Deep Learning with Quantified Uncertainty for Free Electron Laser Scientific Facilities

Lipi Gupta [1] **Aashwin Mishra** [2] **Auralee Edelen** [2]

Abstract

Particle accelerators are essential tools for scientific research, providing charged particle beams with different parameters (e.g. beam energies and durations) for experiments, accomplished by changing accelerator settings in a process called tuning. Tuning can be challenging and time consuming as accelerators have thousands of components, inherent uncertainties, and temporal drift. While machine learning models can aid rapid customization of beams, quantified predictive uncertainties are essential as accelerators are high-regret and safety-critical systems. We address the problem of obtaining uncertainties for models of a noisy, high-dimensional, nonlinear accelerator system using quantile regression neural networks (QRNNs) and Bayesian Neural Networks (BNNs). The QRNN models provide reliable point and interval predictions. BNNs are more computationally expensive and exhibit sensitivity to outliers.

1. Introduction and Motivation

Accelerator physicists and operators at light source facilities provide powerful X-rays to help advance research in manifold disciplines. At the SLAC Linac Coherent Light Source (LCLS) (Bostedt et al., 2013) electron beams are accelerated to the speed of light, to engender X-rays in a Free Electron Laser (FEL). The LCLS facilitates thousands of scientific experiments a year, and many of these have fundamentally improved human understanding of processes such as photosynthesis (Young et al., 2016), electron-phonon interactions (Jiang et al., 2016; Singer et al., 2016), molecular interactions in drug delivery (Colletier et al., 2016), besides others. At light source facilities, each experiment has access to the accelerator for few hours for their research. However, the required beam for each experiment is diametrically different, with respect to the beam energy, shape, duration, etc. Thus, for successful execution and completion of scientific studies, rapid and accurate beam customization is critical. As accelerators have high-dimensional parameter spaces, manual optimization by operators is time consuming and leads to sub-optimal final settings. Machine learning based surrogate models trained on archived data can aid in rapid optimization, design of novel experimental setups, etc. Numerous sources of uncertainty (misalignments in accelerator components, intermittent anomalous behavior and erroneous signals) are present in accelerator systems. This is in addition to noise in system measurements, introducing aleatoric uncertainty. These are exacerbated by temporal drift in the system leading to changes in beam behavior wherein out of distribution robustness is required from the surrogate model. To this end, we investigate quantile regression neural networks (QRNN) and Bayesian neural networks (BNNs) as approaches for producing robust and uncertainty aware predictions of the X-ray pulse energy. We assess model performance on noisy, high-dimensional data covering a broad range of operating configurations.

In this study, the learning dataset was curated from the automated LCLS archive, specifically targeting periods of machine tuning. The full dataset includes 286,923 samples. Each sample consists of 76 scalar inputs, consisting of focusing magnet strengths, accelerating cavity phases and amplitudes, and uncontrolled variables that can change over time. The output is a single scalar value of the photon pulse energy. As the data spans several years of operation and operating modes, there are several significant sources of irreducible uncertainty. One of these is drift in the system's response given input variables over time (due to, for example, equipment aging, part replacement, or other environmental changes). Another source of uncertainty is the sparse sampling of some of the inputs. Additionally, pulse energy is an inherently noisy measurement, due to both jitter in the beam parameters and noise in the detector response. Analysing steady-state operation, the RMS noise in the measured data can be up to 0.3 mJ. The detector also has a sensitivity threshold below which it is not effective, and may even produce a false positive signal when no beam is present. To assess sensitivity, we ascertain performance on datasets where erroneous measurements and outliers are

[1]University of Chicago, Department of Physics, Chicago, IL [2]SLAC National Accelerator Laboratory, Menlo Park, CA. Correspondence to: Lipi Gupta <lipigupta@uchicago.edu>.

Preliminary work. Under review by the ICML 2021 Workshop on Uncertainty and Robustness in Deep Learning. Do not distribute.

present, and on datasets where these are removed.

2. Quantile Regression Neural Networks

Classical deterministic neural networks assume the noise in the data to be Gaussian and homoscedastic. However, this may not be true in real data sets, wherein the target data may be severely skewed, or strictly non-negative. In such cases, a robust alternative is to estimate the point predictions of different quantiles, using sets of quantile neural networks. For a random variable X with a cumulative distribution function, $F_X(x) = P(X \leq x)$, the τ-quantile, q, is given by $min(qF_X(q)) = \tau$. Such quantile regression approaches are more robust as they invoke no assumptions about the parametric form of the final distribution. Estimated quantiles are equivariant to monotonic transformations, thus the quantiles for the target can be used to predict quantiles of derived quantities of interest. Quantile regression approaches apply asymmetric weights to positive and negative errors, using the tilted loss metric. Here we use QRNNs on the measured FEL data set. The model consists of an input layer for all 76 scalar inputs, 8 fully connected hidden layers which decrease in the number of neurons from 80 to 10, by a factor of 10 neurons for each subsequent layer. The activation function for each layer was the hyperbolic tangent function. The final output layer predicts the photon energy. The median prediction (50% quantile), 97.5% quantile prediction, and 2.5% quantile prediction, were each fit using independent models. A custom loss function was written for the tilted loss, and each model was optimized using Adam (Kingma & Ba, 2015) for its given quantile, along with early stopping. The basic QRNN model was trained on 80% of the full dataset, with the remaining 20% split equally into a validation and test set. The results on the test set are shown in Fig. 1. The coverage for the 95% interval of measurements from the test set was 92.84%, which indicates good calibration. The mean absolute error (MAE) is 0.13 mJ. This performance was obtained in the presence of outliers and erroneous measurements below the sensitivity threshold of the detector. We observe that the QRNNs are robust to these outliers and able to capture trends despite the presence of noise. These results suggest that QRNN models are good alternatives for online prediction and sequential retraining.

As a rigorous test for temporal changes in machine behavior, we removed sections of time-ordered data as additional test sets. These results are shown in Fig. 2 and Fig. 3. The learning data is shown in black, and this was spilt into training and validation sets which were held constant for each set of QRNNs. For the models shown in Fig. 2, the output data spanned the full range of potential photon energies, 0 to 5 mJ. For the models shown in Fig. 3, photon energies below 0.2 mJ were removed. First, a portion of chronological data in which the photon energy measurements are fairly stable throughout the segment was chosen. By observing the performance on this portion of the data, the ability for the model to predict on stable operating parameters can be observed. Another portion in which the measurement value changes significantly (more than 50%) was also chosen. Similarly, the ability for the model to predict different operational energies can be observed in regions where the measured values change drastically. The coverage was calculated on the regions of data removed from training. For Fig. 2, the coverage for both removed portions of data was 53.33%, and for Fig. 3 the coverage was 42.83%. While this demonstrates that the confidence interval is not capturing the full extent of possible measured values, the median prediction performs well. For Fig. 2, the MAE is 0.34 mJ, which is commensurate with the RMS noise in the measured data (and thus is about as good as we expect to be able to predict the output). Similarly for the models in Fig. 3, MAE is 0.60 mJ. The results suggest that these models are able to capture the general trends present in the data, in spite of the high-dimensional, noisy, and sparse dataset. Large accelerators like the LCLS are historically difficult to model empirically, thus this is a very encouraging result.

3. Bayesian Neural Networks

In contrast to deterministic neural networks, in Bayesian Neural Networks the weights and biases are assumed to be random variables with corresponding probability distributions. As the Bayesian Neural Network's parameters are random variables, they engender predictions with associated probability distributions. We utilize approximate variational inference, with the Bayes By Backprop algorithm (Blundell et al., 2015), along with the Adam optimizer wherein the rates are set using cross-validation. The network architectures are selected using Bayesian Optimization (O'Malley et al., 2019). The network architecture had 6 dense layers with ReLU activations, and weights and biases are initialized with standard normal priors.

Despite hyperparameter optimization, the BNN was not able to accurately or consistently predict the true value. Further, the interval predicted by the BNN does not provide 95% coverage and is substantially undercalibrated. While Fig. 4 does display that the BNN is beginning to capture trends in the data, it is unsatisfactory compared to the QRNN result shown in Fig. 1. Shown in Fig. 5 is the same range of data as is shown in Fig. 3. While portions of data were removed from training for the models shown in Fig. 3, no portion of the data was explicitly excluded from being sampled in the train/test split to train the BNN. Despite all of the data being available for training, the BNN performance for the last 10,000 samples (which were removed from training for the models in 3) was worse. The coverage for these samples was 17.46%, and the MAE was 0.77mJ. The results reflect

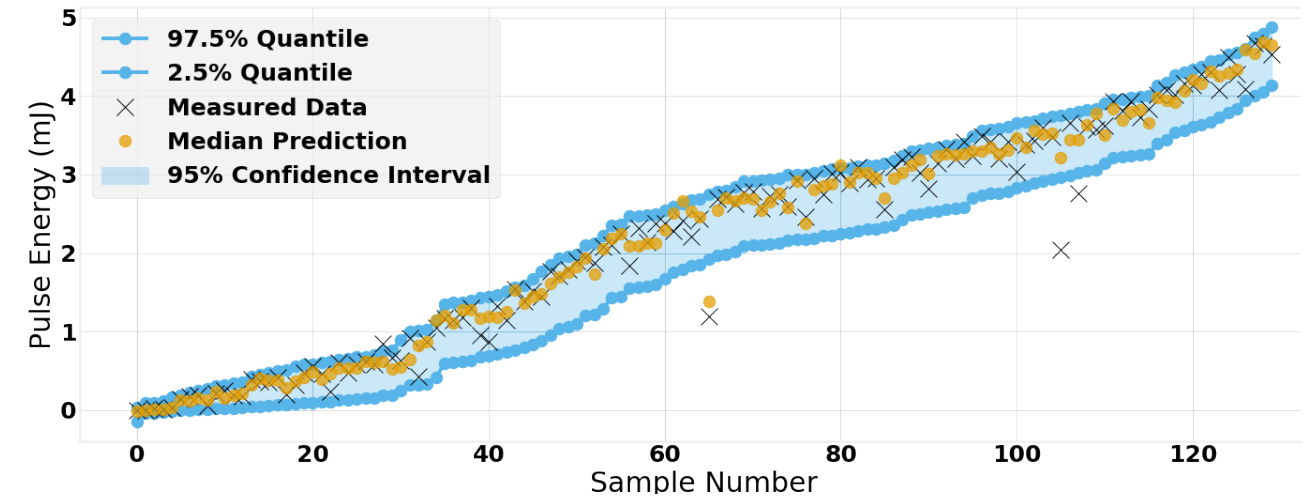

Figure 1. The QRNN model results for the test set. The prediction coverage for these models is 92.84%. The test set mean absolute error (MAE) is 0.13 mJ.

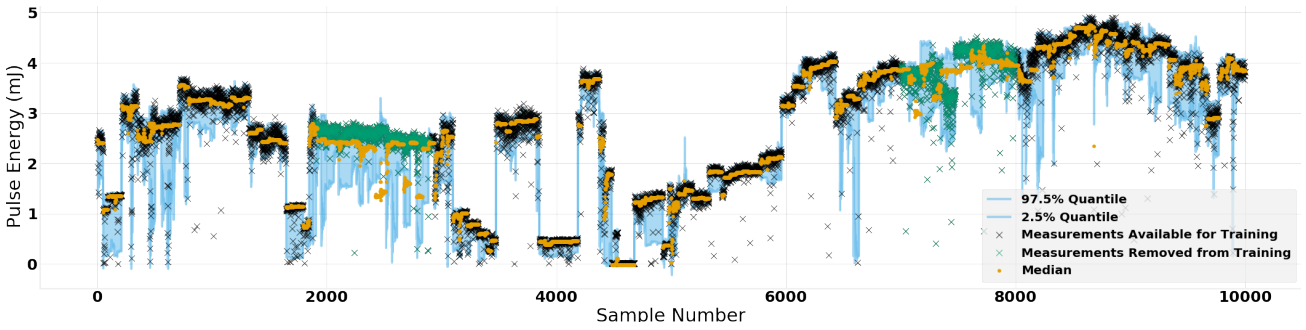

Figure 2. QRNN predictions on holdout chronological data (coverage: 53.35%, MAE: 0.34 mJ)

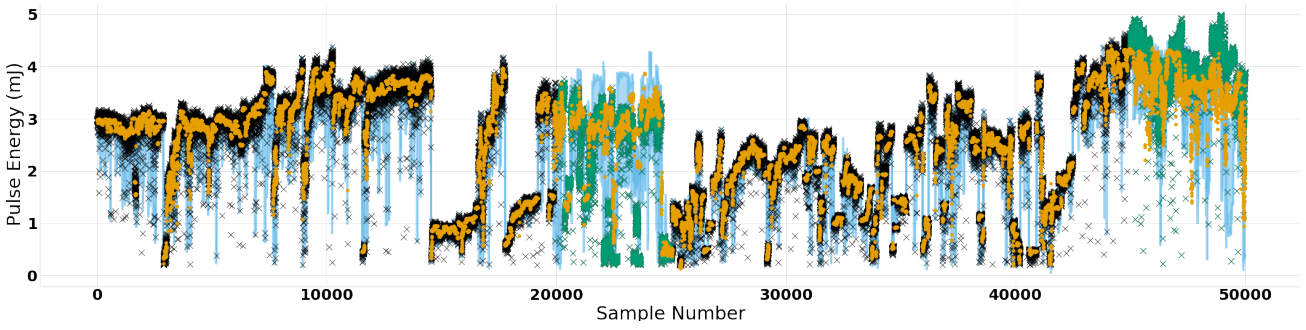

Figure 3. QRNN predictions on holdout spans of chronological data, with legend provided in Fig. 2. (coverage: 42.83%, MAE: 0.60) The training data spans photon energies of 0.2 - 5 mJ. The cutoff at 0.2 mJ removes low-energy samples, where detector noise dominates signals.

an inherent schism in BNN inference at present. While sampling based approaches for inference like Hamiltonian Monte Carlo provide reliable solutions, they do not scale to high dimensions. Approximate variational inference approaches have better scaling. However, they are sensitive to hyperparameter choice and may rely on substantial ap-

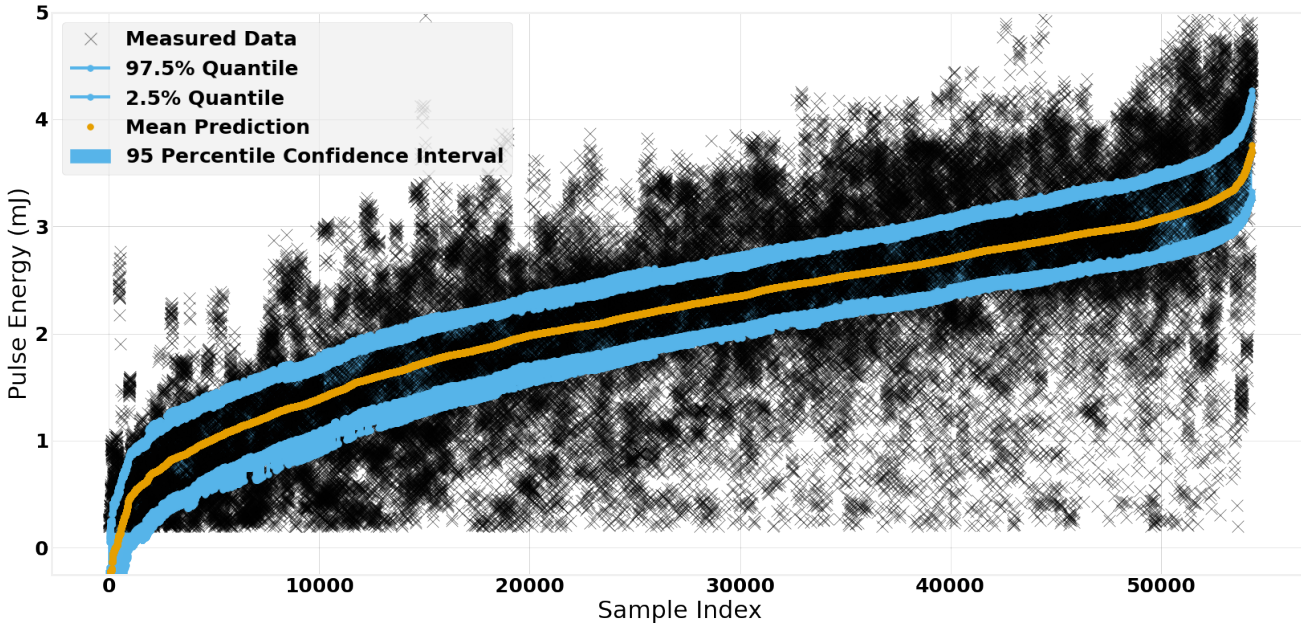

Figure 4. BNN results on the test set, with coverage at 44.64%. The MAE is 0.56 mJ.

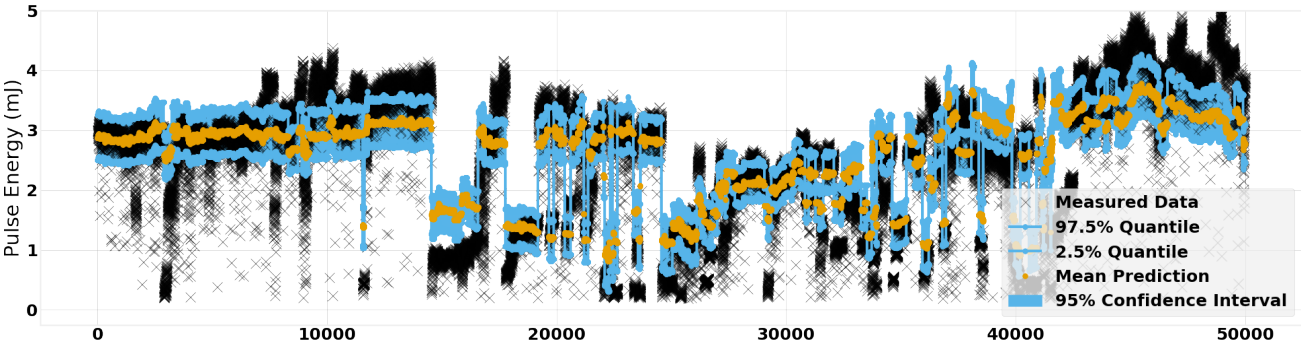

Figure 5. Range of data with the BNN median prediction and confidence intervals (time-ordered sample numbers).

proximations to achieve scalability, resulting in limited or even unreliable uncertainty estimates (Foong et al., 2019a; Antorán et al., 2020; Foong et al., 2019b).

4. Summary & Conclusions

For high-regret and safety critical systems such as particle accelerators, surrogate models need to be robust and uncertainty aware. We examined the ability of quantile regression and Bayesian neural networks to provide predictions and uncertainty estimates for the X-ray pulse energy of the LCLS, given historical data spanning several years, a wide variety of operating modes. Owing to the data sparsity, high dimensionality, inherent noise and system drift, generating a robust, accurate and uncertainty aware model has been a historically challenging problem. We find that QRNNs provide reliable predictions with prediction intervals that are well calibrated, and show robustness to outliers and anomalous behavior in the data. This is essential for continuous re-training of models. While BNNs have shown excellent results on benchmark datasets, we find that in such complex problems approximate variation inference approaches may have limitations. All code used in this study is available at: https://github.com/lipigupta/FEL-UQ.

References

Antorán, J., Allingham, J. U., and Hernández-Lobato, J. M. Depth uncertainty in neural networks. *arXiv preprint arXiv:2006.08437*, 2020.

Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D. Weight uncertainty in neural networks. In *International Conference on Machine Learning*, pp. 1613–1622, 2015.

Bostedt, C., Bozek, J. D., Bucksbaum, P. H., Coffee, R. N., Hastings, J. B., Huang, Z., Lee, R. W., Schorb, S., Corlett, J. N., Denes, P., Emma, P., Falcone, R. W., Schoenlein, R. W., Doumy, G., Kanter, E. P., Kraessig, B., Southworth, S., Young, L., Fang, L., Hoener, M., Berrah, N., Roedig, C., and DiMauro, L. F. Ultra-fast and ultra-intense x-ray sciences: first results from the linac coherent light source free-electron laser. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 46(16):164003, aug 2013. doi: 10.1088/0953-4075/46/16/164003. URL https://doi.org/10.1088%2F0953-4075%2F46%2F16%2F164003.

Colletier, J.-P., Sawaya, M. R., Gingery, M., Rodriguez, J. A., Cascio, D., Brewster, A. S., Michels-Clark, T., Hice, R. H., Coquelle, N., Boutet, S., Williams, G. J., Messerschmidt, M., DePonte, D. P., Sierra, R. G., Laksmono, H., Koglin, J. E., Hunter, M. S., Park, H.-W., Uervirojnangkoorn, M., Bideshi, D. K., Brunger, A. T., Federici, B. A., Sauter, N. K., and Eisenberg, D. S. De novo phasing with x-ray laser reveals mosquito larvicide binab structure, 2016. URL https://doi.org/10.1038/nature19825.

Foong, A. Y., Burt, D. R., Li, Y., and Turner, R. E. Pathologies of factorised gaussian and mc dropout posteriors in bayesian neural networks. *CoRR, abs/1909.00719*, 2019a.

Foong, A. Y., Li, Y., Hernández-Lobato, J. M., and Turner, R. E. 'in-between'uncertainty in bayesian neural networks. *arXiv preprint arXiv:1906.11537*, 2019b.

Jiang, M. P., Trigo, M., Savić, I., Fahy, S., Murray, É. D., Bray, C., Clark, J., Henighan, T., Kozina, M., Chollet, M., Glownia, J. M., Hoffmann, M. C., Zhu, D., Delaire, O., May, A. F., Sales, B. C., Lindenberg, A. M., Zalden, P., Sato, T., Merlin, R., and Reis, D. A. The origin of incipient ferroelectricity in lead telluride, 2016. URL https://doi.org/10.1038/ncomms12291.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y. (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. Keras Tuner. https://github.com/keras-team/keras-tuner, 2019.

Singer, A., Patel, S. K. K., Kukreja, R., Uhlíř, V., Wingert, J., Festersen, S., Zhu, D., Glownia, J. M., Lemke, H. T., Nelson, S., Kozina, M., Rossnagel, K., Bauer, M., Murphy, B. M., Magnussen, O. M., Fullerton, E. E., and Shpyrko, O. G. Photoinduced enhancement of the charge density wave amplitude. *Phys. Rev. Lett.*, 117:056401, Jul 2016. doi: 10.1103/PhysRevLett.117.056401. URL https://link.aps.org/doi/10.1103/PhysRevLett.117.056401.

Young, I. D., Ibrahim, M., Chatterjee, R., Gul, S., Fuller, F. D., Koroidov, S., Brewster, A. S., Tran, R., Alonso-Mori, R., Kroll, T., Michels-Clark, T., Laksmono, H., Sierra, R. G., Stan, C. A., Hussein, R., Zhang, M., Douthit, L., Kubin, M., de Lichtenberg, C., Vo Pham, L., Nilsson, H., Cheah, M. H., Shevela, D., Saracini, C., Bean, M. A., Seuffert, I., Sokaras, D., Weng, T.-C., Pastor, E., Weninger, C., Fransson, T., Lassalle, L., Bräuer, P., Aller, P., Docker, P. T., Andi, B., Orville, A. M., Glownia, J. M., Nelson, S., Sikorski, M., Zhu, D., Hunter, M. S., Lane, T. J., Aquila, A., Koglin, J. E., Robinson, J., Liang, M., Boutet, S., Lyubimov, A. Y., Uervirojnangkoorn, M., Moriarty, N. W., Liebschner, D., Afonine, P. V., Waterman, D. G., Evans, G., Wernet, P., Dobbek, H., Weis, W. I., Brunger, A. T., Zwart, P. H., Adams, P. D., Zouni, A., Messinger, J., Bergmann, U., Sauter, N. K., Kern, J., Yachandra, V. K., and Yano, J. Structure of photosystem ii and substrate binding at room temperature, 2016. URL https://doi.org/10.1038/nature20161.