Novel analysis methods & techniques in CMS

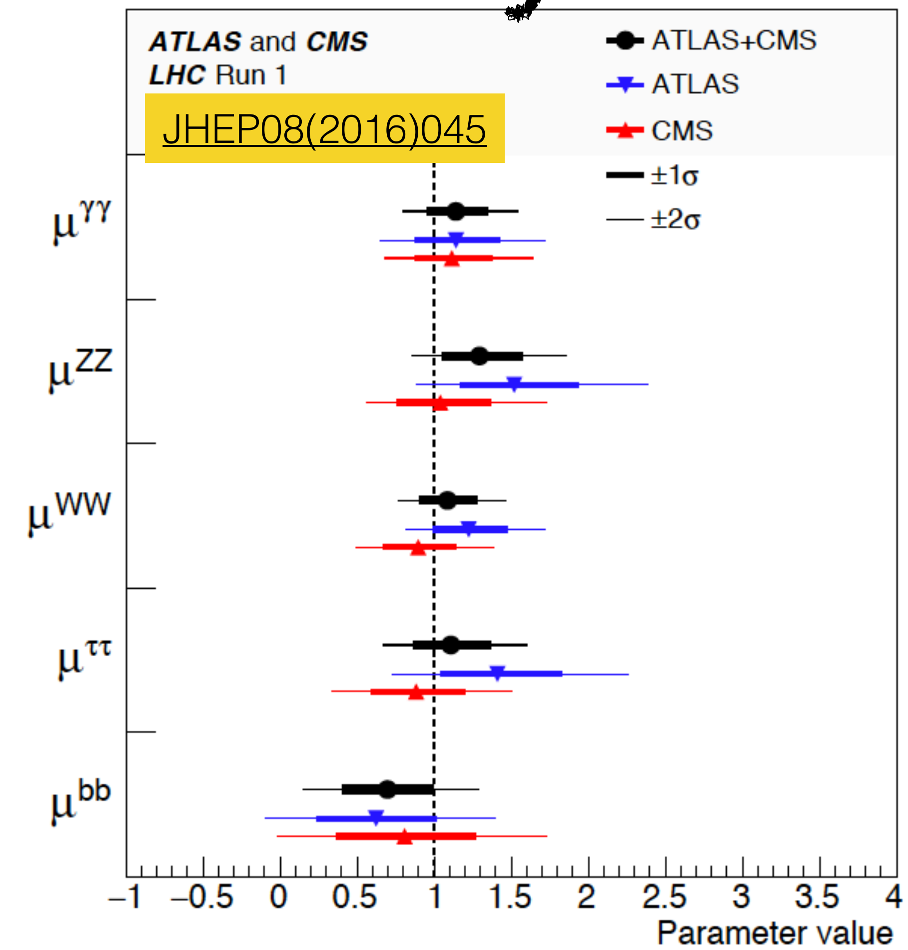
Alessandro Calandri - ETH Zürich
on behalf of the CMS Collaboration

Higgs Hunting 2023 - September 11-13, Orsay-Paris

10+ years after the Higgs Boson discovery

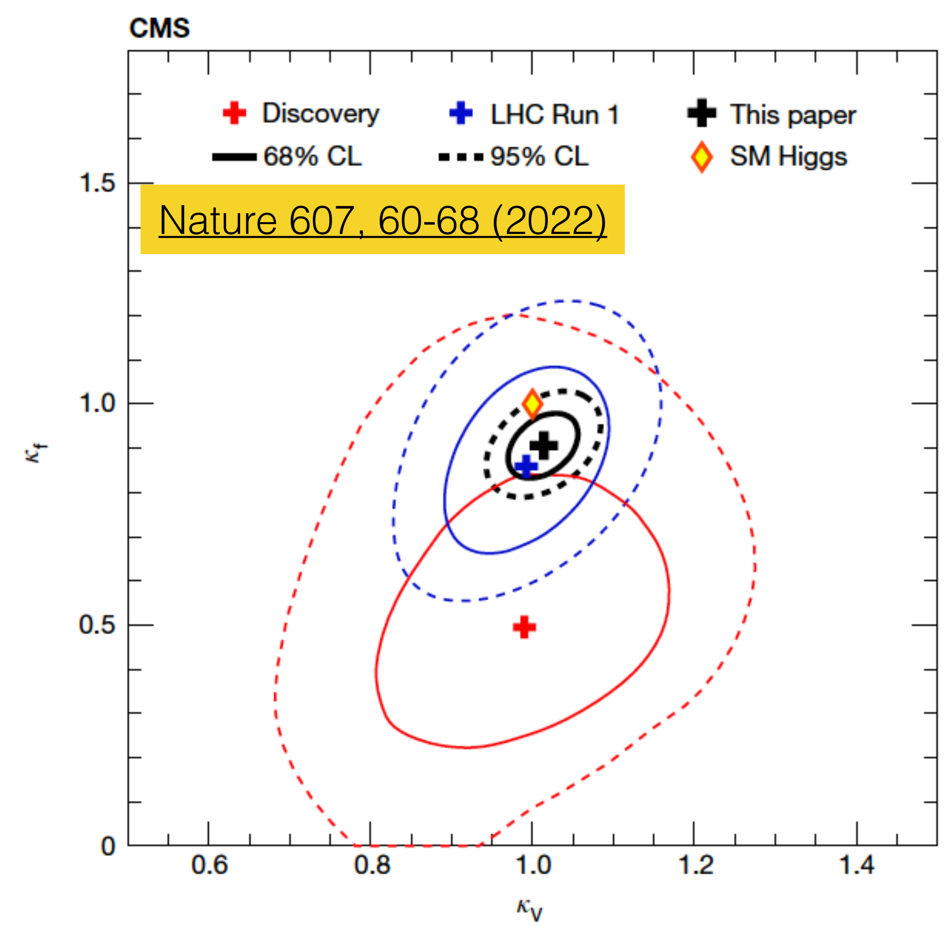
➡ **Higgs discovery using 5 (H→ZZ→4l, H→γγ, H→WW→lν, H→ττ, VH→bb) main decay channels and cut-based analyses**

- signal extraction performed on simple/physics observables (invariant mass, transverse mass, …)
- results provided inclusively in production modes, mostly dominated by gluon-fusion



➡ **10 years later in the 2022 ATLAS/CMS Nature paper measurement techniques have evolved significantly**

- ML techniques used as backbone of Higgs analyses to perform several tasks (object reconstruction, calibration, signal extraction)
- cross-section measured in several production modes with excellent precision (6% inclusive cross-section)
- accessed several kinematic regions with dedicated differential and coupling measurements, STXS
  - evolution of measurements: from inclusive signal strength to STXS, from kappa-framework to EFT to reduce model dependency



Outline of the talk

➡ **Highlights of <u>analysis technique developments/novel approaches</u> for object performance, signal extraction, background predictions for CMS Higgs physics**

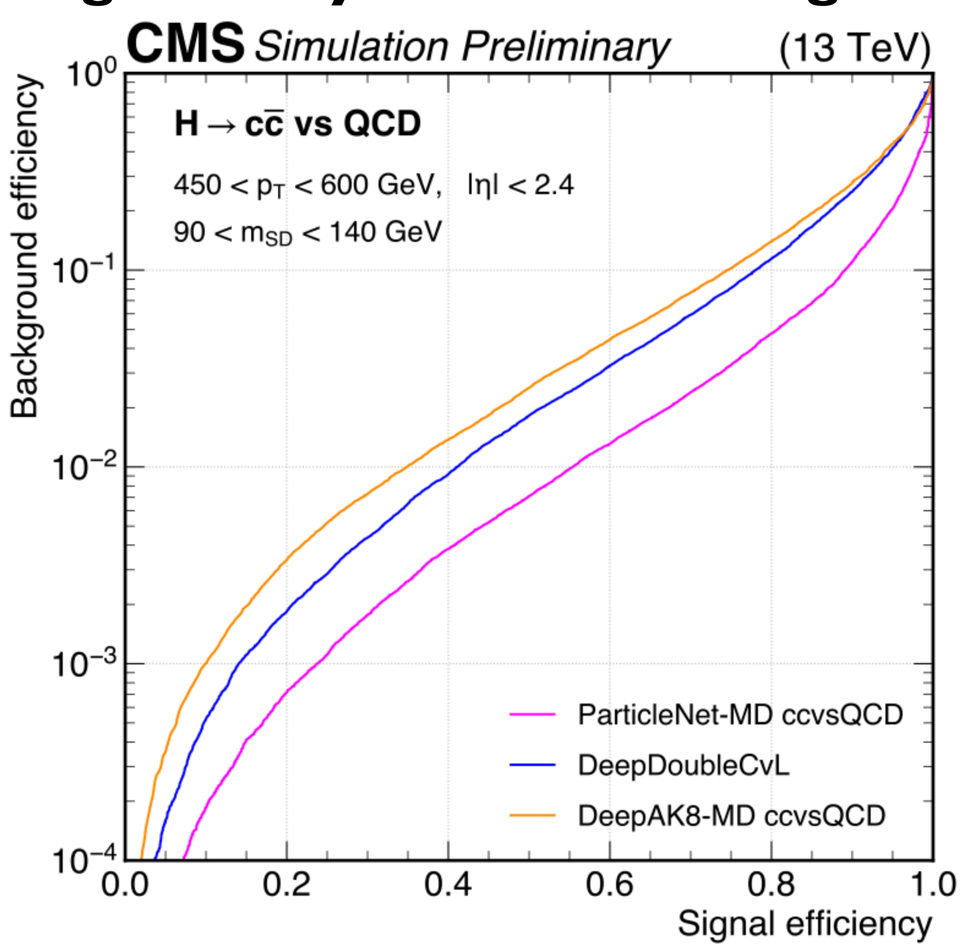
- <u>Object performance</u>
  - improvements of jet flavour tagging and impact on full Run 2 Higgs legacy analyses
  - accessing new Higgs topologies: merged jet topologies employing jet substructure techniques
  - advancements in calibration techniques especially for jet flavour tagging in boosted topologies
- <u>Improvements on object precision</u>
  - per-event resolution and energy regressions (backup material)
  - mitigating systematics uncertainties due to mismodeling in simulation
- <u>Evolution on signal extraction methods and background modelling</u>
  - ML techniques in Higgs analysis strategies
  - improved background modeling (motivations, the ttH→bb example and going beyond CMS)
- <u>several advancements on Higgs framework interpretations: STXS, differentials, enhancing sensitivity to anomalous/EFT</u> (not discussed in this talk)

Jet tagging techniques and impact on Higgs measurements

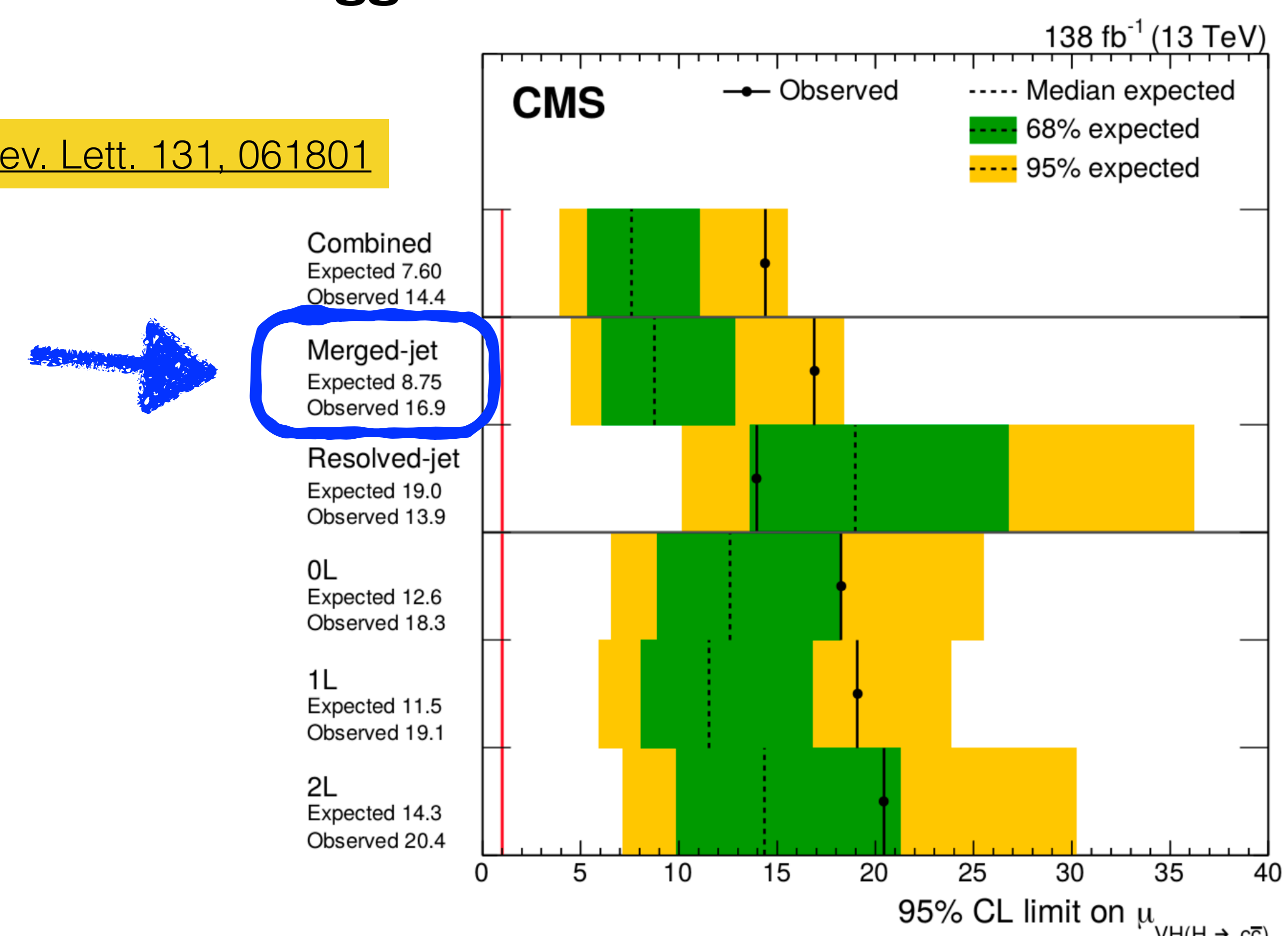
➡ **Flavour tagging (FT) represents a critical feature for Higgs to heavy-flavour - change of paradigm for CMS Run 2 flavour tagging has paid off**

- from high-level features (IP, SV, soft-leptons) to low-level observables (PF candidates)
  - DeepCSV/DeepFlavour for resolved (RNN); ParticleNet for boosted (point cloud architecture)
  - more ideas and algorithms to further improve classification performance/robustness [CMS DP-2022/050, CMS DP-2022/049, CMS DP-2022/051, CMS-DP-2023-021]
  - DNN-based (PN architecture) jet mass regression to predict AK15 mass with highest resolution (15-20% improvement on expected limit)

➡ **Significant leap in sensitivity in H→cc thanks to ParticleNet tagger performance and merged analysis - most stringent constraints on Higgs-charm Yukawa at LHC**

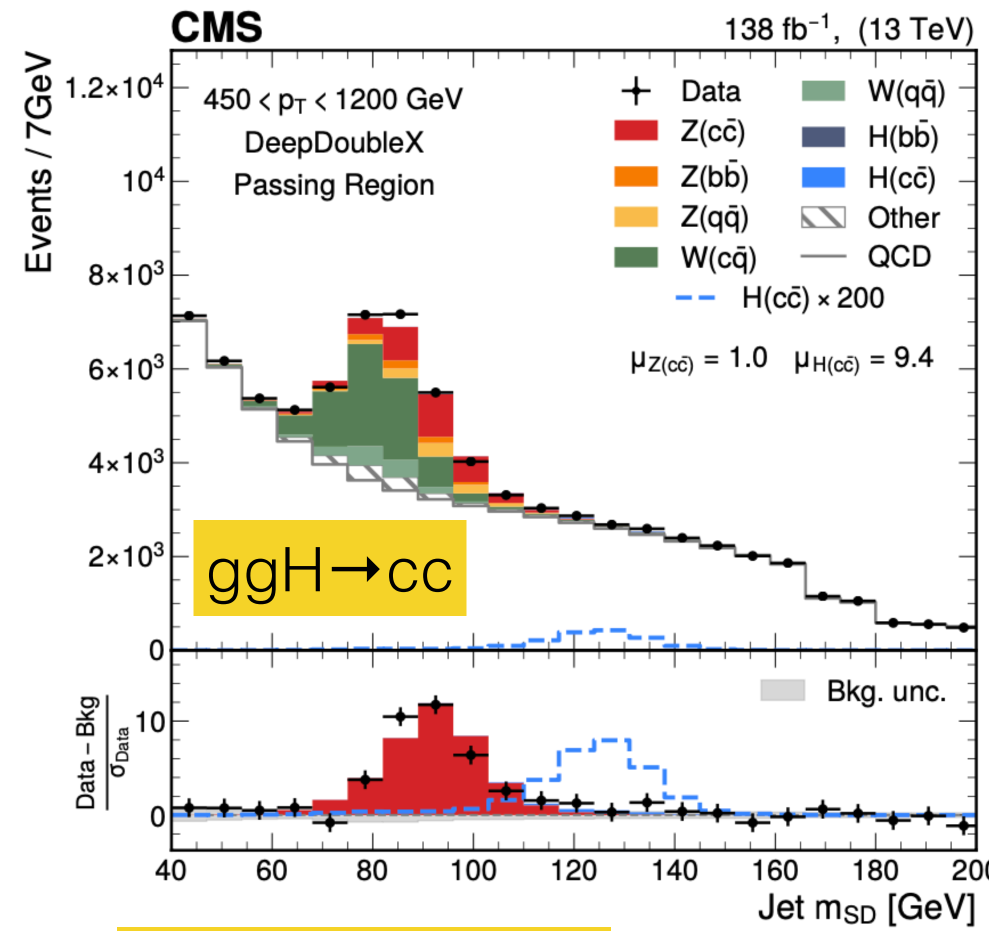


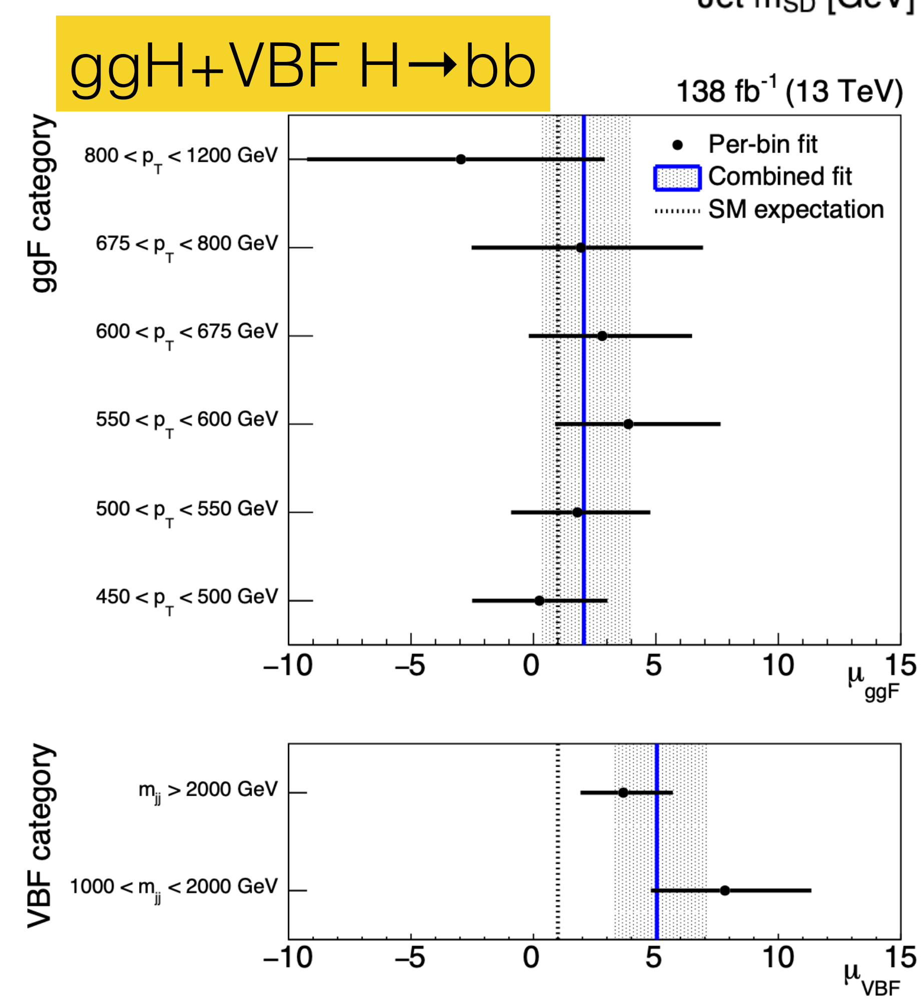
Phys. Rev. Lett. 131, 061801



Accessing new topologies - merged jet and substructures

- ➡ **Significant advancements on understanding jet substructure in boosted Higgs topologies**
    - from calorimeter-based energy to inclusion of PF
    - merged jet reconstruction robust against pile-up - pile-up mitigation techniques (PUPPI)
- ➡ **Crucial developments on grooming techniques for merged jet reconstruction**
    - enhancing soft-drop mass scale and resolution accuracy by removing soft and wide angle radiation in merged jet
- ➡ **Large gain using more complex DNN for double b-tagging in large-radius jets (DeepDoubleX)**
- ➡ **Successfully employed in full Run 2 boosted Hbb/Hcc analyses**
    - accessing high pt Higgs spectrum and sensitivity to gluon-fusion and vector-boson fusion production modes (Hbb)
    - excellent sensitivity for merged/substructure topologies

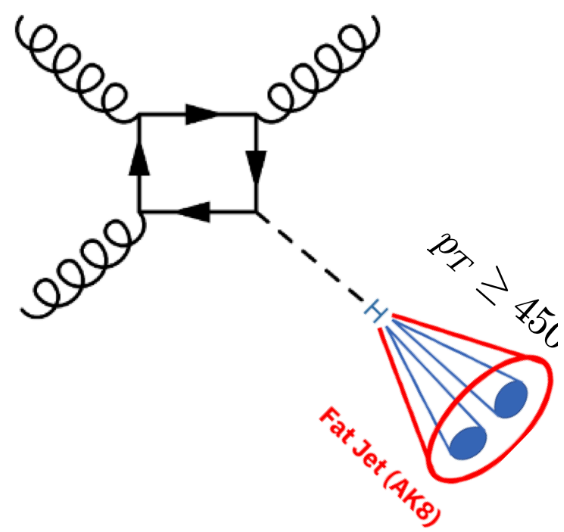




Phys. Rev. Lett. 131, 041801 CMS-PAS-HIG-21-020

Jet tagging calibrations for boosted topologies

➡ **Advancements on new jet flavour tagging goes in parallel with improvements on jet calibration techniques**

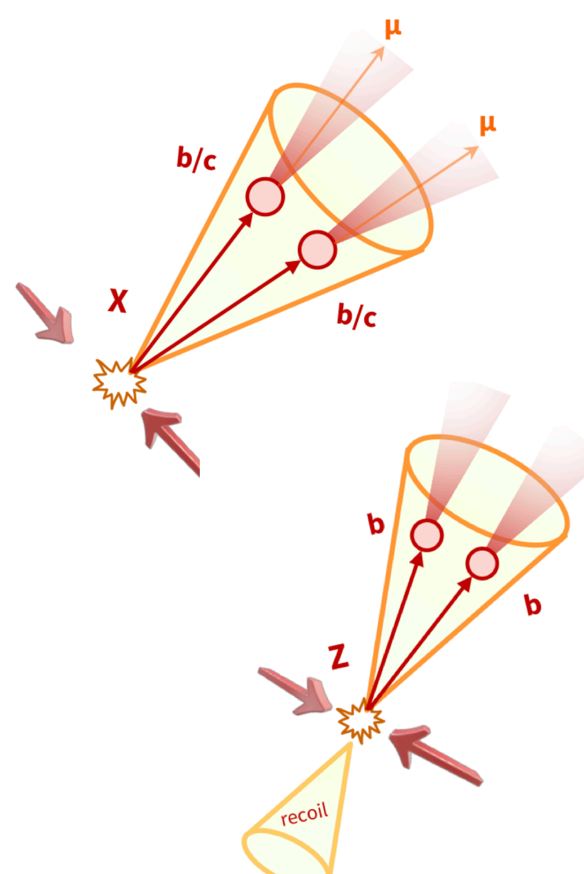


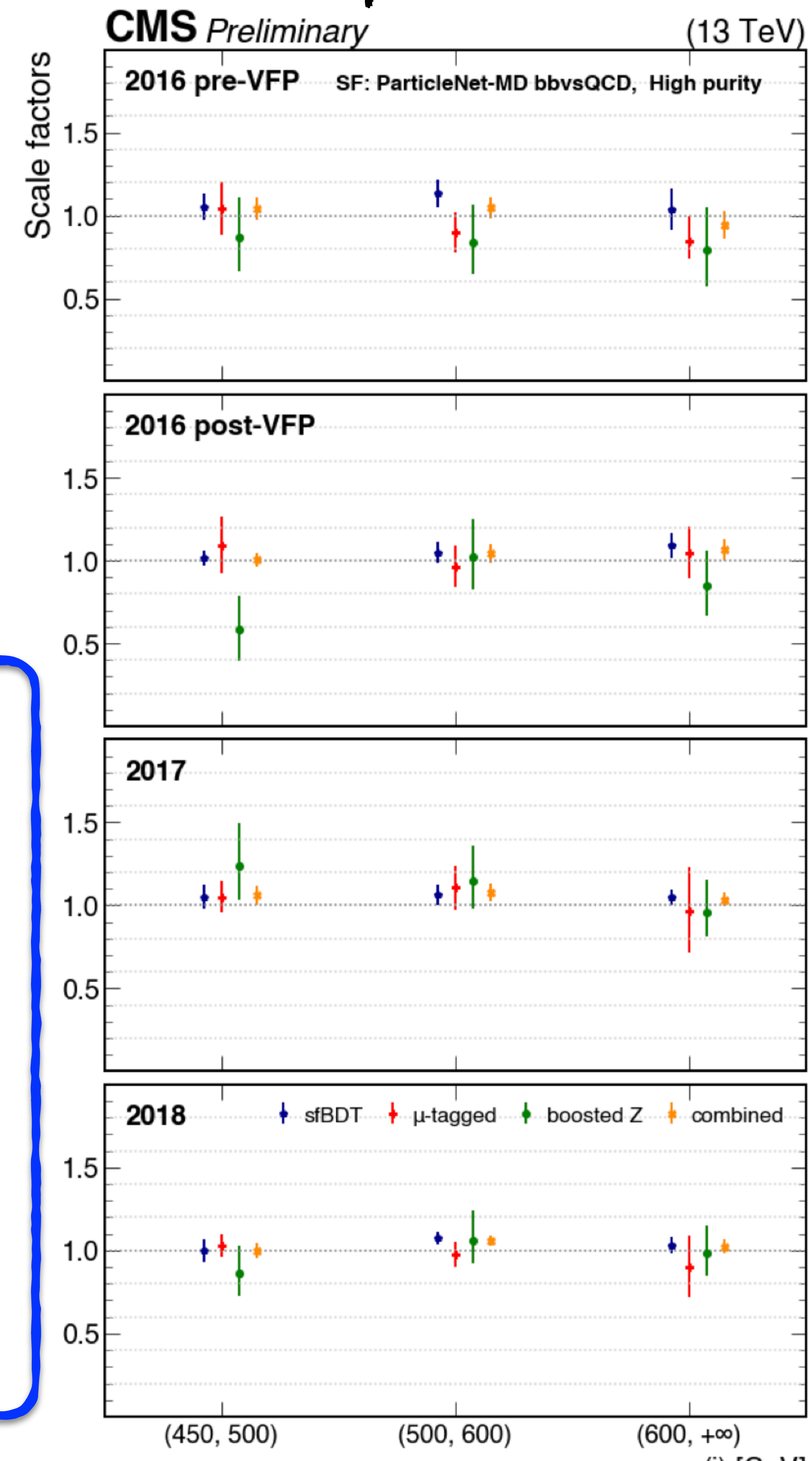
CMS-PAS-BTV-22-001

➡ **Most precise jet tagging calibration corrections to apply on simulation**

- targeting Higgs boosted

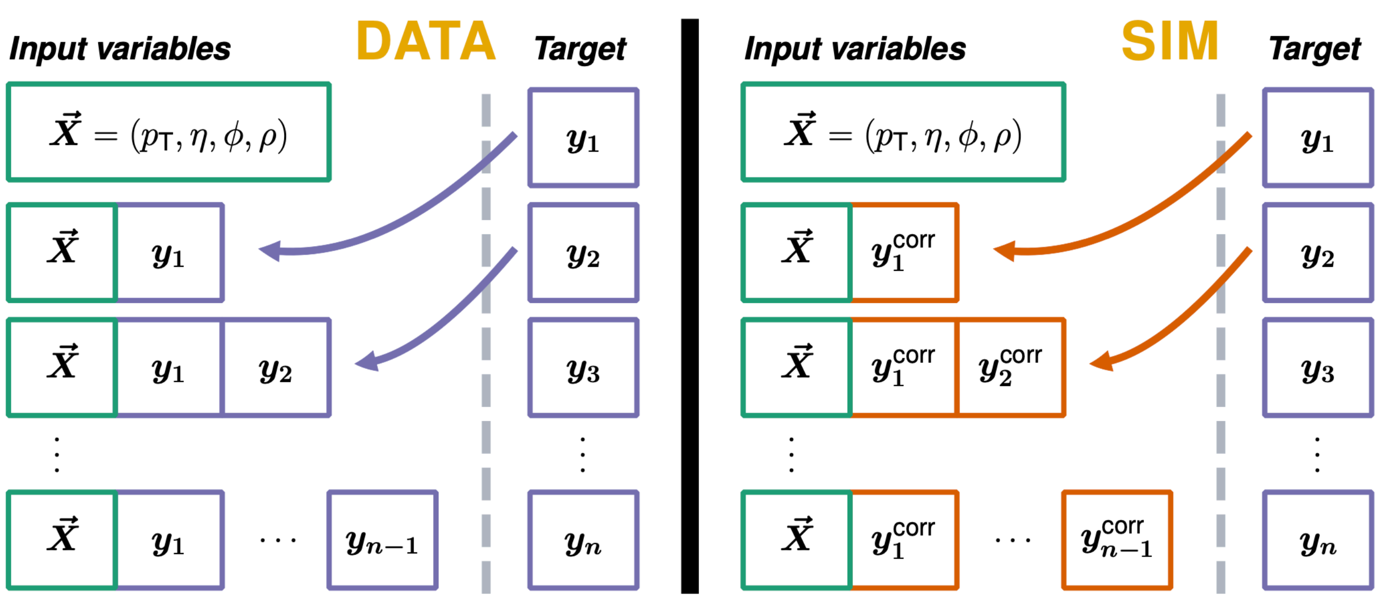
➡ **Using combination of different methods based on high momentum Higgs decay to improve precision and reduce uncertainties**

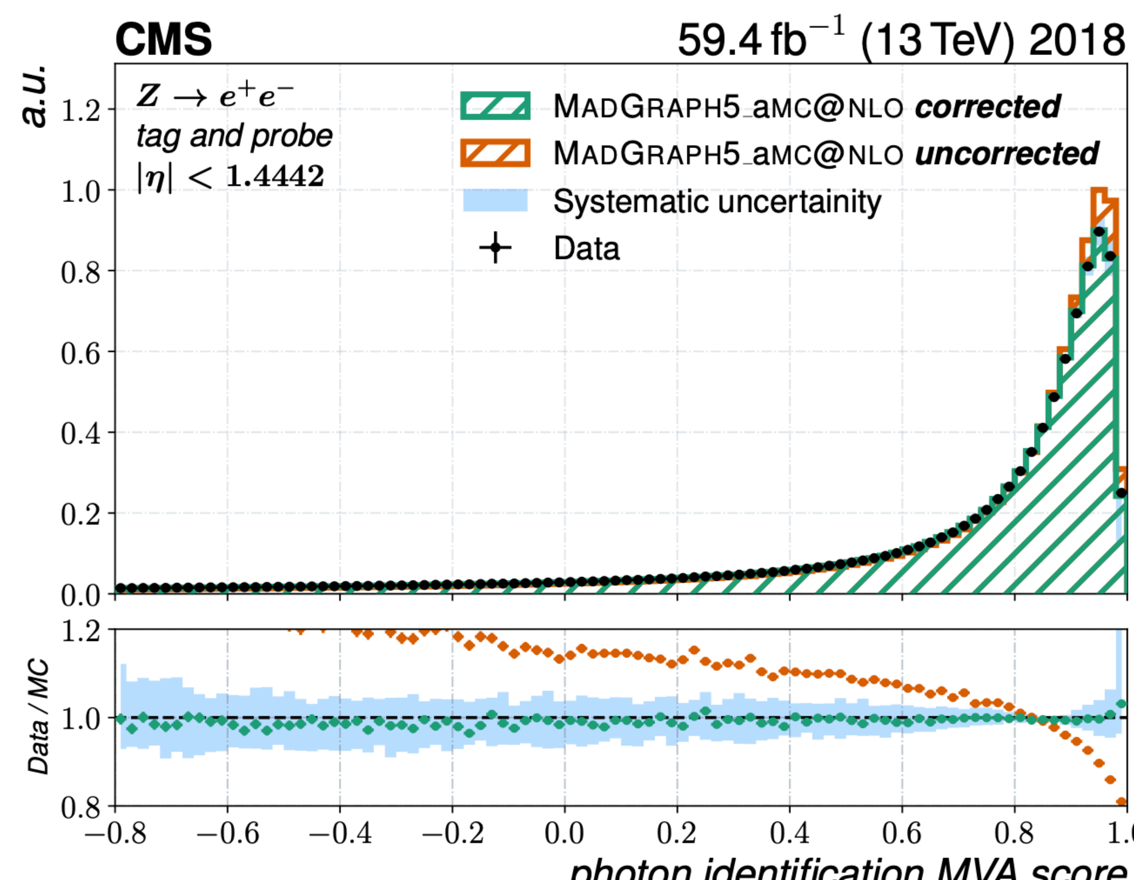




Mitigating systematics due to simulation mismodeling

- **High precision analyses (e.g. H→γγ) suffer from significant uncertainties from object performance**
  - photon ID mismatch in data and simulation account for 5% impact on Higgs cross-section results (aiming to have 1% agreement data/MC)
- **Derive corrections on simulation to electromagnetic shower shape and isolation variables**
  - using chained quantile morphing to map simulation to data
  - extracting shower shape CDF's in bins of photon kinematics in data and MC using BDT's
- **Improvement on data/MC modeling: reduced impact on modeling uncertainty on signal extraction (5%→2%)**





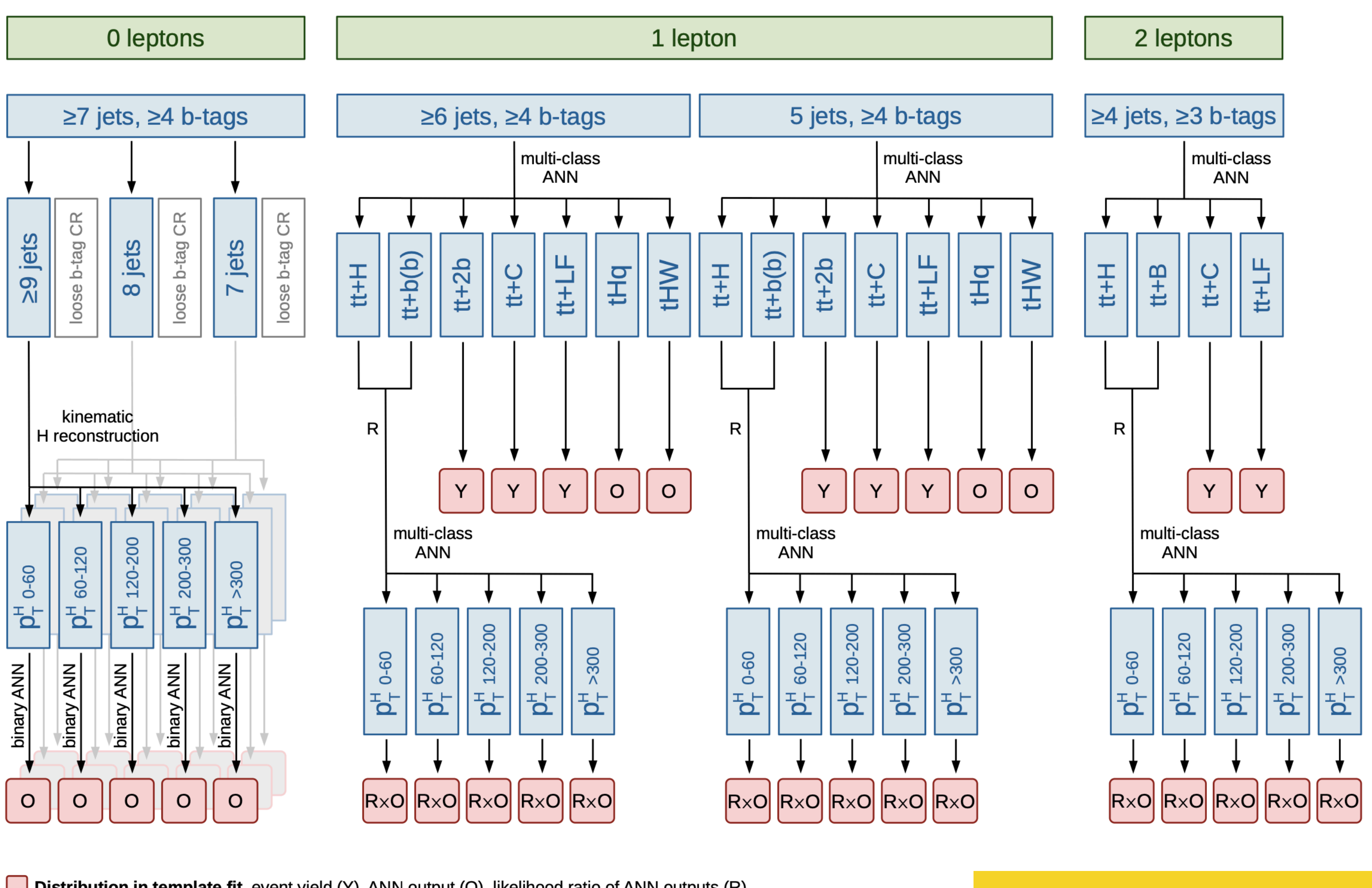
JHEP07(2023)091

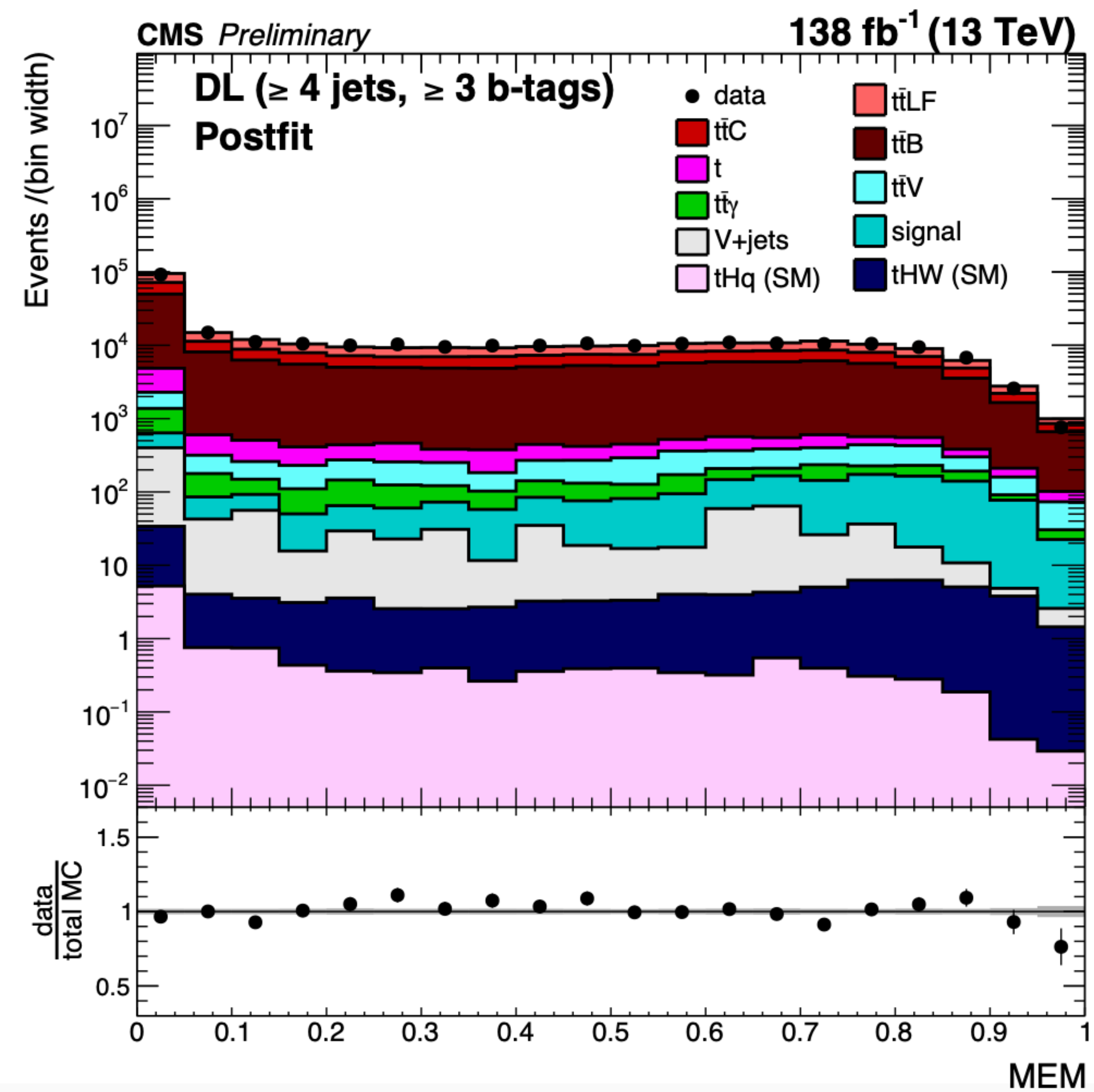
DNN techniques in the analysis strategy

➡ **ML is now backbone of most of Higgs analyses where sensitivity to physics-driven observables (invariant mass, transverse mass, ...) is not enough**

➡ **Using several ML techniques to fully exploit S/B separation and tackle optimal signal extraction and background prediction**

- targeting binary S/B classification in ultra-pure signal regions or multi-output classifier for background extractions in control regions to improve background constraints
- making use of low or high-level (e.g. MEM for ttHbb analysis) input features for enhanced separation - significant improvement over cut-based approaches



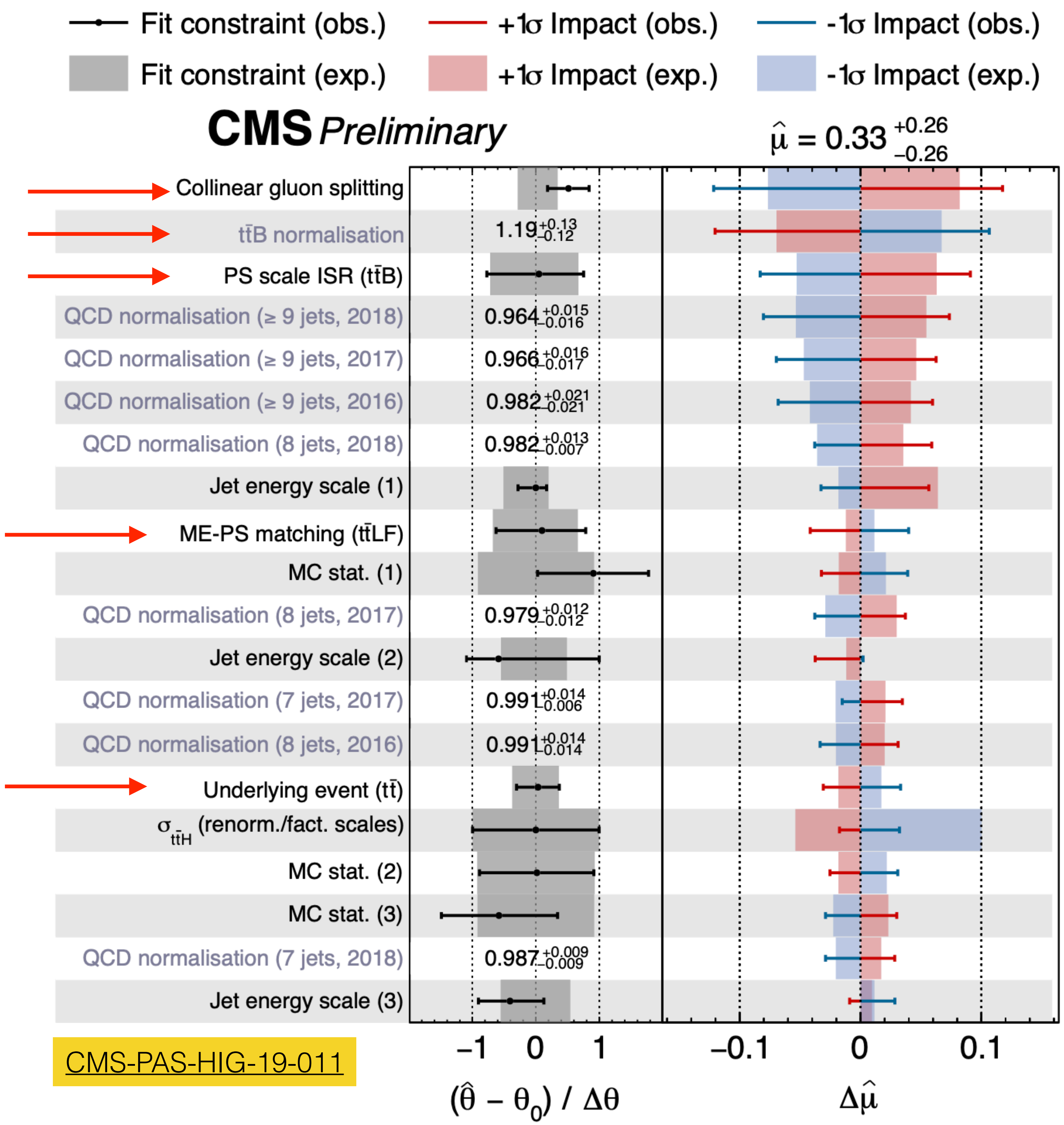


CMS-PAS-HIG-19-011

Improved background modelling: motivations

➡ **Several analyses (e.g. VHbb, ttHbb) have very small phase-space inclusive S/B because of large irreducible (V+jets, tt+bb) backgrounds**

- backgrounds estimated from simulation - plagued by sizeable modeling uncertainties and differences across generator predictions
- prefit modeling used for these backgrounds is very relevant: associated uncertainties are susceptible to constraints when profiled in the fit model
  - even more crucial in light of the usage of shape observables (DNN, BDT) for signal extractions



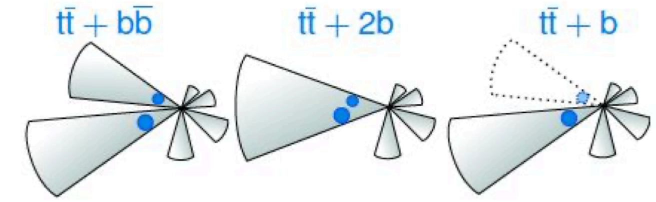
CMS-PAS-HIG-19-011

➡ **Dedicated focus on improving prefit modelling of dominant backgrounds and overall fit model - better model robustness and mitigation of modelling uncertainties**

Improved background modelling: the ttH(->bb) example

➡ **<u>Important to reduce background modeling uncertainties and fit constraints:</u> increased complexity of background model and improve prefit agreement with dedicated generators/tunings**

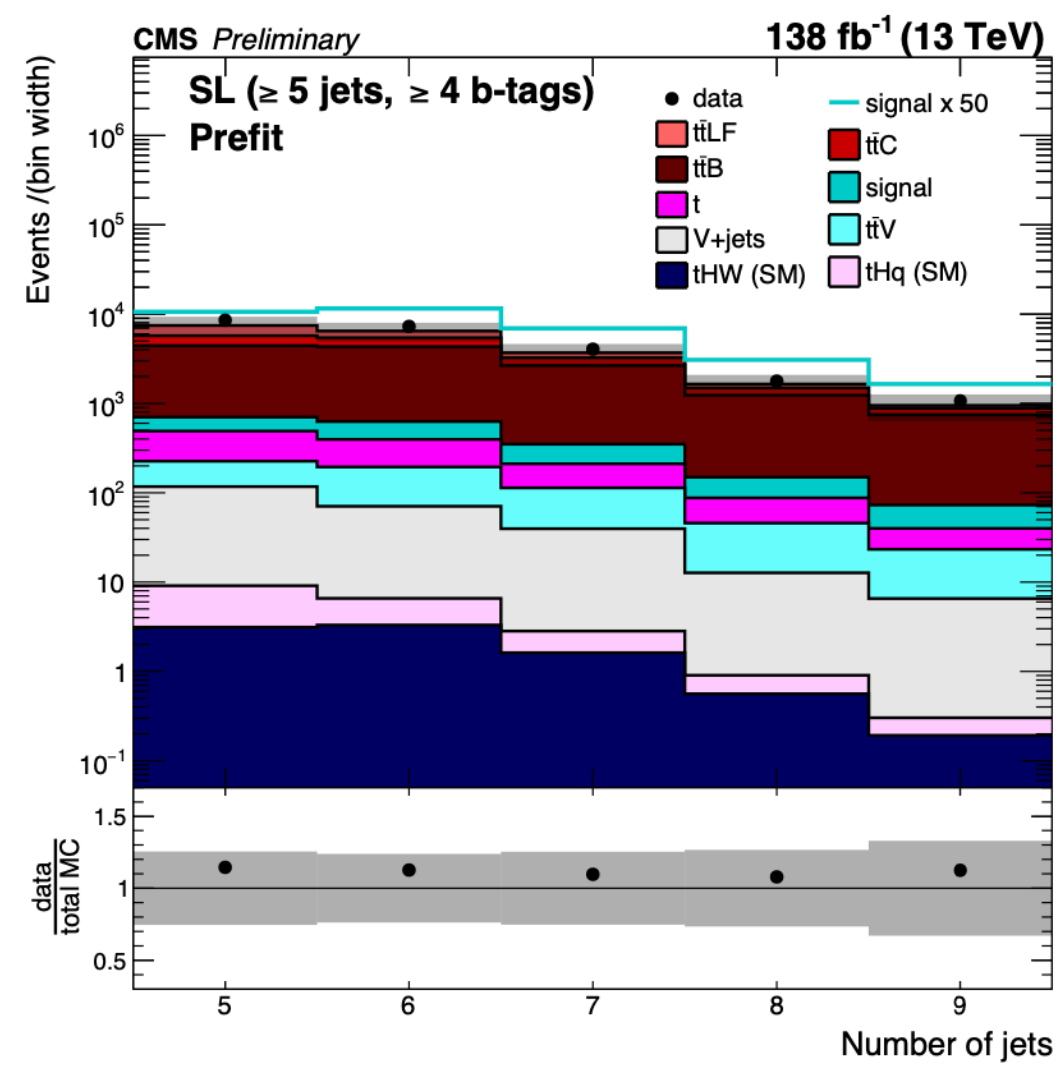
➡ **tt+bb background is largest contribution in leptonic channel**



- state-of-the-art MC simulation for ttB (>=1 additional b-jet) at NLO (4FS) w/ additional b-jets from matrix element instead of parton shower
- updated scheme for merging of ttB with ttC/ttLF contributions (taken from Powheg tt NLO 5FS)
- in-situ constraints on shape and normalisations of ttB via process scale factors

CMS-PAS-HIG-19-011

Uncertainty source	$\Delta\mu_{t\bar{t}H}$ (observed)	$\Delta\mu_{t\bar{t}H}$ (expected)
Total experimental	$+0.10/-0.10$	$+0.11/-0.10$
jet energy scale and resolution	$+0.08/-0.07$	$+0.09/-0.09$
b tagging	$+0.07/-0.06$	$+0.06/-0.02$
luminosity	$+0.02/-0.02$	$+0.01/-0.01$
Total theory	$+0.16/-0.16$	$+0.18/-0.14$
→ $t\bar{t}$ + jets background	$+0.15/-0.16$	$+0.12/-0.11$
signal modelling	$+0.06/-0.01$	$+0.13/-0.06$
Size of the simulated event samples	$+0.13/-0.12$	$+0.10/-0.10$
Total systematic	$+0.20/-0.21$	$+0.23/-0.19$
Statistical	$+0.17/-0.16$	$+0.17/-0.17$
→ background normalisation	$+0.13/-0.13$	$+0.13/-0.13$
→ $t\bar{t}$B and $t\bar{t}$C normalisation	$+0.12/-0.12$	$+0.12/-0.12$
QCD normalisation	$+0.01/-0.01$	$+0.01/-0.01$
Total	$+0.26/-0.26$	$+0.28/-0.25$

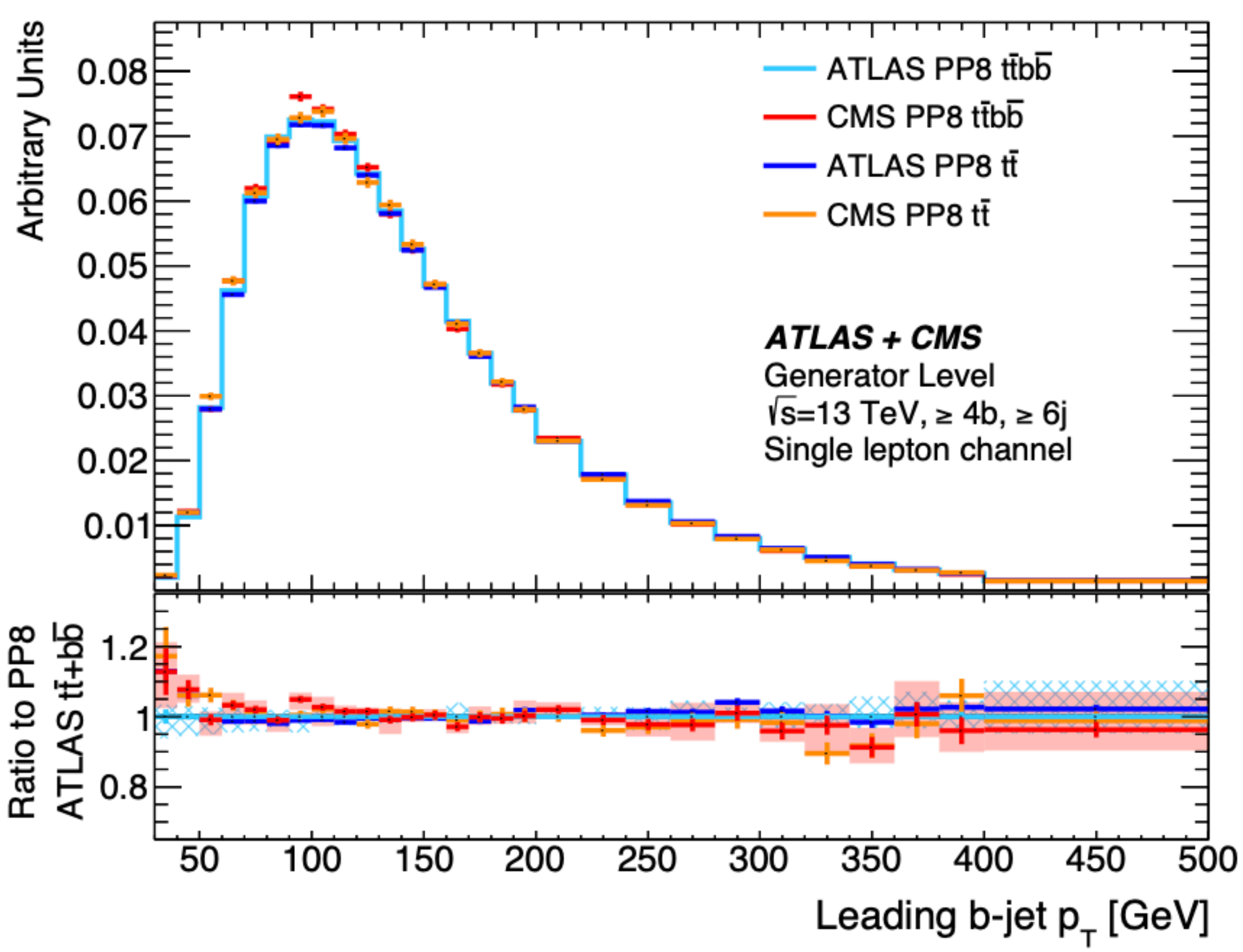


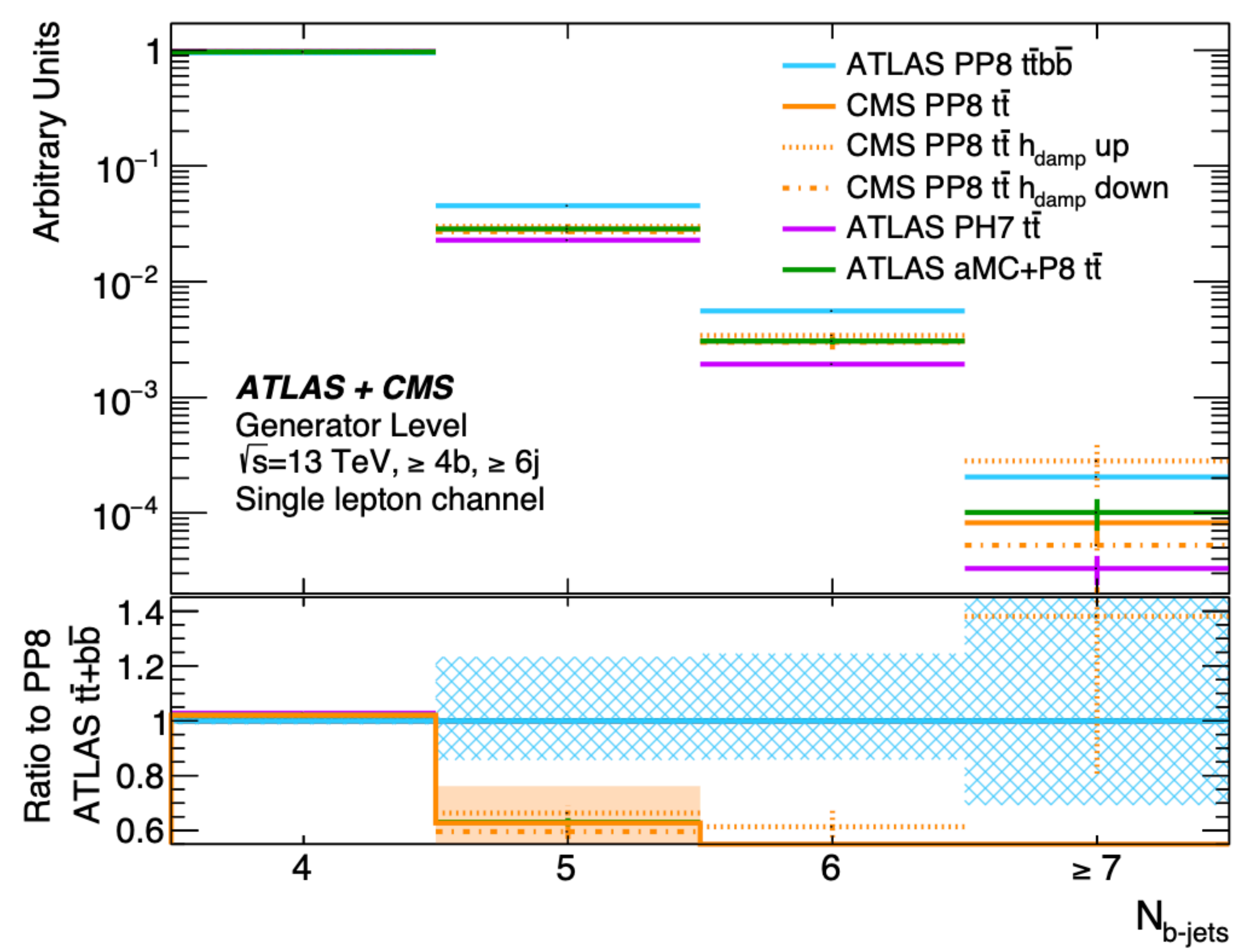
Improved background modelling: next steps

➡ Attempt towards harmonisation of modeling uncertainties treatment for main V+jets and tt+bb background carried out in LHC Higgs WG

- investigating central value differences across generators for tt and tt+bb processes as well as treatment of uncertainties across experiments
  - different assessment of scale uncertainties in CMS/ATLAS (hdamp vs hbzd)
  - variations across generators always exceeds internal scale uncertainties of single MC

LHCHWG-22-003

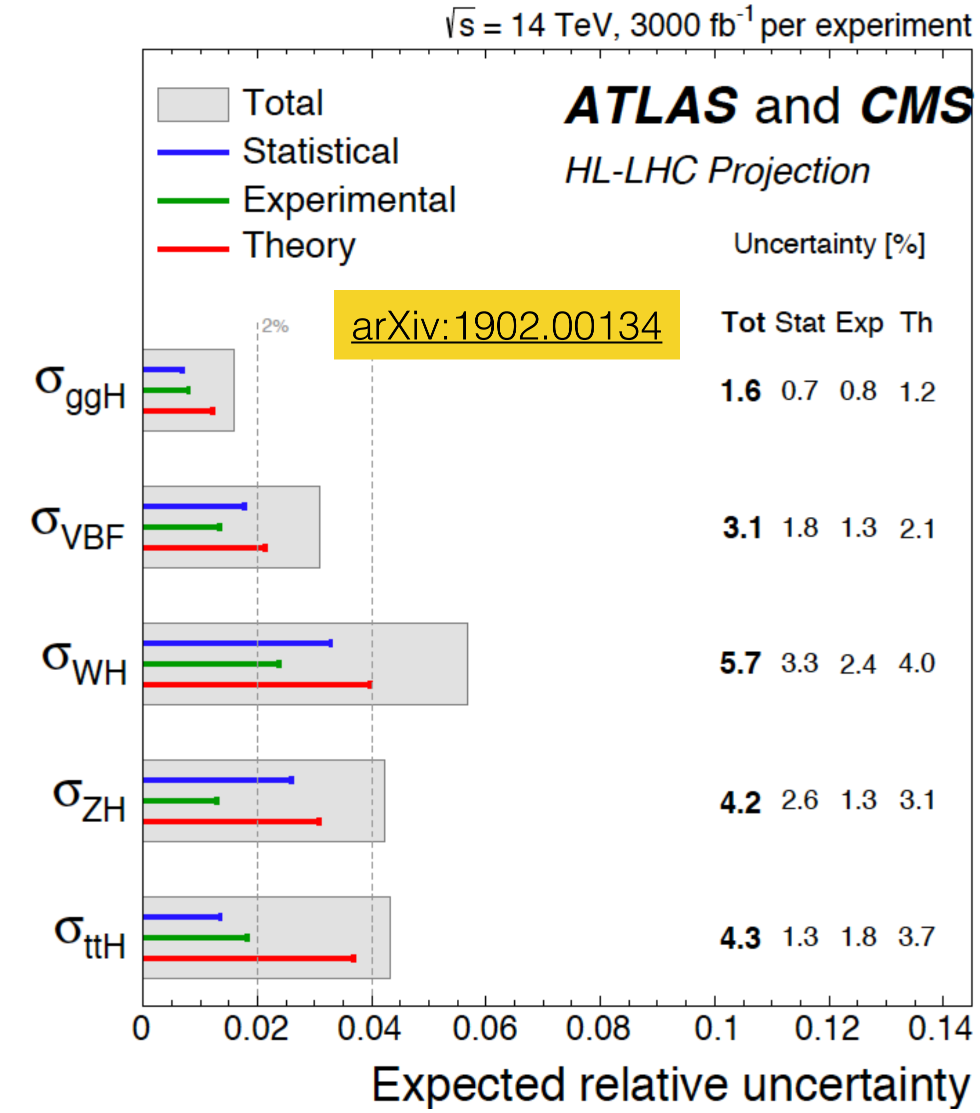


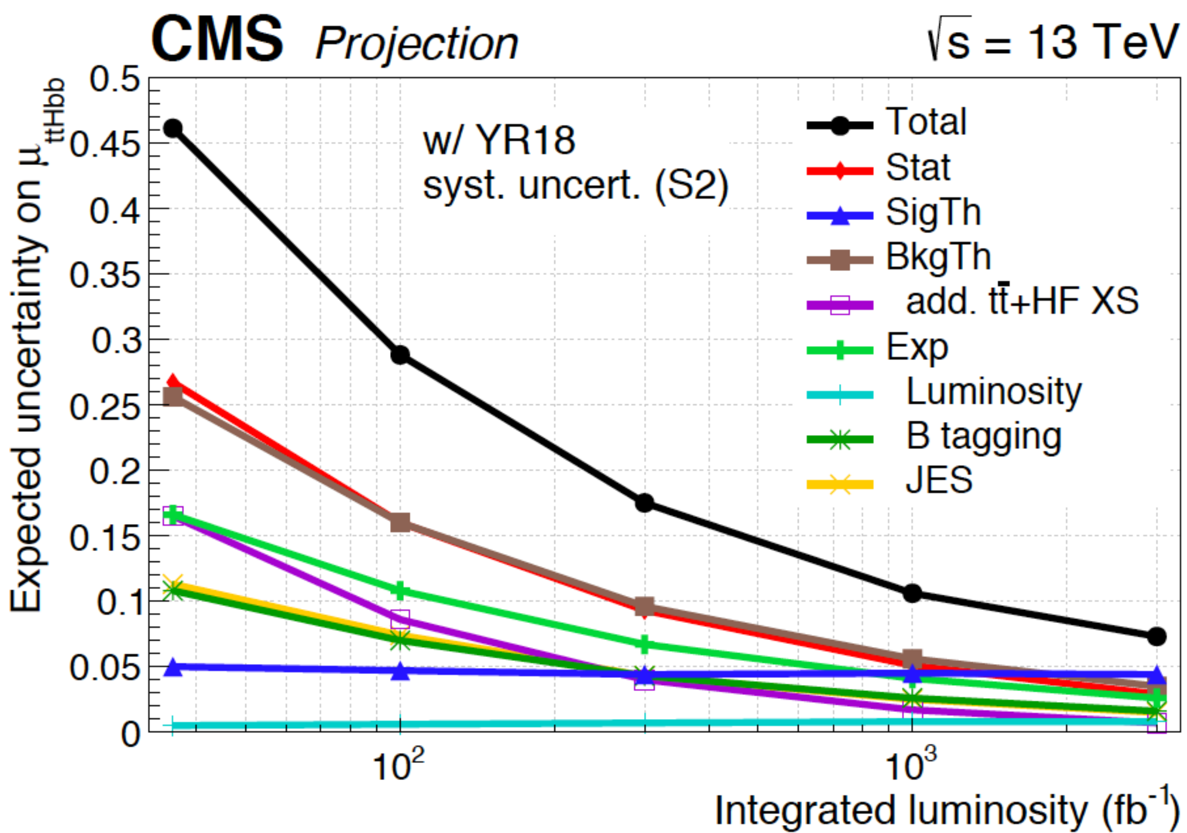


➡ Powheg+Pythia8 setup observed to be consistent across ATLAS and CMS

Wrapping-up and conclusions

- **Novel techniques in CMS Higgs analyses targeting various aspects of measurements fully exploiting physics potential of LHC data**
  - object identification/reconstruction using ML
  - improved ML techniques for signal extraction and background modeling
  - novel analysis interpretations (differential, STXS, SM deviations anomalous couplings and EFT)
  - novel strategies targeting HH and HHH at HLT in Run 3 exploring Data Parking [link]
- **Working on further improving analysis techniques towards High-Luminosity LHC**
  - hitting systematics wall for many Higgs measurements
  - mitigate simulation modeling/theory





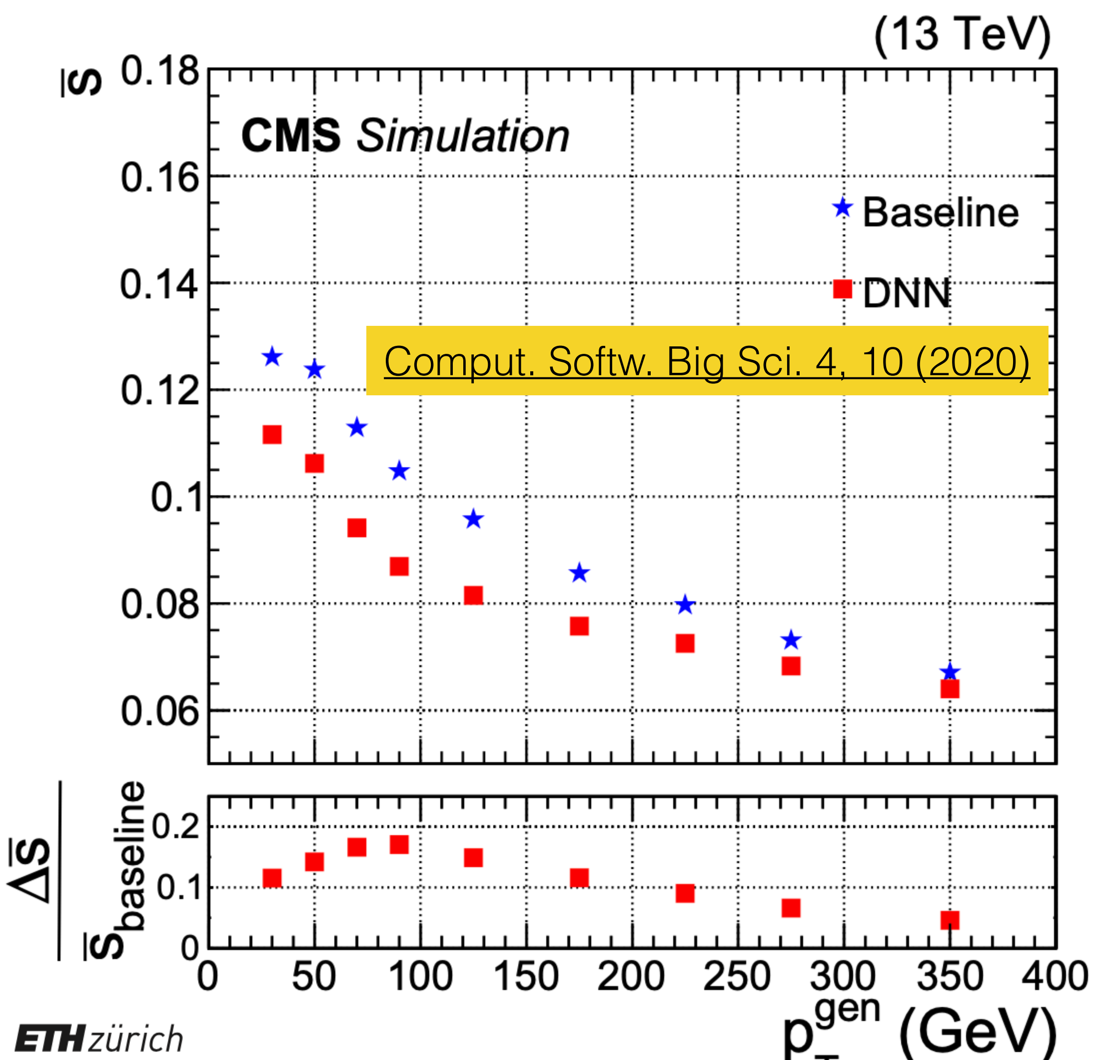
Backup slides

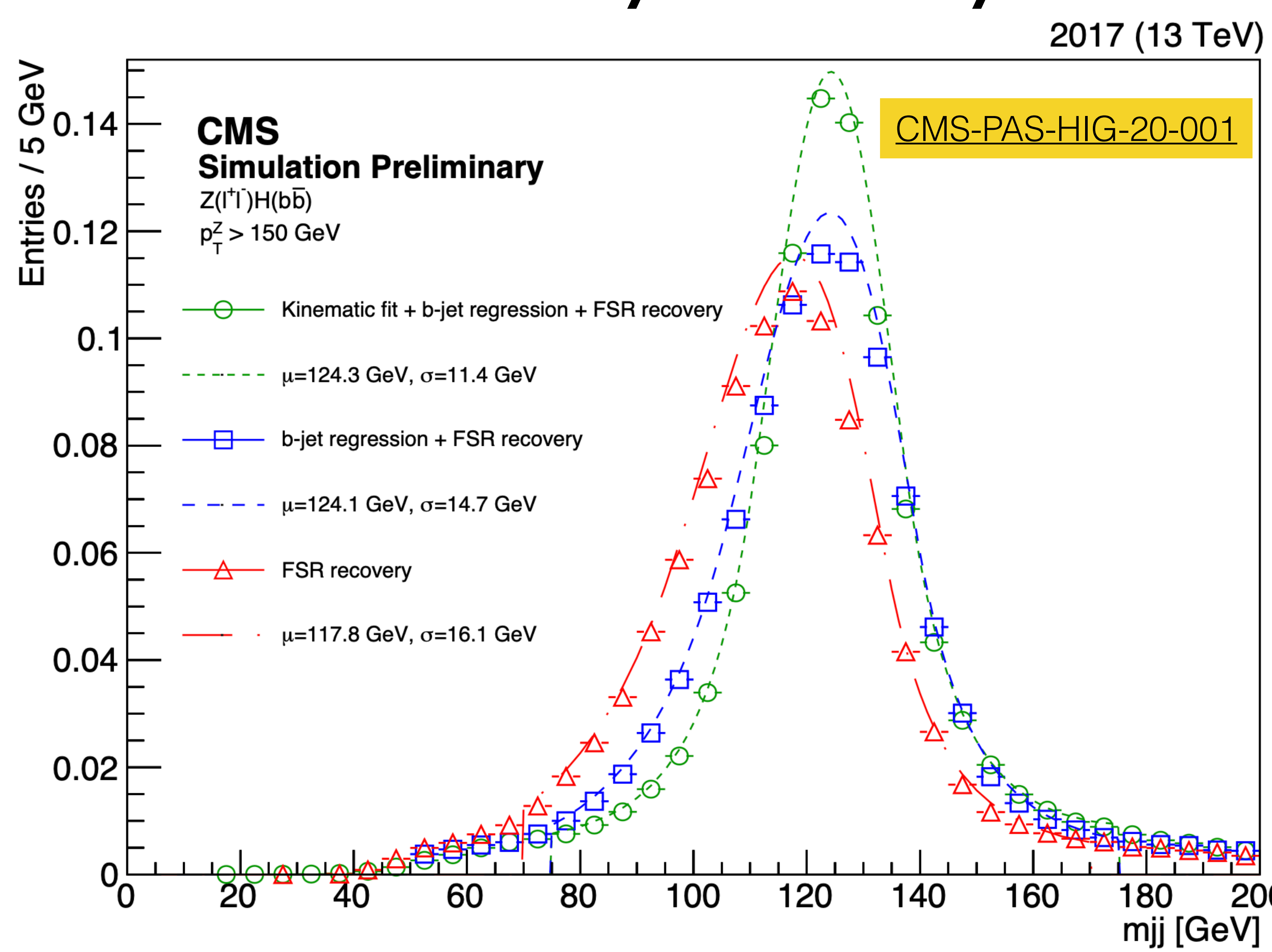
Improving b-jet energy determination

➡ **Crucial to optimally use highest resolution events to measurement precision - excellent estimation of per-object resolution**

- especially relevant for b-jets/c-jets since as final states for several Higgs analyses (VHbb, VHcc): improved energy regression using DNN-based (feed-forward NN) energy regression techniques
- recovers energy loss from escaping neutrinos targeting energy corrections and energy quantiles
- training uses b-jet/c-jet specific kinematic properties to improve reconstruction of jet momentum

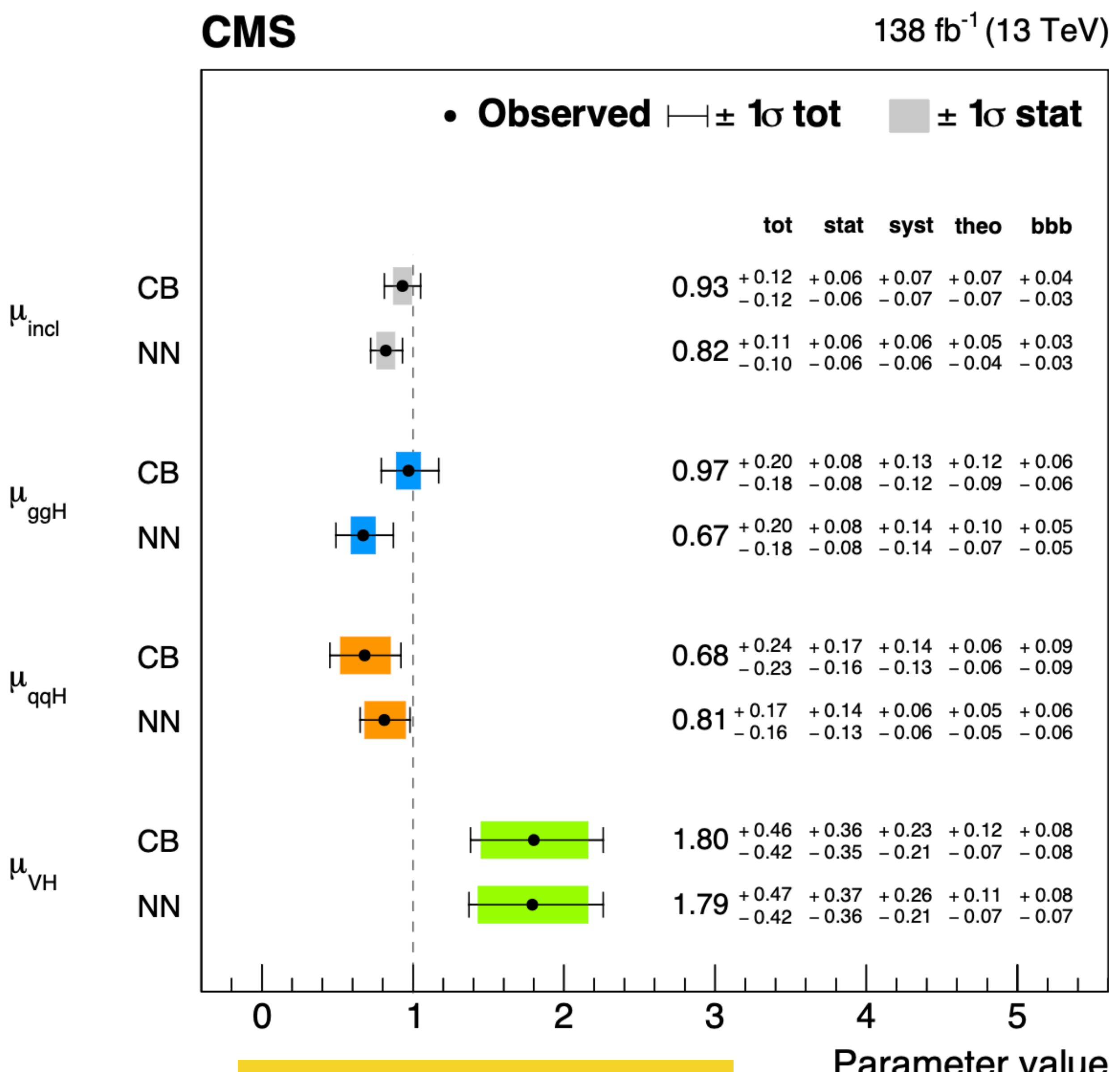
➡ **Successfully employed b/c-jet energy regression in VH(bb)/VH(cc): significant improvement on jet energy scale/resolution better analysis sensitivity**





DNN techniques in the analysis strategy

- ➡ **ML is now backbone of most of Higgs analyses where sensitivity to physics-driven observables (invariant mass, transverse mass, …) is not enough**
- ➡ **Using several ML techniques to fully exploit S/B separation and tackle optimal signal extraction and background prediction**
  - making use of low or high-level (e.g. MEM for ttHbb analysis) input features for enhanced separation - significant improvement over cut-based approaches



Eur. Phys. J. C (2023) 83:562