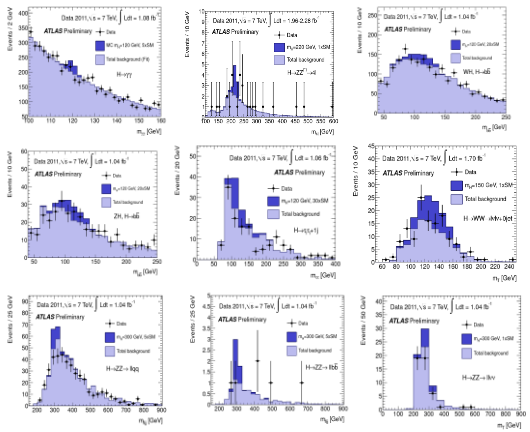
Statistical aspects of Higgs analyses

W. Verkerke

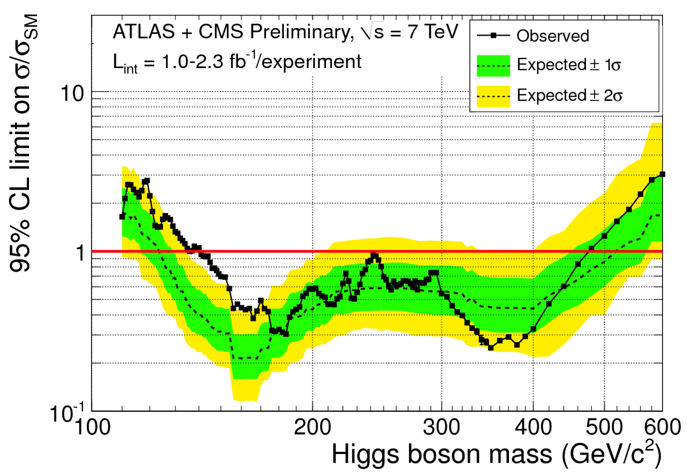
(NIKHEF)

Introduction

- Enormous effort to search for Higgs signature in many decay channels



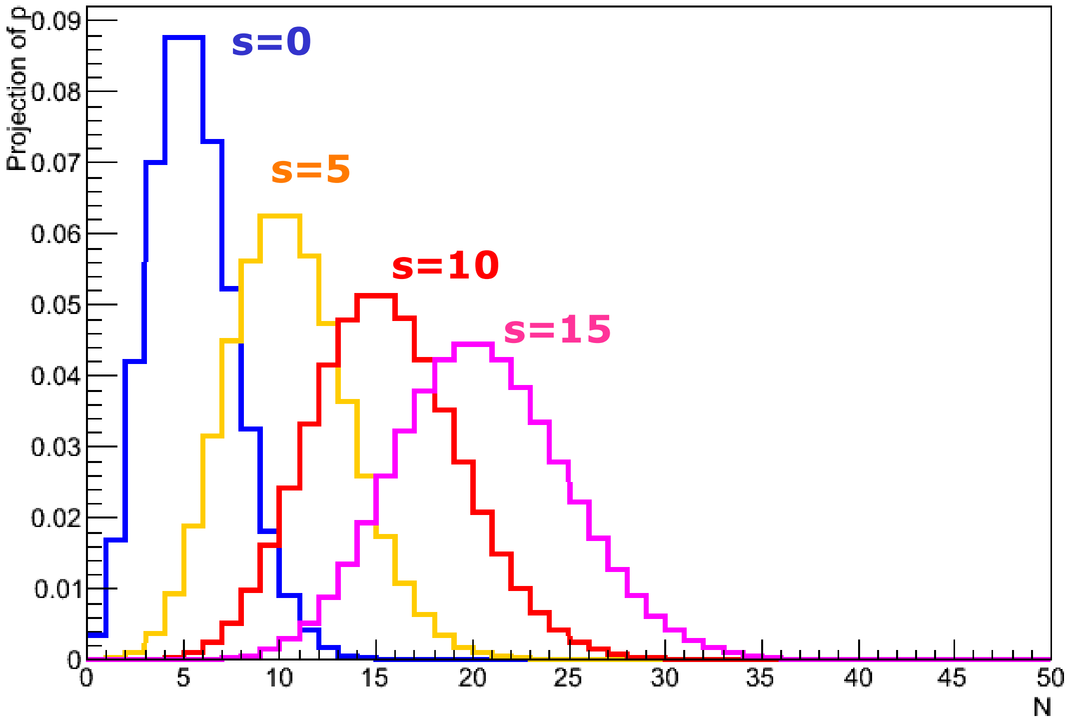
- Results → many plots with signal, background expectations, each with (systematic) uncertainties, and data
- Q: How do you conclude from this that you've seen the Higgs (or not)?
  - Want answer of type:
    'We can exclude that the Higgs exist at 95% CL", or "Probability that background only caused observed excess is $3 \cdot 10^{-7}$



- Here a short guide through how this is (typically) done

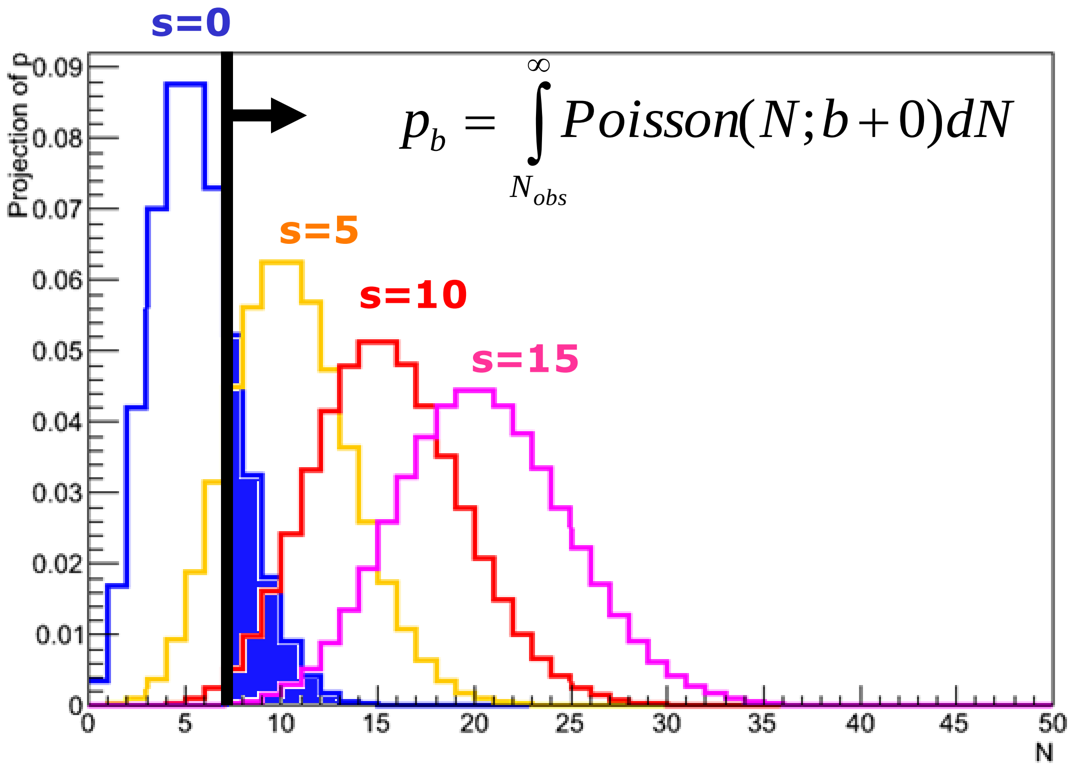
Quantifying discovery and exclusion – Frequentist approach

- Consider the simplest case – a counting experiment
  - Observable: **N** (the event count)
  - Model F(**N**|**s**): Poisson(**N**|**s**+b) with b=5 known exactly
- Predicted distributions of **N** for various values of **s**



Frequentist p-values – excess over expected bkg

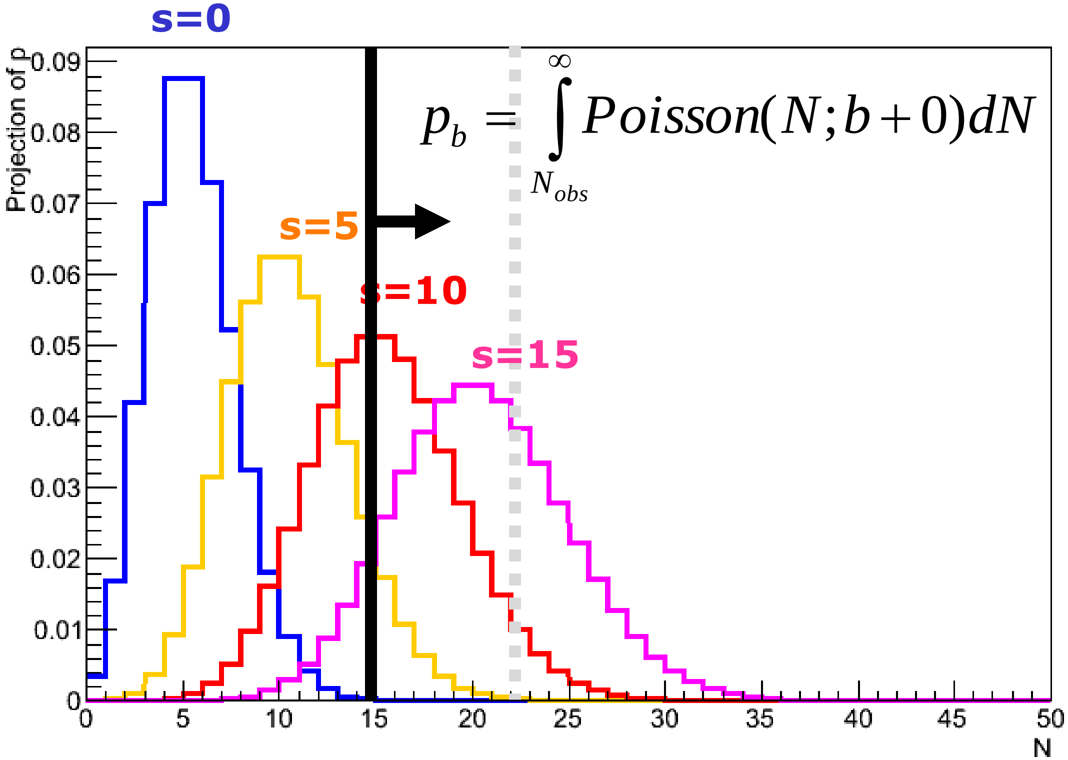
- Now make a measurement N=N_{obs} (example N_{obs}=7)
- Can now define p-value(s), e.g. for bkg hypothesis
  - Fraction of future measurements with N=Nobs (or larger) if s=0



$$p_b = \int_{N_{obs}}^{\infty} Poisson(N; b+0)\,dN \quad (=0.23)$$

Frequentist p-values - excess over expected bkg

- p-values of background hypothesis is used to quantify ‘discovery’ = excess of events over background expectation
- Another example: N_{obs}=15 for same model, what is p_b?

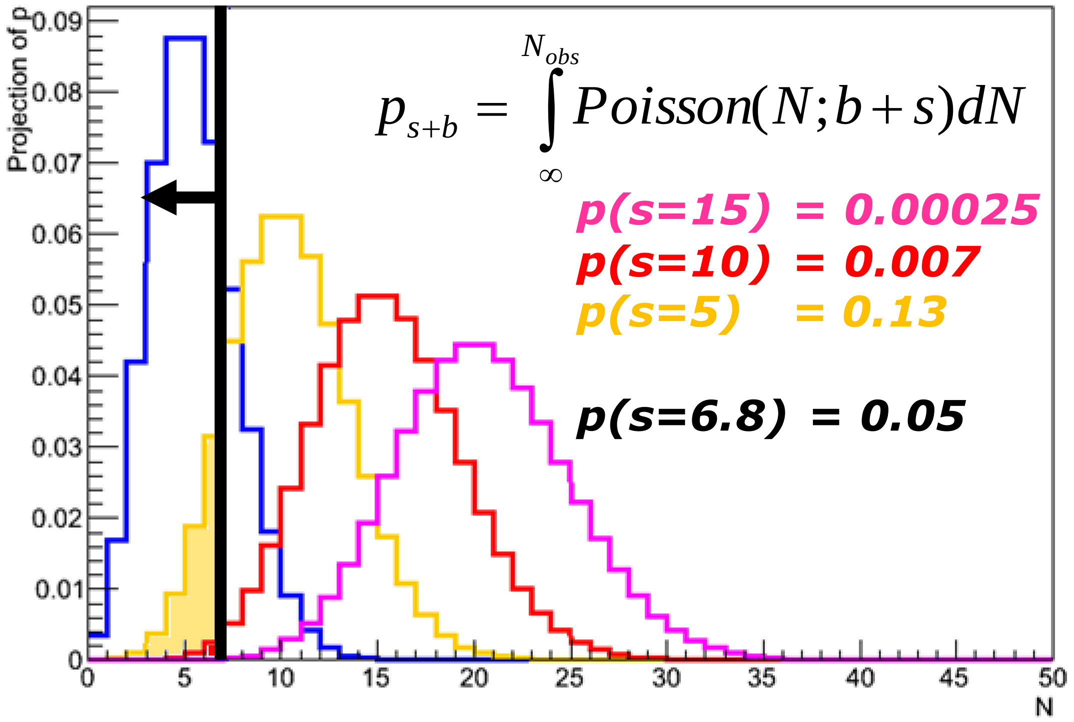


(= 0.00022)

  - Result customarily re-expressed as odds of a *Gaussian fluctuation with equal p-value* (3.5 sigma for above case)
  - NB: N_{obs}=22 gives p_b < 2.810-7 (‘5 sigma’)

Upper limits (one-sided confidence intervals)

- Can also defined p-values for hypothesis with signal: p_{s+b}
  - Note convention: integration range in p_{s+b} is flipped



- Convention: express result as value of **s** for which p(s+b)=5% → **"s>6.8 is excluded at 95% C.L."**

Modified frequentist upper limits

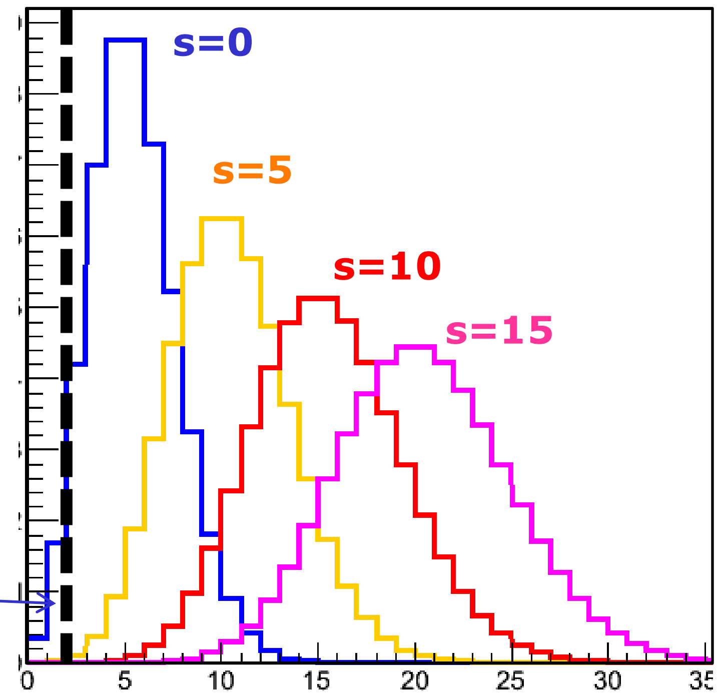
- Need to be careful about interpretation p(s+b) in terms of inference on signal only
  - Since p(s+b) quantifies consistency of signal *plus* background
  - Problem most apparent when observed data has downward stat. fluctations w.r.t background expectation
- Example: $N_{obs}=2$

  → $p_{s+b}(s=0) = 0.04$

  s≥0 excluded at >95% C.L. ?!

- *Modified* approach to protect against such inference on s
  - Instead of requiring p(s+b)=5%, require

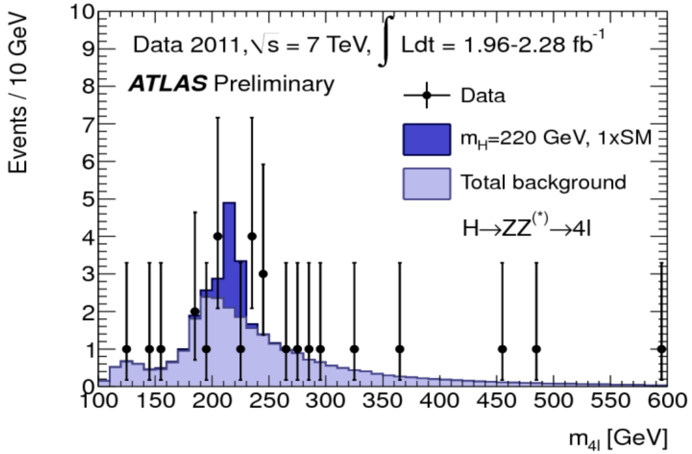
$$CL_S \equiv \frac{p_{s+b}}{1-p_b} = 5\%$$



for N=2 exclude s>3.4 at 95% C.L.s, for large N effect on limit is small as $p_b \to 0$

p-values and limits on non-trivial analysis

- Typical Higgs search result is not a simple number counting experiment, but looks like this:



- Result is a distribution, not a single number
- Models for signal and background have intrinsic uncertainties

- Any type of result can be converted into a single number by constructing a 'test statistic'
  - **A test statistic compresses all signal-to-background discrimination power in a single number**
  - Most powerful discriminators are Likelihood Ratios (Neyman Pearson)

$$q_\mu = -2\ln\frac{L(data\,|\,\mu)}{L(data\,|\,\hat{\mu})}$$

The likelihood ratio test statistic

- Definition: μ = signal strength / signal strength(SM)
  - Choose e.g likelihood with nominal signal strength in numerator (μ=1)

'likelihood assuming nominal signal strength'

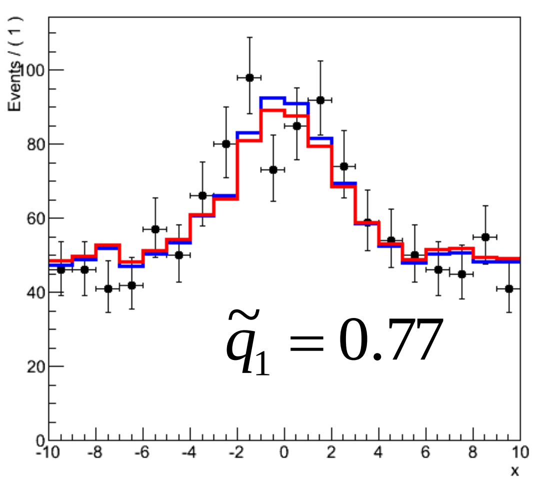
$$q_1 = -2\ln\frac{L(data\,|\,\mu=1)}{L(data\,|\,\hat{\mu})}$$

$\hat{\mu}$ is best fit value of μ

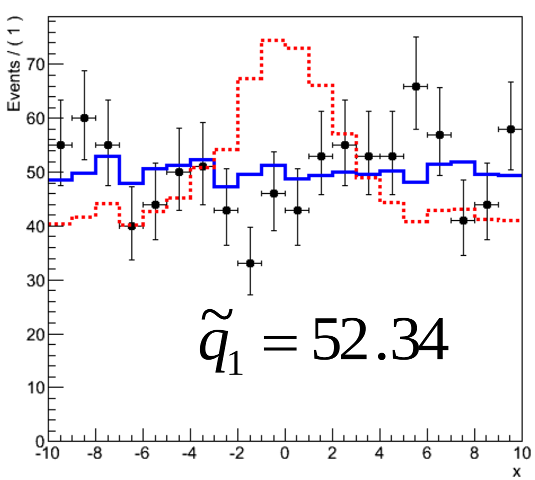
'likelihood of best fit'

- Illustration on model with no shape uncertainties

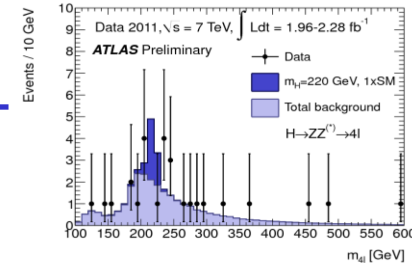
*On signal-like data q1 is **small***



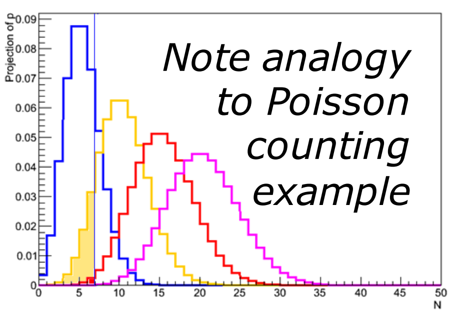
*On background-like data q1 is **large***

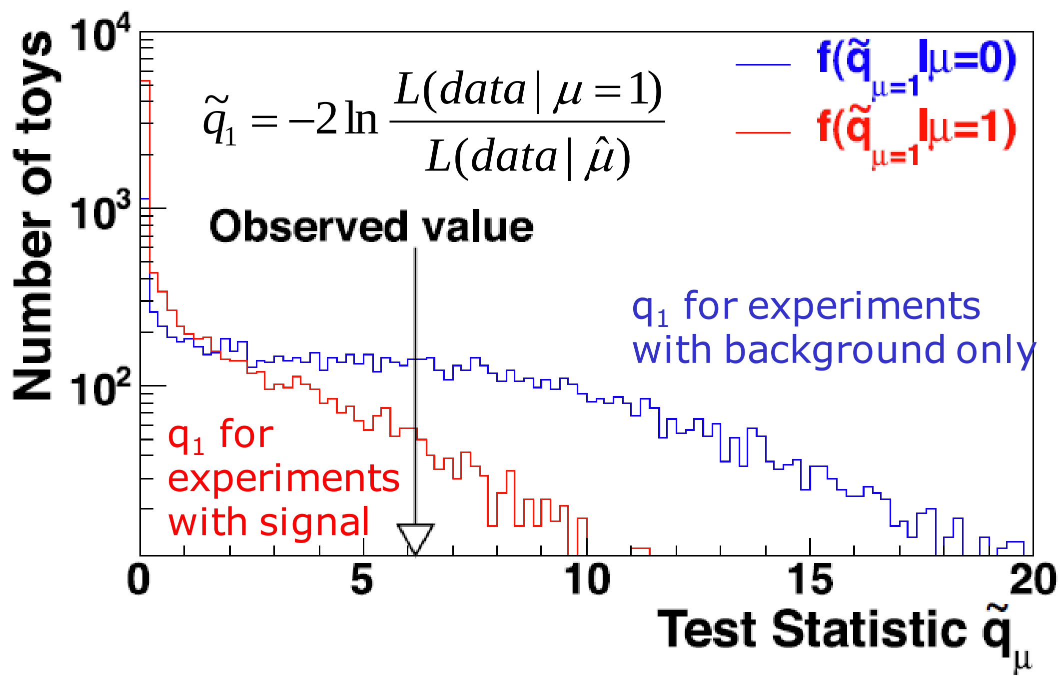


Distributions of test statistics



- Value of q_1 on data is now the 'measurement'
- Distribution of q_1 *not* calculable
  → But can obtain distribution from pseudo-experiments
    - Generate a large number of pseudo-experiments with a given value of mu, calculate q for each, plot distribution





- From $\mathbf{q_{obs}}$ and these distributions, can then set limits similar to what was shown for Poisson counting example

Incorporating systematic uncertainties

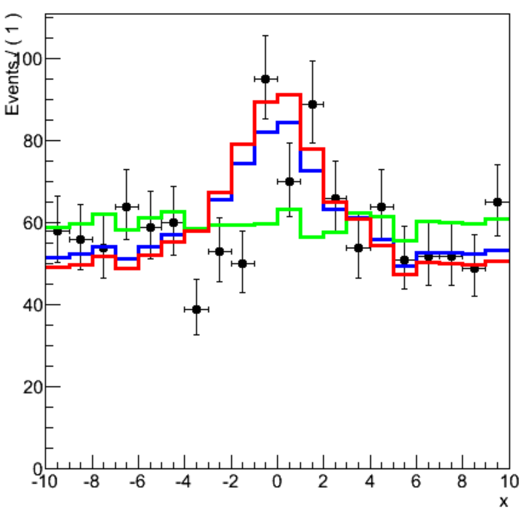
- What happens if models have uncertainties

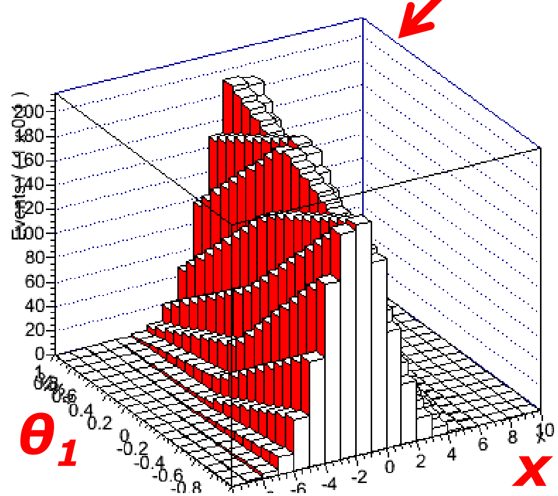
$$L(data \mid \mu) \rightarrow L(data \mid \mu, \vec{\theta})$$

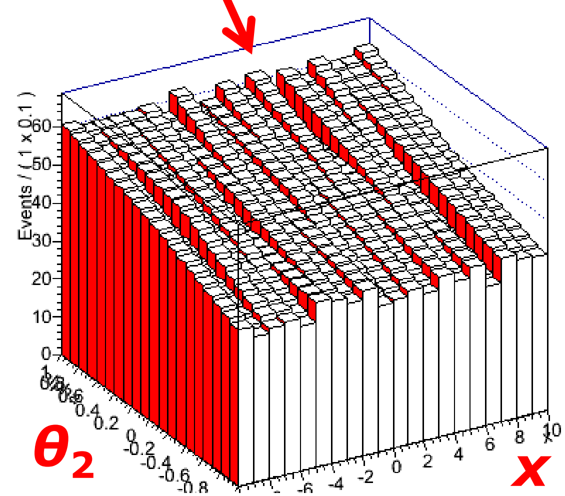
Jet Energy Scale,
QCD scale,
luminosity,
...

  - Introduction of additional model parameters θ that describe effect of uncertainties

$$L(data \mid \mu, \theta) = Poisson(N_i \mid \mu \cdot s_i(\theta) + b_i(\theta)) \cdot p(\tilde{\theta}, \theta)$$

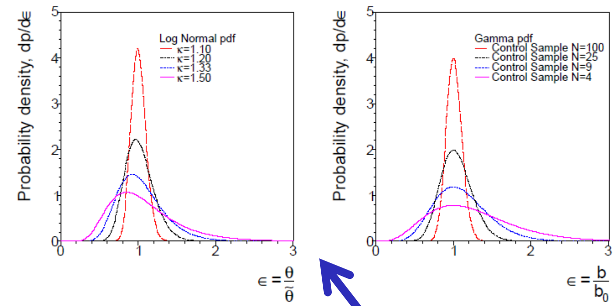




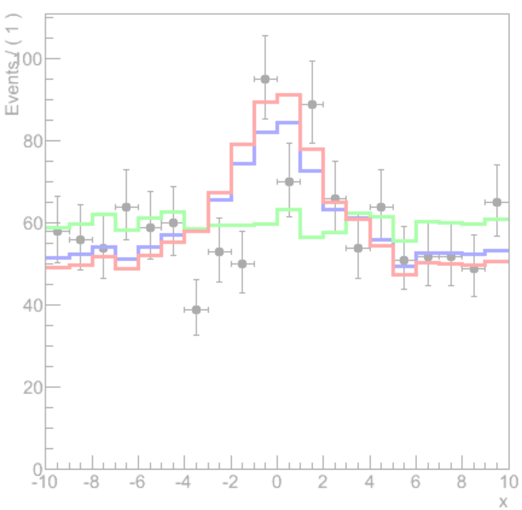


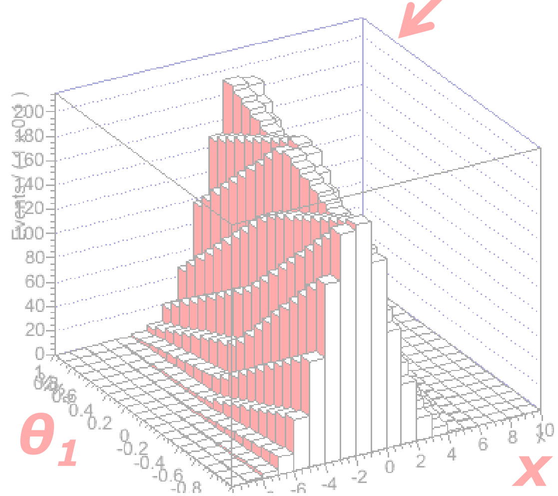
Incorporating systematic uncertainties

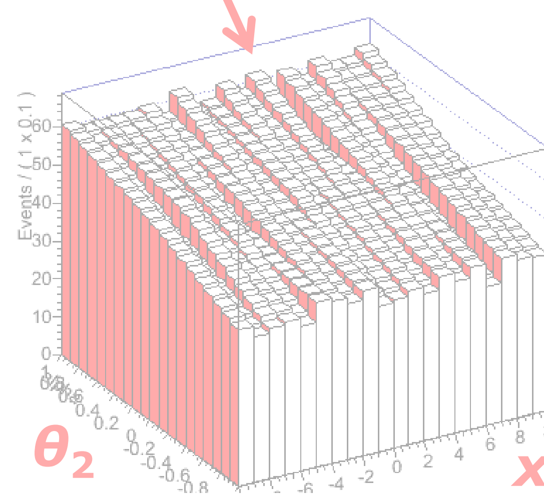
Likelihood includes
auxiliary measurement terms that
constrains the nuisance parameters θ
(shape is flat, log-normal, gamma, or
Gaussian)



$$L(data\,|\,\mu,\theta) = Poisson(N_i\,|\,\mu\cdot s_i(\theta)+b_i(\theta))\cdot p(\tilde{\theta},\theta)$$







Dealing with nuisance parameters in the test statistic

- Uncertainty quantified by nuisance parameters are incorporated in test statistic using a profile likelihood ratio

$$q_\mu = -2\ln\frac{L(data\,|\,\mu)}{L(data\,|\,\hat{\mu})}$$

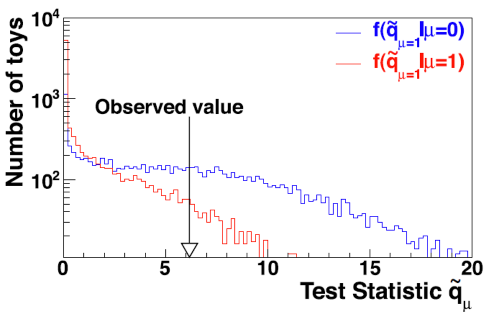
$$\tilde{q}_\mu = -2\ln\frac{L(data\,|\,\mu,\hat{\hat{\theta}}_\mu)}{L(data\,|\,\hat{\mu},\hat{\theta})}$$

'likelihood of best fit for a **given fixed value of μ'**

'likelihood of best fit'

(with a constraint $0 \le \hat{\mu} \le \mu$)

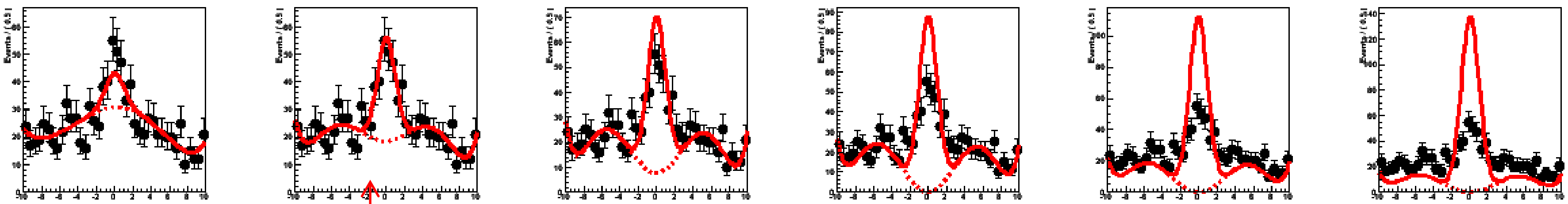
- Interpretation of *observed* value of qμ in terms of p-value again based on expected distribution obtained from pseudo-experiments



Dealing with nuisance parameters in the test statistic

- Uncertainty quantified by nuisance parameters are incorporated using a profile likelihood ratio test statistic

μ=0.35 μ=1.0 μ=1.8 μ=2.6



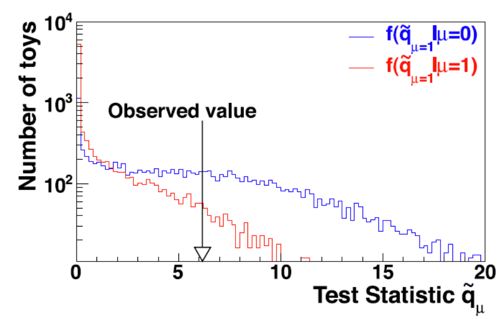
$$\tilde{q}_\mu = -2\ln \frac{L(data\,|\,\mu,\hat{\hat{\theta}}_\mu)}{L(data\,|\,\hat{\mu},\hat{\theta})}$$

'likelihood of best fit for a **given fixed value of μ'**

'likelihood of best fit'

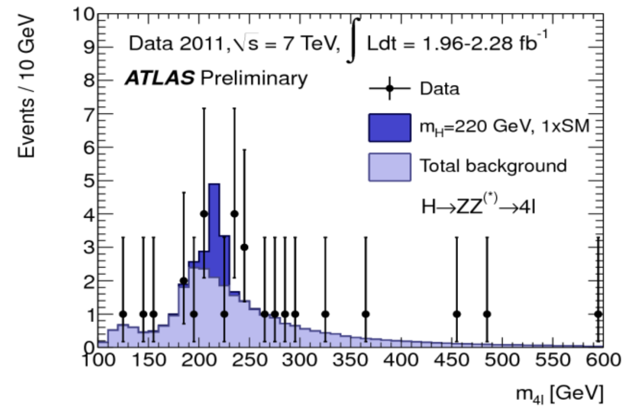
(with a constraint $0 \leq \hat{\mu} \leq \mu$)

- Interpretation of *observed* value of qμ in terms of p-value again based on expected distribution obtained from pseudo-experiments



Putting it all together for one Higgs channel

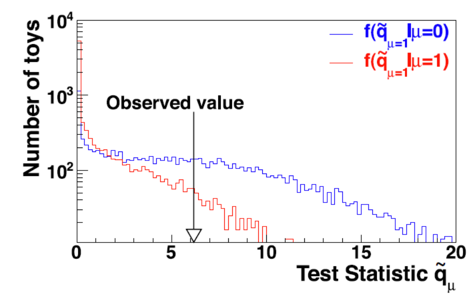
- Result from data D(x)
- PDF $F(x|m_H,\mu,\theta)$ that models the data for a given true Higgs mass hypothesis



- Construct test statistic

$$\tilde{q}_\mu(m_H) = -2\ln \frac{L(data\,|\,\mu, m_H \hat{\hat{\theta}}_\mu)}{L(data\,|\,\hat{\mu}, m_H \hat{\theta})}$$

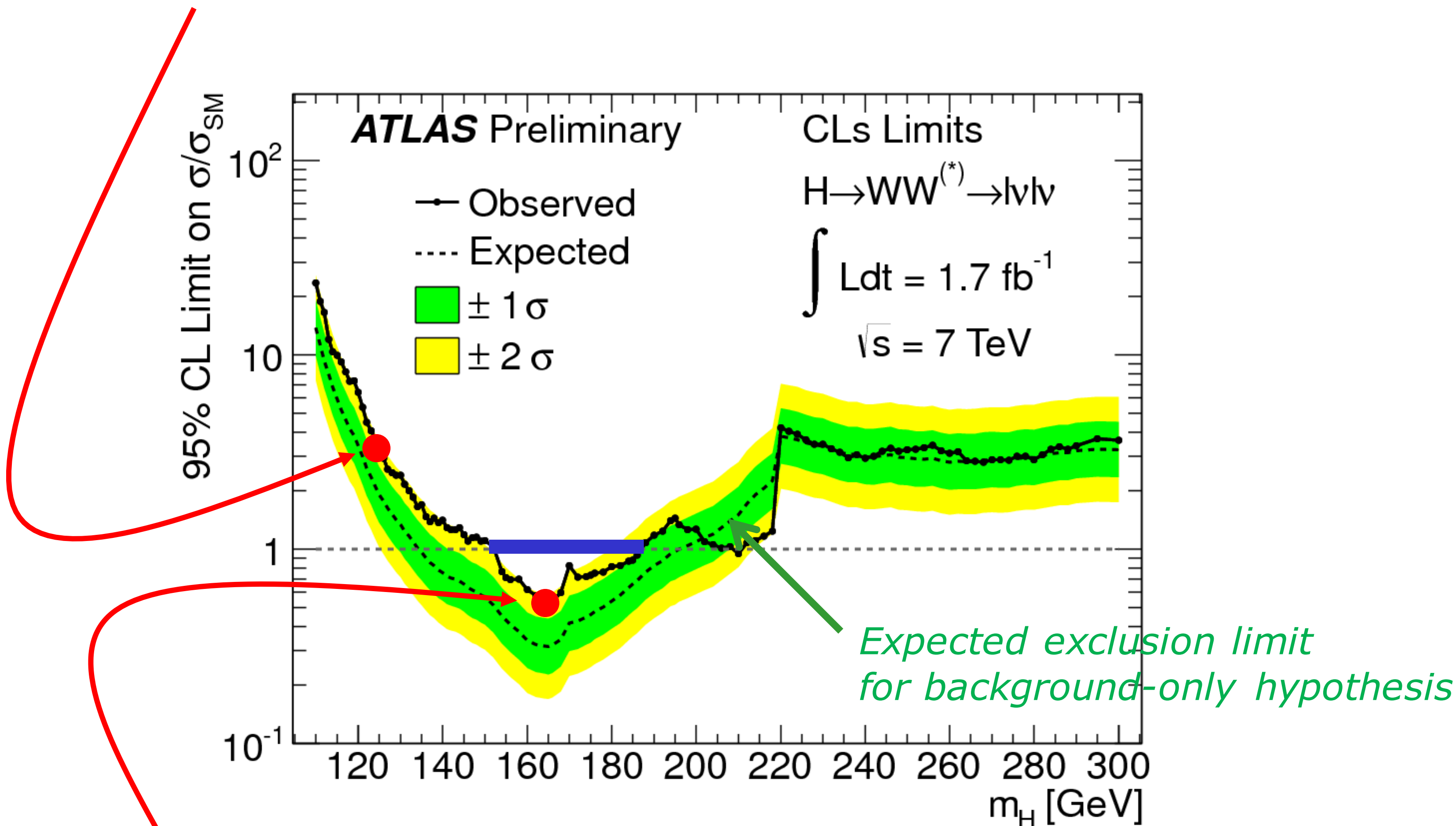
- Obtain expected distributions of q_μ for various μ
  - Determine 'discovery' p-value and signal exclusion limit



- Repeat for each assumed m_H..

Example – 95% Exclusion limit vs m_H for H→WW

Example point: ≈3 x SM H→WW cross-section excluded at m_H=125 GeV

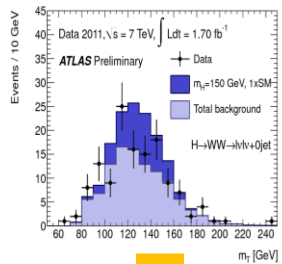


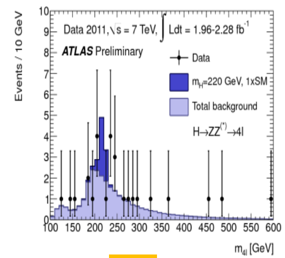
Example point: ≈0.5 x SM H→WW cross-section excluded at m_H=165 GeV

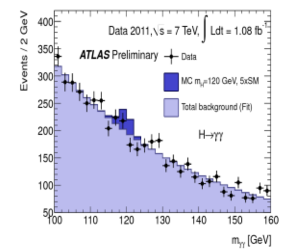
Higgs with 1.0x SM cross-section excluded at 95% CL for m_H in range [150,~187]

Combining Higgs channels (and experiments)

- Procedure: define joint likelihood







$$L(\mu,\theta_{comb}) = L_{H\rightarrow WW}(\mu,\theta_{WW})\cdot L_{H\rightarrow ZZ}(\mu,\theta_{ZZ})\cdot L_{H\rightarrow\gamma\gamma}(\mu,\theta_{\gamma\gamma})\cdot\ldots$$

$$L(\mu,\theta_{LHC}) = L_{ATLAS}(\mu,\theta_{ATLAS})\cdot L_{CMS}(\mu,\theta_{CMS})\cdot\ldots$$

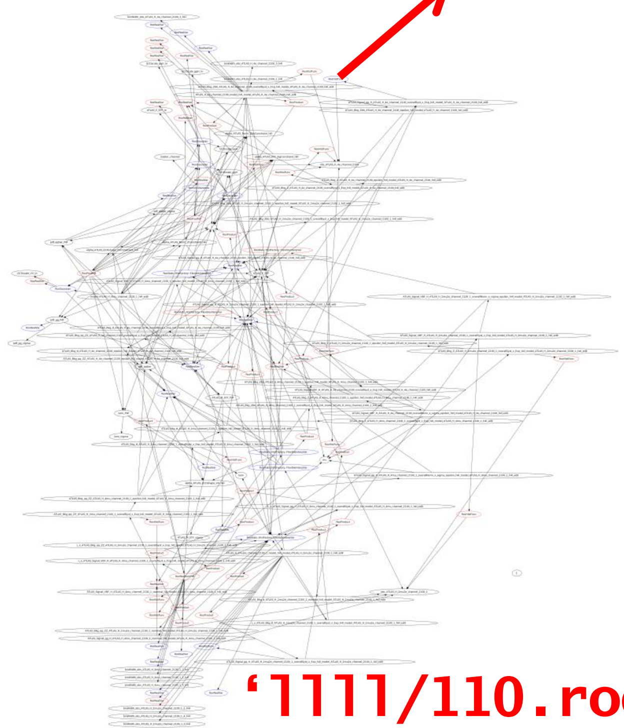
- Correlations between θ_{WW},$\theta_{\gamma\gamma}$ etc and between θ_{ATLAS},θ_{CMS} requires careful consideration!
- The construction profile likelihood ratio test statistic from joint likelihood and proceed as usual

$$\tilde{q}_\mu = -2\ln\frac{L(data\,|\,\mu,\hat{\hat{\theta}}_\mu)}{L(data\,|\,\hat{\mu},\hat{\theta})}$$

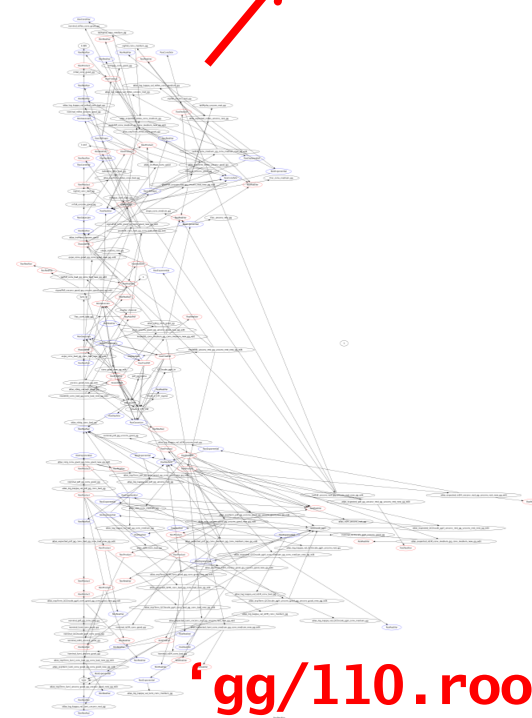
A word on the machinery

- Common tools (RooFit/RooStats - all available in ROOT) have been developed in past 2/3 years to facilitate these combinations
  - Analytical description of likelihood of each component stored and delivered in a uniform language (‘RooFit workspaces’)
  - Construction of joint likelihood *technically* straightforward

$$L(\mu,\theta_{comb}) = L_{H\to WW}(\mu,\theta_{WW})\cdot L_{H\to ZZ}(\mu,\theta_{ZZ})\cdot L_{H\to\gamma\gamma}(\mu,\theta_{\gamma\gamma})\cdot\ldots$$

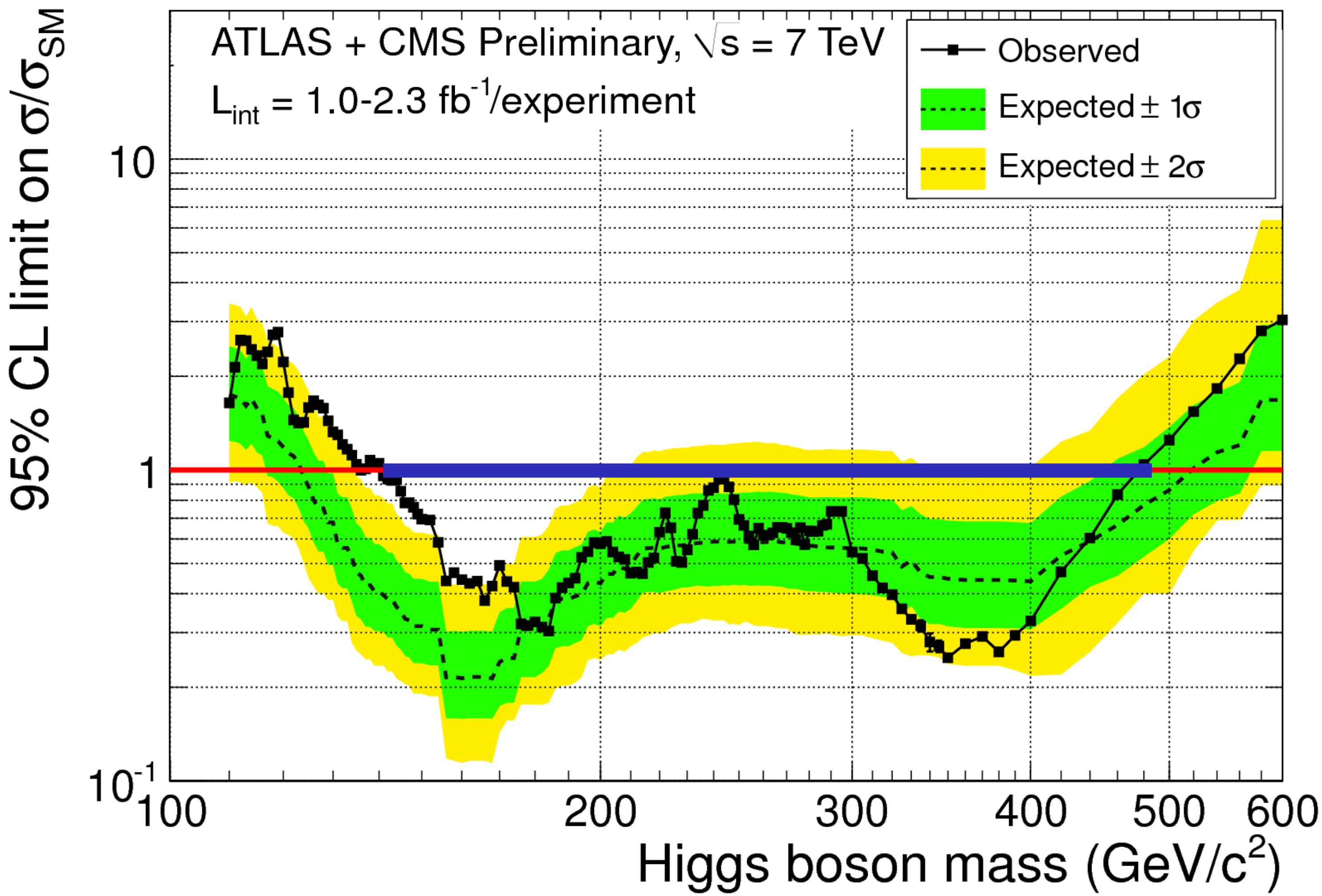


‘llll/110.root’



‘gg/110.root’

Example – joint ATLAS/CMS Higgs exclusion limit



Switching from ‘exclusion’ to ‘discovery’ formulation

‘exclusion’

‘likelihood assuming μ signal strength’

$$\tilde{q}_\mu = -2\ln \frac{L(data \mid \mu, \hat{\hat{\theta}}_\mu)}{L(data \mid \hat{\mu}, \hat{\theta})}$$

‘likelihood of best fit’

‘discovery’

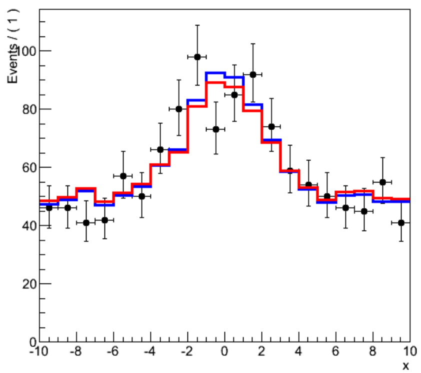
‘likelihood assuming background only’

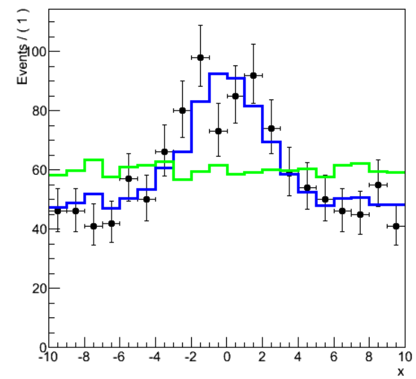
$$\tilde{q}_0 = -2\ln \frac{L(data \mid \mu = 0, \hat{\hat{\theta}}_0)}{L(data \mid \hat{\mu}, \hat{\theta})}$$

‘likelihood of best fit’

simulated data with signal+bkg

$\tilde{q}_1 = 0.77$

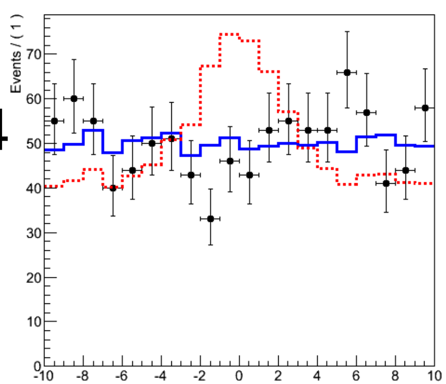


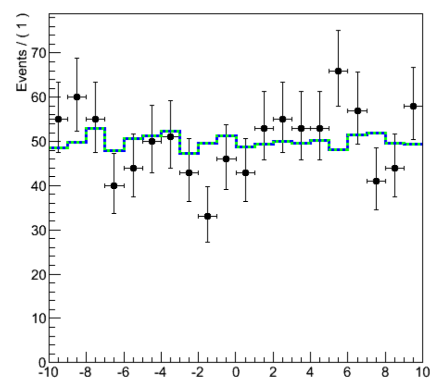


$\tilde{q}_0 = 34.77$

simulated data with bkg only

$\tilde{q}_1 = 52.34$

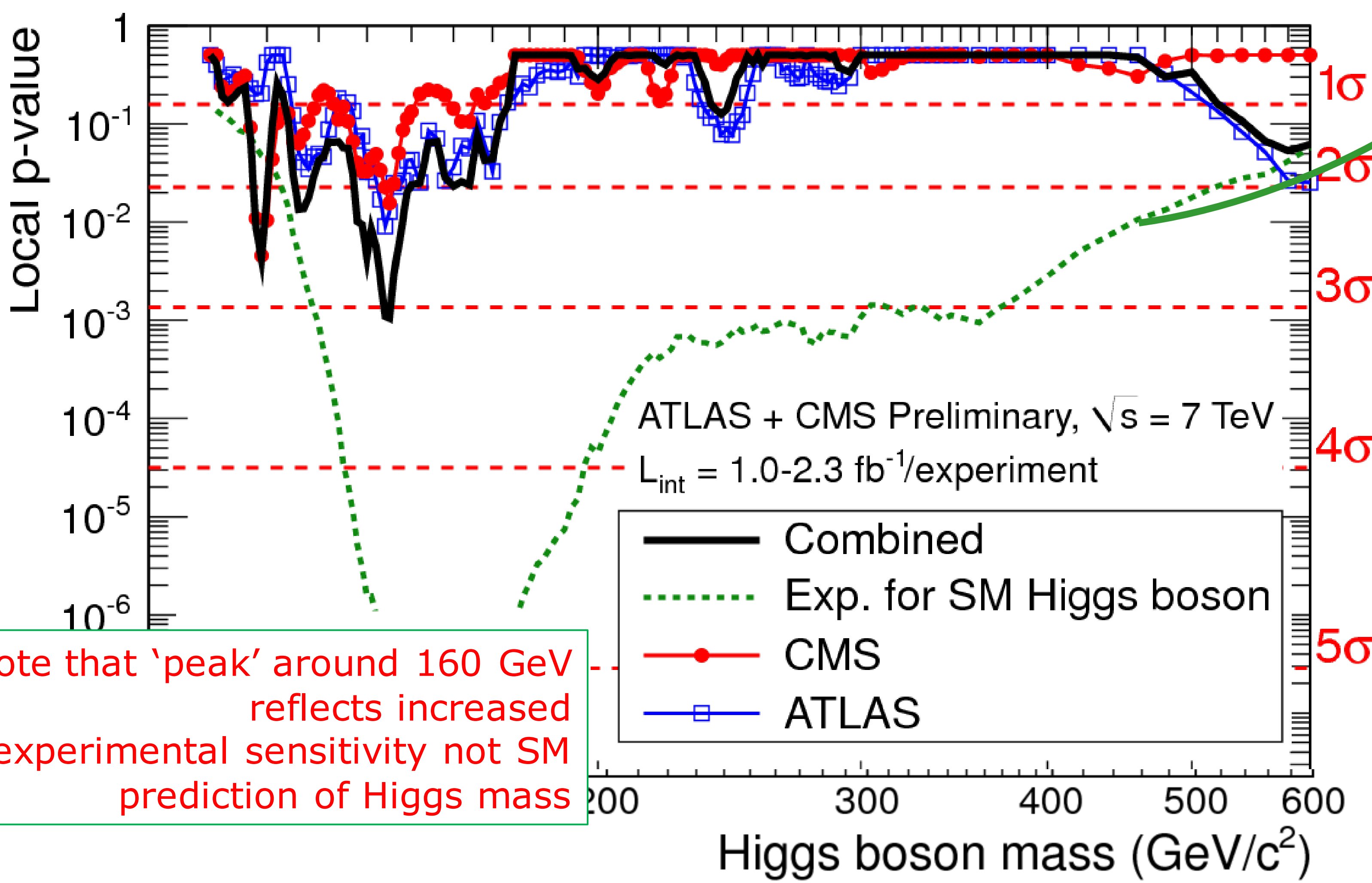




$\tilde{q}_0 = 0$

Comb: p-value of background-only hypothesis ('discovery')

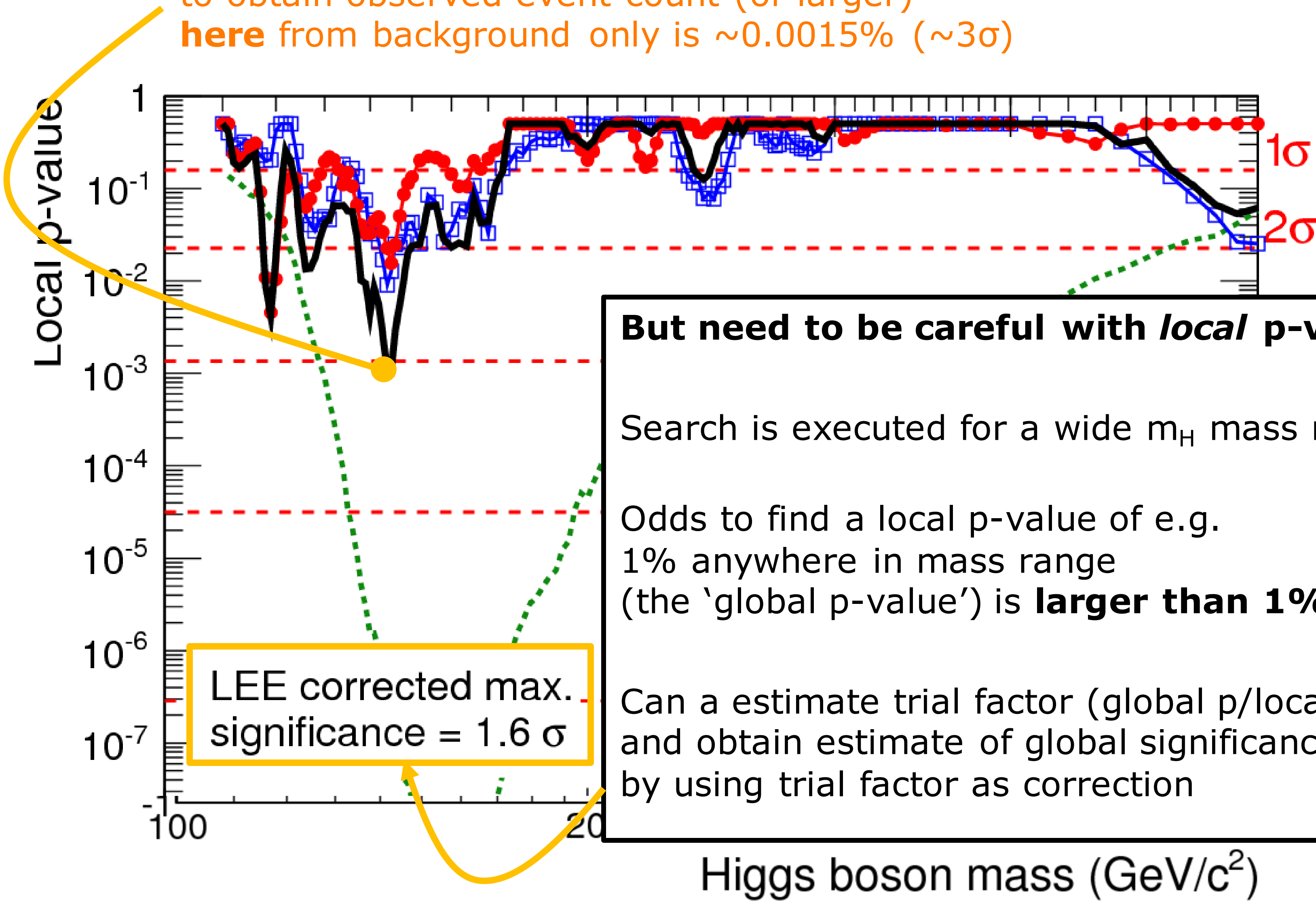
Expected p-value for background hypothesis of a data sample containing SM Higgs boson



Note that 'peak' around 160 GeV reflects increased experimental sensitivity not SM prediction of Higgs mass

Comb: p-value of background-only hypothesis ('discovery')

Example point: at $m_H \approx 150$ GeV probability to obtain observed event count (or larger) **here** from background only is ~0.0015% (~3σ)



But need to be careful with *local* p-values

Search is executed for a wide m_H mass range.

Odds to find a local p-value of e.g. 1% anywhere in mass range (the 'global p-value') is **larger than 1%**

Can a estimate trial factor (global p/local p) and obtain estimate of global significance by using trial factor as correction

Conclusions

- We are early awaiting more data!