AI Experimental Calibration and Control

Thomas Britton[§], Michael Goodrich[§],
Naomi Jarvis*, Torri Jeske[§], Nikhil Kalra[§],
David Lawrence[§], Diana McSpadden[§],
Kishansingh Rajput[§]

[§]Thomas Jefferson National Accelerator Facility
*Carnegie Mellon University

Jefferson Lab
Thomas Jefferson National Accelerator Facility

Summary of expenditures by fiscal year (FY)*:

	FY20 ($k)	FY21 ($k)	Totals ($k)
a.) Funds allocated	270	270	540
b.) Actual Costs to date	192	0	192

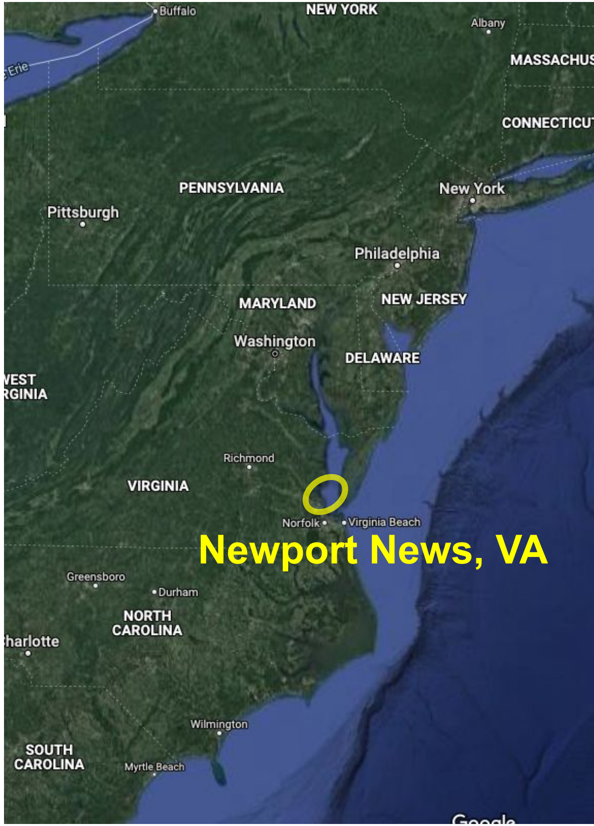
total award for 3 years: $810k

- ~$4.1k for computers
- ~$188k for staffing:
  - 1 physics postdoc (Jeske)
  - 1.25 data scientists (McSpadden, Kalra)
  - 0.1 staff scientist - physics (Britton)

**n.b. funds come at the very end of fiscal year so are actually spent during the following fiscal year*

Aerial photo taken April 6, 2012

Hall-D Complex

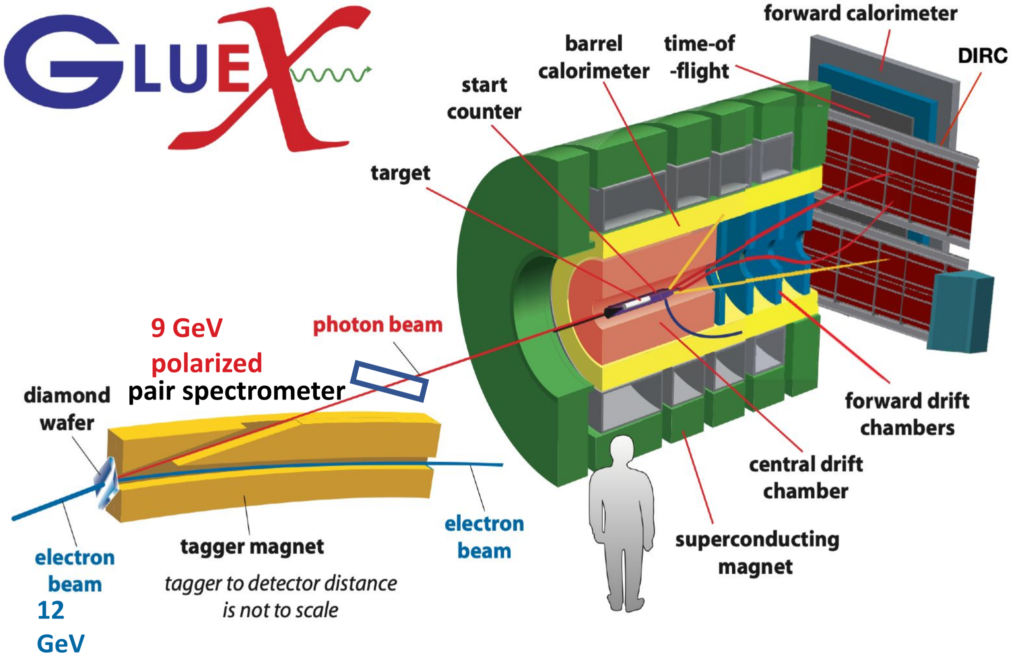


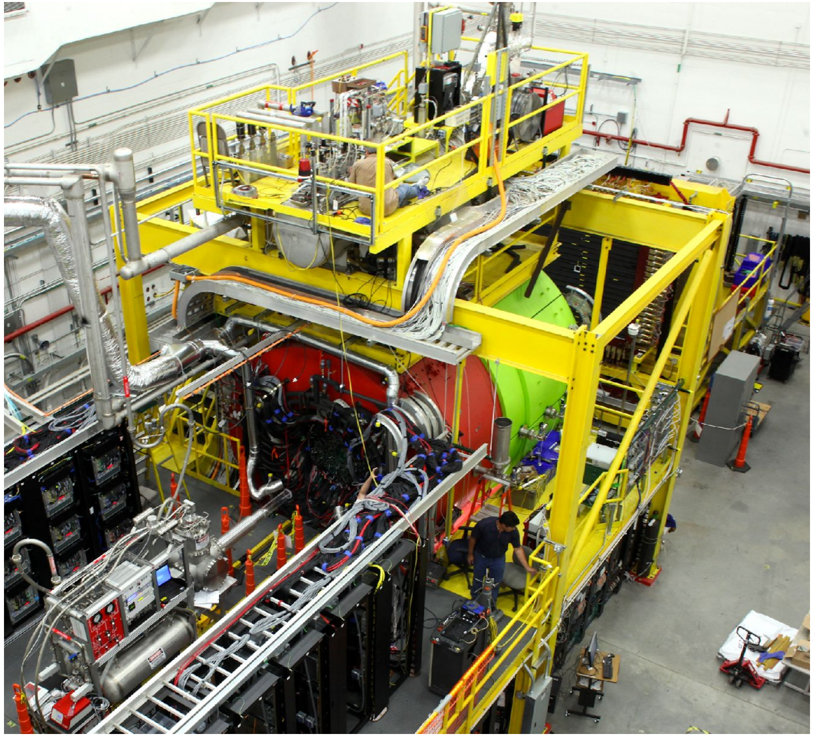
Electron beam accelerator

- continuous-wave (250MHz, 4ns bunch structure in halls)
- Polarized electron beam
- Upgraded to 12GeV (from 6GeV)
- 70 μA max @ 12Gev (200μA max @ 6GeV)

GlueX

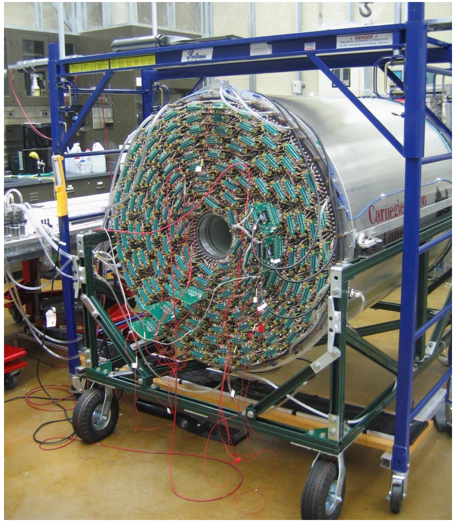
[GlueX detector](#) located in Hall D at Jefferson Lab, VA

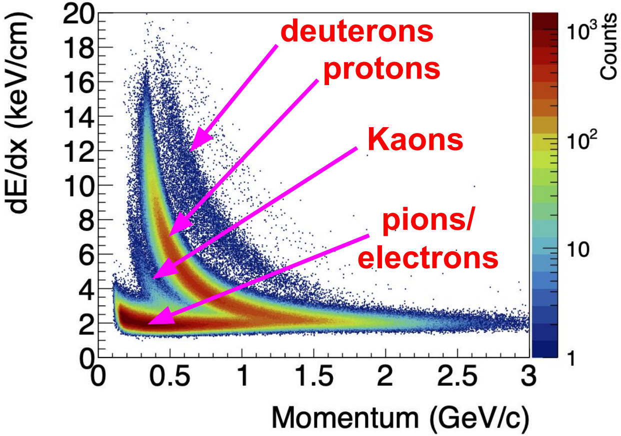




The CDC (= Central Drift Chamber)

- 1.5m long x 1.2m diameter cylinder; central hole for beam, target and start counter scintillators
- 3522 anode wires at 2125V inside 1.6cm diameter straw
- Ar/CO2 gas mix, approx. 30 Pa above atmospheric pressure
- Measures drift time and deposited charge



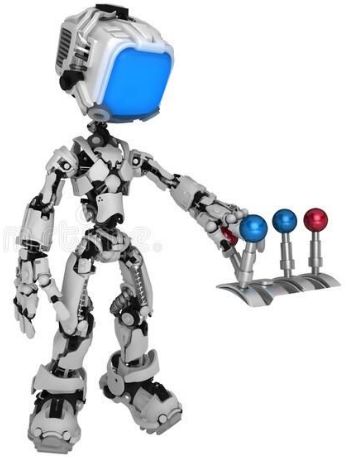


Motivation

- Sensitive detectors need to be calibrated to obtain optimal resolution
- Calibrations cause a delay between data collection and analysis (months)
  - Multiple iterations are needed to converge to final set of constants

Main Goal:

Dynamically adjust the controls of a sensitive detector to reduce or eliminate the need for calibration

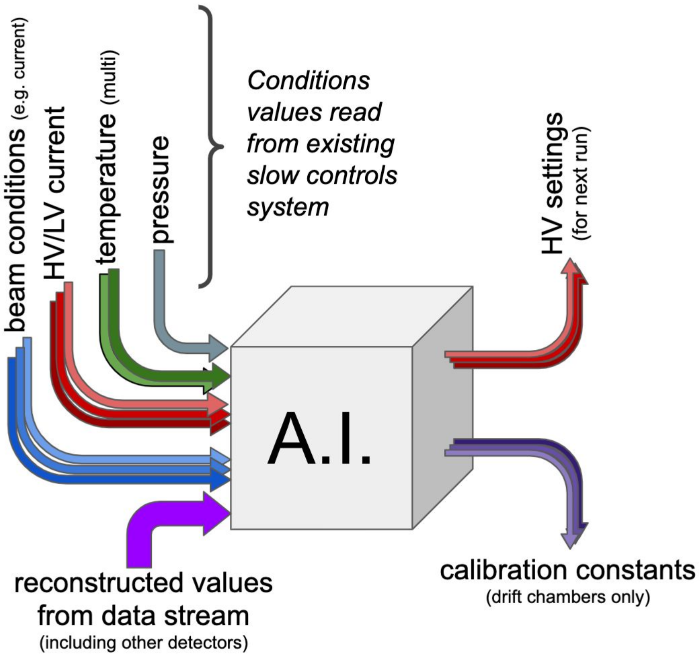


CDC gains

- CDC is sensitive to environmental conditions
  - Atmospheric pressure is one we cannot control (but can measure)
  - Beam conditions change *(usually according to the needs of the experiment)*

- AI is appropriate tool for combining disparate pieces of information to obtain a singular result

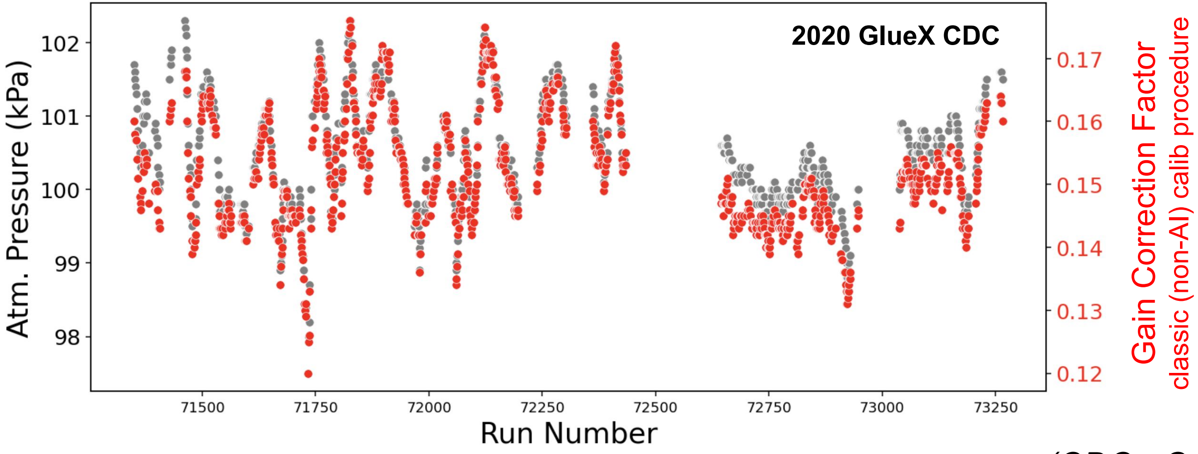
- Ultimately, the model should produce a recommendation for the HV setting of the CDC and a set of predicted calibrations that would correspond to that setting.



CDC = Central Drift Chamber

CDC gains

- Data-taking divided into *runs* (up to 2h), each session of data taking spans several months
- Detector gain sensitive to multiple environmental features
- Atmospheric pressure has strongest correlation



(CDC= Central Drift Chamber)

Bird's Eye View of Plan

1. **Reproduce existing CDC calibrations for the Gain Correction Factor (GCF) from 2018 and 2020 GlueX data using AI**
    - First, using same inputs as classic system (i.e. histograms derived from raw data)
    - Second, using only environmental inputs

2. **Using the AI model, predict GCF *before* data is taken and adjust High Voltage to stabilize calibration**
    - Test with actual beam (PrimEx-η, parasitically)

3. **Extend procedure for time-to-distance calibrations (more complex)**
    - GlueX CDC
    - CLAS12 DC
    - Test over long down time using cosmic rays

4. **Implement semi-automated system in counting house**
    - Shiftworker approval required

5. **Implement fully automated system**

6. **Extend to other detector systems**

Gaussian Process Regression (GPR)

2020 run data (filtered)

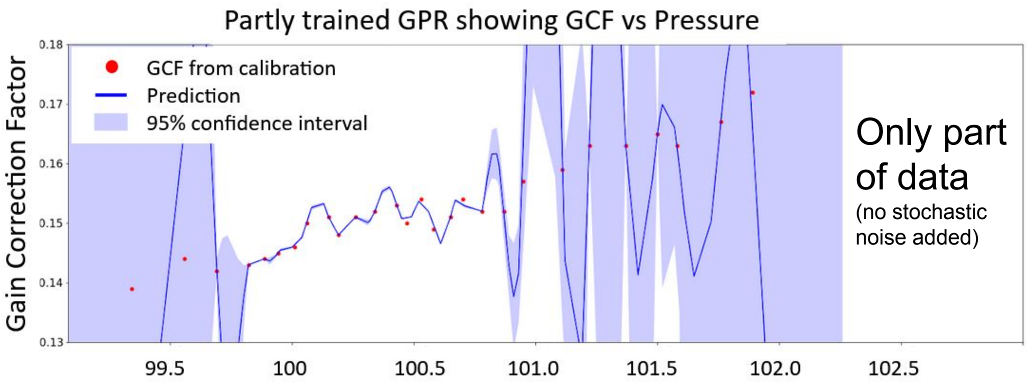
- 430 training observations
- 106 testing observations

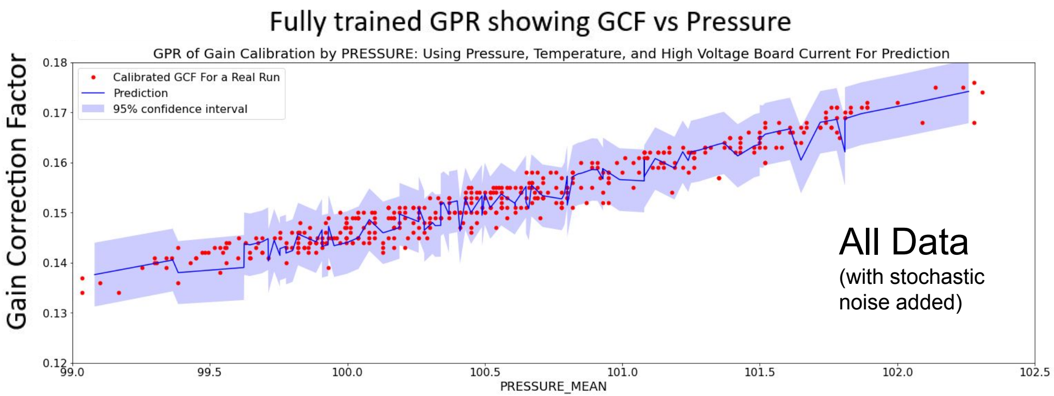
AI/ML methods applied:

NN,
Random Forest,
XGBoost,
GPR

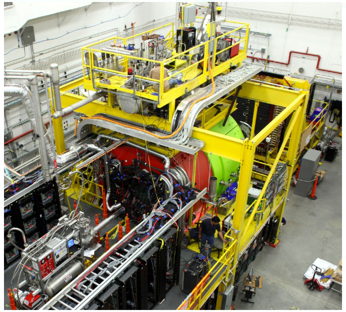
Gaussian Process Regression

- Suited to small data set
- Provides uncertainty quantification



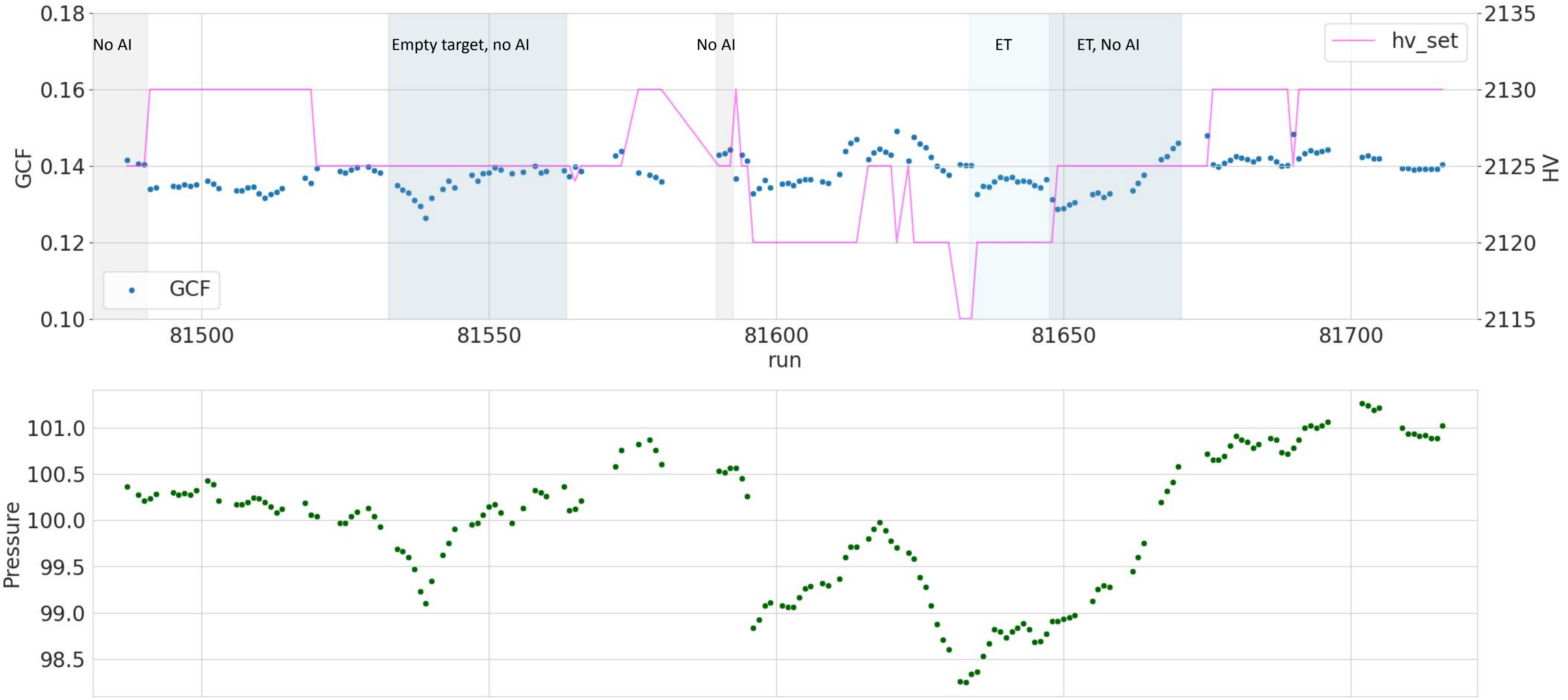


Fall 2021 Beam Test



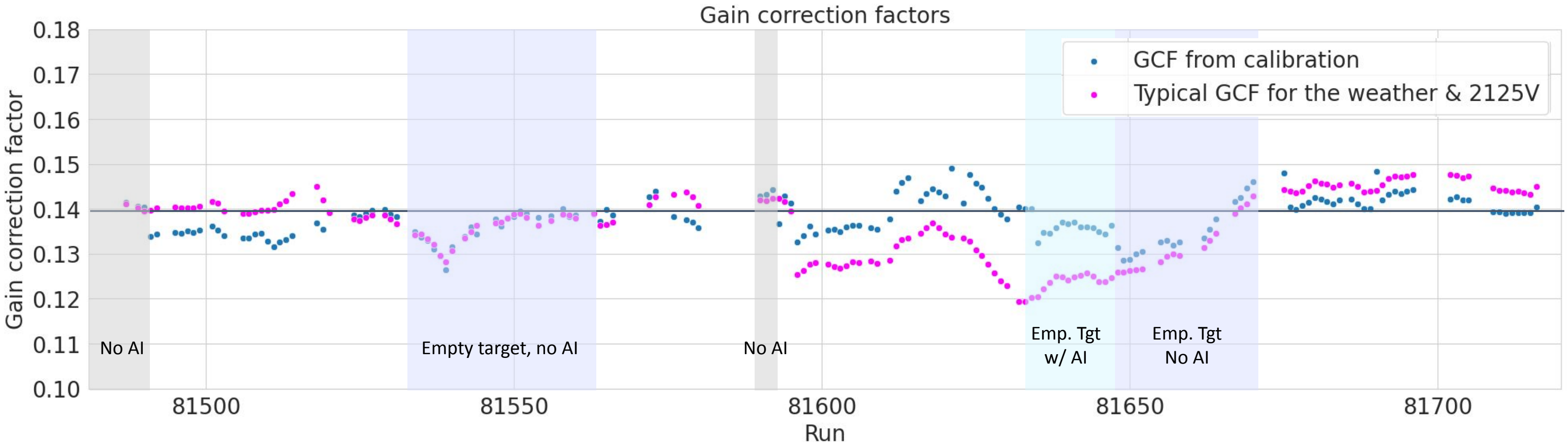
- Mid-October to early November
- PrimEx-η running with GlueX Detector in Hall-D
  - Run plan was to have small amount of data with Solenoid on but most with it off
- Planned to test AI system over 2 days when solenoid was on
- Background levels were improved significantly with solenoid on
  - PI's changed plan and ran with it on for ~2weeks 
  - Atmospheric pressure did not change as much as we wanted 

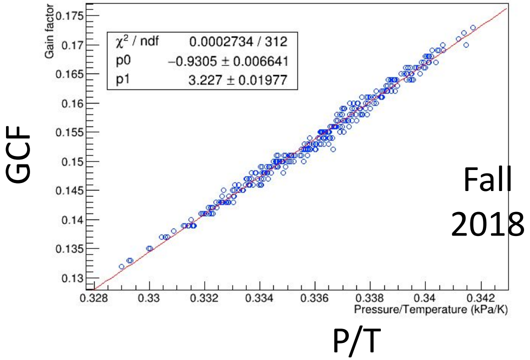
Gain correction factors from conventional calibrations

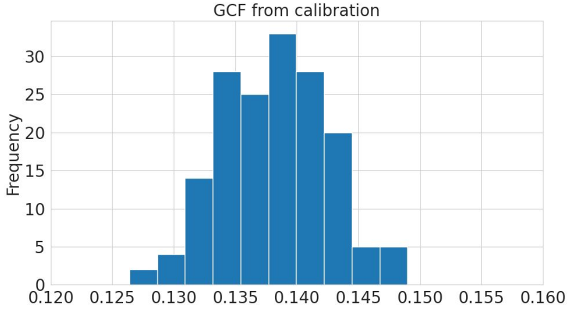


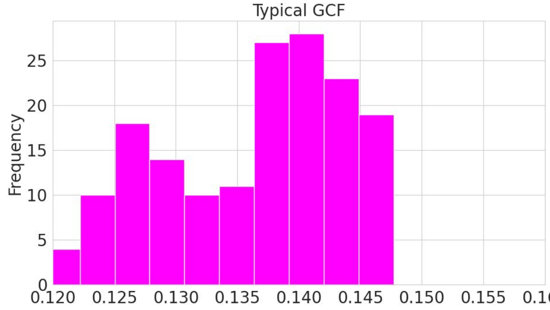
Estimated the GCFs that we'd have found if we had used 2125V all the time

Used 2018's fitted gain vs P/T, scaled to match 2021's mean GCF for runs at mean pressure (x 0.138/0.1475)



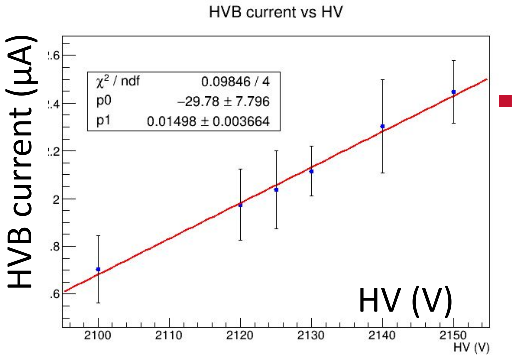


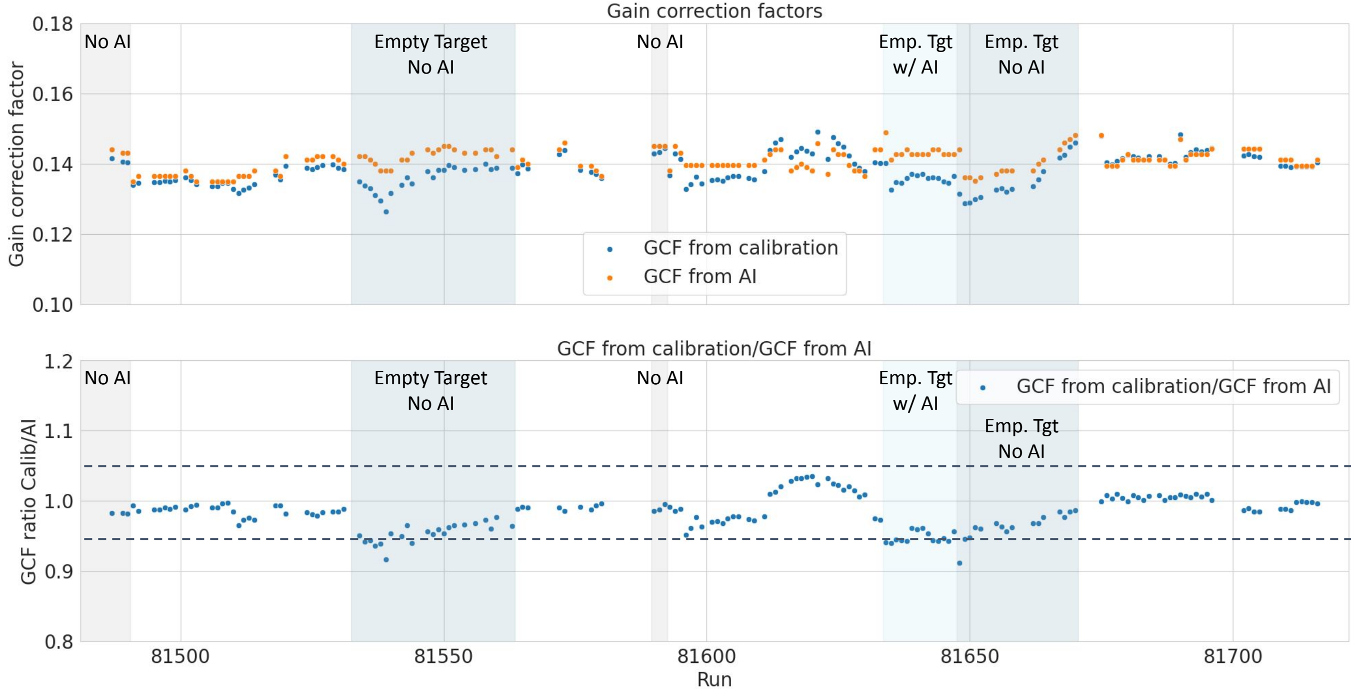


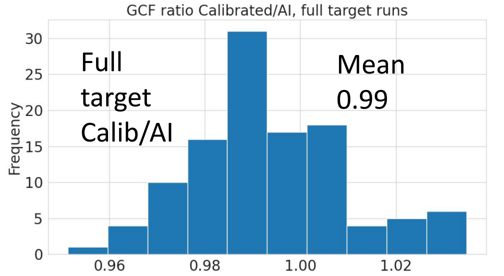


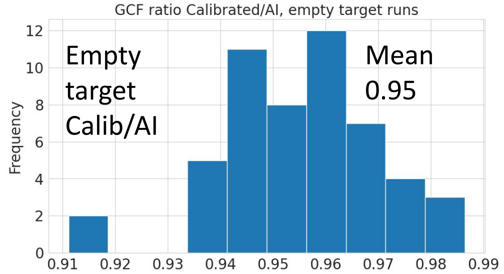
Comparing GCFs from calibration with GCFs from the AI

Harvested EPICS data for each run, scanning to find 30s with beam 200 +/- 5 nA
Scaled the HVB current to find what it would have been at 2125V
Used EPICS + scaled HVB current with the AI to predict new GCFs.

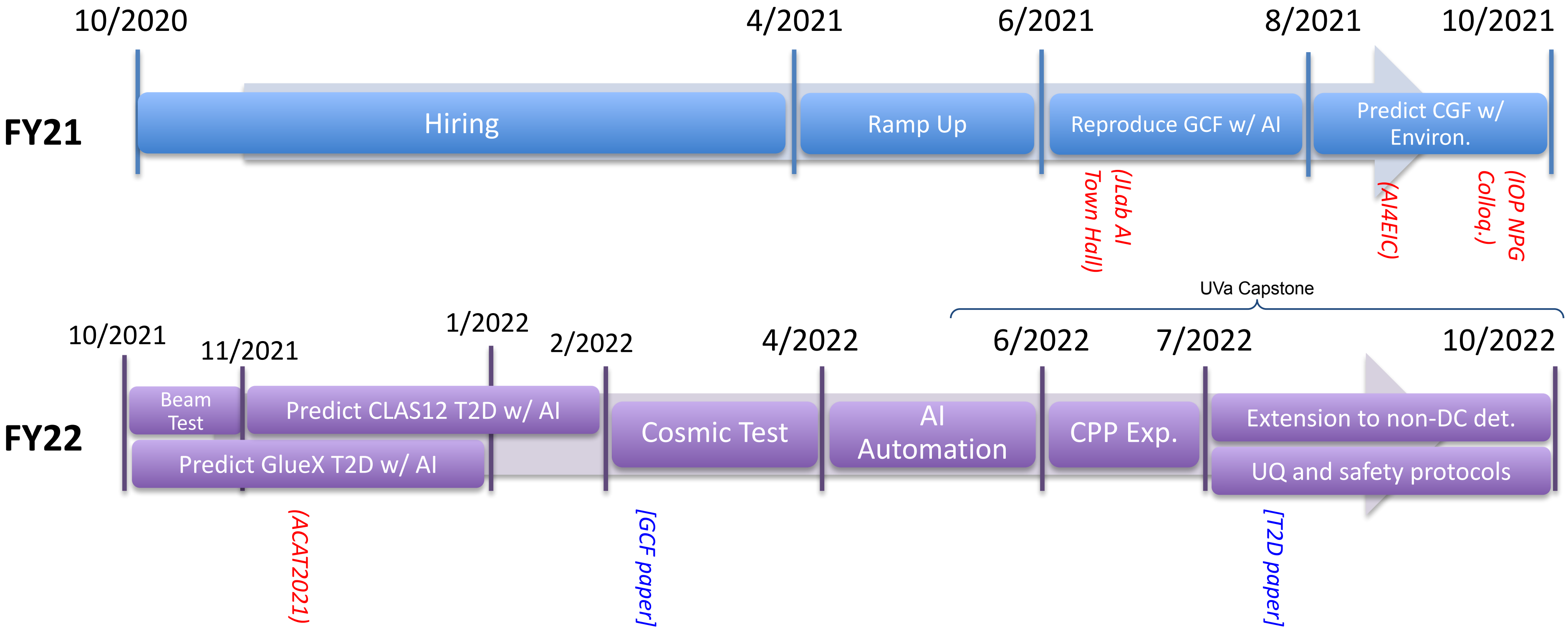








Schedule



*Conference/Workshop presentations

Milestones			
FYQtr	**Description**	**Due Date**	**Status**
FY21Q1	Setup accounts, post job descriptions and form hiring committees.	12/23/2020	✓
FY21Q2	Hiring committee completes interview process and new hires for project are in place	3/31/2021	✓
FY21Q3	New hires able to calibrate GlueX CDC detector using existing system.	5/1/2021	✓
FY21Q3	System in place to extract data from Hall-D EPICS archive and GlueX raw data/reconstructed data into form suitable for machine learning.	6/15/2021	✓
FY21Q4	Candidate network topologies identified along with initial dataset to be used for training.	9/30/2021	✓
FY22Q1	Able to calibrate CLAS12 DC detector using existing systems.	12/23/2021	started
FY22Q2	System in place to extract data from Hall-B EPICS archive and CLAS12 raw data/reconstructed data into form suitable for machine learning.	3/31/2022	✓

Milestones

FYQtr	Description	Due Date	Status
FY22Q3	Plan developed and software machinery in place to test prototype system using current best model for GlueX CDC detector with cosmic rays. System will provide suggestions to shiftworkers for new settings.	6/30/2022	
FY22Q4	Model review and refinement. Performance of initial model choices reviewed and decisions made on whether new model development is needed or refinement of existing models. Plan for next stages of development in place based on results of review.	9/30/2022	
FY23Q1	Plan developed and software machinery in place to test prototype system using current best model for CLAS12 DC detector with cosmic rays. System will provide suggestions to shiftworkers for new settings.	12/23/2022	
FY23Q2	Performance evaluation complete.	3/31/2023	
FY23Q3	Documentation in place for system deployment and operation.	6/30/2023	
FY23Q4	System deployed in standard experimental hall operations in Hall-B and Hall-D (pending collaboration approval)	9/30/2023	

Summary

- Successfully reproduced calibration constants using AI model using same inputs as classic method
- Successfully ***predicted*** GCF calibrations using environmental data only for GlueX 2018 and 2020 data
- Successful preliminary beam test where AI suggested HV settings resulting in more stable GCF
- In process of implementing for time-to-distance calibrations for both GlueX CDC and CLAS12 DC
- Permission obtained to implement semi-automated system for CPP experiment in summer 2022
- UVA Capstone proposal submitted to include DS graduate students in project

Acknowledgements

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

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

Backups

PI Meeting Presentation Guidelines:

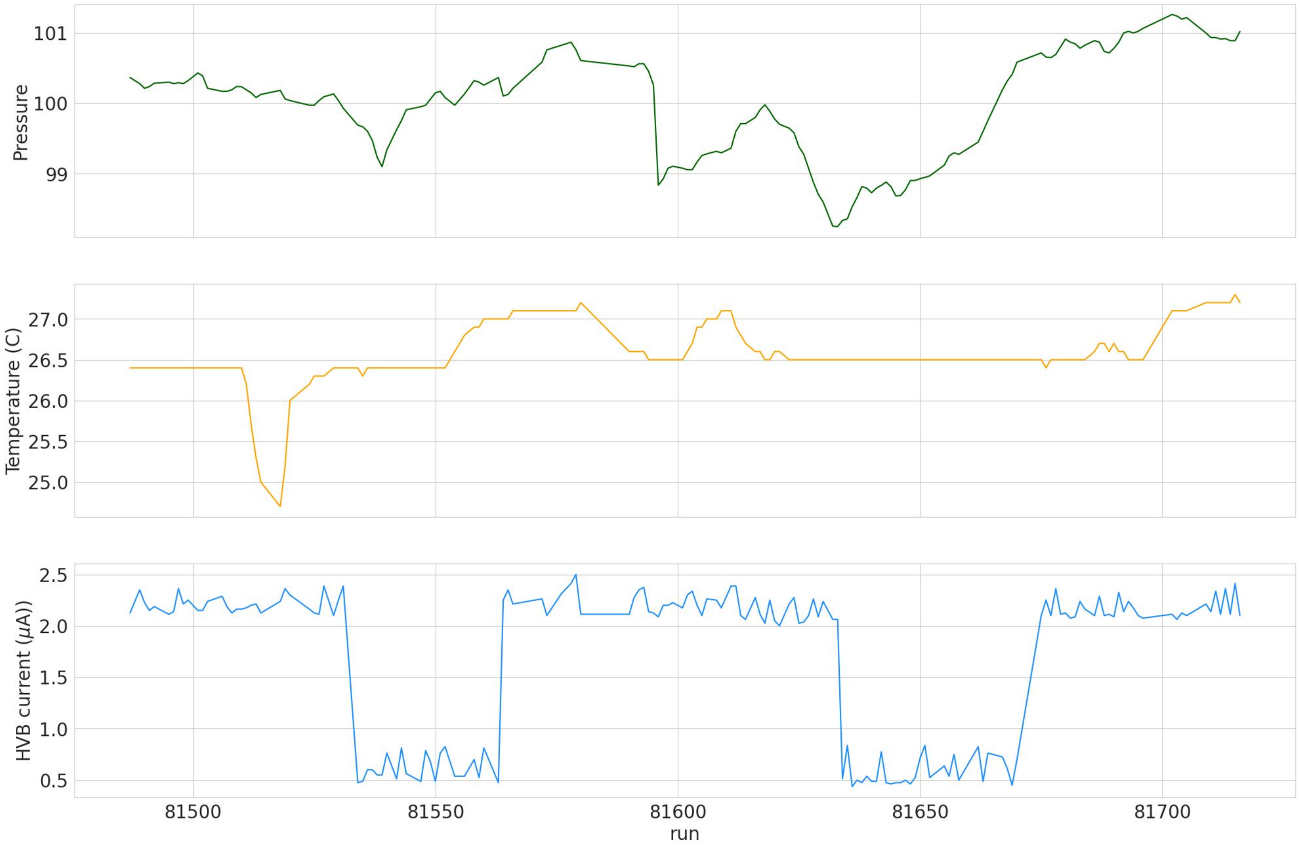
Each presentation should include the following information:

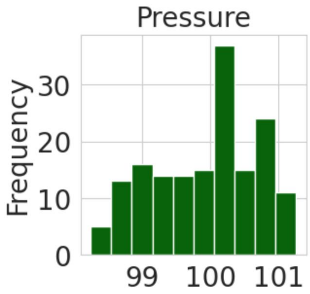
- Description of the project and the current status;
- Main goal of the project for which you received the FY 2020- 121 Accelerator R&D award , and/or Lab Data call for AI/ML award.
- A table showing annual budget and the total received to date (see below);
- A table showing major deliverables and schedule;
- Summary of expenditures by fiscal year (FY):
- There will be no written report or follow up actions required for this meeting.
- All talks will be posted on PI Exchange meeting page on NP website.
- **35 min talks should allow 7 min for Q/A and 30 min talks 5 min for Q/A**

	FY20	FY21	Totals
a) Funds allocated			
b) Actual costs to date			

Measurements from EPICS – pressure, temperature and HVB current

All runs
after HV scan





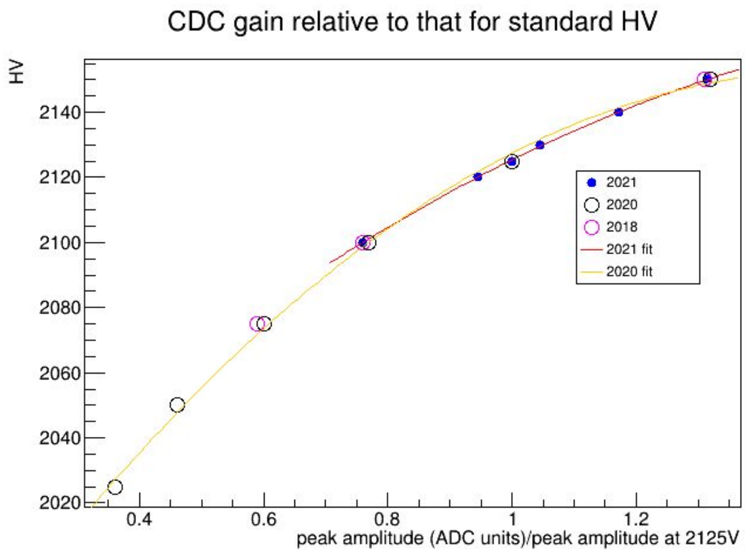
Procedure before each run – scripted except for the HV change

Gathered EPICS data for 1 minute with steady beam current and used AI to predict GCF for conditions at the time. Should have first set HV to 2125V.

Calculated expected GCF/ideal GCF

Find the HV needed to produce that relative gain
(using fitted function from HV scan data)

Rounded the HV to the nearest 5V, in case of calibration issues.



Reverted to 2125V for empty target runs

- not sure if they could be calibrated directly
- if not, need to copy calibrations from other runs
- HVB current will be much lower than training data

Drift times simulated for HV tuned to the pressure

- Drift time to distance conversion uses a table of ideal drift times simulated for standard pressure and nominal HV 2125V (GARFIELD). Calibration accounts for imperfect straws and pressure.

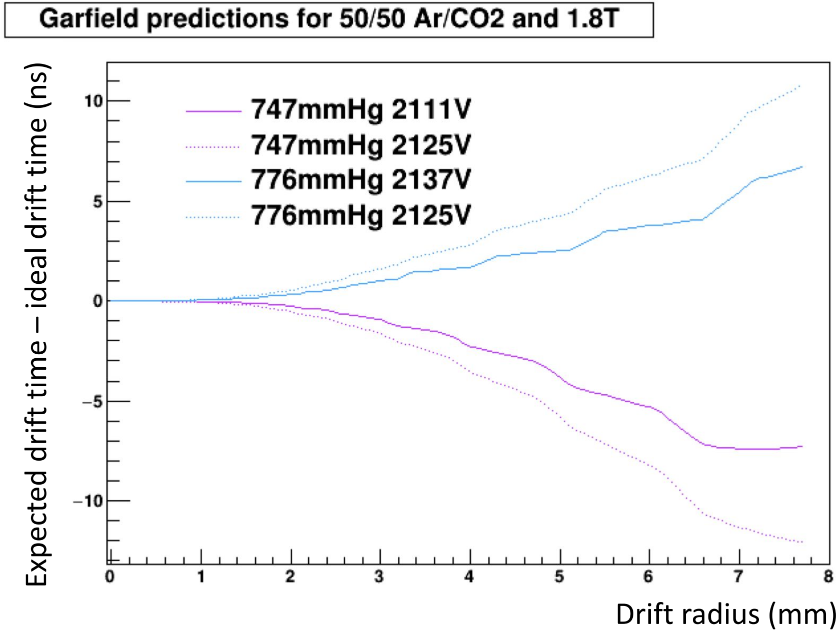
- Calculated difference between expected and ideal drift times at extreme pressure values

  Faint lines: 2125V

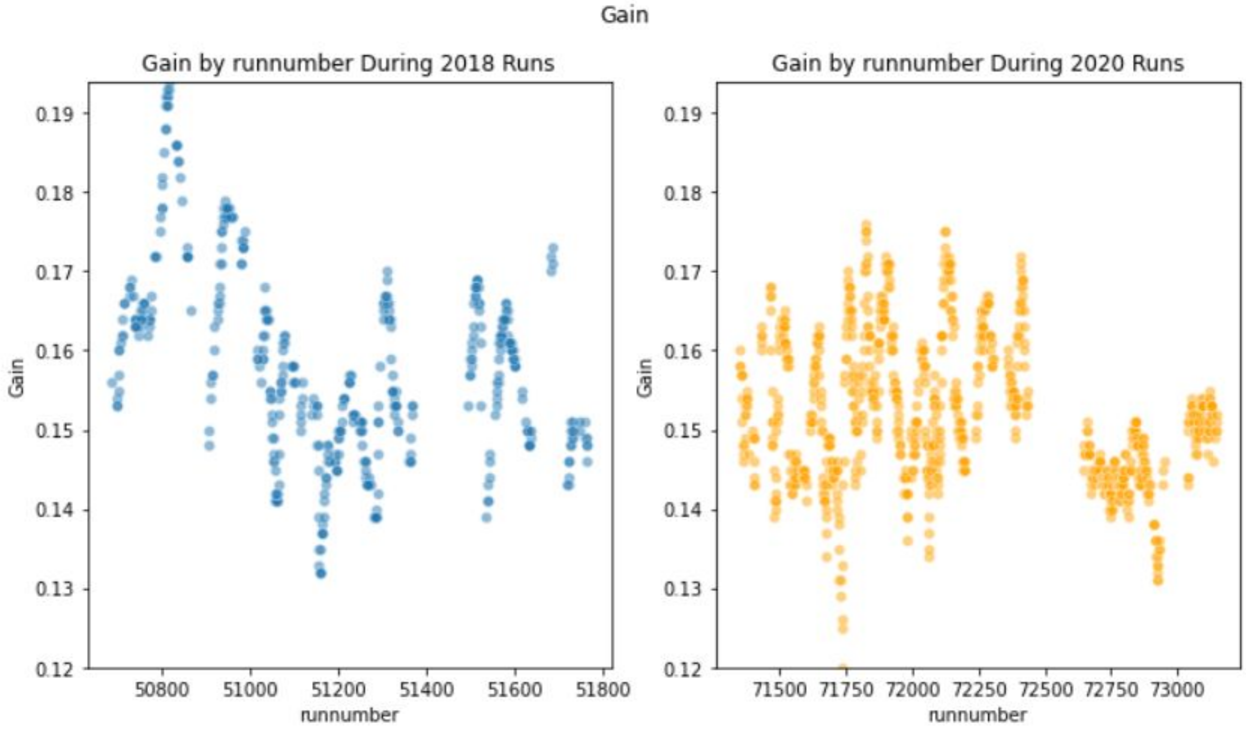
  Solid lines: tuned HV

- Differences are small, smaller for tuned HV

- Tuned HV should improve the position resolution



CDC gains

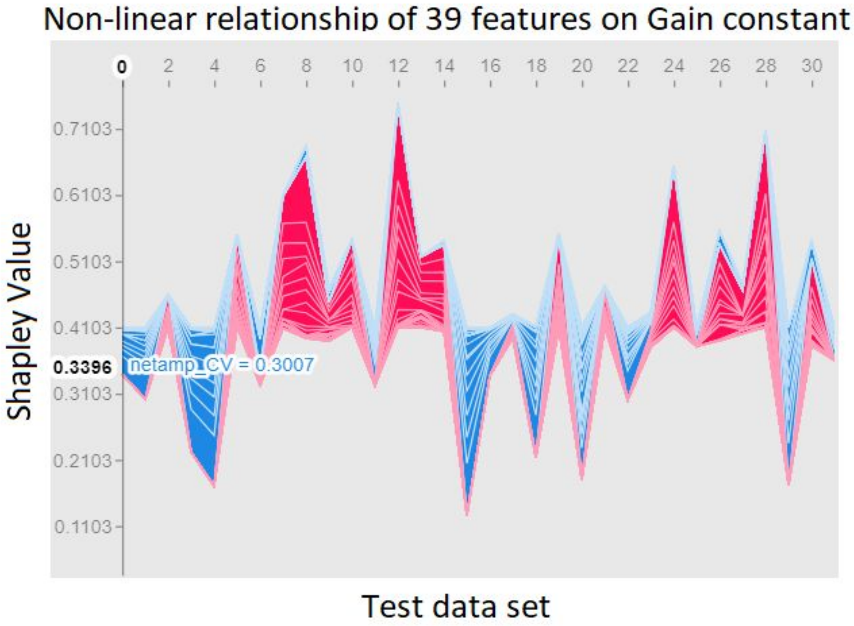


Can we use AI to predict existing gain constants to within ~1%?

CDC gains

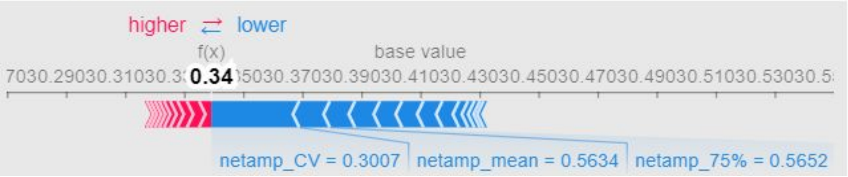
For regression problems, there are a number of available evaluation methods.

We implemented Shapley values.



"The Shapley value is a framework originally proposed in the context of game theory to determine individual contributions of a set of cooperating players" - Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications | IEEE Journals & Magazine | IEEE Xplore

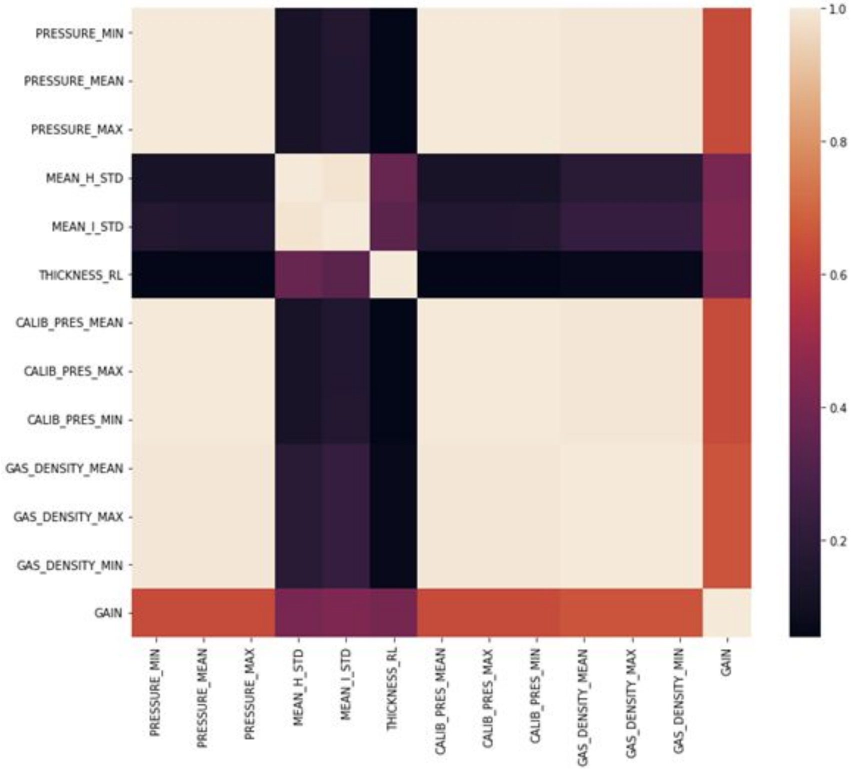
The first test data set run:
effect of features on Gain constant for this run



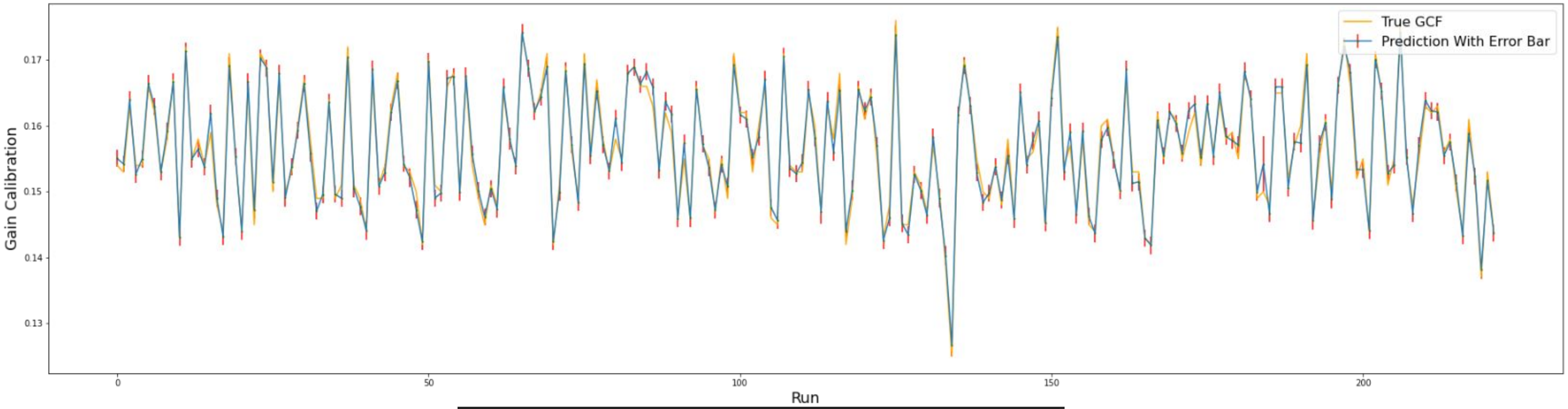
Feature importance

Input Features:

- Aggregate features per run from experimental data and EPICS system:
  - Netamp = pulse height - pedestal, momentum, track angle, drift time
- Split data into train and test sets:
  - 438 runs from 2018
    - 350 train
    - 88 test
  - 897 runs from 2020
    - 717 train
    - 180 test
- **Iterate feature importance** to help with feature engineering and minimize needed data/model size



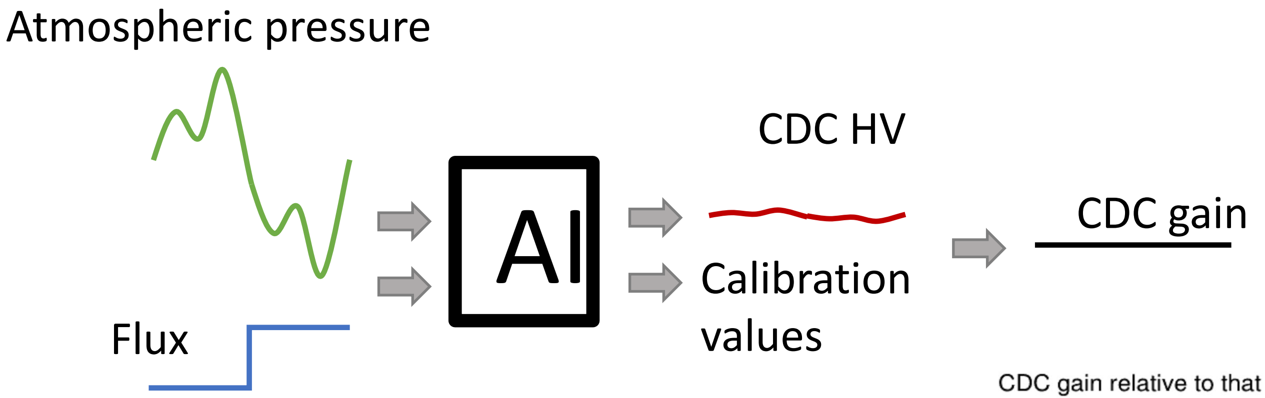
CDC gains results



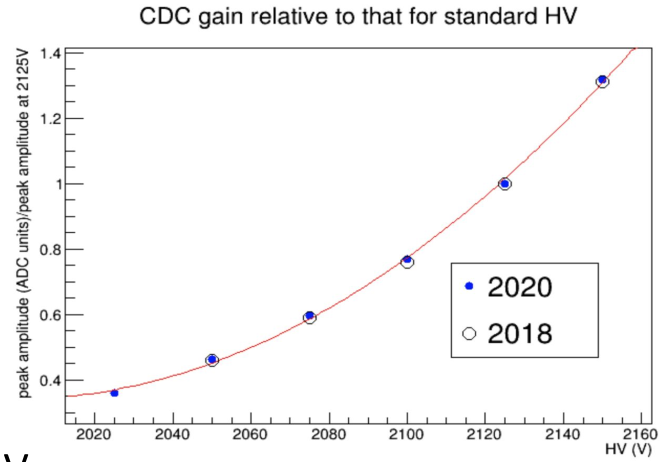

```
Predictions on 222 Training and Test Data Sets
_______________________________________________
Minimum Different Between Truth and Pred: 0.000005
Maximum Different Between Truth and Pred: 0.005996
Mean Different Between Truth and Pred: 0.001026
Minimum Perc Dif: 0.000029
Maximum Perc Dif: 0.039976
Mean Perc Dif: 0.006556
Mean GCF: 0.15697727272727272
------------------------------
IF WE JUST USED THE MEAN GAIN INSTEAD OF PREDICTION?:
Minimum Perc Dif Between Truth and Mean Truth: 0.000145
Maximum Perc Dif Between Truth and Mean Truth: 0.145820
Mean Perc Dif Between Truth and Mean Truth: 0.044877
```

AI solution better than just using the mean gain

HV controls + gains



- AI predicts Gain Correction Factor (GCF) for 2125V
- Ask AI for ideal GCF, at std pressure (101.3 kPa)
- Ask AI for expected GCF at pressure right now
- Calculate relative change in gain needed
- Use known behaviour of relative gain vs HV to find desired HV



time-to-distance

- Current calibration method produces 6 unique calibration constants from fit to data

$$d(t) = f_\delta\left(\frac{d_0(t)}{f_0}P + 1 - P\right)$$

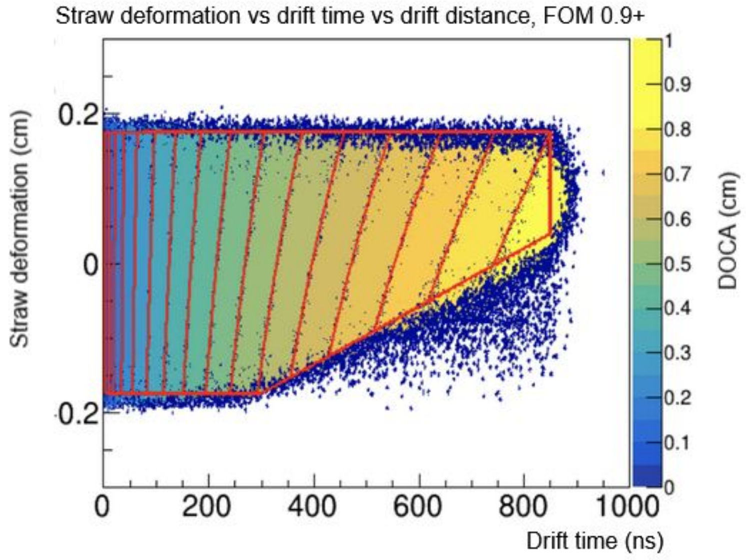
$$f_\delta = a\sqrt{t} + bt + ct^3$$

$$f_0 = a_1\sqrt{t} + b_1 t + c_1 t^3$$

$$a = a_1 + a_2|\delta|$$

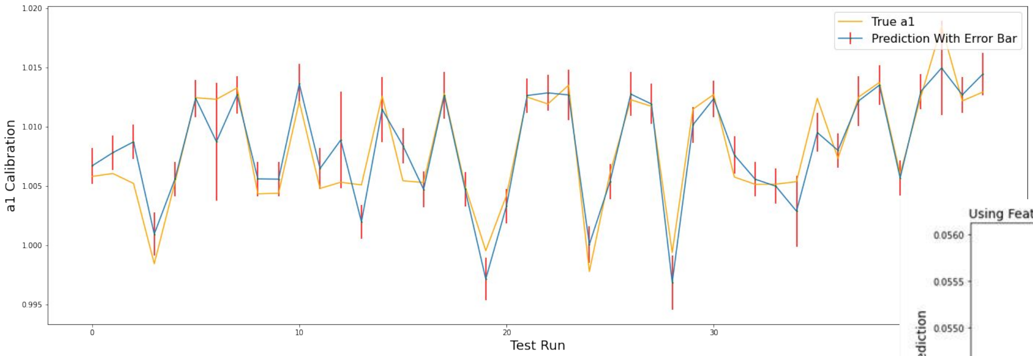
$$b = b_1 + b_2|\delta|$$

$$c = c_1 + c_2|\delta|$$



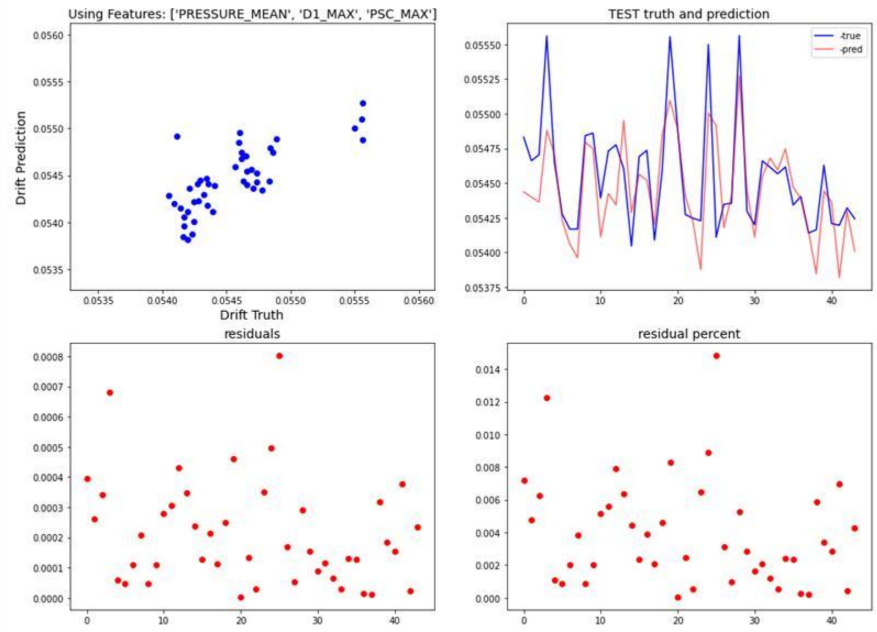
Model development for calibration constants is in very early stages

Early results

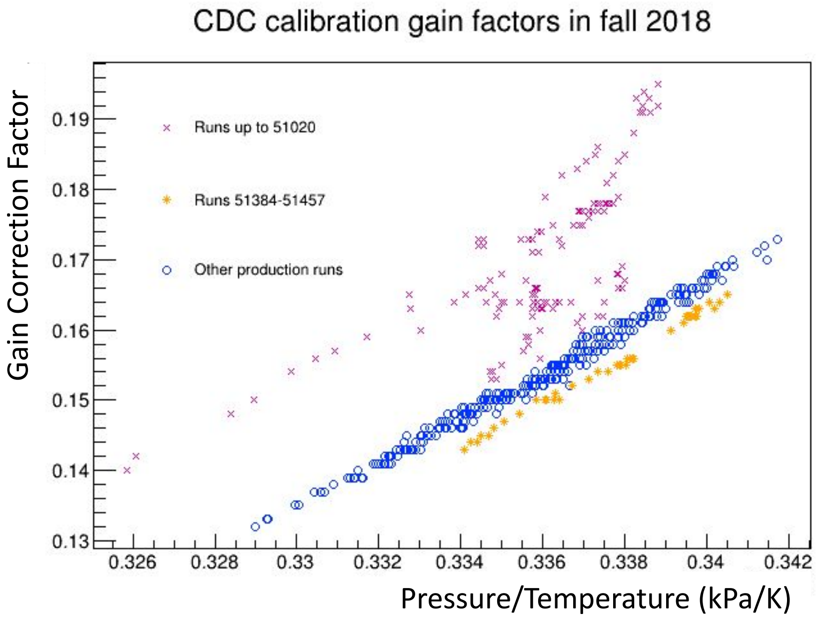


A1 has the biggest effect so concentrating efforts here

First results look promising



Gain factors vs atmospheric pressure/gas temperature



Blue: standard production runs

Gold: low intensity running

Pink: earlier runs, mixed conditions

- Measurements from thermocouples inside the CDC and barometer outside the hall
- Linear relationship between gain and pressure/temperature gives confidence that AI can be trained
- Offset between regions with different flux (beam current or radiator) adds complexity

Gain calibration AI models generation 1

- Used all relevant EPICS data :

  atmospheric pressure, gas temperature

  electron beam current, diamond radiator ID and orientation, pair production rates

  current drawn by CDC’s HV boards

- Also included reconstruction data

  track angle, momentum and dE/dx, aggregated for each run (mean, mode, quartiles, etc)

Data	Max error	Mean abs error
2018 & 2020	2.0%	0.37%

ediction)/truth

- Performance was great, but the setup was impractical (reconstruction is slow)

Gain calibration AI models generation 2

- Used all relevant EPICS data :

    atmospheric pressure, gas temperature

    electron beam current, diamond radiator ID and orientation, pair production rates

    current drawn by CDC's HV boards

- Performance was good for 2020, not so good for 2018, GPR method performed better than NN

Data	Max error	Mean abs error
2018 & 2020	9.1%	1.5%
2020	3.6%	0.7%

- Filtered the data to exclude runs with beam trips

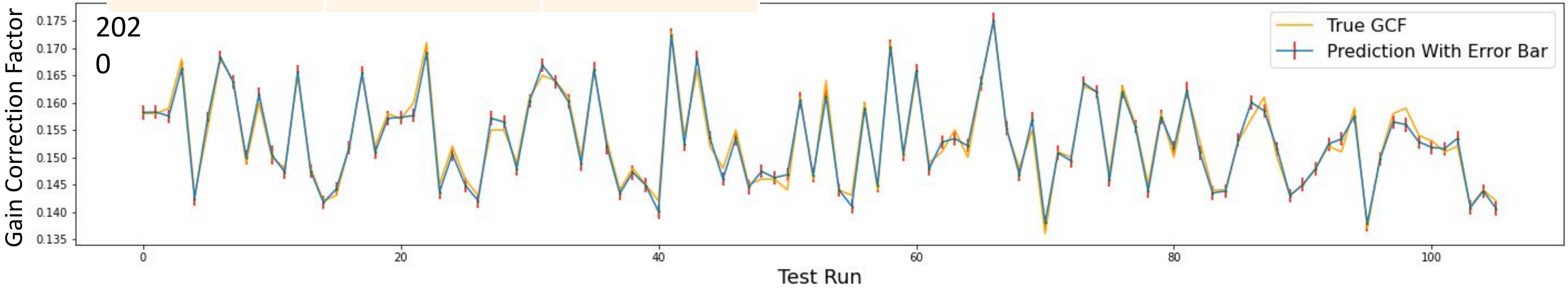
Gain calibration AI models generation 3

- Used drift-chamber-specific EPICS data: atmospheric pressure, gas temperature, CDC's HVB current

  More generic, easier to apply to other drift chambers

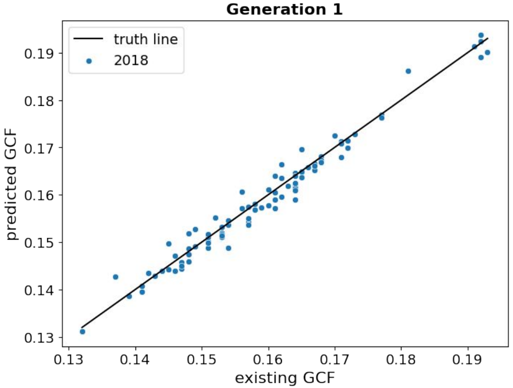
- Performance was very good for 2020, not so good for 2018

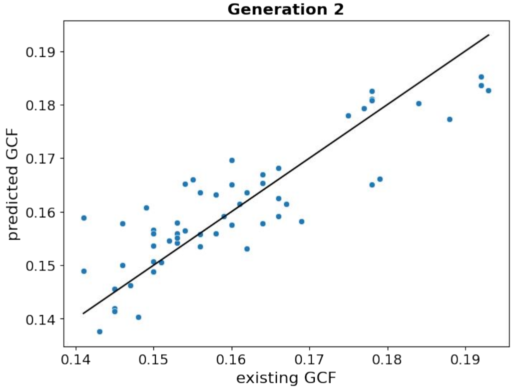
Data	Max error	Mean abs error
2018 & 2020	11.0%	1.7%
2020	1.9%	0.68%

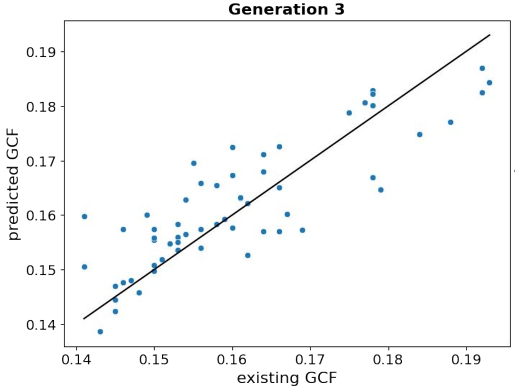


Gain calibration AI model generations comparison

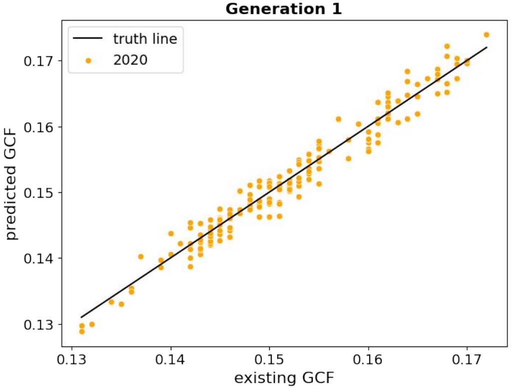
EPICS + reconstruction	EPICS	drift chamber EPICS	

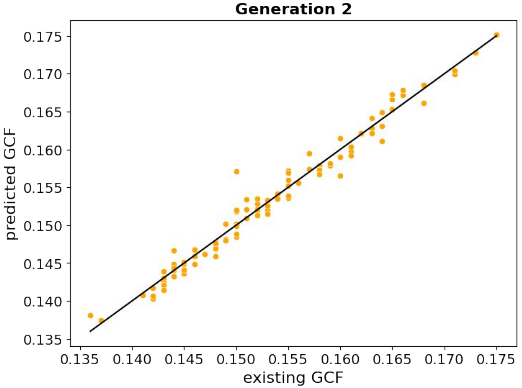


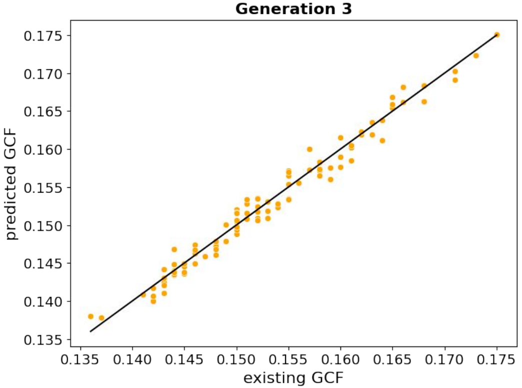




2018







2020

Estimate of the range of HV needed

- Obtained new HV values for several runs spanning the pressure range from 2018

Run	GCF	Pressure from EPICS	Calibrated Pressure (mmHg)	GCF/ideal_GCF	New HV
51687	0.173	102.067	776	1.146	2137
51570	0.160	101.042	768	1.060	2129
51762	**0.151**	**100.016**	**760**	**1.000**	**2125**
51287	0.139	99.1262	753	0.921	2116
51160	0.132	98.4129	747	0.874	2111

New HV obtained from fit to relative GCF as function of HV

- Range of HV needed is within 12V of standard HV

References

GlueX Detector NIM A**987**, 164807 (2021)

GlueX Central Drift Chamber NIM A962, 163727 (2020)

Experimental Physics and Industrial Control System https://epics.anl.gov/

Garfield – Simulation of Gaseous Detectors

https://garfield.web.cern.ch/garfield/