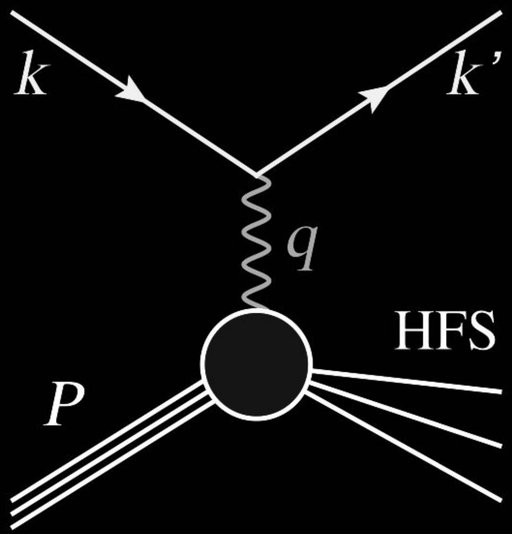
How uncertain your DIS events are? Event-level UQ with Bayesian DL



Cristiano Fanelli, James Giroux

Probing the frontiers of Nuclear physics with AI at EIC, Sep 25, 2023

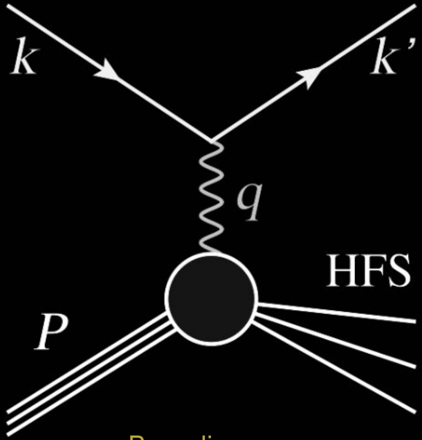
WILLIAM & MARY
CHARTERED 1693

Outline

- Introduction to DL for DIS
  - Previous works
- Quantify uncertainties on inferred kinematics
  - Leveraging UQ at the event-level
- Conclusions

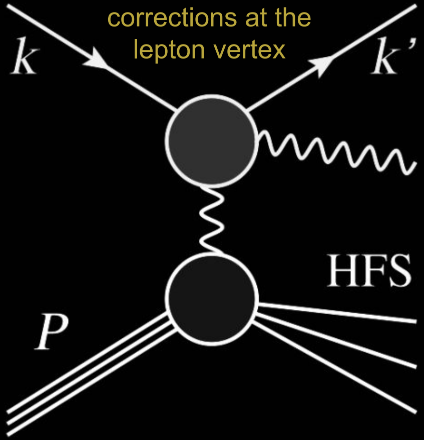
Deep Inelastic Scattering

DIS is governed by the four-momentum transfer squared of the exchanged boson Q^2, the inelasticity y, and the Bjorken scaling variable x.

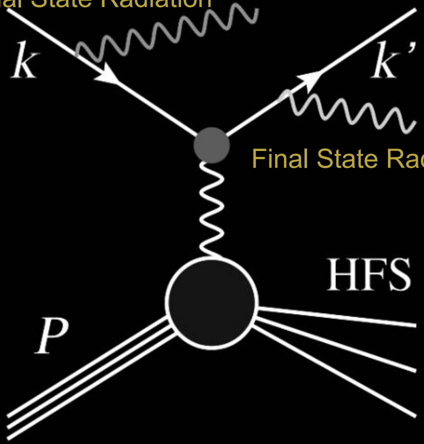


Born diagram

higher-order QED corrections at the lepton vertex



Initial State Radiation



Final State Radiation

These kinematic variables are related via the relation $Q^2 = sxy$, where s is the square of the center-of-mass energy.

$$s = (k+P)^2\,, \qquad Q^2 = -q^2\,, \qquad y = \frac{q \cdot P}{k \cdot P}\,, \qquad \text{and} \qquad x = Q^2/(sy)\,.$$

M. Arratia, D. Britzger, O. Long, B. Nachman, et al., “Reconstructing the kinematics of deep inelastic scattering with deep learning",
NIM-A 1025 (2022): 166164

Previous works

- Conservation of momentum and energy over constrain the DIS kinematics and leads to a freedom to calculate x, Q^2, y from measured quantities
- Each method has advantages and disadvantages, and no single approach is optimal over the entire phase space. Each method exhibits different sensitivity to QED radiative effects
- Once (real) higher-order QED effects are considered, the various methods yield different results and the calculated quantities for Q^2, y and x are not representative for the γ/Z + p scattering process at the hadronic vertex.

Summary of basic reconstruction methods

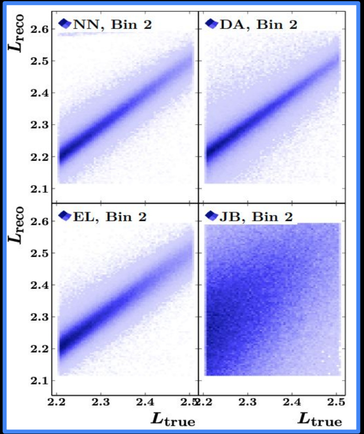
Method name	Observables	y	Q^2	$x \cdot E_p$
Electron (e)	$[E_0,E,\theta]$	$1-\frac{\Sigma_e}{2E_0}$	$\frac{E^2\sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
Double angle (DA) [6, 7]	$[E_0,\theta,\gamma]$	$\frac{\tan\frac{\gamma}{2}}{\tan\frac{\gamma}{2}+\tan\frac{\theta}{2}}$	$4E_0^2\cot^2\frac{\theta}{2}(1-y)$	$\frac{Q^2}{4E_0 y}$
Hadron (h, JB) [4]	$[E_0,\Sigma,\gamma]$	$\frac{\Sigma}{2E_0}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
ISigma (IΣ) [9]	$[E,\theta,\Sigma]$	$\frac{\Sigma}{\Sigma+\Sigma_e}$	$\frac{E^2\sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
IDA [7]	$[E,\theta,\gamma]$	y_{DA}	$\frac{E^2\sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
$E_0E\Sigma$	$[E_0,E,\Sigma]$	y_h	$4E_0E-4E_0^2(1-y)$	$\frac{Q^2}{2\Sigma}$
$E_0\theta\Sigma$	$[E_0,\theta,\Sigma]$	y_h	$4E_0^2\cot^2\frac{\theta}{2}(1-y)$	$\frac{Q^2}{2\Sigma}$
$\theta\Sigma\gamma$ [8]	$[\theta,\Sigma,\gamma]$	y_{DA}	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
Double energy (A4) [7]	$[E_0,E,E_h]$	$\frac{E-E_0}{(xE_p)-E_0}$	$4E_0y(xE_p)$	$E+E_h-E_0$
$E\Sigma T$	$[E,\Sigma,T]$	$\frac{\Sigma}{\Sigma+E\pm\sqrt{E^2+T^2}}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
E_0ET	$[E_0,E,T]$	$\frac{2E_0-E\mp\sqrt{E^2-T^2}}{2E_0}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{4E_0y}$
Sigma (Σ) [9]	$[E_0,E,\Sigma,\theta]$	$y_{\mathrm{I}\Sigma}$	$Q^2_{\mathrm{I}\Sigma}$	$\frac{Q^2}{4E_0y}$
eSigma ($e\Sigma$) [9]	$[E_0,E,\Sigma,\theta]$	$\frac{2E_0\Sigma}{(\Sigma+\Sigma_e)^2}$	$2E_0E(1+\cos\theta)$	$\frac{E(1+\cos\theta)(\Sigma+\Sigma_e)}{2\Sigma}$

Table 1. Summary of basic reconstruction methods that employ only three out of five quantities: E_0 (electron-beam energy), E and θ (scattered electron energy and polar angle), Σ and γ (longitudinal energy-momentum balance, $\Sigma = \sum_{\mathrm{HFS}}(E_i - p_{z,i})$, and the inclusive angle of the HFS). Alternatively, the A4 method makes use of the HFS total energy E_h. Shorthand notations are used

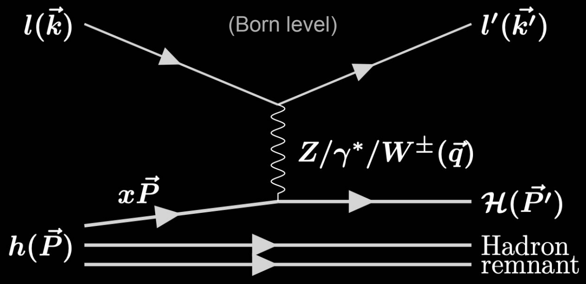
M. Arratia, D. Britzger, O. Long, B. Nachman, et al., “Reconstructing the kinematics of deep inelastic scattering with deep learning", NIM-A 1025 (2022): 166164

Deeply Learning DIS

Example in one specific bin



DIS fundamental process @EIC



DIS beyond the Born approximation has a complicated structure which involve QCD and QED corrections

- Use of DNN to reconstruct the kinematic observable x, Q^2, y in the study of neutral current DIS events at ZEUS and H1 experiments at HERA.
- The performance compared to electron, Jacquet-Blondel and the double-angle methods using data-sets independent of training
- Compared to the classical reconstruction methods, the DNN-based approach enables significant improvements in the resolution of Q^2 and x

Bin	Events	Resolution of $\log x$, $\times 10^3$		Resolution of $\log Q^2/1\,\mathrm{GeV}^2$, $\times 10^3$	
1	301780	**NN: 70**	EL: 83	**NN: 35**	EL: 35
		JB: 180	DA: 103	JB: 203	DA: 62
2	350530	**NN: 69**	EL: 82	**NN: 40**	EL: 43
		JB: 167	DA: 96	JB: 192	DA: 64
3	138456	**NN: 98**	EL: 130	NN: 55	**EL: 53**
		JB: 138	DA: 100	JB: 150	DA: 77
4	74844	**NN: 67**	EL: 84	**NN: 44**	EL: 46
		JB: 117	DA: 77	JB: 138	DA: 63
5	31043	**NN: 64**	EL: 91	**NN: 36**	EL: 41
		JB: 102	DA: 73	JB: 117	DA: 53
6	11475	**NN: 53**	EL: 79	**NN: 33**	EL: 36
		JB: 83	DA: 61	JB: 100	DA: 45
7	3454	**NN: 50**	EL: 69	**NN: 36**	EL: 38
		JB: 74	DA: 55	JB: 93	DA: 42
8	624	**NN: 36**	EL: 55	**NN: 33**	EL: 37
		JB: 67	DA: 45	JB: 95	DA: 41

Table 4: Resolution of the reconstructed kinematic variables in bins of x and Q^2. The resolution for x and Q^2 is defined as the RMS of the distributions $\log(x) - \log(x_{\mathrm{true}})$ and $\log(Q^2) - \log(Q^2_{\mathrm{true}})$ respectively.

M. Diefenthaler, A. Farhat, A. Verbytskyi, Y Xu. "Deeply learning deep inelastic scattering kinematics." EPJ C 82.11 (2022): 1064.

Description of input features

- Define variables to characterize the strength of QED radiation

$$p_T^{\text{bal}} = 1 - \frac{p_{T,e}}{T} = 1 - \frac{\Sigma_e \tan\frac{\gamma}{2}}{\Sigma \tan\frac{\theta}{2}} \quad \text{and} \quad p_z^{\text{bal}} = 1 - \frac{\Sigma_e + \Sigma}{2\, E_0}.$$

7 features to help indicate QED radiation in the event

- The values of p_T^{bal} and p_z^{bal}.
- The energy, η, and $\Delta\phi$ of the reconstructed photon in the event that is closest to the electron-beam direction, where $\Delta\phi$ is with respect to the scattered electron.
- The sum ECAL energy within a cone of $\Delta R < 0.4$ around the scattered electron divided by the scattered-electron track momentum.
- The number of ECAL clusters within a cone of $\Delta R < 0.4$ around the scattered electron.

+ additional 8 features

- Scattered-electron quantities $p_{T,e}$, $p_{z,e}$ and E.
- HFS four-vector quantities T, $p_{z,h}$ and E_h.
- $\Delta\phi(e,h)$ between the scattered electron and the HFS momentum vector.
- The difference $\Sigma_e - \Sigma$.

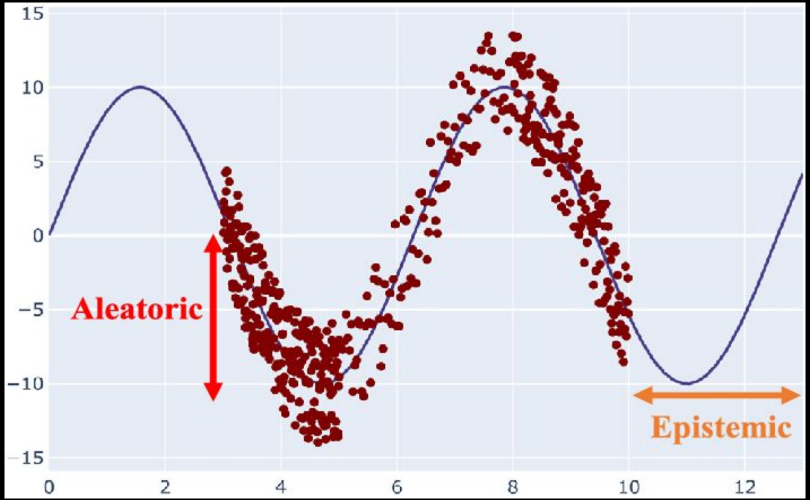
Tot. 15 input features

Dataset	Training Events	Validation Events	Testing Events	Size on Disk
H1	8.7×10^6	1.9×10^6	1.9×10^6	8 GB
Athena	14×10^6	3×10^6	3×10^6	14 GB

M. Arratia, D. Britzger, O. Long, B. Nachman, et al., “Reconstructing the kinematics of deep inelastic scattering with deep learning", NIM-A 1025 (2022): 166164

Epistemic vs Aleatoric

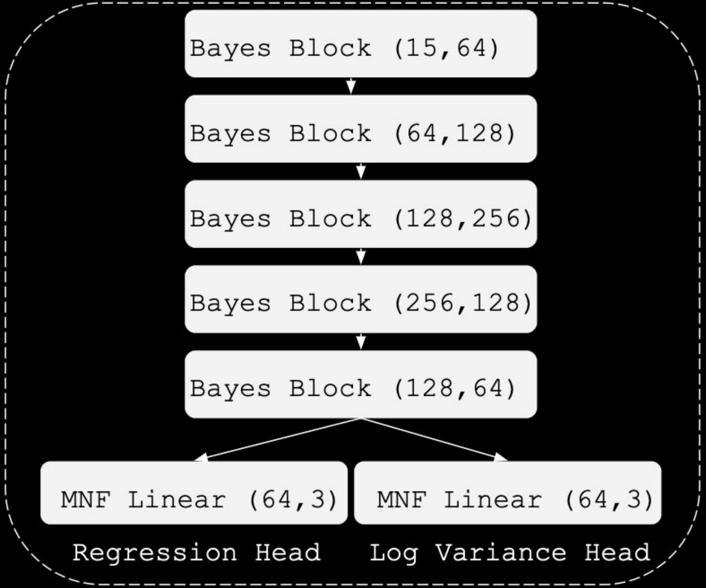
- Epistemic Uncertainty: This type of uncertainty arises from a lack of knowledge which is reflected in the effectiveness of the model in describing the data. It can be reduced as more information or data becomes available, and by improving the model. It can be affected by inaccuracy.
- Aleatoric Uncertainty: This uncertainty is due to inherent variability or randomness in a process or system and cannot be reduced by collecting more data. For example, even if we know the probability of getting heads when flipping a fair coin, the outcome of each individual flip is still uncertain.

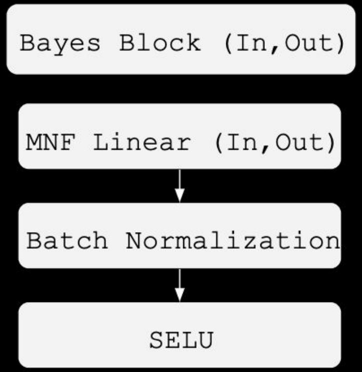


Abdar, Moloud, et al. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges." Information fusion 76 (2021): 243-297.

Network

BayesFlow-DIS





$$\mathcal{L}_{Reg.} = \frac{1}{2}\sum_i e^{-\mathbf{s_i}} \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|^2 + \mathbf{s}_i, \quad \mathbf{s}_i = \log \boldsymbol{\sigma}_i^2$$

epistemic aleatoric

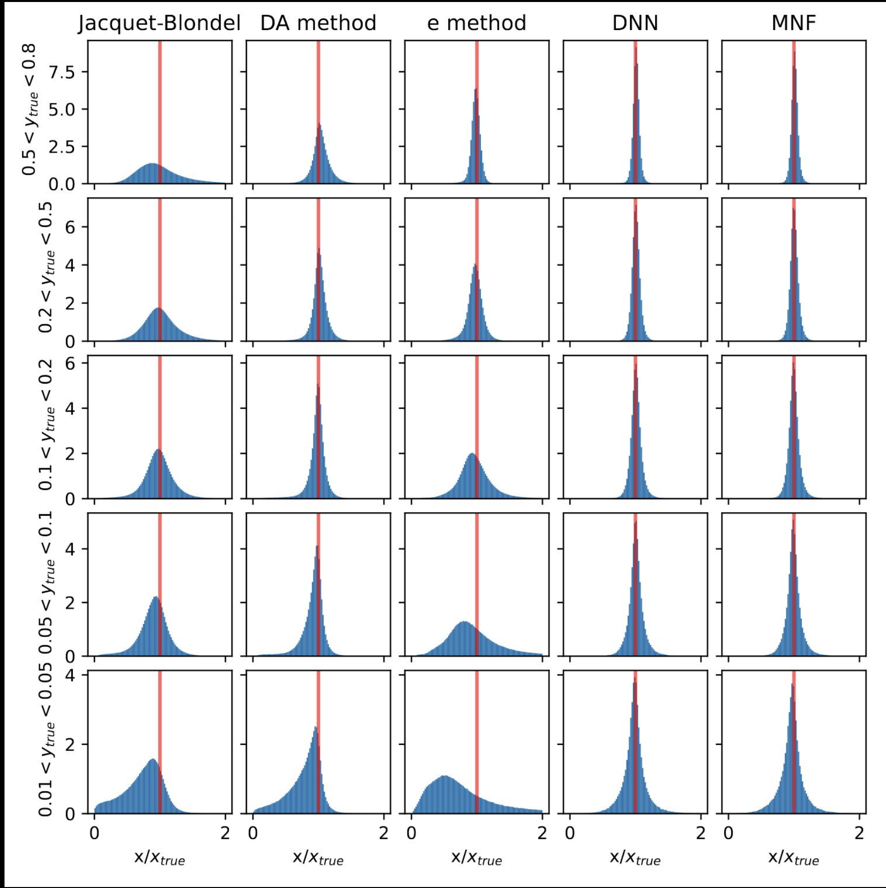
$$\mathcal{L}_{Tot.} = \mathcal{L}_{Reg.} + \gamma\mathcal{L}_{Phys.} + \beta\mathcal{L}_{NF.}$$

$$\mathcal{L}_{Phys.} \overset{\text{DIS}}{=} \sum_i \log \hat{Q}_i^2 - (\log S_i + \log \hat{x}_i + \log \hat{y}_i)$$

$$\mathcal{L}_{NF.} = -KL(q(\mathbf{W})\|p(\mathbf{W})) = \mathbb{E}_{q(\mathbf{W},\mathbf{z}_T)}[-KL(q(\mathbf{W}|\mathbf{z}_{T_f})\|p(\mathbf{W})) + \log r(\mathbf{z}_{T_f}|\mathbf{W}) - \log q(\mathbf{z}_{T_f})]$$

[1] C Louizos, M Welling International Conference on Machine Learning; arXiv:1703.01961 Multiplicative Normalizing Flows for Variational Bayesian Neural Networks
[2] A. Kendall and Y. Gal. "What uncertainties do we need in Bayesian deep learning for computer vision?." Adv. Neural Inf. Process. 30 (2017).

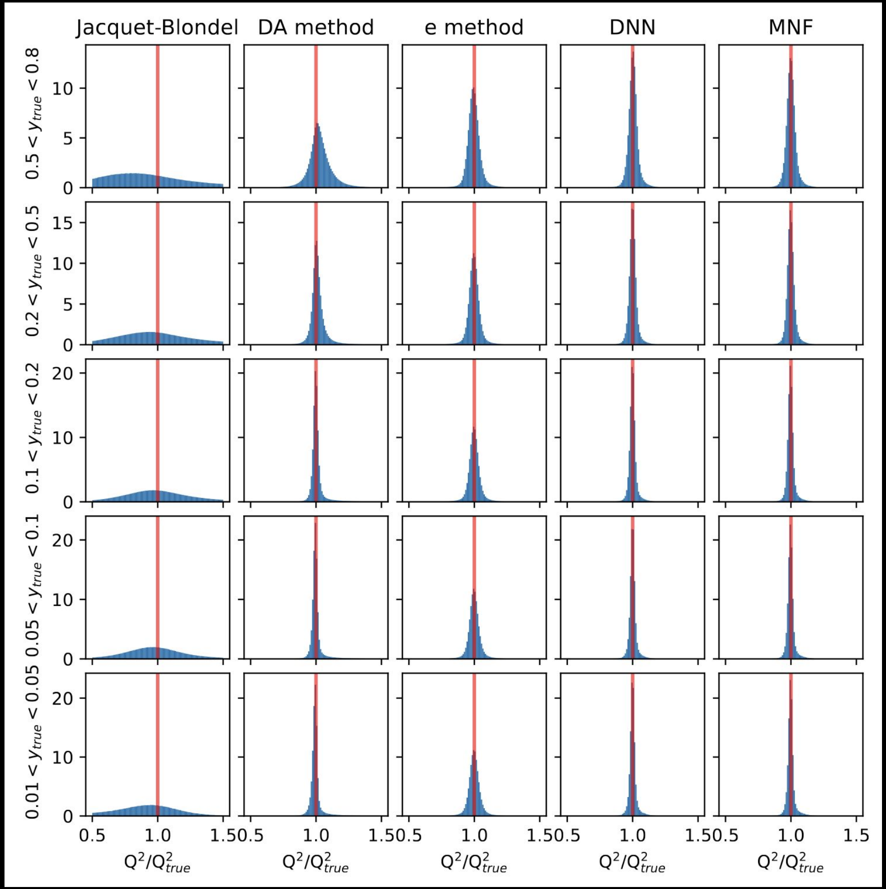
Aleatoric-RMS comparison



Y Bin	DA Method	DNN RMS	Aleatoric
(0.5, 0.8)	0.147955	0.061922	0.057942
(0.2, 0.5)	0.134833	0.075418	0.061706
(0.1, 0.2)	0.145530	0.097903	0.071238
(0.05, 0.1)	0.175290	0.132783	0.082945
(0.01, 0.05)	0.252723	0.184589	0.115453

Table 2: Aleatoric RMS Comparions - X

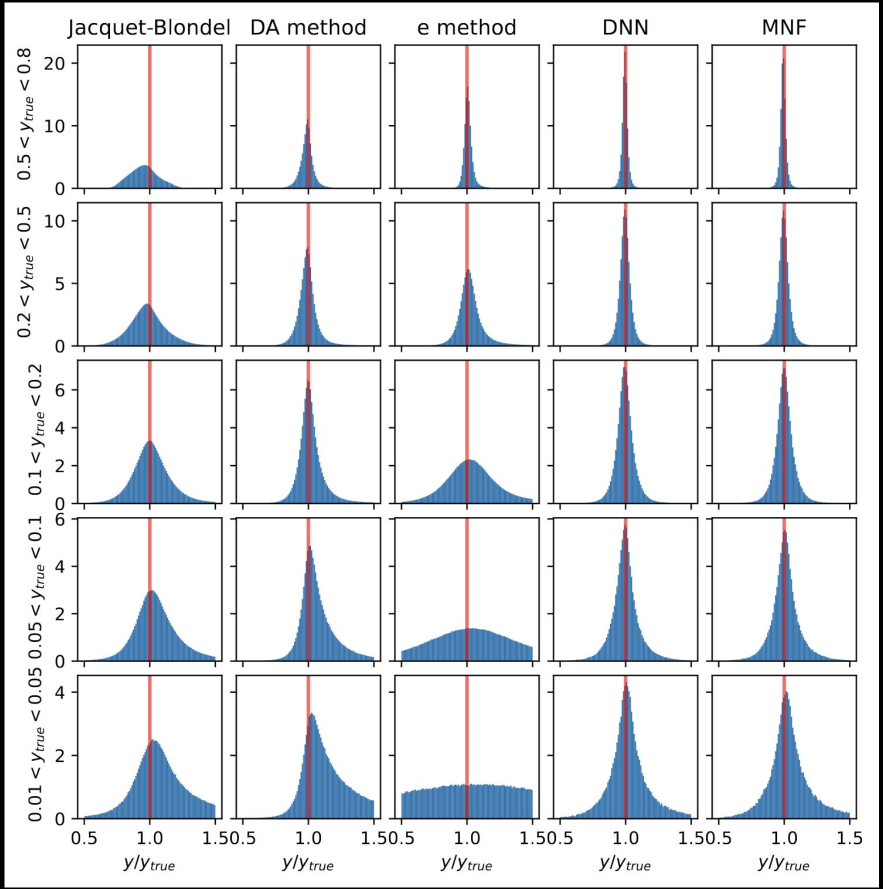
Aleatoric-RMS comparison



Y Bin	e Method	DNN RMS	Aleatoric
(0.5, 0.8)	0.056694	0.044052	0.041349
(0.2, 0.5)	0.055787	0.037505	0.032280
(0.1, 0.2)	0.054219	0.033230	0.029640
(0.05, 0.1)	0.053403	0.032501	0.029411
(0.01, 0.05)	0.053470	0.032139	0.029431

Table 3: Aleatoric RMS Comparison - Q2

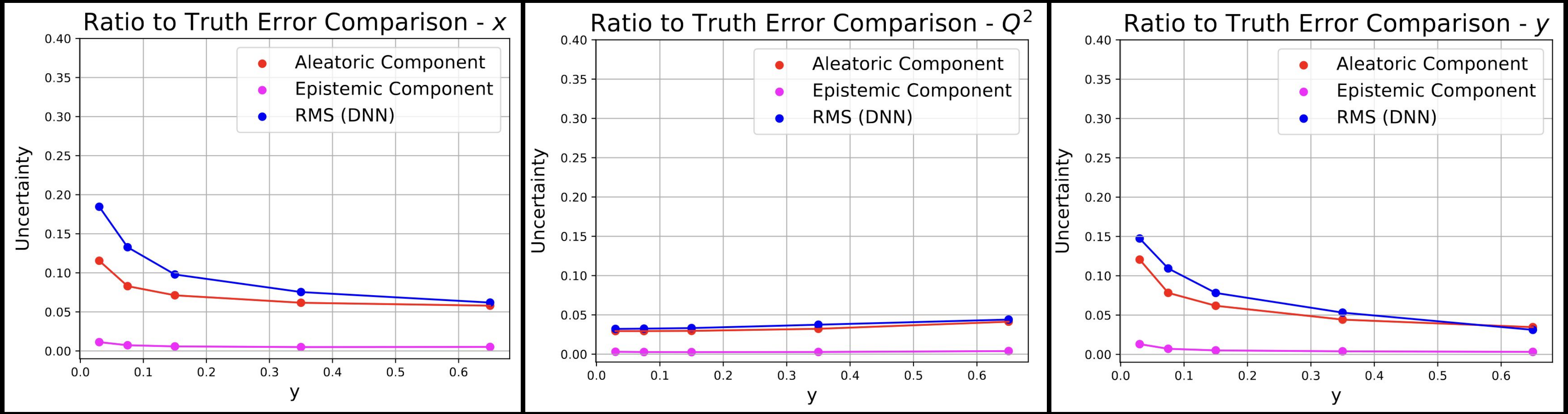
Aleatoric-RMS comparison



Y Bin	DA Method	DNN RMS	Aleatoric
(0.5, 0.8)	0.060537	0.031194	0.034643
(0.2, 0.5)	0.082115	0.053126	0.044249
(0.1, 0.2)	0.098631	0.078143	0.061840
(0.05, 0.1)	0.127276	0.109309	0.078276
(0.01, 0.05)	0.158493	0.147391	0.120546

Table 4: Aleatoric RMS Comparison Y

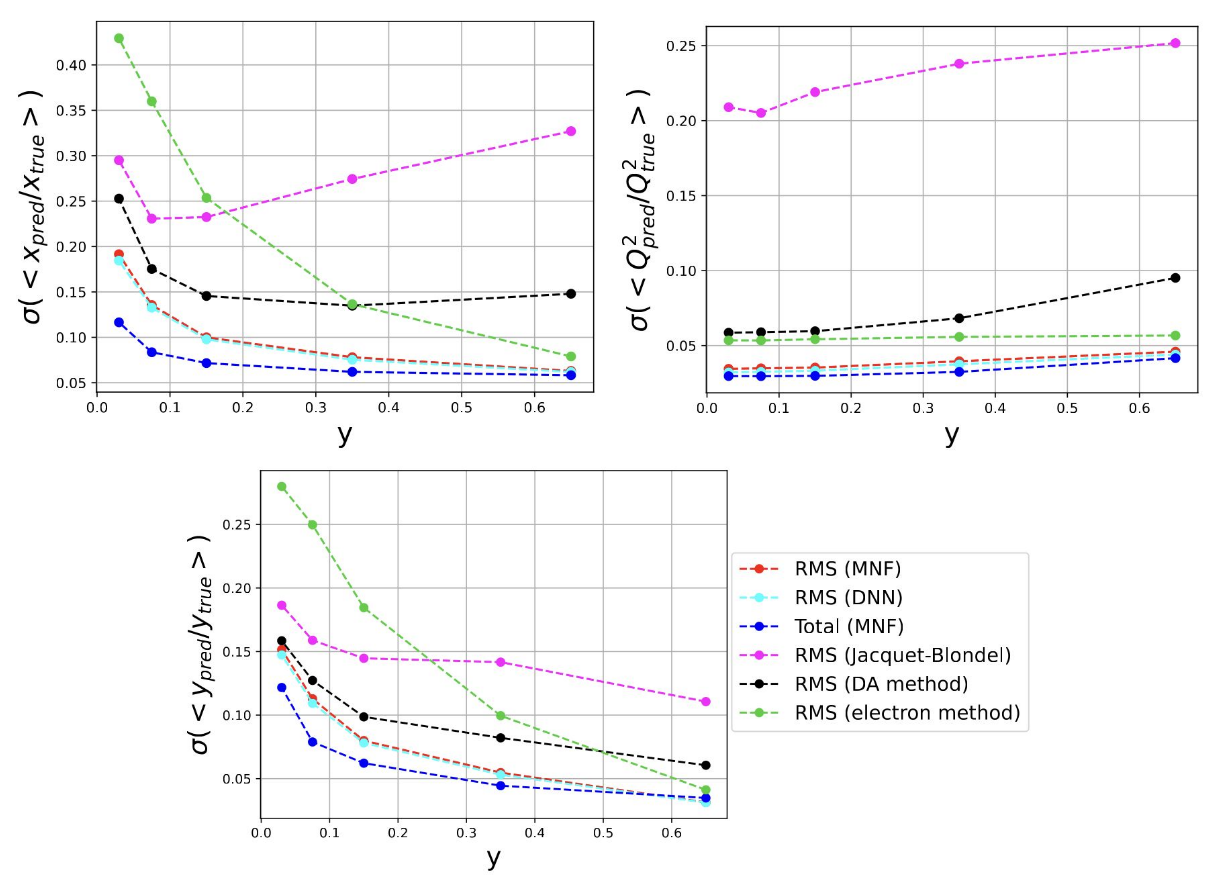
Comparison between DNN and BNN



- The RMS (MNF) roughly coincide with that of DNN as seen previously
- The RMS (DNN) for x and y is larger at low y given the distributions are broader
- The epistemic is systematically smaller than aleatoric component.
- At large y, for x and y the total uncertainty (epistemic+aleatoric) close to RMS of DNN

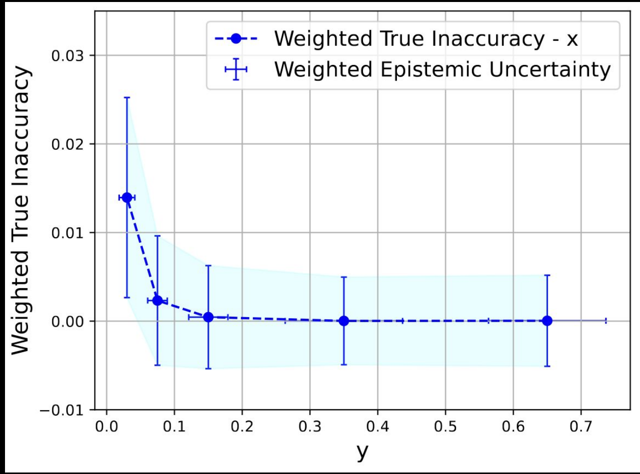
All methods compared

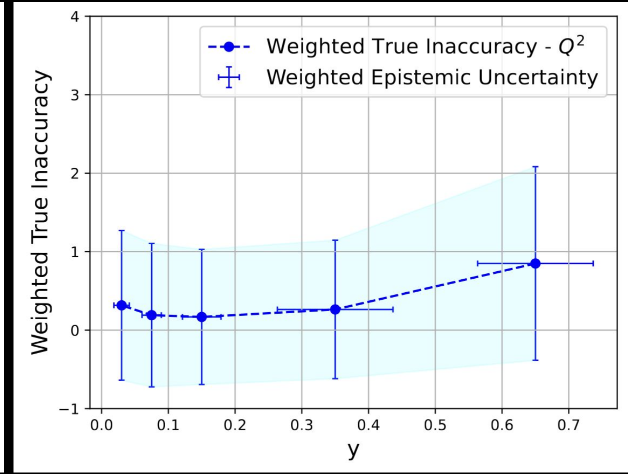
- At low y, the RMS are typically larger due to "broader" distributions
- DNN and MNF have smaller RMS over the whole y range compared to other methods (this was also the finding of NIM-A 1025 (2022): 166164) — "*our method outperforms other methods over a wide kinematics range*"
- "*The RMS resolution for y and x increase at lower y, even for the DNN reconstruction. … This results … may be attributed to further acceptance, noise, or resolution effects that deteriorates the measurement of the HFS*"

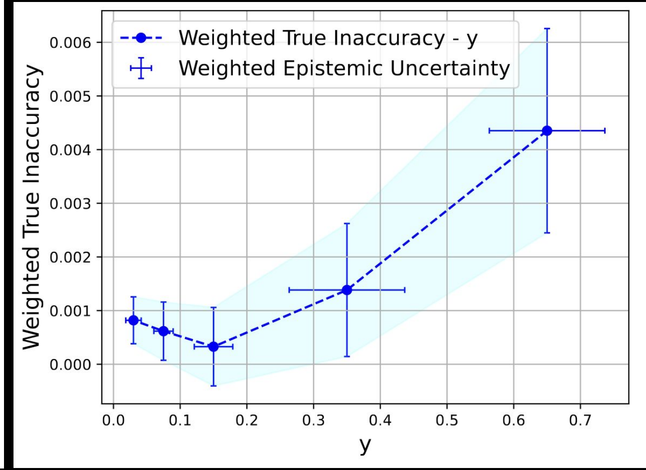


— Reporting uncertainty at the level of the event (e.g., RMS from other methods) —

Epistemic vs True Inaccuracy

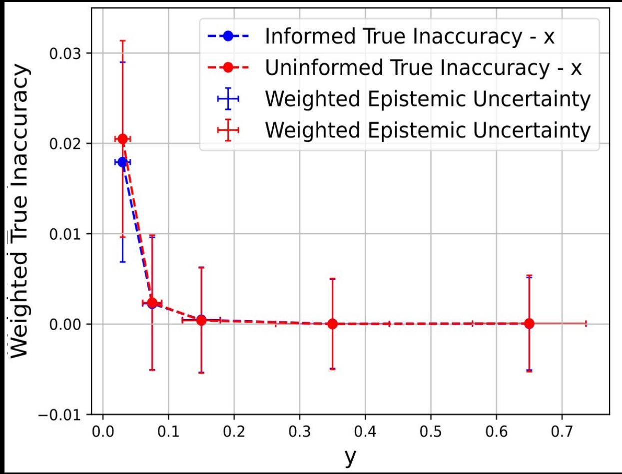


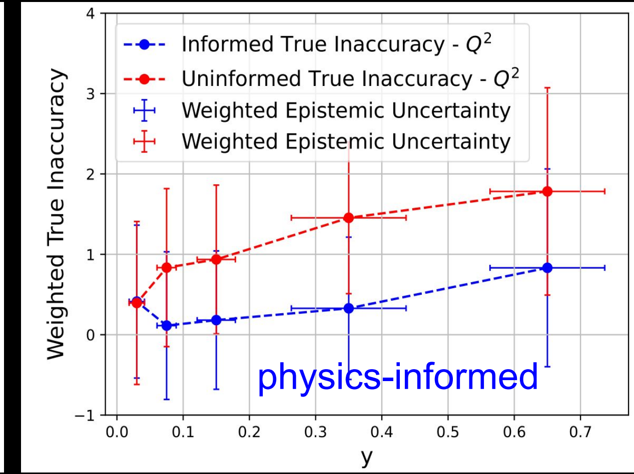


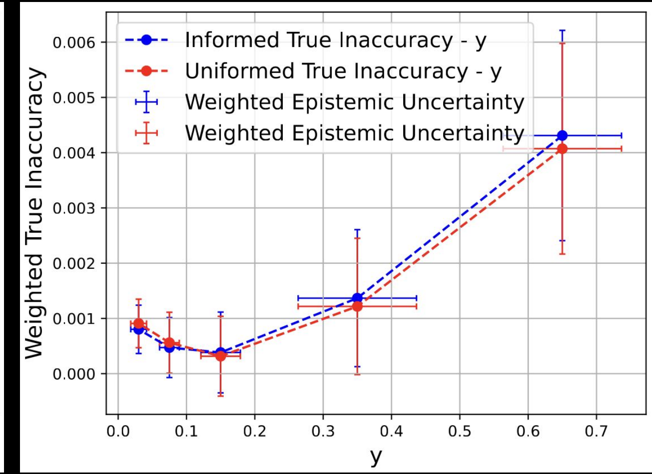


- The plots show that the epistemic uncertainty is larger when the true inaccuracy is larger.

Physics-informed term

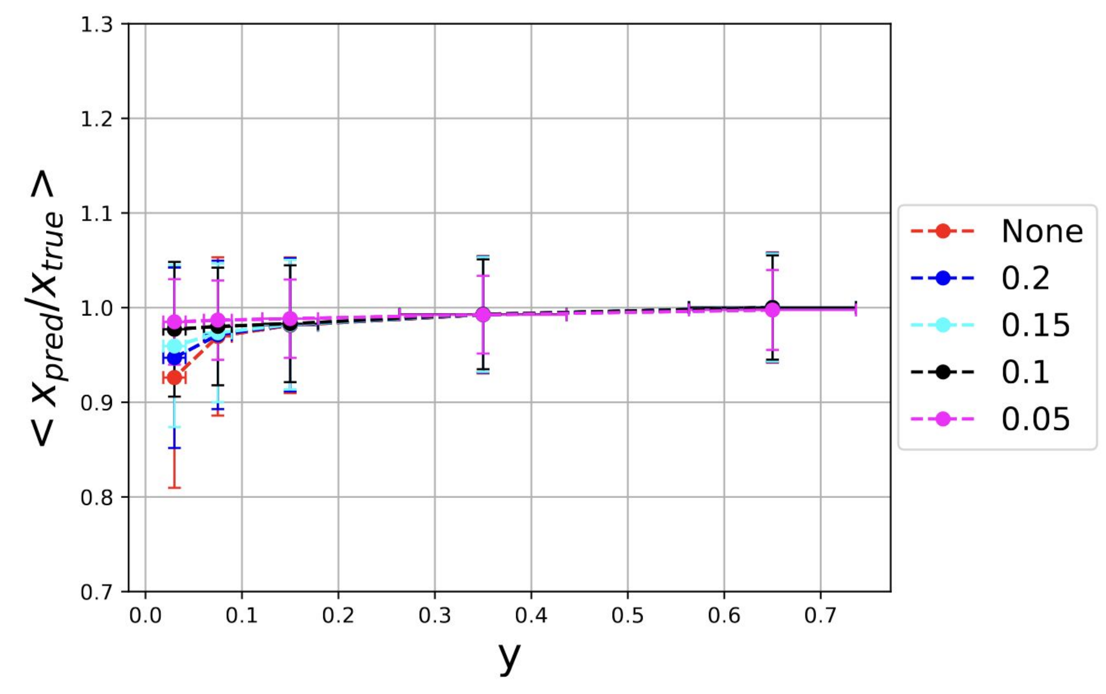






- The plots report the true inaccuracy, and the weighted epistemic uncertainty, which is larger the larger the true inaccuracy is
- The physics-informed term (blue) contributes to decrease the true inaccuracy.

Leveraging event-level information

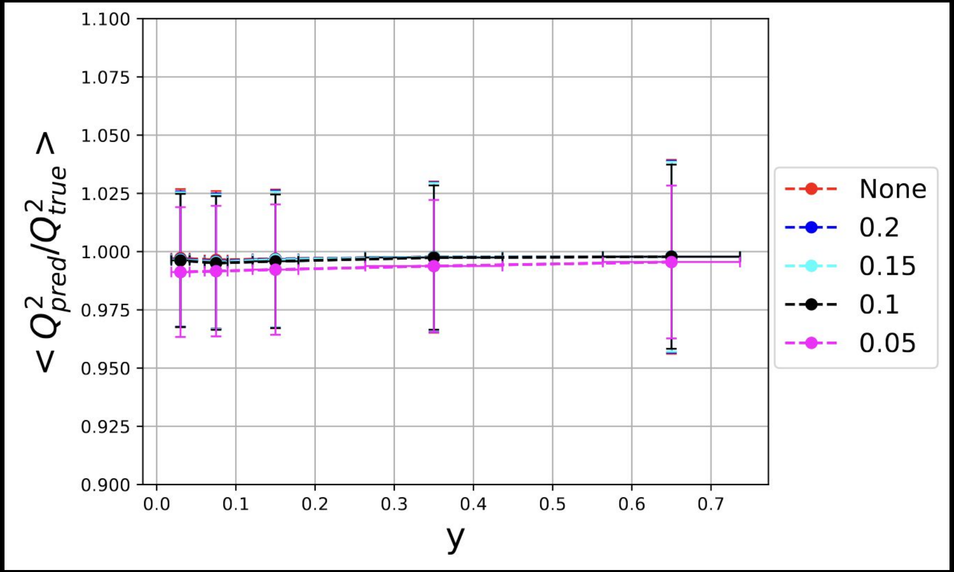


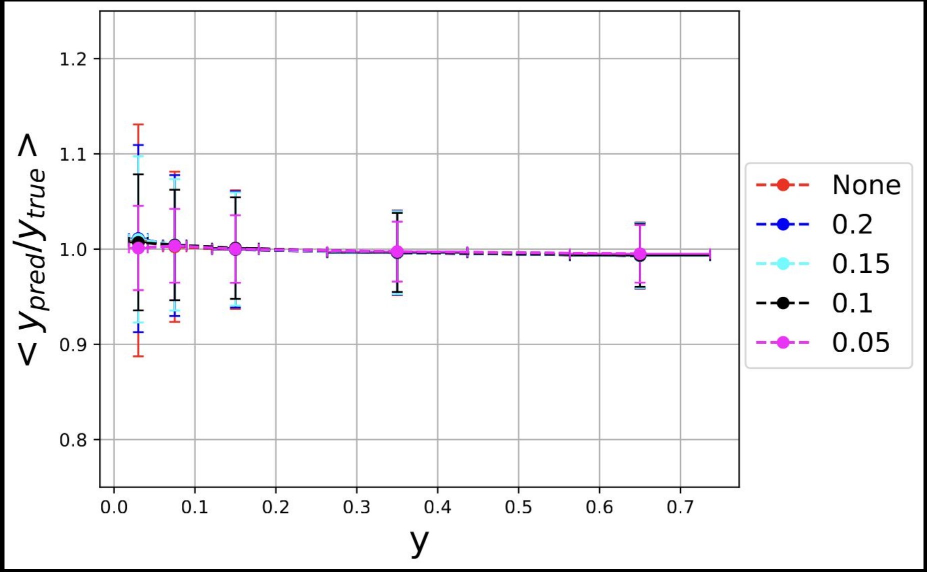
- Removing events with large relative event-level uncertainty (with respect to the network prediction) improve the ratio to truth and reduce inaccuracy
- Notice these cuts do not use any information at the ground truth level

(underway)

— N.b.: events with at least one among x,Q^2, y with relative uncertainty larger than a threshold are removed —

Leveraging event-level information





Negligible impact on Q^2 and y, shown for completeness

(underway)

Conclusions

- **Bayesian Deep Learning and Uncertainty Quantification:** The approach enables detailed uncertainty quantification, encompassing both aleatoric and epistemic uncertainties, at the individual physics event level.
- **Decision-making Advantage:** This level of detailed uncertainty information is instrumental in decision-making processes, such as event filtering, enabling minimization of true inaccuracies without the need for accessing ground truth.
- **Application to DIS Simulation & EIC:** Our findings are corroborated by results from a complete DIS simulation using the H1 detector at HERA, indicating that the same methodology is transferable for applications in EIC, including data quality monitoring and anomaly detection.
- **Network Training and Complexity:** A successful strategy involves initial training of a DNN to achieve satisfactory performance with minimal complexity. The architecture for our MNF and DNN remains consistent in terms of layer sizes.
- **Speed and Efficiency:** The proposed methodology showcases remarkable speed, managing 10,000 samples per event in just 22 milliseconds (RTX 3090).

A comprehensive paper detailing these findings
(validation studies are underway) is currently under development.

AI4EIC Workshop (CUA, Washington,D.C., Nov 28-Dec1, 2023)

Nov 28 - Nov 30

AI/ML for ePIC and Beyond

Calibration, Monitoring, and Experimental Control in Streaming

AI/ML for Accelerators

AI/ML for Data Analysis and Theory

Foundation Models and Trends in Data Science

AI/ML in Production, Distributed ML

Can help spotlight the valuable discussions that occur at this event

Dec 1: Hackathon

Abstract submission open until October 6!

Proceedings will be published in the Journal of Instrumentation.

https://indico.bnl.gov/event/AI4EIC2023
https://eic.ai