Presentation

Master Thesis Work

Federica Scarcella

November 23, 2017

Outline

1. ORCA
2. Toolkit for Multivariate Analysis
3. My Plots

ORCA

ORCA

Oscillation Research with Cosmic in the Abyss (ORCA):

$\Downarrow$

Measure of NMH with **Atmospheric Neutrinos** in the Mediterranean

$\swarrow$

- Produced by the interaction of primary cosmic rays (tipically protons) with nuclei in the atmosphere
- $\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}(\bar{\nu}_{\mu})$
- $\mu^{\pm} \rightarrow \bar{\nu}_{\mu}(\nu_{mu}) + \nu_{e}(\bar{\nu}_{e}) + e^{\pm}$

The detector is a array of multi-PMT digital optical modules (DOMs), 31 PMTs, that are distributed along flexible strings anchored to the sea bottom.

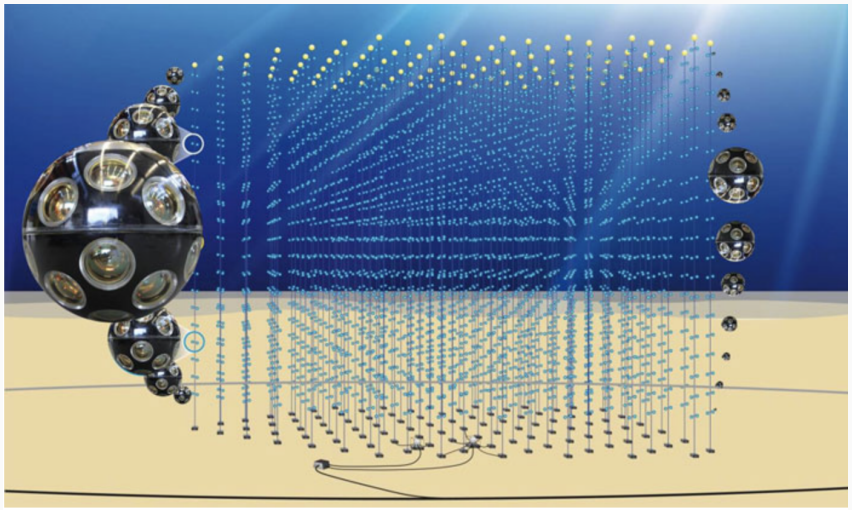


Figure 1: View of the detection string developed for the KM3NeT project.

The detection of neutrinos is based on the detection of Cherenhov light produced by relativistic particles emerging from a neutrino interaction.

Events Signatures

- **Charged Current neutrino interactions** :
  - $\nu_e + N \rightarrow e + X$: Shower-like
  - $\nu_\mu + N \rightarrow \mu + X$: Track-like
  - $\nu_\tau + N \rightarrow \tau + X$: Track-like
- **Neutral Current neutrino interactions**, where $x = e, \mu, \tau$:
  - $\nu_x + N \rightarrow \nu_x + X$: Shower -like
- **Background**: *Atmospheric muons* (Events induced by the passage of downward-going muon bundles coming from a cosmic ray air shower) misreconstructed as upgoing.

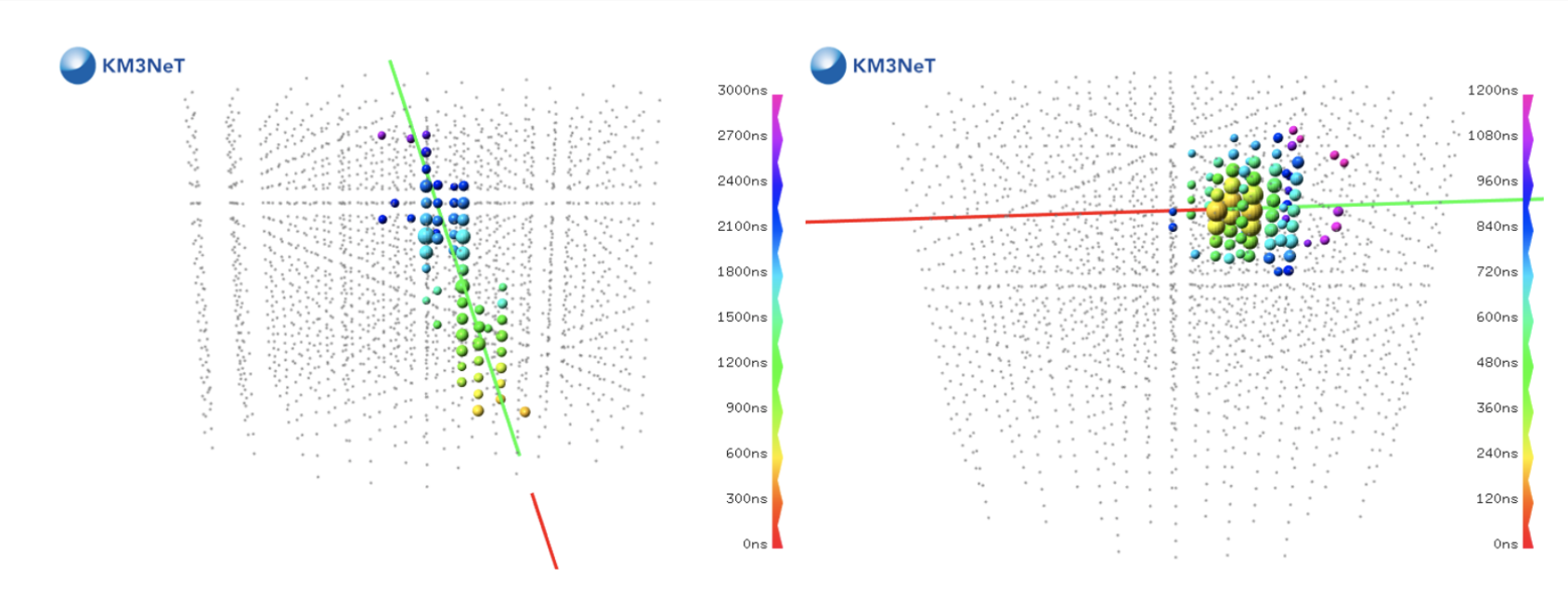


Figure 2: Event Signature: Track-like event (left), Shower-like event (right)

Toolkit for Multivariate Analysis

TMVA

A common problem in the High-Energy Physics is obtain a better separation between **Signal** and **Background**. For this aim, multivariate classification methods based on machine learning techniques, are used.

The *Toolkit for Multivariate Analysis* (TMVA) is integrated into ROOT. It includes many algorithms (BDT, Likelihood, ...) and consists of two main steps:

- Training/Testing Phase
- Application Phase

Decision Trees

A Decision Tree is a machine learning algorithm mostly used in classification problems. The goal is split the data, according to certain parameters, into two or more sub-nodes. The decision tree consists of two entities:

- Decision nodes
- Terminal Nodes: *Leaves*

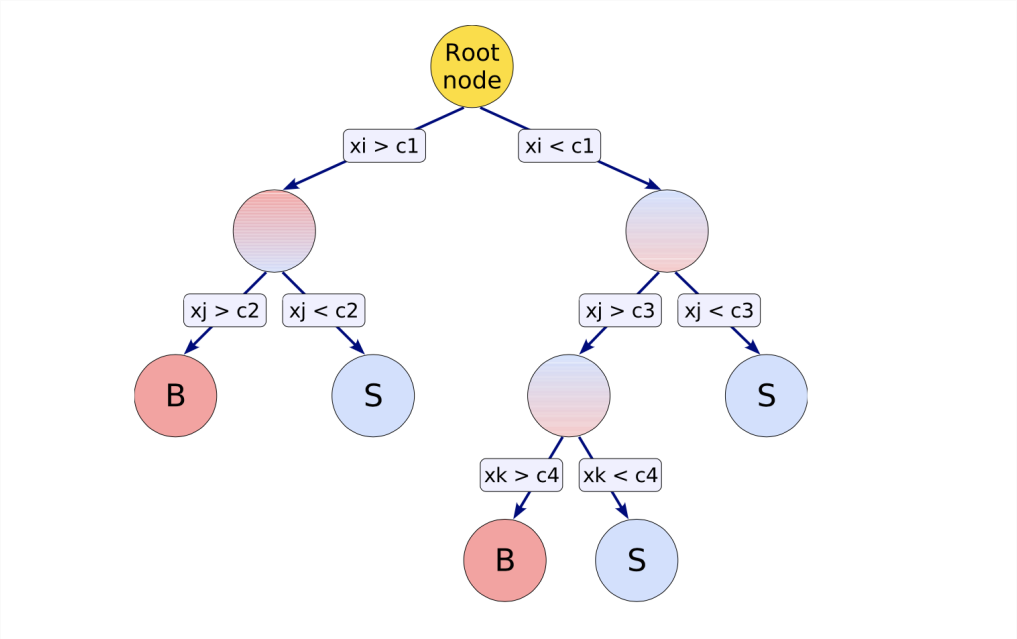


Figure 3: Schematic view of a decision tree.

Decision Trees

Decision trees have several advantages:

- Simple to understand and to interpreter
- No information is lost

Disadvantages of decision trees:

- Decision trees are not stable $=>$ Boosting!

Boosting means that each tree is dependent on prior trees and learns by fitting the residual of the trees that preceded it. It stabilizes the reponse of the decision trees with respect to fluctuation in the training sample.
$\Longrightarrow$ **BOOSTED DECISION TREES!!**

ORCA's Analysis

Track-Shower separation $\Rightarrow$ **Random Decision Forest** (RDF).

The process is repeated until the event reaches a *leaf* and it is classified as a *track*, a *shower* or an *atmospheric muon*.

The reconstruction strategies can produce a rejection of the atmospheric background without significantly reducing the amount of good neutrino events.

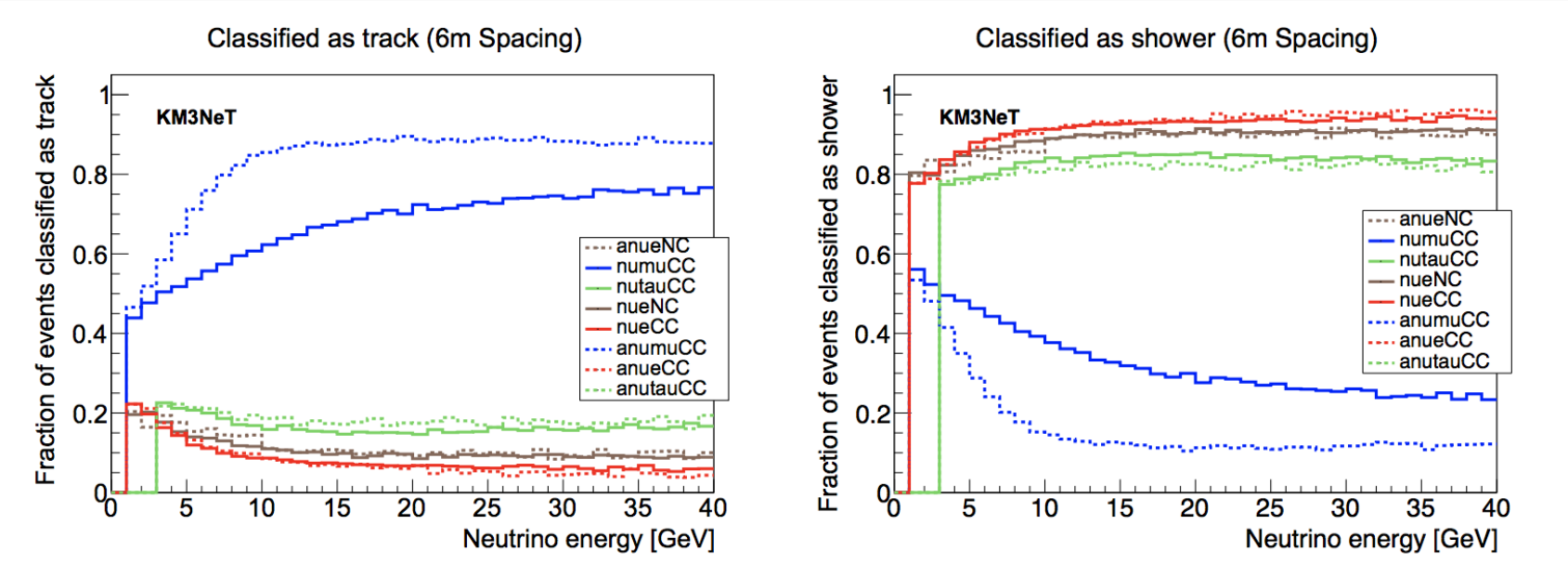
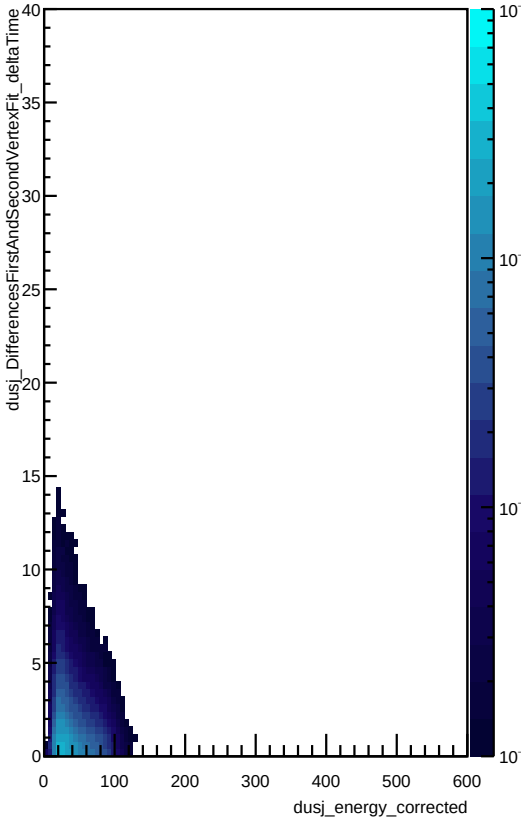


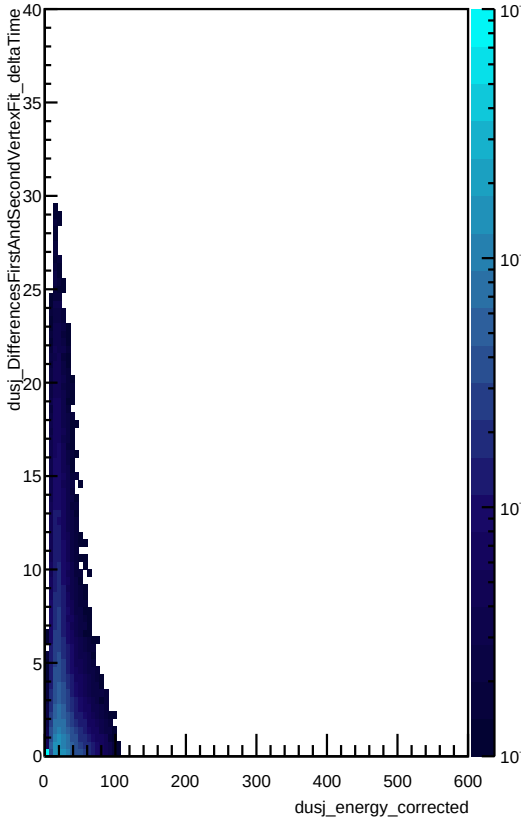
Figure 4: Fraction of events classified as tracks (left) or shower (right) for a detector with 6 m vertical DOM spacing.

My Plots

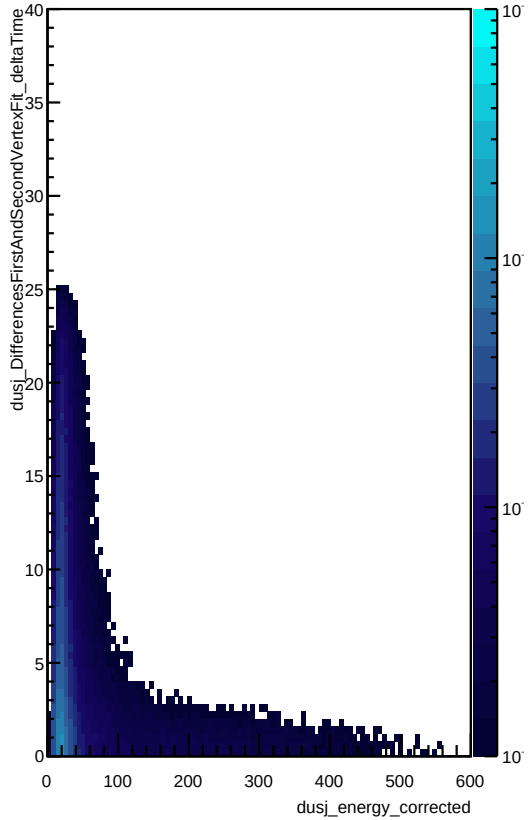
is_cc&&(type==12||type==-12)&&energy>20



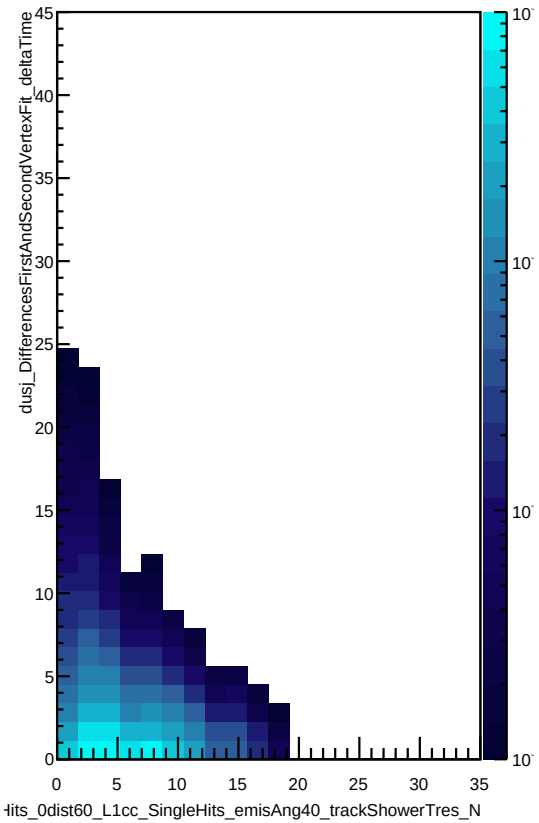
is_cc&&(type==14||type==-14)&&energy>20



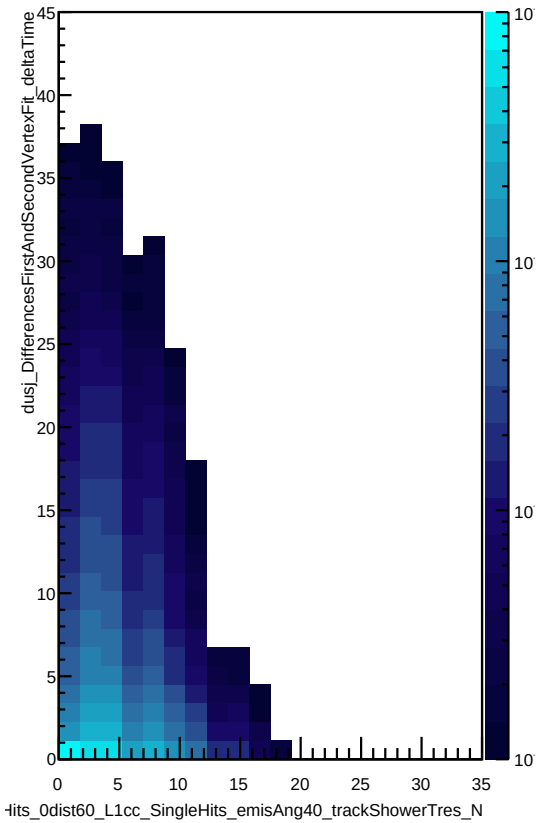
is_cc&&type==5



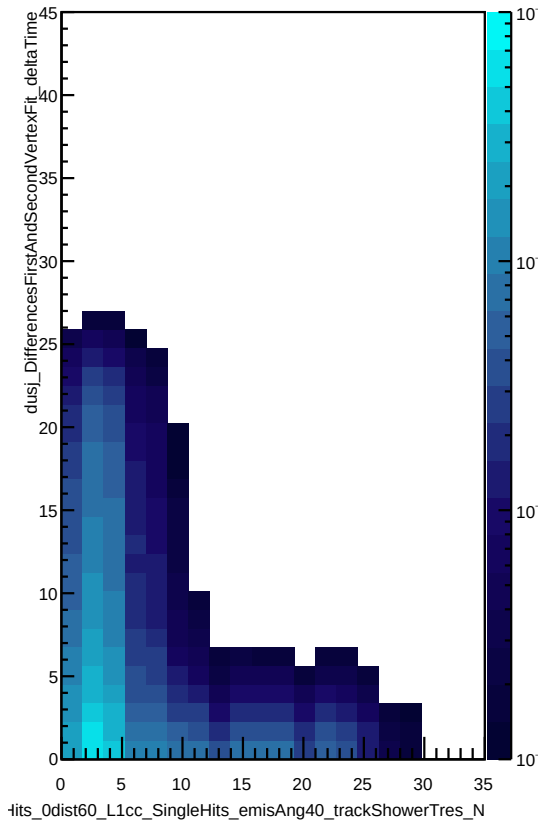
is_cc&&(type==12||type==-12)&&energy>20



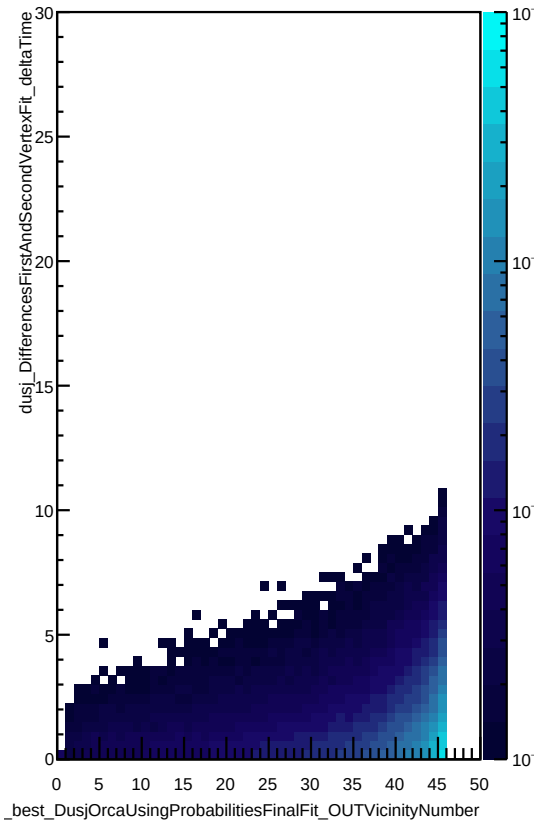
is_cc&&(type==14||type==-14)&&energy>20



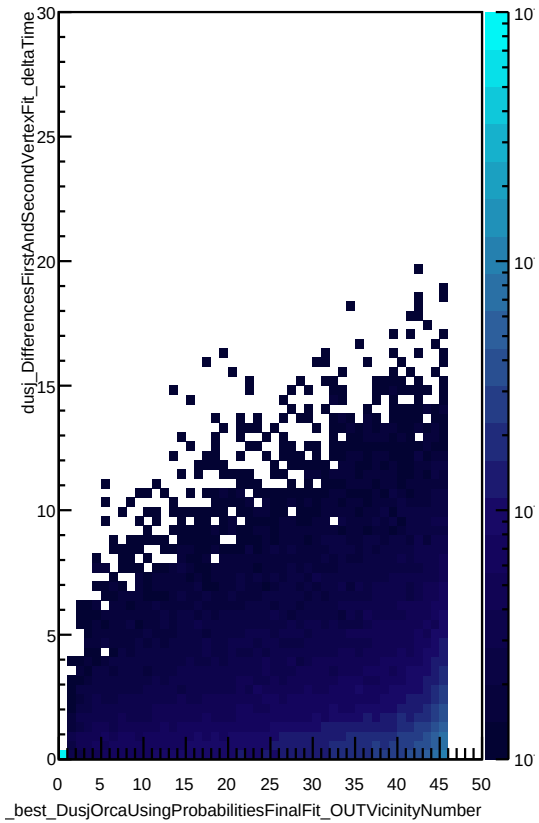
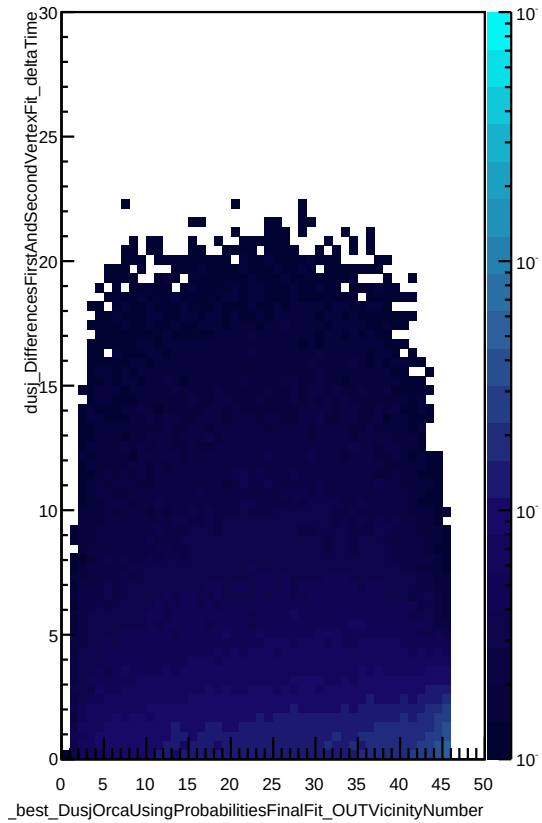
is_cc&&type==5



is_cc&&(type==12||type==-12)&&energy>20



is_cc&&(type==14||type==-14)&&energy>20



is_cc&&type==5