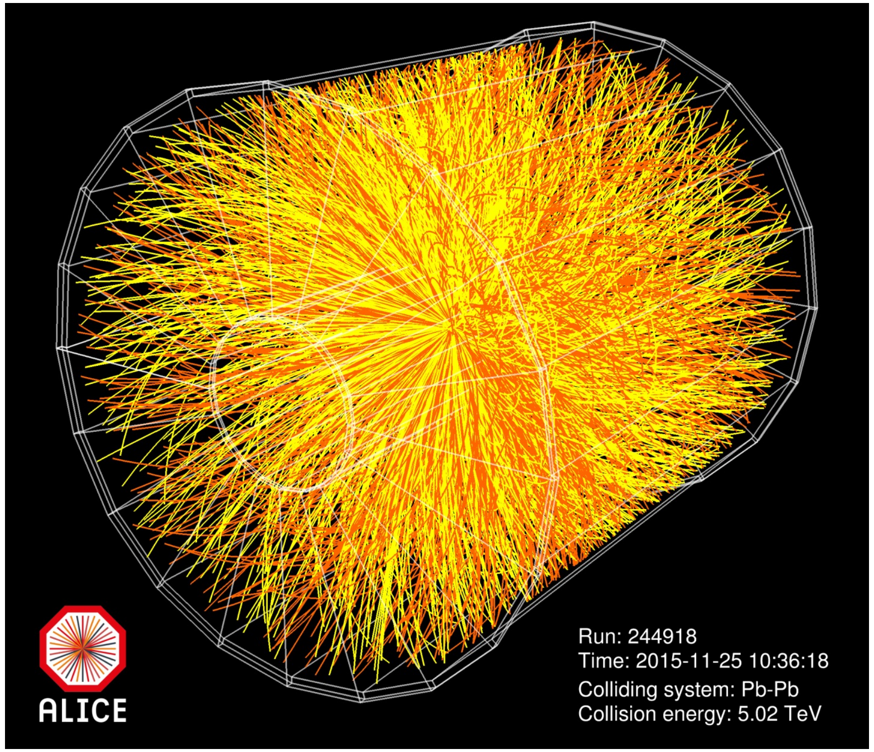
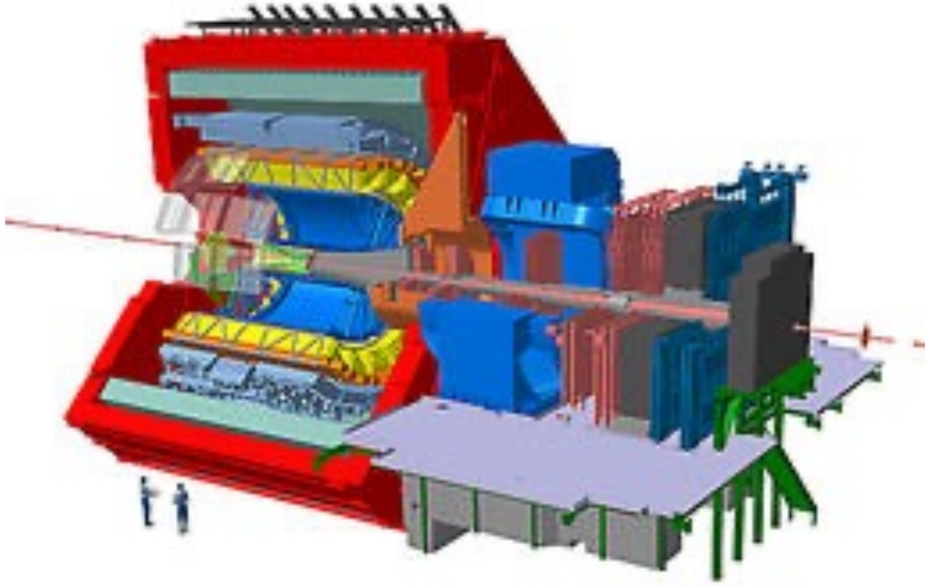
Studies of space-charge distortion corrections in the ALICE TPC using machine learning techniques

Hitoshi Baba (University of Tokyo)
for the ALICE Collaboration

LHC-ALICE Experiment

- Experiment at the CERN LHC which aims to understand the detailed characteristics of the quark-gluon plasma.



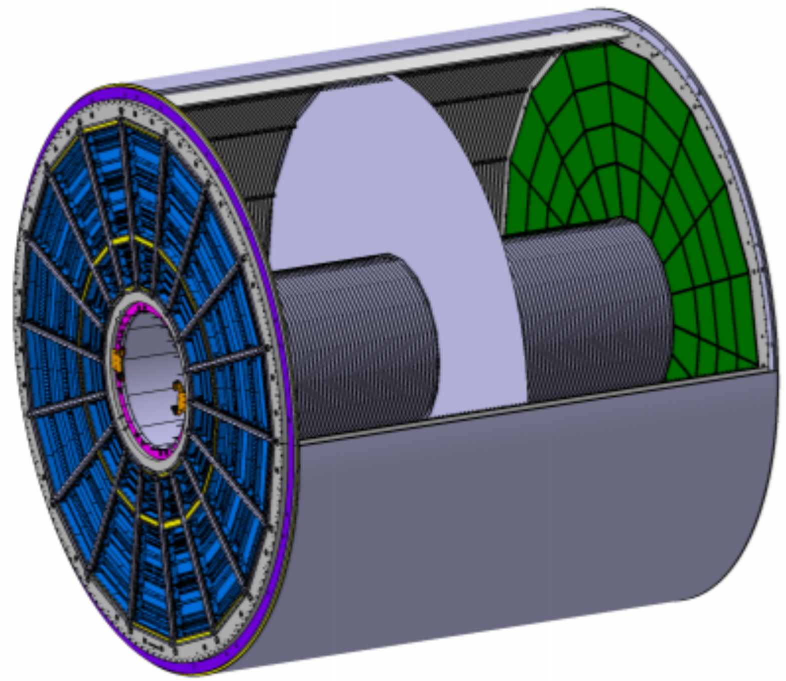


- Pb-Pb collisions with a 50 kHz interaction rate planned for LHC Run 3

(Approx. 5 times larger than the Run 2 interaction rate (10 kHz))

The ALICE-TPC Detector

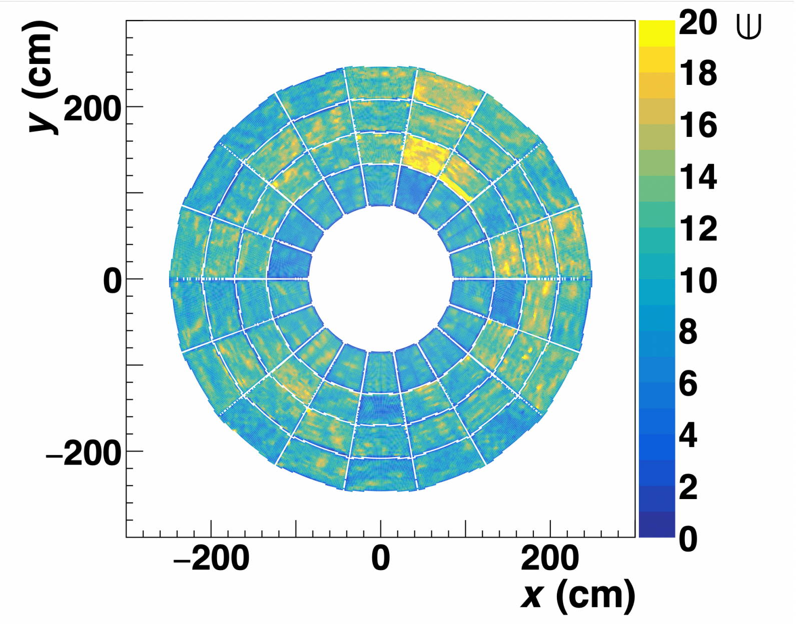
- Main tracking detector of ALICE
- Charge read out through the pads installed to the Readout Chambers (ROCs)
- Multiplication technology at the ROCs changed from multiwire proportional chambers to Gas Electron Multipliers
- Intrinsic spatial resolution: ~200 μm



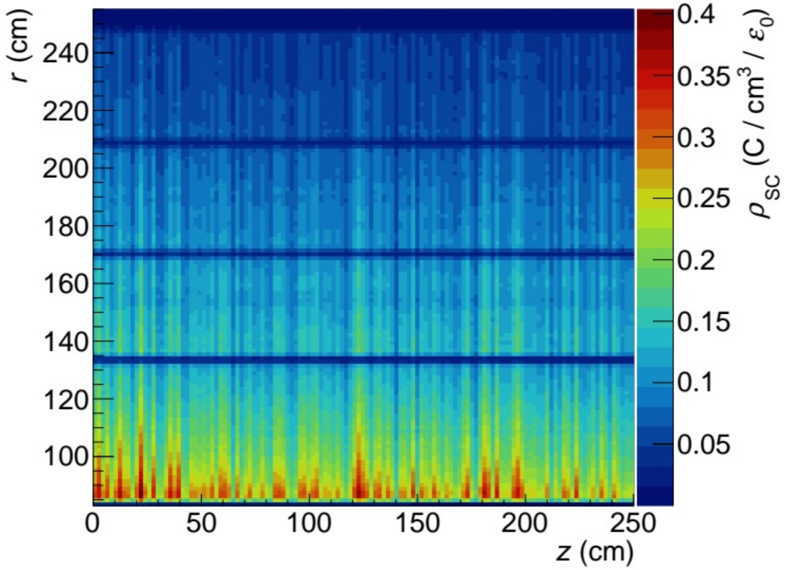
- Gas: $Ne\text{-}CO_2\text{-}N_2$ (90-10-5)
- Drift electric field: 400 V/cm
- Electron properties:
  - Drift velocity: 2.58 cm/μs
  - Max. drift time: 97 μs
- Positive ion (Ne^+) properties:
  - Drift velocity: 1.632 cm/ms
  - Max. drift time: 153 ms

The Space-charge effect

- Combined properties of the four GEM layers:
  - Positive Ion Backflow (IBF):<1%
  - Amplification factor: 2000



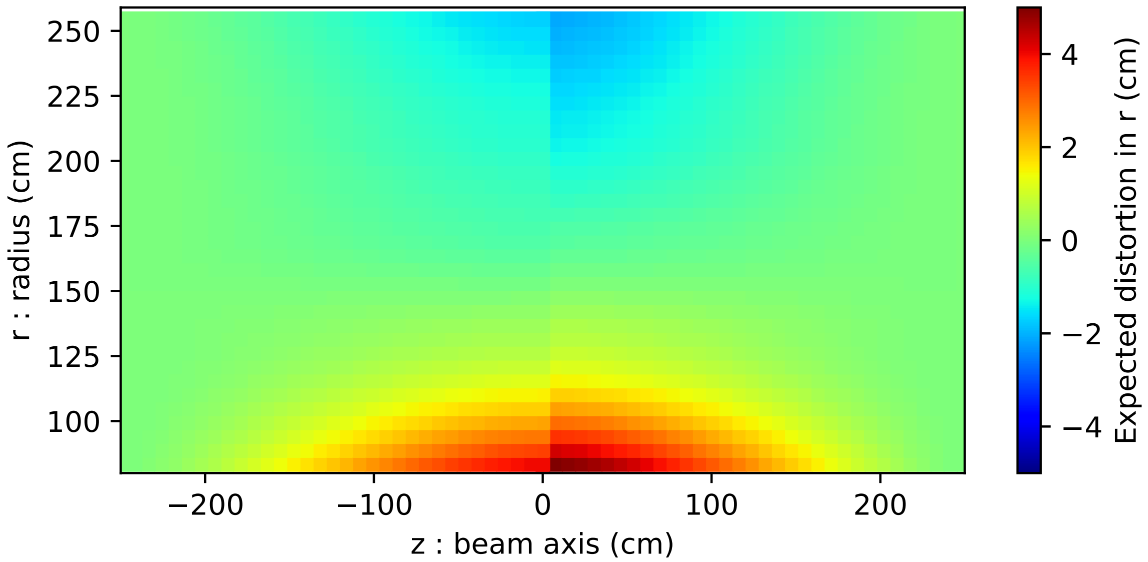
Measured value of ε
(# of backflowing ions per single incoming electron)



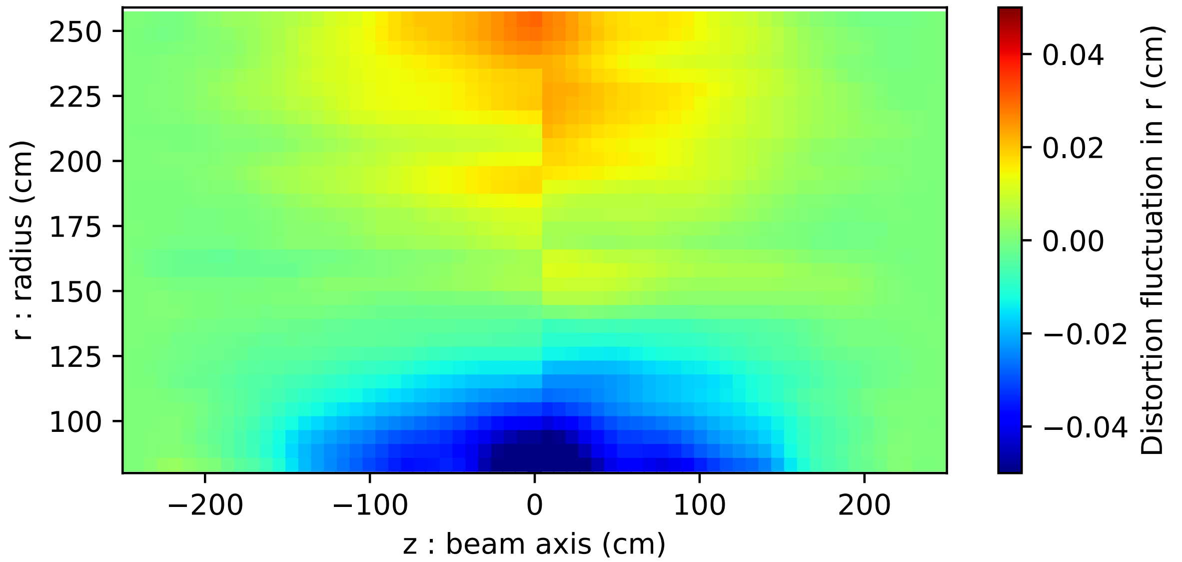
Space-charge density

50 kHz interaction rate, ~160 ms ion drift time
→Space-charge from ~8,000 interactions constantly present inside the TPC drift volume

Simulated Size of Space-point Distortions



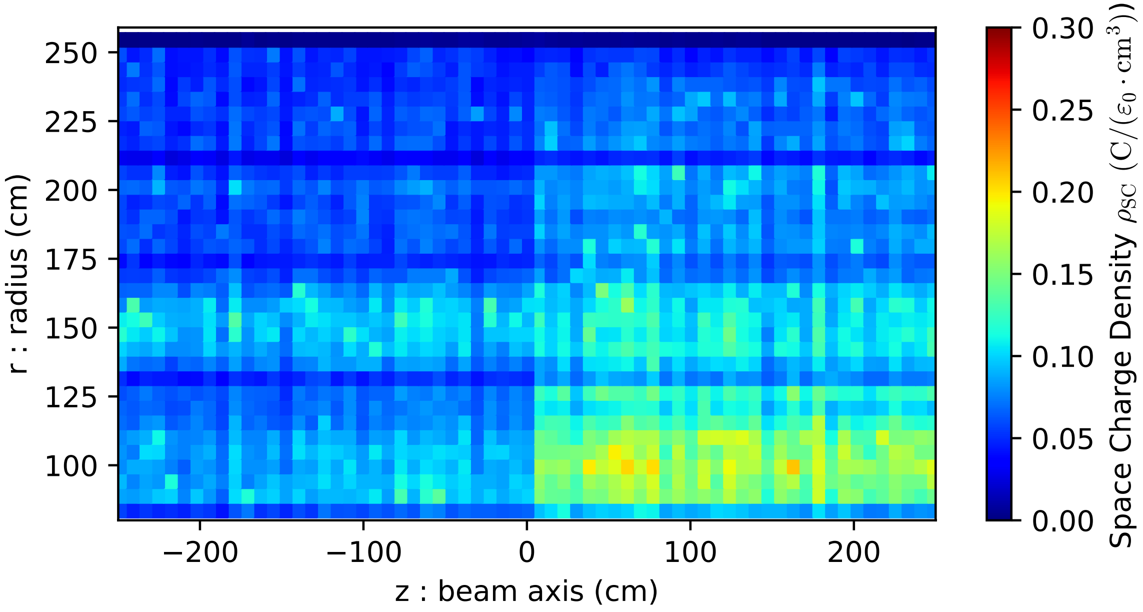
Space-point distortion ~O(cm)



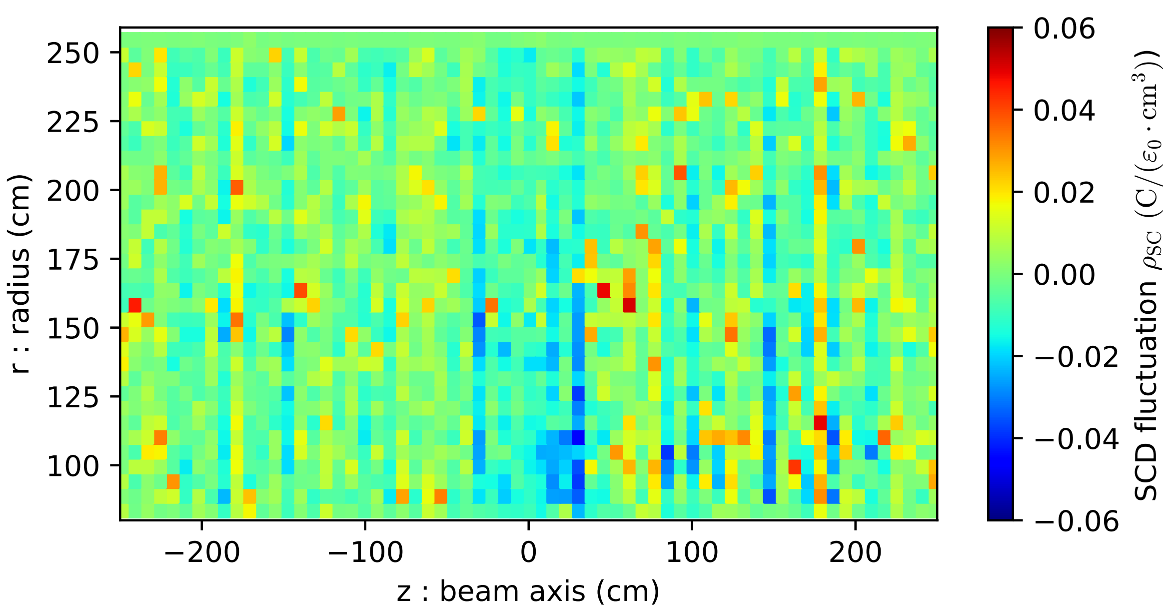
Residual space-point distortion after performing the average correction ~O(500 μm)

Space-point distortion fluctuation correction also necessary to achieve the intrinsic spatial resolution of O(200 μm)

Z dependence of the space-charge density



Space-charge density



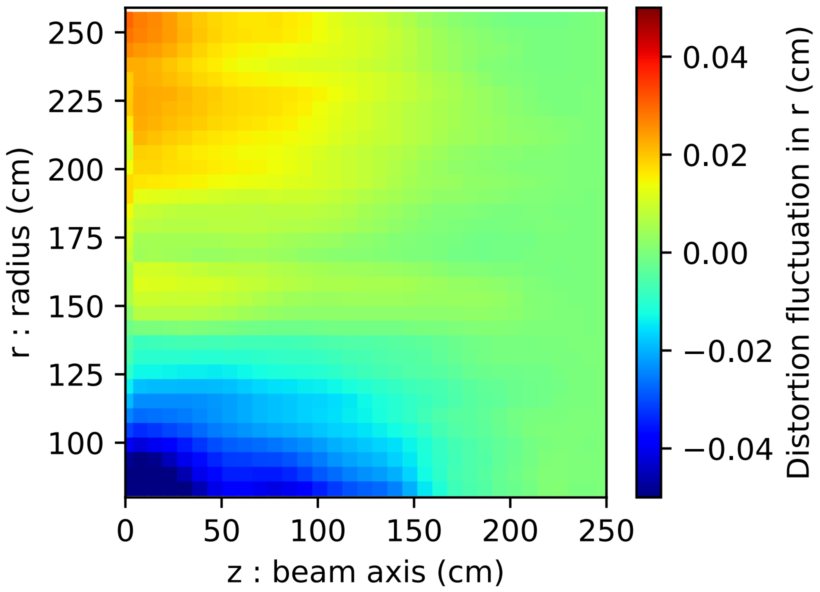
Residual SCD after subtracting the average SCD

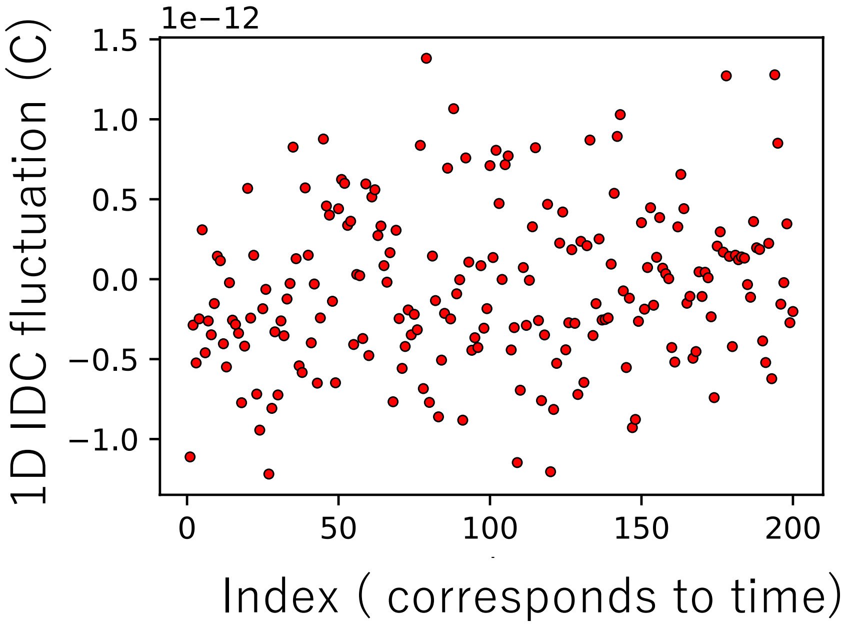
→Time direction information important for space-charge distortion fluctuation correction
Try to perform this correction in a time-efficient way using machine learning to achieve O(200 μm) resolution

Training data

- Usage of the Integrated Digital Current (IDC)
- The digitized signal from the pad is integrated over ~1ms (=3D IDC (r, φ, t))
- 1D IDC $I_1(t)$
  - 3D IDC integrated over $(r,\ \varphi)$
  - Contains information about the time direction fluctuations
  - Fourier decomposed, and the coefficients used for training.
- Derivative $\delta\Delta/\delta I_1(t)$ of the space-point distortion Δ
  - Distortion $\Delta = \langle\Delta\rangle + \delta I_1 \cdot (\delta\Delta/\delta I_1)$
  - Can be estimated using Δ averaged over longer time intervals

Space-point distortion fluctuation at each space point
+
Fourier coefficients of the **1D IDC fluctuation**
+
Derivative of the space-charge distortion $\delta\Delta/\delta I_1(t)$





Machine learning models

Currently considering the Random Forest, XGBoost and the neural network

Models which utilize multiple decision trees

Random Forest

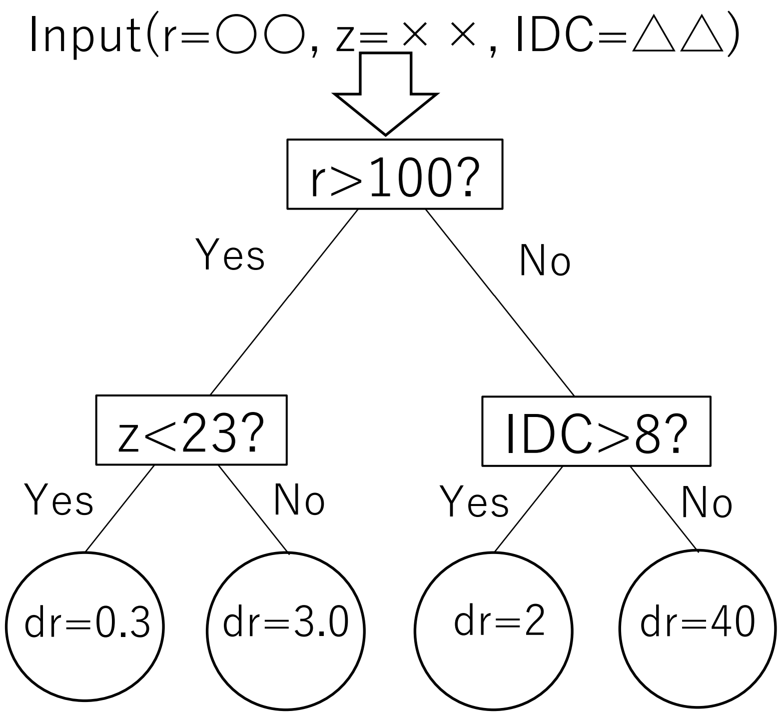
XGBoost

Neural network

Same model trained in a different way

The Random Forest

- Average the results of multiple decision trees
- Each decision tree trained parallelly

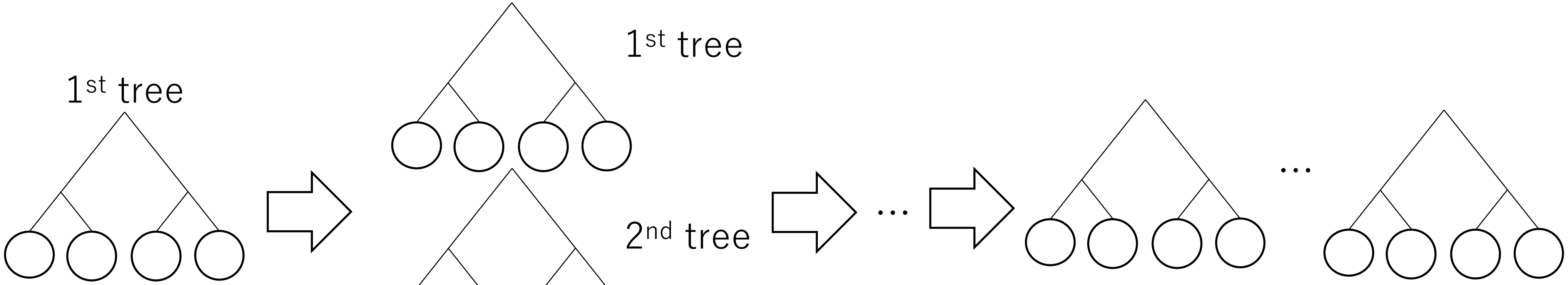


Model parameters:
number of trees (fixed to 100 in this study), maximum depth of the trees, etc.

XGBoost

https://arxiv.org/abs/1603.02754

- Sum the results of multiple decision trees
- Training done in an *additive manner*



$\widehat{y^1} = f_1(\boldsymbol{x})$

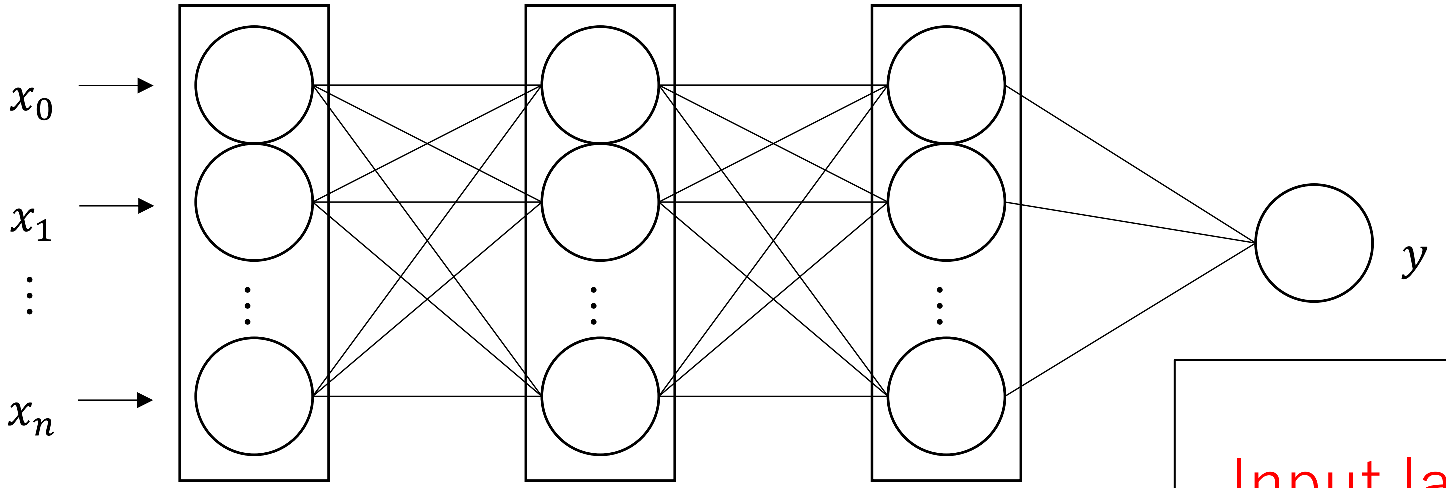
$\widehat{y^2} = \widehat{y^1} + f_2(\boldsymbol{x})$

$\widehat{y^t} = \widehat{y^{(t-1)}} + f_t(\boldsymbol{x})$

Model parameters:
number of trees, maximum depth of the trees, etc.

Neural Networks

- “Neurons” organized in several layers react to the input



Input to the k-th layer: $x_i^{(k)} = \varphi\left(y_i^{(k-1)}\right)$

Output from the k-th layer: $y_i^{(k)} = \sum_l w_l x_l$

($\varphi(x)$: Activation function)

Layers used:
Input layer, output layer and some hidden layers
Activation function for hidden layer: ReLU

Values evaluated in this study

- Used **200,000 maps** for training
- Made the model predict the distortion fluctuations for **400 maps**
- Calculated the **mean** & **std. dev.** of

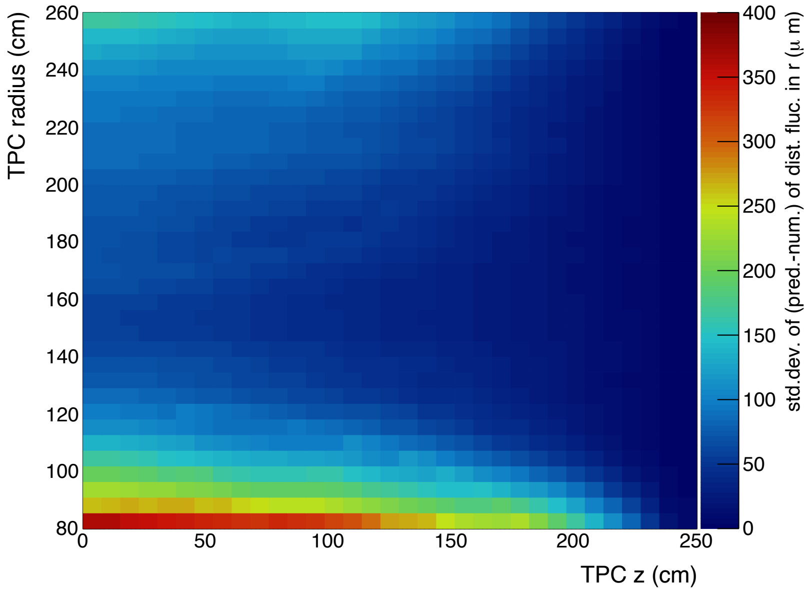
(Predicted distortion fluctuation)-(simulated distortion fluctuation)

- Checked the values at $\varphi=0$

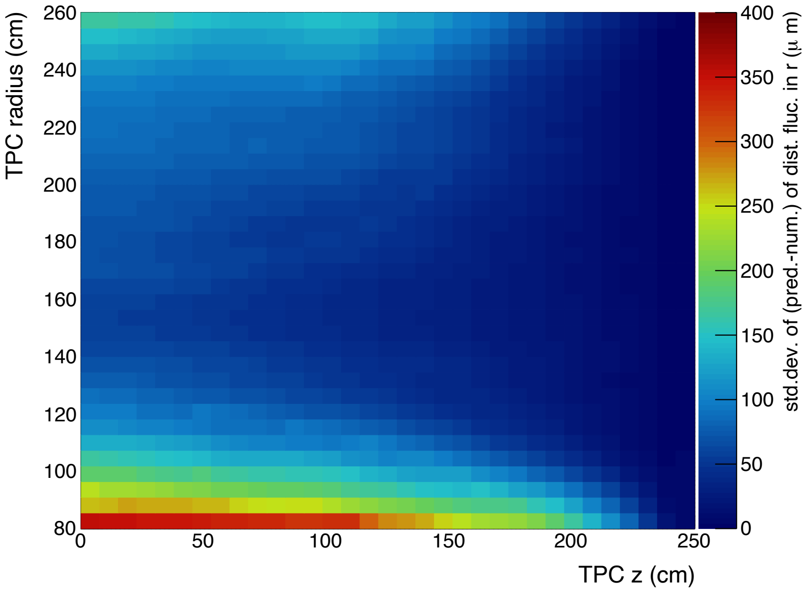
- Observed the model predictions with different values for the number of Fourier coefficients, as well as some of the model parameters (maximum depth and number of trees (for RF& XGB), number of hidden layers (for NN))
- Results with different number of Fourier coefficients will be presented today

Results with different # of Fourier coeffs (RF, max. depth=10)

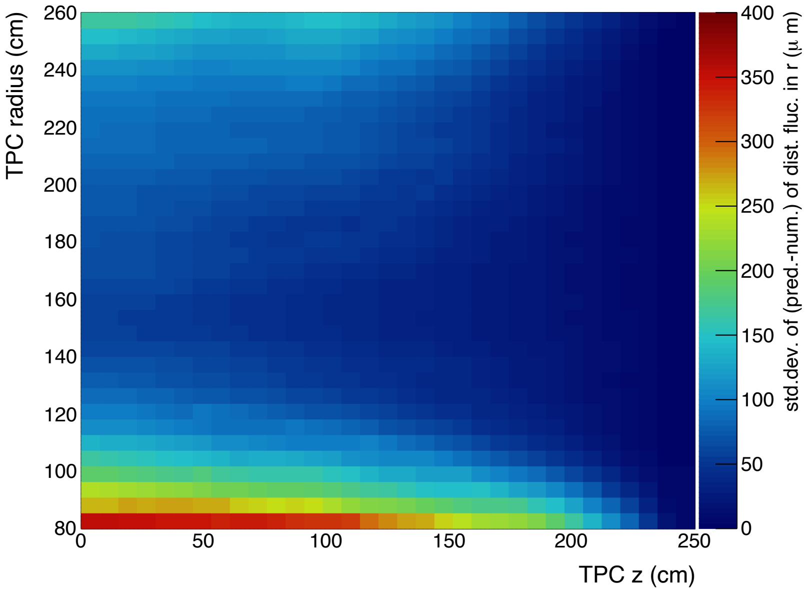
5 coefficients



10 coefficients

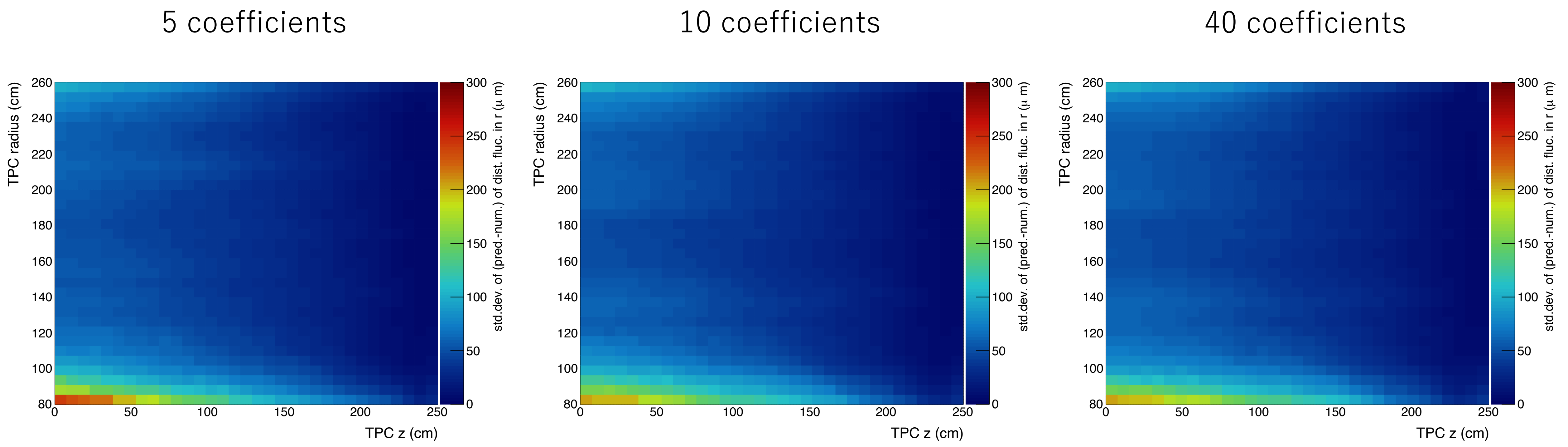


40 coefficients



No big difference in predictive performance
Standard deviation ~300 μm, which is above the required resolution of 200 μm

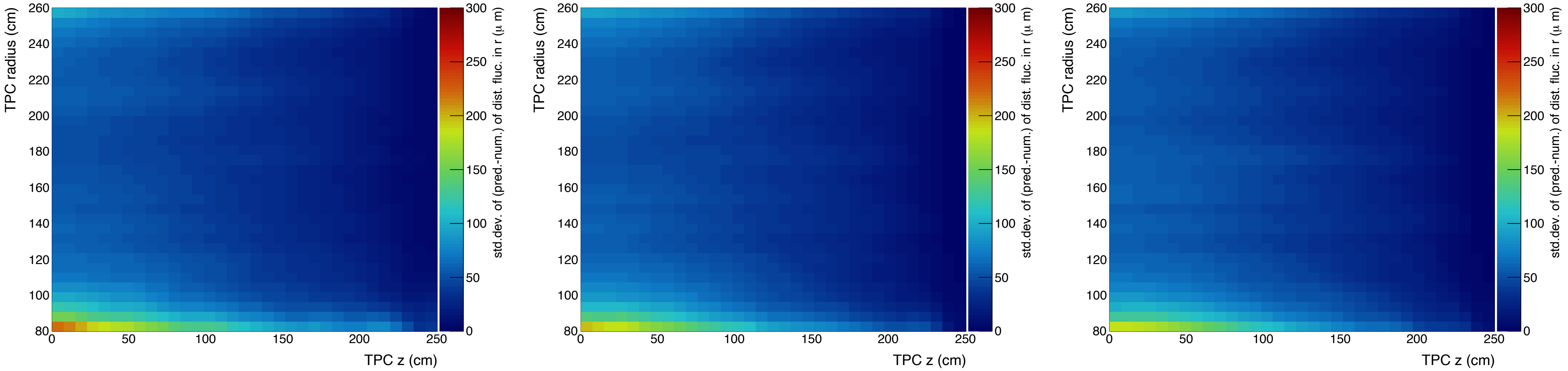
Results with different # of Fourier coeffs (XGB, 490 trees, max. depth=10)



Standard deviation improves from 250 to 200 μm when the number of coefficients are increased from 5 to 10.
Not much improvement for # of coefficients 10→40

Results with different # of Fourier coeffs (NN, 10 hidden layers)

5 coefficients　　10 coefficients　　40 coefficients



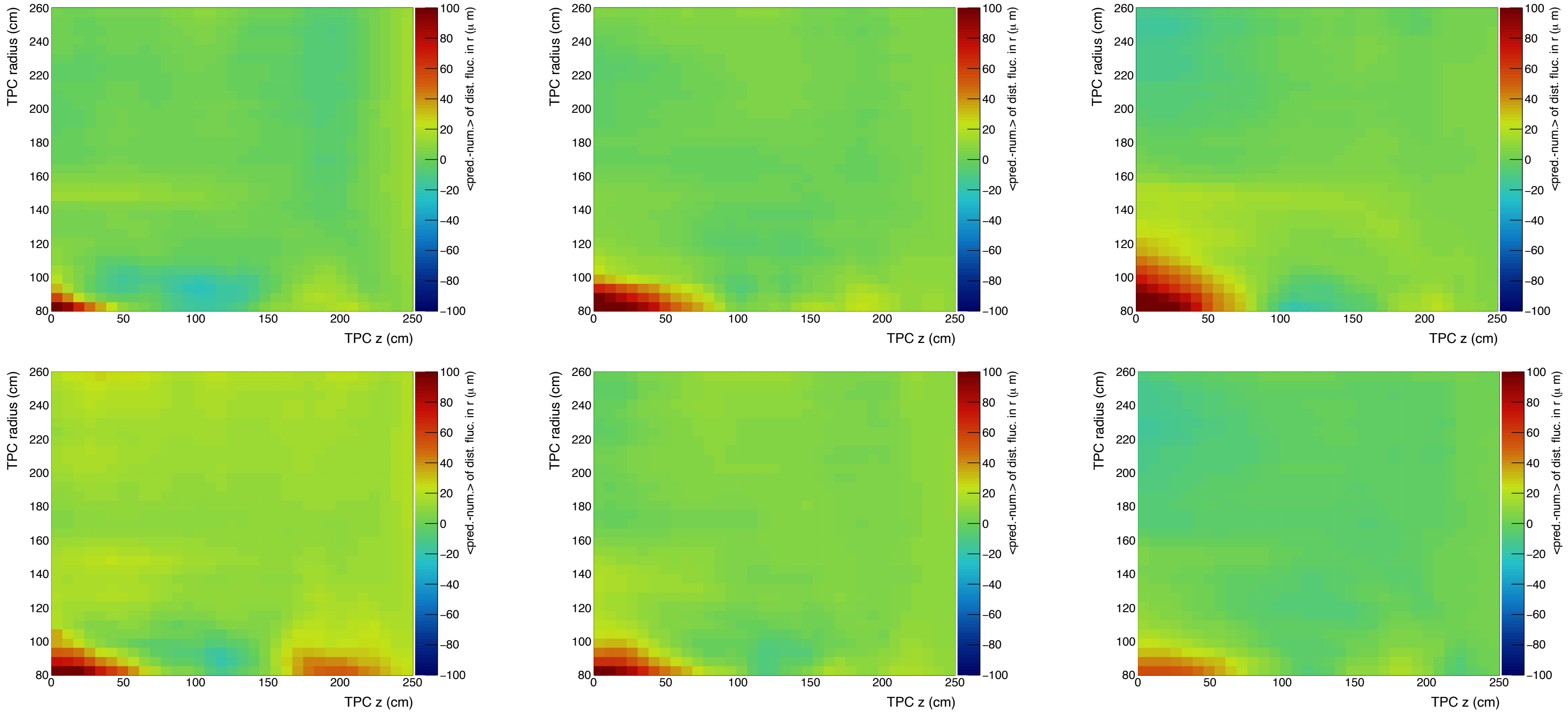
The predictive performance improves as the number of coefficients is increased.
The standard deviation reaches around 180 μm with 40 coefficients
However, the mean fluctuates (shown in a later slide).

Behavior of the NN mean plots

5 coefficients | 10 coefficients | 40 coefficients

8 layers

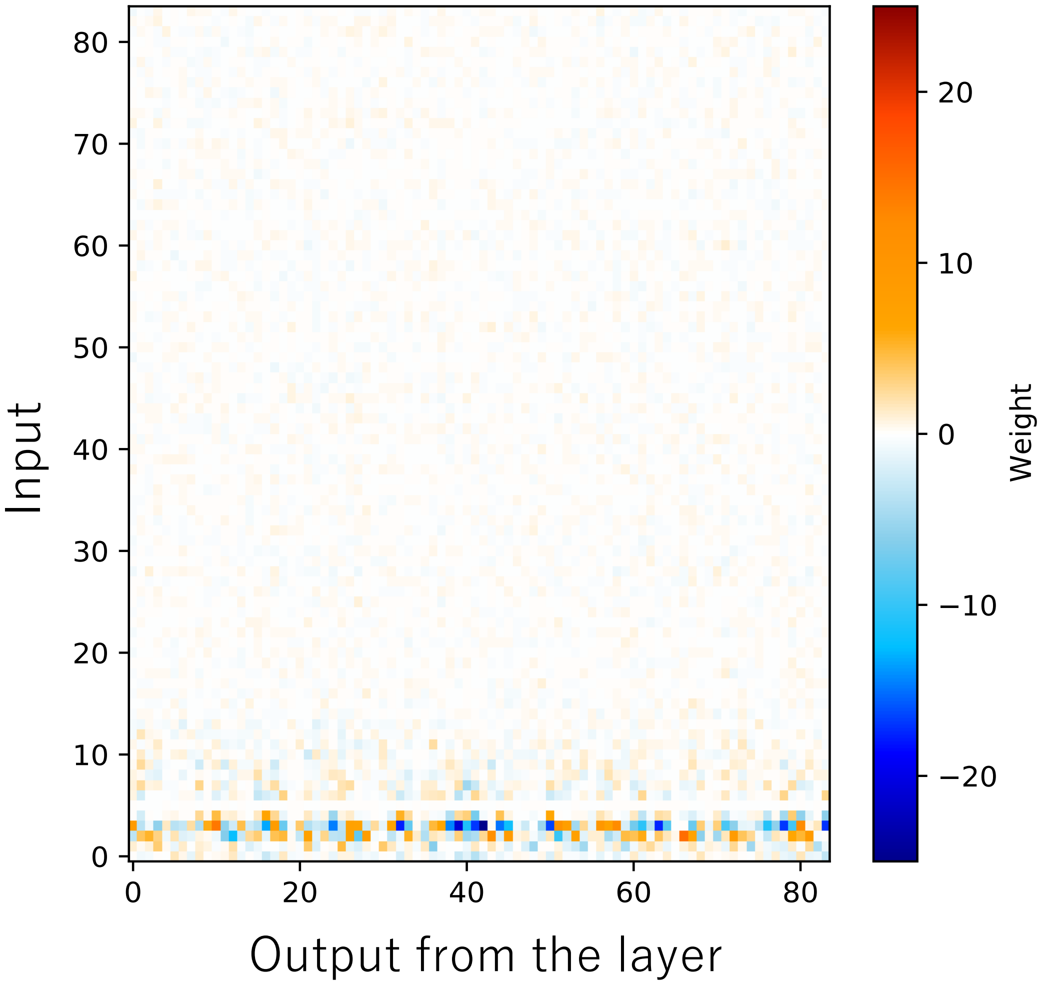
10 layers



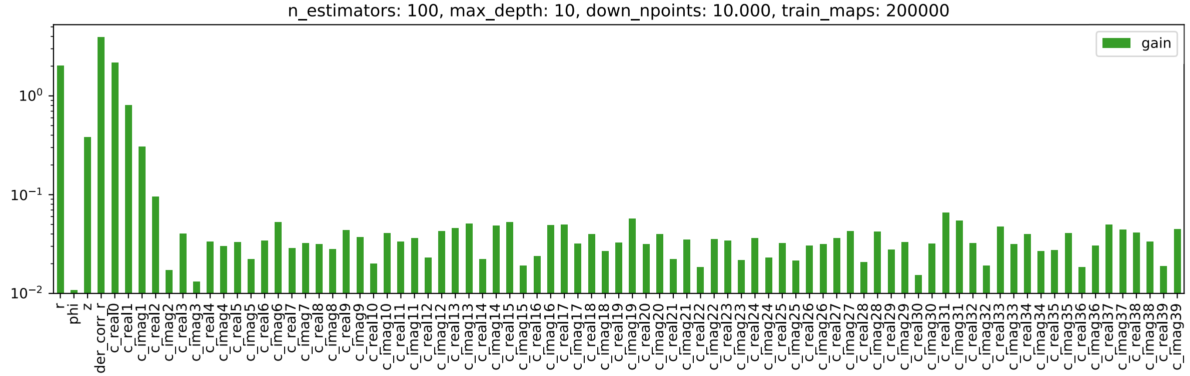
Shows a complex behavior that still needs to be further investigated

Feature Importance / Weights

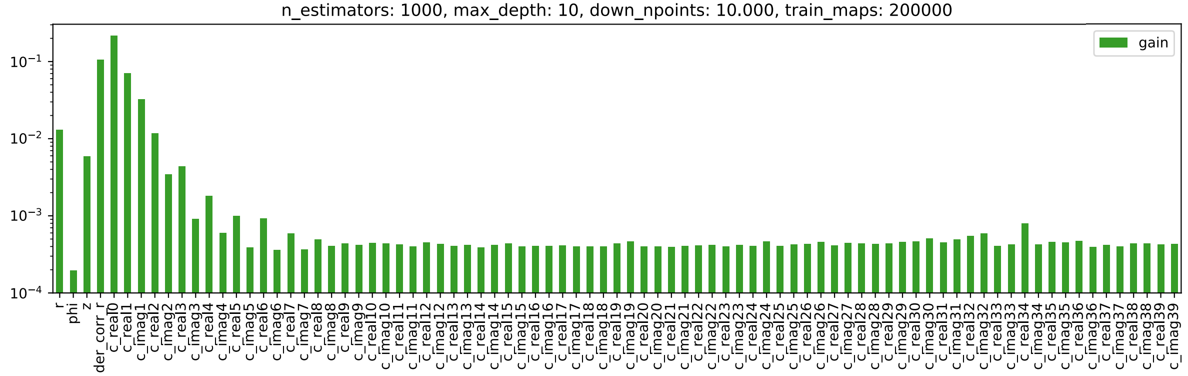
Weight value of the NN input layer



RF



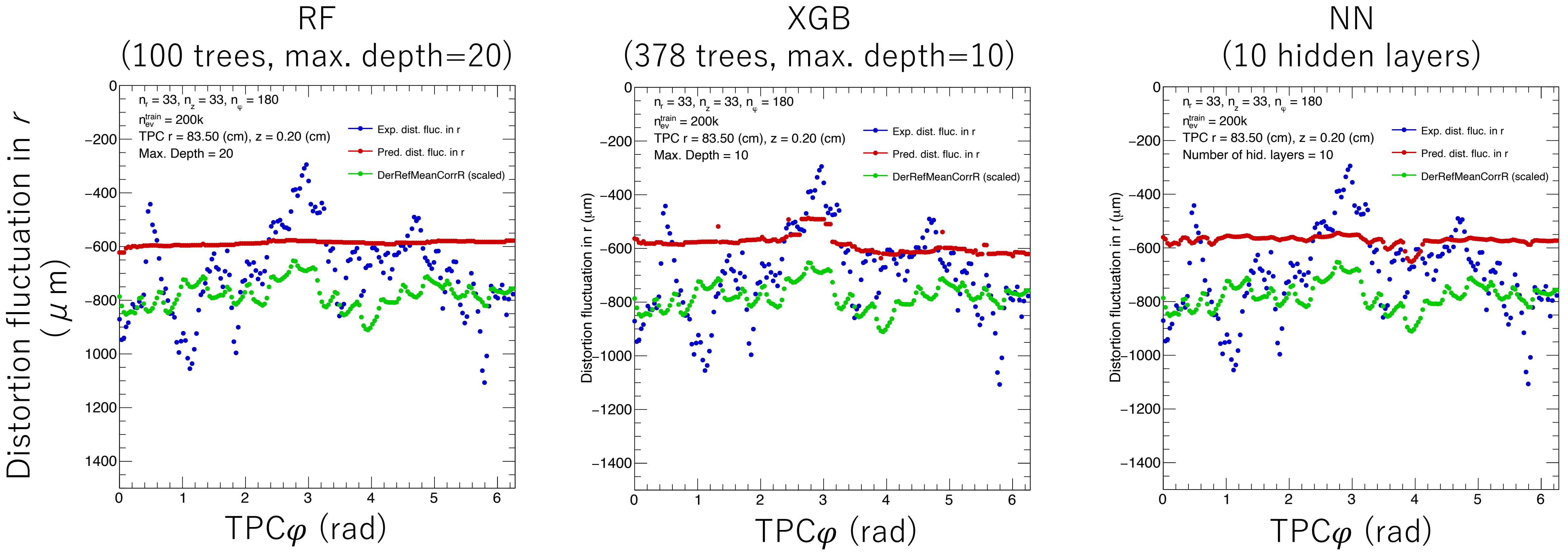
XGB



- Every model values the first ~10 coefficients

φ dependency of the results

φ dependency using 10 Fourier coeffs at r=83.5 cm and z=0.2 cm



- All the models learn part of the φ dependency, likely coming from the derivatives.
- Education of the φ dependency is something one should work on next.

Summary and next steps

- First ~10 Fourier coefficients important for the distortion fluctuation correction
- Standard deviation ~250 μm at minimum using Random Forest
- XGBoost showed performances within expectations in a wide area of the TPC
- The neural network showed exceptional performance in standard deviation, but might be too "sensitive" to the input values
- φ dependency still needs to be educated

- Should be possible to educate the φ dependencies by developing an additional 3D correction
- Should start adapting the framework so that this correction could be applied to actual data

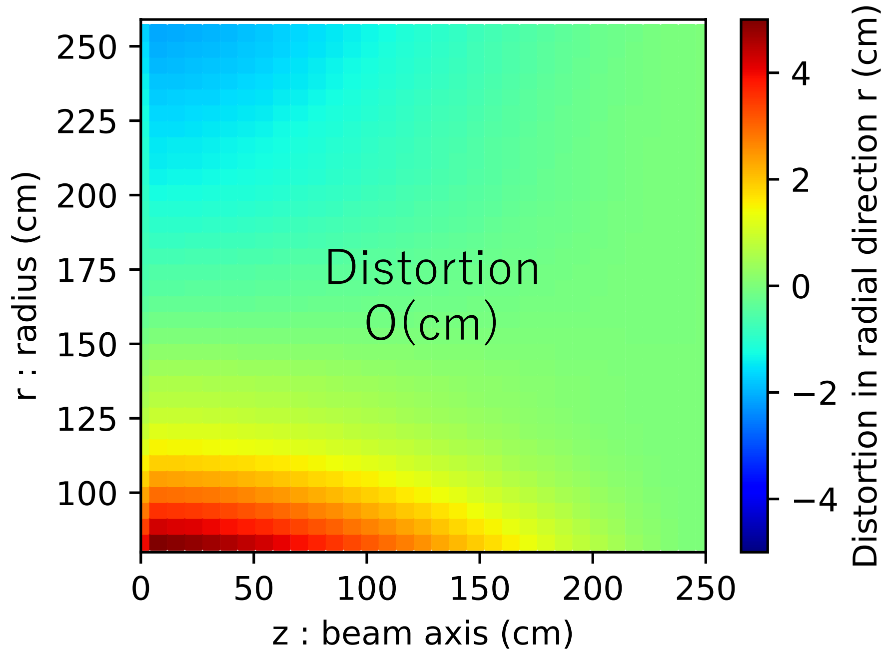
Backup

Simulation of the space-charge effect

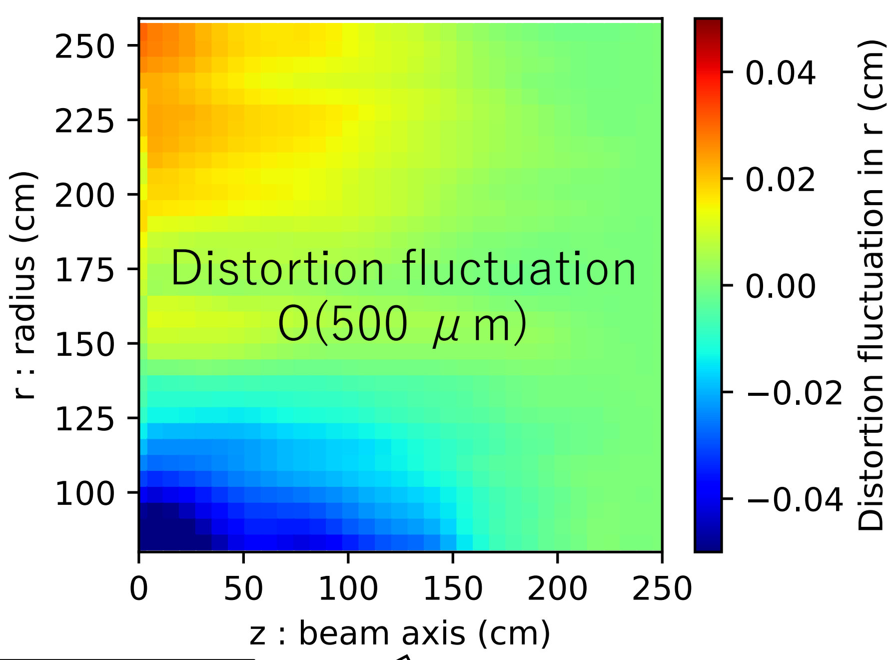
1. Generated 100,000 Pb-Pb collisions using the PYTHIA 8 Angantyr model
2. Chose a random set of 8,000 collisions
3. Calculated the charge observed at the readout pads as well as the space-charge density using the measured amplification factor & IBF of the GEM
4. Calculated the track distortions using the Poisson & Langevin equations
5. Created 1,000 of these SCD/distortion maps
6. Averaged these 1,000 maps, creating an average SCD/distortion map (which correspond to an average over ~2 min.)

Distortion Correction Strategy

Correct the O(cm) distortion to O(200 μm) through **synchronous** and **asynchronous** correction



Interpolation from external detectors (synchronous, O(min))

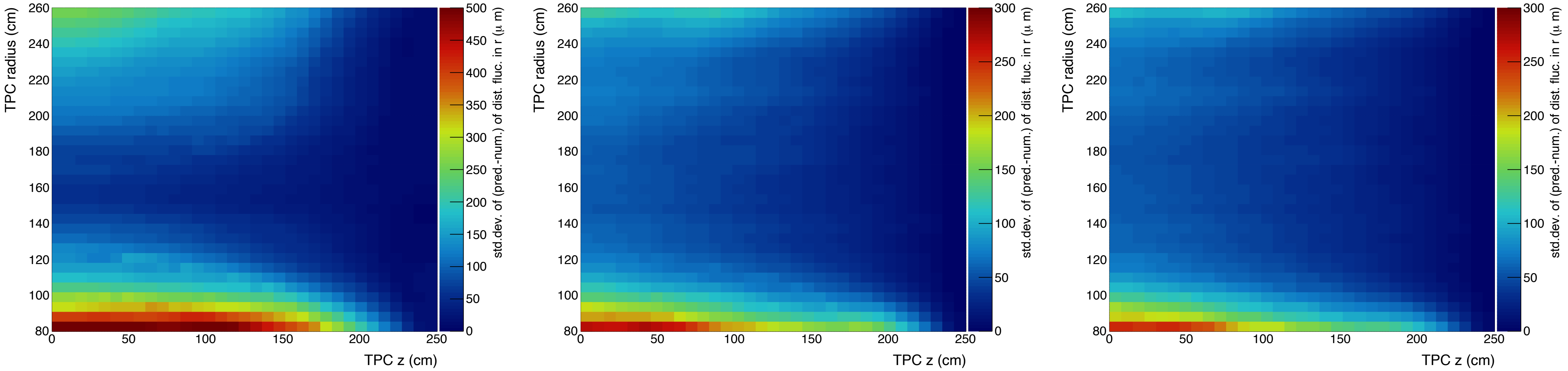


Lower the drift track distortion to the TPC intrinsic spatial resolution (~200 μm)

Use machine learning during asynchronous processing O(5~10 ms)

Dependency on maximum depth (RF)

Max. depth = 6 | Max. depth=14 | Max. depth=20

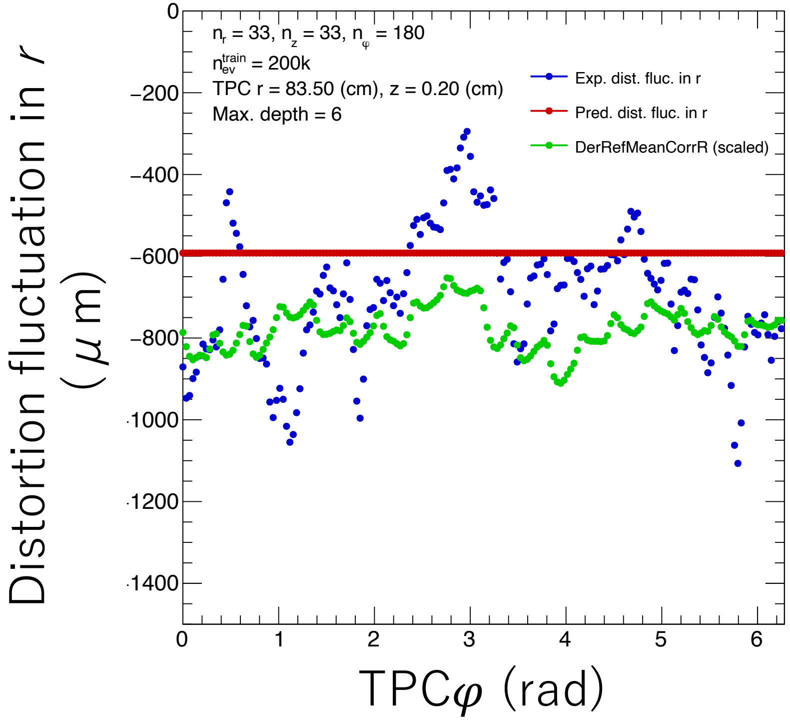


- Performance improves as max. depth is increased
- Standard deviation at around 250 μm when max. depth = 20

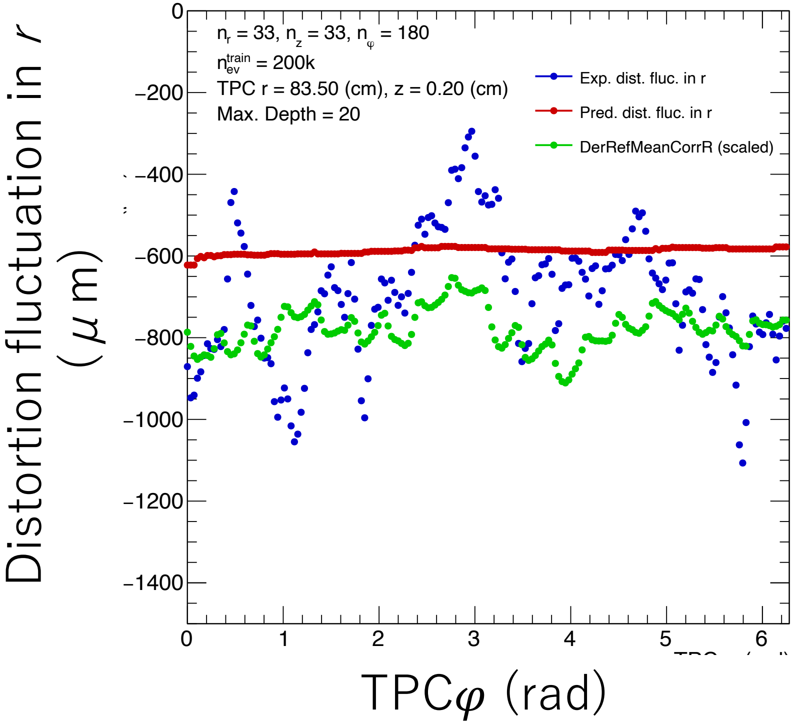
φ dependencies of the results (RF)

φ dependency at r=83.5 cm, z=0.2 cm

Max. depth = 6

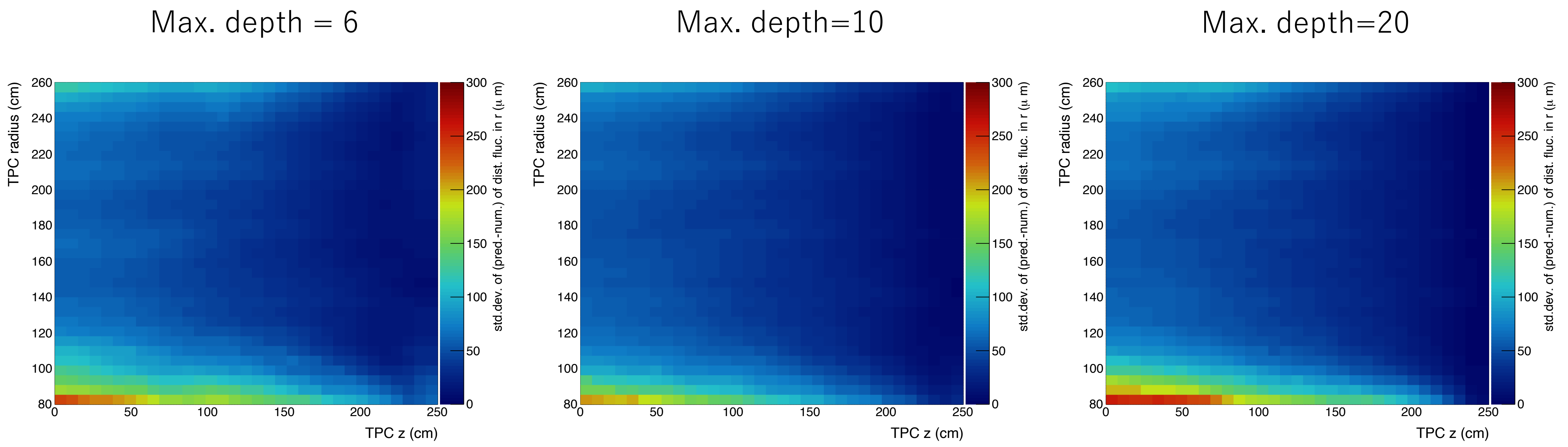


Max. depth = 20



- The model is not learning the φ dependency at all

Dependency on maximum depth (XGB, 100 trees)

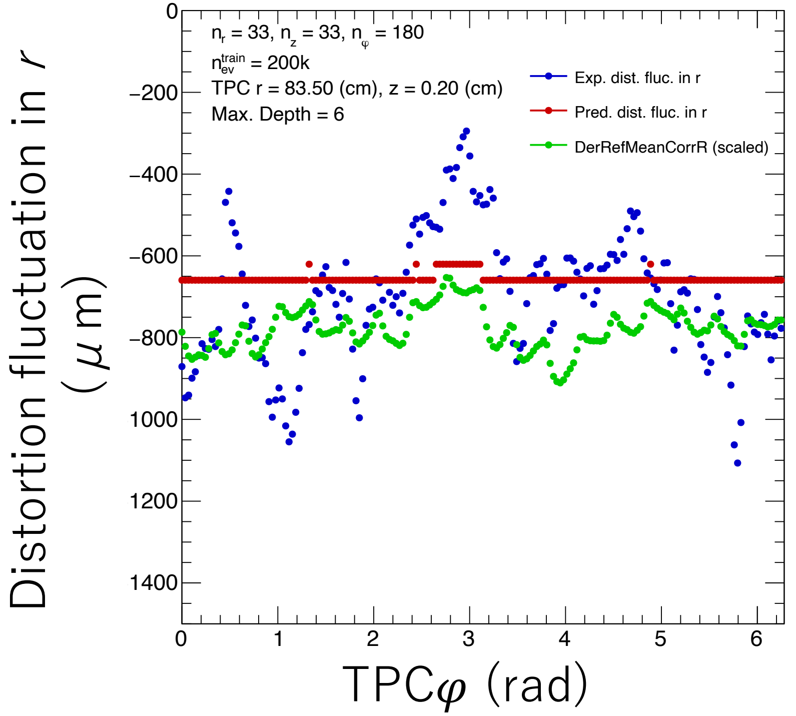


- The model shows best performance (std. dev. ~200 μm) at max. depth = 10

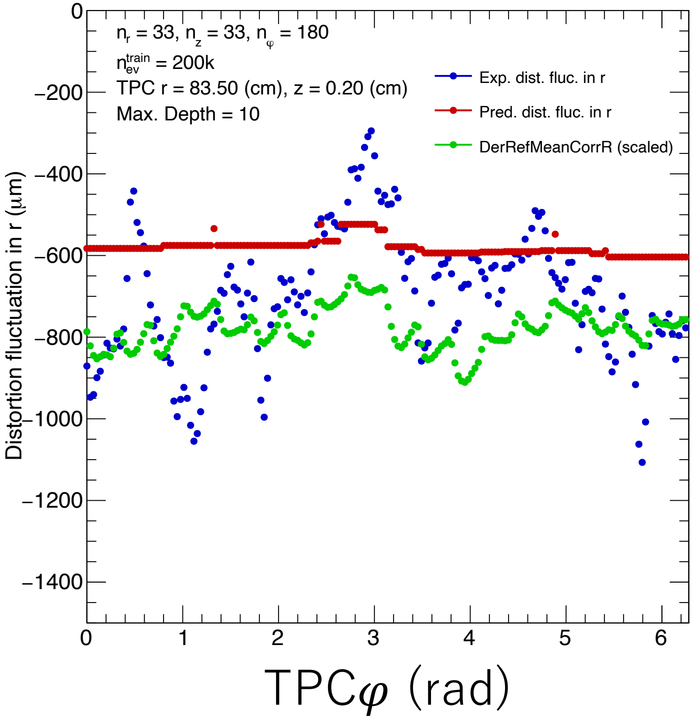
φ dependency(XGB, 100 trees)

φ dependency of the predictive performance with max. depth = 10 at r=83.5 cm, z=0.2 cm

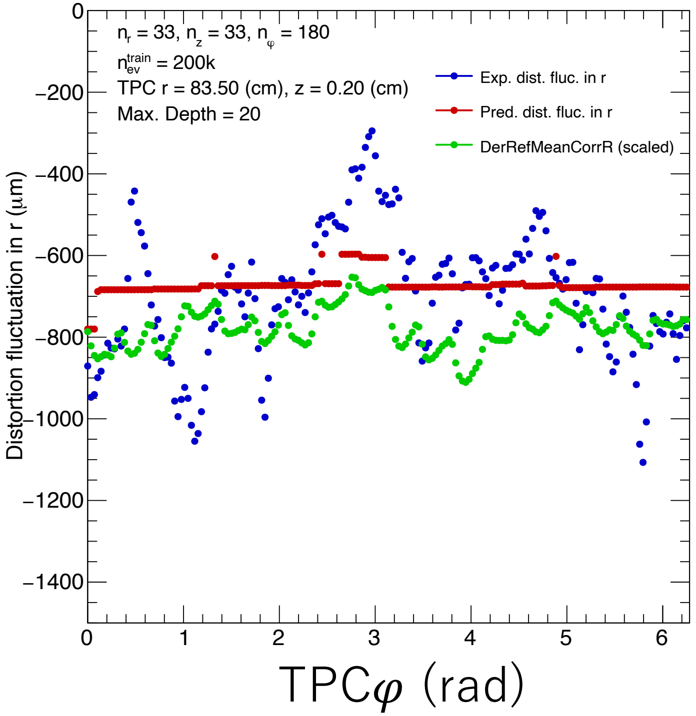
Max. depth = 6



Max. depth = 10

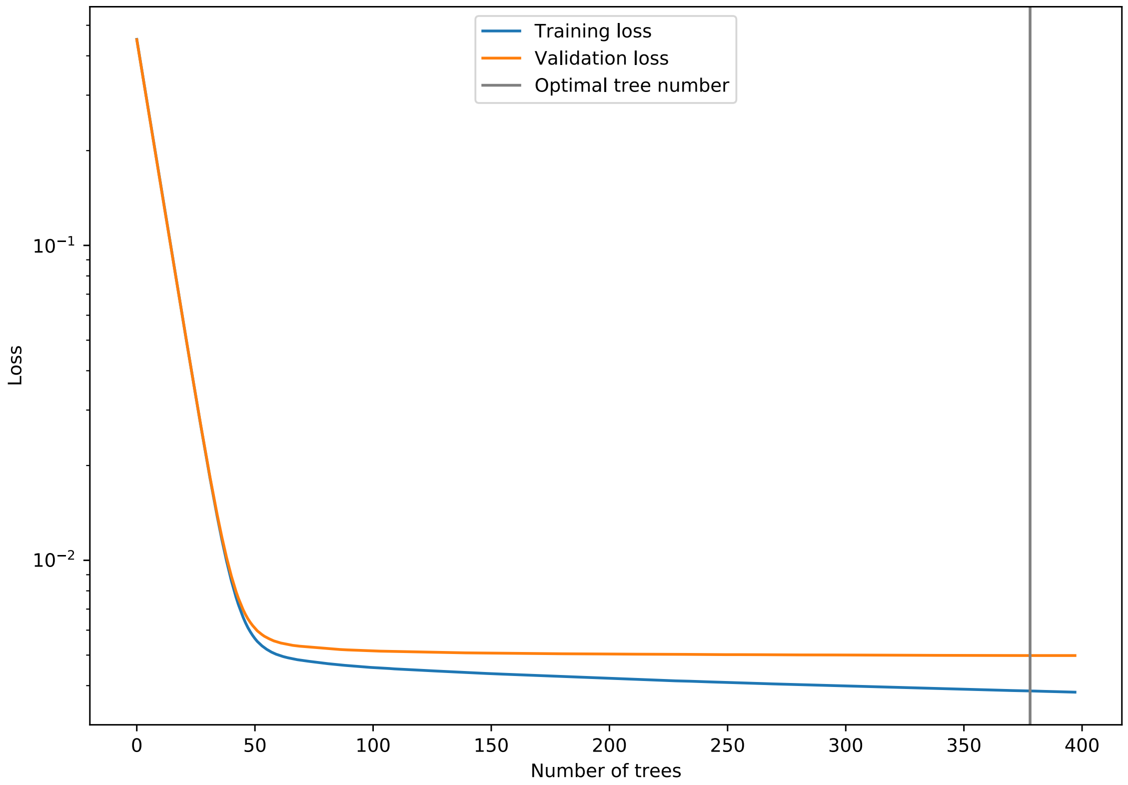


Max. depth = 20



- XGB learns φ dependencies at smaller scales as max. depth is increased from 6 to 10
- No significant difference between max. depth=10 and 20

Learning curve of XGBoost



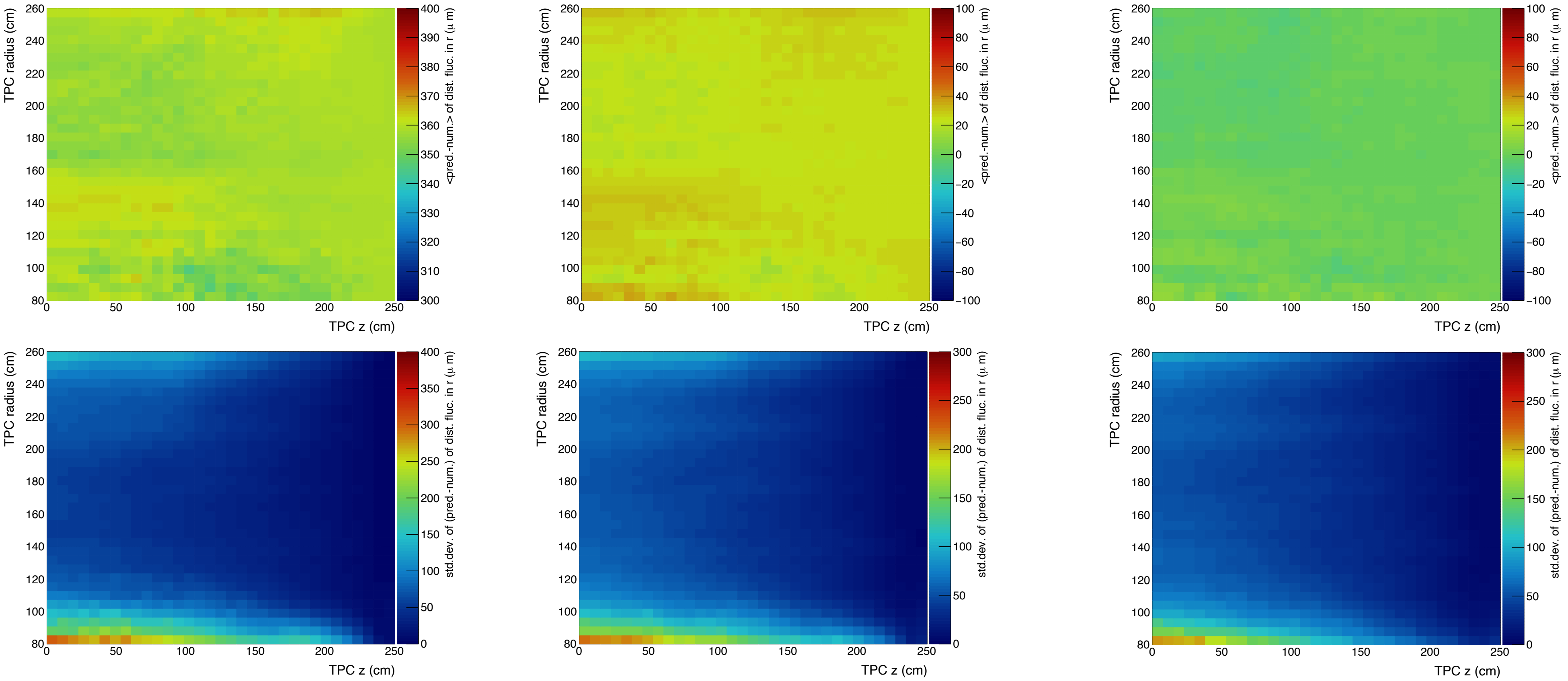
Validation error stops improving at ~50 trees

Dependency on # of trees (XGB)

25 estimators | 50 estimators | 378 estimators

Mean

Standard deviation



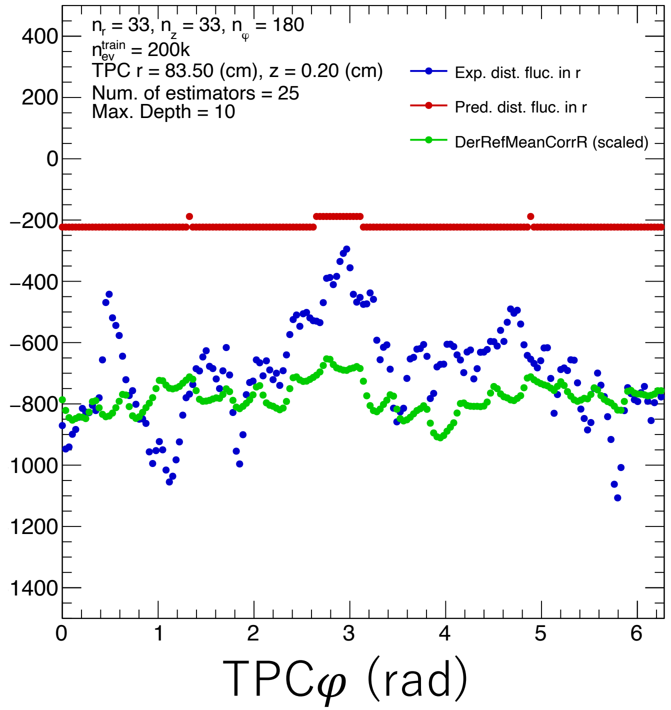
The mean improves from 360 μm→20 μm→ 0 μm with num. of trees 25→50→378
The standard deviation goes from 300 μm→250 μm→200 μm

Dependency on # of trees (XGB, φ direction)

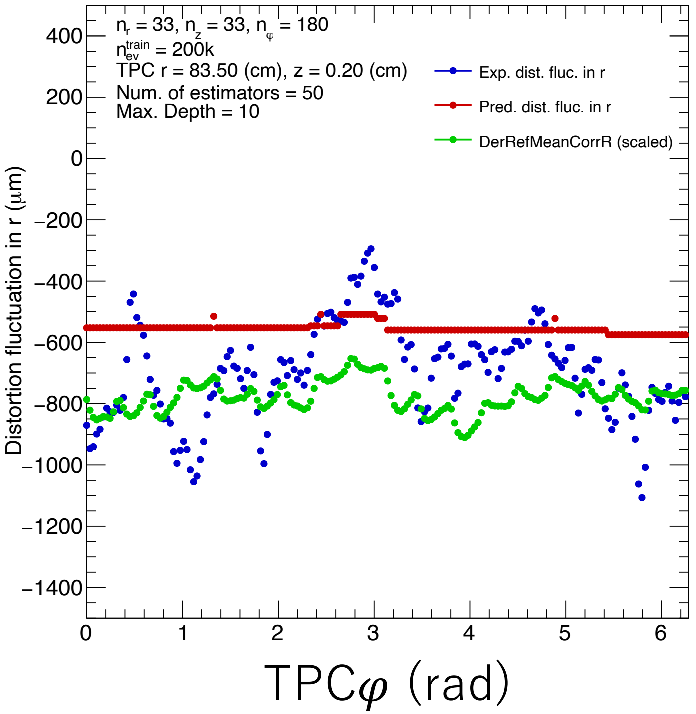
φ dependency with max. depth = 10, r=83.5 cm and z=0.2 cm

Num. of estimators = 25

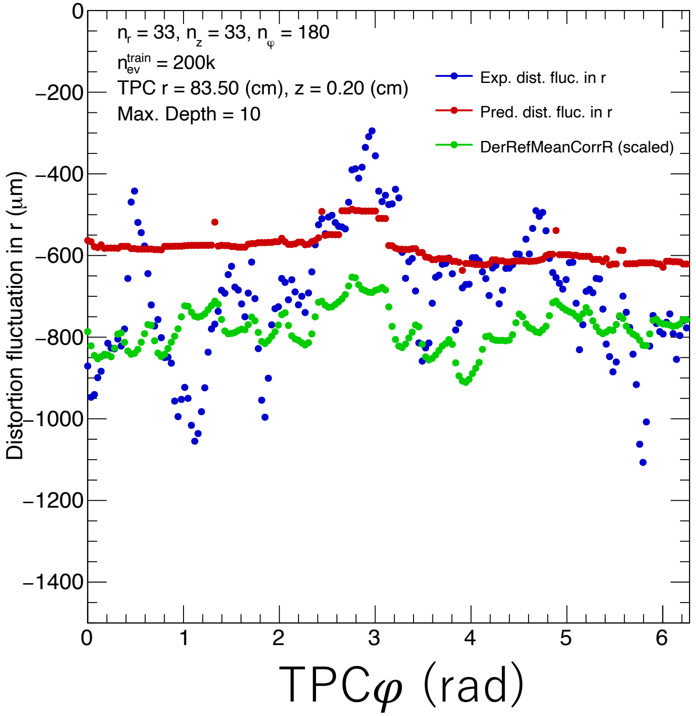
Distortion fluctuation in r (μm)



Num. of estimators = 50



Num. of estimators = 378

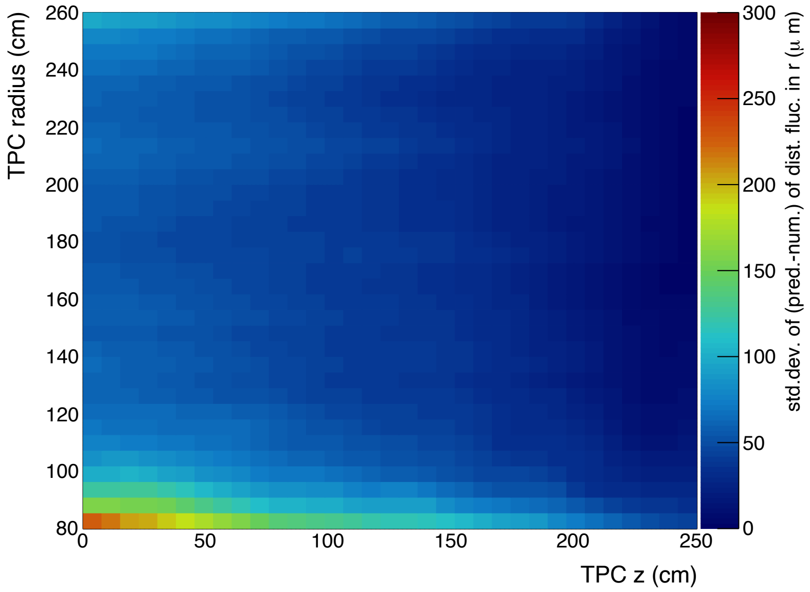


- Learning order: (φ dependency in a larger scale)→(Average size)→(φ dependency in a smaller scale)
- Learns part of the φ dependency likely coming from the derivatives.

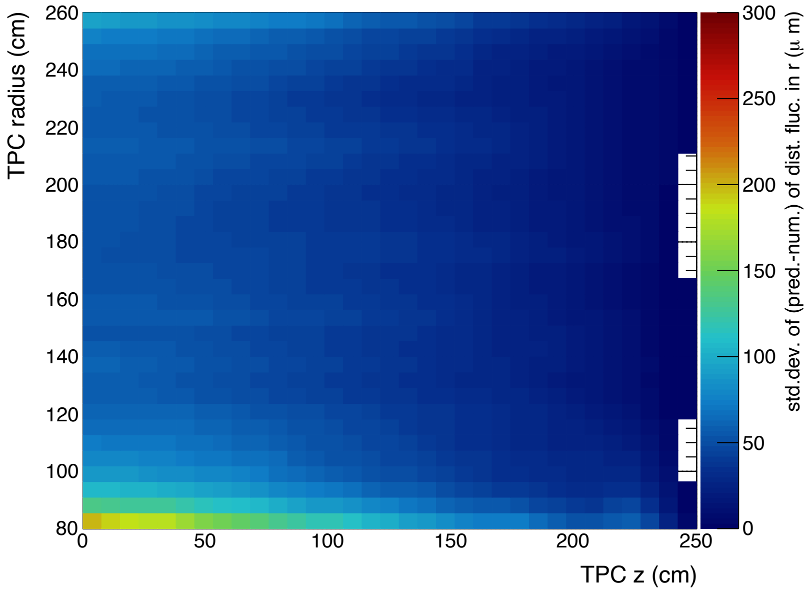
Dependency on # of hidden layers (NN)

Using 10 Fourier coefficients

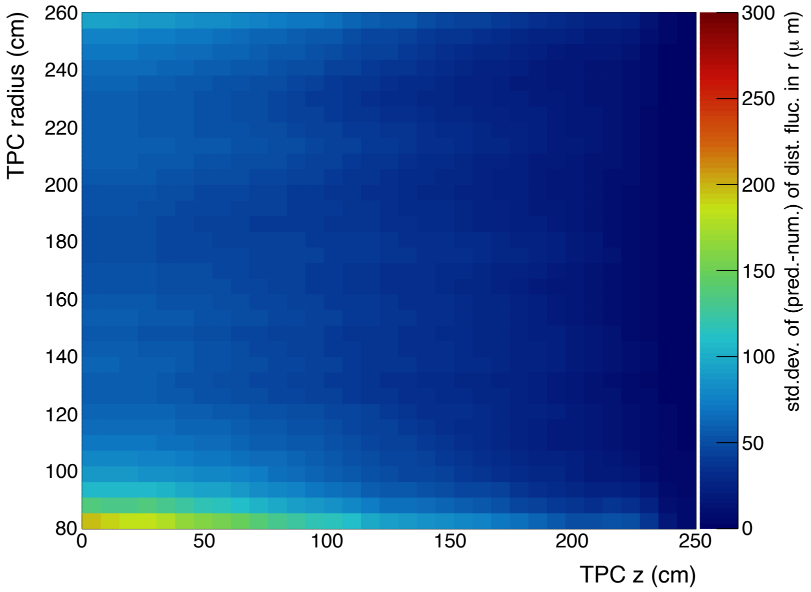
2 hidden layers



6 hidden layers



10 hidden layers

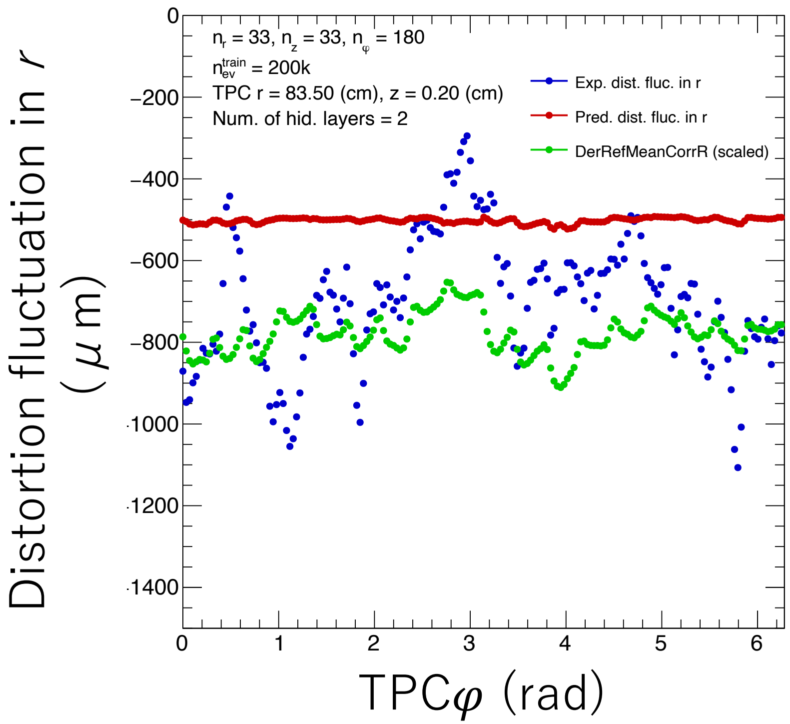


- Stops improving at ~6 hidden layers
- Standard deviation ~200 μm

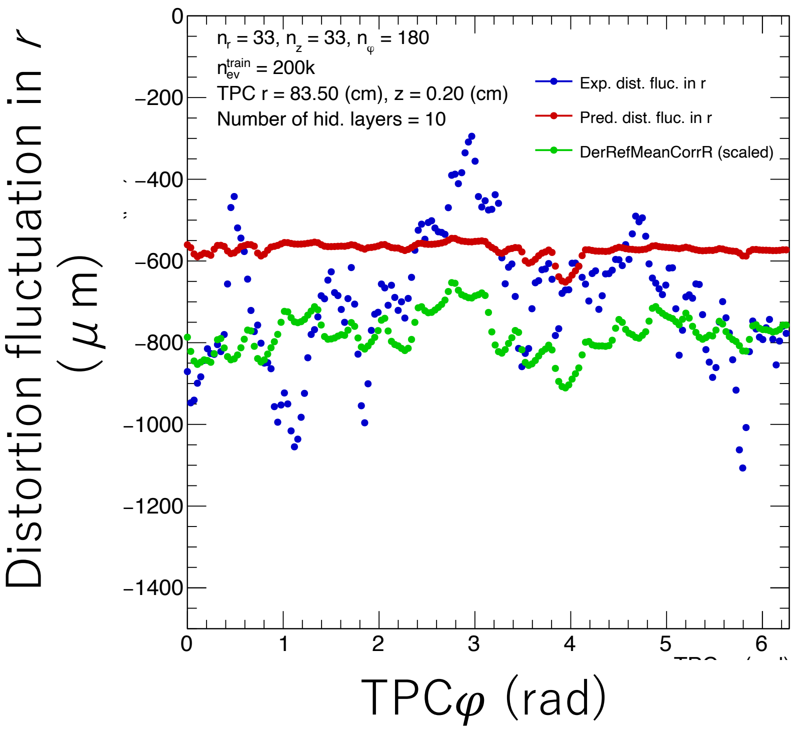
φ dependency of the results (NN)

φ dependency at r=83.5 cm, z=0.2 cm

2 hidden layers

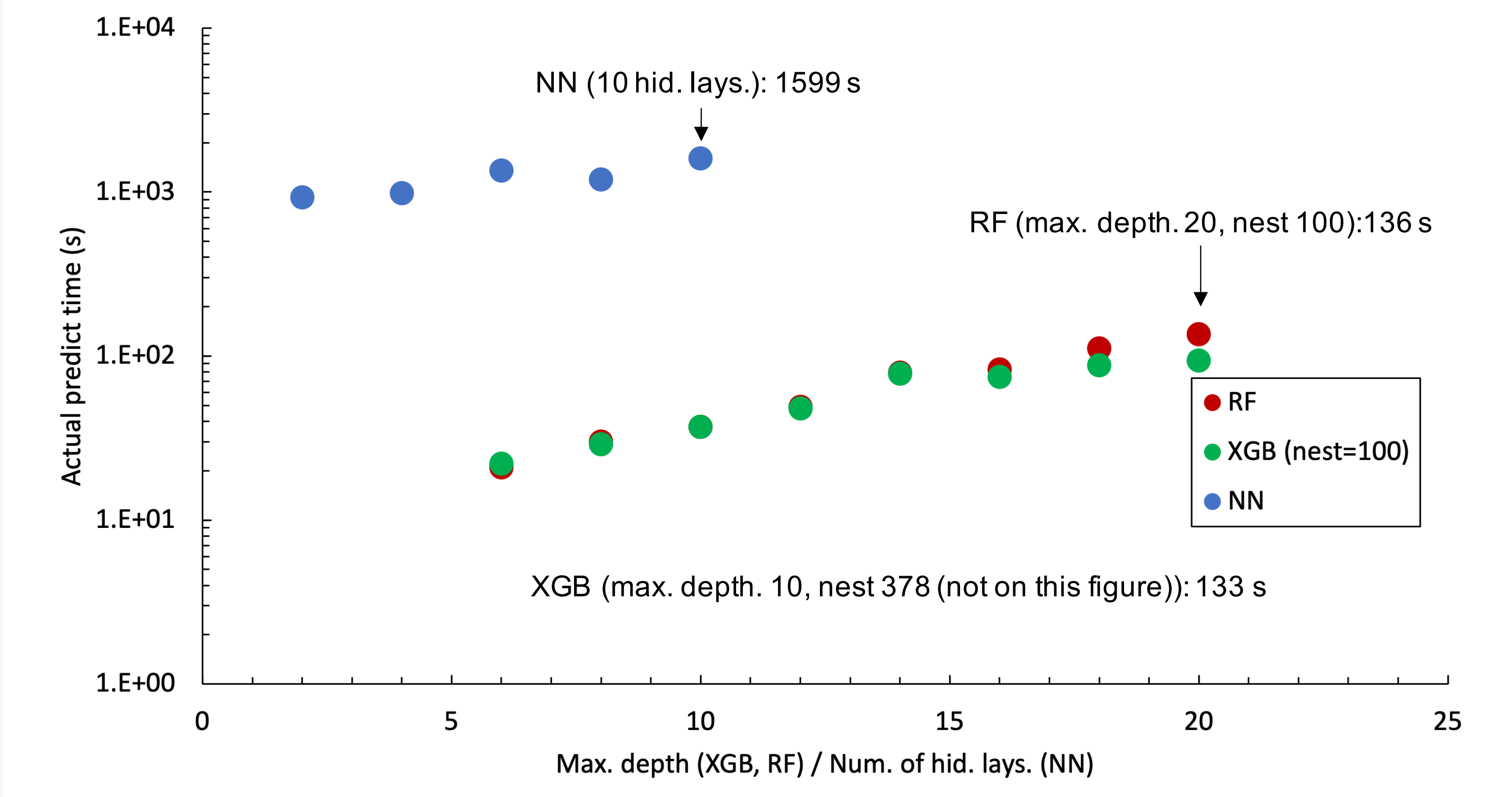


10 hidden layers



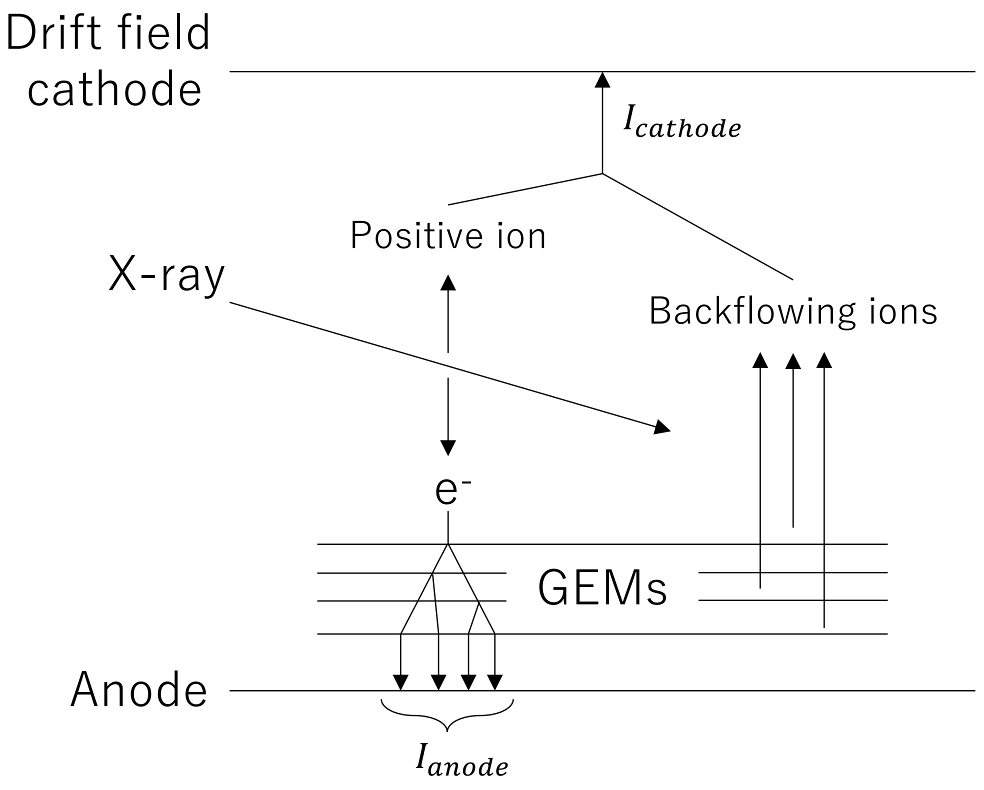
- NN learns more of the φ dependency from the derivative as number of hidden layers are increased
- Recreates the φ dependency from the derivative most precisely compared to the other models

Comparison of calculation speed



- Takes RF & XGB 35 weeks and the NN 400 weeks to process a 1 week run using 1 CPU only
- 200 EPNs all equipped with 8 GPUs available, which would enable calculation in a more realistic speed

Measurement of the IBF and ε



- Measured the IBF と ε map by moving the X-ray source
- IBF $\equiv$ $|I_{cathode}/I_{anode}|$ ≈(backflowing ions)/(number of ions created through the process)
- Adjusted the GEM voltage to make the gain ≈ 2000

→Calculated the relative gain compared to the average gain by

Rel. Gain $\equiv$ $|I_{anode}/\langle I_{anode}\rangle|$

- Calculated ε by

$$\varepsilon = IBF \cdot 2000 \cdot rel.gain$$

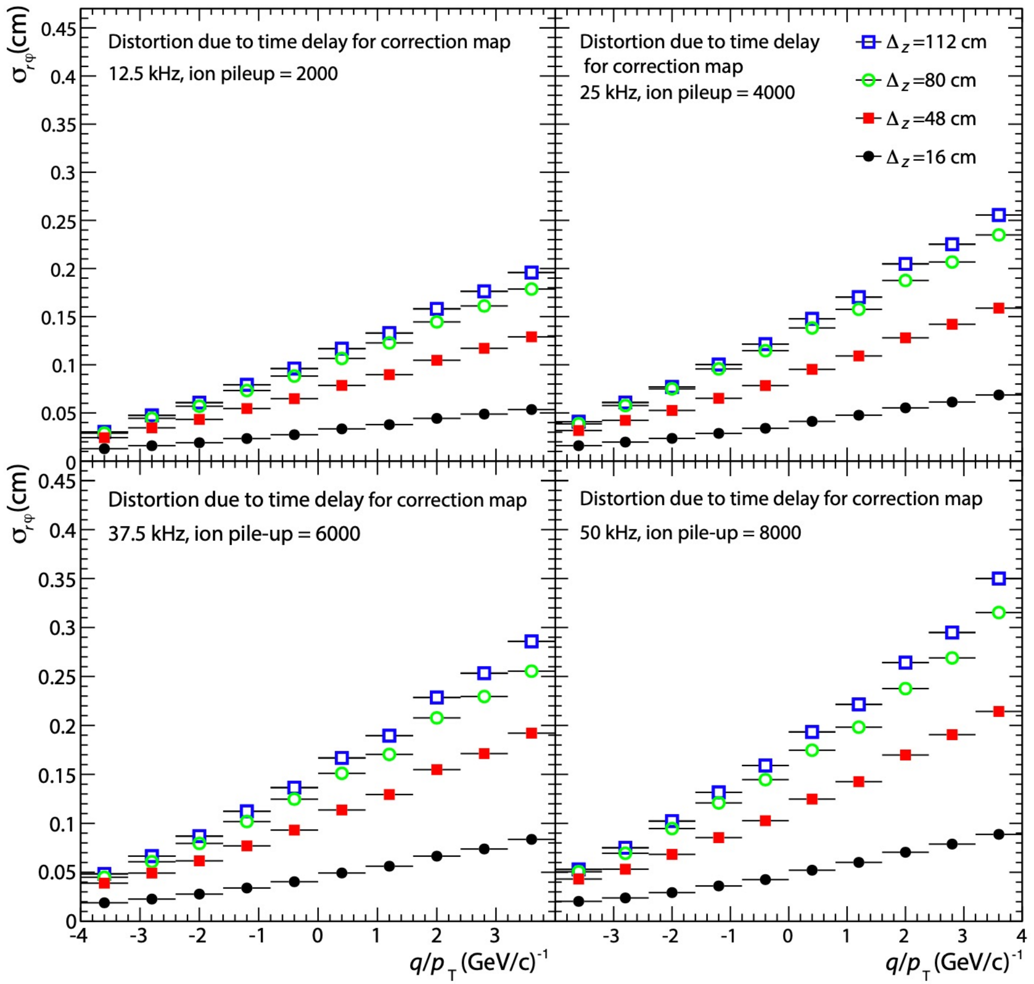
Time scales of the distortion fluctuation

- Calculate the distortion using a certain space-charge density map
- Then, calculate the distortion using the same density map shifted by Δ_z
- For 8,000 pileup events, Δ_z needed to be dropped to 16 cm for the difference between the two to become O(100 μ m)

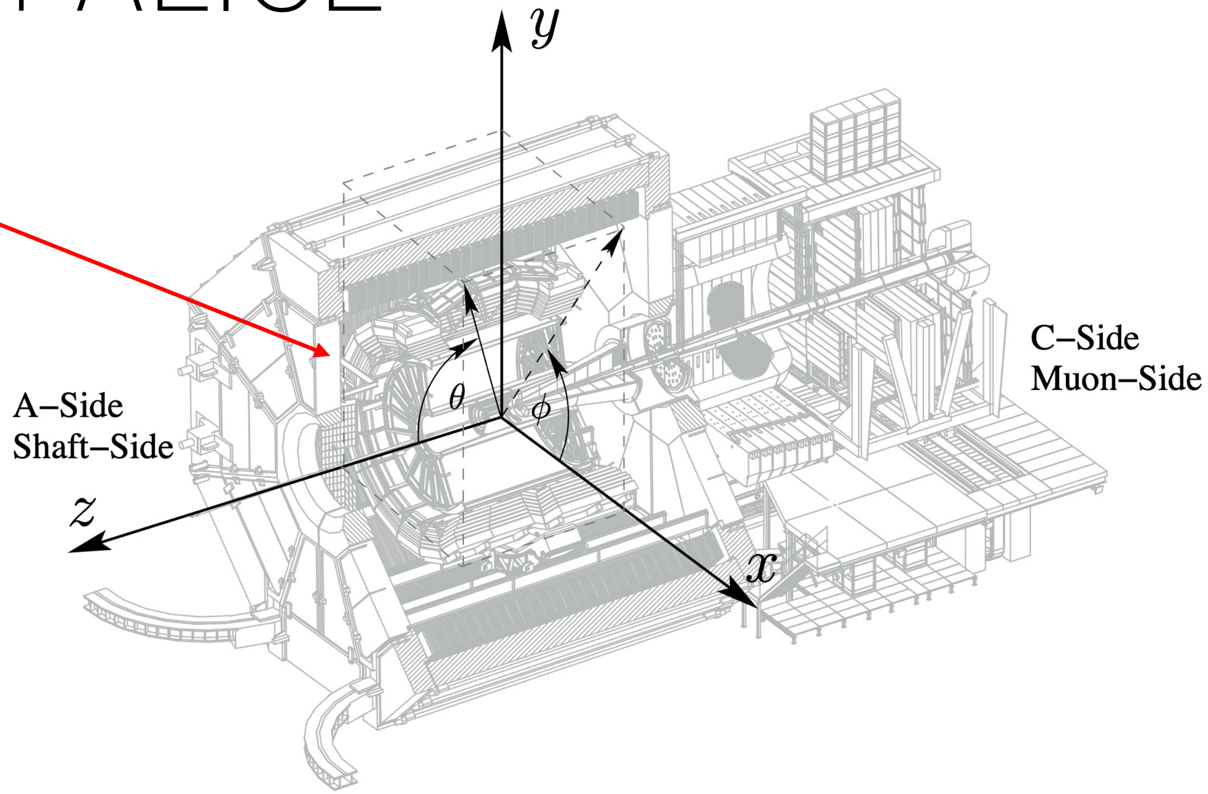
→Relevant time scale at around 5~10 ms

(calculated using the positive ion drift speed ~1.6 cm/ms)

Figure from ALICE-TDR-016



Coordinate system of ALICE



- Right-hand system with the beamline as the *z*-axis
- The Muon Arm installed on the negative *z* direction

L. Betev and P. Chochula. *Definition of the ALICE Coordinate System and basic rules for Sub- Detector Components numbering*. ALICE-INT-2003-038. 2003. (https://edms.cern.ch/document/406391/2).