Machine Learning Techniques to Probe Heavy Neutral leptons in the electron channel at FCC-ee

Pantelis Kontaxakis
on behalf of the HNLs (evjj) team*

October 9, 2024

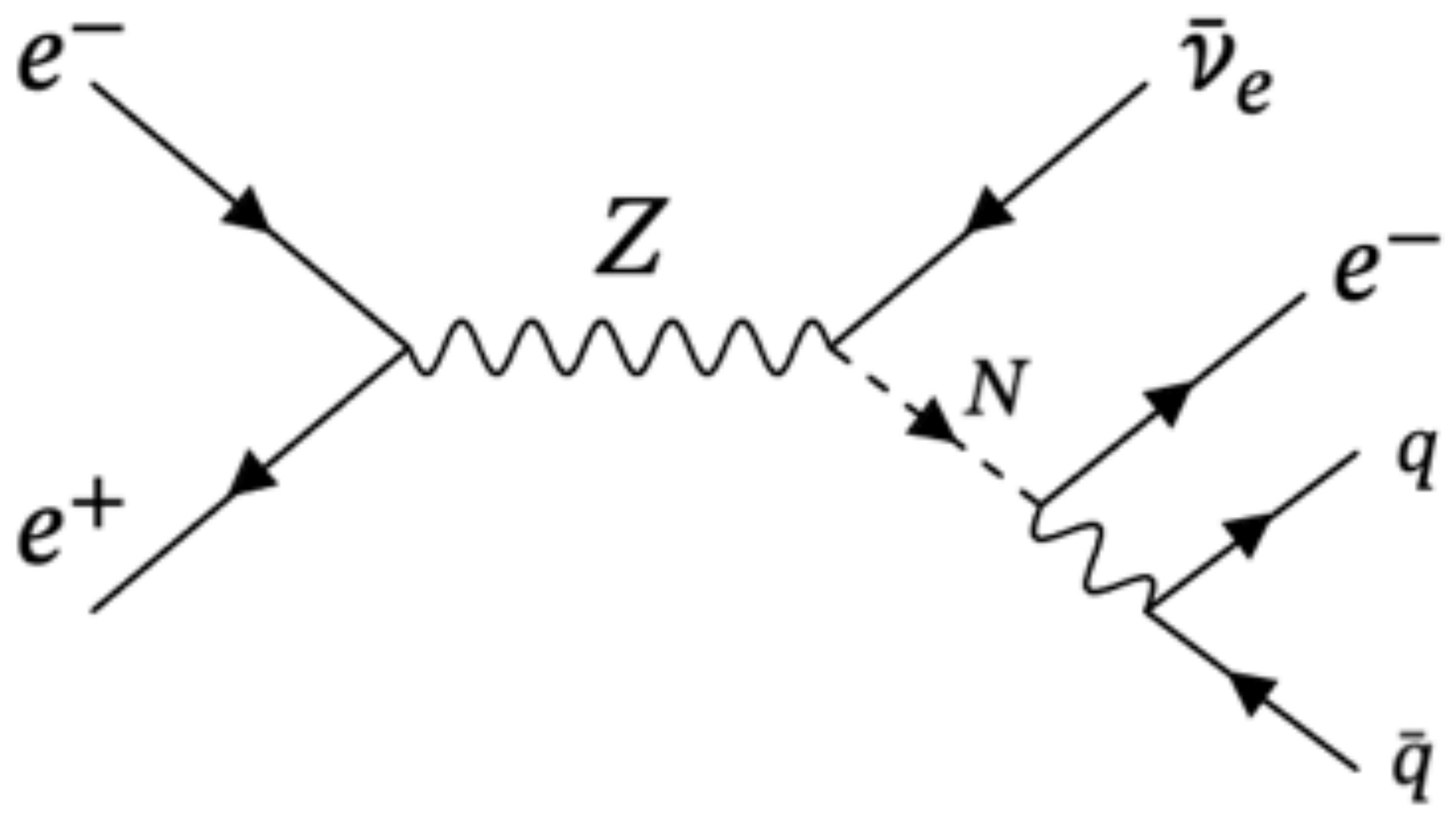
* Work presented of behalf of the following people from the University of Geneva: Dimitri Moulin, Thomas Critchley, Pantelis Kontaxakis, Anna Sfyrla

Introduction

- Neutrinos have **very small, but non-zero masses**, as demonstrated by baseline neutrino oscillation experiments

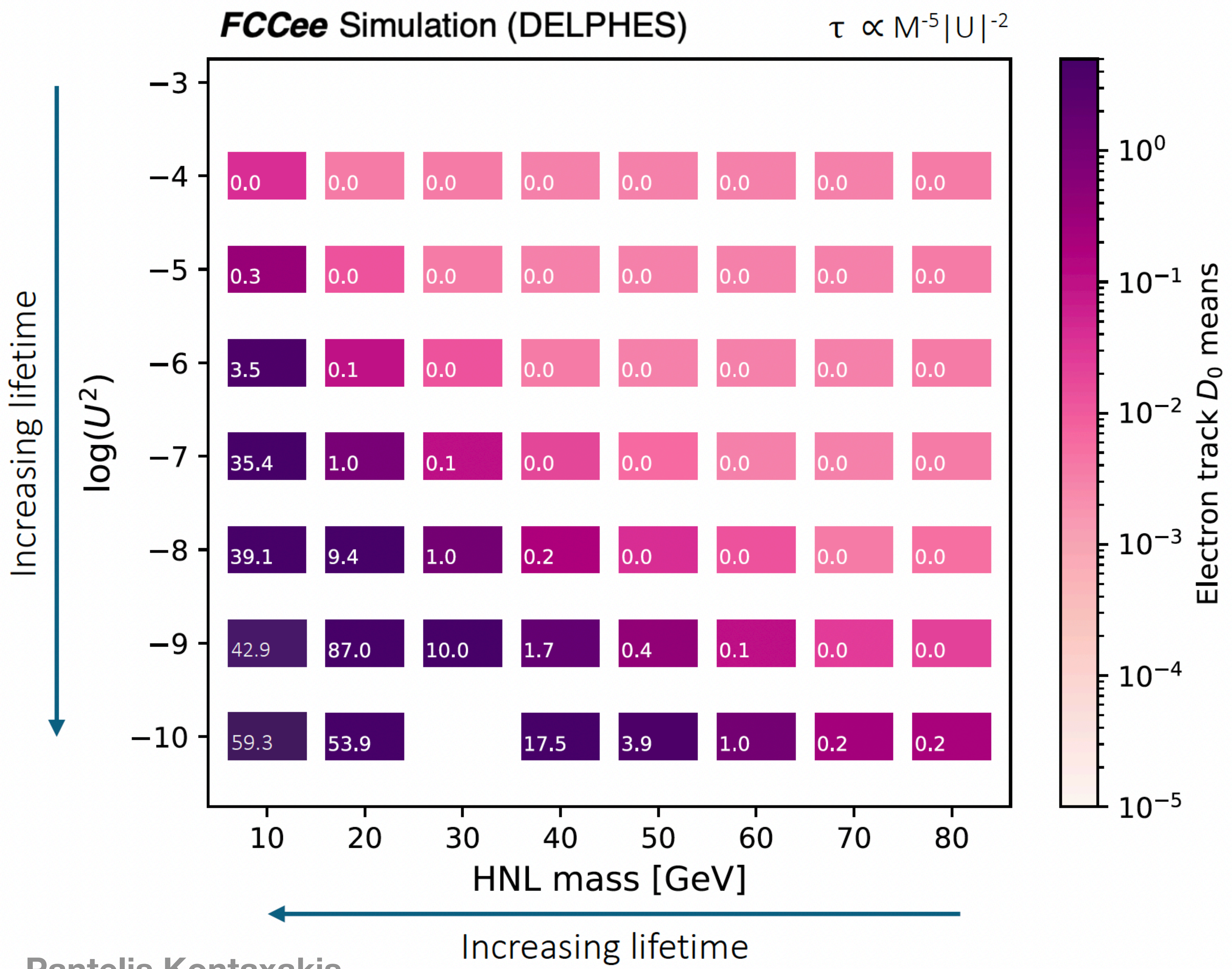
- **Low-scale** inverse seesaw mechanism enables the search for heavy right-handed neutrinos with Yukawa couplings $O(10^{-6})$ in the mass range of **10** to **100 GeV**

- Our analysis focuses on **the electron final state with two jets**, investigating the (pseudo-) **Dirac** HNL model between **10-80 GeV** with mixing angles between $10^{-4} < |U_{eN}|^2 < 10^{-10}$

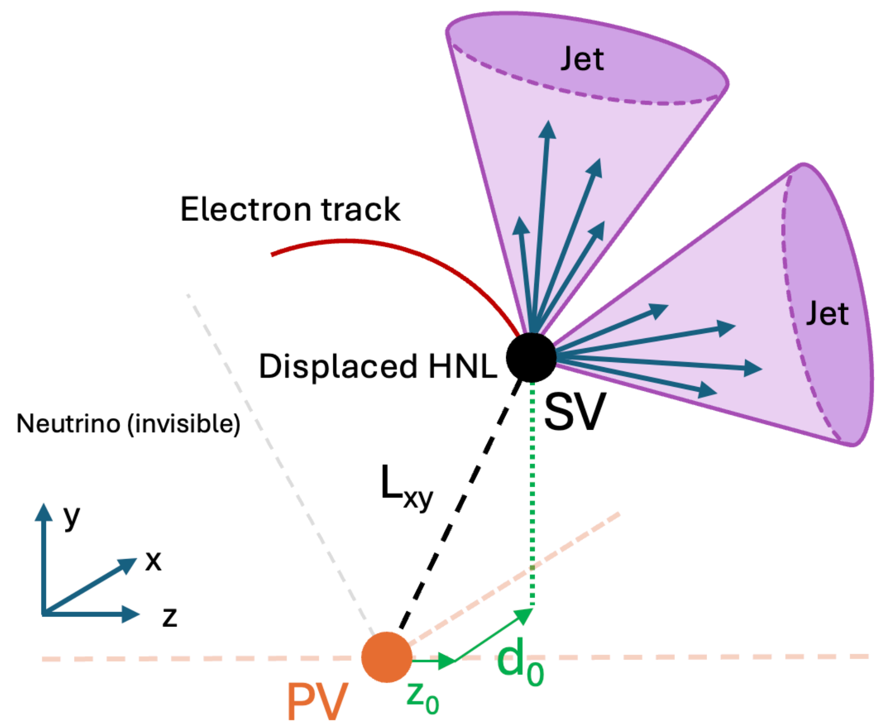


HNL Phenomenology at the FCCee

- **The FCC-ee is expected to produce approximately 10^{12} Z bosons during its Z-channel run (spanning ~3 years of data collection)**
  - High-luminosity & pileup-free environment for the search for HNLs
  - Aim to improve upon the limits previously set by the LEP



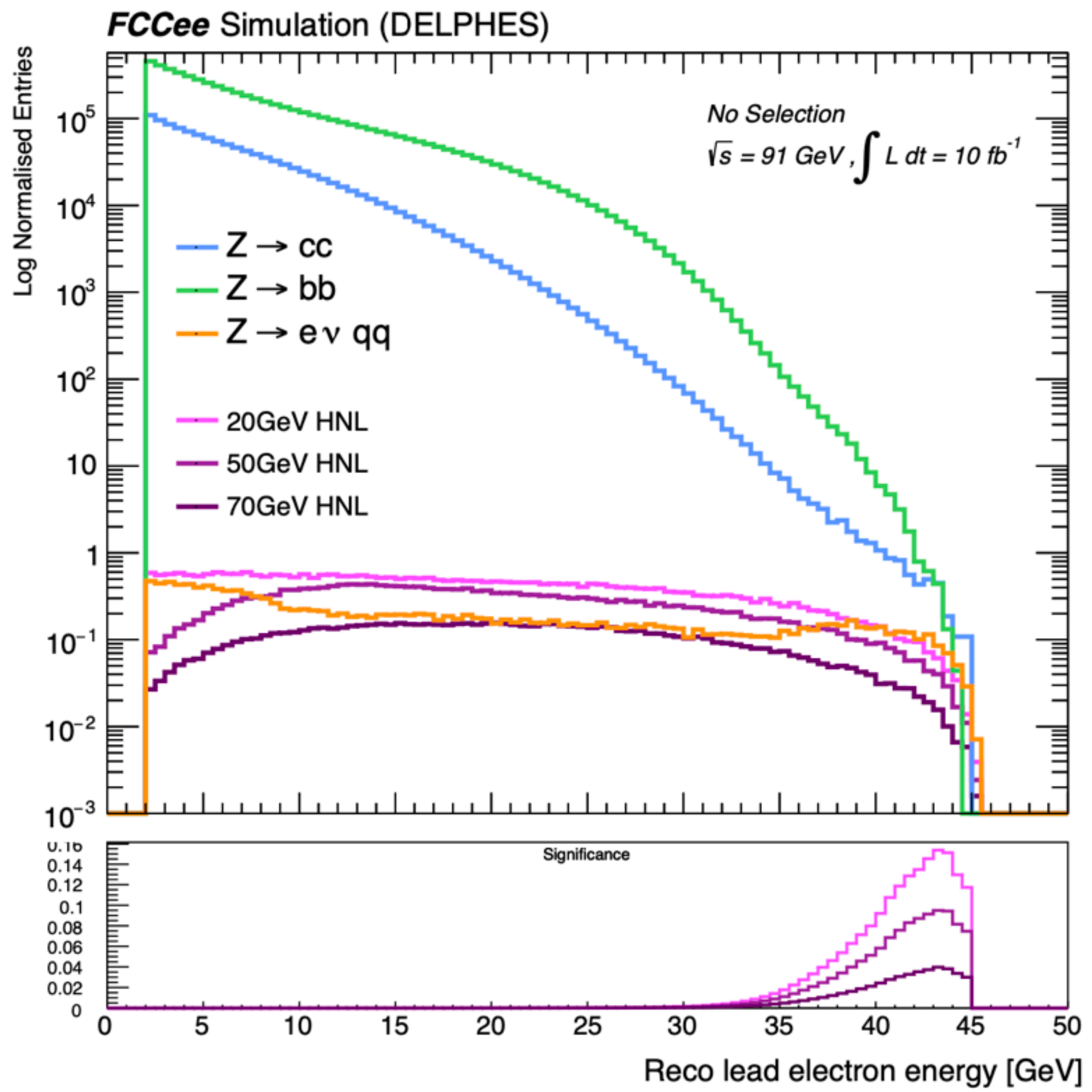
- For many of the mass points under consideration:
  - A displaced topology emerges due to the significant lifetime ($\tau \propto M^{-5}|U|^{-2}$)
  - Can be distinguished from promptly decaying mass points using lifetime metrics, such as decay length or D_0



Background Processes

◉ **Three dominant SM background processes considered:**

- Z→bb, cc or Z→4 body final state (instead of heavier quark final states like Z→ττ)
- The 4-body bkg and all signal samples are privately generated using MadGraph



Process	σ(pb)	Monte-Carlo events	Production $\mathcal{L}$ (fb^{-1})
$Z \to b\bar{b}$	6.65×10^3	4.39×10^8	6.60×10^1
$Z \to c\bar{c}$	5.22×10^3	4.98×10^8	1.15×10^2
$Z \to$ 4body	1.40×10^{-2}	1.00×10^5	7.14×10^3

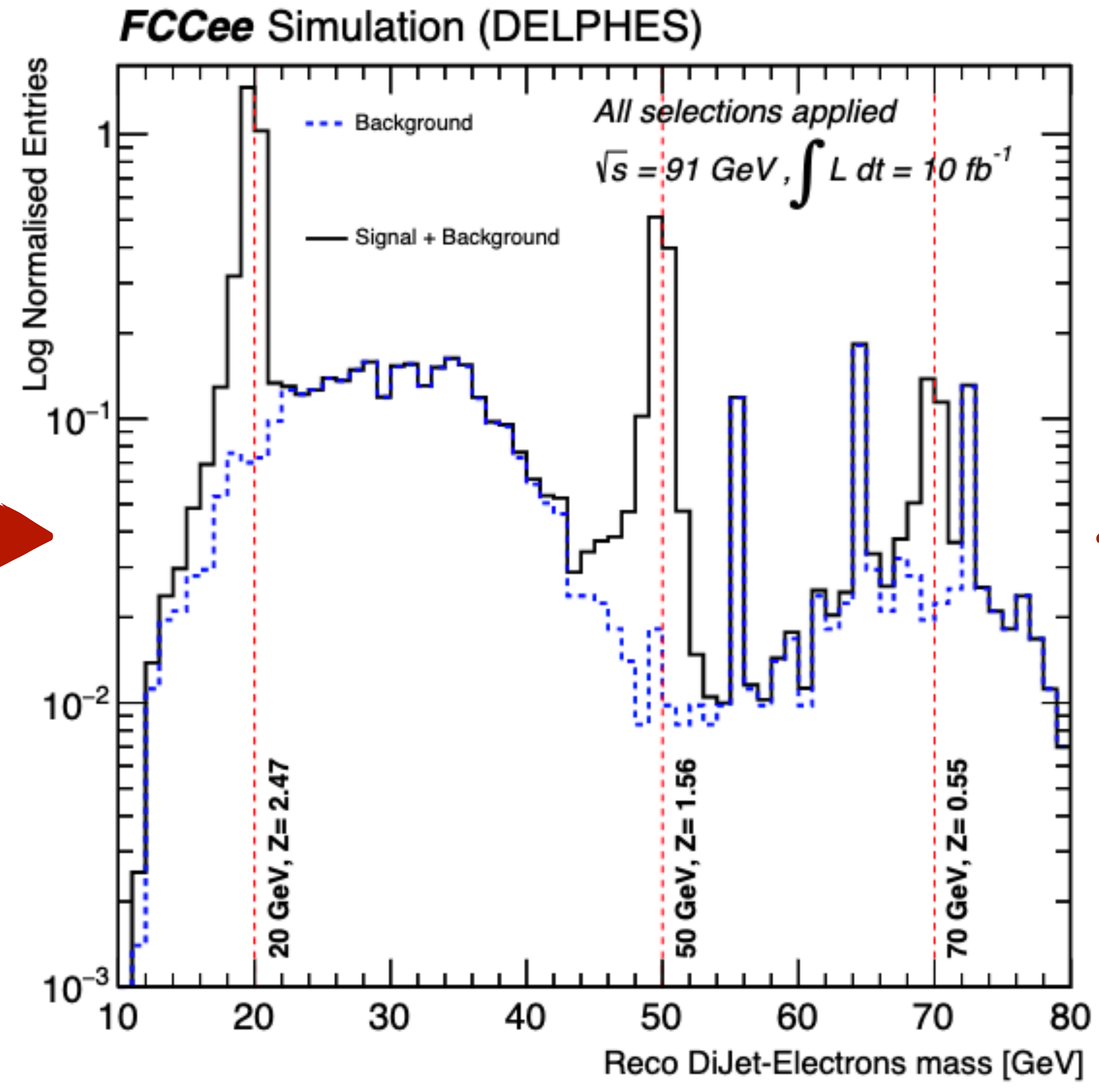
Limited MC statistics for central backgrounds; the analysis is conducted at **10 fb^{-1}** and scaled to **150 ab^{-1}** for the final result

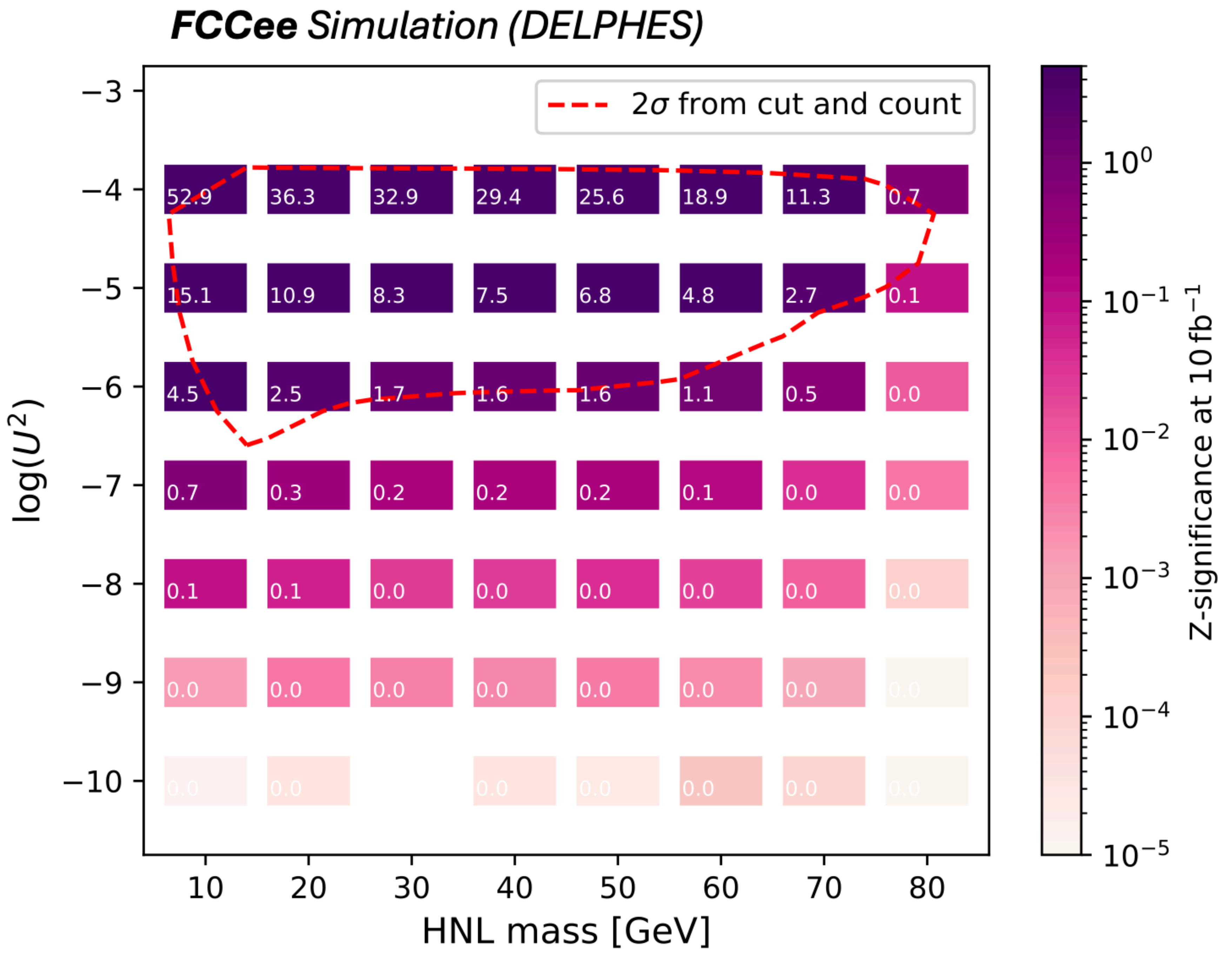
Analysis Methods

◉ Cut and Count Method

- Started with a Cut & Count (C&C) analysis which used as baseline for further optimization strategies

Variable	Selection
Missing energy	$> 12\,\text{GeV}$
Leading electron energy	$> 35\,\text{GeV}$
3D di-jet Angle	$< 2.4\,\text{rad}$
Di-jet – Electron ΔR	< 3





Significance: $Z = \sqrt{2\left(n \cdot \ln\left[\frac{n\,(b+\sigma^2)}{b^2+n\sigma^2}\right] - \frac{b^2}{\sigma^2}\ln\left[1+\frac{\sigma^2(n-b)}{b\,(b+\sigma^2)}\right]\right)}$

D. Moulin master thesis | T. Critchley master thesis

Analysis Methods

T. Critchley master thesis

◉ Machine Learning Method(s)

- Explored multivariate methods trying to increase the sensitivity from the C&C method

✦ BDT Method:

- XGBoost in conjunction with TMVA (binary classification)

✦ DNN Method:

- Keras in Tensorflow with hyperparameter optimization (binary classification)

❖ For both methods:

- Individual training for every mass point trying to reach the full sensitivity
- The following variables were used for the training →

Object	Variables
Leading electron	$E, \phi, d_0, \sigma_{d_0}, \Delta R_{ejj}$
Neutrino	E_{miss}, θ
Di-jet system	$\Delta R_{jj}, \phi$
Vertex and tracks	$n_{\text{tracks}}, n_{\text{primary tracks}}, \chi^2_{\text{vertex}}$

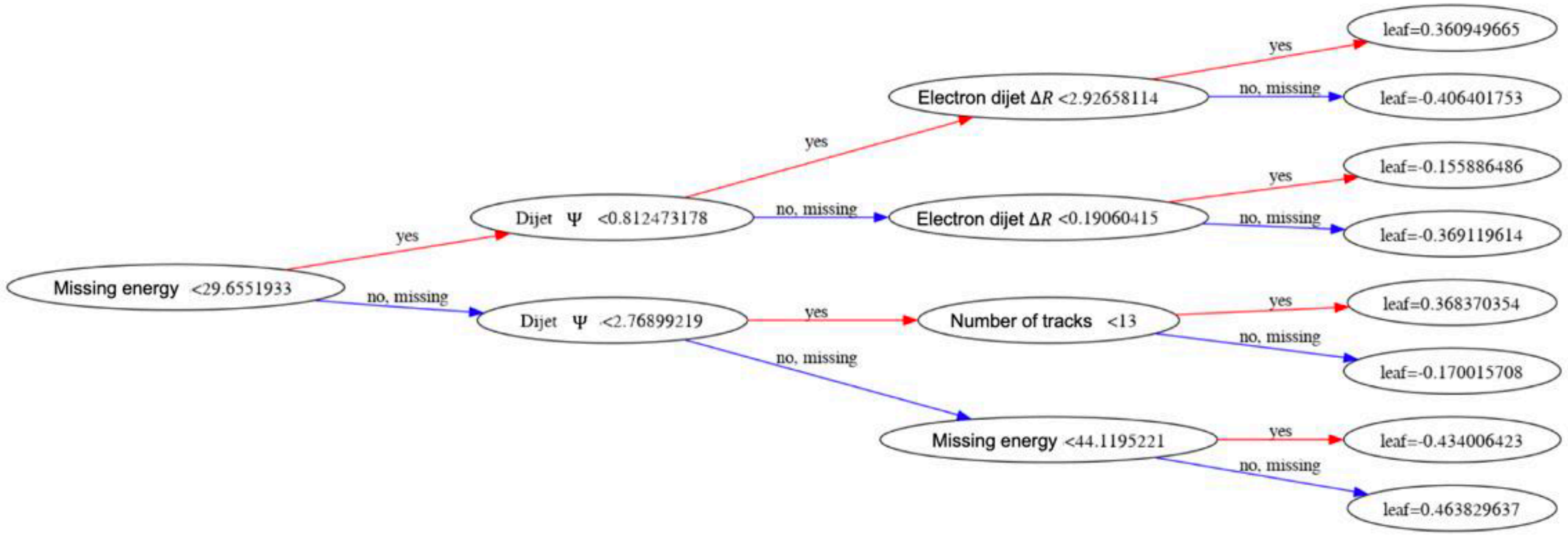
BDT Workflow

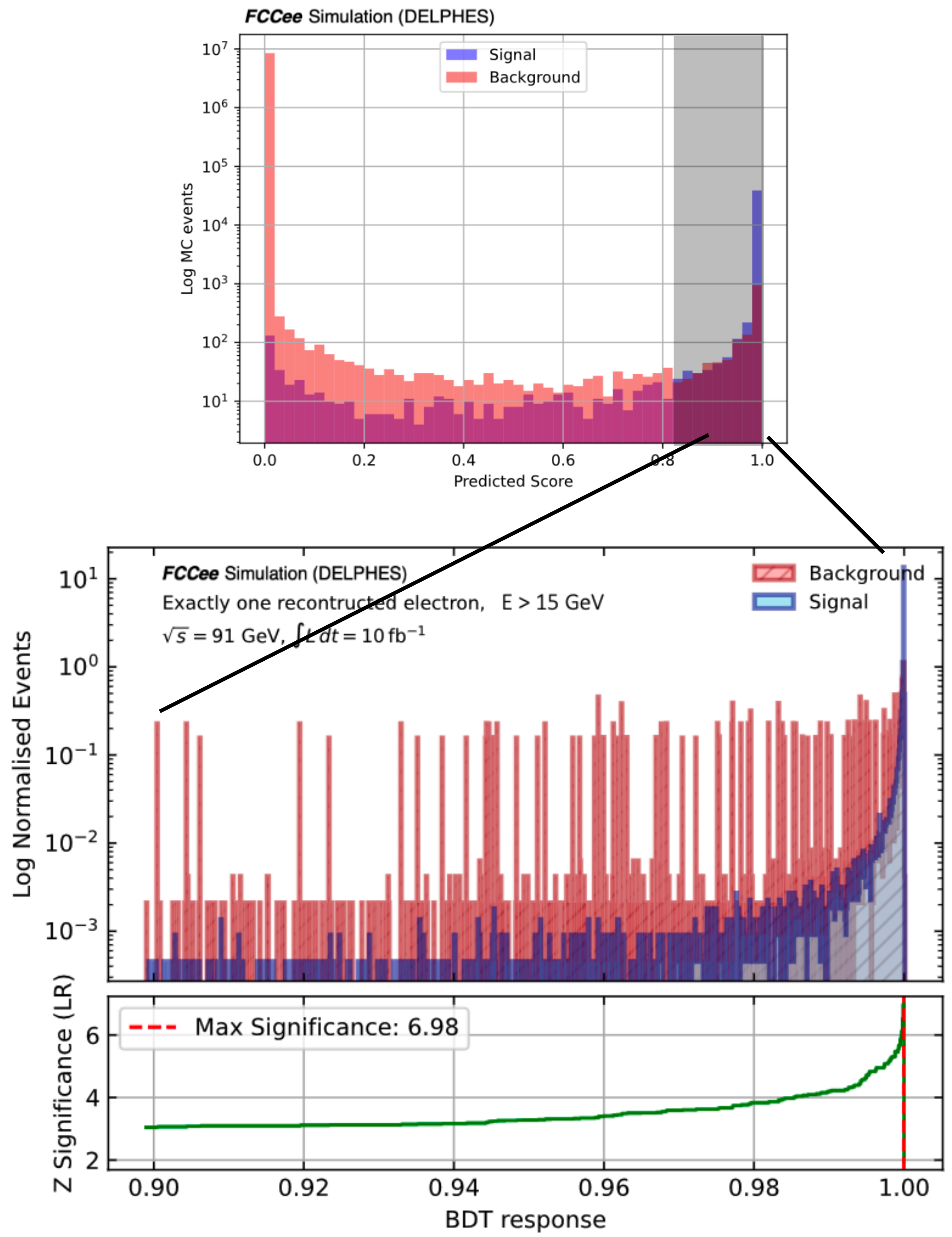
Data Preparation
- Stage 1 flat ntuples (FCC SW)
- Filter applied $E^{\text{leading electron}} > 15$ GeV
- Training and Testing split

Model Training
- GridSearch CV
- Decision trees made

Model Predictions
- Normalization to 10 fb^{-1}
- BDT cut chosen based on optimal significance

Process	Training Events	Testing Events
Total Background	5,655,708	11,311,415
20 GeV, $\|U^2\| = 10^{-6}$	26,254	26,254
50 GeV, $\|U^2\| = 10^{-6}$	29,991	29,991
70 GeV, $\|U^2\| = 10^{-6}$	32,194	32,193





DNN Workflow

Data Preparation
- Stage 1 flat ntuples (FCC SW)
- Filter applied $E^{\text{leading electron}} > 20$ GeV
- Training and Testing split
- Feature flattening

→

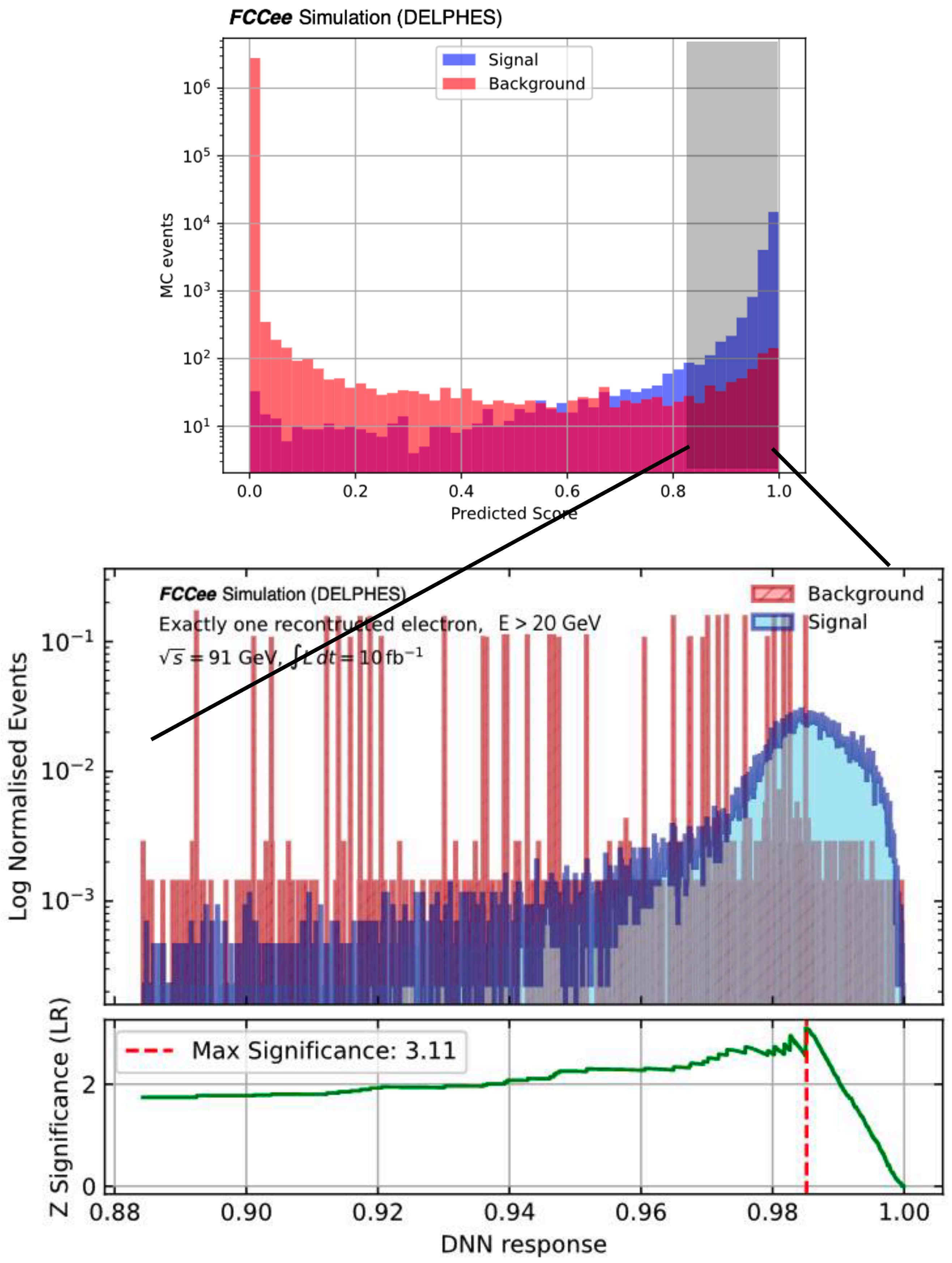
Model Training
- HP optimization using Random search
- Trained for 100 epochs

→

Model Predictions
- Normalization to 10 fb^{-1}
- DNN cut chosen based on optimal significance

Process	Training Events	Testing Events
Total Background	2,792,099	2,792,099
20 GeV, $\lvert U^2 \rvert = 10^{-6}$	19,601	19,600
50 GeV, $\lvert U^2 \rvert = 10^{-6}$	21,471	21,471
70 GeV, $\lvert U^2 \rvert = 10^{-6}$	23,951	23,951

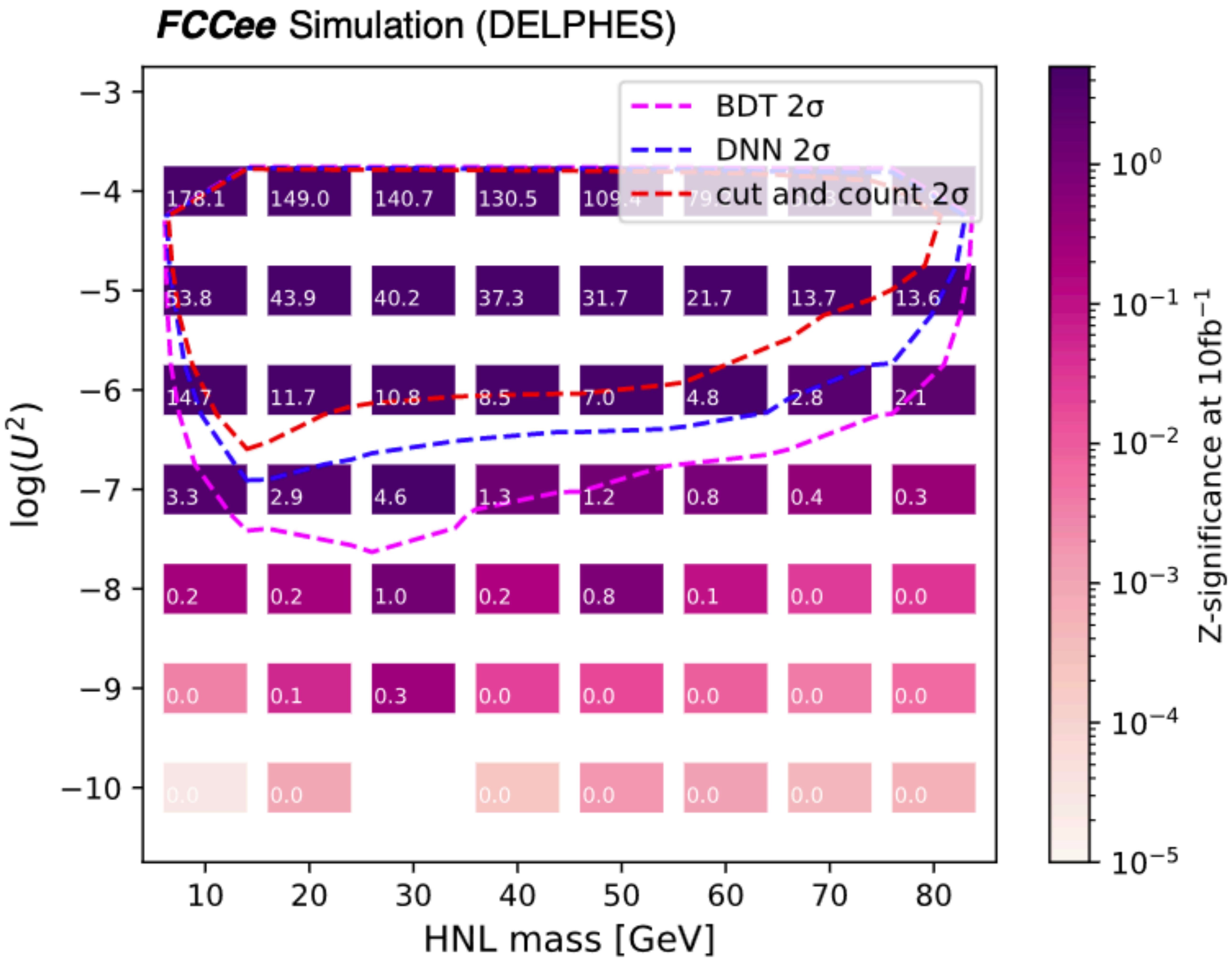
Hyperparameter	**Range**	**Step**
Units in Input Layer	32 to 512	32
Number of Hidden Layers	1 to 5	1
Units in Hidden Layers	32 to 512	32
Learning Rate	1×10^{-5} to 1×10^{-2}	Log scale
Dropout Rate	0.2	Fixed
Activation Function	ReLU	Fixed
Output Activation Function	Sigmoid	Fixed
Optimizer	Adam	Fixed
Loss Function	Binary Crossentropy	Fixed
Metrics	Accuracy, Precision, Recall, AUC	Fixed



Sensitivity Comparison

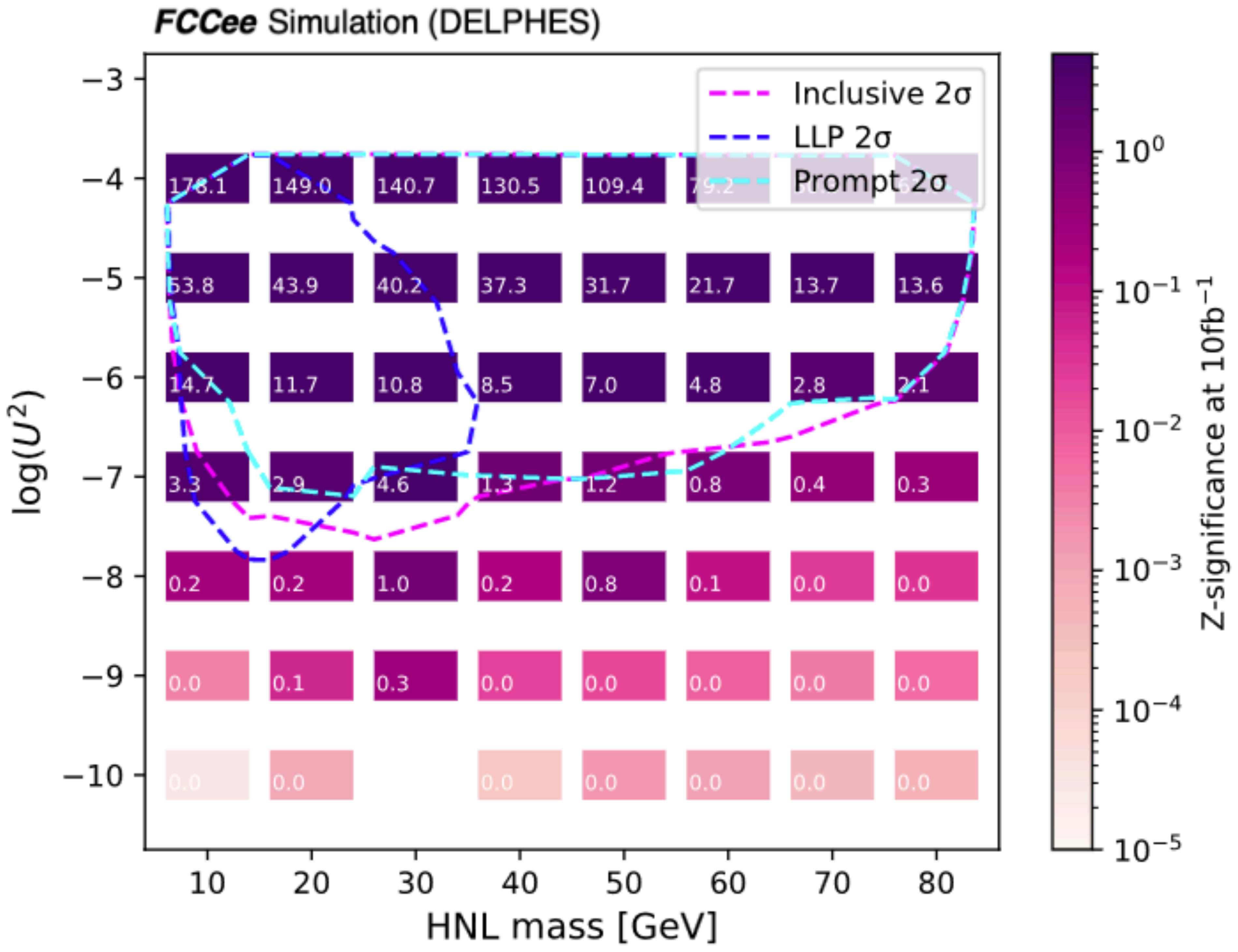
- **BDT model provides ~2 orders of magnitude more sensitivity compared to the C&C method and outpeforms DNN**

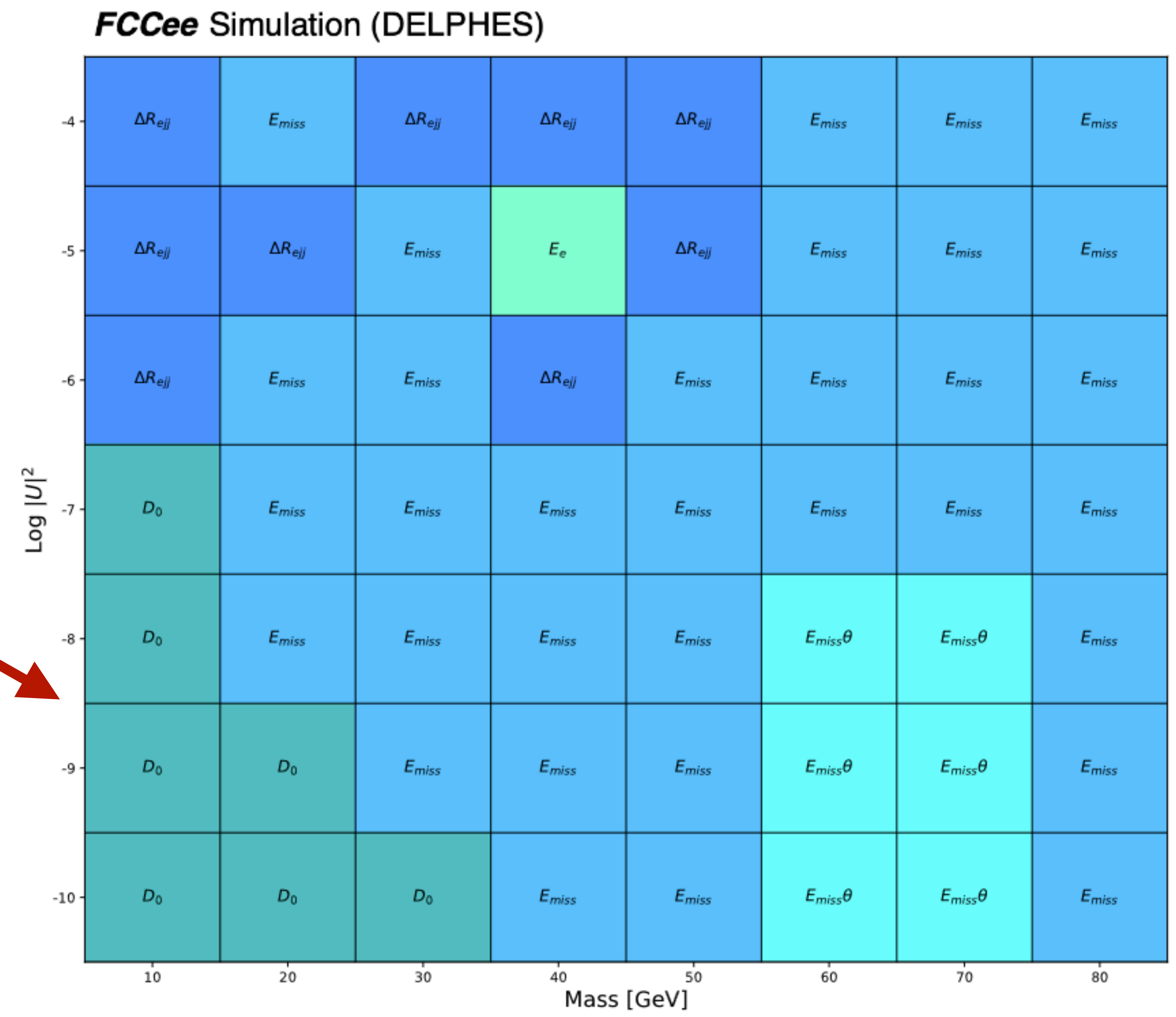
- **The (current) DNN approach offers ~1 order of magnitude improvement**
  - Hard to optimize but…
  - …implementing more sophisticated DNN architectures and robust hyperparameter optimization could in principle improve the performance



LLP study with BDTs

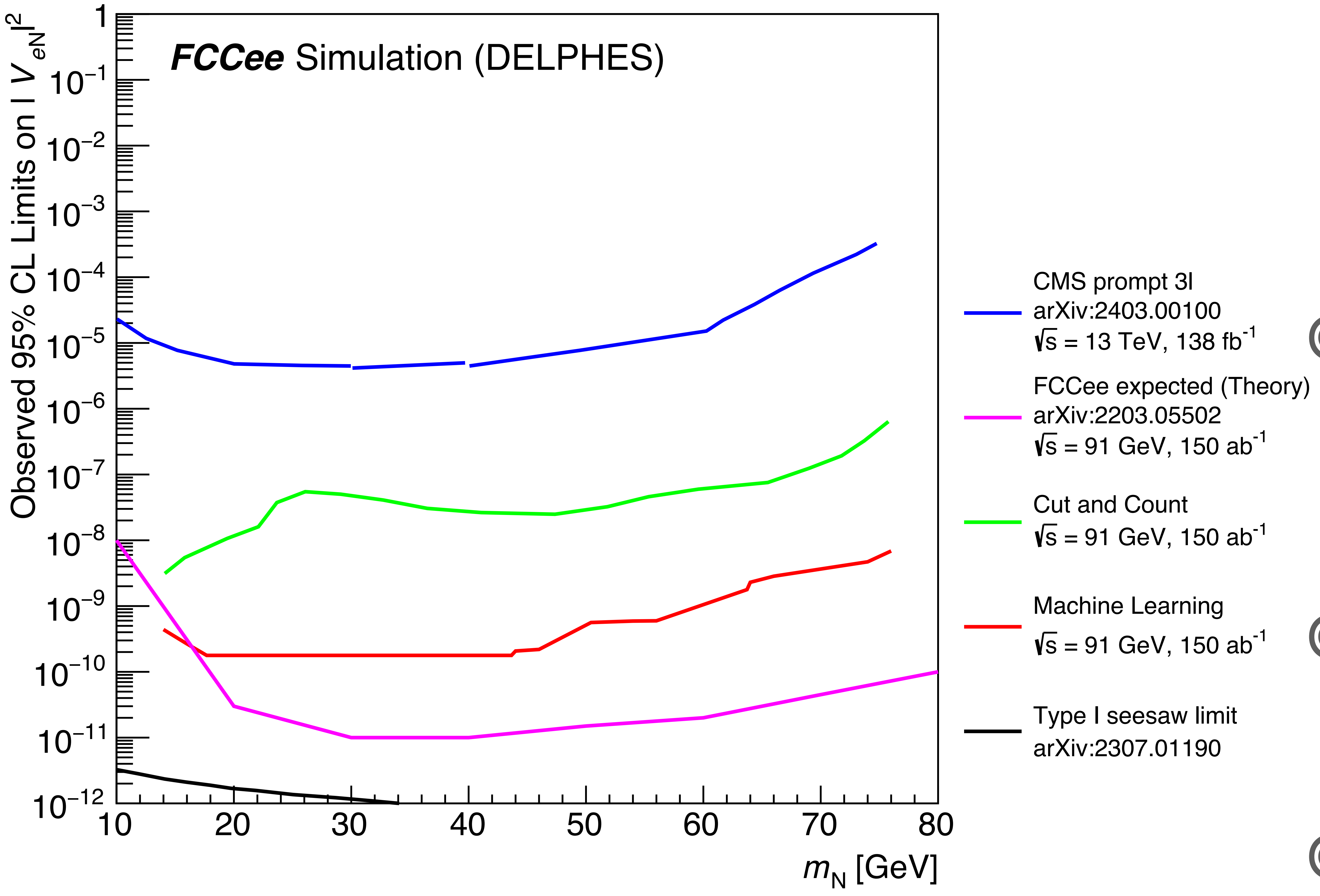
- Attempted to separate the signal using filter based on the IP significance
- Prompt decays are targeting using $\sigma_{d0}<5$ (cyan) and LLPs for $\sigma_{d0}>5$ (blue)
- BDT already utilizes d_0 as the most significant variable, thus explicit selections did not have additional impact





Summary

*CMS result only for 138 fb-1

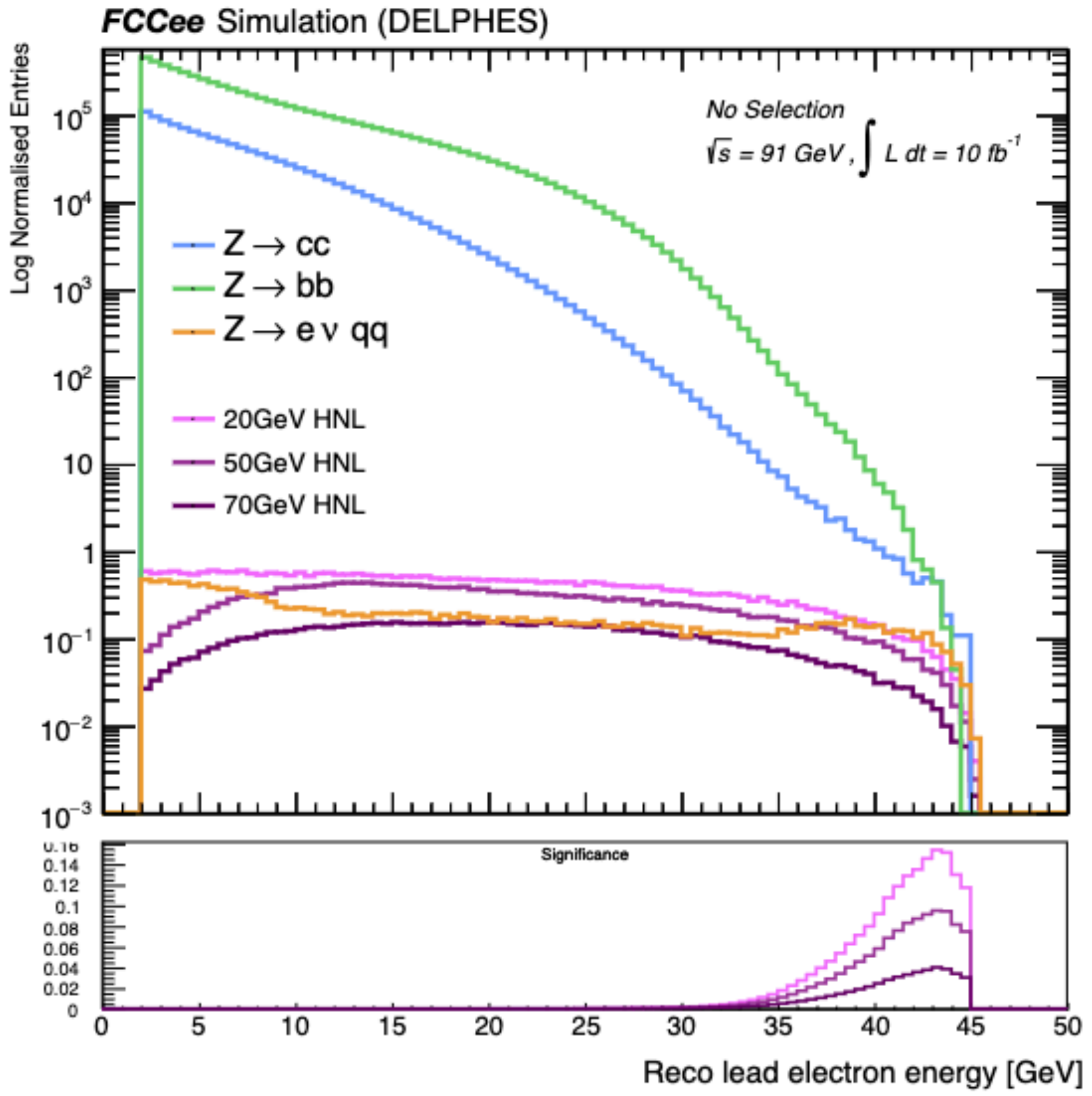


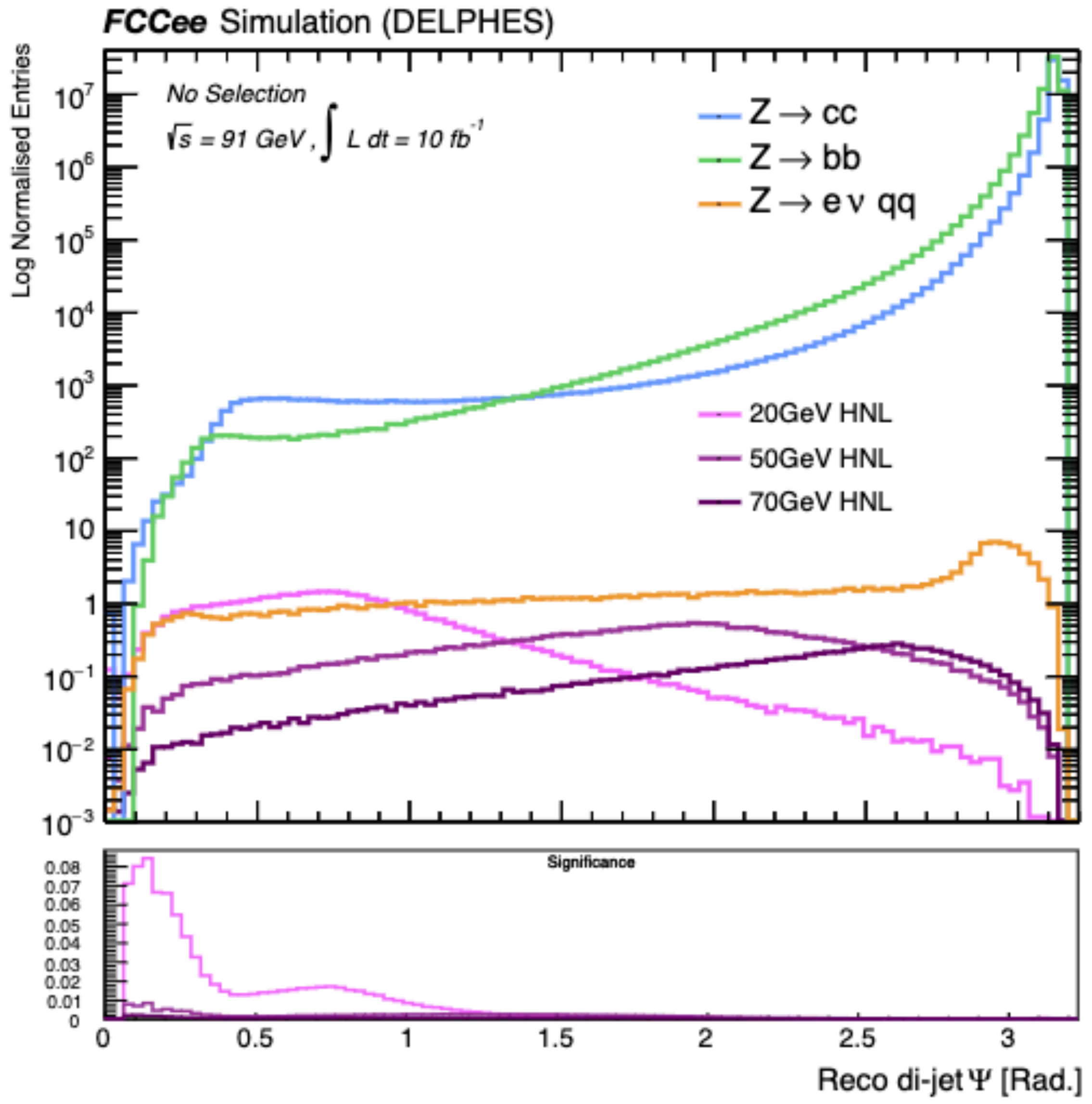
- Scaled to 150 ab^{-1} without accounting for uncertainties, the plot **shows broader phase space coverage compared to the C&C**
- Nearing FCC-ee limits with ~50% of the branching ratio; serves as a guide for improvement
- **ML shows strong potential to improve limits**
- **Increasing MC statistics** in the signal region is **essential** for robust analysis

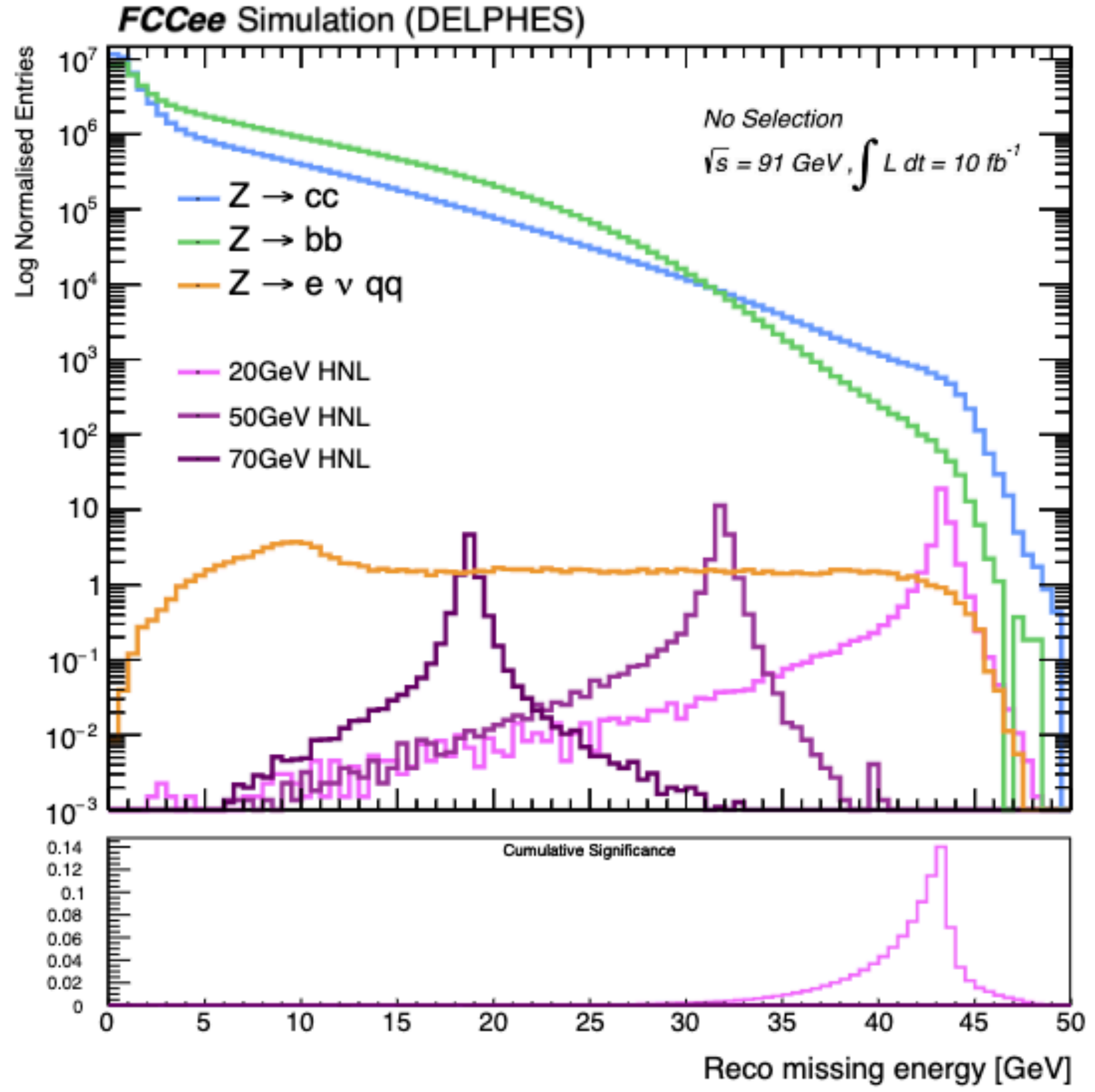
Thank you for your attention !

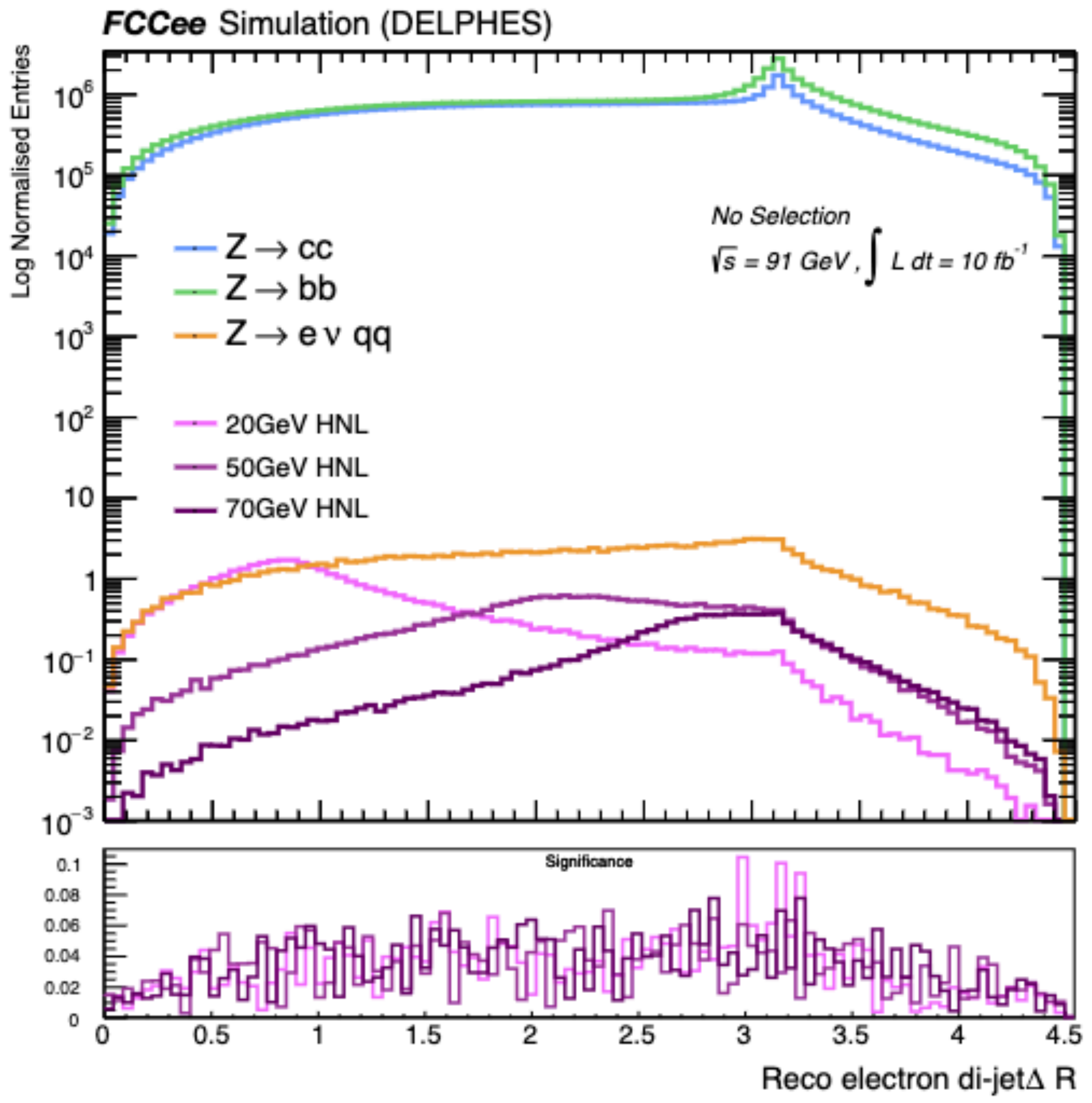
Backup Slides

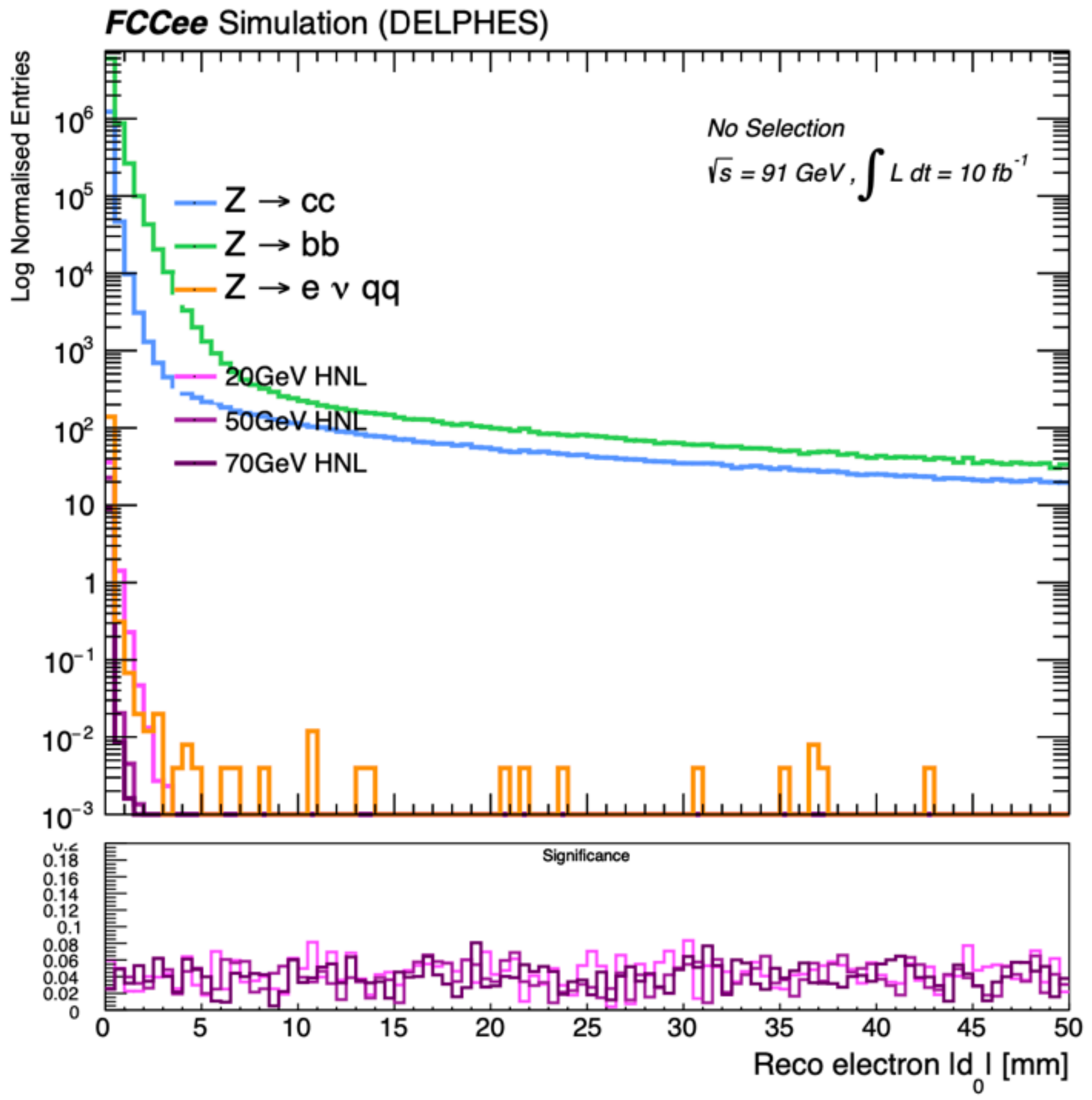
Variable distributions



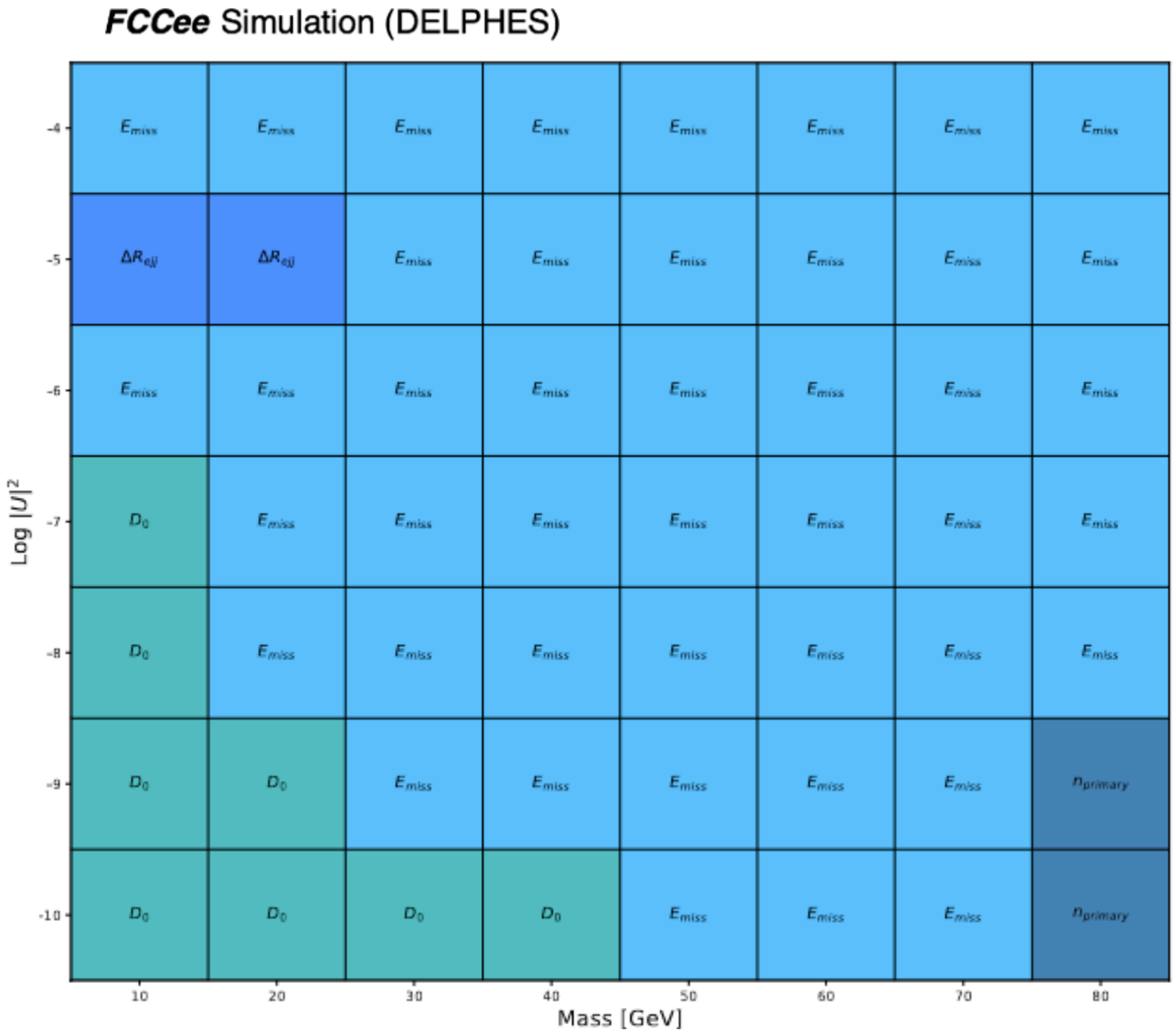


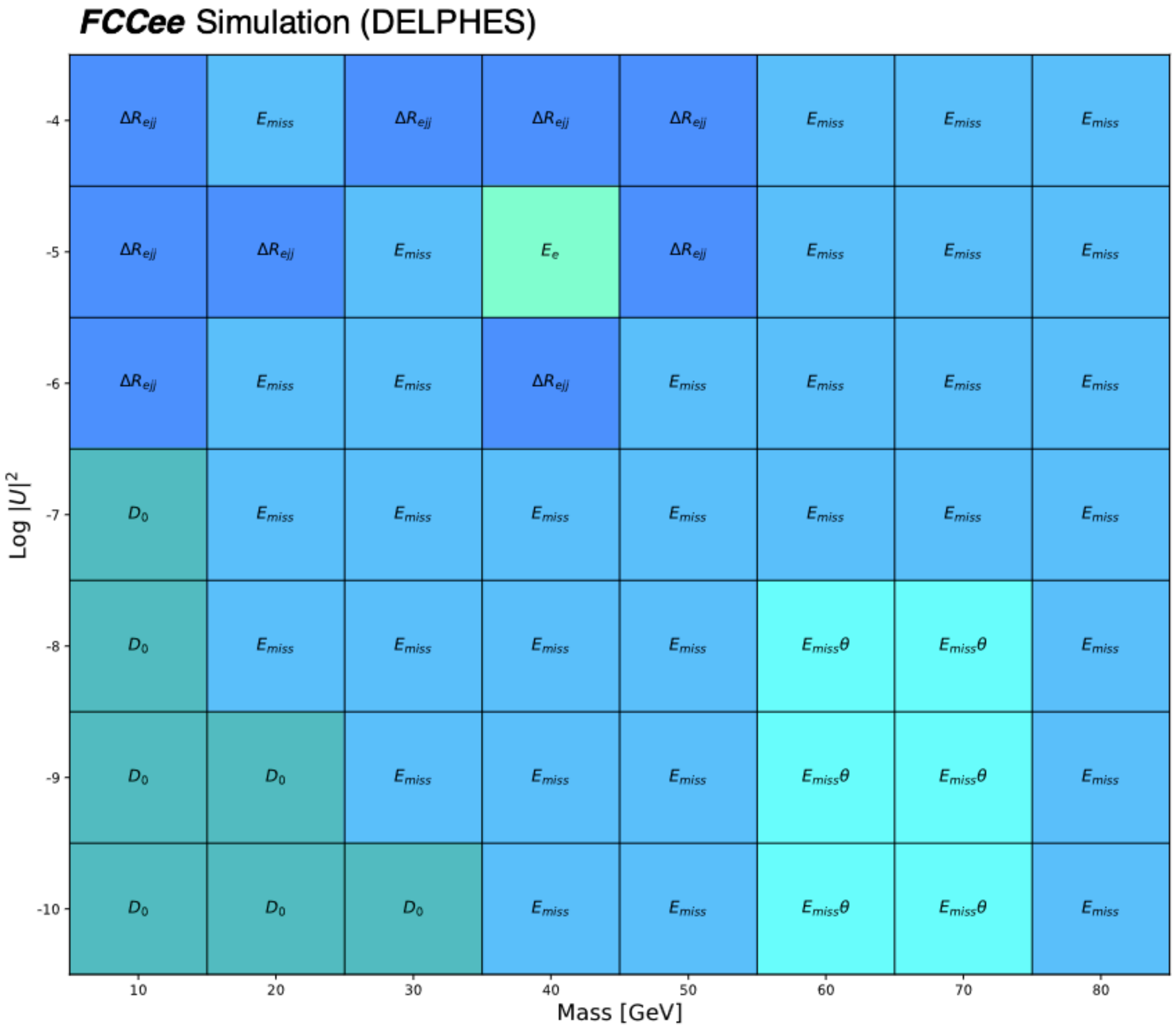






DNN vs BDT feature importance





Jet algorithms

- Jet reconstruction was primarily conducted using the FastJet software, rather the initial event generation phase with Pythia
- This approach was chosen for the enhances control and adaptability it provides when working directly with particle data from the EDMHEP files
- The **Durham jet algorithm** was used for the clustering jets