Machine learning opportunities and applications in BNL ATLAS(Higgs) analyses

Gaetano Αθανάσιος Barone

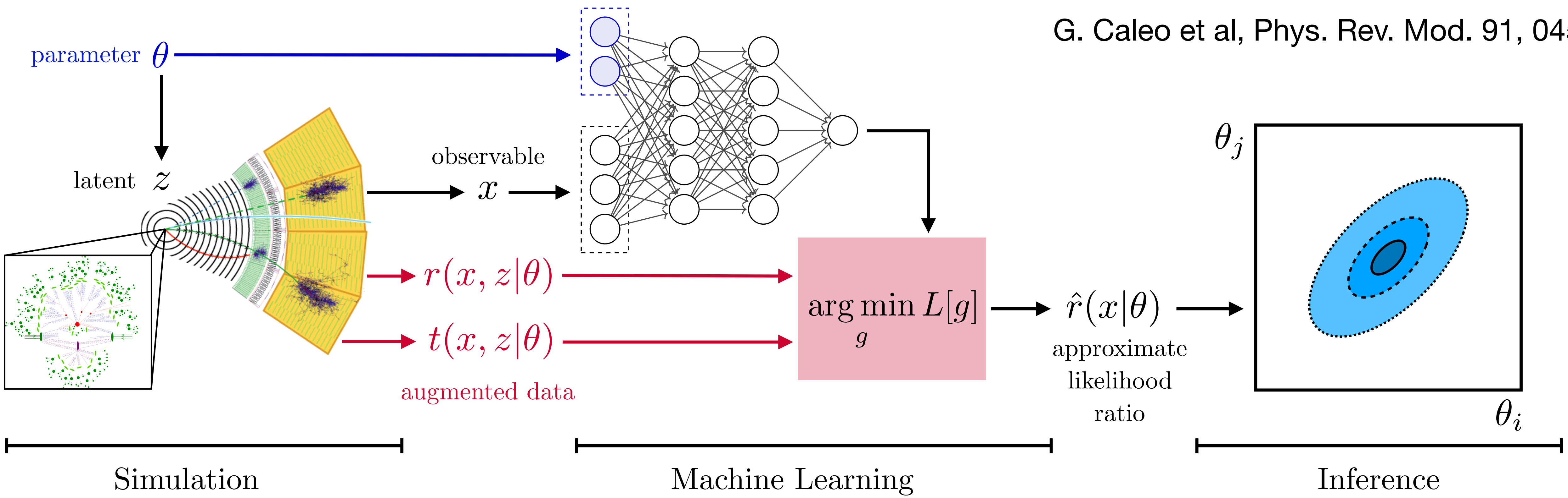
Brookhaven National Laboratory

April 2021

Introduction

- Flourishing of Complex Systems research in the recent years.
  - Machine-learning (ML) tools available for application.
  - Driving ML research from both research in Complex Systems and application needs.

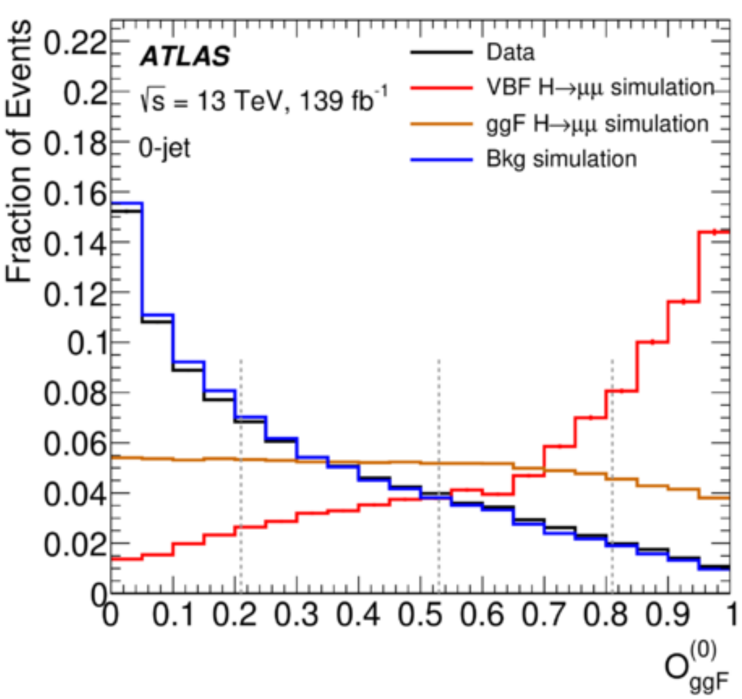
G. Caleo et al, Phys. Rev. Mod. 91, 045002

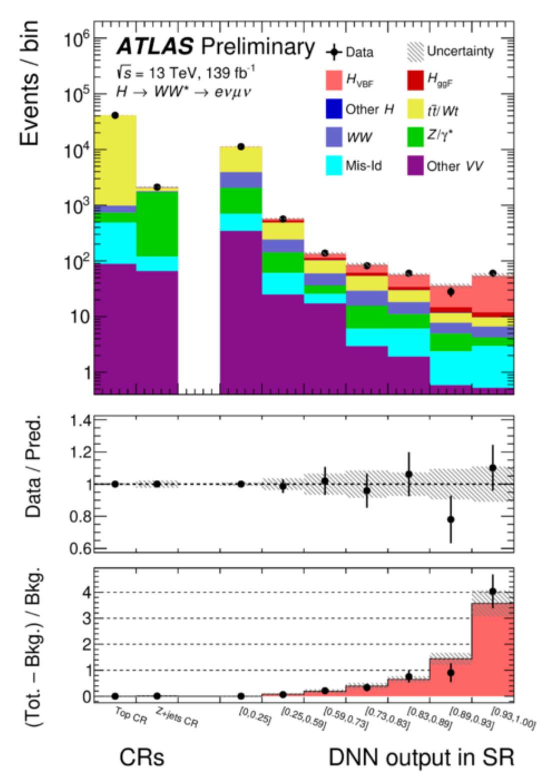


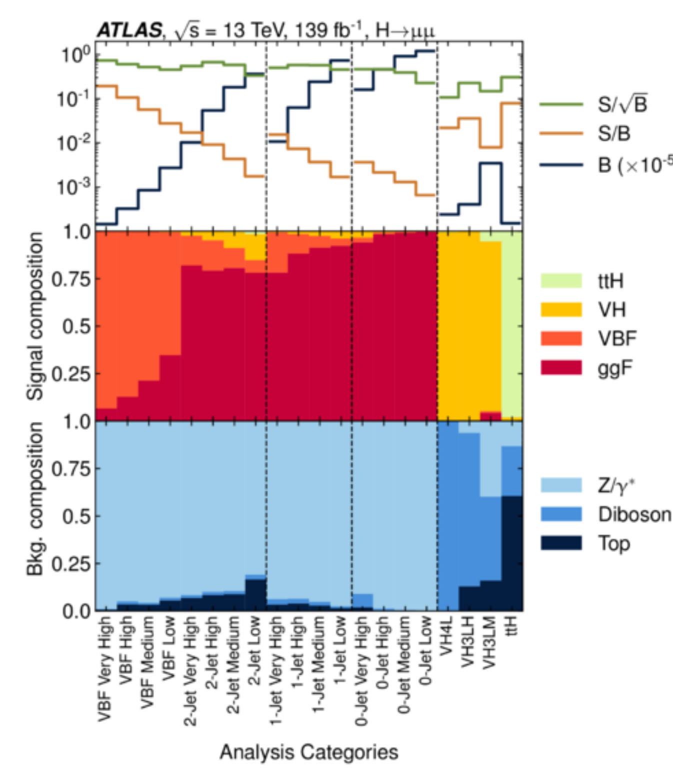
- Hadron collider detectors involve huge number of variables who's *information* needs maximisation.
  - In many analyses complexity of degrees of freedom approaches that of complex systems.
  - ML is seen as a tool to maximise the *information* out of these degrees of freedom thus improving the capability of our detector.

Introduction

- Machine Learning (ML) used widely in the analyses within the Omega Group:
  - Snowmass studies:
    - VBS $V_L V_L$ production,
    - Polarisation discrimination in offshell regime of *ZZ* and *WW.*
  - Higgs physics:
    - $HH \rightarrow \gamma\gamma bb$,
    - $H \rightarrow \mu\mu$,
    - Higgs tagging in $H \rightarrow bb$,
    - $H \rightarrow$ invisible.
    - $H \rightarrow ZZ^*$: couplings, offshell cross section measurement, mass measurement,
    - $H \rightarrow WW^*$: couplings and differential cross section measurements,
  - Searches:
    - $H \rightarrow ZZ_d$,
    - High mass *ZZ* search,
    - Low mass dijet resonances



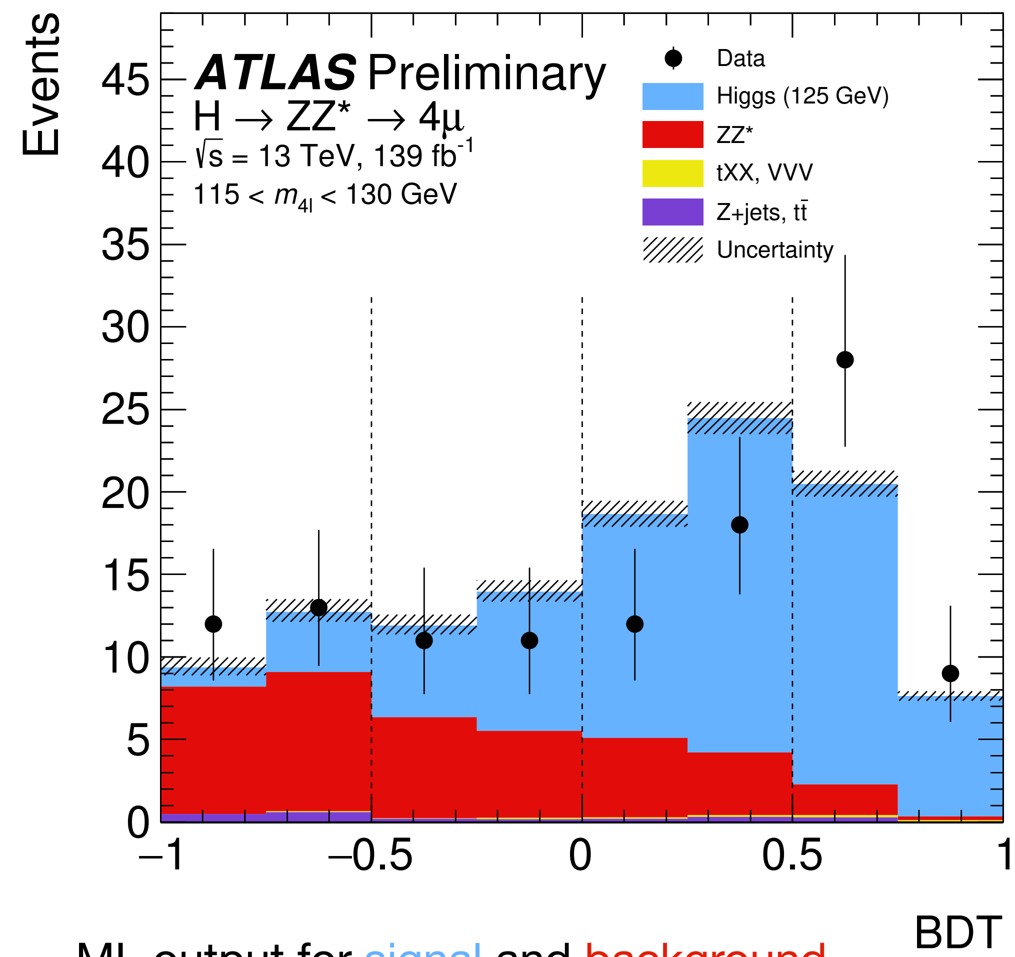




Not an extensive list

Types of applications

- Mostly driven by discrete classification problems:
  - Typically, ML used in targeting more effective event selections.
  - Event classification according to d.o.f. of many input features.
    - ✦ Allows for enhancing *s*/*b*, thus improving on sensitivity.
    - ✦ Applications from Boosted Decision Trees to Deep Neural Networks.



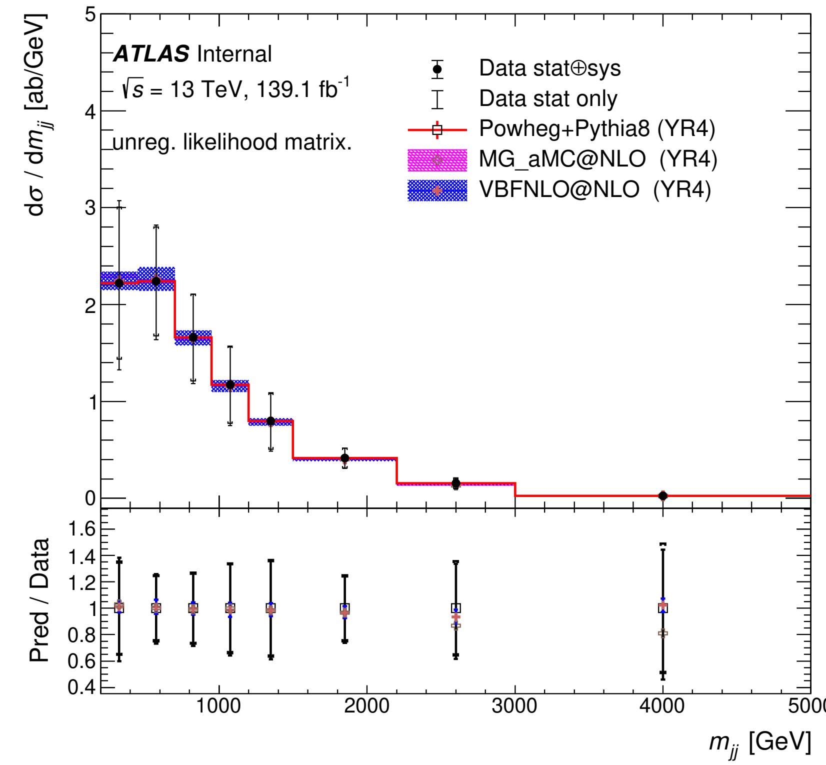
ML output for signal and background

- Continuous and functional classification being evermore employed in our works.
  - Precise prediction of real-valued physics quantities depending on several detector characteristics at a time.

- Will detail in the following the two different aspects, highlighting their different analysis requirements:
  1. Differential and fiducial cross section in $H \to WW^*$
  2. $H \to ZZ^* \to 4\ell$ mass measurement.

Differential cross section

- Goal to measure VBF $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ fiducial and differential cross section.
  - Main backgrounds arising from multiple sources:
    - $t\bar{t}+Wt$, WW, Z+jets, and misidentified leptons.
  - Goal is model-independence: as little as possible dependance on a priori knowledge of the signal.

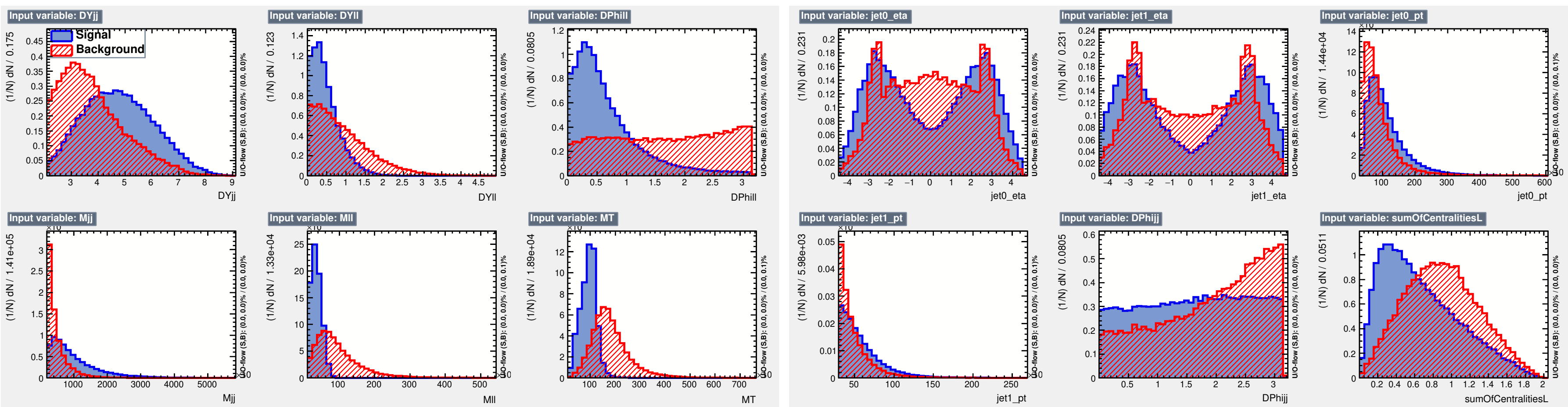


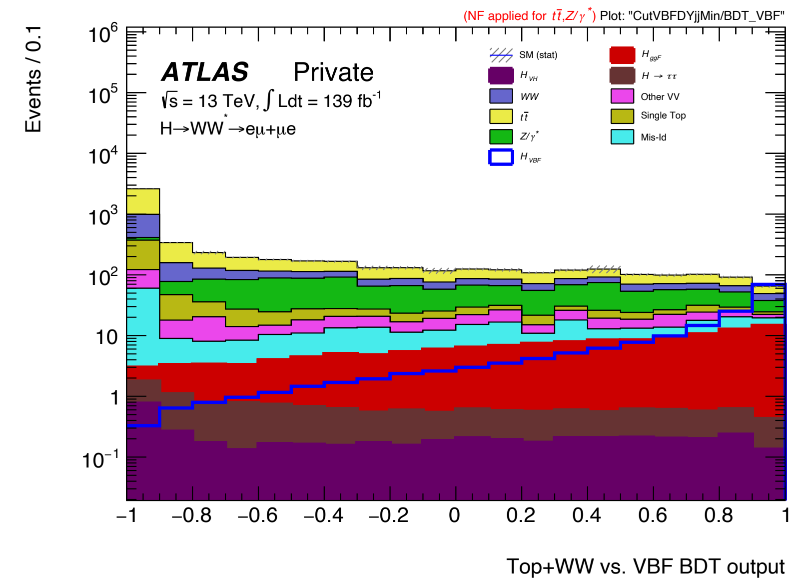
- Precision is challenged by small s/b of ~0.03.
- Precise discrete classification, typically, solves this.
  - At a cost of increased model dependance.
  - Measured value depends on simulated finite training set.

- Structure M.L. to maximize information while relying on minimally-biasing sets:
  - (i) Multidimensional fit on multidimensional M.L. process classification.
  - (ii) Boosted Decision Tree for exploiting only linear (smooth) correlations between input variables.
  - (iii) Careful choice of variables the value of which minimises the bias of the measured quantity.

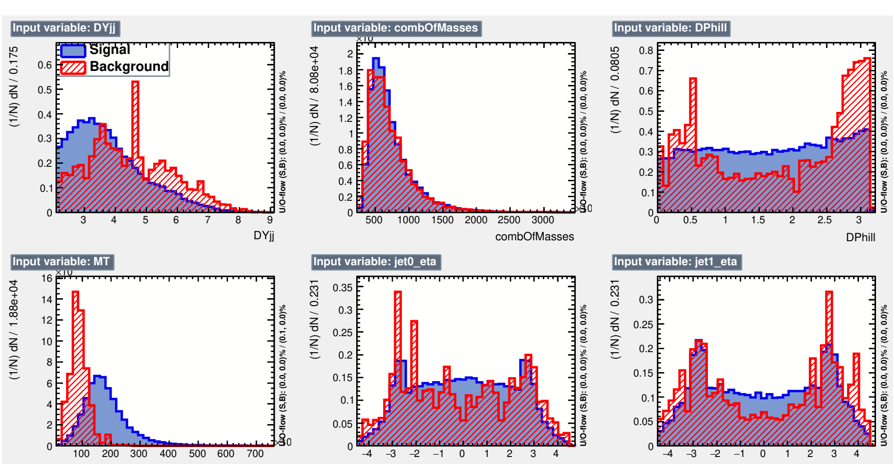
Signal region discriminant

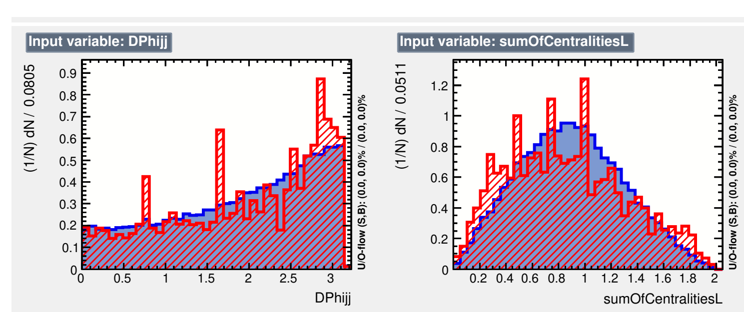
- Two sets of independently-trained BDTs:
  - Minimum-biasing n-trainings in differential fits, by excluding variables highly correlated with the to be unfolded observable.
  - D_{VBF} : train vbf signal against $t\bar{t}+WW$ contributions.

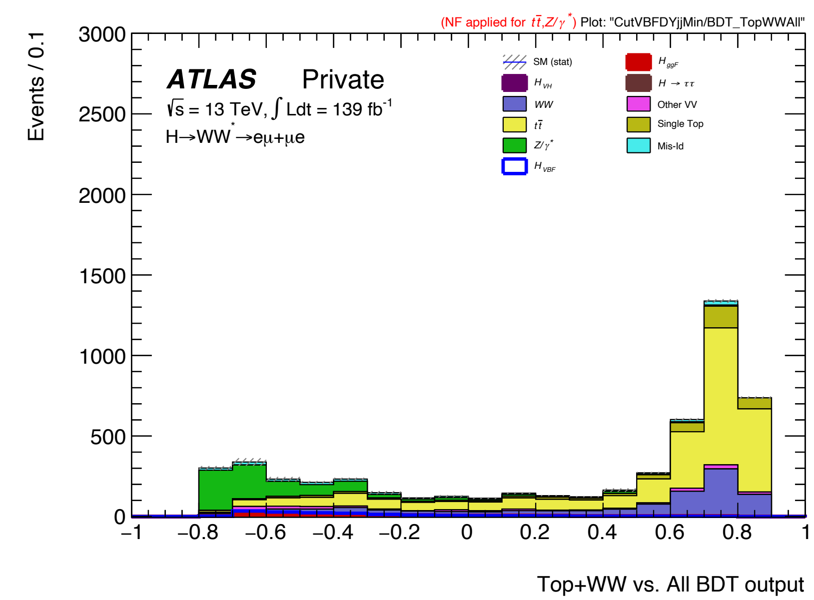




  - D_{TopWW} : train $t\bar{t}+WW$ against all other remaining backgrounds.



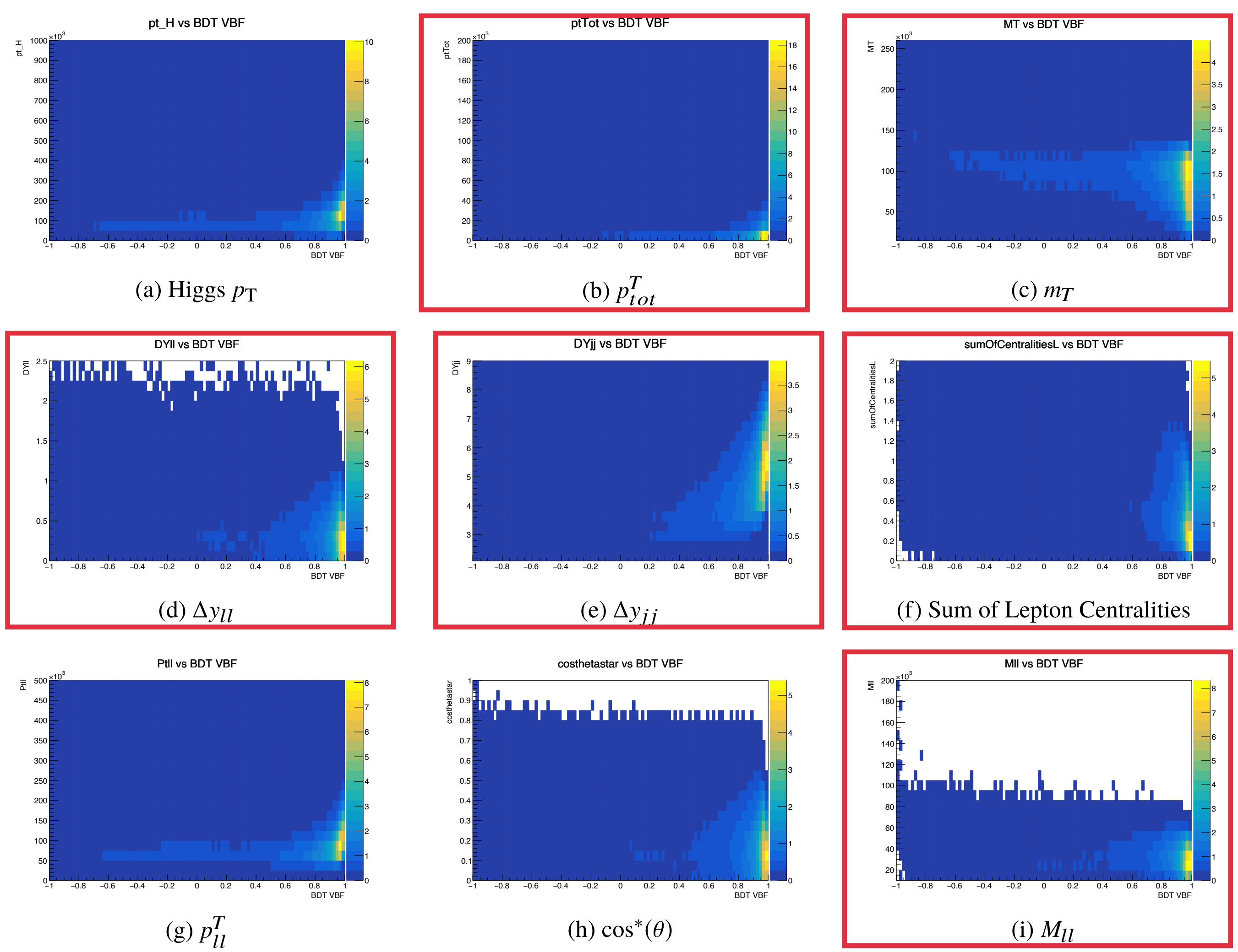




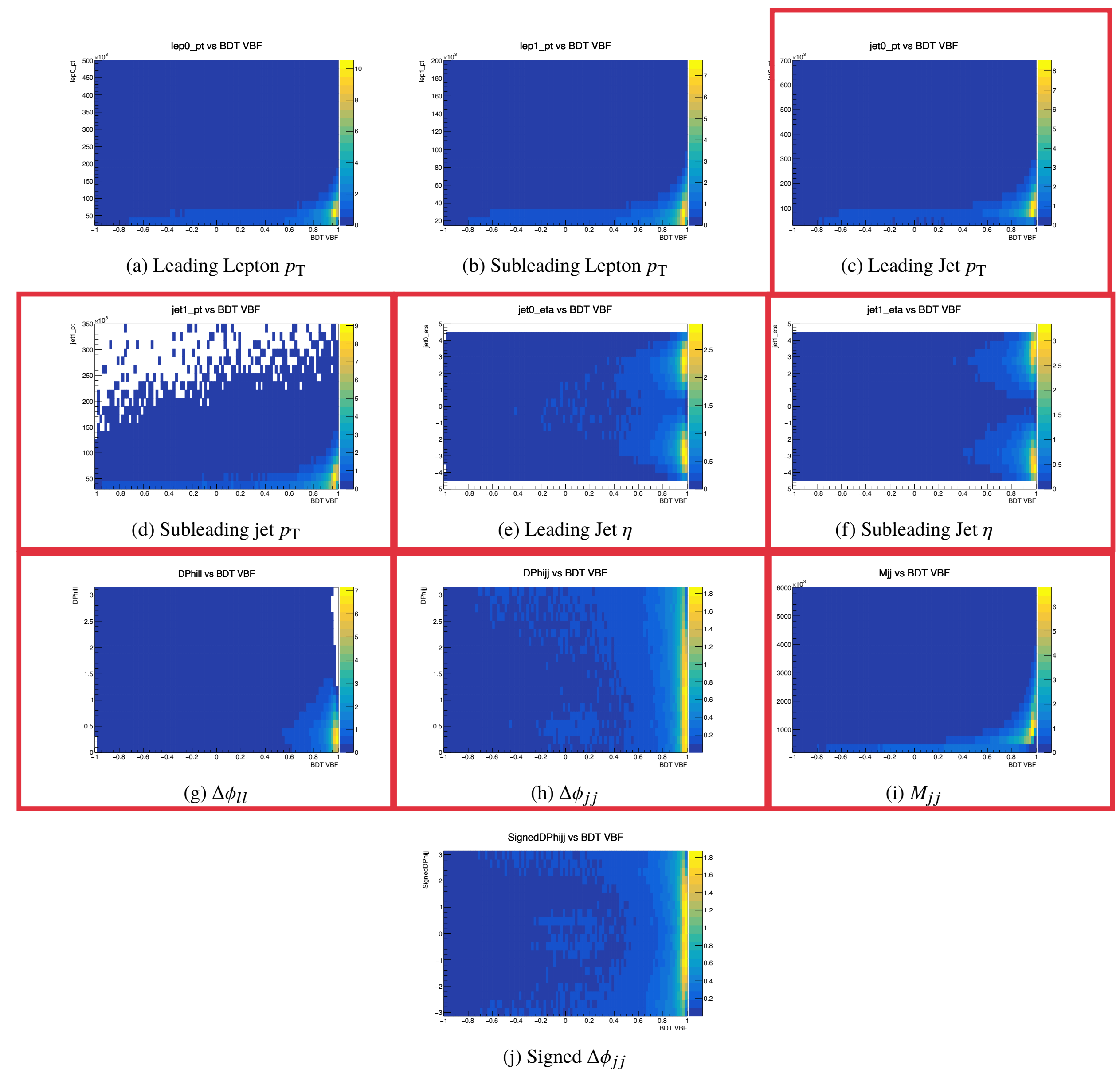
Discriminant bias

- BDT may induce model dependency.
  - Keep variables characterising the background, by dropping variables that sculpt significantly the signal.

- Further discriminant bias reduction:
  - Smart-binning: most of the D_{VBF} dependency concentrated > 0.9.



(a) Higgs p_T (b) p_{tot}^T (c) m_T
(d) Δy_{ll} (e) Δy_{jj} (f) Sum of Lepton Centralities
(g) p_{ll}^T (h) $\cos^*(\theta)$ (i) M_{ll}



(a) Leading Lepton p_T (b) Subleading Lepton p_T (c) Leading Jet p_T
(d) Subleading jet p_T (e) Leading Jet η (f) Subleading Jet η
(g) $\Delta\phi_{ll}$ (h) $\Delta\phi_{jj}$ (i) M_{jj}
(j) Signed $\Delta\phi_{jj}$

Variables used in D_{VBF}

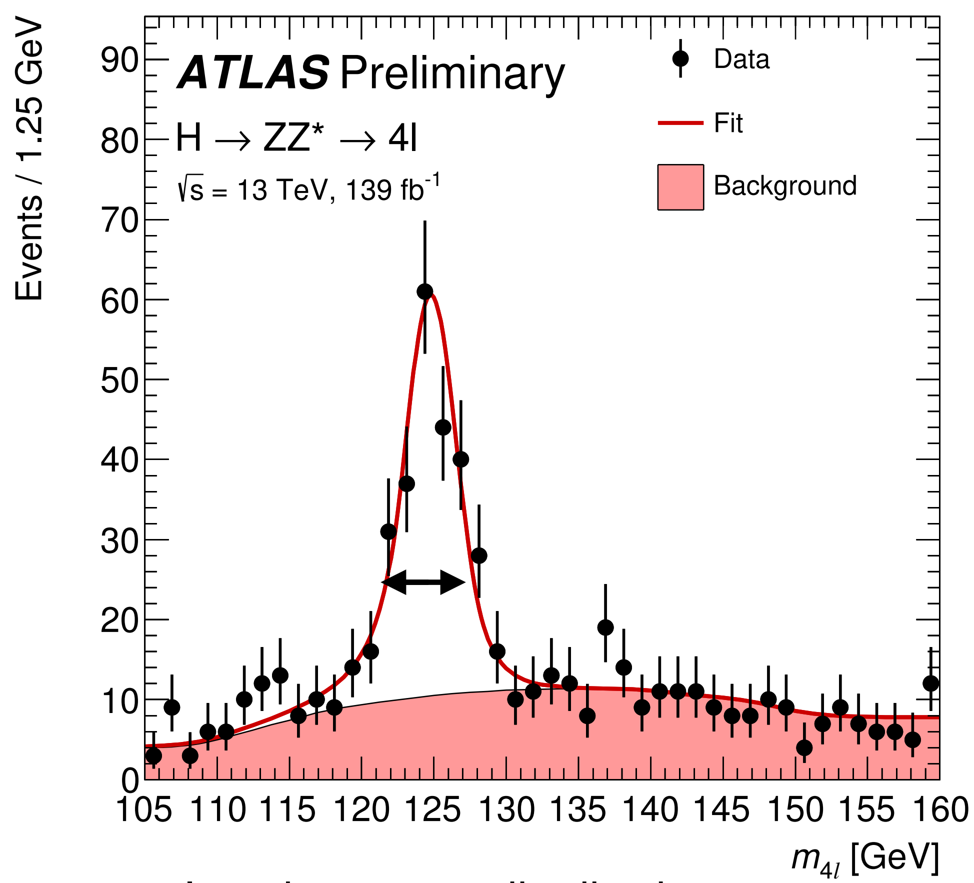
Ingredients for precision

- In the $H \to ZZ \to 4\ell$ the signal is a narrow resonant peak above a background continuum.

ℓ = electron or muon

Higgs signal, resonant at the value of m_H

Non resonant background from other non H production



Invariant mass distribution

$$\delta m_H \simeq \frac{\sigma(m_{4\ell,\gamma\gamma})}{\sqrt{N - N_\mathrm{b}}}$$

Uncertainty on m_H approximated by the uncertainty on the mean of the mass distribution

(I) Statistical precision precision depends upon:

- resolution of the reconstructed final state and number of signal events.

(II) Systematic uncertainty from understanding of detector performance:

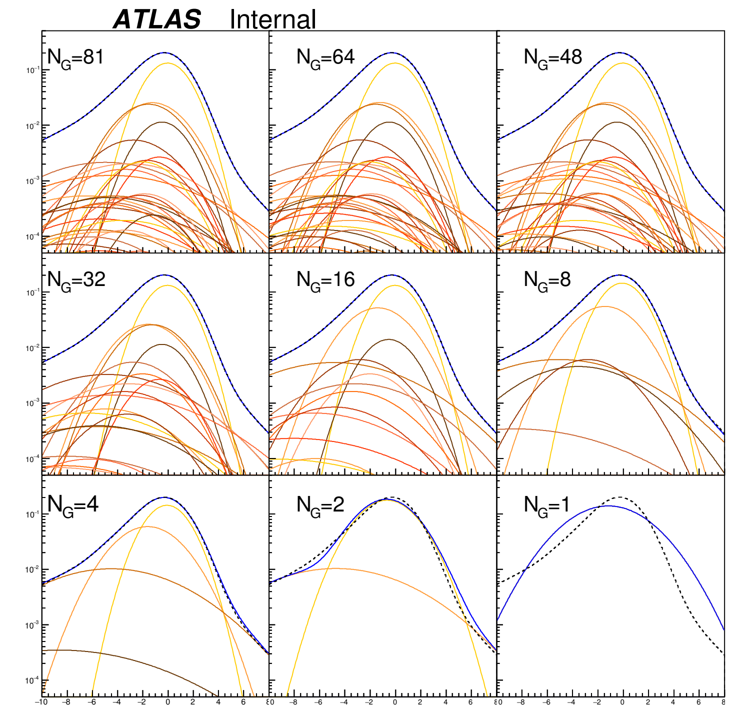
- Multi-prong approach to reduce uncertainty at analysis level:

(i) ~15% from constraint of mass of two leading leptons, which form a well known resonance.

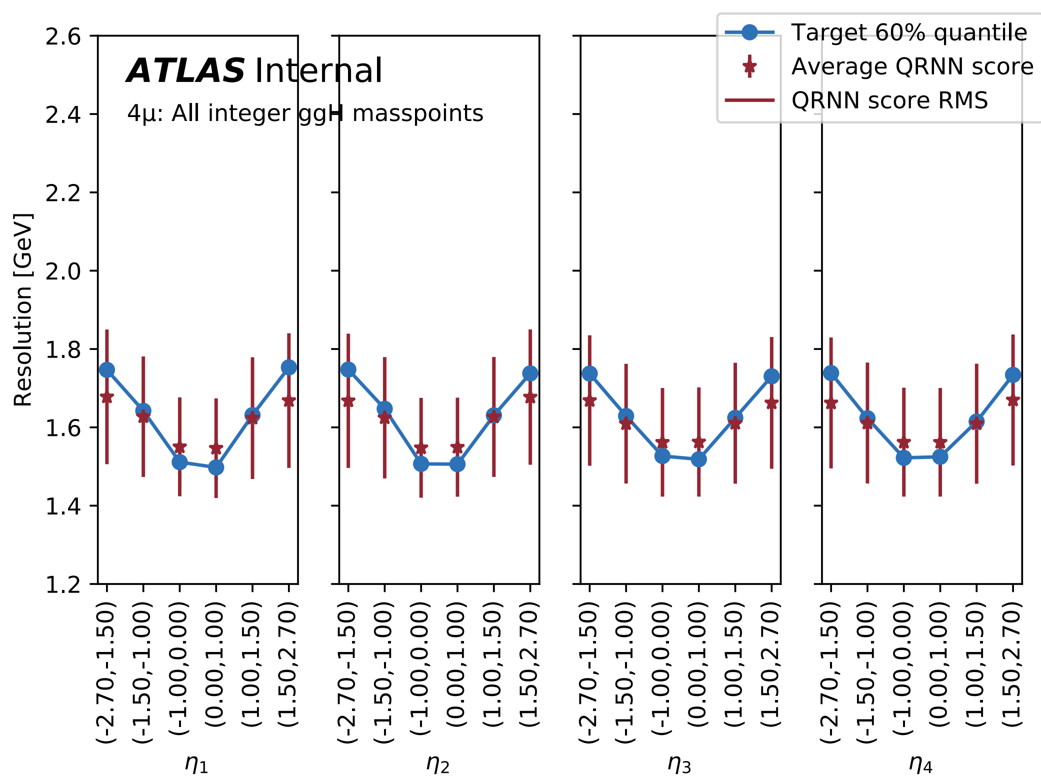
(ii) ~2% from kinematic discriminant between signal and background events.

(iii) Precise knowledge of the detector's resolution curial for ultimate precision

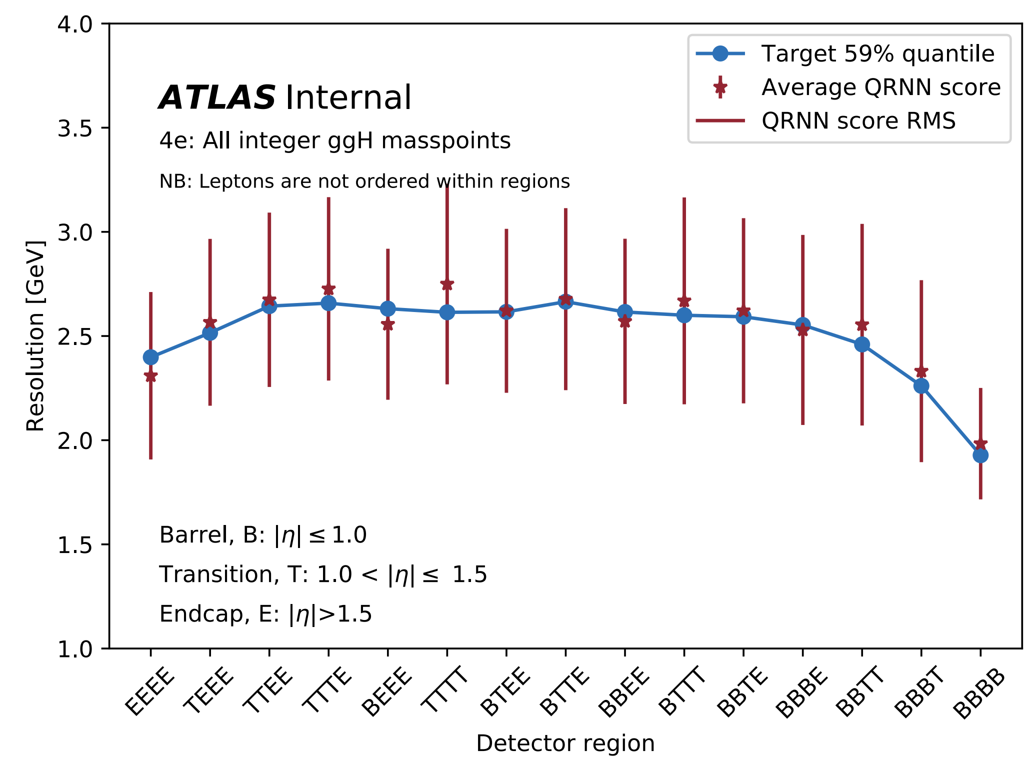
Improvements with ML



- In a per-event error method detector's resolution response crucial:
  - events with smaller uncertainty than others;
  - hidden correlations in detector quantities;
    - ✦ track fitting precision: magnetic filed, detector inefficiencies, material upstream of calorimeters, interlayer movements,…
- Constraint to m_Z pair induces non trivial energy correlation:
  - Analytical model would correspond to ~243 d.o.f per event.
    - Gaussian Sum Filter reduces d.o.f, at a cost of information loss.
- Parametric ML event classification as solution for targeting each event's uncertainty:
  - (i) Identifying patterns of variable track quality in the detector.
  - (ii) Solve complex energy correlation problem from m_Z constraint.



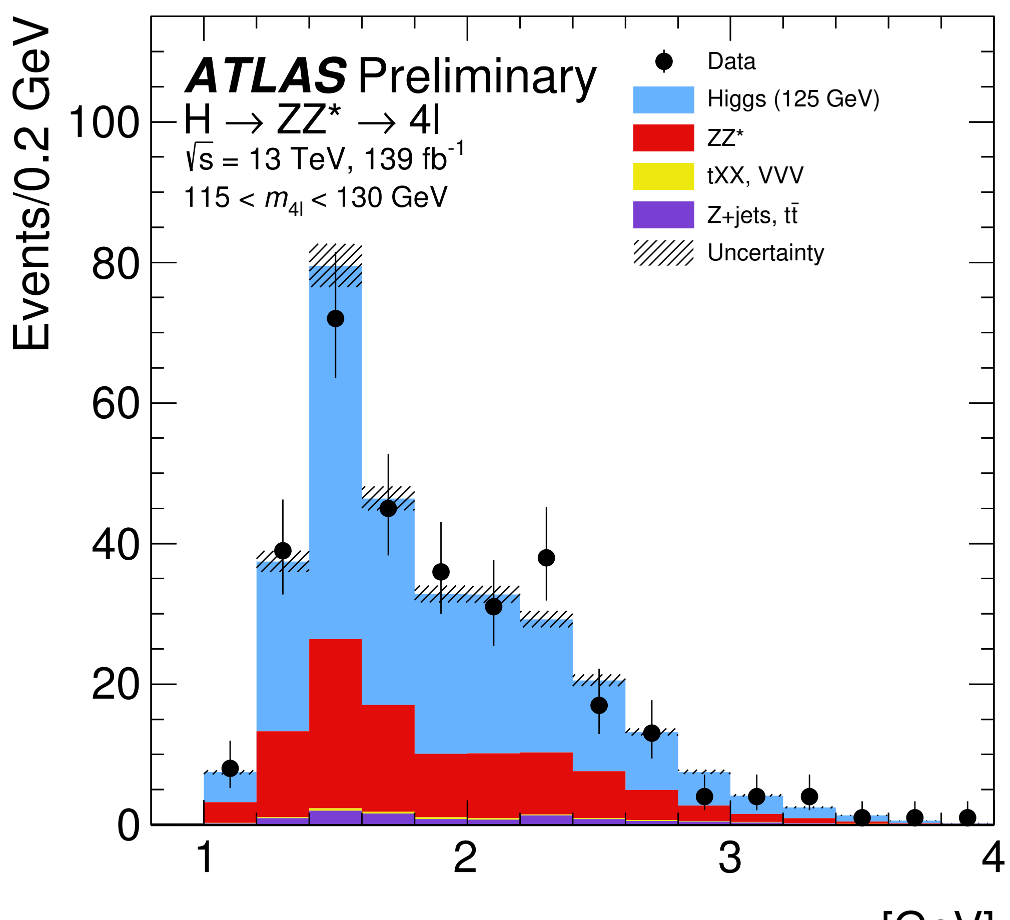
Quantile Neural Network Regression



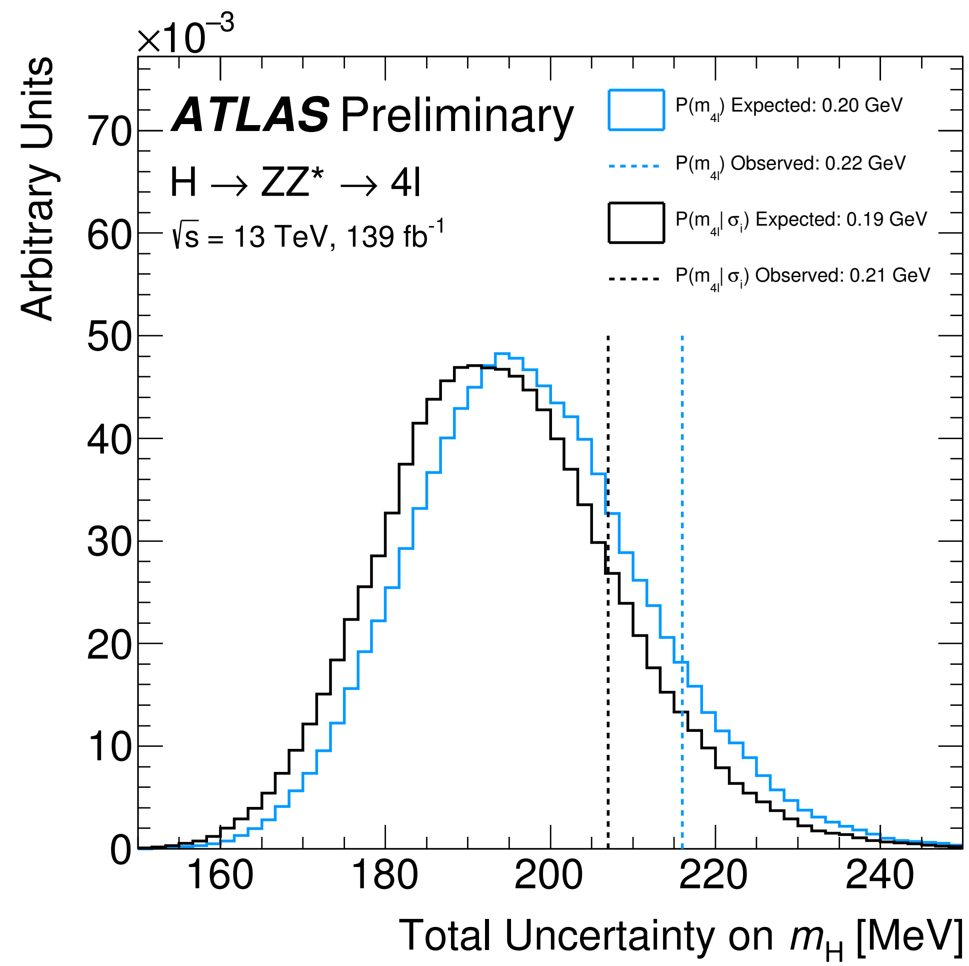
Improvements with ML

- Output used in multidimensional fit, improving with respect to average detector response.
  - Tailored to each event's characteristics.
- Smaller bias towards training assumptions, compared to previous example:
  - ML dependent on detector's response rather properties of different physics models.
  - ML parametrisation as a function of event's properties that define any underling physics.
- Preliminary uncertainty of 0.16%
  - 61% improvement w.r.t $m_{\mathrm{H}}{}^{H\rightarrow ZZ,\mathrm{Run1}}$
  - Working on further improvements in the network's prediction.

$$m_H = 124.92^{+0.21}_{-0.20}\ \mathrm{GeV}$$



Neural network output of the event's uncertainty

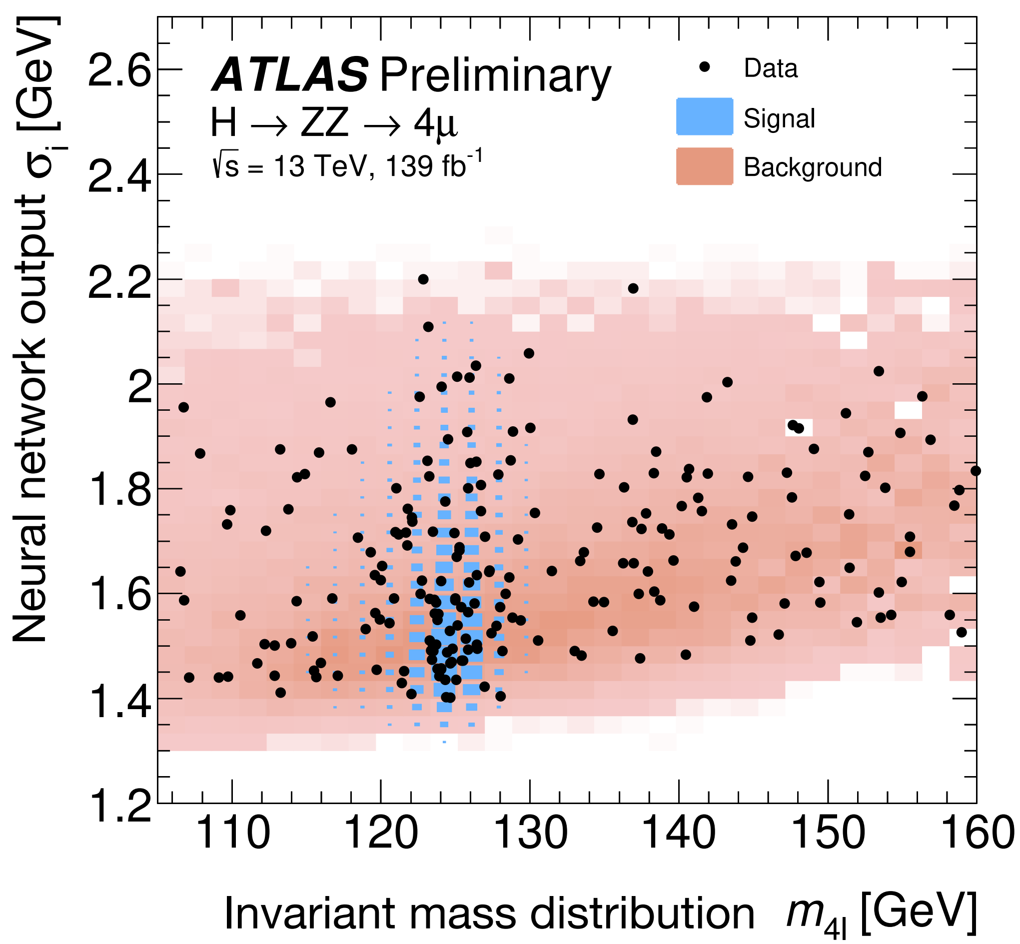


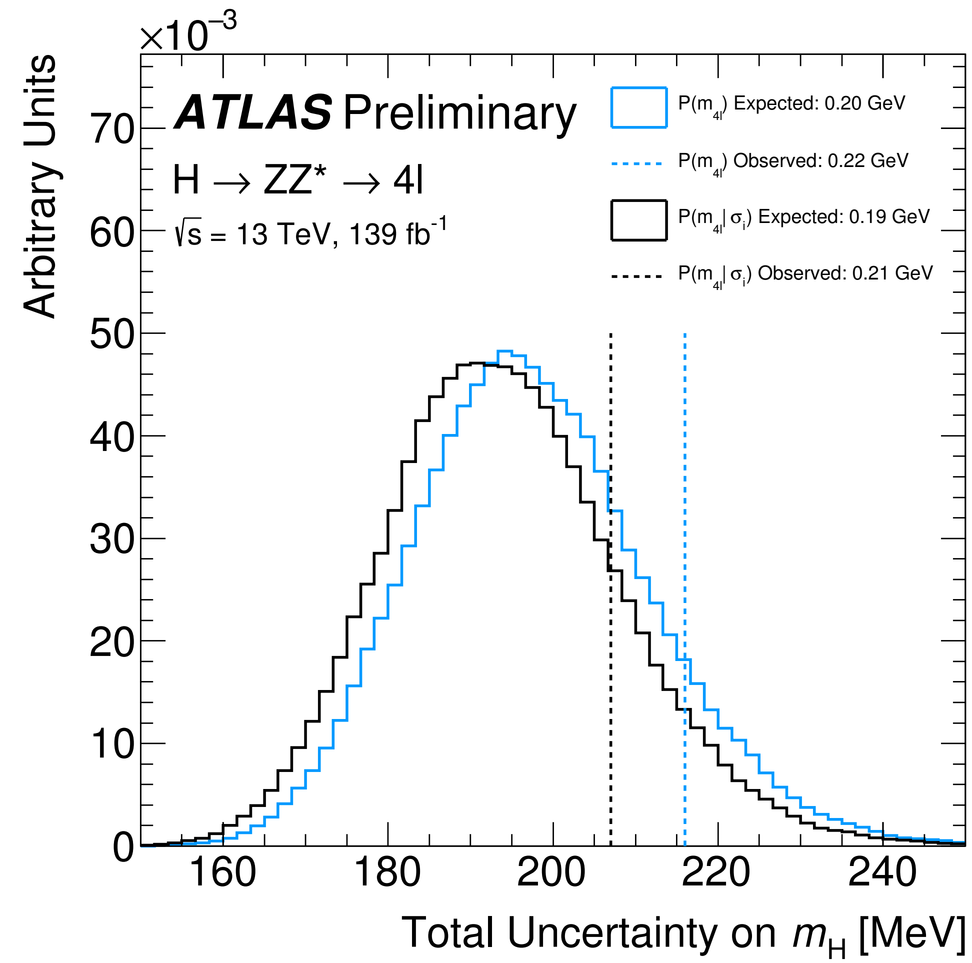
Peak of multidimensional model at lower uncertainties values than without the per-event errors.

Improvements with ML

- Output used in multidimensional fit, improving with respect to average detector response.
  - Tailored to each event's characteristics.
- Smaller bias towards training assumptions, compared to previous example:
  - ML dependent on detector's response rather properties of different physics models.
  - ML parametrisation as a function of event's properties that define any underling physics.
- Preliminary uncertainty of 0.16%
  - 61% improvement w.r.t $m_{\mathrm{H}}{}^{H\to ZZ,\mathrm{Run\,I}}$
  - Working on further improvements in the network's prediction.

$$m_H = 124.92^{+0.21}_{-0.20}\ \mathrm{GeV}$$





Peak of multidimensional model at lower uncertainties values than without the per-event errors.

Conclusions

- Machine learning techniques widely used in our current and planned analyses:
    - ✦ VBS $V_L V_L$ production, $HH\to\gamma\gamma bb$, $H\to\mu\mu$, Higgs tagging in $H\to bb$, $H\to$ invisible. $H\to ZZ^*$, $H\to WW^*$,…
    - Increase of luminosity (Run-3 and HL-LHC) will allow to explore evermore complex datasets:
        - ✦ ML key in widening the scope in complexity of our analyses.

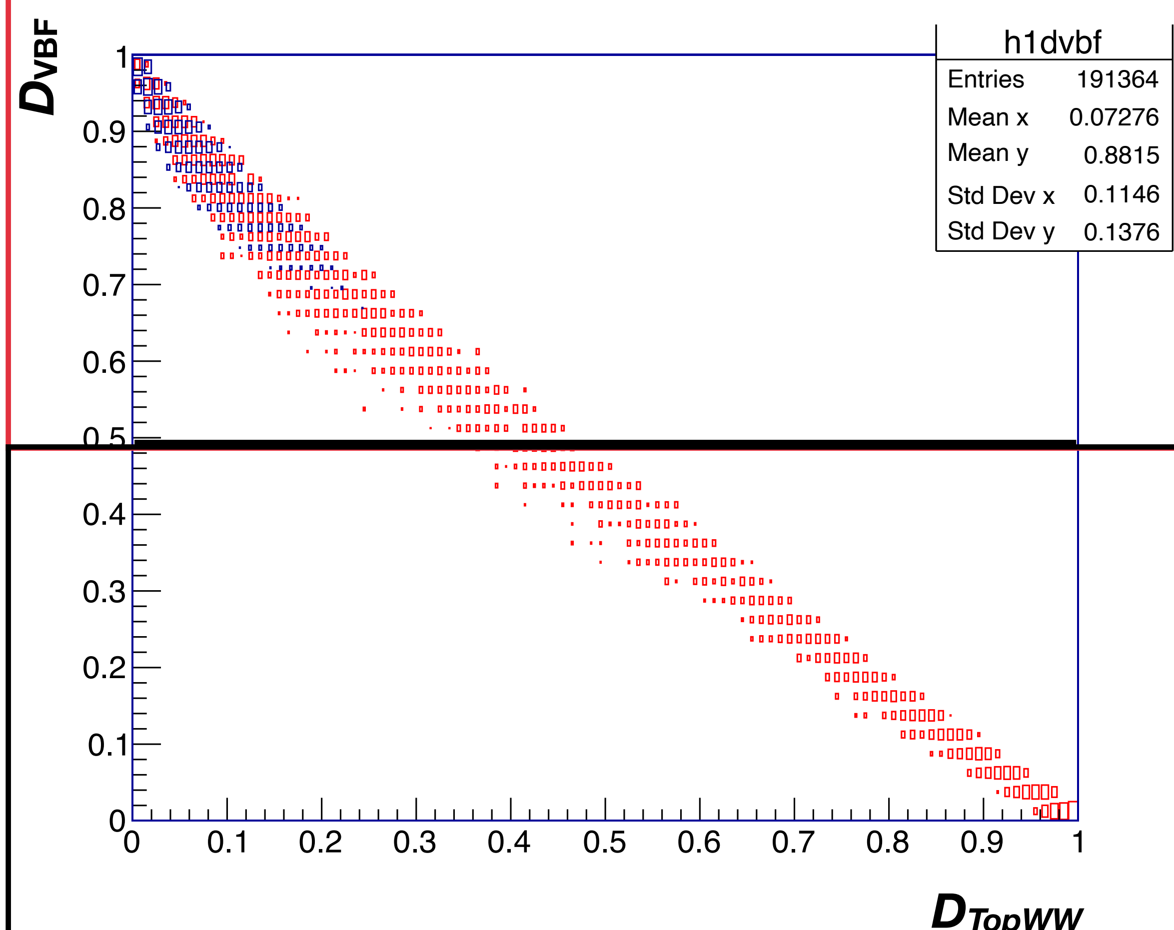
- Complex Systems knowledge application to our research:
    - we benefit from the recent advances in the field,
        - ✦ both from application tools: *Keras, TMVA, xGboost,* .. and in *fundamental research*.
    - Our complex analyses are often well suited to the these applications:
        - Multitude of (correlated) degrees of freedom *detector response* and *physics properties*.

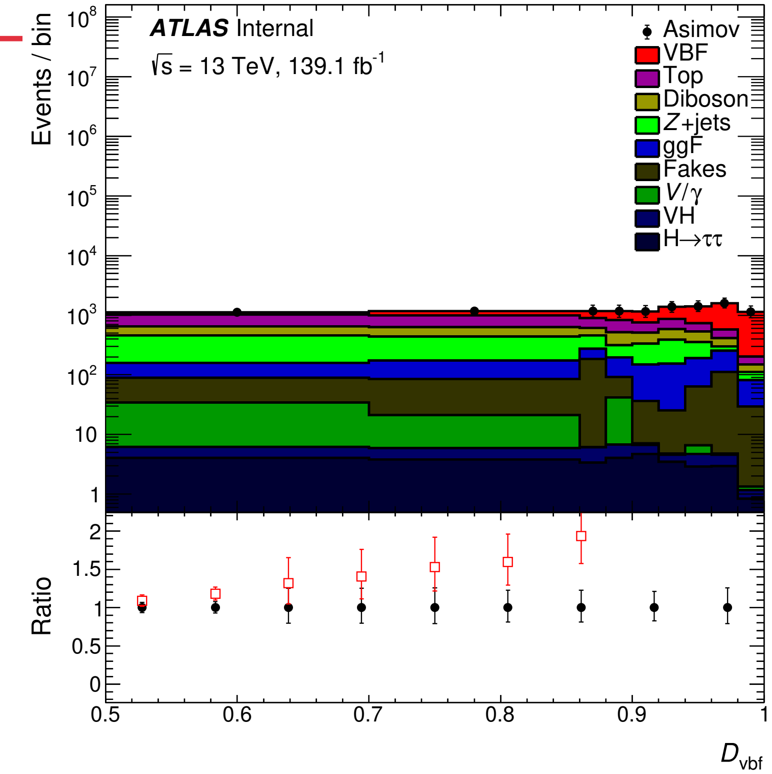
- There is no *unique* application to our analyses
    - Shown two distinct examples of applications with different needs.
    - ML applications in analyses correlate with lower level software / reconstruction quantities.

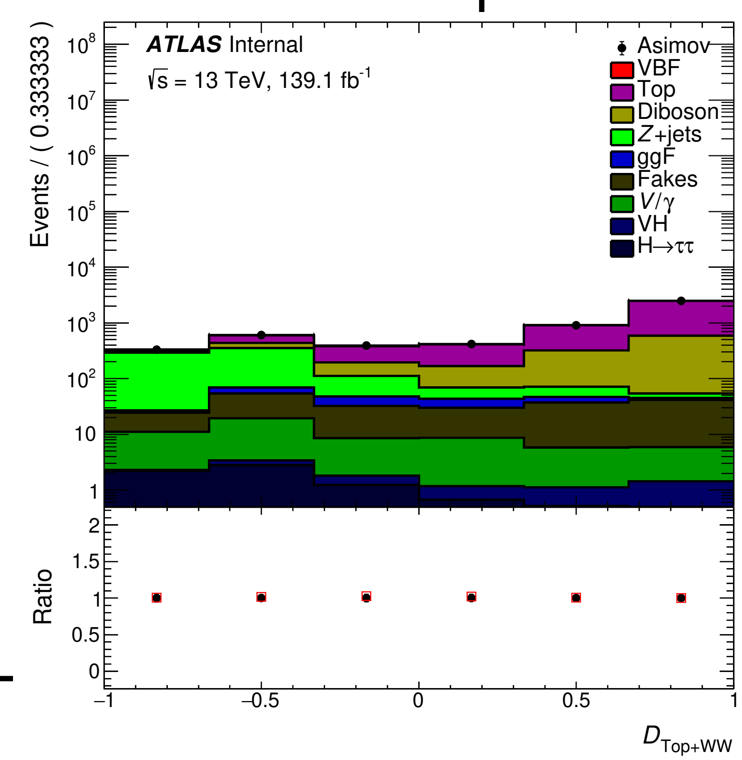
- Intra-BNL collaborations in ML to strengthen the software to analysis bond.

Additional material

- Multidimensional split of the two classifiers:
  - Perform bi-dimensional fit on the D_{VBF} vs D_{TopWW} plane.
  - Independent d.o.f for sets of differential bins in each distribution.







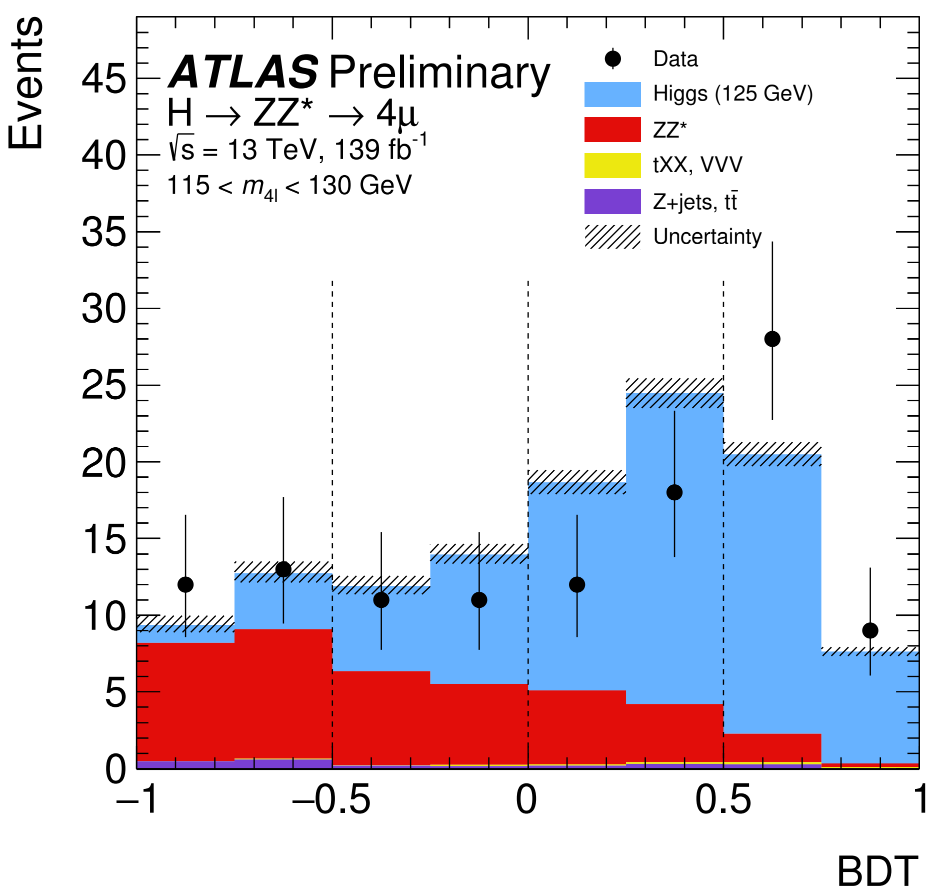
Higher level improvements

- Multi-prong approach to reduce uncertainty at analysis level:

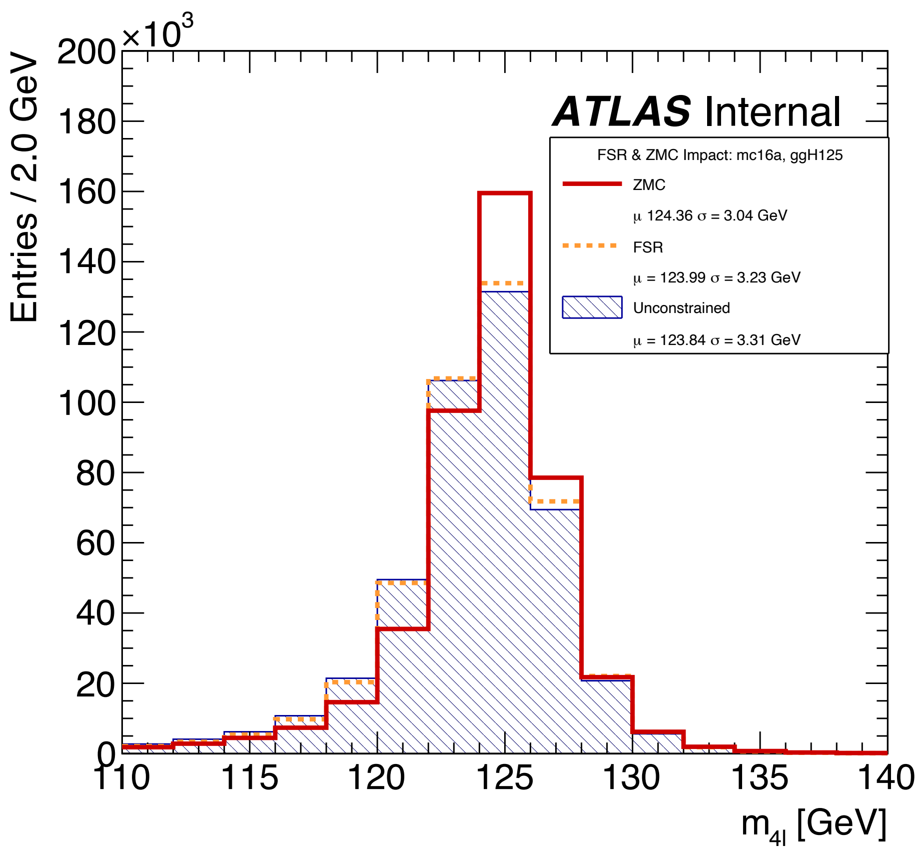
  Increase the signal to background separation:

  (i) ~15% from constraint of mass of two leading leptons, which form a well known resonance.

  (ii) ~2% from kinematic discriminant between signal and background events.



ML output for signal and background

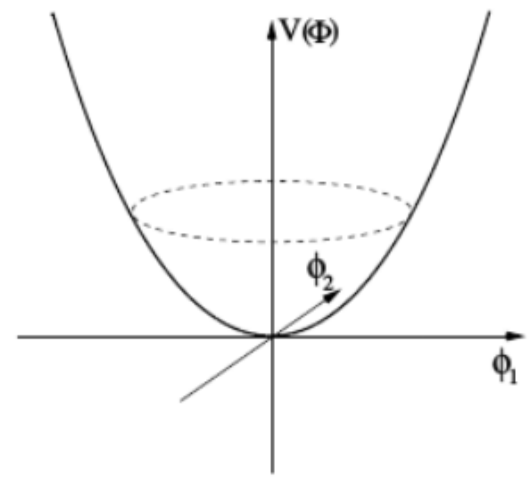


Invariant mass distribution

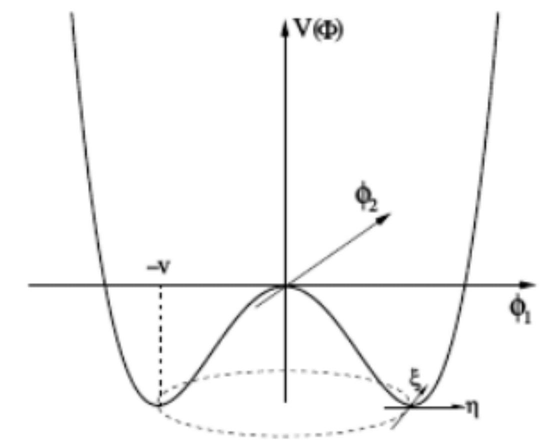
- Categorize events into four further exclusive regions.
  - Based on the output of a Machine Learning Algorithm (Boosted Decision Tree / Neural Network),
  - Using kinematic properties of the event.

- Importance of m_H in several aspects of our understanding of fundamental physics.

Unbroken symmetry



Broken symmetry

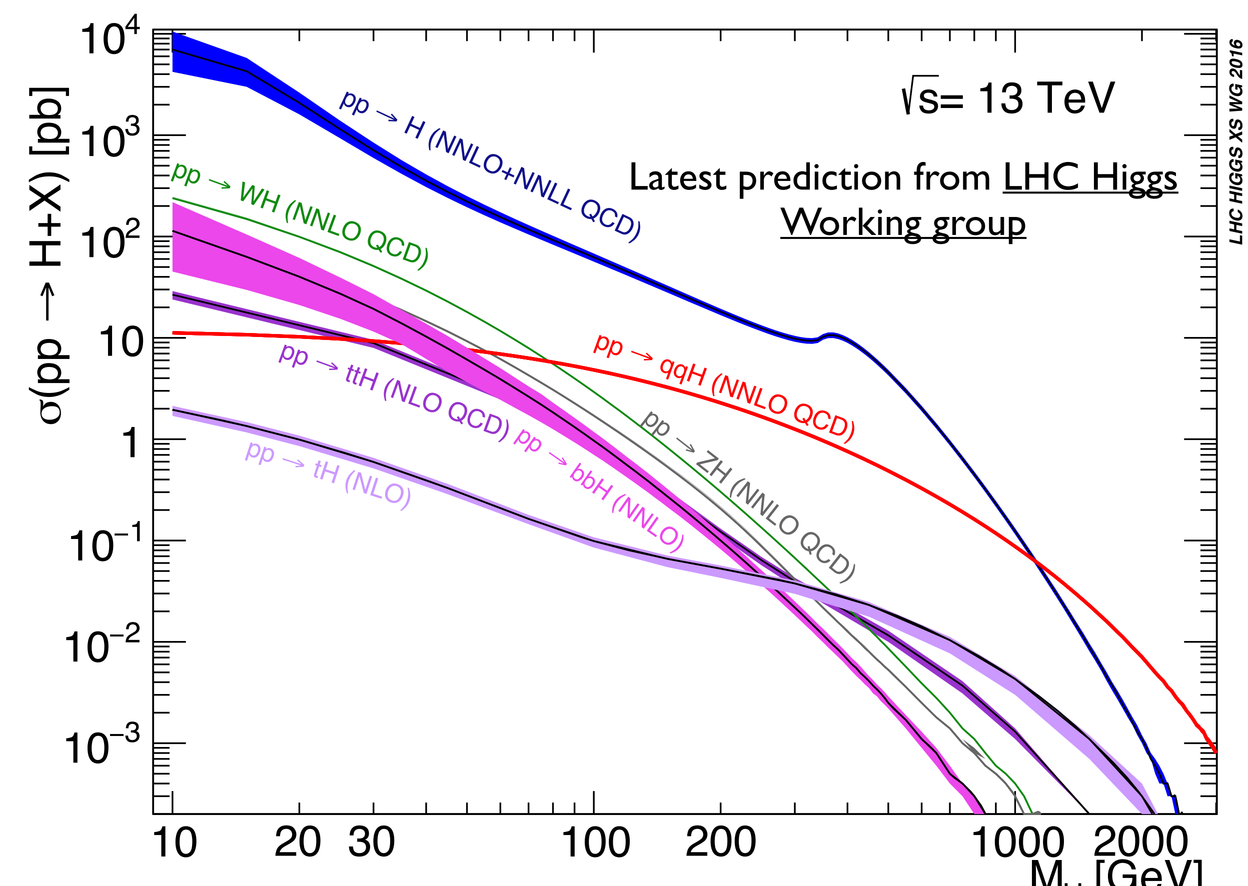


Power law expansion of the potential

$$V(h) \quad = \quad \frac{1}{4}\lambda h^4 + \lambda v h^3 + \lambda v^2 h^2$$

- Understanding the perturbative expansion of its potential ($\lambda v^2 h^2$).
- Precise higher order corrections to the theory predictions of the Higgs interactions depend on the value of m_H.
- Input to precision global fit of the Standard Model.

Aim at improving significantly on the experimental precision on m_H



Prediction and uncertainties of Higgs production processes as a function of the m_H

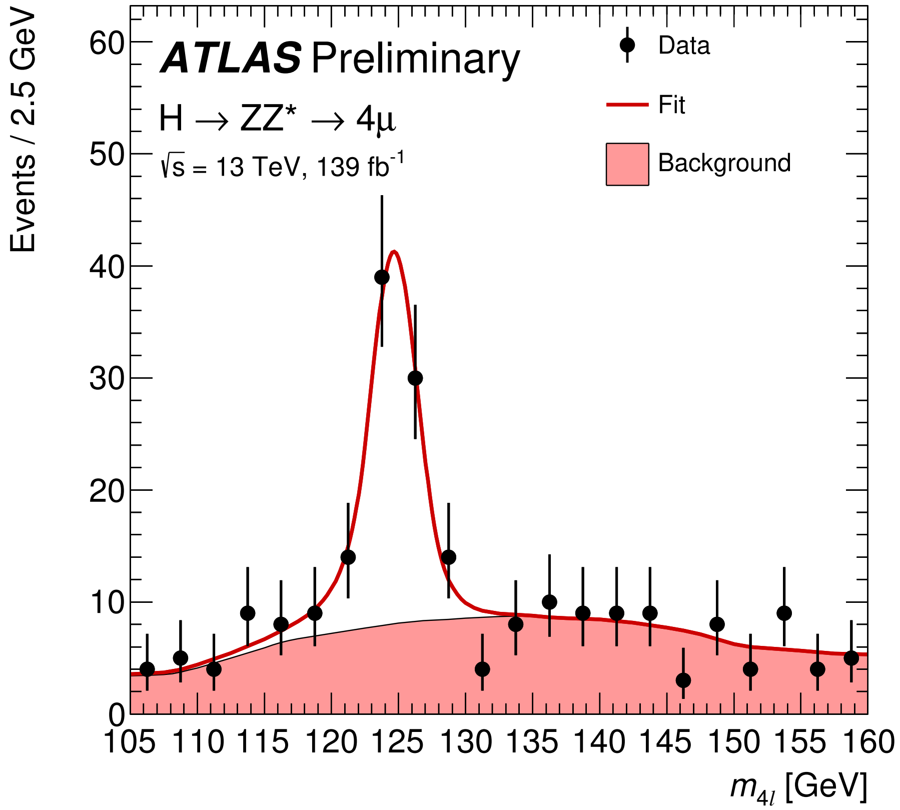
Projections

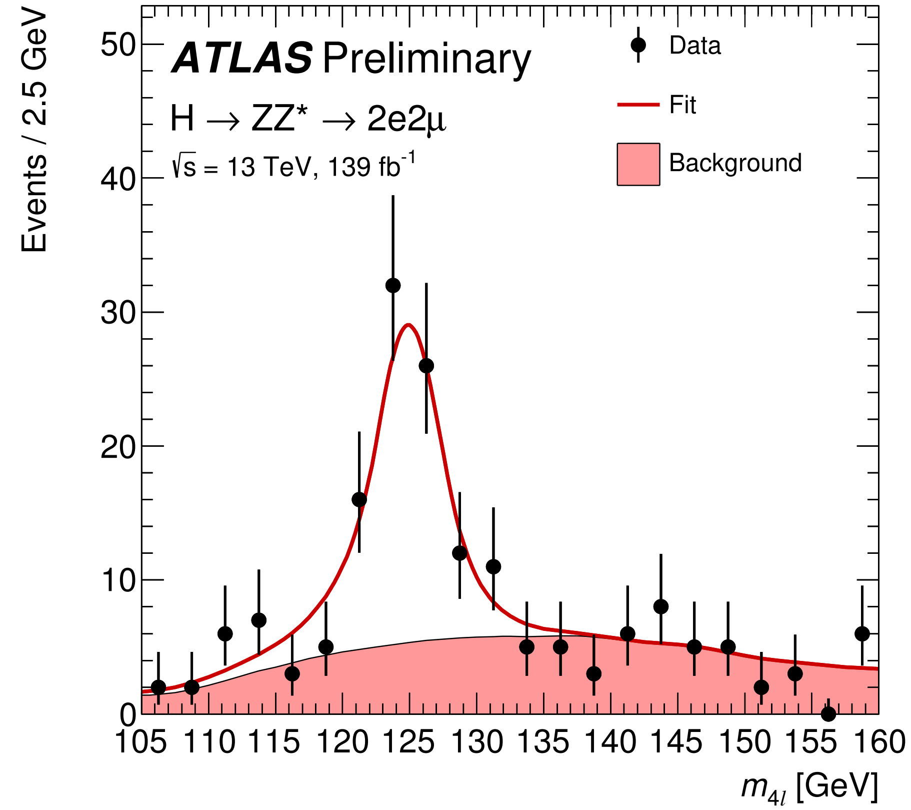
- Expected projections for m_H at High Luminosity LHC
  - With the abundance of data, it is expected that systematics will be reduced further.

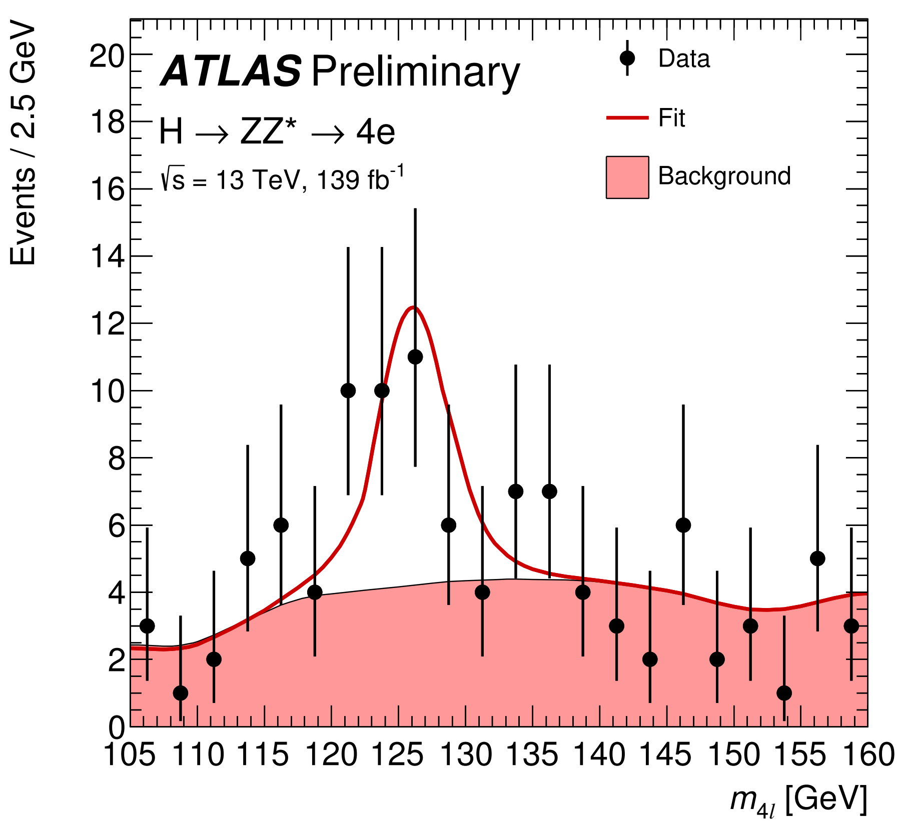
	Δ_{tot} (MeV)	Δ_{stat} (MeV)	Δ_{syst} (MeV)
Current Detector	52	39	35
μ momentum resolution improvement by 30% or similar	47	30	37
μ momentum resolution/scale improvement of 30% / 50%	38	30	24
μ momentum resolution/scale improvement 30% / 80%	33	30	14

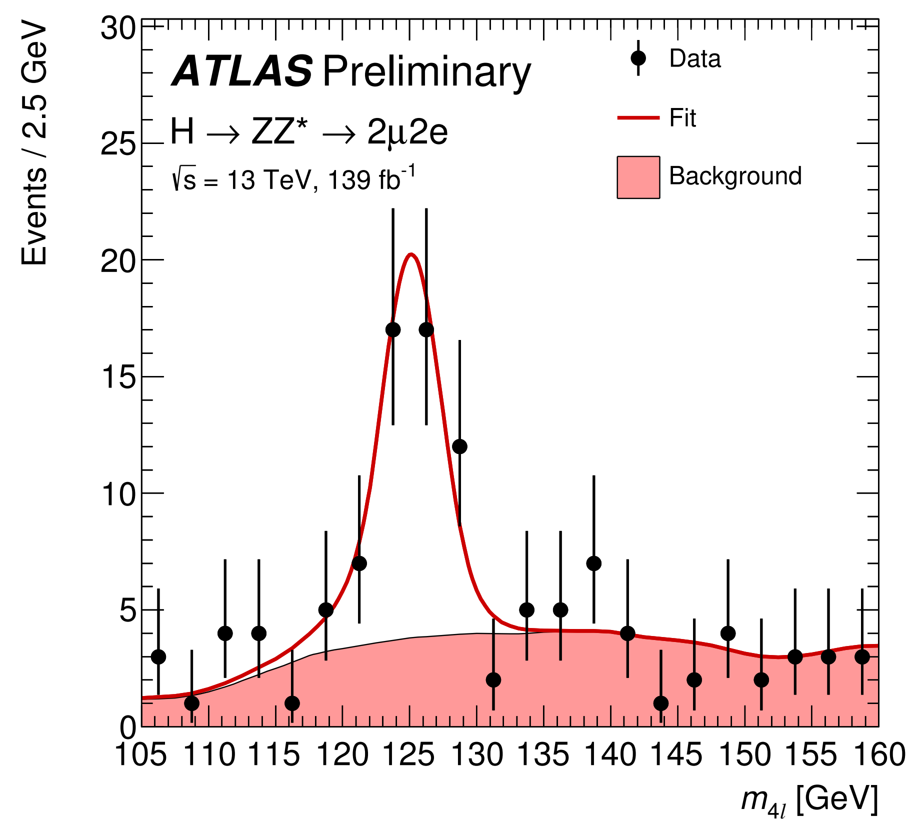
- Lepton colliders
  - In FCC-ee total uncertainty at MeV level
    - ~5.4 MeV
  - ILC about 14 MeV to 30 MeV
    - With a recoil mass of 250 GeV.

$H \rightarrow ZZ \rightarrow 4\ell$ results









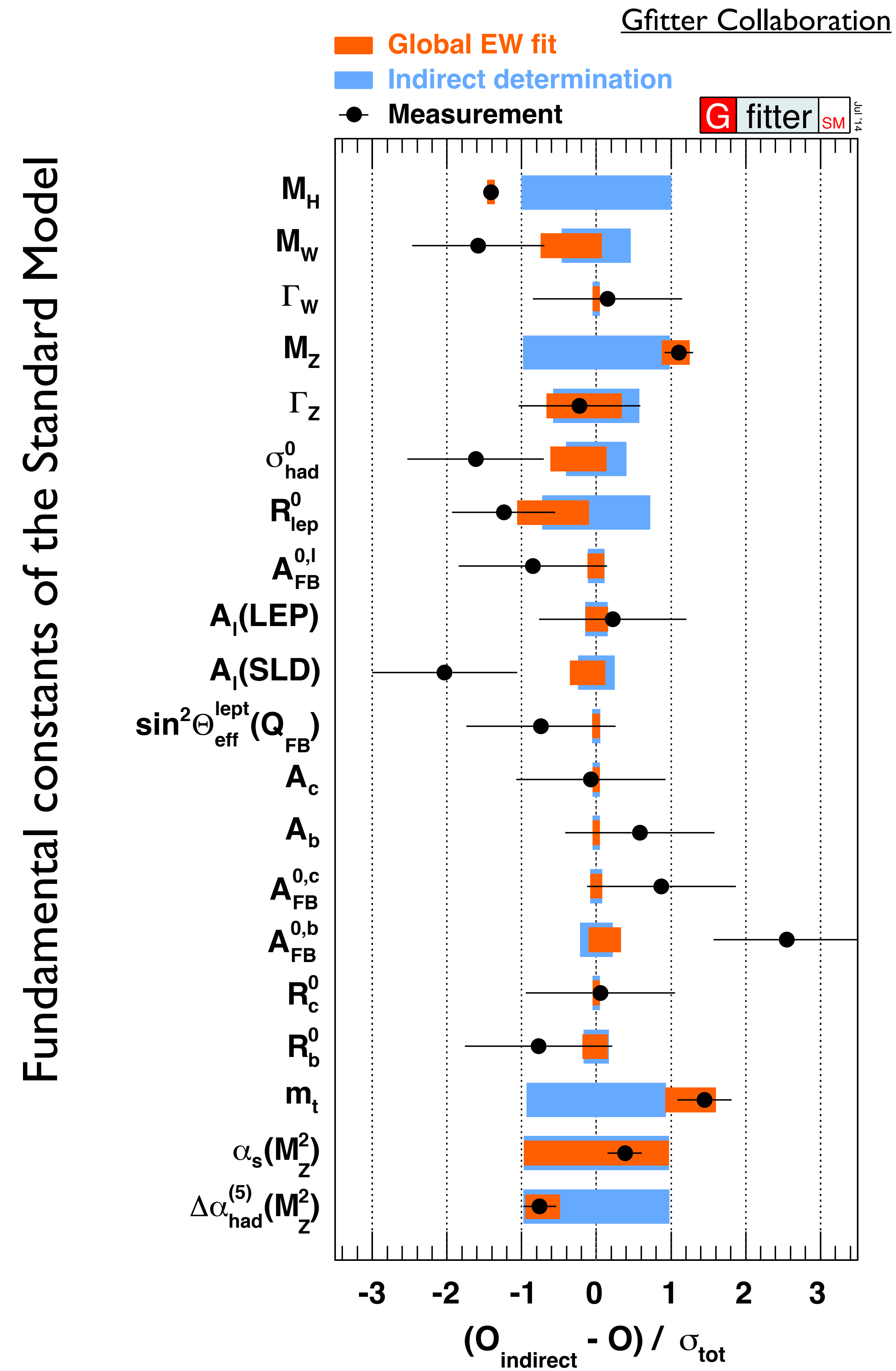
ATLAS-CONF-2020-005

- Importance of m_H in several aspects of our understanding of fundamental physics.

Power law expansion of the potential

$$V(h) \quad = \quad \frac{1}{4}\lambda h^4 + \lambda v h^3 + \lambda v^2 h^2$$

- Understanding the perturbative expansion of its potential ($\lambda v^2 h^2$).
- Precise higher order corrections to the theory predictions of the Higgs interactions depend on the value of m_H.
- Input to precision global fit of the Standard Model.

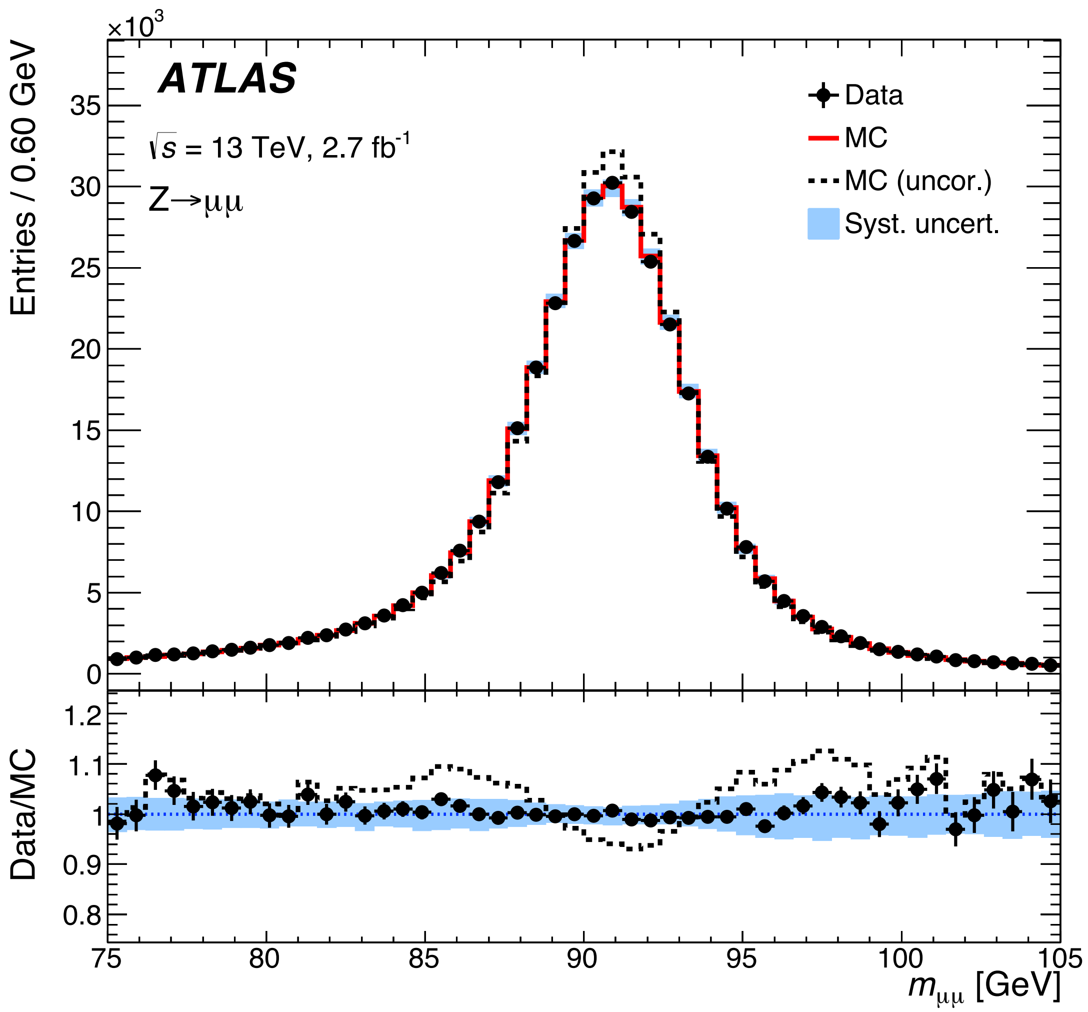


Aim at improving significantly on the experimental precision on m_H

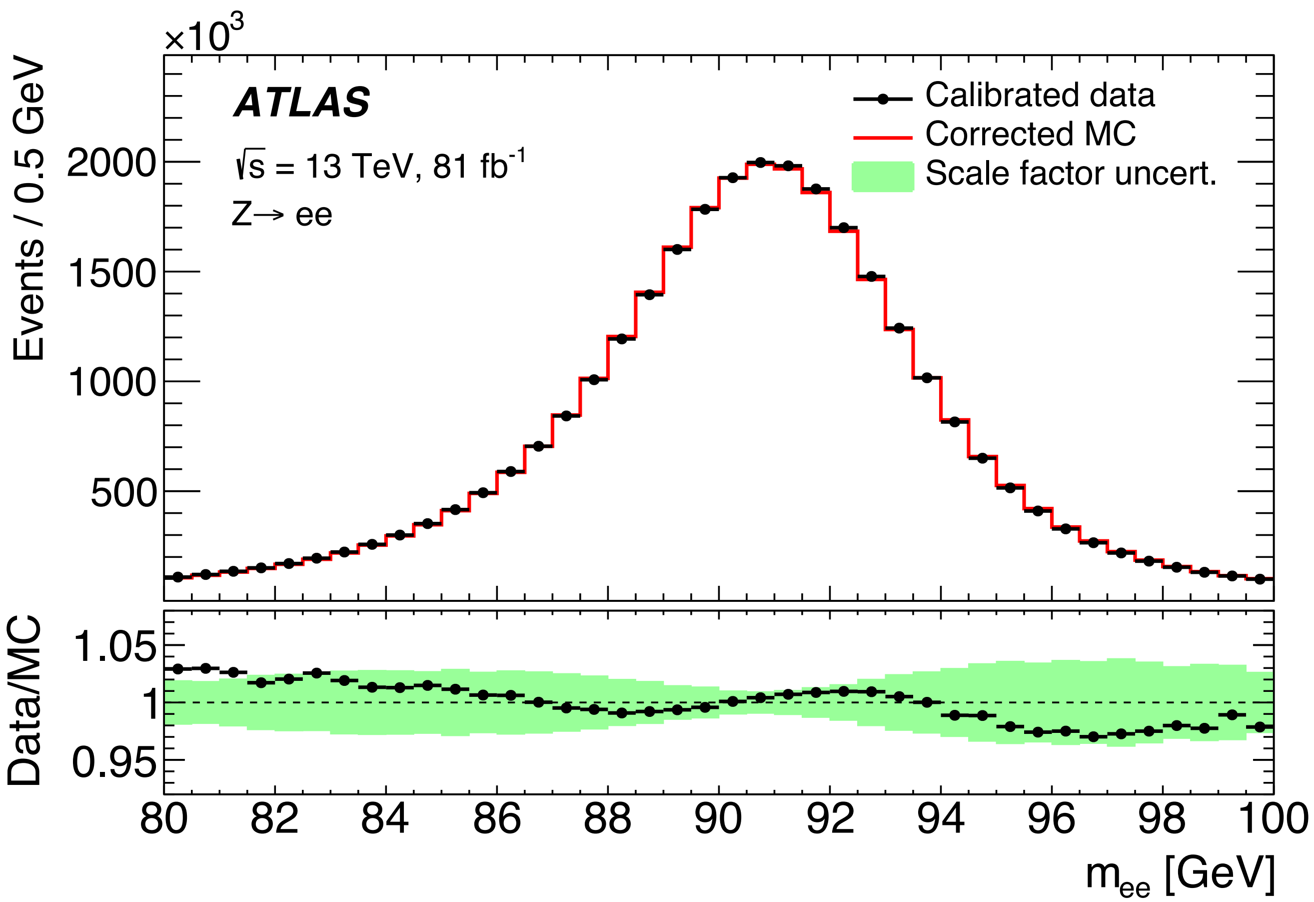
Energy resolution

- Resolution in electron and muon reconstruction crucial for m_H uncertainty.

- We used well known processes to calibrate the detector response.
  - Resonant process of J/ψ, Υ and Z,
  - for modelling of calorimeters deposits, alignment precision, etc.



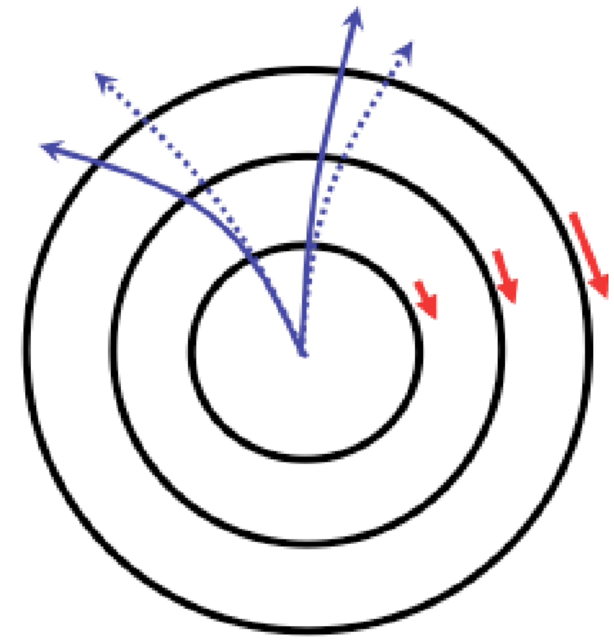
$Z \to \mu\mu$ resonant line shape for data and detector simulation



$Z \to ee$ resonant line shape for data and detector simulation

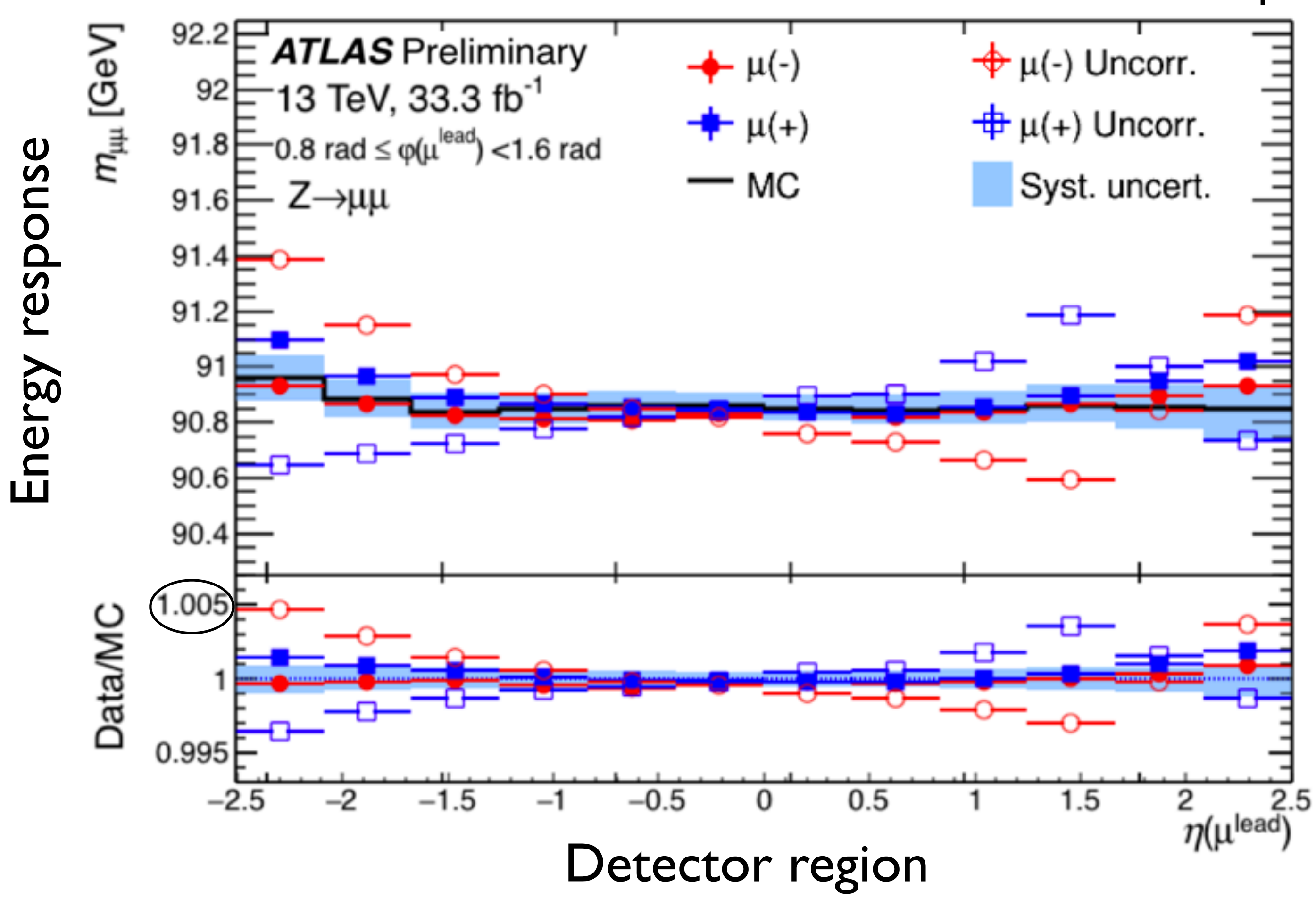
Second order effects

- High precision mandates for studying second order effects:
  - charge-dependent bias because of detector movements.
  - I created an innovative a*d-hoc* correction based on $Z \to \mu\mu$, recovering up to 5% in precision.
  - Allows for per-mille level understating of detector's systematic uncertainties.



Detector layer movements biasing the measurement of the bending of the particle

Biased positive and negative tracks
Corrected positive and negative tracks

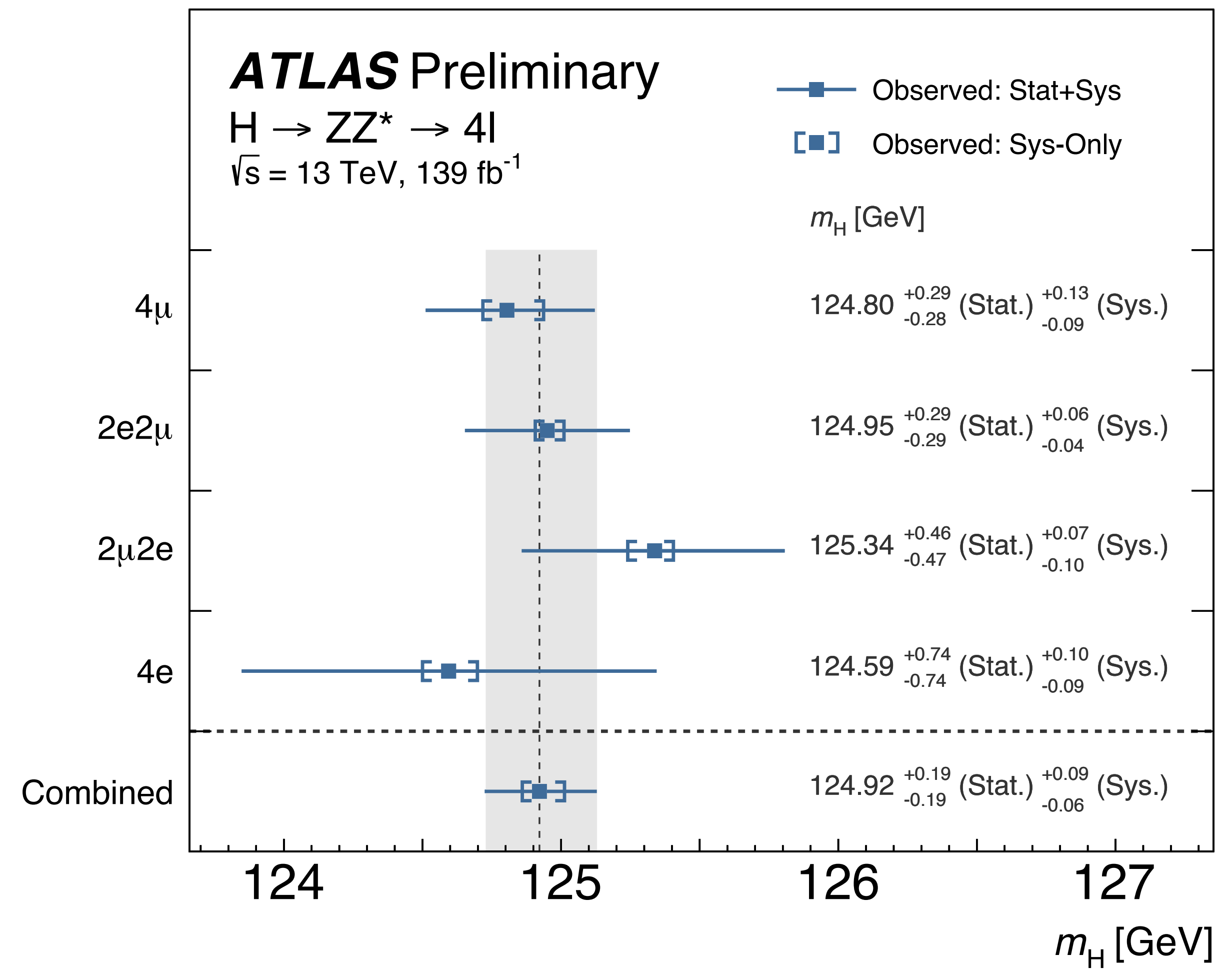


$H \to ZZ \to 4\ell$ results

- ML outputs used in multidimensional fit, improving with respect to average detector response.

- Total uncertainty of 0.16%
- Systematic uncertainty of 0.06%
  - 61% improvement w.r.t $m_{\mathrm{H}}^{H \to ZZ,\mathrm{Run\,I}}$
  - 15% improved precision w.r.t $m_{\mathrm{H}}^{\mathrm{ATLAS+CMS,Run\,I}}$
  - Most precise measurement by ATLAS, so far.

$$\boxed{m_H = 124.92^{+0.21}_{-0.20}\ \mathrm{GeV}}$$



Measured m_{H} for all channels and combined.

ATLAS-CONF-2020-005

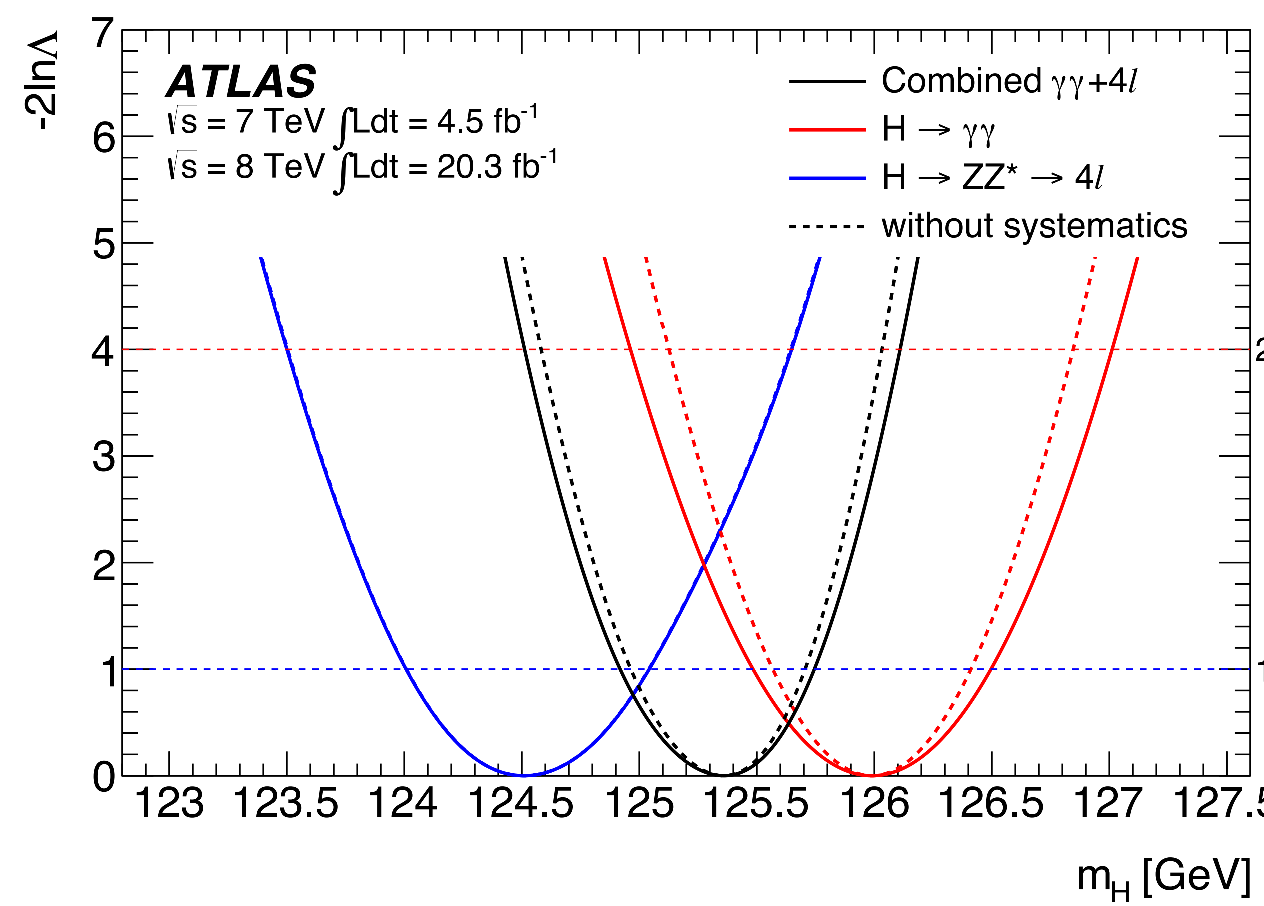
Run I status

- ATLAS run I precision on m_H of 0.33%
  - combined measurement from $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$.

Channel	Mass measurement [GeV]
$H \to \gamma\gamma$	$125.98 \pm 0.42\,(\text{stat}) \pm 0.28\,(\text{syst}) = 125.98 \pm 0.50$
$H \to ZZ^* \to 4\ell$	$124.51 \pm 0.52\,(\text{stat}) \pm 0.06\,(\text{syst}) = 124.51 \pm 0.52$
Combined	$125.36 \pm 0.37\,(\text{stat}) \pm 0.18\,(\text{syst}) = 125.36 \pm 0.41$

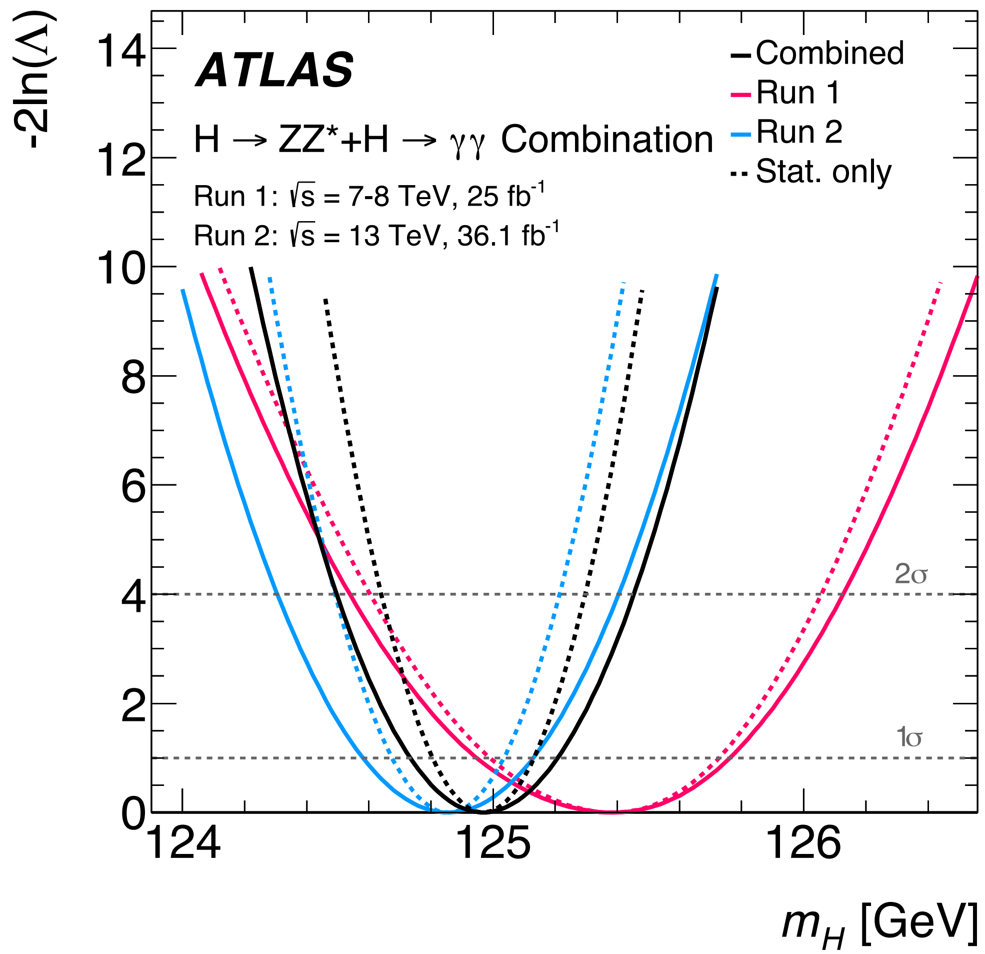
  - For both channels dominated by statistical uncertainty

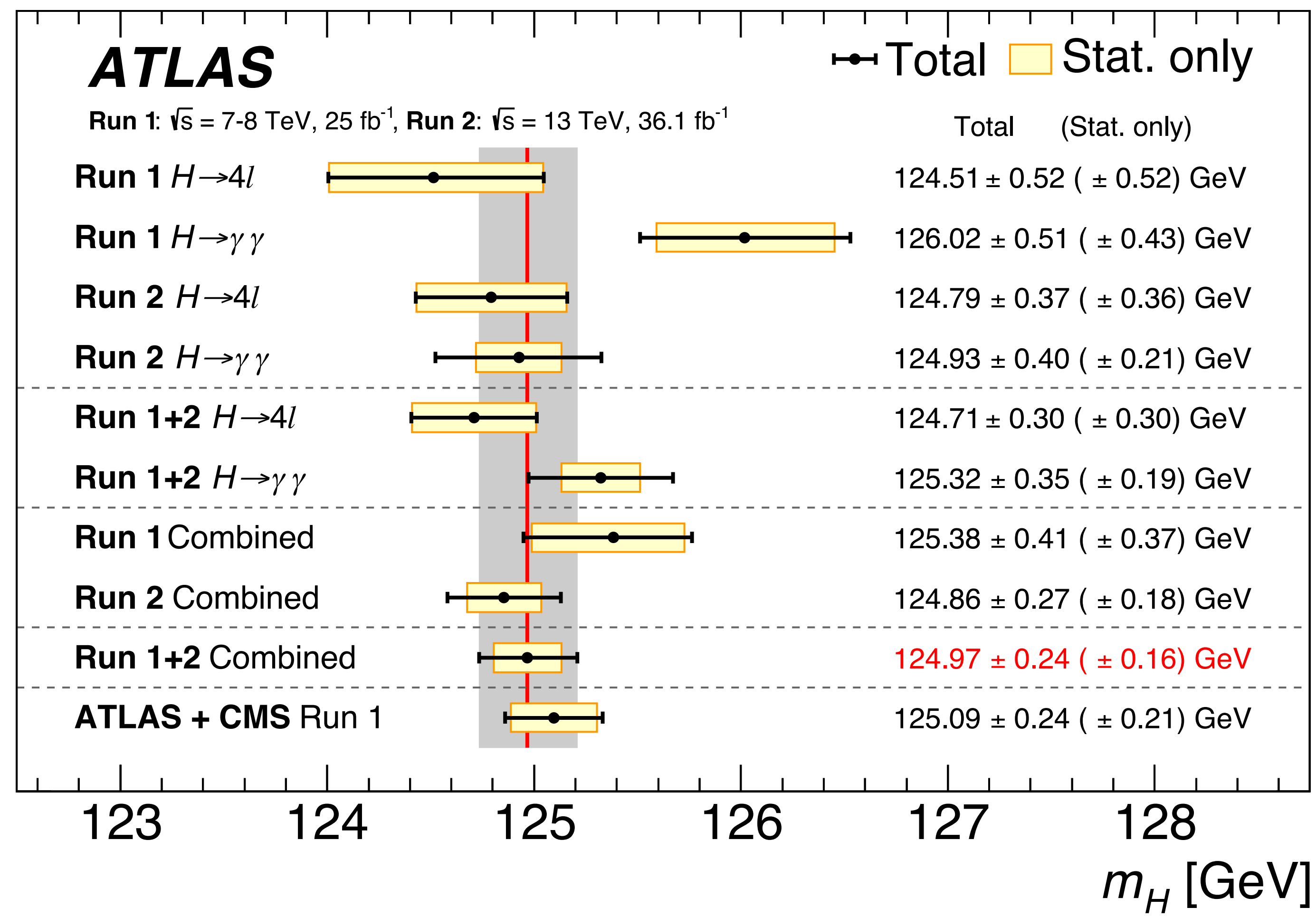
- Aim in improving significantly on δm_H
  - Expect 1.7 times more candidates, with 36 fb^{-1} at $\sqrt{s}$=13 TeV



- 4ℓ and $\gamma\gamma$ measurements are combined with ATLAS Run 1 result

arXiv:1806.00242





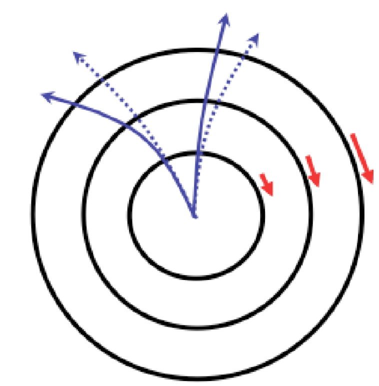
- Run 2 precision improved w.r.t Run 1.

$$m_H = 124.86 \ \pm \ 0.27(\pm 0.18 \text{ stat only}) \text{ GeV}$$

- ATLAS Run 1 + 2 comparable precision to LHC Run 1 combination.

$$m_H = 124.97 \ \pm \ 0.24(\pm 0.16 \text{ stat only}) \text{ GeV}$$

- Correction for local misalignments
  - Charge dependent bias, with net effect of worsening resolution
  - In-situ correction based on Z→μμ data, recovers up to 5% in resolution.
  - Iteratively removing the bias δ_s:



$$p_{\mathrm{T}}^{\mathrm{corr}}(\mu) = \frac{p_{\mathrm{T}}^{\mathrm{bias}}(\mu)}{1 - q(\mu)\delta_s(\eta,\phi)p_{\mathrm{T}}^{\mathrm{bias}}(\mu)}$$