Improvement of the energy resolution for neutral hadrons in highly granular calorimeters based on a neural network approach

Sergey Korpachev and Marina Chadeeva

Outline

- Goals and motivation
- Software, simulations and event selection
- Deep neural network
- Preliminary results
- Summary

Goals and motivation

- Highly granular calorimeters provide some additional information about the structure of hadronic showers.
- In non-compensating hadronic calorimeters the energy resolution for hadronic showers can be improved by applying software compensation techniques.
- The goal is to improve the energy resolution using the information about shower substructure in highly granular calorimeter.
- The current presentation focuses on implementation of global compensation method based on neural network technology in Particle Flow Approach. The global compensation means that variables used characterise a shower as a whole.
- (Local) Software compensation in Particle Flow reconstruction

Software, simulations and event selection

- iLCSoft
- Geant4 hadronic model: FTFP_BERT
- Isotropic Geant4 Particle Gun
- Particles are single K0L with energies: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, 120 GeV

- There are 5000 and 10000 events for each energy point
- E_{event} in event is a sum of hit energies (in GeV) in ECAL and HCAL $E_{event} = \sum_{i=1}^{N_{hits}} Ehit_i$
- The calibration is standard from iLCSoft
- No clustering is applied

Cuts:

- Absolute value of pseudorapidity is up to the end of the calorimeter system ($|\eta| < 3.0$)
- Events rejected when both (ecal + hcal) CalorimeterHit collections are empty
- If primary particle decays before calorimeter system $\Rightarrow$ skip an event (about 5-10% particles from full set are interacting or decaying before calorimeter)

Input features for neural network

ECAL and HCAL separately:

- Reconstructed energy in each calorimeter
- Number of hits in each calorimeter
- Average energy of hits in each calorimeter
- Average time of hits in each calorimeter
- Radius of shower in each calorimeter
- Average number of layers in each calorimeter

MUON SYSTEM:

- Reconstructed energy in muon system
- Average energy of hits in muon system
- Number of hits in muon system
- Average time of hits in muon system
- Average number of layers in muon system

Preprocessing

- 17 input features and 1 target (true energy from mc collection)
- no data normalization

- 26 energies $\cdot$ 10000 $=$ 260k events
- after cuts we have about 228k events (full set)

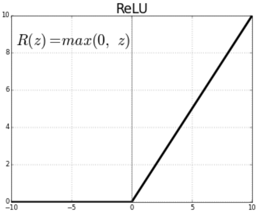
- train is about 160k events (70% of full set)
- validation is about 68k events (30% of full set)
- events are selected randomly without intersections

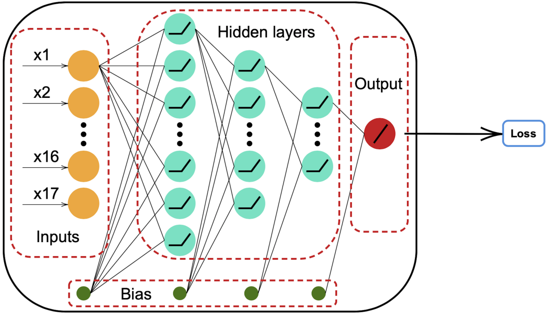
- test is 29 energy samples with 5000 events on each energy point
- further results of DNN performance are for test set
- events are selected randomly without intersections

The full sample contains single hadron events in all energy range studied (except for 7, 45 and 120 GeV). The last three energies only for test of DNN.
Results are shown for the test subsample.

Neural network structure and hyperparameters

- Keras library
- Architecture:
  - 1 input layer, 3 hidden layers, 1 output layer
  - Number of neurons: 17 / 128 / 64 / 32 / 1
  - Activation function: ReLU for hidden layers; linear $(\mathrm{f}(\mathrm{y}) = \mathrm{y})$ for output layer
- Optimizer: ADAM or NADAM
- Learning rate (lr): from 0.1 to 0.0000001
- Batch size (bs): 1, 2, 4, 8, 16 and 32
  $\Rightarrow$ Events come in batches iteratively
- Number of epochs: 10-200
- Optimized DNN: NADAM, bs=4, lr=0.000001, epochs=50





Loss function: modified MSE

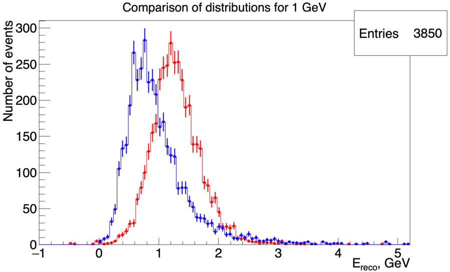
$$\mathrm{Loss} = \frac{1}{\mathrm{N}} \cdot \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{(\mathrm{Xpred_i} - \mathrm{Xtrue_i})^2}{0.44^2 \cdot \mathrm{Xtrue_i} + 0.04^2 \cdot \mathrm{Xtrue_i^2}},$$

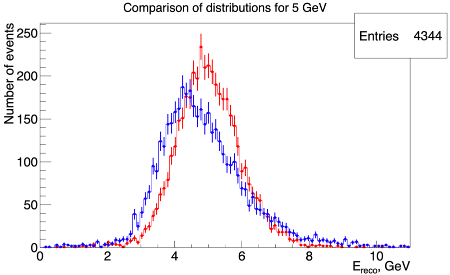
Xpred – prediction, Xtrue – from MC collection.

Further results shown for optimized DNN

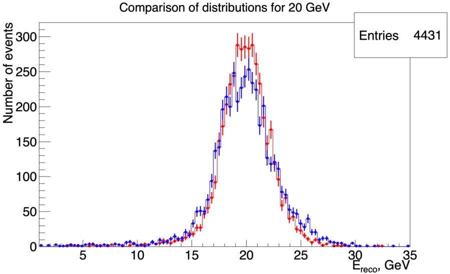
DNN performance for 1, 5, 20 and 40 GeV

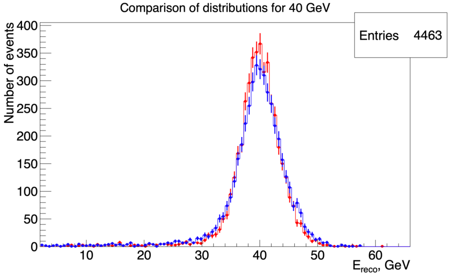
Standard reconstruction, **DNN reconstruction** (test sample)





The distribution width is improved. The mean has shifted to the right.

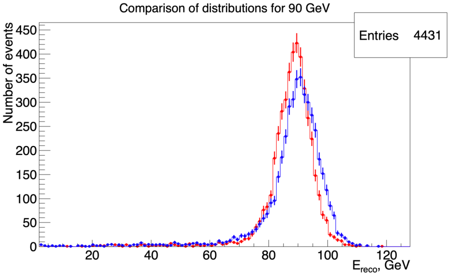


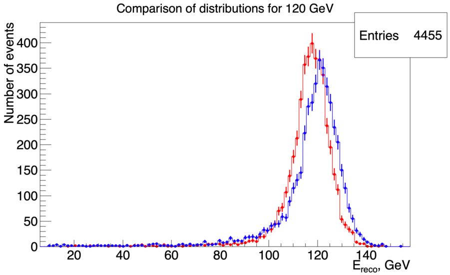


The distribution width is improved. The mean has not shifted.

DNN performance for 90 and 120 GeV

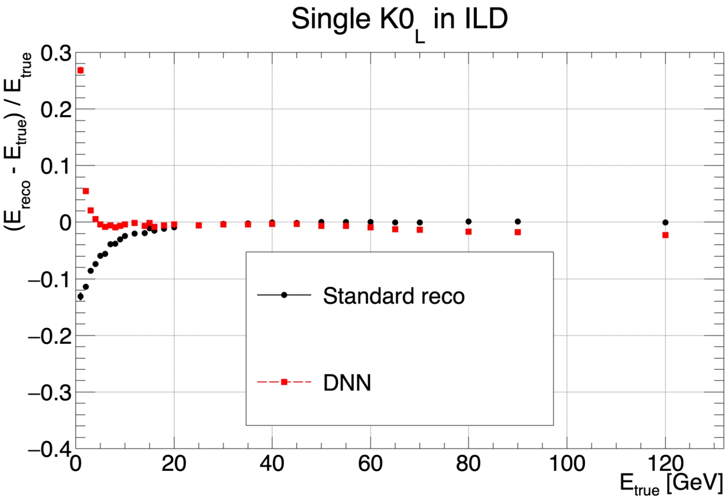
Standard reconstruction, **DNN reconstruction** (test sample)





The distribution width is improved. The mean has shifted to the left.

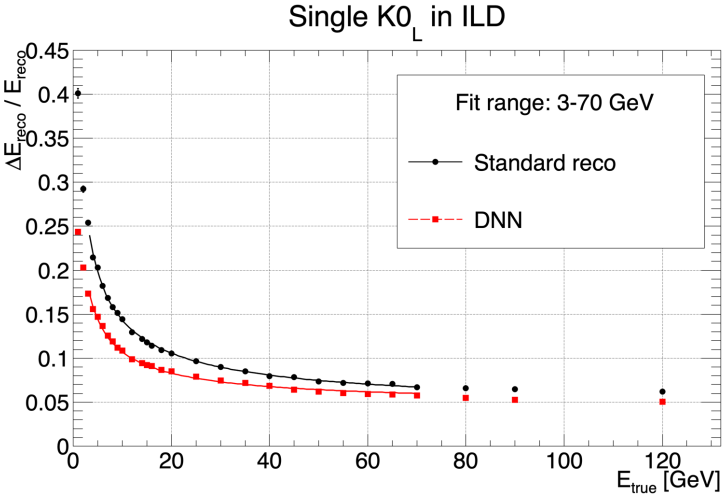
Linearity



Mean and sigma from hist90 of the energy distributions.

The training set didn't have energies: 7, 45 and 120 GeV. But they are in good agreement with other energy points. The neural network improves linearity of the response in the energy range 2–60 GeV.

Relative energy resolution



Mean and sigma from hist90 of the energy distributions.
The training set didn't have energies: 7, 45 and 120 GeV. But they are in good agreement with other energy points. The neural network improves relative energy resolution by about 25%.

$$\frac{\Delta \mathrm{E}}{\mathrm{E}} = \sqrt{(\frac{\mathrm{A}}{\sqrt{\mathrm{E}}})^2 + (\mathrm{B})^2 + (\frac{\mathrm{C}}{\mathrm{E}})^2}\ .$$

Preliminary results

Reconstruction	A, $\sqrt{\mathrm{GeV}}$	B	C, GeV
Standard	0.431$\pm$0.002	0.0432$\pm$0.0013	0.0
DNN	0.303$\pm$0.003	0.0478$\pm$0.0013	0.0

- The neural network shows noticeable improvement in energy resolution
- **Ideal case, which not will be used in real reconstruction**

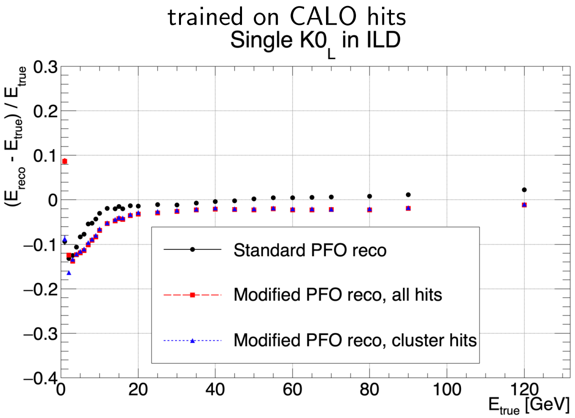
Particle Flow Approach study

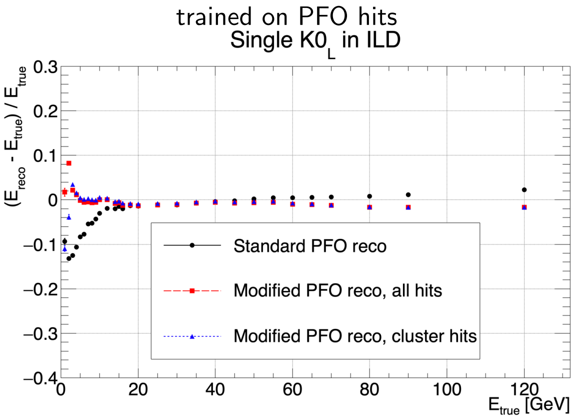
- It is necessary to implement the neural network technique in PandoraPFA processor to extract calorimeter hits from PFO objects (GitHub: PandoraPFA/LCPandoraAnalysis)
- Two new functions in the PandoraPFA:
  - input features for DNN application
  - DNN architecture from scratch - for weights from trained DNN
- Neutral hadron PFO's only
- Several events are empty after PandoraPFA processor ($\approx$100 events per energy point)
- New DNN was trained on all PFO hits
- The same structure of the DNN: architecture and hyperparameters
- Two options were studied:
  - DNN applied to joint neutral hadron PFO hits
  - DNN applied to each neutral hadron PFO and summed up (if energy of neutral hadron PFO is 1.5 GeV and more)

Cuts:

- Full set is about 228k events, about 210k is set after cut for low energies: $E_{event} >= (E_{true} - 3 \cdot 0.6 \cdot \sqrt{E_{true}})$ and $E_{event} >= 0.2$
- Train is about 147k events (70% of set) and validation is about 63k (30% of set)
- Cut before hist90 procedure is $E_{event} >= 0.3 \cdot E_{true}$

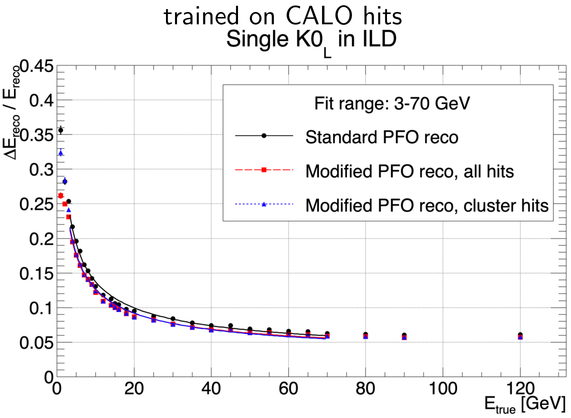
Linearity for PFA

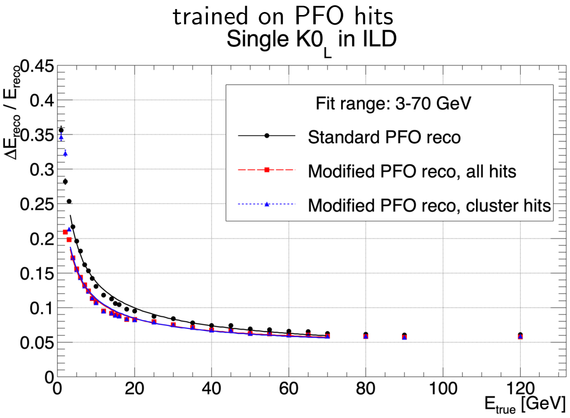




Mean and sigma from hist90 of the energy distributions.
The training set didn't have energies: 7, 45 and 120 GeV. But they are in good agreement with other energy points. The neural network techhique improves linearity of the response in case of trained on PFO hits (right plot). The method shows additive behavior, except low energy points.

Relative energy resolution for PFA





Mean and sigma from hist90 of the energy distributions.

The training set didn't have energies: 7, 45 and 120 GeV. But they are in good agreement with other energy points. The neural network techique improves relative energy resolution in both cases, especially for training on PFO hits (right plot). The method shows additive behavior.

$$\frac{\Delta \mathrm{E}}{\mathrm{E}} = \sqrt{(\frac{\mathrm{A}}{\sqrt{\mathrm{E}}})^2 + (\mathrm{B})^2 + (\frac{\mathrm{C}}{\mathrm{E}})^2} .$$

Summary

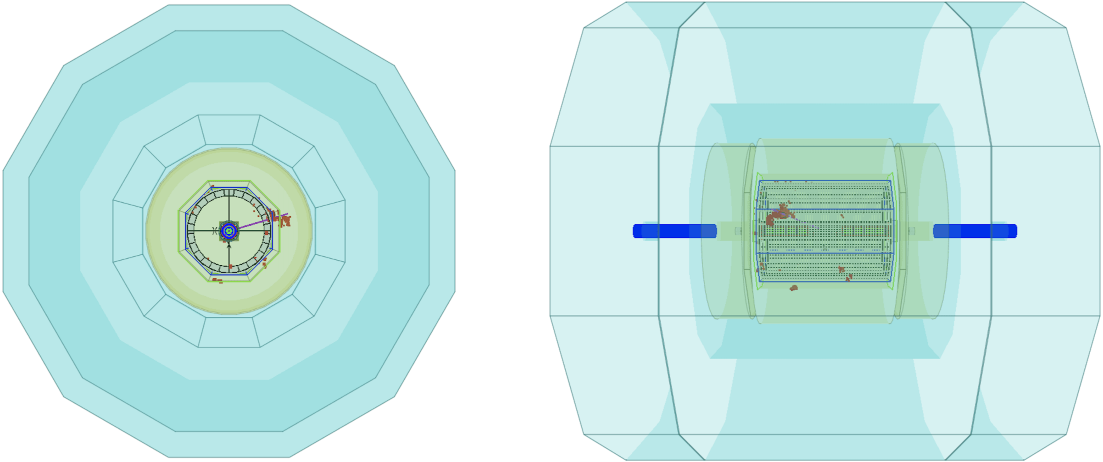
PFA reconstruction	**A,** $\sqrt{\mathrm{GeV}}$	**B**	**C,** GeV
Standard	0.423$\pm$0.005	0.0311$\pm$0.0038	0.0
DNN (trained on CALO hits)			
Applied to all PFO hits	0.383$\pm$0.003	0.0322$\pm$0.0024	0.0
Applied to separate clusters	0.391$\pm$0.004	0.0287$\pm$0.0036	0.0
DNN (trained on PFO hits)			
Applied to all PFO hits	0.328$\pm$0.004	0.0415$\pm$0.0020	0.0
Applied to separate clusters	0.335$\pm$0.005	0.0394$\pm$0.0029	0.0

- Hadronic showers from K0L in the range 1-120 GeV are simulated in ILD
- The neural network from Keras package was trained and tested
- The neural network shows noticeable improvement in energy resolution

- The neural network technique implemented in PandoraPFA processor
- The preliminary results also show improvement in energy resolution
- Suggested method shows additive behavior

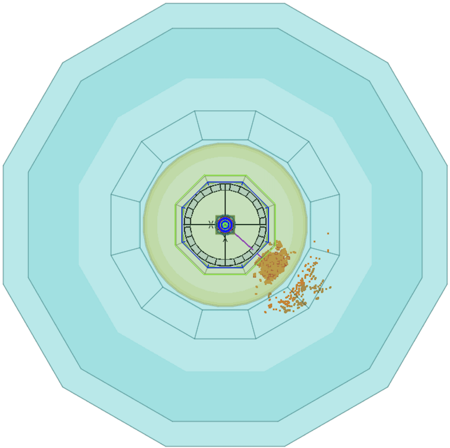
Backup slides

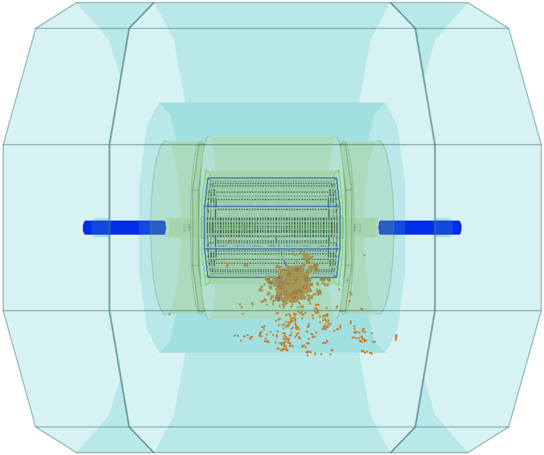
Event display for 5 GeV



The primary particle produces several clusters.

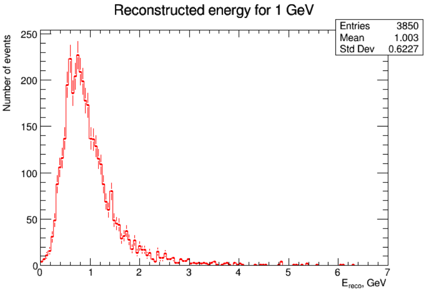
Event display for 90 GeV

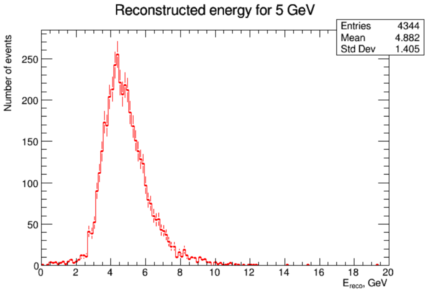




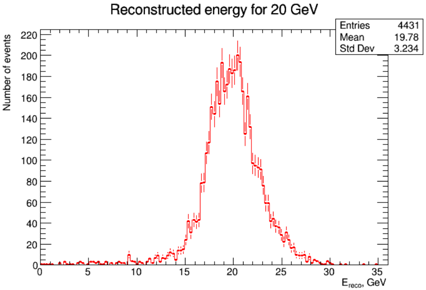
The hadronic shower leaks into muon system.

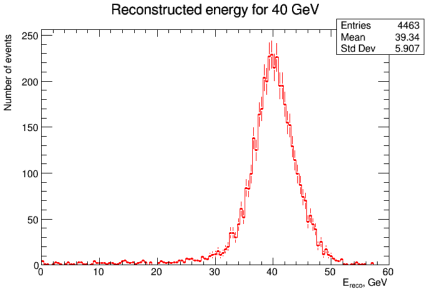
Energy distribution for single hadron: 1, 5, 20 and 40 GeV





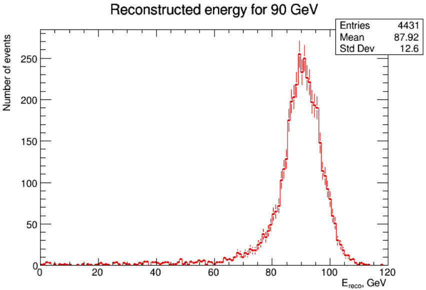
Asymmetric distributions with right tail for energies below 10 GeV

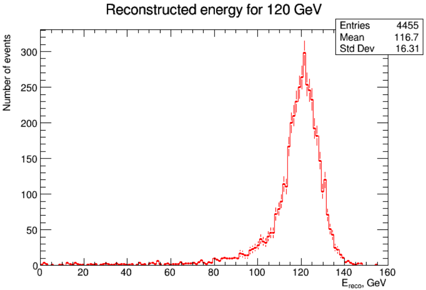




Gaussian shape in the intermediate range

Energy distribution for single hadron: 90 and 120 GeV

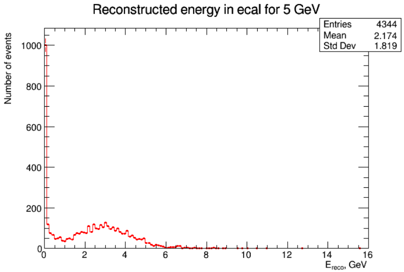


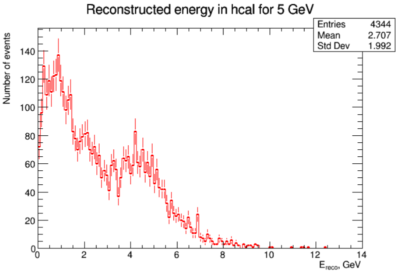


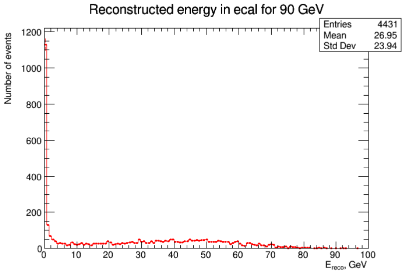
Asymmetric distributions with left tail above 60 GeV (leakage)

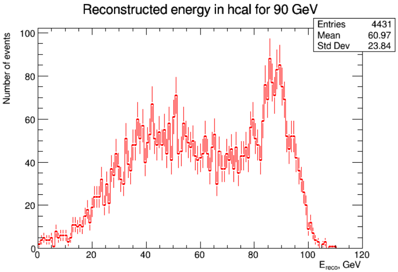
Energy distributions in ecal and hcal: 5 and 90 GeV

For 5000 sample







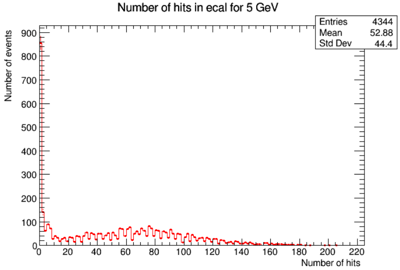


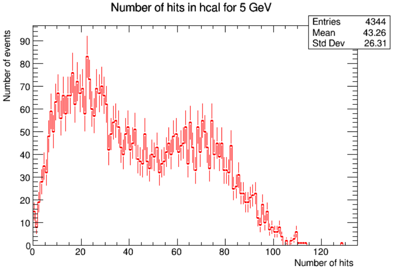
Reconstructed energy for each calorimeter in GeV

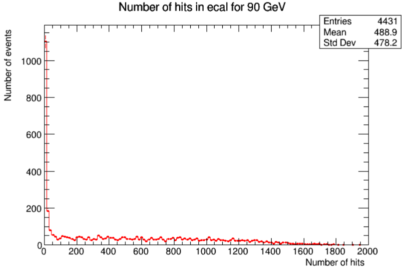
Peak at zero in ecal means hadrons start in hcal

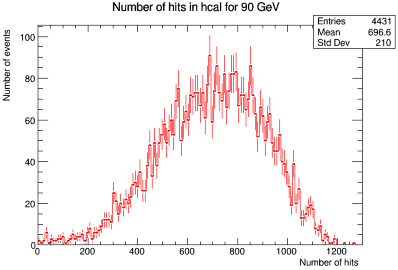
Distribution of number of hits: 5 and 90 GeV

For 5000 sample







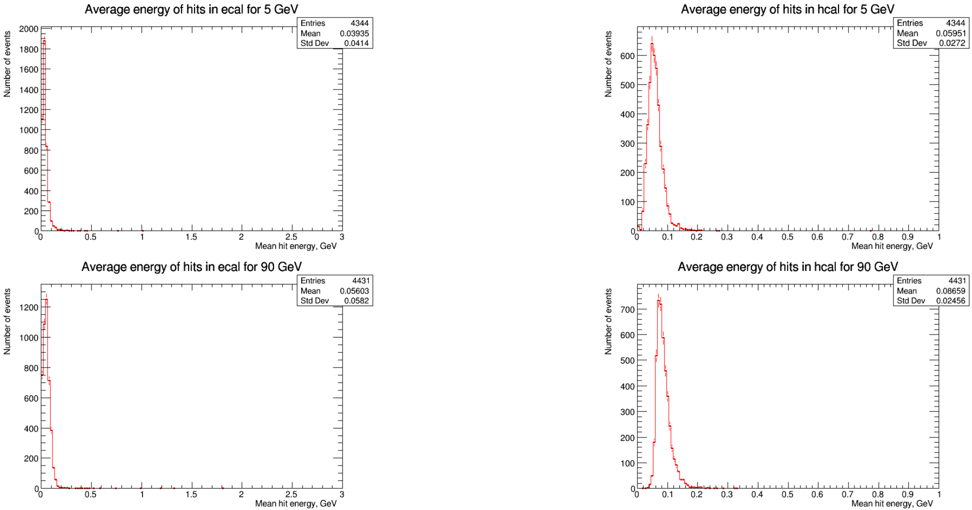


Number of hits in ecal and hcal

Peak at zero in ecal means hadrons start in hcal

Mean hit energy in ecal and hcal: 5 and 90 GeV

For 5000 sample



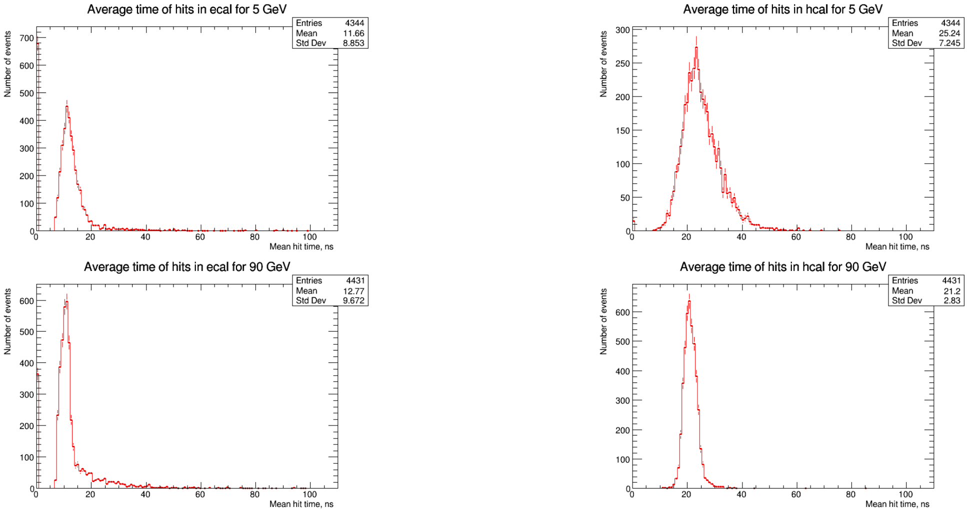
Average energy of hits in ecal and hcal in GeV

Peak at zero in ecal means hadrons start in hcal

$$\text{Mean hit energy} = \frac{\sum_{i=1}^{N_{hits}} Ehit_i}{N_{hits}}$$

Mean hit time: 5 and 90 GeV

For 5000 sample



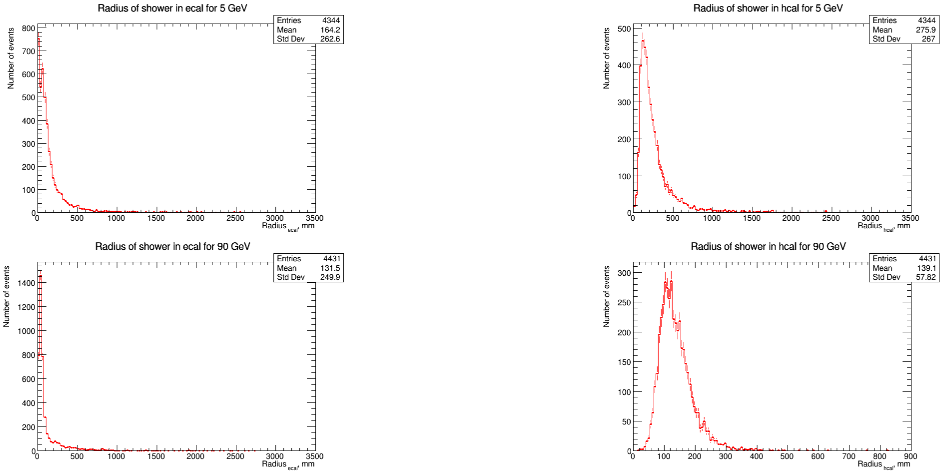
Mean time of hits in ecal and hcal in ns

Peak at zero in ecal means hadrons start in hcal

$$\text{Mean hit time} = \frac{\sum_{i=1}^{N_{hits}} Thit_i}{N_{hits}}$$

Energy weighted radius of shower: 5 and 90 GeV

For 5000 sample



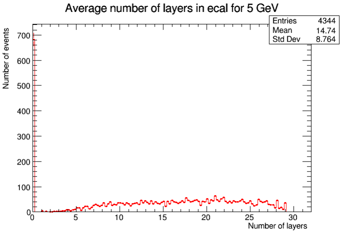
Radius of shower for ecal and hcal in mm

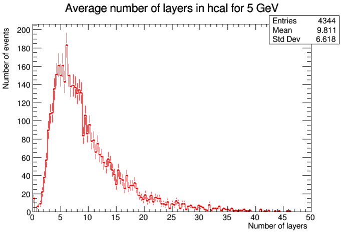
Distance between each hit position and straight line of IP (0,0,0) and CoG (x,y,z)

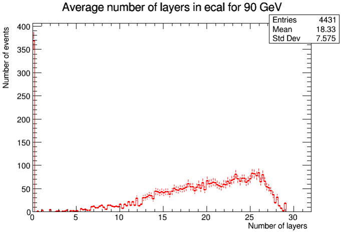
$$\text{Energy weighted radius of shower} = \frac{\sum_{i=1}^{N_{hits}} Distance\Big((IP,CoG),hit\_position_i\Big)\cdot Ehit_i}{\sum_{i=1}^{N_{hits}} Ehit_i}$$

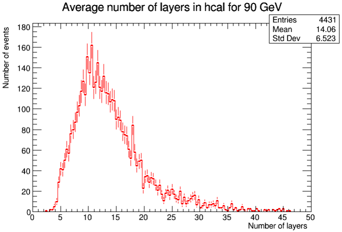
Weighted number of layer: 5 and 90 GeV

For 5000 sample









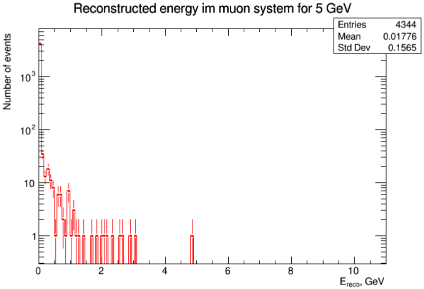
Average number of layer for each events in ecal and hcal

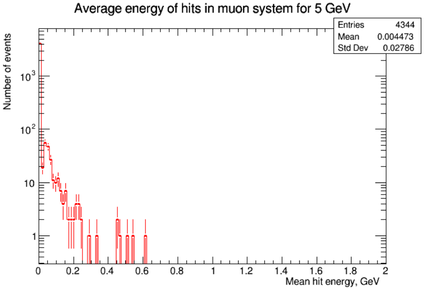
Peak at zero in ecal means hadrons start in hcal

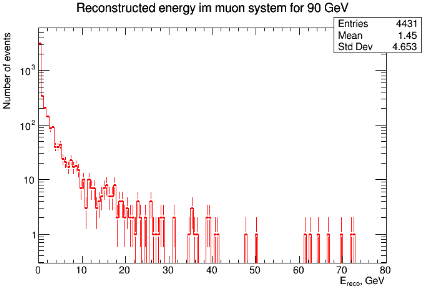
$$\text{Weighted number of layer} = \frac{\sum_{i=1}^{N_{hits}} Number\_hit\_layer_i \cdot Ehit_i}{\sum_{i=1}^{N_{hits}} Ehit_i}$$

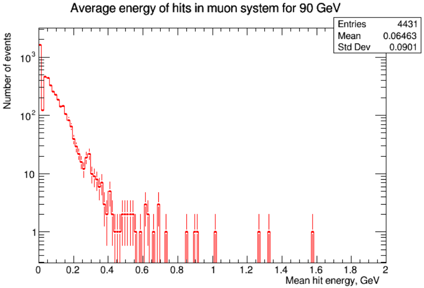
Energy and hit energy in muon system: 5 and 90 GeV

For 5000 sample







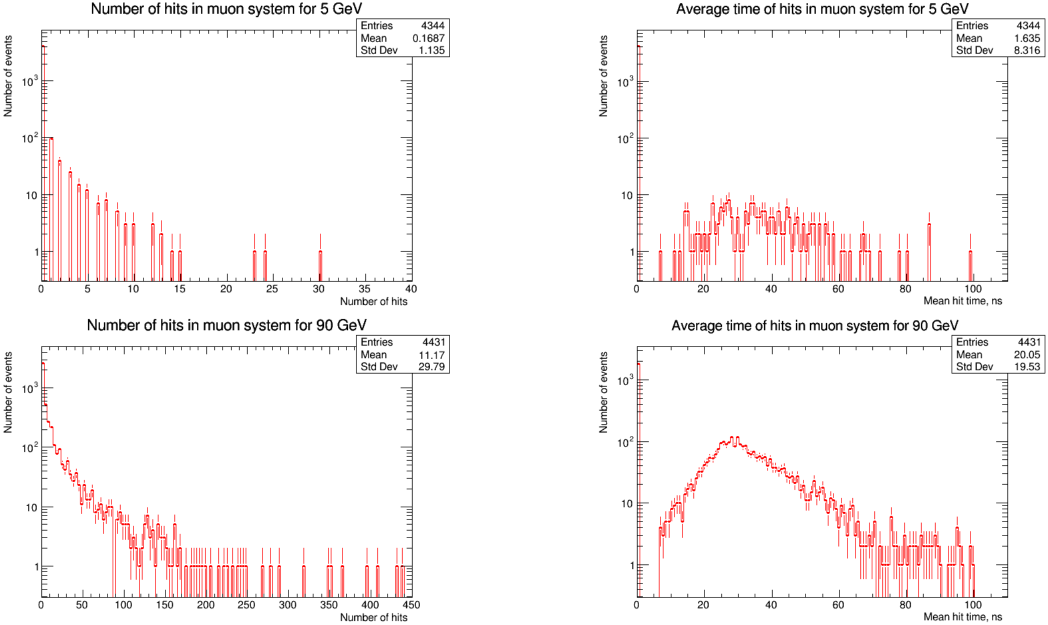


Reconstructed energy in GeV (left) and average energy of hits in GeV (right)

About 5% (63%) of hadrons reach muon system for 5 (90) GeV

Number of hits and hit time in muon system: 5 and 90 GeV

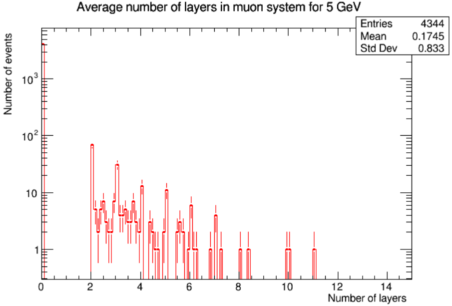
For 5000 sample

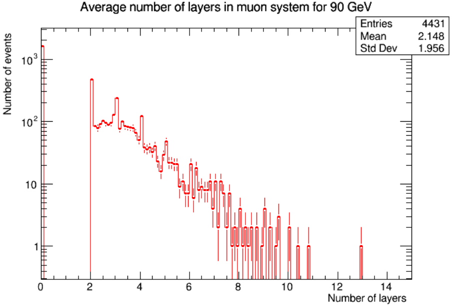


Number of hits (left) and mean time of hits in ns (right)

Weighted number of layer in muon system: 5 and 90 GeV

For 5000 sample

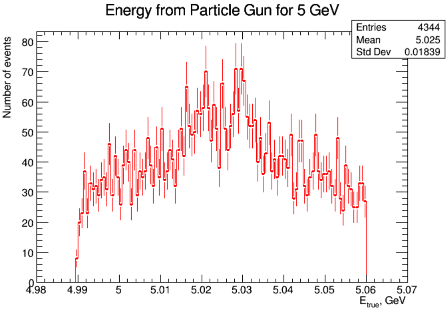


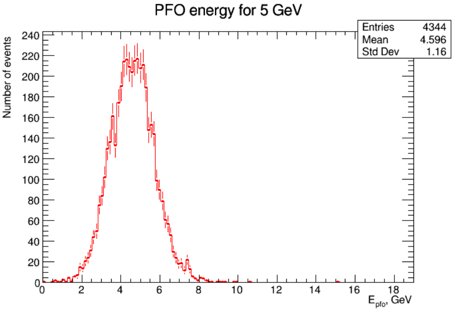


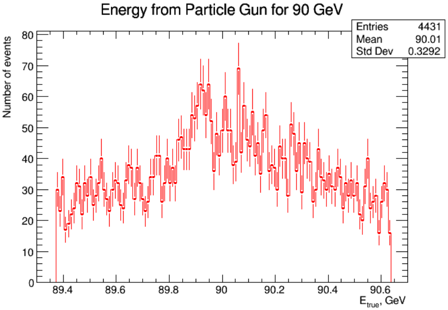
Average number of layer

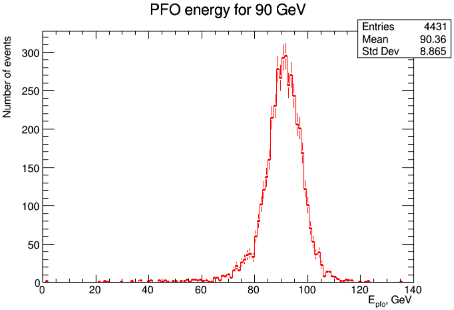
MC and PFO for analysis

For 5000 sample



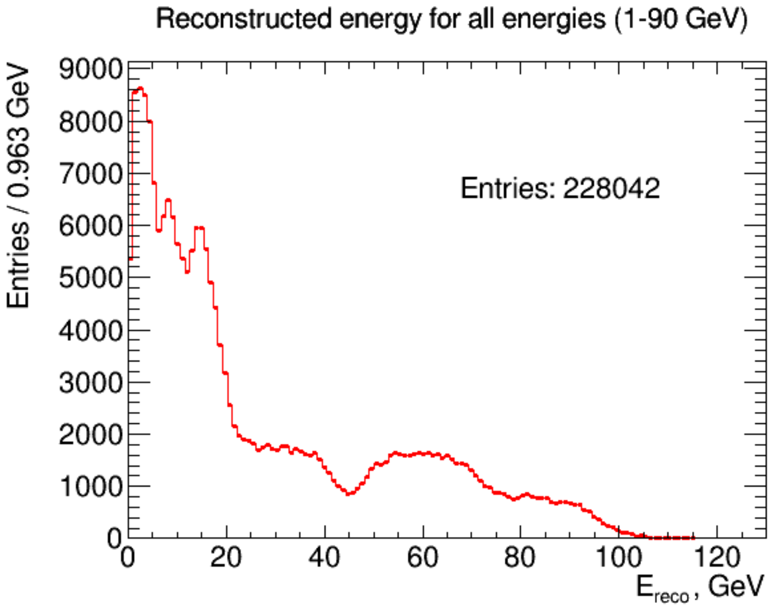






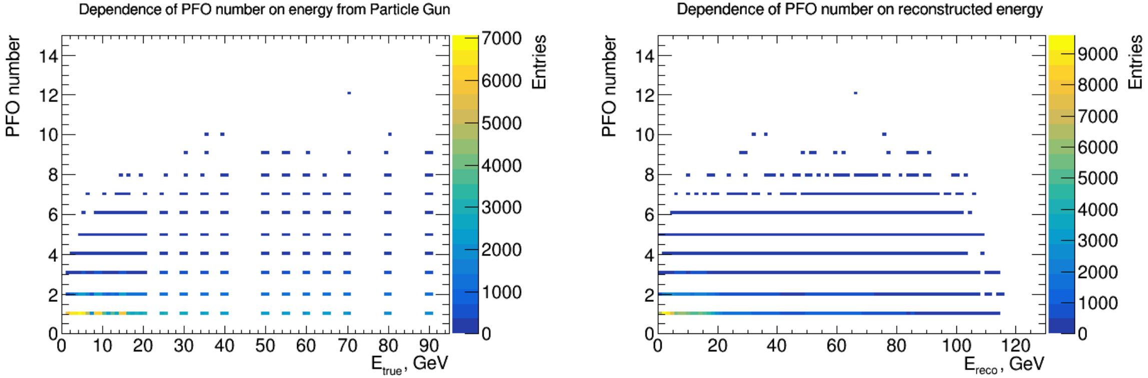
No clustering, PFO's only for cross check

Mixed energy distribution for training of DNN



Except for 7, 45 and 120 GeV.

Dependence PFO number on different types of energy



Correlations are weak for both cases.
Events with one PFO object dominate.

Number of events after cuts

Full set of 10000 events

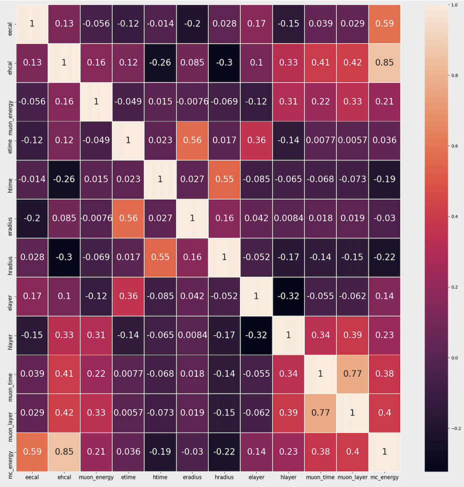
Energy	1	3	5	10	30	40	50	70	90	120
N	7727	8492	8632	8814	8912	8858	8897	8895	8950	8964
%	77.27	84.92	86.32	88.14	89.12	88.58	88.97	88.95	89.50	89.64

Similar efficiencies for all energies, slightly lower for 1 GeV

Full set of 5000 events

Energy	1	3	5	10	30	40	50	70	90	120
N	3850	4242	4344	4342	4446	4463	4439	4456	4431	4455
%	77.00	84.84	86.88	86.84	88.92	89.26	88.78	89.12	88.62	89.10

Correlation between features and target



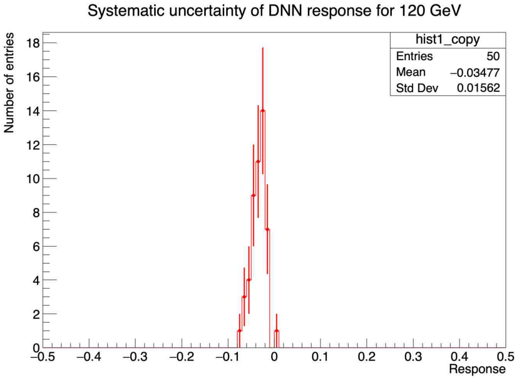
Target: energy from Particle Gun (MC collection)

The largest positive correlation with energy in ECAL, energy in HCAL and average time of hits in muon system; the largest negative correlation with shower radius in HCAL and average time of hits in HCAL

Correlation between variables

	eecal	ehcal	muon_energy	etime	htime	eradius	hradius	elayer	hlayer	muon_time	muon_layer	mc_energy
eecal	1.000000	0.131429	-0.055831	-0.115008	-0.014099	-0.202531	0.027832	0.172889	-0.152756	0.039403	0.028724	0.585834
ehcal	0.131429	1.000000	0.155594	0.121166	-0.257103	0.085121	-0.299339	0.101742	0.327006	0.405211	0.418303	0.852807
muon_energy	-0.055831	0.155594	1.000000	-0.049307	0.014769	-0.007633	-0.068726	-0.120089	0.306266	0.224953	0.334404	0.212033
etime	-0.115008	0.121166	-0.049307	1.000000	0.023430	0.561189	0.016539	0.364091	-0.137274	0.007689	0.005676	0.035826
htime	-0.014099	-0.257103	0.014769	0.023430	1.000000	0.026714	0.553583	-0.085131	-0.065480	-0.067692	-0.072625	-0.194386
eradius	-0.202531	0.085121	-0.007633	0.561189	0.026714	1.000000	0.163117	0.041665	0.008397	0.018068	0.019396	-0.030379
hradius	0.027832	-0.299339	-0.068726	0.016539	0.553583	0.163117	1.000000	-0.052394	-0.166443	-0.141630	-0.150124	-0.222829
elayer	0.172889	0.101742	-0.120089	0.364091	-0.085131	0.041665	-0.052394	1.000000	-0.323224	-0.054740	-0.062230	0.144776
hlayer	-0.152756	0.327006	0.306266	-0.137274	-0.065480	0.008397	-0.166443	-0.323224	1.000000	0.336597	0.390446	0.229130
muon_time	0.039403	0.405211	0.224953	0.007689	-0.067692	0.018068	-0.141630	-0.054740	0.336597	1.000000	0.769852	0.375865
muon_layer	0.028724	0.418303	0.334404	0.005676	-0.072625	0.019396	-0.150124	-0.062230	0.390446	0.769852	1.000000	0.400483
mc_energy	0.585834	0.852807	0.212033	0.035826	-0.194386	-0.030379	-0.222829	0.144776	0.229130	0.375865	0.400483	1.000000

Example of systematic uncertainty



Example histogram for response at 120 GeV

Response $= \frac{Ereco - Etrue}{Etrue}$

- 50 runs with the same DNN
- Fluctuations due to random selection of train/validation samples and random initialization of weights of DNN

- Response (1)
- Absolute energy resolution (2)
- Relative energy resolution (3)

- 29 energies $\cdot 3 = 87$ histograms
- Separate uncertainty for each energy point

Techniques of resolution estimate

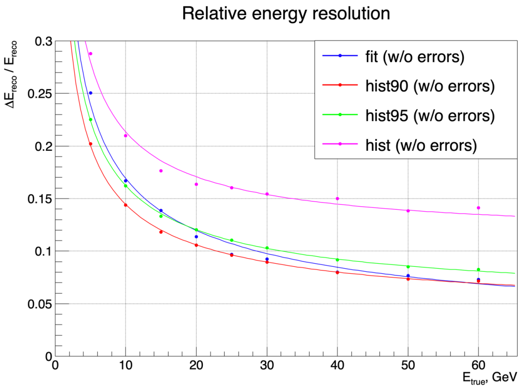
Techniques:

- HIST: mean and rms of the full distribution (standard ROOT procedure is used)
- HIST90: mean and rms of the 90% of the full distribution
- HIST95: mean and rms of the 95% of the full distribution
- Fit: mean and sigma of Gaussian fit of the full distribution

Legends:

- E_{reco} is mean from fit or histogram
- ΔE_{reco} is σ from fit or rms from histogram
- E_{true} is energy from generator

Relative energy resolution for different techniques



	fit	hist90	hist95	hist
A, sqrt(GeV)	0.53	0.44	0.49	0.57
B, %	0.0	4.0	5.1	11.3
C, GeV	0.0	0.0	0.0	0.0

$$\sqrt{(\frac{A}{\sqrt{E}})^2 + B^2}$$

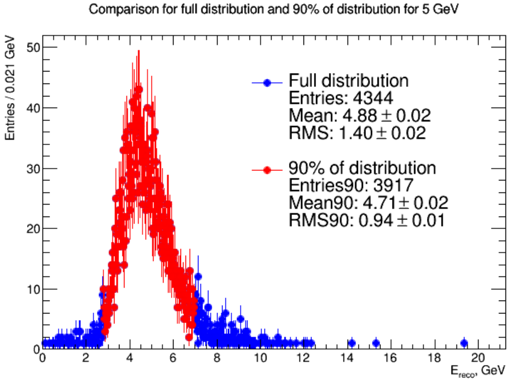
The better resolution is for RMS90 (hist90). No noise is added (C = 0).
Relative resolution fit with 3 terms in backup.

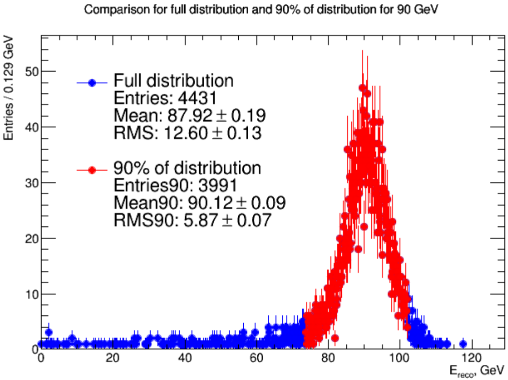
RMS90 procedure

RMS90:

- Find a bin of a mean of the histogram
- Define 90% of the histogram as $N_{90} = 0.9 \cdot (hist \rightarrow GetEntries)$
- RMS formula $= \sqrt{\frac{\sum w \cdot x^2}{\sum w} - (\frac{\sum w \cdot x}{\sum w})^2}$,
  where x is GetBinCenter (bin of mean plus/minus step of iteration) and w is GetBinContent (bin of mean plus/minus step of iteration)
- Sums are calculated by moving symmetrically to the left and to the right bin-by-bin from the mean. The calculation stops when number of events reaches N_{90}

Example of RMS90 for 5 and 90 GeV





This method extracts the true values of the mean and width of the distribution.

Fit procedure

- Take mean and rms from full histogram
- Perform Gaussian fit in [mean $\pm$ (range $\cdot$ rms)], where range is values from 1.0 to 2.5 $\Rightarrow$ array of means and σ's from Gauss fits
- Fit is accepted if $\frac{\chi^2}{NDF} < 1.5$
- Final fit result is that with minimum $\frac{\chi^2}{NDF}$

RMS90 code example

It depends how you define the central 90% (model dependent).
Below a brute force example.

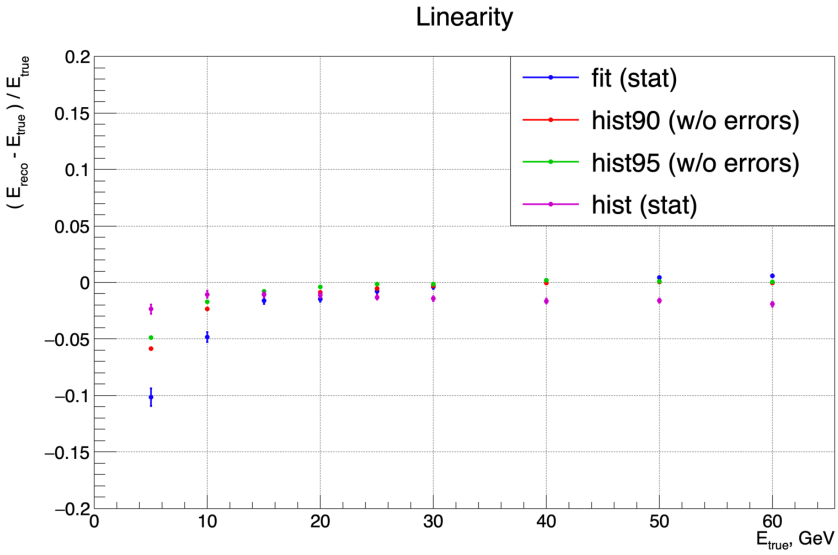
Rene

```
[code]void rms90(TH1 h) {
TAxis axis = h->GetXaxis();
Int_t nbins = axis->GetNbins();
Int_t imean = axis->FindBin(h->GetMean());
Double_t entries =0.9h->GetEntries();
Double_t w = h->GetBinContent(imean);
Double_t x = h->GetBinCenter(imean);
Double_t sumw = w;
Double_t sumwx = wx;
Double_t sumwx2 = wxx;
for (Int_t i=1;i<nbins;i++) {
if (i> 0) {
w = h->GetBinContent(imean-i);
x = h->GetBinCenter(imean-i);
sumw += w;
sumwx += wx;
sumwx2 += wxx;
}
if (i<= nbins) {
w = h->GetBinContent(imean+i);
x = h->GetBinCenter(imean+i);
sumw += w;
sumwx += wx;
sumwx2 += wxx;
}
if (sumw > entries) break;
}
x = sumwx/sumw;
Double_t rms2 = TMath::Abs(sumwx2/sumw -x*x);
Double_t result = TMath::Sqrt(rms2);
printf("RMS of central 90% = %g, RMS total = %g\n",result,h->GetRMS());
}

void central90() {
TH1F *h = new TH1F("h","test",100,-4,2);
h->FillRandom("gaus",10000);
rms90(h);
}[/code]
```

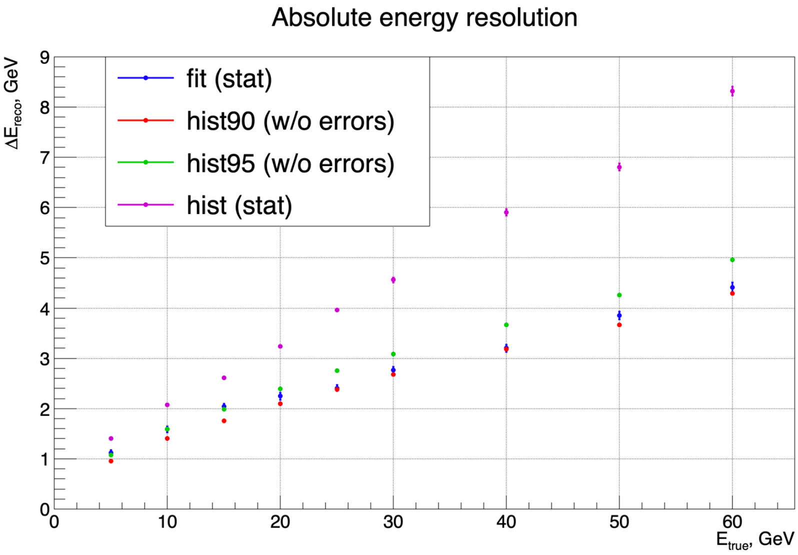
The code is from: https://root-forum.cern.ch/t/rms90

Linearity for different techniques



Good coincidence of fit, RMS90 and RMS95 above 15 GeV.
The worst linearity for fit (in agreement with physics).

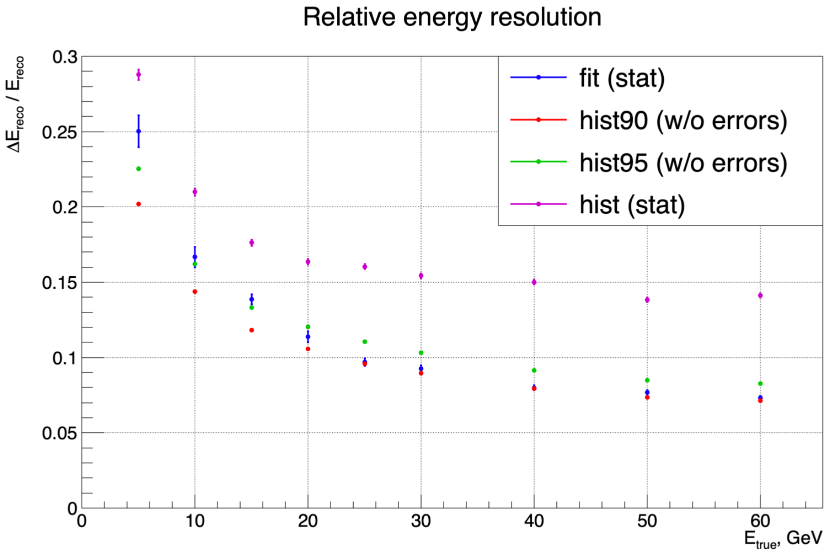
Absolute energy resolution for different techniques



Fit and RMS90 look similar.

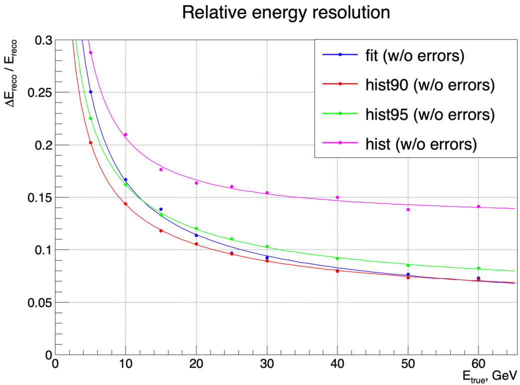
RMS95 in agreement with fit and RMS90 before 20 GeV.

Relative energy resolution for different techniques



The better resolution is for RMS90. (RMS90 is hist90)

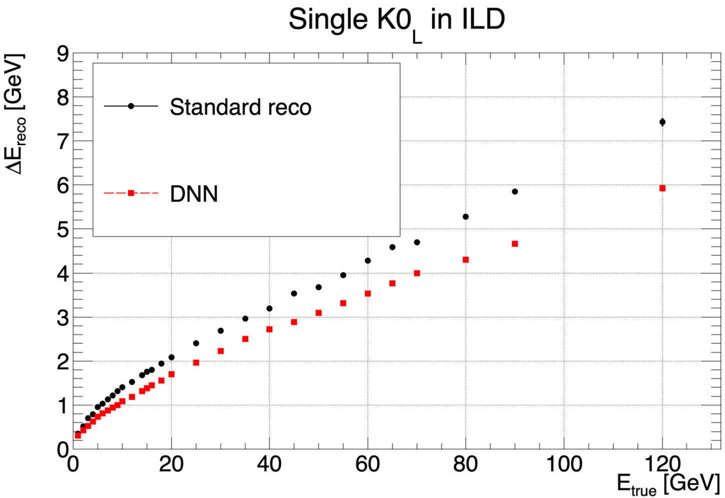
Fit for relative energy resolution: 3 terms



	fit	hist90	hist95	hist
A, sqrt(GeV)	0.46	0.42	0.47	0.44
B, %	3.5	4.5	5.4	12.8
C, GeV	0.7	0.3	0.3	0.8

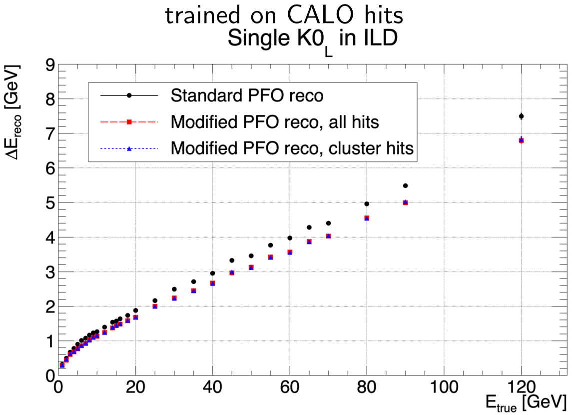
$$\sqrt{\left(\frac{A}{\sqrt{E}}\right)^2 + B^2 + \left(\frac{C}{E}\right)^2}$$

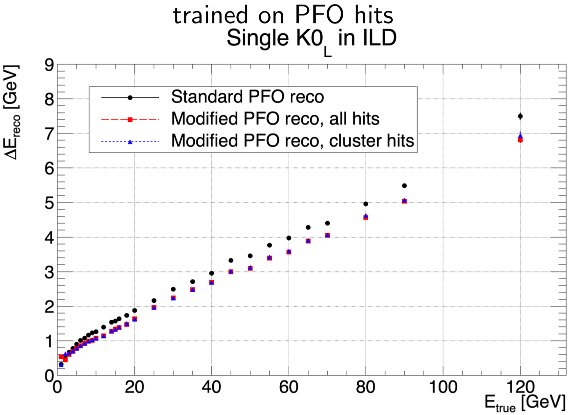
Absolute energy resolution



Mean and sigma from hist90 of the energy distributions.
The training set didn't have energies: 7, 45 and 120 GeV. But they are in good agreement with other energy points. The neural network improves absolute resolution.

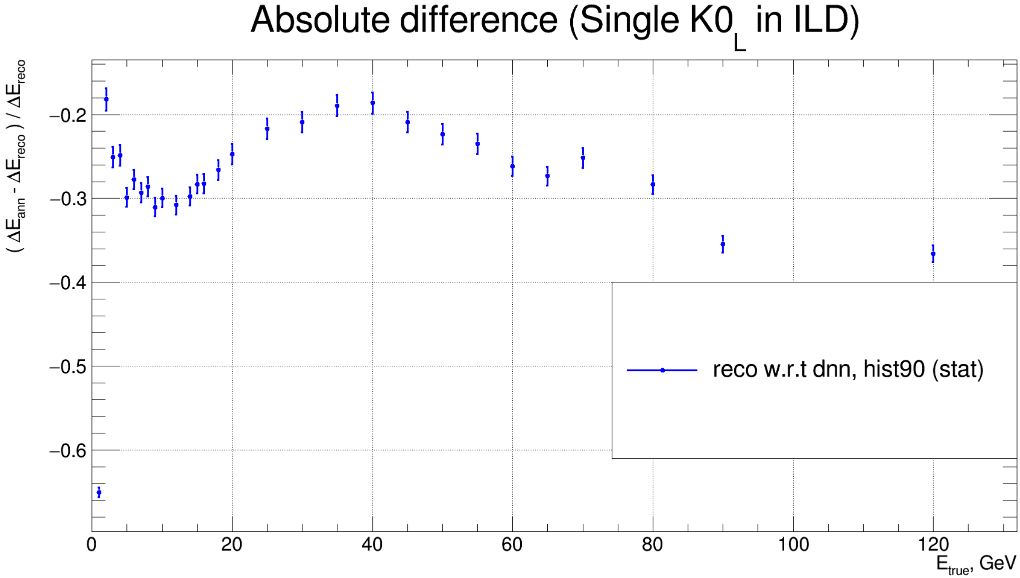
Absolute energy resolution for PFA



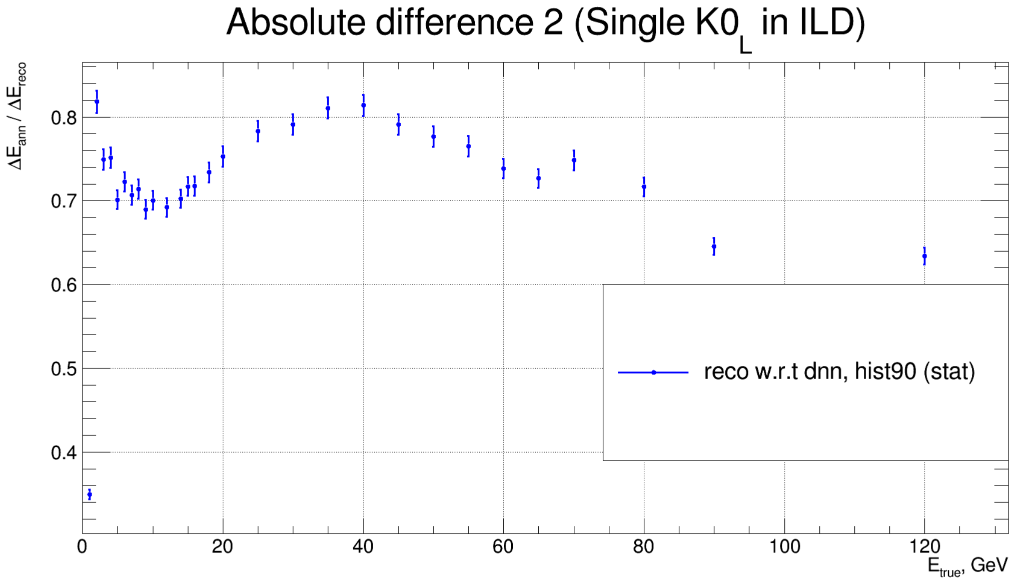


Mean and sigma from hist90 of the energy distributions.
The training set didn't have energies: 7, 45 and 120 GeV. But they are in good agreement with other energy points. The neural network technique improves absolute resolution in both cases.
The method shows additive behavior.

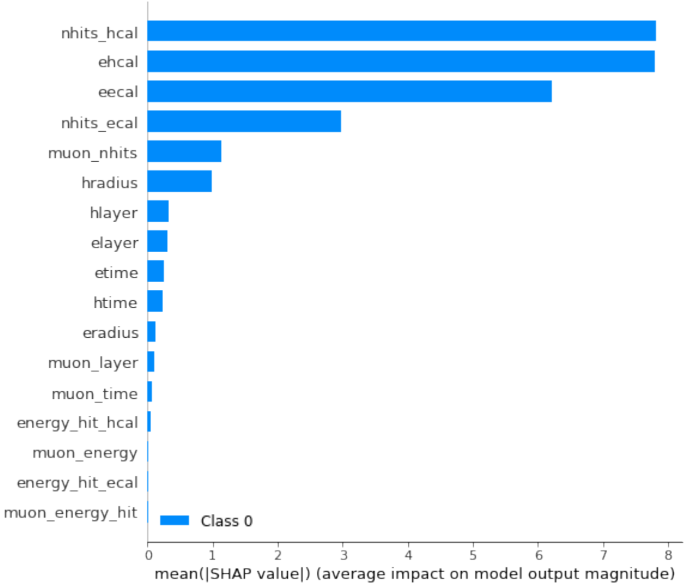
Absolute difference between RECO and DNN



Absolute difference between RECO and DNN



SHAP



The most important are the first six features

TO DO

- Try new version of iLCSoft
- Understand figures of merit for DNN
- Study the effect on jet energy resolution
- Try the global compensation variables from CALICE Analysis Notes (https://arxiv.org/abs/1207.4210)