b-jet energy regression for the CMS experiment

Nadya Chernyavskaya - ETH Zurich

on behalf of the CMS collaboration

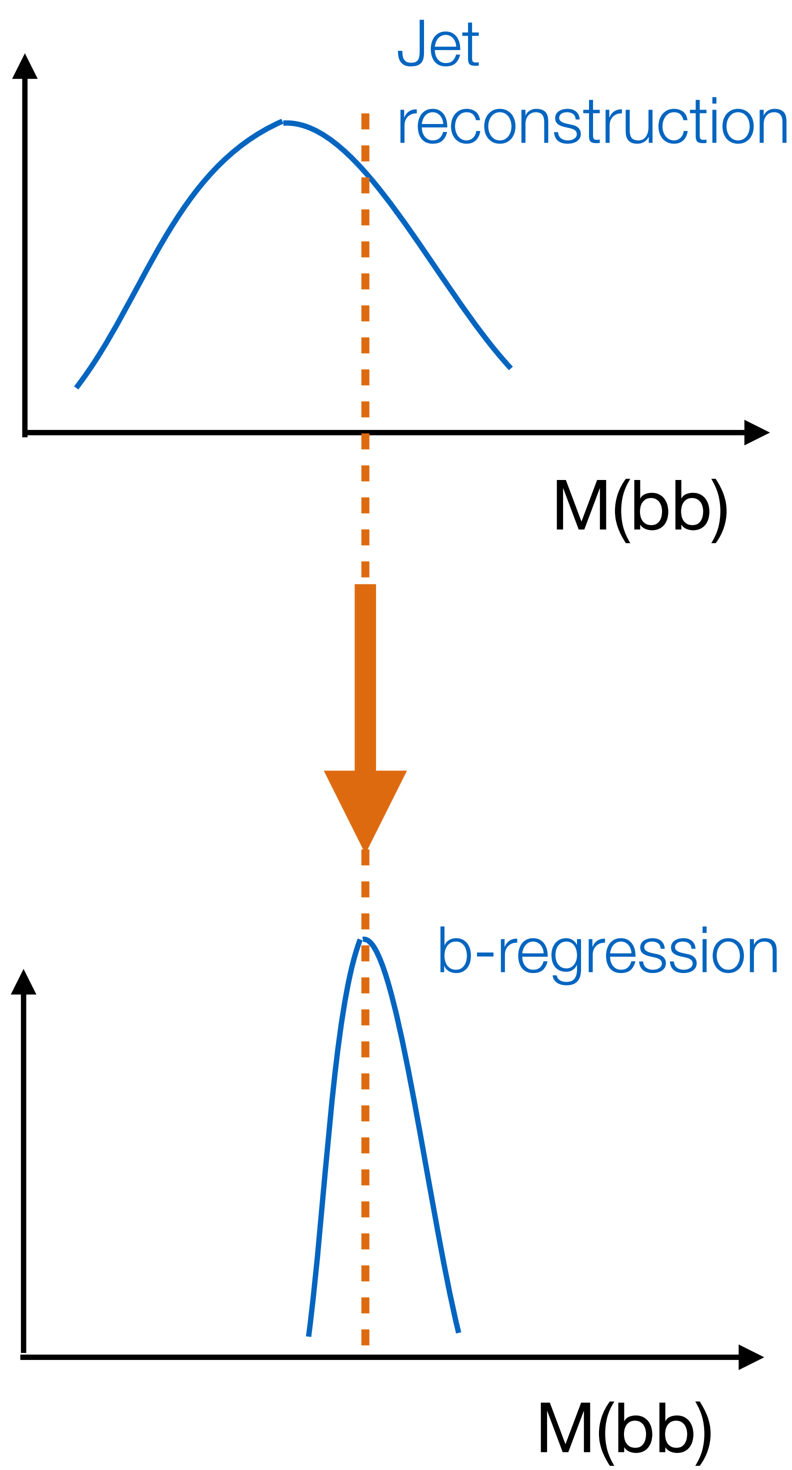
ATLAS ML workshop
CERN
11 - 15 Nov, 2019

Outline

- Introduction
- Historical overview
- CMS b-jet regression
- Performance in simulation
- Validation on data

Introduction

- b jets are important for many LHC analyses
- Many different analyses can benefit from a **momentum scale correction and improved resolution for b jets**
  - Higgs → bb
  - BSM analyses with b jets in the final state
  - Di-Higgs H(bb)H(xx)
    - most sensitive channels where one H →bb

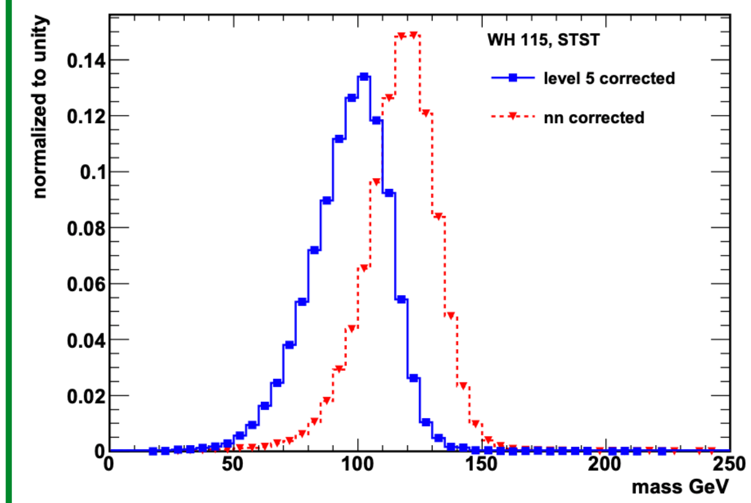


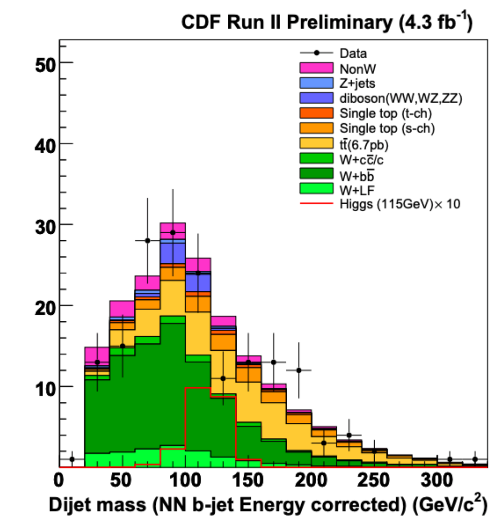
- **goals of b-jet energy regression :**
  - To improve detector response for all b jets (hadronic, semi-leptonic, leptonic)
  - To correct for (semi)leptonic b decays that lead to mismeasurement of p_T due to undetected neutrino

Historical overview of techniques

b-jet energy regression :

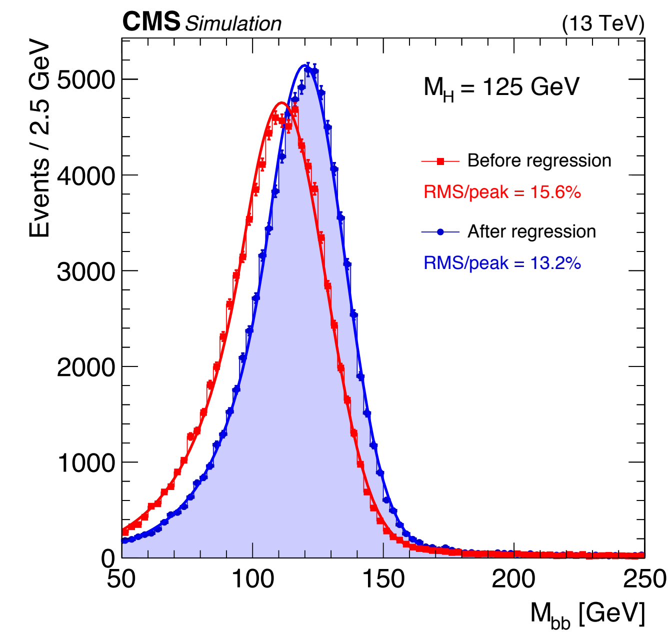
- Early applications at **Tevatron**
  - Tool to improve H → bb searches
  - Shallow neural network (NN) with 1 hidden layer and 9 neurons to estimate energy of b jets
  - Input variables include information about jet kinematics and composition





arxiv : 1107.3026

- LHC **CMS Run I** and 2016
  - BDT based regression
  - Similar input variables
  - Employed in VH→ bb and resonant Di-Higgs analyses

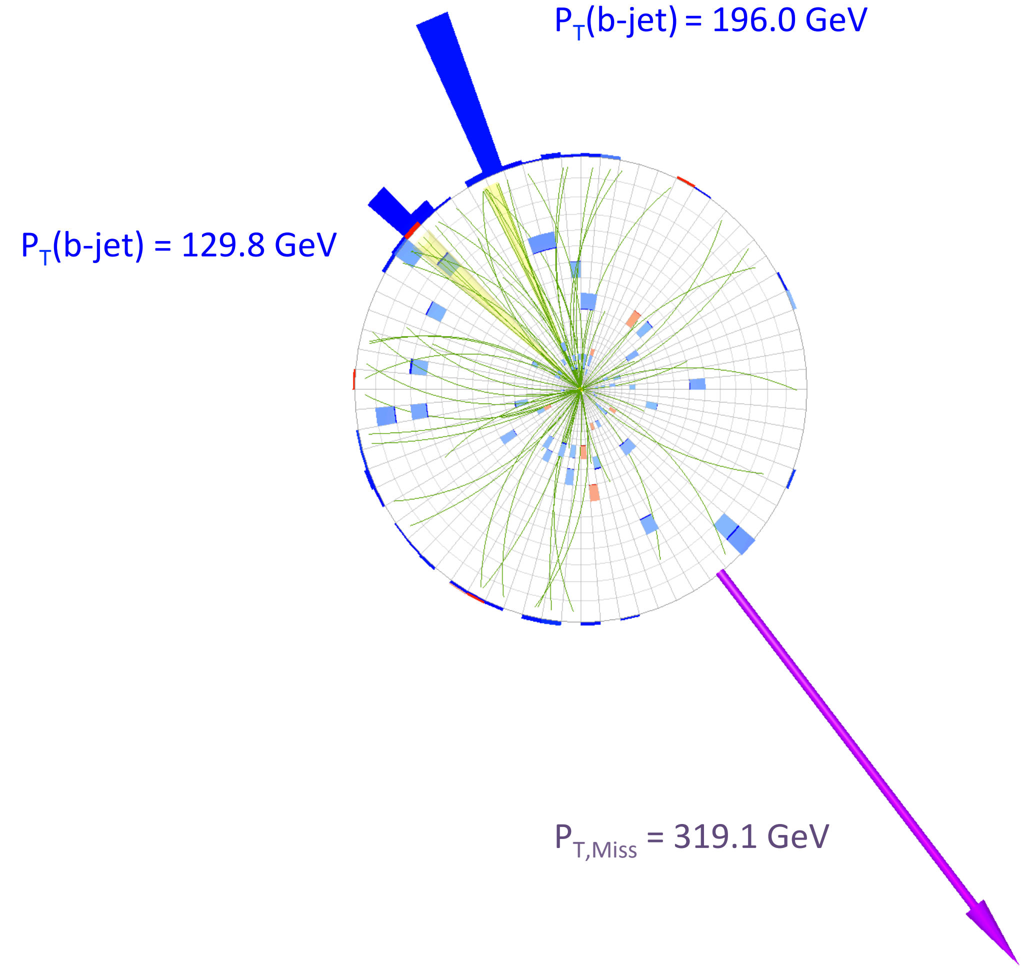


Phys. Lett. B 780 (2018) 501

New b-jet energy regression in CMS :

- Implemented in a Deep Neural Network
- Trained per jet (<u>not</u> per event)
- Developed to **improve resolution of b jets regardless of the final state** of a process
- Provides **jet energy resolution estimator** on jet-by-jet basis
- Improvement in dijet mass resolution brought by this regression **helped to reach observation of H → bb**

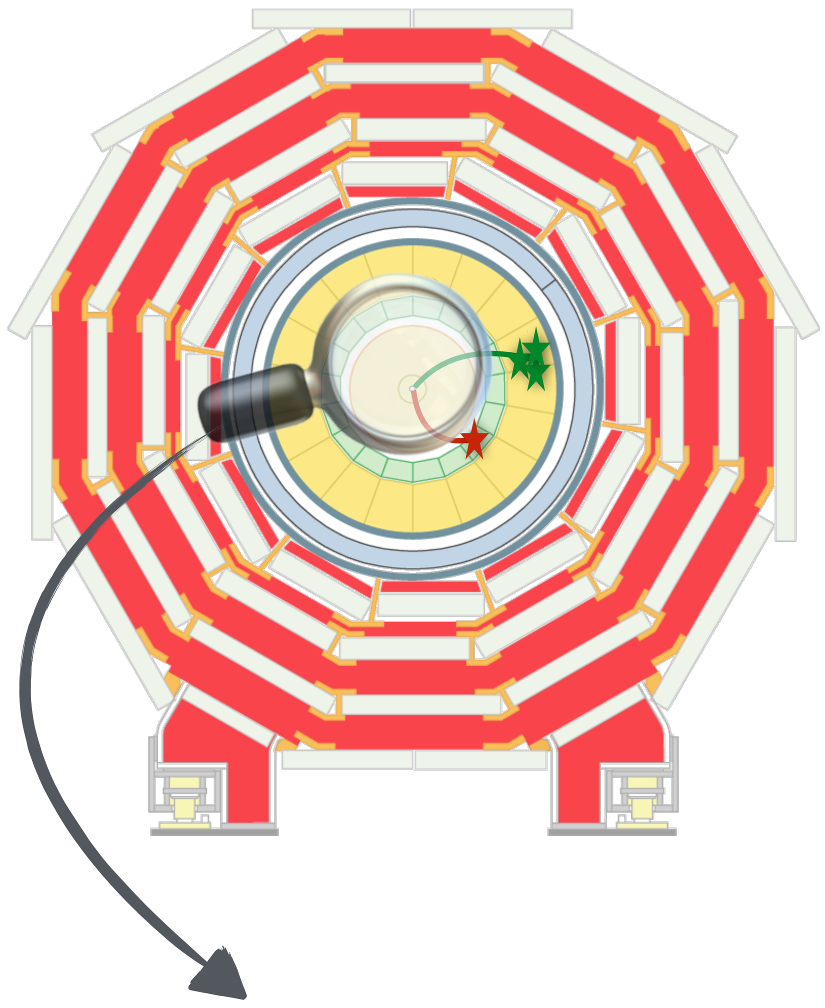
Z(vv)H(bb)

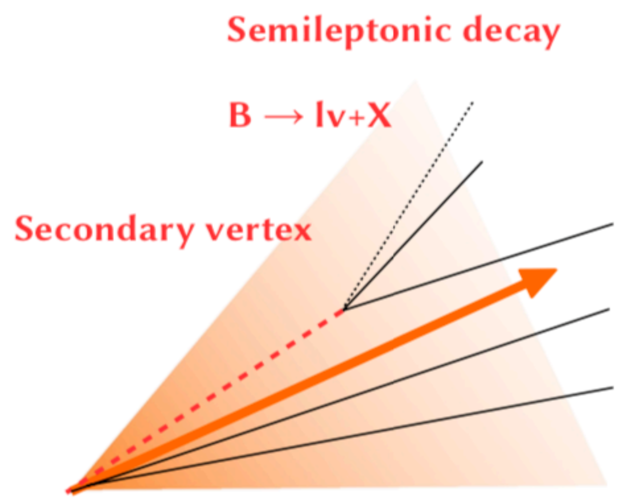


Phys. Rev. Lett. 121 (2018) 121801

New b-jet energy regression in CMS

- Reconstruct b-jet energy using a **multidimensional regression.**
  - Combine information about jet's :
    - **kinematics**
    - **constituents** : tracks, secondary vertices, and individual energy deposits reconstructed by the different subdetectors
  - use as **target** true b-jet energy at generator level from the simulated events
    - include missing energy from neutrinos to the gen jet 4-vector
  - As a regressor use a deep neural network(DNN)



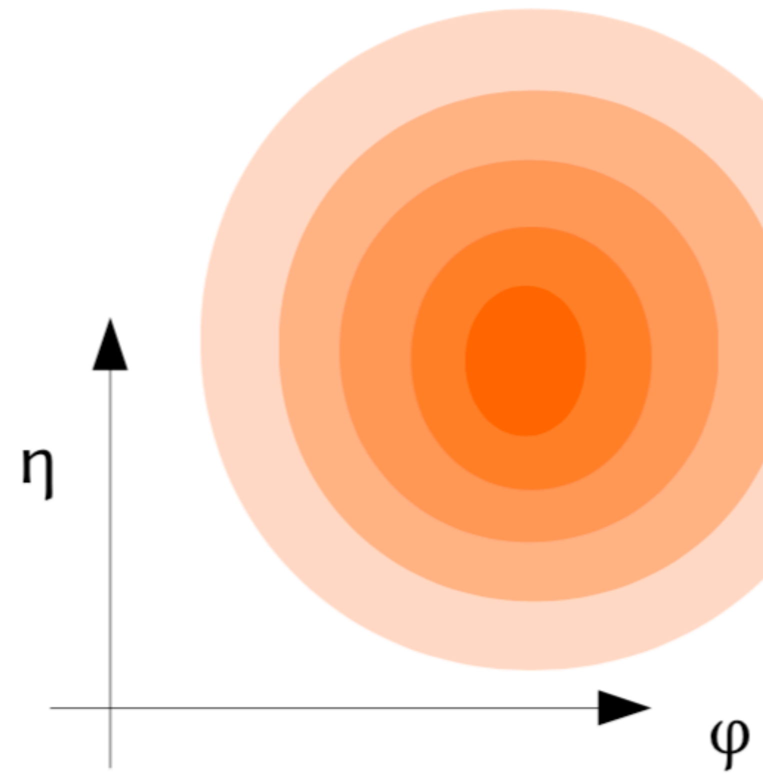


- **Train regression per jet**
  - Large sample of b jets needed : 100 M b jets from $t\bar{t}$ sample

Study of additional inputs

- **Jet shapes** (proxy to individual jet constituents which are difficult to model):
  - energy fractions in rings of dR
  - split the composition by origin : em, charged, neutral and muons
  - energy spread
- Multiplicity of jet constituents
- Lepton ID (e/μ)
- Jet p_T rel wrt to lepton, jet mass

Jet energy rings



Jet rings in $dR = \sqrt{d\phi^2 + d\eta^2}$:
(0 → 0.05 → 0.1 → 0.2 → 0.3 → 0.4)

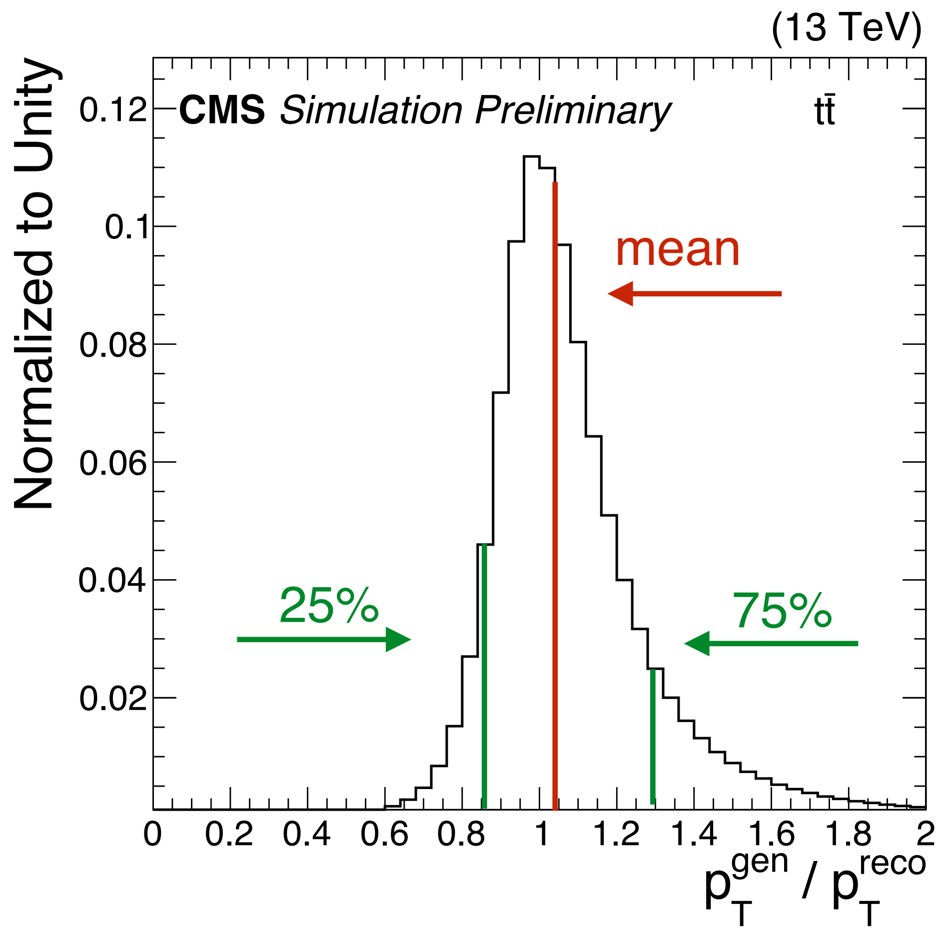
Good Data/MC agreement for all input variables

Resolution estimator

Any analysis can **increase sensitivity** by adding **information about jet resolution**

- **Goal** : provide jet resolution estimator on jet-by-jet basis

How to get a jet resolution estimator?



- Analytical shape is not trivial in this case
- Alternatively, we can be agnostic of the shape of the target and try to estimate quantiles positions
- As a **resolution estimator** use half difference of 25% and 75% quantiles (for a Gaussian distribution σ = 1.482 * IQR)

$$\mathbf{IQR = (\tau_{75\%} - \tau_{25\%}) / 2}$$

- Easy to implement in a simple loss function

Regression Loss function

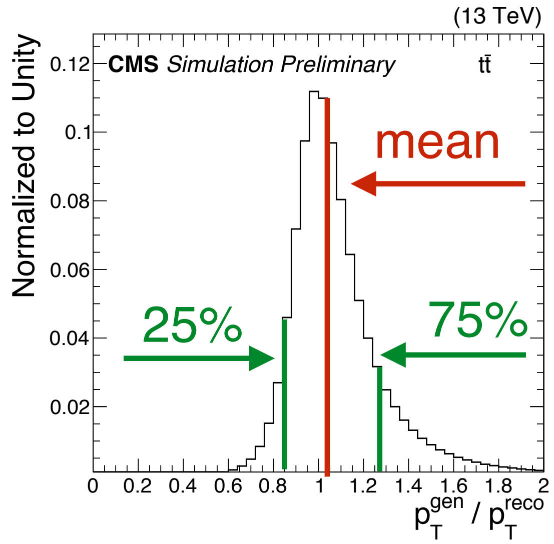
Loss function for DNN regression

- Regression task : **energy correction** to improve resolution and provide a **jet resolution estimator per-jet**
- Regression target $y = \frac{p_T^{gen+\nu}}{p_T^{reco}}$
- To get energy correction we use the Huber loss :

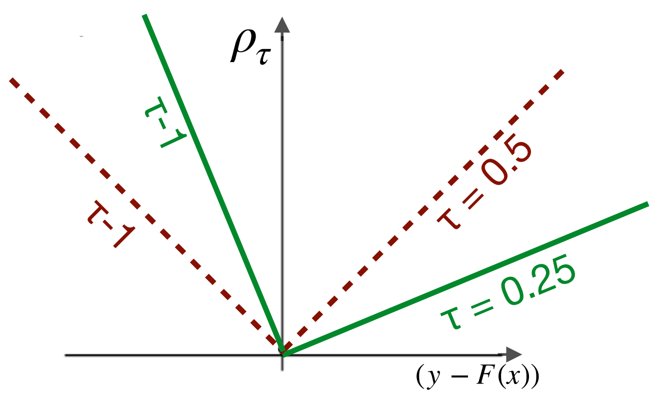
$$Huber(y, F(x)) = \begin{cases} \sum_i \frac{1}{2}(y_i - F(x_i))^2, \text{ for } |y_i - F(x_i)| < 1 \\ \sum_i |y_i - F(x_i)| - \frac{1}{2}, \text{ otherwise.} \end{cases}$$

- As resolution estimator use two quantile loss functions for 25% and 75% quantiles, τ - quantile :

$$\rho_\tau(y, F(x)) = \begin{cases} \sum_i \tau \cdot (y_i - F(x_i)), \text{ for } (y_i - F(x_i)) > 0 \\ \sum_i (\tau - 1) \cdot (y_i - F(x_i)), \text{ otherwise.} \end{cases}$$



Quantile loss $\rho_\tau(y, F(x))$

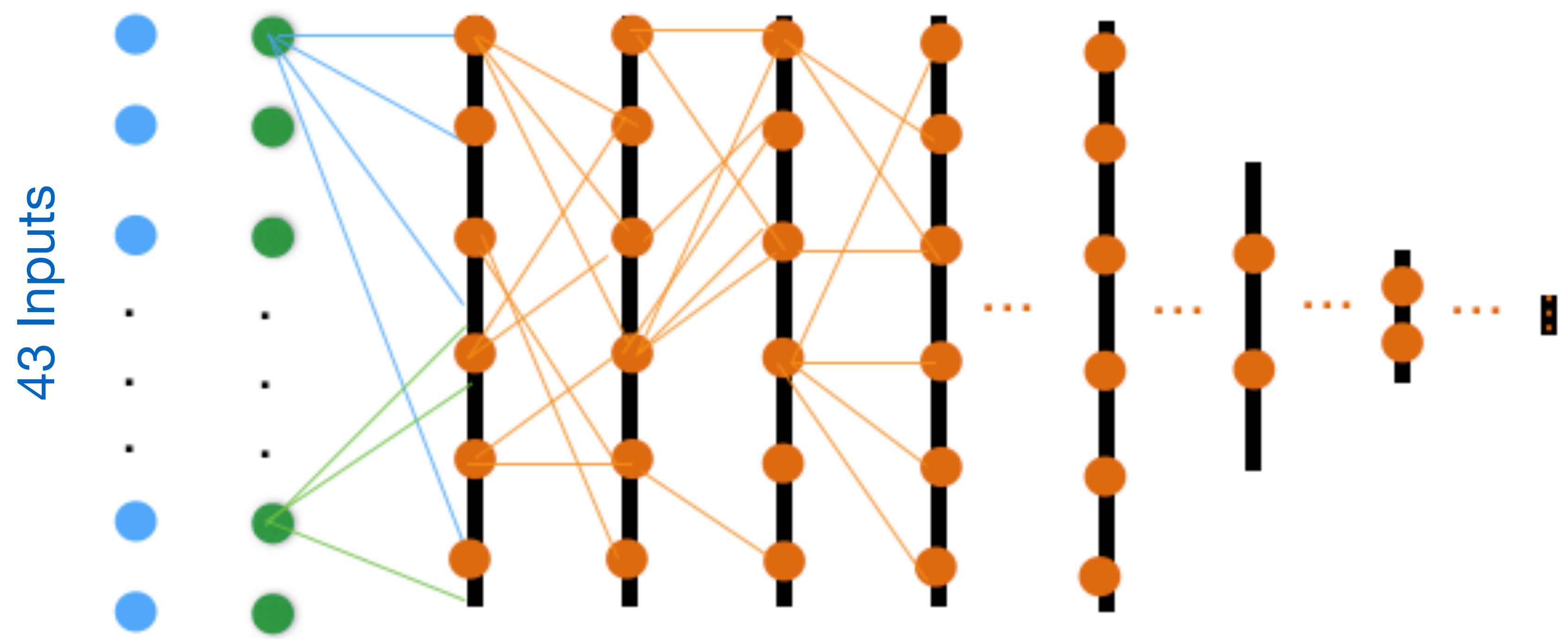


median - 50% quantile : τ = 0.5
25% quantile : τ = 0.25

Joint loss function for correction (Huber) and resolution (quantiles) :

$$\boxed{Loss = Huber(y, F(x)) + \rho_{0.75}(y - F(x)) + \rho_{0.25}(y - F(x))}$$

DNN architecture : Feed-forward fully connected NN



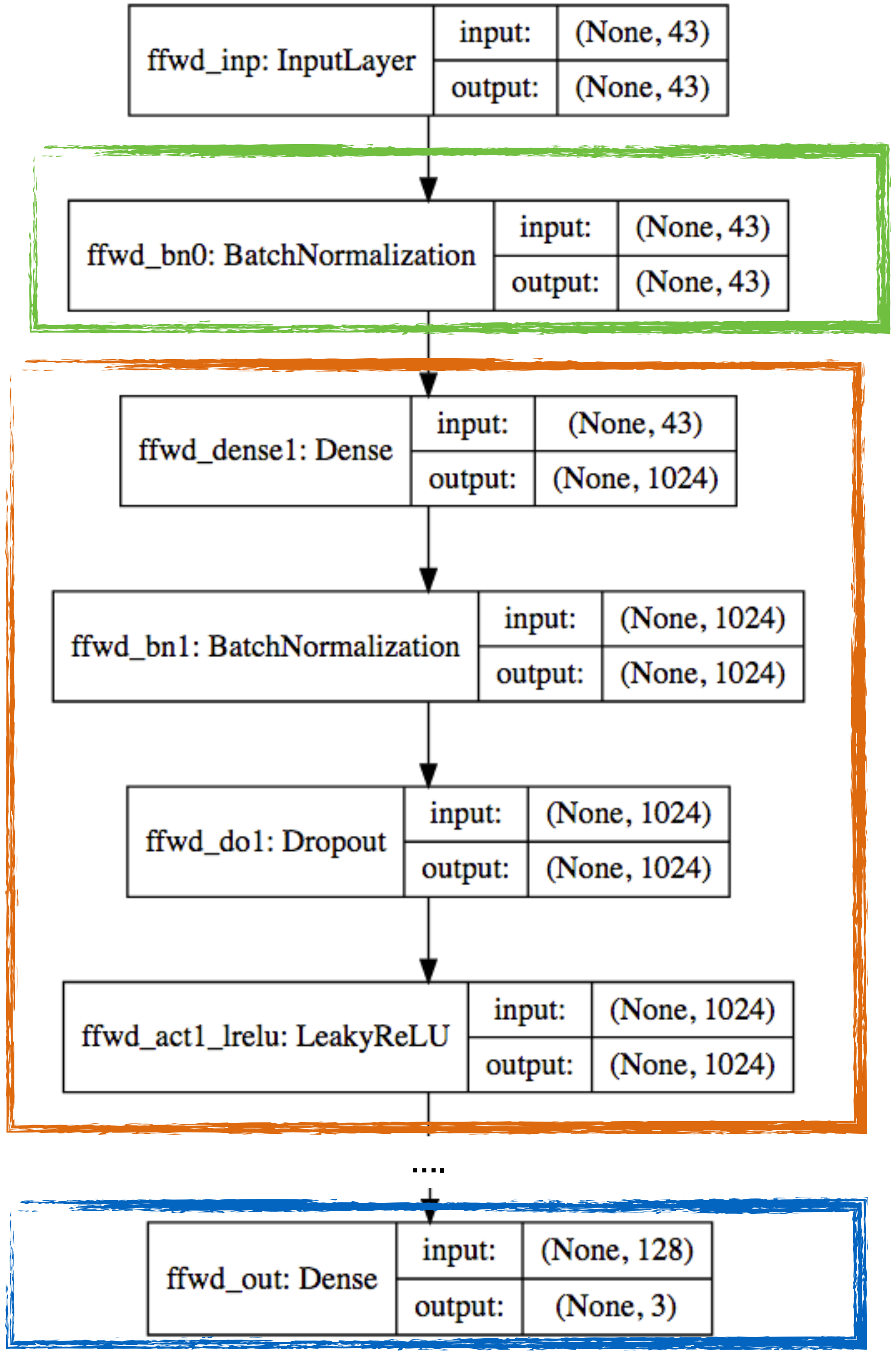
3 Outputs

- $\hat{y}$ - correction
- $\hat{y}_{25\%}$ - 25% quantile
- $\hat{y}_{75\%}$ - 75% quantile

resolution estimator $\hat{\sigma} = \dfrac{\hat{y}_{75\%} - \hat{y}_{25\%}}{2}$

- DNN is implemented in Keras with TensorFlow backend
- Back-propagation using stochastic gradient descent with Adam optimizer
- Hyperparameters and architectures were optimized using randomized grid search
- 6 layers with # neurons : [1024, 1024, 1024, 512, 256, 128]
- The network was trained on a single NVIDIA GeForce GTX 1080 Ti

DNN architecture : Feed-forward fully connected NN



- Input layer
- Batch normalization → internal data standardization

- Each hidden layer has 4 operations :
  - Linear transformation
  - Batch normalization
  - Dropout
  - Non-linear activation function
    - Leaky ReLU activation with α = 0.2

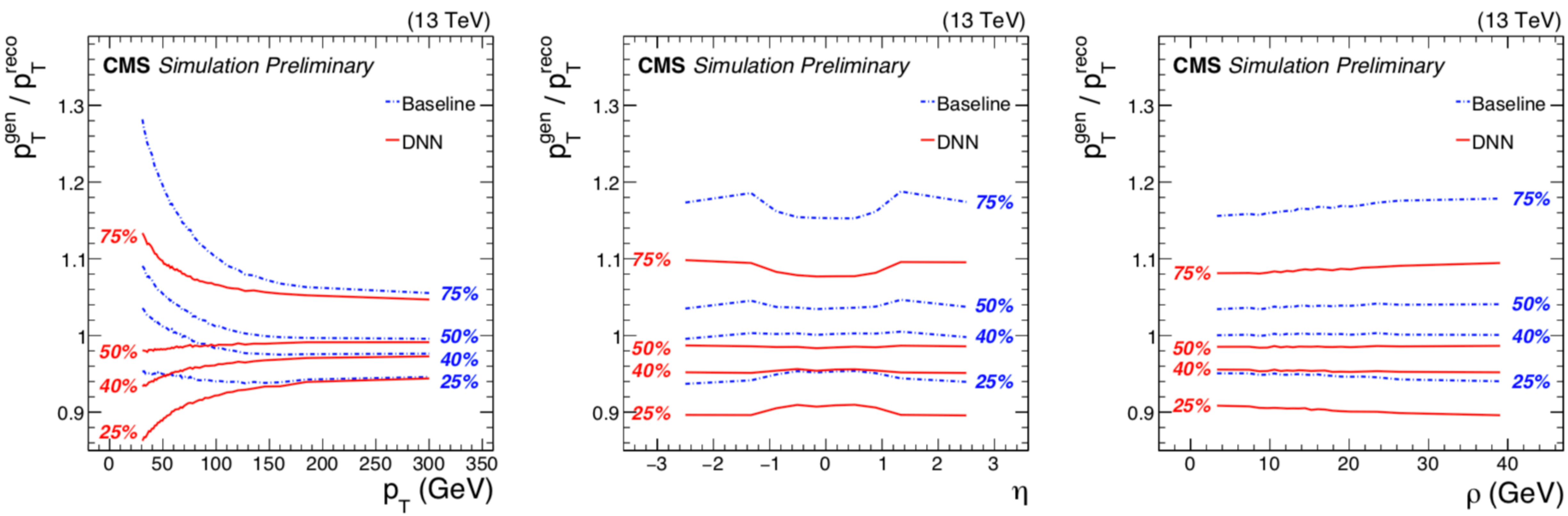
....

- Output : target is standardized (to zero-mean unit-variance)

Results

- Evaluate b-jet energy scale p_T^{gen}/p_T^{reco} after the application of the regression correction as a function of jet p_T , η and average event energy density ρ (quantiles 25%, 40%, 50%, 75%)

- Compare to baseline before-regression p_T^{gen}/p_T^{reco}
  - narrower distributions
  - flatter response



Results

The regression energy correction :

- Helps recovering the neutrino missing energy
- Improves the resolution for all jets

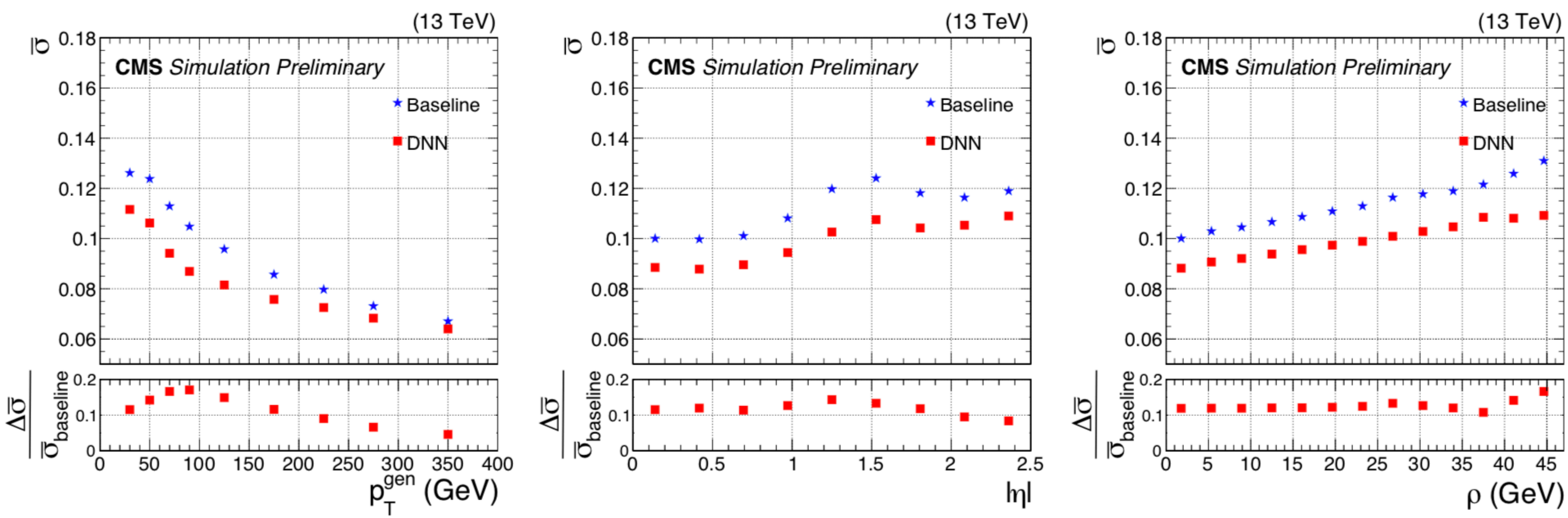
Quantify relative resolution improvement:

- Relative resolution estimated as $\bar{\sigma} = \dfrac{\sigma}{q_{40\%}} = \dfrac{q_{75\%} - q_{25\%}}{2q_{40\%}}$
- After regression **per-jet** relative resolution is improved by **~13%**
- Very similar performance achieved for b jets arising from different physics processes

MC sample	Improvement
$t\bar{t}$	12.2%
$Z(\to \ell^+\ell^-)H(\to b\bar{b})$	12.8%
$H(\to b\bar{b})H(\to \gamma\gamma)$ SM	13.1%
$H(\to b\bar{b})H(\to \gamma\gamma)$ resonant 500 GeV	14.5%
$H(\to b\bar{b})H(\to \gamma\gamma)$ resonant 700 GeV	13.1%

Results

- Improved relative resolution as a function of jet p_T , η and ρ and for simulated $t\bar{t}$
- For all physics processes considered, the per jet relative resolution improvement is around **12-18%** for p_T < 100 GeV and down to around 5-9% for p_T> 200 GeV

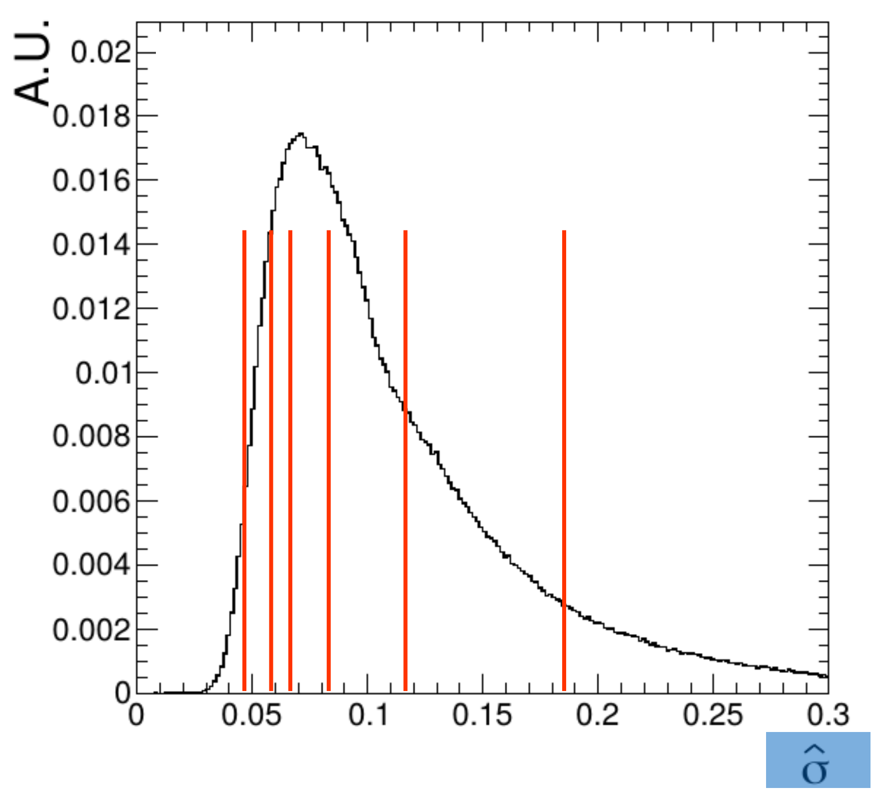


Resolution estimator

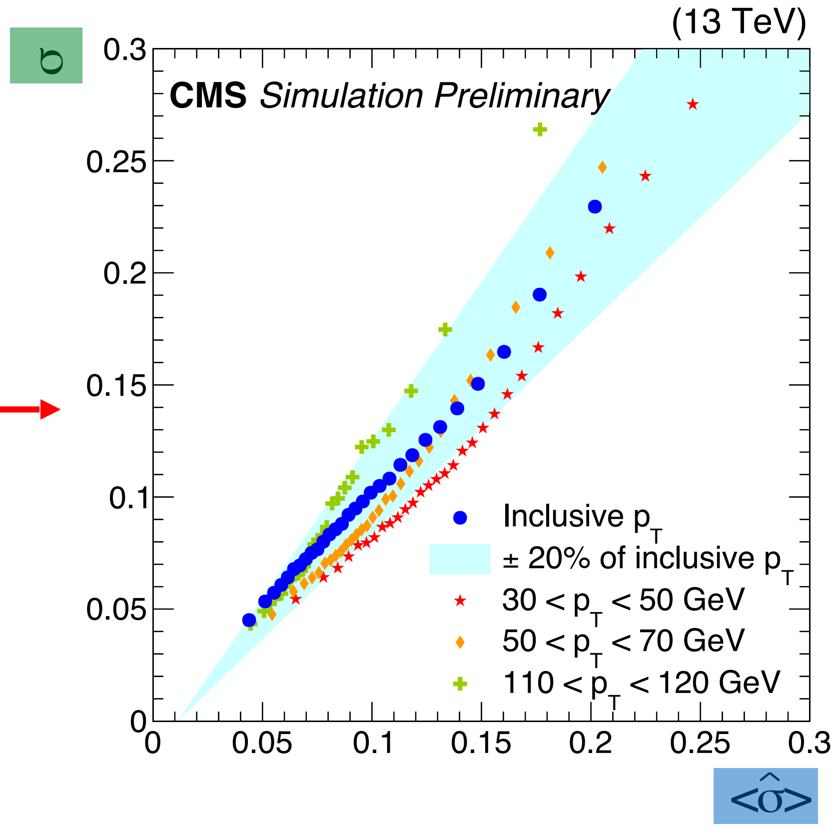
- For each jet a **resolution estimator** is provided as an output of DNN
  - How does it map to the actual resolution of the b jets σ ? $\sigma = \dfrac{q_{75\%} - q_{25\%}}{2}$

Cross-check:

- Split the sample of jets into several equidistant quantiles of jet resolution estimator $\hat{\sigma}$
- In each bin quantify the resolution σ using gen-level information
- Check if the two correspond to each other
- Repeat the same test in bins of jet p_T. Deviations from linear behavior do not exceed 20%

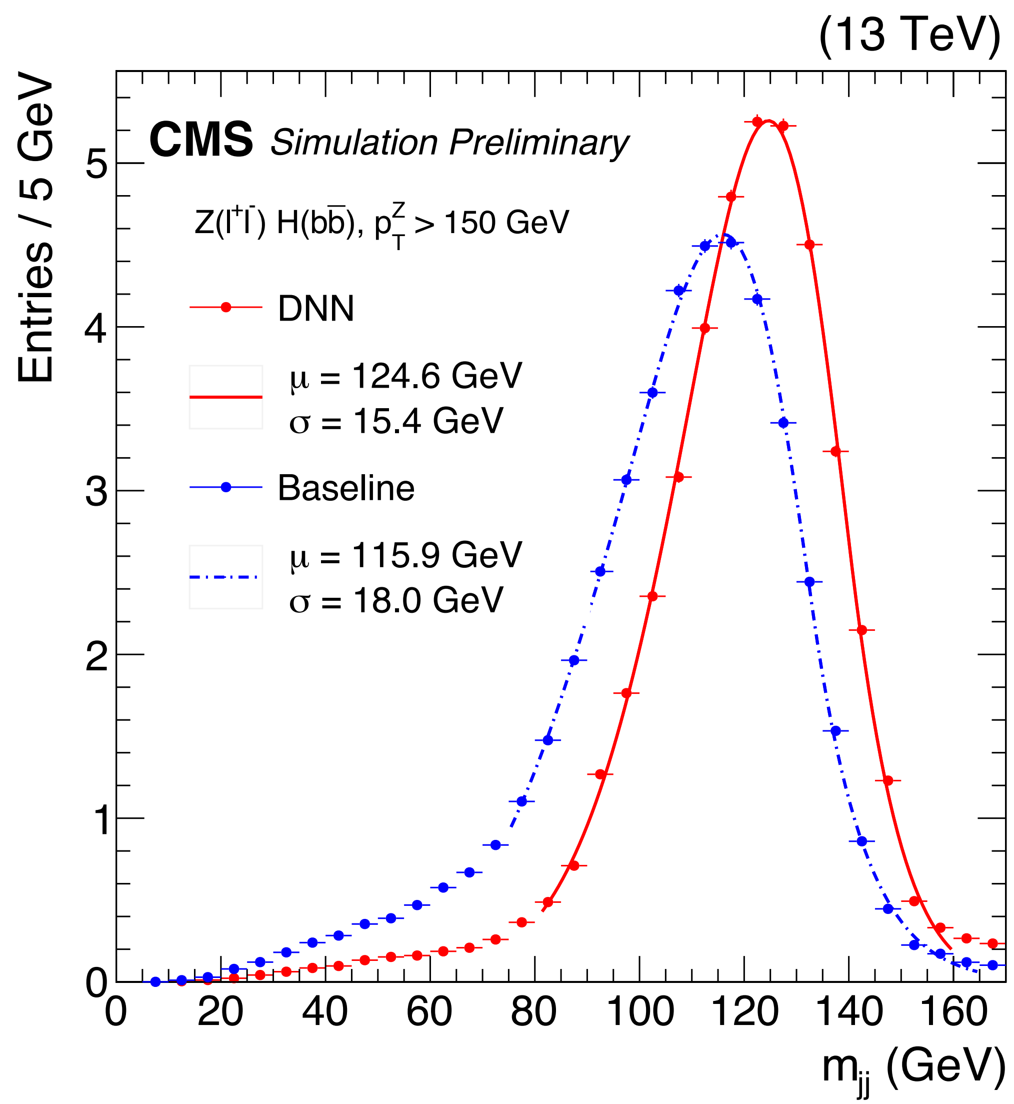


Linear dependence: Resolution estimator OK



Dijet resolution improvement

- Improvements so far are quoted at single-jet level, however many analyses use invariant mass of b jets as a discriminating variable
- **Resolution improvement for dijet inv. mass is larger than for a single jet**
- Improvements to dijet mass resolution come from **2 factors** :
  - improvement in jet resolution
  - effective equalization of the energy scale in all regions of phase space

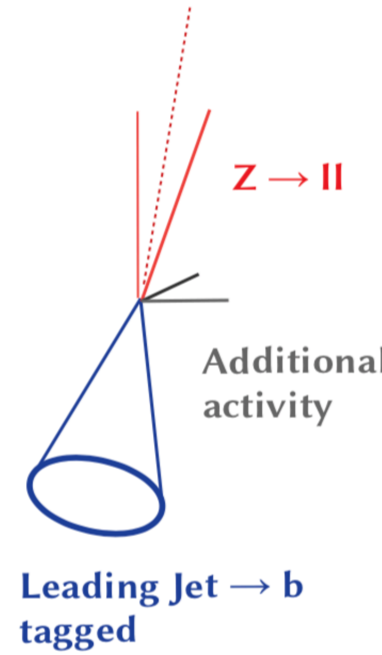


Z(→ll)H(→bb) :
b jets p_T > 20 GeV
leptons p_T > 20 GeV
$p_T(Z)$ > 150 GeV

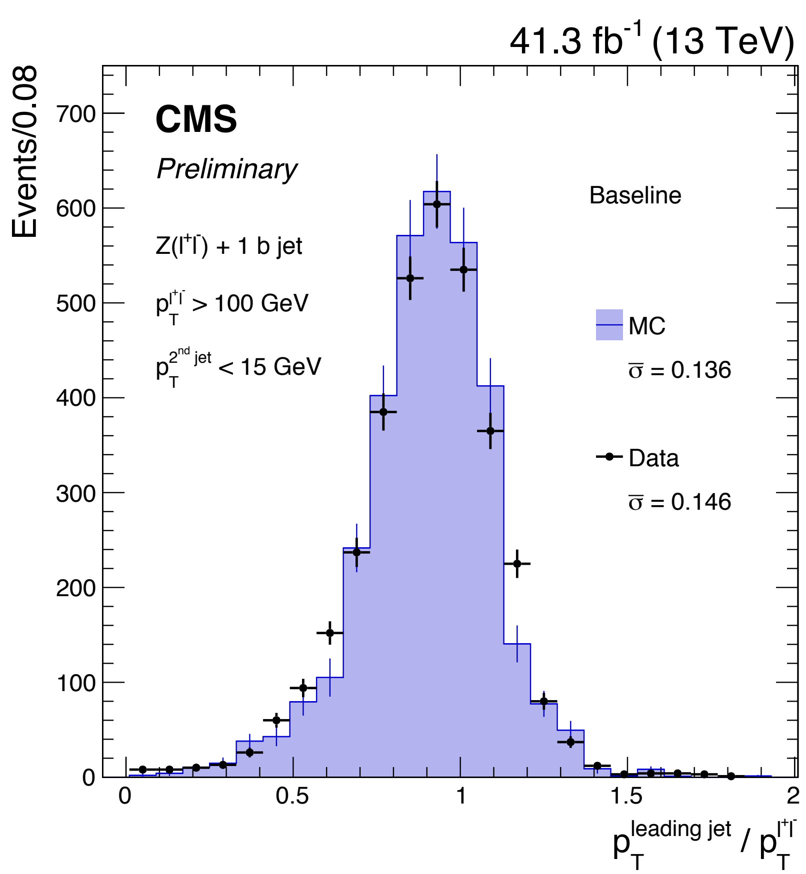
20% improvement in dijet mass resolution

Validation on data

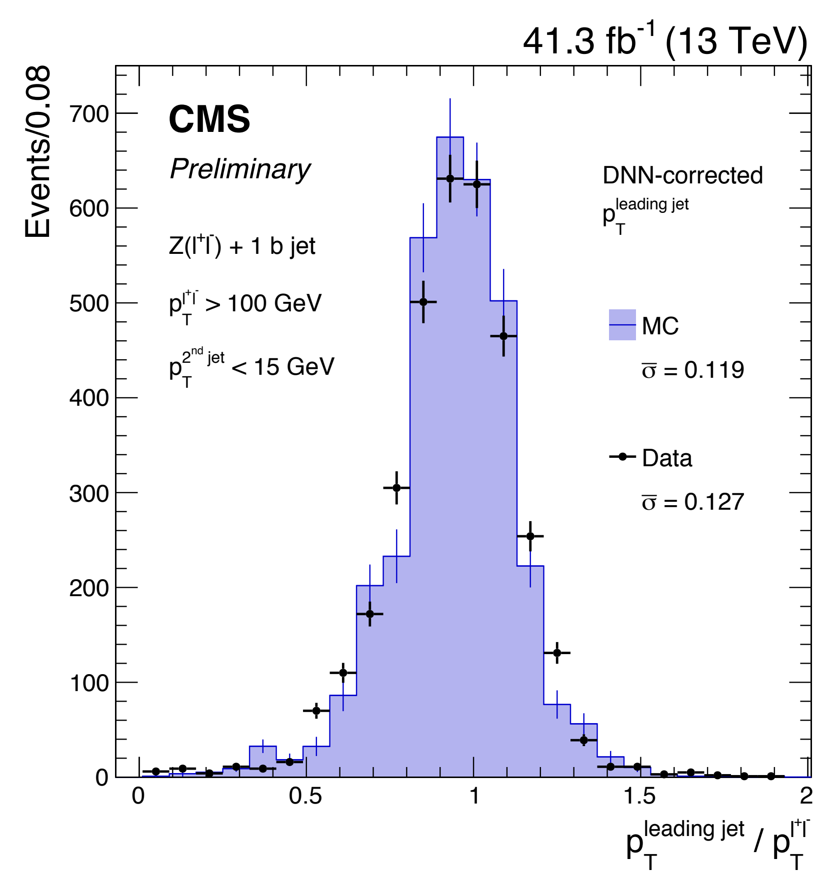
- Performance in data evaluated with p_T balance in $Z \rightarrow \mu\mu/ee$ + b jet topology
- **Resolution improvement** is consistent for MC and data, and is **13 %**



Resolution improvement achieved in MC is successfully transferred to the data domain!



Regression →



Summary

- DNN based b-jet energy regression was developed for the CMS analyses with b jets in final states
- b-jet regression was trained using jet composition information
- Both energy correction and jet resolution estimator are provided
- The technique was validated on data, and the regression was successfully applied to reach the observation of $H \rightarrow bb$
- Resolution improvements are ~13% per-jet inclusively, and phase space dependent for the dijet mass (20-25% for $H \rightarrow bb$)
- **CMS-PAS-HIG-18-027**
- Paper is in the final steps of CMS approval

Additional Material