# FTX ANA Meeting

## Status Update

Konrad Helms

21st September 2023

# Checks for Data Leakage - Truth Information:

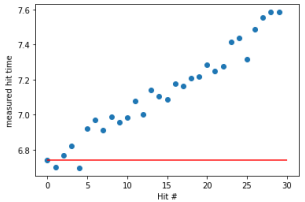
```python
#cut 3:
# shower quality cut, i.e. fraction of deposited energy from the MC
# truth in the shower should be larger than some threshold, here 0.9
data = data[data['clean shower'] >= 0.9]
```

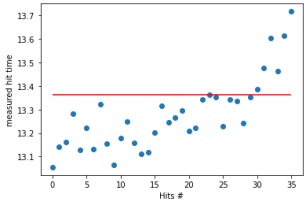
clean shower is calculated by

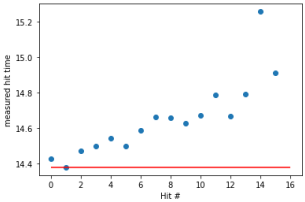
```
auto isCleanShower = [&navEcalHit](ReconstructedParticle* pfo, MCParticle* mcTrue) -> float {
        /* Clean shower == true particle contributes >= 90% energies */
        Cluster* cluster = pfo->getClusters()[0];
        float totalEnergy = 0.;
        float trueEnergy = 0.;
        for ( auto hit : cluster->getCalorimeterHits() ){
            const std::vector<LCObject*>& simHits = navEcalHit.getRelatedToObjects(hit);
            const std::vector<float>& weights = navEcalHit.getRelatedToWeights(hit);
            // it should really be always 1, but just in case
            if ( simHits.size() == 0 ) continue;
            SimCalorimeterHit* simHit = static_cast<SimCalorimeterHit*>( simHits[ std::max_element(weights.begin(), weights.end()) - weights.begin() ] );
            int nCont = simHit->getNMCContributions();
            for(int i = 0; i < nCont; i++){
                totalEnergy += simHit->getEnergyCont(i);
                if ( simHit->getParticleCont(i) == mcTrue ) trueEnergy += simHit->getEnergyCont(i);
            }
        }
        if (totalEnergy == 0.){
            return 0.;
        } else {
            return (trueEnergy/totalEnergy);
        }
    };
```

# Checks for Data Leakage - True TOF in Shower Data:

- find the true TOF exactly once in training, once in testing, once in validation data




- in 0.004% of pfos, the true TOF is equal to at one measured hit time

# Math:

hit times and the true tof are exported with precision:

```
ss<<std::scientific<<std::setprecision(6)<<CLHEP::RandGauss::shoot(hit->getTime(), 0.05)<<", "; // smeared 50 ps
ss<<std::scientific<<std::setprecision(6)<<tof<<", ";
```

Math:

$$\frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} = 1\times10^{-6} \text{ with } \sigma = 0.05 \implies x \simeq \pm 0.281889 \hat{=} 5.6\sigma \implies 1 - \int_{-5.6}^{5.6} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \mathrm{d}x = 2.1435180519446817 \times 10^{-8}$$

# Checks for Data Leakage - True Hit Times in Shower Data #1:

- first check:

- take all true hit times of hits in a shower
- check **every** measured hit time in the shower, if measured hit time == **some** true hit time

  $\Longrightarrow$ this means that: $\text{true\_hit\_time}_{\text{hit4}} == \text{measured\_hit\_time}_{\text{hit30}}$

- find:
train: in 0.03% of hits we find **one true hit time** of the hits in the shower
test: in 0.02% of hits we find **one true hit time** of the hits in the shower
validation: in 0.02% of hits we find **one true hit time** of the hits in the shower

# Checks for Data Leakage - True Hit Times in Shower Data #2:

- first check:

- take all true hit times of hits in a shower
- check **the corresponding** measured hit time in the shower, if measured hit time == **some** true hit time

  $\Longrightarrow$ this means that: $\text{true\_hit\_time}_{\text{hit4}} \neq \text{measured\_hit\_time}_{\text{hit30}}$

- find:
train: in 0.003% of hits we find **true hit time=measured hit time**
test: in 0.002% of hits we find **one true hit time=measured hit time**
validation: in 0.003% of hits we find **one true hit time=measured hit time**

# Other:

Dear Konrad Helms,

We're pleased to announce that your abstract "Time-of-Flight Estimation using Machine Learning Techniques" with ID #115 has been accepted in track "" ().

See below a summary of your submitted abstract:
Conference: ML4Jets2023
Submitted by: Konrad Helms
Title: Time-of-Flight Estimation using Machine Learning Techniques
Primary Authors: Konrad Helms
Co-authors: Bohdan Dudar, Frank-Dieter Gaede, Anatolii Korol, Peter McKeown, Steffen Schumann, Fabian Sinz
Track classification:
Presentation type:

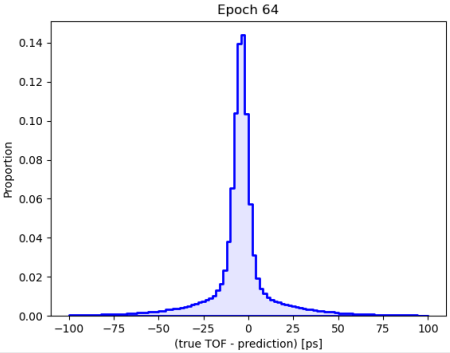
For a more detailed summary please visit the page of your abstract:
https://indico.cern.ch/event/1253794/abstracts/164359/

Kind regards,
The organizers of ML4Jets2023

--
Indico :: Call for Abstracts
https://indico.cern.ch/event/1253794/

# Current CNN Performance:

- RMS90: 10.65 ps



# Outlook:

- other checks for data leakage ongoing, nothing suspicious! - to complete the checks $\simeq$ 2 more days