Development of novel experimental techniques to improve the understanding of the Higgs sector by the ATLAS experiment

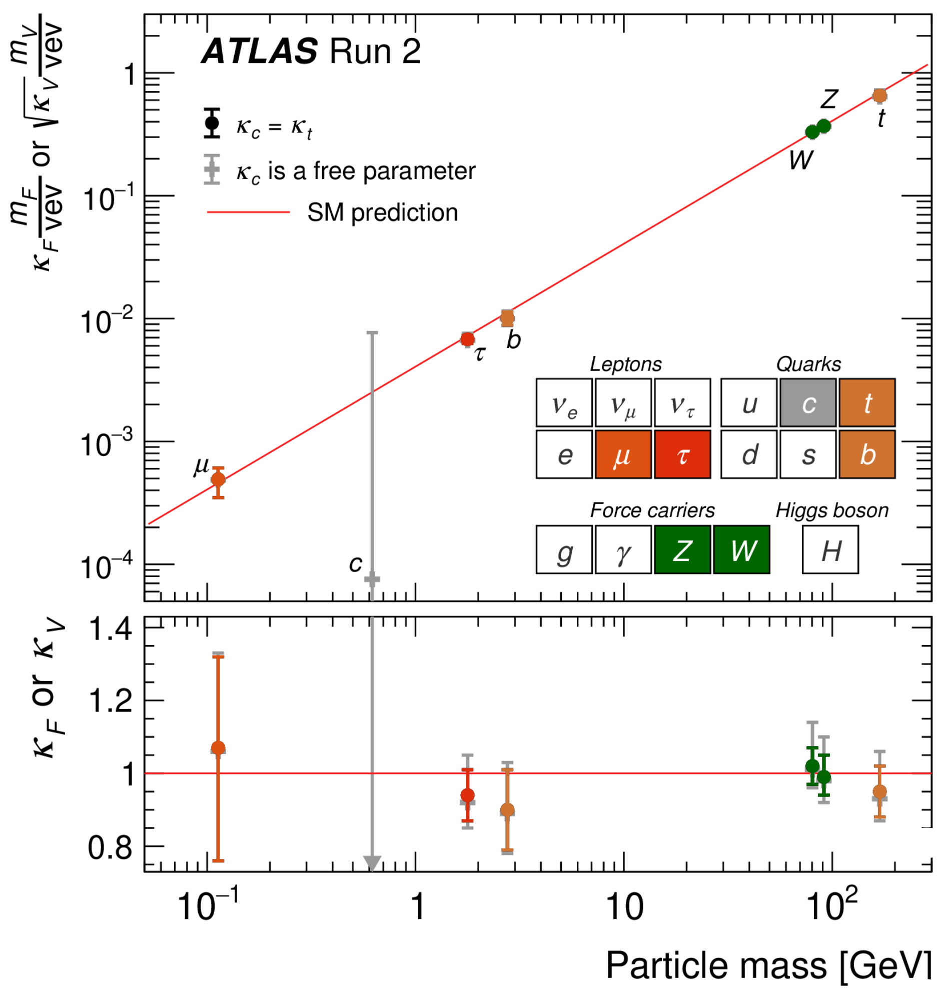
On behalf of the ATLAS Collaboration
Stephen Jiggins
07/07/22

Contents

→ **Goal:** Summarise new/uncommon analysis techniques used by ATLAS data analyses

→ **Focus** on the details of a few key Higgs measurements:

Source: HIGG-2021-23



H→ γγ:
→ Multi-MVA model fit *D-optimality*

Machine Learning *- Classification*

H→ZZ→llll:
→ Recurrent neural network quantile regress of m_{4l} mass resolution

Machine Learning *- Error Estimation*

VBF, H→bb:
→ Mass independent event classification
→ Z(bb)+jets *b*-jet object embedding

Machine Learning *&* **Object Embedding**

VBF+VH, H→ ττ:
→ Z(ττ)+jets kinematic object embedding

Object Embedding

VH, H→bb/cc:
→ b-/c-jet *truth* tagging
→ Multi-dimensional parameterisation of MC theory uncertainties

Likelihood Re-weighting *- Statistical Precision Enhancement*

H→llγ – *(Backup only)*
→ Near-by electron identification

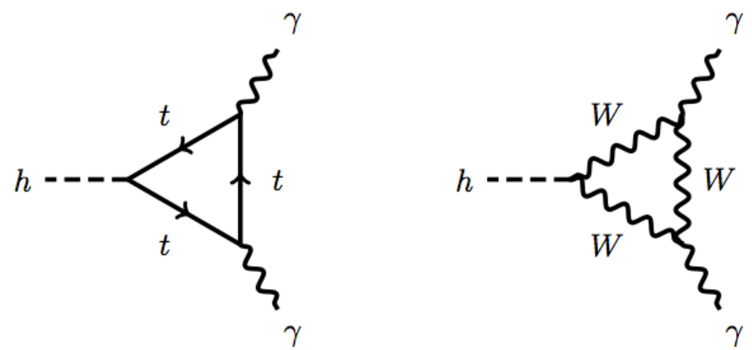
Object Reconstruction

Machine Learning

- H($\gamma\gamma$) STXS Multi-class BDT
- H(ZZ^*) Quantile Regression Neural Network
- VBF H(bb) Adversarial Neural Network

H(γγ) – Multi-Class D-optimality

→ H(γγ) valuable for measuring Higgs properties with precision:



→ New publication addresses Stage 1.2 **Simplified Template Cross-section Measurement** (STXS)

- → ATLAS-CONF-2020-026 superceded by HIGG-2020-16
- → 44 STXS regions in total
- → 28 *merged* STXS regions used in final fit

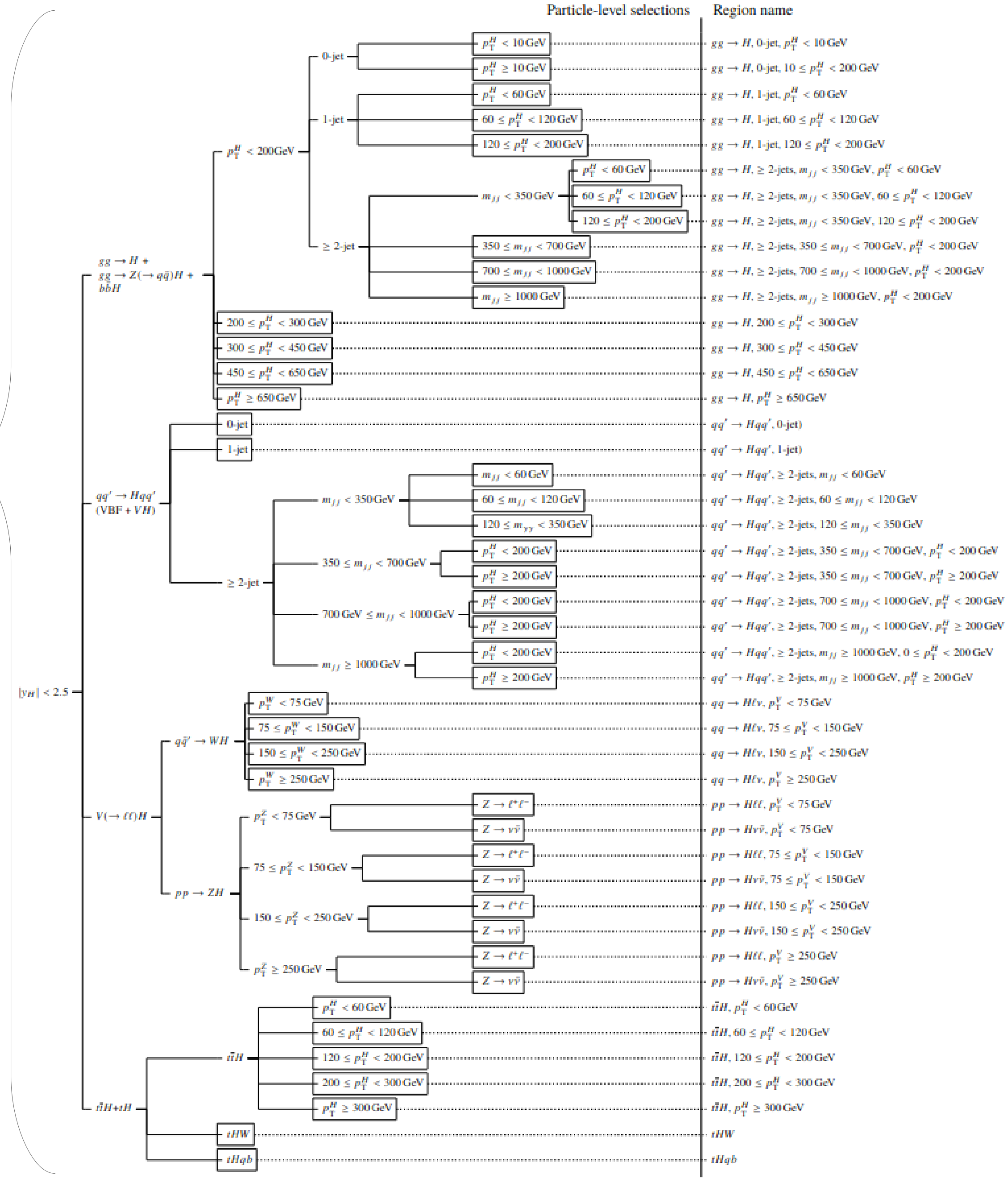
→ **Multi-class** BDT used classify events into STXS regions y_i:

- → Trained using categorical cross-entropy:

$$L=\sum_i^{N=44} y_i \log(\hat{y}_i)$$

- → $\hat{y}_i$ is defined as the softmax converted score of the BDT (z_i):

$$\hat{y}_i=\frac{e^{z_i}}{\sum_j e^{z_j}}$$



H(γγ) – Multi-Class D-optimality

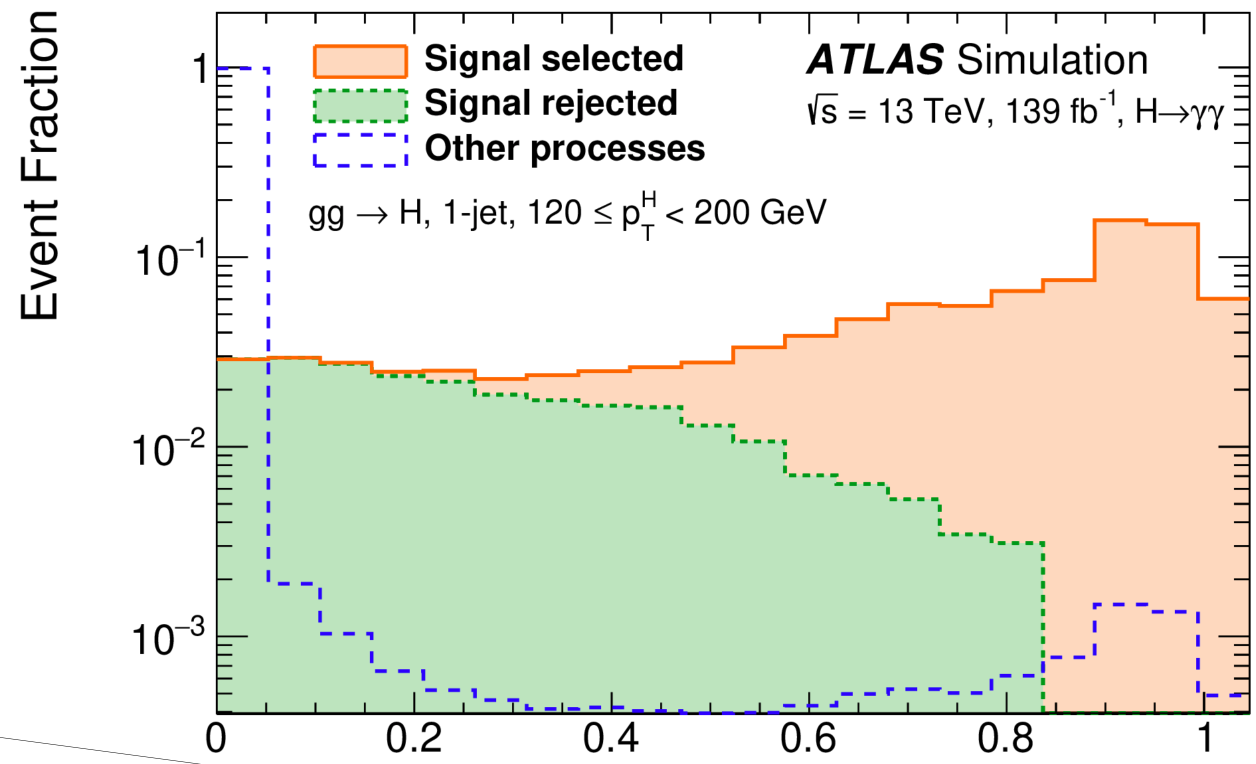
⟶ **Goal:** Single scalar metric for fitting, not 44 class labels

⟶ **Ideally:** Metric to be optimised should be:
- ⟶ Sensitivity aware due to statistical errors
- ⟶ Reduce STXS bin correlations

⟶ **D-optimality:** Determinant of the covariance matrix (C_{xx}):

$$D_{opt} = \frac{1}{2} \cdot \log \frac{|C_{\exp} + C_{theo.}|}{|C_{\exp}|}$$

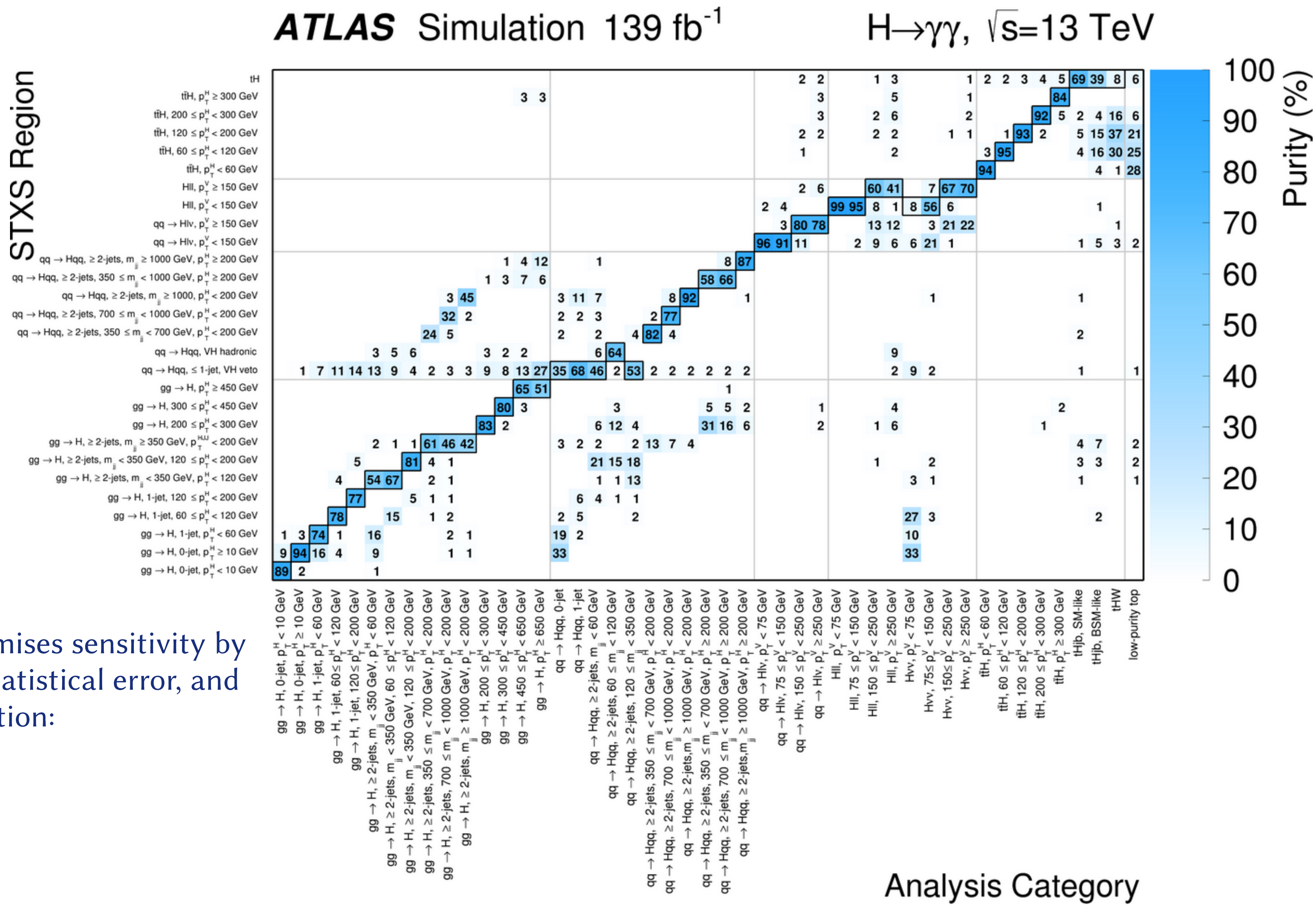
⟶ **Transform:** Transform output of BDT (z_i) to a new more expressive discriminant and take maximum: $max(\hat{z}_i = w_i z_i)$



Multiclass BDT output

⟶ **Optimisation:** The transformation weights w_i determined by iteratively minimising D_{opt} after successive 1-bin Asimov fits of STXS binnings

H(γγ) – Multi-Class D-optimality



→ **Outcome:** Maximises sensitivity by accounting for MC statistical error, and reduces STXS correlation:

$H \to ZZ^* \to llll$ – Mass resolution uncertainty

→ New Higgs mass (m_H) measurement using $H \to ZZ^* \to l^+l^-l^+l^-$

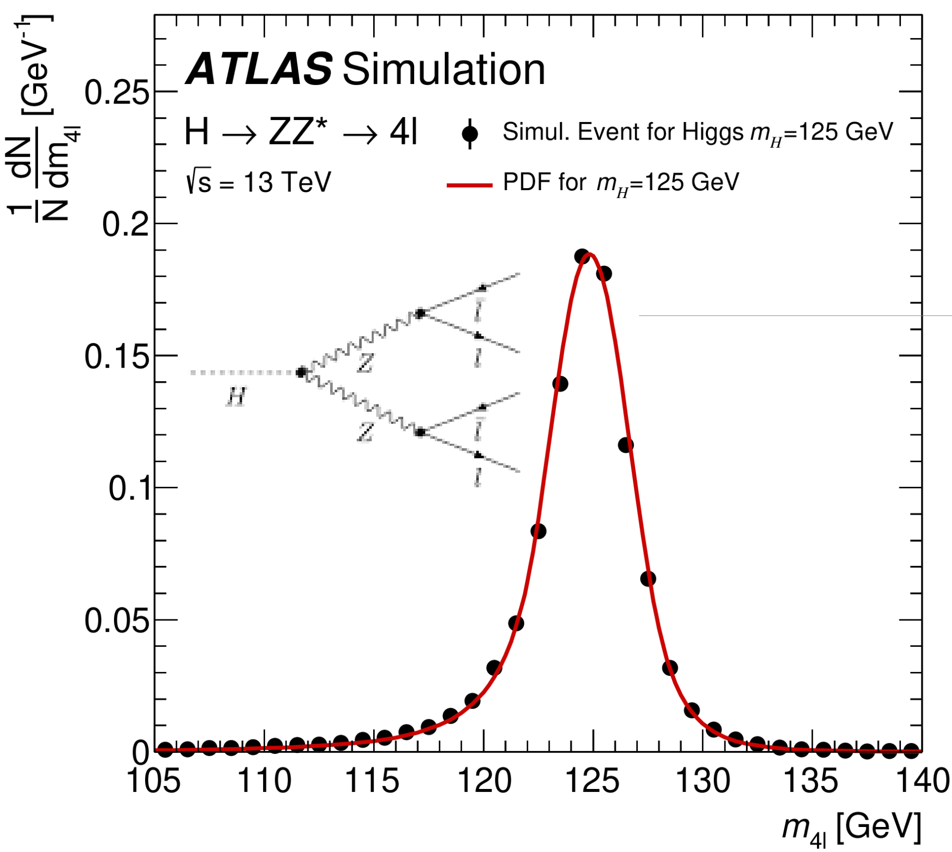
→ Unbinned maximum likelihood fit to data

$$\mathcal{L}(m_H|\boldsymbol{x}) = \mathcal{L}(m_H|m_{4\ell}, D, \sigma) = \prod_i P(m_{4\ell}^i, D^i, \sigma^i|m_H)$$

→ Signal *probability density function* (pdf) conditional on:

1. $\mathbf{m_H}$ = Higgs mass
2. $\mathbf{D_{NN}}$ = Dense Neural Network classifier
3. $\boldsymbol{\sigma_i}$ = Event level m_{4l} resolution

$$\left.\begin{array}{l} 1. \\ 2. \\ 3. \end{array}\right\} P_s(m_{4\ell}, D, \sigma|m_H)$$



→ P_S constructed assuming a *Double Sided Crystal Ball* (DCB)

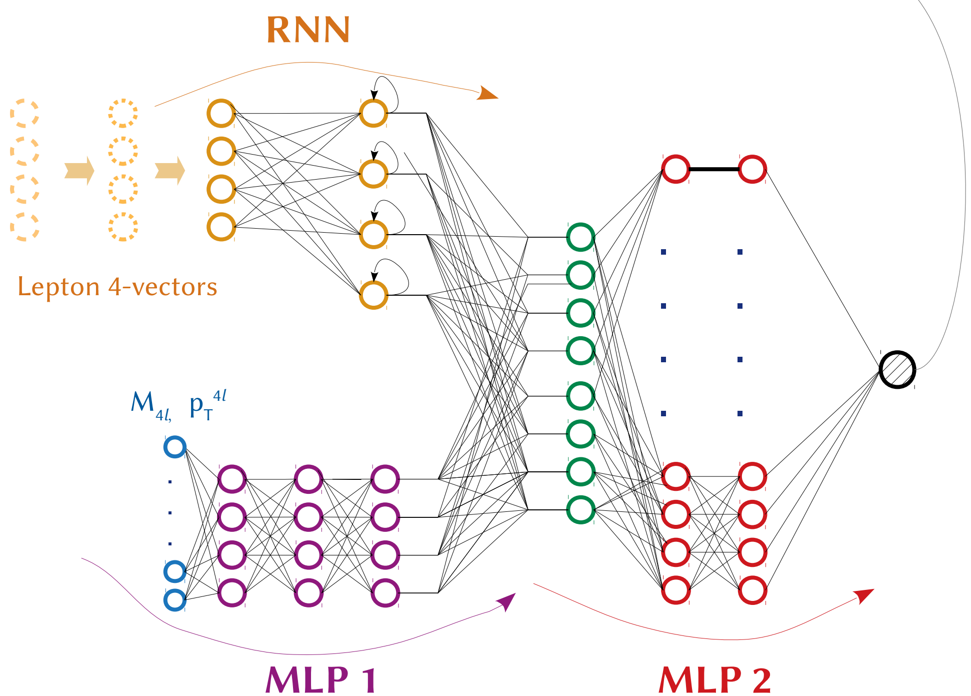
→ Propagating ±1σ (68% quantile) lepton reconstruction uncertainties to m_{4l} does not correspond to 68% quantile of DCB

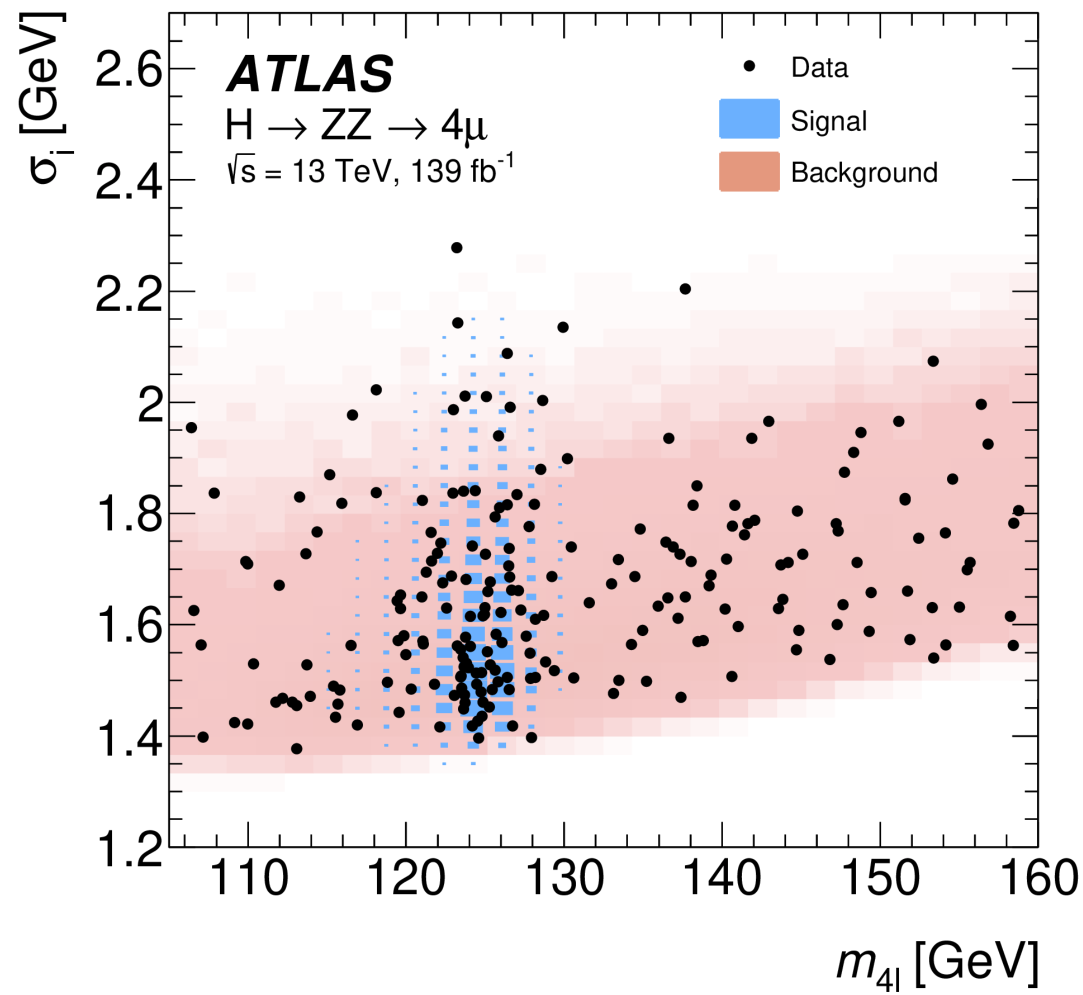
H→ZZ*→ llll – Mass resolution uncertainty

→ **Recurent Neural Network** + MLP used to estimate the **per-event** quantile of the difference between **truth** and **reconstructed** 4-lepton mass:

$P_s(m_{4\ell}, D, \sigma | m_H)$

$|m_{4l}^{\text{constrained}} - m_{4l}^{\text{truth,born}}|$





→ Loss function is the square difference between true and predicted m_{4l} quantile:

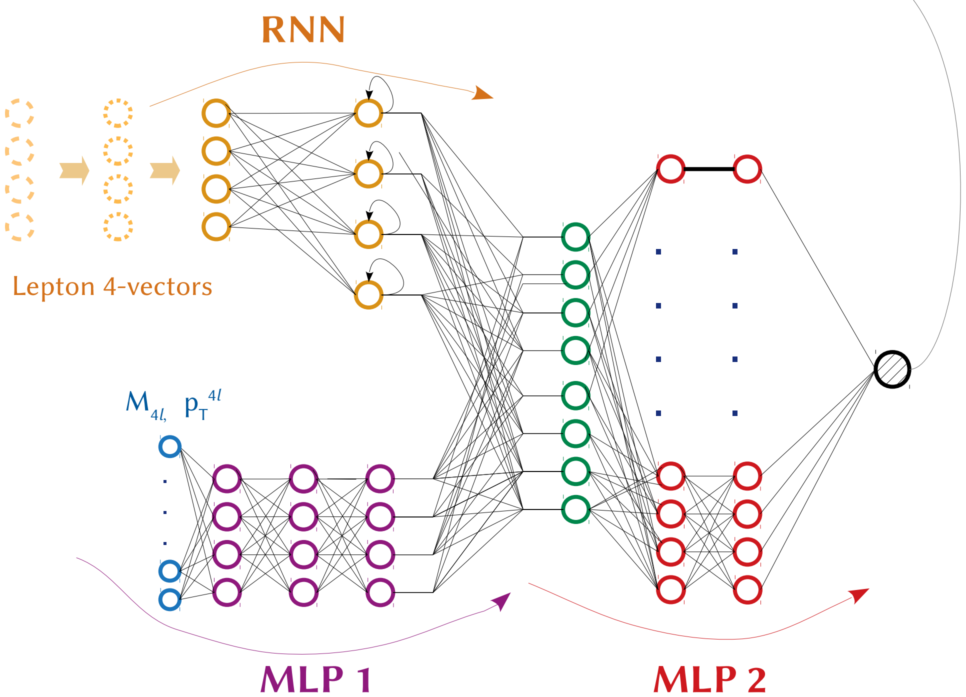
$$\sum_N \max(q \times (\text{true-predicted}), (1-q) \times (\text{true-predicted})) + \frac{1}{w} \sum_N \left(\frac{predicted}{\sigma_{4\ell}^{\text{constrained}}} - -\mu \right)^2$$

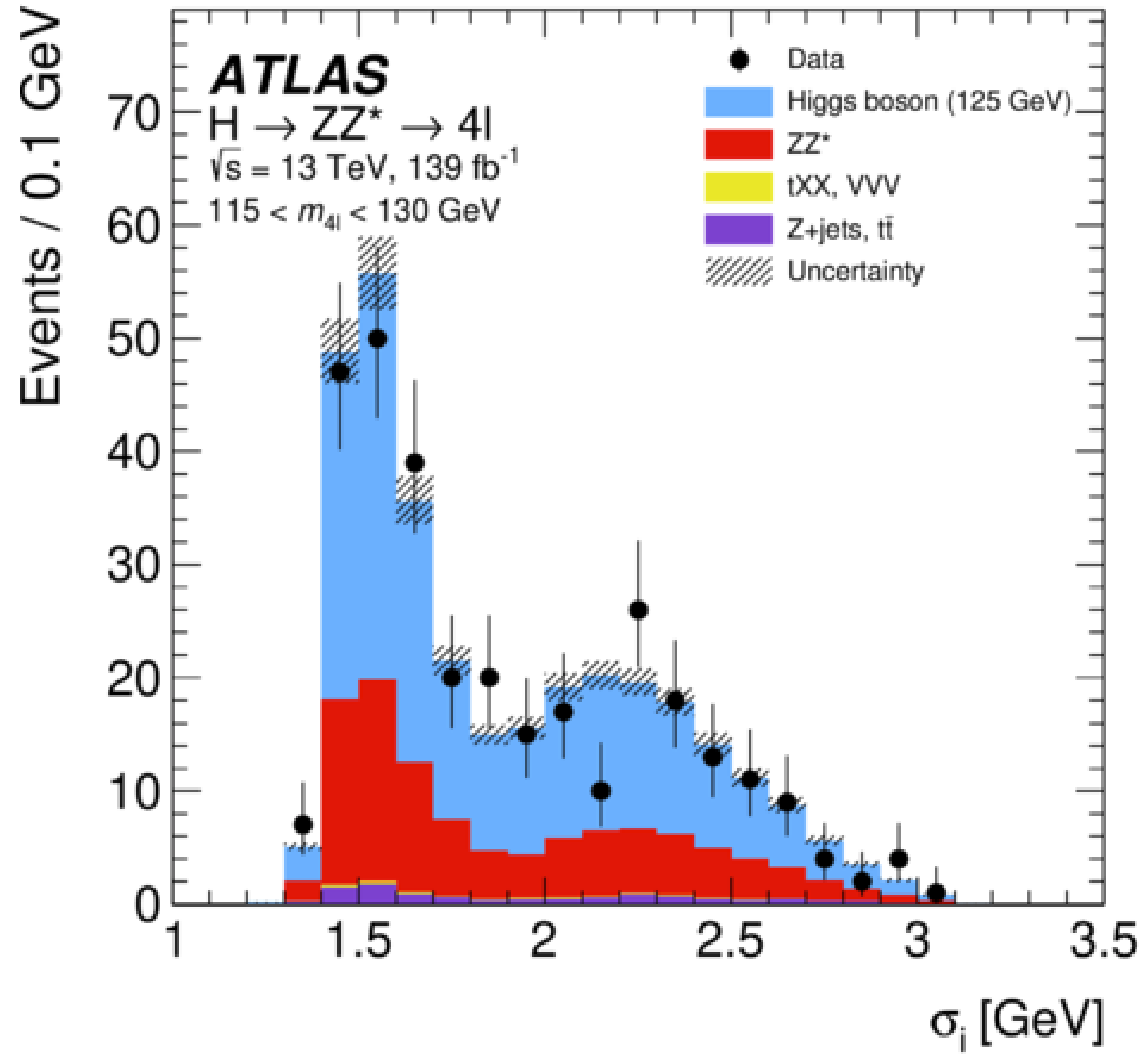
$H \rightarrow ZZ^* \rightarrow llll$ – Mass resolution uncertainty

→ **Recurent Neural Network** + MLP used to estimate the **per-event** quantile of the difference between **truth** and **reconstructed** 4-lepton mass:

$$P_s(m_{4\ell}, D, \sigma | m_H)$$

$$|m_{4l}^{\text{constrained}} - m_{4l}^{\text{truth,born}}|$$





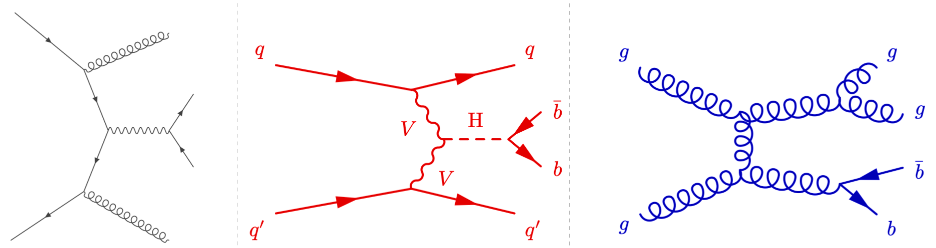
→ Loss function is the square difference between true and predicted m_{4l} quantile:

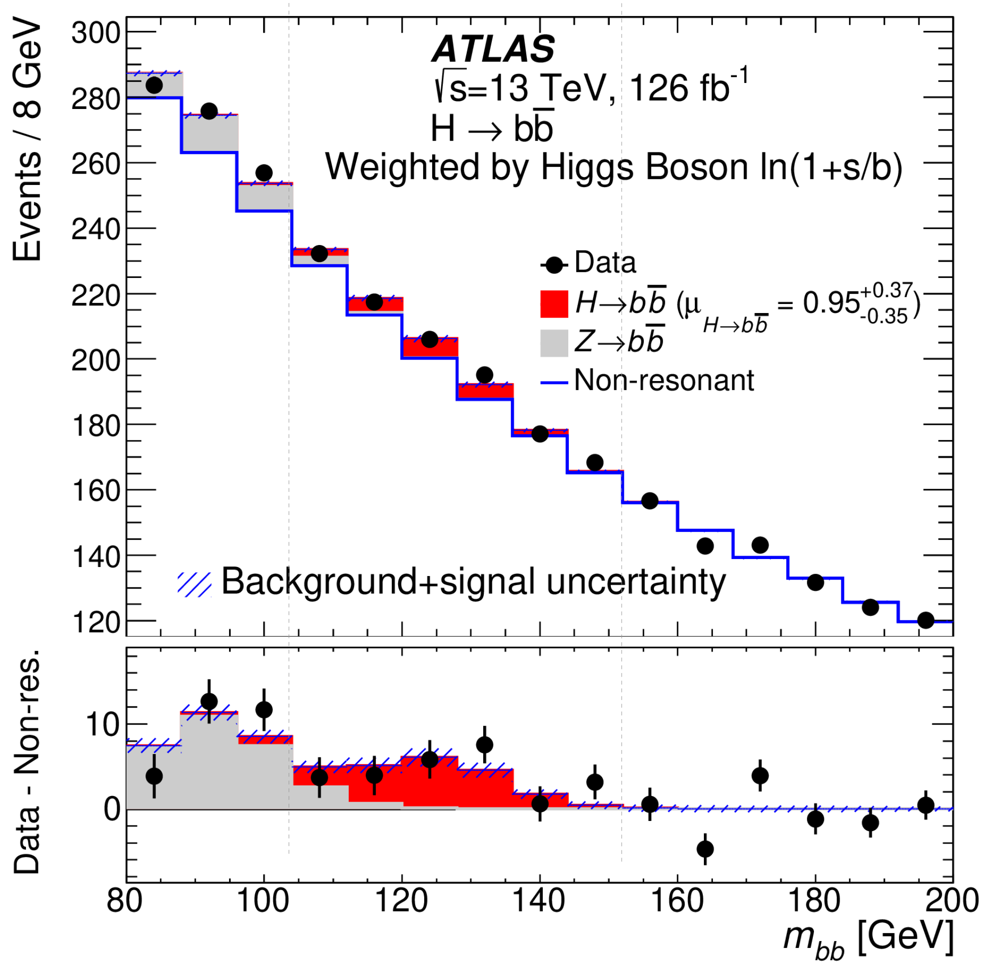
$$\sum_N \max(q \times (\text{true-predicted}), (1-q) \times (\text{true-predicted})) + \frac{1}{w} \sum_N \left(\frac{predicted}{\sigma_{4\ell}^{\text{constrained}}} - -\mu \right)^2$$

VBF H(bb)

→ Search for H(bb) via *vector boson fusion*:

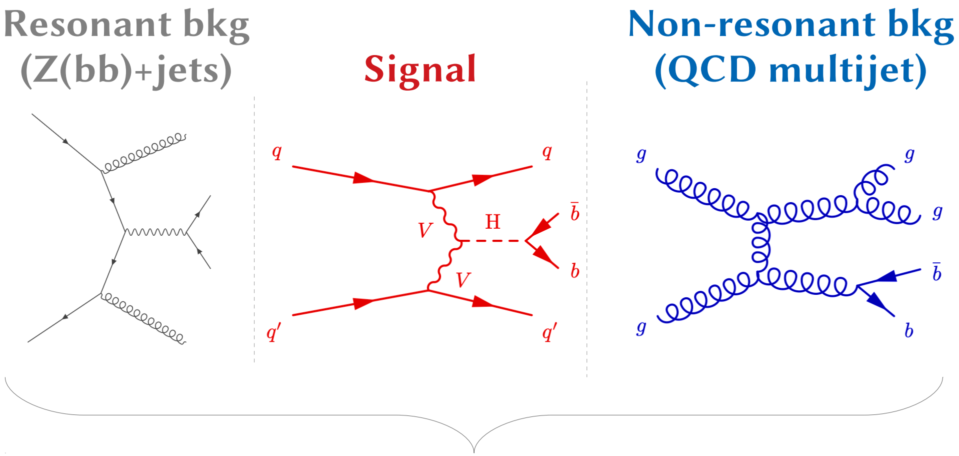
Resonant bkg (Z(bb)+jets) | **Signal** | **Non-resonant bkg (QCD multijet)**





VBF H(bb) – Adversarial Neural Network

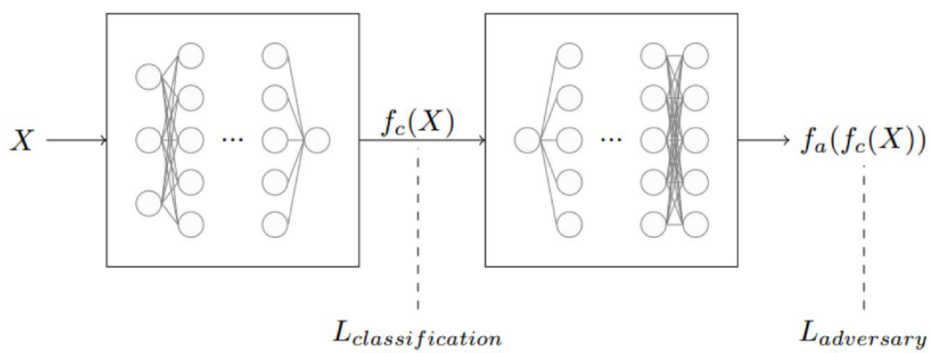
→ Search for H(bb) via *vector boson fusion*:

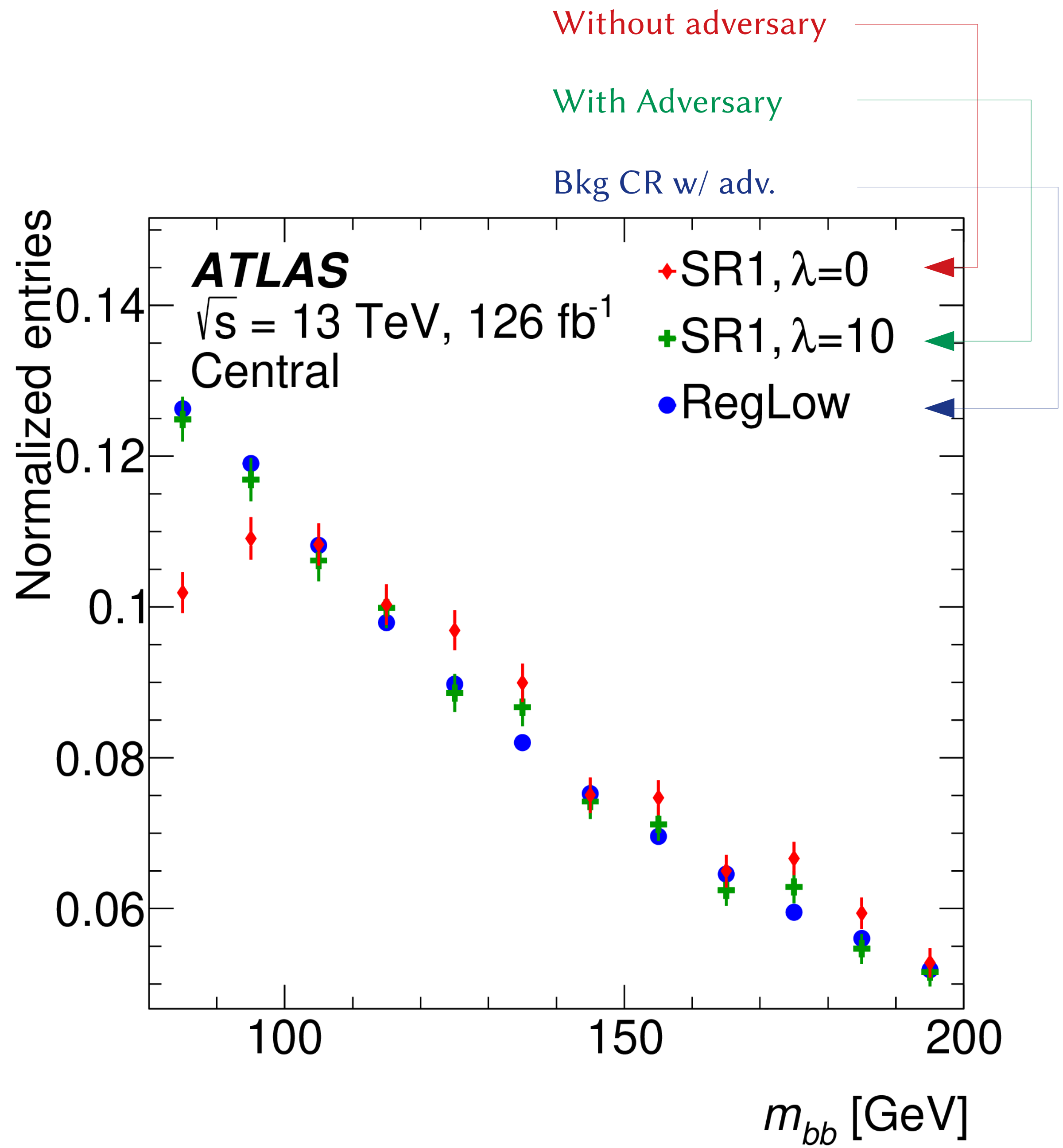


→ Separating background from signal achieved via an *Adversarial Neural Network*:

→ **Classifier** attempts to determine signal vs background

→ **Adversary** attempts to predict the m_{bb} bin of each event:





→ **Adversary penalises** the total loss to ensure classification without *sculpting* m_{bb} of background or signal

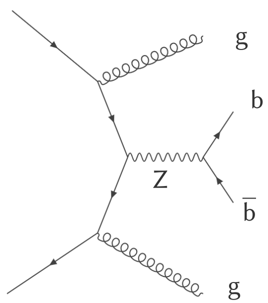
$$L = L_{\text{classifier}} - \lambda L_{\text{adversary}}$$

Object Embedding

- VBF (Hbb) *b*-jet object embedding
- H($\tau\tau$) τ-decay kinematic embedding

VBF H(bb) – Z(bb) object embedding

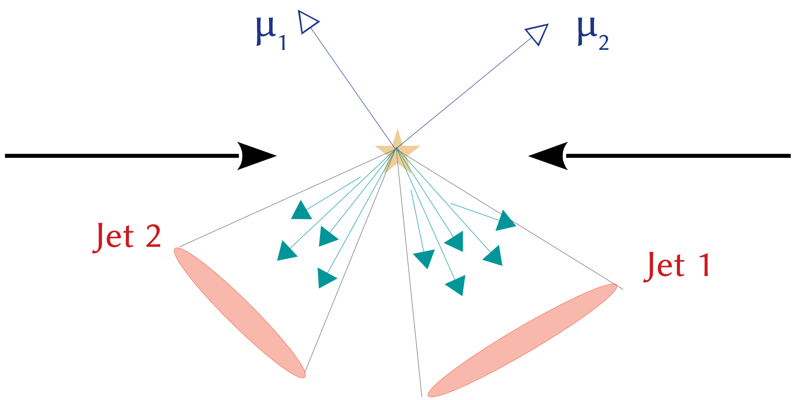
→ **Constraining** resonant Z(bb)+jets **difficult** in fit:



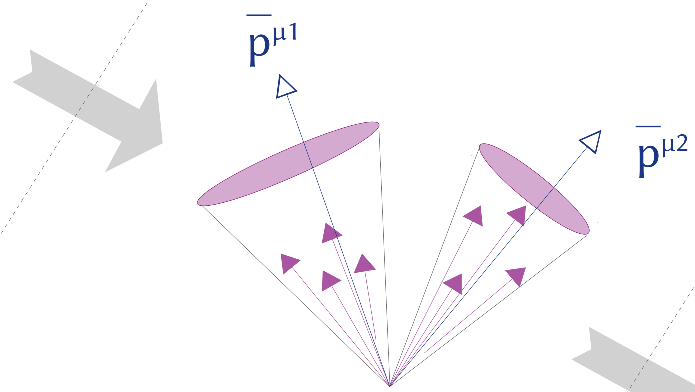
→ Z(bb)+jets & QCD multijet catgorised as bkg by *ANN*

→ Z(bb)+jets poorly modelled by MC in analysis phase space

→ **Solution:** Use Z($\mu\mu$)+jet events from data & replace μ's by simulated b-quark pairs

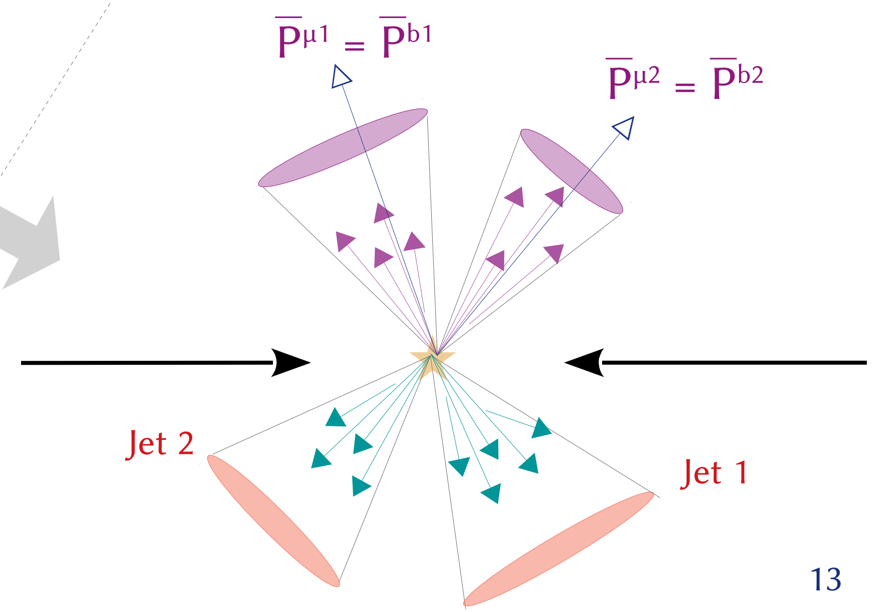


Step 1: Select data events with a Z(μμ) around the pole mass of the Z-boson



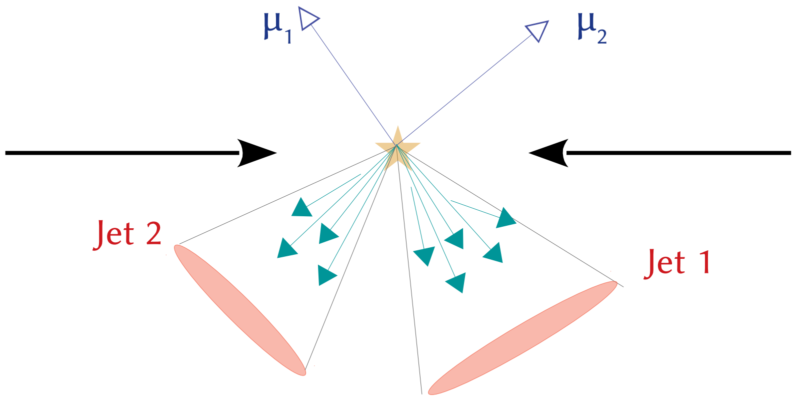
Step 2: Simulate di-b-quark pairs using detector simulation with same muon 4-vector

Step 3: Replace the muons with the simulated jets

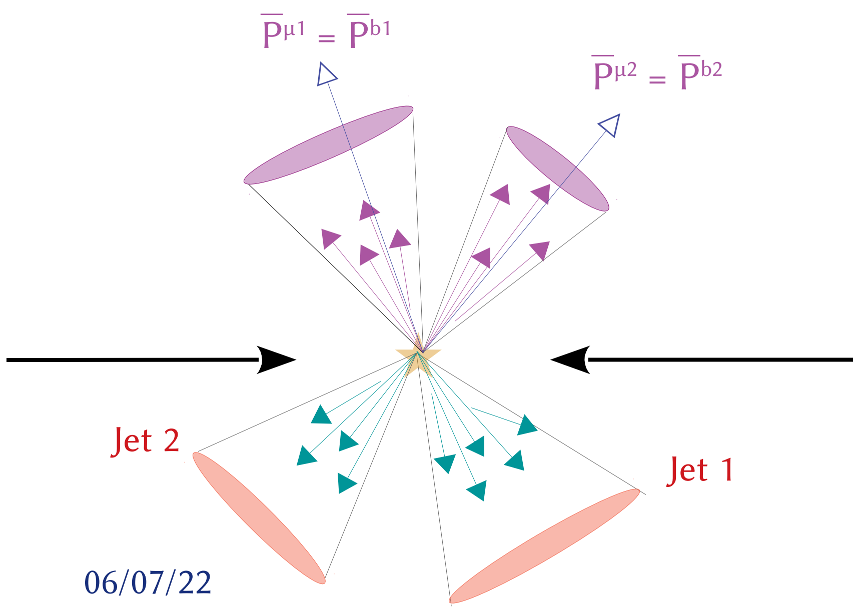


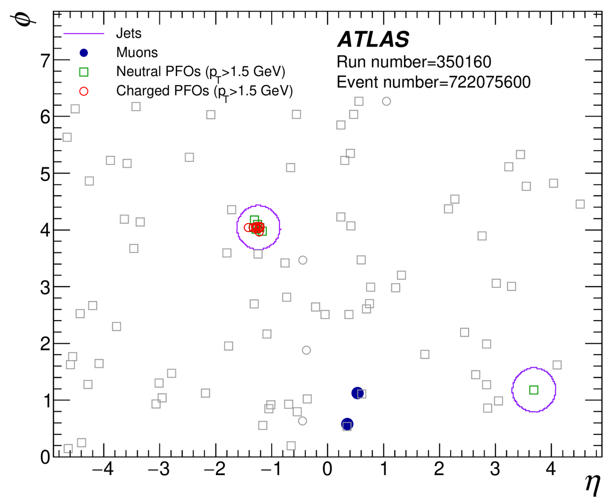
VBF H(bb) – Z(bb) object embedding

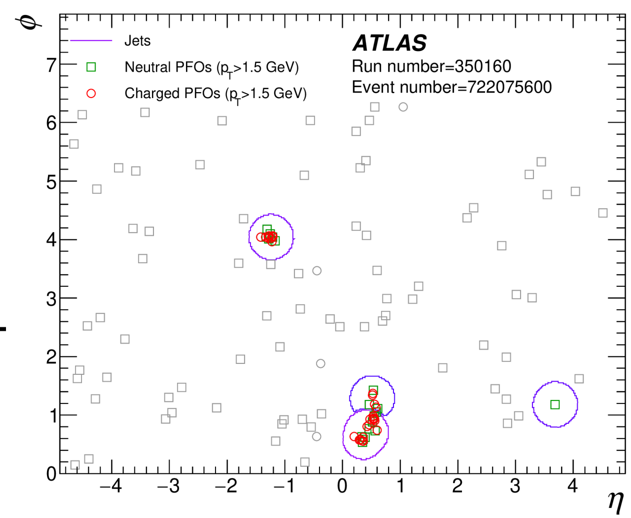
→ How this look? How well does it perform?



Embed MC simulated *b*-jets

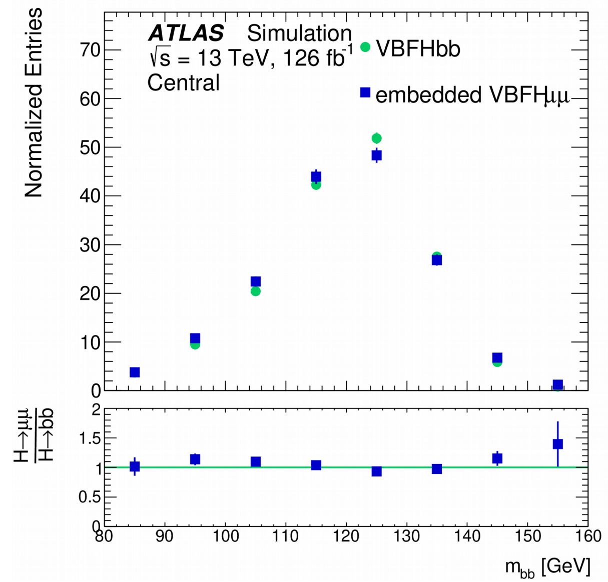


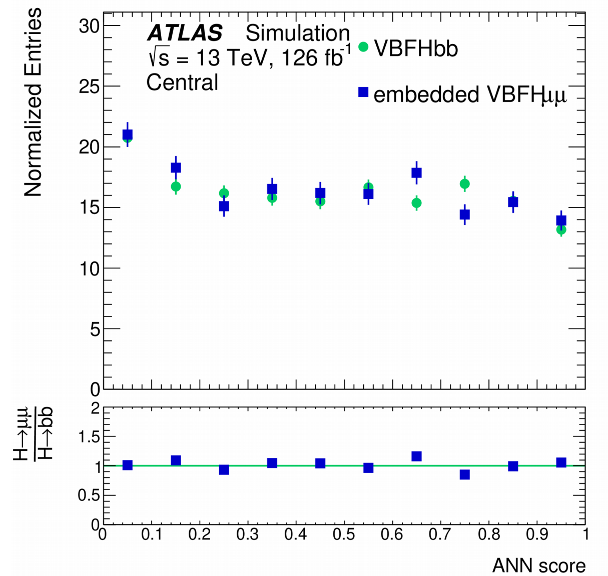




Vaidated using VBF H(μμ)

in which the muons are switched for *b*-jets

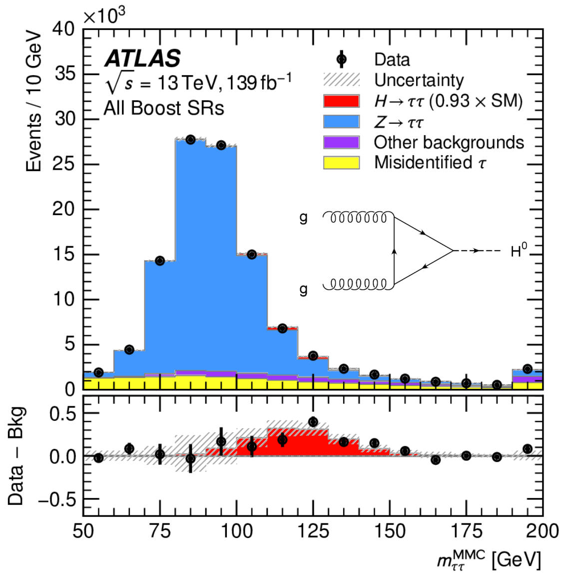




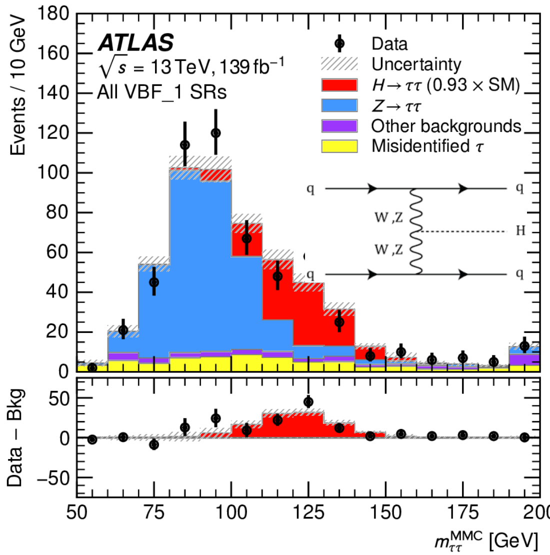
VH+VBF H(ττ) – Z(ττ)+jets background

⟶ Measurement of H(ττ) using all 4 production modes:

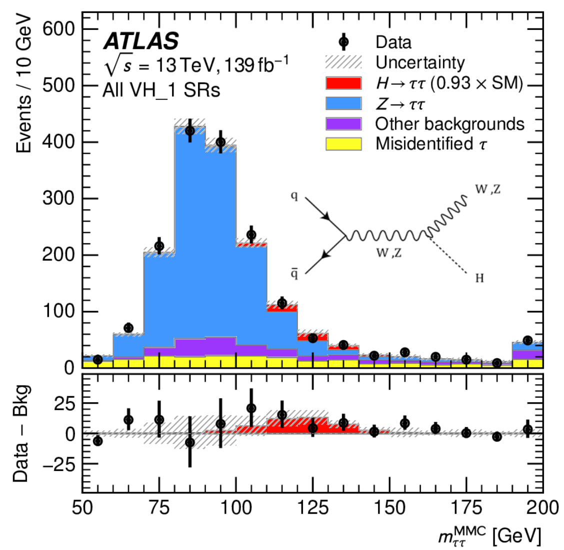
ggF: σ(pp⟶H+X)



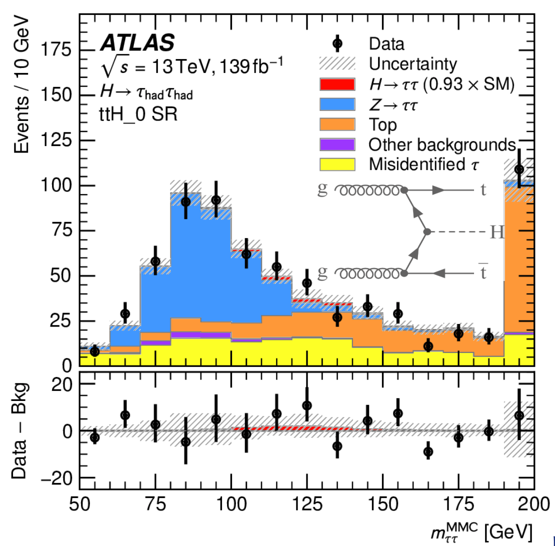
VBF: σ(pp⟶H+X)



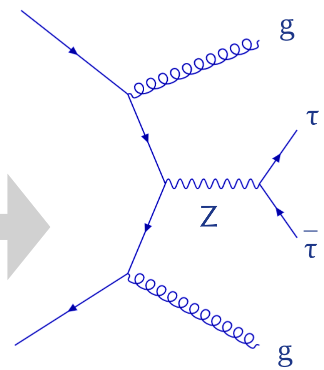
VH: σ(pp⟶H+X)



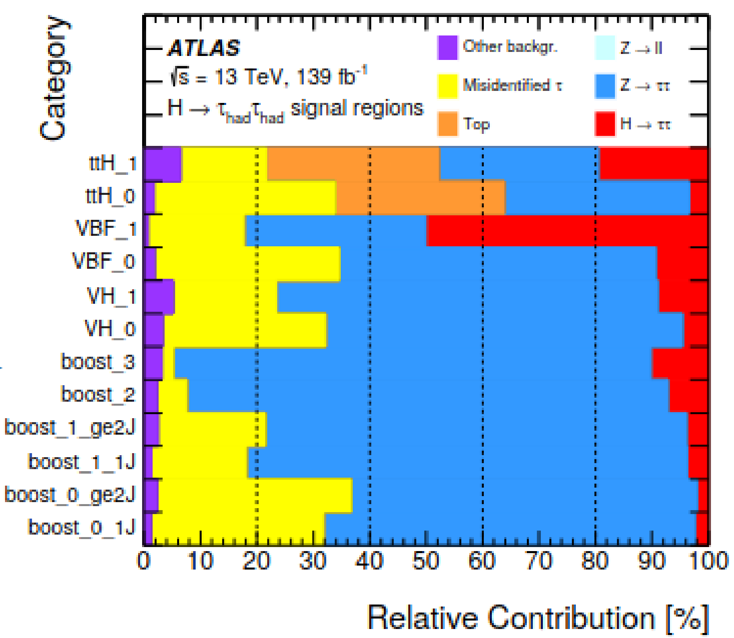
ttH: σ(pp⟶H+X)



⟶ Z(ττ)+jets forms 79% of backround processes:



⟶ Most extreme case it accounts for 90%

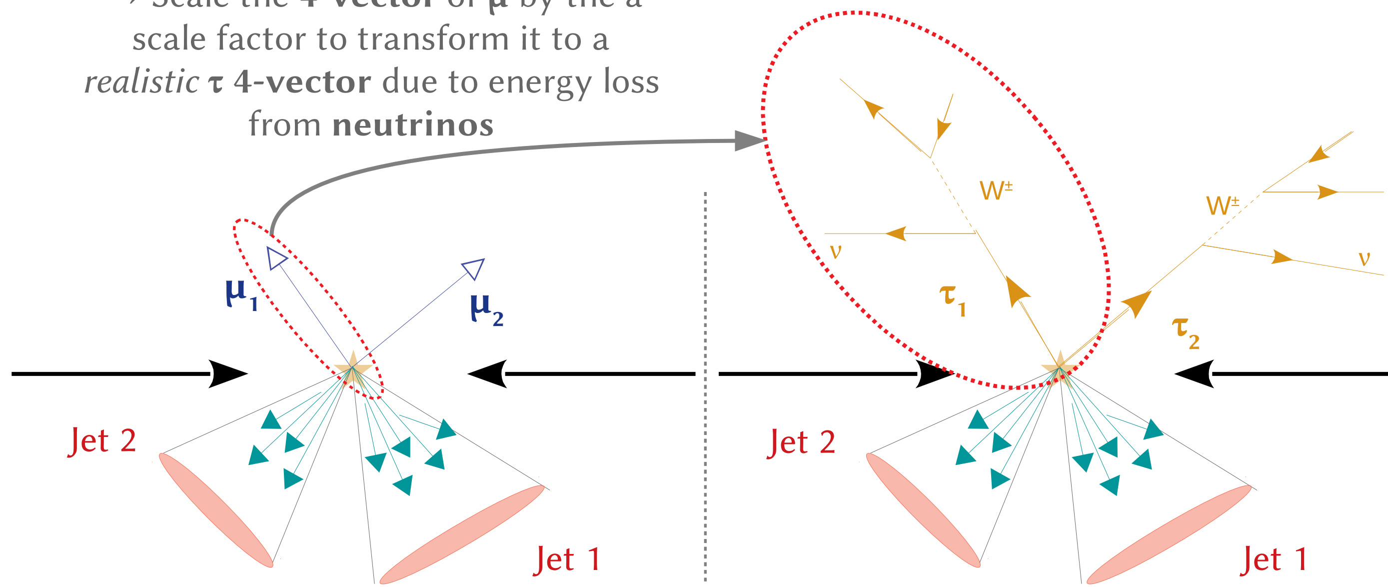


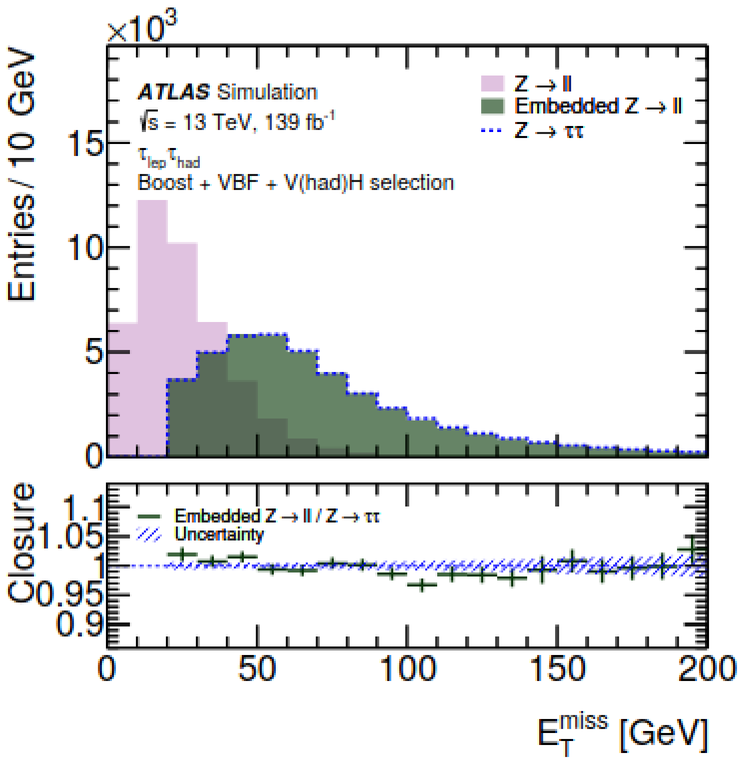
VH+VBF H(ττ) – Z(ττ)+jets background

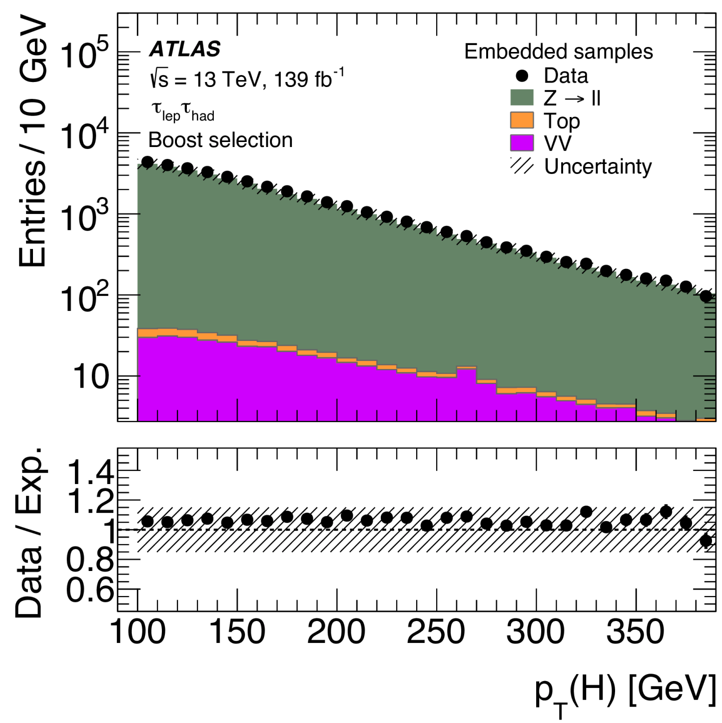
→ Z(ττ)+jets modelled using **kinematic** embedding:

1. Select Z(ee/μμ)+jets events from data
2. Correct for e-/μ-reconstruction/trigger efficiencies
3. Apply a scale to the e/μ four vectors to *mimic* the energy lost due to invisible τ-decay components
4. Apply an event weight to *mimic* the τ-reco/trigger efficiencies for the kinematic topology of the event

→ Scale the **4-vector** of **μ** by the a scale factor to transform it to a *realistic* **τ 4-vector** due to energy loss from **neutrinos**





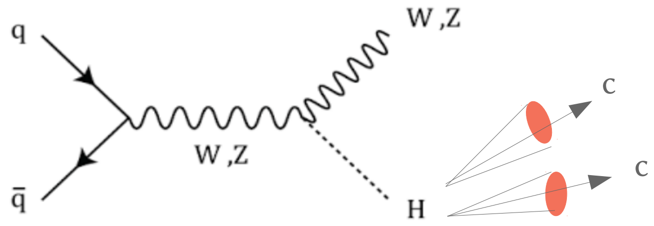


Likelihood Re-weighting

- VH(cc) Jet Flavour Truth Tagging
- VH(bb) Multi-Dimensional MC Uncert. Parameterisation

V(ll,lν,νν)H(cc) – Truth Tagging

→ Search for H(cc) via *associated vector boson* production:

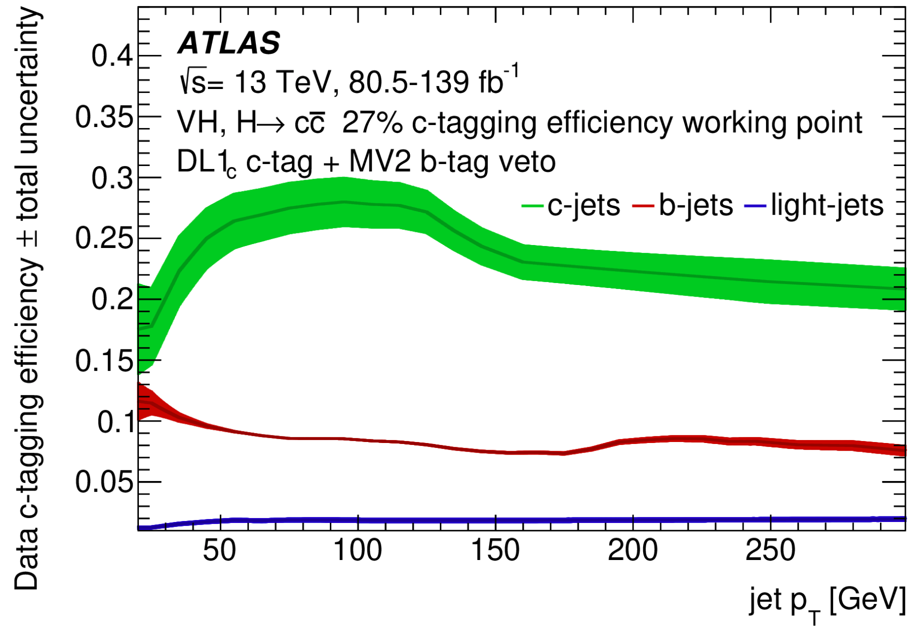


→ Signal strength extracted via profile likelihood fit using:

- $\mathbf{m_{cc}}$: Invariant mass of di-jet system

→ Efficiency of 2 c-jets the product of single jet tagging efficiencies ε^f_i:

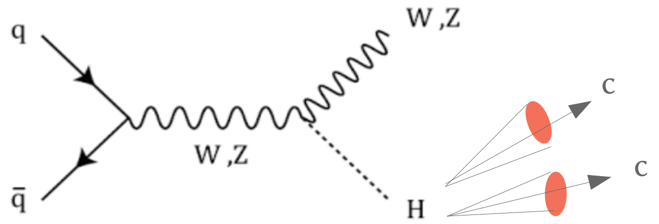
$$\varepsilon\left(\binom{m}{n}_i, x, f\right) = \prod_{j\in n} \epsilon^f_x(j) \approx 27\%, f = c\text{-}jet \longrightarrow \text{x2 } c\text{-jets} \sim 7\%$$



→ **Lot of Monte Carlo is lost as a result of the tagging efficiencies!**

V(ll,lν,νν)H(cc) – Truth Tagging

→ Search for H(cc) via *associated vector boson* production:



→ Signal strength extracted via profile likelihood fit using:
- **m_{cc}** : Invariant mass of di-jet system

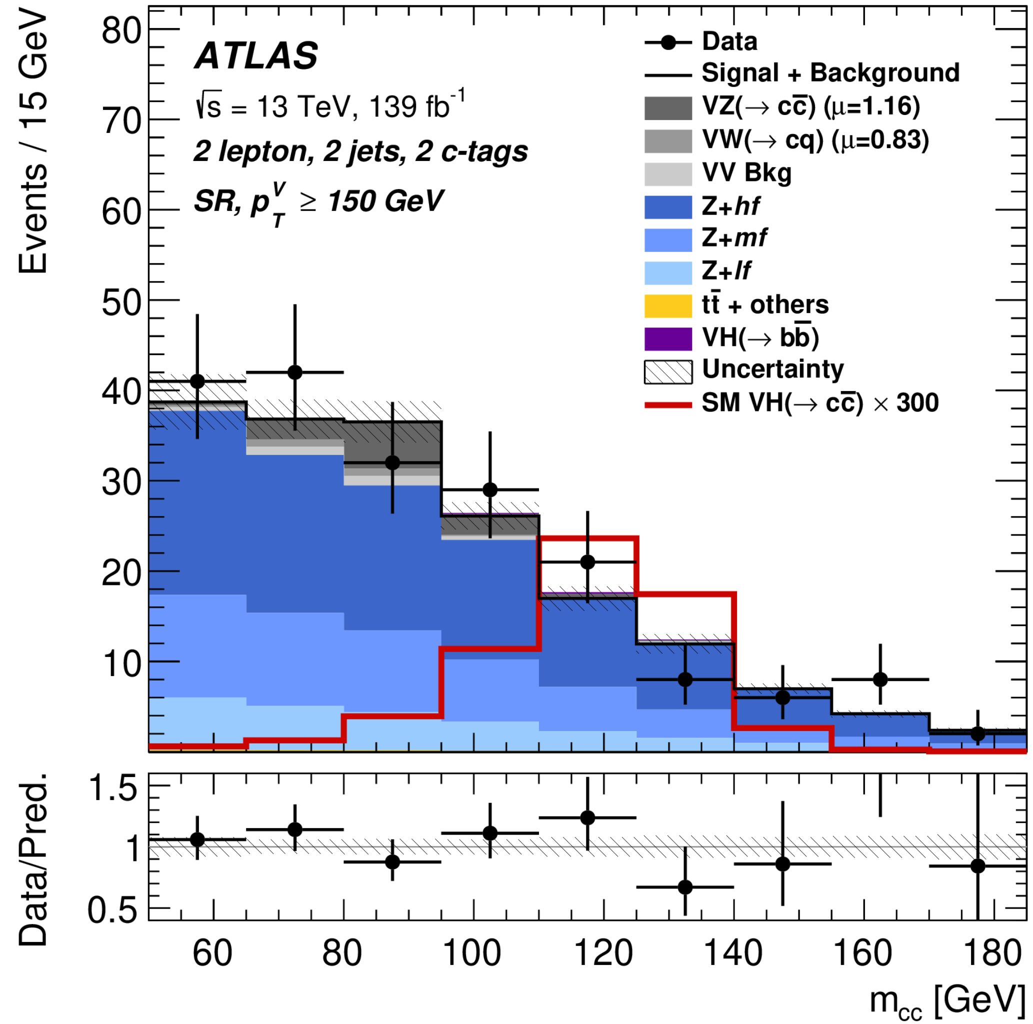
→ Efficiency of 2 c-jets the product of single jet tagging efficiencies ε^f_i:

$$\varepsilon\left(\binom{m}{n}_i, x, f\right) = \prod_{j \in n} \epsilon^f_x(j) \approx 27\%, f = c\text{-}jet \longrightarrow \text{x2 } c\text{-jets} \sim 7\%$$

→ **Solution:** Select all events, but weight event based on probability of tagging *m-jets* out of *n* total jets:

$$P_i(x) = \frac{\varepsilon\left(\binom{m}{n}_i, x\right) \cdot \varepsilon_{in}\left(\overline{\binom{m}{n}}_i, x\right)}{w_{TT}}$$

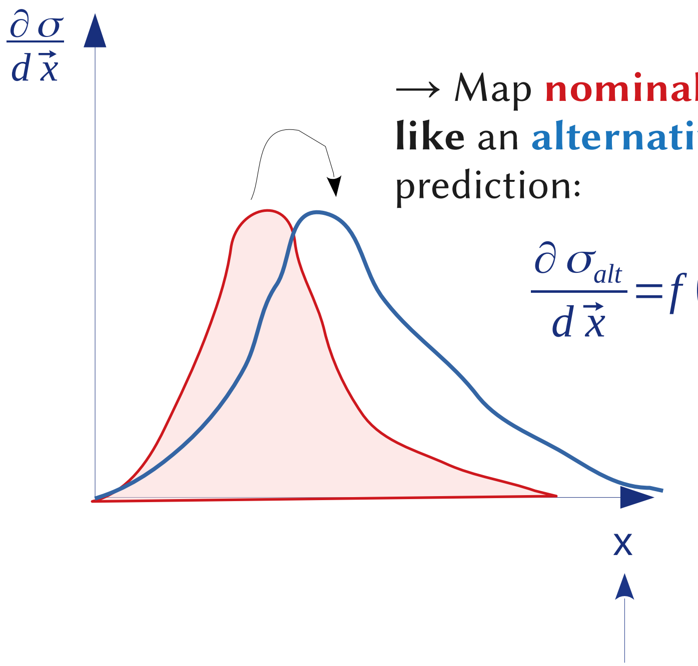
$$w_{TT}(x) = \sum_i^{|\binom{m}{n}|} \varepsilon\left(\binom{m}{n}_i, x\right) \cdot \varepsilon_{in}\left(\overline{\binom{m}{n}}_i, x\right)$$



→ **Truth tagging improves limit on $\mu_{VH(cc)}$ by ~ 10%**

VH(bb) – MC Systematic Uncertainties

→ MC uncertainties from **MC model changes** (e.g. *matrix element generator models)* induce kinematic variations:



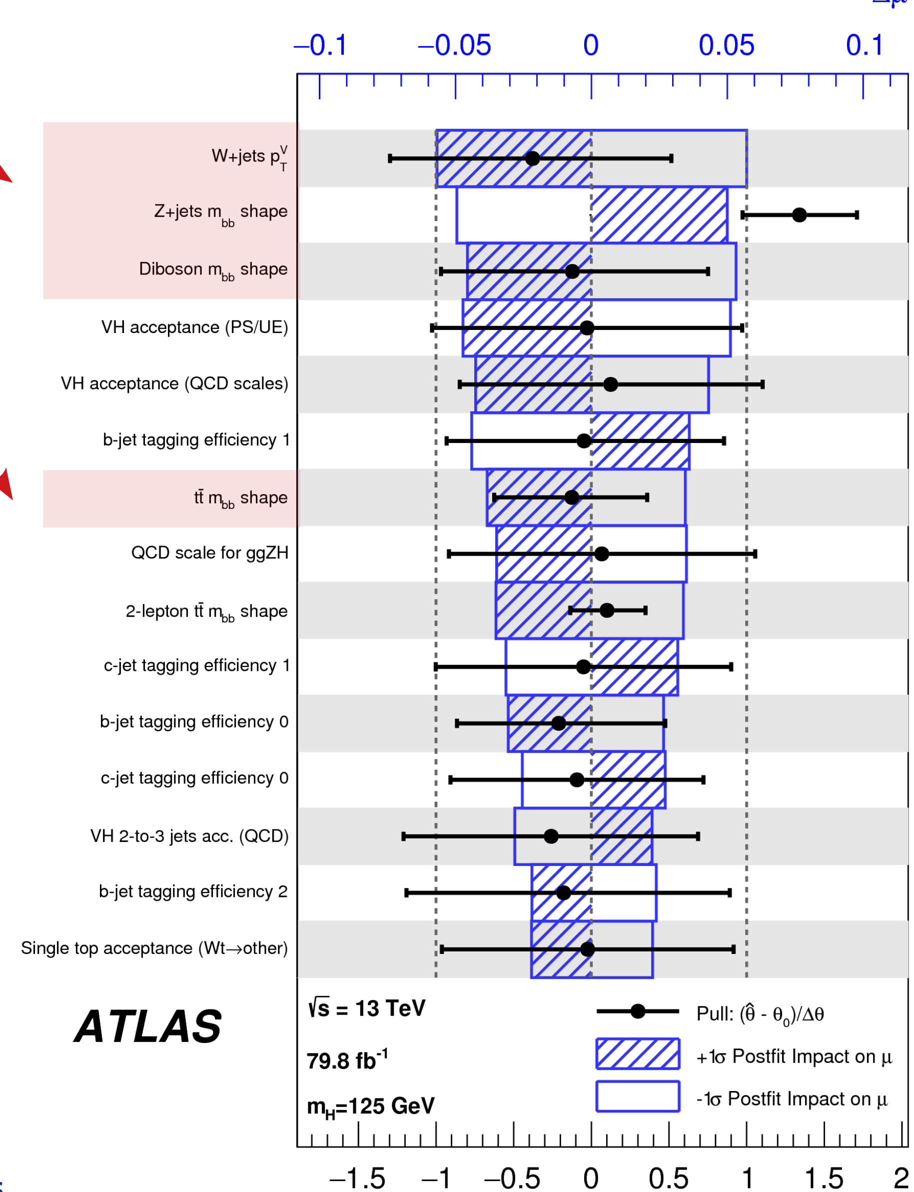
→ Map **nominal** to **look-like** an **alternative** MC prediction:

$$\frac{\partial \sigma_{alt}}{d\vec{x}} = f(\vec{x}) \cdot \frac{\partial \sigma_{nom}}{d\vec{x}}$$

→ Analysis feature space is **13-dimensional**

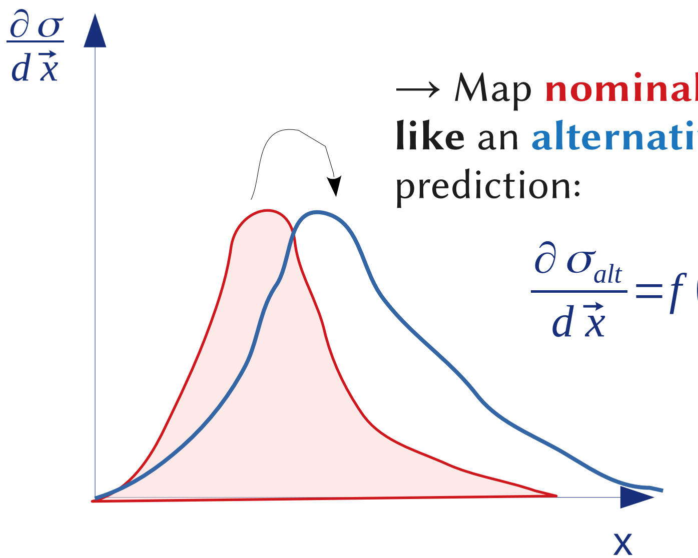
VH(bb) Observation Paper:

HIGG-2018-04



VH(bb) – MC Systematic Uncertainties

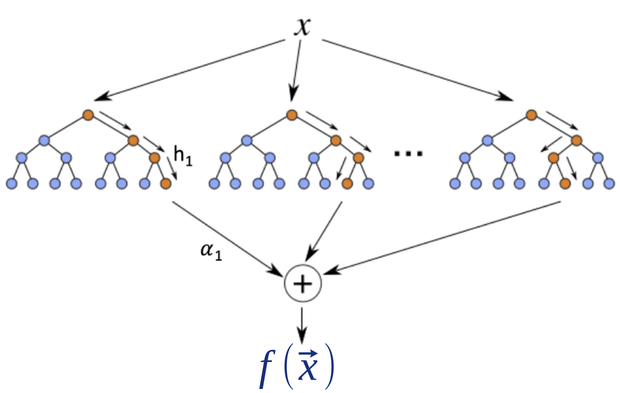
→ MC uncertainties from **MC model changes** (e.g. *matrix element generator models)* induce kinematic variations:

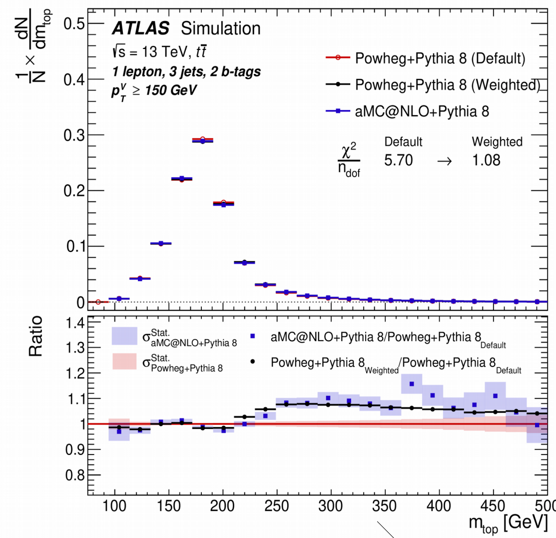


→ Map **nominal** to **look-like** an **alternative** MC prediction:

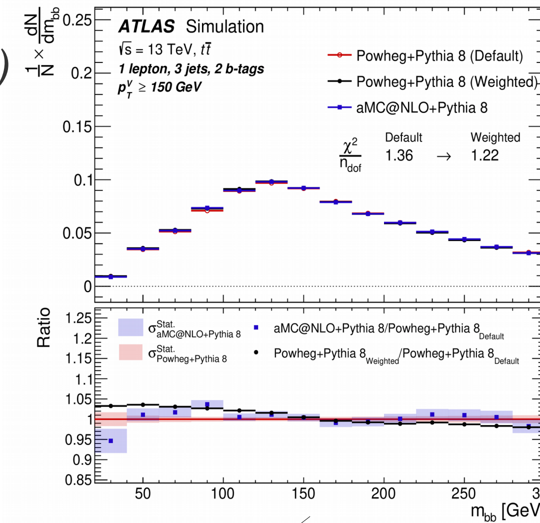
$$\frac{\partial \sigma_{alt}}{d\vec{x}} = f(\vec{x}) \cdot \frac{\partial \sigma_{nom}}{d\vec{x}}$$

→ $f(\bar{x})$ = a surjective map built from a BDT that simultaneous maps all the features of **nominal** ($\bar{x}_{nom}$) to the **alternative** MC ($\bar{x}_{alt}$)

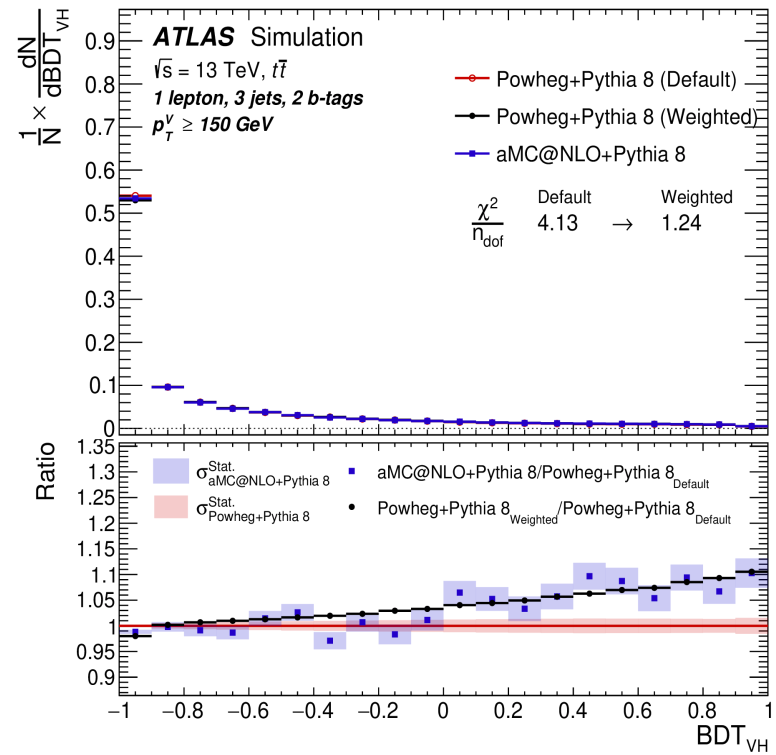




$\otimes f(\bar{x})$



$\otimes f(\bar{x})$



$\otimes f(\bar{x})$

Conclusion

→ Pushing scope of Higgs analyses key to understanding the SM and potentially disovering BSM physics

→ Detector development life cycles are often **10-20** years long

→ Development life cycles of new machine learning, statistical inference models, and accelerator techniques often **2-3 years**

→ **Squeezing** every ounce of information out of the **already** collected data is **unmined gold…**

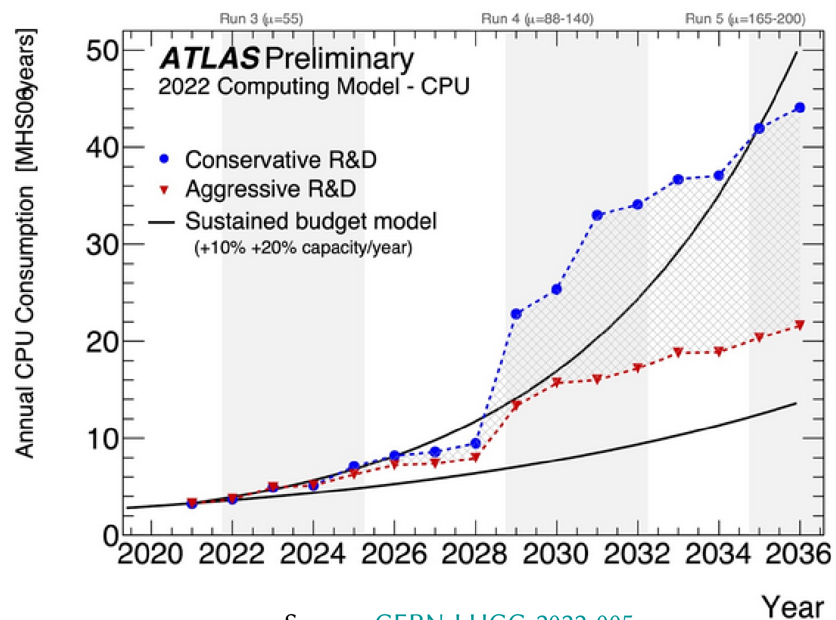
→ For more information about the relevant measurements see:

- **H(γγ):** D.Mungo (Thurs. 09:00)
- **H(ZZ):** G.Barone (Thurs. 11:45)
- **VH(bb/cc):** G.Di Gregorio (Thurs. 09:15)
- **VBF H(bb):** G.Di Gregorio (Thurs. 09:15)
- **H(ττ):** G.Di Gregorio (Thurs. 09:15)

Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \rightarrow \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \rightarrow \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \rightarrow bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \rightarrow bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \rightarrow bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

Source: A. Radovic et al., Nature 560(2018) no. 7716,41

Machine Learning algorithms have helped push the discovery potential of analyses



Source: CERN-LHCC-2022-005

Large data/MC data volumes pose problem many future HL-LHC analyses

Optional

Necessity

Backup

Context

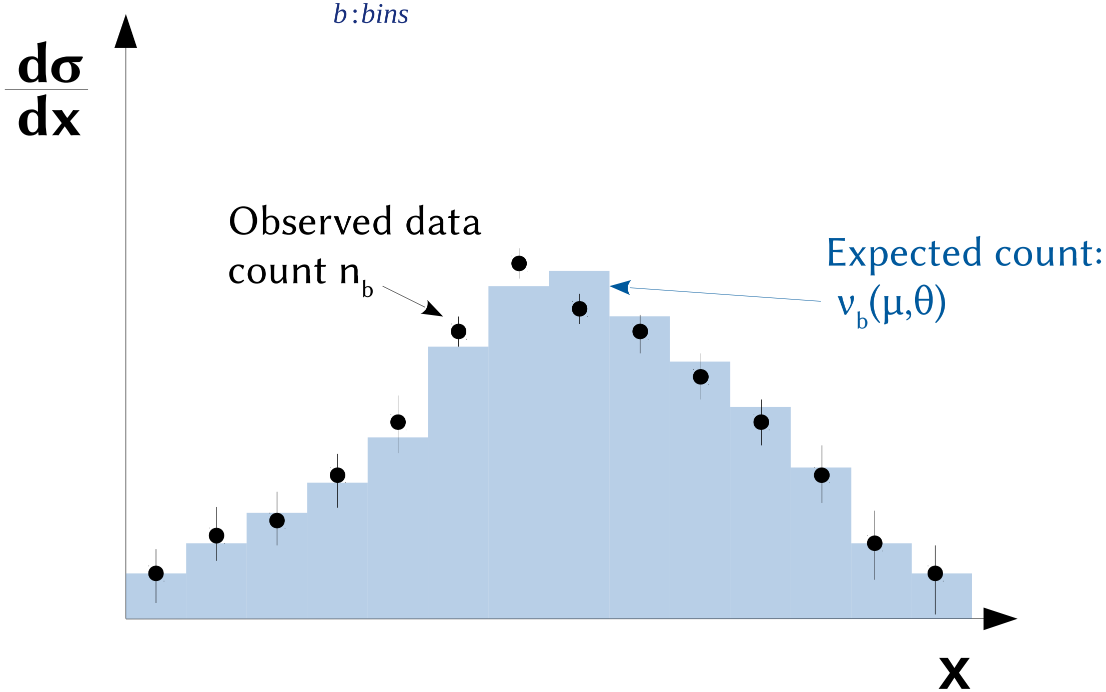
→ **Goal:** Summarise new/uncommon analysis techniques used by ATLAS data analyses

Context

→ **Goal:** Summarise new/uncommon analysis techniques used by ATLAS data analyses

→ **Data Analysis 101:** At its core ATLAS data analyses are *counting* experiments interpreted predominantly in a *frequentist* likelihood paradigm:

$$L(n|\vec{\mu},\vec{\theta}) = \prod_{b:bins} P(n_b|\nu_b(\vec{\mu},\vec{\theta})) \cdot \prod_{\theta}^{N_\theta} f_\theta(\tilde{\theta}|\theta) \cdot \prod_{b:bins} P(\nu_b(\vec{\mu},\vec{\theta})|\lambda_b,\tau_b)$$

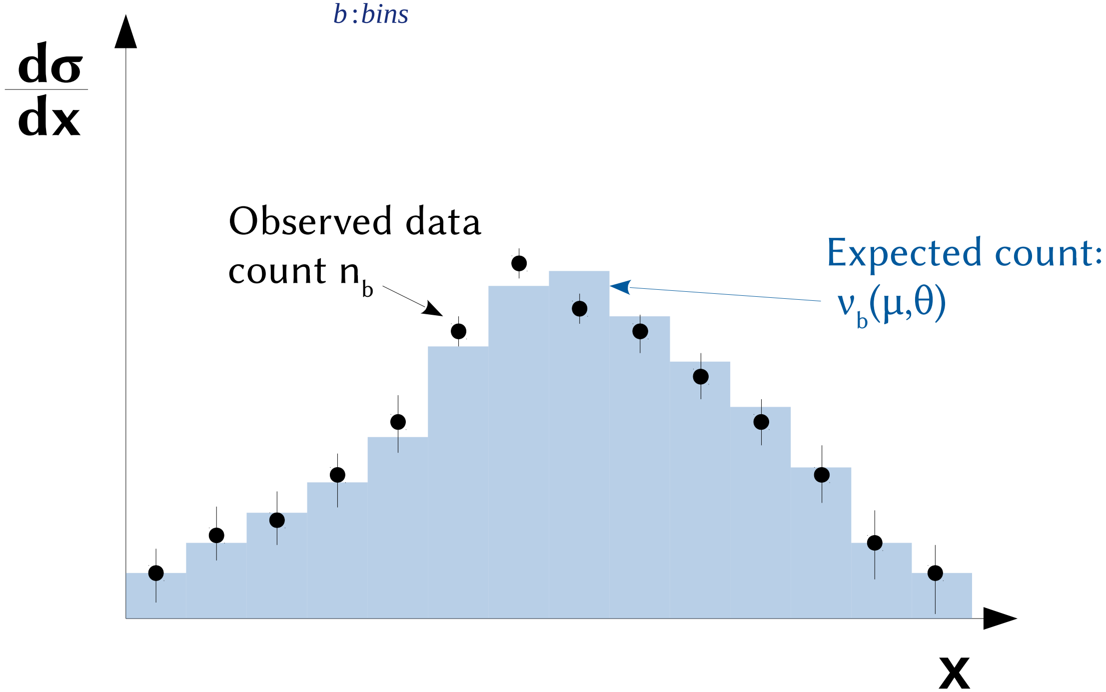


Context

→ **Goal:** Summarise new/uncommon analysis techniques used by ATLAS data analyses

→ **Data Analysis 101:** At its core ATLAS data analyses are *counting* experiments interpreted predominantly in a *frequentist* likelihood paradigm:

$$L(n|\vec{\mu},\vec{\theta}) = \prod_{b:bins} P(n_b|\nu_b(\vec{\mu},\vec{\theta})) \cdot \prod_{\theta}^{N_\theta} f_\theta(\tilde{\theta}|\theta) \cdot \prod_{b:bins} P(\nu_b(\vec{\mu},\vec{\theta})|\lambda_b,\tau_b)$$



→ **Why todays talk?**

Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \to \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \to \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \to bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \to bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \to bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

Source: A. Radovic et al., Nature 560(2018) no. 7716,41

Machine Learning algorithms have helped push the discovery potential of analyses

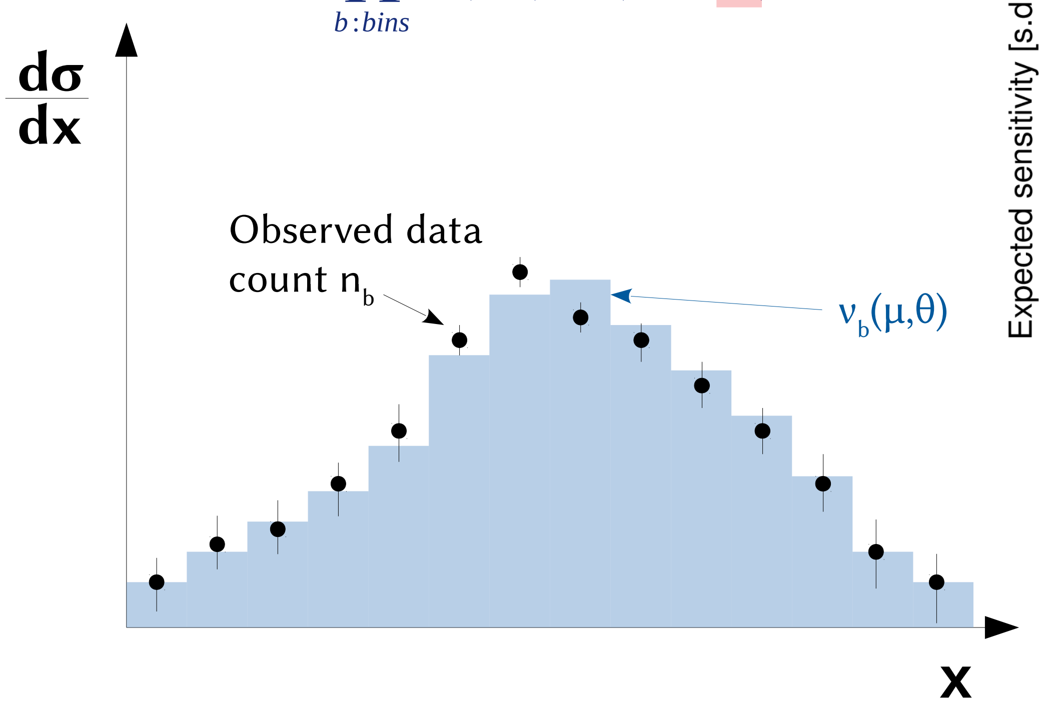
Context

→ **Goal:** Summarise new/uncommon analysis techniques used by ATLAS data analyses

→ **Data Analysis 101:** At its core ATLAS data analyses are *counting* experiments interpreted predominantly in a *frequentist* likelihood paradigm:

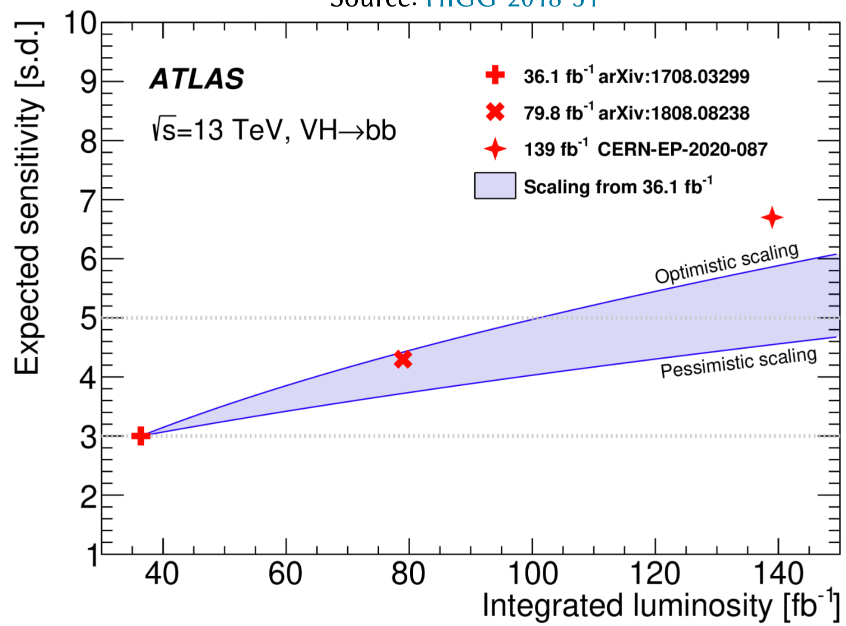
$$L(n|\vec{\mu},\vec{\theta}) = \prod_{b:bins} P\left(\overset{+}{n_b} \,\middle|\, \overset{+}{\nu_b(\vec{\mu},\vec{\theta})}\right) \cdot \prod_{\theta}^{N_\theta} \overset{-}{f_\theta(\tilde{\theta}|\theta)}$$

$$\cdot \prod_{b:bins} P\left(\nu_b(\vec{\mu},\vec{\theta}) \,\middle|\, \lambda_b, \overset{-}{\tau_b}\right)$$



→ **Why todays talk?**

Source: HIGG-2018-51



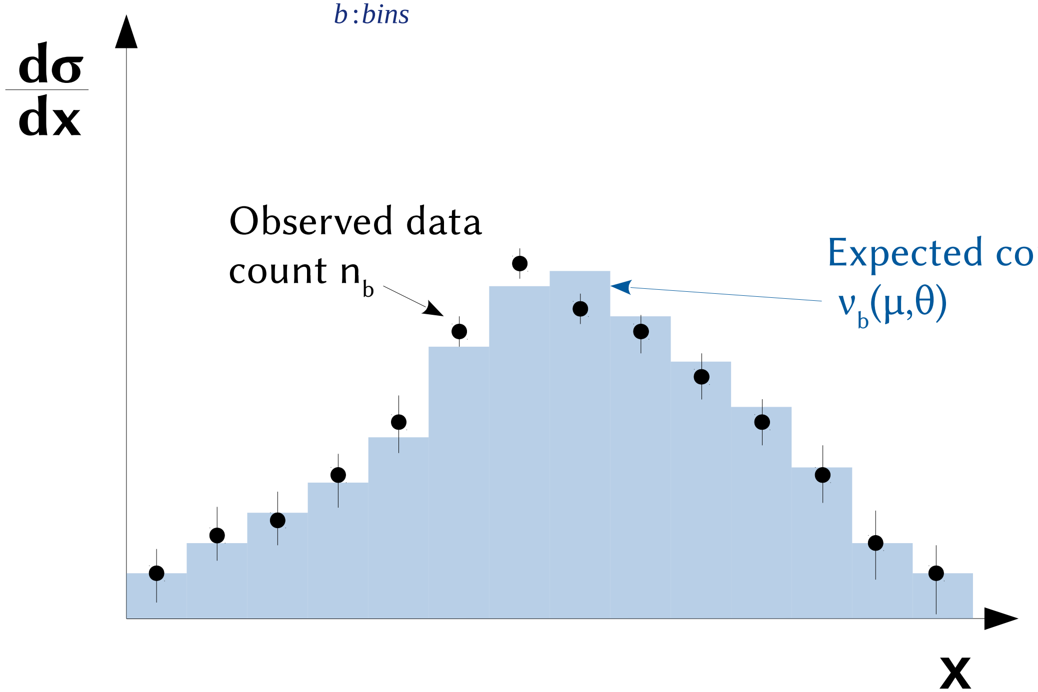
ML models, specialised jet tagging, opitimised mass resolution fitting etc... needed to make specific Higgs decay modes viable

V

→ **Goal:** Summarise new/uncommon analysis techniques used by ATLAS data analyses

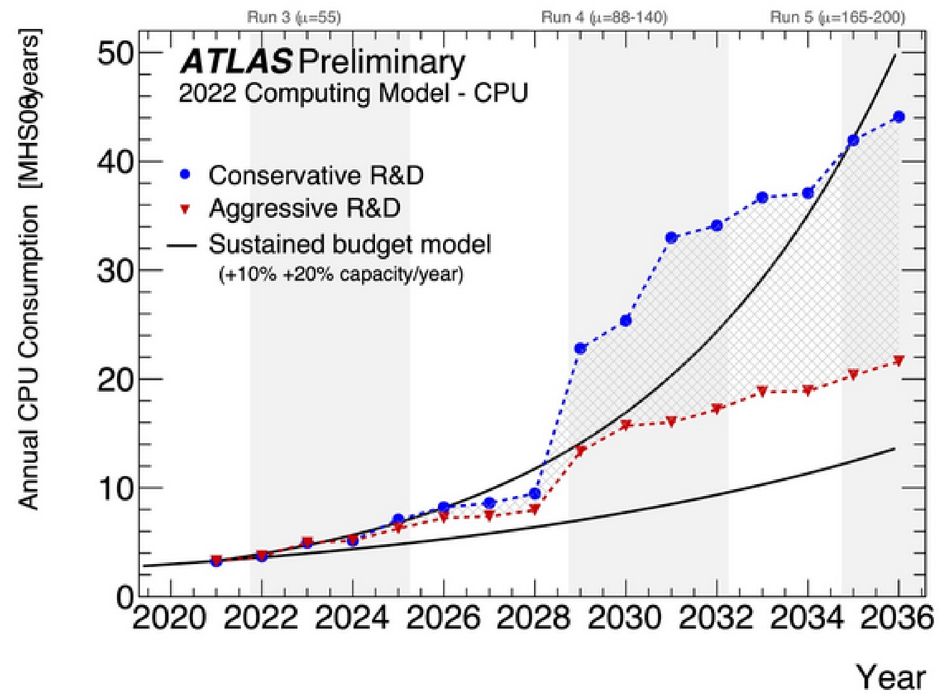
→ **Data Analysis 101:** At its core ATLAS data analyses are *counting* experiments interpreted predominantly in a *frequentist* likelihood paradigm:

$$L(n|\vec{\mu},\vec{\theta}) = \prod_{b:bins} P(n_b|\overset{+}{\boxed{\nu_b(\vec{\mu},\vec{\theta})}}) \cdot \prod_{\theta}^{N_\theta} \overset{?}{\boxed{f_\theta(\tilde{\theta}|\theta)}}$$
$$\cdot \prod_{b:bins} P(\nu_b(\vec{\mu},\vec{\theta})|\lambda_b, \overset{-}{\boxed{\tau_b}})$$



→ **Why todays talk?**

Source: CERN-LHCC-2022-005



Numerical techniques to address data/MC data volumes that pose problem for some current and many future HL-LHC analyses

$H(\gamma\gamma)$

H(γγ) – Multi-Class D-optimality

→ **Goal:** Single metric for optimisation that:

- → Sensitivity aware due to statistical errors
- → Reduces STXS bin correlations

→ **D-optimality:** Determinant of the covariance matrix (C_{xx}):

$$D_{opt} = \frac{1}{2} \cdot \log \frac{|C_{exp} + C_{theo.}|}{|C_{exp}|}$$

→ **Algorithm:** Conduct a 1-bin counting experiment by assigning events to pseudo-STXS bins according based on the maximum multi-class score:

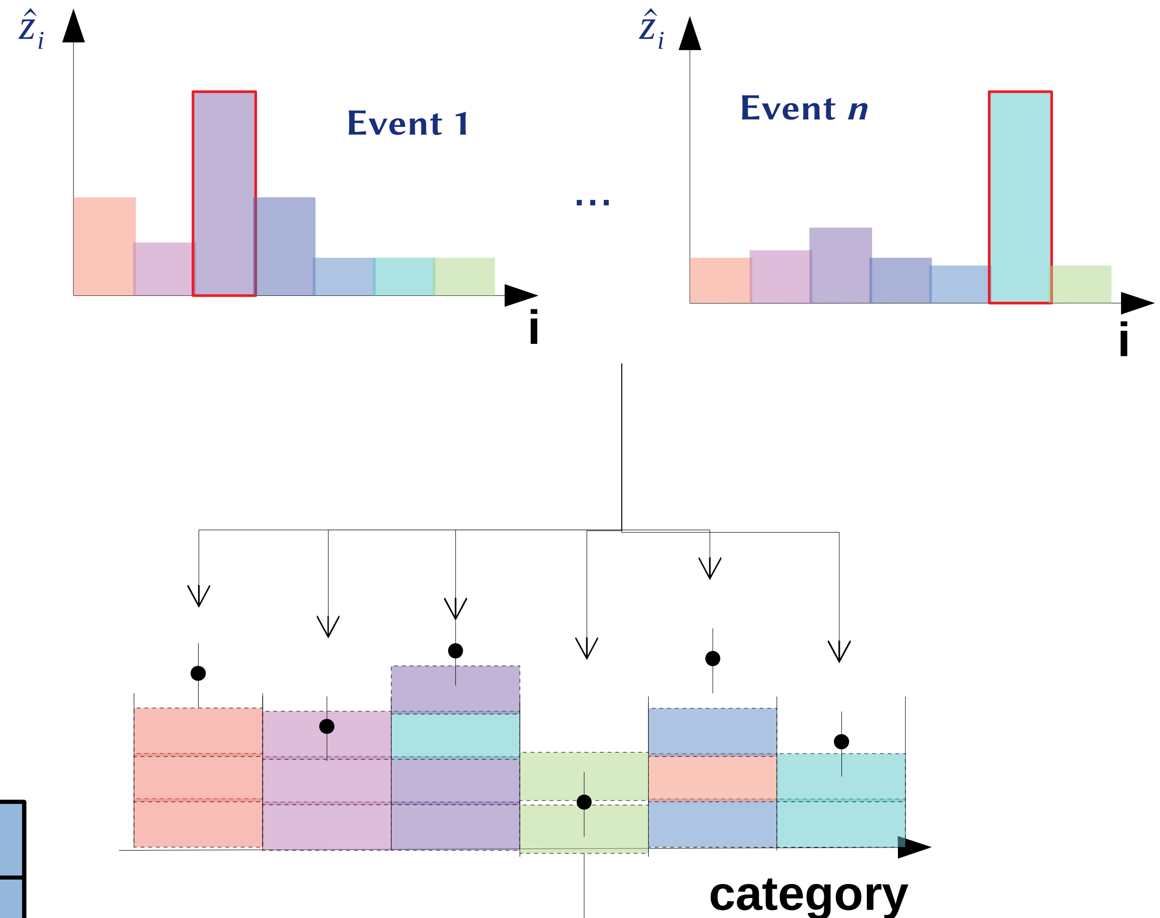
$$max(\hat{z}_i = w_i z_i)$$

Step 3: Minimise covariance matrix by changing BDT transformation weights:

$$w_i$$

1					
0.2	1				
0.11	0.3	1			
0.01	0.21	0.34	1		
0.08	0.5	0.2	0.4	1	
0.87	0.16	0.29	0.36	0.24	1

Step 1: Assign event to a STXS category



Step 2: Perform 1-bin Asimov fit obtaining a covariance matrix

$H(ZZ) \rightarrow 4l$

$H \to ZZ^* \to llll$ – Mass resolution uncertainty

→ New Higgs mass (m_H) measurement using $H \to ZZ^* \to l^+l^-l^+l^-$

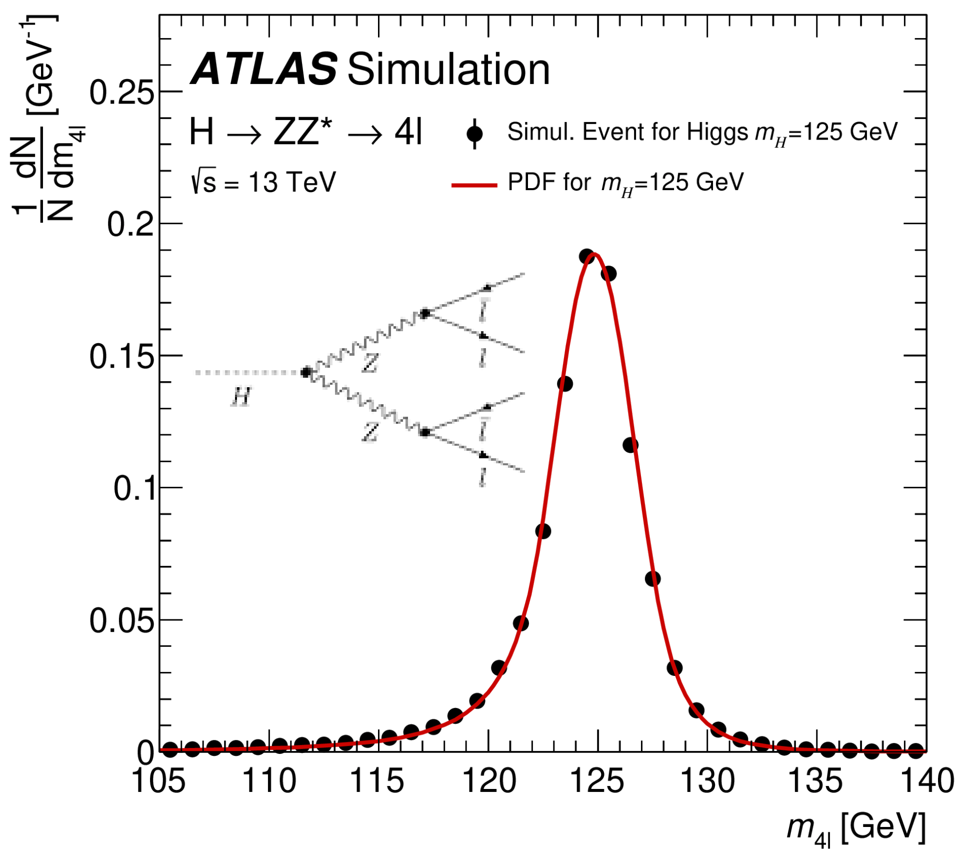
→ Unbinned maximum likelihood fit to data

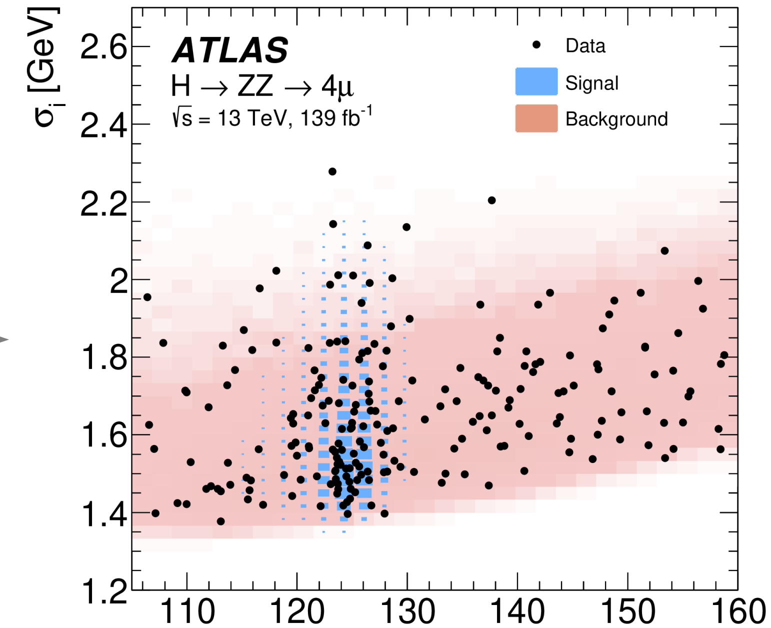
$$\mathcal{L}(m_H|\boldsymbol{x}) = \mathcal{L}(m_H|m_{4\ell}, D, \sigma) = \prod_i P(m^i_{4\ell}, D^i, \sigma^i|m_H)$$

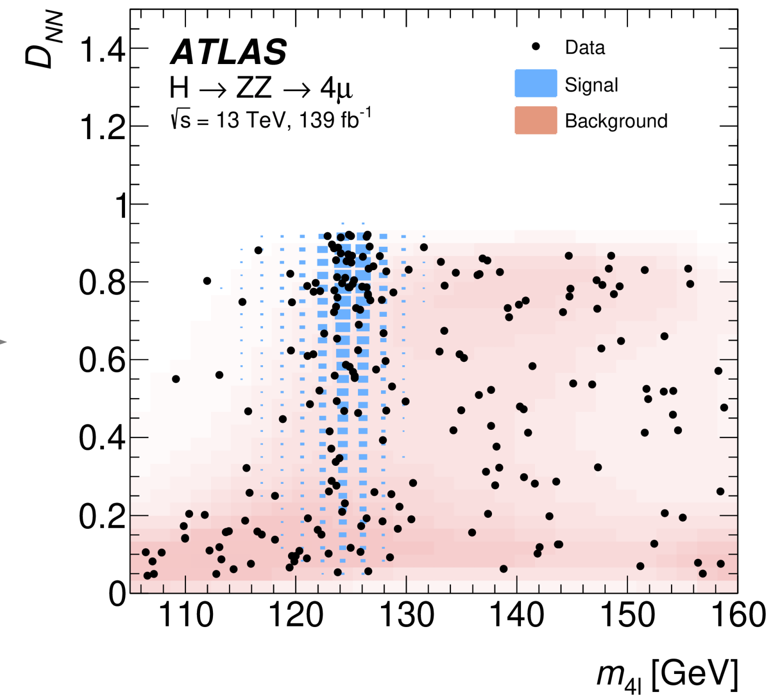
→ Signal *probability density function* (pdf) conditional on:

1. $\mathbf{m_H}$ = Higgs mass
2. $\mathbf{D_{NN}}$ = Dense Neural Network classifier
3. $\boldsymbol{\sigma_i}$ = Event level m_{4l} resolution

$$P_s(m_{4\ell}, D, \sigma|m_H)$$



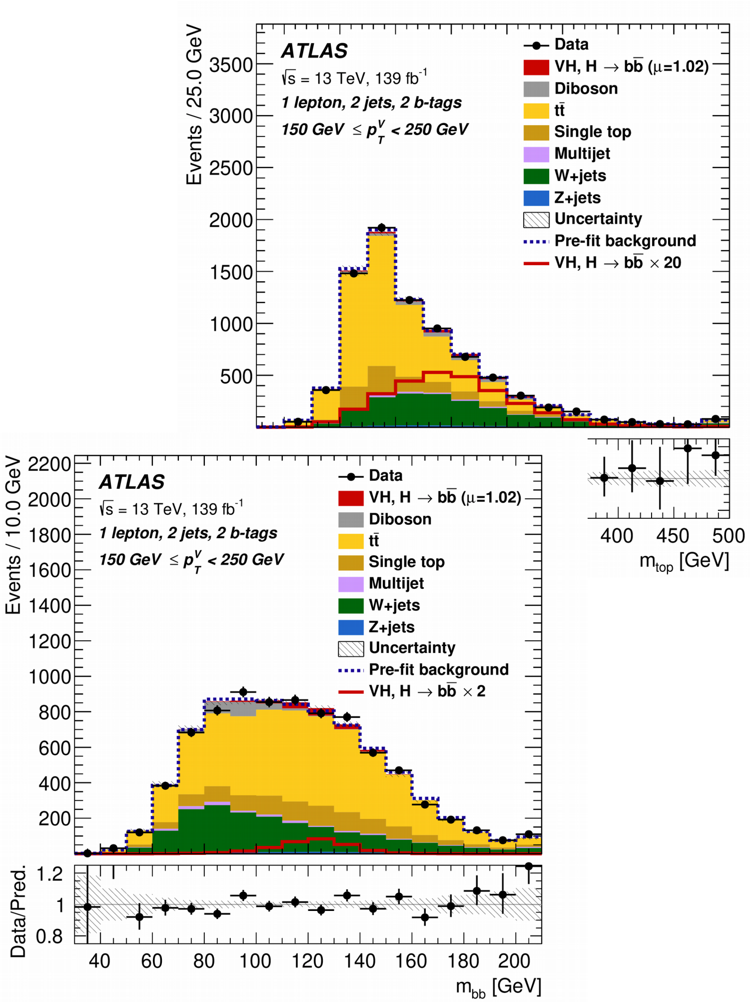


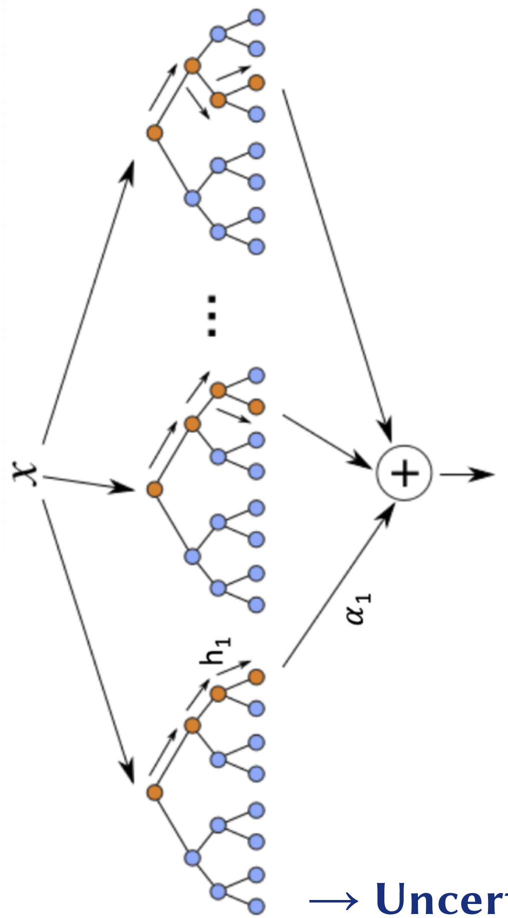


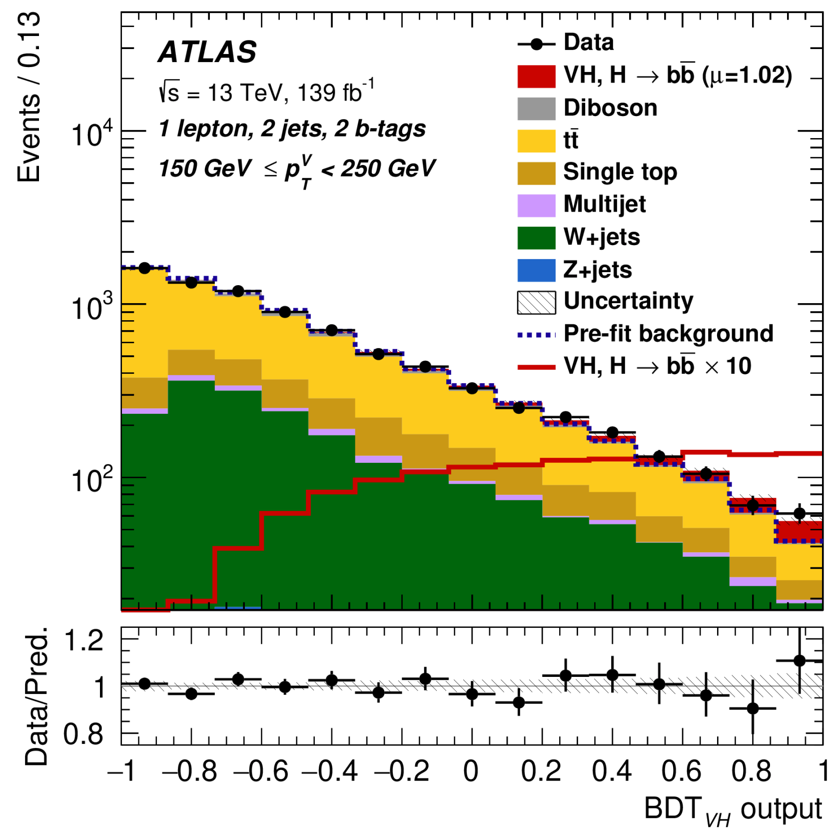
VH(bb/cc)

VH(bb) – MC Systematic Uncertainties

→ Signal strength extracted via profile likelihood fit using a **Boosted Decision Tree multi-variate proxy:**







+

9-11 extra kinematic observables

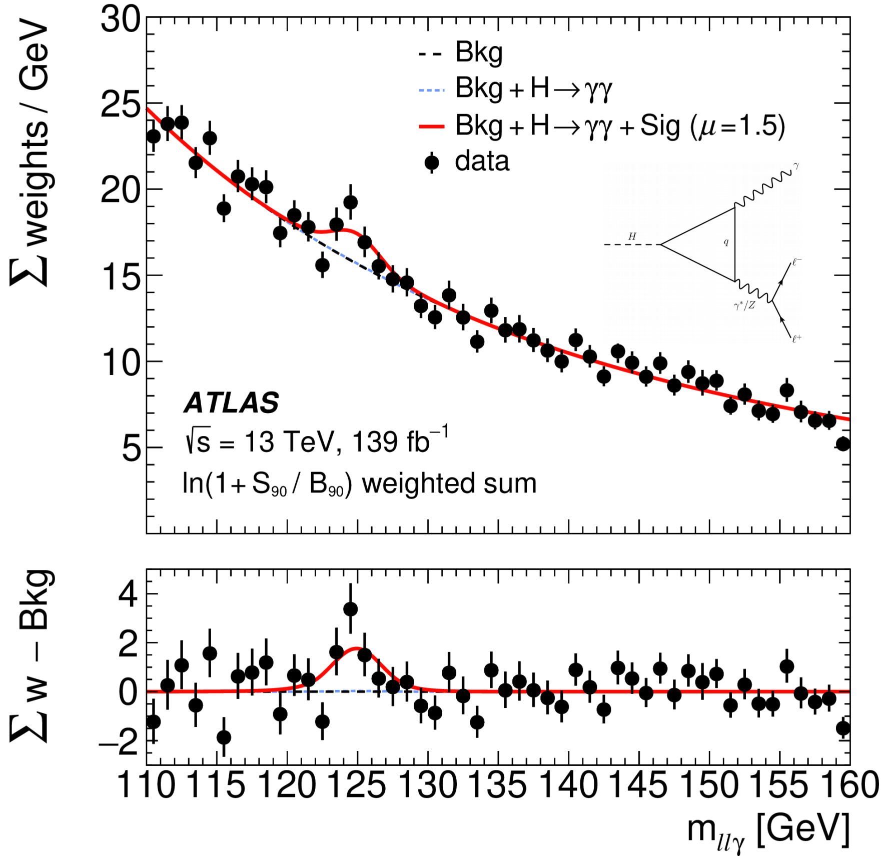
→ **Uncertainties** on the MC expectation included as nuisance parameters in the profile likelihood fit paradigm:

$$L(n|\vec{\mu},\vec{\theta}) = \prod_{b:bins} P(n_b|\nu_b(\vec{\mu},\vec{\theta})) \boxed{\cdot \prod_{\theta}^{N_\theta} f_\theta(\tilde{\theta}|\theta)} \cdot \prod_{b:bins} P(\nu_b(\vec{\mu},\vec{\theta})|\lambda_b,\tau_b)$$

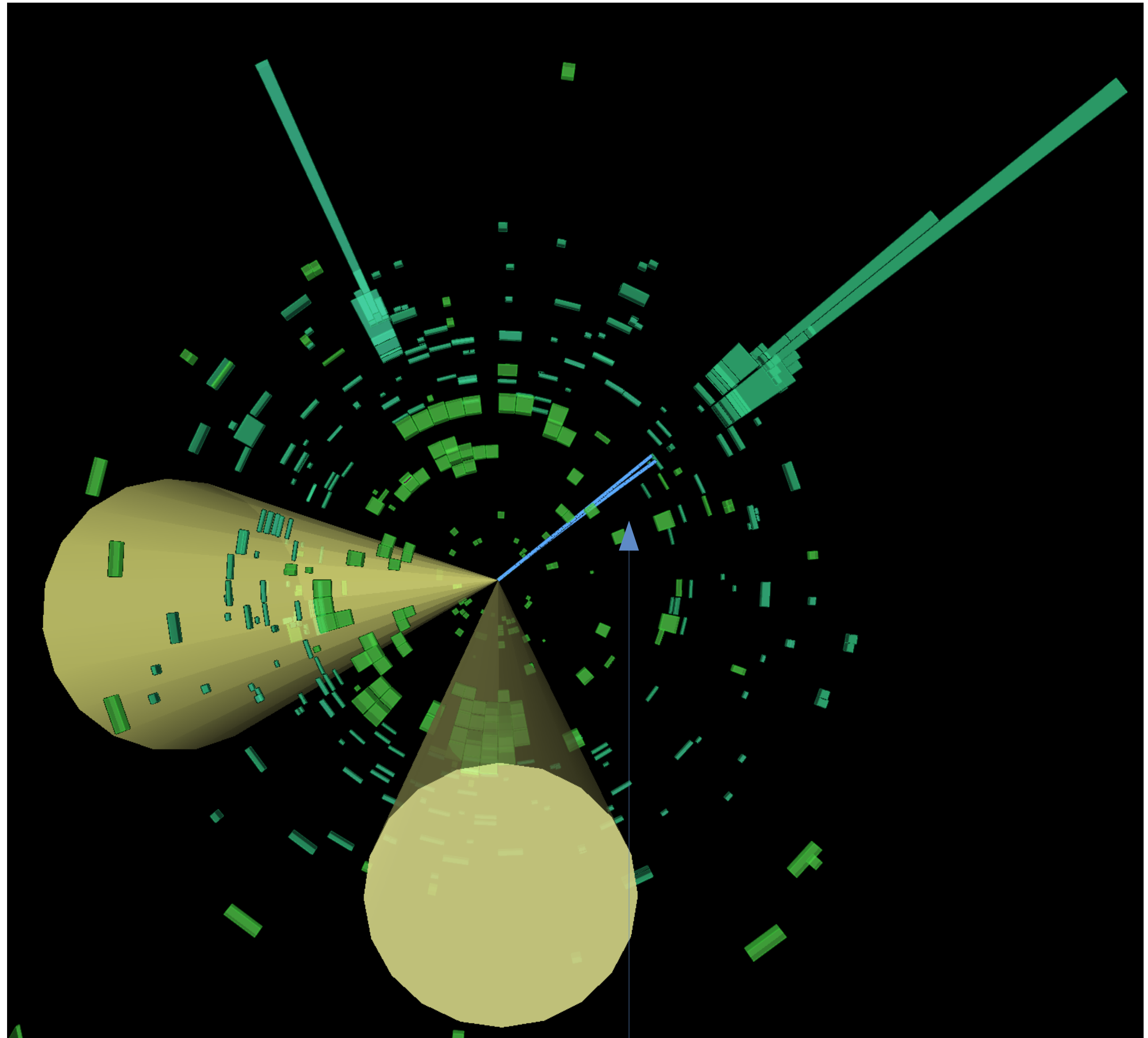
H(II)

$H(ll)+\gamma$ – Near-by electron identification

→ Measurement of $H(ll)+\gamma$ using ggF/VBF/VH production modes:



→ Search in m_{ll} < 30GeV to mitigate resonant backgrounds
- → Dielectron system prone to being collinear
- → Resolution of ECAL means two electrons *merge* at calorimeter cell level



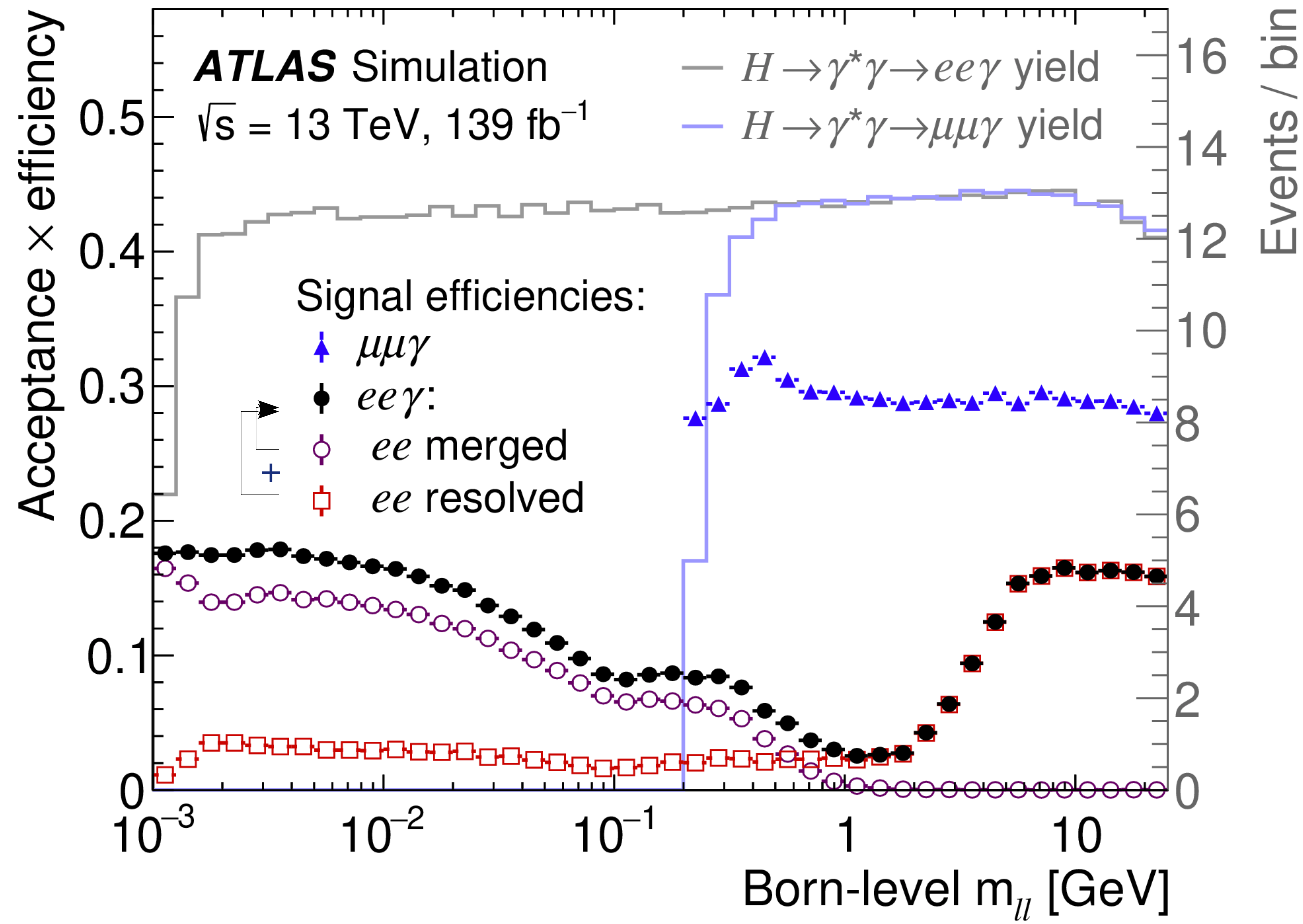
→ Electron inner detector tracks separable
→ Associated calo. cells **merge**

$H(ll)+\gamma$ – Near-by electron identification

⟶ Dedicated trigger, identification, and reconstruction algorithm for *ee-merged* signatures:

- If conversion vertex match, radius < 20mm
- Multivariate discriminant using:
  - Shower Shape
  - TRT signals
  - Cluster/ID track kinematics
- Calibrated using photons with conversion radius of 30 mm

⟶ ee-*merged* extends reach of analysis to low m_{ll} and thus highly boosted topologies



H(ll)+γ – Near-by electron identification

→ Dedicated trigger, identification, and reconstruction algorithm for *ee-merged* signatures:

- If conversion vertex match, radius < 20mm
- Multivariate discriminant using:
  - Shower Shape
  - TRT signals
  - Cluster/ID track kinematics
- Calibrated using photons with conversion radius of 30 mm

→ ee-*merged* extends reach of analysis to low m_{ll} and thus highly boosted topologies

→ Observed(expected) significance of $H \rightarrow ll\gamma$:
3.2(2.1)σ for m_H = 125.09 GeV