# Report for $K_s^0$ analysis method with machine learning

# Introduction

- $K_s^0$ is important to understand the vertical developing for the air shower.
- The decay mode of $K_s^0$ measured by LHCf detector is following:
  - $K_s^0 \to 2\pi^0 \to 4\gamma$ (branching ratio:30.7%)
  - In this case, 4 photon incident on LHCf detector and event criteria is very hard. So events reconstructed of 4 photon is very small with current method.
- I am developing the position reconstruction method of 4photon with using machine learning for $K_s^0$ analysis and report the status .

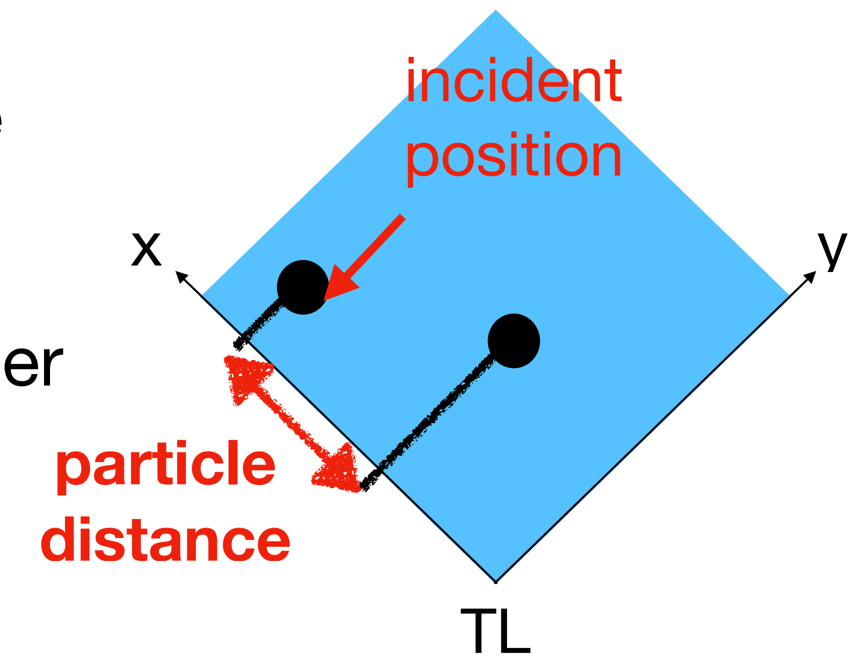
# Position reconstruction method

- Position reconstruction method is structured by the following programs.
  - 1.Search number of peak with TSpectrum
  - 2.Fitting using double lorentzian function corresponding to the number of peak.
- To reconstruct the positions of the four photons, the peaks must be far enough interval to be found by TSpectrum.
- I create each machine learning models for the peak search and the position predict.
  - Verified how accurately independent machine learning models for 1 and 2 can predict

# Position reconstruction method

- Position reconstruction method is structed by the following programs.
  - 1.Search number of peak with TSpectrum
  - 2.Fit to double lorentzian function corresponding number of peak

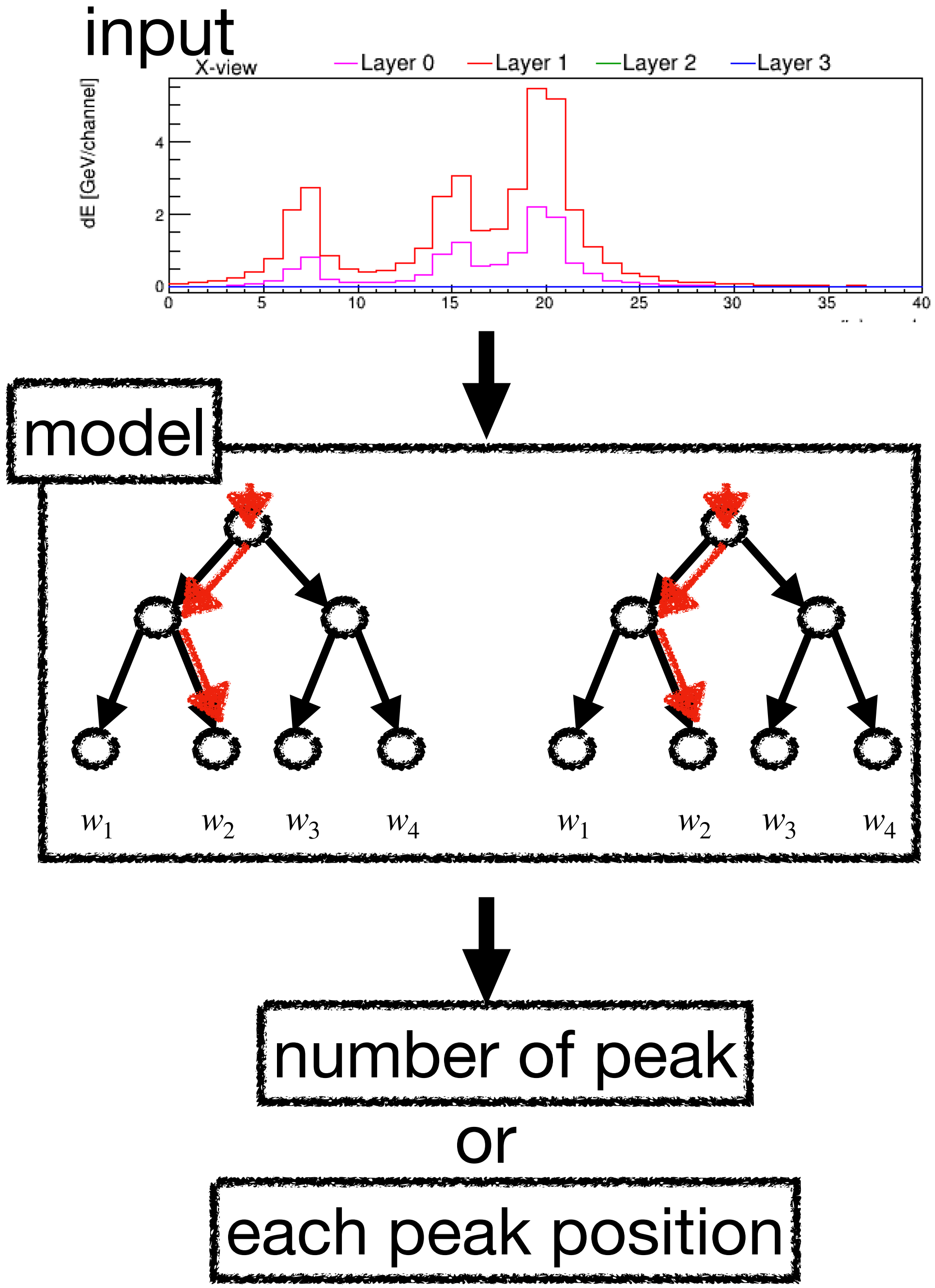
- I defined the parameter ‘particle distance’ as shown the particle apart on right fig.
  - Interval of x-axis in calorimeter coordinate at TL



# Gradient Boosted Decision Tree(GBDT)

- GBDT is a kind of machine learning model. It increase accuracy to be structured several Decision Tree not high performance.

- In this work, I used the library XGBoost both peak search and prediction of peak position.
    - For prediction of peak position, I make each model for each peak.



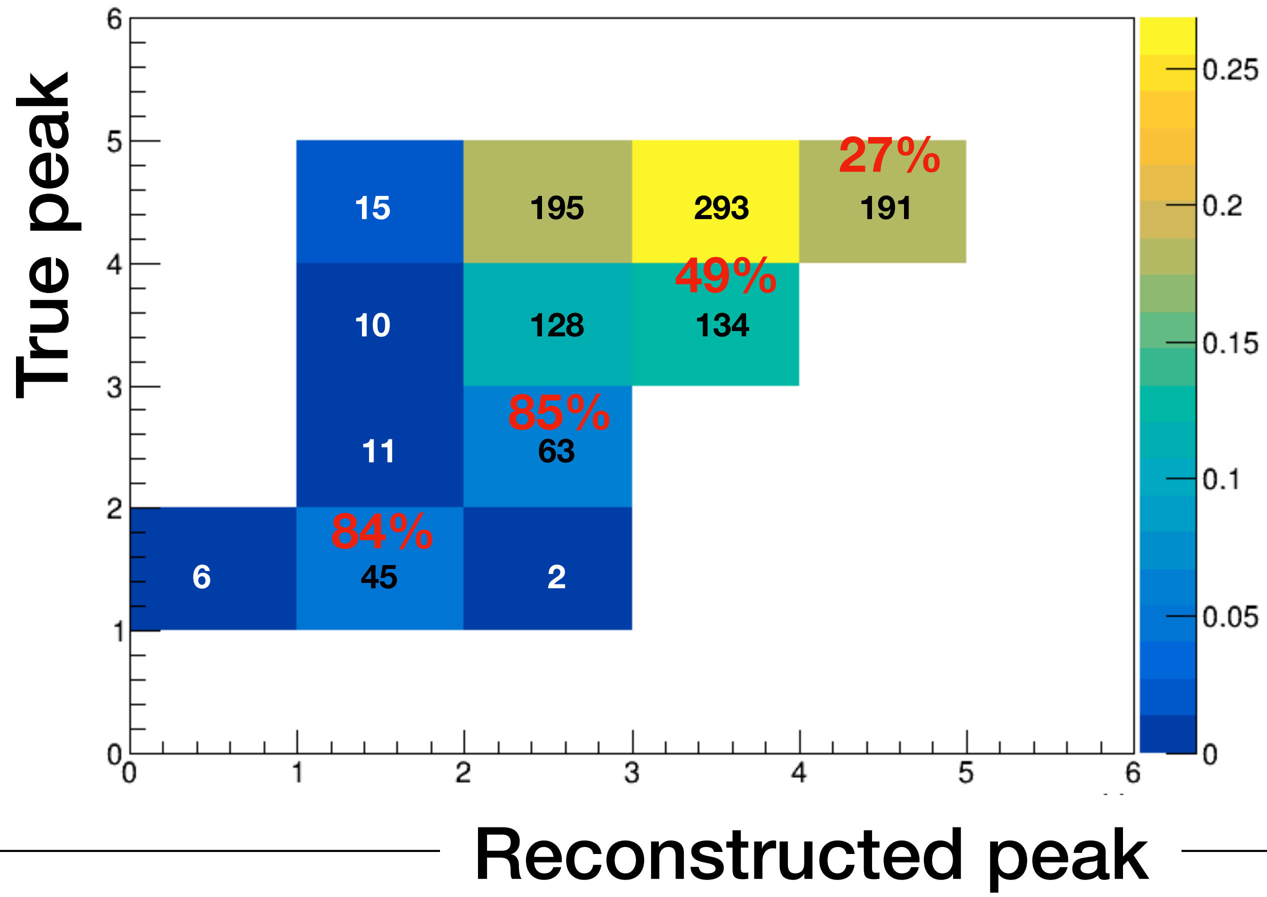
# Number of hit with current method

- true data
  - 4hit 694 event
  - 3hit 272 event
  - 2hit 74 event
  - 1hit 53 event

- Criteria
  - a photon energy > 100GeV
  - particle distance > 3.0mm

- Many 4hit events are mistaken for 3hit events, but never identify 3hit or 2hit events as 4hit.

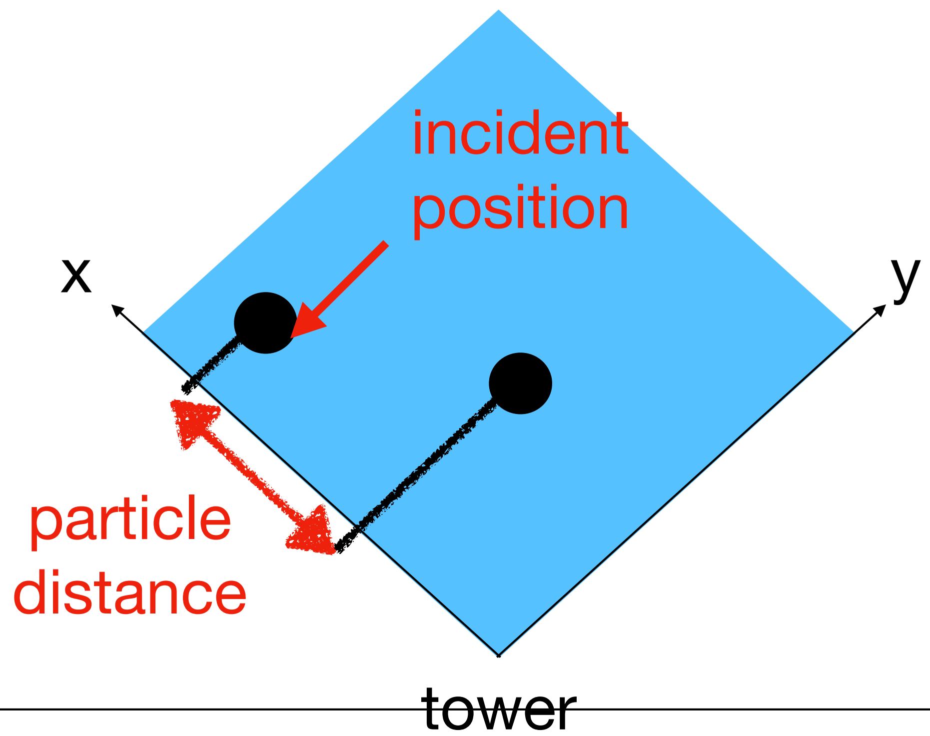
True peak vs Reconstructed peak (TL, X )



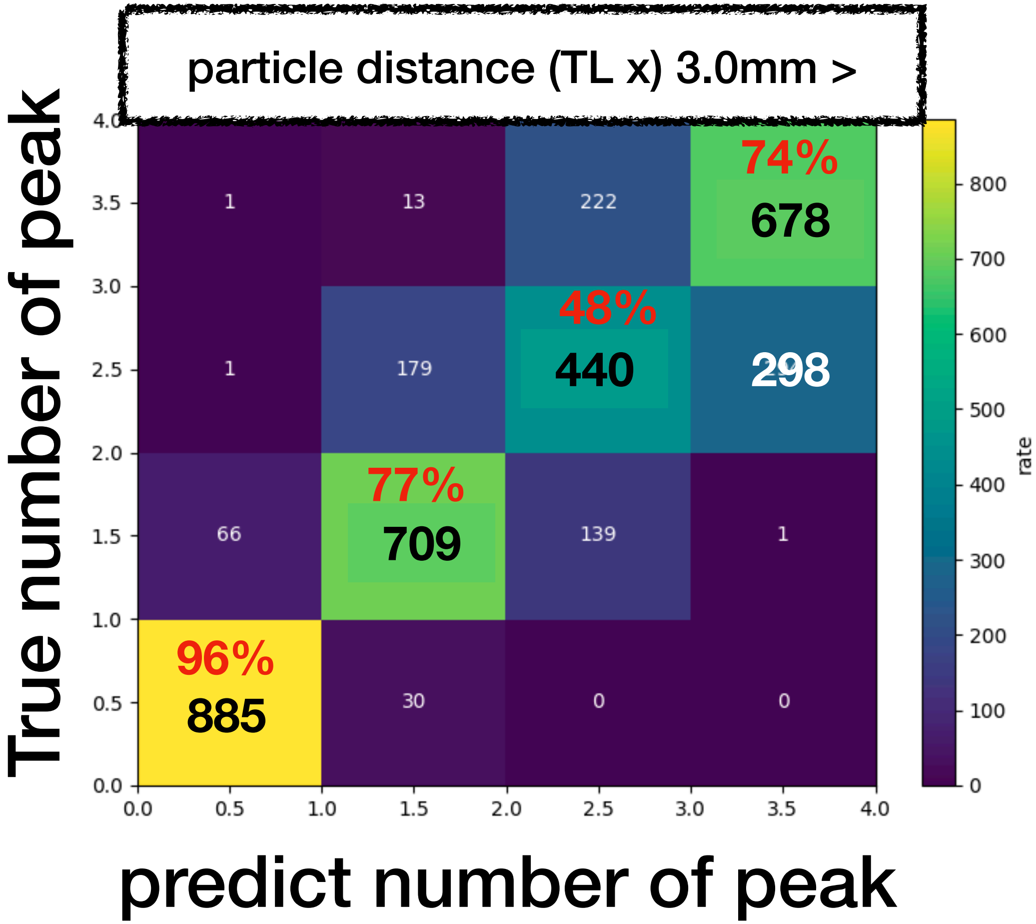
# number of peak with machine learning

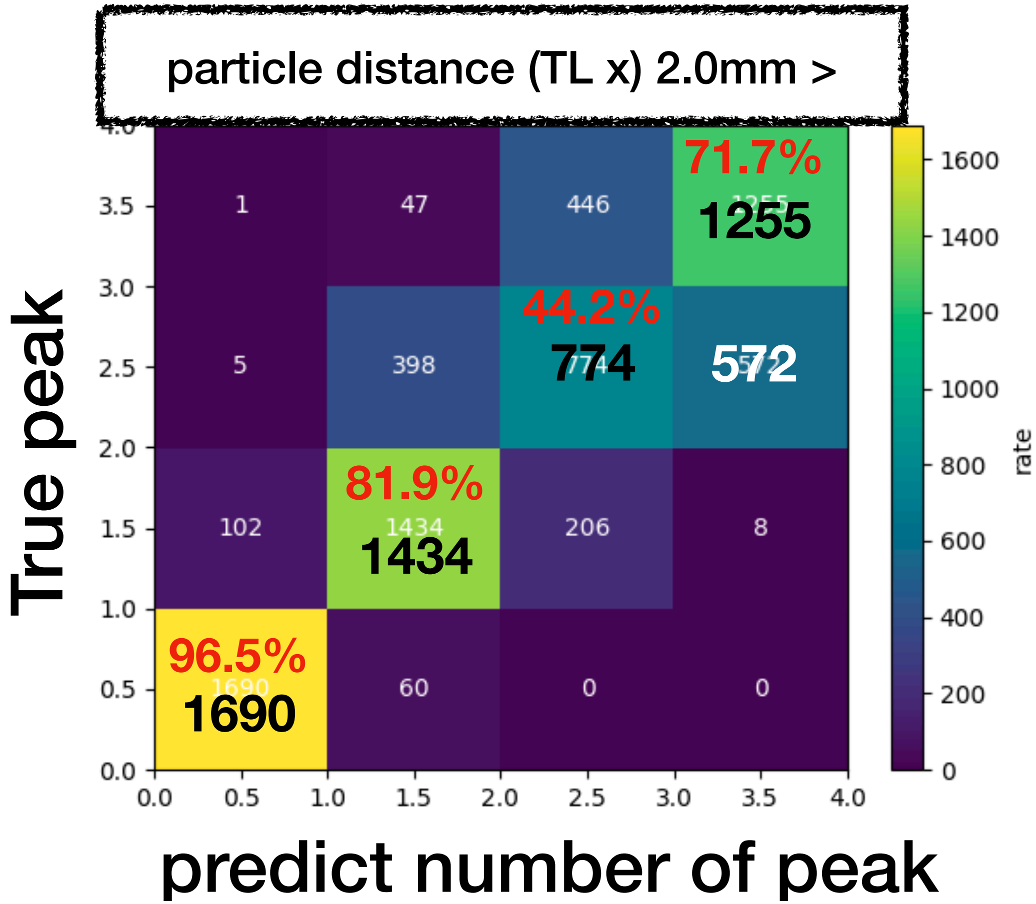
- Data set
  - energy deposit on position layer (the 1st & 2nd layer)
  - Large Tower (x axis)
  - one photon energy > 100GeV
  - particle distance > 3mm(3658) , 2mm(6998), 1mm(13624)

- hyper parameter
  - study rate 0.1
  - max_tree 4
  - number of study100



# Predict number of peak (xgboost)

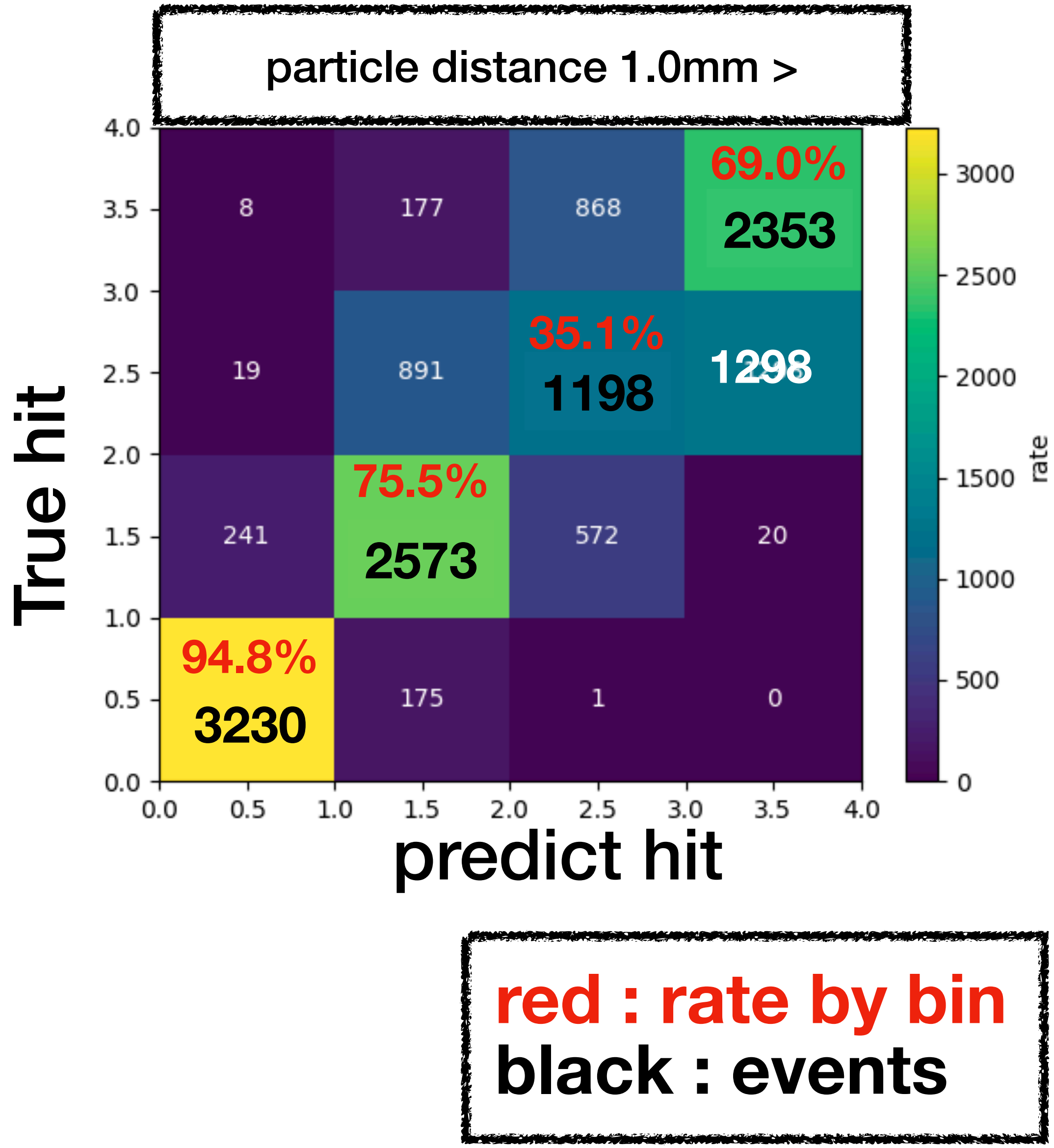



red : rate by bin
black : events

# Predict number of hit (xgboost)

- Compared with the results of the current reconstruction method, 4hit can be correctly determined even when the particle distance is more than 1 mm.
- Note the appearance of a 3-hit event that is mistaken for a 4-hit event.
  - For right figure, 30% of the events predicted to be 4hit are 3hit events.

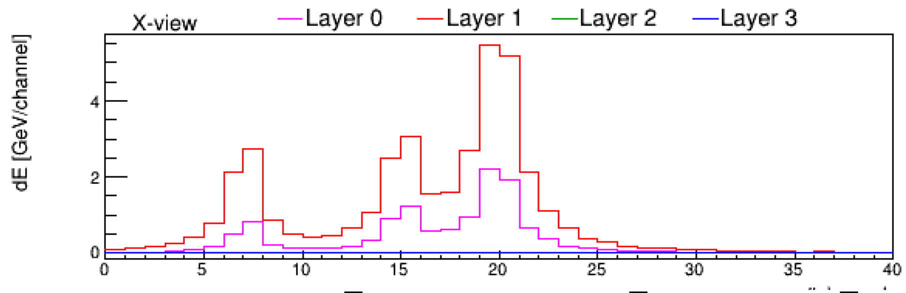


# predict position with machine learning

- Data set
  - Large Tower (x axis)
  - position layer (the 1st & 2nd layer)
  - one photon energy > 100GeV
  - particle distance > 1mm

- **I do not optimize the hyper parameter**

- Instead of predicting four peak locations from a single model, a model was created for each peak.



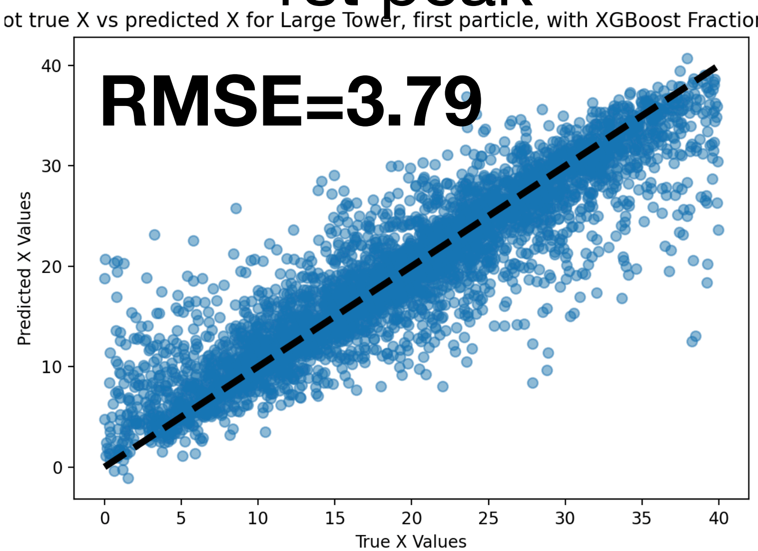

model4 model3 model2 model1

peak position

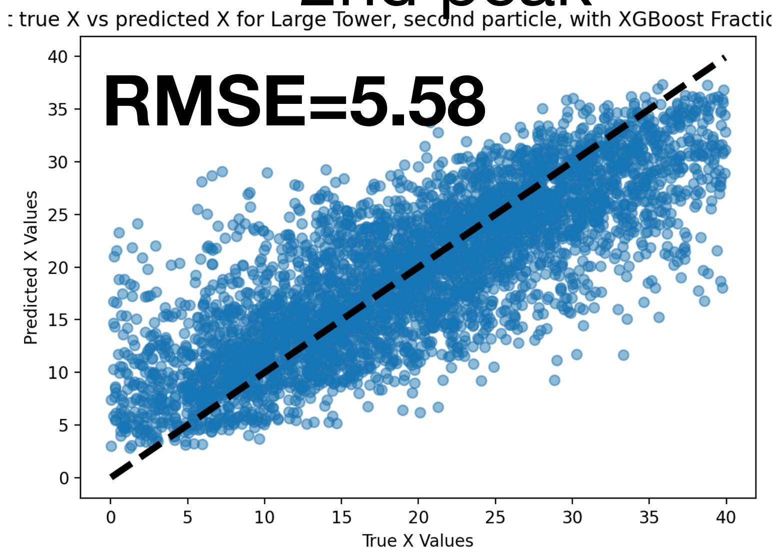
# Predict position result (xgboost)

- Named 1st, 2nd, 3rd, 4th peak in order of energy.

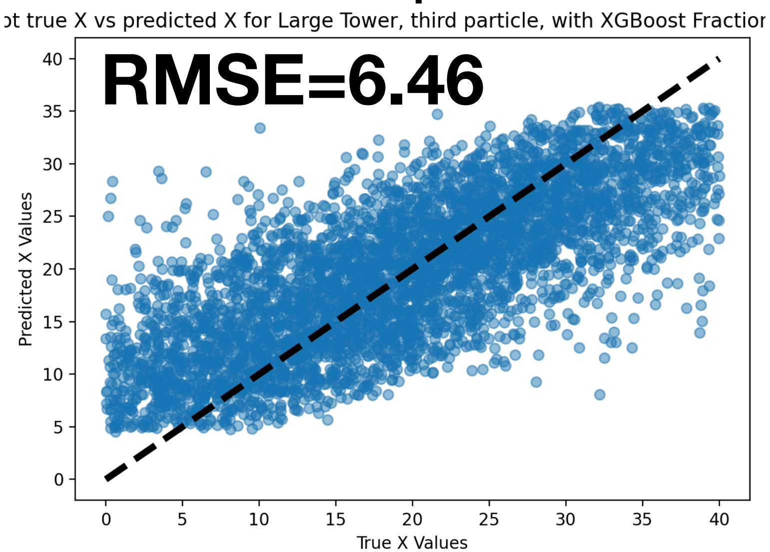
1st peak



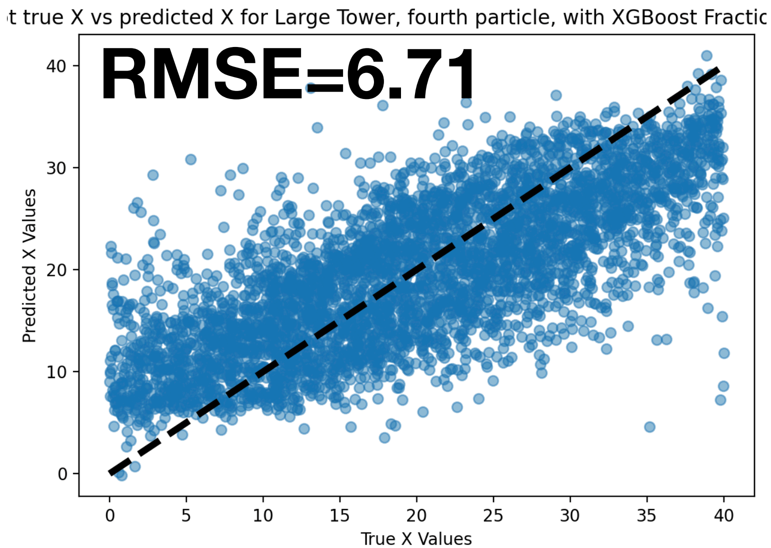

2nd peak




3rd peak



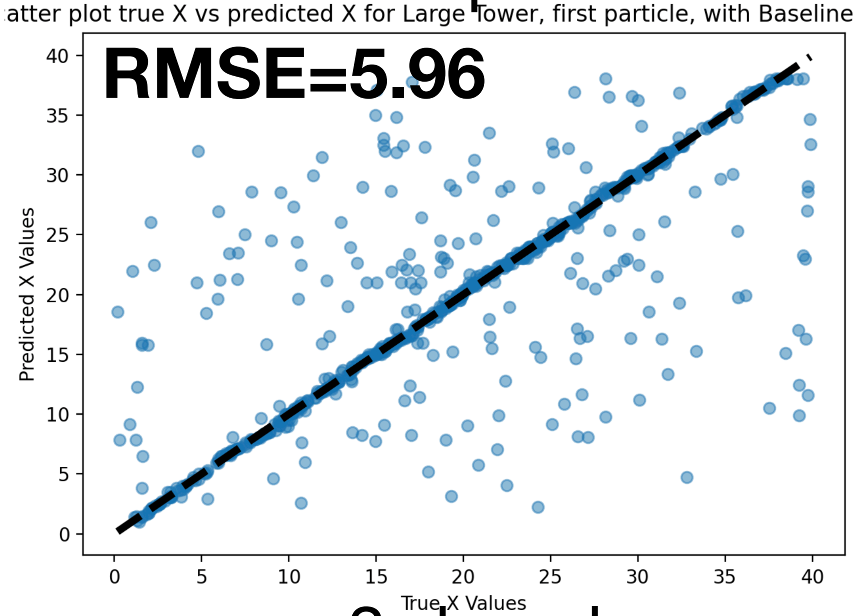

4th peak



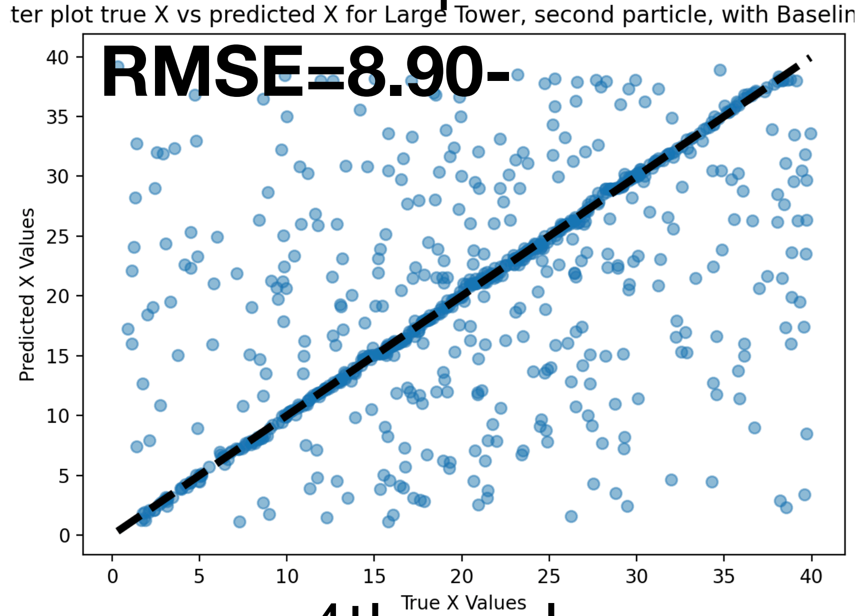

- Predicted peak position by machine learning deviates from baseline as peak height decreases

# Predict position result (reconstruction)

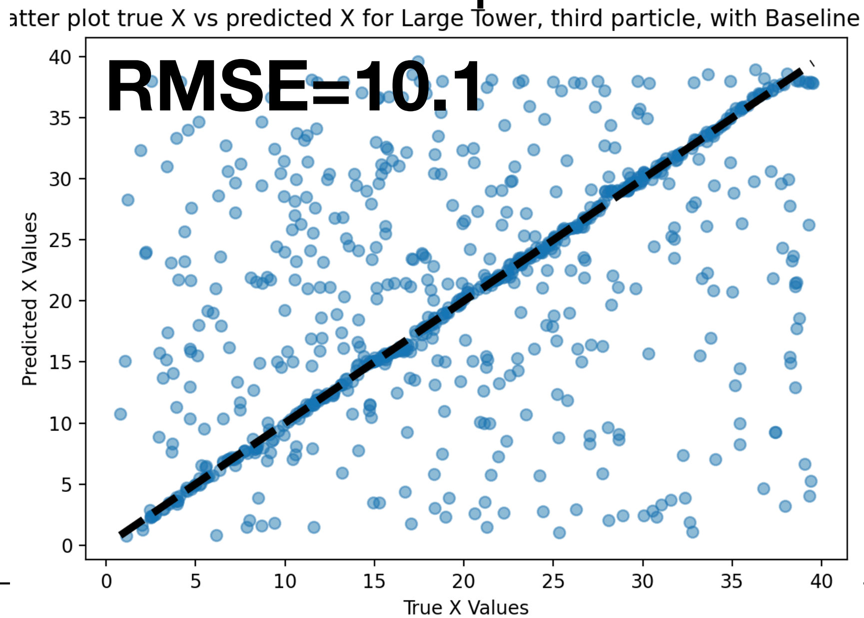
1st peak



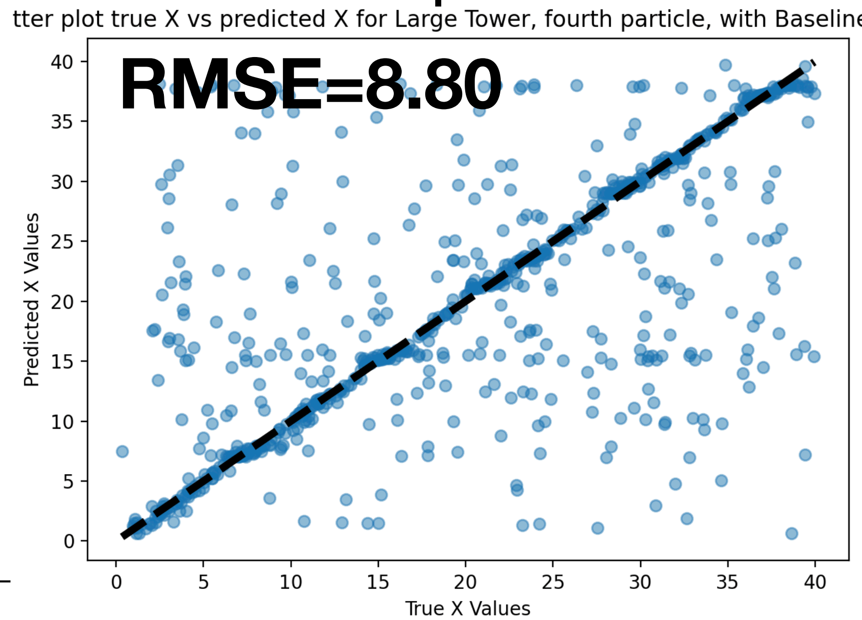

2nd peak




3rd peak




4th peak



# Conclusion

- In peak search using machine learning, the percentage of correctly judged 4-hit events was 69% even when the particle distance was 1.0 mm.
  - It should be noted that the number of events in which a 3-hit is judged as a 4-hit increased.
- Predicted peak position by machine learning deviates from baseline as peak height decreases
- I think it would be better to use machine learning for peak search algorithm and develop a method to get peak position by fitting.
  - How to remove the result deviating from baseline.
  - How to set initial parameter on Minuit2.
  - How to implement number of peak search method with machine learning in NewLibrary.

# hyper parameter of predict peak position model

1st peak

- hyper parameter
  - study rate 0.02
  - lambda 4
  - number of study 800

2nd peak

- hyper parameter
  - study rate 0.001
  - lambda 2.4
  - number of study 500

3rd peak

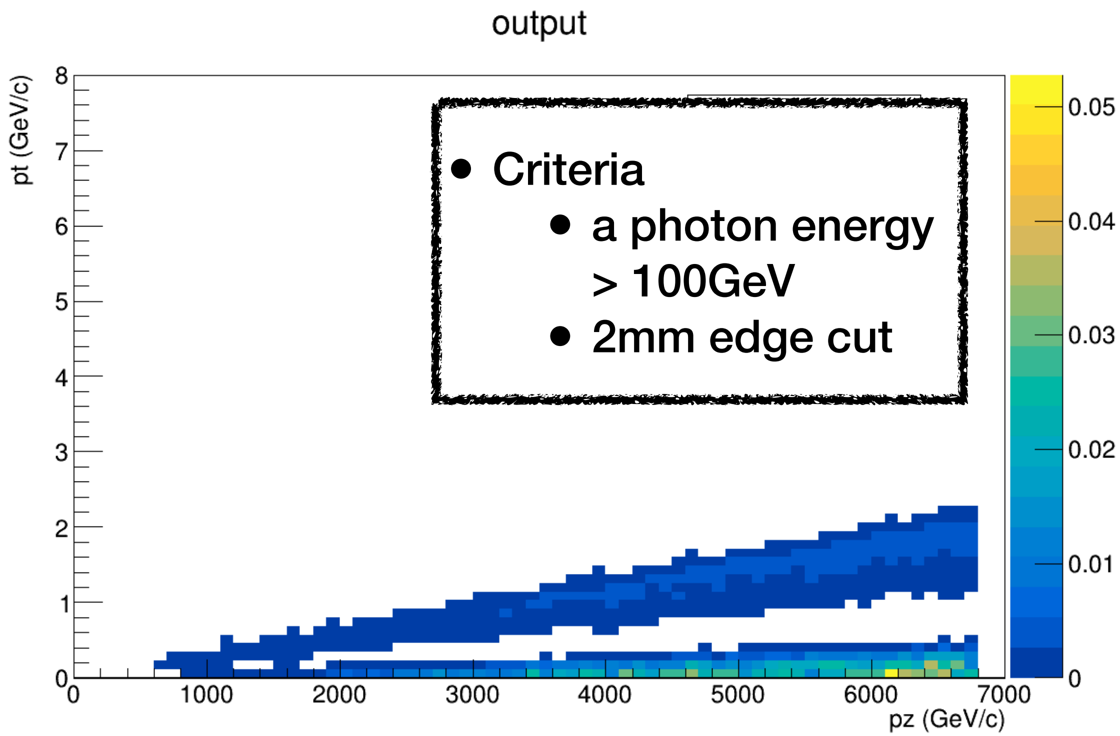
- hyper parameter
  - study rate 0.001
  - lambda 3
  - number of study 400

4th peak

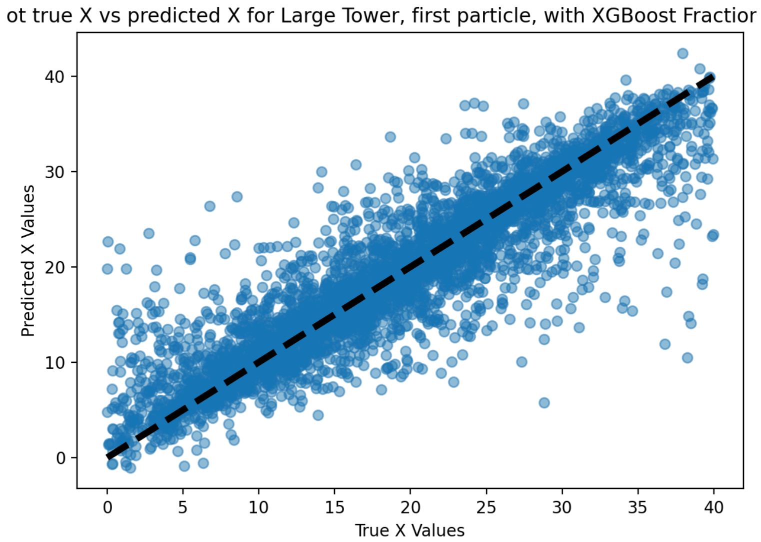
- hyper parameter
  - study rate 0.001
  - lambda 2.4
  - number of study 1000

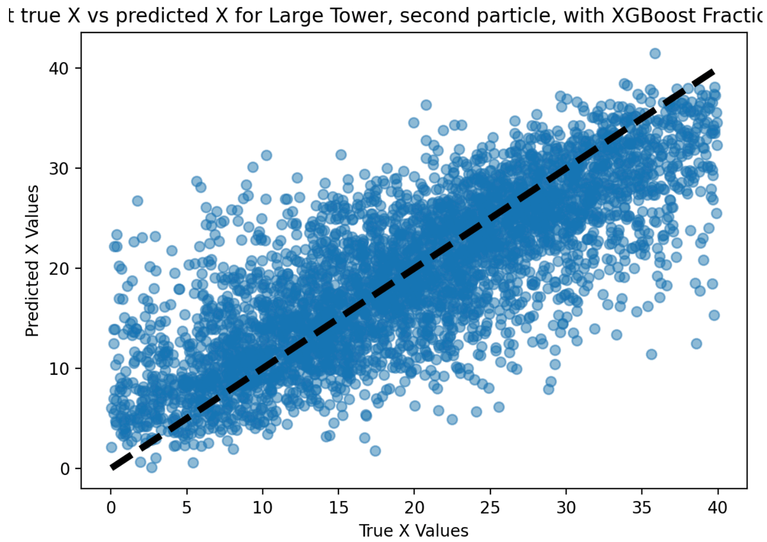
# Issue for $K_s^0$ analysis

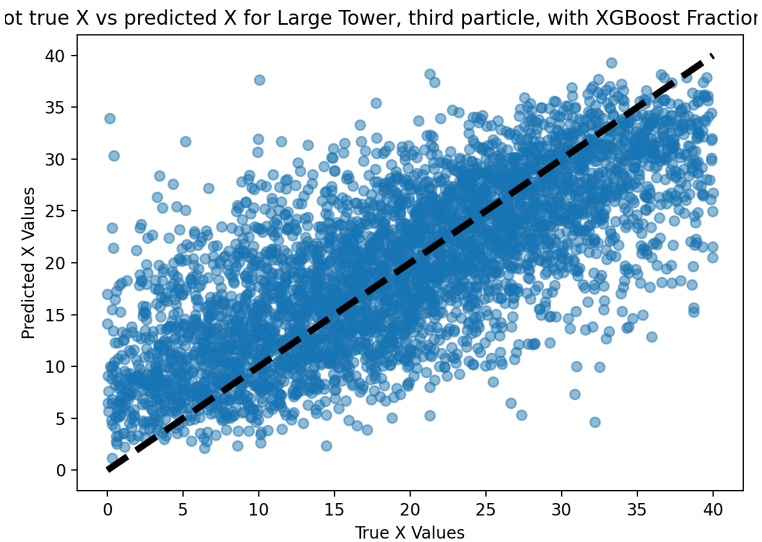
- The decay mode of $K_s^0$ measured by LHCf detector is following:
  - $K_s^0 \rightarrow 2\pi^0 \rightarrow 4\gamma$ (branching ratio:30.7%)
  - In this case, 4 photon incident on LHCf detector and event criteria is very hard. So events reconstructed of 4 photon is very small with current method.

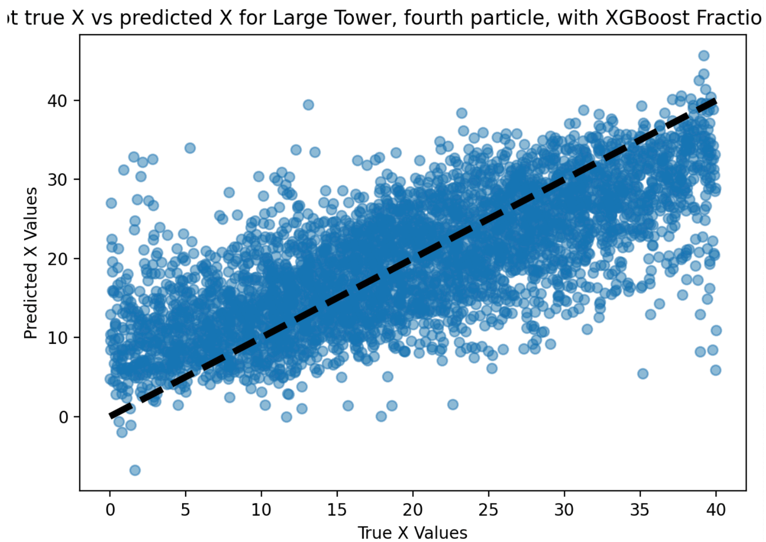


# default hyper parameter



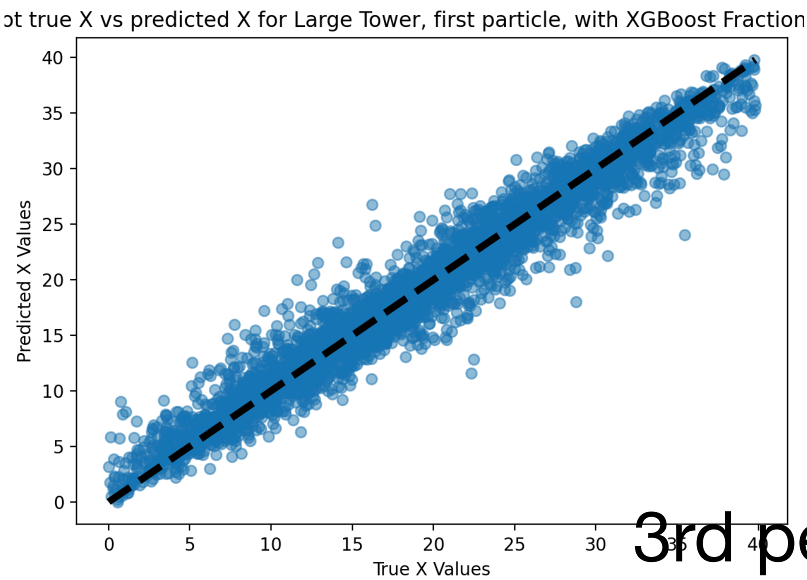



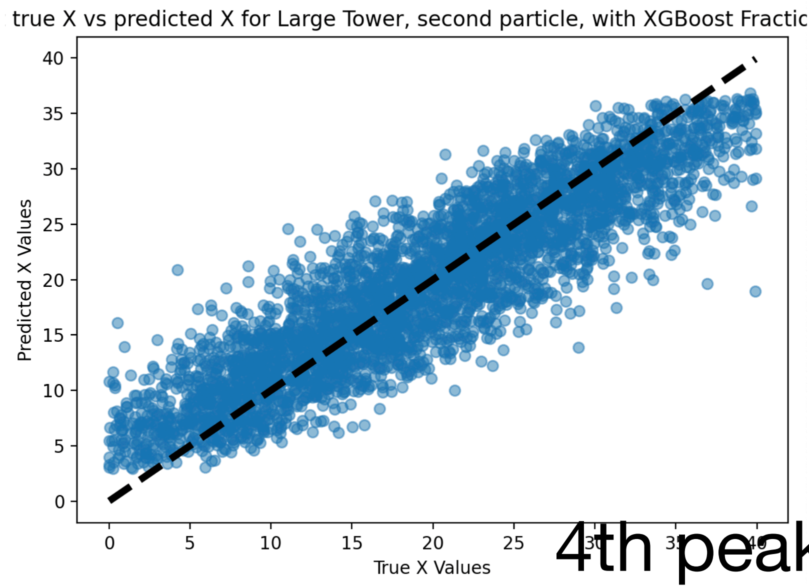



# Only available reconstruction data

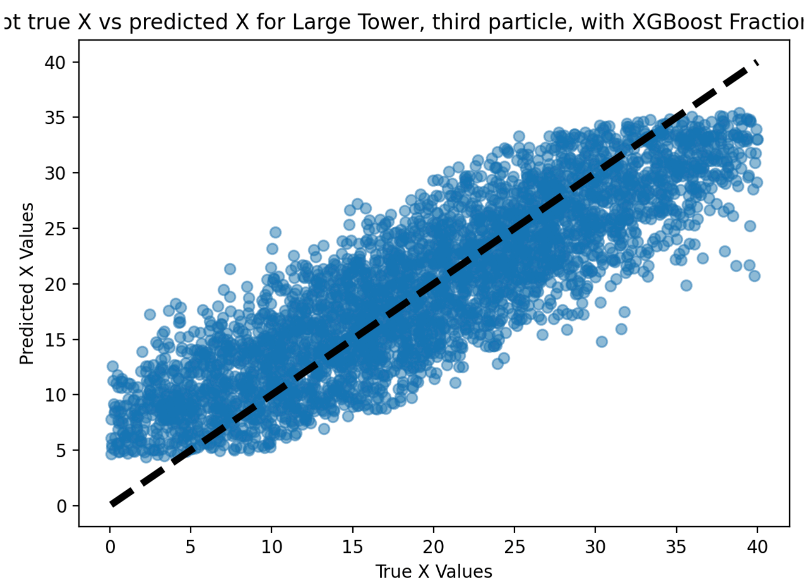
1st peak



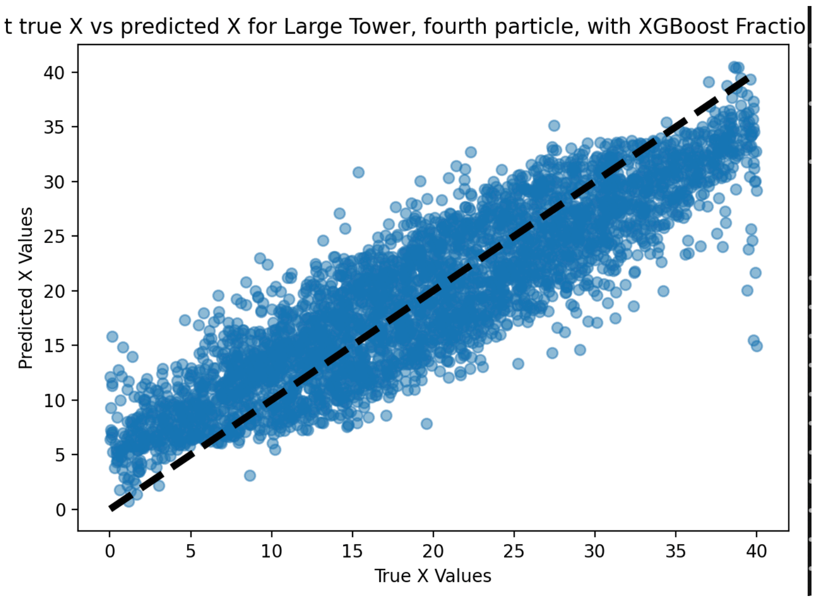

2nd peak



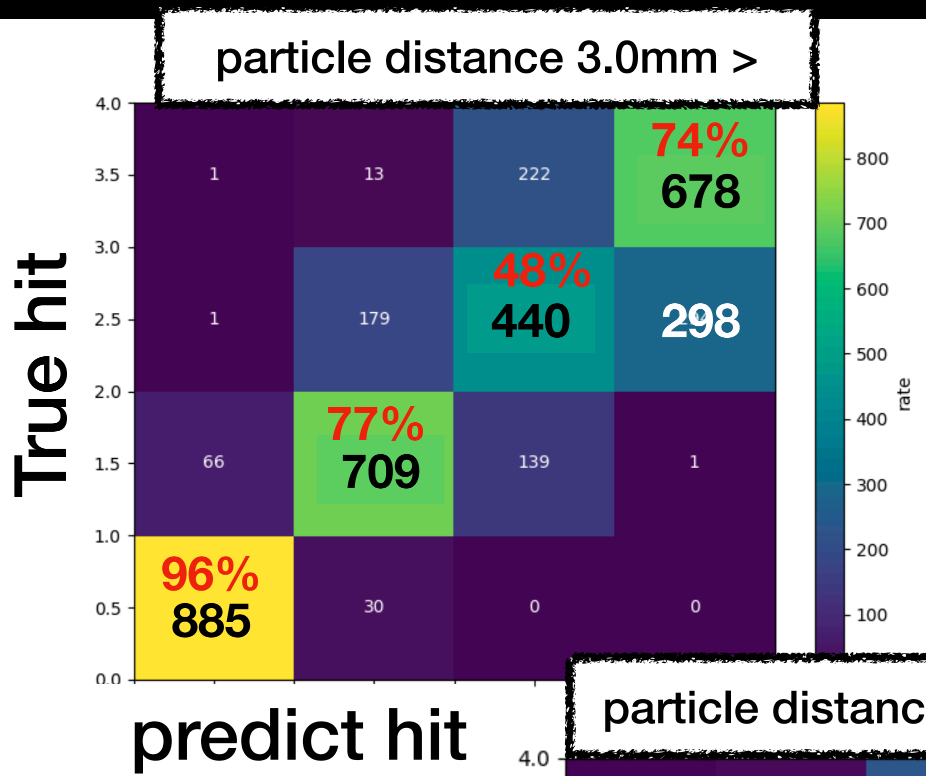

3rd peak

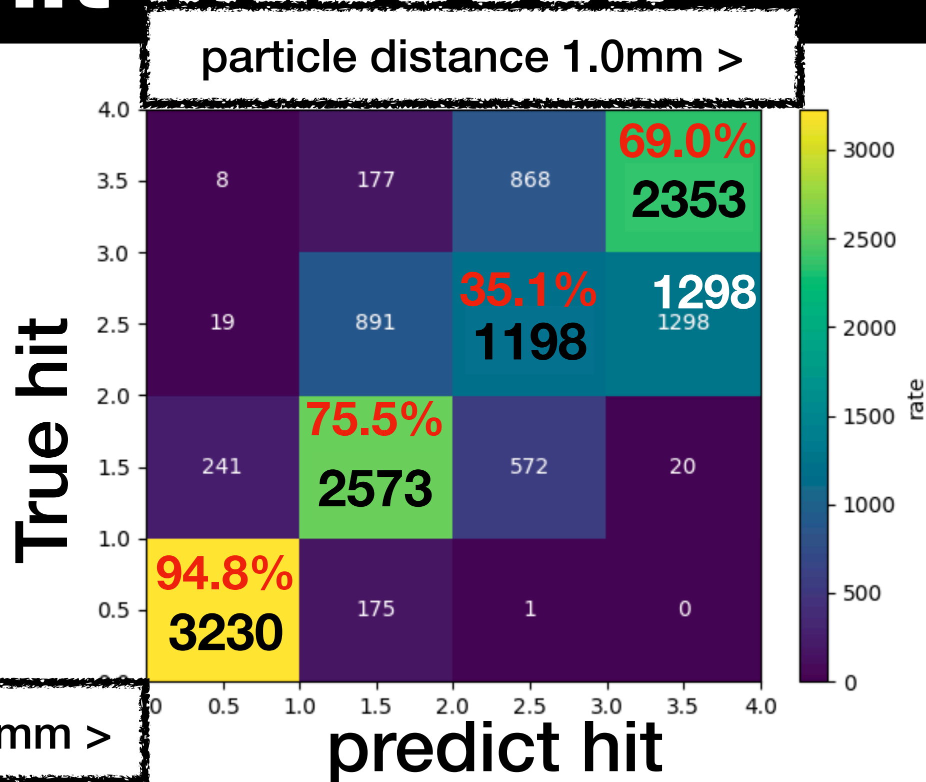


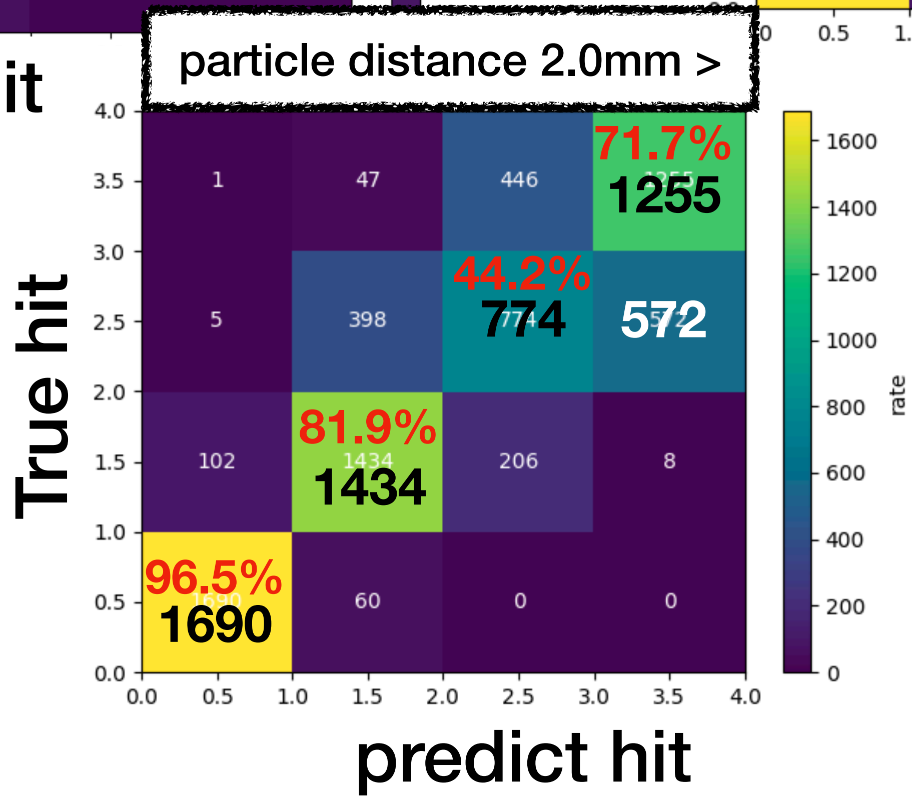

4th peak



# Predict number of hit (xgboost)




- 精度に改善が見られる
- 3hitのイベントを4hitと間違えてしまうものが目立つことには注意