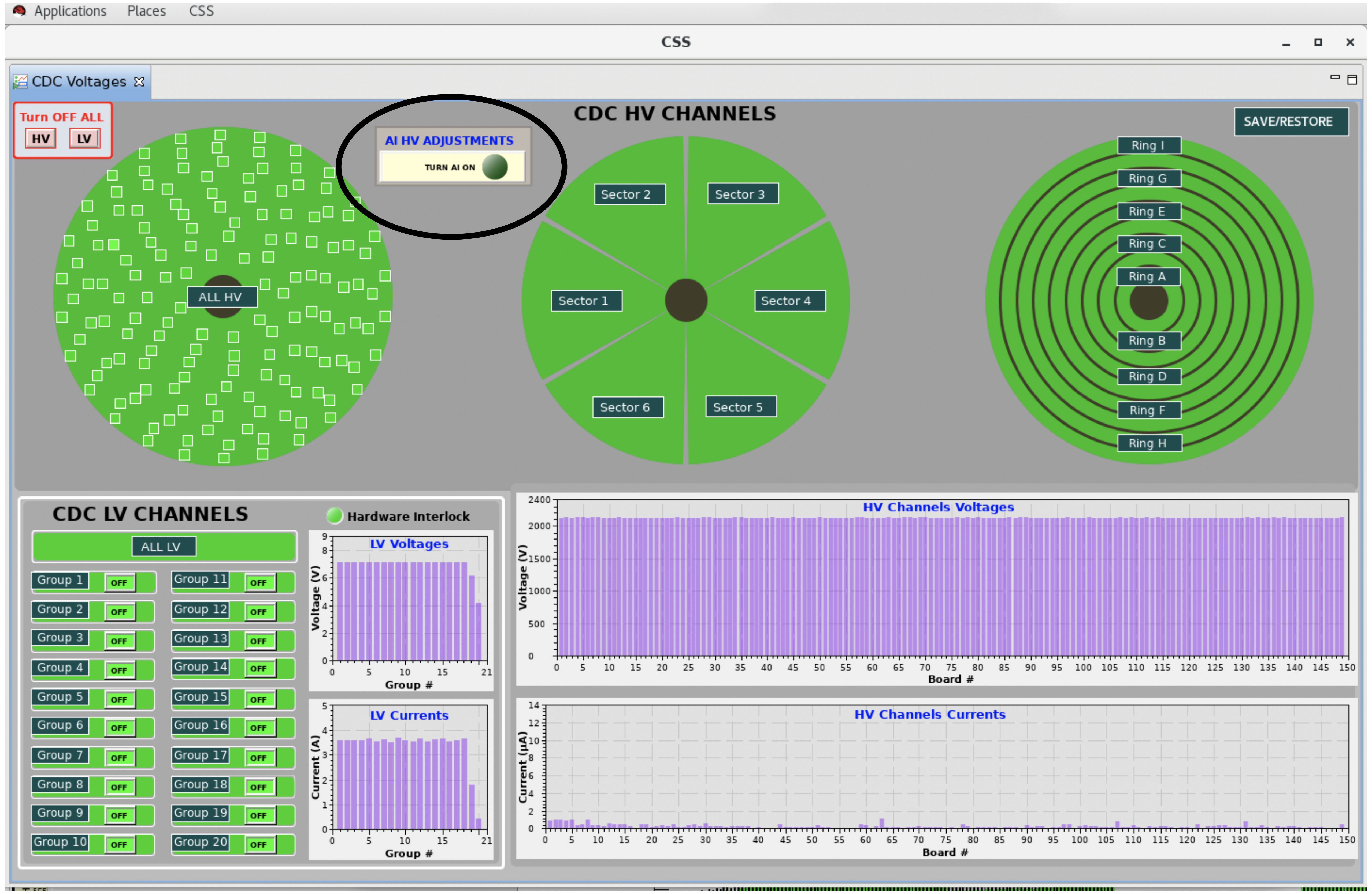
High voltage scans are conducted at regular intervals during each operational period.

This data is used to establish the optimal HV setting required to maintain chamber gain stability and performance.



Monitoring

CDC Control GUI



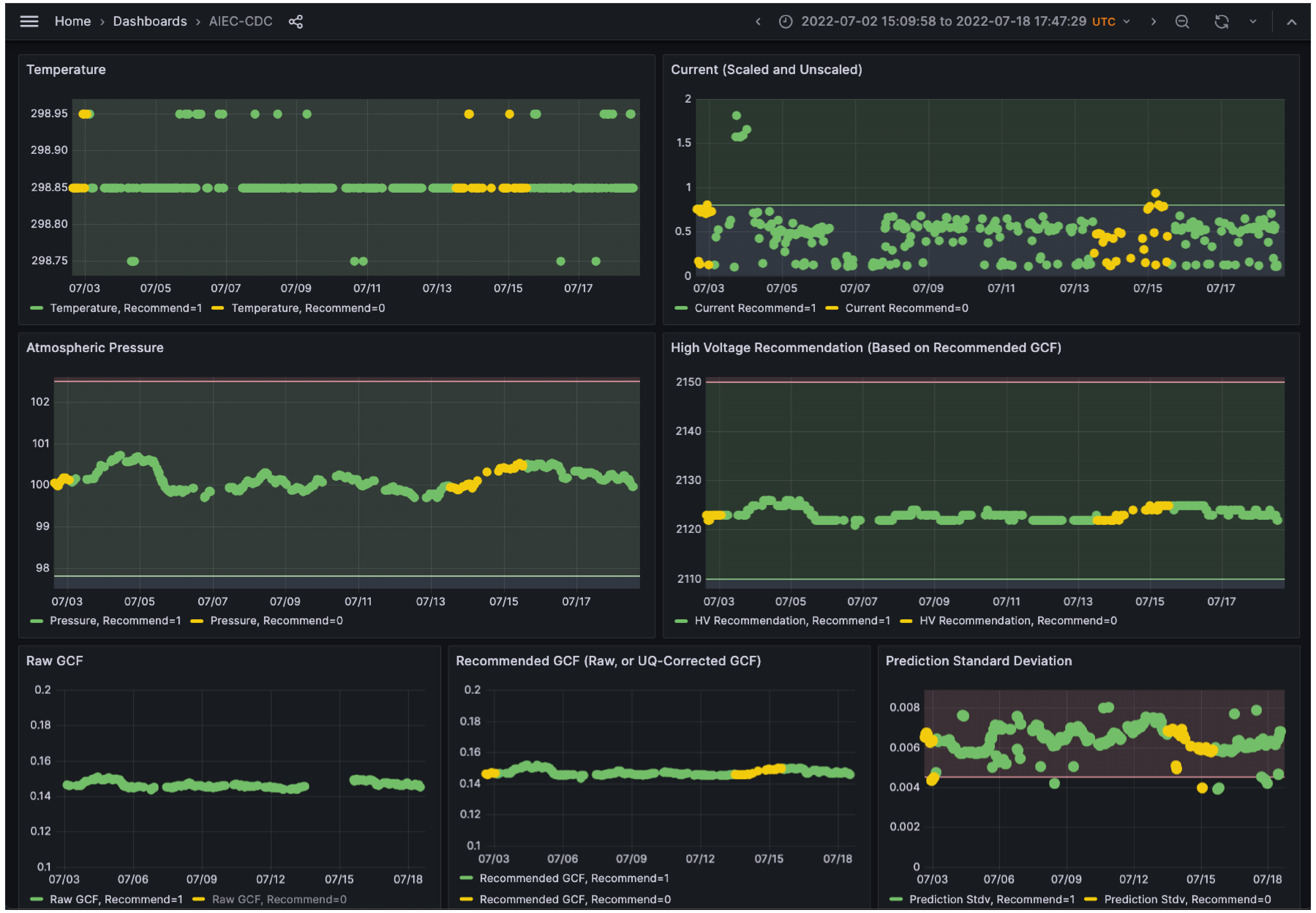
Shift takers are able to toggle the control aspect ON/OFF

Even when control is OFF, we record the actions the system *would* take.

Monitoring

Grafana Interface

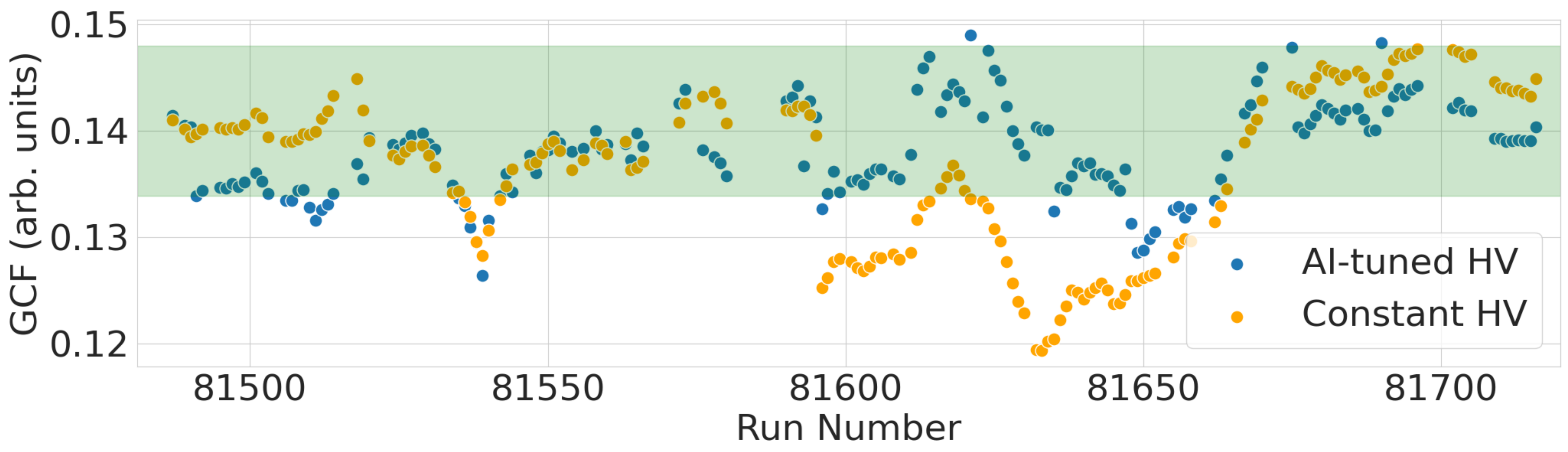
All input features, inferences, and *actions* are logged for further analysis regardless of whether control is ON or OFF

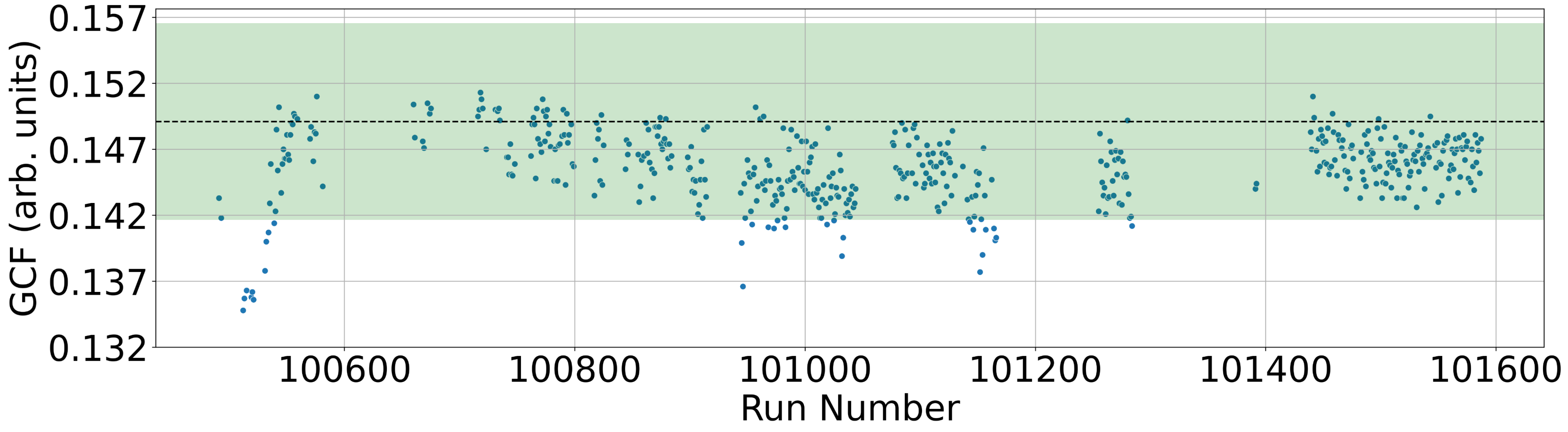


Jefferson Lab EPSCI Carnegie Mellon University

First deployments

5% threshold





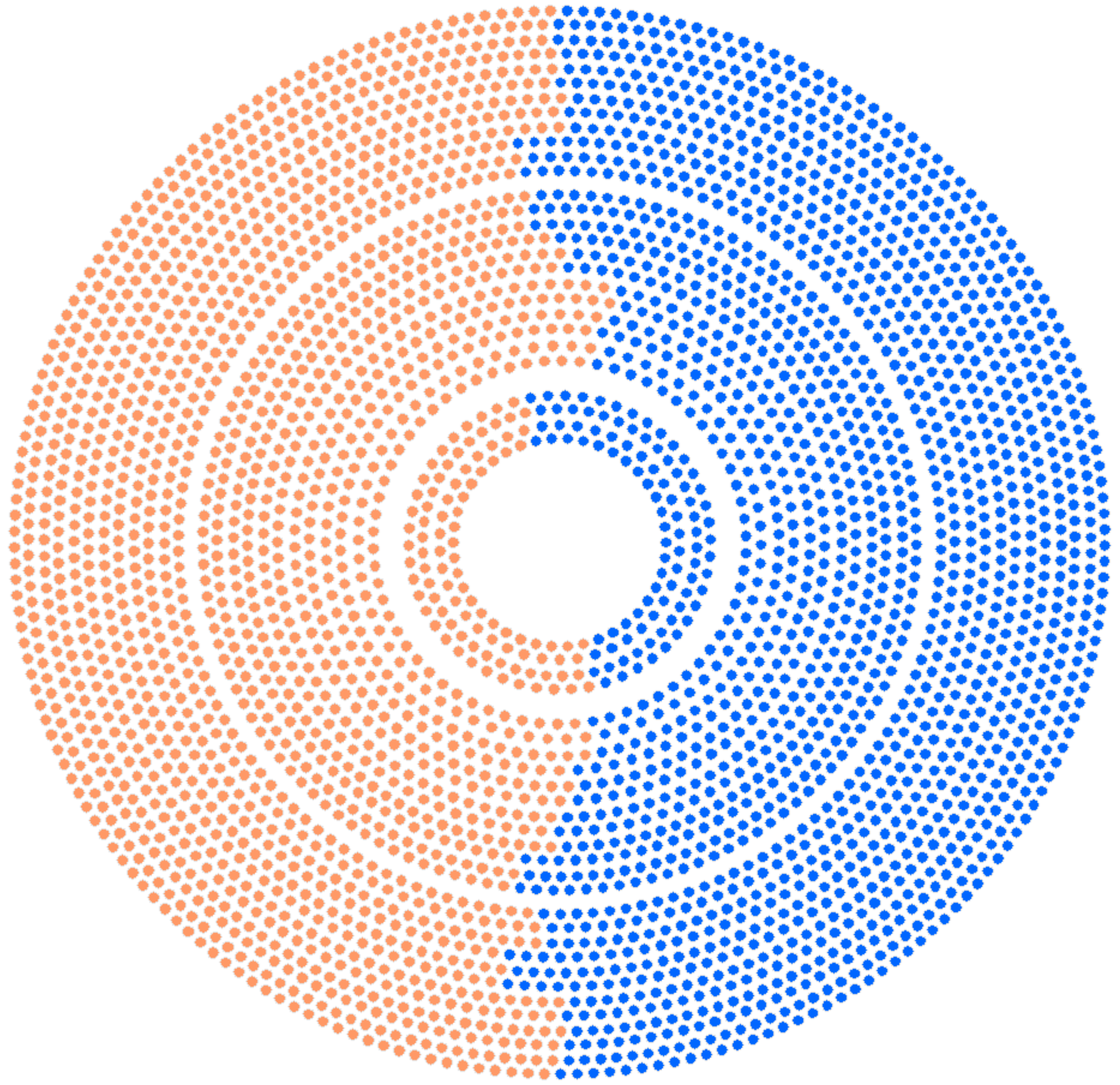
PrimEx - 2021

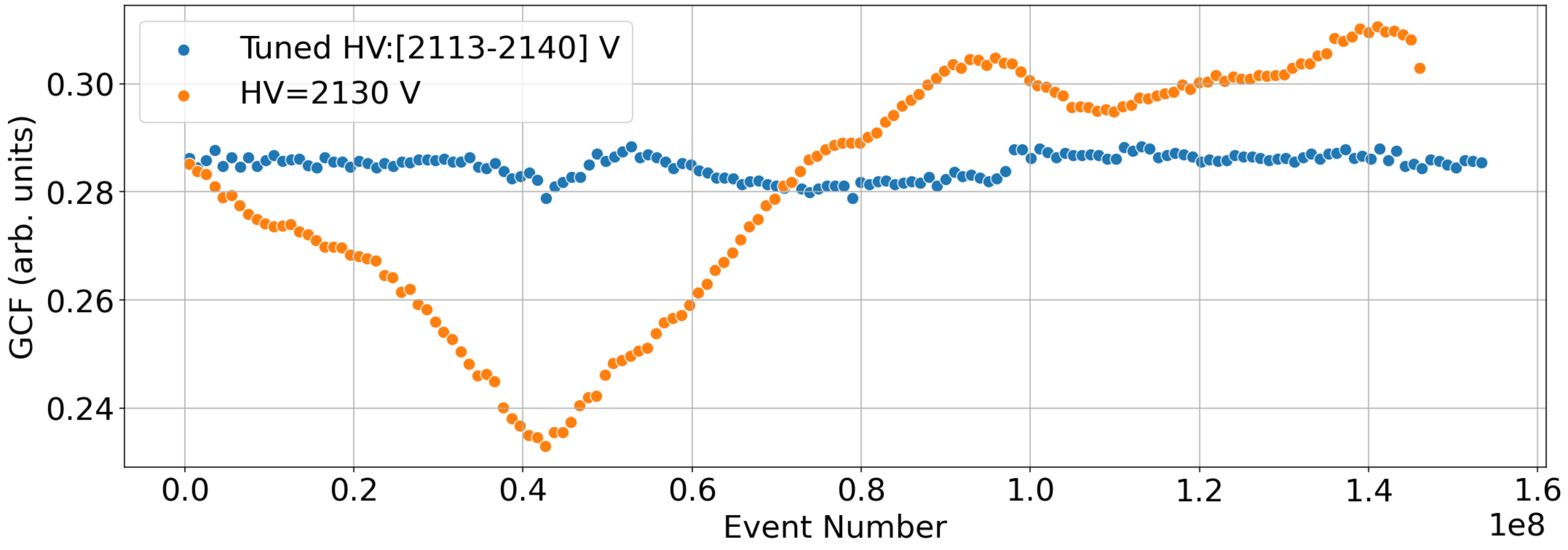
Shift crews ran script to run model inference and adjusted HV (rounded to nearest 5V) manually before starting new run.

Charged Pion Polarizability - 2022

Script was run via DAQ GO. Experimental conditions (beam current, target type and position) were quite different from our training data.

Cosmic ray tests





Split chamber in half via software

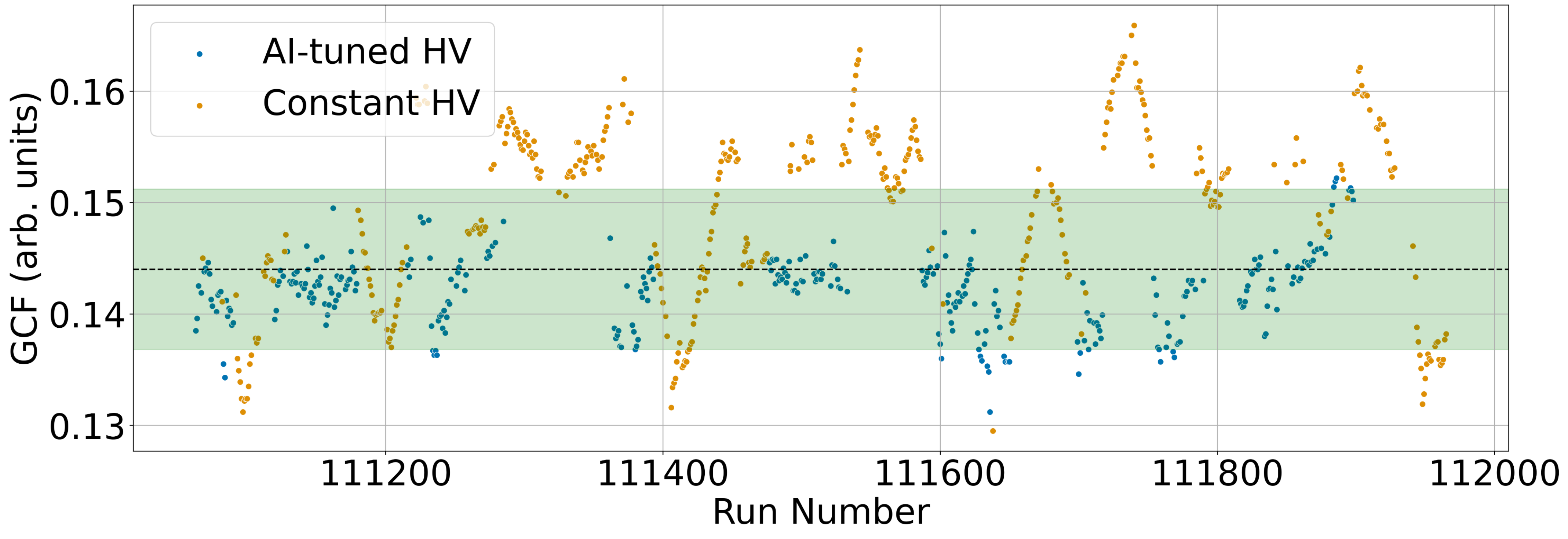
One side held at **fixed HV**, other side **adjusted** every 5 minutes

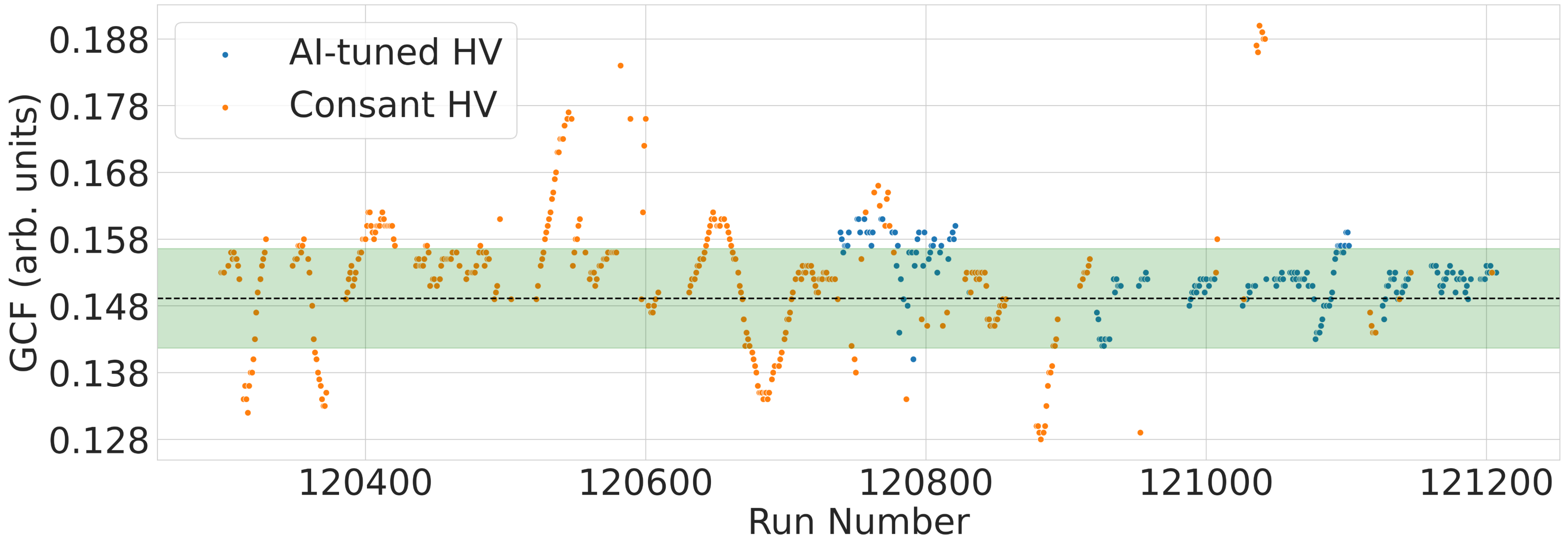
Stabilized gain

The variation in the gain for the "constant HV" side was caused by a well-timed thunderstorm.

More deployments

5% threshold





PrimEx II - 2022

Nearly all controlled runs fall within our 5% threshold.

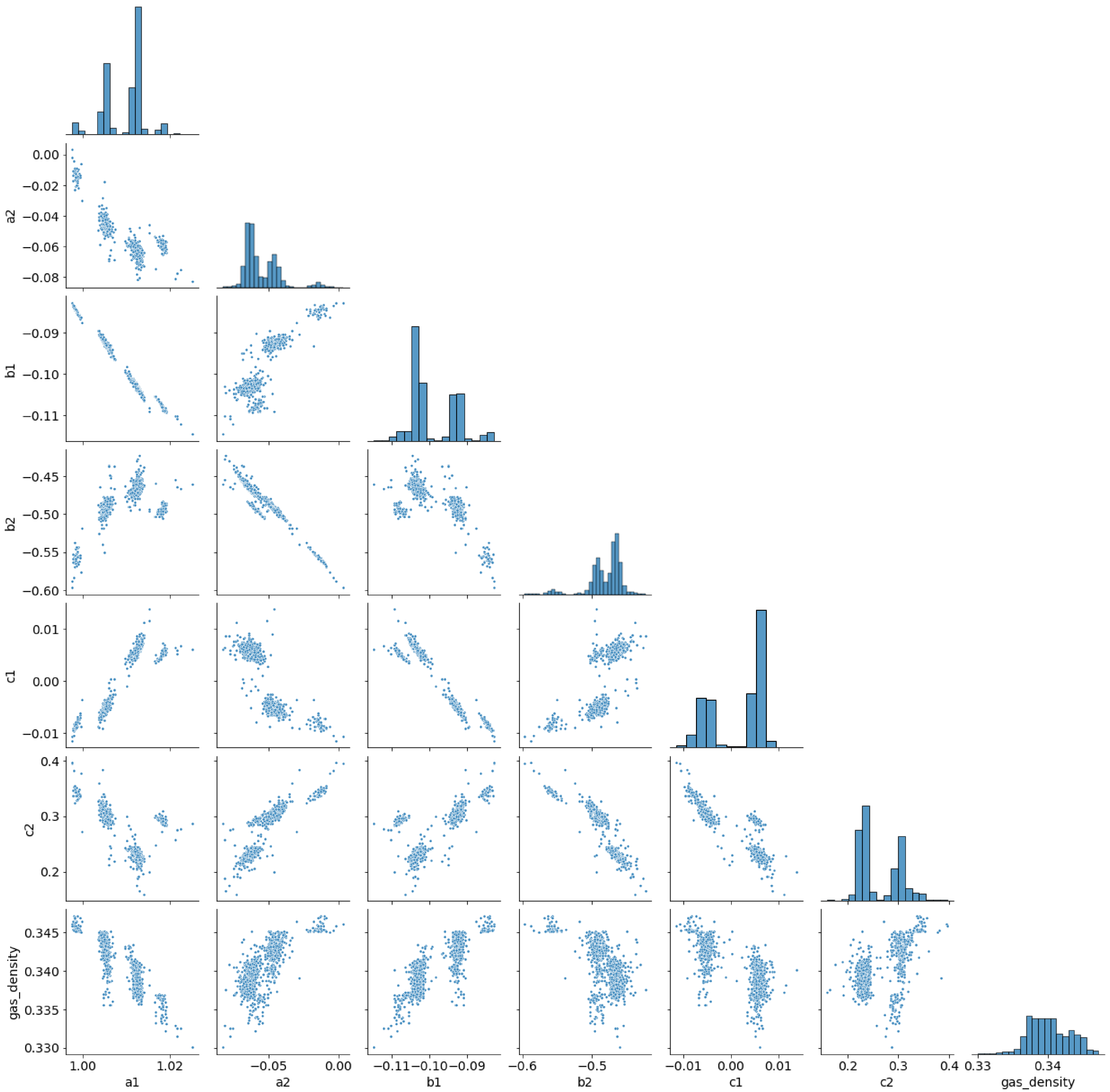
GlueX 2023

Faulty pressure sensor, nearly all controlled runs were within our 5% threshold.

Time-to-distance

Existing drift time to drift distance parameters are strongly correlated with the gas density

We can automatically generate these calibration values from fits to the gas density and reduce the number of iterations required

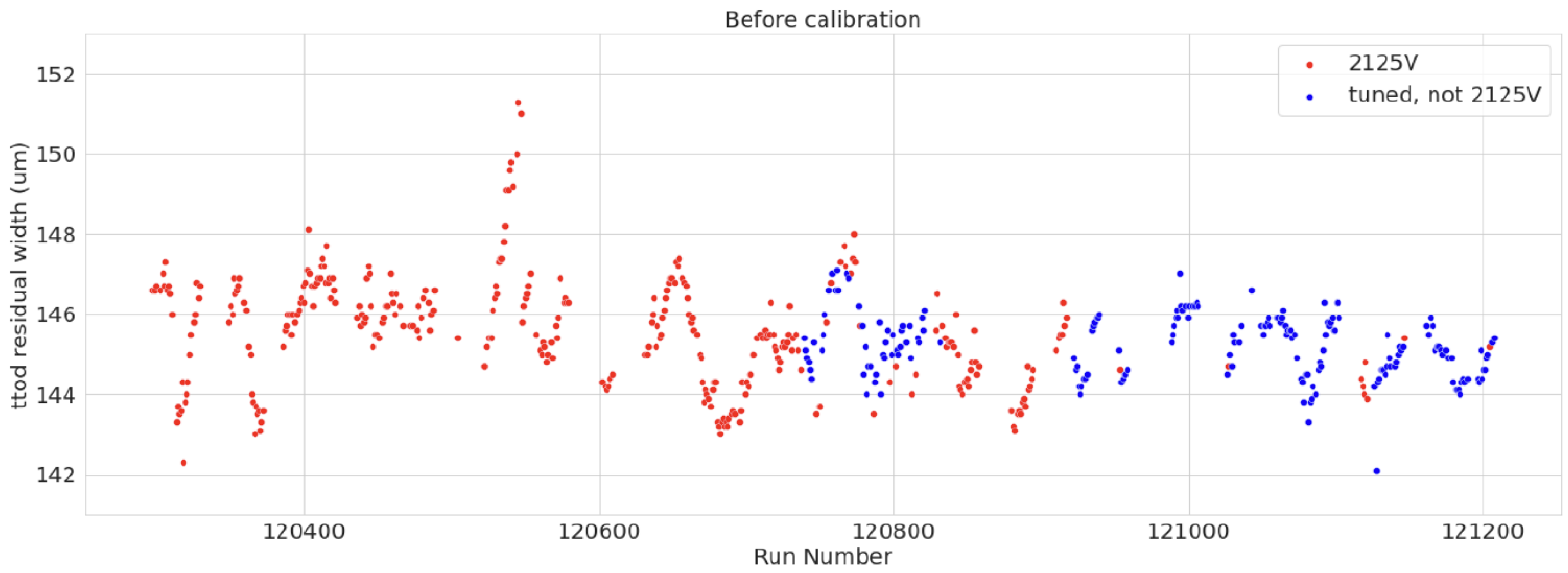


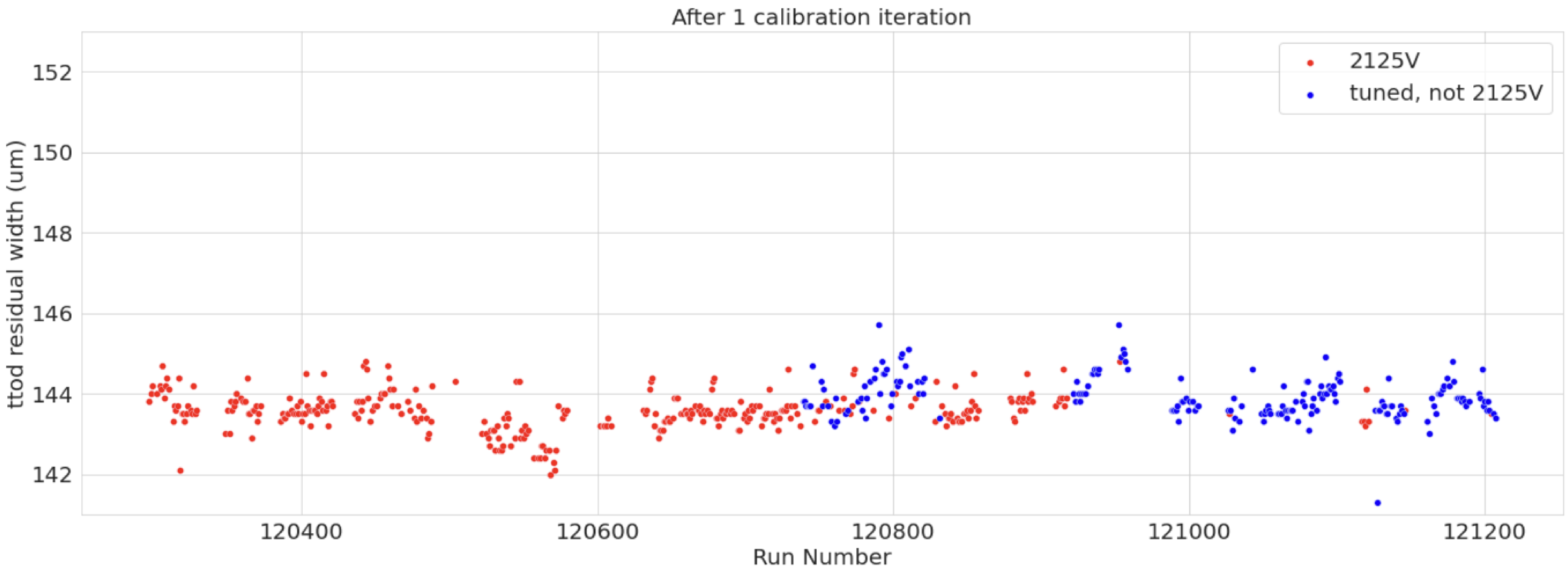
Jefferson Lab EPSCI Carnegie Mellon University

Time-to-distance

Tuning HV to stabilize the gain results in comparable performance to those runs taken at 2125 V

This alleviates any concern that adjusting the HV to stabilize the gain might negatively impact the timing resolution





Extensions to other detector systems

GlueX Forward Calorimeter

Aimed to generate calibration values using the light monitoring system, unsuccessful due to lack of correlations between existing gain values and amplitudes

GlueX Forward Drift Chambers

Chamber gain is also strongly correlated with atmospheric pressure. Can use linear regression or a GP to obtain calibration values at the start of each run.

CLAS12 Drift Chambers

Not pursued due to lack of enough historical calibration data

In Summary

Predict calibration values without relying on track reconstruction

This enables us to generate calibration values during data taking while ensuring stable detector operation and performance

An UQ-aware control system is now standard operation for GlueX

MLOps becomes critical to ensure model performance, detect data drift, and performance degradations

A team of scientists with complementary skills is critical

Continued collaboration will be essential for new AI/ML based solutions for current and future experiments

“The mostly uniform gain is great because it ensures that the yields in our monitoring plots don’t change drastically because the proton band moved below the analysis selection, and also because it ensures that minimum ionizing particles stay above the detection threshold. The biggest effect of data quality is not having hits from the same type of tracks disappear below threshold or saturate the adc *depending on the weather of the day*”

Naomi Jarvis · CDC Expert

Resources

AIEC Final Report

Extensive paper covering entire project: https://arxiv.org/pdf/2402.13261

AI4EIC2023 Proceedings

Shorter description: https://arxiv.org/pdf/2403.13823

Acknowledgements

JSA

Jefferson Science Associates, LLC operated Thomas Jefferson National Accelerator Facility for the United States Department of Energy under U.S. DOE Contract No. DE-AC05-06OR23177

AIEC

This work was supported by the US DOE as LAB 20-2261

CMU

The Carnegie Mellon Group is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, DOE Grant No. DE-FG02-87ER40315

GlueX

GlueX acknowledges the support of several funding agencies and computing facilities: www.gluex.org/thanks

Jefferson Lab EPSCI Carnegie Mellon University

AIEC Team

Work started in April 2021

Strong ties to Data Science Department and Physics Division

Torri Jeske

EPSCI

Diana McSpadden

Data Science

Nikhil Kalra

Data Science

Thomas Britton

EPSCI

Naomi Jarvis

Carnegie Mellon University

David Lawrence

EPSCI

Deployment timelines

PrimEx 2021

Shift crew manually ran roboCDC script and adjusted HV

Cosmics

Full automation tests with half of detector HV adjusted autonomously every 5 min

Observed stable gain for ML controlled side of CDC

CPP 2022

RoboCDC run from DAQ “GO” processes without intervention from shift crew

Stable performance despite out of domain input features

Control ON/OFF toggle added to CDC HV GUI

PrimEx-II 2022

Added an “auto-off” function to roboCDC to detect empty target runs and human tests

GlueX 2023

Problems with atmospheric pressure sensor resulted in less controlled runs.

Stable performance was achieved with new pressure sensor installation.