Measurement of off-shell Higgs boson production in the $H \rightarrow ZZ \rightarrow 4\ell$ decay channel using Neural Simulation-Based Inference

Jay Sandesara
on behalf of the ATLAS collaboration

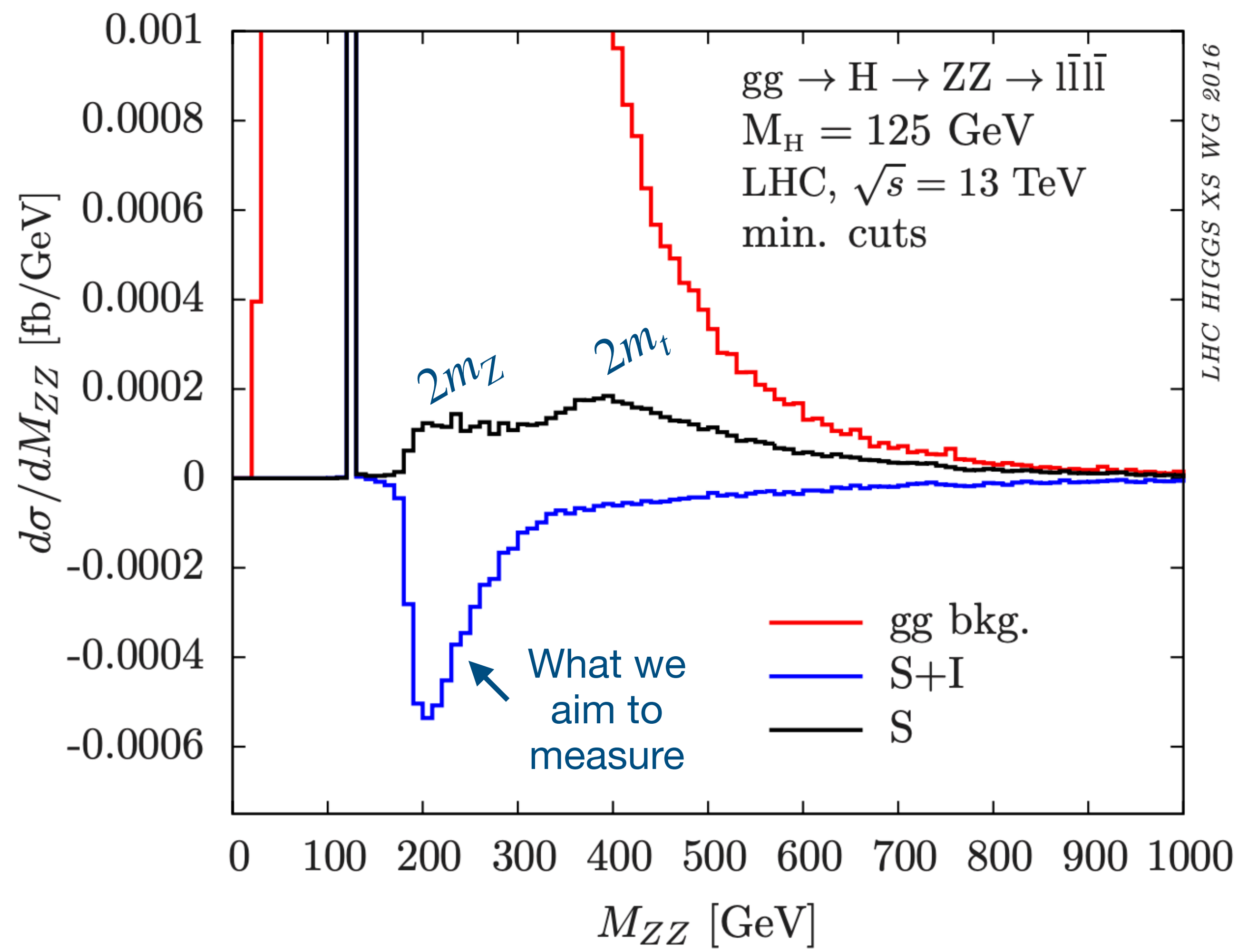
HIGGS 2024
Uppsala, Sweden
4-8 November 2024

https://indico.cern.ch/e/higgs2024

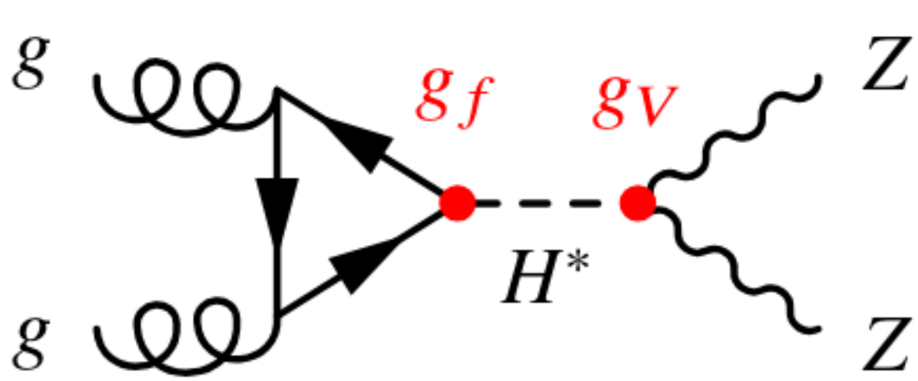
Introduction

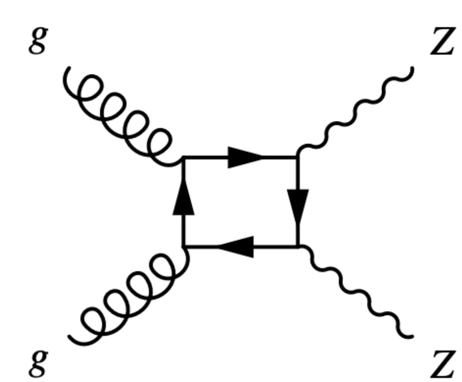
Off-shell $H \to ZZ$

- First evidence for production in 2022 [ATLAS, CMS]
- Small signal, most visible in $H \to VV$, due to enhanced cross-section from V-bosons in the decay channel and t-quarks in the quark loop going on-shell.



Off-shell Higgs signal (S) strongly interferes (S+I) with background (gg bkg.) and thus cannot be measured independently.





Interfering background

The off-shell Higgs boson

The probability model of the off-shell Higgs boson:

$$p(x\,|\,\mu) = \frac{1}{\nu(\mu)}\left[\underbrace{\mu}\cdot\nu_S\cdot\underbrace{p_S(x)} + \underbrace{\sqrt{\mu}}\cdot\nu_I\cdot\underbrace{p_I(x)} + \nu_B\cdot\underbrace{p_B(x)} + \nu_{NI}\cdot p_{NI}(x)\right]$$

$\nu \rightarrow$ Exp events

NI $\rightarrow$ Non-Interfering backgrounds

$p_S(x)$ $p_B(x)$

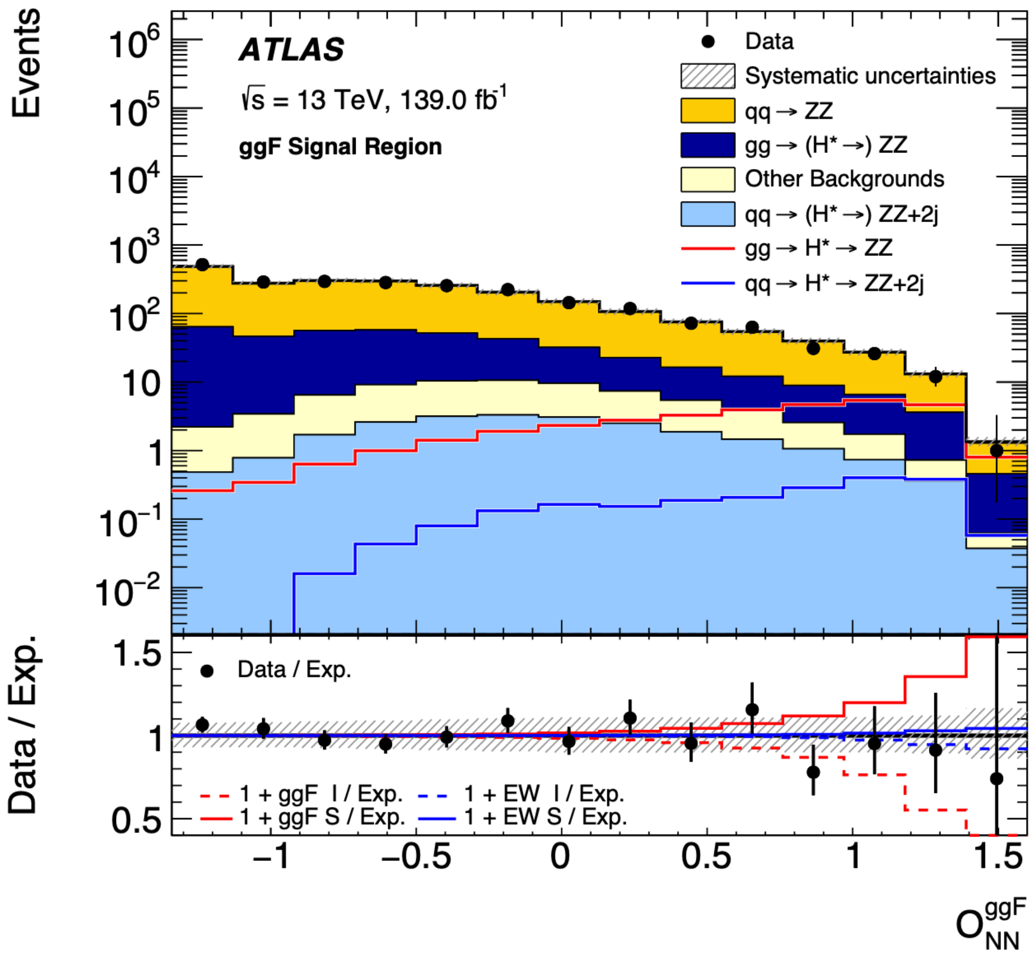
$p_I(x) = 2\cdot Re$

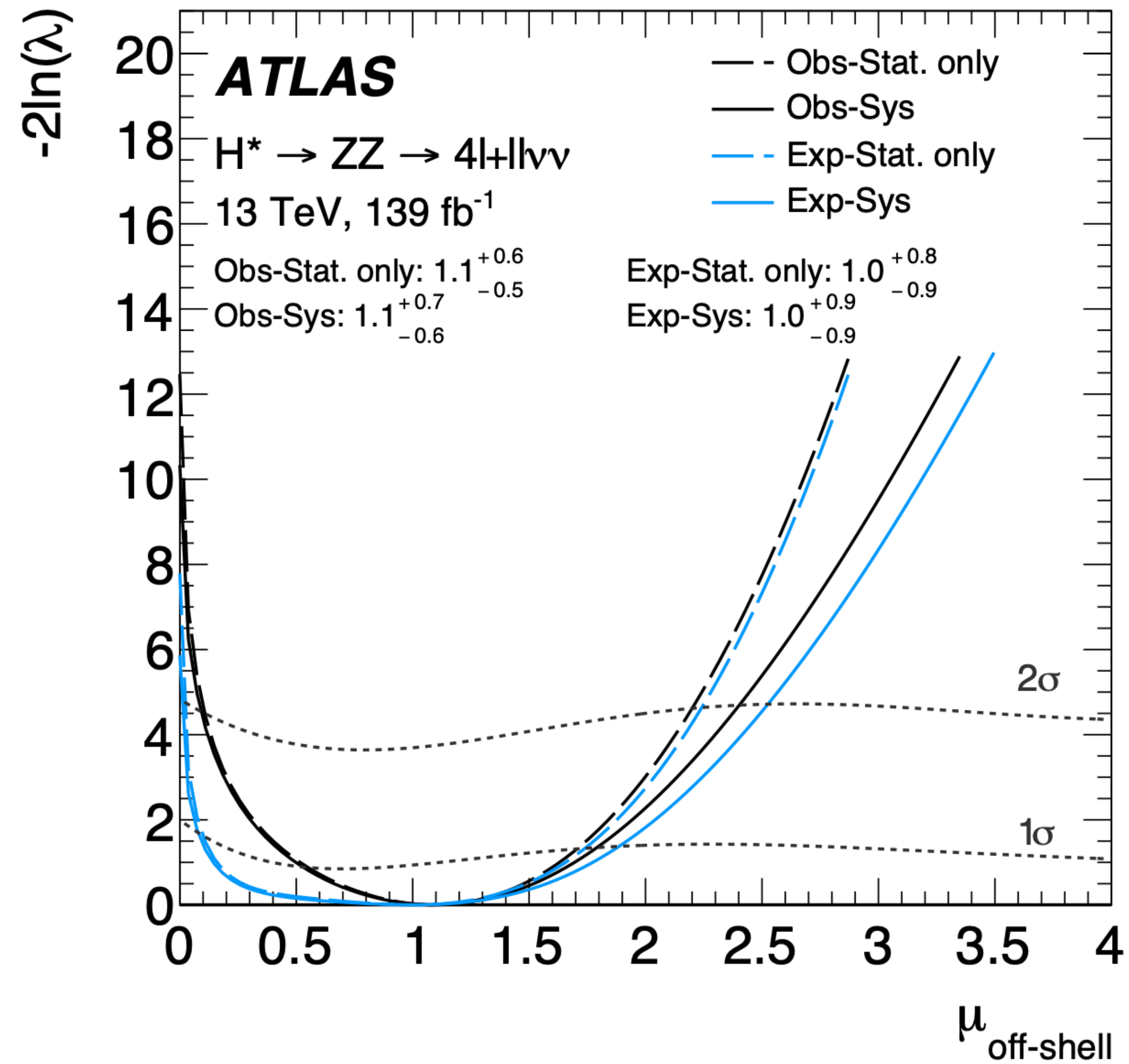
ggF Signal

ggF Background

We aim to measure the off-shell signal strength $\mu = \dfrac{\sigma_{obs}^{H\rightarrow ZZ\rightarrow 4\ell}}{\sigma_{exp}^{H\rightarrow ZZ\rightarrow 4\ell}}$

Previous Measurement





$$O_{NN} = \frac{p_S}{p_B + 0.1 \cdot p_{NI}}$$

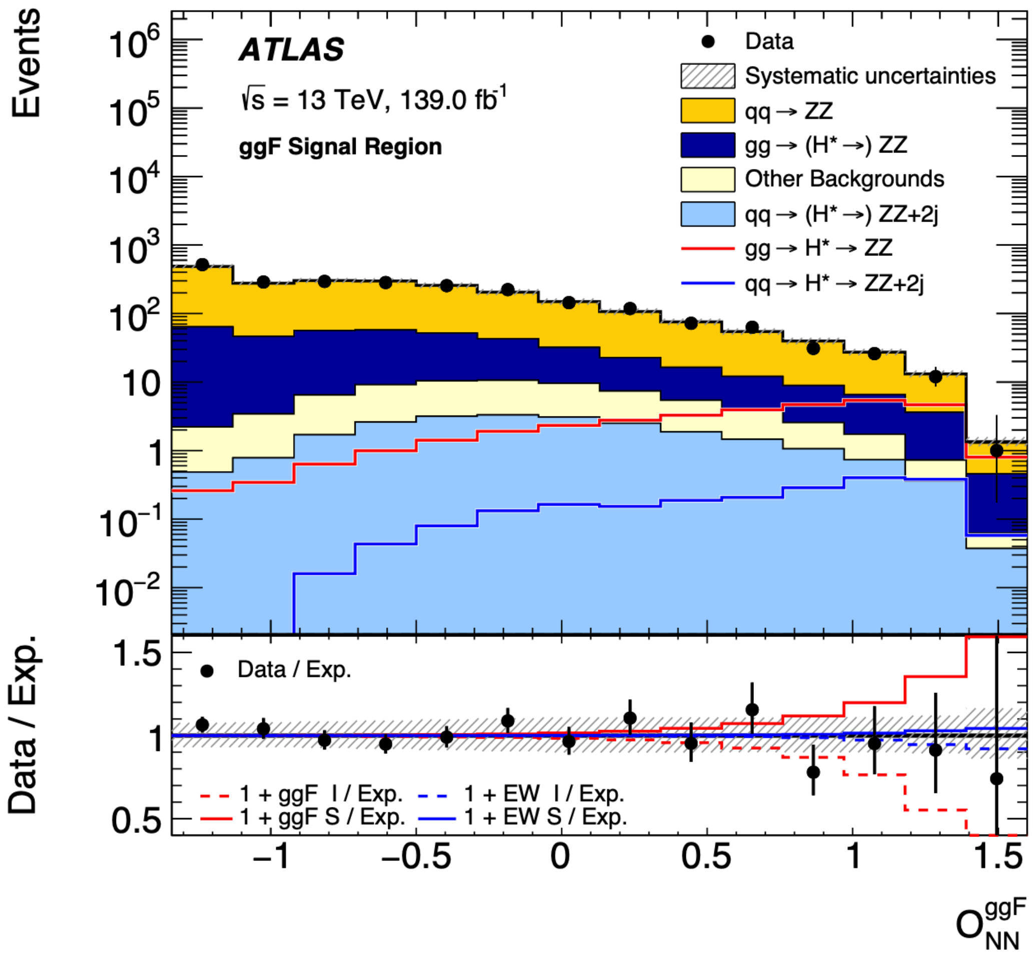
$$t_\mu = -2 \cdot \sum_{bin \in O_{NN}} \log \frac{e^{-\nu_{bin}(\mu,\hat{\hat{\alpha}})} \cdot \nu_{bin}(\mu,\hat{\hat{\alpha}})^{N_{bin}}}{e^{-\nu_{bin}(\mu,\hat{\alpha})} \cdot \nu_{bin}(\mu,\hat{\alpha})^{N_{bin}}}$$

Standard Signal vs Background classification

Binned Poisson Likelihood fit is performed

Previous Measurement

Signal vs Background discriminant optimal ONLY when signal **linearly** scales with parameter.



$$\frac{p(x\,|\,\mu)}{p_B(x)} = \mu \cdot \frac{p_S(x)}{p_B(x)} + \frac{p_B(x)}{p_B(x)}$$

Neyman pearson lemma

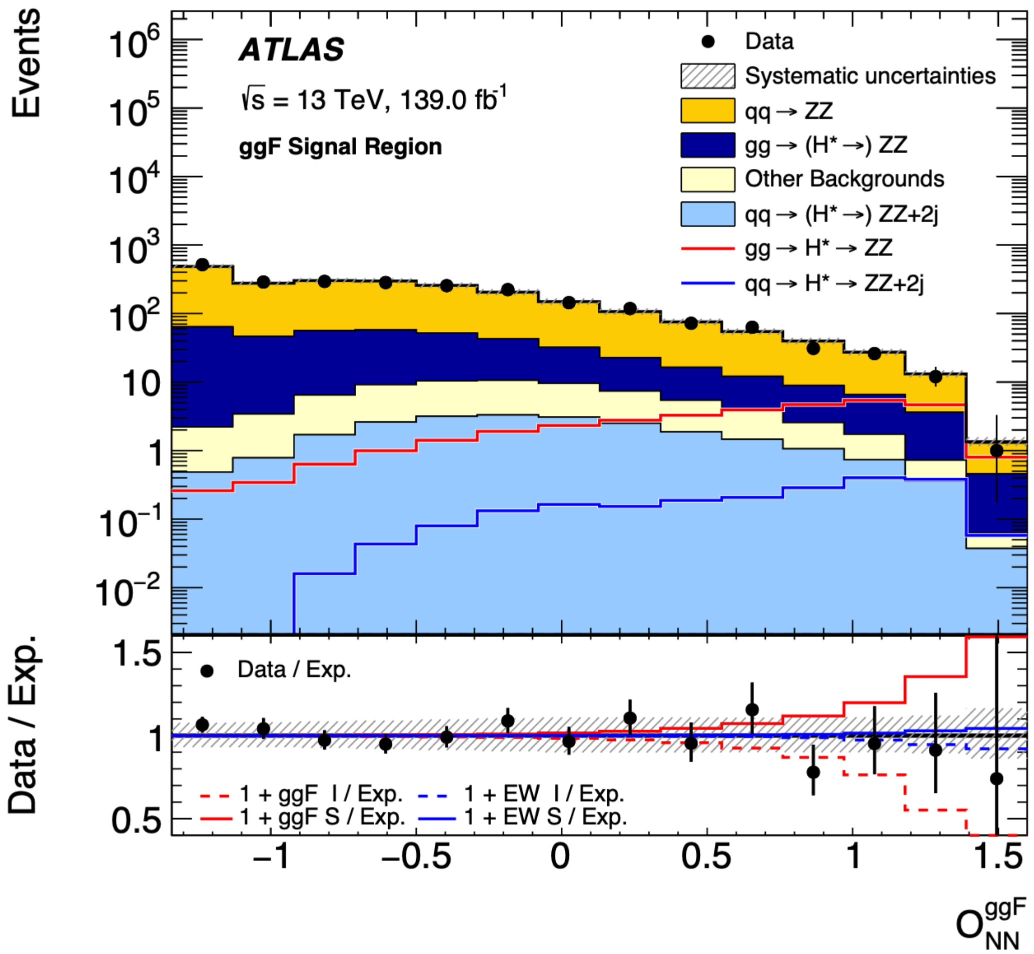
i.e. maximally optimal across the parameter range

$$O_{NN} = \frac{p_S}{p_B + 0.1 \cdot p_{NI}}$$

Standard Signal vs Background classification

Previous Measurement

Signal vs Background discriminant optimal ONLY when signal **linearly** scales with parameter.



$$\frac{p(x\,|\,\mu)}{p_B(x)} = \mu \cdot \frac{p_S(x)}{p_B(x)} + \frac{p_B(x)}{p_B(x)}$$

Neyman pearson lemma

But what if there is large non-linearity?

E.g.: interference effects of off-shell Higgs boson production.

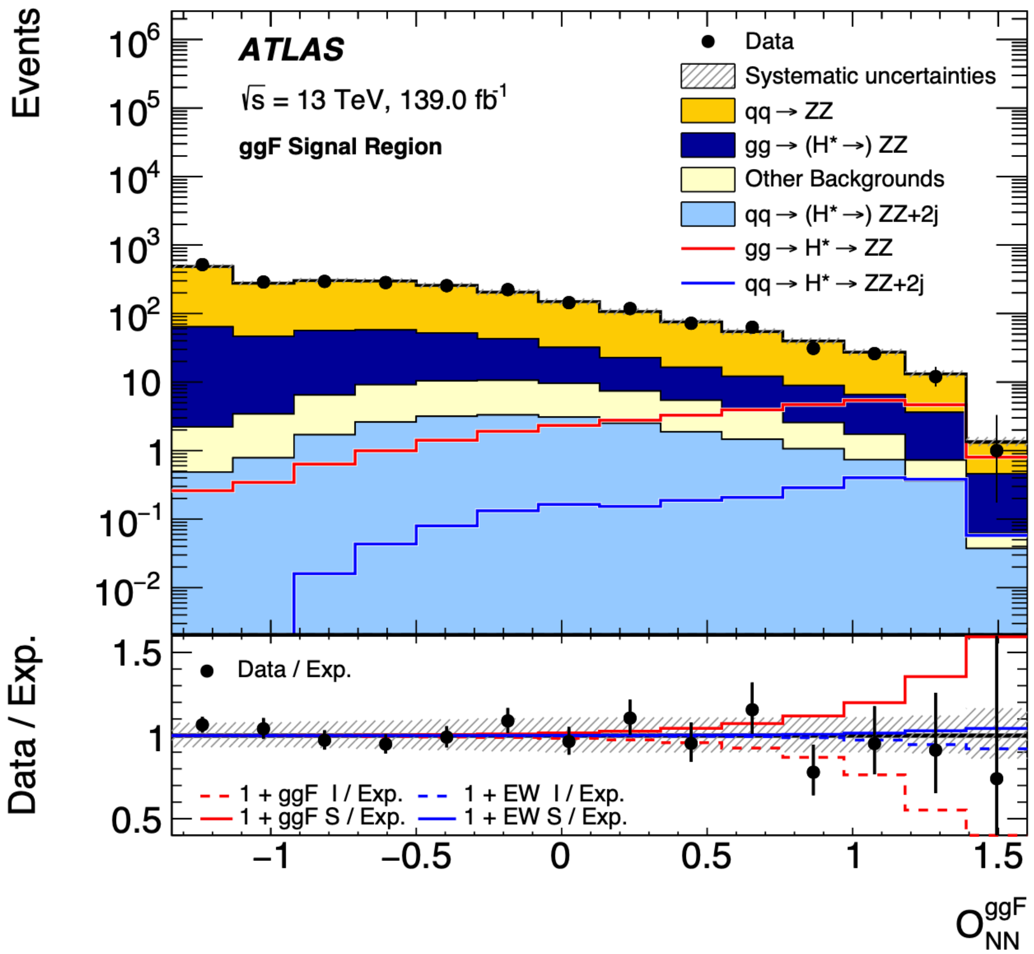
$$O_{NN} = \frac{p_S}{p_B + 0.1 \cdot p_{NI}}$$

Standard Signal vs Background classification

$$\frac{p(x\,|\,\mu)}{p_B(x)} = \mu \cdot \frac{p_S(x)}{p_B(x)} + \sqrt{\mu} \cdot \frac{p_I(x)}{p_B(x)} + \frac{p_B(x)}{p_B(x)}$$

What about optimally discriminating interference from background?

Previous Measurement



$$O_{NN} = \frac{p_S}{p_B + 0.1 \cdot p_{NI}}$$

Standard Signal vs Background classification

Signal vs Background discriminant optimal ONLY when signal **linearly** scales with parameter.

$$\frac{p(x\,|\,\mu)}{p_B(x)} = \mu \cdot \frac{p_S(x)}{p_B(x)} + \frac{p_B(x)}{p_B(x)}$$

Neyman pearson lemma ✔

But what if there is large non-linearity?

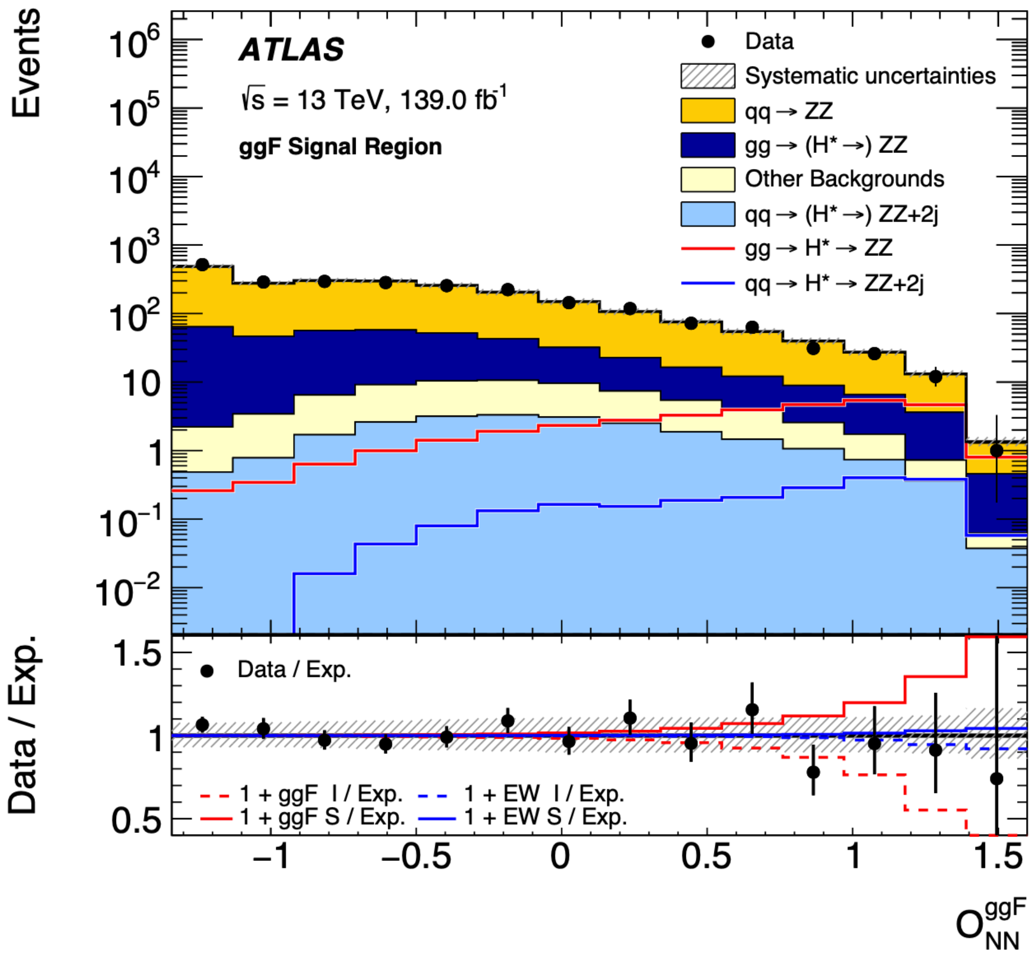
E.g.: interference effects of off-shell Higgs boson production.

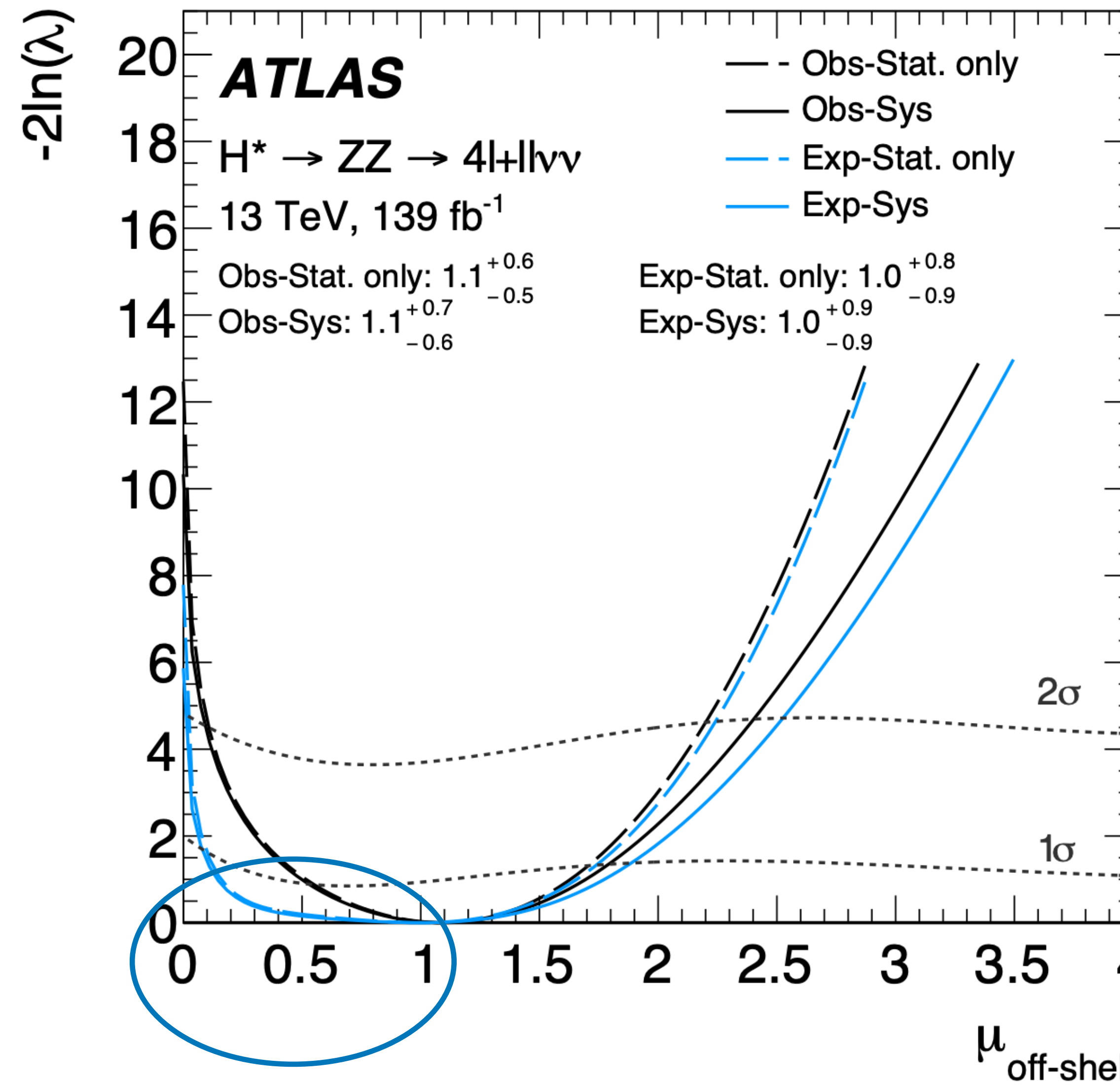
$$\frac{p(x\,|\,\mu)}{p_B(x)} = \mu \cdot \frac{p_S(x)}{p_B(x)} + \sqrt{\mu} \cdot \frac{p_I(x)}{p_B(x)} + \frac{p_B(x)}{p_B(x)}$$

Neyman pearson lemma ✘

What about optimally discriminating interference from background?

Previous Measurement





$$O_{NN} = \frac{p_S}{p_B + 0.1 \cdot p_{NI}}$$

Flat NLL region implies sub-optimality in regions with $\sqrt{\mu} \cdot p_I \gg \mu \cdot p_S$

Standard Signal vs Background classification

New Measurement

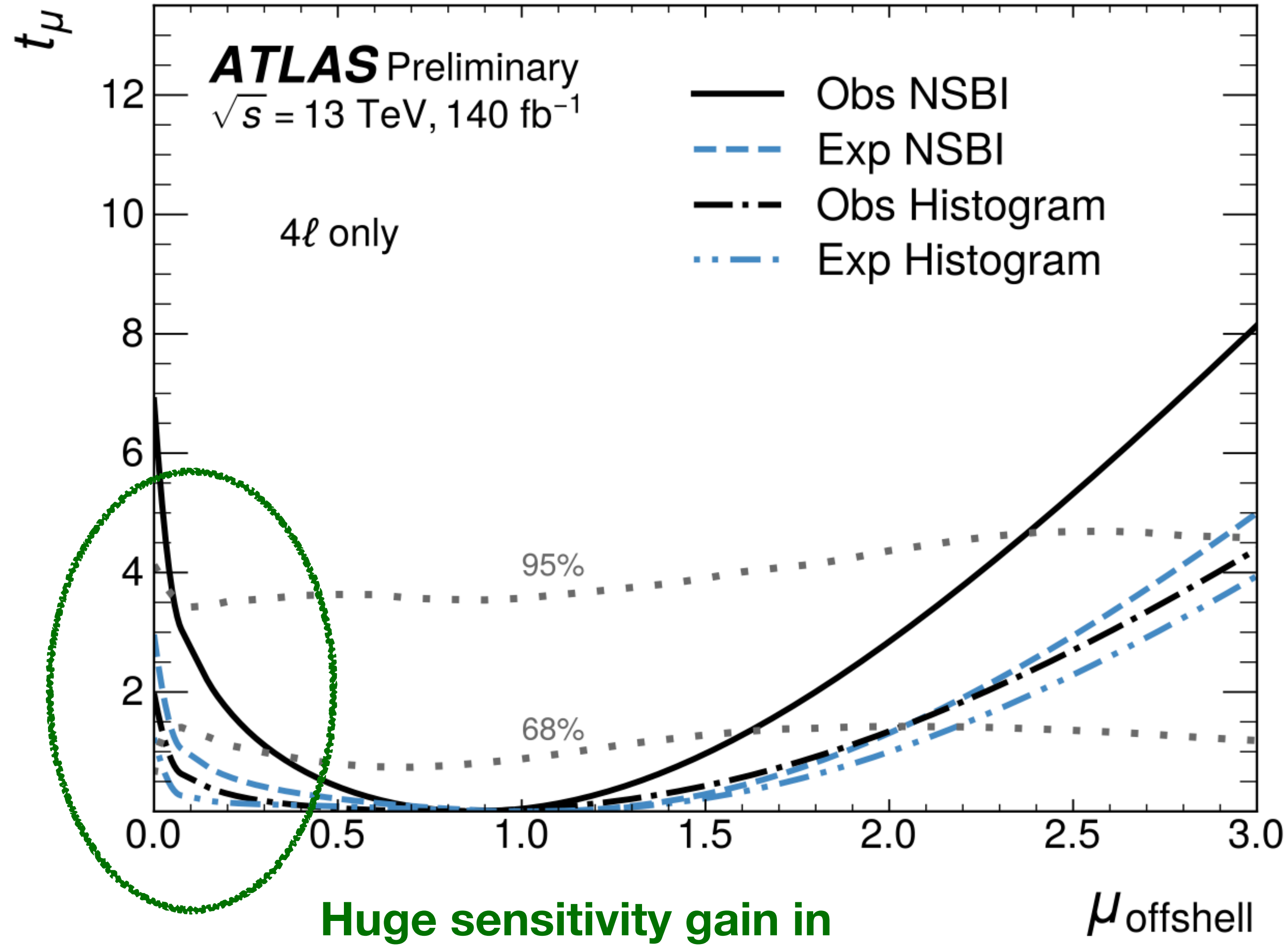
Carefully trained **parameterized per-event density ratios** are now used to build the test statistic:

$$t_\mu \sim -2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \mid \mu)}{p(x_i \mid \hat{\mu})}$$

No fixed observable - maximal optimality throughout μ space.

Neyman pearson lemma ✓

Additional sensitivity from unbinned nature (no Poisson fits) ✓



Huge sensitivity gain in interference rich regions

$$\sqrt{\mu} \cdot p_I(x) \gg \mu \cdot p_S(x)$$

Note: we use the same pre-selections, Monte Carlo samples, background normalization, and systematic uncertainty model as the previously published analysis [link to paper for details]

New Measurement

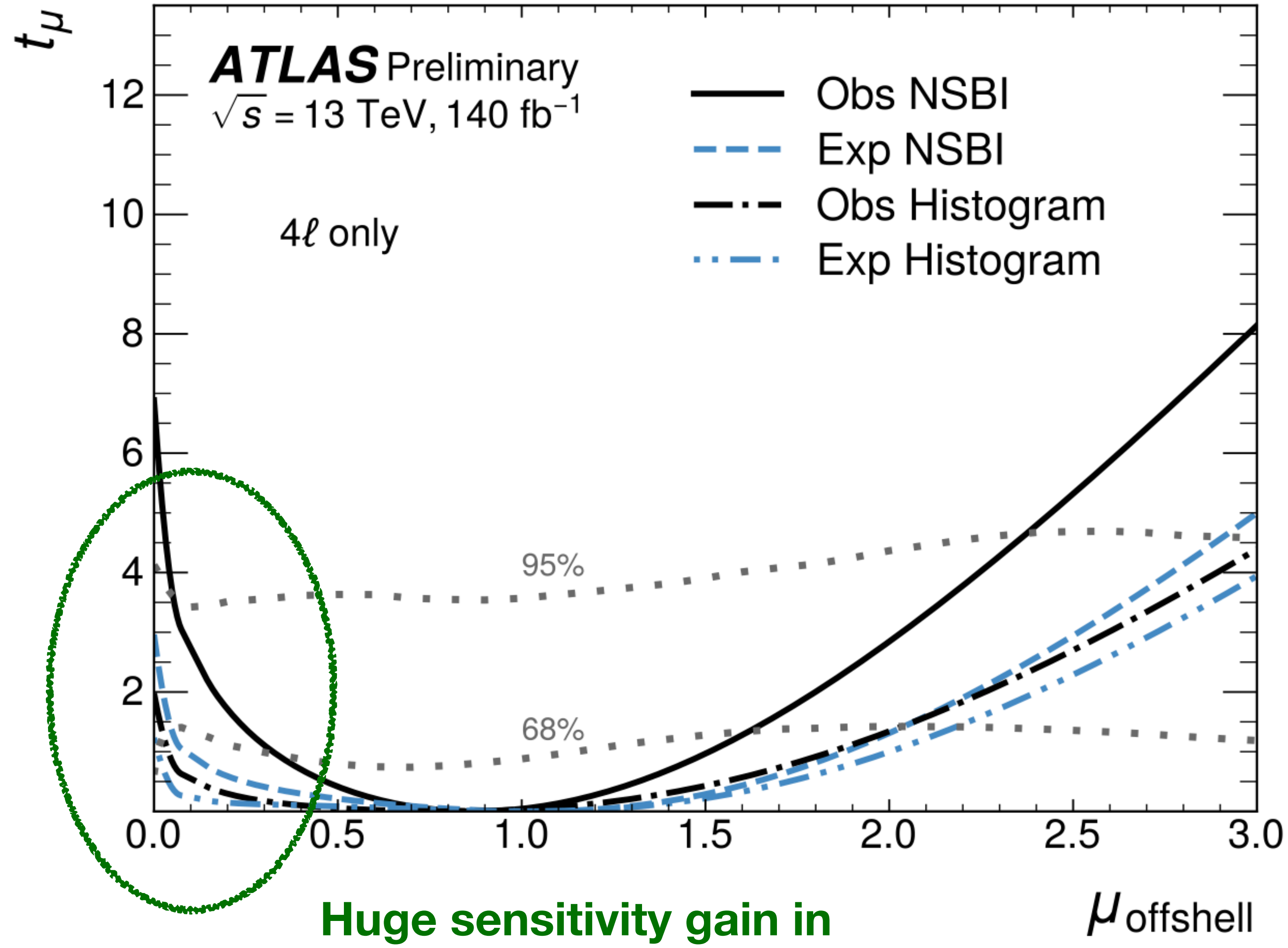
Carefully trained **parameterized per-event density ratios** are now used to build the test statistic:

$$t_\mu \sim -2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i\,|\,\mu)}{p(x_i\,|\,\hat{\mu})}$$

No fixed observable - maximal optimality throughout μ space.

Neyman pearson lemma ✓

Additional sensitivity from unbinned nature (no Poisson fits) ✓



Huge sensitivity gain in interference rich regions

$$\sqrt{\mu} \cdot p_I(x) \gg \mu \cdot p_S(x)$$

Exploiting the known analytical formula - we break down the parameterized ratio into simpler parts:

$$\frac{p(x\,|\,\mu)}{p(x\,|\,\hat{\mu})} = \frac{p(x\,|\,\mu)/p_{ref}(x)}{p(x\,|\,\hat{\mu})/p_{ref}(x)} \longrightarrow \frac{p(x\,|\,\mu)}{p_{ref}(x)} = \mu \cdot \frac{p_S(x)}{p_{ref}(x)} + \sqrt{\mu} \cdot \frac{p_I(x)}{p_{ref}(x)} + \frac{p_B(x)}{p_{ref}(x)} + \frac{p_{NI}(x)}{p_{ref}(x)}$$

p_{ref} is a carefully chosen **parameter-independent** hypothesis

We learn everything, including interference effects

Overview: Neural Simulation-Based Inference

Full test statistic function with nuisance parameters α:

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs} \mid \mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs} \mid \hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \mid \mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}}(x_i)}{p(x_i \mid \hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}}(x_i)} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{L_{subs}(\hat{\hat{\alpha}})}{L_{subs}(\hat{\alpha})}$$

Extended Poisson term

Sum of event-by-event log-likelihood ratios

Constraint terms

$N_{obs} \rightarrow$ total observed events

$L_{subs} \rightarrow$ likelihood from subsidiary measurements of the nuisance parameters

Overview: Neural Simulation-Based Inference

Full test statistic function with nuisance parameters α:

$$t(\mu) = -2 \cdot \log \frac{\mathrm{Pois}(N_{obs}\,|\,\mu, \hat{\hat{\alpha}})}{\mathrm{Pois}(N_{obs}\,|\,\hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i\,|\,\mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}}(x_i)}{p(x_i\,|\,\hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}}(x_i)} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{L_{subs}(\hat{\hat{\alpha}})}{L_{subs}(\hat{\alpha})}$$

sum over processes $c = S, B,$ **etc.**

parameter-independent ratio

$$\frac{p(x_i\,|\,\mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i\,|\,\alpha) \cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Parameterized per-event ratios

Factorized **nuisance parameter** α-dependence:

$$g_c(x\,|\,\alpha) = \prod_m \frac{p_c(x\,|\,\alpha_m)}{p_c(x)}$$

Parameter dependancies are factorized out (see slide 10)

Overview: Neural Simulation-Based Inference

Full test statistic function with nuisance parameters α:

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs} \,|\, \mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs} \,|\, \hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}}(x_i)}{p(x_i \,|\, \hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}}(x_i)} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{L_{subs}(\hat{\hat{\alpha}})}{L_{subs}(\hat{\alpha})}$$

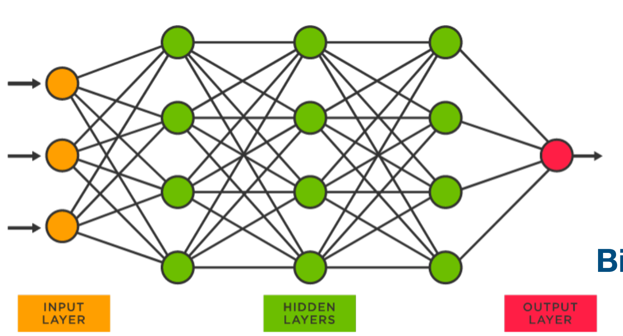
sum over processes
$c = S, B,$ **etc.**

$$\boxed{\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x_i)}} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i \,|\, \alpha) \cdot \nu_c \cdot \boxed{\frac{p_c(x_i)}{p_{ref}(x_i)}} \right]$$

$$\boxed{\begin{matrix} x \sim p_c \\ S = 1 \end{matrix}}$$

$$\boxed{\begin{matrix} x \sim p_{ref} \\ S = 0 \end{matrix}}$$

Two hypothesis:
p_c and p_{ref}



Classification NN

$$\underset{\omega}{\text{argmin}} \; \text{L} \longrightarrow \hat{s}(x) = \frac{p_c}{p_{ref} + p_c}(x) \longrightarrow \boxed{\frac{p_c}{p_{ref}}(x) = \frac{\hat{s}(x)}{1.0 - \hat{s}(x)}}$$

Binary Cross-Entropy loss

"Likelihood ratio trick"

Many examples in ATLAS - HH4b background estimation, Omnifold, etc.

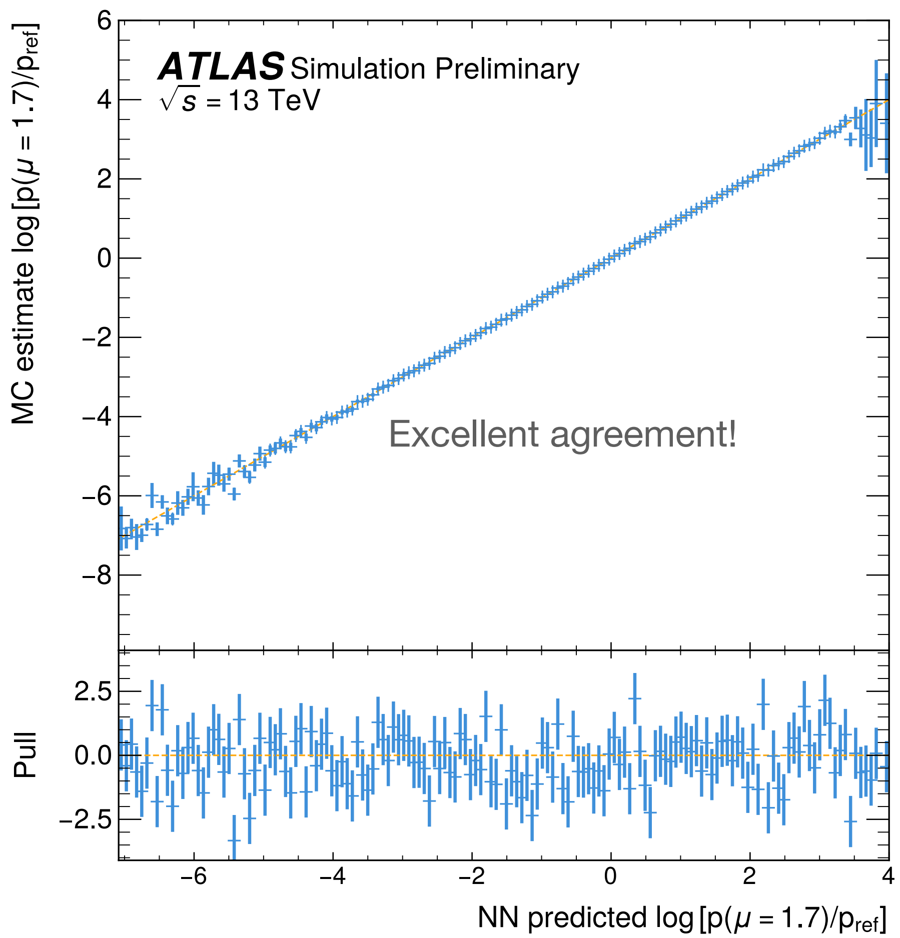
Probability Calibration Test

The NN ratios are meticulously trained to be **true representations of the density ratios**

Does the NN output correspond to real probabilities?

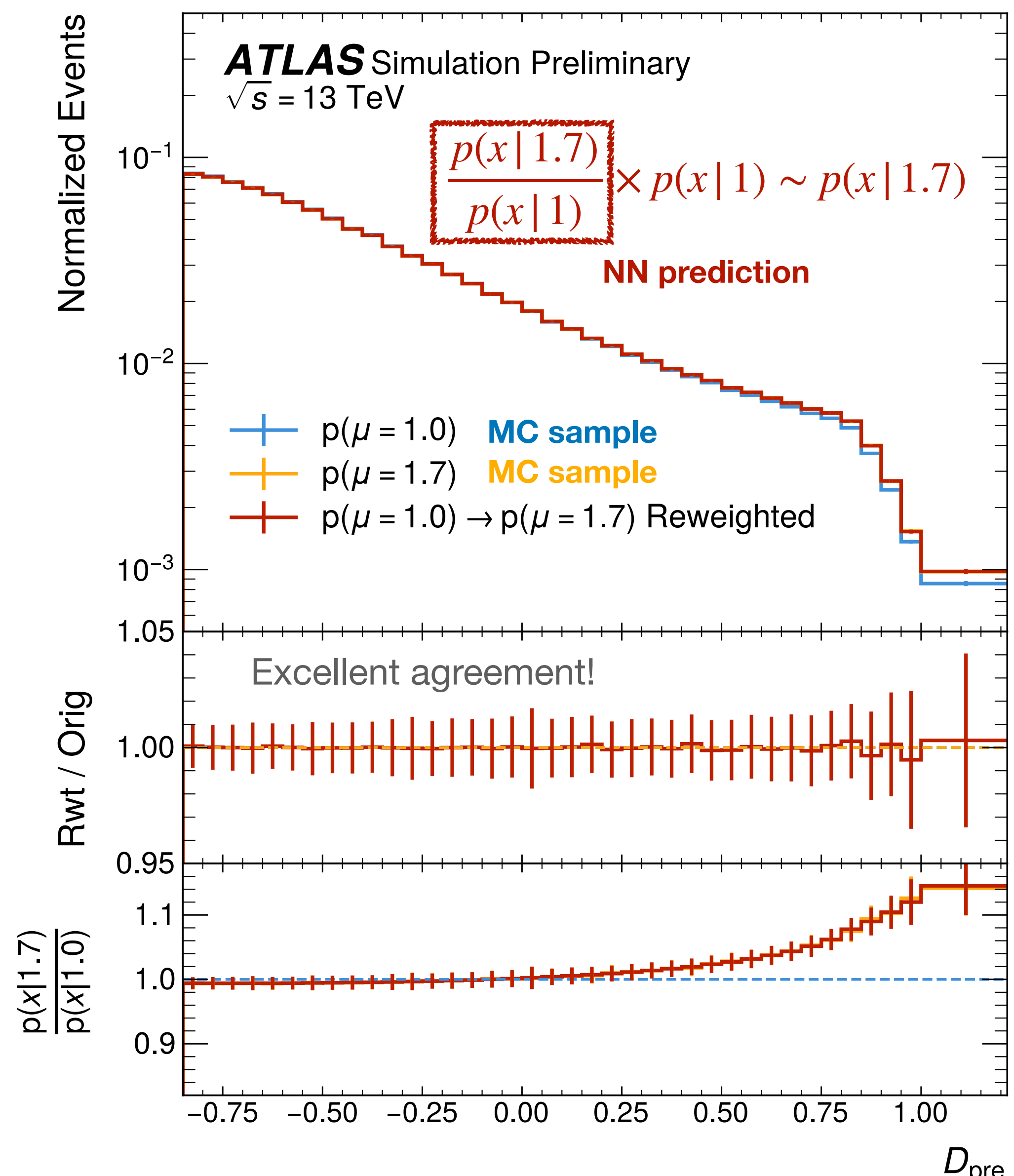
$$\log \frac{p(x\,|\,\mu)}{p_{ref}(x)}(x) \leftrightarrow \log \frac{N^{bin}(x\,|\,\mu)}{N^{bin}_{ref}(x)} \; ?$$

MC estimate $\log \frac{N^{bin}(x\,|\,\mu)}{N^{bin}_{ref}(x)}$



NN predicted $\log \frac{p(x\,|\,\mu)}{p_{ref}(x)}(x)$

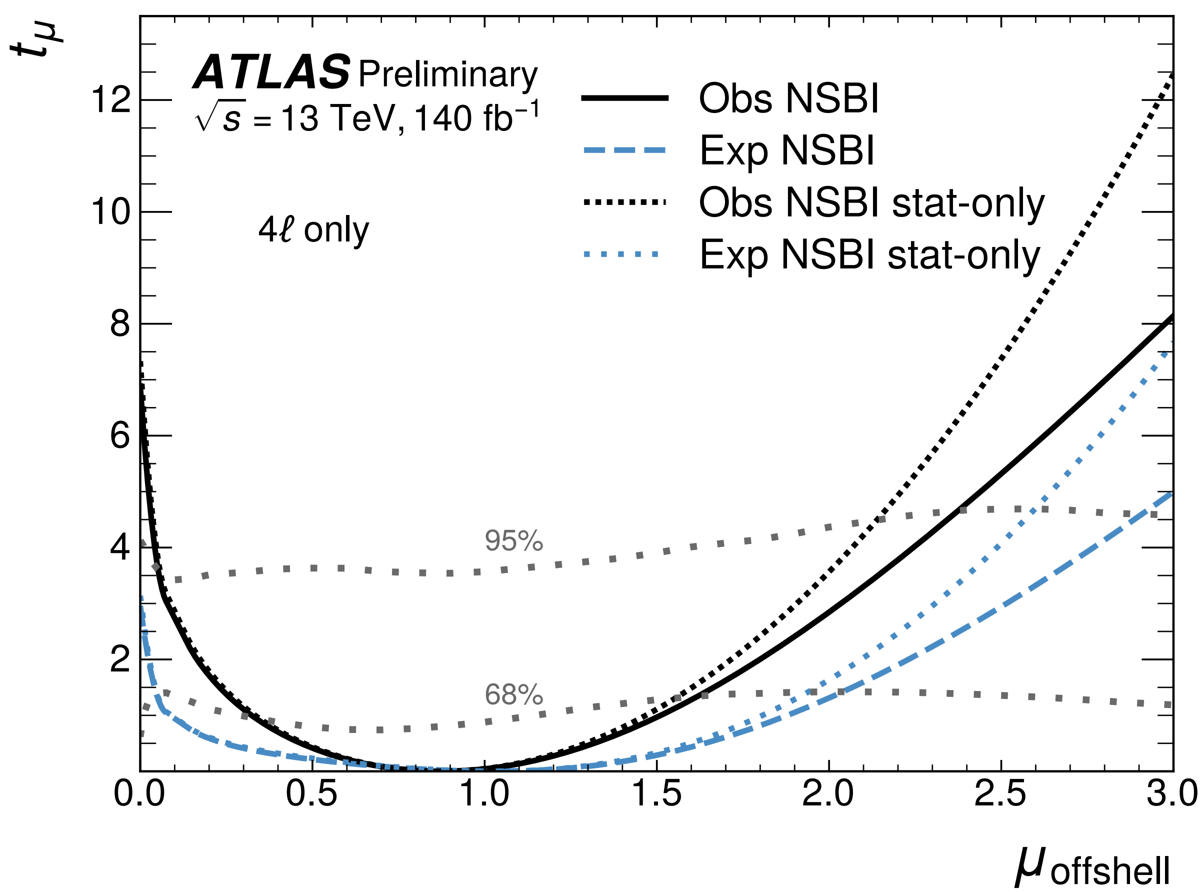
Do the ratios capture the full un-biased dependence of the multi-dimensional feature space x ?

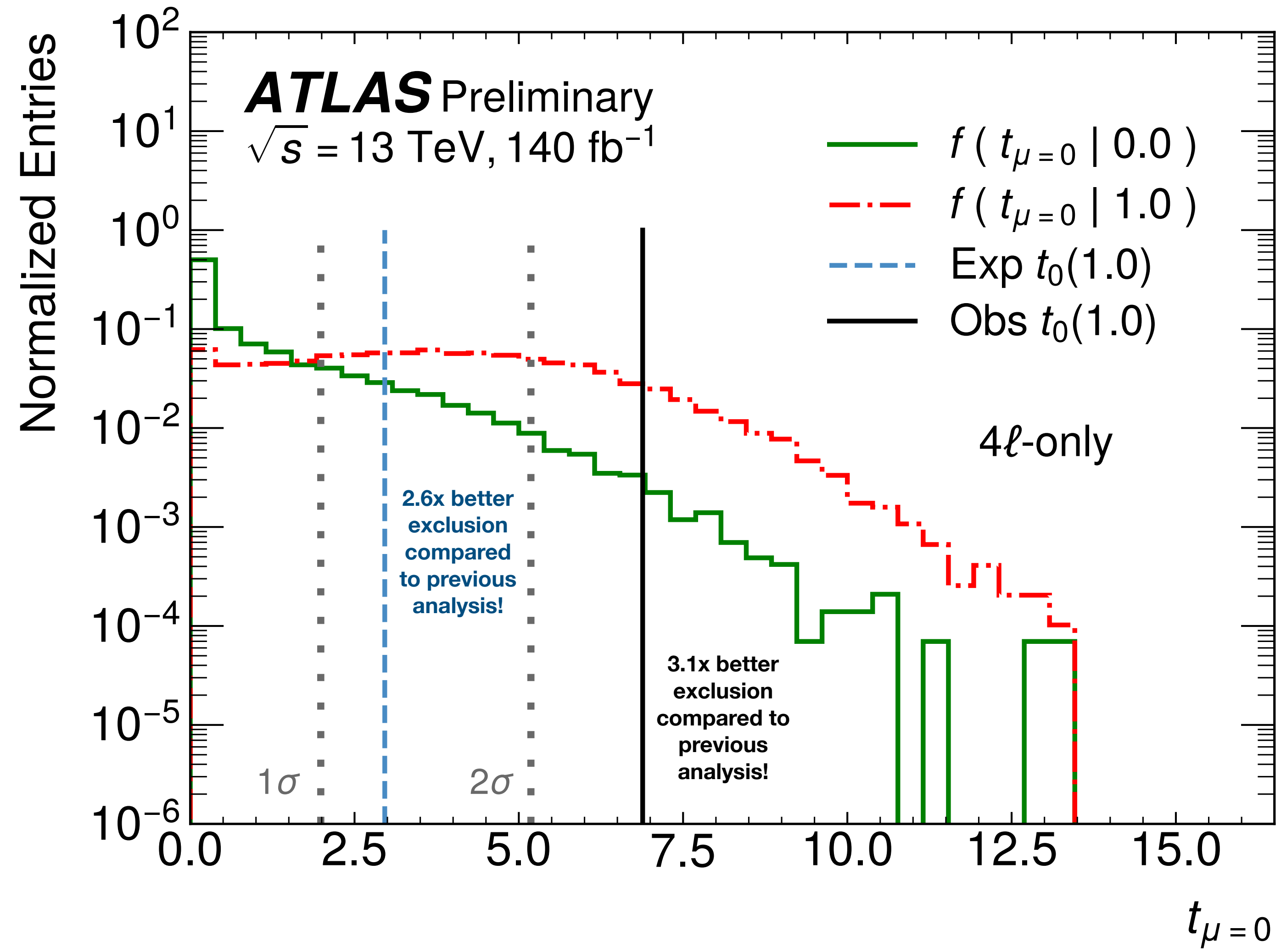


High-level multivariate observable

Unblinded Results - Parameter scans

Having validated the parameterized density ratios we build the test statistic scan for $\mu_{offshell}$





Neyman Construction, **essential due to the non-linear parameterization**, requires sampling pseudo-experiments from the PDF $p(x \mid \mu, \alpha)$, unlike histogram analysis which rely on Poisson bin-by-bin sampling.

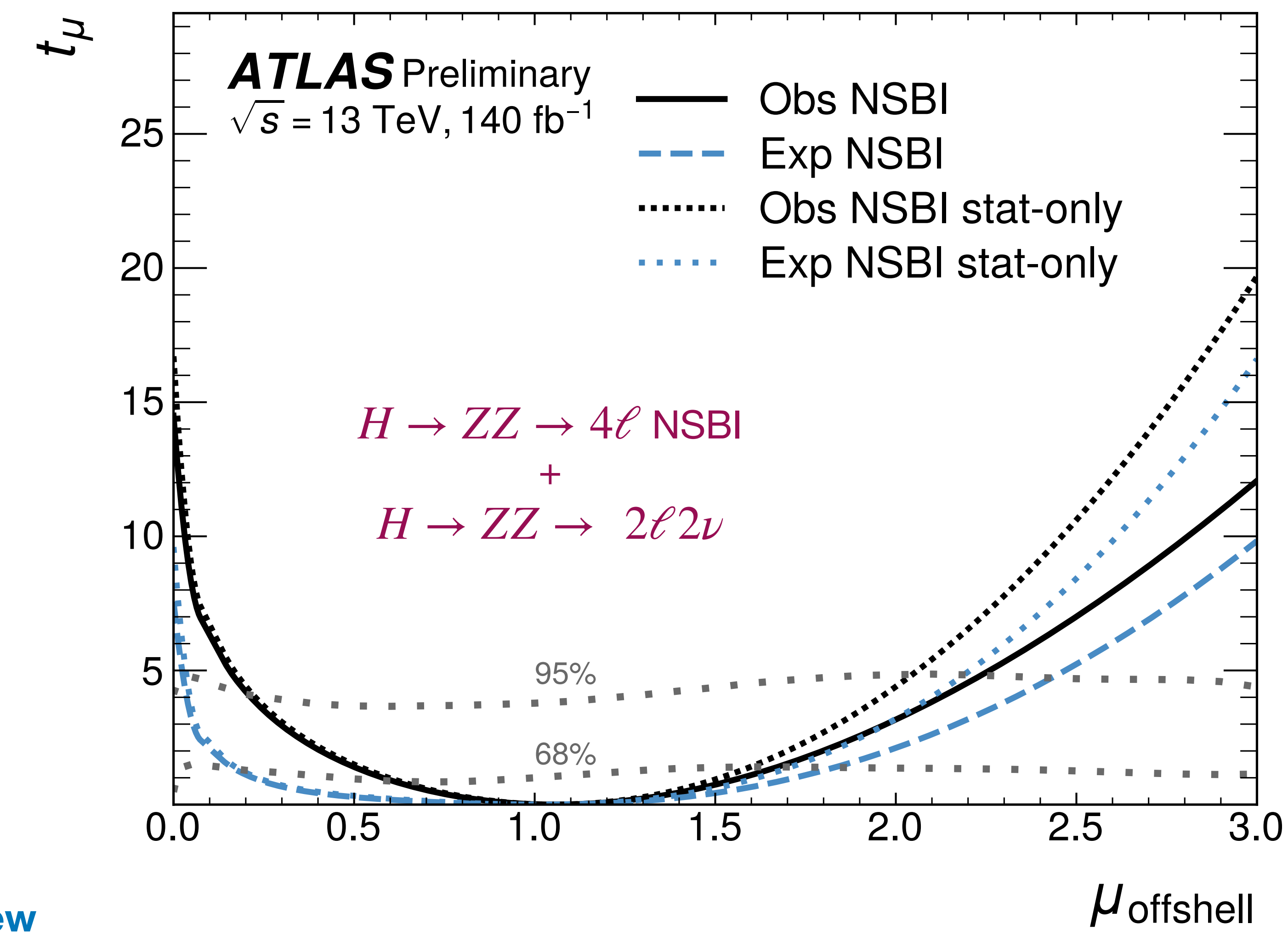
Pseudo-experiments sampled using the newly developed techniques developed have been used to calculate the exact confidence intervals and background exclusion significance.

Conclusions and Outlook

The analysis has been combined with other channels like:

1. off-shell $H \to ZZ \to 2\ell 2\nu$ for powerful inference on $\mu_{offshell}$ and

2. on-shell $H \to ZZ \to 4\ell$ to give precise estimate on the total decay width Γ_H of the Higgs boson.



Will's talk tomorrow will show more of the exciting new results from physics re-interpretations of this powerful new measurement

The no off-shell production hypothesis is rejected with a significance of 3.7σ

Two new papers from ATLAS open up the possibility of wide applications in ATLAS, CMS and beyond - potential to maximise the optimality of many analysis:

- Paper measuring the off-shell Higgs boson: [link]
- Paper with general method: [link]

Papers will be out soon, and current CONF note links will be replaced with paper links.

Backup

Uncertainty Parameterization

$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c}\sum_c \left[f_c(\mu)\cdot g_c(x_i|\alpha)\cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Factorized **yield** α-dependence:

$$G_c(\alpha) = \prod_k \frac{\nu_c(\alpha_k)}{\nu_c}$$

with $\nu_c(\alpha_k)/\nu_c$ estimated using **analytic interpolation techniques:**

Available from simulations at $\alpha_k = 0,\ \alpha_k^+,\ \alpha_k^-$

$$\frac{\nu_c(\alpha_k)}{\nu_c} = \begin{cases} \left(\boxed{\frac{\nu_c(\alpha_k^+)}{\nu_c}}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n\alpha_k^n & -1 \le \alpha_k \le 1, \\ \left(\boxed{\frac{\nu_c(\alpha_k^-)}{\nu_c}}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

Per-event analog of standard techniques

Factorized **per-event** α-dependence:

$$g_c(x|\alpha) = \prod_k \frac{p_c(x|\alpha_k)}{p_c(x)}$$

with $p_c(x|\alpha_k)/p_c(x)$ estimated using a **mix of NNs and analytic interpolation techniques:**

Density ratios trained using NNs from simulations at $\alpha_k = 0,\ \alpha_k^+,\ \alpha_k^-$

$$\frac{p_c(x|\alpha_k)}{p_c(x)} = \begin{cases} \left(\boxed{\frac{p_c(x|\alpha_k^+)}{p_c(x)}}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n\alpha_k^n & -1 \le \alpha_k \le 1\ . \\ \left(\boxed{\frac{p_c(x|\alpha_k^-)}{p_c(x)}}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

Ref: HistFactory

Impact of Systematic Uncertainties

Table 5: Absolute systematic uncertainties on the measurement of $\mu_{\text{off-shell}}$ in the $H \to ZZ \to 4\ell$ decay channel. Two methods of estimation are presented: based on the variation of nuisance parameters and on the variation of global observables. Total uncertainties are given using the global observables methods since it allows variations to be summed in quadrature. The total uncertainty is independent of the method used to estimate systematic uncertainties.

Uncertainty source	**Absolute impact on $\mu_{\text{off-shell}}$**	
	Nuisance Parameter	**Global Observable**
Electron uncertainties	(−0.05, +0.06)	(−0.05, +0.06)
Muon uncertainties	(−0.03, +0.03)	(−0.02, +0.03)
Jet uncertainties	(−0.10, +0.10)	(−0.09, +0.11)
Luminosity	(−0.01, +0.01)	(−0.01, +0.01)
Total experimental	(−0.12, +0.13)	(−0.11, +0.12)
$q\bar{q} \to ZZ$ modeling	(−0.06, +0.07)	(−0.06, +0.07)
$gg \to ZZ$ modeling	(−0.08, +0.13)	(−0.07, +0.09)
EW $q\bar{q} \to ZZ + 2j$ modeling	(−0.01, +0.01)	(−0.01, +0.01)
Total modeling	(−0.10, +0.15)	(−0.09, +0.12)
Systematic uncertainty	(−0.16, +0.19)	(−0.14, +0.17)
Statistical uncertainty	(−0.49, +0.72)	(−0.50, +0.73)
Total uncertainty	(−0.54, +0.75)	

Compatibility Tests

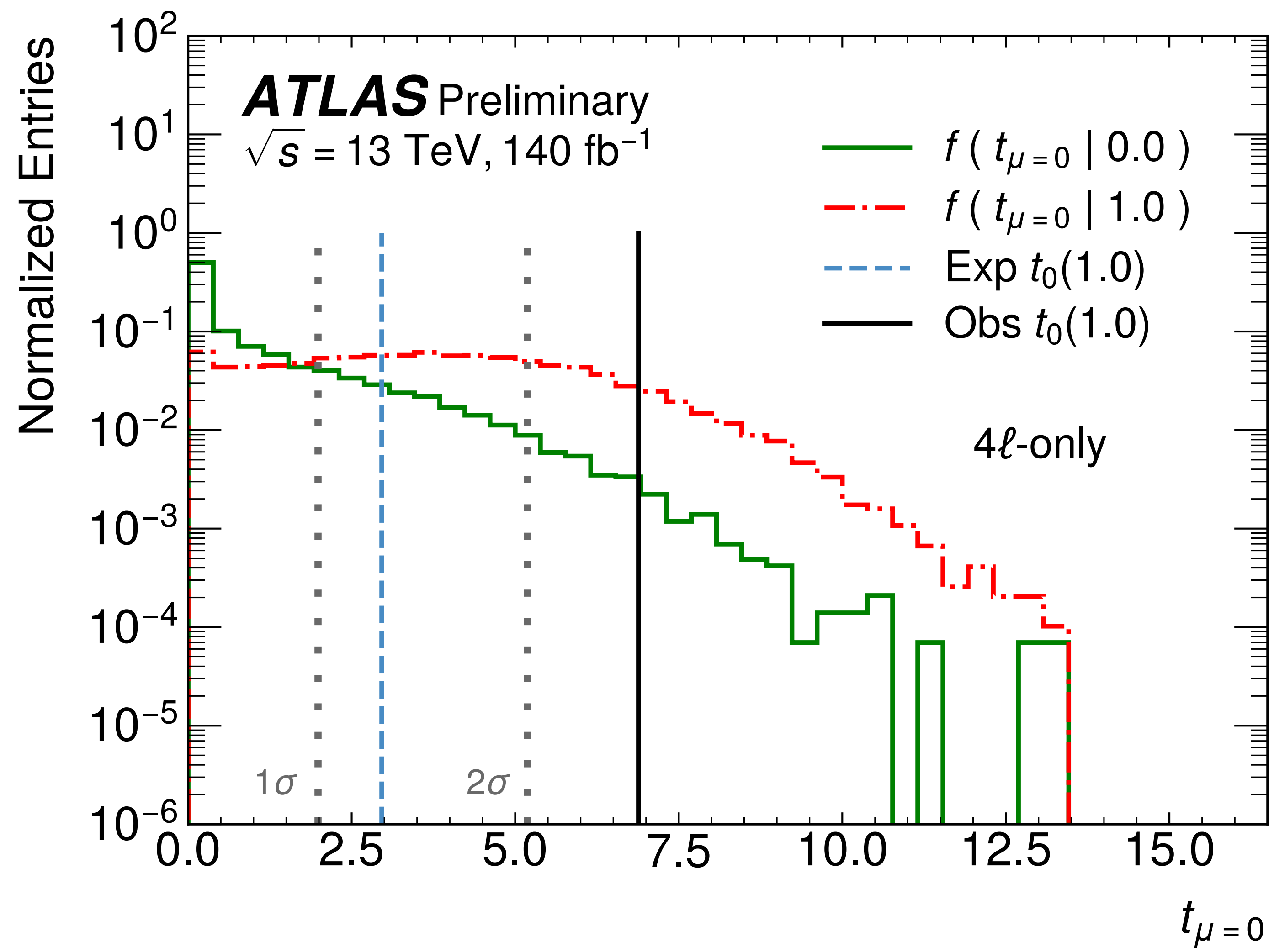
Testing the background exclusion sensitivity (in σ)

	Expected	Observed
Bkg Exclusion	1.3	2.5

Testing the SM compatibility of t_0^{obs} (in σ)

	Observed
SM compatibility	1.26

p-value = $\int_{t_0^{obs}}^{\infty} f(t_{\mu=0} \mid 1.0) = 0.11$



Pseudo-experiments are sampled from the nominal ($\mu = 1$) and bkg-only ($\mu = 0$) hypothesis to set the bkg-exclusion limits and test the SM compatibility of the observed t_0 result.

Where does the sensitivity come from? Not the tails

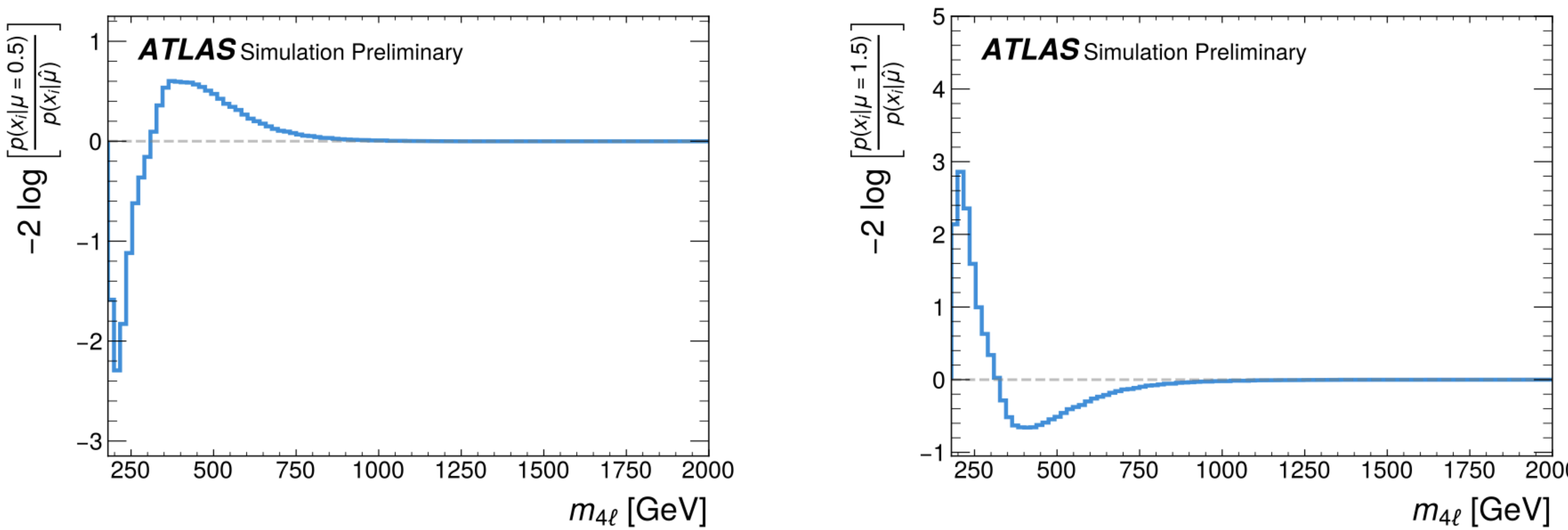


Figure 7: The sum of log density-ratios $-2\log(p(x_i|\mu')/p(x_i|\hat{\mu}))$ for events in bins of $m_{4\ell}$, for a hypothesis $\mu' = 0.5$ (left) or a hypothesis $\mu' = 1.5$ (right), with $\hat{\mu} = 1$ as the maximum likelihood estimate on an Asimov dataset generated at $\mu = 1$. This represents the per-event contribution to the test statistic for a given hypothesis, as a function of $m_{4\ell}$. Events in regions with a sum greater than zero are collectively more consistent with a $\mu = \mu'$ hypthesis over a $\mu = \hat{\mu}$ hypothesis, while regions with a sum less than zero are collectively less consistent. The very high mass region ($m_{4\ell} > 1000$ GeV) is equally consistent with both hypotheses and provides no additional sensitivity.

Parameterized Observables and Unbinning

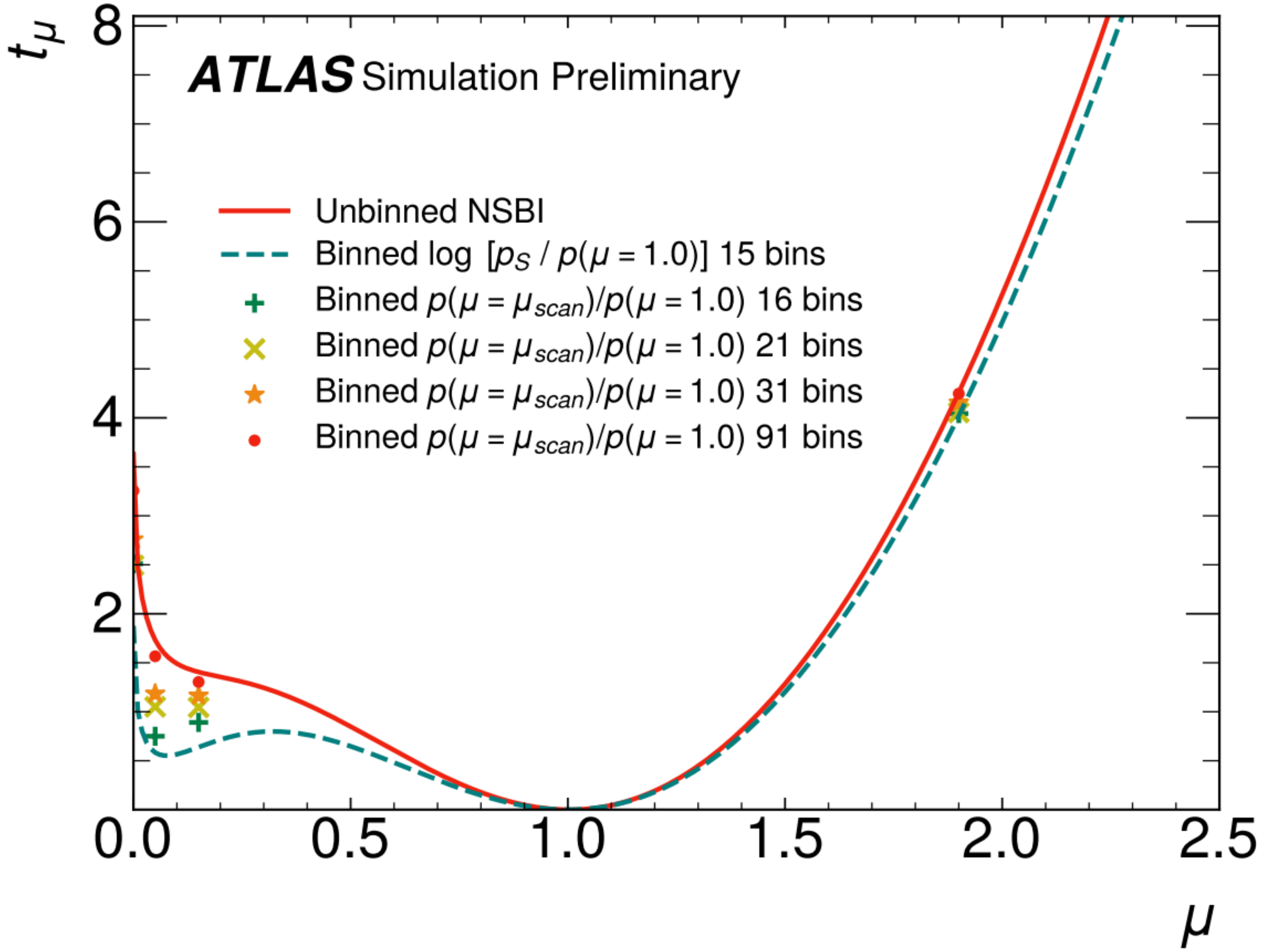
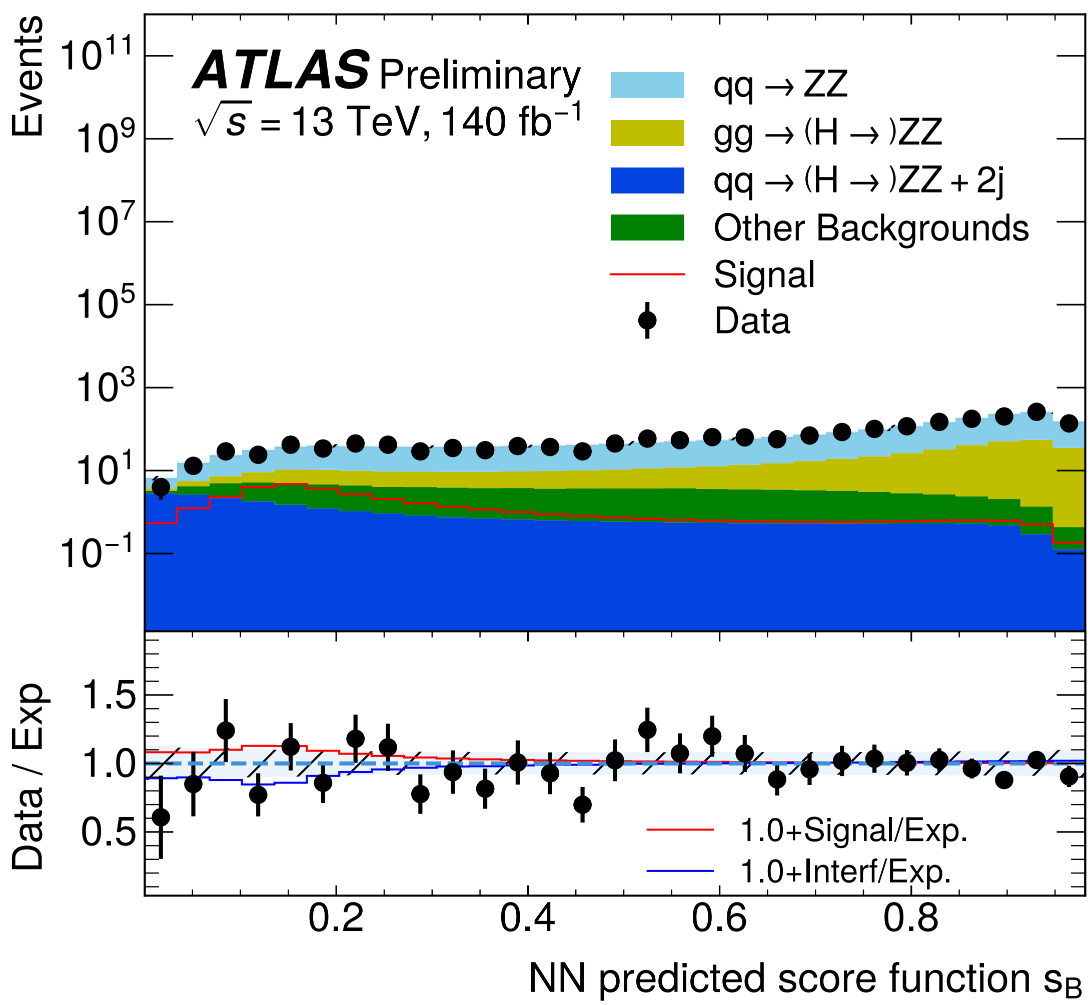
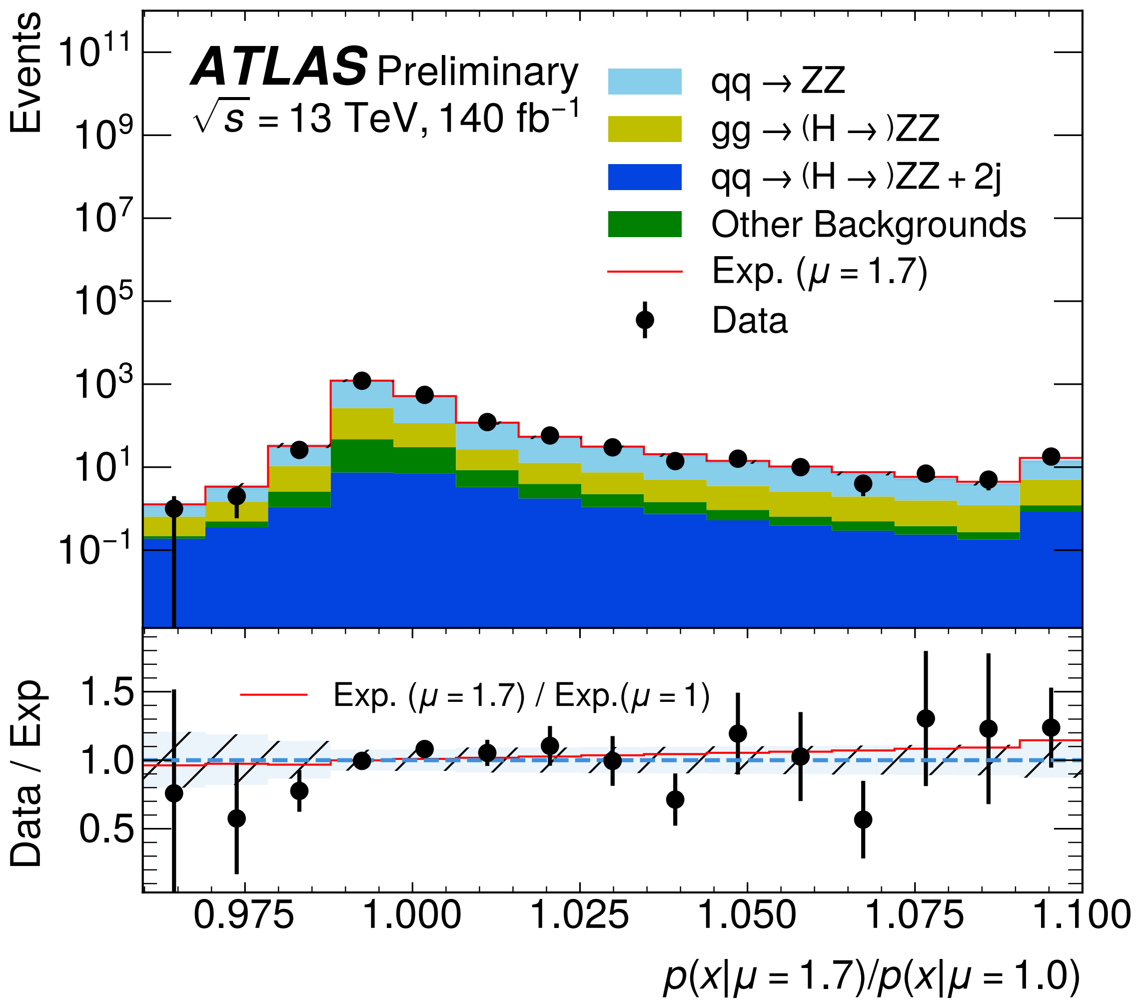


Figure 6: A comparison of expected sensitivity from various analysis strategies using the log-likelihood ratio test statistic t_μ, as a function of μ. The evaluation is performed on an Asimov dataset generated with $\mu = 1$. The red curve represents NSBI. The green curve represents a typical histogram analysis that uses a fixed observable, $\log p_s/p(x|\mu = 1)$, as a discriminant, with 15 bins. The markers show the sensitivity for various histogram analyses that use specific discriminants, $p(x_i|\mu)/p(x_i|\mu = 1)$, for specific values of $\mu(= 0.0, 0.05, 0.15, 1.9)$, with 15 (green pluses), 20 (yellow crosses), 30 (orange stars) or 90 (red dots) bins. The improved sensitivity of the green dots over the green curve (both using 15 bins) is due to the use of a parameterised observable.

Data/MC checks - NN outputs

We test the robustness of the multi-dimensional NN mapping by performing detailed data-MC validations



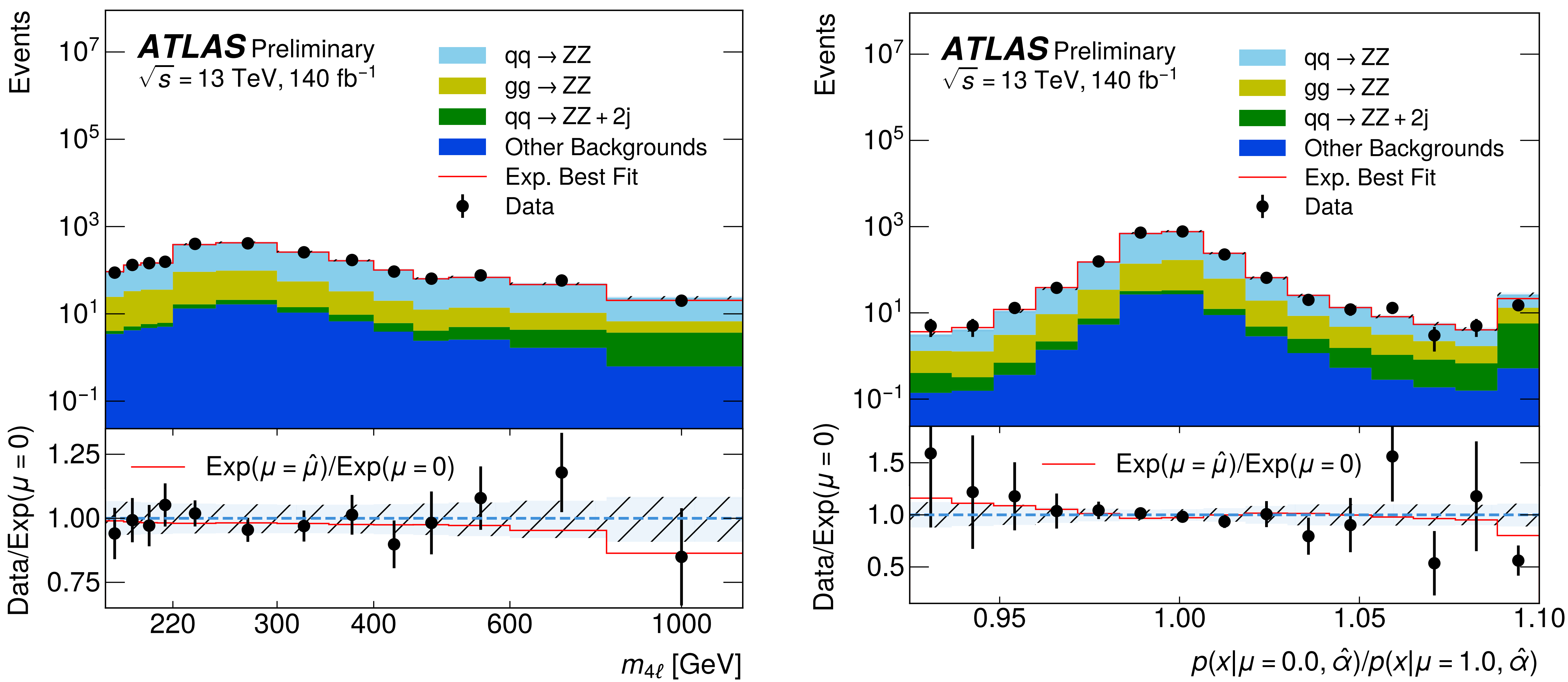


The NN output is also verified on data events from a orthogonal Control Region phase space, ensuring robustness of the mapping function.

Data/MC checks - Optimality

Following plots have data-MC comparisons with $\mu = 0$ background-only histogram stacks.

Red curves depict the distribution at best fit value $\mu \sim 0.9$



The "optimal" observable at $\mu = 0$