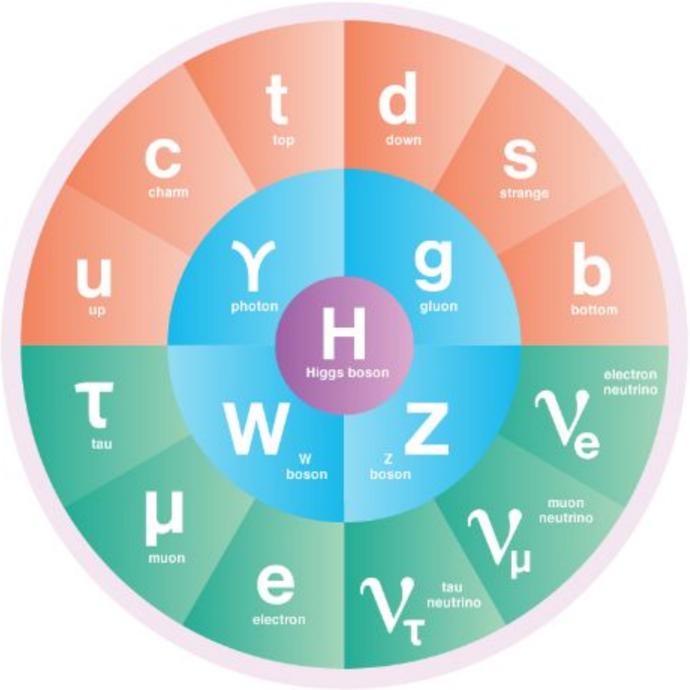
CSI : Improvement of the electron detection performance with AI for the ATLAS experiment measurements of the Drell-Yan scattering at high dilepton masses

Maxime Zumbihl, LAPP, IN2P3, CNRS
Supervision : Tetiana Berger-Hryn'ova (LAPP), Alexandre Benoit (LISTIC)

LISTIC ATLAS EXPERIMENT LAPP Laboratoire d'Annecy de Physique des Particules

Funded by the European Union
European Research Council Established by the European Commission

Standard Model and Beyond



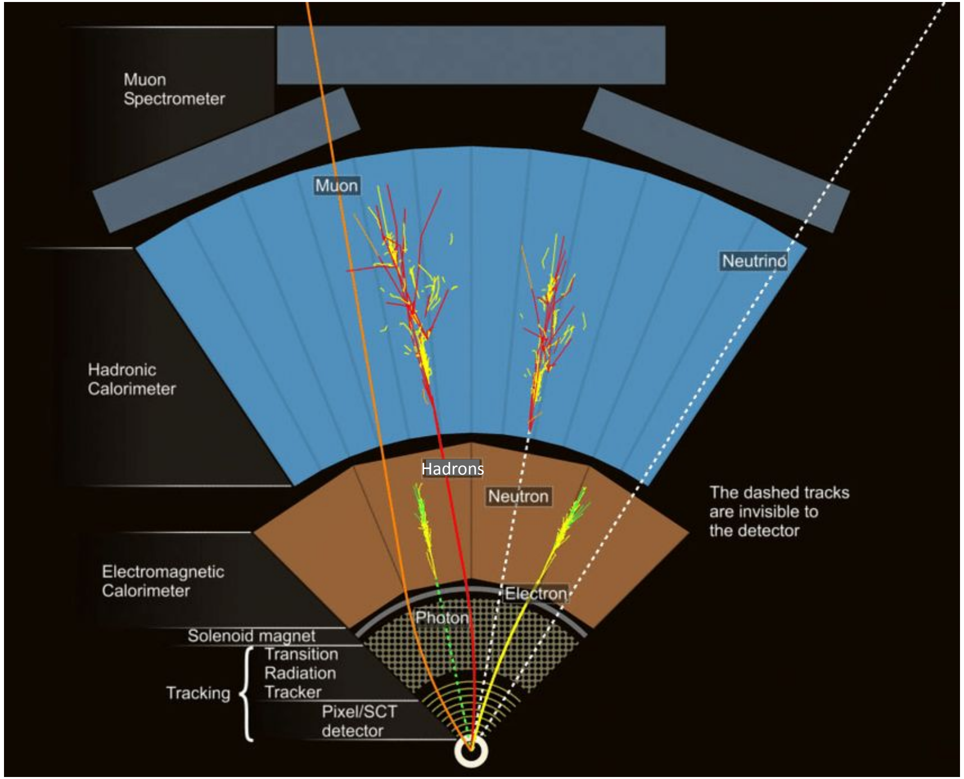
- Description of elementary particles and their interactions
- Incomplete model:
  - Dark matter?
  - Gravitation?
  - Matter-antimatter asymmetry?

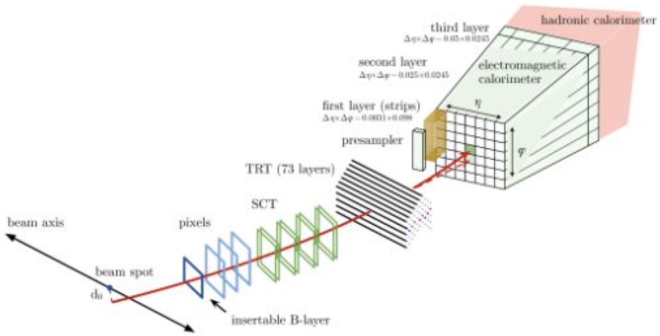
→ Necessity to go beyond SM

Improvement of Electron ID performance

- Electrons are important final state particles in the physics programme of the ATLAS experiment at the LHC
- Found in many final states of proton-proton collisions
- Important for study of the Standard model, but also for measurements and searches of physics Beyond the Standard Model (BSM)
- Improving electron identification → Better sensitivity to new physics

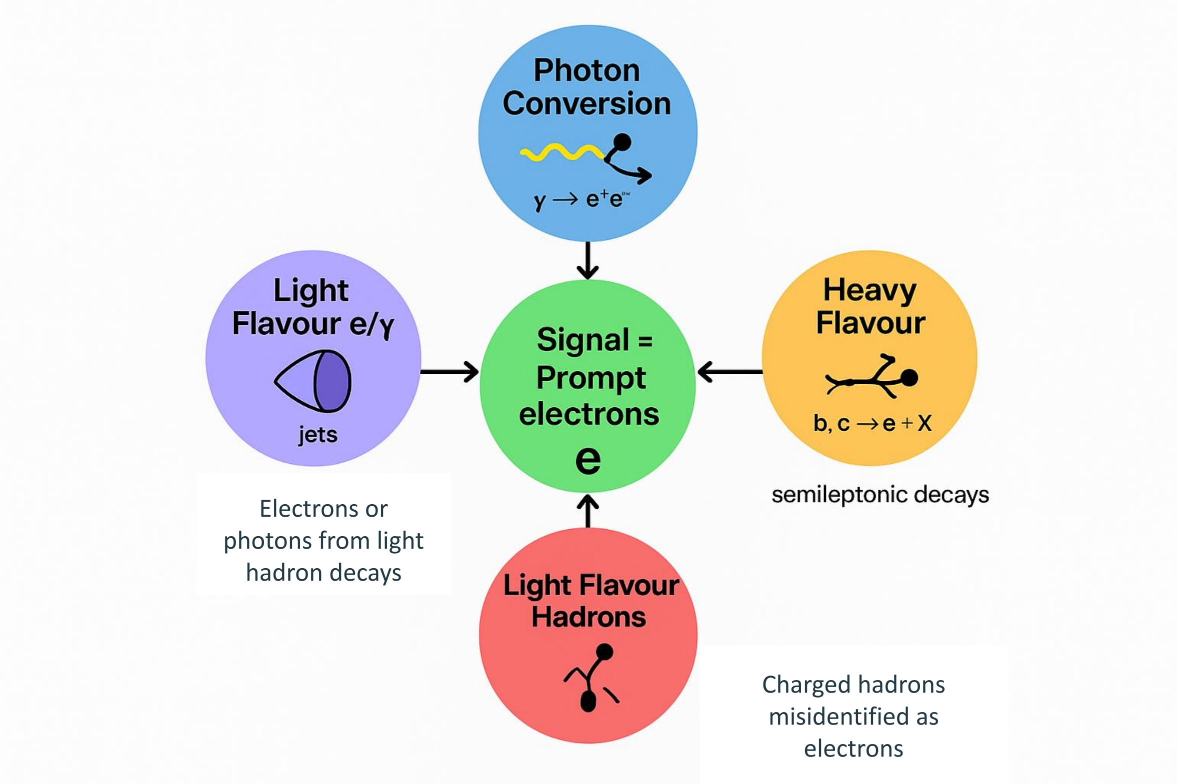
ATLAS Detector : Electromagnetic shower shape





- Important subdetectors for electron identification:
  - Electromagnetic calorimeter (energy)
  - Inner detector (tracks)
- Production of new particles after interaction with the detector = Electromagnetic shower shape
- Necessity to distinguish Signal (= Electrons) from background

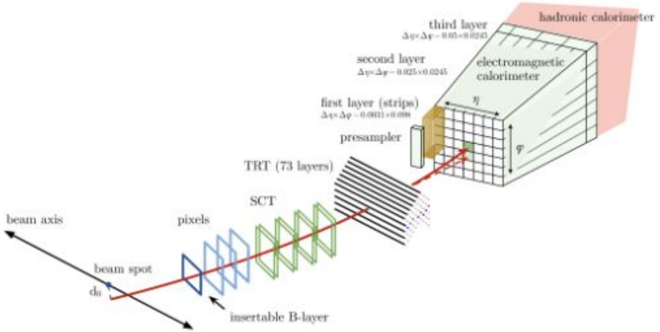
Background types

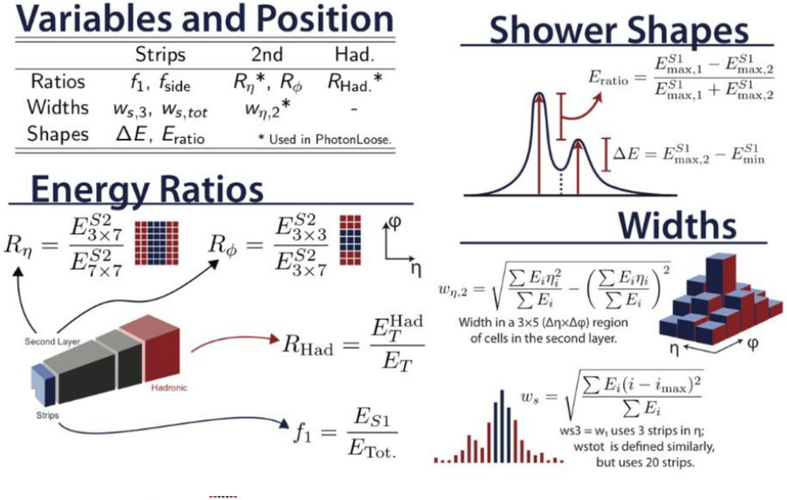


Electron reconstruction

- Electrons reconstructed using energy deposits (clusters) in the calorimeter and tracks.
- Track-Cluster matching

- Identification → Variables designed to distinguish signal from background.





Likelihood Identification

Identification method used so far ⟶ Likelihood (LH) ID :

- 1D Probabilty Density Functions (PDFs) are formed for all input variables without taking correlations into account, separately for signal, L_s and background L_b :

$$L_{\mathrm{S(B)}}(\mathbf{x}) = \prod_i P_{\mathrm{S(B)},i}(x_i)$$

- Signal = Prompt electrons, Background = jets mimicking the signature of prompt electrons, electrons from photon conversions, and non-prompt electrons from the decay of hadrons containing heavy flavours
- We formed, for each electron candidate a discriminant :

$$d_{\mathrm{L}} = \frac{L_{\mathrm{S}}}{L_{\mathrm{S}} + L_{\mathrm{B}}}$$

- Electrons are selected by applying cuts on the LH discriminant
- Different working points (VeryLoose, Loose, Medium, Tight), corresponding to increasing thresholds on the discriminant, are designed to achieve target identification efficiencies

Deep Neural Network

- A new technique of identification based on DNN is developed to improve electron ID, and eventually replace the LH-based ID method.
- Output layer has 6 outputs:
  $\longrightarrow$ signal (prompt electrons) + 5 background classes : Charge Flip (though considered in many cases as signal), Photon Conversion, heavy flavour, Light Flavor e/γ , Light Flavour Hadrons
- Advantage of multinomial classification: Defining different discriminants optimised to reject a specific background
- Disadvantage: Inability to train on real data for all classes (yet…)

$$\mathcal{D}_{\mathrm{el}} = \ln\left(\frac{f_{\mathrm{El}}p_{\mathrm{El}} + (1 - f_{\mathrm{El}})p_{\mathrm{CF}}}{f_{\mathrm{PC}}p_{\mathrm{PC}} + f_{\mathrm{HF}}p_{\mathrm{HF}} + f_{\mathrm{LFEg}}p_{\mathrm{LFEg}} + (1 - f_{\mathrm{PC}} - f_{\mathrm{HF}} - f_{\mathrm{LFEg}})p_{\mathrm{LFH}}}\right)$$

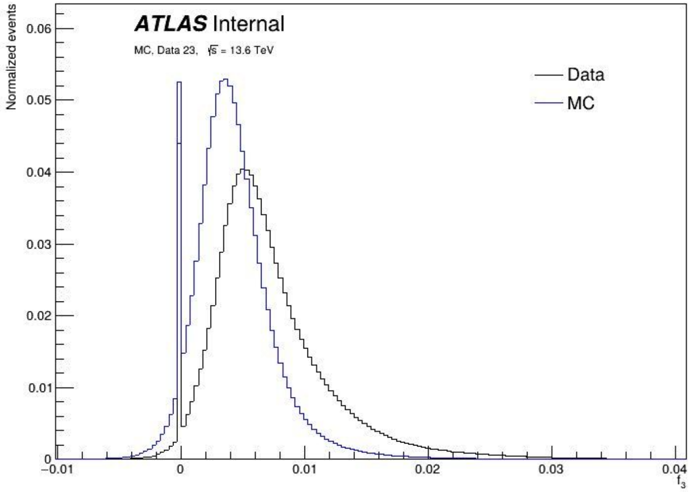
p_X: output scores; f_X: free parameters, can be optimised to enhance rejection of specific backgrounds

Monte-Carlo simulation

- Discrepancies exist in shower shape variables distributions between data and Monte-Carlo simulation

- Corrections are needed, for instance to optimize DNN identification algorithm, since it is trained on MC samples. If corrections are applied to variables, they are called fudged.

- How well should we correct these distributions? Which remaining level of discrepancy is acceptable (without impact on performance)? → Need of detailed studies of these shower shape variables



Shower Shape Studies

- MC/Data comparisons for inclusive distributions (for electrons with p_T > 15 GeV, and $|\eta|$ < 2.47 to cover tracker and EM calorimeter acceptance, Very Loose ID working point) to check the level of agreement between fudged MC and data. For this presentation, focus on Zee signal and f_3 (Ratio of the energy in the third layer to the total energy in the EM calorimeter).

- Datasets considered for these studies : Data 2022, Data 2023 and Data 2024 + Corresponding MC simulations

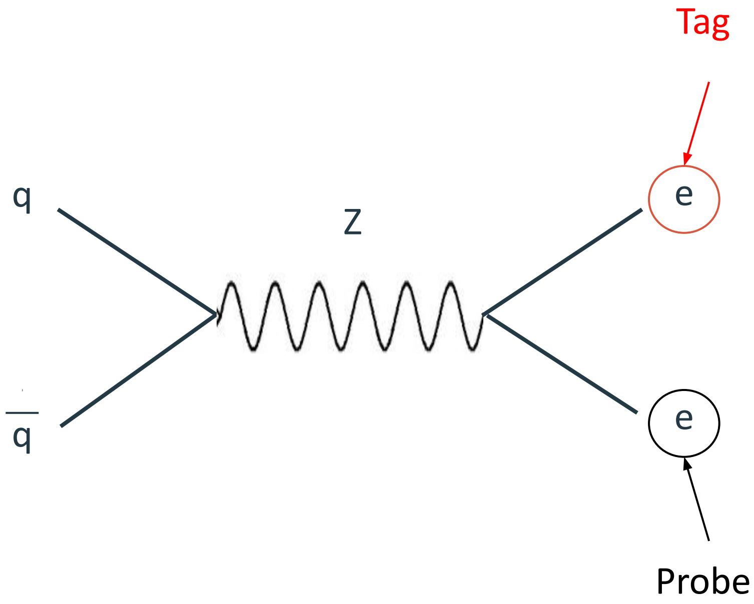
- MC corrected using set of run 3 Fudge Factors

- Study of kinematical dependency :

  - Statistical distance
  - Parameters evolution

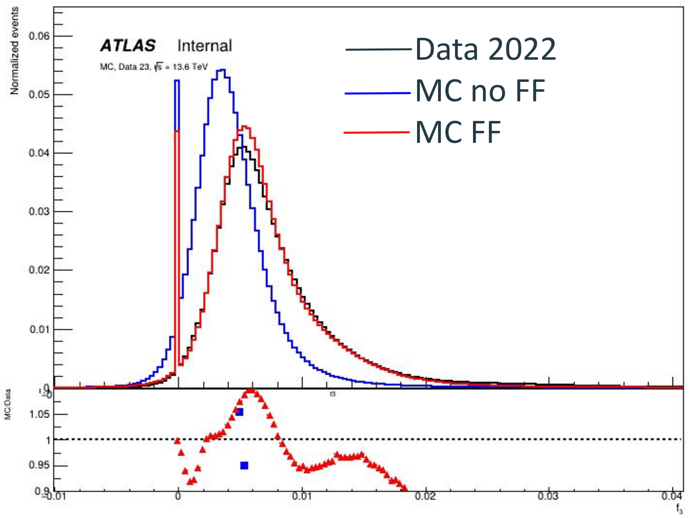
Event Selection : Signal and Tag&Probe Method

- $Z \rightarrow ee$ process :
  - Two isolated, opposite-sign, high-p_T electrons
  - Well-Known resonance (m_Z~ 91 GeV)
- Tag&Probe method:
  - Select an event around Zee resonance (invariant mass m_{ee} close to m_Z)
  - Tag = electron passing tight identification requirement
  - **Probe** = electron with no identification criteria (no biais). Used to compute efficiencies
- Efficiency depends on p_T and η



Data/MC Comparison : f_3

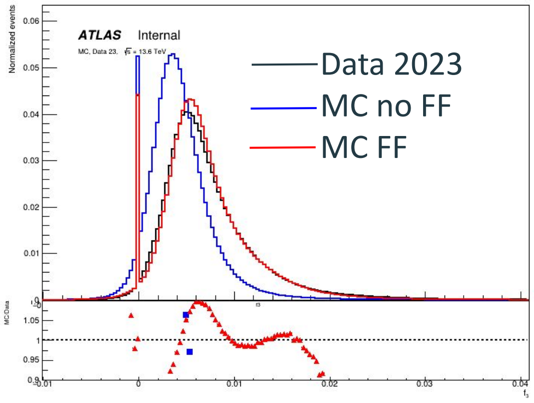
2022



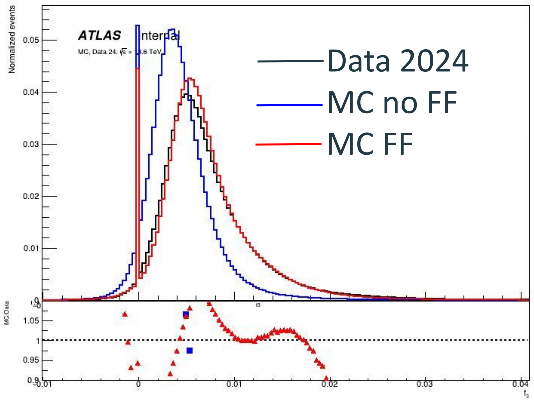
- Significant discrepancy between unfudged MC and Data.
- Applying the fudging correction significantly improves the agreement in the peak region, reducing the discrepancy to about 10% overestimation at the peak maximum and less than 5% in some areas around it.
- A large disagreement between the distributions remains in the lower statistics regions (tails)

Data/MC Comparison : f_3

2023



2024



- Significant discrepancy between unfudged MC and Data.
- Applying the fudging correction significantly improves the agreement in the peak region, reducing the discrepancy to about 10% overestimation at the peak maximum and less than 5% in some areas around it.
- A large disagreement between the distributions remains in the lower statistics regions (tails)

Statistical Distances

A way to quantify the differences between distributions : Statistical distances, that give the distance between two statistical objects, like probability distributions.

Many metrics can be used, sensitive to different aspects of the probability distribution variations (tails, mean, peaks, ...)

- "L1" distance that represents the total shift needed to transform one distribution into another, making it sensitive to the parameters of the distribution, especially **mean and peak position**:

$$d_{\mathrm{L1}}(\mathbf{obs}, \mathbf{ref}) = \sum_{i=1}^{n} |x_i - y_i|$$

- Hellinger distance : For two Probability Distribution Functions P and Q,

$$H(P,Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_i \left(\sqrt{P(i)} - \sqrt{Q(i)}\right)^2}.$$

Particularly sensitive to small probability region changes, making it useful to compare **tails distributions**.

Statistical Distances

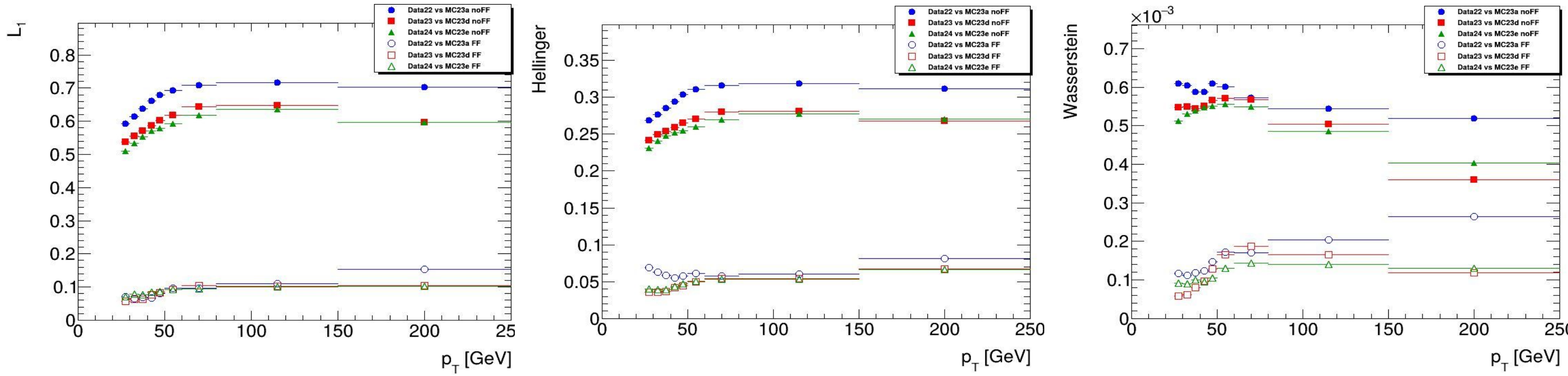
- Wasserstein Distance : For 1D distributions with cumulative distribution functions F and G :

$$W_1(P,Q) = \int_{-\infty}^{\infty} |F(x) - G(x)|\, dx$$

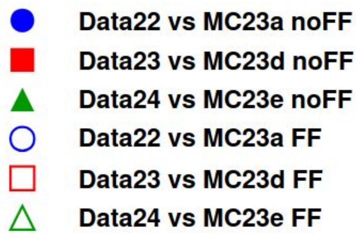
It can be understood as the minimal cost to move one distribution to another

These distances can also be used to derive MC/Data corrections using ML-based techniques (cost functions to minimize in Optimal transport for instance)

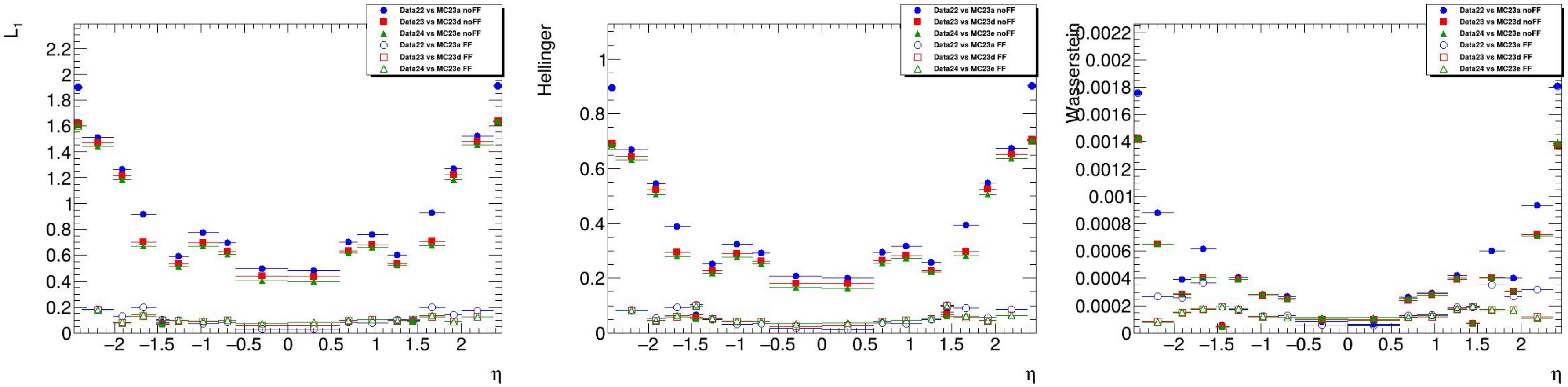
Statistical Distance : f_3



- Distance much lower for MC fudged than unfudged, confirming the good correction introduced by the fudge factors
- Higher differences in the low statistics p_T region (p_T > 80 GeV) for the fudged MC (except for Wasserstein)
- Differences between 2022 and 2023 datasets : Larger discrepancies Data22/MC23a unfudged than Data23/MC23d, but these differences tend to disappear when the fudging is applied

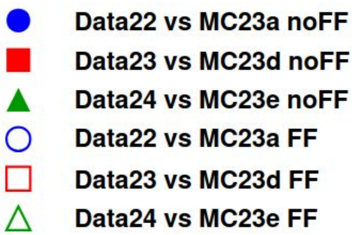


Statistical Distance : f_3



- Distance much lower for MC fudged than unfudged, confirming the good correction introduced by the fudge factors, except in the last bin, not even corrected

- Better agreement between the distributions in the central η region

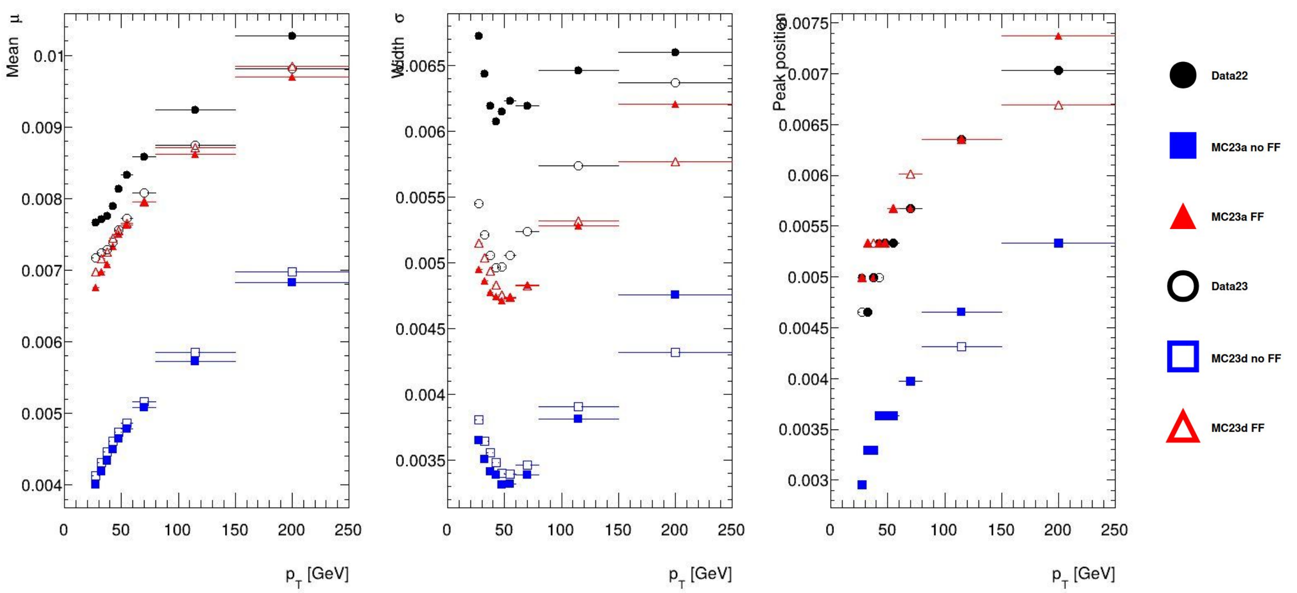


Parameters Evolution

A simple way to characterise the shower shape variables evolution as a function of p_T and η : Depict the evolution of the distributions' parameters : the mean μ, the width σ and the peak position, obtained by a gaussian fit

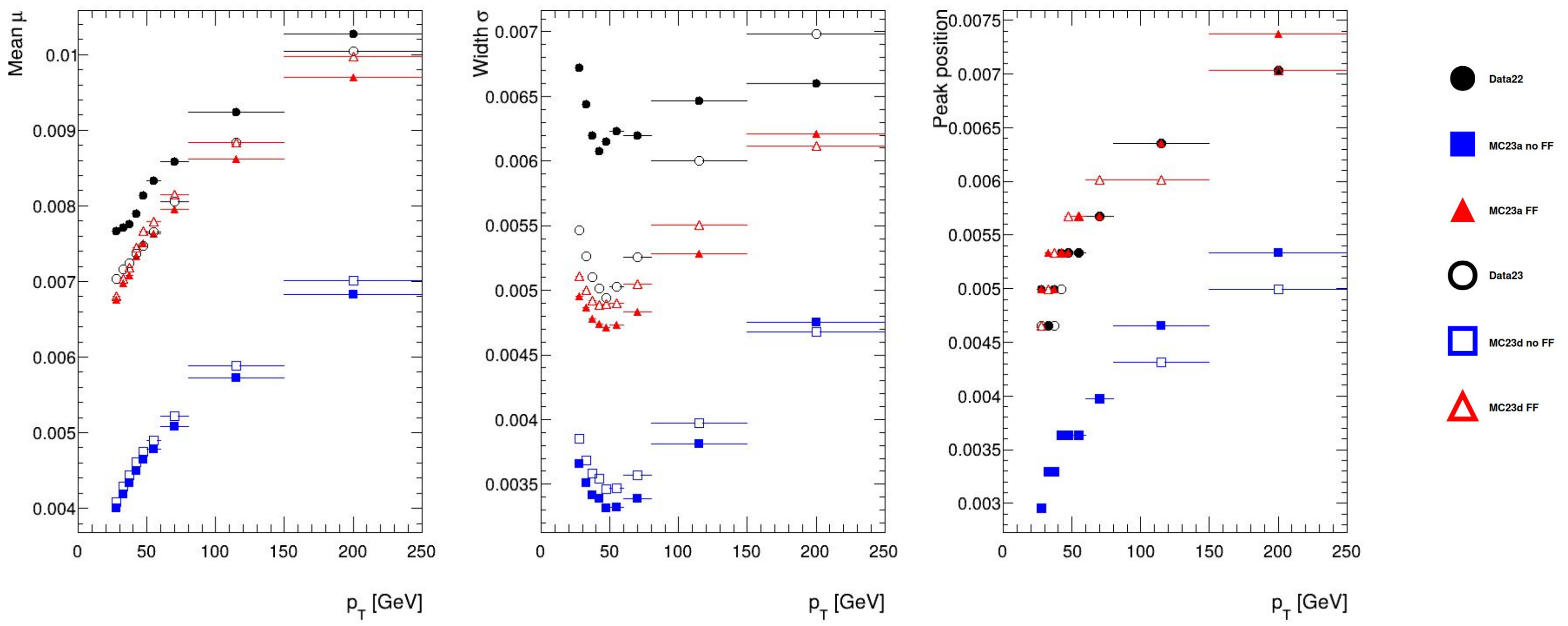
Using these parameters could be an easier way to extrapolate corrections, at high p_T for example, a region with less data available.

Parameters Evolution : f_3



- Differences between data 2022 and data 2023
- Fudged distributions (red) close to data 2023 → Consistent with the distances plots

Parameters Evolution (2024): f_3



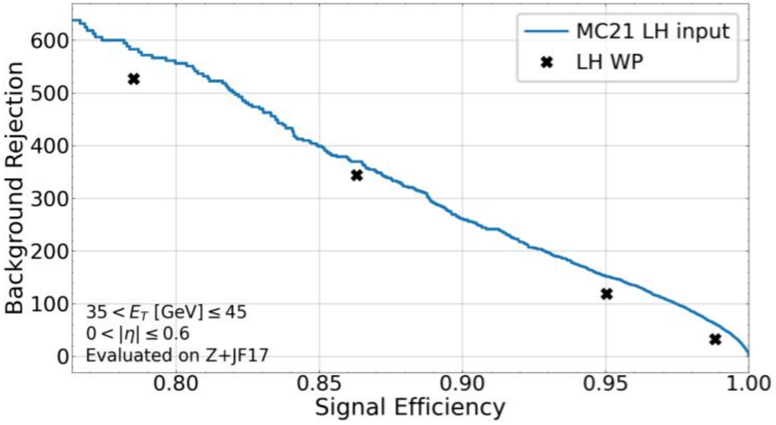
Improvement of DNN algorithm

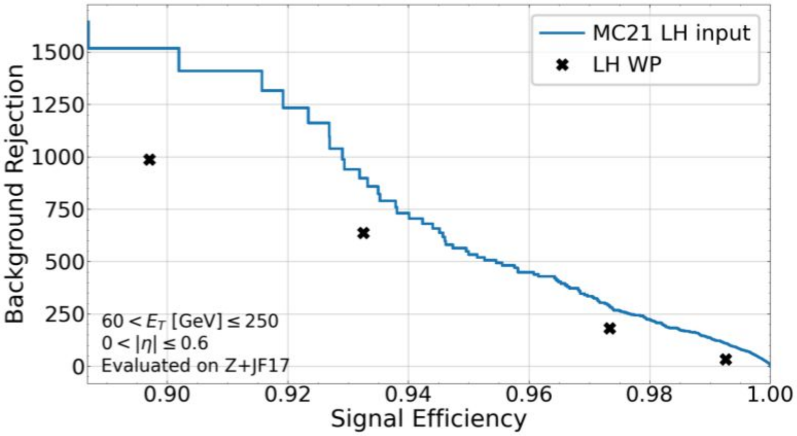
- Objective : Ensure more robust DNN electron identification performance (in collaboration with LISTIC)

- First step : Re-implement and update the current DNN (from Tensorflow to pytorch) in a way that it can improve flexibility for future developments

Validation : Performances

Nihal Brahimi





DNN implementation : Next steps

- Several leads to enhance the network architecture:
  - Conserve MC samples as inputs to the DNN but with a preliminary step consisting in correcting MC to data using optimal transport (no fudging needed anymore)
  - Adopt an unsupervised approach using data as inputs for the training of the DNN $\longrightarrow$ Need a good understanding of the background (Background shower shape variables, Timoty Duong)
  - Design a two-step network : First, a binary classification (Signal vs Background) and, second, multiclassification with 6 outputs.

Training and Scientific Production

- Trainings :
  - Scientific Python (12h, Cross-disciplinary training)
  - Numerical Optimization : theory and application (20h, Scientific training)
  - Research ethics (15h, Cross-disciplinary training)
  - Scientific integrity (15h, Cross-disciplinary training)
  - Summer school Machine learning for Physics, Ljubljana, Slovenia

- Scientific production :
  - Presentation during EGamma Workshop
  - These studies are being compiled in a public note
  - Shifts at the Calo/Fwd desk in the ATLAS control room

Backup

DNN Input Variables

- R_{had} : Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster
- R_{had1}: Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster
- f_3: Ratio of the energy in the third layer to the total energy in the EM calorimeter
- $w_{\eta 2}$: Lateral shower width
- R_ϕ: Ratio of the energy in 3x3 cells over the energy in 3x7 cells centred at the electron cluster position
- R_η: Ratio of the energy in 3x7 cells over the energy in 7x7 cells centred at the electron cluster position
- E_{ratio}: Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies
- f_1: Ratio of the energy in the first layer to the total energy in the EM calorimeter
- eProbabilityHT (TRTPID): Likelihood probability based on transition radiation in the TRT.
- $\Delta\eta_1$: $\Delta\eta$ between the cluster position in the first layer and the matching extrapolated track
- $\Delta\phi_{res}$: $\Delta\phi$ between the cluster position in the second layer of the EM calorimeter and the momentum-rescaled track, extrapolated from the perigee, times the charge q
- w_{stot} : Shower width

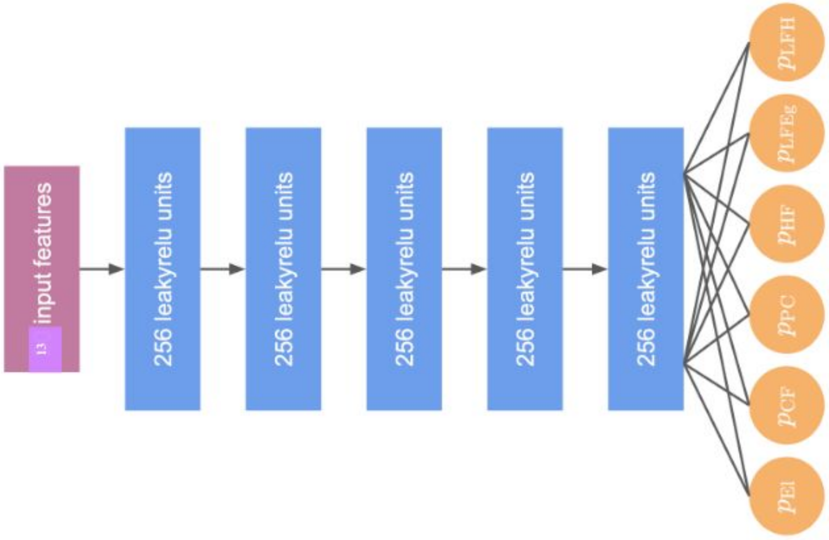
Shower shape variables

- d_0 : Transverse impact parameter relative to the beam-line
- $d_0/\sigma(d_0)$: Significance of transverse impact parameter defined as the ratio of d_0 to its uncertainty
- $\Delta p/p$: Momentum lost by the track between the perigee (~center of the detector) and the last measurement point divided by the momentum at perigee

Tracking variables

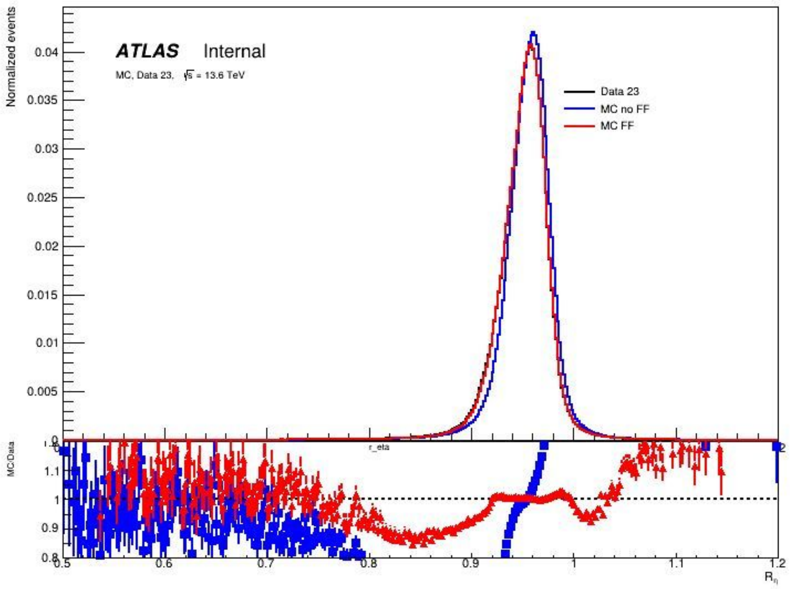
DNN architecture

- Implementation of DNN in pytorch
- 5 hidden layers with 256 nodes each
- 1 output layer with either 2 nodes (binary classification) or 6 nodes (multiclassification)
- Leakyrelu activation function
- L2 regularisation (Apply an additional parameter to the loss function to prevent from overfitting)
- Signal ($Z \rightarrow ee$) and background (JF17, ttbar*) MC (fudged)
- Six outputs classes : Prompt electrons, CF, PC, HF, LFEg, LFH or Two : Signal/Background
- 13 input features (Shower shape variables for trigger electrons) : Eratio, Reta, Rhad, Rhad1, Rphi, TRTPID, deltaEta1, deltaPhiRescaled, Et, eta, f1, f3 and Weta2
- Datasets preprocessed : Downsampling and reweighting



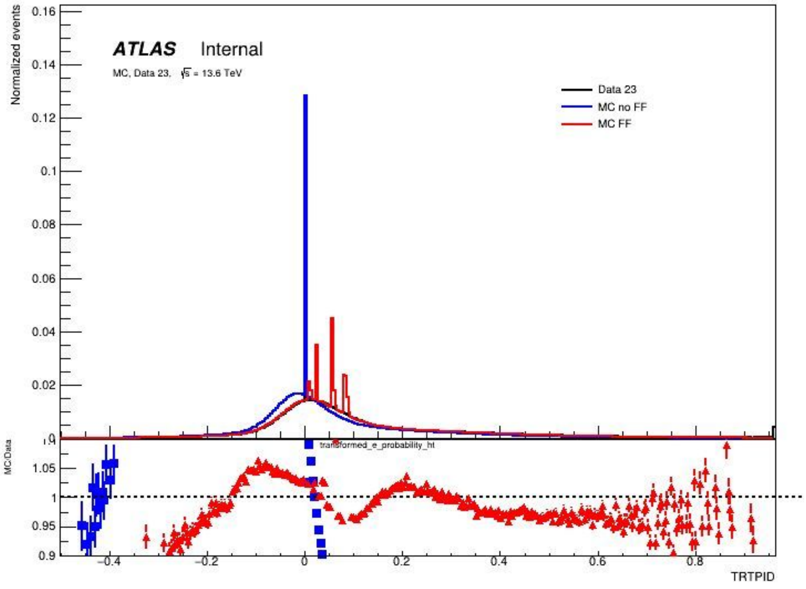
*Could be considered as signal

Data/MC Comparison : R_η



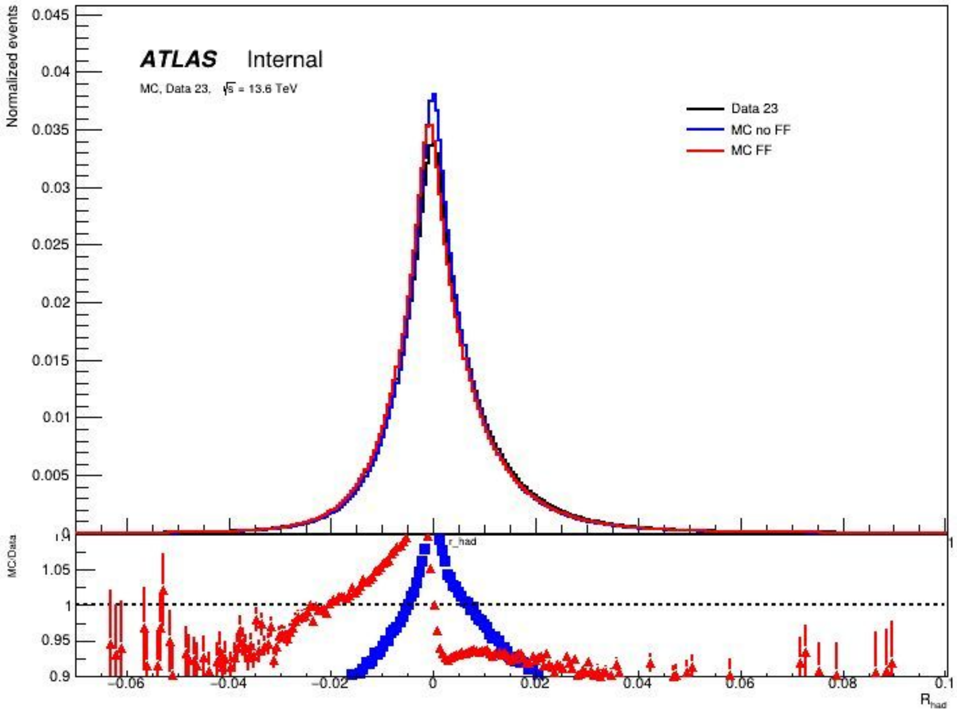
- Very good agreement around the maximum
- Remaining discrepancies in the tails of the distribution (up to 15%)

Data/MC Comparison : TRTPID

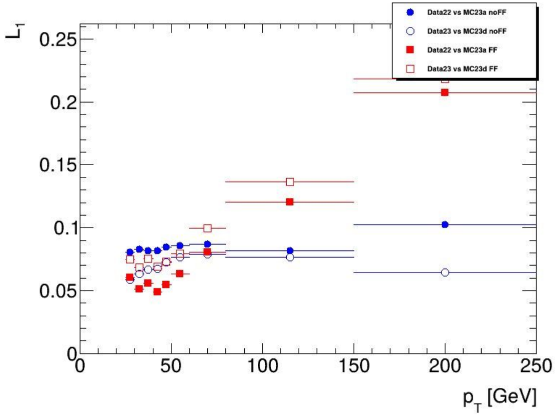


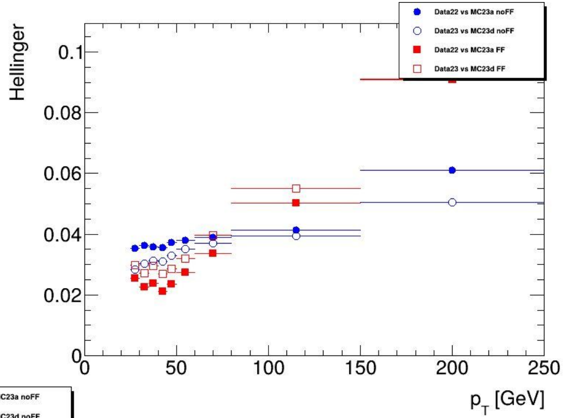
- Fudging of TRTPID introduced with run 3 FFs derivation
- Good agreement between data and MC fudged (within ~10%)
- Corrections introduce spikes

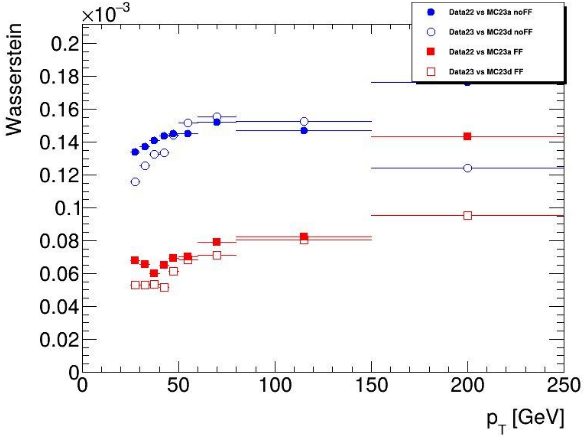
Data/MC Comparison : R_{had}



Statistical Distance : R_{had}







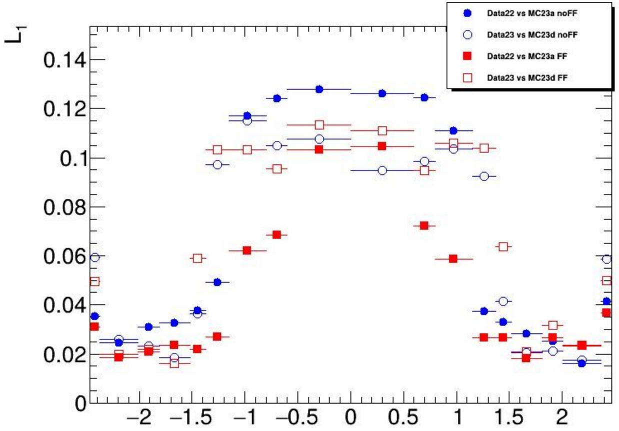
Data22 vs MC23a noFF

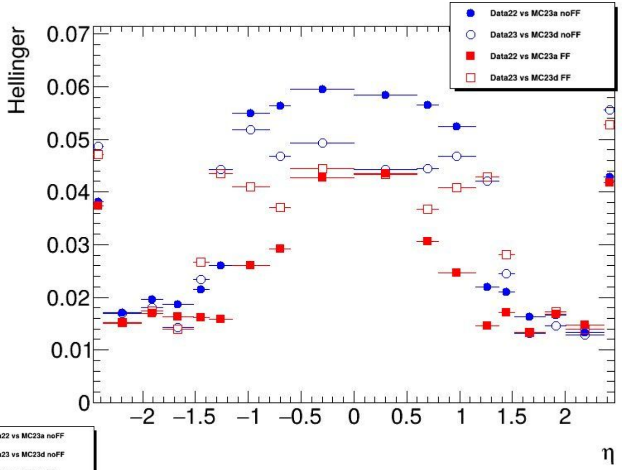
Data23 vs MC23d noFF

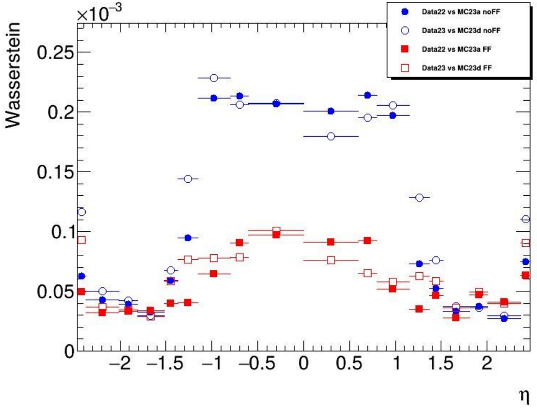
Data22 vs MC23a FF

Data23 vs MC23d FF

Statistical Distance : R_{had}







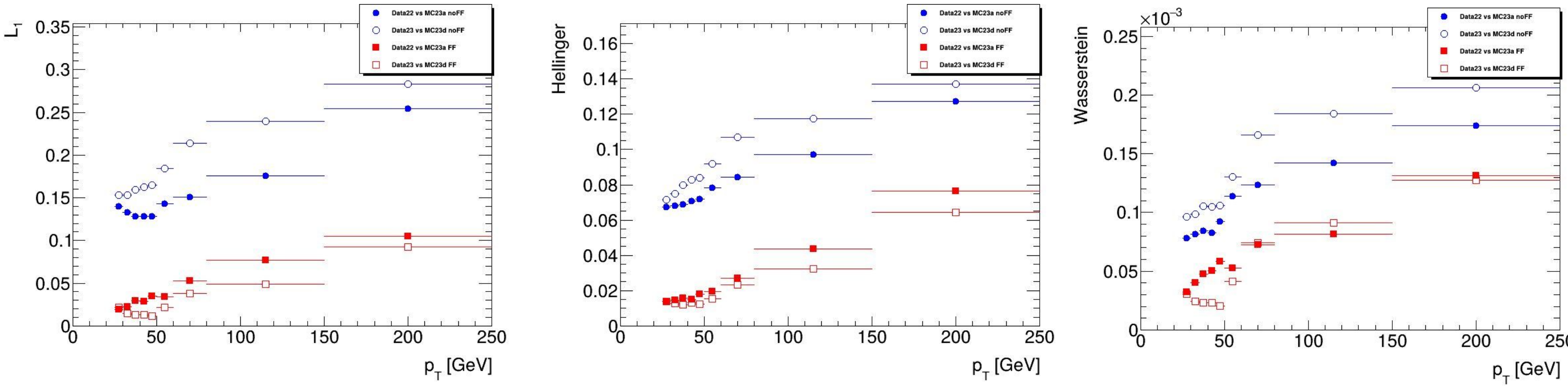
Data22 vs MC23a noFF

Data23 vs MC23d noFF

Data22 vs MC23a FF

Data23 vs MC23d FF

Statistical Distance : R_η



- As for f_3, distances increase with p_T
- Larger distances between data 23 and MC23d uncorrected than between data 22 and MC23a uncorrected but it is the opposite when the fudging is applied

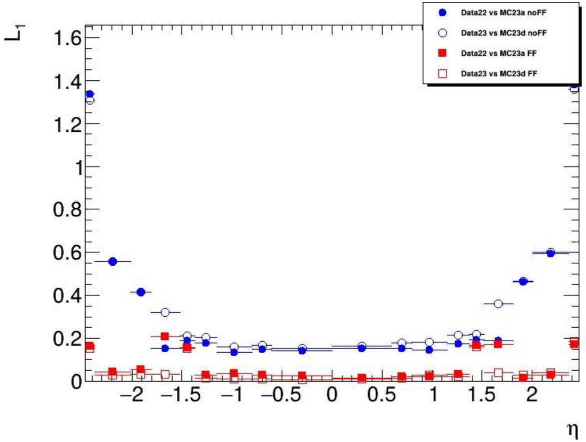
Data22 vs MC23a noFF

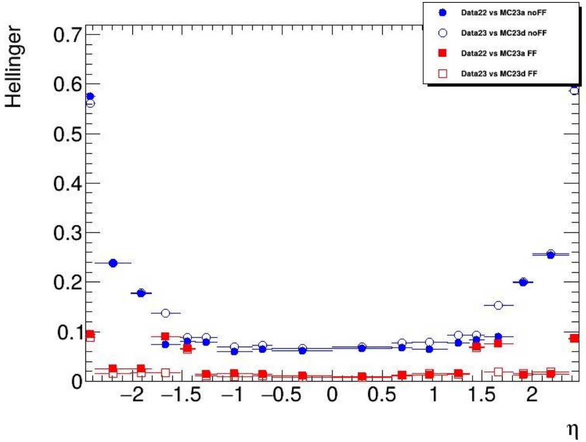
Data23 vs MC23d noFF

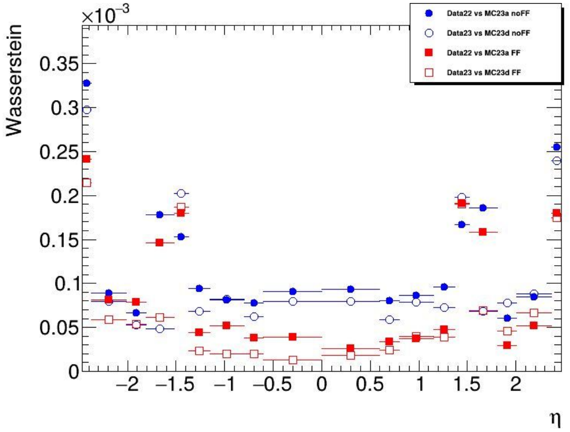
Data22 vs MC23a FF

Data23 vs MC23d FF

Statistical Distance : R_η







- For L_1 and Hellinger distances, no significant difference between the years
- R_η is corrected in the last bin of η

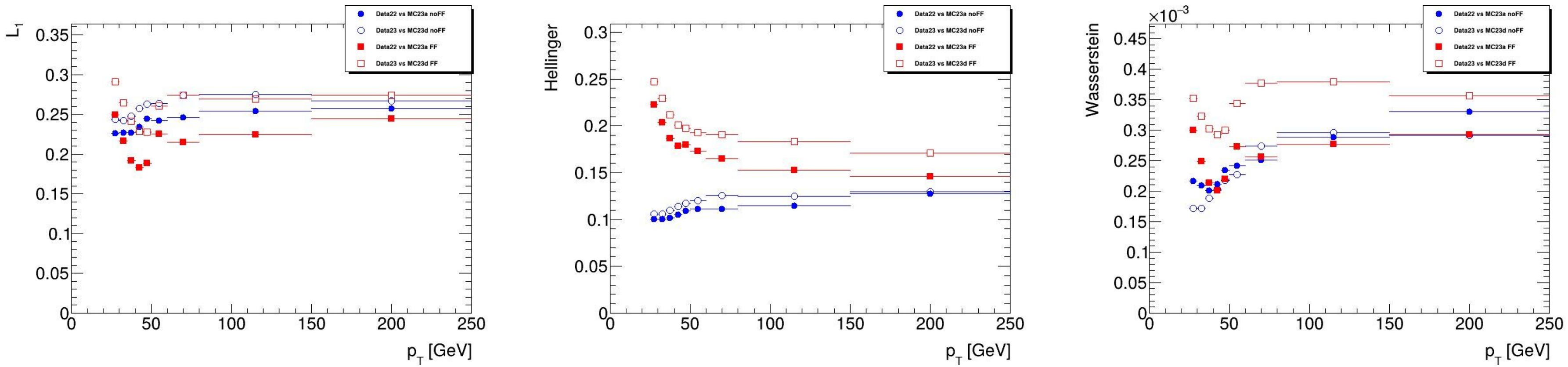
Data22 vs MC23a noFF

Data23 vs MC23d noFF

Data22 vs MC23a FF

Data23 vs MC23d FF

Statistical Distance : TRTPID



- L1 smaller for MC23a fudged w.r.t. to unfudged case. Not seen for MC23d
- Hellinger has an opposite behaviour : Smaller discrepancies in the unfudged case, especially at low p_T
- Mismodelling more pronounced in data 23/MC23d

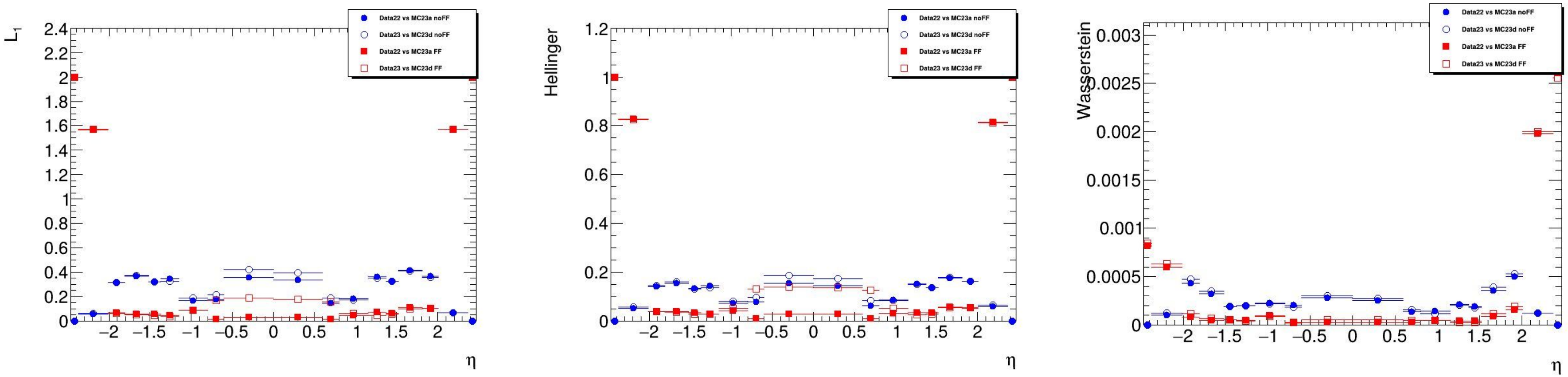
Data22 vs MC23a noFF

Data23 vs MC23d noFF

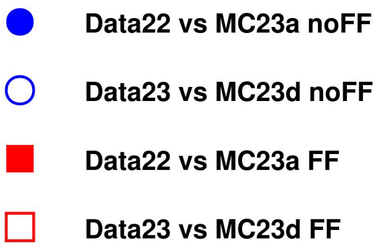
Data22 vs MC23a FF

Data23 vs MC23d FF

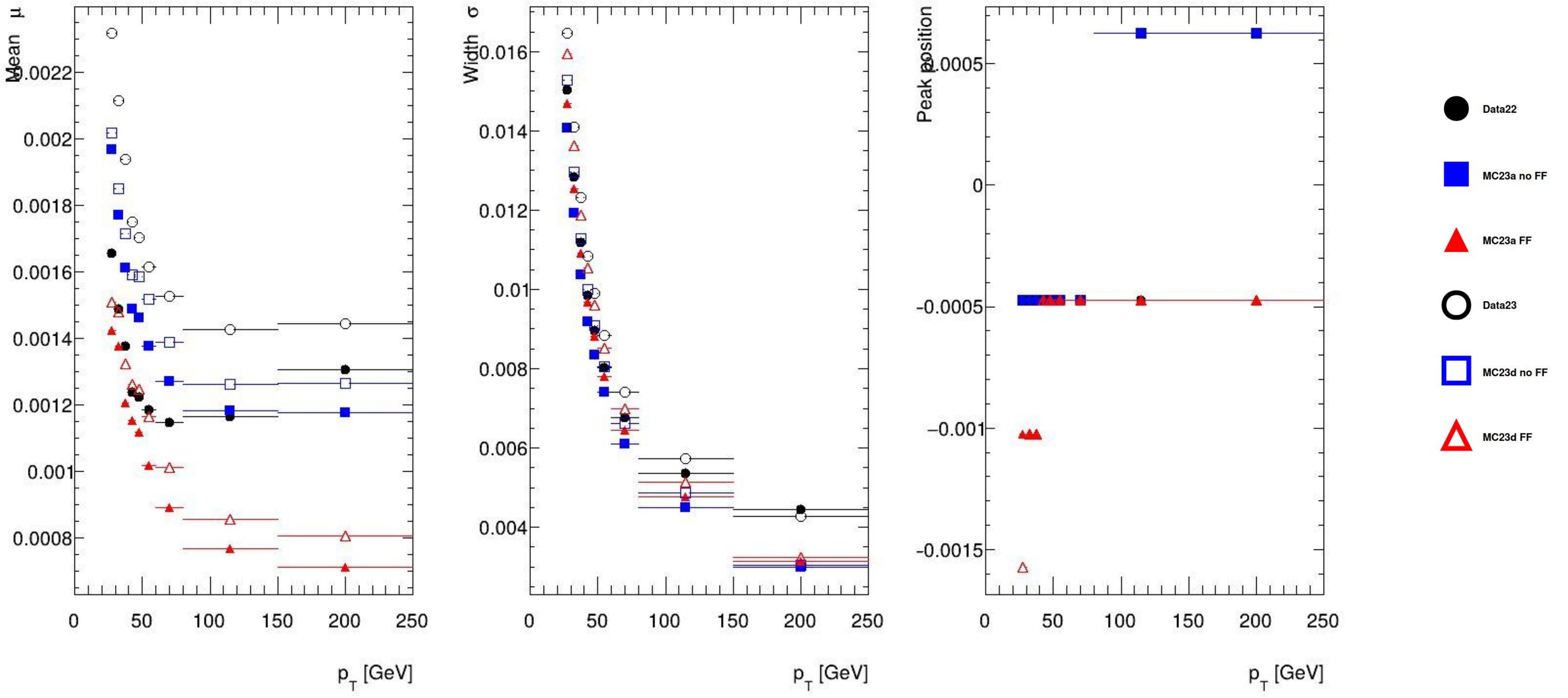
Statistical Distance : TRTPID



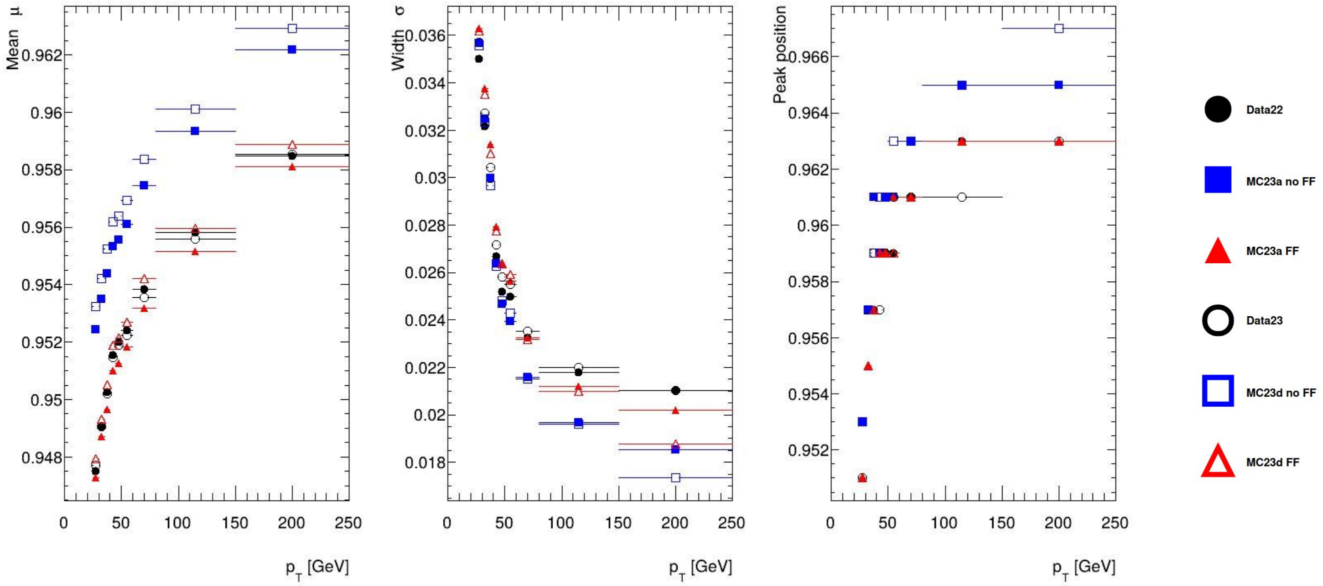
- TRT doesn't cover |η| > 2 region but fudging still applied and introduce a maximum discrepancy in the last two |η| bins for MC23a
- In the covering range of the TRT, Good agreement , especially for fudged MC23a



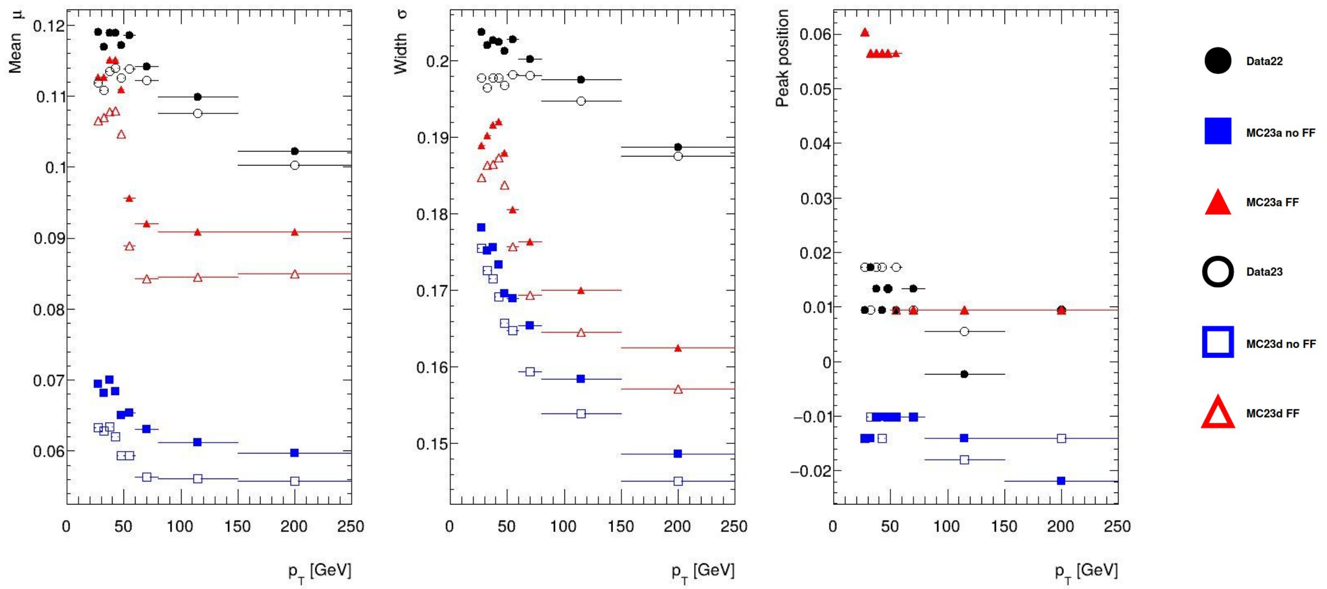
Parameters Evolution : R_{had}



Parameters Evolution : R_η



Parameters Evolution : TRTPID



- MC23a fudged closer to data than MC23d → Consistent with L1 distance