Faculty of Physics
Lomonosov Moscow
State University

НИИЯФ
МГУ

clas
CEBAF Large Acceptance Spectrometer

AI-driven evaluation of $\pi^+ n$ electroproduction cross sections and structure functions from the CLAS data

Andrey Golda

2025

Objectives for AI-driven studies of the $\pi^+ n$ electroproduction data.

- Explore AI capabilities to evaluate two-fold differential $\pi^+ n$ electroproduction cross section and to provide interpolation/extrapolation on a user-defined kinematic grid. Assess accuracy of data prediction.

- Explore AI capabilities to establish the dynamics of $\pi^+ n$ electroproduction reaction from the studies of multi-dimensional correlations in the measured observables, in particular, the ability to reproduce the φ-dependence (see Eq in slide #4) for an algorithm with no prior constrains.

- Explore AI capabilities to reproduce two-fold differential $\pi^+ n$ electroproduction cross section in the areas of small or zero acceptance.

Dataset from CLAS experiment

The measured cross sections are stored in the CLAS Physics Database created in collaboration between Hall B at Jefferson Lab and SINP MSU
https://clas.sinp.msu.ru/cgi-bin/jlab/db.cgi

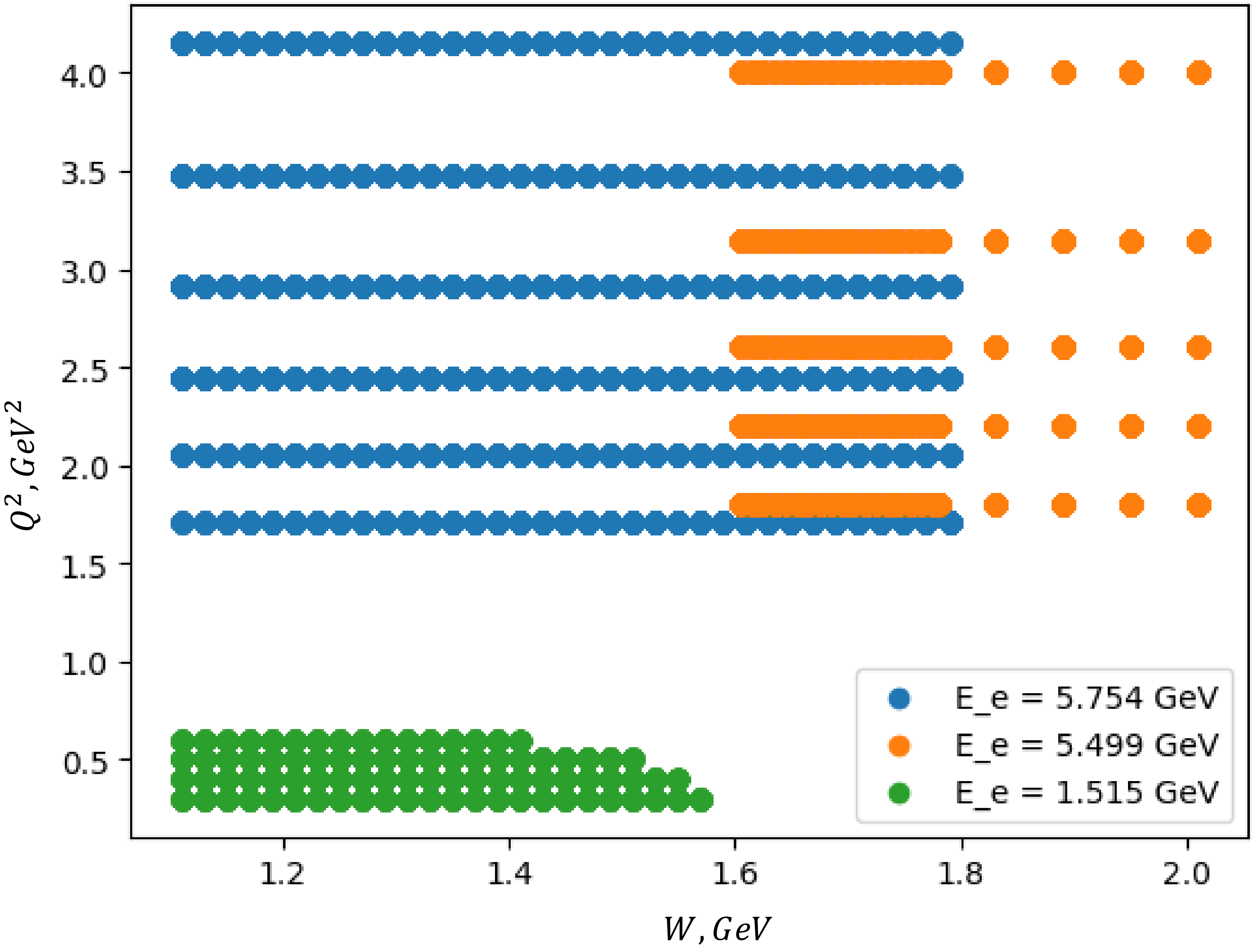
	Ebeam	W	Q2	cos_theta	phi	dsigma_dOmega	error	id
0	1.515	1.11	0.3	0.991445	0.261799	15.3700	5.264366	E8M1
1	1.515	1.11	0.3	0.991445	0.785398	4.5110	1.743136	E8M1
2	1.515	1.11	0.3	0.991445	1.308997	4.4780	1.611260	E8M1
3	1.515	1.11	0.3	0.991445	1.832596	5.1360	1.523529	E8M1
4	1.515	1.11	0.3	0.991445	2.356194	5.0780	1.219442	E8M1
...	...	...	...	...	...	...	...	...
98022	5.499	2.01	4.0	0.975000	3.730641	0.1012	0.043165	E141M160
98023	5.499	2.01	4.0	0.975000	3.992441	0.1199	0.076638	E141M160
98024	5.499	2.01	4.0	0.975000	4.646939	0.1578	0.095391	E141M160
98025	5.499	2.01	4.0	0.975000	4.777839	0.2346	0.158557	E141M160
98026	5.499	2.01	4.0	0.975000	6.086836	0.1250	0.077753	E141M160

93435 rows × 8 columns

Beam energies for the data included in AI/ML studies:

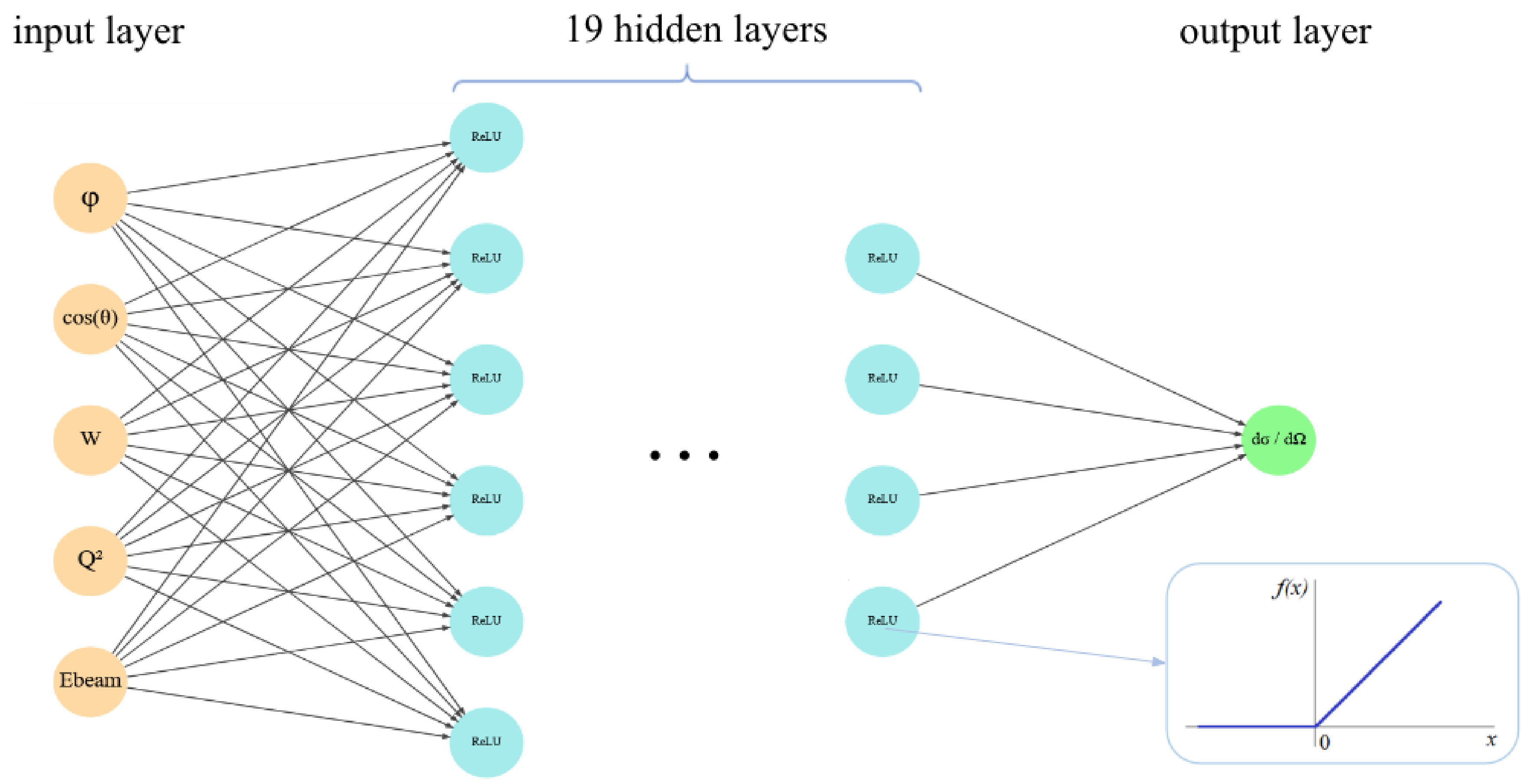
$E, GeV \quad \in \{5.754, 5.499, 1.515\}$

Kinematic coverage



Network architecture

Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. *Nature* 323, 533–536 (1986). https://doi.org/10.1038/323533a0



Validation – baseline: Quality of the training

$$MAE = \frac{1}{N}\sum_{i}^{i=N}\left|y_i^{experimental} - \hat{y}_i^{predicted}\right|$$

Average value and error of cross sections (experimental):

$$\mathrm{AVG}(\mathrm{d}\sigma/\mathrm{d}\Omega) \; = \; 1.158 \; \pm \; 0.2 \, mcb/sr$$

Mean absolute error of AI algorithm of cross-section prediction on a test samples:

$$MAE(\mathrm{d}\sigma/\mathrm{d}\Omega) \; \cong \; 0.08 \, mcb/sr$$

MAE error is smaller than average error of differential cross-section.
Good quality of training!

Validation

$$E = 5.754\ GeV; Q^2 = 2.915\ GeV^2$$

1st resonance region

$E = 5.754\ GeV; Q^2 = 2.915\ GeV^2$ - 1st resonance maximum

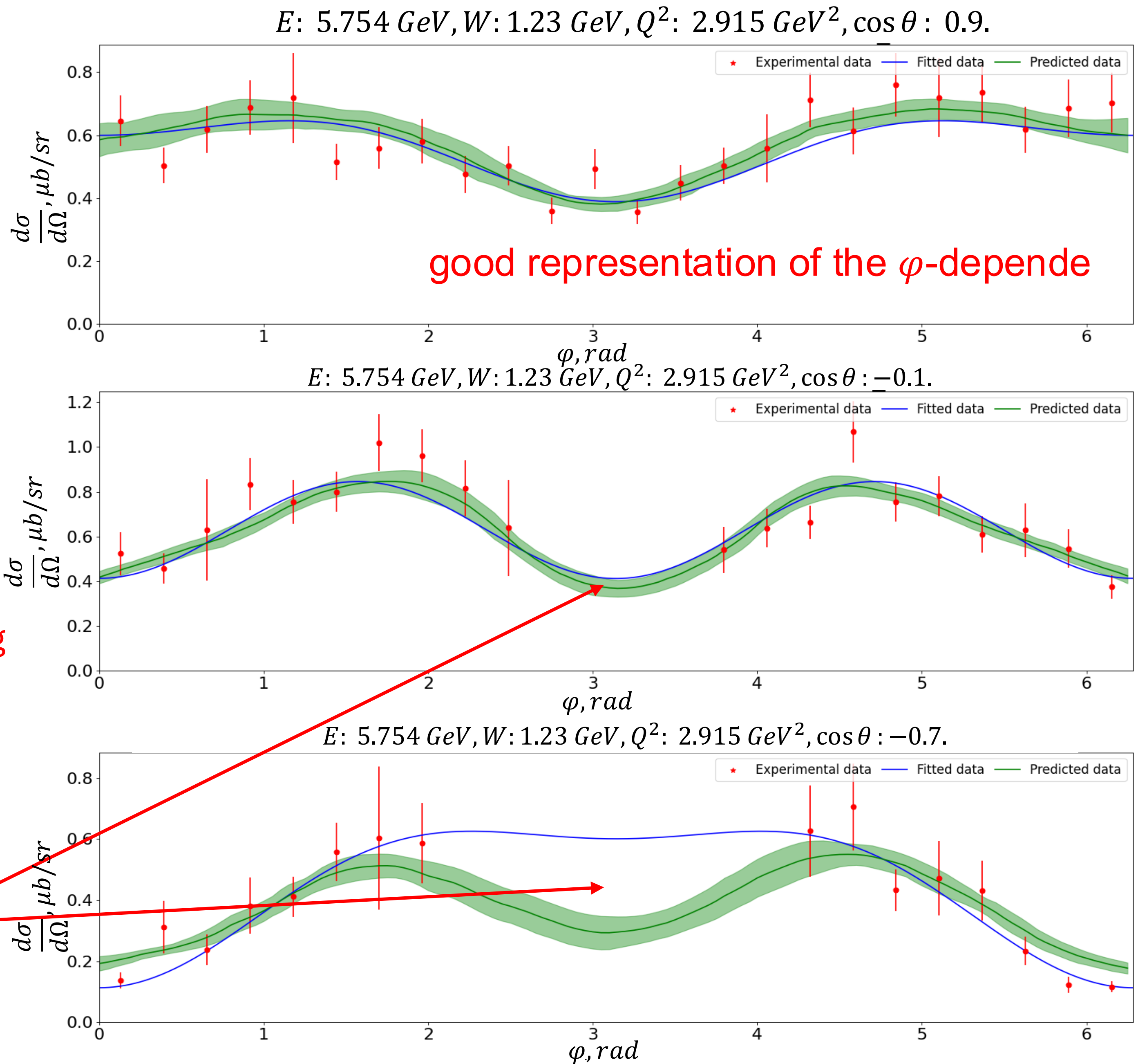
Red dots with error bars – experimental data

Blue line – fit of the experimental data by Eq. in slide #4

Green line - Neural network predictions with uncertainties shown by green areas obtained from bootstrap

AI algorithm is capable of determining φ-depende of π^+n differential cross sections from the multi-dimensional data analysis

AI algorithm provides predictions for π^+n differential cross sections within φ-ranges where the data are not available



$E = 5.754\ GeV; Q^2 = 2.915\ GeV^2$ - structure functions

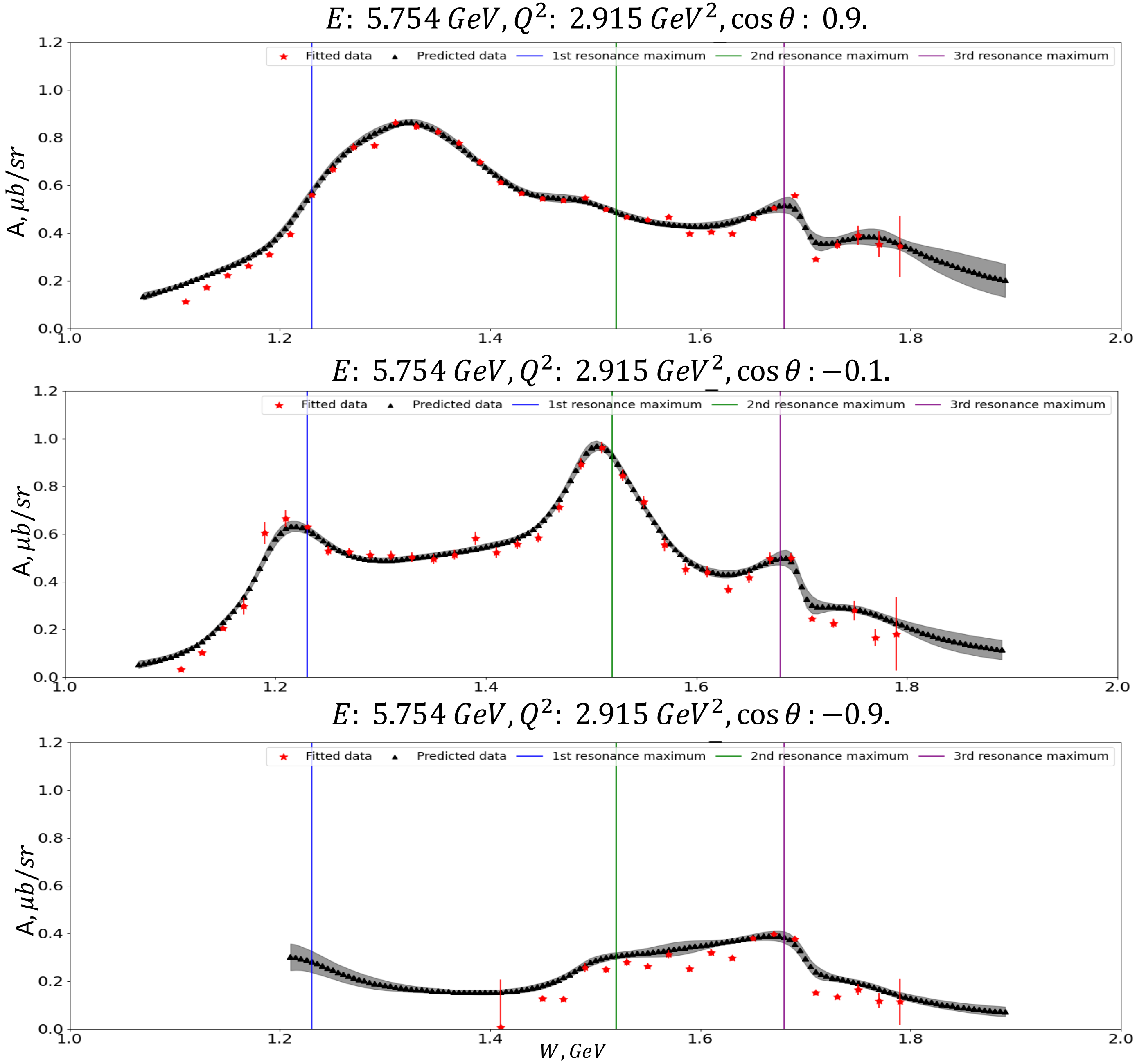
Red dots with error bars – structure functions from the experimental data fit.

Black line – predicted structure function obtained from the moments of the predicted within AI/ML cross sections with uncertainties from bootstrap

Blue line – the center of the 1st resonance region (W=1.23 GeV)

Green line – the center of the 2nd resonance region (W=1.52 GeV)

Purple line – the center of 3rd resonance region for $\pi^+ n$ (W=1.68 GeV)



Replica method

$\boldsymbol{i \in [0, 99], j \in [0, 5], k \in [0, 98207]}$

1. Fix experimental dataset → $\boldsymbol{DS_{exp}^{j,k}}$, where $\boldsymbol{DS_{exp}^{j,k} = DS_{exp}^{j-1,k}} \cup \boldsymbol{\frac{d\sigma}{d\Omega}_{exp}^{k}}$
2. Train a model $\boldsymbol{M_{repl.gen.}}$ $(on\ \boldsymbol{DS_{exp}^{j,k}})$ and make predictions → $\boldsymbol{\frac{d\sigma}{d\Omega}_{repl.gen.}^{k}}$
3. Generate **100** pseudodata vectors of cross-sections $\boldsymbol{\frac{d\sigma}{d\Omega}_{repl}^{i,k}}$ via Gaussian blurring technique with params:

$$\boldsymbol{\mu^k} = \boldsymbol{\frac{d\sigma}{d\Omega}_{repl.gen.}^{k}}, \boldsymbol{\sigma^k} = \boldsymbol{\frac{d\sigma}{d\Omega}_{error}^{k}}, \text{ where } \boldsymbol{\frac{d\sigma}{d\Omega}_{error}^{k}} \subset \boldsymbol{DS_{exp}^{j,k}}$$

4. Train **100** models $\boldsymbol{M_{repl}^{i}}$ on corresponding $\boldsymbol{DS_{repl}^{i,j,k} = DS_{exp}^{i,j-1,k}} \cup \boldsymbol{\frac{d\sigma}{d\Omega}_{repl}^{i,k}}$ and make **100** cross-sections prediction vectors $\boldsymbol{\frac{d\sigma}{d\Omega}_{repl.preds.}^{i,k}}$
5. Take standard deviation $\boldsymbol{\sigma}$ and average $\boldsymbol{\mu}$ from $\boldsymbol{\frac{d\sigma}{d\Omega}_{repl.preds.}^{i,k}}$ compare with $\boldsymbol{\frac{d\sigma}{d\Omega}_{exp}^{k}}$.

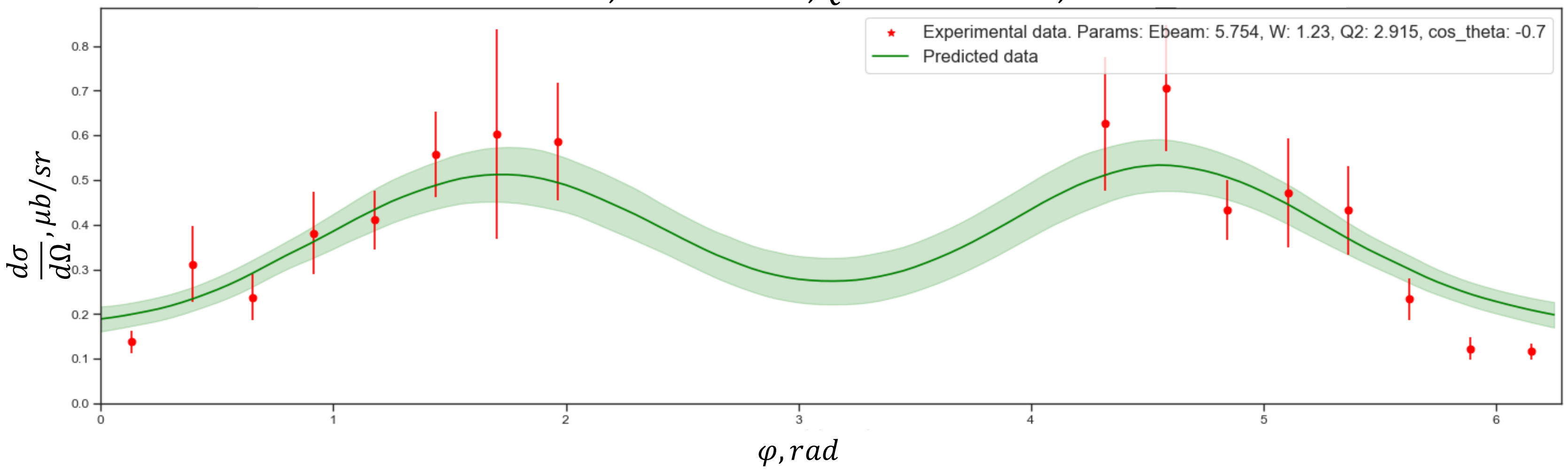
Replicas prediction

experimental data								replicas generator	replicas			replicas predictions		
id	Ebeam	W	Q2	cos_theta	phi	dσ/dΩ	error	dσ/dΩ_prediction	dσ/dΩ_replica_1	...	dσ/dΩ_replica_100	dσ/dΩ_replica_prediction_1	...	dσ/dΩ_replica_prediction_100
0	1.515	1.11	0.3	0.99145	0.268998	15.37	5.26	15.17	15.208	.	15.081	15.133	.	15.16
1	1.515	1.11	0.3	0.99145	0.785623	4.532	1.743	4.331	4.337	.	4.25	4.342	.	4.276
2	1.515	1.11	0.3	0.99145	1.308992	4.478	1.622	4.538	4.596	.	4.456	4.53	.	4.502
3	1.515	1.11	0.3	0.99145	1.861923	5.136	1.53	5.333	5.395	.	5.291	5.316	.	5.375
4	1.515	1.11	0.3	0.99145	2.356194	5.078	1.22	5.571	5.576	.	5.514	5.598	.	5.564
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
98024	5.754	1.53	1.715	0.7	0.3927	1.304	0.042	1.337	1.337	.	1.248	1.288	.	1.286
98025	5.754	1.53	1.715	0.7	0.6544	1.424	0.043	1.464	1.539	.	1.37	1.374	.	1.463
98026	5.754	1.53	1.715	0.7	0.9162	1.605	0.063	1.666	1.721	.	1.652	1.718	.	1.738
98027	5.754	1.53	1.715	0.7	1.178	1.975	0.111	1.907	1.936	.	1.877	1.927	.	1.91

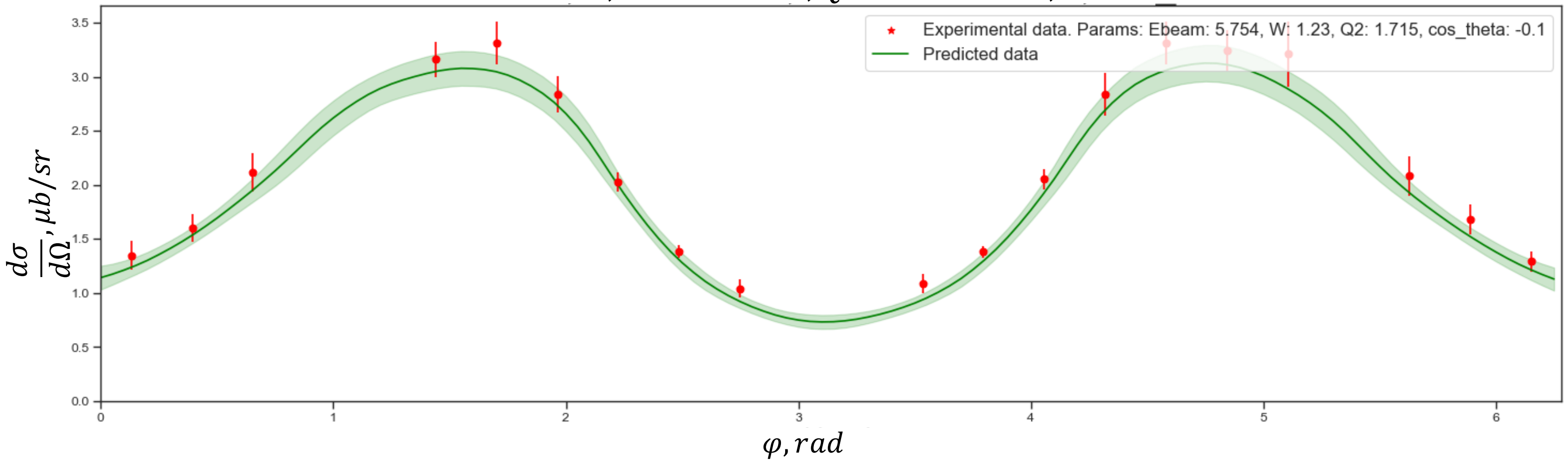
$$\frac{d\sigma}{d\Omega_{diff}}_i = \frac{\frac{d\sigma}{d\Omega_{prediction}} - \frac{d\sigma}{d\Omega_{replica\_prediction}}_i}{\frac{d\sigma}{d\Omega_{prediction}}}$$

Step 3 – replicas prediction validation

E: 5.754 GeV, W: 1.23 GeV, Q^2: 2.915 GeV^2, $\cos\theta$: −0.7.



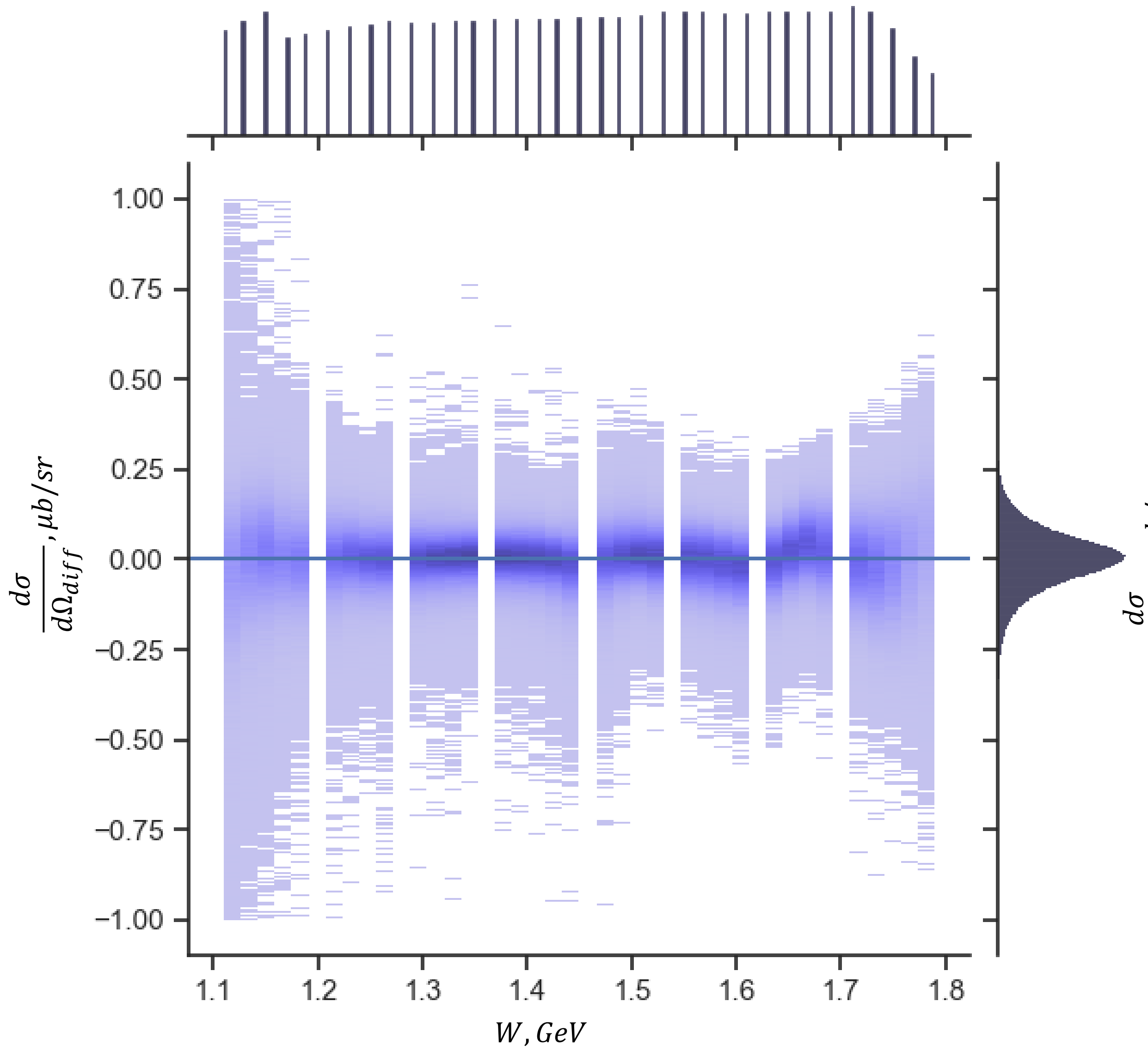
E: 5.754 GeV, W: 1.23 GeV, Q^2: 1.715 GeV^2, $\cos\theta$: −0.1.

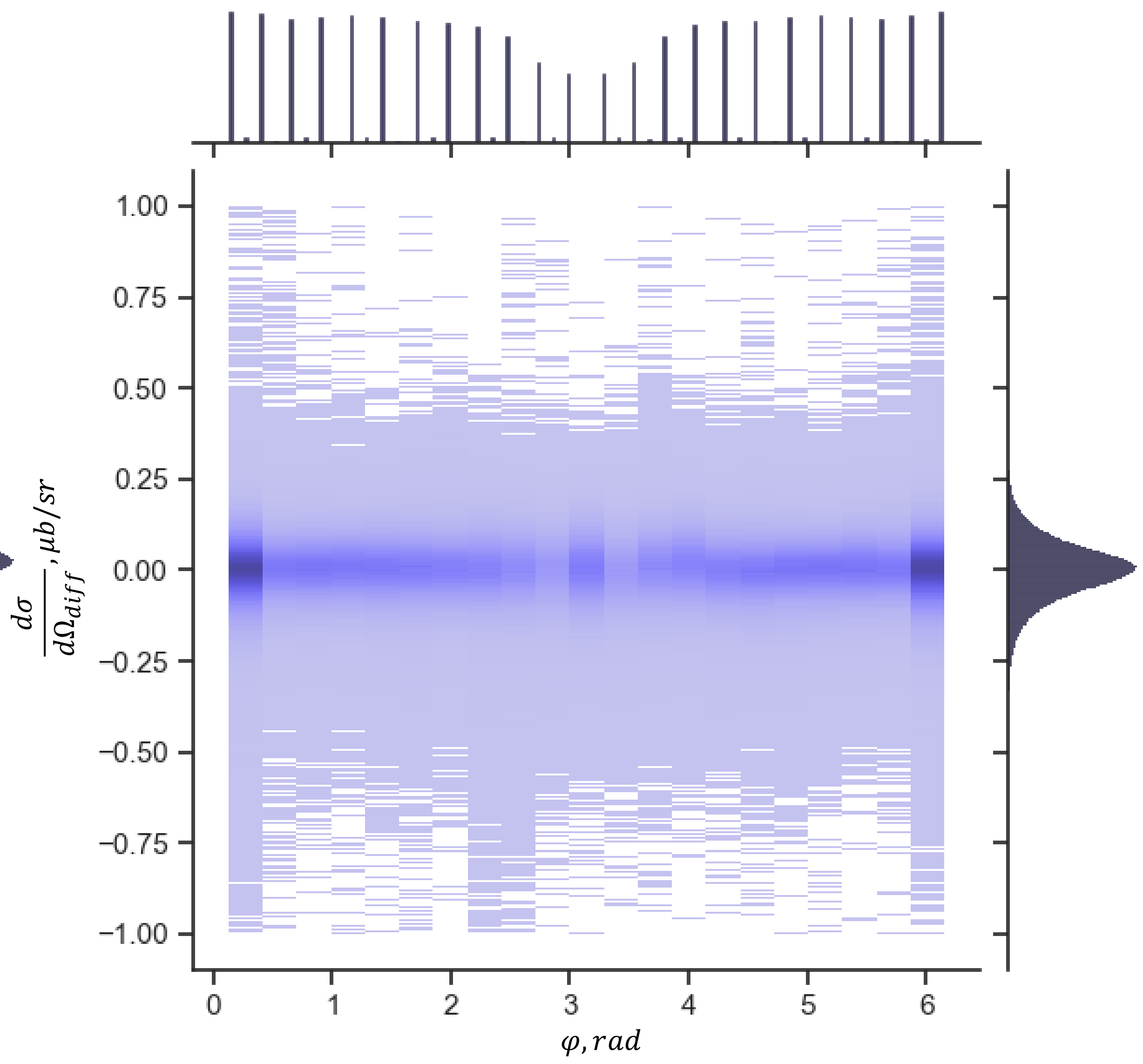


For each point in phase space, 100 predictions were made on 100 replica models and the green area represents 1 standard deviation over these 100 predictions for each point.

Replicas prediction validation

$$\frac{d\sigma}{d\Omega_{diff\,i}} = \frac{\frac{d\sigma}{d\Omega_{prediction}} - \frac{d\sigma}{d\Omega_{replica\_prediction_i}}}{\frac{d\sigma}{d\Omega_{prediction}}}$$





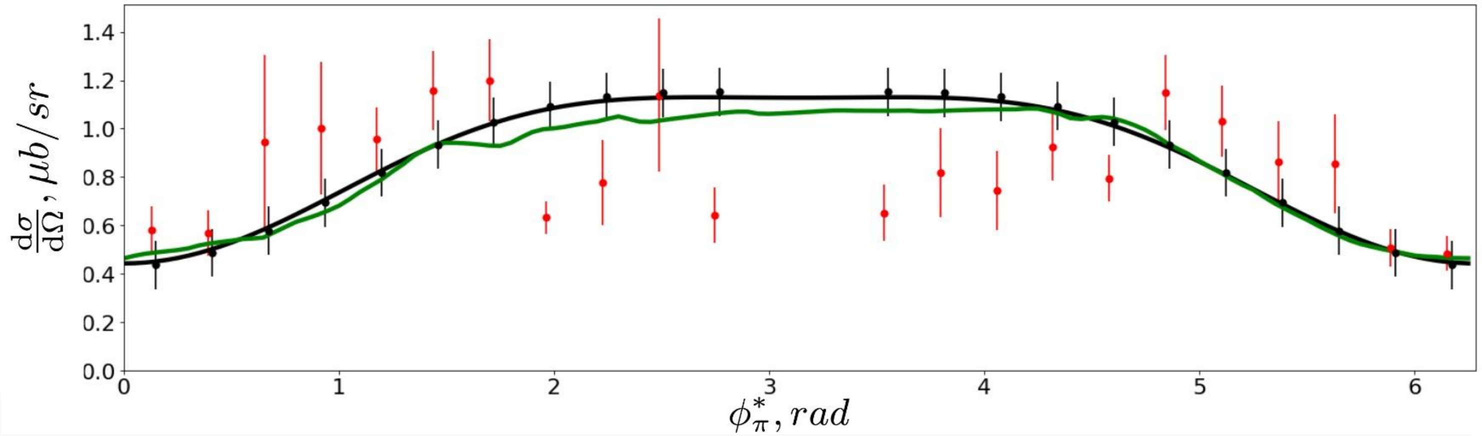
MAID comparisons

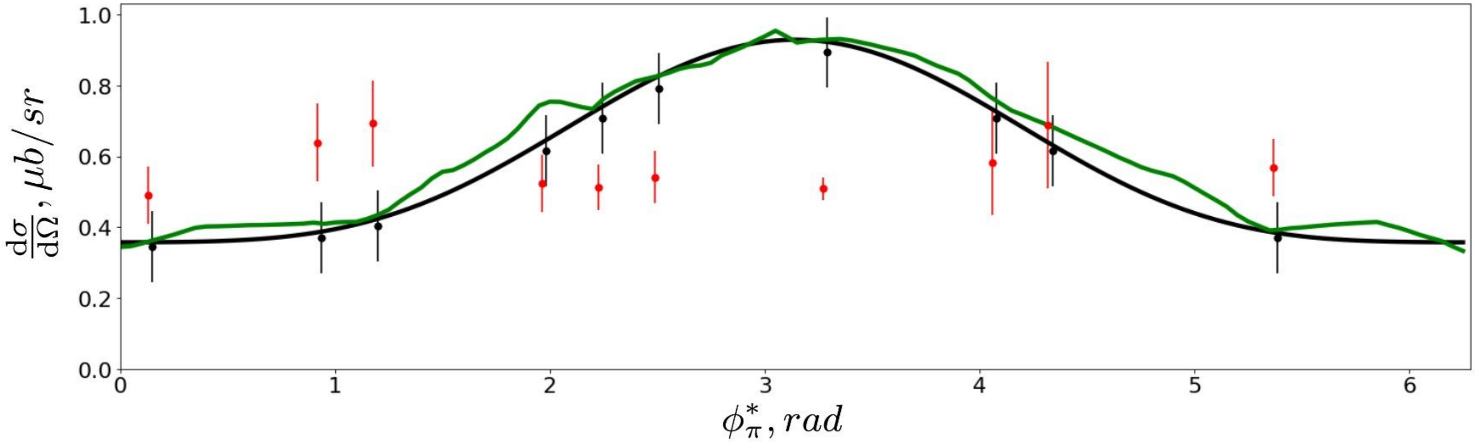
Replicas prediction validation

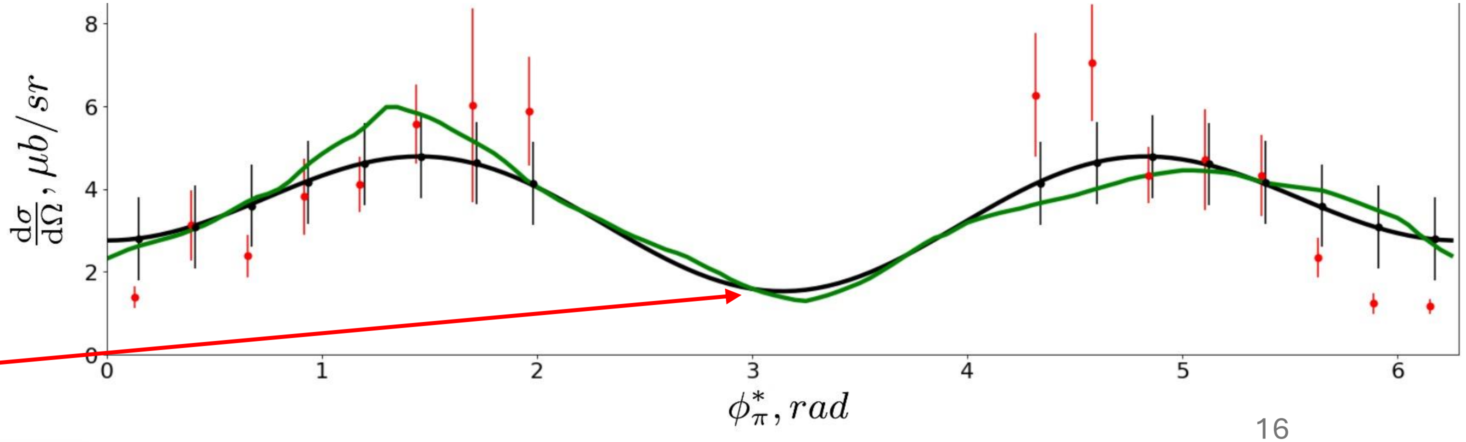
- The model $\pi^+ n$ cross sections were computed from MAID07 within the entire kinematic area (black lines)

- AI was trained on the grid, where CLAS cross sections (red data points) are available, to reproduce the MAID07 quasi-data (black points). The AI estimates (green lines) are compared with MAID07 results within the entire reaction phase space.

- AI algorithm well reproduce the MAID cross sections within areas of zero acceptance







Conclusions

- The AI-driven approach for evaluations of the π^+n differential cross sections and structure functions from the data measured with CLAS has been developed. The approach reproduces the available experimental results with good accuracy and offers predictions for π^+n electroproduction observables within the area of $1.1\ GeV < W < 2\ GeV$ and $0.5\ GeV^2 < Q^2 < 4\ GeV^2$.

- AI algorithms establish the φ-dependence of π^+n cross sections from multi-dimensional data analysis without prior knowledge.

- AI algorithms allow us to evaluate π^+n differential cross section within the reaction phase space where the experimental data are not available (in particular, in the areas of zero acceptance) from accounting for multi-dimensional correlations.

- The structure functions for π^+n electroproduction have been deduced from the CLAS data at $1.1\ GeV < W < 2\ GeV\ \ and\ 0.5\ GeV^2 < Q^2 < 4\ GeV^2$

Thank you
for your attention!