Measurement of the top quark pole mass using $t\bar{t}$+1 jet events with the CMS experiment

DPG Frühjahrstagung Dortmund 2021

Matteo Defranchis[1], Katerina Lipka[2], Sebastian Wuchterl[2]

[1]CERN

[2]DESY

16 March 2021

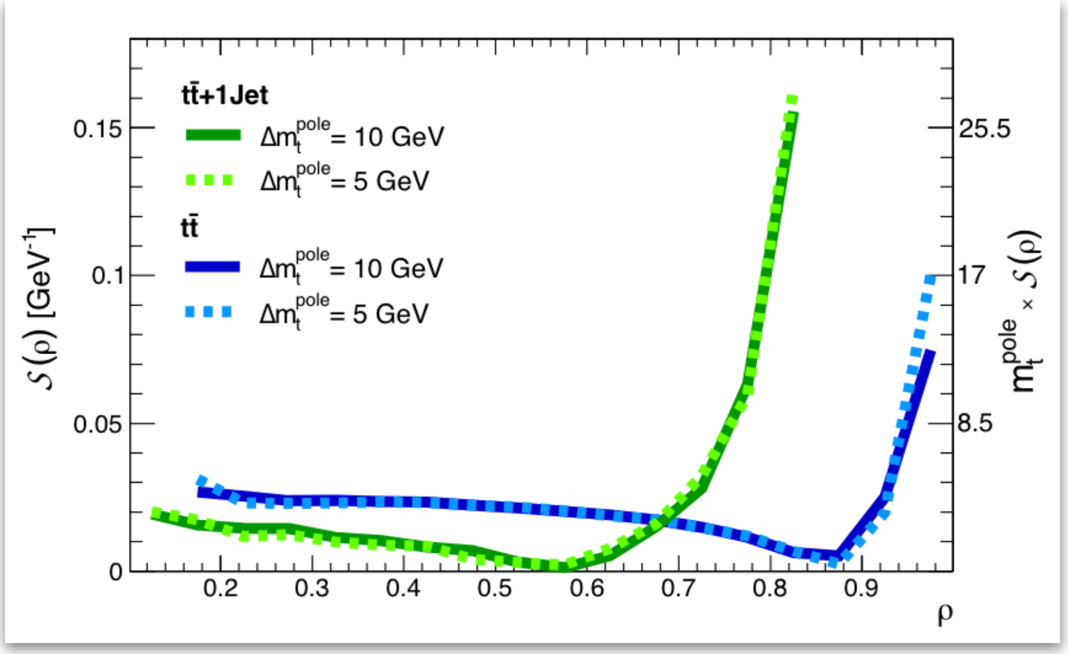
Top quark mass from $t\bar{t}$+1jet

Motivation

- use $t\bar{t}$+1jet events for precision measurement of m_t

$$\mathscr{R}(m_t, \rho) = \frac{1}{\sigma_{t\bar{t}+1jet}} \frac{d\sigma_{t\bar{t}+1jet}}{d\rho}$$

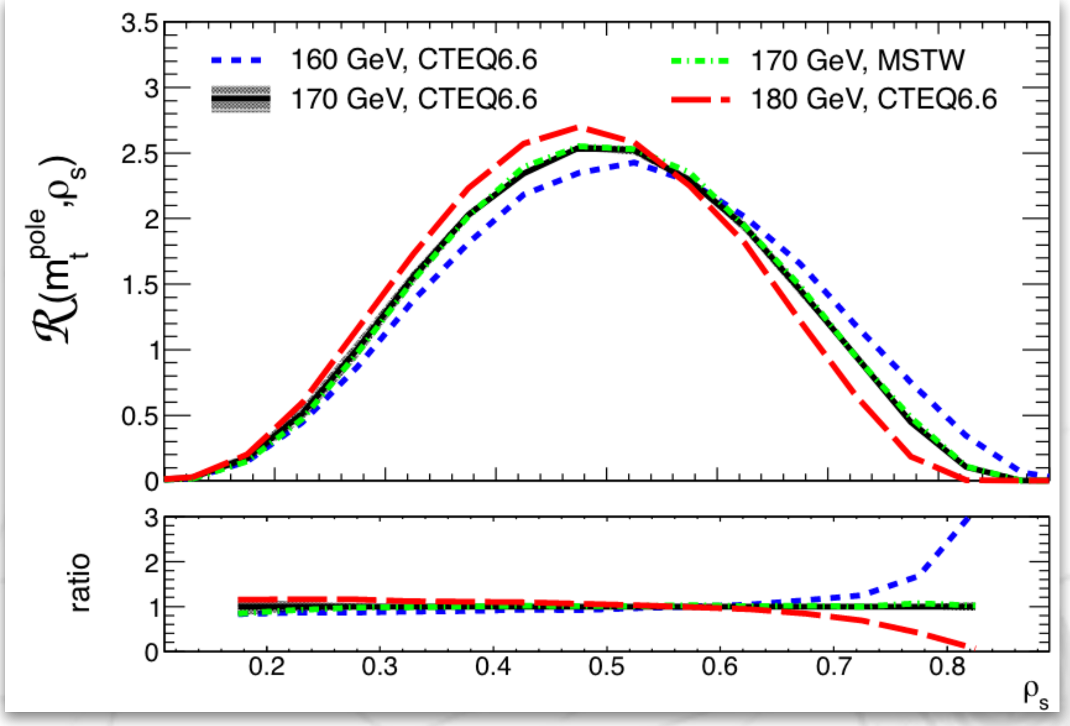
with $\rho = \frac{2m_0}{m_{t\bar{t}+1jet}}, m_0 = 170GeV$



theory paper: arXiv:1303.6415

7+8 TeV calculations

- higher sensitivity compared to inclusive $t\bar{t}$
- compare $\mathscr{R}$ to NLO+PS theory prediction
  - extract top quark mass in pole-mass scheme (+ MSbar & MSR scheme) ⇒ unfolding to parton level

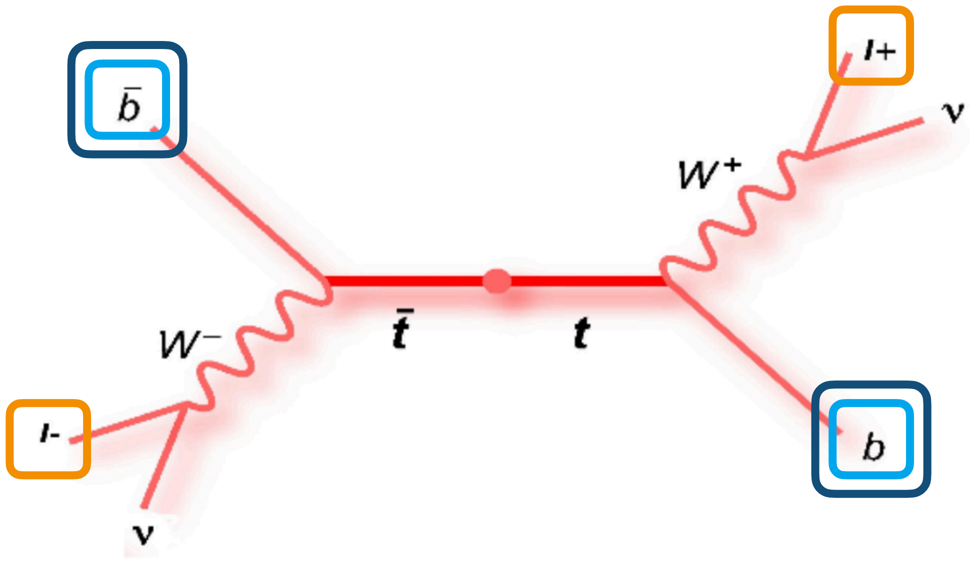


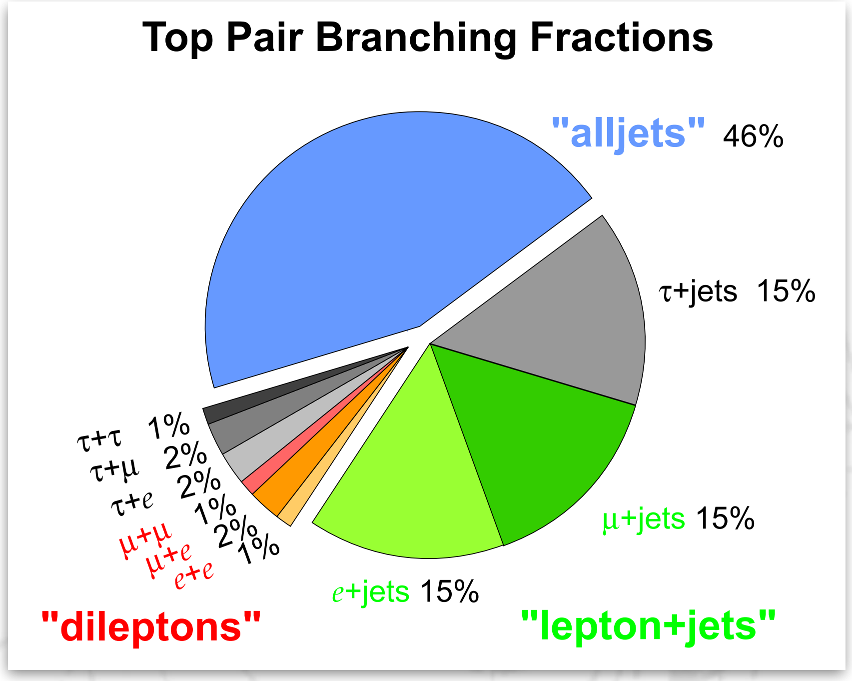
- full Run II analysis (2016-2018 data)
  - total $\mathscr{L}$ = 35.9 + 41.5 +59.7 fb^{-1}

Analysis

Event Selection

- 2 opposite-charged leptons: μμ, eμ, ee
  - tight lepton requirements
  - p_T > 25 (20) GeV & $|\eta|$ < 2.4
  - m_{ll} > 20 GeV
- jets
  - anti-kt R=0.4
  - b-Jet energy regression
  - p_T > 30 GeV & $|\eta|$ < 2.4
  - b-tagging criterion (loose)
  - third jet: p_T > 30 GeV
- ee, μμ channel:
  - MET > 40 GeV
  - Z-window veto: $|m_Z - m_{ll}|$ >15 GeV
- kinematic reconstruction of observable of interest

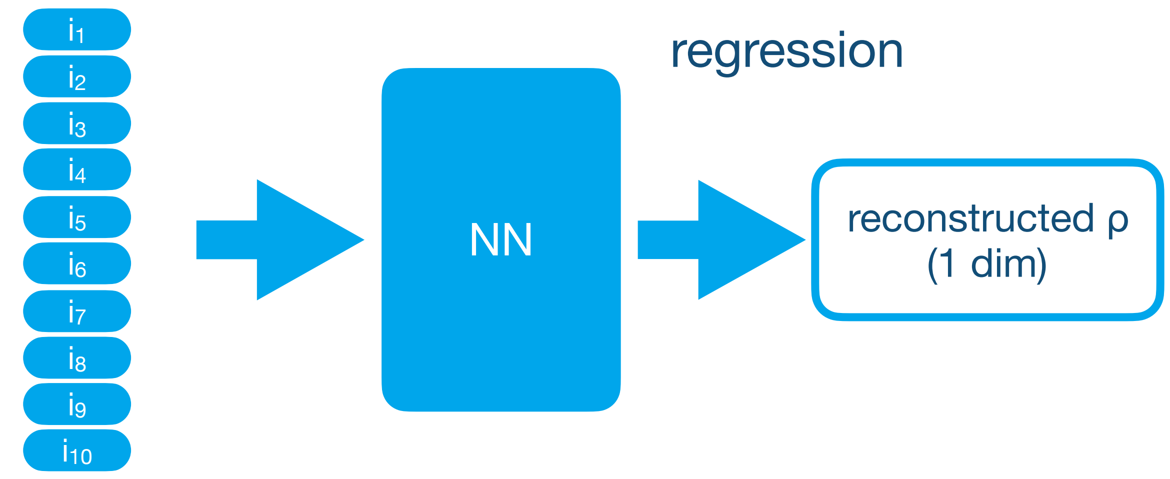


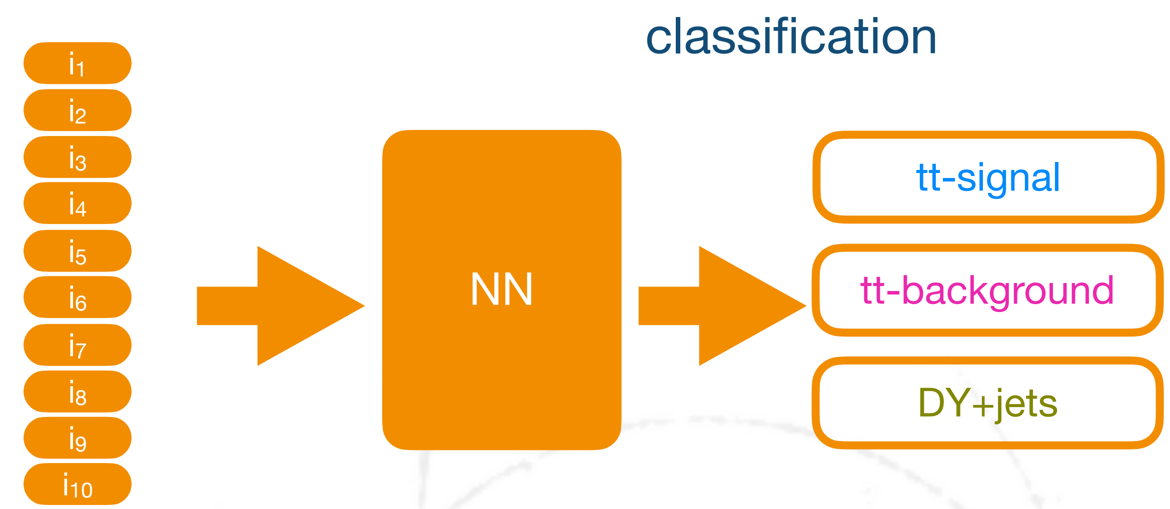


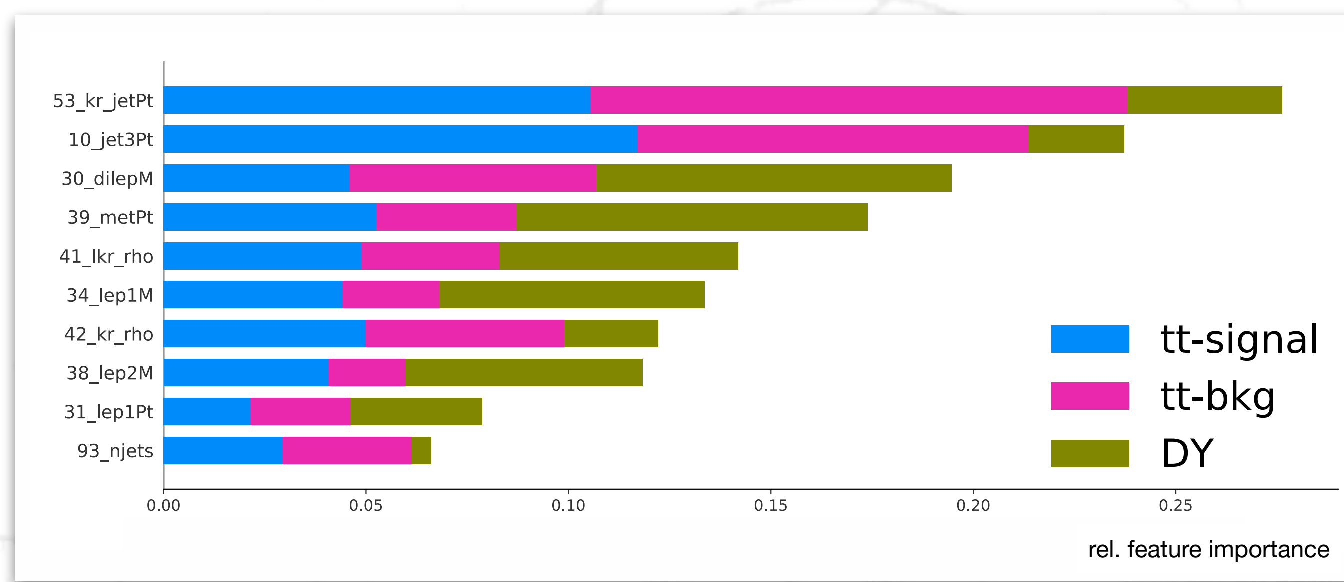
Neural network applications

General procedure & optimisation

- combined training for years & lepton channels
- training/testing/validation steps performed using independent Madgraph+Pythia $t\bar{t}$ simulation
    - performance cross evaluated on Powheg+Pythia simulation
- reduction of input features O(120) to 10 most relevant ones
- bayesian hyperparameter optimisation performed
- input features validated using 1D saturated goodness of fit test



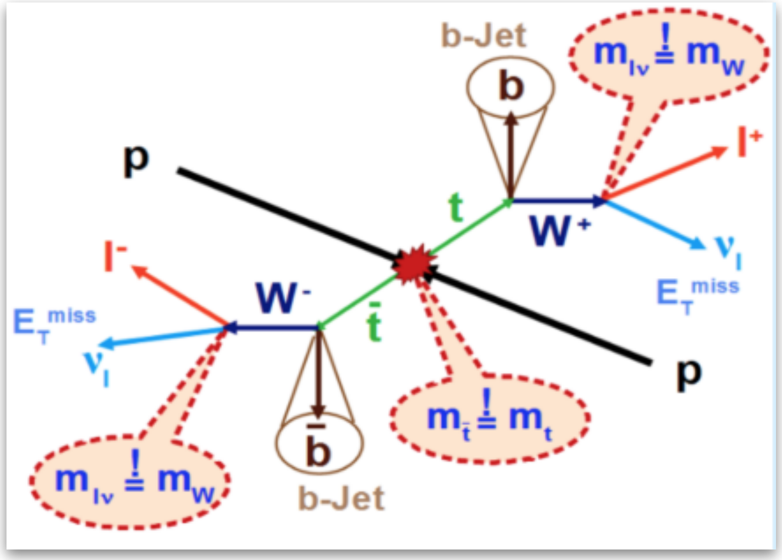


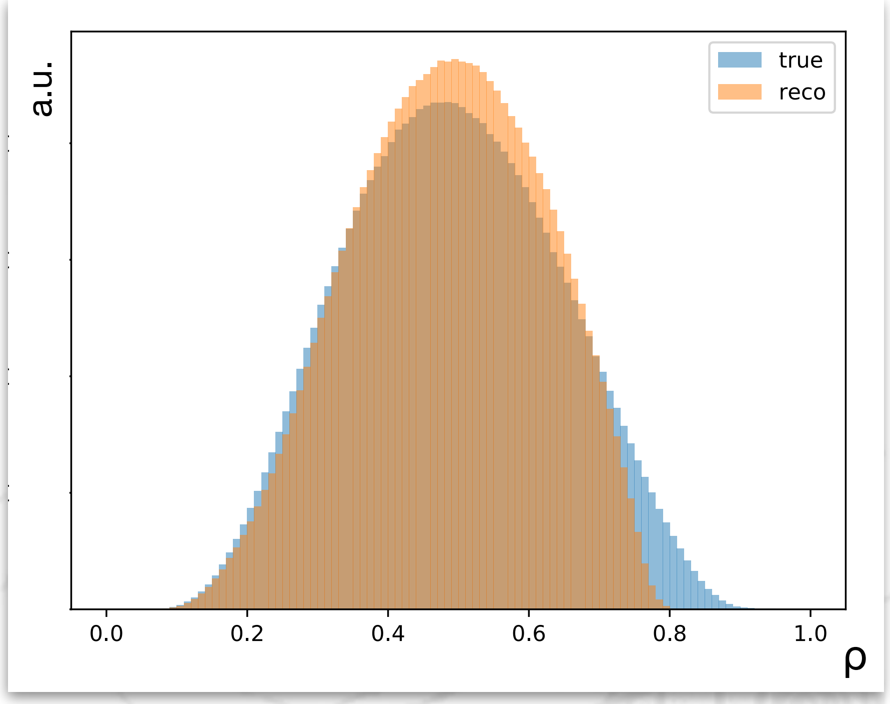


Reconstruction of main observable

Studying different approaches

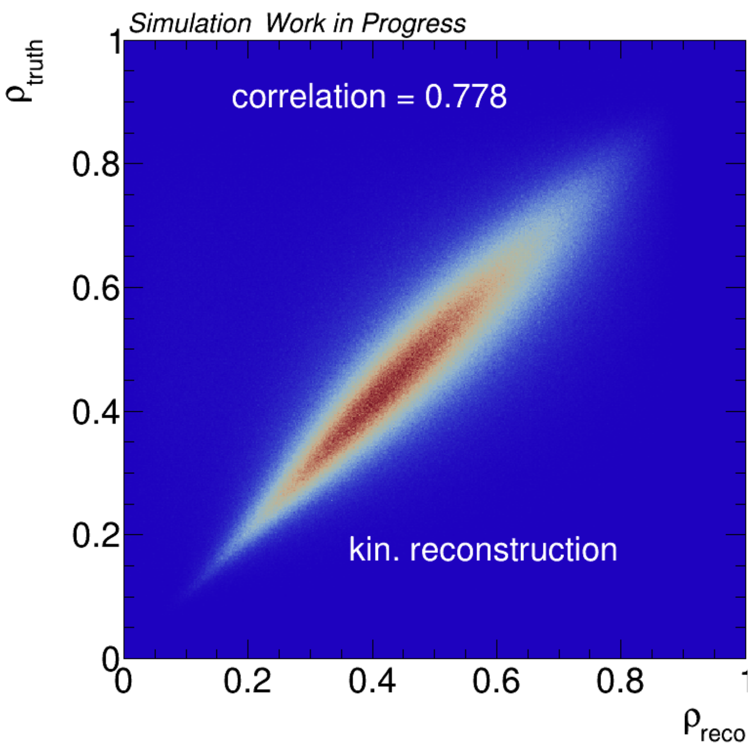
- unfolding of ρ distribution requires kinematic reconstruction of $t\bar{t}$ system
- two common methods in CMS:
  - "full" kin. reconstruction
    - top/antitop solved individually using constraints
  - "loose" kin. reconstruction
    - solve $t\bar{t}$ system without m_t constraint
- studied NN based reconstruction
  - 4-momenta + high level inputs
  - trained on $t\bar{t}$ simulation
    - events that pass basic selection and have at least three rec. jets
    - feed forward & fully connected architecture

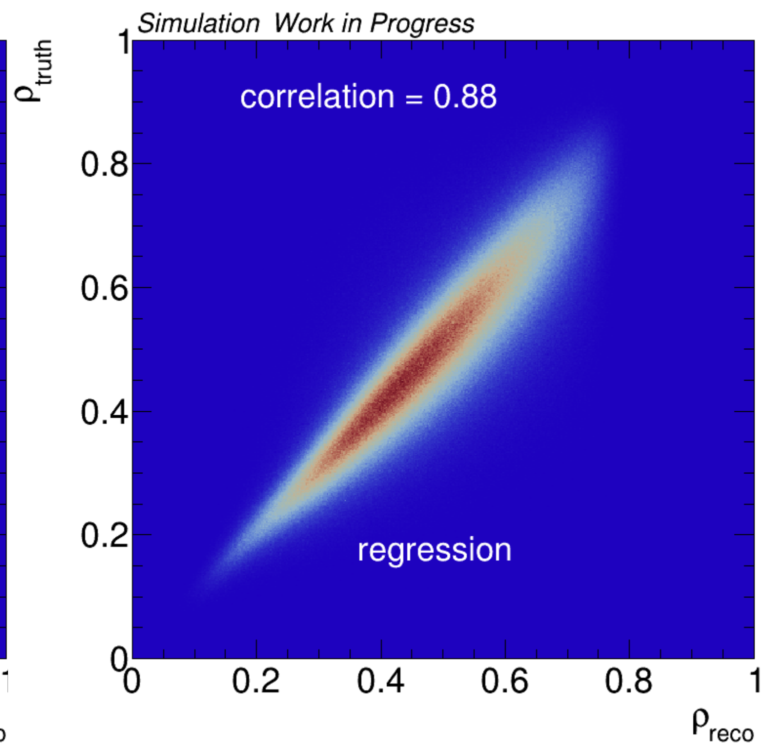


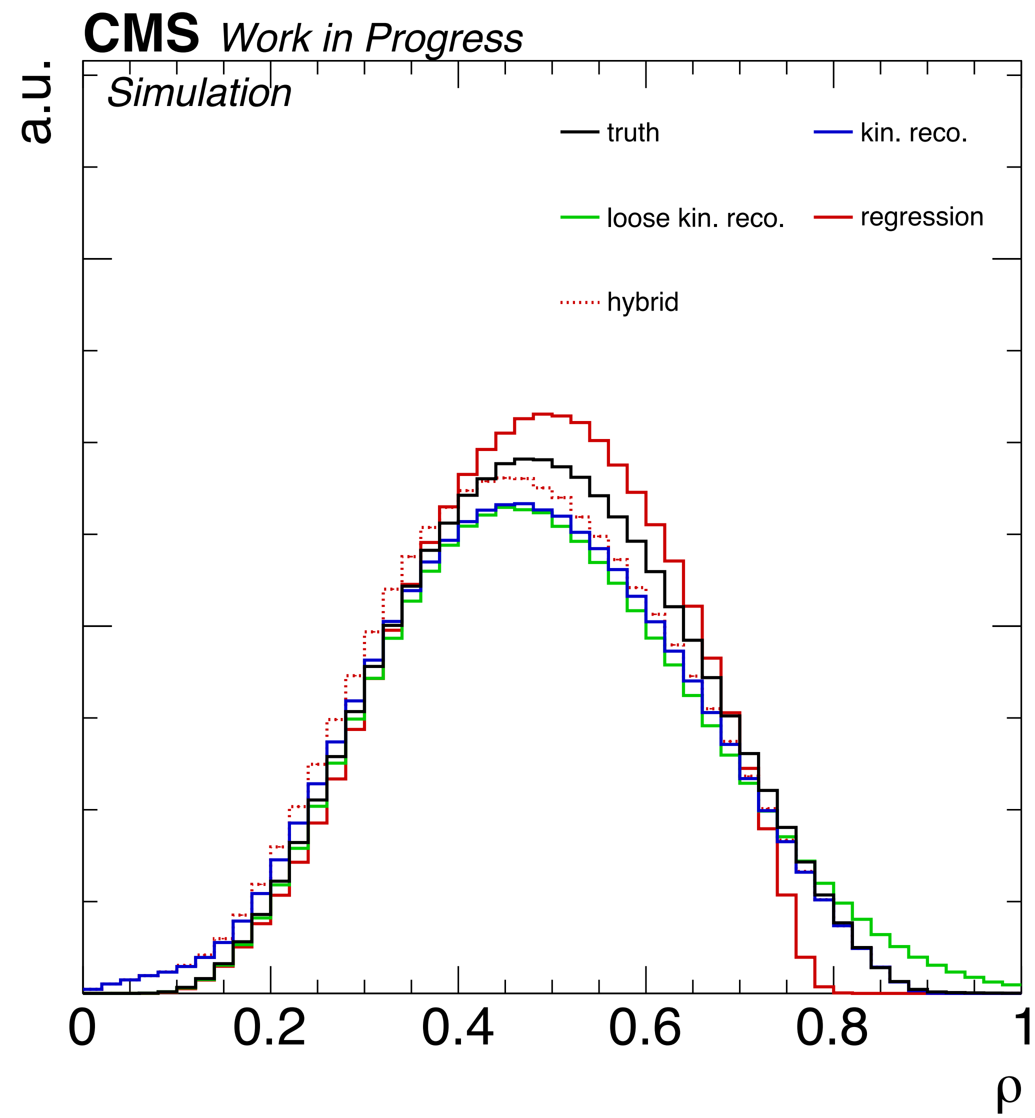


Reconstruction of main observable

Studying different approaches







NN-based reconstruction

pro's

- better resolution
- correlation between truth/reco. closer to 1
- fully efficient

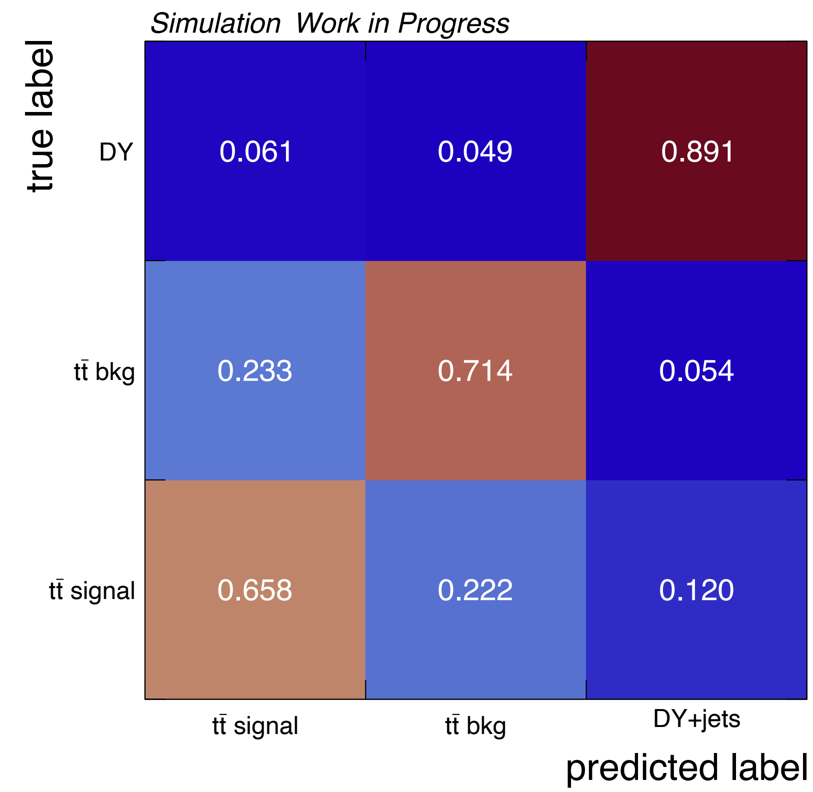
con's

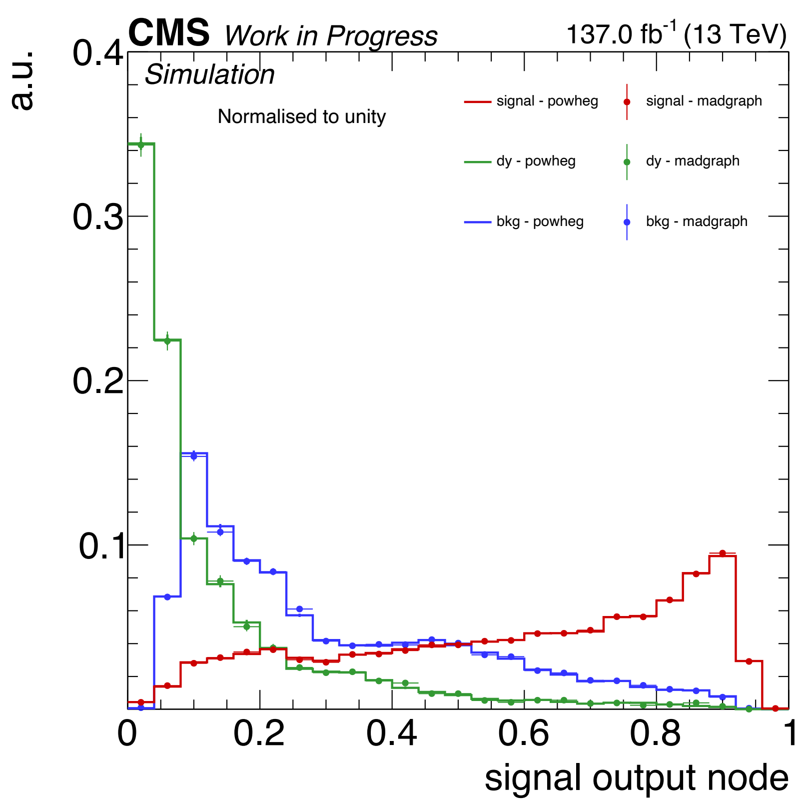
- bias towards bulk of distribution
- not enough statistics in tails

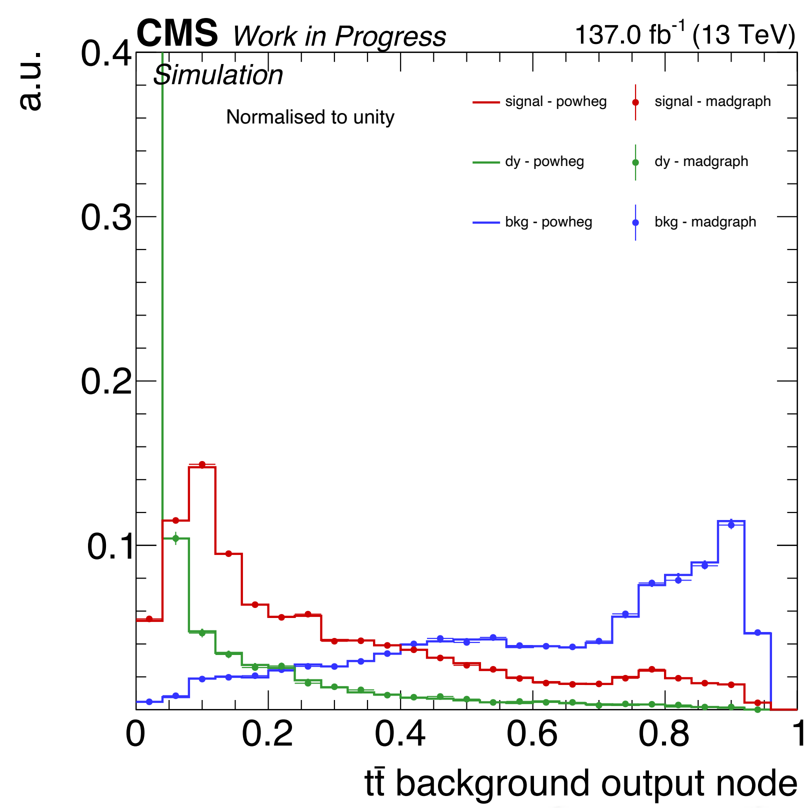
⇒ employ hybrid-approach:
kin. reco. + DNN (where kin. reco. fails)

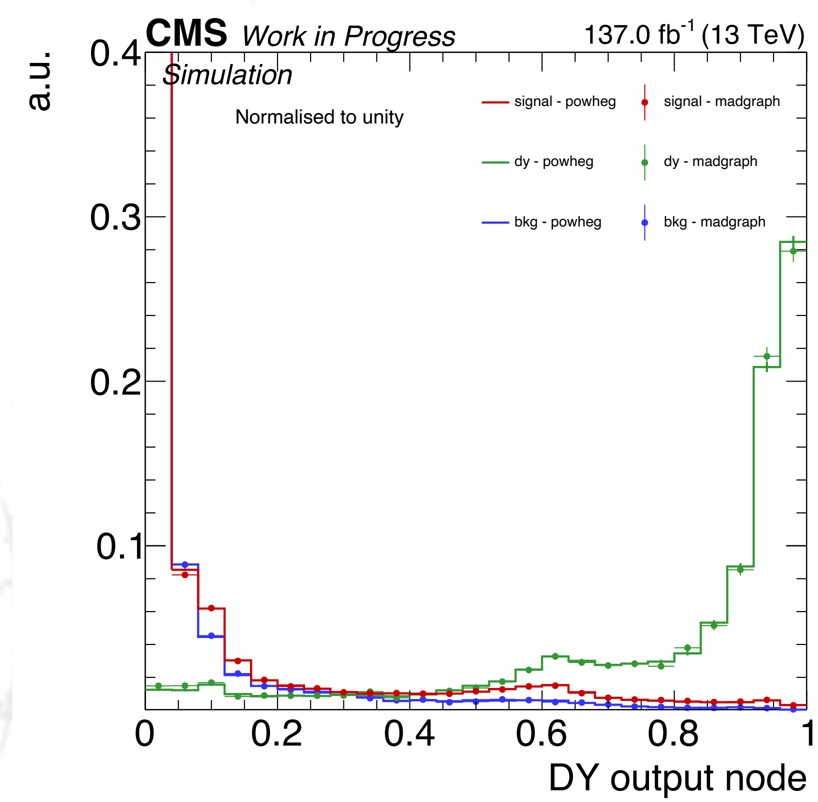
Event classification

- chosen reconstruction technique (hybrid) reconstructs ρ for all signal and background events
- discriminate between three classes:
  - signal, $t\bar{t}$ background, Drell-Yan
- trained on independent Madgraph $t\bar{t}$ and Drell-Yan samples
- threshold for signal region defined using:
  - overall background and signal statistics
  - Asimov significance $\sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right)-s\right)}$
  - Purity x Efficiency





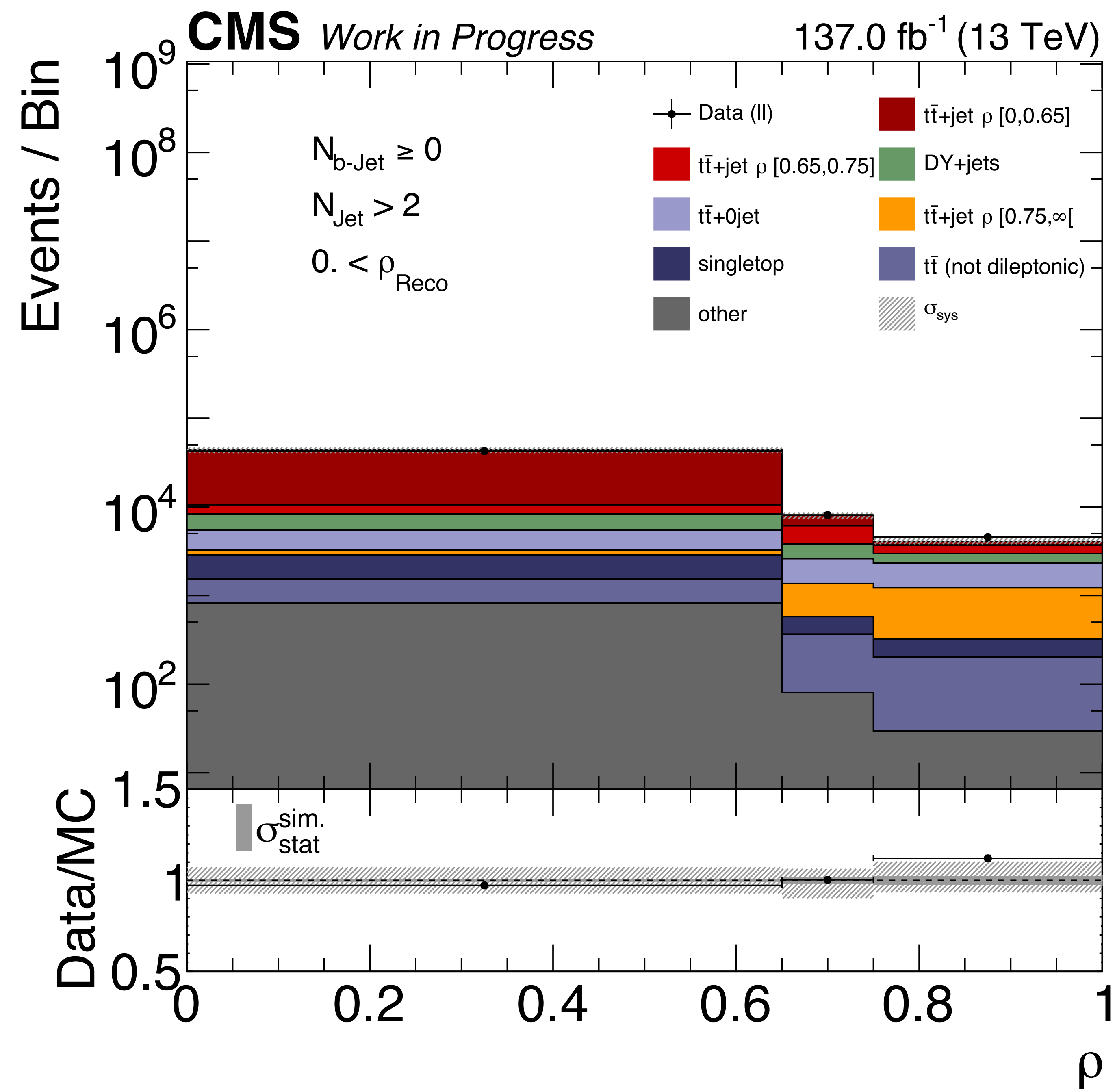




Reconstructed ρ

Data/MC

- using kin. reco+DNN reconstruction
- binning in ρ (gen+reco):
  3 bins:
  [0, 0.65, 0.75, ∞]

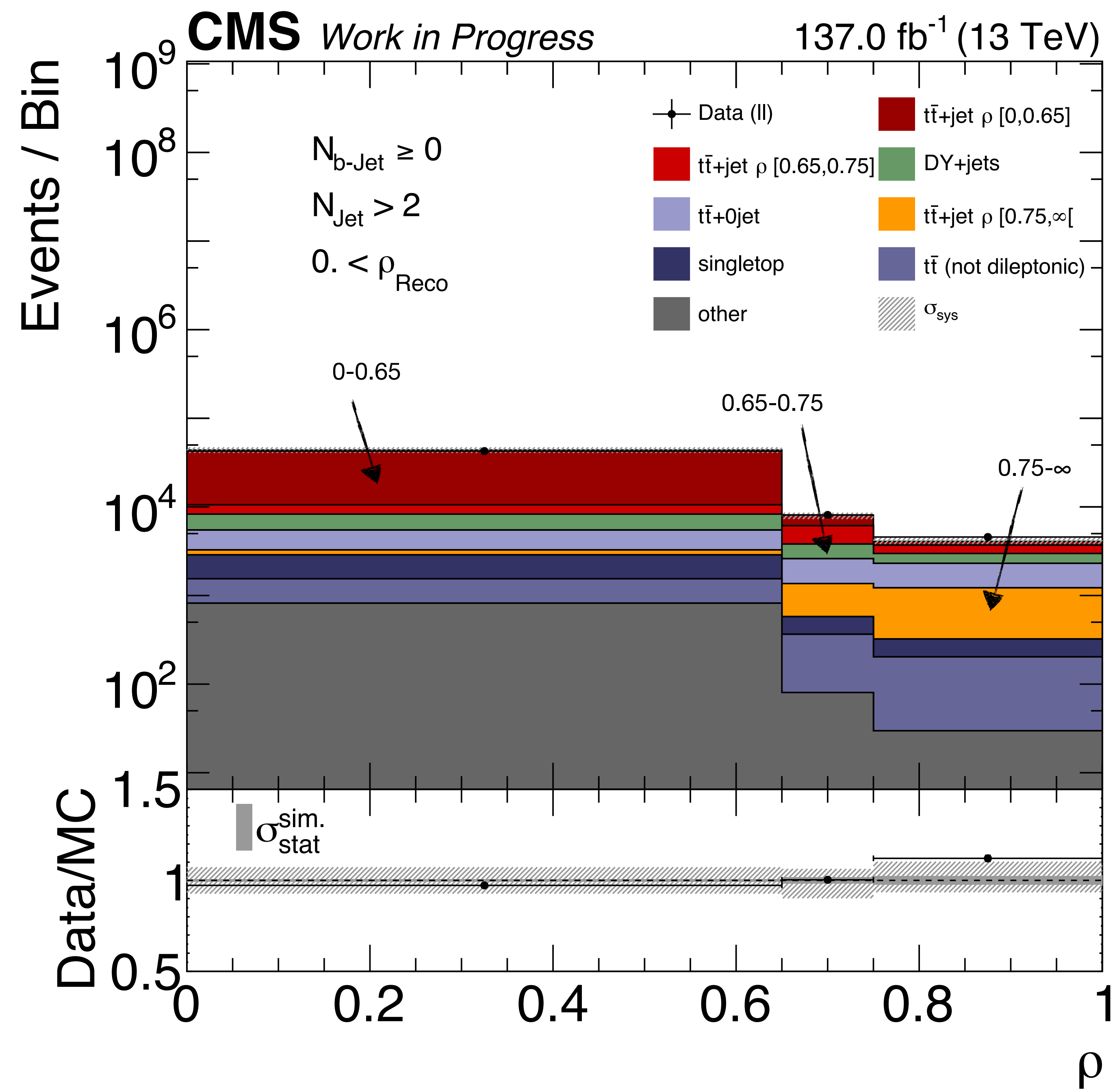


$$\rho = \frac{2m_0}{m_{t\bar{t}+1jet}}, m_0 = 170GeV$$

Reconstructed ρ

Data/MC

- using kin. reco+DNN reconstruction
- binning in ρ (gen+reco):
  3 bins:
  [0, 0.65, 0.75, ∞]



$$\rho = \frac{2m_0}{m_{t\bar{t}+1jet}}, m_0 = 170GeV$$

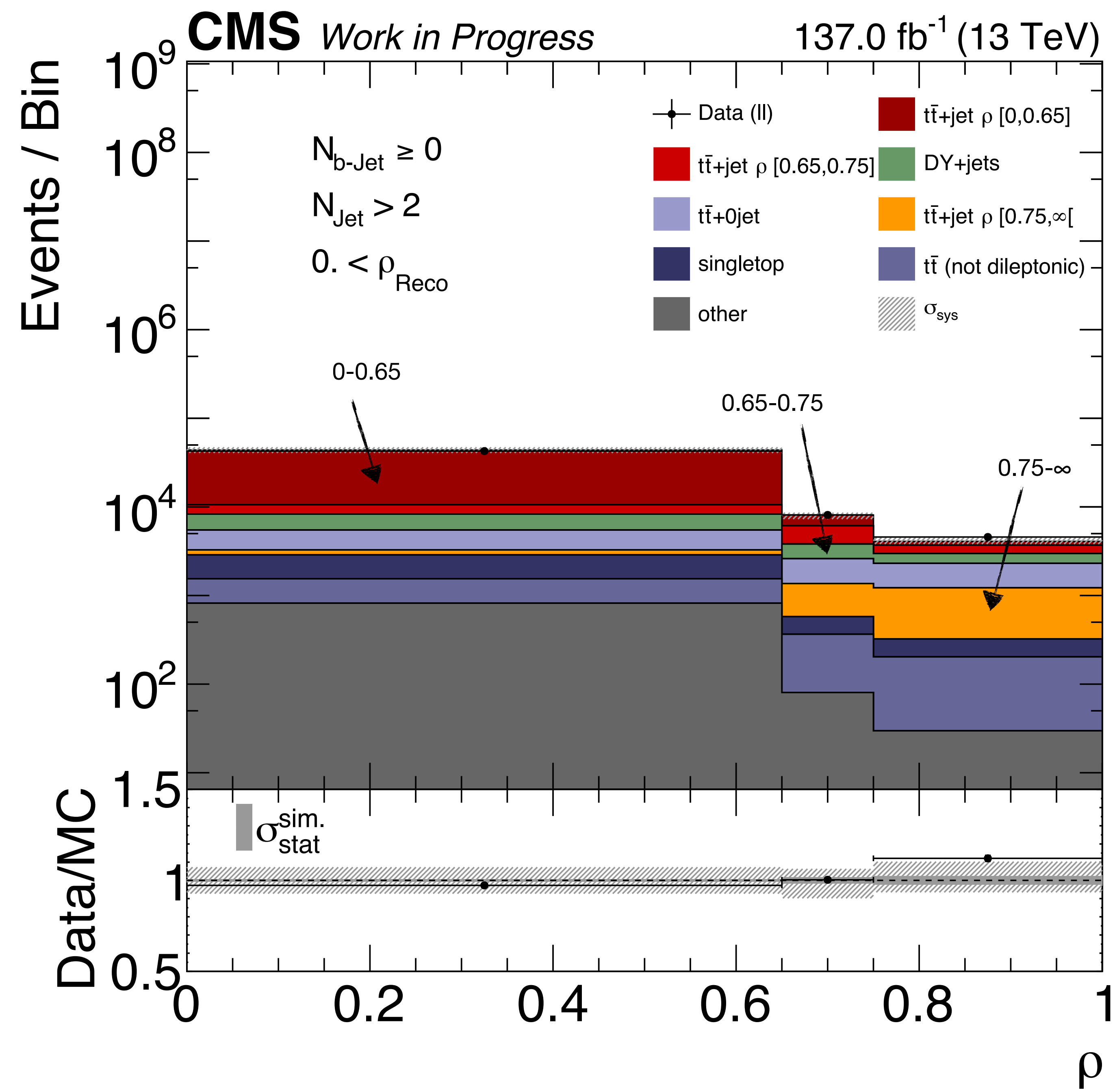
Reconstructed ρ

Data/MC

- using kin. reco+DNN reconstruction
- binning in ρ (gen+reco):
  3 bins:
  [0, 0.65, 0.75, ∞]

- $t\bar{t}$ separated into:
  - $t\bar{t}$ bkg ($t\bar{t} \rightarrow$ l+jets, $t\bar{t} \rightarrow$ hadrons, $t\bar{t} \rightarrow \tau\bar{\tau}$ (not dileptonic))
  - $t\bar{t}$+0jet bkg ($t\bar{t}$ without add. jet)

- $t\bar{t} \rightarrow \tau\bar{\tau} \rightarrow$ ee, μμ, eμ treated as signal



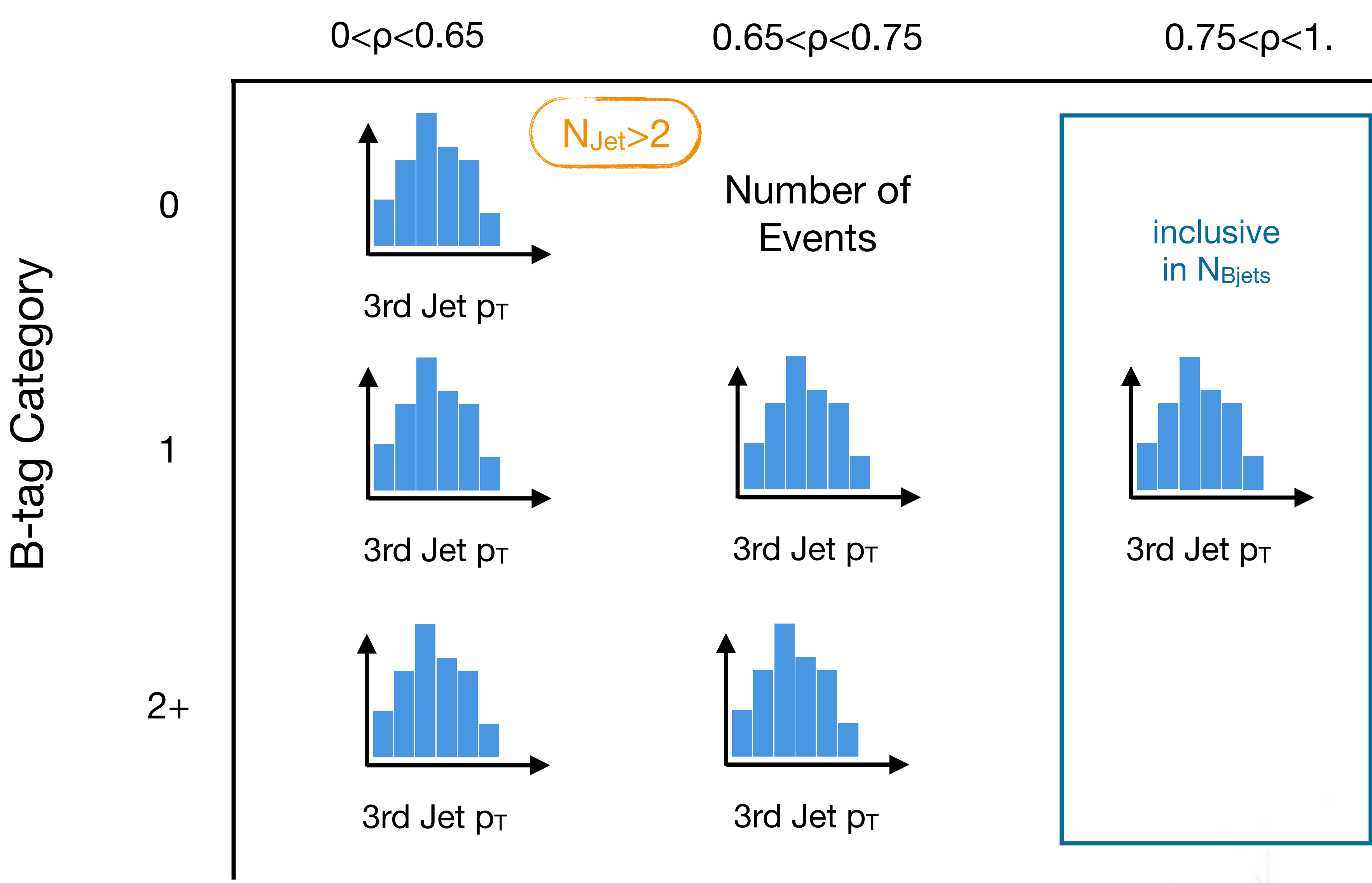
$$\rho = \frac{2m_0}{m_{t\bar{t}+1jet}}, m_0 = 170GeV$$

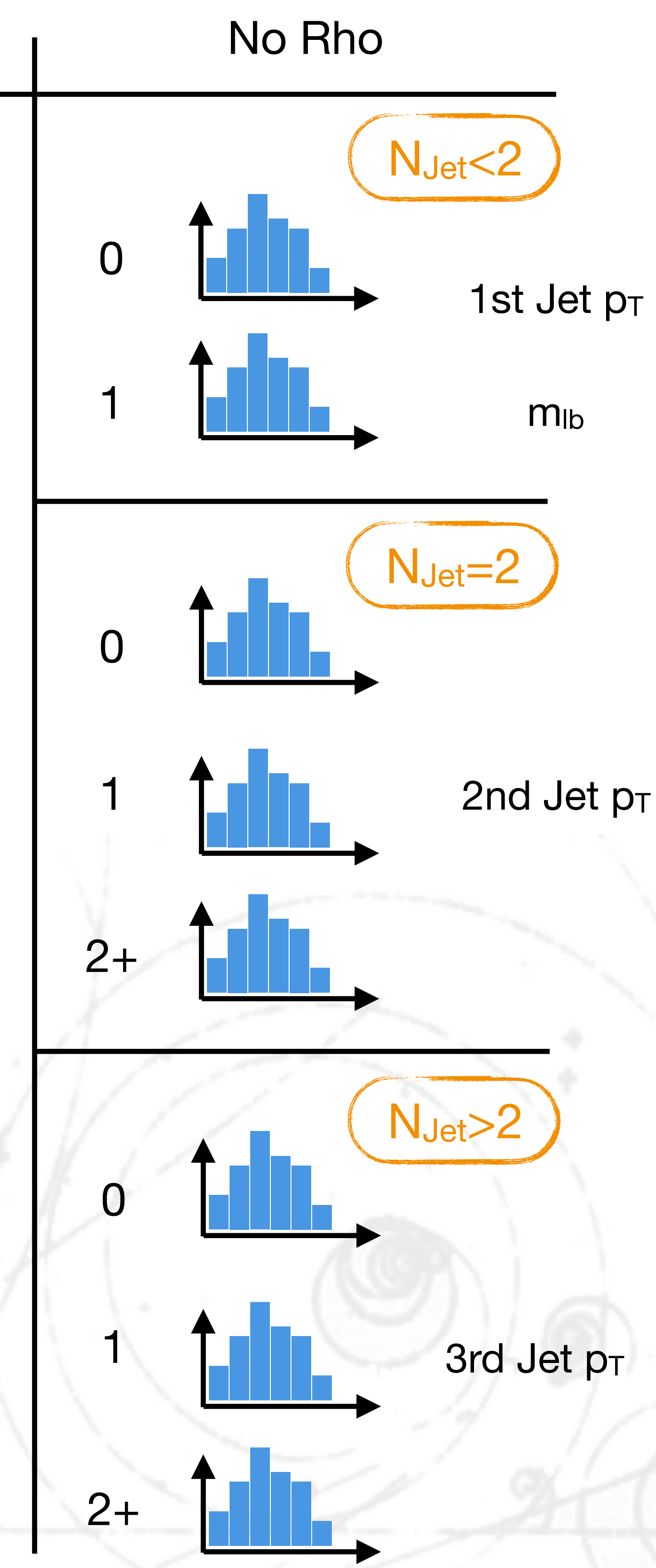
Maximum Likelihood Unfolding

Categorisation

ρ_{Reco} Category

constrain backgrounds & systematics





- fit method accounts for:
  - bin-to-bin migrations
  - background subtraction
  - correlations of uncertainties
- introduce signal strength parameter for each generator bin (n=3)
- eμ and same flavor lepton categories
- fit normalisation of DY and $t\bar{t}$ background simultaneously

Results using Asimov dataset

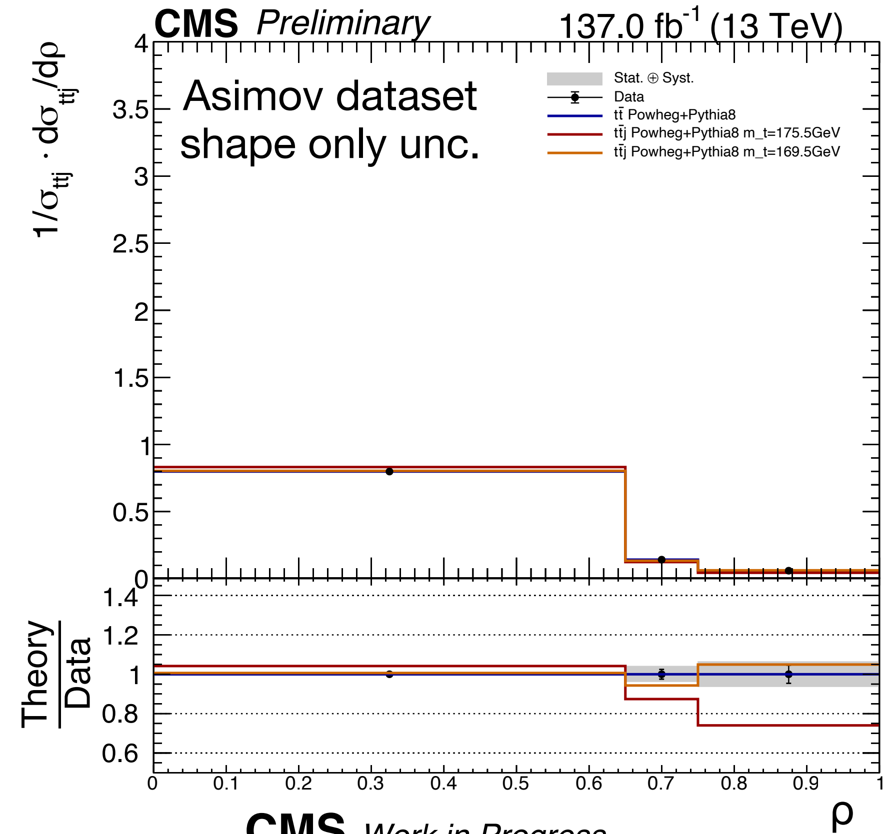
Differential cross section and mass extraction

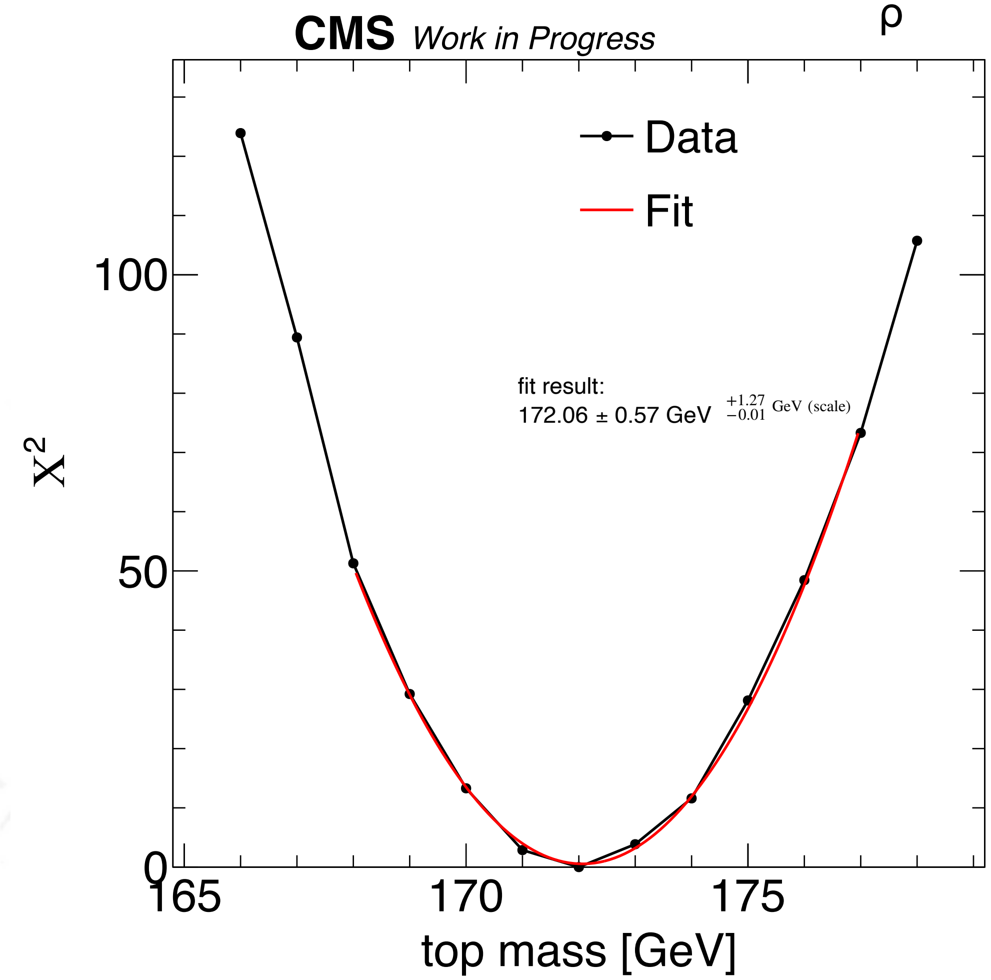
- precision estimate on top mass extraction using:
  - fixed order NLO theory prediction
  - relative estimated uncertainty from Asimov fit on normalised distribution
- χ^2 scan for top masses using XFitter
  - make use of statistical + systematic errors and bin-to-bin correlations
- fit parabola of form $\chi^2(m_t) = p_0 + (\frac{m_t - p_2}{p_1})^2$
  - $p_0 \triangleq \chi^2$ at minimum
  - $p_1 \triangleq$ estimated uncertainty
  - $p_2 \triangleq$ estimated top mass

⇒ 0.57 GeV experimental uncertainty

vs.

1.28 GeV scale uncertainty

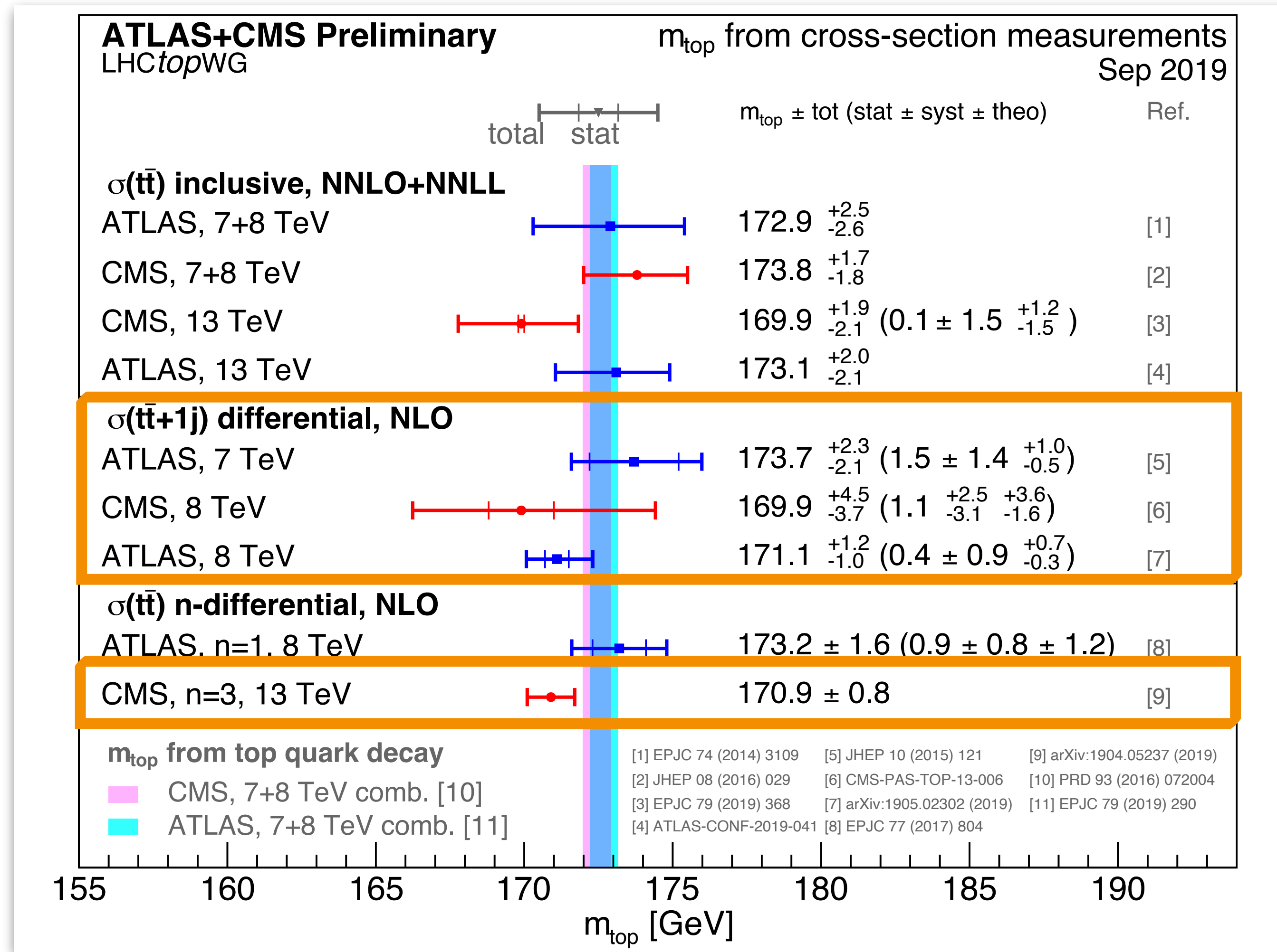




Summary

Conclusion & Outlook

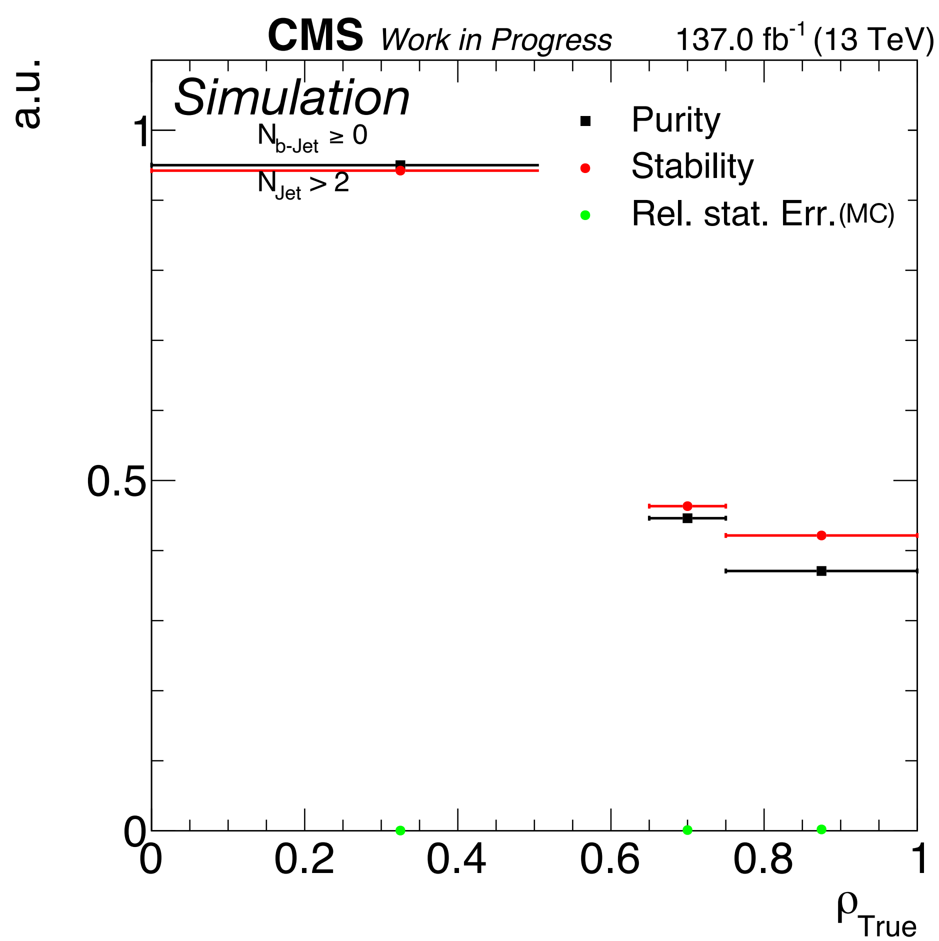
- presented "new" analysis to measure top quark mass in different mass schemes using differential $t\bar{t}$+1jet cross section
- sensitivity improvements thanks to:
  - new reconstruction techniques
  - high FR2 statistics
  - optimised unfolding
- measurement of m_t by comparing to theoretical calculations with comparable precision

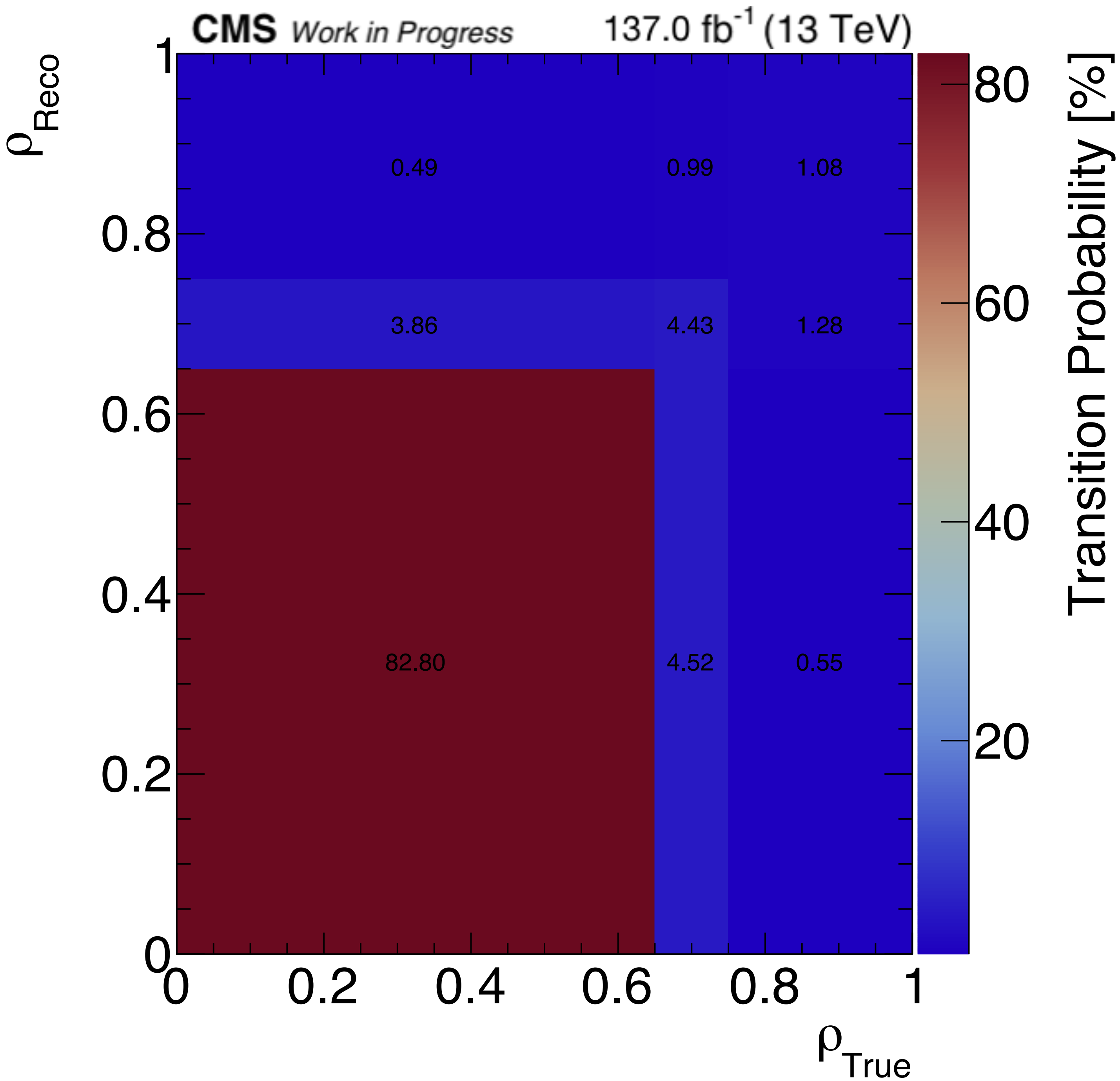


Backup

Reconstructed ρ

Choice of binning for unfolding





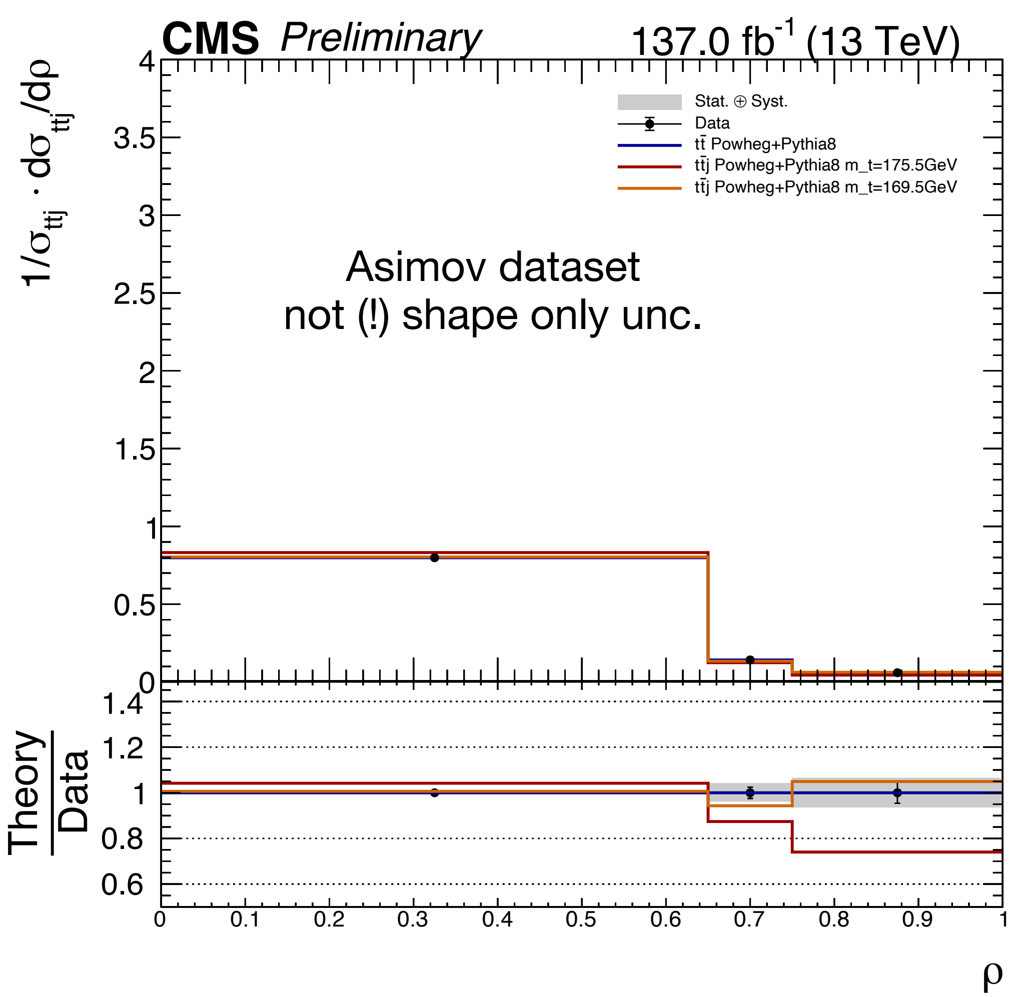
- binning in ρ based on following criteria:
  - purity and stability not below ~25%
  - reasonable condition number of response matrix
    → no regularisation in the unfolding step needed

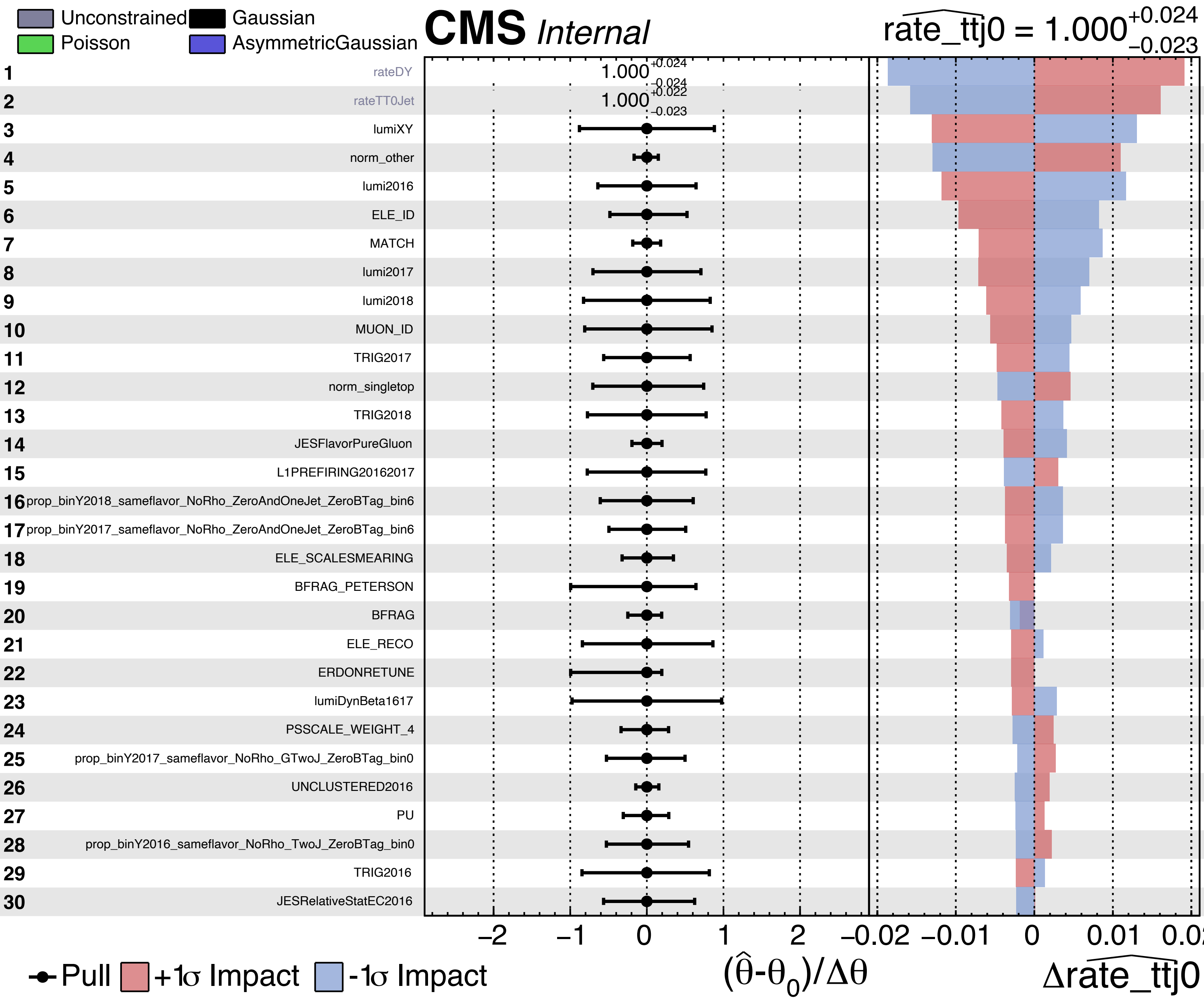
$$stability = \frac{N_i^{rec\&gen}}{N_i^{gen}} \qquad purity = \frac{N_i^{rec\&gen}}{N_i^{rec}}$$

Results

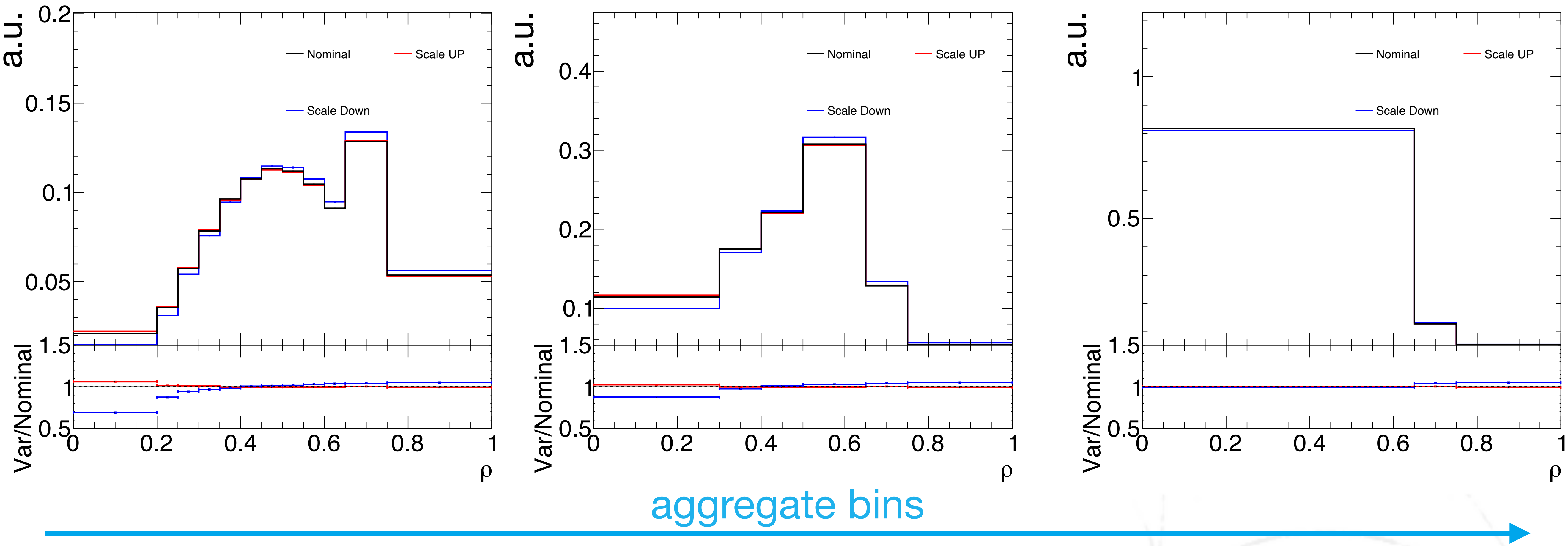
Asimov dataset

- Text bblabla





Scale uncertainty



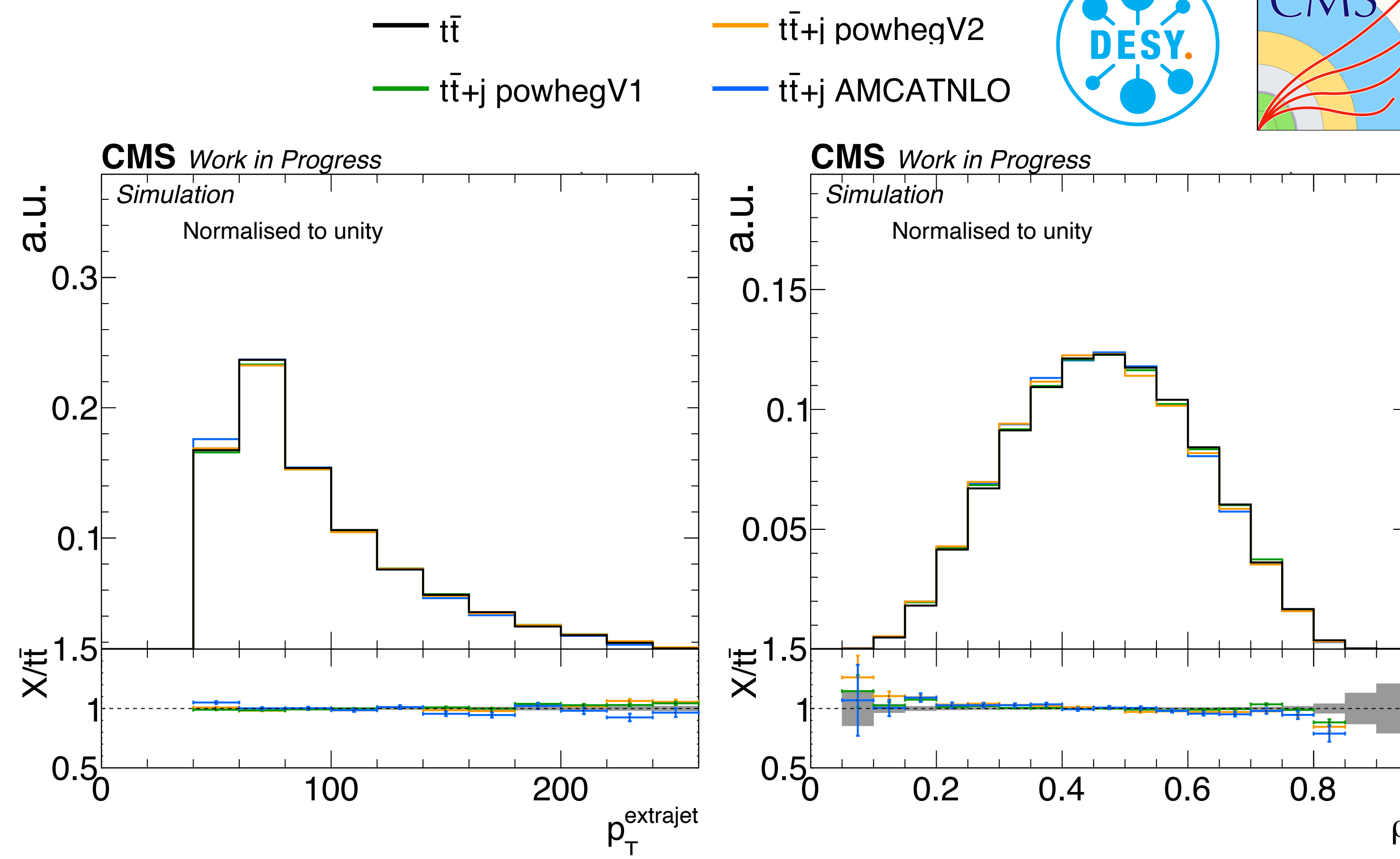
- using prediction in finest binning received by theorists:
  - 20 bins of size 0.05
  - rebin to mitigate scale variation effect at low rho
    - first bin needs to include at least the crossing point

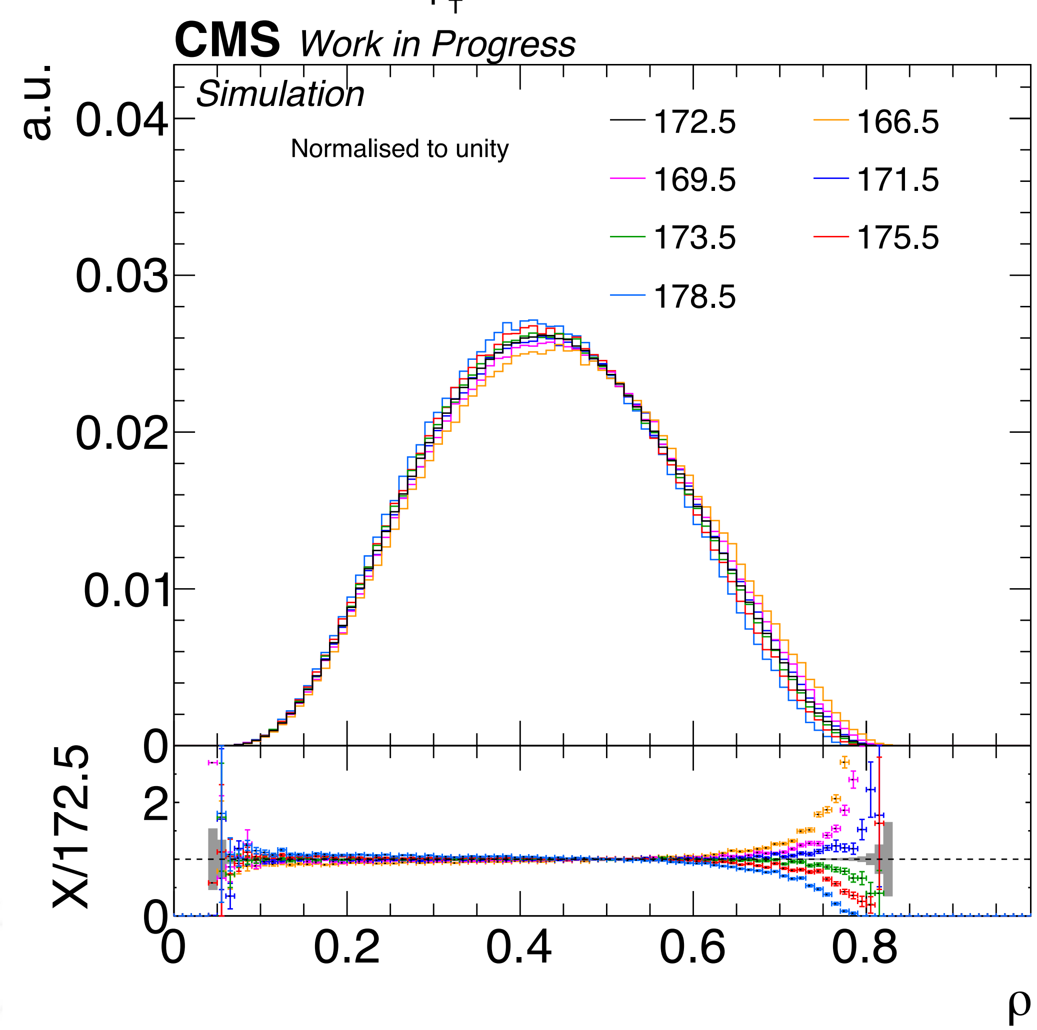
- private Powheg+Pythia8 $t\bar{t}$+jet samples in production to cross-check the effect
  - also for multiple PDF sets

$t\bar{t}$+1jet generator studies

- studies on Powheg NLO $t\bar{t}$+jet prediction:
  - compatibility with NLO Powheg $t\bar{t}$ simulation
    - good agreement after PS
  - mass sensitivity at 13 TeV similar as at 8 TeV

- consistent definition of ρ:
  - "stable" tops at parton level + AK4 jets clustered on stable gen particles before hadronisation





Analysis setup

Selection

Electrons

- Cut based Tight ID
- $|\eta| < 2.4$
- exclude:
  $1.4442 < |\eta| < 1.5660$

Muons

- Tight ID
- $|\eta| < 2.4$
- Tight pfRelIso ($I < 0.15$)

Jets

- Loose PF ID
- $p_T > 30$ GeV
- $\Delta R(l, \text{jet}) > 0.4$
- b-Jet energy regression

Lepton Pair

- $p_T > 25$ (20) GeV for leading (trailing) lepton
- $m_{ll} > 20$ GeV
- Veto events with add. leptons
- For ee/μμ exclude:
  76 GeV $< m_{ll} <$ 106 GeV

p_T^{miss}

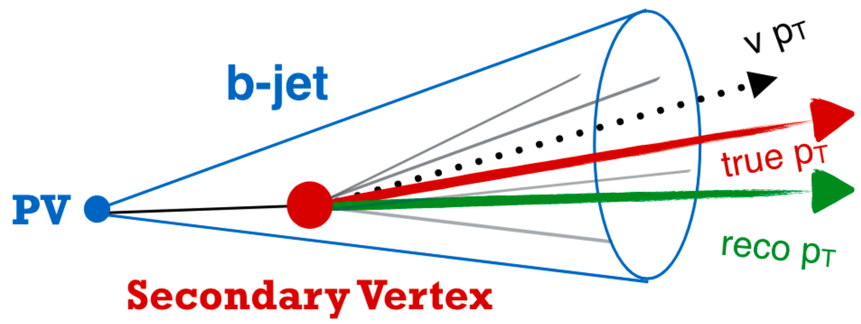
- PUPPI-MET
- MET filters
- MET > 40 GeV (ee, μμ)

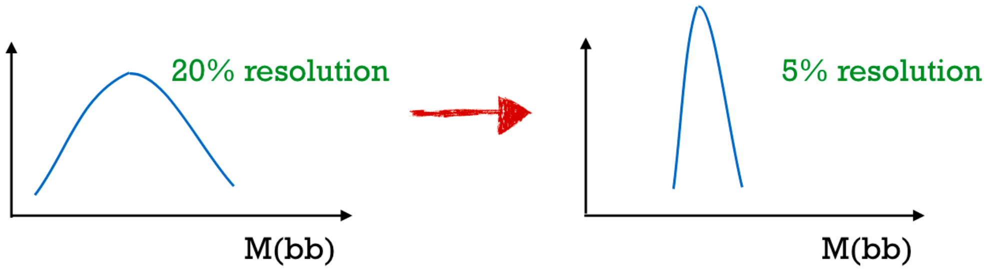
b-jets

- DeepCSV
  Loose WP

BJet Energy Regression

- developed by ETH et al. for HH→ggbb
  - see paper and AN-18-092
  - goal: correct mismeasured energy of b-jets because of escaping neutrinos
  - dijet mass resolution: 20%→5%
  - no change in handling of systematic uncertainties
  - calculated from uncorrected jets
  - apply e.g. JES→Regression→JER





- idea: improve $t\bar{t}$ mass resolution

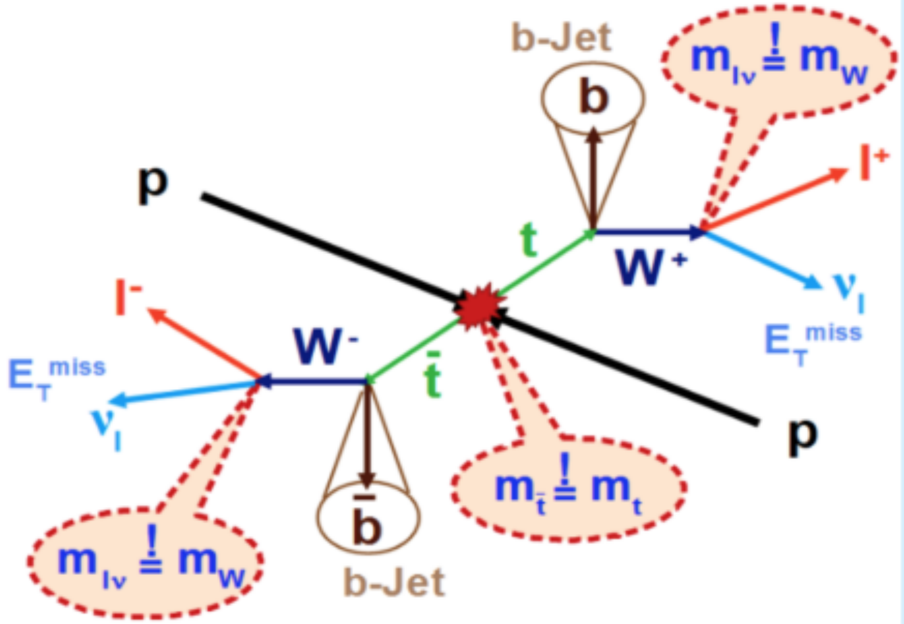
- changed event yields:
  - after full kinematic reconstruction:
    - $t\bar{t}$ signal: 301636 (nom.) vs. 316393 (Regression) (+5%)
    - estimate for $t\bar{t}$ +j: +25%

Analysis setup

Kinematic Reconstruction of the $t\bar{t}$-system

arXiv:1811.06625

- inputs:
  - 2 jets
  - 2 leptons
  - MET
- constraints:
  - m_t, $m_{\bar{t}}$ =172.5 GeV
  - m_{W+}, m_{W-} = 80.4 GeV
  - $p_T(\nu,\bar{\nu})$ = MET
- unknowns:
  - 3-momenta: ν, $\bar{\nu}$
- solutions with b-tagged jets are preferred
- reconstruct event 100 times:
  - W mass smeared according to Breit-Wigner distribution
  - Lepton, b-Jet energies smeared according to detector resolution
  - Weights are calculated based on m_{lb} spectrum
  - Take weighted average as solution
- Efficiency > 90%



$$p^{top}_{x,y,z} = \frac{1}{w} \sum_{i=0}^{100} w_i \cdot (p^{top}_{x,y,z})_i$$

Analysis setup

Kinematic Reconstruction of the $t\bar{t}$-system - another approach

arXiv:1904.05237

- "Loose kinematic reconstruction":
  - drop m_t requirement
    - no bias on top mass
  - reconstruct $\nu\bar{\nu}$ system as a whole
    - only total $t\bar{t}$-system reconstruction
  - $p_T(\nu\bar{\nu}) = p_T(MET)$
  - $p_z(\nu\bar{\nu}) = p_z(l\bar{l})$, $E(\nu\bar{\nu}) = E(l\bar{l})$
  - requirements: $M(\nu\bar{\nu}) \geq 0$, $M(W^+W^-) \geq 2m_W$

  - prefer solutions with higher b-tagged jet multiplicity over others

  - Obtained kinematics similar to full kinematic reconstruction
  - Efficiency similar

NN setup

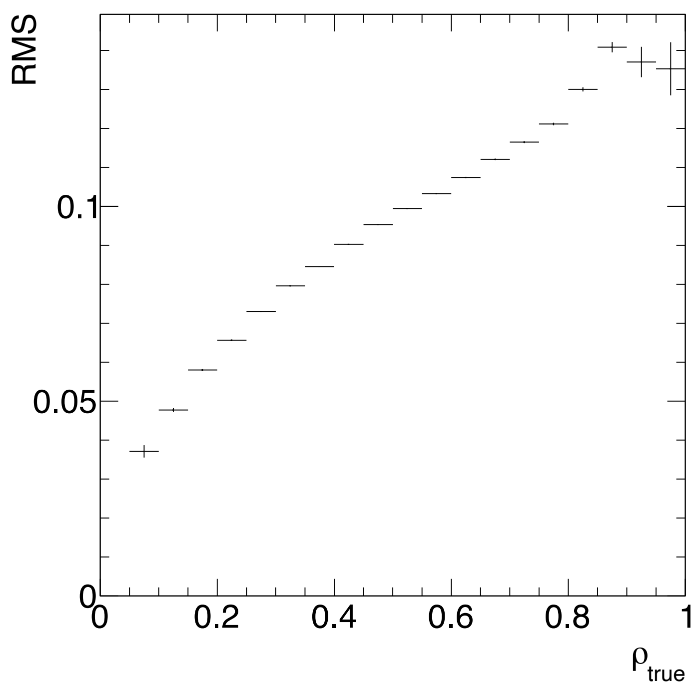
Technical details

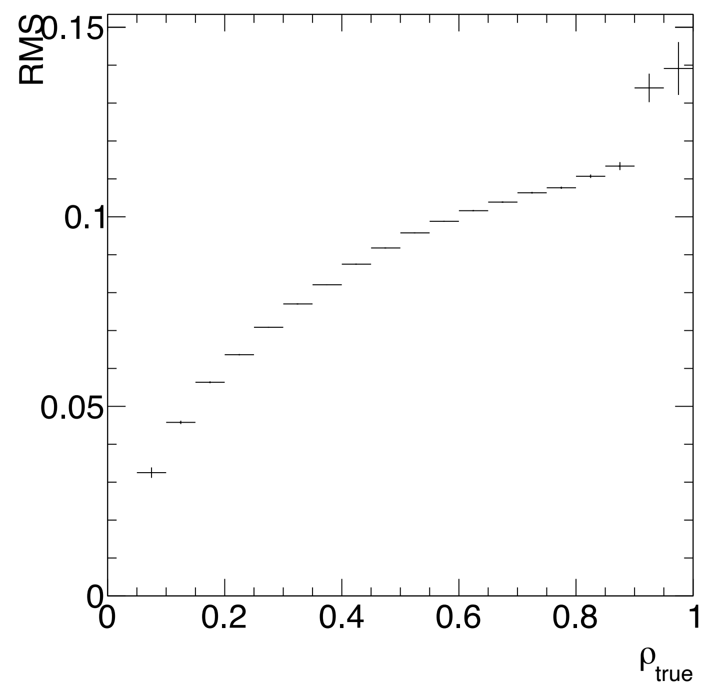
- using tensorflow with keras backend
- Adam optimizer
- decreasing learning rate with epochs
- early stopping
- train-val-test split: 0.6-0.2-0.2
- using Bayesian optimisation for hyper parameter optimisation

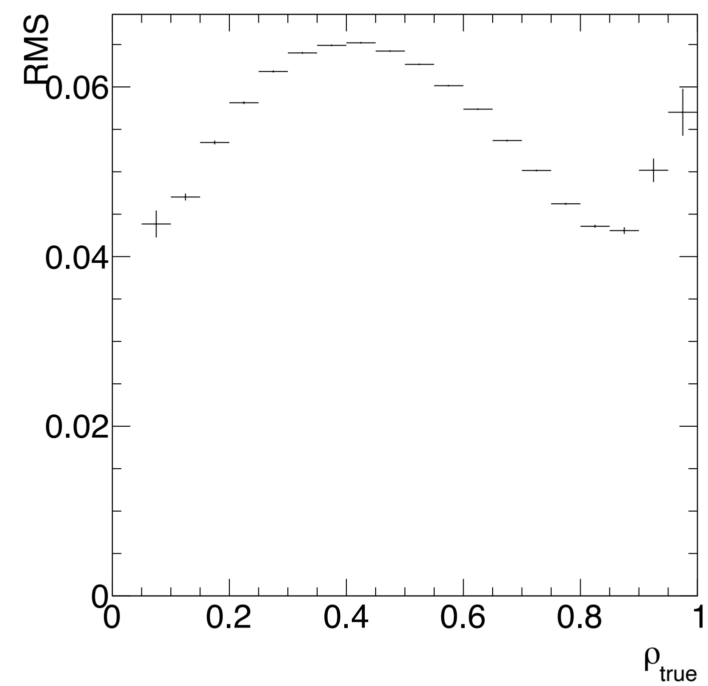
- regression task:
  - LR: 0.007034
  - L2 regularisation: 0.000179
  - batch size: 6250
  - dropout: 0.3673
  - 3 dense layers with 500 nodes
  - SeLU activation per layer
  - final linear activation
  - loss function: mean squared error

Reconstructed ρ

Comparison of various methods



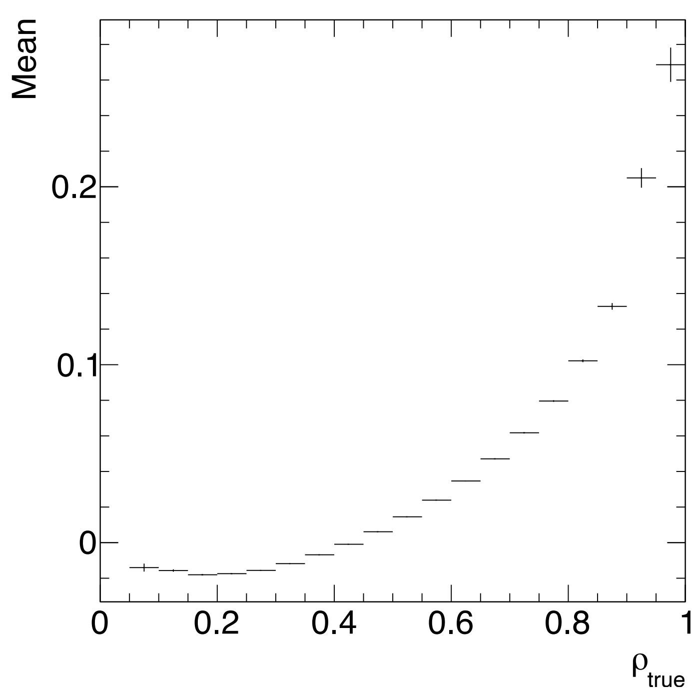


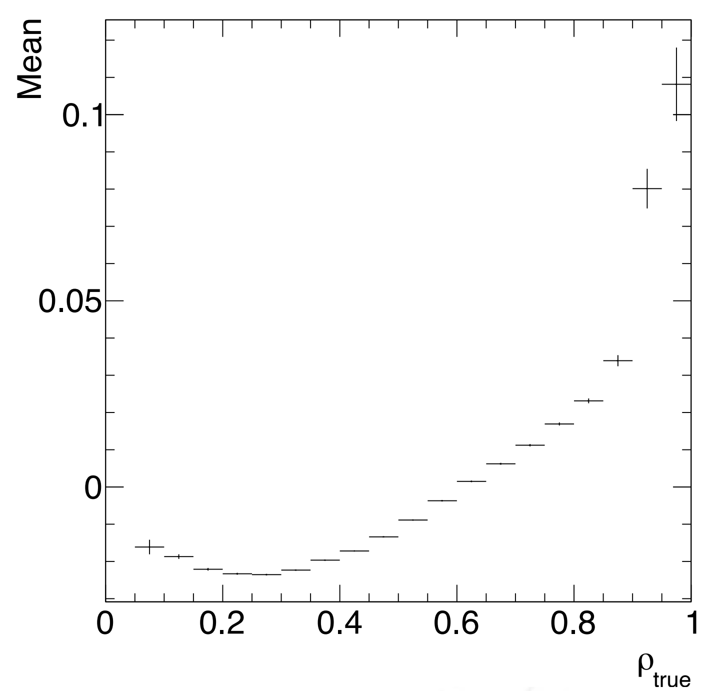


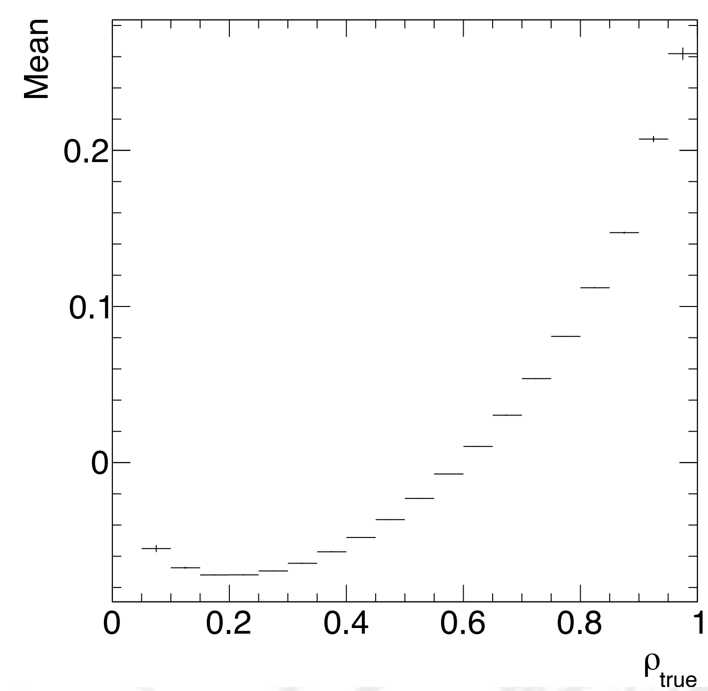
"full" kin. reconstruction

"loose" kin. reconstruction

DNN



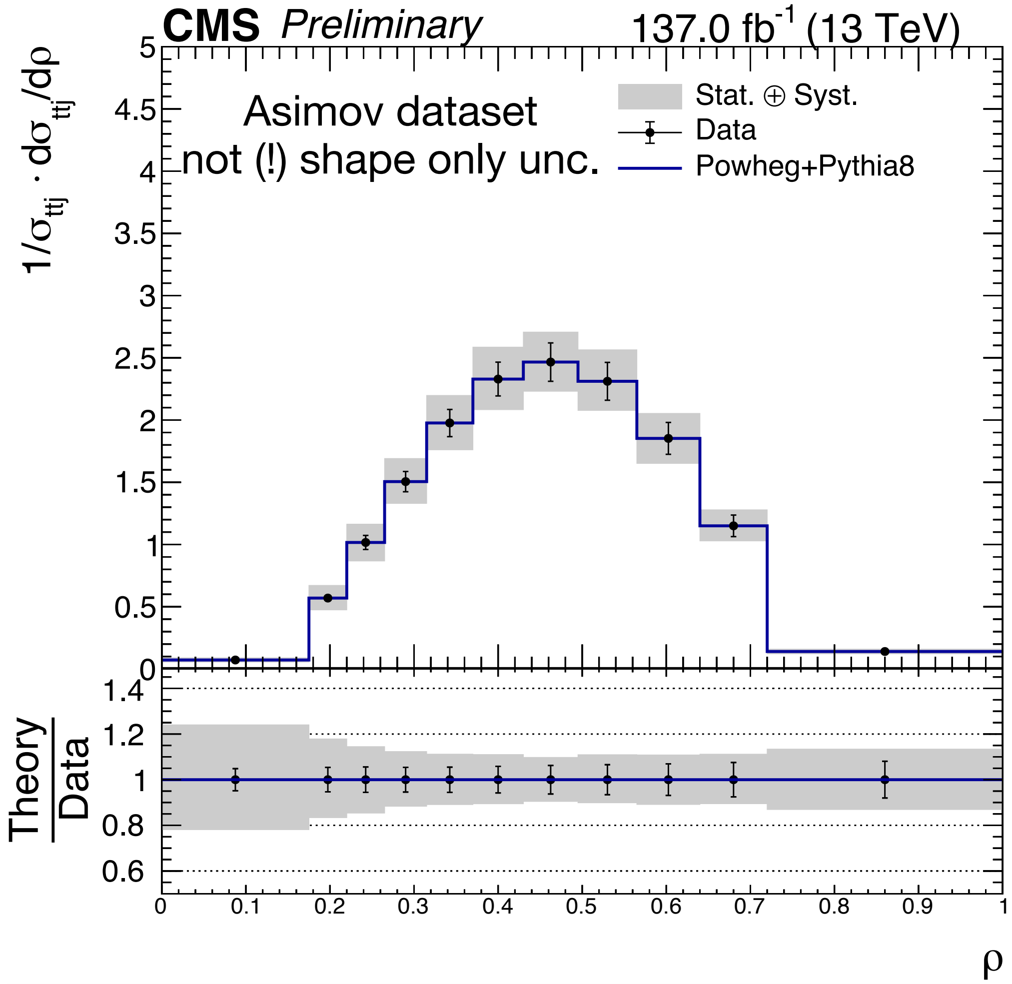




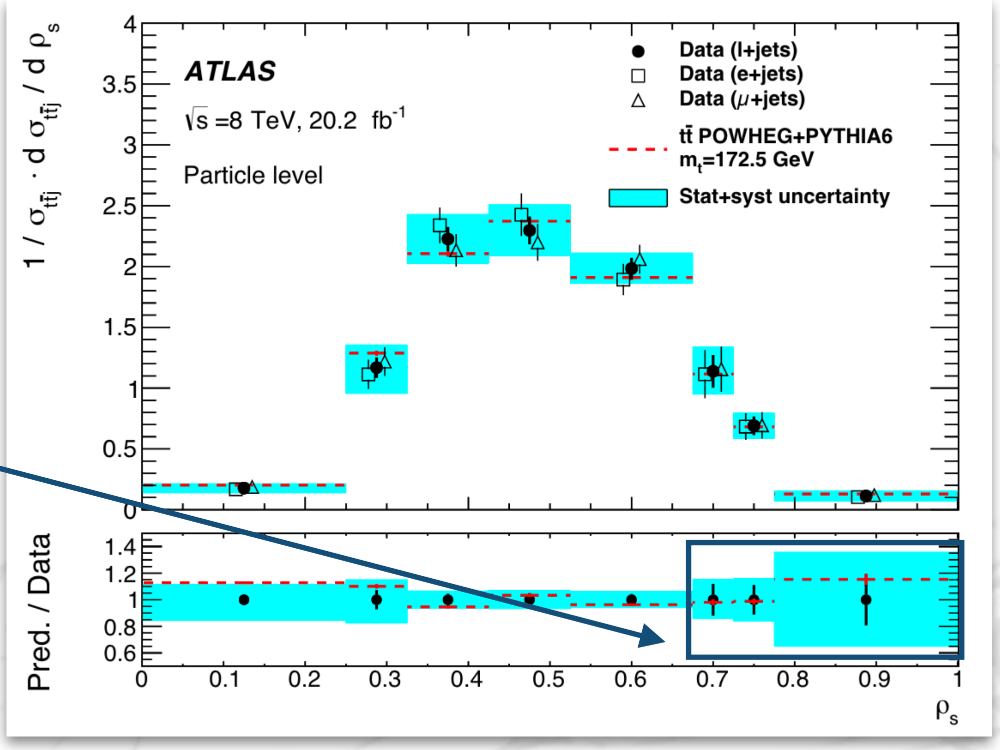
Sensitivity Comparison

ATLAS

- with optimised truth/reco binning and categorisation
  → rel. uncertainty of <15% per xSec bin seems feasible
  ⇒ uncertainty on extracted mass < 1GeV



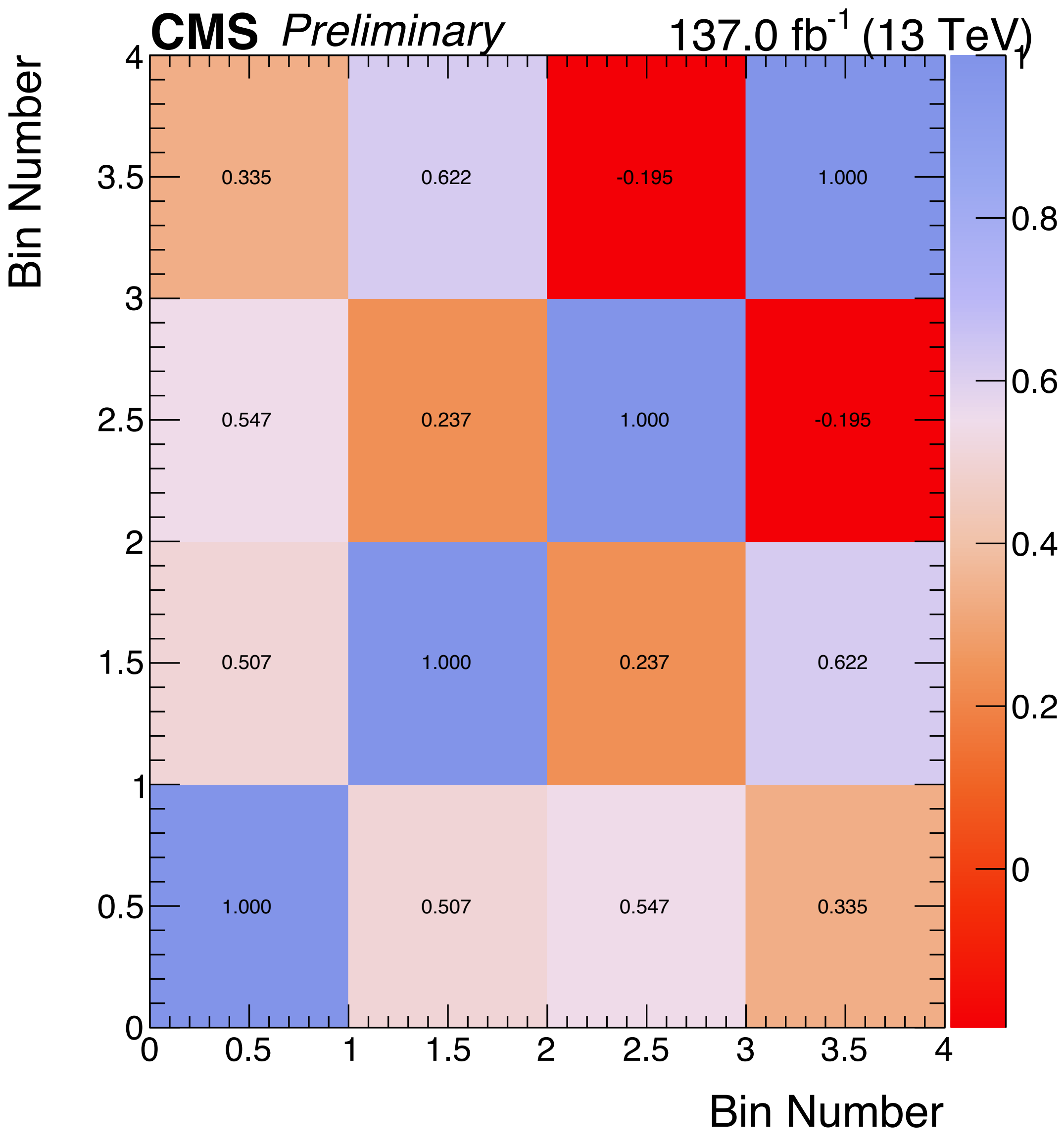
- ATLAS: arXiv:1905.02302
  - $20.2 fb^{-1}$ @ TeV
  - $m_t^{pole} = 171.1 \pm 0.4(\text{stat}) \pm 0.9(\text{syst})^{+0.7}_{-0.3}(\text{theo})$
- based on uncertainty on normalised(!) cross section:
  - 15-35% in highest sensitivity bins
  - theory uncertainties will be comparable



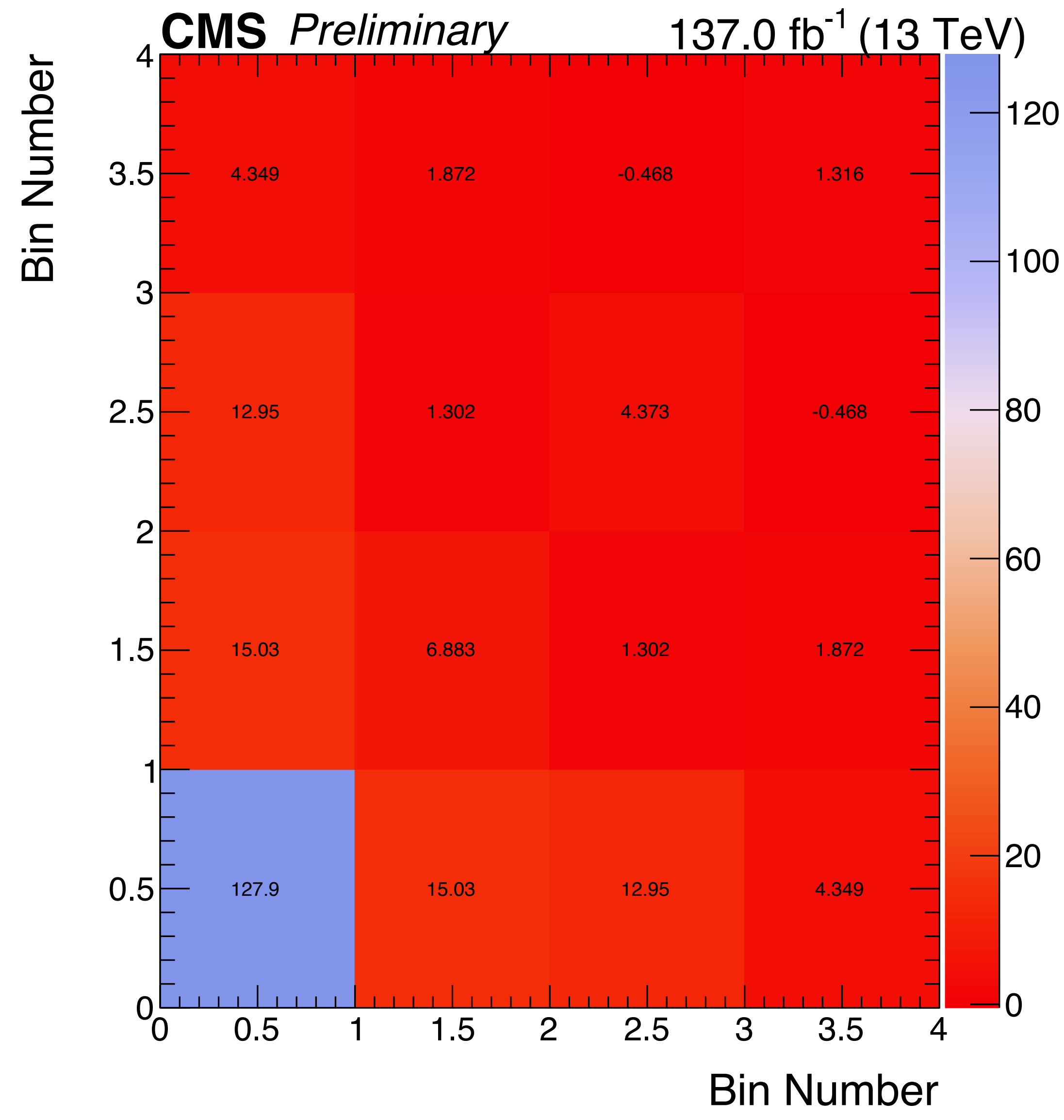
Unfolding

Correlation and Covariance Matrix

Correlation Matrix



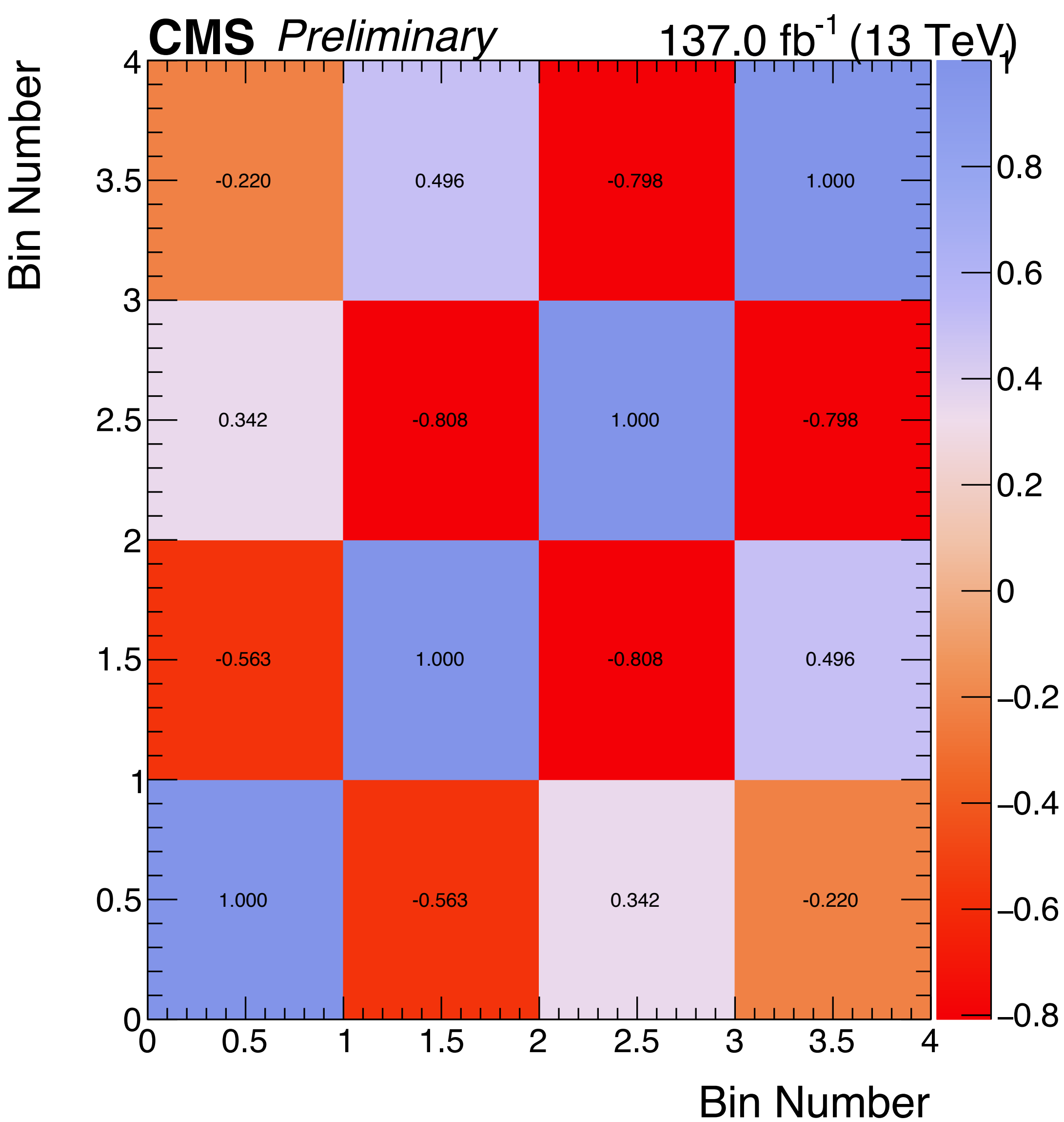
Covariance Matrix



Unfolding

Correlation and Covariance Matrix - stat. only

Correlation Matrix



Covariance Matrix