Analysis of test-beam data from the W-AHCAL prototype

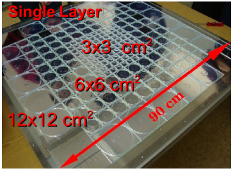
Angela Burger (CERN, LMU)

17.09.2014

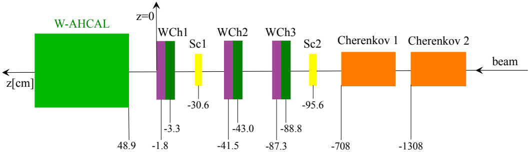
Test-beam measurement

Test-beam measurement performed on the **Calice W-AHCAL**:

- Sampling calorimeter (38 layers)
- absorber material: tungsten
- active layers: scintillator tiles coupled to SiPM readout



Experimental set-up:



- Beam energies: 1-300 GeV (of interest for this studies: 25-150 GeV)
- Particles: electrons, muons, **pions** and **protons**

Goal of the analysis

- **Comparison of Data/Monte-Carlo** with emphasis on variables describing **hadronic shower fluctuations**
  Comparison measurement data to two models (GEANT 4, version 9.6.p01)
  - QGSP_BERT_HP (quark-gluon string precompound + Bertini Cascade models + neutron high precision)
  - FTFP_BERT_HP (Fritiof precompound + Bertini Cascade models + neutron high precision)

- Investigate **separation potential** of **protons and pions** in the W-AHCAL

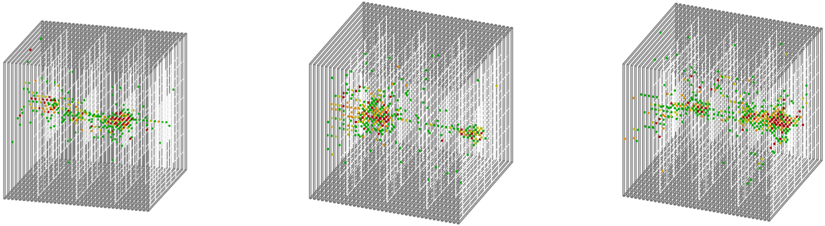
Part 1

Study of fluctuations of hadronic showers
Comparison of data and Monte-Carlo

Fluctuation of hadronic showers: Motivation

Investigate fluctiations in radial and longitudinal shower development

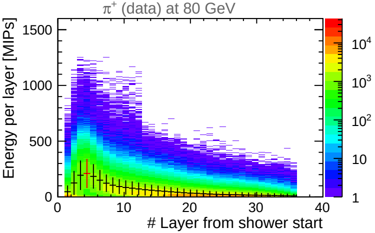
- Not yet discussed in detail in ▸ CAN-044
- Shape of hadronic showers vary very much (Example: Pions, beam energy = 80 GeV):



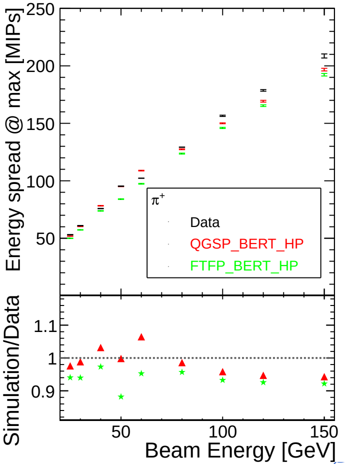
Do the simulations describe hadronic shower fluctuations correctly?

Fluctuations of longitudinal shower development

Fluctuation of energy loss in the calorimeter layers:
Quantified using spread of the energy deposition in mean layer of shower maximum:



Pions:

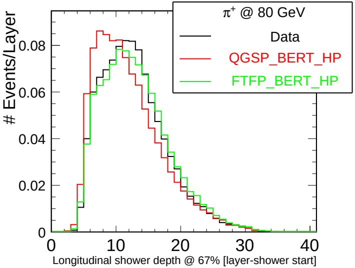


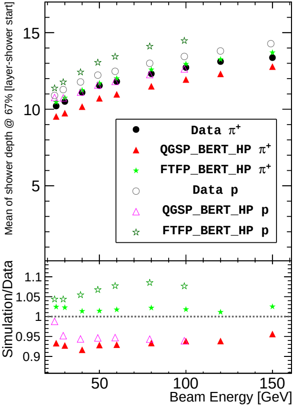
- MC underestimate spread by up to 10%
- QGSP_BERT_HP in this case better option
- Larger deviation data/MC @50-60 GeV ⇒ position of maximum changes, but at a different beam energy for data and MC
- Results similar for protons

Longitudinal shower depth @ 67%: Mean

Calculate mean number of layer it takes to absorb 67% of full shower energy.

Example for the distribution of shower depth @ 67%:
pions, beam energy: 80 GeV:

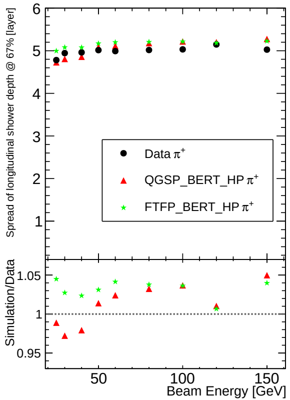




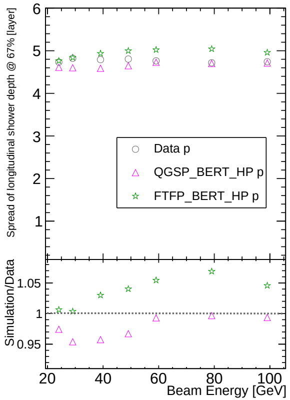
- Goodness of description model and particle dependent
- Agreement within 10%
- Proton showers are deeper than pion showers

Longitudinal shower depth @ 67%: Spread

Pions:



Protons:

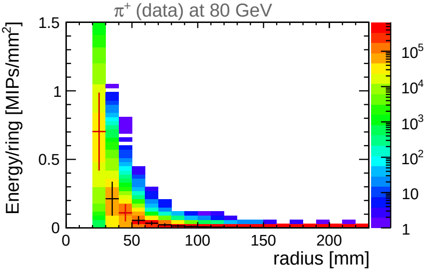


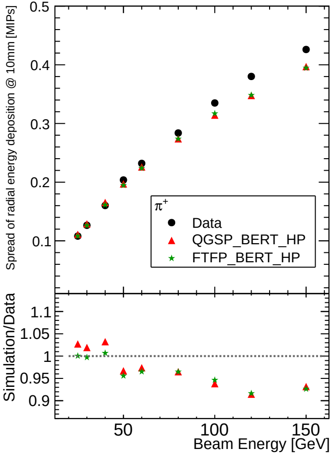
- Agreement data and MC within 5%
- QGSP_BERT_HP better option for protons and pions

Fluctuations of the radial shower development

Fluctuations of energy deposition in rings around the center of gravity
⇒ quantified by spread of energy deposition in innermost ring

Pions:



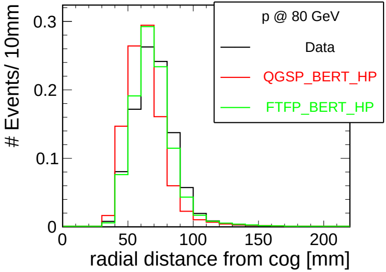


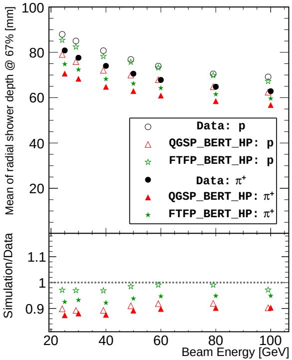
- Agreement data/MC within 10% for pions (for protons even 5%)
- Similar behaviour @ higher radii

Radial shower depth @67%: Mean

Radius of the cylinder containing 67% of the full shower energy

Example of the distribution of the radial shower depth @ 67% (Proton, beam energy = 80 GeV):

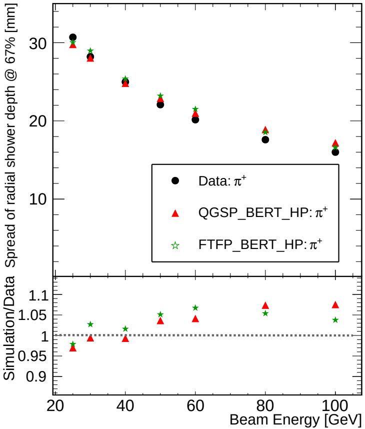




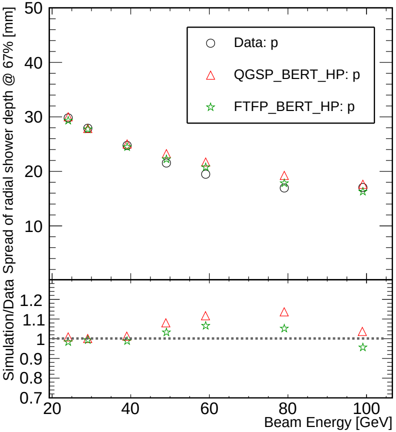
- FTFP_BERT_HP describes data better in both cases $\Rightarrow$ agreement within 5% (Protons) and 10% (Pions)
- QGSP_BERT_HP: agreement in the order of 10-15%
- Both MC underestimate radial shower depth
- Pion shower deposit more energy close to center axis

Radial shower depth @67%: Spread

Pions:



Protons:



- Agreement data/MC of the order of 10%

Part 2

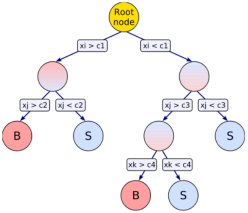
Proton and pion distinction using
a Boosted Decision Tree

Method

- **Goal:** Testing possibility to distinguish between protons and pions in the W-AHCAL prototype using shower properties
- **Tool:** Boosted decision tree (BDT)
  $\Rightarrow$ Machine learning tool to classify events in a sample
- BDT is implemented in TMVA (Toolkit for Multivariate Analysis), for more information see TMVA manual

Method continued

Trees: Consist of subsequent yes/no ("is proton?") decisions based on cuts on variables



"Growing" the tree:

- BDT finds typical properties for protons/π^+ based on training sample with known properties (cut on variable)
- Training sample split at each node into two sub-samples: pion- or proton-like
- Procedure repeated until sample has highest purity or minimum size

- "Leaf node" defined as proton node if contains more proton events from the training sample
- Incorrecly classified events are given larger weight $\Rightarrow$ reweighting $\Rightarrow$ new BDTs are "grown" ("boosting")
- Apply BDTs on testing sample $\Rightarrow$ classification according to how often event ended up in signal and background nodes
- Each event assigned a BDT value dependent whether if event more pion-/proton-like $\Rightarrow$ BDT proposes optimal cut for highest purity

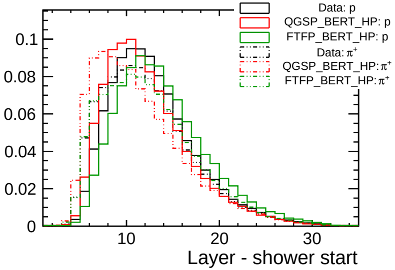
Approach

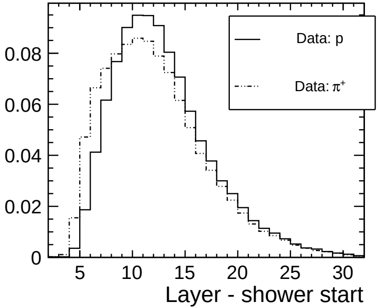
- Proton and pion showers are very similar ⇒ separation very challenging
- Distinction based on 30 variables, describing shower properties and shower shape
- The larger the difference proton/pion, the more important the variable in the BDT
- Important variables are used at many decision nodes, contribute therefore much to the separation
- Unimportant and strongly correlated variables are ignored by the BDT
- Sample beam energy = 60 GeV (highest purity for protons and pions was archived in measurement for 60 GeV, response of Cherenkov-Detector energy dependent) LCD-Note-2013-006

Combinations of training and testing samples used in these studies

training sample	testing sample
QGSP_BERT_HP	QGSP_BERT_HP
FTFP_BERT_HP	FTFP_BERT_HP
data	data
QGSP_BERT_HP	data

BDT input: Longitudinal shower depth @67%

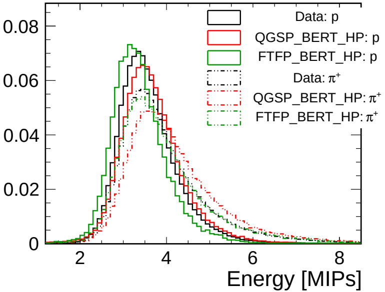


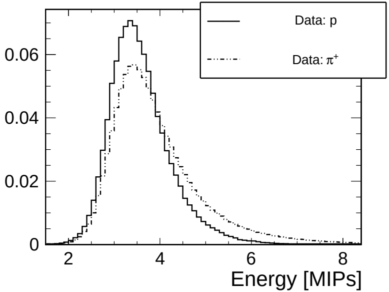


- Shower properties are very similar for protons and pions
- Difference between protons and pions is slightly larger in MC than in data

Note: Beam energy = 60 GeV

BDT input: Mean hit energy/cell

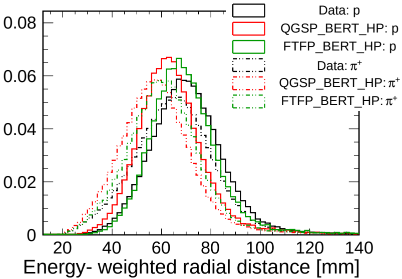


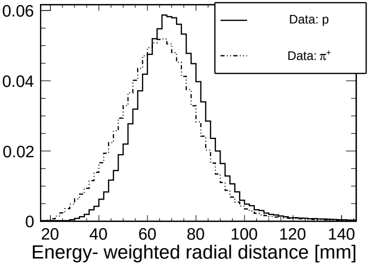


- By far important variable in separation (very large weight in all cases)
  $\Rightarrow$ Large impact on result
- Difference between protons and pions is slightly larger in MC than in data

Remark: Mean value of quantity X: $\bar{X} = \frac{\Sigma_{i=0}^{N} X_i}{N}$

BDT input: Energy weighted radial distance



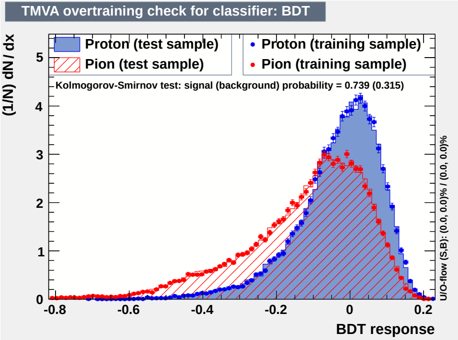


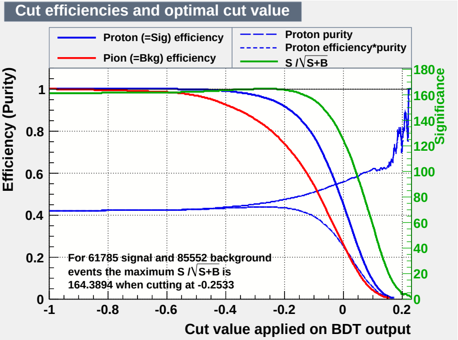
- Discrepancies between data and MC
- Important variable in most cases

Remark: Energy weighted radial distance:

$$d = \frac{\sum_{i=0}^{N_{\text{hits}}} E_i \sqrt{(x_i - x_{cog})^2 + (y_i - y_{cog})^2}}{\sum_{i=0}^{N_{\text{hits}}} E_i}$$

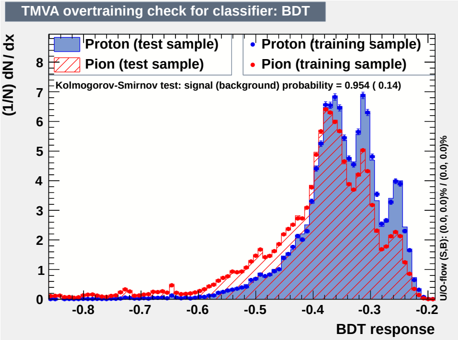
BDT-Response and Efficiency Cut: QGSP_BERT_HP training and testing

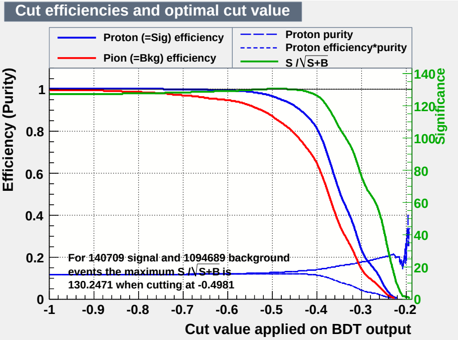




- BDT values for protons and pions similar
- But still: clear difference between proton efficiency (95%) and pion efficiency (80%) @ optimal cut value
  $\Rightarrow$ possible to accumulate protons in sample
- Results similar for FTFP_BERT_HP training and testing

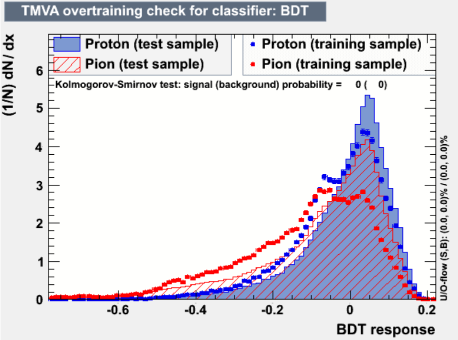
BDT-Response and Efficiency Cut: Data training and testing

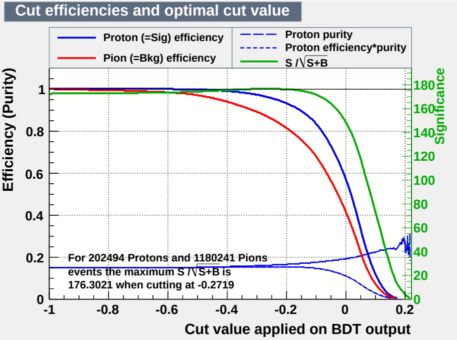




- In data, the difference in BDT-response is lower than in the case of MC
- The potential to separate Protons and Pions is overestimated in both MC
- But still: clear difference between proton efficiency (97%) and pion efficiency (87%) @ optimal cut value **still visible**
  $\Rightarrow$ possible to accumulate protons in sample

BDT-Response and Efficiency Cut: QGSP_BERT_HP training, Data testing





- BDT response plot shows discrepancy between data and MC
- Caution: MC training information in disagreement with data ⇒ possible bias in data selection
- Proton efficiency: 97%, pion efficiency: 87%

Conclusions

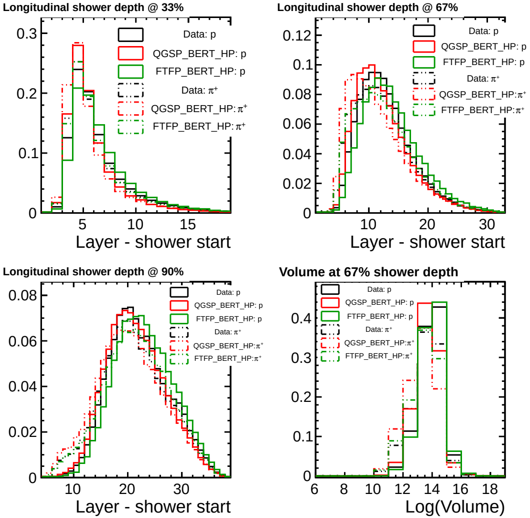
- Investigated longitudinal and radial fluctuations of hadronic showers
- For new variables, data and MC agree on percent level
- Agreement MC/data model and particle dependent

- Analyzed whether it is possible to separate protons and pions only using information provided by the HCAL
- Protons and pions very similar
  $\Rightarrow$ for reliable distinction use information of other parts of the detector
  $\Rightarrow$ Nevertheless, it is possible to accumulate protons/pions in sample

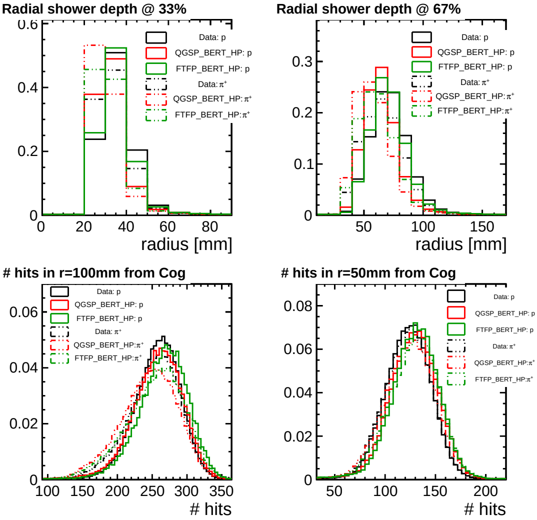
- MC shows larger difference between p/π^+ $\Rightarrow$ p/π^+ separation harder in reality than predicted in MC

Thank you for your attention!

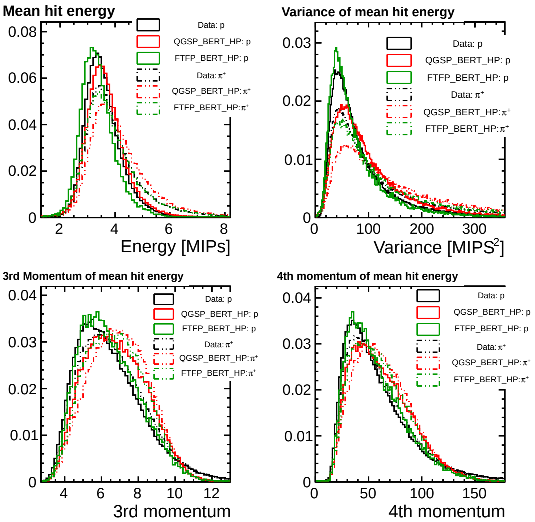
BDT variables part 1



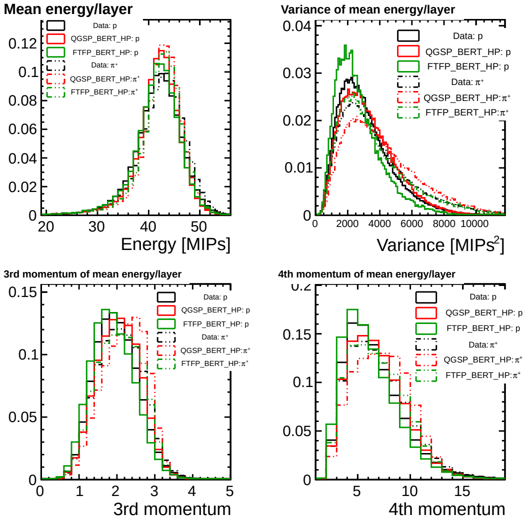
BDT variables part 2



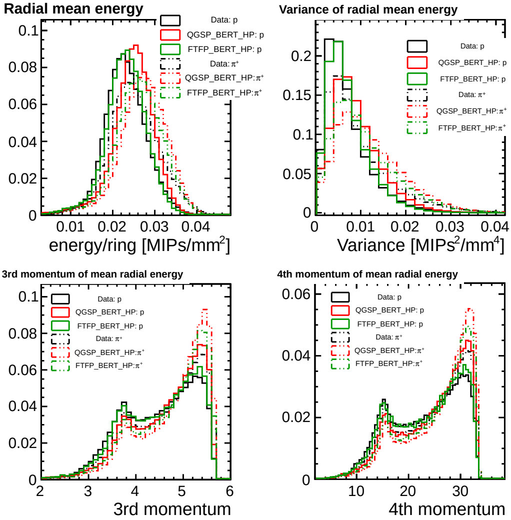
BDT variables part 3



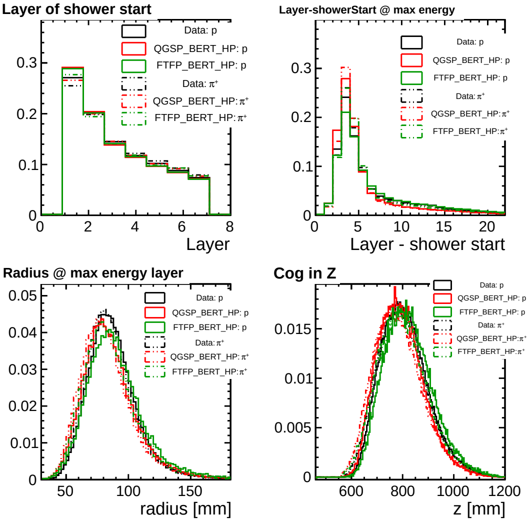
BDT variables part 4



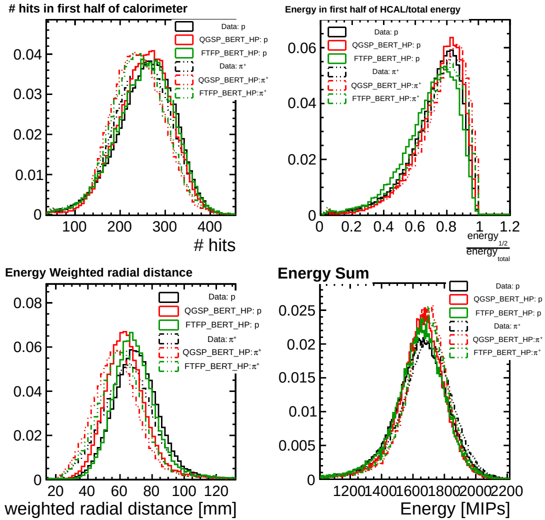
BDT variables part 5



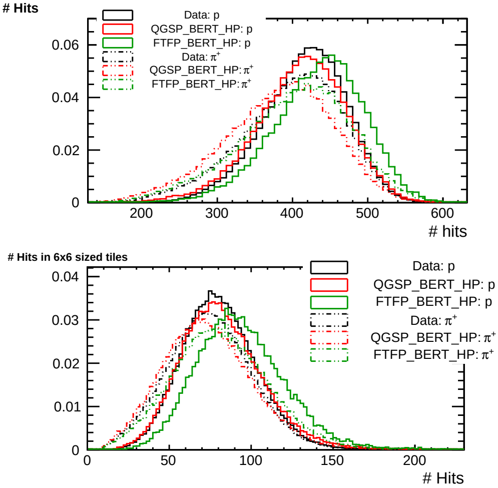
BDT variables part 6



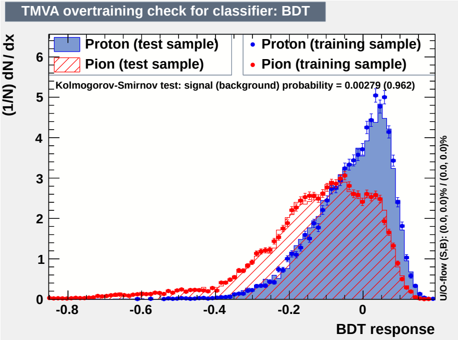
BDT variables part 7

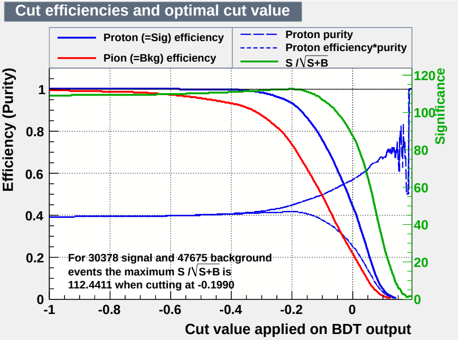


BDT variables part 8



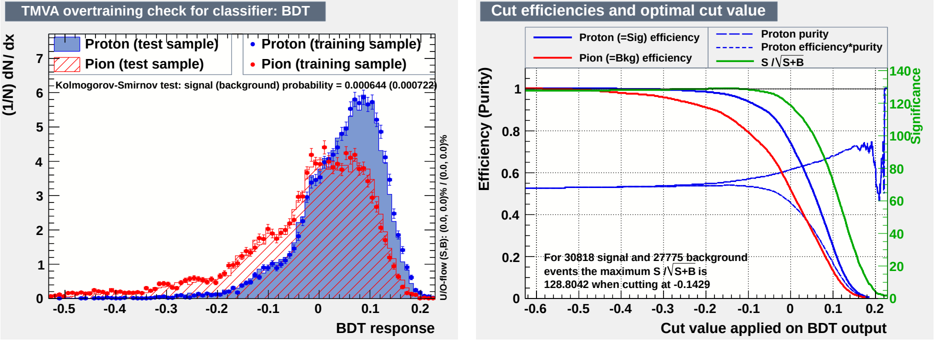
BDT-Response and Efficiency Cut: FTFP_BERT_HP training and testing





- proton efficiency = 95%, pion efficiency = 75% @ optimal cut value

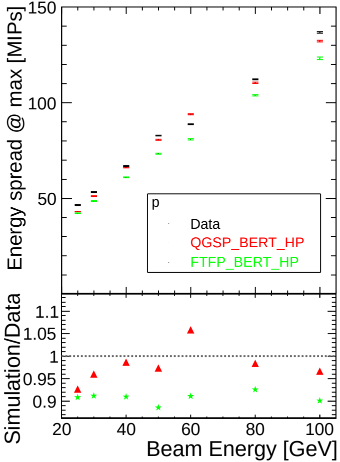
BDT-Response and Efficiency Cut: QGSP_BERT_HP training and testing, $E_{available}$



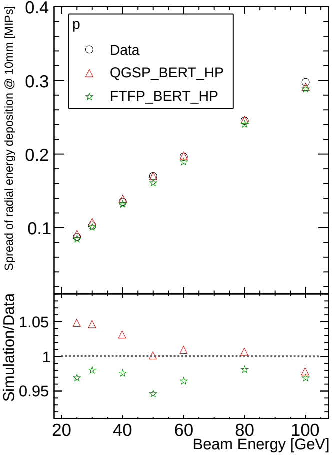
- Consider $E_{available}$: Proton does not decay in contrast to $\pi^+ \Rightarrow$ rest mass does not contribute in shower energy
- $E_{available} = 60 GeV$ for protons and pions
- proton efficiency: 97%, pion efficiency: 87%
- sample purity is lower compared with other MC samples $\Rightarrow$ $E_{available}$ seems to have some impact!

Longitudinal and radial shower fluctuations for protons

Longitudinal fluctuations
Spread @ maximum energy deposition:



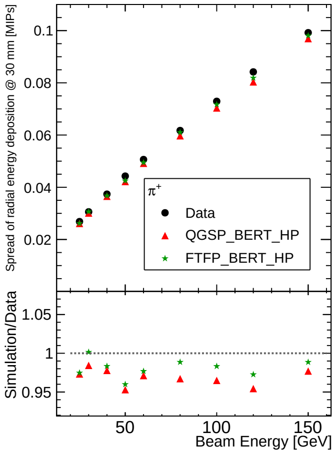
Radial fluctuations
Spread @ center:



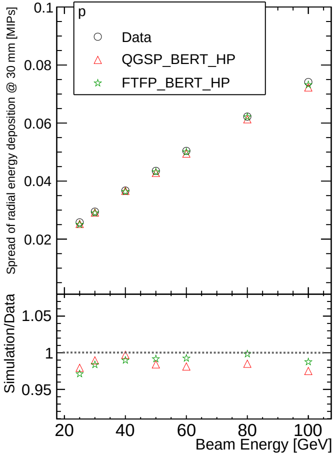
Radial shower fluctuations for higher radius

Spread of energy deposition in ring with r =20-30mm from center of gravity

Pions:



Protons:



Layer of shower start