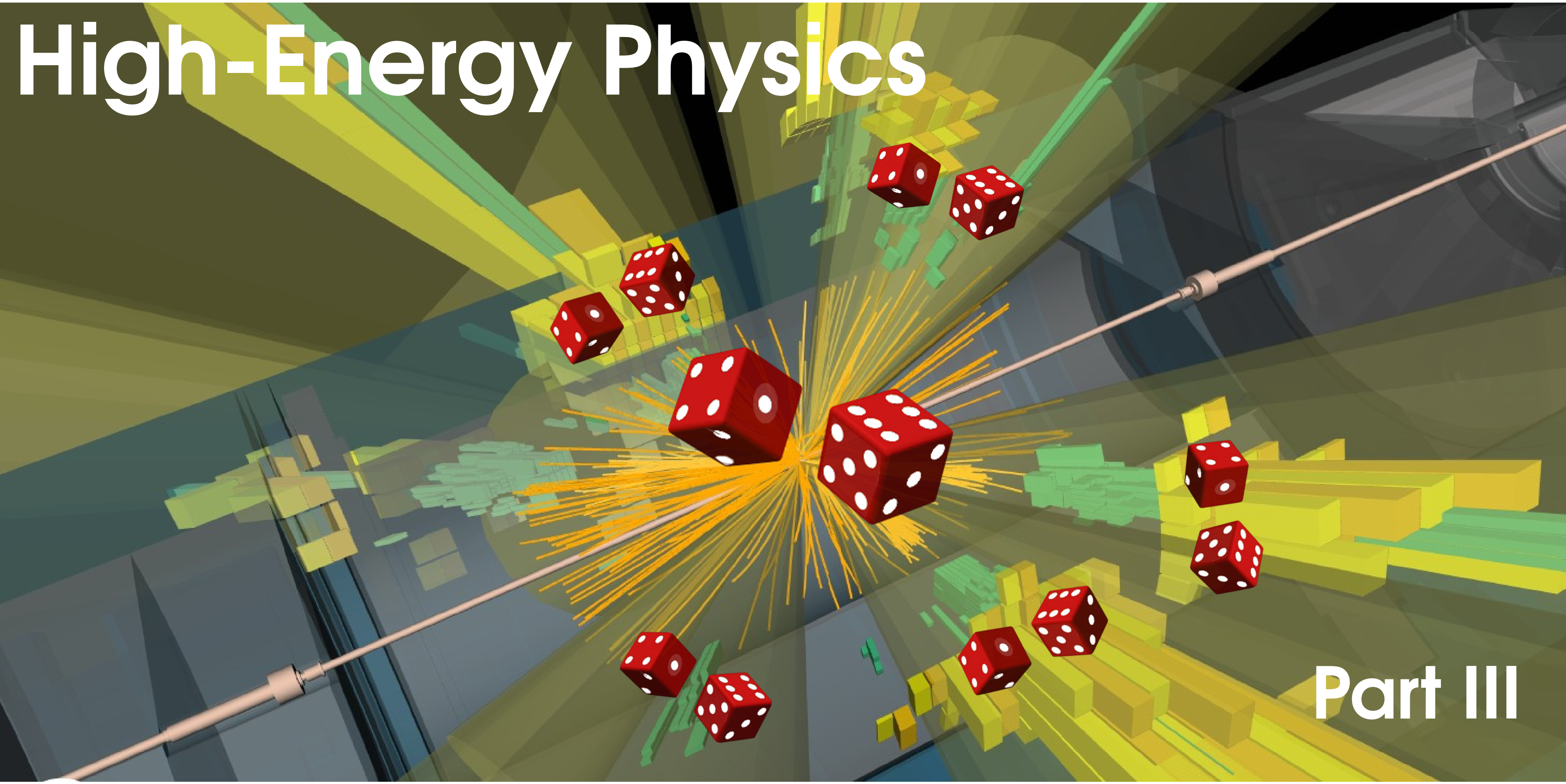
Statistical analysis methods in High-Energy Physics



Part III

Nicolas Berger (LAPP Annecy)

Outline

Profiling

Look-Elsewhere Effect

Bayesian methods

Statistical modeling in practice

- Building binned likelihoods
- Choosing PDFs in unbinned likelihoods
- Implementing systematics

BLUE

Frequentist Constraints

Prototype: NP measured in a separate ***auxiliary* experiment**
e.g. luminosity measurement

→ Build the combined likelihood of the main+auxiliary measurements

$$L(\mu,\theta;\mathbf{data}) = L_{\mathbf{main}}(\mu,\theta;\mathbf{main\ data})\ L_{\mathbf{aux}}(\theta;\mathbf{aux.\ data})$$

Independent measurements:
⇒ just a product

Gaussian form often used by default: $L_{\mathbf{aux}}(\theta;\mathbf{aux.\ data}) = G(\theta^{\mathbf{obs}};\theta,\sigma_{\mathbf{syst}})$

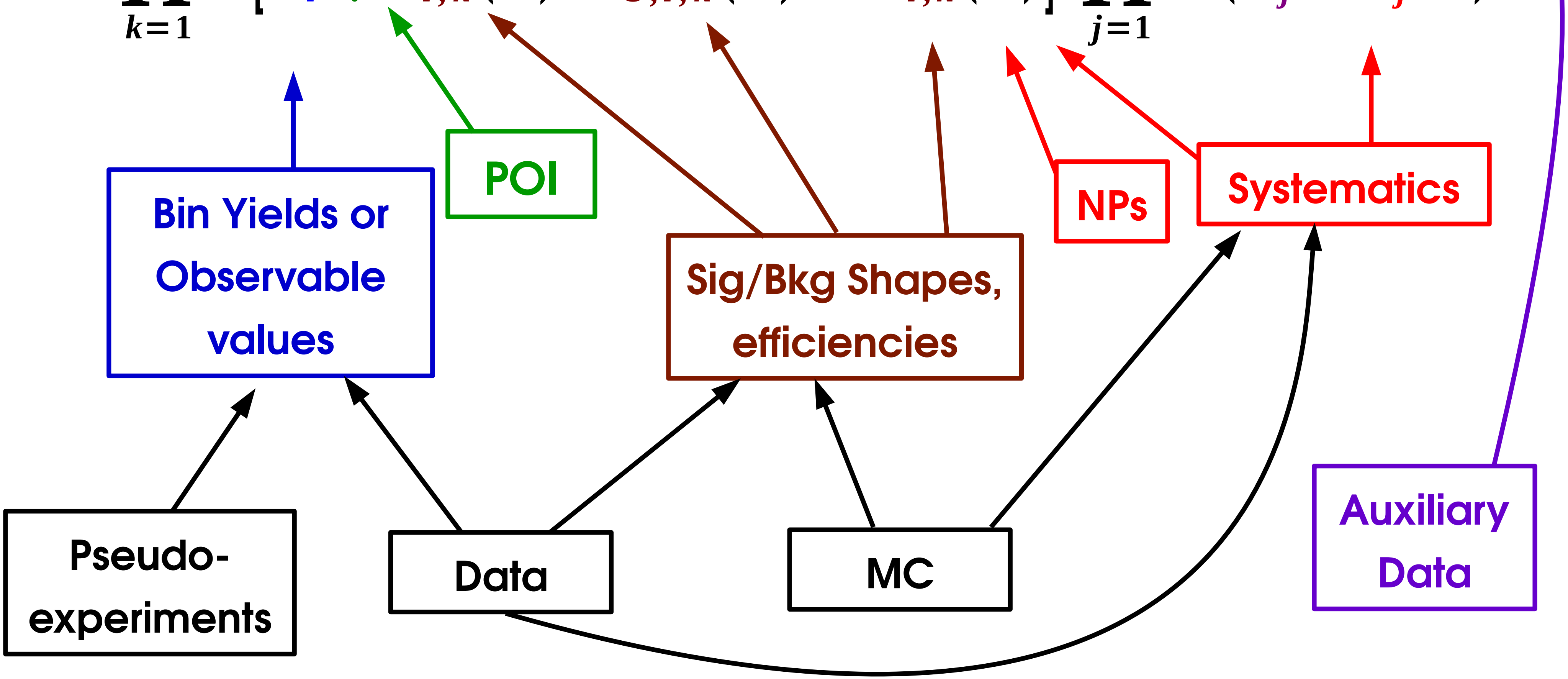
In the combined likelihood, **systematic NPs are constrained**
→ now same as other NPs: **all uncertainties statistical in nature**

→ Often no clear setup for auxiliary measurements
e.g. theory uncertainties on missing HO terms from scale variations
→ Implemented in the same way nevertheless ("pseudo-measurement")

Likelihood, the full version (binned case)

$$L\left(\mu,\{\theta_j\}_{j=1\ldots n_{NP}};\{n_i^{(k)}\}_{i=1\ldots n_{data}^{(k)}}^{k=1\ldots n_{cat}},\{\theta_j^{obs}\}_{j=1..n_{NP}}\right)=$$

$$\prod_{k=1}^{n_{cat}} P\left[n_i;\overbrace{\mu\,\epsilon_{i,k}(\vec{\theta})\;N_{S,i,k}(\vec{\theta})+B_{i,k}(\vec{\theta})}^{\text{Expected bin yield}}\right]\prod_{j=1}^{n_{syst}} G\left(\theta_j^{obs};\theta_j;1\right)$$



× number of categories!

Wilks' Theorem

The likelihood usually has NPs:

- **Systematics**
- Parameters fitted in data

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

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

⇒ **Profile Likelihood Ratio** (PLR)

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

$\hat{\hat{\theta}}_{\mu_0}$ 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 ! $f\left(t_{\mu_0} \mid \mu=\mu_0\right) = f_{\chi^2(n_{dof}=1)}\left(t_{\mu_0}\right)$

also with NPs present

→ Profiling "builds in" the effect of the NPs

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

Gaussian Profiling

Measure **N(S,θ) = S + θ** :

→ Main measurement **n ~ G(S + θ, σ_n)**

→ constraint (aux. meas.) : **θ^{obs} ~ G(θ, σ_θ)**

$$L(S,\theta) = G(n; S+\theta, \sigma_n)\, G(\theta^{obs}; \theta, \sigma_\theta)$$

Then: $$\lambda(S,\theta) = \left(\frac{n-(S+\theta)}{\sigma_n}\right)^2 + \left(\frac{\theta^{obs}-\theta}{\sigma_\theta}\right)^2$$

MLEs: $$\hat{S} = n - \theta^{obs}$$
$$\hat{\theta} = \theta^{obs}$$

Conditional MLE: $$\hat{\hat{\theta}}(S) = \theta^{obs} + \frac{\sigma_\theta^2}{\sigma_n^2+\sigma_\theta^2}(\hat{S}-S)$$

For S = Ŝ, matches MLE as it should

PLR: $$t_{S_0} = -2\log\frac{L(S=S_0, \hat{\hat{\theta}}_{S_0})}{L(\hat{S},\hat{\theta})}$$

$$= \lambda(S_0, \hat{\hat{\theta}}(S_0)) - \lambda(\hat{S},\hat{\theta}) = \frac{(S_0-\hat{S})^2}{\sigma_n^2+\sigma_\theta^2}$$

Statistical Uncertainty **Systematic Uncertainty**

$$\sigma_S = \sqrt{\sigma_n^{\,2} + \sigma_\theta^{\,2}}$$

Stat uncertainty (on n) and syst (on θ) add in quadrature as expected

Effect of Profiling

Systematics still affect the result even after profiling their NPs!

e.g. **again counting experiment: N(S,θ) = S + θ**, measure n, constraint on θ~0.

1. No NP: **N(S) = S**

$$t_{S_0} = -2 \log \frac{L(S_0)}{L(\hat{S})}$$

→ **Ŝ fit**: adjust S to **N(Ŝ) = Ŝ = n**

→ **S=S$_0$ fit**: S=S$_0$ fixed ⇒ N(S$_0$) = S$_0$, **cannot adjust**

⇒ **tension** between N(S$_0$)=S$_0$ and n ⇒ large t$_{S0}$ ⇒ **strong exclusion of H(S$_0$)**

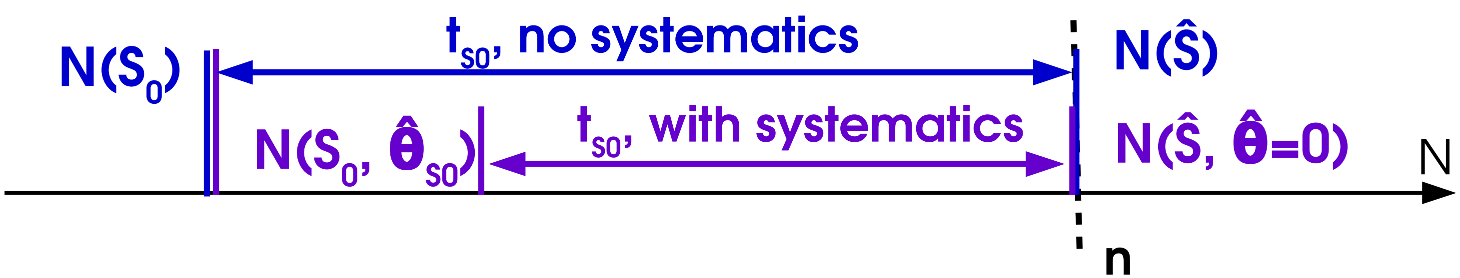
$$t_{S_0} = -2 \log \frac{L(S=S_0, \hat{\hat{\theta}}_{S_0})}{L(\hat{S}, \hat{\theta})}$$

2. With NP: **N(μ,θ) = S + θ**

→ **Ŝ fit**: adjust **N(Ŝ, θ̂) = N(Ŝ, θ̂=0) = n** using S only (avoid penalty on θ)

→ **S=S$_0$ fit**: S=S$_0$ fixed, but $\hat{\hat{\theta}}$(S$_0$) **can still pull N(S$_0$,$\hat{\hat{\theta}}$(S$_0$)) towards n**

⇒ smaller t$_{S0}$ ⇒ **reduced exclusion of H(S$_0$)**



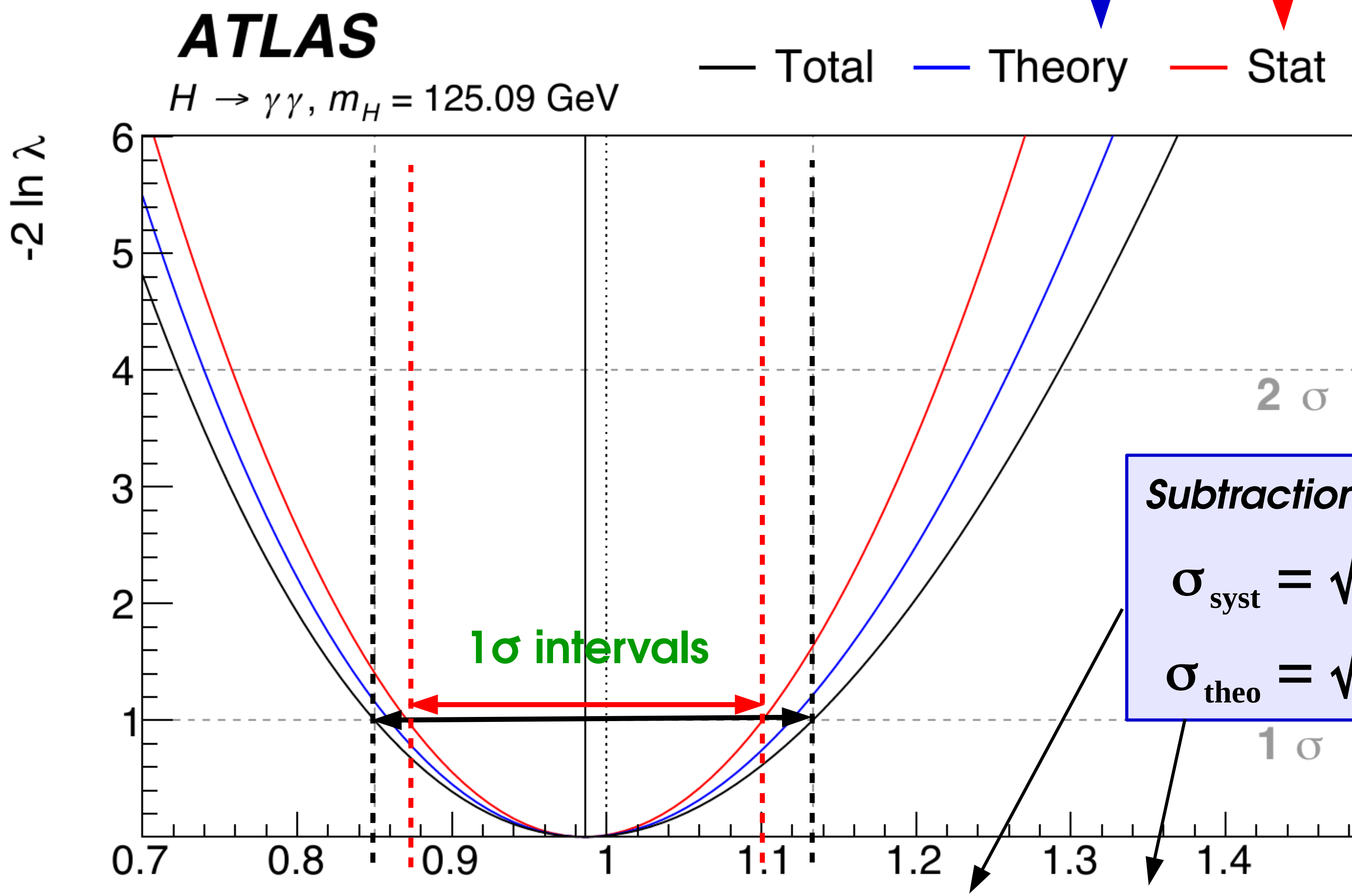
More freedom

⇒ Weaker exclusion

Uncertainty decomposition

All systematics NPs fixed : statistical uncertainty only

exp. syst. NPs fixed : stat+theory uncertainty



Subtraction in quadrature

$$\sigma_{syst} = \sqrt{\sigma^2_{total} - \sigma^2_{stat}}$$

$$\sigma_{theo} = \sqrt{\sigma^2_{total} - \sigma^2_{theo}}$$

$$\mu = 0.99 \pm 0.12\,(\text{stat}) \pm 0.06\,(\text{syst}) \pm 0.06\,(\text{theo})$$

Gaussian Profiling

Gaussian measurement with 1 POI μ and 1 NP θ:

$$L(\mu,\theta;\hat{\mu},\hat{\theta}) = \exp\left[-\frac{1}{2}\begin{pmatrix}\mu-\hat{\mu}\\ \theta-\hat{\theta}\end{pmatrix}^T C^{-1}\begin{pmatrix}\mu-\hat{\mu}\\ \theta-\hat{\theta}\end{pmatrix}\right] \qquad C = \begin{bmatrix}\sigma_\mu^2 & \gamma\sigma_\mu\sigma_\theta\\ \gamma\sigma_\mu\sigma_\theta & \sigma_\theta^2\end{bmatrix}$$

"data"

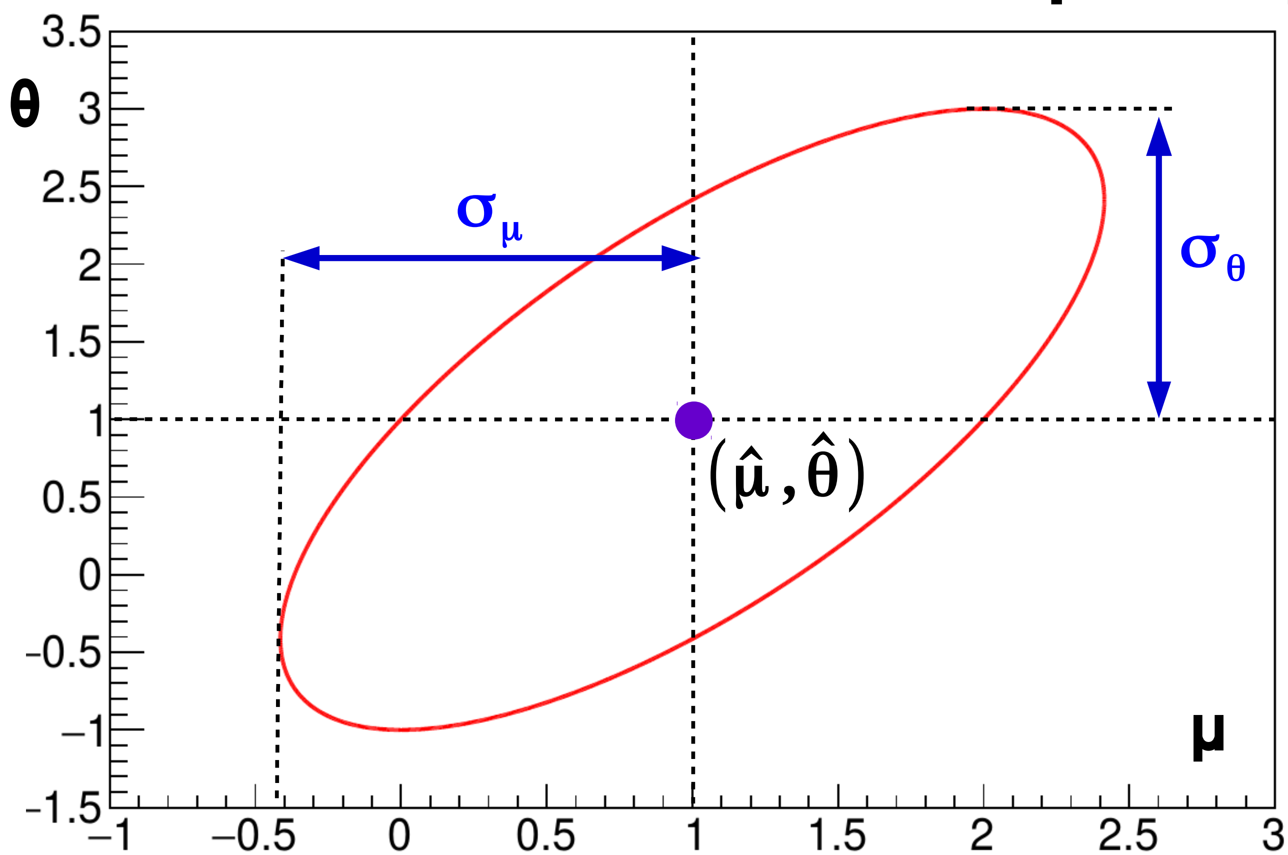
→ λ(μ, θ) defines an **ellipse:**

$$F \equiv C^{-1}$$

$$\lambda(\mu,\theta;\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 + 2F_{\mu\theta}(\mu-\hat{\mu})(\theta-\hat{\theta}) + F_{\theta\theta}(\theta-\hat{\theta})^2 \qquad = \begin{bmatrix}F_{\mu\mu} & F_{\mu\theta}\\ F_{\mu\theta} & F_{\theta\theta}\end{bmatrix}$$

Uncertainty on μ:

- **From C, with θ included:** σ_μ



Gaussian Profiling

$$C = \begin{bmatrix} \sigma_\mu^2 & \gamma\sigma_\mu\sigma_\theta \\ \gamma\sigma_\mu\sigma_\theta & \sigma_\theta^2 \end{bmatrix}$$

$$F = \begin{bmatrix} F_{\mu\mu} & F_{\mu\theta} \\ F_{\mu\theta} & F_{\theta\theta} \end{bmatrix}$$

$$\lambda(\mu,\theta;\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 + 2F_{\mu\theta}(\mu-\hat{\mu})(\theta-\hat{\theta}) + F_{\theta\theta}(\theta-\hat{\theta})^2$$

Profiled θ (minimize λ at fixed μ) :

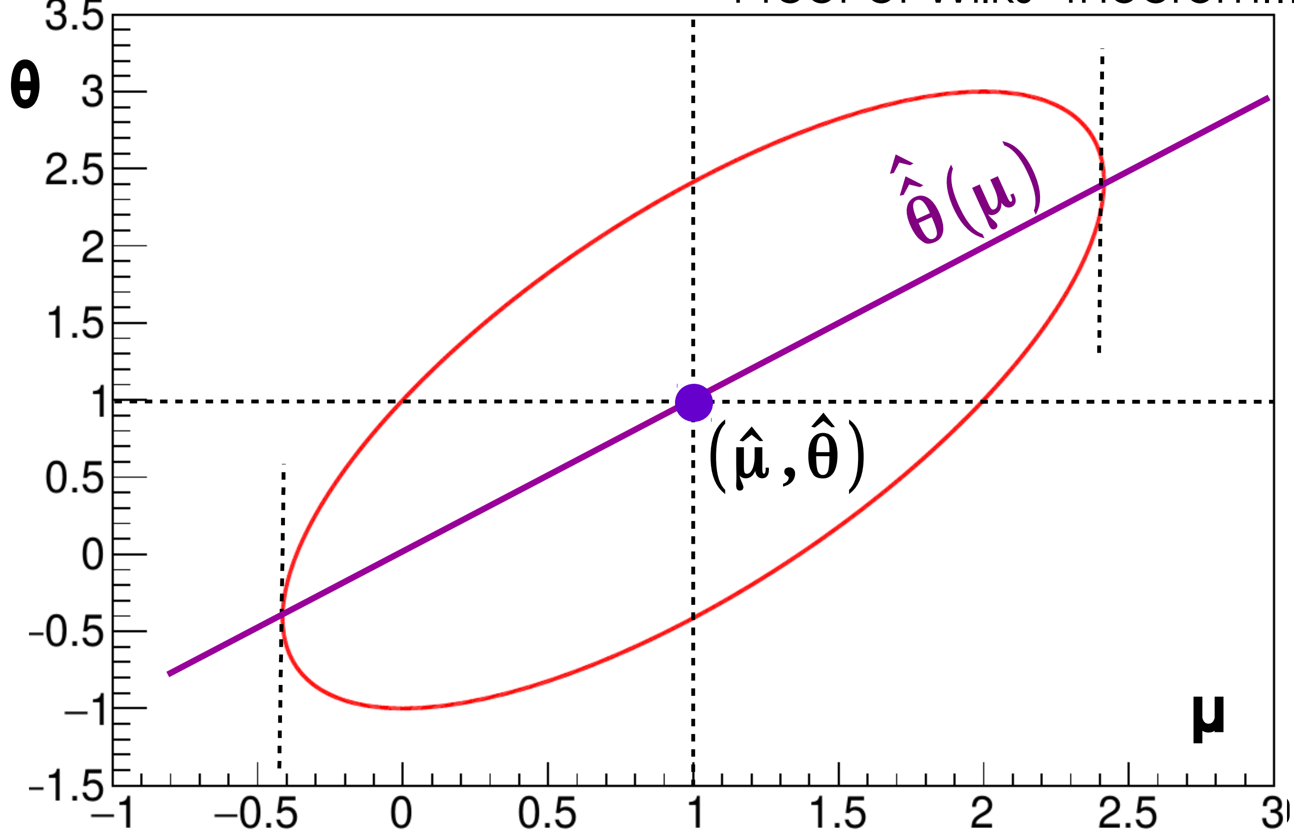
$$\hat{\hat{\theta}}(\mu) = \hat{\theta} - F_{\theta\theta}^{-1} F_{\theta\mu}(\mu - \hat{\mu})$$

Profile likelihood ratio:

$$\lambda(\mu,\hat{\hat{\theta}}(\mu);\hat{\mu},\hat{\theta}) = \left(F_{\mu\mu} - F_{\mu\theta}F_{\theta\theta}^{-1}F_{\theta\mu}\right)(\mu-\hat{\mu})^2 = C_{\mu\mu}^{-1}(\mu-\hat{\mu})^2 = \left(\frac{\mu-\hat{\mu}}{\sigma_\mu}\right)^2$$

$F_{\mu\mu} \neq C_{\mu\mu}^{-1}$!!

Proof of Wilks' theorem...



Uncertainty on μ:

- **From C:** σ_μ
- **From PLR:** σ_μ

Profiled θ **crosses ellipse at vertical tangents** by definition (L is lower at other points on the tangent)

Gaussian Profiling

$$\lambda(\mu,\theta;\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 + 2F_{\mu\theta}(\mu-\hat{\mu})(\theta-\hat{\theta}) + F_{\theta\theta}(\theta-\hat{\theta})^2$$

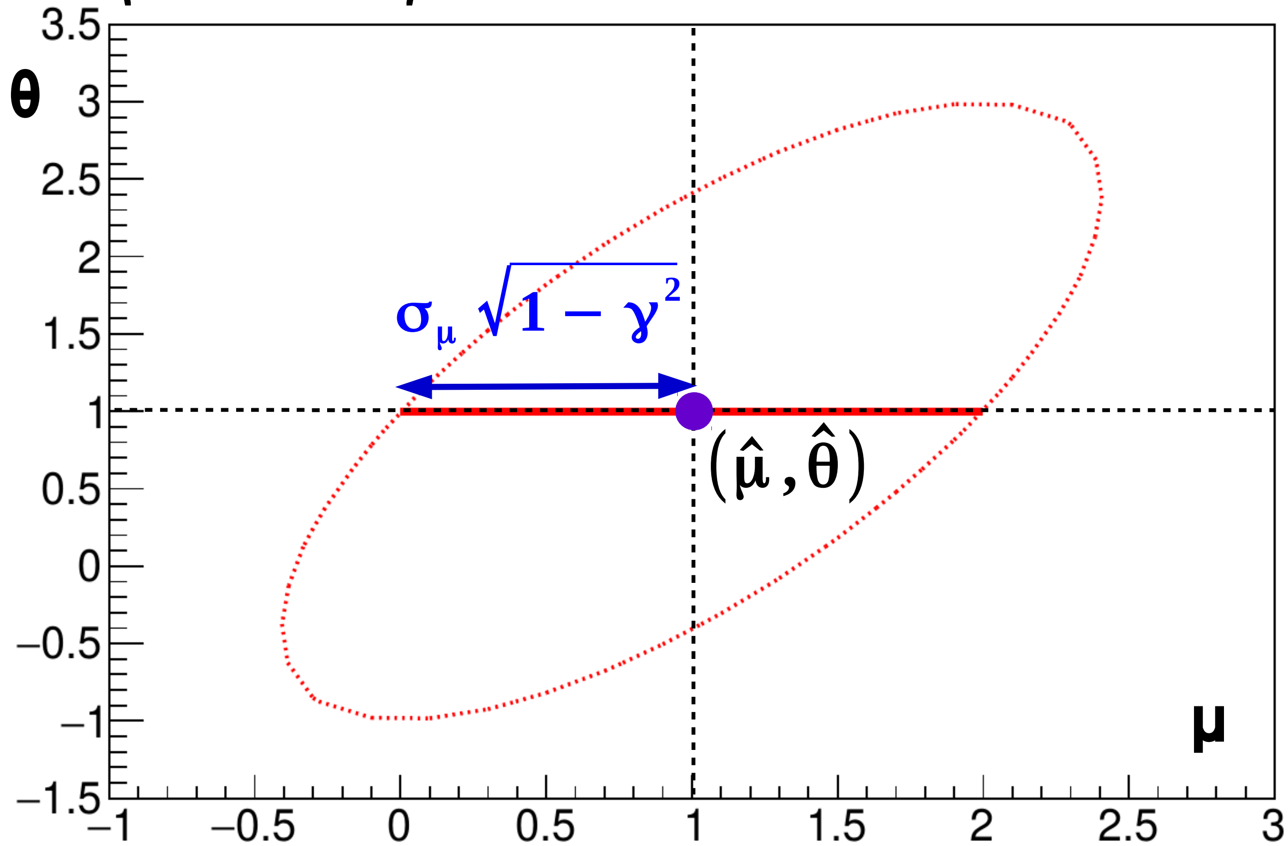
$$F \equiv C^{-1} = \frac{1}{1-\gamma^2}\begin{bmatrix} \frac{1}{\sigma_\mu^2} & \frac{\gamma}{\sigma_\mu\sigma_\theta} \\ \frac{\gamma}{\sigma_\mu\sigma_\theta} & \frac{1}{\sigma_\theta^2} \end{bmatrix}$$

→ For fixed $\theta = \hat{\theta}$, $\lambda(\mu)$ defines an interval:

$$\lambda(\mu,\theta=\hat{\theta};\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 = \left(\frac{\mu-\hat{\mu}}{\sigma_\mu\sqrt{1-\gamma^2}}\right)^2$$

Uncertainty on μ:

- **From C:** σ_μ
- **From PLR:** σ_μ
- **From λ(μ):** $\sigma_\mu\sqrt{1-\gamma^2}$



Gaussian Profiling

$$\lambda(\mu,\theta;\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 + 2F_{\mu\theta}(\mu-\hat{\mu})(\theta-\hat{\theta}) + F_{\theta\theta}(\theta-\hat{\theta})^2$$

$$F \equiv C^{-1} = \frac{1}{1-\gamma^2}\begin{bmatrix} \frac{1}{\sigma_\mu^2} & \frac{\gamma}{\sigma_\mu\sigma_\theta} \\ \frac{\gamma}{\sigma_\mu\sigma_\theta} & \frac{1}{\sigma_\theta^2} \end{bmatrix}$$

→ For fixed $\theta = \hat{\theta}$, $\lambda(\mu)$ defines an interval:

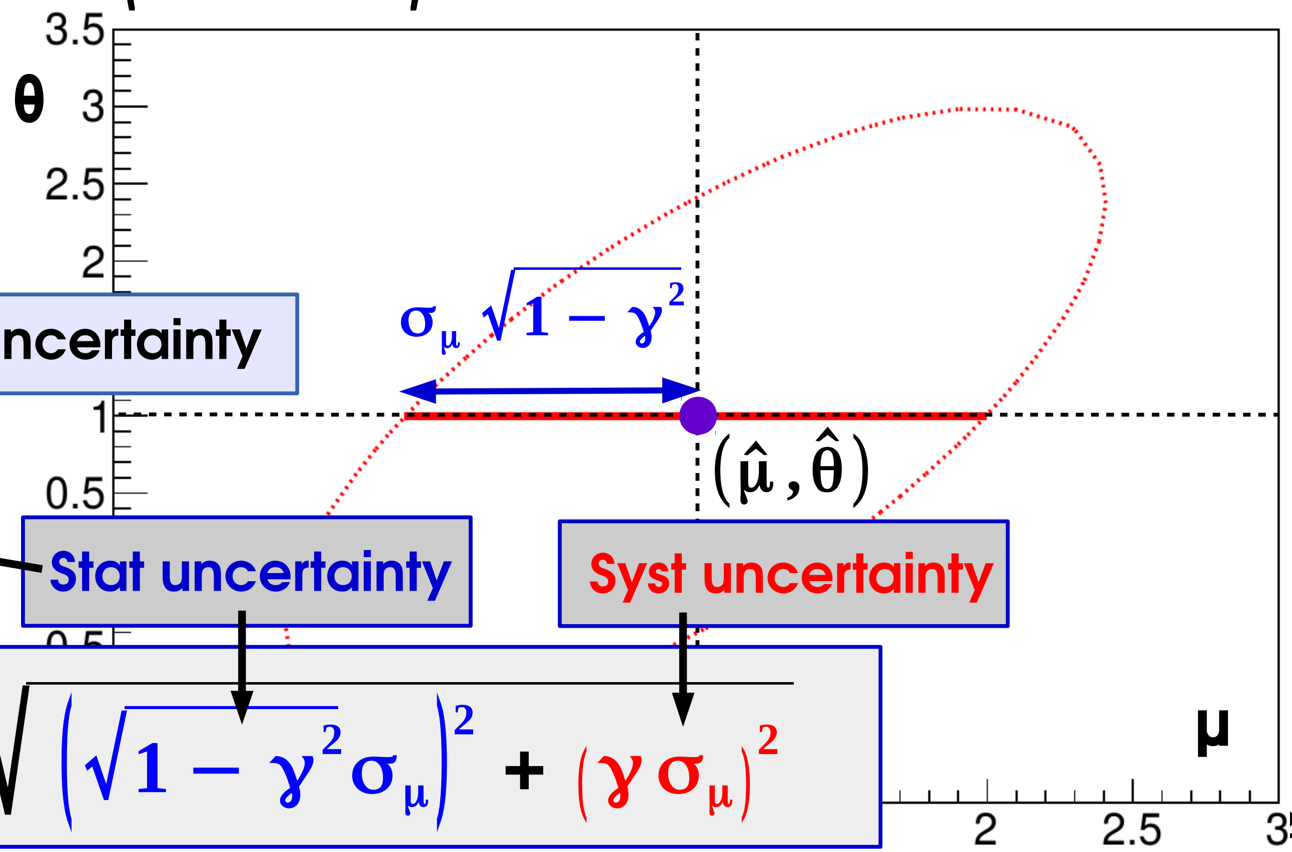
$$\lambda(\mu,\theta=\hat{\theta};\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 = \left(\frac{\mu-\hat{\mu}}{\sigma_\mu\sqrt{1-\gamma^2}}\right)^2$$

Uncertainty on μ:

- **From C:** σ_μ
- **From PLR:** σ_μ

Total uncertainty

- **From λ(μ):** $\sigma_\mu\sqrt{1-\gamma^2}$

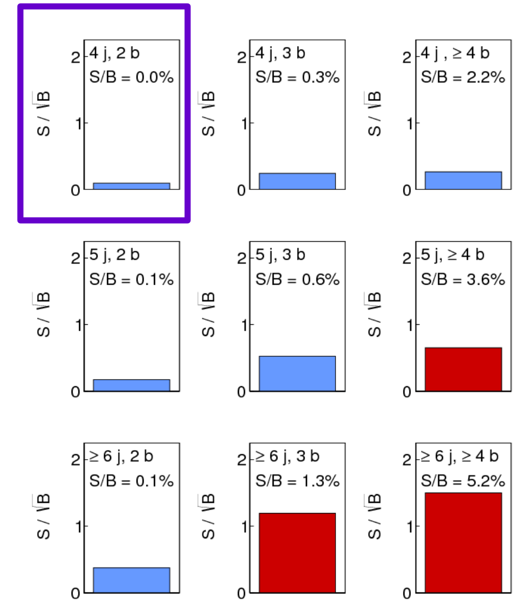


Stat uncertainty

Syst uncertainty

$$\sigma_\mu = \sqrt{\left(\sqrt{1-\gamma^2}\,\sigma_\mu\right)^2 + (\gamma\,\sigma_\mu)^2}$$

Profiling Example: ttH→bb

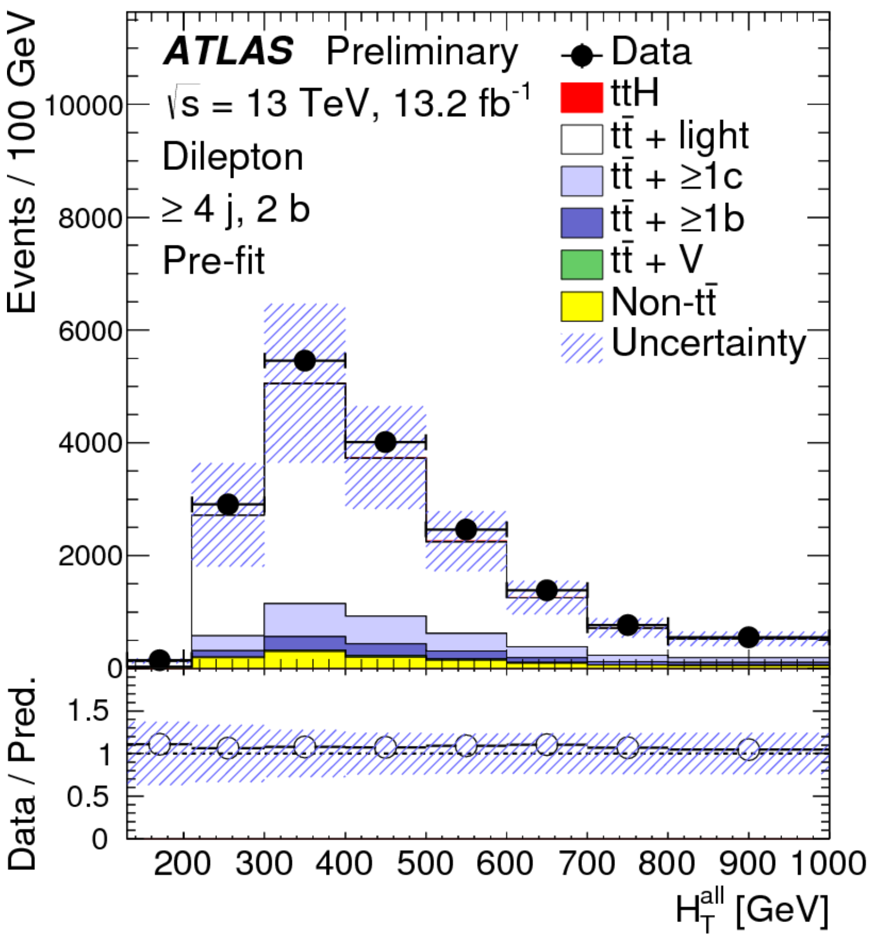


Analysis uses low-S/B categories to constrain backgrounds.

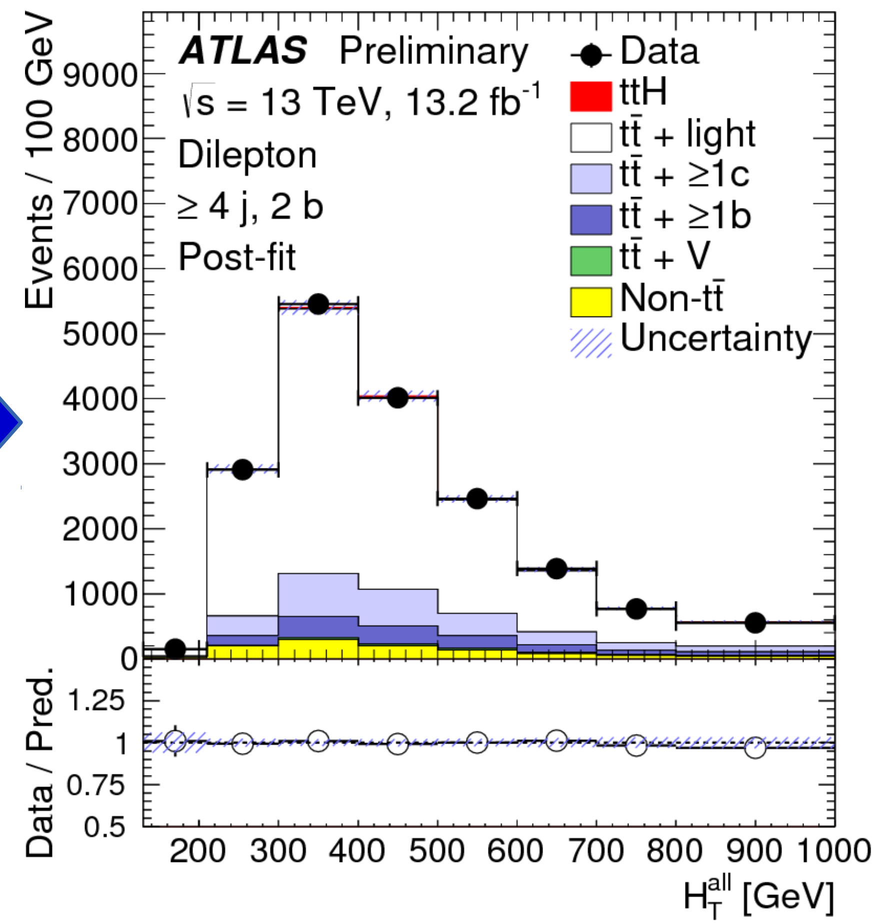
→ **Reduction in large uncertainties on tt bkg**

→ **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

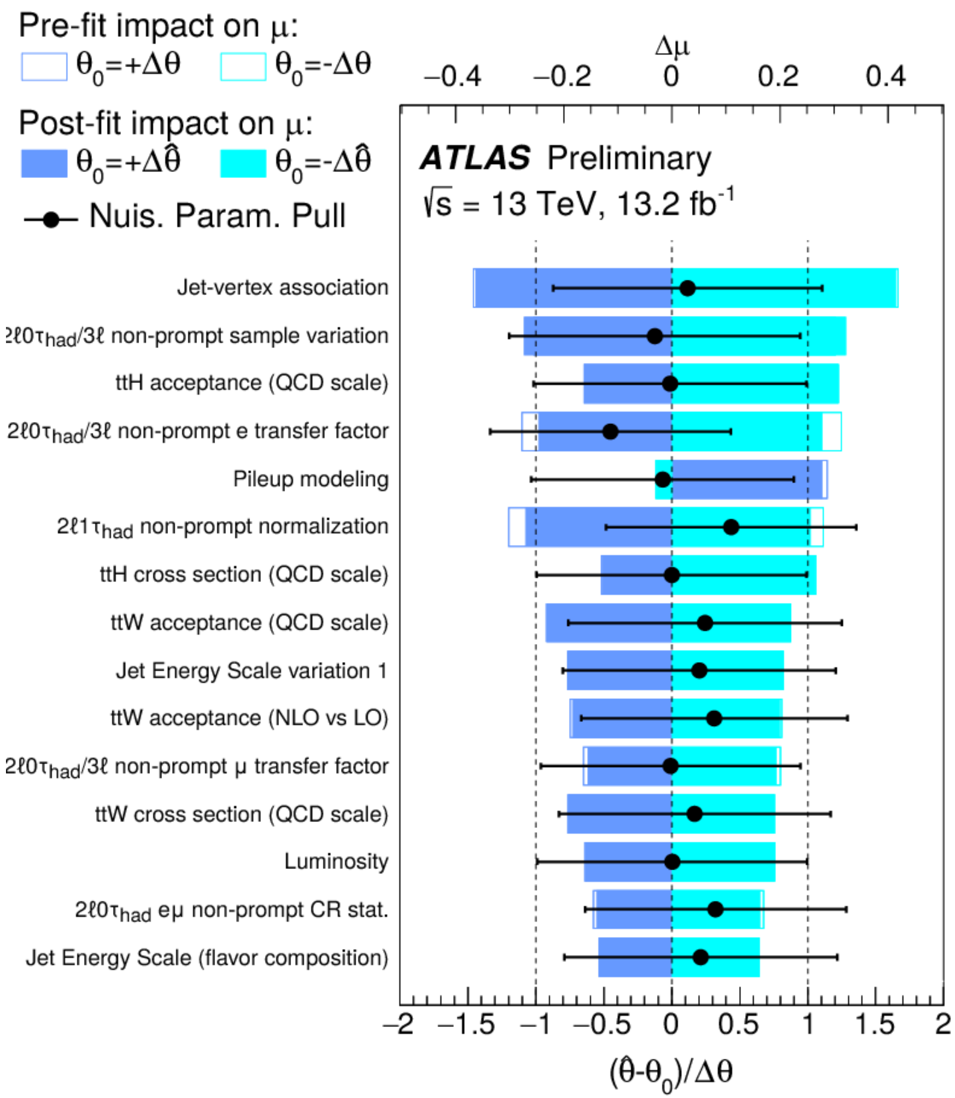
Systematics are described by NPs included in the fit. Nominally:

- **NP central value = 0** : corresponds to the pre-fit expectation (usually MC)
- **NP uncertainty = 1** : since NPs normalized to the value of the syst. :

$$N = N_0\left(1 + \sigma_{\text{syst}}\,\theta\right),\ \theta \sim G(0,1)$$

Fit results provide information on impact of the systematic on the result:

- **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 ±1σ shift of NP



Pull/Impact plots

13 TeV single-t XS (arXiv:1612.07231)

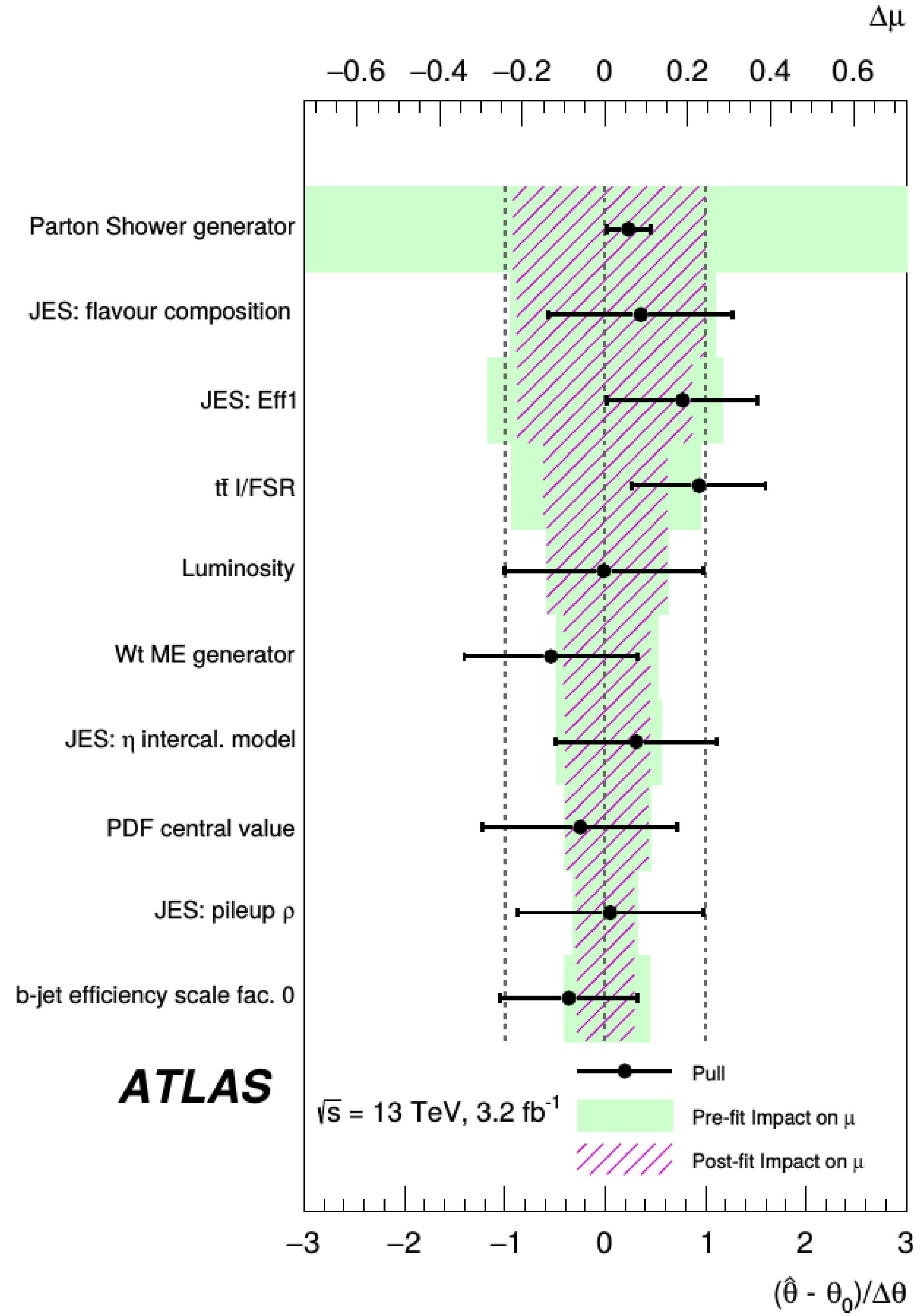
Systematics are described by NPs included in the fit. Nominally:

- **NP central value = 0** : corresponds to the pre-fit expectation (usually MC)
- **NP uncertainty = 1** : since NPs normalized to the value of the syst. :

$$N = N_0\left(1 + \sigma_{\text{syst}}\,\theta\right),\ \theta \sim G(0,1)$$

Fit results provide information on impact of the systematic on the result:

- **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 ±1σ shift of NP



Profiling Takeaways

Systematic = NP with an external constraint (auxiliary measurement).
→ No special treatment, treated like any other NP: statistical and systematic uncertainties represented in the same way.

When testing a hypothesis, use the best-fit values of the nuisance parameters: ***Profile Likelihood Ratio***.

$$\frac{L(\mu=\mu_0, \hat{\hat{\theta}}_{\mu_0})}{L(\hat{\mu}, \hat{\theta})}$$

Wilks' Theorem: the PLR has the same asymptotic properties as the LR without systematics: can profile out NPs and just deal with POIs.

Profiling systematics includes their effect into the total uncertainty. Gaussian:

$$\sigma_{total} = \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}$$

Guaranteed to work only as long as everything is Gaussian, but typically robust against non-Gaussian behavior.

Profiling can have unintended effects – need to carefully check behavior

Beyond Asymptotics: Toys

CMS-PAS-HIG-11-022

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

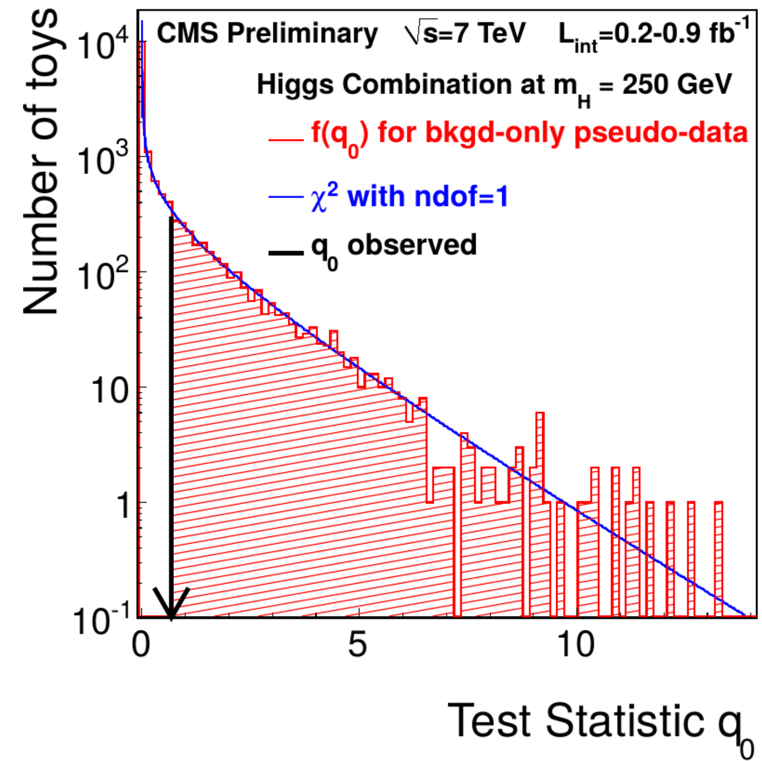
→ Also randomize the observable ($\boldsymbol{\theta}^{obs}$) of each auxiliary experiment: $G(\theta^{obs};\theta,\sigma_{syst})$

→ Samples the true distribution of the PLR

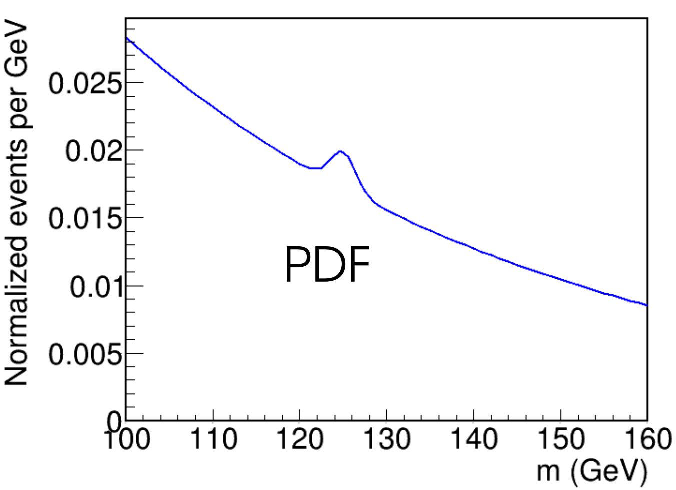
⇒ Integrate above observed PLR to get the p-value

→ Precision limited by number of generated toys,

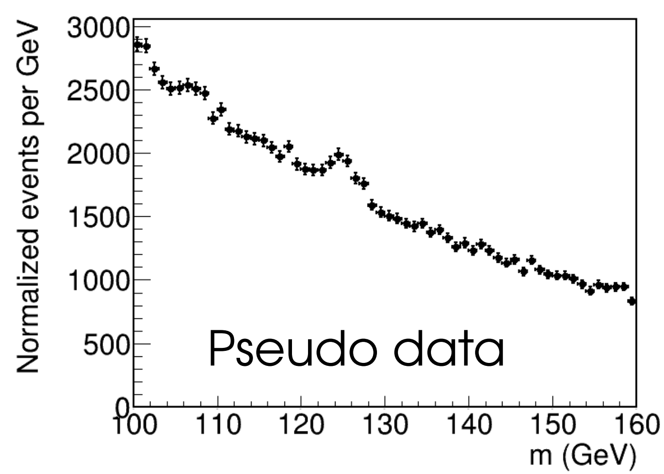
Small p-values (5σ : $p\sim10^{-7}$!) ⇒ **large toy samples**



Repeat N_{toys} times



p(data | x)



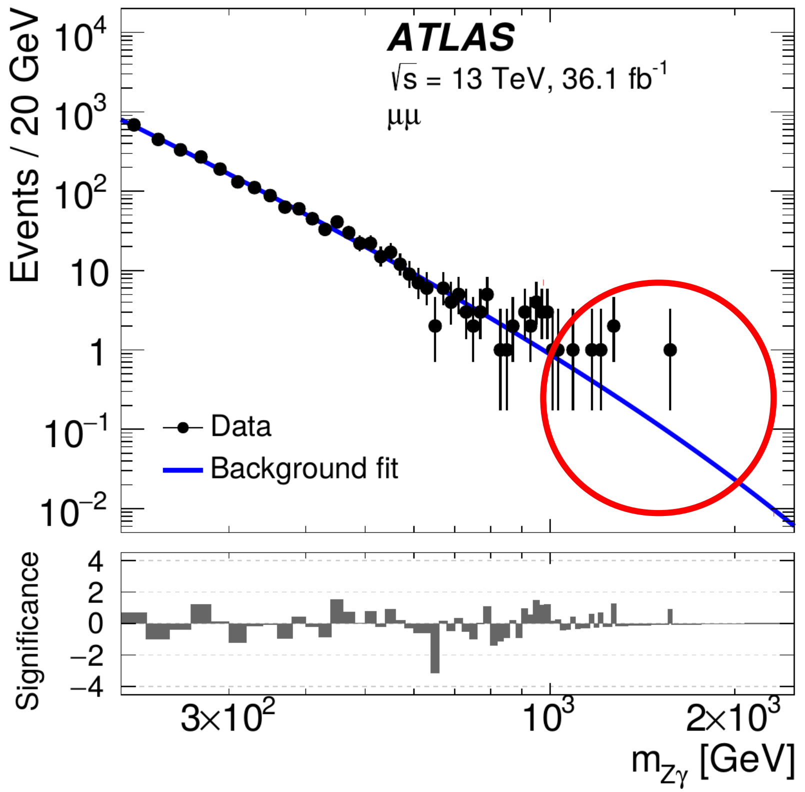
$\mathbf{q_0}$

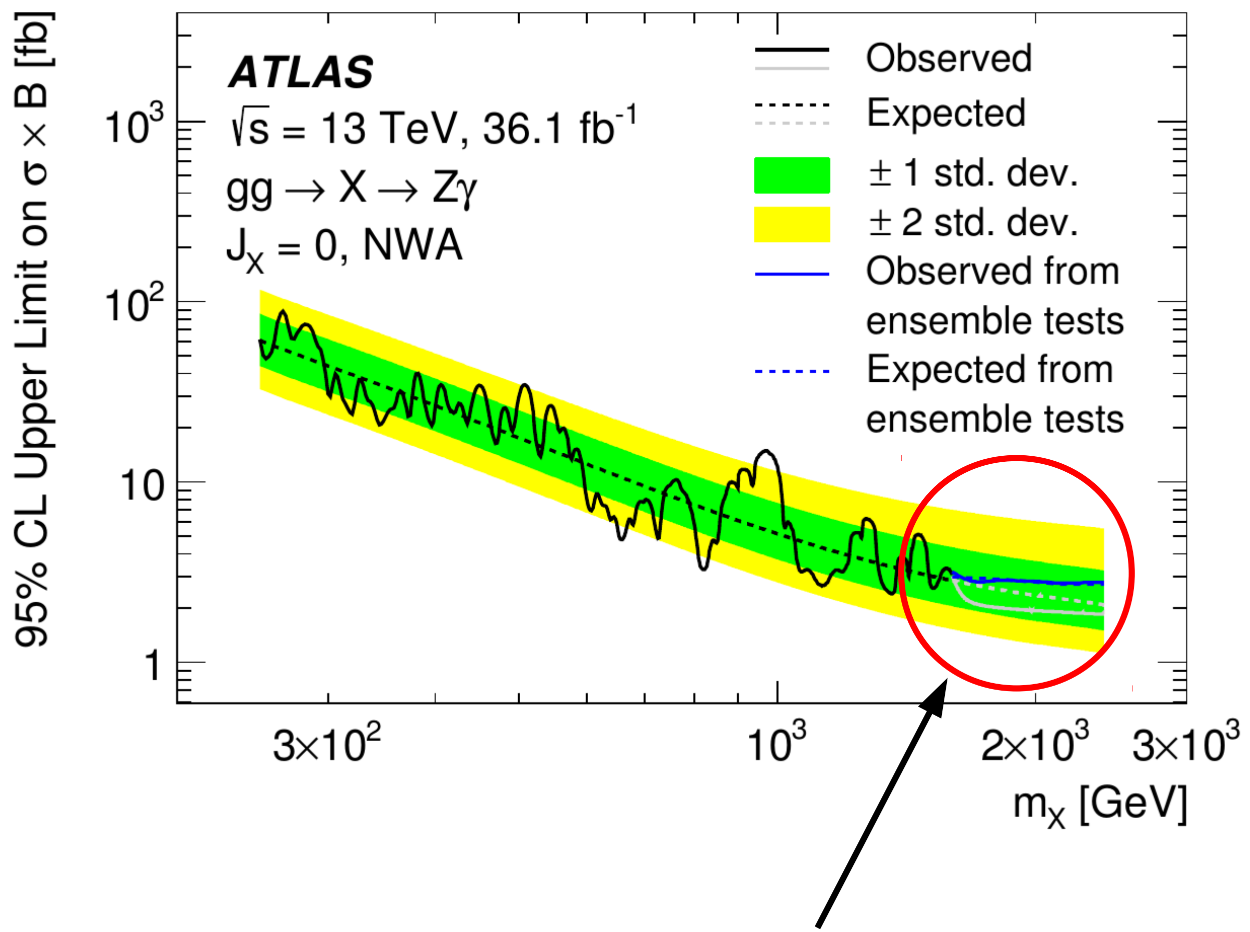
Toys: Example

arXiv:1708.00212

ATLAS $X \to Z\gamma$ 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**)

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

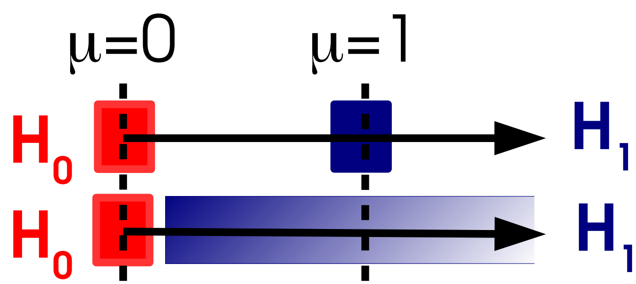
$$q_{LEP} = -2\log\frac{L(\mu=0,\tilde{\theta})}{L(\mu=1,\tilde{\theta})}$$

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

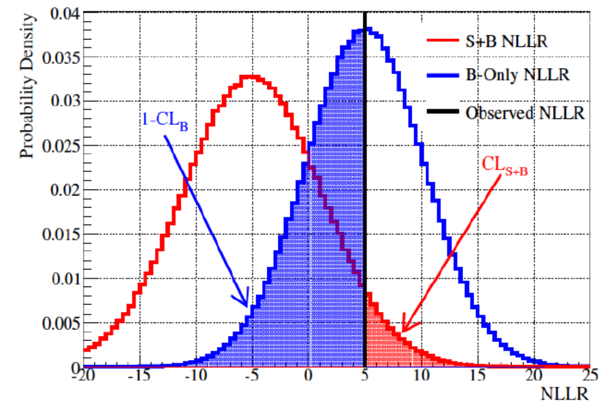
Both compare to **μ=1** instead of best-fit **$\hat{\mu}$**

LEP/Tevatron

LHC

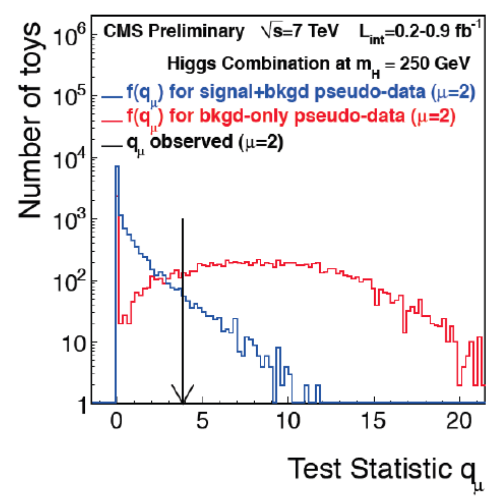


Example from Collie documentation (D0 note 5595)



→ Asymptotically:

- **LEP/Tevaton**: q linear in μ ⇒ **~Gaussian**
- **LHC**: q quadratic in μ ⇒ **~χ2**

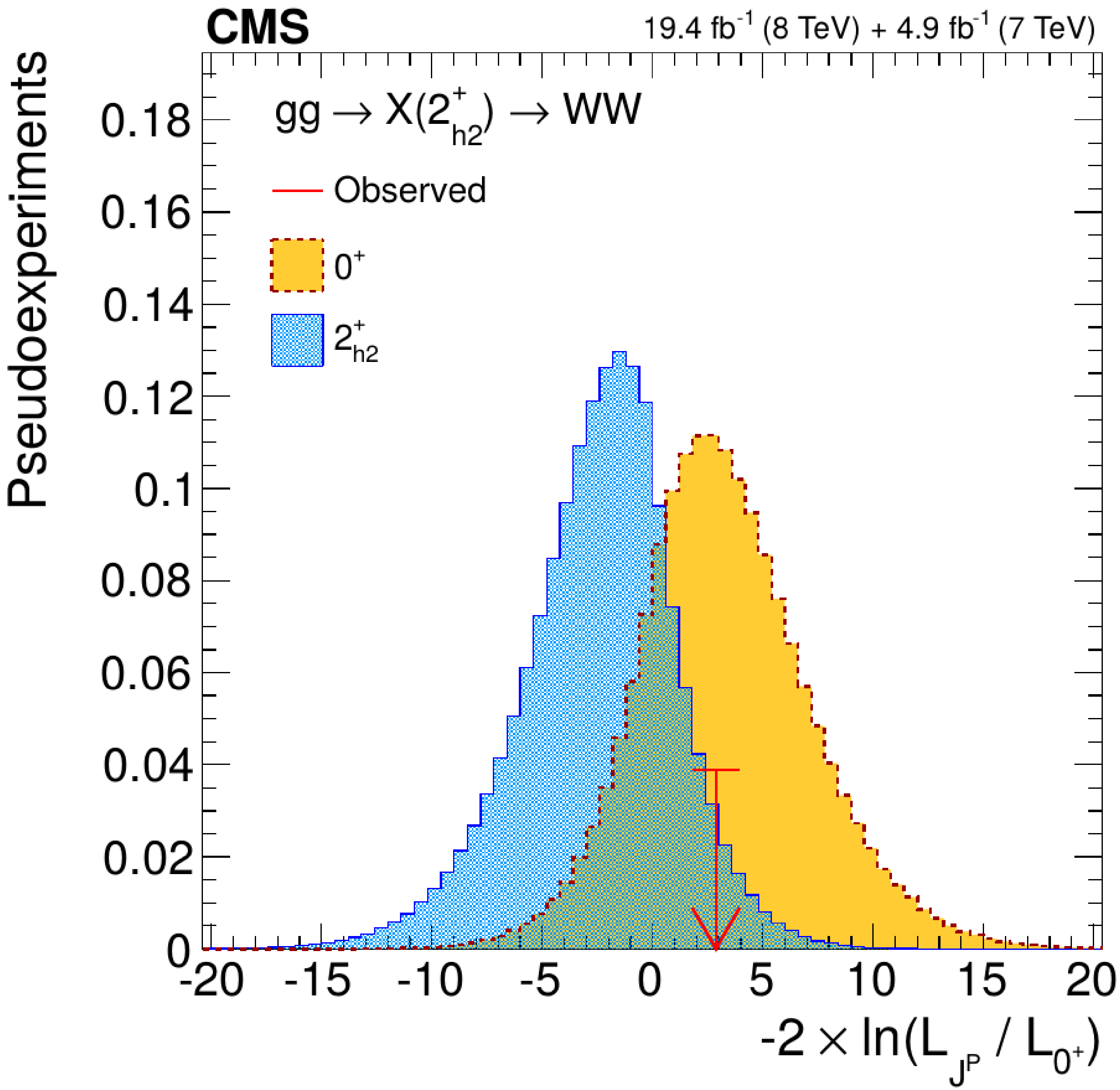


→ Still use TeVatron-style for discrete cases

Andrey Korytov, EPS 2011

Spin/Parity Measurements

Phys. Rev. D 92 (2015) 012004



Summary of Statistical Results Computation

Methods provide:

→ **Optimal use of information from the data under general hypotheses**

→ **Arbitrarily complex/realistic models (up to computing constraints...)**

→ **No Gaussian assumptions in the measurements**

Still often assume Gaussian behavior of PLR – but weaker assumption and can be lifted with toys

Systematics treated as auxiliary measurements – modeling can be tailored as needed

→ **Single PLR-based framework for all usual classes of measurements**

Discovery testing

Upper limits on signal yields

Parameter estimation

Outline

Profiling

Look-Elsewhere Effect

Bayesian methods

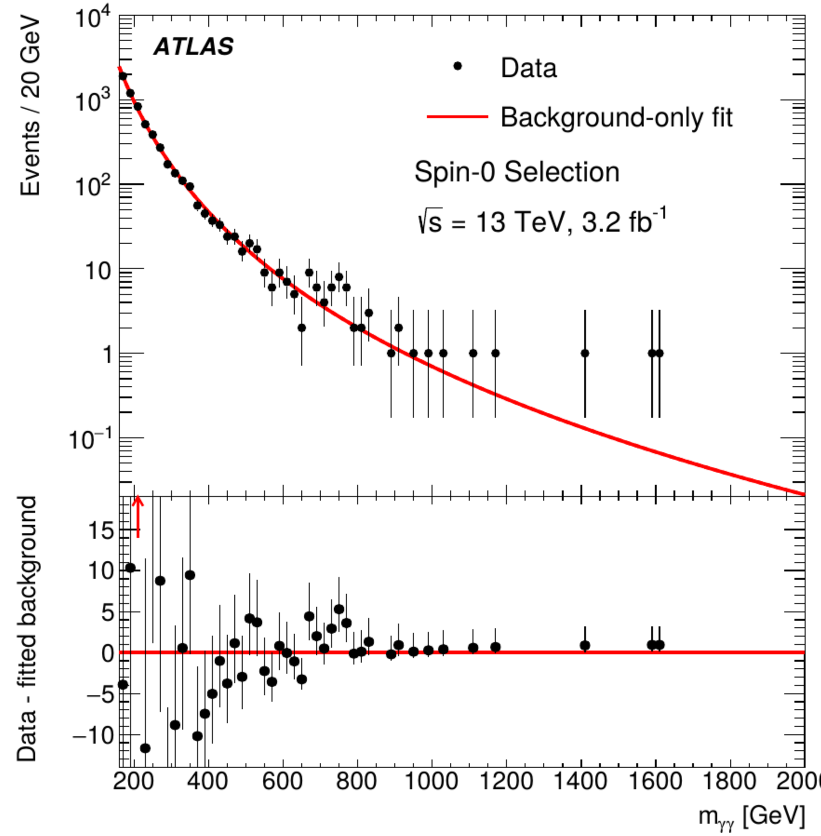
Statistical modeling in practice

- Building binned likelihoods
- Choosing PDFs in unbinned likelihoods
- Implementing systematics

BLUE

Look-Elsewhere Effect

Look-Elsewhere effect

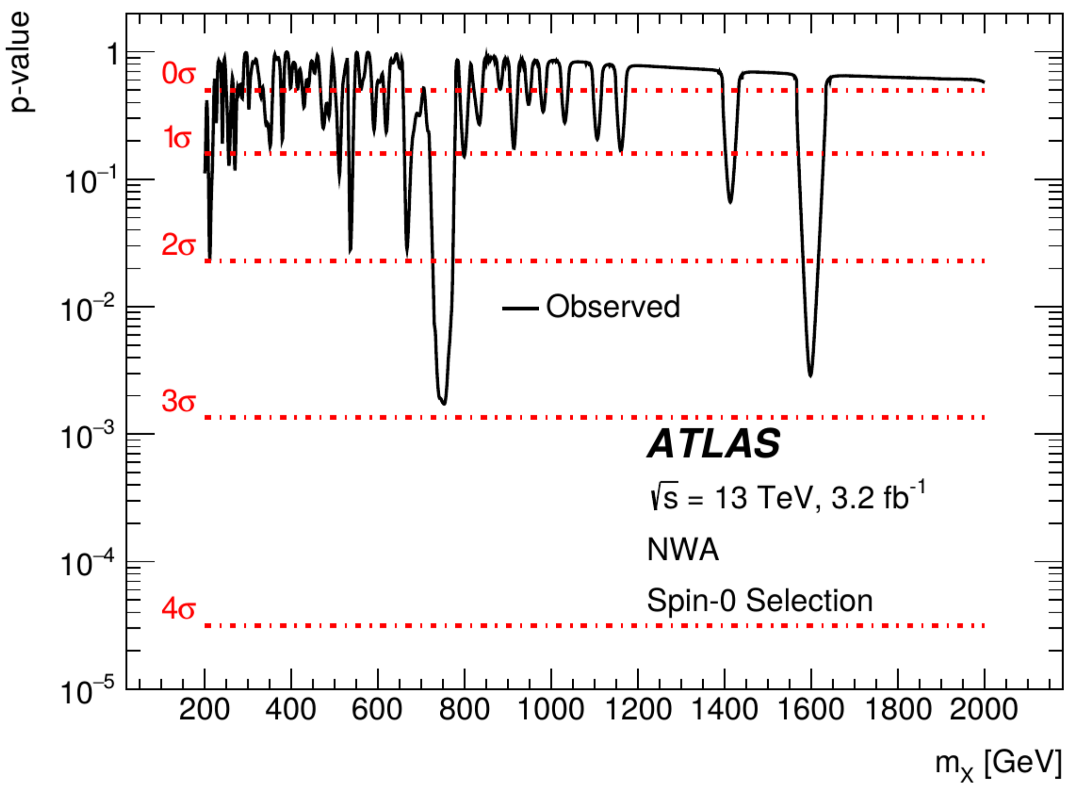


Sometimes, unknown parameters in signal model

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 **anywhere in the range**, rather than in a **predefined** location

⇒ ***Look-elsewhere effect*** (LEE)

Testing the same H_0, but against different alternatives

⇒ different p-values

Global Significance

Probability for a fluctuation ***anywhere*** in the range → **Global** p-value.
at a given location → **Local** p-value

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

Trials factor ↑ (N)

→ $\mathbf{p_{global} > p_{local}}$ ⇒ $\mathbf{Z_{global} < Z_{local}}$ – global fluctuation more likely ⇒ less significant

Trials factor : **naively** = # of independent intervals:
However this is usually **wrong** – more on this later

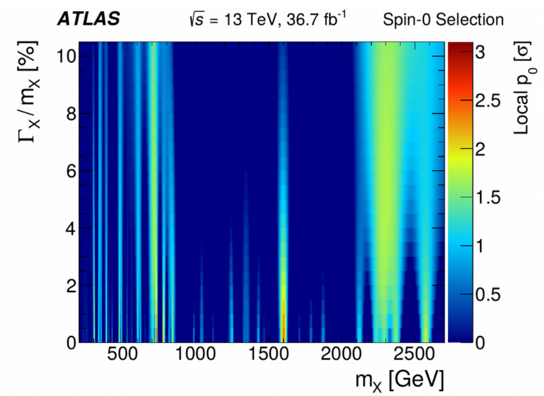
$$N_{\text{trials}} \overset{??}{=} N_{\text{indep}} = \frac{\text{scan range}}{\text{peak width}}$$

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

→ Depends on the scanned parameter ranges

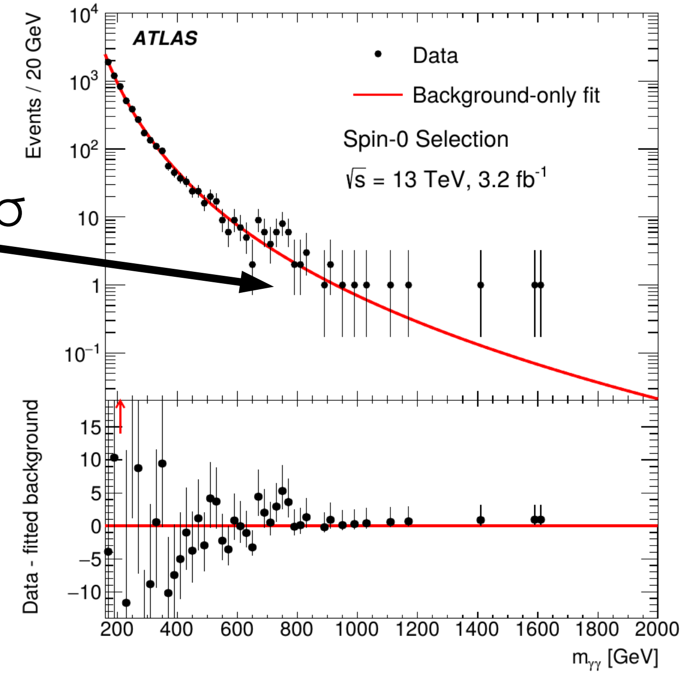
e.g. $X\to\gamma\gamma$: $200 < m_X < 2000$ GeV, $0 < \Gamma_X < 10\%\ m_X$.

→ However what comes out of the usual asymptotic formulas is p_{local}.



How to compute $\mathbf{p_{global}}$ **?** → **Toys** (brute force) or **asymptotic formulas.**

Global Significance from Toys



Local 3.9σ

Principle: repeat the analysis in toy data:

→ generate pseudo-dataset

→ perform the search, scanning over parameters as in the data

→ report the largest significance found

→ repeat many times

⇒ The frequency at which a given Z_0 is found ***is*** the global p-value

e.g. **X→γγ Search**: **Z_{local} = 3.9σ** (⇒ p_{local} ~ 5 10^{-5}), scanning 200 < m_X< 2000 GeV and 0 < Γ_X < 10% m_X

→ In toys, find such an excess 2% of the time

⇒ p_{global} ~ 2 10^{-2}, **Z_{global} = 2.1σ** Less exciting...

⊕ **Exact treatment**

⊖ **CPU-intensive** especially for large Z (need ~O(100)/p_{global} toys)

Global Significance from Asymptotics

Principle: approximate the global p-value in the asymptotic limit
→ reference paper: **Gross & Vitells, EPJ.C70:525-530,2010**

$$N_{indep} = \frac{\text{scan range}}{\text{peak width}}$$

Asymptotic trials factor (1 POI):

$$N_{trials} = 1 + \sqrt{\frac{\pi}{2}}\, N_{indep}\, Z_{local}$$

→ Trials factor is **not just N_{indep},**
also depends on Z_{local} !

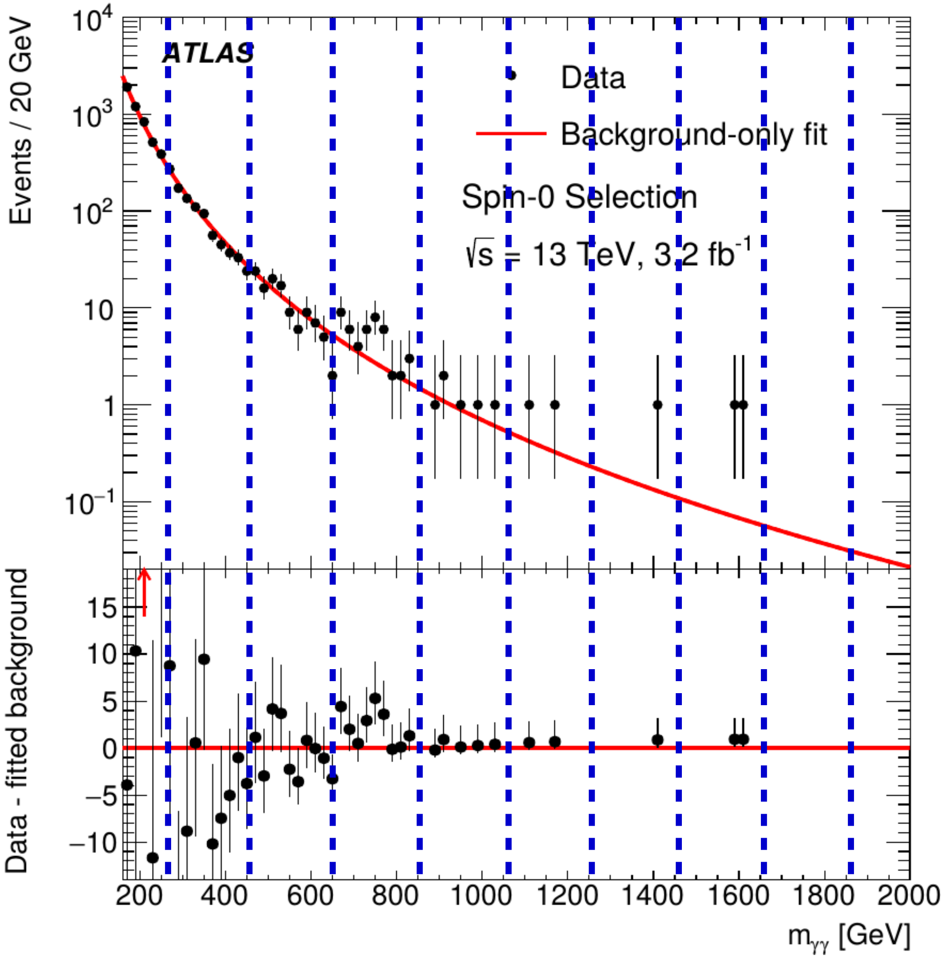
Why ?

→ slice scan range into N_{indep} regions
of size ~ peak width
→ search for a peak in each region

⇒ Indeed gives $N_{trials}=N_{indep}$.

However this misses peaks sitting on
edges between regions

⇒ true N_{trials} is **> N_{indep}**!



Global Significance from Asymptotics

Principle: approximate the global p-value in the asymptotic limit

→ reference paper: **Gross & Vitells, EPJ.C70:525-530,2010**

$$N_{indep} = \frac{\text{scan range}}{\text{peak width}}$$

Asymptotic trials factor (1 POI):

$$N_{trials} = 1 + \sqrt{\frac{\pi}{2}}\, N_{indep}\, Z_{local}$$

→ Trials factor is **not just N_{indep},** also depends on Z_{local} !

Why ?

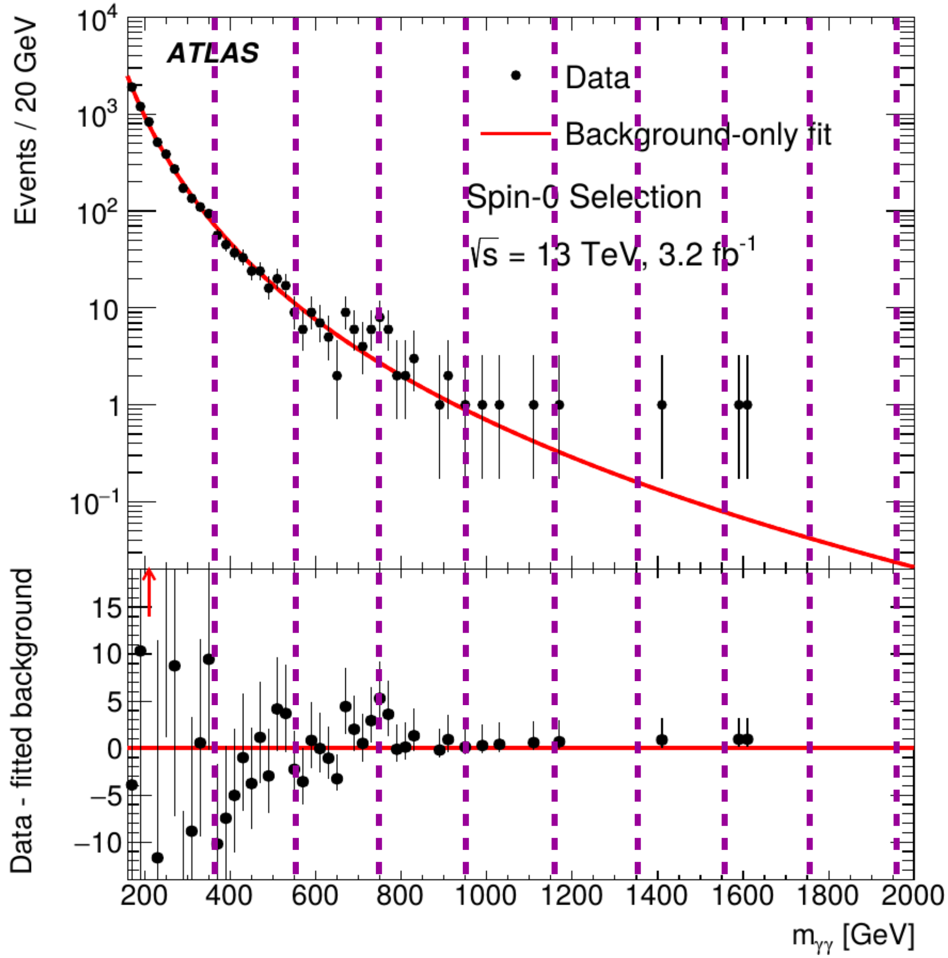
→ slice scan range into N_{indep} regions of size ~ peak width

→ search for a peak in each region

⇒ Indeed gives $N_{trials}=N_{indep}$.

However this misses peaks sitting on **edges between regions**

⇒ true N_{trials} is **>** N_{indep}!



Illustrative Example

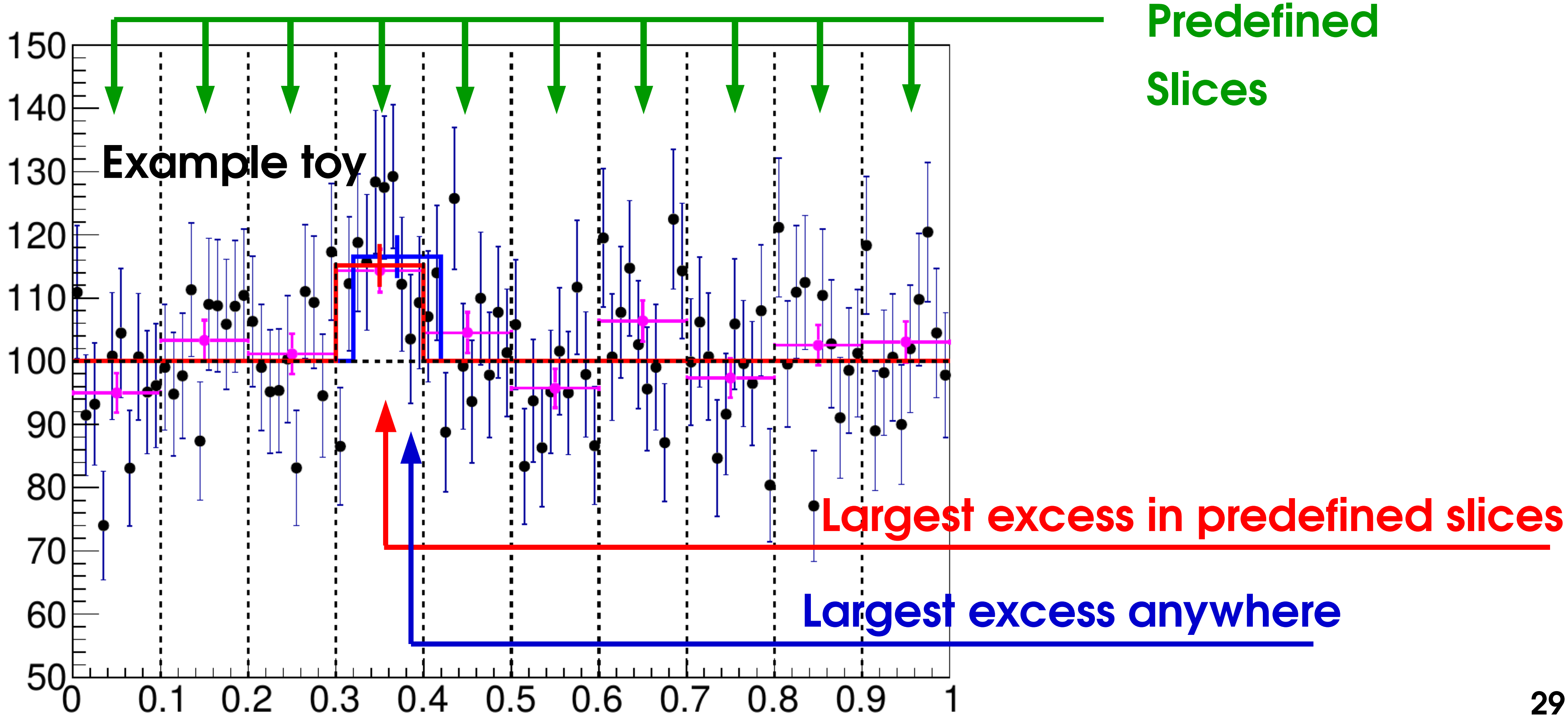
Test on a simple example: generate toys with

→ flat background (100 events/bin)

→ count events in a fixed-size sliding window, look for excesses

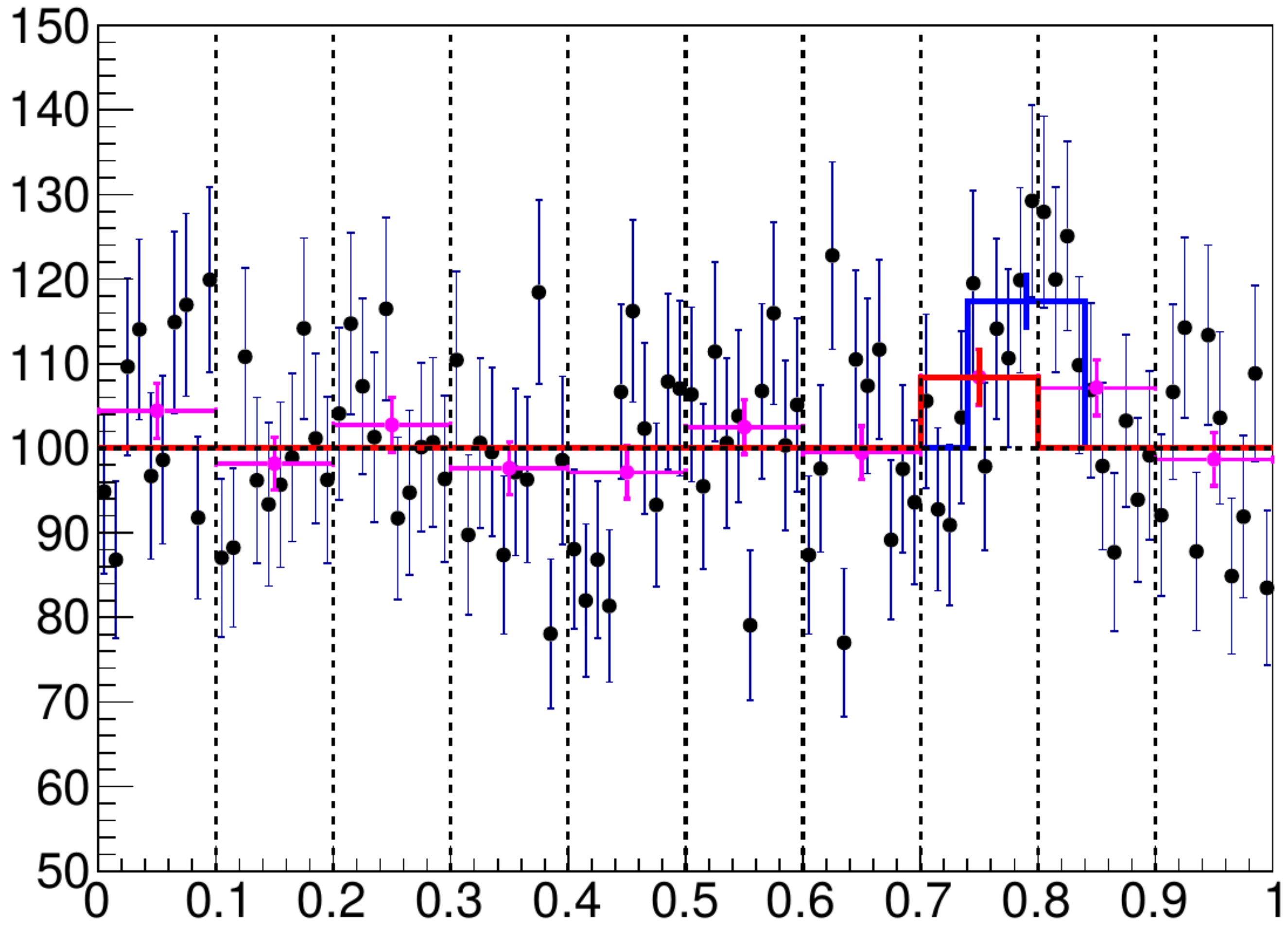
Two configurations:

1. Look only in 10 slices of the full spectrum
2. Look in any window of same size as above, anywhere in the spectrum



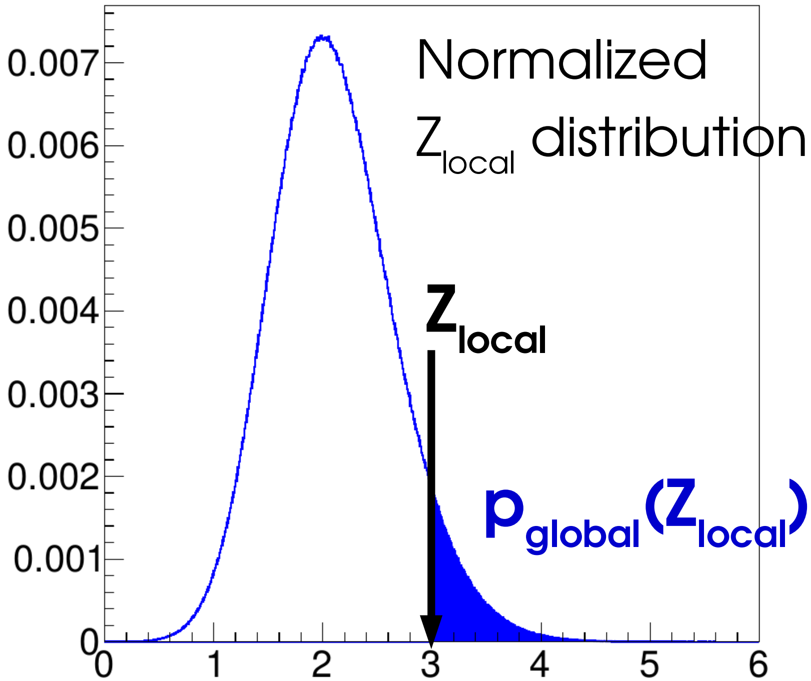
Illustrative Example (2)

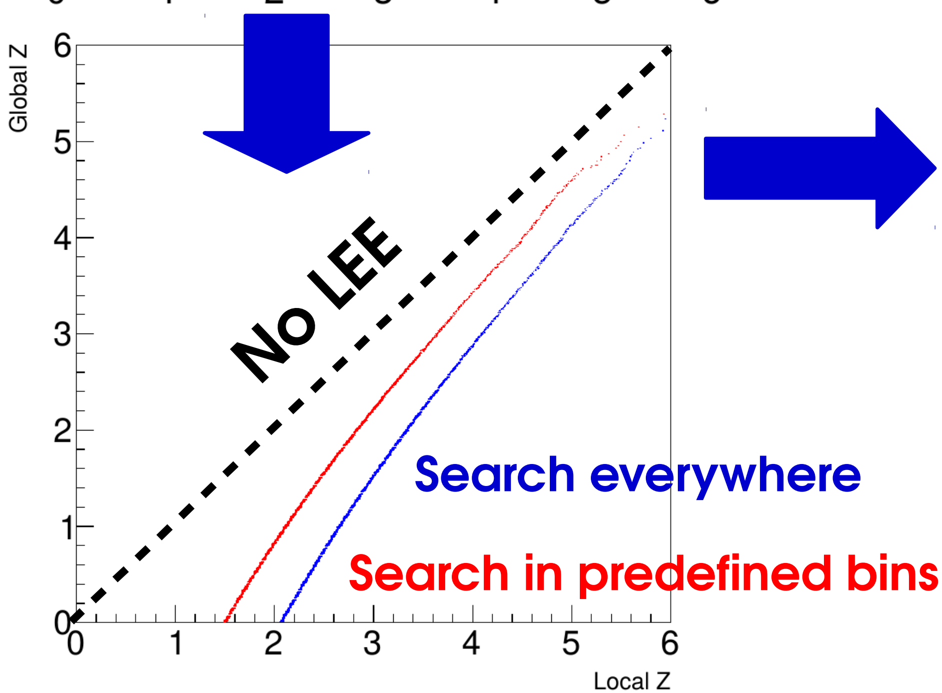
Very different results if the excess is **near a boundary :**

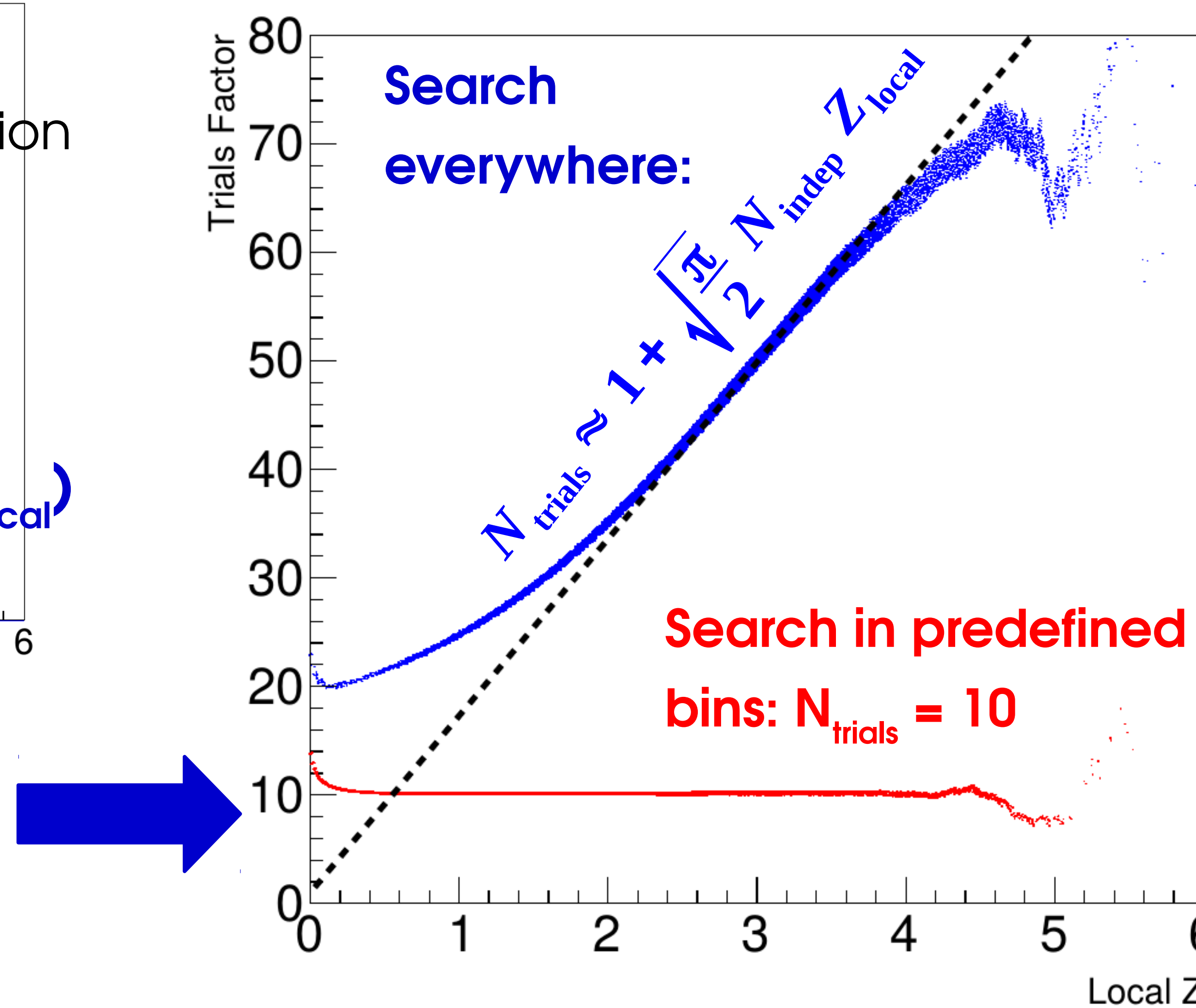


1. Look only in 10 slices of the full spectrum
2. Look in any window of same size as above, anywhere in the spectrum

Illustrative Example (3)







Searching everywhere gives the extra Z_{local} dependence

Z_{Global} Asymptotics Extrapolation

Asymptotic trials factor (1 POI): $$N_{trials} = 1 + \sqrt{\frac{\pi}{2}}\, N_{indep}\, Z_{local}$$

How to get $\mathbf{N_{indep}}$? Usually work with a slightly different formula:

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

Number of excesses with $Z > Z_{test}$

⇒ calibrate for small Z_{test}, apply result to higher Z_{local}.

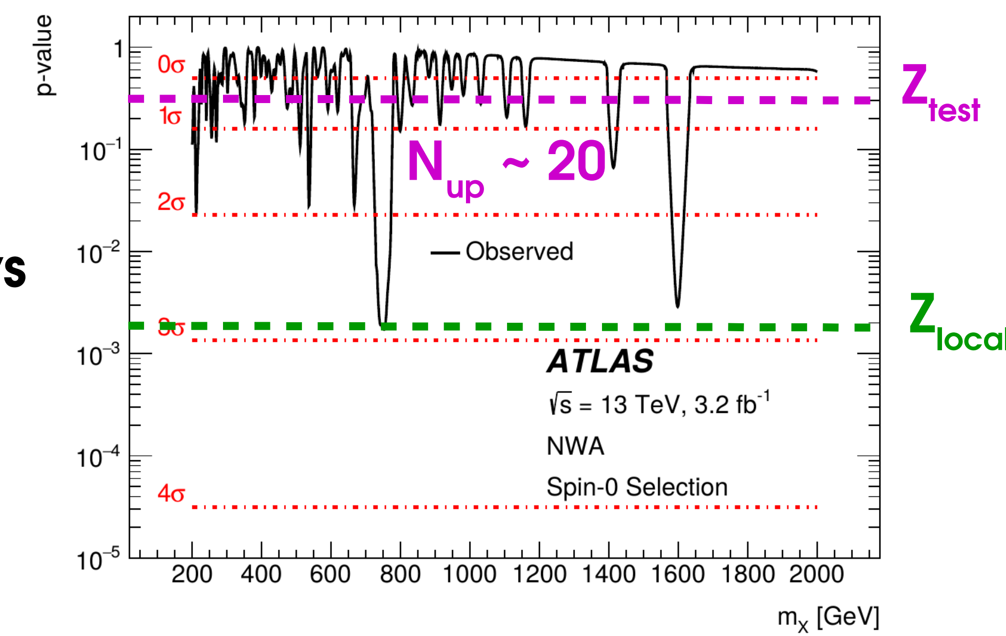
Can choose arbitrarily small Z_{test}

⇒ many excesses

⇒ can measure N_{up} in data (1 "toy")

Can also measure $\langle N_{up} \rangle$ in multiple toys

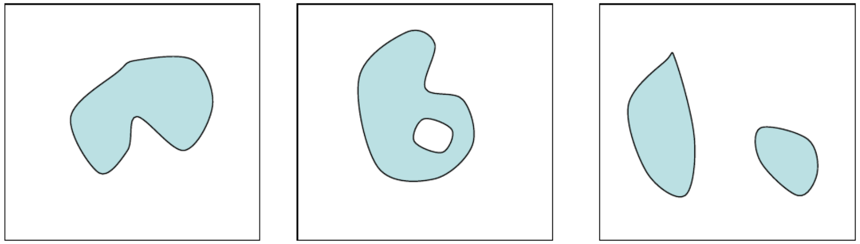
if large stat uncertainty from too few excesses



In 2D

O. Vitells and E. Gross, Astropart. Phys. 35 (2011) 230

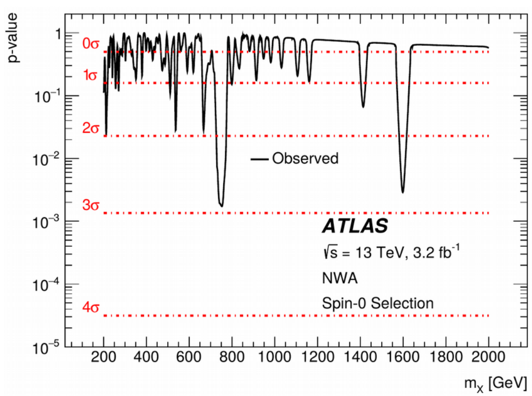
Generalization to 2D scans: consider sections at a fixed Z_{test}, compute its ***Euler characteristic*** φ, and use

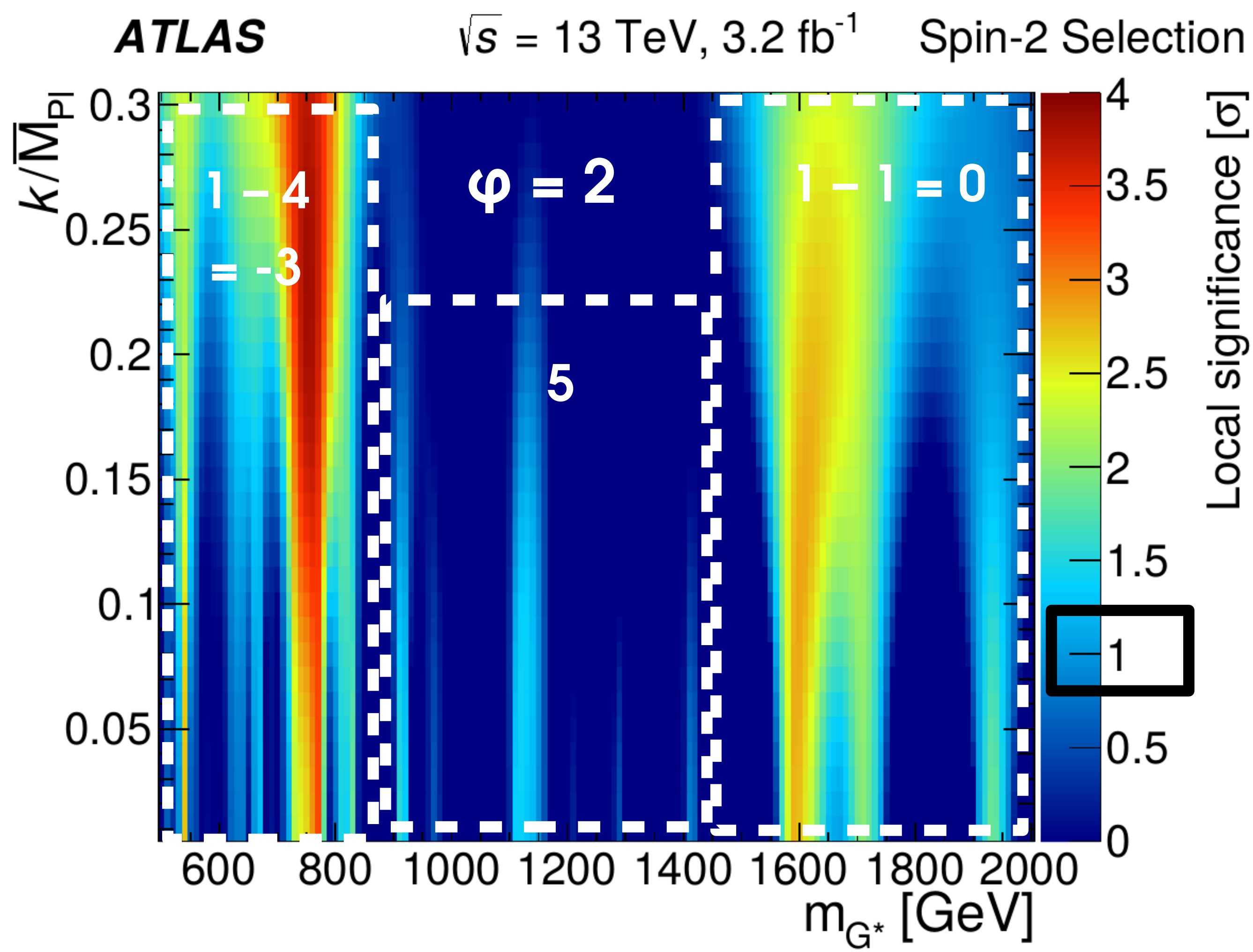


$\phi=1$ $\phi=0$ $\phi=2$

$$p_{\mathrm{global}} \approx E[\phi(A_u)] = p_{\mathrm{local}} + e^{-u/2}(N_1 + \sqrt{u}N_2)$$

→ Generalizes 1D bump counting





Now need to determine 2 constants N_1 and N_2, from Euler φ measurements at 2 different Z_{test} values.

Outline

Profiling

Look-Elsewhere Effect

Bayesian methods

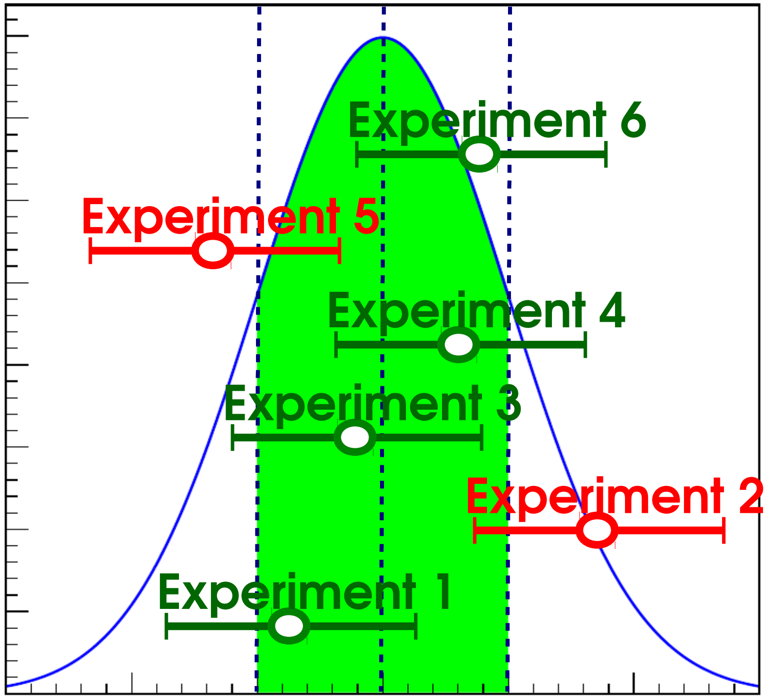
Statistical modeling in practice

- Building binned likelihoods
- Choosing PDFs in unbinned likelihoods
- Implementing systematics

BLUE

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 **fixed but unknown**

- Probabilities apply to measurements:

→ **"m_H = 125.09 ± 0.24 GeV"** :

→ i.e. $[125.09 - 0.24 \;;\; 125.09 + 0.24\;]$ GeV has p=68% 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 we would really like to know is

What is the probability that the excess we see is a fluctuation

→ we want **P(no Higgs | data)** – but all we have is **P(data | no Higgs)**

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. Frequentist 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 systematics NPs**, P(θ)
not auxiliary measurements $P(\theta^{mes}; \theta)$

⊕ Simply integrate them out, no need for profiling: $P(\mu) = \int P(\mu, \theta)\, d\theta$

→ Use probability distribution P(μ) directly for limits, credibility intervals

e.g. define 68% CL ("Credibility Level") interval (A, B) by: $\int_A^B P(\mu)\, d\mu = 68\,\%$

⊖ No simple way to test for discovery

⊖ Integration over NPs can be CPU-intensive

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 | μ) in a frequentist way

  → "Bayesian NPs, frequentist POIs"

- Some use in Run 1, now phased out in favor of frequentist PLR.

Bayesian methods and CL_s: CL_s computation

Gaussian counting with systematic on background: $n = S + B + \sigma_{syst}\theta$

$$L(n;S,\theta) = G(n;S+B+\sigma_{syst}\theta,\sigma_{stat})\ G(\theta_{obs}=0;\theta,1)$$

$$\left.\begin{array}{l}\textbf{MLE: } \hat{S} = n-B \\[2em] \textbf{Conditional MLE: } \hat{\hat{\theta}}(\mu) = \dfrac{\sigma_{syst}}{\sigma_{stat}^2+\sigma_{syst}^2}(n-S-B)\end{array}\right\} \quad \textbf{PLR : } \lambda(\mu) = \left(\frac{S+B-n}{\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}}\right)^2$$

Gaussian ⇒ from previous studies, CL_s limit is

$$\mathbf{CL_s:}\quad S_{up}^{CL_s} = n-B+\left[\Phi^{-1}\left(1-0.05\,\Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}}\right)\right)\right]\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}$$

Bayesian methods and CL$_s$: Bayesian case

Gaussian counting with systematic on background: $n = S + B + \sigma_{syst}\theta$

$$P(n|S,\theta) = G(n; S+B+\sigma_{syst}\theta, \sigma_{stat})\ G(\theta|0,1)$$

Bayesian: G(θ) is actually a ***prior*** on θ ⇒ perform integral (***marginalization***)

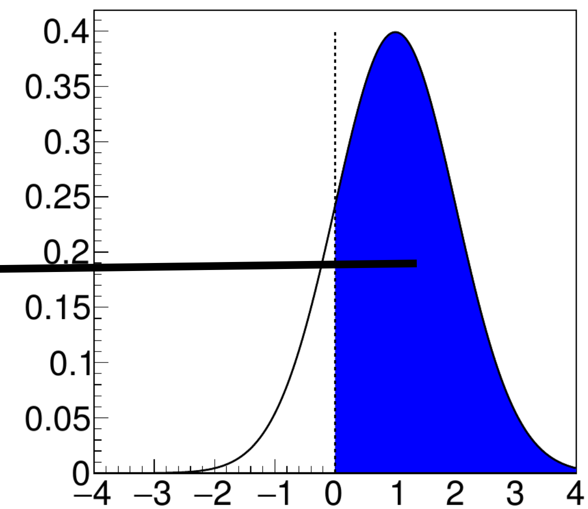
$$P(n|S) = G\left(S; n-B, \sqrt{\sigma_{stat}^2+\sigma_{syst}^2}\right)$$

same effect as profiling!

Need P(S | n) ⇒ a prior for S – take flat PDF over S > 0

⇒ Truncate Gaussian at S=0: $P(S|n) = P(n|S)\ P(S)$

$$P(S|n) = G\left(S; n-B, \sqrt{\sigma_{stat}^2+\sigma_{syst}^2}\right)\left[\Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}}\right)\right]^{-1}$$



Bayesian Limit:

$$\int_{S_{up}}^{\infty} P(S|n)\,dS = 5\,\% = \left[1-\Phi\left(\frac{S_{up}-(n-B)}{\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}}\right)\right]\left[\Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}}\right)\right]^{-1}$$

$$S_{up}^{Bayes} = n-B+\left[\Phi^{-1}\left(1-0.05\,\Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}}\right)\right)\right]\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}$$

same result as CL$_s$!

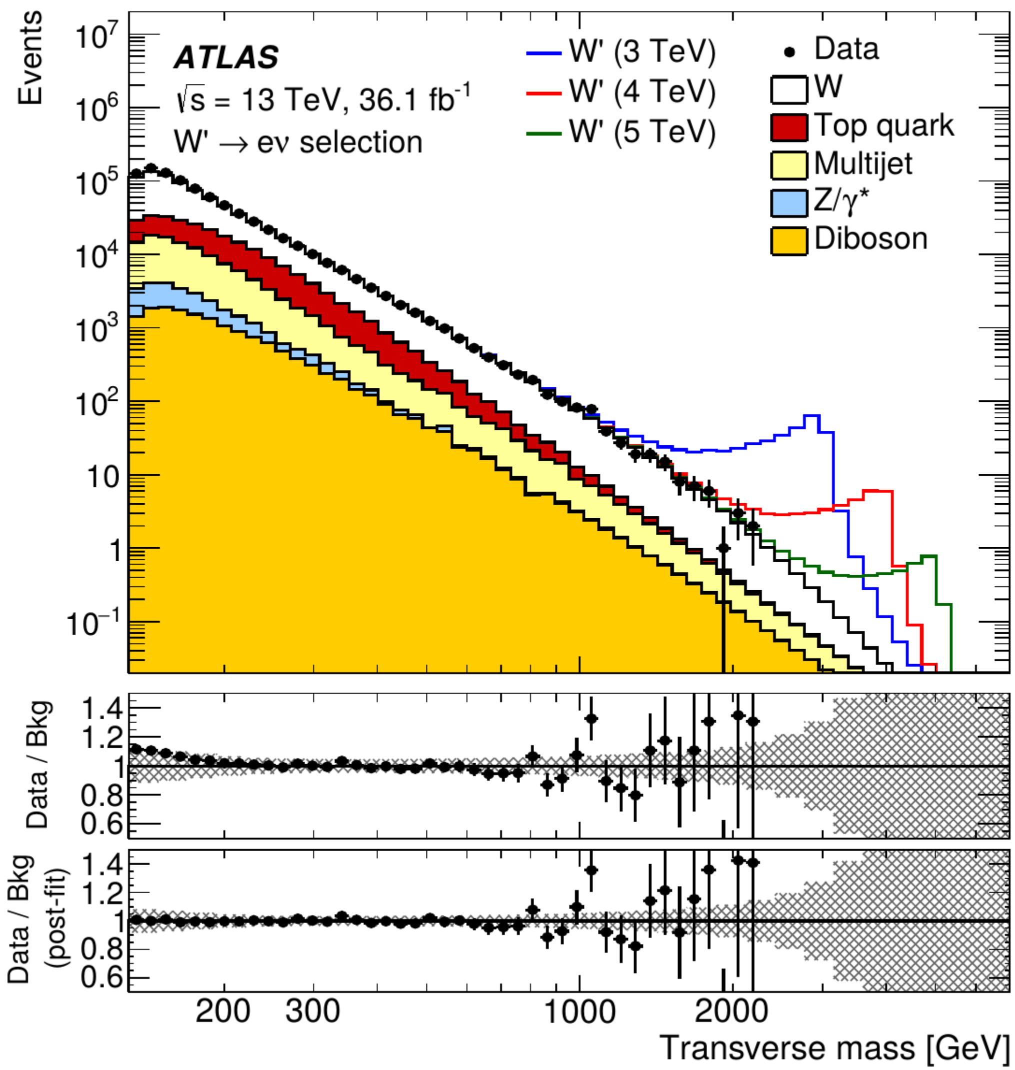
Example: W'→lν Search

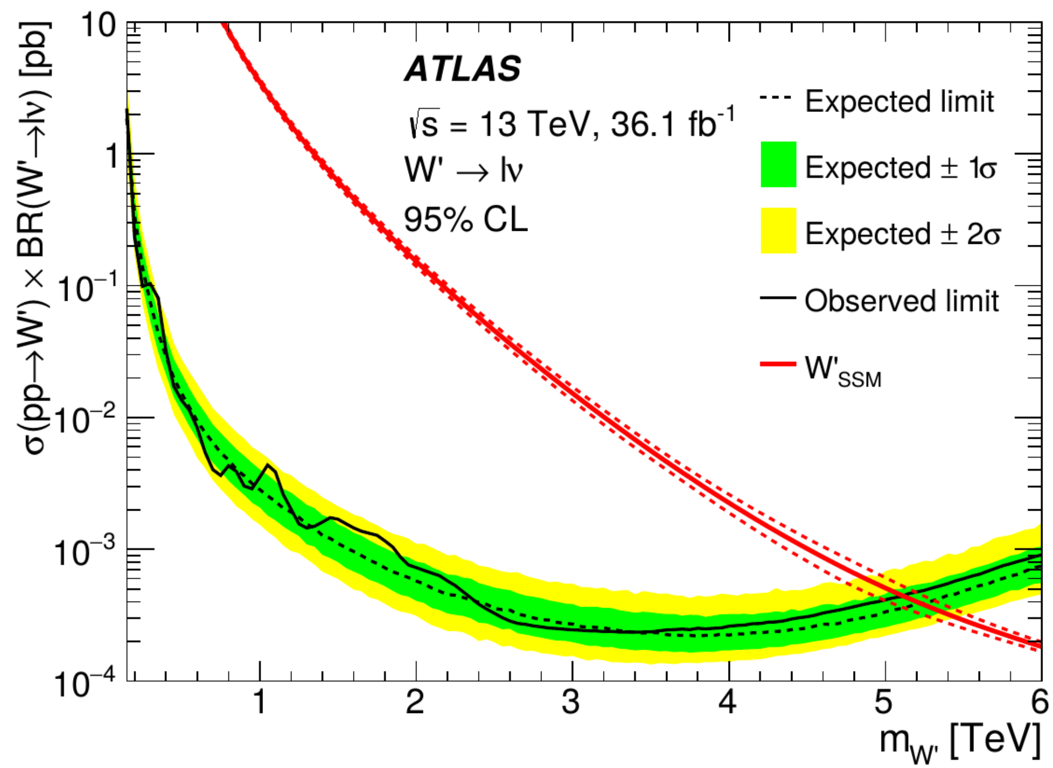
arXiv:1706.04786

- **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
Lepton reconstruction and identification
Lepton momentum scale and resolution
$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and scale
Jet energy resolution
Pile-up

Multijet background
Top extrapolation
Diboson extrapolation
PDF choice for DY
PDF variation for DY
EW corrections for DY
Luminosity





Why 5σ ?

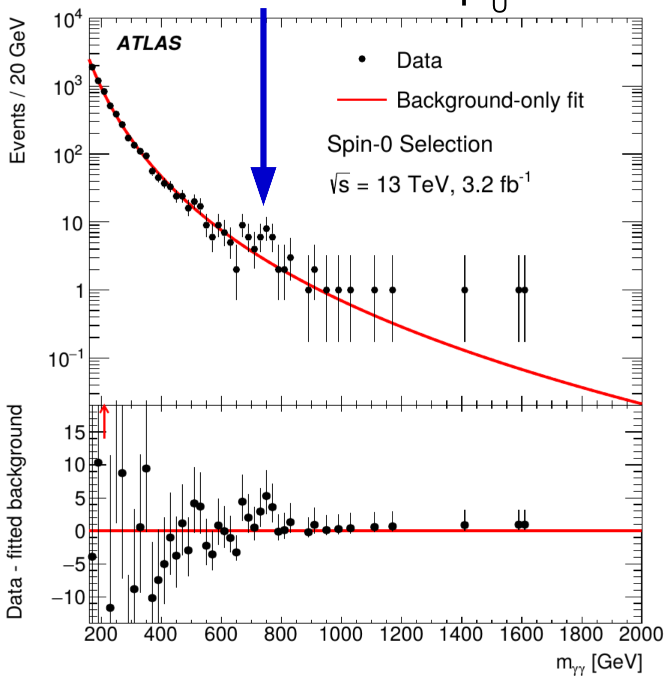
One-sided discovery: **$5\sigma \Leftrightarrow p_0 = 3\ 10\text{-}7 \Leftrightarrow$ 1 chance in 3.5M**

→ Overly conservative ?

→ Do we even know the sampling distributions so far out ?

Local 3.9σ, p_0 = 5E-5

Global 2.1σ, p_0 = 2E-2



Reasons for sticking with 5σ (from Louis Lyons):

- **LEE** : searches typically cover multiple independent regions
  ⇒ Global p-value is the relevant one

  N_{trials} ~ 1000 : local 5σ ⇔ $O(10^{-4})$ more reasonable
- **Mismodeled systematics**: factor 2 error in syst-dominated analysis ⇒ factor 2 error on Z…
- **History**: 3σ and 4σ excesses do occur regularly, for the reasons above
- **"Subconscious Bayes Factor"** : p-value should be at least as small as the subjective p(S):

$$P(\text{fluct}) = \frac{P(\text{fluct}|B)\,P(B)}{P(\text{fluct}|S)\,P(S) + P(\text{fluct}|B)\,P(B)}$$

Extraordinary claims require extraodinary evidence

⇒ Stay with 5σ...

Outline

Profiling

Look-Elsewhere Effect

Bayesian methods

Statistical modeling in practice

- Building binned likelihoods
- Choosing PDFs in unbinned likelihoods
- Implementing systematics

BLUE

Statistical Modeling: in Practice

Bulding statistical models

So far focus has been on concepts, but building a statistical model also requires **numerical** inputs:

- **Data PDFs** for all model components
- **Constraint PDFs** for all sources systematics
- **Impact** of each systematic uncertainty on all relevant model parameters

→ *Statistical methods are only as accurate (and/or optimal) as the description provided by the model!*

Technically, MC simulation provides most of these inputs. However 2 problematic issues:

- **Potential MC/data differences**
- **Limited MC statistics**

Which need to be addressed with (yet more) systematics.

Statistical Modeling:
I. Component PDFs

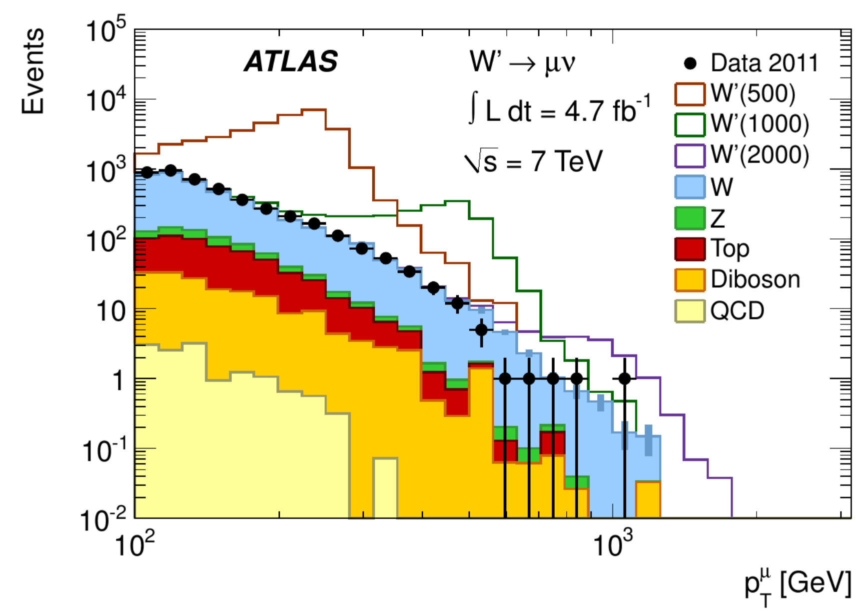
PDFs : Binned likelihood

Binned case:

→ PDF usually just a normalized histogram, from

- **MC sample** or
- **Data control region (CR)**

Eur. Phys. J. C (2012) 72: 2241



⇒ **Statistical uncertainties** on the prediction:

- **Data CR:** counts as **statistical** uncertainty
- **MC sample**: uncertainty can be reduced without collecting more data (just need more CPU!) ⇒ Counted as **systematic**

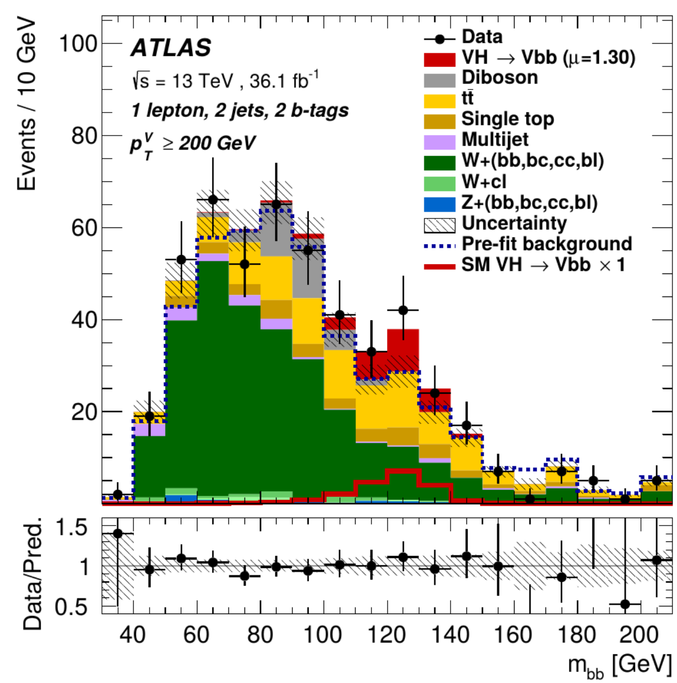
JHEP 12 (2017) 024

Independent counts in each bin

⇒ a separate ***MC statistics*** **NP** in each bin

→ Poisson constraints **Pois**(N_i^{MC}; N_i^{true})

Total uncertainty ~ $\sqrt{\sigma^2_{\text{data stats}} + \sigma^2_{\text{MC stats}} + \ldots}$



⇒ need enough MC to avoid spoiling the sensitivity!

MC Statistics Requirements

e.g. **Discovery**: Total uncertainty: $\sigma_S^2 \sim \sqrt{\sigma^2_{\text{data stats}} + \sigma^2_{\text{MC stats}} + \ldots}$

⇒ need $\sigma_{\text{MC stats}} \ll \sigma_{\text{data stats}}$

$B_{\text{MC}} \gg B_{\text{data}}$

By how much ?

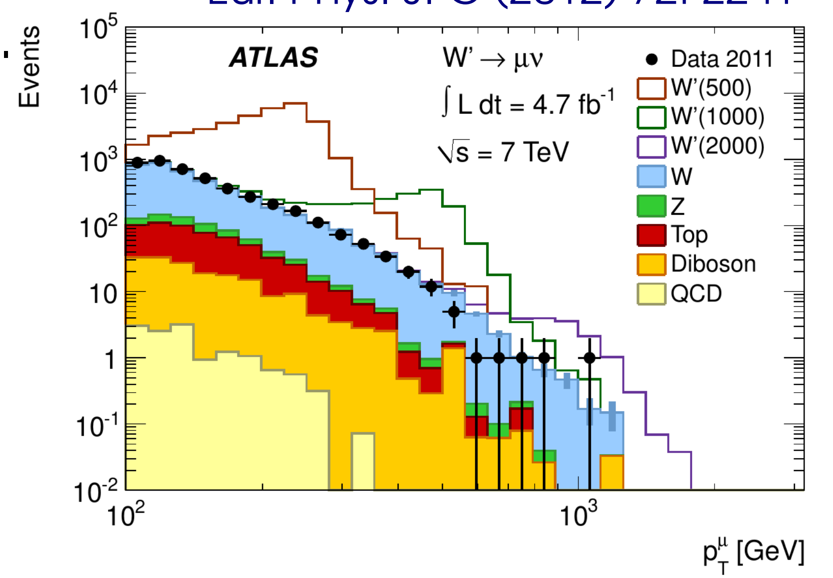
B_{MC}/B_{data} (α)	$\sigma_{\text{MC stats}}/\sigma_{\text{data stats}}$ ($1/\sqrt{\alpha}$)	$\sigma_{\text{data+MC stats}}/\sigma_{\text{data stats}}$ ($\sqrt{(1+\alpha^{-1})}$)
1	1	**1.41**
4	0.5	**1.12**
25	0.2	**1.02**

In the presence of a signal (e.g. limit-setting, N_{sig} measurement), relevant uncertainty is √(S+B).
⇒ **S/B** also matters:

$$\frac{\sigma_S}{S} \sim \sqrt{1 + \frac{S}{B} + \frac{B_{\text{data}}}{B_{\text{MC}}}\frac{1}{1+S/B}}$$

- **low S/B** : same problem as for discovery
- **high S/B** : no issue, dominated by uncertainty on signal itself.

Eur. Phys. J. C (2012) 72: 2241



PDF shapes: Unbinned likelihood

Smooth backgrounds : Describe distribution using appropriate **function**
⇒ **Unbinned likelihood.** Describes sideband + signal region in one fit.

Phys. Lett. B241 (1990) 278-282

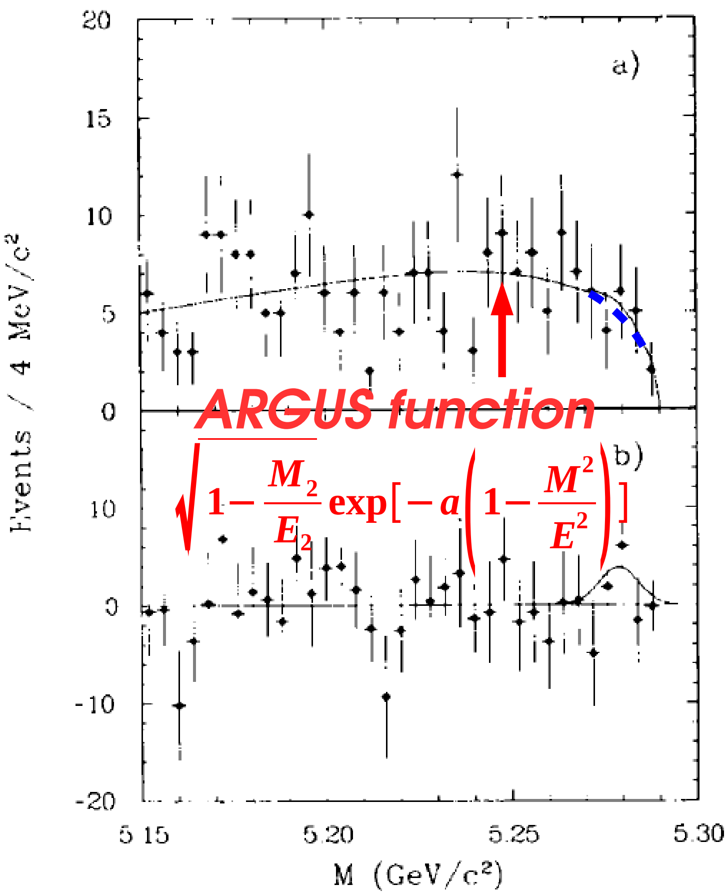
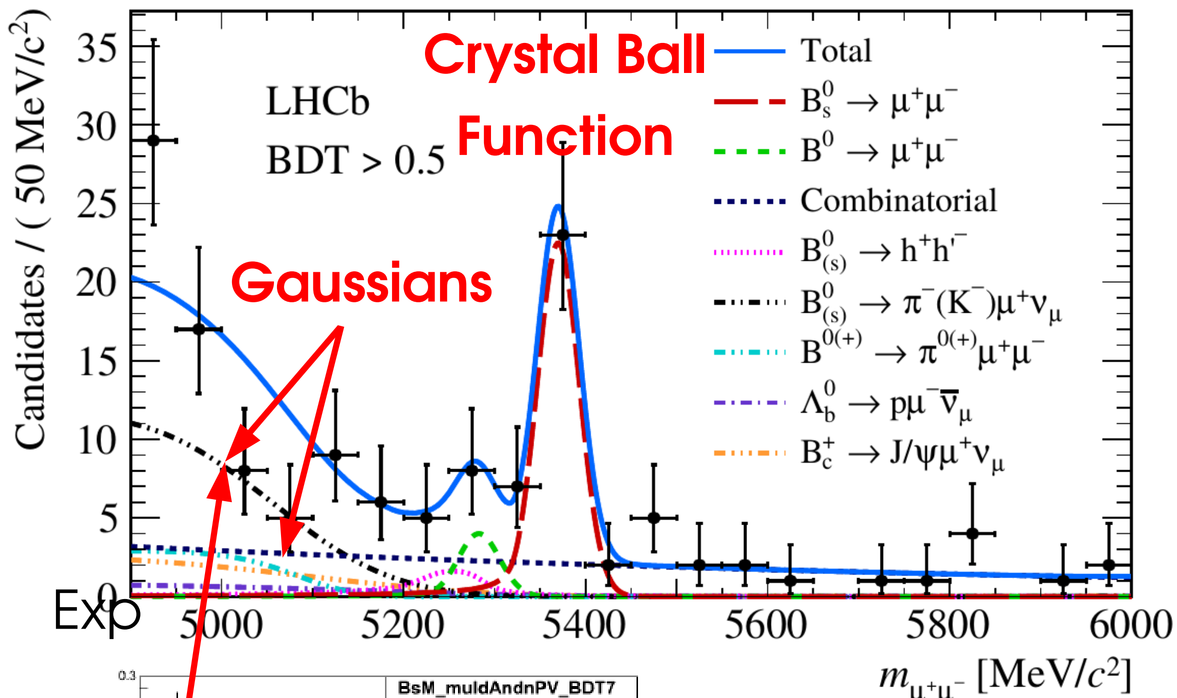
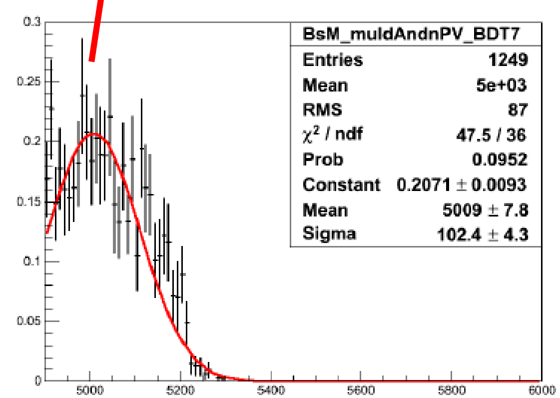


Fig. 1. Invariant mass distribution of the decay $B^+ \to \pi^+\pi^0$. (a) At the $\Upsilon(4S)$; the curve shows the result of the maximum likelihood fit described in the text. (b) After subtraction of the continuum contribution. The gaussian curve represents the 90% CL upper limit on the signal from the above fit (see table 1).

Phys. Rev. Lett. 118 (2017), 191801





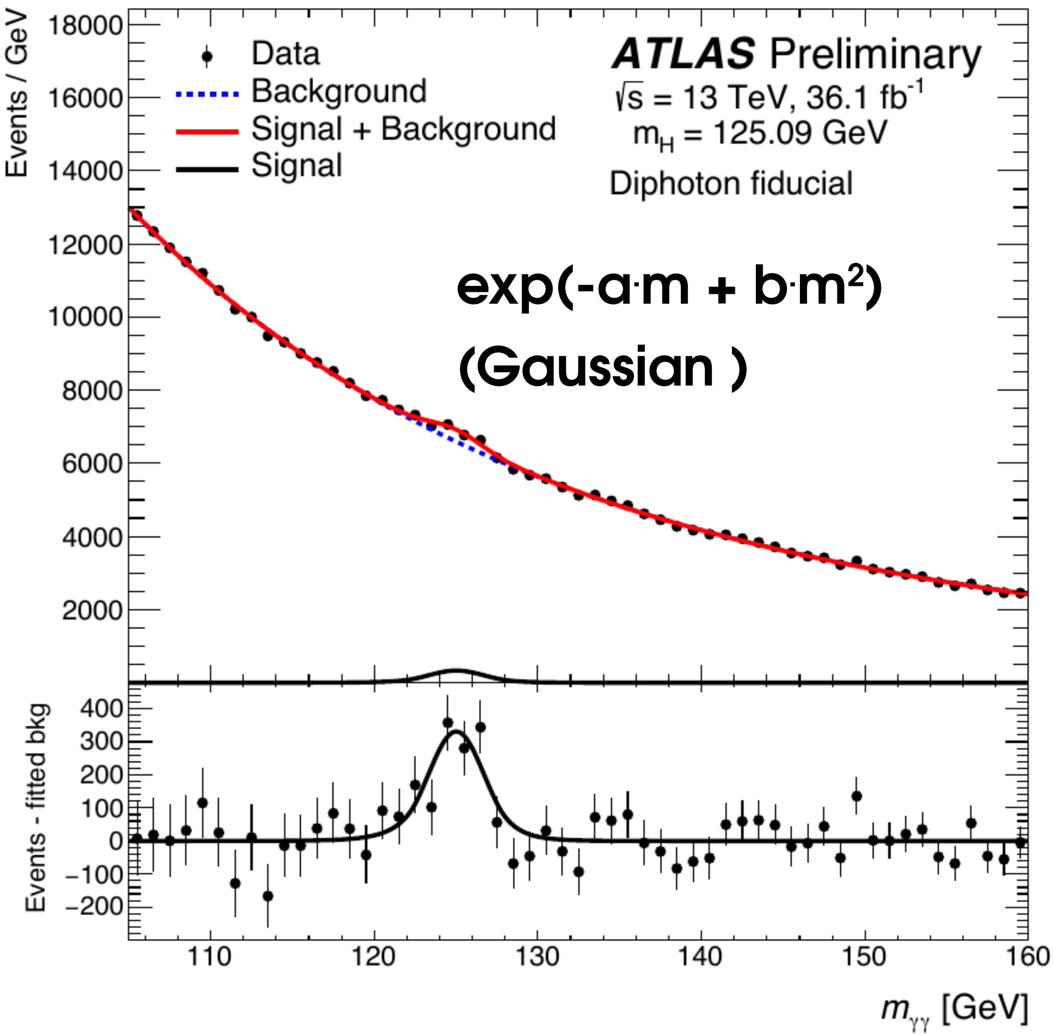
S. Oggero Ph. D. Thesis

Functions help smooth MC stats fluctuations

PDF Shapes: Unbinned likelihood

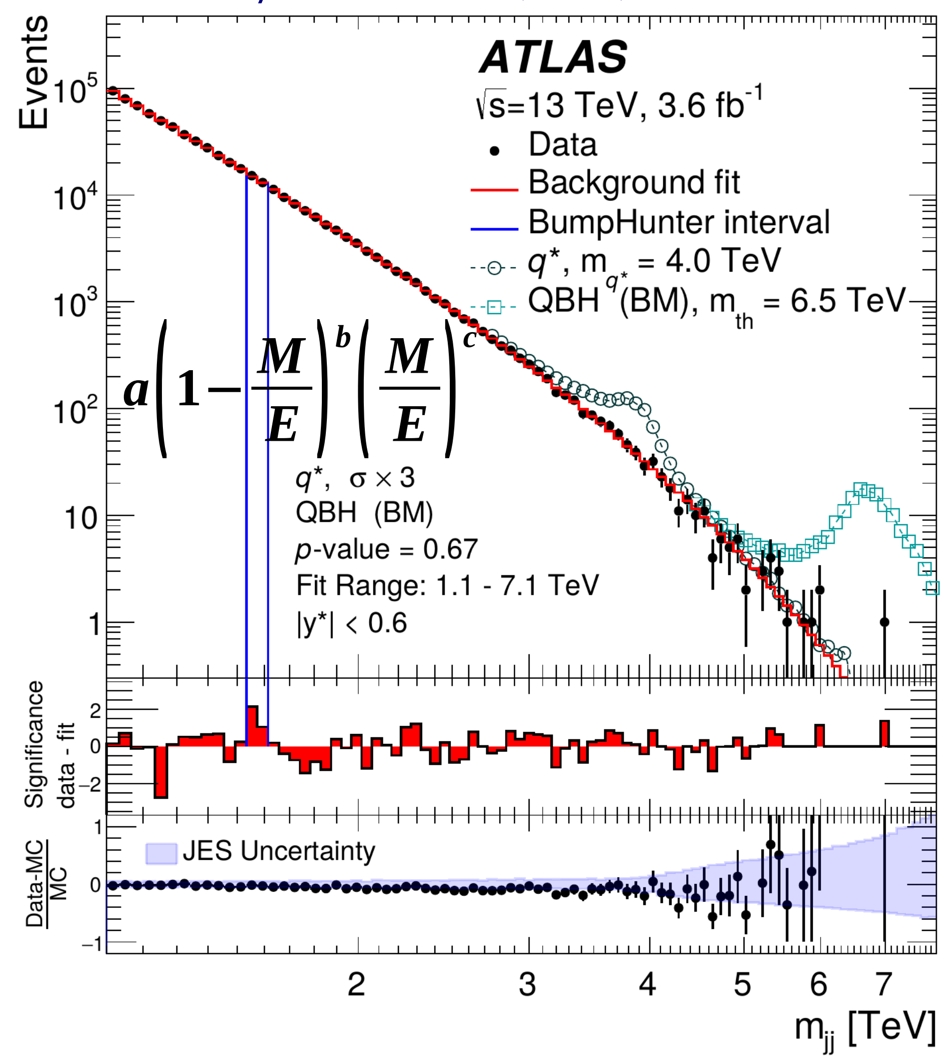
Widely used in HEP for smooth backgrounds (→ no resonances or threshold)

H→ γγ Measurements



X→ jj Search

Phys.Lett. B754 (2016) 302-322



Signal Bias in Unbinned likelihoods

Function usually ad-hoc (no closed form expression for (theory ⊗ detector effects), or usually even theory by itself…)

→ **may not accurately describe the data**

⇒ **Introduce free parameters, fit in sidebands**

→ **Biases may still remain due to functional form itself**

Problematic especially for **low S/B**

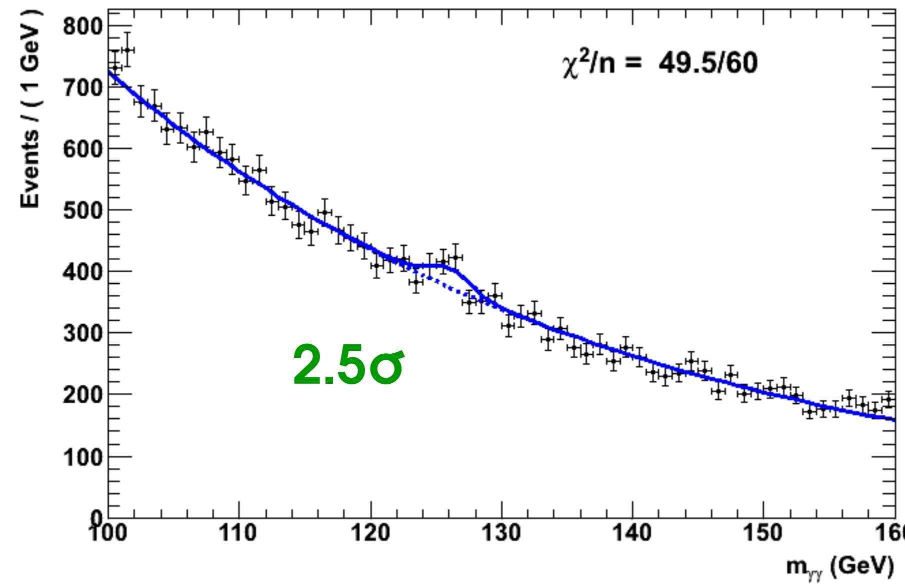
→ small mismodelings of B can be large compared to S.

→ **χ^2 test in sideband may not help**: even a large bias on the scale of S (≪ B) may remain within stat errors in the sideband!

Jan 2012 Higgs search paper

(4.9 fb^{-1} of 2011 data)

exponential



Situation doesn't improve with more luminosity:

→ Reduced statistical uncertainties in sideband, but

→ Also reduced σ_S, in the same proportion

Signal Bias in Unbinned likelihoods

Function usually ad-hoc (no closed form expression for (theory ⊗ detector effects), or usually even theory by itself…)

→ **may not accurately describe the data**

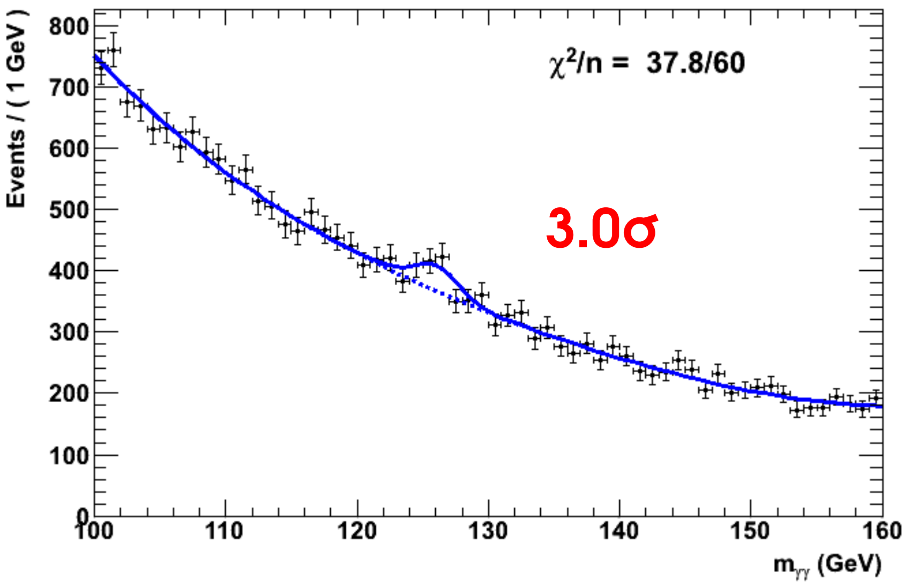
⇒ **Introduce free parameters, fit in sidebands**

→ **Biases may still remain due to functional form itself**

Jan 2012 Higgs search paper

(4.9 fb^{-1} of 2011 data)

polynomial



Problematic especially for **low S/B**

→ small mismodelings of B can be large compared to S.

→ **χ^2 test in sideband may not help**: even a large bias on the scale of S (≪ B) may remain within stat errors in the sideband!

Situation doesn't improve with more luminosity:

→ Reduced statistical uncertainties in sideband, but

→ Also reduced σ_S, in the same proportion

Signal Bias in Unbinned likelihoods

If data cannot fix B shape, **use MC**
→ Measure signal bias N_{SS} on "credible" shapes taken from MC (***Spurious signal***)
→ take the maximum bias as systematic

Works well if the true distribution is somewhere in the space of MC distributions scanned…

Also **Impose:**

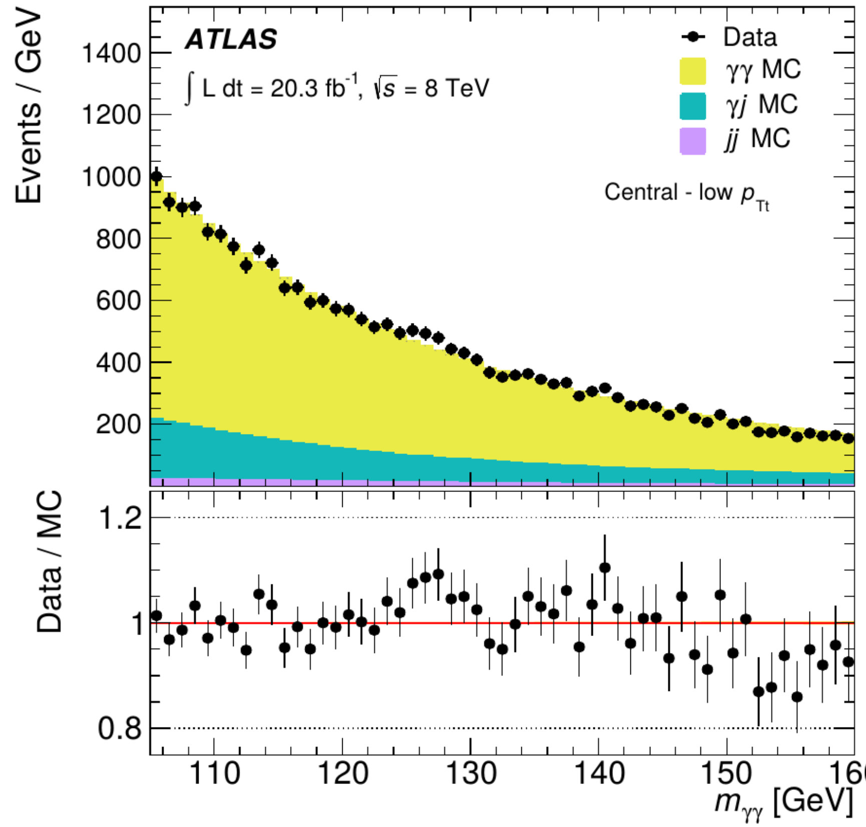
N_{SS} < 20% σ_{stat} (small contribution to σ_{total})

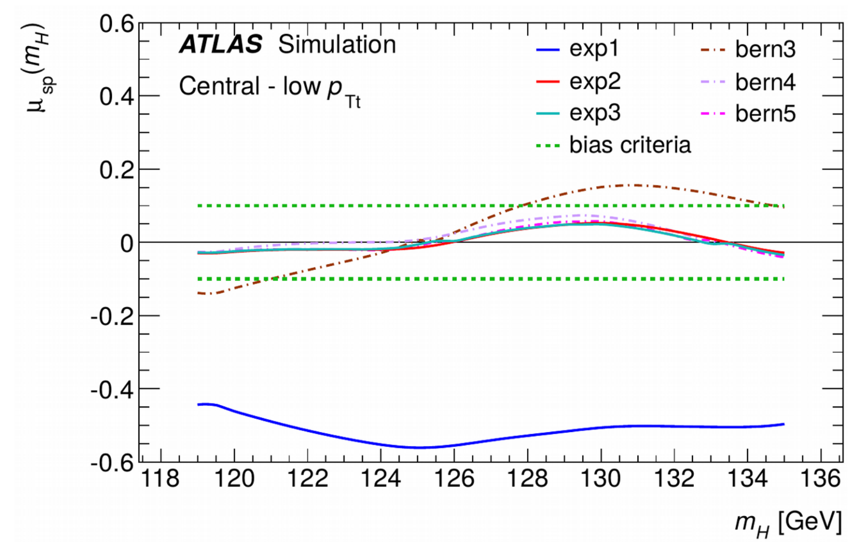
OR

N_{SS} < 10% S_{exp} (small bias on measured S)

Second criterion more stringent at higher S/√B.

If criteria are not met, move to more complex functions (→ more free parameters)





Signal Bias in Unbinned likelihoods

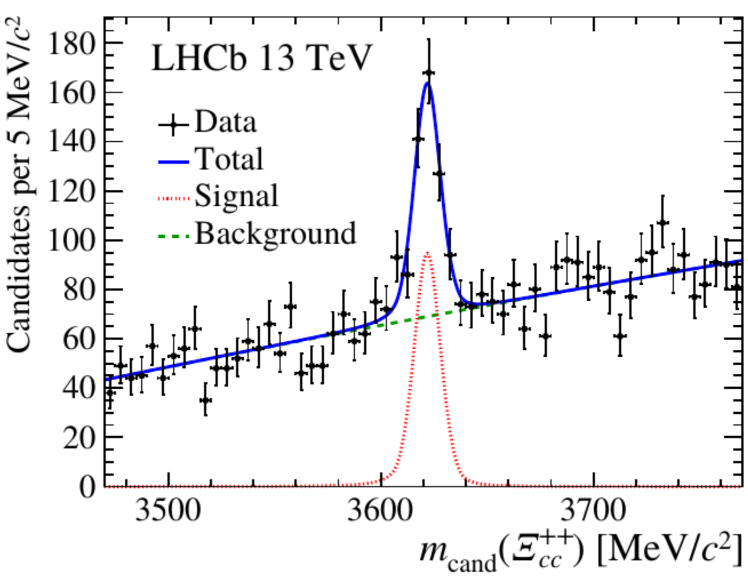
Problem: for small MC stats, measured bias dominated by fluctuations
→ again, need high MC stats (B_{MC} > 25 B_{data}) when S/B is low.

B_{MC}/B_{data} (α)	$\sigma_{\text{MC stats}}/\sigma_{\text{data stats}}$ ($1/\sqrt{\alpha}$)	$\sigma_{\text{data+MC stats}}/\sigma_{\text{data stats}}$ ($\sqrt{(1+\alpha^{-1})}$)
1	100%	1.41
4	50%	1.12
25	20%	1.02

← N_{SS} < 20% σ_{stat}

→ Can compromise on criterion level
(50% instead of 20% ?)

→ As before, better situation at at high S/B



Phys. Rev. Lett. 118, 182001 (2017)

Usual Functions

Comm. Soc. Math. Kharkov 13, 1-2, 1912.

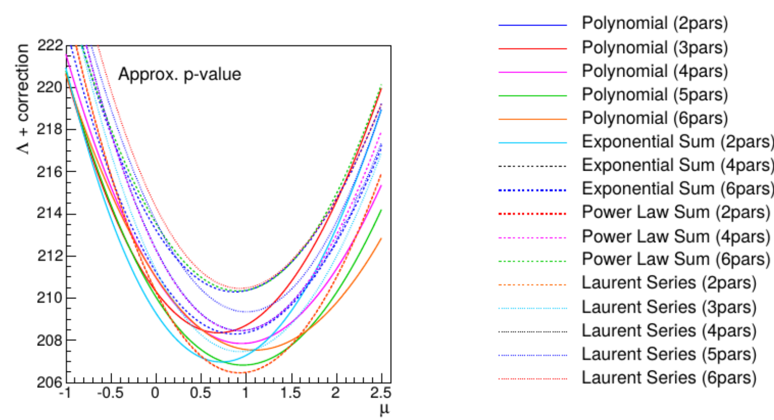
Polynomials: various basis choices (Chebyshev, Bernstein,…)

Bernstein basis:

$$B_{k,n}(x)=\binom{k}{n}x^k(1-x)^{n-k} \text{ for } 0\leq x\leq 1$$

→ **Positive coefficients ⇒ positive polynomial everywhere**, useful to avoid numerical issues in -2 log(PDF) computation

JINST 10 (2015) no.04, P04015



Exponential family : exp(polynomial)

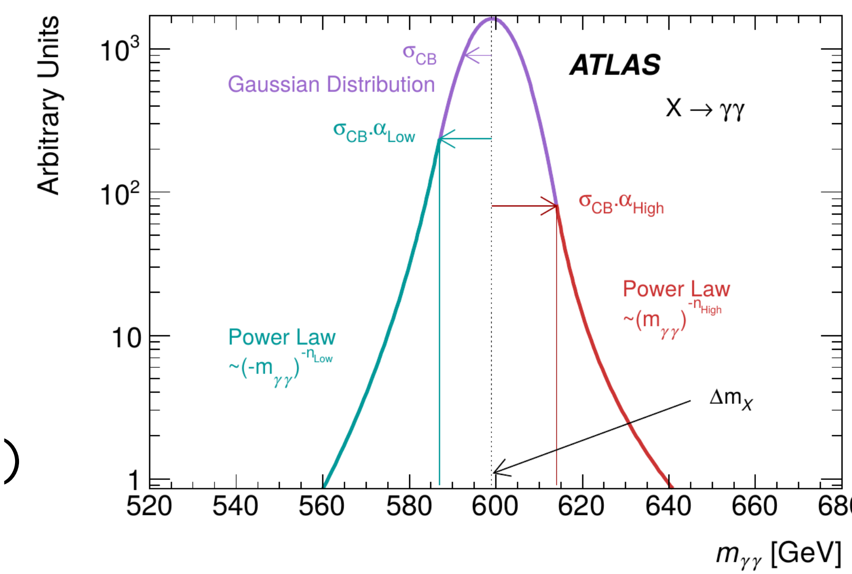
Power laws : x^α, $x^\alpha(1-x)^\beta$, …

Gaussians

Crystal Ball Functions

$$t = (m_{\gamma\gamma} - \mu_{\mathrm{CB}})/\sigma_{\mathrm{CB}}$$

$$N\cdot\begin{cases} e^{-0.5t^2} & \text{if } -\alpha_{\mathrm{low}} \leq t \leq \alpha_{\mathrm{high}} \\ e^{-0.5\alpha_{\mathrm{low}}^2}\left[\frac{\alpha_{\mathrm{low}}}{n_{\mathrm{low}}}\left(\frac{n_{\mathrm{low}}}{\alpha_{\mathrm{low}}}-\alpha_{\mathrm{low}}-t\right)\right]^{-n_{\mathrm{low}}} & \text{if } t < -\alpha_{\mathrm{low}} \\ e^{-0.5\alpha_{\mathrm{high}}^2}\left[\frac{\alpha_{\mathrm{high}}}{n_{\mathrm{high}}}\left(\frac{n_{\mathrm{high}}}{\alpha_{\mathrm{high}}}-\alpha_{\mathrm{high}}+t\right)\right]^{-n_{\mathrm{high}}} & \text{if } t > \alpha_{\mathrm{high}}, \end{cases}$$



→ **Sums of the above**

→ **Convolutions** (resolution ⊗ Breit-Wigner, ...)

Discrete Profiling

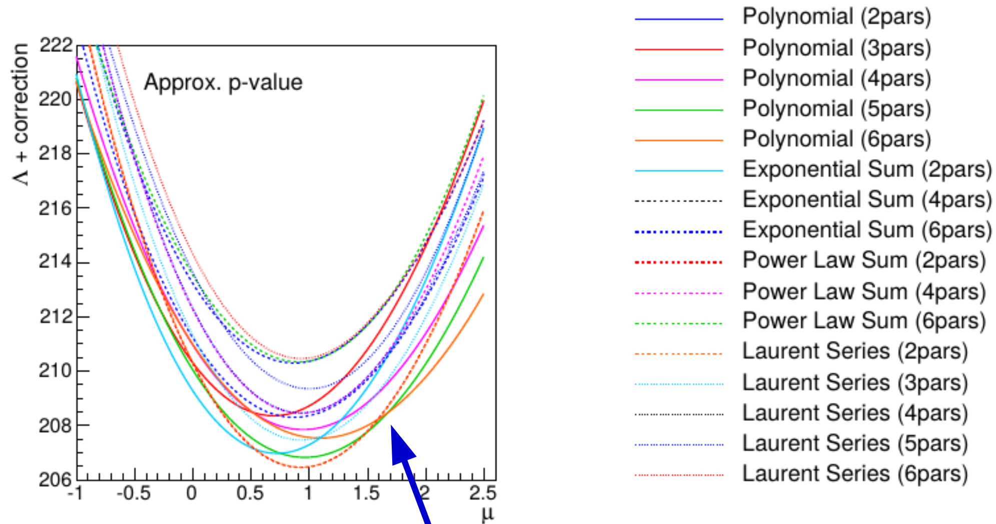
JINST 10 (2015) no.04, P04015

Idea: treat the **type of function** and **number of parameters** as discrete NPs, profiled in data

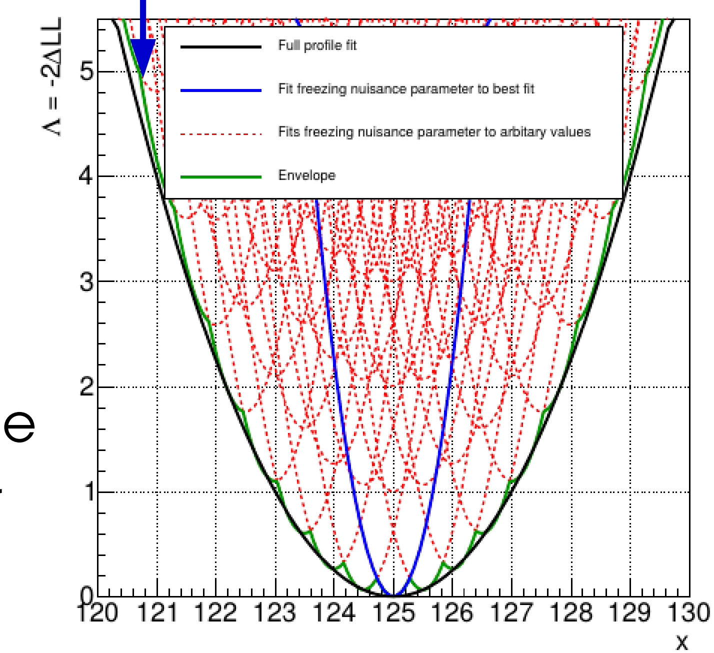
→ **Let data choose the best shape**

→ Similar principle as other NPs, except for discrete nature

→ Need a **penalty on N_{pars}** to avoid always choosing functions with high N_{pars}



Take lower envelope of all functions when profiling



→ Used in the CMS H→γγ analysis, works well in this context.

Caveats:

→ for N categories and M functional forms, **M^N possibilities to check** in principle – difficult in practice

→ Need a **well-chosen pool of sensible functions** for the method to work

→ **Large MC samples** for selection and checks

Gaussian Processes: 1-slide Summary

the data are drawn from one

HUGE*

Gaussian

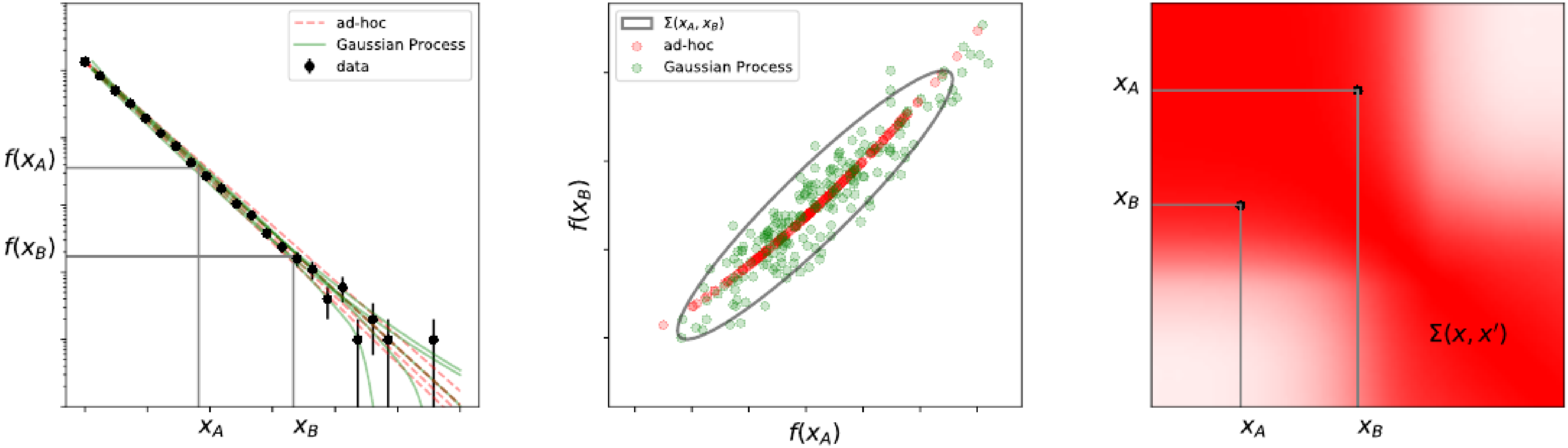
* the dimension is the number of data points.

Image Credits:
K. Cranmer

Gaussian Processes: Longer 1-slide Summary

- Describe background distribution through the **correlations between values at different points**.
- More flexible than a functional form
- **Correlation function** (*Kernel*) can be
  - Defined using a length scale, to ignore narrow peaks
  - Obtained from first principles (e.g. from known JES/PDF effects)

$$K(x_1, x_2) = \exp\left[-\frac{(x_1 - x_2)^2}{2L^2}\right]$$



arXiv:1709.05681

⊕ More flexible than functional form, degrees of freedom less ad-hoc

⊖ Still need large MC samples to check for signal bias

⊖ Mainly for Gaussian processes, not well-adapted to Poisson regime

Statistical Modeling:
II. Systematics

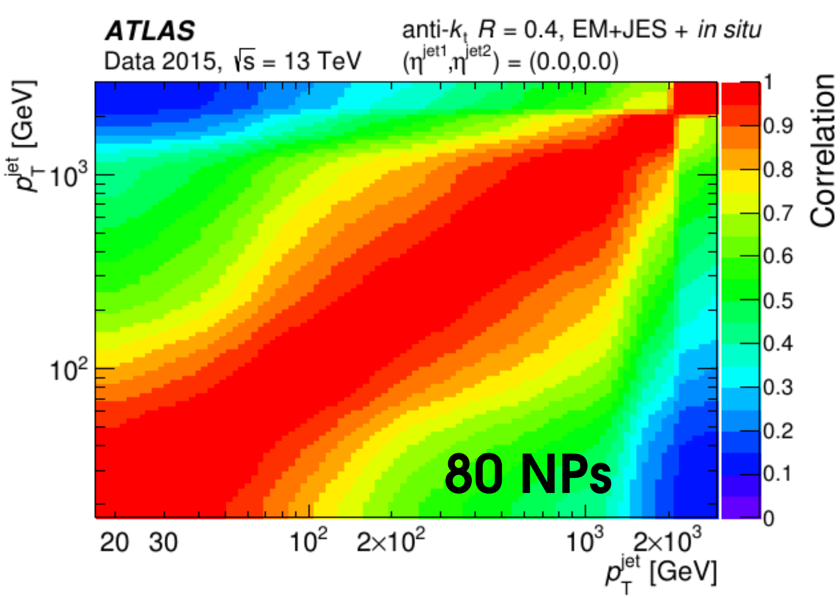
Systematics NPs

Each systematics NP represent **an independent source of uncertainty**
⇒ Usually constrained by a single 1-D PDF (Gaussian, etc.)

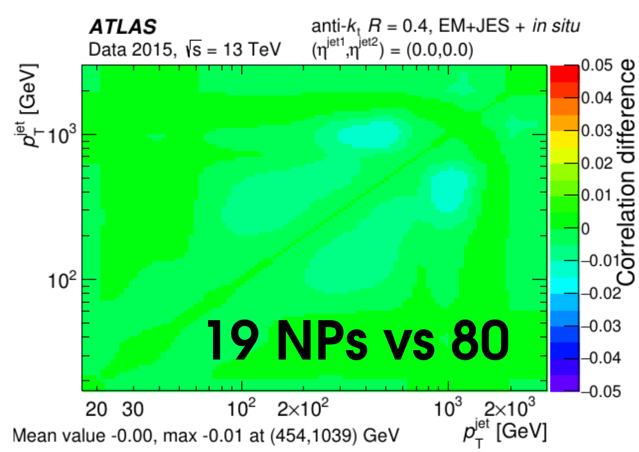
Sometimes multiple parameters **conjointly constrained** by an n-dim. PDF.
→ multiple measurements constraining multiple NPs

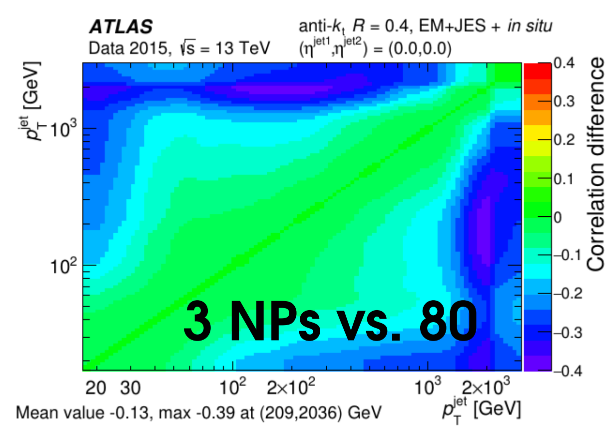
Assume **n-dim Gaussian** PDF: then can **diagonalize the covariance matrix C** and re-express the uncertainties in basis of eigenvector NPs ⇒ **n 1-dim PDFs**

Can also diagonalize to **prune** irrelevant uncertainties: keep NPs with large eigenvalues, combine in quadrature the others



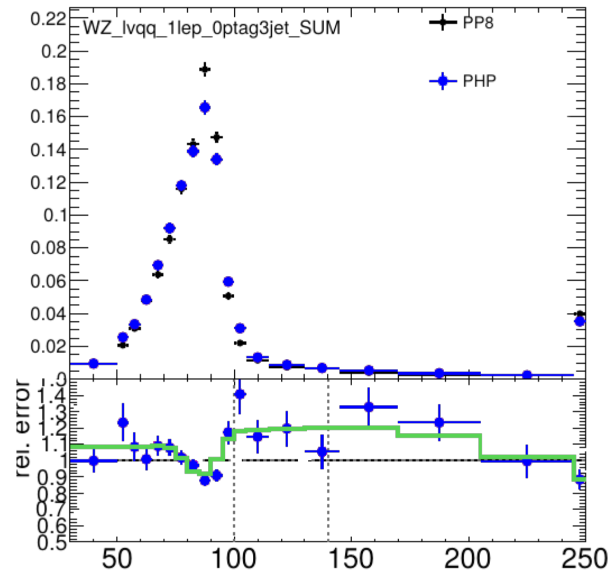
Phys.Rev. D96 (2017) no.7, 072002





Systematics : Impact on Model

The effect of **each NP θ_i** should be propagated to all the relevant **model parameters X_j**.



→ **Propagation through MC:**

1. Apply ±1σ systematic variations in MC,

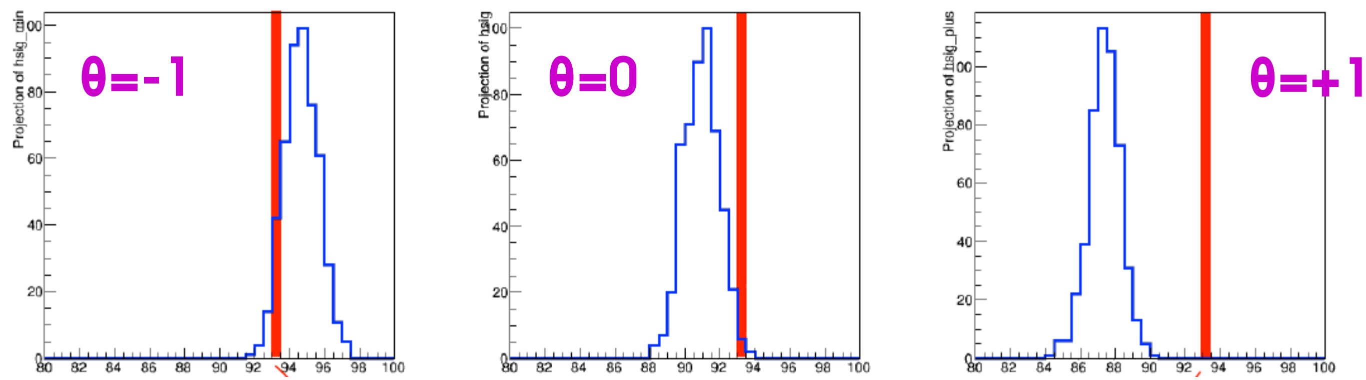
   ⇒ obtain shifted values $\mathbf{X_j^{\pm} = X_j^0\,(1 \pm \Delta_{ij})}$.

   → Possibly smooth out MC stats effects

2. Implement systematic in model, e.g. replace or morph shapes:

Constrained by unit Gaussian

$$X_j \rightarrow X_j^0\left(1 + \Delta_{ij}\theta_i\right)$$



→ can affect event yields, shapes, etc.

Assumes **Gaussian uncertainties** and **linear impact** on model parameters

Systematics : Constraints

Ideally, constraint = **likelihood of auxiliary measurement**
⇒ e.g. Poisson for constraint from counting in a low-stat CR.

Sometimes no clear auxiliary measurement
⇒ Semi-arbitrary "pseudo-measurement" motivated by Central Limit Theorem:

- **Gaussian** for additive corrections
- **Log-normal** for multiplicative corrections

Gaussian:

Constrained by unit Gaussian

- represent impact as $X_j \rightarrow X_j^0 \left(1 + \Delta_{ij} \theta_i\right)$

  → or similar morphing for distributions

Can include asymmetric variations Δ^+, Δ^-: $X_j \rightarrow X_j^0 \left(1 + \begin{cases} \Delta_{ij}^+ \theta_i & \theta_i > 0 \\ \Delta_{ij}^- \theta_i & \theta_i < 0 \end{cases}\right)$

However discontinuity in derivative at 0, so use smooth interpolation instead,
e.g. implementation in `RooStats::HistFactory::FlexibleInterpVar`.

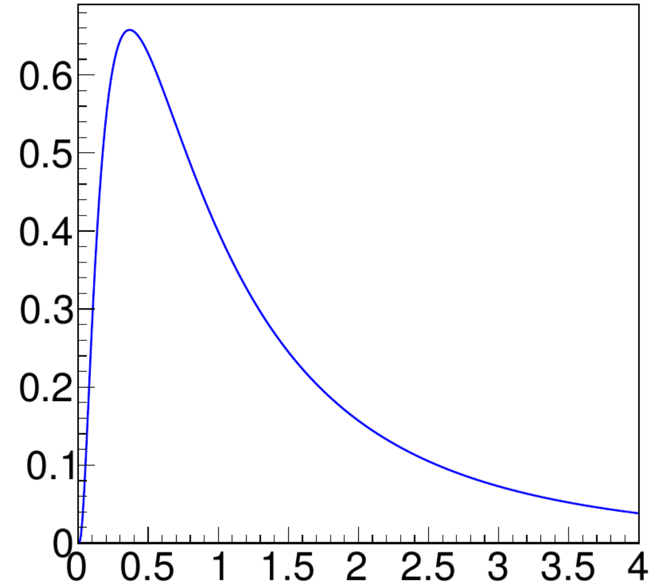
Systematics : Log-normal Constraint

Log-normal: x ~ log-normal if log(x) is normal

→ **always > 0**, useful to avoid numerical issues

→ **PDF**:

$$P(s; X_{0,}\kappa) = \frac{1}{x\,\kappa\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\log(x) - X_0}{\kappa}\right)^2\right)$$



However usually simpler to implement as :

$$X_j \rightarrow X_j^0\,\exp(\kappa_{ij}\theta_i)$$

where θ_i is constrained by a unit Gaussian as usual

→ Correct form for a multiplicative uncertainty:

$$\log\sqrt[n]{(X_0 k_1)(X_0 k_2)\ldots(X_0 k_n)} = \frac{1}{n}\sum_{i=1}^{n}\log(X_0 k_i) \overset{n\to\infty}{\sim} G\left(\log X_0,\ \frac{RMS(\log(k))}{\sqrt{n}} = \kappa\right)$$

Similarly to Gaussian → represent $X = X_0\, e^{\kappa\theta}$ ~ G(log X_0, κ) if θ ~ G(0,1)

Which κ to use ? κ = RMS(X) only at first order. For larger uncertainties,

e.g. **Match ±1σ variations**: $X_j(\theta=\pm 1) = X_j^{\pm} \Rightarrow \quad \kappa_{\pm} = \pm\log(X_j^{\pm}/X_j^0)$

Implemented in `RooStats::HistFactory::FlexibleInterpVar`.

Systematics : Theory Constraints

Missing high-order terms in perturbative calculations: evaluate from scale variations – but no underlying random process. Possible constraint shapes:

- **Gaussians** (ATLAS/CMS Higgs analyses, see Yellow Report 4, I.4.1.d)

  → Usually several independent "sources" of uncertainty(QCD/EW/resummation)

  ⇒ overall uncertainty may be rather Gaussian

  → Numerically well-behaved

  → Uncertainties add in quadrature as usual

- **Flat constraints** : "100% confidence" intervals

  → no preference for any value in the range

  → Need regularization to avoid numerical issues

  → uncertainties add linearly

→ For Higgs cross-sections, rather similar results for both cases

Constraints : Two-point systematics

Sometimes differences between 2 discrete cases → e.g. Pythia vs. Herwig

Solutions:

- Results for one case only
- Full results for both cases
- Single result with an uncertainty that covers the difference

  → ***Two-point* uncertainty**

Usually implemented as **1D linear interpolations** between the two cases

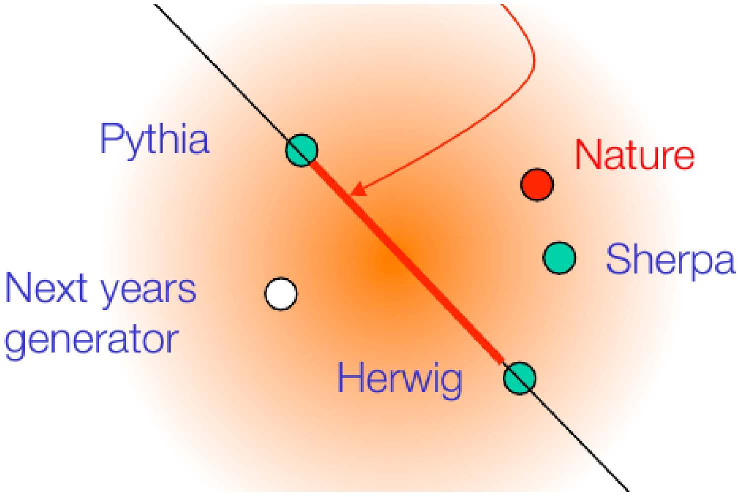
→ However cannot guarantee this covers the space of possible configurations

⇒ This is not even a pseudo-measurement…

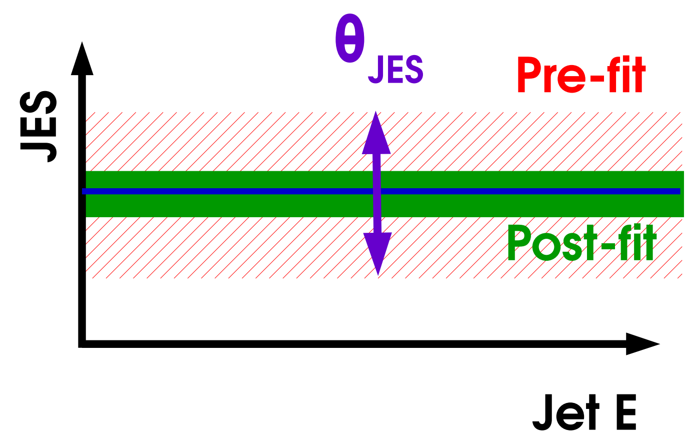
Ideally, need to define proper uncertainties within a single model, which would cover the other cases

→ e.g. showering uncertainties within Pythia, covering Herwig

→ Usually a difficult task

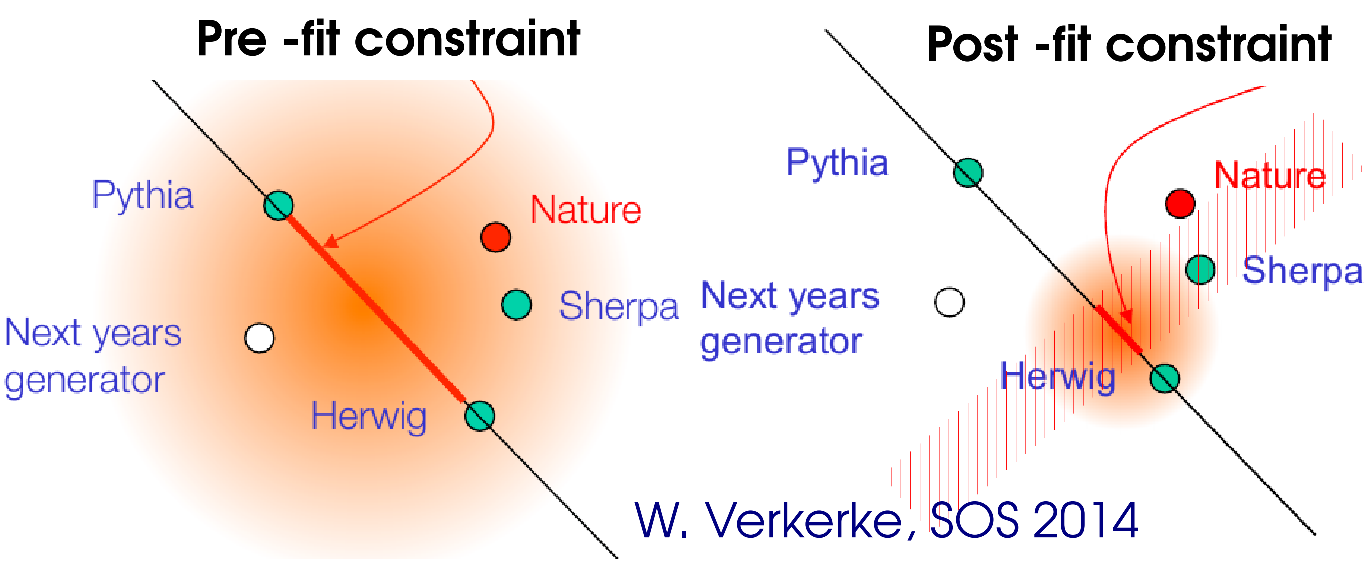


Profiling Issues

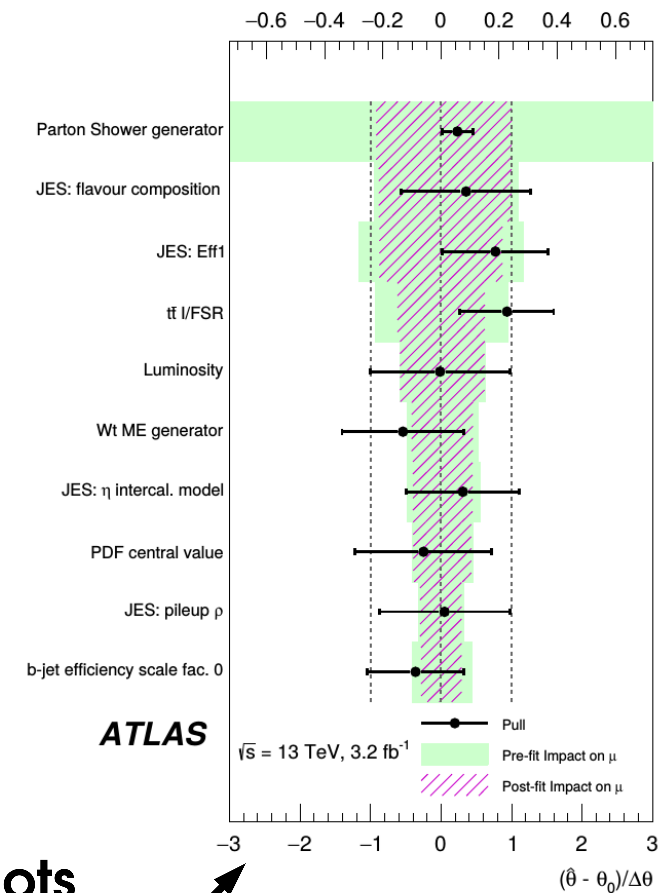


Too simple modeling can have unintended effects
→ e.g. single Jet E scale parameter:
⇒ Low-E jets calibrate high-E jets – intended ?

Two-point **uncertainties**:
→ Interpolation may not cover full configuration space, can lead to too-strong constraints



W. Verkerke, SOS 2014



NP central values and uncertainties in pull/impact plots provide important "debugging" information for profiling

Outline

Profiling

Look-Elsewhere Effect

Bayesian methods

Statistical modeling in practice

- Building binned likelihoods
- Choosing PDFs in unbinned likelihoods
- Implementing systematics

BLUE

BLUE

BLUE

C_{ij} : covariance matrix of measurements:

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

ρ: correlation coefficients

Commonly-used ansatz for combination of measurements:

1. **Build a χ^2:** (same as -2logL for Gaussian L)

$$\chi^2(X) = \sum_i \left(X_i^{obs} - X\right) C_{ij}^{-1} \left(X_j^{obs} - X\right)$$

$$\lambda = \frac{C^{-1} J}{J^T C^{-1} J}, \quad J = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}$$

2. **Estimate combined X from minimum of $\chi^2(X)$**

- In the Gaussian case, equivalent to ML solution ⇒ inherits good properties:
  - **Unbiased** : $\langle \hat{X} \rangle = X^*$
  - **Optimal**: minimizes the combined uncertainty
- Solution is a linear combination of the inputs:

λ_i = combination weight of measurement i

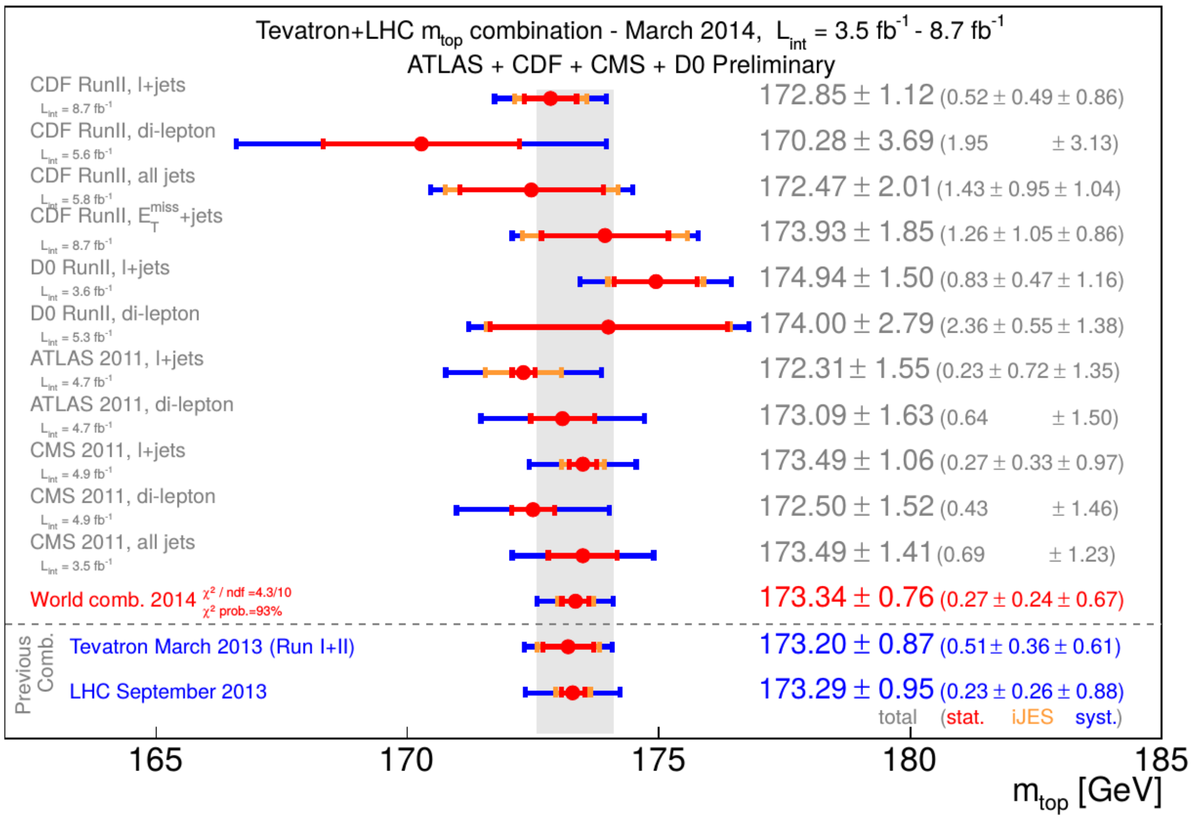
$$\hat{X} = \sum_i \lambda_i X^{obs,i}$$

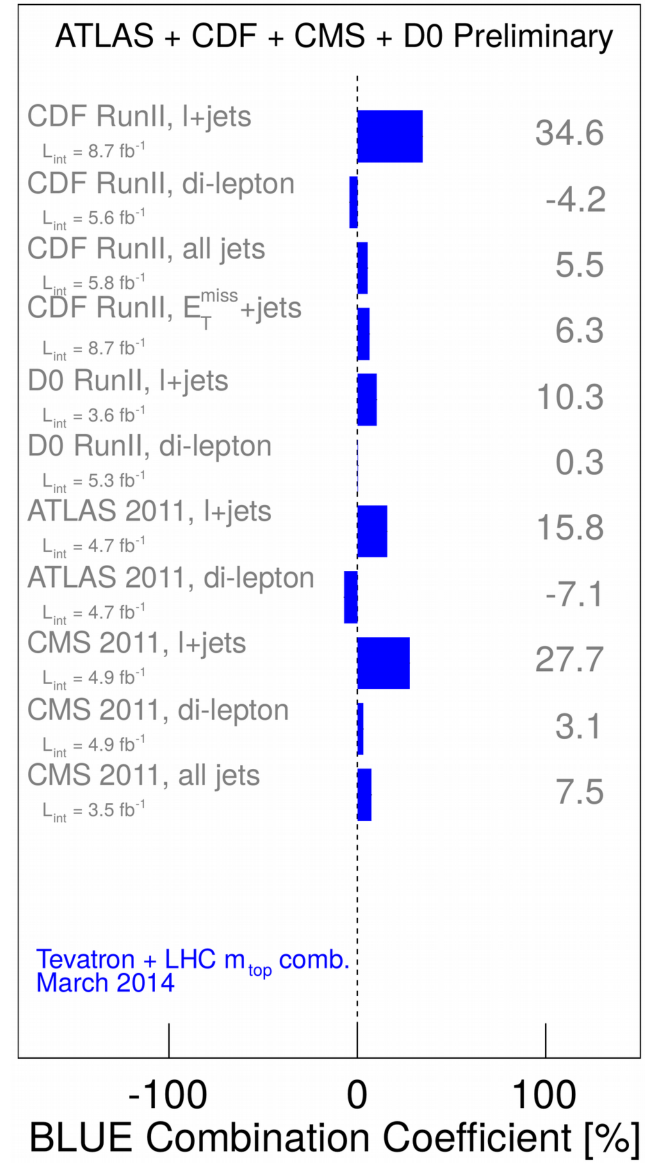
⇒ **"Best Linear Unbiased Estimator"** (**BLUE**)

BLUE Example

ATLAS-CONF-2014-008

Example: World m_{top} combination





Limitation: relies on **Gaussian assumptions** (satisfied in this case!)

Negative weights possible! (for large correlations, see Eur. Phy. J. C 74 (2014), 2717)

BLUE and PLR

$$X_1 = X + \Delta_1\theta \sim G(X^*, \sigma_1)$$
$$X_2 = X + \Delta_2\theta \sim G(X^*, \sigma_2)$$
$$\theta \sim G(0,\ 1)$$

PLR Computation: 2 measurements
+ 1 auxiliary measurement

Single measurement: $\lambda(X,\theta) = \frac{1}{\sigma_1^2}(X + \Delta_1\theta - X_1^{\text{obs}})^2 + (\theta - \theta^{\text{obs}})^2$

MLEs: $\begin{cases} \hat{\theta} = \theta^{\text{obs}} \\ \hat{X} = X_1^{\text{obs}} - \Delta_1\theta^{\text{obs}} \end{cases}$

PLR: $\lambda(X) = \frac{(X - \hat{X})^2}{\sigma_{1,\text{tot}}^2}$ $\quad \sigma_{1,\text{tot}}^2 = \sigma_1^2 + \Delta_1^2$

Combination: $\lambda(X,\theta) = \frac{1}{\sigma_1^2}(X + \Delta_1\theta - X_1^{\text{obs}})^2 + \frac{1}{\sigma_2^2}(X + \Delta_2\theta - X_2^{\text{obs}})^2 + (\theta - \theta^{\text{obs}})^2$

MLE: $\hat{X} = \lambda_1 X_1^{\text{obs}} + \lambda_2 X_2^{\text{obs}} + \lambda_\theta \theta^{\text{obs}}$ $\quad \lambda_{1(2)} = \frac{\sigma_{2(1),\text{tot}}^2 - \Delta_1\Delta_2}{\sigma_{1,\text{tot}}^2 + \sigma_{2,\text{tot}}^2 - 2\Delta_1\Delta_2}$

PLR: $\lambda(X) = \frac{(X - \hat{X})^2}{\sigma_{X,\text{tot}}^2}$ $\quad \sigma_{X,\text{tot}}^2 = \frac{\sigma_{1,\text{tot}}^2\sigma_{2,\text{tot}}^2 - \Delta_1^2\Delta_2^2}{\sigma_{1,\text{tot}}^2 + \sigma_{2,\text{tot}}^2 - 2\Delta_1\Delta_2}$

BLUE and PLR

Eur. Phys. J. C, 74 (2014) 2717

BLUE computation: measurements X_1 and X_2 with uncorrelated statistical uncertainties σ_1 and σ_2, correlated systematics Δ_1 and Δ_2.

Single measurement: stat uncertainty $\boldsymbol{\sigma_1}$, systematic $\boldsymbol{\Delta_1}$

- Uncorrelated uncertainties
- Assume everything is Gaussian

⇒ Uncertainties add in quadrature:

$$\sigma^2_{1,\mathrm{tot}} = \sigma_1^2 + \Delta_1^2$$

Combination:

$$C = \begin{bmatrix} \sigma^2_{1,\mathrm{tot}} & \rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}} \\ \rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}} & \sigma^2_{2,\mathrm{tot}} \end{bmatrix} \qquad \rho = \frac{\Delta_1 \Delta_2}{\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}}}$$

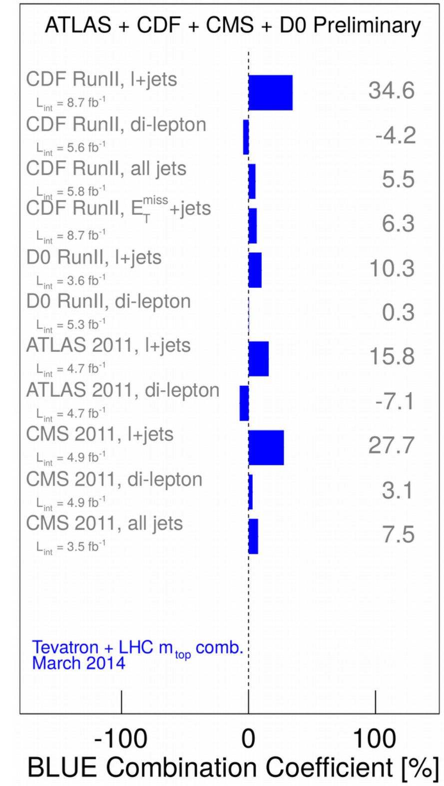
BLUE weights

$$\hat{X} = \lambda_1 X_1^{\mathrm{obs}} + \lambda_2 X_2^{\mathrm{obs}} \qquad \lambda_{1(2)} = \frac{\sigma^2_{2(1),\mathrm{tot}} - \rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}}}{\sigma^2_{1,\mathrm{tot}} + \sigma^2_{2,\mathrm{tot}} - 2\rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}}}$$

Propagate uncertainties from C:

$$\sigma^2_{X,\mathrm{tot}} = \frac{\sigma^2_{1,\mathrm{tot}}\sigma^2_{2,\mathrm{tot}}\left(1-\rho^2\right)}{\sigma^2_{1,\mathrm{tot}} + \sigma^2_{2,\mathrm{tot}} - 2\rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}}}$$

Negative BLUE Weights



Occasionally, negative BLUE weights:

Can happen if $\rho \sim 1$:

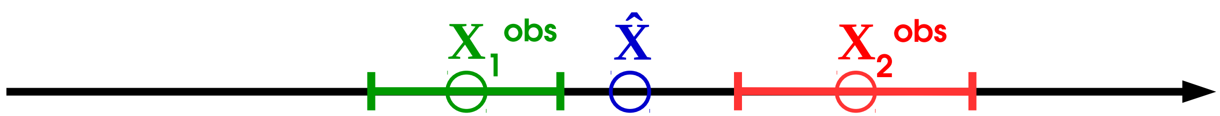
$$\lambda_2 = \frac{\sigma_{1,tot}\left(\sigma_{1,tot} - \rho\,\sigma_{2,tot}\right)}{\sigma^2_{1,tot} + \sigma^2_{2,tot} - 2\rho\,\sigma_{1,tot}\,\sigma_{2,tot}} < 0 \quad \text{for} \quad \rho > \frac{\sigma_{1,tot}}{\sigma_{2,tot}}$$

Not intuitive! (Can also have $\lambda_2 = 0$ for $\sigma_{1,tot} = \rho$ σ2,tot…)

Can be explained in the PLR picture:

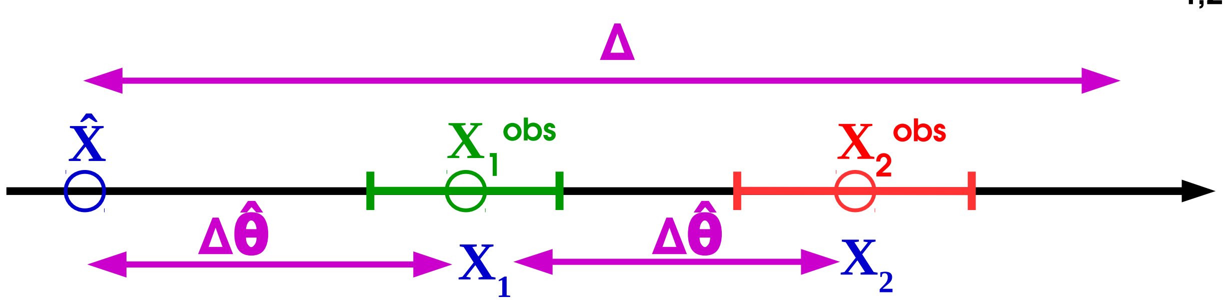
$$X_1 = X + \Delta\theta$$
$$X_2 = X + 2\Delta\theta$$

Without correlated systematics (Δ = 0):



$$\lambda_{1(2)} = \frac{\sigma^2_{2(1)}}{\sigma_1^2 + \sigma_2^2} > 0$$

With large correlated systematics ($\Delta \gg \sigma_{1,2}$)



$\hat{\theta}$ value makes X_1 and X_2 match observations, small pull on θ if Δ is large

$$\lambda_1 < 0$$

$\rho \sim 1 \Rightarrow \theta$ measurement is important ⇒ possibly very different MLE than $X_1 \oplus X_2$..

Uncertainty Decomposition

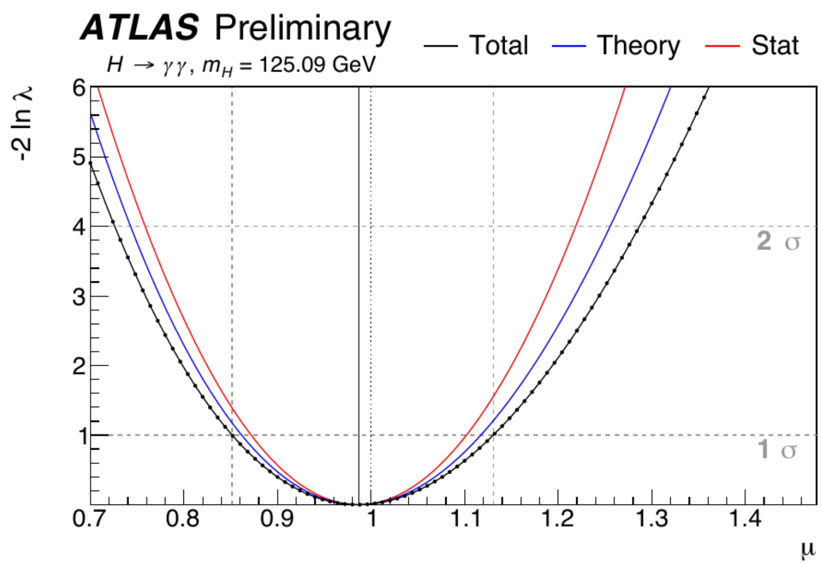
Often useful to break down uncertainties into components (stat + syst, etc.)

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

$$\hat{X} = \sum_i \lambda_i X^{obs,i}$$

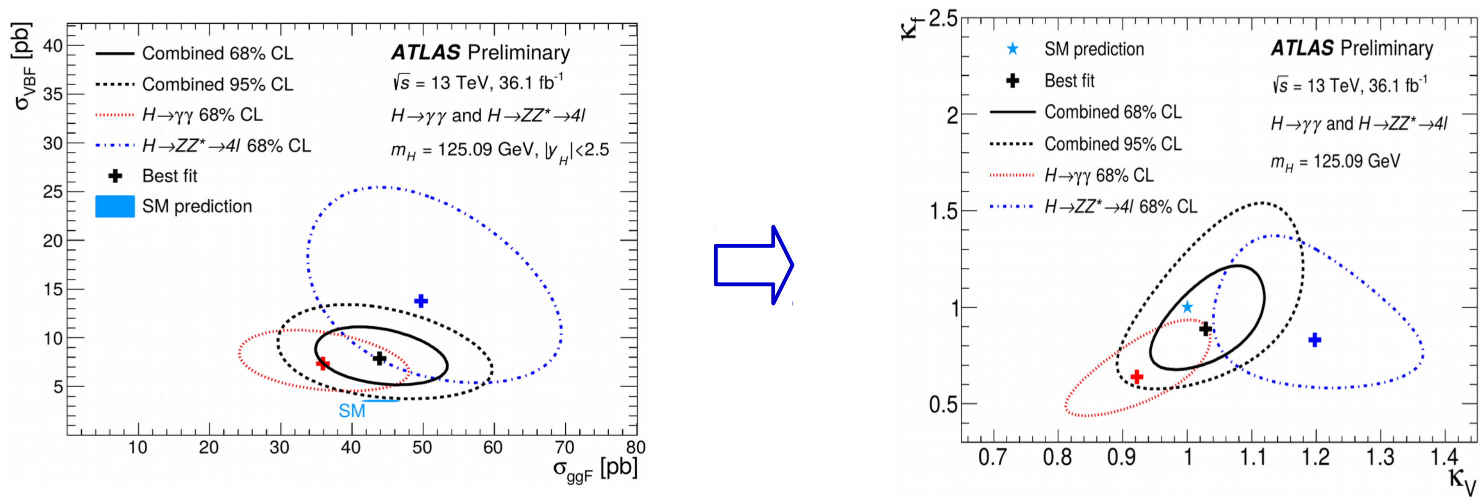
The two methods are not completely equivalent (recently discovered!)
→ In the BLUE case, weights still computed including systematics effects

Presentation of Results

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**

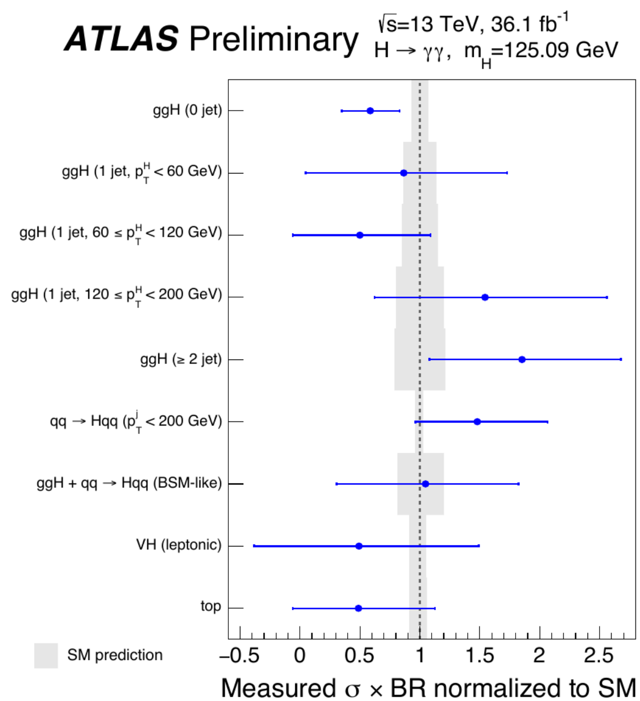
→ However, **often too complex to implement widely:**

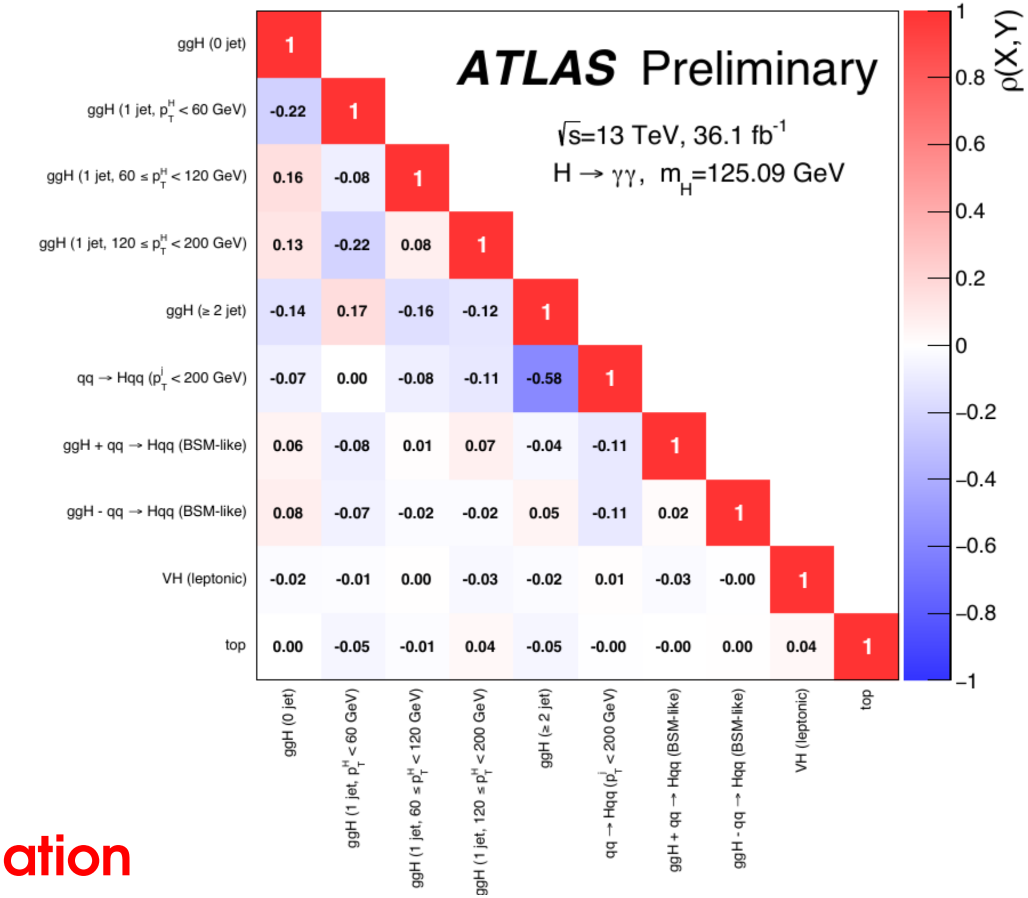
- Full likelihood typically not published
- theorists typically do not want to deal with 4000 NPs...

→ **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**: publish covariance matrices in HEPData, together with the individual measurements





→ **Only valid in the Gaussian approximation**

→ To go further, need some form of **simplified likelihoods**

- **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

- Significant evolution in the statistical methods used in HEP
- Variety of methods, adapted to various situations and target results
- Allow to
  - model the statistical process with high precision in difficult situations (large systematics, small signals)
  - make optimal use of available information
- Implemented in standard RooFit/RooStat toolkits within the ROOT framework, as well as other tools (BAT)
- Improvement and uniformization efforts are still ongoing
- Still many open questions and areas that could use improvement
  → e.g. how to present results with all available information to the "outside"