



# Neutrino Oscillation

June 08, 2023    NeutrinoNet

Yuan-Yen Peng, Siang-Yuan Lin  
PHYS 591000 Hands-on Artificial Intelligence for Physics  
National Tsing Hua University (NTHU)  
Hsinchu, Taiwan

# Outline

## 1. Recap of physics

- Neutrino oscillation

## 2. Methodology and Results

- Data description
- Preprocessing and models
- Other models

## 3. Conclusion and discussion

- Summary

## 4. Backup

- Supplementary materials

# 1.1 Neutrino Oscillation

---

- Mass Eigenstates:
    - The eigenstates to the free Hamiltonian are mass eigenstates.
  - Flavor Eigenstates:
    - The eigenstates to the interaction Hamiltonian are flavor eigenstates.
  - Superposition:
    - The standard model neutrinos are theorized to be a superposition of mass eigenstates.
  - Time evolution:
- $$|\nu_i\rangle = \sum_{\alpha} U_{\alpha i}^* |\nu_{\alpha}\rangle,$$
- $$|\nu_{\alpha}\rangle = \sum_i U_{\alpha i} |\nu_i\rangle,$$
- $$|\nu_j(t)\rangle = e^{-i(E_j t - \vec{p}_j \cdot \vec{x})} |\nu_j(0)\rangle$$

- The Pontecorvo–Maki–Nakagawa–Sakata (PMNS) Matrix

$$\begin{aligned}
 U &= \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13} e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} c_{12} c_{13} & s_{12} c_{13} & s_{13} e^{-i\delta} \\ -s_{12} c_{23} - c_{12} s_{23} s_{13} e^{i\delta} & c_{12} c_{23} - s_{12} s_{23} s_{13} e^{i\delta} & s_{23} c_{13} \\ s_{12} s_{23} - c_{12} c_{23} s_{13} e^{i\delta} & -c_{12} s_{23} - s_{12} c_{23} s_{13} e^{i\delta} & c_{23} c_{13} \end{bmatrix}
 \end{aligned}$$

- The **unitary transformation** between the flavor and mass eigenstates can be determined with 4 parameters:  $\theta_{12}$ ,  $\theta_{13}$ ,  $\theta_{23}$ , and  $\delta_{CP}$ .

- General form of the probability:

$$P_{\alpha \rightarrow \beta} = \left| \langle \nu_\beta | \nu_\alpha(L) \rangle \right|^2 = \left| \sum_j U_{\alpha j}^* U_{\beta j} e^{-i \frac{m_j^2 L}{2E}} \right|^2$$

$$\begin{aligned} P_{\alpha \rightarrow \beta} &= \delta_{\alpha\beta} - 4 \sum_{j>k} \mathcal{R}_e \left\{ U_{\alpha j}^* U_{\beta j} U_{\alpha k} U_{\beta k}^* \right\} \sin^2 \left( \frac{\Delta_{jk} m^2 L}{4E} \right) \\ &\quad + 2 \sum_{j>k} \mathcal{I}_m \left\{ U_{\alpha j}^* U_{\beta j} U_{\alpha k} U_{\beta k}^* \right\} \sin \left( \frac{\Delta_{jk} m^2 L}{2E} \right) \end{aligned}$$

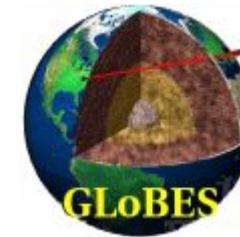
- The main probability formulas we used:

$$P_{\mu e} = \underbrace{4 s_{13}^2 s_{23}^2 \frac{\sin^2(A-1)\Delta}{(A-1)^2}}_{\mathcal{O}_o} + \underbrace{\alpha^2 \cos^2 \theta_{23} \sin^2 2\theta_{12} \frac{\sin^2 A\Delta}{A^2}}_{\mathcal{O}_2} + \underbrace{\alpha s_{13} \sin 2\theta_{12} \sin 2\theta_{23} \cos(\Delta + \delta_{cp}) \frac{\sin(A-1)\Delta}{(A-1)} \frac{\sin A\Delta}{A}}_{\mathcal{O}_1}$$

$$P_{\mu\mu} = 1 - \sin^2 2\theta_{23} \sin^2 \Delta + \mathcal{O}(\alpha, s_{13})$$

## 2.1 Data description

Data source: [GLoBES](#) simulation data.



Event rate

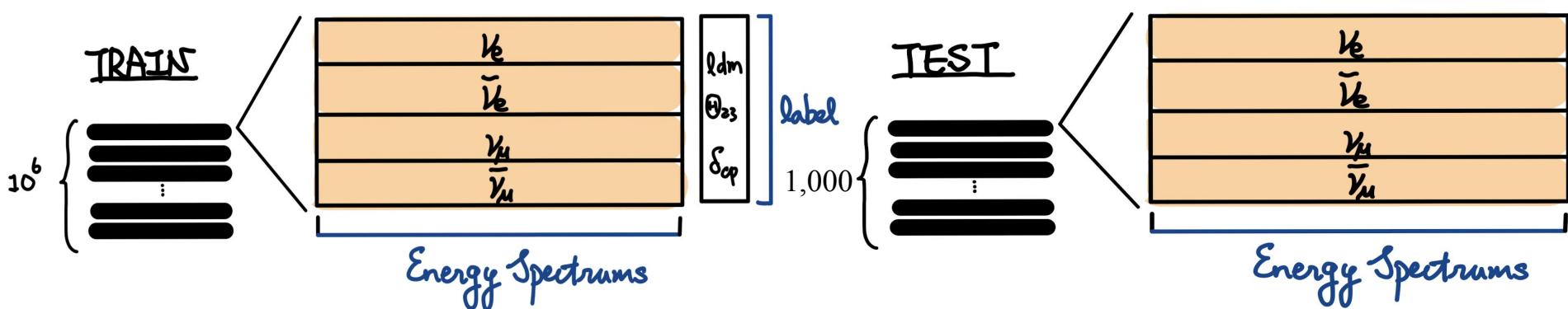
Neutrino flux

Cross section

Detector smearing

Oscillation probability

$$R(\vec{x}) = \Phi(E_\nu) \times \sigma(E_\nu, \vec{x}) \times \epsilon(\vec{x}) \times P(\nu_A \rightarrow \nu_B)$$

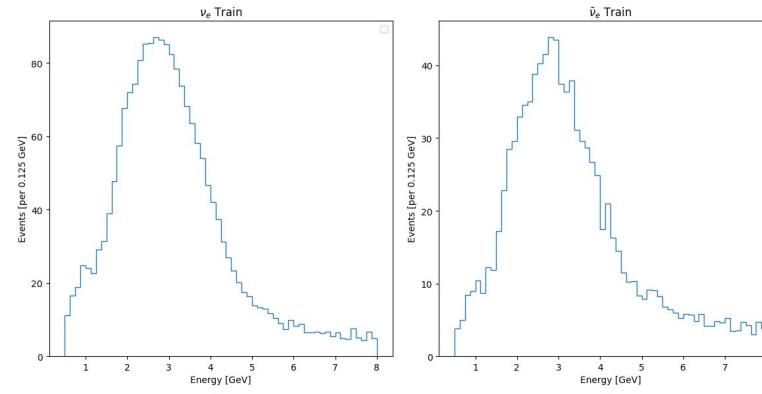


In this project we have 4 channels (2 appearances and 2 disappearances), and 3 labels ( $ldm$ ,  $\theta_{23}$ ,  $\delta_{CP}$ ).

- Four channels

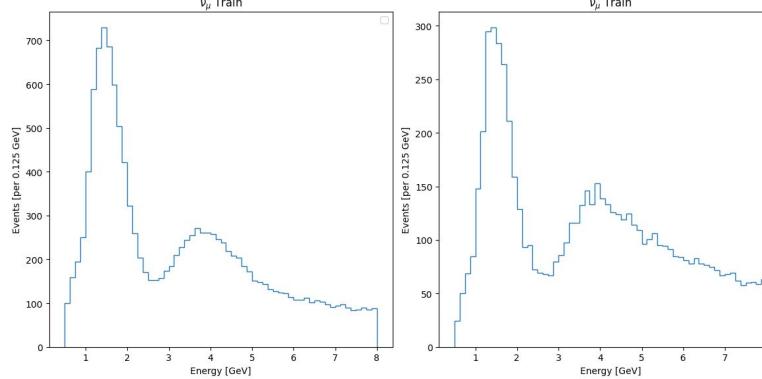
$$\nu_\mu \rightarrow \nu_e$$

$$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$$



$$\nu_\mu \rightarrow \nu_\mu$$

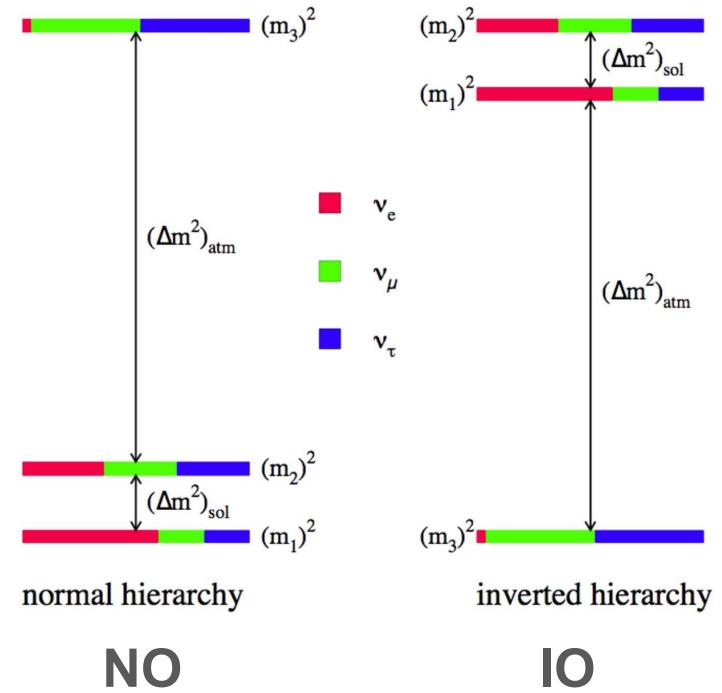
$$\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$$

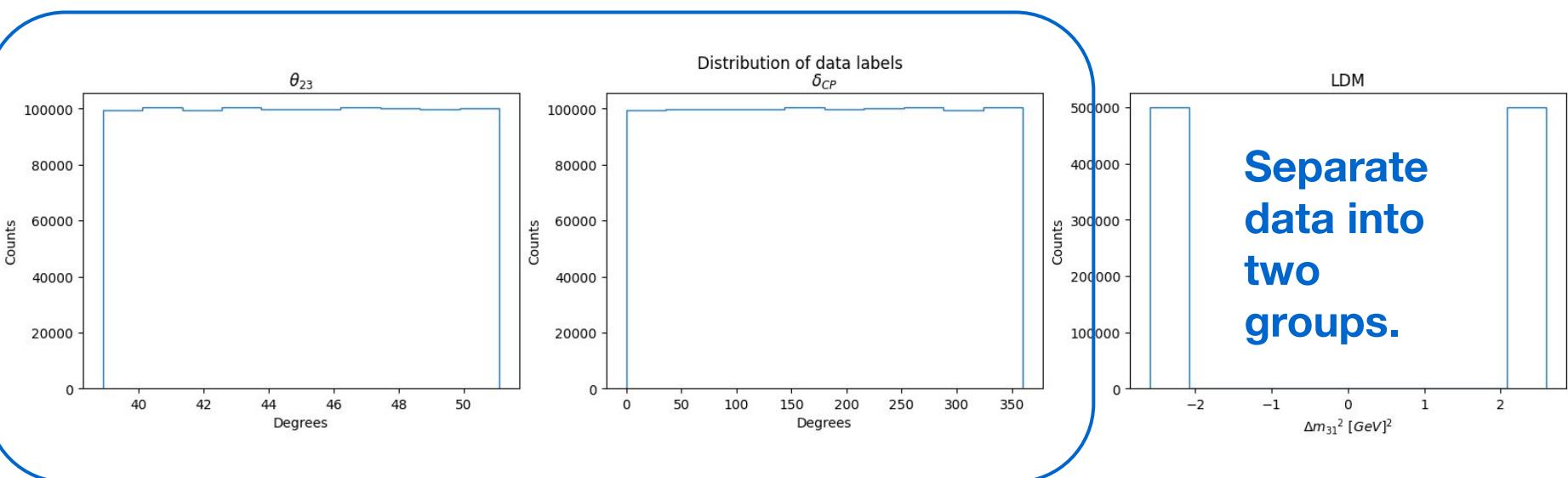


**Appearance channels**

**Disappearance channels**

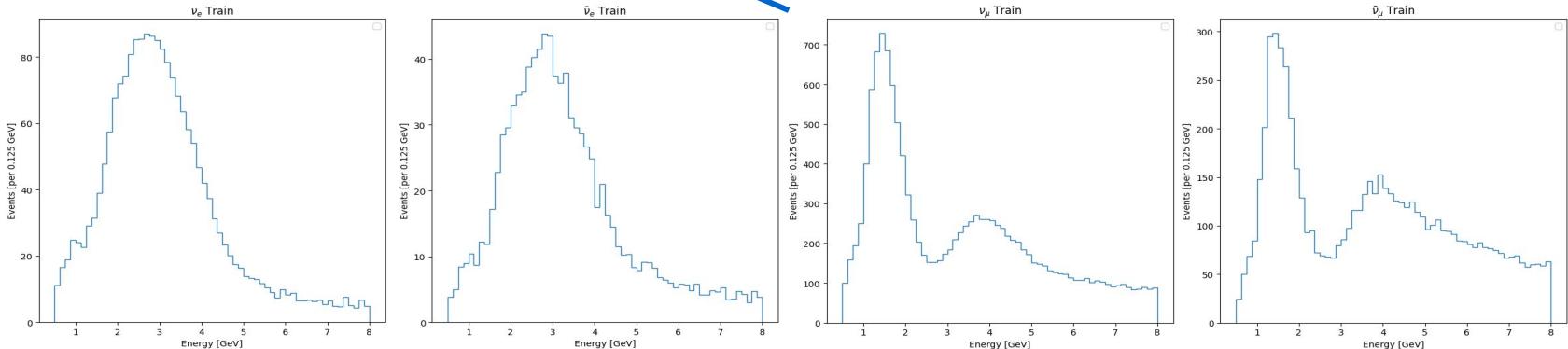
- Mixing angle  $\theta_{23}$ , and CP phase  $\delta_{CP}$   
 $\theta_{23}$  ranges from approximately 39 to 51 degrees,  $\delta_{CP}$  ranges from 0 to 360 degrees.
- Mass hierarchy  $ldm$   
We are only able to know the difference and its orderings.  
 $ldm$  only has two significant peaks, which tells us that we can split our train data into NO and IO two parts.





**Separate data into two groups.**

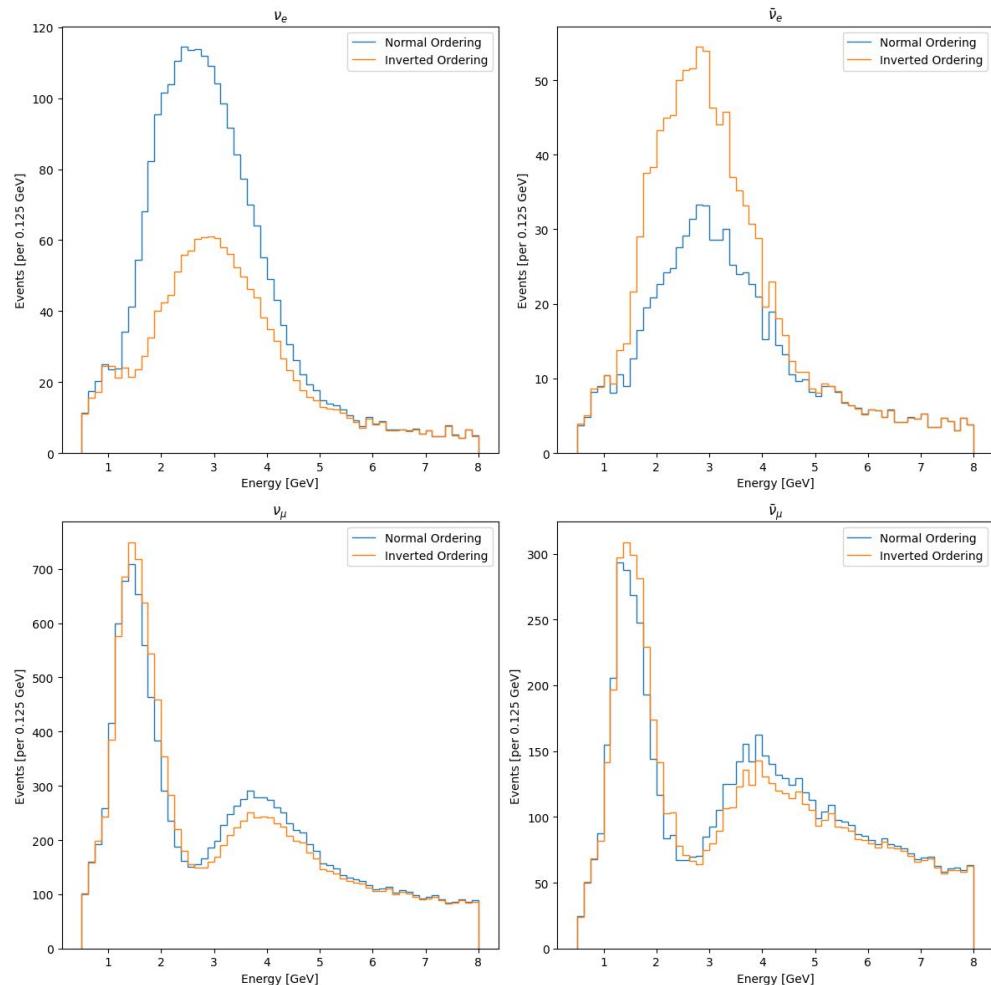
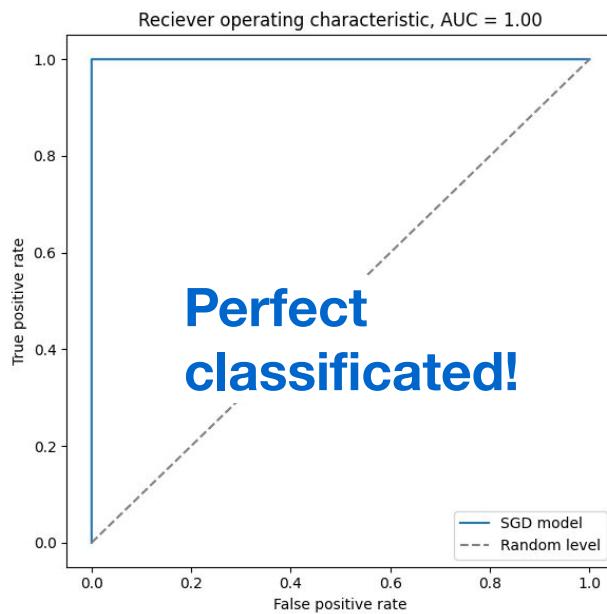
{ Tasked to find the correlation between **2 ANGLES** and **4 SPECTRUMS**. }



# 2.2 Preprocessing and models

## I. LDM classification

- Distinct difference in these distributions.
- SVM with SGD  
→ binary classification.



averaged spectrum of the NO and IO



## II. DCNN regression $\{\theta_{23}\}$

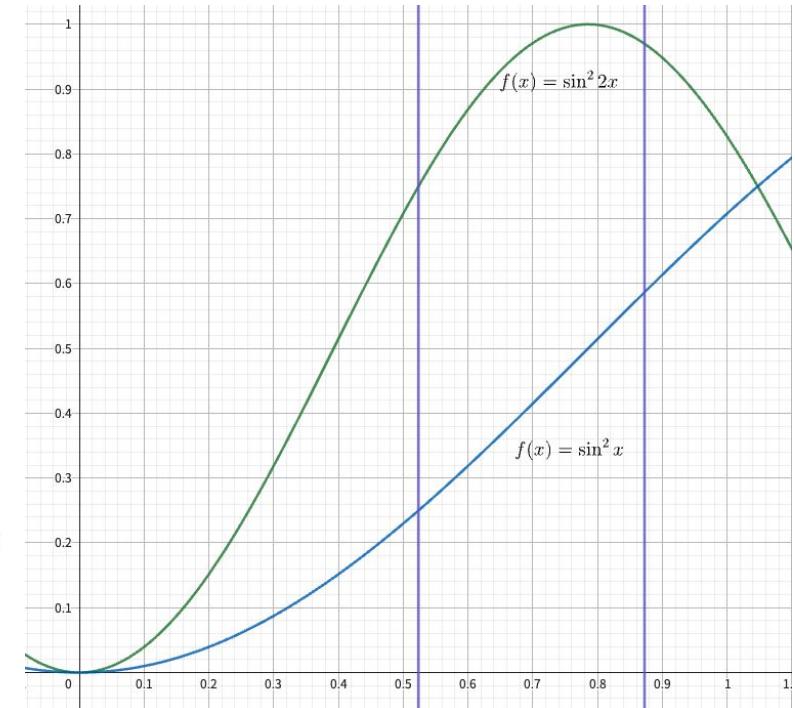
Theoretically,  $\theta_{23}$  in (with non-trivial leading order):

- **appearance channel:**  $\sim \sin^2 \theta_{23}$
- **disappearance channel:**  $\sim \sin^2 2\theta_{23}$

$$P_{\mu e} = \underbrace{4s_{13}^2 s_{23}^2 \frac{\sin^2(A-1)\Delta}{(A-1)^2}}_{\mathcal{O}_0} + \underbrace{\alpha^2 \cos^2 \theta_{23} \sin^2 2\theta_{12} \frac{\sin^2 A\Delta}{A^2}}_{\mathcal{O}_2}$$

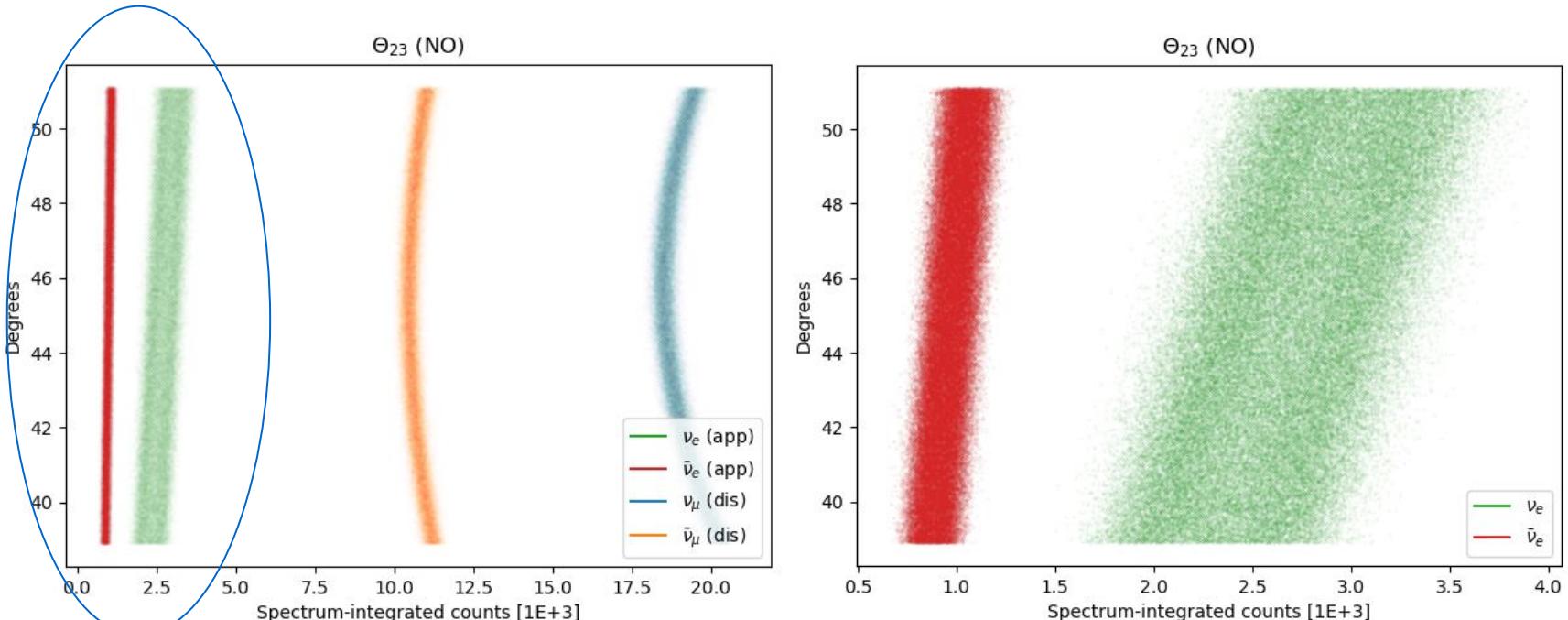
$$+ \underbrace{\alpha s_{13} \sin 2\theta_{12} \sin 2\theta_{23} \cos(\Delta + \delta_{cp}) \frac{\sin(A-1)\Delta}{(A-1)} \frac{\sin A\Delta}{A}}_{\mathcal{O}_1}$$

$$P_{\mu\mu} = 1 - \underbrace{\sin^2 2\theta_{23} \sin^2 \Delta}_{\text{circled term}} + \mathcal{O}(\alpha, s_{13})$$

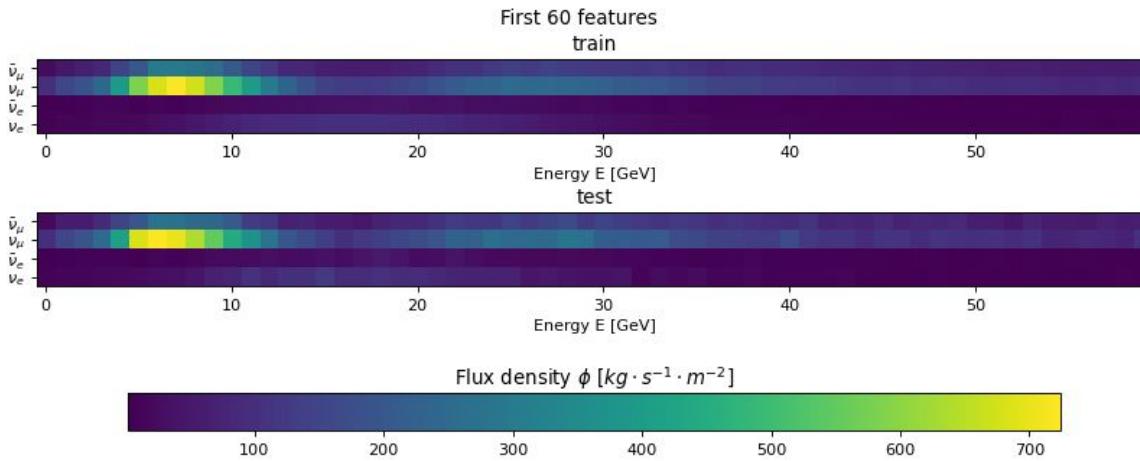


In this interval (30 ~ 50 degrees), the probability can be determined, due to the monotonicity of the function (for appearance channels).

# Expect good predictability in $\theta_{23}$ !



No degeneracy in the appearance channels.

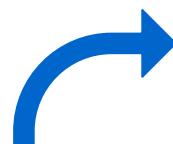


**IN: MinMax  
Normalized  
CNN images  
+label**



$\theta_{23}$

**OUT: MinMax  
Transpose**



**DCNN**

[Conv2D(64, 4)  
+BatchNorm  
+LeakyReLU(0.2)]\*2  
+AvgPooling2D  
{all 2D with same pad}

[Conv2D(32, 4)  
+BatchNorm  
+LeakyReLU(0.2)]\*2  
+AvgPooling2D  
{all 2D with same pad}

**Loss: huber**

Flatten()  
+Dropout(0.2)

dense(64)  
+dense(16)  
+dense(8)  
+dense(1)

[Conv2D(16, 4)  
+BatchNorm  
+LeakyReLU(0.2)]\*2  
+AvgPooling2D  
{all 2D with same pad}



## III-A. CNN regression $\{\delta_{CP}\}$

Theoretically,  $\delta_{CP}$  relates to:

- appearance:

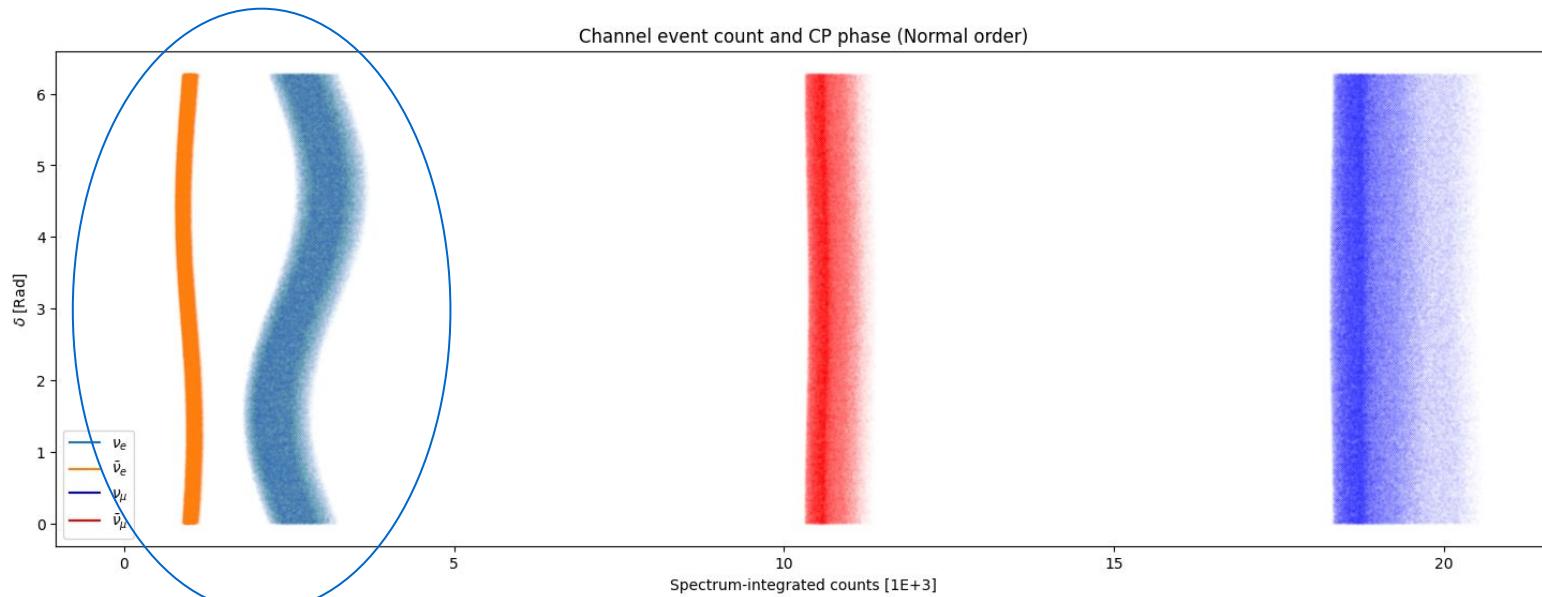
$$\sim \cos(\Delta + \delta_{CP}) \sim \sin \delta_{CP}$$

In the vacuum oscillation maxima,  $\Delta \sim 90^\circ$

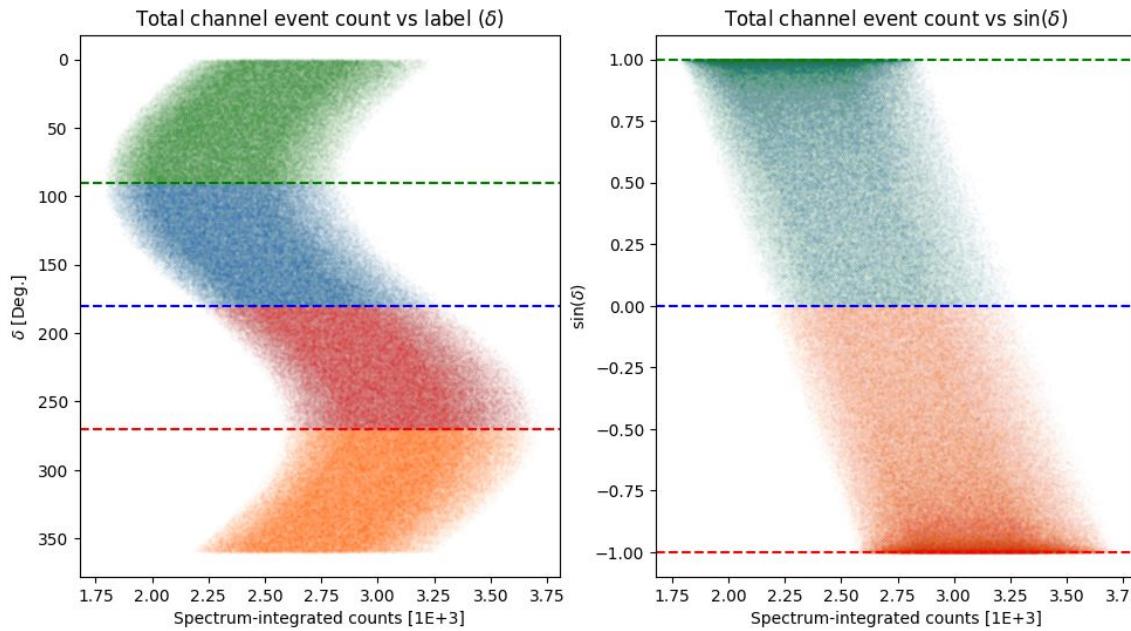
$$P_{\mu e} = \underbrace{4s_{13}^2 s_{23}^2 \frac{\sin^2(A-1)\Delta}{(A-1)^2}}_{\mathcal{O}_o} + \underbrace{\alpha^2 \cos^2 \theta_{23} \sin^2 2\theta_{12} \frac{\sin^2 A\Delta}{A^2}}_{\mathcal{O}_2} \\ + \underbrace{\alpha s_{13} \sin 2\theta_{12} \sin 2\theta_{23} \cos(\Delta + \delta_{cp})}_{\mathcal{O}_1} \frac{\sin(A-1)\Delta}{(A-1)} \frac{\sin A\Delta}{A}$$

$$P_{\mu\mu} = 1 - \sin^2 2\theta_{23} \sin^2 \Delta + \mathcal{O}(\alpha, s_{13})$$

- disappearance: X



N. Nath, M. Ghosh, and S. Goswami, *The Physics of Antineutrinos in DUNE and Determination of Octant and  $\delta_{CP}$* , Nuclear Physics B **913**, 381 (2016).



IN: MinMax  
Normalized  
CNN images



OUT: sine  
values

$\sin \delta_{CP}$

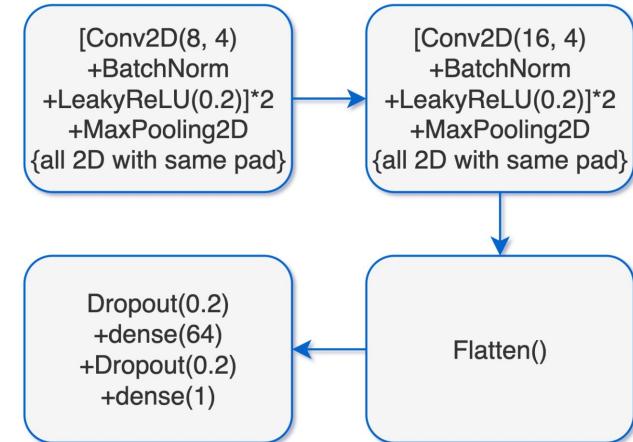


Reduce the degeneracy!

- Use sine and cosine with 2 transpose arc functions.
- Use sine and cosine with tangent and one transpose arc function.
- **Use one sine and one arc; then, do the classification.**

Reg

Loss: huber

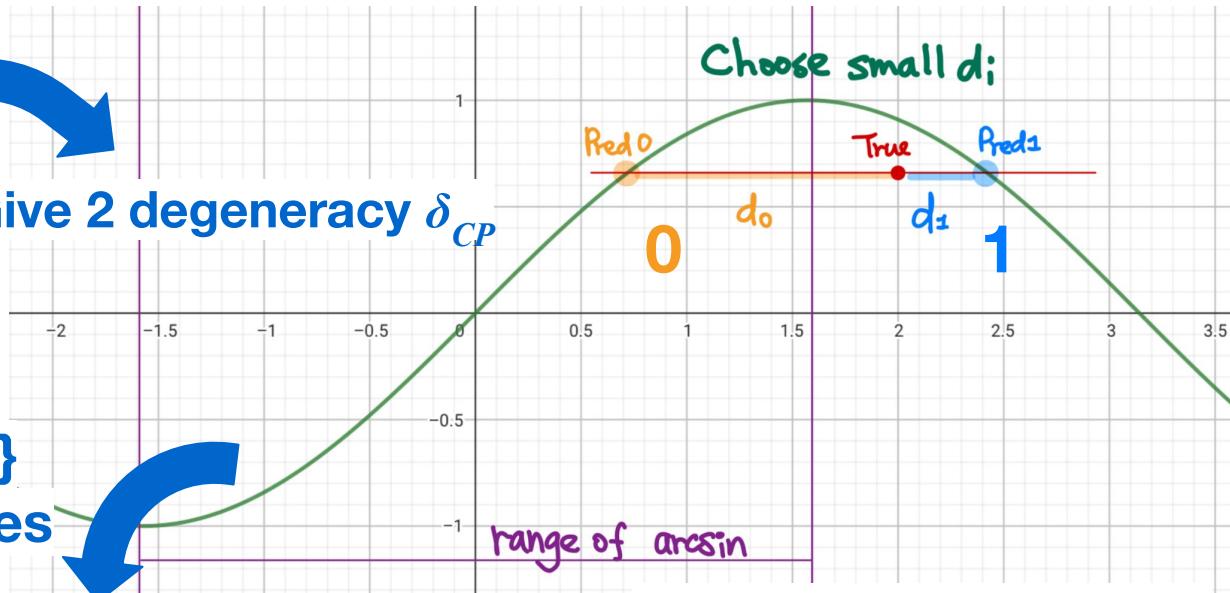


### III-B. CNN classification $\{\delta_{CP}\}$

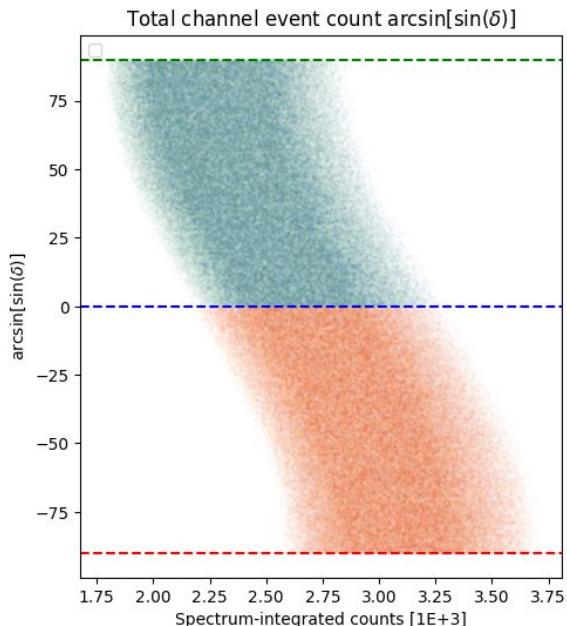
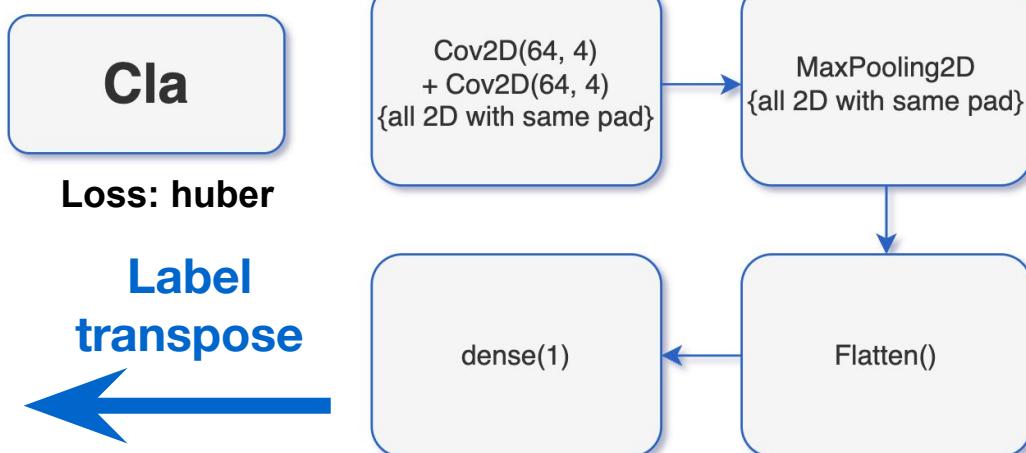
#### TRAIN procedure

$\sin^{-1}(\sin \delta_{cp})$

Give 2 degeneracy  $\delta_{CP}$



Label array with {0,1}  
+MinMax CNN images



# TEST procedure

$\sin^{-1}(\sin \delta_{cp})$

Give 2 degeneracy  $\delta_{CP}$

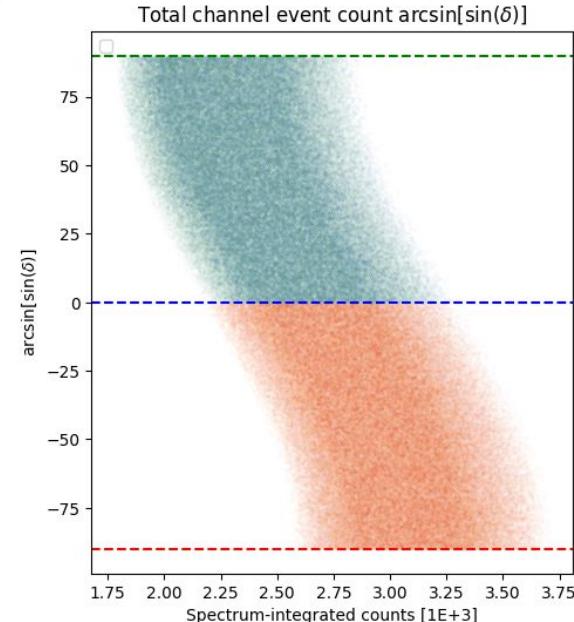
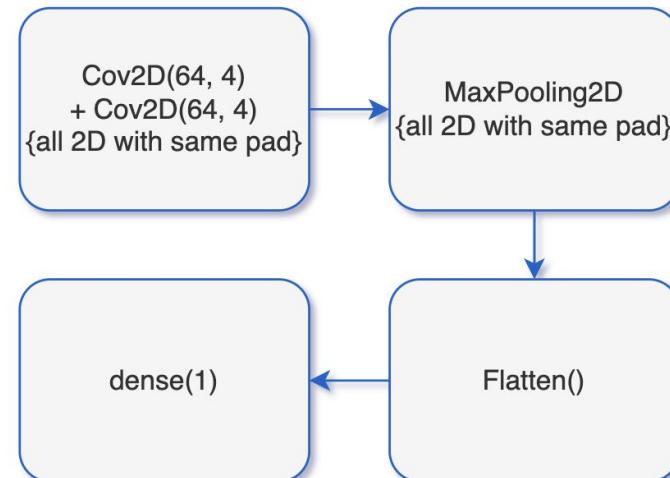
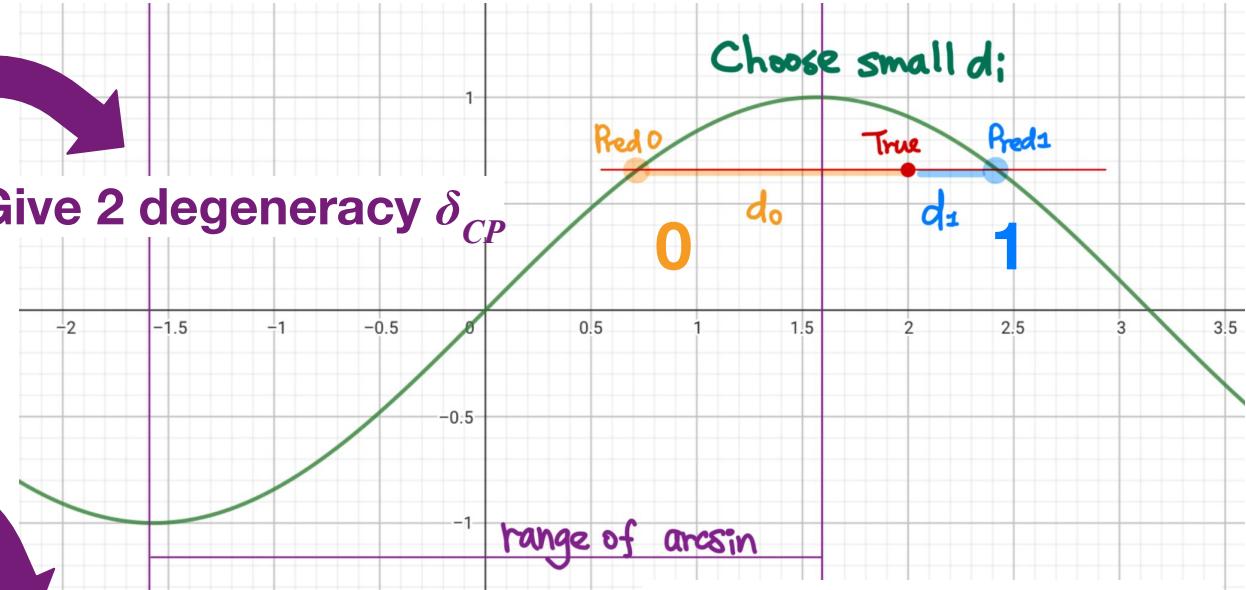
MinMax  
Normalized  
CNN images

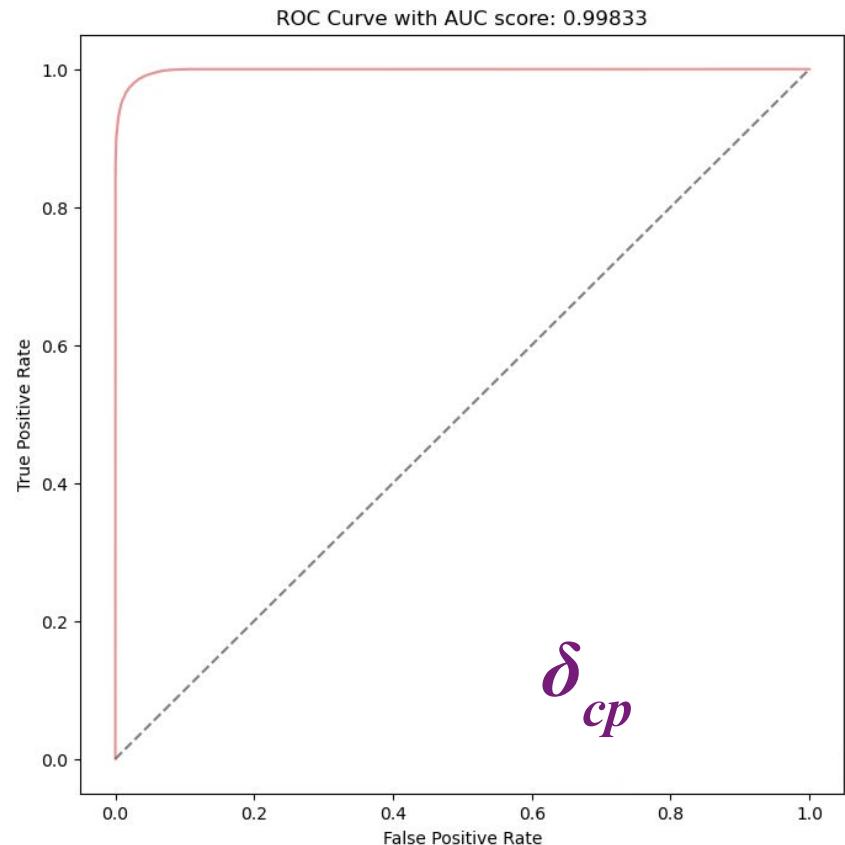
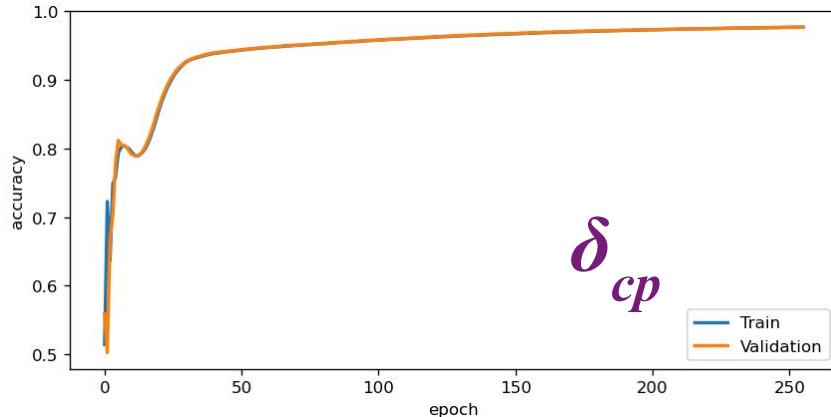
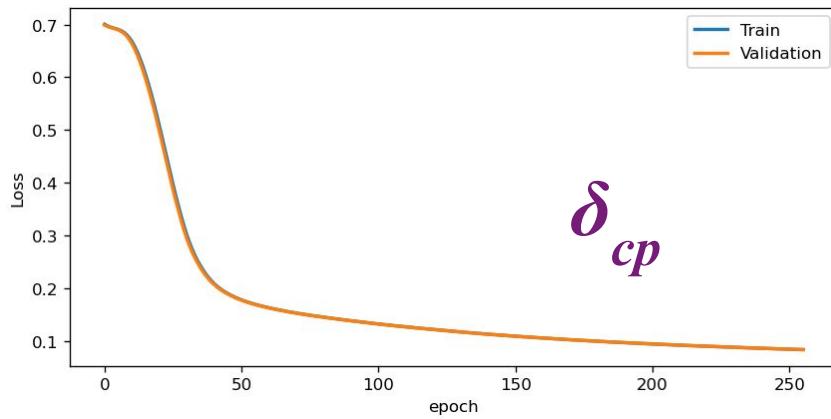
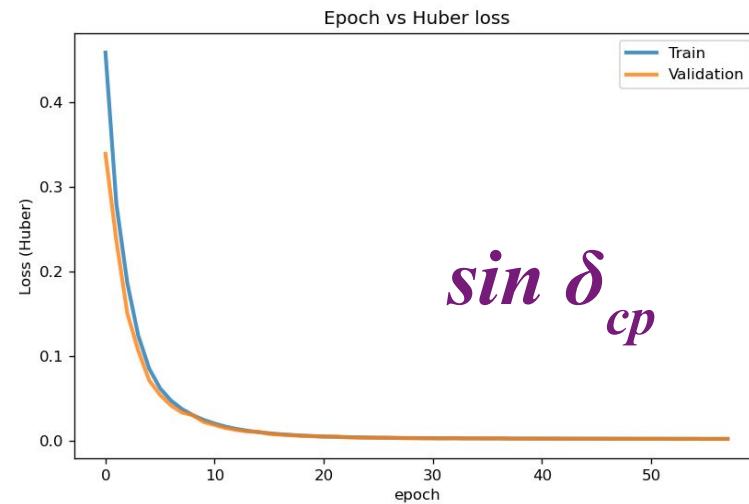
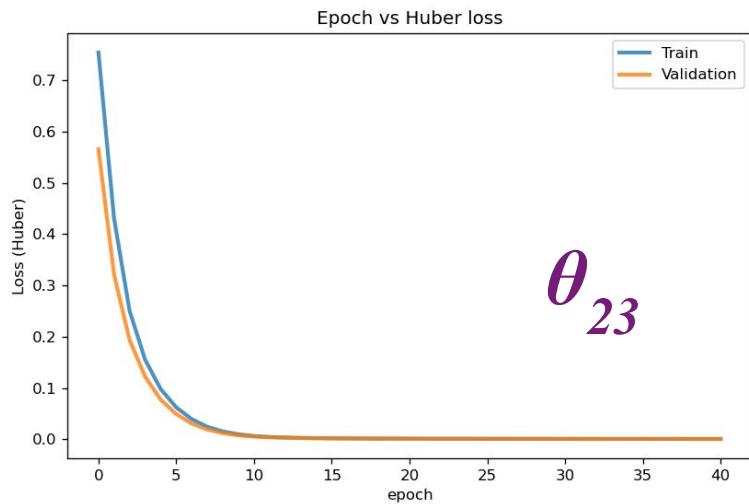
Cla

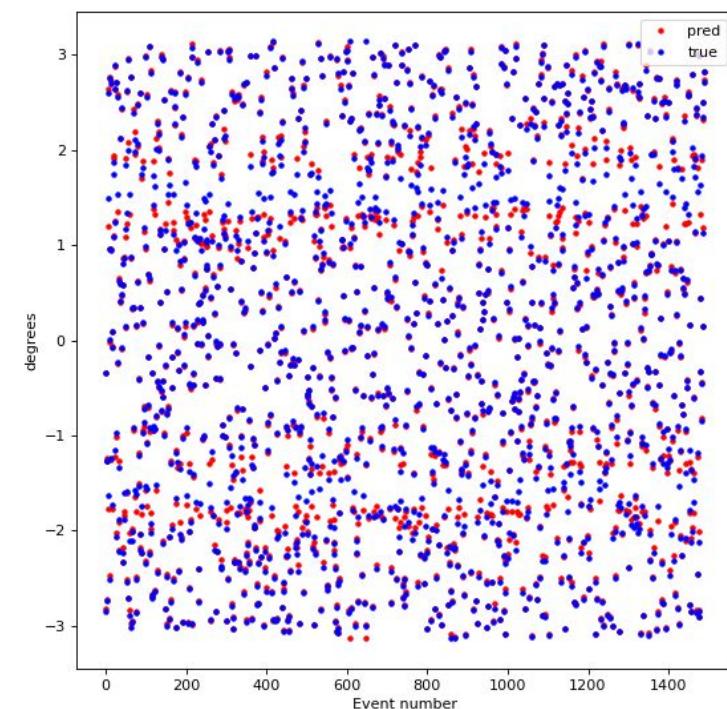
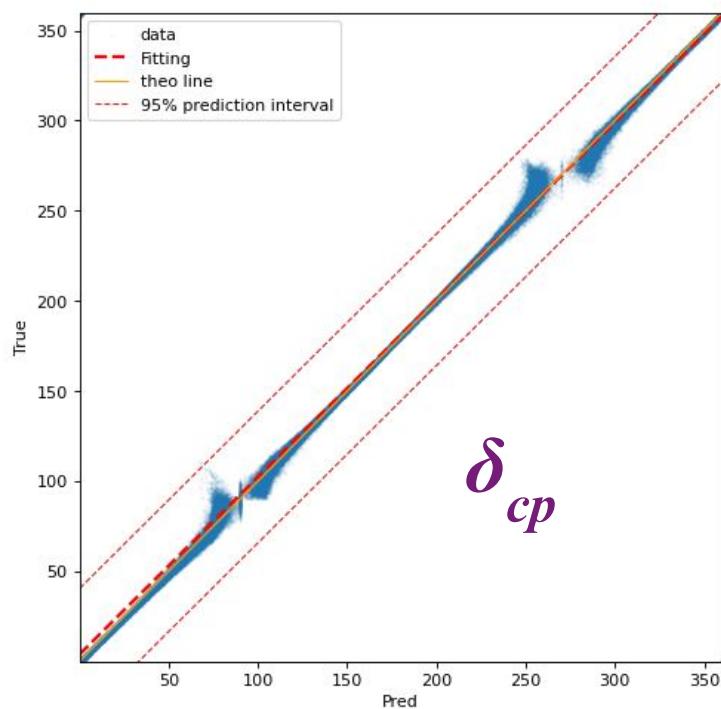
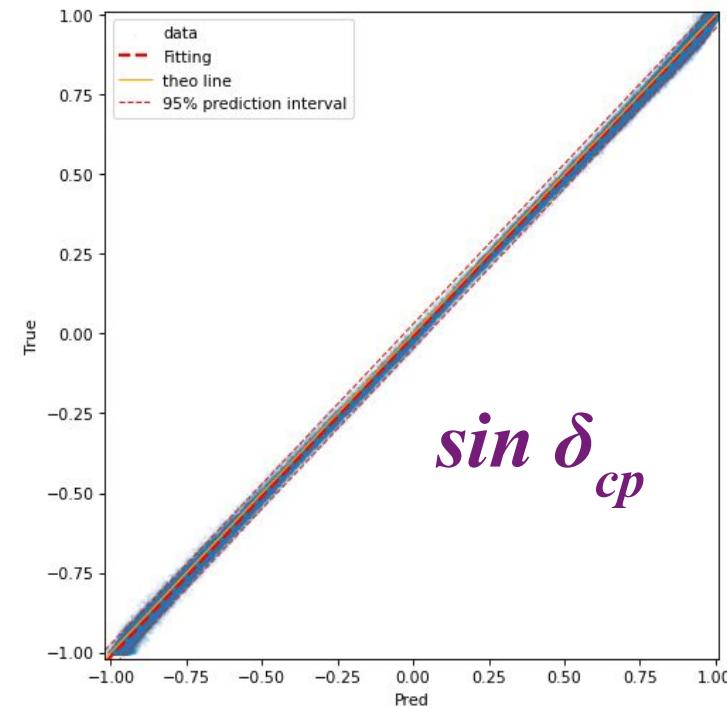
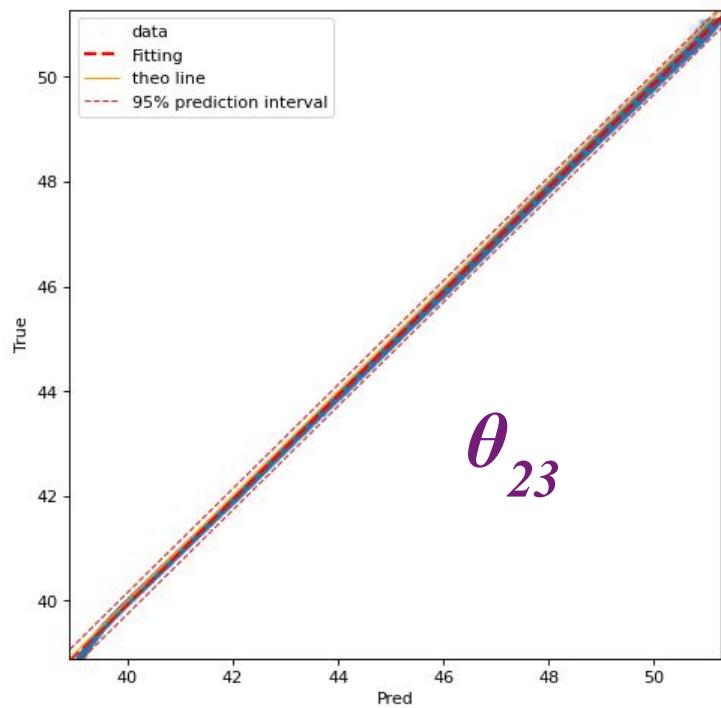
Loss: huber

$\delta_C$   
Label transpose

P






 $\delta_{cp}$

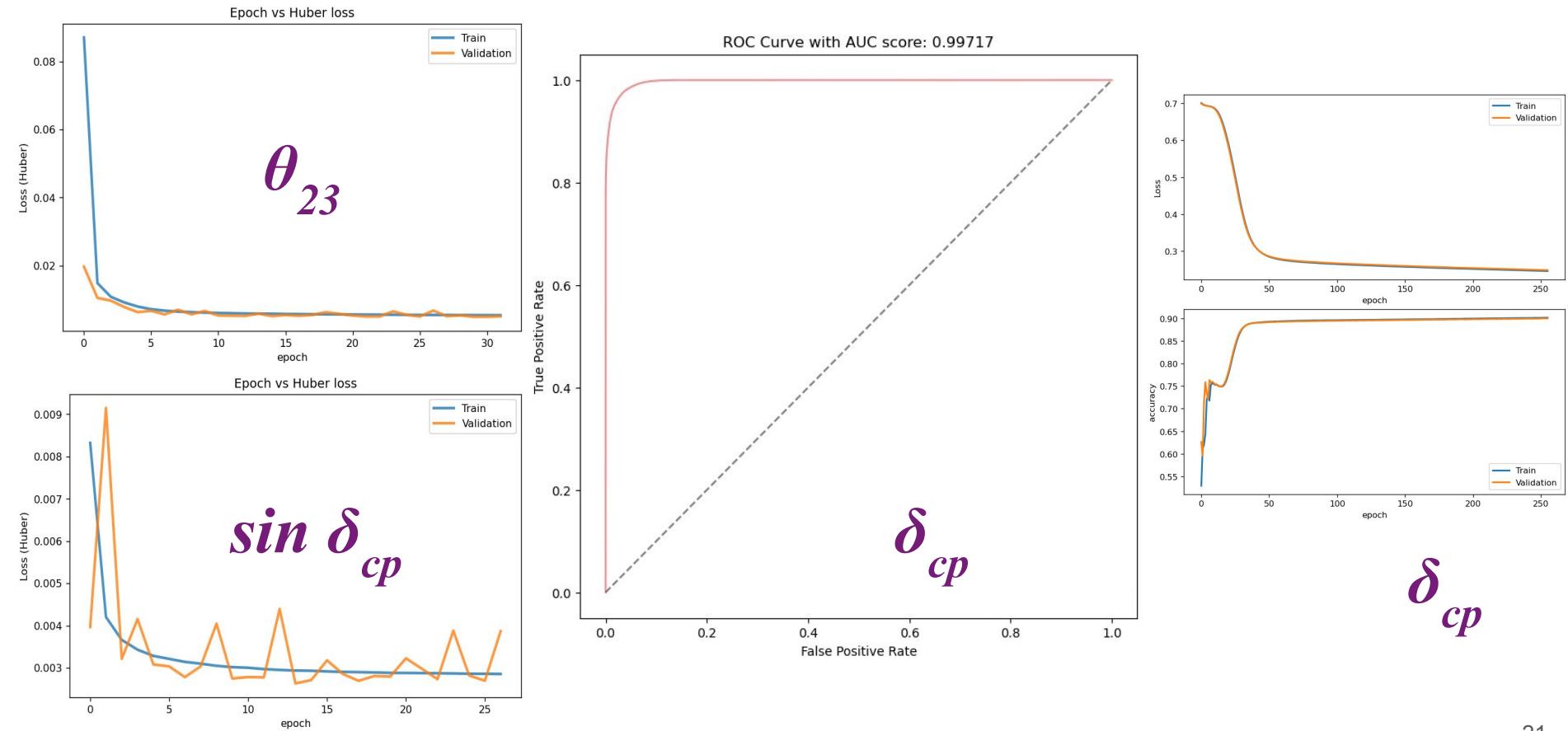
## 2.3 Other models

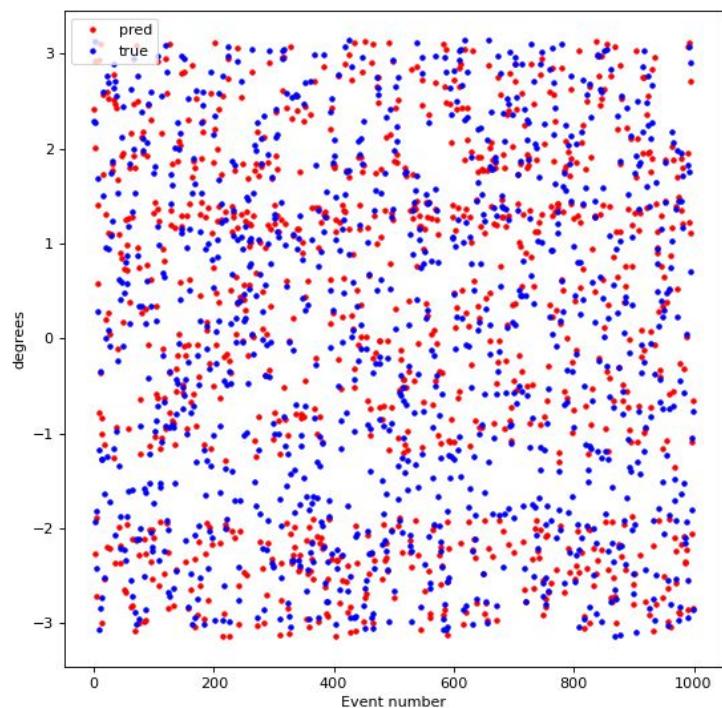
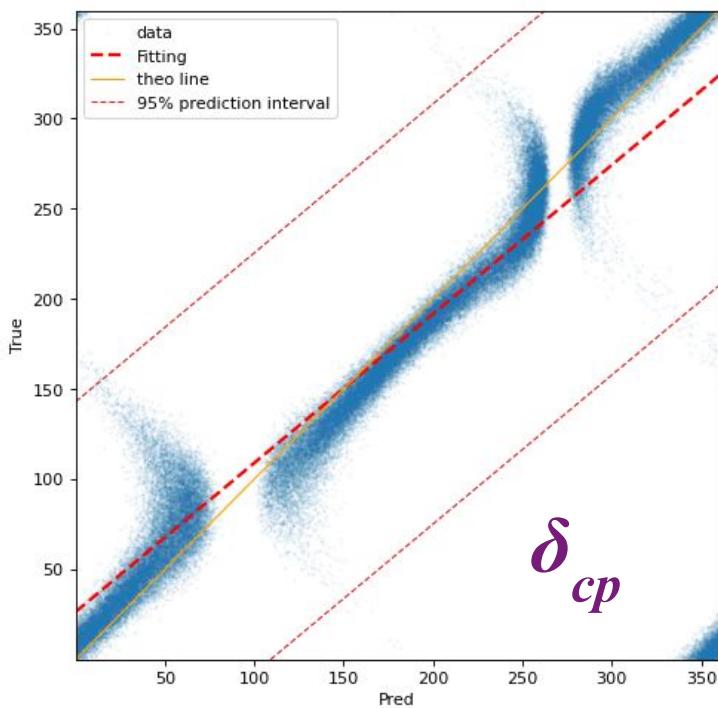
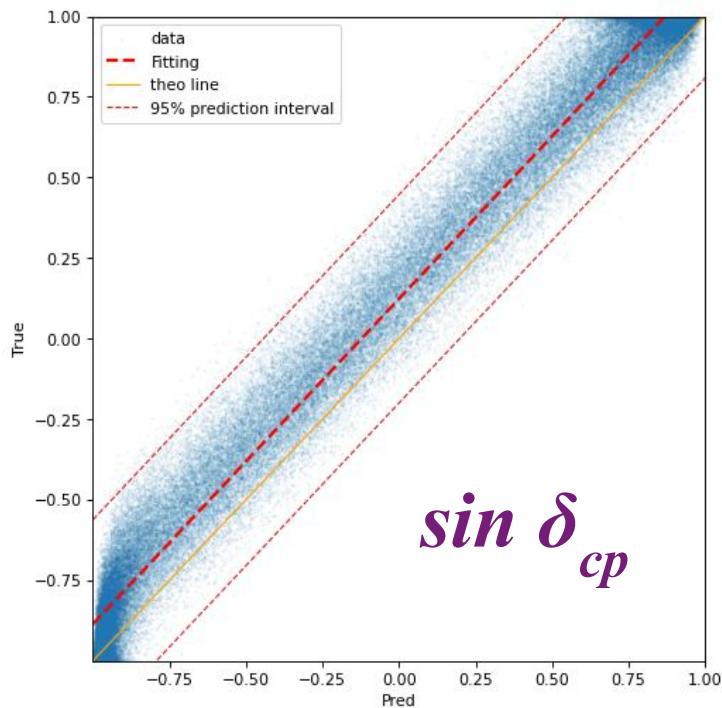
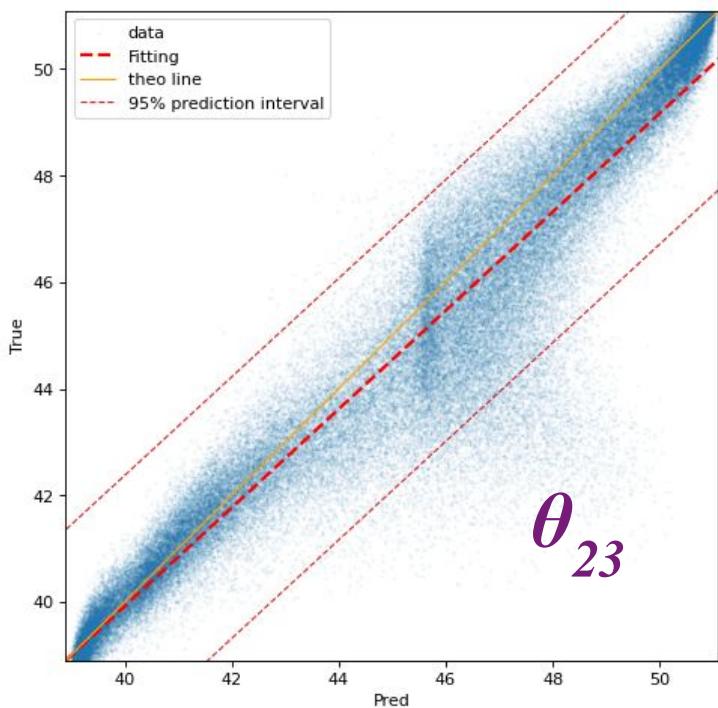
---

- Owing to the overfit with the test sample (base on public and private score), we add **noise** into the training datasets.  
Implementing similar models and methods.  
[improve a little bit]
- Due to the arcsine classification distortion, we try to use the models reported in the [last presentation](#).  
⇒ Classificate  $\delta_{cp}$  training data into 4 groups (each group is monotonic); then, do the regression. (**reversed procedure** of the main  $\delta_{cp}$  model structure)  
[do not have obvious improvement]

## 2.3.A Noisy samples

Owing to the overfit with the test sample (base on public and private score), we add noise into the training datasets. Implementing similar models and methods.





$\delta_{cp}$

## 2.3.B Reversed $\delta_{cp}$ procedure

---

Due to the arcsine classification distortion, we try to use the models reported in the [last presentation](#).

⇒ Classificate  $\delta_{cp}$  training data into 4 groups (each group is monocity); then, do the regression. (reversed procedure of the main model structure)

<LOSS  
CURVE>

時間太長 QQ 放結果就好 <RESULT>

# 3.1 Summary

---

- Physically, we recap some essential concepts of neutrino oscillation.
- For models, we first use the Idm binary model to split data into IO and NO two groups; after that, separating the regression of  $\theta_{23}$  and  $\delta_{cp}$  into several models.  $\theta_{23}$ : regression;  $\delta_{cp}$ : sine regression + arcsine classification.
- The results fit too well to tolerate the Poisson fluctuations in the test sample.
- Apply additional models to rule out the overfitting problems, such as the noisy model and the reversed  $\delta_{cp}$  procedure model.
- Learn from the other group: implement AE to denoise; predict parameters at the same time.

# BACK UP

# Codes

General backup repo in GitHub:

<https://github.com/yygarypeng/Neutrino-Oscillation-DUNE>

Submission backup in Kaggle:

<https://www.kaggle.com/competitions/phys591000-2023-final-project-i>

# The physics of antineutrinos in DUNE and determination of octant and $\delta_{CP}$

$\cos(\pi/2 + \delta) = -\sin(\delta)$ , so this formula is also proportional to  $\sin(\delta)$  if in the vacuum ideal situations. (ref page 386, 387)

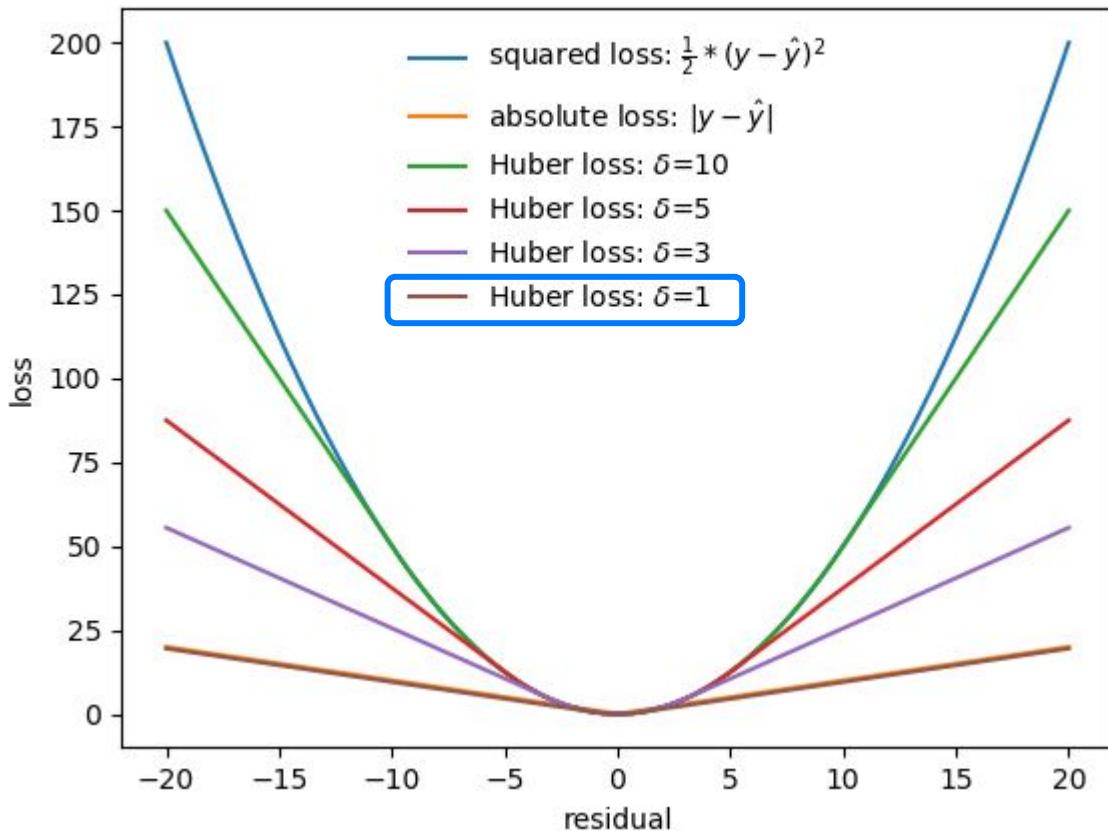
$$P_{\mu e} = \underbrace{4s_{13}^2 s_{23}^2 \frac{\sin^2(A-1)\Delta}{(A-1)^2}}_{\mathcal{O}_o} + \underbrace{\alpha^2 \cos^2 \theta_{23} \sin^2 2\theta_{12} \frac{\sin^2 A\Delta}{A^2}}_{\mathcal{O}_2} + \underbrace{\alpha s_{13} \sin 2\theta_{12} \sin 2\theta_{23} \cos(\Delta + \delta_{cp}) \frac{\sin(A-1)\Delta}{(A-1)} \frac{\sin A\Delta}{A}}_{\mathcal{O}_1}$$

$$P_{\mu\mu} = 1 - \sin^2 2\theta_{23} \sin^2 \Delta + \mathcal{O}(\alpha, s_{13})$$

Note that for vacuum oscillation maxima,  $\Delta$  corresponds to  $90^\circ$ . Thus in the appearance channel probability (cf. Eq. (2)),  $\delta_{CP} = -90^\circ (+90^\circ)$  correspond to maximum (minimum) point in the probability for neutrinos. For antineutrinos it is the opposite. Thus, for these values of  $\delta_{CP}$ , octant sensitivity is expected to be maximum if there is no degeneracy. Note that with the inclusion of matter effect, the appearance channel probability maxima does not coincide with the vacuum maxima and in that case the maximum and minimum points in the probability do not come exactly at  $\pm 90^\circ$  but gets slightly shifted. This can be seen from Fig. 1. However for illustration, we will take  $\delta_{CP} = \pm 90^\circ$  as the reference points to describe the physics of octant in DUNE.

# Huber loss

$$Loss(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2, & |y - \hat{y}| \leq \delta \\ \delta(|y - \hat{y}| - \frac{1}{2}\delta), & o.w. \end{cases}$$



Huber = MSE + MAE

MSE is more sensitive to outliers.  
⇒ improve the

**robustness** of the MSE  
to outliers using MAE.

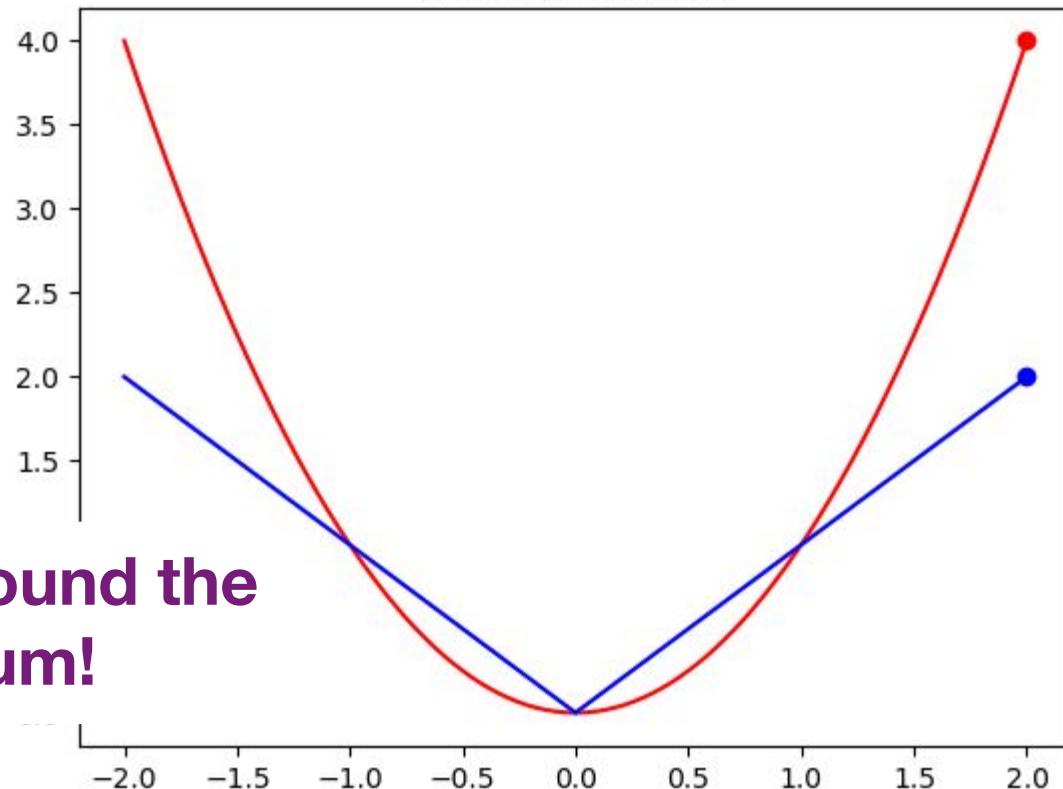
However, MAE is not  
stable around the  
minimum.

Huber loss

## Loss functions

### MAE vs MSE

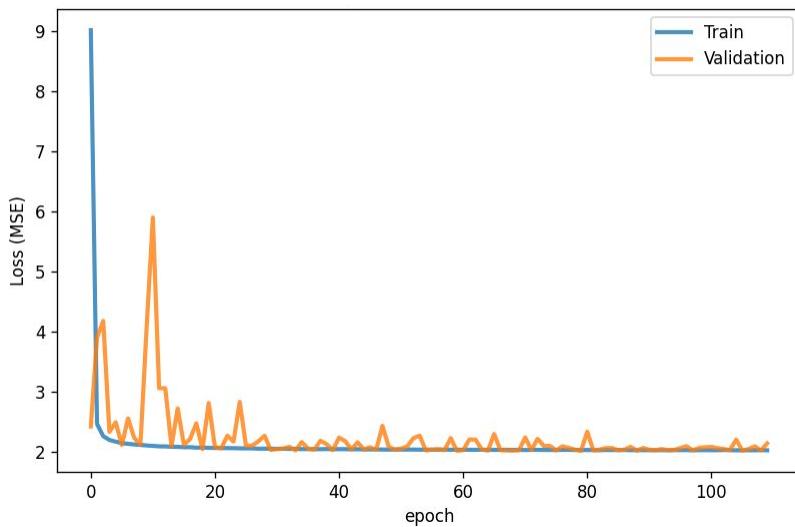
iter:0  
x1(red):2.000000  
x2(blue):2.000000



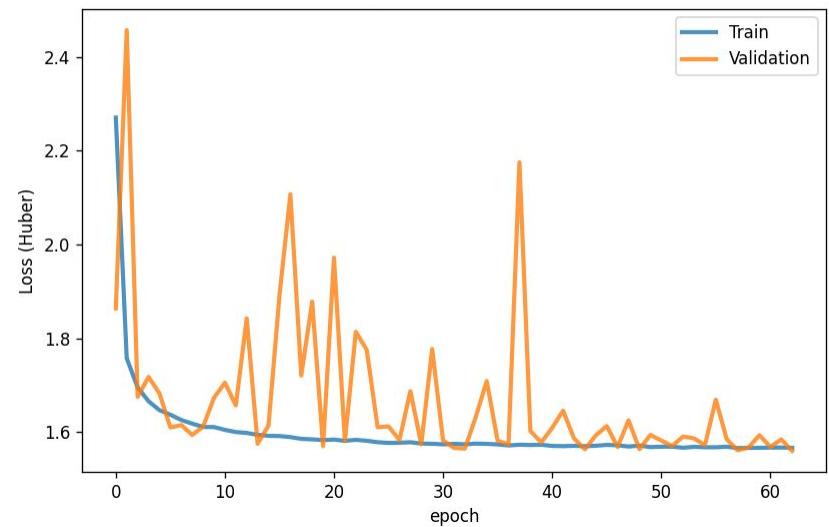
**Unstable around the local minimum!**

# $\theta_{23}$ Loss curves (before normalization)

MSE



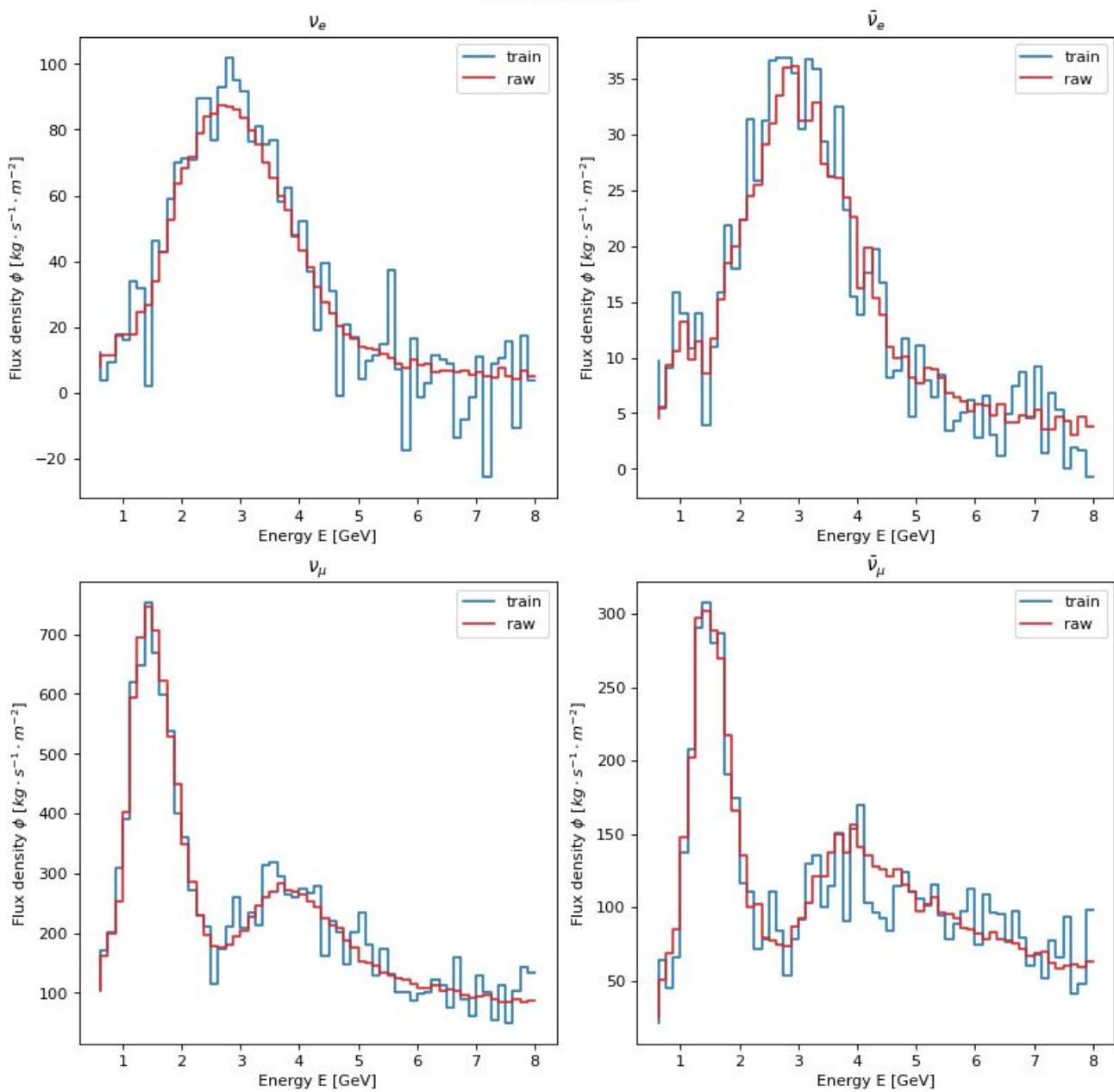
Huber



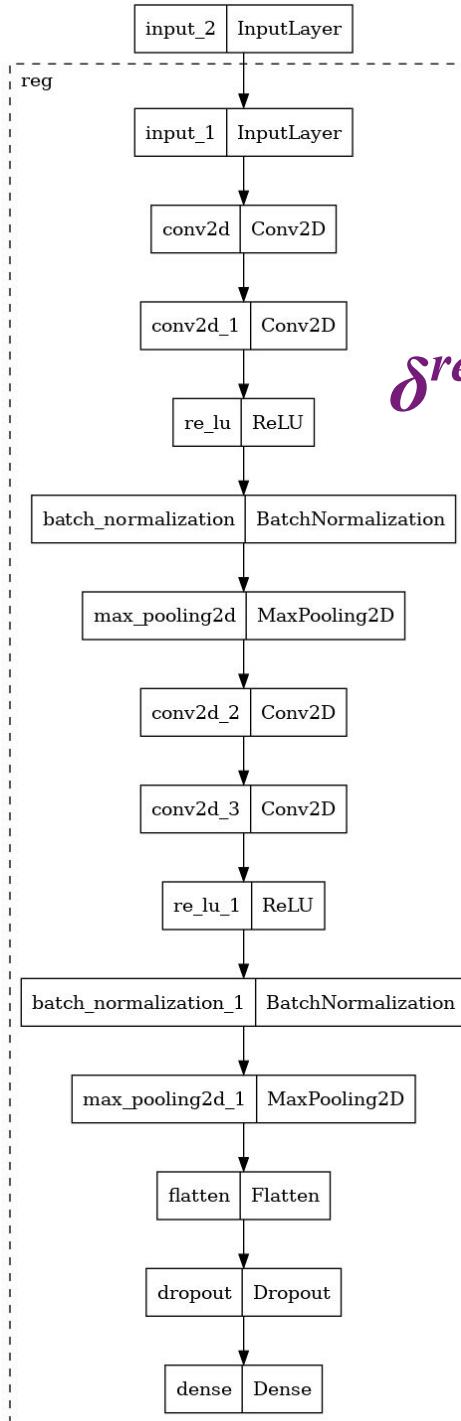
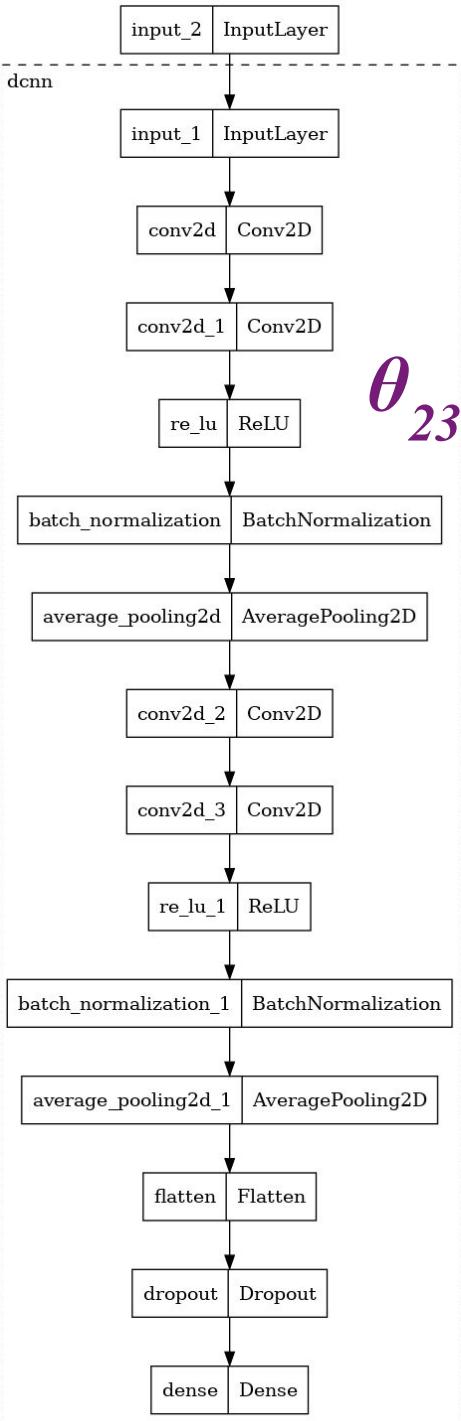
- Large fluctuations if no normalization!
- Huber loss can zoom out the large loss!

# Noisy figures

First 60 features



# Noisy model structures

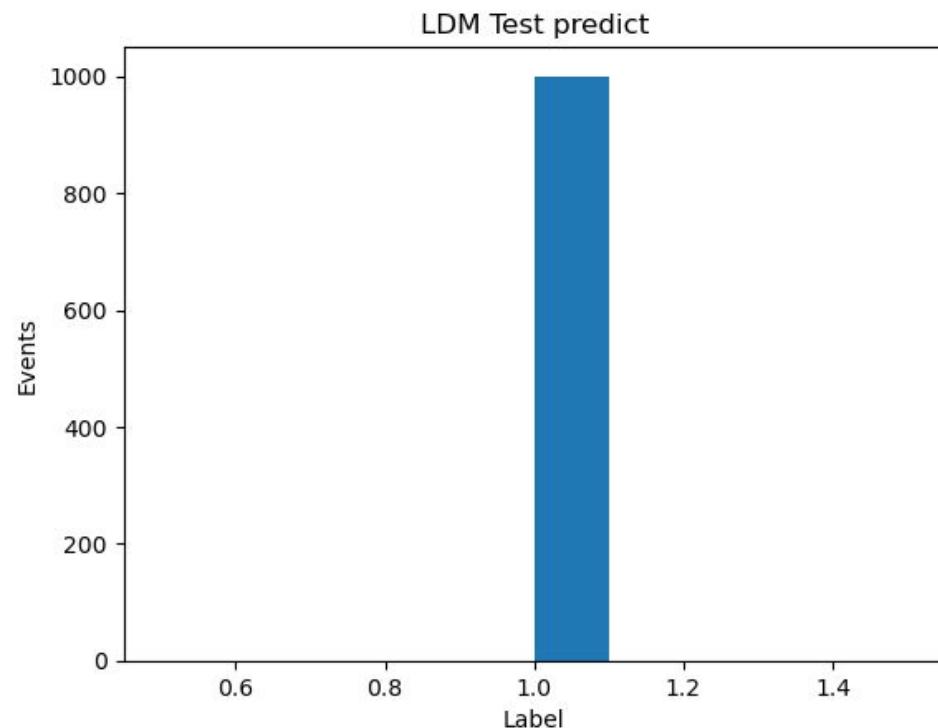


Layer (type)	Output Shape	Param #
<hr/>		
input_1 (InputLayer)	[ (None, 65, 4, 1) ]	0
conv2d_2 (Conv2D)	(None, 33, 2, 64)	320
conv2d_3 (Conv2D)	(None, 17, 1, 64)	16448
max_pooling2d_1 (MaxPooling2D)	(None, 9, 1, 64)	0
flatten (Flatten)	(None, 576)	0
dense (Dense)	(None, 1)	577
<hr/>		
Total params:	17,345	
Trainable params:	17,345	
Non-trainable params:	0	

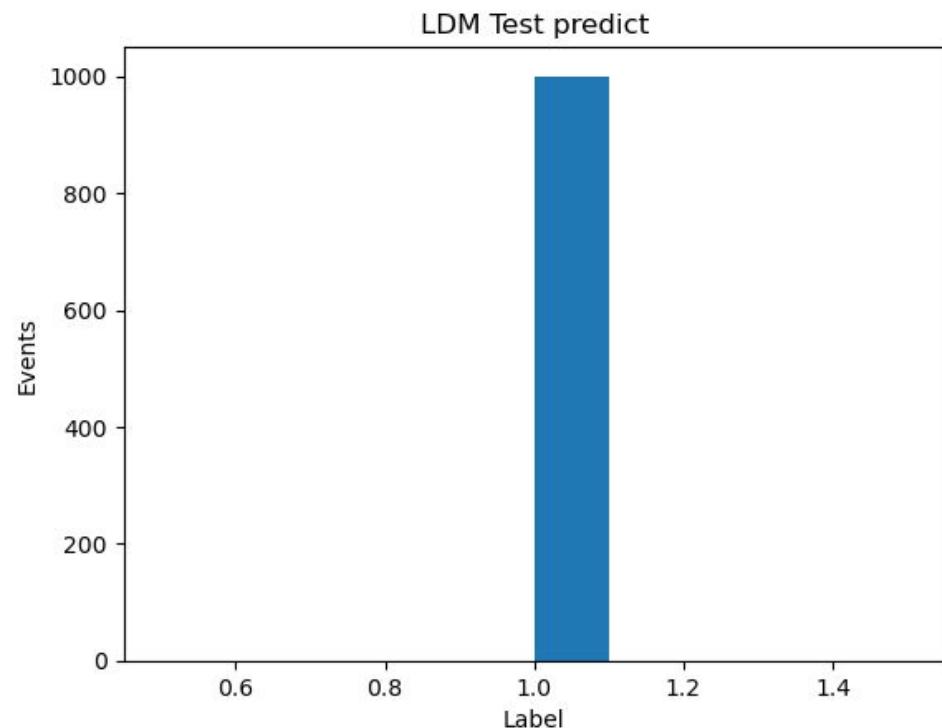
$\delta^{cla}_{cp}$

# Test prediction of LDM classification

With noise



Without noise



# What is being measured?

---

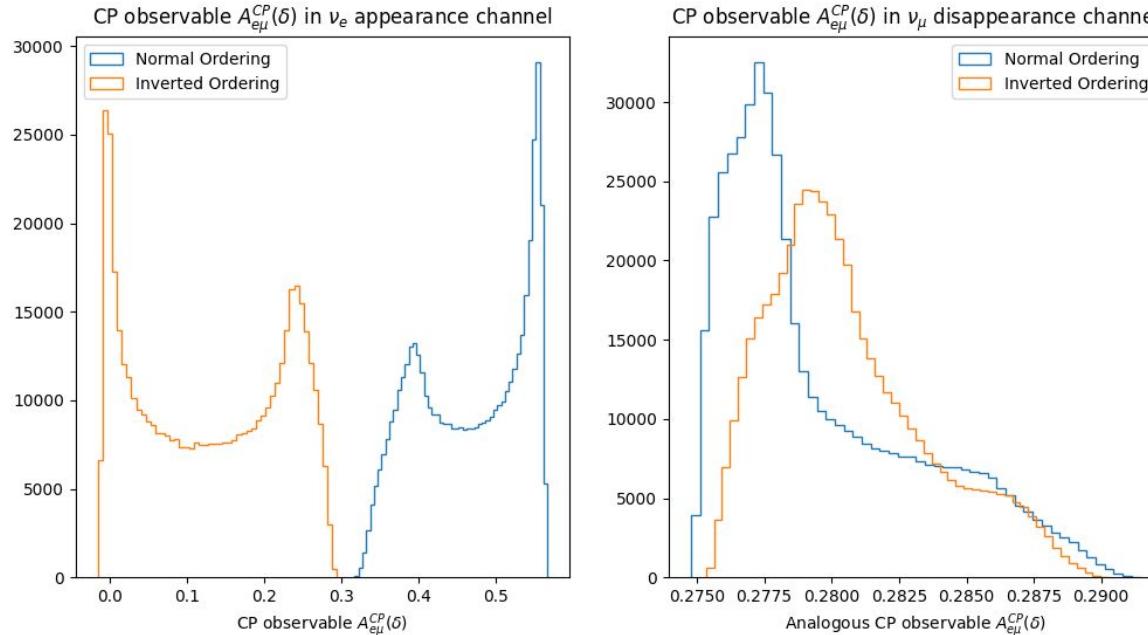
- The experiment aims to probe the effects of neutrino oscillations and CP violation measurements. Consider the CP asymmetry observable:

$$A_{ab}^{CP}(\delta) \equiv \frac{P(\nu_a \rightarrow \nu_b) - P(\bar{\nu}_a \rightarrow \bar{\nu}_b)}{P(\nu_a \rightarrow \nu_b) + P(\bar{\nu}_a \rightarrow \bar{\nu}_b)}$$

- In practice, this observable is constructed from the integrated spectrum of decaying leptons  $N_{\ell^\pm}$  and wrong-sign leptons  $N_{\text{wrong}\ell^\pm}$

$$A_{e\mu}^{CP}(\delta) = \frac{N_{\mu^-}/N_{e^-}^{\text{wrong}} - N_{\mu^+}/N_{e^+}^{\text{wrong}}}{N_{\mu^-}/N_{e^-}^{\text{wrong}} + N_{\mu^+}/N_{e^+}^{\text{wrong}}}$$

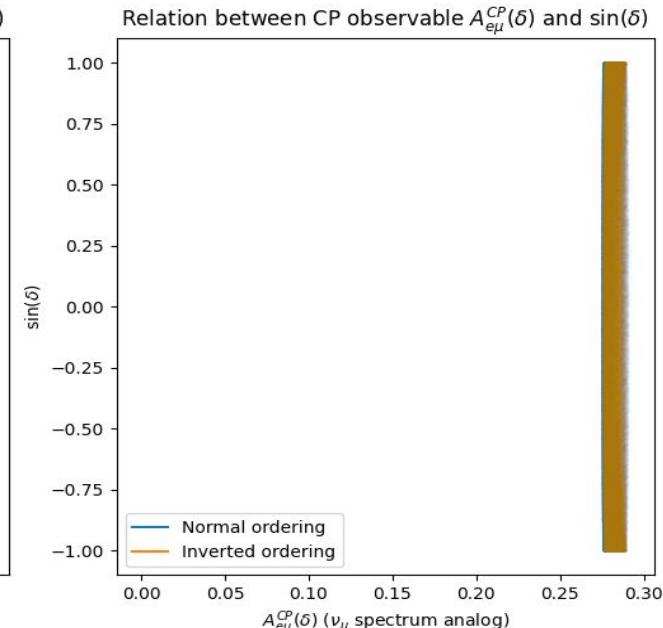
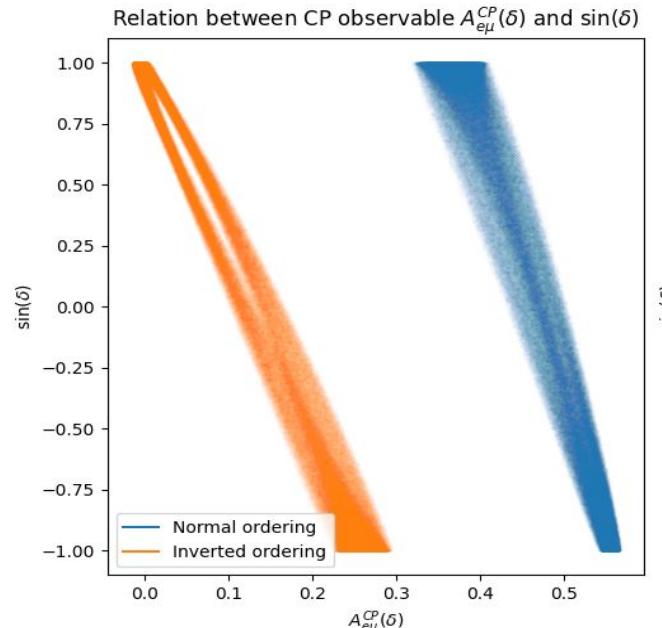
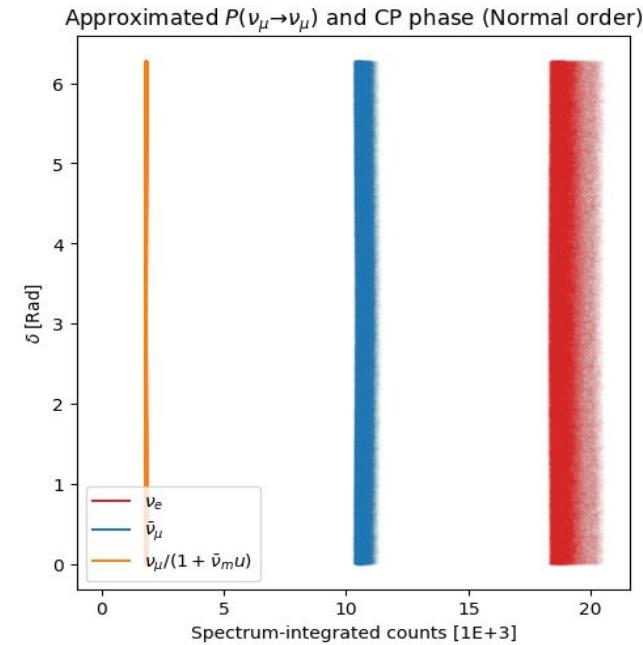
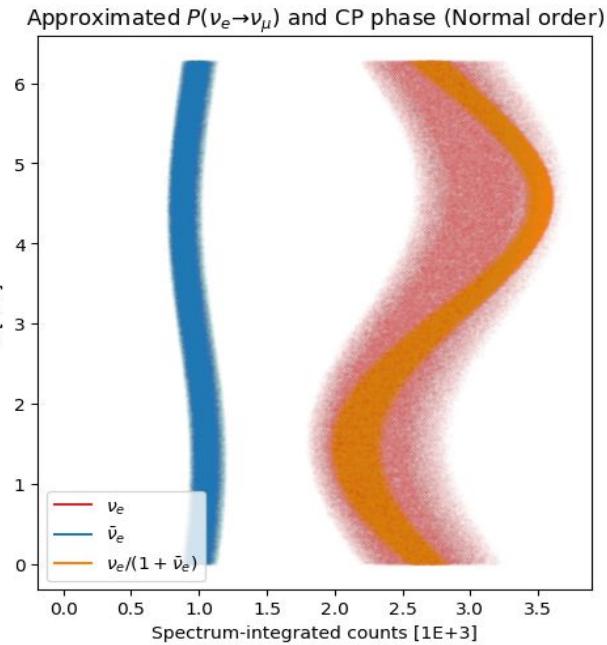
# The CP phase and asymmetry observable



- As mentioned earlier, the CP phase can be quantified by the level of asymmetry.
- Here we present the distribution of CP phase and this observable in both the appearance channel and disappearance channel.

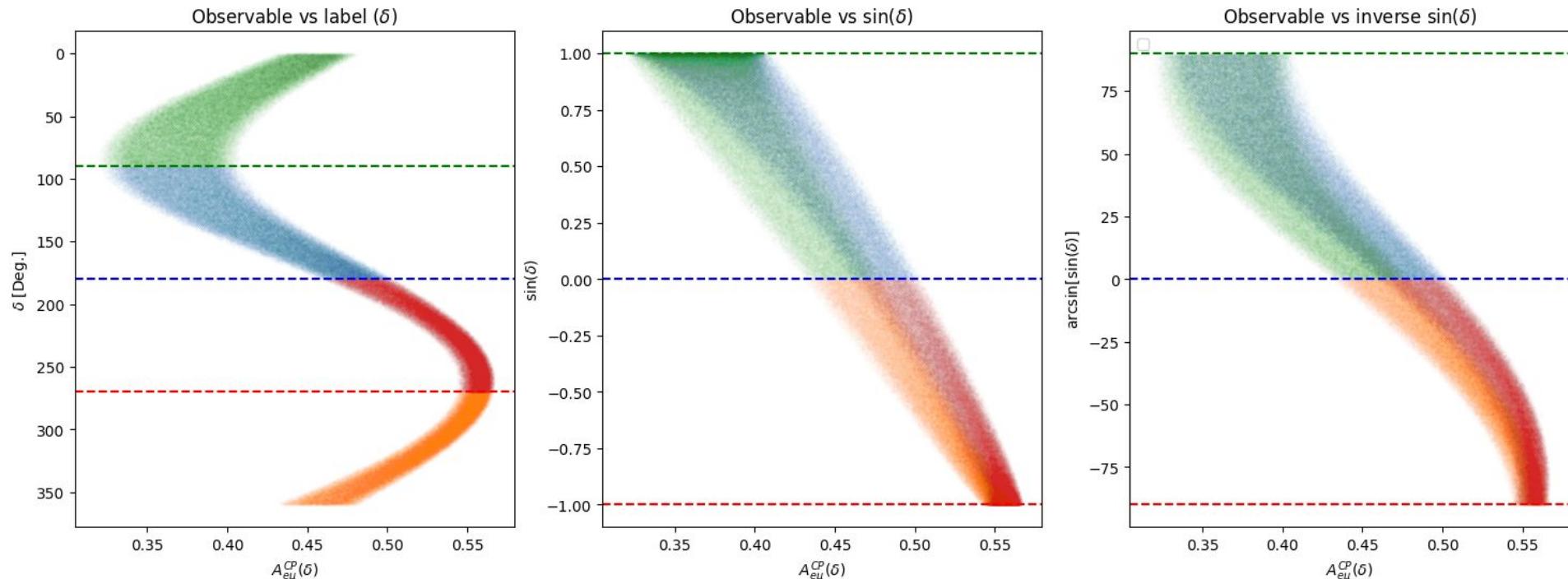
The upper and lower plots present the distribution of CP phase and its relationship with total event counts and the CP asymmetry observable respectively.

Plots of the  $\nu_e$  and  $\nu_\mu$  spectrums are on the left and right side respectively.  
 Our observations show that the event counts in  $\nu_\mu$  spectrums are not correlated with the CP phase.



# The CP phase Regression

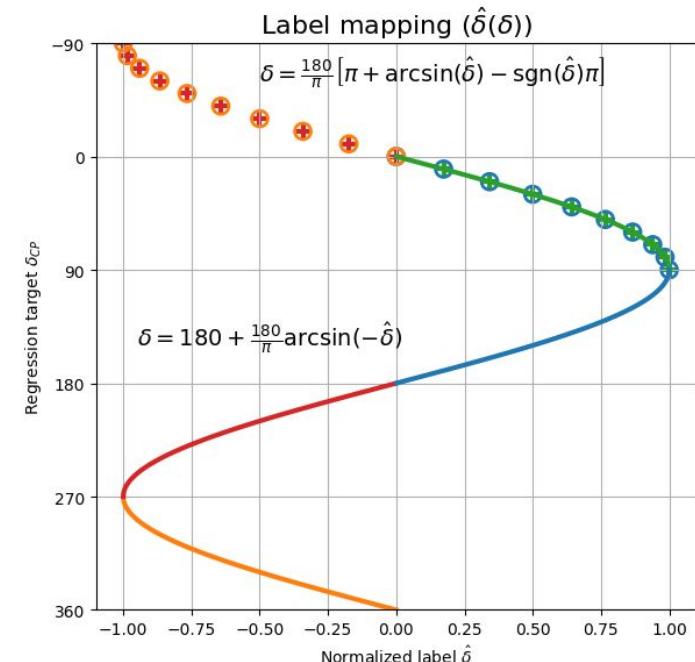
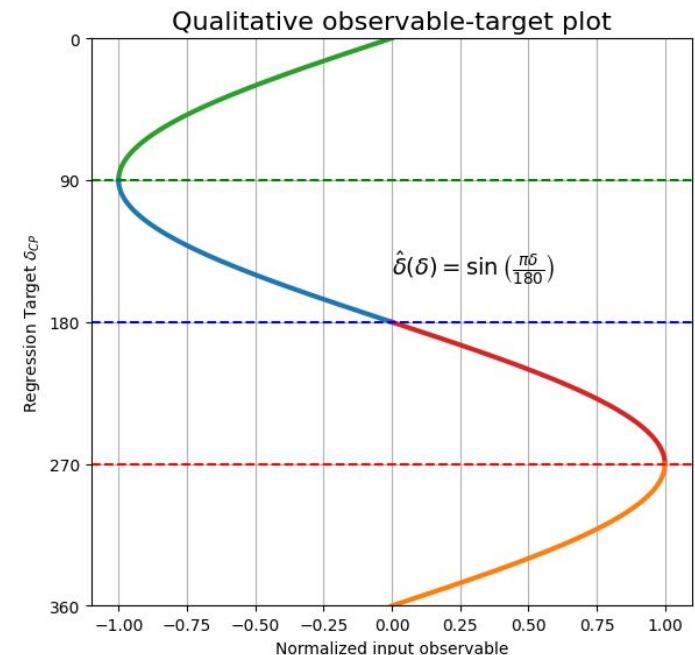
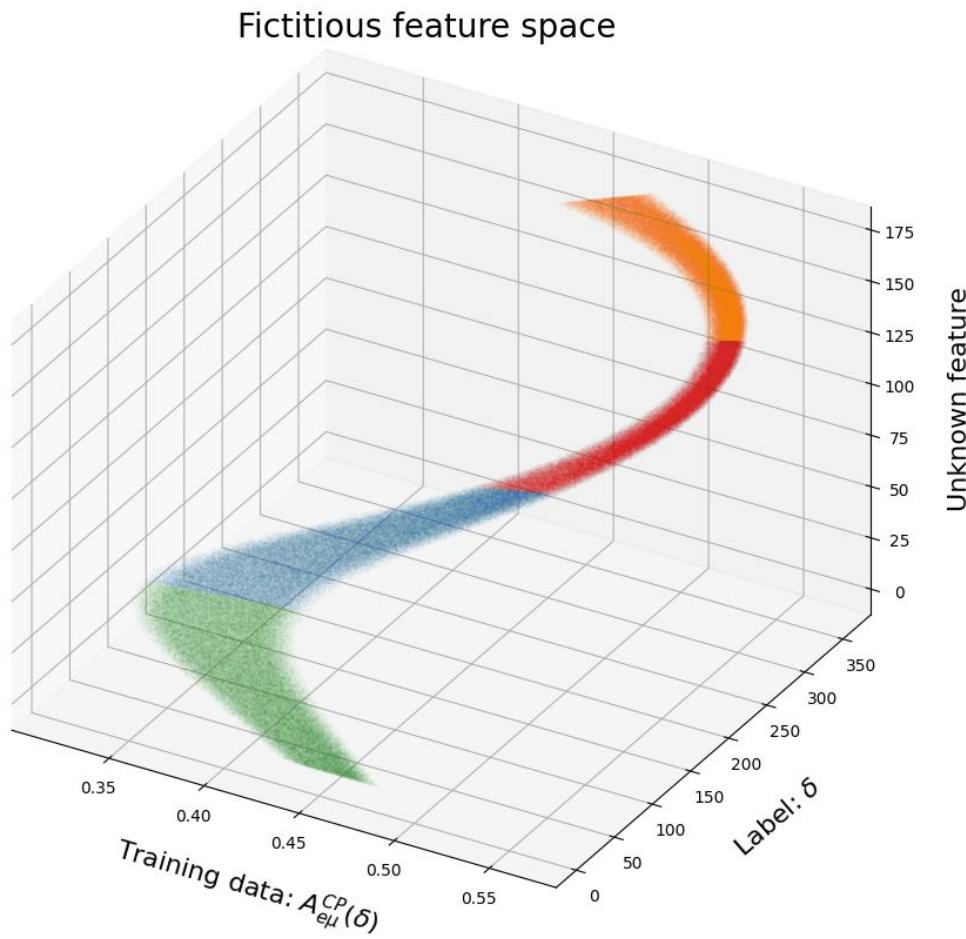
CP observable  $A_{e\mu}^{CP}(\delta)$  and target of regression



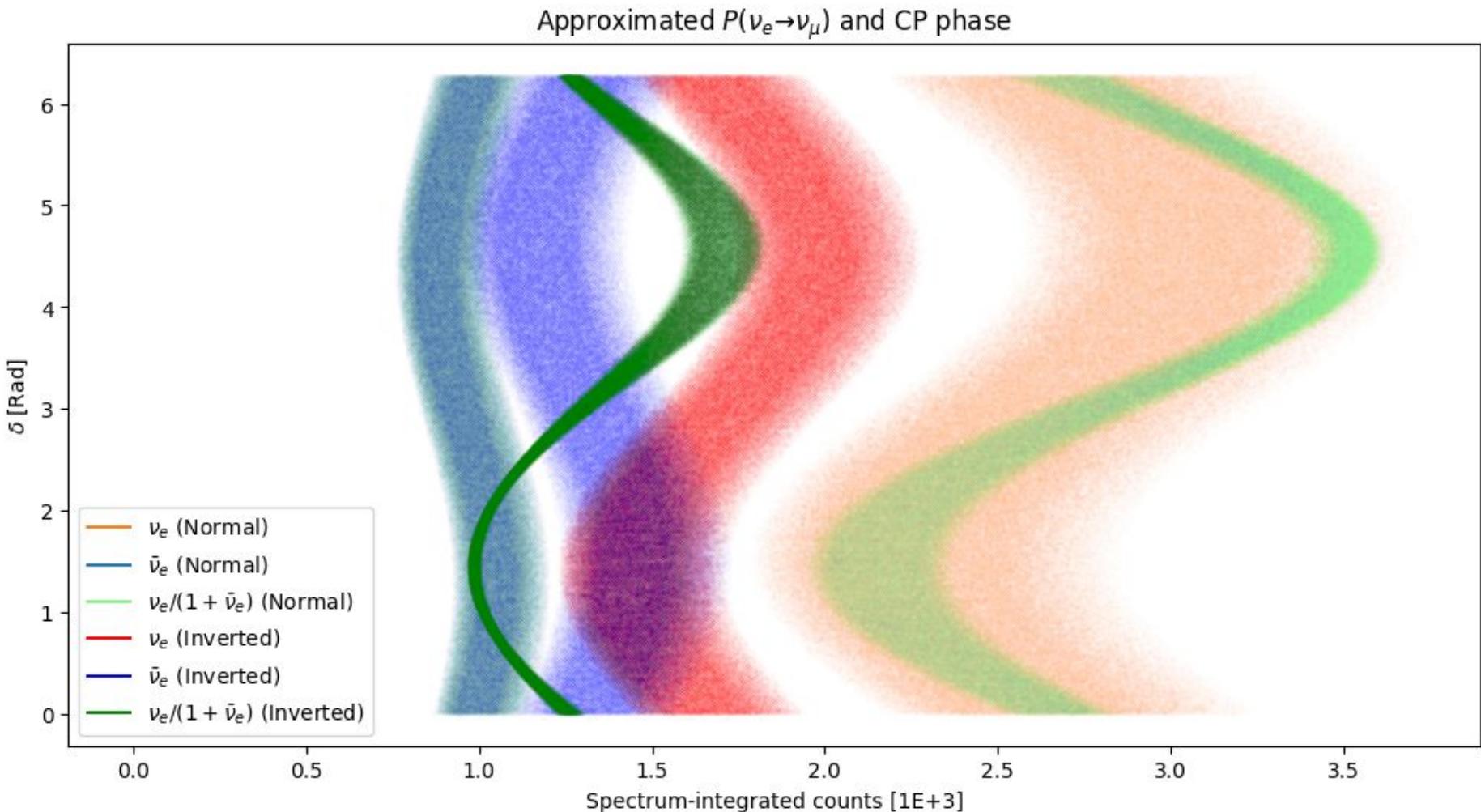
- By taking the sine value of the CP phase, we see a clear one-to-one relationship between the asymmetry observable and the sine value.
- However, transforming back using the arcsine function will only return values in the interval [-90,90].

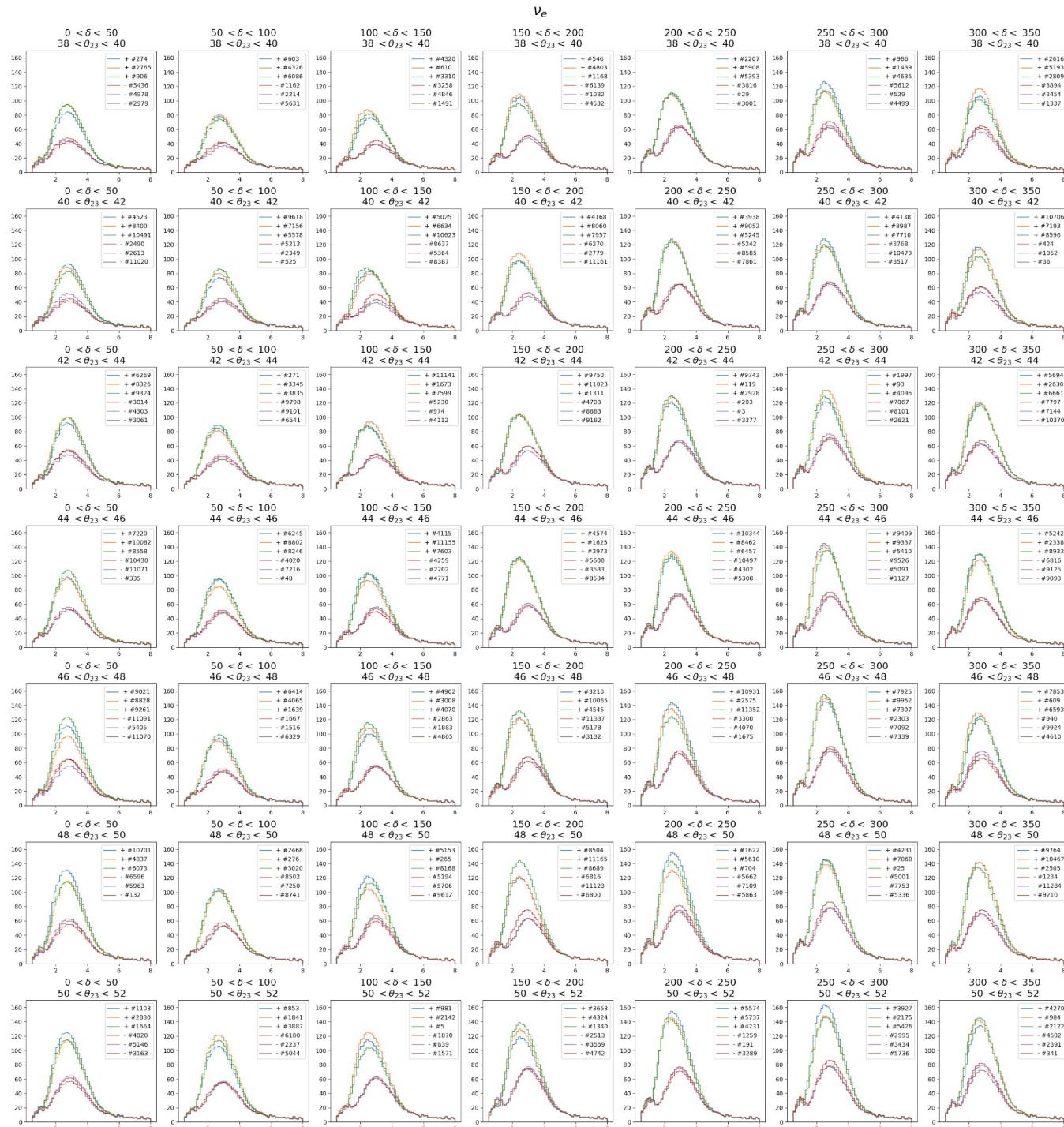


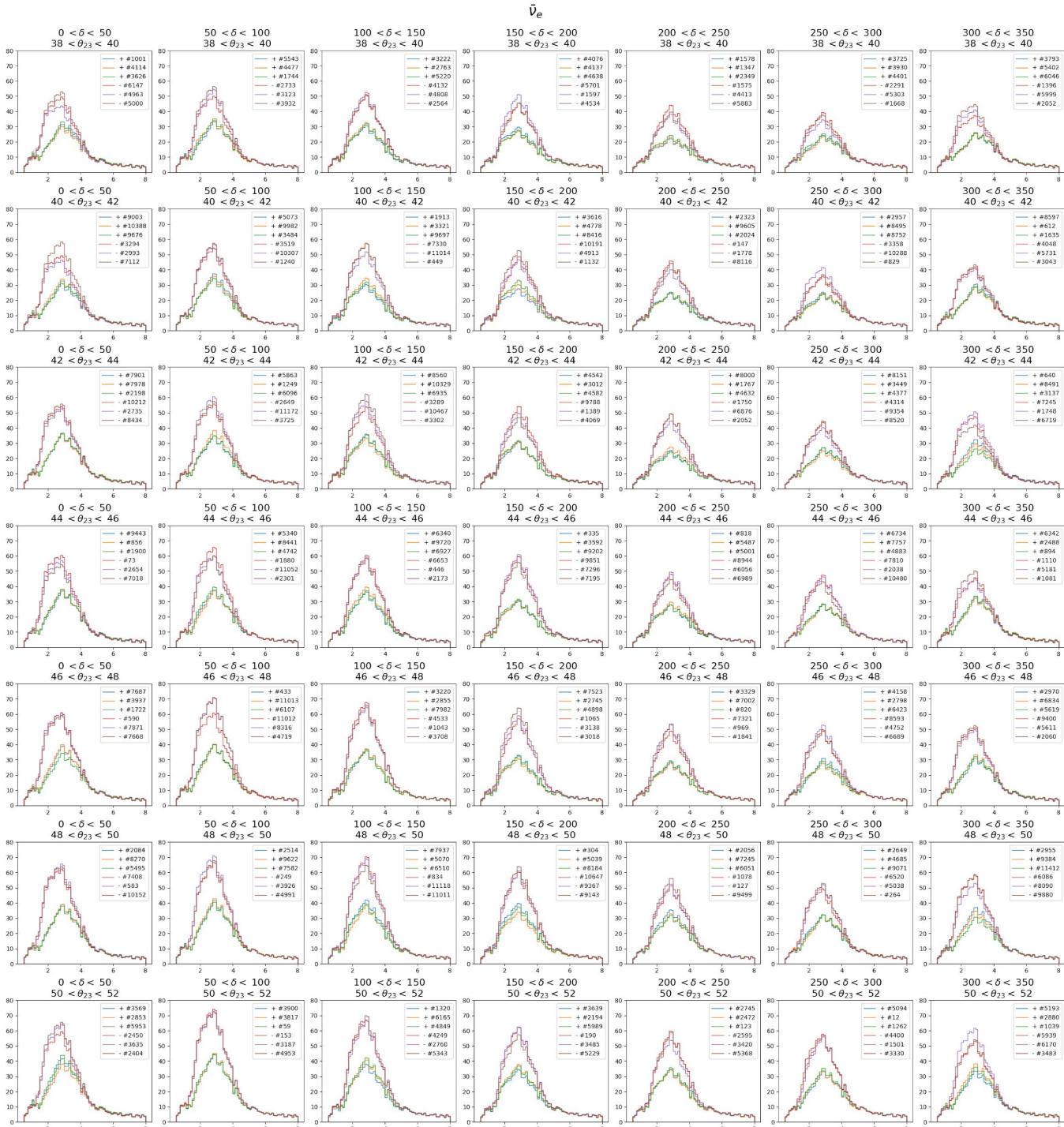
# Idealized feature space

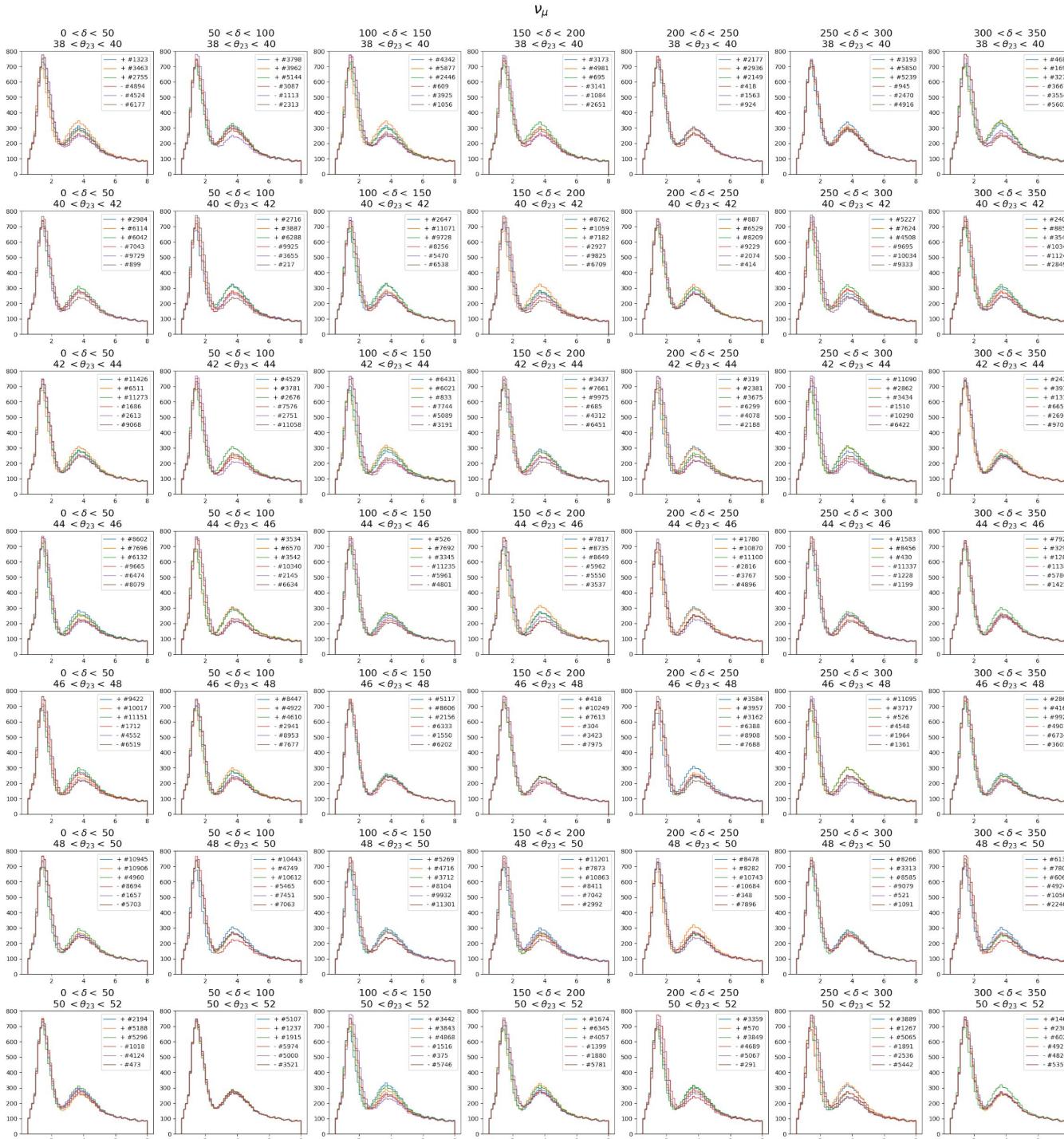


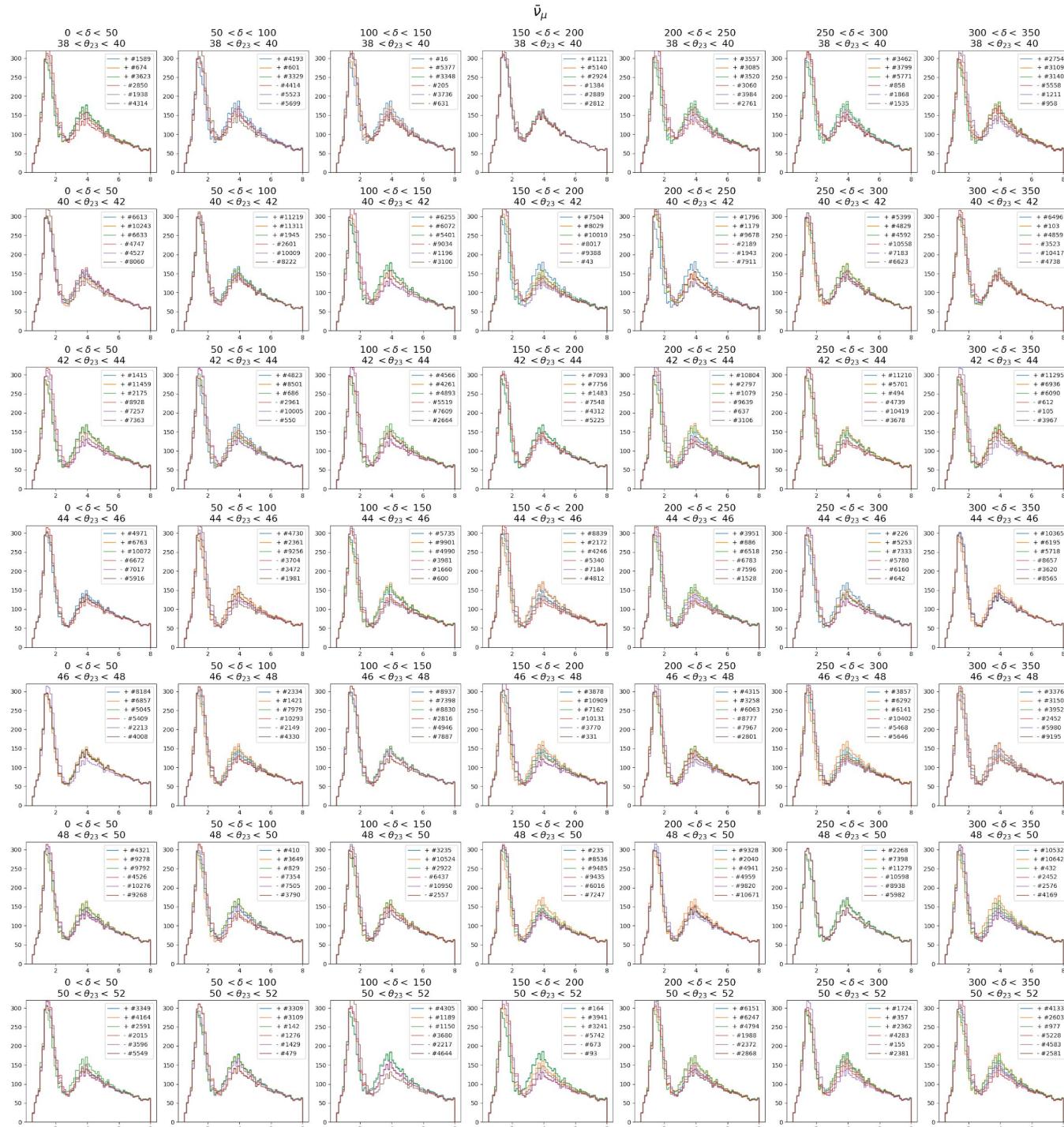
# The CP phase Regression



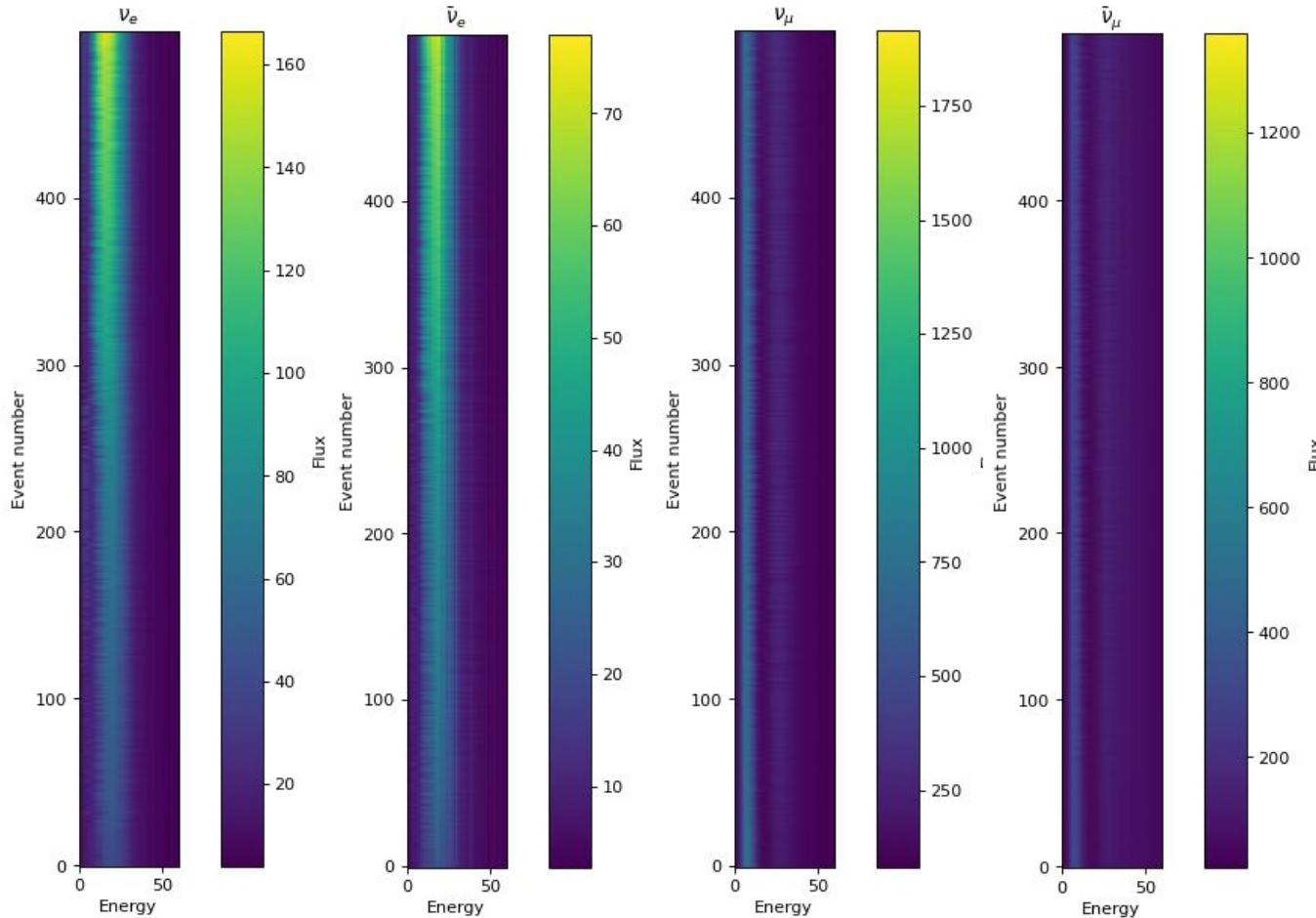




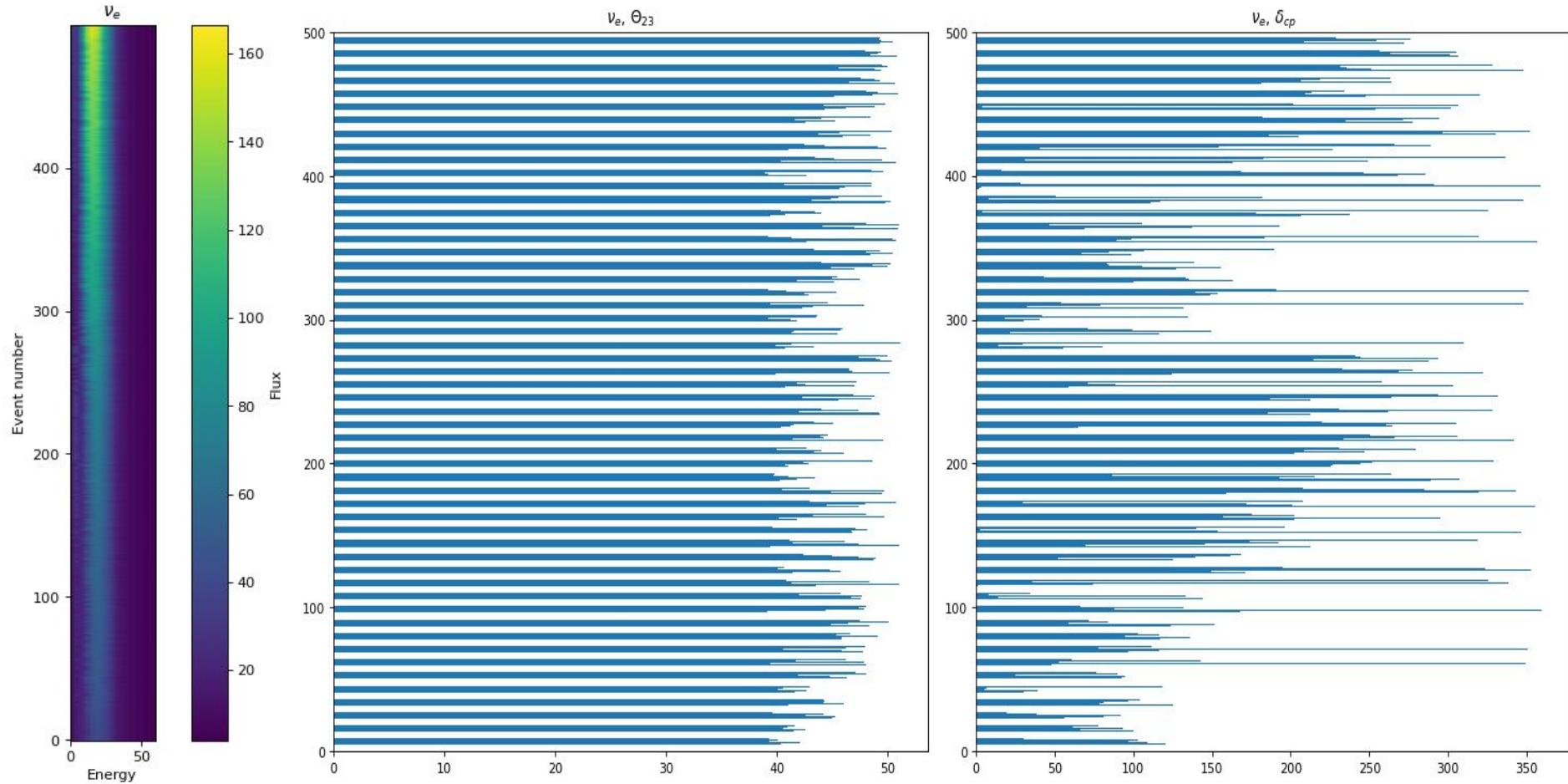


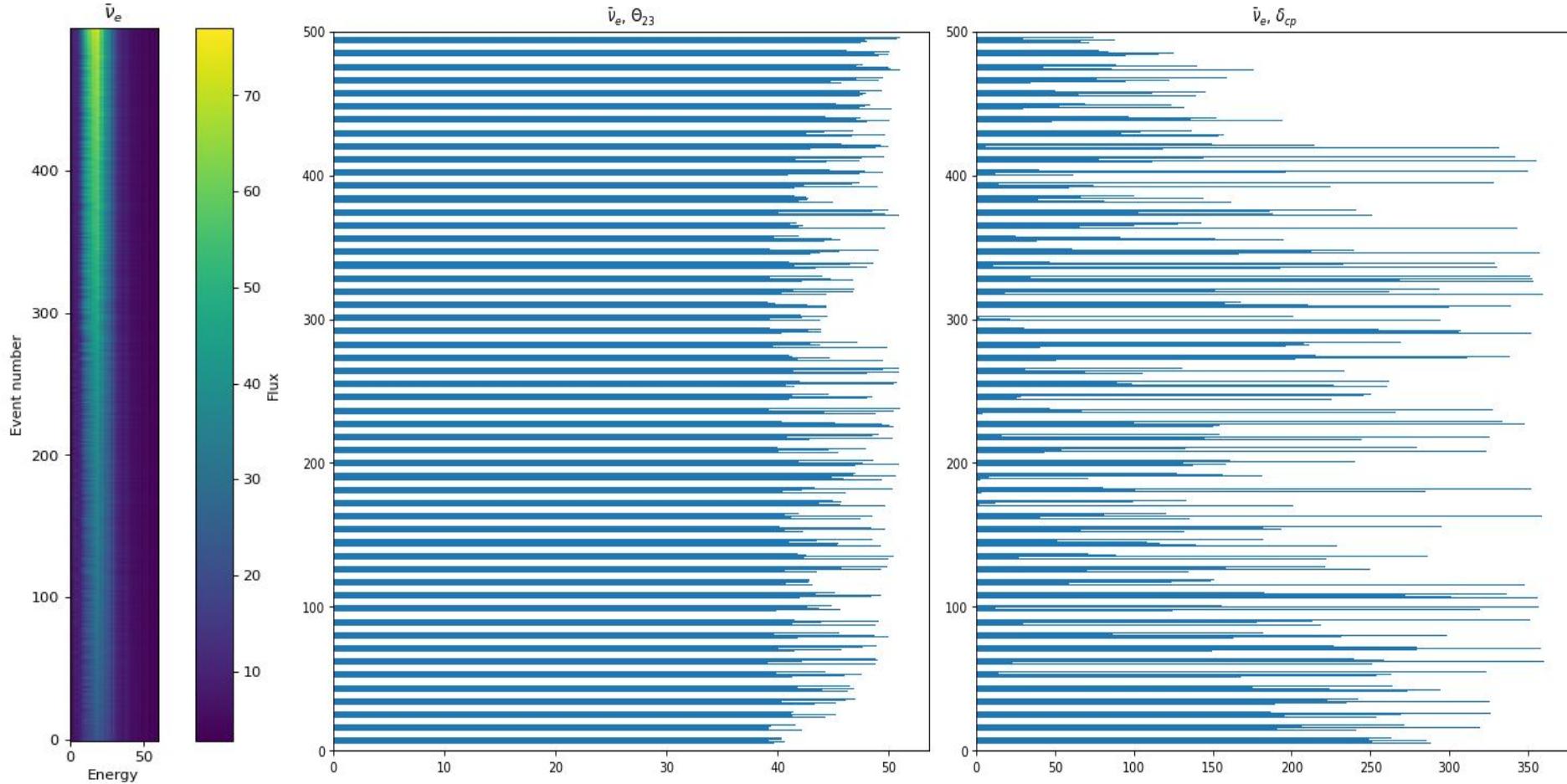


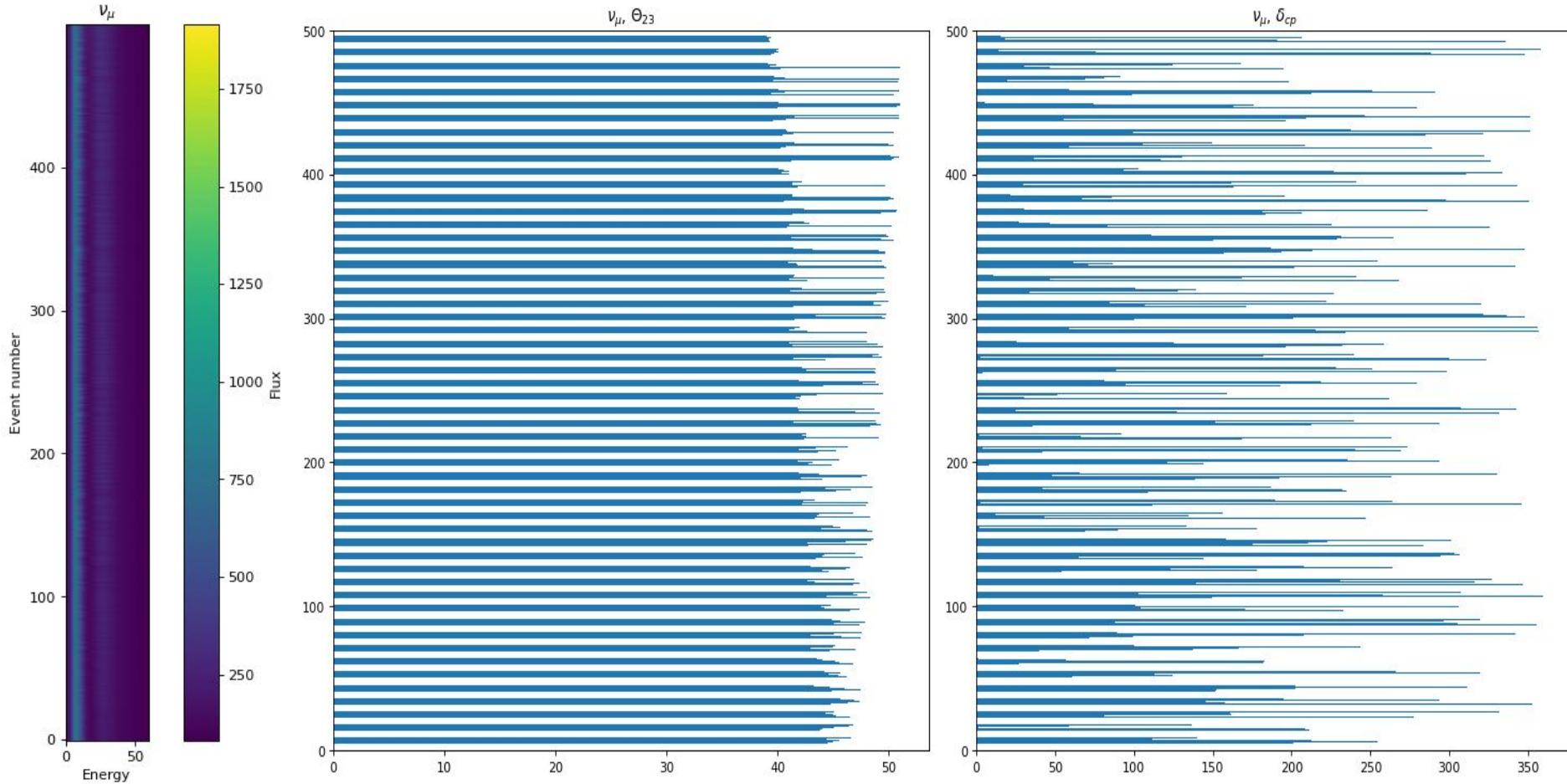
# Sorted four channels with 500 events

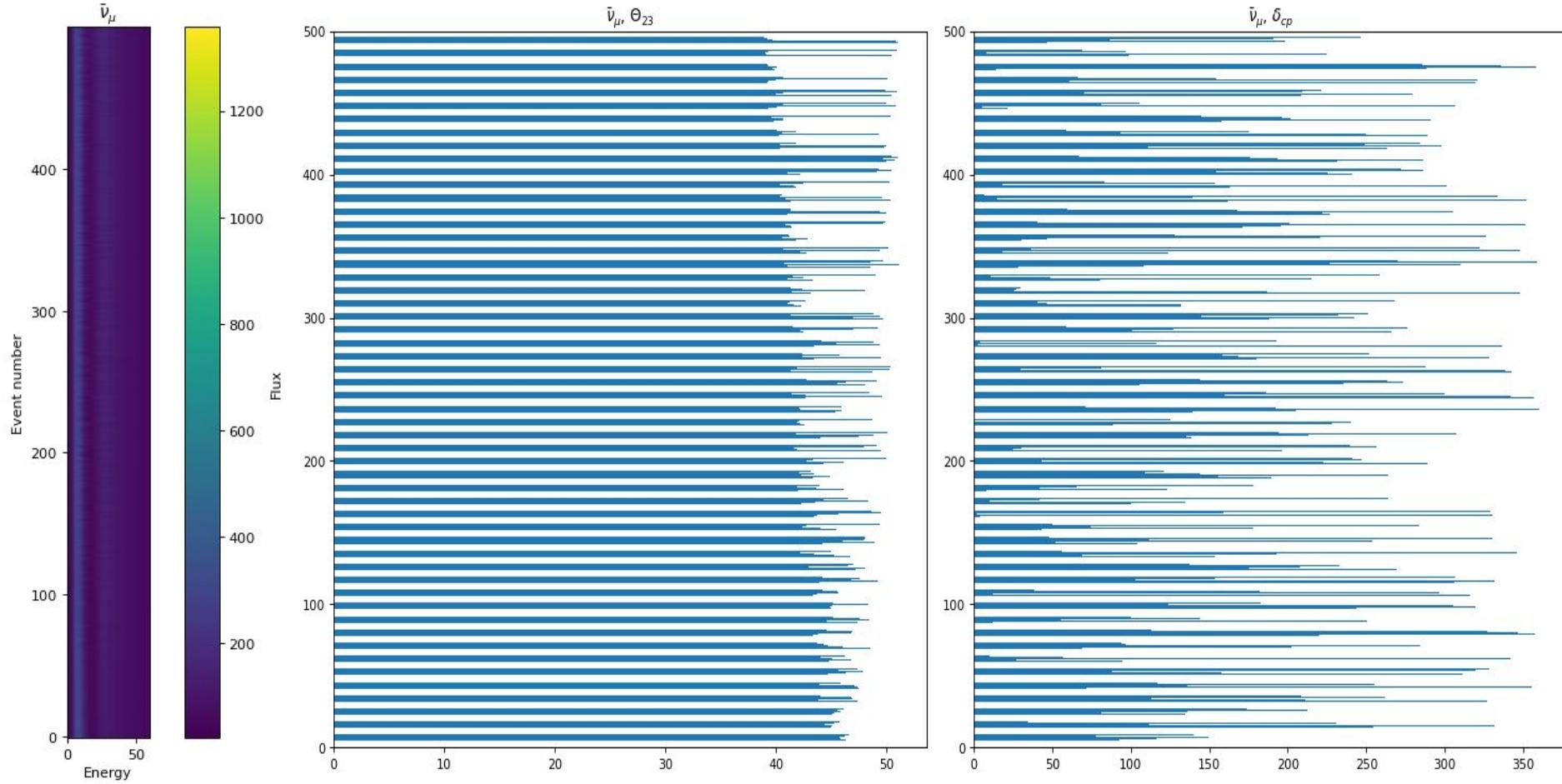


The results are sort by the sum of the specific channel.  
(In this page, the y-axis is sorted by the sum of “event number ” by ve channel)









© Yuan-Yen Peng and Siang-Yuan Lin, Dept. of Physics, National Tsing Hua University, Hsinchu, Taiwan.

Email: [garyphys0915@gapp.nthu.edu.tw](mailto:garyphys0915@gapp.nthu.edu.tw), [sylin@gapp.nthu.edu.tw](mailto:sylin@gapp.nthu.edu.tw)