So, you want to do Machine Learning on high-energy physics?

A KM3net state of the art review

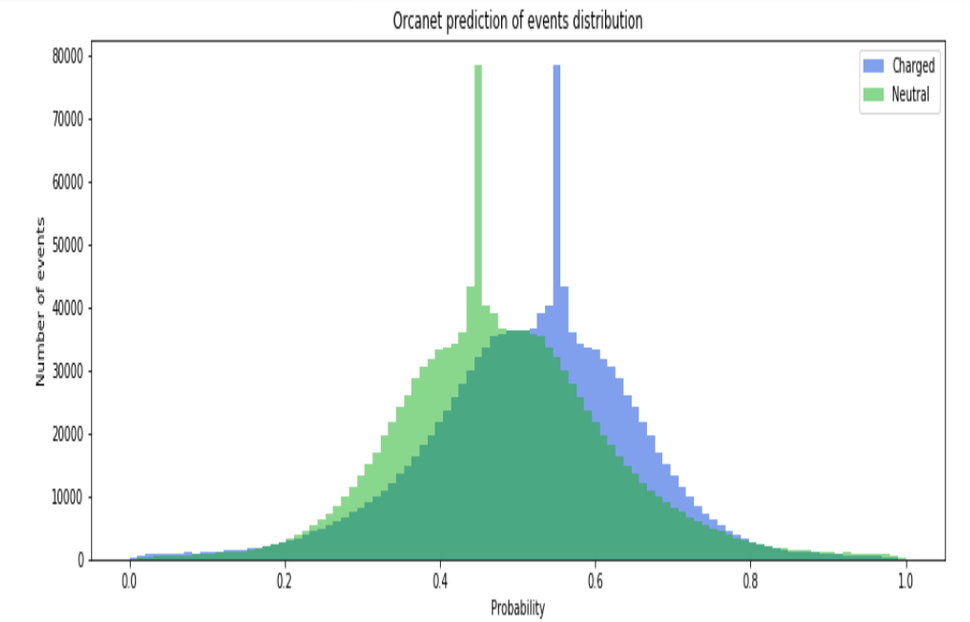
No.

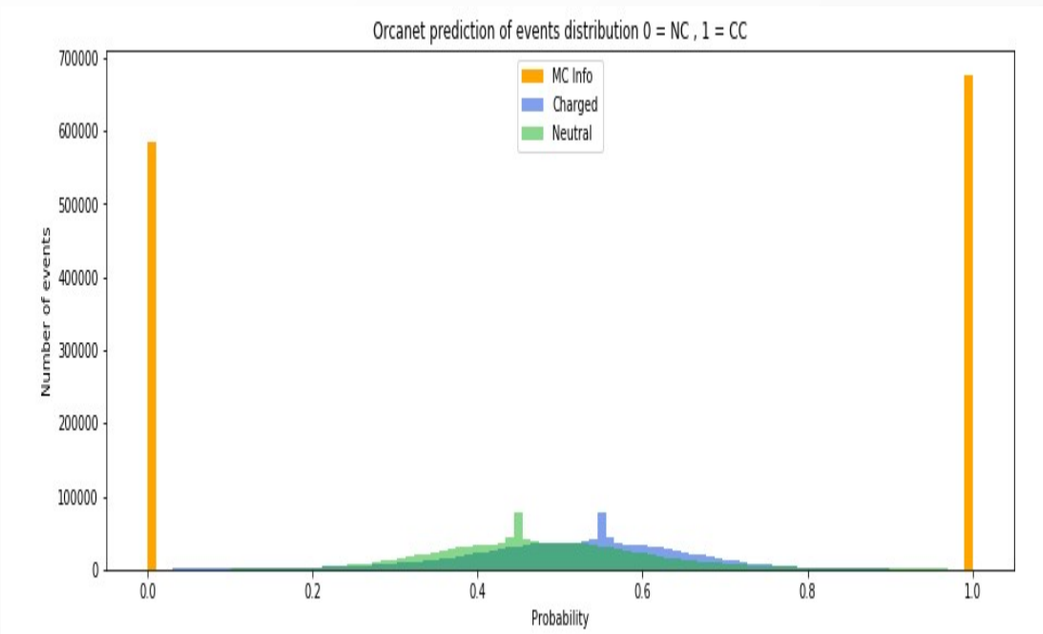
1) I don’t believe in a machine giving me answers without knowing how it gets them.

2) Nothing is better than RooFit.

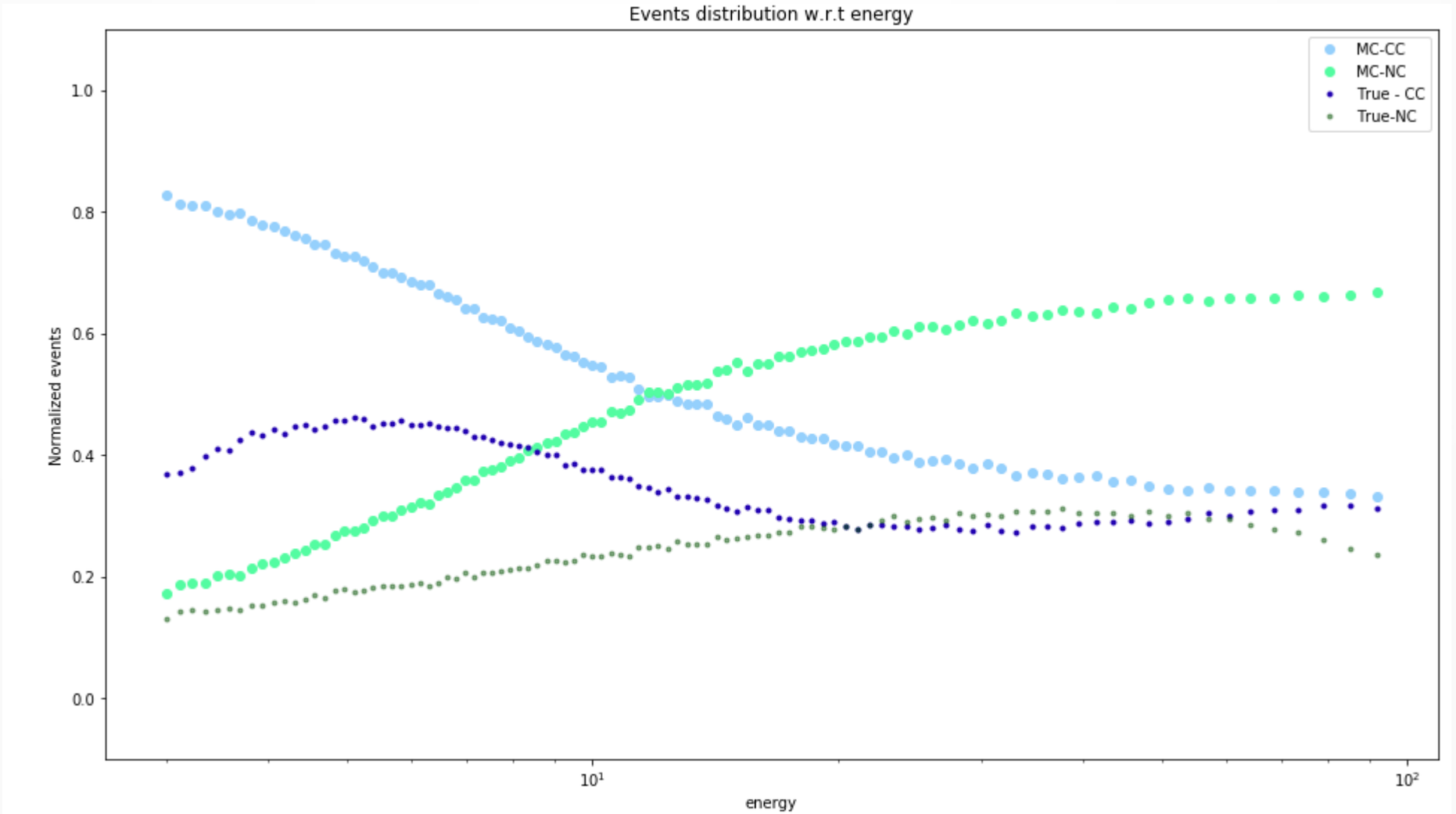
(The computer will not learn for us, but we can learn from the computer.)

My first results

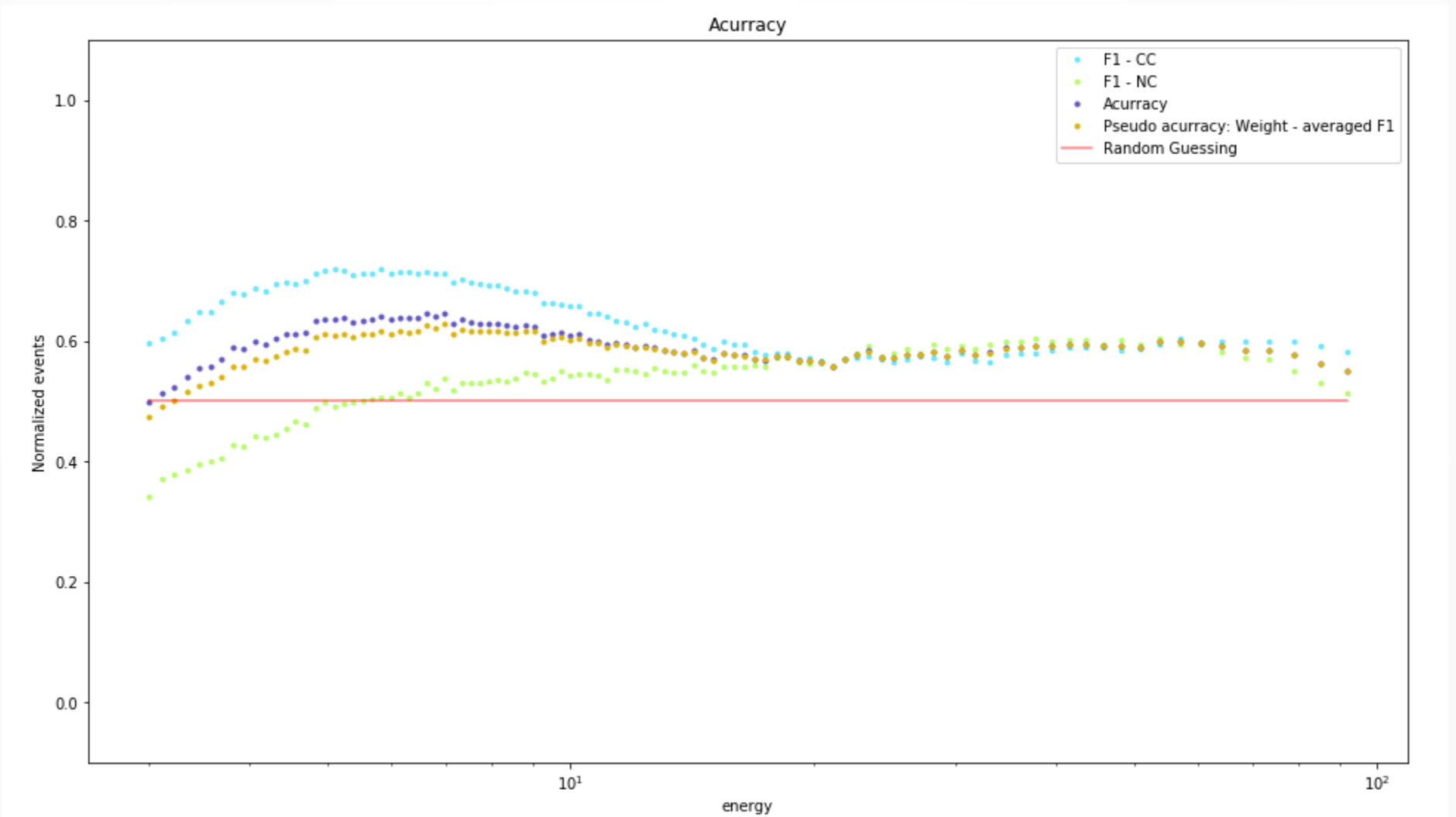




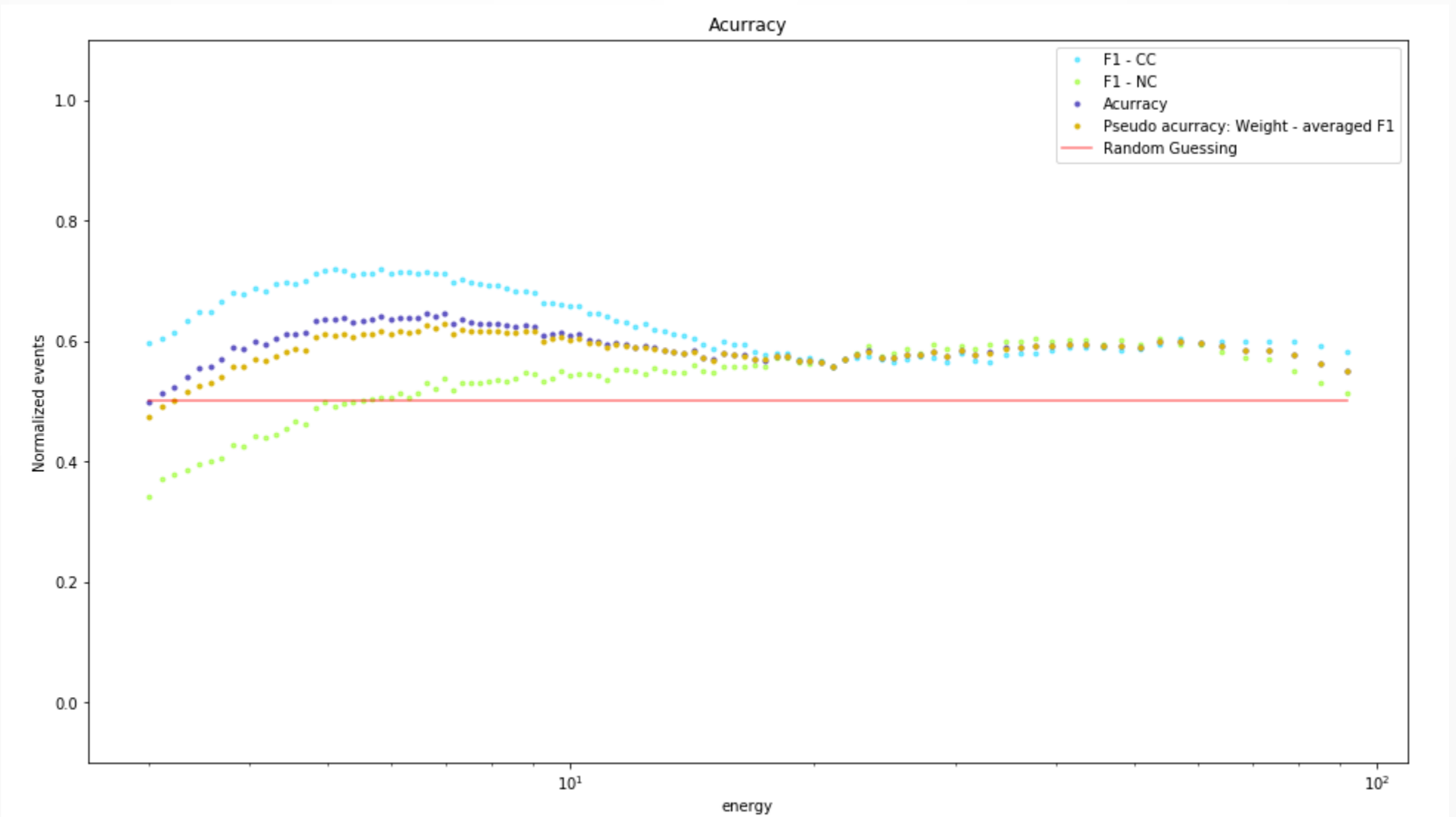
My first results



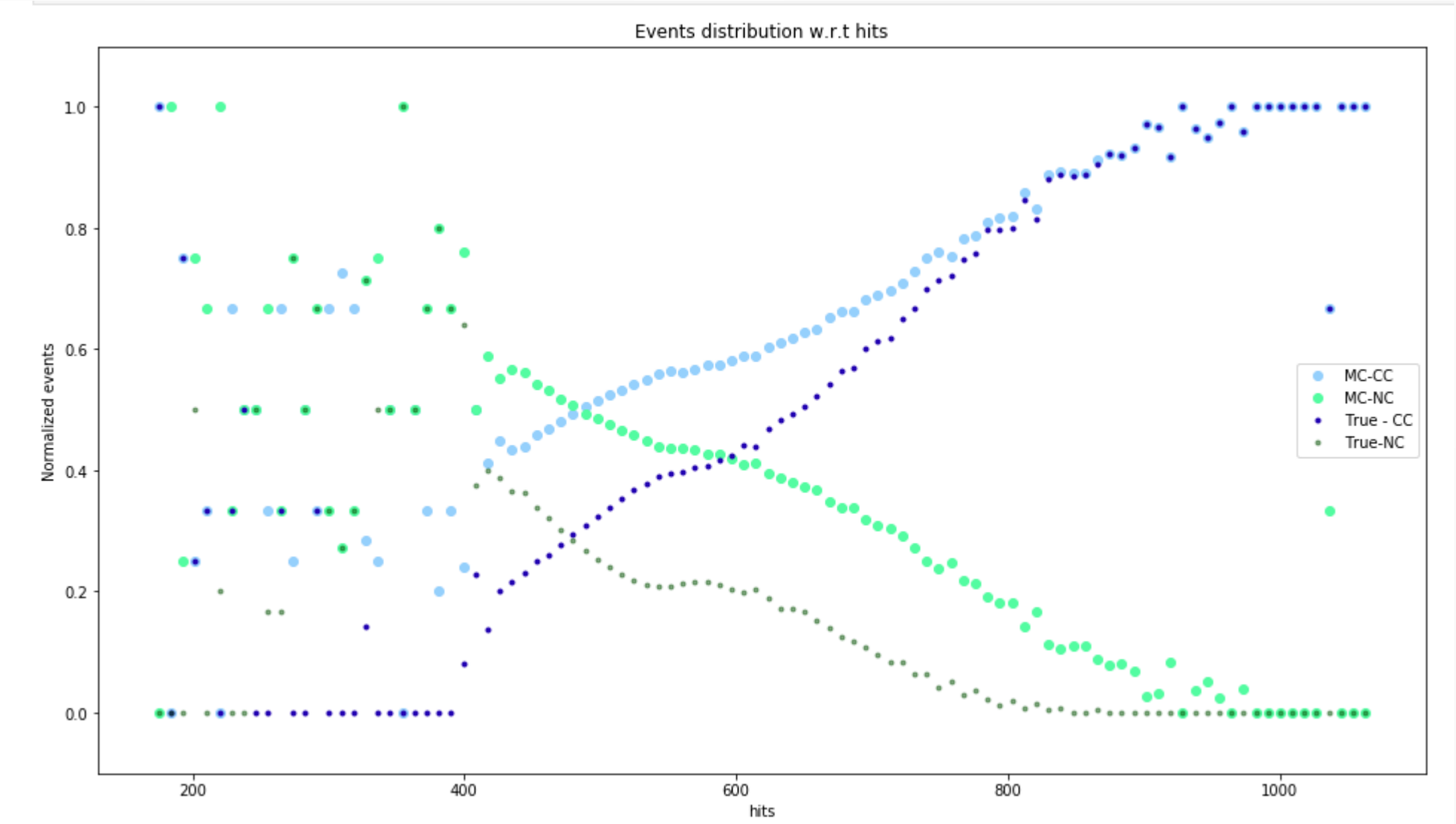
My first results



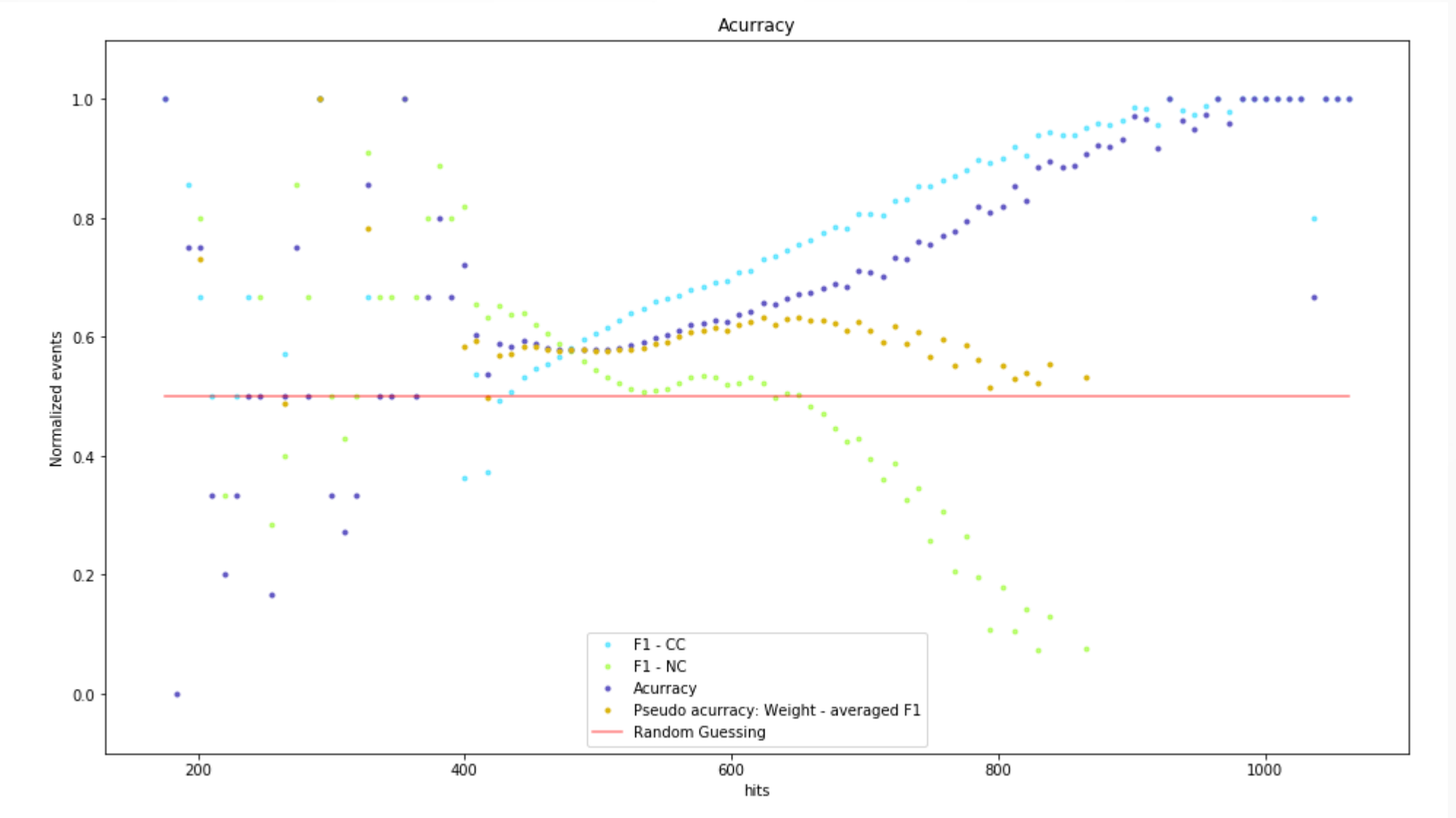
My first results



My first results



My first results



My first conclusions

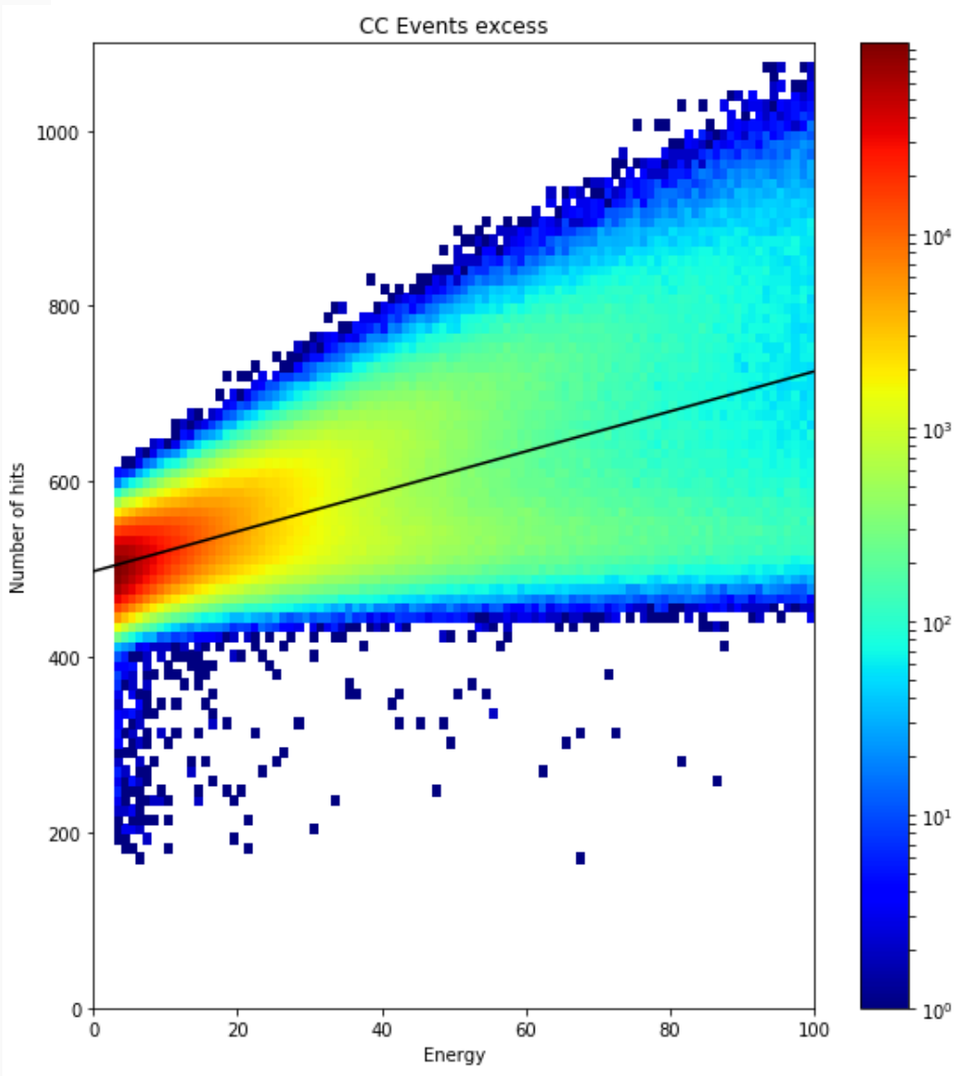
1) It’s clearly just counting the number of hits (visible energy) per event, and getting very good at just that.

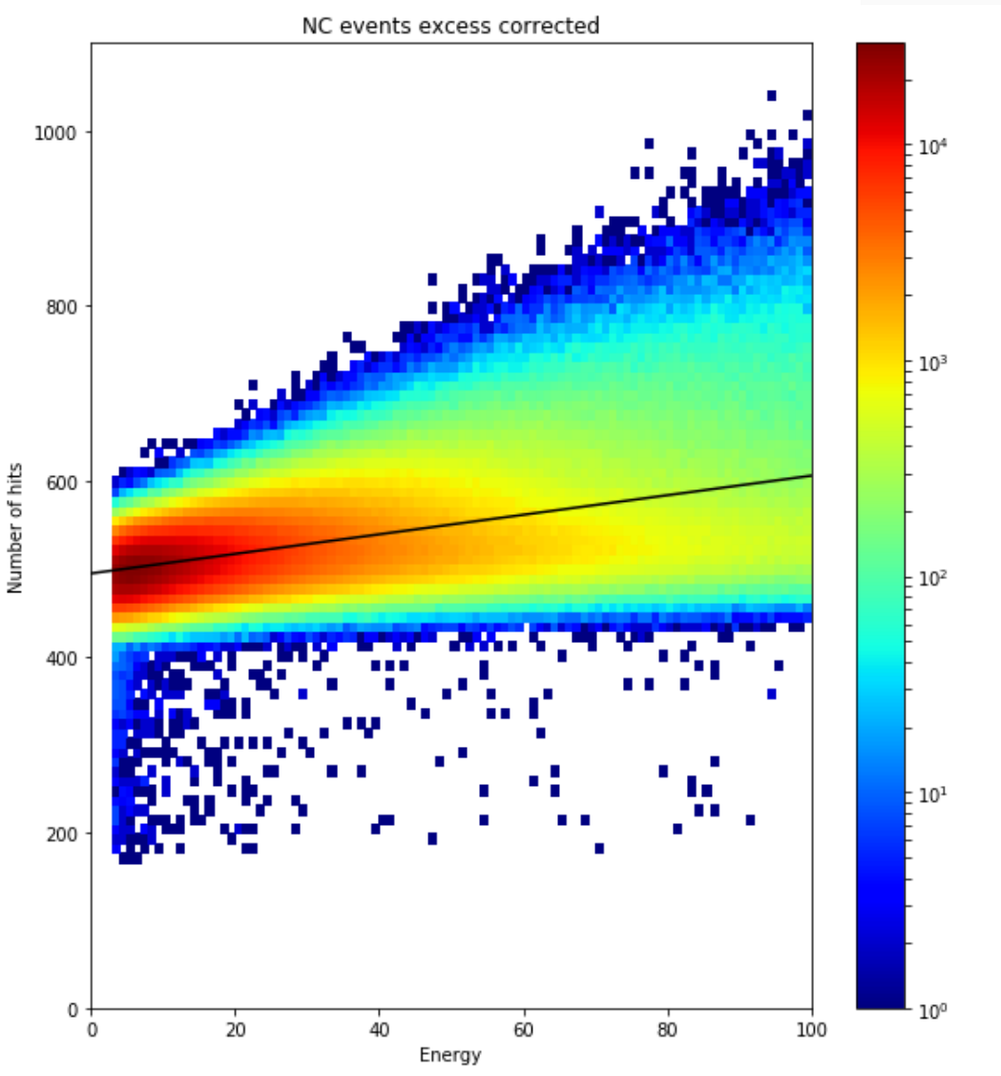
2) It is paying a lot of attention to the extremes: The (marginal) low hit events and the lack of high energy – CC events

3) My data is not equivalent!

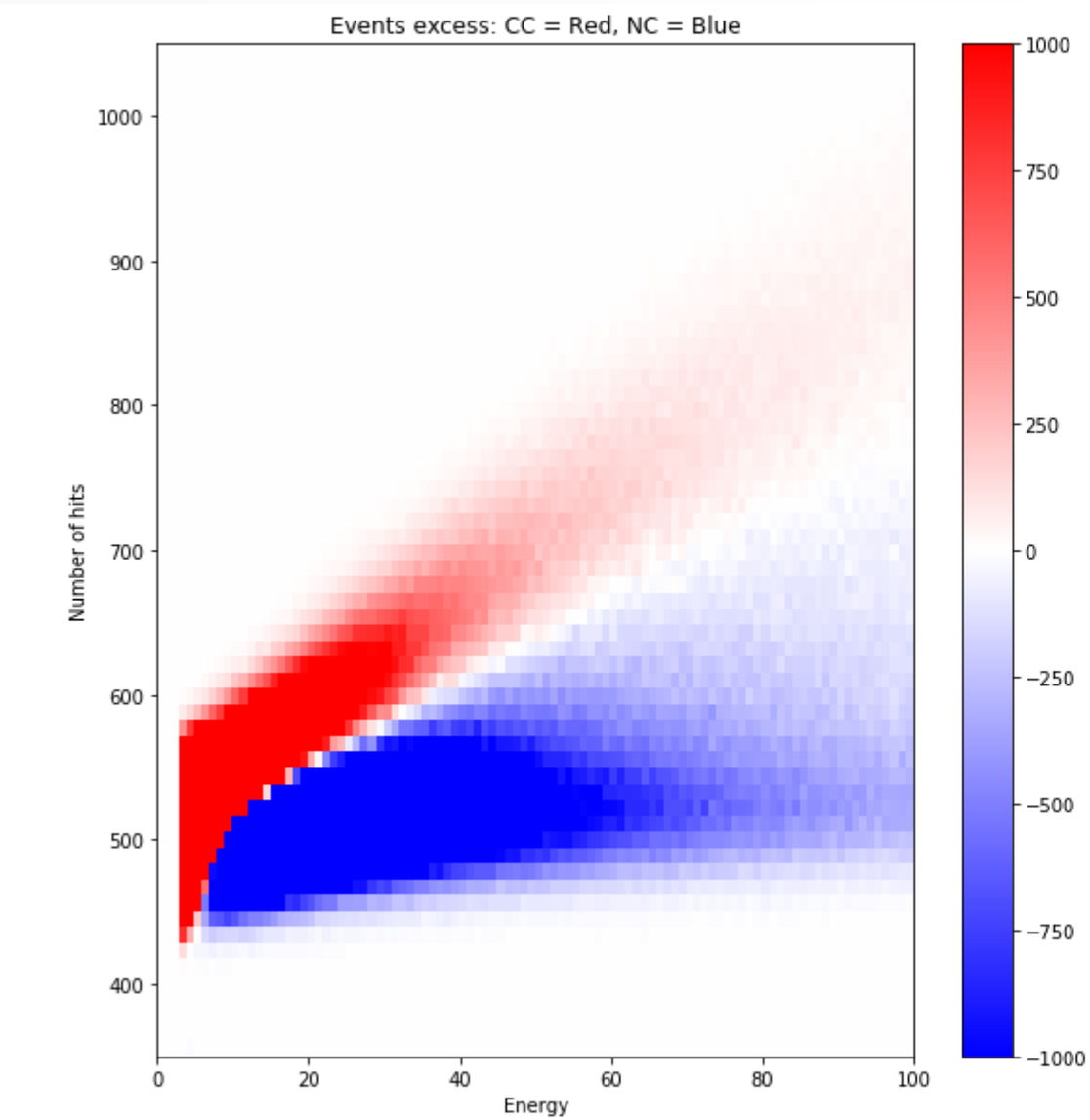
4) Also, acurracy is not a well-behaved metric.

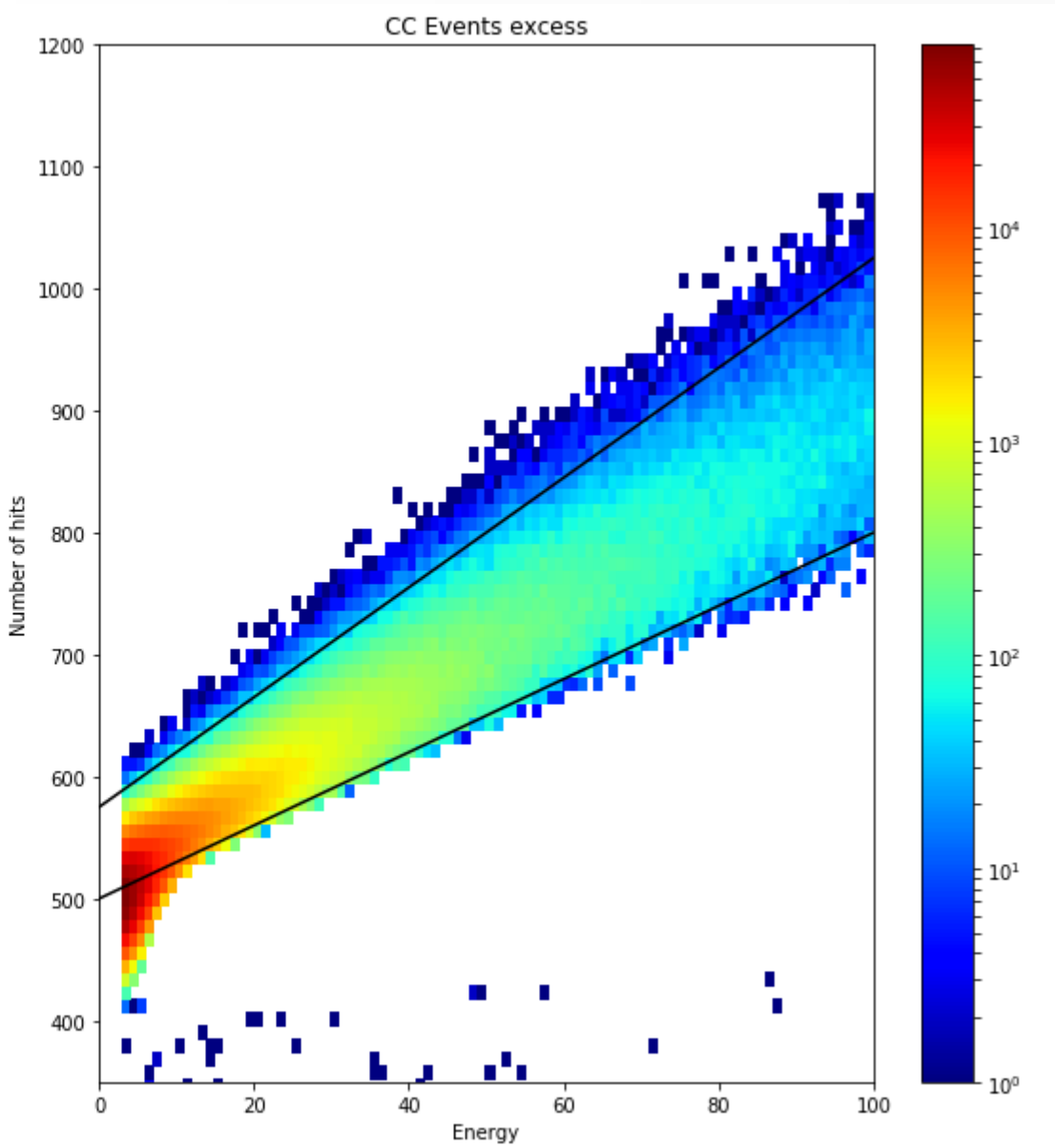
A look at the data



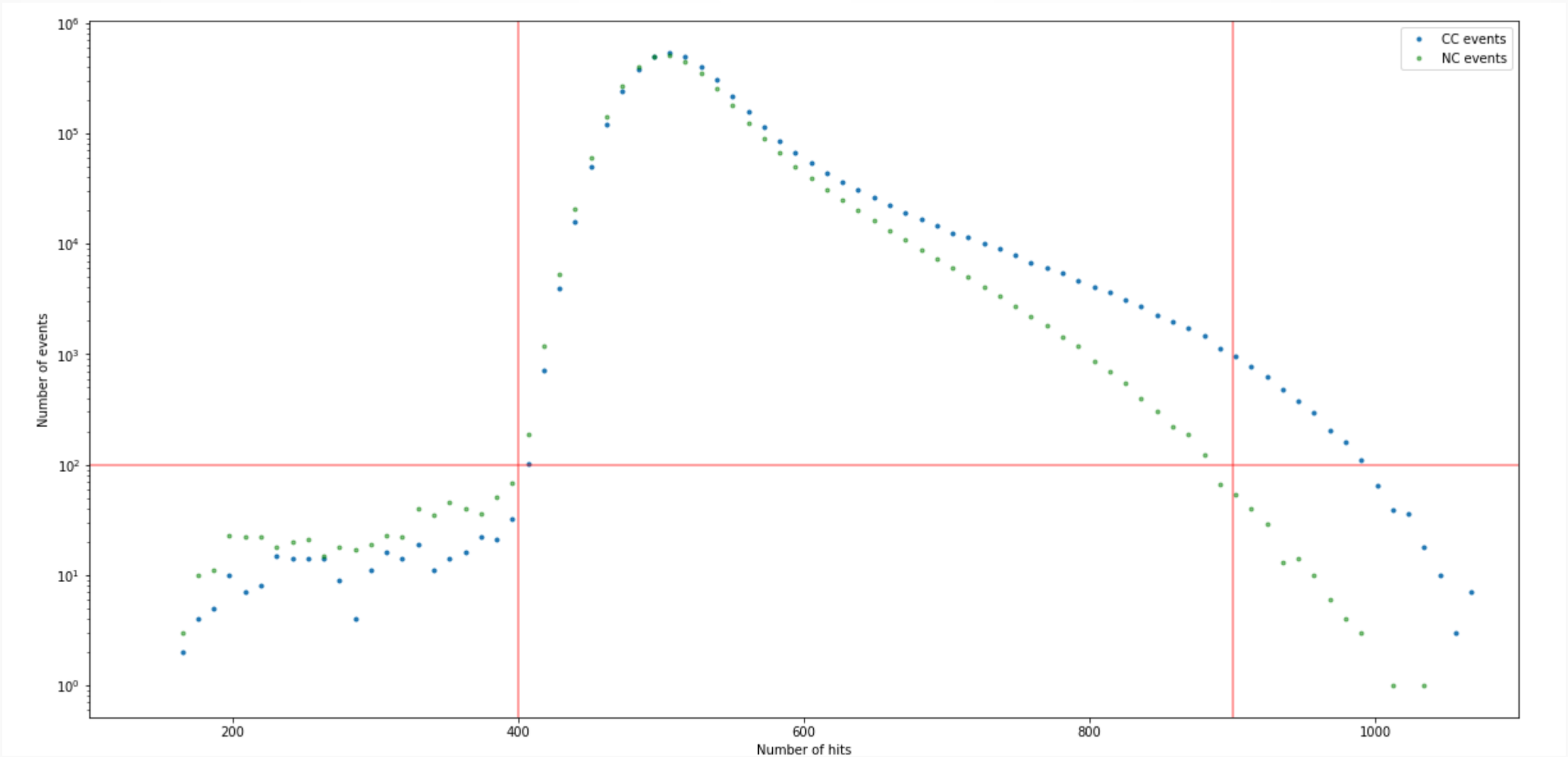


A look at the data





A look at the data



Current steps:

Run corrected (cropped, with flattened spectrum) datasets to see if we can archive some (good) performance without any hidden features.

This requires additions to my preparation process that I’m currently dealing with.

How we get to do DL with ROOT data.

- All the preprocessing steps compose what I call a pipeline (à la KM3Pipe).
  - Any DL analysis needs an appropiate pipeline.
  - (Reco) analysis can become modular and streamlined.
  - Ideally, a pipeline can be written in terms of the existing tools (KM3pipe and OrcaNet).
- Pipeline: Retrieval, format conversion, selection, image production, (architecture selection).

Moser's 5 steps pipeline

1) Lyon: Get triggered data from IRODS.
2) Convert to h5 with KM3Pipe's to_hdf5 (OrcaFin).
2.b) Calibrate with detector file.
3) Orcasong: Make histograms out of the events.
   - Dimension, timecuts, ...
4) Concatenate + Shuffle.
5) OrcaNet (Actual deep learning happens here)