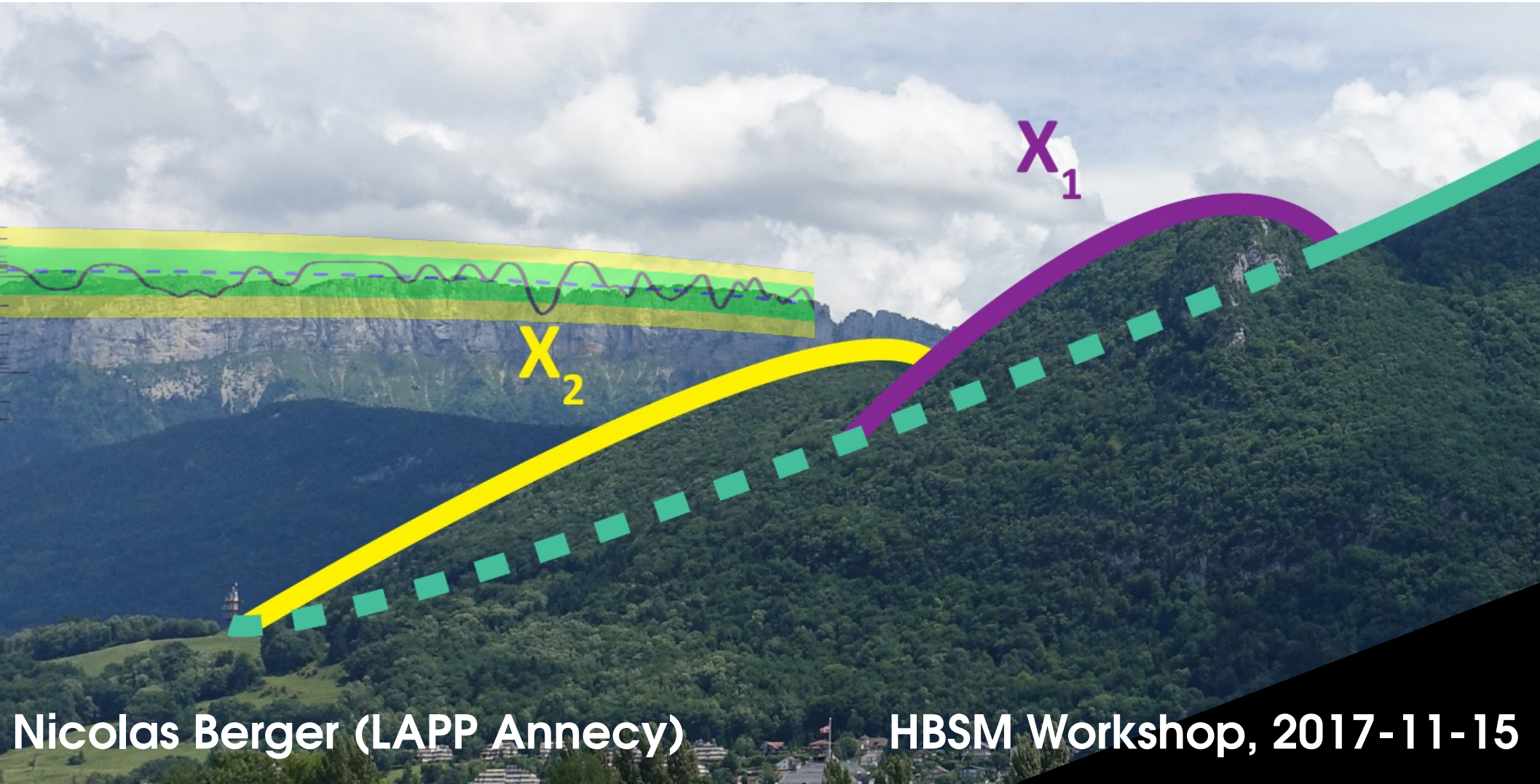
Statistical Analysis Methods



Nicolas Berger (LAPP Annecy)

HBSM Workshop, 2017-11-15

Outline

- **Profile-likelihood-based frequentist methods**
  - Likelihoods, Test Statistics, Profiling
  - Discovery – toys, LEE
  - Upper limits
- **Combination**
- **Presentation of results**

Likelihoods

Building Likelihoods

Most analyses are "shape analyses" : based on the **distribution** of a continuous observable **m**.

Two main implementations:

- **Binned Likelihood**

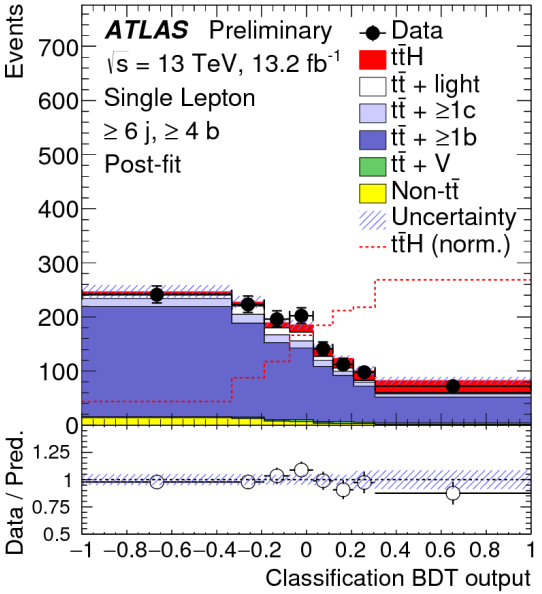
Poisson distribution in each bin

$$L\left(N_S, N_B; \{n_i\}_{i=1\ldots n_{bins}}\right) = \prod_{i=1}^{n_{bins}} e^{-N_S s_i + N_B b_i} \frac{\left(N_S s_i + N_B b_i\right)^{n_i}}{n_i!}$$

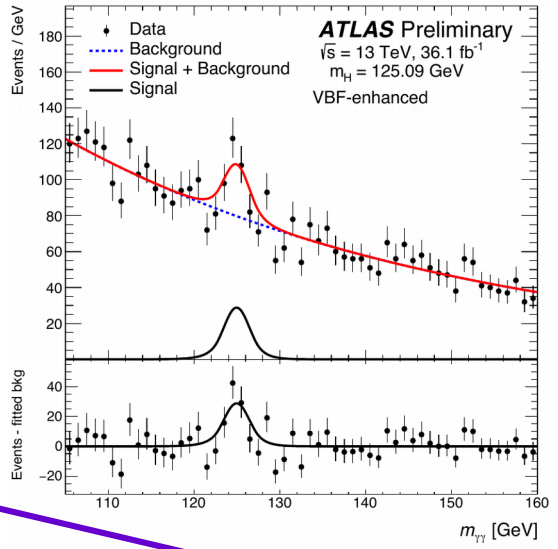
Signal yield — Bkg yield — Observed yields per bin — Per-bin fractions (=shapes) for Signal and Bkg

→ Counting analysis in the limit of $n_{bins}=1$

ATLAS-CONF-2016-080



ATLAS-CONF-2017-045



- **Unbinned Likelihood**

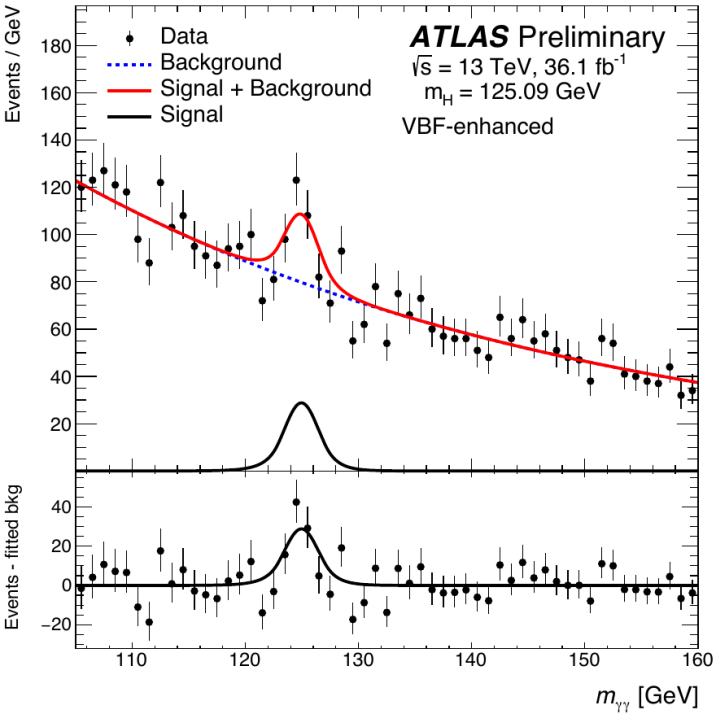
Signal yield — Bkg yield — Per-event m values

PDFs (=function describing shape) for Signal and Bkg

$$L\left(N_S, N_B, \theta; \{m_i\}_{1\ldots n_{obs}}\right) = e^{-\left(N_S+N_B\right)} \frac{\left(N_S+N_B\right)^{n_{obs}}}{n_{obs}!} \prod_{i=1}^{n_{obs}} \frac{N_S}{\left(N_S+N_B\right)} P_S\left(m_i;\theta\right) + \frac{N_B}{\left(N_S+N_B\right)} P_B\left(m_i,\theta\right.$$

Nuisances and Systematics

ATLAS-CONF-2017-045



Likelihood typically includes

- **Parameters of interest** (POIs) : N_S, $\sigma \times B$, m_W, …
- **Nuisance parameters** (NPs) : other parameters needed to define the model
  → Ideally, constrained by data like the POI
  e.g. shape of H→γγ continuum bkg

What about systematics ?

→ Unknown parameters ⇒ **simply additional NPs**

→ By definition, **not constrained by the data** – need external input

– e.g. a σ×B measurement needs a priori knowledge of the luminosity

⇒ **For systematics, need to add an external constraint in the likelihood:**

$$L(\mu, \theta; data) = L_{measurement}(\mu, \theta; data)\ C(\theta)$$

POI

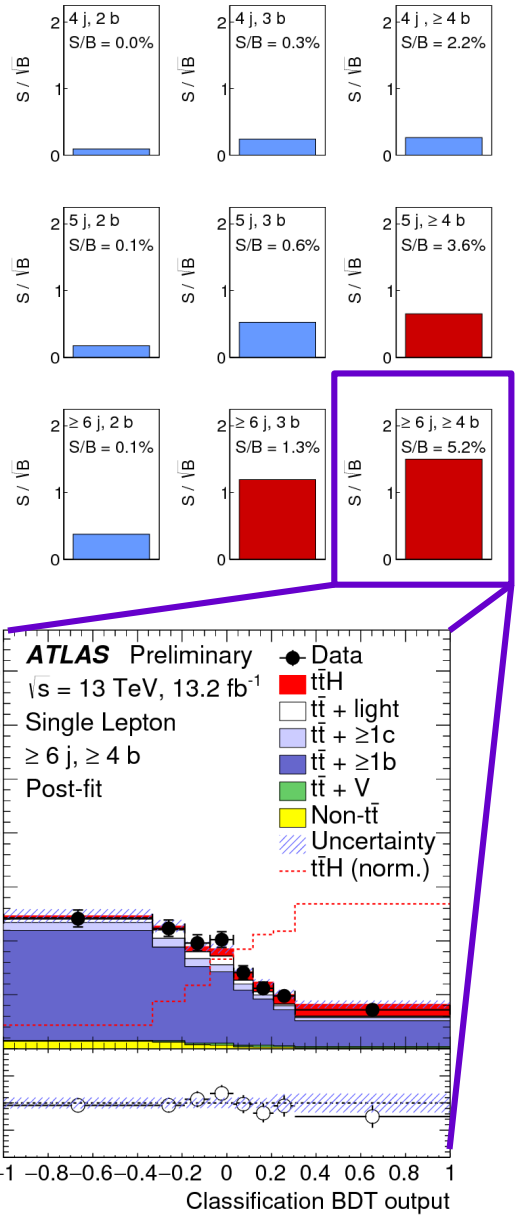
Systematics NP

Measurement Likelihood

Constraint term: "Auxiliary experiment" – typically Gaussian

Categories

Example: ttH→bb
(ATLAS-CONF-2016-080)



Multiple analysis regions often used:

- Decay modes
- Kinematic selections, etc.

→ Useful to model these **separately** if

- Sensitivity is better in some regions (avoids dilution)
- Some regions can constrain NPs
  - e.g. ***Control regions*** for backgrounds

→ However analyze them **simultaneously** to model correlations between the regions (common NPs)

- Better than a-posteriori combination

No stat correlations ⇒ can simply take likelihood products

Likelihood for category k

$$L\left(\mu,\theta;\{n_i^{(k)}\}_{i=1\ldots n_{data}^{(k)}}^{k=1\ldots n_{cat}}\right)=\prod_{k=1}^{n_{cat}} L_k\left(\mu,\theta;\{n_i^{(k)}\}_{i=1\ldots n_{data}^{(k)}}\right)$$

Combinations

- A combination is just another another analysis with categories, no fundamental difference.
- However some work often needed to make sure originally-distinct analyses work well together:
  - **No selection overlaps**: generally assume stat uncertainties uncorrelated
    → Need to make sure analysis selections are orthogonal (~ or < %-level)
    → Checking only data not enough! (especially if few events selected)
  - **Same POI model**
  - **Compatible NP model**: all systematics need to be either **fully correlated** (same NP) or **uncorrelated** (different NPs) across analyses
    → often different, incompatible NP sets chosen by different analysis (e.g. JES, FT) ⇒ Need advance planning to ensure compatibility
    → common backgrounds may also need to be correlated
- **Tools:**
  - can work directly with e.g. roofit/roostats/histfactory
  - tools already developed to deal with such cases, some listed here
    - Recent Higgs combination using the **workspaceCombiner** tool to edit workspaces, "stitch them together" into a single object, etc.

Test Statistics and Profiling

Cowan, Cranmer, Gross & Vitells, Eur.Phys.J.C71:1554,2011

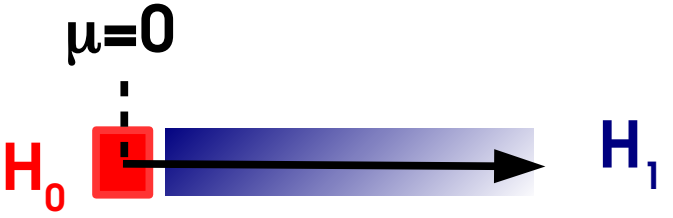
Test Statistic & Asmptotics

Cowan, Cranmer, Gross & Vitells,
Eur.Phys.J.C71:1554,2011

Neyman-Person lemma : the **likelihood ratio** $L(H_0)/L(H_1)$ is the optimal discriminator when testing hypothesis H_1 vs. H_0.

→ e.g. **New physics search –** test:

- H_1: presence of a signal (any **μ>0**) against
- H_0 : background only (**μ=0**)

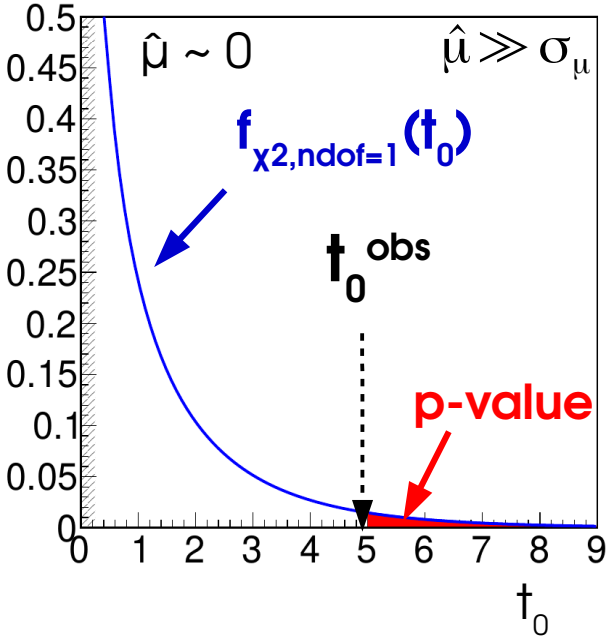


→ Use as test statistic $$t_0 = -2\log\frac{L(\mu=0)}{L(\hat{\mu})}$$

Large values of t_0 ⇔ large observed $\hat{\mu}$.

– Compatibility with H_0: **p-value**

Asymptotic approximation = Gaussian regime for μ
⇒ t_0 is **distributed as a χ^2** under the hypothesis μ=0:



Profiling & Wilks' Theorem

The likelihood usually has NPs

- Systematics
- Parameters fitted in data

→ **What values to use when defining the hypotheses ?** → H(μ=0, θ=?)

Answer: **let the data choose ⇒ use the best-fit values (*Profiling*)**

⇒ **Profile Likelihood Ratio** (PLR)

$$t_{\mu_0} = -2\log\frac{L\left(\mu=\mu_0, \hat{\hat{\theta}}_{\mu_0}\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}$$

$\hat{\hat{\theta}}_{\mu_0}$ best-fit value for $\mu=\mu_0$ (conditional MLE)

$\hat{\theta}$ overall best-fit value (unconditional MLE)

Wilks' Theorem: PLR also follows a χ^2 !

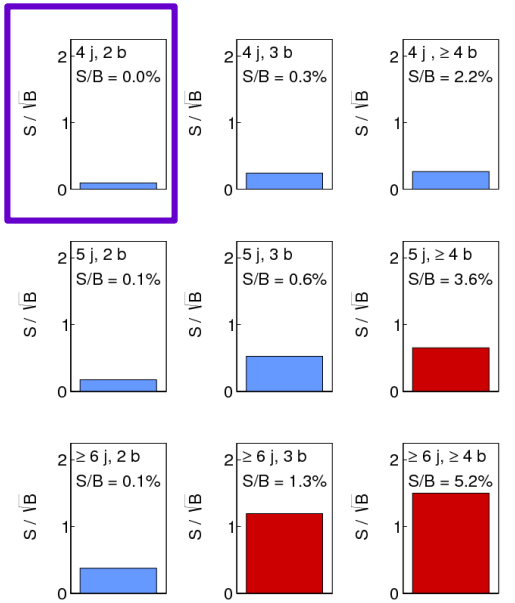
→ Profiling "builds in" the effect of the NPs

⇒ Can treat the PLR as a function of the POI only

$$f\left(t_{\mu_0} \,|\, \mu=\mu_0\right) = f_{\chi^2(n_{dof}=1)}\left(t_{\mu_0}\right)$$

also with NPs present

Profiling Example: ttH→bb

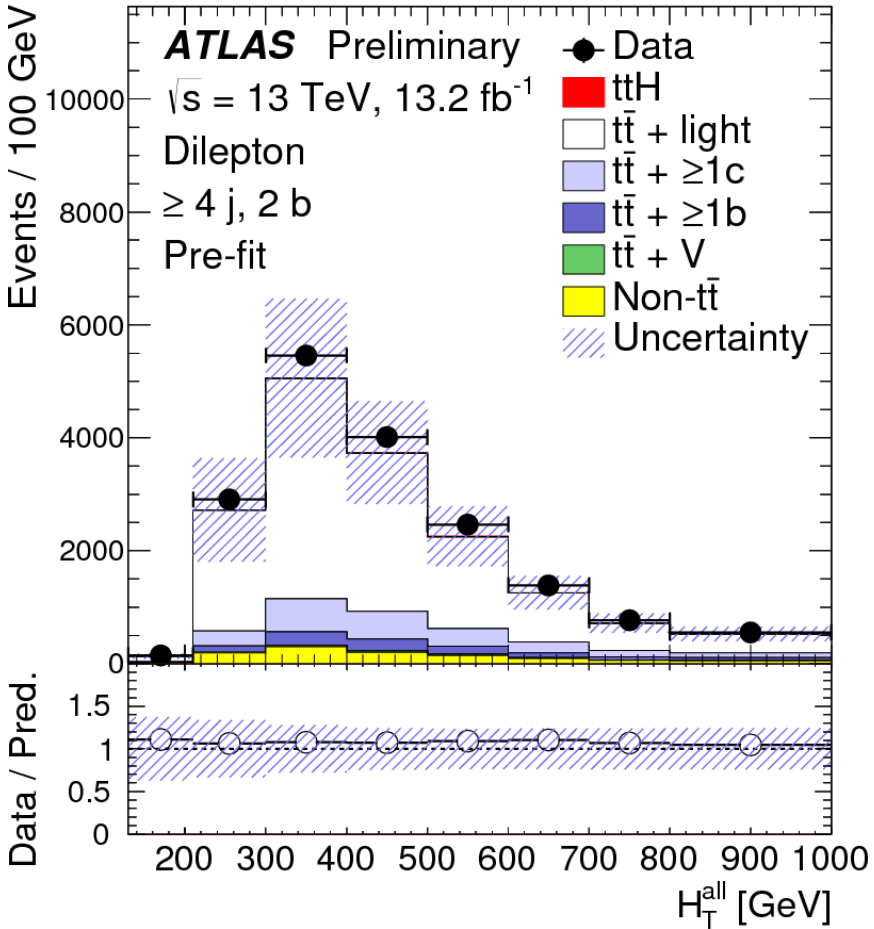


Analysis uses low-S/B categories to constrain backgrounds.

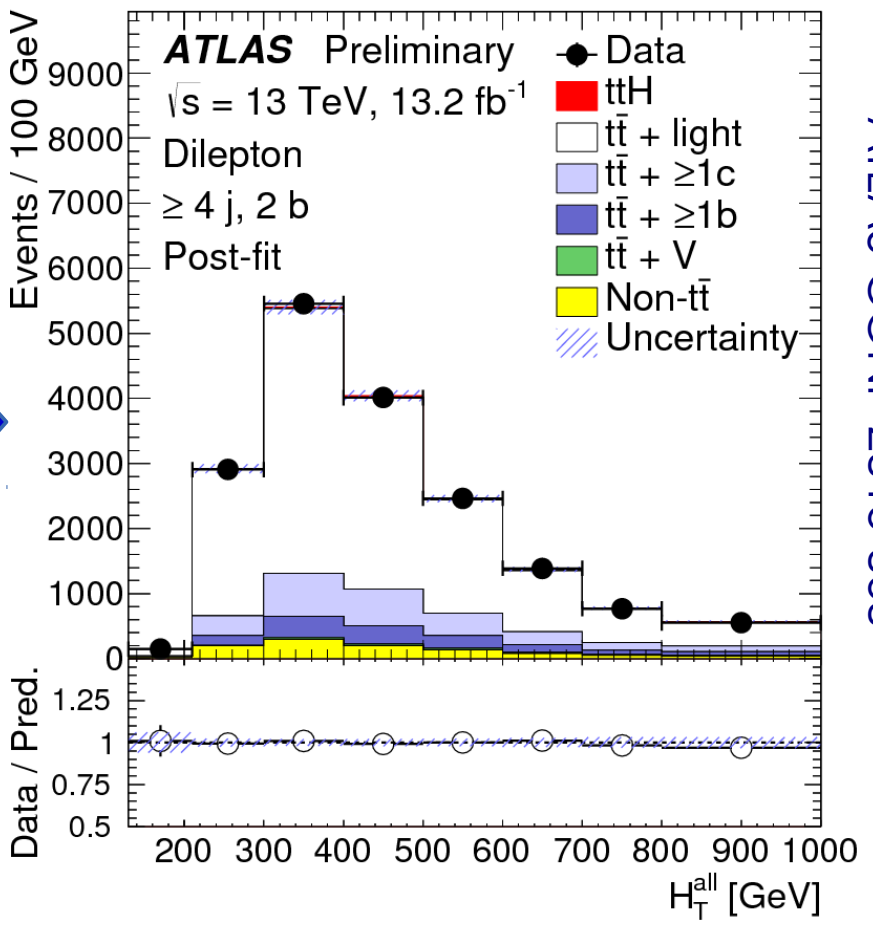
→ **Reduction in large uncertainties on tt bkg NPs**

→ **Propagates to the high-S/B categories** through the statistical modeling

⇒ **Care needed in the propagation** (e.g. different kinematic regimes)



Fit



ATLAS-CONF-2016-080

Pull/Impact plots

ATLAS-CONF-2016-058

Nominally systematics NPs have:

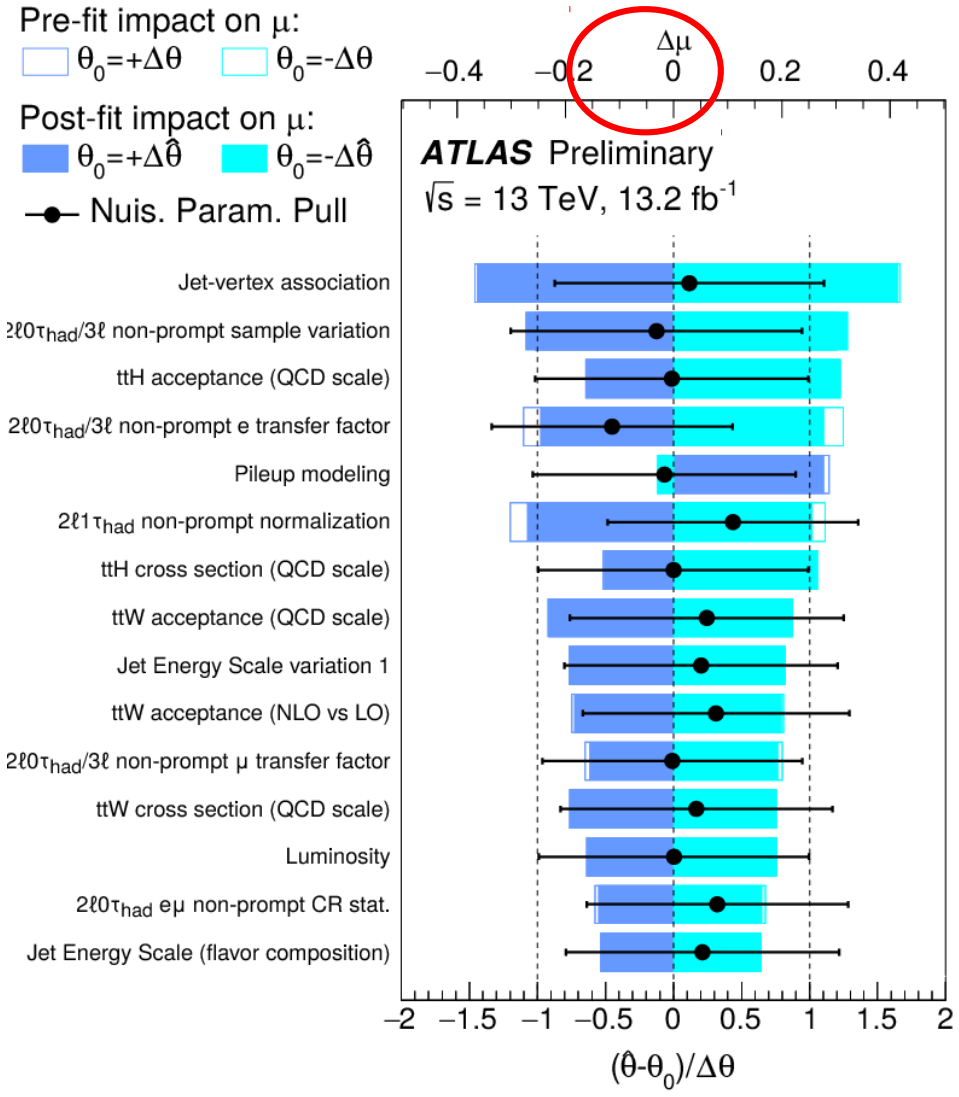
- **Central value = 0** : i.e. the pre-fit expectation
- **Uncertainty = 1** : NPs normalized to the value of the systematic

From fit results:

- **If central value ≠ 0**: some data feature absorbed by nonzero value
  ⇒ Need investigation if large pull

- **If uncertainty < 1** : systematic is constrained by the data
  ⇒ Needs checking if this legitimate or a modeling issue

- **Impact on result** of a ±1σ NP shift



Pull/Impact plots

Nominally systematics NPs have:

- **Central value = 0** : i.e. the pre-fit expectation
- **Uncertainty = 1** : NPs normalized to the value of the systematic

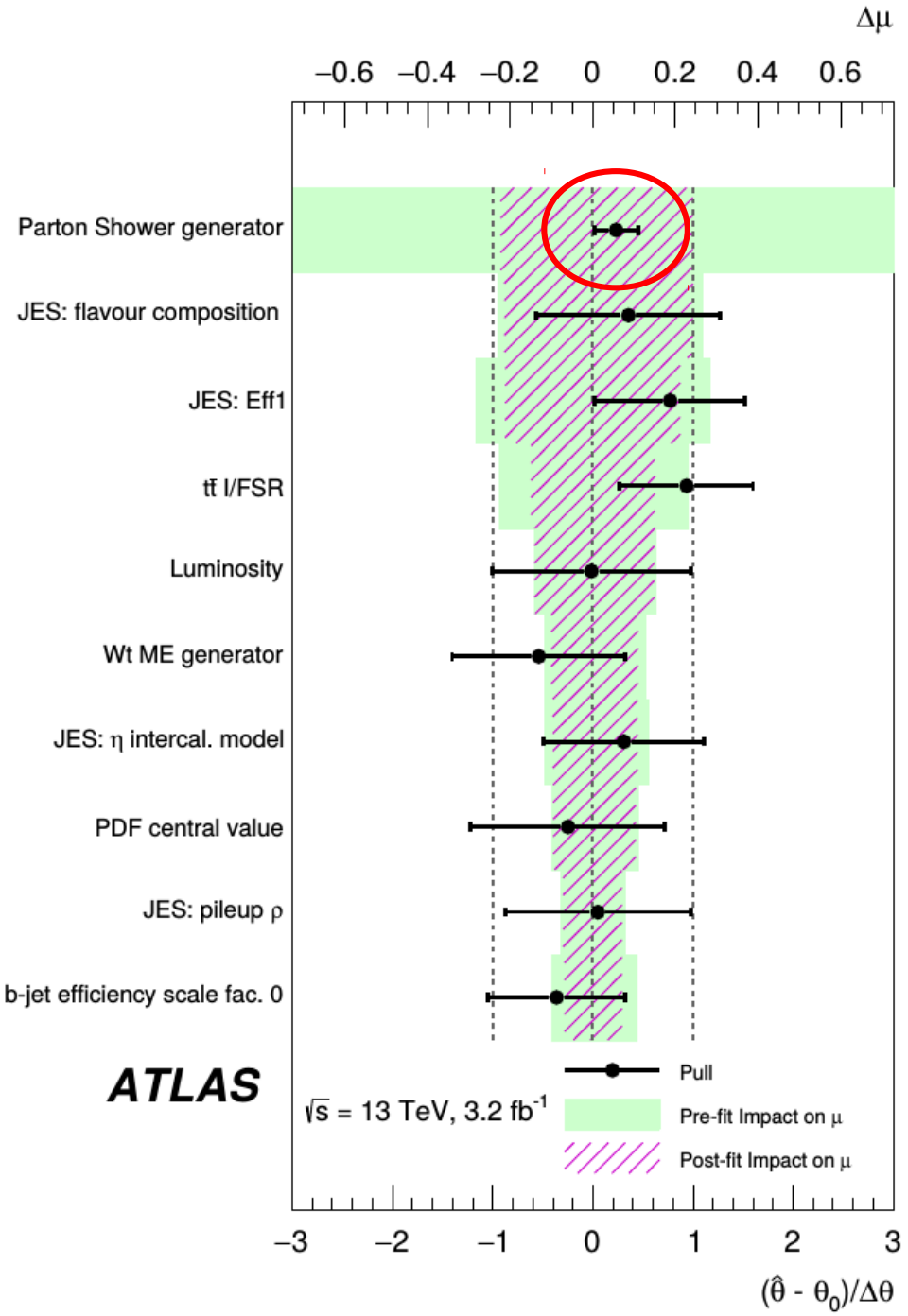
From fit results:

- **If central value ≠ 0**: some data feature absorbed by nonzero value
  ⇒ Need investigation if large pull

- **If uncertainty < 1** : systematic is constrained by the data
  ⇒ Needs checking if this legitimate or a modeling issue

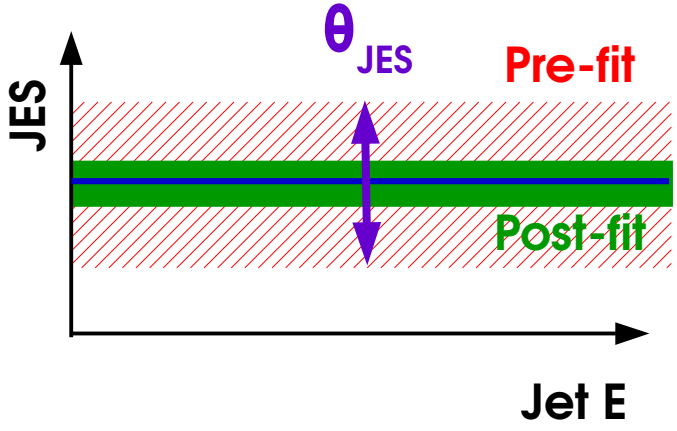
- **Impact on result** of a ±1σ NP shift

13 TeV single-t XS (arXiv:1612.07231)

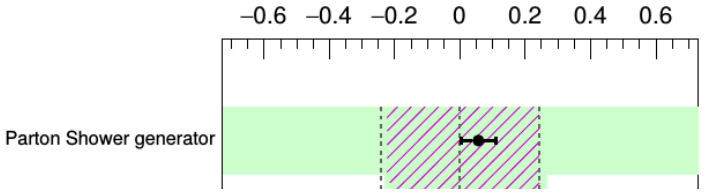


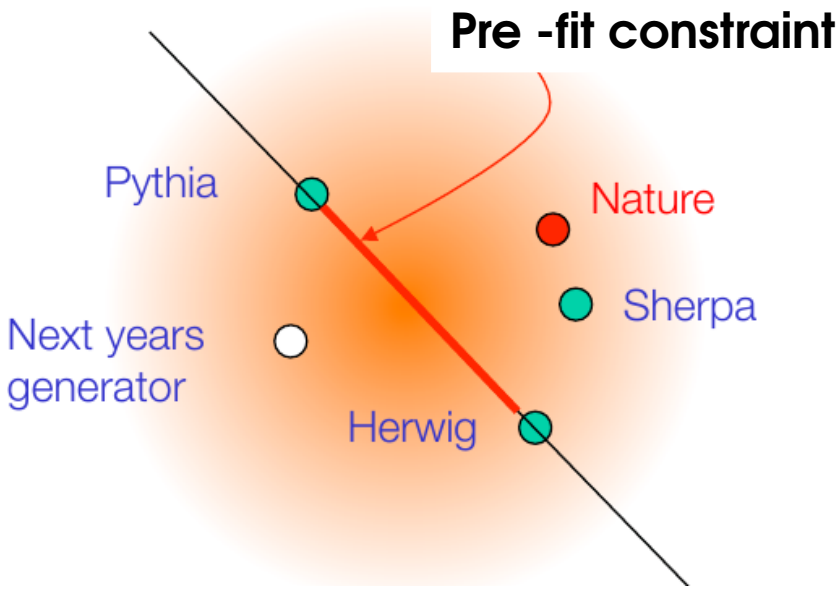
Profiling Issues

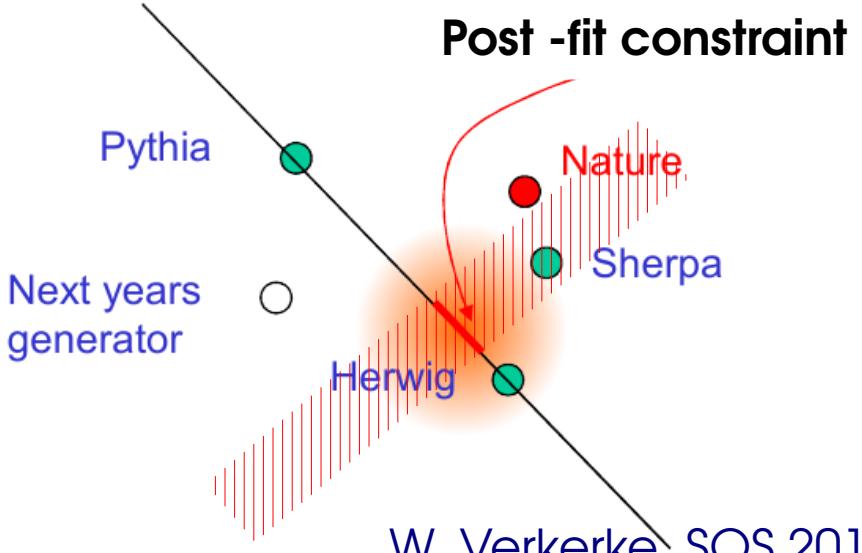
Too simple modeling can have unintended effects
→ e.g. single Jet E scale parameter:
⇒ Low-E jets calibrate high-E jets – intended ?



***Two-point* uncertainties**: interpolate between 2 discrete cases
→ e.g. Pythia vs. Herwig
→ Interpolation may not cover full configuration space, can lead to too-strong constraints



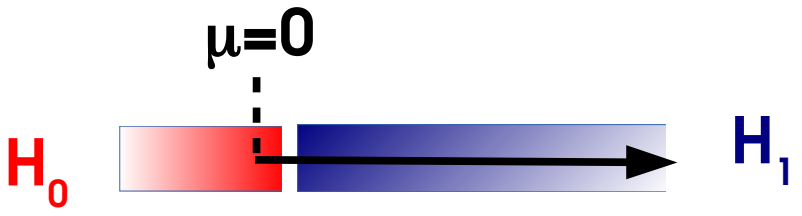




Computing Statistical Results: Discovery

Cowan, Cranmer, Gross & Vitells, Eur.Phys.J.C71:1554,2011

Test Statistic for Discovery



Discovery: test against $H_0(\mu=0)$

→ **One-sided test**: only consider μ>0 as a legitimate signal

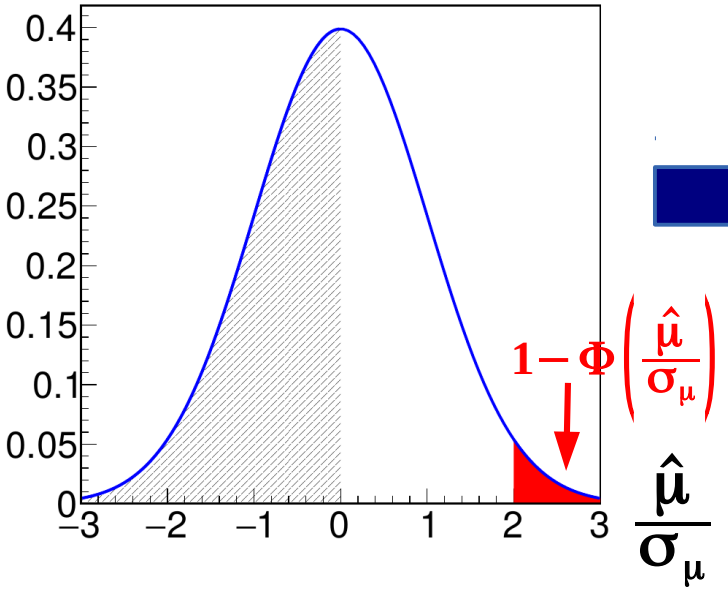
- $\hat{\mu}$ **<0** q_0 set to 0 → perfect agreement with μ=0 hypothesis

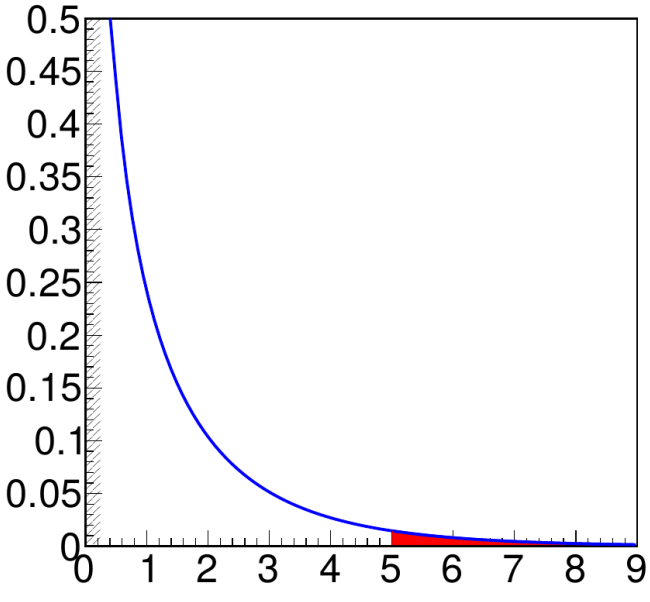
$$q_0 = \begin{cases} -2\log\dfrac{L(\mu=0,\hat{\hat{\theta}}_0)}{L(\hat{\mu},\hat{\theta})} & \hat{\mu}\geq 0 \\ 0 & \hat{\mu}<0 \end{cases}$$

Asymptotics: "half-χ^2" distribution: $f(q_0|\mu=0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}f_{\chi^2(n_{dof}=1)}(q_0)$

Discovery p-value: $p_0 = 1-\Phi\left(\sqrt{q_0}\right)$ **Significance**: $Z = \Phi^{-1}(1-p_0) = \sqrt{q_0}$

Φ : Gaussian CDF

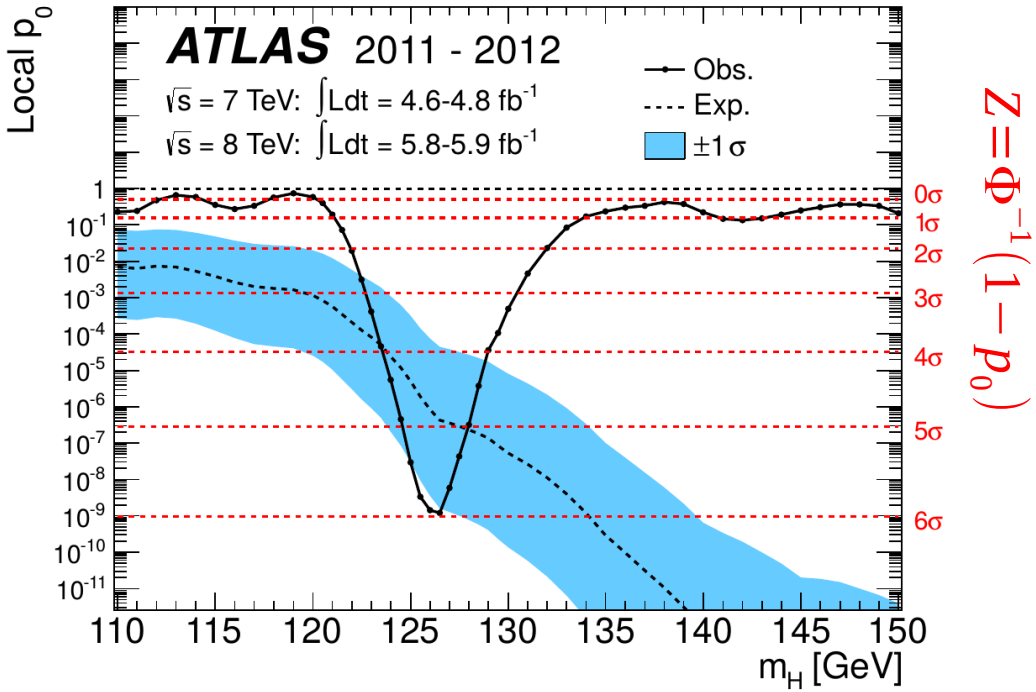




$$q_0 = \left(\frac{\hat{\mu}}{\sigma_\mu}\right)^2$$

Some Examples

Higgs Discovery: Phys. Lett. B 716 (2012) 1-29

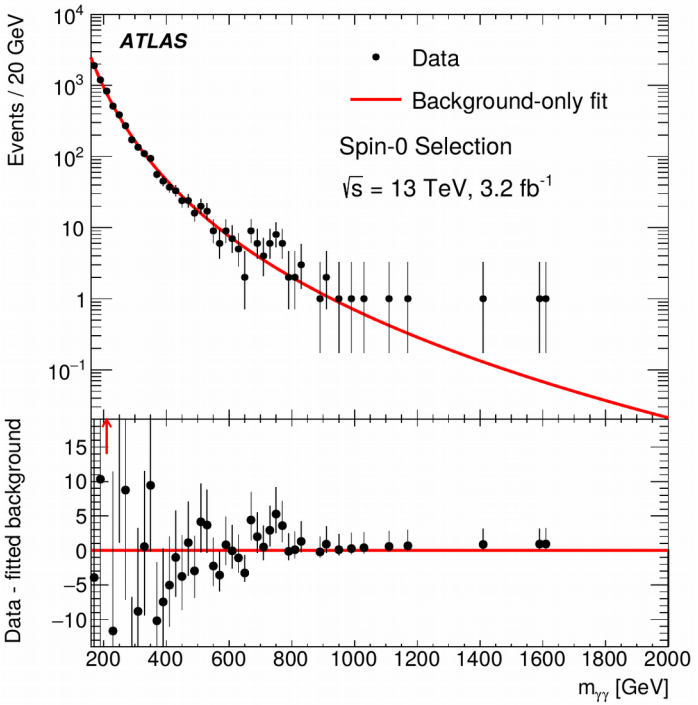


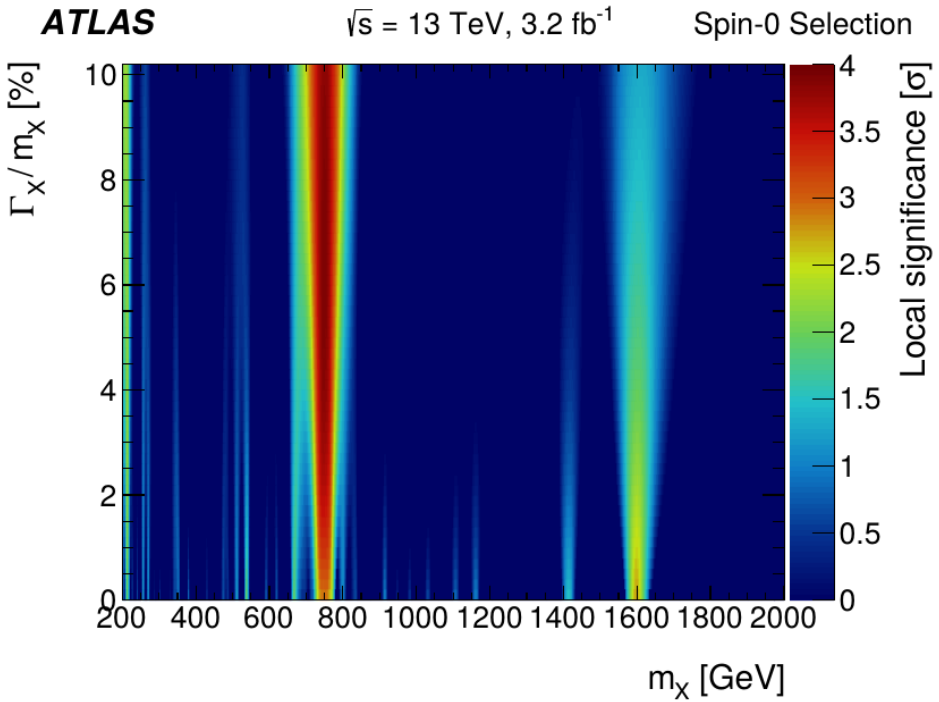
$p_0 = 1.8 \times 10^{-9} \Leftrightarrow Z = 5.9\sigma$

"Uncapped" p-values

$$q_0^{\text{uncapped}} = \begin{cases} -2\log\dfrac{L(\mu=0,\hat{\hat{\theta}}_0)}{L(\hat{\mu},\hat{\theta})} & \hat{\mu}\geq 0 \\ +2\log\dfrac{L(\mu=0,\hat{\hat{\theta}}_0)}{L(\hat{\mu},\hat{\theta})} & \hat{\mu}<0 \end{cases}$$

High-mass X→γγ Search: JHEP 09 (2016) 1





Beyond Asymptotics: Toys

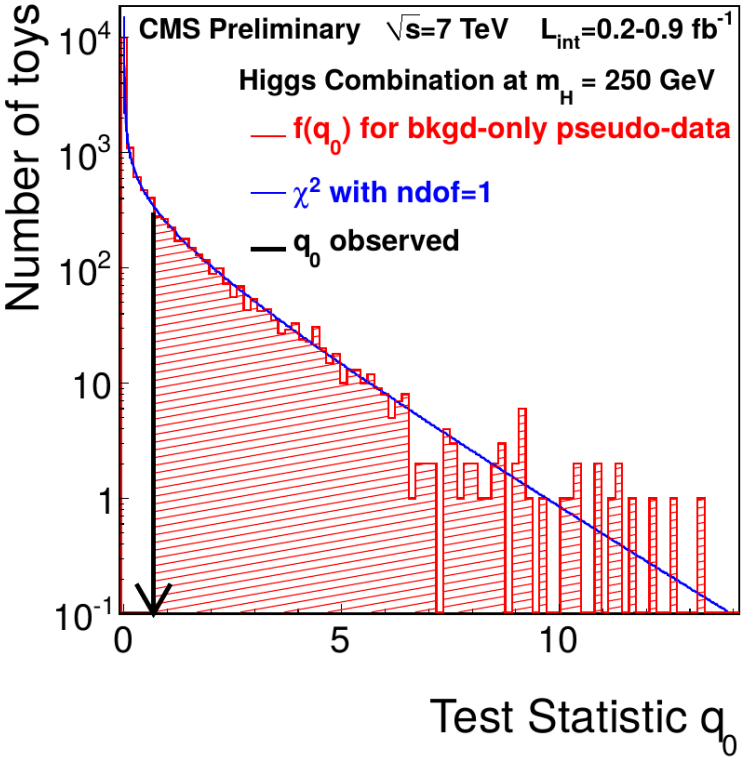
Asymptotics usually work well, but break down in some cases – e.g. **small event counts**

⇒ **Solution**: generate ***pseudo data*** ("Toys") using the PDF, under the tested hypothesis

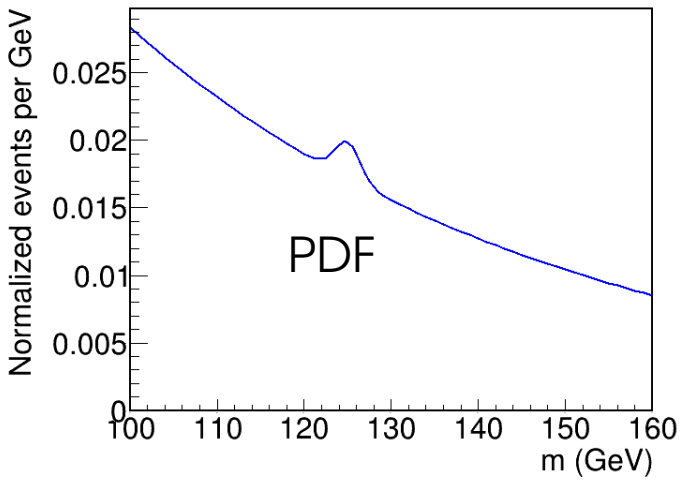
→ Samples the true distribution of the PLR

→ Integrate above observed PLR to get the p-value

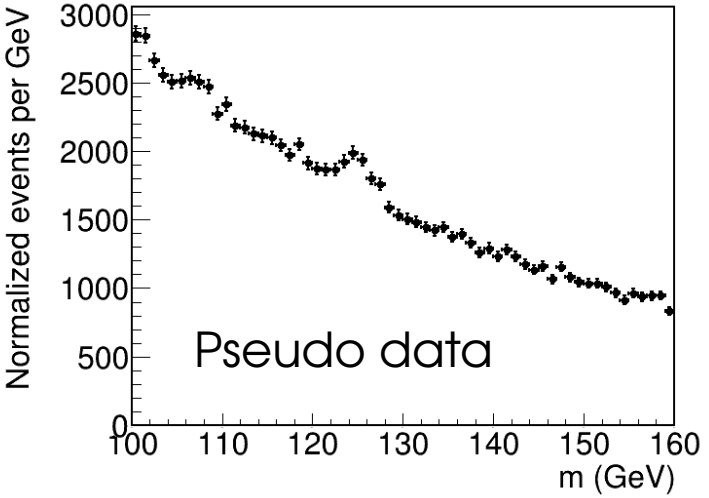
→ Need large toy samples, especially for small p-values (5σ : $p \sim 10^{-7}$!)



Repeat N_{toys} times



P(data | x)



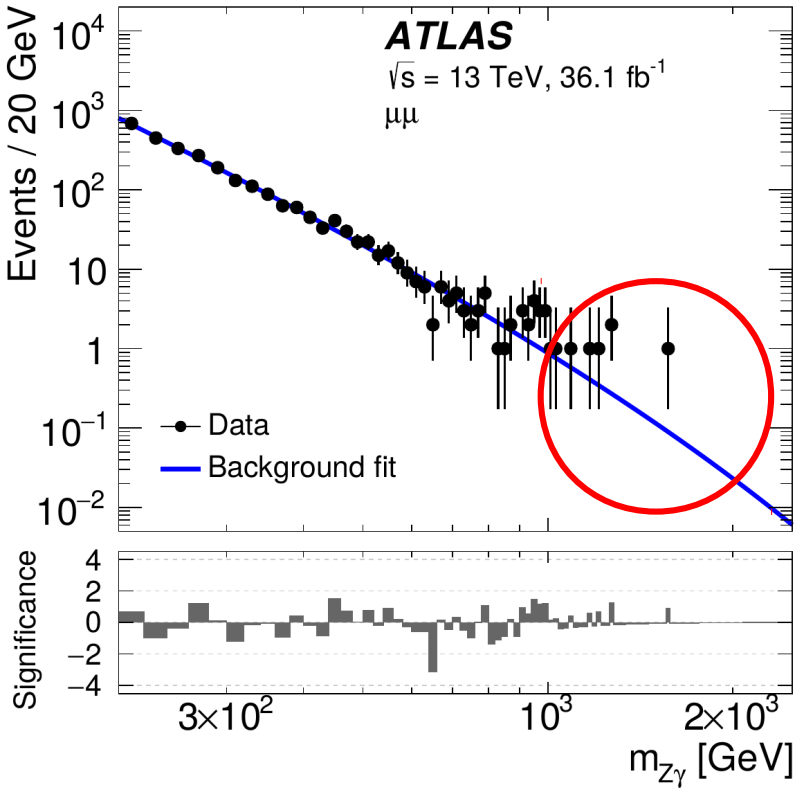
q_0

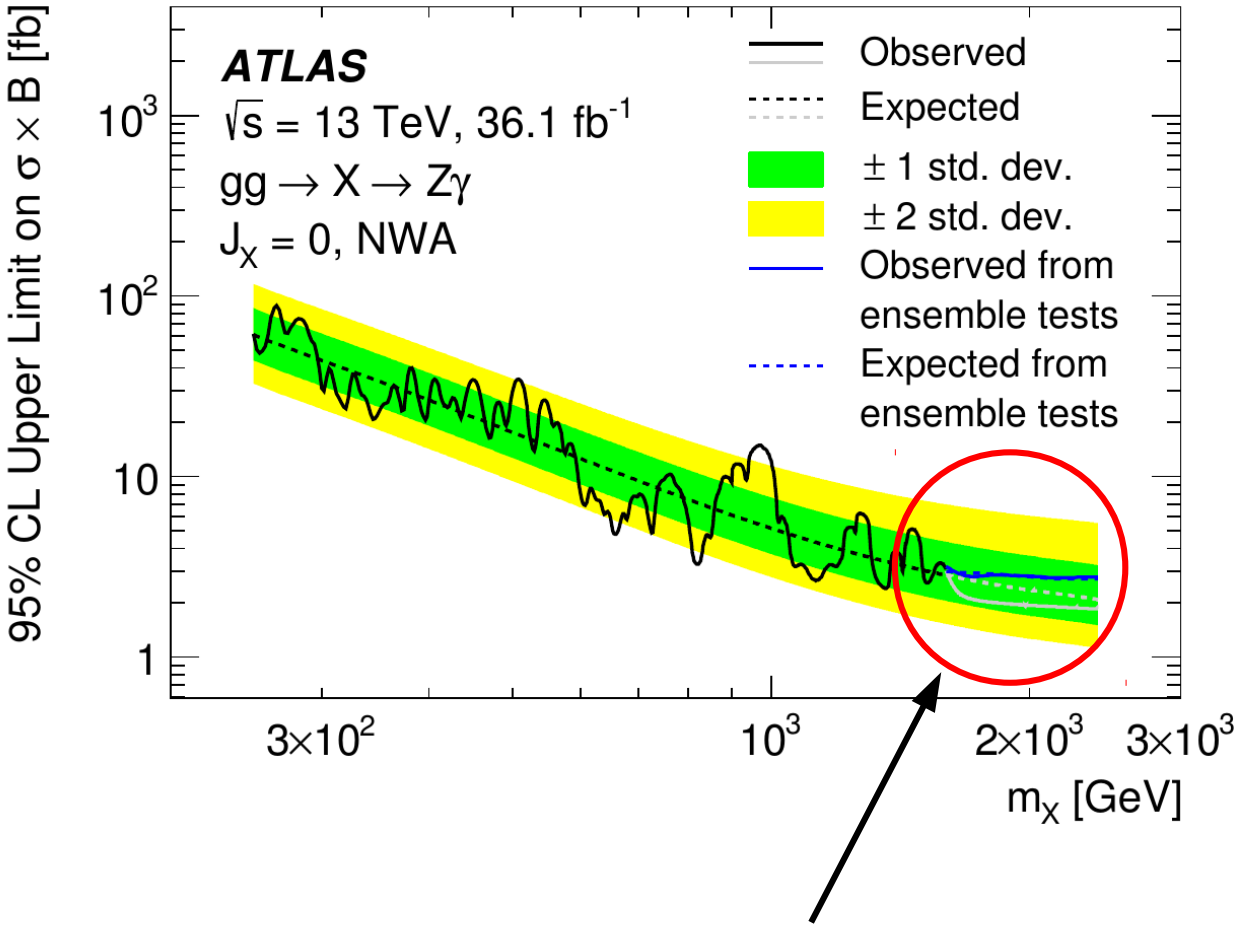
Toys: Example

arXiv:1708.00212

X→Zγ Search: covers 200 GeV < m_X < 2.5 TeV

→ for m_X > 1.6 TeV, low event counts ⇒ derive results from toys





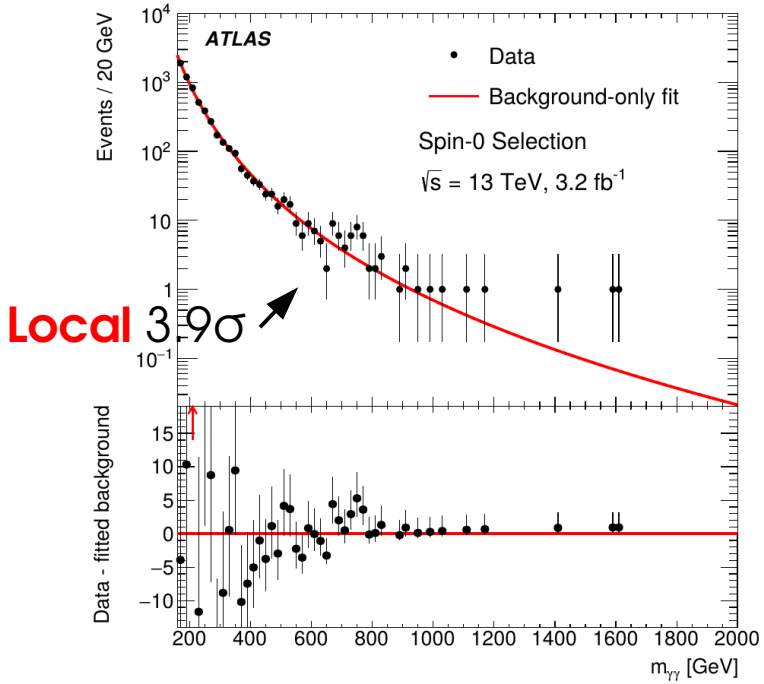
Asymptotic results (**in gray**) give optimistic result compared to toys (**in blue**)

Look-Elsewhere effect

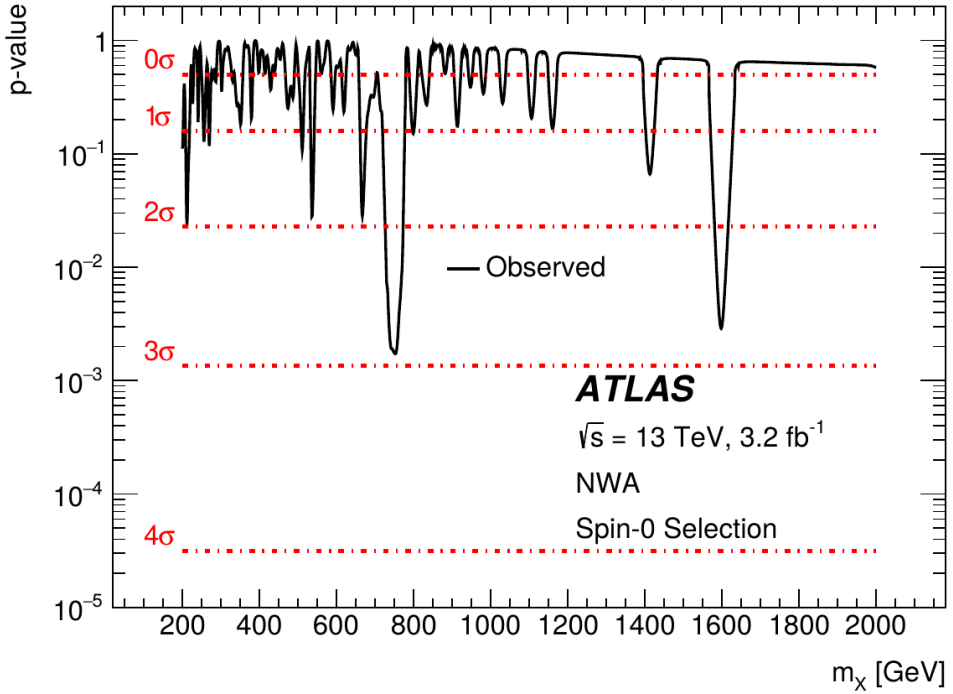
Sometimes, unknown parameters in signal model e.g. p-values as a function of m_X

⇒ Effectively performing **multiple, simultaneous searches**

→ If e.g. small resolution and large scan range, **many independent experiments**



→ More likely to find an excess **somewhere:** ***Look-elsewhere effect*** (LEE)



Global p-value (and Z_{global}) for finding an excess somewhere in the search range

→ relevant measure for searches over a range

→ $N = p_{global}/p_{local}$: ***Trials factor***

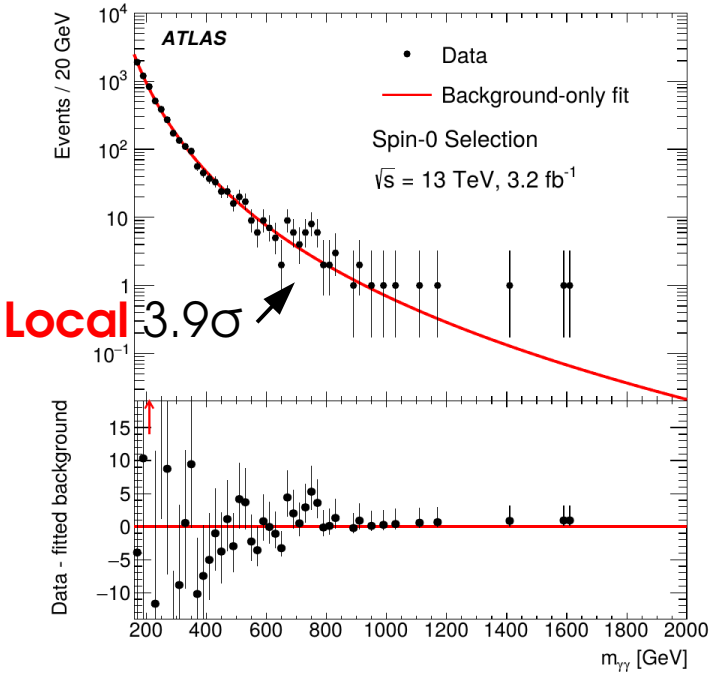
naively (range)/(resolution)

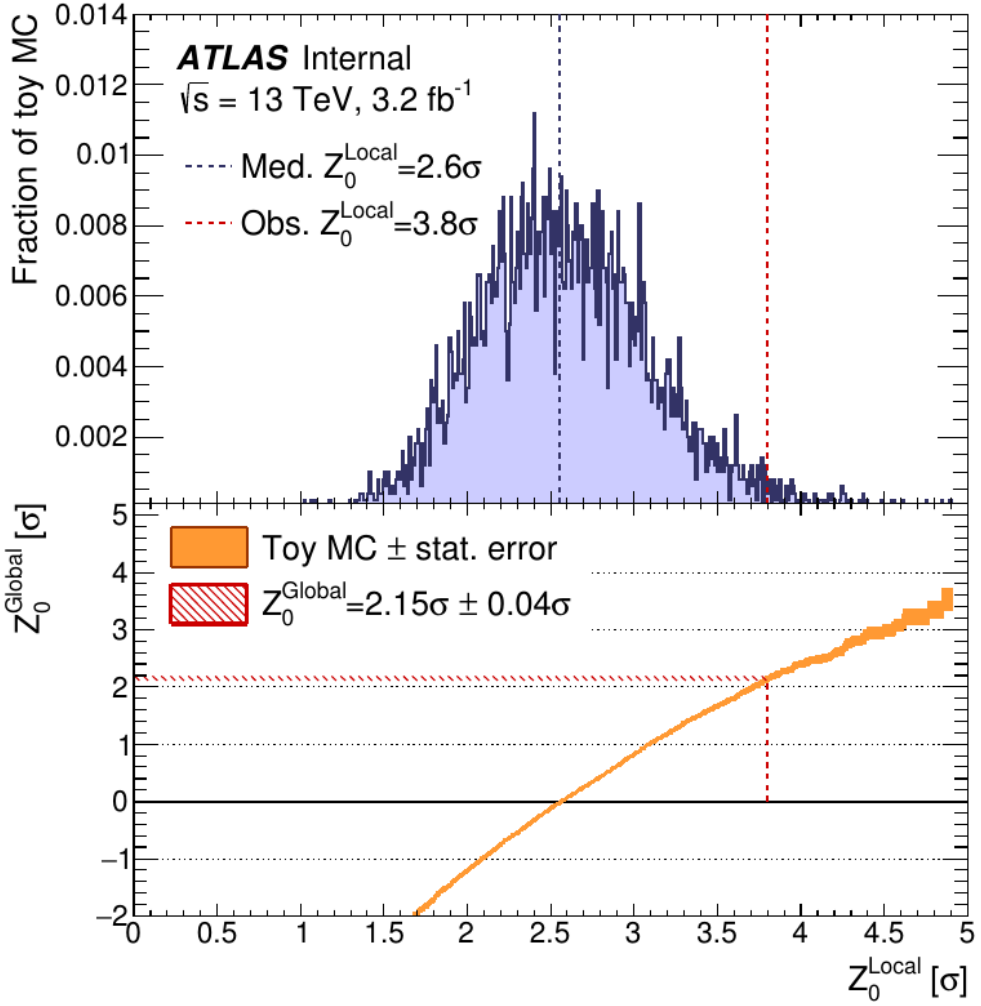
Global Significance Computation

Two main methods:

- **Brute-force toys**:
  - generate pseudo-data,
  - repeat analysis each time (scanning over parameters to find largest excess).

→ **Exact treatment but CPU-intensive**, especially for large Z





X→γγ Search: 200 < m_X < 2000 GeV, 0 < Γ_X < 10% m_X

→ Z_{local} = **3.9σ** (p_{local} ~ 5 10^{-5})

→ Z_{global} = **2.1σ** (p_{global} ~ 2 10^{-2})

(**N_{trials} ~ 400**) Less exciting...

Global Significance Computations

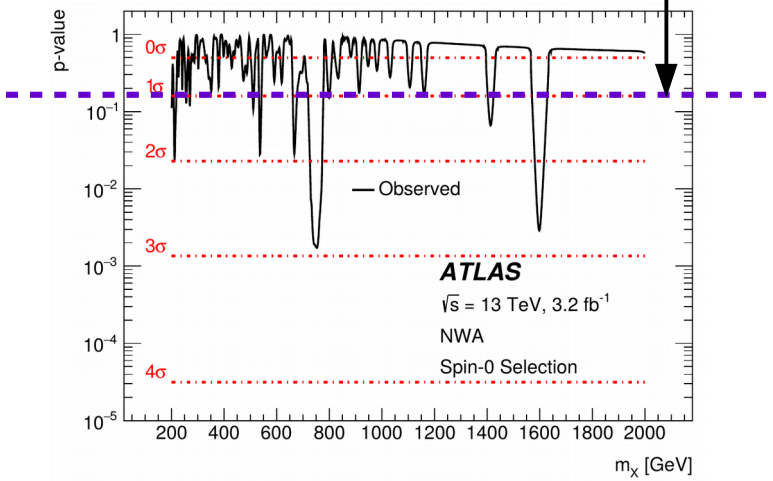
Two main methods:

- **Asymptotic approximation of the trials factor**: Gross & Vitells, EPJ.C70:525-530,2010

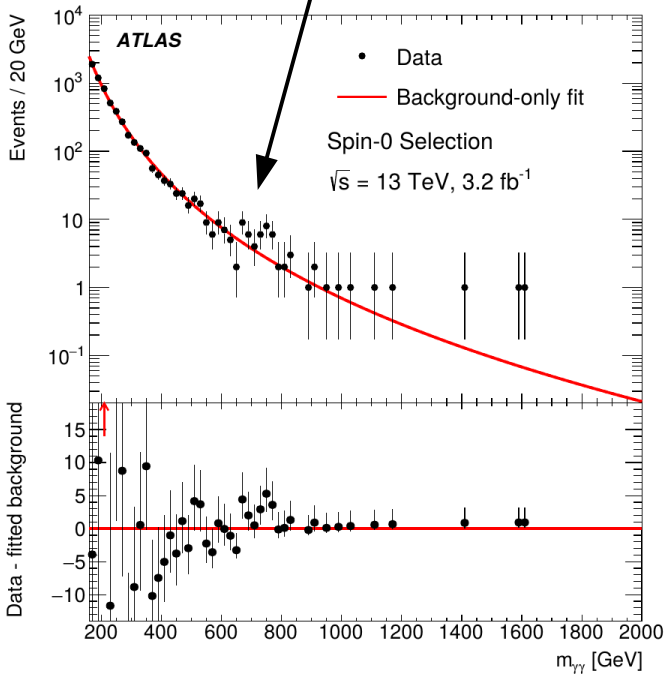
  → Use asymptotics to extrapolate to desired Z_{local} from a lower significance Z_{test} (less toys!)

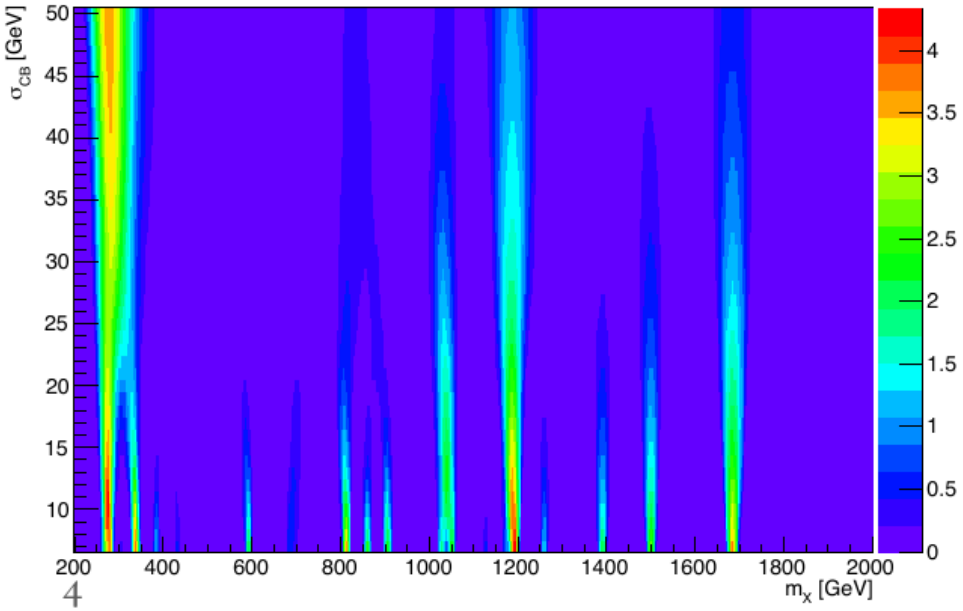
$$N_{trials} = 1 + \frac{1}{p_{local}} \langle N_{bump}(Z_{test}) \rangle \, e^{\frac{Z_{local}^2 - Z_{test}^2}{2}}$$

Average number of excesses with $Z > Z_{test}$



Local 3.9σ





2D: use Euler χ of the regions $Z > Z_{test}$

Computing Statistical Results
Upper Limits

Cowan, Cranmer, Gross & Vitells, Eur.Phys.J.C71:1554,2011

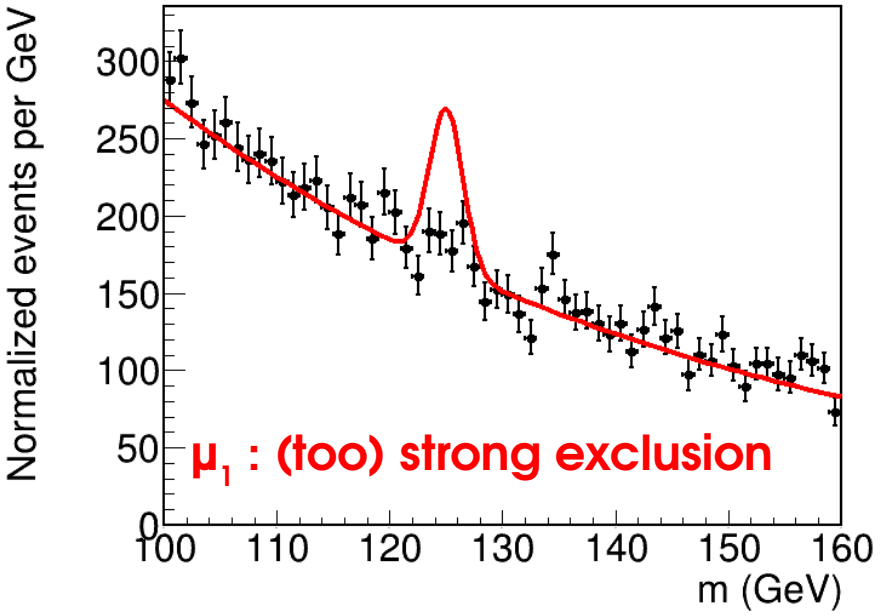
Upper Limits

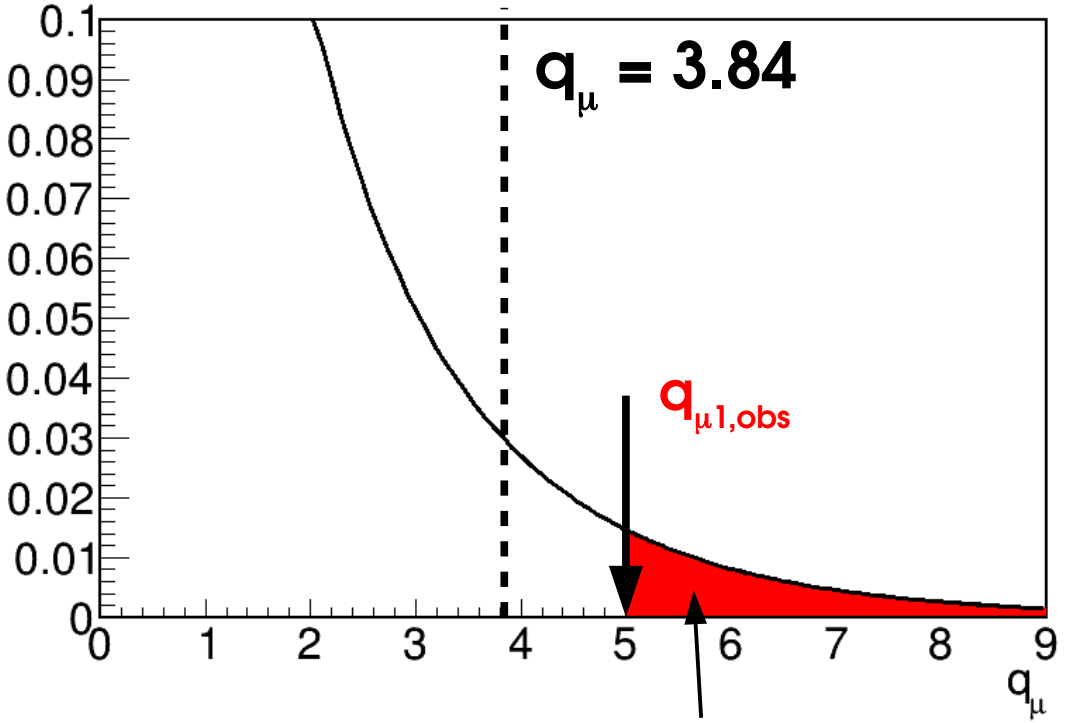
Upper limits: small signal observe – Can we exclude $\mu >$ some μ_0 ?

→ Consider H_0 : $H(\mu=\mu_0)$ – alternative H_1 : $H(\hat{\mu} < \mu_0)$

→ Compute $q_{\mu 0}$, exclusion p-value $p_{\mu 0}$.

→ Raise μ_0 until 95% CL exclusion ($p_{\mu 0}$ = 5%) is reached





→ Also use a 1-sided definition of the PLR, as for discovery

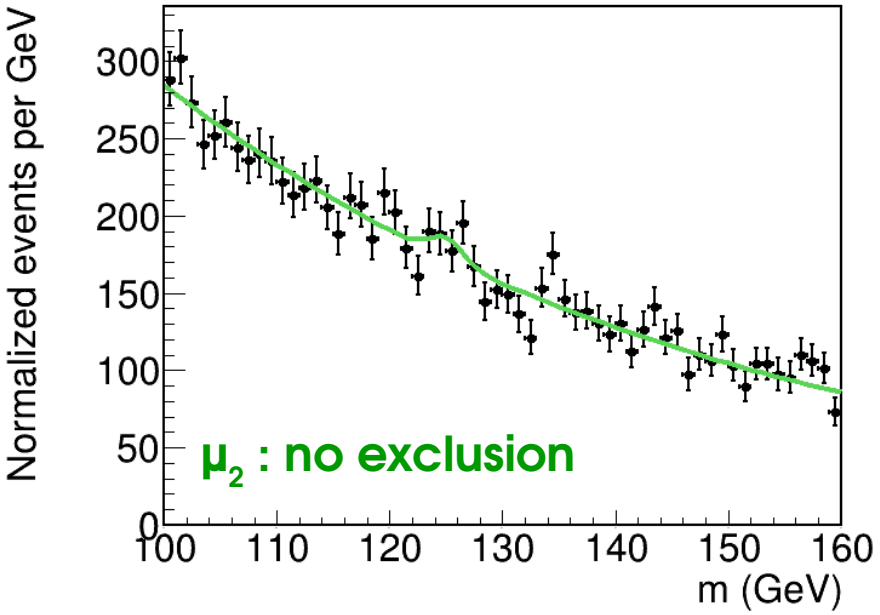
Upper Limits

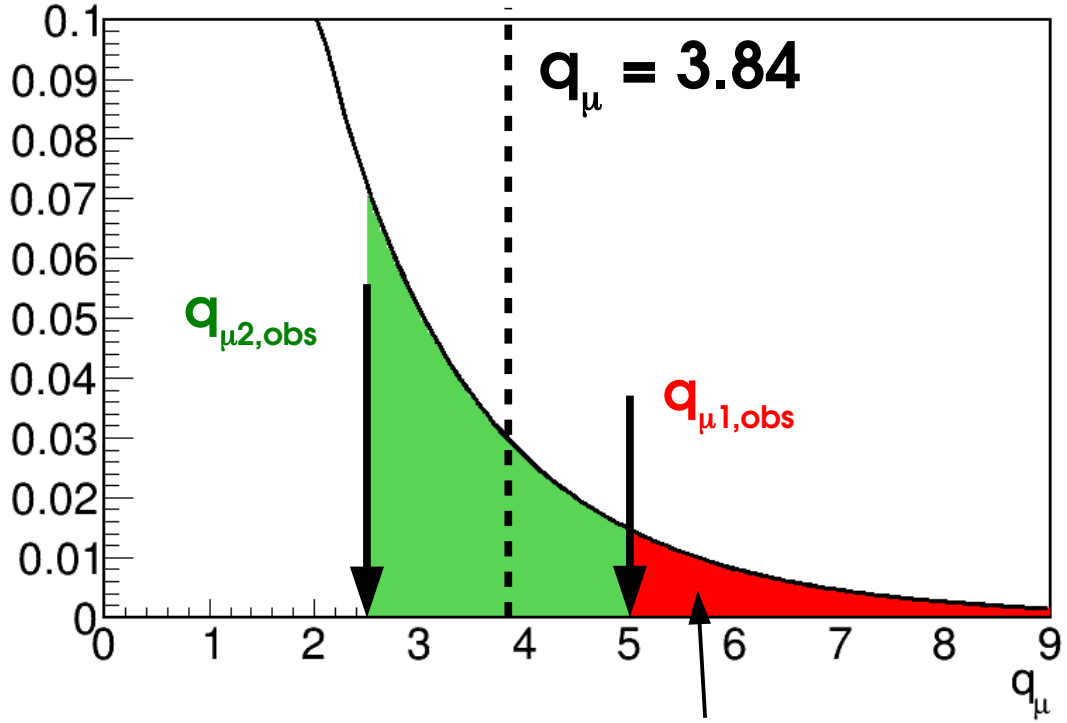
Upper limits: small signal observe – Can we exclude $\mu >$ some μ_0 ?

→ Consider H_0 : $H(\mu=\mu_0)$ – alternative H_1 : $H(\hat{\mu} < \mu_0)$

→ Compute $q_{\mu 0}$, exclusion p-value $p_{\mu 0}$.

→ Raise μ_0 until 95% CL exclusion ($p_{\mu 0}$ = 5%) is reached





→ Also use a 1-sided definition of the PLR, as for discovery

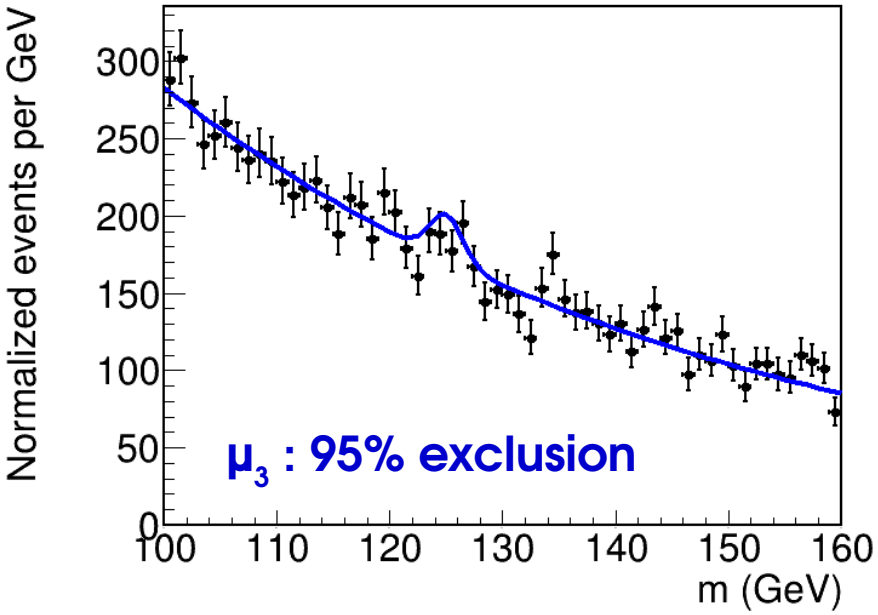
Upper Limits

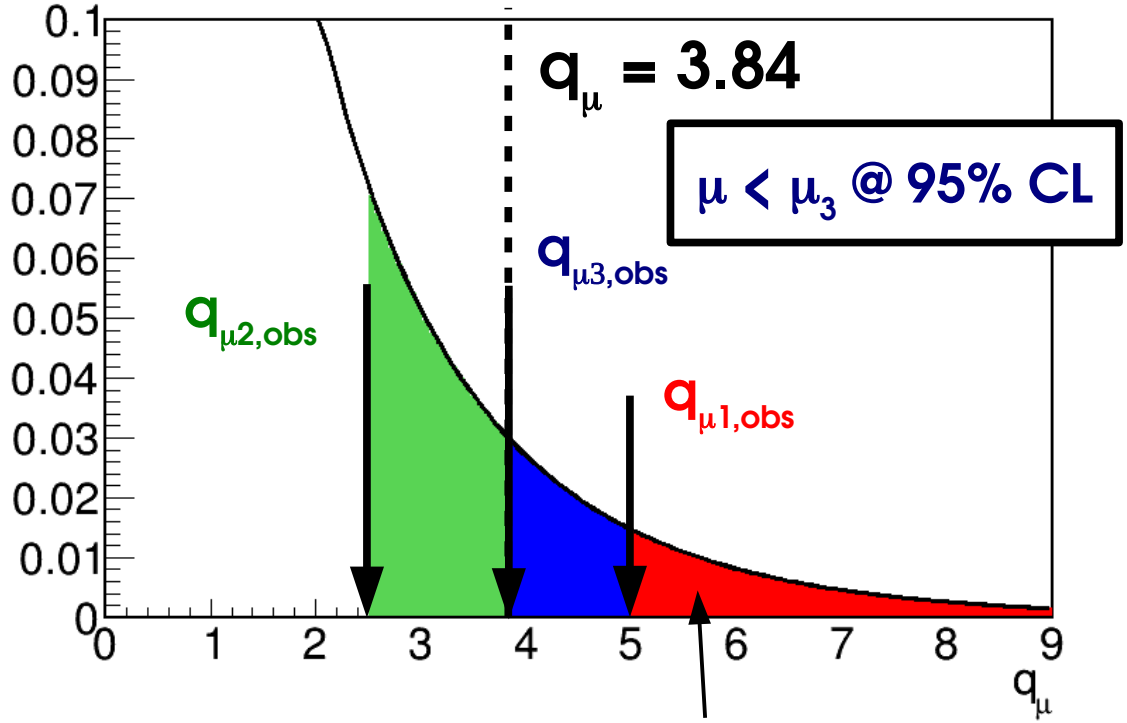
Upper limits: small signal observe – Can we exclude $\mu >$ some μ_0 ?

→ Consider H_0 : $H(\mu=\mu_0)$ – alternative H_1 : $H(\hat{\mu} < \mu_0)$

→ Compute $q_{\mu 0}$, exclusion p-value $p_{\mu 0}$.

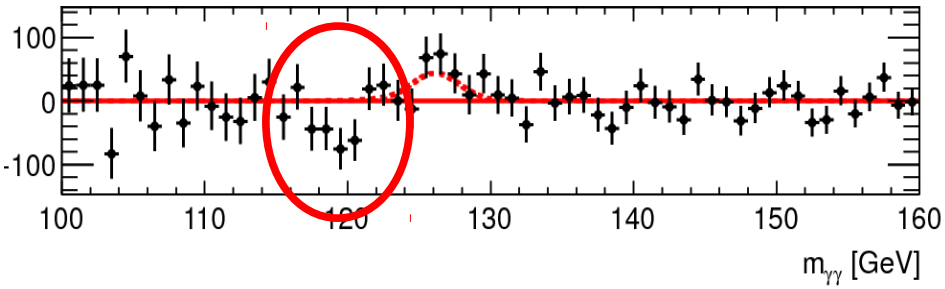
→ Raise μ_0 until 95% CL exclusion ($p_{\mu 0}$ = 5%) is reached





→ Also use a 1-sided definition of the PLR, as for discovery

CL_s



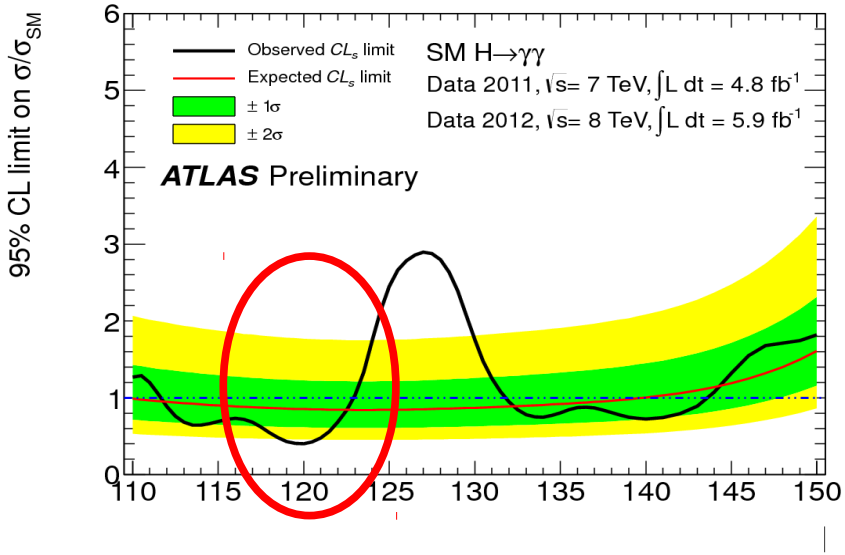
Problem: for negative $\hat{\mu}$, get very good observed limit (can even be negative!)

→ expected: 95% C.L. ⇒ miss true value with p=5%

Usual solution to get more intuitive results: **CL_s**.

→ Compute modified p-value **$p_{\mu 0}$ / $p_{\mu=0}$**

- **$p_{\mu 0}$** is the usual p-value (5%)
- **$p_{\mu=0}$** is the p-value computed for μ=0

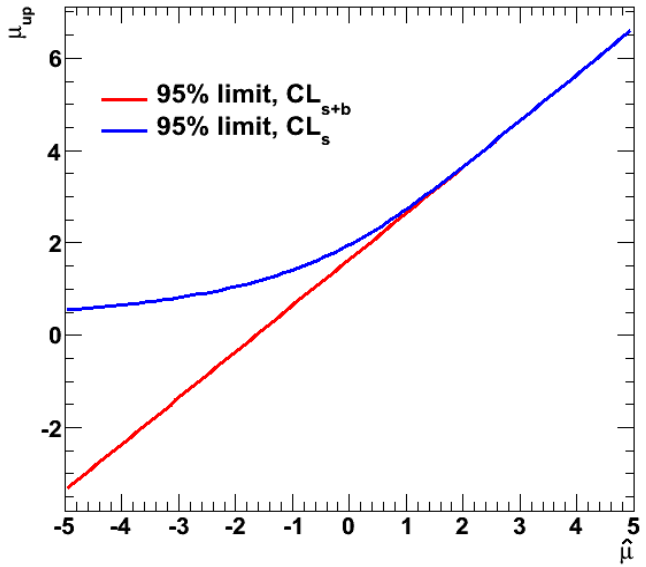


→ **"Good cases"** ($\hat{\mu}$~0) : $p_{\mu=0}$ ~ 1

CL_s ~ $p_{\mu 0}$, no change.

→ **"Pathological case"** (very negative $\hat{\mu}$) : $p_{\mu=0} \ll 1$

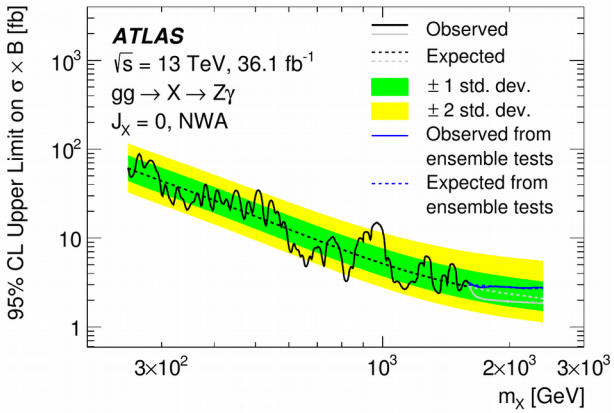
CL_s~ $p_{\mu 0}/p_{\mu=0} \gg 5\%$ so worse limit, as desired



Drawback: coverage is not maintained
(e.g. a 95% C.L. limit is actually 98% C.L.)

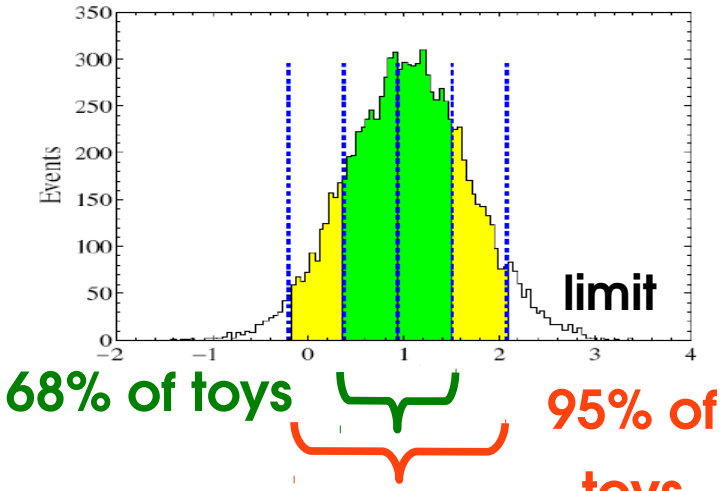
Expected Limits

***Expected* results**: median outcome under given hypothesis, usually B-only. Computed using:



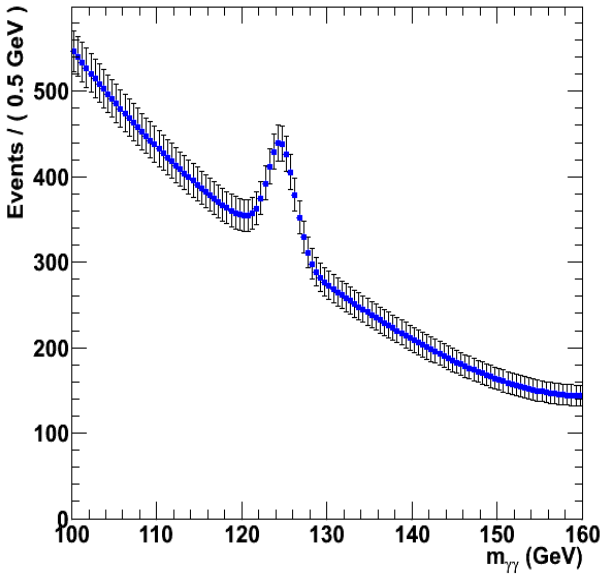
→ **Toys:**

- Generate pseudo-data in B-only hypo
- Compute limit
- Repeat and histogram the results, report median & quantiles



→ ***Asimov Datasets***

- Generate a "perfect dataset", no fluctuations (by setting bin contents carefully)
- Gives the median result immediately
- Get bands from asymptotic formulas (assuming Gaussianity)
  - ⊖ relies on Gaussian approximation
  - ⊕ Much faster (1 "toy")



Cowan, Cranmer, Gross & Vitells, Eur.Phys.J.C71:1554,2011

Presentation of Results

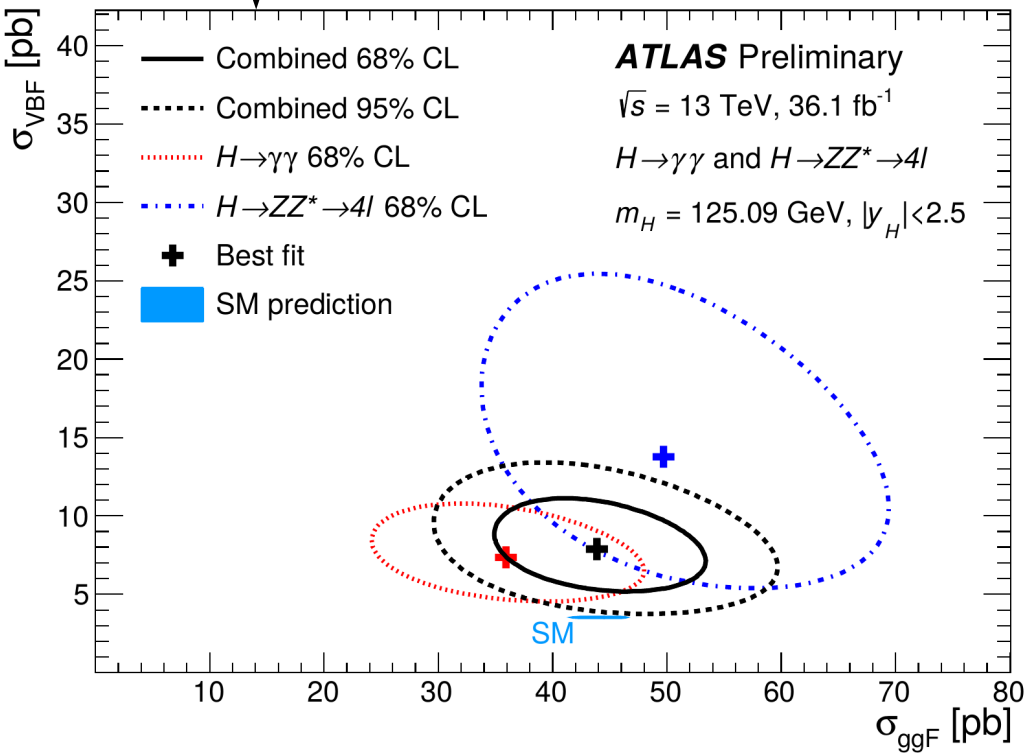
Reparameterization

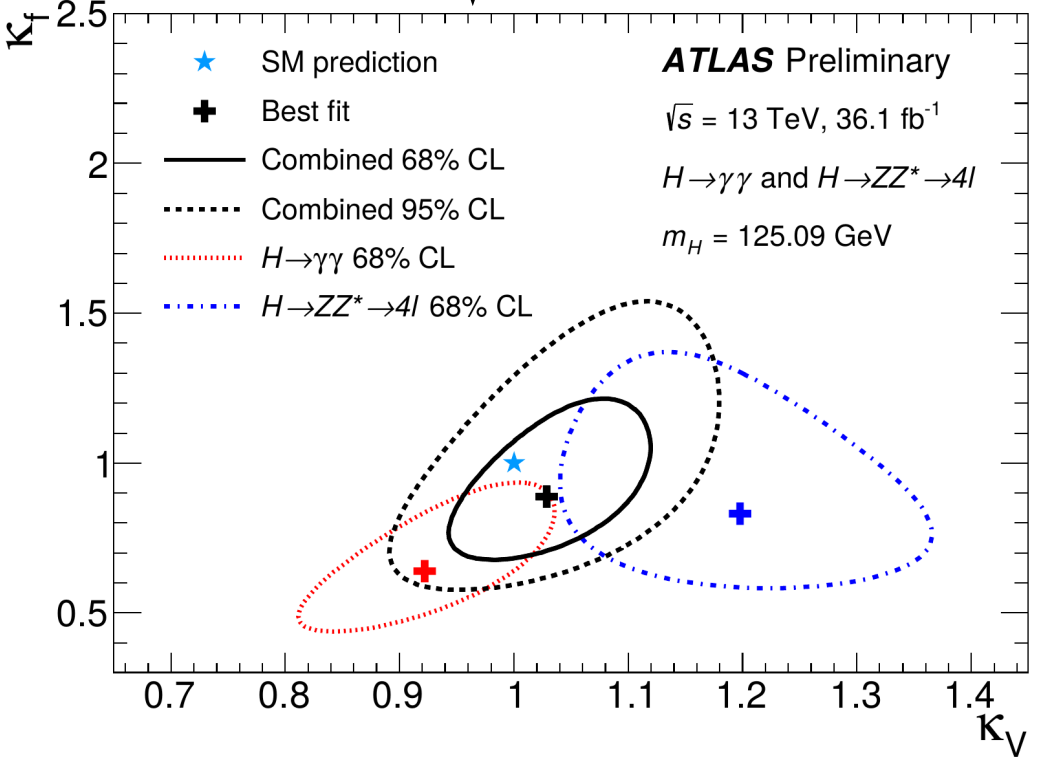
Start with basic measurement in terms of e.g. **σ×B**

→ How to measure derived quantities (couplings, parameters in some theory model, etc.) ? **→ just reparameterize the likelihood:**

e.g. Higgs couplings: σ_{ggF}, σ_{VBF} sensitive to Higgs coupling modifiers κ_V, κ_F.

$$L(\sigma_{ggF}, \sigma_{VBF}) \xrightarrow{\substack{\sigma_{ggF} \to \sigma_{ggF}(\kappa_V, \kappa_F) \\ \sigma_{VBF} \to \sigma_{VBF}(\kappa_V, \kappa_F)}} L(\sigma_{ggF}(\kappa_V, \kappa_F), \sigma_{VBF}(\kappa_V, \kappa_F)) \equiv L'(\kappa_V, \kappa_F)$$

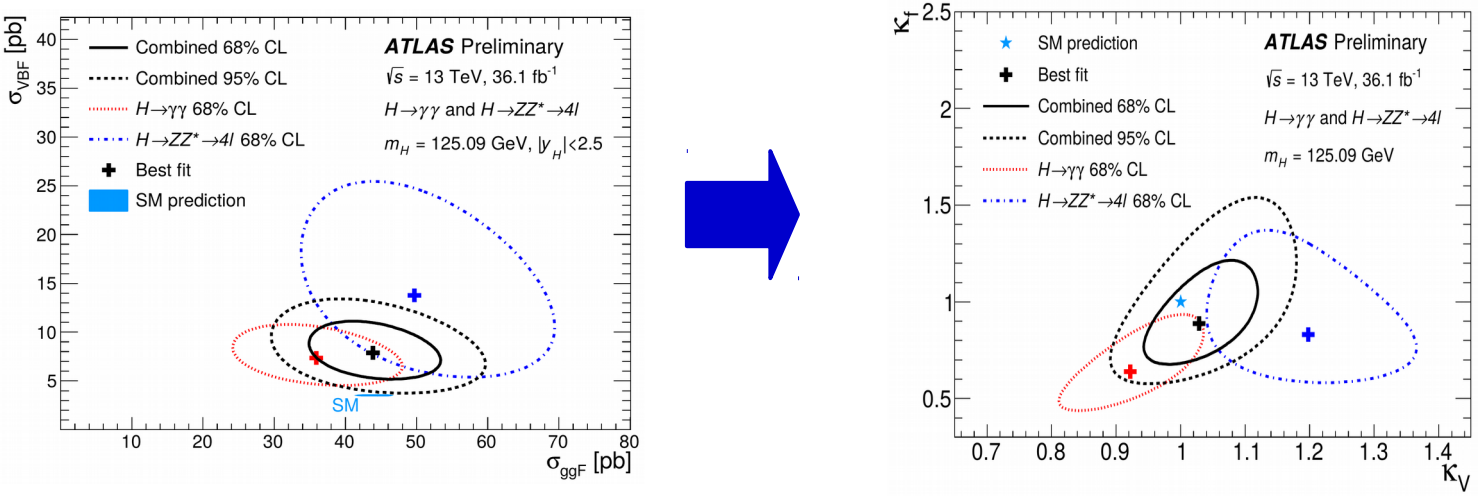




Presentation of Results

Measurements often recast to constrain a particular theory model.

→ Ideally, by **reparameterizing the likelihood** and repeating the measurement



⇒ **Done by experiments for selected benchmark models**

→ However, **not usually possible to do a posteriori**:

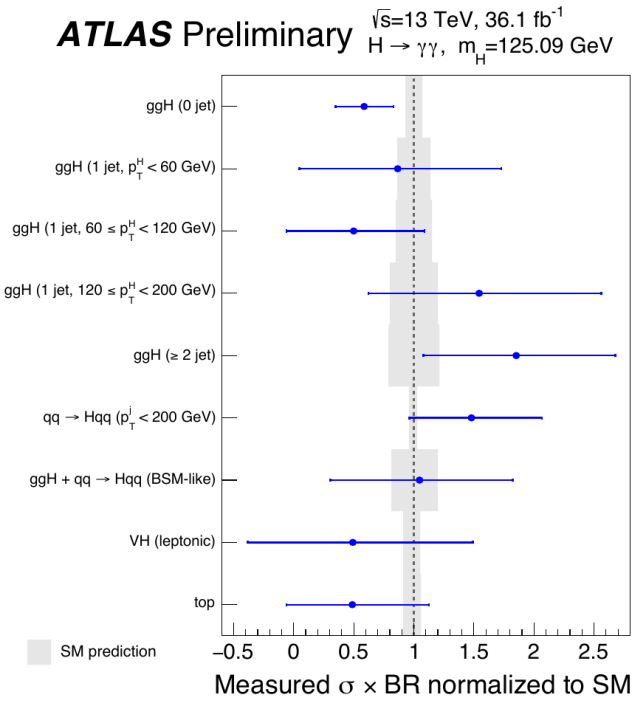
- Full likelihood typically not published
- theorists typically do not want to deal with 4000 NPs…

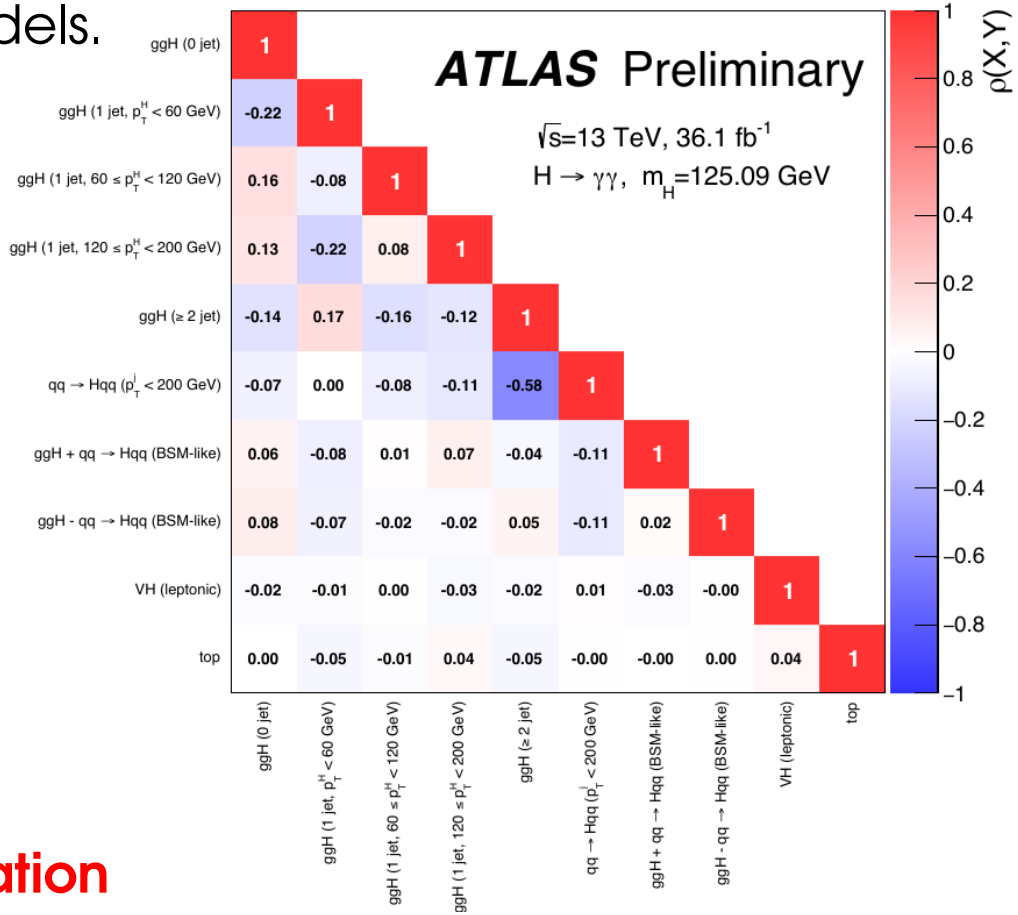
→ **Other approaches**: e.g. reimplementing the analysis in a public fast-simulation framework (e.g. SUSY searches). However clear accuracy limitations

Presentation of Results

→ **Current solution**: correlation matrices in HEPData, together with central values

→ use **BLUE** to combine in alternate models.





→ **Only valid in the Gaussian approximation**

→ To go further, need some **simplified likelihood** including non-Gaussian effects

- **Profile likelihood** – function of POI only (NPs profiled out)
- **Additional terms** for non-Gaussian effects

→ Significantly more complex (many dimensions!)

→ Will be needed eventually as measurements become syst-dominated

Conclusion

- **Statistical methods in ATLAS now rather mature**
  - some not discussed in this talk: Bayesian, BLUE, etc.

- **Standard implementations** within the RooFit/RooStat toolkits shipped with ROOT framework, as well as other tools (BAT for Bayesian methods)
  - Many listed here
  - Documentation already quite good usually, but if you need help please ask the Stat Forum

- **Improvement and uniformization efforts are still ongoing**

- **Open questions on how to present results in a way that preserves the available information for future interpreations.**

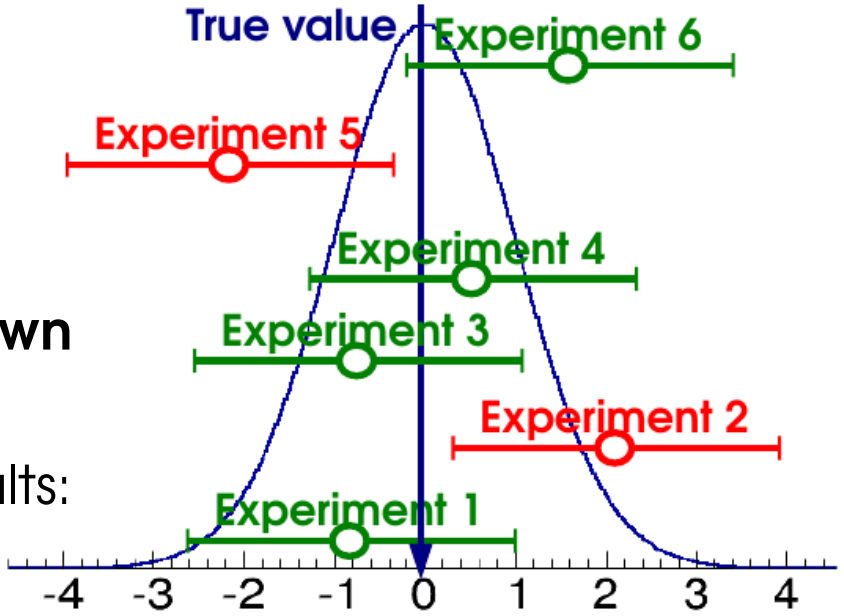
Extra Material

Bayesian Methods

Frequentist vs. Bayesian

All methods described so far are **frequentist**

- Probabilities (p-values) refer to outcomes if the experiment were **repeated identically many times**
- Parameters value are **well-defined but unknown**



- Probabilities apply to measurements and results:

→ **"m_H = 125.09 ± 0.24 GeV"** :

→ i.e. $[125.09 - 0.24 \; ; \; 125.09 + 0.24]$ GeV has 68% probability to contain the true m_H.

→ if we repeated the experiment many times, we would get different intervals, 68% of which would contain the true m_H.

→ **"5σ Higgs discovery"**

- if there is really no Higgs, such fluctuations observed in 3.10^{-7} of experiments
- Not exactly the crucial question: "What is the probability that the excess we see is a fluctuation"

→ What we have is **P(data | no Higgs)** --
→ What we would really want **P(no Higgs |data).**

Frequentist vs. Bayesian

Can use **Bayes' theorem** to address this:

$$P(\mu|data) = \frac{P(data|\mu)}{P(data)} P(\mu)$$

same as in the frequentist formalism (=likelihood)

Prior Probability

irrelevant normalization factor

Can compute P(μ | data), ***if we provide P(μ)***

→ Implicitly, we have now made μ into a random variable

- Is m_H, or the presence of H(125), randomly chosen ?
- In fact, different definition of p: ***degree of belief***, not from frequencies.
- P(μ) ***Prior degree of belief*** – critical ingredient in the computation

Compared to frequentist PLR:

⊕ answers the "right" question

⊖ answer depends on the prior

"Bayesians address the questions everyone is interested in by using assumptions that no one believes. Frequentists use impeccable logic to deal with an issue that is of no interest to anyone." - **Louis Lyons**

Bayesian Methods

Probability distribution (= likelihood) : same form as frequentist case, but

P(θ) constraints now **priors for the syst. NPs**, not aux. measurement $P(\theta^{mes} \mid \theta)$

⊕ Integrate them out, no need for profiling. Then use probability distribution P(μ) directly for limits, credibility intervals

⊖ No simple way to test for discovery

Priors : most analyses still using flat priors in the analysis variable(s)
⇒ **Parameterization-dependent**: if flat in σ×B, then not flat in κ...

→ Can use the Jeffreys' or reference priors, but difficult in practice

Frequentist-Bayesian Hybrid methods ("Cousins-Highland")

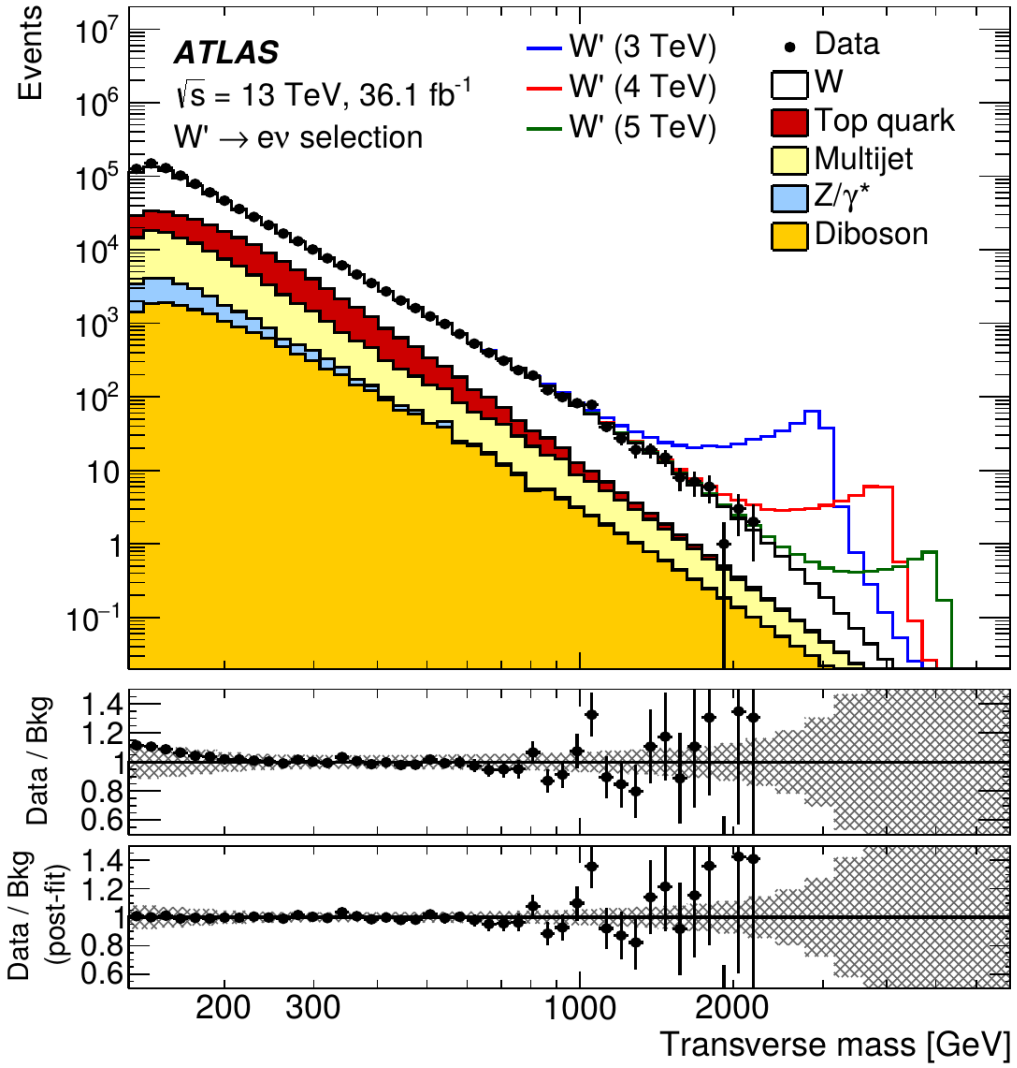
- Integrate out NPs as in Bayesian measurements
- Once only POIs left, Use P(data | x) in a frequentist way

  → "Bayesian NPs, frequentist POIs"
- Some use in Run 1, now phased out in favor of frequentist PLR.

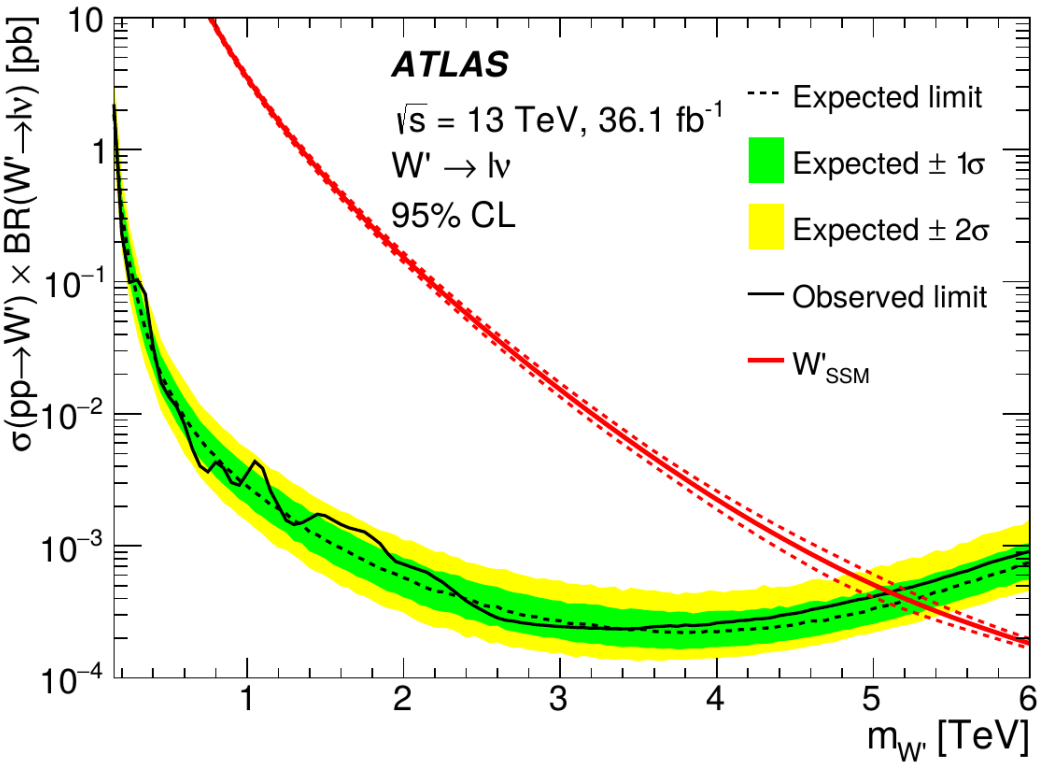
Example: W'→lν Search

arXiv:1706.04786

- **POI**: **W' σ×B** → use **flat prior over** [0, +∞[.
- **NPs**: syst on **signal ε** (6 NPs), **bkg** (6), **lumi** (1) → integrate over Gaussian priors

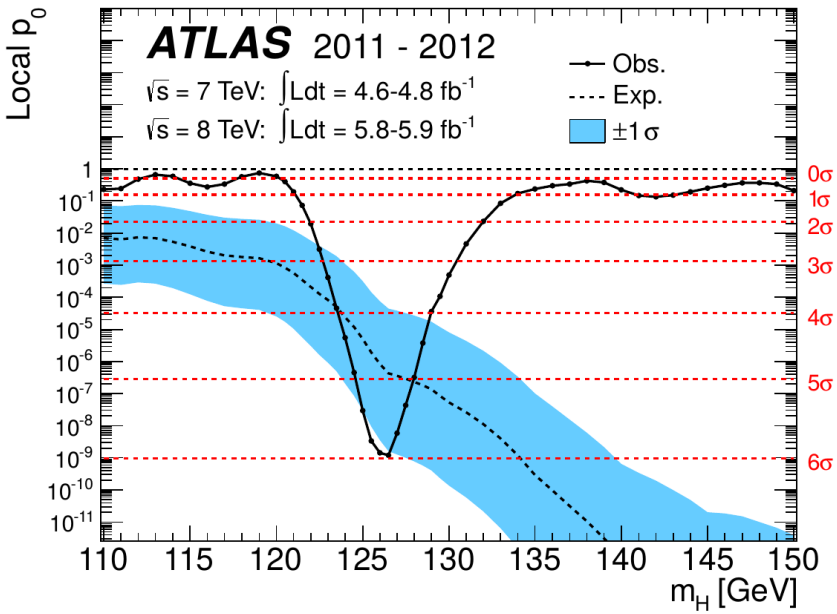


Trigger	Multijet background
Lepton reconstruction and identification	Top extrapolation
Lepton momentum scale and resolution	Diboson extrapolation
$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and scale	PDF choice for DY
Jet energy resolution	PDF variation for DY
Pile-up	EW corrections for DY
	Luminosity



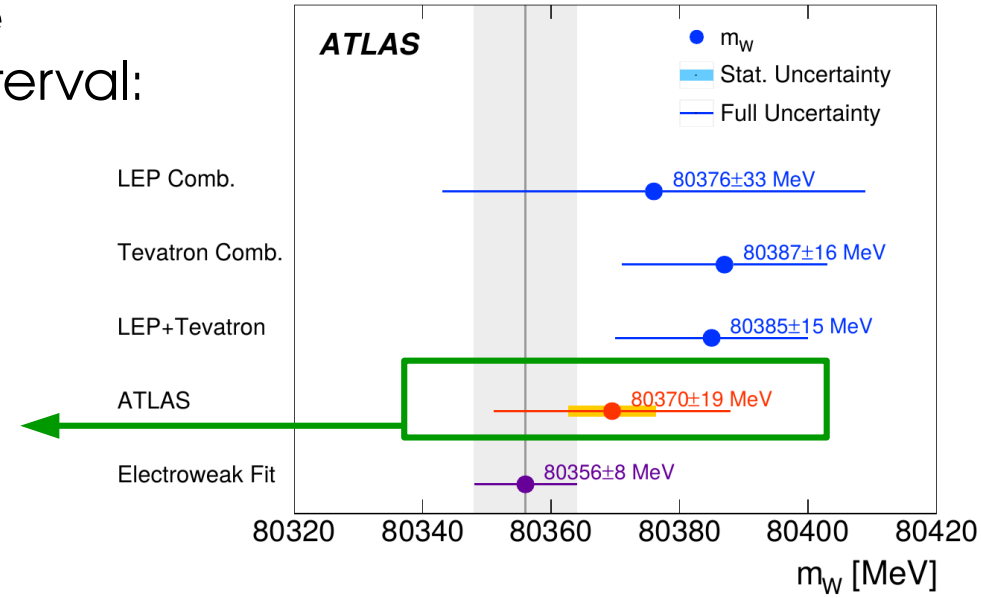
Randomness in High-Energy Physics

All HEP results are probabilistic in nature



- **Can a background fluctuation cause a Higgs-like excess ?**
  → probability ~ 10^{-9}
  ⇒ the famous (and conventional) "5σ"

- For measurements: probabilities that the **true value** of a parameter is within an interval:



68% chance that the true m_W is here

Hypothesis testing

Usual HEP results can be recast in terms of **hypothesis testing**:

- **Discovery**: is the data compatible with background-only ?
  → H_0 ("null hypothesis") : only background is present
  → How well can we **reject** H_0 ? → p-value (or significance)

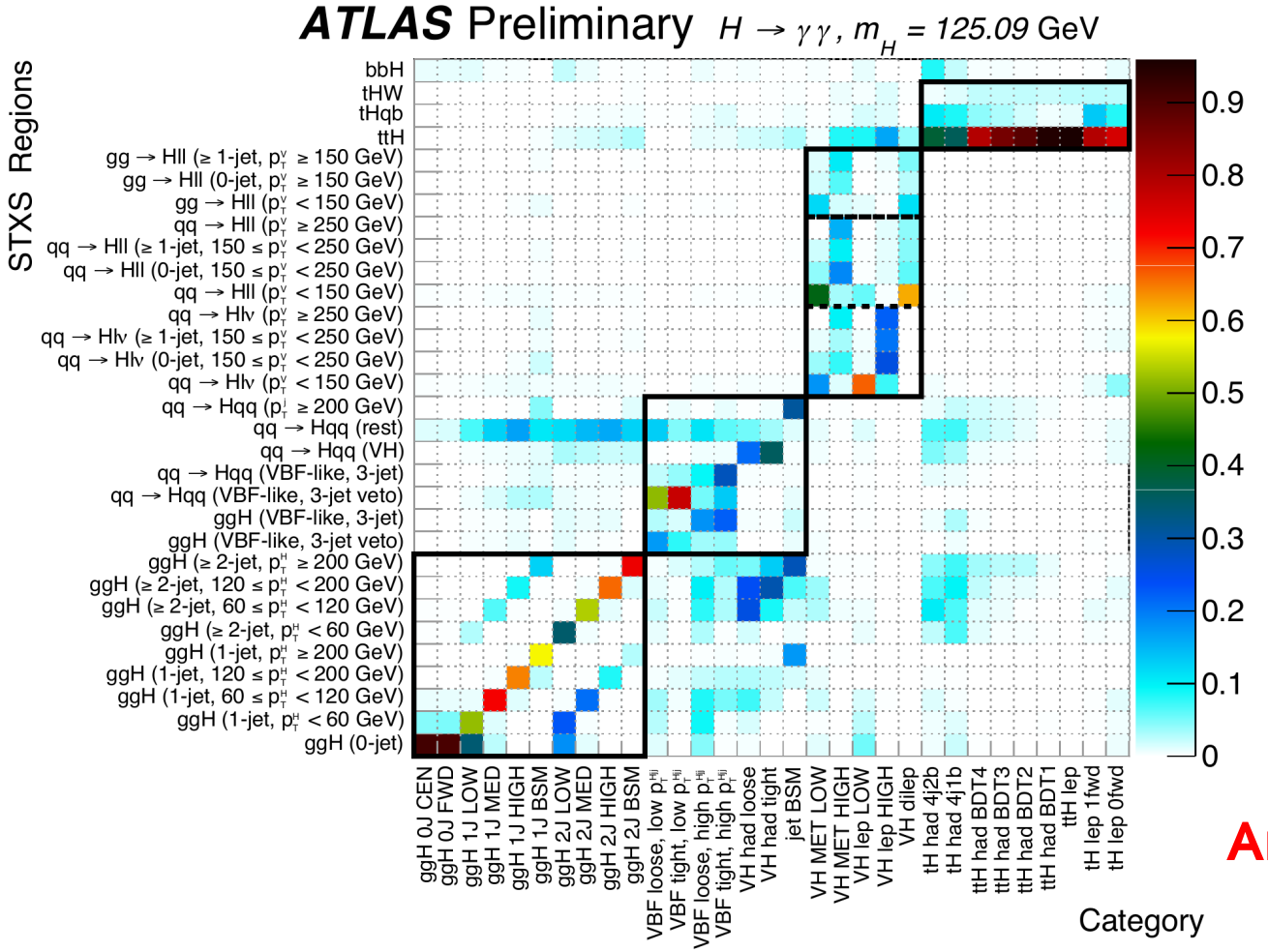
- **Upper limits**: no excess observed – how small must the signal be ?
  → $H_0(S)$: B + some signal S
  → How small can we make S, and still reject $H_0(S)$ at 95% C.L. (p=5%) ?

- **Parameter measurement**
  → $H_0(\mu)$: some parameter value μ
  → What values μ are <u>not</u> rejected at 68% C.L. (p=32%) ?
  ⇒ 1σ confidence interval on μ

Categories for Property Measurements

Categories also useful to provide measurements of separate kinematic regions
→ e.g. differential cross-section measurements

Targeted truth regions



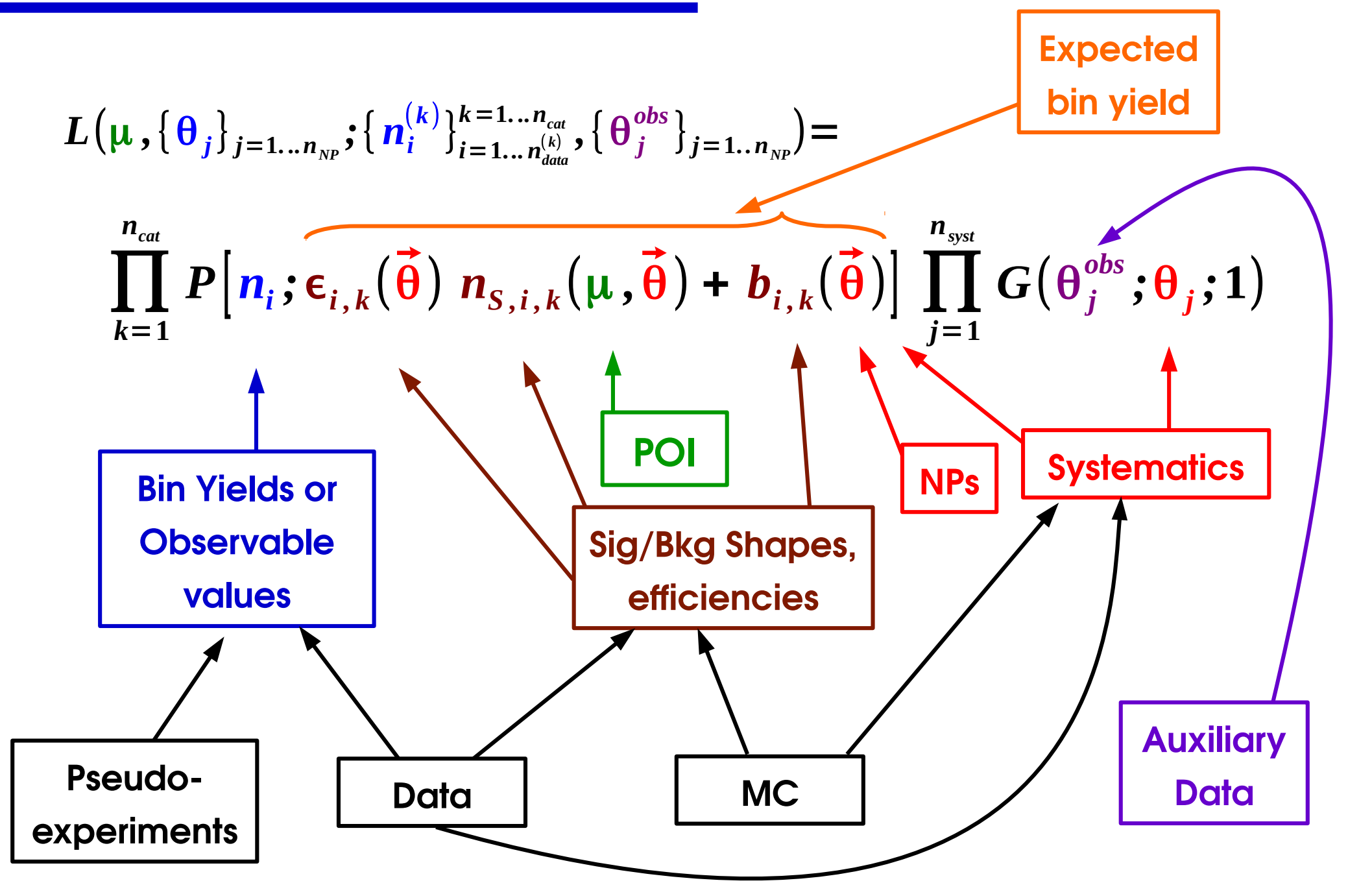
H→ γγ Properties Measurement (ATLAS-CONF-2017-045)

Analysis Selections

Most categories aimed at one particular truth region, but cross-feed from detector effects (acceptance, pileup, etc.)

Correlation of systematics treated naturally as NPs affecting one or more categories

Likelihood, the full version (binned case)



Effect of Profiling

Systematics still affect the result even after profiling their NPs!

e.g. **Simple counting experiment: $N(\mu,\theta) = \mu N_0 + \theta$**, measure $N = N_{obs}$.

1. No systematics: $N(\mu) = \mu N_0$

→ $\hat{\mu}$ **fit** adjust $\hat{\mu}$ so that $N(\hat{\mu}) = N_{obs}$

$$t_{\mu_0} = -2\log\frac{L\left(\mu=\mu_0, \hat{\hat{\theta}}_{\mu_0}; N_{obs}\right)}{L\left(\hat{\mu}, \hat{\theta}; N_{obs}\right)}$$

→ **$\mu=\mu_0$ fit**: $\mu=\mu_0$ fixed ⇒ $N(\mu_0) = \mu_0 N_0$, cannot adjust

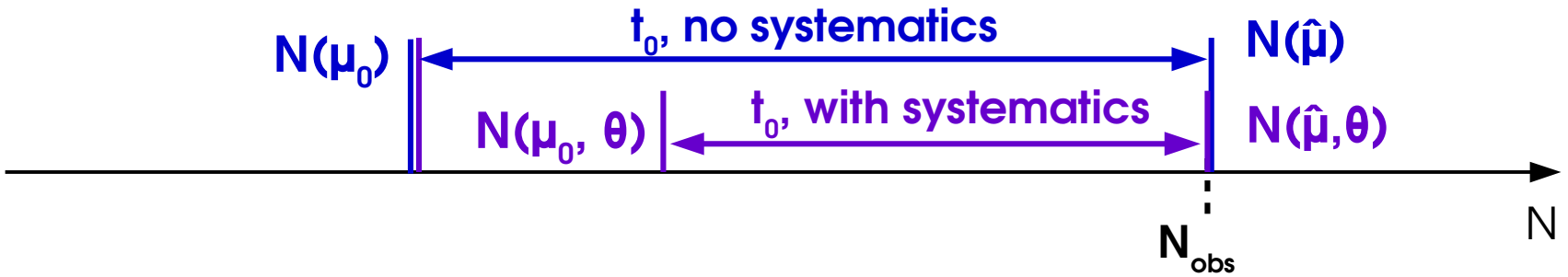
⇒ tension between $N(\mu_0)$ and N_{obs} ⇒ large $t_{\mu 0}$ ⇒ **strong exclusion of $H(\mu_0)$**

2. With Systematics: $N(\mu,\theta) = \mu N_0 + \theta$

→ $\hat{\mu}$ **fit** adjust $N(\hat{\mu}, \hat{\theta}) = N(\hat{\mu}, \hat{\theta}=0) = N_{obs}$ using μ only (avoid penalty on θ)

→ **$\mu=\mu_0$ fit**: $\mu=\mu_0$ fixed, but $\hat{\theta}_{\mu 0}$ **can still pull $N(\mu_0,\hat{\theta}_{\mu 0})$ towards $N(\hat{\mu},0)=N_{obs}$**

⇒ smaller $t_{\mu 0}$ ⇒ **reduced exclusion of $H(\mu_0)$**



Comparison with LEP/TeVatron definitions

Likelihood ratios are not a new idea:

- **LEP**: Simple LR with NPs from MC
  - Compare μ=0 and μ=1
- **Tevatron**: PLR with profiled NPs

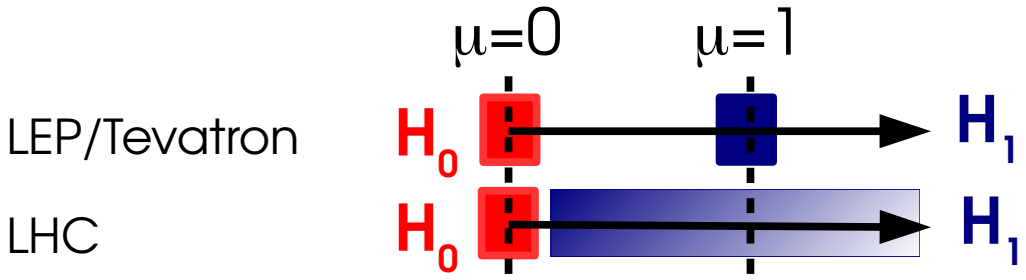
$$q_{LEP} = -2\log\frac{L(\mu=0,\tilde{\theta})}{L(\mu=1,\tilde{\theta})}$$

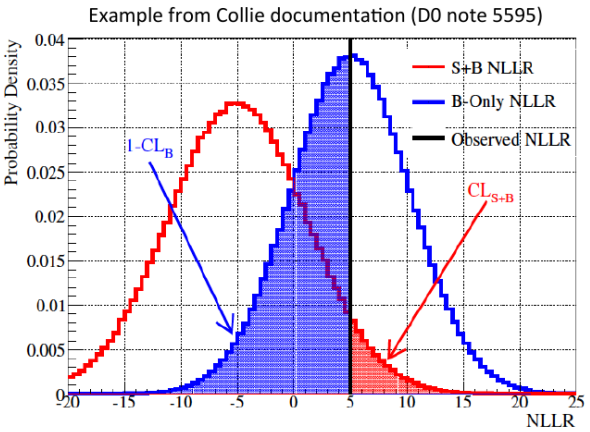
$$q_{Tevatron} = -2\log\frac{L(\mu=0,\hat{\hat{\theta}}_0)}{L(\mu=1,\hat{\hat{\theta}}_1)}$$

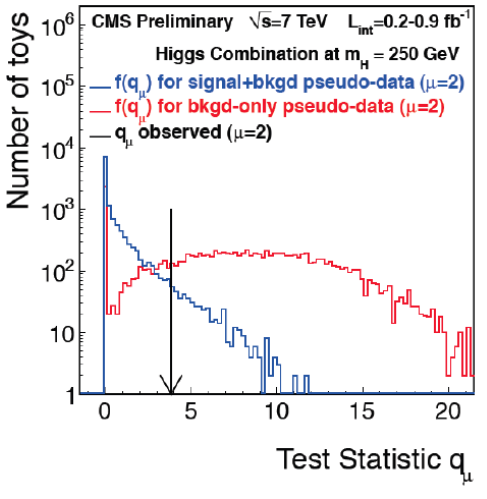
Both compare to μ=1 instead of best-fit
→ Asymptotically:

- **LEP/Tevaton**: q linear in μ ⇒ **~Gaussian**

- **LHC**: q quadratic in μ ⇒ **~χ2**







Andrey Korytov, EPS 2011

Global Significance

Probability to obtain a fluctuation somewhere: **Global** p-value. Schematically:

Global p-value → $p_{global} = 1 - (1 - p_{local})^N \approx N \; p_{local}$ ← **Local** p-value

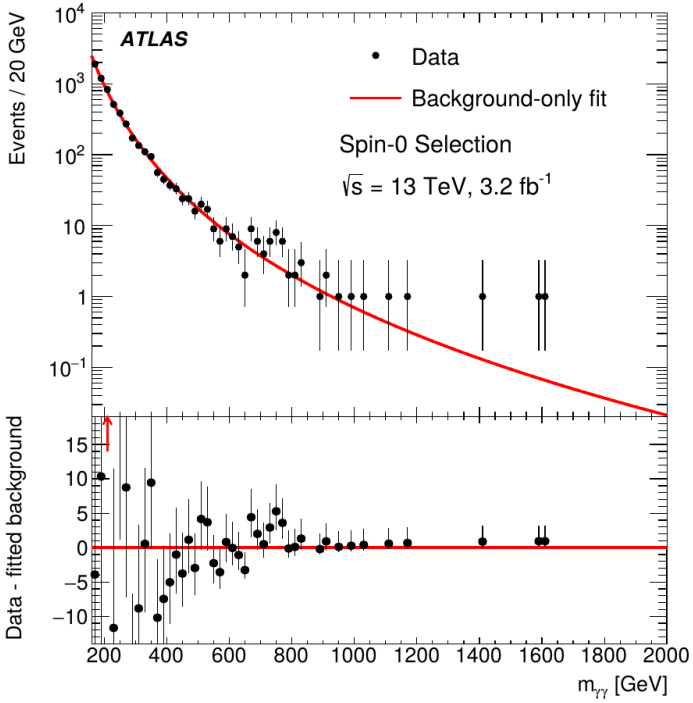
Trials factor = number of independent experiments

→ $p_{global} > p_{local} \Rightarrow Z_{global} < Z_{local}$ – fluctuations are more likely, so less significant

For searches over a parameter range,
p_{global} is the relevant p-value

Depends on the scanned parameter ranges
→ e.g. $X \to \gamma\gamma$: $200 < m_X < 2000$ GeV, $0 < \Gamma_X < 10\% \, m_X$.

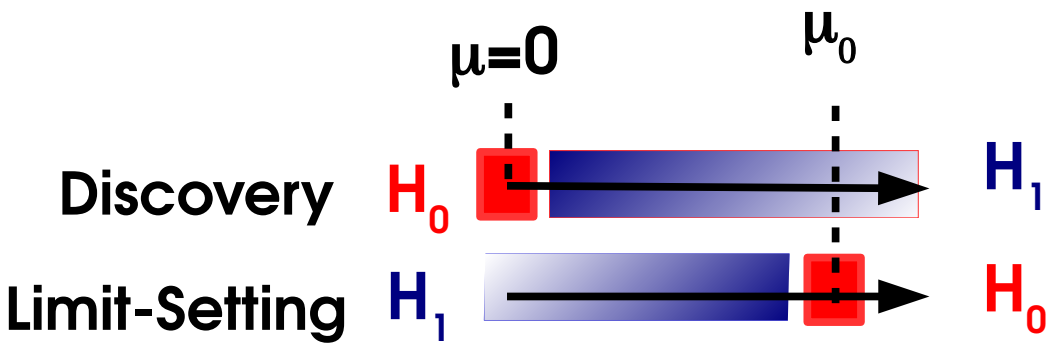
→ However what comes out of the usual asymptotic formulas is p_{local}.
How to compute p_{global} ?



Upper Limits

Again one sided:

→ if **large** signal observed with $\hat{\mu} \gg \mu_0$, this should not help set an upper limit at μ_0 – the actual upper limit should be higher than μ!



⇒ Set $q_{\mu 0} = 0$ for $\hat{\mu} \gg \mu_0$ – only small signals with $\hat{\mu} < \mu_0$ help lower the limit.

→ Also treat separately the case μ < 0 since this can lead to technical issues in fitting

$$q_{\mu_0} = \begin{cases} -2\log\dfrac{L(\mu=\mu_0)}{L(\hat{\mu})} & 0 \le \hat{\mu} \le \mu_0 \\ 0 & \hat{\mu} > \mu_0 \\ -2\log\dfrac{L(\mu=\mu_0)}{L(\mu=0)} & \hat{\mu} < 0 \end{cases}$$

Sensitivity Issues

Limit ~ $\hat{\mu} + 1.96\sigma_\mu$

Problem: for negative $\hat{\mu}$, get very good (**too good!**) observed limit.

→ For $\hat{\mu}$ sufficiently negative, can have limit < 0 !

How can this be ?

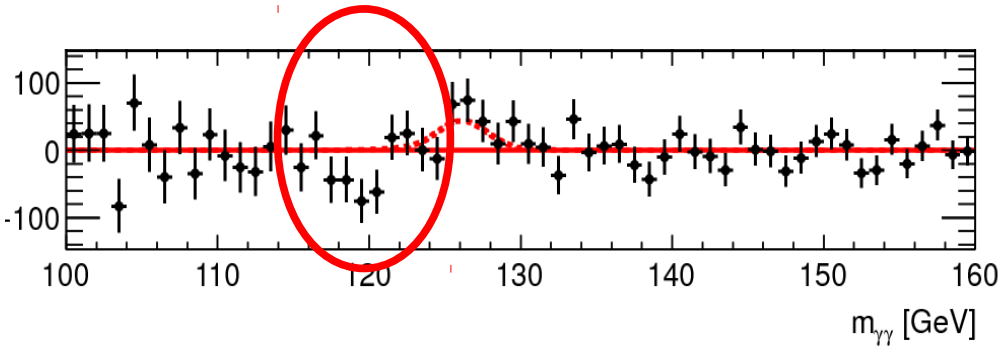
→ Background modeling issue ?...

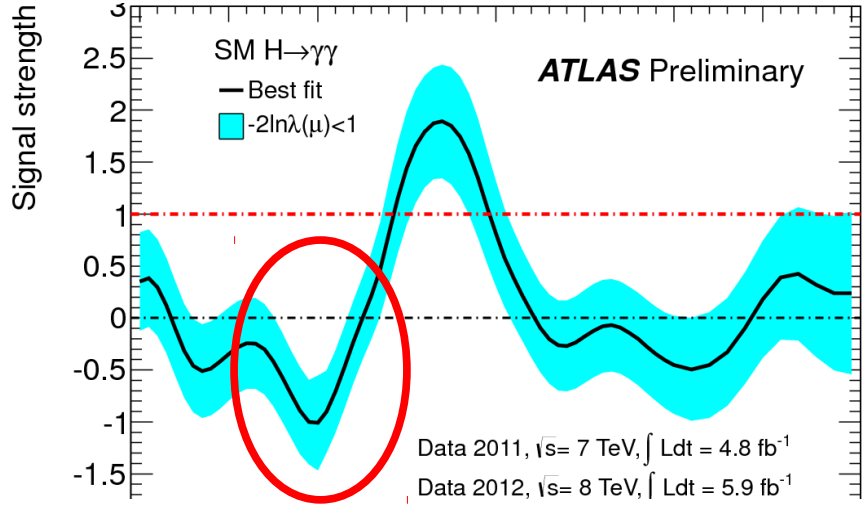
→ This is a **95%** limit

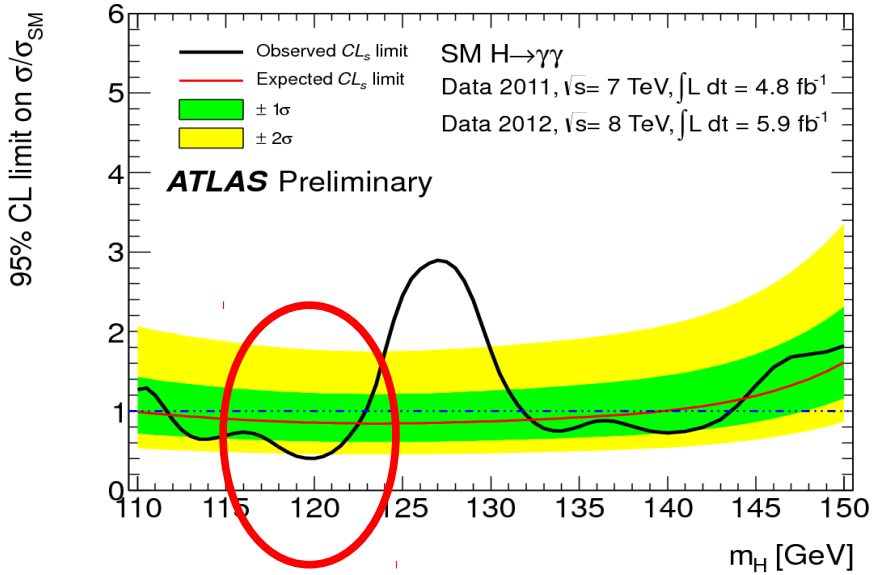
⇒ **5% of the time, the limit wrongly excludes the true value.**

→ If we assume μ must be >0, we know these are just fluctuations

⇒ **Special procedure for $\hat{\mu}$<0**





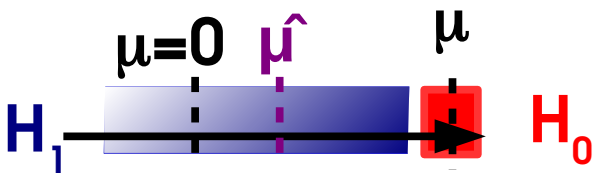


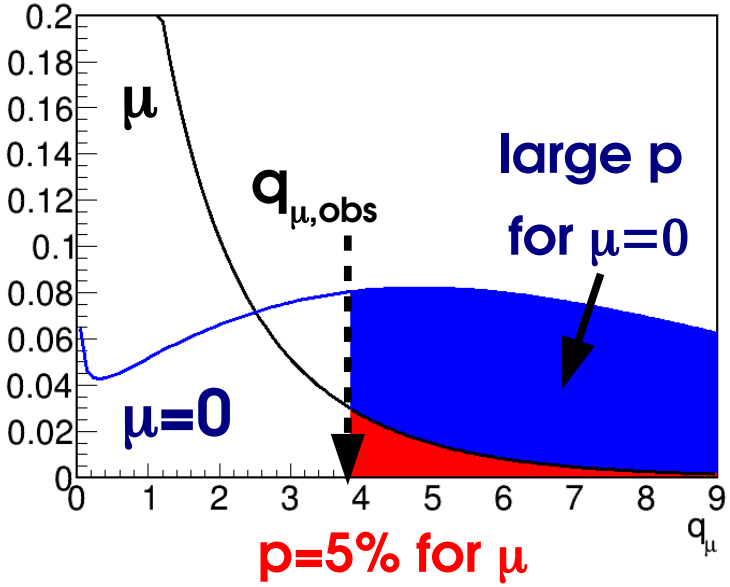
Sensitivity to μ~0

When setting limits, goal is to exclude large μ, **to indicate that μ~0.**

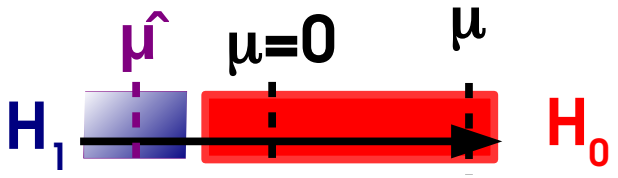
Investigate the μ=0 hypothesis:

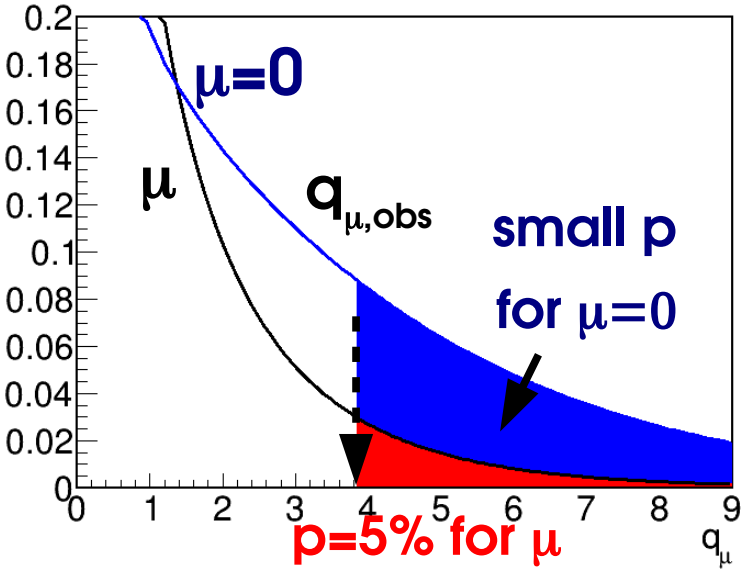
- **Normal case**: μ̂~0, so **μ=0 not excluded** : large p-value for μ=0.





  - **Pathological case**, very negative μ̂, so **μ=0 also excluded** : p-value for μ=0 also small





- Bad case: **large μ and μ~0 both excluded** : **no sensitivity in μ>0 region** !

CL_s

Need a fix to get more reasonable limits – Usual solution in HEP: CL_s.

→ Compute modified p-value $p_{\mu 0}$ / $p_{\mu=0}$

- $p_{\mu 0}$ is the usual p-value (5%)
- $p_{\mu=0}$ is the p-value computed for μ=0

⇒ **Rescale $p_{\mu 0}$ by the p-value $p_{\mu=0}$ for μ=0**

use μ=0 exclusion as reference

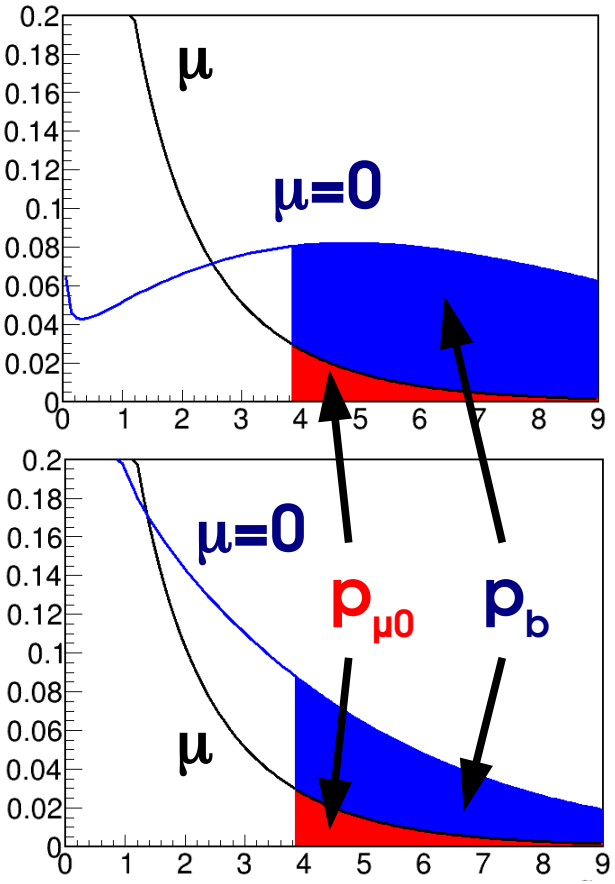
→ **"Good case"** : $p_{\mu=0} \sim 1$

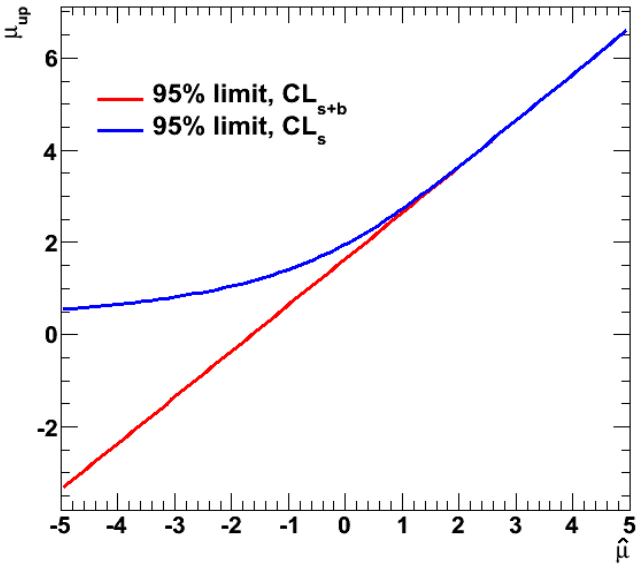
$CL_s \sim p_{\mu 0} \sim 5\%$, no change.

→ **"Pathological case"** : $p_{\mu=0} \ll 1$

$CL_s \sim p_{\mu 0}/p_{\mu=0} \gg 5\%$

- So worse limit, as desired
- **Drawback**: overcoverage (e.g. limit is actually 98% C.L.)



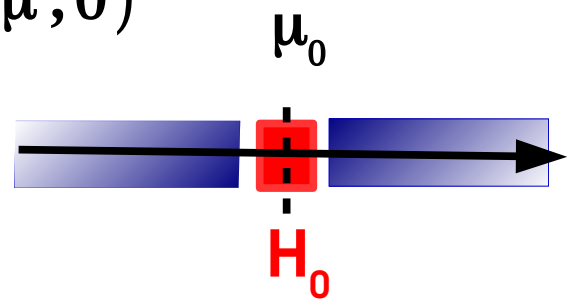


Confidence Intervals

Confidence intervals:

$$t_{\mu_0} = -2\log\frac{L\left(\mu=\mu_0,\hat{\hat{\theta}}_{\mu_0}\right)}{L\left(\hat{\mu},\hat{\theta}\right)}$$

- Test $H(\mu_0)$
- Excluded value are **outside** the interval
- Two-sided test since true value can be higher or lower than observed

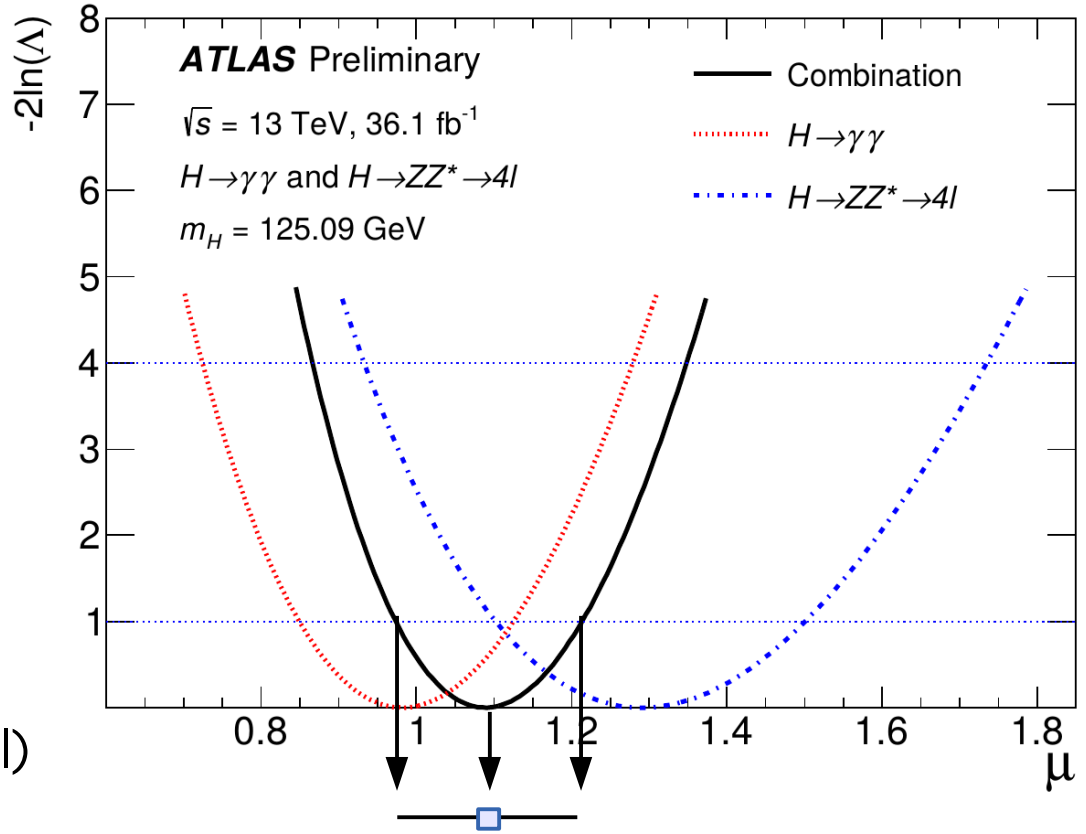


Asymptotics:

- t_μ distributed as a $\chi^2(N_{dof})$
- N_{dof} is number of POIs (1 or more)

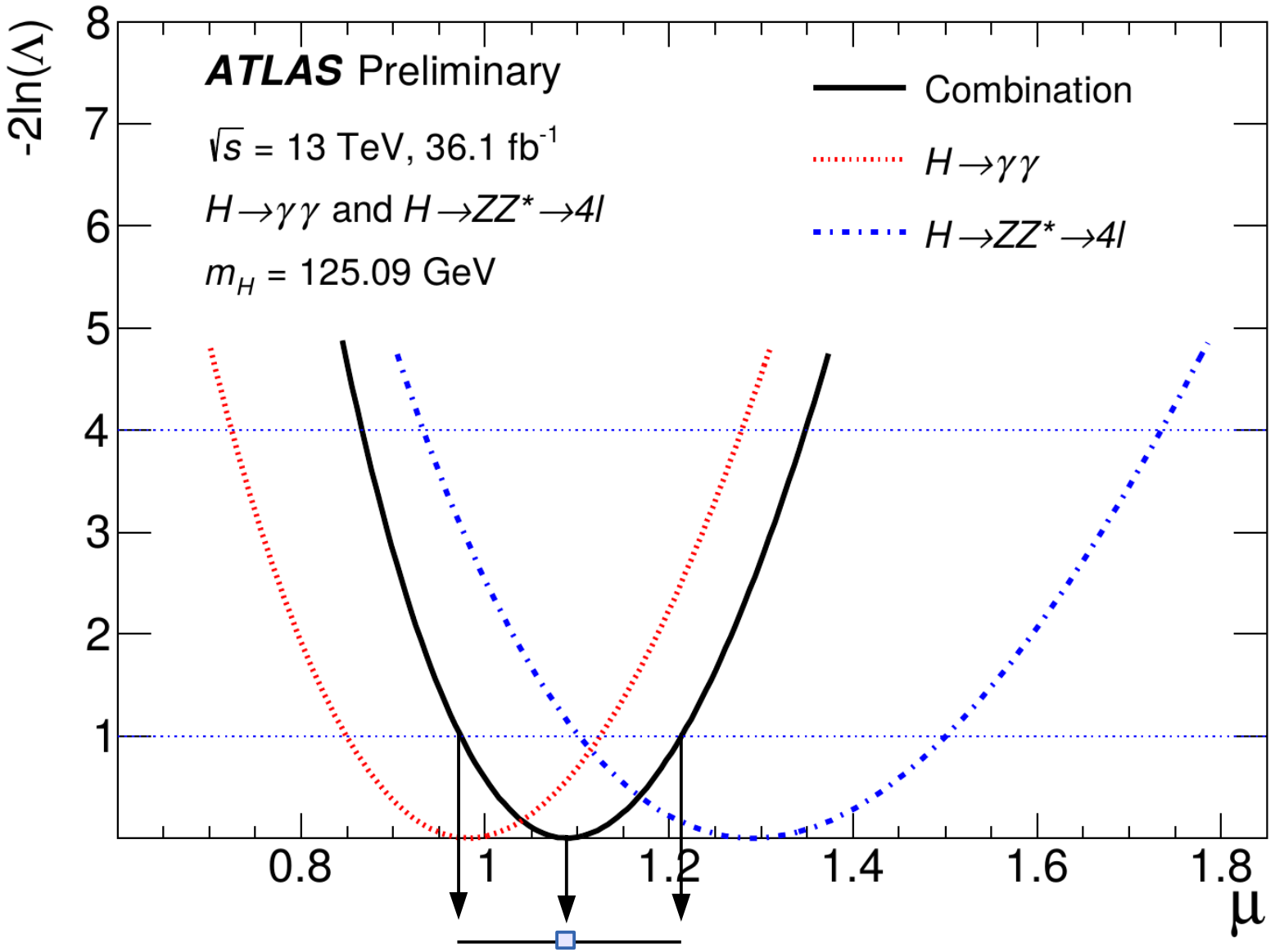
In practice:

- Plot t_μ vs. μ
- The minimum occurs at $\mu=\hat{\mu}$
- Crossings with $t_\mu = 1$ give the $\pm1\sigma$ uncertainties (68% CL interval)



1D Example: μ from H→γγ + H→4l

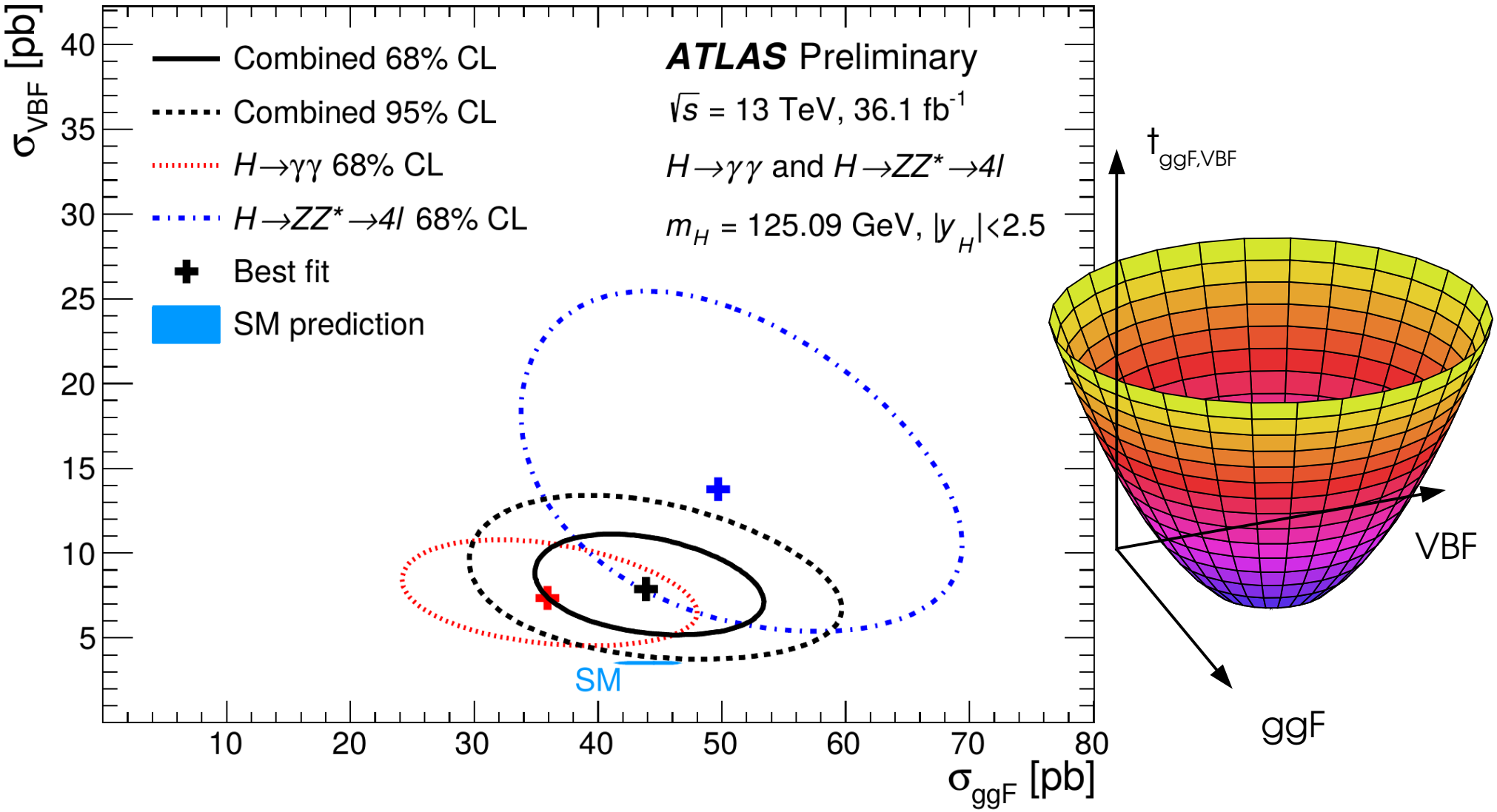
ATLAS-CONF-2017-047



$$\mu = 1.09 \pm 0.12$$

2D Example: Higgs σ_{VBF} vs. σ_{ggF}

ATLAS-CONF-2017-047



χ^2 (n_{dof}=2): **68% C.L.** : $t < 2.3$ ("1σ")

95% C.L. : $t < 6.2$

Uncertainty Decomposition

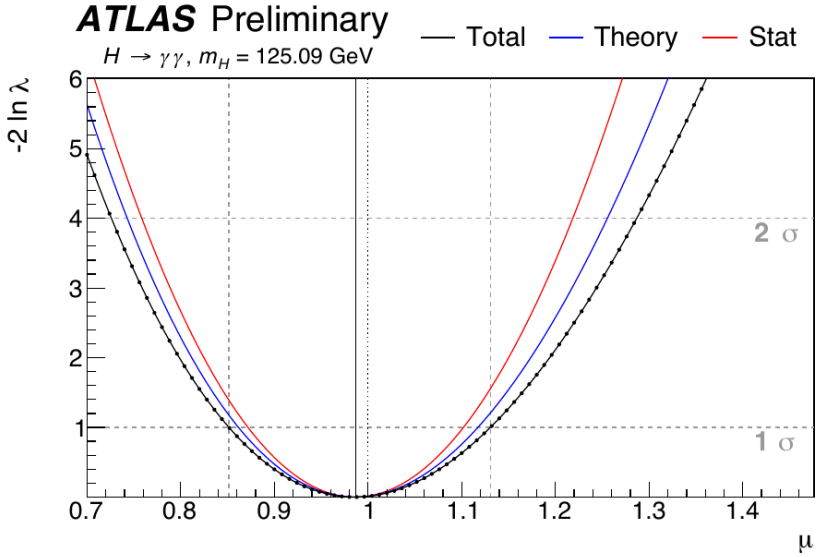
Often useful to break down uncertainties into components (stat + syst, etc.)

PLR approach: perform measurement twice

1. With all uncertainties included
   → **nominal uncertainty σ_{total}.**
2. Removing some uncertainties (e.g. all syst uncertainties) → $\sigma_{no\text{-}syst}$

⇒ Subtract in quadrature:

$$\sigma_{syst} = \sqrt{\sigma^2_{total} - \sigma^2_{no\text{-}syst}}$$



BLUE-based approach:

1. Propagate each source of uncertainty (stat & syst) to the observables
2. Propagate through to the measurement using the BLUE weights

$$\hat{m}_W = \sum_i \lambda_i m_W^{obs,i}$$

The two methods are not completely equivalent (recently discovered!)

→ In the BLUE case, weights still include systematics effects