Control and Calibration of GlueX Central Drift Chamber Using Gaussian Process Regression

Diana McSpadden
Thomas Jefferson National Accelerator Facility
Newport News, VA 23606
dianam@jlab.org

Torri Jeske
Thomas Jefferson National Accelerator Facility
Newport News, VA 23606
roark@jlab.org

Naomi Jarvis
Carnegie Mellon University
Pittsburgh, PA 15213
nsj@cmu.edu

Thomas Britton
Thomas Jefferson National Accelerator Facility
tbritton@jlab.org

David Lawrence
Thomas Jefferson National Accelerator Facility
Newport News, VA 23606
davidl@jlab.org

Nikhil Kalra
Thomas Jefferson National Accelerator Facility
Newport News, VA 23606

Abstract

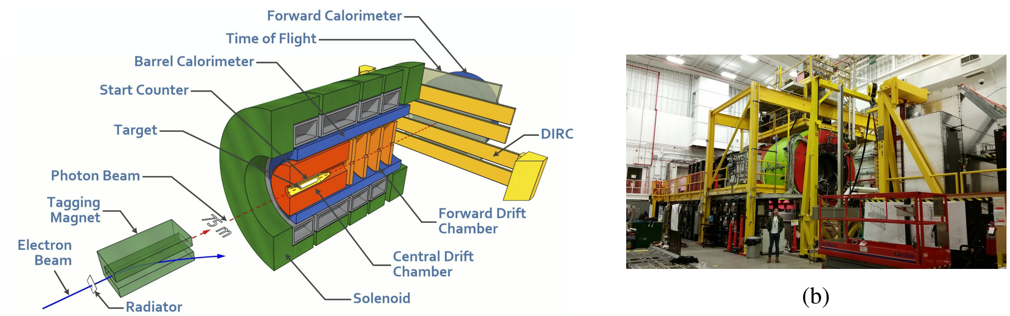
The Gluonic Excitations (GlueX) experiment is designed to search for exotic hybrid mesons using photoproduction, and to study the hybrid meson spectrum predicted from Lattice Quantum Chromodynamics. For the first time, the GlueX Central Drift Chamber was controlled autonomously using machine learning (ML) to calibrate in real time while recording cosmic ray tracks. We demonstrate the ability of a Gaussian Process to predict the gain correction calibration factor used to determine a high voltage setting that will stabilize the CDC gain in response to changing environmental conditions; this is in contrast to the traditional, computationally expensive method of calibrating raw data after data collection is complete.

1 Introduction

The Gluonic Excitations (GlueX) detector [1] is located at the Thomas Jefferson National Accelerator Facility. It is a complex detector comprised of several sub-detector systems shown in Fig. 1. The GlueX experiment typically collects data for a few months at a time. Each detector system is calibrated individually followed by multiple rounds of inter-dependent calibrations that are performed *after* data collection has completed. This process is repeated for each independent 'run', during which

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

data were collected for up to 2 hours. Proper calibration of the detector requires significant time from the individual detector experts and computing resources (on the order of months of calendar time). In contrast, machine learning (ML) can approximate these calibrations in seconds while simultaneously taking data. Adoption of a ML approach for detector control requires a method that (1) satisfies accuracy requirements, (2) reduces the time and compute resources to calibrate data, and (3) accurately estimates the model uncertainty.



(a)

(b)

Figure 1: The GlueX detector system. (a) Schematic of the GlueX spectrometer with labels for most of the major detector systems. (b) The GlueX Spectrometer at Jefferson Lab with a human in front for scale.

2 The Central Drift Chamber and Traditional Calibration

The CDC [2] is a 150 cm long, 120 cm diameter, cylindrical straw tube drift chamber used for charged particle tracking and identification. A 50:50 Ar:CO_2 gas mixture flows through the chamber. Charged particles passing through the chamber ionize the gas along their path, and the liberated avalanche of electrons drifts towards the anode wires, which are maintained at high positive voltage (HV), and creates a pulse 'hit' on the wire which is read out by the detector's electronics. The pattern of hits on the 3522 wires is used to reconstruct the track(s) of the particle(s). The pulse amplitude is used together with the path length of the track through the straw tubes of the CDC to calculate dE/dx, a measure of energy deposited per unit length as a particle travels through the detector, an important feature of particle identification. The gas gain, the ratio of final to initial number of electrons in the avalanche, is sensitive to the gas density. The CDC is operated at atmospheric pressure, thus the variation in gain with atmospheric pressure and gas temperature can be accounted for via calibration. This calibration is usually performed after data taking and relies on reconstructing particle tracks, a slow and computationally expensive process [1].

3 Calibration and control of the CDC using a Gaussian process regression

3.1 Central Drift Chamber data

Data were extracted from two sources: the Experimental Physics Industrial Controls System (EPICS) [3], and the Hall-D Calibration Constants Database (CCDB) [4]. Data specific to the CDC were extracted from the EPICS archive for use as **input** variables. The CCDB contains the calibration values for all sub-detector elements and is the source of our **target** value, the Gain Correction Factor (GCF), for each experimental run. Features from EPICS are readily available throughout data-taking, which is important for online control.

Over 120 variables were retrieved from EPICS for a trailing 15-second look-back time before the start of each run. Some features, e.g. particle momentum, while correlated with gain, were excluded as their calculation requires computationally expensive track reconstruction, making them unsuitable for real-time control. Additionally, features from other sub-detectors were found to be correlated with GCF, but were excluded to ensure the CDC control system is self-contained, ensuring no data from other detectors are used. This is especially important in case some sub-detectors may not be used in future experiments.

Shapley [5] feature importance evaluation identified three variables to predict GCF. Reassuringly, these were consistent with known drift chamber behavior: the mean atmospheric pressure, the maximum gas temperature, and the sum of the maximum gas temperatures and mean current drawn by the innermost high voltage (HV) boards on the CDC (a CDC-specific proxy for the rate of charged particles passing through the detector). For training purposes, data were used from experiments performed during 2020 and 2021. The data included 536 and 65 runs from the 2020 and 2021 experiments.

Balancing the distributions of atmospheric pressure between training and test data sets was motivated by the scarcity of extreme values in atmospheric pressure during data taking. A low pressure threshold of $<=$ 99.27 kPa and a high pressure threshold of $>=$ 101.57 kPa provided similar pressure distributions and adequate diversity of gas temperature and HV board current values for low, medium and high pressure data within the *pressure balanced* training and test data sets. Data were split into 80/20 pressure-balanced training (480 runs) and test (121 runs) datasets. Input features and target values were set to zero mean and unit variance using Scikit-Learn's [6] *StandardScaler*.

3.2 Using Gaussian process regression to predict the Gain Correction Factor

A **Gaussian process** (GP) with a single output parameter infers a latent function $\mathbb{R}^d \to \mathbb{R}$ to map training input X and corresponding targets y [7]. Where X is the matrix of N observations of the selected features, $x_n \in \mathbb{R}^p$, and where n is the n^{th} observation and p is the number of features, in our case three.

The mapping uses a selected covariance function to control the smoothness of the mapping output from one training point to the next closest training point. Here the covariance function k is applied to N measurements x to produce the $N \times N$ Gram matrix, $K_{n,n'} = k(x^n, x^{n'})$. [7]. In our case, $k(\cdot, \cdot)$ is a **sum** kernel of Scikit-learn's [6] Radial Basis Function (RBF), as shown in Eq. 1, and White Noise (WN), as shown in Eq. 2. The RBF explains the covariance of the data, where l is the learned length scale parameter used to scale the difference in distance between training observations.

$$k_{RBF}(x, x') = \exp\left(-\frac{(x - x')^2}{2l^2}\right) \tag{1}$$

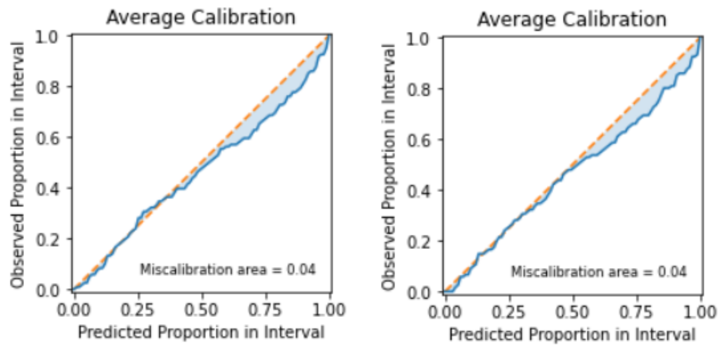
The WN is used to describe the combined global noise in the data, where σ^2 is the variance of the noise, and I_n is the identity matrix.

$$k_{WN}(x, x') = \sigma^2 I_n, \tag{2}$$

Percent error was selected as the evaluation metric based on the use of GCF to scale CDC gain. [6]'s *GaussianProcessRegressor* was trained for 100 iterations to learn the RBF length scale and a white noise parameter. The mean percent error of the test set runs was 0.8% (0.59% for training data). For reference, our target was 5% error. The maximum percent error of the test set runs was 5.9% (2.6% for training data). 36 of the 121 test runs had an error over 1%. Only eight test runs had error over 2%.

We evaluated two versions of the kernel, one with a single length scale that accounts for all features (isotropic), and a second with three length scales, one per feature (anisotropic). The anisotropic kernel had negligible effect on either max (anisotropic 5.8% v. isotropic 5.9%), or mean absolute percent error (anisotropic 0.8% v. isotropic 0.8%), most likely due to the strong known relation of atmospheric pressure (our first feature) to GCF. This is demonstrated by the GP's learned length scales; the first length scale of the anisotropic kernel, 1.400, is consistent with the single length scale, 1.412, of the isotropic kernel. The isotropic kernel was used for the cosmic ray experiment.

The degree to which a predicted uncertainty matched the true underlying uncertainty in the data was evaluated using standard metrics from the Uncertainty Toolbox[8]. Fig. 2 shows [8]'s "Average Calibration" plot for both kernel types. Our model was marginally underconfident with a 4% miscalibration area for both kernels. R-square, Root Mean Square Error (RMSE), Mean Absolute Calibration Error (MACE), and Root Mean Square Calibration Error (RMSCE) are shown in Table 1.



(a) (b)

Figure 2: for the isotropic kernel (a), and anisotropic kernel (b), shows the predicted proportion of the test data expected to lie inside the prediction interval (x-axis) and the proportion of the test data observed inside the prediction interval on the y-axis. For example, the 0.75 prediction interval aims to include observed values 75% of the time.

Table 1: Uncertainty Toolbox accuracy and average calibration metrics for GP models. Metrics were similar for both the isotropic and the anisotropic kernel.

RBF kernel (length scale(s))	noise kernel variance	R^2	RMSE	MACE	RMSCE
isotropic (1.412)	0.0154	0.97	0.002	0.040	0.051
anisotropic (1.400,1.17,1.71)	0.0153	0.97	0.002	0.038	0.049

4 Cosmic Ray Test Setup, Autonomous CDC Control, and Results

To test the performance of the model, cosmic ray data were collected for 2 weeks. This made use of cosmic rays passing through the detector instead of charged particles emerging from collisions between the photon beam and target. As such, we set the HV board current given to the model to be 9.0 nA in accordance with the mode of HV board current in our training data. In order to compare the behavior of the CDC operating at a constant HV to a GP-tuned HV, we used software to divide the operation of the CDC into two sections: one was held at 2130 V and the other was updated to the HV setting determined using the GP prediction every five minutes.

Using the predicted GCFs, the HV setting was determined from a second order polynomial fit to the HV as a function of the peak amplitude relative to that obtained at the nominal operating voltage. By exploiting the empirically derived correlation of peak height and HV we could use the GP-predicted GCF and an "ideal" GCF to obtain a HV setting that stabilized the CDC gain.

After retrieving the input features from EPICS, a prediction is returned in 3 milliseconds utilizing an x86_64 CPU. GCF inference was conducted every minute. The HV was adjusted every 5 minutes. The atmospheric pressure does not usually change fast enough for more frequent measurements to be interesting, and also we needed to collect sufficient data at each setting to obtain the GCF. All inferences and HV settings were logged for further analysis.

In Fig. 3, a subset of the cosmic ray data corresponding to the largest change and subsequent stabilization of the atmospheric pressure is shown. We compare traditionally calculated GCF for the GP-controlled (blue) and constant HV (orange) sections of the CDC as a function of the event number. The orange distribution displays the usual behavior of the CDC at a constant HV, which would be compensated for via calibration after data taking. For the model-controlled side of the CDC, the CDC is unaffected by changes in environmental and experimental conditions.

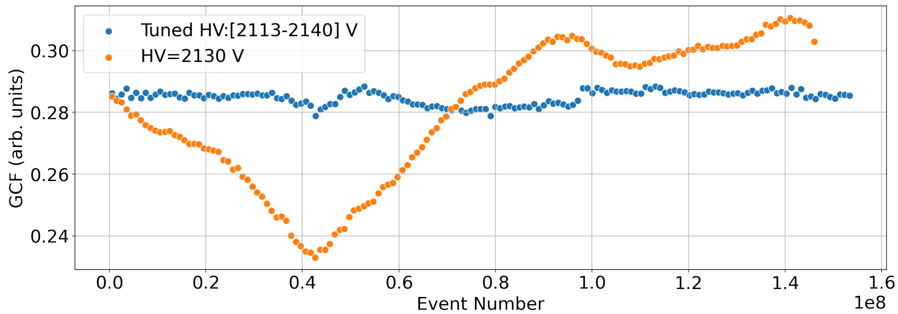


Figure 3: Traditionally calculated GCF for the model-controlled (blue) and constant 2130 HV (orange) sections of the CDC for a subset of the cosmic ray experiment corresponding to the largest change in atmospheric pressure.

5 Ongoing work and Conclusions

Developments are underway to improve the system's robustness to out-of-domain inferences using a learned prior and evaluate the assumption of homogeneous noise. During the course of this work, we started to explore the possibilities for the control software to make an action based on the uncertainty quantification from the model. This will be useful for deployment in experiments using the accelerator, where the CDC's output is vital to physics analysis but was unnecessary for this cosmic ray experiment.

In this work we have shown that a simple Gaussian process model with uncertainty quantification was accepted for use and implemented to control the HV of the GlueX Central Drift Chamber, an expensive particle detector critical to experiments that take place in Hall D at the Thomas Jefferson National Accelerator Facility. This work represents a dramatic shift from calibrating terabytes of data *after* an experiment has completed to near-real-time calibrations, resulting in significant time and computational savings. This system is expected to be directly applicable to other detector systems used in Nuclear and High Energy Physics experiments.

6 Impact Statement

We hope this demonstration of a data-driven technique that (1) satisfies accuracy requirements, (2) reduces the time and compute resources to calibrate data, and (3) accurately estimates uncertainty inspires particle detector experts and experimentalists who must calibrate massive amounts of raw sensor data to investigate machine learning methods to replace computationally expensive and time intensive traditional methods. *Trustworthy* data-driven methods are required for physicists to accept data-driven techniques as a replacement for traditional calculations, and trustworthy machine learning must include the ability of a model to say, "I don't know," which is possible with well-calibrated UQ. As [9] states, "Predictions without UQ are neither predictions nor actionable." Additionally, high performance computing resources are in demand, and calibration-via-online-control reduces the computational and energy footprint of the resulting physics analyses.

7 Acknowledgements

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.

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`.

References

[1] S. Adhikari et al. The GlueX beamline and detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 987:164807, jan 2021. doi:10.1016/j.nima.2020.164807.

[2] N.S. Jarvis, C.A. Meyer, B. Zihlmann, M. Staib, A. Austregesilo, F. Barbosa, C. Dickover, V. Razmyslovich, S. Taylor, Y. Van Haarlem, G. Visser, and T. Whitlatch. The Central Drift Chamber for GlueX. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 962:163727, may 2020. doi:10.1016/j.nima.2020.163727.

[3] AJ Kozubal, LR Dalesio, JO Hill, and DM Kerstiens. Experimental Physics and Industrial Control System. In *ICALEPCS89 Proceedings*, volume 288, 1989.

[4] JeffersonLab. Jefferson Lab Calibration Constants Database Wiki. URL: https://github.com/JeffersonLab/ccdb/wiki.

[5] Sean Gillies et al. Shapely: manipulation and analysis of geometric objects, 2007–. URL: https://github.com/shapely/shapely.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[7] David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[8] Youngseog Chung, Ian Char, Han Guo, Jeff Schneider, and Willie Neiswanger. Uncertainty toolbox: an open-source library for assessing, visualizing, and improving uncertainty quantification. *arXiv preprint arXiv:2109.10254*, 2021.

[9] Edmon Begoli, Tanmoy Bhattacharya, and Dimitri Kusnezov. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence*, 1(1):20–23, 2019.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] As of May 2022 the answer is 'No'; however, we have started the process of asking the GlueX collaboration for permission to share the data and code, and hope to be able to share by the camera-ready deadline in October.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We cite scikit-learn.

(b) Did you mention the license of the assets? [N/A]

(c) Did you include any new assets either in the supplemental material or as a URL? [N/A]

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] Our work does not involve human subjects.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] Our work does not involve human subjects.

5. If you used crowdsourcing or conducted research with human subjects...

   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] Our work does not involve human subjects.

   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A] Our work does not involve human subjects.

   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] Our work does not involve human subjects.